UNIVERSITAT DE BARCELONA

Modular evolution of domain repeat proteins. Metal-binding and domain repeats of Metallothioneins in molluscs and chordates

Sara Calatayud Robert

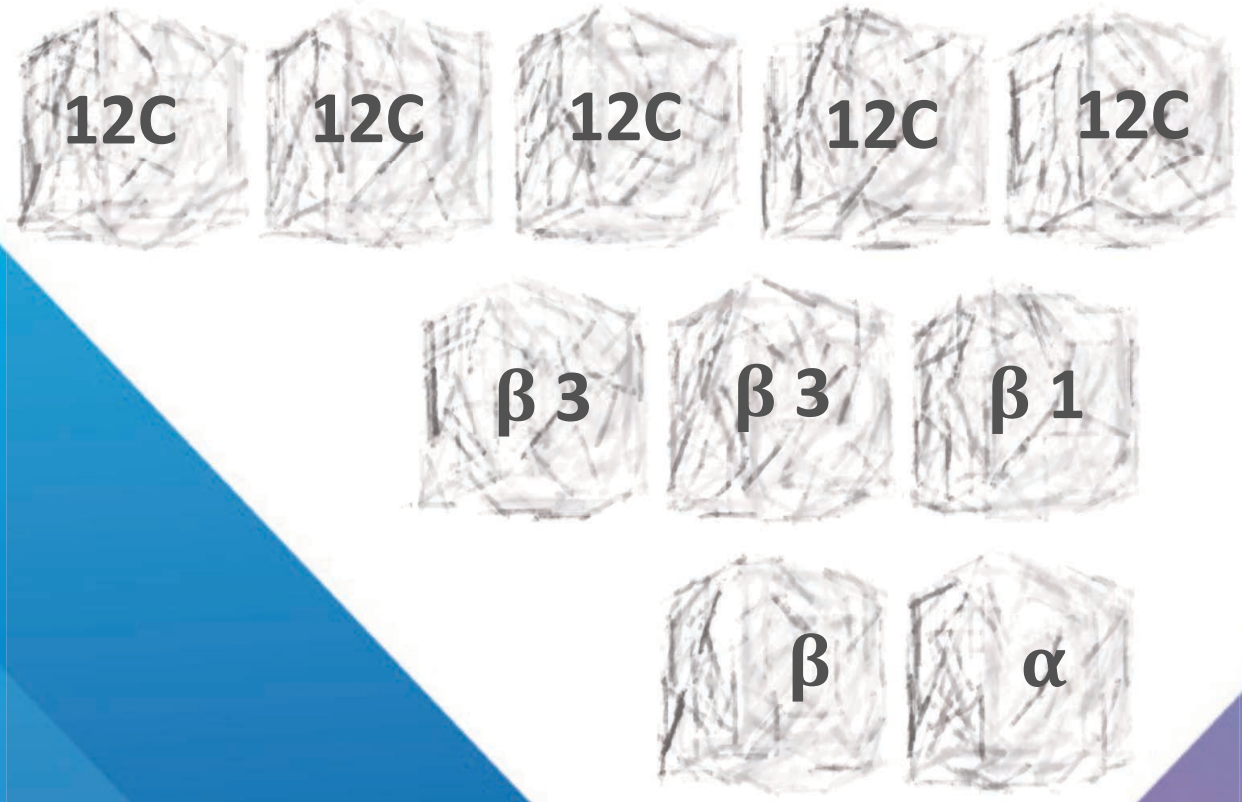
Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento 4.0. España de Creative Commons***.

This doctoral thesis is licensed under the ***Creative Commons Attribution 4.0. Spain License***.

UNIVERSITAT DE BARCELONA

Modular evolution of domain repeat proteins. Metal-binding and domain repeats of Metallothioneins in molluscs and chordates.



2021

SARA CALATAYUD ROBERT

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Biotechnology doctoral program

Modular evolution of domain repeat proteins.

Metal-binding and domain repeats of Metallothioneins in molluscs and chordates

Sara Calatayud Robert

.

Director

Cristian Cañestro Garcia

Director

Ricard Albalat Rodriguez

Tutor

Ricard Albalat Rodriguez

Student

Sara Calatayud Robert

There once were two watchmakers, named Hora and Tempus, who manufactured very fine watches. Both of them were highly regarded, and the phones in their workshops rang frequently — new customers were constantly calling them. However, Hora prospered, while Tempus became poorer and poorer and finally lost his shop. What was the reason?

The watches the men made consisted of about 1,000 parts each. Tempus had so constructed his that if he had one partly assembled and had to put it down — to answer the phone say— it immediately fell to pieces and had to be reassembled from the elements. The better the customers liked his watches, the more they phoned him, the more difficult it became for him to find enough uninterrupted time to finish a watch.

The watches that Hora made were no less complex than those of Tempus. But he had designed them so that he could put together subassemblies of about ten elements each. Ten of these subassemblies, again, could be put together into a larger subassembly; and a system of ten of the latter sub-assemblies constituted the whole watch. Hence, when Hora had to put down a partly assembled watch in order to answer the phone, he lost only a small part of his work, and he assembled his watches in only a fraction of the man-hours it took Tempus."

H. A. Simon, The Architecture of Complexity, 1962

ACKNOWLEDGMENTS

“No journey is too long with the right company”. I have tons of people to thank for in this long journey, but I will try to be short.

First, the instigator, Silvia Atrian my first PhD mentor who encouraged me to do this and settle Metallothioneins in my heart for so many years. Then Ricard, which had the courage to take me “under his wing” and accept this project, and of course, me. I thank him for all the things that I learned from science to life. Sebas, our lab technician, not enough words to thank him, without him this will not be possible at all. Then to Cristian and the rest of the old and former members of lab 5 and 7 for all the help and tips and for the discovery of the “Oiko world” which turns out to be part of mine too. The word “oikos” means house in Greek and that's what I felt all this time with all of you, at home.

Additionally, I want to acknowledge the great work of the UAB group, Mercè, Oscar and Mario as our counterpart in the chemical world of Metallothioneins. Also, to the Innsbruck group: Reinhard, Veronika and Michael for the insights into snails and also for the good scientific and non-scientific discussions. Another special mention is for my work mates of ZeClinics for all their support, encouragement and scientific help.

I want to specially thank all my friends because most of them are not part of this world, but they always rooted for me, gave me faith and the strength to finish this adventure. They are like my shadow always there, but without you seeing it. I am grateful to all: sardines i tonyines, estartiteros, les reines del lab, al Basvé … and much more that I will not have space here.

At the end, one of the most important rock in my life is my family. I want to specially dedicate this work to those that are not here with us anymore who taught me patience, hard work and humility. To the rest of my family, thanks for standing beside me in this long journey and believe in me, this is also yours!

INDEX

ABREVIATIONS

MTs Metallothioneins

GARs Glycinamide ribonucleotide synthetase

AIRs Aminoimidazole ribonucleotide synthetase

GARt Glycinamide ribonucleotide transformylase

SDA Single Domain Arrangement

MDA Multi-domain Arrangement

HMM profiles Hidden Markov models profile

SD Supradomains

IG Immunoglobulin

EGF Epidermal growth factor

Zn Zinc

Fe iron

Cu Copper

Cd Cadmium

Hg Mercury

Pb Lead

C/Cys Cysteine

ICP-AES Inductively Coupled Plasma Atomic Emission Spectroscopy

ESI-MS Electrospray Ionization Time-of-Flight Mass Spectrometry

MTF-1 Metal-responsive Transcription Factor-1

MRE Metal regulation elements

PKC Protein kinase C

GRE Glucocorticoid Response Elements

AP-1 Activator Protein-1

HSE Heat shock elements

ARE Antioxidant response elements

ROS Reactive oxigen species

HSF Heat shock factors

Hsp70 70 kilodalton heat shock proteins

Hsp90 90 kilodalton heat shock proteins

NMR Nuclear magnetic resonance spectroscopy

X-ray X-ray crystallography

Cd(II)-MT/Cd-MT Cadmium methallothionein

Zn(II)-MT/Zn-MT Zinc methallothionein

Cu(I)-MT/Cu-MT Copper methallothionein

SRA Sequence read archive

RNA Ribonucleic acid

NCBI national center for biotechnological information

IPTG isopropylb-D-thiogalactopyranoside

GST Glutatió-S-transferasa

LRR Leucine rich repeats

INTRODUCTION

1. MODULARITY

Modularity is an organizational feature of some systems constructed from the combination of independent units or 'modules', which can be joined and recombined to form larger compositions, often providing the benefit of flexibility and versatility to the system. The concept of modularity can be extended to multiple disciplines, each with their own nuances (Garud et al., 2002). Fields of science, technology, industry, and culture are just some of them. For example, in industrial design, modularity refers to the use of exchangeable parts or options in the fabrication of an object that has been designed and is manufactured with modular components.

1.1 Modularity in Biology

In biology, modularity is an organizational property of biological systems –from ecosystems and organisms to metabolic pathways and developmental networks– in which different modules can be recognized. The concept of modularity in biology was already used in the 18th century, when for instance, Darwin used the *correlation of growth* "I mean by this expression that the whole organisation is so tied together during its growth and development, that when slight variations in any one part occur, and are accumulated through natural selection, other parts become modified. This is a very important subject, most imperfectly understood" (Darwin, 1859), to build up the theory of evolution according to the identification of parts or quantitative repetitive criteria.

Later on, Needham postulated the theory of the "dissociability", in which the biological organization was modular and selection could consequently act on one module without adversely affecting others (Needham, 1933). In 1962, Herbert Simon defined a "complex system" as one that is shaped in a big number of parts that interact in a non-simple way (Simon, 1962). Later on, John Taylor Bonner helped establishing the study of complexity by the number of parts (or units) from which a system is composed (Bonner, 1988). Eventually, Wagner postulated that any biological entity has some degree of modularity and its units are able to independently evolve (Wagner, G.P., 2001).

Evidence that biological systems are modular comes, for instance, from studies related to metabolic pathways and developmental networks (Clune et al., 2013). Metabolic pathways display a hierarchical network organization formed by the assembly of functional modules made of enzymes and substrates that favours a complex web of molecular interactions within a pathway (Peregrín-Alvarez et al., 2009; Ravasz et al.,

1995). Developmental networks are also organized by modules that have a discrete genetic specification, hierarchical organization, interactions with other modules, a particular physical localization within the organism, and the ability to undergo transformations on both developmental and evolutionary time scales (Bolker, 2000; Raff and Raff, 2000). To sum up, at different levels, biological systems are modular, and this organizational trait allows them to combine and/or rearrange modules in several ways without interfering with the rest of the organism. In this way, modularity favours evolvability because facilitates the generation of 'adaptive' genetic diversity, and thereby the evolution by natural selection (Hansen, 2003; Lorenz et al., 2011). Modularity is both the result of evolution as well as a feature that facilitates evolution—an idea that shares a marked resemblance to the work on modularity in technological and organizational fields.

1.2 Protein Modularity

At the molecular level, the structure of proteins also show a modular organization since they are normally built by the combination of two or more 'domains' that can be separated and recombined to form new organizations, behaving therefore as modules (Buchner et al., 2004). A protein domain is a well-defined region of a protein that constitutes a stable, independently folding, compact structural unit, and that might usually perform a specific function (Björklund et al., 2005; Han et al., 2007; Heringa, 2005; Moore et al., 2008; Moore and Bornberg-Bauer, 2012; Toll-Riera and Albà, 2013; Wang and Caetano-Anollés, 2009). For instance, the homeobox domain is a highly conserved 60 amino acid protein domain that function as a DNA binding domain and are found in many transcription factors that control development and cell fate decisions. Domains are therefore the functional and structural constituents of proteins, and according to Structural Classification of Proteins (SCOP) database, there are 68,816 non-redundant domains classified in 5,702 families that group related protein domains with either >30% sequence identity or low sequence identity (e.g. 15%) but performing the same function (Andreeva et al., 2020). Many domain families are shared by Archaea, Bacteria and Eukarya, which has led to conclude they likely are very ancient (Apic et al., 2001). The size of protein domains range from 25 to 200 amino acids (Andreeva et al., 2020), and they can manifest singly (20-35%), but the vast majority (65-80%) are combined in multi-domain proteins (Apic et al., 2001). Many domains in eukaryotic multi-domain proteins can be, indeed, found as single domains of independent proteins in prokaryotes, suggesting that protein domains have once existed as independent proteins. For example, vertebrates have a multi-domain enzyme containing the GAR synthetase, AIR synthetase and GAR transformylase domains. In yeast GARs-AIRs is

encoded separately from GARt, and in bacteria, each domain is encoded separately (Henikoff et al., 1997; Heringa, 2005).

1.2.1 Modular Evolution of Proteins

The modular organization of proteins has important evolutionary implications, because in contrast to point mutations that accumulate at a slow rate and often only result in minor changes in the function of proteins, the recruitment of entire modules by DNA rearrangements can greatly impact on the properties of the protein, favouring a fast evolutionary process of adaptive selection. The evolutionary success of such rearrangements is favoured by the modular organization of proteins because domains might be fused and recombined to efficiently create proteins with new functions or specificities. The main events underlying domain rearrangements are domain fusion and fission, as well as loss of domains, either loss of a single domain or terminal loss of one or more domains (**Figure 1**) (Dohmen et al., 2020). Depending on the groups of organisms (i.e. animals, fungi and plants), each one of these events accounts for ≈10% to ≈64% of the relative contribution of the rearrangement events. Together, these events are sufficient to explain a large proportion (>95%) of the arrangements occurred in multi-domain proteins generated through shuffling of different domains. This combinations has been considered crucial for example in the evolution of the extracellular matrix of the metazoans as well as in the blood coagulation cascade (Cromar et al., 2014; Patthy, 1985). In sharp contrast, domain emergence (**Figure1**), either single domain emergence or terminal domain emergence, little contributes to the domain rearrangements (≈4% and ≈1% at most, respectively). It seems therefore that functional novelty in modular proteins is primarily generated by the four main rearrangements listed above, as opposed to the formation of novel domains.

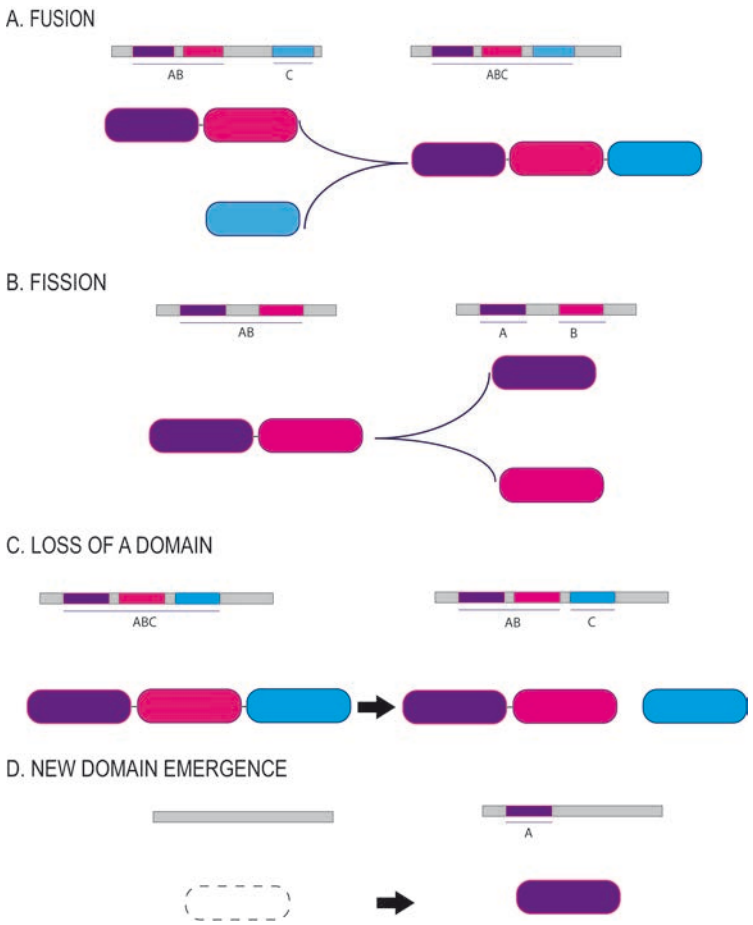


Figure 1. Schematic representation of domain rearrangements. Genome structure represented in grey in which coloured bars code for a domain (purple, blue and pink). Letters code for a protein structure of one (one letter), two (two letters) or tree (three letters) domain protein structure rearranged by different events: (A) Fusion; (B) Fission; (C) Loss of a domain; and (D) New domain emergence.

1.2.2 *De novo* Emergence of Domains

There is however evidence that rearrangements of domains do not strongly alter arrangement functionality. These mechanisms alone are not enough for certain evolutionary innovations such as the required in the wake of environmental shifts, instead, this may be facilitated by the emergence of novel domains. Emergence of new domains, albeit rare, would be therefore the result of functional challenges not met by modular rearrangements. But how do novel domains appear from “scratch”? Despite it remains unclear, Klasberg et al. proposed two main mechanisms; a) *de novo* emergence from a previously non-coding DNA that would lead to a single domain arrangement (SDA); and b) the extension of the coding sequence leading to a novel terminal domain in a multi-domain arrangement (MDA) (Klasberg et al., 2018) (**Figure 2**).

The emergence of new domains would have a great adaptive value and, accordingly, it has frequently happened in the context of abiotic stress, biotic defense, reproduction, and development processes (Moore and Bornberg-Bauer, 2012). For instance, the origin of hair and fur in mammals and the adaptation and restructuration of their skin have been associated to the emergence of several 'keratin-associated' domains (Dohmen et al., 2020). The skin acts as a first barrier against environmental damage and pathogen infestation and it would be therefore related to biotic defense. Also related to biotic defense, the evolution of mammalian immune system has also been associated with the emergence of new domains, particularly with the emergence of the 'Interleukin' and the 'Lymphocyte activation family X' domains (Dohmen et al., 2020).

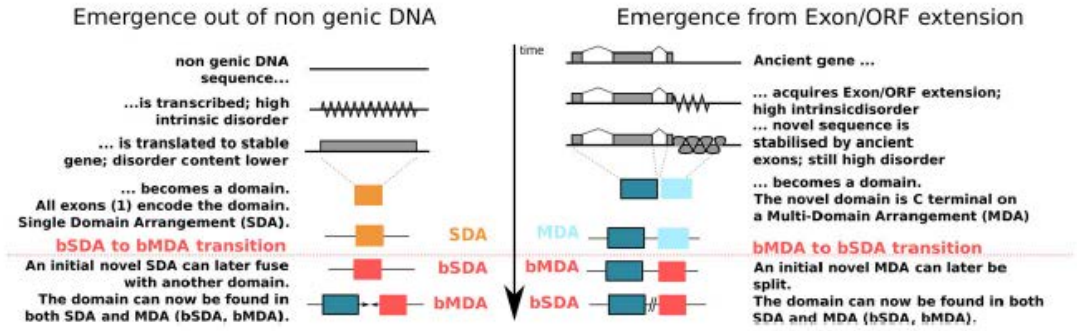


Figure 2. Possible mechanisms of de novo protein domain emergence by Klasberg et al. a) de novo emergence from a non-coding DNA; b) Extension of the coding sequence to produce a multi-domain protein.

De novo evolution of domains or proteins has been considered rare. However, recent analyses have shown that *de novo* evolution is more frequent than previously thought (Levy, 2019; Neme and Tautz, 2013; Toll-Riera and Albà, 2013; Weisman and Eddy, 2017). Studies of human proteins revealed more than 400 "young" domains, 164 of which are found combined with older ones and preferentially located at the N-terminus of the proteins (Toll-Riera and Albà, 2013). These new domains are rich in low-complexity sequences (Toll-Riera et al., 2012) and tend to be structurally disordered (Moore and Bornberg-Bauer, 2012). In fact, new domains appear to share some characteristics: (1) they are poorly annotated in the genomes; (2) they tend to be short in average and (3) they evolve particularly rapidly (Toll-Riera and Albà, 2013). Novel domains are more difficult to identify in genomes since algorithms (e.g. HMM profiles) that check for already known domains will hardly identify them. They seem to be short but tend to present more flexible and charged protein ends that may help the fusion with an "old" domain to produce large and more specialized proteins. Finally, novel domains seem to evolve rapidly because their non-synonymous to synonymous substitution rate is higher than older ones (Toll-Riera and Albà, 2013).

1.2.3 Domain Repeat Proteins

Multi-domain proteins might result from rearrangements of domains from different domain families, or from domains of the same family repeated in tandem. While the former evolve by recombination and fusion processes, and have been extensively analysed (Apic et al., 2001; Björklund et al., 2005; Dohmen et al., 2020), the latter seems to mostly evolve by internal duplications, and they deserve to be considered separately (Apic et al., 2001; Björklund, 2006).

Domain repeat proteins represent a small proportion, 10-20%, of the proteomes (Han et al., 2007; Mary et al., 2015; Reshef et al., 2010), and they are present across all tree of life, especially in eukaryotes. They consist of two or more adjacent domains of the same domain family with less than 30 residues between them (Apic et al., 2001). The number of domain repeats is highly variable between proteins (Björklund, 2006; Moore and Bornberg-Bauer, 2012; Schüler and Bornberg-Bauer, 2016), and even orthologous proteins might have different numbers of repeats depending on the species. For example, the spindle-like microcephaly-associated proteins consist of 71 IQ calmodulin-binding domains in human, but 62 in mouse and rat, 63 in zebrafish, 53 in chicken, 22 in the fruit fly and only 6 in worms (Björklund, 2006).These differences are indicative of gain and/or loss of domains through evolution (Andrade et al., 2001a).

The mechanism for repeat expansion is not fully understood, but nonhomologous recombination leading to internal domain duplications is thought to be the main process (Björklund, 2006). Comparative analyses of sequence similarity between domains may thereby provide information about duplication dynamic. These analyses have revealed, for instance, that in many proteins a group of domains is duplicated in tandem at a time, and these groups behave as evolutionary modules known as “supradomains” (SD) (Moore et al., 2008). For example, the chicken nebulin protein (an actin-binding protein from the skeletal muscle) has duplicated a SD of seven domains at a time, and similar patterns are seen in most nebulin proteins (Björklund et al., 2010). Similarly, C2H2 zinc fingers (responsible to bind proteins to DNA) appear also to have mainly duplicated through SD of two, three or four domains. In contrast, in other proteins, domains show the highest similarity to the neighbouring domain (and the similarity then decreases with distance), which have been taken as a proof that duplication involved one domain at a time is the most likely scenario for their expansion. Immunoglobulin (IG) and the epidermal growth factor (EGF) domains exemplify expansions by a duplication of one domain at a time. It is therefore clear that the size of the duplicated unit in domain repeat proteins might vary greatly, from one up to as many as nine domains at a time (Björklund, 2006). It is still unknown, however, what determines the distinct patterns of domain

repeats, whether there is a preference for duplication of certain sizes due to functional constraints, or whether domain expansion is a random process or a controlled mechanism where specific segments are selectively duplicated.

2. METALLOTHIONEINS

Although in the last decades the use of new computational tools and large proteome databases has improved our understanding of the domain repeat proteins (Andrade et al., 2001a; Björklund, 2006; Moore and Bornberg-Bauer, 2012; Schüler and Bornberg-Bauer, 2016), still many challenges lie ahead before their evolution can be fully understood. Under this perspective, extensive analyses of proteins with well-defined functional domains that are experimentally tractable, and that can be easily recognized in wide phylogenetic contexts may contribute to better understand the origin, dynamics and relevance of modularity in protein evolution. Metallothionein proteins fulfill these criteria.

Metallothioneins (MT) are a group of metal-binding proteins first discovered in mammals by Marghose and Valle in 1957 (Margoshes and Vallee, 1957). MTs are modular proteins made of a variable number of tandem domain repeats of the 'metallothionein domain' family. Their modular organization is functionally and structurally relevant because the domains determine the metal-binding capacity and preference of each MT. The origin, function, organization and diversification of 'metallothionein domains' are therefore key for understanding the functional evolution of MTs.

2.1. Metallothionein function

Heavy metals such as zinc (Zn), iron (Fe) or copper (Cu) are essential for several biological processes, but toxic at high concentrations, while others such as cadmium (Cd), mercury (Hg) or lead (Pb) are highly poisonous even at low concentrations. Living beings have different physiological mechanisms to control the homeostasis of essential metals as well as to counteract the harmful effects of the non-essential ones. One of these mechanisms is based on MTs, a group of metal-binding proteins classically considered a diverse family of Cys-rich (up to 30% content) and low molecular weight (<100 amino acids) proteins found in almost all organisms (reviewed in(Calvo et al., 2017; Capdevila and Atrian, 2011)). Their richness in cysteine (Cys, C) residues confers them the capability of binding metals via the formation of metal thiolate clusters.

MTs have been shown to regulate Zn homeostasis through a direct interaction with different groups of proteins like mitochondrial aconitase (m-aconitase) that accepts zinc

ions from the metal loaded Zn(II)-MT of mice (Feng et al., 2006) or the zinc finger transcription factor Sp1 that accepts or gives Zn to rabbit MT (Zeng et al., 1991; Zhang et al., 2003). For Cu, MTs are also capable to maintain the homeostasis. For example, in mammals the ATP7A and B Cu-transporting ATPases are responsible for the homeostatic response against Cu accumulation by exporting the excess of Cu to plasma membrane or lysosomes. In the absence of this ATP7A, MTs can step up and protect cells of Cu-overload by sequestering it. Also, metallothioneins can use this sequestration capacity to withheld Cu from essential metabolic processes when the extracellular Cu is limiting to maintain cell homeostasis (Gudekar et al., 2020).

MTs have been also associated with metal detoxification. In mammals, for instance, it has been reported that MT-null mice are highly susceptible to Cd exposure producing liver injuries, and eventuality, lethality. In contrast, when wild-type animals are exposed to Cd, constitutive expression of MT provides protection against Cd toxicity, denoting the protective role of MTs (Liu et al., 1996, 1995). Similarly, the gastropod *Pomatia elegans* has two MTs, PeMT1 and PeMT2, that are upregulated when exposed to Cd ions “in vivo”. PeMT2 is related to short term Cd exposure, whereas PeMT1 is expressed when there is a chronic exposure to Cd. The levels of expression of these MTs have been shown to be good bioindicators for the presence of Cd in the terrestrial habitats (Schmielau et al., 2019).

MTs bind with different preference a wide range of metal ions, being Zn, Cu and Cd the most habitual partners *in vivo* (Blindauer, 2014, 2010). The metal-binding preference of a given MT, however, cannot be directly predicted from its amino acid sequence, and metal-binding assays have to be performed to characterize newly discovered forms. The most usual metal-binding assays are based on the recombinant expression of the MTs in *E. coli* growth in media supplemented with metals, typically with Zn(II), Cd(II) and Cu(II) salts, in order to produce metal-MT complexes. These complexes are purified and analysed by inductively coupled plasma atomic emission spectrometer (ICP-AES) and electrospray ionization mass spectrometry (ESI-MS). ICP-AES is an analytical technique that allows protein quantification and metal-to-protein stoichiometry determination through the measurement of element composition of the samples (Zn, Cd, and Cu) (Bongers et al., 1988). ESI-MS is used to determine the molecular mass of the species formed, that is, the speciation of the samples (Capdevila et al., 2012). This is a well-established procedure for MT analysis that determines the metallated species in the complexes from which is possible to infer the metal preference and capacity of newly discovered MTs (or MT domains). With this procedure the metal-binding features of many MTs have been determined and used to categorize the MTs in a stepwise

gradation, from genuine Zn/Cd-thioneins to genuine Cu-thioneins (Bofill et al., 2009; Capdevila and Atrian, 2011; Palacios et al., 2011). The genuine Zn-thioneins being those who (i) render a unique species when are synthetized as Zn(II)-MT complexes, (ii) mixed Zn, Cd(II)-MT species when are synthetized with Cd, and (iii) and yield a mixture of heteronuclear Zn(II), Cu(I)-MT species in presence of Cu. In contrast, genuine Cu-thioneins (i) render unique homonuclear Cu(I)-MT species in the presence of that metal, (ii) contain sulphide ligands in the Cd(II)-MT preparations, and (iii) produce a mixture of Zn species in the presence of Zn(II) (Bofill et al., 2009; Palacios et al., 2011).

Besides metal binding, MTs have been also associated to other physiological functions like radical scavenging, oxidative stress protection or antiapoptotic defense ((Capdevila and Atrian, 2011), reviewed in (Capdevila et al., 2012)). For instance, it has been demonstrated that MT provides protection in ischemia-reperfusion injury (tissue damage when blood supply returns to the heart after a lack of oxygen) in the heart. Studies suggest that oxidative stress plays a major role on producing this type of injury by the generation of reactive oxigen species that produces myocardial damage. MT overexpression in the heart of transgenic mice provides protection against the ischemia-reperfusion injury (Kang et al., 1999).

2.2 Regulation of Metallothionein expression

In agreement with its biological function, metals upregulate MT expression. For instance, injection of $CdCl_2$ salts in different adult rat tissues produced an increase in the MT levels in the liver (Onosaka et al., 1984), and similar results of activation of *MT* expression due to metal exposure have been observed in many other species such as *Mytilus galloprovincialis*, *Clethrionomys glareolus* and *Tetrahymena borealis* (de Francisco et al., 2016; Mikowska et al., 2018; Zorita et al., 2007). This transcriptional activation is mediated by the Metal-responsive Transcription Factor-1 (MTF-1), a protein with six zinc-finger domains and several trans-activation domains conserved from mammals to insects (Giedroc et al., 2001; Zhang et al., 2003) (**Figure 3**). Intracellular increase of Cd(II) ions displace Zn(II) from MTs, which leads to an increase of the intracellular concentration of free Zn(II). The displaced Zn(II) ions are picked up by MTF-1, which is then translocated to the nuclei of the cell and binds to Metal Regulation Elements (MRE) in the promoter region of *MT* genes (**Figure 3**). MRE are conserved seven-nucleotide sequences (TGCRCNC) present in promoter regions of nearly all *MT* genes. MRE have been experimentally and bioinformatically identified in proximal *MT* promoters of most metazoans, from insects and molluscs to fish and mammals ((Auf der Maur et al., 1999; Egg et al., 2009; Michael Niederwanger et al., 2017; Silar et al., 1990; Westin and Schaffner, 1988; Zafarullah et al., 1988) reviewed in (Giedroc et al., 2001)).

Besides activation modulated by Zn ions, MTF-1 seems to interact with more than one signal for its transcriptional activation through MRE. Several studies have shown that MTF-1 is phosphorylated, and in response to metal exposure, the phosphorylation level increases. One of the major intracellular mediator of this signal capable of phosphorylating is the protein kinase C that shows several consensus sites in the MTF-1. This is supported by the observation that several kinase inhibitors block or attenuate metal inducible metallothionein transcription. Therefore, another possible way to activate the MTF-1/MRE pathways is by means of phosphorylation controlled by different transduction signals that involve PKC, casein kinase II, tyrosine kinase, and calcium (Isani and Carpenè, 2014; Lichtlen and Schaffner, 2001; Saydam et al., 2002; Yu et al., 1997). These different mechanisms to activate the MTF-1/MRE pathway to activate metallothionein transcriptions indicates an intriguing complexity of the metallothionein metal transcription process that is still worth pursuing.

Transcription of *MT* genes is not only regulated by the MTF-1/MRE system, but other transcriptional regulatory elements are located in the *MT* promoter regions, including, Glucocorticoid Response Elements (GRE), elements for the Activator Protein-1 (AP-1), Heat Shock Elements (HSE) and Antioxidant Response Elements (ARE) (Ryvolova et al., 2011; Saydam et al., 2002; Sigel et al., 2009) (**Figure 3**). For example, reactive oxygen species (ROS) activate *MT* transcription by promoting the binding to AREs of different transcription factors such as nuclear factor erythroid 2-related factors 1 and 2 (Nrf1, Nrf2) (Haq, 2003; Raghunath et al., 2018; Sabolić et al., 2010). Heat shock stress also activate *MT* expression by phosphorylation of the Heat Shock Factor (HSF) on serine and threonine residues. In *Saccharomyces cerevisiae* Cup1 metallothionein, the MT gene is activated in response to heat shock through HSF and HSE_{Cup1} (**Figure 3**) but this response is transient since this system is thought to be promptly regulated by a feedback mechanism in which the heat shock proteins (hsp) increase and interact with HSF to inhibit the activity (Silar et al., 1991; Tamai et al., 1994).

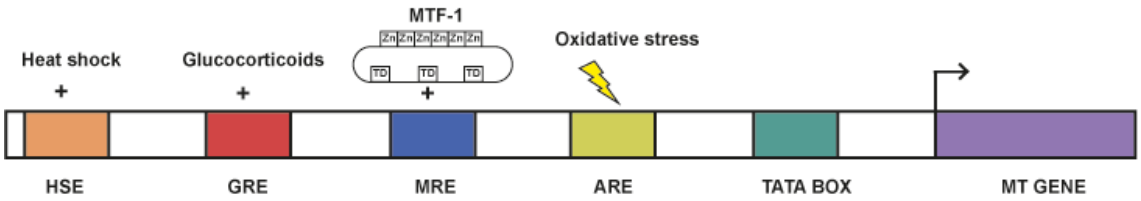


Figure 3. Metallothionein gene regulation scheme. The MT promoter has many response elements that upregulate transcription: (1) metal response elements (MRE), which are activated by the metal-responsive transcription factor (MTF-1) with six zinc finger domains and several trans-activation domains (TD); (2) glucocorticoid response elements (GRE); (3) the antioxidant response element (ARE), activated in response to redox status; and (4) the Heat shock element (HSE) inactivated by an increment in heat shock proteins.

2.3 Metallothionein Structure and Domain Organization

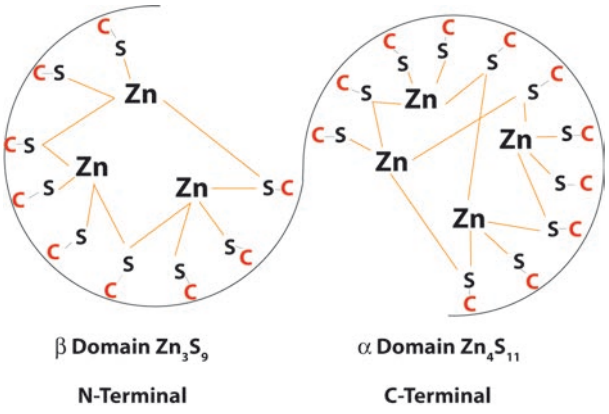


Figure 4 Schematic representation of the Mammalian MT β-domain and α-domain structure. The β-domain composed of 9 cysteines capable of binding 3 Zn(II) metal ions and the α-domain with 11 cysteines capable of binding 4 Zn(II) metal ions.

MTs are rich in Cys residues (reviewed in (Calvo et al., 2017; Capdevila and Atrian, 2011)) that are arranged in CxC, CC, and CCC motifs (Frey, 2014; Hynes, 2003).The number and distribution of motifs led to the original definition of two functional domains in vertebrate MTs (Braun et al., 1986): an α domain (with 11/12-Cys at the C-terminal region) and a β domain (with 9-Cys at the N-terminal region), joined by a short linker sequence. Vertebrate MTs, as well as many other MTs, are therefore bi-domain proteins made of two non-symmetric domains, each one capable to independently coordinate a number of metal ions through the formation of metal–thiolate clusters: vertebrate β domain coordinates three divalent metal ions, while the α domain clusters four divalent metal ions (Blindauer, 2014; Otvos and Armitage, 1980; Schultze et al., 1988; Winge et al., 1984; Winge and Miklossy, 1982) (**Figure 4**). Noteworthy, in absence of metals, MT proteins have a pronounced disordered organization and the so-called apo-MTs, that lack a defined tertiary conformation (Capdevila and Atrian, 2011; Juárez-Rebollar et al., 2017). The NMR or X-ray diffraction analyses used to determine the 3D organization of the MTs require, therefore, MTs bound to metal ions (Škutková et al., 2010; Ziller and Fraissinet-Tachet, 2018).

The 3D organization of metal clusters of other bi-domain MTs has been resolved. In crustaceans, for instance, *Callinectes sapidus* and *Homarus americanus* MTs have two domains similar to the vertebrate β ones that coordinate 6 divalent metal ions forming two M^{II}_3Cys9 metal clusters (M^{II} stands for divalent metal ions) (Muñoz et al., 2002; Narula et al., 1995). In gastropods, the *Helix pomatia* MT and the *Pomatia elegans* MT2 exhibit also a bi-domain organization with each 9-Cys domain binding also three divalent metal ions (Beil et al., 2019; Schmielau et al., 2019). Interestingly, some gastropod MTs show a tri-domain organization. *Pomatia elegans* MT1 and *Littorina littorea* MT have three 9-Cys domains, each one forming M^{II}_3Cys9 metal clusters (Baumann et al., 2017; Schmielau et al., 2019).

In contrast, other MTs have a monodomain structure. The cyanobacterial SmtA MT from *Synechococcus,* for example, forms a single metal cluster of nine cysteines and two histidines that coordinates four Zn(II) metal ions, resembling the organization of a vertebrate α domain. This is a special case where the histidine residues have been demonstrated to help in the coordination of the metal ions (Blindauer et al., 2001). Other examples of single domain metallothioneins are those of fungi. *Saccharomyces cerevisiae* Cup1 MT binds seven Cu ions forming a single cluster with ten cysteines (Calderone et al., 2005; Peroza et al., 2009; Peterson et al., 1996; Romero-Isart and Vasák, 2002), and *Neurospora crassa* MT is able to bind six metal Cu ions in a single domain cluster with seven cysteines (Calderone et al., 2005; Peroza et al., 2009).

2.4 Metallothionein Evolution

The first attempt to classify MTs was made by Fowler and Kojima (Fowler et al., 1987; Kojima et al., 1999). They grouped MTs in three classes based on their similarity with the first MT discovered, the mammal MT, but this classification became insufficient as more MTs were discovered. Kägi and Binz proposed a new classification based on sequence similarity, which grouped MTs in 15 families (Binz and Kägi, 1999a), and these further divided in subfamilies, subgroups and isoforms (Binz and Kägi, 1999a; Blindauer, 2014; Palacios et al., 2011). In this classification, the evolutionary relationships between families were unclear, implicitly meaning that the so-called “metallothioneins” would not have evolved from a single common ancestor but they would be polyphyletic, arising from independent events of *de novo* evolution at least in clades of high taxonomic ranks, p.e. kingdoms or phyla (Blindauer, 2014; Capdevila and Atrian, 2011; Isani and Carpenè, 2014; Ziller and Fraissinet-Tachet, 2018). In contrast, at low taxonomic ranks, MTs seem to have followed typical evolutionary patterns based on progressive accumulations of changes in the sequence, and on processes of duplications, losses or rearrangements of internal domains or genes. These processes would have led to the evolution of MT multiplicity and of large multi-domain forms in some current animal species.

2.4.1 Evolution of MT Multiplicity

Multiple MTs have been observed in many animal lineages. In vertebrates, for instance, there are four MT types, MT1, MT2, MT3, and MT4, and phylogenetic and synteny analyses have shown the existence of mammalian MT1s and MT2s, amniote (mammals, birds and reptiles) MT4s, tetrapod MT3s, and fish MTs (Serén et al., 2014). Even within mammals, there are significant differences of multiplicity between species: humans have 11 *MT* genes and six pseudogenes (Mao et al., 2012; Stennard et al., 1994; West et al., 1990), while mice only have four *MT* genes, one for each MT type

(Juárez-Rebollar et al., 2017). Multiplicity has also been described in many invertebrate species. *Drosophila melanogaster*, for instance, has six genes that code for the MtnA, MtnB, MtnC, MtnD, MtnE and MtnF proteins (Luo et al., 2020). In molluscs, MT multiplicity has been detected in Bivalvia, from two (*Ostrea edulis*) to ten (*Crassostrea virginica*) MTs (Jenny et al., 2016), and in Gastopoda, from two (*Pomatia elegans*)(Schmielau et al., 2019) to four (*Alinda biplicata*) MTs (Pedrini-Martha et al., 2020).

MT multiplicity has been associated to neofunctionalization or subfunctionalization processes in which different paralogs have distinct metal-binding preferences and/or expression patterns. Mammalian MT1 and MT2 are ubiquitous metal-induced forms with preference for Zn(II), whereas MT3 and MT4 are restrictedly expressed and have intermediate Zn/Cu and Cu-thionein characters (Artells et al., 2013; Capdevila and Atrian, 2011; Vašák and Meloni, 2017). In the Heterobranchia clade of Gastropoda, the snails *Cepaea hortensis*, *Cantareus aspersus* (*Helix aspersa*) and *Helix pomatia* have three MTs, each one with distinct metal selectivity –that is, Cd(II)-MT, Cd(II)/Cu(I)-MT, and Cu(I)-MT– and differential transcription patterns. During colonization processes, animal linages have adapted to different environmental conditions –in the case of gastropods from marine to terrestrial or freshwater habitats– coping with fluctuating situations of metal bioavailability (Baumann et al., 2017; Beil et al., 2019). MT multiplicity may have favoured these processes allowing each paralog to be optimized by changes in the coding region affecting residues not directly involved in metal coordination, but that may influence the 3D structure , and thereby, the metal preference of the protein (Blindauer et al., 2001; Pedrini-Martha et al., 2020; Pérez-Rafael et al., 2014). Changes might also affect the regulatory regions providing distinct expression patterns of the paralogs in different organs or tissues. In the *Helix* genus of gastropod, for example, two metallothioneins display different tissue expression patterns. The *Helix* Cu-MT is exclusively expressed in rhogocyte cells, whereas the Cd-MT is expressed in the midgut gland (Chabicovsky, 2003; Dallinger et al., 1997; Pedrini-Martha et al., 2021).

2.4.2 Evolution of Multi-domain MTs

While MT multiplicity has been recognized long time ago, it had been classically assumed that MTs were small proteins made of one or two domains. Only recently, large multi-domain MTs have been described in some mollusc, ciliate and fungi species (de Francisco et al., 2016; Iturbe-Espinoza et al., 2016; Palacios et al., 2014). In Mollusca, bivalve *Crassostrea virginica* has a multi-domain MT, whose α domain might be tandem repeated up to four times (CvMT-IIG,H) (Jenny et al., 2016). Gastropod *Littorina littorea* and *Pomatias elegans* have a tri-domain MTs(Palacios et al., 2017; Schmielau et al.,

2019), whereas *Alinda biplicata*, a terrestrial door snail, have two large multi-domain MTs with nine and ten domain repeats (Pedrini-Martha et al., 2020). It is thought that the evolution of multi-domain MTs in molluscs would have provided these animals the ability to cope with more metal ions, which may have been important for adaptation to stressful conditions in challenging environments exposed by all this species (Dvorak et al., 2019; Jenny et al., 2016).

In ciliates, species of the *Tetrahymena* genus show different multi-domain architectures in their MTs involved in Cd detoxification. These *Tetrahymena* species have MTs with a 17-Cys domain repeated two (9 MTs), three (13 MTs), four (3 MTs) or five (1 MT) times (de Francisco et al., 2016). In Fungi, *Cryptococcus neoformans* and *Tremella mesenterica* have also multi-domain MTs made of a 7-Cys domain tandemly repeated three, four and eight times in CnMT1, CnMT2 and TmMT, respectively. These multi-domain MTs have been considered virulence factors in pathogenic fungus *C. neoformans*, or for Cu-handling for lignin-metabolizing enzymes in saprophytic fungus *T. mesenterica*. Again, these multi-domain MTs may have been a solution for creating large MTs with high capacity of metal binding (Iturbe-Espinoza et al., 2016; Michael Niederwanger et al., 2017; Palacios et al., 2017).

2.5 Metallothionein as a case study of domain repeat proteins

The discovery of multi-domain MTs made these proteins become a good model to study the modular evolution of domain repeated proteins. This was because (1) MT domains could be predicted based on the arrangement of the Cys motifs, and thereby, changes in the domains and in their modular organization could be easily recognized; (2) the functional analysis of multi-domain MTs was experimentally affordable from the characterization of the biochemical features of recombinant produced metal-MT complexes; (3) the functional characterization of single or diverse combinations of domains was also possible thanks to the genetic engineering techniques of recombinant protein expression; and (4) the sequence characteristics of MTs favoured the identification of new MTs and novel domains by *in silico* surveys of genomic and transcriptomic databases, even from unassembled or incomplete projects, opening the possibility of identify MTs in large numbers of species.

From all possible organisms, two phylogenetic groups appeared the most suitable candidates to investigate novel MTs, either ‘classical’ small MTs or ‘unusual’ large multi-domain MTs: Molluscs, particularly gastropods, and Chordates, especially tunicates.

2.5.1 Studies of MTs in Gastropod Molluscs

The Mollusca phylum encompasses a great diversity of animals, with over 85,000 extant species, with only Arthropoda having more species among animal phyla (**Figure 5**). Molluscs are the largest marine phylum with species with very different life cycles, size, body plans and habitats, including gastropods such as snails, slugs and limpets, bivalves such as clams, mussels and oysters, cephalopods such as octopuses and squids, and polyplacophores as the 'primitive' chitons. For millennia, molluscs have been a source of food for humans, as well as important luxury goods, although some species of molluscs have become agricultural pests. Molluscs are fundamental for marine ecosystem as grazing on algae or as filter feeders as well as a source food for other animals, and they are considered excellent ecological indicators, having been assayed as potential biosensors for monitoring marine contaminants such as heavy metals and other environmental stressors (Tran et al., 2007, 2004).

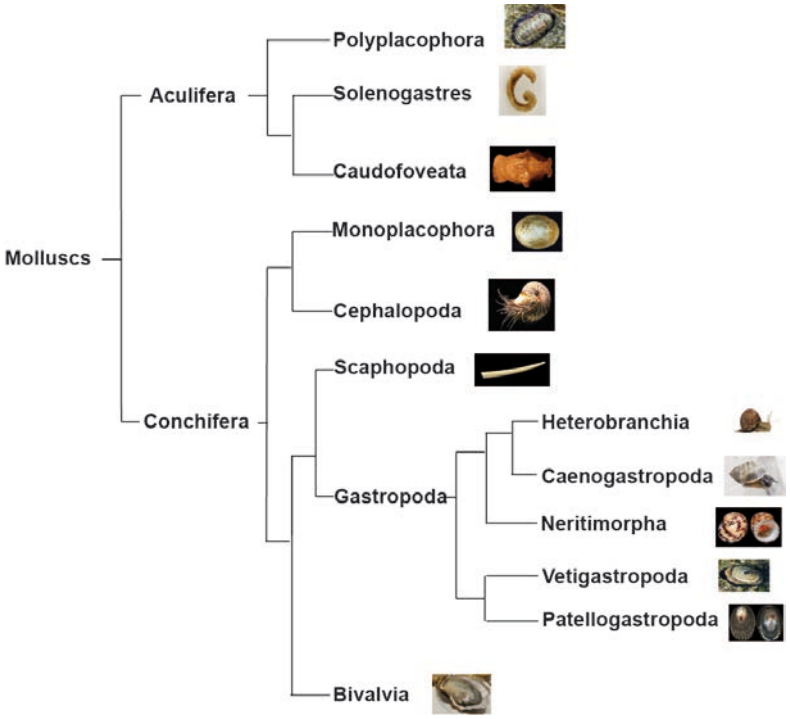


Figure 5. Mollusca phylogeny. Molluscs are divided in aculifera and conchifera clades. The aculifera is divided in the Polyplacophora, Solenogastres, and Caudofoveata classes while the conchifera has Gastropoda, Bivalvia, Cephalopoda, Monoplacophora, and Scaphopoda. The Gastropoda class is also divided in Vetigastropoda, Caenogastropoda, Heterobranchia, Neritimorpha and Patellogastropoda orders.

Although the great attention this organisms have been given, the evolutionary and phylogenetic relationships within the Mollusca phylum are not totally resolved. The most accepted classification divides the phylum into two major clades: Aculifera, which includes those groups that have no conch or shell such as Polyplacophora (a.k.a. Loricata), Solenogastres (a.k.a. Neomeniomorpha), and Caudofoveata (a.k.a. Chaetodermomorpha) classes; and Conchifera, which comprises the shell-bearing classes Gastropoda, Bivalvia, Cephalopoda, Monoplacophora, and Scaphopoda (Kocot et al., 2018).

Gastropoda, with about 400 living families and 40.000 species, is the largest class, and constitutes a substantial part of the biodiversity in the oceans with essential ecological functions in aquatic systems, both as grazers or preys for higher trophic levels. Gastropods possess a single shell and an operculum, although some linages have lost or reduced the shell (Zapata et al., 2014). This class also comprises snails and slugs from freshwater and from land, so they would have colonized nearly every habitat on the planet facing diverse environmental conditions, including different situations of metal bioavailability. From an experimental point of view, 20 gastropod genomes have been published so far, from a total of 39 mollusc genomes (https://www-ncbi-nlm-nih-gov.sire.ub.edu/genome/?term=gastropoda), as well as a large number of transcriptomic projects (1836 SRA of gastropod RNA, 5424 for mollusc species in NCBI), facilitating the *in silico* identification of genes in wide sets of species.

Gastropods are classified in five subclasses: Patellogastropoda, Vetigastropoda, Neritimorpha, Caenogastropoda and Heterobranchia (Cunha and Giribet, 2019; Zapata et al., 2014). Several gastropod MTs had been investigated before this Thesis, but these studies had an uneven phylogenetic distribution. MT identification and metal-binding selectivity had been determined for 12 MTs from five heterobranchia species (with the most extensive studies) (Babenka et al., 2011; Beil et al., 2019; Berger et al., 1995; Hispard et al., 2008; Pedrini-Martha et al., 2020), three MTs from two caenogastropod species (Palacios et al., 2017; Schmielau et al., 2019), and one vetigastropod MT (Lieb, 2003; Pérez-Rafael et al., 2012b). These studies showed that gastropod MTs were bi-domain MTs, with the exception of the multi-domain MTs found in *L. littorea*, *P. elegans* and *A.biplicata* (Palacios et al., 2017; Pedrini-Martha et al., 2020; Schmielau et al., 2019). Gastropod MTs have different metal preferences. Marine and terrestrial species such as *Littorina littorea* and *Alinda biplicata* have Cd-MTs, whereas freshwater snails such as *Lymnaea Stagnalis and Physella acuta* have unspecific MTs (Babenka et al., 2011; Dvorak et al., 2018; Martínez-Paz et al., 2017; Pedrini-Martha et al., 2020). Additionally, some snails possess specialized Cu-MTs for the turnover of hemocyanin, a Cu-metalloprotein responsible of the oxigen transport in these species (Pedrini-Martha et al., 2021, 2020).

The studies of mollusc MTs are highly biased. While MTs have been analysed for a number of species of the Gastropod and Bivalvia classes, nothing is known about the MTs of three classes of the Conchifera clade, nor of the three classes of the Aculifera clade. Even in the Gastropoda class, most of the MT knowledge comes from species of the Heterobranchia subclass, while there is little or no information of MTs from the other four subclasses. To amend this situation and be able to investigate the origin of

gastropod MTs and the emergence of the multi-domain forms, it was necessary to extend the MT studies to all gastropod subclasses. In addition, to provide a wide evolutionary perspective, it was also necessary to identify the MTs in all mollusc classes, from which to be able to recognize common patterns of modular organization across the Mollusca phylum.

2.5.2 Studies of MTs in Tunicate Chordates

The phylum Chordata embraces three subphyla – Vertebrata, Cephalochordata and Tunicata (a.k.a Urochordata) that share five unique features at some point during their larval or adulthood stages, and that distinguish them from all other taxa: a notochord, a dorsal hollow nerve cord, endostyle, pharyngeal slits, and a post-anal tail (Rychel et al., 2006) (**Figure 6**). Vertebrates represent the overwhelming majority of the phylum, with currently about 70,000 species described, including fishes (jawless, cartilaginous and bony fishes), amphibians, reptiles, birds and mammals (**Figure 6**). Vertebrates are present in all habitats and are of great ecological and economical relevance. Cephalochordates, in contrast, are a small group (≈30 species) of filter-feeding marine animals typically distributed in tropical and temperate shallow seas (**Figure 6**). However, cephalochordates are relevant under an evolutionary point of view because due to their morphological, developmental and genetic stasis (Cañestro, 2012; Somorjai et al., 2008), they may represent better the chordate ancestor than the other chordate subphyla (Louis et al., 2012). Finally, tunicates, first described by Alexander Kowalevsky 140 years ago (Berná and Alvarez-Valin, 2014; Kovalevskij,

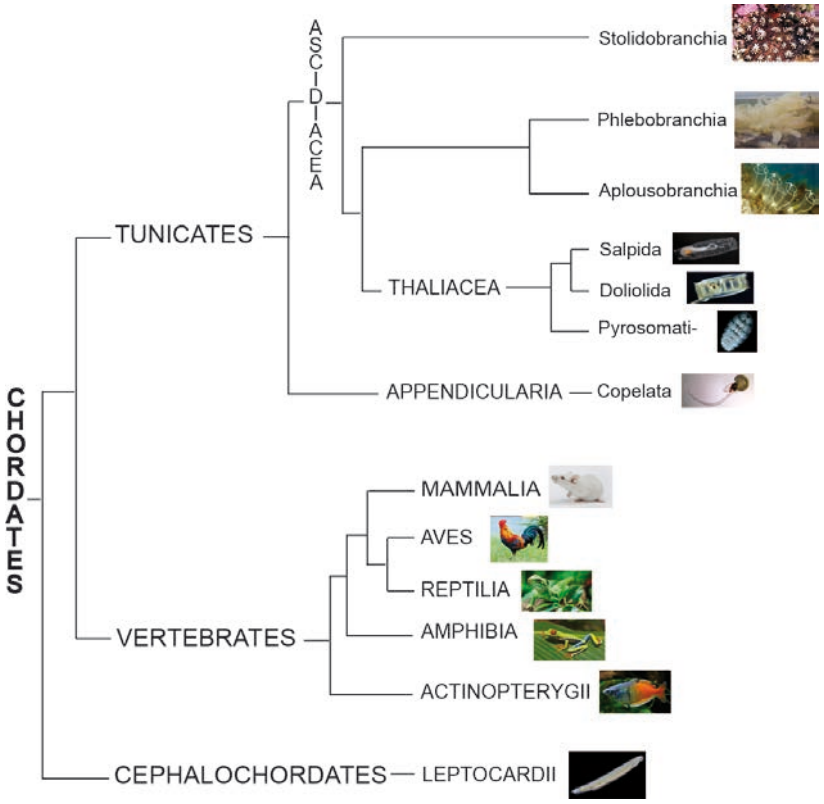


Figure 6. Chordate phylogeny. For tunicates, traditional classifications dividing this subphylum in three classes, Appendicularia, Ascidiacea and Thaliacea. Vertebrates classification with their five classes and Cephalochordata with one class.

1866), are the closest sister group of species to vertebrates. Tunicates comprise around 3,000 species of ecologically diverse filter-feeding animals, including planktonic and benthic specimens as well as solitary and colonial forms. Tunicates occupy an important trophic position in marine food webs (Ferrández-Roldán et al., 2019; Tsagkogeorga et al., 2012), adapted to diverse marine habitats with different conditions of metal bioavailability, and similarly to many molluscs, their filtering lifestyle makes them prone to accumulate metals and other toxics from the seawater ((Papadopoulou and Kanias, 1977; Tzafriri-Milo et al., 2019) and references therein).

Tunicates encompass three classes: Appendicularia (a.k.a Larvacea), Ascidacea and Thaliacea. The ascidians, commonly known as sea squirts, are the most studied group. Ascidian adults are sessile, either solitary or colonial, usually found in shallow water worldwide. Thaliaceans, in contrast, are free-floating (pelagic) animals for their entire lifespan, and the group includes species with complex life cycles with both solitary and colonial forms. Finally, appendicularians are also pelagic animals that represent the second most abundant group of mesozooplankton grazers and an important component of food for fish and zooplankton species. Appendicularians and thaliaceans are also ecologically relevant because they contribute to the vertical transport of carbon to deep ocean through the rapid sinking of fecal pellets and discarded houses. From an experimental point of view, 16 tunicate genomes have been published so far (https://www-ncbi-nlm-nih-gov.sire.ub.edu/genome/?term=Tunicata), as well as a large number of transcriptomic projects (2967 SRA of RNA projects in NCBI), facilitating the *in silico* identification of virtually any gene in a wide set of tunicate species.

Before this thesis, only two MTs had been identified in two tunicate species, *Ciona intestinalis* and *Herdmania curvata* (Franchi et al., 2011), but neither their metal-binding specificity nor their capacity had been determined. In cephalochordates, MTs had been characterized also in two species, *Branchiostoma floridae* and *Branchiostoma lanceolatum* with preference for Cd ions (Guirola et al., 2012). In contrast, the vertebrate MTs had been extensively analysed. The first MT discovered was indeed a vertebrate Cd-binding protein from equine kidney cortex (Margoshes and Vallee, 1957), and since then, more than 125 vertebrate MTs have been studied (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000006/protein/reviewed/?page_size=100#table). The first NMR structure of an MT was also from a vertebrate, the rabbit MT2 (Braun et al., 1986; Otvos et al., 1985; Wagner et al., 1987), and since then other 3D structures of vertebrate MTs have been resolved such as human MT2 and MT3, mouse MT1 and MT3, and rat MT2 (Messerle et al., 1990; Öz et al., 2001; Schultze et al., 1988; Wang et al., 2006; Zangger et al., 1999). As mentioned before, vertebrates

have four MT types, MT1, MT2, MT3, and MT4, with a distinctive species distribution, different expression patterns and diverse metal-binding preferences. All vertebrates MTs are bi-domain MTs with a same domain organization, an amino-terminal β domain clustering three divalent metal ions, and a carboxyl-terminal α domain clustering four divalent metal ions. No large multi-modular MTs have been reported in vertebrates so far.

Previous comparative analyses of tunicate, cephalochordate and vertebrate MTs had showed a high degree of structural diversity between lineages, and the origin of chordate MTs, as well as the evolution of their domain-repeat organization, remained poorly understood. To shed a light on what seemed a complex and enigmatic case of modular protein evolution was therefore necessary to conduct comprehensive studies of diverse chordate MTs, and the MTs of the fast-evolving tunicates lineages appeared particularly attractive.

OBJECTIVES

The general objective of this PhD project has been to study the modular evolution of domain repeat proteins using the MTs as case study. This project aimed to identify, analyse, characterize and compare the MTs form two large groups of organisms – molluscs and chordates– in order to investigate the evolution, function and structural organization of their MTs and domains, and to appreciate the impact of modularity on the evolution of proteins made of domain repeats. To achieve this general objective, the project was divided into four specific objectives:

Objective 1. To perform comprehensive surveys of mollusc and chordate databases and identify the MTs in a large number of phylogenetically diverse species of these two phyla.

Objective 2. To analyse and compare mollusc and chordate MT sequences in order to recognize patterns of amino acid conservation and potential metal-binding domains, and to identify large modular MTs made of tandem domain repeats.

Objective 3. To characterize the metal-protein complexes of the MTs and of the potential domains predicted from sequence analysis, providing experimental evidence of the modular organization of the proteins and allowing to infer the biological function of the MTs and their eventual contribution to the adaptive processes of the species.

Objective 4. To reconstruct the evolutionary history of mollusc and chordate MTs and to reveal the evolutionary events that shaped their MTs during the diversification of these phyla, including gene gains and losses, reorganization, expansion or *de novo* emergence of domains, and processes of parallel or convergent evolution.

SUPERVISORS' REPORT

Thesis supervisors' report about authorship and impact factor of the publications of this doctoral thesis presented by Sara Calatayud Robert

Dr. Ricard Albalat and Dr. Cristian Cañestro, supervisors of the PhD thesis entitled "**Modular evolution of domain repeat proteins. Metal-binding and domain repeats of Metallothineins in mollusks and chordates.**" by Sara Calatayud Robert, certify that the results obtained have been or will be submitted to peer-reviewed international journals. The entire thesis comprises six articles: five of them already published, and one finished and ready to be submitted. None of the articles have been used for the elaboration of other PhD thesis. The details of the papers, journals and their impact factor are detailed below:

Article 1: **Modularity in Protein Evolution: Modular Organization and De Novo Domain Evolution in Mollusk Metallothioneins. Sara Calatayud**, Mario Garcia-Risco, Veronika Pedrini-Martha, Douglas J. Eernisse, Reinhard Dallinger, Òscar Palacios, Mercè Capdevila and Ricard Albalat.

Molecular Biology and Evolution, Volume 38, Issue 2, February 2021, Pages 424–436. doi: 10.1093/molbev/msaa230

Impact factor: 16.240 (2020)

Category: Biochemistry, Genetics and Molecular Biology

Rank SCOPUS: 6/382; Quartile: Q1

The PhD candidate contributed to the identification in the public available genomes of new metallothioneins, and to the analysis of the amino acid sequences and patterns of cysteine motifs in order to determine the domains of the MTs in each mollusk lineage. She also did the gDNA extraction, PCR amplification and sequencing of *Lottia gigantea* MTs, and the recombinant synthesis and purification of all metal-MTs complexes for chemical analyses. She also contributed to the reconstruction of mollusk MT evolution, and to the writing and making of the figures of the manuscript.

Article 2: **Exploring the modular organization of metallothionein in Molluscs. Sara Calatayud**, Mario Garcia-Risco, Mercè Capdevila, Òscar Palacios and Ricard Albalat. Ready to be submitted.

The PhD candidate contributed to the amino acid sequence analysis of the selected MTs, to the study of their homology and evolutionary relationship with the MTs of other mollusk clades, and to the analyses of their cysteine arrangements in order to determine their domain structure. She also did the recombinant synthesis and purification of all metal-MTs complexes for chemical analyses. She also contributed to the writing and making of the figures of the manuscript.

Article 3: **Two Unconventional Metallothioneins in the Apple Snail Pomacea bridgesii Have Lost Their Metal Specificity during Adaptation to Freshwater Habitats.** Mario García-Risco, **Sara Calatayud**, Michael Niederwanger, Ricard Albalat, Òscar Palacios, Mercè Capdevila and Reinhard Dallinger.

Int. J. Mol. Sci. 2021, 22, 95. doi:10.3390/ijms22010095

Impact factor: 5.923 (2020)

Category: Inorganic Chemistry

Rank SCOPUS: 12/69; Quartile: Q1

The PhD candidate contributed to *Pomacea bridgessii* amino acid sequence analysis, to examine their homology and evolutionary relationship with the MTs of other gastropod clades, and to study their cysteine arrangements in order to determine their domain structure. She also did the recombinant synthesis and purification of all metal-MTs complexes for chemical analyses and contributed to the writing and making of the figures of the manuscript.

Article 4: **Metallothioneins of the urochordate Oikopleura dioica have Cys-rich tandem repeats, large size and cadmium-binding preference. Sara Calatayud**, Mario Garcia-Risco, Natalia S. Rojas, Lizethe Espinosa-Sánchez, Sebastián Artime, Òscar Palacios, Cristian Cañestro and Ricard Albalat.

Metallomics, 2018, 10, 1585--1594. doi: 10.1039/c8mt00177d

Impact factor: 3.796 (2020)

Category: Chemistry and Metals

Rank SCOPUS: 59/793; Quartile: Q1

The PhD candidate contributed to the identification of the metallothioneins of *O.dioica* in the genome, to the nucleic acid sequence analysis in order to characterize the promoter regions, and to the study of their cysteine arrangements in order to determine their domain structure. She also did the cloning, recombinant synthesis and purification of all metal-MTs complexes for chemical analyses and contributed to the writing and making of the figures of the manuscript.

Article 5: **Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins. Sara Calatayud**, Mario Garcia-Risco, Mercè Capdevila, Cristian Cañestro, Òscar Palacios and Ricard Albalat. *The first two authors equally contributed.

Front. Cell Dev. Biol. July 2021 | Volume 9 | Article 702688
doi:10.3389/fcell.2021.702688

Impact factor: 6.684 (2020)

Category: Cell Biology and Developmental Biology

Rank SCOPUS: 64/81; Quartile: Q1

The PhD candidate contributed to the identification of the metallothioneins in different *O.dioica* populations, and to their amino acid sequence analysis in order to determine the population variability. She also did the extraction, PCR amplification and cloning of the OdiMTs from the Catalonia population, and the recombinant synthesis and purification of all metal-MTs complexes for chemical analyses. She contributed to the writing and making of the figures of the manuscript.

Article 6: **Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions and functional convergences. Sara Calatayud**, Mario Garcia-Risco, Òscar Palacios, Mercè Capdevila, Cristian Cañestro and Ricard Albalat.

Molecular Biology and Evolution, 2021; msab184; doi:10.1093/molbev/msab184

Impact factor: 16.240 (2020)

Category: Biochemistry, Genetics and Molecular Biology

Rank SCOPUS: 6/382; Quartile: Q1

The PhD candidate contributed to the identification in the public available genomes of new metallothioneins in tunicate, cephalochordate and vertebrate species, and to the analysis of the amino acid sequences and patterns of cysteine motifs in order to determine the domains of the MTs in each chordate lineage. She also did, the gDNA extraction, PCR amplification and sequencing of the MTs from *Salpa thompsoni*, and the recombinant synthesis and purification of all metal-MTs complexes for chemical analyses. She also contributed to the reconstruction of chordate MT evolution and to the writing and making of the figures of the manuscript.

Barcelona, September 30th, 2021

Ricard Albalat Rodriguez

Cristian Cañestro García

SUMMARY OF THE RESULTS

To achieve the objectives of this PhD project I have identified the MTs in the sequence databases of many species of the Mollusca and Chordata phyla. I have analysed their domain organization and characterized the metal-binding features of some selected ones. Notice that the metal-binding characterization of the MTs involves ICP-AES and ESI-MS analyses, and these analyses have been carried out by our collaborating team at the Universitat Autònoma de Barcelona –Mario García-Risco as PhD student, and Òscar Palacios and Mercè Capdevila as PhD supervisors– using the metal-proteins complexes that I have recombinantly produced and purified. From the gathered data, I have been able to infer the evolutionary history of the MTs during the diversification of the Mollusca and Chordata phyla, and to evaluate the impact of the modular structure of the MTs on this history. My results have been described in six articles (five already published and one in preparation), of which I am first author in five of them. I have also participated in four additional articles that have been included in the Appendix Section of this memory.

RESULTS I: MOLLUSC MTs

Regarding the mollusc MTs, the results have been described in three articles. The first article entitled "*Modularity in Protein Evolution: Modular Organization and De Novo Domain Evolution in Mollusk Metallothioneins*" by **Sara Calatayud**, Mario Garcia-Risco, Veronika Pedrini-Martha, Douglas J. Eernisse, Reinhard Dallinger, Òscar Palacios, Mercè Capdevila and Ricard Albalat, has been published in **Molecular Biology and Evolution** 38:2: 424–436 (2021). DOI: 10.1093/molbev/msaa230 (IF_{2020}: 16.240).

In this article, I conducted exhaustive surveys of mollusc public databases and created the most comprehensive catalogue of mollusc MTs to date, 272 MTs from 189 species. I identified for the first time MTs in six out of the eight classes of molluscs (i.e. Scaphopoda, Cephalopoda, Monoplacophora, Polyplacophora, Solenogastres and Caudofoveata), and in two out of the five gastropod subclasses (i.e. Neritimorpha and Patellogastropoda). Sequence analyses of all these sequences allowed me to define six mollusc MT domains –α, β1, β2, β3, γ and δ –, some of them new, that are combined in bi-domain MTs in a lineage-specific manner. I also characterized the metal-binding features of the new MTs of the Neritimorpha and Patellogastropoda subclasses, showing a divalent-metal character for the Neritimorpha MTs and a very strong Cd-thionein character for Patellogastropoda MTs as well as for new Patellogastropoda γ domain when it was expressed alone. Finally, comparative analyses of the mollusc MTs indicated that the ancestral mollusc form probably was a bi-modular Cd-thionein, which either was duplicated leading to MT multiplicity or expanded their number of domains generating multi-domain MTs. These results revealed the impact of gene duplications and losses as well as the emergence or expansion of domains on the modular evolution of MTs in different mollusc lineages.

The second article was entitled "*Exploring the modular organization of metallothionein in Mollusks*" by **Sara Calatayud**, Mario Garcia-Risco, Mercè Capdevila, Òscar Palacios and Ricard Albalat, and it is still **in preparation**.

In this work, I extended the characterization of the predicted mollusc MT domains described in the first article, focusing in the α and δ domains. I analyzed the MTs of two mollusk species, the αβ1-MT of *Nautilus pompilius* of the Cephalopoda class in the Conchifera clade, and the δβ1-MT of *Falcidens caudatus*, a Caudofoveata species of the Aculifera group. I demonstrated that the predicted α and δ domains were genuine MT domains capable to autonomously bind metal ions. I showed the preference of the

NpoMT1 and FcaMT1 for divalent metal ions, which would be at least in part determined by the metal preference of their respective α and δ domains. In conclusion, my results supported the bi-domain organization of the Cephalopoda and Caudofoveata MTs and the independent functionality of their domains, and reinforced the Cd-thionein character of the ancestral mollusc MT.

The third article entitled "*Two Unconventional Metallothioneins in the Apple Snail Pomacea bridgesii Have Lost Their Metal Specificity during Adaptation to Freshwater Habitats*" by Mario García-Risco, **Sara Calatayud**, Michael Niederwanger, Ricard Albalat, Òscar Palacios, Mercè Capdevila and Reinhard Dallinger, has been published in **International Journal of Molecular Sciences** 22:95 (2021). DOI:10.3390/ijms22010095 (IF_{2020}: 5.923).

This article describes the analysis of the two MTs -PbrMT1 and PbrMT2- of *Pomacea bridgesii*, a freshwater snail belonging to the gastropod class of Caenogastropoda. These MTs were selected because they showed deviations to the archetypal gastropod bi-domain organization, since both MTs were multi-modular MTs made of three domains. Thus, in addition to the archetypal β3 and β1 domains, PbrMT1 contained an extra N-terminal H_4C_4 domain, while PbrMT2 would have lost the β1 domain but tandem duplicated the β3 domain twice, resulting in a β3β3β3 MT. I characterized the metal-binding features of these MTs, showing that the increase in their number of domains correlated with an increase of their binding capacity for metal ions. I also showed that although none of these MTs exhibited strong metal specificity, they behaved different, with PbrMT1 resembling a Zn(II)-thionein and PbrMT2 showing a more Cu(I)-thionein character. These results exemplified a case of MT multiplicity in which duplicates diverged in their metal preferences, supporting the idea that during the colonization to freshwater habitats, the metal preference of some mollusc species MTs changed.

ARTICLE I: Modularity in Protein Evolution: Modular Organization and *De Novo* Domain Evolution in Mollusk Metallothioneins

Mol Biol Evol. 2021 Jan 23;38(2):424-436.

DOI: 10.1093/molbev/msaa230 (IF_{2020}: 16.240).

Modularity in Protein Evolution: Modular Organization and De Novo Domain Evolution in Mollusk Metallothioneins

Sara Calatayud,[1] Mario Garcia-Risco,[2] Veronika Pedrini-Martha,[3] Douglas J. Eernisse,[4] Reinhard Dallinger,[3] Òscar Palacios,[2] Mercè Capdevila,[2] and Ricard Albalat*,[1]

[1]Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

[2]Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

[3]Department of Zoology and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innsbruck, Austria

[4]Department of Biological Science, California State University Fullerton, Fullerton, CA

***Corresponding author:** E-mail: ralbalat@ub.edu.

Associate editor: Yoko Satta

Abstract

Metallothioneins (MTs) are proteins devoted to the control of metal homeostasis and detoxification, and therefore, MTs have been crucial for the adaptation of the living beings to variable situations of metal bioavailability. The evolution of MTs is, however, not yet fully understood, and to provide new insights into it, we have investigated the MTs in the diverse classes of Mollusks. We have shown that most molluskan MTs are bimodular proteins that combine six domains—α, β1, β2, β3, γ, and δ—in a lineage-specific manner. We have functionally characterized the Neritimorpha $\beta_3\beta_1$ and the Patellogastropoda $\gamma\beta_1$ MTs, demonstrating the metal-binding capacity of the new γ domain. Our results have revealed a modular organization of mollusk MT, whose evolution has been impacted by duplication, loss, and de novo emergence of domains. MTs represent a paradigmatic example of modular evolution probably driven by the structural and functional requirements of metal binding.

***Key words:* α, β1, β2, β3, γ, and δ domains, bi- and multimodular metallothioneins, cysteine motifs, de novo evolution, metal-binding capacity and preference.**

Introduction

Metallothioneins (MTs) are a superfamily of intracellular, cysteine-rich (≈15–30%), and mostly low-molecular-weight (<100 amino acids) proteins present across eukaryotes, from various protists to plants, fungi, and animals (Capdevila and Atrian 2011; Blindauer 2014). Their cysteine (Cys, C) residues are arranged in distinctive motifs (i.e., CxC, CC, and CCC), whose number and distribution led to the original definition of the two functional domains in vertebrate MTs (Braun et al. 1986) designated as α domain (with 11–12 cysteines at the C-terminal region) and β domain (with nine cysteines at the N-terminal region), joined by a linker sequence. In this bimodular structure made of two domains, the cysteine motifs of each domain are able to form metallic clusters, thus conferring the capacity of binding both essential and nonessential metals (Nielson and Winge 1985). Notice that the α and β nomenclature has been also used for referring to N-terminal (α) or C-terminal (β) domains in some gastropod MTs (Baumann et al. 2017; Niederwanger, Calatayud, et al. 2017; Palacios et al. 2017; Schmielau et al. 2019). In order to avoid confusion, however, we will use here the α/β nomenclature in its primary meaning for classifying domains based on the number and distribution of cysteine motifs (Jenny et al. 2016; Nam and Kim 2017), and not on their N- or C-terminal position.

Thanks to their metal-binding capacity, MTs are directly involved in metal homeostasis and detoxification, but also in radical scavenging, oxidative stress protection, and antiapoptotic defense (Capdevila et al. 2012), leading to their role as a model system for the investigation of the genetic mechanisms by which organisms adapt to diverse metal bioavailabilities or stress. For instance, duplications of *MT* genes (Maroni et al. 1987; Adamo et al. 2012) or expansions of MT domains (Tanguy and Moraga 2001; Jenny et al. 2016; Pedrini-Martha et al. 2020), along with elevated levels of MT expression (Timmermans et al. 2005; Janssens et al. 2008, 2009; Costa et al. 2012; Catalan et al. 2016; de Francisco et al. 2018) or changed metal specificity (Tio et al. 2004; Palacios et al. 2011; de Francisco et al. 2017) have been considered adaptive events contributing to increase the metal and stress tolerance of the organisms in different environments. In particular, a previous thorough analysis investigating the role of cadmium (Cd) on the evolution of gastropod MTs has led us to suggest that lineage-specific changes of metal-selectivity features might have been important during the recurrent colonization of marine gastropods to terrestrial and freshwater habitats, where they had to face the challenge of adapting to

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5912274 by guest on 28 October 2020

Article

Table 1. Cysteine Motifs of Mollusca Bimodular MTs.

	Class	Clade	Domain		Cysteine Motifs		
					N-Terminal	Linker	C-Terminal
Conchifera	Gastropoda	Patellogastropoda	γ	$\beta1$	$CCx_5CCx_4CCx_6CCx_7CC$	X_{5-6}	$[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$
		Vetigastropoda	$\beta3$	$\beta1$	$Cx_3Cx_4[CxC]x_3[CxC]x_4[CxC]x_2C$	X_{3-6}	$[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$
		Neritimorpha	$\beta3$	$\beta1$	$Cx_3Cx_5[CxC]x_3[CxC]x_3[CxC]x_2C$	X_{3-4}	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
		Caenogastropoda	$\beta3$	$\beta1$	$Cx_3Cx_4[CxC]x_5[CxC]x_3[CxC]x_2C$	X_{2-3}	$[CxC]x_4[CxC]x_3Cx_6[CxC]x_3[CxC]$
		Heterobranchia	$\beta3$	$\beta1$	$Cx_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$	X_2	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
	Scaphododa		$\beta2$	$\beta1$	$[CxC]x_5[CxC]x_3[CxC]x_4[CxC]x_4C$	X_{8-10}	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
	Bivalvia		α	$\beta1$	$[CxC]x_5[CxC]x_3Cx_5[CxC]x_3[CxC]x_3[CxC]x_2C$	X_{3-4}	$[CxC]x_6[CxC]x_3Cx_5[CxC]x_3[CxC]$
			$\beta2$	$\beta2$	$[CxC]x_5[CxC]x_3[CxC]x_4[CxC]x_2C$	X_{2-3}	$Cx_2Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$
			$\beta2$ like	$\beta1$	$Cx_4Cx_4[Cx_{0-1}C]x_4[Cx_{0-1}C]x_3[CxC]x_2C$	X_{1-3}	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
	Cephalopoda		α	$\beta1$	$[CxC]x_6[CxC]x_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$	X_{3-4}	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
	Monoplacophora		$\beta2$	$\beta1$	$[CxC]x_6[CxC]x_3[CxC]x_2Cx[CxC]x_2C$	X_3	$[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$
Aculifera	Polyplacophora		α	$\beta1$	$[CxC]x_4[CxC]x_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$	X_3	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
	Solenogastres		$\beta2$	$\beta1$	$[CxC]x_5[CxC]x_3[CxC]x_4[CxC]x_4C$	X_6	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$
			α	$\beta1$	$[CxC]x_5[CxC]x_3Cx_5[CxC]x_3[CxC]x_3[CxC]x_2C$	X_7	$[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$
	Caudofoveata		δ	$\beta1$	$[CxC]x_3[CxC]CCx_4Cx_3[CxC]x_3CCx_4[CxC]C$	X_3	$[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$

ecosystems with different levels of metal bioavailability (Dallinger et al. 2020). In this context, studies on MT evolution have been of interest to evolutionary ecologists (Janssens et al. 2009; Faddeeva-Vakhrusheva et al. 2016; Zhang et al. 2018, 2019; Purać et al. 2019) who have associated environmental factors—that is, concentrations of heavy metals—with the evolution of diverse MTs.

Gastropoda is composed of five distinctive lineages ranked as subclasses, whose interrelationships are still controversial: Patellogastropoda, Vetigastropoda, Neritimorpha, Caenogastropoda, and Heterobranchia (Zapata et al. 2014; Cunha and Giribet 2019). Although several gastropod MTs have been previously investigated, these studies have an uneven phylogenetic distribution. MT identification and metal-binding selectivity have been determined for 15 MTs from seven heterobranch species (with the most extensive studies), for four MTs from three caenogastropod species, and for a single vetigastropod MT. In order to fully understand the evolution of gastropod MTs and their metal-binding specificities, it was therefore necessary to extend studies to include particular species of Patellogastropoda (true limpets) and Neritimorpha (snails known as nerites and their allies). To add both patellogastropods and neritimorphs here, we have conducted an exhaustive survey of gastropod MTs in public databases, including raw sequencing data from transcriptomic and genomic high-throughput sequencing projects. We have been able to reconstruct and identify new MTs from limpet and neritid species. We have characterized the metal-binding properties of new MTs for each, selecting *Lottia gigantea* and *Nerita peloronta* MTs as representatives of each taxon. Our data set extends the knowledge of metal-binding preferences to MTs of previously unstudied gastropod taxa, broadening the scope of known diversification of metal-binding selectivity in MTs for gastropods, the most species-rich, and arguably the most morphologically and ecologically diverse, class of Mollusca.

By extending what we have learned with new genomic searches, we have added novel MTs from the other three gastropod clades (Vetigastropoda, Caenogastropoda, and Heterobranchia) and from all mollusk classes, identifying and classifying new MTs from other conchiferan (i.e., Gastropoda, Bivalvia, Cephalopoda, Monoplacophora, and Scaphopoda) and aculiferan (i.e., Polyplacophora, Solenogastres, and Caudofoveata classes) mollusks. In summary, by collecting 272 MTs from 189 different species, we have created the most comprehensive catalog of mollusk MTs compiled so far, and we have exposed patterns of modular organization across molluskan MTs.

Results

Identification of Patellogastropoda MTs

We surveyed the genome project of the *L. gigantea* (Patellogastropoda: Lottiidae). This search yielded an automatically predicted MT (Gene ID: 20249168; hypothetical protein: XP_009056965.1), whose sequence, structure, and size suspiciously differed from other Gastropoda *MT* genes. The low-quality sequence of the genomic region containing the putative *MT* gene (NW_008709190.1) prompted us to reanalyze this region by polymerase chain reaction (PCR) amplifying, cloning, and resequencing >12 kb of *L. gigantea* genome (supplementary fig. S1*A*, Supplementary Material online). Sequence analysis identified two *MT* genes tandem repeated in this region (accession number MK795721), which we named *LgiMT1* and *LgiMT2*. Our predictions were further supported by PCR amplification of the corresponding cDNAs (MK770430 and MK770431) and by nine *L. gigantea* expressed sequence tags (supplementary table S1, Supplementary Material online), corroborating the misassembly of the *MT* genomic region in the genome project, and revealing some degree of polymorphism at the amino acid level (S/T and S/A at amino acid positions 3 and 68, respectively, supplementary fig. S1*B*, Supplementary Material online). Comparison of the coding region (CDS) of *LgiMT1* and *LgiMT2* genes revealed that they were 97.78% identical at the nucleotide level (220 out of 225 nt) and both encoded for a 75 amino acid protein with 19 cysteines (25.7%). LgiMT1 and LgiMT2 proteins only differed at positions 6, 20, and 49: P,

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

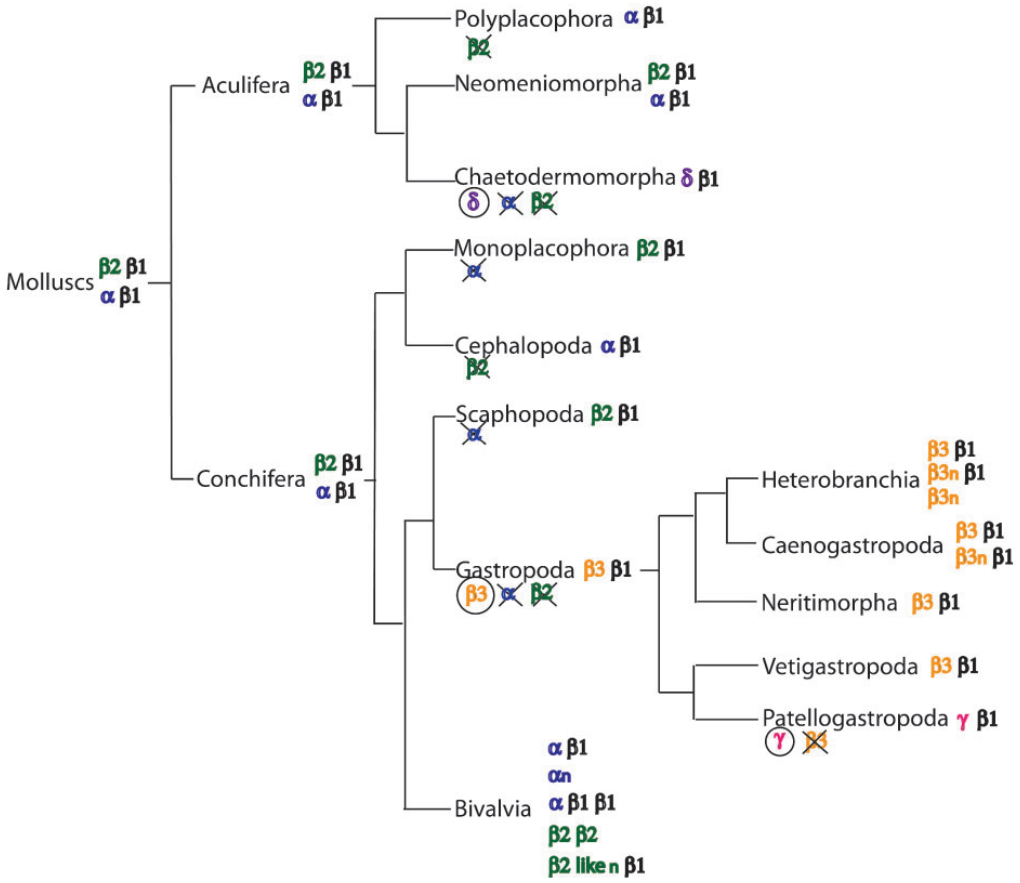


FIG. 1. Structural evolution of MTs along the Mollusca phylum. Most mollusk MTs have a bimodular structure made of a variable taxon-specific N-terminal domain—α, β2, β3, γ or δ—, and a conserved C-terminal β1 domain. Domains are classified based on the number and configuration of the cysteine motifs. Based on the distribution of domains among the different clades, the most parsimonious evolutionary scenario would be that α, β1, and β2 domains are ancient, already present in the MTs of mollusk ancestor. In contrast, β3, γ, and δ domains would be de novo domains arose in Gastropoda, Patellogastropoda, and Caudofoveata, respectively. Domain gains (circled) and losses (crossed) are indicated under each clade. Deviations of the standard bimodular structure are found in MTs of Bivalvia and Gastropoda, including multimodular MTs with more than two domains (α_n [Jenny et al. 2004, 2016]; $\beta1_n$ and β2-like$_n$ [Nam and Kim 2017] in Bivalvia; and $\beta3_n$ in Heterobranchia and Caenogastropoda [Niederwanger, Calatayud, et al. 2017; Palacios et al. 2017]), and MTs lacking the conserved β1 domain (α_n and $\beta2\beta2$ MTs in Bivalvia; $\beta3_n$ MTs in Heterobranchia). Phylogenetic relationships within Mollusca are based on Zapata et al. (2014), Cunha and Giribet (2019), and Kocot et al. (2020)

L, and P residues in LgiMT1, and A, S, and S in LgiMT2 (supplementary fig. S1*B*, Supplementary Material online).

We extended the limpet MT listing to three families and eight additional species of Patellogastropoda (supplementary fig. S2*A* and table S2, Supplementary Material online, and Dallinger et al. 2020). These include five species of Lottiidae: *Lottia digitalis* (LdiMT1 partial), *Lottia kogamogai* (LkoMT1), *Lottia scutum* (LscMT1 partial), *Nipponacmea fuscoviridis* (NfuMT1), and *Patelloida pygmaea* (PpyMT1); two species of Nacellidae: *Cellana rota* (CroMT1) and *Nacella concinna* (NcoMT1); and one species of Patellidae: *Patella vulgata* (PvuMT1 and PvuMT2). Amino acid comparison of the 11 Patellogastropoda MTs revealed high sequence similarity (from 69% to 97% identity) with the distinctive 19 cysteines fully conserved. Interestingly, the 19 cysteines were organized into two domains: a novel MT domain at the N-terminal region with ten cysteines arranged in five CC pairs ($CCx_5CCx_4CCx_6CCx_7CC$) that we named γ domain (notice that this γ domain is not related with the N-terminal, six Cys domain of plant E_c-1 MTs, also named γ domain [Loebus et al. 2011]), which was connected by a linker of four to six amino acids to an archetypal 9-Cys β1 domain ($[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$) (Jenny et al. 2016; Nam and Kim 2017) at the C-terminal region (table 1 and supplementary fig. S2*A*, Supplementary Material online). We concluded therefore that the structure of Patellogastropoda MTs was of γ/β1 domains (fig. 1).

Metal-Binging Capacity of Patellogastropoda MTs

In order to demonstrate the MT nature and to explore the metal-selectivity features of the two *L. gigantea* MTs, we studied the formation of metal-LgiMT1 and metal-LgiMT2 complexes by the proteins heterologously expressed in *Escherichia coli* and grown in media supplemented with copper (Cu), cadmium (Cd), or zinc (Zn) salts by inductively coupled plasma atomic emission spectrometer (ICP-AES) and

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

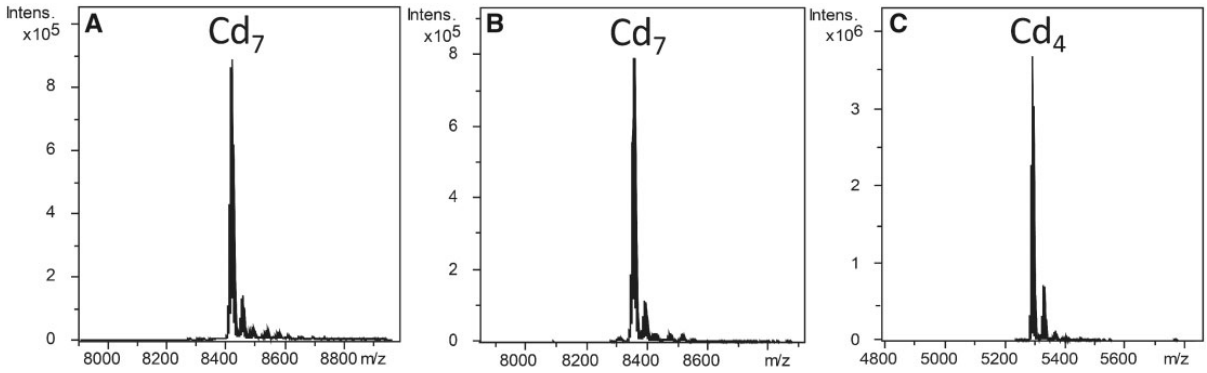


FIG. 2. Deconvoluted ESI-MS spectra LgiMT1 (*A*), LgiMT2 (*B*), and γLgiMT2 (*C*) recombinantly produced by *Escherichia coli* in Cd-enriched media.

electrospray ionization mass spectrometry (ESI-MS) analyses. ICP-AES is an analytical technique that allows for protein quantification and metal-to-protein stoichiometry determination through the measurement of element composition of the samples (S, Zn, Cd, and Cu) (Bongers et al. 1988), and ESI-MS is used to determine the molecular mass of the species formed, that is, the speciation of the samples (Capdevila et al. 2012). The ICP-AES and native ESI-MS (recorded at neutral pH in order to allow the observation of unaltered species) analyses (data not shown) of the recovered samples showed that LgiMT1 and LgiMT2 rendered a mixture of metallated species both in Zn(II)- as well as in Cu(II)-supplemented cultures, thus indicating the absence of preference for either of those metal ions (Zn(II) and Cu(II) stand for divalent Zn^{2+} and Cu^{2+} ions). On the contrary, unique Cd_7-LgiMT1 and Cd_7-LgiMT2 species, that is, MTs loaded with seven Cd(II) ions, were recovered from Cd-enriched culture media (fig. 2*A* and *B*). The identification of single species after recombinant metal-supplemented productions normally reveals the existence of a preferred thermodynamically favored metal cluster, whereas in the absence of any specific species, the formation of a variety of species with similar but not identical metal content is observed (Palacios et al. 2014). The observation of single Cd_7-LgiMT species, together with the observation by acid ESI-MS (recorded at pH 2.4 that lead to partial protonation of thiol Cys groups, which normally results on the release of Zn and Cd while bound Cu is maintained) that both proteins are reluctant to release Cd(II) by acidification at pH 2.4 giving rise to Cd_4-LgiMT species, demonstrated the Cd-thionein character of both Patellogastropoda MTs.

Metal-Binding Functionality of the New γ Domain

In order to determine the independent metal-binding capacity of the new γ domain, we analyzed its ability to form metal complexes when expressed alone. Thus, we heterologously expressed the 10-Cys γ domain of LgiMT2 (from Met_1 to Gln_{45}), hereafter as γLgiMT2, in *E. coli* grown in medium supplemented with Cu, Cd, or Zn salts. The election of the γ domain of LgiMT2 relied on the fact that it was more conserved than that of LgiMT1 when compared across the γ domains of other Patellogastropoda MTs (supplementary fig. S2*A*, Supplementary Material online). The distinct metal-γLgiMT2 preparations obtained after purification were characterized by ICP-AES and ESI-MS. The results showed a clear preference of γLgiMT2 for divalent Zn(II) and Cd(II) ions, as major M_4-γLgiMT2 species were identified for both metal ions by ESI-MS at pH 7. Interestingly, formation of Cd_4-γLgiMT2 complexes was more favored than Zn_4-γLgiMT2 complexes, as suggested by the fact that the former one was obtained as a single species (fig. 2*C*), whereas the latter one coexisted with other minor species (data not shown). This Cd_4-γLgiMT2 cluster also exhibited a high resistance against demetallation at acidic pH levels, as this required an acidification down to pH 1 to obtain the apoprotein. Conversely to the Zn- and Cd-cultures, the preparations obtained from Cu(II)-enriched media yielded mixtures of multiple Cu_x-γLgiMT2 complexes (*x* ranging from 5 to 10), confirming the poor preference of this domain for Cu(I).

Identification of Neritimorpha MTs

We collected 12 MTs from two families and six species of Neritimorpha, including five species of Neritidae: *Clithon retropictum* (CretrMT1 and CretrMT2), *Nerita albicilla* (NalMT1 and NalMT2), *Nerita melanotragus* (NmeMT1 and NmeMT2), *N. peloronta* (NpeMT1 and NpeMT2), and *Neritina pulligera* (NpuMT1 and NpuMT2); one species of Neritopsidae: *Titiscania limacina* (TliMT1 and TliMT2 partial) (supplementary fig. S2*B* and table S2, Supplementary Material online, and Dallinger et al. 2020). All Neritimorpha species possessed two MTs, and this appears to be a lineage-specific gene duplication shared by at least two of the four extant superfamilies within Neritimorpha. As for patellogastropods, amino acid comparison of the 12 Neritimorpha MTs showed high sequence conservation (from 65% to 96% identity) with 19–20 cysteines organized in two MT domains: at the N-terminal region, a 9-Cys β3 domain ($Cx_3Cx_{4-5}[CxC]x_3[CxC]x_3[CxC]x_2C$, formerly known as α1/2 domain in non-Neritimorpha species [Baumann et al. 2017; Niederwanger, Calatayud et al. 2017; Palacios et al. 2017; Schmielau et al. 2019] with an extra cysteine in Neritimorpha MT1s), connected by a linker of three to four amino acids to a 9-Cys β1 domain at the C-terminal region ($[CxC]x_{3-4}[CxC]x_3Cx_5[CxC]x_3[CxC]$) (table 1 and supplementary fig. S2*B*, Supplementary Material online). The

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

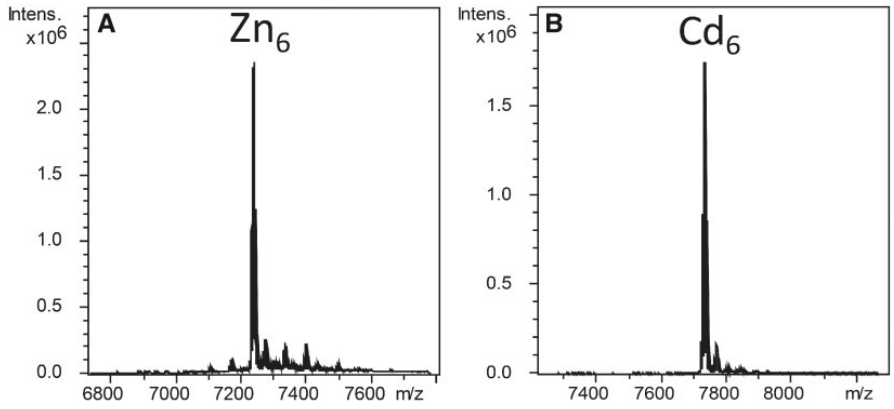


FIG. 3. Deconvoluted ESI-MS spectra of the in vivo preparations of *Escherichia coli* recombinant NpeMT1 obtained from a Zn-enriched medium (*A*) and NpeMT2 synthesized in Cd-enriched cultures (*B*).

structure of Neritimorpha MTs was therefore of $\beta3/\beta1$ domains (fig. 1).

Metal-Binding Capacity of Neritimorpha MTs

In order to analyze the metal-binding abilities of Neritimorpha MTs, we studied the formation of metal-MT complexes of NpeMT1 and NpeMT2 heterologously expressed in *E. coli* grown in medium supplemented with Cu, Cd, or Zn salts. Metal-NpeMT complexes were purified and analyzed by ICP-AES and ESI-MS. Our data showed that both Neritimorpha MTs presented similar specificities for divalent metal ions, Zn(II) and Cd(II). However, their binding preference was not exactly the same, since NpeMT1 was more specific for Zn(II) than NpeMT2, whereas NpeMT2 was more specific for Cd(II) than NpeMT1. Consequently, Zn_6-NpeMT1 and Cd_6-NpeMT2 complexes could be recovered as single species (fig. 3). The Zn-thionein character of NpeMT1 and the Cd-thionein nature of NepMT2 were also supported by the recombinant productions of this protein in Cu-enriched media, which rendered mixtures of heterometallic Zn, Cu-MT complexes for NpeMT1, but mixtures of homometallic Cu-MT species for NpeMT2 (data not shown).

MTs from Other Gastropoda Clades

We extended the surveys of MTs to gastropod species belonging to Caenogastropoda, Heterobranchia, and Vetigastropoda, considered separately below. We collected a total of 163 MT sequences, including newly identified MTs and previously reported ones (supplementary table S2, Supplementary Material online).

Caenogastropoda MTs

We collected 55 MT sequences (43 new) of 43 Caenogastropoda species (supplementary fig. S2*C* and table S2, Supplementary Material online). MT multiplicity was observed in 11 species (ten species with two, and *Pomacea canaliculata* with three MTs). Size diversity within Caenogastropoda was high, with MTs ranging from 66 amino acids (BacMT1) to 251 amino acids (McornMT2), with a cysteine content of 17 (25.8%) and 72 cysteines (28.7%), respectively. Overall, Caenogastropoda MTs were organized in an N-terminal 9-Cys $\beta3$ domain ($Cx_3Cx_4[CxC]x_{3-5}[CxC]x_3[CxC]x_2C$) linked by two to three residues to a C-terminal 9-Cys $\beta1$ domain ($[CxC]x_{3-4}[CxC]x_3Cx_{5-6}[CxC]x_3[CxC]$) (table 1 and supplementary fig. S2*C*, Supplementary Material online). The archetypal structure of Caenogastropoda MTs was therefore of $\beta3/\beta1$ domains (fig. 1). Remarkably, $\beta3$ domain duplications were observed in a number of Caenogastropoda MTs: one single duplication (i.e., $\beta3.1/\beta3.2$) in 13 MTs; two duplications (i.e., $\beta3.1/\beta3.2/\beta3.3$) in JjaMT1, PbrMT2, and PcanMT2 sequences; three duplications (from $\beta3.1$ to $\beta3.4$) in partial EheMT2 and PcanMT1; five duplications (from $\beta3.1$ to $\beta3.6$) in AplaMT2; and six duplications (from $\beta3.1$ to $\beta3.7$) in McornMT2 (fig. 4*A* and supplementary fig. S3*A*, Supplementary Material online). Since these multi-$\beta3$ MTs were unevenly distributed among the Littorinimorpha, Neogastropoda, and "architaenioglossan" representatives, independent events of internal domain duplications were the most plausible origin of such MTs (Schmielau et al. 2019). Also noteworthy, ten MTs (most of them duplicated copies) of the Architaenioglossa order showed an additional $H_{3-4}C_4$ motif ($HxHHHx_2Cx_3Cx_{6-9}Cx_{0-1}C$) at the N-terminal region (supplementary fig. S2*C*, Supplementary Material online). Since these MTs belonged to the same taxonomic group, a lineage-specific event that added the $H_{3-4}C_4$ motif to the N-terminus of an ancestral Architaenioglossa MT duplicate would be the most parsimonious explanation.

Heterobranchia

From an MT perspective, this was the most studied Gastropoda clade. We collected 86 MT sequences (61 new) of 55 Heterobranchia species. MT multiplicity was observed in 22 species with two (14 species), three (seven species), or four (one species) MTs (supplementary fig. S2*D* and table S2, Supplementary Material online). Size diversity of Heterobranchia MTs was the highest within the Gastropoda class, with MTs ranging from 58 (AbaMT1, GscMT1, GcuMT1, GtrMT1, and LstMT1) to 319 (AbiMT4) amino acids, with a cysteine content of 15 (25%) and 85 cysteines (26.2%), respectively. Overall, Heterobranchia MTs were organized in a N-terminal 9-Cys $\beta3$ domain

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

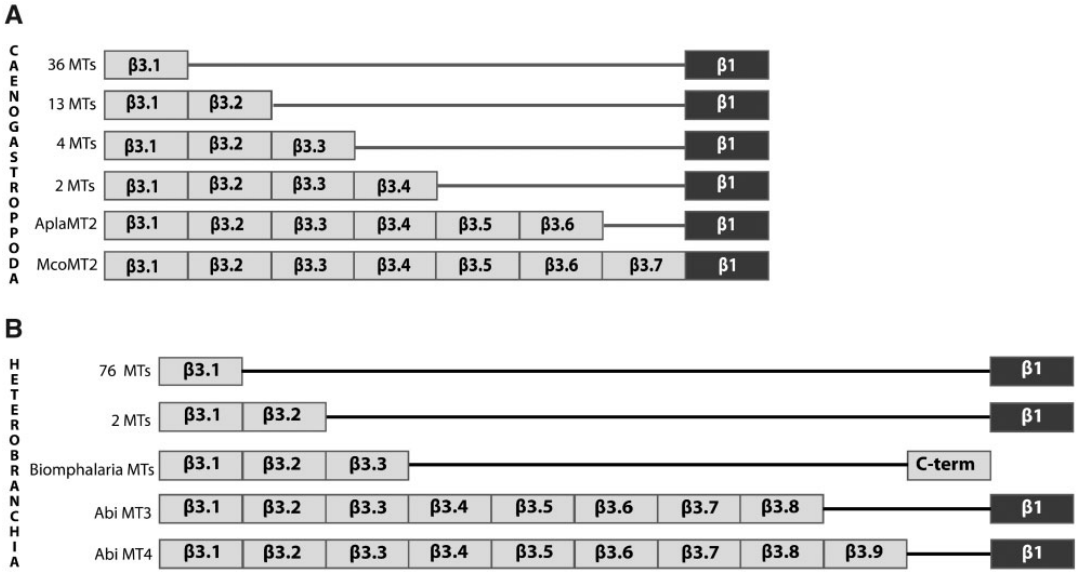


FIG. 4. Schematic representation of multimodular Gastropoda MTs. Internal duplications of the N-terminal β3 domain (gray box) generated multimodular MTs in Caenogastropoda (*A*) and Heterobranchia (*B*) (see supplementary fig. S2 and table S2, Supplementary Material online, for further details of species). A variable number of β3 domains, ranging from 1 (top) to 9 (bottom), is followed the by a conserved β1 domain (black box). In Heterobranchia, *Biomphalaria* MTs have three β3 domains followed by a C-terminal tail of five cysteines (C-term) but lack the conserved β1 domain.

(Cx_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C) linked by two residues to a C-terminal 9-Cys β1 domain ([CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]) (table 1 and supplementary fig. S2*D*, Supplementary Material online). The structure of Heterobranchia MTs was therefore of β3/β1 domains (fig. 1). There were, however, several exceptions to this structure. Species of Lymnaeidae, for instance, lacked the last [CxC] motif of the β1 domain, MTs of *Bradybaena similaris* and *Fiona pinnata* had an additional β3 domain (i.e., β3.1/β3.2) (supplementary fig. S2*D*, Supplementary Material online), and *Alinda biplicata* AbiMT3 and AbiMT4 had eight and nine β3 domains, respectively (Pedrini-Martha et al. 2020). But the most deviant Heterobranchia MTs were those of the *Biomphalaria* species, *B. glabrata* and *B. pfeifferi*. These MTs lacked the β1 domain and had three β3 domains (β3.1/β3.2/β3.3) followed by a C-terminal tail with five cysteines ($Cx_6Cx_5Cx_4CC$) (fig. 4*B* and supplementary fig. S3*B*, Supplementary Material online) (Niederwanger, Calatayud, et al. 2017). The fact that both species shared the same structure suggested that it could be a general MT feature for *Biomphalaria*.

Vetigastropoda

We collected 22 MT sequences (18 new) of 18 Vetigastropoda species (supplementary fig. S2*E* and table S2, Supplementary Material online). MT multiplicity was found in four species: *Clanculus pharaonius* (CphMT1 and CphMT2 partial), *Perotrochus lucaya* (PluMT1 and PluMT2), *Phasianella ventricosa* (PveMT1 and PveMT2), and *Prothalotia lehmanni* (PleMT1 and PleMT2). Overall, Vetigastropoda MTs were 65–71 amino acid long with 18 cysteines (25.4–27.7%) organized in a N-terminal 9-Cys β3 domain (Cx_3Cx_4[CxC]x_3[CxC]x_{3-4}[CxC]x_2C) linked by 3–6 residues to a C-terminal 9-Cys β1 domain ([CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]) (table 1 and supplementary fig. S2*E*, Supplementary Material online). The last common ancestor for Vetigastropoda thus likely had MTs with β3/β1 domains (fig. 1).

Comparisons across all of the phylogenetically diverse Gastropoda have suggested that the ancestral Gastropoda already had a β_3/β_1 MT (fig. 1), which underwent diverse lineage-specific modifications during evolution: 1) gene duplications leading to parallel MT multiplicity in divergent gastropod lineages, 2) internal domain duplications and losses in the Caenogastropoda and Heterobranchia taxa, and 3) acquisitions of novel modules such the H_{3-4}-C_4 motif in certain Caenogastropoda, and the new γ domain in Patellogastropoda.

New MTs in Other Mollusk Classes

Recent phylogenomic analyses have split Mollusks into two major clades (see fig. 1) (Kocot et al. 2011, 2020; Smith et al. 2011): Conchifera (including Gastropoda, Bivalvia, Cephalopoda, Monoplacophora, and Scaphopoda) and Aculifera (including Polyplacophora, Solenogastres, and Caudofoveata). We used publicly available molluskan Sequence Read Archives (SRA) projects to assemble new MTs that, together with previously reported ones, represented an extensive evolutionary list of mollusk MTs. Thus, our analysis has led us to identify for the first time MTs in Scaphopoda, Cephalopoda, and Monoplacophora classes, and the first MTs for chiton, solenogastres and caudofoveatan representatives of Aculifera.

Conchifera Clade

Scaphopoda. This molluskan class has been considered the sister group of Gastropods (Smith et al. 2011), though this phylogenetic relationship is still under debate. We identified

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

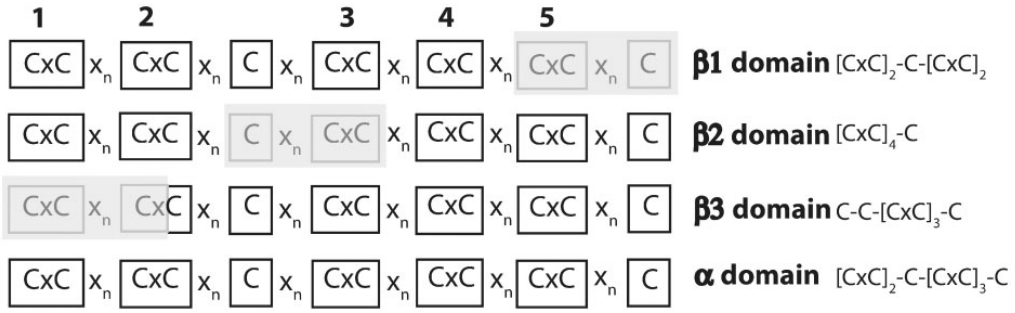


FIG. 5. Hypothetical relationship between α and β domains. Based on the configuration of the C and [CxC] motifs, a putative connection between α and β domains is hypothesized. The five [CxC] motifs and the two additional Cs distributed as $[CxC]_2$-C-$[CxC]_3$-C in the α domain appear to have been trimmed (shadowed) at the C-terminal end in β1 domain: $[CxC]_2$-C-$[CxC]_2$; at the middle in β2: $[CxC]_4$-C; or at the N-terminal end in β3: C-C-$[CxC]_3$-C.

two Scaphopoda MTs from *Antalis entalis* (AenMT1) and *Graptacme eborea* (GebMT1) species (supplementary fig. S2*F* and table S2, Supplementary Material online). Scaphopoda MTs were 68–72 amino acid long with 18–19 cysteines (26.4%) organized in a N-terminal 9-Cys β domain ($[CxC]x_5[CxC]x_3[CxC]x_{2-4}[CxC]x_4C$) linked by 8–10 residues to a C-terminal 9-Cys β1 domain ($[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$) (table 1 and supplementary fig. S2*F*, Supplementary Material online). Interestingly, the organization of the [CxC] motifs in the Scaphopoda N-terminal β domain, $[CxC]_4$-C, resembled that of β2 domain defined in the unconventional bivalve MTIIIs (Jenny et al. 2016), whereas it differed from that of Gastropoda β1 and β3 domains—$[CxC]_2$-C-$[CxC]_2$ and C-C-$[CxC]_3$-C, respectively—(table 1 and fig. 5). The structure of Scaphopoda MTs was therefore of $\beta2/\beta1$ domains (fig. 1).

Bivalvia. In order to have a broad perspective of the MTs within Bivalvia, we analyzed 62 MTs from 36 species (supplementary fig. S2*G* and table S2, Supplementary Material online). For clarity, and since the structural diversity of MTs in oysters and mussels has been extensively described elsewhere (Mackay et al. 1993; Jenny et al. 2004, 2016; Aceto et al. 2011), we focused our analysis on the MTI and MTIV within available species of Ostreidae, and on MT10, MT10B, and MT20 within Mytilidae, which have been considered to have the plesiomorphic condition for bivalve MTs (Nam and Kim 2017). Most Bivalvia MTs were 66–78 amino acid long with 21–25 cysteines (26.9–32%) organized in an N-terminal 12-Cys α domain ($[CxC]x_5[CxC]x_3Cx_{4-5}[CxC]x_3[CxC]x_{3-4}[CxC]x_2C$) linked by three residues to a C-terminal 9-Cys β1 domain ($[CxC]x_{3-6}[CxC]x_3Cx_5[CxC]x_{3-5}[CxC]$) (table 1 and supplementary fig. S2*G*, Supplementary Material online). The likely ancestral structure of these Bivalvia MTs was therefore of $\alpha/\beta1$ domains (fig. 1). Other Bivalvia MTs (e.g., *Crassostrea* MTIIIs in the order Ostreoidea and PmarMT2 of *Pinctada martensii* in the order Pterioida) were organized in two β2 domains. These MTs probably derived from an ancestral $\beta2/\beta1$ form (fig. 1). Deviations from these structures were found, ranging from loss or gain of some cysteines or small protein fragments (e.g., MquMT1 and CgiMTIV), to significant structural modifications, such as those previously described in a sphaeriid clam (*Pisidium coreanum*) PcorMT1 and the oyster (*Alectryonella plicatula*) ApliMT1: $\alpha/\beta1/\beta1$; in *Crassostrea* MTIIs: α_n or $\alpha/\beta1/\beta1$; or in the scallop (*Argopecten irradians*) AirMT1 and AirMT2: β2-like/β2-like/β2-like/β1 and β2-like/β2-like/β1, respectively (notice that organization of the [CxC] motifs in the β2-like domain was more similar to β3 than to β2 domain) (Tanguy and Moraga 2001; Baek et al. 2009; Jenny et al. 2016; Nam and Kim 2017).

Cephalopoda. We identified the MTs of four Cephalopoda species: *Nautilus pompilius* (NpoMT1), *Octopus vulgaris* (OvuMT1), *Octopus bimaculoides* (ObiMT1), and *Sepia esculenta* (SesMT1) (supplementary fig. S2*H* and table S2, Supplementary Material online). Cephalopoda MTs were 69–73 amino acid long with 19–21 cysteines (27.5–28.7%) organized in $\alpha/\beta1$ domains more similar to those of other molluskan classes in the *Nautilus* MT than in the other species. The N-terminal 12-Cys α domain ($[CxC]x_{5-6}[CxC]x_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$) in the *Nautilus* MT, but slightly divergent in the coleoid cephalopods ($[CxC]x_{5-6}[CxC]x_3Cx_4Cx_3[CxC]x_3[CxC]x_2C$ in *Sepia* and *Octopus* MTs), was linked by three or four residues to a C-terminal 9-Cys β1 domain ($[CxC]x_4[CxC]x_3Cx_5[CxC]x_3[CxC]$ in *Nautilus* and *Sepia* MTs, but somewhat modified in *Octopus* MTs; table 1 and supplementary fig. S2*H*, Supplementary Material online). Based on *Nautilus* and *Sepia* MTs, we concluded that the prototypical structure of Cephalopoda MTs was of $\alpha/\beta1$ domains (fig. 1).

Monoplacophora. We identified the first MT in a Monoplacophora species, *Laevipilina hyalina* (LhyMT1) (supplementary fig. S2*I* and table S2, Supplementary Material online). LhyMT1 was 63 amino acid long with 19 cysteines (30.2%) organized in an N-terminal β2 domain ($[CxC]x_6[CxC]x_3[CxC]x_2[CxC]xCx_2C$) linked by three residues to a C-terminal 9-Cys β1 domain ($[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$) (table 1 and supplementary fig. S2*I*, Supplementary Material online). The structure of the Monoplacophora MT was therefore of $\beta2/\beta1$ domains (fig. 1).

Aculifera Clade

Polyplacophora. We identified five MTs in Polyplacophora species: *Acanthochitona crinita* (AcriMT1), *Chaetopleura apiculata* (CapMT1), *Chiton olivaceus* (ColMT1), *Tonicella lineata*

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

(TliMT1), and *Leptochiton rugatus* (LruMT1) (supplementary fig. S2*J* and table S2, Supplementary Material online). Polyplacophora MTs were 70–73 amino acid long with 21 cysteines (28.8–30.0%) organized in a N-terminal 12-Cys α domain ([CxC]x_5[CxC]x_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C) linked by three residues to a C-terminal 9-Cys β1 domain ([CxC]x_{3-4}[CxC]x_3Cx_5[CxC]x_3[CxC]) (table 1 and supplementary fig. S2*J*, Supplementary Material online). The structure of Polyplacophora MTs was therefore of α/β1 domains (fig. 1).

Solenogastres. We identified eight MTs of Solenogastres species: *Alexandromenia crassa* (AcraMT1 partial), *Amphimeniidae* sp. (AspMT1 partial), *Micromenia fodiens* (MfoMT1 partial), *Neomenia carinata* (NcaMT1), *Neomenia megatrapezata* (NmegMT1 and NmegMT2), and *Neomeniomorpha* sp. (NspMT1 and NspMT2 partial) (supplementary fig. S2*K* and table S2, Supplementary Material online). Full-length Solenogastres MTs (i.e., NcaMT1, NmegMT1, and NspMT1) were 65–66 amino acid long with 18–19 cysteines (27.3–29.3%) organized in an N-terminal 9-Cys β2 domain ([CxC]x_5[CxC]x_3[CxC]x_{2-4}[CxC]x_{3-4}C) linked by six residues to a C-terminal 9-Cys β1 domain ([CxC]x_{3-4}[CxC]x_3Cx_5[CxC]x_3[CxC]) with an additional cysteine in NmegMT1 and NspMT1 sequences (table 1 and supplementary fig. S2*K*, Supplementary Material online). The structure of these MTs was therefore of β2/β1 domains (fig. 1). Interestingly, *Neomenia megatrapezata* species had a second MT, NmegMT2, whose N-terminal domain was a 12-Cys α domain ([CxC]x_5[CxC]x_3Cx_5[CxC]x_3[CxC]x_3[CxC]x_2C) linked by seven residues to a C-terminal 9-Cys β1 domain ([CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]) (table 1 and supplementary fig. S2*L*, Supplementary Material online). The NmegMT2 structure was, therefore, of the α/β1 type, meaning that two structurally different MTs coexisted in the same Solenogastres species (fig. 1).

Caudofoveata. We identified four MTs in Caudofoveata species: *Chaetoderma nitidulum* (CniMT1), *Falcidens caudatus* (FcaMT1), *Falcidens sagittiferus* (FsaMT1), and *Scutopus ventrolineatus* (SveMT1) (supplementary fig. S2*M* and table S2, Supplementary Material online). The Caudofoveata MTs were 73–78 amino acid long with 23 cysteines (29.5–31.5%) organized in a novel N-terminal 14-Cys domain ([CxC]x_3[CxC]CCx_4Cx_{3-7}[CxC]x_3CCx_4[CxC]C) that we named δ domain, linked by three residues to a C-terminal 9-Cys β1 domain ([CxC]x_{3-4}[CxC]x_3Cx_5[CxC]x_3[CxC]) (table 1 and supplementary fig. S2*M*, Supplementary Material online). The structure of Caudofoveata MTs was therefore of δ/β1 domains (fig. 1).

In summary, we have identified more than 270 MTs in 189 different species distributed across the eight molluskan classes, ≈64% as single copy sequences (supplementary table S2, Supplementary Material online; the existence of additional MTs in some species cannot be excluded because some sequence databases are still in progress). Our data showed that a single MT with a two-domain (i.e., bimodular) structure was the predominant type of MTs in most Mollusca species (fig. 1 and table 1). Our results revealed, however, many exceptions to this situation, with MT multiplicity and/or multimodular MTs in many Mollusca species, which denoted an intricate and dynamic evolutionary history of mollusk MTs.

Discussion

Functional Evolution of Mollusk MTs

Evolution of the Metal-Binding Capacity

Most mollusk MTs are bimodular proteins with 18 (in β_{2-3}/β_1 MTs), 19 (in γ/β_1), 21 (in α/β_1), or 23 (in δ/β_1) cysteines. Neritimorpha β_3/β_1 MTs render homometallic complexes with six divalent ions, either Zn(II) or Cd(II) (fig. 3), meaning that each 9-Cys domain is designed to allocate three divalent metal ions. This capacity is similar to those reported for other bimodular gastropod MTs (Perez-Rafael et al. 2012, 2014; Palacios et al. 2014; Dvorak et al. 2018) and agrees with 3D structural analysis of *Littorina littorea* (Baumann et al. 2017) and *Helix pomatia* MTs (Beil et al. 2019). The metal-binding capacity of Patellogastropoda γ/β_1 LgiMTs, both with 19 cysteines, is slightly higher since they bind seven divalent metal ions (fig. 2*A* and *B*), similar to the 21-Cys α/β_1 MTs of Bivalvia (Munoz et al. 2002; Orihuela et al. 2008). The successful synthesis of the γ domain of LgiMT2, with ten Cys residues, and the characterization of the species produced in Cd(II)-enriched media indicate that the LgiMTs render metal-aggregates containing seven Cd(II) ions because these MTs are capable of binding three Cd(II) ions in their β1 domain, whereas their γ domain allocates four Cd(II) ions (fig. 2*C*). The metal-binding capacity of MTs appears, therefore, to rely on the number and position of the cysteines in the different domains and, as expected, the higher the cysteine content, the higher the metal-binding capacity.

In that sense, an effective evolutionary strategy for increasing the metal-binding capacity of MTs has been the design of multimodular forms with high cysteine content and a high capacity of metal binding (Niederwanger, Dvorak et al. 2017; Palacios et al. 2017; Calatayud et al. 2018). Multimodular MTs had been identified in a few Bivalve and Gastropoda species (Tanguy and Moraga 2001; Jenny et al. 2004, 2016; Baumann et al. 2017; Nam and Kim 2017; Niederwanger, Calatayud et al. 2017; Palacios et al. 2017; Schmielau et al. 2019; Pedrini-Martha et al. 2020), and our results have increased this list with 16 new proteins containing a variable number of repeated domains. Sequence comparisons (fig. 4 and supplementary fig. S3, Supplementary Material online) and structural analysis (Baumann et al. 2017) indicate that these multimodular MTs originated by N-terminal duplications of the β3 domain, in agreement with the idea that proteins tend to increase in length mainly by the gain of sequences at the 5′-end of their genes (Toll-Riera and Alba 2013). The evolution of such multimodular MTs in some mollusk species suggests that diverse lineages have had to adapt to different conditions of metal bioavailability and stress, despite the ecophysiological determinants that have favored them in only some species remain, however, unknown.

Evolution of the Metal-Binding Preference

The evolution of MTs with different metal preferences has usually been associated with scenarios of MT multiplicity, in

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

which neofunctionalization processes yielded MT duplicates with new metal-binding selectivities. Sixty-seven species patchily distributed across the mollusk phylogeny possess at least two MTs (supplementary table S2, Supplementary Material online), and their sequence (supplementary fig. S2, Supplementary Material online) and domain conservation (fig. 1 and table 1) suggest that most of them originated after the splitting of the main Mollusk groups by lineage-specific duplications. For instance, a gene duplication in the ancestor of the Neritimorpha class resulted in two MTs that according to the analysis of *N. peloronta* NpeMT1 and NpeMT2 diverged in their metal preferences (i.e., a Zn-thionein character for NpeMT1 and a Cd-thionein nature for NpeMT2; fig. 3). In contrast, the duplication found in *L. gigantea* appears more recent since both duplicates are 96% identical (supplementary fig. S1, Supplementary Material online) and still share a Cd binding preference (fig. 2). Our results support the idea that metal-binding preference and hence, functional specificity, does not mainly rely on the number and position of the cysteines (>94% identical between NpeMT1 and NpeMT2) but on the nature of the noncoordinating amino acids (≈70% different between NpeMT1 and MpeMT2) distributed along the protein sequence (Palacios et al. 2011; Perez-Rafael et al. 2014; Dallinger et al. 2020), although we are still far from being able to predict metal preference based on the analysis of the noncoordinating amino acids.

We do not know the biological determinants that favored the evolution of MTs with different metal preferences in certain mollusk lineages, but the colonization of new habitats (Dallinger et al. 2020) and the emergence of physiological novelties (Dallinger et al. 2005; Höckner et al. 2011) have been proposed as significant evolutionary factors. We do not know either the binding selectivity of the ancestral mollusk MT, but the widespread cadmium-binding capacity of many MTs—not only in mollusk but in diverse marine animals (Narula et al. 1995; Riek et al. 1999; Valls et al. 2001; Guirola et al. 2012; Calatayud et al. 2018)—along with the ancient origin α and β domains (see below), lead us to speculate that ancestral MTs might have been a detoxification system that was later co-opted for homeostatic functions for essential metals. Other scenarios are possible but since Cd is a highly toxic metal because it competes for Zn-dependent cellular processes, and Cd is frequently found with Zn in ore deposits of the earth crust, an early evolution of Cd-detoxifying MTs could have conferred a significant physiological advantage to marine organisms. This advantage would be especially important after increased Cd levels during Paleozoic era (Dallinger et al. 2020), concomitantly with the Cambrian explosion and the emergence of most animal phyla. From the ancestral MT, different metal-selective MTs would have independently evolved in different molluskan lineages: Cu-selective MTs in Heterobranchia species (Höckner et al. 2011; Perez-Rafael et al. 2011; Palacios et al. 2014), Zn-selective forms in Neritimorpha gastropods (this work) and bivalves (Orihuela et al. 2008), and metal-unselective MTs in Vetigastropoda and Heterobranchia lineages (Höckner et al. 2011; Perez-Rafael et al. 2011, 2012, 2014; Niederwanger, Calatayud, et al. 2017). This functional diversification might be related to the extraordinary evolutionary success of the phylum, with species that have colonized and adapted to very diverse habitats around the world.

Structural Evolution of Mollusca MTs

Bimodular Structure of Mollusca MTs

Vertebrate MTs have a bimodular structure made of two independent functional domains—an 11–12-Cys α domain and a 9-Cys β domain—each one capable to bind metal ions (Braun et al. 1986; Capdevila et al. 1997; Cols et al. 1999). By comparison with vertebrates, the bimodular structure has been extended to gastropod (Palacios et al. 2011; Perez-Rafael et al. 2012, 2014; Dvorak et al. 2018; Beil et al. 2019) and bivalve MTs (Jenny et al. 2004, 2016; Nam and Kim 2017; Yingprasertchai et al. 2019). Our results spread the bimodular structure to the MTs of all molluskan classes since 90% of their MTs are two-domain proteins. In these bimodular MTs, an N-terminal domain that it is variable depending on the taxon (an α domain in Bivalvia, Cephalopoda, and Polyplacophora MTs; a $\beta 2$ domain in Scaphopoda, Monoplacophora, and Solenogastres MTs; a δ domain in Caudofoveata MTs; and a $\beta 3$ domain in Gastropoda MTs, with the exception of the γ domain in Patellogastropoda MTs) is linked to a conserved $\beta 1$ domain at the C-terminal region (fig. 1 and table 1). The pervasiveness of the C-terminal $\beta 1$ domain suggests a conserved role for this domain, probably related to the stabilization of the 3D cluster structure of the entire protein (Dallinger et al. 2020).

Modular Evolution of MTs: Ancient and Recent Domains

The origin and the evolutionary relationships of the distinct α, $\beta 1$, $\beta 2$, $\beta 3$, γ, and δ domains were intriguing. The finding of α, $\beta 1$, and $\beta 2$ domains in diverse molluskan classes together with the presence of α/β_1 and β_2/β_1 MTs in species of both Conchifera and Aculifera (fig. 1) supports that the origin of the domains predated the diversification of the phylum, which has estimated at more than 545 Ma (Kocot et al. 2020), and suggests that an ancient MT multiplicity was subsequently maintained or lost in a lineage-specific manner. In addition, a possible connection between α and β domains might be envisaged based on the number and configuration of the cysteine motifs (fig. 5). Thus, the three β domains might have derived from an ancestral α domain trimmed at the C-terminal end ($\beta 1$), at the middle ($\beta 2$), and at the N-terminal end ($\beta 3$), respectively. This possibility challenges the classical view that β domains represent ancestral forms, and that α domains evolved later to provide detoxification capacity in front of toxic metals such as cadmium (Cols et al. 1999).

In contrast to the ancient origin of α, $\beta 1$, and $\beta 2$ domains, the restricted distribution of $\beta 3$, γ, and δ domains in some mollusk lineages suggests a more recent origin, probably concomitant with the appearance of these taxonomic groups (fig. 1). Novel N-terminal domains would have replaced former and older ones specifically in some mollusk lineages: the $\beta 3$ domain in Gastropoda, the γ domain in Patellogastropoda, and the δ domain in Caudofoveata.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

Other scenarios of domain evolution cannot be ruled out, but they would require assuming complex processes of parallel or convergent evolution to justify the current distribution of the different domains throughout Mollusca.

De Novo Domain Evolution

Whereas there might be an evolutionary relationship between α and β domains, any link among these domains and the novel γ and δ domains is obscured by their differences. Although it cannot be ruled out that γ and δ domains derived from α or β domains that have diverged too much for homology to be recognized, the sequence and cysteine motifs in the γ and δ domains are so divergent from those in α and β domains that they seem to have evolved de novo. Recent analyses have shown that, indeed, de novo evolution is more frequent than previously thought (Neme and Tautz 2013; Toll-Riera and Alba 2013; Weisman and Eddy 2017; Levy 2019). Studies about the emergence of novel domains in human proteins, for instance, have revealed more than 400 "young" domains, 164 of which are found combined with older ones and preferentially located at the N-terminus of the proteins (Toll-Riera and Alba 2013). These new domains are rich in low-complexity sequences (Toll-Riera et al. 2012) and tend to be structurally disordered (Moore and Bornberg-Bauer 2012). Such structural features match well those of MTs, which are considered as low complexity and intrinsically disordered proteins. Thus, de novo emergence of MT domains might be relatively easy under an evolutionary perspective because the only requirement for a peptide to function as a metal ion chelator would be a high content of coordinating residues (e.g., cysteines) and a relative small length that favored the polypeptide folding (Capdevila and Atrian 2011). De novo evolution of MTs has been, indeed, implicitly stated from diverse evolutionary studies concluding that MTs likely evolved more than once in different animal phyla (Capdevila and Atrian 2011; Blindauer 2014; Isani and Carpene 2014; Ziller and Fraissinet-Tachet 2018).

In summary, the evolution of the mollusk MTs is intriguing. At the short term (at low taxonomic ranks), it appears to have followed the habitual evolutionary patterns based on progressive accumulation of changes in the sequence, and on duplications or losses of internal domains or genes. At the long term (at high taxonomic ranks), in contrast, MT evolution seems to have been mainly impacted by emergence of new structural domains. The modular structure of mollusk MTs makes the analyses of their domain organization and cysteine motifs more informative for inferring their evolution during the diversification of the phylum than the classic comparisons of sequences, which may be biased toward a general cysteine-richness due to the structural and functional requirements of metal binding.

Materials and Methods

Database Searches and MT Identification

Molluskan MT sequences were identified from public databases by Entrez searches using "metallothionein" and "Mollusca" as queries. Retrieved MT sequences were then used as queries in TBlastN searches in expressed sequence tags and genomic NCBI and eSnail (http://soft.bioinfo-minzhao.org/esnail/index.html) databases. In addition, RNA-SRA for each mollusk species deposited in NCBI were Blast searched using as queries MT sequences from the nearest phylogenetically species as well as from different mollusk species covering all the major clades. Raw sequence data were retrieved from the SRA and assembled using SeqMan 8.0.2 (Pro Assembler) software from the DNASTAR Lasergene package, and manually inspected in order to reconstruct new MT sequences. The MT nature of each new identified sequence was evaluated by BlastX searches against metazoan NCBI nonredundant protein sequence database. The amino acid sequences and the accession numbers of the retrieved MTs are provided in supplementary table S2, Supplementary Material online.

Characterization of *L. gigantea* MT Genes

Collection of three *L. gigantea* specimens, dissection of their hepatopancreas, RNA extraction, and storage in RNA later (Thermo Fisher Scientific, Waltham, CA) were performed by one of us (D.J.E.). Genomic DNA was obtained from hepatopancreatic tissue disrupted with the TissueLyser II (Qiagen, Hilden, Germany) procedure, and following the manufacturer's instructions of DNeasy Plant Mini Kit (Qiagen) for DNA extraction. The DNA concentration, purity, and integrity were checked using Tecan Infinite M200 (Tecan Group Ltd, Switzerland) measuring the absorbance at 260 and 280 nm. For the RNA extraction, tissue of the midgut gland was homogenized with glass beads using the Precellys homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France). Total RNA was isolated applying the RNeasy Plant Mini Kit (Qiagen) including on-column DNase I digestion (Qiagen) according to manufacturer's constructions. RNA integrity was checked by visualization on a 1.5% agarose gel (Biozym, Hessisch Oldendorf, Germany). For cDNA synthesis, 450 ng total RNA was used applying the RevertAid Reverse Transcriptase (Fermentas by Thermo Fisher Scientific).

The genomic region containing the putative *MT* genes of *L. gigantea* was PCR amplified, cloned PCR primers (supplementary table S3, Supplementary Material online) were designed based on the *L. gigantea* genome project (Scaffold and resequenced. 35, NW_008709190.1) in order to amplify overlapping fragments of different sizes (from 500 bp to 4 kb), covering the entire putative MT-containing genomic region. For each PCR reaction, 1 ng of genomic DNA was amplified using selected pairs of primers and the Phusion High-Fidelity DNA Polymerase (Invitrogen, Thermo Fisher Scientific) in a final 25 μl reaction. PCR conditions were 98 °C 30 s (s); 35 cycles of 98 °C 10 s, 58 °C 30 s, and 72 °C 3 min; and 72 °C 10 min. PCR products were visualized in 0.7% agarose gels, isolated with the GelElute Plasmid Miniprep Kit (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany), and cloned with TOPO TA Cloning Kit (Invitrogen, Thermo Fisher Scientific). Plasmid DNA was purified from bacteria using the GeneElutet Plasmid Miniprep Kit (Sigma-Aldrich), screened for insert presence by digestion with *Eco*RI (*Eco*RI Fast Digest Restriction Enzyme, Invitrogen, Thermo Fisher

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

Scientific), and sequenced at the Scientific and Technological Centers of the University of Barcelona using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) in an automatic sequencer (ABIPRISM 310, Applied Biosystems). Several internal primers (supplementary table S3, Supplementary Material online) were used to fully sequence the overlapping PCR fragments, which were manually assembled in order to reconstruct the entire genomic region.

The prediction of the *L. gigantea* MT genes was corroborated by the PCR amplification of the corresponding cDNAs using gene specific primers (supplementary table S3, Supplementary Material online) designed with CLC main workbench (Version 6.9), and Advantage 2 polymerase (Clontech, Takara Bio Europe, Saint-Germain-en-Laye, France) in a final 50 μl reaction. Cycling conditions were 95 °C 1 min; 30 cycles 95 °C 30 s, 53 °C (for LgiMT2)/55.5 °C (for LgiMT1) 30 s, and 68 °C 40 s; and 68 °C 5 min. PCR products were visualized on a 1.5% agarose gels, purified using the QIAquick gel extraction kit (Qiagen) and cloned with the TOPO TA Cloning Kit (Invitrogen, Thermo Fisher Scientific). Insert containing plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen) and sent for sequencing to Microsynth (Balgach, Switzerland). Sequences were analyzed via CLC main workbench (Version 6.9).

Production and Purification of Recombinant of Metal-MT Complexes

Synthetic cDNAs codifying the selected MTs (i.e., LgiMT1, LgiMT2, LgiMT2-γ domain, NpeMT1, and NpeMT2) were provided by Synbiotech (Monmouth Junction, NJ, USA) cloned in the pGEX-4T-1 expression vector (GE Healthcare). Recombinant plasmids were transformed in *E. coli* BL21 strain, a protease-deficient strain used for heterologous protein expression. For heterologous protein production, 500 ml of Luria-Bertani (LB) medium with 100 μg/ml ampicillin was inoculated with *E. coli* BL21 cells transformed with the corresponding recombinant plasmid. After overnight growth at 37 °C/250 rpm, the cultures were used to inoculate 5 l of fresh LB-100 μg/ml ampicillin medium. Gene expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside for 3 h. After the first 30 min of induction, cultures were supplemented with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM), or $CuSO_4$ (500 μM) in order to generate metal-MT complexes. Cells were harvested by centrifugation for 5 min at 9,100 × g (7,700 rpm), and bacterial pellets were suspended in 125 ml of ice-cold phosphate-buffered saline (PBS) (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4, and 0.5% v/v β-mercaptoethanol). Resuspended cells were sonicated (Sonifier Ultrasonic Cell Disruptor) 8 min at voltage 6 with pulses of 0.6 s, and then centrifuged for 40 min at 17,200 × g (12,000 rpm) and 4 °C. Soluble protein extracts containing GST-MT fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. GST-MT fusion proteins bound to the sepharose beads were washed with 30 ml of cold 1× PBS bubbled with argon to prevent oxidation. After three washes, GST-MT fusion proteins were digested with thrombin (GE Healthcare, 25 U/l of culture or SERVA, 25 U/l of culture) overnight at 17 °C, thus enabling separation of the metal-MT complexes from the GST that remained bound to the sepharose matrix. The eluted metal-MT complexes were concentrated with a 3-kDa Centripep Low Concentrator (Amicon, Merck) and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris–HCl, pH 7.0. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at −80 °C until use.

Analysis of Metal-MT Complexes

Protein quantification and element composition of all the samples were achieved by ICP-AES measurements performed in a Optima 4300DV (Perkin-Elmer, MA, USA) apparatus (S, 182.040 nm; Zn, 213.856 nm; Cd, 228.802 nm; and Cu, 324.803 nm) under conventional conditions following an already established method (Bongers et al. 1988).

Molecular weights were determined by ESI-MS, in a MicroTof-Q instrument (Bruker Daltonics Gmbh, Bremen, Germany) connected to a Series 1100 HPLC pump (Agilent Technologies) controlled by the Compass Software. The instrument was calibrated with ESI-L Low Concentration Turning Mix (Agilent Technologies, Santa Clara, CA). Metallated forms were detected under native conditions: 20 µl of sample injected through a PEEK tube at 30–50 µl min^{-1} in a 3.5–5.0-kV capillary-counter voltage, at 90–110 °C of desolvation temperature, and with dry gas at 6 l min^{-1}. Spectra were recorded between a m/z range from 800 to 3,000. The liquid carrier was a 90:10 mixture of 15 mM ammonium acetate and acetonitrile at pH 7.0. All molecular masses were calculated according to the bibliography (Fabris et al. 1996).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Parts of this study were financed by a cooperation grant to R.D. from the Austrian Science Fund (DACH grant No I 1482-N28). Researchers from Barcelona want to acknowledge the Spanish Ministerio de Ciencia e Innovación and FEDER for the projects BIO2015-67358-C2-2-P (M.C. and O.P.) and BIO2015-67358-C2-1-P (R.A.). M.C. and O.P. are members of the "Grup de Recerca de la Generalitat de Catalunya," ref. 2017SGR-864, and R.A. of ref. 2017SGR-1665. M.G.-R. acknowledges to the UAB the PIF grant. D.J.E. acknowledges support from the United States National Science Foundation (Grant No. DEB-1355230). We also thank the Servei d'Anàlisi Química (SAQ) at the Universitat Autònoma de Barcelona (ICP-AES, ESI-MS) for allocating instrument time, the Centres Científics i Tecnològics (CCiT) at the Universitat de Barcelona for DNA sequencing, and to Sebastian Artime for the experimental support.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

Data Availability

The data underlying this article are available in NCBI (https://www.ncbi.nlm.nih.gov/) and eSnail (http://soft.bioinfo-minzhao.org/esnail/index.html) databases. The amino acid sequences and the accession numbers of the mollusk MTs are available in supplementary table S2, Supplementary Material online.

References

Aceto S, Formisano G, Carella F, De Vico G, Gaudio L. 2011. The metallothionein genes of *Mytilus galloprovincialis*: genomic organization, tissue expression and evolution. *Mar Genomics*. 4(1):61–68.

Adamo GM, Lotti M, Tamas MJ, Brocca S. 2012. Amplification of the CUP1 gene is associated with evolution of copper tolerance in *Saccharomyces cerevisiae*. *Microbiology* 158(9):2325–2335.

Baek M-K, Lee J-S, Kang S-W, Lee J-B, Kang H-J, Jo YH, Noh M-Y, Han Y, Choi S-H, Chae S-H, et al. 2009. Phylogenetic analysis based on metallothionein gene sequence of an indigenous species *Pisidium* (*Neopisidium*) *coreanum* in Korea. *Korean J Malacol*. 25(2):153–160.

Baumann C, Beil A, Jurt S, Niederwanger M, Palacios O, Capdevila M, Atrian S, Dallinger R, Zerbe O. 2017. Structural adaptation of a protein to increased metal stress: NMR structure of a marine snail metallothionein with an additional domain. *Angew Chem Int Ed Engl*. 56(16):4617–4622.

Beil A, Jurt S, Walser R, Schonhut T, Guntert P, Palacios O, Atrian S, Capdevila M, Dallinger R, Zerbe O. 2019. The solution structure and dynamics of Cd-metallothionein from *Helix pomatia* reveal optimization for binding Cd over Zn. *Biochemistry* 58(45):4570–4581.

Blindauer CA. 2014. Metallothioneins. In: Binding, transport and storage of metal ions in biological cells. Cambridge, United Kingdom: The Royal Society of Chemistry. p. 594–653.

Bongers J, Walton CD, Richardson DE, Bell JU. 1988. Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma atomic emission spectrometric determination of sulfur. *Anal Chem*. 60(24):2683–2686.

Braun W, Wagner G, Worgotter E, Vasak M, Kagi JH, Wuthrich K. 1986. Polypeptide fold in the two metal clusters of metallothionein-2 by nuclear magnetic resonance in solution. *J Mol Biol*. 187(1):125–129.

Calatayud S, Garcia-Risco M, Rojas NS, Espinosa-Sanchez L, Artime S, Palacios O, Cañestro C, Albalat R. 2018. Metallothioneins of the urochordate *Oikopleura dioica* have Cys-rich tandem repeats, large size and cadmium-binding preference. *Metallomics* 10(11):1585–1594.

Capdevila M, Atrian S. 2011. Metallothionein protein evolution: a miniassay. *J Biol Inorg Chem*. 16(7):977–989.

Capdevila M, Bofill R, Palacios Ò, Atrian S. 2012. State-of-the-art of metallothioneins at the beginning of the 21st century. *Coord Chem Rev*. 256(1–2):46–62.

Capdevila M, Cols N, Romero-Isart N, Gonzalez-Duarte R, Atrian S, Gonzalez-Duarte P. 1997. Recombinant synthesis of mouse Zn3-beta and Zn4-alpha metallothionein 1 domains and characterization of their cadmium(II) binding capacity. *Cell Mol Life Sci*. 53(8):681–688.

Catalan A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J. 2016. An indel polymorphism in the MtnA 3′ untranslated region is associated with gene expression variation and local adaptation in Drosophila melanogaster. *PLoS Genet*. 12(4):e1005987.

Cols N, Romero-Isart N, Bofill R, Capdevila M, Gonzalez-Duarte P, Gonzalez-Duarte R, Atrian S. 1999. In vivo copper- and cadmium-binding ability of mammalian metallothionein beta domain. *Protein Eng*. 12(3):265–269.

Costa D, Marien J, Janssens TK, van Gestel CA, Driessen G, Sousa JP, van Straalen NM, Roelofs D. 2012. Influence of adaptive evolution of cadmium tolerance on neutral and functional genetic variation in *Orchesella cincta*. *Ecotoxicology* 21(7):2078–2087.

Cunha TJ, Giribet G. 2019. A congruent topology for deep gastropod relationships. *Proc Biol Sci*. 286(1898):20182776.

Dallinger R, Chabicovsky M, Hödl E, Prem C, Hunziker P, Manzl C. 2005. Copper in *Helix pomatia* (Gastropoda) is regulated by one single cell type: differently responsive metal pools in rhogocytes. *Am J Physiol Regul Integr Comp Physiol*. 289(4):R1185–R1195.

Dallinger R, Zerbe O, Baumann C, Egger B, Capdevila M, Palacios Ò, Albalat R, Calatayud S, Ladurner P, Schlick-Steiner BC, et al. 2020. Metallomics reveal a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins. *Metallomics* 12(5):702–720.

de Francisco P, Martin-Gonzalez A, Turkewitz AP, Gutierrez JC. 2017. Extreme metal adapted, knockout and knockdown strains reveal a coordinated gene expression among different *Tetrahymena thermophila* metallothionein isoforms. *PLoS One* 12(12):e0189076.

de Francisco P, Martin-Gonzalez A, Turkewitz AP, Gutierrez JC. 2018. Genome plasticity in response to stress in *Tetrahymena thermophila*: selective and reversible chromosome amplification and paralogous expansion of metallothionein genes. *Environ Microbiol*. 20(7):2410–2421.

Dvorak M, Lackner R, Niederwanger M, Rotondo C, Schnegg R, Ladurner P, Pedrini-Martha V, Salvenmoser W, Kremser L, Lindner H, et al. 2018. Metal binding functions of metallothioneins in the slug *Arion vulgaris* differ from metal-specific isoforms of terrestrial snails. *Metallomics* 10(11):1638–1654.

Fabris D, Zaia J, Hathout Y, Fenselau C. 1996. Retention of thiol protons in two classes of protein zinc ion coordination centers. *J Am Chem Soc*. 118(48):12242–12243.

Faddeeva-Vakhrusheva A, Derks MF, Anvar SY, Agamennone V, Suring W, Smit S, van Straalen NM, Roelofs D. 2016. Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan *Orchesella cincta*. *Genome Biol Evol*. 8(7):2106–2117.

Guirola M, Perez-Rafael S, Capdevila M, Palacios O, Atrian S. 2012. Metal dealing at the origin of the *Chordata* phylum: the metallothionein system and metal overload response in amphioxus. *PLoS One* 7(8):e43299.

Höckner M, Stefanon K, de Vaufleury A, Monteiro F, Pérez-Rafael S, Palacios O, Capdevila M, Atrian S, Dallinger R. 2011. Physiological relevance and contribution to metal balance of specific and non-specific metallothionein isoforms in the garden snail. *Biometals* 24(6):1079–1092.

Isani G, Carpene E. 2014. Metallothioneins, unconventional proteins from unconventional animals: a long journey from nematodes to mammals. *Biomolecules* 4(2):435–457.

Janssens TKS, Lopéz RDR, Mariën J, Timmermans MJTN, Montagne-Wajer K, van Straalen NM, Roelofs D. 2008. Comparative population analysis of metallothionein promoter alleles suggests stress-induced microevolution in the field. *Environ Sci Technol*. 42(10):3873–3878.

Janssens TKS, Roelofs D, Van Straalen NM. 2009. Molecular mechanisms of heavy metal tolerance and evolution in invertebrates. *Insect Sci*. 16(1):3–18.

Jenny MJ, Payton SL, Baltzegar DA, Lozier JD. 2016. Phylogenetic analysis of molluscan metallothioneins: evolutionary insight from *Crassostrea virginica*. *J Mol Evol*. 83(3–4):110–125.

Jenny MJ, Ringwood AH, Schey K, Warr GW, Chapman RW. 2004. Diversity of metallothioneins in the American oyster, *Crassostrea virginica*, revealed by transcriptomic and proteomic approaches. *Eur J Biochem*. 271(9):1702–1712.

Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, et al. 2011. Phylogenomics reveals deep molluscan relationships. *Nature* 477(7365):452–456.

Kocot KM, Poustka AJ, Stoger I, Halanych KM, Schrodl M. 2020. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci Rep*. 10(1):101.

Levy A. 2019. How evolution builds genes from scratch. *Nature* 574(7778):314–316.

Loebus J, Peroza EA, Bluthgen N, Fox T, Meyer-Klaucke W, Zerbe O, Freisinger E. 2011. Protein and metal cluster structure of the wheat metallothionein domain gamma-E(c)-1: the second part of the puzzle. *J Biol Inorg Chem*. 16(5):683–694.

Mackay EA, Overnell J, Dunbar B, Davidson I, Hunziker PE, Kagi JH, Fothergill JE. 1993. Complete amino acid sequences of five dimeric

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

and four monomeric forms of metallothionein from the edible mussel *Mytilus edulis*. *Eur J Biochem*. 218(1):183–194.
Maroni G, Wise J, Young JE, Otto E. 1987. Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster*. *Genetics* 117(4):739–744.
Moore AD, Bornberg-Bauer E. 2012. The dynamics and evolutionary potential of domain loss and emergence. *Mol Biol Evol*. 29(2):787–796.
Munoz A, Forsterling FH, Shaw CF 3rd, Petering DH. 2002. Structure of the (113)Cd(3)beta domains from *Homarus americanus* metallothionein-1: hydrogen bonding and solvent accessibility of sulfur atoms. *J Biol Inorg Chem*. 7(7–8):713–724.
Nam YK, Kim EJ. 2017. Diversification and domain evolution of molluskan metallothioneins: a mini review. *Fish Aquatic Sci*. 20(1):8.
Narula SS, Brouwer M, Hua Y, Armitage IM. 1995. Three-dimensional solution structure of *Callinectes sapidus* metallothionein-1 determined by homonuclear and heteronuclear magnetic resonance spectroscopy. *Biochemistry* 34(2):620–631.
Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics*. 14(1):117.
Niederwanger M, Calatayud S, Zerbe O, Atrian S, Albalat R, Capdevila M, Palacios O, Dallinger R. 2017. *Biomphalaria glabrata* metallothionein: lacking metal specificity of the protein and missing gene upregulation suggest metal sequestration by exchange instead of through selective binding. *Int J Mol Sci*. 18(7):1457.
Niederwanger M, Dvorak M, Schnegg R, Pedrini-Martha V, Bacher K, Bidoli M, Dallinger R. 2017. Challenging the metallothionein (MT) gene of *Biomphalaria glabrata*: unexpected response patterns due to cadmium exposure and temperature stress. *Int J Mol Sci*. 18(8):1747.
Nielson KB, Winge DR. 1985. Independence of the domains of metallothionein in metal binding. *J Biol Chem*. 260(15):8698–8701.
Orihuela R, Domenech J, Bofill R, You C, Mackay EA, Kagi JH, Capdevila M, Atrian S. 2008. The metal-binding features of the recombinant mussel *Mytilus edulis* MT-10-IV metallothionein. *J Biol Inorg Chem*. 13(5):801–812.
Palacios O, Jimenez-Marti E, Niederwanger M, Gil-Moreno S, Zerbe O, Atrian S, Dallinger R, Capdevila M. 2017. Analysis of metal-binding features of the wild type and two domain-truncated mutant variants of *Littorina littorea* metallothionein reveals its Cd-specific character. *Int J Mol Sci*. 18(7):1452.
Palacios O, Pagani A, Pérez-Rafael S, Egg M, Höckner M, Brandstätter A, Capdevila M, Atrian S, Dallinger R. 2011. Shaping mechanisms of metal specificity in a family of metazoan metallothioneins: evolutionary differentiation of mollusc metallothioneins. *BMC Biol*. 9(1):4.
Palacios O, Perez-Rafael S, Pagani A, Dallinger R, Atrian S, Capdevila M. 2014. Cognate and noncognate metal ion coordination in metal-specific metallothioneins: the *Helix pomatia* system as a model. *J Biol Inorg Chem*. 19(6):923–935.
Pedrini-Martha V, Koll S, Dvorak M, Dallinger R. 2020. Cadmium uptake, MT gene activation and structure of large-sized multi-domain metallothioneins in the terrestrial door snail *Alinda biplicata* (Gastropoda, Clausiliidae). *Int J Mol Sci*. 21(5):1631.
Perez-Rafael S, Atrian S, Capdevila M, Palacios O. 2011. Differential ESI-MS behaviour of highly similar metallothioneins. *Talanta* 83(3):1057–1061.
Perez-Rafael S, Mezger A, Lieb B, Dallinger R, Capdevila M, Palacios O, Atrian S. 2012. The metal binding abilities of *Megathura crenulata* metallothionein (McMT) in the frame of gastropoda MTs. *J Inorg Biochem*. 108:84–90.
Perez-Rafael S, Monteiro F, Dallinger R, Atrian S, Palacios O, Capdevila M. 2014. *Cantareus aspersus* metallothionein metal binding abilities: the unspecific CaCd/CuMT isoform provides hints about the metal preference determinants in metallothioneins. *Biochim Biophys Acta* 1844(9):1694–1707.
Purać J, Nikolić TV, Kojić D, Celić AS, Plavša JJ, Blagojević DP, Petri ET. 2019. Identification of a metallothionein gene in honey bee *Apis mellifera* and its expression profile in response to Cd, Cu and Pb exposure. *Mol Ecol*. 28(4):731–745.
Riek R, Precheur B, Wang Y, Mackay EA, Wider G, Guntert P, Liu A, Kagi JH, Wuthrich K. 1999. NMR structure of the sea urchin (*Strongylocentrotus purpuratus*) metallothionein MTA. *J Mol Biol*. 291(2):417–428.
Schmielau L, Dvorak M, Niederwanger M, Dobieszewski N, Pedrini-Martha V, Ladurner P, Pedregal JR, Marechal JD, Dallinger R. 2019. Differential response to cadmium exposure by expression of a two and a three-domain metallothionein isoform in the land winkle *Pomatias elegans*: valuating the marine heritage of a land snail. *Sci Total Environ*. 648:561–571.
Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SC, Rouse GW, Giribet G, Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480(7377):364–367.
Tanguy A, Moraga D. 2001. Cloning and characterization of a gene coding for a novel metallothionein in the Pacific oyster *Crassostrea gigas* (CgMT2): a case of adaptive response to metal-induced stress? *Gene* 273(1):123–130.
Timmermans MJ, Ellers J, Roelofs D, van Straalen NM. 2005. Metallothionein mRNA expression and cadmium tolerance in metal-stressed and reference populations of the springtail *Orchesella cincta*. *Ecotoxicology* 14(7):727–739.
Tio L, Villarreal L, Atrian S, Capdevila M. 2004. Functional differentiation in the mammalian metallothionein gene family: metal binding features of mouse MT4 and comparison with its paralog MT1. *J Biol Chem*. 279(23):24403–24413.
Toll-Riera M, Alba MM. 2013. Emergence of novel domains in proteins. *BMC Evol Biol*. 13(1):47.
Toll-Riera M, Rado-Trilla N, Martys F, Alba MM. 2012. Role of low-complexity sequences in the formation of novel protein coding sequences. *Mol Biol Evol*. 29(3):883–886.
Valls M, Bofill R, Gonzalez-Duarte R, Gonzalez-Duarte P, Capdevila M, Atrian S. 2001. A new insight into metallothionein (MT) classification and evolution. The in vivo and in vitro metal binding features of *Homarus americanus* recombinant MT. *J Biol Chem*. 276(35):32835–32843.
Weisman CM, Eddy SR. 2017. Gene evolution: getting something from nothing. *Curr Biol*. 27(13):R661–R663.
Yingprasertchai T, Yu RMK, Tran TKA, Chong Kong RY, O'Connor WA, MacFarlane GR. 2019. Characterisation of the metallothionein gene in the Sydney rock oyster and its expression upon metal exposure in oysters with different prior metal exposure histories. *Mar Environ Res*. 151:104775.
Zapata F, Wilson NG, Howison M, Andrade SC, Jorger KM, Schrodl M, Goetz FE, Giribet G, Dunn CW. 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc Biol Sci*. 281(1794):20141739.
Zhang C, Jansen M, De Meester L, Stoks R. 2018. Thermal evolution offsets the elevated toxicity of a contaminant under warming: a resurrection study in *Daphnia magna*. *Evol Appl*. 11(8):1425–1436.
Zhang C, Jansen M, De Meester L, Stoks R. 2019. Rapid evolution in response to warming does not affect the toxicity of a pollutant: insights from experimental evolution in heated mesocosms. *Evol Appl*. 12(5):977–988.
Ziller A, Fraissinet-Tachet L. 2018. Metallothionein diversity and distribution in the tree of life: a multifunctional protein. *Metallomics* 10(11):1549–1559.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa230/5904274 by guest on 28 October 2020

ARTICLE I: SUPPLEMENTARY MATERIAL

MODULARITY IN PROTEIN EVOLUTION: MODULAR ORGANIZATION AND DE NOVO DOMAIN EVOLUTION IN MOLLUSK METALLOTHIONEINS

The file includes:

Table S1

Table S2

Table S3

Figure S1

Figure S2

Figure S3

Table S1. EST Sequences similar to LgiMT1 and LgiMT2.

LgiMT1 GROUP	LgiMT2 GROUP
FC571967.1_2 CAWG692.rev	FC699555.1_3 CAXX7741.fwd
FC571968.1_2 CAWG692.fwd	FC609649.1_3 CAXS4555.fwd
FC616861.1_5 CAXS8508.rev	FC618816.1_3 CAXS9646.fwd
FC603045.1_2 CAXS14463.fwd	FC784314.1_1 CBGC14896.fwd
FC557237.1_2 CAWC462.rev	
FC804908.1_3 CBGC6779.fwd	
FC557238.1_2 CAWC462.fwd	
FC616862.1_2 CAXS8508.fwd	

Table S2. Mollusca MT sequences analyzed in this study.

Class	Clade	Species	MT name	Previous name	Accession number*	Sequence
Conchifera						
Gastropoda	Patellogastropoda	Cellana rota	CroMt1		(1)	MPSEKASCCLAEYDCCKTKSCCKEGPSNCCSDDNPXNCCK2VCNCXGTCACGVGCQGIESCQCPSTCSCK
		Lottia digitalis	LdiMT1		(1)	CLAELECCKTKSCCATGPADCCKPGNKSDCCTPGKTQCKCAGSCACGAGCTGEGNCKCGPNCSCK
		Lottia gigantea	LgiMT1	LgMT1[1]	MK770430	MSSEKPSCCIAEYECCKTKLCCDTGPADCCKPGNKPDCCAPGKLQCKCPGTCACGVGCTGVDNCKCGSGCSCFN
			LgiMT2	LgMT2 [1]	MK770431	MSSEKASCCIAEYECCKTKSCCDTGPADCCKPGNKPDCCAPGKLQCKCSGTCACGVGCTGVDNCKCGAGCSCFN
		Lottia kogamogai	LkaMT1		(1)	MSSEKAACCVADYECCKTKSCCENGPEDCCKPGNKSDCCGPSKNQCKCSGTCACGVGCTGQDNCKCGPNCSCK
		Lottia scutum	LscMT1		(1)	CLAEYECCKTKSCCATGPADCCKPGNKAANCCAPGKNQCKCDGTCGCGVGCTGQXNCKCGPXCSCK
		Nipponacmea fuscoviridis	NfuMT1		(1)	MSSDKAACCVAEYECCKTKSCCANGPEDCCEPGNKSDCCAPGKNQCKCAGSCACGVGCTGQGNCKCGSDCSCK
		Nacella concinna	NcoMT1		(1)	MSSEKAACCIAEYECCKTKSCCKDGPADCCKPGNTTDCCKGKVACKCAGSCACGAGCTGQTPCKCGAGCSCNS
		Patelloida pygmaea	PpyMT1		(1)	MSSEKAACCLAERECCKTKSCCATGPADCCQPGNKSDCCAPGKTQCKCSGSCACGVGCTGEGNCKCGPNCSCK
		Patella vulgata	PvuMT1	MT1 [2]	MK443472	MSSQKASCCLAELECCKTKACCAKGPANCCSPGNDPNCCKSNICKCNGNCACGVGCTGIENCECGTGCSCK
			PvuMT2	MT2 [2]	MK443472	MSSEKAACCLAEHECCKTKSCCANGPADCCKPGKTVDCCKSQNTCKCGESCACGAGCSGVDNCKCGSGCSCK
	Neritimorpha	Clithon retropictum	CretrMT1		(1)	MSDPKGTSCTAECKCDPCNCGTNCKCGSDCTCSSCKKSCKCSGTCDCGKGCTGPDNCSCSTGCSCR
			CretrMT2		(1)	MSGKGPNCTEACKGDPCQCGDNCQCGDGCSCNSCKSCKCSSEGCKCGVGCKGPGTCKCDSSCSCK
		Nerita albicilla	NalMT1		(1)	MSDPKGASCTTECKCNPCACGANCKCAGDCSCNNCNKSSCLCPGTCSCGKGCTGVSTCKCDSSCSCR
			NalMT2		(1)	MPRPKGPGCTEPCVVAEDCQCDTNCKCAKECPCMDCHKASCKCTGACDCGVGCTGPVTCKCEEDCTCH
		Nerita melanotragus	NmelMT1		(1)	MSDPKGASCTTECKCNPCACGTNCKCGTDCTCSXCNKSSCKCAGPCACGKGCTTTATCKCDSGCSCR
			NmelMT2		(1)	MPHPKGPSCTEDCKAVECQCGTNCXCSRECPCKDCHKATCKCSGSCACGEGCSGPQTCKCEEDCSCH
		Nerita peloronta	NpeMT1		(1)	MSDPKGASCTTECKCDPCACGTNCKCGSDCTCSSCKKSSCKCAADSCACGKGCTGPSTCKCDSGCSCR-
			NpeMT2		(1)	MPNPKGPGCTEDCKAAQCQCGTNCRCSRDCPCNDCHKATCKCSGSCACGEGCSGPQTCKCEDDCSCH-
		Neritina pulligera	NpuMT1	MT1 [2]	MK577683	MSDPKGASCTTECKCNPCNCGTNCKCGPDCTCSSCKKSACKCSGTCACGKGCTGPDSCKCGAGCSCR
			NpuMT2	MT2 [2]	MK577684	MPDPKGKGCTKECKADSCQCGANCKCGGDCPCKDCHKPTCSCSGSCACGKGCTGPETCKCADDCSCH
		Titiscania limacina	TlimMT1	MT1 [1]	(1)	MSDTKPAGCTTECRTDPCACGTNCKCTAECPCSACHKPTCKCAGGPCACGKGCTGPASCKCADDCSCH
			TlimMT2 partial		(1)	FRSCKCAADSCACGKGCTGPSTCKCDSGCSCR
	Caenogastropoda	Anentome helena	AheMT1	MT1 [2]	MK577685	SDTEAKHGDGCTDACKETPCGCAASGGCKCTGDCNCTACCCKCDKDECKCDQGCTGPDNCKCEEGCKCKSSD
		Asolene platae	AplaMT1		(1)	MSSSEAHSHHHGECAKECKKSKESCCEGACTDECKKTPCNCGDKCKCSDGCRCQSCSAPCKCDGTCQCGKGCTGPVSCKCDRKCSCK
			AplaMT2 partial		(1)	ECQTCKRDPCTADCKKTPSNCGDKCGCAKGCKCQTCKRDACTAECKKTPCXCGDKCGCADGCECQTCKRDACTAECKKTPCNCGDKCGCADGCECQTCKRDPCTADCKKTPCNCGDKCGCAKGCKCQTCKRDACTAECKKTPCDCGDKCGCADGCNCQTCTG
		Batillaria attramentaria	BatMT1		(1)	MSSAESAKYGAGCTDQCKSDPCGCAGKCECKDDCKCQKCSDPCKCTAGSCKCGVGCQGPDSCRCDRKCACK
		Bithynia siamensis goniomphalos	BsiaMT1 partial		(1)	DCKKDPCGCEGKEKCGCGDNCSCKSCPSSKGKVAGESCTEDCKKDPCGCEGKEKCGCGADCKCKSCAPCKCEENCACGEGCTKEKDDCECNKSCACKG
		Bostrycapulus aculeatus	BacMT		(1)	MSSTGKGCVDLCTEQSCGCAKGGCDCGDLCKCQTCNPCKCGGXCKCGNGCTGPADCHCAKSCSGCK
		Californiconus californicus	CcalMT1		(1)	MSQSSKPREFTAGCTADCKKDPCGCASAGCNCKSDCTCSSCGRGPGCTDACKQSPCGCGSSCKCVGDCKCPACCCQCTKDACKCGKKCEGPGSGCTCDRSCVCFQK
		Cinguloterebra anilis	CanMT1		(1)	MSSAQTTIDNTCCTEACKQTPCGCGDECKCTGKCGCQACSSCQCTSGGCACGKGCSGPGSCKCDKSCSCKTK
			CanMT2		(1)	MSSAQAKFGEGCTDACKQTPCGCGDNCECTGKCGCLMCSSCQCTXGGCACGKGCSGPGSCKCDKSCSCKXK
		Cipangopaludina cathayensis	CcatMT1		(1)	MSSSEAHAHHHGGCAKECKKSKESCCEGACTDECKKTPCNCGDNCKCSDGCRCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
		Colubraria reticulata	CretiMT1		(1)	MSDSEKFGPGCTEECKTTCQCSVSGCKCVGECACPSCACTCTDGKCACGKDCTGPKTCKCDSGCSCKPGKSK
		Concholepas concholepas	CconcMT1		(1)	MSSPVAEMNTGCTDACKTEGGCGCGDKCKCTGKCACQNCSSCQCSDSCNCGKGCTGPASCKCDKGCSCK
		Conus consors	CconsMT1		(1)	MSDTAAYDSGCTEACKQTPCGCATSGGCKCAGPCKCPTCGCQCTDGTCKCGEESCKGPQGCKCDQTCTCKPKK
		Crepidula fornicata	CfoMT		(1)	MSATLGAGCTDACKQIPCECAQTGGCKCTGGICPCPTCNPGGNCTDACNQKPCGCIDAGDCQCGANCPCLTCNPCKCSGACVCGQGCTGRESCKCVRGCSCRK
		Crepidula plana	CplMT1 partial		(1)	GPASTKGAGCTEACKQNPCGCKEAGGCSCAGDCCCQTCSPCXCSGSCACGKGCTGRGNCKCNKGCTCKN
		Echinolittorina malaccana	EmaMT1		(1)	MSSVYGPGCTDVCKQTPCGCASSGCKCTGDCKCPSCAYGAGCTDSCKETPCGCGSGCNCXSDCXCQTCSTSCKCDGSCKCGQGCTGPDNCKCDRTCTGCK
		Euspira heros	EheMT1		(1)	MSSEGAGCTDVCKQDPCGCASTGCKCAGDCQCQSCGSSDGCTDACKQTPCGCATSGCGCKEGCRCKSCTTACKCGDQCKCGQGCTGPSNCKCGSGCSCK
			EheMT2 partial		(1)	CGPGAGCTAACKQSPCGCXGAGCKCTGDCRCKACGFGAGCTDACKQTPCGCAGAGCKCT
		Gemmula speciosa	GspMT1		(1)	MSSAEAKYGPSCTEACRETPCGCKKDGCQCKGDCKCPSCCCKCTKEGGCKCGADCKGPGDCKCGPDCKCKPT
		Hinea brasiliana	HbrMT1		(1)	MSSAEAKYGAGCTDACKNDPCGCTSKCECKEDCKCQQCSSPCKCTPEGCKCSVGCTGPDSCQCDRKCACK
		Janthina janthina	JjaMT1		(1)	AGCTDACKESPCACATKGCXCKDDCRCPSCHKAPGCTDACKENPCACASKGCECKADCPCKSCALCKCPGECKCGVGCTGPDSCNCDKACSCK
		Lanistes nyassanus	LnyMT1		(1)	MSSEGHAHTHAECTEECKKAGEGGESCDCGSACTEECKKTPCNCGDKCGCGEGCKCQDCSAPCKCAGSCQCGHGCTGPDSCKCDRKCACK
		Littorina littorea	LliMT	LliMT/ MT allelic variant 1 [3]	AST14862	MSSVFGAGCTDVCKQTPCGCATSGCNCTDDCKCQSCKYGAGCTDTCKQTPCGCGSECNCKEGCRCQSCSTACKCAAGSCKCGKGCTGPDSCRCDRSCSCK
		Littorina saxatilis	LsaMT1		(1)	MSSVFGAGCTDVCKQTPCGCATSGCKCTDDCKCQSCKYAXGCTDXCKQTPCGCGSGCNCKEDCCCQSCSTACKCAGSCNCGKGCTGPDSCKCDRSCSCK
		Marisa cornuarietis	McornMT1	MT [2]	MK577687	MSSSEAHSHHHGECAKECKKSKESCCGGACTDECKKTPCNCGDNCKCSDGCRCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
			McornMT2	8mdMT [2]	MK577688	MSSANPACTAECKKVPCNCGDTCRCGDGCTCQTCKRDACTADCKKTPCNCGDKCGCAKGCKCQTCKRDACTAECKKTPCNCGDKCKCAKGCNCQTCKRDACTPECKKTPCNCGDKCGCADGCECQTCKRDACTADCKKTPCNCGDKCGCAKGCKCQTCKRDACTAECKKTPCNCGDKCGCADGCECQTCKRDACTAECKKTPCNCGDRCRCVHGCRCQSCSAPCKCRGTCQCGVGCTGATSCKCSRQCSCK
		Marseniopsis mollis	MmoMT1		(1)	MSSAEATYGPGCTDSCKETPCGCVATGECKCTGDCKCMACSSCKCSESCKCGKGCTGSENCKCDKSCSCKK
		Neverita didyma	NdiMT1		(1)	MSSYGPGCTDACKQSPCGCATSGCKCTGDCKCSACGFGSGCTDTCKQTPCGCATAGCKCTDDCRCKSCRCKCGDKCDCGKGCTGPANCKCGSDCTCK
		Nucella lapillus	NlaMT1		(1)	MSDKPTYGAGCTDVCKETPCGCASNGGCKCTGDCNCPACSCKCTDSSCQCGQGCTGPDSCSCGADCSXKSK

	Oncomelania hupensis	OhuMT1		(1)	MSDKAAKDRCGDSCTETCKTDPCACEGGDKCACGDDCRCKSCAGPCKCDANCDCGEGCTKEKDDCDCNKSCDCKQ
	Pila ampullacea	PamMT1		(1)	MSSEKGHAHAHSECTEACKTGDEAGDTCGCKDACTDECKKSPCNCGDQCGCGEDCKCQACSAPCKCSGTCQCGQGCSGPDSCKCDRKCSCK
	Pomacea bridgesi	PbrMT1	MT 1 [2]	AST14869	MSSSEAHAHHHGECAKECKKKSKASCCESACTEECKKTPCNCGDKCKCSDGCKCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
		PbrMT2	MT 2 [2]	AST14870	MSSANPICTAECEKVPCNCGDKCGCADGCECQTCKRDACTAECRKTPCNCGDKCGCGDGCKCQTCKRDACTAECKKTPCNCGDSC
	Pomacea caniculata	PcanMT1	MT20-IIIlike variantX-1 like [1]	XP_025111372	MSSANPACTAECKKTPCNCGDKCGCADGCECQTCKRDACTAECKKTPCNCGDTCGCGDGCKCQTCKRDACTAECKKTPCNCGDTCGCADGCECQTCKRDPCTAECKKTPCSCGDRCQCADGCRCETCSAPCKCRDTCQCGEGCTGTTSCKCPLKCSCK
		PcanMT2	MT 20-III-like isoform X2 [1]	XP_025111374.1	MSSANPACTAECKKTPCNCGDTCGCGDGCKCQTCKRDACTAECKKTPCNCGDTCGCADGCECQTCKRDPCTAECKKTPCSCGDRCQCADGCRCETCSAPCKCRDTCQCGEGCTGTTSCKCPLKCSCK
		PcanMT3		(1)	MSSSEAHAHHHGGCAKECKKSKESCCEGACTDECKKTPCNCGDNCKCSDGCRCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
	Pomacea diffusa	PdiMT1		(1)	MSSSEAHAHHHGECAKECKKSKASCCESACTEECKKTPCNCGDKCKCSDGCKCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
		PdiMT2		(1)	MSSTQASANPACTAECKKTPCNCGEKCRCGDGCTCQSCKSCTTECKRTPCSCGDTCGCSDGCRCQACSAPCKCRGTCQCGVGCTGAASCKCPRKCSCK
	Pomacea maculata	PmaMT1		(1)	MSSSEAHXHHHGGCAKECKKSKESCCEGACTEECKKTPCNCGDNCKCSDGCRCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
		PmaMT2		(1)	MSSANPACTAECKKTPCNCGDKCGCADGCKCQTCKXDACTAECKKTPCNCGDKCGCADGCKCQTCSAPCKCRDTCQCGAGCTGTTSCKCPLKCSCK
	Pomacea scalaris	PscMT1		(1)	MSSSEAHAHHHGECAKECKKSKASCCEGACTEECKKTPCNCGDKCKCSDGCRCQSCSAPCKCDGTCQCGKGCTGADSCKCDRKCSCK
		PscMT2		(1)	MSSTQASANPACTAECKKTPCNCGEKCRCGDGCTCQSCKSCTTECKRTPCSCGDTCGCSDGCRCQACSAPCKCRGTCQCGVGCTGAASCKCPRKCSCK
	Pomatia elegans	PelMT1	MT 1 [2]	ARA71541	MSTSGANVIYGAGCTGTCKQSPCGCKNSAAGCRCKDDCQCPACAKYGAGCTGTCKQSPCGCKNSAAGCGCKDDCRCPACAKSCKCGTCNCGKGCTGPSNCKCDDGCSCK
		PelMT2	MT 2 [2]	ARA71542	MSSSGANATGAGCTETCKESPCGCKNSAAGCKCKDDCQCTTCAKSCKCAGTCNCGKGCTGPNSCKCDGGCPCK
	Potamopyrgus antipodarum	PantMT1		(1)	MSDTAAKDRCGAGCSEECKQDPCGC2ASCKCEAGCACKSCGPACKCGDNCECGKGCTKEKEDCACDKSCDCKS
	Pugilina cochlidium	PcocMT1		(1)	MSSTAAEMNTGSGCTDACKKGDGCGCGKECKCTGNCQCKSCSTCKCSGSCNCGAGCQGPDTCKCDKSCKCK
	Reishia clavigera	RclMT1		(1)	MSSTAGEVNTGCTDACKNAGGCECGDKCKCTGDCQCKSCSTCKCSGSCKCGQGCTGPETCKCDKSCGCK
	Rubyspira osteovora	RosMT1		(1)	MSSTESGATEFGAGCTDTCKKTPCGCASGGCECKEDCKCQSCSSPCKCAGTCKCGVGCTGPDSCKCDRKCACK
	Semisulcospira coreana	ScoMT1		(1)	MSSTDTSKHGAGCTDECKTDPCACAEGCECKDDCKCTTCSAPCKCTEGNCACGEGCEGPDSCGCDRKCDCK
	Terebra subulata	TsuMT1		(1)	MSSTQPTYGPNCTEACKQNPCACGDQCKCGAECPCPVCSNCKCTSGACLCGEGCTGPGTCTCEKSCVCKK
		TsuMT2		(1)	MSSAQPTYEGCTDACKQTPCGCGDKCACAGSCSCQTCSPCQC2GCRCQKDCKGPDSCKCNKECACKKK
	Tritia obsoleta	TobMT		FK716482	MEGTGCAGPCKQTPCGCESSGGCKCGGDCSCTECGRGAGCTDACKETPCGCASSGGCKCTGNCSCPSCSCKCGKGACKCDKGCEGPGSCKCGPGCTCKKS
	Turridrupa cerithina	TceMT1 partial		(1)	ETYGPRCSAACKETPCGCATSPKGCECKEPCKCTSCGCQCTPGSCK
	Tylomelania sarasinorum	TsaMT1		(1)	MSSTEESAKHGESCTESCKKDPCACSDNCECKDDCKCQSCSQACKCEEGNCKCGVGCEGPDSCQCARKCACK
	Urosalpinx cinerea	UciMT1		(1)	MSSVEAPSAPLNPGCTEACKAKDGCQCGDHCKCTGDCQCNSCSKCKCSGTCQCSGGCTGPDNCKCGKSCSCR
Heterobranchia	Alinda biplicata	AbiMT1	CuMT1a [4]	MK648138	SGKGANCTGACNSNPCQCGDDCKCGVGCSCAECNTCKCTNDGCKCGHGCTGAGSCKCGNSCGCK
		AbiMT2	CuMT2 [4]	MK639793	MSGKGANCTGACNSNPCQSGDDCKCGVGCSCAECNTCKCTNDGCKCGHGCTGAGSCKCGNSCGCK
		AbiMT3	9md-MT [4]	MT084760.1	MSGKACTGDCKSDPCKCGANCQCGEDCTCTSCKSTKACTGDCKSDPCKCGDNCQCGVGCTCASCKSTKACTGDCKSDPCKCGDNCQCGVGCTCASCKSTKACTGDCKSDPCKCGDNCQCGEGCTCASCKSTKACTGDCKSDPCKCGANCQCGEGCTCASCKSTKACTGDCKSDPCKCGANCQCGEGCTCASCKSTKACTGDCKSDPCKCGDNCQCGEGCTCASCKNTKACTGDCKSDPCKCGDNCQCGDGCTCASCKTCKCTNEACKCGQECTGPATCKCASGCSCK
		AbiMT4	10md-MT [4]	MT084761.1	MSGKACTGDCKSDPCKCGANCQCGEGCTCASCKSTKACTGDCKSDPCKCGDNCQCGEGCTCASCKSTKACTGDCKSDPCKCGDNCQCGVGCTCASCKSTKACTGDCKSDPCKCGDNCQCGEGCTCASCKNTKACTGDCKSDPCKCGDNCQCGEGCTCASCKSTKACTGDCKSDPCKCGDNCQCGVGCTCASCKNTKACTGDCKSDPCKCGANCQCGEGCTCASCKSTKACTGDCKSDPCKCGDNCQCGEGCTCASCKNTKACTGDCKSDPCKCGDNCQCGDGCTCASCKTCKCTNEGCKCGQECTGPATCKCASGCSCK-
	Arianta arbustorum	AarMT1	MT_ARIAR [5]	P55946	SGKGKGDLCTAACKNEPCQCGSKCQCGEGCACASCKTCNCTSDGCKCGKECTGAASCKCNSSCSCK
	Ampullaceana balthica	AbaMT1		(1)	MSGKGPNCTEACTGEQCTCGDSCKCGEGCNCPSCKTCKCSAEDCKCDAGSKGTGNCQC
	Aplysia californica	AcaMT1		(1)	MSGKGPNCTEACGGDPCNCADSCQCGEGCSCSACKKCLCTAESCKCGVGCQGPASCKCGSSCGCK
	Arion vulgaris	AvuMT1	AvMT1 [6]	AWD77146	MSGKACTGACKSEPCQCGNNCQCGGDCDCSQCKTCKCTNEGCKCGQNCTGQATCSCEKSCSCK
		AvuMT2	AvMT2 [6]	AWD77147	MSGRGCNGTCNSNPCQCEDGCQCGDACSCAQCNTCKCTNDGCKCGNECTATGSCKCGTSCGCN
	Biomphalaria glabrata	BglMT1	MT allelic variant 1 [7]	AST14858	MSGKGPNCTEACTGEQCNCGDSCKCGEGCNCPSCKTTKGPNCTEACTGKQCSCGDSCQCGEGCTCSCCKNACTKECTDTECSCGDSCKCGEGCKCSSCKAGKCTKSDEGCKTEGHCAKGKCCKS
	Biomphalaria pfeifferi	BpfMT1		(1)	MSGKGPNCTEACTGEQCNCGDSCKCGEGCNCPSCKTTKGPNCTEACTGKQCNCGDSCQCGEGCTCSYCKKACTKECTNTECSCGDSCKCGEGCKCSSCKAGKCSKSDEGCKTEGHCAKGKCCKS
	Bradybaena similaris	BsimMT1		(1)	MSGKGTACAGSCNTNPCSCGDDCRCGAACSCSQCQAHHCTNDSCKCGSQCSSSGSCKCGSCGCK
		BsimMT2		(1)	MSGKGKEEACTAACKSDPCQCGADCTCSACKSCQCTKACTAACKSDPCQCGDQCQCGAGCTCSACKSCQCTNDGCKCGKECTGPASCKCDSSCSCK
	Cepaea hortensis	ChoMT1_1	Cd MT variant 1 [6]	AYL40762	MSGKGKGEKCTAACRNEPCQCGSKCQCGEGCACAACKTCNCTSDGCKCGKECTGPDSCKCGSLCSCK
		ChoMT1_2	Cd MT variant 2 [6]	AYL40763	MSGKGKGEKCTAACRNEPCQCGSKCQCGEGCACAACKTCNCTSDGCKCGKECTGPDSCKCGSSCSCK
		ChoMT2	Cd/Cu MT [6]	AYL40764	MSGKASACAGSCNSNPCSCGDDCQCGAGCSCAQCHSCQCNNDTCKCGNQCSASGSCKCGSCGCK
	Cepaea nemoralis	CneMT1		(1)	MSGRGNCGGACNSNPCSCGDDCKCGAACNCXRCSSCHCSNDDCKCGSQCTGSGSCKCGSACGCK
		CneMT2		(1)	MSGKGKGEKCTAACRNEPCQCGSKCQCGEGCACAACKTCNCTSDGCKCGKECTGPDSCKCGSSCSCK
		CneMT3		(1)	MSGKGSACAGSCNSNPCSCGDDCQCGAGCSCAQCQSCQCNNDTCKCGNQCSASGSCKCGTCGCK
	Clione limacina anatarctica	ClaMT1		(1)	MSGKGPNCTESCGVPCNCGDSCQCGEGCSCKACKNCKCTAEGCKCGVGCKGAETCKCDSSCSCK
	Clione limacina	CliMT1		(1)	MSGKGPNCTEACSTPCNCGDSCQCGEGCSCKACKKCKCTAEGCKCGVGCTGAGSCACGDSCSCK
		CliMT2		GESV01103158 (2)	MSGKGTTMTEACDTPCECGDNCQCGEECSCTSCRKCKCTAEAAEGCKCGDDCTCAESCNCSCRGK
	Cochlicella acuta	CacMT1		Unigene65576_Cochlicella_acuta_CNS (3)	MSGKGSACAGSCGNNPCSCGDDCRCGAGCSCAQCNSCQCNNDTCKCGNQCSTSGSCKCGSCGCK
		CacMT2		Unigene52262_Cochlicella_acuta_CNS (3)	MSGKGKAESCTAQCQSNPCQCGDKCQCGEGCACTSCKTCKCTSDGCKCGKECTGPASCKCGSSCSCK

		CacMT3		Unigene18701_Cochlicella_acuta_CNS (3)	MSGRGNCGGACKSNPCSCGQVCKCGGACTCAQCNACRCSGDSCKCGDQCTASGSCQCGSGCGCK
	Deroceras invadens	DinMT1		Unigene4785_Derocerus_invadens_CNS (3)	MSGKGSKCTTDCKSTPCKCGDNCQCGNDCSCSECKTCKCTNEGCKCGVGCTGAATCKCADSCSCK
	Deroceras reticulatum	DreMT1		Unigene30723_Derocerus_reticulata_CNS (3)	MSGKGEKCTGDCKSEPCKCGQNCQCGNDCTCSQCKTCKCSSSGCQCGHGCTGVESCKCGSSCTCK
	Elysia cornigera	EcoMT1		GBRW01068275 (2)	MSGKGASCPPTCRNDPCGCGQDCQCGQDCTCSSCSCLCGNECKC2SCSQGVRCRCDTLCDCRPK
	Elysia crispata	EcrMT1	MT [2]	AST14868	MSGKGLICTAPCTNDPCGCGENCQCGHACSCHSCSCSCGADCKCNQATCHEKSNCKCEASCSCRAK
	Euhadra quaesita	EquMT1		(1)	MSGKGTACAGSCNTNPCNCGDDCRCGAGCSCSQCHAHHCTNDSCKCGSQCSSSGSCKCGSCGCK
	Fiona pinnata	FpiMT1		(1)	MSGRGTCTEDCKSSPCKCGTNCKCGXQCPCXXCXXCXCTADSGCQCGVGCTGSDTCKCNASCSCK
	Galba cubensis	GcuMT1		(1)	MSGRGPNCTEACRGEQCNCGDSCKCGEGCNCPSCKTCKCSAEDCKCDAGSKGTGNCKC
	Galba schirazensis	GscMT1		(1)	MSGKGPNCTEACRGEQCNCGDSCKCGEGCNCPSCKTCKCSAEDCKCDAGSKGTGNCKC
	Galba truncatula	GtrMT1		(1)	MSGKGPNCTEACRGEQCNCGDSCKCGEGCNCPSCKTCKCAAEDCKCDAGSKGTGNCKC
	Haminoea antillarum	HanMT1		(1)	MSGKGPNCTEACKTTPNCQCGDKCACGTECGCNSCSCTDCKCTAASGCGCGKXCTGKSTCACDGPCLCK
	Helix aspersa	HasMT1	copper-MT [8]	ABM55268	MSGRGQNCGGACNSNPCNCGNDCNCGTGCNCDQCSARHCSNDDCKCGSQCTRSGSCKCGNACGCK
		HasMT2	cadmium-MT [8]	ABL73910	MSGKGKGEKCTAACRNEPCQCGSKCQCGEGCTCAACKTCNCTSDGCKCGKACTGPDSCTCGSSCGCK
		HasMT3	Cd,Cu-MT [8]	ABM92276	MSGKGSACAGSCNSNPCSCGDDCKCGAGCSCAQCYSCQCNNDTCKCGSQCSTSGSCKCGSCGCK
	Hermissenda crassicornis	HcrMT1		(1)	MSGRGNCTDACKESPCKCGADCKCGDQCPCNSCKPCQCSASGCKCGVGCTGADSCHCLSGCSCK
	Helix lucorum	HluMT1		(1)	MSGRGKNCGGACNSNPCSCGNDCKCGAGCNCDRCSSCHCSNDDCKCGSQCTGSGSCKCGSACGCK
		HluMT2		(1)	MSGKGSNCAGSCNSNPCSCGDDCKCGAGCSCAQCHSCQCNNDTCKCGNQCSTSGSCKCGSCGCK
		HluMT3		(1)	MSGKGPNCTEACGGDPCNCADSCQCGEGCSCSACKKCLCTAESCKCGVGCQGPASCKCGSSCGCK
	Hydatina physis	HphMT1		(1)	MSGRGPNCTEACKNDPCNCGSDCKCGEGCACNSCTTCECTSGGCKCGVGCTGPESCRCDATCYCKKK
	Helix pomatia	HpoMT1	Cu-MT isoform [9]	AAK84864	MSGRGKNCGGACNSNPCSCGNDCKCGAGCNCDRCSSCHCSNDDCKCGSQCTGSGSCKCGSACGCK
		HpoMT2	Cd-specific MT [9]	ACN66299	MSGKGKGEKCTSACRSEPCQCGSKCQCGEGCTCAACKTCNCTSDGCKCGKECTGPDSCKCGSSCSCK
		HpoMT3	Cd/Cu MT isoform 1 [9]	ACY71053	MSGKGSNCAGSCNSNPCSCGDDCKCGAGCSCVQCHSCQCNNDTCKCGNQCSASGSCKCGSCGCK
	Lehmannia nyctelia	LnyMT1	1	CL1241.Contig1_Lehmannia_nyctelia_CNS (3)	SGKGAKCTGACKSEPCQCGQNCQCGDDCSCSQCKTCKCSAGSTCQCGHGCTGVESCKCGNSCSCK
	Limacina helicina	LheMT1		(1)	MPGTGPNCTENCKSADCKCGADCKCGEGCDCKACKCKCTADGCKCGVGCEGPASCKCANDCSCK
		LheMT2		(1)	MPGTGPNCSETCKAAGDCKCGADCKCGEGCDCKSCKCKCSADACKCGVGCEGPASCKCDASCSCK
	Limacina inflata	LinMT1		(1)	MSGKGPNCTEVCGTEDRCCGENCQCGEGCTCINSKCNCTDDECKCDIGCKGPSSCKCDESCRCK
		LinMT2		(1)	MSGKGANCSEACSCVPCHCKDACKCGTGCECKTCLCNCSSDGCKCGEGCKGADSCKCEAACTGCK
	Limax maximus	LmaMT1	Cd MT [6]	AYL40761	MSGKGAKCTGACKSEPCQCGQNCQCGDDCSCSQCKTCKCSAGSTCQCGHGCTGVESCKCGSSCSCK
	Limacina retroversa	LreMT1		GBXC01058817.1 (2)	MPGTGPNCSETCKAAGDCNCGADCKCGKSCDCTSCNCKCSVDTCKCGVGCEGPESCGCGGSCSCK
	Lymnaea stagnalis	LstMT1	MT [10]	ALA15336	MSGKGPNCTEACTGEQCTCGDSCKCGEGCNCPSCKTCKCSAEDCKCDAGSKGTGNCKC
	Milax gagates	MgagMT1		Unigene22845_Milax_gagates_CNS (3)	MSGKGSKCTTDCKSTPCKCGDNCQCGNDCSCSECKTCKCTNEGCKCGVGCTGAATCKCADSCSCK
	Microhedyle glandulifera	MglMT1		(1)	MSGKGPNCTEACNGNPCQCGSDCKCGDGCPCNSCKTCKCSAGSCQCGVGCSGPGSCKCDSSCSCK
		MglMT2		(1)	MSGKGPNCTQACTGNPCQCGDKCECGDGCPCNTCKICKCSSGCKCGVGCTGPDTCKCDSSCSCK
		MglMT3		(1)	MSGKGPNCTEACKGSGGTCQCGDKCECGDGCPCNTCKTCKCNNGCQCGVGCTGPDTCKCDSSCSCK
	Melibe leonina	MleMT1		(1)	MSGKGNCSENCKQSPCQCGSKCQCGDQCPCNTCKPCQCLANECKCGVGCTGGECHCLASCKCK
		MleMT2		(1)	MSGRGTCSDACKETPCKCGTDCQCGAQCPCNTCKPCHCTAEGCKCGVGCTGGDschclagcsck
	Nesiohelix samarangae	NsaMT1	MT	ACC17831	MSGKGELCTSACKSNPCQCGDKCQCGEGCTCSACKSCHCTNDGCNCGKECTGPTSCKCDTSCSCK
	Onchidella floridana	OflMT1		(1)	MSGKGPNCTQACKGEPCQCGDKCECGDGCSCMACKVCKCSADGCKCGKGCTGKESCACDTSCSCK
		OflMT2		(1)	MSGRGPNCTQTCTPCQCGDKCGCGTGCPCMSCRCTSDGCKCGKGCTGKVTCNCDKSCTCK
	Onchidium reevesii	OreMT1	MT	QCG81602	MSGRGPNCTEACKGSPCQCGDKCECGSGCDCKECKVCKCSAEGCKCGAGCKGKESCGCDSSCSCK
	Ophicardelus sulcatus	OsuMT1		(1)	MSGKGPNCTEACKVDGTCQCGDNCECGDGCPCGSCKSCKCTDEGCKCDAGCKGPGTCKCDSSCKCK
		OsuMT2 partial		(1)	MSGKGPNCTEACNGXXXXXXXXXXXXXXXXXTCTSCKHCTCSSEGCKCGKGCEGPDMCTCDSSCSCK
	Oxynoe viridis	OviMT1		(1)	GRGVPGSYCDEACRTVPCACANDCRCGTGCGCNSCKLCECTDGGCKCGVGCKGIDSCSCDATCTCK
		OviMT2 partial		(1)	MSGKGPNCTEAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTCSSEGCKCGKGCEGPDMCTCDSSCSCK
	Physella acuta	PacMT1		(1)	MSGKGPNCTEACTGEQCTCGDSCKCGEGCNCPSCKTCKCEDNACKCGEGCTGPSTCKCESSDCACK
		PacMT2	MT	ADB29127	MSGKGPNCTEACTGEQCTCGRQLQVCCKTCKCEDNACKCGEGCTGPSTCKCESSDCACK
	Physella carolinae	PcarMT1	MT[1]	(1)	SGKGPNCTEACTGEQCTCGDSCKCGEGCNCPSCKTCKCEDNACKCGEGCTGPSTCKCESSDCACK
	Physella gyrina	PgyMT1		(1)	MSGKGPNCTEACTGEQCNCGDSCKCGEGCNCPSCKTCKCEDNACKCGEGCTGPSTCKCESSDCACK
	Physella hendersoni	PheMT1	MT[1]	(1)	SGKGPNCTEACTGEQCTCGDSCKCGEGCNCPSCKTCKCEDNACKCGEGCTGPSTCKCESSDCACK
	Philine angasi	PangMT1		(1)	MSGKGANCSEACKNDPCNCGTSCQCGDGCDCKDCKTCKCTADXCNCGEGCTGKDTCSCGSGCSCK
		PangMT2		(1)	MSGKGPNCTAACGNPCQCGDNCQCGDGCSCKSCKKCKCTTDGCKCNAGCQGPSSCKCDSS
	Phallomedusa solida	PsoMT1		(1)	MSGKGANCTEACKGQPCKCGENCQCGDGCDCGTCSCKCTSEACKCGVGCKGPGTCKCDSSCSCK
	Rissoella caribaea	RcaMT1		(1)	MSGKGQNCTTSCKNSPCRCGDNCTCGDNCNCKSCVVCKCTESNGCACSIGCTGPDSCSCDSSCTCK
	Siphonaria pectinata	SpeMT1		(1)	MSGRGQNCTDACKGDPCQCGDKCQCGTGCNCNSCKACQCTNESCKCGAGCKGPGSCTCGSSCGCK
	Strubellia wawrai	SwaMT1		(1)	MSGKGPNCTAACGNPCQCGDNCQCGDGCSCKSCKKCKCTTDGCKCNAGCQGPSSCKCDSSCSCK
	Tritonia diomedea	TdiMT1		(1)	MSGKGCSEACQQKPCNCGNDCKCGAECPCNTCKPCTCTDSGCKCGVGCTGGDSCHCLGGCSCK
		TdiMT2		(1)	MSGKGSACAGSCNSNPCSCGDDCQCGAGCSCAQCQSCQCNNDTCKCGSQCSASGSCKCGSCGCN
	Tylodina fungina	TfuMT1		(1)	MSGKGPNCTEACNGDPCACGDKCGCGSGCTCTSCKHCTCSSEGCKCGKGCEGPDMCTCDSSCSCK
		TfuMT2 partial		(1)	GGCKCGVGCTGPESCRCDATCYCKKK
	Theba pisana	TpiMT1		CL8321.Contig1_Theba_pisana_CNS (3)	MSGRGKNCGGACNSNPCNCANNCRCGAGCNCDSCSSCHCSNDDCKCGNQCTTSGSCKCGSACGCK
	Turbonilla sp.	TspMT1		(1)	MSGKGPNCTEECKNDPCNCEDKCQCGDDCPCDSCNKQCKCTGNACGCGVGCSGSERCGCPSSCSCK
		TspMT2		(1)	MSGKGPNCTEACNGNPCQCGSDCKCGDGCPCNSCKTCKCSAGSCQCGVGCSGPDSCKCDSSCSCK
Vetigastropoda	Amphiplica gordensis	AgoMT1		(1)	MSCAKGEACTTECRTDPCACGESCKCGPGCDCNTCKDVKRA
	Chlorostoma funebralis	CfuMT1		(1)	MSSTGEKCTTECKTTPCACGTDCKCGPGCACGSCKDVKRACKCSDSCKCGEGCTGTDTCKCDNSCSCK
	Clanculus margaritarius	CmaMT1		(1)	MSSTGEPGVTCTPECKTTPCACGDDCKCQAGCKCEGCKKECKCSGGCDCGKGCTGGDTCKCDNSCSCK

		Clanculus pharaonius	CphMT1		(1)	MSSTGETCTTECKTTPCSCGDDCKCQAGCKCXSCKKECKCSDACACGKGCTGGDTCKCDNSCSCK
			CphMT2 partial		(1)	PCACNDNCRCMIGCTCVACSDKCKCQGTCECGKGCTGNTTCTCDVSCSCKSGRS
		Granata imbricata	GimMT1		(1)	MSAPNPSCTDACRADPCACGDNCGCKAGCTCTNCRGVVRPCKCSGTCQCGAGCTGSATCTCDNACRCKE
		Haliotis discus discus	HddMT1	MT [11]	APT40583	MSSPQGAGCTGECKTDPCACGTDCKCNPDDCACDTCKVKKTCKCPGSCECGKGCTSGETCKCDDSCTCK
		Haliotis discus hannai	HdhMT1	MT	ASA48039	MSSPQGAGCTGECKTDPCACGTDCKCNPDDCACDTCKVKKTCKCPGSCECGKGCTSGETCKCDDSCTCK
		Haliotis diversicolor supertexta	HdsMT1	MT	ABW04628	MSSPQGPGCTASCKSEPCACGTDCKCNPSDCPCTTCKDKTVCKCSDGCQCGKGCTTGDICKCDDSCSCK
		Haliotis laevigata	HlaMT1		GFTT01185788 (2)	MSSPQGAGCTPECRSNPCACGENCRCNPSDCVCTTCKVKKVCTCSGVCQCGNGCTGGDTCTCDDSCRCK
		Haliotis tuberculata	HtuMT1		GEAU01019532 (2)	MSSSGAGCTAECRSEPCACGDDCRCDPKTCRCTECRKTCTCTEAGCRCGRGCTGPENCRCANACTCKKPATKTYTRTASCHS
		Lepetodrilus fucensis	LfuMT1		(1)	MSSEQGTACTAECKTDPCKCGDACKCGAGCSCSSCKDVKKTCKCSDSCQCGKGCKTGEDCKCDNSCKCK
		Margarella antarctica	ManMT1		(1)	MSSTGETCTTECKTDPCACGADCKCKAGCACVGCKDTTKACKCKDXCGCGKGCTGGESCKCDNSCSCK
		Megathura crenulata	McrMT1	MT [12]	AAM51554	MSGKGENCTAECKSDPCACGDSCKCGEGCACTTCVKTEAKTTCKCGESCKCEGCKEGEACKCESGCASCK
		Monodonta labio	MlaMT1		(1)	MSSTGETCTTECKTDPCGCGDDCKCKPGCACGSCKDTKKACKCSDSCKCGEGCTGSDSCKCDNSCSCK
		Perotrochus lucaya	PluMT1		(1)	MSDVKDKTDCTECTDACKTTPNACGTECKCEAGCTCEKCKPCKCSGKCTCGTGCTDSHSCKDDHSCKDDHSCSSK
			PluMT2		(1)	MSDXKTAGVACTDACKTDPCACGTECKCGAGCACEKCKHVVRPCKCSGKCDCGAGCTGVDSCKCDNSCSCK
		Phasianella ventricosa	PveMT1		(1)	MSSTAATCNVDCQTTPCACGDSCKCKEGCACSTCNDINIDCQCKDPCGCGKGCTGADTCKCGNSCSCK
			PveMT2		(1)	MSSPGSCPPECRSDPCRCGNNCECPSSCQCSRCNGGRRECRCRGRCDCGXGCTGPATCNCDNSCHCKGPITG
		Prothalotia lehmanni	PleMT1		(1)	MSGKGESCKTPCKNSPCSCGDSCKCGTDCACGSCKTVEKDCRCSESCSCGKGCTGNGSCKCGNGCNCK
			PleMT2		(1)	MSSTGETCTTECKTDPCGCGDDCKCKPGCACGSCKDTKKACKCSDSCKCGEGCTGSDSCKCDNSCSCK
		Tegula atra	TatMT1		GFBI01039752 (2)	MSSTGEKCTTECKTTPCACGTDCKCGPGCACDSCKDVKKACKCSDSCKCGIGCTGDDTCKCDNSCSCK
Cephalopoda		Nautilus pompilius	NpoMT1		(1)	MSDRCDCIRTGNCICSDACKGDNCRCGDACKCQKNCSCPSCRVTCKCSGSCGCGAGCTGADSCHCANTCSCK
		Octopus bimaculoides	ObiMT1		KOF91040	MPDSCCCTLEGGKCGCSSSCKTGKCDNTGCKCGDDCKCKDCKGRKCQCSEGNCKCDSGCKHGSEKSCCK
		Octopus vulgaris	OvuMT1		(1)	MPDSCCCTLEGGKCGCSSSCKTGKCDNTACKCGDECKCKDCKGPKCQCSEGNCKCDSGCKHGSEKSCCK
		Sepia esculenta	SesMT1		(1)	MPDSCCCTEGGKCVCSDSCKGGKCESSGCKCSADCTCPKCKGTACKCSNGGCKCGAGCKGNSDCKCGAGCSCK
Scaphopoda		Antalis entalis	AenMT1		(1)	MSDPCNCATGGNCTCGESCDCSKSNCNCDPSKCNSKCSSKGGVCNCPADGCKCGGECTGPSSCSCGPSCSCK
		Graptacme eborea	GebMT1		(1)	MSDPCNCATTGDCKCDESCACTDCKCDASKCKSKSAAHACKCTDSSCQCGKGCTGAASCSCAPDCSCK
Monoplacophora		Levipilina hyalina	LhyMT1		(1)	MPDPCTCSETGESCKCDSPCTCTDCKCACAKCKTVCKCPGTCKCGNDCTGPENCKCKPDCSCK
Bivalvia		Alectryonella plicatula	ApliMT1	MT	ALA14799	MSDPCNCSETGTCVCSDSCPATGCKCGPGCKCGDGCKCSGCKVKCYCSGSCGCGKGCTGPENCKCANDSGCVCKVKCNCSGSCGCGKGCTGPENCKCANDSGCVCKK
		Argopecten irradians	AirMT1	MT [13]	ABO16370	MSSEAVFCAKCPGGKCGCAGPDCC2CRCSGCKVNCACFSAKSSTCPDGVCQMGVACKCGGSCKCSDSCPCPNCKAGCKCAGANFVCSDDCQSGKCCGTSDCKCDGQCACPKCKCNCASGTCTCGVGCTGPNSCKCGTDCKCKQ
			AirMT2	MT [13]	ACI01808	MSEVAALCKLEECLIGTCCSGNDACKCDRACQCQKGAAKCQCSVAAICKLEECQAGNCCGGSIACKCEGACQCSRCGVKCQCSDSCGCGKGCTSLLCNCPVGCTCRPAKS
		Bathymodiolus azoricus	BazMT10	MT, isoform MT-10a [14]	CAF34421	MPAPCNCVETNVCICDTGCSGEGCRCGDACKCSGADCKCSGCKVVCKCSGSCACEGGCTGPSTCKCAPGCSCK
			BazMT20	MT, partial [14]	CAF34424	GFCNCIETNVCICGTGCSGKCCRCGDACKCASGCGCSGCKVVCKCSGTCKCGCDCTGPTNCKCESGCSC
		Bathymodiolus thermophilus	BthMT10	MT, partial [15]	CAE11860	PCNCVETNVCICDTGCSGDGCRCGDACKCSGADCKCSGCKVVCKCSAGSCECGKGCTGPSTCRCAPGCSC
			BthMT20	MT, partial [15]	CAF34425	GFCNCIETNVCICGTGCSGKCCRCGDACKCASGCGCSGCKVVCKCSGTCKCGCDCTGPTNCKCESGCSC
		Cerastoderma edule	CedMT1	MT 1	ACT66292	MGDPCNCAQTGGDCKCAAGNCCSSDTPCRCGSGCGCGSECTCHVKCTCSGSCACGNNCTGPANCTCGAGCSCK
			CedMT2 partial	MT, partial	ABN68955	MSDPCNCTETGTCVCSDSCPATGCKCGPGCKCGDGCKCSGCKVV
		Cerastoderma glaucum	CglMT1	MT, partial [16]	ACB05816	IPCNCVETGSCQCSAQCTSEGGCKCGPNCKCGDGCACPGCKVVCNCNGSCACGQDCTGPSTCKCGGDCSCKK
			CglMT2	MT	CCW28381	MGDPCNCAQTGGDCKCAAGNCCSGETPCRCGSGCGCGSDCTCHVKCSCSGSCACGNNCTGPANCTCGAGCSCK
		Corbicula fluminea	CflMT1	MT [17]	ABM55725	MSDPCDCATTGTCKCSEDCTAESGCRCGTGCNCGDDCKCPGCKVVCKCSGSCDCGQGCTGPSTCKCESDCSCK
		Crassostrea angulata	CanMT1	MT	AAK15581	MSDPCNCIETGTCVCSESCPATGCKCGLGCKCGDGCKCSGCKVKCNCSGSCGCGKGCTGPENCKCANDSGCVCKVKCNCS
		Crassostrea ariakensis	CarMTI	MT	ABC69708	MSDPCNCIEAGTCVCSDSCPPTGCKCGPGCKCGDGCNCSGCKVKCNCSGSCGCGKGCTGPENCKCSNDSGCGCKK
			CarMTIV	MT IV [18]	AEF33387	MSDHCTCAETGKCVCCDTCPPDGNCSCGDKCKCAKASCNCKGCKVKCCCTKEKCCCGKGCHGPETCKCPADCCCKKEHDACSKAGH
		Crassostrea gigas	CgiMTI	MT [19]	CAB64869	MSDPCNCTESGTCVCSDSCPATGCKCGPGCKCGDGCKCSGCKVKCNCSGTCGCGKGCTGPENCKCANDSGCGCKK
			CgiMTII	MT [19]	CAC48045	MSDPCNCTETGTCVCSDSCPATGCKCGPGCKCGDGCKCSGCKVKCNCSGSCGCGKGCTGPENCKCANDSGCGCKVKCNCSGSCGCGKGCTGPENCKCANDSGCGCKK
			CgiMTIIIA	MT	AEG42189	MPIETNCTCANGACNCGETCQCKTTDCACAICNNPCGCTESECNCGAECQCPETCSCKTCKA
			CgiMTIV	MT IV	CAK22381	MSDHCTCAQTGKCVCCDTCPPDGNCSCGDKCKCAKASCNCTCKGKAKCCCTKEKCCCGKGCHGPETCKCPADCCCKKEHDACSKAGH
		Crassostrea virginica	CviMTIA	MT IA [20]	AAQ23904	MSDPCNCIETGTCACSDSCPATGCKCGPGCKCGDDCKCAGCKVKCSCTSEGGCKCGEKCTGPATCKCGSGCSCKK
			CviMTIIIA	MT IIIA [21]	AAQ23917	MPFETSCTCANGACECGENCQCKTTDCACTTCNVTCSCTESECKCGADCNCSAECKCQTCKS
			CviMTIVA	MT IVA [21]	AAZ94897	MSDTCACATTGKCVCCDTCGPDGACSCGEACSCAKKTCNCKGCKVKCCCTKDKCCCGKGCHGPGTCKCDENCCCKKQDKKQAT
		Cyclina sinensis	CsiMT1	MT	AEG19519	MSETCGCTTEVKPDAEAKKCSAAQSKGSCKCGPDCKCGDACKCSGCATACSCTPGPCECAKGGKDCGCSGKPSA
		Dreissena polymorpha	DpoMT1	MT [22]	AAB07548	MSDPCNCVETGDCRCADGSCSDCSNCKCGDSCKCSKPNCCGKNVTCKCGENCQCGVGCTGPDSCTCDSGCSCK
		Hyriopsis cumingii	HcuMT1	MT	ACS44750	MSDACNCLETGECKCCGESTGDCRCGKDCKCGDACKCPGCKVVCKCSSSCDCGKGCTGPSTCKCDSGCSCK
			HcuMT2	MT	ACZ06027	MSDPCNCIETGTCVCSDSCPATGCKCGPGCKCGDGCKCSGCKVKCNCSGSCGCGKGCTGPENCKCANDSGCGCKK
		Hyriopsis schlegelii	HscMT1	MT 1	AHH31398	MSDACNCLETGECKCCGESTGDCRCGKDCKCGDACKCPGCKVVCKCSSSCDCGKGCTGPSTCKCDSGCSCK
			HscMT2	MT 2	AHH31399	MSESCQCLEVADRECTDVCKTQCNCSDTCNCSEDCECTGRCKIQCNCSEEECICGHGCKGPRTCKCDPSCECH
		Laternula elliptica	LelMT10a	MT 10a [23]	ABH03633	MPSPCNCRETGKCTCDGKCSGDACCCGANCNCGEGCKCPGCKTVVCKCSGDCACGKGCTGPDSCKCDAGCSCK
			LelMT10b	MT 10b [23]	ABH03634	MPSPCNCRETGKCTCDGKCSGDACCCGANCNCGEDCKCQGCKTVVCKCSGDCACGKGCTGPDSCKCDAGCSCK
		Mactra quadrangularis	MquMT1	MT [24]	ACU46012	MGDPCDCVKDGQCTCQTSCNCAGKCDCTKCACKCSGSCACGKGCTGPESCKCGTDCSCK
		Mercenaria mercenaria	MmercMT1	MT [25]	AFH73818	MGDPCNCAETGSCNCSDQCTSDGGCRCGSNCKCGADCKCPGCKVVCKCGDSCACGKGCTGPSTCQCDSGCSCK
		Meretrix lamarckii	MlamMT1	MT	ALX35069	MSDPCNCIETGTCKCSEDCSSTTDCRCGAGCKCGDACKCPGCKVVCKCADNNGECKCGKGCTGPSSCKCDQSCSCK
		Meretrix lusoria	MluMT1	MT [26]	AAS92877	MSDPCNCIETGTCKCSEDCSSTTGCRCGAGCKCGDACKCPGCKVVCKCAENNGECKCGKGCTGPDSCKCDQGCSCK

		Meretrix meretrix	MmereMT1	MT [27]	ADM96223	MSDPCNCIETGTCKCSEDCSSTTGCRCGAGCKCGDACKCPGCKVVCKCAENNGECKCGKGCTGPDSCKCDQGCSCK
		Mytilus coruscus	McoruMT1	MT [28]	AUI39236	MPAPCNCIETNVCICDTGCSGQGCRCGDACKCSGDDCKCSGCKVVCKCSGTCACEGGCTGPKTCKCAPGCSCK
			McoruMT2	MT [28]	ATD87395	MPAPCNCIETNVCICATGCSDQGCRCGDACKCSGDDCKCSGCKVVCKCSGTCACEGGCTGPKTCKCEPGCSCK
		Mytilus edulis	MedMT10B	MT [15]	CAE11857	MPAPCNCIETNECICATRCSGEGCRCGECKCSGCKVVSACKCSGSCGCGGGCTPLTCKCAPGCSCK
			MedMT10IV	MT 10IV [29]	CAA07546	MPAPCNCIETNVCICDTGCSGEGCRCGDACKCSGADCKCSGCKVVCKCSGSCACEGGCTGPSTCKCAPGCSCK
			MedMT20II	MT 20 II [30]	CAA06553	MPGPCNCIETNVCICGTGCSGKCCRCGDACKCASGCGCSGCKVVCKCSGTCACGCDCTGPTNCKCESGCSCK
		Mytilus galloprovincialis	MgalMT10B	MT-10B, partial	ABI30643	MPAPCNCIETNECICATGCSGEGCRCGECKCSGCKVVSACKCSGSCGCGGGCTPLTCKCAPGCSCK
			MgalMT10III	MT 10-III [31]	AAT72936	MPAPCNCIESNVCICGTGCSGEGCRCGDACKCSGADCKCSGCKVVCKCSGSCACEAGCTGPSTCRCAPGCSCK
			MgalMT20III	MT 20-III isoform A [32]	P69154	MPGPCNCIETNVCICGTGCSGKCCQCGDACKCASGCGCSGCKVVCRCSGTCACGCDCTGPINCKCESGCSCK
			MgalMT20IV	MT 20-IV [31]	AAT72935	MAGPCNCIATNVCICGTGCSEKCCQCGDACKCESGCGCSGCKVVCRCSGTCACGCGCTGPTNCKCESGCSCN
		Ostrea edulis	OedMT1	MT [33]	CAC83769	MSDPCNCQKRNCQCSGICPSTGCNCASGCKCVAGCKCPGCKIVNTVASFCAYGIECIRPVTCKCAAGCTCK
			OedMT2	MT[33]	CAC83770	MSDPCNCKTTGNCLCSDTCPSTGCNCASGCKCGAGCKCPGCQIVCKCNSTCACGVGCTGPDTCKCAADSGCSCK
		Perna viridis	PviMT1	MT 1 [34]	AAF22486	MPSPCNCIETQVCICGTGCSGEGCRCGDACKCSSGCGCSGCKVVCKCQPGECACGKQCTGPDTCKCDSSCSCK
			PviMT2A	MT IIA	AEX60798	MPSPCNCNETQVCICGSGCSGEGCGCGDACKCDSDCGCSGCKIVCKCSDACACGKQCTEPKTCKCDSSCSCK
		Pinctada martensii	PmarMT1	MT	AGU91654	MPSPCNCATTSDCVCFNNCPTSGCNCAMSCKCSDSCPCPGCKSAVNSTVNCTCPGSCACGVGCTGPSECSCTSCSCKT
			PmarMT2	MT2	AGQ46322	MANVDQVIVVNCGCKLETSTCECGDDCKCGVNCACAICRTCSSCAENCSCGDNCKCGTRCPCKTCKRSCCNVAS
		Pinctada maxima	PmaxMT1	MT	ACJ22893	MSDPCRCAKTKEPCACTDACPASGCCCGSTCQCGDNCKCPNCKVKCACADKGICGCGVGCTTLSDCNCTDCSCKN
		Pisidium coreanum	PcorMT1	MT	ACT53273	MSDSCNCVEAGSCTCGDSCNGCKDCKCGSQCKCAKGCACKASCQCEPGNCTCGKSCDPAKDCSCVGECSCAKSSDCKKCKCKDSCDCDSACTGLSSCLCKVCKCK
		Ruditapes decussatus	RdeMT1partial	MT, partial	CAB96402	PCNCIETGTCVCSCSDSCPATGCKCGPGCKCGDGCKCSGCKVKCNCSGSCGCGKGCTGP
			RdeMT2partial	MT, partial [16]	ABS20116	IPCNCVETGSCQCSAQCTSEGGCKCGPNCKCGDGCACPGCKVVCNCNGSCACGQDCTGQSTCKCGSDCSCKK
		Ruditapes philippinarum	RphMT1	MT	AEB91530	MGDPCNCAETGTCKCSDQC2DGCKCGPNCNCGSDCSCPGCKVATCKCSGSCACGKGCTREGTCECGNDCSCKK
			RphMT2 partial	MT, partial	ABP57063	MSDPCDCAETGSCQCSAQCTSEGGCKCGPNCKCGDGCACPGCKVVCNCNGSCACGQDCTGPETCKC
		Saccostrea glomerata	SglMT1	Saccostrea glomerata MT I [35]	MN037799	MSDPCNCIETGKCLCADSCPASGCKCGPSCKCGPSCKCPGCKVKCNCNGTCGCGKGCTGPENCKCTDSSCACKK
		Scapharca broughtonii	SbrMT1	MT	ACH99846	MSDPCKCIEGGECKCDESCASDNCRCDPAKCKCKAGCKCSGCGVKCKCSGTCDCGQNCTGPADCKCKPNSCPCNQ
		Scapharca inaequivalvis	SinMT1	MT, partial	AAZ76545	MSDPCKCVEGGECKCDESCASDNCRCDPAKCKCKSGCKCSGCCGVKCKCSGNCDCGKSCTGP
		Tegillarca granosa	TgrMT1	MT	AAS75318	MSDPCKCVEGGECKCSDSCAGDNCRCDPEKCKCKDGCACSNCKVKCKCSGTCDPACGKNCTGPKDCKCPPNSCPCNQ
			TgrMT_like	MT-like protein, partial [36]	AAK39563	PCNCVKGGDCTCSSSCSGDSCRCGDSCKCASGCACSNCRVKCSCSGSCACGKGCKGPATCECAPGSCSCK
		Unio tumidus	UtuMT1	MT	ABP01350	MSDPCNCLETGECKCSGDCTGDCRCGDACKCGNACKCPGCKVVCKCSSSCDCGKGCTGPSTCKCDSGCSCK
Aculifera						
	Polyplacophora	Acanthochitona crinita	AcriMT1		(1)	MPDPCQCATTGTCQCGPECSGDGCKCGEGCKCGANCACPNCKVACKCAGTCNCGVGCTGEANCNCPNSCKCK
		Chaetopleura apiculata	CapMT1		(1)	MPDPCNCKETGTCKCTAECQGDACNCGEGCKCGSNCACPSCKKTCKCAGSCACGVGCTGPESCSCXNSCSCK
		Chiton olivaceus	ColMT1		(1)	MPCNCAKTGTCGCSAACQGDNCQCGPGCKCGADCPCPSCKSCKCTADNCTCGVGCKGPDTCKCKNSCGCK
		Leptochiton rugatus	LruMT1		(1)	MSDPCNCIQTGECKCSDACQGEGCKCGDGCKCGDSCKCPGCKKTCKCAGTCKCGADCTGPDTCKCDNSCSCK
		Tonicella lineata	TlinMT1		(1)	MPDPCNCATTGTCKCSSDCASSGCNCGSGCKCGSDCACPGCKKTCNCSAGNCACGAGCTGPDSCKCKSGCSCK
	Solenogastres	Alexandromenia crassa	AcraMT1 partial		(1)	SDPCNCADTGDCKCAADCDCKAKGCKCGAGCK
		Amphimeniidae sp.	AspMT1 partial		(1)	MSDPCKCAETGDCKCAADCDCTAKGCKCGAGC
		Micromenia fodiens	MfoMT1 partial		(1)	MPDKCNCAQGGECQCGSDCNCSDCKCDPDKCKAAQNFTVCKHN
		Neomenia carinata	NcaMT1		(1)	MSDPCKCAETGDCKCAADCDCTAKGCKCGAGCKSTQHKCKCTDSCACGKGCKSGEACKCGTDCSCR
		Neomenia megatrapezata	NmegMT1		(1)	MSDPCNCVETNECVCSKDCTSDSGCKCGPSCKCGDDCKCTGCKVKTAVTCKCSDPCSCEGGCTGPANCKCAAGCSCK
			NmegMT2		(1)	MSDPCNCAETGDCKCEAGCDCSKCNCDPAKCKVKNAKCCTDSACKCADGCKGGDGCKCQSDCGCCK
		Neomeniomorpha sp.	NspMT1		(1)	MSDPCNCAESGECQCKEgCDCTnCKCDPAKcKVKAAKKCCGKCECTDGCKGADGCKCAAGCDCCK
			NspMT2 partial		(1)	MSDPCNCAETGDCKCEAGCDCSKCNCDPAKCKVKNAKCCTDSACKCADGCKGGDGXKCQ
	Caudofoveata	Chaetoderma nitidulum	CniMT1		(1)	MSDVKPCNCADSCNCCCKDGACGDKCECKVKCCQSGDCGCCQDKCKCAGSCQCGKGCTGPGDCXCGSDCSCKKXR
		Falcidens caudatus	FcaMT1		(1)	MSDVKPCNCADSCNCCCKDGACAGKKCECKVKCCQSGDCGCCQDKCKCAGTCNCGKGCTGPGDCTCDSGCSCKK
		Falcidens sagittiferus	FsaMT1		(1)	MADDKPCNCADSCNCCCKSGACGDKCECGVKCCTSGDCGCCPXXCKCSAGSCKCGKGCTGPGDCKCAAGCACR
		Scutopus ventrolineatus	SveMT1		(1)	MSDVKPCDCADSCNCCCKSGPCGDKSGDKCECRVKCCQSGDCGCCKDKCKCTTGSCQCGKGCTGVADCKCDTGCSCRR

*NCBI accession numbers are provided when available.

(1) Sequences assembled from SRA projects deposited in NCBI

(2) Sequences from TSA projects deposited in NCBI

(3) Sequences from eSnail database (http://soft.bioinfo-minzhao.org/esnail/)

1. R. Dallinger *et al.* , Metallomics reveal a persisting impact of cadmium on the evolution of metal-selective metallothioneins. *Metallomics* , 1-26 (2020).
2. M. Dvorak *et al.* , Cadmium pathways in snails follow a complementary strategy between metallothionein detoxification and auxiliary inactivation by phytochelatins. *Int. J. Mol. Sci.* **21**, 1-14 (2020).
3. Ò. Palacios *et al.* , Analysis of metal-binding features of the wild type and two domain-truncated mutant variants of littorina littorea metallothionein reveals its cd-specific character. *Int. J. Mol. Sci.* **18** (2017), doi:10.3390/ijms18071452.
4. V. Pedrini-martha, S. Köll, M. Dvorak, R. Dallinger, Cadmium uptake, MT gene activation and structure of large-sized multi-domain metallothioneins in the terrestrial door snail alinda biplicata (Gastropoda, clausiliidae). *Int. J. Mol. Sci.* **21**, 1-33 (2020).
5. B. Berger, P. E. Hunziker, C. R. Hauer, N. Birchler, R. Dallinger, Mass spectrometry and amino acid sequencing of two cadmium-binding metallothionein isoforms from the terrestrial gastropod Arianta arbustorum. *Biochem. J.* **311**, 951-957 (1995).
6. M. Dvorak *et al.* , Metal binding functions of metallothioneins in the slug Arion vulgaris differ from metal-specific isoforms of terrestrial snails. *Metallomics* (2018), doi:10.1039/c8mt00215k.
7. Biomphalaria glabrata Metallothionein: Lacking Metal Specificity of the Protein and Missing Gene Upregulation Suggest Metal Sequestration by Exchange Instead of through Selective Binding. *Int. J. Mol. Sci.* (2017), doi:10.3390/ijms18071457.
8. F. Hispard *et al.* , Metal distribution and metallothionein induction after cadmium exposure in the terrestrial snail Helix aspersa (Gastropoda, pulmonata). *Environ. Toxicol. Chem.* **27**, 1533-1542 (2008).
9. M. Chabicovsky *et al.* , Localization and quantification of Cd- and Cu-specific metallothionein isoform mRNA in cells and organs of the terrestrial gastropod Helix pomatia. *Toxicol. Appl. Pharmacol.* **190**, 25-36 (2003).
10. A. S. Babenka, S. Dromashko, COPPER SULFATE IMPACTS ON GROWTH , SURVIVAL AND METALLOTHIONEIN EXPRESSION IN FRESHWATER POND SNAIL LYMNAEA (2011).
11. J. Jayasinghe, E. Jo, Q. Wan, S. Bathige, J. Lee, *SPECIFIC BIOLOGICAL RESPONSE MARKER FOR METAL POLLUTION FROM DISK ABALONE; MOLECULAR CHARACTERIZATION AND METAL TOLERANCE STUDY OF METALLOTHIONEIN (AbMT)* (2015).
12. B. Lieb, A new metallothionein gene from the giant keyhole limpet Megathura crenulata. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **134**, 131-137 (2003).

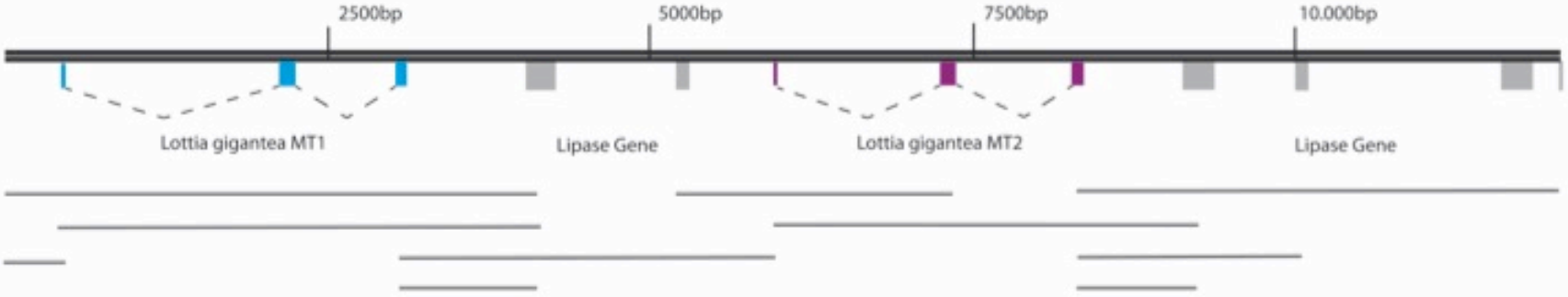
13. L. Wang, L. Song, D. Ni, H. Zhang, W. Liu, Alteration of metallothionein mRNA in bay scallop Argopecten irradians under cadmium exposure and bacteria challenge. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **149**, 50-57 (2009).
14. Y. Hardivillier *et al.* , Do organisms living around hydrothermal vent sites contain specific metallothioneins? The case of the genus Bathymodiolus (Bivalvia, Mytilidae). *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **139**, 111-118 (2004).
15. V. Leignel, Y. Hardivillier, M. Laulier, Small metallothionein MT-10 genes in coastal and hydrothermal mussels. *Mar. Biotechnol.* **7**, 236-244 (2005).
16. R. Ladhar-Chaabouni, R. Mokdad-Gargouri, F. Denis, A. Hamza-Chaffai, Cloning and characterization of cDNA probes for the analysis of metallothionein gene expression in the Mediterranean bivalves: Ruditapes decussatus and Cerastoderma glaucum. *Mol. Biol. Rep.* **36**, 1007-1014 (2009).
17. A. Bigot, P. Doyen, P. Vasseur, F. Rodius, Metallothionein coding sequence identification and seasonal mRNA expression of detoxification genes in the bivalve Corbicula fluminea. *Ecotoxicol. Environ. Saf.* **72**, 382-387 (2009).
18. T. Xu *et al.* , Identification of expressed genes in cDNA library of hemocytes from the RLO-challenged oyster, Crassostrea ariakensis Gould with special functional implication of three complement-related fragments (CaC1q1, CaC1q2 and CaC3). *Fish Shellfish Immunol.* **32**, 1106-1116 (2012).
19. A. Tanguy, D. Moraga, Cloning and characterization of a gene coding for a novel metallothionein in the Pacific oyster Crassostrea gigas (CgMT2): A case of adaptive response to metal-induced stress? *Gene* . **273**, 123-130 (2001).
20. M. J. Jenny, A. H. Ringwood, K. Schey, G. W. Warr, R. W. Chapman, Diversity of metallothioneins in the American oyster, Crassostrea virginica, revealed by transcriptomic and proteomic approaches. *Eur. J. Biochem.* **271**, 1702-1712 (2004).
21. M. J. Jenny, G. W. Warr, A. H. Ringwood, D. A. Baltzegar, R. W. Chapman, Regulation of metallothionein genes in the American oyster (Crassostrea virginica): Ontogeny and differential expression in response to different stressors. **379**, 156-165 (2006).
22. J. Engelken, A. Hildebrandt, cDNA cloning and cadmium-induced expression of metallothionein mRNA in the zebra mussel Dreissena polymorpha. *Biochem. Cell Biol.* (1999), doi:10.1139/o99-026.
23. H. Park, I. Y. Ahn, H. J. Choi, S. H. Pyo, H. E. Lee, Cloning, expression and characterization of metallothionein from the Antarctic clam Laternula elliptica. *Protein Expr. Purif.* **52**, 82-88 (2007).
24. Y. Fang *et al.* , Metallothionein and superoxide dismutase responses to sublethal cadmium exposure in the clam Mactra veneriformis. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **151**, 325-333 (2010).
25. M. Perrigault, B. Allam, Differential immune response in the hard clam (mercenaria mercenaria) against bacteria and the protistan pathogen QPX (quahog parasite unknown). *Fish Shellfish Immunol.* **32**, 1124-1134 (2012).
26. Y. T. Chang, K. J. Jong, B. K. Liao, S. M. Wu, Cloning and expression of metallothionein cDNA in the hard clam (Meretrix lusoria) upon cadmium exposure. *Aquaculture* . **262**, 504-513 (2007).
27. Q. Wang, X. Wang, X. Wang, H. Yang, B. Liu, Analysis of metallotionein expression and antioxidant enzyme activities in: Meretrix meretrix larvae under sublethal cadmium exposure. *Aquat. Toxicol.* **100**, 321-328 (2010).
28. M. Bao *et al.* , A novel biomarker for marine environmental pollution of CAT from Mytilus coruscus. *Mar. Pollut. Bull.* **127**, 717-725 (2018).
29. S. Lemoine, Y. Bigot, D. Sellos, R. P. Cosson, M. Laulier, Metallothionein isoforms in Mytilus edulis (Mollusca, Bivalvia): Complementary DNA characterization and quantification of expression in different organs after exposure to cadmium, zinc, and copper. *Mar. Biotechnol.* **2**, 195-203 (2000).
30. D. Barsyte, K. N. White, D. A. Lovejoy, Cloning and characterization of metallothionein cDNAs in the mussel Mytilus edulis L. digestive gland. *Comp. Biochem. Physiol. C. Pharmacol. Toxicol. Endocrinol.* **122**, 287-296 (1999).
31. F. Dondero *et al.* , Quantitative PCR analysis of two molluscan metallothionein genes unveils differential expression and regulation. *Gene* . **345**, 259-270 (2005).
32. N. Ceratto, F. Dondero, J. W. Van de Loo, B. Burlando, A. Viarengo, Cloning and sequencing of a novel metallothionein gene in Mytilus galloprovincialis Lam. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **131**, 217-222 (2002).
33. A. Tanguy *et al.* , Metallothionein genes in the European flat oyster Ostrea edulis: A potential ecological tool for environmental monitoring? *Mar. Ecol. Prog. Ser.* **257**, 87-97 (2003).
34. H. W. Khoo, K. H. Patel, Metallothionein cDNA, promoter, and genomic sequences of the tropical green mussel, Perna viridis. *J. Exp. Zool.* **284**, 445-453 (1999).
35. T. Yingprasertchai *et al.* , Characterisation of the metallothionein gene in the Sydney rock oyster and its expression upon metal exposure in oysters with different prior metal exposure histories. *Mar. Environ. Res.* **151** (2019), doi:10.1016/j.marenvres.2019.104775.
36. M. K. Chan, R. Othman, D. Zubir, S. Salmijah, Induction of a putative metallothionein gene in the blood cockle, Anadara granosa, exposed to cadmium. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **131**, 123-132 (2002).

Table S3: Specific primers used for gDNA and cDNA amplificationa and cloning. Sequences of the primers, annealing temperatures of single primers as well as applied annealing temperature in the PCR set up are listed.

PRIMER	SEQUENCE 5'3'	Tm (ºC)
Primers for gDNA amplification and sequenciation		
Fw1 extern	GCTCCCCTCCCGCTGTACCGCCAC	80.7
FwMT1	GTTATAAAAAGTACTTCTTAAAGTTTC	53.7
FwE1	GTCTACAGAATAACAGAAAAATGTC	56.7
FwMT2	GTCGTAAAAATTACTTCTCAAAGC	59
RvMTE3uni	GTCTACTCCTGTACAACCAACACC	62.5
RvLPE3	GTCGACTGTATACATTCAGACGGG	65.4
RvLPE1	GAGACAACGATTTTTAGCTATGGT	61.3
RvLPE4	GCTTGGGGTGACATTGATTTCTATC	67.2
Rv E1	GACATTTTTCTGTTATTCTGTAG	54
FwE3uni	GGTGTTGGTTGTACAGGAGTAGAC	62.5
RvLPE5	GCTGTGGAGATGGTCCCTGTCC	71.3
Rv E2 uni	ACCAGGTTTACAGCAATCAGC	63
FwLPE3	GTCTGAATGTATACAGTCGAC	53.5
FwE2 uni	GCTGATTGCTGTAAACCTGGT	63
Fw (FwE1-E2 MT1)	GGAAAATGATGTGTTTTGTGGTAAA	64.4
Fw (FwE1-E2 uni)	ATGATGTGTTTTGTGGTAAA	55.3
Fw (FwE2-E3 uni)	GCAAATTTTAATTAAGTTTGATTTGATC	61.7
Primers for cDNA amplification		
MT1 Fw	CCG CAG AGT ACA GCA ATT CAT A	60.1
MT1 Rev	GGT ATG TGA GTC AAC AAA GAG	57.5
MT2 Fw	CTT CTC AAA GCT TCT ACA GAA T	56.4
MT2 Rev	CAA GAT TGT GTG CAT GAT GAA	55.4

FIGURE S1

A.

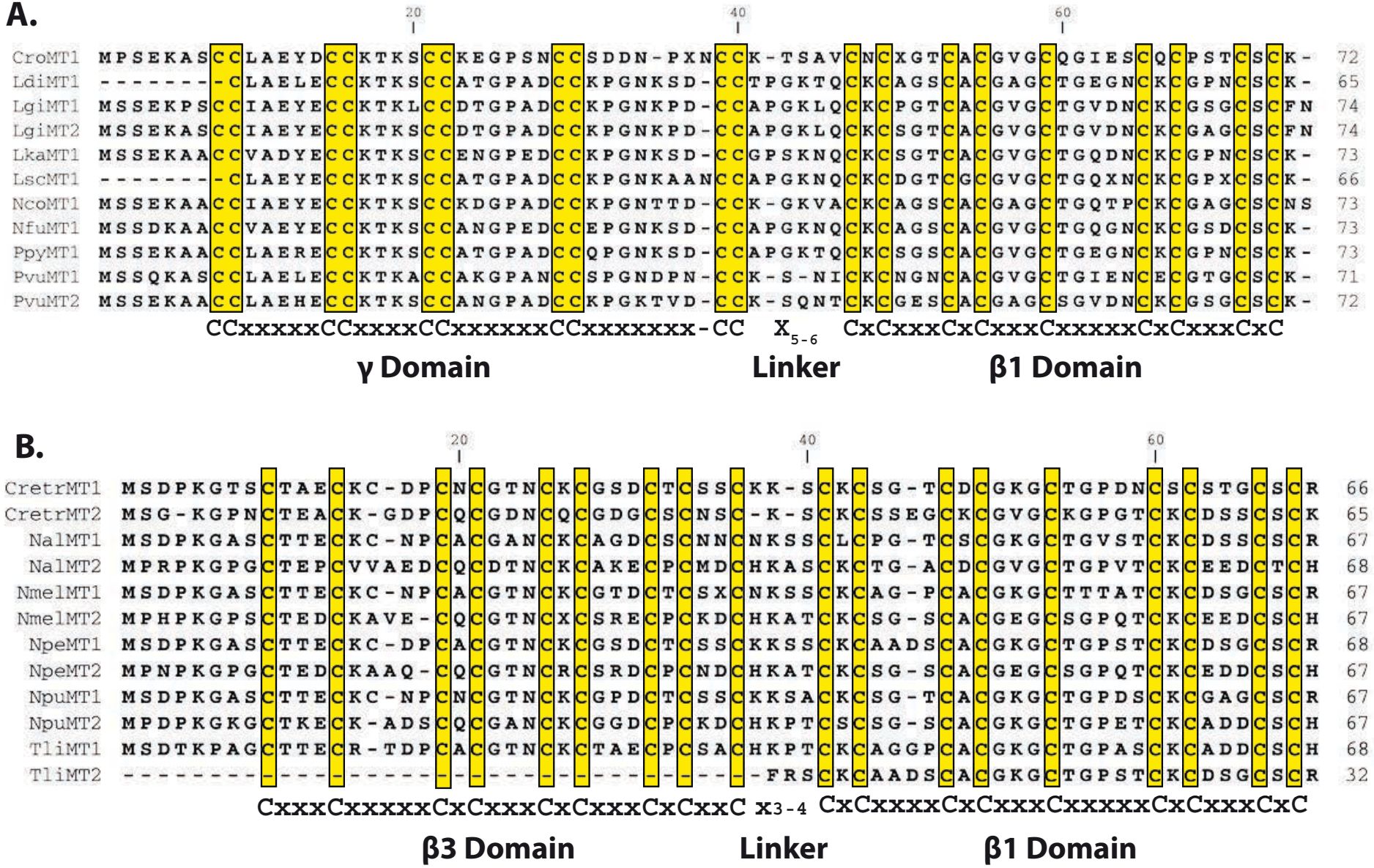


B.

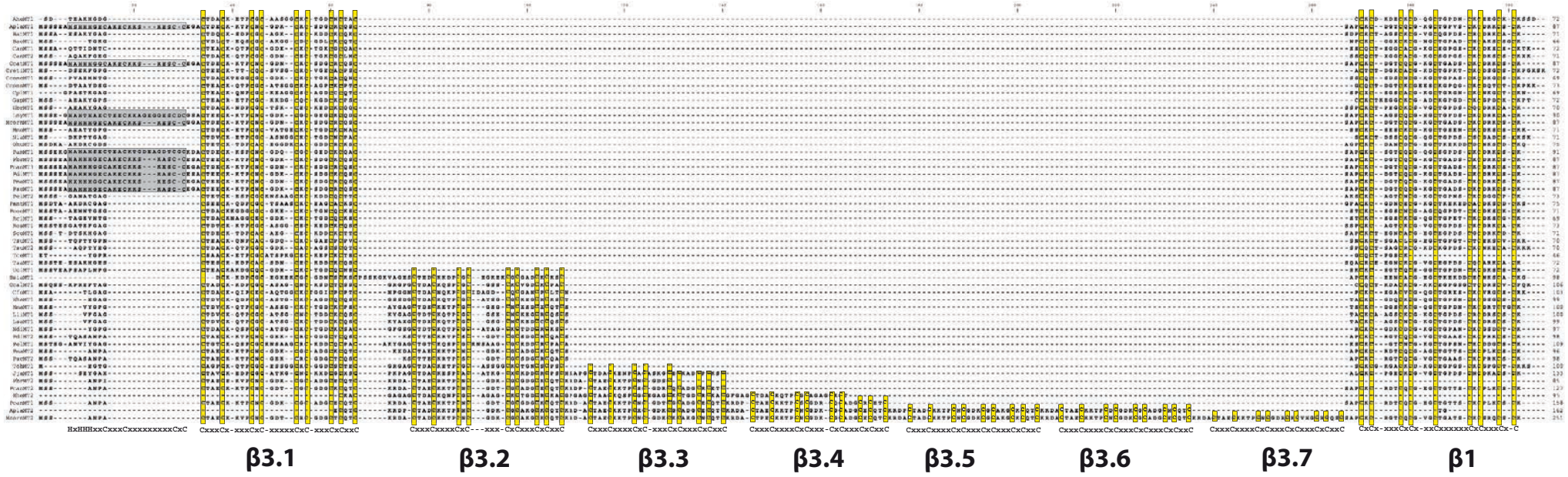
LgiMT1 MS[S/T]EK**P**S**CC**IAEYE**CC**KTK**L**CCDTGPAD**CC**KPGNKPD**CC**APGKLQ**C**K**C**P**G**T**C**A**C**GVG**C**TGVDN**C**K**C**G[S/A]G**C**S**C**FN 74

LgiMT2 MS[S/T]EK**A**S**CC**IAEYE**CC**KTK**S**CCDTGPAD**CC**KPGNKPD**CC**APGKLQ**C**K**C**S**G**T**C**A**C**GVG**C**TGVDN**C**K**C**G[S/A]G**C**S**C**FN 74

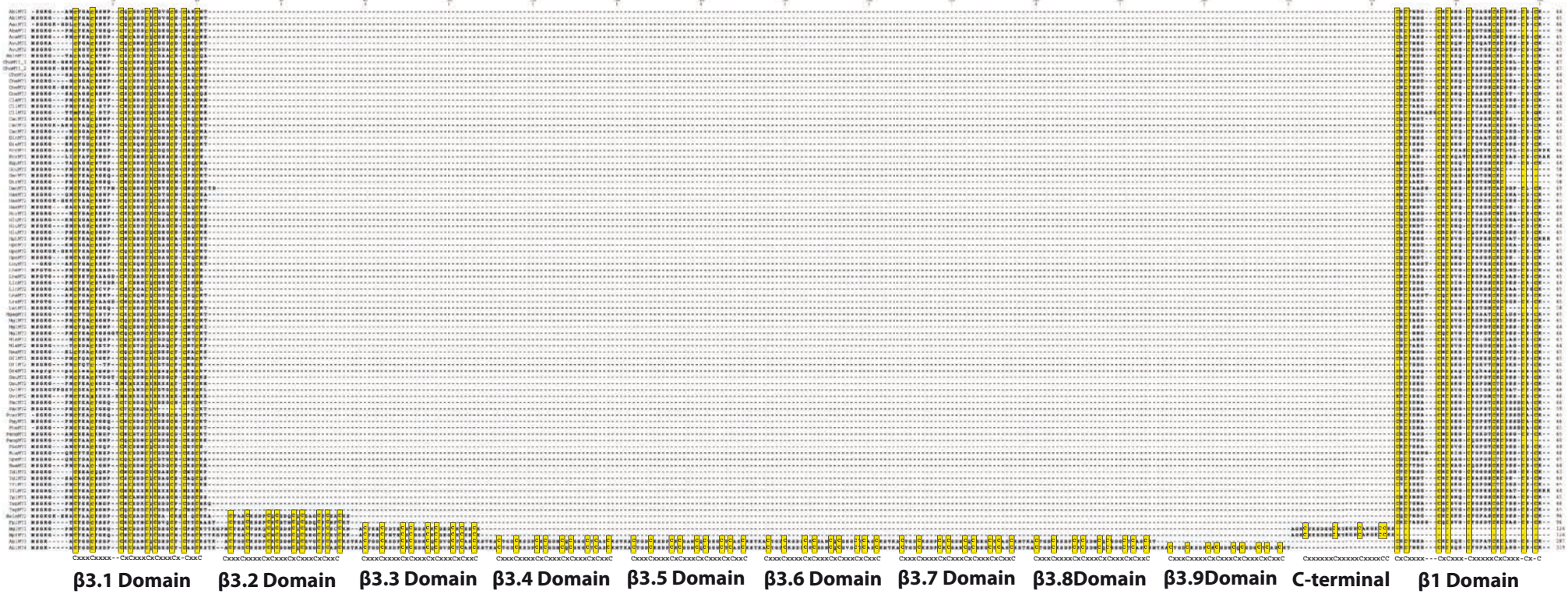
Figure S2


A.
20
40
60
CroMT1 MPSEKASCCLAEYDCCKTKSCCKEGPSNCCSDDN-PXNCCK-TSAVCNCXGTCACGVGCQGIESCQCPSTCSCK- 72
LdiMT1 --------CLAELECCKTKSCCATGPADCCKPGNKSD-CCTPGKTQCKCAGSCACGAGCTGEGNCKCGPNCSCK- 65
LgiMT1 MSSEKPSCCIAEYECCKTKLCCDTGPADCCKPGNKPD-CCAPGKLQCKCPGTCACGVGCTGVDNCKCGSGCSCFN 74
LgiMT2 MSSEKASCCIAEYECCKTKSCCDTGPADCCKPGNKPD-CCAPGKLQCKCSGTCACGVGCTGVDNCKCGAGCSCFN 74
LkaMT1 MSSEKAACCVADYECCKTKSCCENGPEDCCKPGNKSD-CCGPSKNQCKCSGTCACGVGCTGQDNCKCGPNCSCK- 73
LscMT1 --------CLAEYECCKTKSCCATGPADCCKPGNKAANCCAPGKNQCKCDGTCGCGVGCTGQXNCKCGPXCSCK- 66
NcoMT1 MSSEKAACCIAEYECCKTKSCCKDGPADCCKPGNTTD-CCK-GKVACKCAGSCACGAGCTGQTPCKCGAGCSCNS 73
NfuMT1 MSSDKAACCVAEYECCKTKSCCANGPEDCCEPGNKSD-CCAPGKNQCKCAGSCACGVGCTGQGNCKCGSDCSCK- 73
PpyMT1 MSSEKAACCLAERECCKTKSCCATGPADCCQPGNKSD-CCAPGKTQCKCSGSCACGVGCTGEGNCKCGPNCSCK- 73
PvuMT1 MSSQKASCCLAELECCKTKACCAKGPANCCSPGNDPN-CCK-S-NICKCNGNCACGVGCTGIENCECGTGCSCK- 71
PvuMT2 MSSEKAACCLAEHECCKTKSCCANGPADCCKPGKTVD-CCK-SQNTCKCGESCACGAGCSGVDNCKCGSGCSCK- 72
CCxxxxxCCxxxxCCxxxxxxCCxxxxxxx-CC X5-6 CxCxxxCxCxxxCxxxxxCxCxxxCxC
γ Domain
Linker
β1 Domain
B.
20
40
60
CretrMT1 MSDPKGTSCTAECKC-DPCNCGTNCKCGSDCTCSSCKK-SCKCSG-TCDCGKGCTGPDNCSCSTGCSCR 66
CretrMT2 MSG-KGPNCTEACK-GDPCQCGDNCQCGDGCSCNSC-K-SCKCSSEGCKCGVGCKGPGTCKCDSSCSCK 65
NalMT1 MSDPKGASCTTECKC-NPCACGANCKCAGDCSCNNCNKSSCLCPG-TCSCGKGCTGVSTCKCDSSCSCR 67
NalMT2 MPRPKGPGCTEPCVVAEDCQCDTNCKCAKECPCMDCHKASCKCTG-ACDCGVGCTGPVTCKCEEDCTCH 68
NmelMT1 MSDPKGASCTTECKC-NPCACGTNCKCGTDCTCSXCNKSSCKCAG-PCACGKGCTTTATCKCDSGCSCR 67
NmelMT2 MPHPKGPSCTEDCKAVE-CQCGTNCXCSRECPCKDCHKATCKCSG-SCACGEGCSGPQTCKCEEDCSCH 67
NpeMT1 MSDPKGASCTTECKC-DPCACGTNCKCGSDCTCSSCKKSSCKCAADSCACGKGCTGPSTCKCDSGCSCR 68
NpeMT2 MPNPKGPGCTEDCKAAQ-CQCGTNCRCSRDCPCNDCHKATCKCSG-SCACGEGCSGPQTCKCEDDCSCH 67
NpuMT1 MSDPKGASCTTECKC-NPCNCGTNCKCGPDCTCSSCKKSACKCSG-TCACGKGCTGPDSCKCGAGCSCR 67
NpuMT2 MPDPKGKGCTKECK-ADSCQCGANCKCGGDCPCKDCHKPTCSCSG-SCACGKGCTGPETCKCADDCSCH 67
TliMT1 MSDTKPAGCTTECR-TDPCACGTNCKCTAECPCSACHKPTCKCAGGPCACGKGCTGPASCKCADDCSCH 68
TliMT2 ---------------------------------------FRSCKCAADSCACGKGCTGPSTCKCDSGCSCR 32
CxxxCxxxxxCxCxxxCxCxxxCxCxxC x3-4 CxCxxxxCxCxxxCxxxxxCxCxxxCxC
β3 Domain
Linker
β1 Domain

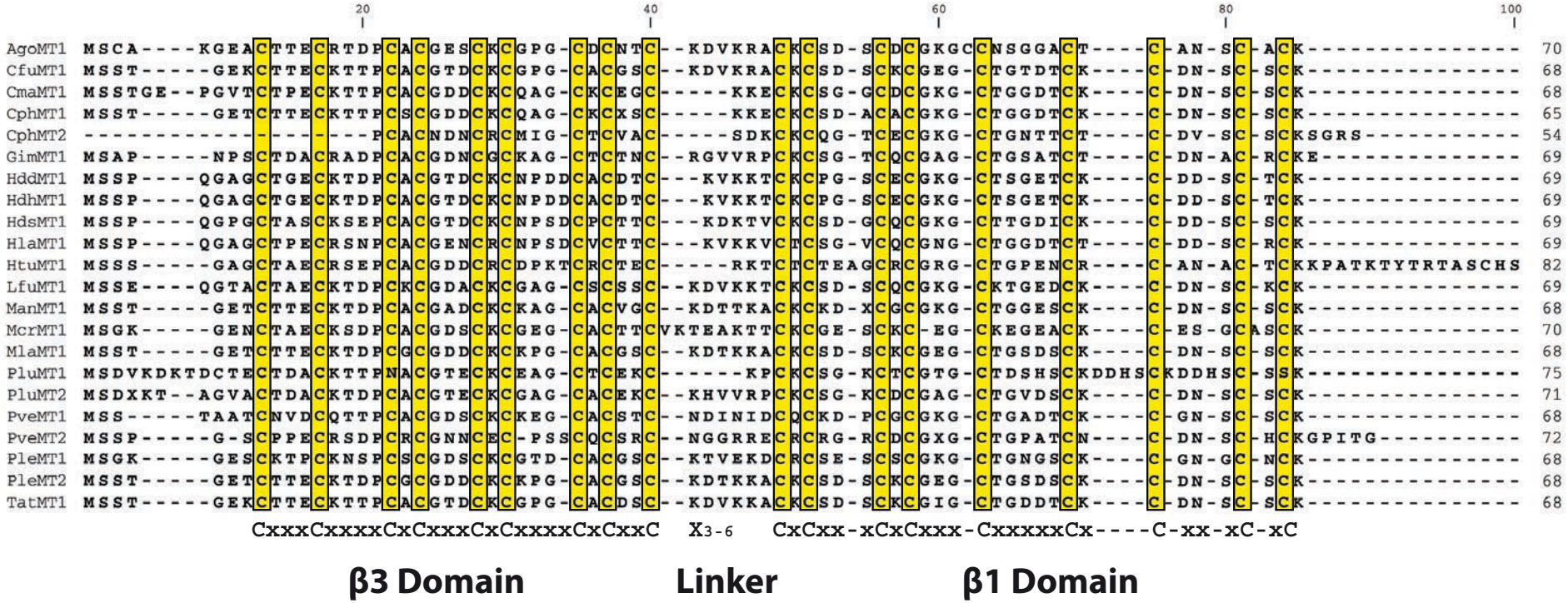
C.



D.

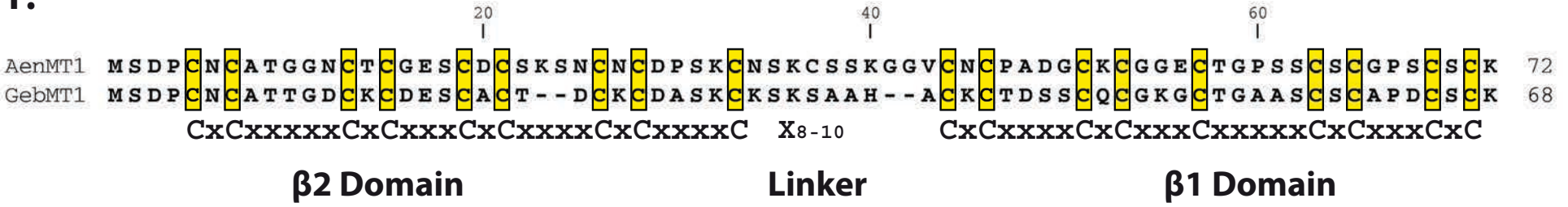

β3.1 Domain
β3.2 Domain
β3.3 Domain
β3.4 Domain
β3.5 Domain
β3.6 Domain
β3.7 Domain
β3.8Domain
β3.9Domain
C-terminal
β1 Domain

E.



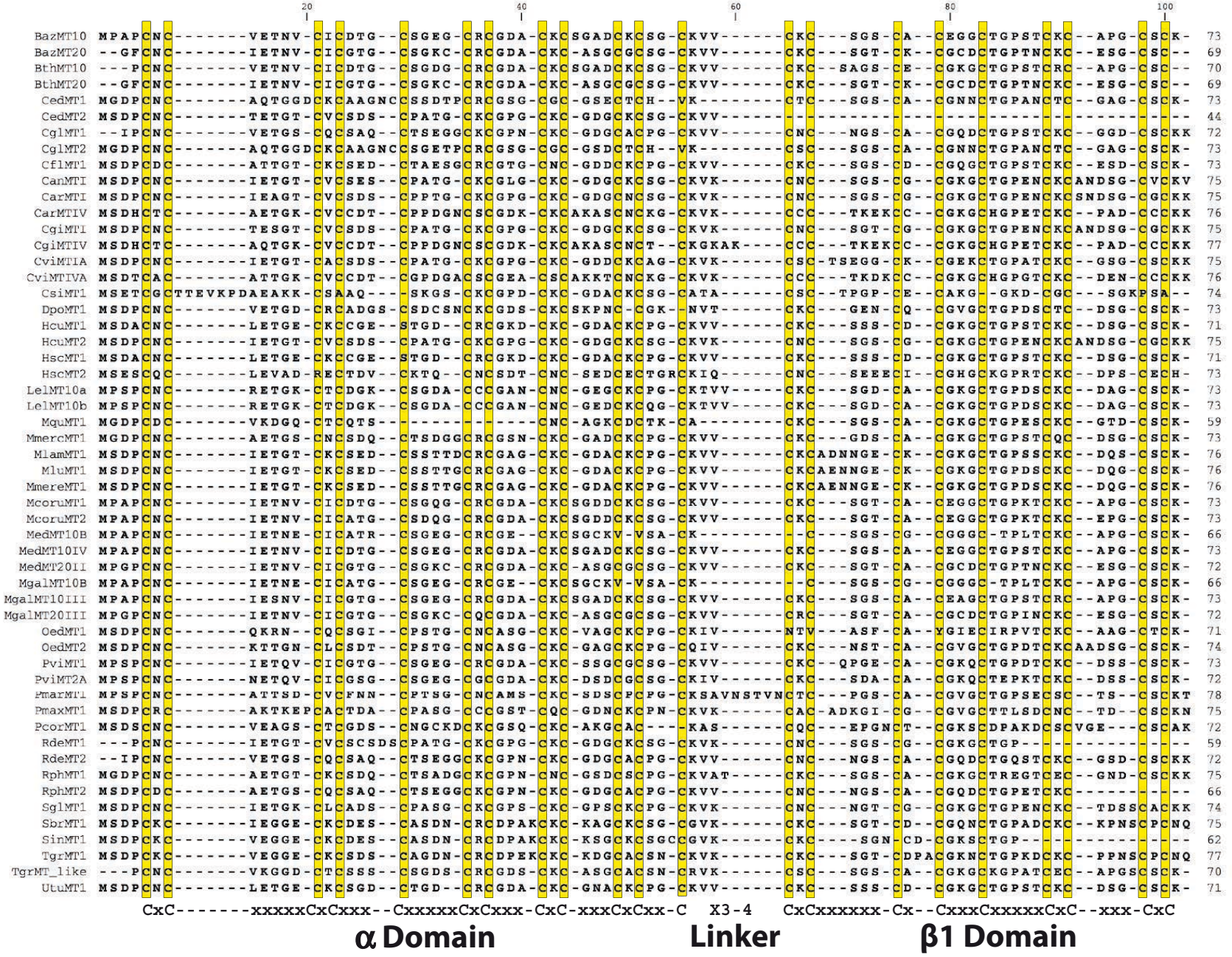
β3 Domain **Linker** **β1 Domain**

F.

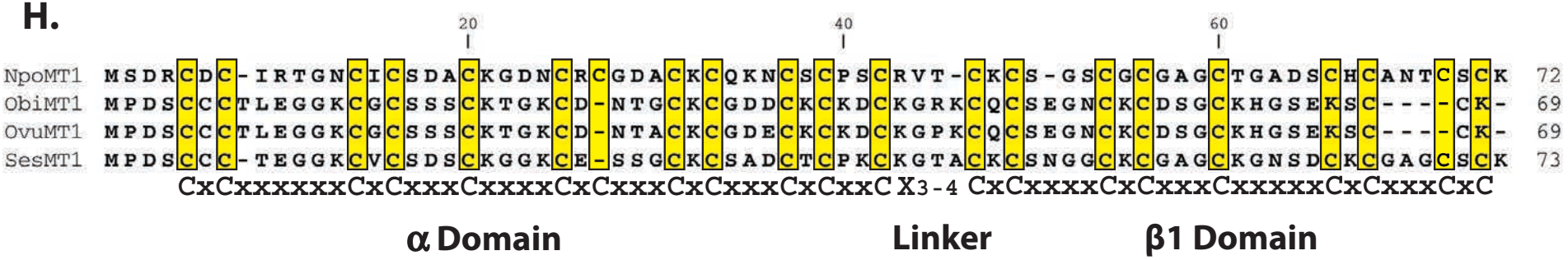


β2 Domain **Linker** **β1 Domain**

G.


20
40
60
80
100
BazMT10 MPAPCNC-------VETNV-CICDTG--CSGEG-CRCGDA-CKCSGADCKCSG-CKVV------CKC---SGS-CA--CEGGCTGPSTCKC--APG-CSCK- 73
BazMT20 --GFCNC-------IETNV-CICGTG--CSGKC-CRCGDA-CKC-ASGCGCSG-CKVV------CKC---SGT-CK--CGCDCTGPTNCKC--ESG-CSC-- 69
BthMT10 ---PCNC-------VETNV-CICDTG--CSGDG-CRCGDA-CKCSGADCKCSG-CKVV------CKC---SAGS-CE--CGKGCTGPSTCRC--APG-CSC-- 70
BthMT20 --GFCNC-------IETNV-CICGTG--CSGKC-CRCGDA-CKC-ASGCGCSG-CKVV------CKC---SGT-CK--CGCDCTGPTNCKC--ESG-CSC-- 69
CedMT1 MGDPCNC-------AQTGGDCKCAAGNCCSSDTPCRCGSG-CGC-GSECTCH--VK--------CTC---SGS-CA--CGNNCTGPANCTC--GAG-CSCK- 73
CedMT2 MSDPCNC-------TETGT-CVCSDS--CPATG-CKCGPG-CKC-GDGCKCSG-CKVV-- 44
CglMT1 --IPCNC-------VETGS-CQCSAQ--CTSEGGCKCGPN-CKC-GDGCACPG-CKVV------CNC---NGS-CA--CGQDCTGPSTCKC--GGD-CSCKK 72
CglMT2 MGDPCNC-------AQTGGDCKCAAGNCCSGETPCRCGSG-CGC-GSDCTCH--VK--------CSC---SGS-CA--CGNNCTGPANCTC--GAG-CSCK- 73
CflMT1 MSDPCDC-------ATTGT-CKCSED--CTAESGCRCGTG-CNC-GDDCKCPG-CKVV------CKC---SGS-CD--CGQGCTGPSTCKC--ESD-CSCK- 73
CanMTI MSDPCNC-------IETGT-CVCSES--CPATG-CKCGLG-CKC-GDGCKCSG-CKVK------CNC---SGS-CG--CGKGCTGPENCKCANDSG-CVCKV 75
CarMTI MSDPCNC-------IEAGT-CVCSDS--CPPTG-CKCGPG-CKC-GDGCNCSG-CKVK------CNC---SGS-CG--CGKGCTGPENCKCSNDSG-CGCKK 75
CarMTIV MSDHCTC-------AETGK-CVCCDT--CPPDGNCSCGDK-CKCAKASCNCKG-CKVK------CCC---TKEKCC--CGKGCHGPETCKC--PAD-CCCKK 76
CgiMT1 MSDPCNC-------TESGT-CVCSDS--CPATG-CKCGPG-CKC-GDGCKCSG-CKVK------CNC---SGT-CG--CGKGCTGPENCKCANDSG-CGCKK 75
CgiMTIV MSDHCTC-------AQTGK-CVCCDT--CPPDGNCSCGDK-CKCAKASCNCT--CKGKAK----CCC---TKEKCC--CGKGCHGPETCKC--PAD-CCCKK 77
CviMT1A MSDPCNC-------IETGT-CACSDS--CPATG-CKCGPG-CKC-GDDCKCAG-CKVK------CSC-TSEGG-CK--CGEKCTGPATCKC--GSG-CSCKK 75
CviMTIVA MSDTCAC-------ATTGK-CVCCDT--CGPDGACSCGEA-CSCAKKTCNCKG-CKVK------CCC---TKDKCC--CGKGCHGPGTCKC--DEN-CCCKK 76
CsiMT1 MSETCGCTTEVKPDAEAKK-CSAAQ----SKGS-CKCGPD-CKC-GDACKCSG-CATA------CSC--TPGP-CE--CAKG--GKD-CGC---SGKPSA-- 74
DpoMT1 MSDPCNC-------VETGD-CRCADGS-CSDCSNCKCGDS-CKCSKPNC-CGK--NVT------CKC---GEN-CQ--CGVGCTGPDSCTC--DSG-CSCK- 73
HcuMT1 MSDACNC-------LETGE-CKCCGE--STGD--CRCGKD-CKC-GDACKCPG-CKVV------CKC---SSS-CD--CGKGCTGPSTCKC--DSG-CSCK- 71
HcuMT2 MSDPCNC-------IETGT-CVCSDS--CPATG-CKCGPG-CKC-GDGCKCSG-CKVK------CNC---SGS-CG--CGKGCTGPENCKCANDSG-CGCKK 75
HscMT1 MSDACNC-------LETGE-CKCCGE--STGD--CRCGKD-CKC-GDACKCPG-CKVV------CKC---SSS-CD--CGKGCTGPSTCKC--DSG-CSCK- 71
HscMT2 MSESCQC-------LEVAD-RECTDV--CKTQ--CNCSDT-CNC-SEDCECTGRCKIQ------CNC---SEEECI--CGHGCKGPRTCKC--DPS-CECH- 73
LelMT10a MPSPCNC-------RETGK-CTCDGK--CSGDA-CCCGAN-CNC-GEGCKCPG-CKTVV-----CKC---SGD-CA--CGKGCTGPDSCKC--DAG-CSCK- 73
LelMT10b MPSPCNC-------RETGK-CTCDGK--CSGDA-CCCGAN-CNC-GEDCKCQG-CKTVV-----CKC---SGD-CA--CGKGCTGPDSCKC--DAG-CSCK- 73
MquMT1 MGDPCDC-------VKDGQ-CTCQTS----------------CNC-AGKCDCTK-CA--------CKC---SGS-CA--CGKGCTGPESCKC--GTD-CSCK- 59
MmercMT1 MGDPCNC-------AETGS-CNCSDQ--CTSDGGCRCGSN-CKC-GADCKCPG-CKVV------CKC---GDS-CA--CGKGCTGPSTCQC--DSG-CSCK- 73
MlamMT1 MSDPCNC-------IETGT-CKCSED--CSSTTDCRCGAG-CKC-GDACKCPG-CKVV------CKCADNNGE-CK--CGKGCTGPSSCKC--DQS-CSCK- 76
MluMT1 MSDPCNC-------IETGT-CKCSED--CSSTTGCRCGAG-CKC-GDACKCPG-CKVV------CKCAENNGE-CK--CGKGCTGPDSCKC--DQG-CSCK- 76
MmereMT1 MSDPCNC-------IETGT-CKCSED--CSSTTGCRCGAG-CKC-GDACKCPG-CKVV------CKCAENNGE-CK--CGKGCTGPDSCKC--DQG-CSCK- 76
McoruMT1 MPAPCNC-------IETNV-CICDTG--CSGQG-CRCGDA-CKCSGDDCKCSG-CKVV------CKC---SGT-CA--CEGGCTGPKTCKC--APG-CSCK- 73
McoruMT2 MPAPCNC-------IETNV-CICATG--CSDQG-CRCGDA-CKCSGDDCKCSG-CKVV------CKC---SGT-CA--CEGGCTGPKTCKC--EPG-CSCK- 73
MedMT10B MPAPCNC-------IETNE-CICATR--CSGEG-CRCGE--CKCSGCKV-VSA-CK---------C---SGS-CG--CGGGC-TPLTCKC--APG-CSCK- 66
MedMT10IV MPAPCNC-------IETNV-CICDTG--CSGEG-CRCGDA-CKCSGADCKCSG-CKVV------CKC---SGS-CA--CEGGCTGPSTCKC--APG-CSCK- 73
MedMT20II MPGPCNC-------IETNV-CICGTG--CSGKC-CRCGDA-CKC-ASGCGCSG-CKVV------CKC---SGT-CA--CGCDCTGPTNCKC--ESG-CSCK- 72
MgalMT10B MPAPCNC-------IETNE-CICATG--CSGEG-CRCGE--CKCSGCKV-VSA-CK---------C---SGS-CG--CGGGC-TPLTCKC--APG-CSCK- 66
MgalMT10III MPAPCNC-------IESNV-CICGTG--CSGEG-CRCGDA-CKCSGADCKCSG-CKVV------CKC---SGS-CA--CEAGCTGPSTCRC--APG-CSCK- 73
MgalMT20III MPGPCNC-------IETNV-CICGTG--CSGKC-CQCGDA-CKC-ASGCGCSG-CKVV------CRC---SGT-CA--CGCDCTGPINCKC--ESG-CSCK- 72
OedMT1 MSDPCNC-------QKRN--CQCSGI--CPSTG-CNCASG-CKC-VAGCKCPG-CKIV------NTV---ASF-CA--YGIECIRPVTCKC--AAG-CTCK- 71
OedMT2 MSDPCNC-------KTTGN-CLCSDT--CPSTG-CNCASG-CKC-GAGCKCPG-CQIV------CKC---NST-CA--CGVGCTGPDTCKCAADSG-CSCK- 74
PviMT1 MPSPCNC-------IETQV-CICGTG--CSGEG-CRCGDA-CKC-SSGCGCSG-CKVV------CKC--QPGE-CA--CGKQCTGPDTCKC--DSS-CSCK- 73
PviMT2A MPSPCNC-------NETQV-CICGSG--CSGEG-CGCGDA-CKC-DSDCGCSG-CKIV------CKC---SDA-CA--CGKQCTEPKTCKC--DSS-CSCK- 72
PmarMT1 MPSPCNC-------ATTSD-CVCFNN--CPTSG-CNCAMS-CKC-SDSCPCPG-CKSAVNSTVNCTC---PGS-CA--CGVGCTGPSECSC--TS--CSCKT 78
PmaxMT1 MSDPCRC-------AKTKEPCACTDA--CPASG-CCCGST-CQC-GDNCKCPN-CKVK------CAC-ADKGI-CG--CGVGCTTLSDCNC--TD--CSCKN 75
PcorMT1 MSDSCNC-------VEAGS-CTCGDS--CNGCKDCKCGSQ-CKC-AKGCAC----KAS------CQC---EPGNCT--CGKSCDPAKDCSCVGE---CSCAK 72
RdeMT1 ---PCNC-------IETGT-CVCSCSDSCPATG-CKCGPG-CKC-GDGCKCSG-CKVK------CNC---SGS-CG--CGKGCTGP---------------- 59
RdeMT2 --IPCNC-------VETGS-CQCSAQ--CTSEGGCKCGPN-CKC-GDGCACPG-CKVV------CNC---NGS-CA--CGQDCTGQSTCKC--GSD-CSCKK 72
RphMT1 MGDPCNC-------AETGT-CKCSDQ--CTSADGCKCGPN-CNC-GSDCSCPG-CKVAT-----CKC---SGS-CA--CGKGCTREGTCEC--GND-CSCKK 75
RphMT2 MSDPCDC-------AETGS-CQCSAQ--CTSEGGCKCGPN-CKC-GDGCACPG-CKVV------CNC---NGS-CA--CGQDCTGPETCKC---------- 66
SglMT1 MSDPCNC-------IETGK-CLCADS--CPASG-CKCGPS-CKC-GPSCKCPG-CKVK------CNC---NGT-CG--CGKGCTGPENCKC--TDSSCACKK 74
SbrMT1 MSDPCKC-------IEGGE-CKCDES--CASDN-CRCDPAKCKC-KAGCKCSG-CGVK------CKC---SGT-CD--CGQNCTGPADCKC--KPNSCPCNQ 75
SinMT1 MSDPCKC-------VEGGE-CKCDES--CASDN-CRCDPAKCKC-KSGCKCSGCCGVK------CKC----SGN-CD-CGKSCTGP---------------- 62
TgrMT1 MSDPCKC-------VEGGE-CKCSDS--CAGDN-CRCDPEKCKC-KDGCACSN-CKVK------CKC---SGT-CDPACGKNCTGPKDCKC--PPNSCPCNQ 77
TgrMT_like ---PCNC-------VKGGD-CTCSSS--CSGDS-CRCGDS-CKC-ASGCACSN-CRVK------CSC---SGS-CA--CGKGCKGPATCEC--APGSCSCK- 70
UtuMT1 MSDPCNC-------LETGE-CKCSGD--CTGD--CRCGDA-CKC-GNACKCPG-CKVV------CKC---SSS-CD--CGKGCTGPSTCKC--DSG-CSCK- 71
CxC-------xxxxxCxCxxx--CxxxxxCxCxxx-CxC-xxxCxCxx-C X3-4 CxCxxxxxx-Cx--CxxxCxxxxxCxC--xxx-CxC
α Domain
Linker
β1 Domain

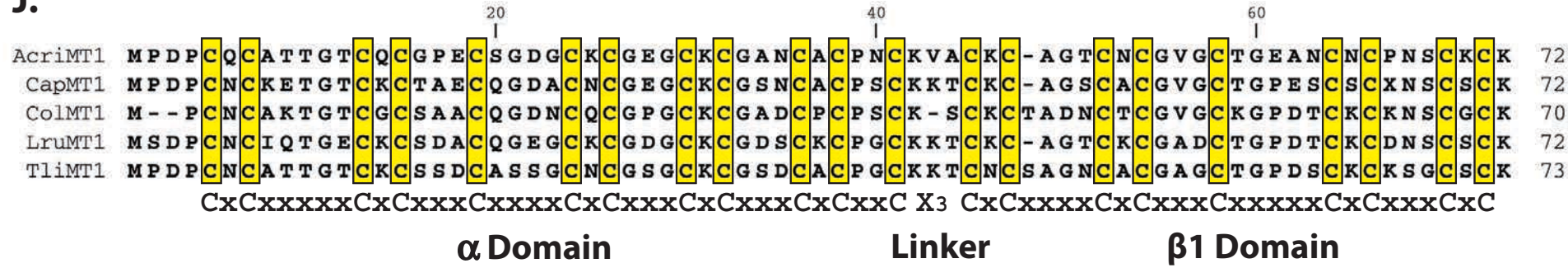
H.


20
40
60
NpoMT1 MSDRCDC-IRTGNCICSDACKGDNCRCGDACKCQKNCSCPSCRVT-CKCS-GSCGCGAGCTGADSCHCANTCSCK 72
ObiMT1 MPDSCCCTLEGGKCGCSSSCKTGKCD-NTGCKCGDDCKCKDCKGRKCQCSEGNCKCDSGCKHGSEKSC---CK- 69
OvuMT1 MPDSCCCTLEGGKCGCSSSCKTGKCD-NTACKCGDECKCKDCKGPKCQCSEGNCKCDSGCKHGSEKSC---CK- 69
SesMT1 MPDSCCC-TEGGKCVCSDSCKGGKCE-SSGCKCSADCTCPKCKGTACKCSNGGCKCGAGCKGNSDCKCGAGCSCK 73
CxCxxxxxxCxCxxxCxxxxCxCxxxCxCxxxCxCxxC X3-4 CxCxxxxCxCxxxCxxxxxCxCxxxCxC
α Domain
Linker
β1 Domain

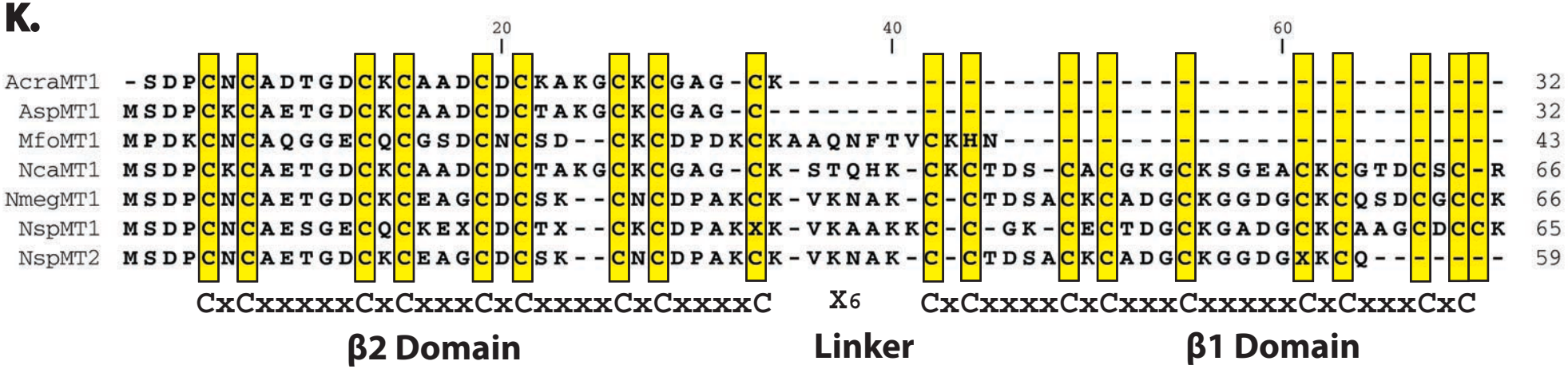
I.


20
40
60
LhyMT1 MPDPCTCSETGESCKCDSPCTCTDCKCACAKCKTVCKCPGTCKCGNDCTGPENCKCKPDCSCK
CxCxxxxxxCxCxxxCxCxxCxCxCxxC X3 CxCxxxCxCxxxCxxxxxCxCxxxCxC
β2 Domain
Linker
β1 Domain

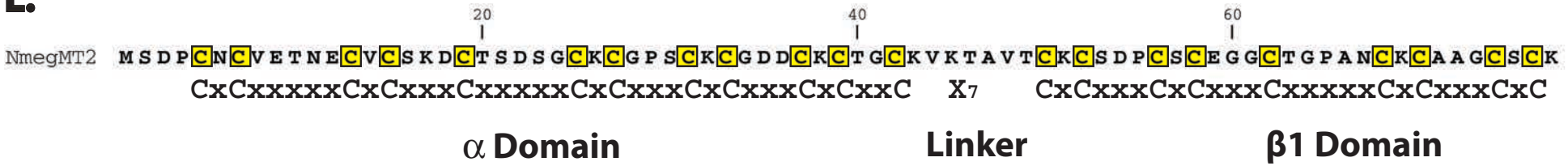
J.


20
40
60
AcriMT1 MPDPCQCATTGTCQCGPECSGDGCKCGEGCKCGANCACPNCKVACKC-AGTCNCGVGCTGEANCNCPNSCKCK 72
CapMT1 MPDPCNCKETGTCKCTAECQGDACNCGEGCKCGSNCACPSCKKTCKC-AGSCACGVGCTGPESCSCXNSCSCK 72
ColMT1 M--PCNCAKTGTCGCSAACQGDNCQCGPGCKCGADCPCPSCK-SCKCTADNCTCGVGCKGPDTCKCKNSCGCK 70
LruMT1 MSDPCNCIQTGECKCSDACQGEGCKCGDGCKCGDSCKCPGCKKTCKC-AGTCKCGADCTGPDTCKCDNSCSCK 72
TliMT1 MPDPCNCATTGTCKCSSDCASSGCNCGSGCKCGSDCACPGCKKTCNCSAGNCACGAGCTGPDSCKCKSGCSCK 73
CxCxxxxxCxCxxxCxxxxCxCxxxCxCxxxCxCxxC X3 CxCxxxxCxCxxxCxxxxxCxCxxxCxC
α Domain
Linker
β1 Domain

K.


20
40
60
AcraMT1 -SDPCNCADTGDCKCAADCDCKAKGCKCGAG-CK-------------------------- 32
AspMT1 MSDPCKCAETGDCKCAADCDCTAKGCKCGAG-C---------------------------- 32
MfoMT1 MPDKCNCAQGGECQCGSDCNCSD--CKCDPDKCKAAQNFTVCKHN---------------- 43
NcaMT1 MSDPCKCAETGDCKCAADCDCTAKGCKCGAG-CK-STQHK-CKCTDS-CACGKGCKSGEACKCGTDCSC-R 66
NmegMT1 MSDPCNCAETGDCKCEAGCDCSK--CNCDPAKCK-VKNAK-C-CTDSACKCADGCKGGDGCKCQSDCGCCK 66
NspMT1 MSDPCNCAESGECQCKEXCDCTX--CKCDPAKXK-VKAAKKC-C-GK-CECTDGCKGADGCKCAAGCDCCK 65
NspMT2 MSDPCNCAETGDCKCEAGCDCSK--CNCDPAKCK-VKNAK-C-CTDSACKCADGCKGGDGXKCQ------ 59
CxCxxxxxCxCxxxCxCxxxxCxCxxxxC X6 CxCxxxxCxCxxxCxxxxxCxCxxxCxC
β2 Domain
Linker
β1 Domain

L.


20
40
60
NmegMT2 MSDPCNCVETNECVCSKDCTSDSGCKCGPSCKCGDDCKCTGCKVKTAVTCKCSDPCSCEGGCTGPANCKCAAGCSCK
CxCxxxxxCxCxxxCxxxxxCxCxxxCxCxxxCxCxxC X7 CxCxxxCxCxxxCxxxxxCxCxxxCxC
α Domain
Linker
β1 Domain

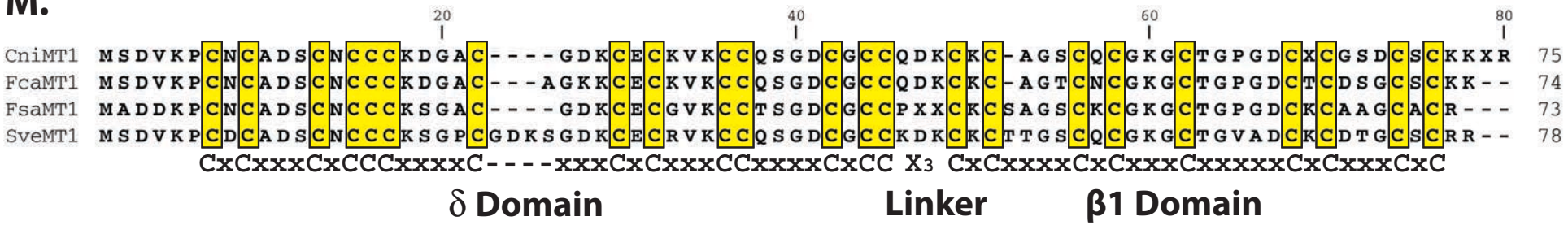
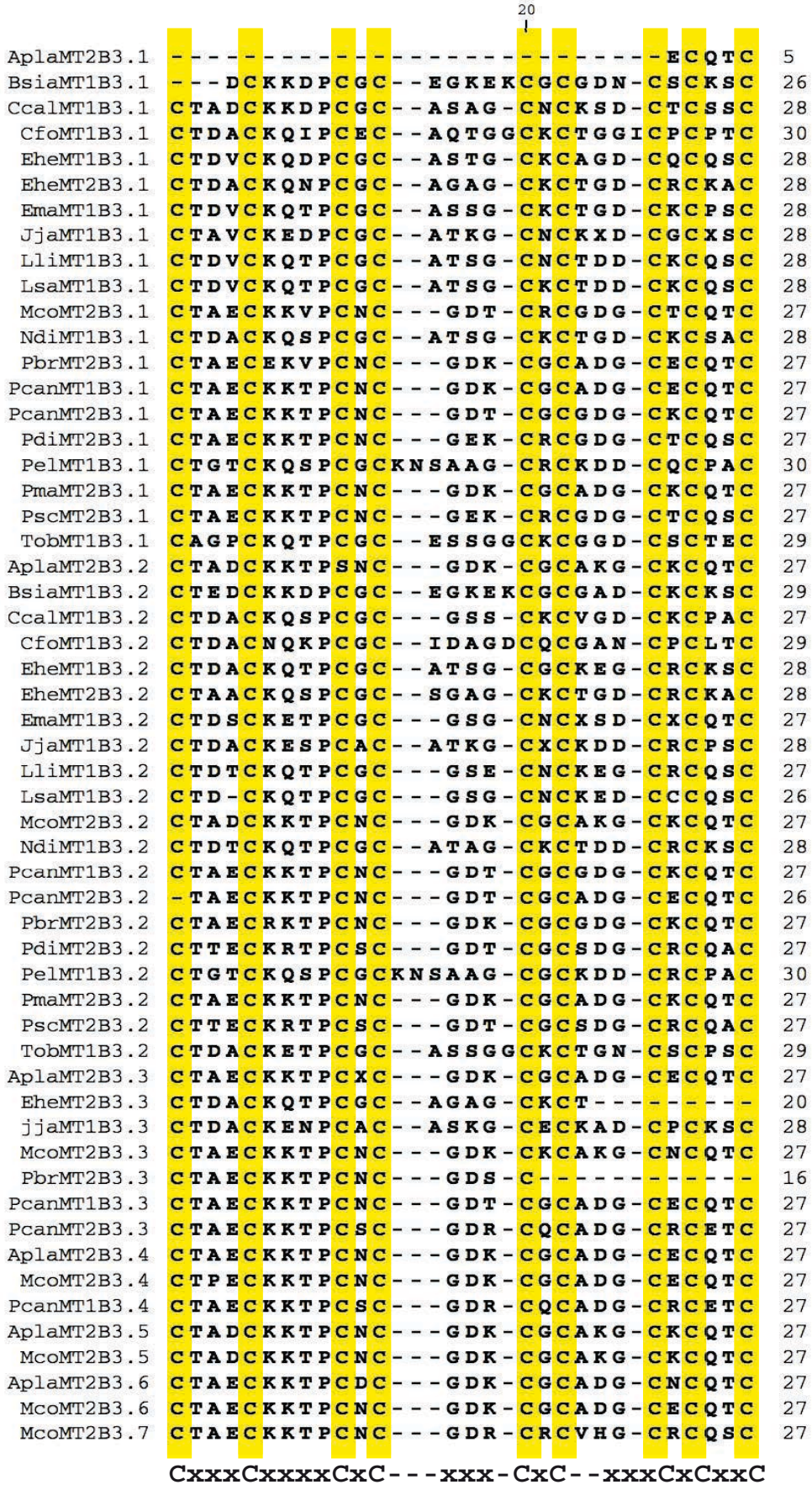


Figure S2. Amino acid alignment of Mollusca MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs at the N-terminal and C-terminal domains are showed below each alignment. Names of MT sequences from different species are as in **Table S2**, and alphabetically ordered. **(A)** Alignment of Patellogastropoda MTs showing a bi-modular γβ1 structure. **(B)** Alignment of Neritimorpha MTs showing a β3β1 structure. **(C)** Alignment of Caenogastropoda MTs showing a bi-modular β3β1 structure, with exception of 19 multi-modular MTs that have a variable number of internal β3 domain duplications, from β3.2 to β3.7 **(Figure S3A)**. MTs have been listed in alphabetical order except multi-modular MTs that have been organized by the number (from 2 to 7) of repeated domains. An additional $H_{3\text{-}4}C_4$ motif at the N-terminal region of 10 Architaenioglossa MTs are also depicted (grey box). **(D)** Alignment of Heterobranchia MTs showing a bi-modular β3β1 structure, with exception of 4 multi-modular MTs that have one or two internal β3 domain duplications **(Figure S3B)**. Notice that the two Biomphalaria MTs, BglMT1 and BpfMT1, have an additional C-terminal tail with 5 cysteines but lack the β1 domain. MTs have been listed in alphabetical order except multi-modular MTs that have been organized by the number of repeated β3 domains. **(E)** Alignment of Vetigastropoda MTs showing a β3β1 structure. **(F)** Alignment of Scaphopoda MTs showing a β2β1 structure. **(G)** Alignment of Bivalvia MTs. Archetypal bivalvia MTs show a αβ1 structure. Bivalvia MTs with other structural organizations have been described elsewhere (Tanguy and Moraga 2001; Baek, et al. 2009; Jenny, et al. 2016; Nam and Kim 2017). **(H)** Alignment of Cephalopoda MTs showing an αβ1 structure. **(I)** Monoplacophora MT from *L. hyalina* has a β2β1 structure. **(J)** Alignment of Polyplacophora MTs showing an αβ1 structure. **(K)** Alignment of Solenogastres MTs showing a β2β1 structure with the exception of **(L)** NmegMT2 sequence from *N. megatrapezata* that has an αβ1 strcuture. **(M)** Alignment of Caudofoveata MTs showing a bi-modular δβ1 structure.

FIGURE S3

A.



β3Domain

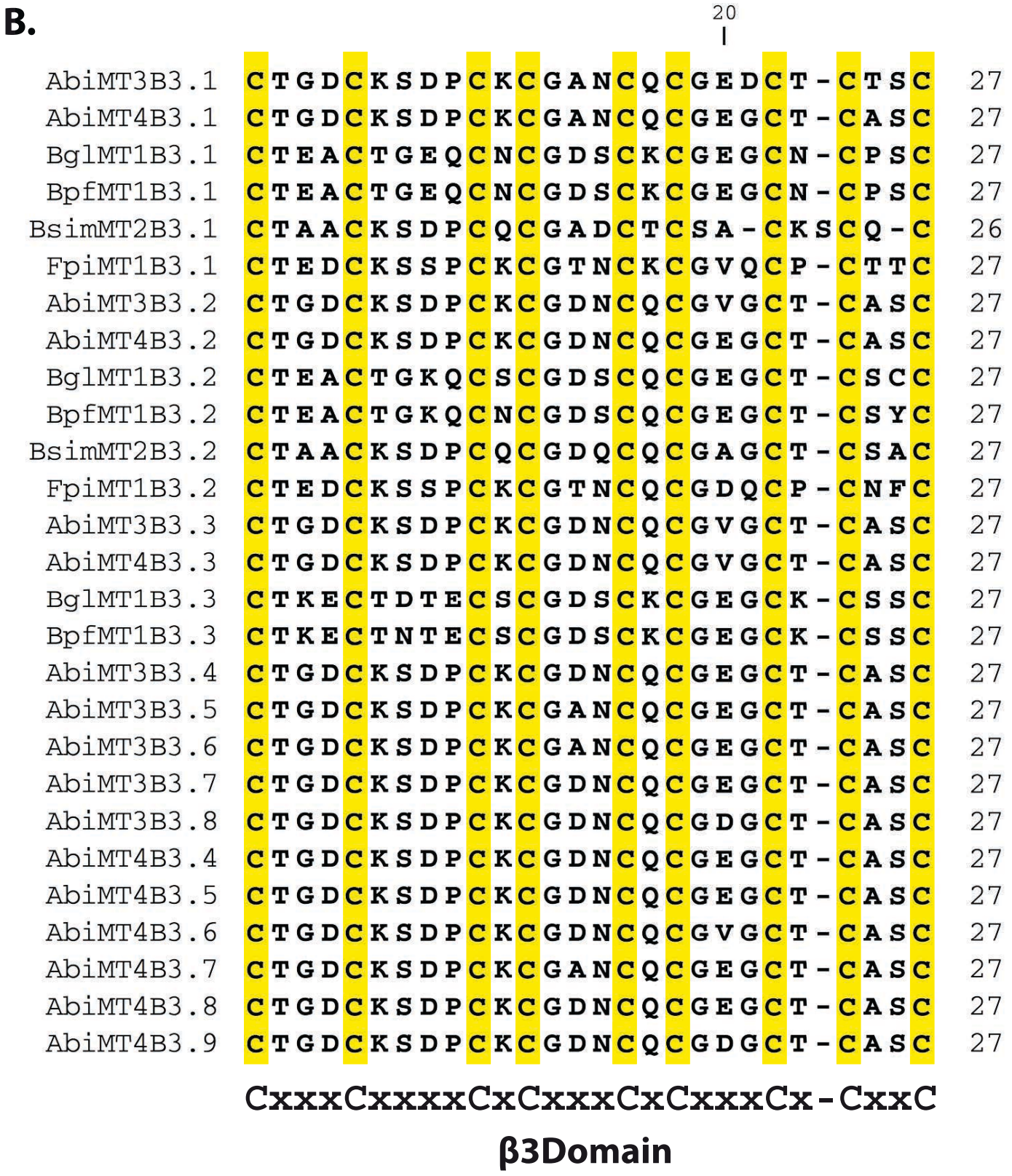


Figure S3. Amino acid alignment of duplicated β3 domains. (A) Repeated β3 domains in Caenogastropoda MTs. **(B)** Repeated β3 domains in Heterobranchia MTs. Conserved cysteines are highlighted in yellow. The configuration of the cysteine motifs for the β3 domains is showed below the alignment. Species abbreviations are as in Table S2.

ARTICLE II: *Exploring the modular organization of metallothionein in Molluscs*

In Preparation

Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ARTICLE TYPE

Exploring the modular organization of metallothionein in Molluscs.

Sara Calatayud[1] | Mario García-Risco[2] | Òscar Palacios[2] | Mercè Capdevila[2] | Ricard Albalat*[1]

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain
[2]Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Correspondence
*Ricard Albalat. Email: ralbalat@ub.edu

Present Address
Av.Diagonal 643, 08028 Barcelona, Spain

Abstract

Many proteins show a modular structure since they are built by the combination of structural and functional domains that can be used and recombined to form new organizations. The study of protein modularity is challenging, and the origin, evolution, function and rearrangements of the domains remain poorly understood. Metallothioneins are modular proteins that have emerged as a suitable model system to study protein modularity because they have well- defined structural domains that are experimentally tractable, and that can be recognized in wide phylogenetic contexts. Comparative analyses of MTs from phylogenetically diverse mollusk species, for instance, have led to propose that most mollusk MTs are bi-modular proteins that combine six putative domains - α, β1, β2, β3, γ and δ - in a lineage-specific manner. Some of these mollusk MTs and several of their domains have been characterized, but these analyses have been restricted to some species of the Bivalvia and Gastropoda classes of the Conchifera clade, and to their β and γ domains. Nothing is known, however, about the MTs from other Conchifera classes such as Cephalopoda, nor from any Caudofoveata species of the Aculifera clade, and the functional features of their α and δ domains are also unknown. To fill this gap, we have now studied the Cephalopoda α β1 and the Caudofoveata δ β1 MTs. We have experimentally confirmed [1] the modular organization of these mollusk MTs demonstrating that α and δ domains are functionally autonomous to bind metal ions; and [2] the preference of these MTs for divalent metal ions, which is determined, at least in part, by the metal preference of their respective α and δ domains. Our results are compatible with a cadmium preference for the ancestral mollusk MT, and with an old origin of the α domain that contrast with the de *novo* evolution of the δ domain..

KEYWORDS:
Modularity, Metallothioneins domains, Molluscs, Modular protein evolution.

1 | INTRODUCTION

Metallothioneins (MTs) have classically been considered low molecular weight (<100 amino acids) and cysteine-rich (15%-30%) proteins that bind metal ions, and therefore, they have been related with the control of metal homeostasis and detoxification (reviewed in (Capdevila and Atrian, 2011)). Their cysteine (Cys, C) residues are arranged in CC, CxC or CCC motifs, whose

[0]**Abbreviations:** MT: metallothioneins, MS: Mass Spectrometry

number and distribution define the structural domains that form metal-thiolate clusters. Vertebrate MTs, for instance, have 20 Cys distributed in two domains: an amino-terminal β-domain made of $[CxC]x_n[CxC]x_n[CxC]x_n[CxC]x_nCx_n$ motifs that binds three divalent metal ions in a M_3Cys_9 cluster, and a carboxyl-end domain with $CCxCCx_nCx_nCx_n[CxC]x_nCx_nCC$ motifs that coordinates four divalent ions in a M_4Cys_{11} cluster (reviewed in (Blindauer, 2014)). These structural domains seem to be the key pieces in MT evolution operating as evolutionary modules that might be gained, lost, modified or combined with other (similar or different) domains to build new MTs. This modular mode of evolution has been showed, for instance, in MTs of Chordata and Molluscan phyla, in which ancestral bi-modular proteins would have given rise highly diverse forms, from small mono-modular MTs made of a single domain, to large multi-modular MTs comprising multiple tandem domain repeats (Calatayud, 2018; Calatayud et al., 2021c, 2021a, 2021b; Guirola et al., 2012; Jenny et al., 2016, 2004; Nam and Kim, 2017; Niederwanger et al., 2017; Palacios et al., 2017). Based on the number and distribution of Cys residues, mollusk MT domains have been classified in four types: α domains made of 11/12 Cys, β domains containing 9 Cys (β1, β2 and β3 with different Cys motif arrangements), γ domains with 10 Cys, and δ domains made of 14 Cys (Calatayud et al., 2021c). Most mollusk bi-modular MTs combine a conserved carboxy-terminal β1 domain with an amino-terminal α, β2, β3, γ or δ domain, which would have been gained or lost in a lineage-specific manner: α domains seem restricted to Bivalvia, Cephalopoda and Solenogastres (a.k.a. Neomeniamorpha) classes, γ domains are only present in the Patellogastropoda subclass of gastropods, and δ domains have been only described in the Caudofoveata, (a.k.a. Chaetodermomorpha) class (Calatayud et al., 2021c). Based on the species distribution of these MT domains, it has been proposed that α and β domains are ancestral, whereas γ and δ would have evolved de novo during mollusk diversification, probably concomitant with the appearance of the Patellogastropoda and Caudofoveata taxonomic groups, respectively (Calatayud et al., 2021c). Mollusk β and γ domains have been functionally characterized, and their metal preferences as well as their binding capacities have been determined (Calatayud et al., 2021c). In contrast, functional information on the α and δ domains is scarce or totally absent. To fill this gap, we have now analyzed the MTs and their domains of two singular mollusk species, the 'living fossil' Nautilus pompilius of the Cephalopoda class, and the Falcidens caudatus, a Caudofoveata species of the Aculifera group that includes mollusc species that have no shell. The MT of N. pompilius, NpoMT1, is a bi-modular MT with α and β1 domains, whereas that of F. caudatus, FcaMT1, is also a bi-modular MT made of δ and β1 domains. We have characterized the metal-binding abilities of both NpoMT1 and FcaMT1 after their heterologous expression in *E. coli*. We have also determined the abilities of the α and δ domains to form metal-protein complexes when expressed alone. Our results extend the MT analysis to two mollusk classes previously uncharacterized confirming the cadmium preference for the ancestral mollusk form, supporting de *novo* emergence of MT domains, and completing the characterization of the metal-biding preferences and capacities of the different MT domains. Overall, these data improve our understanding of the domain organization and modular evolution of mollusk MTs.

2 | RESULTS AND DISCUSSION

2.1 | Heterologous expression and production of metal-MT complexes of NpoMT1 and FcaMT1, and of α and δ domains.

In order to study modular organization of mollusk MTs and characterize their α and δ domains, we selected the bimodular α/β1-MT1 of N. pompilius and the δ/β1-MT1 of F. caudatus as representative forms. Based on comparative analyses, NpoMT1 would exemplify the prototypical cephalopod MT made of an α domain with its 12 Cys arranged in $[CxC]x_5[CxC]x_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$ motifs, and a β1 domain with its 9 Cys arranged in $[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$ motifs (**Figure 1A**). FcaMT1 would be representative of the MTs of the Caudofoveata class and consisted in a δ domain with 14 Cys arranged in $[CxC]x_3[CxC]CCx_4Cx_4[CxC]x_3CCx_4[CxC]C$ motifs, and a β1 domain with its 9 Cys arranged in $[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$ motifs (**Figure 1B**). In order to characterize the biochemical properties and metal-binding features of these MTs, both proteins were expressed as GST-fusion proteins in*E. coli* BL21 cultures supplemented with $ZnCl_2$, $CdCl_2$ or $CuSO_4$. The α and domains from NpoMT1 and FacMT1, respectively, were also expressed alone in *E. coli* in order to determine the ability of these domains to independently form metal complexes.

Metal-protein complexes were purified from *E. coli* total protein extracts by a GST-affinity system, followed by the cleavage with thrombin of the GST tag, and a FPLC chromatography. Notice that the digestion with thrombin of the GST-MT fusion proteins resulted in the addition of two extra residues, Glycine and Serine, at the N-terminal end of the purified MTs. As shown in previous studies, these two amino acids do not interfere with the metal-binding features of recombinant MTs (Espart et al., 2015). The FPLC fractions containing the metal-MT complexes were characterized by ESI-MS analyses (**Figure 2**).

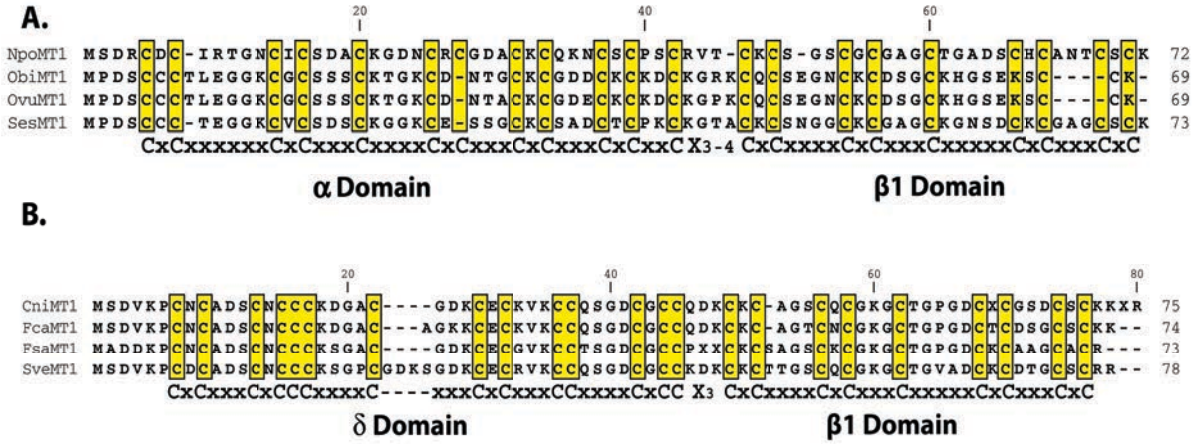


FIGURE 1 Amino acid alignment with conserved cysteines highlighted in yellow. Configurations of cysteine motifs at the N-terminal and C-terminal domains are showed below each alignment. (A) Alignment of Cephalopod MTs showing a bi-modular $\alpha\beta1$ structure. (B) Alignment of Caudofoveata MTs showing a bi-modular $\delta\beta1$ structure.

The experimental masses corresponding to the apo-NpoMT1 and apo-FcaMT1 (7471.8 Da and 7593.6Da, respectively) after demetallation by acidification of Cd-NpoMT1 and Cd-FcaMT21 nicely matched with the theoretical masses (7472.52 Da and 7594.79Da). The same held true for the synthesis of αNpoMT1 and δFcaMT1 domains (**Figure 2**), whose experimental mass (4872.0Da 4821.0Da, respectively) perfectly matched with the expected ones (4872.57 Da and 4821.59 Da).

2.2 | Metal bindings features of NpoMT1 and FcaMT1, and of α and δ domains.

NpoMT1 (21 Cys) and FcaMT1 (23 Cys) produced in Cd(II)-enriched culture medium yielded metallated species with seven and eight Cd(II) ions, respectively (**Figure 3A and C**) indicating that both MTs bind Cd(II) with high efficiency. Interestingly, the ICP-AES and ESI-MS results obtained from the independent expression of α and δ domains showed that these domains mainly rendered Cd_4- and Cd_5-MT complexes, respectively (Figure 3B and D). Considering that the 9-Cys of the $\beta1$ domains binds three divalent metal ions (ref), the metal-to-protein stoichiometries of the α and δ domains fully agreed with those of the full MTs, Cd_7-NpoMT1 (4+3=7 Cd(II) ions) and Cd_8-FcaMT1 (5+3=8 Cd(II) ions), and indicated that both were autonomous domains that can form single and structurally well-defined metal-protein complexes. These data further supported the modular organization and function of mollusk MTs. Our results showed that when synthesized in a Zn(II)-enriched culture medium, FcaMT1 and δFcaMT1 rendered single Zn_8- and Zn_5-protein complexes, respectively, whereas NpoMT1 and αNpoMT1 yielded a mixture of Zn-protein complexes, ranging from four (minor) to seven (major) metal ions for the NpoMT1, and three (minor) or four (major) for the α domain (**Figure 3E, F, G and H**). These differences suggested a stronger Cd-thionein character for NpoMT1 than for FcaMT1, which can also efficiently coordinate Zn(II) ions. FcaMT1 would be therefore a Cd/Zn-thionein. These results agreed with the analysis of the MT productions in a Cu(II)-enriched culture medium (see below). The ICP-AES and ESI-MS analyses of the MT productions in a Cu(II)-enriched culture medium showed that both NpoMT1 and FcaMT1, and their respective α and δ domains, rendered a mixture of Cu-MT complexes (**Figure 3I J, K and L**). The fact that both MTs did not give rise to a unique and therefore thermodynamically favored Cu-MT species unambiguously ascertains that neither NpoMT1 nor FcaMT1 behaves as a Cu-thionein. Interestingly, the metal binding of NpoMT1 and FcaMT1 in front of Cu ions was clearly different. FcaMT1 rendered a mixture of heteronuclear Zn, Cu-MT complexes, while NpoMT1 yields an assortment of homonuclear Cu-MT complexes (**Figure 3I and J**). Homometallic Cu(I) species in the Cu-preparations is characteristic of Cd-thioneins, whereas Zn-thioneins yield heterometallic Zn/Cu-MT complexes when expressed Cu-enriched media (Palacios et al., 2011). The homonuclear nature of the Cu-NpoMT1 complexes supported therefore its strong Cd-thionein character, whereas the detection of Zn(II) ions in the Cu-enriched synthesis of FcaMT1 implied that this MT was also able of coordinating Zn(II) ions. This different behavior – formation of homonuclear or heteronuclear species– was also observed in the analyses of the corresponding α and δ domains (**Figure 3K and L**): αNpoMT1 rendered homonuclear species while δFcaMT1 yielded heteronuclear species. These observations suggested that these α and δ domains significantly contribute to determine the metal preference of the full proteins.

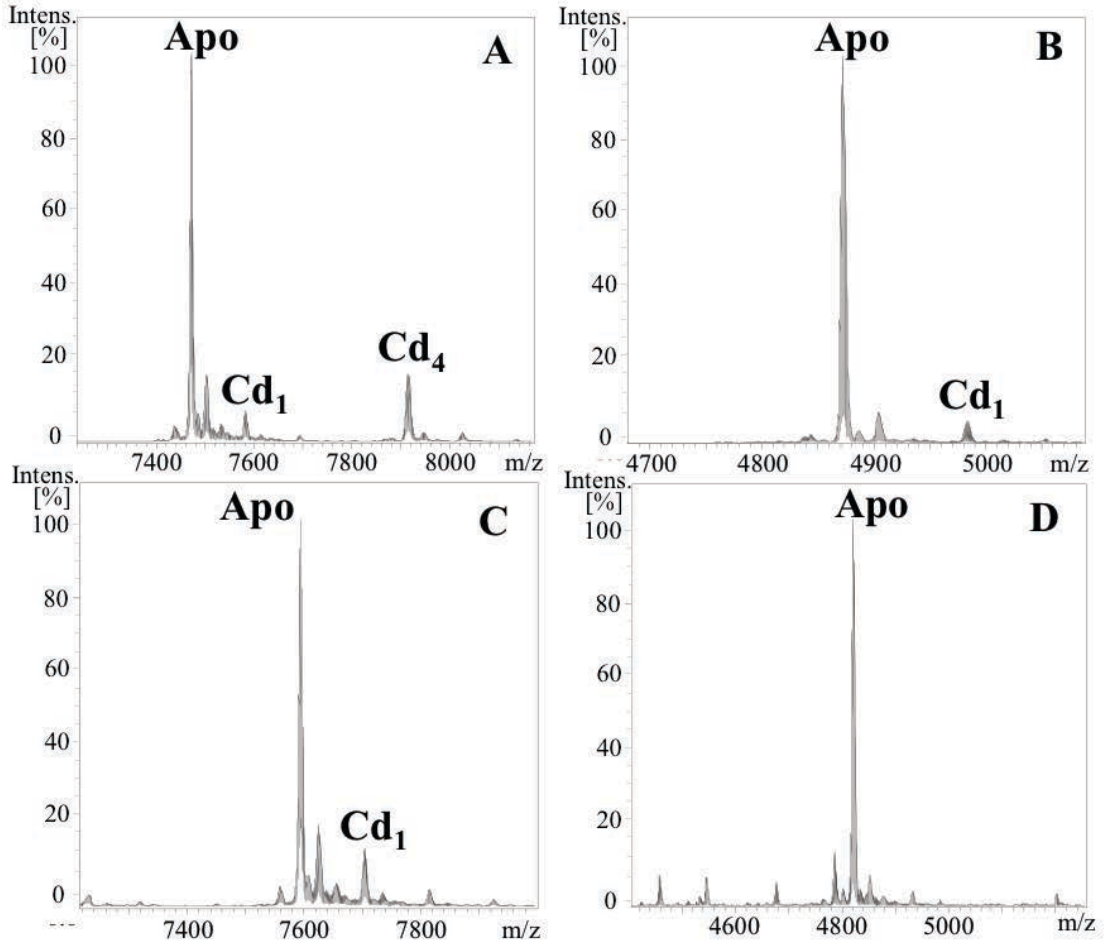


FIGURE 2 Deconvoluted mass spectra measured at pH 2.4 of the products of the recombinant synthesis of (A) Cd-NpoMT1 and (C) Cd-FcaMT1 and their fragments (B) Cd-αNpoMT1 and (D) Cd-δFcaMT1.

3 | CONCLUSIONS

Our results extended the Cd-binding ability of MTs to two new mollusk classes, Cephalopoda and Caudofoveata, further supporting the idea that the ancestral mollusc MT may have had a Cd-thionein character, and that this metal preference have been conserved by the MTs of many marine species likely due to an important influence of Cd in sea ecosystems through geological eras (Calatayud et al., 2021c; Dallinger et al., 2020). Our results also confirmed the modular organization of mollusk MTs demonstrating that the predicted α and δ domains were functional domains that autonomously coordinate divalent metal ions. Interestingly, whereas the α domain probably is one of the ancient domains in the mollusk MTs, the δ domain seems have been emerged de *novo* in the Caudofoveata lineage and have substituted previous N-terminal domains. To elucidate the molecular mechanisms for such domain changes and the adaptive causes that drove them will be a stimulating challenge.

4 | MATERIAL AND METHODS

4.1 | Production and purification of recombinant metal-MT complexes

Production and purification of recombinant metal-MT complexes of NpoMT, FcaMT, αNpoMT and δFcaMT domains was performed as described elsewhere (Calatayud, et al. 2018; Calatayud, et al. 2021b). In brief, synthetic cDNAs codifying the different constructs were provided by Synbio Technologies (Monmouth Junction, NJ, USA), cloned in the pGEX-4T-1 expression vector

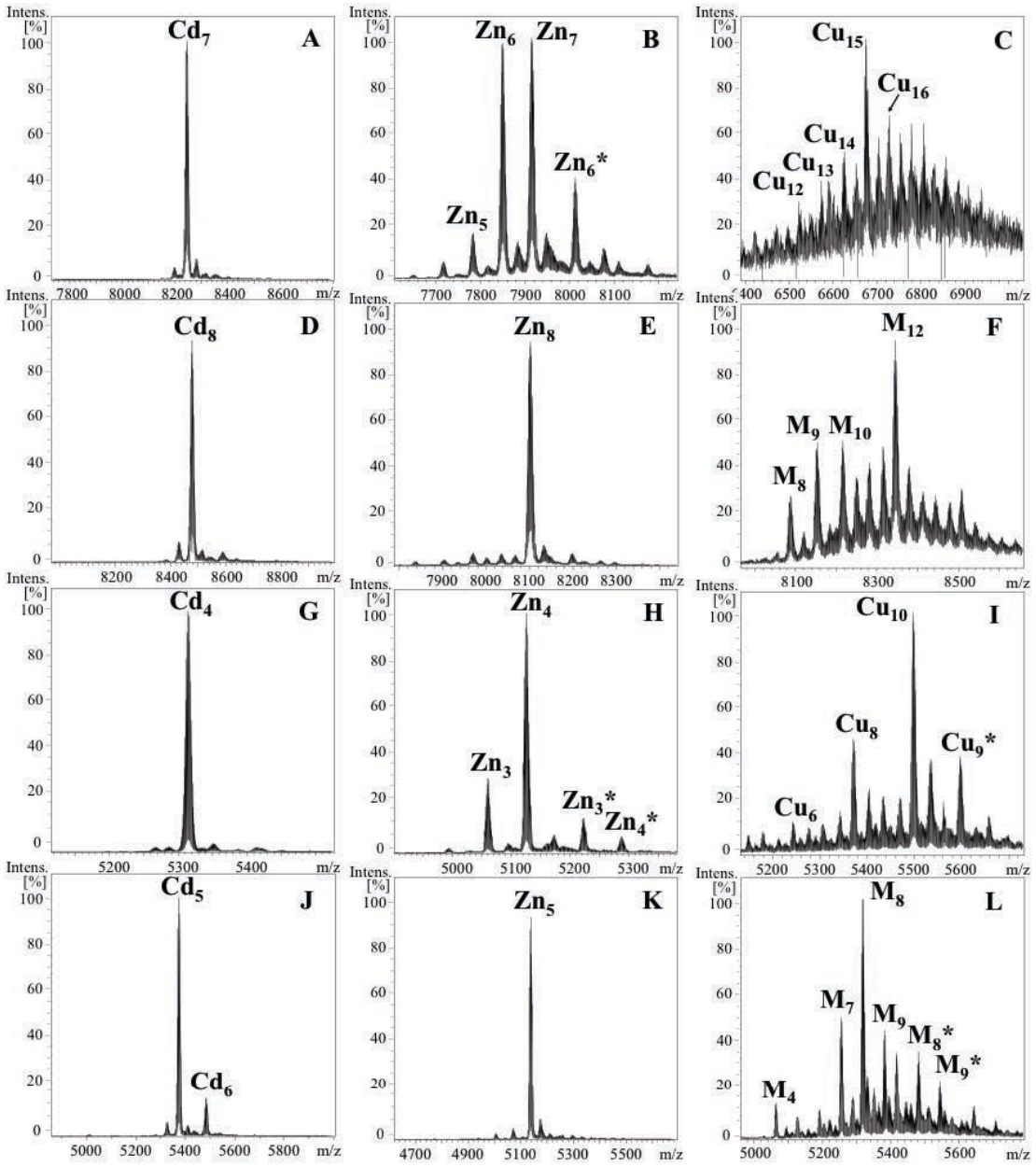


FIGURE 3 Deconvoluted mass spectra measured at pH 7.0 of the products of the recombinant synthesis of the putative MTs NpoMT1 and FcaMT1 (A) Cd-NpoMT1, (B) Zn-NpoMT1 and (C) Cu-NpoMT1, (D) Cd-FcaMT1, (E) Zn-FcaMT1 and (F) Cu-FcaMT1; and the products of the synthesis of the independent NpoMT1 and δFcaMT1 domains (G) Cd-αNpoMT1, (H) Zn-αNpoMT1 and (I) Cu-NpoMT1;(J) Cd-δFcaMT1, (K) Zn-δFcaMT1 and (L) Cu-δFcaMT1. Glycosylated metal-MT species are marked with an asterisk (*).

(GE Healthcare, Chicago, IL, USA) and transformed in protease-deficient *E. coli* BL21 strain. Metal-MT complexes were produced in *E. coli* BL21 cultures expressing the recombinant plasmids, after induction with isopropyl-β-D-thiogalactopyranoside (100 μM) and supplementation with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM) or $CuSO_4$ (500 μM). Metal-MT complexes were purified from the soluble protein fraction of sonicated bacteria by affinity purification of the GST-tagged proteins, and digestion with thrombin. Notice that the digestion with thrombin added two additional residues, Gly and Ser, at the N-terminal end of all purified proteins. These two amino acids do not interfere with the metal-binding features of recombinant MTs (Cols et al., 1997). The metal-MT complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck), and fractionated on a

Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.0 or with fresh 50 mM ammonium acetate, pH 7.0, and run at 0.8 mL min-1. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at -80C until use.

4.2 | Analysis of metal-MT complexes

All designed constructs (NpoMT, FcaMT, αNpoMT and δFcaMT) were characterized by means of mass spectrometry (ESI-MS) and spectroscopy (ICP-AES). An electrospray ionization mass spectrometry (ESI-MS) Micro Tof-Q Instrument (Brucker Daltonics Gmbh, Bremen, Germany) interfaced with a Series 1100 HPLC pump (Agilent Technologies) was used to determine the molecular mass of the recombinant proteins. The instrument was calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies, USA) and the experimental conditions were set up as follows: injection of 10-20 μL of sample through a PEEK long tube (1-1.5m 0.18mm i.d.) at 30-50 μL min-1; capillary-counterelectrode voltage, 3.5-5.0 kV; desolvation temperature, 90–110oC; dry gas, 6 L min-1. Data was acquired over an m/z range of 800 to 3000. The liquid carriers were a 90:10 mixture of 15 mM ammonium acetate and acetonitrile at pH 7.0 and a 95:5 mixture of formic acid and acetonitrile at pH 2.4. Element concentrations of S, Zn, Cd, and Cu in the sample were determined by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) by means of a Perkin-Elmer Optima 4300DV (Waltham, USA) at the correct wavelength (S, 182.04 nm; Zn, 213.86 nm; Cd, 228.80 nm; Cu, 324.80 nm) under conventional conditions (Bongers, et al. 1988). MTs concentration was calculated based on the S concentration obtained by ICP-AES, assuming that all the sulphur measured comes from peptides' Cys and Met residues.

5 | BIBLIOGRAPHY

- Blindauer, C.A., 2014. CHAPTER 21 Metallothioneins, in: Binding, Transport and Storage of Metal Ions in Biological Cells. The Royal Society of Chemistry, pp. 606–665. https://doi.org/10.1039/9781849739979-00606
- Calatayud, S., 2018. Metallothioneins of the urochordate Oikopleura dioica have Cys-rich tandem repeats, large size and cadmium-binding preference 25.
- Calatayud, S., Garcia-Risco, M., Capdevila, M., Cañestro, C., Palacios, Ò., Albalat, R., 2021a. Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins. Front. Cell Dev. Biol. 9, 1733. https://doi.org/10.3389/fcell.2021.702688
- Calatayud, S., Garcia-Risco, M., Palacios, Ò., Capdevila, M., Cañestro, C., Albalat, R., 2021b. Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions and functional convergences. Mol. Biol. Evol. msab184. https://doi.org/10.1093/molbev/msab184
- Calatayud, S., Garcia-Risco, M., Pedrini-Martha, V., Eernisse, D.J., Dallinger, R., Palacios, Ò., Capdevila, M., Albalat, R., 2021c. Modularity in Protein Evolution: Modular Organization and De Novo Domain Evolution in Mollusk Metallothioneins. Mol. Biol. Evol. 38, 424–436. https://doi.org/10.1093/molbev/msaa230
- Capdevila, M., Atrian, S., 2011. Metallothionein protein evolution: a miniassay. JBIC J. Biol. Inorg. Chem. 16, 977–989. https://doi.org/10.1007/s00775-011-0798-3
- Cols, N., Romero-Isart, N., Capdevila, M., Oliva, B., Gonzàlez-Duarte, P., Gonzàlez-Duarte, R., Atrian, S., 1997. Binding of excess cadmium(II) to Cd7-metallothionein from recombinant mouse Zn7-metallothionein 1. UV-VIS absorption and circular dichroism studies and theoretical location approach by surface accessibility analysis. J. Inorg. Biochem. 68, 157–166. https://doi.org/10.1016/S0162-0134(97)00085-8
- Dallinger, R., Zerbe, O., Baumann, C., Egger, B., Capdevila, M., Palacios, Ò., Albalat, R., Calatayud, S., Ladurner, P., Schlick-Steiner, B.C., Steiner, F.M., Pedrini-Martha, V., Lackner, R., Lindner, H., Dvorak, M., Niederwanger, M., Schnegg, R., Atrian, S., 2020. Metallomics reveals a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins. Metallomics 12, 702–720. https://doi.org/10.1039/C9MT00259F

- Espart, A., Marín, M., Gil-Moreno, S., Palacios, Ò., Amaro, F., Martín-González, A., Gutiérrez, J.C., Capdevila, M., Atrian, S., 2015. Hints for Metal-Preference Protein Sequence Determinants: Different Metal Binding Features of the Five Tetrahymena thermophila Metallothioneins. Int. J. Biol. Sci. 11, 456–471. https://doi.org/10.7150/ijbs.11060
- Guirola, M., Pérez-Rafael, S., Capdevila, M., Palacios, Ò., Atrian, S., 2012. Metal Dealing at the Origin of the Chordata Phylum: The Metallothionein System and Metal Overload Response in Amphioxus. PLOS ONE 7, e43299. https://doi.org/10.1371/journal.pone.0043299
- Jenny, M.J., Payton, S.L., Baltzegar, D.A., Lozier, J.D., 2016. Phylogenetic Analysis of Molluscan Metallothioneins: Evolutionary Insight from Crassostrea virginica. J. Mol. Evol. 83, 110–125. https://doi.org/10.1007/s00239-016-9758-4
- Jenny, M.J., Ringwood, A.H., Schey, K., Warr, G.W., Chapman, R.W., 2004. Diversity of metallothioneins in the American oyster, Crassostrea virginica, revealed by transcriptomic and proteomic approaches. Eur. J. Biochem. 271, 1702–1712. https://doi.org/10.1111/j.1432-1033.2004.04071.x
- Nam, Y.K., Kim, E.J., 2017. Diversification and domain evolution of molluskan metallothioneins: a mini review. Fish. Aquat. Sci. 20, 8. https://doi.org/10.1186/s41240-017-0054-z
- Niederwanger, M., Calatayud, S., Zerbe, O., Atrian, S., Albalat, R., Capdevila, M., Palacios, Ò., Dallinger, R., 2017. Biomphalaria glabrata Metallothionein: Lacking Metal Specificity of the Protein and Missing Gene Upregulation Suggest Metal Sequestration by Exchange Instead of through Selective Binding. Int. J. Mol. Sci. 18, 1457. https://doi.org/10.3390/ijms18071457
- Palacios, Ò., Atrian, S., Capdevila, M., 2011. Zn- and Cu-thioneins: a functional classification for metallothioneins? JBIC J. Biol. Inorg. Chem. 16, 991–1009. https://doi.org/10.1007/s00775-011-0827-2
- Palacios, Ò., Jiménez-Martí, E., Niederwanger, M., Gil-Moreno, S., Zerbe, O., Atrian, S., Dallinger, R., Capdevila, M., 2017. Analysis of Metal-Binding Features of the Wild Type and Two Domain-Truncated Mutant Variants of Littorina littorea Metallothionein Reveals Its Cd-Specific Character. Int. J. Mol. Sci. 18, 1452. https://doi.org/10.3390/ijms18071452

ARTICLE III: *Two Unconventional Metallothioneins in the Apple Snail Pomacea bridgesii Have Lost Their Metal Specificity during Adaptation to Freshwater Habitats.*

Int. J. Mol. Sci. 2021, 22(1), 95;

DOI:10.3390/ijms22010095

International Journal of
Molecular Sciences

MDPI

Article

Two Unconventional Metallothioneins in the Apple Snail *Pomacea bridgesii* Have Lost Their Metal Specificity during Adaptation to Freshwater Habitats

Mario García-Risco [1], Sara Calatayud [2], Michael Niederwanger [3], Ricard Albalat [2], Òscar Palacios [1], Mercè Capdevila [1] and Reinhard Dallinger [3,*]

1 Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Cerdanyola del Vallès, Spain; Mario.GarciaRisco@uab.cat (M.G.-R.); oscar.palacios@uab.cat (Ò.P.); merce.capdevila@uab.cat (M.C.)
2 Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, E-08028 Barcelona, Spain; sarak09@hotmail.com (S.C.); ralbalat@ub.edu (R.A.)
3 Institute of Zoology, Center of Molecular Biosciences, University of Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria; michael.niederwanger@uibk.ac.at
* Correspondence: reinhard.dallinger@uibk.ac.at

Abstract: Metallothioneins (MTs) are a diverse group of proteins responsible for the control of metal homeostasis and detoxification. To investigate the impact that environmental conditions might have had on the metal-binding abilities of these proteins, we have characterized the MTs from the apple snail *Pomacea bridgesii*, a gastropod species belonging to the class of Caenogastropoda with an amphibious lifestyle facing diverse situations of metal bioavailability. *P. bridgesii* has two structurally divergent MTs, named PbrMT1 and PbrMT2, that are longer than other gastropod MTs due to the presence of extra sequence motifs and metal-binding domains. We have characterized the Zn(II), Cd(II), and Cu(I) binding abilities of these two MTs after their heterologous expression in *E. coli*. Our results have revealed that despite their structural differences, both MTs share an unspecific metal-binding character, and a great ability to cope with elevated amounts of different metal ions. Our analyses have also revealed slight divergences in their metal-binding features: PbrMT1 shows a more pronounced Zn(II)-thionein character than PbrMT2, while the latter has a stronger Cu(I)-thionein character. The characterization of these two unconventional PbrMTs supports the loss of the metal-binding specificity during the evolution of the MTs of the Ampullariid family, and further suggests an evolutionary link of this loss with the adaptation of these gastropod lineages to metal-poor freshwater habitats.

Keywords: metallothionein; metal-specificity; protein domains; *Pomacea bridgesii*; Mollusca; Gastropoda

check for updates

Citation: García-Risco, M.; Calatayud, S.; Niederwanger, M.; Albalat, R.; Palacios, Ò.; Capdevila, M.; Dallinger, R. Two Unconventional Metallothioneins in the Apple Snail *Pomacea bridgesii* Have Lost Their Metal Specificity during Adaptation to Freshwater Habitats. *Int. J. Mol. Sci.* **2021**, *22*, 95. https://dx.doi.org/10.3390/ijms22010095

Received: 26 November 2020
Accepted: 18 December 2020
Published: 24 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Metallothioneins (MTs) are usually small proteins known for their singularly high content of cysteine residues (Cys), which provide them the capacity for binding essential and non-essential heavy metal ions [1–3]. Since MT samples are natively found in association with metals their functions have been related with biological processes involving metal-binding [4]. Some organisms have achieved metal-specific functions of their MTs by harboring multiple forms, each one with a different metal preference [5,6]. Some gastropods, such as *Cantareus aspersus* (also known as *Helix aspersa* or *Cornu aspersum*) and *Helix pomatia*, for instance, exhibit different highly specific MTs devoted to selectively bind Cd(II) (CdMTs) or Cu(I) (CuMTs) [5,7]. Since the expression of CdMTs is upregulated by Cd^{2+} ions, these MTs have been associated with the physiological control of Cd, specifically with Cd detoxification [8–10], whereas the expression of CuMTs appears to be activated for

controlling Cu homeostasis, which in the case of gastropods is related to the synthesis of the oxygen transporter hemocyanin [5,11].

Interestingly, comparisons of the number and distribution of the Cys motifs (i.e., CxC, CC, and CCC) between MTs have led to conclude that CdMTs and CuMTs of a given group of species derived from duplication events of a common MT ancestor [12], and that changes in the non-coordinating amino acids of the duplicates would have been responsible for the evolution of distinct metal binding preferences [12]. Their frequency and positioning within the sequence, along with their spatial arrangement in the three-dimensional structure determine the metal-selectivity of the protein [7,12–14]. In addition to the CdMT and CuMT isoforms, *C. aspersus* and *H. pomatia* express a third MT isoform with intermediate metallic preferences, yielding mixtures of heterometallic metal-complexes when recombinantly synthesized [5,14] or characterized from native MT preparations [15]. These unspecific MTs seem multipurpose proteins capable of maintaining either the physiological levels of metals in the cell [16], transferring them to other metalloproteins [17], or binding toxic heavy metals in order to avoid their deleterious effects [18]. Paradoxically, the amino acid sequences of the three types of gastropod MTs—CdMT, CuMT and unspecific MT—are extremely similar despite their different functionalities. To overcome the difficulties of classifying these clearly different proteins, two criteria were proposed based on MTs' metal-binding features on the one hand [19,20] and their metal-specific functionality on the other [6,7]. The first criterion is based on metal selectivity of MTs through rendering unique, well-structured, homometallic metal-complexes when binding either divalent (Cd(II) or Zn(II) for Cd/Zn-thioneins) or monovalent metal ions (Cu(I) for Cu-thioneins) [20], and to render a mixture of heterometallic complexes for non-selective MTs somewhere in the middle of this gradual classification [19]. The second criterion is based on physiological evidence indicating a prevalent functionality of the respective MTs in performing metal-specific tasks in favor of the metal ion selectively bound, or in favor of a multi-task performance in case of non-selective MTs [5,7–9].

Intriguingly, some gastropod species possess a full set of MTs with different metal binding preferences (CdMTs, CuMTs and unspecific MTs), whereas others like *Megathura crenulata* [21] and *Biomphalaria glabrata* [22] only express unspecific MTs. The diversity of gastropods, as well as of their MTs, is actually so high that the study of gastropod MTs is becoming more and more a model case in the field of MT research [13,23]. The present study adds to this field novel evidence, being devoted to the characterization of two unconventional MTs found in *P. bridgesii* (PbrMT1 and PbrMT2) [13,23]. This species belongs to the group of freshwater apple snails (family Ampullariidae) of the gastropod class of Caenogastropoda, some of them known for their vast ecological impact as pests [24,25]. These snails are also unusual because they possess both a gill and a lung, situated in separated compartments of their mantle cavity. Owing to this particular adaptation, apple snails have assumed an amphibious lifestyle, by which they live predominantly in freshwater swamps and ponds, being also able to temporarily survive under arid conditions buried in the dried-out mud [26].

In a previous work dealing with evolutionary aspects of gastropod MTs including other freshwater snails of the same family (Ampullariidae), PbrMT1 and PbrMT2 were classified as unselective MTs [13]. Based on preliminary analyses and phylogenetic inferences, this work hypothesized that the MTs of freshwater snails might have lost their metal-selective character during adaptation of gastropod families to freshwater habitats with low Cd background concentrations [13]. However, a detailed comparative characterization of the two PbrMTs providing valuable biochemical information about their metal-binding features was missing. To this end, we have thoroughly analyzed the metal-binding abilities of PbrMT1 and PbrMT2 after their recombinant production in metal-supplemented media, including a further experiment in which initially Zn(II)-loaded MT complexes were replaced by Cd(II) to assess their behavior under in vitro conditions. Our approach is based on convincing evidence that heterologous productions of MTs yield metal-MT complexes structurally and functionally equivalent to those isolated from native sources [7,8,14,27,28].

In the present study, all our preparations were characterized by UV-Vis and circular dichroism spectroscopies and mass spectrometry in order to establish the metal-binding features of the *P. bridgesii* MTs, providing experimental evidence to further confirm or reject the inferences made on the basis of our phylogenetic approach. This information will allow us to correlate the binding abilities of these proteins with the requirements of metal handling of *P. bridgesii*, and to speculate how the evolution of unspecific MTs might have provided an adaptive advantage to this species for tolerating variable and fluctuating metal availabilities, improving its adaptation capacity to an amphibious lifestyle.

2. Results

2.1. Sequence Features of PbrMTs

As already mentioned above, *P. bridgesii* is a freshwater snail belonging to the gastropod class of Caenogastropoda and specifically, to the apple snail family of Ampullariidae. The amino acid sequences of its two MTs (PbrMT1 and PbrMT2) were aligned with other gastropod MTs (Figure 1). The sequences chosen for the alignment comprise an MT from another freshwater apple snail of the same genus, *Pomacea diffusa* (PdiMT1) [23]; two known Cd-selective MTs from other Caenogastropoda such as the marine periwinkle *Littorina littorea* (LliMT) [29,30], and the round-mouthed snail *Pomatias elegans*, a close relative of *L. littorea*, but adapted to terrestrial conditions [31]. *P. diffusa* is a close relative of *P. bridgesii*, and its PdiMT1 shows a high sequence homology to PbrMT1 and the same domain organization as PbrMT1, although the metal selectivity features of PdiMT1 are still unknown. The MTs from *L. littorea* (LliMT) and *P. elegans* (PelMT1) are Cd-selective and comprise three metal-binding domains, one C-terminal β1 and two N-terminal β3 domains [23]. Additionally, shown in the alignment are two Cd-selective (CaCdMT and HpCdMT) and two Cu-selective MTs (CaCuMT and HpCuMT) from the two terrestrial helicid snails *C. aspersus* and *H. pomatia*, respectively [7,14,32] (Figure 1). These snail species belong to the gastropod class of Heterobranchia which constitutes the sister clade to Caenogastropoda. The alignment in Figure 1 reveals the archetypal structure of gastropod MTs consisting of β3/β1 domains, featuring cysteine motifs like $Cx_3Cx_4[CxC]x_{3–5}[CxC]x_3[CxC]x_2C$ in the N-terminal β3 domain, and $[CxC]x_{3–4}[CxC]x_3Cx_{5–6}[CxC]x_3[CxC]$ in the C-terminal β1 domain [23].

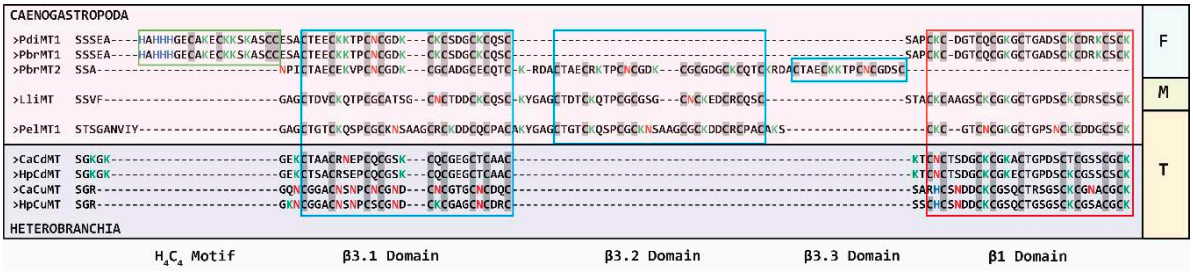


Figure 1. Amino acid alignment of *P. bridgesii* MTs (PbrMT1 and PbrMT2) sequences with other gastropod MTs from species of Caenogastropoda (pink background) and Heterobranchia (purple background) classes (SRA accession number/UniProt accession number): *P. diffusa* PdiMT1 (SRX3488051_2_SRX644696), *L. littorea* LliMT (Q962G0), *P. elegans* PelMT1 (KY646305), *C. aspersus* CaCdMT (A1YZ80; a.k.a. HasMT2) and CaCuMT (A2I9Y4; a.k.a. HasMT1) and *H. pomatia* HpCdMT (P33187; a.k.a. HpoMT2) and HpCuMT (P55947; a.k.a. HpoMT1). β3-domains are boxed in blue, β1-domain in red, and the H_4C_4 motif in green. Cysteines are highlighted with a grey background, histidines are colored in blue, lysines in green, and asparagines in red. Colored boxes at the right side of the alignment indicate the habitats of the selected species: in light blue for freshwater (F) habitats, in light green for marine (M) habitats, and in light brown for terrestrial (T) habitats.

Both PbrMTs show remarkable deviations from the archetypal β3/β1 structure of most other gastropod MTs (Figure 1). PbrMT1 contains additional cysteine (C) and histidine (H) residues in a H_4C_4 motif ($HxHHHx_2Cx_3Cx_6CC$) in its N-terminal region, also shared by PdiMT1 of *P. diffusa* from the same genus, suggesting that the H_4C_4 motif may represent a

lineage-specific sequence divergence in some MTs of closely related ampullariid freshwater snails. PbrMT2 possesses two tandem duplications (one full-length and one partial) of the β3 domain, but it has lost its β1 domain. Therefore, the domain structure of PbrMT2 is β3.1/β3.2/β3.3 (Figure 1). PbrMT1 and PbrMT2 contain, therefore, an extra number of Cys with respect to the archetypal gastropod MTs, probably conferring to them the capability of binding additional metal ions.

Sequence analysis also revealed that the number of lysine residues (K) is higher in PbrMT1 than in PbrMT2 (14 vs. 9), which in turn contains more asparagine residues (N) compared to PbrMT1 (4 vs. 1). These differences may be significant because previous studies had suggested a connection between the K:N ratio of gastropod MTs and their kind of metal preference [12,22]. Hence, snail MTs with a preponderance of K over N residues would have a binding preference for divalent metals such as Zn(II) and Cd(II), whereas MTs with a predominance of N over K would exhibit a Cu-thionein character [12,19,22].

2.2. Production of Metal–PbrMT Complexes

To explore the metal selectivity features of the two *P. bridgesii* MTs, we studied the formation of metal–PbrMT1 and metal–PbrMT2 complexes by the proteins heterologously expressed in *E. coli* and grown in media supplemented with Cu(II), Cd(II) or Zn(II) salts. In order to confirm the identity of the proteins recombinantly produced, apo-PbrMT1 and apo-PbrMT2 forms were obtained from the respective Zn-PbrMT productions under acidic conditions (pH 2.4), and analyzed by ESI-MS. The experimental masses of PbrMT1 and PbrMT2 (9235.5 Da and 8973.3 Da) were coincident with the theoretical masses predicted (9235.51 Da and 8975.20 Da) (Figure 2), ascertaining that the major proportions of the produced proteins corresponded in their masses with the two respective PbrMTs. Here, it should be noticed that recombinant proteins contain two additional amino acids (GS) at their N-term, and therefore the theoretical masses correspond to full length PbrMT1 and PbrMT2 plus those two amino acids. Moreover, in the Zn-supplemented productions, it has to be noted that an additional minor species with an extra mass of 162 Da was detected for PbrMT2. This minor species, named Apo'-PbrMT2 (Figure 2B), did not disappear when acidifying to pH 1, and it was not found neither in PbrMT1 nor in the PbrMT2 productions in media supplemented with Cd(II) or with Cu(II). Preliminary studies suggest that Apo'-PbrMT2 is a minor form artefactually produced only in the PbrMT2 productions with Zn(II) ions, and thereby not affecting the overall characterization of the metal-MTs complexes.

2.3. Zn(II) Binding Abilities of the PbrMT Isoforms

Both PbrMTs rendered mixtures of several species when synthesized in culture media enriched with $ZnCl_2$ (Table 1; Figure 3). In particular, these mixtures consisted of Zn_8- (major) and Zn_7- complexes for PbrMT1, and Zn_7- (major), Zn_8- and Zn_6- complexes for PbrMT2. Concordantly, these results match those obtained by ICP-AES (Table 1), and although PbrMT2 has one Cys more than PbrMT1 (23 vs. 22 Cys), both MTs rendered similar metallated species. However, and despite the degree of heterogeneity in speciation of the Zn-PbrMT1, preparation is lower than that of Zn-PbrMT2, the latter shows a higher folding degree with an exciton coupling at ca. 240 nm in the CD spectra in front of the simple Gaussian band at the same wavelength recorded for the Zn-PbrMT1 preparation (Figure 3C).

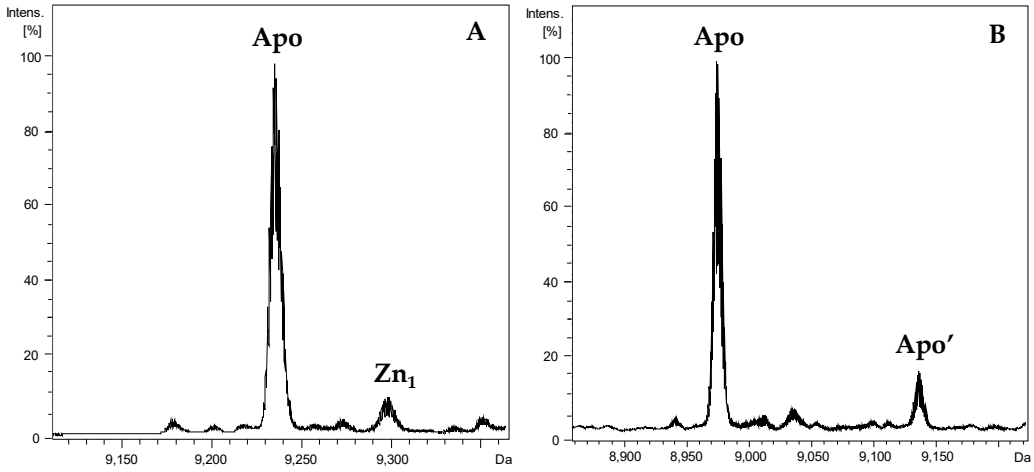


Figure 2. Deconvoluted ESI-MS spectra recorded at acidic conditions (pH 2.4) of (**A**) apo-PbrMT1 and (**B**) apo-PbrMT2 from the Zn-MT productions.

Table 1. Metal-to-protein ratios found by ICP-AES in the recombinantly produced Zn(II)-, Cd(II)- and Cu(I)-PbrMT1 and PbrMT2 preparations. ESI-MS data were recorded at neutral pH and are shown in several figures along this work.

MT	Supplemented Metal	M/Protein [a] (Conventional)	M/Protein [b] (Acidic)	ESI-MS Results [c,d]
PbrMT1 (22 Cys)	Zn	7.4 Zn	n.m.	Zn_8-(major) > Zn_7-
	Cd	5.3 Cd	8.6 Cd	Cd_8- ~ Cd_7S_2-(major) > Cd_7-
	Cu	2.3 Zn;12.4 Cu	n.m.	M_8- to M_{13}-
PbrMT2 (23 Cys)	Zn	6.4 Zn	7.1 Zn	Zn_7-(major) > Zn_8- > Zn_6- > Zn_6-* > Zn_7 *
	Cd	6.9 Cd	9.0 Cd	d-Cd_{18}S- to d-Cd_{22}S- Cd_8-(major) > Cd_9-
	Cu	16.4 Cu	n.m.	Cu_9- to Cu_{17}-

[a] Metal-to-protein ratios obtained from conventional ICP. Zn, Cd and Cu were measured in all cases but only those results different than zero are shown. [b] Metal-to-protein ratio obtained from acidic ICP. Zn, Cd, and Cu were measured in all cases but only those results different than zero are shown. n.m. stands for "not measured". [c] d stands for dimeric species. [d] species with * correspond to metal complexes formed by apo'-PbrMT2.

2.4. Cd(II) Binding Abilities of the PbrMT Isoforms

The recombinant synthesis of PbrMT1 and PbrMT2 in cultures supplemented with $CdCl_2$ also rendered a mixture of species (Figure 4), as in the case of the Zn productions. The difference between the conventional ICP values and the acidic ones in both productions (Table 1), as well as the presence of CD bands with maxima at wavelengths higher than 255 nm (Figure 4D) clearly indicate the presence of additional Cd-S^{2-} chromophores [3,33]. On the one side, PbrMT1 rendered Cd_8- and Cd_7S_2- as major species, together with minor Cd_7-PbrMT1 (Figure 4A). On the other hand, PbrMT2 yielded dimeric species ranging from d-Cd_{18}S- to d-Cd_{22}S-PbrMT2 (Figure 4B), together with a mixture of monomeric Cd_8- and Cd_9-PbrMT2 (Figure 4C) species.

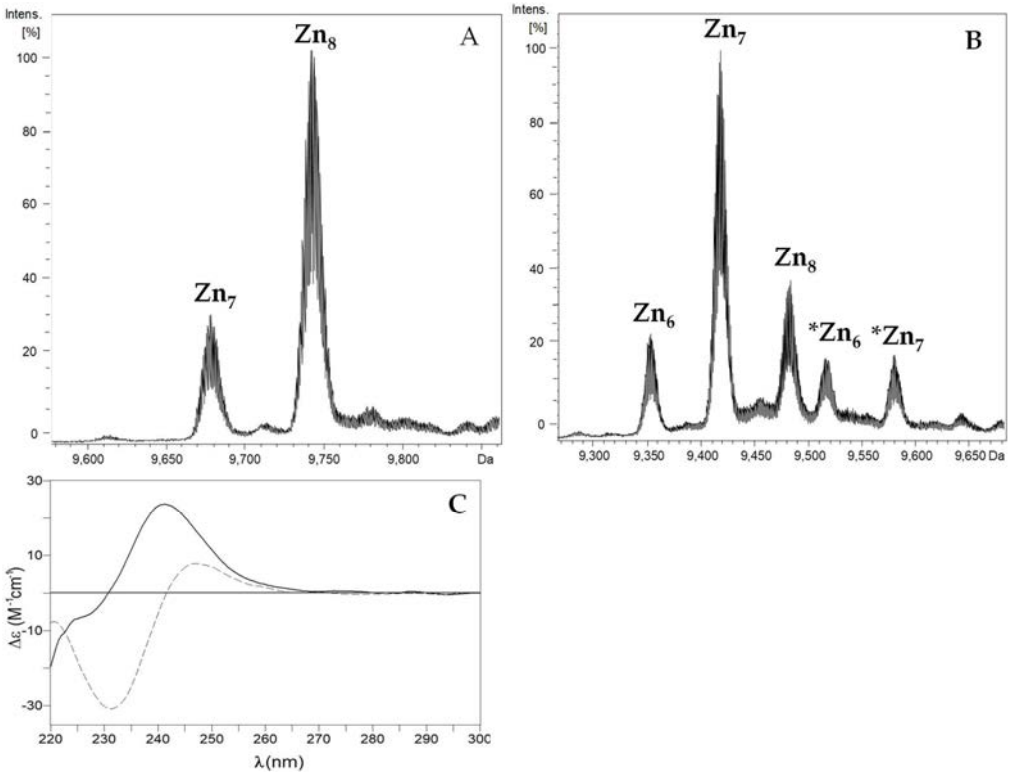


Figure 3. Deconvoluted ESI-MS spectra of the recombinant productions of (**A**) Zn-PbrMT1 and (**B**) Zn-PbrMT2 at neutral pH. (**C**) CD spectra of Zn-PbrMT1 (solid) and Zn-PbrMT2 (dashed) productions. Species with * correspond to metal complexes formed by apo'-PbrMT2.

These results already point out that these two MTs are clearly non-Cd-selective and at the same time, barely reveal any differences in their behavior towards Cd(II). To further investigate the Cd(II) binding capabilities of these two MTs, their in vitro Cd(II)-complexes were obtained by adding Cd(II) aliquots to both Zn-PbrMT preparations. As shown in Figure S1 and Table 2, Zn-PbrMT1 shows a gradual non-cooperative replacement of Zn(II) by Cd(II). In terms of CD absorptions, the first 6 Cd(II) equivalents evoked the same effect in the protein folding, leading to the replacement of the initial Gaussian band at ca. 240 nm to a positive CD absorption centered at ca. 265 nm, very similar to the signal recorded for the recombinant Cd-PbrMT1 preparation (Figure S1E). Further additions of Cd(II) only led to small variations in the CD and UV-vis envelopes (Figure S1), suggesting that after 6–7 Cd(II) equivalents added only slight modifications in the structuration of the Cd-PbrMT1 species take place without the incorporation of further Cd(II) ions. The ESI-MS data of this experiment (Figure S2; Table 2) show a subsequent Zn/Cd replacement from the initial preparation, containing Zn_8- and Zn_7-PbrMT1 species, to mainly major Zn_1Cd_7- and Cd_7-PbrMT1 after the addition of 7 Cd(II) which remain unaltered in solution for 10 Cd(II) added, without reaching the Cd_8- major species achieved in the recombinant preparation.

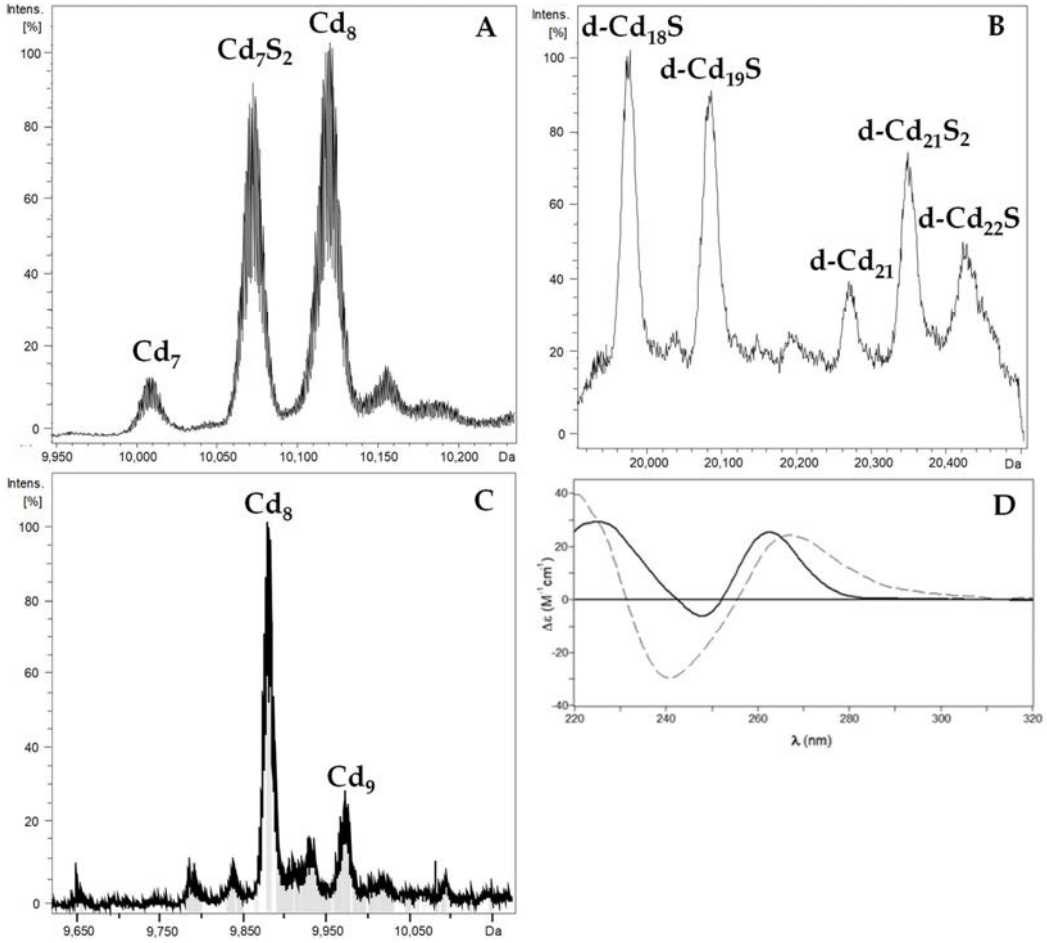


Figure 4. In vivo Cd(II) binding assessment of PbrMT1 and PbrMT2. Deconvoluted ESI-MS spectra of (**A**) Cd-PbrMT1 and (**B**,**C**) Cd-PbrMT2 productions. Dimeric species in the Cd-PbrMT2 production spectrum are marked as "d-". (**D**) CD spectra of Cd-PbrMT1 (solid) and of Cd-PbrMT2 (dashed) productions.

Table 2. Species detected by ESI-MS during the addition of Cd(II) to the Zn-PbrMT1 preparation at pH 7.0. The corresponding ESI-MS spectra are shown in Figure S2.

Zn + Cd		Cd(II) Equivalents					
		0	**1**	**3**	**5**	**7**	**10**
7	Zn_7-PbrMT1	•	✖				
8	Zn_8-PbrMT1	✔	•				
8	Zn_7Cd_1-PbrMT1		✔				
7	Zn_6Cd_1-PbrMT1		•				
8	Zn_6Cd_2-PbrMT1		•	✖			
8	Zn_5Cd_3-PbrMT1		✖	✔			
8	Zn_4Cd_4-PbrMT1			✔			
8	Zn_3Cd_5-PbrMT1			•	✔		
8	Zn_2Cd_6-PbrMT1			✖	✔		
7	Zn_1Cd_6-PbrMT1				•	✖	✖
8	Zn_1Cd_7-PbrMT1				•	✔	✔
7	Cd_7-PbrMT1					✖	•

✔ denotes major species, • denotes intermediate species, ✖ denotes minor species. The terms "major", "intermediate" and "minor" species refer to their signal intensity in the corresponding MS spectra.

The Zn/Cd displacement carried out on the Zn-PbrMT2 preparation (Figure S3; Table 3) shows a different pattern from that observed in Zn-PbrMT1, as expected if considering their different amino acid composition (Figure 1) and the speciation of their respective Zn(II) preparations (Figure 3). The addition of Cd(II) to Zn-PbrMT2 provoked a complex progression from an initial CD spectrum showing the characteristic exciton coupling band at ca. 240 nm of certain Zn–MT complexes, to another exciton coupling at ca. 245 nm after 10 Cd(II) equivalents added. This final stage, as expected, is different from that recorded for the recombinant Cd-PbrMT2 sample (Figure S3F) due to the content of labile S^{2-} ions of the latter which contribute to the signal in the 280–300 nm range. UV-vis differences between additions at the latter stages of the experiment showed barely any differences between chromophores, suggesting that the protein did not accept extra metal ions (Figure S3D,E).

Table 3. Species detected by ESI-MS during the addition of Cd(II) to the Zn-PbrMT2 preparation at pH 7.0. The corresponding ESI-MS spectra are shown in Figure S4.

Zn + Cd		Cd(II) Equivalents					
		0	2	4	6	8	10
7	Zn_7-PbrMT2	✔					
8	Zn_8-PbrMT2	•					
8	Zn_6*-PbrMT2	•					
9	Zn_7*-PbrMT2	✖	•				
8	Zn_5Cd_3-PbrMT2		✔				
7	Zn_4Cd_3-PbrMT2		•				
8	Zn_4Cd_4-PbrMT2		•	•			
8	Zn_3Cd_5-PbrMT2		✖	✔	✖		
8	Zn_2Cd_6-PbrMT2			•	•		
8	Zn_1Cd_7-PbrMT2			•	✔	•	✖
9	Zn_1Cd_8-PbrMT2				•	•	•
8	Cd_8-PbrMT2				✔	✔	✔
10	Zn_1Cd_9-PbrMT2					✖	•
9	Cd_9-PbrMT2				✖	✖	•
10	Cd_{10}-PbrMT2						•

✔ denotes major species, • denotes intermediate species, ✖ denotes minor species. The terms "major", "intermediate" and "minor" species refer to their signal intensity in the corresponding MS spectra.

It is remarkable that, at after 6 Cd(II) equivalents added, PbrMT2 renders not only the same Zn_1Cd_7- species as PbrMT1 but also Cd_8-PbrMT2. This last species is maintained as one of the important peaks during the whole metal-displacement experiment, as can be seen in the ESI-MS (Table 3; Figure S4). Additionally, unlike PbrMT1, PbrMT2 reaches, at the latter stages, high metallated homonuclear species such as Cd_9- and Cd_{10}-PbrMT2. These results are in agreement with those from the recombinant productions in E. coli and suggest that, although none of the PbrMTs are specifically adapted to capture Cd(II), PbrMT2 can bind more Cd(II) than PbrMT1, in accordance with the former's higher content in Cys residues.

2.5. Cu(I) Binding Abilities of the PbrMTs

The recombinant synthesis of both PbrMT peptides in Cu-supplemented media rendered always a mixture of species (Figure 5), but important differences among both proteins were observed. The ICP-AES results of both Cu-supplemented PbrMT preparations (Table 1) show that, while PbrMT1 renders heterometallic Zn,Cu-species, PbrMT2 yields homometallic Cu(I) species. Moreover, a higher Cu content was detected for PbrMT2 (16.4 Cu), as it was expected if considering that this MT has one more Cys residue than PbrMT1 (12.4 Cu and 2.3 Zn). Consistent with these results, ESI-MS spectra reveal the presence of various metallated species, ranging from M_9- to M_{13}-PbrMT1 (Figure 5A), while for PbrMT2 an intense peak of Cu_{15}-PbrMT2 was detected above the minor species ranging from Cu_{11}- to Cu_{17}-PbrMT2 (Figure 5B). Despite the speciation differences observed, their

CD spectra of the two MT preparations are quite similar (Figure 5C), depicting very faint absorption bands that denote a low degree of protein folding about the metal ions.

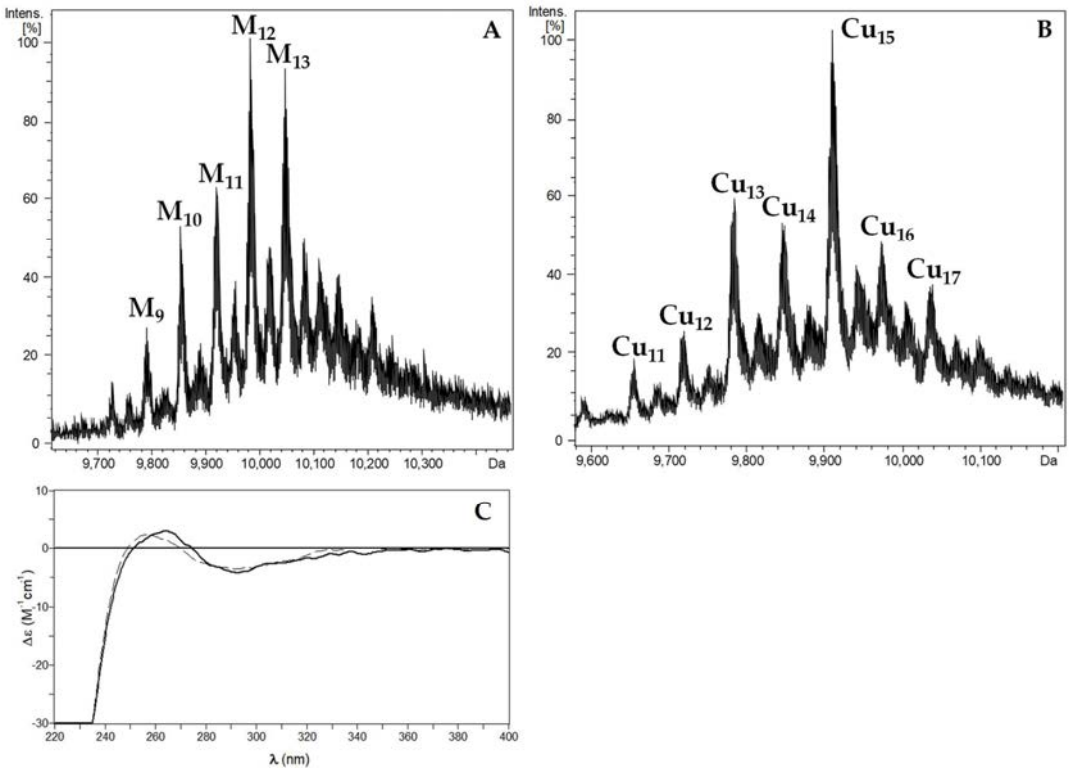


Figure 5. In vivo Cu(I) binding assessment of recombinant PbrMT1 and PbrMT2. Deconvoluted ESI-MS spectra of (**A**) Cu-PbrMT1 and (**B**) Cu-PbrMT2 productions. (**C**) CD spectra of PbrMT1 (solid) and of PbrMT2 (dashed) productions.

3. Discussion

Sequence analyses reveal unconventional primary structures for both multi-modular MTs of *P. bridgesii*, very divergent from that of the archetypal β_3/β_1 MT form of other gastropods (Figure 1). On one hand, PbrMT1 has an additional domain with a H_4C_4 motif in its N-terminal region, which has been also found in other gastropod MTs [23], all of them belonging to the freshwater family of Ampullariidae (informal group of Architaenioglossa) form the gastropod clade of Caenogastropoda. It is assumed that a lineage-specific event may have added the H_4C_4 motif to the N-terminus of an ancestral Architaenioglossan MT duplicate. This motif increases the number of potential metal-coordinating amino acids (histidines and cysteines), and thereby it could also increase its binding capacity for metal ions. In fact, the recombinant Zn(II)- and Cd(II)-PbrMT productions always rendered mixtures of mainly M_7- and M_8-PbrMT species (M = Zn or Cd), which represents up to 33% of increased capacity with respect to the M_6-MT species rendered by archetypal gastropod β_3/β_1 MTs [23]. However, metal binding sites including His residues are known for their preference for Zn over Cd, so this H_4C_4 motif may intervene in metal-selectivity features of PbrMT1 [33–35]. Overall, we found this peculiar motif extremely interesting, and for that reason, the biochemical properties of this H_4C_4 motif in ampullariid MTs and its possible role in modifying metal selectivity features of PbrMT1 will be widely clarified and discussed in a separate study.

On the other hand, structural evolution has led PbrMT2 to lose the C-terminal β1 domain and to tandem-duplicate the β3 domain. These modifications increase the overall number of coordinating cysteines from 18 in the conventional gastropod β_3/β_1 MTs [23] to

23, implying a higher number of divalent metal ions that these MTs may be able to bind compared to other gastropod MTs [13]. Our results clearly revealed an increased metal-binding capacity for PbrMT2, ranging from the M_6-MT species rendered by archetypal gastropod MTs to M_8-PbrMT2 complexes. Other known mollusk MTs with similarly high Cys contents, such as *Mytilus galloprovincialis* MT-10 (21 Cys) and MT-20 (23 Cys), were recovered binding 7 Cd(II) ions [36,37]. Surprisingly, the presence of 22–23 Cys residues in the *P. bridgesii* metallothioneins provoked the binding of one more M(II) ion, in comparison with the mentioned MTs of *Mytilus*, to render M_8-species. Moreover, the presence of labile sulfide ligands in the Cd-preparations of both PbrMTs resulted in the absence of a clearly defined M(II)-thionein character (M = Zn, Cd) for these MTs. Interestingly, the analysis of the CD spectra revealed that the Zn-PbrMT2 (major Zn_7-species, heterogenous speciation) preparation shows an exciton coupling centered at 240 nm which contrasts with the Gaussian band at the same wavelength recorded for the Zn-PbrMT1 (major Zn_8- and minor Zn_7-complexes) production. This probably points out to a formation of more compact clusters with 7 Zn(II) ions in PbrMT2 than with 8 Zn(II) ions in PbrMT1. Overall, our results support that the multi-modular MTs of *P. bridgesii* have significantly diverged from the archetypal β_3/β_1 gastropod MT variants not only with respect to their primary structure, but also by acquiring a higher metal-binding capacity than the standard gastropod MTs.

Complementary to the data from recombinant productions of metal–MT complexes, the Zn-by-Cd in vitro replacement carried out in both Zn-preparations proceeded similarly. The initial Zn-species led, by a gradual non-cooperative replacement, to final Cd-species with similar CD profiles as those observed in the recombinant Cd-productions. Probably, the absence of the additional S^{2-} ligands in the Zn(II)-preparations observed in the recombinant Cd(II) samples could be the main reason for the recording of different CD signals at the end of both Cd(II) titrations. These Zn/Cd replacement reaction in both proteins revealed that PbrMT1 possesses a higher Zn(II)-thionein character than PbrMT2, and that PbrMT2 exhibits a higher Cd(II) than Zn(II) specificity compared to PbrMT1. While Zn-PbrMT1 productions are reluctant to fully exchange the initial Zn(II), even after the addition of 10 Cd(II) equivalents, Zn-PbrMT2 preparations easily release their Zn(II) metal ions at the early stages of the displacement experiment and render highly metallated homometallic Cd(II)-species at the late stages. With regards to Cu(I) binding, the most remarkable difference between the two PbrMTs has been the significant presence of Zn(II) in the Cu(I)-PbrMT1 production (Table 1), which reinforces the fact that PbrMT1 shows a stronger Zn-thionein character than PbrMT2.

Overall, our analyses have revealed that the two PbrMTs do not exhibit strong metal specificities for any of the tested metals, probably because their overall primary structure and special Cys pattern seem to make them incapable of forming unique homometallic Zn- Cd- or Cu-metallated species. We have therefore experimentally demonstrated that both proteins are unspecific, but with similar although not identical biochemical features. PbrMT1 exhibits a more pronounced Zn(II)-thionein character, while PbrMT2 shows a more explicit Cu(I)-thionein character, which questions the possibility to predict the metal preference of gastropod MTs solely based on sequence features such as the K:N ratios. Our results also address an important aspect in the evolution of metal preferences of gastropod MTs. It has been hypothesized that the ancestral MT of Caenogastropoda might have been Cd-selective [13,23]. Our results imply the loss of the metal-binding specificity during the evolutionary adaptation to freshwater habitats in the lineage of *P. bridgesii* of the Ampullariidae family of Caenogastropoda. It can be argued that due to the low cadmium background levels in the freshwater habitats in which apple snails live [13], the selective pressure for maintaining Cd-selective MTs was no longer effective, leading therefore to the evolution of unspecific MTs. Our results, together with previous studies performed with living snails [38], suggest that the evolution of the two PbrMTs might reflect the needs of *P. bridgesii* to adapt to environmental conditions of metal-poor freshwater habitats, paired with fluctuating availabilities of metal ions due to its amphibious lifestyle. The results in

P. bridgesii remind some of our previous findings in *B. glabrata*, a freshwater air-breathing snail of the clade of Hygrophila that belongs to the gastropod class of Heterobranchia, the sister class to Caenogastropoda. The MT of *B. glabrata*, too, has lost its Cd selectivity during adaptation of ancestral Hygrophila to freshwater habitats [13,22]. Apparently, this has also happened to PbrMT1 and PbrMT2 of *P. bridgesii*.

4. Materials and Methods

4.1. Cloning Pomacea bridgesii Metallothioneins for Heterologous Expression

Following an already established methodology in our groups [7,39], cDNAs of PbrMT1 and PbrMT2 (GenBank No. KY963504.1 and KY963505 [13]) were synthesized by Integrated DNA Technologies Company (Coralville, IA, USA). *Bam*HI and *Xho*I restriction sites and 6 or 7 additional 5′-nucleotides were added to both cDNA ends to facilitate the cloning processes. Each synthetic cDNA was PCR-amplified in a 25 µL PCR mixture using Expand High Fidelity PCR system (Roche, Penzberg, Upper Bavaria, Germany) with a common forward primer (5′-TTTTATTGGATCCATGTCTTC-3′ for both *PbrMT1* and *PbrMT2*) and a reverse primer specific for each MT (5′-ATTTTTCTCGAGTCACTTGC-3′ for *PbrMT1*, and 5′-ATTTTTCTCGAGTCAGCAACTG-3′ for *PbrMT2*). PCR conditions for both MT genes were as follows: an initial denaturation step at 94 °C for 5 min was followed by 25 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and a final extension step at 72 °C for 7 min. PCR bands with the expected size were cut and purified with the GenElute™ Gel Extraction Kit (Sigma-Aldrich, St. Louis, MO, USA) following the manufacturer's instructions. PCR products were digested by *Bam*HI and *Xho*I enzymes, cloned into *Bam*HI-*Xho*I digested pGEX-4T-1 expression vector (GE Healthcare, Chicago, IL, USA) with the DNA Ligation kit 2.1 (Takara Bio Inc., Shimogyo-ku, Kyoto, Japan), and transformed into the *E. coli* DH5α strain.

Plasmid DNA was purified from bacteria cultures using the GeneElute™ Plasmid Miniprep Kit (Sigma-Aldrich), screened for insert presence by digestion with *Sca*I enzyme, and sequenced at the Scientific and Technological Centers of the University of Barcelona, using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Waltham, MA, USA) in an automatic sequencer (ABIPRISM 310, Applied Biosystems). DNA from each recombinant *PbrMT*-pGEX plasmid was used to transform *E. coli* BL21 strain, a protease-deficient strain used for heterologous protein expression.

4.2. Production and Purification of Recombinant Metal–PbrMT Complexes

For production of recombinant metal–PbrMT complexes, 500 mL of LB medium with 100 µg/mL ampicillin were inoculated with *E. coli* BL21 cells transformed with the *PbrMT1*-pGEX or *PbrMT2*-pGEX recombinant plasmids. After overnight growth at 37 °C/250 rpm, bacterial cultures were used to inoculate 5 L of fresh LB-100 µg/mL ampicillin medium. MT expression was induced with 100 µM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h at 37 °C/250 rpm. After the first 30 min of induction, cultures were supplemented with $ZnCl_2$ (300 µM), $CdCl_2$ (300 µM), or $CuSO_4$ (500 µM) to generate metal-PbrMT complexes. Cells were harvested by centrifugation for 5 min at 9100× *g* (7700 rpm), and bacterial pellets were suspended in 125 mL of ice-cold PBS (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4) with 0.5% *v/v* β-mercaptoethanol. Suspended cells were sonicated (Sonifier® ultrasonic cell disruptor) 8 min at voltage 6 with pulses of 0.6 s, and then centrifuged for 40 min at 17,200× *g* (12,000 rpm) at 4 °C.

Protein extracts containing GST-PbrMT1 or GST-PbrMT2 fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. After centrifugation at 500× *g* (1553 rpm) for 5 min, pellets of GST-PbrMT fusion proteins bound to the sepharose beads were washed by resuspending them in 30 mL of cold 1xPBS (20 mL for 3 L cultures) bubbled with argon to prevent oxidation. After three washes, GST-MT fusion proteins were digested with thrombin (GE Healthcare, 25 U/L of culture) overnight at 17 °C, enabling separation of the metal–PbrMT complexes from the GST that remained bound to the sepharose matrix. The eluted metal–PbrMT complexes

were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Millipore, MA, USA), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.0. Protein-containing fractions were identified by their absorbance at 254 nm, and pooled and stored at −80 °C until use.

4.3. Characterization of the Metal–PbrMT1 and Metal–PbrMT2 Complexes

The different metal-MT recombinant preparations were characterized by means of Inductively Coupled Plasma Atomic Emission Spectrometer (ICP-AES), performed in an Optima 4300DV (Perkin-Elmer, Waltham, MA, USA) spectrometer by measuring S at 182.04 nm, Zn at 213.85 nm, Cd at 228.80 nm, and Cu at 324.75 nm. The protein concentration was determined assuming that the total content of S in the proteins comes from both cysteine and methionine amino acids [40]. However, some recombinant productions may render species that include labile S^{-2} anions as a third component on the metal–MT complexes [41]. Therefore, two methodologies were performed to detect these labile sulfide anions. The so-called "conventional ICP" [40] that refers to the standard methodology used to measure the samples employing 1% HNO_3 solution as a matrix and the "acid ICP" [41] methodology that stands for those samples measured after incubation in 1 M HNO_3 at 65 °C for 10 min to eliminate labile sulfide anions as H_2S. Thus, the differences in S content measured via "conventional" and "acid" treatment allowed to detect the presence of labile sulfide anions in the samples. Moreover, the global metal to protein ratios were measured.

The molecular weight of Zn-, Cd- and Cu-MT complexes formed, as well as that of the corresponding apo proteins was determined by Electrospray Ionization Time-of-Flight Mass Spectrometry (ESI-TOF MS) on a Micro TOF-Q instrument (Bruker Daltonics, Bremen, Germany) connected to a Series 1200 HPLC Agilent pump and controlled by the Compass Software. Samples were analyzed under neutral (pH 7.0) and/or acidic (pH 2.4) conditions, using as running buffer a 5:95 mixture of acetonitrile:ammonium acetate (15 mM) and a 5:95 mixture of acetonitrile:formic acid solution, respectively. In acidic conditions, Zn(II) and Cd(II) are released while Cu(I) is kept in the protein complex. Therefore, the characterization of the apo proteins was performed at acidic conditions on the Zn- or Cd-loaded forms. For each analysis, 20 µL of protein solution (of concentrations ranging between 20 and 129 µM) were injected at 30–40 $\mu L \cdot min^{-1}$ and analyzed under the following conditions: capillary counter-electrode voltage, 3500–5000 V; dry temperature, 90–110 °C; dry gas, 6 $L \cdot min^{-1}$; m/z range 800–3000. Experimental masses were calculated as previously described [42].

Circular Dichroism (CD) measurements were performed in a Jasco spectropolarimeter (Model J-715; Jasco Inc., Easton, MD, USA) interfaced to a computer (J700 software). Electronic absorption was carried out by means of an HP-8453 Diode array UV-Visible spectrophotometer (Hewlett-Packard, Palo Alto, CA, USA). The resultant spectra from both techniques were processed with GRAMS 32 Software (GRAMS/AI v.7.02; Thermo Scientific, Walthman, MA, USA).

4.4. In Vitro Metal–Protein Binding Studies

Solutions of the Zn-PbrMT1 (6.06 µM) and Zn-PbrMT2 (6.84 µM) preparations were titrated with Cd(II) at pH 7.0 as previously described [27,43], using a $CdCl_2$ standard solution of 1 mM concentration. Essentially, a molar equivalent of metal was added by a stepwise manner to the protein solution. UV-Vis and CD spectra were recorded after each equivalent addition and once the spectrum, and hence the protein folding was stabilized. The procedure continued till metal saturation was reached. All the experiments were performed under strict oxygen-free conditions using argon to saturate the solutions. Aliquots of relevant equivalent additions were taken to execute MS analysis to draw, step by step, the evolution of each MT species throughout the displacement reaction.

Supplementary Materials: Supplementary materials are available online at https://www.mdpi.com/1422-0067/22/1/95/s1.

Author Contributions: Conceptualization, R.D., M.C. and Ò.P.; methodology, R.A., M.C., Ò.P. and R.D.; software, S.C. and M.G.-R.; validation, R.D., M.C., Ò.P. and R.A.; formal analysis, S.C., M.N. and M.G.-R.; investigation, S.C., M.G.-R. and M.N.; resources, R.D., M.C. and R.A.; data curation, S.C., M.G.-R. and M.N.; writing—original draft preparation, M.C.; writing—review and editing, M.C., Ò.P., R.D., S.C., M.N. and M.G.-R.; visualization, S.C., M.G.-R., M.C. and R.D.; supervision, R.D., M.C., Ò.P. and R.A.; project administration, R.D., M.C. and R.A.; funding acquisition, R.D., M.C., Ò.P. and R.A. All authors have read and agreed to the published version of the manuscript.

Funding: Parts of this study were co-financed by a cooperation grant to R.D. from the Austrian Science Fund (DACH grant No I 1482-N28). Researchers from Barcelona want to acknowledge the Spanish Ministerio de Ciencia e Innovación and FEDER for the projects BIO2015-67358-C2-2-P (M.C. and Ò.P.) and BIO2015-67358-C2-1-P (R.A.). M.C. and Ò.P. are members of the "Grup de Recerca de la Generalitat de Catalunya", ref. 2017SGR-864, and R.A. of ref. 2017SGR-1665. M.G.-R. acknowledges to the UAB the PIF grant. We also thank the Servei d'Anàlisi Química (SAQ) at the Universitat Autònoma de Barcelona (ICP-AES, ESI-MS) for allocating instrument time, and the Centres Científics i Tecnològics (CCiT) at the Universitat de Barcelona for DNA sequencing.

Institutional Review Board Statement: Not applicable, no living animals involved.

Informed Consent Statement: Not applicable, no humans involved.

Data Availability Statement: Additional data for this work can be found in the Supplementary material. For access to raw data, please contact the corresponding author.

Acknowledgments: Open Access Funding by the Austrian Science Fund (FWF).

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Kagi, J.H.R.; Vasák, M.; Lerch, K.; Gilg, D.; Hunziker, P.; Bernhard, W.; Good, M. Structure of mammalian metallothionein. *Environ. Health Perspect.* **1984**, *54*, 93–103. [PubMed]
2. Kägi, J.H.R.; Vallee, B.L. Metallothionein: A cadmium and zinc-containing protein from equine renal cortex. *J. Biol. Chem.* **1960**, *235*, 3460–3465. [PubMed]
3. Capdevila, M.; Bofill, R.; Palacios, O.; Atrian, S. State-of-the-art of metallothioneins at the beginning of the 21st century. *Coord. Chem. Rev.* **2012**, *256*, 46–62.
4. Vašák, M.; Bogumil, R.; Faller, P.; Hasler, D.W.; Binz, P.-A.; Klauser, S.; Bruinink, A.; Charnock, J.M.; Garner, C.D. Structural and biological studies on native bovine Cu,Zn-metallothionein-3. In *Metallothionein IV: Advances in Life Sciences*; Klaassen, C.D., Ed.; Birkhäuser: Basel, Switzerland, 1999; pp. 15–21.
5. Höckner, M.; Stefanon, K.; Vaufleury, A.; Monteiro, F.; Pérez-Rafael, S.; Palacios, O.; Capdevila, M.; Atrian, S.; Dallinger, R. Physiological relevance and contribution to metal balance of specific and non-specific Metallothionein isoforms in the garden snail *Cantareus aspersus*. *Biometals* **2011**, *24*, 1079–1092.
6. Dallinger, R.; Berger, B.; Hunziker, P.; Kägi, J.H.R. Metallothionein in snail Cd and Cu metabolism. *Nature* **1997**, *388*, 237–238. [PubMed]
7. Palacios, Ò.; Pagani, A.; Pérez-Rafael, S.; Egg, M.; Höckner, M.; Brandstätter, A.; Capdevila, M.; Atrian, S.; Dallinger, R. Shaping mechanisms of metal specificity in a family of metazoan metallothioneins: Evolutionary differentiation of mollusc metallothioneins. *BMC Biol.* **2011**, *9*, 4.
8. Dvorak, M.; Lackner, R.; Niederwanger, M.; Rotondo, C.; Schnegg, R.; Ladurner, P.; Pedrini-Martha, V.; Salvenmoser, W.; Kremser, L.; Lindner, H.; et al. Metal binding functions of metallothioneins in the slug *Arion vulgaris* differ from metal-specific isoforms of terrestrial snails. *Metallomics* **2018**, *10*, 1638–1654.
9. Chabicovsky, M.; Niederstätter, H.; Thaler, R.; Höld, E.; Parson, W.; Rossmanith, W.; Dallinger, R. Localization and quantification of Cd- and Cu-specific metallothionein isoform mRNA in cells and organs of the terrestrial gastropod *Helix pomatia*. *Toxicol. Appl. Pharmacol.* **2003**, *190*, 25–36.
10. Hödl, E.; Felder, E.; Chabicovsky, M.; Dallinger, R. Cadmium stress stimulates tissue turnover in *Helix pomatia*: Increasing cell proliferation from metal tolerance to exhaustion in molluscan midgut gland. *Cell Tissue Res.* **2010**, *341*, 159–171.
11. Dallinger, R.; Chabicovsky, M.; Hödl, E.; Prem, C.; Hunziker, P.; Manzl, C. Copper in *Helix pomatia* (Gastropoda) is regulated by one single cell type: Differently responsive metal pools in rhogocytes. *Am. J. Physiol. Regul. Comp. Physiol.* **2005**, *289*, 1185–1195.
12. Pérez-Rafael, S.; Monteiro, F.; Dallinger, R.; Atrian, S.; Palacios, Ò.; Capdevila, M. *Cantareus aspersus* metallothionein metal binding abilities: The unspecific CaCd/CuMT isoform provides hints about the metal preference determinants in metallothioneins. *Biochim. Biophys. Acta* **2014**, *1844*, 1694–1707. [PubMed]
13. Dallinger, R.; Zerbe, O.; Baumann, C.; Egger, B.; Capdevila, M.; Palacios, Ò.; Albalat, R.; Calatayud, S.; Ladurner, P.; Schlick-Steiner, B.C.; et al. Metallomics reveals a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins. *Metallomics* **2020**, *12*, 702–720. [PubMed]

14. Palacios, Ò.; Pérez-Rafael, S.; Pagani, A.; Dallinger, R.; Atrian, S.; Capdevila, M. Cognate and noncognate metal ion coordination in metal-specific metallothioneins: The *Helix pomatia* system as a model. *J. Biol. Inorg. Chem.* **2014**, *19*, 923–935. [PubMed]
15. Hispard, F.; Schuler, D.; de Vaufleury, A.; Scheifler, R.; Badot, P.-M.; Dallinger, R. Metal distribution and metallothionein induction after cadmium exposure in the terrestrial snail *Helix aspersa* (Gastropoda, Pulmonata). *Environ. Toxicol. Chem.* **2008**, *27*, 1533–1542. [PubMed]
16. Coyle, P.; Philcox, J.; Carey, L.; Rofe, A. Metallothoinein: The multipurpose protein. *Cell. Mol. Life Sci.* **2002**, *59*, 627–647. [PubMed]
17. Feng, W.; Cai, J.; Pierce, W.; Franklin, R.; Maret, W.; Benz, F.; Kang, J. Metallothionein transfers zinc to mitochondrial aconitase through a direct interaction in mouse hearts. *Biochem. Biophys. Res. Commun.* **2005**, *332*, 853–858.
18. Loebus, J.; Leitenmaier, B.; Meissner, D.; Braha, B.; Krauss, G.-J.; Dobritzsch, D.; Freisinger, E. The major function of a metallothionein from the aquatic fungus *Heliscus lugdunensis* is cadmium detoxification. *J. Inorg. Biochem.* **2013**, *127*, 253–260.
19. Bofill, R.; Capdevila, M.; Atrian, S. Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu-thioneins. *Metallomics* **2009**, *1*, 229–234.
20. Palacios, Ò.; Atrian, S.; Capdevila, M. Zn- and Cu-thioneins: A functional classification for metallothioneins? *J. Biol. Inorg. Chem.* **2011**, *16*, 991–1009.
21. Pérez-Rafael, S.; Mezger, A.; Lieb, B.; Dallinger, R.; Capdevila, M.; Palacios, O.; Atrian, S. The metal binding abilities of *Megathura crenulata* metallothionein (McMT) in the frame of Gastropoda MTs. *J. Inorg. Biochem.* **2012**, *108*, 84–90.
22. Niederwanger, M.; Calatayud, S.; Zerbe, O.; Atrian, S.; Albalat, R.; Capdevila, M.; Palacios, Ò.; Dallinger, R. *Biomphalaria glabrata* metallothionein: Lacking metal specificity of the protein and missing gene upregulation suggest metal sequestration by exchange instead of through selective binding. *Int. J. Mol. Sci.* **2017**, *18*, 1457. [CrossRef] [PubMed]
23. Calatayud, S.; Garcia-Risco, M.; Pedrini-Martha, V.; Eernisse, D.J.; Dallinger, R.; Palacios, Ò.; Capdevila, M.; Albalat, R. Modularity in protein evolution: Modular organization and de novo domain evolution in mollusc metallothioneins. *Mol. Biol. Evol.* **2020**, msaa230. [CrossRef] [PubMed]
24. Yang, Q.-Q.; Yu, X.-P. A new species of apple snail in the genus *Pomacea* (Gastropoda: Caenogastropoda: Ampullariidae). *Zool. Stud.* **2019**, *58*, 1–7.
25. Halwart, M. The golden apple snail *Pomacea canaliculata* in Asian rice farming systems: Present impact and future threat. *Int. J. Pest Manag.* **1994**, *40*, 199–206. [CrossRef]
26. Horgan, F.B. The ecophysiology of apple snails in rice: Implications for crop management and policy. *Ann. Appl. Biol.* **2018**, *172*, 245–267. [CrossRef]
27. Capdevila, M.; Cols, N.; Romero-Isart, N.; González-Duarte, R.; Atrian, S.; González-Duarte, P. Recombinant synthesis of mouse Zn3-β and Zn4-α metallothionein 1 domains and characterization of their cadmium(II) binding capacity. *Cell. Mol. Life Sci.* **1997**, *53*, 681–688. [CrossRef] [PubMed]
28. Cols, N.; Romero-Isart, N.; Capdevila, M.; Oliva, B.; González-Duarte, P.; González-Duarte, R.; Atrian, S. Binding of excess cadmium(II) to Cd7-metallothionein from recombinant mouse Zn7-metallothionein 1. UV-VIS absorption and circular dichroism studies and theoretical location approach by surface accessibility analysis. *J. Inorg. Biochem.* **1997**, *68*, 157–166. [CrossRef]
29. Baumann, C.; Beil, A.; Jurt, S.; Niederwanger, M.; Palacios, O.; Capdevila, M.; Atrian, S.; Dallinger, R.; Zerbe, O. Structural adaptation of a protein to increased metal stress: NMR structure of a marine snail metallothionein with an additional domain. *Angew. Chem. Int. Ed.* **2017**, *56*, 4617–4622. [CrossRef]
30. Palacios, Ò.; Jiménez-Martí, E.; Niederwanger, M.; Gil-Moreno, S.; Zerbe, O.; Atrian, S.; Dallinger, R.; Capdevila, M. Analysis of metal-binding features of the wild type and two domain-truncated mutant variants of *Littorina littorea* metallothionein reveals its Cd-specific character. *Int. J. Mol. Sci.* **2017**, *18*, 1452. [CrossRef]
31. Schmielau, L.; Dvorak, M.; Niederwanger, M.; Dobieszewski, N.; Pedrini-Martha, V.; Ladurner, P.; Rodríguez-Guerra Pedregal, J.; Maréchal, J.-D.; Dallinger, R. Differential response to cadmium exposure by expression of a two and a three-domain metallothionein isoform in the land winkle *Pomatias elegans*: Valuating the marine heritage of a land snail. *Sci. Total Environ.* **2019**, *648*, 561–571. [CrossRef]
32. Beil, A.; Jurt, S.; Walser, R.; Schönhut, T.; Güntert, P.; Palacios, Ò.; Atrian, S.; Capdevila, M.; Dallinger, R.; Zerbe, O. The solution structure and dynamics of Cd-metallothionein from *Helix pomatia* reveal optimization for binding Cd over Zn. *Biochemistry* **2019**, *58*, 4570–4581. [CrossRef]
33. Blindauer, C.A. Metallothioneins with unusual residues: Histidines as modulators of zinc affinity and reactivity. *J. Inorg. Biochem.* **2008**, *102*, 507–521. [CrossRef]
34. Blindauer, C.A.; Razi, M.T.; Campopiano, D.J.; Sadler, P.J. Histidine ligands in bacterial metallothioneins enhance cluster stability. *J. Biol. Inorg. Chem.* **2007**, *12*, 393–405. [CrossRef]
35. Perinelli, M.; Tegoni, M.; Freisinger, E. Different behavior of the histidine residue toward cadmium and zinc in a cadmium-specific metallothionein from an aquatic fungus. *Inorg. Chem.* **2020**, *59*, 16988–16997. [CrossRef] [PubMed]
36. Vergani, L.; Grattarola, M.; Grasselli, E.; Dondero, F.; Viarengo, A. Molecular characterization and function analysis of MT-10 and MT-20 metallothionein isoforms from *Mytilus galloprovincialis*. *Arch. Biochem. Biophys.* **2007**, *465*, 247–253. [CrossRef]
37. Orihuela, R.; Domènech, J.; Bofill, R.; You, C.; Mackay, E.; Kägi, J.H.R.; Capdevila, M.; Atrian, S. The metal-binding features of the recombinant mussel *Mytilus edulis* MT-10-IV metallothionein. *J. Biol. Inorg. Chem.* **2008**, *13*, 801–812. [CrossRef] [PubMed]

38. Maltez, H.F.; Villanueva Tagle, M.; Fernández de la Campa, M.d.R.; Sanz-Mendel, A. Metal–metallothioneins like proteins investigation by heteroatom-tagged proteomics in two different snails as possible sentinel organisms of metal contamination in freshwater ecosystems. *Anal. Chim. Acta* **2009**, *655*, 234–240. [CrossRef] [PubMed]
39. Calatayud, S.; Garcia-Risco, M.; Rojas, N.; Espinosa-Sánchez, L.; Artime, S.; Palacios, Ò.; Cañestro, C.; Albalat, R. Metallothioneins of the urochordate *Oikopleura dioica* have Cys-rich tandem repeats, large size and cadmium-binding preference. *Metallomics* **2018**, *10*, 1585–1594. [CrossRef]
40. Bongers, J.; Walton, C.D.; Richardson, D.E. Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma atomic emission spectrometric determination of sulfur. *Anal. Chem.* **1988**, *60*, 2683–2686. [CrossRef]
41. Capdevila, M.; Domènech, J.; Pagani, A.; Tío, L.; Villarreal, L.; Atrian, S. Zn- and Cd-metallothionein recombinant species from the most diverse phyla may contain sulfide (S2-) ligands. *Angew. Chem. Int. Ed.* **2005**, *44*, 4618–4622. [CrossRef]
42. Fabris, D.; Zaia, J.; Hathout, Y.; Fenselau, C. Retention of thiol protons in two classes of protein zinc ion coordination centers. *J. Am. Chem. Soc.* **1996**, *118*, 12242–12243. [CrossRef]
43. Bofill, R.; Palacios, Ò.; Capdevila, M.; Cols, N.; González-Duarte, R.; Atrian, S.; González-Duarte, P. A new insight into the Ag+ and Cu+ binding sites in the metallothionein beta domain. *J. Inorg. Biochem.* **1999**, *73*, 57–64. [CrossRef]

ARTICLE III: SUPPLEMENTARY MATERIAL

TWO UNCONVENTIONAL METALLOTHIONEINS IN THE APPLE SNAIL POMACEA BRIDGESII HAVE LOST THEIR METAL SPECIFICITY DURING ADAPTATION TO FRESHWATER HABITATS

The file includes:

Figure S1

Figure S2

Figure S3

Figure S4

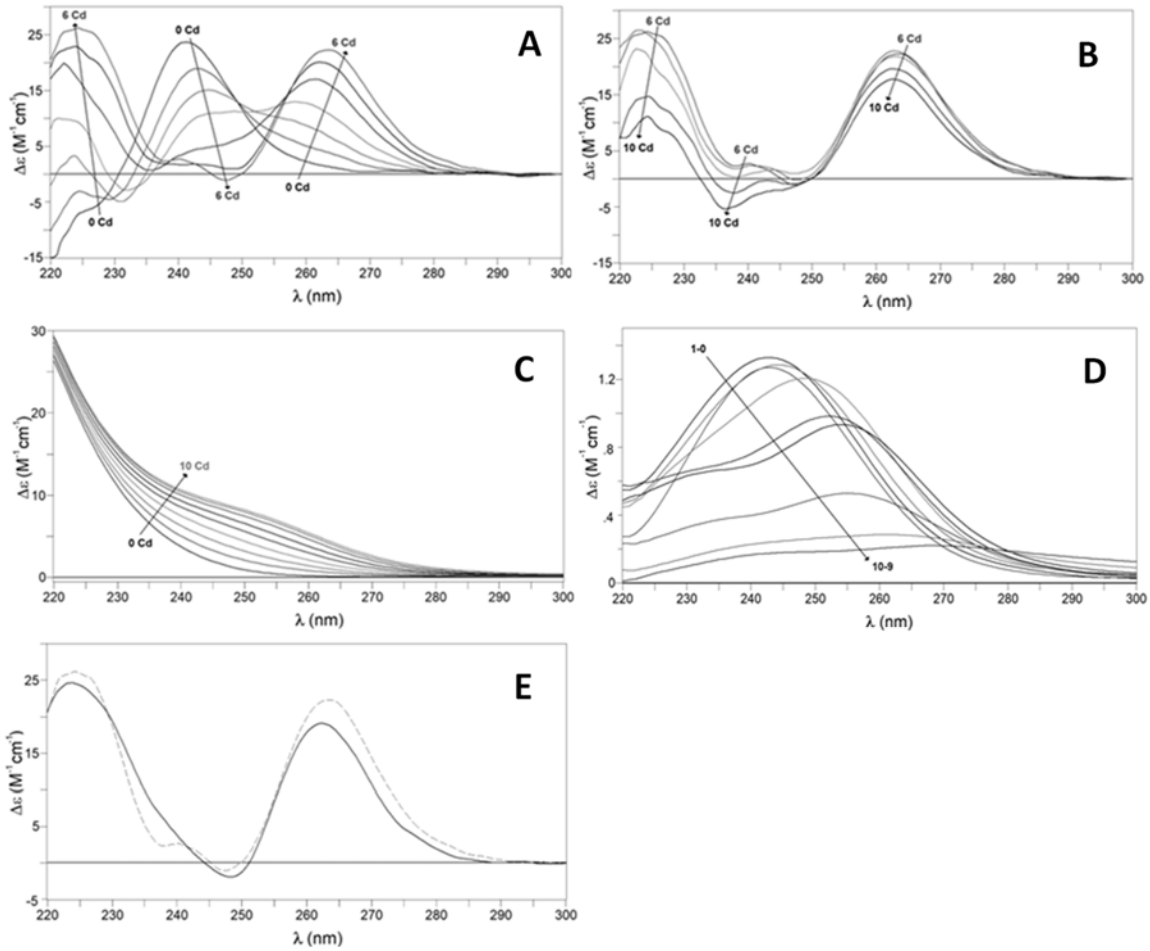


Figure S1. Spectroscopic characterization of the Zn/Cd replacement in the Zn-PbrMT1 preparation followed by (A and B) CD and (C) UV-vis. (D) UV-Vis difference spectra. (E) Comparison of CD spectra of *in vivo* Cd-PbrMT1 preparation (solid line) and the *in vitro* Cd-PbrMT1 spectrum measured after adding 7 equivalents of Cd(II) to the Zn-PbrMT1 preparation (dashed line)

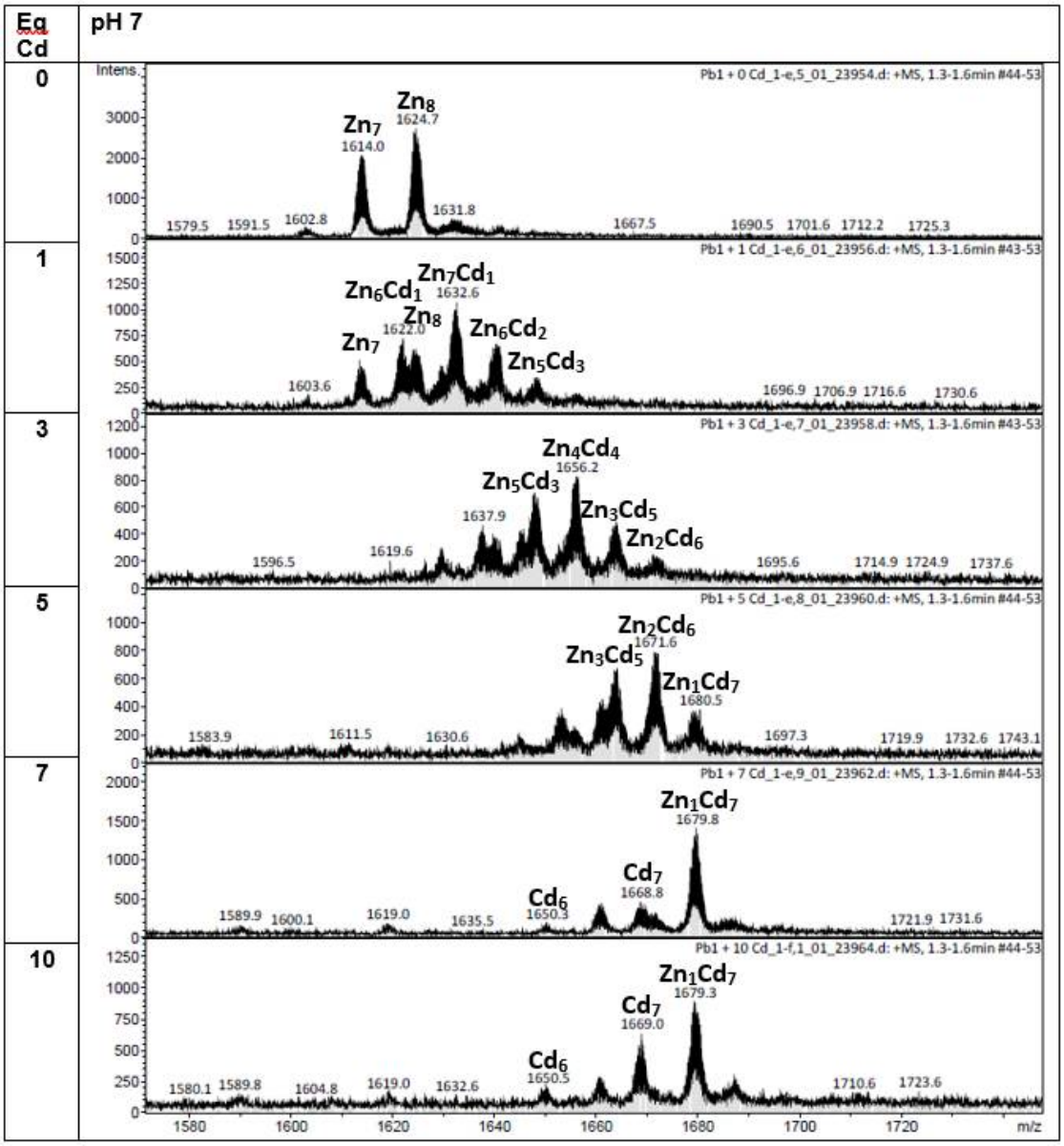


Figure S2. ESI-MS spectra of the Zn/Cd replacement experiment performed on the Zn-PbrMT1 production at pH 7.

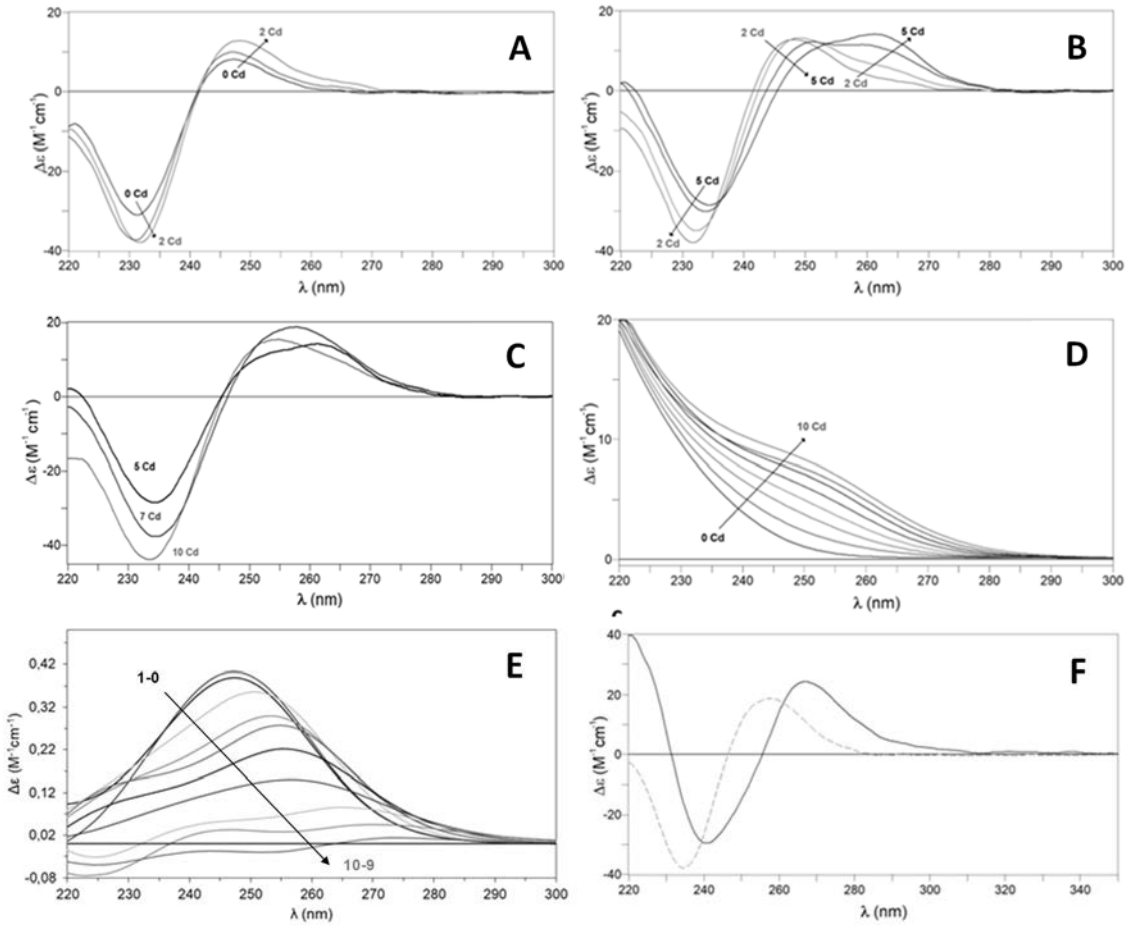


Figure S3. Spectroscopic characterization of the Zn/Cd replacement in the Zn-PbrMT2 preparation followed by (A, B and C) CD and (D) UV-vis. (E) UV-Vis difference spectra. (F) Comparison of CD spectra of *in vivo* Cd-PbrMT2 species (solid line) and the *in vitro* Cd-PbrMT2 species (dashed line) after adding 7 equivalents of Cd(II) to Zn-PbrMT2 production.

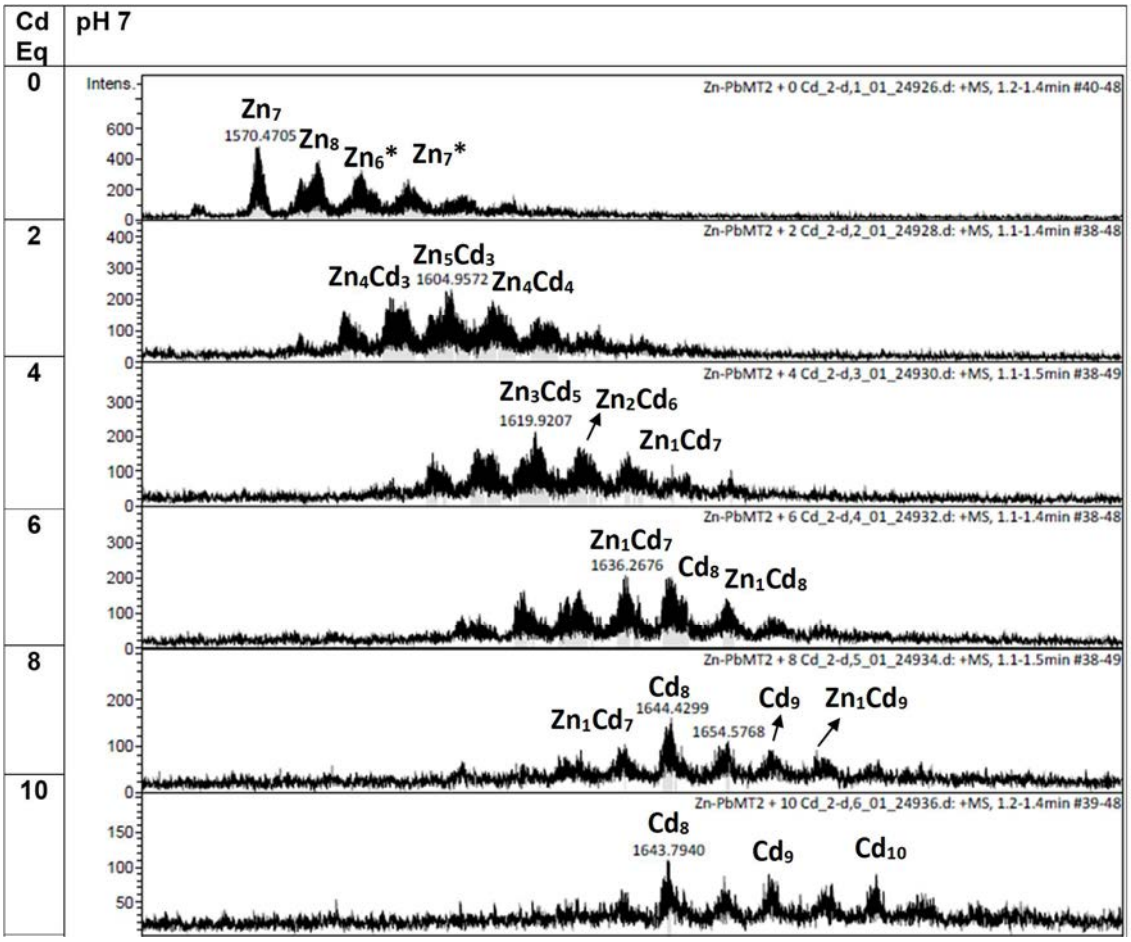


Figure S4. ESI-MS spectra of the Zn/Cd replacement experiment performed on the Zn-PbrMT2 production at pH 7.

RESULTS II: CHORDATE MTs

Regarding chordate MTs, the results have been described in three articles. The first article entitled "*Metallothioneins of the urochordate Oikopleura dioica have Cys-rich tandem repeats, large size and cadmium-binding preference*" by **Sara Calatayud**, Mario Garcia-Risco, Natalia S. Rojas, Lizethe Espinosa-Sánchez, Sebastián Artime, Òscar Palacios, Cristian Cañestro and Ricard Albalat, has been published in **Metallomics** 10: 1585-1594. (2018) DOI: 10.1039/c8mt00177d (IF_{2018}: 3.571).

This article describes the identification of two MTs in the tunicate appendicularian *Oikopleura dioica*, OdiMT1 and OdiMT2 (formerly OdMT1 and OdMT2). A detailed inspection of the sequence revealed a complex modular organization of both MTs. OdiMT1 was a 72 amino acid protein made of two putative domains, one full-length 12C domain (previously C7b+C-tail subunits), and one truncated t-12C form (previously C7a subunit). OdiMT2 was a 399 amino acid protein, also made of several tandem repeats of 12C and t-12C domains, that represented the longest MT reported to date for any living being. Bioinformatic analyses of regulatory regions of both *OdiMT* genes revealed the presence of binding sites for Metal Regulatory Elements Factor (MREF), as well as for different stress-responsive factors, suggesting a role of these MTs in response to heavy metals and other adverse environmental conditions. Finally, the metal-binding characterization of both OdiMTs showed that both MTs were capable to form metal-complexes, with preference for Cd ions.

The second article entitled "*Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins*" by **Sara Calatayud**, Mario Garcia-Risco, Mercè Capdevila, Cristian Cañestro, Òscar Palacios and Ricard Albalat, has been published in **Frontiers of Cell and Developmental Biology** 9: 702688 (2021) DOI: 10.3389/fcell.2021.702688 (IF_{2020}: 6.684).

In this article, I further characterized the two OdiMTs, with special interest in the functional analysis of the previously predicted 12C domain. I determined the metal-binding features of the 12C domain, both the full-length and the trimmed form. I demonstrated that the predicted 12C domain was a genuine domain capable to autonomously coordinate divalent metal ions. Although the 12C domain was functionally autonomous, I also showed that the t-12C/12C pair was an improved form for divalent metal binding. In addition, the results indicated a direct relationship between the number of the t-12C/12C domain repeats and the metal-binding capacity

of OdiMTs. Finally, the results revealed an unexpected variability in the number of t-12C/12C domain repeats in the OdiMT2 from different *O. dioica* populations, suggesting a high structural plasticity of MTs probably facilitated by modular organization of these proteins.

The third article entitled "*Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions and functional convergences*" by **Sara Calatayud**, Mario Garcia-Risco, Òscar Palacios, Mercè Capdevila, Cristian Cañestro and Ricard Albalat, has been published in **Molecular Biology and Evolution** msab184 (2021). DOI: 10.1093/molbev/msaa230 (IF_{2020}: 16.240).

In this article, I analysed the evolution of MT in chordates by identifying and analysing more than 160 MTs from 44 tunicate species. Tunicates include three different classes: Ascidiacea, Thaliacea and Appendicularia. First, I showed that the most ascidian and thaliacean MTs were made from a single 12C domain with a clear Cd-thionein character. In contrast, the MTs of several Stelydae species were large multi-domain forms with tandem repeats of the 12C domain. Appendicularians MT were also multi-domain forms, but they were made of tandem repeats of a different 12C domain. Comparative analyses of tunicate MTs suggested, in fact, that ascidian and thaliacean MTs represented the prototypical tunicate form whereas appendicularian MTs probably had an independent evolutionary origin. Finally, comparisons of tunicate, cephalochordate and vertebrate MTs revealed a complex evolutionary history of chordate MTs from an ancestral bi-domain Cd-thionein. This ancestral form underwent different linage-specific modifications, including loss or readjustment of the domains, duplications that led to protein multiplicity and neo- or subfunctionalization processes, recurrent expansion of the number of domains that generated different multi-domain forms, and *de novo* emergence of genes and domains probably associated with processes of convergent evolution.

ARTICLE IV: *Metallothioneins of the urochordate Oikopleura dioica have Cys-rich tandem repeats, large size and cadmium-binding preference.*

Metallomics, 2018,10, 1585-1594

DOI: 10.1039/C8MT00177D

Metallomics

PAPER

Check for updates

Cite this: *Metallomics*, 2018, **10**, 1585

Metallothioneins of the urochordate *Oikopleura dioica* have Cys-rich tandem repeats, large size and cadmium-binding preference†

Sara Calatayud,[a] Mario Garcia-Risco,[b] Natalia S. Rojas,[a] Lizethe Espinosa-Sanchez,[a] Sebastian Artime,[a] Oscar Palacios,[b] Cristian Canestro[a] and Ricard Albalat*[a]

Received 3rd July 2018,
Accepted 28th September 2018

DOI: 10.1039/c8mt00177d

rsc.li/metallomics

The increasing levels of heavy metals derived from human activity are poisoning marine environments, threating zooplankton and ocean food webs. To protect themselves from the harmful effects of heavy metals, living beings have different physiological mechanisms, one of which is based on metallothioneins (MTs), a group of small cysteine-rich proteins that can bind heavy metals counteracting their toxicity. The MT system of urochordate appendicularians, an ecologically relevant component of the zooplankton, remained, however, unknown. In this work, we have characterized the MTs of the appendicularian species *Oikopleura dioica*, revealing that *O. dioica* has two *MT* genes, named *OdMT1* and *OdMT2*, which encode for Cys-rich proteins, the former with 72 amino acids comparable with the small size MTs of other organisms, but the second with 399 amino acids representing the longest MT reported to date for any living being. Sequence analysis revealed that *OdMT2* gene arose from a duplication of an ancestral *OdMT1* gene followed by up to six tandem duplications of an ancestral repeat unit (RU) in the current *OdMT2* gene. Interestingly, each RU contained, in turn, an internal repeat of a 7-Cys subunit ($X_3CX_3CX_2CX_2CX_{3\ 6}CX_2CXCX$), which is repeated up to 12 times in *OdMT2*. Finally, ICP-AES analyses of heterologously expressed OdMT proteins showed that both MTs were capable to form metal-complexes, with preference for cadmium ions. Collectively, our results provide the first characterization of the MT system in an appendicularian species as an initial step to understand the zooplankton response to metal toxicity and other environmental stress situations.

Significance to metallomics

The analysis of the metallothioneins of *Oikopleura dioica* is important for understanding the evolution and function of the detoxification mechanisms that this zooplankton species might use to counteract the harmful effects of heavy metal exposure. This knowledge is fundamental to estimate the potential impact that an increase of heavy metal amounts may have on marine ecosystems.

1. Introduction

Heavy metals such as zinc, iron or copper are essential for several biological processes, but toxic at high concentrations, while others such as cadmium, mercury or lead are highly poisonous even at low concentrations. Living beings have different physiological mechanisms to control the homeostasis of essential metals as well as to counteract the harmful effects of the non-essential ones. One of these mechanisms is based on metallothioneins (MTs), a group of metal-binding proteins originally discovered in 1957 in the horse kidney cortex,[1] and classically considered a diverse family of Cys-rich (≈30%) and low molecular weight (<60 amino acids) proteins found in almost all organisms (reviewed in ref. 2 and 3). The amino acid sequence of different MTs is highly heterogeneous, particularly when MTs of distantly related taxa are compared. In such comparisons, sequence similarity appears restricted to an overall ``Cys abundance with distinctive cysteine motifs (*i.e.* CXC, CC and CCC), whereas the length of the protein, the number

[a] *Departament de Genetica, Microbiologia i Estadıstica and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, E-08028, Barcelona, Catalonia, Spain. E-mail: ralbalat@ub.edu*

[b] *Departament de Quımica, Facultat de Ciencies, Universitat Autonoma de Barcelona, E-08193 Cerdanyola del Valles, Spain*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8mt00177d

and distribution of cysteines and the intercalating residues are largely variable from one taxon to another.[2] Vertebrate MT sequences, for instance, can be reliably aligned within the subphylum, but they are barely comparable with other MT sequences, even from their closest chordate relatives, amphioxus cephalochordates or ascidian urochordates.[4,5]

From a structural perspective, MTs have been extensively analyzed showing that they are able to coordinate a number of heavy metal ions through the formation of metal thiolate bonds.[6,7] Depending on their metal-binding preferences, MTs can be classified in a stepwise gradation from extreme Zn/Cd-thioneins to extreme Cu-thioneins.[8,9] The metal-binding preference of a given MT cannot be, however, directly predicted from its amino acid sequence, and metal-binding assays have to be performed to classify newly discovered forms.

From a physiological perspective, MTs are activated for controlling metal homeostasis and detoxification roles, but also for radical scavenging, oxidative stress protection or anti-apoptotic defense (reviewed in ref. 6). The presence of Metal Responsive Elements (MRE) and different response elements for transcription factors (TF) involved in stress response (*e.g.* AREs, HSE, StRE) in the promoter regions has been associated to the transcriptional activation of *MT* genes in response to heavy metals and other stress situations (reviewed in ref. 10).

In the last century, what has been called the Anthropocene, human activities such as those related to the industries of mining, metal plating, fertilizer, paper, pesticides or batteries, are increasing the amount of heavy metals in aquatic environments,[11,12] becoming a serious problem for coastal and marine ecosystems.[13] In marine ecosystems, Appendicularians (a.k.a. Larvaceans; phylum Chordata, subphylum Urochordata) represent the second most abundant group of mesozooplankton grazers and an important component of food for fish and zooplankton species.[14] Appendicularians are also ecologically relevant because they contribute to the vertical transport of carbon to deep ocean through the rapid sinking of fecal pellets and discarded houses.[15–17] Among Appendicularians, *Oikopleura dioica* is the most studied species, and it is becoming a new laboratory model for comparative genetic and genomic analyses,[18–22] developmental biology studies,[23–30] as well as ecological and toxicological investigations.[31–34]

The ecological relevance of Appendicularians together with the potential of *O. dioica* as an experimental model[35–38] prompted us to survey the MTs of *O. dioica* to characterize the MT system of an Appendicularian species. Here we report the identification of two *MT* genes in *O. dioica*, named *OdMT1* and *OdMT2*, and the analysis of their gene structures and promoter regions. *OdMT1* encode for a Cys-rich protein of 72 amino acids (theoretical molecular weight 7645 Da) comparable with the small size MTs of other organisms (human MT1 theoretical molecular weight 6120 Da), while *OdMT2* encode for a Cys-rich protein of 399 amino acids (theoretical molecular weight 42 716 Da) representing the longest MT reported to date for any living being. Interestingly, sequence analysis reveals that *OdMT2* gene arose from a duplication of an ancestral *OdMT1* gene followed by up to six tandem duplications of an ancestral repeat unit (RU) in the current *OdMT2* gene. *OdMT* promoters appear rich in stress-responsive elements, including binding sites for MREF, YAP1/CREB, RXR/PPAR and HSF transcription factors, envisaging a role of *MT* genes in adverse environmental conditions. Finally, heterologously expressed OdMT1 and OdMT2 proteins showed that they are capable to form metal complexes, with preference for cadmium ions. Overall, these results pave the way for a better understanding of the homeostasis of the physiological metals in *O. dioica*, as well as the genetic mechanisms that Appendicularians might use against the harmful effects of heavy metal exposure.

2. Material and methods

2.1. Genome survey of *O. dioica MT* genes

To identify *O. dioica MT* homologs we made low-restrictive tblastn searches[39] in the genome database of *O. dioica* (Oikobase, http://oikoarrays.biology.uiowa.edu/Oiko) with a high value of expect threshold (1000) and no filtering for low complexity sequences, using the two only available MTs of Urochordate ascidians *Ciona robusta* and *Herdmania curvata* MTs (accession numbers ACN32211 and AY314949, respectively), as well as Cephalochordate and Vertebrate MTs as queries.[4,5] Gene annotations were manually reviewed and corrected based on ESTs availability: EST FP794470.1 for *OdMT1* and a consensus sequence derived from the assembly of 18 SRA sequences for *OdMT2* (Fig. S1 and Table S1, ESI†).

2.2. Promoter analysis and stress-responsive elements

The analysis of putative transcription factor binding sites for *OdMT1* and *OdMT2* genes was performed by means of the MatInspector 8.4.1 program with default parameters, using the MatBase 11.0 from the Genomatix software suite.[40] The MatBase 11.0 includes 2029 weight matrices of 482 families, representing binding sites descriptions of more than 11 000 TF. We restricted our analysis to a subset of the 33 families that included TF involved in metal- or stress-response in animals, plants or fungi (Table S3, ESI†). A 1000-bp region upstream of predicted CDS of both *OdMT* genes was selected for the promoter analysis.

2.3. Cloning of *OdMT1* and *OdMT2* for recombinant expression

A full-length synthetic cDNA for predicted *OdMT1* gene was synthesized by Integrated DNA Technologies Company (Coralville, IA, USA), while for technical limitations, the cDNA of *OdMT2* was split in two fragments of 610 nt (5′-fragment 1) and 660 nt (3′-fragment 2), overlapping 52 nt among them. *Bam*HI and *Xho*I restriction sites and 6 7 additional 5′-nucleotides were added to both *OdMT* cDNA ends to facilitate the cloning processes. The synthetic cDNA for *OdMT1* was PCR amplified −94 °C 5 minutes (min); 25 cycles of 94 °C 30 seconds (s), 55 °C 30 s and 72 °C 30 s; and 72 °C 7 min with specific primers (5′GGGGGATCCATGGATCCGGTTTGCTCTTTCCGCTG3′ and 5′GGGCTCGAGTTATTCCGCTGTGCTGGTCGGGCAG3′) using Expand High Fidelity PCR system® (Roche). The full-length

This journal is © The Royal Society of Chemistry 2018

synthetic cDNA for *OdMT2* was reconstructed in two steps by adapting previous methods for assembling overlapping DNA fragments.[41 44] In brief, in the first step, the two *OdMT2* fragments were denaturalized at 98 °C for 3 min, annealed at 80 °C for 1 min, and extended at 94 °C for 30 s. Ten cycles of annealing and extension were repeated to generate a single full-length cDNA molecule from the assembly of the two overlapping fragments. In the second step, the resulting cDNA for *OdMT2* was PCR-amplified 15 cycles of 76 °C 30 s and 94 °C 30 s; and 76 °C 5 min with specific end-terminal primers (5′GGGGGATCCATGGAAGTAAAACGACC3′ and 5′GGGCTCGAG TTAACAGCATTTTTTGG3′) using Phusion High Fidelity DNA polymerase (New England Biolabs, Thermo Scientific). *OdMT1* and *OdMT2* PCR products were *BamH*I/*Xho*I-digested and cloned into the *E. coli* pGEX-4T-1 expression vector (GE Healthcare) with the DNA Ligation kit 2.1 (Takara Bio Inc.), and transformed into *E. coli* Dh5α strain. Plasmid DNA was purified from bacteria using the GeneElute™ Plasmid Miniprep Kit (Sigma-Aldrich), screened for insert presence by digestion with *Sca*I enzyme, and sequenced at the Scientific and Technological Centers of the University of Barcelona, using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) in an automatic sequencer (ABIPRISM 310, Applied Biosystems). DNA from each recombinant *OdMT*-pGEX plasmid was used to transform *E. coli* BL21 strain, a protease deficient strain used for heterologous protein expression.

2.4 Production of recombinant metal–*OdMT1* and –*OdMT2* complexes

For heterologous protein production 500 mL of LB medium with 100 μg mL^{-1} ampicillin were inoculated with *E. coli* BL21 cells transformed with the *OdMT1* or *OdMT2* recombinant plasmids. After overnight growth at 37 °C/250 rpm, the cultures were used to inoculate 5 L of fresh LB-100 μg mL^{-1} ampicillin medium. Gene expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 hours (h). After the first 30 min of induction, cultures were supplemented with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM) or $CuSO_4$ (500 μM) in order to generate metal OdMT complexes. Cells were harvested by centrifugation for 5 min at 9100*g* (7700 rpm), and bacterial pellets were suspended in 125 mL of ice-cold PBS (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4 and 0.5% v/v β-mercaptoethanol). Resuspended cells were sonicated (Sonifier Ultrasonic Cell Disruptor) 8 min at voltage 6 with pulses of 0.6 s, and then centrifuged for 40 min at 17 200*g* (12 000 rpm) and 4 °C.

2.5. Purification of recombinant metal–OdMT complexes

Protein extracts containing GST OdMT1 or GST OdMT2 fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. GST OdMT fusion proteins bound to the sepharose beads were washed with 30 mL of cold 1× PBS bubbled with argon to prevent oxidation. After three washes, GST MT fusion proteins were digested with thrombin (GE Healthcare, 25 U L^{-1} of culture) overnight at 17 °C, thus enabling separation of the metal OdMT complexes from the GST that remained bound to the sepharose matrix. The eluted metal OdMT complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris HCl, pH 7.0 for OdMT1, and with fresh 50 mM amonium acetate, pH 7.0 for OdMT2, and run at 0.8 mL min^{-1}. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at −80 °C until use.

2.6. Analysis of metal–OdMT complexes

Metal OdMT complexes were analyzed by different techniques: inductively coupled plasma atomic emission spectroscopy (ICP-AES) for protein quantification and element composition (S, Zn, Cd and Cu) on a Polyscan 61E spectrometer (Thermo Jarrell Ash Corporation, Franklin, MA, USA) at appropriate wavelengths (S, 182.040 nm; Zn, 213.856 nm; Cd, 228.802 nm; Cu, 324.803 nm) under conventional (dilution with 2% HNO_3 (v/v)) conditions.[45] OdMT concentrations in the recombinant preparations were calculated from sulfur measurements, assuming the only contribution to their S content was that made by the OdMT peptides.

Electrospray ionization mass spectrometry with a time-of-flight analyzer (ESI-TOF MS) was used to determine the molecular mass of the species formed, using a Micro Tof-Q Instrument (Bruker Daltonics Gmbh, Bremen, Germany) calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies, Santa Clara, CA, USA), interfaced with a Series 1100 HPLC pump (Agilent Technologies) equipped with an autosampler, both controlled by the Compass Software. The experimental conditions for analyzing the proteins were as follows: 10 20 μL of the sample were injected at 30 50 μL min^{-1} at 3.5 5.0 kV capillary-counter electrode voltage, 90 110 °C desolvation temperature, and dry gas at 6 L min^{-1}. Spectra were collected throughout an *m*/*z* range from 800 to 2000. The liquid carrier was a 90 : 10 mixture of 15 mM ammonium acetate and acetonitrile, pH 7.0. For the analysis at acidic pH the conditions used were the same as those used in the analysis of the divalent metals, except in the composition of the carrier liquid, which in this case was a 95 : 5 mixture of formic acid and acetonitrile at pH 2.4. All samples were injected in duplicates to ensure reproducibility. In all cases, molecular masses were calculated according to the reported method.[46]

3. Results and discussion

3.1. Identification, structural features and evolutionary origin of *O. dioica* MTs

We conducted an exhaustive survey of *MT* genes in the *O. dioica* genome.[19,21] Because standard blast strategies[39] were unsuccessful for identifying putative *MT* genes in *O. dioica* sequence databases, we used low-restrictive tblastn searches using *C. robusta* and *H. curvata* MTs as well as Cephalochordate and Vertebrate MTs as query sequences. We manually inspected all the retrieved sequences for Cys-rich ORFs with MT-distinctive Cys arrangements (*i.e.* CXC, CXXC, CC and CCC) as structural criteria for identification of new putative MT-coding genes.

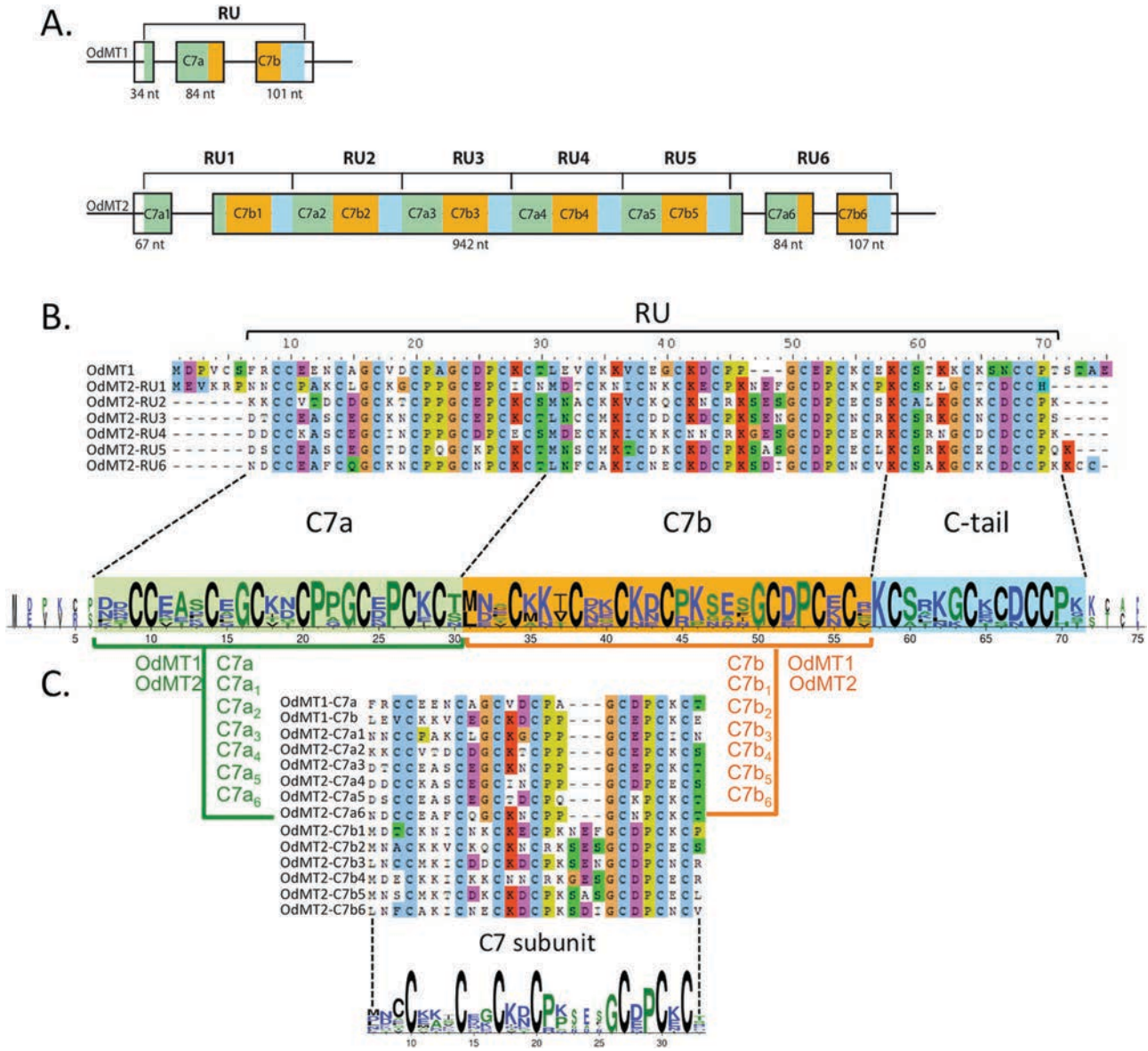


Fig. 1 *O. dioica* metallothioneins. (A) Schematic representation of the exon/intron structure of *OdMT1* and *OdMT2* genes. *OdMT1* is split in three exons and encodes for a protein made of a single repeat unit (RU) with an internal repeat of a C7 subunit C7a (green box) and C7b (orange box) followed by a C-terminal tail (blue box). *OdMT2* is split in four exons and encodes for a protein made of 6 RU (RU1 to RU6) with the same C7a/C7b/C-tail structure. (B) Amino acid alignment of OdMT1 and RU1 RU6 of OdMT2, and graphical representation using WebLogo 3[78] highlighting the 20 conserved Cys in each RU. Each RU contains a repetition of a C7 subunit (C7a in green background, and C7b in orange background) followed by a C-terminal tail (in blue background). (C) Comparison of C7a and C7b defines the C7 subunit as a 24 27 amino acid sequence with 7 Cys ($X_3CX_3CX_2CX_2CX_{3\ 6}CX_2CXCX$). In the WebLogo, the overall height of the stack indicates the conservation degree at a given position, while the height of symbols within the stack indicates the relative frequency of each amino at that position. Color code represents amino acid hydrophobicity scale: hydrophobic amino acids (YVMCLFIW) in black, neutral (SGHTAP) in green, and hydrophilic (RKDENQ) in blue.

Two *O. dioica* genomic sequences satisfied the criteria, one in the scaffold 50 and another in the scaffold 16, which we named *OdMT1* and *OdMT2*, respectively. *OdMT1* gene spanned 317 nt and its coding region (CDS) was organized in 3 exons of 34 nt, 84 nt and 101 nt (Fig. 1A and Fig. S2, ESI†). *OdMT2* gene extended over 1359 nt, and its CDS was organized in three small exons (exons 1, 3 and 4) of 67 nt, 84 nt and 107 nt, respectively, and one large exon (exon 2) of 942 nt (Fig. 1A and Fig. S2, ESI†).

OdMT1 encoded a 72 amino acid protein with 20 Cys (28%), and *OdMT2*, in contrast, encoded a much larger protein, with 399 amino acids containing 123 Cys (31%), being to our knowledge the longest MT so far described in any living being. A detail comparison of OdMT1 and OdMT2 sequences revealed that OdMT2 was made of 6 direct tandem repeat units (RU) of approximately 65 amino acids, each RU resembling to OdMT1 (Fig. 1A and B and Fig. S2, ESI†). Moreover, close inspection revealed that each RU was made of an internal repeat of 24 27

This journal is © The Royal Society of Chemistry 2018

amino acids with 7 conserved Cys (C7a and C7b subunits: $X_2CX_3CX_2CX_2CX_3{}_6CX_2CXCX$; Fig. 1B and C) followed by a C-terminal tail of 14 amino acids with 5 additional Cys residues. The fact that OdMT1 corresponded to a single RU whereas OdMT2 had 6 RUs suggested that the latter might have been originated by duplication of an ancestral *OdMT1* gene with one RU followed by successive tandem duplications of the primeval RU. These results indicated a modular and step-wise evolution of OdMTs, in which the unit of expansion was a RU made of two C7 subunits, likely related with functional and structural constraints that would link the number of RUs and C7 subunits to the metal-binding capacity of OdMTs. Interestingly, a modular structure based on different C7 subunits had been found in several long fungal MTs, that is, in *Cryptococcus neoformans* CnMT1 and CnMT2[47] and in *Tremella mesenterica* TmMT,[48] although it had never been reported in multicellular eukaryotes, and never to the level reached by *O. dioica* MTs. Tandem amplifications of C7 subunits yielding MTs with high metal-binding capacity had been associated with biological adaptations, as virulence factors in the pathogenic fungus *C. neoformans*,[47] or for Cu-providing to lignin-metabolizing enzymes in saprophytic fungus living in decaying woods as *T. mesenterica*.[48] The selective advantage that MTs with high metal-binding capacity might provide to a marine animal as *O. dioica* is still unknown, but it has been suggested that the evolution of multi-domain MTs in some animal species (*e.g. Littorina littorea* and *Pomatias elegans* with 3-domain MTs[49 51] *Crassostrea virginica* with 4-domain MTs,[52] and *Alinda biplicata* with 10-domain MTs[53]) might have improved their response to metal stress and adverse environmental conditions, offering a selective advantage to organisms that experience a greater exposure to metals due to their ecological niche.[50,52]

3.2. Promoter prediction analysis

Metals, stress and hormones regulate transcription of *MT* genes (reviewed in ref. 10). In animals, MT expression can be activated in response to metals by MRE-binding transcription factor 1 (MTF-1) that recognizes Metal Responsive Elements (MRE) present in MT promoters.[54] MRE have been experimentally and bioinformatically identified in proximal *MT* promoters of most metazoans, from insects and mollusks to fish and mammals[55 61] (reviewed in ref. 62). We surveyed for MRE in 1 kb upstream sequence from the predicted CDS of both *OdMT1* and *OdMT2* genes using MatInspector 8.4.1 program but, unexpectedly, we did not find any putative MRE sequence in their promoter regions. In agreement with the absence of MRE sequences, a genome survey for a *O. dioica* homolog of the *MTF-1* gene was also unsuccessful, making thereby unlikely that these elements MRE sequences and MTF-1 proteins contributed to the transcriptional regulation of *MT* genes in this species. In contrast, bioinformatic analysis predicted the presence MREF-binding sequences (ACE1/AMT1) in the promoters of *OdMT1* and *OdMT2* genes (Fig. 2A and Fig. S2, ESI†), which are known to confer metal response to fungal metallothioneins *CUP1* and *CRS5* genes. Binding of metal regulatory elements factors (MREF) such as CUP2 (a.k.a ACE1 or AMT1 transcription factor) to ACE1/AMT1 sequences regulates *CUP1* and *CRS5* in response to copper levels,[63,64] as well as they regulate fungal *copper zinc superoxide dismutase* (*Cu/ZnSOD1*) genes.[65] Although CUP2 orthologous have not been identified outside fungal species, ACE1/AMT1 sequences have been related to transcriptional regulation of *Cu/ZnSOD1* genes in maize[66] and *Arabidopsis*.[67] Therefore, we also analyzed the promoters (1 kb upstream sequence from the CDS) of the five *SOD1* genes predicted in *O. dioica* genome, and interestingly, we identified MREF (ACE1/AMT1) sequences in four of them (Fig. 2A), suggesting that as described in fungi, *MT* and *SOD* genes shared a common transcriptional response in *O. dioica*.

Because MTs are activated in many stress circumstances, we extended our promoter analyses to 33 families of transcription factors with members involved in stress response in animals, plants and fungi (Table S2, ESI†). Interestingly, in addition to MRFE sequences, we discovered that *OdMT1* and *OdMT2* promoters shared binding sites for many stress response elements (Table S3, ESI†). First, *OdMT* promoters included binding sites for the family of fungal YAP1, most of them overlapping with sites for vertebrate (CREB) bZIP transcription factors (Fig. 2B, Fig. S2 and Table S3, ESI†). Fungal YAP1 transcription factor belongs to the bZIP (basic leucine zipper factors) family, and it is required for cadmium and oxidative stress tolerance in yeast.[68] Vertebrate bZIP transcription factors of the CREB family (*e.g.* CREB, ATF-1, XBP1, E4BP4) might act also as activators of vertebrate *MT* genes,[69] and CRE binding sites have been identified in the promoter of earthworm *MT* genes.[70,71] Second, it has also been reported that transcription factors of the nuclear receptor family such as the steroid receptors, retinoid receptors and orphan receptors might be involved either directly (*e.g.* GR;[72] PPAR[56,57]) or indirectly (*e.g.* RXR[73]) in the transcriptional regulation of *MT* expression. Our analysis of *OdMT1* and *OdMT2* genes revealed several binding sites for nuclear receptors, mainly for RXR/PPAR, in the promoter region (Fig. 2B, Fig. S2 and Table S3, ESI†). Third, it has been demonstrated that heat shock stress also induces *MT* expression through the binding of HSF to the heat shock elements (HSE),[74,75] which appear remarkably abundant in some *MT* promoters.[57] Our computational analysis revealed a preponderant abundance of putative HSF binding sites in the *OdMT1* and *OdMT2* promoters (Fig. 2B, Fig. S2 and Table S3, ESI†).

Overall, the promoter region of *OdMT1* and *OdMT2* genes did not appear to contain MRE sequences, which are a common feature of metazoans *MT* genes, but they had a MREF conserved motif, which is typical of fungal *MT* genes.[63,64] The presence of a MREF in the *OdMT* promoters was shared by the promoters of *O. dioica Cu/ZnSOD1* genes, and although the functionality of these sites needs to be experimentally demonstrated, they suggested a coordinated transcriptional activation of both (*MT* and *SOD*) stress-activated genes. *OdMT* promoters appeared, indeed, rich in stress-responsive elements, including binding sites for YAP1/CREB, RXR/PPAR and HSF transcription factors, envisaging a role of *MT* genes in adverse environmental conditions.

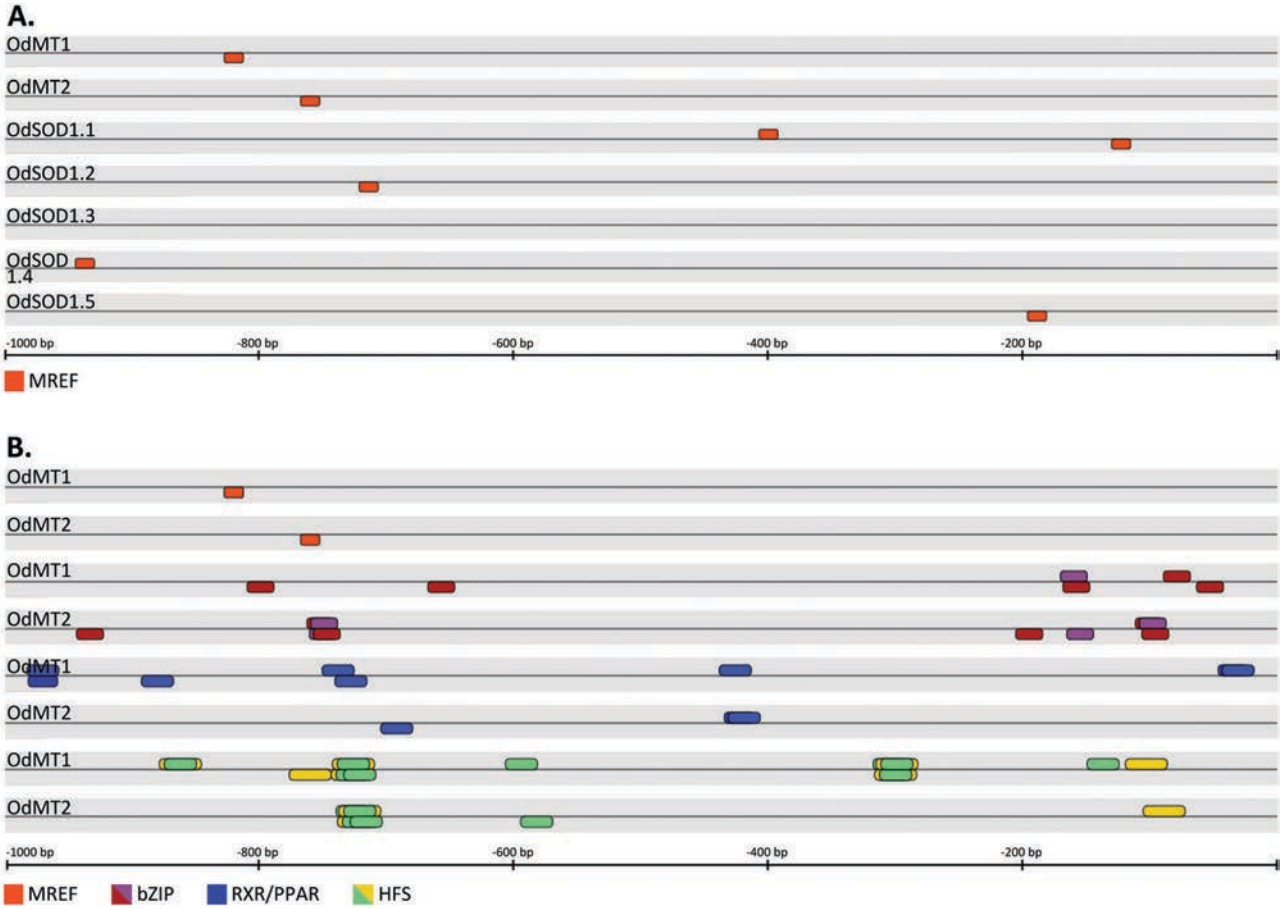


Fig. 2 Schematic representation of putative binding sites for transcription factors in the regulatory regions (1.0 kb upstream of CDS) of *O. dioica* genes. (A) MREF binding sites (orange boxes) were predicted in the *OdMT1* and *OdMT2* promoters as well as in the promoters of four out of five *OdSOD1* genes. (B) In addition of the MREF elements, several stress-response elements were predicted in *OdMT1* and *OdMT2* promoters, including those for bZIP transcription factors such as YAP1 (violet boxes) and CREB (purple boxes), for RXR/PPAR (blue boxes) and for animals or yeast HSF (green and yellow boxes, respectively). A box above the lines indicates that the binding site was in the plus DNA strand, whereas a box below indicates that it was in the minus strand (see Fig. S2 and Table S3 for additional details, ESI†).

3.3. Metal-binding capacity of *OdMT1* and *OdMT2*

In order to verify the MT nature of the two *O. dioica* MTs, and to evaluate their metal-binding affinity and capacity, we studied the formation of metal MT complexes of OdMT1 and OdMT2 proteins heterologously expressed in *E. coli* and grown in medium supplemented with cooper, cadmium or zinc salts. Metal OdMT1 and OdMT2 complexes were purified and analyzed by ICP-AES and ESI-MS. Acidification of the Zn OdMT complexes yielded the corresponding apo-forms, with molecular masses of 7789 Da for OdMT1 and 42 861 Da for OdMT2 (Fig. 3A and B), fully concordant with the calculated average theoretical values for the synthesized products (7789.10 Da and 42860.48 Da, respectively; notice that recombinant proteins had two additional amino acids at N-terminus). This confirmed both the identity and purity of the recombinant proteins.

The ICP-AES analyses showed that both recombinant OdMT1 and OdMT2 proteins could form metal complexes, principally with cadmium ions (9.7 and 55.0 Cd/protein ratios for OdMT1 and OdMT2, respectively; Table 1). ESI-MS analysis of the metal OdMT1 species formed allowed us to observe the formation of a variety of species in the productions in Zn- and Cu-enriched media: from Zn_4- to Zn_7 OdMT1 (Fig. 3C), and from Cu_8 to Cu_{14} OdMT1 (Fig. 3E), respectively. In contrast, the production in Cd-enriched media rendered a single Cd_7 OdMT1 species, indicative of the formation of a highly favored species (Fig. 3D). Both ICP-AES and ESI-MS results indicated, therefore, that OdMT1 exhibits a clear preference for coordinating Cd^{2+}. In agreement with the Cd preference, the yield of recombinant OdMT1 produced in Cd-media was higher than in Zn- or Cu-enriched media (Table 1) because metal binding might be contributing to stabilize recombinantly expressed protein. Regarding OdMT2, the high molecular weight of OdMT2 and its complex structure made of repeated modules may be responsible for the low concentration of the samples recovered from the metal OdMT2 productions, even in the presence of Cd, impairing to obtain valid ESI-MS data of the metallated species. However, ICP data (Table 1), together with the fact that the yield of OdMT2 production was higher in Cd-enriched

This journal is © The Royal Society of Chemistry 2018

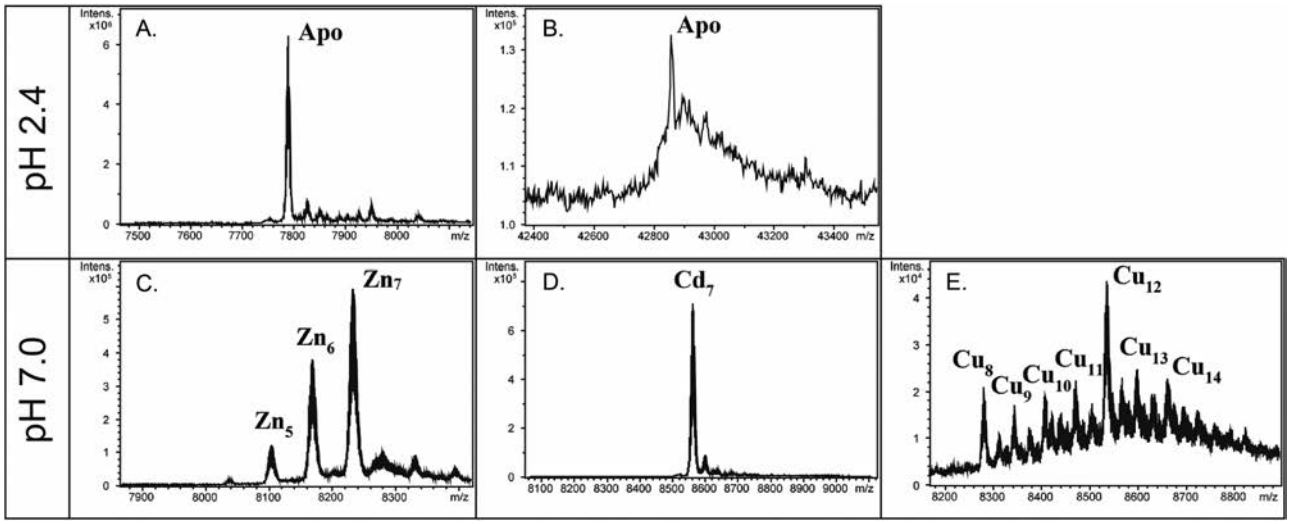


Fig. 3 Deconvoluted ESI-MS spectra of recombinant OdMT productions with different metal-enriched media. The apo-OdMT1 (A) and apo-OdMT2 (B) were recorded from the Zn- or Cd-productions, respectively, analyzed at acidic pH. Metal OdMT1 species were recorded from the Zn-, Cd- or Cu- productions (C, D and E, respectively) at neutral pH.

Table 1 Protein concentration and metal content in metal OdMT complexes

Metal OdMT complex	Protein concentration (μM)	Metal/protein ratio
Zn OdMT1	12	6.4 (Zn)
Cd OdMT1	34	9.7 (Cd)
Cu OdMT1	12	15.3 (Cu)
Zn OdMT2	<1	LDL[a]
Cd OdMT2	15	55.0 (Cd)
Cu OdMT2	<1	LDL

[a] LDL: lower than detection limit.

media than in Zn- and Cu-media (Table 1 and Fig. S3, ESI†) indicated a preference for Cd^{2+}, likewise it has been demonstrated for OdMT1. In fact, OdMT2 seems to bind up to 6 more times higher amounts of Cd than OdMT1 (Table 1) accordingly with its high Cys content.

4. Conclusions

In this work, we have identified two *MT* genes in the chordate *O. dioica* revealing a particular modular structure of the encoded proteins based on repeat units of two C7 subunits. The structure of these chordate MTs together with data from fungal[47,48] and protozoan *MT*s[76,77] suggest that during metallothionein evolution, diverse living beings with different *MT* sequences have used a similar modular and step-wise mechanism for generating long proteins with high metal-binding capacities. In addition, sequence analyses have revealed the presence of metal- and stress-response elements in the regulatory regions of both genes, while ICP-AES analyses have shown the capacity of both *O. dioica* proteins to form metal MT complexes, with preference for cadmium ions. These results support the MT nature of the *O. dioica* genes and proteins, and suggest that they might play a role in the physiological response of this animal against metal toxicity and other stress situations. Overall our data pave the way for better understanding the evolution and function of the detoxification mechanisms of zooplankton species, which might be crucial to counteract the increasing amounts of heavy metals in marine ecosystems.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

In memory of Silvia Atrian, for so many years sharing ideas, projects and meetings in the Departament de Genetica of the Universitat de Barcelona. Her love for metallothioneins inspired this work and prompted us to look for new metallothionein in our favorite animal model. We miss her knowledge and advices, always accurate and opportune. The authors are grateful to Merce Capdevila for her support and scientific advice, and for the opportunity to contribute to this special issue. We also thank all team members of the C. C. and R. A. laboratories for fruitful scientific discussions. We thank Josep Martı-Solans for the *Oikopleura dioica* picture. C. C. was supported by BFU2016-80601-P grant; O. P. was supported by BIO2015-67358-C2-2-P grant; and R. A. was supported by BIO2015-67358-C2-1-P grant from Ministerio de Economıa y Competitividad (Spain). C. C. and R. A. were also supported by grant SGR2017-1665 from Generalitat de Catalunya. M. G.-R. acknowledges to the UAB the PIF grant. We thank the Centres Cientıfics i Tecnologics (CCiT) de la Universitat de Barcelona (DNA sequencing) and the Servei d Analisi Quımica (SAQ) de la Universitat Autonoma de Barcelona (ICP-AES, CD, UV-vis, ESI-MS) for allocating instrument time.

References

1 M. Margoshes and B. L. Vallee, A cadmium protein from equine kidney cortex, *J. Am. Chem. Soc.*, 1957, **79**, 4813 4814.

2 M. Capdevila and S. Atrian, Metallothionein protein evolution: a miniassay, *JBIC, J. Biol. Inorg. Chem.*, 2011, **16**, 977 989.

3 J. Calvo, H. Jung and G. Meloni, Copper metallothioneins, *IUBMB Life*, 2017, **69**, 236 245.

4 N. Franchi, F. Boldrin, L. Ballarin and E. Piccinni, CiMT-1, an unusual chordate metallothionein gene in Ciona intestinalis genome: structure and expression studies, *J. Exp. Zool., Part A*, 2011, **315**, 90 100.

5 M. Guirola, S. Perez-Rafael, M. Capdevila, O. Palacios and S. Atrian, Metal dealing at the origin of the Chordata phylum: the metallothionein system and metal overload response in amphioxus, *PLoS One*, 2012, **7**, e43299.

6 M. Capdevila, R. Bofill, O. Palacios and S. Atrian, State-of-the-art of metallothioneins at the beginning of the 21st century, *Coord. Chem. Rev.*, 2012, **256**, 46 62.

7 J. S. Scheller, G. W. Irvine and M. J. Stillman, Unravelling the mechanistic details of metal binding to mammalian metallothioneins from stoichiometric, kinetic, and binding affinity data, *Dalton Trans.*, 2018, **47**, 3613 3637.

8 R. Bofill, M. Capdevila and S. Atrian, Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu-thioneins, *Metallomics*, 2009, **1**, 229 234.

9 O. Palacios, S. Atrian and M. Capdevila, Zn- and Cu-thioneins: a functional classification for metallothioneins?, *JBIC, J. Biol. Inorg. Chem.*, 2011, **16**, 991 1009.

10 S. R. Davis and R. J. Cousins, Metallothionein expression in animals: a physiological perspective on function, *J. Nutr.*, 2000, **130**, 1085 1088.

11 J. O. Nriagu and J. M. Pacyna, Quantitative assessment of worldwide contamination of air, water and soils by trace metals, *Nature*, 1988, **333**, 134 139.

12 S. Cheng, Heavy metal pollution in China: origin, pattern and control, *Environ. Sci. Pollut. Res.*, 2003, **10**, 192 198.

13 G. W. Bryan, The effects of heavy metals (other than mercury) on marine and estuarine organisms, *Proc. R. Soc. London, Ser. B*, 1971, **177**, 389 410.

14 G. Gorsky and R. Fenaux, in *The Biology of Pelagic Tunicates*, ed. Q. Bone, Oxford University Press, Oxford, 1998.

15 P. Davoll and M. Youngbluth, Heterotrophic activity on appendicularian (Tunicata: Appendicularia) houses in mesopelagic regions and their potential contribution to particle flux, *Deep-Sea Res.*, 1990, **37**, 285 294.

16 B. H. Robison, K. R. Reisenbichler and R. E. Sherlock, Giant larvacean houses: rapid carbon transport to the deep sea floor, *Science*, 2005, **308**, 1609 1611.

17 K. Katija, R. E. Sherlock, A. D. Sherman and B. H. Robison, New technology reveals the role of giant larvaceans in oceanic carbon cycling, *Sci. Adv.*, 2017, **3**, e1602374.

18 H.-C. Seo, M. Kube, R. B. Edvardsen, M. F. Jensen, A. Beck, E. Spriet, G. Gorsky, E. M. Thompson, H. Lehrach, R. Reinhardt and D. Chourrout, Miniature genome in the marine chordate *Oikopleura dioica*, *Science*, 2001, **294**, 2506.

19 F. Denoeud, S. Henriet, S. Mungpakdee, J. M. Aury, C. Da Silva, H. Brinkmann, J. Mikhaleva, L. C. Olsen, C. Jubin, C. Canestro, J. M. Bouquet, G. Danks, J. Poulain, C. Campsteijn, M. Adamski, I. Cross, F. Yadetie, M. Muffato, A. Louis, S. Butcher, G. Tsagkogeorga, A. Konrad, S. Singh, M. F. Jensen, E. H. Cong, H. Eikeseth-Otteraa, B. Noel, V. Anthouard, B. M. Porcel, R. Kachouri-Lafond, A. Nishino, M. Ugolini, P. Chourrout, H. Nishida, R. Aasland, S. Huzurbazar, E. Westhof, F. Delsuc, H. Lehrach, R. Reinhardt, J. Weissenbach, S. W. Roy, F. Artiguenave, J. H. Postlethwait, J. R. Manak, E. M. Thompson, O. Jaillon, L. Du Pasquier, P. Boudinot, D. A. Liberles, J. N. Volff, H. Philippe, B. Lenhard, H. Roest Crollius, P. Wincker and D. Chourrout, Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate, *Science*, 2010, **330**, 1381 1385.

20 J. Marti-Solans, O. V. Belyaeva, N. P. Torres-Aguila, N. Y. Kedishvili, R. Albalat and C. Canestro, Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate, *Mol. Biol. Evol.*, 2016, **33**, 2401 2416.

21 G. Danks, C. Campsteijn, M. Parida, S. Butcher, H. Doddapaneni, B. Fu, R. Petrin, R. Metpally, B. Lenhard, P. Wincker, D. Chourrout, E. M. Thompson and J. R. Manak, OikoBase: a genomics and developmental transcriptomics resource for the urochordate *Oikopleura dioica*, *Nucleic Acids Res.*, 2013, **41**, D845 853.

22 K. Wang, T. Omotezako, K. Kishi, H. Nishida and T. A. Onuma, Maternal and zygotic transcriptomes in the appendicularian, Oikopleura dioica: novel protein-encoding genes, intra-species sequence variations, and trans-spliced RNA leader, *Dev. Genes Evol.*, 2015, **225**, 149 159.

23 S. Bassham, PhD, University of Oregon, 2002.

24 H. C. Seo, R. B. Edvardsen, A. D. Maeland, M. Bjordal, M. F. Jensen, A. Hansen, M. Flaat, J. Weissenbach, H. Lehrach, P. Wincker, R. Reinhardt and D. Chourrout, Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica, *Nature*, 2004, **431**, 67 71.

25 S. Bassham and J. H. Postlethwait, The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate Oikopleura dioica, *Development*, 2005, **132**, 4259 4272.

26 C. Canestro, S. Bassham and J. H. Postlethwait, Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain, *Dev. Biol.*, 2005, **285**, 298 315.

27 A. Nishino, Y. Satou, M. Morisawa and N. Satoh, Brachyury (T) gene expression and notochord development in Oikopleura longicauda (Appendicularia, Urochordata), *Dev. Genes Evol.*, 2001, **211**, 219 231.

28 R. B. Edvardsen, H. C. Seo, M. F. Jensen, A. Mialon, J. Mikhaleva, M. Bjordal, J. Cartry, R. Reinhardt, J. Weissenbach, P. Wincker and D. Chourrout, Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica, *Curr. Biol.*, 2005, **15**, R12 R13.

This journal is © The Royal Society of Chemistry 2018

29 S. Fujii, T. Nishio and H. Nishida, Cleavage pattern, gastrulation, and neurulation in the appendicularian, Oikopleura dioica, *Dev. Genes Evol.*, 2008, **218**, 69 79.
30 C. Canestro, S. Bassham and J. H. Postlethwait, Evolution of the thyroid: anterior posterior regionalization of the Oikopleura endostyle revealed by Otx, Pax2/5/8, and Hox1 expression, *Dev. Dyn.*, 2008, **237**, 1490 1499.
31 C. Troedsson, J.-M. Bouquet, C. Lobon, A. Novac, J. Nejstgaard, S. Dupont, S. Bosak, H. Jakobsen, N. Romanova, L. Pankoke, A. Isla, J. R. Dutz, A. Sazhin and E. Thompson, Effects of ocean acidification, temperature and nutrient regimes on the appendicularian Oikopleura dioica: a mesocosm study, *Mar. Biol.*, 2013, **160**, 2175 2187.
32 F. Yadetie, S. Butcher, H. E. Forde, C. Campsteijn, J. M. Bouquet, O. A. Karlsen, F. Denoeud, R. Metpally, E. M. Thompson, J. R. Manak, A. Goksoyr and D. Chourrout, Conservation and divergence of chemical defense system in the tunicate *Oikopleura dioica* revealed by genome wide response to two xenobiotics, *BMC Genomics*, 2012, **13**, 55.
33 J. M. Bouquet, C. Troedsson, A. Novac, M. Reeve, A. K. Lechtenborger, W. Massart, K. S. Skaar, A. Aasjord, S. Dupont and E. M. Thompson, Increased fitness of a key appendicularian zooplankton species under warmer, acidified seawater conditions, *PLoS One*, 2018, **13**, e0190625.
34 N. P. Torres-Aguila, J. Martı-Solans, A. Ferrandez-Roldan, A. Almazan, V. Roncalli, S. D Aniello, G. Romano, A. Palumbo, R. Albalat and C. Canestro, Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate *Oikopleura dioica*, *Commun. Biol.*, 2018, **1**, 121.
35 J. M. Bouquet, E. Spriet, C. Troedsson, H. Ottera, D. Chourrout and E. M. Thompson, Culture optimization for the emergent zooplanktonic model organism Oikopleura dioica, *J. Plankton Res.*, 2009, **31**, 359 370.
36 H. Nishida, Development of the appendicularian *Oikopleura dioica*: culture, genome, and cell lineages, *Dev., Growth Differ.*, 2008, **50**(Suppl 1), S239 S256.
37 J. Marti-Solans, A. Ferrandez-Roldan, H. Godoy-Marin, J. Badia-Ramentol, N. P. Torres-Aguila, A. Rodriguez-Mari, J. M. Bouquet, D. Chourrout, E. M. Thompson, R. Albalat and C. Canestro, *Oikopleura dioica* culturing made easy: a low-cost facility for an emerging animal model in EvoDevo, *Genesis*, 2015, **53**, 183 193.
38 R. Albalat and C. Canestro, Evolution by gene loss, *Nat. Rev. Genet.*, 2016, **17**, 379 391.
39 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 1997, **25**, 3389 3402.
40 K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein and T. Werner, MatInspector and beyond: promoter analysis based on transcription factor binding sites, *Bioinformatics*, 2005, **21**, 2933 2942.
41 D. G. Gibson, Enzymatic assembly of overlapping DNA fragments, *Methods Enzymol.*, 2011, **498**, 349 361.
42 N. A. Shevchuk, A. V. Bryksin, Y. A. Nusinovich, F. C. Cabello, M. Sutherland and S. Ladisch, Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously, *Nucleic Acids Res.*, 2004, **32**, e19.
43 A. S. Xiong, Q. H. Yao, R. H. Peng, X. Li, H. Q. Fan, Z. M. Cheng and Y. Li, A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences, *Nucleic Acids Res.*, 2004, **32**, e98.
44 A. S. Xiong, Q. H. Yao, R. H. Peng, H. Duan, X. Li, H. Q. Fan, Z. M. Cheng and Y. Li, PCR-based accurate synthesis of long DNA sequences, *Nat. Protoc.*, 2006, **1**, 791 797.
45 J. Bongers, C. D. Walton, D. E. Richardson and J. U. Bell, Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma atomic emission spectrometric determination of sulfur, *Anal. Chem.*, 1988, **60**, 2683 2686.
46 D. Fabris, J. Zaia, Y. Hathout and C. Fenselau, Retention of Thiol Protons in Two Classes of Protein Zinc Ion Coordination Centers, *J. Am. Chem. Soc.*, 1996, **118**, 12242 12243.
47 O. Palacios, A. Espart, J. Espin, C. Ding, D. J. Thiele, S. Atrian and M. Capdevila, Full characterization of the Cu-, Zn-, and Cd-binding properties of CnMT1 and CnMT2, two metallothioneins of the pathogenic fungus Cryptococcus neoformans acting as virulence factors, *Metallomics*, 2014, **6**, 279 291.
48 P. Iturbe-Espinoza, S. Gil-Moreno, W. Lin, S. Calatayud, O. Palacios, M. Capdevila and S. Atrian, The Fungus Tremella mesenterica Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization, *PLoS One*, 2016, **11**, e0148651.
49 O. Palacios, E. Jimenez-Marti, M. Niederwanger, S. Gil-Moreno, O. Zerbe, S. Atrian, R. Dallinger and M. Capdevila, Analysis of Metal-Binding Features of the Wild Type and Two Domain-Truncated Mutant Variants of Littorina littorea Metallothionein Reveals Its Cd-Specific Character, *Int. J. Mol. Sci.*, 2017, **18**, 1452.
50 C. Baumann, A. Beil, S. Jurt, M. Niederwanger, O. Palacios, M. Capdevila, S. Atrian, R. Dallinger and O. Zerbe, Structural Adaptation of a Protein to Increased Metal Stress: NMR Structure of a Marine Snail Metallothionein with an Additional Domain, *Angew. Chem., Int. Ed.*, 2017, **56**, 4617 4622.
51 L. Schmielau, M. Dvorak, M. Niederwanger, N. Dobieszewski, V. Pedrini-Martha, P. Ladurner, J. R. Pedregal, J. D. Marechal and R. Dallinger, Differential response to Cadmium exposure by expression of a two and a three-domain metallothionein isoform in the land winkle *Pomatias elegans*: valuating the marine heritage of a land snail, *Sci. Total Environ.*, 2018, **648**, 561 571.
52 M. J. Jenny, S. L. Payton, D. A. Baltzegar and J. D. Lozier, Phylogenetic Analysis of Molluscan Metallothioneins: Evolutionary Insight from Crassostrea virginica, *J. Mol. Evol.*, 2016, **83**, 110 125.
53 R. Dallinger, *Gastropod metallothioneins in evolution: new rules for an old protein family*, Universitat Innsbruck, 2018.

54 R. Heuchel, F. Radtke, O. Georgiev, G. Stark, M. Aguet and W. Schaffner, The transcription factor MTF-1 is essential for basal and heavy metal-induced metallothionein gene expression, *EMBO J.*, 1994, **13**, 2870 2875.
55 P. Silar, L. Theodore, R. Mokdad, N. E. Erraiss, A. Cadic and M. Wegnez, Metallothionein Mto gene of Drosophila melanogaster: structure and regulation, *J. Mol. Biol.*, 1990, **215**, 217 224.
56 M. Egg, M. Hockner, A. Brandstatter, D. Schuler and R. Dallinger, Structural and bioinformatic analysis of the Roman snail Cd-Metallothionein gene uncovers molecular adaptation towards plasticity in coping with multifarious environmental stress, *Mol. Ecol.*, 2009, **18**, 2426 2443.
57 M. Niederwanger, M. Dvorak, R. Schnegg, V. Pedrini-Martha, K. Bacher, M. Bidoli and R. Dallinger, Challenging the Metallothionein (MT) Gene of Biomphalaria glabrata: Unexpected Response Patterns Due to Cadmium Exposure and Temperature Stress, *Int. J. Mol. Sci.*, 2017, **18**, 1747.
58 M. Zafarullah, K. Bonham and L. Gedamu, Structure of the rainbow trout metallothionein B gene and characterization of its metal-responsive region, *Mol. Cell. Biol.*, 1988, **8**, 4469 4476.
59 A. Auf der Maur, T. Belser, G. Elgar, O. Georgiev and W. Schaffner, Characterization of the transcription factor MTF-1 from the Japanese pufferfish (*Fugu rubripes*) reveals evolutionary conservation of heavy metal stress response, *Biol. Chem.*, 1999, **380**, 175 185.
60 F. Radtke, M. Hug, O. Georgiev, K. Matsuo and W. Schaffner, Differential sensitivity of zinc finger transcription factors MTF-1, Sp1 and Krox-20 to CpG methylation of their binding sites, *Biol. Chem. Hoppe-Seyler*, 1996, **377**, 47 56.
61 G. Westin and W. Schaffner, A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-I gene, *EMBO J.*, 1988, **7**, 3763 3770.
62 D. P. Giedroc, X. Chen and J. L. Apuy, Metal response element (MRE)-binding transcription factor-1 (MTF-1): structure, function, and regulation, *Antioxid. Redox Signaling*, 2001, **3**, 577 596.
63 D. J. Thiele, ACE1 regulates expression of the Saccharomyces cerevisiae metallothionein gene, *Mol. Cell. Biol.*, 1988, **8**, 2745 2752.
64 P. Zhou, M. S. Szczypka, T. Sosinowski and D. J. Thiele, Expression of a yeast metallothionein gene family is activated by a single metalloregulatory transcription factor, *Mol. Cell. Biol.*, 1992, **12**, 3766 3775.
65 E. B. Gralla, D. J. Thiele, P. Silar and J. S. Valentine, ACE1, a copper-dependent transcription factor, activates expression of the yeast copper, zinc superoxide dismutase gene, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 8558 8562.
66 S. M. Ruzsa and J. G. Scandalios, Altered Cu metabolism and differential transcription of Cu/ZnSod genes in a Cu/ZnSOD-deficient mutant of maize: evidence for a Cu-responsive transcription factor, *Biochemistry*, 2003, **42**, 1508 1516.
67 J. Xu, Y. S. Tian, R. H. Peng, A. S. Xiong, B. Zhu, X. F. Jin, J. J. Gao, X. L. Hou and Q. H. Yao, Yeast copper-dependent transcription factor ACE1 enhanced copper stress tolerance in Arabidopsis, *BMB Rep.*, 2009, **42**, 752 757.
68 A. Wu, J. A. Wemmie, N. P. Edgington, M. Goebl, J. L. Guevara and W. S. Moye-Rowley, Yeast bZip proteins mediate pleiotropic drug and metal resistance, *J. Biol. Chem.*, 1993, **268**, 18850 18858.
69 R. K. Uddin and S. M. Singh, Ethanol-responsive genes: identification of transcription factors and their role in metabolomics, *Pharmacogenomics J.*, 2007, **7**, 38 47.
70 M. Hockner, R. Dallinger and S. R. Sturzenbaum, Metallothionein gene activation in the earthworm (Lumbricus rubellus), *Biochem. Biophys. Res. Commun.*, 2015, **460**, 537 542.
71 V. Drechsel, K. Schauer, M. Srut and M. Hockner, Regulatory Plasticity of Earthworm wMT-2 Gene Expression, *Int. J. Mol. Sci.*, 2017, **18**, 1113.
72 M. Karin, A. Haslinger, H. Holtgreve, R. I. Richards, P. Krauter, H. M. Westphal and M. Beato, Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-IIA gene, *Nature*, 1984, **308**, 513 519.
73 P. Martinez-Paz, M. Morales, P. Sanchez-Arguello, G. Morcillo and J. L. Martinez-Guitarte, Cadmium *in vivo* exposure alters stress response and endocrine-related genes in the freshwater snail Physa acuta. New biomarker genes in a new model organism, *Environ. Pollut.*, 2017, **220**, 1488 1497.
74 K. T. Tamai, X. Liu, P. Silar, T. Sosinowski and D. J. Thiele, Heat shock transcription factor activates yeast metallothionein gene expression in response to heat and glucose starvation *via* distinct signalling pathways, *Mol. Cell. Biol.*, 1994, **14**, 8155 8165.
75 W. M. Yang, W. Gahl and D. Hamer, Role of heat shock transcription factor in yeast metallothionein gene expression, *Mol. Cell. Biol.*, 1991, **11**, 3676 3681.
76 S. Diaz, F. Amaro, D. Rico, V. Campos, L. Benitez, A. Martin-Gonzalez, E. P. Hamilton, E. Orias and J. C. Gutierrez, Tetrahymena metallothioneins fall into two discrete subfamilies, *PLoS One*, 2007, **2**, e291.
77 A. Espart, M. Marin, S. Gil-Moreno, O. Palacios, F. Amaro, A. Martin-Gonzalez, J. C. Gutierrez, M. Capdevila and S. Atrian, Hints for metal-preference protein sequence determinants: different metal binding features of the five tetrahymena thermophila metallothioneins, *Int. J. Biol. Sci.*, 2015, **11**, 456 471.
78 G. E. Crooks, G. Hon, J. M. Chandonia and S. E. Brenner, WebLogo: a sequence logo generator, *Genome Res.*, 2004, **14**, 1188 1190.

This journal is © The Royal Society of Chemistry 2018

ARTICLE IV: SUPPLEMENATRY MATERIAL

METALLOTHIONEINS OF THE UROCHORDATE OIKOPLEURA DIOICA HAVE CYS-RICH TANDEM REPEATS, LARGE SIZE AND CADMIUM-BINDING PREFERENCE

The file includes:

Figure S1

Figure S2

Figure S3

Table S1

Table S2

Table S3

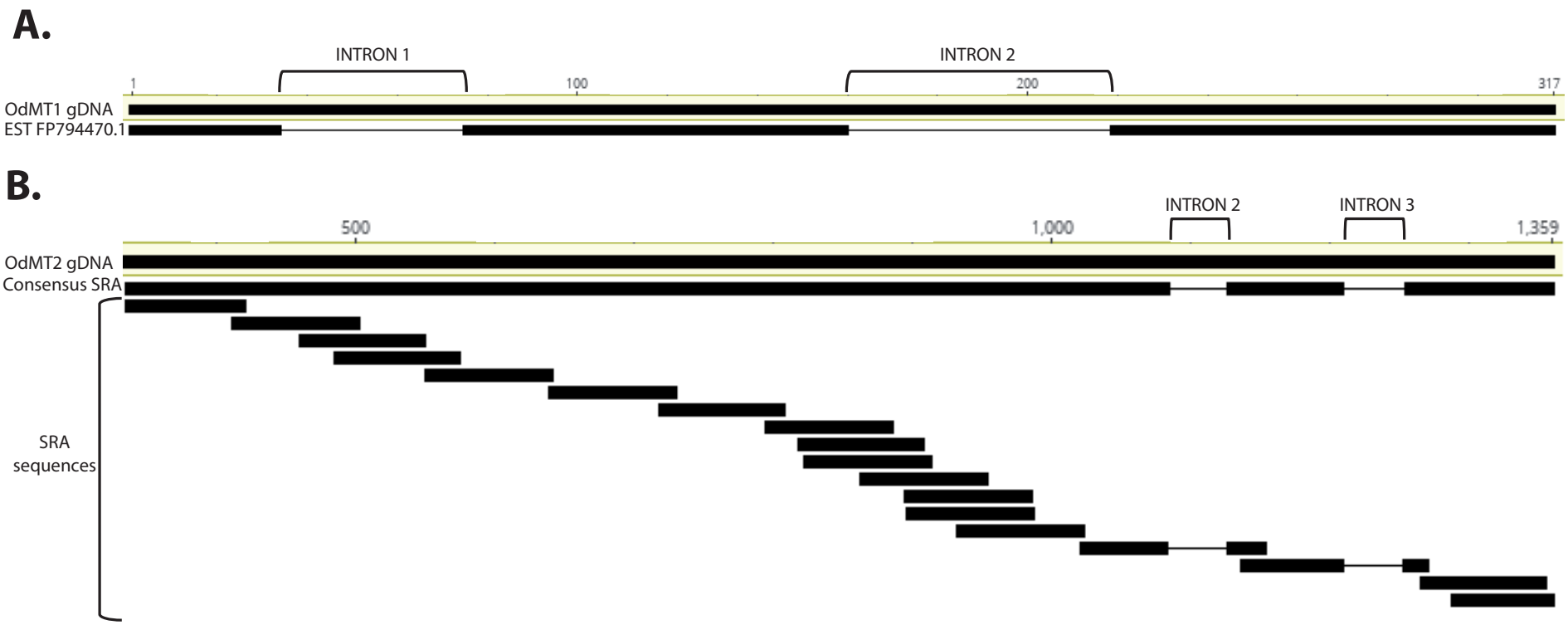


Figure S1. Annotation of *OdMT1* and *OdMT2* genes were supported by EST and SRA sequences. (**A**) *OdMT1* gene was annotated as 'GSOIDP00013076001-unnamed product' in the Oikobase, scaffold_50 from 61,593 nt to 61,997 nt (plus strand). Comparison of the genomic sequence (top line) with the EST FP794470.1 (bottom line) supported the annotated 3 exons / 2 introns structure. (**B**) *OdMT2* gene was not annotated in the Oikobase, scaffold_16: from 829,667 nt to 831,025 nt (minus strand). *OdMT2* gene structure was deduced by comparison of the genomic sequence (top line) with a consensus cDNA (bottom line) reconstructed from 18 SRA sequences (short lines; **Table S1**) covering 75% of the gene, from position 335 (exon 2) to 1359 (exon 4). Exon 1 and the 5' end of exon 2 were deduced by comparison with *OdMT1* and manual inspection of the ORFs.

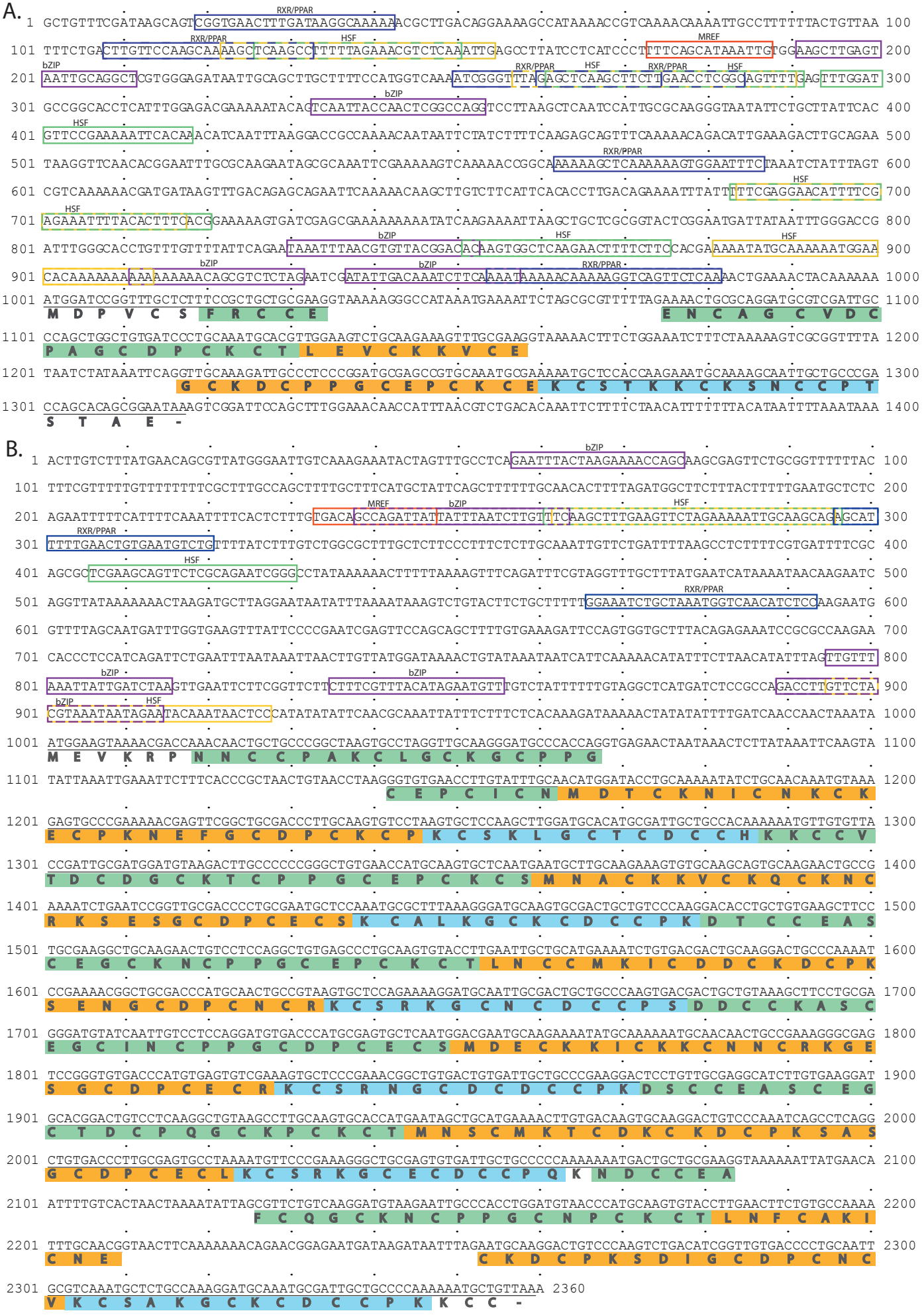

A.
RXR/PPAR
HSF
MREF
bZIP
RXR/PPAR
HSF
RXR/PPAR
HSF
bZIP
RXR/PPAR
HSF
HSF
HSF
bZIP
HSF
HSF
bZIP
bZIP
RXR/PPAR
M D P V C S F R C C E
E N C A G C V D C
P A G C D P C K C T L E V C K K V C E
G C K D C P P G C E P C K C E K C S T K K C K S N C C P T
S T A E -
B.
bZIP
MREF
bZIP
HSF
RXR/PPAR
RXR/PPAR
HSF
bZIP
bZIP
bZIP
HSF
M E V K R P N N C C P A K C L G C K G C P P G
C E P C I C N M D T C K N I C N K C K
E C P K N E F G C D P C K C P K C S K L G C T C D C C H K K C C V
T D C D G C K T C P P G C E P C K C S M N A C K K V C K Q C K N C
R K S E S G C D P C E C S K C A L K G C K C D C C P K D T C C E A S
C E G C K N C P P G C E P C K C T L N C C M K I C D D C K D C P K
S E N G C D P C N C R K C S R K G C N C D C C P S D D C C K A S C
E G C I N C P P G C D P C E C S M D E C K K I C K K C N N C R K G E
S G C D P C E C R K C S R N G C D C D C C P K D S C C E A S C E G
C T D C P Q G C K P C K C T M N S C M K T C D K C K D C P K S A S
G C D P C E C L K C S R K G C E C D C C P Q K N D C C E A
F C Q G C K N C P P G C N P C K C T L N F C A K I
C N E C K D C P K S D I G C D P C N C
V K C S A K G C K C D C C P K K C C -

Figure S2. *OdMTs* genomic regions. (**A**) Genomic region of *OdMT1* gene along with 1 Kb of the 5' flanking promoter region and *in silico* translation of CDS. Deduced OdMT1 protein is made of a single repeat unit (RU) with an internal repeat –C7a (green background) and C7b (orange background)– and a C-terminal tail (blue background). (**B**) Genomic region of *OdMT2* gene along with 1 Kb of the 5' flanking promoter region and *in silico* translation of CDS. Deduced OdMT2 protein is made of 6 RUs, each one with the same C7a/C7b/C-tail structure (green, blue and orange backgrounds, respectively). Putative binding sites for transcription factors in the promoter regions of both genes are depicted with the same colour code than in Figure 2: MREF (orange boxes), bZIP transcription factors (purple boxes), RXR/PPAR (blue boxes) and animal or yeast HSF (green and yellow boxes, respectively). Notice that some boxes overlap (see **Table S3** for additional details).

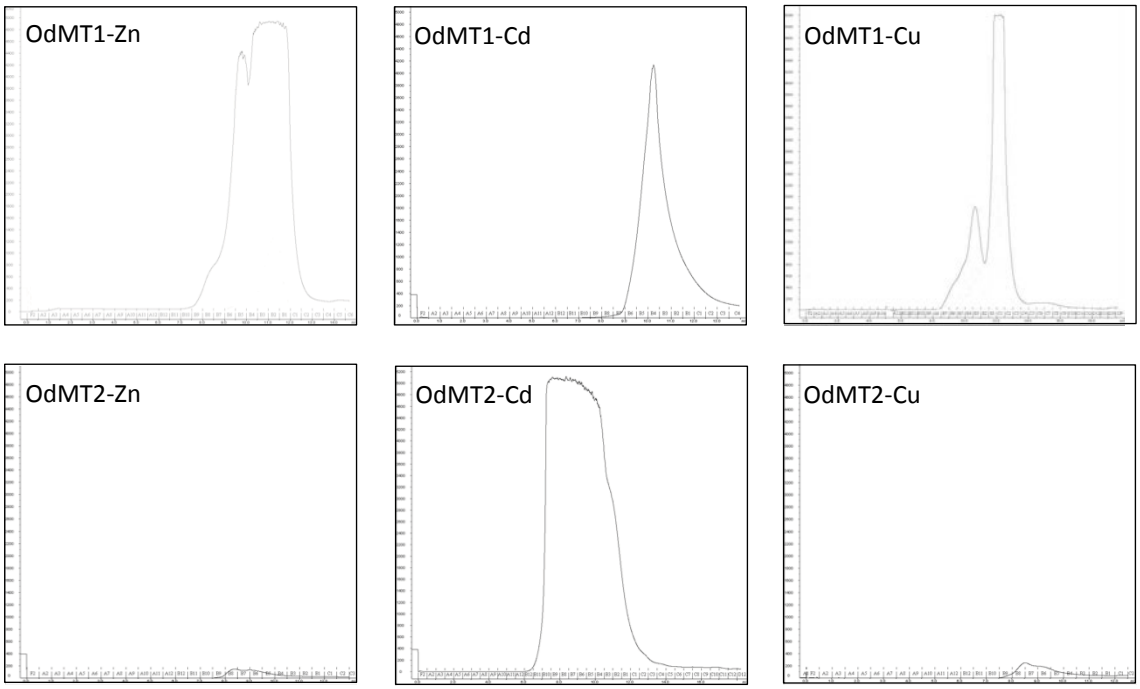


Figure S3. FPLC elution profiles at 254 nm of recombinant metal-OdMT1 (top panels) and metal-OdMT2 (low panels) complexes with Zn, Cd and Cu. Notice that whereas metal-OdMT1 complexes were obtained from *E. coli* cultures grown in medium supplemented with Zn, Cd or Cu, metal-OdMT2 complexes were only obtained from cultures grown in Cd-enriched media (central low panel). The intensity range was kept at the same values for better comparison among the chromatograms.

Table S1. Accession numbers of SRA sequences used for *OdMT2* cDNA reconstruction

SRR1693766.108467.2
SRR1693766.3371994.1
SRR1693766.3441030.1
SRR1693766.3698004.1
SRR1693766.4545152.1
SRR1693766.4545152.2
SRR1693766.4883353.1
SRR1693766.4883353.2
SRR1693766.6889256.2
SRR1693766.7376164.1
SRR1693767.211882.1
SRR1693767.461625.1
SRR1693767.461625.2
SRR1693767.5982171.1
SRR1693767.7626983.2
SRR1693767.7849746.1
SRR1693767.8124551.1
SRR1693767.11209530.1

Table S2. Families of transcription factors involved in stress response based on Matrix Family Library Version 11.0 from MatInspector program analyzed in this work.

FAMILY	FAMILY INFORMATION	MATRIX NAME	INFORMATION
F$ASG1	Activator of stress genes	F$ASG1 .01	Fungal zinc cluster transcription factor Asg1
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance
		F$CST6.01	Chromosome stability, bZIP transcription factor of the ATF/CREB family (ACA2)
		F$HAC1.01	bZIP transcription factor (ATF/CREB1 homolog) that regulates the unfolded protein response
		F$HAC1.02	bZIP transcription factor (ATF/CREB1 homolog) that regulates the unfolded protein response
		F$YAP1.01	Yeast activator protein of the basic leucine zipper (bZIP) family
		F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family
F$MREF	Metal regulatory element factors	F$CUSE.01	Copper-signaling element, AMT1/ACE1 recognition sequence
F$SKN7	Skn7 response regulator of S. cerevisiae	F$SKN7.01	SKN7, a transcription factor contributing to the oxidative stress response
F$SXBP	S.cerevisiae, XhoI site-binding protein I	F$XBP1.01	S.cerevisae XhoI site-binding protein I, stressinduced expression
		F$XBP1.02	Stress-induced transcriptional repressor
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
		F$HSF1.01	Trimeric heat shock transcription factor
		F$HSF1.02	Trimeric heat shock transcription factor
		F$MGA1.01	Heat shock transcription factor Mga1
F$YNIT	Asperg./Neurospora-activ. of genes induced by nitrogen	F$NIT2.01	Activator of nitrogen-regulated genes
F$YSTR	Yeast stress response elements	F$GIS1.01	Glg1-2 suppressor, involved in the expression of genes during nutrient limitation, JmjC domain-containing histone demethylase
		F$MSN2.01	Transcriptional activator for genes in multistress response
		F$MSN4.01	Transcriptional activator related to Msn2p, activated in stress conditions
		F$RPH1.01	Jumonji-like transcription factor
		F$STRE.01	MSN2/MSN4, STRE element, S. cerevisiae
		F$STRE.02	Stress-response element
		F$USV1.01	Usv (Up in StarVation), Nsf1 (nutrient and stress factor 1)
I$CEBP	Drosophila C/EBP like bZIP transcription factors	I$SLBO.01	Slow border cells
I$DCF1	Drosophila Chorion Factor 1 /Ultraspiracle	I$CF1.01	Chorion factor 1, homologous to retinoid X receptor
I$DHSF	Drosophila heat shock factors	I$HSF.01	Heat shock factor (Drosophila)
P$AREF	Auxin response element	P$ARE.01	Auxin Response Element
		P$SEBF.01	Silencing element binding factor - transcriptional repressor
P$ARF3	Auxin Response Factor 3	P$ETT.01	ETTIN (Auxin Response Factor 3)
		P$ETT.02	ETTIN (secondary DNA binding preference)
P$ERSE	ER stress-response elements	P$ERSE_I.01	ERSE I (ER stress-response element I)-like motif
P$HEAT	Heat shock factors	P$AT3G09735.01	DNA-binding protein S1FA3
		P$HSE.01	Heat shock element
		P$HSFA1A.01	Arabidopsis thaliana class A heat shock factor 1a
		P$HSFA1B.01	Heat stress transcription factor A-1b (HSF3)
		P$HSFA4A.01	Heat stress transcription factor A-4a (HSF21)
		P$HSFA6B.01	Heat stress transcription factor A-6b
		P$HSFB2A.01	Heat shock transcription factor B2A
		P$HSFB2A.02	Heat stress transcription factor B-2a (HSF6)
		P$HSFB2B.01	Heat stress transcription factor B-2b (HSF7)

	P$HSFB3.01	Heat stress transcription factor B-3
	P$HSFC1.01	Heat shock transcription factor C1
	P$HSFC1.02	Heat shock transcription factor C1 (secondary DNA binding preference)
	P$HSFC1.03	Heat stress transcription factor C-1
V$AHRR AHR-arnt heterodimers and AHR-related factors	V$AHR.01	Aryl hydrocarbon / dioxin receptor
	V$AHRARNT.01	Aryl hydrocarbon receptor / Arnt heterodimers
	V$AHRARNT.02	Aryl hydrocarbon / Arnt heterodimers, fixed core
	V$AHRARNT.03	DRE (dioxin response elements), XRE (xenobiotic response elements) bound by AHR/ARNT heterodimers
	V$NXF_ARNT.01	bHLH-PAS type transcription factors NXF/ARNT heterodimer
V$AP1R MAF and AP1 related factors	V$BACH1.01	BTB/POZ-bZIP transcription factor BACH1 forms heterodimers with the small Maf protein family
	V$BACH1.02	BTB and CNC homology 1, basic leucine zipper transcription factor 1
	V$BACH2.01	Bach2 bound TRE
	V$BACH2.02	BTB and CNC homology 1, basic leucine zipper transcription factor 2
	V$MAFA.01	Lens-specific Maf/MafA-sites
	V$MAFB.01	MAFB/Kreisler basic region/leucine zipper transcription factor (half site)
	V$MAFF.01	Transcription factor MafF
	V$MAFK.01	V-maf musculoaponeurotic fibrosarcoma oncogene homolog K (half site)
	V$MAFK.02	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog K (NFE2U)
	V$MARE.01	Maf response elements, half sites
	V$MARE.02	Binding sites for homodimers of large Maf-proteins
	V$MARE.03	Binding sites for heterodimers with small Maf-proteins
	V$MARE_ARE.01	Antioxidant response elements
	V$NFE2.01	NF-E2 p45
	V$NFE2.02	Leucine zipper protein NF-E2 (nuclear factor, erythroid-derived)
	V$NFE2L2.01	Nuclear factor, erythroid 2-like 2 (NFE2L2)
	V$NRL.01	Neural retinal basic leucine zipper factor (bZIP)
	V$NRL.02	Neural retinal basic leucine zipper factor (bZIP)
	V$TCF11MAFG.01	TCF11/MafG heterodimers, binding to subclass of AP1 sites
	V$VMAF.01	v-Maf
V$CEBP Ccaat/Enhancer Binding Protein	V$CEBP.02	CCAAT/enhancer binding protein
	V$CEBPA.01	CCAAT/enhancer binding protein alpha
	V$CEBPA.02	CCAAT/enhancer binding protein alpha
	V$CEBPB.01	CCAAT/enhancer binding protein beta
	V$CEBPB.02	CCAAT/enhancer binding protein beta
	V$CEBPD.01	CCAAT/enhancer binding protein (C/EBP), delta
	V$CEBPE.01	CCAAT/enhancer binding protein (C/EBP), epsilon
	V$CEBPE.02	CCAAT/enhancer binding protein (C/EBP), epsilon
	V$CEBPG.01	CCAAT/enhancer binding protein (C/EBP), gamma
V$CHOP C/EBP homologous protein (CHOP)	V$CHOP.01	Heterodimers of CHOP and C/EBPalpha
	V$CHOP.02	Heterodimers of CHOP and C/EBPalpha

V$CREB cAMP-responsive element binding proteins	V$ATF.01	Activating transcription factor
	V$ATF.02	Activating transcription factor
	V$ATF1.01	Activating transcription factor 1
	V$ATF1.02	Activating transcription factor 1
	V$ATF2.01	Activating transcription factor 2
	V$ATF6.02	Activating transcription factor 6, member of b-zip family, induced by ER stress
	V$BATF3.01	Basic leucine zipper transcription factor, ATF-like 3
	V$CJUN_ATF2.01	c-Jun/ATF2 heterodimers
	V$CREB.02	cAMP-responsive element binding protein
	V$CREB.03	cAMP-response element-binding protein
	V$CREB1.01	cAMP-responsive element binding protein 1
	V$CREB1.02	cAMP-responsive element binding protein 1
	V$CREB2.01	cAMP-responsive element binding protein 2
	V$CREB2CJUN.01	CRE-binding protein 2/c-Jun heterodimer
	V$CREB3.01	cAMP responsive element binding protein 3, dimeric binding site
	V$CREB3L2.01	cAMP responsive element binding protein 3-like 2, BBF2H7, dimeric binding site
	V$E4BP4.01	E4BP4, bZIP domain, transcriptional repressor
	V$JUNDM2.01	Jun dimerization protein 2
	V$TAXCREB.01	Tax/CREB complex
	V$TAXCREB.02	Tax/CREB complex
	V$XBP1.01	X-box-binding protein 1
V$EBOX E-box binding factors	V$ATF6.01	Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y
	V$CMYC.01	Myelocytomatosis oncogene (c-myc proto-oncogene)
	V$CMYC.02	Myelocytomatosis oncogene (c-myc proto-oncogene)
	V$MAX.01	Max/Max dimer
	V$MAX.02	MYC associated factor X
	V$MAX.03	MYC associated factor X
	V$MNT.01	MAX binding protein
	V$MYCMAX.01	c-Myc/Max heterodimer
	V$MYCMAX.02	c-Myc/Max heterodimer
	V$MYCMAX.03	MYC-MAX binding sites
	V$NMYC.01	N-Myc
	V$NMYC.02	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived
	V$USF.01	Upstream stimulating factor
	V$USF.02	Upstream stimulating factor
	V$USF.03	Upstream stimulating factor
	V$USF.04	Upstream stimulating factor 1/2
	V$USF1.01	Upstream stimulating factor 1
	V$USF1.02	Upstream stimulating factor 1
	V$USF2.01	Upstream transcription factor 2, c-fos interacting
V$FXRE Farnesoid X - activated receptor response elements	V$FXRE.01	Farnesoid X - activated receptor (RXR/FXR dimer), IR1 sites

V$GMEB	Glucocorticoid modulatory element binding proteins	V$GMEB2.01	Glucocorticoid modulatory element binding protein
V$GREF	Glucocorticoid responsive and related elements	V$ARE.01	Androgene receptor binding site, IR3 sites
		V$ARE.02	Androgene receptor binding site, IR3 sites
		V$ARE.03	Androgene receptor binding site, IR3 sites
		V$ARE.04	Androgene receptor binding site, IR3 sites
		V$GRE.01	Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs, IR3 sites
		V$GRE.02	Glucocorticoid receptor, IR3 sites
		V$GRE.03	Glucocorticoid receptor, IR3 sites
		V$MRE.01	Mineralocorticoid receptor response element
		V$PRE.01	Progesterone receptor binding site, IR3 sites
V$HEAT	Heat shock factors	V$HSF1.01	Heat shock factor 1
		V$HSF1.02	Heat shock factor 1
		V$HSF1.03	Heat shock factor 1
		V$HSF1.04	Heat shock factor 1
		V$HSF1.05	Heat shock factor 1
		V$HSF2.01	Heat shock factor 2
		V$HSF2.02	Heat shock factor 2
		V$HSF2.03	Heat shock factor 2
V$IRFF	Interferon regulatory factors	V$IRF1.01	Interferon regulatory factor 1
		V$IRF2.01	Interferon regulatory factor 2
		V$IRF2.02	Interferon regulatory factor 2
		V$IRF3.01	Interferon regulatory factor 3 (IRF-3)
		V$IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)
		V$IRF4.02	Interferon regulatory factor 4
		V$IRF4.03	Interferon regulatory factor 4
		V$IRF5.01	Interferon regulatory factor 5
		V$IRF6.01	Interferon regulatory factor 6
		V$IRF7.01	Interferon regulatory factor 7 (IRF-7)
		V$IRF8.01	Interferon consensus sequence binding protein 1, interferon regulatory factor 8, dimeric binding site
		V$ISGF3G.01	Interferon-stimulated transcription factor 3, gamma (IRF9)
		V$ISRE.01	Interferon-stimulated response element
		V$ISRE.02	Interferon-stimulated response element
V$MTF1	Metal induced transcription factor	V$MTF-1.01	Metal transcription factor 1, MRE
		V$MTF-1.02	Metal-regulatory transcription factor 1
V$NFAT	Nuclear factor of activated T-cells	V$NFAT.01	Nuclear factor of activated T-cells
		V$NFAT5.01	Nuclear factor of activated T-cells 5
		V$NFAT5.02	Nuclear factor of activated T-cells 5
		V$NFATC1.01	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1, dimeric binding site
V$NGRE	"Negative" glucocoticoid response elements	V$IR1_NGRE.01	Repressive binding sites for glucocorticoid receptor (IR1)
		V$IR2_NGRE.01	Repressive binding sites for glucocorticoid receptor (IR2)
		V$PPARA.01	PPAR/RXR heterodimers, DR1 sites

V$PERO	Peroxisome proliferator-activated receptor	V$PPARG.02	Peroxisome proliferator-activated receptor gamma
		V$PPARG.03	Peroxisome proliferator-activated receptor gamma, DR1 sites
		V$PPAR_RXR.01	PPAR/RXR heterodimers, DR1 sites
		V$PPAR_RXR.02	PPAR/RXR heterodimers, DR1 sites
V$PPAR	Peroxisome proliferator activated receptor homodimers	V$PPARG.01	Pal3 motif, bound by a PPAR-gamma homodimer, IR3 sites
V$RORA	v-ERB and RAR-related orphan receptor alpha	V$REV-ERBA.01	Orphan nuclear receptor rev-erb alpha (NR1D1)
		V$REV-ERBA.02	Orphan nuclear receptor rev-erb alpha (NR1D1), monomer binding site
		V$REV-ERBA.03	Orphan nuclear receptor rev-erb alpha (NR1D1), homodimer DR2 binding site
		V$RORA.01	RAR-related orphan receptor alpha
		V$RORA.02	RAR-related orphan receptor alpha, homodimer DR5 binding site
		V$RORA.03	RAR-related orphan receptor alpha, homodimer DR2 binding site
		V$RORA1.01	RAR-related orphan receptor alpha1
		V$RORA2.01	RAR-related orphan receptor alpha2
		V$RORGAMMA.01	RAR-related orphan receptor gamma
		V$VERBA.01	vErbA, viral homolog of thyroid hormone receptor alpha1
V$RXRF	RXR heterodimer binding sites	V$CAR_RXR.01	Constitutive androstane receptor / retinoid X receptor heterodimer, DR4 sites
		V$LXRE.01	Nuclear receptor involved in the regulation of lipid homeostasis, DR4 element
		V$LXRE.02	Highly conserved DR1 element selected by LXRbeta/RXR heterodimers
		V$PXR_RXR.01	Pregnane X receptor / retinoid X receptor heterodimer, DR4 sites
		V$RARA.01	Retinoic acid receptor alpha, homodimer DR4 binding site
		V$RARG.01	Retinoic acid receptor gamma, homodimer DR2 binding site
		V$RAR_RXR.01	Retinoic acid receptor / retinoid X receptor heterodimer, DR1 sites
		V$RAR_RXR.02	Retinoic acid receptor / retinoid X receptor heterodimer, DR2 sites
		V$RAR_RXR.03	Retinoic acid receptor / retinoid X receptor heterodimer, DR5 sites
		V$RXRA.01	Retinoid X receptor alpha homodimer, DR1 sites
		V$RXR_RXR.01	Retinoid X receptor homodimer, DR1 sites
		V$THR.01	THR/RXR heterodimer and THR homodimer DR4 binding sites
		V$THRA.01	Thyroid hormone receptor, alpha (ER4 - everted repeat, spacer 4)
		V$THRB.01	Thyroid hormone receptor, beta (ER4 - everted repeat, spacer 4)
		V$THRB.02	Thyroid hormone receptor, beta (ER5 - everted repeat, spacer 5)
		V$THRB.03	Thyroid hormone receptor, beta (ER6 - everted repeat, spacer 6)
		V$VDR_RXR.01	VDR/RXR Vitamin D receptor RXR heterodimer, DR3 sites
		V$VDR_RXR.03	Bipartite binding site of VDR/RXR heterodimers, DR1 sites
		V$VDR_RXR.04	Bipartite binding site of VDR/RXR heterodimers, DR3 sites
		V$VDR_RXR.05	Bipartite binding site of VDR/RXR heterodimers, DR4 sites
		V$VDR_RXR.06	Bipartite binding site of VDR/RXR heterodimers, DR5 sites

Table S3. Putative binding sites for transcription factors involved in stress response in *OdMT1* and *OdMT2* regulatory regions[1].

OdMT1			
Matrix Family	**Detailed Family Information**	**Matrix**	**Detailed Matrix Information**
F$YNIT	Asperg./Neurospora-activ. of genes induced by nitrogen	F$NIT2.01	Activator of nitrogen-regulated genes
V$PERO	Peroxisome proliferator-activated receptor	V$PPARG.03	Peroxisome proliferator-activated receptor gamma, DR1 sites
V$RXRF	RXR heterodimer binding sites	V$THRB.03	Thyroid hormone receptor, beta (ER6 - everted repeat, spacer 6)
V$NFAT	Nuclear factor of activated T-cells	V$NFATC1.01	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1, dimeric binding site
V$RXRF	RXR heterodimer binding sites	V$RARG.01	Retinoic acid receptor gamma, homodimer DR2 binding site
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFB3.01	Heat stress transcription factor B-3
V$HEAT	Heat shock factors	V$HSF2.02	Heat shock factor 2
F$MREF	Metal regulatory element factors	F$CUSE.01	Copper-signaling element, AMT1/ACE1 recognition sequence
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPD.01	CCAAT/enhancer binding protein (C/EBP), delta
P$HEAT	Heat shock factors	P$HSFB3.01	Heat stress transcription factor B-3
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPE.01	CCAAT/enhancer binding protein (C/EBP), epsilon
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.01	CCAAT/enhancer binding protein beta
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPE.01	CCAAT/enhancer binding protein (C/EBP), epsilon
F$BZIP	Fungal basic leucine zipper family	F$YAP1.01	Yeast activator protein of the basic leucine zipper (bZIP) family
V$EBOX	E-box binding factors	V$MAX.03	MYC associated factor X
F$YNIT	Asperg./Neurospora-activ. of genes induced by nitrogen	F$NIT2.01	Activator of nitrogen-regulated genes
V$NFAT	Nuclear factor of activated T-cells	V$NFAT5.02	Nuclear factor of activated T-cells 5
V$NFAT	Nuclear factor of activated T-cells	V$NFATC1.01	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1, dimeric binding site
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor
V$NFAT	Nuclear factor of activated T-cells	V$NFATC1.01	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1, dimeric binding site
V$RORA	v-ERB and RAR-related orphan receptor alpha	V$RORA.02	RAR-related orphan receptor alpha, homodimer DR5 binding site
V$RXRF	RXR heterodimer binding sites	V$RXR_RXR.01	Retinoid X receptor homodimer, DR1 sites
I$DHSF	Drosophila heat shock factors	I$HSF.01	Heat shock factor (Drosophila)
V$RXRF	RXR heterodimer binding sites	V$VDR_RXR.05	Bipartite binding site of VDR/RXR heterodimers, DR4 sites
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFA4A.01	Heat stress transcription factor A-4a (HSF21)
V$HEAT	Heat shock factors	V$HSF2.01	Heat shock factor 2
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
V$HEAT	Heat shock factors	V$HSF2.03	Heat shock factor 2
V$HEAT	Heat shock factors	V$HSF1.05	Heat shock factor 1
F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family
V$FXRE	Farnesoid X - activated receptor response elements	V$FXRE.01	Farnesoid X - activated receptor (RXR/FXR dimer), IR1 sites
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPG.01	CCAAT/enhancer binding protein (C/EBP), gamma
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPG.01	CCAAT/enhancer binding protein (C/EBP), gamma
V$AP1R	MAF and AP1 related factors	V$MARE.03	Binding sites for heterodimers with small Maf-proteins
V$HEAT	Heat shock factors	V$HSF1.04	Heat shock factor 1
P$HEAT	Heat shock factors	P$HSFB3.01	Heat stress transcription factor B-3
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPA.02	CCAAT/enhancer binding protein alpha
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPA.01	CCAAT/enhancer binding protein alpha
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.02	Aryl hydrocarbon / Arnt heterodimers, fixed core
V$GMEB	Glucocorticoid modulatory element binding proteins	V$GMEB2.01	Glucocorticoid modulatory element binding protein 2
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPA.01	CCAAT/enhancer binding protein alpha
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPE.02	CCAAT/enhancer binding protein (C/EBP), epsilon
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPG.01	CCAAT/enhancer binding protein (C/EBP), gamma
V$RXRF	RXR heterodimer binding sites	V$PXR_RXR.01	Pregnane X receptor / retinoid X receptor heterodimer, DR4 sites
P$HEAT	Heat shock factors	P$HSFB3.01	Heat stress transcription factor B-3
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.01	CCAAT/enhancer binding protein beta
V$AP1R	MAF and AP1 related factors	V$MARE_ARE.01	Antioxidant response elements
F$SXBP	S.cerevisiae, XhoI site-binding protein I	F$XBP1.02	Stress-induced transcriptional repressor
F$SXBP	S.cerevisiae, XhoI site-binding protein I	F$XBP1.02	Stress-induced transcriptional repressor
V$HEAT	Heat shock factors	V$HSF1.05	Heat shock factor 1
V$HEAT	Heat shock factors	V$HSF1.04	Heat shock factor 1
F$SXBP	S.cerevisiae, XhoI site-binding protein I	F$XBP1.02	Stress-induced transcriptional repressor
F$SXBP	S.cerevisiae, XhoI site-binding protein I	F$XBP1.02	Stress-induced transcriptional repressor
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
V$HEAT	Heat shock factors	V$HSF2.03	Heat shock factor 2
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
V$HEAT	Heat shock factors	V$HSF1.01	Heat shock factor 1
I$DHSF	Drosophila heat shock factors	I$HSF.01	Heat shock factor (Drosophila)
F$YNIT	Asperg./Neurospora-activ. of genes induced by nitrogen	F$NIT2.01	Activator of nitrogen-regulated genes
P$HEAT	Heat shock factors	P$HSE.01	Heat shock element
P$HEAT	Heat shock factors	P$HSFA6B.01	Heat stress transcription factor A-6b
V$EBOX	E-box binding factors	V$MNT.01	MAX binding protein
V$CREB	cAMP-responsive element binding proteins	V$XBP1.01	X-box-binding protein 1
F$BZIP	Fungal basic leucine zipper family	F$HAC1.01	bZIP transcription factor (ATF/CREB1 homolog) that regulates the unfolded protein response

Tissue	Opt.	Start position	End position	Anchor position	Strand	Core sim.	Matrix sim.
	0,99	27	33	30	-	1	0,994
Adipose Tissue Connective Tissue Digestive System Liver	0,83	19	41	30	-	1	0,849
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,75	18	42	30	+	0,796	0,791
Blood Cells Cardiovascular System Cartilage Connective Tissue Heart Immune System Leukocytes Lymphocytes Myeloid Cells	0,78	105	123	114	-	0,831	0,794
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,84	108	132	120	-	1	0,935
	0,85	122	154	138	+	0,8	0,89
	0,83	127	149	138	-	1	0,864
ubiquitous	0,95	126	150	138	+	1	0,959
	0,88	173	187	180	-	1	0,94
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,92	181	195	188	+	1	0,972
	0,83	178	200	189	-	0,845	0,886
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,79	182	196	189	-	0,816	0,822
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,94	191	205	198	+	0,94	0,957
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,79	192	206	199	-	0,77	0,799
	0,76	191	211	201	-	1	0,771
ubiquitous	0,9	204	220	212	-	1	0,911
	0,99	216	222	219	-	1	0,994
Blood Cells Cardiovascular System Cartilage Connective Tissue Heart Immune System Leukocytes Lymphocytes Myeloid Cells	0,87	226	244	235	-	1	0,878
Blood Cells Cardiovascular System Cartilage Connective Tissue Heart Immune System Leukocytes Lymphocytes Myeloid Cells	0,78	230	248	239	-	1	0,841
	0,95	224	256	240	-	1	0,969
Blood Cells Cardiovascular System Cartilage Connective Tissue Heart Immune System Leukocytes Lymphocytes Myeloid Cells	0,78	235	253	244	+	1	0,898
Blood Cells Brain Central Nervous System Endocrine System Immune System Leukocytes Lymphocytes Nervous System Pineal Gland	0,75	234	258	246	+	1	0,816
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,78	250	274	262	+	0,833	0,785
	0,77	254	278	266	-	1	0,778
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,79	260	284	272	-	0,952	0,811
	0,85	257	289	273	-	0,8	0,872
	0,81	262	284	273	+	1	0,892
ubiquitous	0,88	261	285	273	-	1	0,888
	0,85	258	290	274	+	1	0,872
	0,86	263	285	274	-	0,762	0,883
ubiquitous	0,75	262	286	274	+	0,764	0,78
ubiquitous	0,92	267	291	279	-	0,807	0,932
	0,93	333	353	343	-	1	0,934
Digestive System Gallbladder Liver	0,8	351	363	357	+	1	0,842
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,84	367	381	374	+	1	0,982
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,84	368	382	375	-	1	0,961
Antibody-Producing Cells Blood Cells Blood Platelets Bone Marrow Cells Digestive System Ear Endocrine System Eye Hematopoietic System Immune System Islets of Langerhans Liver Pancreas	0,82	385	409	397	+	0,81	0,868
ubiquitous	0,76	394	418	406	+	0,85	0,817
	0,83	462	484	473	+	0,801	0,843
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,89	466	480	473	-	0,904	0,902
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,94	490	504	497	-	1	0,942
Digestive System Liver	0,77	501	525	513	-	0,75	0,792
	0,75	511	525	518	+	0,768	0,786
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,94	515	529	522	+	1	0,945
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,97	516	530	523	-	1	0,976
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,84	526	540	533	+	0,75	0,847
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,8	562	586	574	+	0,791	0,837
	0,83	576	598	587	+	1	0,841
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,94	580	594	587	-	1	0,969
Antibody-Producing Cells Blood Cells Blood Platelets Bone Marrow Cells Digestive System Ear Endocrine System Eye Hematopoietic System Immune System Islets of Langerhans Liver Pancreas	0,76	613	637	625	-	1	0,891
	0,91	680	692	686	-	1	0,937
	0,91	681	693	687	+	1	0,937
ubiquitous	0,92	683	707	695	+	0,801	0,926
ubiquitous	0,76	686	710	698	-	0,973	0,779
	0,91	693	705	699	-	1	0,935
	0,91	694	706	700	+	1	0,949
	0,85	684	716	700	-	1	0,931
	0,86	689	711	700	+	0,775	0,902
ubiquitous	0,75	688	712	700	-	0,768	0,763
	0,85	685	717	701	+	0,8	0,904
	0,86	690	712	701	-	0,775	0,902
ubiquitous	0,84	689	713	701	+	1	0,889
	0,77	696	720	708	+	1	0,774
	0,99	747	753	750	+	1	0,994
	0,81	740	762	751	+	1	0,831
	0,81	741	763	752	-	1	0,868
ubiquitous	0,84	802	818	810	-	0,753	0,841
Brain Central Nervous System Endocrine System Nervous System Pineal Gland	0,88	830	850	840	+	1	0,904
	0,93	832	852	842	-	1	0,974

P$HEAT	Heat shock factors	P$HSFA6B.01	Heat stress transcription factor A-6b
V$HEAT	Heat shock factors	V$HSF1.01	Heat shock factor 1
V$NFAT	Nuclear factor of activated T-cells	V$NFATC1.01	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1, dimeric binding site
V$GMEB	Glucocorticoid modulatory element binding proteins	V$GMEB2.01	Glucocorticoid modulatory element binding protein 2
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor
F$BZIP	Fungal basic leucine zipper family	F$HAC1.02	bZIP transcription factor (ATF/CREB1 homolog) that regulates the unfolded protein response
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
I$CEBP	Drosophila C/EBP like bZIP transcription factors	I$SLBO.01	Slow border cells
F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family
P$HEAT	Heat shock factors	P$HSFB3.01	Heat stress transcription factor B-3
V$PERO	Peroxisome proliferator-activated receptor	V$PPARG.03	Peroxisome proliferator-activated receptor gamma, DR1 sites
V$RXRF	RXR heterodimer binding sites	V$RXRA.01	Retinoid X receptor alpha homodimer, DR1 sites
V$RORA	v-ERB and RAR-related orphan receptor alpha	V$RORA1.01	RAR-related orphan receptor alpha1
V$IRFF	Interferon regulatory factors	V$IRF1.01	Interferon regulatory factor 1
OdMT2			
I$CEBP	Drosophila C/EBP like bZIP transcription factors	I$SLBO.01	Slow border cells
F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family
V$IRFF	Interferon regulatory factors	V$IRF2.02	Interferon regulatory factor 2
V$IRFF	Interferon regulatory factors	V$IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)
V$IRFF	Interferon regulatory factors	V$IRF1.01	Interferon regulatory factor 1
V$IRFF	Interferon regulatory factors	V$IRF7.01	Interferon regulatory factor 7 (IRF-7)
V$AP1R	MAF and AP1 related factors	V$MAFK.01	V-maf musculoaponeurotic fibrosarcoma oncogene homolog K (half site)
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPE.02	CCAAT/enhancer binding protein (C/EBP), epsilon
P$HEAT	Heat shock factors	P$HSE.01	Heat shock element
P$HEAT	Heat shock factors	P$HSFB2A.02	Heat stress transcription factor B-2a (HSF6)
V$IRFF	Interferon regulatory factors	V$IRF3.01	Interferon regulatory factor 3 (IRF-3)
P$AREF	Auxin response element	P$SEBF.01	Silencing element binding factor - transcriptional repressor
F$MREF	Metal regulatory element factors	F$CUSE.01	Copper-signaling element, AMT1/ACE1 recognition sequence
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance
V$CREB	cAMP-responsive element binding proteins	V$E4BP4.01	E4BP4, bZIP domain, transcriptional repressor
V$CREB	cAMP-responsive element binding proteins	V$E4BP4.01	E4BP4, bZIP domain, transcriptional repressor
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance
V$HEAT	Heat shock factors	V$HSF1.05	Heat shock factor 1
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
V$HEAT	Heat shock factors	V$HSF2.03	Heat shock factor 2
F$YHSF	Yeast heat shock factors	F$HSF.01	Heat shock factor (yeast)
P$HEAT	Heat shock factors	P$HSFC1.01	Heat shock transcription factor C1
V$HEAT	Heat shock factors	V$HSF2.03	Heat shock factor 2
V$HEAT	Heat shock factors	V$HSF1.05	Heat shock factor 1
V$RXRF	RXR heterodimer binding sites	V$RXR_RXR.01	Retinoid X receptor homodimer, DR1 sites
V$GREF	Glucocorticoid responsive and related elements	V$GRE.03	Glucocorticoid receptor, IR3 sites
F$SXBP	S.cerevisiae, XhoI site-binding protein I	F$XBP1.01	S.cerevisae XhoI site-binding protein I, stressinduced expression
V$HEAT	Heat shock factors	V$HSF1.02	Heat shock factor 1
V$IRFF	Interferon regulatory factors	V$ISRE.01	Interferon-stimulated response element
V$AP1R	MAF and AP1 related factors	V$BACH1.01	BTB/POZ-bZIP transcription factor BACH1 forms heterodimers with the small Maf protein family
V$AP1R	MAF and AP1 related factors	V$BACH1.01	BTB/POZ-bZIP transcription factor BACH1 forms heterodimers with the small Maf protein family
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPE.02	CCAAT/enhancer binding protein (C/EBP), epsilon
V$NFAT	Nuclear factor of activated T-cells	V$NFAT5.02	Nuclear factor of activated T-cells 5
V$PERO	Peroxisome proliferator-activated receptor	V$PPAR_RXR.01	PPAR/RXR heterodimers, DR1 sites
V$RXRF	RXR heterodimer binding sites	V$RXRA.01	Retinoid X receptor alpha homodimer, DR1 sites
V$NGRE	"Negative" glucocoticoid response elements	V$IR1_NGRE.01	Repressive binding sites for glucocorticoid receptor (IR1)
I$CEBP	Drosophila C/EBP like bZIP transcription factors	I$SLBO.01	Slow border cells
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.01	CCAAT/enhancer binding protein beta
P$HEAT	Heat shock factors	P$HSFA1A.01	Arabidopsis thaliana class A heat shock factor 1a
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.02	CCAAT/enhancer binding protein beta
V$GREF	Glucocorticoid responsive and related elements	V$PRE.01	Progesterone receptor binding site, IR3 sites
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.01	CCAAT/enhancer binding protein beta
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance
V$IRFF	Interferon regulatory factors	V$IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)
V$CREB	cAMP-responsive element binding proteins	V$E4BP4.01	E4BP4, bZIP domain, transcriptional repressor
V$RORA	v-ERB and RAR-related orphan receptor alpha	V$RORA.02	RAR-related orphan receptor alpha, homodimer DR5 binding site
V$GREF	Glucocorticoid responsive and related elements	V$GRE.03	Glucocorticoid receptor, IR3 sites
F$BZIP	Fungal basic leucine zipper family	F$CST6.01	Chromosome stability, bZIP transcription factor of the ATF/CREB family (ACA2)
V$CREB	cAMP-responsive element binding proteins	V$E4BP4.01	E4BP4, bZIP domain, transcriptional repressor
F$BZIP	Fungal basic leucine zipper family	F$YAP1.01	Yeast activator protein of the basic leucine zipper (bZIP) family
F$YHSF	Yeast heat shock factors	F$MGA1.01	Heat shock transcription factor Mga1

[1]Colour code highlights the binding sites for transcription factors showed in Figure 2 and Figure S1. Overlapping binding sites are boxed.

	0,81	852	874	863	-	1	0,843
ubiquitous	0,84	851	875	863	+	1	0,85
Blood Cells Cardiovascular System Cartilage Connective Tissue Heart Immune System Leukocytes Lymphocytes Myeloid Cells	0,78	866	884	875	-	0,834	0,794
	0,75	876	890	883	+	0,798	0,756
	0,95	881	913	897	+	1	1
	0,83	911	931	921	+	1	0,87
	0,86	918	940	929	+	1	0,876
	0,86	919	941	930	-	1	0,877
	0,91	938	950	944	+	1	0,941
	0,93	937	957	947	-	0,967	0,974
	0,83	942	964	953	+	0,801	0,832
Adipose Tissue Connective Tissue Digestive System Liver	0,83	954	976	965	+	1	0,898
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,83	957	981	969	+	1	0,852
Blood Cells Brain Central Nervous System Endocrine System Immune System Leukocytes Lymphocytes Nervous System Pineal Gland	0,93	960	984	972	+	1	0,957
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,87	972	996	984	+	1	0,932
	0,91	28	40	34	+	1	0,92
	0,93	57	77	67	-	1	1
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,87	88	112	100	-	1	0,88
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,94	94	118	106	-	1	0,969
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,87	110	134	122	-	1	0,887
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,86	125	149	137	-	1	0,876
Antibody-Producing Cells Blood Cells Blood Platelets Bone Marrow Cells Digestive System Ear Endocrine System Eye Hematopoietic System Immune System Islets of Langerhans Liver Pancreas	0,82	137	161	149	+	0,767	0,824
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,97	152	166	159	+	1	0,987
	0,81	171	193	182	-	1	0,822
	0,81	179	201	190	-	0,833	0,81
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,85	200	224	212	-	1	0,934
	0,96	228	240	234	-	1	0,964
	0,88	233	247	240	-	1	0,902
	0,89	238	258	248	+	1	0,956
Brain Central Nervous System Endocrine System Nervous System Pineal Gland	0,8	240	260	250	-	0,769	0,817
Brain Central Nervous System Endocrine System Nervous System Pineal Gland	0,8	241	261	251	+	0,758	0,865
	0,89	243	263	253	-	1	0,909
ubiquitous	0,92	261	285	273	+	0,844	0,941
	0,85	262	294	278	-	1	1
	0,86	267	289	278	+	1	0,995
ubiquitous	0,75	266	290	278	-	1	0,755
	0,85	263	295	279	+	1	0,917
	0,86	268	290	279	-	1	0,991
ubiquitous	0,75	267	291	279	+	1	0,763
ubiquitous	0,92	272	296	284	-	0,844	0,941
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,78	296	320	308	-	0,889	0,786
Breast Endocrine System Leydig Cells Ovary Prostate Testis Urogenital System	0,88	355	373	364	+	1	0,907
	0,88	402	414	408	+	1	0,944
ubiquitous	0,75	406	430	418	-	1	0,768
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,81	446	470	458	-	1	0,864
Antibody-Producing Cells Blood Cells Blood Platelets Bone Marrow Cells Digestive System Ear Endocrine System Eye Hematopoietic System Immune System Islets of Langerhans Liver Pancreas	0,82	468	492	480	-	0,75	0,842
Antibody-Producing Cells Blood Cells Blood Platelets Bone Marrow Cells Digestive System Ear Endocrine System Eye Hematopoietic System Immune System Islets of Langerhans Liver Pancreas	0,82	470	494	482	+	0,75	0,842
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,97	558	572	565	+	0,96	0,975
Blood Cells Cardiovascular System Cartilage Connective Tissue Heart Immune System Leukocytes Lymphocytes Myeloid Cells	0,87	562	580	571	+	1	0,925
Adipose Tissue Connective Tissue Digestive System Liver	0,76	566	588	577	+	0,769	0,762
Bone and Bones Brain Central Nervous System Connective Tissue Digestive System Ear Endocrine System Integumentary System Liver Nervous System Parathyroid Glands Pituitary Gland Skeleton Thyroid Gland	0,83	569	593	581	+	1	0,846
Adrenal Glands Brain Central Nervous System Endocrine System Nervous System	0,81	587	601	594	-	1	0,816
	0,91	600	612	606	-	1	0,982
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,94	599	613	606	+	1	0,957
	0,75	608	630	619	-	1	0,784
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,91	650	664	657	+	1	0,995
Breast Endocrine System Leydig Cells Ovary Prostate Testis Urogenital System	0,84	695	713	704	-	1	0,867
Adipose Tissue Bone Marrow Cells Connective Tissue Digestive System Hematopoietic System Immune System Liver Myeloid Cells Phagocytes	0,94	775	789	782	-	1	0,984
	0,89	795	815	805	-	0,867	0,904
Antibody-Producing Cells Antigen-Presenting Cells Blood Cells Bone Marrow Cells Hematopoietic System Immune System Leukocytes Lymphocytes Monocytes Myeloid Cells Phagocytes	0,94	829	853	841	-	1	0,96
Brain Central Nervous System Endocrine System Nervous System Pineal Gland	0,8	835	855	845	-	1	0,827
Blood Cells Brain Central Nervous System Endocrine System Immune System Leukocytes Lymphocytes Nervous System Pineal Gland	0,75	877	901	889	-	1	0,779
Breast Endocrine System Leydig Cells Ovary Prostate Testis Urogenital System	0,88	883	901	892	+	1	0,883
	0,85	889	909	899	+	1	0,861
Brain Central Nervous System Endocrine System Nervous System Pineal Gland	0,8	892	912	902	+	1	0,802
	0,76	894	914	904	-	1	0,816
	0,88	895	927	911	+	0,826	0,898

ARTICLE V: *Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins.*

Front. Cell Dev. Biol., 02 July 2021

DOI: 10.3389/fcell.2021.702688

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fcell.2021.702688

Modular Evolution and Population Variability of *Oikopleura dioica* Metallothioneins

Sara Calatayud[1†], Mario Garcia-Risco[2†], Mercè Capdevila[2], Cristian Cañestro[1], Òscar Palacios[2] and Ricard Albalat[1*]*

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain, [2] Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Chordate *Oikopleura dioica* probably is the fastest evolving metazoan reported so far, and thereby, a suitable system in which to explore the limits of evolutionary processes. For this reason, and in order to gain new insights on the evolution of protein modularity, we have investigated the organization, function and evolution of multi-modular metallothionein (MT) proteins in *O. dioica*. MTs are a heterogeneous group of modular proteins defined by their cysteine (C)-rich domains, which confer the capacity of coordinating different transition metal ions. *O. dioica* has two MTs, a bi-modular OdiMT1 consisting of two domains (t-12C and 12C), and a multi-modular OdiMT2 with six t-12C/12C repeats. By means of mass spectrometry and spectroscopy of metal-protein complexes, we have shown that the 12C domain is able to autonomously bind four divalent metal ions, although the t-12C/12C pair –as it is found in OdiMT1– is the optimized unit for divalent metal binding. We have also shown a direct relationship between the number of the t-12C/12C repeats and the metal-binding capacity of the MTs, which means a stepwise mode of functional and structural evolution for OdiMT2. Finally, after analyzing four different *O. dioica* populations worldwide distributed, we have detected several OdiMT2 variants with changes in their number of t-12C/12C domain repeats. This finding reveals that the number of repeats fluctuates between current *O. dioica* populations, which provides a new perspective on the evolution of domain repeat proteins.

Keywords: appendicularian tunicate chordates, modular protein evolution, metallothionein domains, tandem domain repeats, population variants, intra-species variability

OPEN ACCESS

Edited by:
Juan Pascual-Anaya, Malaga University, Spain

Reviewed by:
María Ayelén Pagani, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
Filipe Castro, University of Porto, Portugal

***Correspondence:**
Òscar Palacios
Oscar.Palacios@uab.cat
Ricard Albalat
ralbalat@ub.edu

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Evolutionary Developmental Biology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *29 April 2021*
Accepted: *09 June 2021*
Published: *02 July 2021*

Citation:
Calatayud S, Garcia-Risco M, Capdevila M, Cañestro C, Palacios Ò and Albalat R (2021) Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins. Front. Cell Dev. Biol. 9:702688. doi: 10.3389/fcell.2021.702688

INTRODUCTION

Oikopleura dioica is a tunicate species of the appendicularian class in the chordate phylum. This species is emerging as a non-classical animal model in the field of evolutionary developmental biology (a.k.a. evo-devo) especially attractive for its unusually dynamic gene and genome evolution (reviewed in Ferrández-Roldán et al., 2019). At genome level, *O. dioica* has suffered numerous chromosomal rearrangements accompanied by a drastic process of compaction, becoming the smallest known chordate genome (Denoeud et al., 2010). At gene level, besides an extraordinary amount of gene duplications and losses, *O. dioica* sequences show high evolutionary rates, which

are on average two-three times higher than in ascidians and vertebrates (Berna et al., 2012; Berna and Alvarez-Valin, 2014). *O. dioica* probably is the fastest evolving metazoan reported so far (Edvardsen et al., 2005; Denoeud et al., 2010). In addition, its pattern of amino acid substitution also shows some unusual traits in comparison with other chordates. Cysteines (Cys, C), for instance, are the less conserved amino acids in *O. dioica* proteins (Berna et al., 2012; Berná and Alvarez-Valin, 2015), whereas Cys are one of the most conserved amino acids according with classical analyses of protein evolution (Henikoff and Henikoff, 1992; Marino and Gladyshev, 2010). These exceptional evolutionary features make comparative studies between *O. dioica* and other chordate species suitable for understanding the functional and structural limits to which chordate genes and proteins can evolve.

In protein evolution, domains are considered evolutionary modules, and the majority of proteins of all living beings are multi-modular proteins that consist of several domains (Apic et al., 2001). While the creation of multi-modular proteins through shuffling of different domain types has been extensively analyzed (Apic et al., 2001; Björklund et al., 2005; Dohmen et al., 2020), the evolution of proteins made of tandem domain repeats is less understood (Björklund et al., 2006). To get new insights into the functional and structural evolution of these domain repeat proteins, we have focused on the multi-modular metallothioneins (MTs), using those of the fast evolving *O. dioica* species as case study.

Metallothioneins are proteins known for their high percentage of cysteines (Kojima et al., 1999), which confers them the capability of binding both essential and non-essential transition metals (reviewed in Capdevila et al., 2012; Blindauer, 2014). The Cys residues of MTs are arranged in distinctive motifs (i.e., CxC, CC, and CCC), whose number and distribution led to define different functional domains, originally for mammalian MTs (Braun et al., 1986), and later, in other animal MTs (Riek et al., 1999; Munoz et al., 2002; Baumann et al., 2017; Beil et al., 2019; Calatayud et al., 2021b). In chordates, for instance, vertebrate and cephalochordate MTs are bi-modular proteins with two domains that have diverse preferences and capacities for binding zinc (Zn), copper (Cu), or cadmium (Cd) ions (Capdevila and Atrian, 2011; Vasak and Meloni, 2011; Guirola et al., 2012; Artells et al., 2013). In contrast, most tunicate MTs are mono-modular proteins, whose single domain has a pervasive preference for Cd(II) ions (Calatayud et al., 2021a). The domain configuration of each MT is, indeed, functionally and structurally relevant because domains determine the formation of metal-thiolate clusters: domains with 9 Cys cluster with three divalent metal ions, while 11/12 Cys domains cluster with four divalent metal ions (e.g., mammalian β and α domains, respectively, Otvos and Armitage, 1980; Schultze et al., 1988). In addition, domain analyses have been shown to be helpful for elucidating the origin and evolutionary relationships of MTs in diverse groups of mollusks (Jenny et al., 2016; Nam and Kim, 2017; Calatayud et al., 2021b), and to reconstruct the complex evolutionary history of chordate MTs (Calatayud et al., 2021a).

O. dioica has two MTs, a bi-modular OdiMT1 and a multi-modular OdiMT2 (formerly OdMT1 and OdMT2) made of different number of domain repeats (Calatayud et al., 2018). The arrangement of Cys motifs in *O. dioica* domains diverges from that found in the MTs of other tunicates belonging to the ascidian and thaliacean classes, but it is similar to that of other appendicularian species of the same genus, *O. albicans* and *O. vanhoeffeni* (Calatayud et al., 2021a). Comparison of the appendicularian MTs show that the original *Oikopleura* MT domain had twelve cysteines (12C), and that this domain corresponds to previously described C7 + C5 subunits (Calatayud et al., 2018, 2021a). OdiMT1 would have therefore two 12C domains, but its amino-terminal domain was "trimmed" to become a t-12C domain that lacks the C5 subunit. OdiMT2 would be a multi-modular MT derived from an ancestral copy with a t-12C/12C domain organization, similar to the current OdiMT1, that was tandem duplicated five times yielding its domain repeat organization (Calatayud et al., 2018). In this work, we have pursued the analysis of *O. dioica* MTs, paying special attention to their modular configuration. First, we have characterized the metal-binding features of the original *Oikopleura* 12C domains –both the full-length (12C) and the trimmed (t-12C)–, revealing that although the 12C domain autonomously coordinates divalent metal ions, the t-12C/12C pair seems an improved form for divalent metal binding. Second, we have shown a direct relationship between the number of the t-12C/12C domain repeats and the metal-binding capacity of OdiMTs. Finally, taking advantage of the high level of genetic variation among *O. dioica* populations (Wang et al., 2020; Bliznina et al., 2021), we have detected population variants of OdiMT2 with changes in their number of t-12C/12C domains. Our data expose a high structural plasticity of MTs in *O. dioica* that, as if it was a natural test-bench, seems to be exploring the chordate limits of MT modularity.

MATERIALS AND METHODS

Production and Purification of Recombinant Metal-MT Complexes

Production and purification of recombinant metal-MT complexes of proteins with different number and combinations of t-12C and 12C domains (see **Table 1** for details; sequences of the domains are from Norwegian OdiMT1 and OdiMT2 sequences) was performed as described elsewhere (Calatayud et al., 2018, 2021b). In brief, synthetic cDNAs codifying the different constructs of 12-Cys domains were provided by Synbio Technologies (Monmouth Junction, NJ, United States), cloned in the pGEX-4T-1 expression vector (GE Healthcare, Chicago, IL, United States) and transformed in protease-deficient *E. coli* BL21 strain. Metal-MT complexes were produced in *E. coli* BL21 cultures expressing the recombinant plasmids, after induction with isopropyl-β-D-thiogalactopyranoside (100 μM) and supplementation with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM), or $CuSO_4$ (500 μM). Metal-MT complexes were purified from the soluble protein fraction of sonicated bacteria by affinity purification of the GST-tagged proteins, and digestion with thrombin. Notice that the digestion with thrombin added two additional residues, Gly and Ser, at the N-terminal end of all

purified proteins. These two amino acids do not interfere with the metal-binding features of recombinant MTs (Cols et al., 1997). The metal-MT complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.0 or with fresh 50 mM ammonium acetate, pH 7.0, and run at 0.8 mL min^{-1}. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at −80°C until use.

Analysis of Metal-MT Complexes

All designed constructs of OdiMTs were characterized by means of mass spectrometry (ESI-MS) and spectroscopy (ICP-AES). An electrospray ionization mass spectrometry (ESI-MS) Micro Tof-Q Instrument (Brucker Daltonics Gmbh, Bremen, Germany) interfaced with a Series 1100 HPLC pump (Agilent Technologies) was used to determine the molecular mass of the recombinant proteins. The instrument was calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies, United States) and the experimental conditions were set up as follows: injection of 10–20 μL of sample through a PEEK long tube (1–1.5 m × 0.18 mm i.d.) at 30–50 μL min^{-1}; capillary-counterelectrode voltage, 3.5–5.0 kV; desolvation temperature, 90–110°C; dry gas, 6 L min^{-1}. Data was acquired over an *m/z* range of 800–3,000. The liquid carriers were a 90:10 mixture of 15 mM ammonium acetate and acetonitrile at pH 7.0 and a 95:5 mixture of formic acid and acetonitrile at pH 2.4.

Element concentrations of S, Zn, Cd, and Cu in the sample were determined by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) by means of a Perkin-Elmer Optima 4300 DV (Waltham, United States) at the correct wavelength (S, 182.04 nm; Zn, 213.86 nm; Cd, 228.80 nm; Cu, 324.80 nm) under conventional conditions (Bongers et al., 1988). MTs concentration was calculated based on the S concentration obtained by ICP-AES, assuming that all the sulfur measured comes from peptides' Cys and Met residues.

Analysis of MT Variation

We analyzed the MT sequence variations of four geographically distant *O. dioica* populations: Norway, Japan (Osaka and Okinawa), Oregon, and Catalonia. We used Norwegian OdiMT1 (NCBI accession number CABV01001936.1) and OdiMT2 (NCBI accession number CABV01001042.1) sequences retrieved from the Oikobase genome database[1] (Denoeud et al., 2010) as reference for blast searches in ANISEED[2] and NCBI Sequence Read Archives[3] for Japanese population (Osaka and Okinawa, respectively). Raw sequence data from SRA searches was assembled using SeqMan 8.0.2 (Pro Assembler) software from the DNASTAR Lasergene package, and manually inspected in order to reconstruct the MT sequences.

For Catalonian sequences, we PCR amplified the *MT* genes from specimens captured in the Mediterranean coast of Barcelona and cultured in our animal facility at the University of Barcelona (Martí-Solans et al., 2015). Primers and PCR conditions are listed in **Supplementary Table 1**. PCR products were cloned using Topo TA Cloning® Kit of Invitrogen and sequenced at the Scientific and Technological Centers of the University of Barcelona, using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) in an automatic sequencer (ABIPRISM 310, Applied Biosystems). For Oregonian sequences, we took advantage of a data of a genomic shotgun approach. All new MT sequences have been deposited in Genbank database at NCBI[4].

For comparisons, MT sequences from different *O. dioica* populations were aligned with Aliview program (Larsson, 2014) and reviewed manually. Phylogenetic reconstructions were based on ML inferences calculated with PhyML v3.0 and automatic mode of selection of substitution model (Guindon et al., 2010) using protein sequence alignments. Sequences and the accession numbers used in this study are provided in **Supplementary Table 2**.

RESULTS AND DISCUSSION

Functional Analysis of the Modularity in OdiMTs

Metal-Binding Capabilities of 12C and t-12C Domains

OdiMT1 is a 72 amino acid long protein (20 Cys, 28%) that can be divided in three parts: an N-terminal peptide made of six amino acids; next, an amino-terminal t-12C domain; and finally, a carboxyl-terminal 12C domain. OdiMT2 is a protein of 399 amino acids (123 Cys, 31%) that consists of the N-terminal peptide, and six tandem repeat units (RU1–RU6), each one made of a t-12C domain and a 12C domain (**Figure 1**; Calatayud et al., 2018). To investigate the functionality of this modular organization, we studied the formation of metal-MT complexes of recombinant proteins containing different combinations of t-12C and 12C domains (**Table 1**). The metal-binding capability of each construct was analyzed by inductively coupled plasma atomic emission spectroscopy (ICP-AES), and the species recovered identified by electrospray ionization mass spectrometry (ESI-MS) (Capdevila et al., 2012). ICP-AES was also used for metal-to-protein stoichiometry determination through the measurement of element composition of the samples (S, Zn, Cd, and Cu) (Bongers et al., 1988).

We first characterized the metal-binding features of the 12C domain by analyzing its ability to form metal complexes when expressed alone. Based on OdiMT1 sequence, we designed construct 1.1, which encoded a 42 amino acid peptide (residues from 31 to 72) comprising the 12C domain of OdiMT1 (**Table 1**). Notice that the 12C domain of OdiMT1 is, indeed, an 11C domain due to a Cys to Ser substitution in the carboxyl terminal region in comparison with prototypical 12C domain (thus, we named this domain as 11C/12C; **Figure 1**). Our results showed that this 11C/12C domain was able to yield almost unique Zn_4- and Cd_4-protein species (**Figure 2**), in a similar way that other domains with 11 or 12 Cys: the 11C mammal and echinoderm

[1] http://oikoarrays.biology.uiowa.edu/Oiko

[2] https://www.aniseed.cnrs.fr/

[3] https://www.ncbi.nlm.nih.gov/sra/

[4] https://www.ncbi.nlm.nih.gov/genbank/

TABLE 1 | Heterogously produced proteins containing different number and combinations of t-12C and 12C domains.

Construct	Scheme[a]	Domains	Amino acid included (Num. C)	Num.aa[b]	Major metal-protein species
OdiMT (full protein)[c]		**N-term + t-12C + 12C**	From 1 to 72 (20)	72	$Cd_7/Zn_7/Cu_{12}$
OdiMT1-Construct 1.1		**12C**	From 31 to 72 (11)	42	$Cd_4/Zn_4/Cu_9$
OdiMT1-Construct 1.2		**N-term + 12C**	From 1 to 6 (1) + from 31 to 72 (11)	48	$Cd_5/Zn_2/Cu_4$
OdiMT1-Construct 1.3		**N-term + 12C**	From 1 to 30 (9) + from 55 to 72 (4)	48	$Cd_5/Zn_4/Cu_{11}$
OdiMT1-Construct 1.4		**t-12C**	From 7 to 30 (8)	24	Cd_3/Zn_3/nd[d]
OdiMT1-Construct 1.5		**t-12C**	From 31 to 54 (7)	24	Cd_3/Zn_3/nd
OdiMT2 (full protein)[c]		**N-term + [t-12C+12C] ×6**	From 1 to 399 (123)	399	nd/nd/nd
OdiMT2-Construct 2.1		**N-term + [t-12C+12C] ×3**	From 1 to 199 (61) + from 396 to 399 (2)	203	nd/Cd_{22}/nd
OdiMT2-Construct 2.2		**[t-12C+12C] ×3**	From 201 to 399 (62)	199	Zn_{20}/Cd_{22}/nd

[a] *Colored boxes represents the six amino acids of the N-terminus (black), a "trimmed" 12C domain (green) and a complete 12C domain (orange).*
[b] *This number does not include two additional amino acids (Gly and Ser) added at the N-terminus of the recombinant proteins as consequence of the experimental design for the expression and purification of the metal-protein complexes (see section "Materials and Methods" for details).*
[c] *OdiMT1 and OdiMT2 full proteins were analyzed in Calatayud et al. (2018).*
[d] *nd, no detected.*

α domains (Stillman et al., 1987; Tomas et al., 2013), the 12C mollusk α domains (Digilio et al., 2009), the insect 12C MTs (Egli et al., 2006) and the ascidian and thaliacean 12C MTs (Calatayud et al., 2021a). These results revealed a significant structural and functional autonomy of *O. dioica* 11C/12C domain, which was able to form stable metal-protein clusters with Zn(II) and Cd(II). Regarding the metal preference of this domain, the mixture of multiple Cu_n-protein complexes (n ranging from 4 to 10) in Cu(II) surplus productions (**Figure 2**) discarded a preference of this domain for monovalent Cu(I) ions. In addition, a preference for Cd(II) over Zn(II) might be indirectly inferred not only from the neatness of the ESI-MS spectra of the Cd-preparation but also from the fact that the domain rendered homometallic Cu(I) species in the Cu-preparations (**Figure 2**), which is characteristic of Cd-thioneins, whereas Zn-thioneins yield heterometallic Zn/Cu-MT complexes when expressed Cu-enriched media (Palacios et al., 2011). Overall, we concluded that the 11C/12C domain of OdiMTs formed stable clusters with four Cd(II) ions, which was in agreement with the biochemical features reported for the full OdiMT1 protein (Calatayud et al., 2018).

Although we obtained reliable results for the 11C/12C domain of OdiMT1, we wondered if the extra Cys found the N-terminal peptide of OdiMT1 could compensated the loss of one Cys in this domain, significantly improving its metal coordination features (i.e., enhancing specificity, increasing stability or metal-to-protein stoichiometries). To test this possibility, we designed two constructs: construct 1.2, which expressed a 48 amino acid peptide comprising the six N-term residues of OdiMT1 (which included the extra Cys) fused to the 11C/12C domain (**Table 1**); and construct 1.3, which also expressed a 48 amino acid peptide comprising the six N-term residues fused to a 12C domain resulting from the combination of the t-12C domain of OdiMT1 (from 7 to 30) with the 18 last residues of the carboxyl-terminus of the protein (from 55 to 72) (**Table 1**). Although both constructs rendered Cd_5-protein complexes as major species according to ESI-MS data (**Figure 2**), the samples recovered from both, Zn- and Cd-supplemented cultures, resulted to be a significant mixture of metal-protein complexes and some Cd-protein species contained sulfide labile ligands (**Figure 2**). The presence of S^{2-} ions is probably due to the incapability of these artificially designed peptides to build a stable metal cluster as these "extra" ligands can aid in the stabilization of the metal clusters (Capdevila et al., 2005). This, together with the heterogeneity of the samples suggested, therefore, that the 1.2 and 1.3 constructs have not improved the metal binding abilities of construct 1.1, questioning the contribution of the extra Cys in the N-terminal peptide to the metal coordination, and reinforcing the functional entity of the 11C/12C domain as an efficient solution for coordinating divalent metal ions emerged during the evolution of the *Oikopleura* lineage.

Next, we investigated the functionality of the t-12C domain, which is a 12C domain lacking the carboxyl C5 subunit, and thereby, containing only 7/8 cysteines. We analyzed the metal-binding features of two different t-12C domains expressed by two constructs. Construct 1.4 produced the t-12C domain of OdiMT1, from residues 7 to 30. Construct 1.5 produced the t-12C domain resulting from the truncation of the last 18 residues of the 11/12C domain of OdiMT1, from residues 31 to 54 (**Table 1**). Our analyses showed that both constructs mainly bind three divalent metal ions, either Zn(II) or Cd(II) (**Figure 2**). This metal-to-protein stoichiometry agreed with the possibility that the seven divalent metal ions (M_7) coordinated by the full OdiMT1 protein are organized in two metal clusters: an 11C/12C domain cluster with four metal ions ($M^{II}{}_4$) at the carboxyl-end of the MT, and a t-12C domain cluster with three metal ions ($M^{II}{}_3$) at the amino-term region. The t-12C domains rendered, however, multiple metallospecies and some Cd-protein species

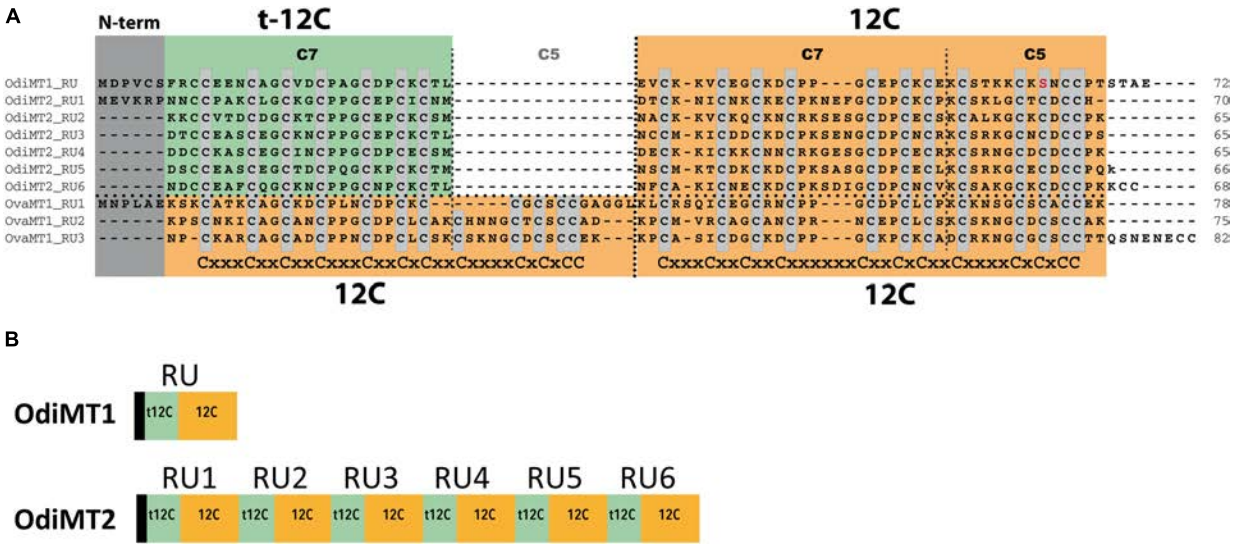


FIGURE 1 | Domain organization of *O. dioica* metallothioneins. **(A)** Amino acid alignment of *O. dioica* OdiMT1 and OdiMT2 split in its repeat units (RU1–RU6). The sequence of *O. vanhoeffeni* MT (OvaMT1) split in its repeat units (RU1–RU3) has been included to show that the amino-terminal 12C domains of OdiMT1 and of the RU of OdiMT2 are "trimmed" (t-12C; green background), and lacks the C5 subunit. Conserved cysteines are highlighted in gray. The arrangement of Cys motifs of each 12C domain is showed below the alignment. Notice that the 12C domain (orange background) corresponds to previously described C7 + C5 subunits (Calatayud et al., 2018). **(B)** Schematic representation of the modular organization of OdiMT1 and OdiMT2 proteins. OdiMT1 consists of 6 amino acids of the N-terminus (black box) followed by a single repeat unit (RU) with a t-12C domain (green box) and a 12C domain (orange box), while OdiMT2 is made of six RU, RU1–RU6.

with sulfide ligands (**Figure 2**), which suggested that the t-12C domain expressed alone did not efficiently bind the metals by itself, and that it would require the neighboring 11C/12C domain to properly coordinate the seven metal ions in the full MT protein (Calatayud et al., 2018).

Metal-Binding Capabilities of t-12C/12C Repeats

We also analyzed the metal-binding features of tandem repeats of t-12C/12C domains as they are found in OdiMT2. We designed two constructs. Construct 2.1, encoding a 203 amino acid protein, comprised the three amino-terminal RU (i.e., $RU_1RU_2RU_3$; residues 1–199), plus the four last amino acids (two of them Cys) of the carboxyl-end (from 396 to 399). Construct 2.2, encoding a 199 amino acid protein, comprised the three carboxyl-terminal RU (i.e., $RU_4RU_5RU_6$) plus the four last amino acids (from 201 to 399) (**Figure 1** and **Table 1**). Recombinant synthesis of the two partial OdiMT2 proteins –$RU_1RU_2RU_3$ and $RU_4RU_5RU_6$– in Zn(II)- and Cd(II)-enriched *E. coli* cultures rendered a low yield of metal-protein complexes, but ICP-AES analyses showed that both produced complexes with divalent metal ions, mostly with Cd(II) ions. ESI-MS analyses (**Figure 3**) showed mixtures of metal-protein complexes with different stoichiometries, and Cd-protein complexes with sulfide ligands, being Zn_{20-21}, Cd_{22-23}, and $Cd_{23-24}S$-complexes the major species (**Figure 3**). The metal-to-protein stoichiometry of these species was close to what was expected considering that the partial proteins consist of three RU, each one equivalent to a full-length OdiMT1 that binds seven Zn(II)/Cd(II) ions, i.e., 3 RU × 7 M^{II} = 21 total divalent metal ions. The two additional Cys at the carboxyl-end, the presence of sulfide ligands, and the fact that multiple metallospecies are common in multi-modular MT productions (Palacios et al., 2014; Iturbe-Espinoza et al., 2016) could account for the slightly high metal content (up to 24 ions) observed in some of them. Interestingly, ESI-MS spectra of both constructs at pH 2.4 showed species loaded with 4 and 8 divalent metal ions (**Figure 3**). These species might be reflecting the t-12C/12C organization of each RU, in which the "standard" 12C domain would be more reluctant to release its four metal ions than the "trimmed" t-C12 domain. Overall, the metal stoichiometry of the partial MTs pointed to a direct relationship between the number of the t-12C/12C repeats and the metal-binding capacity of multi-modular OdiMTs.

In conclusion, our analyses suggested that during the evolution of MTs in *O. dioica*, an ancestral 12C domain was tandem duplicated. The N-terminal 12C copy was trimmed (t-12C domain), partially losing its autonomy for metal binding, and the t-12C/12C pair became the optimized functional unit. Afterward, this primeval t-12C/12C *OdiMT* gene was duplicated. While one of the duplicates remained unaltered as the current *OdiMT1*, the other copy suffered several internal tandem duplications of the functional t-12C/12C pair in an evolutionary process that stepwise changed the number of domain repeats affecting the metal binding capacity of the new multi-modular OdiMT2. Domain expansions that generate high metal-binding capacity MTs such as OdiMT2, gene duplications that lead to the amplification of the number of *MT* genes such as those of *CUP1* in yeast (Adamo et al., 2012), and mutations in regulatory regions that lead to the overexpression of MTs in insects (Costa et al., 2012; Catalan et al., 2016) appear to be different ways of increasing the

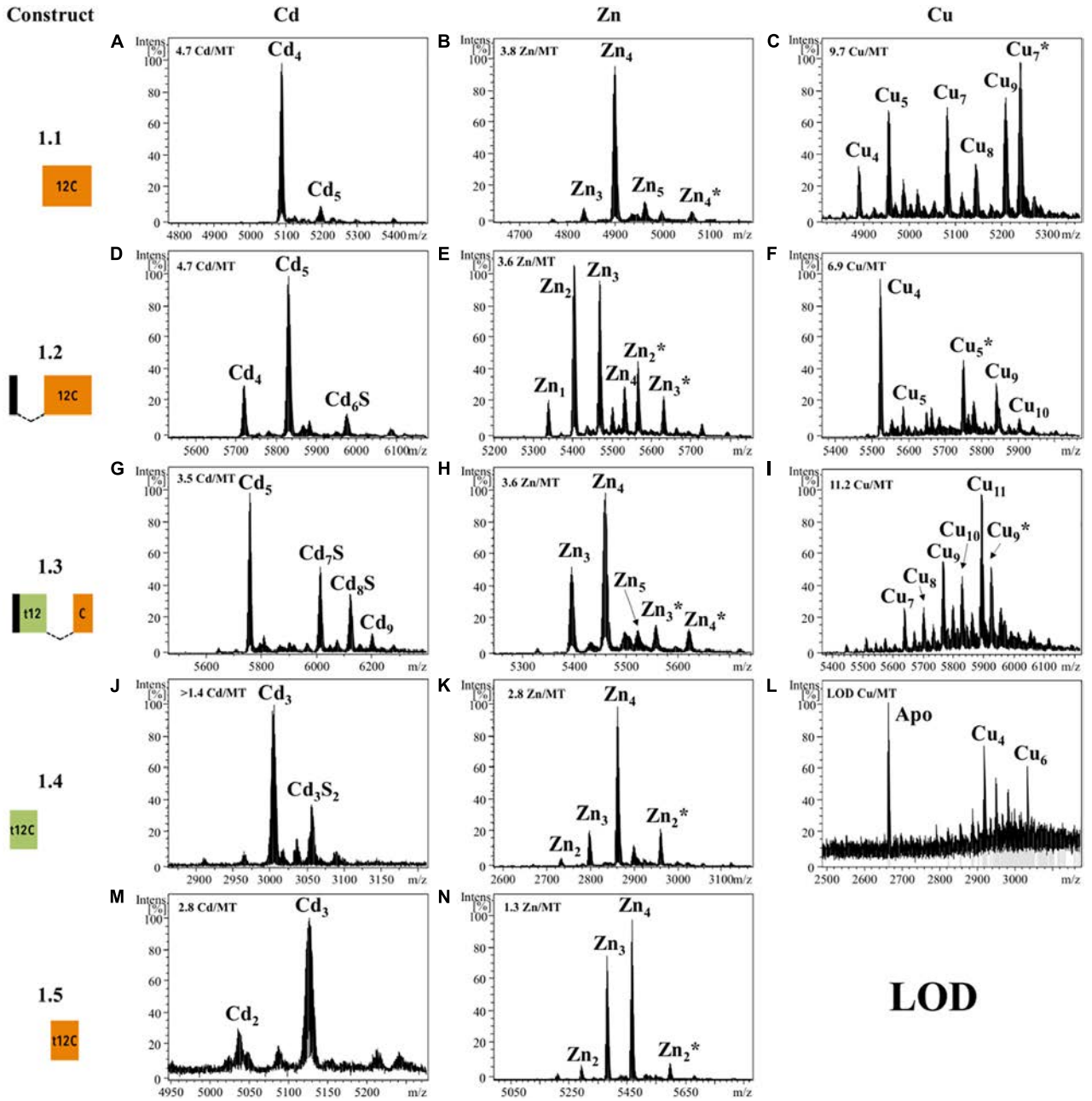


FIGURE 2 | Deconvoluted ESI-MS spectra recorded at pH 7.0 of *OdiMT1* constructs 1.1 **(A–C)**, 1.2 **(D–F)**, 1.3 **(G–I)**, 1.4 **(J–L)**, and 1.5 **(M,N)** bioproduced with Cd **(A,D,G,J,M)**, Zn **(B,E,H,K,N)**, or Cu **(C,F,I,L)**. *Y*-axis represents the relative intensities of the peaks while the *X*-axis represents the m/z ratio equivalent to the MW (Da). The metal-to-protein ratio obtained by ICP-AES is included on the top left of each panel. It is indicated those cases in which the sample concentration was lower than the limit of detection (LOD). Glycosylated species are marked with an asterisk (*).

physiological capabilities of the organisms to adapt to diverse conditions of metal bioavailability and other environmental stress situations.

Genetic Variation in *O. dioica* MTs

The peculiar structural and evolutionary characteristics of OdiMTs together with the high evolutionary rate of *O. dioica* (Denoeud et al., 2010; Berna et al., 2012; Berna and Alvarez-Valin, 2014), and significant level of sequence variation detected between *O. dioica* populations (Wang et al., 2015, 2020; Bliznina et al., 2021), prompted us to investigate the OdiMT sequences in several *O. dioica* populations worldwide distributed. We analyzed the MT sequences from animals from Norway (NOR), Oregon (ORE), Japan (Japanese specimens were from two different localizations, Osaka (OSA) and Okinawa (OKI), and we have analyzed

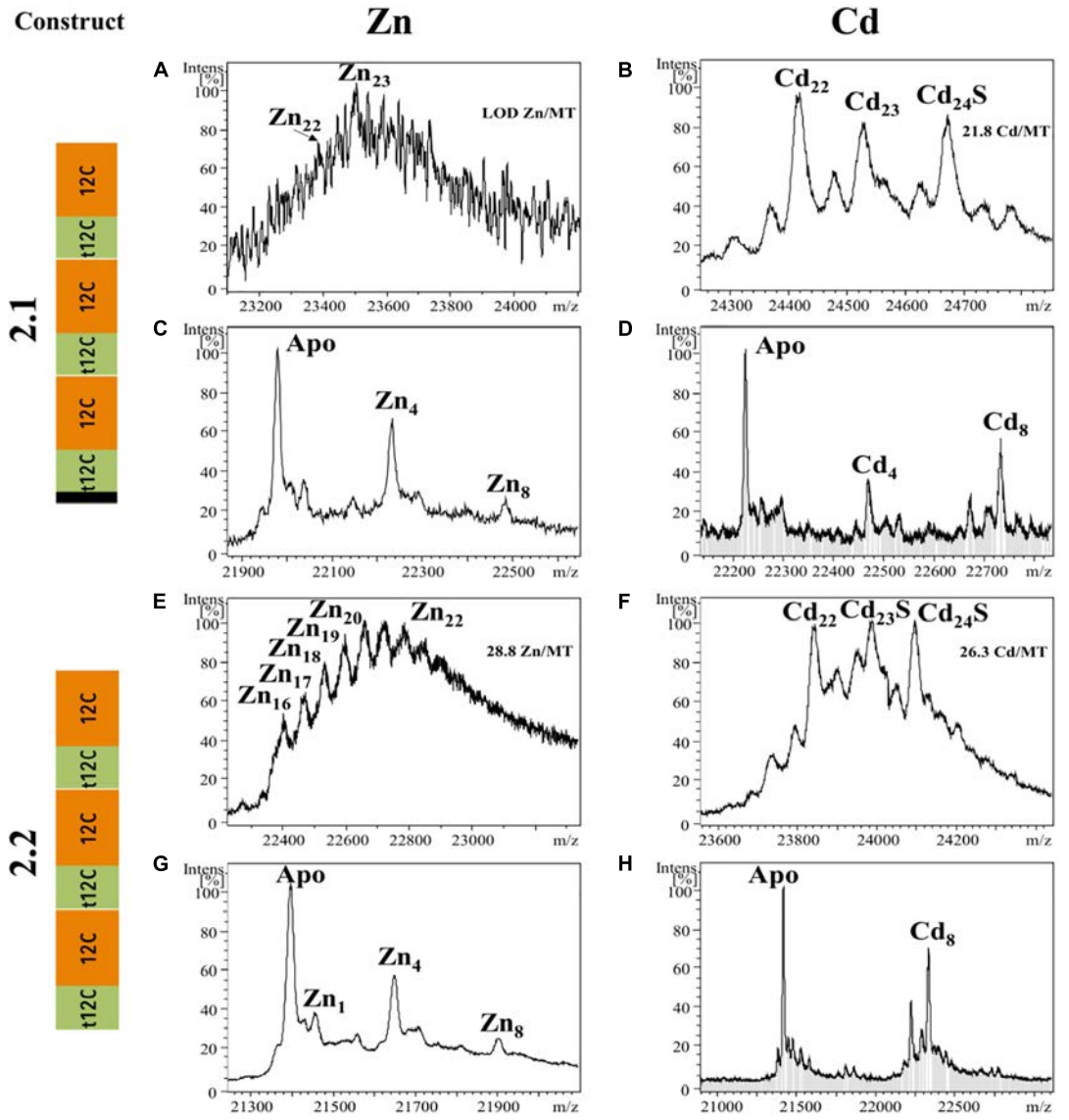


FIGURE 3 | Deconvoluted ESI-MS spectra recorded at pH 7.0 **(A,B,E,F)** and pH 2.4 **(C,D,G,H)** of *OdiMT2* constructs 2.1 **(A–D)** and 2.2 **(E–H)** bioproduced with Zn **(A,C,E,G)** or Cd **(B,D,F,H)**. *Y*-axis represents the relative intensities of the peaks while the *X*-axis represents the m/z ratio equivalent to the MW (Da). The metal-to-protein ratio obtained by ICP-AES is included on the top right of panels **(A,B,E,F)**. It is indicated those cases in which the sample concentration was lower than the limit of detection (LOD).

them separately), and Catalonia (CAT), representing four geographically distant *O. dioica* populations: north Atlantic, eastern Pacific, western Pacific and Mediterranean populations, respectively. We used as guiding reference the Norwegian sequences ($OdiMT_{NOR}$) (**Figure 4** and **Supplementary Figure 1**) because they were the first sequences identified (Calatayud et al., 2018).

Both *OdiMT1* and *OdiMT2* genes were present in all analyzed populations (**Figure 3**), but comparison among populations revealed important differences affecting three aspect: (i) sequence variability, (ii) presence of non-functional allelic variants, and (iii) differences in the number of RU in the case of OdiMT2. Regarding sequence variability, protein sequence comparisons reveal amino-acid identities ranging from 94.4% between Norway and Barcelona, up to 63.9% when compared with the sequence of Okinawa, which appeared as the more distant population to any other one (**Table 2**). In contrast to the overall sequence variation – 82.6 and 87.7% of average amino acid identity (excluding the divergent Okinawa sequences) for OdiMT1 and OdiMT2, respectively, Cys residues were nearly invariants, with 98.3% preservation. This high Cys conservation contrasted with the fact that Cys are the less conserved amino acids in *O. dioica* proteins (Berna et al., 2012; Berná and Alvarez-Valin, 2015), suggesting that Cys substitutions are negatively selected due to the functional restrictions imposed by metal coordination.

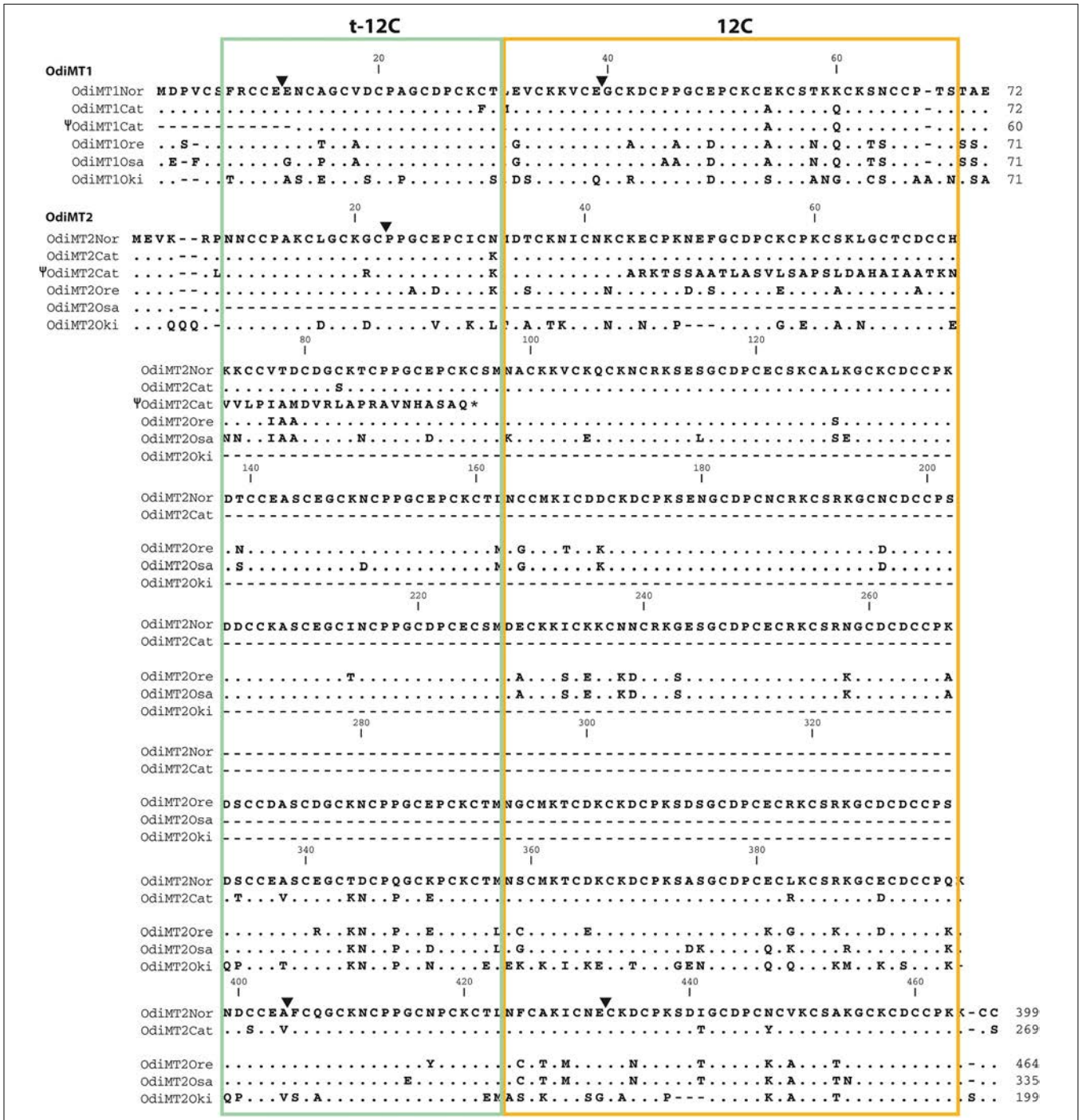


FIGURE 4 | Amino acid alignments of OdiMT1 and OdiMT2 variants form Norwegian (NOR), Catalonian (CAT), Oregonian (ORE), Osaka (OSA), and Okinawa (OKI) *O. dioica* populations. Norwegian sequences are used as reference; dots and dashes indicate amino acid identity and gaps, respectively. Black arrowheads indicate the intron positions relative to the amino acid sequences. The sequences of two hypothetical OdiMT pseudogenes –i.e., Ψ $OdiMT1_{CAT}$, lacking the first 12 residues, and Ψ $OdiMT2_{CAT}$, with a premature stop codon (*)– are also included. Trimmed 12-Cys domains (green box) and full-length 12-Cys domains (orange box) at the N- and C-terminal regions of the RU, respectively, are depicted.

Moreover, the fact that amino acid identities persistently were slightly lower than nucleotide identities (**Table 2**) indicated that nucleotide substitutions are significantly affecting non-synonymous positions, which can be considered an indication that MT variability among populations might probably be under positive selection.

Noteworthy, during the identification of the Catalonian MTs, we detected some non-functional allelic variants of *OdiMT1*

TABLE 2 | Percentage identity[a] of nucleotide (above diagonal) and amino acid sequences (below diagonal) of the OdiMT coding regions from five different *O. dioica* populations, Norway (NOR), Catalonia (CAT), Osaka (OSA), Okinawa (OKI), and Oregon (ORE).

	OdiMT1$_{NOR}$	**OdiMT1$_{CAT}$**	**OdiMT1$_{OSA}$**	**OdiMT1$_{OKI}$**	**OdiMT1$_{ORE}$**
OdiMT1$_{NOR}$	–	97.26	84.02	73.87	83.56
OdiMT1$_{CAT}$	94.44	–	84.47	72.97	83.56
OdiMT1$_{OSA}$	76.38	76.38	–	70.58	93.52
OdiMT1$_{OKI}$	67.12	64.32	63.88	–	69.09
OdiMT1$_{ORE}$	79.16	79.16	90.27	66.66	–
	OdiMT2$_{NOR}$	**OdiMT2$_{CAT}$**	**OdiMT2$_{OSA}$**	**OdiMT2$_{OKI}$**	**OdiMT2$_{ORE}$**
OdiMT2$_{NOR}$	–	96.42	86.62	70.40	86.68
OdiMT2$_{CAT}$	94.07	–	85.81	70.72	87.42
OdiMT2$_{OSA}$	85.42	81.16	–	73.95	92.77
OdiMT2$_{OKI}$	67.79	68.75	67.59	–	73.76
OdiMT2$_{ORE}$	87.00	86.67	91.69	69.23	–

[a]*For the calculation of sequence identities, a gap between two sequences has been considered as a single difference.*

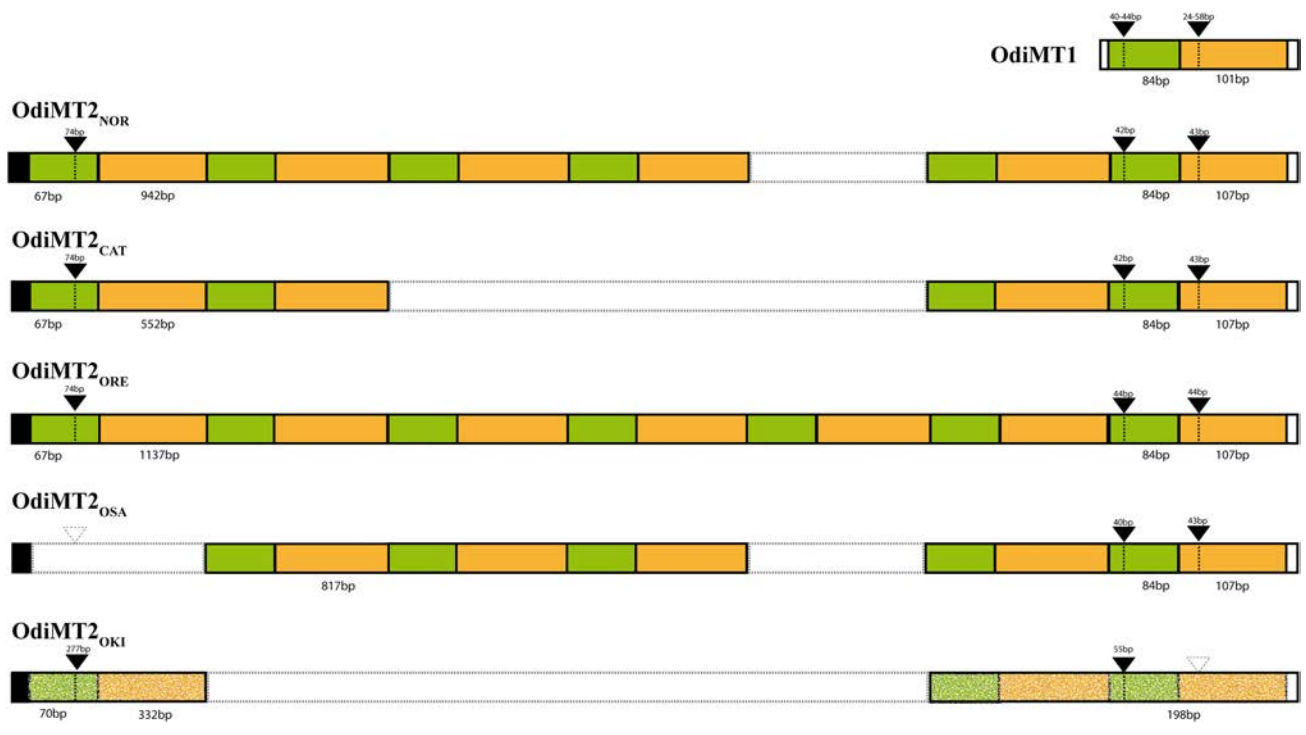


FIGURE 5 | Schematic representation of the organization of RU in OdiMTs. The t-12C/12C RU in OdiMT1s (top) and in OdiMT2 variants with six (OdiMT2$_{NOR}$), four (OdiMT2$_{CAT}$), seven (OdiMT2$_{ORE}$), five (OdiMT2$_{OSA}$), and three (OdiMT2$_{OKI}$) RU are depicted following the color code of **Figure 1**. Repeats are positioned on the basis of intron positions (black arrowheads; dashed arrowheads denote absent introns) and phylogenetic analysis (**Supplementary Figure 2**). Exon and intron sizes are indicated. The high divergence of the RU of OdiMT2$_{OKI}$ is highlighted by the dotted pattern.

and *OdiMT2* genes. One variant displayed a 35 nucleotide (nt) deletion at the 5'-end of the *OdiMT1* genomic region that eliminated the first exon, and therefore likely resulted in a pseudogene, Ψ-*OdiMT1$_{CAT}$* (**Figure 4** and **Supplementary Figure 1**). In another variant, we found a 7 nt deletion at the beginning of the exon 2 of the *OdiMT2* that caused a frame-shift mutation that resulted in a change of the amino acid sequence and a premature stop codon (**Figure 4** and **Supplementary Figure 1**). The functionality of the truncated protein codified by Ψ-*OdiMT2$_{CAT}$* was unlikely because only 44% (41 out of 92) of the amino acids were similar to the OdiMT2 sequence (**Figure 4**). The presence of these non-functional alleles for both *MTs,*Ψ-*OdiMT1$_{CAT}$,* and Ψ-*OdiMT2$_{CAT}$*, in the Mediterranean population opened the possibility that some *O. dioica* specimens might lack functional MTs. Further analyses of the presence and frequencies of non-functional variants in different populations could reveal different selective pressures related to variations in heavy-metal amounts of different environments.

The most conspicuous difference of *OdiMT* genes between *O. dioica* populations was the identification of *OdiMT2* encoding proteins with variable number of RUs (i.e., t-12C/12 pairs) (**Figures 4**, **5**). OdiMT2$_{ORE}$ with seven repeats (RU_1–RU_7) was the longest one, followed by OdiMT2$_{NOR}$, with six repeats, OdiMT2$_{OSA}$, with five repeats, OdiMT2$_{CAT}$, with four repeats, and OdiMT2$_{OKI}$, with three repeats. The conservation in the *OdiMT2* alleles of the introns both at the 5′- and 3′-ends of the gene (**Figure 5**), together with phylogenetic reconstructions based on the sequence alignments of the repeats (**Supplementary Figure 2**), suggested that the increase/decrease in the number of repeats was the result of internal tandem duplications/losses due to recombination events. Phylogenetic analysis suggested that some expansions of the RU preceded the divergence of populations, while some independent gains or losses also occurred (**Supplementary Figure 2**). In OdiMT2$_{OKI}$, for instance, duplications of RU seemed to have taken place independently form other OdiMT2s, suggesting again that this population is the most divergent among all populations. Finally, the fact that in all variants the repeat units comprised complete t-12C/12C pairs supported a modular, step-wise evolution for OdiMTs, and agreed with the functional analysis of the domains that demonstrated that despite the autonomy of the 12C domain, the t-12C/12C pair appeared as an improved functional unit for divalent metal binding (see section "Functional Analysis of the Modularity in OdiMTs").

Proteins with domain repeats have been observed to be particularly common in multicellular species (Apic et al., 2001), especially in vertebrates (Björklund et al., 2006). The exact mechanism for repeat expansion remains to be discovered, but evidence of the expansion of repeats come from the fact that orthologous proteins might have different numbers of domain repeats in different species (Björklund et al., 2006). The number of domain repeats is, however, well conserved within a species, and it does not present intraspecific variability in terms of repeat unit gains/losses (Schaper et al., 2014). The population variability of OdiMT2s here exposed is, therefore, surprising, and *O. dioica* challenges again the standard patterns of gene and protein evolution, opening a new perspective in the functional and structural evolution of domain repeat proteins.

In summary, our results suggested that the modular organization provides MTs with a high structural and functional plasticity and dynamism, as it demonstrates the detection of variants with variable number of t-12C/12C repeats. These features seem to have facilitated the creation of large multi-modular MTs with high cysteine content and a high capacity of metal binding. Large multi-modular MTs have been described in other organisms (Palacios et al., 2014; Iturbe-Espinoza et al., 2016; Dallinger et al., 2020; Calatayud et al., 2021a,b), some of them associated with biological adaptations (Palacios et al., 2014; Iturbe-Espinoza et al., 2016; Jenny et al., 2016; Baumann et al., 2017). These large MTs did not show, however, the complexity level reached by *O. dioica* proteins, nor the inter-population variability described here. It seems, therefore, that *O. dioica* would be exploring the limits of chordate MT evolvability since, although we still do not know if adaptive selection to different environmental conditions would be driving the changes in the number of the t-12C/12C repeats, the more genetic variation there is, the greater the capacity for adaptive evolution of a biological system.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

RA was responsible for the project coordination. RA, CC, ÒP, and MC conceived and designed the experiments. SC, CC, and RA collected the MT sequences from databases and elaborated the evolutionary inferences. SC performed the cloning and recombinant synthesis of the analyzed proteins. MG-R performed the ICP-AES, CD, UV-vis, and ESI-MS experiments. MG-R, ÒP, and MC analyzed the metal-binding data. SC and MG-R were responsible of the figures and tables. All authors discussed the experimental results, were responsible for writing the manuscript, commented on the manuscript, and agreed to its final version.

FUNDING

RA was supported by BIO2015-67358-C2-1-P. CC was supported by BFU2016-80601-P and PID2019-110562GB-I00. MC and ÒP were supported by BIO2015-67358-C2-2-P from the Spanish Ministerio de Ciencia e Innovación. MC and ÒP were members of the "Grup de Recerca de la Generalitat de Catalunya," ref. 2017SGR-864, and RA and CC of ref. 2017SGR-1665. MG-R acknowledges to the UAB the PIF grant.

ACKNOWLEDGMENTS

We acknowledge the Centres Científics i Tecnològics (CCiT) de la Universitat de Barcelona (DNA sequencing) and the Servei d'Anàlisi Química (SAQ) de la Universitat Autònoma de Barcelona (ICP-AES, ESI-MS) for allocating instrument time. We also thank to Sebastian Artime for experimental support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.702688/full#supplementary-material

REFERENCES

Adamo, G. M., Lotti, M., Tamas, M. J., and Brocca, S. (2012). Amplification of the CUP1 gene is associated with evolution of copper tolerance in Saccharomyces cerevisiae. *Microbiology* 158, 2325–2335. doi: 10.1099/mic.0.058024-0

Apic, G., Gough, J., and Teichmann, S. A. (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes11Edited by G. von Heijne. *J. Mole. Biol.* 310, 311–325. doi: 10.1006/jmbi.2001.4776

Artells, E., Palacios, O., Capdevila, M., and Atrian, S. (2013). Mammalian MT1 and MT2 metallothioneins differ in their metal binding abilities. *Metallomics* 5, 1397–1410. doi: 10.1039/c3mt00123g

Baumann, C., Beil, A., Jurt, S., Niederwanger, M., Palacios, O., Capdevila, M., et al. (2017). Structural Adaptation of a Protein to Increased Metal Stress: NMR Structure of a Marine Snail Metallothionein with an Additional Domain. *Angew. Chem. Int. Ed. Engl.* 56, 4617–4622. doi: 10.1002/anie.201611873

Beil, A., Jurt, S., Walser, R., Schonhut, T., Guntert, P., Palacios, O., et al. (2019). The Solution Structure and Dynamics of Cd-Metallothionein from Helix pomatia Reveal Optimization for Binding Cd over Zn. *Biochemistry* 58, 4570–4581. doi: 10.1021/acs.biochem.9b00830

Berna, L., and Alvarez-Valin, F. (2014). Evolutionary genomics of fast evolving tunicates. *Genome Biol. Evol.* 6, 1724–1738. doi: 10.1093/gbe/evu122

Berná, L., and Alvarez-Valin, F. (2015). Evolutionary volatile Cysteines and protein disorder in the fast evolving tunicate *Oikopleura dioica*. *Mar. Genomics* 24, 47–54. doi: 10.1016/j.margen.2015.07.007

Berna, L., D'Onofrio, G., and Alvarez-Valin, F. (2012). Peculiar patterns of amino acid substitution and conservation in the fast evolving tunicate Oikopleura dioica. *Mol. Phylogenet. Evol.* 62, 708–717. doi: 10.1016/j.ympev.2011.11.013

Björklund, A. K., Ekman, D., and Elofsson, A. (2006). Expansion of protein domain repeats. *PLoS Comput. Biol.* 2:e114–e114. doi: 10.1371/journal.pcbi.0020114

Björklund, Åk, Ekman, D., Light, S., Frey-Skött, J., and Elofsson, A. (2005). Domain Rearrangements in Protein Evolution. *J. Mole. Biol.* 353, 911–923. doi: 10.1016/j.jmb.2005.08.067

Blindauer, C. A. (2014). "Metallothioneins," in *Binding, Transport and Storage of Metal Ions in Biological Cells*, eds W. Maret and A. Wedd (Cambridge, UK: The Royal Society of Chemistry), 594–653.

Bliznina, A., Masunaga, A., Mansfield, M. J., Tan, Y., Liu, A. W., West, C., et al. (2021). Telomere-to-telomere assembly of the genome of an individual Oikopleura dioica from Okinawa using Nanopore-based sequencing. *BMC Genomics* 22:222. doi: 10.1186/s12864-021-07512-6

Bongers, J., Walton, C. D., Richardson, D. E., and Bell, J. U. (1988). Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma atomic emission spectrometric determination of sulfur. *Anal. Chem.* 60, 2683–2686.

Braun, W., Wagner, G., Worgotter, E., Vasak, M., Kagi, J. H., and Wuthrich, K. (1986). Polypeptide fold in the two metal clusters of metallothionein-2 by nuclear magnetic resonance in solution. *J. Mol. Biol.* 187, 125–129. doi: 10.1016/0022-2836(86)90412-2

Calatayud, S., Garcia-Risco, M., Palacios, O., Capdevila, M., Cañestro, C., and Albalat, R. (2021a). Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions and functional convergences. *Mole. Biol. Evol.* doi: 10.1093/molbev/msab184

Calatayud, S., Garcia-Risco, M., Pedrini-Martha, V., Eernisse, D. J., Dallinger, R., Capdevila, M., et al. (2021b). Modularity in Protein Evolution: Modular Organization and De Novo Domain Evolution in Mollusk Metallothioneins. *Mole. Biol. Evol.* 38, 424–436. doi: 10.1093/molbev/msaa230

Calatayud, S., Garcia-Risco, M., Rojas, N. S., Espinosa-Sanchez, L., Artime, S., Palacios, O., et al. (2018). Metallothioneins of the urochordate Oikopleura dioica have Cys-rich tandem repeats, large size and cadmium-binding preference. *Metallomics* 10, 1585–1594. doi: 10.1039/c8mt00177d

Capdevila, M., and Atrian, S. (2011). Metallothionein protein evolution: a miniassay. *J. Biol. Inorg. Chem.* 16, 977–989. doi: 10.1007/s00775-011-0798-3

Capdevila, M., Bofill, R., Palacios, Ò, and Atrian, S. (2012). State-of-the-art of metallothioneins at the beginning of the 21st century. *Coordinat. Chem. Rev.* 256, 46–62. doi: 10.1016/j.ccr.2011.07.006

Capdevila, M., Domènech, J., Pagani, A., Tío, L., Villarreal, L., and Atrian, S. (2005). Zn– and Cd–Metallothionein Recombinant Species from the Most Diverse Phyla May Contain Sulfide (S2−) Ligands. *Angew. Chem. Internat. Edn.* 44, 4618–4622. doi: 10.1002/anie.200501183

Catalan, A., Glaser-Schmitt, A., Argyridou, E., Duchen, P., and Parsch, J. (2016). An Indel Polymorphism in the MtnA 3′ Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in *Drosophila melanogaster*. *PLoS Genet.* 12:e1005987. doi: 10.1371/journal.pgen.1005987

Cols, N., Romero-Isart, N., Capdevila, M., Oliva, B., Gonzalez-Duarte, P., Gonzalez-Duarte, R., et al. (1997). Binding of excess cadmium(II) to Cd7-metallothionein from recombinant mouse Zn7-metallothionein 1. UV-VIS absorption and circular dichroism studies and theoretical location approach by surface accessibility analysis. *J. Inorg. Biochem.* 68, 157–166. doi: 10.1016/s0162-0134(97)00085-8

Costa, D., Marien, J., Janssens, T. K., van Gestel, C. A., Driessen, G., Sousa, J. P., et al. (2012). Influence of adaptive evolution of cadmium tolerance on neutral and functional genetic variation in Orchesella cincta. *Ecotoxicology* 21, 2078–2087. doi: 10.1007/s10646-012-0961-9

Dallinger, R., Zerbe, O., Baumann, C., Egger, B., Capdevila, M., Palacios, O., et al. (2020). Metallomics reveal a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins. *Metallomics* 12, 702–720. doi: 10.1039/c9mt00259f

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J. M., Da Silva, C., Brinkmann, H., et al. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 330, 1381–1385. doi: 10.1126/science.1194167

Digilio, G., Bracco, C., Vergani, L., Botta, M., Osella, D., and Viarengo, A. (2009). The cadmium binding domains in the metallothionein isoform Cd7-MT10 from Mytilus galloprovincialis revealed by NMR spectroscopy. *JBIC J. Biol. Inorgan. Chem.* 14, 167–178. doi: 10.1007/s00775-008-0435-y

Dohmen, E., Klasberg, S., Bornberg-Bauer, E., Perrey, S., and Kemena, C. (2020). The modular nature of protein evolution: domain rearrangement rates across eukaryotic life. *BMC Evol. Biol.* 20:30. doi: 10.1186/s12862-020-1591-0

Edvardsen, R. B., Seo, H. C., Jensen, M. F., Mialon, A., Mikhaleva, J., Bjordal, M., et al. (2005). Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. *Curr. Biol.* 15, R12–R13.

Egli, D., Domènech, J., Selvaraj, A., Balamurugan, K., Hua, H., Capdevila, M., et al. (2006). The four members of the Drosophila metallothionein family exhibit distinct yet overlapping roles in heavy metal homeostasis and detoxification. *Genes Cells* 11, 647–658. doi: 10.1111/j.1365-2443.2006.00971.x

Ferrández-Roldán, A., Martí-Solans, J., Cañestro, C., and Albalat, R. (2019). "Oikopleura dioica: An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development," in *Evo-Devo: Non-model Species in Cell and Developmental Biology*, eds W. Tworzydlo and S. M. Bilinski (Berlin: Springer International Publishing), 63–105.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Guirola, M., Perez-Rafael, S., Capdevila, M., Palacios, O., and Atrian, S. (2012). Metal dealing at the origin of the Chordata phylum: the metallothionein system and metal overload response in amphioxus. *PLoS One* 7:e43299. doi: 10.1371/journal.pone.0043299

Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* 89:10915. doi: 10.1073/pnas.89.22.10915

Iturbe-Espinoza, P., Gil-Moreno, S., Lin, W., Calatayud, S., Palacios, O., Capdevila, M., et al. (2016). The Fungus Tremella mesenterica Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization. *PLoS One* 11:e0148651. doi: 10.1371/journal.pone.0148651

Jenny, M. J., Payton, S. L., Baltzegar, D. A., and Lozier, J. D. (2016). Phylogenetic Analysis of Molluscan Metallothioneins: Evolutionary Insight from *Crassostrea virginica*. *J. Mol. Evol.* 83, 110–125. doi: 10.1007/s00239-016-9758-4

Kojima, Y., Binz, P.-A., and Kägi, J. H. R. (1999). "Nomenclature of metallothionein: Proposal for a revision," in *Metallothionein IV. Basel: Birkhäuser Basel*, ed. C. D. Klaassen (Basel: Birkhäuser Basel), 3–6.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Marino, S. M., and Gladyshev, V. N. (2010). Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *J. Mole. Biol.* 404, 902–916. doi: 10.1016/j.jmb.2010.09.027

Martí-Solans, J., Ferrández-Roldán, A., Godoy-Marín, H., Badia-Ramentol, J., Torres-Aguila, N. P., Rodríguez-Marí, A., et al. (2015). Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* 53, 183–193. doi: 10.1002/dvg.22800

Munoz, A., Forsterling, F. H., Shaw, C. F. III, and Petering, D. H. (2002). Structure of the (113)Cd(3)beta domains from Homarus americanus metallothionein-1: hydrogen bonding and solvent accessibility of sulfur atoms. *J. Biol. Inorg. Chem.* 7, 713–724. doi: 10.1007/s00775-002-0345-3

Nam, Y. K., and Kim, E. J. (2017). Diversification and domain evolution of molluskan metallothioneins: a mini review. *Fish. Aquatic Sci.* 20:8. doi: 10.1186/s41240-017-0054-z

Otvos, J. D., and Armitage, I. M. (1980). Structure of the metal clusters in rabbit liver metallothionein. *Proc. Natl. Acad. Sci.* 77:7094. doi: 10.1073/pnas.77.12.7094

Palacios, O., Atrian, S., and Capdevila, M. (2011). Zn- and Cu-thioneins: a functional classification for metallothioneins? *J. Biol. Inorg. Chem.* 16, 991–1009. doi: 10.1007/s00775-011-0827-2

Palacios, O., Espart, A., Espin, J., Ding, C., Thiele, D. J., Atrian, S., et al. (2014). Full characterization of the Cu-, Zn-, and Cd-binding properties of CnMT1 and CnMT2, two metallothioneins of the pathogenic fungus Cryptococcus neoformans acting as virulence factors. *Metallomics* 6, 279–291. doi: 10.1039/c3mt00266g

Riek, R., Precheur, B., Wang, Y., Mackay, E. A., Wider, G., Guntert, P., et al. (1999). NMR structure of the sea urchin (Strongylocentrotus purpuratus) metallothionein MTA. *J. Mol. Biol.* 291, 417–428. doi: 10.1006/jmbi.1999.2967

Schaper, E., Gascuel, O., and Anisimova, M. (2014). Deep Conservation of Human Protein Tandem Repeats within the Eukaryotes. *Mole. Biol. Evol.* 31, 1132–1148. doi: 10.1093/molbev/msu062

Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R., et al. (1988). Conformation of [Cd7]-metallothionein-2 from rat liver in aqueous solution determined by nuclear magnetic resonance spectroscopy. *J. Mole. Biol.* 203, 251–268. doi: 10.1016/0022-2836(88)90106-4

Stillman, M. J., Cai, W., and Zelazowski, A. J. (1987). Cadmium binding to metallothioneins. Domain specificity in reactions of alpha and beta fragments, apometallothionein, and zinc metallothionein with Cd2+. *J. Biol. Chem.* 262, 4538–4548. doi: 10.1016/S0021-9258(18)61226-8

Tomas, M., Domènech, J., Capdevila, M., Bofill, R., and Atrian, S. (2013). The sea urchin metallothionein system: Comparative evaluation of the SpMTA and SpMTB metal-binding preferences. *FEBS Open Bio* 3, 89–100. doi: 10.1016/j.fob.2013.01.005

Vasak, M., and Meloni, G. (2011). Chemistry and biology of mammalian metallothioneins. *J. Biol. Inorg. Chem.* 16, 1067–1078. doi: 10.1007/s00775-011-0799-2

Wang, K., Omotezako, T., Kishi, K., Nishida, H., and Onuma, T. A. (2015). Maternal and zygotic transcriptomes in the appendicularian, Oikopleura dioica: novel protein-encoding genes, intra-species sequence variations, and trans-spliced RNA leader. *Dev. Genes Evol.* 225, 149–159. doi: 10.1007/s00427-015-0502-7

Wang, K., Tomura, R., Chen, W., Kiyooka, M., Ishizaki, H., Aizu, T., et al. (2020). A genome database for a Japanese population of the larvacean Oikopleura dioica. *Dev. Growth Differ.* 62, 450–461. doi: 10.1111/dgd.12689

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Calatayud, Garcia-Risco, Capdevila, Cañestro, Palacios and Albalat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ARTICLE V: SUPPLEMENATRY MATERIAL

MODULAR EVOLUTION AND POPULATION VARIABILITY OF OIKOPLEURA DIOICA METALLOTHIONEINS

The file includes:

Figure S1

Figure S2

Table S1

Table S2

Supplementary Figure 1

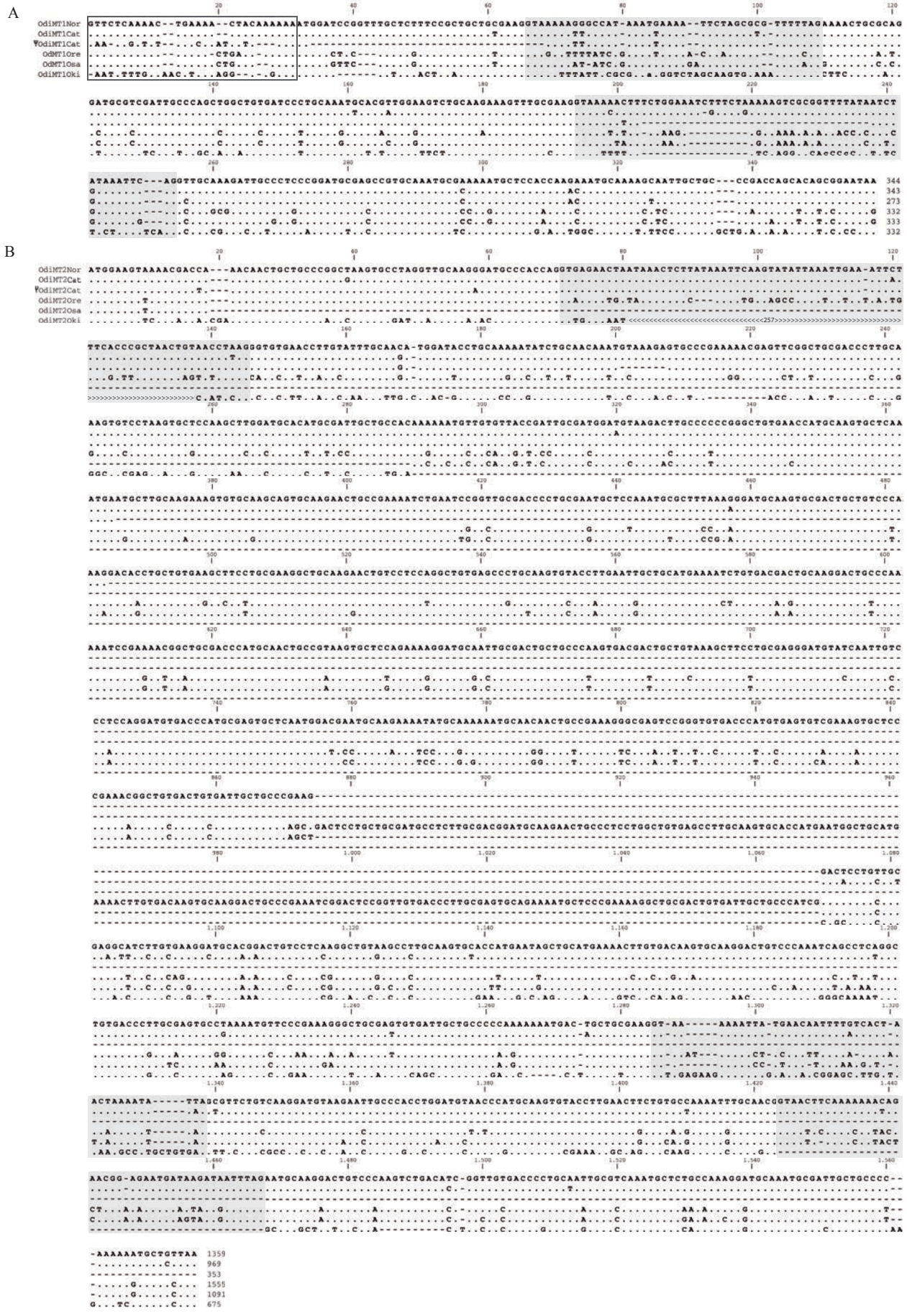
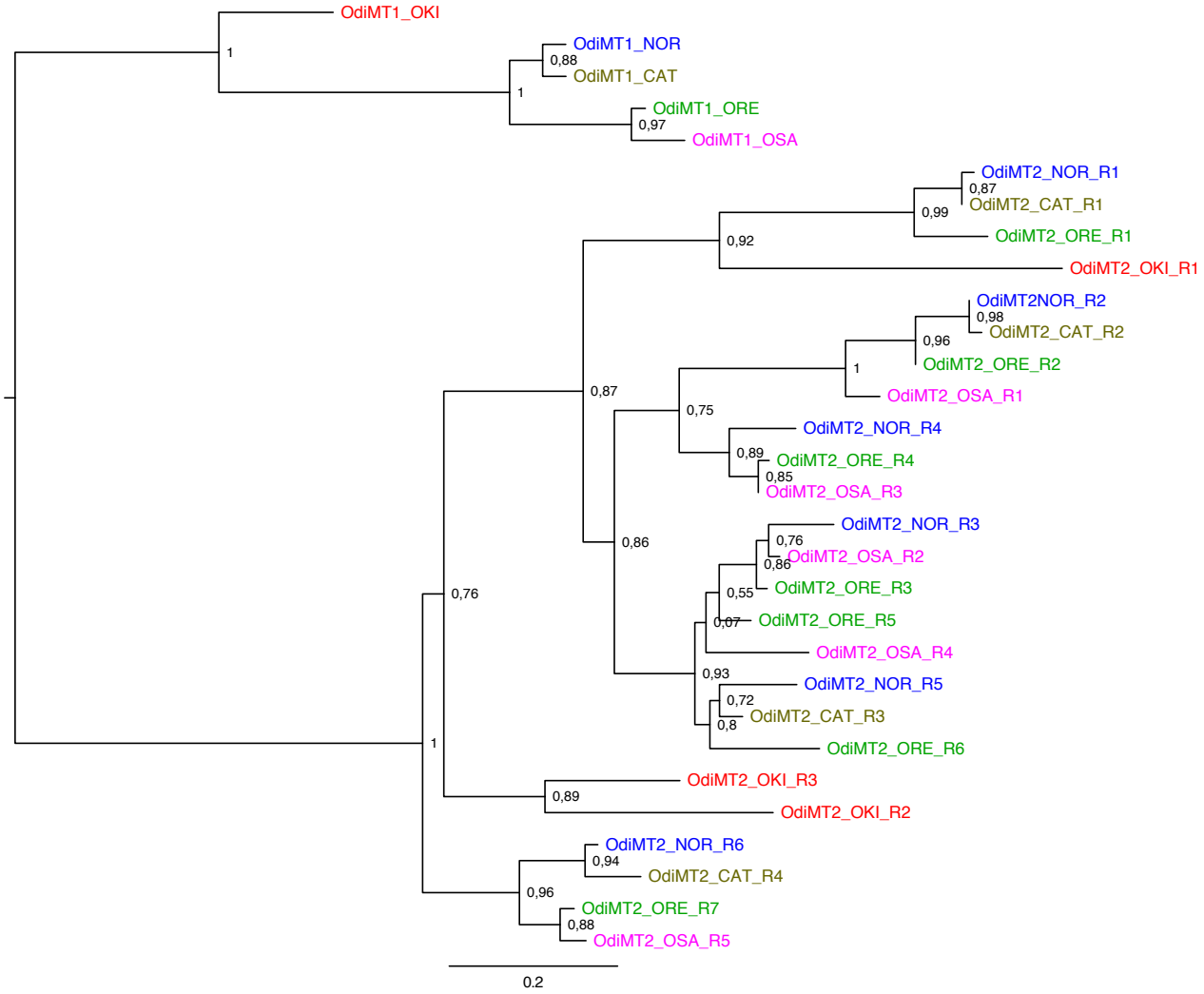


Figure S1. Nucleotide alignments of OdiMT gene variants form Norwegian (NOR), Catalonian (CAT), Oregonian (ORE), Osaka (OSA) and Okinawa (OKI) O. dioica populations. Norwegian sequences used as reference; dots and dashes indicate nucleotide identity and gaps, respectively. Grey boxes indicate introns relative to the genomic sequence alignment. (A) OdiMT1 genes. Black box depicts the sequence upstream of the ATG, revealing that a 35 nt deletion at the 5'-end of the Ψ-OdiMT1CAT removed the first exon. (B). OdiMT2 genes. Arrows in first intron of OdiMT2OKI indicate that the full nucleotide sequence is not included because the large size of the intron.

Supplementary Figure 2.



Supplementary Figure 2. Phylogenetic tree of the RU of OdiMT1 and OdiMT2 proteins from Norwegian (NOR; blue), Catalonian (CAT; olive), Oregonian (ORE; green), Osaka (OSA; purple) and Okinawa (OKI; red) *O. dioica* populations. RU are numbered from the amino- to the carboxyl-end of each MT variant as represented in **Figure 5**. Values for the approximate likelihood ratio test (aLRT) are shown at nodes The scale bar indicates amino acid substitutions.

Table S1. Primers used for *OdiMT1*$_{CAT}$ and *OdiMT2*$_{CAT}$ *PCR-amplification and PCR conditions.*

Gene	Forward	Reverse	PCR conditions
OdiMT1$_{CAT}$	5' GGGGAATCGATATTGACAAATCTTCAA 3'	5' GGGCTCGAGTTATTCCGCTGTGCTGGTCGGGCAG 3'	94°C 3 min; 94°C 20 sec, 55ºC 30 sec and 72°C 1.30 min x40 cycles; 72°C 10 min
OdiMT2$_{CAT}$	5' GGGGGATCCATGGAAGTAAAACGACC 3'	5' GGTTCAGACGAAATTTGTCCCG 3'	98°C 30sec; 98°C 10 sec, 60°c 30 sec and 72°C 3 min x35 cycles; 72°C 10 min

Table S2. Accession numbers and protein sequence of OdiMTs from different *O. dioica* populations.

OdiMT1

$OdiMT1_{NOR}$	CABV01001936.1	MDPVCSFRCCEENCAGCVDCPAGCDPCKCTLEVCKKVCEGCKDCPPGCEPCKCEKCSTKKCKSNCCPTSTAE
$OdiMT1_{CAT}$	MH547307	MDPVCSFRCCEENCAGCVDCPAGCDPCKFTMEVCKKVCEGCKDCPPGCEPCKCAKCSTKQCKSNCCPTSTAE
$\Psi OdiMT1_{CAT}$	MH577048	NCAGCVDCPAGCDPCKCTLEVCKKVCEGCKDCPPGCEPCKCAKCSTKQCKSNCCPTSTAE
$OdiMT1_{ORE}$	MH577047	MDSCSFRCCEENCTGCADCPAGCDPCKCTLGVCKKVCEGCADCPAGCDPCKCAKCSNKQCKTSCCPTSSSE
$OdiMT1_{OKI}$	SRA Project	MDCSTRCCEASCEGCVSCPPGCDPCKCSLDSCKKVCQGCRDCPPGCDPCKCSKCSANGCKCSCCAATNTSA
$OdiMT1_{OSA}$	SRA Project	MEFCSFRCCEGNCPGCADCPAGCDPCKCTLGVCKKVCEGCKDCAAGCDPCKCAKCSNKQCKTSCCPTSSSE

OdiMT2

$OdiMT2_{NOR}$	CABV01001042.1	MEVKRPNNCCPAKCLGCKGCPPGCEPCICNMDTCKNICNKCKECPKNEFGCDPCKCPKCSKLGCTCDCCHKKCCVTDCDG CKTCPPGCEPCKCSMNACKKVCKQCKNCRKSESGCDPCECSKCALKGCKCDCCPKDTCCEASCEGCKNCPPGCEPCKCTLN CCMKICDDCKDCPKSENGCDPCNCRKCSRKGCNCDCCPSDDCCKASCEGCINCPPGCDPCECSMDECKKICKKCNNCRKG ESGCDPCECRKCSRNGCDCDCCPKDSCCEASCEGCTDCPQGCKPCKCTMNSCMKTCDKCKDCPKSASGCDPCECLKCSRK GCECDCCPQKNDCCEAFCQGCKNCPPGCNPCKCTLNFCAKICNECKDCPKSDIGCDPCNCVKCSAKGCKCDCCPKKCC
$OdiMT2_{CAT}$	MH577044	MEVKRPNNCCPAKCLGCKGCPPGCEPCICKMDTCKNICNKCKECPKNEFGCDPCKCPKCSKLGCTCDCCHKKCCVTDCDG SKTCPPGCEPCKCSMNACKKVCKQCKNCRKSESGCDPCECSKCALKGCKCDCCPKDTCCEVSCEGCKNCPPGCEPCKCTM NSCMKTCDKCKDCPKSASGCDPCECRKCSRKGCDCDCCPQKNDSCEVFCQGCKNCPPGCNPCKCTLNFCAKICNECKDCP KSDTGCDPCYCVKCSAKGCKCDCCPKKCS
$\Psi OdiMT2_{CAT}$	MH577045	MEVKRLNNCCPAKCLGCKRCPPGCEPCICKMDTCKNICNKCARKTSSAATLASVLSAPSLDAHAIAATKNVVLPIAMDVR LAPRAVNHASAQ
$OdiMT2_{ORE}$	MH577046	MEVKRPNNCCPAKCLGCKGCPPACDPCICKMDSCKNICNNCKECPKDESGCDPCECPKCAKLGCTCACCHKKCCIAACDG CKTCPPGCEPCKCSMNACKKVCKQCKNCRKSESGCDPCECSKCASKGCKCDCCPKDNCCEASCEGCKNCPPGCEPCKCTM NGCMKTCDKCKDCPKSENGCDPCNCRKCSRKGCDCDCCPSDDCCKASCEGCTNCPPGCDPCECSMDACKKSCEKCKDCR KSESGCDPCECRKCSRKGCDCDCCPADSCCDASCDGCKNCPPGCEPCKCTMNGCMKTCDKCKDCPKSDSGCDPCECRKC SRKGCDCDCCPSDSCCEASCRGCKNCPPGCEPCKCTLNCCMKTCEKCKDCPKSASGCDPCKCGKCSKKGCDCDCCPKKND CCEAFCQGCKNCPPGCYPCKCTLNCCTKMCNECKNCPKSDTGCDPCKCAKCSTKGCKCDCCPKKCC
$OdiMT2_{OKI}$	SRA Project	MEVQQQRNNCCPAKCDGCKDCPPGCVPCKCLTDACTKICNNCKNCPPGCDPCGCEKCAKNGCTCDCCEQPCCETSCEG CKNCPPGCNPCKCEMEKCKKICKECKTCPKGENGCDPCQCQKCSKMGCKCSCCPKQPCCEVSCAGCKNCPPGCNPCKCE MASCKKICSGCADCPPGCDPCKCAKCSTKGCKCDCCPKKSCC
$OdiMT1_{OSA}$	SRA Project	MEVKRPNNCCIAACDGCKNCPPGCDPCKCSMKACKKVCEQCKNCRKSELGCDPCECSKCASEGCKCDCCPKDSCCEASCE GCKDCPPGCEPCKCTMNGCMKICDKCKDCPKSENGCDPCNCRKCSRKGCDCDCCPSDDCCKASCEGCINCPPGCDPCECS MDACKKSCEKCKDCRKSESGCDPCECRKCSRKGCDCDCCPADSCCEASCEGCKNCPPGCDPCKCTLNGCMKTCDKCKDCP KSDKGCDPCQCKKCSRRGCECDCCPKKNDCCEAFCQGCKNCPPECNPCKCTLNCCTKMCNECKNCPKSDTGCDPCKCAK CSTNGCKCDCCPKKCC

ARTICLE VI: *Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions and functional convergences.*

Molecular Biology and Evolution, msab184

DOI: 10.1093/molbev/msab184

Tunicates Illuminate the Enigmatic Evolution of Chordate Metallothioneins by Gene Gains and Losses, Independent Modular Expansions, and Functional Convergences

Sara Calatayud,[1] Mario Garcia-Risco,[2] Òscar Palacios,[2] Mercè Capdevila,[2] Cristian Cañestro,[1] and Ricard Albalat*,[1]

[1]Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

[2]Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

*Corresponding author: E-mail: ralbalat@ub.edu.

Associate editor: Naruya Saitou

Abstract

To investigate novel patterns and processes of protein evolution, we have focused in the metallothioneins (MTs), a singular group of metal-binding, cysteine-rich proteins that, due to their high degree of sequence diversity, still represents a "black hole" in Evolutionary Biology. We have identified and analyzed more than 160 new MTs in nonvertebrate chordates (especially in 37 species of ascidians, 4 thaliaceans, and 3 appendicularians) showing that prototypic tunicate MTs are mono-modular proteins with a pervasive preference for cadmium ions, whereas vertebrate and cephalochordate MTs are bimodular proteins with diverse metal preferences. These structural and functional differences imply a complex evolutionary history of chordate MTs—including de novo emergence of genes and domains, processes of convergent evolution, events of gene gains and losses, and recurrent amplifications of functional domains—that would stand for an unprecedented case in the field of protein evolution.

Key words: **metallothionein domains, modular proteins, *Chordata*, *Tunicata*, ascidians/thaliaceans/appendicularians, metallothionein evolution.**

Introduction

Heavy metals such as zinc (Zn) and copper (Cu) are integral components of numerous proteins involved in multiple cellular functions (reviewed in Vallee and Falchuk [1993]; Maret [2013]; Vest et al. [2013]), whereas other metals such as cadmium (Cd) have no known biological functions and are highly toxic for living beings. Organisms have developed different molecular mechanisms that regulate the homeostasis of the physiological metals and counteract the harmful effects of the toxic ones. One of such mechanisms is based on an extremely diverse group of metal-binding proteins known as metallothioneins (MTs), which are engaged in the physiological control of metals operating as ion reservoirs, metal transporters and/or metal deliverers to target metalloproteins, but also in radical scavenging, oxidative stress protection, and antiapoptotic defense (reviewed in Capdevila et al. [2012]; Ziller and Fraissinet-Tachet [2018]).

Most MTs are low molecular weight (<100 amino acids) and cysteine-rich (≈15–30%) proteins. Their cysteine (Cys, C) residues are arranged in CxC, CC, and CCC motifs, whose number and distribution define functional domains. Many MTs have a bimodular organization made of two functional domains, in which each domain enables the coordination of a number of metal ions through metal–thiolate bonds (reviewed in Blindauer [2014]). MTs have been classified in a stepwise gradation, from extreme Zn/Cd-thioneins to extreme Cu-thioneins depending on their metal-binding preference (Bofill et al. 2009; Palacios, Atrian, et al. 2011). There have been attempts to predict the metal preference of a given MT from the analysis of their non-Cys amino acids, and in the case of gastropod MTs, for instance, it has been proposed that the lysine-asparagine (*K*/*N*) ratio could be important for metal preference, so that a preponderance of *K* over *N* residues would be a feature of Cd selectivity, whereas a predominance of *N* over *K* would be typical of Cu-MTs (Perez-Rafael et al. 2014; Pedrini-Martha et al. 2020). Due to the high sequence and structural diversity of MTs, however, it is unclear whether the rule of *K*/*N* ratio applies to nongastropod MTs, and thereby, metal-binding assays are needed to experimentally determine the metal preference of newly discovered MTs.

Comparative analyses of the MTs across diverse species have been useful to better understand the function and evolution of these proteins, and to illustrate unusual modes of

Article

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

molecular evolution of these genes and proteins. MT research has exposed, for instance, pervasive events of internal sequence duplications (Tanguy and Moraga 2001; Palacios, Espart, et al. 2014; Iturbe-Espinoza et al. 2016; Jenny et al. 2016; Calatayud et al. 2018), de novo emergence of domains (Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021), and a high bias in the amino acid composition probably as a result of the 3D structural requirements for metal binding. MT research has also provided useful information for many other research fields. In the field of Evolutionary Ecology and Environmental Sciences, for instance, MTs are considered a good model to investigate the mechanisms by which different organisms—from protists and fungi to insects, mollusks, and mammals—have adapted to diverse metal bioavailability and other environmental stress conditions. These mechanisms include duplications of *MT* genes (Maroni et al. 1987; Adamo et al. 2012; Luo et al. 2020), increases in the levels of MT expression (Timmermans et al. 2005; Janssens et al. 2008, 2009; Costa et al. 2012; Catalan et al. 2016; de Francisco et al. 2018), and changes in metal specificity (Tio et al. 2004; Palacios, Pagani, et al. 2011; de Francisco et al. 2017; Dallinger et al. 2020; García-Risco et al. 2020). In the fields of Bioinorganic Chemistry and Metallomics, MTs have been used to analyze metal–protein interactions, analyzing the qualitative and quantitative characteristics of the metal-binding capacities of these proteins (Blindauer and Leszczyszyn 2010; Palacios, Pagani, et al. 2011; Beil et al. 2019) and by extension, of a broad range of metalloproteins (Waldron et al. 2009).

In this work, we aim to explore the MTs found in a large number of species of the three tunicate classes—Ascidiacea, Thaliacea, and Appendicularia—in order to investigate the intriguing origin and evolution of MTs in chordates. Tunicates (*Tunicata* subphylum, a.k.a. *Urochordata*) are the sister group of vertebrates, and together with basally branching cephalochordates form the phylum *Chordata*. Tunicates have undergone a rapid and "liberal" pattern of evolution in comparison with the "conservative" pattern of vertebrates and cephalochordates (Somorjai et al. 2018; Ferrández-Roldán et al. 2019), yielding a group of ecologically diverse filter-feeding marine animals—including planktonic and benthic specimens as well as solitary and colonial forms–, which have adapted to different conditions of metal bioavailability. In addition, the filtering lifestyle makes tunicates prone to accumulate metals from the seawater (Papadopoulou and Kanias 1977; Tzafriri-Milo et al. 2019; and references therein), and thereby, molecular mechanisms such as the MTs that regulate metal homeostasis and detoxification are physiologically relevant to them.

Our understanding of the MTs in tunicates is, however, scarce, limited to two MT sequences from the ascidian *Ciona robusta* (formerly *C. intestinalis*) and *Herdmania curvata* (Franchi et al. 2011), and two sequences from the appendicularian *Oikopleura dioica* recently described by our group (Calatayud et al. 2018). To overcome this situation, we have conducted an exhaustive survey of the MTs in tunicate databases, identifying 168 MTs in 37 species of ascidians, 4 of thaliaceans, and 3 of appendicularians, representing planktonic and benthic species, as well as solitary and colonial forms. We have also functionally characterized the metal-binding properties of eight tunicate MTs, which were selected on the basis of their taxon representativeness, sequence composition, degree of multiplicity, or modular organization. Comparative analyses of tunicate MTs with those of cephalochordates and vertebrates have exposed a high degree of structural diversity between lineages, which would not have been accompanied, however, by differences in their metal preferences. This diversity seems the result of numerous and intricate evolutionary events, and therefore, chordate MTs would represent one of the most complex and enigmatic examples of protein evolution.

Results

The biological interest for tunicates in the last decades has given rise to many genomic and transcriptomic resources, facilitating the in silico identification of virtually any gene in a wide set of species. This advantageous situation allowed us to perform an extensive survey of the MT system in tunicate databases.

MTs in Ascidians

Ascidians, commonly known as sea squirts, are the most studied group of tunicates. Adults are sessile, either solitary or colonial, usually found in shallow water worldwide. We surveyed ascidian databases and identified 145 MT sequences from 37 species distributed among Aplousobranchia (three species), Phlebobranchia (eight species), and Stolidobranchia (3 molgulidae + 5 pyuridae + 18 styelidae species) orders (fig. 1; see supplementary table S1, Supplementary Material online, for the species list). Sequence comparison showed that most ascidian MTs are 38–45 amino acids long with 12 cysteines (31–27%) organized in a single functional domain. In this domain, the C-motifs were distributed as $[CxC]x_5[CxC]x_2Cx_2Cx_2[CxC]x_4[CxC]x_{3\text{-}6}CC$, which would represent the prototypical organization of the monomodular ascidian MTs (supplementary fig. S1, Supplementary Material online, and table 1).

Twenty-seven out of the 37 analyzed species showed multiple MT genes: seven species with two sequences, six with three, two with four, three with six, two with seven, four with eight, one with nine, and two species with up to ten sequences. Despite in some cases, due to the nature of the data source—mainly sequence read archive (SRA) projects–, allelic variants of a particular gene could not be discriminated from gene duplicates, overestimating thereby the degree of MT multiplicity in our catalog, our results suggested that, overall, the MT multiplicity in ascidians is pervasive, although a definitive evaluation of the degree of multiplicity must await more complete and deeper genomic databases.

Interestingly, there were some exceptions to the prototypical organization, ranging from MTs that have lost or gained some cysteines or small protein fragments (supplementary fig. S1, Supplementary Material online) to multimodular forms containing tandem repeats of the prototypical domain (supplementary figs. S2 and S3, Supplementary Material online). We were able to reconstruct 12 multimodular MTs (five partial

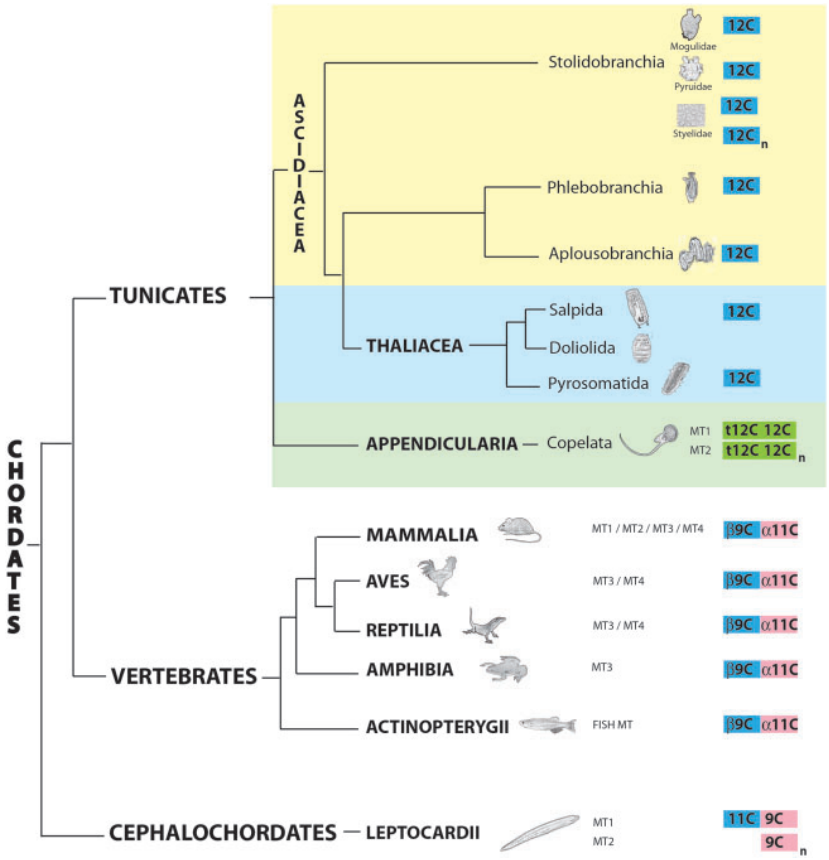


Fig. 1. Phylogenetic relationships in the *Chordata* phylum. Regarding tunicates, traditional classifications divided this subphylum in three classes, Appendicularia (green background), Ascidiacea (yellow background), and Thaliacea (blue background), but recent analyses propose Thaliacea species (salps, doliolids, and pyrosomes) are nested within the Ascidiacea class, closer to the Aplousobranchia and Phlebobranchia orders than to the Stolidobranchia order (Delsuc et al. 2018; Kocot et al. 2018). A schematic representation of the MTs of each taxonomic group, with a mono-, bi-, or multimodular organization with distinct domains (color coded) containing seven, nine, 11, or 12 cysteines, is showed at the right (see fig. 7 and text for additional details).

Table 1. Cysteine Motifs in 11/12-Cys Domains from Different Animal Taxa.

Taxon	11/12-Cys Domains	Cysteine Motifs	Metal Ions
Ascidians/thaliaceans	**Full-length MT**	$[CxC]x_5[CxC]x_2Cx_2Cx_2[CxC]x_4[CxC]x_{3-6}CC$	**4 (this work)**
Appendicularians	**12C domain**	$Cx_3Cx_2Cx_2Cx_3Cx_2[CxC]x_2Cx_4[CxC]xCC$	**4 (Calatayud, Garcia-Risco, Capdevila, et al. 2021)**
Vertebrates	**α-domain[a]**	$CCxCCx_3Cx_2Cx_3[CxC]x_{4-12}Cx_{1-3}CC$	**4 (Stillman et al. 1987)**
Cephalochordates	**11C domain**	$[CxC]x_5[CxC]x_{3-4}[CxC]x_3[CxC]x_3[CxC]x_2C$	**nd**
Mollusks	**α-domain**	$[CxC]x_{4-6}[CxC]x_3Cx_{4-5}[CxC]x_3[CxC]x_3[CxC]x_2C$	**4 (Digilio et al. 2009)**
Echinoderms	**α-domain**	$CxCCx_5[CxC]x_4CCx_4CCx_4CC$	**4 (Tomas et al. 2013)**
Insects	**MTnB/C/D**	$Cx_2Cx_3[CxC]x_4Cx_3[CxC]x_3[CxC]xC]x_7CC$	**4 (Egli et al. 2006)**

Note.—nd, not determined.
[a]Notice that although vertebrate, echinoderm, and mollusk MTs have α-domains, the origin and evolutionary relationship of these α-domains remains uncertain.

and seven full-length sequences), whose number of repeats ranged from two to nine (fig. 2). Noteworthy, all the species with multimodular MTs belonged to the styelidae family in the stolidobranchia order (fig. 1), suggesting that multimodular MTs arose by a lineage-specific evolutionary event.

Metal-Binding Features of Mono-Modular Ascidian MTs

In order to analyze the metal-binding capacity and preference of the ascidian MTs, we selected three MTs of two species belonging to distinct orders within the Ascidiacea class: CroMT1 from *C. robusta* (phlebobranchia order), and HroMT1 and HroMT2 from *Halocynthia roretzi* (stolidobranchia order). These MTs were chosen because CroMT1 was the only MT in *C. robusta*, whereas HroMT1 and HroMT2 represented duplicate forms in *H. roretzi*. These MTs were also selected because of their *K*/*N* ratios, which in the CroMT1 was 3:5, opposite of that in HroMT1 and HroMT2 sequences, 6:3 and 8:2, respectively. HroMT2 showed indeed one of the highest *K*/*N* ratios among tunicate MTs (supplementary table S1, Supplementary Material online). The *K*/*N* ratio seems important for metal preference of gastropod MTs (Perez-Rafael

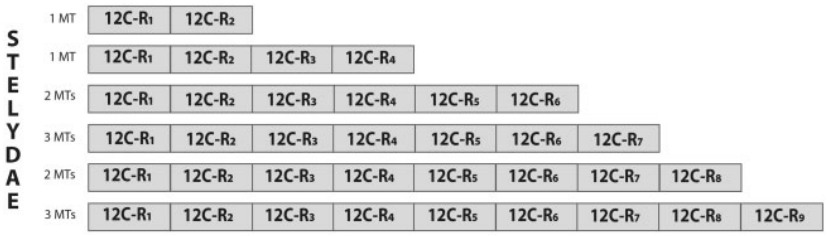


Fig. 2. Schematic representation of multimodular Stelydae MTs. Internal duplications of a 12-Cys domain (grey box) generated multimodular MTs with variable number of repeats (R), from 2 to 9. The number of MTs with different number of repeats (R_1–R_9) is indicated (DgrMT7 = 2 repeats; EtiMT8part ≥4 repeats; ScaMT6part and SsoMT8part ≥6 repeats; PmiMT10part, BleMT1, and BleMT2 ≥7 repeats; PauMT10part and PspMT7 ≥8 repeats; BscMT1, PmamiMT3, and PpoMT7 = 9 repeats) (see supplementary fig. S2, Supplementary Material online, for further details).

et al. 2014; Pedrini-Martha et al. 2020), and the study of the metal selectivity of the CroMT1, HroMT1, and HroMT2 would allow us to experimentally evaluate the *K*/*N* rule in the ascidian system.

We characterized the formation of metal–MT complexes of CroMT1, HroMT1, and HroMT2 proteins heterologously expressed in *Escherichia coli* cultures supplemented with Cu, Cd, or Zn salts by inductively coupled plasma atomic emission spectroscopy (ICP-AES) and electrospray ionization mass spectrometry (ESI-MS). ICP-AES was used for protein quantification and metal-to-protein stoichiometry determination through the measurement of element composition of the samples (S, Zn, Cd, and Cu) (Bongers et al. 1988), and ESI-MS allowed us to determine the molecular mass of the species formed, that is, the speciation of the samples (Capdevila et al. 2012). Single well-folded Cd_4-MT species were recovered from CroMT1, HroMT1, and HroMT2 Cd-productions (fig. 3*A*–*C*), and a major Zn_4-MT species were found in their recombinant Zn productions (supplementary fig. S4*A*, *C*, and *E*, Supplementary Material online), which contrasted with the mixtures of homometallic Cu-MT species purified from Cu-supplemented *E. coli* cultures (supplementary fig. S4*B*, *D*, and *F*, Supplementary Material online). It is also noteworthy the resemblance of the CD spectra of the Cd–MT complexes of all ascidian MTs, denoting structurally equivalent metal clusters for these proteins (fig. 4). The exemplary exciton coupling centered at ∼250 nm exhibited in all the CD envelopes confirms high robustness and compactness of the $Cd_4S(Cys)_{12}$ clusters and, therefore, high Cd specificity. Altogether, these results showed that Zn_4- or Cd_4–MT complexes were energetically the most favorable forms of these proteins, suggesting that their function would be related to binding divalent metal ions, either as Zn- or Cd-thioneins [8]. Genuine Zn-thioneins, however, render heterometallic Cu, Zn–MT complexes when synthesized under Cu surplus, meaning that the ascidian MTs deviated toward a more specific Cd-thionein behavior (Palacios, Atrian, et al. 2011). Regarding the *K*/*N* rule used to predict the metal-binding preference in gastropod MTs, our results demonstrated that this rule did not apply to tunicate MTs, since the three of them displayed contrasted *K*/*N* ratios and their metal preferences were exactly the same.

Metal-Binding Features of Multimodular Ascidian MTs

In order to determine the metal-binding features of multimodular MTs, we selected BscMT1 (ID: ATSW01006707.1; chromosome 6) of *Botryllus schlosseri* (stolidobranchia order) because this MT was not only identified in an accurately assembled genome (Voskoboynik et al. 2013), but also because it was supported by three EST sequences (supplementary table S1, Supplementary Material online). BscMT1 was a multimodular MT of 365 amino acids with 105 cysteines (29%) and a *K*/*N* ratio of 1:1 (31*K* vs. 31*N*) (supplementary fig. S2 and table S1, Supplementary Material online). BscMT1 was one of the longest MT described so far, and a detailed analysis of its sequence revealed that it was made of nine tandem repeats (R_1–R_9, the last one only partially complete) of a 12-Cys domain, each one resembling a prototypical ascidian form (supplementary fig. S3, Supplementary Material online).

The metal–BscMT1 complexes purified from the heterologous expression in *E. coli* cultures supplemented with Cu, Cd, or Zn salts were characterized by the IPC-AES and ESI-MS. The concentrations of metal–BscMT1 productions, despite the several attempts, were always low, probably due to the high molecular weight of BscMT1 and its structural complexity made of repeated domains. Only the samples recovered from the Zn(II)-enriched culture media rendered informative data through these techniques. The IPC-AES and ESI-MS analyses showed that this preparation contained Zn_{36}–MT complexes (fig. 3*D*). The binding of 36 metal ions agreed with its multimodular organization made of nine repeated domains, each one similar to a prototypical M_4-MT: 4 metal ions×9 domains = 36 metal ions. To further verify the functional correspondence between one of the 12-Cys tandem repeats of BscMT1 and the mono-modular ascidian MTs, we analyzed the ability of the R_4 domain (R_4-BscMT1) to form metal complexes when it was expressed alone. The election of the R_4 domain was due to the fact that it was highly similar to the prototypical ascidian MTs, including some additional amino acids after the carboxyl-terminal CC doublet (supplementary fig. S3, Supplementary Material online). We heterologously expressed the R_4 domain in *E. coli* grown in media supplemented with Cu, Cd, or Zn salts. The ICP-AES and ESI-MS analyses of the recovered samples confirmed the ability of this fragment to autonomously form metal complexes by itself,

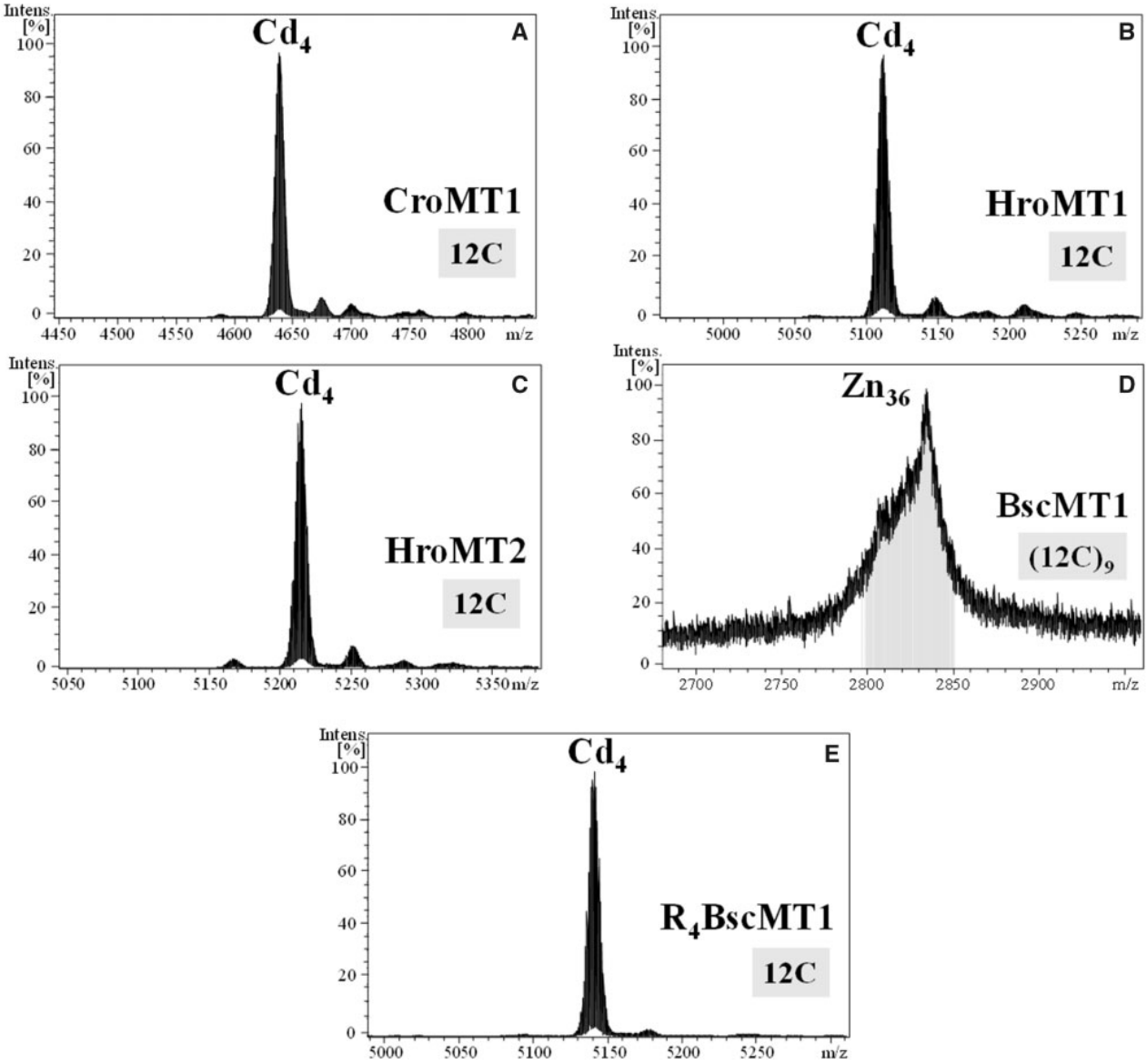


Fig. 3. Deconvoluted ESI-MS spectra of the Cd(II)-productions of ascidian CroMT1 (*A*), HroMT1 (*B*), HroMT2 (*C*), and +14 charge state ESI-MS spectra of the Zn(II)-production of BscMT1 (*D*), as well as the deconvoluted ESI-MS spectra of the Cd(II)-production of its fragment R_4 BscMT1 (*E*). Notice that BscMT1 is capable to bind up to 36 divalent ions, which nicely match with a multimodular organization made of nine repeated domains $(12C)_9$, each one binding four divalent ions alike the mono-modular (12C) MTs (A–C) or the R_4 (12C) domain (E).

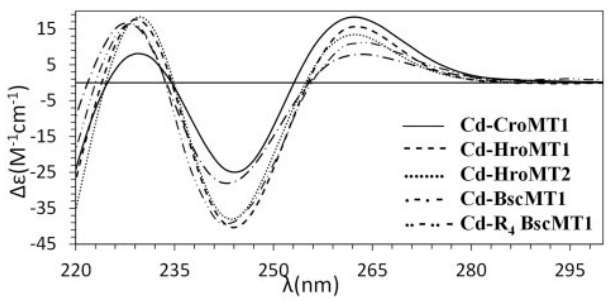


Fig. 4. Circular dichroism spectra of the Cd(II)-productions of CroMT1, HroMT1, HroMT2, BscMT1, and R_4 BscMT1, all exhibiting very similar CD signals. The exciton coupling centered at ~250 nm confirms high robustness and compactness of the clusters formed.

and revealed that the biochemical features of R_4 domain were highly similar to the mono-modular ascidian forms (compare fig. 3A–C and *E*). Thus, this domain yielded major M_4–R_4 complexes with Zn(II) and Cd(II), being unique species only in the latter, whereas rendered a homometallic mixture of species in the presence of copper, as the ascidian MTs discussed above (supplementary fig. S4*G* and *H*, Supplementary Material online). Based on this R_4 fragment as a representation of the whole protein, the multidomain BscMT1 would be a Cd-thionein capable to bind up to 36 divalent ions, and its stoichiometry—that is, M_{36}-BscMT1—would nicely match with what would be expected from the repetition of nine domains, that is, M_4-R_1 to M_4-R_9. These results supported a direct relationship between the binding capacity and the number of the repeated domains in multimodular MTs.

MTs in Thaliaceans

We analyzed the thaliacean MTs by surveying the databases of the five available species. We identified 14 MTs from four thaliacean species, three from Salpida (*Iasis cylindrica*, *Salpa fusiformis*, and *Salpa thompsoni*), and one from Pyrosomata (*Pyrosomella verticillata*) orders. Sequence comparison showed that thaliacean MTs were 36–41 amino acids long, with 12 cysteines (33–29%) organized in a single functional domain with the C-motifs distributed as $[CxC]x_5[CxC]x_2Cx_2Cx_2[CxC]x_4[CxC]x_{3-4}CC$ (noteworthy, we did not find multimodular MTs in thaliacean databases) (table 1). This motif arrangement was identical to that of the prototypical ascidian MTs, denoting a common origin of that thaliacean and ascidian MTs, and supporting the recently suggested phylogenetic relationship of these classes (Delsuc et al. 2018; Kocot et al. 2018) (fig. 1).

Importantly, MT multiplicity was found in the four thaliacean species although as in ascidians, it was challenging to determine whether the multiple sequences represented different genes or allelic variants. In *S. thompsoni*, for instance, we identified six MT sequences, five full-length (SthMT1, MT2, MT3, MT5, and MT6), and one partial (SthMT4), which coded for proteins 48% to 95% identical at amino acid level (supplementary table S2, Supplementary Material online). The highest identities were between SthMT1 and SthMT5 (88%) and SthMT2 and SthMT6 (95%) sequences, and therefore, these pairs were the most plausible cases of allelic variants. To investigate this possibility, we analyzed the conservation of 5′ and 3′ flanking regions, as well as of intron sequences in PCR-amplified genomic regions from two *S. thompsoni* specimens. Nucleotide identity (excluding gaps) in SthMT1–SthMT5 and SthMT2–SthMT6 alignments was 91–100%, 87–82% and 83–86% at 5′ regions, introns and 3′ regions, respectively (supplementary table S3, Supplementary Material online). These results would be compatible with that SthMT5 and SthMT6 were allelic variants of SthMT1 and SthMT2, respectively, although the possibility they may be recently duplicated genes cannot be fully discarded.

Metal-Binding Features of Thaliacean MTs

We analyzed the metal-binding capacity and metal preference of the thaliacean MTs, and compared them with those of ascidian MTs in order to investigate whether the differences in lifestyles—planktonic versus benthic—might have been key for MT evolution. We characterized the metal–MT complexes rendered for the four MTs from *S. thompsoni*, SthMT1 to SthMT4, heterologously expressed in Zn-, Cd-, or Cu-supplemented *E. coli* cultures. The clear Cd-thionein character of *S. thompsoni* MTs was reflected, once again, by the ICP-AES and ESI-MS analyses of the recovered samples. SthMTs yielded single Cd_4-MT species under Cd(II) surplus (fig. 5), as well as major M_4–MT complexes with Zn(II) (supplementary fig. S5*A–D*, Supplementary Material online). A mixture of species (for SthMT1 and SthMT3) or not detection of the metallated species (for SthMT2 and StMT4) were obtained in the presence of copper (supplementary fig. S5*E* and *F*, Supplementary Material online), evidencing that these proteins struggle to build metal clusters with Cu. Our results demonstrated that the four SthMTs were functionally very similar among them (compare panels *A–D* in fig. 5), as well as to the mono-modular ascidian MT (compare fig. 3 and fig. 5). The fact that thaliacean (SthMTs) and ascidian (CroMT1, HroMT1, HroMT2, and BscMT1) MTs exhibited a similar Cd-thionein character strongly supported that the prototypical ascidian/thaliacean MT (and most likely the ancestral tunicate MT, see below) was a mono-modular protein with a 12C domain capable to bind four metal ions with preference for Cd.

MTs in Appendicularians

We investigated the MTs in the Appendicularia (a.k.a. Larvacea) class. In a previous work, we had characterized the *OdiMT1* and *OdiMT2* genes (formerly *OdMT1* and *OdMT2*) of *O. dioica* (Calatayud et al. 2018). We showed that both genes encoded MTs very different from all MTs described so far because their cysteines motifs were arranged in a novel pattern forming a "repeat unit" (RU) made of two C7 subunits (C7a and C7b) plus a carboxyl-terminal tail with five cysteines (C5) (Calatayud et al. 2018) (supplementary fig. S6*A*, Supplementary Material online). We concluded that OdiMT1 was made of a single RU, whereas OdiMT2 was a multimodular protein made of six direct tandem RU (Calatayud et al. 2018). We wondered whether these unconventional MTs were specific of *O. dioica*—gene and protein sequences are known to have considerably diverged during *O. dioica* evolution (reviewed in Ferrández-Roldán et al. [2019])—or they were widespread in the appendicularian lineage. We took advantage of recent sequence projects of several appendicularian species (Naville et al. 2019) and identified seven new MT sequences in two additional *Oikopleura* species: *O. albicans* (three sequences) and *O. vanhoeffeni* (four sequences) (supplementary table S1, Supplementary Material online). We did not identify any putative MT in databases of other oikopleurids such as *O. longicauda*, *Bathochordaeus* sp., and *Mesochordaeus erythrocephalus*, nor in the fritillarid species *Fritillaria borealis*, although any conclusion about the loss of the MTs in these lineages must await deeper sequence projects.

Regarding *Oikopleura* MTs, although most of the sequences were partial, they were clearly similar to *O. dioica* MTs. Sequence analyses showed that these new MTs were multimodular proteins made of repeats of a conserved 12-Cys domain with C-motifs distributed as $CX_3Cx_2Cx_2Cx_3Cx_2[CxC]x_2Cx_4[CxC]xCC$ (table 1 and supplementary fig. S6*B*, Supplementary Material online). This 12-Cys domain corresponded to the previously described C7b+C5 subunits (Calatayud et al. 2018). OdiMT1 consist, therefore, of two 12-Cys domains, one full-length (12C = C7b + C5), and one 12-Cys domain that has been trimmed (t-12C, formerly known as C7a) and lacks the five last cysteines (C5) (supplementary fig. S6*A*, Supplementary Material online and fig. 7). Interestingly, the trimmed domain when expressed alone did not efficiently bind divalent metals by itself, suggesting that it has lost its capacity to independently work (Calatayud, Garcia-Risco, Capdevila, et al. 2021). OdiMT2 would be a multimodular MT made of twelve 12-Cys domains, whose

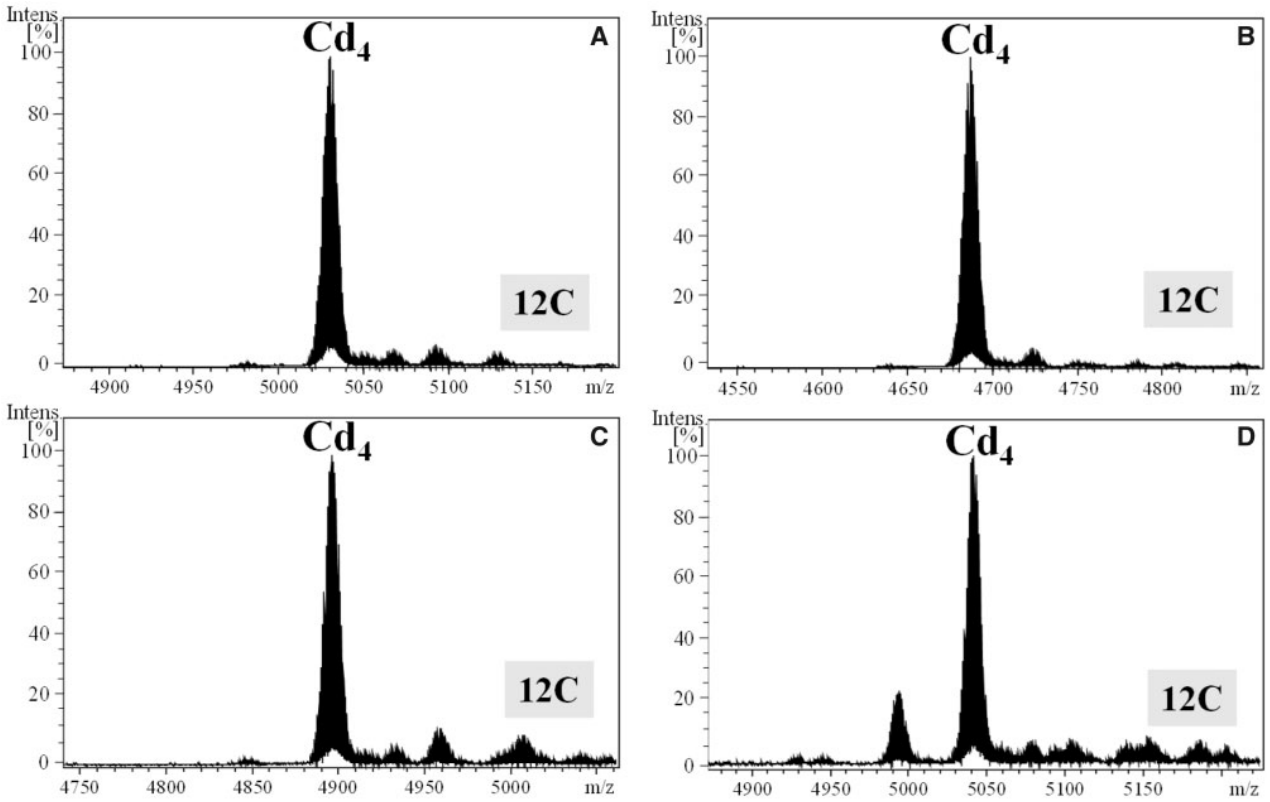


Fig. 5. Deconvoluted ESI-MS of the Cd(II)-production of *S. thompsoni* SthMT1 (*A*), SthMT2 (*B*), SthMT3 (*C*), and to SthMT4 (*D*). Thaliacean MTs are mono-modular proteins made of a single 12C domain capable to bind four divalent ions, identically to ascidian MTs.

odd domains would have been also trimmed as t-12C. The identification of this conserved 12-Cys domain in all *Oikopleura* MTs suggested that this domain appeared in the ancestor of the *Oikopleura* genus, and that it was duplicated generating multimodular MTs independently in the diverse appendicularian species (supplementary fig. S6*B*, Supplementary Material online).

MTs in Other Chordates

To provide a broad evolutionary perspective to our analyses we compared the tunicate MTs with those from the other chordate subphyla: cephalochordates and vertebrates.

In cephalochordates, MTs had been characterized in two amphioxus species, *Branchiostoma floridae* and *Branchiostoma lanceolatum* (Guirola et al. 2012). We extended the surveys of MTs to two additional amphioxus species, *Branchiostoma belcheri* (BbeMT1 and BbeMT2) and *Branchiostoma japonicum* (BjaMT1 and BjaMT2), and to *Asymmetron lucayanum* (AluMT1 and AluMT2), a cephalochordate distantly related to the other Branchiostoma species (Yue et al. 2014). We identified two MTs, MT1 and MT2, in all the cephalochordate species. *B. lanceolatum* MT2 showed two alternatively spliced forms, MT2-L (for the long form, with 67–70 amino acids) and MT2-S (for the short form, with 46–47 amino acids) (supplementary fig. S7, Supplementary Material online). Splice variants were also identified in *B. floridae, B. japonicum*, and *A. lucayanum* cDNA databases (i.e., BflMT2L and S, BjaMT2L and S, and AluMT2L and S), and alternative splicing was predicted from genomic sequence in *B. belcheri* species (BbeMT2L and S).

Sequence comparisons revealed that cephalochordate MT2s were multimodular proteins made of two (MT2S) or three (MT2L) 9-Cys domains ($CCx_2Cx_2Cx_3Cx[CxC]xCC$) depending on an alternative splicing on exon 2 (supplementary figs. S7 and S8*B*, Supplementary Material online). In contrast, MT1s were bimodular proteins with two different domains: a 11-Cys, N-terminal domain with the C-motifs distributed as $[CxC]x_5[CxC]x_{3-4}[CxC]x_3[CxC]x_3[CxC]x_2C$ (table 1 and supplementary fig. S8*A*, Supplementary Material online) linked by 9–11 residues to a domain similar to the 9-Cys domain of MT2 (we named this domain 9-Cys_{Like} because an Ile substitutes the fourth Cys: $CCx_2Cx_2Ix_3Cx[CxC]xCC$), followed by a carboxyl-terminal tail of 16–17 amino acids devoid of Cys residues. Interestingly, MT2 9-Cys and the MT1 9-Cys_{Like} domains showed a clear sequence similarity (supplementary fig. S8C, Supplementary Material online), which suggested that cephalochordate MT2s might have arisen from the duplication of an ancestral MT1 that lost the 11-Cys domain but amplified its 9-Cys domain.

In vertebrates, MTs had been classified in four different MT types: MT1, MT2, MT3, and MT4. Vertebrate MTs had been characterized in many species with significant differences of multiplicity in a species-dependent manner (Serén et al. 2014). Phylogenetic and synteny analyses had suggested the existence of mammalian MT1s and MT2s, amniote MT4s, tetrapod MT3s, and fish MTs (fig. 1) (Serén et al. 2014). The appearance of these four types would not be related with the two rounds

of whole-genome duplication events that occurred during early vertebrate evolution since they are not localized in paralogous chromosomes but are located in tandem on a same chromosome (Serén et al. 2014). At protein level, all vertebrate MTs (even an MT we identified in the early divergent group of hagfish *Eptatretus burgeri*), showed a conserved bimodular organization made of a 9-Cys N-terminal domain (the so-called β-domain: $[CxC]x_5[CxC]x_3[CxC]x_2[CxC]x_2Cx_3$) linked to a 11-Cys domain (the so-called α-domain: $CCxCCx_3Cx_2Cx_3[CxC]x_6CxCC$) at the carboxyl-terminal end (supplementary table S1 and fig. S9, Supplementary Material online). The arrangement of the C-motifs in the α and β domains did not show any obvious resemblance between them that might suggest of a common ancestry of these domains, though vertebrate β-domain appeared somewhat similar to cephalochordate and tunicate 11/12-Cys domains.

Discussion

Origin and Convergent Evolution of Tunicate MTs

Ascidian and thaliacean MTs share the same arrangement of their C-motifs (table 1) and are clearly similar in their amino acid sequence, also in the noncysteine residues: ascidian MTs are at least ≈40% identical in their noncysteine residues to any of the thaliacean MTs, and ≈60% identical when considering the full sequence. In contrast, their similarity with appendicularian MTs is negligible at the arrangement of the C-motifs (table 1) and in the composition of the noncysteine residues. For instance, OdiMT1 only is, at most, 13% identical to ascidian (e.g., CroMT1 and HroMT1) or thaliacean (e.g., SthMT1) MTs in noncysteine residues (notice that 25–30% sequence identity is generally taken as the minimal threshold for presumption of homology; Rost 1999). Therefore, although it cannot be ruled out that appendicularian MTs derive from ascidian and thaliacean MTs that have diverged too much for orthology to be recognized, their sequence and the C-motifs are so divergent that they seem to have had independent evolutionary origins. This possibility agrees with the idea that de novo emergence of MTs might be more habitual than previously thought because the only requirement for a peptide to function as a metal ion chelator would be a high content of coordinating residues (i.e., cysteines) and a relative small length that would favor its proper folding (Capdevila and Atrian 2011; Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021). The independent emergence of MTs in distant animal groups have been, indeed, implicitly suggested in diverse evolutionary studies (Capdevila and Atrian 2011; Blindauer 2014; Isani and Carpene 2014; Ziller and Fraissinet-Tachet 2018), and we have recently showed de novo evolution of MT domains within the Mollusca phylum (Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021).

Intriguingly, ascidian/thaliacean MTs and appendicularian MTs seem to have converged in the number of cysteines—12—of their basic functional organizations: the prototypical 12-Cys MTs of ascidians and thaliaceans, and the repeated 12-Cys domains of appendicularian MTs. Functional domains with 11/12 cysteines organized in different patterns of C-motifs but converging to similar 3D architectures (Beil et al. 2019) appear a recurrent evolutionary solution in MTs of phylogenetically distant animal clades, from vertebrates and tunicates to mollusks and insects (table 1). 3D structural constraints to efficiently coordinate four metal ions through the formation of metal–thiolate bonds might have led to the convergent evolution of 11/12-Cys domains in MTs of different animal phyla (Stillman et al. 1987; Egli et al. 2006; Digilio et al. 2009; Tomas et al. 2013; Calatayud, Garcia-Risco, Capdevila, et al. 2021; and this work).

Evolution of the Metal-Binding Preference

Ascidian/thaliacean MTs and appendicularian MTs also share the Cd-thionein character. The evolution of Cd-MTs seems, indeed, a common adaptive event in marine animals, including mollusk (Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021), crustacean (Narula et al. 1995; Valls et al. 2001; Munoz et al. 2002), echinoderm (Riek et al. 1999), and chordate species (Guirola et al. 2012; Calatayud et al. 2018). The preference of tunicate MTs for Cd further support the idea that ancestral MTs arose and evolved probably driven by the Cd(II) concentrations in the seawater because Cd is highly toxic by interfering in Zn-dependent cellular processes (Dallinger et al. 2020; Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021). In addition to sporadic episodes of increases in Cd emissions throughout geological eras (from Paleozoic to Mesozoic and Cenozoic; Dallinger et al. [2020] and references therein), Cd is frequently found with Zn in ore deposits of the earth crust. The natural Cd(II) concentrations are, therefore, higher where seawater comes into contact with the continental earth crust (≈0.8 nmol/kg) than in surface pelagic waters (≈0.002 nmol/kg) (Bruland 1980). Cd(II) concentrations are also higher in eutrophic coastal regions of the littoral and its close neritic zone (≈0.16 nmol/kg) (Bruland 1980), which are usual habitats for tunicates. It is noteworthy that all tunicate MTs analyzed in this work show a similar sequence and metal preference, either they are from benthic ascidians or from planktonic thaliaceans, or from solitary (*C. intestinalis* and *H. roretzi*) or colonial species (*B. schlosseri*). This similarity suggests that the different lifestyles of tunicates might have not been a significant factor in the evolution of metal-binding preference.

Evolution of MT Multiplicity and Multimodular MTs

Increasing the metal-binding capacity enhance the physiological capabilities of the organisms to adapt to diverse conditions of metal bioavailability and other environmental stress situations (Jenny et al. 2016; Baumann et al. 2017). In a simplified view, such increments might be attained through two strategies: increasing the number of *MT* genes (i.e., augmenting MT multiplicity) (Maroni et al. 1987; Adamo et al. 2012; Luo et al. 2020) or increasing the number of metal-binding domains in a given MT (i.e., generating multimodular MTs) (Jenny et al. 2016; Baumann et al. 2017). Tunicates have used both strategies.

Evolution of MT Multiplicity

Multiplicity in the MT system has been associated to neofunctionalization or subfunctionalization processes, in which

different paralogs have distinct metal-binding preferences and/or expression patterns. For instance, in mollusks, several MTs can be found in a given species, each one with distinct metal selectivity—that is, Cd-MT, Cd/Cu-MT, and Cu-MT—and differential transcription patterns (Palacios, Perez-Rafael, et al. 2014; Dvorak et al. 2018; Dallinger et al. 2020). Similarly, in the mammalian system, ubiquitous metal-induced forms MT1 and MT2 have preference for Zn(II), whereas restrictedly expressed MT3 and MT4 have intermediate Zn/Cu- and Cu-thionein characters (Capdevila and Atrian 2011; Vasak and Meloni 2011; Artells et al. 2013). MT multiplicity has been observed in 36 tunicate species, patchily distributed among the three tunicate classes. However, in contrast with the observed in the MT system of other animal groups, the analyzed tunicate MT paralogs do not seem to have essentially diverged in their metal-binding selectivity. The two paralogs of the ascidian *H. roretzi* (HroMT1 and HroMT2), the four paralogs of the thaliacean *S. thompsoni* (SthMT1 to SthMT4), and the two paralogs of the appendicularian *O. dioica* (OdiMT1 and OdiMT2; Calatayud et al. 2018) have a conserved Cd selectivity. These results suggest that tunicate paralogs have undergone a subfunctionalization process partitioning the expression of the predecessor, rather than a neofunctionalization event acquiring new metal preferences. Future detailed spatiotemporal analyses of the transcription patterns of each paralog will provide a definitive answer to this question, and will allow us to understand the ecophysiological needs that led the evolution of multiple MTs in these species.

Evolution of Multimodular MTs

Evolution of multimodular MTs seems a recurrent solution for creating large MTs with high cysteine content and a high capacity of metal binding (Niederwanger et al. 2017; Palacios et al. 2017; Calatayud et al. 2018; and this work). Multimodular MTs have been found in very diverse organisms, from fungi (Palacios, Espart, et al. 2014; Iturbe-Espinoza et al. 2016) and mollusks (Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021; and references therein), to chordates (Guirola et al. 2012; Calatayud et al. 2018). Here, we have identified multimodular MTs in 20 species of ascidian and appendicularian classes (thaliaceans appear devoid of multimodular MTs). These multimodular MTs are made of tandem repeats of a prototypical organization within each tunicate class—$[CxC]x_5[CxC]x_2Cx_2Cx_2[CxC]x_4[CxC]x_{3-6}CC$ for ascidian and $CX_3Cx_2Cx_2Cx_3Cx_2[CxC]x_2Cx_4[CxC]xCC$ for appendicularian MTs–, implying independent evolutionary origins for ascidian and appendicularian multimodular MTs. In ascidians, the multimodular MTs arose during the evolution of the Styelidae family in the Stolidobranchia order, after its split from the Pyuridae family, 277 Ma (Delsuc et al. 2018). In appendicularians, multimodular MTs arose after its divergence from the other tunicate classes, 447 Ma (Delsuc et al. 2018). The adaptive causes that favored the evolution of large MT in some benthic ascidian species but not in others, or in some planktonic tunicates such as appendicularians but not in thaliaceans, are still an unsolved question.

The Enigmatic Origin of the MTs in Chordates

The extensive catalog of tunicate MTs collected in this work together with the analysis of cephalochordate and vertebrate MTs has allowed us to investigate the origin and evolution of the MTs in the chordate phylum. This has been a particularly difficult challenge because when comparing distant taxonomic groups (e.g., at phylum, subphylum, or class level), one has to be cautious with the MT sequence alignments. The general cysteine richness of the MTs distorts the comparisons by forcing the alignment of Cys positions and inserting many gaps in the aligned sequences. This distortion results in a false impression of similarity between nonhomologous MT sequences, and if the alignments are used for phylogenetic reconstructions, they can lead to wrong evolutionary inferences. To circumvent this drawback, we focused our analyses on the patterns of the cysteine motifs (CxC, CC, and CCC) in the MT domains that, as it has been proved for other molecular markers (Rokas and Holland 2000), might have enormous potential as a tool for phylogenetic inferences.

Regarding the ancestral chordate MT, it is generally accepted that due to their substantial genomic stasis (Cañestro and Albalat 2012; Somorjai et al. 2018; Coppola et al. 2019), cephalochordate genes and genomes are more similar to those of the chordate ancestor than those of tunicates and vertebrates (Louis et al. 2012). It can be argued, therefore, that the MT of the chordate ancestor was a bimodular protein with an amino-terminal 11/12-Cys domain and a carboxyl-terminal 9-Cys domain (fig. 6), resembling the organization of the current cephalochordate MT1s. The similarity of cephalochordate and vertebrate domains (fig. 7) supports the bimodular organization of the ancestral chordate MT. In addition, the fact that cephalochordate MT1s are also similar to the α and β1 domains of mollusks (Calatayud, Garcia-Risco, Pedrini-Martha, et al. 2021) would push back in evolutionary time this MT organization to the origins of the bilaterian animals.

The evolutionary scenario of a chordate MT ancestor similar to the current cephalochordate MT1 implies a number of evolutionary events independently shaping the MTs in each chordate subphylum: (I) In cephalochordates, an ancient tandem duplication of the ancestral MT followed by the loss of the 11-Cys domain and the amplification of the 9-Cys domain in one of the copies would have led the MT2 forms of the subphylum (fig. 6); (II) In vertebrates, the ancestral MT would have readjusted its Cys motifs, leading to the current amino-terminal β- and the carboxyl-terminal α-domains (fig. 6). Successive lineage-specific tandem duplications would have led to a variable degree of the MT multiplicity in the different vertebrate species (Serén et al. 2014); (III) In tunicates, the ancestral MT would have lost the 9-Cys domain in the ascidian/thaliacean lineage. This mono-modular 12-Cys MT was patchily duplicated, giving rise to a variable number of MTs in different ascidian/thaliacean species, or internally repeated in the multimodular MTs of styelidae species (fig. 6); (IV) Finally, in the fast–evolving appendicularians, the ancestral chordate MT would have been lost, and a new MT with a different pattern of Cys motifs seems to have evolved de novo (fig. 6).

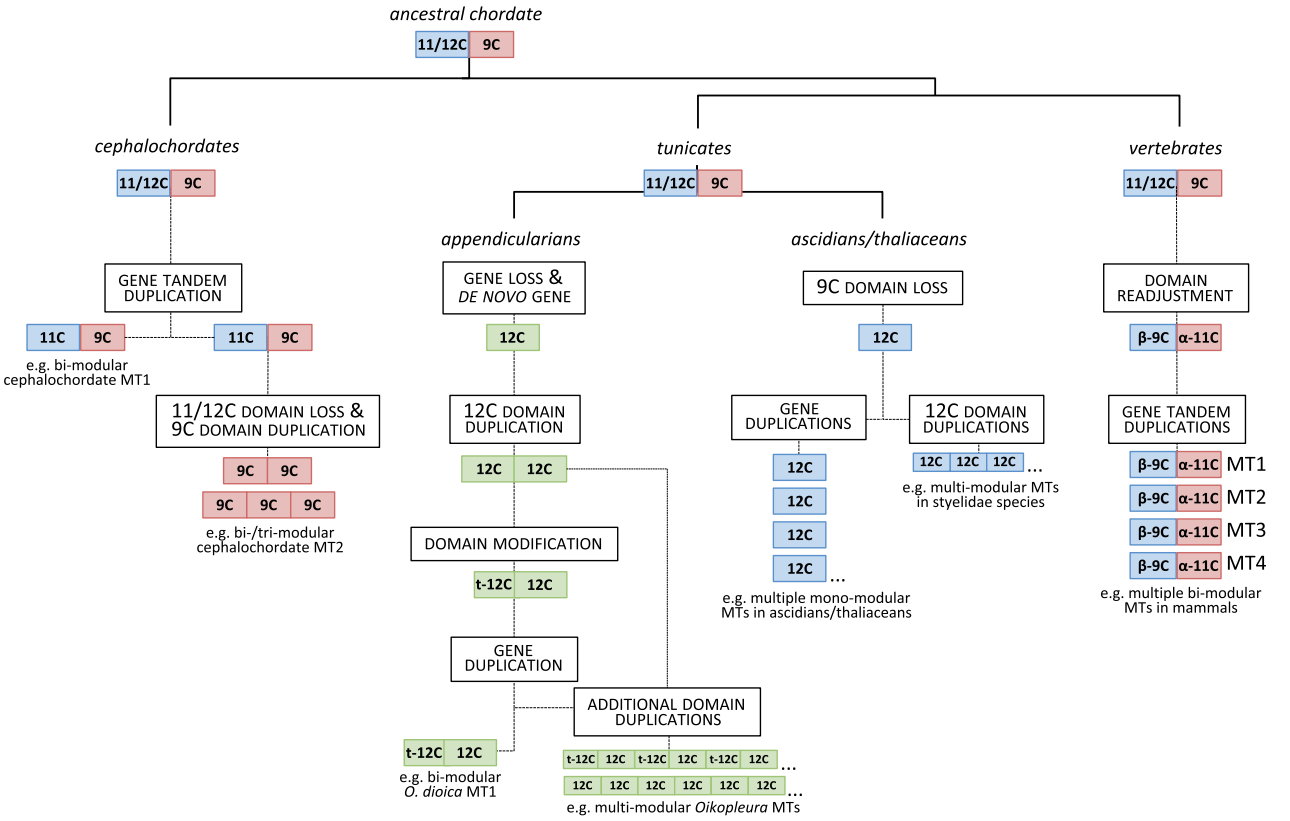


Fig. 6. Reconstruction of MT evolution in the *Chordata* phyla. The most parsimonious scenario would be that the ancestral chordate MT was a bimodular MT made of an N-terminal 11/12-Cys domain (in blue) and a C-terminal 9-Cys domain (in red), similar to the current cephalochordate MT1. In cephalochordates, this ancestral form was tandem duplicated. One copy lost the 11/12-Cys domain and expanded the 9-Cys domain, yielding the current bi/trimodular MT2 form. In vertebrates, the number of cysteines in each domain was readjusted to the current 9 Cys of the N-terminal β-domains, and 11 Cys of the C-terminal α domain. Successive lineage-specific tandem duplications of this primeval vertebrate MT led to different MT types (MT1, MT2, MT3, and MT4) with a variable degree of multiplicity in different species. In tunicates, the ancestral form lost the C-terminal 9-Cys domain in ascidian/thaliacean lineages, leading a mono-modular. This MT was frequently duplicated generating a high degree of MT multiplicity in many ascidian and thaliacean species. In contrast, internal duplications led to the emergence of multimodular MTs in the stelydae lineage. Finally, in the appendicularian clade, the ancestral MT appears to have been lost, and a new MT would have emerged de novo by the duplication of a 12-Cys domain (in green). During *O. dioica* evolution, the new bimodular MT lost five Cys in the N-terminal domain, yielding a 7-Cys/12-Cys arrangement (as it is found in current OdiMT1). Next, this 7-Cys/12-Cys MT was internal duplicated generating multimodular MTs (as it is OdiMT2).

The primeval appendicularian MT would have been a bimodular form made of two 12-Cys domains, apparently unrelated with the other 12-Cys domains of chordates. This ancestral appendicularian MT lost five Cys in the N-terminal domain in the *O. dioica* lineage (but remained intact in other *Oikopleura* lineages), yielding the t-12C + 12C organization of the current OdiMT1. This MT was later duplicated, and internal duplications in one of the copies of a repeat unit (RU) made of the t-12C + 12C domains gave rise to the large multimodular OdiMT2 (Calatayud et al. 2018).

In summary, our analyses have shown: 1) de novo emergence of MT genes and domains during chordate diversification; 2) convergent evolution to 12-Cys domains in chordate MTs, but also in other animals, probably favored by essential constraints of the metal coordination; 3) convergent evolution in the metal preference for Cd ions, probably triggered by an ancestral requirement for cadmium-selective forms in many marine animals; 4) lineage-specific events of gene duplications yielding significant levels of MT multiplicity; and 5) independent amplifications of domains that generated multimodular MTs with high metal-binding capacities. All these results have led us to reconstruct an intricate evolutionary history for the chordate MTs, from a bimodular, cadmium-selective form in the last common ancestor, to a high structural and functional MT diversity currently found in the diverse species of the phylum. To discover the adaptive causes that shaped this history and the molecular mechanisms that made it possible will certainly be a fascinating challenge.

Materials and Methods

Chordate MT Search in Databases

Tunicate and cephalochordate MT sequences were retrieved from public databases (NCBI database: https://www.ncbi.nlm.nih.gov/; and Octopus database: http://octopus.obs-vlfr.fr/public/botryllus/blastbotryllus.php) by BlastP and TBlastN searches using known chordate MTs as queries: tunicate CroMT1 (ACN32211), HcuMTa (AY314949), OdiMT1 (AYN64372), and OdiMT2 (AYN64376); cephalochordate

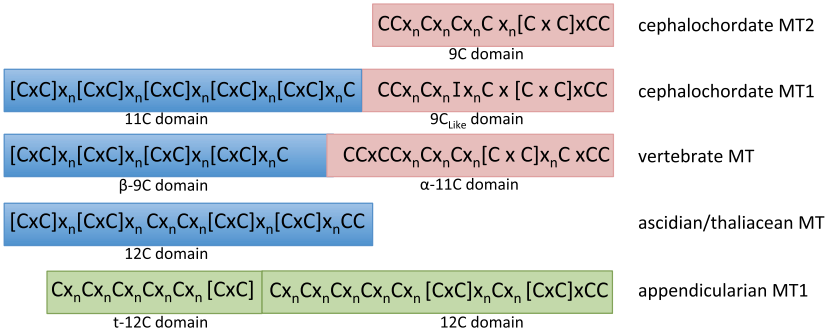


Fig. 7. Comparison of the arrangement of Cys motifs in different MT domains. The arrangement of the Cys motifs in the 11-Cys amino-terminal domain of cephalochordate MT1s are somewhat similar to that in the vertebrate 9-Cys β domains and in the ascidian/thaliacean 12-Cys MTs (blue boxes), whereas the Cys motifs in the carboxyl-terminal domains of cephalochordate MT1s and MT2s (9C and 9C$_{Like}$ domains) are vaguely similar to those in vertebrate 11-Cys α domains (pink boxes). In contrast, the arrangement of the Cys motifs in the appendicularian MT1 seems totally different (green boxes).

BflMT1 (XP_035672826), BflMT2 (BW841405), BlaMT1 (BL11229), and BlaMT2 (JT872034); and vertebrate HsaMT1A (NP_005937), HsaMT2A (NP_005944), HsaMT3 (NP_005945), and HsaMT4 (NP_116324). In addition, RNA-SRA for each tunicate species deposited in NCBI were BLAST searched using as queries MT sequences from the nearest phylogenetic species as well as from different species covering the main tunicate clades. Raw sequence data were retrieved and assembled using SeqMan 8.0.2 (Pro Assembler) software from the DNASTAR Lasergene package, and manually inspected in order to reconstruct new MT sequences. The MT nature of each new identified sequence was evaluated by BlastX searches against metazoan NCBI nonredundant protein sequence database. The amino acid sequences and the accession numbers of the retrieved MTs are provided in supplementary table S1, Supplementary Material online. Sequence alignments were generated with Aliview program (Larsson 2014) and reviewed manually.

Characterization of *S. thompsoni* Genes

Salpa thompsoni specimens were kindly provided by Nora-Charlotte Pauli from the Bettina Meyer's group at Alfred Wegener Institute for Polar and Marine Research (AWI). Genomic DNA was extracted using Quick-DNA/RNA Microprep Plus Kit Protocol (Zymo Research, CA) using 5 mg of tissue and following the manufacturer's instructions. The DNA concentration, purity, and integrity were checked using Tecan Infinite M200 (Tecan Group Ltd, Switzerland) measuring the absorbance at 260 and 280 nm.

Based on *S. thompsoni* genome and transcriptome projects (PRJNA318929 and GFCC00000000.1, respectively) and the raw data from *S. thompsoni* SRA projects, several pairs of primers were designed in order to PCR amplify the 5′ and 3′ flanking regions, as well as of intron sequences of the different SthMTs (supplementary table S4, Supplementary Material online).

For SthMT1 we used the primer pairs FwMT1/MT5 with RvMT1/MT2; for SthMT5, FwMT1/MT5 with RvMT5, and finally, for SthMT6 we used FwMT2/MT6 with RvMT6.

For each PCR reaction, 1 ng of genomic DNA was amplified using selected pairs of primers and the Phusion High-Fidelity DNA Polymerase (Invitrogen, Thermo Fisher Scientific, Waltham, MA) in a final 25 μl reaction. PCR conditions were 98 °C 30 s (s); 35 cycles of 98 °C 10 s, 50–54 °C 30 s (supplementary table S4, Supplementary Material online) and 72 °C 3 min; and 72 °C 10 min. PCR products were visualized in 0.8% agarose gels, isolated with the GelElute Plasmid Miniprep Kit (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany), and cloned with TOPO TA Cloning Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA). Plasmid DNA was purified from bacteria using the GeneElute Plasmid Miniprep Kit (Sigma-Aldrich), screened for insert presence by digestion with Fast Digest *Eco*RI (Invitrogen, Thermo Fisher Scientific, Waltham, MA), and sequenced at the Scientific and Technological Centers of the University of Barcelona using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) in an automatic sequencer (ABIPRISM 310, Applied Biosystems). Sequences were analyzed with Unipro UGENE software (Okonechnikov et al. 2012).

Production and Purification of Recombinant Metal–MT Complexes

Synthetic cDNAs codifying the selected MTs (i.e., CroMT1, HroMT1, HroMT2, BscMT1, R_4-BscMT1, SthMT1, SthMT2, SthMT3, and SthMT4) were provided by Synbio Technologies (Monmouth Junction, NJ), cloned in the pGEX-4T-1 expression vector (GE Healthcare, Chicago, IL), and transformed in protease-deficient *E. coli* BL21 strain for heterologous expression. For heterologous protein production, 500 ml of Luria–Bertani (LB) medium with 100 μg/ml ampicillin was inoculated with *E. coli* BL21 cells transformed with the corresponding recombinant plasmid. After overnight growth at 37 °C/250 rpm, the cultures were used to inoculate 5 l of fresh LB-100 μg/ml ampicillin medium. Gene expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h. After the first 30 min of induction, cultures were supplemented with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM), or $CuSO_4$ (500 μM) in order to generate metal–MT complexes. Cells were harvested by centrifugation for 5 min at 9,100 × g (7,700 rpm), and bacterial pellets were suspended in 125 ml of ice-cold phosphate-buffered saline

(PBS: 1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4, and 0.5% v/v β-mercaptoethanol). Resuspended cells were sonicated (Sonifier Ultrasonic Cell Disruptor) 8 min at voltage 6 with pulses of 0.6 s, and then centrifuged for 40 min at 17,200 × g (12,000 rpm) and 4 °C.

Soluble protein extracts containing GST-MT fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. GST-MT fusion proteins bound to the sepharose beads were washed with 30 ml of cold 1xPBS bubbled with argon to prevent oxidation. After three washes, GST-MT fusion proteins were digested with thrombin (SERVA Electrophoresis GmbH Heidelberg, Germany, 25 U/l of culture) overnight at 17 °C, thus enabling separation of the metal–MT complexes from the GST that remained bound to the sepharose matrix. The eluted metal–MT complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck-Millipore, Darmstadt, Germany), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris–HCl, pH 7.0. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at −80 °C until use.

Analysis of Metal–MT Complexes

Protein quantification and element composition of all the samples were achieved by ICP-AES measurements performed in a Optima 4300DV apparatus (Perkin-Elmer, MA) (S, 182.040 nm; Zn, 213.856 nm; Cd, 228.802 nm; Cu, 324.803 nm) under conventional conditions following an already stablished method (Bongers et al. 1988). Molecular weights were determined by ESI-MS, in a MicroTof-Q instrument (Bruker Daltonics Gmbh, Bremen, Germany) connected to a Series 1100 HPLC pump (Agilent Technologies) controlled by the Compass Software. The instrument was calibrated with ESI-L Low Concentration Turning Mix (Agilent Technologies, Santa Clara, CA). Metallated forms were detected under native conditions: 20 µl of sample injected through a PEEK tube at 30–50 µl min^{-1} in a 3.5–5.0 kV capillary-counter voltage, at 90–110 °C of desolvation temperature, and with dry gas at 6 l min^{-1}. The spectra were recorded between a *m/z* range from 800 to 3,000. The liquid carrier was a 90:10 mixture of 15 mM ammonium acetate and acetonitrile at pH 7.0. All molecular masses were calculated according to the bibliography (Fabris et al. 1996).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors thank Nora-Charlotte Pauli and Bettina Meyer at Alfred Wegener Institute for Polar and Marine Research for providing *Salpa thompsoni* specimens, and all tunicate laboratories for sharing sequence results ahead of their publication. They thank Sebastian Artime for experimental support. They also thank the Centres Científics i Tecnològics (CCiT) de la Universitat de Barcelona (DNA sequencing) and the Servei d'Anàlisi Química (SAQ) de la Universitat Autònoma de Barcelona (ICP-AES, ESI-MS) for allocating instrument time. R.A. was supported by BIO2015-67358-C2-1-P, C.C. was supported by BFU2016-80601-P and PID2019-110562GB-I00, and M.C. and O.P. were supported by BIO2015-67358-C2-2-P, the Spanish Ministerio de Ciencia e Innovación. M.C. and O.P. are members of the "Grup de Recerca de la Generalitat de Catalunya," ref. 2017SGR-864, and R.A. and CC. of ref. 2017SGR-1665. M.G.-R. acknowledges the UAB the PIF grant.

Data Availability

The data underlying this article are available in NCBI (https://www.ncbi.nlm.nih.gov/) and Octopus (http://octopus.obs-vlfr.fr/public/botryllus/blastbotryllus.php) databases. The amino acid sequences and the accession numbers of the MTs used in this work are available in supplementary table S1, Supplementary Material online.

References

Adamo GM, Lotti M, Tamas MJ, Brocca S. 2012. Amplification of the *CUP1* gene is associated with evolution of copper tolerance in *Saccharomyces cerevisiae*. *Microbiology* 158(Pt 9):2325–2335.

Artells E, Palacios O, Capdevila M, Atrian S. 2013. Mammalian MT1 and MT2 metallothioneins differ in their metal binding abilities. *Metallomics* 5(10):1397–1410.

Baumann C, Beil A, Jurt S, Niederwanger M, Palacios O, Capdevila M, Atrian S, Dallinger R, Zerbe O. 2017. Structural adaptation of a protein to increased metal stress: NMR structure of a marine snail metallothionein with an additional domain. *Angew Chem Int Ed Engl*. 56(16):4617–4622.

Beil A, Jurt S, Walser R, Schonhut T, Guntert P, Palacios O, Atrian S, Capdevila M, Dallinger R, Zerbe O. 2019. The solution structure and dynamics of Cd-metallothionein from helix pomatia reveal optimization for binding Cd over Zn. *Biochemistry* 58(45):4570–4581.

Blindauer CA. 2014. Metallothioneins. In: Maret W, Wedd A, editors. Binding, transport and storage of metal ions in biological cells. Cambridge: The Royal Society of Chemistry. p. 594–653.

Blindauer CA, Leszczyszyn OI. 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. *Nat Prod Rep*. 27(5):720–741.

Bofill R, Capdevila M, Atrian S. 2009. Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu-thioneins. *Metallomics* 1(3):229–234.

Bongers J, Walton CD, Richardson DE, Bell JU. 1988. Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma atomic emission spectrometric determination of sulfur. *Anal Chem*. 60(24):2683–2686.

Bruland KW. 1980. Oceanographic distributions of cadmium, zinc, nickel, and copper in the North Pacific. *Earth Planetary Sci Lett*. 47(2):176–198.

Calatayud S, Garcia-Risco M, Capdevila M, Cañestro C, Palacios O, Albalat R. Forthcoming 2021. Modular evolution and population variability of *Oikopleura dioica* metallothioneins. *Front Cell Dev Biol*. doi: 10.3389/fcell.2021.702688.

Calatayud S, Garcia-Risco M, Pedrini-Martha V, Eernisse DJ, Dallinger R, Palacios Ò, Capdevila M, Albalat R. 2021. Modularity in protein evolution: modular organization and de novo domain evolution in mollusk metallothioneins. *Mol Biol Evol*. 38(2):424–436.

Calatayud S, Garcia-Risco M, Rojas NS, Espinosa-Sanchez L, Artime S, Palacios O, Cañestro C, Albalat R. 2018. Metallothioneins of the urochordate *Oikopleura dioica* have Cys-rich tandem repeats, large size and cadmium-binding preference. *Metallomics* 10(11):1585–1594.

Cañestro C, Albalat R. 2012. Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. *Brief Funct Genomics*. 11(2):131–141.
Capdevila M, Atrian S. 2011. Metallothionein protein evolution: a miniassay. *J Biol Inorg Chem*. 16(7):977–989.
Capdevila M, Bofill R, Palacios Ò, Atrian S. 2012. State-of-the-art of metallothioneins at the beginning of the 21st century. *Coordination Chem Rev*. 256(1–2):46–62.
Catalan A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J. 2016. An indel polymorphism in the MtnA 3′ untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*. *PLoS Genet*. 12(4):e1005987.
Coppola U, Ristoratore F, Albalat R, D'Aniello S. 2019. The evolutionary landscape of the Rab family in chordates. *Cell Mol Life Sci*. 76(20):4117–4130.
Costa D, Marien J, Janssens TK, van Gestel CA, Driessen G, Sousa JP, van Straalen NM, Roelofs D. 2012. Influence of adaptive evolution of cadmium tolerance on neutral and functional genetic variation in *Orchesella cincta*. *Ecotoxicology* 21(7):2078–2087.
Dallinger R, Zerbe O, Baumann C, Egger B, Capdevila M, Palacios O, Albalat R, Calatayud S, Ladurner P, Schlick-Steiner B, et al. 2020. Metallomics reveal a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins. *Metallomics* 12(5):702–720.
de Francisco P, Martin-Gonzalez A, Turkewitz AP, Gutierrez JC. 2017. Extreme metal adapted, knockout and knockdown strains reveal a coordinated gene expression among different *Tetrahymena thermophila* metallothionein isoforms. *PLoS One* 12(12):e0189076.
de Francisco P, Martin-Gonzalez A, Turkewitz AP, Gutierrez JC. 2018. Genome plasticity in response to stress in *Tetrahymena thermophila*: selective and reversible chromosome amplification and paralogous expansion of metallothionein genes. *Environ Microbiol*. 20(7):2410–2421.
Delsuc F, Philippe H, Tsagkogeorga G, Simion P, Tilak MK, Turon X, Lopez-Legentil S, Piette J, Lemaire P, Douzery EJP. 2018. A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol*. 16(1):39.
Digilio G, Bracco C, Vergani L, Botta M, Osella D, Viarengo A. 2009. The cadmium binding domains in the metallothionein isoform Cd7-MT10 from *Mytilus galloprovincialis* revealed by NMR spectroscopy. *J Biol Inorg Chem*. 14(2):167–178.
Dvorak M, Lackner R, Niederwanger M, Rotondo C, Schnegg R, Ladurner P, Pedrini-Martha V, Salvenmoser W, Kremser L, Lindner H, et al. 2018. Metal binding functions of metallothioneins in the slug *Arion vulgaris* differ from metal-specific isoforms of terrestrial snails. *Metallomics* 10(11):1638–1654.
Egli D, Domènech J, Selvaraj A, Balamurugan K, Hua H, Capdevila M, Georgiev O, Schaffner W, Atrian S. 2006. The four members of the *Drosophila* metallothionein family exhibit distinct yet overlapping roles in heavy metal homeostasis and detoxification. *Genes Cells* 11(6):647–658.
Fabris D, Zaia J, Hathout Y, Fenselau C. 1996. Retention of thiol protons in two classes of protein zinc ion coordination centers. *J Am Chem Soc*. 118(48):12242–12243.
Ferrández-Roldán A, Martí-Solans J, Cañestro C, Albalat R. 2019. *Oikopleura dioica*: an emergent chordate model to study the impact of gene loss on the evolution of the mechanisms of development. In: Tworzydlo W, Bilinski SM, editors. Evo-Devo: non-model species in cell and developmental biology. Cham: Springer International Publishing. p. 63–105.
Franchi N, Boldrin F, Ballarin L, Piccinni E. 2011. CiMT-1, an unusual chordate metallothionein gene in *Ciona intestinalis* genome: structure and expression studies. *J Exp Zool A Ecol Genet Physiol*. 315A(2):90–100.
García-Risco M, Calatayud S, Niederwanger M, Albalat R, Palacios Ò, Capdevila M, Dallinger R. 2020. Two unconventional metallothioneins in the apple snail *Pomacea bridgesii* have lost their metal specificity during adaptation to freshwater habitats. *Int J Mol Sci*. 22(1):95.
Guirola M, Perez-Rafael S, Capdevila M, Palacios O, Atrian S. 2012. Metal dealing at the origin of the *Chordata* phylum: the metallothionein system and metal overload response in amphioxus. *PLoS One* 7(8):e43299.
Isani G, Carpene E. 2014. Metallothioneins, unconventional proteins from unconventional animals: a long journey from nematodes to mammals. *Biomolecules* 4(2):435–457.
Iturbe-Espinoza P, Gil-Moreno S, Lin W, Calatayud S, Palacios O, Capdevila M, Atrian S. 2016. The fungus *Tremella mesenterica* encodes the longest metallothionein currently known: gene, protein and metal binding characterization. *PLoS One* 11(2):e0148651.
Janssens TKS, Lopéz R. D R, Mariën J, Timmermans MJTN, Montagne-Wajer K, van Straalen NM, Roelofs D. 2008. Comparative population analysis of metallothionein promoter alleles suggests stress-induced microevolution in the field. *Environ Sci Technol*. 42(10):3873–3878.
Janssens TKS, Roelofs D, Van Straalen NM. 2009. Molecular mechanisms of heavy metal tolerance and evolution in invertebrates. *Insect Sci*. 16(1):3–18.
Jenny MJ, Payton SL, Baltzegar DA, Lozier JD. 2016. Phylogenetic analysis of molluscan metallothioneins: evolutionary insight from *Crassostrea virginica*. *J Mol Evol*. 83(3–4):110–125.
Kocot KM, Tassia MG, Halanych KM, Swalla BJ. 2018. Phylogenomics offers resolution of major tunicate relationships. *Mol Phylogenet Evol*. 121:166–173.
Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276–3278.
Louis A, Roest Crollius H, Robinson-Rechavi M. 2012. How much does the amphioxus genome represent the ancestor of chordates? *Brief Funct Genomics*. 11(2):89–95.
Luo M, Finet C, Cong H, Wei HY, Chung H. 2020. The evolution of insect metallothioneins. *Proc Biol Sci*. 287: 20202189.
Maret W. 2013. Zinc and the zinc proteome. In: Banci L, editor. Metallomics and the cell. Dordrecht: Springer Netherlands. p. 479–501.
Maroni G, Wise J, Young JE, Otto E. 1987. Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster*. *Genetics* 117(4):739–744.
Munoz A, Forsterling FH, Shaw CF III, Petering DH. 2002. Structure of the (113)Cd(3)beta domains from *Homarus americanus* metallothionein-1: hydrogen bonding and solvent accessibility of sulfur atoms. *J Biol Inorg Chem*. 7(7–8):713–724.
Narula SS, Brouwer M, Hua Y, Armitage IM. 1995. Three-dimensional solution structure of *Callinectes sapidus* metallothionein-1 determined by homonuclear and heteronuclear magnetic resonance spectroscopy. *Biochemistry* 34(2):620–631.
Naville M, Henriet S, Warren I, Sumic S, Reeve M, Volff JN, Chourrout D. 2019. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr Biol*. 29(7):1161–1168.e1166.
Niederwanger M, Dvorak M, Schnegg R, Pedrini-Martha V, Bacher K, Bidoli M, Dallinger R. 2017. Challenging the metallothionein (MT) gene of *Biomphalaria glabrata*: unexpected response patterns due to cadmium exposure and temperature stress. *Int J Mol Sci*. 18:1747.
Okonechnikov K, Golosova O, Fursov M, UGENE Team. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28(8):1166–1167.
Palacios O, Atrian S, Capdevila M. 2011. Zn- and Cu-thioneins: a functional classification for metallothioneins? *J Biol Inorg Chem*. 16(7):991–1009.
Palacios O, Espart A, Espin J, Ding C, Thiele DJ, Atrian S, Capdevila M. 2014. Full characterization of the Cu-, Zn-, and Cd-binding properties of CnMT1 and CnMT2, two metallothioneins of the pathogenic fungus *Cryptococcus neoformans* acting as virulence factors. *Metallomics* 6(2):279–291.
Palacios O, Jimenez-Marti E, Niederwanger M, Gil-Moreno S, Zerbe O, Atrian S, Dallinger R, Capdevila M. 2017. Analysis of metal-binding features of the wild type and two domain-truncated mutant variants of *Littorina littorea* metallothionein reveals its Cd-specific character. *Int J Mol Sci*. 18(7):1452.

Palacios O, Pagani A, Pérez-Rafael S, Egg M, Höckner M, Brandstätter A, Capdevila M, Atrian S, Dallinger R. 2011. Shaping mechanisms of metal specificity in a family of metazoan metallothioneins: evolutionary differentiation of mollusc metallothioneins. *BMC Biol*. 9:4.

Palacios O, Perez-Rafael S, Pagani A, Dallinger R, Atrian S, Capdevila M. 2014. Cognate and noncognate metal ion coordination in metal-specific metallothioneins: the *Helix pomatia* system as a model. *J Biol Inorg Chem*. 19(6):923–935.

Papadopoulou C, Kanias GD. 1977. Tunicate species as marine pollution indicators. *Mar Pollut Bull*. 8(10):229–231.

Pedrini-Martha V, Koll S, Dvorak M, Dallinger R. 2020. Cadmium uptake, MT gene activation and structure of large-sized multi-domain metallothioneins in the terrestrial door snail *Alinda biplicata* (Gastropoda, Clausiliidae). *Int J Mol Sci*. 21(5):1631.

Perez-Rafael S, Monteiro F, Dallinger R, Atrian S, Palacios O, Capdevila M. 2014. *Cantareus aspersus* metallothionein metal binding abilities: the unspecific CaCd/CuMT isoform provides hints about the metal preference determinants in metallothioneins. *Biochim Biophys Acta*. 1844(9):1694–1707.

Riek R, Precheur B, Wang Y, Mackay EA, Wider G, Guntert P, Liu A, Kagi JH, Wuthrich K. 1999. NMR structure of the sea urchin (*Strongylocentrotus purpuratus*) metallothionein MTA. *J Mol Biol*. 291(2):417–428.

Rokas A, Holland PW. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol*. 15(11):454–459.

Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Eng*. 12(2):85–94.

Serén N, Glaberman S, Carretero MA, Chiari Y. 2014. Molecular evolution and functional divergence of the metallothionein gene family in vertebrates. *J Mol Evol*. 78(3–4):217–233.

Somorjai IML, Martí-Solans J, Diaz-Gracia M, Nishida H, Imai KS, Escriva H, Cañestro C, Albalat R. 2018. Wnt evolution and function shuffling in liberal and conservative chordate genomes. *Genome Biol*. 19(1):98.

Stillman MJ, Cai W, Zelazowski AJ. 1987. Cadmium binding to metallothioneins. Domain specificity in reactions of alpha and beta fragments, apometallothionein, and zinc metallothionein with Cd2. *J Biol Chem*. 262(10):4538–4548.

Tanguy A, Moraga D. 2001. Cloning and characterization of a gene coding for a novel metallothionein in the Pacific oyster *Crassostrea gigas* (CgMT2): a case of adaptive response to metal-induced stress? *Gene* 273(1):123–130.

Timmermans MJ, Ellers J, Roelofs D, van Straalen NM. 2005. Metallothionein mRNA expression and cadmium tolerance in metal-stressed and reference populations of the springtail *Orchesella cincta*. *Ecotoxicology* 14(7):727–739.

Tio L, Villarreal L, Atrian S, Capdevila M. 2004. Functional differentiation in the mammalian metallothionein gene family: metal binding features of mouse MT4 and comparison with its paralog MT1. *J Biol Chem*. 279(23):24403–24413.

Tomas M, Domènech J, Capdevila M, Bofill R, Atrian S. 2013. The sea urchin metallothionein system: comparative evaluation of the SpMTA and SpMTB metal-binding preferences. *FEBS Open Bio*. 3:89–100.

Tzafriri-Milo R, Benaltabet T, Torfstein A, Shenkar N. 2019. The potential use of invasive Ascidians for biomonitoring heavy metal pollution. *Front Mar Sci*. 6:611.

Vallee BL, Falchuk KH. 1993. The biochemical basis of zinc physiology. *Physiol Rev*. 73(1):79–118.

Valls M, Bofill R, Gonzalez-Duarte R, Gonzalez-Duarte P, Capdevila M, Atrian S. 2001. A new insight into metallothionein (MT) classification and evolution. The in vivo and in vitro metal binding features of *Homarus americanus* recombinant MT. *J Biol Chem*. 276(35):32835–32843.

Vasak M, Meloni G. 2011. Chemistry and biology of mammalian metallothioneins. *J Biol Inorg Chem*. 16(7):1067–1078.

Vest KE, Hashemi HF, Cobine PA. 2013. The copper metallome in eukaryotic cells. In: Banci L, editor. Metallomics and the cell. Dordrecht (The Netherlands): Springer Netherlands. p. 451–478.

Voskoboynik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, Koh W, Passarelli B, Fan HC, Mantalas GL, Palmeri KJ, et al. 2013. The genome sequence of the colonial chordate, *Botryllus schlosseri*. *Elife* 2:e00569.

Waldron KJ, Rutherford JC, Ford D, Robinson NJ. 2009. Metalloproteins and metal sensing. *Nature* 460(7257):823–830.

Yue JX, Yu JK, Putnam NH, Holland LZ. 2014. The transcriptome of an amphioxus, *Asymmetron lucayanum*, from the Bahamas: a window into chordate evolution. *Genome Biol Evol*. 6(10):2681–2696.

Ziller A, Fraissinet-Tachet L. 2018. Metallothionein diversity and distribution in the tree of life: a multifunctional protein. *Metallomics* 10(11):1549–1559.

ARTICLE VI: SUPPLEMENATRY MATERIAL

TUNICATES ILLUMINATE THE ENIGMATIC EVOLUTION OF CHORDATE METALLOTHIONEINS BY GENE GAINS AND LOSSES, INDEPENDENT MODULAR EXPANSIONS, AND FUNCTIONAL CONVERGENCES

The file includes:

Figure S1

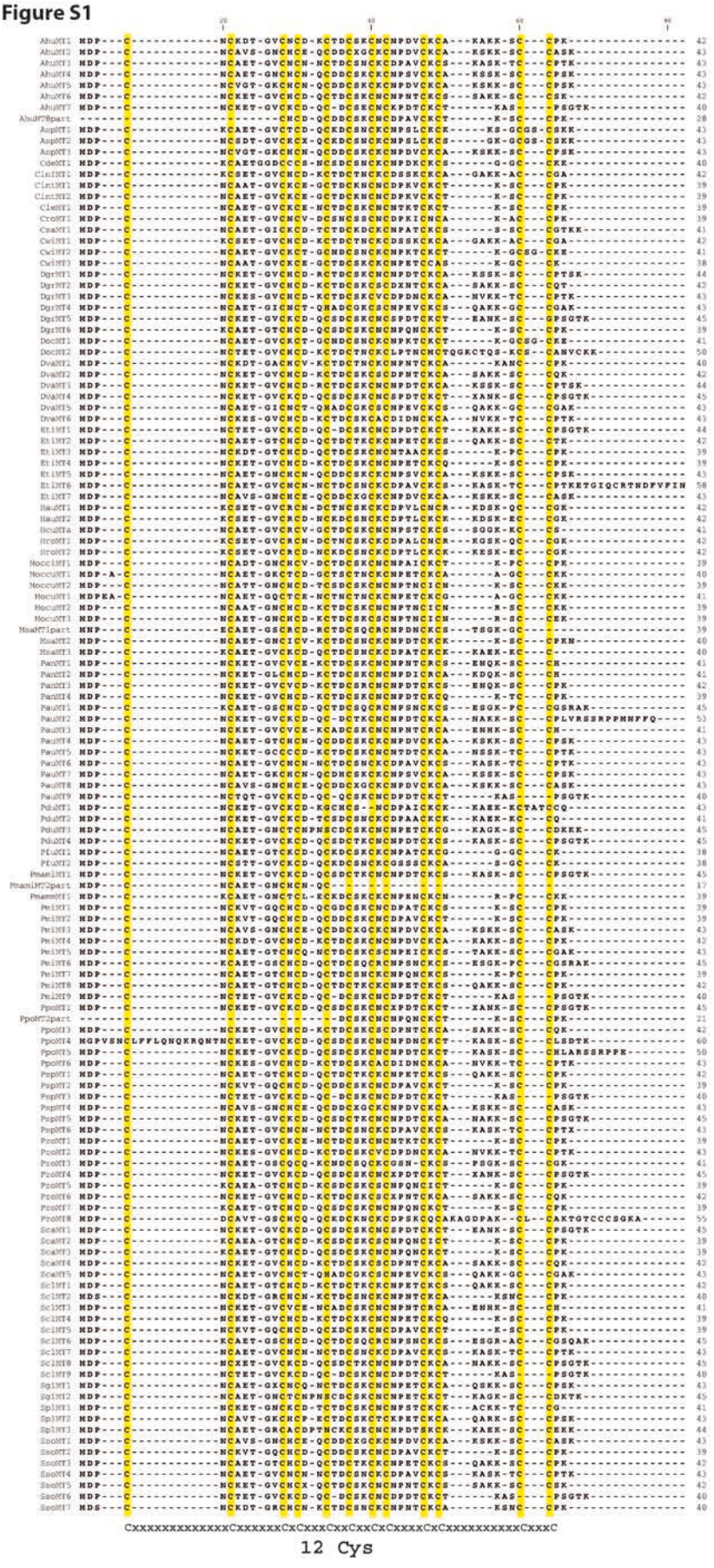


Figure S1. Amino acid alignment of Ascidian MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs are showed below each alignment. Names of MT sequences from different species are as in Supplemental_Table_S1.pdf, and alphabetically ordered.

FIGURE S2

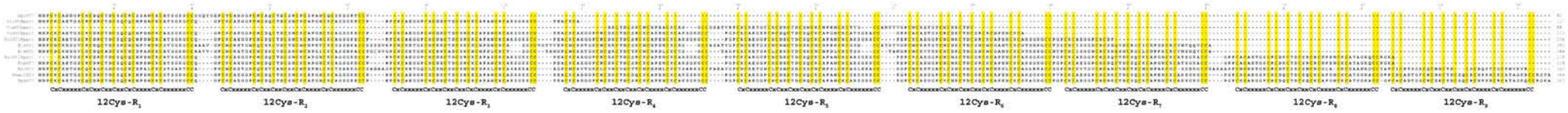


Figure S2. Amino acid alignment of multimodular Stelydae Ascidian MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs are showed below each alignment. Names of MT sequences from different species are as in Supplemental_Table_S1.pdf.

FIGURE S3

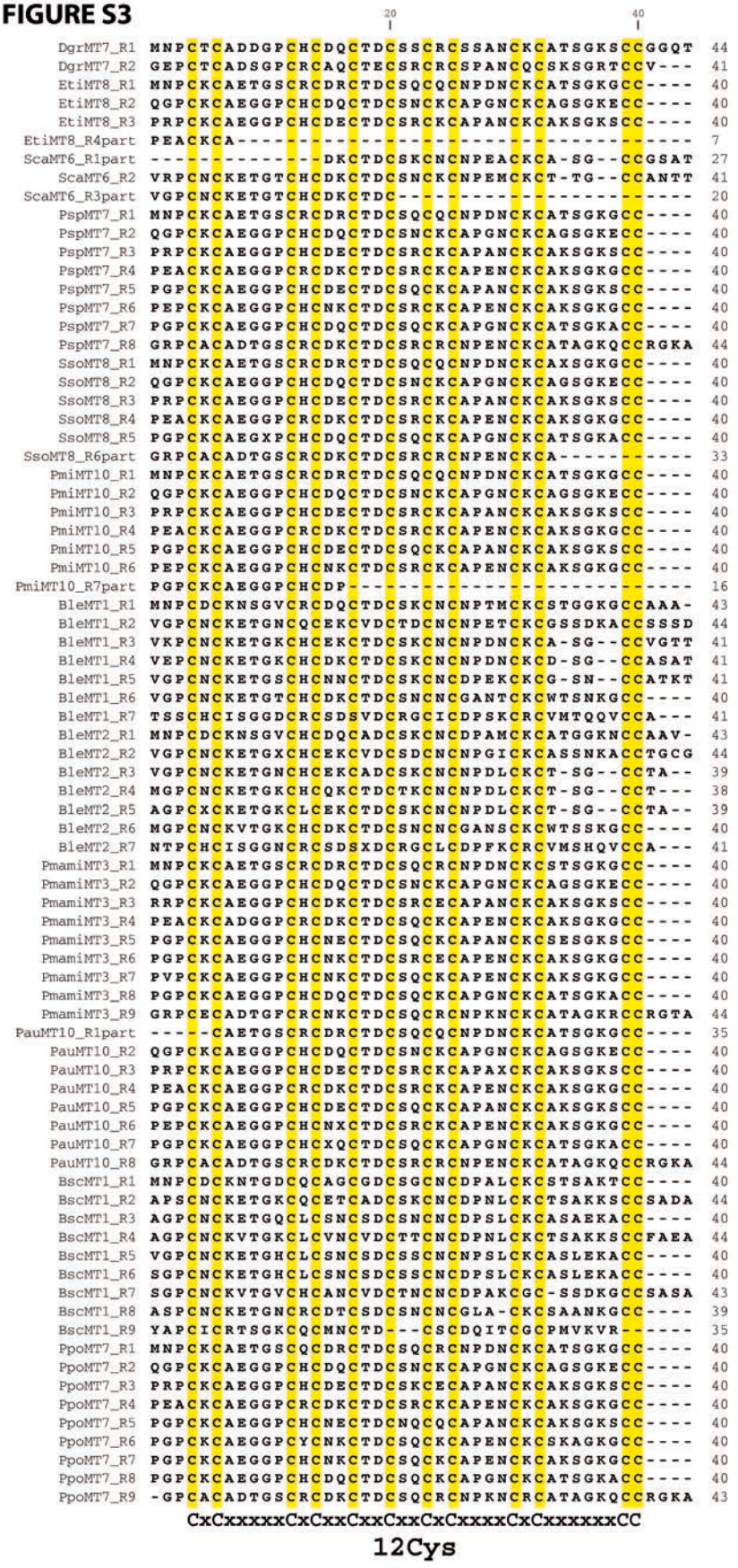


Figure S3. Amino acid alignment of multimodular Stelydae repeats of Ascidian MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs are showed below each alignment. Names of MT sequences from different species are as in Supplemental_Table_S1.pdf.

Figure S4

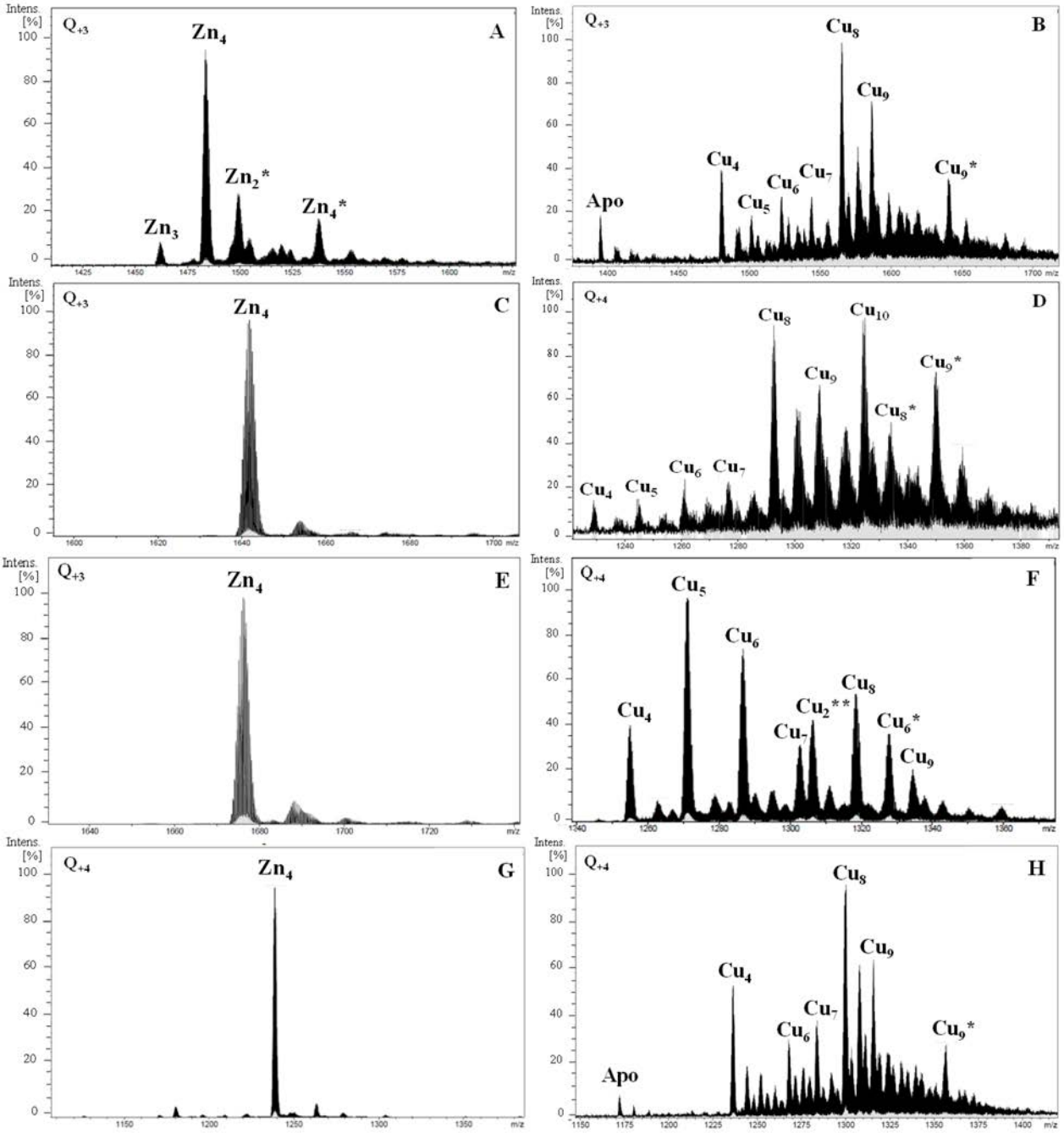


Figure S4. ESI-MS spectra recorded at neutral pH showing the corresponding charge state (+3 or +4) of CroMT1, HroMT1, HroMT2 and BscMT1 (R4) synthesized in Zn-enriched media (A, C, E and G, respectively), or in Cu-enriched media (B, D, F and H, respectively). Glycosylated metallated-species are marked with asterisk (*).

Figure S5

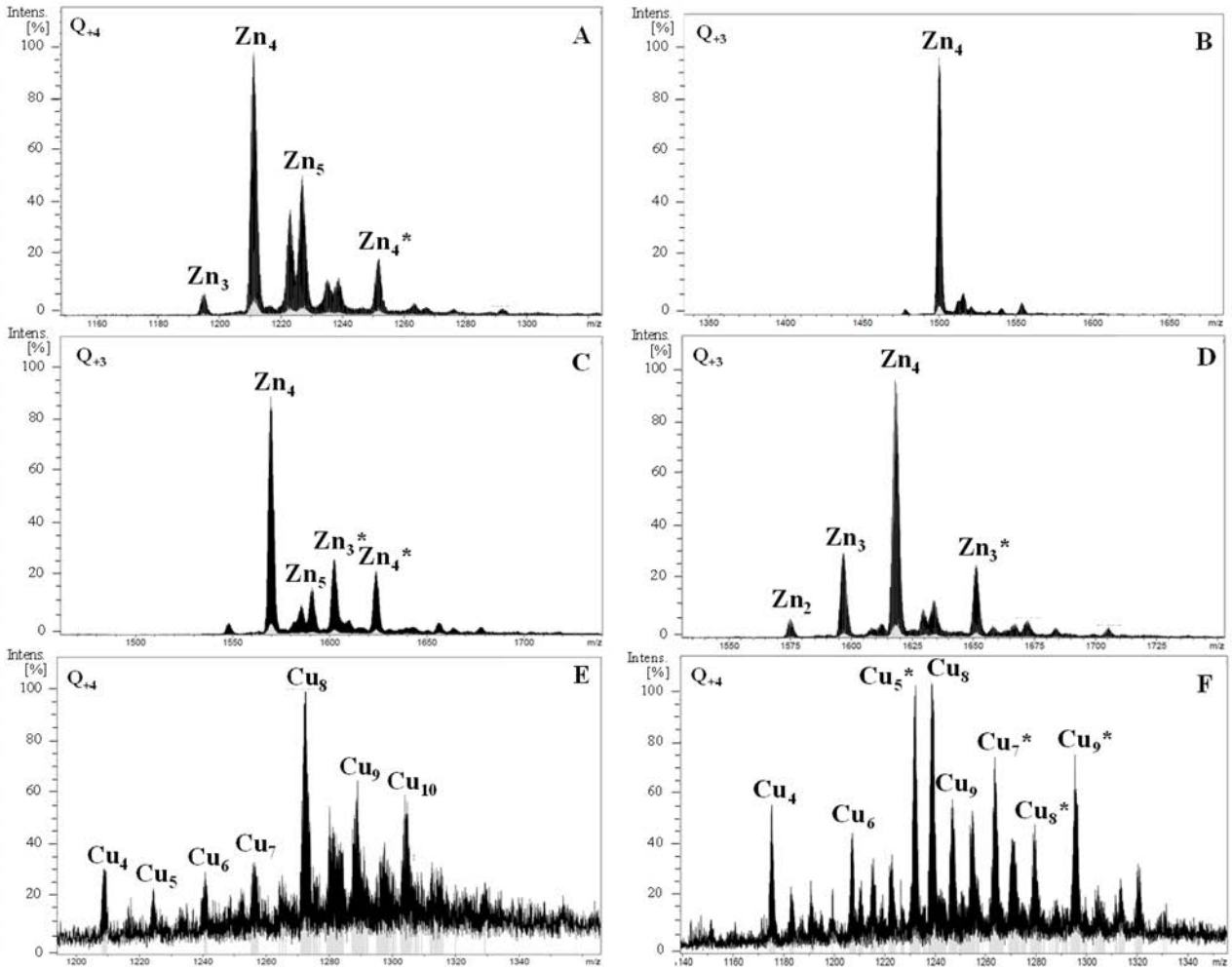


Figure S5. ESI-MS spectra recorded at neutral pH showing the corresponding charge state (+3 or +4) of SthMT1, SthMT2, SthMT3 and SthMT4 synthesized in Zn-enriched media (A, B, C and D), or SthMT1, SthMT3 produced in Cu-enriched media (E and F). Glycosylated metallated-species are marked with asterisk (*).

Figure S6

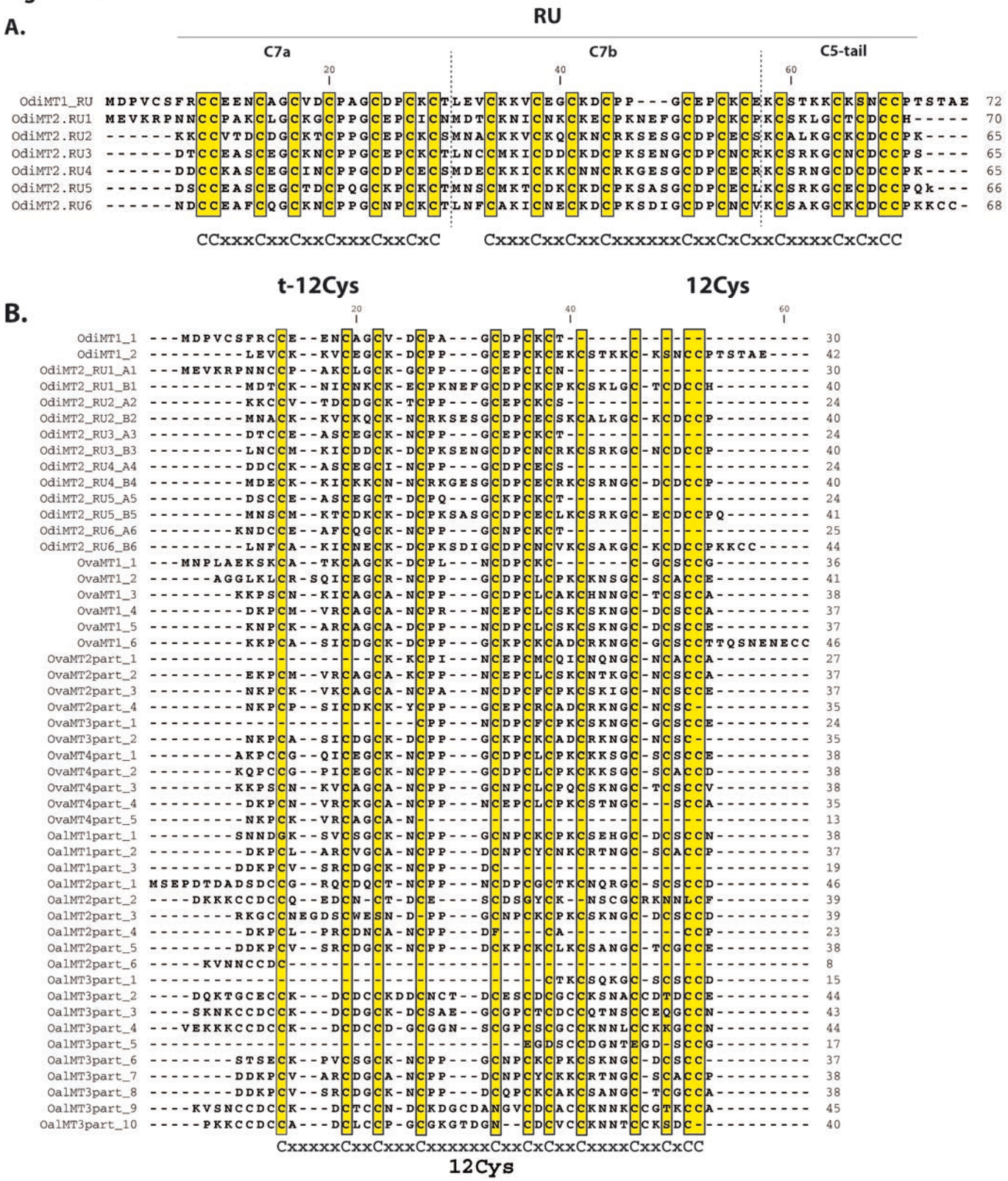


Figure S6. Amino acid alignment of Appendicularia MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs are showed below the alignment. Names of MT sequences from different species are as in Supplemental_Table_S1.pdf. **(A)** Alignment of OdiMT1 with the six repeats units (RU1 to 6) of OdiMT2. Notice that a RU is made of two C7 subunits (C7a and C7b) plus a carboxyl-terminal tail with 5 cysteines (C5). Comparison of OdiMTs with MTs from other *Oikopleura* species revealed that appendicularian MTs were multi-modular proteins made of a conserved 12-Cys domain that corresponded to the previously described C7b+C5 subunits (see text for details). OdiMT1 would be therefore a bi-modular MT, whose first 12-Cys domain (formerly known as C7a) would be a trimmed 12-Cys domain (t-12C) lacking the five last cysteines (C5). OdiMT2 would be a multi-modular MT made of twelve 12-Cys domains, whose odd domains would have been also trimmed as t-12C. **(B)** Alignment of the C12 domains of several multi-modular appendicularian MTs.

Figure S7

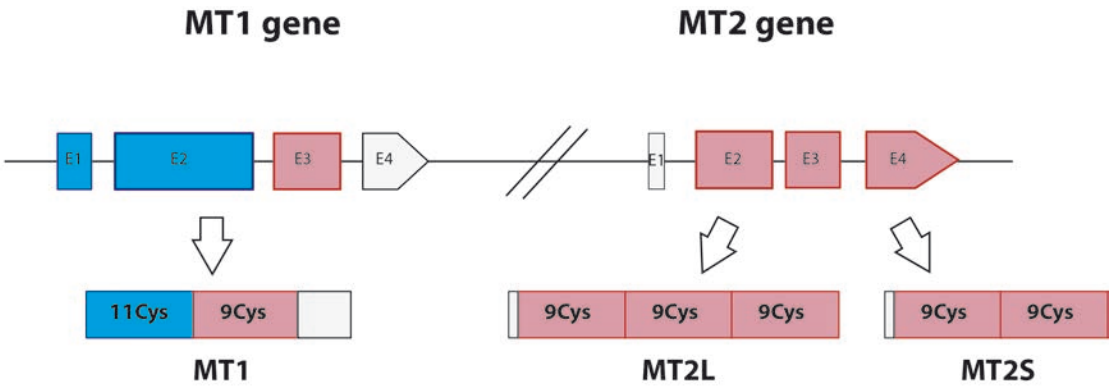


Figure S7. Cephalochordate gene structure. Cephalochordate *MT1* genes have four exons: exons 1 and 2 codify for the 11 Cys domain (blue box), exon 3 codifies for the 9 Cys_{Like} domain (red box), and exon 4 codifies for the carboxyl-terminal end devoid of Cys residues. Cephalochordate *MT2* genes have four exons, a tiny exon 1, followed by three similar exons (exon 2, 3 and 4), each one encoding for a 9-Cys domain (red box). Exon 2 is included in three-domain MT2L isoforms but skipped in the two-domain MT2S isoforms. .

Figure S8

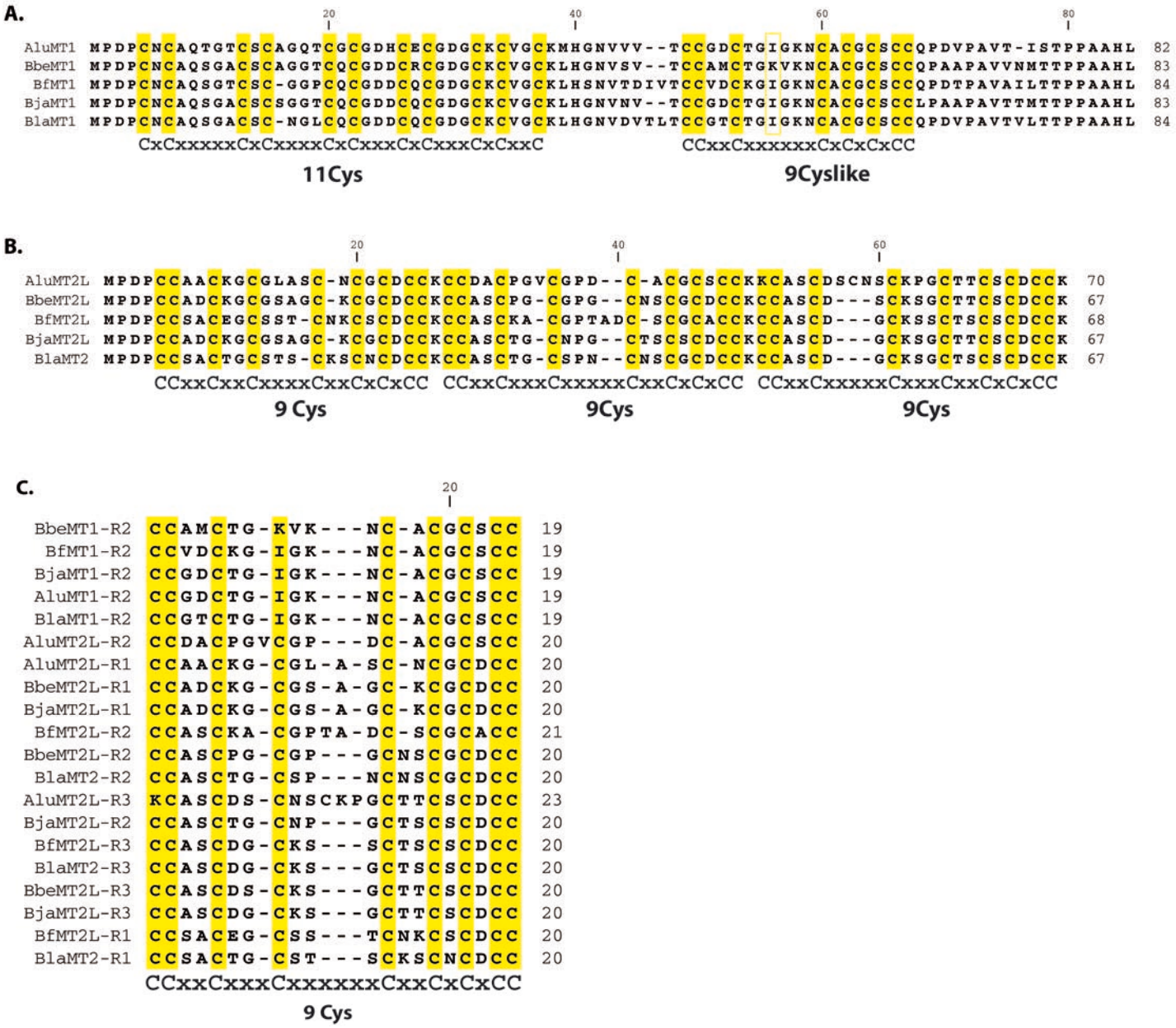


Figure S8. Amino acid alignment of Cephalochordate MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs are showed below each alignment. Names of MT sequences from different species are as in Supplemental_Table_S1.pdf, and alphabetically ordered. **(A)** Cephalochordate MT1 alignment. **(B)** Cephalochordate MT2L alignment. (C) Alignment of 9-Cys domains of MT2s and 9-Cys_{Like} domains of MT1s.

Figure S9

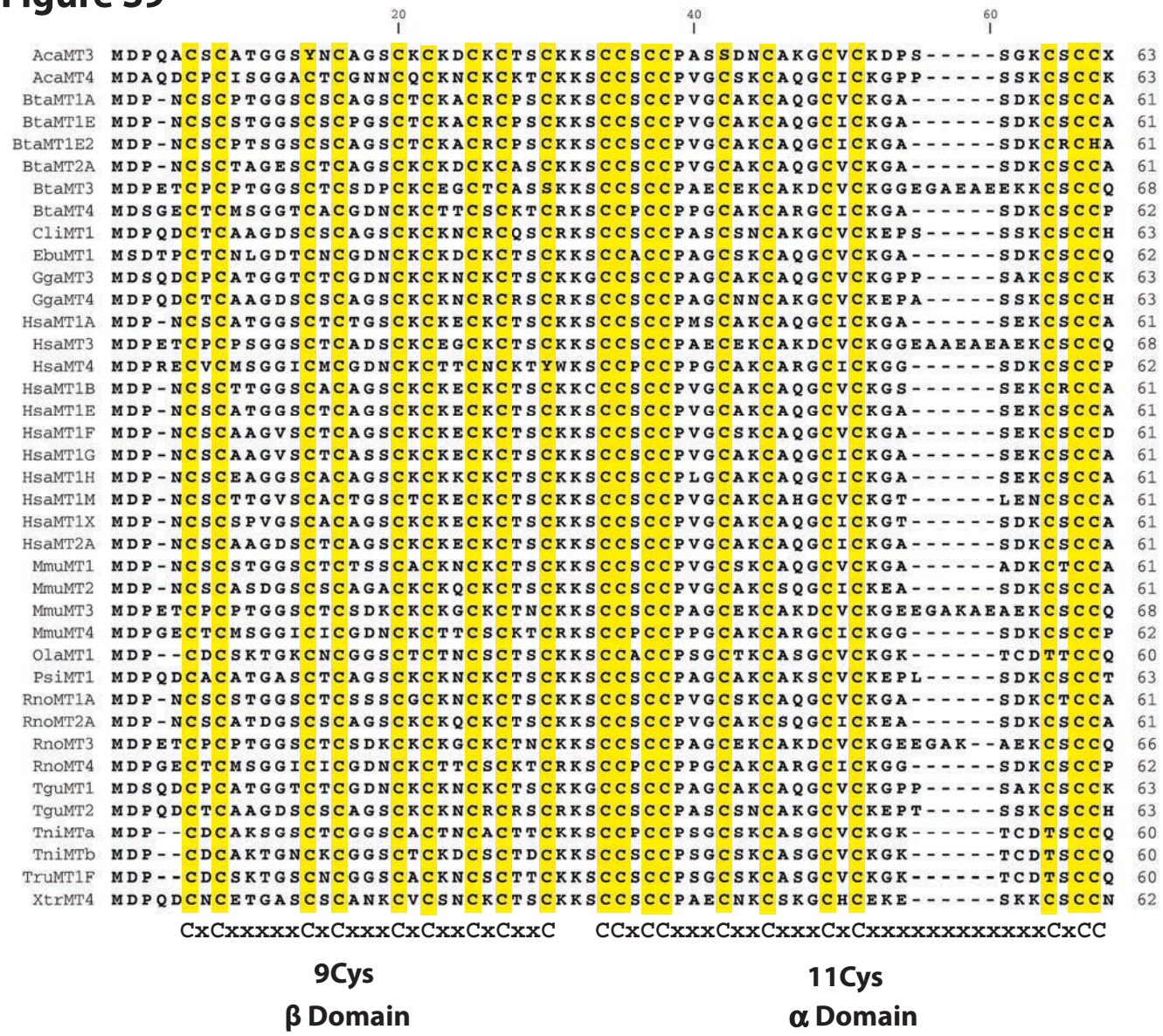


Figure S9. Amino acid alignment of vertebrate MTs. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs are showed below each alignment. Names of MT sequences from different species are as in Supplemental_Table_S1.pdf, and alphabetically ordered.

Table S1. Chordate MT sequences analyzed in this study.

Subphylum	Class	Order	Family	Species	MT name	Accession Number	Sequence	K/N ratio
VERTEBRATES								
	Mammalia							
		Artiodactyla	Bovidae	Bos taurus	BtaMT1A	XP_024833736.1	MDPNCSCPTGGSCSCAGSCTCKACRCPSCKKSCCSCCPVGCAKCAQGCVCKGASDKCSCCA	6:1
					BtaMT1E	NP_001108329.1	MDPNCSCSTGGSCSCPGSCTCKACRCPSCKKSCCSCCPVGCAKCAQGCICKGASDKCSCCA	6:1
					BtaMT1E2	NP_001071602.1	MDPNCSCPTSGSCSCAGSCTCKACRCPSCKKSCCSCCPVGCAKCAQGCICKGASDKCRCHA	6:1
					BtaMT2A	NP_001068608.1	MDPNCSCTAGESCTCAGSCKCKDCKCASCKKSCCSCCPVGCAKCAQGCVCKGASDKCSCCA	8:1
					BtaMT3	NP_001106775.1	MDPETCPCPTGGSCTCSDPCKCEGCTCASSKKSCCSCCPAECEKCAKDCVCKGGEGAEAEEKKCSCCQ	8:0
					BtaMT4	NP_001108331.1	MDSGECTCMSGGTCACGDNCKCTTCSCKTCRKSCCPCCPPGCAKCARGCICKGASDKCSCCP	6:1
		Primates	Hominidae	Homo Sapiens	HsaMT1A	NP_005937.2	MDPNCSCATGGSCTCTGSCKCKECKCTSCKKSCCSCCPMSCAKCAQGCICKGASEKCSCCA	8:1
					HsaMT1B	NP_005938.1	MDPNCSCTTGGSCACAGSCKCKECKCTSCKKCCCSCCPVGCAKCAQGCVCKGSSEKCRCCA	8:1
					HsaMT1E	NP_783316.2	MDPNCSCATGGSCTCAGSCKCKECKCTSCKKSCCSCCPVGCAKCAQGCVCKGASEKCSCCA	8:1
					HsaMT1F	NP_005940.1	MDPNCSCAAGVSCTCAGSCKCKECKCTSCKKSCCSCCPVGCSKCAQGCVCKGASEKCSCCD	8:1
					HsaMT1G	NP_005941.1	MDPNCSCAAGVSCTCASSCKCKECKCTSCKKSCCSCCPVGCAKCAQGCICKGASEKCSCCA	8:1
					HsaMT1H	NP_005942.1	MDPNCSCEAGGSCACAGSCKCKKCKCTSCKKSCCSCCPLGCAKCAQGCICKGASEKCSCCA	9:1
					HsaMT1M	NP_789846.2	MDPNCSCTTGVSCACTGSCTCKECKCTSCKKSCCSCCPVGCAKCAHGCVCKGTLENCSCCA	6:2
					HsaMT1X	NP_005943.1	MDPNCSCSPVGSCACAGSCKCKECKCTSCKKSCCSCCPVGCAKCAQGCICKGTSDKCSCCA	8:1
					HsaMT2A	NP_005944.1	MDPNCSCAAGDSCTCAGSCKCKECKCTSCKKSCCSCCPVGCAKCAQGCICKGASDKCSCCA	8:1
					HsaMT3	NP_005945.1	MDPETCPCPSGGSCTCADSCKCEGCKCTSCKKSCCSCCPAECEKCAKDCVCKGGEAAEAEAEKCSCCQ	8:0
					HsaMT4	NP_116324.2	MDPRECVCMSGGICMCGDNCKCTTCNCKTYWKSCCPCCPPGCAKCARGCICKGGSDKCSCCP	6:2
		Rodentia	Muridae	Mus Musculus	MmuMT1	NP_038630.1	MDPNCSCSTGGSCTCTSSCACKNCKCTSCKKSCCSCCPVGCSKCAQGCVCKGAADKCTCCA	7:2
					MmuMT2	NP_032656.1	MDPNCSCASDGSCSCAGACKCKQCKCTSCKKSCCSCCPVGCAKCSQGCICKEASDKCSCCA	8:1
					MmuMT3	NP_038631.1	MDPETCPCPTGGSCTCSDKCKCKGCKCTNCKKSCCSCCPAGCEKCAKDCVCKGEEGAKAEAEKCSCCQ	11:1
					MmuMT4	NP_032657.1	MDPGECTCMSGGICICGDNCKCTTCSCKTCRKSCCPCCPPGCAKCARGCICKGGSDKCSCCP	6:1
				Rattus norvegicus	RnoMT1A	NP_620181.1	MDPNCSCSTGGSCTCSSSCGCKNCKCTSCKKSCCSCCPVGCSKCAQGCVCKGASDKCTCCA	7:2
					RnoMT2A	NP_001131036.1	MDPNCSCATDGSCSCAGSCKCKQCKCTSCKKSCCSCCPVGCAKCSQGCICKEASDKCSCCA	8:1
					RnoMT3	NP_446420.1	MDPETCPCPTGGSCTCSDKCKCKGCKCTNCKKSCCSCCPAGCEKCAKDCVCKGEEGAKAEKCSCCQ	11:1
					RnoMT4	NP_001119556.1	MDPGECTCMSGGICICGDNCKCTTCSCKTCRKSCCPCCPPGCAKCARGCICKGGSDKCSCCP	6:1
	Aves							
		Galliformes	Phasianidae	Gallus gallus	GgaMT3	NP_001091007.1	MDSQDCPCATGGTCTCGDNCKCKNCKCTSCKKGCCSCCPAGCAKCAQGCVCKGPPSAKCSCCK	9:2
					GgaMT4	NP_990606.1	MDPQDCTCAAGDSCSCAGSCKCKNCRCRSCRKSCCSCCPAGCNNCAKGCVCKEPASSKCSCCH	6:3
		Columbiformes	Columbinae	Columba livia	CliMT1	P15787	MDPQDCTCAAGDSCSCAGSCKCKNCRCQSCRKSCCSCCPASCSNCAKGCVCKEPSSSKCSCCH	6:2
		Passeriformes	Estrildidae	Taeniopygia guttata	TguMT1	NP_001232450.1	MDSQDCPCATGGTCTCGDNCKCKNCKCTSCKKGCCSCCPAGCAKCAQGCVCKGPPSAKCSCCK	9:2
					TguMT2	NP_001232079.1	MDPQDCTCAAGDSCSCAGSCKCKNCRCRSCRKSCCSCCPASCSNCAKGCVCKEPTSSKCSCCH	6:2
	Reptilia							
		Squamata	Dactyloidae	Anolis carolinensis	AcaMT3	H9GCQ4	MDPQACSCATGGSYNCAGSCKCKDCKCTSCKKSCCSCCPASSDNCAKGCVCKDPSSGKCSCCX	8:2
					AcaMT4	H9GQC7	MDAQDCPCISGGACTCGNNCQCKNCKCKTCKKSCCSCCPVGCSKCAQGCICKGPPSSKCSCCK	9:3
			Lacertidae	Podarcis siculus	PsiMT1	Q708T3	MDPQDCACATGASCTCAGSCKCKNCKCTSCKKSCCSCCPAGCAKCAKSCVCKEPLSDKCSCCT	9:1
	Amphibia							
		Anura	Pipidae	Xenopus tropicalis	XtrMT4	NP_001165150.1	MDPQDCNCETGASCSCANKCVCSNCKCTSCKKSCCSCCPAECNKCSKGCHCEKESKKCSCCN	9:5
	Actinopterygii							
		Beloniformes	Adrianichthyidae	Oryzias latipes	OlaMT1	NP_001098255.1	MDPCDCSKTGKCNCGGSCTCTNCSCTSCKKSCCACCPSGCTKCASGCVCKGKTCDTTCCQ	7:2
		Tetraodontiformes	Tetraodontidae	Takifugu rubripes	TruMT1F	XP_003967042.1	MDPCDCSKTGSCNCGGSCACKNCSCTTCKKSCCSCCPSGCSKCASGCVCKGKTCDTSCCQ	7:2
				Tetraodon nigroviridis	TniMTa	Q4SKC8	MDPCDCAKSGSCTCGGSCACTNCACTTCKKSCCPCCPSGCSKCASGCVCKGKTCDTSCCQ	6:1
					TniMTb	Q4SKC7	MDPCDCAKTGNCKCGGSCTCKDCSCTDCKKSCCSCCPSGCSKCASGCVCKGKTCDTSCCQ	8:1
	Myxini							
		Myxiniformes	Myxinidae	Eptatretus burgeri	EbuMT1	SRA	MSDTPCTCNLGDTCNCGDNCKCKDCKCTSCKKSCCACCPAGCSKCAQGCVCKGASDKCSCCQ	8:3
CEPHALOCHORDATES								
	Leptocardii							

	Amphioxiformes	Asymmetronidae	Asymmetron lucayanum	AluMT1	SRA	MPDPCNCAQTGTCSCAGGQTCGCGDHCECGDGCKCVGCKMHGNVVVTCCGDCTGIGKNCACGCSCCQPDVPAVTISTPPAAHL	3:3
				AluMT2 (S/L)	SRA	MPDPCCAACKGCGLASCNCGCDCCKCCDACPGVCGPDCACGCSCCKKCASCDSCNSCKPGCTTCSCDCCK	6:2
		Branchiostomidae	Branchiostoma belcheri	BbeMT1	XP_019631158	MPDPCNCAQSGACSCAGGTCQCGDDCRCGDGCKCVGCKLHGNVSVTCCAMCTGKVKNCACGCSCCQPAAPAVVNMTTPPAAHL	4:4
				BbeMT2 (S/L)	JZ816971	MPDPCCADCKGCGSAGCKCGCDCCKCCASCPGCGPGCNSCGCDCCKCCASCDSCKSGCTTCSCDCCK	6:1
			Branchiostoma japonicum	BjaMT1	SRA	MPDPCNCAQSGACSCSGGTCQCGDDCQCGDGCKCVGCKLHGNVNVTCCGDCTGIGKNCACGCSCCLPAAPAVTTMTTPPAAHL	3:4
				BjaMT2 (S/L)	SRA	MPDPCCADCKGCGSAGCKCGCDCCKCCASCTGCNPGCTSCSCDCCKCCASCDGCKSGCTTCSCDCCK	6:1
			Branchiostoma lanceolatum	BlaMT1	JT862963	MPDPCNCAQSGACSCNGLCQCGDDCQCGDGCKCVGCKLHGNVDVTLTCCGTCTGIGKNCACGCSCCQPDVPAVTVLTTPPAAHL	3:4
				BlaMT2 (S/L)	JT872034	MPDPCCSACTGCSTSCKSCNCDCCKCCASCTGCSPNCNSCGCDCCKCCASCDGCKSGCTSCSCDCCK	5:3
			Branchiostoma floridae	BflMT1	SRA	MPDPCNCAQSGTCSCGGPCQCGDDCQCGDGCKCVGCKLHSNVTDIVTCCVDCKGIGKNCACGCSCCQPDTPAVAILTTPPAAHL	4:3
				BflMT2(S/L)	BW841405	MPDPCCSACEGCSSTCNKCSCDCCKCCASCKACGPTADCSCGCACCKCCASCDGCKSSCTSCSCDCCK	6:1
TUNICATES							
	Ascidiacea						
		Aplousobranchia	Clavelinidae	Clavelina lepadiformis	CleMT1	SRA	MDPCNCAETGVCKCENCTDCSKCNCNTKTCKCTKSCCPK	6:4
			Polycitoridae	Cystodytes dellechiajei	CdeMT1	SRA	MDPCKCAETGGDCCCSNCSDCSNCKCNPDKCKCSGGCCKK	6:3
			Holozoidae	Distaplia occidentalis	DocMT1	SRA	MDPCNCAETGVCKCTGCNDCSNCKCNPKTCKCTKGCSGCKE	6:4
					DocMT2	SRA	MDPCNCTETGVCHCDKCTDCTNCKCLPTNCMCTQGKCTQSKCSCANVCKK	6:4
		Phlebobranchia	Ascidiidae	Ascidia sp T160	AspMT1	SRA	MDPCKCAETGVCTCDQCKDCSNCKCNPSLCKCKKSGCGSCSKK	8:2
					AspMT2	SRA	MDPCNCSDTGVCKCXQCKDCSNCKCNPSLCKCSGKGCGSCSKK	7:3
					AspMT3	SRA	MDPCNCVGTGKCHCNQCDDCSKCKCNPDVCKCAKSKKSCCPSK	8:3
				Phallusia fumigata	PfuMT1	SRA	MDPCNCAETGTCKCDQCKDCSKCKCNPATCKCGGGCCK	6:2
					PfuMT2	SRA	MDPCNCSTTGVCKCDQCKDCSNCKCGSSSCKCASGCCK	5:2
				Phallusia mammilata	PmammMT1	SRA	MDPCKCAETGNCTCLECKDCSKCKCNPENCKCNRPCCKK	7:4
			Cionidae	Ciona robusta[1]	CroMT1	ACN32211	MDPCNCAETGVCNCVDCSNCSSCNCDPKICNCAKACCPK	3:5
				Ciona intestinalis[2]	CintMT1	SRA	MDPCNCAATGVCKCEGCTDCKNCNCDPKVCKCTKSCCPK	6:3
					CintMT2	SRA	MDPCNCAETGVCKCEGCTDCKNCNCDPKVCKCTKSCCPK	6:3
				Ciona savignyi	CsaMT1	SRA	MDPCNCAETGICKCDTCTDCKDCKCNPATCKCSSSCCGTKK	6:2
			Corellidae	Corella inflata	CinfMT1	SRA	MDPCKCSETGVCHCDKCTDCTNCKCDSSKCKCAGAKKACCGA	7:1
				Corella willmeriana	CwiMT1	SRA	MDPCNCAATGVCKCEGCTDCSKCKCNPETCCASKGCCK	5:2
					CwiMT2	SRA	MDPCNCAETGVCKCTGCNDCSNCKCNPKTCKCTKGCSGCKE	6:4
					CwiMT3	SRA	MDPCNCAATGVCKCEGCTDCSKCKCNPETCCASKGCCK	5:2
		Stolidobranchia	Molgulidae	Mogula occidentalis	MocciMT1	S539075	MDPCNCADTGNCHCVDCTDCSKCNCNPAICKCTKPCCPK	4:4
				Mogula occulta	MoccuMT1	SRA	MDPACNCAETGKCTCDGCTSCTNCKCNPETCKCAAGCCKK	5:3
					MoccuMT2	SRA	MDPCNCATTGNCHCDTCSDCSKCNCNPTNCICNKSCCKK	4:6
				Mogula oculata	MocuMT1	S112679	MDPEACNCAETGQCTCENCTNCTNCKCNPETCKCGAGCCKK	4:5
					MocuMT2	S94991	MDPCNCAATGNCHCDKCTDCSKCNCNPTNCICNRSCCKK	4:6
					MocuMT3	S103250	MDPCNCAETGNCHCDKCTDCSKCSCNPTNCICNRSCCEK	3:5
			Pyuridae	Halocynthia aurantium	HauMT1	SRA	MDPCKCSETGVCRCNDCTNCSKCKCDPVLCNCRKDSKQCCGK	6:3
					HauMT2	SRA	MDPCKCSETGVCRCDNCKDCSNCKCDPTLCKCKKDSKECCGK	8:2
				Halocynthia roretzi	HroMT1	SRA	MDPCKCSETGVCRCNDCTNCSKCKCDPALCNCRKGSKQCCGK	6:3
					HroMT2	SRA	MDPCKCSETGVCRCDNCKDCSNCKCDPTLCKCKKESKECCGK	8:2
				Herdmania curvata	HcuMT1	AY314949	MDPCNCAETGVCRCVGCTDCSKCNCNPSTCKCSSGGKKCCS	4:3
				Microcosmus sabatieri	MsaMT1part	SRA	MNPCECAETGSCRCDRCTDCSQCRCNPDNCKCSTSGKGC	2:3
					MsaMT2	SRA	MDPCNCAETGNCICVKCTDCSNCKCNPDTCKCKKSCCPKN	6:5
					MsaMT3	SRA	MDPCKCAETGVCKCDQCTDCSNCKCDPATCKCKKAEKKCC	8:1
				Pyura dura	PduMT1	SRA	MDPCNCKETGVCKCDKGCHCSNCDPAICKCKKAEKKCTATCCQ	8:2
					PduMT2	SRA	MDPCNCKETGVCKCDTCSDCSNCKCDPAACKCKKAEKKCCQ	8:2
					PduMT3	SRA	MDPCNCAETGNCTCNPNSCDCSKCNCNPETCKCGKAGKSCCDKKK	7:6
					PduMT4	SRA	MDPCNCKETGVCKCDQCSDCTKCNCNPDTCXCSKASKSCCPSGTK	6:3

Styelidae	Asterocarpa humilis	AhuMT1	SRA	MDPCNCKDTGVCNCDKCTDCSKCNCNPDVCKCAKAKKSCCPK	8:4
		AhuMT2	SRA	MDPCNCAVSGNCHCEQCDDCXGCKCNPDVCKCAKSKKSCCASK	6:3
		AhuMT3	SRA	MDPCNCAETGVCNCNNCTDCSNCKCDPAVCKCSKASKTCCPTK	5:5
		AhuMT4	SRA	MDPCNCAETGNCHCNQCDDCSKCKCNPSVCKCAKSSKSCCPSK	6:4
		AhuMT5	SRA	MDPCNCVGTGKCHCNQCDDCSKCKCNPDVCKCAKSKKSCCPSK	8:3
		AhuMT6	SRA	MDPCNCKETGVCHCDQCTDCSKCNCNPNTCKCSSAKKSCCSK	6:4
		AhuMT7	SRA	MDPCNCKETGVCKCDQCDCSECNCKPDTCKCTKASPSGTK	6:2
		AhuMT7part	SRA	CHCDQCDDCSKCNCDPAVCKCTKSCCPK	4:1
	Botrylloides leachii	BleMT1	SRA	MNPCDCKNSGVCRCDQCTDCSKCNCNPTMCKCSTGGKGCCAAAVGPCNCKETGNCQCEKCVDCTDCNCNPETCKCGSSDKACCSSSDVKPCNCKETGKCHCEKCTDCSKCNCNPDNCKCASGCCVGTTVEPCNCKETGKCHCDKCTDCSKCNCNPDNCKCDSGCCASATVGPCNCKETGSCHCNNCTDCSKCNCDPEKCKCGSNCCATKTVGPCNCKETGTCHCDKCTDCSNCNCGANTCKCWTSNKGCCTSSCHCISGGDCRCSDSVDCRGCICDPSKCRCVMTQQVCCA	20:26
		BleMT2	SRA	MNPCDCKNSGVCHCDQCADCSKCNCDPAMCKCATGGKNCCAAVVGPCNCKETG?CHCEKCVDCSDCNCNPGICKCASSNKACCTGCGVGPCNCKETGNCHCEKCADCSKCNCNPDLCKCTSGCCTAMGPCNCKETGKCHCQKCTDCTKCNCNPDLCKCTSGCCTAGPC?CKETGKCLCEKCTDCSKCNCNPDLCKCTSGCCTAMGPCNCKVTGKCHCDKCTDCSNCNCGANSCKCWTSSKGCCNTPCHCISGGNCRCSDS?DCRGCLCDPFKCRCVMSHQVCCA	28:23
	Botryllus schlosseri	BscMT1	ATSW01006707.1; chromosome 6 / JG326099.1 CCAO15503.b1*; JG321680.1 CCAO13237.b1*; JG326100.1 CCAO15503.g1*	MNPCDCKNTGDCQCAGCGDCSGCNCDPALCKCSTSAKTCCAPSCNCKETGKCQCETCADCSKCNCDPNLCKCTSAKKSCCSADAAGPCNCKETGQCLCSNCSDCSNCNCDPSLCKCASAEKACCAGPCNCKVTGKCLCVNCVDCTTCNCDPNLCKCTSAKKSCCFAEAVGPCNCKETGHCLCSNCSDCSSCNCNPSLCKCASLEKACCSGPCNCKETGHCLCSNCSDCSSCNCDPSLCKCASLEKACCSGPCNCKVTGVCHCANCVDCTNCNCDPAKCGCSSDKGCCSASAASPCNCKETGNCRCDTCSDCSNCNCGLACKCSAANKGCCYAPCICRTSGKCQCMNCTDCSCDQITCGCPMVKVR	31:31
	Dendrodoa grossularia	DgrMT1	SRA	MDPCNCKETGVCHCDRCTDCSKCNCNPDTCKCAKSSKSCCPTSK	6:3
		DgrMT2	SRA	MDPCNCKETGVCHCDKCTDCSKCSCDXNTCKCASAKKSCCQT	6:2
		DgrMT3	SRA	MDPCNCKESGVCHCDKCTDCSKCVCDPDNCKCANVKKTCCPTK	7:3
		DgrMT4	SRA	MDPCNCAETGICNCTQHADCGKCSCNPEVCKCSQAKKGCCGAK	5:3
		DgrMT5	SRA	MDPCNCKETGVCKCDQCSDCSKCNCSPDTCKCTEANKSCGPSGTK	6:3
		DgrMT6	SRA	MDPCKCAETGTCHCDQCSDCSKCNCNPQNCKCTKSCCPK	5:3
		DgrMT7	SRA	MNPCTCADDGPCHCDQCTDCSSCRCSSANCKCATSGKSCCGGQTGEPCTCADSGPCRCAQCTECSRCRCSPANCQCSKSGRTCCV	3:3
	Distomus variolosus	DvaMT1	SRA	MDPCNCKDTGACHCVKCTDCTNCKCNPNTCKCAKANCCPK	6:5
		DvaMT2	SRA	MDPCNCKETGVCHCDKCTDCSKCSCDPNTCKCASAKKSCCQK	7:2
		DvaMT3	SRA	MDPCNCKETGVCHCDRCTDCSKCNCNPDTCKCAKSSKSCCPTSK	6:3
		DvaMT4	SRA	MDPCNCKETGVCKCDQCSDCSKCNCSPDTCKCTXANKSCCPSGTK	6:3
		DvaMT5	SRA	MDPCNCAETGICNCTQHADCGKCSCNPEVCKCSQAKKGCCGAK	5:3
		DvaMT6	SRA	MDPCNCKESGVCHCDKCTDCSKCACDIDNCKCANVKKTCCPTK	7:3
	Eusynstyela tincta	EtiMT1	SRA	MDPCNCTETGVCKCDQCDCSKCNCDPDTCKCTKAKKSCCPSGTK	7:2
		EtiMT2	SRA	MDPCNCAETGTCHCDQCTDCTKCKCNPETCKCSQAKKSCCTK	6:2
		EtiMT3	SRA	MDPCNCKDTGTCHCDQCTDCSKCNCNTAACKCSKPCCPK	5:3
		EtiMT4	SRA	MDPCNCKETGVCHCDKCTDCSKCNCNPETCKCQKSCCPK	6:3
		EtiMT5	SRA	MDPCNCAETGNCHCNQCDDCSKCKCNPSVCKCAKSKKSCCPSK	7:4
		EtiMT6	SRA	MDPCNCAETGVCNCNNCTDCSNCKCDPAVCKCSKASKTCCPTKETGIQCRTNDFVFIN	5:7
		EtiMT7	SRA	MDPCNCAVSGNCHCEQCDDCXGCKCNPDVCKCAKSKKSCCASK	6:3
		EtiMT8part	SRA	MNPCKCAETGSCRCDRCTDCSQCQCNPDNCKCATSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCPRPCKCAEGGPCHCDECTDCSRCKCAPANCKCAKSGKSCCPEACKCA	13:6
	Polyandrocarpa anguinea	PanMT1	SRA	MDPCNCKETGVCVCEKCTDCSKCNCNPNTCRCSENQKSCCH	4:5
		PanMT2	SRA	MDPCNCKETGLCHCDKCTDCSKCNCNPDICRCAKDQKSCCH	5:3
		PanMT3	SRA	MDPCNCKETGVCVCDKCTDCSRCNCNPDTCKCSENQKSCCPk	5:4
		PanMT4	SRA	MDPCNCKETGVCHCDKCTDCSKCNCNPDTCKCQKTCCPK	6:3
	Polyandrocarpa misakiensis	PmiMT1	SRA	MDPCNCKVTGQCHCDQCGDCSKCNCDPATCKCSKSCCPK	5:2
		PmiMT2	SRA	MDPCNCKVTGQCHCDQCDDCSKCNCDPAVCKCTKSCCPK	5:2

	PmiMT3	SRA	MDPCNCAVSGNCHCEQCDDCXGCKCNPDVCKCAKSKKSCCASK	6:3
	PmiMT4	SRA	MDPCNCKDTGVCNCDKCTDCSKCNCNPDVCKCAKAKKSCCPK	8:4
	PmiMT5	SRA	MDPCNCAETGTCNCQNCTDCSKCSCNPEICKCSTAKKSCCGAK	5:4
	PmiMT6	SRA	MDPCKCAETGSCHCDQCTDCSQCRCNPSNCKCSESGKPCCGSRAK	4:2
	PmiMT7	SRA	MDPCNCAETGTCHCDQCTDCSNCKCNPQNCKCSKPCCPK	4:4
	PmiMT8	SRA	MDPCNCAETGTCHCDQCTDCTKCKCNPETCKCSQAKKSCCPK	6:2
	PmiMT9	SRA	MDPCNCTETGVCKCDQCDCSKCNCDPDTCKCTKASPSGTK	5:2
	PmiMT10part	SRA	MNPCKCAETGSCRCDRCTDCSQCQCNPDNCKCATSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCPRPCKCAEGGPCHCDECTDCSRCKCAPANCKCAKSGKSCCPEACKCAEGGPCRCDKCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCDECTDCSQCKCAPANCKCAKSGKSCCPEPCKCAEGGPCHCNKCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCDP	30:10
Polyandrocarpa zorritensis	PzoMT1	SRA	MDPCNCAETGVCKCENCTDCSKCNCNTKTCKCTKSCCPK	6:4
	PzoMT2	SRA	MDPCNCKESGVCHCDKCTDCSKCVCDPDNCKCANVKKTCCPTK	7:3
	PzoMT3	SRA	MDPCNCAETGSCQCQKCNDCSQCKCGSNCKCSPSGKSCCGK	5:3
	PzoMT4	SRA	MDPCNCKETGVCKCDQCSDCSKCNCXPDTCKCTXANKSCCPSGTK	6:3
	PzoMT5	SRA	MDPCKCAEAGTCHCDKCSDCSKCSCNPQNCICTKSCCPK	5:2
	PzoMT6	SRA	MDPCNCKETGVCHCDKCTDCSKCSCXPNTCKCASAKKSCCQK	7:2
	PzoMT7	SRA	MDPCKCAETGTCHCDQCSDCSKCNCNPQNCKCTKSCCPK	5:3
	PzoMT8	SRA	MDPCDCAVTGSCHCQQCKDCKNCKCDPSKCQCAKAGDPAKCLCAKTGTCCCSGKA	8:1
Polycarpa aurata	PauMT1	SRA	MDPCKCAETGSCHCDQCTDCSQCRCNPSNCKCSESGKPCCGSRAK	4:2
	PauMT2	SRA	MDPCNCKETGVCKCDQCDCTKCNCNPDTCKCANAKKSCCPLVRSSRPPMNFFQ	6:5
	PauMT3	SRA	MDPCNCKETGVCVCEKCADCSKCNCNPNTCRCAENHKSCCH	4:5
	PauMT4	SRA	MDPCNCAETGTCHCNQCDDCSKCKCNPDVCKCAKSKKSCCPSK	7:3
	PauMT5	SRA	MDPCNCKETGCCCCDKCTDCSKCNCNTDTCKCANSSKTCCPTK	6:4
	PauMT6	SRA	MDPCNCAETGVCNCNNCTDCSNCKCDPAVCKCSKASKTCCPTK	5:5
	PauMT7	SRA	MDPCNCAETGKCHCNQCDHCSKCKCNPSVCKCAKSSKSCCPSK	7:3
	PauMT8	SRA	MDPCNCAVSGNCHCEQCDDCXGCKCNPDVCKCAKSKKSCCASK	6:3
	PauMT9	SRA	MDPCNCTQTGVCKCDQCQCSKCNCDPDTCKCTKASPSGTK	5:2
	PauMT10part	SRA	CAETGSCRCDRCTDCSQCQCNPDNCKCATSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCPRPCKCAEGGPCHCDECTDCSRCKCAPAXCKCAKSGKSCCPEACKCAEGGPCRCDKCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCDECTDCSQCKCAPANCKCAKSGKSCCPEPCKCAEGGPCHCNXCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCXQCTDCSQCKCAPGNCKCATSGKACCGRPCACADTGSCRCDKCTDCSRCRCNPENCKCATAGKQCCRGKA	35:11
Polycarpa mamillaris	PmamiMT1	SRA	MDPCNCKETGVCKCDQCSDCTKCNCNPDTCKCSKASKSCCPSGTK	7:3
	PmamiMT2part	SRA	MDPCNCAETGNCHCNQC	0:3
	PmamiMT3	SRA	MNPCKCAETGSCRCDRCTDCSQCRCNPDNCKCSTSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCRRPCKCAEGGPCHCDKCTDCSRCECAPANCKCAKSGKSCCPEACKCADGGPCRCDKCTDCSQCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCNECTDCSQCKCAPANCKCSESGKSCCPGPCKCAEGGPCHCNKCTDCSRCECAPENCKCAKSGKGCCPVPCKCAEGGPCHCNKCTDCSQCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCDQCTDCSQCKCAPGNCKCATSGKACCGRPCECADTGFCRCNKCTD	41:17
Polycarpa pomaria	PpoMT1	SRA	CSQCRCNPKNCKCATAGKRCCRGTA	6:3
	PpoMT2part	SRA	MDPCNCKETGVCKCDQCSDCSKCNCXPDTCKCTXANKSCCPSGTK	4:3
	PpoMT3	SRA	DCSKCNCNPQNCKCTKSCCPK	7:2
	PpoMT4	SRA	MDPCNCKETGVCHCDKCTDCSKCSCDPNTCKCASAKKSCCQK	8:7
	PpoMT5	SRA	MGPVSNCLFFLQNQKRQNTNCKETGVCKCDQCSDCSKCNCNPDNCKCTKASKSCCLSDTK	6:3
	PpoMT6	SRA	MDPCNCKETGVCKCDQCSDCSKCNCNPDTCKCTKASKSCCHLARSSRPPE	7:3
	PpoMT7	SRA	MDPCNCKESGVCHCDKCTDCSKCACDIDNCKCANVKKTCCPTK	42:17
Polycarpa sp.	PspMT1	SRA	MNPCKCAETGSCQCDRCTDCSQCRCNPDNCKCATSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCPRPCKCAEGGPCHCDECTDCSKCECAPANCKCAKSGKSCCPEACKCAEGGPCRCDKCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCNECTDCNQCQCAPANCKCAKSGKSCCPGPCKCAEGGPCYCNKCTDCSQCKCAPENCKCSKAGKGCCPGPCKCAEGGPCHCNKCTDCSQCKCAPENCKCAKSGKGCCPGPCKCAEGGPCHCDQCTDCSQCKCAPGNCKCATSGKACCGPCACADTGSCRCDKCTD	42:17
	PspMT1	SRA	MDPCNCAETGTCHCDQCTDCTKCKCNPETCKCSQAKKSCCPK	6:2
	PspMT2	SRA	MDPCNCKVTGQCHCDQCDDCSKCNCDPAVCKCTKSCCPK	5:2

				PspMT3	SRA	MDPCNCTETGVCKCDQC-DCSKCNCDPDTCKCTKASPSGTK	5:2
				PspMT4	SRA	MDPCNCAVSGNCHCEQCDDCXGCKCNPDVCKCAKSKKSCCASK	6:3
				PspMT5	SRA	MDPCNCKETGVCKCDQCSDCTKCNCNPDTCKCANAKKSCCPSGTK	7:4
				PspMT6	SRA	MDPCNCAETGVCNCNNCTDCSNCKCDPAVCKCSKASKTCCPTX	4:5
				PspMT7part	SRA	MNPCKCAETGSCRCDRCTDCSQCQCNPDNCKCATSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCPRPC KCAEGGPCHCDECTDCSRCKCAPANCKCAKSGKSCCPEACKCAEGGPCRCDKCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEGG PCHCDECTDCSQCKCAPANCKCAKSGKSCCPEPCKCAEGGPCHCNKCTDCSRCKCAP	26:9
			Stolonica socialis	SsoMT1	SRA	MDPCNCAVSGNCHCEQCDDCXGCKCNPDVCKCAKSKKSCCASK	6:3
				SsoMT2	SRA	MDPCNCKVTGQCHCDQCDDCSKCNCDPAVCKCTKSCCPK	5:2
				SsoMT3	SRA	MDPCNCAETGTCHCDQCTDCTKCKCNPETCKCSQAKKSCCPK	6:2
				SsoMT4	SRA	MDPCNCAETGVCNCNNCTDCSNCKCDPAVCKCSKASKTCCPTK	5:5
				SsoMT5	SRA	MDPCNCKETGVCHCXQCTDCSKCNCNPNTCKCSSAKKSCCSK	6:4
				SsoMT6	SRA	MDPCNCTETGVCKCDQCDCSKCNCDPDTCKCTKASPSGTK	5:2
				SsoMT7	SRA	MDSCNCKDTGRCHCNKCTDCSNCKCNPNTCKCAKSNCCPK	6:6
				SsoMT8part	SRA	MNPCKCAETGSCRCDRCTDCSQCQCNPDNCKCAXSGKGCCQGPCKCAEGGPCHCDQCTDCSNCKCAPGNCKCAGSGKECCPRP CKCAEGGPCHCDECTDCSRCKCAPANCKCAKSGKSCCPEACKCAEGGPCRCDKCTDCSRCKCAPENCKCAKSGKGCCPGPCKCAEG XPCHCDQCTDCSQCKCAPGNCKCATSGKACCGRPCACADTGSCRCDKCTDCSRCRCNPENCKCA	24:10
			Styela canopus	ScaMT1	SRA	MDPCNCKETGVCKCDQCSDCSKCNCSPDTCKCTEANKSCCPSGTK	6:3
				ScaMT2	SRA	MDPCKCAEAGTCHCDKCSDCSKCSCNPQNCICT	3:2
				ScaMT3	SRA	MDPCKCAETGTCHCDQCSDCSKCNCNPQNCKCTKSCCPK	5:3
				ScaMT4	SRA	MDPCNCKETGVCHCDKCTDCSKCSCDPNTCKCASAKKSCCQK	7:2
				ScaMT5	SRA	MDPCNCAETGVCNCTQHADCGKCSCNPEVCKCSQAKKGCCGAK	5:3
				ScaMT6part	SRA	DKCTDCSKCNCNPEACKCASGCCGSATVRPCNCKETGTCHCDKCTDCSNCKCNPEMCKCTTGCCANTTVGPCNCKETGTCHCDKC TDC	9:7
			Styela clava	SclMT1	SRA	MDPCNCAETGTCHCDKCTDCTKCKCNPETCKCSQAKKSCCPK	7:2
				SclMT2	SRA	MDSCNCKDTGRCHCNKCTDCSNCKCNPNTCKCAKSNCCPK	6:6
				SclMT3	SRA	MDPCNCKETGVCVCENCADCSKCNCNPNTCRCAENHKSCCH	3:6
				SclMT4	SRA	MDPCNCKETGVCHCDKCTDCXKCNCNPETCKCQKSCCPK	6:3
				SclMT5	SRA	MDPCNCKVTGQCHCDQCXDCSKCNCDPAVCKCTKSCCPK	5:2
				SclMT6	SRA	MDPCKCAETGSCHCDQCTDCSQCRCNPSNCKCSESGRACCGSQAK	3:2
				SclMT7	SRA	MDPCNCAETGVCNCNNCTDCSNCKCDPAVCKCSKASKTCCPTK	5:5
				SclMT8	SRA	MDPCNCXETGVCKCDQCSDCTKCNCNPDTCKCANAKKSCCPSGTK	6:4
				SclMT9	SRA	MDPCNCTETGVCKCDQCDCSKCNCDPDTCKCTKASPSGTK	5:2
			Styela gibbsii	SgiMT1	SRA	MDPCNCAETGXCNCQNCTDCSKCNCNPETCKCAQSKKSCCPSK	5:5
				SgiMT2	SRA	MDPCNCAETGNCTCNPNSCDCSKCNCNPETCKCTKAGKSCCDKTK	6:6
			Styela plicata	SplMT1	SRA	MDPCNCKETGNCKCDTCTDCSKCNCNPSTCKCKACKKTCCG	7:4
				SplMT2	SRA	MDPCNCAVTGKCHCPECTDCSKCTCKPETCKCAQARKSCCPSK	6:1
				SplMT3	SRA	MDPCNCAETGRCACDPTNCKCSECNCHPDTSKCKKAEKSCCEKK	7:3
Thaliaceans							
	Pyrosomatida	Pyrosomatidae	Pyrosomella verticillata	PveMT1	SRA	MDPCNCKVTGKCHCDQCTDCSKCNCASPCKCTKSCC	5:2
				PveMT2	SRA	MDPCNCKITGNCHCAQCTDCSNCSCAAPCKCTKSCC	3:3
	Salpida	Salpidae	Iasis cylindrica	IcyMT1	SRA	MDPCNCETTRDCHCQXCTDCSRCKCDLKNCNCSKPCCPK	4:3
				IcyMT2	SRA	MAYCNCLSNSNCHCNTCTDCSKCACAKNTCSCSGKTCCSPQ	3:5
			Salpa fusiformis	SfuMT1	SRA	MDPCNCNTSDMCHCESCNDCSKCNCARQTCKCSTKRCCSPK	4:4
				SfuMT2	SRA	MDPCNCQVTGVCHCNQCDDCSKCSCKPSNCKCNKPCCPK	5:4
				SfuMT3	SRA	MDPCNCQDTQSCYCSSCTDCSKCTCAKTTCKCSAKGCCSPN	4:2
				SfuMT4	SRA	MDSCNCKGTQSCHCNTCQDCSKCTCDKRTCKCSGKGCCSPE	5:2
			Salpa Thompsoni	SthMT1	MW558888	MDPCNCNTSDMCHCDSCKDCSKCNCARKTCKCSTKGCCSPK	6:3
				SthMT2	MKHR01072568	MDPCNCKVTGACHCDQCTDCGKCSCNPANCKCSKPCCPK	5:3
				SthMT3	MKHR01021011	MDPCNCQDTQSCYCNSCTDCSKCACAKTTCKCSAKGCCSPI	4:2
				SthMT4part	MKHR01216936	MDPCNCNTSVMCHCDTCESCSESNCAKQTCKSSTKRCCSPQ	3:3
				SthMT5	MW558889	MDPCNCNTSDMCHCESCKDCSKCNCAKQTCKCSTKRCCSSK	6:3

				SthMT6	MW558890	MDPCNCKVTGACHCDQCTDCSKCNCNPANCKCSKPCCPK	5:4
Larvaceans							
	Copelata	Oikopleuridae	Oikopleura albicans	OalMT1part	sc7	SNNDGKSVCSGCKNCPPGCNPCKCPKCSEHGCDCSCCNDKPCLARCVGCANCPPDCNPCYCNKCRTNGCSCACCPDDKPCVSRCD GCKNCPPDC	8:10
				OalMT2part	sc8914	MSEPDTDADSDCCGRQCDQCTNCPPNCDPCGCTKCNQRGCSCSCCDDKKKCCDCCQEDCNCTDCESCDSGYCKNSCGCRKNNLC FRKGCCNEGDSCWESNDPPGCNPCKCPKCSKNGCDCSCCDDKPCLPRCDNCANCPPDFCACCPDDKPCVSRCDGCKNCPPDCKPC KCLKCSANGCTCGCCEKVNNCCDC	17:17
				OalMT3part	sc11769	CTKCSQKGCSCSCCDDQKTGCECCKDCDCCKDDCNCTDCESCDCGCCKSNACCDTDCCESKNKCCDCCKDCDGCKDCSAEGCGPC TCDCCQTNSCCEQGCCNVEKKKCCDCCKDCDCCDGCGGNSCGPCSCGCCKNNLCCKKGCCNEGDSCCDGNTEGDSCCGSTSECK PVCSGCKNCPPGCNPCKCPKCSKNGCDCSCCDDKPCVARCDGCANCPPDCNPCYCKKCRTNGCSCACCPDDKPCVSRCDGCKNCP PDCQPCKCAKCSANGCTCGCCAKVSNCCDCCKDCTCCNDCKDGCDANGVCDCACCKNNKCCGTKCCAPKKCCDCCADCLCCPGC GKGTDGNCDCVCCKNNTCCKSDC	40:26
			Oikopleura dioica	OdiMT1	AYN64376.1	MDPVCSFRCCEENCAGCVDCPAGCDPCKCTLEVCKKVCEGCKDCPPGCEPCKCEKCSTKKCKSNCCPTSTAE	9:2
				OdiMT2	AYN64373.1	MEVKRPNNCCPAKCLGCKGCPPGCEPCICNMDTCKNICNKCKECPKNEFGCDPCKCPKCSKLGCTCDCCHKKCCVTDCDGCKTCPP GCEPCKCSMNACKKVCKQCKNCRKSESGCDPCECSKCALKGCKCDCCPKDTCCEASCEGCKNCPPGCEPCKCTLNCCMKICDDCKD CPKSENGCDPCNCRKCSRKGCNCDCCPSDDCCKASCEGCINCPPGCDPCECSMDECKKICKKCNNCRKGESGCDPCECRKCSRNGC DCDCCPKDSCCEASCEGCTDCPQGCKPCKCTMNSCMKTCDKCKDCPKSASGCDPCECLKCSRKGCECDCCPQKNDCCEAFCQGCK NCPPGCNPCKCTLNFCAKICNECKDCPKSDIGCDPCNCVKCSAKGCKCDCCPKKCC	57:24
			Oikopleura vanhoeffeni	OvaMT1	sc728	MNPLAEKSKCATKCAGCKDCPLNCDPCKCCGCSCCGAGGLKLCRSQICEGCRNCPPGCDPCLCPKCKNSGCSCACCEKKPSCNKICA GCANCPPGCDPCLCAKCHNNGCTCSCCADKPCMVRCAGCANCPRNCEPCLCSKCSKNGCDCSCCAKNPCKARCAGCADCPPNCD PCLCSKCSKNGCDCSCCEKKPCASICDGCKDCPPGCKPCKCADCRKNGCGCSCCTTQSNENECC	25:17
				OvaMT2part	sc8397	CKKCPINCEPCMCQICNQNGCNCACCAEKPCMVRCAGCAKCPPNCEPCLCSKCNTKGCNCSCCANKPCKVKCAGCANCPANCDP CFCPKCSKIGCNCSCCENKPCPSICDKCKYCPPGCEPCRCADCRKNGCNCSC	15:14
				OvaMT3part	sc62250	CPPNCDPCFCPKCSKNGCGCSCCENKPCASICDGCKDCPPGCKPCKCADCRKNGCNCSC	7:5
				OvaMT4part	sc79129	AKPCCGQICEGCKNCPPGCDPCLCPKCKKSGCSCSCCEKQPCCGPICEGCKNCPPGCDPCLCPKCKKSGCSCACCDKKPSCNKVCAG CANCPPGCNPCLCPQCSKNGCTCSCCVDKPCNVRCKGCANCPPNCEPCLCPKCSTNGCSCCANKPCKVRCAGCAN	19:12

[1]C. robusta formerly C. intestinalis type A
[2]C. intestinalis formerly C. intestinalis type B
* EST sequences

Table S2. Percentatge of Similarity (down) and identity (up) in protein sequence obtained with GENEIOUS prime (Geneious Prime® 2020.1.2)

	SthMT1	SthMT2	SthMT3	SthMT4_part	SthMT5	SthMT6	
SthMT1		60,98%	68,29%	66,67%	87,80%	65,85%	**Identity**
SthMT2	65,85%		58,54%	42,42%	56,10%	94,87%	
SthMT3	78,05%	70,73%		52,52%	65,85%	60,98%	
SthMT4_part	84,85%	54,55%	66,67%		69,70%	48,48%	
SthMT5	95,12%	63,41%	70,73%	84,85%		60,98%	
SthMT6	68,29%	97,44%	70,73%	57,58%	65,85%		
	Similarity						

TABLE S3. Number of Identical bases in front of the total alignment and percentage of Identity between SthMT1/5 and SthMT2/6.

	5' UTR		INTRON		3' UTR	
	Identity number/Total alignment	% identity	Identity number/Total alignment	% identity	Identity number/Total alignment	% identity
SthMT1 vs SthMT5	41/45	91,11%	339/389	87,14%	186/225	82,66%
SthMT2 vs SthMT6	28/28	100,00%	323/395	81,77%	61/71	85,91%

Table S4: PCR amplification primers for Salpa gDNA.

	Primer	Sequence	Tª Annealing
SthMT1	FwMT1/MT5	CGTCATACTTTTGTGCTTCAAAC	**55**
	RvMT1/MT2	TACACKRATCRCAATGRCA	
SthMT5	FwMT1/MT5	CGTCATACTTTTGTGCTTCAAAC	**50**
	RvMT5	AATCAATTCACAATTGCAAATC	
SthMT6	FwMT2/MT6	CTATCGAACGAATCACTTCAAC	**54**
	RvMT6	GGAAAAAGACAAGACAATTC	

DISCUSSION

1. The enigmatic function of MTs

1.1. The persisting impact of Cadmium on MT evolution

Six decades since MTs first discovery, and after literally thousands of studies describing their structure, biochemical characteristics and tissue distributions, the primary biological function(s) of MTs and the evolutionary history behind this function still remains enigmatic. In fact, it has been recently proposed that '…*searching for "the function" of MTs might not be a productive research strategy. We might be better off asking where the MT family comes from and how it has developed'* (Abdin et al., 2021). In this sense, the systematic characterization of MTs in numerous species of mollusc and chordate phyla carried out by previous works and substantially further extended by the results of this thesis (Calatayud, 2018; Calatayud et al., 2021c, 2021b; Dallinger et al., 2020; Dvorak et al., 2018, 2019; García-Risco et al., 2020; Gil-Moreno et al., 2015; Michael Niederwanger et al., 2017; Pedrini-Martha et al., 2020, 2021)(**see Results I and II, and Annex II, III and IV from this thesis**) might provide a suitable framework to investigate the origin and evolution of the MTs and of their biological function. Our studies have highlighted, for instance, a pervasive Cd(II)-thionein character for most MTs of mollusc/gastropod or chordate/tunicate species we have analysed (Calatayud, 2018; Calatayud et al., 2021c, 2021b; García-Risco et al., 2020; Palacios and Albalat, 2021), which may be related with the evolutionary impact of Cd as an important abiotic factor.

In natural environments, Cd represents a tiny portion (~0.00001%) of the composition of the earth crust (Dallinger et al., 2020; Hans Wedepohl, 1995)(**Annex IV**), and therefore, it may be seen as a negligible metal. The biological significance of Cd is, however, distinctly greater because its high toxicity, even at very low concentrations, and its high ability to compete with Zn in Zn-dependent cellular processes. Cd also competes with calcium (Ca), and hence it might also affect animals like molluscs that depend on Ca pathways for bio-mineralization of their shells (Choong et al., 2014). Evidence for increased Cd emissions through geological eras is provided by data from bedrock formation of different geological origin, from Paleozoic through Mesozoic and Cenozoic (Dallinger et al., 2020)(**Annex IV**). Recent studies in volcanic areas revealed an increasing bioavailability of Zn and Cd in soils (Amaral et al., 2007), suggesting that local concentrations of Cd, even nowadays, may vary.

Currently in the oceans, Cd co-occurs with Zn in ore deposits of the earth crust. Natural Cd concentrations are, therefore, higher where seawater comes into contact with the continental earth crust (10^{-10}M) than in surface oceanic waters (10^{-12}M) (Bruland, 1980)(**Figure 7**). Importantly, Cd concentrations are also high in eutrophic coastal regions of the littoral (10^{-10}M) and its close neritic zone (10^{-10}–10^{-11}M) (Bruland, 1980; Sigel et al., 2013), where water also comes in contact with the seabed, displaying up to 100 times higher Cd(II) than those of superficial oceanic seawater (10^{-12}M) (Bruland, 1980; Dallinger et al., 2020)(**Annex IV**) (**Figure 7**). These areas just off of the coast, from the shoreline to the edge of the continental shelf, bear most of the marine life, including molluscs/gastropods and chordates/tunicates, and the Cd concentrations in them would explain the Cd-thionein character of most MTs of species we have analysed (Calatayud, 2018; Calatayud et al., 2021c, 2021b, 2021a; García-Risco et al., 2020)(**Results I and II**).

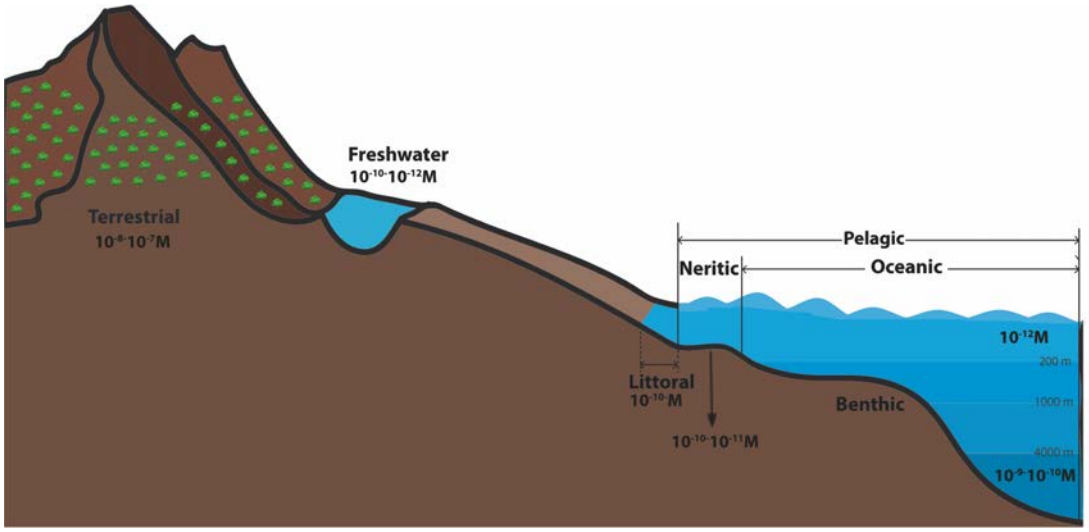


Figure 7. Natural cadmium concentrations presented in molarity (M) in several earth habitats.

The pervasive Cd preference of the MTs studied in this thesis (**Results I, II and Annex IV**) –a feature shared by MTs from many other marine animals, including crustaceans (Narula et al., 1995; Otvos et al., 1985; Valls et al., 2001), annelids (Won et al., 2012) and echinoderms (Riek et al., 1999)(**Table 1**)–, together with the fact that many MTs from aquatic animals are induced as a defence mechanism to Cd exposure (reviewed in (Amiard et al., 2006)) have led us to hypothesize that MTs arose as a detoxification system against the poisonous Cd, and that this ancient function has been preserved in most marine lineages (**Results I, II and Annex IV**). This primeval MT function, however, seems to have changed during the species colonization to other habitats depending on the Cd availability. For instance, several gastropod lineages from the Caenogastropoda and Heterobranchia subclasses have independently abandoned marine realms and successfully adapted to freshwater environments, in which natural

Cd background concentrations (10^{-10}–10^{-12} M) are the lowest of any snail habitat on earth (Boyle et al., 1976; Dallinger et al., 2020; Sigel et al., 2013)(**Annex IV**) (**Figure 7**). The MTs from these freshwater species would have independently lost their Cd-thionein character probably because it was no longer required, and would have acquired new metal preferences and novel biological roles (Dallinger et al., 2020; García-Risco et al., 2020; Michael Niederwanger et al., 2017)(**Results I, Annex II and IV**).

1.2. MT multiplicity and evolution of new functions

Besides the ancestral Cd-detoxification role, the MT system has been diversified and co-opted for other functions as animals colonized new habitats with different metal bioavailabilities and requirements. This process has been frequently associated with scenarios of MT multiplicity, in which neofunctionalization or subfunctionalization events of the MT duplicates have led paralogs to acquire different metal-binding preferences and/or expression patterns. We have observed this process in the MTs of several gastropod species, including *Arion vulgaris* MT1 and MT2 (**Annex III**), *Nerita Peloronta* MT1 and MT2 (**Results I**) and *Pomacea bridgesii* MT1 and MT2 (**Results I**).

One of the best examples of MT duplication followed by changes in metal preference and in expression patterns is that of some helicid land snails such as *Helix pomatia* and *H. aspersa*, in which the transition from aquatic to terrestrial habitats concurred with the duplication and evolution of different MTs: in addition to the primal Cd(II)-MT, these species have a Cu(I)-specific MT and an 'unspecific' Cd/Cu-MT that is able to bind both Cu(I) and Cd(II) ions indistinctly. Interestingly, the new Cu-MT is expressed in rhogocytes, a cell type that produce hemocyanin, the respiratory pigment of gastropods. Hemocyanin is a metalloprotein that contain two copper atoms that reversibly bind a single oxygen molecule, and thereby, the Cu-MT has been related with copper storage and delivery for hemocyanin synthesis in these species (Chabicovsky, 2003; Gil-Moreno et al., 2015; Pérez-Rafael et al., 2014). The function of the unspecific Cd/Cu-MT is still uncertain. Metal exposure does not induce its expression in the hepatopancreas of adult snails, and therefore, its contribution to metal homeostasis and detoxification would be marginal. In contrast, Cd/Cu-MT is highly expressed during embryogenesis, which would indicate that it might have a role in snail development (Pedrini-Martha et al., 2017).

In other invertebrates, MT multiplicity has been also associated with the evolution of new metal preferences from an ancient Cd-MT: the crustacean *Daphnia pulex* has four MTs, two of them induced by Cu(I) (Asselman et al., 2013; Isani and Carpenè, 2014); the insect *Drosophila melanogaster* has six MTs, all of them with a Cu(I)-thionein

character except one form that would have Zn(II) preference (Capdevila and Atrian, 2011; Luo et al., 2020; Silar et al., 1990).

Finally, in vertebrates, the mammalian MT system is a well-characterized example of MT multiplicity. Mammals have four distinct MT subtypes –MT1, MT2, MT3 and MT4– and a lineage-specific number of paralogs, from 16 MTs in humans to four in mouse (Mao et al., 2012; Moffatt and Séguin, 1998; Palmiter, 1998; Quaife et al., 1994; Stennard et al., 1994; West et al., 1990). Mammalian MT1 and MT2 are widely distributed forms with preference for Zn(II), and they are expressed in many cell types in different tissues and organs that have been related with metal homeostasis, transport and detoxification (Artells et al., 2013; Capdevila and Atrian, 2011; Vašák and Meloni, 2011). MT3 shows an intermediate Zn/Cu- thionein character, and it is restricted to neurons where it acts as a neuronal growth-inhibitory factor (Babula et al., 2012; Tío et al., 2004). Finally, MT4 would be a Cu(I)-thionein expressed in the cornified and stratified squamous epithelia with an uncertain function, perhaps related with a homeostatic Cu(I) function (Blindauer, 2014; Tío et al., 2004). Interestingly, a Cd detoxifying function has been questioned for mammalian MTs (Blindauer, 2014). It has been argued that although MT expression is up-regulated in response to exposure to Cd(II), mammalian MTs do not protect from effects of acute Cd toxicity because they would be responsible for prolonged retention of Cd in the body, especially in liver and kidney, with the detrimental long-term effects. The Cd-binding ability of mammalian MTs and their transcriptional activation by Cd would be, therefore, an “accidental feature” of mammalian MTs rather than the result of the evolutionary response to a toxin, to which this animal lineage is rarely exposed (Blindauer, 2014). Our analyses about the ancestral chordate MT (Calatayud et al., 2021b)(**Results II**) and the fact that several marine vertebrate species such as *Pleuronectes platessa, Lithognathus mormyrus and Notothenia coriiceps* have Cd(II)-inducible MTs (Capasso et al., 2003; D’Auria et al., 2001; Overnell and Coombs, 1979; Tom et al., 1998) (**Table 1**) open the possibility that even though the current functions of mammalian MTs may not be related with Cd detoxification, this capability would not be an “accidental feature” but an evolutionary remnant of the ancestral chordate MT.

Table 1. Metallothioneins and their metal binding preference in several phyla collected from other studies.

Phylum	Class	Specie	MT	Habitat	Metal preference	Article
Chordata	Actinopterygii	Oncorhynchus mykiss	MT-A	Feshwater	Cd-thionein	(Isani and Carpenè, 2014)
			MT-B	Freshwater	Cd-thionein	(Isani and Carpenè, 2014)
Chordata	Actinopterygii	Pleuronectes platessa	MT	Marine	Cd-induced	(Overnell and Coombs, 1979)
Chordata	Actinopterygii	Clarias gariepinus	CgMT	Freshwater	Zn-thionein	(M'kandawire et al., 2017)
Chordata	Actinopterygii	Cyprinus carpio	MT-I	Freshwater	Homeostasis	(Van Campenhout et al., 2004)
			MT-II	Feshwater	Cd-thionein	(Van Campenhout et al., 2004)
Chordata	Actinopterygii	Lithognathus mormyrus	MT	Marine	Cd-induced	(Tom et al., 1998)
Chordata	Actinopterygii	Notothenia coriiceps	MT	Marine	Mass and NMR Cd	(Capasso et al., 2003; D'Auria et al., 2001)
Chordata	Mammalia	Mus musculus	MT1	Terrestrial	Cd-thioneins	(Artells et al., 2013)
			MT2		Zn-thionein	(Artells et al., 2013)
			MT3			
			MT4		Cu-thionein	(Capdevila and Atrian, 2011; Tío et al., 2004)
Chordata	Amphibia	Xenopus laevis	MTA	Freshwater		(Saint-Jacques and Séguin, 1993)
			XlMT-C17			(Scudiero et al., 2017)
Chordata	Aves	Gallus gallus	ckMT1	Terrestrial	Zn-thionein	(Villarreal et al., 2006)
			ckMT2		N/D	(Villarreal et al., 2006)
Chordata	Leptocardii	Branchiostoma floridae	BfMT1	Marine	multipurpouse	(Guirola et al., 2012)
			BfMT2	Marine	Cd-thionein	(Guirola et al., 2012)
Chordata	Reptilia	Podarcis sicula	MT	Terrestrial	Cd-thionein	(Scudiero et al., 2017)
Annelida	Clitellata	Lumbricus rubellus	MT1	Terrestrial		
			MT2	Terrestrial	Cd-thionein	(Stürzenbaum et al., 1998)
			MT3	Terrestrial		
Annelida	Clitellata	Eisenia foetida	MT	Terrestrial	Cd-thionein	(Gehrig et al., 2000)
Annelida	Polychaeta	Perinereis nuntia	Pn-MT	Marine	Cd-thionein	(Won et al., 2012)
Nematoda	Chromadorea	Caenorhabditis elegans	MT1	Terrestrial	Cd-thionein	(Isani and Carpenè, 2014)
			MT2	Terrestrial	Cd-thionein	(Isani and Carpenè, 2014)
Echinodermata	Echinoidea	Strongylocentrotus purpuratus	SpMTA	Marine	Divalent Thionein	(Tomas et al., 2013)
			SpMTB	Marine	Cu-thionein	(Tomas et al., 2013)
Arthropoda	Malacostraca	Callinectes sapidus	CdMT-I	Marine	Cd-thionein	(Gehrig et al., 2000; Syring et al., 2000)
			CdMT-II	Marine	Cd-thionein	(Syring et al., 2000)
			CuMT-II	Marine	Cu-thionein	(Syring et al., 2000)
Arthropoda	Branchiopoda	Daphnia pulex	MT1	Freshwater	Cd-stress	(Isani and Carpenè, 2014)
			MT2	Freshwater	Cu-stress	(Isani and Carpenè, 2014)
			MT3	Freshwater	Cd-stress	(Isani and Carpenè, 2014)
			MT4	Freshwater	Cu-stress	(Isani and Carpenè, 2014)
Arthropoda	Malacostraca	Scylla serrata	MT1	Marine	Cd-thionein	(Otvos et al., 1982)
			MT2	Marine	Cd-thionein	(Otvos et al., 1982)
Arthropoda	Malacostraca	Cancer pagurus	MT1	Marine	Cd-thionein	(Overnell et al., 1988)
Arthropoda	Insecta	Drosophila melanogaster	MtnA	Air	Cu	(Capdevila and Atrian, 2011; Pérez-Rafael et al., 2012a)
			MtnB	Air	Cu	(Capdevila and Atrian, 2011; Pérez-Rafael et al., 2012a)
			MtnC	Air	Cu	(Capdevila and Atrian, 2011; Pérez-Rafael et al., 2012a)
			MtnD	Air	Cu	(Capdevila and Atrian, 2011; Pérez-Rafael et al., 2012a)
			MtnE	Air	Cu	(Capdevila and Atrian, 2011; Pérez-Rafael et al., 2012a)
			MtnF	Air	Zn-thionein	(Luo et al., 2020)
Mollusca	Bivalvia	Mytilus edulis	MT-10	Marine	Induced by essential metals	(Blindauer, 2010; Lemoine et al., 2000; Orihuela et al., 2008)
			MT-20	Marine	Cd-thionein	(Lemoine et al., 2000; Orihuela et al., 2008)
Mollusca	Bivalvia	Mytilus galloprovincialis	MT-10	Marine	Induced by essential metals	(Vergani et al., 2007)
			MT-20	Marine	Cd-thionein	(Vergani et al., 2007)

2. The unusual evolution of MTs

2.1. The MT classification dilemma

The first attempts to develop a classification system for MTs was based on sequence similarity with the first MT discovered, i.e. equine MT (Margoshes and Vallee, 1957). Thus, three classes were defined: class I, essentially comprising vertebrate, crustacean and mollusc MTs; class II, comprising MTs that did not display any clear sequence similarity to vertebrate MTs; and class III, comprising biosynthetic polypeptides such as the phytochelatins (Fowler et al., 1987; Kägi and Kojima, 1979). This classification system, however, was questioned because the enormous diversity of primary and 3D structures and biological functions of these proteins. A refined classification system was developed not only based on similarity of the coding sequences, but also including intronic sequences and analysis of upstream 5' regions of the gene that include the metal-response elements (Binz and Kägi, 1999b). The MT sequences known in 1997 were divided into 15 families, almost one family for each taxonomic group with available MTs at that time. This classification, however, was of little use because it did not provide any clear phylogenetic relationship between families.

To overcome this drawback, alternative evolutionary markers that do not rely on sequence alignments –e.g. intron indels, retroposon integrations, signature sequences, gene order changes, gene duplications or genetic code changes– have been shown to have an enormous potential for testing the strength of the evolutionary history of gene families whose phylogenetic analyses based on primary sequence data are doubtful (Rokas and Holland, 2000). With this purpose, we have used the patterns of cysteines, that is, the different number and distribution of CxC, CC, and CCC motifs in the MT domains, as a molecular marker for untangling MT evolution. Cysteine patterns are under functional selection because they determine the organization of the metal-thiolate clusters since identical locations of coordinating cysteine residues result in clusters with the same domain topology. From comparisons of a large number of MT sequences belonging to a particular taxon, the patterns of cysteines in the MT domains might be determined, and based on this, we have been able to infer the evolution of mollusc/gastropod and chordate/tunicate MTs (**Results I and II**).

2.2. Patterns of cysteine motifs and MT domains in molluscs

From comparisons of a large number of MT sequences, the patterns of cysteines in the molluscan MT domains have been determined, allowing us to classify them in six types: the α domain with 11/12 cysteines ($[CxC]x_6[CxC]x_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$); the β1, β2 and β3 domains with 9 cysteines organized in different patterns of C motifs ($[CxC]x_3[CxC]x_3Cx_5[CxC]x_3[CxC]$, $[CxC]x_5[CxC]x_3[CxC]x_4[CxC]x_4C$) and

($Cx_3Cx_4[CxC]x_3[CxC]x_3[CxC]x_2C$, respectively); the γ domain with 10 cysteines ($CCx_5CCx_4CCx_6CCx_7CC$); and the δ domain with 14 cysteines ($[CxC]x_3[CxC]CCx_4Cx_3[CxC]x_3CCx_4[CxC]C$). Most Mollusca MTs are bi-modular proteins that combine a conserved C-terminal β1 domain with one out the five distinct N-terminal domains –α, β2, β3, γ or δ– in a lineage-specific manner.

Interestingly, α and β domains show similarities in the patterns of cysteine motifs, so we proposed they share a common ancestor (**Results I: see Figure 5 of Article I**). In contrast, the patterns C motifs of the γ and δ domains were completely different, which led us to propose that these domains would have emerged *de novo*. Finally, from the analysis of the domains in the different molluscan lineages (**Results I: see Figure 1 of Article I**), we concluded that the mollusc ancestor probably had two MTs, one α/β1 and one β2/β1, that could undergo several linage-specific changes. These changes included duplications or losses of the MTs, as well as modifications of their domains that could have been expanded, swapped or emerged de novo during the diversification of the phylum (Calatayud et al., 2021c; García-Risco et al., 2020) (**Results I**).

2.3. Patterns of cysteine motifs and MT domains in chordates

Similar analyses of many chordate MTs have allowed us to describe the patterns of cysteine motifs and the domain structures of cephalochordate, tunicate and vertebrate MTs. Previous studies had described vertebrate MTs as bi-domain proteins made of an amino-terminal 9Cys domain, called β-domain ($[CxC]x_5[CxC]x_3[CxC]x_2[CxC]x_2Cx_3$), and a carboxyl-terminal 11Cys domain, called α-domain ($CCxCCx_3Cx_2Cx_3[CxC]x_6CxCC$). In cephalochordates, we extended the MT analyses to several species, and showed cephalochordates have two MTs, MT1 and MT2, and that MT2 had two splice variants (MT2S and MT2L, for short and long, respectively). The MT1 was a bi-modular MT with a 11Cys domain ($[CxC]x_5[CxC]x_{3\text{-}4}[CxC]x_3[CxC]x_3[CxC]x_2C$) and a 9Cys-$_{LIKE}$ domain ($CCx_2Cx_2Ix_3Cx[CxC]xCC$), whereas MT2 had two (MT2S) or three (MT2L) 9Cys domains ($CCx_2Cx_2Cx_3Cx[CxC]xCC$). Finally, the analysis of the MTs of tunicates, which encompass three different classes –ascidiacea, thaliacea and appendicularia–, showed that ascidians and thaliaceans possessed mono-modular 12Cys MTs ($[CxC]x_5[CxC]x_2Cx_2Cx_2[CxC]x_4[CxC]x_{3-6}CC$), while appendicularian MTs were bi-modular proteins made also of 12Cys domains but with a distinct Cys pattern ($Cx_3Cx_2Cx_2Cx_3Cx_2[CxC]x_2Cx_4[CxC]xCC$). A detail analysis of the MTs of the appendicularian *O. dioica* revealed that this species displays two MTs, a bi-modular OdiMT1 and a multi-modular OdiMT2. Sequence comparisons showed that both MTs were made of a repeat unit (RU) or 'supradomain' (SD; see also Section 3 of Discussion)

encompassing a trimmed 12C (t-12C) domain ($Cx_3Cx_2Cx_2Cx_3Cx_2$[CxC]) at the N-terminal, and a 12C domain ($Cx_3Cx_2Cx_2Cx_6Cx_2$[CxC]x_2Cx_4[CxC]xCC) at their C-terminal. This RU was repeated one time in OdiMT1 and a variable number of times (from three to seven) in OdiMT2 (**Results II: see Figure 1 of Article IV, and Figures 1 and 5 of Article V**; see also Section 3 of Discussion).

With all this data, we reconstructed the evolutionary history of MT in chordates. We proposed that the ancestral chordate MT might have been similar to the current cephalochordate MT1, made of two domains: an amino-terminal 11/12C domain (α domain) and a carboxyl-terminal 9C domain (β domain). In cephalochordates, the original MT was duplicated and one copy was mostly preserved, as it is the MT1, but the other copy was modified losing the 11Cys domain and expanding the 9Cys domain, as it is in the MT2. Vertebrate MTs suffered domain readjustments, shortening the N-terminal domain to 9Cys and enlarging the C-terminal domain to 11Cys. In addition, successive lineage-specific tandem duplications would have led to a variable degree of the MT multiplicity in the different vertebrate species. Finally, in tunicates, the ancestral chordate MT would have lost the 9Cys domain in the ascidian/thaliacean linages to become a mono-modular MT. In appendicularians, the changes would have been even more drastic: the ancestral MT would have been lost and appendicular MTs would have emerged de novo, first as a bi-modular form that was later duplicated and the RU expanded in one of the copies (Calatayud, 2018; Calatayud et al., 2021b, 2021a)(**Results II: see Figures 6 and 7 of Article VI**).

2.4. Beyond mollusc and chordate MTs

The patterns of cysteines motifs of other animal phyla have not been yet systematically analysed and resolved. In an attempt to extend our discussion from molluscs and chordates to other metazoan MTs, we have investigated the cysteine patterns of MTs in other animal phyla using the available sequences in current databases, paying special attention to the deuterostome Echinodermata phylum, and to the protostome lophotrochozoa Annelida and ecdysozoa Arthropoda phyla (**Figure 8**).

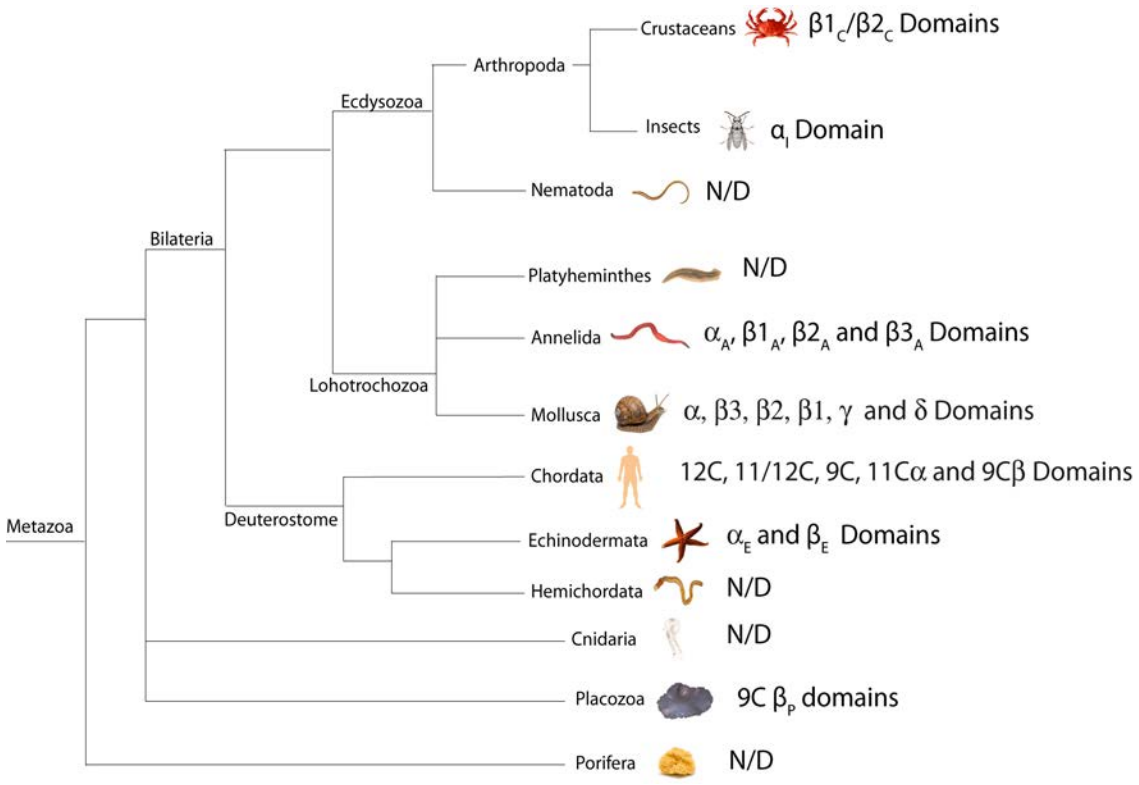


Figure 8 ***Phylogenetic relationships in the metazoan kingdom.*** *At each side of the phyla is a representative image of them and the identified domains. If no cysteine pattern is found and no domain can be determined, we indicate it (N/D).*

2.4.1 Echinoderm MTs

Echinodermata and Hemichordata phyla constitute the Ambulacraria clade, which is the sister group of the Chordata phylum. Echinodermata contains about 7000 living species of exclusively marine animals, including starfish, sea urchins and sea cucumbers. In echinoderms, MTs were initially identified in sea urchin species, *Strongylocentrotus purpuratus* (Harlow et al., 1989; Riek et al., 1999; Tomas et al., 2013) and *Paracentrotus lividus* (Ragusa et al., 2017, 2013), and described as bi-domain proteins with an amino-terminal 11/12C domain (α_E type, E stands for echinoderm; [CxCC]x_5[CxC]x_4CCx_4CC$x_{4\text{-}6}$CC), and a carboxyl-terminal 9C domain (β_E type; Cx_4[CxC]x_3[CxC]x_3[CxC]x_4[CxC])(**Figure 9A**). It was suggested that the αβ organization of echinoderm MTs was a specular reflection of vertebrate (chordate) βα MTs (Riek et al., 1999), although the molecular mechanisms that have might led to this domain reorganization remain unknown. Interestingly, the echinoderm αβ organization is conserved in the MTs of Echinoidea class (sea urchins), Holothuroidea (sea cucumbers) and Ophiuroidea (brittle stars) but has clearly diverged in Asteroidea (sea stars) classes. In sea stars, MTs are made of two α domains (**Figure 9B**), meaning that the β domain has been lost during the evolution of this lineage. These observations further support the

flexible modular organization of the MT proteins, for which domains might be gain or lost in a lineage-specific manner.

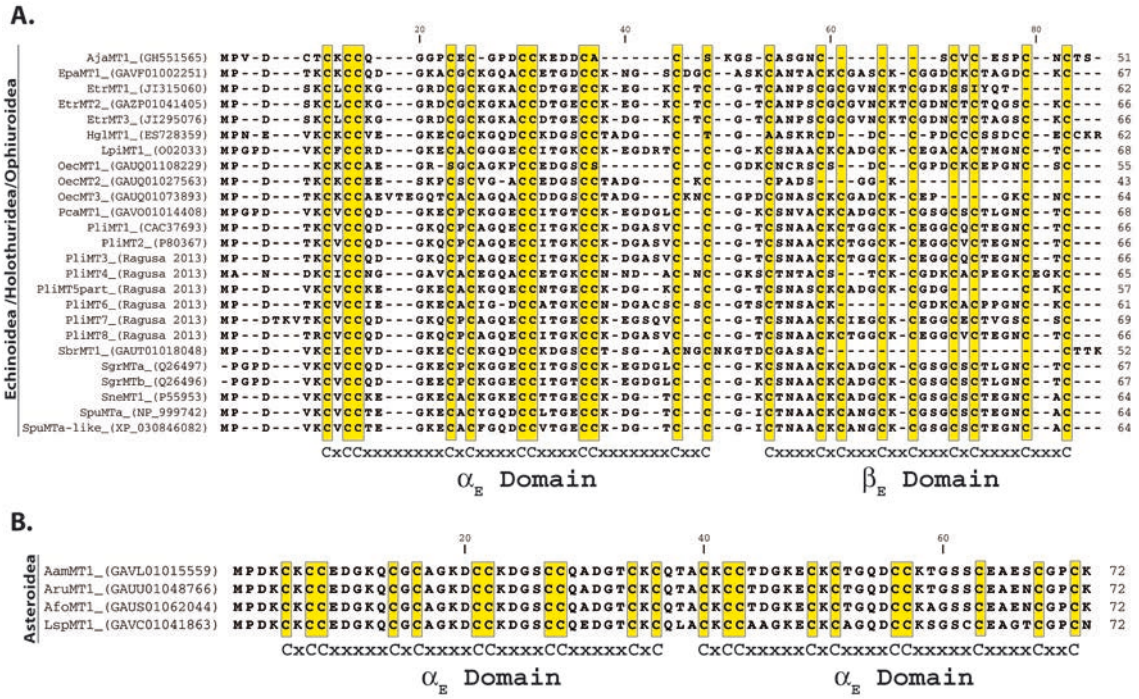


Figure 9 ***Echinoderm amino acid alignment of MTs.*** *Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs at the N-terminal and C-terminal domains are showed below each alignment. Names of MT sequences are alphabetically ordered. Alignment of Echinoderm MTs showing a bi-modular structure (A) in Echinoidea, Holoturoidea and Ophiuroidea with an α_E/β_E structure and (B) in Asteroidea an α_E/α_E structure.*

2.4.2 Annelid MTs

Annelida is a large phylum within the lophotrochozoa clade, a group of protostomes that also includes the Mollusca phylum. With over 22,000 extant species of segmented worms, this phylum has been traditionally classified into two classes, Polychaeta (mostly marine worms) and Citellata (earthworms and leeches), although this classification is under revision (Struck et al., 2011, 2007). Annelid MTs typically combine two out of four possible domains with different cysteine patterns: an 11/12C (α_A type, A stands for Annelid) domain ($CCx_4Cx_6[CxC]X_2[CxC]x_4Cx_3Cx_3CC$), and three 9C ($\beta_A$ type) domains, named $\beta1_A$ ($Cx_4[CxC]X_3[CxC]x_{4\text{-}5}Cx_3Cx_3CC$), $\beta2_A$ ($[CxC]x_3[CxC]x_9[CxC]x_3[CxC]x_3C$), and $\beta3_A$ ($Cx_4Cx_4[CxC]X_3[CxC]x_{4\text{-}5}Cx_3Cx_3C$). MTs from polychaete species are bi-domain $\beta2_A\beta1_A$ MTs, whereas MTs from clitellates are $\alpha_A\beta1_A$ or $\alpha_A\beta3_A$ MTs (**Figure 10A, 10B and 10C**). Remarkably, the MT from clitellate *Enchytraeus buchholzi* is a multi-domain MT of 251 amino acids, made of eight tandem repeats of a $\beta1_A$ domain (R1-R8, **Figure 10B**) (Schmitt-Wrede et al., 2004), and its identification in this species extends the presence of multi-domain MTs to the Annelida phylum. Although the patterns of C motifs in α_A, $\beta1_A$, and $\beta3_A$ domains are different, they still retain some degree of similarity between them, suggesting a common evolutionary

origin from an ancestral annelid α_A domain. The number of annelid available MTs is, however, too small and the taxonomic diversity is too narrow to have a complete perspective of the origin and evolution of the MTs in this phylum.

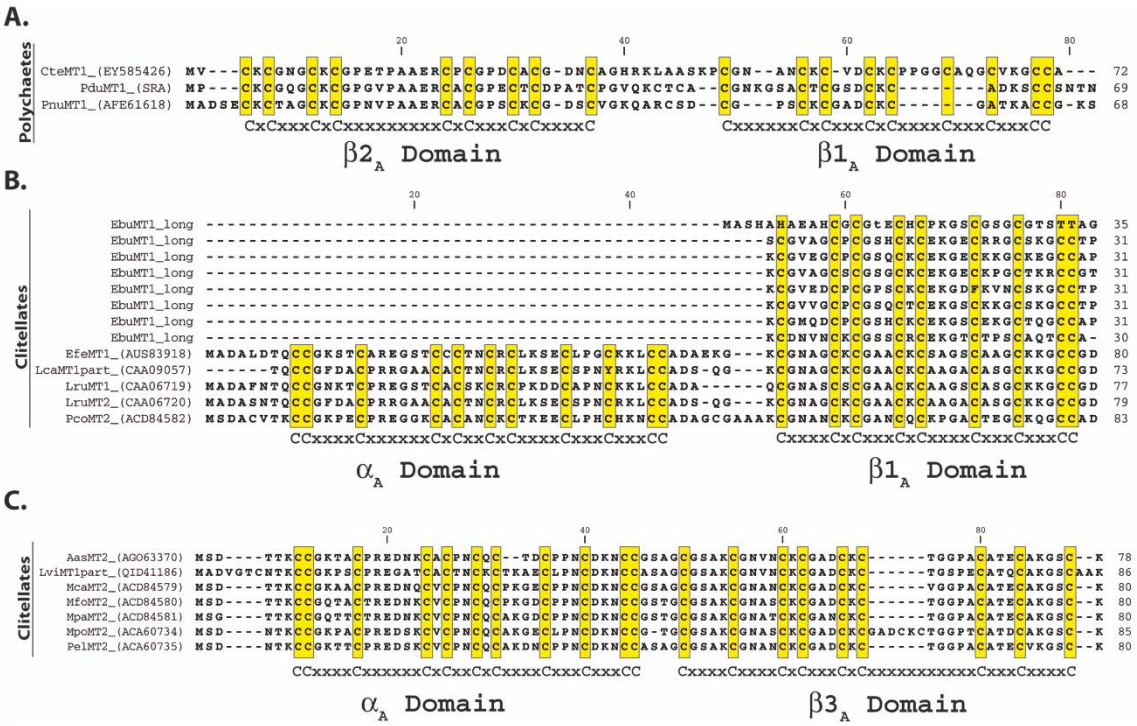


Figure 10 ***Annelid amino acid alignment of MTs.*** *Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs at the N-terminal and C-terminal domains are showed below each alignment. Names of MT sequences are alphabetically ordered. Alignment of Annelid MTs showing a bi-modular structure (A) in Polychaetes with an $\beta2_A/\beta1_A$ structure; (B) in Citellates an $\alpha_A/\beta1_A$ structure and (C) in other Citellates an $\alpha_A/\beta3_A$ structure.*

2.4.3 Arthropod MTs

The phylum Arthropoda comprises >80% of all known living animal species and includes insects, arachnids, myriapods, and crustaceans. The two main subphyla, Crustacea and Hexapoda (insects), have different MTs. Crustacean MTs (**Figure 11)** consist in two 9C domains (β_C type; C stands for crustacean): an amino-terminal $\beta1_C$ (a.k.a. ββ or β_N domain; $CCx_3[CxC]x_4Cx_3[CxC]x_2[CxC]$), and a carboxyl-terminal $\beta2_C$ (a.k.a. βα or β_C domain ($Cx_2Cx_3[CxC]x_5Cx_3Cx_3[CxCC]$) (Muñoz et al., 2002; Narula et al., 1995; Valls et al., 2001). In contrast, insect MTs (**Figure 12)** have a single 12C domain (α_I type; I stands for insect) ($Cx_2Cx_3[CxC]x_4Cx_3[CxC]x_3[CxCxC]x_7CC$) (Liu et al., 2021, 2014; Luo et al., 2020; Pérez-Rafael et al., 2012a), without any obvious similarity with the crustacean organization.

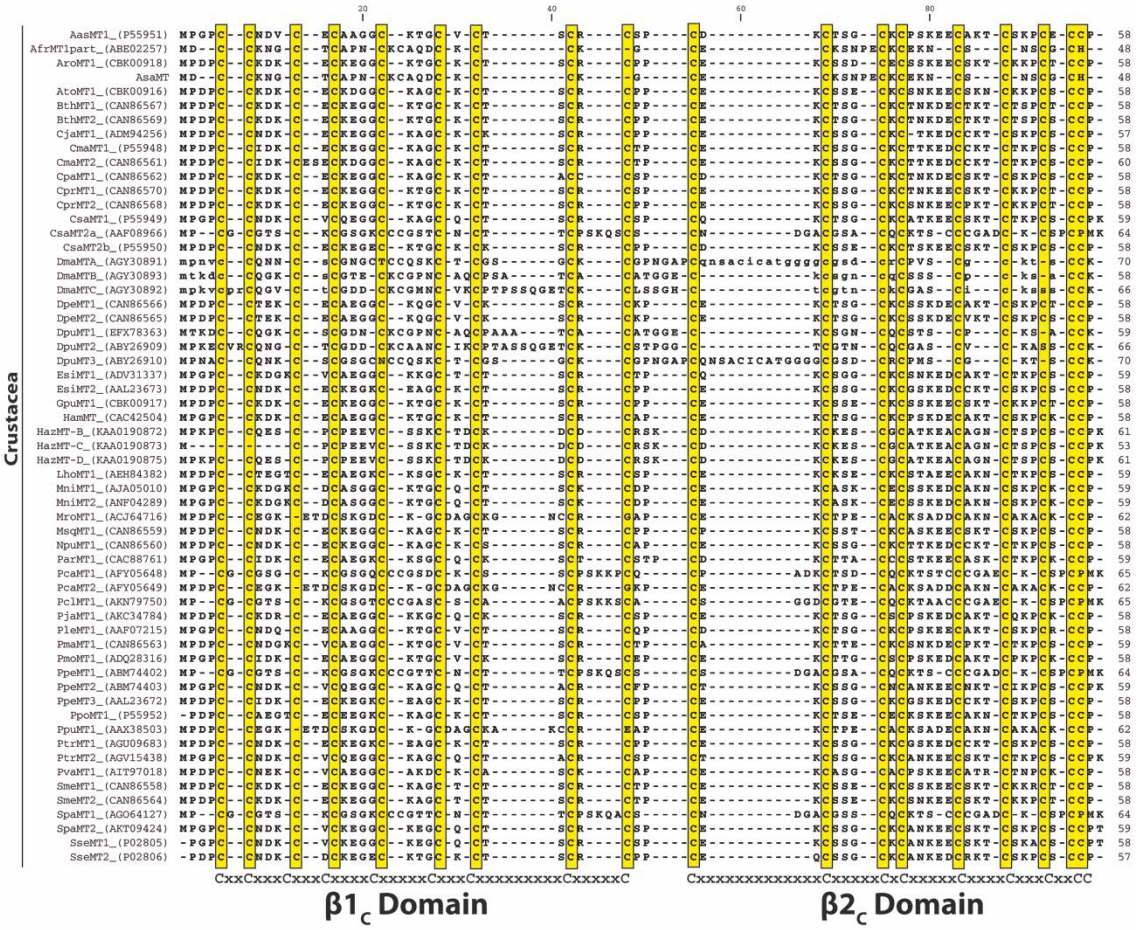


Figure 11 ***Crustacea amino acid alignment of MTs****. Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs at the N-terminal and C-terminal domains are showed below each alignment. Names of MT sequences are alphabetically ordered. Alignment of Crustacea MTs showing a bi-modular* $\beta 1_C/\beta 2_C$ *structure.*

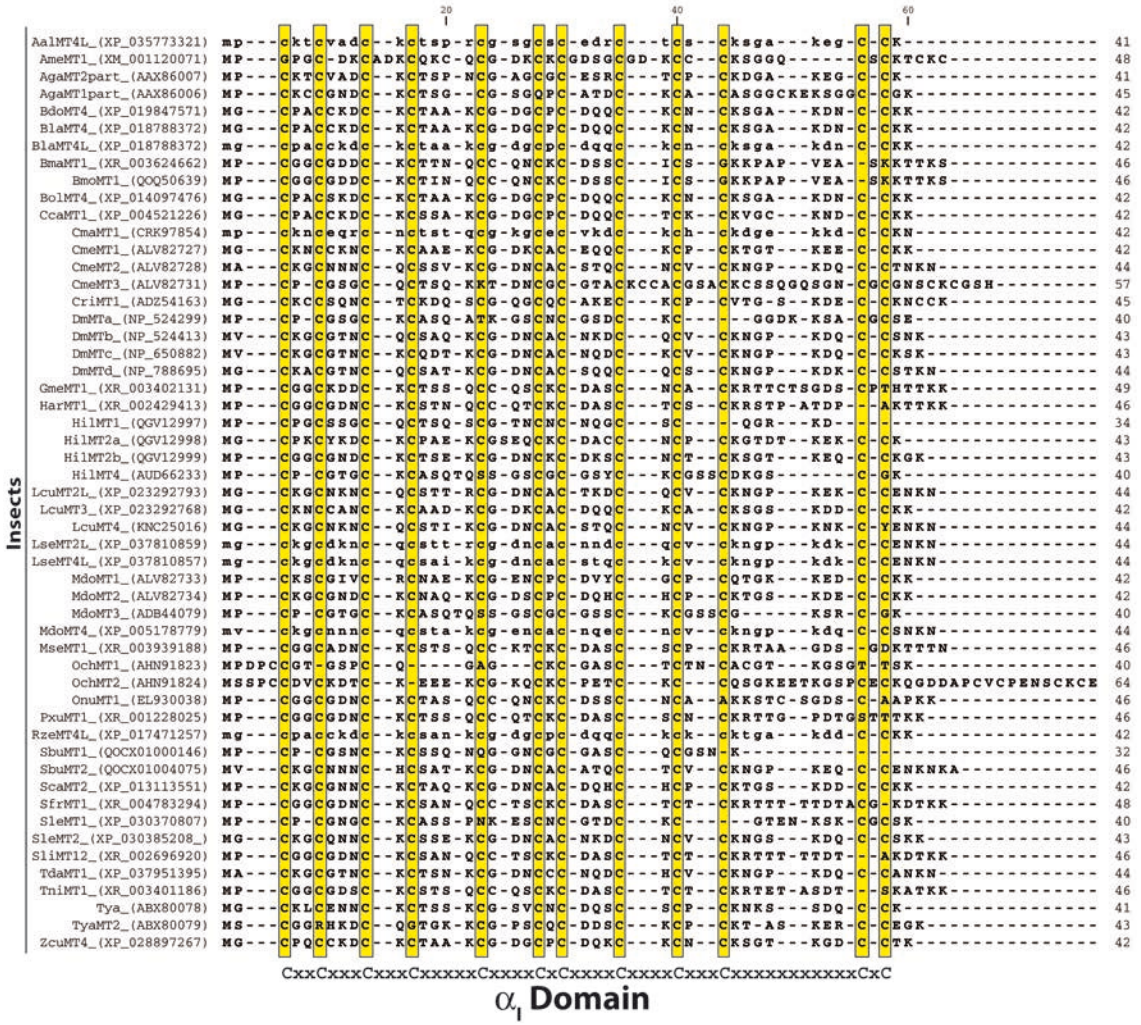


Figure 12. ***Insect amino acid alignment of MTs.*** *Conserved cysteines are highlighted in yellow. Configurations of cysteine motifs at the N-terminal and C-terminal domains are showed below each alignment. Names of MT sequences are alphabetically ordered. Alignment of Insect MTs showing a mono-modular α_1 structure.*

2.4.4 Other animal MTs

Information about MTs from other bilaterian animals is scarce. MTs have been reported in the nematode *Caenorhabditis elegans* and in the platyhelminthe *Schistosoma mansoni. C. elegans* presents two similar MTs, MTL-1 and MTL-2, with 19 and 18 cysteines, respectively, organized as $[CxCxC]x_4[CxC]x_{3\text{-}6}[CxC]x_4CCx_3CCx_7CCx_3Cx_3[CxC]x_2[CxC]x_5C$ motifs (the last C is absent in MTL-2) (Liu et al., 2021, 2014; Luo et al., 2020; Pérez-Rafael et al., 2012a), whereas *S. mansoni* MT is a large MT with 226 amino acids and 34 cysteines, which domain organization is unknown (Blindauer, 2014).

In non-bilaterian phyla, a partial putative MT has been reported in the cnidarian *Acropora cervicornis,* and a potential MT has been identified in the placozoan *Trichoplax adhaerens* (Blindauer, 2014). Interestingly, placozoan MT has 19 cysteines probably organized into two 9C β_P domains (P stands for placozoa; $[CxC]x_{5\text{-}6}[CxC]x_{2\text{-}3}[CxC]x_2[CxC]x_3C$), with one extra C at the end of the N-terminal domain.

2.5 The origin of the MTs

Comparisons of the amino acid sequences have been used to reconstruct the evolutionary histories of many genes and proteins (Doolittle, 1981). For MTs, however, the cysteine richness in the sequences distorts these comparisons forcing the alignment of cysteine positions by inserting many gaps and resulting in a false impression of similarity between non-homologous MT sequences. To overcome this problem, we have used the comparison of the patterns of Cys motifs as evolutionary markers since these patterns are under functional selection, and thereby, it is reasonable to assume that MTs sharing a common ancestor also share the pattern of cysteines.

Comparisons of the MTs of the different metazoan phyla did not reveal a pattern of Cys motifs common to all the animal lineages, but we were able to find a similar pattern in some phyla made of four or five [CxC] motifs and one or two additional Cs arranged in a $[CxC]x_n[CxC]x_nCx_n[CxC]x_n[CxC]x_n[CxC]x_nC$ pattern. To some extent, this is the pattern of the mollusc α domain (and of the related mollusc β domains) (**Results I, see Figure 5 of Article I**), the chordate 11/12C domain (and the vertebrate β-9C domain) (**Results II, see Figure 7 of Article VI**), the echinoderm β_E domain (**Figure 9A**), the annelid $\beta2_A$ domain (**Figure 10A)**, and the placozoan β_P domain. The phylogenetic distribution of these domains suggests that the most plausible cysteine pattern in the MT of the last common ancestor of metazoans may be 11/12 cysteines arranged in C motifs resembling the current mollusc α domains (**Figure 13**). In some animal lineages, this ancestral 11/12C α-type domain would have been duplicated and trimmed to 9C β-type domains. In other lineages, however, this domain appears to have been lost and replaced by other MT domains that did not show any obvious similarity with the ancestral form: 12C α_A, 9C $\beta1_A$ and 9C $\beta3_A$ domains in annelids (**Figure 10**), 10C γ and 14C δ domains in molluscs (patellogastropods and caudofoveates, respectively) (**Results I**), 9C $\beta1_C$ and $\beta2_C$ domains in crustaceans (**Figure 11)**, 12C α_I domain in insects (**Figure 12)**, 12C- 9C- and 11C- domains in chordates (appendicularians, cephalochordates and vertebrates, respectively) (**Results II, see Figure 7 of Article VI**). Although we cannot rule out that these domains derived from an ancestral 11/12C α-type domain that have diverged too much for homology to be recognized, their pattern of the Cys motifs are so divergent from the ancestral form that they seem to have evolved *de novo* (**Figure 13**). It has been proposed that *de novo* domains are low-complexity sequences (Toll-Riera et al., 2012) and tend to be structurally disordered (Moore and Bornberg-Bauer, 2012). Such structural features match well those of MTs, which are considered as low complexity and intrinsically disordered proteins. Thus, *de novo* emergence of MT domains might be relatively easy under an evolutionary perspective because the only

requirement for a peptide to function as a metal ion chelator would be a high content of coordinating residues (e.g., Cys) and a relatively small length that favoured the polypeptide folding (Capdevila and Atrian, 2011).

Interestingly, despite being different in the pattern of Cys motifs, many MT domains have 11/12 Cys (**Results II, see Table 1 of Article VI**), probably because different patterns of Cys motifs converge to similar 3D architectures (Beil et al., 2019). It might therefore be argued that 3D structural constraints to efficiently coordinate four metal ions through the formation of metal–thiolate bonds have led to the convergent evolution of 11/12 Cys domains in MTs of different animal phyla.

In summary, our data suggest that although a metazoan ancestral domain seems to have been preserved in several animal lineages, diverse *de novo* domains have emerged throughout MT evolution. In some lineages, the new domains have been combined with the old ones (e.g. in molluscs and chordates), whereas in others, the old domains have been lost, being replaced by the novel ones in brand new MTs (e.g. in crustaceans and insects). This means that 'metallothioneins' are a polyphyletic group with an extraordinarily complex evolutionary history that have merged de novo emergence of genes and domains, processes of convergent evolution, events of gene gains and losses, and recurrent amplifications of functional domains. MTs stand for a fascinating case in the field of protein evolution that we are just beginning to understand.

3 Multi-domain MTs

3.1. Repeat number and supradomains in multi-domain MTs

MTs have usually been considered small proteins made of one or two domains, and therefore, the discovery of large multi-domain MTs in a few species (de Francisco et al., 2016; Iturbe-Espinoza et al., 2016; Jenny et al., 2016; Palacios et al., 2014) was unexpected. The results reported in this thesis, identifying multi-domain MTs in many gastropods (Calatayud et al., 2021c) (**Results I**), ascidians (Calatayud et al., 2021b) (**Results II**) and appendicularians (Calatayud, 2018; Calatayud et al., 2021a) (**Results II**), enlarge the catalogue of species with large multi-domain MTs, and reveal that these multi-domain forms have recurrently emerged in different taxa across the tree of life. The number of domain repeats in MTs is variable between species, ranking from the two domains of the classic forms, up to ten domains in the snail AbiMT4 (**Results I : See Figure 4 of Article I**), which is the MT with the highest number of domains reported so far (Calatayud et al., 2021c; Pedrini-Martha et al., 2020). Interspecies differences in the number of repeats are frequent between orthologs of domain repeat proteins. The abnormal spindle-like microcephaly-associated proteins (ASPM), for instance, parade 71

calmodulin binding domains (IQ) in human, 62 in mouse, 63 in zebrafish, 22 in the fruit fly and 6 in worms (Björklund, 2006). Similarly, human Titin, the largest domain repeat protein reported so far, has 166 immunoglobin (Ig) and 132 fibronectin III (FnIII) domains, whereas *C.elegans* titin only has 56 Ig and 11 FnIII domains (Ferrara et al., 2006; Flaherty et al., 2002; Kenny et al., 1999).

Regarding the unit of variation in the number of repeats, it is interesting to notice that some domain families have been found to duplicate consistent numbers of domains instead of duplicating one repeat at a time (Björklund, 2006; Björklund et al., 2010, 2005). The giant muscle protein Nebulin, for instance, presents duplications either of a single domain or a group of 7 domains (Björklund et al., 2010; Donner et al., 2004; McElhinny, 2003). In this last case, the unit of variation is not a single domain but a group of a fixed number of domains called supradomain (SD). The evolution of SD in proteins can be considered as a strategy to easily expand domain repeat proteins in length that would only be restricted by functional and structural constraints of the protein (Fraternali and Pastore, 1999; Higgins et al., 1994; Kenny et al., 1999). We have identified one SD in the multi-domain OdiMT2 of *O. dioica*. This species has two MTs, a bi-modular OdiMT1 consisting of two domains (t-12C and 12C), and a multi-modular OdiMT2 with six t-12C/12C repeats (Calatayud, 2018; Calatayud et al., 2021a)(**Results II**). We have shown that the t-12C/12C pair is a SD that has been duplicated several times during OdiMT2 evolution, which even show interpopulation variability (see below). We argued that this SD is the unit of variation because, although the 12C domain is able to autonomously coordinate metal ions, the t-12C/12C pair might have been selected as an optimized unit for binding divalent metal ions (Calatayud et al., 2021a) (**Results II**).

3.2. Inter- and intra-species variability in repeat number

Variations in repeat number are presumed to occur through tandem repeat domain duplications via recombination events resulting in repeat expansions (Andrade et al., 2001b; Marcotte et al., 1999). Marcotte et al. found that the longer is a protein repeat, the more likely it is to be expanded further, so, it is plausible to argue that once a duplication has occurred, it is also likely to occur again (Marcotte et al., 1999). Furthermore, it seems that duplications of domains are an ongoing process and possibly they are more common than deletions since the repeats have expanded so rapidly (Björklund et al., 2010). It is noteworthy that, however, the number of repeats is well conserved within a species, and it does not show intraspecific variability in terms of domain duplications or deletions (Schaper et al., 2014). A recent analysis of human proteins found that, in fact, none of the analysed domain repeat proteins underwent

recent changes in their repeat number, and a majority of human domain repeat proteins were strongly conserved among all mammals (Schaper et al., 2014).

These observations strongly contrast with our results in *O. dioica* and make the case of appendicularian MTs unique and exceptionally interesting. We have revealed the presence of variations in the number of domain repeats in OdiMT2 from four geographically distant *O. dioica* populations, that is, north Atlantic, eastern Pacific, western Pacific (two populations, Okinawa and Osaka) and Mediterranean populations (Calatayud et al., 2021a) (**Results II**). The number of domain repeats was six, seven, three (Okinawa) or five (Osaka), and four, respectively (**Results II: See Figure 5 of Article V**). Gene-structure analyses and phylogenetic reconstructions of these variants suggested that the increase/decrease in the number of repeats was the result of internal tandem duplications/losses probably due to recombination events. In addition, the fact that in all variants the repeat units comprised a SD made of the t-12C and12C domains supports a modular, step-wise mode of evolution for OdiMTs as discussed above (Calatayud et al., 2021a) (**Results II**). We do not know, however, if the intraspecies variability of OdiMT2 may be a general trait of multi-modular MTs, or it results from the peculiar evolution of *O. dioica*, which is the fastest evolving metazoan reported so far ((Denoeud et al., 2010; Edvardsen et al., 2005) reviewed in (Ferrández-Roldán et al., 2019)). In summary, our findings revealed that the number of repeats fluctuates between current *O. dioica* populations, which provides a new perspective on the intraspecies variability of domain repeat proteins and opens the door to study the presence of selection signatures reflecting adaptive processes to different environmental conditions in each marine habitat (e.g. presence of different heavy metals or environmental redox hazards).

3.3. Emergence and sequence evolution of domain repeat proteins

The molecular mechanism behind the emergence of domain repeat proteins is not fully understood, but exon shuffling and DNA slippage events have been proposed responsible for the rise of tandem duplications within a gene (Andrade et al., 2001b). However, once domain repeat proteins have arisen, it appears that domain repeats might confer several advantages on proteins, and therefore, to the organisms that accommodate them. It has been proposed, for instance, that domain duplication in multi-domain proteins is error prone and allows for more rapid evolution than non-repeat containing proteins (Marcotte et al., 1999). Domain repeat proteins provide therefore the organisms with a relatively quick way of adapting to new environments. For example, the rapid evolution of leucine-rich repeats (LRR), a repeat domain present in many plant

resistance genes, appears to drive adaptations in plant innate immune systems (Ellis et al., 2000; Padmanabhan et al., 2009).

In MTs, the emergence of multi-domain forms had been associated with biological adaptations, as virulence factors in the pathogenic fungus *C. neoformans* or for Cu-providing to lignin-metabolizing enzymes in saprophytic fungus living in decaying woods for *T. mesenterica* (Iturbe-Espinoza et al., 2016; Palacios et al., 2014) (**Annex I**). Our analyses of tunicate MTs have shown that the evolution of multi-domain MTs would be a way for creating large MTs with high cysteine content and a high capacity of metal binding (Calatayud, 2018; Calatayud et al., 2021b, 2021a) (**Results II**), and thereby, it might have contributed to the adaptation of the species to different conditions of metal bioavailability and stress. The ecophysiological determinants that have favoured the evolution of large MT in only some animal species are, however, still an open question.

Regarding sequence evolution, the repeat 'purity' in several domain repeat proteins has been studied. Repeat purity is defined as the average pairwise sequence identity between all individual repeat units within one protein. Fraternali et al., for instance, showed that most Ig domains in the titin protein were 20-35 % identical and 40-60 % similar, and only a tiny fraction (<10%) of adjacent domains were more than 40% identical (Fraternali and Pastore, 1999). In general, the repeat purity in most of domain repeat proteins is low, probably due to the evolutionary pressure to avoid misfolding and aggregation of the domain repeats (Borgia et al., 2015). It has been demonstrated that above 40% sequence identity, protein aggregation and misfolding become very likely (Wright et al., 2005), and this may be the reason for why repeats above that degree of similarity are exceedingly rare in naturally occurring proteins (Reshef et al., 2010; Street et al., 2006; Wright et al., 2005).

Multi-domain MTs are, however, an exception to this general rule. The domain repeats of *Tetrahymena malaccensis* MT (TmalaMTT3) are ≈90% identical (de Francisco et al., 2016), and analyses of sequence conservation between OdiMT2 repeats show a striking 50-90% identity (unpublished data). It may be argued that the higher purity of domains in MTs than in other domain repeat proteins would be due to their different biological constraints. Most domain repeats are found almost always associated with additional protein domains, and the repeats are mainly involved in protein-protein interactions for which proper domain organization and folding are essential (Björklund, 2006; Björklund et al., 2005; Heringa, 2005; Jernigan and Bordenstein, 2015; Marcotte et al., 1999). In contrast, MTs are exclusively made of 'metallothionein' domains that are folded when they bind metals. Misfolding and aggregation issues of the domain repeats

would not be, therefore, a major problem for multi-domain MT proteins, whose main biological role is not the interaction with other cellular components, but the stress response, and the homeostasis and detoxification of heavy metals.

Overall, the results from this thesis reveal the evolution of multi-domain structures and combination of old and new domains as pervasive features of the history of MTs. These features have probably been favoured by the modular organization of these proteins, being a single structural domain or a set of domains that conform a SD the basic evolutionary module. This thesis, therefore, highlights the MT protein family as a paradigmatic case study to better understand modular evolution of proteins, including different levels of analysis such structure, functional activity, evolutionary adaptability and genetic population variability.

CONCLUSIONS

1- Our studies in molluscs and chordates have shown an extraordinary dynamism in their MT proteins, in which their modular organization would have made it easier for MT domains to have been gained, lost, mixed, exchanged, expanded or *de novo* created several times throughout their evolution.

2- Our results have validated the usefulness of molecular markers that do not rely on sequence alignments for inferring the evolutionary histories of protein families whose phylogenetic analyses are equivocal or incorrect. The analysis of the patterns of Cys motifs as an evolutionary marker has revealed that although some MTs might share a common ancestor, metazoan MTs are polyphyletic.

3- Our data have also revealed a pervasive preference for cadmium for most of the MTs from different marine species. These results have led us to propose that MTs arose as a detoxification system against the Cd concentrations in the seawater, exemplifying how an abiotic factor might have favoured the convergent evolution in protein function.

4- Our analyses have revealed numerous cases of MT multiplicity and multi-domain MTs. MT multiplicity has been associated to the acquisition of new metal preferences and/or of novel expression patterns. Multi-domian MTs have a high capacity of metal, and thereby, they might have contributed to the adaptation of the species to different conditions of metal bioavailability and stress.

5- Our studies of *O. dioica* MTs have revealed two unusual features of these proteins. First, while the number of repeats in domain repeat proteins is strongly conserved within a species, OdiMT2 shows population variability in the number of repeat units. The repeat unit in this MT is a supradomian that seems to be an optimized combination for metal binding. Second, whereas the repeat purity is low in most domain repeat proteins, it is high in OdiMT2, probably because multi-domain MTs seem not to be affected by misfolding or aggregation problems.

6- In summary, the results of this thesis illustrate how extensive analyses in wide phylogenetic contexts of proteins that, as metallothioneins, are organized in structural and functional domains are useful to understand the origin, dynamics and relevance of modularity in protein evolution.

ANNEX I:

The Fungus Tremella mesenterica Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization

PLoS ONE 11(2): e0148651

DOI: 10.1371/journal.pone.0148651

PLOS ONE

RESEARCH ARTICLE

The Fungus *Tremella mesenterica* Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization

Paul Iturbe-Espinoza[1☯], Selene Gil-Moreno[2☯], Weiyu Lin[1], Sara Calatayud[1], Òscar Palacios[2], Mercè Capdevila[2], Sílvia Atrian[1]*

1 Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain,
2 Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain

☯ These authors contributed equally to this work.
* satrian@ub.edu

CrossMark
click for updates

OPEN ACCESS

Citation: Iturbe-Espinoza P, Gil-Moreno S, Lin W, Calatayud S, Palacios Ò, Capdevila M, et al. (2016) The Fungus *Tremella mesenterica* Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization. PLoS ONE 11(2): e0148651. doi:10.1371/journal.pone.0148651

Editor: Yong-Sun Bahn, Yonsei University, REPUBLIC OF KOREA

Received: October 19, 2015

Accepted: January 21, 2016

Published: February 16, 2016

Copyright: © 2016 Iturbe-Espinoza et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: GenBank: KM244758.1 GenBank: AJK28606.1.

Funding: This work was supported by the Spanish Ministerio de Economía y Competitividad (MINECO), grants BIO2012-39682-C02-01 (to SA) and -02 (to MC), which are co-financed by the European Union through the FEDER program. Authors are members of the 2014SGR-423 Grup de Recerca de la Generalitat de Catalunya.

Abstract

Fungal Cu-thioneins, and among them, the paradigmatic *Neurospora crassa* metallothionein (MT) (26 residues), were once considered as the shortest MTs -the ubiquitous, versatile metal-binding proteins- among all organisms, and thus representatives of their primeval forms. Nowadays, fungal MTs of diverse lengths and sequence features are known, following the huge heterogeneity of the Kingdom of Fungi. At the opposite end of *N*. *crassa* MT, the recently reported *Cryptococcus neoformans* CnMT1 and CnMT2 (122 and 186 aa) constitute the longest reported fungal MTs, having been identified as virulence factors of this pathogen. CnMTs are high-capacity Cu-thioneins that appear to be built by tandem amplification of a basic unit, a 7-Cys segment homologous to *N*. *crassa* MT. Here, we report the *in silico*, *in vivo* and *in vitro* study of a still longer fungal MT, belonging to *Tremella mesenterica* (TmMT), a saprophytic ascomycete. The *TmMT* gene has 10 exons, and it yields a 779-bp mature transcript that encodes a 257 residue-long protein. This MT is also built by repeated fragments, but of variable number of Cys: six units of the 7-Cys building blocks-$CXCX_3CSCPPGXCXCAXCP$-, two fragments of six Cys, plus three Cys at the N-terminus. TmMT metal binding abilities have been analyzed through the spectrophotometric and spectrometric characterization of its recombinant Zn-, Cd- and Cu-complexes. Results allow it to be unambiguous classified as a Cu-thionein, also of extraordinary coordinating capacity. According to this feature, when the TmMT cDNA is expressed in MT-devoid yeast cells, it is capable of restoring a high Cu tolerance level. Since it is not obvious that *T*. *mesenterica* shares the same physiological needs for a high capacity Cu-binding protein with *C*. *neoformans*, the existence of this peculiar MT might be better explained on the basis of a possible role in Cu-handling for the Cu-enzymes responsible in lignin degradation pathways.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Metallothioneins (MTs) constitute a heterogeneous superfamily of ubiquitously occurring, low molecular weight, cysteine rich proteins that natively coordinate divalent (Zn^{2+}, Cd^{2+}) or monovalent (Cu^{+}) metal ions through metal-thiolate bonds, which imposes a definite polypeptide folding (see [1,2,3] for recent MT reviews). No single biological role has been assigned to these peptides, but, instead, several functions have been proposed [4], ranging from physiological metal handling to toxic metal protection. MTs are highly polymorphic proteins exhibiting a low degree of sequence similarity, so that in fact they can be considered different homology groups along the Tree of Life [5]. It is precisely this sequence heterogeneity what explains why different classification criteria have emerged during the progressive discovery and characterization of new MTs. Binz and Kägi, in the late nineties [6], proposed a taxonomy-based MT classification, in which essentially each MT group (so-called *family*) included the MTs of a taxonomic group of organisms, this ensuring their homology. In contrast to this *sequence-based* criterion, we proposed a MTs *function-based* classification that assigns a Zn-/Cd (*i.e.* divalent metal ion) or Cu-thionein (*i.e.* monovalent metal ion) character to each peptide, according to its metal-binding preference [7], a classification that was latter modulated as a step gradation between these two extreme metal preferences [8,9]. Thereafter, we showed that a unique, energetically optimized complex results when a MT polypeptide folds about its cognate metal ions, while with non-cognate metal ions, it renders a mixture of species, none of them specially favored [10].

The Kingdom of Fungi is an extremely large and heterogeneous group of organisms, comprising between 1.5 million to 5 million species. This same heterogeneity applies to their MTs, because as for no other taxon fungal MTs are distributed in 6 different families (from Family #8 to #13) in the Binz and Kägi classification [6] (Table 1). At the time of such proposal, the fungal non-yeast MTs (Family #8) were restricted to those of the ascomycete *Neurospora crassa* [11] and the basidiomycete *Agaricus bisporus* (the edible champignon or white mushroom) [12]. Both were the shortest MTs reported (26 amino acids, encompassing 7 Cys residues), and shared with Cup1 (the yeast *Saccharomyces cerevisiae* MT [13,14]) its definite Cu-thionein character. From these results, the idea that fungal non-yeast MTs were representative of short, archetypical MTs, which would have evolved to yield the higher invertebrate and vertebrate forms by domain duplication and specialization towards divalent metal ion binding, gained full support [15].

This picture was completely turned over when the two MTs of the human pathogenic fungus *Cryptococcus neoformans* (CnMT1 and CnMT2) were recently reported as infection, virulence, and pathogenicity factors [16]. CnMT1 and CnMT2 counteract the Cu(I) ions diffused by macrophages in the infected tissues through the extraordinary Cu-binding capacity derived from their unusual length: 122 (CnMT1) and 183 amino acids (CnMT2) [17]. These two MTs revealed an unexpected modular structure, being respectively constituted by three and five

Table 1. Families of the Binz & Kägi MT classification that include the fungal MTs.

Family	Group	Example	Sequence	UniProtKB
8	**Fungal 1**	*N. crassa* MT	MGDCGCSGASSCNCGSGCSCSNCGSK	P02807
9	**Fungal 2**	*C. glabrata* MT1	MANDCKCPNGCSCPNCANGGCQCGDKCECKKQSCHGCGEQCKCGSHGSSCHGSCGCGDKCECK	P15113
10	**Fungal 3**	*C. glabrata* MT2	MPEQVNCQYDCHCSNCACENTCNCCAKPACACTNSASNECSCQTCKCQTCKC	P15114
11	**Fungal 4**	*Y. lipolytica* MT3	MEFTTAMLGASLISTTSTQSKHNLVNNCCCSSSTSESSMPASCACTKCGCKTCKC	Q9HFD0
12	**Fungal 5**	*S. cerevisiae* Cup1	MFSELINFQNEGHECQCQCGSCKNNEQCQKSCSCPTGCNSDDKCPCGNKSEETKKSCCSGK	P0CX80
13	**Fungal 6**	*S. cerevisiae* Crs5	MTVKICDCEGECCKDSCHCGSTCLPSCSGGEKCKCDHSTGSPQCKSCGEKCKCETTCTCEKSKCNCEKC	P41902

doi:10.1371/journal.pone.0148651.t001

A)

```
C. neoformans   -MACNCPP-QKNTACCSTSEAQDKCTCQKGNCECKACPNSTKTSESGGKASTCNCGGSGE 58
T. mesenterica  GEECSCPPGQCSCANCPAKEKKDACSCSEG-CSCPPGQCACANCPHKDKAKGCSCG---E 56
                   *.*** * . * *.:.* :* *:*.:* *.* .   :  ..   .**. *.**   *
C. neoformans   ACTCPPGQCACDKCPKKAKSVSTCGCGGSGAACSCPPGKCACDNCPKQAQEKVSSCACSG 118
T. mesenterica  SCSCPPGECKCANCPKKSEPAKACACG---DECSCPPGQCGCADCP----GKTSS----- 104
                :*:****:* * :****::...:*.**     ******:*.* :**     *.**
C. neoformans   SGAA 122
T. mesenterica  ----
```

B)

```
AGATGTGGAGAAGAATGTTCATGCCCCCCCGGACAATGTTCTTGTGCCAATTGTCCCGCTAAAGAGAAGAAAGATGCTTGCAGGTATGTCTTTC
GCCCTCTTCTTGTCACATCTCCCGTCTTGAATGACGGGACGCGTAGTATCGTGACACTCGTATTTGTCATCAGAGGAACGACGCCCGACATCTT
GCATTTGATTGACAGGCGGACGCGAGATATGCGGTGTGAGAAATGCTGCATTTAGCAGGGGTTCAACACTGCAGTATGAGAAGAGGGCTGATAT
CAAAGCTGCTCTGAAGGGTGTTCTTGCCCTCCTGGTCAATGCGCATGCGCCAACTGTCCGCACAAGGACGAAGCGAAGGGGTGCAGGTGAGCTA
TGTTGTCCTTGACCTCATTCCTAACCCTGCTTGTCAAGTTGTCACGTTCTGATGACTGATAGTGTAACCTGATTTTGAGACAAACGCTGATGAT
AGCTGCGGCGAGTCATGTTCATGCCCACCTGGAGAATGTAAATGTGCGAACTGCCCCAAGAAAACCGAACCTGCCAAAGCCTGCGCGTGAGTGG
CTCCTTCCCTCAGTCTGCAACCTTTCCCCTCTCGACTACGCTTGAAGACATGACGTTCTCTCCCTACTATAGAAGTATAGAGCGGGGAGAGGGA
AGTTAGAAAAGATGTTGACCAATAGTTGCGGTGACGAATGTTCTTGTCCTCCTGGACAGTGCGGATGTGCGGATTGTCCAGGGAAGACTTCATC
TTGATTGGCACCTTGAGCACGGATGAACCTCAGCAAAATTCCCGACACCACAATGACCCCAA
```

Fig 1. Partial *T. mesenterica* MT protein and cDNA sequences retrieved from data banks. (A) Clustal W2 alignment of the *C. neoformans* CnMT2 protein sequence and that retrieved from the *T. mesenterica* NCBI EIW70699 ORF annotation. (B) cDNA corresponding to the sequence coding for the protein shown in (A) localized in the TREMEscaffold_3 (access code JH711530.1) of the *T. mesenterica* genome. The putative exons are shown in yellow and the conventional splicing donor/acceptor sites in gray.

doi:10.1371/journal.pone.0148651.g001

7-Cys regions separated by spacer stretches. Therefore, their origin was hypothesized to be the result of ancient tandem repetitions of a primeval fungal MT unit comprising seven Cys residues, with the same Cys pattern of the *Neurospora* and *Agaricus* MTs (X_2-[CXC]-X_5-[CXC]-X_3-[CXC]-X_2-C-X_3) ([11,12], respectively) [16]. The analysis of the Cu-binding features of CnMT1 and CnMT2 further supported this hypothesis, because the homometallic Cu-CnMT species folded *in vivo* at high Cu concentrations could be readily explained by the respective three- and five-fold presence of basic Cu_5-(7-Cys) clusters [17]. The other known "long MTs" are the five MT isoforms present in the ciliate *Tetrahymena* species, thus another unicellular eukaryote, with lengths ranging between 96 and 181 amino acids, most of them also with modular primary structures [18].

During the study of *C. neoformans CnMT1* and *CnMT2*, there was evidence that both genes had been wrongly annotated [17]. Miss annotation of MTs or MT-like's often occurs as the result of automatic analysis of raw genome sequences, due to the common short exon length of the *MT* genes, and the Cys repetition in MT proteins. In this scenario, we aimed at examining the ensemble of available fungus genomes, transcript and EST databases, in order to identify other putative long-length MTs, and to analyze their putative modular structure and metal binding abilities. Among all the retrieved hits, a partial cDNA from *Tremella mesenterica*, devoid of translation starting codon, and annotated as coding a hypothetical protein (Fig 1) attracted our attention. This sequence showed a Cys pattern characteristic of MT polypeptides, it aligned well with the *C. neoformans* MTs, and it belonged to a fungus genus also member of the Tremellales order [19]. *T. mesenterica* is typically considered as a saprophytic fungus, but it has also been reported as parasitic to other fungi. Hence, starting from the partial and

unassigned sequence, we were able to define, both bioinformatically and experimentally, the whole polypeptide coding sequence, showing that it had all the requirements to be considered as an MT protein, which in fact turned out to be the longest MT ever reported (257 amino acids), and it also exhibited a modular structure. The *T. mesenterica* MT (TmMT) was then recombinantly synthesized as Zn-, Cd- and Cu-complexes, which were spectroscopically and spectrometrically characterized. The results showed that TmMT was also a Cu-thionein, with an extremely high Cu coordination capacity. This is in concordance with the high tolerance exhibited by this fungus to Cu. Our results open the possibility to ascertain the biological significance that this protein may exhibit in the *T. mesenterica* physiology, but most significantly, they show how tandem amplification of basic MT units seem to be a common trait in the evolution of several MTs in the Tremellales fungi.

Materials and Methods

Bioinformatic methods for identification of MT genes in the *Tremella mesenterica* genome

DNA and protein sequences were retrieved, analyzed and compared using the online versions of BLAST (at www.ncbi.nim.nih.gov) and Clustal Omega2 (W2) (at http://www.ebi.ac.uk). Screened databases were NCBI (National Center for Biotechnology Information, at www.ncbi.nim.nih.gov) the JCI (Joint Genome Institute of the USA Department of Energy (at genomeportal.jgi.doe.gov) [20].

Tremella mesenterica cultures

Tremella mesenterica strain was obtained from PYCC® (Portuguese Yeast Culture Collection, ref # PYCC 5472). Liquid cultures were performed at 25°C in YM (Yeast Mold) medium (yeast extract 3 g/L, malt extract 3 g/L, peptone 5 g/L, and glucose 10 g/L) [21], supplemented with copper salt when necessary ($CuSO_4$, at concentrations ranging from 0.1 mM to 5 mM, as indicated in each experiment). For plate cultures, YPD (2% glucose, 1% yeast extract, 2% peptone and 2% agar) was used and supplemented with $CuSO_4$ when necessary.

Tremella mesenterica RNA isolation and retrotranscription

Total RNA was extracted from *T. mesenterica* 10-mL YM cultures, following an adaptation of yeast RNA isolation rationale [22]. All the used material was sterilized and treated to be RNAase free. 1-mL aliquots of grown cultures were centrifuged in Eppendorf tubes for 3 min at 2500 rpm, frozen in liquid N_2 and re-suspended in 0.5 mL of LETS buffer (0.1 mM LiCl, 0.01 mM EDTA, 0.01 mM Tris, 2% SDS, pH 4.3). Further 0.5 mL of Tris-HCl saturated phenol (pH 4.3) and 0.5 mL of glass beads (425–600 μm diameter) were added to each Eppendorf and cells were disrupted in a TissueLyser® (Qiagen) by two 30-s series of 30 pulse/s. A mixture of phenol:chlorophorm:isoamyl alcohol (25:24:1) was added 1:1 v:v to the supernatant of a 15-min centrifugation at 12000 rpm for a first extraction. A second extraction was performed from the previous supernatant with 24:1 chlorophorm:isoamyl alcohol, and samples were precipitated with 5 M LiCl and kept at -80°C for at least 3 h. After centrifugation at 12000 rpm for 15 min, the pellet was washed with 200 μL of 70% ethanol per Eppendorf, and re-precipitated. Finally, the RNA was re-suspended in 30 μL of milliQ water per Eppendorf and its concentration assessed by A_{260} in a NanoQuant® (Tecan) equipment, and by agarose gel electrophoresis (1% agarose in TAE (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) buffer. The isolated RNA was treated with RQ1® RNAse-Free DNAse (Promega) to avoid DNA contamination (digestion with 1 u/μL of the enzyme at 37°C for 30 min) and stored at -80°C until needed. Total RNA

was retrotranscribed using the Transcriptor First Strand cDNA Synthesis® kit (Roche), which includes oligo dT and random hexamer primers. 1 ng of total RNA was denatured for 5 min at 65°C, and then RT buffer (8 mM $MgCl_2$), RNAse inhibitor, RT enzyme and the oligonucleotide mix were added in a 20-μL final volume. The sample was incubated for 10 min at 25°C, and retrotranscription was allowed for 30 min at 55°C and finally stopped by a 5-min incubation at 85°C. The obtained cDNA was quantified by A_{260} measurement in a NanoQuant® (Tecan).

RACE amplification of the *Tremella mesenterica* mRNA

To obtain the full *T. mesenterica* MT cDNA 5' end sequence, the RACE (Rapid Amplification of cDNA Ends) strategy was applied, through the 5' RACE v12® kit from Roche. To this end, three antisense primers were designed from the already known TmMT sequence: R1 (5' TCAA GATGAAGTCTTCCCTG 3'), R2 (5' GACAATCCGCACATCCGCAC 3') and R3 (5' ACATT CGTCACCGCAAGCGC 3'). RACE reactions were performed following the supplier instructions, starting from 2 μg of the total *T. mesenterica* RNA preparation (147.5 μg RNA/μL) obtained from fresh fungus cultures in YM medium. The products of the nested PCR of the RACE steps were followed by agarose gel electrophoresis, and the final product was directly sequenced.

Synthesis of the *Tremella mesenterica* cDNA and cloning in the *E. coli* expression vector

cDNA cloning procedures were performed essentially as described previously in detail for the *C. neoformans* [17] and *Amphioxus* [23] MTs. Hence, the complete TmMT cDNA was amplified by PCR from the total *T. mesenterica* cDNA preparation obtained as previously described. The reaction was performed in a final volume of 25 μL, using the Expand High Fidelity PCR system® (Roche), and the specific primers: 5' 5'AAAAGGATCCATGTCTGCTCCTGTCGA AAC 3' (upstream) and 5'AAAACTCGAGGATTTGACGTTAGAGCAACC 3' (downstream), to respectively add the *BamH*I/*Xho*I sites necessary for in-frame cloning into the *E. coli* pGEX-4T-1® expression vector (GE Healthcare). Expression from this system yields GST-fusion proteins, from which the MT portion is isolated by thrombin cleavage [24]. The 35-cycle amplification reaction was performed under the following conditions: 15 s at 94°C (denaturation), 50 s at 55°C (annealing), and 50 s at 72°C (elongation). The final product was analyzed by 2% agarose gel electrophoresis/ethidium bromide staining, and directly purified from the PCR reaction with the Illustra DNA Purification Kit® (GE Healthcare). The amplified DNA and the pGEX-4T-1 vector were digested with *BamH*I and *Xho*I (Fast Digest®, Thermo Scientific), and ligated using the DNA Ligation Kit 2.1® (Takara Bio Inc.). Finally, the recombinant plasmids were transformed into *E. coli* Mach I strain for sequencing, which was attained using the Big Dye Terminator 3.1 Cycle Sequencing Kit® (Applied Biosystems) and the pGEX-4T-1 5' and 3' primers, in a automatic sequencer (ABIPRISM 310, Applied Biosystems) of the CCiTUB (Genomics Services of the University of Barcelona). The correct recombinant plasmids (pGEX-TmMT) were then transformed into the *E. coli* BL21 protease deficient strain (GE Healthcare) for protein synthesis.

Cloning of the *Tremella mesenterica* cDNA in yeast expression vector and complementation assays

The *Saccharomyces cerevisiae* 51.2cΔc5 strain (*MAT*a, *trp*1-1, *ura*3-52, *ade-*, *his-*, CAN^R, *gal*1, *leu2-3*, *112 met13*, *cup1*Δ::*URA*3 *crs5*Δ::*LEU2*), derived from VC-sp6 [25] was used for copper tolerance assays as described in detail previously for the CnMT1-derived peptides [26]. The TmMT cDNA was inserted into the *BamH*I/*Xho*I sites of the yeast vector p424-GPD, which

contains TRP1 as a selection marker, the constitutive glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter for gene expression, and the cytochrome-c-oxidase (CYC1) transcriptional terminator [27]. The recombinant TmMT-p424 plasmid was transformed into 51.2cΔc5 cells using the LiAc/SS-DNA/PEG procedure [28], and positive transformants were selected by their capacity to grow in synthetic complete medium (SC) devoid of Trp, Leu, and Ura. For comparative purposes, yeast transformants with CUP1-p424 and mMT1-p424, encoding for the yeast CUP1 MT and the mouse MT1 isoform respectively, and with the void p424 plasmid, were also assayed. For copper tolerance tests, the transformants were initially grown in selective SC-Trp-Ura medium at 30°C and 220 rpm until saturation, and then a cell suspension of OD_{600} = 0.01 was used to inoculate 3-mL fresh cultures supplemented with $CuSO_4$ at 0, 2, 4, 7, 10, 15, 20, and 30 μM final concentrations. After 18 h of growth, the final OD_{600} was measured and plotted as a percentage of the OD_{600} reached by the culture grown without Cu. Three replicates were run for each Cu concentration, and for each transformant.

Preparation of recombinant and *in vitro*-constituted metal-TmMT complexes

5-L LB (Luria-Bertani) cultures of pGEX-TmMT transformed BL21 cells were induced with 100 μM (final concentration) of IPTG (isopropyl β-D-thiogalactopyranoside), and after 30-min growth, they were supplemented with 300 μM $ZnCl_2$, 300 μM $CdCl_2$ or 500 μM $CuSO_4$ (final concentrations) to respectively allow the synthesis of the Zn-, Cd- or Cu-TmMT complexes, and further grown for additional 2.5 h. Cu-supplemented cultures were performed at two aeration conditions: regular (*i.e.* 1-L of LB media in a 2-L Erlenmeyer flask, at 250 rpm) or low (1.5-L of LB media in a 2-L Erlenmeyer flask, at 150 rpm), since available oxygen determines the intracellular copper levels in the host cells, as described in [29]. To prevent oxidation of the metal–TmMT complexes, argon was bubbled in all the steps of the purification protocol. Cells recovered from the bacterial cultures by centrifugation were resuspended in ice-cold PBS (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4)-0.5% v/v β-mercaptoethanol and disrupted by sonication in a Sonifier® ultrasonic cell disruptor (8 min, at 0.6 pulse/s). The total protein extract was centrifuged at 12,000 x*g*, 40 min, and the supernatant was incubated with Glutathione-Sepharose 4B® (GE Healthcare) beads at gentle agitation for 1 h, room temperature. After three washes in cold PBS, the matrix-bound GST-MT protein was split by thrombin digestion (10 u per mg of fusion protein, overnight at 17°C). The solution containing the metal-TmMT complexes, which had consequently been released from the matrix, was concentrated using Ultracel® YM-3 (Millipore) filters, and finally fractionated through a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.0, and run at 0.8 mL min^{-1}. The protein-containing fractions, identified by their absorbance at 254 and 280 nm, were later analyzed in 15% SDS-PAGE gels stained with Coomassie Blue, and they were pooled and stored at -80°C until further use. Due to the pGEX vector cloning specificities, the recombinant TmMT exhibited two additional N-term residues (Gly-Ser), but they have been shown to have no effect on the MT metal-binding features [30]. Further details about the synthesis and purification procedures can be found in previous publications [17,23,26].

The so-called "*in vitro* complexes" were those prepared *via* metal replacement by adding the corresponding metal ions (Cd(II) or Cu(I)) to the recombinant Zn-TmMT samples. These reactions were performed at pH 7.0 using $CdCl_2$ or $[Cu(CH_3CN)_4]ClO_4$ solutions, respectively, as described earlier in detail for mammalian MTs [24,31]. During all the experiments, strict oxygen-free conditions were maintained by saturating the solutions with argon. All the *in vitro*-obtained metal-MT samples were analyzed following the same rationale as for the recombinant samples.

Characterization of the metal-TmMT complexes

ICP-AES (inductively coupled plasma atomic emission spectroscopy) analysis of the purified metal-TmMT complexes was essentially performed as previously described for other MTs [17,23,26]. S, Zn, Cd and Cu contents were measured in a Polyscan 61E (Thermo Jarrell Ash) spectrometer, reading S at 182.040 nm, Zn at 213.856 nm, Cd at 228.802 and Cu at 324.803 nm. Samples were treated as in [32], and were also alternatively incubated in 1 M HCl at 65°C for 15 min (acid ICP conditions) to eliminate any labile sulfide ions, as described in [33]. Protein concentration was calculated from the acid ICP-AES sulfur content, assuming that the measured S atoms were contributed by the MT peptide. CD spectra were recorded in a Jasco spectropolarimeter (model J-715) interfaced to a computer (J700 software) at a constant temperature of 25°C, maintained by a Peltier PTC-351S apparatus. Electronic absorption was measured in a HP-8453 Diode array UV–visible spectrophotometer. All spectra were recorded using 1-cm capped quartz cuvettes, corrected for the dilution effects and processed using the GRAMS 32 program.

ESI-MS (electrospray ionization mass spectrometry) analyses of the metal-TmMT complexes

Electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) conditions for the analysis of the metal-TmMT complexes were adapted from those reported in detail in [17] and [26]. The equipment used was a Micro TOF-Q (Bruker) interfaced with a Series 1200 HPLC Agilent pump and an autosampler, all of them controlled by the Compass Software. The mass spectrophotometer was calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies). The Zn- and Cd-TmMT samples were analyzed under the following conditions: 20 μL of protein solution injected through PEEK (polyether heteroketone) tubing (1.5 m x 0.18 mm i.d.) at 40 μL min^{-1}; capillary counter-electrode voltage 5 kV; desolvation temperature 90–110°C; dry gas 6 L min^{-1}; spectra collection range 800–2500 m/z. The carrier buffer was a 5:95 mixture of acetonitrile:ammonium acetate (15 mM, pH 7.0). The Cu-TmMT samples were analyzed by injecting 20 μL of protein solution at 40 μL min^{-1}; capillary counter-electrode voltage 3.5 kV; lens counter-electrode voltage 4 kV; dry temperature 80°C; dry gas 6 L min^{-1}. Here, the carrier was a 10:90 mixture of acetonitrile:ammonium acetate, 15 mM, pH 7.0. Acidic-MS conditions, which cause the release of the divalent metal ions from the MT complexes, but keeps the Cu(I) ions, were used to generate apo-TmMT forms and to analyze the Cu-containing samples. To this end, 20 μL of the preparation were injected under the same conditions described previously, but using a 5:95 mixture of acetonitrile:formic acid, pH 2.4, as liquid carrier. For all the ESI-MS results, the error associated with the mass measurements was always inferior to 0.1%. Masses for the holo-species were calculated as described in [34].

Results and Discussion

Identification and analysis of MT coding sequences in *Tremella mesenterica* databases

The discovery of the *C. neoformans* long MTs [16] and their relevance in pathogenesis, prompted us to analyze several fungal genomes for the presence of unusually lengthy MTs. Among all the retrieved BLAST matches using *C. neoformans* CnMT1 and CnMT2 as queries in the NCBI databases, the sequence derived from the *T. mesenterica* NCBI EIW70699 ORF (Fig 1A) was the clearest MT-like candidate, as 25% of its 104 residues were cysteine and it contained no aromatic amino acids. Nevertheless, since it constituted an incomplete protein sequence in view of its lack of a N-terminal initial methionine, the *T. mesenterica* genome was

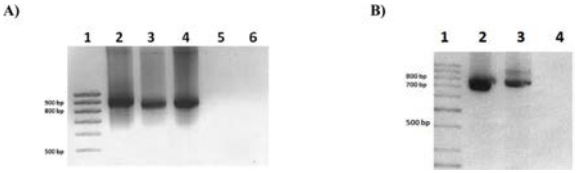


Fig 2. (A) PCR products of the TmMT 5' RACE reaction. Lanes are: (1) DNA size markers, (2) first RACE amplification reaction, (3) second RACE amplification reaction 1:20 dilution, (4) second RACE amplification reaction, (5 and 6) negative control RACE reactions (no template DNA included). **(B) PCR products of the RT PCR and 5' RACE TmMT amplification reactions.** Lanes are: (1) DNA size markers, (2) the PCR product of the RT-PCR reaction using TmMT specific primers, (3) final product of the RACE amplification, (4) negative control PCR reaction (no template DNA included).

doi:10.1371/journal.pone.0148651.g002

searched in the JCI BioProject (PRJNA32829), and its encoding sequence was localized in the TREMEscaffold_3 (access code JH711530.1). This 750 bp-long sequence was located between positions 1234899 and 1235652 in this scaffold and it included unambiguous exon/intron limits, according to the canonical splicing signals (GT/AG). In concordance with the protein sequence, this gene fragment did not show any ATG starting codon in frame, but it exhibited a stop codon (TGA) after the last serine triplet. All these observations were indicative of a 5'-truncated gene and cDNA in the *T. mesenterica* genome annotation, so it was necessary to further investigate the *TmMT* cDNA start.

Determination of the full-length *T. mesenterica* MT cDNA and gene

The full TmMT transcript 5' sequence was determined by RACE. Sequencing of the final RACE product (Fig 2) allowed to unambiguously identify the *TmMT* ATG start codon nearly 1200 bp upstream from the start of the truncated sequence initially retrieved (Fig 1B) from the *T. mesenterica* genome, and also revealed the presence of a 69- bp 5' UTR region in the corresponding cDNA (Fig 3).

Analysis of the *T. mesenterica* MT gene, cDNA and protein sequences

The full-length *T. mesenterica* MT protein and cDNA sequences (Fig 3) were submitted to GenBank, and are available under the accession numbers: AJK28606.1 and KM244758.1, respectively. The TmMT cDNA coding portion (from the ATG to the STOP codons) encompasses 774 bp, and encodes a 257 aa-long protein with a molecular weight of 25.37 kDa, including 57 Cys (23% of the polypeptide) and 1 His (Fig 4A). A RT-PCR reaction on total *T. mesenterica* cDNA (retrotranscribed from total RNA preparation) corroborated that it was not a cloning or amplification artifact. The result of this reaction (Fig 2B) yielded a unique band of the expected size (approximately 834 bp), the sequence of which matched those of the corresponding regions in the genomic DNA database (Fig 3A). Therefore, it was established that *TmMT* was a *real* coding gene, with a sequence split into 10 exons interrupted by 9 introns (Fig 3B). The cDNA 5' UTR had 69 bp, and an *in silico* analysis of 1500 bp of the scaffold sequence upstream the *TmMT* translation start site revealed the presence of a putative TATA box at -38 bp, an also a putative MRE at -1457 bp. It is worth noting that when the retro-PCR was carried out using RNA preparations obtained from cells grown in Cu-enriched cultures, no difference was observed in the intensity of the TmMT product, which indirectly suggests a rather constitutive gene expression pattern (data not shown).

The isolated TmMT cDNA presented ten site variations regarding the sequence in the JH11530.1 scaffold, five of which represented a missense mutation, involving the following substitutions: N22S, S35P, T48I, K201E and S207T, neither of them altering the expected metal

A)

```
5′ GAAAAAGCACCTCTCTAGGAGTGCA(-1475)...CAAAGGAAAGTCAAAAATGTATCAAAGGATCTTTTCTTTGACGCCCTAAGTGCGCC
AAATCGACGCAAACCGCAGAATTGGCTCAACCACTCTCTTCTACTTTAAATTCCTTCTCTTCTTTCTTCTTATCCTCATCTCGACCCAGTCTCA
TCTCAGTCTAGTTCAGTCCTATCCAGTTCAGTTCAGTTTAGTCATCTAAGAATCAACATGTCTGCGTGAGTATCTTCTCAAGCTCTCGGGTAGA
                                                  M  S  A
TAAATGAGCTGACCATTAGTCCTGTCGAAACCAAGGAGAAATCTTGTGGGTGCCAACCTGCTCCTGCTGTGCAAAGTTGCAAGTGAGTCGTTCA
                    P  V  E  T  K  E  K  S  C  G  C  Q  P  A  P  A  V  Q  S  C  N
TTCTCCTTCATCTCTAGTCTCTACACTCCTCATCTCATAGGCTCTTGTATGGTCAACTTCCATGTCAGACTTTCAGAAGGAAACGACCAGAACG
TCCGAGAAAGCGGGGTTGGAATATTCTCATAGTACCGACCCTTAAGACCTGTGTTCGAAGACGAAGGCTTATCTTTAGTTGCTCGAACGAAGGA
                                                                          C  S  N  E  G
AATTGTACCTGTGCTCCGGGAAAATGCGCTTGCTCTTCGTGTTCCTCGGACTCTATTAAGAAAACAGGCAAGTGCGGGTGAGTCATACTTGACT
N  C  T  C  A  P  G  K  C  A  C  S  S  C  S  S  D  S  I  K  K  T  G  K  C  G
TCTCAATATGATCCAAAATCACTTTACTTTTCTTGACTACCCCGTGCCCATGGGTGTCTTCCTCGTCAAGATCTGACCCTGTGATCTGGCACAA
AGAATGGGATGATTTGACAGAGAATGAGGGCTTATACATAGAGGGTCTGAGGGATGTACTTGCGAAGCCGGTAAATGCGATTGTGCTTCTTGTC
                                                  G  S  E  G  C  T  C  E  A  G  K  C  D  C  A  S  C  P
CTGGTTCTAGCGGGCAAGTAAAGGCATGCACGTGAGTCTTCTCTTGTTTCCTCTCTCCATATTGATGACCTTTCCTGCATCCCGCTCGGACACT
  G  S  S  G  Q  V  K  A  C  T
CCCTATCGGTCGTGCTAAATTGAGGTAACCAGCGGTATGAAAAAAAAAGTATATTTATATTTGGGCTCAGATTGGTCGCAGGAGAGTCAAGTAAG
GAAGATCGACCATGACAACTGATCTCCAGTTGCGGCACCTCTTGCTCTTGTCCCCCTGGCGAATGCACTTGTGCCGGATGTCCAAACAACAAAG
                                  C  G  T  S  C  S  C  P  P  G  E  C  T  C  A  G  C  P  N  N  K  G
GCAAAGAAAAGGCAAAGGACGAAAAAGCCGGGGAATGCAGGTGAGCAGTTTCGTTCATACCTAGGTAAATCTGACGGGTAGTTGCGGTCCATCC
  K  E  K  A  K  D  E  K  A  G  E  C  S                                    C  G  P  S
TGTTCTTGTCCGCCGGGAGAATGTTCCTGTGCAGGTTGCTCTAACGTCAAATCCACAGGGAAAGAAAAGGCTCCTGCTAAGGCATGCGAGTGAG
C  S  C  P  P  G  E  C  S  C  A  G  C  S  N  V  K  S  T  G  K  E  K  A  P  A  K  A  C  E
TGACCAAATTCTCCTCAGTTGTACCATCCCCCATCTCAGAACGATACGCCCACCTCTGTCAGGTGAACATGAAAGGGAGAGTGGAGTTGCCGCT
CACAGTGGGGAACAAATTGAATAGAGGCTGACGTCAGATGTGGAGAAGAATGTTCATGCCCCCCCGGACAATGTTCTTGTGCCAATTGTCCCGC
                                      C  G  E  E  C  S  C  P  P  G  Q  C  S  C  A  N  C  P  A
TAAAGAGAAGAAAGATGCTTGCAGGTATGTCTTTCGCCCTCTTCTTGTCACATCTCCCGTCTTGAATGACGGGACGCGTAGTATCGTGACACTC
 K  E  K  K  D  A  C  S
GTATTTGTCATCAGAGGAACGACGCCCGACATCTTGCATTTGATTGACAGGCGGACGCGAGATATGCGGTGTGAGAAATGCTGCATTTAGCAGG
GGTTCAACACTGCAGTATGAGAAGAGGGCTGATATCAAAGCTGCTCTGAAGGGTGTTCTTGCCCTCCTGGTCAATGCGCATGCGCCAACTGTCC
                                          C  S  E  G  C  S  C  P  P  G  Q  C  A  C  A  N  C  P
GCACAAGGACGAAGCGAAGGGGTGCAGGTGAGCTATGTTGTCCTTGACCTCATTCCTAACCCTGCTTGTCAAGTTGTCACGTTCTGATGACTGA
 H  K  D  E  A  K  G  C  S
TAGTGTAACCTGATTTTGAGACAAACGCTGATGATAGCTGCGGCGAGTCATGTTCATGCCCACCTGGAGAATGTAAATGTGCGAACTGCCCCAA
                                      C  G  E  S  C  S  C  P  P  G  E  C  K  C  A  N  C  P  K
GAAAACCGAACCTGCCAAAGCCTGCGCGTGAGTGGCTCCTTCCCTCAGTCTGCAACCTTTCCCCTCTCGACTACGCTTGAAGACATGACGTTCT
 K  T  E  P  A  K  A  C  A
CTCCCTACTATAGAAGTATAGAGCGGGGAGAGGGAAGTTAGAAAAGATGTTGACCAATAGTTGCGGTGACGAATGTTCTTGTCCTCCTGGACAG
                                                  C  G  D  E  C  S  C  P  P  G  Q
TGCGGATGTGCGGATTGTCCAGGGAAGACTTCATCTTGATTGGCACCTTGAGCACGGATGA 3′
C  G  C  A  D  C  P  G  K  T  S  S  *
```

B)

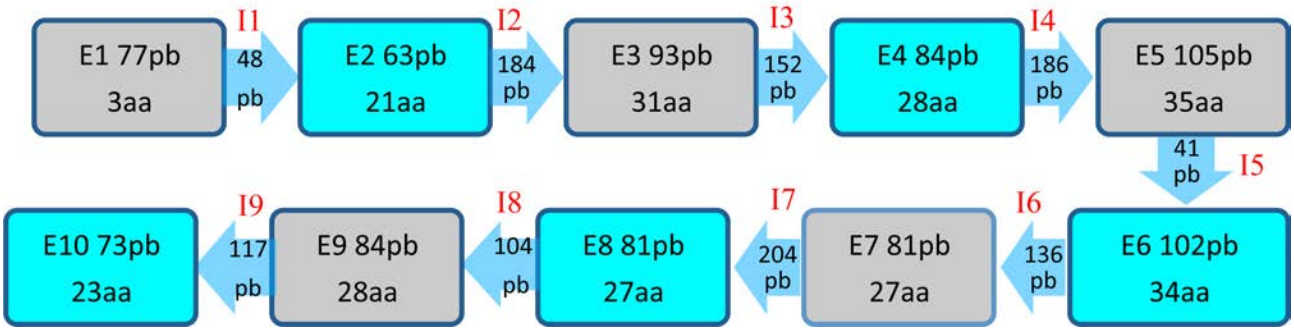


Fig 3. (A) Complete TmMT cDNA and protein sequence. The TmMT cDNA sequence obtained from the RACE reactions has been localized into the *T. mesenterica* genomic sequence (BioProject PRJNA32829, access code # JH711530.1), and the corresponding gene and protein significant elements have been localized. Exons are shown in yellow and the corresponding splicing donor/acceptor sites in gray, and the translational initiation codon is in red. A putative TATA is boxed in blue and a putative MRE in red. In the protein sequence, Cys are in red, His in green and Met in blue. **(B)** Scheme showing the *TmMT* gene structure (exon and intron sizes). This sequence has been submitted to GenBank and has the accession number TPA: BK008867.

doi:10.1371/journal.pone.0148651.g003

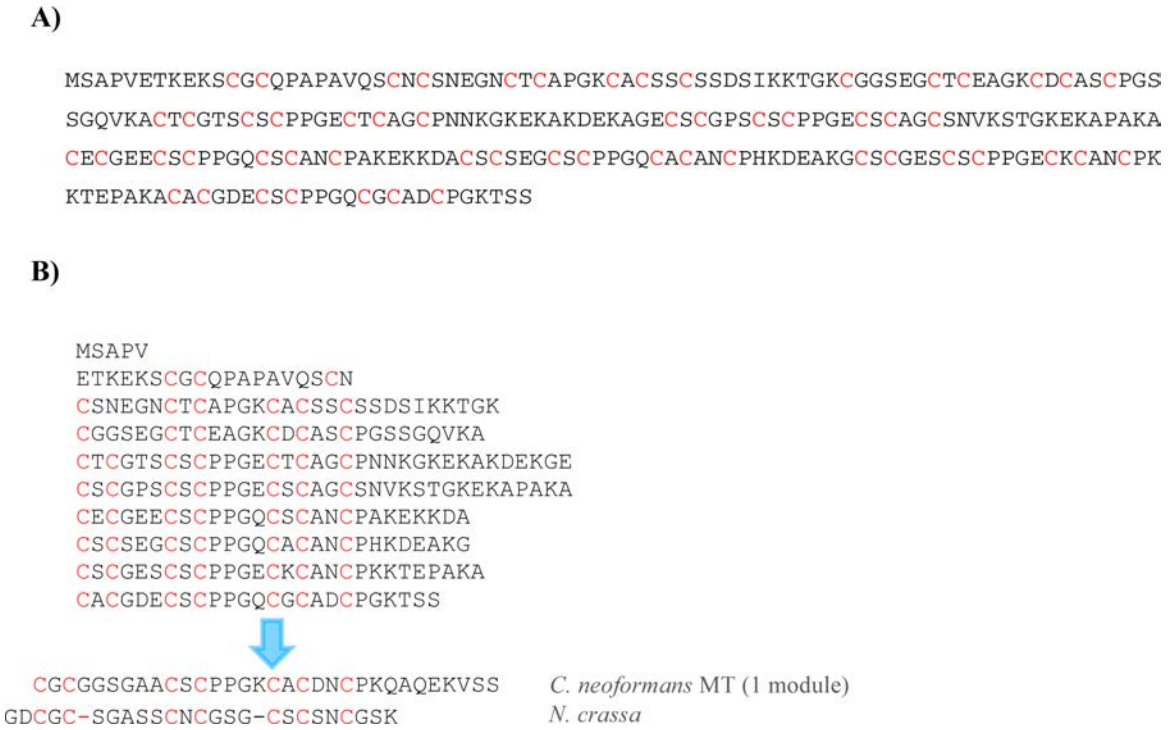


Fig 4. (A) Protein sequence of TmMT and (B) proposed modular structure of TmMT. The alignment with one 7-Cys segment of *C. neoformans* CnMT2 and *N. crassa* MT is also displayed.

doi:10.1371/journal.pone.0148651.g004

ion coordination properties of a MT polypeptide. Overall, these substitutions (representing a 1.25% of the total coding sequence) were fully compatible with a natural polymorphism between the *T. mesenterica* strain used in this work and that used for genomic sequencing an annotation in the JCI project. TmMT, with 257 amino acids and 57 Cys, is the longest MT protein ever reported, longer than the *C. neoformans* CnMT2 isoform (183 aa), the *Tetrahymena thermophila* MTT1 isoform (162 aa) and the *Paramecium sp*. PMCd1 protein (203 aa, [35]), all of them belonging to unicellular eukaryotes. Forty-eight of its 57 Cys are organized in CXC motifs, the most extended arrangement of the Cys residues in the MT sequences, but furthermore, the analysis of TmMT readily revealed an internal repetition pattern (Fig 4B), which suggests that it was originated by a modular amplification process, as shown for *Cryptococcus neoformans* [16,17] *and Tetrahymena* [18,36] MTs. In the case of TmMT, the most obvious delimitation of hypothetical internal repetitions involves six units with a -CXCX$_3$CSCPPGXCXCAXCP- sequence including 7 Cys in an arrangement alignable to that of the paradigmatic *Neurospora* MT, two units with six Cys, and a N-terminal segment with only three Cys. Other sequence features also support the homology between the TmMT building blocks and those proposed for the *C. neoformans* MTs, such as the occurrence of a proline doublet after the second CXC motif, a single proline after the last Cys of the segment, and the high similarity of the Cys interspersed amino acids, with a clear predominance of small residues (glycine, alanine), and the conservation of charged residues in key positions (lysine).

The recombinant synthesis of the TmMT protein

DNA sequencing of several pGEX-TmMT constructs confirmed that they included no nucleotide substitutions, and that the cDNA was cloned in correct frame after the GST coding

sequence. Protein extracts of small-scale (3 mL) cultures of pGEX-*TmMT*-transformed BL21 cells yielded an exclusive band a *ca*. 51 kDa, concordantly with the size of the expected fusion protein (*ca*. 26 kDa of the GST plus *ca*. 25 kDa of TmMT) (data not shown). Consequently, synthesis of TmMT was then performed in large scale cultures (5-L) supplemented with Zn (II), Cd(II) or Cu(II). Aliquots of the complexes purified from Zn- and Cd-enriched cultures (Zn-TmMT and Cd-TmMT, respectively) were first analyzed by acid (pH 2.4) ESI-MS, which revealed an almost unique peak of 25377.0 Da (Fig 5A) corresponding well with the TmMT MW of 25377.62 Da theoretically calculated for the recombinant peptide, including N-terminal Gly and Ser residues derived from the GST-fusion construct. Thus, the identity, purity and integrity of the recombinant TmMT were fully confirmed.

The divalent metal ion (Zn(II) and Cd(II)) binding abilities of TmMT

The recombinant synthesis of TmMT in cells grown under Zn-supplementation yielded a mixture of Zn-complexes of different stoichiometries, with Zn_{17}- to Zn_{21}-TmMT as major species (Table 2 and S1 Table, and Fig 5B). This matched well with the Zn-per-protein mean content calculated by ICP-AES, which gave values between 14.0 and 20.1 (Table 2). The CD spectra of the three samples preparations were practically identical, exhibiting a Gaussian band centered at *ca*. 240 nm typical of the Zn-Cys chromophores (Fig 5B).

The biosynthesis of TmMT in Cd-supplemented cultures revealed its poor ability for coordination of this metal ion, even lower than for Zn(II). The synthesis was repeated several times, and practically always it was impossible to discriminate individual peaks (Table 2 and S1 Table, and Fig 5C). The only data that can be suggested from the ESI-MS spectra was that most probably the major complexes were those including between 23 and 25 Cd(II), this coinciding with results of ICP-AES that yielded an average content of 24.2 to 24.5 Cd(II) per TmMT (Table 2). The CD spectrum showed the typical fingerprint attributable to tetrahedral $Cd(SCys)_4$ chromophores absorbing at *ca*. 250 nm, and a slight absorption in the region of 280–300 nm, probably contributed by the presence of some Cd-S^{2-} chromophores (Fig 5C). Overall, all the previous results lead to the conclusion that TmMT is far from exhibiting a binding preference for divalent (Zn(II) or Cd(II) [8]). This is reflected in the formation of a continuum of species, none of them energetically favored, with both Zn(II) and Cd(II) ions, and also accounts for the lack of reproducibility of the recombinant syntheses, which is patent in this work for the Zn-complexes, and which has been observed also before, among others, for the *Tetrahymena pyriformis* MT1 [37] and *C. neoformans* MT1 isoforms [17]. The presence of S^{2-} ligands in the Cd-TmMT complexes also corroborates this inability [33].

Cu-binding abilities of TmMT

TmMT was recombinantly synthesized in Cu(II)-supplemented cultures grown under normal and low aeration, because this modulates the amount of Cu available inside the host cells for recombinant Cu-MT complex formation [29]. Hence, when TmMT was synthesized in normally aerated cultures, it yielded a mixture of Zn,Cu-TmMT species ranging from M_8- to M_{19}-TmMT (M = Zn+Cu), with those from M_{12}- to M_{17}-TmMT as major products, as shown by ESI-MS at neutral pH (Table 2 and S2 Table, and Fig 6A). ESI-MS at pH 2.4 revealed that this continuum was formed by heterometallic complexes with Cu(I) contents showing a peculiar periodicity: multiple of 4, and multiple of 4 plus one, that is: Cu_4 and Cu_5, Cu_8 and Cu_9, and Cu_{12} and Cu_{13} (*cf*. Table 2 and S2 Table, and Fig 6A). In contrast, low aeration yielded a mixture of homometallic Cu-TmMT species -since ICP-AES measurements only detected this metal-, which ranged between Cu_{36}- and Cu_{45}-TmMT, with major species being Cu_{42}- and Cu_{41}-TmMT (Table 2 and S2 Table, Fig 6B). Acid ESI-MS of this sample revealed a similar

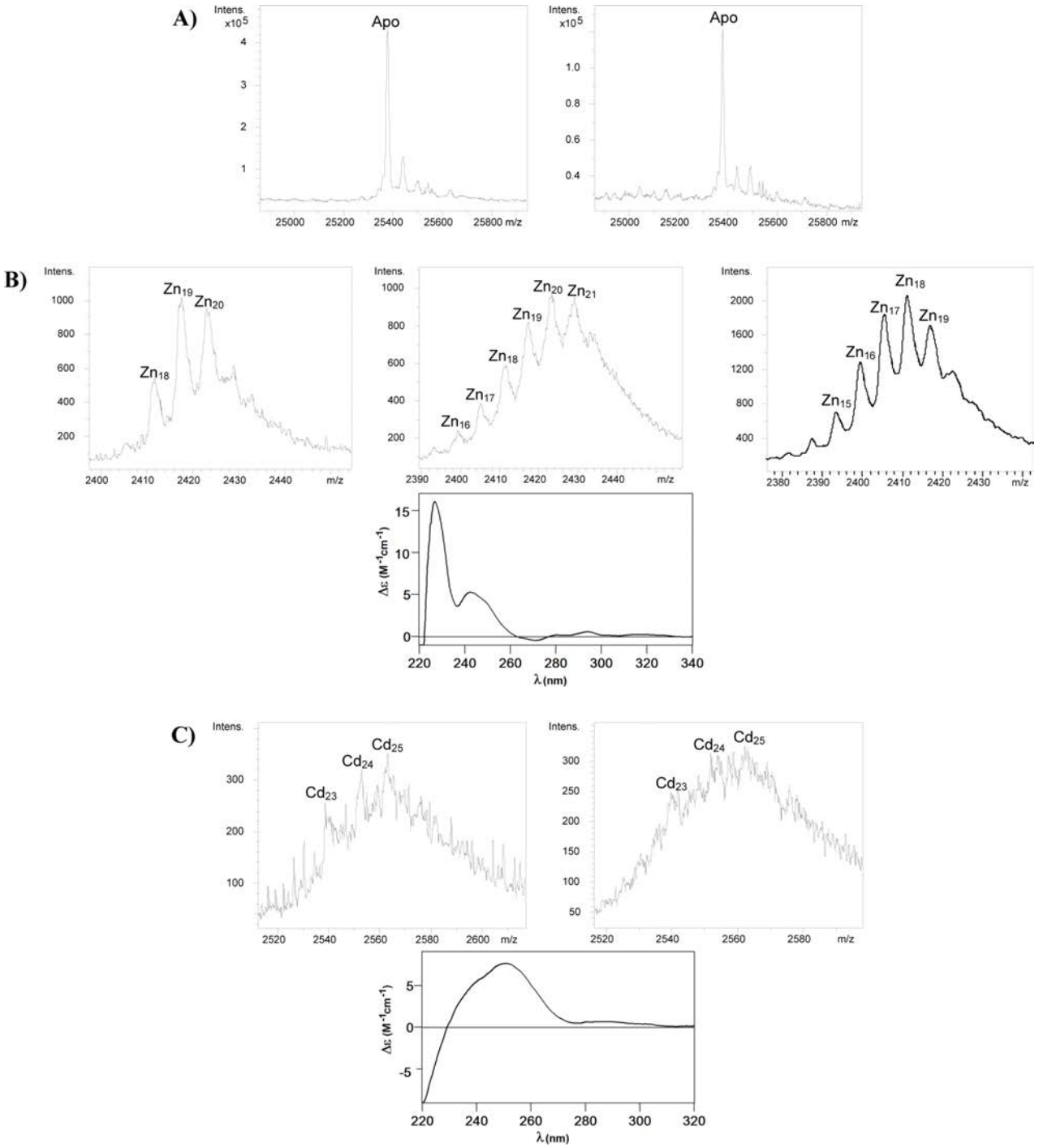


Fig 5. Spectroscopic and spectrometric analyses of the Zn- and Cd-TmMT preparations. (A) Deconvoluted ESI-MS spectra of the recombinant Zn- and Cd-TmMT samples, run at acid pH (2.4). A practically unique peak, corresponding to the expected molecular weight of the protein is observed in each spectrum. (B) ESI-MS spectra of three different Zn-TmMT syntheses (at the +11 charge state) recorded at neutral pH. The CD spectra corresponding to Zn-$TmMT_2$ is included, those for Zn-$TmMT_1$ and Zn-$TmMT_3$ being completely similar. (C) ESI-MS spectra of two different Cd-TmMT syntheses (at the +11 charge state) recorded at neutral pH. The CD spectra corresponding to the Cd-$TmMT_2$ is included; that for Cd-$TmMT_1$ being completely similar.

doi:10.1371/journal.pone.0148651.g005

distribution of species, of slightly lower stoichiometry (Cu_{29} and Cu_{38}) probably attributable to some loosely bound Cu(I) ions. Interestingly, both types of preparations rendered well defined and closely related CD spectra (Fig 6A and 6B). The CD fingerprint of the sample purified

Table 2. Recombinant synthesis yield and metal-to-protein ratios of the purified Zn-, Cd- and Cu-TmMT complexes, according to neutral ICP-AES and ESI-MS measurements.

Metal-TmMT syntheses[a]	[TmMT] (10^{-4} M)	ICP-AES[b]			Neutral ESI-MS[c]				Acid ESI-MS[c] Cu species
		Zn/MT	Cd/MT	Cu/MT	Zn species	Cd species	Zn/Cu species	Cu species	
Zn-TmMT$_1$	0.71	20.1			Zn_{21}				
					Zn_{20}				
					Zn_{19}				
					Zn_{18}				
Zn-TmMT$_2$	0.74	14.0			**Zn_{21}**				
					Zn_{20}				
					Zn_{19}				
					Zn_{18}				
					Zn_{17}				
					Zn_{16}				
Zn-TmMT$_3$	0.94	16.8			Zn_{19}				
					Zn_{18}				
					Zn_{17}				
					Zn_{16}				
					Zn_{15}				
Cd-TmMT$_1$	0.54		24.2			**Cd_{25}**			
						Cd_{24}			
						Cd_{23}			
						multiple			
Cd-TmMT$_2$	0.62		24.5			**Cd_{25}**			
						Cd_{24}			
						Cd_{23}			
						multiple			
Cu-TmMT (normal aeration)	0.55	6.6		5.8			**M_{15}**		Cu_{12}
							M_{19}-M_8		Cu_{10}
									Cu_9
									Cu_8
									Cu_6
									Cu_5
									Cu_4
									Apo
Cu-TmMT (low aeration)	0.05			40.0				**Cu_{42}**	**Cu_{38}**
								Cu_{41}	**Cu_{36}**
								Cu_{45}-Cu_{38}	Cu_{38}-Cu_{29}

[a] Several syntheses were performed in each metal supplemented medium, which are numbered by subscripts.

[b] In all cases, the Zn, Cd, Cu and S content was measured by ICP-AES, but only the detectable contents are shown. Protein concentration was calculated from the S content in normal ICP-AES measurements.

[c] Stoichiometries were calculated from the mass difference between the holo- and apo-proteins. Major species are in bold. M = Zn or Cu which have a molecular mass indiscernible by ESI-MS. *Multiple* means that a continuum of peaks are observed between the indicated values.

doi:10.1371/journal.pone.0148651.t002

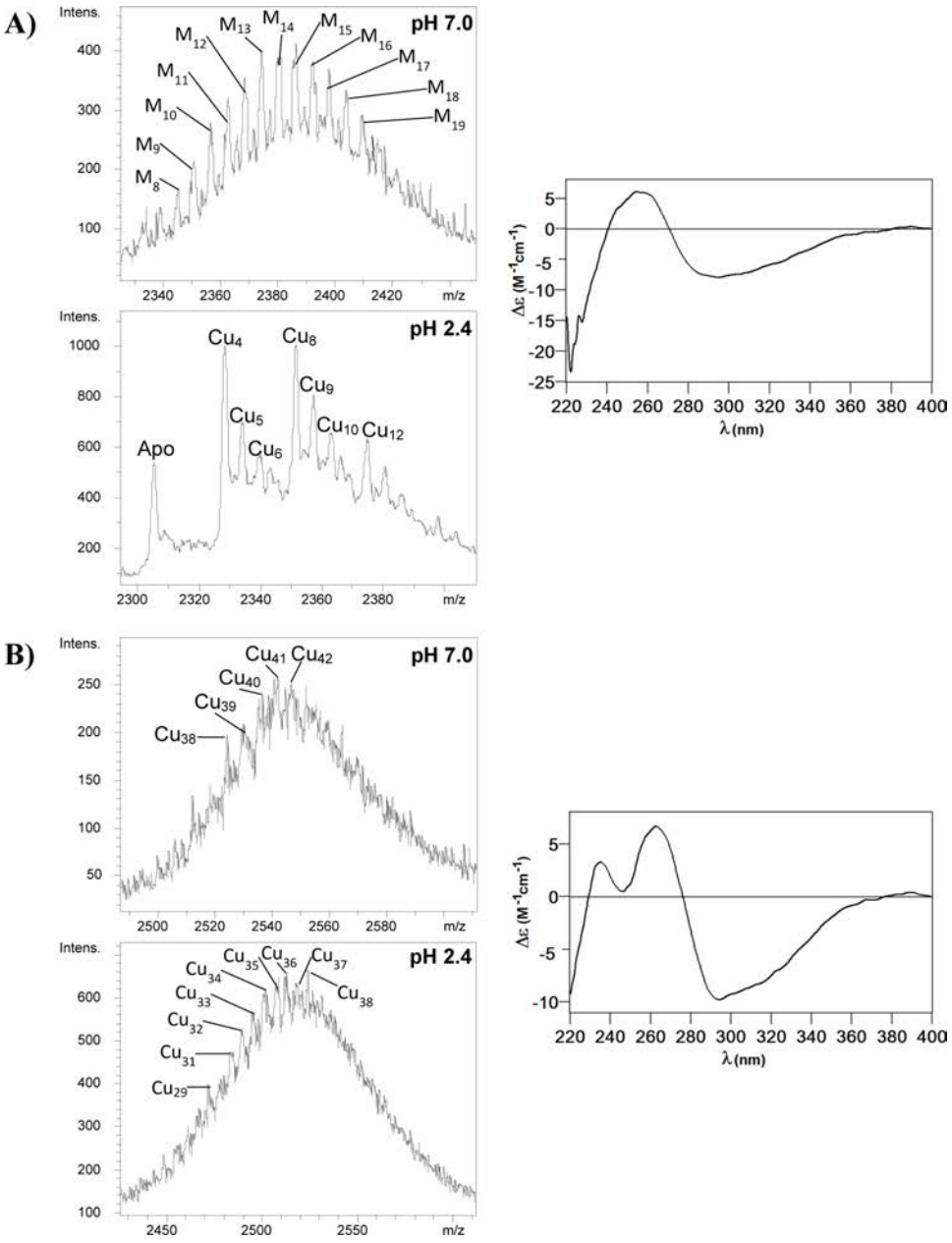


Fig 6. Spectroscopic and spectrometric analyses of the Cu-TmMT preparations. (A) ESI-MS spectra of the recombinant Cu-TmMT synthesized in normally-aerated cultures (at the +11 charge state), run at neutral pH (7.0) and acid pH (2.4), and CD spectrum of the same sample. (B) ESI-MS spectra of the Cu-TmMT synthesized in low-aerated cultures (at the +11 charge state), run at neutral pH (7.0) and acid pH (2.4), and CD spectrum of the same sample.

doi:10.1371/journal.pone.0148651.g006

from normally aerated bacteria showed a wide absorption between 240 and 270 nm, while that of the low-aerated productions precisely showed two maxima with no absorption at 240 nm, which reinforces the presence of Zn(II) ions in the former.

Most informative results about TmMT Cu(I) binding abilities came from the study of the species constituted *in vitro* by Zn/Cu exchange. Nevertheless, these experiments revealed an

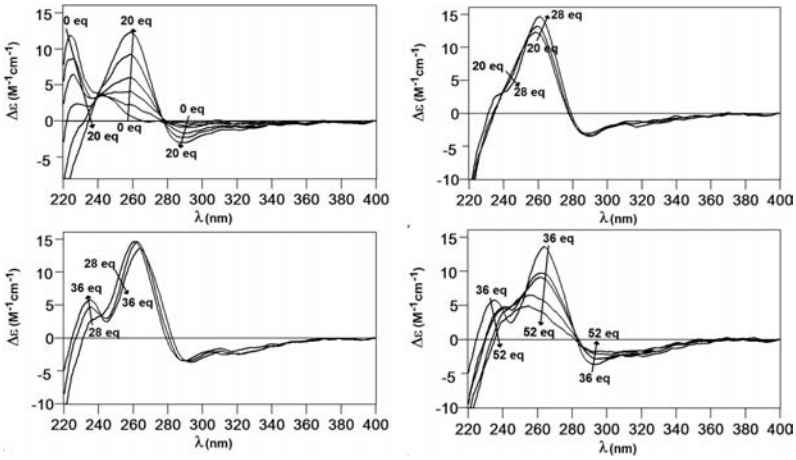


Fig 7. CD spectrum of the Zn/Cu exchange reaction in Zn-TmMT. The CD spectra were recorded every 4 Cu(I) eq added to a 3 μM solution of Zn-TmMT.

doi:10.1371/journal.pone.0148651.g007

unprecedented feature regarding the standard methodological procedure used [31], because spectroscopic (CD, UV-vis) and spectrometric (ESI-MS) measurements could not be performed on the same sample due to incompatible concentration requirements. Thus the CD and UV-vis data was collected on a 3 μM solution of Zn-TmMT with additions of 4 Cu(I) eq each to avoid detector saturation (Fig 7) and the ESI-MS data on a *ca*. 100 μM solution, which enabled the detection of the high number of species present in solution (Fig 8). The CD spectra of the successive reaction steps draw very neat isodichroic points, which strongly suggested a cooperative copper loading and zinc displacement process. These isodichroic points arise by the decrease of the *ca*. 230 nm band and the increase of the 260(+) and 290(-) nm CD absorptions, owing to the creation of Cu-SCys chromophores from 0 to 28 Cu(I) eq added. After a small rearrangement between 28 and 36 Cu(I) eq, further additions provoke the collapse of the CD signal.

It is evident that, from the beginning, the addition of Cu(I) enhances the already high complexity of the initial Zn-TmMT sample, generating a larger number of species in solution (Fig 8). The heterometallic species formed increase in nuclearity up to a maximum of M_{38}-TmMT (M = Zn or Cu), after the addition of 30 Cu(I) eq, which is the step yielding one of the most intense CD spectra (Fig 7). However, the most remarkable hints about the replacement reaction are revealed by the acid ESI-MS analyses, because they suggest how TmMT builds its copper complexes. Hence, the addition of just 4 Cu(I) eq at the beginning of the experiment already gave rise to a predominant Cu_4-core, accompanied by a minor Cu_5-cluster, and species of higher nuclearity which will become significant later. Thence, at the following step, Cu_4- and Cu_5-TmMT are still significant, but the doublet Cu_8- and Cu_9- is almost as important. In the next step, Cu_4-, Cu_5- and Cu_6- Cu_8- have practically disappeared, and the predominant cores are the Cu_{12}- and Cu_{13}-clusters. From this point on (*i.e*. 10 Cu(I) eq added), the complexity of the Cu(I)-containing species is very high, with a continuum of Cu(I)-cores, the Cu(I) content of which increases until approximately 40. It is worthwhile noting that the addition of 25 Cu(I) eq was necessary to totally displace the Zn(II) initially bound to TmMT. Finally, in the presence of an excess of copper beyond 30 Cu(I) eq added, the Cu-TmMT complexes become unstable, so that the Cu(I) nuclearity of the detected species decreases, and at the end of the reaction (Cu overload conditions) only the apo-TmMT peptide is detected, which has been reported also for other Cu-thioneins [38,39]. From these data, a certain periodicity of the predominant Cu(I)

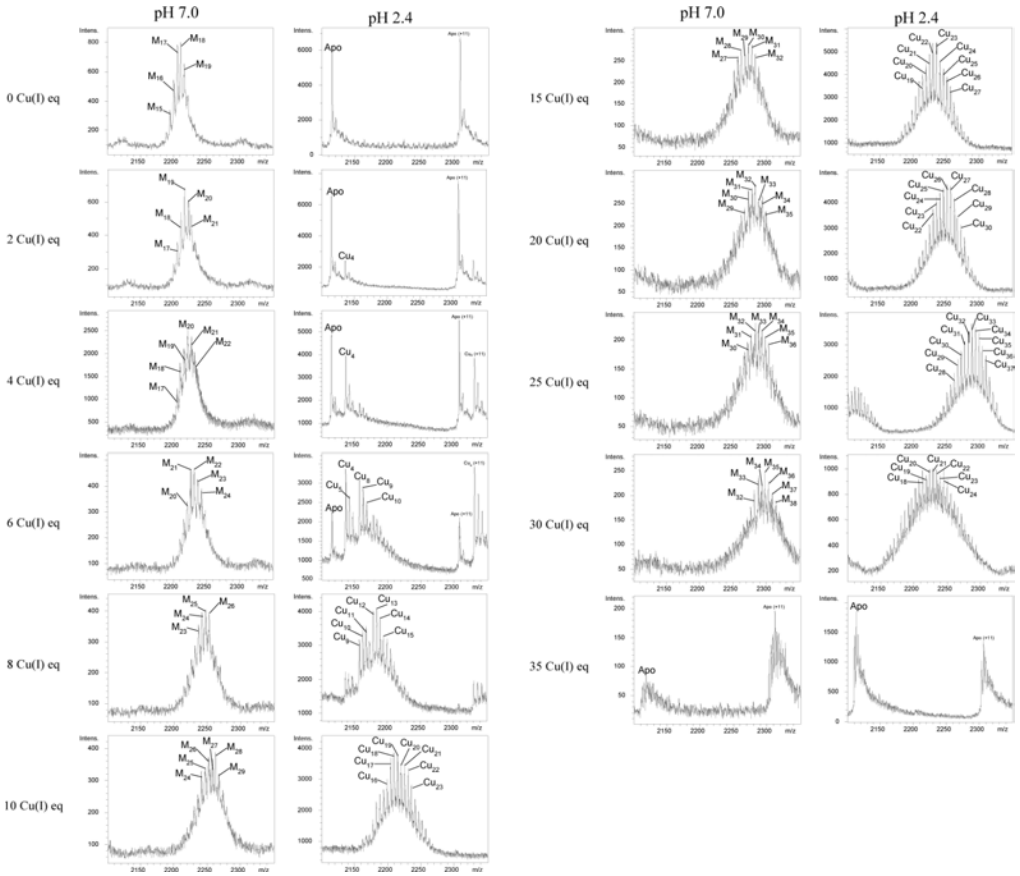


Fig 8. ESI-MS spectra of the Zn/Cu exchange reaction in Zn-TmMT. The ESI-MS spectra were measured, both at pH 7.0 and pH 2.4, on aliquots of a 94 μM solution of TmMT every 2 Cu(I) eq added at the beginning of the reaction and every 5 Cu(I) between 10 and 35 equivalents added.

doi:10.1371/journal.pone.0148651.g008

clusters can be deduced, so that in the two first steps, they were Cu_4-units always accompanied by a more or less important peak with an additional Cu(I) (*i.e.* Cu_4- and Cu_5-, and then Cu_8- and Cu_9), and from here onwards, the following predominant peaks result from adding 5 Cu(I) to the previous species (*i.e.* Cu_{12}- and Cu_{13}-; Cu_{17}- and Cu_{18}-; Cu_{22}- and Cu_{23}-; Cu_{27}- and Cu_{28}- and Cu_{33}-TmMT). Although a strict cooperative process for the Zn/Cu displacement in Zn-TmMT has to be ruled out owing to the many different species coexisting during the experiment, a clear periodicity for the major Cu(I) content of created species is observed, which can be related to the modular structure of the TmMT polypeptide (*cf.* Fig 4). Characterization of the Cu-binding abilities of the also modular *C. neoformans* CnMTs showed that their 7-Cys boxes render Cu_5-clusters of additive behavior [16,17,26], and coincidently the TmMT results suggest that its 7-Cys stretches may exhibit a similar behavior. Additionally, TmMT has two 6-Cys boxes, which can be assumed to optimally bind 4 Cu(I) ions. The remaining 3 Cys would also contribute to bind extra Cu(I) ions. This will explain the series observed for the Cu(I)-core composition in the Zn/Cu exchange reaction. Precisely, both $[Cu_4S_6]$ and $[Cu_5S_7]$ clusters were characterized by classic inorganic chemistry model studies as the most relevant flexible cores in relation to Cu-MT complexes [40,41,42].

Finally, it is worth noting that, also as first noticed for the yeast Crs5 MT [29] and later reported for other Cu-thioneins, the results of TmMT synthesis in Cu-supplemented media at both assayed aerations is reproduced in two different steps of the Zn/Cu exchange process, at

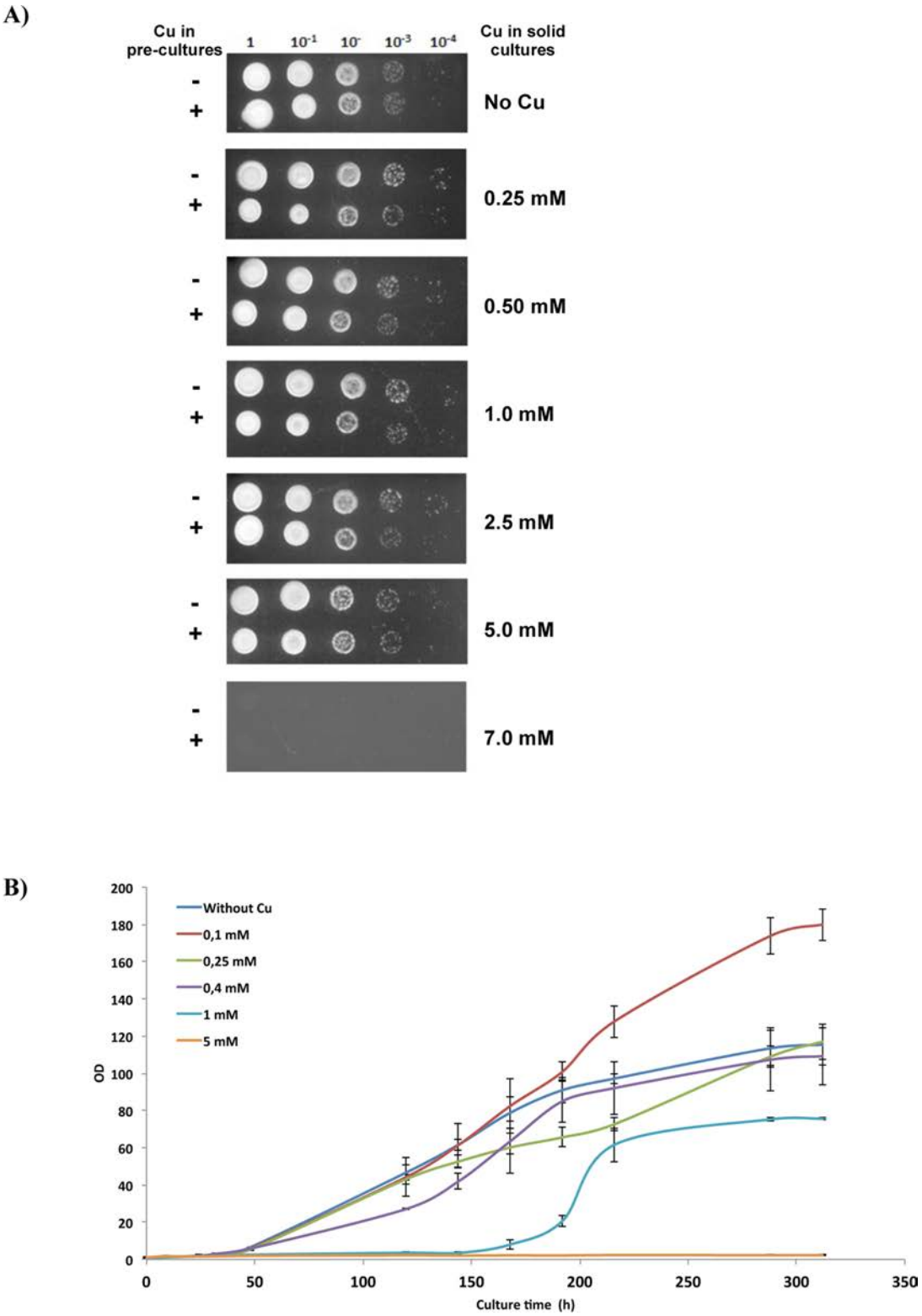


Fig 9. *Tremella mesenterica* growth under Cu supplementation conditions. (A) *T*. *mesenterica* solid cultures. Pre-cultures were grown in YPD liquid medium overnight at 25°C (first row in each condition without added Cu and second row, with 0.1 mM Cu added), then diluted to OD_{600} 0.5 and spotted into the YPD agar plates at 4 serial dilutions. Plates were allowed to grow at 30°C for 3 days. For more details, see the Experimental Procedures section. (B)

Growth curve representing the normalized mean values of the OD_{600} exhibited by the liquid cultures grown in MY medium supplemented with the indicated Cu concentrations. Results represent the mean and standard deviation (vertical bars) of at least three replicates.

doi:10.1371/journal.pone.0148651.g009

least in relation to the Cu(I) cores present. Hence, the results of the Cu-TmMT production in regularly aerated cultures correspond to the Zn-TmMT+6 Cu(I) eq added stage (*cf*. CD features and ESI-MS spectra in Figs 6, 7 and 8), while the results from low aerated syntheses correspond to the last steps of the titration, just before the unfolding of the Cu-TmMT complexes caused by an excess of Cu(I) ions (*cf*. CD features and ESI-MS spectra in Figs 6, 7 and 8). This indicates that the results of these recombinant syntheses effectively correspond to two different situations of Cu-availability.

Copper tolerance of native *Tremella mesenterica* and *Saccharomyces cerevisiae* cells transformed with TmMT

T. *mesenterica* copper tolerance assayed in solid cultures revealed a rather unaltered growth up to 5.0 mM Cu, and at 7.0 mM Cu cells were unable to grow. In this experiment, each Cu concentration was assayed in two conditions, depending on whether the pre-culture spotted on the plates had been supplemented with Cu (0.1 mM) or not (Fig 9A). No significant differences could be observed. When Cu tolerance was measured in liquid cultures, two different phases were observed in long-term cultures (up to 300 h), taking into account that this fungus exhibits a slow growth, and a clear increase in the OD_{600} was not observed until 50 h after inoculation (Fig 9B). Hence, until 150 h growth, all the cultures grew at an equivalent rate, except those supplemented with 1 mM Cu, which exhibited a lengthened lag phase in concordance with the stress imposed by the excess of Cu. All these results confirm a high Cu tolerance for this fungus, comparable to that of the CUP1 multicopy *S*. *cerevisiae* strains [43], considering that only one copy of the *TmMT* gene was detected in the genome by the *in silico* screening. However, it has to be noted that this increased tolerance is manifested in lasting cultures, maybe in relation to the slow *T*. *mesenterica* growth dynamics. Noteworthy, the culture supplemented with 0.1 mM Cu exhibited a maintained increased growth in relation to all other conditions, so that it could be hypothesized that the optimal growth of *T*. *mesenterica* depends on the presence of this metal ion in its natural environment.

To confirm the ability of the *TmMT* to confer high-Cu resistance, the effect of its heterologous expression in the *S*. *cerevisiae* 51-2c-Δc5 strain, which lacks its two MT genes (*CUP1* and *CRS5*) was assayed. The ability of the cells transformed with *TmMT*-p424, to grow in liquid media supplemented with increasing copper concentrations was represented as percentage of the growth exhibited in cultures with no Cu supplementation (Fig 10). For comparative purposes, the *CUP1*-p424 and *mMT1*-p424 constructs, containing respectively the *S*. *cerevisiae* and the mouse MT1 isoforms, as well as the void p424 plasmids were also assayed. The results unambiguously show the extraordinary tolerance to Cu conferred by the expression of *TmMT* in yeast cells, in all the assayed Cu concentrations, up to 30 μM. What is more, it allows cell survival, and at considerable levels, beyond 10 μM Cu, when expression of either *CUP1* or *mMT1* fails to maintain cell growth (Fig 10).

Conclusions

One gene encoding a metallothionein polypeptide is present in the *Tremella mesenterica* genome, hereinafter called, *TmMT*. It has been identified *in silico* from a partial protein sequence found in a *T*. *mesenterica* data bank as starting query, and also from total mRNA retrotranscription and specific amplification, which has confirmed that the predicted *TmMT*

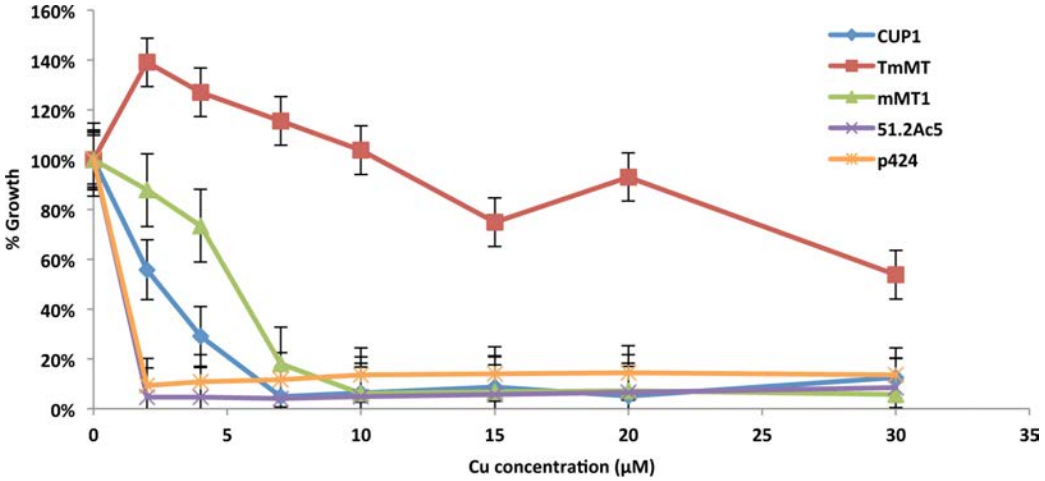


Fig 10. Effect of the heterologous expression of the *T. mesenterica* TmMT in *S. cerevisiae* 51.2cΔc5 MT-null strain. Yeast cells were transformed with the different p424-MT constructs. Besides p424-TmMT, cells transformed with the void p424, and the p424-CUP1 (the yeast MT) and p424-mMT1 (the mammalian isoform MT1) plasmids were also included in the assay for comparative purposes. 3-mL cultures grown overnight at 30°C were diluted to OD_{600} 0.1 in 3-mL fresh SC-Trp-Ura medium supplemented with 0, 2, 4, 7, 10, 15, 20, and 30 μM $CuSO_4$ and grown for 18h at 30°C. The growth of the host 51.2cΔc5 strain cultured in complete medium was also assayed. Growth was evaluated in liquid cultures and it is represented as the percentage of the growth rate attained in a non-Cu supplemented medium. All the experiments were carried out at least by triplicate and the plotted points represent the mean value for each condition associated to its corresponding error bar. The data unambiguously show the extraordinary copper tolerance conferred by TmMT. For more details, see the Experimental Procedures section.

doi:10.1371/journal.pone.0148651.g010

ORF was a *real* coding sequence of uncommon length for a *MT* gene. The *TmMT* gene covers more than 2 kb of DNA, and its coding sequence is separated in 10 small exons -sizes from 63 bp to 105 bp-, interrupted by 9 introns, from 48 bp to 204 bp. The 5' UTR of this cDNA is 69-bp long, and in the upstream gene sequence, a putative TATA box and an MRE element have been identified. The cDNA coding portion encompasses 774 bp, and the resulting protein has 257 amino acids, 57 of which are Cys distributed in CXXC motifs as corresponding to the common hallmarks of Metallothioneins. Therefore, this is the longest MT polypeptide reported up to now, beyond the *Paramecium sp*. PMCd1 protein (203 aa) the fungal *C. neoformans* CnMT2 (183 aa) and CnMT1 (122 aa) isoforms. Coincidently with the *C. neoformans* MTs, the TmMT polypeptide exhibits clear internal repetition patterns, which suggests that modular amplification processes of a primeval fungal MT -represented by the paradigmatic short MTs reported in the *Neurospora* or *Agaricus* genus- would be in the bases of the genesis of long MTs in the Tremellales order of Ascomycota fungi. In the case of TmMT, the most obvious arrangement of hypothetical internal repetitions is the alignment shown in Fig 4B, for which six of the units depict a -$CXCX_3$CSCPPGXCXCAXCP- motif including 7 Cys, as for paradigmatic fungal MTs (*i.e. Neurospora* MT), and includes two units with 6 Cys and a N-terminal segment with only 3 Cys.

The analysis of the metal binding abilities of TmMT confirmed that it exhibits all the features of a high-capacity Cu-thionein. Hence, when TmMT coordinates Zn(II) ions, it renders a mixture of species of different stoichiometries, which is also the case for Cd(II) coordination. This patent inability for divalent metal ion coordination is accompanied by a clear behavior of Cu-thionein, although its large size entails that when synthesized in regular Cu concentrations, the obtained complexes are heterometallic (Zn, Cu) species, just as the *C. neoformans* CnMT1 and CnMT2 were. However, at high Cu concentrations (*i.e.* produced in host cells grown under poor aeration), TmMT is able to render homometallic complexes of as high nuclearity as 45 Cu(I), with the major species being Cu_{40} to Cu_{42}-TmMT. The modular composition

reported for the *C. neoformans* Cu-CnMT complexes is probably also re-encountered in Cu-TmMT, since its polypeptide sequence is easily dissected in alignable stretches: six 7-Cys boxes, two 6-Cys boxes, and three N-terminal Cys. This assumption would also explain the Cu contents of the most favored Cu-species. Strong similarity is also observed for some residues located between the coordinating Cys (*i.e.* the proline doublet after the second CXC motif, a single proline after the last Cys of the segment, a clear predominance of small (glycine, alanine) residues, and the conservation of charged residues in other key positions (lysine)). Thus, it is tempting to associate these features, not only to the evolutionary origin and conservation of this protein pattern, but also to their optimization for Cu(I) coordination geometry.

Of course, the most interesting question raised by these results concerns the physiological role that this MT protein may play in its native surrounding. A first approach is to consider the natural habitat where *Tremella* grows, mainly decaying wood, suggests that Cu-handling may be of the utmost importance to provide active lignin-metabolizing enzymes, such as laccases and peroxidases, which are well-known Cu-containing oxidases. This would be in concordance with the constitutive, rather than inducible, *TmMT* pattern expression, since it could be considered a housekeeping gene functionality. Appealingly, this hypothesis might, in the end, link the biological function of TmMT with that proposed for the Cu-MTs of the pulmonate snails, which are supposed to serve as storage/chaperones of Cu for the hemocyanins, the respiratory O_2 carriers of these organisms [44].

Supporting Information

S1 Table. Experimental molecular masses (ESI-MS results) and calculated molecular masses for Zn-TmMT and Cd-TmMT species. Subindexes describe different syntheses. Major species are in bold.
(PDF)

S2 Table. Experimental molecular masses (ESI-MS results) and calculated molecular masses for Cu-TmMT species synthesized in regular and low-aerated *E. coli* cultures. Major species are in bold. M = Zn or Cu.
(PDF)

Acknowledgments

Technical advise of Dra. E. Jimenez and Dra. A. Espart and experiment support of M. Sebastian Artime are greatly appreciated. We thank the Centres Científics i Tecnològics (CCiT) de la Universitat de Barcelona (ICP-AES, DNA sequencing) and the Servei d'Anàlisi Química (SAQ) de la Universitat Autònoma de Barcelona (CD, UV-vis, ESI-MS) for allocating instrument time.

Author Contributions

Conceived and designed the experiments: MC SA. Performed the experiments: PIE SGM WL SC OP. Analyzed the data: OP MC SA. Wrote the paper: OP MC SA.

References

1. Sigel A, Sigel H, Sigel RKO. Metal ions in Life Sciences; vol. 5: Metallothioneins and Related Chelators. 1st ed. Cambridge: The Royal Society of Chemistry; 2009.
2. Capdevila M, Bofill R, Palacios O, Atrian S. State-of-the-art of metallothioneins at the beginning of the 21st century. Coord Chem Rev. 2012; 256: 46–62.

3. Blindauer C. Binding, Transport and Storage of Metal Ions in Biological Cells. In: Maret W, Wedd A, editors.) Metallothioneins. (RSC Metallobiology Series 2). Cambridge: The Royal Society of Chemistry; 2014. pp. 594–653.
4. Vasak M, Hasler DW. Metallothioneins: new functional and structural insights. Curr Opin Chem Biol. 2000; 4: 177–183. PMID: 10742189
5. Capdevila M, Atrian S. Metallothionein protein evolution: a miniassay. J Biol Inorg Chem. 2011; 16: 977–989. doi: 10.1007/s00775-011-0798-3 PMID: 21633816
6. Binz PA, Kägi JHR. () Metallothionein: Molecular evolution and classification. In: Klaassen C, editor. Metallothionein IV. Basel: Birkhäuser Verlag; 1999. pp.7–13.
7. Valls M, Bofill R, Gonzalez-Duarte R, Gonzalez-Duarte P, Capdevila M, Atrian S. A New Insight into Metallothionein (MT) Classification and Evolution. J Biol Chem. 2001; 276: 32835–32843. PMID: 11413132
8. Bofill R, Capdevila M, Atrian S. Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu-thioneins. Metallomics 2009; 1: 229–234. doi: 10.1039/b904953c PMID: 21305119
9. Palacios O, Atrian S, Capdevila M. Zn- and Cu-thioneins: a functional classification for metallothioneins?. J Biol Inorg Chem. 2011; 16: 991–1009. doi: 10.1007/s00775-011-0827-2 PMID: 21823038
10. Palacios O, Pagani A, Pérez-Rafael S, Egg M, Höckner M, Brandstätter A, et al. How Metazoan Metallothioneins achieved Metal Specificity: Evolutionary Differentiation in the Mollusc MT Gene Families. BMC Biol. 2011; 9: 4.
11. Lerch K. Copper metallothionein, a copper-binding protein from *Neurospora crassa*. Nature. 1980; 284: 368–370. PMID: 6444697
12. Munger K, Lerch K. Copper metallothionein from the fungus *Agaricus bisporus*: chemical and spectroscopic properties. Biochemistry. 1985; 24: 6751–6756.
13. Prinz R, Weser U. A naturally occurring Cu-thionein in *Saccharomyces cerevisiae*. Hoppe Seylers Z Physiol Chem. 1975; 356: 767–776. PMID: 1102411
14. Winge DR, Nielson KB, Gray WR, Hamer DH. Yeast metallothionein. Sequence and metal-binding properties. J Biol Chem. 1985; 260: 14464–14470. PMID: 3902832
15. Nemer M, Wilkinson DG, Travaglini EC, Sternberg EJ, Butt TR. Sea urchin metallothionein sequence: Key to an evolutionary diversity. Proc Natl Acad Sci USA. 1985; 82: 4992–4994. PMID: 3860837
16. Ding C, Festa RA, Chen Y- L, Espart A, Palacios O, Espín J, et al. *Cryptococcus neoformans* copper detoxification machinery is critical for fungal virulence. Cell Host Microbe. 2013; 13: 265–276. doi: 10.1016/j.chom.2013.02.002 PMID: 23498952
17. Palacios O, Espart A, Espín J, Ding C, Thiele DJ, Atrian S, et al. Full characterization of the Cu-, Zn- and Cd-binding properties of CnMT1 and CnMT2, two metallothioneins of the pathogenic fungus *Cryptococcus neoformans* acting as virulence factors. Metallomics. (2014; 6: 279–291. doi: 10.1039/c3mt00266g PMID: 24317230
18. Espart A, Marín M, Gil-Moreno S, Palacios P, Amaro F, Martín-González A, et al. Hints for Metal-Preference Protein Sequence Determinants: Different Metal Binding Features of the Five *Tetrahymena thermophila* Metallothioneins. Int J Biol Sci. 2015; 11: 456–471. doi: 10.7150/ijbs.11060 PMID: 25798065
19. Findley K, Rodriguez-Carres M, Metin B, Kroiss J, Fonseca A, Vilgalys R, et al. Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. Eukariot Cell. 2009; 8: 353–361.
20. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D et al. The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. 2012; 40: D26–D32. doi: 10.1093/nar/gkr947 PMID: 22110030
21. Wickerman L. Taxonomy of yeasts. Dept Agric Tech Bull. 2012; 1029: 1–56.
22. Sherman F, Fink GF, Hicks J. Techniques and protocols: 6. Yeast RNA Isolation. In: Kaiser C, Michaelis S, Mitchell A, editors. Methods in Yeast Genetics. A Cold Spring Harbor Laboratory Course Manual, 1994 edition. New York: Cold Spring Harbor Laboratory Press; 1994. pp. 151–155.
23. Guirola M, Perez-Rafael S, Capdevila M, Palacios O, Atrian S. Metal Dealing at the Origin of the Chordata Phylum: The Metallothionein System and Metal Overload Response in *Amphioxus*. PLoS ONE 2012; 7: e43299. doi: 10.1371/journal.pone.0043299 PMID: 22905252
24. Capdevila M, Cols N, Romero-Isart N, González-Duarte R, Atrian S, González-Duarte P. Recombinant synthesis of mouse Zn_3-ß and Zn_4-α metallothionein 1 domains and characterization of their cadmium (II) binding capacity. Cell Mol Life Sci. 1997; 53: 681–688. PMID: 9351472
25. Culotta VC, Howard WR, Liu XF. CRS5 encodes a metallothionein-like protein in *Saccharomyces cerevisiae*. J Biol Chem. 1994; 269: 25295–25302. PMID: 7929222

26. Espart A, Gil-Moreno S, Palacios P, Capdevila M, Atrian S. Understanding the internal architecture of long metallothioneins: 7-Cys building blocks in fungal (*C. neoformans*) MTs. Mol Microbiol. 2015; 98: 977–992. doi: 10.1111/mmi.13171 PMID: 26287377

27. Mumberg D, Müller R, Funk M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene. 1995; 156: 119–122. PMID: 7737504

28. Gietz RD, Woods RA. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol. 2002; 350: 87–96. PMID: 12073338

29. Pagani A, Villarreal L, Capdevila M, Atrian S. The *Saccharomyces cerevisiae* Crs5 metallothionein metal-binding abilities and its role in the response to zinc overload. Mol Microbiol. 2007; 63: 256–269. PMID: 17163970

30. Cols N, Romero-Isart N, Capdevila M, Oliva B, Gonzàlez-Duarte P, Gonzàlez-Duarte R, et al. Binding of excess cadmium(II) to Cd_7-metallothionein from recombinant mouse Zn_7-metallothionein 1. UV-VIS absorption and circular dichroism studies and theoretical location approach by surface accessibility analysis. J Inorg Biochem. 1997; 68: 157–166. PMID: 9352652

31. Bofill R, Palacios O, Capdevila M, Cols N, Gonzalez-Duarte R, Atrian S, et al. A new insight into the Ag^+ and Cu^+ binding sites in the metallothionein beta domain. J Inorg Biochem. 1999; 73: 57–64. PMID: 10212995

32. Bongers J, Walton CD, Richardson DE, Bell JU. Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma atomic emission spectrometric determination of sulfur. Anal Chem. 1988; 60: 2683–2686. PMID: 3245594

33. Capdevila M, Domenech J, Pagani A, Tio L, Villarreal L, Atrian S. Zn- and Cd-metallothionein recombinant species from the most diverse phyla may contain sulfide (S^{2-}) ligands. Angew Chem Int Ed Engl. 2005; 44: 4618–4622. PMID: 15991200

34. Fabris D, Zaia J, Hatout Y, Fenselau C. Retention of thiol protons in two classes of protein zinc coordination centers. J Am Chem Soc. 1996; 118: 12242–12243.

35. Shuja RN, Shakoori AR. Identification and cloning of first cadmium metallothionein like gene from locally isolated ciliate, *Paramecium* sp. Mol Biol Rep. 2009; 36: 549–560. doi: 10.1007/s11033-008-9213-8 PMID: 18273691

36. Diaz S, Amaro F, Rico D, Campos V, Benitez L, Martín-González A, et al. *Tetrahymena* metallothioneins fall into two discrete subfamilies. PLoS ONE. 2007; 3: e291.

37. Domenech J, Bofill R, Tinti A, Torreggiani A, Atrian S, Capdevila M. Comparative insight into the Zn(II)-, Cd(II)- and Cu(I)-binding features of the protozoan *Tetrahymena pyriformis* MT1 metallothionein. Biochim Biophys Acta. 2008; 1784: 693–704. doi: 10.1016/j.bbapap.2008.01.008 PMID: 18280816

38. Perez-Rafael S, Kurz A, Guirola M, Capdevila M, Palacios O, Atrian S.X Is MtnE, the fifth *Drosophila metallothionein*, functionally distinct from the other members of this polymorphic protein family? Metallomics. 2008; 4: 342–349.

39. Artells E, Palacios O, Capdevila M, Atrian S. In vivo-folded metal-MT3 complexes reveal the Cu- rather than Zn-thionein character of this brain-specific mammalian metallothionein. Metallomics. 2013; 5: 1397–1410. doi: 10.1039/c3mt00123g PMID: 23925449

40. Winge D, Dameron CT, George GN, Pickering IJ, Dance IG. Cuprous-thiolate polymetallic clusters in biology. In: Karin KD, Tyeklar Z, editors. Bioinorganic Chemistry of Copper. New York: Chapman & Hall; 1994. pp. 110–122.

41. Maiti BK, Pal K, Sarkar S. Flexible Cu^I-thiolate clusters with relevance to Metallothioneins. Eur J Inor Chem. 2007; 2007: 5548–5555.

42. Pushie MJ, Zhang L, Pickering IJ, George GN. The fictile coordination chemistry of cuprous-thiolate sites in copper chaperones. Biochim Biophys Acta. 2012; 1817: 938–947. doi: 10.1016/j.bbabio.2011.10.004 PMID: 22056518

43. Adamo GM, Lotti M, Tamas MJ, Brocca S. Amplification of the *CUP1* gene is associated with evolution of copper tolerance in *Saccharomyces cerevisiae*. Microbiol. 2012; 158: 2325–2335.

44. Dallinger R, Chabicovsky M, Hödl E, Prem C, Hunziker P, Manzl C. Copper in *Helix pomatia* (Gastropoda) is regulated by one single cell type: differently responsive metal pools in rhogocytes. Am J Physiol. 2005; 189: R1185–R1195.

ANNEX I: SUPPLEMENTARY MATERIAL

THE FUNGUS TREMELLA MESENTERICA ENCODES THE LONGEST METALLOTHIONEIN CURRENTLY KNOWN: GENE, PROTEIN AND METAL BINDING CHARACTERIZATION

The file includes:

Table S1

Table S2

S1 Table. Experimental molecular masses (ESI-MS results) and calculated molecular masses for Zn-TmMT and Cd-TmMT species. Subindexes describe different syntheses. Major species are in bold.

Synthesis (metal supplemented)	ESI-MS pH	Identified species	Experimental mass (Da)	Theoretical mass (Da)
Zn-TmMT$_1$	7.0	Zn_{21}	26708	26707.8
		Zn_{20}	**26644**	**26645.4**
		Zn_{19}	**26578**	**26582.0**
		Zn_{18}	26515	26518.6
	2.4	apo	25377	25377.6
Zn-TmMT$_2$	7.0	**Zn_{21}**	**26711**	**26707.8**
		Zn_{20}	**26646**	**26645.4**
		Zn_{19}	26575	26582.0
		Zn_{18}	26514	26518.6
		Zn_{17}	26448	26455.2
		Zn_{16}	26384	26391.9
	2.4	apo	25375	25377.6
Zn-TmMT$_3$	7.0	Zn_{19}	26570	26582.0
		Zn_{18}	**26513**	**26518.6**
		Zn_{17}	**26448**	**26455.2**
		Zn_{16}	26384	26391.9
		Zn_{15}	26322	26328.6
	2.4	apo	25375	25377.6
Cd-TmMT$_1$	7.0	**Cd_{25}**	**28136**	**28137.9**
		Cd_{24}	28025	28026.6
		Cd_{23}	27915	27916.2
	2.4	apo	25376	25377.6
Cd-TmMT$_2$	7.0	**Cd_{25}**	**28136**	**28137.9**
		Cd_{24}	28025	28026.6
		Cd_{23}	27915	27916.2
	2.4	apo	25377	25377.6

Table S2. Experimental molecular masses (ESI-MS results) and calculated molecular masses for Cu-TmMT species synthesized in regular and low-aerated *E. coli* cultures. Major species are in bold. M= Zn or Cu.

Synthesis (metal supplemented)	ESI-MS pH	Identified species	Experimental mass (Da)	Theoretical mass (Da)
Cu-TmMT (regular aeration)	7.0	M_{19}	26557	26564.5
		M_{18}	26493	26502.0
		M_{17}	26428	26440.9
		M_{16}	26363	26378.4
		M_{15}	**26302**	**26315.8**
		M_{14}	26238	26253.3
		M_{13}	26176	26190.7
		M_{12}	26104	26128.1
		M_{11}	26046	26065.6
		M_{10}	25978	26003.0
		M_9	25907	25940.5
		M_8	25845	25877.9
	2.4	Cu_{12}	26107	26128.1
		Cu_{10}	25971	26003.0
		Cu_9	25917	25940.5
		Cu_8	**25853**	**25877.9**
		Cu_6	25728	25752.9
		Cu_5	25665	25690.3
		Cu_4	**25600**	**25627.8**
		apo	25347	25377.6
Cu-TmMT (low aeration)	7.0	Cu_{45}	28182	28192.2
		Cu_{44}	28138	28129.6
		Cu_{43}	28061	28067.1
		Cu_{42}	**28006**	**28004.6**
		Cu_{41}	**27936**	**27942.0**
		Cu_{40}	27874	27879.5
		Cu_{39}	27819	27816.9
		Cu_{38}		
	2.4	**Cu_{38}**		
		Cu_{37}		
		Cu_{36}		
		Cu_{35}		
		Cu_{34}		
		Cu_{33}		
		Cu_{32}		
		Cu_{31}		
		Cu_{30}		
		Cu_{29}		

ANNEX II:

Biomphalaria glabrata Metallothionein: Lacking Metal Specificity of the Protein and Missing Gene Upregulation Suggest Metal Sequestration by Exchange Instead of through Selective Binding

Int. J. Mol. Sci. 2017, 18, 1457

DOI: 10.3390/ijms18071457

International Journal of
Molecular Sciences

MDPI

Article

Biomphalaria glabrata Metallothionein: Lacking Metal Specificity of the Protein and Missing Gene Upregulation Suggest Metal Sequestration by Exchange Instead of through Selective Binding

Michael Niederwanger [1], Sara Calatayud [2], Oliver Zerbe [3], Sílvia Atrian [2,†], Ricard Albalat [2], Mercè Capdevila [4], Òscar Palacios [4] and Reinhard Dallinger [1,*]

1 Institute of Zoology and Center of Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria; michael.niederwanger@uibk.ac.at
2 Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, E-08028 Barcelona, Spain; scalatro7@alumnes.ub.edu (S.C.); ralbalat@ub.edu (R.A.)
3 Department of Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; oliver.zerbe@chem.uzh.ch
4 Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Cerdanyola del Vallès, Spain; merce.capdevila@uab.cat (M.C.); oscar.palacios@uab.cat (Ò.P.)
* Correspondence: reinhard.dallinger@uibk.ac.at; Tel.: +43-512-5075-1861; Fax: +43-512-5075-1899
† Deceased.

Received: 30 May 2017; Accepted: 1 July 2017; Published: 6 July 2017

Abstract: The wild-type metallothionein (MT) of the freshwater snail *Biomphalaria glabrata* and a natural allelic mutant of it in which a lysine residue was replaced by an asparagine residue, were recombinantly expressed and analyzed for their metal-binding features with respect to Cd^{2+}, Zn^{2+} and Cu^{+}, applying spectroscopic and mass-spectrometric methods. In addition, the upregulation of the *Biomphalaria glabrata MT* gene was assessed by quantitative real-time detection PCR. The two recombinant proteins revealed to be very similar in most of their metal binding features. They lacked a clear metal-binding preference for any of the three metal ions assayed—which, to this degree, is clearly unprecedented in the world of *Gastropoda* MTs. There were, however, slight differences in copper-binding abilities between the two allelic variants. Overall, the missing metal specificity of the two recombinant MTs goes hand in hand with lacking upregulation of the respective *MT* gene. This suggests that in vivo, the *Biomphalaria glabrata* MT may be more important for metal replacement reactions through a constitutively abundant form, rather than for metal sequestration by high binding specificity. There are indications that the MT of *Biomphalaria glabrata* may share its unspecific features with MTs from other freshwater snails of the *Hygrophila* family.

Keywords: metallothionein; metal-binding; zinc; copper; cadmium; *Biomphalaria glabrata*; *Gastropoda*; *Hygrophila*

1. Introduction

In many species of *Gastropoda* (snails and slugs), cadmium (Cd) and copper (Cu) metabolism and detoxification are apparently linked to the expression of metal-specific metallothionein (MT) isoforms. Terrestrial snails like the Roman snail (*Helix pomatia*) of the *Stylommatophora* phylum, for example, possess Cd and Cu-specific *MT* genes whose transcriptional activation can be induced by metal exposure, leading to the expression of two-domain MT proteins which preferentially bind and inactivate the respective cognate metal ion [1–3]. Likewise, snails of the taxonomic

clade of *Caenogastropoda* such as the marine periwinkle (*Littorina littorea*), possess a Cd-specific MT whose expression is strongly induced by Cd exposure [4] and environmental stressors [5]. I contrast to terrestrial snails from the *Stylommatophora* phylum, however, the MTs of *Littorina littorea* and other species of the *Caenogastropoda* clade possess three metal-binding domains [4]. Overall, the metal-binding performance of most of these MTs can be attributed to an intact domain structure with an optimized metal-binding stoichiometry, where every domain provides nine cysteine-linked sulfur atoms that coordinate with high affinity three Cd^{2+} or four to six Cu^{+} ions. In accordance with the metal-specific binding preference of the expressed proteins, the respective *MT* genes can be upregulated in response to exposure to the corresponding metal ion [2,6]. It has been suggested that the high binding specificity of some snail MTs for Cd^{2+} may serve detoxification of this harmful metal by keeping Cd^{2+} pathways within the snail organism strictly separated from pathways of other, essential metal ions [1].

In freshwater pulmonate snails of the *Hygrophila* clade, the situation may be different. Although they possess MTs [7], an important pathway of metal detoxification in these species is represented by metal-binding to phytochelatins as shown, for example, for the Great Pond snail, *Lymnaea stagnalis* [8]. Accordingly, the capacity of *Hygrophila* to express *MT*s in response to metal stress is apparently inhibited or strongly reduced when compared to other gastropod species, as recently demonstrated in the Bladder snail, *Physa acuta* [9]. *Biomphalaria glabrata* is another species of *Hygrophila* that lives in tropical and subtropical rivers and ponds, being one of the most important intermediate hosts of the trematode, *Schistosoma mansoni* [10,11] that infects with schistosomiasis millions of people worldwide [12]. In the present study we have examined the metal-binding performance of two MT allelic variants of *Biomphalaria glabrata*, the wildtype form (BgwtMT) and a natural mutant in which a lysine residue has been replaced by an asparagine residue (BgKNMT). Last but not least, it was also explored how the *Biomphalaria glabrata* MT variants compare with structural and functional features of MTs from other *Hygrophila* species. In addition, the expression of the *Biomphalaria glabrata MT* gene was assessed under control and Cd exposure conditions.

2. Results and Discussion

2.1. The BgwtMT and BgKNMT Recombinant Proteins

The primary structure of the wildtype MT from *Biomphalaria glabrata* (BgwtMT) was originally obtained by translation from the nucleotide sequence of the corresponding gene (see below), and afterwards confirmed by mRNA isolation and sequencing from laboratory-grown living snails [13] (in preparation). The second sequence originates from a natural allelic MT variant with an amino acid replacement (K→N) (BgKNMT), and was chosen from a number of different allelic mutants, all of them characterized from individuals of the same laboratory-grown population. The primary structure of the two recombinant proteins is shown in Figure 1. Due to the specific recombinant expression conditions, the two proteins contain two additional amino acid residues (GS) at their N-termini in relation to their native isoforms previously characterized [13] (in preparation). As previously shown [14], these modifications do not interfere with the metal-binding capacity of the recombinant proteins.

The mass spectra of the recombinant productions of Zn-BgwtMT and Zn-BgKNMT recorded under acidic conditions (Figure 2) allowed confirmation of the expected masses of both proteins according to their sequences.

```
BgwtMT       1 GSMSGKGPNCTEACTGEQCNCGDSCKCGEGCNCPSCKTTKGPNCTEACTG    50
BgKNMT       1 GSMSGKGPNCTEACTGEQCNCGDSCKCGEGCNCPSCKTTKGPNCTEACTG    50

BgwtMT      51 KQCSCGDSCQCGEGCTCSCCKKACTKECTDTECSCGDSCKCGEGCKCSSC   100
BgKNMT      51 KQCSCGDSCQCGEGCTCSCCKNACTKECTDTECSCGDSCKCGEGCKCSSC   100

BgwtMT     101 KAGKCTKSDEGCKTEGHCAKGKCCKS    126
BgKNMT     101 KAGKCTKSDEGCKTEGHCAKGKCCKS    126
```

Figure 1. Amino acid sequences of the studied recombinant metallothioneins: Wild type (BgwtMT) and the naturally mutated (BgKNMT) proteins. **Red**: cysteine residues; underlined **blue**: N-terminal additional residues introduced due to the recombinant expression conditions (see material and methods section); shaded in **green**: wildtype position of K (BgwtMT); shaded in **yellow**: its replacement by N in the natural allelic mutant (BgNKMT). The two protein sequences can be found in the GenBank under the following accession numbers: KT697617 (BgwtMT); and KY963493 (BgKNMT).

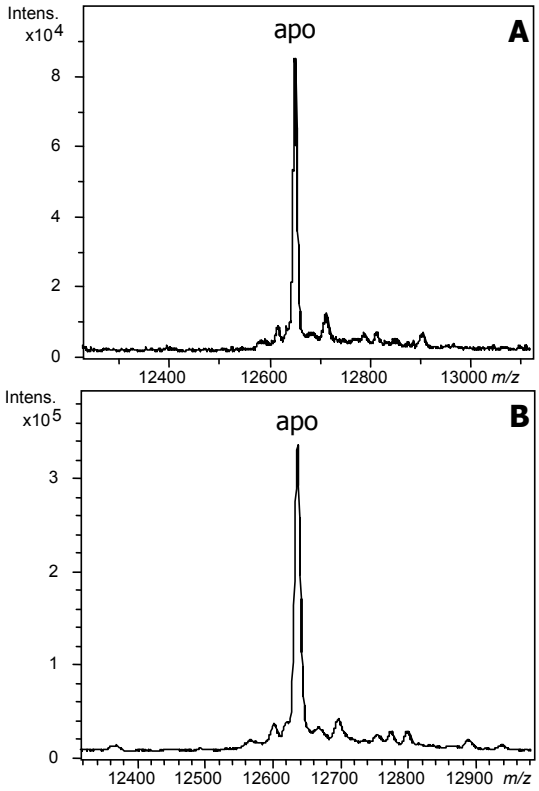


Figure 2. MS spectra of BgMT proteins: Deconvoluted electrospray ionization mass spectrometry (ESI-MS) spectra of the bacterial recombinant production of (**A**) BgwtMT (Experimental Molecular Mass: 12,652) and (**B**) BgKNMT (Experimental Mass: 12,639) in Zn-enriched media, recorded at pH 2.4.

2.2. Divalent Zn(II) and Cd(II)-Binding Features of BgwtMT and BgKNMT

The recombinant production of BgwtMT in Zn-enriched media rendered a mixture of several metallated species (Figure 3, Table 1) with Zn_{11}- and Zn_{10}-BgwtMT being the most abundant ones (Figure 3A). Similarly, the production of the BgKNMT mutant form rendered the same mixture of metal-loaded species, even if the most intense peak correlates with Zn_{10}-BgKNMT (Figure 3B). The CD spectra of both preparations (Figure 3C) are very similar (an exciton coupling band centered at approx.

240 nm, corresponding to the expected Zn-$(SCys)_4$ binding chromophores), thus suggesting a similar folding of both proteins about the Zn(II) ions.

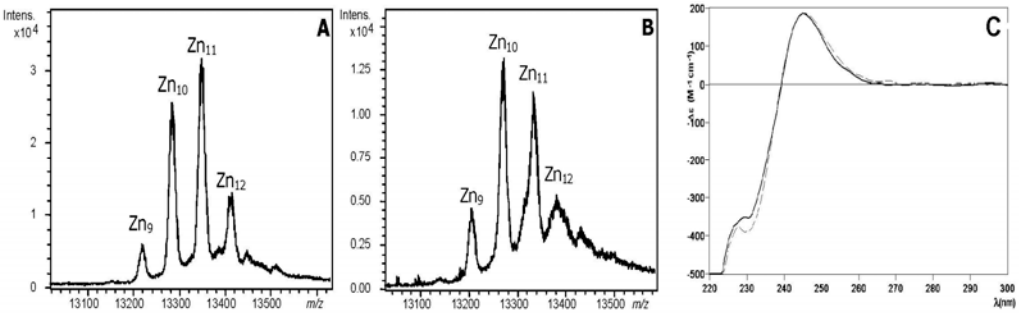


Figure 3. MS and circular dichroism (CD) spectra of Zn-BgMT: Deconvoluted ESI-MS spectra of the recombinant (**A**) Zn-BgwtMT and (**B**) Zn-BgKNMT, recorded at neutral (7.0) pH; (**C**) CD spectra of each Zn(II)-preparation: BgwtMT (solid line) and BgKNMT (dashed).

Table 1. Analytical characterization of the recombinant Zn-, Cd- and Cu-preparations of BgwtMT and BgKNMT. All data for the copper supplemented cultures correspond to normal aeration conditions, since no complexes could be recovered from low aerated cultures. Abbreviations: MT: Metallothionein, ICP-AES: Inductively coupled plasma atomic emission spectroscopy, ESI-MS: electrospray ionization mass spectrometry, Exp. MM: expected molecular mass, Calc. MM: calculated molecular mass.

Supplemented Metal	MT	ICP-AES [a]	ESI-MS [b] pH 7.0	Exp. MM [c]	Calc. MM [d]	ESI-MS [b] pH 2.4	Exp. MM [c]	Calc. MM [d]
Zn	BgwtMT	9.8 Zn 0 Cd 0 Cu	Zn_9- **Zn_{10}-** **Zn_{11}-** Zn_{12}-	13,221 **13,286** **13,350** 13,414	13,222.8 **13,286.2** **13,349.6** 13,413.0	**apo-**	12,652	12,652.3
	BgKNMT	9.4 Zn 0 Cd 0 Cu	Zn_9- **Zn_{10}-** **Zn_{11}-** Zn_{12}-	13,207 **13,272** **13,336** 13,400	13,208.7 **13,272.1** **13,335.5** 13,398.9	**apo-**	12,639	12,638.2
Cd	BgwtMT	0 Zn 12.9 Cd 0 Cu	Cd_{12}- Cd_{13}- **$Cd_{14}S$-** $Cd_{15}S$- Cd_{16}-	13,980 14,090 **14,229** 14,341 14,418	13,977.2 14,087.6 **14,230.0** 14,340.5 14,418.9	apo- Cd_7S- **Cd_8S-**	12,652 13,457 **13,567**	12,652.3 13,457.2 **13,567.6**
	BgKNMT	0 Zn 13.3 Cd 0 Cu	Cd_{12}- Cd_{13}- **$Cd_{14}S$-** $Cd_{15}S$- Cd_{16}-	13,960 14,075 **14,215** 14,327 14,406	13,963.2 14,073.6 **14,216.0** 14,326.4 14,404.8	**apo-** Cd_7S- **Cd_8S-**	**12,638** 13,442 **13,553**	**12,638.2** 13,443.1 **13,553.5**
Cu	BgwtMT	3.7 Zn 0 Cd 9.5 Cu	M_{11}-MT M_{12}-MT **M_{13}-MT** **M_{14}-MT** M_{15}-MT M_{16}-MT	13,341 13,403 **13,467** **13,529** 13,588 13,653	13,340.4 13,402.9 **13,465.5** **13,528.0** 13,590.6 13,653.1	Cu_8-MT Cu_9-MT Cu_{10}-MT Cu_{11}-MT **Cu_{12}-MT** Cu_{13}-MT Cu_{14}-MT	13,153 13,215 13,277 13,340 **13,403** 13,465 13,526	13,152.7 13,215.3 13,277.8 13,340.4 **13,402.9** 13,465.5 13,528.0
	BgKNMT	1.9 Zn 0 Cd 9.9 Cu	M_{13}-MT M_{14}-MT **M_{15}-MT** **M_{16}-MT** M_{17}-MT M_{18}-MT	13,452 13,514 **13,577** **13,639** 13,701 13,762	13,451.4 13,513.9 **13,576.5** **13,639.0** 13,701.6 13,764.1	Cu_{12}-MT Cu_{13}-MT **Cu_{14}-MT** Cu_{15}-MT Cu_{16}-MT Cu_{17}-MT	13,388 13,450 **13,514** 13,575 13,637 13,697	13,388.8 13,451.4 **13,513.9** 13,576.5 13,639.0 13,701.6

[a] metal-to-peptide ratio calculated from S, Zn, Cd and Cu content (ICP-AES) data); [b] The stoichiometry of the metal-loaded complex was calculated from the mass difference between the holo- and the apo-peptides; Major species are highlighted in bold. M denotes mixtures of Zn and Cu; [c] Experimental molecular masses corresponding to the detected metal-complexes. The corresponding ESI-MS spectra are shown in Figures 2–4 and Figure 6; [d] Theoretical molecular masses corresponding to the metal-complexes.

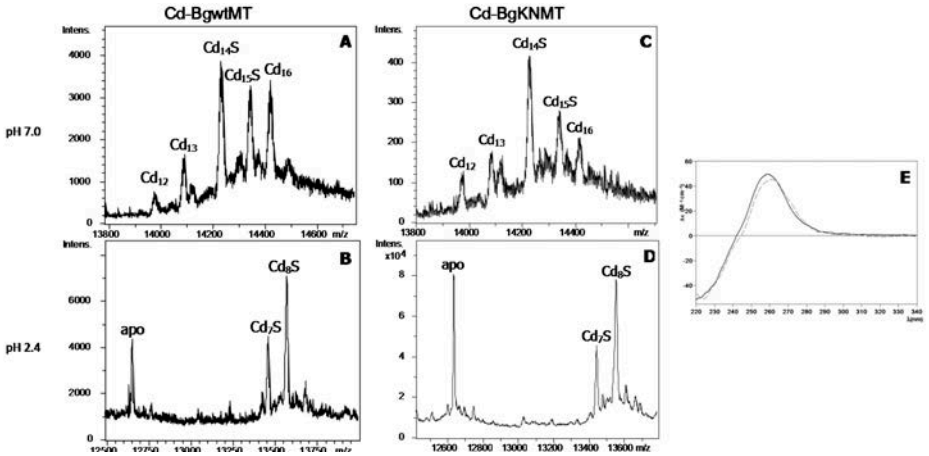


Figure 4. MS and CD spectra of Cd-BgMT: (**A**–**D**) Deconvoluted ESI-MS spectra of recombinant Cd-BgwtMT and Cd-BgKNMT, recorded at neutral (7.0) and acidic (2.4) pH; (**E**) CD spectra of each Cd(II) preparation: BgwtMT (solid line) and BgKNMT (dashed).

The ESI-MS analyses at neutral pH of the recombinant production of BgwtMT and BgKNMT in Cd-supplemented *E. coli* cultures (Figure 4, Table 1) revealed the formation of the same species in both preparations (with similar relative abundances), with the important presence of one major sulfide-containing species ($Cd_{14}S$-BgMT) accompanied by minor amounts of Cd_{15}-S, Cd_{16}-, Cd_{13}- and Cd_{12}-MT. Interestingly, the mass spectra of both samples recorded at pH 2.4 revealed the remaining presence of Cd_7S- and Cd_8S-BgMT complexes coexisting with the apo-form. In every case, the decrease of pH down to 0.9 was necessary to completely remove all Cd(II) initially bound to proteins (data not shown). The CD spectra of the Cd-BgwtMT and Cd-BgKNMT preparations (Figure 4) display the contribution of both (1) a Gaussian band centered at ca. 250 nm, corresponding to the Cd-$(SCys)_4$ chromophores and (2) a contribution of a further absorption at ca. 280 nm in agreement with the presence of Cd-S^{2-} chromophores in these samples. Furthermore, the congruence of the respective CD fingerprints confirmed equivalent folds of BgwtMT and BgKNMT when coordinating Cd(II) ions.

In order to further study the capabilities of the BgwtMT and BgKNMT proteins for Cd(II) binding, the recombinant Zn-BgwtMT and Zn-BgKNMT preparations were titrated with Cd(II) and the Zn/Cd metal displacement reactions were followed in parallel by ESI-MS and CD spectroscopy (Figure 5). CD signals from both titrations proceeded very similarly: the addition of Cd(II) provoked a red shift of the maxima until 12 Cd(II) equivalents were added; further additions led to a decrease of the intensity suggesting unfolding of the formed clusters. These data are in agreement with the measured mass data. They show the subsequent replacement of the initial Zn(II) through formation of heteronuclear Zn, Cd-species, until the system is saturated after the addition 12 Cd(II) equivalents. Even when adding more Cd(II) the main species were still Cd_{12}- and Cd_{13}-BgMT. Other species (Cd_{11}- and Cd_{14}-BgMT as well as $Cd_{11}Zn_1$- and $Cd_{12}Zn_1$-BgMT) were also present to minor amounts.

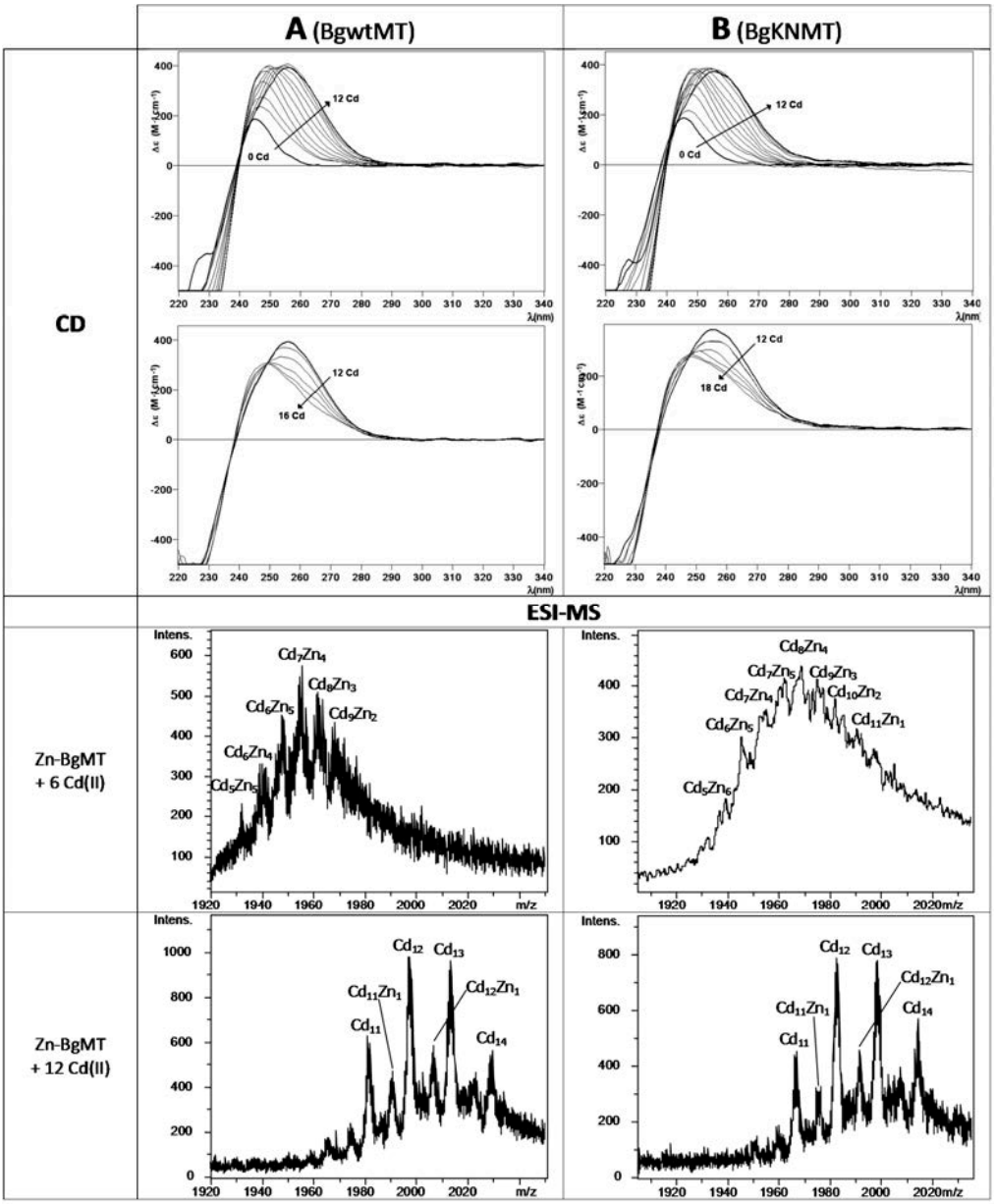


Figure 5. CD and ESI-MS spectra of Zn/Cd replacement of BgMT: Shown are spectra (at the +7 charge state) corresponding to the Zn(II)/Cd(II) replacement reaction of recombinant (**A**) Zn-BgwtMT and (**B**) Zn-BgKNMT, recorded at neutral pH. In both cases, a 5 μM solution of Zn-BgMT was titrated with up to 16–18 equivalents of $CdCl_2$ at neutral pH.

2.3. Monovalent -Cu(I)- Binding Features of BgwtMT and BgKNMT

The main differences between BgwtMT and BgKNMT were observed when producing them recombinantly in Cu-enriched media under standard or low aeration conditions (Figure 6). Although the two conventional types of Cu-supplemented productions [15] at standard (low intracellular copper content) or at low aeration (high intracellular copper content) conditions were assayed, several efforts to purify BgwtMT and BgKNMT from *E. coli* cultures grown at low oxygen conditions failed. Contrarily, both proteins could be isolated under normal aeration conditions, and were analyzed

by ESI-MS and CD for comparison. Interestingly, both preparations exhibited a similar pattern: No stable species were observed, neither under the low aeration conditions nor under normal aeration conditions—both yielded a mixture of heterometallic Zn, Cu-species (Figure 6, Table 1).

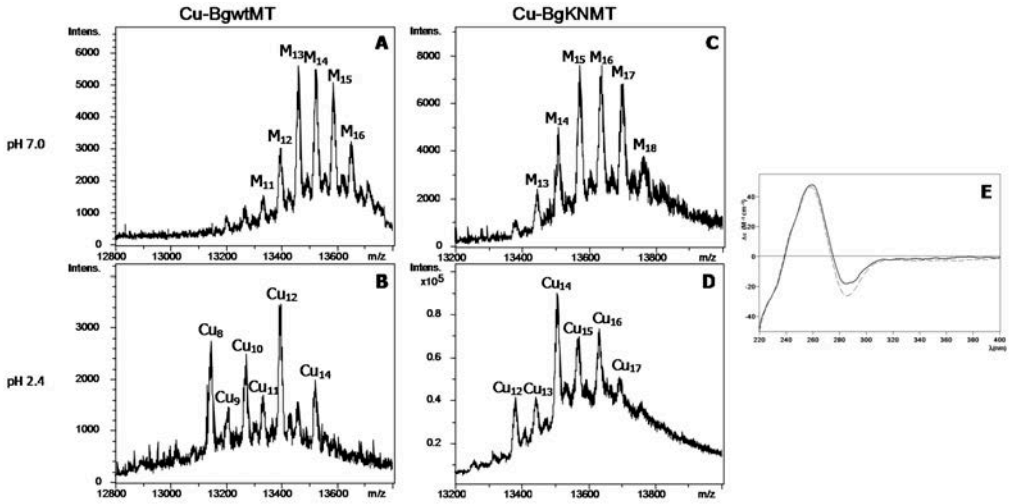


Figure 6. MS and CD spectra of Cu-BgMT: (**A**–**D**) Deconvoluted ESI-MS spectra of recombinant Cu-BgwtMT and Cu-BgKNMT, recorded at neutral (7.0) and acidic (2.4) pH. M denotes mixture of Zn and Cu (**E**) CD spectra of each Cu-preparation: BgwtMT (solid line) and BgKNMT (dashed).

However, while for BgwtMT the main species at neutral pH contains 13 metal ions (M_{13}-BgwtMT, with M = Zn + Cu), and Cu_8- and Cu_{12}-BgwtMT at acidic pH, the natural BgKNMT mutant protein can bind more metal. This was observed at both pH values, with $M_{15\text{-}16}$- and Cu_{14}-BgKNMT being the main peaks at neutral and acidic pH, respectively (Figure 6C,D). The CD spectra were again very similar for both preparations (Figure 6E) and showed the typical absorbances at ca. 260 and 280 nm corresponding to Cu-loaded MTs. Even if considering that the metal-speciation was slightly different in both preparations, the coexistence of several heterometallic (Zn,Cu-BgwtMT and Zn,Cu-BgKNMT) species probably resulted in a similar folding of both proteins about the metal ions, always dominated by the characteristic signals of the more abundant Cu(I). Zn just revealed its presence as a faint shoulder at ca. 240 nm.

2.4. Lacking Metal-Binding Specificity is an Eminent Feature of Wild-Type and Allelic Biomphalaria Glabrata MTs

The fact that no single metallated species (with Zn^{2+}, Cd^{2+} or even Cu^{+}) could be obtained, neither in the recombinant productions, nor in the in vitro Zn/Cd replacement, indicates the lack of a metal specificity for both studied proteins. The absence of metal-binding specificity is thus considered as an inherent feature of BgwtMT and BgKNMT. Interestingly, the natural mutation K/N of the wild-type protein did not significantly affect its binding of divalent metal ions, and both proteins (BgwtMT and BgKNMT) behaved very similarly when binding Zn(II) and Cd(II) (similar speciation and very similar folding) (Figure 4). Contrarily, there were remarkable differences between the wild-type and the mutant variant with respect to binding of Cu(I). In fact, the replacement of one single amino acid (K by N) altered the binding properties of the protein by significantly increasing its Cu-binding capability, leading to species with a higher Cu(I) content than in the wild-type protein (Figure 6). This suggests that the replacement of K by N may have increased the Cu-thionein character of the unspecific BgwtMT protein, in agreement with previously reported data concerning the CaCdCuMT isoform of the *Cantareus aspersus* MTs [16]. Overall, however, this did not significantly change the unspecific binding character of both BgMT proteins.

The present findings are in contrast to the high metal-binding specificity of MTs from other gastropods, e.g., in case of the well characterized CdMT and CuMT isoforms of terrestrial snails from the Helicid family, including the Roman snail (*Helix pomatia*) [1,3] and the garden snail (*Cantareus aspersus*) [17], all belonging to the gastropod phylum of *Stylommatophora*. It is also true for the Cd-specific MT of the marine periwinkle (*Littorina littorea*), that belongs to the gastropod phylum of *Caenogastropoda* [4]. Altogether, our data indicate that metal-binding properties of gastropod MTs may vary in a lineage-specific manner.

2.5. Missing Metal Specificity of Recombinant BgMT Proteins Goes Hand in Hand with Lacking Upregulation of the BgMT Gene

While many genes of metal-specific *MT*s from other gastropod species can specifically be upregulated by exposure to their cognate metal ions [2,6], this was clearly not the case for the *MT* gene of *Biomphalaria glabrata (BgwtMT)*, at least upon exposure to Cd^{2+}. While the metal itself was strongly accumulated in the midgut gland of Cd-exposed individuals with a concentration factor about 500 times above control levels (Figure 7A), there was no concomitant upregulation of the *MT* mRNA concentration at all (Figure 7B). Instead, it appeared that mRNA levels of the Bg*MT* in both untreated (control) and Cd-treated animals were already highly elevated, compared to transcription levels of other gastropod *MT* genes under control conditions. mRNA concentrations of the metal-specific *CdMT* genes of the terrestrial snails *Helix pomatia* and *Cantareus aspersus* under control conditions, for example [2,6], are about 8–15 times lower than the mRNA levels of the *BgMT* gene in untreated animals (Figure 7B). This strongly suggests that the *BgMT* gene and protein may function in a different manner than the metal-specific *MT* genes and proteins of *Helix pomatia* and *Cantareus aspersus* species. Considering the concomitance of the lack of metal specificity and the missing Cd-dependent inducibility of the *BgMT* gene, which is already highly expressed in controls, one obvious hypothesis is that the metal sequestration potential of BgMT may primarily be based on its metal replacement capacity rather than on a strong binding selectivity after metal-dependent induction and de novo synthesis. In fact, metal replacement reactions of the recombinant BgwtMT and BgKNMT proteins indicate that, at least in vitro, such replacement reactions (e.g., Cd^{2+} versus Zn^{2+}) may take place without significantly impairing the protein integrity and structure (Figure 5).

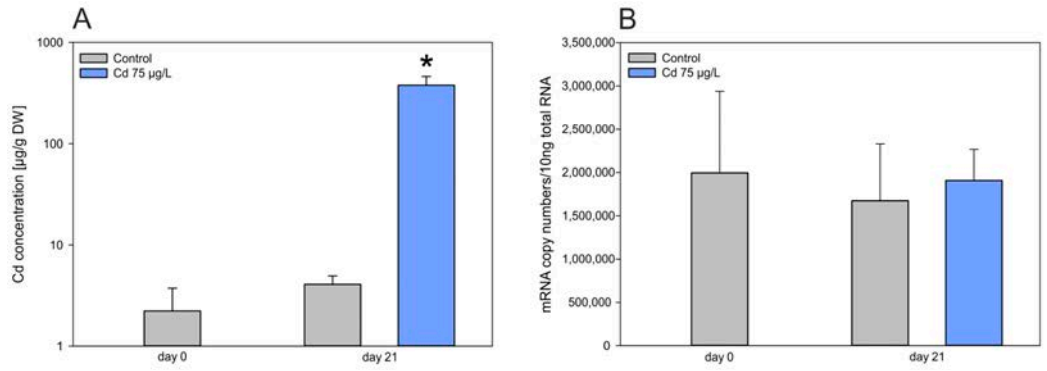


Figure 7. (**A**) Cd concentration in midgut gland tissue and qRT-PCR of mRNA after Cd exposure. (**A**) Cd concentration in midgut gland tissue of *Biomphalaria glabrata* on days 0 and day 21 of Cd exposure (75 µg/L); (* $p < 0.001$); the y-axis shows values in decadic logarithm; (**B**) Quantitative Real-Time PCR data of *BgwtMT* gene transcription on days 0 and 21 of Cd exposure (75 µg/L). Grey bars: control values; Blue bars: Cd exposure values.

2.6. The Biomphalaria glabrata MT in the Context of Other Hygrophila MTs: Deviant Primary Structures and Metal Stoichiometries

As indicated above, the unspecific metal-binding properties of recombinant BgMT and the lacking upregulation capacity of the corresponding gene grossly deviate from features of the rather metal-specific MT proteins and their responsive cognate genes of many other gastropod species. Indeed, *Biomphalaria glabrata* seems to share some of its deviating MT features with snail species belonging, just as *Biomphalaria glabrata* does, to the monophyletic gastropod clade of *Hygrophila* that comprises, according to the current phylogeny, all air-breathing ("pulmonate") freshwater snails of the super-phylum of *Panpulmonata* [18]. While all so-far studied *Hygrophila* species seem to possess MTs [7,19], their cognate *MT* genes may only weakly or even hardly at all respond to heavy metal stress upon exposure. This was shown for Cd-exposed *Physa acuta* [9], metal-stressed *Lymnaea stagnalis* [20] or Cd-stressed *Biomphalaria glabrata* (this study). Instead, phytochelatin (PC) synthesis was recently demonstrated as a responsive reaction of *Lymnaea stagnalis* to Cd stress [8]. This suggests that apart from MTs, PC metal complexes may be formed in these aquatic snails in response to metal stress. It is not clear, however, if and how the PC system in these animals may interfere with the MT pool.

Interestingly, a primary sequence alignment of the Cd-specific MT isoform of *Helix pomatia*, HpCdMT, with several MT sequences of *Hygrophila* species (Figure 8) shows that, in contrast to the former, the MT sequences of *Hygrophila* species (including *Physa acuta*, *Lymnaea stagnalis* and *Biomphalaria glabrata*), do not show a clear two-domain organization with a straight forward Cys:divalent metal-binding ratio of 9:3 per domain. Instead, the *Hygrophila* MT sequences seem to suffer from structural "degeneration" by deletion, truncation or extension of their primary sequences, often with deviations from the above-mentioned model of the 9:3 Cys:divalent metal ratio [21] (Figure 8). It is not known if and how all these structural deviations may influence the metal-binding behavior of the respective single proteins, but it may be assumed that overall, these structural "degenerations" from the classical gastropod MT model may contribute to impairment of their metal-specific binding properties.

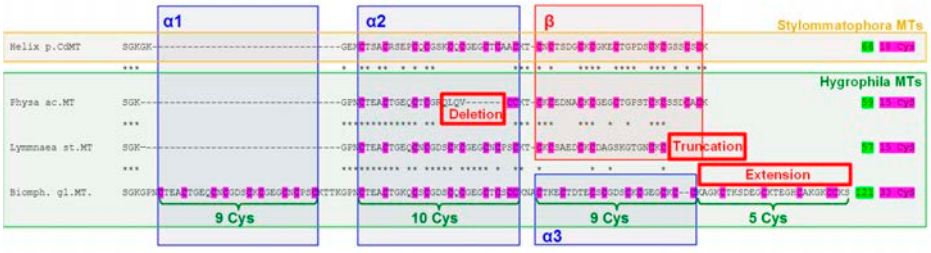


Figure 8. Sequence Alignment of sequences of various snail MTs: Alignment of *Helix pomatia* CdMT (Helix p.CdMT) (*Stylommatophora* MTs, **orange**-framed box) sequence with those of MTs of *Hygrophila* (*Hygrophila* MTs, **green**-framed box), including MTs of *Physa acuta* (Physa ac.MT), *Lymnaea stagnalis* MT (Lymnaea st.MT) and *Biomphalaria glabrata* MT (Biomph. gl.MT), showing the presumed domain organization with up to three α domains (α1, α2 and α3, **blue**-framed boxes) and one β domain (**red**-framed box). Conserved cysteine positions in the MT sequences are underlaid in pink. Also indicated in green letters are the numbers of Cys residues in every single domain of the *Biomphalaria gl.* MT, along with primary structure deviations (Deletion, Truncation, Extension) in *Hygrophila* MTs from the *Helix pomatia* CdMT model. Identical amino acid residues between the aligned sequences are indicated by stars. The respective amino acid chain lengths of the MT peptides are highlighted in **green** near their C-terminal ends. Also indicated next to the C-termini are the numbers of Cys residues in each MT sequence (underlaid in pink). GenBank Accession numbers of the shown sequences are as follows: *Helix pomatia* CdMT, ACN66299.1; *Physa acuta* MT, GU259686; *Lymnaea stagnalis* MT, KT253648; *Biomphalaria glabrata* MT KT697617.

3. Materials and Methods

3.1. Primary Structure of BglMT

The primary structure of the BglwtMT was elucidated by genome analysis within the *Biomphalaria glabrata* genome project (VectorBase, National Institutes of Health, Bethesda, MD, USA). The BglKNMT was obtained by screening for allelic variations and both sequences were experimentally verified by sequencing PCR-amplified and cloned individuals [13]. The sequences were submitted to the GenBank, and are available under the accession numbers KT697617 (BgwtMT) and KY963493 (BgKNMT).

3.2. Cloning and Heterologous Expression of Biomphalaria glabrata BgwtMT and BgKNMT

Synthetic cDNAs for the two allelic variants of the predicted *Biomphalaria glabarata* MTs were provided by Integrated DNA Technologies Company (Coralville, IA, USA). *Bam*HI and *Xho*I restriction sites and 6–7 additional 5′-nucleotides were added to the BgMT cDNA ends to facilitate the cloning processes. The synthetic cDNAs were PCR amplified with specific primers 5′-TTTTATT*GGATCC*ATGAGTGGCAAAG-3′ (forward) and 5′-ATTTTT*CTCGAG*TCAACTCTTAC-3′ (reverse), using Expand High Fidelity PCR system® (Roche, Penzberg, Upper Bavaria, Germany). A 25-cycle amplification reaction was performed under the following conditions: 30 s at 94 °C, 30 s at 55 °C, and 45 s at 72 °C, in a 25 µL PCR mixture containing 25 ng of template DNA, 0.01 mM of each primer, and 0.0125 mM of each dNTP. The amplified products were analyzed on a 1% agarose gel stained with Gel red (Biotum Inc., Bay Area, CA, USA). The *BgMT* cDNAs were digested with *Bam*HI and *Xho*I enzymes, cloned into a *Bam*HI-*Xho*I digested pGEX-4T-1 vector (GE Healthcare, Chicago, IL, USA) with the DNA Ligation Kit 2.1® (Takara Bio Inc., Shimogyo-ku, Kyoto, Japan), and transformed into *E. coli* Dh5α strain. Plasmid DNA was purified from bacteria using the GeneElute™ Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA), screened for insert presence by digestion with ScaI enzyme, and sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Waltham, MA, USA) in an ABIPRISM 310 automatic sequencer (Applied Biosystems). DNA from each recombinant *BgMT*-pGEX plasmid was used to transform *E. coli* BL21 strain, a protease deficient strain used for heterologous protein expression.

The recombinant expression of *BgMTs* was assayed by growing 3 mL of LB-25 mg/mL ampicillin medium inoculated with BgMT-producing *E. coli* BL21 strains. After growth overnight at 37 °C and 250 rpm, 0.3 mL of the culture was used to inoculate 3 mL of fresh medium and new cultures were grown for 2 h. The expression of the *BgMT*s was induced with 100 µM (final concentration) of isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h. After 30 min of induction, cultures were supplemented with 500 µM of $CuSO_4$, 300 µM of $CdCl_2$ or 300 µM of $ZnCl_2$ (final concentrations), and allowed to grow for further 2.5 h for the synthesis of the respective metal complexes. Cells were harvested by centrifugation for 1 min at 13,000 rpm, and bacterial pellets were suspended by vortexing in 150 µL of Phosphate Buffered Saline (PBS1X) (140 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.8 mM Kh_2PO_4). Suspended cells were sonicated (Sonifier® ultrasonic cell disruptor, Ferguson, MO, USA) at voltage 2 with 9 pulses of 0.6 s, and then centrifuged for 10 min at 12,500 rpm at 4 °C. Supernatant was recovered, and protein content was measured by a Bradford assay (Bio-Rad, Hercules, CA, USA) in a Nanoquant infinite M200 microplate reader (Infinite M200 TECAN). Expression of both *BgMTs* was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) on 12.5% gels stained with Coomassie Blue.

3.3. Synthesis and Purification of Recombinant BgwtMT and BgKNMT

The BgMT metal complexes were biosynthesized in 5 L Erlenmeyer cultures of the corresponding transformed *E. coli* Bl21 cells grown in LB medium containing 100 mg/mL ampicillin supplemented with $ZnCl_2$ (300 µM), $CdCl_2$ (300 µM) or $CuSO_4$ (500 µM). For copper cultures, two different aeration conditions were applied (250 rpm—standard aeration and 150 rpm—low aeration), of which only the standard aeration culture was successful. BgMT synthesis was induced with

isopropyl-1-thio-β-D-galactopyranoside (IPTG) at a final concentration of 100 μM for 30 min. After adding the metal solution followed by a 2.5 h induction, cells were harvested by centrifugation. The protein containing pellets were re-suspended in 4 °C PBS 1× supplied with 0.5% (*v*/*v*) ß-mercaptoethanol, sonicated and centrifuged at 12,000× *g* for 40 min at 4 °C. Glutathione S-transferase (GST)-BgMT polypeptides were purified using batch affinity chromatography with glutathione sepharose (GE Healthcare, Buckinghamshire, UK). After incubation for one hour at room temperature applying constant agitation, the mix was washed in PBS three times. Argon was bubbled through all the washing steps following cell disruption to avert oxidation of the metal-BgMT complexes. Thrombin (GE Healthcare, Buckinghamshire, UK) (1 μL/mg of fusion protein) was added to the mixture and digestion was carried out overnight at 16 °C. This enabled separation of the GST fragment, which remained bound to the gel matrix, from the fusion protein. The eluted solution was applied to Centriprep low Concentrators (Amicon, Millipore, MA, USA) with a cut-off of 3 kDa and subsequently fractionated by means of fast protein liquid chromatography (FPLC), using a Superdex-75 column (GE Healthcare) equilibrated with 50 mM Tris-HCl, pH 7.0, and run at 1 mL/min. Fractions were collected and the protein content was analyzed by its absorbance at 254 nm. Samples containing BgMT were pooled and stored at −80 °C.

3.4. Zn(II)/Cd(II) Replacement Reactions in the Zn(II)-BgMT Proteins

Reactions of Zn(II) displacement by Cd(II) on the recombinant Zn-BgMT preparations were performed as described elsewhere [22]. This allowed the formation of the Cd-"in vitro complexes" by addition of several molar equivalents of Cd^{2+} from a standard solution to the corresponding Zn-BgMT preparation. These experiments were performed at constant pH 7.0 without the addition of any extra buffers, and under argon atmosphere.

3.5. Spectroscopic Analyses (ICP-AES, UV-Vis and CD) of the Metal Complexes Formed by the BgwtMT and BgKNMT Proteins

Determination of the sulfur and metal content of all the metal-BgMT samples was performed by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) in a Polyscan 61E (Thermo Jarrel Ash, Franklin, MA, USA) spectrometer by measuring S at 182.040 nm, Zn at 213.856 nm, Cd at 228.802, and Cu at 324.803 nm. This allowed determination of the protein concentration by considering that all S atoms were provided by the BgMT peptides. A Jasco spectropolarimeter (Model J-715, JASCO, Groß-Umstadt, Germany) interfaced to a computer (J700 software, JASCO, Groβ-Umstadt, Germany) was used for circular dichroism (CD) measurements. The electronic absorption measurements were performed in an HP-8453 Diode array UV-vis spectrophotometer (GIM, Ramsey, MN, USA) in 1-cm capped quartz cuvettes. In all spectroscopic measurements the dilution effects were corrected and processed using the GRAMS 32 software (Thermo Fisher Scientific, Waltham, MA, USA).

3.6. Electrospray Ionization Time-of-Flight Mass Spectrometry (ESI-TOF MS) of the Metal Complexes Obtained from the BgMT proteins

The M_W determinations by electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) were carried out in a Micro TOF-Q instrument (Bruker Daltonics, Bremen, Germany) interfaced with a Series 1200 HPLC Agilent pump and equipped with an autosampler, all of which were controlled by the Compass Software. ESI-L Low Concentration Tuning Mix (Agilent Technologies, Santa Clara, CA, USA) was used for calibration. The samples were analyzed using a 5:95 mixture of acetonitrile:ammonium acetate (15 mM, pH 7.0) or a 5:95 acetonitrile:formic acid mixture (at pH 2.4) as carrier buffers. The neutral pH buffer allowed to detect all the metallated species while the acidic conditions provoke the release of Zn(II) and Cd(II), but keep the Cu(I) ions bound to the proteins.

Experimental mass values were calculated as described in [23] and the error associated with the measurements resulted to be always smaller than 0.1%.

3.7. Experimental Set-Up for BgMT Gene Induction Studies

Individuals of *Biomphalaria glabrata* originated from a laboratory-grown culture at the Institute of Zoology in Innsbruck, where the snails were kept in freshwater aquarium tanks at 25 °C with a 12:12 h photoperiod. Snails were fed *ad libitum* with commercially available lettuce (*Lactuca sativa*) every third day.

Prior to the experiment, forty individuals of *Biomphalaria glabrata* were acclimatized for two weeks in reconstituted water (KCl 18 mg/L, $MgSO_4$ 190 mg/L, $NaHCO_3$ 98.5 mg/L, $CaCl_2$ 450 mg/L and NaCl 430 mg/L in milliQ water). Afterwards, snails were separated into different tanks and a Cd exposure regime was applied by adding $CdCl_2$ to a final Cd concentration of 75 µg/L. A control group of 20 individuals was kept in Cd-free reconstituted water as a reference. Resulting Cd concentrations in the water were as follows (mean ± standard deviation, n = 5): Control, 0.24 ± 0.14 µg/L; Cd exposure, 63 ± 7.6 µg/L. Throughout the experiment the snails were fed with lettuce *ad libitum*. Four snails of each group were sampled on day 0 and 21. All sampled individuals were dissected and the midgut gland tissue was used for RNA isolation and tissue Cd analysis as described below.

3.8. mRNA Isolation, Reverse Transcription and BgMT qRT-PCR

Biomphalaria glabrata individuals were dissected on an ice-cooled stainless steel plate and ~10 mg (fresh weight) of midgut gland tissue was used for RNA isolation. The remaining midgut gland tissue of each animal was processed further for Cd analysis as described below. Tissue samples were homogenized (Precellys, Bertin instruments, France) and total RNA was isolated with the RNeasy®Plant Mini Kit (Qiagen) applying on-column DNase 1 digestion (Qiagen). Quantification was achieved by means of the RiboGreen®RNA Quantification Kit from Molecular Probes (Invitrogen, Karlsruhe, Germany) on a VICTOR™X4 2030 Multilabel Reader (PerkinElmer, Waltham, MA, USA). 250 ng RNA were subjected to cDNA synthesis (Superscript® IV Reverse Transcriptase, Invitrogen, Life Technologies, Waltham, MA, USA) in a 20 µL approach for subsequent Real-time Detection PCR.

Quantitative Real-time Detection PCR of *BgwtMT* cDNA was performed on a Quant studio 3 (Applied Biosystems, Thermo Fisher Scientific) using Power SYBR Green (Applied Biosystems). RT Primers were designed using the Primer Express 3.0 software (Applied Biosystems) and optimal primer concentrations were assessed with a primer-matrix followed by dissociation curves. The *BgMT* transcript with the defined amplicon length (107 bp) was amplified with the following primers and concentrations: *BgMT* sense, 900 nM; 5′-GCACTGACACAGAATGCAGTTG-3′ and *BgMT* antisense, 900 nM; 5′-TTTGCACCCTTCATCTGACTTAGT-3′ applying the following protocol of 40 cycles: denaturation at 95 °C for 15 s, annealing and extension combined at 60 °C for 60 s. Subsequently calibration curves from amplicons were generated to determine C_q values for copy number analysis (PCR efficiency ~96%) using the Thermo Fisher Cloud Software, Version 1.0 (Life Technologies Corporation, Waltham, MA, USA). The 10-µL PCR reaction contained 1 µL of cDNA and 1× Power SYBR Green PCR master mix, 1× U-BSA and sense and antisense primer.

3.9. Metal Analysis

Cd concentrations in the midgut gland tissues and the medium were assessed by flame atomic absorption spectrophotometry. Dry weight was determined after oven-drying the samples at 65 °C. Dry samples were pressure-digested in 2 mL tubes (Eppendorf, Hamburg, Germany) with a 1:1 mixture of nitric acid (Suprapure, Merck, Darmstadt, Germany) and deionized water in an aluminum oven covered with a heated lid at 69 °C until a clear solution was obtained. All samples were diluted to 2 mL with deionized water and Cd concentrations measured in the flame of an atomic absorption spectrophotometer (model Z-8200, Hitachi, Tokyo, Japan). Calibration was achieved using standard metal solutions in 1% nitric acid. Accuracy of metal measurements of the midgut gland was verified with certified standard reference material (TORT-2, Lobster Hepatopancreas Reference Material for Trace Metals; National Research Council Canada).

3.10. Statistical Methods

Data of qRT PCR and metal analysis were evaluated statistically by means of Sigma Plot 12.5. For normal-distributed data, the t-test was applied. For data failing equal distribution the Mann-Whitney rank sum test was used. Statistical significance was set at $p \leq 0.05$.

4. Conclusions

The metal-binding capabilities of BgwtMT and its natural mutant BgKNMT, as studied in this work, revealed to be very similar in most of the investigated aspects. This is not of much surprise, as they differ by only one single mutation of a non-coordinating amino acid residue. To summarize, the two proteins share the following features: (1) They lack a clear metal-binding preference for any of the three metal ions assayed—Zn(II), Cd(II) or Cu(I). To this degree, this feature is clearly unprecedented in the world of *Gastropoda* MTs, even if some other MT isoforms of this sub-family are also characterized by a relatively small, but when compared to BgMT still significant, degree of metal-binding specificity [16,24]. (2) Consequently, the two MT variants of *Biomphalaria glabrata* (BgwtMT and BgKNMT) presented, in all the cases, mixtures of several differently metallated species. Accordingly we suspect that none of the uniquely metallated species presents a system significantly lower in energy compared to mixed-metal species. Hence, the metal-BgMT species found are almost equivalent in their Zn- and Cd-BgMT preparations. (3) The CD profiles of both protein variants are very similar when complexed to the same metal ion. (4) There are, however, slight differences in copper-binding between BgwtMT and BgKNMT: the Cu-BgKNMT preparations contain higher nuclearity (M = Zn + Cu) and higher copper content than the Cu-BgwtMT samples. (5) These observations can be explained if we consider the previously reported role of Lys and Asn residues for metal-binding preferences of snail MTs [16]: the replacement of a Lys residue (highly present in the MT isoforms with higher Zn- and/or Cd-thionein character) by an Asn residue (abundant in the so-called Cu-thioneins) reduces the Cd specificity in favor of Cu specificity. This increase of the Cu-thionein character is however rather limited, which we attribute to the high number of amino acids and Cys content (33 Cys residues among 126 amino acids) that reduces the influence of a single amino acid mutation. (6) The results from qRT-PCR do not show significant *BgMT* gene upregulation upon induction by Cd despite a strong accumulation of Cd in the midgut gland tissue. (7) This agrees well with the ESI-MS results that fail to show a clear metal-binding preference for BgMTs, while at the same time, indicate a high potential for metal replacement. This suggests that BgMT(s) in living cells may be more important for metal exchange (e.g., replacement of Zn^{2+} by Cd^{2+}) through a constitutively abundant form, rather than for metal sequestration by an MT species that is highly upregulated in presence of a specific metal. (8) A comparison of the *Biomphalaria glabrata* MTs with the CdMT isoform of *Helix pomatia* and MTs of the gastropod clade of *Hygrophila* (to which *Biomphalaria glabrata* belongs) shows several deviations in primary structure from the classical domain organization of metal-specific gastropod MTs. These include primary sequence aberrations such as deletions, truncations and chain extensions, along with altering combinations of α and β domains. (9) Perhaps as a consequence, different ratios of Cys:divalent metal ions seem to be a common feature of *Hygrophila* MTs compared to the CdMT of *Helix pomatia*. (10) Thus, it is hypothesized that these structural "degenerations" of *Hygrophila* MTs from the classical gastropod MT model may contribute to the impairment of their metal-binding specificity and response properties. At the same time, this may clear the way for the activation of alternative detoxification strategies such as metal complexation by low molecular weight ligands.

Acknowledgments: This work was funded by the Tiroler Wissenschaftsfonds (TWF) (Project ref. 214457) granted to Michael Niederwanger, the Austrian Science foundation (FWF) and the Swiss National Science Foundation (Project ref. I 1482-N28 DACH) granted to Reinhard Dallinger and Oliver Zerbe. Authors from both Barcelona universities are members of the 2014SGR-423 Grup de Recerca de la Generalitat de Catalunya, and they are recipients of MINECO-FEDER grants (BIO2015-67358-C2-1-P to SA and RA, and BIO2015-67358-C2-2-P to MC). We thank the Centres Científics i Tecnològics (CCiT) de la Universitat de Barcelona (DNA sequencing) and the

Servei d'Anàlisi Química (SAQ) de la Universitat Autònoma de Barcelona (ICP-AES, CD, UV-vis, ESI-MS) for allocating instrument time. Experimental support of Sebastian Artime and Mireia Mata is greatly appreciated.

Author Contributions: Reinhard Dallinger was responsible for the project coordination. Michael Niederwanger, Reinhard Dallinger, Oliver Zerbe, Mercè Capdevila and Sílvia Atrian conceived and designed experiments. Mercè Capdevila, Òscar Palacios, Ricard Albalat and Sílvia Atrian analyzed the data and discussed the experimental results. Sara Calatayud and Michael Niederwanger performed the cloning and recombinant synthesis of the analyzed proteins and Òscar Palacios performed their ESI-MS and CD characterization. Òscar Palacios, Mercè Capdevila, Michael Niederwanger, Oliver Zerbe and Reinhard Dallinger were primarily responsible for writing the manuscript. All authors—except Silvia Atrian that passed away in December the 5th, 2016 and is sorely missed—edited and approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dallinger, R.; Berger, B.; Hunziker, P.; Kägi, H.R.J. Metallothionein in snail Cd and Cu metabolism. *Nature* **1997**, *143*, 831–833. [CrossRef]
2. Palacios, Ò.; Pagani, A.; Pérez-Rafael, S.; Egg, M.; Höckner, M.; Brandstätter, A.; Capdevila, M.; Atrian, S.; Dallinger, R. Shaping mechanisms of metal specificity in a family of metazoan metallothioneins: Evolutionary differentiation of mollusc metallothioneins. *BMC Biol.* **2011**, *9*, 4. [CrossRef] [PubMed]
3. Palacios, Ò.; Pérez-Rafael, S.; Pagani, A.; Dallinger, R.; Atrian, S.; Capdevila, M. Cognate and noncognate metal ion coordination in metal-specific metallothioneins: The Helix pomatia system as a model. *J. Biol. Inorg. Chem.* **2014**, *19*, 923–935. [CrossRef] [PubMed]
4. Baumann, C.; Beil, A.; Jurt, S.; Niederwanger, M.; Palacios, Ò.; Capdevila, M.; Atrian, S.; Dallinger, R.; Zerbe, O. Structural adaptation of a protein to increased metal stress: NMR structure of a marine snail metallothionein with an additional domain. *Angew. Chem.* **2017**, *56*, 4617–4622. [CrossRef] [PubMed]
5. English, T.E.; Storey, K.B. Freezing and anoxia stresses induce expression of metallothionein in the foot muscle and hepatopancreas of the marine gastropod *Littorina littorea*. *J. Exp. Biol.* **2003**, *206*, 2517–2524. [CrossRef] [PubMed]
6. Pedrini-Martha, V.; Niederwanger, M.; Kopp, R.; Schnegg, R.; Dallinger, R. Physiological, diurnal and stress-related variability of cadmium-metallothionein gene expression in land snails. *PLoS ONE* **2016**, *11*, e0150442. [CrossRef] [PubMed]
7. Gnatyshyna, L.L.; Fal'fushinskaya, G.I.; Golubev, O.P.; Dallinger, R.; Stoliar, O.B. Role of metallothioneins in adaptation of *Lymnaea stagnalis* (Mollusca: Pulmonata) to environment pollution. *Hydrobiol. J.* **2011**, *47*, 56–66. [CrossRef]
8. Gonçalves, S.F.; Davies, S.K.; Bennett, M.; Raab, A.; Feldmann, J.; Kille, P.; Loureiro, S.; Spurgeon, D.J.; Bundy, J.G. Sub-lethal Cadmium exposure increases phytochelatin concentrations in the aquatic snail Lymnaea stagnalis. *Sci. Total Environ.* **2016**, *568*, 1054–1058. [CrossRef]
9. Martínez-Paz, P.; Morales, M.; Sánchez-Argüello, P.; Morcillo, G.; Martínez-Guitarte, J.L. Cadmium in vivo exposure alters stress response and endocrine-related genes in the freshwater snail *Physa acuta*. New biomarker genes in a new model organism. *Environ. Pollut.* **2017**, *220*, 1488–1497. [CrossRef] [PubMed]
10. DeJong, R.J.; Morgan, J.A.T.; Paraense, W.L.; Pointier, J.P.; Amarista, M.; Ayeh-Kumi, P.F.K.; Babiker, A.; Barbosa, C.S.; Brémond, P.; Canese, A.P.; et al. Evolutionary relationships and biogeography of Biomphalaria (*Gastropoda*: Planorbidae) with implications regarding its role as host of the human bloodfluke, Schistosoma mansoni. *Mol. Biol. Evol.* **2001**, *18*, 2225–2239. [CrossRef] [PubMed]
11. Morgan, J.A.T.; Dejong, R.J.; Snyder, S.D.; Mkoji, G.M.; Loker, E.S. *Schistosoma mansoni* and *Biomphalaria*: Past history and future trends. *Parasitology* **2001**, *123*, S211–S228. [CrossRef] [PubMed]
12. Gryseels, B.; Polman, K.; Clerinx, J.; Kestens, L. Human schistosomiasis. *Lancet* **2006**, *368*, 1106–1118. [CrossRef]
13. Niederwanger, M.; Dvorak, M.; Schnegg, R.; Bacher, K.; Bidoli, M.; Dallinger, R. The metallothionein gene of *Biomphalaria glabrata*: A complex structure with unexpected response patterns. Manuscript in preparation.
14. Espart, A.; Marín, M.; Gil-Moreno, S.; Palacios, Ò.; Amaro, F.; Martín-González, A.; Gutiérrez, J.C.; Capdevila, M.; Atrian, S. Hints for metal-preference protein sequence determinants: Different metal binding features of the five tetrahymena thermophila metallothioneins. *Int. J. Biol. Sci.* **2015**, *11*, 456–471. [CrossRef] [PubMed]

15. Pagani, A.; Villarreal, L.; Capdevila, M.; Atrian, S. The *Saccharomyces cerevisiae* Crs5 Metallothionein metal-binding abilities and its role in the response to zinc overload. *Mol. Microbiol.* **2007**, *63*, 256–269. [CrossRef] [PubMed]
16. Pérez-Rafael, S.; Monteiro, F.; Dallinger, R.; Atrian, S.; Palacios, Ò.; Capdevila, M. Cantareus aspersus metallothionein metal binding abilities: The unspecific CaCd/CuMT isoform provides hints about the metal preference determinants in metallothioneins. *Biochim. Biophys. Acta Proteins Proteom.* **2014**, *1844*, 1694–1707. [CrossRef] [PubMed]
17. Höckner, M.; Stefanon, K.; de Vaufleury, A.; Monteiro, F.; Pérez-Rafael, S.; Palacios, Ò.; Capdevila, M.; Atrian, S.; Dallinger, R. Physiological relevance and contribution to metal balance of specific and non-specific metallothionein isoforms in the garden snail, *Cantareus aspersus*. *BioMetals* **2011**, *24*, 1079–1092. [CrossRef] [PubMed]
18. Jörger, K.M.; Stöger, I.; Kano, Y.; Fukuda, H.; Knebelsberger, T.; Schrödl, M. On the origin of *Acochlidia* and other enigmatic euthyneuran gastropods, with implications for the systematics of *Heterobranchia*. *BMC Evol. Biol.* **2010**, *10*, 323. [CrossRef] [PubMed]
19. Jo, Y.H.; Baek, M.K.; Kang, S.W.; Lee, J.B.; In-Seon, B.; Choi, S.-H.; Chae, S.-H.; Kang, J.-H.; Han, Y.S.; Park, H.-S.; et al. Molecular cloning and expression pattern of metallothionein gene from left-handed shell, *Physa acuta*. *Korean J. Malacol.* **2009**, *25*, 223–230.
20. Niederwanger, M.; Dallinger, R.; Stoliar, O. University of Innsbruck, Austria. Unpublished work, 2013.
21. Dallinger, R.; Wang, Y.; Berger, B.; Mackay, E.A.; Kägi, J.H.R. Spectroscopic characterization of metallothionein from the terrestrial snail, *Helix pomatia*. *Eur. J. Biochem.* **2001**, *268*, 4126–4133. [CrossRef] [PubMed]
22. Capdevila, M.; Cols, N.; Romero-isart, N.; Gonzàlez-Duarte, R.; Atrian, S.; Gonzàlez-Duarte, P. Recombinant synthesis of mouse Zn_3-β and Zn_4-α metallothionein 1 domains and characterization of their cadmium (II) binding capacity. *Cell. Mol. Life Sci.* **1997**, *53*, 681–688. [CrossRef] [PubMed]
23. Pérez-Rafael, S.; Kurz, A.; Guirola, M.; Capdevila, M.; Palacios, Ò.; Atrian, S. Is MtnE, the fifth *Drosophila* metallothionein, functionally distinct from the other members of this polymorphic protein family? *Metallomics* **2012**, *4*, 342. [CrossRef] [PubMed]
24. Pérez-Rafael, S.; Mezger, A.; Lieb, B.; Dallinger, R.; Capdevila, M.; Palacios, Ò.; Atrian, S. The metal binding abilities of *Megathura crenulata* metallothionein (McMT) in the frame of Gastropoda MTs. *J. Inorg. Biochem.* **2012**, *108*, 84–90. [CrossRef] [PubMed]

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ANNEX III:

Metal binding functions of metallothioneins in the slug Arion vulgaris differ from metal-specific isoforms of terrestrial snails

Metallomics, Volume 10, Issue 11, November 2018, Pages 1638–1654

DOI: 10.1039/c8mt00215k

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

Metallomics

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/c8mt00215k

Metal binding functions of metallothioneins in the slug *Arion vulgaris* differ from metal-specific isoforms of terrestrial snails†

Martin Dvorak,[a] Reinhard Lackner,[a] Michael Niederwanger,[a] Claire Rotondo,[a] Raimund Schnegg,[a] Peter Ladurner,[a] Veronika Pedrini-Martha,[a] Willi Salvenmoser,[a] Leopold Kremser,[b] Herbert Lindner,[b] Mario García-Risco,[c] Sara Calatayud,[d] Ricard Albalat,[d] Oscar Palacios,[c] Merce Capdevila[c] and Reinhard Dallinger*[a]

Received 10th August 2018,
Accepted 27th September 2018

DOI: 10.1039/c8mt00215k

rsc.li/metallomics

Arion vulgaris is a land-living European slug belonging to the gastropod clade of *Stylommatophora*. The species is known as an efficient pest organism in vegetable gardening and horticulture, which may in part be the consequence of its genetically based innate immunity, along with its high ability to withstand toxic metal stress by intracellular detoxification. Like many species of terrestrial snails, slugs possess a distinct capacity for Cd accumulation in their midgut gland, where the metal is stored and inactivated, conferring to these animals an increased metal tolerance. Although midgut gland Cd fractions in slugs have been shown to be variably allocated between different metal-binding protein pools, depending on the level of environmental metal contamination, a true metallothionein (MT) was so far never characterized from slugs. Instead, the Cd binding proteins identified so far were described as Metallothionein-like proteins (MTLPs). In the present study, the slug *A. vulgaris* was used as a model organism, in order to verify the presence of true MTs in experimentally metal-exposed slugs. We wanted to find out if these suggested slug MTs have similar metal binding properties and metal-selective features like those previously reported from helicid snails. To this aim, two *MT* isoform genes (*AvMT1* and *AvMT2*) were characterized from midgut gland extracts and localized in the cells of this tissue. The AvMT1 and AvMT2 proteins were purified and partially sequenced, and their metal-binding features analysed after recombinant expression. Eventually, we wanted to understand if and by how much the metal binding features of the two MT isoforms of *A. vulgaris* may be related, owing to their reciprocal amino acid sequence similarities, to the binding properties of metal-specific MTs from terrestrial snails.

Significance to metallomics

The present study aims at elucidating the handling of environmentally significant trace metals (cadmium, copper) by the metallothionein system of a terrestrial slug (*Arion vulgaris*). This species is an important pest organism, showing a high degree of tolerance against these metals by accumulating them in its midgut gland. In the present study, we focus on the molecular, protein and phylogenetic levels of *Arion vulgaris* metallothioneins and their metal binding features, in order to understand their physiological significance for the metal metabolism in the slug organism *in vivo*. The introduction of an evolutionary comprehension of the slug s metallothionein system within the monophyletic clade of Stylommatophora (to which *Arion vulgaris* belongs) may promote a better understanding of the differences in the metal binding behavior of metallothioneins between terrestrial slugs and snails.

Introduction

A. vulgaris is a land-living European slug belonging to the gastropod clade of *Stylommatophora*. It is comprised in the clade of *Arionoidea* which, together with the sister clade of *Limacoidea*, forms the terrestrial group of slugs, and is thus considered as a distant relative of the edible and shell bearing terrestrial gastropods like the Roman snail (*Helix pomatia*) and

[a] Institute of Zoology and Center of Molecular Biosciences, University of Innsbruck, Technikerstr 25, A-6020 Innsbruck, Austria. E-mail: reinhard.dallinger@uibk.ac.at
[b] Innsbruck Medical University, Biocenter, Division of Clinical Biochemistry, Innrain 80, A-6020 Innsbruck, Austria
[c] Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Cerdanyola del Vallès, Barcelona, Spain
[d] Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, E-08028 Barcelona, Spain

† Electronic supplementary information (ESI) available: Table S1. Characterization by ESI-MS of the recombinant productions of AvMT1 in the corresponding culture media, showing the kind of metal exposure, the obtained metal species, as well as respective experimental and theoretical masses. See DOI: 10.1039/c8mt00215k

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

the garden snail (*Cantareus aspersus*, syn. *Helix aspersa*).[1] Some authors consider *A. vulgaris* as an invasive species of uncertain geographical origin that is actually spreading across European countries,[2,3] whereas other studies suggest that this species is with high probability native to Central Europe.[4] One of the reasons for these contradicting views may be the close relationship of *A. vulgaris* to other European Arionid species such as *Arion ater* and *Arion rufus*, with which *A. vulgaris* forms a species cluster through introgressive hybridization.[5,6] However, *A. vulgaris* is an extremely efficient pest organism in vegetable gardening and horticulture.[7,8] The recent invasion success of this species may rely on its extensive genetic repertoire of novel proteins with an innate immunity which are thought to protect the slug efficiently against parasite infections.[9] An additional adaptive advantage of *A. vulgaris* may be its capacity to withstand toxic metal stress by intracellular detoxification within the midgut gland,[10,11] a feature shared with several other terrestrial gastropods.[12,13]

The uptake of Cd by terrestrial gastropods normally leads to an activation of genes of the metallothionein (MT) family.[14,15] MTs are mostly low-molecular-weight, cysteine-rich proteins involved in the binding, sequestration, regulation, and detoxification of Cd^{2+} and other transition metal ions such as Zn^{2+}, Hg^{2+}, Cu^{+} and Ag^{+};[16] their encoding genes can be upregulated by metal ions.[17] In addition, certain organic chemicals, oxidative stress and exposure to ionizing radiation can also induce MT synthesis.[18,19] The metal ions are specifically bound to the MTs by the sulphur atoms of Cys residues, forming characteristic metal thiolate clusters.[20] Gastropods express MTs in a tissue-specific manner, with MT isoforms often differing from each other by the position of few amino acids,[14] but affecting the affinity to different metal ion species.[15] In the helicid land snails *H. pomatia* and *C. aspersus*, for example, three different MT isoforms have been discovered.[21,22] A Cd-specific isoform is devoted to the detoxification of Cd^{2+};[23] a Cu-selective isoform is responsible for homeostatic Cu regulation;[24] and a metal-unspecific isoform of uncertain but probably minor importance is expressed only at very low levels and can, therefore, hardly be detected at all.[21]

The situation may be different in land slugs of the *Stylommatophora* clade. In contrast to their shell-bearing relatives, land slugs have lost or largely reduced their calcareous shell during evolution. Like many species of terrestrial snails, slugs possess a distinct capacity for Cd accumulation,[25] particularly in their midgut gland.[26 28] Cd accumulation in these animals is accompanied by an enhanced storage capacity for this metal and an increased metal tolerance.[29] It was suggested that most of the accumulated Cd may be bound to metal-binding proteins designated as ``metallothionein-like proteins (MTLPs)[10,11] or even as MTs.[30] Interestingly, Cd concentrations in the midgut gland of land slugs seem to be lower than those detected in terrestrial snails.[27] In addition, high individual and seasonal variabilities of metal concentrations in the midgut gland of slugs were observed,[25,31] sometimes to such an extent as to compromise the value of slugs as biological indicators of soil metal pollution.[32] Midgut gland Cd fractions in *Arion lusitanicus* have been shown to be variably allocated between different metal-binding protein pools, depending on the level of environmental metal contamination.[33] This is in contrast to the very stable allocation of Cd in the midgut gland of terrestrial snails, where the metal is strongly and exclusively bound to a highly Cd-selective MT isoform.[14] This isoform is expressed at rising concentrations upon increasing Cd exposure,[34] and independent of variable amounts of Cu that may also be present in the midgut gland tissue.[15]

In the present study, the slug *A. vulgaris* was used to shed light on the apparent differences in the retention and binding capacity of Cd between terrestrial snails and slugs. Thus, our first aim was to detect the presence of MTLPs or MTs in experimentally metal-exposed slugs, and to identify the cDNA and primary structure of these longly predicted metal binding proteins. Second, we wanted to find out if these suggested slug MTs (or MTLPs) are similarly metal-selective as those reported from helicid snails. Additionally, we wanted to understand whether and how these supposed protein-based metal handling mechanisms may explain the inconsistencies in the metabolism of Cd and Cu between terrestrial snails and slugs *in vivo*. Thanks to the strong improvement of analytical methods and the diversity of modern approaches, such an analysis has nowadays become feasible and is featured in the present study.

Material and methods

Chemicals and reagents

Unless otherwise stated, all reagents and solvents were purchased from Carl Roth GmbH (Karlsruhe, Baden-Wurttemberg, Germany), and were of analytical or HPLC grade.

Animals, experimental design and metal exposure

Individuals of *A. vulgaris* were collected in the surroundings of Innsbruck during early summer 2014 and kept on moistened garden soil under constant conditions at a temperature of 18 °C and a photoperiod of 12:12 hours. They were fed on commercially available lettuce (*Lactuca sativa*).

After an acclimatization period of 2 weeks (for details see ref. 35), four different groups of slugs (with 90 individuals each) were used for metal exposure experiments (Cd and Cu) over a period of 15 days (see Table 1). Depending on the experimental design, animals were either fed on uncontaminated lettuce or on metal-enriched lettuce leaves which had been soaked for one hour in a metal containing solution of $CdCl_2$ (1 mg l^{-1}) or $CuCl_2$ (10 mg l^{-1}) respectively, as described previously.[36] Resulting metal concentrations (mean ± standard deviation) in enriched lettuce leaves were 46.67 ± 22.57 μg g^{-1} dry wt for Cd and 241.60 ± 137.39 μg g^{-1} dry wt for Cu. The respective concentrations (mean ± standard deviation) of uncontaminated lettuce were 5.39 ± 3.27 μg g^{-1} dry wt, for Cd and 15.43 ± 6.48 μg g^{-1} dry wt for Cu. In all cases, the number of analysed samples (n) was 6.

Two metal exposure groups (with 50 individuals each) were applied to a Cd and a Cu feeding experiment, respectively. For each metal exposure group, a respective control group

This journal is © The Royal Society of Chemistry 2018

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

Table 1 Experimental setup for metal feeding experiments applied to *A. vulgaris*, showing experimental endpoints (first column), exposure groups for both metals with numbers of involved individuals (second column), metal concentrations in the feed (third column), duration of exposure experiments (fourth column) and kind of organs sampled (last column). For details see description in the text

Experiments (endpoints)	Exposure groups (no. of animals)	Metal concentration in the feed [$\mu g\ g^{-1}$ dry wt] (molar concentration)	Duration (sampling day)	Organs sampled
Cd-feeding (*in situ* hybridization, quantitative RT-PCR, Cd accumulation)	Control (30) Cd-exposed (30)	Control, Cd: 5.39 ± 3.27 (0.05 $\mu mol\ g^{-1}$)	15 days (0, 3, 5, 8, 11, 15)	Midgut gland Mantle Gut
Cd-feeding (MT purification)	Control (20) Cd-exposed (20)	Cadmium exposure, Cd: 46.67 ± 22.57 (0.42 $\mu mol\ g^{-1}$)	15 days (15)	Midgut gland
Cu-feeding (quantitative RT-PCR, Cu accumulation)	Control (30) Cu-exposed (30)	Control, Cu: 15.43 ± 6.48 (0.24 $\mu mol\ g^{-1}$)	15 days (0, 3, 5, 8, 11, 15)	Midgut gland Mantle Gut
Cu-feeding (MT purification)	Control (20) Cu-exposed (20)	Copper exposure, Cu: 241.60 ± 137.39 (3.80 $\mu mol\ g^{-1}$)	15 days (15)	Midgut gland Mantle Foot

(with 50 animals each) was treated with uncontaminated feed (Table 1). A subset of 30 individuals from each group was used for a time course analysis of metal accumulation, MT mRNA upregulation, and (for the Cd feeding experiment and respective control snails only) *in situ* hybridization (ISH) of MT mRNA. Another subset of 20 animals from each group was used for MT purification upon gel chromatography, HPLC and MS/MS spectrometry for protein identification, respectively (Table 1).

Assessment of mortality

Throughout all experiments, no mortality of control or metal-exposed slugs was detected. Moreover, no change in feeding or locomotion activity was observed in slugs of all experimental groups.

Sampling

Slugs were dissected on an ice-cooled aluminum plate intermittently cleansed with 70% EtOH and RNase AWAY (Sigma-Aldrich, Vienna, Austria). Midgut gland, mantle and gut (foregut, midgut and hindgut together) of metal-exposed and control groups were sampled on day 0 (controls only), as well as on days 3, 5, 8, 11 and 15 of metal exposure. Small tissue aliquots (1 3 mg fresh wt) (see Table 1) were removed and stored at −80 °C in RNAlater™ solution (Thermo Fisher Scientific, Waltham, CA, USA) for succeeding mRNA isolation. For *in situ* hybridization, pieces of midgut gland were fixed for 24 hours in a solution of 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4) at 4 °C, dehydrated in methanol, cleared in methylbenzoate and benzene, and finally embedded in paraffin. Tissue aliquots for metal analyses were sampled in screw-capped polyethylene tubes (Greiner, Kremsmunster, Austria) and oven-dried at 60 °C. For MT protein purification, remaining midgut gland tissue of control or metal-exposed slugs was removed on days 0, 8 and 15 of exposure, pooled and frozen at −80 °C.

Metal analysis in lettuce and slug tissues

After dry weight determination, samples were heat-digested for several days at 70 °C in 12 ml screw-capped polyethylene tubes (Greiner) using a mixture (1 : 1) of nitric acid (suprapur, Merck, Darmstadt, Germany) and deionized water. The samples were digested until a clear solution was obtained, which was eventually spiked with a few drops of H_2O_2 for complete oxidation. Residues of digested solutions were filled up to a final volume of 11.5 ml by addition of deionized water and 5% nitric acid (suprapur, Merck). Sample metal concentrations (Cd and Cu) were measured by flame (model 2380 Perkin Elmer, Boston, MA, USA) or graphite furnace atomic absorption spectrophotometry with polarized Zeeman background correction (model Z-8200, Hitachi, Tokyo, Japan). Instrument calibration was achieved by using standard metal solutions (Titrisol, Merck) diluted with deionized water and 5% nitric acid (suprapur, Merck). Accuracy of metal analyses was verified by means of certified standard reference material (Lobster Hepatopancreas TORT-2 NRC-CNRC, Canada and Polish Virginia Tobacco Leaves INCT-PVTL-6, Poland), digested and treated in the same way as the tissue samples. Means of Cd and Cu concentrations assessed in standard reference materials were within a range of ± 12% of certified metal values of standard reference materials.

RNA isolation and cDNA synthesis

Midgut gland tissue was homogenized with glass beads using a Precellys 24 ball mill (Bertin Corp., Rockville, MD, USA). Total RNA was isolated applying the RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) including on-column DNase I digestion with the RNase-Free DNase Set (Qiagen). RNA was quantified with the Quant-iT™ RiboGreen® RNA Assay Kit (Life Technologies Corporation, Carlsbad, USA) using the Victor™ X4 2030 Multilable Reader (Perkin Elmer, Waltham, USA).

For quantitative Real Time PCR (qRT PCR), 450 ng of total RNA was transcribed into cDNA with SuperScript™ IV First-Strand Synthesis System (Invitrogen, Waltham, MA, USA) and stored at −20 °C. For sequence confirmation of the two MT isoforms (AvMT1 and AvMT2), 5′- and 3′-RACE-Ready cDNA s were generated by applying the SMARTer™ RACE cDNA amplification kit (Clontech, Takara Bio Europe, Saint-Germain-en-Laye, France).

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

Transcriptome and genome generation and screening

Isolated RNA from midgut gland tissue (see above) of one individual exposed to Cd through 15 days was sent to the Duke Center for Genomic and Computational Biology (GBC, Duke University, Durham, NC, USA) for transcriptome generation. The RNA was subjected to 125 bp paired-end Illumina sequencing. One library and 30 731 238 reads were generated. Raw data (available at the SRA database, NCBI: PRJNA490798) were assembled at the Institute of Zoology (University of Innsbruck) using Trinity Version v2.1.1 (GitHub Inc., San Francisco, USA)[37] and provided for analysis on a local TBlast page.

Nucleotide sequences coding for diverse snail MTs and in particular, for the *CdMT*, *CuMT* and *Cd*/*CuMT* isoforms of the related species *H. pomatias* and *C. aspersum* were blasted against the transcriptome data set of *A. vulgaris* using TBlast of the SequenceServer software (Priyam *et al.*, 2015).[69] Nucleotide sequences were analysed with CLC Main workbench 6.9 (Qiagen, Aarhus, Denmark) and aligned with online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/), in order to identify respective reading frames and characteristic MT cysteine motifs.

For genome sequencing, a small aliquot of foot tissue (approx. 100 mg) from *A. vulgaris* was homogenized using Precellys CKMix beads in 360 µl lysis buffer AP1 (Qiagen). Subsequently, 40 µl of Proteinase K (20 mg ml^{-1}) (Invitrogen) were added to the homogenized tissue and digestion was performed at 50 °C for 2 hours. RNAse A digestion was performed after completion of lysis. Genomic DNA was extracted by means of the DNeasy Plant Mini Kit (Qiagen). DNA was sequenced with Illumina s NextSeq 500 (150 nucleotides, paired-end, 170 Mio reads; StarSEQ, Mainz, Germany). Raw data (available at the SRA database, NCBI: PRJNA491624) were trimmed, formatted and analyzed with Geneios R10 (Biomatters Ltd., Auckland, New Zealand).

Screening was achieved by blasting of similar sequences from closely related species and the newly discovered MT sequences from the transcriptomic database against the genomic database. In both cases, the screening suggested the presence of two *MT* isoforms genes, called *AvMT1* and *AvMT2*.

Sequence confirmation of *AvMT1* and *AvMT2 via* PCR

For confirmation of *AvMT1*, gene-specific primers for the 5′- and 3′-RACE reaction were developed by incorporating both, the information from *de novo* protein sequencing of AvMT1 *via* ESI-MS (see below), and respective hits from transcriptomic data screening (see above). Consequently, a forward primer GSP2 with the sequence 5′-AATTGTCAGTGTGGAGGTGACTGCG-3′ and a reverse primer GSP1 with the sequence 5′-TGCATTGTGAACAATC GCAGTCACC-3′ were produced (Microsynth, Balgach, Switzerland). Rapid amplification of cDNA ends was performed according to the SMARTer™ RACE (Clontech) cDNA amplification kit protocol.

Amplified fragments were separated by size through electrophoresis on an 1.5% agarose gel in TAE buffer (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA) and extracted from gel with the QIAquick Gel extraction kit (Qiagen). Cloning of PCR fragments was performed with the TOPO TA Cloning Kit for Sequencing (Invitrogen). Plasmids were purified with the EasyPrep Pro Plasmid Miniprep Kit (Biozym, Hessisch Oldendorf, Germany) and aliquots were sequenced by Microsynth (Switzerland).

For confirmation of transcriptomic *AvMT2* sequences, primers covering the whole coding region of the respective MT were designed. PCR was performed with the Titanium® Taq DNA Polymerase kit (Clontech), using the forward primer 5′-CGTTAAGACCAACAGATCC-3′ and the reverse primer, 5′-TTCTCTTCCCACTCCTCAG-3′, both of them designed from untranslated regions according to the recommended protocol. A denaturation cycle of 95 °C for 1 min was followed by 30 cycles with 95 °C for 30 s, 55 °C for 30 s and 68 °C for 40 s, with a final 10 min extension step at 70 °C. Clean plasmids for subsequent sequencing by Microsynth (Switzerland) were obtained from amplified fragments as mentioned above for RACE PCR.

Quantitative real time PCR (qRT PCR)

qRT PCR was performed with QuantStudio® 3 Real-Time PCR Instrument (96-Well 0.1 ml Block; Thermo Fisher Scientific, Waltham, USA) using Power SYBR® Green PCR Master Mix (Applied Biosystems by Thermo Fisher Scientific). Primers were designed with the Primer Express 3.0 software (Applied Biosystems) and optimal concentrations were determined by performing dissociation curves. PCR efficiency and calibration curve were 72.41%, with a CT value = −4.2273 log copies# +42.456 for *AvMT1* and 71.53%, with a CT value = −4.2672 log copies# +43.197 for *AvMT2*. *AvMT1* (amplicon length 65 bp) and *AvMT2* (amplicon length 75 bp) transcripts were amplified with the following concentrations and primers.

AvMT1 isoform: sense, 300 nM: 5′-ACTGGAGCTTGTAAAAGT GAACCA-3′; antisense, 900 nM: 5′-TCGCAGTCACCTCCACATTG-3′.

AvMT2 isoform: sense, 900 nM: 5′-TGTGAGGATGGATGCCA GTGT-3′; antisense, 900 nM: 5′-GCCATCATTGGTGCATTTACAA-3′.

PCR conditions were: 10 min denaturation at 95 °C, followed by 40 cycles of 15 s at 95 °C, and 60 s at 60 °C. Samples were measured in triplicates.

In situ hybridization (ISH)

AvMT1 and *AvMT2* cDNA fragments were amplified by PCR using tagged primers with a T7 tag (5′-GGATCCTAATACGAC TCACTATAGG-3′), and an Sp6 tag (5′-CATTTAGGTGACACTATA GAAG-3′) for the *AvMT1* forward primer, 5′-CATTTAGGTG ACACTATAGAAGAAAAGTGAACCATGTCAGTGTGG-3′ and the *AvMT1* reverse primer, 5′-GATCCTAATACGACTCACTATAGGCT GCAAGATTTTTCACATGAGC-3′, as well as the *AvMT2* forward primer, 5′-CATTTAGGTGACACTATAGAAGCAACGGTACCTGCA ACAGC-3′ and the *AvMT2* reverse primer, 5′-GGATCCTAATA CGACTCACTATAGGTAATTGCAGCCACAAGAAGTGC-3′ (tag sequences underlined).

PCR products were purified by QIAquick PCR purification kit (Qiagen). The Digoxigenin labelled *in situ* probes were synthesized with Roche DIG labelling kit (Roche Diagnostics, Basel, Switzerland), following the manufacturers protocol.

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

Probes were purified with Micro Bio Spin Chromatography Columns (Biorad, Hercules, CA, USA), quantified with Nanodrop 2000 (Thermo Fisher) and diluted to a final concentration 5 ng μl^{-1} with the HybMix (without Tween, see ref. 38).

Hybridization was performed according to Chabicovsky *et al.* (2003)[39] with minor modifications. Briefly, after deparaffinization and rehydration, 7 μm thick sections of midgut gland from up to six Cd-exposed and control slugs were rinsed in 1× PBS (Phosphate-buffered saline, pH 7.4), treated 5 min in 20 μg ml^{-1} Proteinase K in 1× PBS at 37 °C, followed by 5 min in 1× PBS at 4 °C. Subsequently, sections were incubated for 2 × 15 min in 1× PBS with 0.1% active DEPC, followed by 15 min in 5× SSC (Saline sodium citrate, pH 7.0). Sections were pre-hybridized for 2 h at 50 °C (*AvMT1*) or 53 °C (*AvMT2*) in a mixture consisting of 50% (v/v) formamide (Sigma-Aldrich Chemie GmbH, Steinheim, Germany), 5× SSC buffer, and 40 μg ml^{-1} salmon sperm DNA (Sigma-Aldrich Chemie GmbH). The diluted probes (0.3 ng μl^{-1}) were denaturated for 8 min at 80 °C, applied to dried sections and covered with cover slip. Hybridization was achieved overnight at 50 °C (*AvMT1*) or 53 °C (*AvMT2*) in a sealed humidity chamber containing 50% formamide (Sigma Aldrich Chemie GmbH) in 5× SSC at 50 °C (*AvMT1*) or 53 °C (*AvMT2*). After incubation, cover slips were removed by immersing slides to 2× SSC for 30 min. Sections were washed 1 hour in 2× SSC, 1 hour in 0.1% SSC at 53 °C (*AvMT1*) or 58 °C (*AvMT2*) and equilibrated for 5 min in 0.1 M Tris, 0.15 M NaCl (pH 7.5). The sections were incubated in anti-digoxigenin-alkaline phosphate antibodies (Roche Diagnostic) diluted 1 : 500 in 1× blocking solution and MAB (Maleic acid buffer, pH 7.5) (v/v 1 : 1) in the dark. Staining was achieved with 5-bromo-4-chloro-3indolyl phosphate/nitro blue tetrazolium chloride (Roche Diagnostics) for 4 7 hours for *AvMT1* and 8 14 hours for *AvMT2*. For counterstaining, sections were incubated 5 min in Nuclear Fast Red Stain solution. Control sections were processed in the absence of probes.

Chromatography and MT purification

Frozen midgut gland samples were homogenized (1 : 3 w/v) in a buffer containing 25 mM Tris HCl (pH 7.5), 5 mM β-mercaptoethanol, and 0.1 mM phenyl-methyl sulfonyl fluoride (PMSF). After an incubation on ice for 10 min, the homogenate was centrifuged for 15 min at 15 000 rpm at 4 °C. About 8 12 ml of filtrated supernatant (0.2 μm syringe filter) was applied to a Sephacryl S-100 column (GE Healthcare, Little Chalfont, UK) (2.5 × 30 cm) calibrated with a mixture of standard components (Blue dextran, >100 kDa; Myoglobin, 17.5 kDa; Vitamin B12, 1.5 kDa) (Sigma Aldrich Chemie GmbH). Elution was performed with 25 mM Tris HCl buffer (pH 7.5) containing 5 mM β-mercaptoethanol at a flow rate of 2 ml min^{-1}. Fractions of 4 ml were collected. MT-containing fractions were pooled and further purified by ion exchange chromatography (HiTrap® Q 1 ml (GE Healthcare)) with a dual buffer system (A: 25 mM Tris HCl, pH 7.5, 5 mM β-ME; and B: A + 1 M NaCl) at a flow rate of 1 ml min^{-1}. Interesting fractions were pooled and fractionated by Reversed-Phase HPLC (LiChrospher® 100 RP-18(Merck)) with a dual buffer system of either buffer A, consisting of 25 mM Tris (pH 7.5), 5 mM β-ME and buffer B, with 60% acetonitrile in A; or, alternatively, buffer A with 0.1% TFA and buffer B, 60% acetonitrile in A, at a flow rate of 1 ml min^{-1}. For purification of AvMT2, 500 μl of pooled MT-containing fractions from ion exchange chromatography were mixed with 125 μl 20 mM ammonium-tetrathiomolybdate (TTM) and shaked for 15 min at 1000 rpm in order to remove the Cu^{+} ions from the MT. Further, the sample was fractionated by HPLC with a Superdex peptide 10/300 GL column (GE Healthcare) with a 10 mM NH_4HCO_3, calibrated with the same standard components as mentioned above.

De novo sequencing of AvMT1 and AvMT2

Samples for *de novo* sequencing were reduced with 50 μl 10 mM dithiothreitol at 56 °C for 30 min, alkylated with 50 μl 55 mM iodoacetamide at room temperature for 20 min, and digested either with arginine C, peptidase V8 or lysine C for 5 hours at 37 °C in 100 mM NH_4HCO_3 buffer at pH 8.0. Samples were analysed using an UltiMate 3000 nano-HPLC system coupled to a Q Exactive Plus mass spectrometer (both Thermo Scientific, Bremen, Germany) equipped with a Nanospray Flex ionization source. The peptides were separated on a homemade fretless fused-silica microcapillary column (75 μm i.d. × 280 μm o.d. × 10 cm length) packed with 3 μm reversed-phase C18 material (Reprosil). Solvents for HPLC were 0.1% formic acid (solvent A) and 0.1% formic acid in 85% acetonitrile (solvent B). The gradient profile was as follows: 0 2 min, 4% B; 2 55 min, 4 50% B; 55 60 min, 50 100% B, and 60 65 min, 100% B. The flow rate was 250 nL min^{-1}. The Q Exacitve Plus mass spectrometer was operating in the data dependent mode selecting the top 12 most abundant isotope patterns with charge >1 from the survey scan with an isolation window of 1.6 mass-to-charge ratio (m/z). Survey full scan MS spectra were acquired from 300 to 1750 m/z at a resolution of 70 000 with a maximum injection time (IT) of 120 ms, and automatic gain control (AGC) target 1e6. The selected isotope patterns were fragmented by higher-energy collisional dissociation (HCD) with normalized collision energy of 28 at a resolution of 35 000 with a maximum IT of 120 ms, and AGC target 5e5.

De novo sequencing was performed with Peaks Studio 7.0 (Bioinformatic Solutions Inc., Waterloo, ON, Canada). The software calculates a local confidence score for each amino acid that extends the accuracy of amino acid assignment in a resultant peptide. Only local confidence scores higher than 90% were considered for *de novo* sequencing. The *AvMT1* sequence was determined by alignment of the resultant peptides from samples digested with different enzymes. The results from *de novo* sequencing were verified by adding the sequence of AvMT to uniprot database performing database search using Proteome Discoverer 1.4.1.14 (Thermo Scientific) with search engine Sequest. Precursor and fragment mass tolerance was set to 10 ppm and 0.02 Da, respectively, and up to two missed cleavages and ten modifications per peptide were allowed. Carbamidomethylation of cysteine, and acetylation of peptide N-terminus were set as variable modifications. Peptide identifications were filtered at 1% false discovery rate.

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

Cloning of *AvMT1* for recombinant expression

AvMT1 DNA was synthesized by Integrated DNA Technologies (Leuven, Belgium), based on the known *AvMT1* sequence. 10 ng of the synthetic cDNA were PCR-amplified in a 25 μl PCR mixture using the Expand High Fidelity PCR system (Roche Diagnostics) and a pair of specifics primers (forward, 5′-AAAAGGATCCATGAGCGGAAAAGCCTGC-3′ and reverse, 5′-GG GGCTCGAGCTACTTGCAACTGCAAGATT-3′) to add suitable terminal *Bam*HI/*Xho*I sites for cloning. The PCR conditions were as follows: 1 cycle at 94 °C for 5 min, 25 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and a final extension step at 72 °C for 7 min. PCR products were *BamHI*/*Xho*I-digested and cloned into the *E. coli* pGEX-4T-1 expression vector (GE Healthcare, Little Chalfont, UK) with the DNA Ligation kit 2.1 (Takara Bio Inc.) for glutathione S-transferase (GST)-fusion protein production. Sequencing was performed at the Scientific and Technological Centers of the University of Barcelona, using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) in an automatic sequencer (ABIPRISM 310, Applied Biosystems).

Production of recombinant metal AvMT1 complexes

For heterologous protein production, 300 ml or 500 ml of LB medium with 100 μg ml^{-1} ampicillin were inoculated with protease-deficient *E. coli* BL21 cells transformed with the *AvMT1* recombinant plasmid. After growth overnight at 37 °C, the culture inoculated with 3 5 l of fresh LB-100 μg ml^{-1} ampicillin medium for production of Cu-, Cd- and Cd/Cu AvMT1 complexes. Gene expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h. After the first 30 min of induction, cultures were supplemented with Cu or Cd (500 μM of $CuSO_4$ or 300 μM of $CdCl_2$) or with an 1 : 1 mixture of both metals (150 μM of $CuSO_4$ + 150 μM of $CdCl_2$), in order to generate metal AvMT1 complexes at standard aeration conditions (shaking at 250 rpm). Cells were harvested by centrifugation for 5 min at 7700 rpm, and bacterial pellets were suspended in 75 125 ml (for 3 l or 5 l cultures, respectively) of ice-cold PBS (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4 and 0.5% v/v β-mercaptoethanol). Suspended cells were sonicated (Sonifier Ultrasonic Cell Disruptor) 8 min at voltage 6 with pulses of 0.6 s, and then centrifuged for 40 min at 12 000 rpm and 4 °C.

Purification of recombinant metal AvMT1 complexes

Protein extracts containing GST AvMT1 fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. GST AvMT1 fusion proteins bound to the sepharose beads were washed with 20 or 30 ml (for 3 l or 5 l cultures, respectively) of cold 1× PBS bubbled with argon to prevent oxidation. After three washes, GST AvMT1 fusion proteins were digested with thrombin (GE Healthcare, 25 U l^{-1} of culture) overnight at 17 °C, thus enabling separation of the metal AvMT1 complexes from the GST that remained bound to the sepharose matrix. The eluted metal AvMT1 complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris HCl, pH 7.0, and run at 0.8 ml min^{-1}. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at −80 °C until use.

Analysis of recombinantly expressed and *in vitro* prepared metal AvMT1 complexes

The recombinantly expressed metal AvMT1 complexes were analyzed for element composition (S, Zn, Cd and Cu) by inductively coupled plasma atomic emission spectroscopy (ICP-AES) on a Polyscan 61E spectrometer (Thermo Jarrell Ash Corporation, Franklin, MA, USA) at appropriate wavelengths (S, 182.040 nm; Zn, 213.856 nm; Cd, 228.802 nm; Cu, 324.803 nm) under conventional (dilution with 2% HNO_3 (v/v)) conditions.[40] AvMT1 concentration in the recombinant preparations was calculated from the acidic ICP sulfur measurements, assuming the only contribution to their S content was that made by the Cu AvMT1, Cd AvMT1 and Cd/Cu AvMT1 peptides. Molecular mass was determined by electrospray ionization mass spectrometry with a time-of-flight analyzer (ESI-TOF MS) using a Micro Tof-Q Instrument (Bruker Daltonics Gmbh, Bremen, Germany) calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies, Santa Clara, CA, USA), interfaced with a Series 1100 HPLC pump (Agilent Technologies) equipped with an autosampler, both controlled by the Compass Software. The experimental conditions for analyzing Cd proteins were as follows: 20 μl of the sample were injected through a PEEK tube (1.5 m × 0.18 mm i.d.) at 40 μl min^{-1} under the following conditions: capillary-counter electrode voltage, 5.0 kV; desolvation temperature, 90 110 °C; dry gas 6 l min^{-1}. Spectra were collected throughout an *m*/*z* range from 800 to 2000. Cu containing proteins were analyzed by injecting 20 μl of the sample at 30 μl min^{-1}; capillary-counter electrode voltage, 4.0 kV; desolvation temperature, 80 °C; *m*/*z* range from 800 to 2000. The liquid carrier was a 90 : 10 mixture of 15 mM ammonium acetate and acetonitrile, pH 7.0. For the analysis at acidic pH the conditions used were the same as those used in the analysis of the divalent metals, except in the composition of the carrier liquid, which in this case was a 95 : 5 mixture of formic acid and acetonitrile at pH 2.4. All samples were injected in duplicates to ensure reproducibility. In all cases, molecular masses were calculated according to the reported method.[41]

Cladogram construction

New MT sequences from the species *Cochlicella acuta*, *Lehmannia nyctelia* and *Theba pisana* were identified by applying the Blast tool (default parameters) and datasets of the eSnail database (http://soft.bioinfo-minzhao.org/esnail/).[42]

For confirmation of the MT isoforms from *Cepaea hortensis (C.h.)* and *Limax maximus (L.m.)*, hepatopancreatic tissue of three adults was dissected on an ice-cooled aluminum plate intermittently cleansed with water and RNase AWAY (Sigma-Aldrich, Vienna) and stored in RNAlater (Fisher Scientific, Vienna, Austria) at −80 °C. Tissue disruption, total RNA isolation and cDNA synthesis were carried out as described above (see RNA Isolation and cDNA synthesis). The coding region was amplified with the Advantage 2 PCR System (Clontech, Takara Bio Europe)

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

and following gene specific primers: L.m.CdMT: sense, 5′-TTA CAATGAGCGGAAAAGGAG-3′; antisense, 5′GAGTTCCTCCTTCC TGTGG-3′. C.h.CdCuMT: Sense, 5′-AGTCAGCAGTGATCCAGC-3′; antisense, 5′-GAGTAATCCCACACATCCTTG-3′. C.h.CdMT: Sense, 5′-ATCAAGGCCAATTTTACACCC-3′; antisense, 5′-TGGTTGTGGT TGATAATGTGG-3′. The cycling parameters were as follows: Initial denaturation 95 °C for 1 min, 30 cycles of denaturation at 95 °C for 30 s, annealing at 54.5 °C (C.h.Cd/CuMT), 55.5 °C (C.h.CdMT) or 55 °C (L.m.CdMT) for 30 s, extension at 68 °C for 1 min and a final extension at 68 °C for 5 min. PCR products were analysed on a 1.5% agarose gel (Biozym). Respective bands were cut out, cleaned with the QIAquick gel extraction kit (Qiagen) and sent to Microsynth (Switzerland) for Sanger sequencing.

Alignment and cladogram construction of gastropod MTs were performed using the publicly accessible computational tool SeaView (seaview4.exe) (http://doua.prabi.fr/software/seaview)[43] available *via* the PRABI portal (Rhone-Alpes Bioinformatics Center) (http://www.prabi.fr/). Precisely, sequences were first automatically aligned using the ``muscle option and then manually adjusted. Tree calculation was performed by a maximum likelihood approach using ``PhyML with a bootstrap analysis based on 500 replications.

In particular, the MT sequences used in the present study were from the following species (with GenBank, UniProt or eSnail database accession numbers). *Helix pomatia*: CdMT (AF399740.1); CdCuMT1 (KY420172.1); CdCuMT2 (KY420171.1); CuMT (AF399741.1); *Cantareus aspersus*: CdMT (EF152281.1); CdCuMT (EF206312.1); CuMT (EF178297.2); *Ceapaea hortensis*: CdMT1 (MH574545); CdMT2 (MH574546); CdCuMT (MH574547); *Cochlicella acuta*: CdCuMT (Unigene65576); CuMT (Unigene18701); *Arion vulgaris*: *AvMT1* (MF155618.1); AvMT2 (MF155619.1; MF155620.1); *Theba pisana* CuMT (CL8321.Contig1); *Arianta arbustorum* CdMT (UniProt P55946*); *Lehmannia nyctelia* CdMT (CL1241.Contig1); *Limax maximus* CdMT (MH 574544).

Statistical analysis

Since most values for tissue metal concentrations and qRT PCR failed to pass the Shapiro Wilk normality test and the equal variance test, non-parametric statistical methods were applied as mentioned below. A two-way analysis of variance (ANOVA) was performed in Sigmaplot 12.5 (SYSTAT software, San Jose, CA, USA) applying the Holm Sidak method for pairwise and multiple comparisons, with a significance level of $p \leq 0.05$. A home-made graphics program (RLplot 1.5.6a) (https://www.uibk.ac.at/zoology/download/rlsoft/index.html.en) was applied to design the plots which were finally edited with Photoshop CS4.

Results

Metal uptake and organ-specific accumulation

Upon Cd exposure, *A. vulgaris* accumulated the metal mainly in the midgut gland and the gut tissues (Fig. 1A and B). Individuals from the control group fed on uncontaminated lettuce did not show any significant change of tissue Cd concentrations. When fed on Cd-enriched diet, however, the highest metal accumulation was found in the midgut gland reaching concentrations of about 300 μg per g dry weight (Fig. 1A). Compared to the Cd concentration of metal-enriched lettuce (46.67 μg Cd per g dry wt, see above), this is an increase by a factor of 6.4. A significant accumulation, although much lower than in the midgut gland, was also observed in the gut, while no change of Cd-concentration at all was observed in the mantle tissue (Fig. 1B).

Upon Cu exposure, a strong accumulation of Cu was again observed in the midgut gland of metal-exposed slugs (Fig. 1C), where metal concentrations rose to more than 650 μg g^{-1} dry wt Compared to the Cu concentration in Cu-enriched lettuce (241.60 μg g^{-1} dry wt, see above), this corresponds to a concentration factor of 2.7. A considerable accumulation of Cu was also seen in the gut tissue, while the lowest, yet significant, increase of Cu concentration was noticed in the mantle (Fig. 1D).

Overall, these data clearly show that upon metal exposure, the midgut gland of *A. vulgaris* is the main target organ for both, Cd and Cu accumulation.

Involvement and upregulation of *Metallothionein* genes

Based on molecular evidence through evaluation of our generated transcriptomic and genomic data sets (see Material and methods), two distinct *MT* genes have been identified in *A. vulgaris* (*AvMT1* and *AvMT2)*. Extensive screening of both databases for possible additional *MT* genes remained unsuccessful. The cDNAs of the two genes were confirmed and characterized by PCR from midgut gland tissue aliquots of metal-exposed slugs. For the cDNA of AvMT2, two variants were identified differing from each other by a single silent nucleotide polymorphism (SNP). They were published in GenBank and can be found under the accession numbers MF155618.1 (*AvMT1*), MF155619.1 (*AvMT2 variant 1*) and MF155620.1 (*AvMT2 variant 2*).

The significances of all curves shapes were tested with ANOVA ($p \leq 0.05$) and were found to be significant for Cd concentrations in midgut gland (A) and gut of Cd-exposed slugs (B), as well as in midgut gland (C) and gut and mantle (D) of Cu-exposed slugs. The asterisks above single values indicate significant differences compared to respective control values Holm Sidak method of all pairwise multiple comparison ($p \leq 0.05$).

As shown by quantitative Real-time PCR, the two *MT* genes of *A. vulgaris* (*AvMT1* and *AvMT2*) differ significantly in their response towards metal exposure (Fig. 2). The most striking response of transcriptional upregulation to metal exposure was shown by *AvMT1* upon Cd treatment, compared to the steady low level of *AvMT1* in control animals. In contrast, no reaction to Cd stress at all was observed for *AvMT2* (Fig. 2A). Interestingly, *AvMT1* was also responsive to Cu exposure, even more so than *AvMT2*, which reacted to Cu exposure by an only slight increase of transcriptional expression, compared to respective control levels (Fig. 2B). The strong responsiveness of *AvMT1* to exposure of either of the two metal treatments (Cd and Cu), can be appraised more clearly by comparing its high induction levels (referred to respective control values) with to the nearly unchanged or only modestly increased fold induction levels of *AvMT2* (Fig. 2C and D).

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

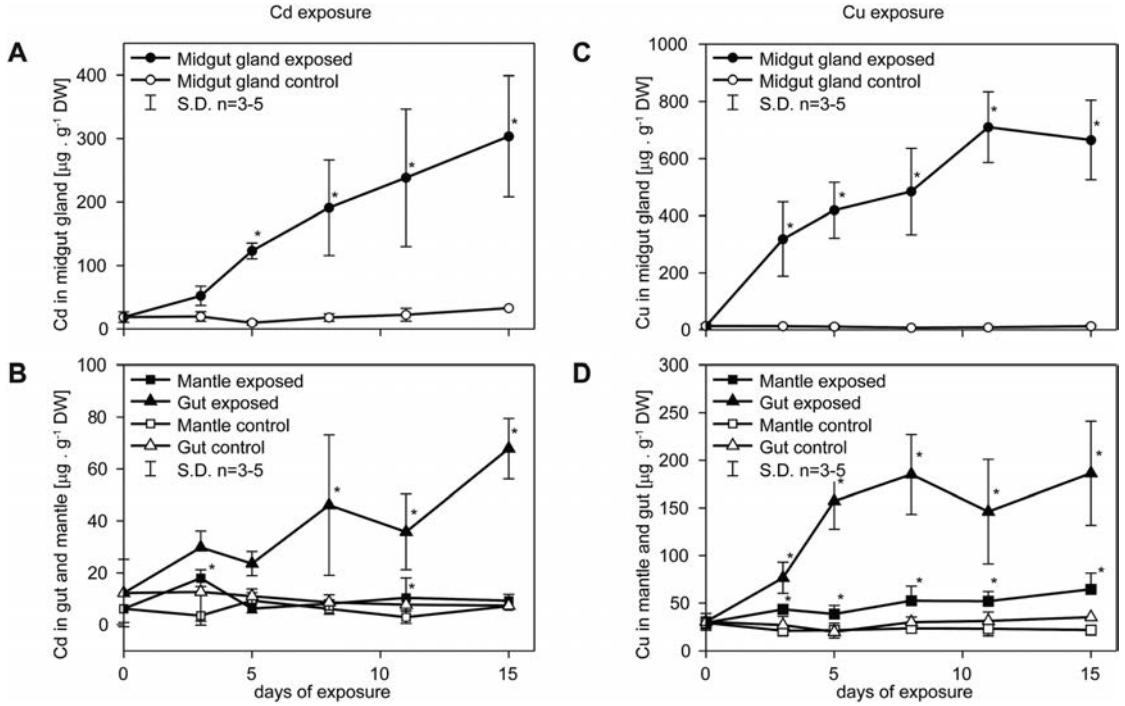


Fig. 1 Metal concentrations (means and standard deviations, n = 4 5) expressed in µg per g dry wt, shown in organs of slugs exposed to of Cd or Cu over a period of 15 days. (A) Course of Cd concentrations through 15 days in midgut gland of control (empty circle symbols) and Cd-exposed slugs (full circle symbols). (B) Cd concentrations through 15 days in gut (triangles) and mantle (squares) of control (empty symbols) and Cd-exposed slugs (full symbols). (C) Course of Cu concentrations through 15 days in midgut gland of control (empty circle symbols) and Cu-exposed slugs (full circle symbols). (D) Cu concentrations through 15 days in gut (triangles) and mantle (squares) of control (empty symbols) and Cu-exposed slugs (full symbols).

Localization of *AvMT1* mRNA by *in situ* hybridization

In situ hybridization (ISH) was applied for cellular localization of *AvMT1* and *AvMT2* mRNA transcription in the midgut gland of unexposed and Cd-treated slugs. Overall, positive mRNA transcript signals were only found for *AvMT1*. No ISH-signals at all were detected for *AvMT2*, independent of the kind of exposure. For AvMT1, there was a clear difference of ISH-staining between Cd-exposed and control samples, with positive transcript signals observed only in Cd-exposed individuals. This corresponds to our findings in Real-time detection PCR (see Fig. 2), showing that a significant upregulation capacity was only found for the *AvMT1* gene. ISH also revealed a cell specific gene expression for *AvMT1*. As seen by the blueish precipitations in Fig. 3A, the main sites of *AvMT1* expression in Cd-exposed slugs were the digestive (Dc) and the excretory cells (Ec) of the midgut gland tubular epithelium. No *AvMT1* mRNA was observed in tubular calcium cells (Cc). Also the rhogocytes (Rh) surrounding a transversal section of a blood vessel beside the tubular epithelium (Fig. 3A) were devoid of *AvMT1* mRNA. The positive reaction products for *AvMT1* mRNA in Fig. 3A contrast with a control section of the same tissue region treated with the sense probe for *AvMT1* (Fig. 3B), thereby confirming the successful ISH reaction in midgut gland cells of Cd-exposed slugs. A higher magnification of a *tubulus* section from another Cd-exposed slug further supports the observation that positive ISH products for *AvMT1* mRNA were mainly located in elongated digestive cells and in roundish or polygonal excretory cells, but not in calcium cells (Fig. 3C and D). In contrast to Cd-exposed slugs, no reaction product of *AvMT1* mRNA was observed in the midgut gland of control animals (Fig. 3E and F).

MT purification and identification of AvMT1 at the protein level

In order to purify and identify AvMT1 and AvMT2 at the protein level, different chromatographic purification steps were applied. First, supernatants of midgut gland homogenates from control and metal-exposed (Cd or Cu) slugs were filtered and applied to gel permeation chromatography (see Material and methods). Fractionation of midgut gland supernatants from control (uncontaminated) slugs revealed the presence of very little Cu ($\sim$90 µg l^{-1}) and Cd ($\sim$40 µg l^{-1}) in the high molecular weight protein range ($\geq$100 000 Da), and little Cd in the presumed MT-containing fractions (Fig. 4A). After exposure to Cd, a distinct Cd peak appeared in the presumed MT containing fractions along with a somewhat increased Cu peak (Fig. 4B). Interestingly, however, considerable amounts of Cd and Cu remained associated with high molecular weight protein fractions of $\geq$100 000 Da at similarly moderate concentrations as already observed in controls (see Fig. 4A).

Cd containing MT fractions from gel chromatography (Fig. 4B, open arrow) were further purified by ion exchange chromatography and Reversed-Phase HPLC (Fig. S1, ESI†). After ion exchange chromatography in a Tris HCl/NaCl gradient, a homogeneous Cd peak coinciding with a high absorption

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

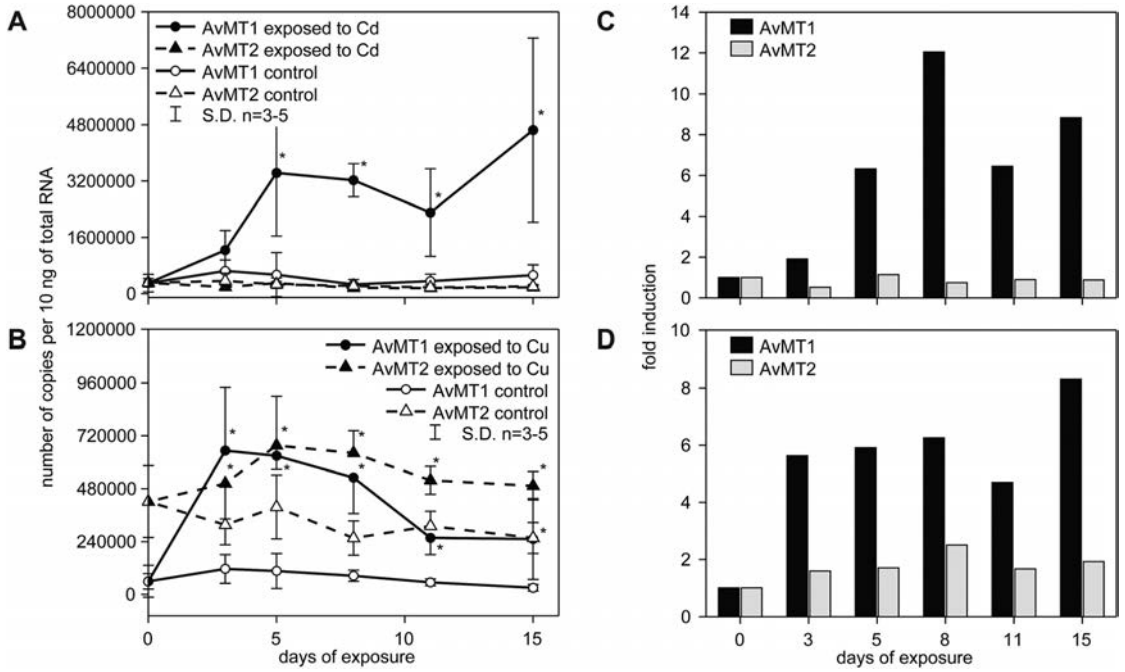


Fig. 2 Transcriptional upregulation and fold induction of *MT* genes of *A. vulgaris* (*AvMT1* and *AvMT2*) upon Cd or Cu exposure, expressed as mRNA copy numbers/10 ng of total RNA (means and standard deviations, $n = 5$). (A) mRNA copy numbers of *AvMT1* (circles) and *AvMT2* (triangles) in the midgut gland of *A. vulgaris* through 15 days of Cd exposure in control slugs (empty symbols) and metal-treated animals (full symbols). (B) mRNA copy numbers of *AvMT1* (circles) and *AvMT2* (triangles) in the midgut gland of *A. vulgaris* through 15 days of Cu exposure in control slugs (empty symbols) and Cu-exposed individuals (full symbols). The significances of all curves were tested with ANOVA ($p \leq 0.05$) and were found to be significant for AvMT1 in Cd-exposed slugs (A) and for AvMT1 and AvMT2 in Cu-exposed slugs (B). The asterisks above single values indicate significant differences compared to respective control values (Holm Sidak method of all pairwise multiple comparison) ($p \leq 0.05$). (C) Fold induction of *AvMT1* (black bars) and *AvMT2* (grey bars) upon Cd exposure. (D) Fold induction of *AvMT1* and *AvMT2* upon Cu exposure. In every case, -fold induction was derived from the mean ratio of copy numbers in metal-exposed slugs to copy numbers of control slugs at each respective day of exposure.

signal at 254 nm was obtained (Fig. S1A, ESI†). After further purification of the respective Cd-containing fractions by Reverse-Phase HPLC under acidic conditions (Fig. S1B, ESI†), two distinct peaks were finally obtained (peak A and peak B, see Fig. 5A).

Of these two peaks, peak A did not contain MTs or MT-like components, whereas peak B was identified as an MT by *de novo* sequencing on ESI-MS. The respective amino acid sequence retrieved is compared with the respective cDNA translations below (see Fig. 7). It is not complete, since it lacks one N-terminal (S) and five C-terminal amino acids (SCSCK). Nevertheless, its identity with AvMT1 is obvious.

Cu-containing fractions from gel chromatography of Cu-exposed slugs (Fig. 4C, open arrow) were also fractionated by ion exchange chromatography (Fig. S2, ESI†), revealing a distinct peak containing most of the Cu, besides several smaller Cu peaks indicating a rather broad distribution pattern of Cu along the elution gradient. After removal of Cu by treatment with the strong Cu chelator ammonium-tetrathiomolybdate (TTM),[24] a metal-free apo-thionein was obtained (Fig. 5B, Peak C) as compared with the Cu-containing MT not subjected to TTM treatment (Fig. 5B, peak D). Fractions from both peaks (peak C and peak D) were collected for further analysis upon ESI-MS. No MT sequence was identified in peak C. Surprisingly, small amounts of AvMT1 were retrieved in peak D, but no traces of the supposed AvMT2 at all.

This indicates that at the protein level, AvMT1 may be associated with both, Cd and Cu, while AvMT2 was virtually absent. Overall, it appears that Cd is more tightly bound to AvMT1 than Cu, which is also indicated by the fact that a great part of Cu was lost during the chromatographic purification steps. In fact, a quantitative balance of metal recovery from midgut gland supernatants of metal-exposed slugs through all steps of protein purification (Fig. S3, ESI†) suggested that in Cu-exposed slugs, only 5.6% of the of the total Cu applied at the beginning of fractionation were recovered at the end of purification. This suggests that most of the Cu was lost. In contrast, in Cd-exposed animals about 29% of the Cd applied at the beginning of MT purification was still recovered at the end. This indicates that in *A. vulgaris*, Cd is bound more specifically and stronger to the MT system than Cu.

Metal binding properties of recombinantly produced AvMT1

Since only AvMT1 was found to be present in *A. vulgaris* at detectable concentrations, the following studies on metal binding abilities were focused on AvMT1. A comparative study involving the metal binding features of both isoforms is in preparation (unpublished data). As shown in Fig. 6, the recombinant synthesis of AvMT1 in Cd-supplemented media led to a unique Cd_6 AvMT1 species, which highlights its clear Cd-thionein nature (Fig. 6A). On the other hand, when the

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

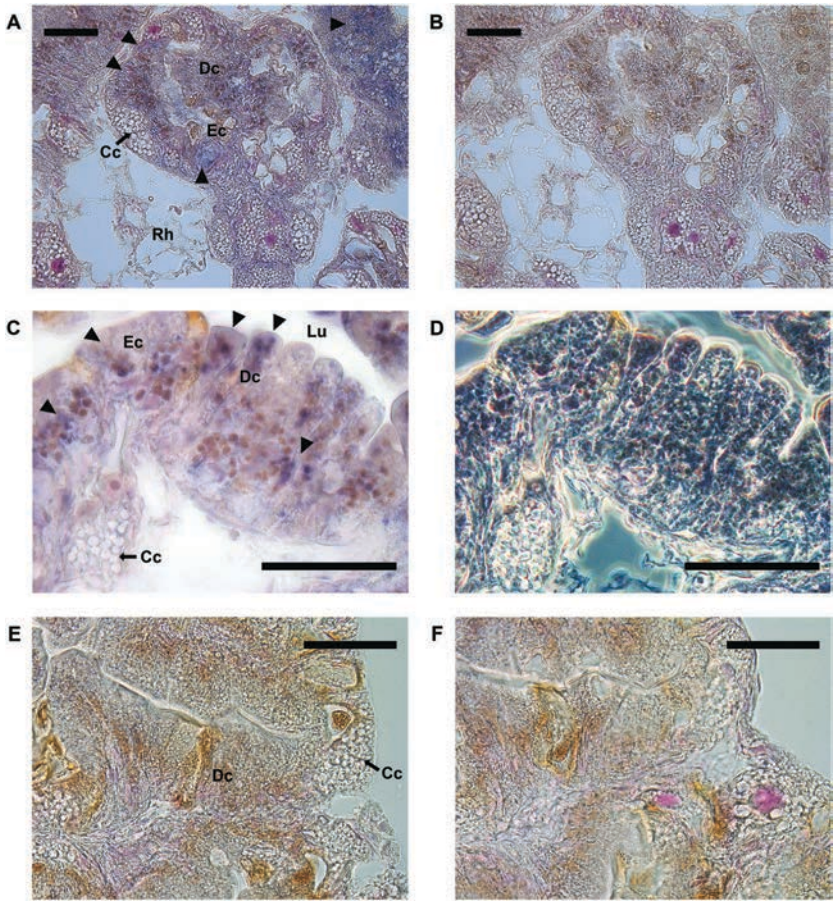


Fig. 3 *In situ* hybridization of *AvMT1* mRNA in paraffin midgut gland sections and cells of *A. vulgaris* exposed to Cd for 8 days (A D) and of unexposed control slugs (E and F). (A) Midgut gland cross section treated with antisense probe, showing a *tubulus* with constricted central lumen surrounded by digestive cells (Dc) and excretory cells (Ec) with blueish precipitations of *AvMT1* mRNA (arrowheads). Calcium cells (Cc) and rhogocytes (Rh) are devoid of reaction products (B) control cross section of the same tissue region as in (A), treated with the sense probe. (C) Detail of a *tubulus* treated with antisense probe, showing part of the lumen (Lu), confined by digestive cells (Dc) and excretory cells (Ec) with *AvMT1* mRNA staining (arrowheads), but not in calcium cells (Cc). (D) Same tissue section and same cells as in (C), photographed upon phase contrast, showing the elongated shape of digestive cells and the roundish shape of excretory cells. (E) Digestive gland *tubulus* of a control (uncontaminated) slug treated with antisense probe, showing digestive cells (Dc) and calcium cells (Cc) devoid of *AvMT1* mRNA. (F) Same tissue region as in (E), but after treatment with sense probe. The black scale bar corresponds to a size of 50 μm.

expression of the *AvMT1* gene was induced in Cu-supplemented cultures of *E. coli*, a complex mixture of heteronuclear Zn, Cu AvMT1 complexes was obtained (Fig. 6C). This reveals the non Cu-thionein features of AvMT1 and confirms its preferential selectivity for Cd.

Interestingly, the recombinant synthesis of the AvMT1 protein in a 1:1 mixed Cd/Cu-supplemented medium did not yield a mixture of both former results (see above), but a very major Cd6 AvMT1 species plus a very minor Cd6Cu1 AvMT1 species in agreement with a small content of Cu in the preparation, as suggested by the ICP data (Fig. 6B). The theoretical and experimentally observed masses of all MT metal species are summarized in the Table S1 (ESI†). Overall, the data obtained can be explained by the theoretical capability of the protein to bind Cu(I) with higher affinity than Cd(II) following the Irwing Williams rules. At the same time, AvMT1 does not behave as a true Cu-thionein, since the Cu^{+} incorporation into the protein should be higher, if these rules were strictly followed. These findings will be the object of an intense future study, but at this stage, the results allow to explain the association of AvMT1 with some Cu *in vivo*.

Phylogenetic classification of AvMT1 and AvMT2 in the frame of Stylommatophora MTs

AvMT1 and AvMT2 are highly similar among each other, indicating that they may have arisen from a previous event of gene duplication of an ancestral *MT* gene (Fig. 7A), a feature that they apparently share with all other stylommatophoran metal- and non-specific MTs. At the same time, the two proteins seem to be closely related to the metal-specific MT isoforms (HpCdMT and HpCuMT) of *H. pomatia*. AvMT1 has a higher degree of similarity with HpCdMT, whereas AvMT2 resembles more closely the sequence of HpCuMT. These sequence similarities can be extended to other MT sequences of terrestrial snails by blast comparisons of AvMT1 and AvMT2 with other MT sequences of terrestrial pulmonates: the AvMT1

This journal is © The Royal Society of Chemistry 2018

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

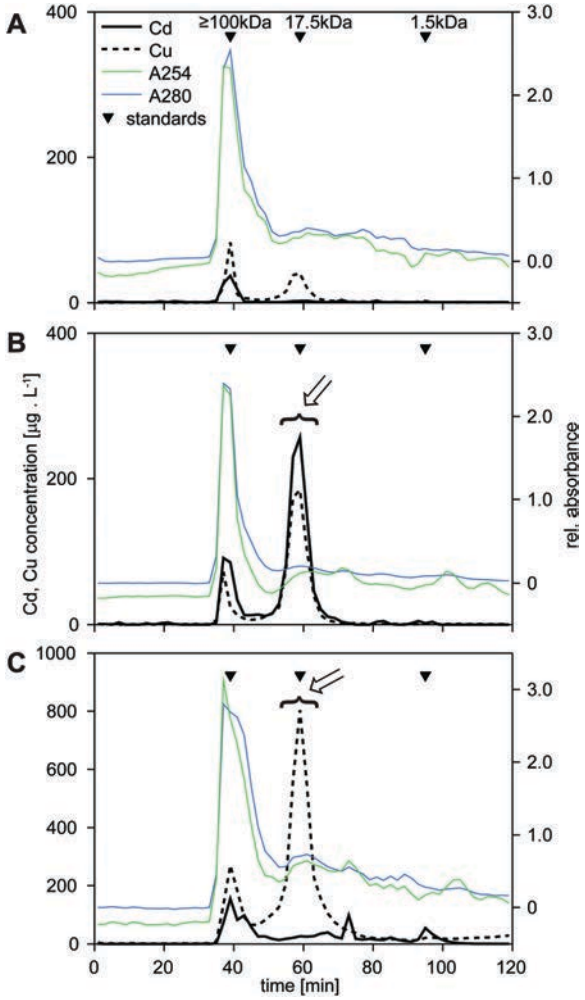


Fig. 4 Gel permeation chromatography (Sephacryl S-100, 25 × 300 mm) profiles of midgut gland homogenate supernatants from control (A), Cd (B) and Cu-exposed *A. vulgaris* (C), showing absorptions at 280 and 254 nm, as well as concentrations of Cd and Cu, as specified in Fig. 4A. Elution peaks of calibration standards (blue dextran, ≥100 kDa; myoglobin, 17.5 kDa; and vitamin B12, 1.5 kDa) are marked by inverted black triangles above the elution profiles. Fractions collected and pooled for subsequent purification (Fig. 4B and C) are indicated by braces below the empty arrow.

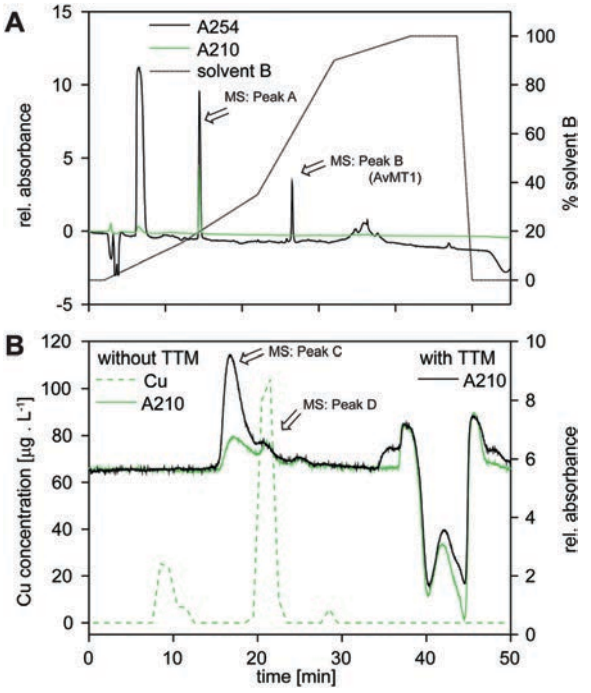


Fig. 5 Last purification steps by High Performance Liquid Chromatography (HPLC) of Cd and Cu-containing MT fractions from Cd-exposed (A) and Cu-exposed (B) *Arion vulgaris*, with MT peaks (arrows) purified for subsequent Mass spectrometry (MS). (A) Denaturing Reversed phase HPLC (0.1% TFA in 60% acetonitrile) of Cd fractions derived from ion exchange chromatography and HPLC (Fig. S1A and B, ESI†), showing metal-free MT peaks due to acidic conditions, with absorptions at 254 and 280 nm, and elution gradient. Open arrows point to the two peaks (A and B) recovered for MS, with only peak B containing AvMT1 verified by *de novo* sequencing (see Material and methods). (B) Elution profiles of Cu containing pooled fractions from previous ion exchange chromatography (Fig. S2, ESI†), showing absorption at 210 nm in samples treated with the Cu chelator ammonium-tetrathiomolybdate (TTM, black line) or absorption at 210 nm (green line) and Cu concentrations (green dotted line) in samples without TTM. Open arrows point to peaks C and D, collected for subsequent *de novo* sequencing by MS.

sequence is more similar to stlylommatophoran CdMT isoforms, while AvMT2 is more similar to the CuMT and CdCuMT isoforms. Correspondingly, AvMT1 shares with other stylommatophoran CdMTs its elevated K : N ratio with values typically above 1, whereas AvMT2 exhibits, like most other CuMTs, a K : N ratio below 1 (Fig. 7A). In agreement with these observations, a cladogram (Fig. 7B) including 20 metal-specific and non-specific MTs from 10 species of stylommatophoran gastropods (7 of them from snails and 3 from slugs) shows that AvMT1 clusters with CdMTs identified from the arionoid and limacoid slugs (bootstrap value 81%; Fig. 7B), within the stylommatophoran Cd-specific MT group. In contrast, AvMT2 appears to be the slug-specific sister group (bootstrap value 99%; Fig. 7B) of the *Stylommatophora* CuMT and CdCuMT clades (bootstrap value 95%; Fig. 7B).

Discussion

A. vulgaris copes with excessive metal exposure by specific accumulation in the midgut gland

In a mass spectrometric imaging study applying laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS), Becker *et al.* (2007) explored the distribution of some trace elements (Se, Zn, Cu) within longitudinal sections of an arionid slug species. They found a rather inhomogeneous distribution for the trace elements detected, with elevated signals for Cu in the heart and the mucous ventral skin, whereas Zn was mainly located in the digestive gland and the dorsal skin.[44] A rather inhomogeneous distribution of metals (Cu, Zn, Fe, Mn, Pb and Cd) between different tissues was also reported for *Arion rufus* from a supposedly low contaminated semi-rural site in Northern Italy.[27] Such an inhomogeneous distribution of metallic trace elements in slugs may reflect different cellular and tissue-specific affinities of metals under normal physiological conditions. However, the situation changes under metal stress. As indicated in the present study,

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

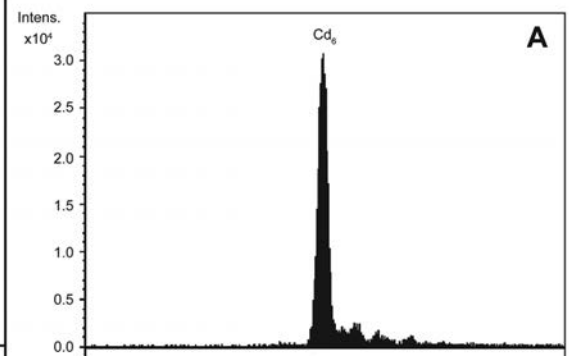
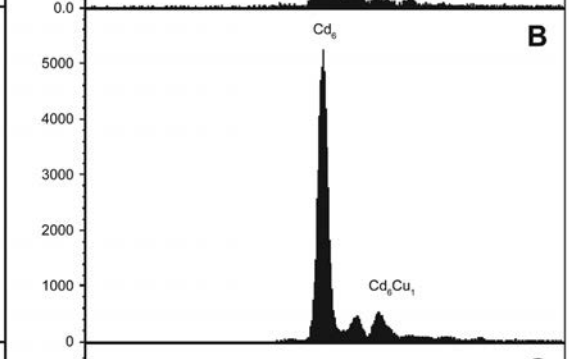
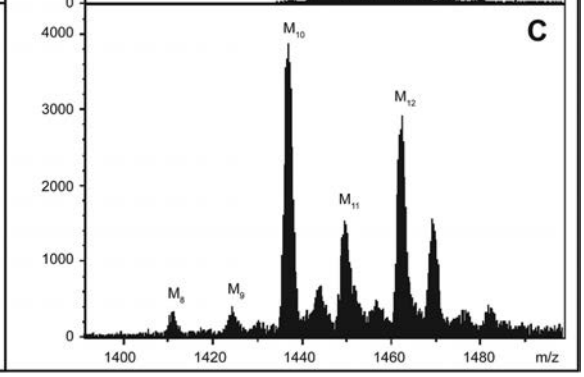
Metal supplemented media	*ICP-AES quantification*[a]	*ESI-MS data*[b]
300 μM $CdCl_2$	Protein: 0.58 10^{-4} M Cd/protein: 4.7 Cu/protein: *LDL* Zn/protein: *LDL*	
150 μM $CdCl_2$ + 150 μM $CuSO_4$	Protein: 0.82 10^{-4} M Cd/protein: 4.9 Cu/protein: 0.78 Zn/protein: *LDL*	
500 μM $CuSO_4$	Protein: 0.48 10^{-4} M Cd/protein: *LDL* Cu/protein: 10.9 Zn/protein: 0.9	

Fig. 6 Analytical characterization of the metal complexes formed by the recombinantly expressed AvMT1. Shown are metal concentrations applied to the expressing *E. coli* BL21 cultures (Cd^{2+}, Cu^{2+} or mixtures of Cd^{2+} and Cu^{2+}) (first column), protein concentrations calculated from the S content of the solutions and metal/protein ratios of the recovered complexes (second column), as well as signals of resulting ESI-MS spectra (last column) after Cd exposure (A), exposure to an equimolar proportion of Cd and Cu (B) and exposure to Cu alone (C). The metal complex species were detected at neutral pH, showing the +5 charge state. Abbreviations: LDL, lower than detection limit; M, mixture of Cu + Zn due to the impossibility to differentiate between the two metals.

metal-exposed individuals of *A. vulgaris* apparently accumulate huge amounts of Cd and Cu predominantly in the midgut gland (see Fig. 1A and C), followed distantly by some accumulation in the gut tissue (Fig. 1B and D), and some minor increase of Cu concentrations in the mantle of Cu-exposed slugs (Fig. 1D). Since the metals were provided to the slugs *via* feeding, the increased metal concentrations in the gut epithelium may be a consequence of metal uptake from the gut bolus. Eventually, excess amounts of Cd and Cu are diverted into the slug s midgut gland for storage and detoxification. This is in line with previous studies reporting a high accumulation capacity of the midgut gland for metallic trace elements (Cd, Cu, Zn, Hg) in various slug species like *Arion ater*,[10,45,46] *Arion lusitanicus*,[11,33] and *Deroceras reticulatum*.[47] Thus, with respect to Cd, terrestrial slugs behave like most stylommatophoran snails studied so far, where the midgut gland consistently appears to be the central turntable of Cd metabolism and detoxification.[12,48 50] A somewhat different picture is observed for Cu, which in most Cu-exposed shell-bearing land snails is distributed in contrast to *A. vulgaris* among different tissues, including the mantle and the foot, besides the midgut gland.[48,51]

In spite of the enormous Cd and Cu load in the slug s midgut gland achieved in the present study, all animals survived without visible signs of decreased activity. Obviously, this high metal tolerance[29] can be attributed to efficient

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

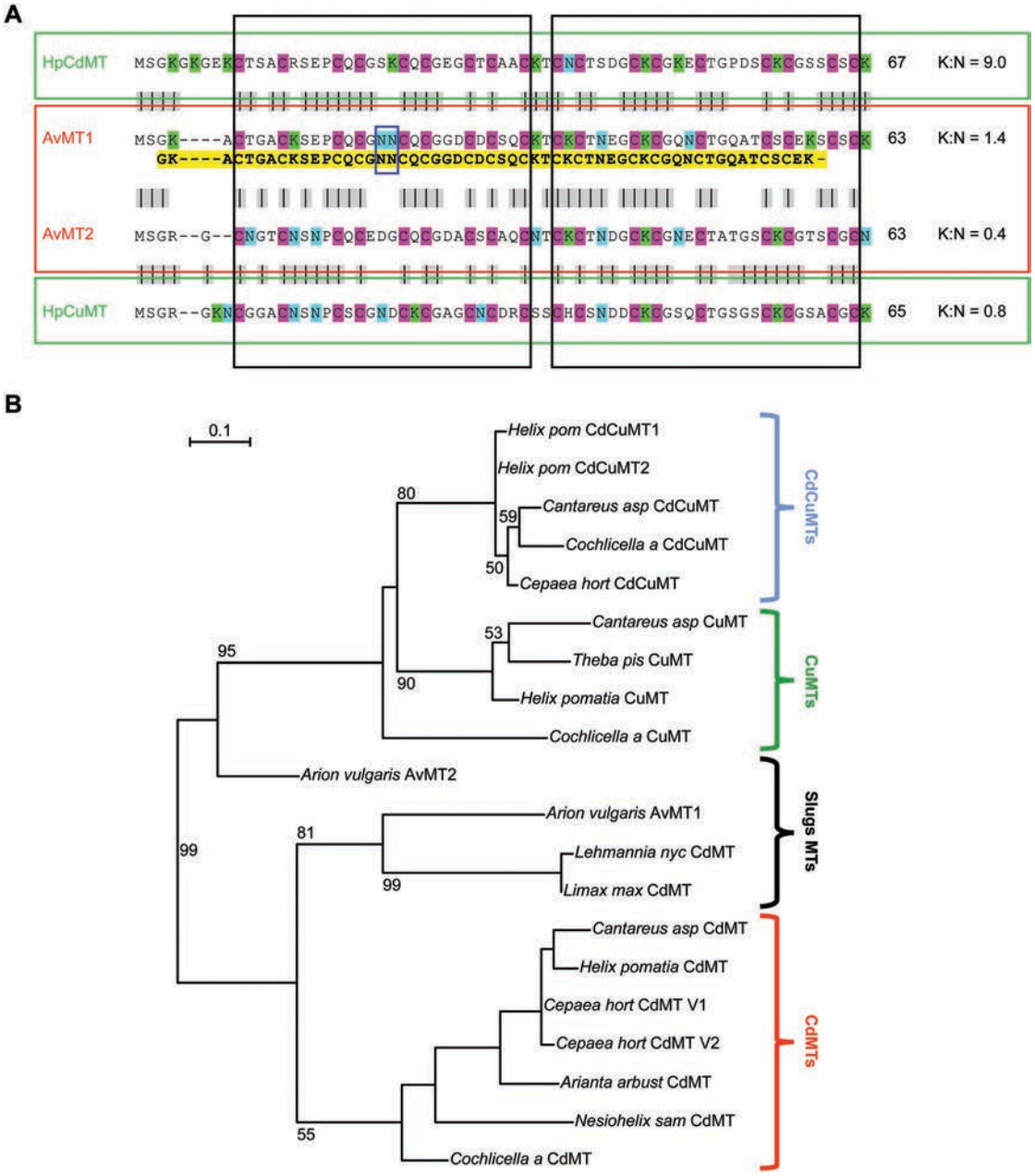


Fig. 7 Phylogenetic allocation of AvMT1 and AvMT2 within the sub-families of *Stylommatophora* MTs. (A) Reciprocal alignment of amino acid sequences of AvMT1 and AvMT2 (red-framed box), and of AvMT1 *versus* sequences of *H. pomatia* CdMT (green-framed upper box), or AvMT2 *versus* *H. pomatia* CuMT (green-framed lower box). Identical residues between the amino acid chains are indicated by grey-shadowed dashes. The sequence of AvMT1 shaded in yellow was identified by *de novo* protein sequencing. Conserved C (cysteine) positions are shadowed in pink, K (lysine) positions in green, and N (asparagine) positions in blue. Also reported on the right-hand margin are the amino acid length of the respective MT peptides, and their K/N ratios. (B) Maximum likelihood cladogram of *Stylommatophora* MTs, including Cd-specific CdMTs (indicated by red brace), Cu-specific CuMTs (green brace) and less specific Cd/CuMTs (blue brace), with AvMT1 and AvMT2 of *A. vulgaris* together with other slug MTs (black brace) at an intermediate position between metal-specific CdMTs and CuMTs. Only bootstrap values ≥50 are shown. For tree calculation parameters, see material and methods.

mechanisms of detoxification in the midgut gland cells.[12] In fact, at the cellular and tissue-specific level, slugs react to metal-induced stress by extensive cell-specific renovation and restructuring of the midgut gland epithelia, apparently embedded into a rhythmic pattern of activity.[28,52] Arionid slugs share this cellular plasticity in response to environmental challenges or stressors with many other gastropod species from marine or terrestrial environments, like the intertidal periwinkle, *Littorina littorea*[53,54] or the pulmonate land snail *H. pomatia*.[55] Upon metal exposure, in particular, such short-term cellular alterations go hand in hand with the activation of efficient buffering systems in order to temper fluctuating concentrations of unwanted metal ions that are potentially liberated and relocated during cellular processes of degradation, cell death

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

and renovation. It may be in this context that MTs are of essential significance for the survival of metal-exposed snails and slugs.[56]

Two *MT* isoform genes with differential metal and cell-specific transcription patterns

Two MT cDNAs were identified from the midgut gland of *A. vulgaris*, called *AvMT1* and *AvMT2* (see Fig. 7). As revealed in the present study by qRT PCR, *AvMT1* is, with an up to 12-fold induction compared to control levels, highly responsive to Cd exposure (Fig. 2A and C). This suggests that it may play an important role in Cd detoxification of the slug s midgut gland by Cd-dependent transcriptional activation. At the same time, *AvMT1* is activated upon Cu exposure, too, up to a level about 8-fold higher than in the control state (Fig. 2B and D). This is a feature that was occasionally observed in some gastropod *CuMT* genes,[21] and indicates that *AvMT1* may also be involved in Cu binding and detoxification. Hence, this gene exhibits ambivalent metal induction behaviour, and seems to combine the transcriptional response pattern of stylommatophoran metal-specific *MT* genes, with *CdMT* genes being responsive to Cd^{2+} on the one hand, and some Cu-responsive *CuMT* genes on the other.[21]

The presence of *MT* isoform genes with differential metal-specific transcription patterns is a common phenomenon in animals. In most cases, this genetic multiplicity is reflected in splitting of functions among the different expressed isoforms, either in a cell or organ-specific, or in a metal-specific manner.[16,57,58] In a number of terrestrial gastropod species, particularly, distinct *MT* isoform genes encoding for proteins with different metal-specific preferences, are transcribed differentially in accordance with their metal-specific functions.[14]

Localization of *AvMT1* transcripts by *in situ* hybridization shows that in Cd-exposed slugs, this gene is expressed in two cell types of the midgut gland, namely the digestive cells and the excretory cells of the midgut gland tubular epithelium (Fig. 3). Calcium cells constitute the third cell type present in this epithelium.[28] They were devoid of *AvMT1* mRNA. The expression in digestive and excretory cells of midgut gland *tubuli* is clearly an indication of the detoxifying function of *AvMT1*, since these two gastropod cell types are involved in absorption and degradation of digestive nutrients and in storage and excretion of cellular waste products.[59] A fourth cell type present in the midgut gland connective tissue, the so-called rhogocytes, was also devoid of *AvMT1* transcripts (Fig. 3). This is interesting, considering that true CuMTs of stylommatophoran gastropods are exclusively expressed in rhogocytes, where they may exert a homeostatic function, possibly in connection with hemocyanin synthesis.[39,60] This is not the case for *AvMT1*. Thus, one can conclude that overall, the role of *AvMT1* in *A. vulgaris* may prevalently be important in a detoxifying rather than a homeostatic regulatory context, for both, Cd and Cu. Such a detoxifying function may be all the more important, as metal exposure in arionid slugs is generally accompanied by extensive renovation and reshuffling of midgut gland cell types,[52,61] which is a feature that slugs share with various other gastropod species.[47,55,56]

In contrast to *AvMT1*, *AvMT2* appears to be unresponsive to Cd exposure (Fig. 2B), and less inducible by Cu treatment than *AvMT1*, reaching -fold induction levels of up to 2 3, at most (Fig. 2D). This already indicates that its significance for Cd and Cu handling and detoxification in the slug midgut gland may be subsidiary, compared with *AvMT1*. This suggestion is supported by the fact that at the histochemical level, no *AvMT2* transcripts at all were detected in any of the midgut gland cell types upon any kind of metal exposure. This does not mean that *AvMT2* transcripts may be completely absent, but rather demonstrates the limitation in sensitivity of the applied *in situ* hybridization approach at very low concentrations of *AvMT2* mRNA. However, this underpins the assumption that *AvMT2* plays a minor role in the processing of Cd and Cu metabolism of *A. vulgaris*, compared to *AvMT1*. The possibility remains that *AvMT2* may be present at higher transcript levels in other slug organs or during embryonic development, as shown, for example, for a non-specific *MT* isoform gene in juveniles of the helicid snail *C. aspersus*.[62]

One MT isoform, two metals: how to combine Cd specificity with Cu binding requirements *in vivo*

The results of MT recovery and identification by means of chromatographic separation and MS techniques suggest that AvMT1 is the only expressed MT isoform detectable in the midgut gland of *A. vulgaris*. After repeated chromatographic fractionation and purification steps (Fig. 4, 5A and Fig. S1, ESI†), two distinct and homogeneous fractions were isolated from the midgut gland supernatant of Cd-exposed slugs (Fig. 5A, peaks A and B). These fractions were subjected to *de novo* sequencing *via* MS, yielding eventually for peak B an amino acid sequence lacking one N-terminal and five C-terminal amino acid residues, but otherwise matching to 100% with the predicted primary structure derived from *AvMT1* cDNA (Fig. 7A). Interestingly, a small Cd peak in the MT-containing molecular weight range was also observed after gel chromatography of midgut gland supernatants from control slugs (Fig. 4A), suggesting that small amounts of AvMT1 may also be expressed under control conditions. This may be due to some low-level, pre-existing contamination of slugs collected from the field. Recombinant expression of this isoform in *E. coli* BL21 cultures with different metal incubations (Cd^{2+}, Cu^{2+} or mixtures of Cd^{2+} and Cu^{2+}) and subsequent analysis of binding features of the resulting MT-metal complexes (see Material and methods section) suggested that AvMT1 is a predominantly Cd-specific MT isoform (Fig. 6). This means that with respect to its Cd binding preferences, it behaves in a similar way as the true CdMT isoforms so far characterized from gastropod species of different clades, including *H. pomatia*,[63] *C. aspersus*[21] or *L. littorea*.[64] As expected, and particularly after Cu exposure, the MT-containing fractions after gel chromatography were always associated with some Cu at moderately high to very high concentrations (Fig. 4B). Therefore, our first assumption was that the respective purified fractions were composed of a mixture of both, AvMT1 and AvMT2, with the presumption that the detected amounts of Cu were bound to the supposedly expressed

This journal is © The Royal Society of Chemistry 2018

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

AvMT2. In spite of many efforts, we were never successful in identifying from these fractions an AvMT2 sequence, even after further purification upon ion exchange chromatography or HPLC (Fig. S2, ESI† and Fig. 5B). Instead, in one of these isolated fractions traces of AvMT1 were retrieved by MS analysis (Fig. 5B, peak D). These findings led us to the conclusion that AvMT2, if expressed in the midgut gland of *A. vulgaris* at all, may there be present only in trace amounts below detection limits of MS analysis. Hence, the metal binding role of AvMT1 *in vivo* must be re-considered, suggesting that in spite of its prevalent Cd-specific character, this isoform might also be able to bind some Cu. In fact, results from recombinantly expressed AvMT1 in *E. coli* cells exposed to an equimolar mixture of both, Cd and Cu, shows that besides Cd^{2+}, AvMT1 is also capable of binding low amounts of Cu^{+} at a stoichiometric ratio of 1 Cu : 5 Cd ions (Fig. 6B). This ratio may even increase under conditions *in vivo*, where Cu (as an essential trace element) is normally and even after Cd exposure present in higher concentrations than the non-essential Cd. It appeared that at higher Cu : Cd ratios up to 90 : 10, the yield of the binding reaction decreased (data not shown). We believe that this enforced association of the Cd-specific AvMT1 with some Cu^{+} may happen at the expense of its structural firmness, with the consequence that probably both, Cd^{2+} and Cu^{+}, but particularly the latter one, become more loosely bound to the protein structure. This may explain why for both metals, but largely for Cu, high losses in recovery were observed through all chromatographic purification steps until final isolation (Fig. S3, ESI†). At the same time, this may also explain why some Cd and Cu are always found to be associated with high molecular protein fractions. In previous studies with *A. lusitanicus*, this phenomenon has been described as a ``spillover effect[33] and was explained due to an assumed exhaustion of the detoxification capacity of the slug s MT system after excessively high metal exposure. In the light of the present study, a second possible interpretation may be that the observed ``spillover is an artificial effect. It may occur when, during purification loosely, thus rather mobile, metal fractions swap from AvMT1 to other proteins and components present in the homogenate mixture. *In vivo*, the loss of Cu from AvMT1 may be compensated by the uptake of the available Cu ions by the spherical concretions of calcium cells (perhaps in the form of a low-molecular weight Cu complex), which seem to be the primary site of Cu detoxification in arionid slugs.[52]

Overall, our study has shown that in spite of two *MT* genes (*AvMT1* and *AvMT2*) present and transcribed in the midgut gland of *A. vulgaris*, only one of these genes (*AvMT1*) is expressed at detectable levels. The observation that transcripts of *MT* isoform genes are expressed at very low or undetectable concentrations at the protein level, as for AvMT2 in the present study, is a phenomenon that has been described for MT isoforms from other gastropod species, too.[22] Although AvMT1 is predominantly Cd-specific, it can also bind some Cu^{+}, depending apparently on the Cu:Cd ratio accumulated in the midgut gland. This makes sense, considering that AvMT1 is actually the only MT isoform available for metal detoxification in the slug s midgut gland. Hence, the corresponding transcriptional activation of the *AvMT1* gene in response to both, Cd and Cu exposure, is an adaptation to its dual metal binding role.

The ambivalent features of AvMT1 and AvMT2 as a reflection of their primary structures

The amino acid sequences of AvMT1 and AvMT2 show characteristic features of molluscan MTs. In particular, the two proteins possess a high cysteine content, while aromatic amino acids or histidine are missing. In addition, they contain the MT-specific conserved Cys motifs as observed in nearly all MT sequences throughout the animal kingdom.[65,66] Both proteins are similar in size consisting of 63 amino acids with a calculated molecular mass of 6.41 kDa (AvMT1) and 6.26 kDa (AvMT2), respectively. From their number and arrangement of Cys residues we conclude by homology,[67] that they must possess a dumbbell-like two-domain structure. A close inspection of the HpCdMT and AvMT1 sequences reveals that AvMT1 is more closely related to the stylommatophoran CdMTs, while AvMT2 resembles more closely the group of stylommatophoran CuMTs. In agreement with this, the K : N (lysine to asparagine) ratio of AvMT1 is 1.4, in a like manner to the HpCdMT K : N ratio, which with a value of 9 is even higher. In contrast, the K : N ratio of AvMT2 is, similarly with that of other CuMTs, lower than 1 (Fig. 7A). This is probably significant, since it was shown previously that in metal-specific gastropod MTs, a preponderance, among others, of K (lysine) over N (asparagine) is typical for Cd-specific isoforms, whereas a decreased proportion of K in favour of N turns the metal preference of gastropod MTs toward Cu specificity.[68] The fact that in the cladogram of Fig. 7B, the MTs of other slug species (*Lehmannia nyctelia* and *Limax maximus*) belong to the same sub-cluster as AvMT1, strongly suggests that they share Cd specificity with AvMT1.

Altogether, these results explain to a certain degree why AvMT1 behaves at both, the gene and the protein levels, like a CdMT that has tentatively assumed some Cu-specific tasks. Hence, the often-quoted concept of ``multifunctionality applied to the MT family as a whole may be realized, in the case of *A. vulgaris*, by this one uniquely behaving AvMT1 isoform.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We would like to dedicate this article to Silvia Atrian, with whom we had the privilege to cooperate for many years. She was to us an excellent scientist and expert in metallothionein research. Moreover, she was an obliging colleague, a gentle person, and an unforgettable friend. We will keep her memory in our minds for the rest of our life. This work was funded by the Fonds zur Forderung der wissenschaftlichen Forschung (Austrian Science foundation, FWF), Project ref. I 1482-N28 (DACH) granted to Reinhard Dallinger, and by a doctorate scholarship (2017/2/Bio-3) of the University of Innsbruck,

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

granted to Martin Dvorak. Authors from both Barcelona universities are recipients of MINECO-FEDER grants (BIO2015-67358-C2-1-P to SA and RA, and BIO2015-67358-C2-2-P to MC). Ricard Albalat was also supported by SGR2017-1665 from Generalitat de Catalunya. Mario Garcıa-Risco acknowledges the PIF grant to the Universitat Autonoma de Barcelona. We thank Thomas Ostermann for help with the transcriptome assembly and Julia Wunderer with *in situ* hybridization experiments. The computational results presented have been achieved in part using the HPC infrastructure of the University of Innsbruck. We also thank the Servei d Analisi Quımica (SAQ) de la Universitat Autonoma de Barcelona (ICP-AES, CD, UV-vis, ESI-MS) for allocating instrument time.

References

1 C. M. Wade, P. B. Mordan and F. Naggs, *Biol. J. Linn. Soc.*, 2006, **87**, 593 610.
2 V. Adam, I. Fabrik, R. Kizek, V. Adam, T. Eckschlager, M. Stiborova and L. Trnkova, *Trends Anal. Chem.*, 2010, **29**, 409 418.
3 M. A. Zemanova, E. Knop and G. Heckel, *Mol. Ecol.*, 2016, **25**, 5747 5764.
4 M. Pfenninger, A. Weigand, M. Balint and A. Klussmann-Kolb, *Evol. Appl.*, 2014, **7**, 702 713.
5 B. A. Hatteland, T. Solh y, C. Schander, T. Von Proschwitz and L. R. Noble, *Malacologia*, 2015, **58**, 303 321.
6 M. A. Zemanova, E. Knop and G. Heckel, *Sci. Rep.*, 2017, **7**, 1 11.
7 D. Godan, *Pest slugs and snails – Biology and control*, Springer-Verlag Berlin, Heidelberg, 1983.
8 A. Peltanova, A. Petrusek, P. Kment and L. Jurickova, *Biol. Invasions*, 2012, **14**, 759 764.
9 T. Bulat, R. Smidak, F. J. Sialana, G. Jung, T. Rattei, M. Bilban, H. Sattmann, G. Lubec and J. Aradska, *PLoS One*, 2016, **11**, e0150614.
10 M. P. Ireland, *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.*, 1981, **68**, 37 41.
11 R. Dallinger, H. H. Janssen, A. Bauer-Hilty and B. Berger, *Comp. Biochem. Physiol., Part C: Pharmacol., Toxicol. Endocrinol.*, 1989, **92**, 355 360.
12 R. Dallinger, in *Ecotoxicology of metals in invertebrates*, ed. R. Dallinger and P. S. Rainbow, Lewis Publisher, Boca Raton, USA, 1993, pp. 245 289.
13 R. Dallinger, B. Berger, R. Triebskorn-Kohler and H.-R. Kohler, in *The Biology of Terrestrial Molluscs*, ed. G. Barker, CAB International, Wallingford, England, 2001, pp. 489 525.
14 R. Dallinger, B. Berger, P. Hunziker and J. H. R. Kagi, *Nature*, 1997, **388**, 237 238.
15 O. Palacios, A. Pagani, S. Perez-Rafael, M. Egg, M. Hockner, A. Brandstatter, M. Capdevila, S. Atrian and R. Dallinger, *BMC Biol.*, 2011, **9**, 4.
16 M. Capdevila, R. Bofill, O. Palacios and S. Atrian, *Coord. Chem. Rev.*, 2012, **256**, 46 62.
17 D. H. Hamer, *Annu. Rev. Biochem.*, 1986, **55**, 913 951.
18 M. Sato and I. Bremner, *Free Radical Biol. Med.*, 1993, **14**, 325 337.
19 T. E. English, *J. Exp. Biol.*, 2003, **206**, 2517 2524.
20 J. H. R. Kagi and A. Schaffer, *Biochemistry*, 1988, **27**, 8509 8515.
21 M. Hockner, K. Stefanon, A. De Vaufleury, F. Monteiro, S. Perez-Rafael, O. Palacios, M. Capdevila, S. Atrian and R. Dallinger, *Biometals*, 2011, **24**, 1079 1092.
22 V. Pedrini-Martha, R. Schnegg, P. E. Baurand, A. De Vaufleury and R. Dallinger, *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.*, 2017, **199**, 38 47.
23 R. Dallinger and B. Berger, *Sci. Total Environ.*, 1993, **134**, 607 615.
24 B. Berger, R. Dallinger, P. Gehrig and P. E. Hunziker, *J. Biochem.*, 1997, **328**, 219 224.
25 R. W. Greville and A. J. Morgan, *J. Molluscan Stud.*, 1990, **56**, 355 362.
26 M. P. Ireland, *Environ. Pollut.*, 1979, **20**, 271 278.
27 C. Menta and V. Parisi, *Environ. Pollut.*, 2001, **115**, 205 208.
28 B. Zaldibar, I. Cancio, M. Soto and I. Marigomez, *Chemosphere*, 2007, **70**, 144 154.
29 R. W. Greville and A. J. Morgan, *Environ. Pollut.*, 1991, **74**, 115 127.
30 H. Greven, U. Ruther and J. D Haese, *Andrias*, 2001, **15**, 193 204.
31 R. W. Greville and A. J. Morgan, *Environ. Pollut.*, 1989, **59**, 287 303.
32 D. Crespo, M. Bolon, J. R. Aboal, J. A. Fernandez and A. Carballeira, *Environ. Sci. Pollut. Res.*, 2015, **22**, 7622 7627.
33 H. H. Janssen and R. Dallinger, *Arch. Environ. Contam. Toxicol.*, 1991, **20**, 132 137.
34 R. Dallinger, B. Berger, C. Gruber, P. Hunziker and S. Sturzenbaum, *Cell. Mol. Biol.*, 2000, **46**, 331 346.
35 R. Dallinger, B. Berger, P. E. Hunziker, N. Birchler, C. R. Hauer and J. H. R. Kagi, *Eur. J. Biochem.*, 1993, **216**, 739 746.
36 R. Dallinger, M. Chabicovsky and B. Berger, *Environ. Toxicol. Chem.*, 2004, **23**, 890 901.
37 M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev, *Nat. Biotechnol.*, 2011, **29**, 644 652.
38 M. Plusquin, K. De Mulder, F. Van Belleghem, O. DeGheselle, N. Pirotte, M. Willems, A. Cuypers, W. Salvenmoser, P. Ladurner, T. Artois and K. Smeets, *Environ. Toxicol.*, 2016, **31**, 1217 1228.
39 M. Chabicovsky, H. Niederstatter, R. Thaler, E. Hodl, W. Parson, W. Rossmanith and R. Dallinger, *Toxicol. Appl. Pharmacol.*, 2003, **190**, 25 36.
40 J. Bongers, C. D. Walton, D. E. Richardson and J. U. Bell, *Anal. Chem.*, 1988, **60**, 2683 2686.
41 D. Fabris, J. Zaia, Y. Hathout and C. Fenselau, *J. Am. Chem. Soc.*, 1996, **118**, 12242 12243.
42 M. Zhao, T. Wang, M. J. Stewart, U. Bose, S. Suwansa-ard, K. B. Storev and S. F. Cummins, *Mol. Ecol. Resour.*, 2018, **18**, 147 158.
43 M. Gouy, S. Guindon and O. Gascuel, *Mol. Biol. Evol.*, 2010, **27**, 221 224.

Published on 04 October 2018. Downloaded by Universitat de Barcelona on 10/6/2018 12:46:16 PM.

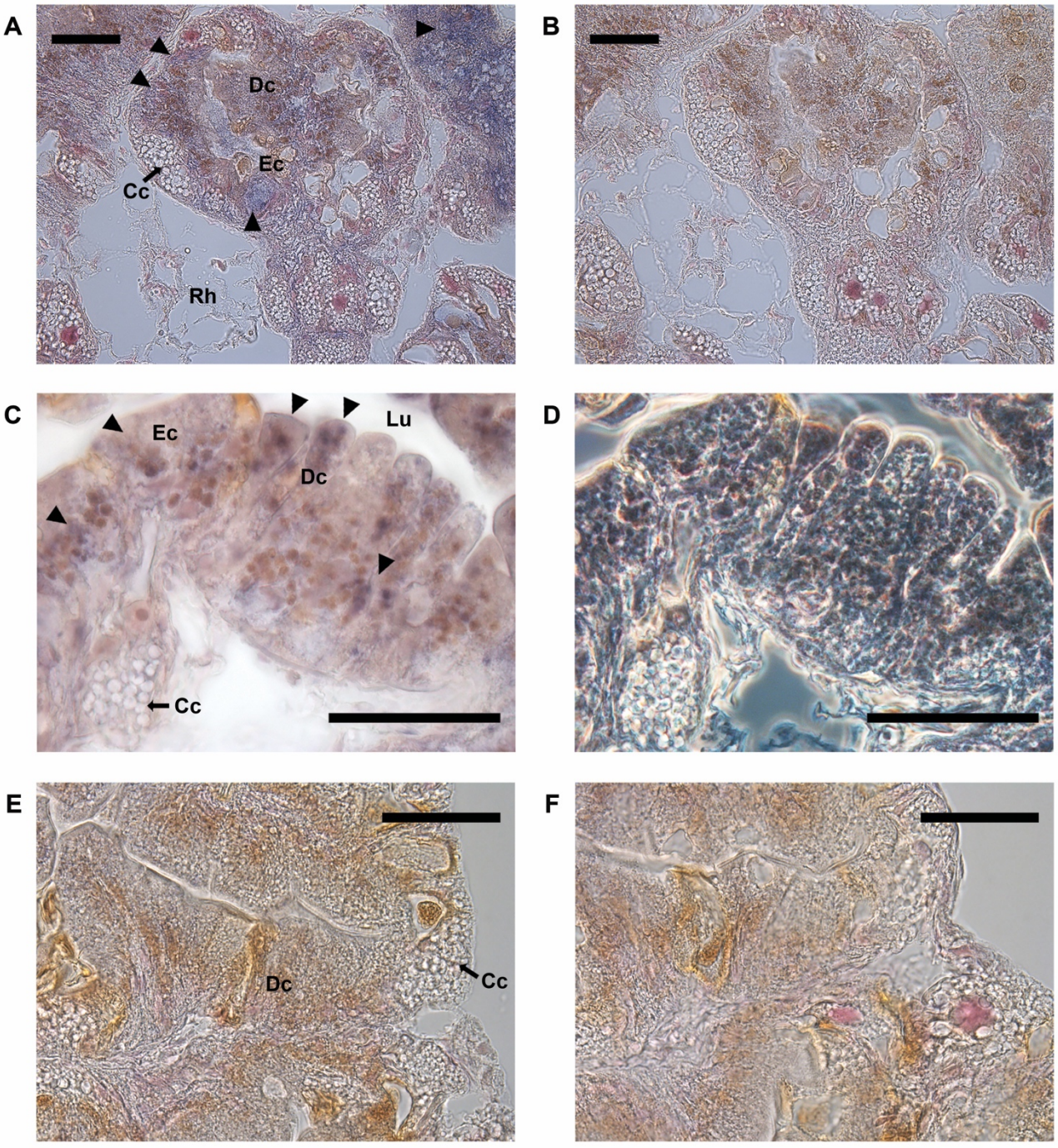
44 J. S. Becker, A. Matusch, C. Depboylu, J. Dobrowolska and M. V. Zoriy, *Anal. Chem.*, 2007, **79**, 6074 6080.
45 M. P. Ireland, *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.*, 1982, **73**, 217 221.
46 A. Recio, J. A. Marigomez, E. Angulo and J. Moya, *Bull. Environ. Contam. Toxicol.*, 1988, **41**, 858 864.
47 R. Triebskorn and H.-R. Kohler, *Environ. Pollut.*, 1996, **93**, 327 343.
48 R. Dallinger and W. Wieser, *Comp. Biochem. Physiol., Part C: Pharmacol., Toxicol. Endocrinol.*, 1984, **79**, 117 124.
49 R. Dallinger, B. Lagg, M. Egg, R. Schipflinger and M. Chabicovsky, *Ecotoxicology*, 2004, **13**, 757 772.
50 F. Hispard, D. Schuler, A. De Vaufleury, R. Scheifler, P. M. Badot and R. Dallinger, *Environ. Toxicol. Chem.*, 2008, **27**, 1533 1542.
51 B. Berger and R. Dallinger, *Oecologia*, 1989, **79**, 60 65.
52 J. A. Marigomez, E. Angulo and J. Moya, *Bull. Environ. Contam. Toxicol.*, 1986, **36**, 600 607.
53 D. M. Lowe, M. N. Moore and J. W. Readman, *Mar. Environ. Res.*, 2006, **61**, 457 470.
54 B. Zaldibar, I. Cancio and I. Marigomez, *Aquat. Toxicol.*, 2007, **81**, 183 196.
55 E. Hodl, E. Felder, M. Chabicovsky and R. Dallinger, *Cell Tissue Res.*, 2010, **341**, 159 171.
56 D. Benito, M. Niederwanger, U. Izagirre, R. Dallinger and M. Soto, *Int. J. Mol. Sci.*, 2017, **18**, 1815.
57 S. Lemoine, Y. Bigot, D. Sellos, R. Cosson and M. Laulier, *Mar. Biotechnol.*, 2000, **2**, 195 203.
58 M. Vasak, *J. Trace Elem. Med. Biol.*, 2005, **19**, 13 17.
59 A. T. Sumner, *Q. J. Microsc. Sci.*, 1965, **106**, 173 192.
60 R. Dallinger, M. Chabicovsky, E. Hodl, C. Prem, P. Hunziker and C. Manzl, *Am. J. Physiol.: Regul., Integr. Comp. Physiol.*, 2005, **289**, R1185 R1195.
61 I. Marigomez, M. Soto and M. Kortabitarte, *Arch. Environ. Contam. Toxicol.*, 1996, **31**, 54 62.
62 P. E. Baurand, V. Pedrini-Martha, A. De Vaufleury, M. Niederwanger, N. Capelli, R. Scheifler and R. Dallinger, *PLoS One*, 2015, **10**, 1 14.
63 O. Palacios, S. Perez-Rafael, A. Pagani, R. Dallinger, S. Atrian and M. Capdevila, *J. Biol. Inorg. Chem.*, 2014, **19**, 923 935.
64 O. Palacios, E. Jimenez-Martı, M. Niederwanger, S. Gil-Moreno, O. Zerbe, S. Atrian, R. Dallinger and M. Capdevila, *Int. J. Mol. Sci.*, 2017, **18**, 1 16.
65 Y. Kojima, *Methods Enzymol.*, 1991, **205**, 8 10.
66 M. J. Jenny, S. L. Payton, D. A. Baltzegar and J. D. Lozier, *J. Mol. Evol.*, 2016, **83**, 110 125.
67 L. Schmielau, M. Dvorak, M. Niederwanger, N. Dobieszewski, V. Pedrini-Martha, P. Ladurner, J. Rodrıguez-Guerra Pedregal, J. D. Marechal and R. Dallinger, *Sci. Total Environ.*, 2018, **648**, 561 571.
68 S. Perez-Rafael, F. Monteiro, R. Dallinger, S. Atrian, O. Palacios and M. Capdevila, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 1694 1707.
69 A. Priyam, B. J. Woodcroft, V. Rai, A. Munagala, I. Moghul, F. Ter, M. A. Gibbins, H. Moon, G. Leonard, W. Rumpf and Y. Wurm, *bioRxiv*, 2015, 033142.

ANNEX III: SUPPLEMENTARY MATERIAL

METAL BINDING FUNCTIONS OF METALLOTHIONEINS IN THE SLUG ARION VULGARIS DIFFER FROM METAL-SPECIFIC ISOFORMS OF TERRESTRIAL SNAILS

The file includes:

Figure 3 (high-resolution imatge)


A
Dc
Ec
Cc
Rh
B
C
Ec
Lu
Dc
Cc
D
E
Dc
Cc
F

ANNEX IV:

Metallomics reveals a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins

Metallomics, 2020,12, 702

DOI: 10.1039/c9mt00259f

Metallomics

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Metallomics*, 2020, **12**, 702

Metallomics reveals a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins†‡

Reinhard Dallinger,*[ab] Oliver Zerbe,*[c] Christian Baumann,[c] Bernhard Egger,[a] Merce Capdevila,[d] Oscar Palacios,[d] Ricard Albalat,[e] Sara Calatayud,[e] Peter Ladurner,[ab] Birgit C. Schlick-Steiner,[f] Florian M. Steiner,[f] Veronika Pedrini-Martha,[a] Reinhard Lackner,[a] Herbert Lindner,[g] Martin Dvorak,[a] Michael Niederwanger,[a] Raimund Schnegg[a] and Silvia Atrian§[e]

The tiny contribution of cadmium (Cd) to the composition of the earth s crust contrasts with its high biological significance, owing mainly to the competition of Cd with the essential zinc (Zn) for suitable metal binding sites in proteins. In this context it was speculated that in several animal lineages, the protein family of metallothioneins (MTs) has evolved to specifically detoxify Cd. Although the multi-functionality and heterometallic composition of MTs in most animal species does not support such an assumption, there are some exceptions to this role, particularly in animal lineages at the roots of animal evolution. In order to substantiate this hypothesis and to further understand MT evolution, we have studied MTs of different snails that exhibit clear Cd-binding preferences in a lineage-specific manner. By applying a metallomics approach including 74 MT sequences from 47 gastropod species, and by combining phylogenomic methods with molecular, biochemical, and spectroscopic techniques, we show that Cd selectivity of snail MTs has resulted from convergent evolution of metal-binding domains that significantly differ in their primary structure. We also demonstrate how their Cd selectivity and specificity has been optimized by the persistent impact of Cd through 430 million years of MT evolution, modifying them upon lineage-specific adaptation of snails to different habitats. Overall, our results support the role of Cd for MT evolution in snails, and provide an interesting example of a vestigial abiotic factor directly driving gene evolution. Finally, we discuss the potential implications of our findings for studies devoted to the understanding of mechanisms leading to metal specificity in proteins, which is important when designing metal-selective peptides.

Received 23rd October 2019,
Accepted 9th March 2020

DOI: 10.1039/c9mt00259f

rsc.li/metallomics

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

Significance to metallomics

In the present study we apply gastropod (snail) metallothioneins at a lineage level as model molecules, trying to track the evolution, structural/functional optimization and diversification of metal-selectivity under the persistent influence of cadmium since early gastropod evolution. To this aim, we applied an `Eco -Metallomics approach including 74 MT sequences from 47 gastropod species, combining phylogenomic methods with molecular, biochemical, and spectroscopic techniques. This allows us to demonstrate that Cd binding selectivity paired with Cd-specific tasks has emerged repeatedly in Gastropoda clades since 430 million years ago. We believe that our article may be particularly significant to metallomics, because it demonstrates how differing techniques such as molecular and biochemical methods, combined with ecological and evolutionary approaches, can prove how a rare metallic trace element like cadmium has shaped the structure, metal-binding behavior and physiological function of an important protein family.

[a] *Department of Zoology, University of Innsbruck, Austria. E-mail: reinhard.dallinger@uibk.ac.at*
[b] *Center for Molecular Biosciences Innsbruck, Austria*
[c] *Department of Chemistry, University of Zürich, Switzerland. E-mail: oliver.zerbe@chem.uzh.ch*
[d] *Departament de Química, Universitat Autònoma de Barcelona, Spain*
[e] *Department of Genetics, University of Barcelona, Spain*
[f] *Department of Ecology, University of Innsbruck, Austria*
[g] *Division of Clinical Biochemistry, Innsbruck Medical University, Austria*

† We want to dedicate this work to our longtime cooperation partner, Silvia Atrian.
‡ Electronic supplementary information (ESI) available. See DOI: 10.1039/c9mt00259f
§ Deceased.

This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

Introduction

With a tiny amount of about 0.00001%, the contribution of cadmium (Cd) to the composition of the earth s crust is seemingly negligible.[1] In spite of this, the biological significance of Cd is distinctly higher, owing to its particular patchy distribution, enrichment and circulation in the biosphere.[2] In some diatom marine algae, for example, Cd has achieved an essential importance as a constituent of the algal enzyme carbonic anhydrase, a fact that has been explained by the relative preponderance of Cd at the cost of lowly available Zn in oceanic environments inhabited by these algae.[3,4] In most organisms, however, Cd is highly toxic at very low concentrations, due to its physico-chemical similarity and competition with zinc (Zn),[5] one of the most important essential trace elements. Because of this, most organisms have developed strategies for Cd handling and detoxification,[6] and it has been hypothesized that metallothioneins (MTs), a ubiquitous protein family with a high affinity to transition metal ions, may have been developed by organisms to clear this highly toxic metal.[7] Yet, this hypothesis has been questioned because of the apparent involvement of most MTs in a variety of functions, and their often heterometallic and metamorphic composition with binding affinities to different metal ions.[8–10] However, MTs form a huge and diverse gene superfamily present in most kingdoms of organisms, from bacteria through fungi, plants and animals.[11,12] This suggests that their origins may go back to the primal evolutionary roots of life on earth, although the metal preference of the ancestral MT remains unknown. In contrast to modern vertebrates some MTs at the roots of, for example, Chordata are Cd-selective, as recently reported for MTs of the tunicate *Oikopleura dioica*.[13] Cd- and Cu-selective MTs have also been discovered in several species of the ancient mollusk class of Gastropoda (snails and slugs).[14,15] This suggests that in early evolution of life, Cd-selectivity of MTs might have been more common than today, and this feature has evidently been preserved to the present in diverse animal clades while it disappeared in others.

To support this hypothesis, we have taken advantage of the Cd-specific gastropod MTs, which provide an ideal model system to study the evolutionary influence of Cd on MT evolution along more of 400 million years (MY) of Gastropoda diversification. Unlike many other modern animals, snails possess metal-selective MTs, such as Cd- (CdMTs) and Cu-selective (CuMTs) isoforms[15] that perform Cd- or Cu-specific tasks. Thereby they exhibit a straightforward relationship between metal binding features and related physiological functions. Interestingly, Cd-specific snail MTs bind this metal with a strength and exclusive preference hardly observed in any other protein family. They are expressed in a multitude of isoforms that vary in a clade-specific manner allowing us to compare and evaluate similar proteins and protein variants (and their metal-binding modifications) in a large number of species that have adapted to different habitats. These spread from marine through terrestrial to freshwater environments with significantly different Cd concentrations. This comparative approach is central to understand how MTs have been optimized for Cd binding during gastropod evolution by the continuous impact of Cd, and how its influence is modulated by habitat-specific environmental constraints.

In our work, we applied a comprehensive metallomics approach by characterizing 74 novel or known MT sequences from 47 species across all major gastropod clades.[14–23] We used phylogenomic methods based on next generation sequencing to obtain transcriptomic data for evolutionary analyses and construction of phylogenetic trees. We also analyzed neutral DNA markers to compare the resulting phylogenetic tree with MT-derived trees. In addition, we provide data on metal-selective features of recombinant snail MTs and their metal-binding domains, based on molecular, biochemical and spectroscopic methods. Our data indicate that Cd selectivity has evolved since 430 million years ago (MYA) in gastropod MTs through convergent evolution of metal-binding domains with diverging primary structures. We study the mechanisms by which their Cd binding features have been optimized, and illustrate how they have diversified into different kinds with altered or even lost metal selectivity through lineage-specific transition into novel habitats that differ in their natural Cd background concentrations. Overall, we have been able to demonstrate a continuous impact of Cd on evolution of one of the most important metal-binding protein families, and describe a paradigmatic case of how an abiotic factor directly drives gene evolution. Finally, we discuss possible implications of our findings to better understand how metal-selectivity has been achieved in nature, and how this knowledge can help in designing metal-selectivity in synthetic peptides.

Material and methods

Animal collection, rearing and Cd exposure

A list of gastropod species involved in experimental work for the present study along with methodical applications is reported in Table 1.

Individuals of *Alinda biplicata* and *Deroceras reticulatum* were collected in suburbs of Innsbruck (Tyrol, Austria) in 2017 and 2018. Individuals of *Patella vulgata* were collected in Barcelona, Spain in summer 2016. Snails of the helicid species *Cornu aspersum* were bought from a commercial dealer (Wiener Schnecken Manufaktur, Vienna, Austria), as were the aquatic species *Marisa cornuaretis*, *Anentome helena*, *Physa acuta* and *Aplysia californica* (Aquaristikzentrum Innsbruck, Tyrol, Austria). Adult individuals of *Lottia gigantea* were collected and kindly provided to us by Dr Douglas J. Eernisse (California State University, Fullerton, CA, USA).

For Cd exposure of *Cornu aspersum*, adult snails were acclimatized on garden earth substrate containing lime powder ($CaCO_3$) in groups of 30 individuals each under stable conditions in a climate chamber (18 °C, 12 h light/dark cycle) and were fed regularly with uncontaminated lettuce (*Lactuca sativa*) under moistened conditions for one week. For Cd exposure, control snails were fed with uncontaminated lettuce whereas Cd-exposed snails were fed with Cd-enriched lettuce which had been incubated for one hour in a $CdCl_2$-solution containing 2 mg l^{-1} Cd^{2+}.[24] After five days of exposure, five individuals of each group were sacrificed.

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

Table 1 List of gastropod species and their use for different methodical applications (red check marks) within the present study. Reported are all species acquired (first column) and their utilization for Cd exposure (second column), RNA sequencing and transcriptome assembly (second column), RNA isolation and cDNA transcription (third column), quantitative real-time PCR (fourth column), protein purification from tissues *in vivo* (fifth column), recombinant expression (sixth column), MS analysis (seventh column), NMR analysis and metal titration (eighth column), and construction of neutral marker phylogeny (ninth column)

Animal collection, purchasing and rearing (species)	Cd exposure	RNA seq and transcriptome assembly	RNA isolation and cDNA	Quantitative RT-PCR	*In vivo* protein purification	Recombinant expression	MS analysis	NMR and metal titration	Neutral marker phylogeny
Lottia gigantea			√			√	√		√
Patella vulgata		√	√						√
Neritina pulligera		√	√						√
Littorina littorea						√		√	√
Pomatias elegans									√
Marisa cornuarietis			√						√
Pomacea bridgesii						√	√		√
Anentome helena		√	√						√
Aplysia californica			√						√
Elysia crispata		√							√
Physella acuta			√						√
Lymnaea stagnalis	√		√	√					
Biomphalaria glabrata									√
Arion vulgaris					√	√	√		√
Deroceras reticulatum			√						√
Limax maximus									√
Helix pomatia						√		√	√
Cepaea hortensis			√						√
Cornu aspersum	√		√	√	√				√
Alinda biplicata		√	√						√

Individuals of *Lymnaea stagnalis* were collected from an unpolluted freshwater pond in the Ternopil region, Ukraine (49°49/N, 25°23/E) and were kindly provided to us by Dr Oksana B. Stoliar and Dr Halina I. Falfushynska (Ternopil National Pedagogical University, Ukraine). For metal exposure, snails were kept in 80 l tanks of aerated tap water during 14 days, and exposed to a Cd concentration of 15 μg l^{-1} (in tap water). Water and Cd solutions were renewed every two days, lettuce feed was provided before water exchange. Control snails without Cd addition were kept in the same manner as exposed individuals. At the end of the exposure, three individuals per group were dissected for mRNA extraction from midgut gland.

Dissection, RNA/DNA isolation and cDNA synthesis

Snails were sacrificed and midgut gland tissue of individual snails (n = 3 5) (*Patella vulgata*, *Lottia gigantea*, *Anentome helena*, *Marisa cornuarietis*, *Aplysia californica*, *Cornu aspersum*, *Deroceras reticulatum*, *Lymnaea stagnalis*) or due to the small size of some species mixed tissue parts (*Alinda biplicata*, *Physella acuta*) were dissected and stored in RNAlater® (Fisher Scientific, Vienna, Austria) at −80 °C. For quantitative Real Time PCR (qPCR) after metal exposure, small aliquots of midgut gland tissue (approx. 1 mg fresh weight) of control and Cd-exposed *Cornu aspersum* and *Lymnaea stagnalis* (n = 3 5) were transferred to RNAlater® (Fisher Scientific, Vienna, Austria) whereas the remaining part of the tissue was collected for metal measurement.

RNA tissue samples were homogenized with a Precellys® homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France). Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) including the on-column DNase I digestion according to the manufacturer s instructions (Qiagen). RNA integrity was checked by agarose gel electrophoresis and concentrations were estimated with Nanodrop (Thermo Fisher Scientific, Waltham, CA, USA). For qPCR, RNA samples were measured in triplicates with the Quant-iT™ Ribogreen® RNA Assay Kit (Life Technologies Corporation, Carlsbad, USA) applying the Victor™ X4 Multilable Reades (PerkinElmer, Waltham, USA). 450 ng of total RNA was transcribed to cDNA in a 50 μl approach with the RevertAid Reverse Transcriptase (Thermo Fisher Scientific). For amplification of the multidomain MT sequences (*Alinda biplicata*, *Marisa cornuarietis*) AccuScript Hi-Fi Reverse transcriptase (Agilent, Santa Clara, CA, USA) was used in a 20 μl approach for cDNA synthesis.

For phylogenetic reconstruction based on neutral markers, DNA of the same specimens mentioned above was extracted using GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich). A ~590 bp stretch of the mitochondrial gene Cytochrome *C* oxidase I (COI) was amplified using the standard primers LCO1490 and HCO2198 suggested by ref. 25 and the degenerated primers LoboF1 and LoboR1[26] for some species.

PCR products of *Neritina pulligera* and *Littorina littorea* showed multiple bands and therefore were cloned before sequencing using the pTZ57R/T InsTAclone Kit (Thermo Fisher, Waltham, USA).

This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

To obtain ~1000 bp of 18SrDNA the following primers of ref. 25 were used in various combinations: 18A1, 470F, 1155F 700R, 1500R, 1800. PCR products were purified and Sanger sequenced by the facilities of Eurofins (MWG Operon, Germany). For *Helix pomatia* and *Patella vulgate*, only ~500 bp of 18SrDNA sequence were available; thus, full sequences were obtained from GenBank (FJ977750, FJ977632, AF239734, AY145373, MF544434, AY427527). Conditions for thermal cycling, polymerase and PCR are shown in Table S1 (ESI‡), newly generated sequences have been deposited at GenBank (MK919674-MK919701).

RNA seq and transcriptome assembly

Isolated RNA from an individual midgut gland (*Patella vulgata*, *Neritina pulligera*, *Littorina littorea*, *Pomatias elegans*, *Pomacea bridgesii*, *Marisa cornuarietis*, *Anentome helena*, *Elysia crispata*, and *Limax maximus*) or of pooled soft-tissue (*Alinda biplicata*) was sent to the Duke Center for Genomic and Computational Biology (GBC, Duke University, Durham, NC, USA) and sequenced with Hi-Seq 4000 Illumina sequencing. A separate library was sequenced for each species. Raw data were assembled using Trinity[27] version: v2.1.1 with default settings. Assemblies were provided for analysis on a local BLAST server `SequenceServer ,[28] where cDNA sequences encoding for diverse snail MTs were blasted against the transcriptomic data sets to identify MT sequences. Raw sequence reads data were deposited as Bioproject data base under the accession number PRJNA604693.

Collection and processing of transcriptomic data

For the species *Nacella polaris* and *Cepaea nemoralis*, raw reads from the SRA database (NCBI) were imported to Geneious R10 (Biomatters Ltd, Auckland, New Zealand) to assemble transcriptomes. New MT sequences were identified by blasting already known MT sequences from close relatives against the new transcriptomes. For most other species, MT peptide sequences of the diverse gastropod families were identified using the blastn tool at the NCBI platform (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the database transcriptome shotgun assembly for gastropod species (taxid: 6448).

MT sequence confirmation *via* long distance (LD) PCR and quantitative real-time PCR

Gene-specific primers (Table S2A, ESI‡) were designed from identified MT sequences derived from transcriptomic data (see above). For PCR, a 50 μl approach was set up using the Advantage 2 PCR System (Clontech, Takara Bio Europe, Saint-Germain-en-Laye, France) (Table S2B, ESI‡). PCR products were separated on a 1.5% agarose gel (Biozym, Hessisch Oldendorf, Germany) and gene specific bands were excised. DNA was purified applying the QIAquick™ Gel Extraction Kit (Qiagen, Hilden, Germany), and cleaned samples were sent to Microysnth AG (Balgach, Switzerland) for Sanger-sequencing. When necessary, subsequent cloning was performed with the TOPO® TA Cloning® Kit for sequencing (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Insert containing plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and sent to Microysnth AG (Balgach, Switzerland) for Sanger sequencing. Primer design and sequence analysis were performed applying CLC Main workbench 6.9 (Quiagen, Aarhus, Denmark).

For CdMT quantification of *Cornu aspersum*, cDNA of the controls and Cd-exposed individuals were measured in triplicates using a 7500 Real Time PCR Analyzer with Power SYBR® Green detection (Applied Biosystems™ by ThermoFisher Scientific, USA). Details on primer design and concentrations as well as establishment of the calibration curve are described in ref. 26. Total RNA was used as a reference for transcriptional quantification (see ref. 29).

Phylogenetic analysis

Alignments of MT amino acid sequences were done with MUSCLE v3.8.31,[30] and manual corrections were applied if deemed appropriate. Alignment length was variable and species specific, with protein lengths between 50 and 180 amino acids. All alignments applied for the present tree calculations are reported as FASTA alignments (see Alignments S1 S5, ESI‡). Phylogenetic tree reconstructions were performed with RAxML v8.2.8 (maximum likelihood ML[31]) and MrBayes v3.2.6 (Bayesian inference BI[32]). For ML with the model PROTGAMMAIWAG, 1000 10 000 inferences were calculated, and 1000 bootstrap replicates. For BI, 10 million generations were calculated with rates = invgamma and aamodelpr = mixed, average standard deviation of split frequencies.

In addition, phylogenetic trees were also computed with a maximum likelihood (ML) approach with 500 bootstrap replicates, using the freely accessible programme platform SeaView (version 4.7) of PRABI-Doua, using default settings. Overall topologies between BI and ML trees were very similar, and the trees with the lowest number of polytomies are shown.

Mitochondrial COI sequences were manually aligned and checked for correct amino acid translation; the ribosomal 18SrDNA sequences were aligned using the SINA Alignment tool v. 1.2.11, based on the SILVA database[33] (Alignment S1, ESI‡). In all phylogenetic reconstructions gaps were treated as missing data. Four partitions were defined in the concatenated data, one for each codon position of COI and one for 18SrDNA. ML analysis was performed using IQ-tree[34] allowing for model estimation in each partition; node supports were calculated using 1000 non-parametric UltraFast Bootstraps. For BI the best-fitting substitution models were obtained with Modeltest 3.7.:[35] GTR + I + G achieved the best AIC and BIC values in all four partitions. BI was performed with MrBayes v3.2.6 allowing for unlinked parameter estimation in each partition. Five million generations were performed and 25% burnin was chosen to discard data prior to convergence of runs (standard deviation of split-frequencies below 0.01). The ML tree and the BI tree (data not shown) revealed the same topology.

Cloning and recombinant expression of *MT* genes from *Pomacea bridgesii* and *Lottia gigantea*

Full-length synthetic cDNAs for *PbMT1* and *LgMT1* genes were synthesized by Integrated DNA Technologies Company (Coralville, IA, USA) and by Synbiotech (Monmouth Junction NJ, USA), respectively. Both cDNAs were cloned into the *E. coli* pGEX-4T-1 expression vector (GE Healthcare) as described elsewhere[13] with minor modifications. Cloned *PbMT1* and *LgMT1* cDNAs were sequenced

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

with the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) at the Scientific and Technological Centers of the University of Barcelona (ABIPRISM 310, Applied Biosystems).

For heterologous expression of GST-MT fusion proteins, 500 mL of LB medium with 100 μg mL^{-1} ampicillin were inoculated with protease-deficient *E. coli* BL21 cells previously transformed with the *PbMT1* pGEX-4T-1 or *LgMT1* pGEX-4T-1 recombinant plasmids. After overnight growth at 37 °C/250 rpm, the cultures were used to inoculate 5 L of fresh LB-100 μg mL^{-1} ampicillin medium. Gene expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 hours (h). After the first 30 minutes (min) of induction, cultures were supplemented with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM) or $CuSO_4$ (500 μM) in order to generate metal MT complexes. Cells were harvested by centrifugation for 5 min at 9100*g* (7700 rpm), and bacterial pellets were suspended in 125 mL of ice-cold PBS (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4 and 0.5% v/v β-mercaptoethanol). Resuspended cells were sonicated (Sonifier Ultrasonic Cell Disruptor) 8 min at voltage 6 with pulses of 0.6 seconds, and then centrifuged for 40 min at 17 200*g* (12 000 rpm) and 4 °C.

Purification of recombinant metal–MT complexes

Protein extracts containing GST PbMT1 or GST LgMT1 fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. GST MT fusion proteins bound to the sepharose beads were washed with 30 mL of cold 1× PBS bubbled with argon to prevent oxidation. After three washes, GST MT fusion proteins were digested with thrombin (GE Healthcare, 25 U L^{-1} of culture) overnight at 17 °C, thus enabling separation of the metal MT complexes from the GST that remained bound to the sepharose matrix. The eluted metal MT complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris HCl, pH 7.0, and run at 0.8 mL min^{-1}. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at −80 °C until use.

Analysis of recombinantly expressed metal–MT complexes

For determination of the molecular mass of the metal complex species in solution, the metal MT complexes produced by recombinant expression were analyzed by electrospray ionization mass spectrometry (ESI-MS). For that purpose, a Micro Tof-Q Instrument (Bruker Daltonics Gmbh, Bremen, Germany) with a time-of-flight analyzer (ESI-TOF MS) was utilized, calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies, Santa Clara, CA, USA), and interfaced with a Series 1100 HPLC pump (Agilent Technologies) equipped with an autosampler, both controlled by the Compass Software. The experimental conditions for analysis of Zn and Cd proteins were as follows. 10 20 μL of the sample were injected at 40 μL min^{-1} using the capillary-counter electrode voltage at 5.0 kV and the desolvation temperature in the 90 110 °C range. For Cu containing samples the conditions used were milder, applying the capillary-counter electrode voltage at 4.0 kV and the desolvation temperature at 80 °C. Spectra were collected throughout an *m*/*z* range from 800 to 2500. The liquid carrier was a 90 : 10 mixture of 15 mM ammonium acetate and acetonitrile, pH 7.0. All samples were injected in duplicates to ensure reproducibility.

NMR and metal titration

Fully cadmium-loaded forms of *Littorina littorea* and *Helix pomatia* MTs were produced by recombinant expression and uniformly ^{15}N-labelled in *E. coli* cells as described previously.[20] To demetallate the proteins their solutions were acidified in three buffer exchange steps, adding demetallation solutions (pH 2.0, 10 or 20 mM MES or TRIS, 10 mM TCEP) using Amicon Ultra 3k Centrifugal Filter Devices (EMD Millipore). All solutions were carefully purged with argon prior to use. Titrations were performed in 20 or 50 mM MES (pH 6.0), MES (pH 7.0) and Tris (pH 7.0) buffers with 10 mM TCEP yielding very similar results. Metallation was followed by recording [^{15}N,^{1}H]-HSQC spectra or best-type [^{15}N,^{1}H]-HSQC spectra.[36] Spectra were analyzed and peaks integrated applying the program CcpNmr v.2.4.2.[37]

Measurements of ^{15}N transverse relaxation rates (*R*2) were performed using a HSQC-type version of the Carr Purcell Meiboom Gill (CPMG) experiment.[38] 32 scans were performed per increment and T2 delays of 0, 17, 34, 51, 68, 102, 119, 204 and 305 ms were used. The relaxation delay in all these experiments was set to 2 s. Spectra were recorded using Zn_6- or Cd_6-HpMT and Zn_9- or Cd_9-LlMT samples. Zn-Loaded MTs were generated by adding Zn to demetallated MTs. Peaks were integrated batchwise using the program SPSCAN and R2 rates extracted from least square fits to the standard exponential decay function with gnuplot.

All spectra were recorded at 298 K on a Bruker NEON 600 MHz or 700 MHz NMR spectrometer using a PRODIGY triple-resonance probe for samples at a concentration range of 0.1 0.5 mM.

Tissue sample digestion and metal analyses

For Cd analysis, midgut gland tissue samples and lettuce leaves (*Lactuca sativa*) were oven-dried at 60 °C. After dry weight (d.w.) determination, samples were wet-digested at 70 °C with a mixture of HNO_3 (Suprapur, Merck, Darmstadt, Germany) and deionized water (1 : 1) in 12 ml screw-capped polyethylene tubes (Greiner, Kremsmunster, Austria). For complete oxidation, a few drops of H_2O_2 were added to the hot digested samples. They were filled up with deionized water to a final volume of 11.5 ml. Cd concentrations were measured by flame (Model 2380, PerkinElmer, Boston, MA) or graphite furnace atomic absorption spectrophotometry with polarized Zeeman background correction (Model Z-8200, Hitachi, Japan) and $Pd(NO_3)_2$ as a matrix modifier, depending on concentration levels in the samples. Calibration was performed with diluted titrisol standard solutions (Merck) prepared with de-ionized water and 5% HNO_3 (suprapur, Merck). Lobster hepatopancreas powder (TORT-2, National Research Council, Canada) was used as a standard reference material and processed in the same way as the samples (n = 5).

This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

Preparation and chromatography of *in vivo* MTs

Purification and preparation of *in vivo* MTs for determination of molar metal ratios were performed on centrifuged supernatants of midgut gland homogenates obtained from Cd-exposed snails (*Helix pomatia*, *Cornu aspersum*) und slugs (*Arion vulgaris*), by applying successive fractionation steps on gel permeation chromatography, anion exchange chromatography, ultrafiltration and Reverse phase HPLC.[23] For each species, HPLC fractions of the eluted MT peak were pooled and diluted 1:10 with deionized water under addition of 1% HNO_3. Metal concentrations (Cd, Cu, Zn) were analysed in triplicate in 1 ml aliquots by graphite furnace atomic absorption spectrophotometry with polarized Zeeman background correction (Model Z-8200, Hitachi, Japan) as described above.

Statistics

Data from q-RT-PCR and metal analyses were evaluated using SigmaPlot 12.5 (SYSTAT software, San Jose, CA, USA). Values were tested for normal distribution with the Shapiro Wilk normality test and the equal variance test. Outliers of normally distributed data were assessed with the Grubbs test (https://www.graphpad.com/quickcalcs/Grubbs1.cfm). For not normally distributed data, non-parametric methods (Mann Whitney *U* test) were applied. Significance levels were set at $p \leq 0.05$.

Results and discussion

In this work we propose that Cd acts as a driver in the evolution of gastropod metallothioneins. In what follows we first describe the variety of gastropod MTs (Section 1), structural features of MTs (Section 2), and our phylogenetic analysis of how gastropod MTs changed during evolution to gain (Section 3) or loose (Section 4) Cd-binding selectivity. Moreover, we describe how Cd-selectivity was achieved during evolution (Section 5) and conclude with how changes in environmental Cd levels influenced Cd selectivity (Section 6).

1. Gastropod diversity

Modern Gastropoda represent five distinct clades with about 80 000 species: Patellogastropoda, Neritimorpha, Vetigastropoda, Caenogastropoda and Heterobranchia. Their phylogenetic relationships[25,39 44] served as a reference for this study. Interestingly, several gastropod lineages have independently abandoned marine realms and successfully adapted to semi-terrestrial, terrestrial and freshwater environments.[45] This manifold colonization of non-marine habitats has promoted the huge diversity of gastropod traits, including the structural and functional diversity of their *MT* genes and proteins. Species and their MT sequences used for phylogenetic tree constructions of the present study are reported in Table S3 (ESI‡).

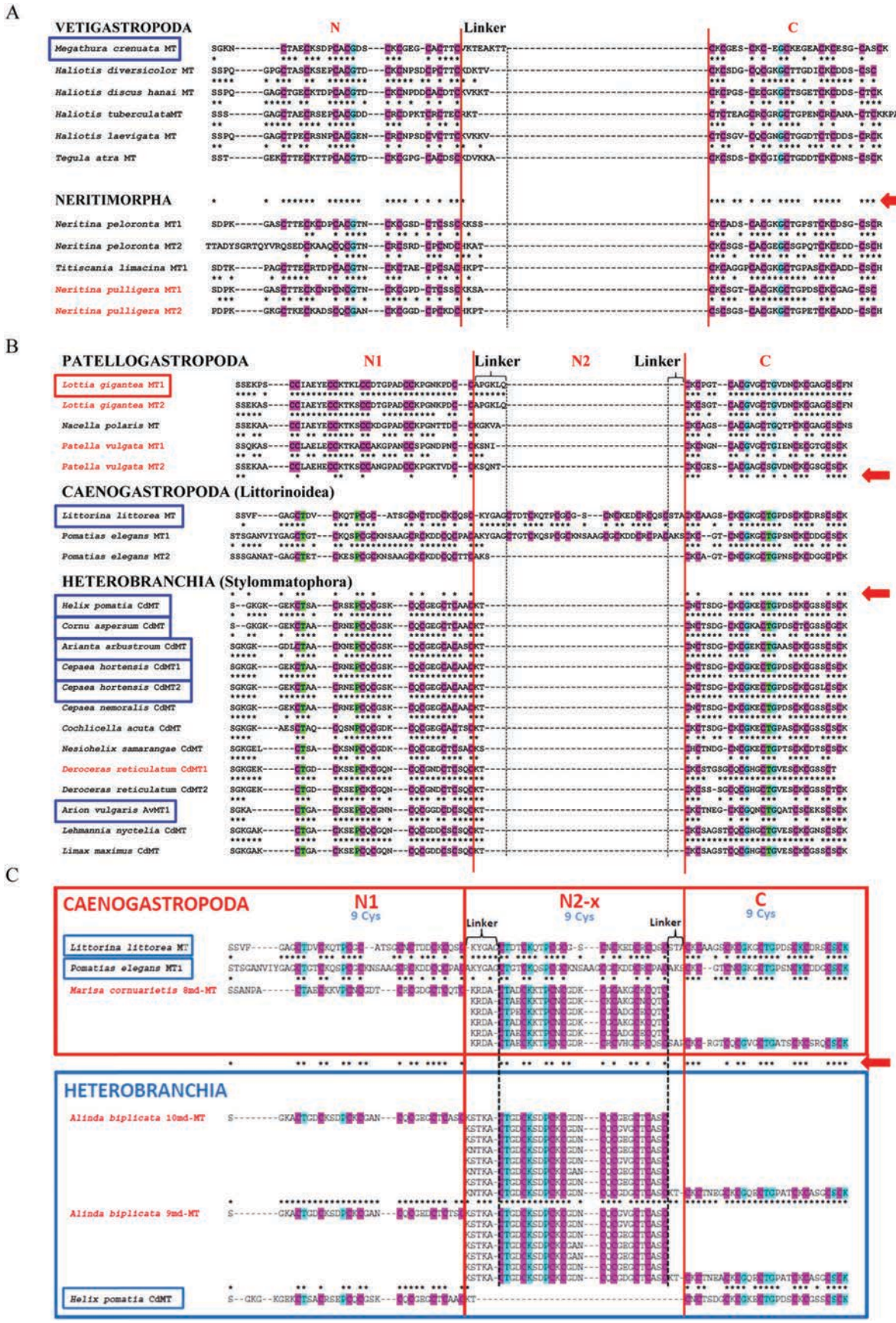
2. Gastropod MTs: structures, domain organization and metal binding features

Examples of primary MT structures across all major gastropod clades are shown in Fig. 1A E. Amino acid sequences of most gastropod MTs reflect a bipartite organization of one N-terminal metal-binding domain linked to a distinctly different C-terminal metal-binding domain (Fig. 1A, B and E). This kind of structural organization has been confirmed by NMR studies and molecular modeling.[16,46] It is therefore assumed that the primordial gastropod MT was a bidominial MT. Both N-terminal and C-terminal domains, contain nine Cys residues each which bind in a stoichiometric ratio, three divalent (mainly Cd^{2+}, Zn^{2+}), or six monovalent (mainly Cu^{+}) metal ions, such that a prototypical two-domain snail MT can accommodate six divalent or 12 monovalent metal ions, respectively.[14] An exception from this rule is observed in MTs of Patellogastropoda such as *Lottia gigantea* and *Patella vulgata*, which possess a deviating N-terminal domain that contains 10 instead of nine Cys residues, most of them arranged in form of double (Cys Cys) motifs (Fig. 1B). In addition to this, the N-terminal MT domain in some snail species has been duplicated once or several times independently, as seen in *Littorina littorea* and *Pomatias elegans*[16,20] (Fig. 1B). In the land snail *Alinda biplicata* and in some other species, tandem duplications generated multi-domain MTs (md-MTs) consisting of modular strings of up to nine N-terminal domain repeats, always linked to a single C-terminal domain that has, to the best of our knowledge, never been duplicated[47] (Fig. 1C). Domain duplications were also reported from bivalve MTs,[48] which have probably emerged independently from those in gastropods. In gastropods, md-MTs can bind additional metal ions according to the number of added domains within the protein chain. For example, in In the three-domain MT of *Littorina littorea*, the metal binding ratio for Cd^{2+} has been extended to a number of nine Cd^{2+} ions as compared with six Cd^{2+} ions in normal bidominial snail MTs.[20,22] Apart from this, N- and/or C-terminal domains are modified in some species by deletions at specific positions or through premature chain truncations (Fig. 1D and E).

Across all gastropod MTs, primary structures of C-terminal domains appear to be higher conserved compared with N-terminal domains. A BLAST comparison of the N-terminal domain of *Littorina littorea* with those of all other gastropod CdMTs reveals low degrees of homology in a clade-specific gradation from Heterobranchia through Caenogastropoda to Patellogastropoda (Fig. 2). In contrast, similarity scores between C-terminal domains of the same MT sequences are much more significant. This clearly demonstrates the higher degree of conservation of C-terminal against N-terminal metal binding domains, which is also confirmed by a distance matrix derived from single domain alignments (Table S4, ESI‡). We used *Littorina littorea* as a reference because this species possesses a well characterized CdMT[20,22] and occupies a central position between ancient and modern Gastropoda.[41]

The higher evolutionary pressure for sequence conservation of the C-terminal domain in snail MTs is probably related to preferred Cd^{2+} loading into that part of the protein. This is demonstrated by NMR data of experiments, in which Cd^{2+} equivalents were added stepwise to the apo-MT of *Littorina littorea* (Fig. 3). Unlike the Cd-loaded MT (Fig. 3B), the apo-MT is unfolded and does not assume a specific three-dimensional shape (Fig. 3A).[11,49] Added Cd^{2+} is initially cooperatively incorporated into the C-terminal domain to build the C-terminal cluster (Fig. 3C), before the two N-terminal domains form

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.



This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

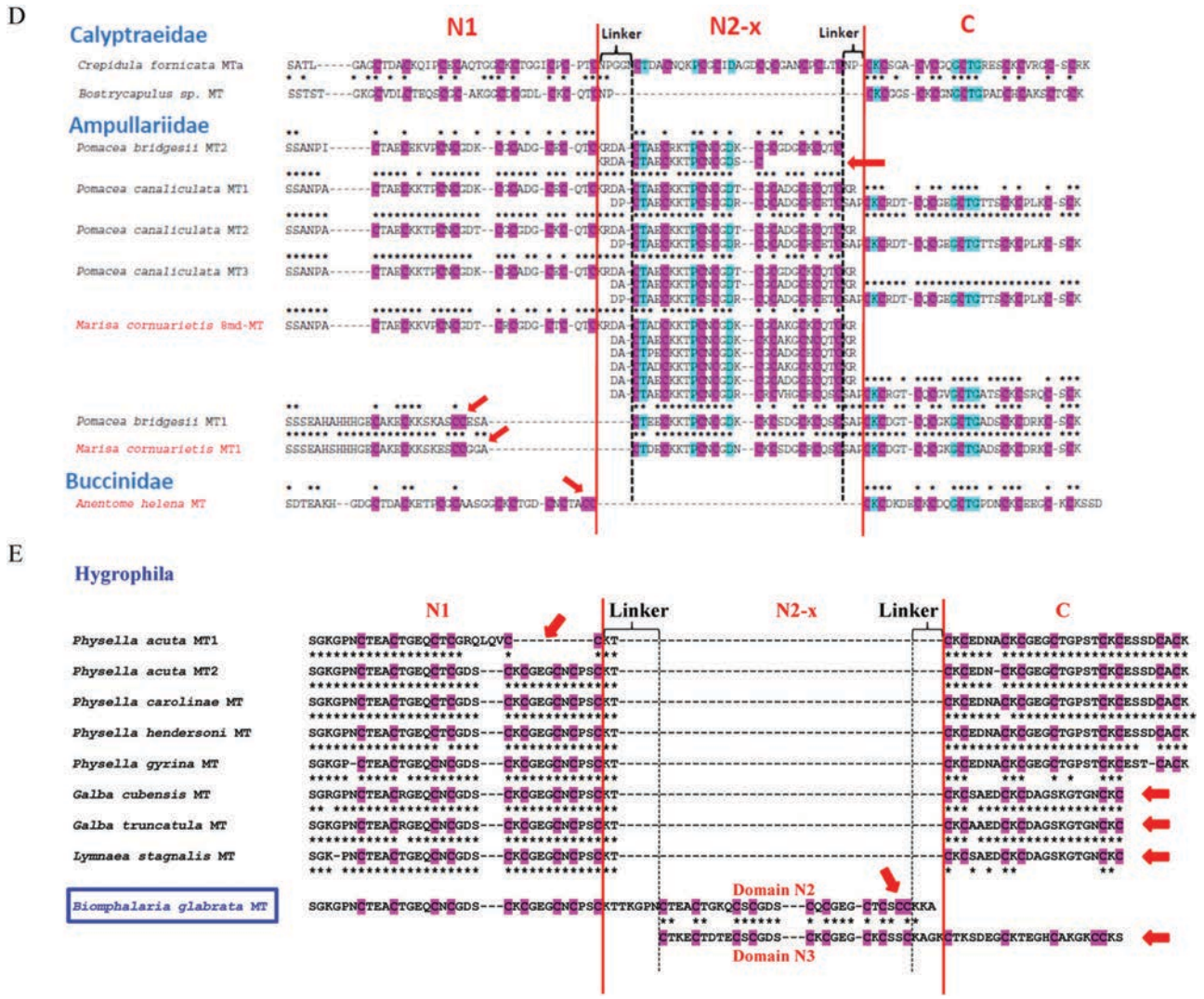


Fig. 1 (A–E) Sequence alignments of snail metallothioneins. Cys positions are underlaid in pink, conserved non-Cys positions through sequences of both clades are underlaid in light blue. Identical amino acid positions between pairwise aligned sequences are indicated by black stars. Domain boundaries of the N-terminal and the C-terminal domain (designated above the alignments with N and C) are indicated by bold red lines. The linker between the two domains is shown in black letters, its boundary is symbolized by a dotted line. The gaps between the two domains were inserted indicating the lack of a second N-terminal domain (present in other gastropod MTs). MTs of species shown in red letters were sequenced in this study for the first time while sequences in black letters were downloaded from publications or databases. Species for which metal selectivity features of respective MTs were documented experimentally by us through MS or NMR methods elsewhere are framed in blue. (A) Metal-binding domain organization and amino acid sequence alignment of unspecific MTs from the gastropod clade of Vetigastropoda and MTs with still unknown metal binding features from Neritimorpha. The bold red arrow on the right hand of the alignments points to sequence identities between MTs of Tegula atra (Vetigastropoda) and Neritina peloronta MT1 (Neritimorpha). (B) Metal-binding domain organization and amino acid sequence alignment of Cd-selective MTs from the gastropod clades of Patellogastropoda, Caenogastropoda and Heterobranchia using the same annotations as described in Fig. 1A. The bold red arrows on the right hand of the alignments points to sequence differences or similarities between MTs of the three clades. (C) Metal-binding domain organization and amino acid sequence alignment of multi-domain MTs from the gastropod clades of Caenogastropoda (framed in red) and Heterobranchia (framed in blue), compared with the bidominial CdMT of *Helix pomatia* (last sequence within Heterobranchia). Numbers of Cys positions per domain are given in blue letters above the alignments. The bold red arrow on the right hand of the alignments points to sequence similarities between MT sequences of Caenogastropoda and Heterobranchia. (D) Metal-binding domain organization and amino acid sequence alignment of two and multi-domain MTs from freshwater families (with Calyptraeidae, Ampullariidae and Buccinidae) of the clade of Caenogastropoda. The bold red arrows above or besides some sequences point to sequence irregularities such as truncations or Cys replacements. (E) Metal-binding domain organization and amino acid sequence alignment of unspecific two and multi-domain MTs from the freshwater snail order of Hygrophila of the clade of Heterobranchia. N and C-terminal domains are designated in red letters above the alignments by N1, N2-x and C, and as Domain N2 and Domain N3 for *Biomphalaria glabrata*. The bold red arrows above or besides some sequences point to sequence irregularities such as gaps, truncations or Cys replacements.

simultaneously (Fig. 3D). This proves a clear priority for Cd^{2+} uptake into the C-terminal domain.

So far, the tertiary structure of two snail MTs has been disclosed by solution NMR, namely for the bidominial CdMT

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

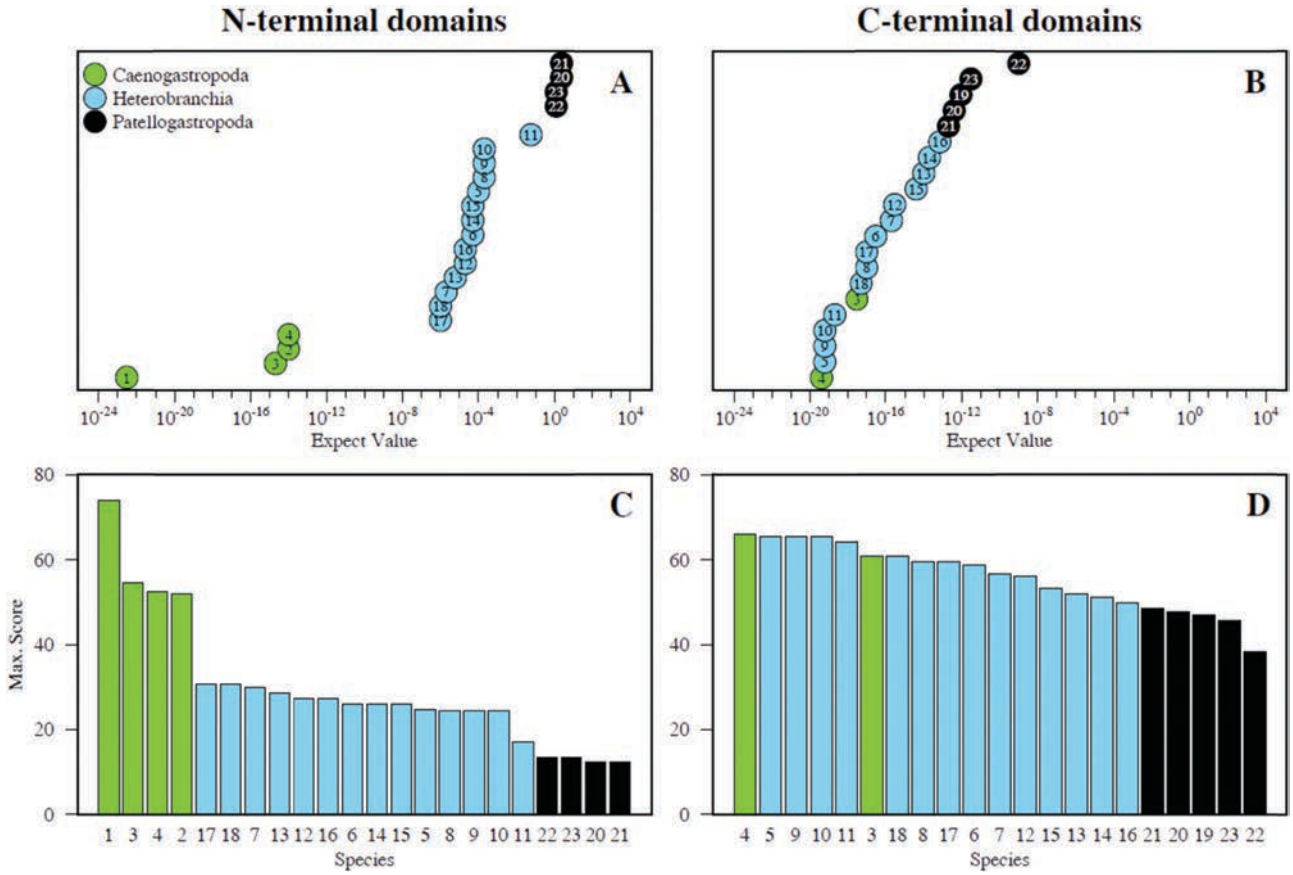


Fig. 2 Conservation between N- and C-terminal domains in CdMTs of Caenogastropoda (green symbols or bars), Heterobranchia (blue symbols or bars) and Patellogastropoda (black symbols or bars). Shown are expect values in ascending order (A and B) and homology scores in descending order (C and D), calculated with BLASTp. N-Terminal domains (A and C) and C-terminal domains (B and D) of Cd-selective MTs were compared to the N-terminal (N1-domain) and to the C-terminal domain of *Littorina littorea* MT. N-Terminal domains are less conserved through evolution (higher differences in e-values and scores) than the C-terminal domains. Species labels: (1) *Littorina littorea* MT (N2-domain); (2) *Pomatias elegans* MT1 (N1-domain); (3) *Pomatias elegans* MT1 (N2- & C-domain); (4) *Pomatias elegans* MT2; (5) *Helix pomatia* CdMT; (6) *Cornu aspersum* CdMT; (7) *Arianta arbustorum* CdMT; (8) *Cepaea hortensis* CdMT1; (9) *Cepaea hortensis* CdMT2; (10) *Cepaea nemoralis* CdMT; (11) *Cochlicella acuta* CdMT; (12) *Nesiohelix samarangae* CdMT; (13) *Alinda biplicata* CdMT; (14) *Deroceras reticulatum* CdMT1; (15) *Deroceras reticulatum* CdMT2; (16) *Arion vulgaris* AvMT1; (17) *Lehmannia nyctelia* CdMT; (18) *Limax maximus* CdMT; (19) *Lottia gigantea* MT1; (20) *Lottia gigantea* MT2; (21) *Nacella polaris* MT; (22) *Patella vulgata* MT1; (23) *Patella vulgata* MT2.

of the Roman snail, *Helix pomatia*,[46] and for the three-domain CdMT of the periwinkle, *Littorina littorea*.[20] The tertiary structure of the Roman snail CdMT in its dumbbell shape resembles the very similar structures of vertebrate MTs.[50,51] However, the metal-binding stoichiometry of the snail MT with six Cd^{2+} ions for the entire protein and three Cd^{2+} ions coordinated by nine Cys residues within each of the two domains, respectively, differs from the well-known metal binding stoichiometry of MTs from most other animal clades.[46] In mammalian MTs, four divalent metal ions are coordinated by 11 Cys residues in the C-terminal cluster (called alpha domain), whereas three divalent metal ions are bound by nine Cys residues in the N-terminal domain (called beta domain).[51] The MT of *Littorina littorea* is the first reported animal MT ever that exhibits a three-domain partition.[20]

Many snail MTs possess Cd- or Cu-selective binding preferences, and can be isolated as stable, homometallic metal complexes from native snail tissues.[14,52] Although the exact prerequisits for metal-selectivity are not yet fully understood, it appears that the frequency and position of certain non-coordinating amino acid residues in the primary sequence and their spatial arrangement in the tertiary structure are crucial determinants in conferring metal-selectivity to snail MTs.[15,19 21]

The homometallic composition of metal-selective snail MTs was demonstrated by electrospray ionization mass spectrometry (ESI-MS) in recombinantly expressed and purified MT proteins.[15,22] Thereby, metal-selective MTs can be detected as homometallic complexes with their cognate metal ions (mainly Cd^{2+} or Cu^{+}) but appear as a heterometallic mixture of complexes with variable stoichiometry when forced to associate with non-cognate metal species.[15,18] In contrast, metal-unspecific snail MTs form heterometallic mixtures of complexes with variable stoichiometry in presence of any metal ions.[17] According to this definition, it appears that Cd-selective MTs have not equally evolved in all gastropod clades. For example, vetigastropod species like *Megathura crenulata* possess an unspecific MT that produces a mixture of sulfide containing heterometallic

This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

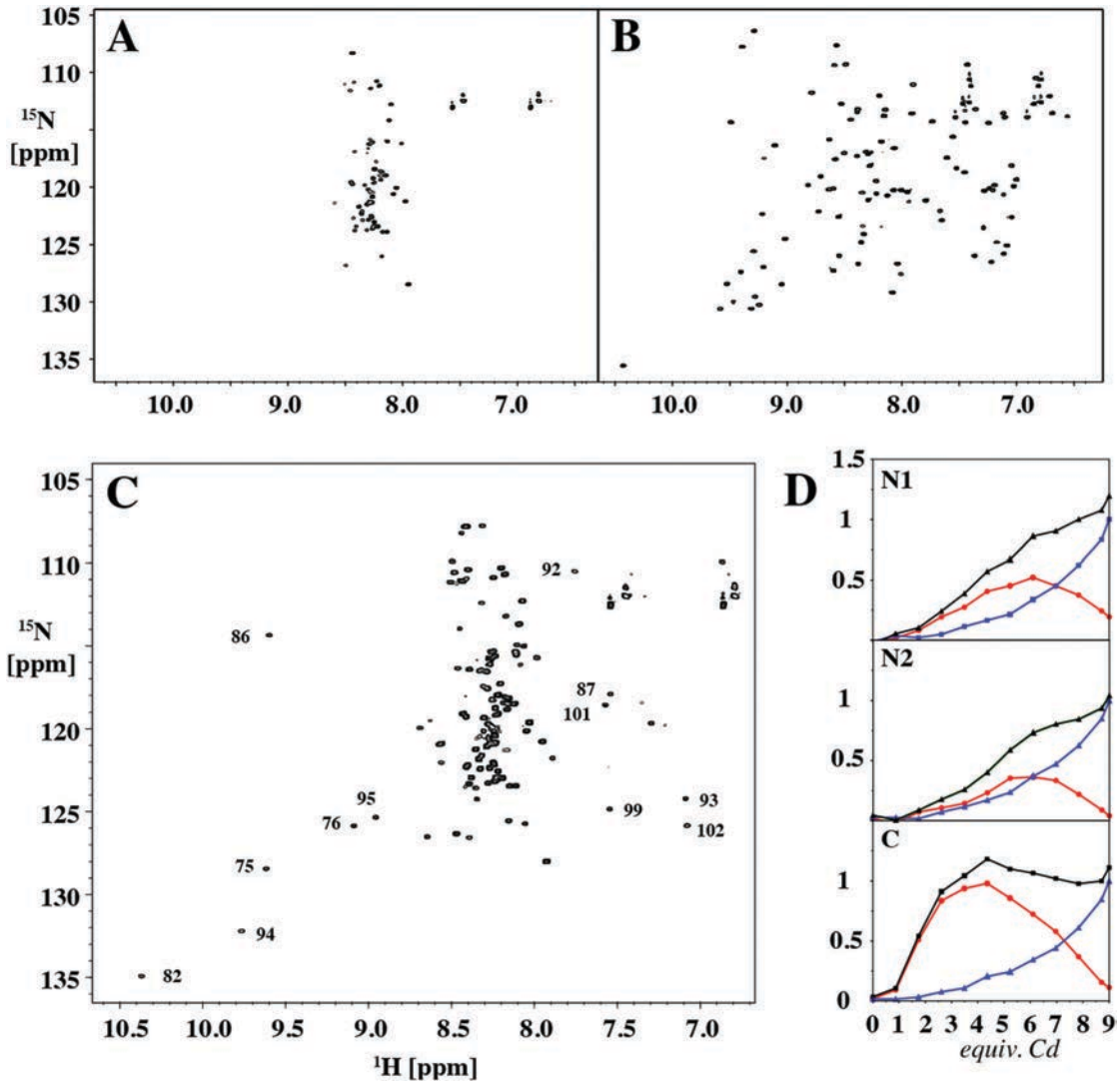


Fig. 3 Metallation of apo-MT from *Littorina littorea* (LIMT) with Cd^{2+}. [^{15}N,^{1}H]-HSQC spectrum of apo (A) and fully-metallated (B) (Cd_9)-LIMT. (C) [^{15}N,^{1}H]-HSQC spectrum after addition of 2 equiv. of Cd^{2+} to apo-LIMT. Peaks close to positions in the fully-metallated form are annotated, and exclusively stem from the metallated C-terminal domain. (D) Normalized average relative peak volumes of peaks from the first (top) and second (center) N-terminal as well as the C-terminal (bottom) domains. Mostly, two peaks are observed for each residue corresponding to apo (red) and metallated (blue) neighboring domains (in the case of peaks from the C-terminal domain that corresponds to species in which the N2 domain is already metallated). The black line corresponds to the sum intensity of both peaks.

complexes with reduced stability when recombinantly expressed in Cd-enriched media[17] (Fig. 4). In contrast, recombinant CdMTs of some Caenogastropoda (like *Littorina littorea*) and Heterobranchia (like *Helix pomatia*) form homometallic Cd^{2+} complexes (Fig. 5). MTs of Patellogastropoda are also Cd-selective. However, because of the divergent primary structure of their N-terminal domain with 10 Cys residues (Fig. 1B), CdMTs of Patellogastropoda bind seven instead of six Cd^{2+} ions per protein molecule, as demonstrated for *Lottia gigantea* (Fig. 5). CdMTs of *Helix pomatia* and *Arion vulgaris* (Heterobranchia) bind six Cd^{2+} ions per protein molecule or nine Cd^{2+} ions in three-domain CdMTs like that of *Littorina littorea* (Caenogastropoda) (Fig. 5). All gastropod CdMTs are incapable to form homometallic Cu^{+} complexes (Fig. 5). However, due to the chemical similarity between Zn and Cd, some recombinant gastropod CdMTs can form homometallic complexes with divalent Zn^{2+} ions. This Zn^{2+}-binding selectivity is low in CdMTs of Patellogastropoda, as demonstrated for *Lottia gigantea* (Fig. 5). In contrast, Zn preference is high for Caenogastropoda CdMTs (*Littorina littorea*) and Stylommatophora CdMTs (*Helix pomatia* and *Arion vulgaris*), which are able to form homometallic Zn^{2+} complexes in the presence of excessive Zn^{2+} concentrations, with the same stoichiometry as for Cd^{2+} (Fig. 5). Interestingly, some evidence indicates Zn specificity in MTs of some mussels (Bivalvia), the mollusk sister class of gastropods.[44]

3. Phylogeny suggests convergent evolution of CdMTs in early gastropod clades

The multitude of published and collected primary MT sequences from species across all clades of Gastropoda (see

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

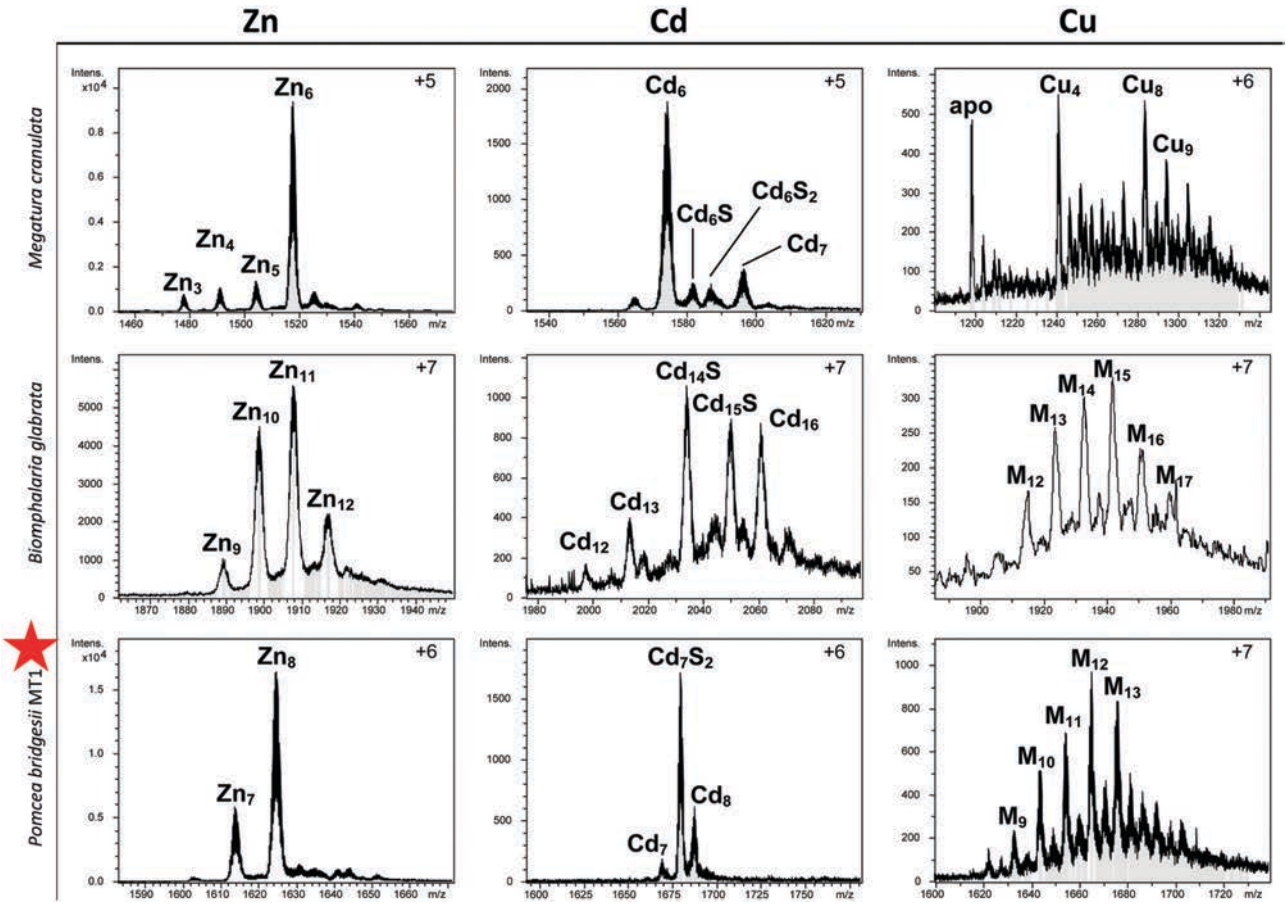


Fig. 4 ESI-MS spectra of metal-unselective gastropod MTs recombinantly produced in media containing Cd, Zn and Cu ions. Data are shown for MTs from *Megathura crenulata* (Vetigastropoda), *Biomphalaria glabrata* (freshwater Heterobranchia) and *Pomacea bridgesii* (freshwater Caenogastropoda). The corresponding charge state is indicated in the upper right corner. In Cu productions, M denotes mixtures of Zn + Cu. Spectra of *Pomea bridgesii* MT1 are shown here for the first time and are marked with a red star. Spectra of other MTs are re-drawn from data reported in ref. 17 and 21.

Table S3, ESI‡) and basic knowledge about their structure and metal-binding features (see above) fosters an attempt towards establishing a phylogeny of gastropod MTs and, in particular, Cd-selective snail MTs. The smallness of most MT proteins and the fact that the abundance of conserved cysteine residues and repeat motifs do not bear much phylogenetically evaluable information creates a challenge in such an analysis. In the present study, we have developed a domain and metal-specific approach to compensate somewhat for these handicaps.

Yet, confronting a phylogeny of neutral DNA markers with one based on Cd-selective MTs (Fig. 6) reflects the evolution of Cd selectivity in MTs of three gastropod clades: Patellogastropoda, Caenogastropoda and Heterobranchia. It appears that Cd-selective MTs are predominantly observed in species that have adapted to littoral and terrestrial environments (Fig. 6). A closer phylogenetic view in which MTs of Panpulmonata (a taxon of Heterobranchia comprising the lineages of Sacoglossa, Syphonariodea, Hygrophila and Stylommatophora)[53] are rooted with MTs of Caneogastropoda (Fig. 7) reveals the loss of Cd-selective MTs in freshwater snails of Caenogastropoda and Heterobranchia, and the initial emergence and subsequent loss of Cu-selective MT isoforms (CuMTs) in the lineage of Stylommatophora (terrestrial snails and slugs). In that context it is of interest that it was previously shown that snail CuMTs are involved in Cu regulation, possibly linked to hemocyanin synthesis.[54,55]

Chronograms show that Cd selectivity developed from ancestral MTs twice independently. CdMTs evolved first in Patellogastropoda, about 430 million years (My) ago (Fig. 8A). A second line of CdMTs emerged in the two sister clades Caenogastropoda and Heterobranchia, before 418 My ago (Fig. 8A). Apart from the phylogenetic evidence, another clear indication of this independent evolution is the emergence of a new kind of N-terminal metal-binding domain in CdMTs of Patellogastropoda (see above), which differs fundamentally from N-terminal domains in CdMTs of all other gastropod lineages (Fig. 1B). The two sister clades Caenogastropoda and Heterobranchia have shaped their CdMTs through parallel evolution (Fig. 8A and B). This is reflected by sequence similarities and homologous domain organization across primary structures of their CdMTs (Fig. 1B).

Also indicated in the chronogram are some of the main mass extinction events through the evolutionary history of the earth (Fig. 8B). Fluctuating emissions of Cd through continental and

This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

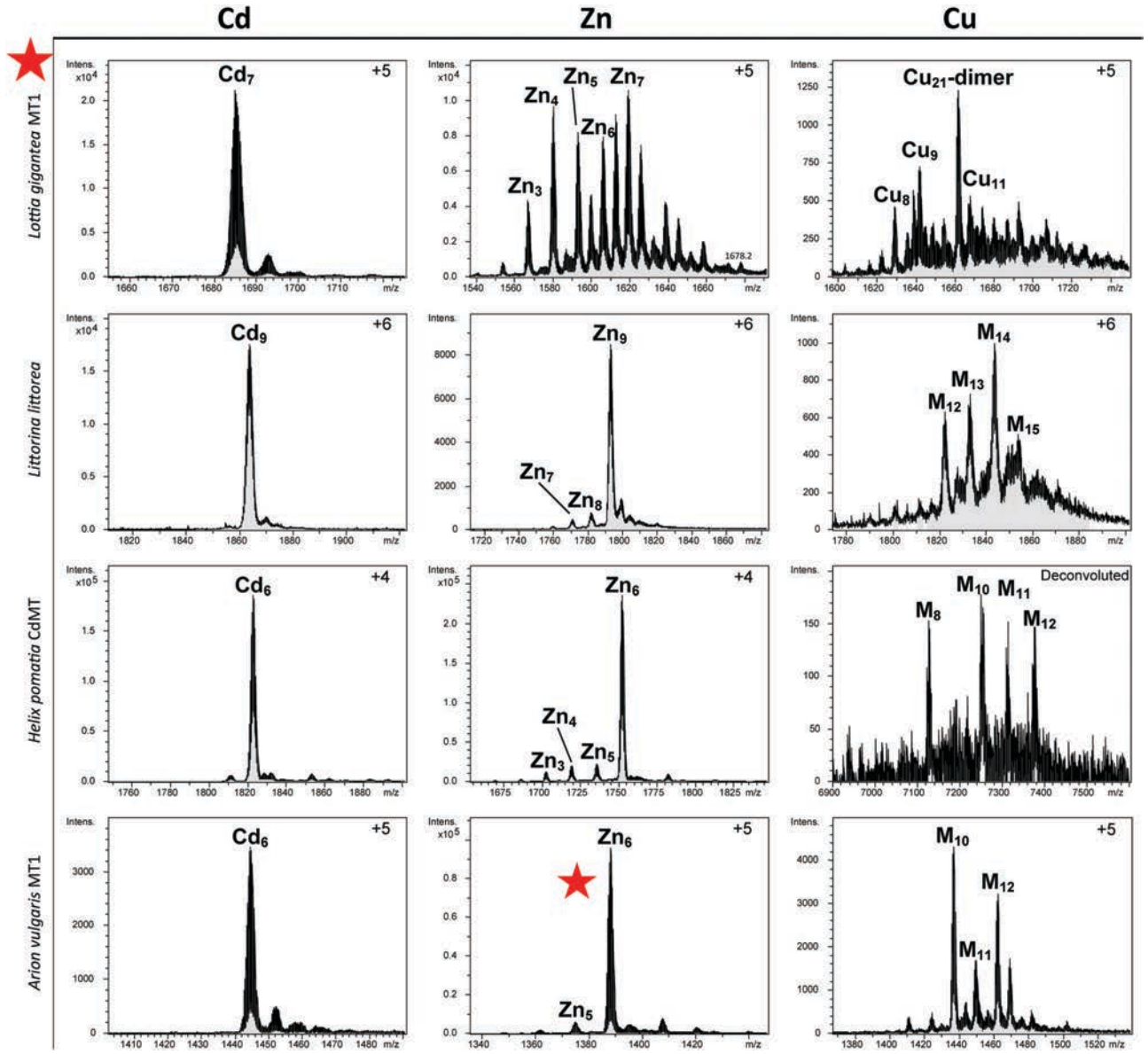


Fig. 5 ESI-MS spectra of Cd-selective gastropod MTs recombinantly produced in media containing Cd, Zn and Cu ions. Data are shown for MTs from *Lottia gigantea* (Patellogastropoda), *Littorina littorea* (Caenogastropoda), *Helix pomatia* (terrestrial snail, Heterobranchia) and *Arion vulgaris* (terrestrial slug, Heterobranchia). The corresponding charge state is indicated in the upper right corner. In Cu productions, M denotes mixtures of Zn + Cu. Spectra for which metal selectivity features are shown here for the first time are marked with a red star. Spectra of other MTs are re-drawn from data reported in ref. 15, 22 and 23.

super-volcanic emissions in combination with these catastrophic extinction events[2,56 59] may have triggered convergent evolution of Cd-selective MTs in gastropod clades since 430 My ago (Fig. 8B). Evidence for increased Cd emissions through geological eras is provided by elevated Cd concentrations in worldwide bedrock formations of different geological origin, from Paleozoic[60 63] through Mesozoic[63,64] and Cenozoic.[65]

Based on experimental data with recombinant proteins,[17] it appears that Cd selectivity is lacking in MTs of Vetigastropoda (Fig. 8A), which forms a sister clade to Patellogastropoda.[43] The metal-specific character in MTs of Neritimorpha, on the other hand, is still unknown (Fig. 8A). Since Neritimorpha from a sister clade to Caenogastropoda and Heterobranchia,[43] it could be speculated that they share Cd-specific features with their two sister clades. On the other hand, a high degree of identity in primary sequence and domain organization between MTs of Vetigastropoda and Neritimorpha (Fig. 1A) suggests the possibility that Neritimorpha MTs share some of their metal-binding features with those of Vetigastropoda. Future experiments through recombinant expression and ESI-MS analyses will probably resolve this question. The supposed zinc (Zn) specificity in MTs of some mussels (Bivalvia), the mollusk sister class of gastropods, is also indicated in Fig. 8A. However, this evidence is scarce, being derived from one single experimental study.[44]

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

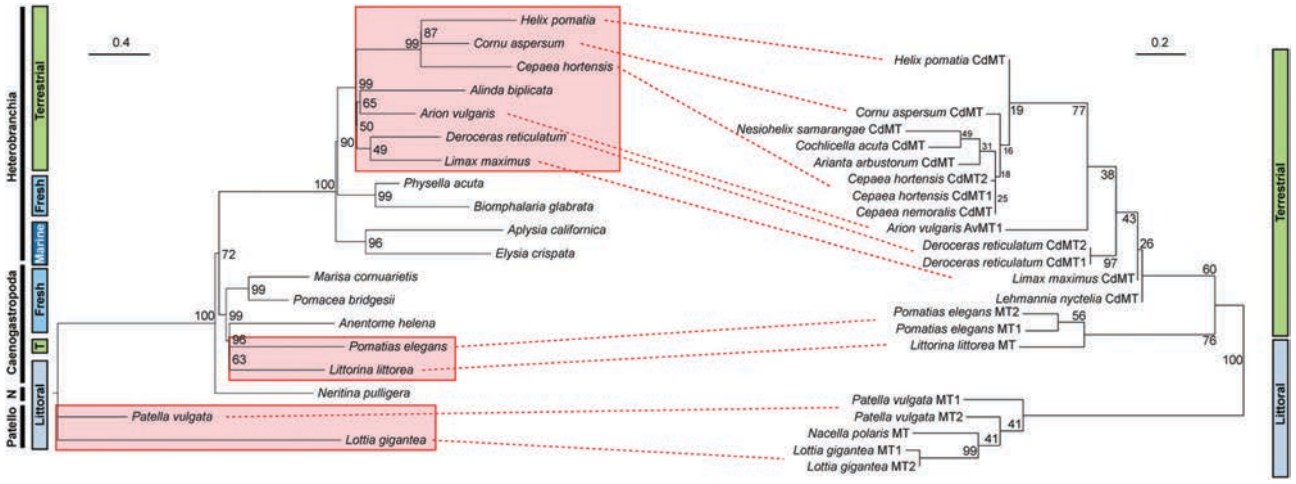


Fig. 6 Mirrored phylogenetic trees (maximum likelihood) of investigated species of the four major gastropod clades of Patellogastropoda, Neritimorpha, Caenogastropoda and Heterobranchia. Right: Phylogeny (maximum likelihood) showing the separated lineage clusters of only Cd-selective MTs. Bootstrap values (500 replicates) are given at nodes. Left: Neutral marker phylogeny based on concatenated CO1-18SrDNA data. Bootstrap values (1000 repetitions) are given at nodes. Mirrored species possessing Cd-selective MTs are shown within red-colored frames. Identical species between the two mirroring trees are connected by dotted red lines. On outside margins of the trees, habitats of the represented species are shown with colored bars. On the left outer margin of the neutral marker tree, major taxonomic clades are indicated by black bold lines. Abbreviations: T, Terrestrial; Fresh, Freshwater; Patello, Patellogastropoda; N, Neritimorpha.

4. Diversification and loss of cadmium selectivity during late gastropod radiation

Since the Cretaceous period, Cd selectivity of MTs was apparently lost in snail lineages that adapted to freshwater habitats. Accordingly, metal-binding features of MTs from *Pomatia bridgesii* (family of Ampullariidae, Caenogastropoda) and *Biomphalaria glabrata* (Hygrophila, Heterobranchia) resemble those of the unspecific *Megathura crenulata* MT[17] (Fig. 4). As indicated by their primary sequence and domain organization (Fig. 1D and E), a loss of Cd selectivity may also have occurred in caenogastropod species of freshwater Calyptreidae and Buccinidae (Fig. 6). The loss of Cd selectivity in these MTs is a derived character (Fig. 7), suggesting that metal selectivity was no longer required in MTs of freshwater snails. In some freshwater species of Caenogastropoda such as *Pomacea canaliculata*, MTs have developed N-terminal repeats, similar to some snail CdMTs (Fig. 1C).

In terrestrial snails of Stylommatophora (Heterobranchia), gene duplications of the primordial CdMT led to the emergence of three MT isoforms, each of them devoted to different, metal-specific tasks.[14,55,66] First, a gene duplication of *CdMT* gave rise to Cu-selective MTs, which form homometallic Cu^{+} complexes at a ratio of 12 Cu^{+} ions per protein molecule, but neither bind Cd^{2+} nor Zn^{2+} (Fig. S1 and S2, ESI‡). In a second event of gene duplication, *CuMT* genes lost their Cu selectivity in the so-called CdCuMT isoforms[18,19,55,66 68] (Fig. S1 and S2, ESI‡). Phylogenetic trees (Fig. S3, ESI‡) support the chronological succession of these evolutionary steps.

5. Evolutionary optimization of Cd selectivity and specificity

For the sake of clarity, we like to distinguish between Cd (or metal) selectivity and specificity of MTs. We define Cd selectivity as the binding preference of an MT for Cd^{2+} ions in presence of other metal ions, mainly Zn^{2+} and Cu^{+}. We define Cd (or metal) specificity as the involvement of the respective MT into a Cd- or metal-specific physiological function, which is often the consequence of its metal binding selectivity. For example, Cd-selective snail MTs are predominantly involved in Cd-specific functions like detoxification.[14,69,70]

Accordingly, we can observe that during gastropod evolution both, metal selectivity and physiological specificity of snail CdMTs have been optimized in favor of Cd^{2+}. The CdMT of *Littorina littorea*, for example, has been optimized for Cd^{2+} complexation to the disadvantage of Zn^{2+} binding. This can be concluded indirectly from the better fit of the protein backbone to the Cd *vs.* the Zn cluster. To this end we measured ^{15}N dynamics NMR data that probe for rigidity of the polypeptide backbone. Transverse relaxation (*R*2) rates of Zn^{2+}-loaded CdMTs are increased by 14 and 8 Hz in the N-terminal N1 and N2 domains of the CdMT of *Littorina littorea*, respectively, and by up to 5 Hz in the C-terminal domain of the *Helix pomatia* CdMT (Fig. 9) when compared to the Cd^{2+}-loaded forms. The increase in transverse relaxation rates reflects additional contributions from conformational exchange only for the Zn^{2+} species, indicating that the complexes with the cognate Cd^{2+} ion are conformationally more stable (Fig. 9). Similarly, NMR studies of the CdMT of *Helix pomatia* indicate a structural optimization for Cd^{2+} rather than Zn^{2+} binding.[46] Further evidence for an evolutionary optimization for Cd binding in these MTs comes from the fact that Cd^{2+} ions are incorporated into these proteins cooperatively (Fig. 3). The fact that peaks in the [^{15}N,^{1}H]-HSQC spectra occur at the same positions as in the fully-metallated domains indicates that no partially metallated domains form in situations of substoichiometric metal content. Strikingly, the Cd-selective character of gastropod CdMTs is also maintained in the presence of Cu^{+} ions at equimolar

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

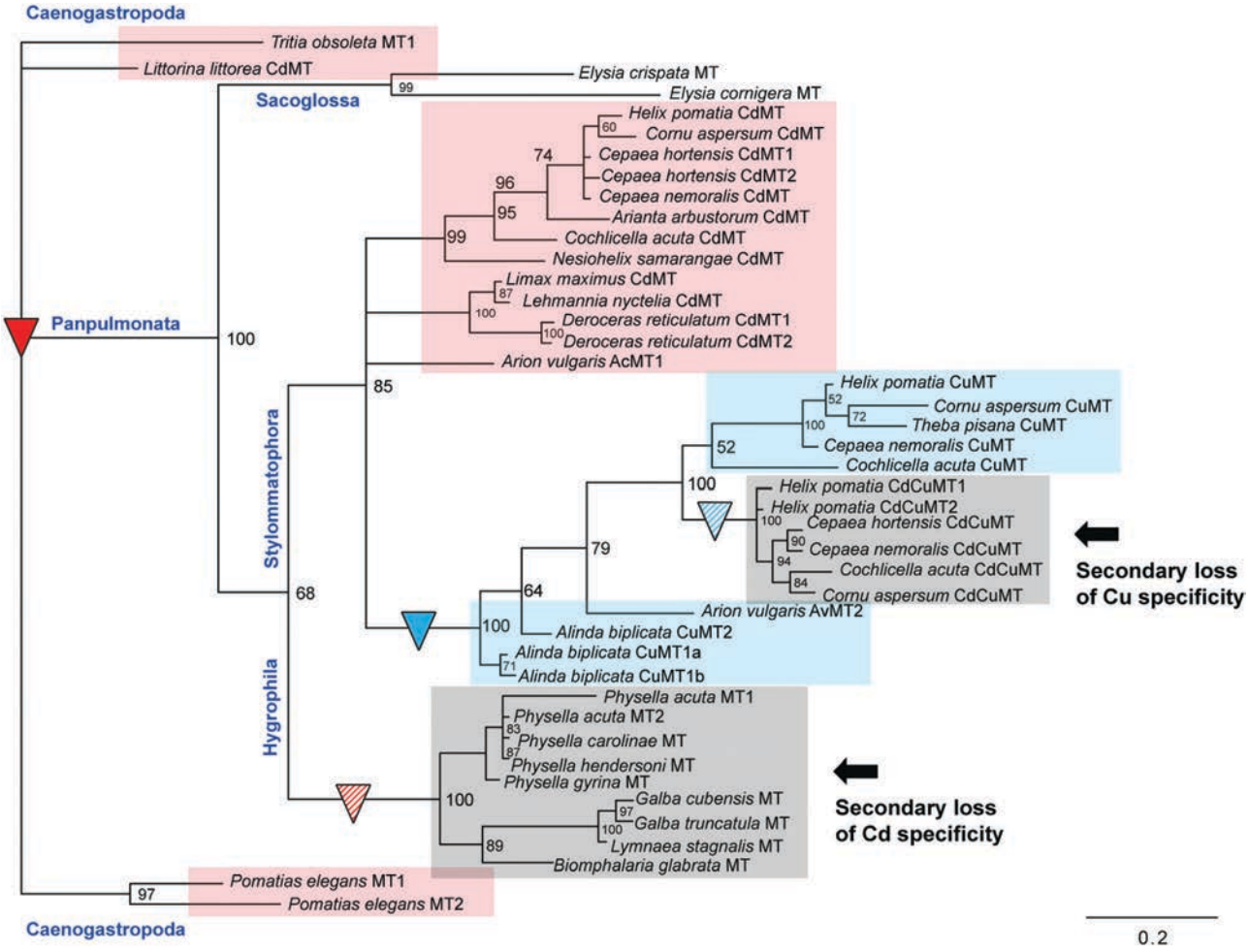


Fig. 7 Bayesian Inference tree with posterior probabilities of metal selectivity features in MTs of Panpulmonata (a taxon of Heterobranchia) *versus* Caenogastropoda. Shown are the gain of Cd selectivity (red triangle) in MTs at the root of Caenogastropoda and Heterobranchia, with species possessing Cd-selective MTs underlaid in pink, and the gain of Cu selectivity (blue triangle) in MTs of Stylommatophora, with species possessing Cu-selective MTs underlaid in blue. Also illustrated are the secondary loss of ancestral Cd selectivity in MTs of Hygrophila (red hatched triangle), and the secondary loss of Cu selectivity (blue hatched triangle) in CdCuMTs of Stylommatophora, with respective species clusters underlaid in blue. Bayesian inference calculations were made based on a manually edited MUSCLE alignment (see Alignment S4, ESI‡) using the free software MrBayes (see Material and methods).

concentrations with Cd^{2+}, as demonstrated recently for the recombinantly produced CdMT (AvMT1) of the terrestrial slug, *Arion vulgaris*.[23] This is remarkable since the evolution of thiolate proteins with an apparent preference for binding Cd^{2+} over Cu^{+} is a particular feature of snail CdMTs which is otherwise not observed in other animal MTs,[46,50,51] and seems to contradict the chemical rules of the Irving Williams series.[71] These rules predict that the stability constants of transition-metal ion complexes increase by a factor of 100 to 1000 from Cd- towards Cu-thiolates.[72,73] However, The Irving Williams rules may not apply to metal-selective snail MTs, considering that they do not contain simple binary metal-thiolate complexes. In the CdMTs of *Littorina littorea* and *Helix pomatia*, for example, divalent Cd^{2+} ions are tetrahedrally coordinated,[20,46] forming Cd-thiolate clusters that most likely differ in their structural configuration from the Cu-thiolate clusters of snail CuMTs.[14] Importantly, it was demonstrated that the replacement of a few amino acid positions in the near vicinity to the metal-coordinating Cys residues can have a strong impact on the metal binding preferences of snail MTs,[15,19,21] probably due to spatial and charge-dependent constraints upon formation of protein metal complexes. We suspect that such amino acid replacements must have gradually improved/modified the Cd-binding selectivity of snail MTs during evolution. Apart from this, the capacity for Cd-loading of many snail CdMTs has been increased through evolutionary multiplication of Cd-binding domains as demonstrated for the littoral periwinkle, *Littorina littorea*, and the land snails *Pomatias elegans* and *Alinda biplicata*.[16,20,47] At the functional level, evolutionary optimization for Cd binding in CdMTs has resulted in Cd-specific detoxification pathways within snail tissues. This is reflected by the fact that native purified gastropod CdMTs contain mainly Cd^{2+}, but only small amounts of Zn^{2+} or Cu^{+} (Fig. 10A). Concomitantly, Cd inactivation in these species is enhanced by metal-dependent upregulation of the

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

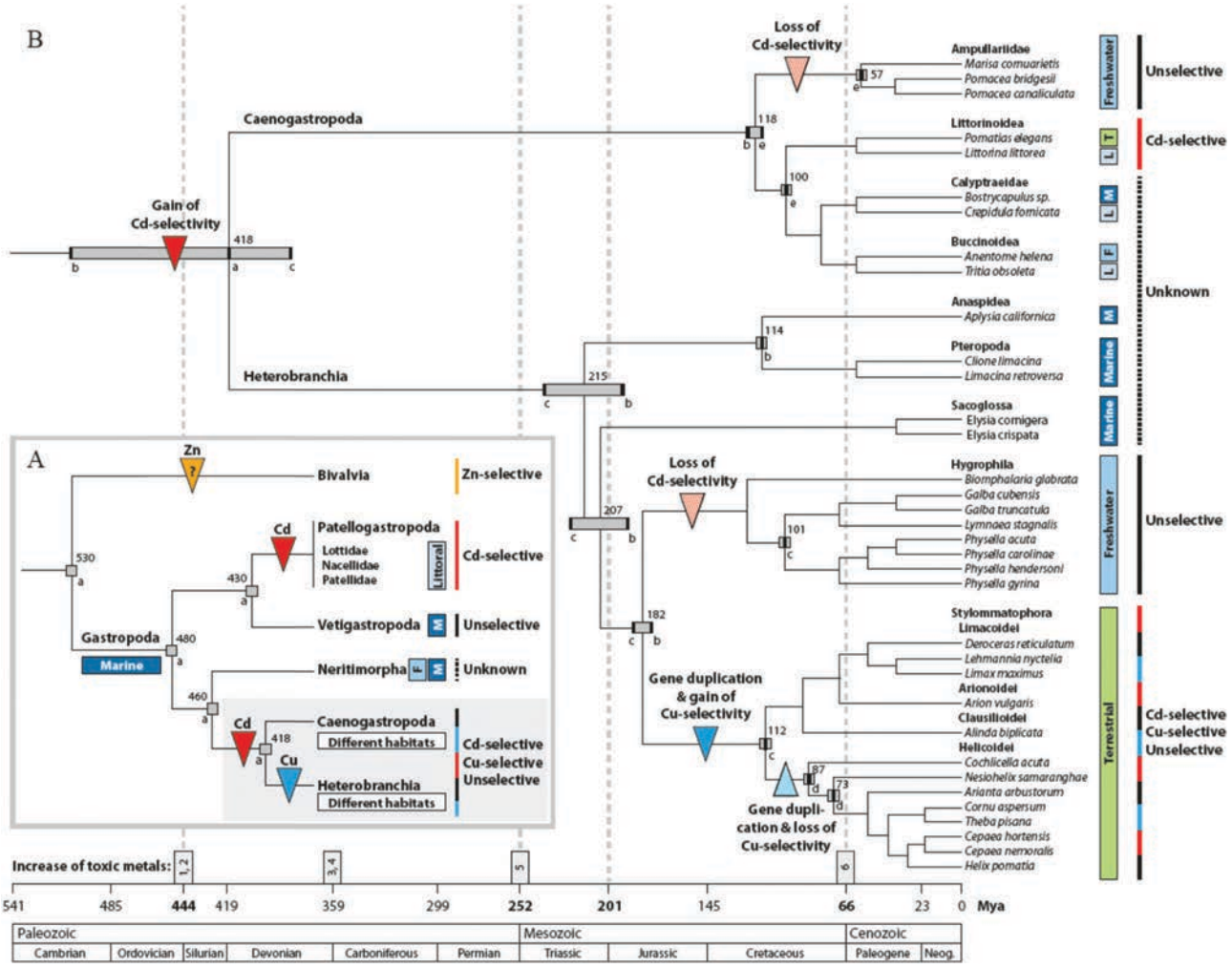


Fig. 8 Phylogenetic tree (A) and chronogram (B) of Cd and Cu selectivity gain and loss in metallothioneins of Gastropoda. (A) (inset), Phylogenetic tree of Gastropoda (reconstructed after[41,43]) with most probable relationships of gastropod clades (Patellogastropoda, Vetigastropoda, Neritimorpha, Caenogastropoda and Heterobranchia), rooted against the Gastropod sister class of Bivalvia (mussels).[92] Gain of Cd and Cu selectivity is indicated by red and blue triangles. The kind of metal selectivity in Neritimorpha MTs is still unknown. The possible gain of Zn selectivity in Bivalvia is indicated by an orange triangle with a query. Approximate divergence times (with references) of gastropod lineages are given in million years. Marine (M), littoral (L), freshwater (F) and terrestrial (T) habitats are specified in colored framed boxes. Metal selectivities are indicated by red (Cd-selective), blue (Cu-selective), orange (Zn-selective) and black (unselective) bars. (B) Chronogram showing gains and losses of metal selectivity in MTs of the two gastropod sister clades Caenogastropoda and Heterobranchia (enhanced from grey area in A), with their splits into major lineages, including investigated species. Cd-Selective MTs (red triangle) appeared prior to the divergence of Canogastropoda and Heterobranchia, and Cu selectivity (blue triangle) in MT isoforms of Stylommatophora. Cu selectivity was lost in novel MT isoforms of Stylommatophora (hatched blue triangle), and Cd selectivity was lost (hatched red triangles) in freshwater lineages of Ampullariidae (Caenogastropoda) and Hygrophila (Heterobranchia). Approximate divergence times of gastropod lineages are given in million years ago. Grey bars indicate published mean values for the divergence times (references a e). Vertical, grey dashed lines indicate four of the major mass extinction events. Elevated levels of toxic metals (including Cd) are indicated in grey boxes (references 1 6) above the time axis. Chronogram construction was based on: (a) ref. 41; (b) ref. 39; (c) ref. 25; (d) ref. 42; (e) ref. 44 (additional ref. 40, 43 and 93 95). Dating of increased volcanic Cd or metal emissions are based on information from the following studies: (1) ref. 60; (2) ref. 63; (3) ref. 62; (4) ref. 58; (5) ref. 57; (6) ref. 59 (additional ref. 2 and 56).

respective *CdMT* genes (Fig. 10B), and tissue or cell-specific expression of *CdMT* mRNA.[23,69,74]

6. Environmental Cd levels through the earth history: an important evolutionary driver

Cd is carcinogenic and highly toxic in animals, even at low concentrations.[75] The chemical similarity of this metal and its frequent co-occurrence with Zn in ore deposits of the earth crust make Cd a dangerous competitor for Zn-dependent cellular processes.[5] Cd can also compete with calcium (Ca),[76] and hence affect gastropods that depend on Ca pathways for bio-mineralization of their shells.[77]

Gain of Cd-selective MTs may have provided an advantage particularly for gastropod lineages that have adapted to littoral, semi-terrestrial, and terrestrial conditions. Recent natural Cd concentrations in seawater follow those of algal nutrients such

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

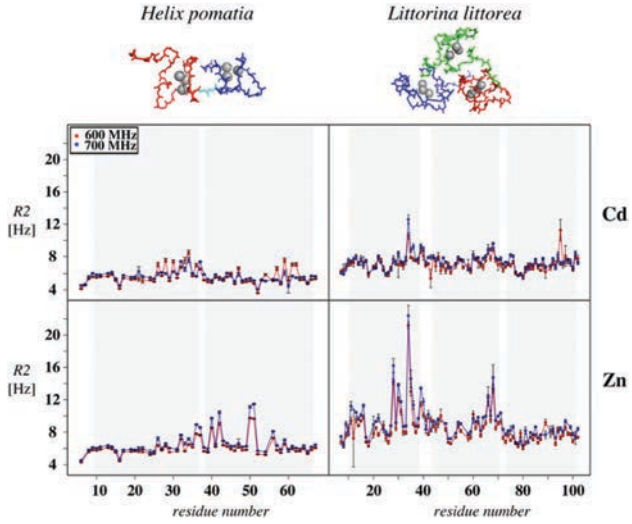


Fig. 9 Conformational exchange effects in MTs. ^{15}N *R2* rates of the Cd (top) and Zn complexes (bottom) of CdMTs of *Helix pomatia* (left) and *Littorina littorea* (right), recorded at 600 (red) and 700 (blue) MHz. Contributions from conformational exchange can be detected for residues with largely increased *R2* rates.

as phosphate,[3] displaying higher concentrations in deeper oceanic waters and exhibit a depletion towards neritic surface waters down to concentrations of 10^{-9} 10^{-10} M[78,79] (Fig. 11). Complex formation of Cd^{2+} with inorganic and organic ligands further decreases its biological availability in neritic seawater realms.[80] The situation changes drastically in the littoral zone, where marine habitats come into contact with the continental earth crust, in which natural Cd background concentrations, at 10^{-8} 10^{-7} M, can be up to 100 times higher than those of superficial seawater[81] (Fig. 11). Decreasing seawater salinities in the supra-littoral zone can even enhance the availability of Cd^{2+} for animals.[82,83]

Gastropods of these habitats have adapted to fluctuating environmental conditions[84] but also had to cope with increasing Cd concentrations. Inactivation of Cd^{2+} ions by metal-selective MTs would, therefore, confer on them a physiological advantage[85] over energy-consuming activities for continuously re-adjusting intracellular Cd concentrations.[86] Upon adaptation to terrestrial life, gastropods have learned to cope with alternating and adverse environmental conditions.[87,88] Hence, the conservation of Cd-selective MTs may also be beneficial for land snails[16] (Fig. 11).

In contrast to terrestrial snails, freshwater species of Caenogastropoda and Heterobranchia have lost their Cd binding selectivity, likely because natural Cd background concentrations in freshwater habitats with about 10^{-10} 10^{-12} M are the lowest of any snail habitat on earth[89] (Fig. 11).

The multitude of metal-selective MT variants naturally occurring in snails offers the unique possibility to apply them as models for optimization of MT metal binding features through experiments in the laboratory. This may promote our understanding of about how amino acid replacements modify metal selectivity in MTs, and could have implications for the design of novel artificial Cd-binding proteins for the sake of basic research or for application in environmental

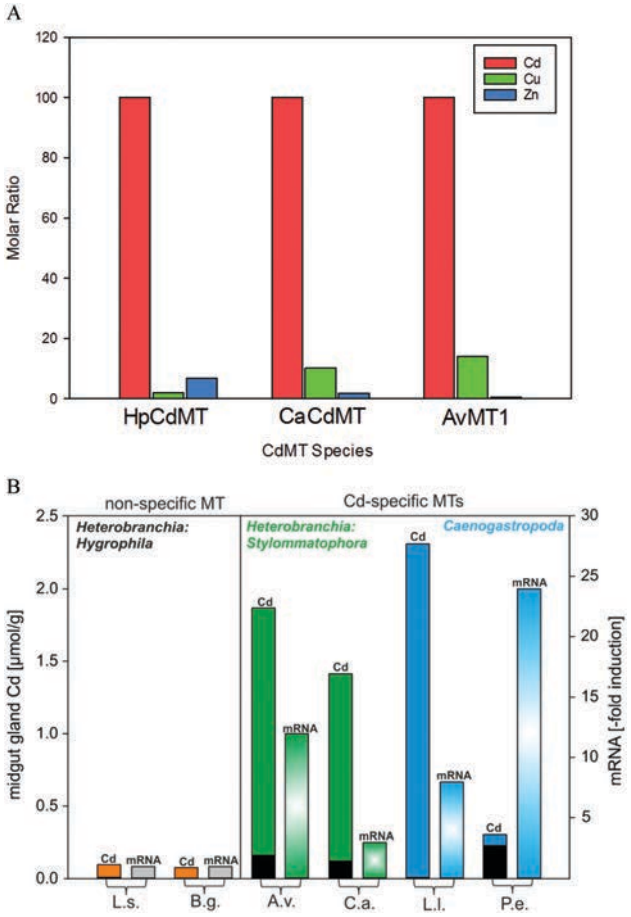


Fig. 10 (A) Zinc and copper content in native CdMTs isolated from midgut gland preparations of Cd-exposed snails. Values are given as molar ratios in % of Cd content. HpCdMT, *Helix pomatia* CdMT; CaCdMT, *Cornu aspersum* CdMT; AvMT1, *Arion vulgaris* AvMT1 (CdMT). MTs of species, for which metal contents were analyzed for the first time in this study are marked with a red star. Molar ratios for HpCdMT were re-drawn from data reported in ref. 14. (B) Cd accumulation and -fold induction of *MT* gene mRNA transcription in midgut gland of Cd-exposed snails with unselective and with Cd-selective MTs. Left-hand part of the graph: Cd accumulation (orange bars) and -fold *MT mRNA* induction (grey bars) in two freshwater snails (*Lymnaea stagnalis* and *Biomphalaria glabrata*, both *Hygrophila*) with unspecific MTs. Right-hand part of the graph: Cd accumulation (Cd) and -fold *MT mRNA* induction for snails possessing Cd-selective MTs, with respective values for *Arion vulgaris* and *Cornu aspersum* belonging to the clade of Stylommatophora (black/green and light green bars), and for *Littorina littorea* and *Pomatias elegans* belonging to the clade of Caenogastropoda (black/blue and light blue bars). Cd contents and mRNA induction data of species analysed *de novo* for the present study are marked with a red star. The other values were re-drawn from data reported in ref. 96, 23, 97 and 16. Species abbreviations: L.s., *Lymnaea stagnalis*; B.g., *Biomphalaria glabrata*; A.v., *Arion vulgaris*; C.a., *Cournu aspersum*; L.l., *Littorina littorea*; P.e., *Pomatias elegans*.

bioremediation.[90,91] It underscores once more the true model character of metal-selective snail MTs.

Conclusions

Some important conclusions are derived from our findings: first, presence of Cd has been a continuous evolutionary

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

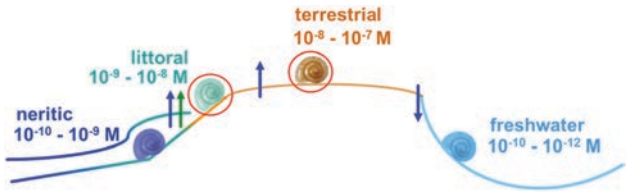


Fig. 11 Molar Cd background concentrations along the axis of evolutionary habitat adaptation in Gastropoda. Data start from neritic superficial seawater realms, through littoral and terrestrial habitats up to freshwater environments. Blue arrows indicate an increase (upward) or a decrease (downward) of molar Cd concentrations along the habitat axis. The green upward arrow at the transition zone between neritic and littoral habitats symbolizes the increasing availability of Cd due to decreasing concentrations of complexing ligands and decreasing salinity. Snail symbols encircled in red indicate the gain of Cd-selective MTs.

stimulus through the last 430 million years, driving convergent evolution and optimization of Cd-selective MTs in gastropod clades. Second, the C-terminal domain of Cd-selective gastropod MTs has been subjected to a high pressure for evolutionary conservation, which we attribute to its important role for immobilizing Cd^{2+}. Third, gastropod adaptation to habitats with different Cd background levels has triggered MT diversification towards partial or complete loss of metal selectivity. Fourth, the natural evolution in snails of an array of differently metal-selective MT variants designates them MTs as model molecules and indicates that it is possible to design artificial Cd-selective peptides.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by a cooperation grant to R. D. and O. Z. from the Austrian Science Fund and the Swiss National Science Foundation (DACH grant no. I 1482-N28), and a grant from the Forschungskredit of the University of Zurich to C. B. (grant no. FK-18-083). B. E. was supported by a grant for young scientists at the University of Innsbruck. Researchers from Barcelona want to acknowledge the Spanish Ministerio de Ciencia e Innovacion and FEDER for the projects BIO2015-67358-C2-2-P (M. C. and O. P.) and BIO2015-67358-C2-1-P (R. A.). M. C. and O. P. are members of the ``Grup de Recerca de la Generalitat de Catalunya , ref. 2017SGR-864. We thank Thomas Ostermann for help with the transcriptome assembly and Philipp Andesner (both: University of Innsbruck, Austria) for producing the neutral marker sequences. Transcriptomic data have been achieved in part using the HPC infrastructure of the University of Innsbruck. We also thank the Servei d Analisi Quımica (SAQ) at the Universitat Autonoma de Barcelona (ICP-AES, CD, UV-vis, ESI-MS) for allocating instrument time and three anonymous reviewers for constructive comments on an earlier version of the manuscript.

References

1 K. H. Wedepohl, *Geochim. Cosmochim. Acta*, 1995, **59**, 1217 1232.
2 J. T. Cullen and M. T. Maldonado, in *Cadmium: From Toxicity to Essentiality. Metal Ions in Life Sciences*, ed. A. Sigel, H. Sigel and R. Sigel, Springer, Dordrecht, 2013, pp. 31 62.
3 T. W. Lane and F. M. M. Morel, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 4627 4631.
4 T. W. Lane, M. A. Saito, G. N. George, I. J. Pickering, R. C. Prince and F. M. M. Morel, *Nature*, 2005, **434**, 42.
5 A. Martelli, E. Rousselet, C. Dycke, A. Bouron and J. M. Moulis, *Biochimie*, 2006, **88**, 1807 1814.
6 A. M. Sandbichler and M. Hockner, *Int. J. Mol. Sci.*, 2016, **17**(1), 1 22, DOI: 10.3390/ijms17010139.
7 C. D. Klaassen, J. Liu and S. Choudhuri, *Annu. Rev. Pharmacol. Toxicol.*, 1999, **39**, 267 294.
8 R. D. Palmiter, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 8428 8430.
9 S. G. Bell and B. L. Vallee, *ChemBioChem*, 2009, **10**, 55 62.
10 A. Krezel and W. Maret, *Int. J. Mol. Sci.*, 2017, **18**, 1237.
11 C. A. Blindauer and O. I. Leszczyszyn, *Nat. Prod. Rep.*, 2010, **27**, 720 741.
12 G. Isani and E. Carpene, *Biomolecules*, 2014, **4**, 435 457.
13 S. Calatayud, M. Garcia-Risco, N. S. Rojas, L. Espinosa-Sanchez, S. Artime, O. Palacios, C. Canestro and R. Albalat, *Metallomics*, 2018, **10**, 1585 1594.
14 R. Dallinger, B. Berger, P. Hunziker and J. H. R. Kagi, *Nature*, 1997, **388**, 237 238.
15 O. Palacios, A. Pagani, S. Perez-Rafael, M. Egg, M. Hockner, A. Brandstatter, M. Capdevila, S. Atrian and R. Dallinger, *BMC Biol.*, 2001, **9**(4), 1 20, DOI: 10.1186/1741-7007-9-4.
16 L. Schmielau, M. Dvorak, M. Niederwanger, N. Dobieszewski, V. Pedrini-Martha, P. Ladurner, J. Rodrıguez-Guerra Pedregal, J. D. Marechal and R. Dallinger, *Sci. Total Environ*, 2019, **648**, 561 571.
17 S. Perez-Rafael, A. Mezger, B. Lieb, R. Dallinger, M. Capdevila, O. Palacios and S. Atrian, *J. Inorg. Biochem.*, 2012, **108**, 84 90.
18 O. Palacios, S. Perez-Rafael, A. Pagani, R. Dallinger, S. Atrian and M. Capdevila, *J. Biol. Inorg. Chem.*, 2014, **19**, 923 935.
19 S. Perez-Rafael, F. Monteiro, R. Dallinger, S. Atrian, O. Palacios and M. Capdevila, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 1694 1707.
20 C. Baumann, A. Beil, S. Jurt, M. Niederwanger, O. Palacios, M. Capdevila, S. Atrian, R. Dallinger and O. Zerbe, *Angew. Chem., Int. Ed.*, 2017, **56**, 4617 4622.
21 M. Niederwanger, S. Calatayud, O. Zerbe, S. Atrian, R. Albalat, M. Capdevila, O. Palacios and R. Dallinger, *Int. J. Mol. Sci.*, 2017, **18**(7), 1 15, DOI: 10.3390/ijms18071457.
22 O. Palacios, E. Jimenez-Martı, M. Niederwanger, S. Gil-Moreno, O. Zerbe, S. Atrian, R. Dallinger and M. Capdevila, *Int. J. Mol. Sci.*, 2017, **18**, 1 16.
23 M. Dvorak, R. Lackner, M. Niederwanger, C. Rotondo, R. Schnegg, P. Ladurner, V. Pedrini-Martha, W. Salvenmoser, L. Kremser, H. Lindner, M. Garcia-Risco, S. Calatayud, R. Albalat,

This journal is © The Royal Society of Chemistry 2020

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

O. Palacios, M. Capdevila and R. Dallinger, *Metallomics*, 2018, **10**, 1638 1654.
24 R. Dallinger, M. Chabicovsky and B. Berger, *Environ. Toxicol. Chem.*, 2004, **23**, 890 901.
25 K. M. Jorger, I. Stoger, Y. Kano, H. Fukuda, T. Knebelsberger and M. Schrodl, *BMC Evol. Biol.*, 2010, **10**(323), 1 20, DOI: 10.1186/1471-2148-10-323.
26 M. Hockner, K. Stefanon, D. Schuler, R. Fantur, A. De Vaufleury and R. Dallinger, *J. Exp. Zool., Part A*, 2009, **311**, 776 787.
27 M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev, *Nat. Biotechnol.*, 2011, **29**, 644 652.
28 A. Priyam, B. J. Woodcroft, V. Rai, A. Munagala, I. Moghul, F. Ter, M. A. Gibbins, H. Moon, G. Leonard, W. Rumpf and Y. Wurm, *bioRxiv*, 2015, 033142.
29 V. Pedrini-Martha, M. Niederwanger, R. Kopp, R. Schnegg and R. Dallinger, *PLoS One*, 2016, **11**, 1 19.
30 R. C. Edgar, *Nucleic Acids Res.*, 2004, **32**, 1792 1797.
31 A. Stamatakis, *Bioinformatics*, 2014, **30**, 1312 1313.
32 F. Ronquist, M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck, *Syst. Biol.*, 2012, **61**, 539 542.
33 E. Pruesse, J. Peplies and F. O. Glockner, *Bioinformatics*, 2012, **28**, 1823 1829.
34 J. Trifinopoulos, L. T. Nguyen, A. von Haeseler and B. Q. Minh, *Nucleic Acids Res.*, 2016, **44**, W232 W235.
35 D. Posada and K. A. Crandall, *Bioinformatics*, 1998, **14**, 817 818.
36 P. Schanda, E. Kupce and B. Brutscher, *J. Biomol. NMR*, 2005, **33**, 199 211.
37 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins: Struct., Funct., Bioinf.*, 2005, **59**, 687 696.
38 S. Meiboom and D. Gill, *Rev. Sci. Instrum.*, 1958, **29**, 688 691.
39 A. Dinapoli and A. Klussmann-Kolb, *Mol. Phylogenet. Evol.*, 2010, **55**, 60 76.
40 B. Dayrat, M. Conrad, S. Balayan, T. R. White, C. Albrecht, R. Golding, S. R. Gomes, M. G. Harasewych and A. M. de Frias Martins, *Mol. Phylogenet. Evol.*, 2011, **59**, 425 437.
41 F. Zapata, N. G. Wilson, M. Howison, S. C. S. Andrade, K. M. Jorger, M. Schrodl, F. E. Goetz, G. Giribet and C. W. Dunn, *Proc. R. Soc. B*, 2014, **281**, 20141739.
42 O. Razkin, B. J. Gomez-Moliner, C. E. Prieto, A. Martınez-Ortı, J. R. Arrebola, B. Munoz, L. J. Chueca and M. J. Madeira, *Mol. Phylogenet. Evol.*, 2015, **83**, 99 117.
43 T. J. Cunha and G. Giribet, *Proc. R. Soc. London, Ser. B*, 2019, **286**, 20182776.
44 H. Yang, J. Zhang, J. Xia, J. Yang, J. Guo, Z. Deng and M. Luo, *Int. J. Mol. Sci.*, 2018, **19**, 3646.
45 A. Klussmann-Kolb, A. Dinapoli, K. Kuhn, B. Streit and C. Albrecht, *BMC Evol. Biol.*, 2008, **8**(57), 1 16, DOI: 10.1186/1471-2148-8-57.
46 A. Beil, S. Jurt, R. Walser, T. Schonhut, P. Guntert, O. Palacios, S. Atrian, M. Capdevila, R. Dallinger and O. Zerbe, *Biochemistry*, 2019, **58**, 4570 4581.
47 V. Pedrini-Martha, S. Koll, M. Dvorak and R. Dallinger, *Int. J. Mol. Sci.*, 2020, **21**(5), 1 21, DOI: 10.3390/ijms21051631.
48 R. Orihuela, J. Domenech, R. Bofill, C. You, E. A. Mackay, J. H. R. Kagi, M. Capdevila and S. Atrian, *J. Biol. Inorg. Chem.*, 2008, **13**, 801 812.
49 N. Romero-Isart and M. Vasak, *J. Inorg. Biochem.*, 2002, **88**, 388 396.
50 J. H. R. Kagi, in Metallothionein III: Biological Roles and Medical Implications, Proceedings of the Metallothionein International Conference, 1992, Tsukuba, Japan, Birkhauser Verlag, Boston, 1993, vol. 3, pp. 29 56.
51 E. H. Fischer and E. W. Davie, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 3333 3334.
52 P. M. Gehrig, C. You, R. Dallinger, C. Gruber, M. Brouwer, J. H. R. Kagi and P. E. Hunziker, *Protein Sci.*, 2000, **9**, 395 402.
53 K. M. Kocot, K. M. Halanych and P. J. Krug, *Mol. Phylogenet. Evol.*, 2013, **69**, 764 771.
54 R. Dallinger, M. Chabicovsky, E. Hodl, C. Prem, P. Hunziker and C. Manzl, *Am. J. Physiol.: Regul., Integr. Comp. Physiol.*, 2005, **289**, R1185 R1195.
55 M. Hockner, K. Stefanon, A. De Vaufleury, F. Monteiro, S. Perez-Rafael, O. Palacios, M. Capdevila, S. Atrian and R. Dallinger, *Biometals*, 2011, **24**, 1079 1092.
56 E. B. Naimark and A. V. Markov, *Biol. Bull. Rev.*, 2011, **1**, 71 81.
57 S. E. Grasby, B. Beauchamp, D. P. G. Bond, P. Wignall, C. Talavera, J. M. Galloway, K. Piepjohn, L. Reinhardt and D. Blomeier, *Geol. Soc. Am. Bull.*, 2015, **127**, 1331 1347.
58 D. P. G. Bond and S. E. Grasby, *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2017, **478**, 3 29.
59 G. Keller, P. Mateo, J. Punekar, H. Khozyem, B. Gertsch, J. Spangenberg, A. M. Bitchong and T. Adatte, *Gondwana Res.*, 2018, **56**, 69 89.
60 S. Aizawa and H. Akaiwa, *Chem. Geol.*, 1992, **98**, 103 110.
61 E. Porebska and Z. Sawłowicz, *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 1997, **132**, 343 354.
62 M. L. W. Tuttle, G. N. Breit and M. B. Goldhaber, *Appl. Geochem.*, 2009, **24**, 1565 1578.
63 Y. Liu, T. Xiao, R. B. Perkins, J. Zhu, Z. Zhu, Y. Xiong and Z. Ning, *J. Geochem. Explor.*, 2017, **176**, 42 49.
64 H. J. Brumsack, *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2006, **232**, 344 361.
65 F. Batifol, C. Boutron and M. De Angelis, *Nature*, 1989, **337**, 544 546.
66 V. Pedrini-Martha, R. Schnegg, P. E. Baurand, A. deVaufleury and R. Dallinger, *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.*, 2017, **199**, 38 47.
67 P. E. Baurand, V. Pedrini-Martha, A. De Vaufleury, M. Niederwanger, N. Capelli, R. Scheifler and R. Dallinger, *PLoS One*, 2015, **10**, 1 14.
68 P. E. Baurand, R. Dallinger, M. Niederwanger, N. Capelli, V. Pedrini-Martha and A. de Vaufleury, *Environ. Sci. Pollut. Res.*, 2016, **23**, 3062 3067.

Published on 10 March 2020. Downloaded by Universitat de Barcelona on 7/5/2021 10:11:43 AM.

69 M. Chabicovsky, H. Niederstatter, R. Thaler, E. Hodl, W. Parson, W. Rossmanith and R. Dallinger, *Toxicol. Appl. Pharmacol.*, 2003, **190**, 25 36.
70 F. Hispard, D. Schuler, A. De Vaufleury, R. Scheifler, P. M. Badot and R. Dallinger, *Environ. Toxicol. Chem.*, 2008, **27**, 1533 1542.
71 H. Irving and R. J. P. Williams, *J. Chem. Soc.*, 1953, 3192 3210.
72 M. Erk and B. Raspor, *Anal. Chim. Acta*, 1998, **360**, 189 194.
73 G. Meloni, V. Sonois, T. Delaine, L. Guilloreau, A. Gillet, J. Teissie, P. Faller and M. Vasak, *Nat. Chem. Biol.*, 2008, **4**, 366 372.
74 E. Hodl, E. Felder, M. Chabicovsky and R. Dallinger, *Cell Tissue Res.*, 2010, **341**, 159 171.
75 W. Maret and J. M. Moulis, in *Cadmium: From Toxicity to Essentiality. Metal Ions in Life Sciences*, ed. A. Sigel, H. Sigel and R. Sigel, Springer, Dordrecht, 2013, pp. 1 29.
76 G. Choong, Y. Liu and D. M. Templeton, *Chem. –Biol. Interact.*, 2014, **211**, 54 65.
77 D. Faubel, M. Lopes-Lima, S. Freitas, L. Pereira, J. Andrade, A. Checa, H. Frank, T. Matsuda and J. Machado, *Mar. Freshwater Behav. Physiol.*, 2008, **41**, 93 108.
78 E. A. Boyle, F. Sclater and J. M. Edmond, *Nature*, 1976, **263**, 42 44.
79 W. Abouchami, S. J. G. Galer, H. J. W. De Baar, R. Middag, D. Vance, Y. Zhao, M. Klunder, K. Mezger, H. Feldmann and M. O. Andreae, *Geochim. Cosmochim. Acta*, 2014, **127**, 348 367.
80 K. W. Bruland, *Limnol. Oceanogr.*, 1992, **37**, 1008 1017.
81 S. Sposito, in *Metal Ioins in Biological Systems, Concepts on Metal ion Toxicity*, ed. H. Sigel and A. Sigel, Marcel Dekker Inc., New York, 1986, vol. 20, pp. 1 20.
82 P. Bjerregaard and M. H. Depledge, *Mar. Biol.*, 1994, **119**, 385 395.
83 G. Roesijadi, *Mar. Environ. Res.*, 1994, **38**, 147 168.
84 K. B. Storey, B. Lant, O. O. Anozie and J. M. Storey, *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.*, 2013, **165**, 448 459.
85 T. E. English and K. B. Storey, *J. Exp. Biol.*, 2003, **206**, 2517 2524.
86 M. H. Depledge and P. S. Rainbow, *Comp. Biochem. Physiol.*, 1990, **97C**, 1 7.
87 S. Slotsbo, K. V. Fisker, L. M. Hansen and M. Holmstrup, *J. Comp. Physiol., B*, 2011, **181**, 1001 1009.
88 T. Bishop and M. D. Brand, *J. Exp. Biol.*, 2000, **203**, 3603 3612.
89 C. A. Mebane, in US Department of the Interior & US Geological Survey, Scientific Investigations Report 2006-5245, Reston, Virginia, v. 1.2., 2010, p. 130.
90 A. F. A. Peacock and V. L. Pecoraro, in *Cadmium: From Toxicity to Essentiality (Metal Ions in Life Sciences)*, ed. A. Sigel, H. Sigel and R. Sigel, Springer, Dordrecht, 2013, vol. 11, pp. 303 338.
91 M. Mejare and L. Bulow, *Trends Biotechnol.*, 2001, **19**, 67 73.
92 K. M. Kocot, J. T. Cannon, C. Todt, M. R. Citarella, A. B. Kohn, A. Meyer, S. R. Santos, C. Schander, L. L. Moroz, B. Lieb and K. M. Halanych, *Nature*, 2011, **477**, 452 456.
93 D. J. Colgan, W. F. Ponder, E. Beacham and J. Macaranas, *Mol. Phylogenet. Evol.*, 2007, **42**, 717 737.
94 C. M. Wade, P. B. Mordan and B. Clarke, *Proc. R. Soc. B*, 2001, **268**, 413 422.
95 E. V. Soldatenko and A. A. Petrov, *J. Morphol.*, 2019, **280**, 508 525.
96 M. Niederwanger, M. Dvorak, R. Schnegg, V. Pedrini-Martha, K. Bacher, M. Bidoli and R. Dallinger, *Int. J. Mol. Sci.*, 2017, **18**, 1747.
97 D. Benito, M. Niederwanger, U. Izagirre, R. Dallinger and M. Soto, *Int. J. Mol. Sci.*, 2017, **18**, 1815.

This journal is © The Royal Society of Chemistry 2020

ANNEX IV: SUPPLEMENTARY MATERIAL

METALLOMICS REVEALS A PERSISTING IMPACT OF CADMIUM ON THE EVOLUTION OF METAL-SELECTIVE SNAIL METALLOTHIONEINS

The file includes:

Table S1: PCR conditions for the neutral markers COI and 18SrDNA: a) Thermal cycles followed [1] with the annealing temperatures adjusted according to the particular primers, samples, and polymerase used. b) PCR mix for MyTaq Polymerase c) PCR mix for Rotorgene probe-PCR for samples which did not amplify with MyTaq Polymerase

a) Thermal cycles:	**step**	**temperatur e**	**time**	**cycles**
	Intitial step	95°C	4 min	1x
_	denaturation per cycle	95°C	30 s	38x
	annealing per cycle	50 - 65°C	45 s	
	elongation per cycle	72°C	150 s	
	final step	72°C	10 min	1x
b) PCR mix for MyTaq Polymerase	**reagent**	**volume (µl)**		
	dH_2O	5.55		
	5x PCR Buffer	2.00		
	BSA 100µg/µl (final conc. 10µg/µl)	1.00		
	each primer (final conc. 0.2µM)	0.20		
	MyTaq polymerase Bioline	0.05		
	template DNA	1.00		
c) PCR mix for Rotorgene probe-PCR	**reagent**	**volume (µl)**		
	dH_2O	3.6		
	2x Rotorgene Mastermix Qiagen	5.00		
	each primer (final conc. 0.2µM)	0.20		
	template DNA	1.00		

Table S2 – Lists of primers and cycling parameters applied for PCR confirmation of MT sequences

2A: Description of applied gene-specific primers used for Long Distance PCR

Species	*MT Isoform*	*Primer*	*Sequence 5'-3'*	*Tm [°C]*	*Tm_{An} [°C]*
Neritina pulligera	MT1	Fw	TCA TCG TAC TCA CCA CTC CA	58.4	57
		Rev	TGG ACA AGG GGG TAA ACG A	57.5	
	MT2	Fw	CAG TGT CGT GTG CAG AGA GA	60.5	58
		Rev	CAG TCC AGA ATG TCT TCT TCC	59.5	
Deroceras reticulatum	CdMT	Fw	GGC TTA AAC GAG GAC CTG	56.3	54
		Rev	AGC TTG ATT ATT TGT CAG GTA C	56.4	
	CdMT2	Fw	CAG TCC AGA ATG TCT TCT TCC	59.5	54.5
		Rev	CTT GAA TCT TGA GAG TGG CA	56.4	
Marisa cornuarietis	MT1	Fw	CGT GTT GTT GTG GTT GTT	51.4	49.5
		Rev	GCA CAT ACA GAC AGG AAT	51.4	
	MT2	Fw	GTC GAC AAC CGA ACC ATC CT	60.5	58.5
		Rev	GCA GAC ATC GTG TCA CTT GC	60.5	
Anentome helena	MT	Fw	TCA CTT TTC CGT GAC TTG CT	56.4	54.5
		Rev	AAA GGC TTA CAA GGC GGG G	59.5	
Aplysia californica	MT	Fw	CGT ACG TCA CCA CCA GAG AT	60.5	54
		Rev	AAT GCT CTC CGC ACG CTA	56.3	
Alinda biplicata	MT1	Fw	GTC AAG TTT ACG ATC TGC G	52.9	52
		Rev	TCA ATC CCA TCC AAC CAA C	52.9	
	MT2	Fw	AGG TCA AGT TTA CGA TCT GCG T	60.1	58
		Rev	CCT AAC CAC ACA ACT CCC GT	60.5	
	9md-MT	Fw	TCC TCG TAC CGT ATC TCA GC	60.5	58
		Rev	ACG CTC AAT GTA GTC TTA TGT TGC	62.0	
	10Md-MT	Fw	TCC TCG TAC CGT ATC TCA GC	60.5	58
		Rev	ACG CTC AAT GTA GTC TTA TGT TGC	62.0	
Patella vulgata	MT1	Fw	GAC GCA CAC AAT CAG A	48.1	47
		Rev	AGC CAT TGC ACG AGA C	50.6	
	MT2	Fw	AGT AAG ACA CCC ACA GTC AG	58.4	53
		Rev	ACG GGG ATA GAA GAC ACA A	55.0	
Lottia gigantea	MT1	Fw	CTT CTC AAA GCT TCT ACA GAA T	56.4	53
		Rev	CAA GAT TGT GTG CAT GAT GAA	55.4	
	MT2	Fw	CCG CAG AGT ACA GCA ATT CAT A	60.1	55.5
		Rev	GGT ATG TGA GTC AAC AAA GAG	57.5	

2B: Applied cycling conditions, using the Advantage 2 polymerase PCR system (see Material and Methods).

	Temp [°C]	*Time*	*Cycles*
Denaturation	95	1 min	1
Denaturation	95	30 sec	30
Annealing	See table 1	30 sec	
Extension	68	40 sec – 1 min	
Final Extension	68	5 min	1

Table S3: List of gastropod Metallothionein sequences applied for the present study. Gastropod clades and major lineages (first column) with family names (second column), species names with consecutive numbers (bold blue letter fond in square brackets) (third column), MT designations (fourth column), sequence data source and databank deposition (fifth column), PCR confirmation (sixth column), kind of metal selectivity (seventh column), reference numbers (eighth column) and consecutive sequence numbers (ninth column) are shown. MT sequences with red consecutive numbers (last column) were sequenced *de novo* and are published in the present study for the first time. All other MT sequences (last column, black consecutive numbers) are reported from previous studies (see references) or were identified from publically available databases (see material and methods). Metal selectivity features (Cd – red; Cu – blue; yellow - unspecific) of single MTs (seventh column) were either directly determined by MS methods (🖙), derived from other kinds of experimental evidence combined with sequence homology (●), or only deduced from sequence homology by alignment with related sequences of known metal-binding character (no symbol).

Clade	*Family*	*Species* **[Species Number]**	*MT Name*	*Data source and Acc. No.*	*PCR-confirmed*	*Kind of Metal selectivity*	*References*	*Sequence No.*
Patellogastropoda	*Lottiidae*	*Lottia gigantea* **[1]**	MT1	GenBank MK770430	yes	Cd 🖙		**1**
			MT2	GenBank MK770431	yes	Cd 🖙		**2**
	Nacellidae	*Nacella polaris* **[2]**	MT	SRA; SRR2061227	no	Cd		3
	Patellidae	*Patella vulgata* **[3]**	MT1	GenBank MK443472	yes	Cd		**4**
			MT2	GenBank MK443472	yes	Cd		**5**
Vetigastropoda	*Fissurellidae*	*Megathura crenulata* **[4]**	MT	GenBank AY102647.1	-	Unspecific 🖙	[2,3]	6
	Haliotidae	*Haliotis diversicolor* **[5]**	MT	GenBank EU071824.1	-	Unspecific		7
		Haliotis discus hannai **[6]**	MT	GenBank KT895222.1	-	Unspecific		8
		Haliotis tuberculata **[7]**	MT	GenBank GEAU01019532.1	no	Unspecific		9
		Haliotis laevigata **[8]**	MT	GenBank GFTT01185788.1	no	Unspecific		10
	Tegulidae	*Tegula atra* **[9]**	MT	GenBank GFBI01039752.1	no	Unspecific		11
Neritimorpha	*Neritidae*	*Nerita peloronta* **[10]**	MT1	SRA; SRX644686	no	Unknown		12
			MT2	SRA; SRX644686	no	Unknown		13
		Neritina pulligera **[11]**	MT1	GenBank MK577683	yes	Unknown		**14**
			MT2	GenBank MK577684	yes	Unknown		**15**
	Titiscaniidae	*Titiscania limacina* **[12]**	MT	SRA; SRX644702	no	Unknown		16

Caenogastropoda	*Littorinidae*	*Littorina littorea* **[13]**	MT	GenBank AY034179.1	-	Cd	[3,4]	17
	Pomatiidae	*Pomatias elegans* **[14]**	MT1	GenBank KY636305.1	-	Cd ●	[3]	18
			MT2	GenBank KY646306.1	-	Cd ●	[3]	19
	Nassariidae	*Tritia obsoleta* **[15]**	MT	GenBank FK716482.1	no	Cd	[3]	20
	Calyptraeidae	*Bostrycapulus sp.* **[16]**	MT	SRA; SRP042651	no	Unknown		21
		Crepidula fornicata **[17]**	MT	SRA; SRP042651	no	Unknown		22
	Ampullariidae	*Pomacea bridgesii* **[18]**	MT1	GenBank KY963504.1	-	Unspecific		23
			MT2	GenBank KY963505.1	-	Unspecific		24
		Pomacea canaliculata **[19]**	MT20-III like variant X-1 like	GenBank XM_025255587.1	-	Unspecific		25
			MT20-III like variant X-2 like	GenBank XM_025255589.1	-	Unspecific		26
			MT20-III like variant X-3 like	GenBank XM_025255590.1	-	Unspecific		27
		Marisa cornuarietis **[20]**	MT1	GenBank MK577687	yes	Unspecific		**28**
			8md-MT	GenBank MK577688	yes	Unspecific		**29**
	Buccinidae	*Anentome helena* **[21]**	MT	GenBank MK577685	yes	Unspecific		**30**
Heterobranchia	*Aplysiidae*	*Aplysia californica* **[22]**	MT	GenBank MK577686	yes	Unknown		**31**
	Clionidae	*Clione limacina* **[23]**	MT	GenBank GESV01103158.1	no	Unknown		32
	Limacinidae	*Limacina retroversa* **[24]**	MT	GenBank GBXC01058817.1	no	Unknown		33

Heterobranchia ***Panpulmonata*** ***Hygrophila***	*Plakobranchidae*	*Elysia crispata* [25]	MT	GenBank KY963503	-	Unknown		**34**
		Elysia cornigera [26]	MT	GenBank GBRW01068275.1	no	Unknown		35
	Physidae	*Physella acuta* [27]	MT1	GenBank GU259686.1	-	Unspecific	5	36
			MT2	SRA; SRP042651	no	Unspecific		37
		Physella carolinae [28]	MT	SRA; SRP042651	no	Unspecific		38
		Physella hendersoni [29]	MT	SRA; SRP042651	no	Unspecific		39
		Physella gyrina [30]	MT	SRA; SRP042651	no	Unspecific		40
	Lymnaeidae	*Galba cubensis* [31]	MT	SRA; SRP042651	no	Unspecific		41
		Galba trunculata [32]	MT	SRA; SRP042651	no	Unspecific		42
		Lymnaea stagnalis [33]	MT	GenBank FX201894.1	yes	Unspecific ●		43
	Planorbidae	*Biomphalaria glabrata* [34]	MT + genomic variants	GenBank KY963495.1	-	Unspecific	5,6	44
Panpulmonata ***Stylommatophora***	*Agriolimacidae*	*Deroceras reticulatum* [35]	CdMT1	GenBank MK672823	yes	Cd		**45**
			CdMT2	eSnail database CL8321.Contig1	yes	Cd		46
	Arionidae	*Arion vulgaris* [36]	AvMT1	GenBank MF155618.1	-	Cd	7	47
			AvMT2	GenBank MF155619.1 MF155620.1	-	Cu	7	48
	Limacidae	*Lehmannia nyctelia* [37]	CdMT	eSnail database CL1241.Contig1	no	Cd		49
		Limax maximus [38]	CdMT	GenBank MH574544	-	Cd		50
Panpulmonata	*Clausiliidae*	*Alinda biplicata* [39]	CuMT1a	GenBank MK648138	yes	Cu		**51**

Stylommatophora			CuMT1b	SRX7671047	no	Cu		**52**
			CuMT2	GenBank MK639793	yes	Cu		**53**
			9md-MT	GenBank MK648139	yes	Cd		**54**
			10md-MT	GenBank MK648140	yes	Cd		**55**
	Cochlicellidae	*Cochlicella acuta* **[40]**	CdMT	eSnail database Unigene52262	no	Cd	[7]	56
			CuMT	eSnail database Unigene18701	no	Cu	[7]	57
			CdCuMT	eSnail database Unigene65576	no	Unspecific	[7]	58
	Helicidae	*Arianta arbustorum* **[41]**	CdMT (MTa)	Uniprot P55946	-	Cd ●		59
		Cornu aspersum **[42]**	CdMT	GenBank EF152281.1	-	Cd ●		60
			CuMT	GenBank EF178297.2	-	Cu ●		61
			CdCuMT	GenBank EF206312.1	-	Unspecific ●	[8,9]	62
		Cepaea hortensis **[43]**	CdMTV1	GenBank MH574545	-	Cd ●	[7]	63
			CdMTV2	GenBank MH574546	-	Cd ●	[7]	64
			CdCuMT	GenBank MH574547	-	Unspecific ●	[7]	65
		Cepaea nemoralis **[44]**	CdMT	GenBank GFLU01086707.1	no	Cd		66
			CuMT	GenBank GFLU01065470.1	no	Cu		67
			CdCuMT	GenBank GFLU01003885.1	no	Unspecific		68
		Helix pomatia **[45]**	CdMT	GenBank AF399740.1	-	Cd	[10–12]	69
			CuMT	GenBank AF399741.1	-	Cu	[10–12]	70
			CdCuMT1	GenBank KY420172.1	-	Unspecific		71
			CdCuMT2	GenBank KY420171.1	-	Unspecific		72
		Theba pisana **[46]**	MT	eSnail database CL8321.Contig1	no	Cu		73
	Helicoidea	*Nesiohelix samarangae* **[47]**	CdMT (MT)	GenBank EU437399.1	-	Cd		74

Table S4: Distance analysis with Identity and Similarity values (Means and Standard Deviations) of N-terminal versus C-terminal domains of Cd-selective MTs of Patellogastropoda, Caenogastropoda and Heterobranchia with the same sequences as adopted in **Fig. 3**, and for unspecific gastropod MTs**.** For N-terminal domains, calculation was performed without the N-terminal extension up to the first Cys position. For C-terminal domains, the linker region was included. Note that both similarity and identity scores are significantly higher for C-terminal domains. This supports the hypothesis of a higher degree of conservation for C-terminal domains of gastropod Cd-selective MTs. Calculations were performed using the openly accessible program package SIAS (Sequence Identity And Similarity) of the Immunomedicine group of the Universidad Complutense Madrid (http://imed.med.ucm.es/Tools/sias.html).

S4A: Cd-selective MTs

	N-terminal Domain	**C-terminal Domain**	**Difference (mean)**
Identity:			
All Cd-selective MTs	53.47% (± 20.78%)	66.17% (± 12.02%)	-12.70%
Patellogastropoda versus Littorinoidea	34.68% (± 3.07%)	66.56% (± 5.89%)	-31.88%
Patellogastropoda versus Stylommatophora	31.62% (± 2.69%)	59.10% (± 6.19%)	-27.47%
Littorinoidea versus Stylommatophora	47.49% (± 4.35%)	62.79% (± 5.97%)	-15.30%
Similarity:			
All Cd-selective MTs	58.01% (± 20.05)	71.47 (± 11.36%)	-13.46%
Patellogastropoda versus Littorinoidea	36.97% (± 3.82%)	69.31% (± 5.31%)	-32.34%
Patellogastropoda versus Stylommatophora	36.39% (± 2.99%)	63.17% (± 7.82%)	-26.78%
Littorinoidea versus Stylommatophora	54.70% (± 4.37%)	66.86% (± 7.64%)	-12.16%

S4B: Unspecific MTs

	N-terminal Domain	C-terminal Domain	Difference (mean)
Identity:			
All unspecific MTs	61.91% (± 16.11%)	59.10% (± 15.30%)	+2.80%
Vetigastropoda versus Ampullariidae	57.64% (± 5.43%)	55.86% (± 7.94%)	+1.78%
Vetigastropoda versus Hygrophila	49.36% (± 7.44%)	51.85% (± 8.33%)	-2.49%
Ampullariidae versus Hygrophila	52.75% (± 5.92%)	54.54% (± 5.63%)	-1.79%
Similarity:			
All unspecific MTs	67.087% (± 14.63%)	63.60% (± 14.60%)	+3.48%
Vetigastropoda versus Ampullariidae	62.35% (± 5.55%)	63.06% (± 7.54%)	-0.71%
Vetigastropoda versus Hygrophila	55.02% (± 7.87%)	56.14% (± 8.90%)	-1.12%
Ampullariidae versus Hygrophila	60.24% (± 7.30%)	60.03% (± 8.41%)	+0.21%

Figure S1:
Metal-binding domain organization and amino acid sequence alignment of Cu-selective (CuMTs, (framed in blue) and unspecific MT isoforms (called CdCuMTs, framed in green) from terrestrial snail and slug species of the order of Stylommatophora (clade Heterobranchia). Cys positions are underlaid in pink, conserved non-Cys positions through sequences of all species are underlaid in blue. Identical amino acid positions between pairwise aligned sequences are indicated by black stars. Domain boundaries of the N-terminal and the C-terminal domains (designated above the alignments as N and C in red letters) are symbolized by perpendicular bold red lines. The linker region between the N and the C-terminal domains are designated in black letters, their boundaries are symbolized by perpendicular dotted black lines. The gaps between the two domains indicate the lack of a second N-terminal domain (present in other gastropod MTs). MTs of species shown in red letters were sequenced for this study *de novo* and are published here for the first time. Sequences of all other species (in black letters) were identified from publically available publications or databases. Species for which metal selectivity features of respective MTs were documented experimentally by us through MS or NMR methods elsewhere are framed in blue. The bold red arrow on the right hand of the alignments points to sequence similarities between Cu-selective and unspecific MTs.

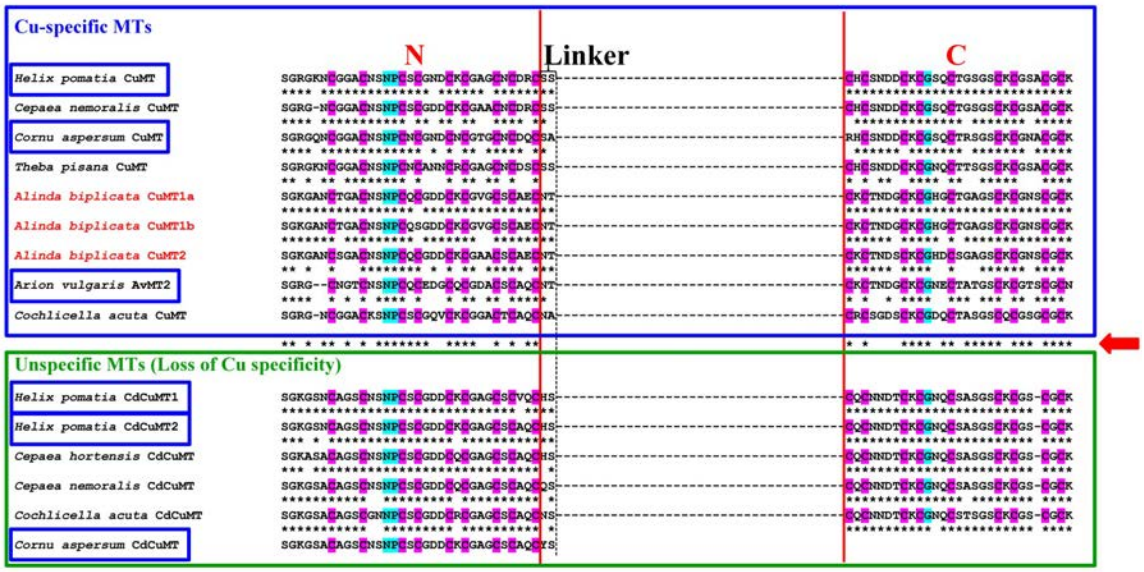


Figure S2:
Cu-selective MTs from terrestrial snails *Helix pomatia* and *Cornu aspersum* of the order of Stylommatophora (Heterobranchia). ESI-MS spectra of MT species recombinantly produced in different meal-enriched media (Cd, Zn and Cu) at the corresponding charge state (as indicated in the upper right corner) or deconvoluted. Spectra are re-drawn from data reported in [11] and [9].

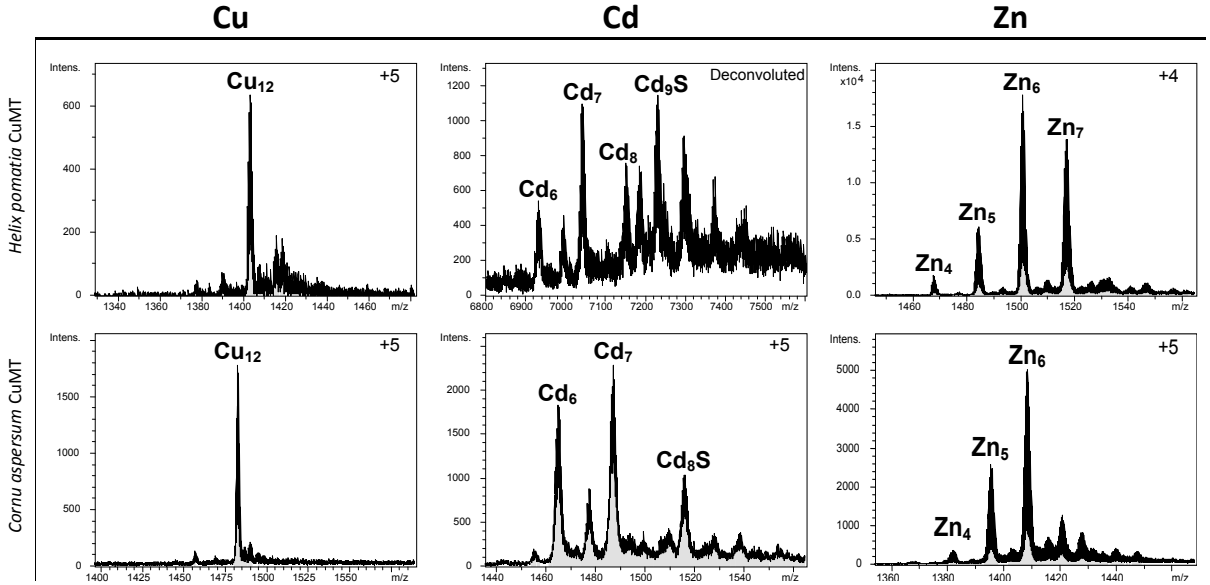
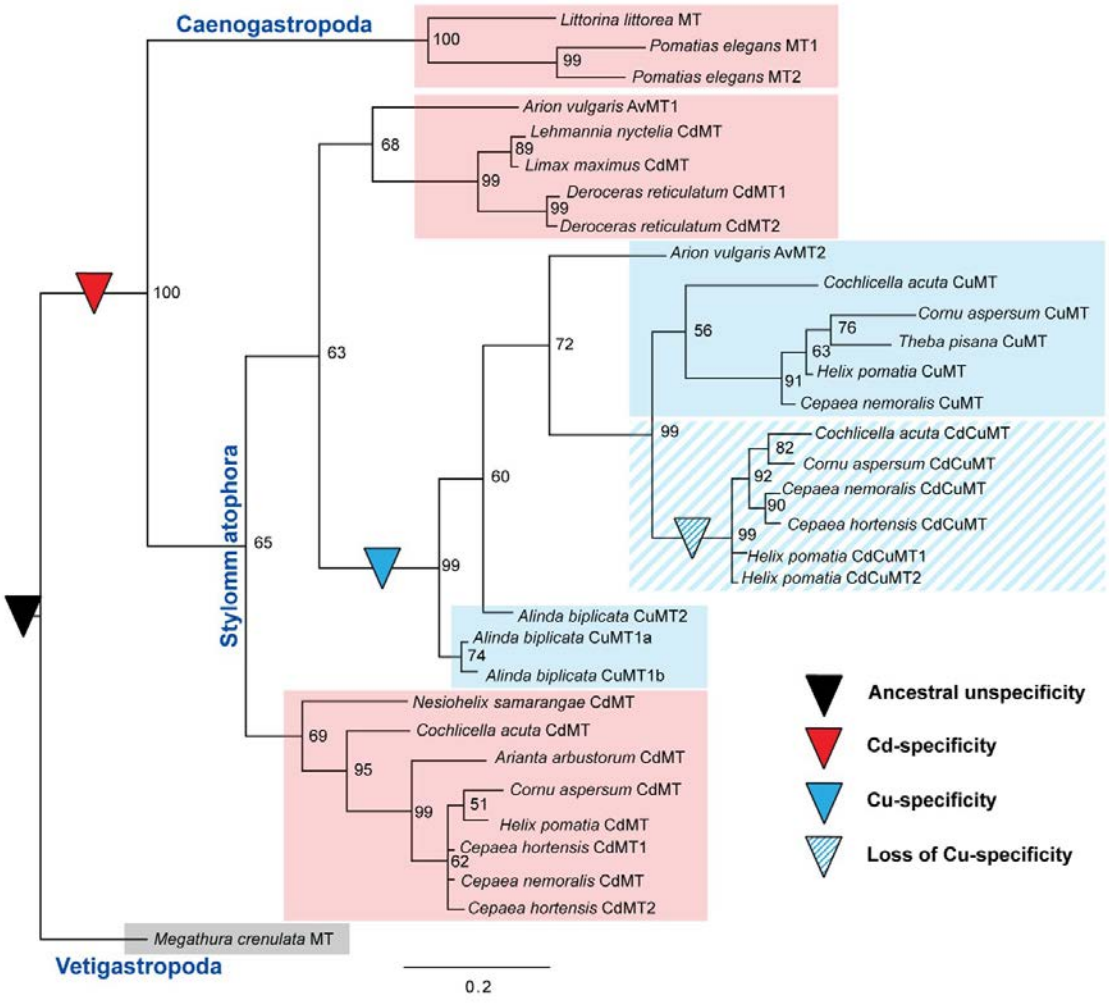


Figure S3:
Bayesian Inference tree with bootstrap support values of metal selectivity features in MTs of the clade of Caenogastropoda and in the order of Stylommatophora (a taxon of Heterobranchia with land snails and slugs), rooted against the unspecific MTs of Vetigastropoda (represented by *Megathura crenulata*, underlaid in gray). Ancient unspecificity of MTs at the root of Vetigastropoda is indicated by a black triangle. Gain of Cd selectivity in MTs at the root of divergence between Caenogastropoda and Heterobranchia (represented only by Stylommatophora) is indicated by a red triangle, species poossessing Cd-selective MTs are underlaid in pink. Gain of Cu selectivity in MTs of Stylommatophora is indicated by a blue triangle, respective species possessing Cu-selective MTs are underlaid in blue. Secondary loss of Cu selectivity in CdCuMTs of land snails and slugs is indicated by a blue hatched triangle, species with unspecific MT isoforms are highlighted within a blue hatched square. Bayesian inference calculations were made based on a manually edited MUSCLE alignment (see **alignment S5**) using the free software MrBayes (see Material and Methods)



Alignment S1:

Concatenated NEXUS alignment of the mitochondrial COI sequences (position 1-580, start=3rd codon position) and the ribosomal 18SrDNA sequences (position 581-2158; aligned using the SINA Alignment tool v. 1.2.11) of the nineteen gastropod species analysed for neutral phylogeny markers, applied for the mirrored Maximum Likelihood tree represented in **Figure 6**

Patella vulgata

CGGCCTTAGTATATTAATTCGAGCTGAACTTGGTCAGCCAGGCGCT---CTTTTAGGCGATGACCAACTTTATAATGTTATTGTTACTGCTCACGCTTTTGTTATAATTTTCTTTTTAGTAATGCCTATAATAATTGGGGGTTTTGGTAATTGATTAG
TGCCACTTATGCTTGGGGCCCCTGATATAGCTTTTCCCCGACTTAACAACATAAGATTTTGACTTCTCCCTCCCTCTCTATTTCTTCTTTTAGCCAGAAGAGCTGTTGAAGGTGGGGTAGGTACAGGTTGAACAGTTTATCCTCCTTTATCTTCAAAT
ATTGCCCACTCAGGCCCTTCGGTAGATTTAGCTATTTTTTCTCTCCACCTAGCTGGTGCCTCTTCAATTCTGGGAGCTGCTAATTTTATTAGAACGGCCGCAAACATTCGTTGAGAAGCTATGCAACTTGAACGAATGCCTTTGTTTGTCTGATCTGT
AAAGGTAACCGCTATTTTATTACTTCTATCTTTACCTGTTTTAGCCGGGGCAATTACTATACTATTAACTGACCGTAACTTTAATACTTGTTTCTTTGACCCTAGGGTTCA-GGCT-----TGTTCCTTTCGGGGAGCGAGCCGAAACTGCGAACGGC
TCATTAG-TCAGATAAGGTTCCTTGGCGAAAA-----GCGGGTCGGTTTAATGGATAACTGTGGTAATTCTAGAGCTAATACATGC--AA-CGCACCGT-GGTC-CCCCCC-T-TTCGTTTCTCGACAAAGAGGG------GG-AAACGGC-ATTTAT
TCTCTATACCAGATCGCC-CTAGCCTGCGAAGCTACCGTC--A------AAAGTA-GC-------------GACGGGGCGAA-GCGACAAGT-GTTGAA-ATCCGAATAACTGTG--CCGATCGCGGGGTCTTTC-CGGGC-----CCCCGACGACTT
TGCCATGAAGTGTCTGTCCCATCAATTGGCGATGGTCGGCGACCTGCCTACCACGGTGTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGCGAAACGGCTACCACCTCCAAGGAGGGCAGCAGGCGCGCAACTTACCCAATCC
CGACACGGGGAGGTAGCGACGAAAAATAACGTGGCGGGGCCCTTTCGTGGTCGCCGCGGGCGGAATGAGCGCAATTTAAAAGAGTGTGCGAGGAGCTATTGGAGGGCAAGCCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCCTATATCG
GAGTTGCTGCGGTTAAAAAGCTCGTCGTTGGATCTCGGAGGCGGCACGGTCGGGCGCGATACGCT--GCGGGTCCGACC--------------------------GGCGAT-TTCT--C--G----------------------------GCC--GT-C-
------C-C-T--AA-CC-G-GCGCGCGGG----GCGAGGTGG--ACTTCACGG--TC--CTGCCTCGTTTTTCTCCTT-C-GCGG-GAGAGA----ACCCGCCG—CATTTTCTGTCTACCTTGAAAAAATTAGAGTGCTAAAGGCAGGCTAAAACCC
-CGA------GCCTGAATATTTCGTGCATGGAATAAGCGGGGATGCTCTTTTCCGGAGGGTGGTCCGGCCGACTTGG-----CGACGTCGCTTTCGCTGTCGGTTTCGCCGGAGA—GCTAGGATTCGAGAGGAACGGTGGGGGGCGAGCGTACCGGGC
GGGGAGAGGTGAAATTTTGTGATCCGCTCGGGACGACTCGAA-GCGAAAGCGCTGGCCCCGGACGTCTTCCTTAATCCAGAACGAAG-GTGGGTGGAGCAAAGACGATCAGATACCGTCGTAGTACCGACGGTAAACGCTGCCGACCGGGCCGCGGAC
GA-GGCTCGAGAACAGCGGTTATGGACCCGTCCGTGGCGGCTT-CGGGCAACCTGAGAGTTTACGGGTTCCGGGGGTAGTATGGTCGCAAGGCTGAAACTTGAAGGAATTGACGGAAGGGCACCACAAGGAGTGGAGCCTGCGGCTTAATTTGACTCA
ACACGGGGCAACTTACCCGGTCCGGACATCGTCAGGATTGACAGACGCTCGTGCAGTTCTTTCTCGATTCGATGGGTGGTGGTGCATGGCCGTTCGTGGCTG

Alinda biplicata

AGGGCTTTCTTTATTAATTCGATTAGAGTTAGGAACAAGAGGG------GTATTACTTGATGATCATTTTTTTAATGTAATTGTCACAGCGCACGCCTTCGTTATAATTTTTTTTATAGTTATACCGGTAATAATTGGGGGGTTTGGCAATTGAATAG
TACCCCTTCTCATTGCTGCACCTGATATAAGATTTCCTCGAATAAATAATATAAGCTTTTGGTTATTACCTCCGGCTTTCCTTTTATTAATTAGTAGTAGAATAGTGGAAGGGGGGGCTGGTACAGGGTGAACCGTATATCCGCCCTTAAGAGGTATT
GCCGCCCATGGTGGCGCCTCTGTTGATTTGGCAATTTTTTCTTTGCACCTTGCAGGAATATCCTCTATTTTAGGGGCTATTAATTTTATCACTACGGTATTTAATATACGTAGAAGAGGGATAAATATAGAACGAGTAAGTCTTTTTGTATGATCTAT
TTTAGTGACGGTTTTTTTATTACTACTTAGCTTACCAGTATTAGCTGGTGCGATTACTATGTTATTAACTGATCGAAATTTTAATACTTGTTTTTTTGATCCCGCAGTTCACA-CT-------GTCTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GC----------------------GG-AA--AGA-CGCTTT
TATTAGTTC-AA--AAC-CAAT-----------CGG-CGTCGCCCT-----TCG-C-------GGGGTG-GCG-TCGTCC------AAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTCACGT-----GCC-GGCGACGCA
TCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCCC
GGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAA
AGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGCGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGT-CACT--G--C----------------------------CCT-G--T--
-----T-TCC--TG--CCC-T-ACCCGTC----TGCCGGCT-----CTCTCCCGCGG--GTGC----TCTTCACTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CGC
-AGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGCG
GCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCGC
AGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC--AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Aplysia californica

AGGTTTAAGTCTTTTAATTCGATTTGAGCTTGGTACCGCTGGAGCT---TTCCTAGGGGATGATCACTTCTATAATGTTATTGTAACTGCTCACGCTTTTGTAATGATTTTTTTTATAGTTATGCCTATTATAATTGGTGGATTCGGGAATTGAATGG
TACCTTTATTAATTGGTGCTCCTGATATAAGATTTCCTCGTATAAATAACATAAGTTTCTGACTTCTTCCCCCTTCCTTTTTACTTCTTTTAGTCTCTAGACTAATAGAAGGTGGAGCTGGAACAGGATGAACAGTTTATCCTCCTCTATCAGGTCCT
ATTGCACACGGAGGAACTTCTGTTGATTTAGCTATTTTTTCTTTACATTTAGCAGGGATGTCTTCAATTTTAGGTGCTATTAATTTTATTACTACTATTTTTAATATACGATCGCCAGGAATTACTTTTGAGCGATTAAGCTTATTTGTTTGATCTGT
TCTAGTAACAGCTTTCTTACTTCTTCTTTCTTTACCAGTATTAGCTGGTGCTATTACGATGCTTTTAACGGATCGTAATTTTAATACGAGCTTCTTCGACCCTGCGGTTCACA-CT-------GTTTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTCG----CA-AG--CTC-CGACC--C-TC-GT----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGTTTCGCCGC---TTC-G-----GCGGCGG--GCG-GCGTCCC-----AAT-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTTTTGT-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGATATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGTGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGTTCACT--G--C----------------------------CC--G--T-
-----ATCTCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGCGG--GTGC----TCTTCGCTGAG-C-GTCCCGGGTG------GCCG--G-------TGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-GGTA-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAAC-GGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCG

```
CAGGAGT-T--------GCTT-TGATGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTTTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Arion vulgaris
AGGTCTATCTCTTTTAATTCGCTTAGAGTTAGGAACAACTGGT------ATTTTAACTGATGATCAATTTTTTAATGTTATTATCACTGCACACGCTTTTGTAATAATTTTTTTTATAGTTATACCACTAATAATTGGAGGTTTCGGTAATTGAATAG
TTCCTCTACTTATTGGAGCTCCCGATATGAGATTTCCACGTATAAATAATATAAGATTTTGATTATTGCCCCCATCTTTACTTTTACTTATTTGCTCCAGTATAGTAGAAGGAGGTGCGGGGACAGGGTGAACGGTTTACCCCCCCTTAAGAGGTTCC
TTAGCACATAGTGGAGCTTCAGTAGACTTAGCAATTTTTTCATTACACTTGGCTGGTATATCTTCTATTTTAGGTGCAATTAATTTTATTTCTACTATTTTTAATATACGCCCTAAAGCCTTATCATTGGAGCGTATAAGTTTGTTTGTGTGATCCAT
CTTGATTACGGTTTTTTTATTACTACTATCTCTGCCCGTATTGGCGGGGGCTATTACCATACTATTAACAGACCGTAATTTTAATACTAGATTTTTTGATCCTGCGGTTCACA-CT-------GTCTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCAATAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GT----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGC-CGTTACCCT-----TTG-C-------GGGGT--GCG-GCGTCC------CAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTCATGT-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGCGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGTTCACT--G--C-----------------------------CC--GT-T-
-----GTCTCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGTGG--GTGC----TCTTCACTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-AGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Biomphalaria glabrata
TGGATTATCATTATTGATTCGTTTGGAGCTTGGAACTACTCTT------GTTTTGATAGATGAACATTTTTATAATGTTGTTGTTACAGCTCATGCTTTTATTATAATTTTTTTTATAGTTATACCTATAATAATTGGTGGTTTTGGTAATTGAATGG
TTCCTTTATTAATTGGTGCTCCAGATATAAGTTTTCCTCGGATAAATAATATGTCTTTTTGATTACTTCCGCCTTCATTTATTTTATTGTTAGTATCTAGTATAGTTGAGGGTGGAGTAGGTACAGGTTGAACTGTATATCCTCCCTTAAGTGGTCCT
ATTGCTCATGGNGGTGCTTCTGTGGATTTAGCTATTTTTTCTTTACATTTAGCAGGAATAAGTTCAATTTTAGGTGCTATTAATTTTATTACTACAATTTTTAATATGCGTGCTCCCGGTATTACAATGGAACGTTTATCATTATTTGTATGATCTGT
ATTAGTTACAGCATTTTTACTTTTATTATCTTTACCTGTTTTAGCAGGGGCTATTACAATGTTATTAACAGATCGAAATTTTAATACTAGTTTTTTTGATCCTGCAGTTCACA-CT-------GTCTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGATGTTTATTAGATGAAA------CTA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GT----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGC-CGTTGCCCT-----TCG-C-------GGGGC--GCG-GCGTCCC-----GAT-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CCTA-GT-----GCT-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGATATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCCAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCAATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGCGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGCTCACT--G--C-----------------------------CC--GT-T-
-----GTCTCC--TG--CCC-T-ACCCGTC----TGCCGGCT-----CTCTCCCGTGG--GTGC----TCTTCGCTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----C-
G-AGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAACTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACTATGCCACCTGGCCATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCGGCCTCCGGGAAACC—AAAGGTTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Cepaea hortensis
CGGGCTGTCTCTATTAATCCGTTTAGAACTTGGTACCGCTGGT------GTATTAACTGACGATCATTTTTATAATGTTGTTGTTACTGCTCATGCGTTTGTTATGATTTTTTTTATAGTTATACCAATTATAATTGGGGGGTTTGGAAATTGAATAG
TTCCTTTGCTAATTGGTGCTCCTGACATAAGCTTTCCTCGGATAAATAATATAAGTTTTTGGTTATTACCTCCAGCATTTTTATTGCTTATTAGTAGTAGTCTCGTAGAAGGAGGAGCTGGAACCGGTTGAACCGTTTATCCGCCGCTAAGAAGCTTA
ATTGGGCATAGGGGTGCATCAGTTGACTTAGCTATCTTTTCCTTGCACCTCGCAGGTGTGTCATCTATTTTAGGGGCTATTAATTTTATTACTACTATTTTTAACATGCGAGCACCTGGAATAACACTTGAACGTATGAGCCTTTTTGTTTGATCAAT
TCTAGTAACTGTTTTTTTACTTCTTCTTTCGCTTCCTGTTTTAGCAGGTGCTATTACAATATTGTTAACTGATCGAAATTTCAGCACTTCGTTTTTTGATCCTGCTGTTCACA-CT-------GTCTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GC----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGTTGCCCT-----TCG-C-------GGGGC--GCG-GCGTTCC-----CAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTTACGT-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTATCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGTGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGTTCACT--G--C-----------------------------CC--GT-T-
-----GT-TCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGTGG--GTGC----TCTTCGCTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-GGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGT-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAATTGGAGGA--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCGACTAGCGATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Anentome helena
AGGATTAAGTTTATTAATTCGTGCTGAGCTAGGACAACCTGGAGCT---TTACTTGGTGATGATCAACTTTATAACGTGATTGTTACAGCTCATGCTTTTGTTATAATTTTTTTTTTAGTTATACCTATAATAATTGGCGGTTTTGGAAATTGATTGA
```

```
TTCCATTGATATTAGGAGCACCAGATATGGCATTTCCTCGTTTAAATAATATAAGATTTTGATTATTACCACCTGCTTTACTTCTTTTATTATCTTCACCAGCTGTTGAGAGTGGTGTTGGAACTGGGTGAACTGTTTATCCACCTCTTGCAGGAAAT
ATAGCTCATGCTGGCGGTTCTGTAGATTTAGCTATTTTTTCATTACATCTTGCAGGAGTATCATCAATTTTAGGTGCGGTAAATTTTATTACAACTATTATTAATATACGATGACGTGGTATGCAATTTGAGCGACTTCCTTTATTTGTTTGATCTGT
AAAGATTACTGCTATTTTACTACTTTTATCATTACCTGTATTAGCTGGAGCTATTACTATACTTTTAACTGATCGAAATTTTAATACTTCTTTTTTTGACCCTGCAGTTCACA-CC-------CTCGT-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGATC------CGA-A-AT-TACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC--CC----AACAG--CTC-CGACCC-T-TA-C----------------G------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AGC-CAGT-----------CGGGT-------------TCT-----------------GC-CCGTTC------T-T-T-GGTG-A-CTCTGGATAACTTTGTGCCGATCGCATGGC----C-TC-GA-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCATATGACGATGGTACGTGATCTGCCTACCATGTTAGCAACGGGTAGCGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCCAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
TGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGAACTCTTTTGAGGCT-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATACTA
AAGTTGTTGCGATTAAAAAGCTCGTAGTTGGATCTC-----AGGCATGGGCGCACG-GTCC--GC--------CTCGC----------------------------GGCGGT-CACT--G--T----------------------------GC--GT-T-
------T-TGT--TT--CCCATCCTACGCT----TCCCGGTTG---TTCAGCTCA-TG--GTGC----TCTTCGCTGAG-C-GCTATGGGTG------GCCG--G-------AACGTTTACTTTGAAGAAATTAGAGTGTTCAAAGCAGGC------G
C-GTC------GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GCTGGTT----TTCGGAACACGAGGT--A-----ATGATTA-AGAGGGACAGACGGGGGCATCCGTATTGC
GGTGTTAGAGGTGAAATTCTTGGATCATCGCAAGACGAACAACT-GCGAAAGCATTTGCCAAGCATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATTCG
CTGGTGT-T--------GCTT-CATCGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTTTTCGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Cornu aspersum
AGGCTTGTCTTTACTCATTCGTTTTGAACTAGGTACGTCGGGT------GTATTAACTGATGACCATTTTTATAATGTTATTGTTACAGCGCACGCATTTGTTATAATTTTTTTCATAGTTATACCTATTATGATTGGTGGCTTTGGTAACTGAATAG
TACCTTTATTAATTGGCGCTCCGGATATAAGTTTTTCGCGCATAAACAACATAAGTTTTTGATTATTACCTCCGTCTTTTTTGTTACTTATCAGTAGCAGCTTAGTTGAAGGAGGGGCTGGAACTGGATGAACTGTCTATCCACCTTTGAGAGGTCTA
AGCGGCCATAGTGGTGCCTCCGTAGACTTAGCAATTTTCTCGCTCCATCTTGCTGGTATGTCGTCTATTTTAGGTGCGATTAATTTTATTACAACTATTTTCAACATACGTTCGCCTGGTGTGAGATTAGAACGCATAAGTTTGTTTGTTTGGTCTAT
TTTAGTGACTGTGTTTTTACTATTACTGTCCTTGCCTGTTCTCGCAGGCGCAATTACTATGCTATTAACTGATCGAAATTTTAACACTTCATTTTTTGATCCCGCCGTTCACA-CT-------GTCTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GC----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGTTGCCCT-----TCA-C-------GGGGC--GCG-GCGTTCC-----CAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTTATGT-----GCC-GGCGACGC
ACCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAAAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTATCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGTGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGTTCACT--G--C----------------------------CC--G--T-
-----GTCTCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGTGG--GTGC----TCTTCGCTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-GGCTA----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACA-GAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTACTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCGACTAGCGATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Deroceras reticulatum
GGGGCTTTCTCTTTTAATTCGATTAGAATTAGGAACAGCGGGA------GTTTTACTAGATAATCATTTTTTTAATGTAGTAGTAACTGCACATGCATTTGTTATAATTTTTTTTATGGTGATACCTATTATAATTGGAGGCTTTGGTAATTGGATGG
TACCCTTGTTAATTGGAGCTCCGGATATAAGTTTTCCTCGAATAAATAATATAAGGTTTTGATTACTTCCCCCTTCATTTCTTTTACTAATTTGTTCTAGTATAGTAGAGGGTGGGGCGGGAACTGGGTGAACAGTTTACCCCCCTTTAAGAGGGCCA
TTAGGTCATGCTGGGGCATCTGTGGATTTAGCTATTTTCTCATTACATTTAGCAGGGATGTCTTCTATTTTAGGTGCTATTAATTTCATTACTACAATTTTTAATATACGGTCACCTGGGATGAGAATAGAACGTTTAAGATTATTTGTTTGATCAAT
TTTAGTTACTGTGTTTTTACTTTTATTATCTTTACCTGTTTTAGCGGGGGCAATTACTATACTTTTAACTGATCGAAATTTTAATACAAGGTTTTTTGACCCTGCGGTTCACA-CT-------GTCCC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GT----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGTCGCCCT-----TCG-C-------GGGGTG-GCG-GCGTCC------CAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCACTGA----CTT--GC-----TCA-GGCGACAC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGTGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGT-CACT--G--C----------------------------CC--GT-T-
-----GT-TCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGCGG--GTGC----TCTTTGCTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-GGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCGACTAGCGATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Elysia crispata
TGGTTTAAGGCTCTTAATTCGATTTGAATTAGGGACTTCCGGTGCC---TTTTTAGGTGACGACCACTTTTATAATGTTATTGTTACAGCACATGCTTTTGTTATAATTTTTTTCATAGTTATACCTCTAATAATTGGAGGATTTGGAAATTGAATAG
TTCCTATTCTTATTGGGGCTCCAGATATAAGTTTTCCTCGTATGAACAATATGAGGTTTTGACTTTTACCTCCATCTTTTATTTTTCTTCTGTCTTCAAGCCTTGTGGAAGGAGGTGCTGGAACTGGATGGACAGTGTATCCACCCCTTAGGGGACCT
ATTGGCCACGGTGGTGCCTCAGTAGATTTAGCAATTTTCTCACTACATCTTGCTGGTATATCTTCTATCTTAGGTGCAGTTAATTTTATTACTACAATTTTTAATATACGATCTCCGGGGATAAGAATAGAACGACTAAGACTCTTTGTTTGATCAGT
TTTAGTTACGGCATTTTTACTATTGCTATCACTACCTGTACTTGCCGGAGCGATTACGATACTGCTCACCGATCGTAACTTTAATACTAGCTTTTTTGATCCTGCGGTTCACA-CT-------GTTCC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCATCG----CA-AG--CTC-CGACC--T-TC-G----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGG-TCG-------CTC-G---------CGG--CCG-GCGTCCT-----CAC-T-GGTG-A-CTCTGGATAACTTTTTGCTGATCGCATGGC----CTT--GC-----GCC-GGCGACGC
```

```
ATCTTTCAAATGTCTGCCCTATCAAATTTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGACCCATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGCGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGCTCACT--G--C-----------------------------CC--GT-T-
-----GTTTCC--TG--CCC-T-ACCCGTC----TGCCGGCT-----CTCTCCCG-CGCGGTGC----TCTTCGCTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
T-GGTC-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATGGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCTAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAGCTCGCGATCCG
GGGGAGT-T--------GATT-CGATGACTCCGCG--GGCAGCGTCCGGGAAACC—AAAGCTTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Helix pomatia
GGGTCTCTCTTTGTTAATTCGTTGAGAGTTAGGAACATCCGGC------GTATTAAGCGACGATCATTTCTATAATGTGATTGTAACTGCCCATGCATTTGTCATAATTTTTTTTATGGTAATACCTATTATAATTGGTGGATTTGGAAATTGGATAG
TCCCCTTACTTATTGGTGCACCTGATATAAGGTTTCCTCGTATAAATAACATAAGCTTTTGACTGCTGCCCCCGTCTTTTATATTATTAATTAGGAGTAGTTTAGTTGAAGGTGGGGCAGGGACGGGTTGGACTGTTTATCCACCTCTTAGCAGTTTA
ATTGGCCATAGTGGAGCATCTGTTGATCTCGCTATTTTTTCTTTACACCTTGCTGGAATGTCTTCTATTCTGGGTGCTATTAATTTTATTACTACTATTTTTAATATACGGTCTCCAGGTATAAGTTTAGAACGTATAAGGTTATTTGTGTGGTCTAT
TCTAGTAACTGTGTTTTTACTCTTATTATCCTTACCTGTTTTAGCTGGTGCTATTACTATATTGTTGACTGATCGAAATTTTAGTACTTCTTTTTTTGATCCTGCTGTTCACA-CT-------GTCTC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GC----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGTTGCCCT-----TCA-C-------GRGGC--GCG-GCGTTCC-----CAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTTAYGT-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTATCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCACG-AGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGYAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGTGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGTTCACT--G--C-----------------------------CC--G--T-
-----GT-TCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGTGG--GTGC----TCTTCGCTGA--CGGTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-GGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACA-GAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTACTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTGCT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAAAGTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCGACTAGCGATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC--AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGC-GCTTAATTTGACT
CAACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTWG

Limax maximus
AGGTTTATCTTTATTAATTCGGTTAGAACTAGGGACAGCTGGT------GTTTTAATAGATGATCATTTTTTCAACGTTATTGTAACGGCTCATGCTTTTGTAATAATTTTTTTTATAGTTATACCTATTATGATTGGGGGTTTTGGTAATTGAATAG
TTCCTCTTTTAATTGGGGCTCCTGATATAAGGTTTCCTCGTATAAATAATATAAGGTTTTGGTTACTTCCTCCATCTTTTATTTTGCTTATTTGTTCTAGTATAGTGGAAGGGGGTGCAGGTACAGGGTGAACTGTATATCCTCCTTTGAGAGGTCCT
TTAGGGCATGGAGGAGCTTCTGTTGATTTAGCTATTTTTTCACTTCATTTAGCTGGTATATCTTCTATTTTAGGTGCAATTAATTTTATTACTACAATTTTTAATATACGAACGTCTGGGATAACCATAGAACGTGTGAGATTATTTGTATGGTCAAT
TTTAGTCACGGTATTTTTACTTCTTCTTTCGCTTCCTGTATTGGCGGGGGCAATTACTATGCTTTTAACAGACCGAAATTTTAATACTAGCTTTTTTGATCCAGCTGTTCACA-CT-------GTCCC-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTAC----CA-AG--CTC-CGACC--C-TC-GT----------------------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGT-CGTTACCCT-----TCG-C-------GGGGT--GCG-GCGTCCC-----CAC-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----CTCACGT-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGACATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGCGCAGGCGGGCG-GTCC-GGC---------TCGC----------------------------GCCGGTTCACT--G--C-----------------------------CC--GT-T-
-----GT-TCC--TG--CCC-T-ACC-GTC----TGCCGGCT-----CTCTCCCGCGG--GTGC----TCTTTGCTGAG-C-GTCCCGGGTG------GCCG--G-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-GGCT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
GGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAGCTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCGACTAGCGATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTGTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Littorina littorea
TGCCTTAAGTCTACTTATTCGGGCTGAATTAGGTCAACCTGGCGCT---CTCCTGGGAGATGACCAGCTGTACAACGTTATCGTTACAGCCCACGCCTTTGTAATAATTTTCTTTCTTGTTATGCCTATAATAATTGGTGGGTTTGGAAATTGACTTG
TCCCCTTAATATTAGGAGCACCCGATATAGCGTTCCCTCGCTTAAATAACATAAGCTTTTGATTACTCCCCCCCGCCTTGTTGTTGTTACTATCTTCTGCTGCGGTAGAAAGTGGTGCAGGGACGGGTTGAACTGTATATCCTCCTTTATCCGGAAAT
TTAGCCCATGCCGGAGGCTCTGTGGACTTAGCCATTTTCTCTCTTCATTTGGCCGGTGTCTCATCTATTTTAGGGGCCGTAAATTTTATTACAACTATTATTAATATACGATGACGAGGGATGCAATTCGAGCGATTACCTCTTTTTGTTTGATCTGT
AAAAATTACAGCCATTCTTTTACTTTTATCCCTTCCAGTTTTAGCGGGAGCCATTACAATATTGTTAACTGATCGAAATTTTAACACTGCCTTCTTCGATCCTGCTGTTCACA-CC-------CTCGT-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGATC------CCA-A-TC-TACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC--CA----ACCAG--CTC-CGACCCGT-TA-C----------------TG------GG-AA--AGA-GCGCTT
TTATCAGTTC-AA--AAC-CAGT-----------CGGGCCC-G---------TTA-A----------AC---GGTCCGTCC------C-T-T-GGTG-A-CTCTGGATAACTTTGTGCCGATCGCATGGC----C-TC-GA-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGACGATGGTACGTGATCTGCCTACCATGTTGGCAACGGGTAGCGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
TGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGAACTCTTTTGAGGCT-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATACTA
AAGTTGTTGCGATTAAAAAGCTCGTAGTTGGATCTC-----AGGCATGGGCGCACG-GTCC--GC--------CTCGC----------------------------GGCGGT-CACT--G--T----------------------------GT--G--T-
------A-TGT--TT--CCCATCCTACGCT----TCCCGGTTG----TTAGCCCA-TG--GTGC----TCTTCATTGAG-C-GTTTTGGGTG------GCCG--G-------AACGTTTACTTTGAAGAAATTAGAGTGTTCAAAGCAGGC------A
C-GTC------GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GCTGGTT----TTCGGAACACGAGGT--A-----ATGATTA-AGAGGGACAGACGGGGGCATCCGTATTGC
```

```
GGTGTTAGAGGTGAAATTCTTGGATCATCGCAAGACGAACTACT-GCGAAAGCATTTGCCAAGAATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCG
CTGGTGT-T--------GCTT-CATCGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTCTATGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGGAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Lottia gigantea
CACTTTGAGACTTTTAATTCGAGTTCATTTAATACACTCTGGGAGTGGGCTACTGGTTAGAGGTACGCTGTATAACGCAGTTGTTACGGCCCATGCTTTTTTAATAATTTTCTTTATGGTAATGCCGGTCTTAATTGGAGGGTTCGGGAATTGGTTAG
TGCCATTAATGCTCCCGGCTCCTGATTTAGGTTTGCCCCGGATAAACAACTTAAGGTTTTGGTTAGTTCCTCATTCGTTGGTATTTTTATTAGTTTCCACGTTAACCGATACTGCAGTGGGGACAGGCTGAACCTTGTATCCTCCGTTGTCTTCTCTT
GAGGGTCACCATTCTCCTTGTGTGGACGAGGCAATTTTTTCTTTGCATATATCAGGGATTGCTTCAATTTTTGCTAGAATTAACTTTATCACTACGATTAAACACGCTCGGGGGCCTCATAAGAAGATGTCTTCTATACCACTTTTTGTAGCTGCTAT
TGGTGTAACGAGGGTTCTACTTTTGTTGTCTGTTCCTGTTTTGGCTGGGGGGCTTACTATACTTATTACTGATCGTAATATTAACACCGCCTTTTTTGACCCAGAGGTTCAGG-CTTGTTCCTTTCGG-GG----AGCGAGCCGAAACTGCGAACGGC
TCATTAG-TCAGATAAGGTTCCTTGGCTAAAAGCGGGTCGG-T-TT-TA-ATGGATAACTGTGGTAATTCTAGAGCTAATACATGC--AACGCACCGTGGTCCCCGTAGTC-T-TCTC----------------GTCTCGGGGAAA—CGGCATTTATT
CTCTGGACC-AG--ATCGCCCT-----AGCCTGCGA-AGC-----------TACCGTCAAAAGTAGCGACGGG-GCGAAGC---GACAA-GTGTTG-AAATCCGAATAACTGTG--CCGATCGCGGGGT----C-TT-TCCGGGCCCCCGACGACTTT
GCCATGAAGTGTCTGTCCCATCAATTGGCGATGGTCGGCGACCTGCCTACCACGGTGTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGCGAAACGGCTACCACCTCCAAGGAGGGCAGCAGGCGCGCAACTTACCCAATCCC
GACACGGGGAGGTAGTGACGAAAAATAACGTGGCGGGGCCCTTTCGTGGTCGCCGCGGGCGGAATGAGCGCAATTTAAAAGAGTGTGCGAGGAGCTATTGGAGGGCAAGCCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCCTATATCGG
AGTTGCTGCGGTTAAAAAGCTCGTCGTTGGATCTCGGAGGCGGCACGGTCGGGCGCGATACGCTGC----GGGTCCG----------------------ACC-GGCGAT-TTCT--C--G----------------------------GCC—GT-CC--
----C-T-A---A--C-C-A-GCG--------CGCGGGGCGAGGTGGACTTCG-CG--GTCC---TGCCTCGTTTTTCT-CCTTCACGGGAGAGAACCCG--CCGCATTTTCTGTCTACCTTGAAAAAATTAGAGTGCTAAAGGCAGGC------TAA
AACC-CCGAGCCTGAATATTTCGTGCATGGAATAAGCGGGGATGCTCTTTTCCGGAGGGCGGTCCGACCGA-CTGGCGACGTCGCTTTCGCT----GTCGGTTT-CGCCGG—AGAGCTAGGATTCGAGAGGAACGGTGGGGGGCGAGCGTACCGGGCG
GGGAGAGGTGAAATTTTGTGATCCGCTCGGGACG-ACTCGAAGCGAAAGCGCTGGCCCCGGACGTCTTCCTTAATCCAGAACGAAG-GTGGGTGGAGCAAAGACGATCAGATACCGTCGTAGTACCGACGGTAAACGCTGCCGACCGGGCCGCGGACG
AGGC-TCGAGAACAGCGGTTATGGACCCGTCCGTGGCGGC-TTCGGGCAACCTGAGAGTTTACGGGTTCCGGGGGTAGTATGGTCGCAAGGCTGAAACTTGAAGGAATTGACGGAAGGGCACCACAAGGAGTGGAGCCTGCGGCTTAATTTGACTCAA
CACGGGGCAACTTACCCGGTCCGGACATCGTCAGGATTGACAGACGCTCGTGCAGTTCTTTCTCGATTCGATGGGTGGTGGTGCATGGCCGTTCGTGGCTG

Marisa cornuarietis
TGCTTTAAGAATGCTTATTCGAGCTGAACTTGGGCAACCGGGAGCC---TTGTTAGGTGATGATCAGCTTTACAATGTCATTGTTACAGCTCATGCTTTTGTAATGATTTTCTTTTTAGTTATGCCTATAATAATTGGTGGATTTGGAAATTGACTTG
TGCCATTAATATTAGGTGCTCCTGACATGGCTTTTCCTCGTTTAAATAATATAAGTTTTTGGTTGCTTCCGCCTTCTTTATTACTTCTTTTGTCTTCAGCTGCTGTAGAAAGTGGAGTTGGGACAGGTTGAACAGTTTATCCTCCTTTAGCTGGAAAT
CTAGCACATGCTGGTGGTTCTGTAGATTTAGCAATTTTTTCTTTACATTTAGCTGGTGCTTCTTCTATTTTAGGGGCTGTTAATTTTATCACAACAGTAATTAATATACGTTGACGAGGCATGCAATTTGAGCGATTACCTTTGTTTGTGTGATCTGT
TAAAATTACGGCTATTTTATTACTTTTATCTTTACCTGTTTTAGCCGGTGCTATTACTATATTACTTACGGACCGAAATTTTAATACTGCTTTTTTTGATCCTGCTGTTCACA-CC-------CTCGT-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGATC------CAT-T-TC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC--AA----CCAAG--CTC-CGACCC-GGTG-TC--------AA-AGCCG------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAGT-----------CGG-GGT-----------CTC----------------GGCCCCGTCCCT----T-T-T-GGTG-A-CTCTGGATAACTTTGTGCCGATCGCATGGC----C-TC-GA-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGATAGGCCTACCATGTTGACTACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CAGCTCGGGGAGGTAGTGACGAAAAATAACAATACGGAACTCTTTTGAGGCT-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCGATTAAAAAGCTCGTAGTTGGATCTC-----AGGCATGGGCGTGCG-GTCC-GTC---------TCGC----------------------------GACGGT-CACT--GTGCGAAGCCGGGGGCGGGCGGGGCCCTTCGCGGG-GTTT-
CGTTCGTCTTTTGGCGTCCCATCCTCCGTTTATCTCC-GGTCT---CTCCGTCCA-TG--GTGC----TCTTCGCTGAG-C-GTTATGGGCG------GCCG--G--------AACGTTTACTTTGAAGAAATTAGAGTGTTCAAAGCAGGC------T
C-GTTA-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GCTGGTT----TTCGGAACACGAGGT--A-----ATGATTA-AGAGGGACAGACGGGGGCATCCGTATTGC
GGTGTTAGAGGTGAAATTCTTGGATCATCGCAAGACGAACAACT-GCGAAAGCATTTGCCAAGCATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCG
CAGGTGT-T--------GCTT-CATCGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTTTTCGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Neritina pulligera
TGCTCTTAGTTTGCTGATTCGAGCAGAGCTTGGGCAACCTGGTGCT---TTATTGGGGGATGATCAGCTTTATAATGTAATTGTGACTGCTCATGCATTTGTAATAATTTTCTTTTTGGTTATGCCAATGATAATTGGTGGGTTTGGTAACTGATTGG
TTCCTTTAATGTTAGGGGCTCCTGATATGGCTTTTCCTCGACTAAATAACATGAGTTTTTGACTTCTTCCTCCTTCATTGACTTTGTTGTTAGCCTCTTCTGCCGTTGAGAGTGGAGTAGGTACTGGTTGAACTGTTTATCCTCCACTTTCTGGGAAT
TTAGCTCATGCTGGTGGATCTGTTGATTTAGCTATTTTTTCTTTACACTTGGCTGGTGTTTCATCAATTCTAGGTGCGGTAAACTTTATTACTACTATTATTAATATGCGATGACAGGGAATGCAATTTGAGCGACTTCCTTTGTTCGTATGATCTGT
TAAAATTACCGCAATTTTGTTGTTGCTTTCTCTTCCTGTGCTTGCAGGCGCAATTACTATGTTGCTAACTGATCGAAATTTTAATACTTCGTTTTTTGATCCAGCTGTACAAA-CC-------TTCAC-AT----GGTGA-----AACCGCGAATGGC
TCATTAGATCAGTTATGGTTCCTTAGATCGTA------CAA-C-TC-CACTCGGATAACTGTGGCAATTCTAGAGCTAATACGTGC--AA----GAAAG--CTC-TGACC--T-CG-C----------------------GG-GA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT-----------CGGGCC------------GCA-A---------------GGCTCGTCC------GTT-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCACGGC----C-TC-GA-----GCC-GGCGACGT
ATCTTTCAAATGTCTGCCCTATCAACTTTAGATGGTACGTGATATGCCTACCATGGTTATAACGGGTAGCGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
TGGTACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-TCGTAATTGGAATGAGTACACTTCAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----GGGTTTGGGCGGGCG-GTCC-GCC---------TCGC----------------------------GGCGGT-TACT--G--C----------------------------AC--GA-C-
------C-C-G---A--C-C-T-ACC--------TCCCGGTTT-------TCCCT-TG--GTGC----TCTTGACTGAG-T-GCTTCGGGTG------GCCG--G-------AACGTTTACTTTGAAAAAATTAGAGTGCTCAAAGCAGGC------G
C-T-TC-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAACTCGAGGT--A-----ATGATTA-AGAGGGACAGACGGGGGCATTCGTATTAC
GGTGTTAGAGGTGAAATTCTTGGATCGCCGTAAGACGAACTACT-GCGAAAGCATTTGCCAAGCATGTTTTCATTAATCAAGAACGAAA-GTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCG
CTGGAGT-T--------GCTT-CTATGACTCAGCG--GGCAGCTTCCGGGAAACC—AAAGTTTTTGGGTTCTGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG
```

Physella acuta

AGGTTTAAGCTTGTTAATTCGTTTGGAATTAGGAACATCTCTT------GTACTGTTGGATGAACATTTTTATAATGTAATTGTTACAGCACATGCTTTTGTAATGATTTTTTTTATAGTTATACCTATAATAATTGGAGGGTTTGGGAATTGAATAG
TACCTATATTAATTGGTGCTCCCGATATAAGATTTCCTCGAATAAATAATATAAGATTTTGACTTTTACCGCCTTCATTTATCTTATTATTATGTAGGTCTATAGTTGAGGGTGGAGTTGGAACTGGGTGAACTGTTTACCCCCCTCTATCAGGACCT
GTAGCTCACTCTGGTTCATCAGTAGATCTTGCTATTTTCTCATTACACTTAGCTGGGTTATCATCTATTCTAGGTGCTATTAATTTTATTACTACCATTTTTAATATACGTTCACCTGGTATTACACTGGAACGAATAAGCTTATTTGTTTGATCGGT
GTTAATTACTGCATTTTTATTATTATTGTCATTGCCTGTTTTAGCAGGGGCTATTACTATACTATTAACTGATCGAAATTTTAATACTAGGTTCTTTGATCCAAGAGTTCACA-CT-------GTCCC-AT----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGACA------CGA-T-CC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCAATC----GA-AG--CTC-CGACC--T-TTATC----------GG-GA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAAT----------CGG-CGC-GGCC------TCG-C-----AAGGGGTT-GCG-TCGTCCC-----ATT-T-GGTG-A-CTCTGGATAACTTTGTGCTGATCGCATGGC----C-TC-GC-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGATATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCGAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCC-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGTATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTC-----AGGCGCAGGCGGGCG-GTCC-GGC---------TCGC----------GCCGGCTCACT--G--C----------CC--G--T-
-----ATCTCC--TG--CCC-T-ACC-GTC----TGTCGGCT-----CTCTCCCGCGG--GTGC----TCTTCACTGAG-C-GTCCCGGGTG------GCCG--A-------CGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGC-----CT
C-G-CT-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GTTGGTT----TTCGGAATTGGAGGT--A-----ATGATTA-ACAGGGACAAACGGGGGCATTCGTATTGC
TGCGTTAGAGGTGAAATTCTTGGATCGCAGCAAGACGAACTACT-GCGAAAGCATTTGTCAAGAATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACTATGCCACCTGGCCATCCG
CAGGAGT-T--------GCTT-CGATGACTCTGCG--GGCGGCCTCCGGGAAACC—AAAGGTTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Pomacea bridgesii

AGCTTTAAGATTGCTTATTCGTGCTGAACTAGGGCAACCTGGTGCT---TTATTAGGAGATGACCAGCTTTATAATGTAATTGTGACGGCTCATGCTTTTGTAATAATTTTCTTTCTAGTTATGCCAATAATGATTGGAGGATTTGGTAACTGATTGG
TTCCATTGATACTAGGGGCCCCAGATATGGCTTTTCCACGGCTTAATAACATAAGATTTTGGTTATTGCCTCCTTCTTTATTACTATTATTGTCTTCGGCTGCTGTCGAAAGTGGTGTTGGTACTGGTTGAACAGTTTACCCCCCTTTAGCTGGAAAT
TTAGCTCATGCTGGTGGTTCTGTAGACTTAGCAATTTTTTCTTTACATTTAGCTGGGGCCTCCTCTATCCTAGGTGCGGTAAATTTTATTACAACGGTAATTAATATACGGTGACGAGGTATGCAATTTGAGCGTCTTCCTCTATTCGTGTGGTCAGT
TAAAATTACTGCTATTTTATTATTATTATCATTGCCAGTTCTTGCTGGGGCTATTACTATACTATTAACTGACCGAAATTTTAATACGGCTTTTTTTGACCCGGCTGTTCACA-CC-------CTCGT-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGATC------CAA-A-TC-TACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC--AA----ACCAG--CTC-CGACCC-GGTG-TC--------AC-AGCCG------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAGT----------CGG-GGT----------CTC---------------GGCCCCGTCCCT----T-T-T-GGTG-A-CTCTGGATAACTTTGTGCCGATCGCATGGC----C-TC-GA-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGATAGGCCTACCATGTTGACTACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
CAGCTCGGGGAGGTAGTGACGAAAAATAACAATACGGAACTCTTTTGAGGCT-CCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTA
AAGTTGTTGCGATTAAAAAGCTCGTAGTTGGATCTC-----AGGCATGGGCGTGCG-GTCC-GTC---------TCGCGACGGTCACTGTGCGAAGCCGGGGCGGGCGGTGCCCTTCG--C----------GGGGTGT-CGT-T-
CGCTTCTTTTGGCGT--CCCATCCTCCGTTTATCTCC-GGTCT---CTCCGTCCA-TG--GTGC----TCTTCGCTGAG-C-GTTATGGGCG------GCCG--G-------AACGTTTACTTTGAAGAAATTAGAGTGTTCAAAGCAGGC------T
C-GTCA-----GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GCTGGTT----TTCGGAACACGAGGT--A-----ATGATTA-AGAGGGACAGACGGGGGCATCCGTATTGC
GGTGTTAGAGGTGAAATTCTTGGATCATCGCAAGACGAACAACT-GCGAAAGCATTTGCCAAGCATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATCCG
CAGGTGT-T--------GCTT-CATCGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTTTTCGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Pomatias elegans

TGCTTTAAGGCTTTTAATTCGGGCTGAATTGGGGCAGCCTGGGAGT---TTGCTAGGTGATGATCAACTTTATAATGTTATTGTTACTGCTCATGCATTTGTTATAATTTTTTTTTTGGTTATGCCAATAATGATTGGTGGTTTTGGCAATTGGTTAG
TTCCTTTAATGTTGGGGGCTCCTGATATGGCTTTTCCTCGGTTAAATAATATAAGATTTTGATTACTTCCTCCTGCGTTATTTTTATTGTTAATATCTGCGGCTGTAAAGAGGGGAGCTGGTACTGGGTGAACAGTATATCCCCCTTTGGCTGGTAAT
TTAGCACATTCGGGTGGTTCTGTTGATCTTGCTATTTTTTCGTTACATTTGGCTGGGGCATCTTCTATTTTAGGTGCTGTGAATTTCATTACTACTATTGTTAATATACAATGGTGTGGTGTTCAGTTGGAGCGGTTATCTTTGTTTGTTTGGTCGGT
TAAAATTACTGCTATTTTGCTTTTACTTTCTTTACCAGTTTTGGCTGGGGCAGTTACTATGCTTTTAACAGATCGGAATTTTAATACTACTTTTTTTGATCCTGCTGTTCACA-CC-------CTCGT-AC----GGTGA-----AACCGCGAATGGC
TCATTAAATCAGTCGAGGTTCCTTAGATGATC------CAA-A-TC-TACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC--CA----ACCAG--CTC-CGACCCTT-TC-G----------GG------GG-AA--AGA-GCGCTT
TTATTAGTTC-AA--AAC-CAGT----------CGGGTC----------GCA-A----------GAC-TCGTCC------T-T-T-GGTG-A-CTCTGGATAACTTTGTGCCGATCGCATGGC----C-TC-GA-----GCC-GGCGACGC
ATCTTTCAAATGTCTGCCCTATCAAATGACGATGGTACGTGATCTGCCTACCATGTTAGCAACGGGTAGCGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCC
TGGCACGGGGAGGTAGTGACGAAAAATAACAATACGGAACTCTTTTGAGGCT-CCGTAATTGGAATGAGTACACTTTAAAGCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCA-GCAGCCGCGGTAATTCCAGCTCCAATAGCGTATACTA
AAGTTGTTGCGATTAAAAAGCTCGTAGTTGGATCTC-----AGGCATGGGCGCACG-GTCC--GC--------CTCGC----------GGCGGT-CACT--G--T----------GT--GG-T-
------TTTGT--TT--CCCATCCTACGCT----TCCCGGTTG---TTCAGCCCA-TG--GTGC----TCTTGACTGAG-C-GTTTTGGGTG------GCCG--G-------AACGTTTACTTTGAAGAAATTAGAGTGTTCAAAGCAGGC------A
C-GTC------GCCTGAATAAT-GGTGCATGGAATAATGGAATAGGA-------CC----TCGGTTCT----A-TTTT---------GCTGGTT----TTCGGAACACGAGGT--A-----ATGATTA-AGAGGGACAGACGGGGGCATCCGTATTGC
GGTGTTAGAGGTGAAATTCTTGGATCATCGCAAGACGAACAACT-GCGAAAGCATTTGCCAAGCATGTTTTCATTAGTCAAGAACGAAA-GTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAACTAGCGATTCG
CTGGTGT-T--------GCTT-CATCGACTCTGCG--GGCAGCTTCCGGGAAACC—AAAGTTTTCGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTC
AACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATTGACAGA---TTGATA-GCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTG

Alignment S2:
MUSCLE alignment (manually edited) of gastropod CdMT protein sequences applied for the mirrored Maximum Likelihood tree represented in **Figure 6**.

```
>Lottia gigantea MT1
-SSEK----------------------------------------PSCCIAEYECCKTKLCCDTGPADCCKPGNKPDCCAPGKLQCKCP-G-
TCACGVGCTGVDNCKCGAGCS-CFN

>Lottia gigantea MT2
-SSEK----------------------------------------ASCCIAEYECCKTKSCCDTGPADCCKPGNKPDCCAPGKLQCKCS-G-
TCACGVGCTGVDNCKCGAGCS-CFN

>Nacella polaris MT
-SSEK----------------------------------------AACCIAEYECCKTKSCCKDGPADCCKPGNTTDCCKGKV-ACKCA-G-
SCACGAGCTGQTPCKCGAGCS-CNS

>Patella vulgata MT1
-SSQK----------------------------------------ASCCLAELECCKTKACCAKGPANCCSPGNDPNCCKS-N-ICKCN-G-
NCACGVGCTGIENCECGTGCS-CK-

>Patella vulgata MT2
-SSEK----------------------------------------AACCLAEHECCKTKSCCANGPADCCKPGKTVDCCKSQN-TCKCG-E-
SCACGAGCSGVDNCKCGSGCS-CK-

>Littorina littea MT
-SS---VF-GAGCTDVCKQTPCGCATS--GCNCTDDCKCQSC-KYGAGCT---DTCKQTPCGC---GSGCNCKEDCRCQSCST-ACKCAAG-
SCKCGKGCTGPDSCKCDRSCS-CK-

>Pomatias elegans MT1
STSGANVIYGAGCTGTCKQSPCGCKNSAAGCRCKDDCQCPACAKYGAGCT---GTCKQSPCGCKNSAAGCGCKDDCRCPACAK-SCKC--G-
TCNCGKGCTGPSNCKCDDGCS-CK-

>Pomatias elegans MT2
SSSGANAT-------------------------------------GAGCT---ETCKESPCGCKNSAAGCKCKDDCQCTTCAK-SCKCA-G-
TCNCGKGCTGPNSCKCDGGCP-CK-

>Helix pomatia CdMT
SGKGK----------------------------------------GEKCT---SACRSEPCQC---GSKCQCGEGCTCAAC-K-TCNCTSD-
GCKCGKECTGPDSCKCGSSCS-CK-

>Cornu aspersum CdMT
SGKGK----------------------------------------GEKCT---AACRNEPCQC---GSKCQCGEGCTCAAC-K-TCNCTSD-
GCKCGKACTGPDSCTCGSSCG-CK-

>Arianta arbustorum CdMT
SGKGK----------------------------------------GDLCT---AACKNEPCQC---GSKCQCGEGCACASC-K-TCNCTSD-
GCKCGEKCTGAASCKCGSSCS-CK-

>Cepaea hortensis CdMT1
SGKGK----------------------------------------GEKCT---AACRNEPCQC---GSKCQCGEGCACAAC-K-TCNCTSD-
GCKCGKECTGPDSCKCGSSCS-CK-

>Cepaea hortensis CdMT2
SGKGK----------------------------------------GEKCT---AACRNEPCQC---GSKCQCGEGCACAAC-K-TCNCTSD-
GCKCGKECTGPDSCKCGSLCS-CK-

>Cepaea nemoralis CdMT
SGKGK----------------------------------------GEKCT---AACRNEPCQC---GSKCQCGEGCACAAC-K-TCNCTSD-
GCKCGKECTGPDSCKCGSSCS-CK-

>Cochlicella acuta CdMT
SGKGK----------------------------------------AESCT---AQCQSNPCQC---GDKCQCGEGCACTSC-K-TCKCTSD-
GCKCGKECTGPASCKCGSSCS-CK-

>Nesiohelix samarangae CdMT
SGK-------------------------------------------GELCT---SACKSNPCQC---GDKCQCGEGCTCSAC-K-SCHCTND-
GCNCGKECTGPTSCKCDTSCS-CK-

>Lehmannia nyctelia CdMT
SGK-------------------------------------------GAKCT---GACKSEPCQC---GQNCQCGDDCSCSQC-K-TCKCSAGS
TCQCGHGCTGVESCKCGNSCS-CK-

>Limax maximus CdMT
SGK-------------------------------------------GAKCT---GACKSEPCQC---GQNCQCGDDCSCSQC-K-TCKCSAGS
TCQCGHGCTGVESCKCGSSCS-CK-
```

595 >*Deroceras reticulatum* CdMT1
596 SGK----------GEKCT---GDCKSEPCKC---GQNCQCGNDCTCSQC-K-TCKCSSS-
597 GCQCGHGCTGVESCKCGSSCT-CK-
598
599 >*Deroceras reticulatum* CdMT2
600 SGK----------GEKCT---GDCKSEPCKC---GQNCQCGNDCTCSQC-K-TCKCSSS-
601 GCQCGHGCTGVESCKCGSSCT-CK-
602
603 >*Arion vulgaris* AvMT1
604 SGK----------A-CT---GACKSEPCQC---GNNCQCGGDCDCSQC-K-TCKCTNE-
605 GCKCGQNCTGQATCSCEKSCS-CK-
606
607
608
609
610
611
612
613

Alignment S3:

MUSCLE alignment (manually edited) of MTs of Panpulmonata (Hygrophila + Stylommatophora + Sacoglossa) applied for the Bayesian Inference Tree represented in **Figure 7**.

```
>Tritia obsoleta MT1
--------EGTGCAGPCKQTPCGCE-SSGGCKCGGDCSCTECGR-GAGCTDACKETPCGC-ASSGGCKCTGNCSCPSCS----------------CKCGK------
-----------------------GACKCDKG-CEGPGSCKCGPG-C-----TCKKS-----------

>Littorina littorea CdMT
-----SSVFGAGCTDVCKQTPCGC--ATSGCNCTDDCKCQSC-KYGAGCTDTCKQTPCGC---GSGCNCKEDCRCQSCSTA-------------CKCAA------
-----------------------GSCKCGKG-CTGPDSCKCDRS-C-----SCK-------------

>Pomatias elegans MT1
STSGANVIYGAGCTGTCKQSPCGCKNSAAGCRCKDDCQCPACAKYGAGCTGTCKQSPCGCKNSAAGCGCKDDCRCPACAKS-------------CKCGT------
-------------------------CNCGKG-CTGPSNCKCDDG-C-----SCK-------------

>Pomatias elegans MT2
-------------------------------------SSSGANATGAGCTETCKESPCGCKNSAAGCKCKDDCQCTTCAKS-------------CKCA-------
-----------------------GTCNCGKG-CTGPNSCKCDGG-C-----PCK-------------

>Elysia crispata MT
----------------------------------------SGK-GLICTAPCTNDPCGC---GENCQCGHACSCHSCS----------------CSCGA------
------------------------DCKCNQATCHEKSNCKCEAS-C-----SCRAK-----------

>Elysia cornigera MT
----------------------------------------SGK-GASCPPTCRNDPCGC---GQDCQCGQDCTCSSCS----------------CLCGN------
------------------------ECKCTSASCSQGVRCRCDTL-C-----DCRPK-----------

>Helix pomatia CdMT
---------------------------------------SGKGK-GEKCTSACRSEPCQC---GSKCQCGEGCTCAACKT--------------CNCTS------
-----------------------DGCKCGKE-CTGPDSCKCGSS-C-----SCK-------------

>Cornu aspersum_CdMT
---------------------------------------SGKGK-GEKCTAACRNEPCQC---GSKCQCGEGCTCAACKT--------------CNCTS------
-----------------------DGCKCGKA-CTGPDSCTCGSS-C-----GCK-------------

>Arianta arbustorum CdMT
---------------------------------------SGKGK-GDLCTAACKNEPCQC---GSKCQCGEGCACASCKT--------------CNCTS------
-----------------------DGCKCGEK-CTGAASCKCGSS-C-----SCK-------------

>Cepaea hortensis CdMT1
---------------------------------------SGKGK-GEKCTAACRNEPCQC---GSKCQCGEGCACAACKT--------------CNCTS------
-----------------------DGCKCGKE-CTGPDSCKCGSS-C-----SCK-------------

>Cepaea hortensis CdMT2
---------------------------------------SGKGK-GEKCTAACRNEPCQC---GSKCQCGEGCACAACKT--------------CNCTS------
-----------------------DGCKCGKE-CTGPDSCKCGSL-C-----SCK-------------

>Cepaea nemoralis CdMT
---------------------------------------SGKGK-GEKCTAACRNEPCQC---GSKCQCGEGCACAACKT--------------CNCTS------
-----------------------DGCKCGKE-CTGPDSCKCGSS-C-----SCK-------------

>Cochlicella acuta CdMT
---------------------------------------SGKGK-AESCTAQCQSNPCQC---GDKCQCGEGCACTSCKT--------------CKCTS------
-----------------------DGCKCGKE-CTGPASCKCGSS-C-----SCK-------------

>Nesiohelix samarangae CdMT
-------------------------------------------SGK-GELCTSACKSNPCQC---GDKCQCGEGCTCSACKS--------------CHCTN------
-----------------------DGCNCGKE-CTGPTSCKCDTS-C-----SCK-------------

>Limax maximus CdMT
-----------------------------------------SGK-GAKCTGACKSEPCQC---GQNCQCGDDCSCSQCKT--------------CKCSA------
----------------------GSTCQCGHG-CTGVESCKCGSS-C-----SCK-------------

>Deroceras reticulatum CdMT1
-----------------------------------------SGK-GEKCTGDCKSEPCKC---GQNCQCGNDCTCSQCKT--------------CKCST------
----------------------GSGCQCGHG-CTGVESCKCGSS-C-----T---------------

>Deroceras reticulatum CdMT2
-----------------------------------------SGK-GEKCTGDCKSEPCKC---GQNCQCGNDCTCSQCKT--------------CKCSS------
----------------------SGCQCGHG-CTGVESCKCGSS-C-----TCK-------------
```

```
>Lehmannia nyctelia CdMT
----------------------------------------SGK-GAKCTGACKSEPCQC---GQNCQCGDDCSCSQCKT--------------CKCSA------
----------------------GSTCQCGHG-CTGVESCKCGNS-C-----SCK-------------

>Arion vulgaris AvMT1
----------------------------------------SGK-A--CTGACKSEPCQC---GNNCQCGGDCDCSQCKT--------------CKCTN------
----------------------EGCKCGQN-CTGQATCSCEKS-C-----SCK-------------

>Physella acuta MT1
----------------------------------------SGK-GPNCTEACTGEQCTC---GRQLQV-------CCKT--------------CKCED------
----------------------NACKCGEG-CTGPSTCKCESSDC-----ACK-------------

>Physella acuta MT2
----------------------------------------SGK-GPNCTEACTGEQCTC---GDSCKCGEGCNCPSCKT--------------CKCED------
----------------------NACKCGEG-CTGPSTCKCESSDC-----ACK-------------

>Physella carolinae MT
----------------------------------------SGK-GPNCTEACTGEQCTC---GDSCKCGEGCNCPSCKT--------------CKCED------
----------------------NACKCGEG-CTGPSTCKCESSDC-----ACK-------------

>Physella hendersoni MT
----------------------------------------SGK-GPNCTEACTGEQCTC---GDSCKCGEGCNCPSCKT--------------CKCED------
----------------------NACKCGEG-CTGPSTCKCESSDC-----ACK-------------

>Physella gyrina MT
----------------------------------------SGK-GPNCTEACTGEQCNC---GDSCKCGEGCNCPSCKT--------------CKCED------
----------------------NACKCGEG-CTGPSTCKCEST-C-----ACK-------------

>Galba cubensis MT
----------------------------------------SGR-GPNCTEACRGEQCNC---GDSCKCGEGCNCPSCKT--------------CKCSA------
----------------------EDCKCDAG-SKGTGNCKC--------------------------

>Galba truncatula MT
----------------------------------------SGK-GPNCTEACRGEQCNC---GDSCKCGEGCNCPSCKT--------------CKCAA------
----------------------EDCKCDAG-SKGTGNCKC--------------------------

>Lymnaea stagnalis MT
----------------------------------------SGK-GPNCTEACTGEQCNC---GDSCKCGEGCNCPSCKT--------------CKCSA------
----------------------EDCKCDAG-SKGTGNCKC--------------------------

>Biomphalaria glabrata MT
----------------------------------------SGK-GPNCTEACTGEQCNC---GDSCKCGEGCNCPSCKTTKGPNCTEACTGKQ
CSCGDSCQCGEGCTCSCCKKACTKECTDTECSCGDSCKCGEG-CK-CSSCKAGK--CTKSDEGCKTEGHCAKGKCCKS

>Helix pomatia CuMT
----------------------------------------SGR-GKNCGGACNSNPCSC---GNDCKCGAGCNCDRCSS--------------CHCSN------
----------------------DDCKCGSQ-CTGSGSCKCGSA-C-----GCK-------------

>Cornu aspersum CuMT
----------------------------------------SGR-GQNCGGACNSNPCNC---GNDCNCGTGCNCDQCSA--------------RHCSN------
----------------------DDCKCGSQ-CTRSGSCKCGNA-C-----GCK-------------

>Cepaea nemoralis CuMT
----------------------------------------SGR-G-NCGGACNSNPCSC---GDDCKCGAACNCDRCSS--------------CHCSN------
----------------------DDCKCGSQ-CTGSGSCKCGSA-C-----GCK-------------

>Theba pisana CuMT
----------------------------------------SGR-GKNCGGACNSNPCNC---ANNCRCGAGCNCDSCSS--------------CHCSN------
----------------------DDCKCGNQ-CTTSGSCKCGSA-C-----GCK-------------

>Alinda biplicata CuMT1a
----------------------------------------SGK-GANCTGACNSNPCQC---GDDCKCGVGCSCAECNT--------------CKCTN------
----------------------DGCKCGHG-CTGAGSCKCGNS-C-----GCK-------------

>Alinda biplicata CuMT1b
----------------------------------------SGK-GANCTGACNSNPCQS---GDDCKCGVGCSCAECNT--------------CKCTN------
----------------------DGCKCGHG-CTGAGSCKCGNS-C-----GCK-------------

>Alinda biplicata CuMT2
----------------------------------------SGK-GANCSGACNSNPCQC---GDDCKCGAACSCAECNT--------------CKCTN------
----------------------DSCKCGHD-CSGAGSCKCGNS-C-----GCK-------------

>Cochlicella acuta CuMT
----------------------------------------SGR-G-NCGGACKSNPCSC---GQVCKCGGACTCAQCNA--------------CRCSG------
----------------------DSCKCGDQ-CTASGSCQCGSG-C-----GCK-------------
```

766 >*Arion vulgaris* AvMT2
767 ----------SGR-G--CNGTCNSNPCQC---EDGCQCGDACSCAQCNT----------CKCTN------
768 ----------DGCKCGNE-CTATGSCKCGTS-C-----GCN----------
769
770 >*Helix pomatia* CdCuMT1
771 ----------SGK-GSNCAGSCNSNPCSC---GDDCKCGAGCSCVQCHS----------CQCNN------
772 ----------DTCKCGNQ-CSASGSCKCGS--C-----GCK----------
773
774 >*Helix pomatia* CdCuMT2
775 ----------SGK-GSNCAGSCNSNPCSC---GDDCKCGAGCSCAQCHS----------CQCNN------
776 ----------DTCKCGNQ-CSASGSCKCGS--C-----GCK----------
777
778 >*Cepaea hortensis* CdCuMT
779 ----------SGK-ASACAGSCNSNPCSC---GDDCQCGAGCSCAQCHS----------CQCNN------
780 ----------DTCKCGNQ-CSASGSCKCGS--C-----GCK----------
781
782 >*Cepaea nemoralis* CdCuMT
783 ----------SGK-GSACAGSCNSNPCSC---GDDCQCGAGCSCAQCQS----------CQCNN------
784 ----------DTCKCGNQ-CSASGSCKCGS--C-----GCK----------
785
786 >*Cochlicella acuta* CdCuMT
787 ----------SGK-GSACAGSCGNNPCSC---GDDCRCGAGCSCAQCNS----------CQCNN------
788 ----------DTCKCGNQ-CSTSGSCKCGS--C-----GCK----------
789
790 >*Cornu aspersum* CdCuMT
791 ----------SGK-GSACAGSCNSNPCSC---GDDCKCGAGCSCAQCYS----------CQCNN------
792 ----------DTCKCGSQ-CSTSGSCKCGS--C-----GCK----------
793

794

Alignment S4:

MUSCLE alignment (manually edited) of MT protein sequences of Stylommatophora versus Caenogastropoda and Vetigastropoda applied for the Bayesian Inference Tree represented in **Figure S3**.

```
>Megathura crenulata MT
--SGK-----------------------------------------GENCTAECKSDPCACGD---SCKCGEGCACTTCVKTEAKTTCKCG--ESCKC-EG
CKEGEACKCESGCASCK

>Arion vulgaris AvMT1
--SGK-----------------------------------------ACTGACKSEPCQCGN---NCQCGGDCDCSQC------KTCKCTN-EGCKCGQN
CTGQATCSCEKSC-SCK

>Lehmannia nyctelia CdMT
--SGK-----------------------------------------GAKCTGACKSEPCQCGQ---NCQCGDDCSCSQC------KTCKCSAGSTCQCGHG
CTGVESCKCGNSC-SCK

>Limax maximus CdMT
--SGK-----------------------------------------GAKCTGACKSEPCQCGQ---NCQCGDDCSCSQC------KTCKCSAGSTCQCGHG
CTGVESCKCGSSC-SCK

>Deroceras reticulatum CdMT1
--SGK-----------------------------------------GEKCTGDCKSEPCKCGQ---NCQCGNDCTCSQC------KTCKCSTGSGCQCGHG
CTGVESCKCGSSC-T--

>Deroceras reticulatum CdMT2
--SGK-----------------------------------------GEKCTGDCKSEPCKCGQ---NCQCGNDCTCSQC------KTCKCSS-SGCQCGHG
CTGVESCKCGSSC-T--

>Nesiohelix samarangae CdMT
--SGK-----------------------------------------GELCTSACKSNPCQCGD---KCQCGEGCTCSAC------KSCHCTN-DGCNCGKE
CTGPTSCKCDTSC-SCK

>Cochlicella acuta CdMT
SGKGK-----------------------------------------AESCTAQCQSNPCQCGD---KCQCGEGCACTSC------KTCKCTS-DGCKCGKE
CTGPASCKCGSSC-SCK

>Arianta arbustorum CdMT
SGKGK-----------------------------------------GDLCTAACKNEPCQCGS---KCQCGEGCACASC------KTCNCTS-DGCKCGEK
CTGAASCKCGSSC-SCK

>Cornu aspersum CdMT
SGKGK-----------------------------------------GEKCTAACRNEPCQCGS---KCQCGEGCTCAAC------KTCNCTS-DGCKCGKA
CTGPDSCTCGSSC-GCK

>Helix pomatia CdMT
SGKGK-----------------------------------------GEKCTSACRSEPCQCGS---KCQCGEGCTCAAC------KTCNCTS-DGCKCGKE
CTGPDSCKCGSSC-SCK

>Cepaea_hortensis_CdMT1
SGKGK-----------------------------------------GEKCTAACRNEPCQCGS---KCQCGEGCACAAC------KTCNCTS-DGCKCGKE
CTGPDSCKCGSSC-SCK

>Cepaea nemoralis CdMT
SGKGK-----------------------------------------GEKCTAACRNEPCQCGS---KCQCGEGCACAAC------KTCNCTS-DGCKCGKE
CTGPDSCKCGSSC-SCK

>Cepaea hortensis CdMT2
SGKGK-----------------------------------------GEKCTAACRNEPCQCGS---KCQCGEGCACAAC------KTCNCTS-DGCKCGKE
CTGPDSCKCGSLC-SCK

>Arion vulgaris AvMT2
--SGR-----------------------------------------G--CNGTCNSNPCQCED---GCQCGDACSCAQC------NTCKCTN-DGCKCGNE
CTATGSCKCGTSC-GCN

>Alinda biplicata CuMT2
--SGK-----------------------------------------GANCSGACNSNPCQCGD---DCKCGAACSCAEC------NTCKCTN-DSCKCGHD
CSGAGSCKCGNSC-GCK

>Alinda biplicata CuMT1a
--SGK-----------------------------------------GANCTGACNSNPCQCGD---DCKCGVGCSCAEC------NTCKCTN-DGCKCGHG
CTGAGSCKCGNSC-GCK

>Alinda biplicata CuMT1b
--SGK-----------------------------------------GANCTGACNSNPCQSGD---DCKCGVGCSCAEC------NTCKCTN-DGCKCGHG
CTGAGSCKCGNSC-GCK
```

>*Cochlicella acuta* CuMT
--SGR----------G-NCGGACKSNPCSCGQ---VCKCGGACTCAQC------NACRCSG-DSCKCGDQ
CTASGSCQCGSGC-GCK

>*Cochlicella acuta* CdCuMT
--SGK----------GSACAGSCGNNPCSCGD---DCRCGAGCSCAQC------NSCQCNN-DTCKCGNQ
CSTSGSCKCGS-C-GCK

>*Cornu aspersum*_CdCuMT
--SGK----------GSACAGSCNSNPCSCGD---DCKCGAGCSCAQC------YSCQCNN-DTCKCGSQ
CSTSGSCKCGS-C-GCK

>*Cepaea nemoralis* CdCuMT
--SGK----------GSACAGSCNSNPCSCGD---DCQCGAGCSCAQC------QSCQCNN-DTCKCGNQ
CSASGSCKCGS-C-GCK

>*Cepaea hortensis* CdCuMT
--SGK----------ASACAGSCNSNPCSCGD---DCQCGAGCSCAQC------HSCQCNN-DTCKCGNQ
CSASGSCKCGS-C-GCK

>*Helix pomatia* CdCuMT1
--SGK----------GSNCAGSCNSNPCSCGD---DCKCGAGCSCVQC------HSCQCNN-DTCKCGNQ
CSASGSCKCGS-C-GCK

>*Helix pomatia* CdCuMT2
--SGK----------GSNCAGSCNSNPCSCGD---DCKCGAGCSCAQC------HSCQCNN-DTCKCGNQ
CSASGSCKCGS-C-GCK

>*Cornu aspersum* CuMT
--SGR----------GQNCGGACNSNPCNCGN---DCNCGTGCNCDQC------SARHCSN-DDCKCGSQ
CTRSGSCKCGNAC-GCK

>*Theba pisana* CuMT
--SGR----------GKNCGGACNSNPCNCAN---NCRCGAGCNCDSC------SSCHCSN-DDCKCGNQ
CTTSGSCKCGSAC-GCK

>*Helix pomatia* CuMT
--SGR----------GKNCGGACNSNPCSCGN---DCKCGAGCNCDRC------SSCHCSN-DDCKCGSQ
CTGSGSCKCGSAC-GCK

>*Cepaea nemoralis* CuMT
--SGR----------G-NCGGACNSNPCSCGD---DCKCGAACNCDRC------SSCHCSN-DDCKCGSQ
CTGSGSCKCGSAC-GCK

>*Littorina littorea* MT
-----SSVFGAGCTDVCKQTPCGC--ATSGCNCTDDCKCQSC-KYGAGCTDTCKQTPCGCGS---GCNCKEDCRCQSCS-----TACKCAA-GSCKCGKG
CTGPDSCKCDRSC-SCK

>*Pomatias elegans* CdMT1
STSGANVIYGAGCTGTCKQSPCGCKNSAAGCRCKDDCQCPACAKYGAGCTGTCKQSPCGCKNSAAGCGCKDDCRCPACA-----KSCKC---GTCNCGKG
CTGPSNCKCDDGC-SCK

>*Pomatias elegans* CdMT2
SSSGANAT----------GAGCTETCKESPCGCKNSAAGCKCKDDCQCTTCA-----KSCKCA--GTCNCGKG
CTGPNSCKCDGGC-PCK

References

1 V. Vonnemann, M. Schrödl, A. Klussmann-Kolb and H. Wägele, *J. Molluscan Stud.*, 2005, **71**, 113–125.

2 S. Pérez-Rafael, A. Mezger, B. Lieb, R. Dallinger, M. Capdevila, Ò. Palacios and S. Atrian, *J. Inorg. Biochem.*, 2012, **108**, 84–90.

3 C. Baumann, A. Beil, S. Jurt, M. Niederwanger, O. Palacios, M. Capdevila, S. Atrian, R. Dallinger and O. Zerbe, *Angew. Chemie - Int. Ed.*, 2017, **56**, 4617–4622.

4 Ò. Palacios, E. Jiménez-Martí, M. Niederwanger, S. Gil-Moreno, O. Zerbe, S. Atrian, R. Dallinger and M. Capdevila, *Int. J. Mol. Sci.*, 2017, **18**, 1–16.

5 M. Niederwanger, S. Calatayud, O. Zerbe, S. Atrian, R. Albalat, M. Capdevila, Ò. Palacios and R. Dallinger, *Int. J. Mol. Sci.*, , DOI:10.3390/ijms18071457.

6 M. Niederwanger, M. Dvorak, R. Schnegg, V. Pedrini-Martha, K. Bacher, M. Bidoli and R. Dallinger, *Int. J. Mol. Sci.*, 2017, **18**, 1747.

7 M. Dvorak, R. Lackner, M. Niederwanger, C. Rotondo, R. Schnegg, P. Ladurner, V. Pedrini-Martha, W. Salvenmoser, L. Kremser, H. Lindner, M. Garcia-Risco, S. Calatayud, R. Albalat, Ò. Palacios, M. Capdevila and R. Dallinger, *Metallomics*, 2018, **10**, 1638–1654.

8 F. Hispard, D. Schuler, A. De Vaufleury, R. Scheifler, P. M. Badot and R. Dallinger, *Environ. Toxicol. Chem.*, 2008, **27**, 1533–1542.

9 S. Pérez-Rafael, F. Monteiro, R. Dallinger, S. Atrian, Ò. Palacios and M. Capdevila, *Biochim. Biophys. Acta - Proteins Proteomics*, 2014, **1844**, 1694–1707.

10 R. Dallinger, B. Berger, P. Hunziker and J. H. R. Kägi, *Nature*, 1997, **388**, 237–238.

11 Ò. Palacios, A. Pagani, S. Pérez-Rafael, M. Egg, M. Höckner, A. Brandstätter, M. Capdevila, S. Atrian and R. Dallinger, *BMC Biol.*, , DOI:10.1186/1741-7007-9-4.

12 Ò. Palacios, S. Pérez-Rafael, A. Pagani, R. Dallinger, S. Atrian and M. Capdevila, *J. Biol. Inorg. Chem.*, 2014, **19**, 923–935.

BIBLIOGRAPHY

Abdin, A.Y., Jacob, C., Kästner, L., 2021. The Enigmatic Metallothioneins: A Case of Upward-Looking Research. Int. J. Mol. Sci. 22, 5984. https://doi.org/10.3390/ijms22115984

Amaral, A., Cabral, C., Guedes, C., Rodrigues, A., 2007. Apoptosis, metallothionein, and bioavailable metals in domestic mice (Mus musculus L.) from a human-inhabited volcanic area. Ecotoxicology 16, 475–482. https://doi.org/10.1007/s10646-007-0156-y

Amiard, J.-C., Amiard-Triquet, C., Barka, S., Pellerin, J., Rainbow, P.S., 2006. Metallothioneins in aquatic invertebrates: Their role in metal detoxification and their use as biomarkers. Aquat. Toxicol. 76, 160–202. https://doi.org/10.1016/j.aquatox.2005.08.015

Andrade, M.A., Perez-Iratxeta, C., Ponting, C.P., 2001a. Protein Repeats: Structures, Functions, and Evolution. J. Struct. Biol. 134, 117–131. https://doi.org/10.1006/jsbi.2001.4392

Andrade, M.A., Perez-Iratxeta, C., Ponting, C.P., 2001b. Protein Repeats: Structures, Functions, and Evolution. J. Struct. Biol. 134, 117–131. https://doi.org/10.1006/jsbi.2001.4392

Andreeva, A., Kulesha, E., Gough, J., Murzin, A.G., 2020. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. Nucleic Acids Res. 48, D376–D382. https://doi.org/10.1093/nar/gkz1064

Apic, G., Gough, J., Teichmann, S.A., 2001. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. J. Mol. Biol. 310, 311–325. https://doi.org/10.1006/jmbi.2001.4776

Artells, E., Palacios, Ò., Capdevila, M., Atrian, S., 2013. Mammalian MT1 and MT2 metallothioneins differ in their metal binding abilities†. Metallomics 5, 1397–1410. https://doi.org/10.1039/c3mt00123g

Asselman, J., Shaw, J.R., Glaholt, S.P., Colbourne, J.K., De Schamphelaere, K.A.C., 2013. Transcription patterns of genes encoding four metallothionein homologs in Daphnia pulex exposed to copper and cadmium are time- and homolog-dependent. Aquat. Toxicol. 142–143, 422–430. https://doi.org/10.1016/j.aquatox.2013.09.010

Auf der Maur, A., Belser, T., Elgar, G., Georgiev, O., Schaffner, W., 1999. Characterization of the transcription factor MTF-1 from the Japanese pufferfish (Fugu rubripes) reveals evolutionary conservation of heavy metal stress response. Biol. Chem. 380, 175–185. https://doi.org/10.1515/BC.1999.026

Babenka, A., Shevtsova, S., Dromashko, S., 2011. COPPER SULFATE IMPACTS ON GROWTH, SURVIVAL AND METALLOTHIONEIN EXPRESSION IN FRESHWATER POND SNAIL LYMNAEA STAGNALIS. Proc. Belarusian State Univ. 6, 152–162.

Babula, P., Masarik, M., Adam, V., Eckschlager, T., Stiborova, M., Trnkova, L., Skutkova, H., Provaznik, I., Hubalek, J., Kizek, R., 2012. Mammalian metallothioneins: properties and functions. Metallomics 4, 739. https://doi.org/10.1039/c2mt20081c

Baumann, C., Beil, A., Jurt, S., Niederwanger, M., Palacios, O., Capdevila, M., Atrian, S., Dallinger, R., Zerbe, O., 2017. Structural Adaptation of a Protein to Increased Metal Stress: NMR Structure of a Marine Snail Metallothionein with an Additional Domain. Angew. Chem. Int. Ed. 56, 4617–4622. https://doi.org/10.1002/anie.201611873

Beil, A., Jurt, S., Walser, R., Schönhut, T., Güntert, P., Palacios, Ò., Atrian, S., Capdevila, M.,

Dallinger, R., Zerbe, O., 2019. The Solution Structure and Dynamics of Cd-Metallothionein from *Helix pomatia* Reveal Optimization for Binding Cd over Zn. Biochemistry 58, 4570–4581. https://doi.org/10.1021/acs.biochem.9b00830

Berger, B., Hunziker, P.E., Hauer, C.R., Birchler, N., Dallinger, R., 1995. Mass spectrometry and amino acid sequencing of two cadmium-binding metallothionein isoforms from the terrestrial gastropod Arianta arbustorum. Biochem. J. 311, 951–957. https://doi.org/10.1042/bj3110951

Berná, L., Alvarez-Valin, F., 2014. Evolutionary Genomics of Fast Evolving Tunicates. Genome Biol. Evol. 6, 1724–1738. https://doi.org/10.1093/gbe/evu122

Binz, P.-A., Kägi, J.H.R., 1999a. Metallothionein: Molecular evolution and classification, in: Klaassen, C.D. (Ed.), Metallothionein IV, Advances in Life Sciences. Birkhäuser, Basel, pp. 7–13. https://doi.org/10.1007/978-3-0348-8847-9_2

Binz, P.-A., Kägi, J.H.R., 1999b. Metallothionein: Molecular evolution and classification, in: Klaassen, C.D. (Ed.), Metallothionein IV. Birkhäuser Basel, Basel, pp. 7–13. https://doi.org/10.1007/978-3-0348-8847-9_2

Björklund, Å., 2006. Expansion of Protein Domain Repeats. https://doi.org/10.1371/journal.pcbi.0020114

Björklund, Å.K., Ekman, D., Light, S., Frey-Skött, J., Elofsson, A., 2005. Domain Rearrangements in Protein Evolution. J. Mol. Biol. 353, 911–923. https://doi.org/10.1016/j.jmb.2005.08.067

Björklund, Å.K., Light, S., Sagit, R., Elofsson, A., 2010. Nebulin: A Study of Protein Repeat Evolution. J. Mol. Biol. 402, 38–51. https://doi.org/10.1016/j.jmb.2010.07.011

Blindauer, C.A., 2014. CHAPTER 21 Metallothioneins, in: Binding, Transport and Storage of Metal Ions in Biological Cells. The Royal Society of Chemistry, pp. 606–665. https://doi.org/10.1039/9781849739979-00606

Blindauer, C.A., 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. Nat. Prod. Rep. 27, 720–741. https://doi.org/10.1039/B906685N

Blindauer, C.A., Harrison, M.D., Parkinson, J.A., Robinson, A.K., Cavet, J.S., Robinson, N.J., Sadler, P.J., 2001. A metallothionein containing a zinc finger within a four-metal cluster protects a bacterium from zinc toxicity. Proc. Natl. Acad. Sci. U. S. A. 98, 9593–9598. https://doi.org/10.1073/pnas.171120098

Bofill, R., Capdevila, M., Atrian, S., 2009. Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu-thioneins. Metallomics 1, 229. https://doi.org/10.1039/b904953c

Bolker, J.A., 2000. Modularity in Development and Why It Matters to Evo-Devo 1.

Bongers, J., Walton, C.D., Richardson, D.E., Bell, J.U., 1988. Micromolar Protein Concentrations and Metalloprotein Stoichiometries Obtained by Inductively Coupled Plasma Atomic Emission Spectrometric Determination of Sulfur. Anal. Chem. 60, 2683–2686. https://doi.org/10.1021/ac00175a008

Bonner, J.T., 1988. The Evolution of Complexity by Means of Natural Selection. Princeton University Press. https://doi.org/10.2307/j.ctv173f26c

Borgia, A., Kemplen, K.R., Borgia, M.B., Soranno, A., Shammas, S., Wunderlich, B., Nettels, D., Best, R.B., Clarke, J., Schuler, B., 2015. Transient misfolding dominates multidomain protein folding. Nat. Commun. 6, 8861. https://doi.org/10.1038/ncomms9861

Boyle, E.A., Sclater, F., Edmond, J.M., 1976. On the marine geochemistry of cadmium. Nature 263, 42–44. https://doi.org/10.1038/263042a0

Braun, W., Wagner, G., Wörgötter, E., Vašák, M., Kägi, J.H.R., Wüthrich, K., 1986. Polypeptide fold in the two metal clusters of metallothionein-2 by nuclear magnetic resonance in solution. J. Mol. Biol. 187, 125–129. https://doi.org/10.1016/0022-2836(86)90412-2

Bruland, K.W., 1980. Oceanographic distributions of cadmium, zinc, nickel, and copper in the North Pacific. Earth Planet. Sci. Lett. 47, 176–198. https://doi.org/10.1016/0012-821X(80)90035-7

Buchner, J., Kiefhaber, T., Mayer, R.J., Ciechanover, A.J., Rechsteiner, M., Nierhaus, K.H., Wilson, D.N., Brakmann, S., Schwienhorst, A., 2004. Modular Protein Domains, Modular Protein Domains. Wiley. https://doi.org/10.1002/3527603611

Calatayud, S., 2018. Metallothioneins of the urochordate Oikopleura dioica have Cys-rich tandem repeats, large size and cadmium-binding preference 25.

Calatayud, S., Garcia-Risco, M., Capdevila, M., Cañestro, C., Palacios, Ò., Albalat, R., 2021a. Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins. Front. Cell Dev. Biol. 9, 1733. https://doi.org/10.3389/fcell.2021.702688

Calatayud, S., Garcia-Risco, M., Palacios, Ò., Capdevila, M., Cañestro, C., Albalat, R., 2021b. Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions and functional convergences. Mol. Biol. Evol. msab184. https://doi.org/10.1093/molbev/msab184

Calatayud, S., Garcia-Risco, M., Pedrini-Martha, V., Eernisse, D.J., Dallinger, R., Palacios, Ò., Capdevila, M., Albalat, R., 2021c. Modularity in Protein Evolution: Modular Organization and De Novo Domain Evolution in Mollusk Metallothioneins. Mol. Biol. Evol. 38, 424–436. https://doi.org/10.1093/molbev/msaa230

Calderone, V., Dolderer, B., Hartmann, H.-J., Echner, H., Luchinat, C., Del Bianco, C., Mangani, S., Weser, U., 2005. The crystal structure of yeast copper thionein: The solution of a long-lasting enigma. Proc. Natl. Acad. Sci. 102, 51–56. https://doi.org/10.1073/pnas.0408254101

Calvo, J., Jung, H., Meloni, G., 2017. Copper metallothioneins. IUBMB Life 69, 236–245. https://doi.org/10.1002/iub.1618

Cañestro, C., 2012. Two Rounds of Whole-Genome Duplication: Evidence and Impact on the Evolution of Vertebrate Innovations, in: Soltis, P.S., Soltis, D.E. (Eds.), Polyploidy and Genome Evolution. Springer, Berlin, Heidelberg, pp. 309–339. https://doi.org/10.1007/978-3-642-31442-1_16

Capasso, C., Carginale, V., Crescenzi, O., Maro, D.D., Parisi, E., Spadaccini, R., Temussi, P.A., 2003. Solution Structure of MT_nc, a Novel Metallothionein from the Antarctic Fish Notothenia coriiceps. Structure 11, 435–443. https://doi.org/10.1016/S0969-2126(03)00044-3

Capdevila, M., Atrian, S., 2011. Metallothionein protein evolution: a miniassay. JBIC J. Biol. Inorg. Chem. 16, 977–989. https://doi.org/10.1007/s00775-011-0798-3

Capdevila, M., Bofill, R., Palacios, Ò., Atrian, S., 2012. State-of-the-art of metallothioneins at

the beginning of the 21st century. Coord. Chem. Rev., European Group of Thermodynamics of Metal Complexes 256, 46–62. https://doi.org/10.1016/j.ccr.2011.07.006

Chabicovsky, M., 2003. Localization and quantification of Cd- and Cu-specific metallothionein isoform mRNA in cells and organs of the terrestrial gastropod Helix pomatia. Toxicol. Appl. Pharmacol. 190, 25–36. https://doi.org/10.1016/S0041-008X(03)00148-0

Choong, G., Liu, Y., Templeton, D.M., 2014. Interplay of calcium and cadmium in mediating cadmium toxicity. Chem. Biol. Interact. 211, 54–65. https://doi.org/10.1016/j.cbi.2014.01.007

Clune, J., Mouret, J.-B., Lipson, H., 2013. The evolutionary origins of modularity. Proc. R. Soc. B Biol. Sci. 280, 20122863. https://doi.org/10.1098/rspb.2012.2863

Cromar, G., Wong, K.-C., Loughran, N., On, T., Song, H., Xiong, X., Zhang, Z., Parkinson, J., 2014. New Tricks for "Old" Domains: How Novel Architectures and Promiscuous Hubs Contributed to the Organization and Evolution of the ECM. Genome Biol. Evol. 6, 2897–2917. https://doi.org/10.1093/gbe/evu228

Cunha, T.J., Giribet, G., 2019. A congruent topology for deep gastropod relationships. Proc. R. Soc. B Biol. Sci. 286, 20182776. https://doi.org/10.1098/rspb.2018.2776

Dallinger, R., Berger, B., Hunziger, P., Kgi, J.H.R., 1997. Metallothionein in snail Cd and Cu metabolism. Nature 388, 237–238. https://doi.org/10.1038/40785

Dallinger, R., Zerbe, O., Baumann, C., Egger, B., Capdevila, M., Palacios, Ò., Albalat, R., Calatayud, S., Ladurner, P., Schlick-Steiner, B.C., Steiner, F.M., Pedrini-Martha, V., Lackner, R., Lindner, H., Dvorak, M., Niederwanger, M., Schnegg, R., Atrian, S., 2020. Metallomics reveals a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins. Metallomics 12, 702–720. https://doi.org/10.1039/C9MT00259F

Darwin, C., 1809-1882, 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859.

D'Auria, S., Carginale, V., Scudiero, R., Crescenzi, O., Di Maro, D., Temussi, P.A., Parisi, E., Capasso, C., 2001. Structural characterization and thermal stability of Notothenia coriiceps metallothionein. Biochem. J. 354, 291–299. https://doi.org/10.1042/bj3540291

de Francisco, P., Melgar, L.M., Díaz, S., Martín-González, A., Gutiérrez, J.C., 2016. The Tetrahymena metallothionein gene family: twenty-one new cDNAs, molecular characterization, phylogenetic study and comparative analysis of the gene expression under different abiotic stressors. BMC Genomics 17, 346. https://doi.org/10.1186/s12864-016-2658-6

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.-M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L.C., Jubin, C., Canestro, C., Bouquet, J.-M., Danks, G., Poulain, J., Campsteijn, C., Adamski, M., Cross, I., Yadetie, F., Muffato, M., Louis, A., Butcher, S., Tsagkogeorga, G., Konrad, A., Singh, S., Jensen, M.F., Cong, E.H., Eikeseth-Otteraa, H., Noel, B., Anthouard, V., Porcel, B.M., Kachouri-Lafond, R., Nishino, A., Ugolini, M., Chourrout, P., Nishida, H., Aasland, R., Huzurbazar, S., Westhof, E., Delsuc, F., Lehrach, H., Reinhardt, R., Weissenbach, J., Roy, S.W., Artiguenave, F., Postlethwait, J.H., Manak, J.R., Thompson, E.M., Jaillon, O., Du Pasquier, L., Boudinot, P., Liberles, D.A., Volff, J.-N., Philippe, H., Lenhard, B., Crollius, H.R., Wincker, P., Chourrout, D., 2010. Plasticity of Animal Genome Architecture Unmasked by Rapid Evolution of a Pelagic Tunicate. Science 330, 1381–1385. https://doi.org/10.1126/science.1194167

Dohmen, E., Klasberg, S., Bornberg-Bauer, E., Perrey, S., Kemena, C., 2020. The modular nature of protein evolution: domain rearrangement rates across eukaryotic life. BMC Evol. Biol.

20, 30. https://doi.org/10.1186/s12862-020-1591-0

Donner, K., Sandbacka, M., Lehtokari, V.-L., Wallgren-Pettersson, C., Pelin, K., 2004. Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. Eur. J. Hum. Genet. 12, 744–751. https://doi.org/10.1038/sj.ejhg.5201242

Doolittle, R., 1981. Similar amino acid sequences: chance or common ancestry? Science 214, 149–159. https://doi.org/10.1126/science.7280687

Dvorak, M., Lackner, R., Niederwanger, M., Rotondo, C., Schnegg, R., Ladurner, P., Pedrini-Martha, V., Salvenmoser, W., Kremser, L., Lindner, H., García-Risco, M., Calatayud, S., Albalat, R., Palacios, Ò., Capdevila, M., Dallinger, R., 2018. Metal binding functions of metallothioneins in the slug *Arion vulgaris* differ from metal-specific isoforms of terrestrial snails. Metallomics 10, 1638–1654. https://doi.org/10.1039/C8MT00215K

Dvorak, M., Schnegg, R., Niederwanger, M., Pedrini-Martha, V., Ladurner, P., Lindner, H., Kremser, L., Lackner, R., Dallinger, R., 2019. Cadmium Pathways in Snails Follow a Complementary Strategy between Metallothionein Detoxification and Auxiliary Inactivation by Phytochelatins. Int. J. Mol. Sci. 21, 7. https://doi.org/10.3390/ijms21010007

Edvardsen, R.B., Seo, H.-C., Jensen, M.F., Mialon, A., Mikhaleva, J., Bjordal, M., Cartry, J., Reinhardt, R., Weissenbach, J., Wincker, P., Chourrout, D., 2005. Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. Curr. Biol. 15, R12–R13. https://doi.org/10.1016/j.cub.2004.12.010

Egg, M., Höckner, M., Brandstätter, A., Schuler, D., Dallinger, R., 2009. Structural and bioinformatic analysis of the Roman snail Cd-Metallothionein gene uncovers molecular adaptation towards plasticity in coping with multifarious environmental stress. Mol. Ecol. 18, 2426–2443. https://doi.org/10.1111/j.1365-294X.2009.04191.x

Ellis, J., Dodds, P., Pryor, T., 2000. Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. 3, 278–284. https://doi.org/10.1016/S1369-5266(00)00080-7

Feng, W., Benz, F.W., Cai, J., Pierce, W.M., Kang, Y.J., 2006. Metallothionein Disulfides Are Present in Metallothionein-overexpressing Transgenic Mouse Heart and Increase under Conditions of Oxidative Stress. J. Biol. Chem. 281, 681–687. https://doi.org/10.1074/jbc.M506956200

Ferrández-Roldán, A., Martí-Solans, J., Cañestro, C., Albalat, R., 2019. Oikopleura dioica: An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development. Results Probl. Cell Differ. 68, 63–105. https://doi.org/10.1007/978-3-030-23459-1_4

Ferrara, T.M., Flaherty, D.B., Benian, G.M., 2006. Titin/connectin-related proteins in C. elegans: a review and new findings. J. Muscle Res. Cell Motil. 26, 435–447. https://doi.org/10.1007/s10974-005-9027-4

Flaherty, D.B., Gernert, K.M., Shmeleva, N., Tang, X., Mercer, K.B., Borodovsky, M., Benian, G.M., 2002. Titins in C.elegans with Unusual Features: Coiled-coil Domains, Novel Regulation of Kinase Activity and Two New Possible Elastic Regions. J. Mol. Biol. 323, 533–549. https://doi.org/10.1016/S0022-2836(02)00970-1

Fowler, B.A., Hildebrand, C.E., Kojima, Y., Webb, M., 1987. Nomenclature of Metallothionein, in: Kägi, J.H.R., Kojima, Yutaka (Eds.), Metallothionein II: Proceedings of the «Second International Meeting on Metallothionein and Other Low Molecular Weight Metalbinding Proteins», Zürich, August 21–24, 1985, Experientia Supplementum. Birkhäuser, Basel, pp. 19–

22. https://doi.org/10.1007/978-3-0348-6784-9_2

Franchi, N., Boldrin, F., Ballarin, L., Piccinni, E., 2011. CiMT-1, an unusual chordate metallothionein gene in Ciona intestinalis genome: structure and expression studies. J. Exp. Zool. Part Ecol. Genet. Physiol. 315A, 90–100. https://doi.org/10.1002/jez.653

Fraternali, F., Pastore, A., 1999. Modularity and homology: modelling of the type II module family from titin. J. Mol. Biol. 290, 581–593. https://doi.org/10.1006/jmbi.1999.2876

Frey, P.A., 2014. Advances in Research on Metalloproteins, in: Fontecilla-Camps, J.C., Nicolet, Y. (Eds.), Metalloproteins: Methods and Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 1–3. https://doi.org/10.1007/978-1-62703-794-5_1

García-Risco, M., Calatayud, S., Niederwanger, M., Albalat, R., Palacios, Ò., Capdevila, M., Dallinger, R., 2020. Two Unconventional Metallothioneins in the Apple Snail Pomacea bridgesii Have Lost Their Metal Specificity during Adaptation to Freshwater Habitats. Int. J. Mol. Sci. 22, 95. https://doi.org/10.3390/ijms22010095

Garud, R., Kumaraswamy, A., Langlois, R., 2002. Managing in the modular age : architectures, networks, and organizations, College of Business & Public Management Faculty Books. Wiley.

Gehrig, P.M., You, C., Dallinger, R., Gruber, C., Brouwer, M., Kägi, J.H., Hunziker, P.E., 2000. Electrospray ionization mass spectrometry of zinc, cadmium, and copper metallothioneins: evidence for metal-binding cooperativity. Protein Sci. Publ. Protein Soc. 9, 395–402.

Giedroc, D.P., Chen, X., Apuy, J.L., 2001. Metal response element (MRE)-binding transcription factor-1 (MTF-1): structure, function, and regulation. Antioxid. Redox Signal. 3, 577–596. https://doi.org/10.1089/15230860152542943

Gil-Moreno, S., Jiménez-Martí, E., Palacios, Ò., Zerbe, O., Dallinger, R., Capdevila, M., Atrian, S., 2015. Does Variation of the Inter-Domain Linker Sequence Modulate the Metal Binding Behaviour of Helix pomatia Cd-Metallothionein? Int. J. Mol. Sci. 17, 6. https://doi.org/10.3390/ijms17010006

Gudekar, N., Shanbhag, V., Wang, Y., Ralle, M., Weisman, G.A., Petris, M.J., 2020. Metallothioneins regulate ATP7A trafficking and control cell viability during copper deficiency and excess. Sci. Rep. 10, 1–11. https://doi.org/10.1038/s41598-020-64521-3

Guirola, M., Pérez-Rafael, S., Capdevila, M., Palacios, Ò., Atrian, S., 2012. Metal Dealing at the Origin of the Chordata Phylum: The Metallothionein System and Metal Overload Response in Amphioxus. PLOS ONE 7, e43299. https://doi.org/10.1371/journal.pone.0043299

Han, J.-H., Batey, S., Nickson, A.A., Teichmann, S.A., Clarke, J., 2007. The folding and evolution of multidomain proteins. Nat. Rev. Mol. Cell Biol. 8, 319–330. https://doi.org/10.1038/nrm2144

Hans Wedepohl, K., 1995. The composition of the continental crust. Geochim. Cosmochim. Acta 59, 1217–1232. https://doi.org/10.1016/0016-7037(95)00038-2

Hansen, T.F., 2003. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability 12.

Haq, F., 2003. Signaling events for metallothionein induction. Mutat. Res. Mol. Mech. Mutagen. 533, 211–226. https://doi.org/10.1016/j.mrfmmm.2003.07.014

Harlow, P., Watkins, E., Thornton, R.D., Nemer, M., 1989. Structure of an ectodermally expressed sea urchin metallothionein gene and characterization of its metal-responsive region.

Mol. Cell. Biol. 9, 5445–5455.

Henikoff, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K., Hood, L., 1997. Gene Families: The Taxonomy of Protein Paralogs and Chimeras. Science 278, 609–614.

Heringa, J., 2005. Protein domains, in: Jorde, L.B., Little, P.F.R., Dunn, M.J., Subramaniam, S. (Eds.), Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics. John Wiley & Sons, Ltd, Chichester, p. g406204. https://doi.org/10.1002/047001153X.g406204

Higgins, D.G., Labeit, S., Gautel, M., Gibson, T.J., 1994. The evolution of titin and related giant muscle proteins. J. Mol. Evol. 38, 395–404. https://doi.org/10.1007/BF00163156

Hispard, F., Schuler, D., de Vaufleury, A., Scheifler, R., Badot, P.-M., Dallinger, R., 2008. METAL DISTRIBUTION AND METALLOTHIONEIN INDUCTION AFTER CADMIUM EXPOSURE IN THE TERRESTRIAL SNAIL HELIX ASPERSA (GASTROPODA, PULMONATA). Environ. Toxicol. Chem. 27, 1533. https://doi.org/10.1897/07-232.1

Hynes, M.J., 2003. The Neurospora crassagenome opens up the world of filamentous fungi. Genome Biol. 4, 217. https://doi.org/10.1186/gb-2003-4-6-217

Imagawa, M., Onozawa, T., Okumura, K., Osada, S., Nishihara, T., Kondo, M., 1990. Characterization of metallothionein cDNAs induced by cadmium in the nematode Caenorhabditis elegans. Biochem. J. 268, 237–240.

Isani, G., Carpenè, E., 2014. Metallothioneins, Unconventional Proteins from Unconventional Animals: A Long Journey from Nematodes to Mammals. Biomolecules 4, 435–457. https://doi.org/10.3390/biom4020435

Iturbe-Espinoza, P., Gil-Moreno, S., Lin, W., Calatayud, S., Palacios, Ò., Capdevila, M., Atrian, S., 2016. The Fungus Tremella mesenterica Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization. PLOS ONE 11, e0148651. https://doi.org/10.1371/journal.pone.0148651

Jenny, M.J., Payton, S.L., Baltzegar, D.A., Lozier, J.D., 2016. Phylogenetic Analysis of Molluscan Metallothioneins: Evolutionary Insight from Crassostrea virginica. J. Mol. Evol. 83, 110–125. https://doi.org/10.1007/s00239-016-9758-4

Jernigan, K.K., Bordenstein, S.R., 2015. Tandem-repeat protein domains across the tree of life. PeerJ 3, e732. https://doi.org/10.7717/peerj.732

Juárez-Rebollar, D., Rios, C., Nava-Ruíz, C., Méndez-Armenta, M., 2017. Metallothionein in Brain Disorders. Oxid. Med. Cell. Longev. 2017, 1–12. https://doi.org/10.1155/2017/5828056

Kägi, J.H.R., Kojima, Y., 1979. Nomenclature of Metallothionein: A Proposal, in: Kägi, J.H.R., Nordberg, M. (Eds.), Metallothionein: Proceedings of the »First International Meeting on Metallothionein and Other Low Molecular Weight Metal-Binding Proteins« Zürich, July 17–22, 1978, Experientia Supplementum. Birkhäuser, Basel, pp. 141–142. https://doi.org/10.1007/978-3-0348-6493-0_4

Kang, Y.J., Li, G., Saari, J.T., 1999. Metallothionein inhibits ischemia-reperfusion injury in mouse heart. Am. J. Physiol.-Heart Circ. Physiol. 276, H993–H997. https://doi.org/10.1152/ajpheart.1999.276.3.H993

Kenny, P.A., Liston, E.M., Higgins, D.G., 1999. Molecular evolution of immunoglobulin and fibronectin domains in titin and related muscle proteins. Gene 232, 11–23. https://doi.org/10.1016/S0378-1119(99)00122-5

Klasberg, S., Bitard-Feildel, T., Callebaut, I., Bornberg-Bauer, E., 2018. Origins and structural properties of novel and *de novo* protein domains during insect evolution. FEBS J. 285, 2605–2625. https://doi.org/10.1111/febs.14504

Kocot, K.M., Tassia, M.G., Halanych, K.M., Swalla, B.J., 2018. Phylogenomics offers resolution of major tunicate relationships. Mol. Phylogenet. Evol. 121, 166–173. https://doi.org/10.1016/j.ympev.2018.01.005

Kojima, Y., Binz, P.-A., Kägi, J.H.R., 1999. Nomenclature of metallothionein: Proposal for a revision, in: Klaassen, C.D. (Ed.), Metallothionein IV, Advances in Life Sciences. Birkhäuser, Basel, pp. 3–6. https://doi.org/10.1007/978-3-0348-8847-9_1

Kovalevskij, A.O., 1866. Entwickelungsgeschichte der einfachen Ascidien. Eggers & Schmitzdorff.

Lemoine, S., Bigot, Y., Sellos, D., Cosson, R.P., Laulier, M., 2000. Metallothionein Isoforms in Mytilus edulis (Mollusca, Bivalvia): Complementary DNA Characterization and Quantification of Expression in Different Organs after Exposure to Cadmium, Zinc, and Copper. Mar. Biotechnol. 2, 195–203. https://doi.org/10.1007/s101269900026

Levy, A., 2019. How evolution builds genes from scratch. Nature 574, 314–316. https://doi.org/10.1038/d41586-019-03061-x

Lichtlen, P., Schaffner, W., 2001. Putting its fingers on stressful situations: the heavy metal-regulatory transcription factor MTF-1. BioEssays 23, 1010–1017. https://doi.org/10.1002/bies.1146

Lieb, B., 2003. A new metallothionein gene from the giant keyhole limpet Megathura crenulata. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 134, 131–137. https://doi.org/10.1016/S1532-0456(02)00215-6

Liu, J., Liu, Y., Michalska, A.E., Choo, K.H., Klaassen, C.D., 1996. Metallothionein plays less of a protective role in cadmium-metallothionein-induced nephrotoxicity than in cadmium chloride-induced hepatotoxicity. J. Pharmacol. Exp. Ther. 276, 1216–1223.

Liu, Y., Liu, J., Iszard, M.B., Andrews, G.K., Palmiter, R.D., Klaassen, C.D., 1995. Transgenic mice that overexpress metallothionein-I are protected from cadmium lethality and hepatotoxicity. Toxicol. Appl. Pharmacol. 135, 222–228. https://doi.org/10.1006/taap.1995.1227

Liu, Y., Wu, H., Kou, L., Liu, X., Zhang, J., Guo, Y., Ma, E., 2014. Two Metallothionein Genes in Oxya chinensis: Molecular Characteristics, Expression Patterns and Roles in Heavy Metal Stress. PLoS ONE 9, e112759. https://doi.org/10.1371/journal.pone.0112759

Liu, Y., Yang, C., Sun, L., Wang, A., Lan, X., Xu, W., Liang, Y., Ma, S., Xia, Q., 2021. In-depth transcriptome unveils the cadmium toxicology and a novel metallothionein in silkworm. Chemosphere 273, 128522. https://doi.org/10.1016/j.chemosphere.2020.128522

Lorenz, D.M., Jeng, A., Deem, M.W., 2011. The emergence of modularity in biological systems. Phys. Life Rev. S1571064511000170. https://doi.org/10.1016/j.plrev.2011.02.003

Louis, A., Roest Crollius, H., Robinson-Rechavi, M., 2012. How much does the amphioxus genome represent the ancestor of chordates? Brief. Funct. Genomics 11, 89–95. https://doi.org/10.1093/bfgp/els003

Luo, M., Finet, C., Cong, H., Wei, H., Chung, H., 2020. The evolution of insect metallothioneins. Proc. R. Soc. B Biol. Sci. 287, 20202189.

https://doi.org/10.1098/rspb.2020.2189

Mao, J., Yu, H., Wang, C., Sun, L., Jiang, W., Zhang, P., Xiao, Q., Han, D., Saiyin, H., Zhu, J., Chen, T., Roberts, L.R., Huang, H., Yu, L., 2012. Metallothionein MT1M is a tumor suppressor of human hepatocellular carcinomas. Carcinogenesis 33, 2568–2577. https://doi.org/10.1093/carcin/bgs287

Marcotte, E.M., Pellegrini, M., Yeates, T.O., Eisenberg, D., 1999. A census of protein repeats. J. Mol. Biol. 293, 151–160. https://doi.org/10.1006/jmbi.1999.3136

Margoshes, M., Vallee, B.L., 1957. A CADMIUM PROTEIN FROM EQUINE KIDNEY CORTEX. J. Am. Chem. Soc. 79, 4813–4814. https://doi.org/10.1021/ja01574a064

Martínez-Paz, P., Morales, M., Sánchez-Argüello, P., Morcillo, G., Martínez-Guitarte, J.L., 2017. Cadmium in vivo exposure alters stress response and endocrine-related genes in the freshwater snail Physa acuta. New biomarker genes in a new model organism. Environ. Pollut. 220, 1488–1497. https://doi.org/10.1016/j.envpol.2016.10.012

Mary, R.D., Saravanan, M.K., Selvaraj, S., 2015. Conservation of inter-residue interactions and prediction of folding rates of domain repeats. J. Biomol. Struct. Dyn. 33, 534–551. https://doi.org/10.1080/07391102.2014.894944

McElhinny, A., 2003. Nebulin The Nebulous, Multifunctional Giant of Striated Muscle. Trends Cardiovasc. Med. 13, 195–201. https://doi.org/10.1016/S1050-1738(03)00076-8

Messerle, B.A., Schäffer, A., Vašák, M., Kägi, J.H.R., Wüthrich, K., 1990. Three-dimensional structure of human [113Cd7]metallothionein-2 in solution determined by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 214, 765–779. https://doi.org/10.1016/0022-2836(90)90291-S

Michael Niederwanger, Martin Dvorak, Raimund Schnegg, Veronika Pedrini-Martha, Katharina Bacher, Massimo Bidoli, Reinhard Dallinger, 2017. Challenging the Metallothionein (MT) Gene of Biomphalaria glabrata: Unexpected Response Patterns Due to Cadmium Exposure and Temperature Stress. Int. J. Mol. Sci. 18, 1747. https://doi.org/10.3390/ijms18081747

Mikowska, M., Dziublińska, B., Świergosz-Kowalewska, R., 2018. Variation of Metallothionein I and II Gene Expression in the Bank Vole (Clethrionomys glareolus) Under Environmental Zinc and Cadmium Exposure. Arch. Environ. Contam. Toxicol. 75, 66–74. https://doi.org/10.1007/s00244-017-0485-7

M'kandawire, E., Mierek-Adamska, A., Stürzenbaum, S.R., Choongo, K., Yabe, J., Mwase, M., Saasa, N., Blindauer, C.A., 2017. Metallothionein from Wild Populations of the African Catfish Clarias gariepinus: From Sequence, Protein Expression and Metal Binding Properties to Transcriptional Biomarker of Metal Pollution. Int. J. Mol. Sci. 18, 1548. https://doi.org/10.3390/ijms18071548

Moffatt, P., Séguin, C., 1998. Expression of the gene encoding metallothionein-3 in organs of the reproductive system. DNA Cell Biol. 17, 501–510. https://doi.org/10.1089/dna.1998.17.501

Moore, A.D., Björklund, Å.K., Ekman, D., Bornberg-Bauer, E., Elofsson, A., 2008. Arrangements in the modular evolution of proteins. Trends Biochem. Sci. 33, 444–451. https://doi.org/10.1016/j.tibs.2008.05.008

Moore, A.D., Bornberg-Bauer, E., 2012. The Dynamics and Evolutionary Potential of Domain Loss and Emergence. Mol. Biol. Evol. 29, 787–796. https://doi.org/10.1093/molbev/msr250

Muñoz, A., Försterling, H.F., Shaw, F.C., Petering, D.H., 2002. Structure of the 113Cd3β

domains from Homarus americanus metallothionein-1: hydrogen bonding and solvent accessibility of sulfur atoms. JBIC J. Biol. Inorg. Chem. 7, 713–724. https://doi.org/10.1007/s00775-002-0345-3

Narula, S.S., Brouwer, M., Hua, Y., Armitage, I.M., 1995. Three-Dimensional Solution Structure of Callinectes sapidus Metallothionein-1 Determined by Homonuclear and Heteronuclear Magnetic Resonance Spectroscopy. Biochemistry 34, 620–631. https://doi.org/10.1021/bi00002a029

Needham, J., 1933. ON THE DISSOCIABILITY OF THE FUNDAMENTAL PROCESSES IN ONTOGENESIS. Biol. Rev. 8, 180–223. https://doi.org/10.1111/j.1469-185X.1933.tb01153.x

Neme, R., Tautz, D., 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics 14, 117. https://doi.org/10.1186/1471-2164-14-117

Onosaka, S., Tanaka, K., Cheriant, M.G., 1984. Effects of Cadmium and Zinc on Tissue Levels of Metallothionein. Environmental Health Perspectives Vol. 54, 7. https://doi.org/161.116.100.31

Orihuela, R., Domènech, J., Bofill, R., You, C., Mackay, E.A., Kägi, J.H.R., Capdevila, M., Atrian, S., 2008. The metal-binding features of the recombinant mussel Mytilus edulis MT-10-IV metallothionein. JBIC J. Biol. Inorg. Chem. 13, 801–812. https://doi.org/10.1007/s00775-008-0367-6

Otvos, J.D., Armitage, I.M., 1980. Structure of the metal clusters in rabbit liver metallothionein. Proc. Natl. Acad. Sci. U. S. A. 77, 7094–7098.

Otvos, J.D., Engeseth, H.R., Wehrli, S., 1985. Preparation and 113Cd NMR studies of homogeneous reconstituted metallothionein: reaffirmation of the two-cluster arrangement of metals. Biochemistry 24, 6735–6740. https://doi.org/10.1021/bi00345a001

Otvos, J.D., Olafson, R.W., Armitage, I.M., 1982. Structure of an invertebrate metallothionein from Scylla serrata. J. Biol. Chem. 257, 2427–2431. https://doi.org/10.1016/S0021-9258(18)34941-X

Overnell, J., Coombs, T.L., 1979. Purification and properties of plaice metallothionein, a cadmium-binding protein from the liver of the plaice (Pleuronectes platessa). Biochem. J. 183, 277–283. https://doi.org/10.1042/bj1830277

Overnell, J., Good, M., Vašàk, M., 1988. Spectroscopic studies on cadmium(II)- and cobalt(II)-substituted metallothionein from the crab Cancer pagurus. Eur. J. Biochem. 172, 171–177. https://doi.org/10.1111/j.1432-1033.1988.tb13869.x

Öz, G., Zangger, K., Armitage, I.M., 2001. Three-Dimensional Structure and Dynamics of a Brain Specific Growth Inhibitory Factor: Metallothionein-3. Biochemistry 40, 11433–11441. https://doi.org/10.1021/bi010827l

Padmanabhan, M., Cournoyer, P., Dinesh-Kumar, S.P., 2009. The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. Cell. Microbiol. 11, 191–198. https://doi.org/10.1111/j.1462-5822.2008.01260.x

Palacios, Ò., Albalat, R., 2021. Modular Evolution and Population Variability of Oikopleura dioica Metallothioneins. Front. Cell Dev. Biol. 9, 15.

Palacios, Ò., Atrian, S., Capdevila, M., 2011. Zn- and Cu-thioneins: a functional classification for metallothioneins? JBIC J. Biol. Inorg. Chem. 16, 991–1009. https://doi.org/10.1007/s00775-

011-0827-2

Palacios, Ò., Espart, A., Espín, J., Ding, C., Thiele, D.J., Atrian, S., Capdevila, M., 2014. Full characterization of the Cu-, Zn-, and Cd-binding properties of CnMT1 and CnMT2, two metallothioneins of the pathogenic fungus Cryptococcus neoformans acting as virulence factors. Metallomics 6, 279–291. https://doi.org/10.1039/C3MT00266G

Palacios, Ò., Jiménez-Martí, E., Niederwanger, M., Gil-Moreno, S., Zerbe, O., Atrian, S., Dallinger, R., Capdevila, M., 2017. Analysis of Metal-Binding Features of the Wild Type and Two Domain-Truncated Mutant Variants of Littorina littorea Metallothionein Reveals Its Cd-Specific Character. Int. J. Mol. Sci. 18, 1452. https://doi.org/10.3390/ijms18071452

Palmiter, R.D., 1998. The elusive function of metallothioneins. Proc. Natl. Acad. Sci. 95, 8428–8430. https://doi.org/10.1073/pnas.95.15.8428

Papadopoulou, C., Kanias, G.D., 1977. Tunicate species as marine pollution indicators. Mar. Pollut. Bull. 8, 229–231. https://doi.org/10.1016/0025-326X(77)90431-3

Patthy, L., 1985. Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. Cell 41, 657–663. https://doi.org/10.1016/S0092-8674(85)80046-5

Pedrini-Martha, V., Köll, S., Dvorak, M., Dallinger, R., 2020. Cadmium Uptake, MT Gene Activation and Structure of Large-Sized Multi-Domain Metallothioneins in the Terrestrial Door Snail Alinda biplicata (Gastropoda, Clausiliidae). Int. J. Mol. Sci. 21, 1631. https://doi.org/10.3390/ijms21051631

Pedrini-Martha, V., Schnegg, R., Baurand, P.-E., deVaufleury, A., Dallinger, R., 2017. The physiological role and toxicological significance of the non-metal-selective cadmium/copper-metallothionein isoform differ between embryonic and adult helicid snails. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol., Unraveling complexity: from molecules to ecosystems 199, 38–47. https://doi.org/10.1016/j.cbpc.2017.02.009

Pedrini-Martha, V., Schnegg, R., Schäfer, G.G., Lieb, B., Salvenmoser, W., Dallinger, R., 2021. Responsiveness of metallothionein and hemocyanin genes to cadmium and copper exposure in the garden snail Cornu aspersum. J. Exp. Zool. Part Ecol. Integr. Physiol. 335, 228–238. https://doi.org/10.1002/jez.2425

Peregrín-Alvarez, J.M., Sanford, C., Parkinson, J., 2009. The conservation and evolutionary modularity of metabolism. Genome Biol. 10, R63. https://doi.org/10.1186/gb-2009-10-6-r63

Pérez-Rafael, S., Kurz, A., Guirola, M., Capdevila, M., Palacios, Ò., Atrian, S., 2012a. Is MtnE, the fifth Drosophila metallothionein, functionally distinct from the other members of this polymorphic protein family? Metallomics 4, 342. https://doi.org/10.1039/c2mt00182a

Pérez-Rafael, S., Mezger, A., Lieb, B., Dallinger, R., Capdevila, M., Palacios, Ò., Atrian, S., 2012b. The metal binding abilities of Megathura crenulata metallothionein (McMT) in the frame of Gastropoda MTs. J. Inorg. Biochem., CanBIC Special Issue: 3rd Georgian Bay International Conference on Bioinorganic Chemistry 108, 84–90. https://doi.org/10.1016/j.jinorgbio.2011.11.025

Pérez-Rafael, S., Monteiro, F., Dallinger, R., Atrian, S., Palacios, Ò., Capdevila, M., 2014. Cantareus aspersus metallothionein metal binding abilities: The unspecific CaCd/CuMT isoform provides hints about the metal preference determinants in metallothioneins. Biochim. Biophys. Acta BBA - Proteins Proteomics 1844, 1694–1707. https://doi.org/10.1016/j.bbapap.2014.06.018

Peroza, E.A., Schmucki, R., Güntert, P., Freisinger, E., Zerbe, O., 2009. The βE-Domain of Wheat Ec-1 Metallothionein: A Metal-Binding Domain with a Distinctive Structure. J. Mol. Biol. 387, 207–218. https://doi.org/10.1016/j.jmb.2009.01.035

Peterson, C.W., Narula, S.S., Armitage, I.M., 1996. 3D solution structure of copper and silver-substituted yeast metallothioneins. FEBS Lett. 379, 85–93. https://doi.org/10.1016/0014-5793(95)01492-6

Quaife, C.J., Findley, S.D., Erickson, J.C., Froelick, G.J., Kelly, E.J., Zambrowicz, B.P., Palmiter, R.D., 1994. Induction of a new metallothionein isoform (MT-IV) occurs during differentiation of stratified squamous epithelia. Biochemistry 33, 7250–7259. https://doi.org/10.1021/bi00189a029

Raff, E.C., Raff, R.A., 2000. Dissociability, modularity, evolvability. Evol. Dev. 2, 235–237. https://doi.org/10.1046/j.1525-142x.2000.00069.x

Raghunath, A., Sundarraj, K., Nagarajan, R., Arfuso, F., Bian, J., Kumar, A.P., Sethi, G., Perumal, E., 2018. Antioxidant response elements: Discovery, classes, regulation and potential applications. Redox Biol. 17, 297–314. https://doi.org/10.1016/j.redox.2018.05.002

Ragusa, M.A., Costa, S., Gianguzza, M., Roccheri, M.C., Gianguzza, F., 2013. Effects of cadmium exposure on sea urchin development assessed by SSH and RT-qPCR: metallothionein genes and their differential induction. Mol. Biol. Rep. 40, 2157–2167. https://doi.org/10.1007/s11033-012-2275-7

Ragusa, M.A., Nicosia, A., Costa, S., Cuttitta, A., Gianguzza, F., 2017. Metallothionein Gene Family in the Sea Urchin Paracentrotus lividus: Gene Structure, Differential Expression and Phylogenetic Analysis. Int. J. Mol. Sci. 18, 812. https://doi.org/10.3390/ijms18040812

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A., 1995. 14. W. Jetz, C. Rahbek, Proc. Natl. Acad. Sci. U.S.A. Chica-go Univ. Press.

Reshef, D., Itzhaki, Z., Schueler-Furman, O., 2010. Increased sequence conservation of domain repeats in prokaryotic proteins. Trends Genet. 26, 383–387. https://doi.org/10.1016/j.tig.2010.06.003

Riek, R., Prêcheur, B., Wang, Y., Mackay, E.A., Wider, G., Güntert, P., Liu, A., Kägi, J.H., Wüthrich, K., 1999. NMR structure of the sea urchin (Strongylocentrotus purpuratus) metallothionein MTA. J. Mol. Biol. 291, 417–428. https://doi.org/10.1006/jmbi.1999.2967

Rokas, A., Holland, P.W.H., 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15, 454–459. https://doi.org/10.1016/S0169-5347(00)01967-4

Romero-Isart, N., Vasák, M., 2002. Advances in the structure and chemistry of metallothioneins. J. Inorg. Biochem. 88, 388–396. https://doi.org/10.1016/s0162-0134(01)00347-6

Rychel, A.L., Smith, S.E., Shimamoto, H.T., Swalla, B.J., 2006. Evolution and Development of the Chordates: Collagen and Pharyngeal Cartilage. Mol. Biol. Evol. 23, 541–549. https://doi.org/10.1093/molbev/msj055

Ryvolova, M., Krizkova, S., Adam, V., Beklova, M., Trnkova, L., Hubalek, J., Kizek, R., 2011. Analytical Methods for Metallothionein Detection. Curr. Anal. Chem. 7, 243–261. https://doi.org/10.2174/157341011107030243

Sabolić, I., Breljak, D., Škarica, M., Herak-Kramberger, C.M., 2010. Role of metallothionein in cadmium traffic and toxicity in kidneys and other mammalian organs. BioMetals 23, 897–926.

https://doi.org/10.1007/s10534-010-9351-z

Saint-Jacques, É., Séguin, C., 1993. Cloning and Nucleotide Sequence of a Complementary DNA Encoding Xenopus laevis Metallothionein: mRNA Accumulation in Response to Heavy Metals. DNA Cell Biol. 12, 329–340. https://doi.org/10.1089/dna.1993.12.329

Saydam, N., Adams, T.K., Steiner, F., Schaffner, W., Freedman, J.H., 2002. Regulation of Metallothionein Transcription by the Metal-responsive Transcription Factor MTF-1. J. Biol. Chem. 277, 20438–20445. https://doi.org/10.1074/jbc.M110631200

Schaper, E., Gascuel, O., Anisimova, M., 2014. Deep Conservation of Human Protein Tandem Repeats within the Eukaryotes. Mol. Biol. Evol. 31, 1132–1148. https://doi.org/10.1093/molbev/msu062

Schmielau, L., Dvorak, M., Niederwanger, M., Dobieszewski, N., Pedrini-Martha, V., Ladurner, P., Pedregal, J.R.-G., Maréchal, J.-D., Dallinger, R., 2019. Differential response to Cadmium exposure by expression of a two and a three-domain metallothionein isoform in the land winkle Pomatias elegans: Valuating the marine heritage of a land snail. Sci. Total Environ. 648, 561–571. https://doi.org/10.1016/j.scitotenv.2018.07.426

Schmitt-Wrede, H.-P., Koewius, H., Tschuschke, S., Greven, H., Wunderlich, F., 2004. Genomic organization of the cadmium-inducible tandem repeat 25-kDa metallothionein of the oligochaete worm Enchytraeus buchholzi. Biochim. Biophys. Acta BBA - Gene Struct. Expr. 1680, 24–33. https://doi.org/10.1016/j.bbaexp.2004.08.007

Schüler, A., Bornberg-Bauer, E., 2016. Evolution of Protein Domain Repeats in Metazoa. Mol. Biol. Evol. 33, 3170–3182. https://doi.org/10.1093/molbev/msw194

Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J.H.R., Wüthrich, K., 1988. Conformation of [Cd7]-metallothionein-2 from rat liver in aqueous solution determined by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 203, 251–268. https://doi.org/10.1016/0022-2836(88)90106-4

Scudiero, R., Verderame, M., Motta, C.M., Simoniello, P., 2017. Unravelling the Role of Metallothionein on Development, Reproduction and Detoxification in the Wall Lizard Podarcis sicula. Int. J. Mol. Sci. 18, 1569. https://doi.org/10.3390/ijms18071569

Serén, N., Glaberman, S., Carretero, M.A., Chiari, Y., 2014. Molecular Evolution and Functional Divergence of the Metallothionein Gene Family in Vertebrates. J. Mol. Evol. 78, 217–233. https://doi.org/10.1007/s00239-014-9612-5

Sigel, A., Sigel, H., Sigel, R.K. (Eds.), 2013. Cadmium: From Toxicity to Essentiality, Metal Ions in Life Sciences. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-5179-8

Sigel, A., Sigel, H., Sigel, R.K.O., Royal Society of Chemistry (Great Britain) (Eds.), 2009. Metallothioneins and related chelators, Metal ions in life sciences. RSC Pub, Cambridge, UK.

Silar, P., Butler, G., Thiele, D.J., 1991. Heat shock transcription factor activates transcription of the yeast metallothionein gene. Mol. Cell. Biol. 11, 1232–1238.

Silar, P., Théodore, L., Mokdad, R., Erraïss, N.-E., Cadic, A., Wegnez, M., 1990. Metallothionein Mto gene of Drosophila melanogaster: Structure and regulation. J. Mol. Biol. 215, 217–224. https://doi.org/10.1016/S0022-2836(05)80340-7

Simon, H.A., 1962. The Architecture of Complexity (No. 6), Proceedings of the American Philosophical Society.

Škutková, H., Provazník, I., Kizek, R., 2010. Similarity analysis of primary and secondary structures of metallothioneins, in: Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine. Presented at the Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine, pp. 1–4. https://doi.org/10.1109/ITAB.2010.5687624

Somorjai, I., Bertrand, S., Camasses, A., Haguenauer, A., Escriva, H., 2008. Evidence for stasis and not genetic piracy in developmental expression patterns of Branchiostoma lanceolatum and Branchiostoma floridae, two amphioxus species that have evolved independently over the course of 200 Myr. Dev. Genes Evol. 218, 703–713. https://doi.org/10.1007/s00427-008-0256-6

Stennard, F.A., Holloway, A.F., Hamilton, J., West, A.K., 1994. Characterisation of six additional human metallothionein genes. Biochim. Biophys. Acta BBA - Gene Struct. Expr. 1218, 357–365. https://doi.org/10.1016/0167-4781(94)90189-9

Street, T.O., Rose, G.D., Barrick, D., 2006. The Role of Introns in Repeat Protein Gene Formation. J. Mol. Biol. 360, 258–266. https://doi.org/10.1016/j.jmb.2006.05.024

Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95–98. https://doi.org/10.1038/nature09864

Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M., 2007. Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol. Biol. 7, 57. https://doi.org/10.1186/1471-2148-7-57

Stürzenbaum, S. r, Kille, P., Morgan, A. j, 1998. The identification, cloning and characterization of earthworm metallothionein. FEBS Lett. 431, 437–442. https://doi.org/10.1016/S0014-5793(98)00809-6

Syring, R.A., Hoexum Brouwer, T., Brouwer, M., 2000. Cloning and sequencing of cDNAs encoding for a novel copper-specific metallothionein and two cadmium-inducible metallothioneins from the blue crab Callinectes sapidus. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. 125, 325–332. https://doi.org/10.1016/S0742-8413(99)00114-0

Tamai, K.T., Liu, X., Silar, P., Sosinowski, T., Thiele, D.J., 1994. Heat shock transcription factor activates yeast metallothionein gene expression in response to heat and glucose starvation via distinct signalling pathways. Mol. Cell. Biol. 14, 8155–8165. https://doi.org/10.1128/MCB.14.12.8155

Tío, L., Villarreal, L., Atrian, S., Capdevila, M., 2004. Functional Differentiation in the Mammalian Metallothionein Gene Family: METAL BINDING FEATURES OF MOUSE MT4 AND COMPARISON WITH ITS PARALOG MT1*. J. Biol. Chem. 279, 24403–24413. https://doi.org/10.1074/jbc.M401346200

Toll-Riera, M., Albà, M.M., 2013. Emergence of novel domains in proteins. BMC Evol. Biol. 13, 47. https://doi.org/10.1186/1471-2148-13-47

Toll-Riera, M., Rado-Trilla, N., Martys, F., Alba, M.M., 2012. Role of Low-Complexity Sequences in the Formation of Novel Protein Coding Sequences. Mol. Biol. Evol. 29, 883–886. https://doi.org/10.1093/molbev/msr263

Tom, M., Moran, O., Jakubov, E., Cavari, B., Rinkevich, B., 1998. Molecular characterization of Metallothionein-cDNA of Sparus aurata used for detecting heavy metal pollution along the Mediterranean coast of Israel. Mar. Pollut. Bull. 36, 131–137. https://doi.org/10.1016/S0025-

326X(97)87421-8

Tomas, M., Domènech, J., Capdevila, M., Bofill, R., Atrian, S., 2013. The sea urchin metallothionein system: Comparative evaluation of the SpMTA and SpMTB metal-binding preferences. FEBS Open Bio 3, 89–100. https://doi.org/10.1016/j.fob.2013.01.005

Tran, D., Fournier, E., Durrieu, G., Massabuau, J.-C., 2007. Inorganic mercury detection by valve closure response in the freshwater clam Corbicula fluminea: integration of time and water metal concentration changes. Environ. Toxicol. Chem. 26, 1545–1551. https://doi.org/10.1897/06-390r1.1

Tran, D., Fournier, E., Durrieu, G., Massabuau, J.-C., 2004. Copper detection in the Asiatic clam Corbicula fluminea: optimum valve closure response. Aquat. Toxicol. Amst. Neth. 66, 333–343. https://doi.org/10.1016/j.aquatox.2004.01.006

Tsagkogeorga, G., Cahais, V., Galtier, N., 2012. The Population Genomics of a Fast Evolver: High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate Ciona intestinalis. Genome Biol. Evol. 4, 852–861. https://doi.org/10.1093/gbe/evs054

Tzafriri-Milo, R., Benaltabet, T., Torfstein, A., Shenkar, N., 2019. The Potential Use of Invasive Ascidians for Biomonitoring Heavy Metal Pollution. Front. Mar. Sci. 6, 611. https://doi.org/10.3389/fmars.2019.00611

Valls, M., Bofill, R., González-Duarte, R., González-Duarte, P., Capdevila, M., Atrian, S., 2001. A New Insight into Metallothionein (MT) Classification and Evolution. J. Biol. Chem. 276, 32835–32843. https://doi.org/10.1074/jbc.M102151200

Van Campenhout, K., Infante, H.G., Adams, F., Blust, R., 2004. Induction and Binding of Cd, Cu, and Zn to Metallothionein in Carp (Cyprinus carpio) Using HPLC-ICP-TOFMS. Toxicol. Sci. 80, 276–287. https://doi.org/10.1093/toxsci/kfh149

Vašák, M., Meloni, G., 2017. Mammalian Metallothionein-3: New Functional and Structural Insights. Int. J. Mol. Sci. 18, 1117. https://doi.org/10.3390/ijms18061117

Vašák, M., Meloni, G., 2011. Chemistry and biology of mammalian metallothioneins. JBIC J. Biol. Inorg. Chem. 16, 1067. https://doi.org/10.1007/s00775-011-0799-2

Vergani, L., Grattarola, M., Grasselli, E., Dondero, F., Viarengo, A., 2007. Molecular characterization and function analysis of MT-10 and MT-20 metallothionein isoforms from Mytilus galloprovincialis. Arch. Biochem. Biophys. 465, 247–253. https://doi.org/10.1016/j.abb.2007.05.023

Villarreal, L., Tío, L., Capdevila, M., Atrian, S., 2006. Comparative metal binding and genomic analysis of the avian (chicken) and mammalian metallothionein. FEBS J. 273, 523–535. https://doi.org/10.1111/j.1742-4658.2005.05086.x

Wagner, G., Frey, M.H., Neuhaus, D., Wörgötter, E., Braun, W., Vasak, M., Kägi, J.H.R., Wüthrich, K., 1987. Spatial Structure of Rabbit Liver Metallothionein-2 in Solution by NMR, in: Kägi, Jeremias H. R., Kojima, Y. (Eds.), Metallothionein II: Proceedings of the «Second International Meeting on Metallothionein and Other Low Molecular Weight Metalbinding Proteins», Zürich, August 21–24, 1985, Experientia Supplementum. Birkhäuser, Basel, pp. 149–157. https://doi.org/10.1007/978-3-0348-6784-9_8

Wagner, G.P., 2001. The Character Concept in Evolutionary Biology. Academic Press.

Wang, H., Zhang, Q., Cai, B., Li, H., Sze, K.-H., Huang, Z.-X., Wu, H.-M., Sun, H., 2006. Solution structure and dynamics of human metallothionein-3 (MT-3). FEBS Lett. 580, 795–800.

https://doi.org/10.1016/j.febslet.2005.12.099

Wang, M., Caetano-Anollés, G., 2009. The Evolutionary Mechanics of Domain Organization in Proteomes and the Rise of Modularity in the Protein World. Structure 17, 66–78. https://doi.org/10.1016/j.str.2008.11.008

Weisman, C.M., Eddy, S.R., 2017. Gene Evolution: Getting Something from Nothing. Curr. Biol. 27, R661–R663. https://doi.org/10.1016/j.cub.2017.05.056

West, A.K., Stallings, R., Hildebrand, C.E., Chiu, R., Karin, M., Richards, R.I., 1990. Human metallothionein genes: Structure of the functional locus at 16q13. Genomics 8, 513–518. https://doi.org/10.1016/0888-7543(90)90038-V

Westin, G., Schaffner, W., 1988. A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-I gene. EMBO J. 7, 3763–3770.

Winge, D.R., Miklossy, K.A., 1982. Domain nature of metallothionein. J. Biol. Chem. 257, 3471–3476. https://doi.org/10.1016/S0021-9258(18)34802-6

Winge, D.R., Nielson, K.B., Zeikus, R.D., Gray, W.R., 1984. Structural characterization of the isoforms of neonatal and adult rat liver metallothionein. J. Biol. Chem. 259, 11419–11425. https://doi.org/10.1016/S0021-9258(18)90877-X

Won, E.-J., Rhee, J.-S., Ra, K., Kim, K.-T., Au, D.W.T., Shin, K.-H., Lee, J.-S., 2012. Molecular cloning and expression of novel metallothionein (MT) gene in the polychaete Perinereis nuntia exposed to metals. Environ. Sci. Pollut. Res. 19, 2606–2618. https://doi.org/10.1007/s11356-012-0905-1

Wright, C.F., Teichmann, S.A., Clarke, J., Dobson, C.M., 2005. The importance of sequence diversity in the aggregation and evolution of proteins. Nature 438, 878–881. https://doi.org/10.1038/nature04195

Yu, C.-W., Chen, J.-H., Lin, L.-Y., 1997. Metal-induced metallothionein gene expression can be inactivated by protein kinase C inhibitor. FEBS Lett. 420, 69–73. https://doi.org/10.1016/S0014-5793(97)01489-0

Zafarullah, M., Bonham, K., Gedamu, L., 1988. Structure of the rainbow trout metallothionein B gene and characterization of its metal-responsive region. Mol. Cell. Biol. 8, 4469–4476.

Zangger, K., Öz, G., Armitage, I.M., Otvos, J.D., 1999. Three-dimensional solution structure of mouse [Cd7]-metallothionein-l by homonuclear and heteronuclear NMR spectroscopy. Protein Sci. 8, 2630–2638. https://doi.org/10.1110/ps.8.12.2630

Zapata, F., Wilson, N.G., Howison, M., Andrade, S.C.S., Jörger, K.M., Schrödl, M., Goetz, F.E., Giribet, G., Dunn, C.W., 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. Proc. R. Soc. B Biol. Sci. 281, 20141739. https://doi.org/10.1098/rspb.2014.1739

Zeitoun-Ghandour, S., Charnock, J.M., Hodson, M.E., Leszczyszyn, O.I., Blindauer, C.A., Stürzenbaum, S.R., 2010. The two Caenorhabditis elegans metallothioneins (CeMT-1 and CeMT-2) discriminate between essential zinc and toxic cadmium. FEBS J. 277, 2531–2542. https://doi.org/10.1111/j.1742-4658.2010.07667.x

Zeng, J., Heuchel, R., Schaffner, W., Kägi, J.H.R., 1991. Thionein (apometallothionein) can modulate DNA binding and transcription activation by zinc finger containing factor Spl. FEBS Lett. 279, 310–312. https://doi.org/10.1016/0014-5793(91)80175-3

Zhang, B., Georgiev, O., Hagmann, M., Günes, Ç., Cramer, M., Faller, P., Vasák, M., Schaffner, W., 2003. Activity of Metal-Responsive Transcription Factor 1 by Toxic Heavy Metals and H_2O_2 In Vitro Is Modulated by Metallothionein. Mol. Cell. Biol. 23, 8471–8485. https://doi.org/10.1128/MCB.23.23.8471-8485.2003

Ziller, A., Fraissinet-Tachet, L., 2018. Metallothionein diversity and distribution in the tree of life: a multifunctional protein. Metallomics 10, 1549–1559. https://doi.org/10.1039/C8MT00165K

Zorita, I., Bilbao, E., Schad, A., Cancio, I., Soto, M., Cajaraville, M.P., 2007. Tissue- and cell-specific expression of metallothionein genes in cadmium- and copper-exposed mussels analyzed by in situ hybridization and RT–PCR. Toxicol. Appl. Pharmacol. 220, 186–196. https://doi.org/10.1016/j.taap.2007.01.003