



UNIVERSITAT DE  
BARCELONA

## Functional interplays of proteins and proteases: AD13-VWF, TGF $\beta$ 2- $\alpha$ <sub>2</sub>M, and the proteolysis of gliadin by neprosin

Laura del Amo Maestro

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE  
BARCELONA



**UNIVERSITAT DE BARCELONA**

Departamento de Bioquímica y Biología Molecular

Facultad de Farmacia

**CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS  
INSTITUTO DE BIOLOGÍA MOLECULAR DE BARCELONA**

Departamento de Biología Estructural

Proteolysis lab

**Functional interplays of proteins and  
proteases: AD13-VWF, TGF $\beta$ 2- $\alpha$ <sub>2</sub>M, and the  
proteolysis of gliadin by neprosin**

**Laura del Amo Maestro**

Barcelona, febrero 2021



**UNIVERSITAT DE BARCELONA**

Facultad de Farmacia

Programa de Doctorado de Biotecnología 2016-2021

**Functional interplays of proteins and  
proteases: AD13-VWF, TGF $\beta$ 2- $\alpha$ <sub>2</sub>M, and the  
proteolysis of gliadin by neprosin**

Memoria presentada por Laura del Amo Maestro para optar al título de doctor por  
la Universidad de Barcelona.

**Laura del Amo Maestro**

*Supervisor:*

**Prof. F. Xavier Gomis-Rüth**

*Tutora:*

**Dr. Josefa Badia Palacín**

Barcelona, febrero 2021





*A todas esas personas que desde el inicio  
me brindaron todo su apoyo incondicional,*



*Un científico en su laboratorio no es un simple técnico:  
también es un niño que se enfrenta a fenómenos naturales que  
lo impresionan como si fueran cuentos de hadas.*

*Marie Curie*



## ACKNOWLEDGEMENTS

---

Al finalizar un trabajo tan arduo y lleno de dificultades como es el desarrollo de una tesis doctoral te demuestra inmediatamente que el aporte y dedicación que has realizado hubiese sido imposible sin la participación de personas e instituciones que han facilitado las cosas para que este trabajo llegue a un feliz término. Por ello, es para mí un verdadero placer utilizar este espacio para ser justa y consecuente con ellas, expresándoles mis agradecimientos.

En primer lugar, me gustaría agradecer de manera especial y sincera al Prof. Xavier Gomis-Rüth la oportunidad que me ha brindado al aceptarme en su laboratorio bajo su dirección. Su apoyo y confianza en mi trabajo y su capacidad de guiar mis ideas y motivación ha sido un aporte invaluable, no solamente en el desarrollo de esta tesis, sino también en mi formación como investigadora. Le agradezco también el haberme facilitado siempre los medios suficientes para llevar a cabo todos los experimentos propuestos durante el desarrollo de la tesis.

Quiero expresar también mi más sincero agradecimiento al Dr. Ulrich Eckhard y al Dr. Diego Sebastian Ferrero por su importante aporte, participación y corrección de esta tesis. Debo de destacar, por encima de todo su disponibilidad y paciencia que hizo que nuestras discusiones científicas redundaran benéficamente tanto a nivel profesional como personal. Gracias Diego por los innumerables cristales pescados tanto a 20 como a 4°C. Aprovecho para agradecer a Arturo Rodríguez sus siempre atentas y rápidas respuestas a las diferentes inquietudes surgidas durante el desarrollo de este trabajo, muchas de ellas acompañadas de risas y humor.

Agradezco de manera especial al Prof. Jan Enghild de la Universidad de Aarhus, Dinamarca y en particular a la Dr. Nadia Sukusu por su disponibilidad y generosidad para compartir su experiencia y amplio conocimiento en proteómica durante mi estancia en su grupo, durante la cual obtuve todo el soporte profesional y logístico para alcanzar los objetivos perseguidos. Muchas gracias Nadia por tu apoyo y tus gestos de generosidad fuera del laboratorio, siempre conseguías sacarme una sonrisa al mismo tiempo que me has hecho sentir muy cómoda durante toda mi estancia.

Para mis compañeros de grupo, doctorandas y postdocs, tengo solo palabras de cariño y agradecimiento, especialmente por los consejos y soporte moral, así como los buenos momentos compartidos como los videos de final de tesis, los cafés de emergencia, nuestros refranes y frases míticas, las cervezas, las cenas fuera del laboratorio y nuestras escapadas y

viajes. También por el apoyo en momentos no tan buenos, que han sido parte del proceso necesario para mi desarrollo personal. Destaco a Laura Mariño y Soraia Mendes, siempre dispuestas a ayudarme en todo lo que necesitase, compartiendo siempre conocimientos y experiencias de tipo profesional y personal. En definitiva, unas amigas para toda la vida.

Quiero expresar mi agradecimiento a todos los compañeros del IBMB con los que he tenido la oportunidad de trabajar, y aquellos que han estado presentes en mi día a día. Antonia, gracias por tus calurosos buenos días y tu risa contagiosa. Inés y Andrea gracias por haber formado parte de esta aventura y haber enriquecido de manera significativa estos años de doctorado.

Para aquellos amigos que han compartido “ires y venires” durante esta larga estancia en la ciudad de Barcelona. A mis mejores amigas en Madrid, Carlota, Alicia e Irene de quienes siempre he recibido palabras de aliento. Gracias por animarme, confiar en mí y comprender la importancia de mis estudios.

Y, por supuesto, el agradecimiento más profundo y sentido va para mi familia. Sin su apoyo, colaboración e inspiración habría sido imposible llevar a cabo esta dura etapa. A mis padres, Manuel y Paloma por su ejemplo de lucha y honestidad; a mi hermana Paloma por su tenacidad y superación; y por último a mi marido Luca por la paciencia y comprensión, por estar a mi lado en todo momento y por la fuerza que siempre me da para seguir adelante. Gracias por ser un pilar en todas las etapas que hemos vivido juntos, y las que nos quedan por vivir.

## ABSTRACT

---

The present thesis has established a new and state-of-the-art mammalian protein expression system at the laboratory, which has led to high-yield production procedures for the purification and crystallization of three glycoproteins, namely TGF $\beta$ 2, ADAMTS13 and neprosin. These proteins adopt important biomedical roles in human physiology, cardiovascular disorders, and high potential in the treatment of coeliac disease, respectively.

Proteases are major players in the physiology and pathology of all living organisms. Their often highly sophisticated regulation mechanisms are therefore essential for proper function, and to prevent misdirected spatial and/or temporal proteolytic activity, which in turn may trigger disease. This regulation is achieved through a wide variety of mechanisms including protein-protein interaction, substrate allosteric activation, or their biosynthesis as inactive precursor, so called zymogens. In the present thesis, a variety of regulatory mechanisms of proteases and inhibitors were studied by using a combination of biochemical, biophysical and structural techniques.

In the first project, we comprehensively illuminated the biochemistry of the VWF:ADAMTS13 axis, which is involved in haemostasis maintenance. ADAMTS13 is a metalloprotease that regulates the multimeric form of von Willebrand Factor (VWF) in blood circulation by specific, shear-dependent proteolysis. Despite circulating in a seemingly latent form with a very long active plasma half-life, ADAMTS13 is resistant to plasma inhibitors and VWF is the only known substrate, which causes an allosteric activation of ADAMTS13 upon binding. This unprecedented specificity has been attributed to extensive exosite interactions between peptidase domains, where the disintegrin-like domain (D) approaches the metalloprotease (M), cysteine-rich (C), and spacer (S) domains once a 73-peptide spanning the relevant fragment of VWF that is efficiently cleaved by the peptidase is bound, thus achieving a “tight” conformation. Structural analysis in solution revealed that the enzyme adopts a highly-flexible unbound structure that is latent. Through the integration of our experimental studies with computational approaches, we concluded that the active VWF-peptide-bound structure significantly differs from the AD13-MDTCS crystal structure previously published. Thus, we hypothesize that the interaction gives rise to a ‘fuzzy complex’ that follows a ‘dynamic zipper’ mechanism involving numerous reversible, weak but additive interactions that result in strong binding and, ultimately, in cleavage.



In the second project, a novel family of glutamate peptidases was studied in the form of the endopeptidase neprosin, and the crystal structure of both the active and zymogenic forms was solved. We ascertained the active site, catalytic mechanism and protease class adscription, and described a unique catalytic dyad composed by two glutamates. Neprosin has been recently discovered in the fluid of the carnivorous pitcher plant *Nepenthes ventrata*. Given its high activity at low pH, thermic stability and great specificity, neprosin is potentially useful in detoxifying gluten proteins, including wheat gliadin and its 33-meric major immunogenic fragment, into non-toxic peptides in an environment mimicking gastric conditions. Additionally, we determined the inhibitory profile of neprosin against a cohort of peptidase inhibitors. Future approaches will study the effect of neprosin processing of the 33-mer on inflammatory responses in cells and tissue explants to elucidate if neprosin indeed is suited for an effective and harmless oral supplement for gluten-sensitive patients.

In the third project, we characterised the interaction between the main protease inhibitor of the blood plasma, human  $\alpha$ 2-macroglobulin (h $\alpha$ 2Ms), with a cytokine that displays high potential in therapeutic applications, human TGF $\beta$ 2, through a variety of biochemical, biophysical, and binding techniques. Intriguingly, during our crystallographic studies of the complex, we obtained the crystal structure of a TGF $\beta$ 2 dimer in a novel dimeric assembly of biological unknown function.

Overall, the present thesis contributes significantly to the field of structural biochemistry by expanding previous knowledge at the molecular level and adding new regulatory mechanisms, which will hopefully pave the way for the design of specific drugs for therapeutic interventions.

## ABSTRACT (Castellano)

---

En el presente trabajo de tesis se ha puesto a punto un sistema de expresión de proteínas de mamífero de última generación en el laboratorio. Esto ha contribuido a la expresión de tres glicoproteínas con un alto rendimiento, así como su purificación y cristalización, a saber, TGF $\beta$ 2, ADAMTS13 y neprosina. Estas proteínas llevan a cabo, respectivamente, funciones esenciales en la fisiología humana, los trastornos cardiovasculares y, potencialmente, en el tratamiento de la enfermedad celíaca.

Las proteasas participan de manera decisiva en la fisiología y patología de todos los organismos vivos. Su estricta regulación es, por tanto, esencial para el correcto funcionamiento y para evitar una actividad proteolítica inadecuada en el tiempo y/o el espacio que a su vez puede desencadenar enfermedades. Esta regulación se consigue mediante un amplio abanico de mecanismos, entre ellos la interacción proteína-proteína, la activación alostérica por sustrato o su biosíntesis como precursores inactivos, también conocido como zimógenos. En el presente trabajo se estudiaron diversos mecanismos de regulación de las proteasas e inhibidores mediante una combinación de técnicas bioquímicas, biofísicas y estructurales.

En el primer proyecto se estudió de forma exhaustiva la bioquímica del eje VWF:ADAMTS13, implicado en el mantenimiento de la hemostasia. ADAMTS13 es una metaloproteasa responsable de la regulación de la forma multimérica del Factor von Willebrand (VWF) en la circulación sanguínea mediante proteólisis específica y dependiente de fuerzas de cizallamiento. A pesar de que circula en una forma aparentemente latente con una vida media muy larga, ADAMTS13 es resistente a los inhibidores de proteinasas de plasma y VWF se ha descrito como su único sustrato conocido. Éste último a su vez lleva a cabo una activación alostérica de ADAMTS13 a través de su unión. Esta especificidad se atribuye a extensas interacciones entre los dominios MDTCS de la proteinasa, donde el dominio D se aproxima al dominio M, C y al dominio espaciador S, logrando de esta manera una conformación globular del complejo. Se han realizado análisis estructurales que demuestran que la proteasa adopta una estructura latente altamente flexible en solución. Mediante la integración de nuestros estudios experimentales con enfoques computacionales concluimos que las estructuras activas ligadas a un péptido de 73 residuos, que constituye la parte relevante de VWF, difieren de forma significativa de la estructura cristalina AD13-MDTCS publicada anteriormente. Por consiguiente, hemos llegado a la hipótesis de que la

interacción culmina en un "complejo difuso" que sigue un mecanismo dinámico que implica numerosas interacciones reversibles, débiles pero aditivas, que dan lugar a una fuerte unión y, en última instancia, a la proteólisis.

En el segundo proyecto se llevó a cabo el estudio de una novedosa familia de glutamato peptidasas centrada en neprosina con el fin de determinar la estructura cristalográfica, tanto de la forma activa como de la forma zimogénica. De esta forma también definimos el sitio activo, el mecanismo catalítico y la asignación de la clase de proteasa correspondiente, ya que describe una díada catalítica única compuesta por dos glutamatos, pero a su vez se ha visto que muestra una estrecha homología estructural con la familia de las eqolisinas.

La neprosina se descubrió recientemente en el fluido de la planta carnívora *Nepenthes ventrata*. Gracias a su extraordinaria actividad a pH ácido, su alta estabilidad térmica y su gran especificidad, la neprosina cumple todos los requisitos para ser especialmente útil para la proteólisis y degradación de las proteínas del gluten en péptidos no tóxicos. Entre ellos podemos destacar la gliadina de trigo, y un fragmento inmunogénico derivado, el 33-mer, responsables del desarrollo de la enfermedad celiaca en individuos susceptibles. Además, también evaluamos el perfil inhibitorio de la neprosina frente a un grupo de inhibidores de la peptidasa. En futuros ensayos se estudiará el efecto del procesamiento de la gliadina y del 33-mer por la neprosina sobre las respuestas inflamatorias en las células y los explantes de los tejidos para dilucidar si dicha proteasa presenta realmente un perfil de suplemento alimenticio eficaz e inócuo para los pacientes con intolerancia al gluten.

En el tercer proyecto, se caracterizó la interacción entre el principal inhibidor de proteasas en el plasma, la  $\alpha_2$  macroglobulina humana ( $h\alpha_2Ms$ ), con una citoquina con gran potencial en aplicaciones terapéuticas, la TGF $\beta$ 2 humana, mediante una variedad de técnicas bioquímicas, biofísicas y estudios de interacción. La  $h\alpha_2Ms$  tiene un gran impacto en la fisiología humana modulando la actividad de citoquinas como TGF $\beta$ 2. Curiosamente, durante nuestros estudios de cristalización del complejo, obtuvimos la estructura cristalográfica de un dímero de la forma madura de TGF $\beta$ 2 en una conformación novedosa y de función biológica desconocida.

En definitiva, esta tesis contribuye de manera significativa a ampliar el conocimiento previo de la bioquímica estructural a nivel molecular y entender los mecanismos que regulan la actividad, facilitando así el diseño de fármacos específicos como parte de aproximaciones terapéuticas.

# TABLE OF CONTENTS

---

ABSTRACT .....	XI
ABSTRACT (Castellano).....	XII
INDEX OF FIGURES AND TABLES .....	IXI
ABBREVIATIONS .....	XXI

## INTRODUCTION

1. Recent advances in the production of recombinant proteins in mammalian cells for structural biology .....	3
1.1. Production of recombinant therapeutic proteins .....	3
1.2. Expression system selection .....	4
1.3. Development of a mammalian expression system .....	5
1.3.1. Cell line selection.....	6
1.3.2. Glycosylation differences in mammalian cell lines.....	7
1.3.3. Transient vs Stable Systems .....	8
1.3.4. Transfection .....	9
1.3.4.1. Chemical transfection methods .....	9
1.3.4.2. Physical transfection methods.....	10
1.3.4.3. Biological transfection methods.....	10
1.3.5. Expression vector design .....	10
1.3.6. Signal peptide selection .....	11
1.4. Seleno-methionine (Se-Met) incorporation .....	12
1.5. Control of N-linked glycosylation.....	13
1.5.1. Inhibitors .....	13
2. Proteolytic enzymes .....	15
2.1. Active site .....	16
2.2. Classification .....	16
2.2.1. Specificity.....	16
2.2.2. Catalytic mechanism .....	17
2.2.3. MEROPS: The peptidase database.....	18

2.3. Biological relevance.....	19
2.4. Proteolytic regulation.....	19
2.4.1. Zymogenicity .....	19
2.4.2. Post-translational modifications.....	21
2.4.3. Protease Inhibitors.....	21
2.4.3.1. Reversible inhibitors .....	22
2.4.3.2. Irreversible inhibitors .....	23
2.4.4. Cofactor.....	24
2.4.5. Allosteric regulation .....	24
2.4.6. Protein compartmentalisation and protease trafficking.....	24
3. The intriguing interplay of ADAMTS13 and VWF.....	26
3.1. Circulatory system and shear stress .....	26
3.2. Maintenance of Haemostasis .....	26
a. Primary haemostasis .....	26
b. Secondary haemostasis .....	27
c. Fibrinolysis.....	28
d. The balance of haemostasis .....	28
3.3. Von Willebrand factor (VWF).....	28
3.3.1. Biosynthesis and secretion .....	28
3.3.2. Function and structure of VWF .....	29
3.3.3. Regulation of size and conformation of VWF .....	30
3.4. ADAMTS13.....	32
3.4.1. ADAMTS13 protease is part of a family of metalloproteases.....	32
3.4.2. Gen, synthesis and secretion .....	33
3.4.3. AD13 domain organization and structure.....	34
3.4.3.1. The propeptide (P <sup>30</sup> -R <sup>74</sup> ) .....	34
3.4.3.2. The metalloprotease domain (A <sup>75</sup> -A <sup>293</sup> ) .....	34
3.4.3.3. The disintegrin-like domain (G <sup>294</sup> -P <sup>379</sup> ) .....	37
3.4.3.4. The thrombospondin type 1-like domain (TSP1) (I <sup>380</sup> -E <sup>439</sup> ).....	37
3.4.3.5. The cysteine-rich domain (K <sup>440</sup> -C <sup>555</sup> ).....	37

3.4.3.6. The spacer domain (S <sup>556</sup> -P <sup>682</sup> ) .....	38
3.4.3.7. CUB domains .....	38
3.4.4. Recognition and cleavage of VWF by AD13 .....	38
3.4.4.1. AD13-VWF interactions .....	38
3.4.4.2. The conformational activation of ADAMTS13 .....	40
3.4.4.3. Factors affecting cleavage of VWF by ADAMTS13.....	41
3.4.5. ADAMTS13 in disease .....	42
3.4.5.1. Thrombotic thrombocytopenic purpura (TTP).....	42
3.4.5.2. Congenital TTP .....	42
3.4.5.3. Acquired TTP .....	44
4. Prolyl-endopeptidase, neprosin .....	45
4.1. Coeliac disease .....	45
4.1.1. Epidemiology and pathophysiology .....	46
4.2. Wheat composition: Gluten .....	47
4.2.1. Gliadins .....	48
4.3. Proteases involved in protein digestion .....	49
4.4. Gluten immunogenic peptides .....	50
4.4.1. Properties of the 33-mer peptide from $\alpha$ -gliadin .....	51
4.5. Novel coeliac disease therapies .....	52
4.5.1. Detoxification of gluten proteins with enzymatic therapy .....	52
4.6. Digestive secretion from carnivorous plants .....	53
4.6.1. Plant aspartic proteases: nephenthesins .....	54
4.6.2. A novel plant prolyl endoprotease: neprosin .....	55
4.6.2.1. Neprosin characterization: recombinant protein .....	55
4.6.2.2. Biological roles of neprosin.....	56
4.6.2.3. Molecular structure of neprosin.....	57
5. Protein-protein interaction of TGF $\beta$ 2 with $\alpha$ <sub>2</sub> M .....	58
5.1. TGF $\beta$ Overview .....	58
5.2. TGF $\beta$ cytokines .....	58

5.2.1. Processing of TGF $\beta$ .....	59
5.2.2. Mechanism of activation .....	61
5.3. Soluble TGF $\beta$ binding proteins .....	62
5.3.1. Alpha-2 macroglobulin ( $\alpha_2$ M).....	62
5.4. Biological interactions of h $\alpha_2$ M/TGF $\beta$ .....	63
<b>OBJECTIVES</b> .....	67
<b>RESULTS</b>	
Informe del Director.....	71
Project 1 - <i>An integrative structural biology analysis of von Willebrand factor binding and processing by ADAMTS13 in homeostasis</i> .....	75
Project 2 - <i>Structure, function, zymogenic latency and catalytic mechanism of a rare plant glutamate peptidase with therapeutic potential in coeliac disease</i> .....	107
Project 3 - <i>Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor <math>\beta</math>2 in a new conformation</i> .....	133
<i>Recombinant production of human <math>\alpha_2</math>-macroglobulin variants and interaction studies with recombinant G-related <math>\alpha_2</math>-macroglobulin binding protein and latent TGF-<math>\beta</math>2</i> .....	149
<b>GENERAL DISCUSSION</b> .....	163
G.1. Functional characterization: cleavage and binding studies.....	164
G.1.1. Protein dynamics: structural proteomics (CLMS and HDX-MS).....	165
G.1.2. Integrative analysis: in-solution structural analysis.....	168
G.2. Structural and functional characterization <i>in vitro</i> of neprosin.....	171
G.2.1. Gluten degradation: gliadin and 33-mer .....	172
G.2.2. Structural similarity with glutamate peptidases.....	173
G.2.3. Mechanism of latency.....	174
G.2.4. Mutational analysis.....	176
G.2.5. Inactivation/Inhibition trials.....	177
G.3. Interaction studies .....	178
G.3.1. TGF $\beta$ 2 Maturation .....	178
G.3.2. Protein-protein interaction analysis.....	179
G.3.3. Crystal structure of mature TGF $\beta$ 2.....	179
<b>CONCLUDING REMARKS</b> .....	183
<b>GENERAL REFERENCES</b> .....	189

## INDEX OF FIGURES AND TABLES

---

<b>Fig. 1.</b> Elements to be considered for the development of a mammalian expression system.....	6
<b>Fig. 2.</b> Mammalian cell line classification .....	7
<b>Fig. 3.</b> Schematic representation of two different transfections .....	9
<b>Fig. 4.</b> Schematic design of a mammalian's expression vector.....	11
<b>Fig. 5.</b> pCMV-Sport6: expression vector map.....	12
<b>Fig. 6.</b> Schematic view of the glycosylation pathway in mammalian cells .....	14
<b>Fig. 7.</b> Representation of the N-glycans attached to proteins .....	14
<b>Fig.8.</b> Schematic representation of an enzyme-substrate complex and subsite nomenclature .....	16
<b>Fig. 9.</b> Classification of proteases based on the chemical groups involved in catalysis.....	17
<b>Fig. 10.</b> Proteolytic activity regulation through zymogenicity .....	20
<b>Fig. 11.</b> Proteolytic activity regulation through reversible and irreversible inhibitors.....	22
<b>Fig. 12.</b> Shear forces in vessels .....	26
<b>Fig. 13.</b> Overview of primary haemostasis .....	27
<b>Fig. 14.</b> Synthesis of VWF in endothelial cells.....	29
<b>Fig. 15.</b> VWF domain structure .....	30
<b>Fig. 16.</b> Shear-induced unfolding of VWF and cleavage through ADAMTS13.....	31
<b>Fig. 17.</b> Domain structure of the ADAM and ADAMTS protease family as well as AD13.....	32
<b>Fig. 18.</b> Post-translational modifications of AD13 .....	34
<b>Fig. 19.</b> Structure of the AD13 metalloprotease (M) domain.....	35
<b>Fig. 20.</b> Crystal structure of the 3H9 Fab-MD'TCS (MP to Spacer) complex .....	36
<b>Fig. 21.</b> AD13-mediated cleavage of VWF .....	39
<b>Fig. 22.</b> Conformational activation of AD13.....	40
<b>Fig. 23.</b> The pathophysiology of thrombotic thrombocytopenic purpura (TTP) .....	43
<b>Fig. 24.</b> Schematic representation of A) healthy and B) damaged villi .....	45



<b>Fig. 25.</b> Key steps in CD pathogenesis .....	47
<b>Fig. 26.</b> Schematic representation of the wheat gluten composition .....	48
<b>Fig. 27.</b> Fragment of the $\alpha$ -gliadin protein sequence .....	51
<b>Fig. 28.</b> <i>Nepenthes ventrata</i> : carnivorous pitcher plant.....	54
<b>Fig. 29.</b> Cleavage analysis of endogenous neprosin expressed as iceLogo.....	56
<b>Fig. 30.</b> Multistep process enabling the release of active TGF $\beta$ .....	60
<b>Fig. 31.</b> Schematic representation of a conformational change of 63	
$\alpha_2$ -Macroglobulin.....	63
<b>Fig. 32.</b> Schematic representation of the interaction of $\alpha_2$ M with the	
LRP1 receptor .....	64
<b>Fig. 33.</b> Quantitative intramolecular crosslinking using DSA .....	166
<b>Fig. 34.</b> Ligand-induce conformational dynamics of AD13-MDTCS E <sup>225</sup> Q.....	168
<b>Fig. 35.</b> Surface electrostatic potential map of active site at different	
environmental pH.....	175
<b>Fig. 36.</b> Close up view of the pro-neprosin active site cleft and the bound	
prodomain in a standard orientation.....	176
<b>Table 1.</b> Characteristics of gastric, intestinal, and pancreatic proteases.....	49

## ABBREVIATIONS

---

<b>Å</b>	Ångstrom
<b>aa</b>	Amino acid
<b>AEBSF</b>	4-(2-aminoethyl)-benzenesulfonyl fluoride
<b>ADAMTS13</b>	<u>A</u> <u>D</u> isintegrin <u>A</u> nd <u>M</u> etalloproteinase with a <u>T</u> hrombo <u>S</u> pondin type 1 motif, member <u>13</u> )
<b>α<sub>2</sub>M</b>	Alpha 2-Macroglobulin
<b>AMP</b>	2-acetyl-1-methylpyrrole
<b>AGP</b>	Aspergiloglutamic peptidase
<b>AN-PEP</b>	<i>Aspergillus Niger</i> prolyl peptidase
<b>AP</b>	Aminopeptidase
<b>APC</b>	Antigen presenting cells
<b>BHK</b>	Baby hamster kidney cell
<b>BEOPC</b>	( <i>S</i> )-tert-butyl-2-(3-ethoxy-3-oxopropanoyl)pyrrolidine-1-carboxylate
<b>BMPs</b>	Bone morphogenetic proteins
<b>BGP</b>	<i>N</i> -boc-glycylproline
<b>BSA</b>	Bovine serum albumin
<b>BS<sup>3</sup></b>	Bis(sulfosuccinimidy) suberate
<b>C-</b>	Carboxy
<b>C</b>	Cysteine rich domain
<b>CAP</b>	CEVEC's Amniocyte Production
<b>CD</b>	Coeliac disease
<b>CHO</b>	Chinese human hamster ovary cells
<b>CK</b>	Cysteine knot

<b>CLMS</b>	Chemical crosslinking mass spectrometry
<b>CMV</b>	Citomegalovirus
<b>CP</b>	Carboxipeptidase
<b>CUB</b>	Complement C1r/C1s, Uegf, Bmp1
<b>D</b>	Disintegrin-like domain
<b>DAN</b>	Methyl-2-[(2-diazoacetyl)amino]hexanoate
<b>DSA</b>	Disuccinimidyl adipate
<b>DSG</b>	Disuccinimidyl glutarate
<b>DTT</b>	1,4-Dithiothreitol
<b>DUF</b>	Domain of unknown function
<b>EC</b>	Enzyme commission
<b>ECM</b>	Extracellular matrix
<b>EDC</b>	1-ethyl-3-[3-dimethylaminopropyl]carbodiimida hidrocloreuro
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>EGF</b>	Epidermal growth factor
<b>EOM</b>	Ensemble Optimization Methods
<b>ENPN</b>	2-[(4-nitrophenoxy)methyl]oxirane
<b>ER</b>	Endoplasmatic reticulum
<b>Fab</b>	Monoclonal antibody antigen-binding fragment
<b>FDA</b>	US-Food and Drug Administration
<b>Fuc</b>	Fucose
<b>Gal</b>	Galactose
<b>GF</b>	Growth factor
<b>GI</b>	Gastrointestinal
<b>GIP</b>	Gluten immunogenic peptides
<b>XX</b>	

<b>Glc</b>	Glucose
<b>Gp</b>	Glycoprotein
<b>HDX MS</b>	Hydrogen/Deuterium exchange mass spectrometry
<b>HEK 293</b>	Human embryonic kidney 293 cells
<b>HLA</b>	Human leukocyte antigen
<b>IC<sub>50</sub></b>	Half-maximal inhibitory concentration
<b>IEC</b>	Ion exchange chromatography
<b>IEL</b>	Intraepithelial lymphocyte
<b>Ig<math>\kappa</math></b>	Immunoglobulin kappa
<b>INF-<math>\gamma</math></b>	Interferon gamma
<b>IL</b>	Interleukin
<b>IMAC</b>	Immobilized metal affinity chromatography
<b>IPTG</b>	Isopropyl- $\beta$ -D thiogalactopyranoside
<b>k<sub>a</sub></b>	Association rate constant
<b>k<sub>d</sub></b>	Dissociation rate constant
<b>K<sub>D</sub></b>	Equilibrium dissociation constant
<b>K<sub>M</sub></b>	Michaelis-Menten constant
<b>K<sub>cat</sub></b>	Catalytic constant
<b>LAP</b>	Latency-associated peptide
<b>LRP</b>	Lipoprotein receptor-related protein
<b>LTBP</b>	Latent TGF $\beta$ binding protein
<b>M</b>	Metalloprotease domain
<b>Mabs</b>	Monoclonal antibodies
<b>Man</b>	Mannose
<b>MALDI-TOF</b>	Matrix-assisted laser desorption/ionization time of flight

<b>MALLS</b>	Multiple-angle laser light scattering
<b>Mca</b>	7 - Methoxycoumarin - 4 - acetic acid
<b>MCS</b>	Multicloning site
<b>MHC</b>	Major histocompatibility complex
<b>MMP</b>	Matrix Metalloproteinase
<b>MW</b>	Molecular weight
<b>N-</b>	Amino
<b>Ni-NTA</b>	Nickel-nitrilotriacetic acid
<b>NB-DNJ</b>	N-butyldeoxynojirimycin
<b>NS0</b>	Nonsecreting murine myeloma cells
<b>LLC</b>	Large latent complex
<b>Neu5Gc</b>	Glycosylneuraminic acid
<b>O/N</b>	Overnight
<b>PCR</b>	Polymer chain reaction
<b>PEI</b>	Polyethylenimine
<b>PEG</b>	Polyethylene glycol
<b>PEP</b>	Prolyl endopeptidase
<b>PDB</b>	Protein Data Bank
<b>PD</b>	Prodomain
<b>PMF</b>	Peptide mass fingerprint
<b>PMSF</b>	Phenylmethylsulfonyl fluoride
<b>Poly A</b>	Polyadenylation
<b>PT</b>	Post-translational
<b>SeMet</b>	Seleno-methionine
<b>SMIs</b>	Small molecule inhibitors

<b>R<sub>g</sub></b>	Radius of gyration
<b>R<sub>max</sub></b>	Maximum response
<b>Rmsd</b>	Root-mean-square distance
<b>r-protein</b>	Recombinant protein
<b>RT</b>	Room temperature
<b>RU</b>	Response units
<b>S</b>	Spacer domain
<b>SAD</b>	Single-wavelength anomalous diffraction
<b>SAXS</b>	Small-angle X-ray scattering
<b>SCP-B</b>	Scytalidoglutamic peptidase B
<b>SCX</b>	Strong cationic exchange
<b>SDS</b>	Sodium dodecylsulfate
<b>SEC</b>	Size-exclusion chromatography
<b>SLC</b>	Small latent complex
<b>SP</b>	Signal peptide
<b>SPR</b>	Surface plasmon resonance
<b>T</b>	Thrombospondin type 1 repeat
<b>TCA</b>	Trichloroacetyl acid
<b>TCEP</b>	Tris (2-carboxyethyl) phosphine
<b>TEV</b>	Tobacco etch virus
<b>TG2</b>	Tissue transglutaminase 2
<b>TGFβ</b>	Transforming growth factor beta
<b>TIMP</b>	Tissue inhibitors of matrix metalloproteinase
<b>T<sub>m</sub></b>	Melting temperature or temperature of midtransition
<b>TNF-α</b>	Tumor necrosis factor-alpha

<b>TSA</b>	Thermal shift assay
<b>TSP</b>	Thrombospondin domain
<b>TTP</b>	Thrombotic thrombocytopenic purpura
<b>UL-VWF</b>	Ultra large multimers of VWF
<b>UP</b>	Uniprot
<b>VWF</b>	Von Willebrand factor
<b>WB</b>	Western blot
<b>WT</b>	Wild type
$\lambda_{em}$	Emission wavelength
$\lambda_{ex}$	Excitation wavelength

The twenty physiological and proteinogenic amino acids with respective three-letter and one letter codes

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K

Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V





# **INTRODUCCION**

---



# 1. Recent advances in the production of recombinant proteins in mammalian cells for structural biology.

---

## 1.1 Production of recombinant therapeutic proteins

Since the early 1980s, over 250 therapeutic proteins including several notable monoclonal antibodies (mabs) and peptides have been approved for clinical use by the US-Food and Drug Administration (FDA) ([www.fda.gov](http://www.fda.gov)). With the introduction of human insulin (Goeddel et al. 1979), the first recombinant therapeutic protein, a major new class of therapeutics was introduced. Today, almost 380-marketed pharmaceutical products of this kind exist.

Proteins functioning as therapeutics have gained large scientific attention because they are part of many receptors and membrane channels, which contribute to molecular transport within the body. Despite their involvement in several biochemical processes, proteins also promote high specificity with the molecular target (Leader, Baca, and Golan 2008). Conversely, peptides have also been used extensively as therapeutics in several areas, including cancer, diabetes, infectious diseases and many other disorders, which have long been investigated for therapeutics (Vlieghe et al. 2010; Boohaker et al. 2012).

In the pharmaceutical industry, there is often a requirement to produce proteins on a large scale in order to support structural chemistry approaches, high-throughput screenings, and target validations. To produce large quantities of protein in a soluble form, it is usually necessary to screen multiple expression constructs in parallel, for example, protein homologues, mutants or truncated versions of the protein. To speed up the process, this has led to a move towards high-throughput protein expression strategies used in a range of expression systems that can produce milligram quantities of protein in a short time (Aricescu et al. 2006).

This thesis focuses on the practical aspects of developing a strategy for the rapid, economical production of proteins in a mammalian cell line, not only in a small-scale screening of a large number of constructs but allowing trustworthy upscaling for milligram-quantity production for functional, biophysical and structural experiments. Additionally, the mammalian cell line system permits protein labelling with seleno-methionine (SeMet) and provides control over the N-linked glycosylation levels. The system is suitable for use in

conventional molecular biology laboratories and can be easily implemented in a medium- or high-performance pipeline.

## 1.2 Expression system selection

The production of genetically engineered, recombinant proteins (r-proteins) on a large scale is an essential process for preclinical, biophysical, biochemical, structural and drug discovery studies used by several scientific groups and biopharmaceutical companies. The selection of the expression system depends on the type of protein, the requirements for functional activity and the yield required. Each system has advantages and challenges, and choosing the right one for a specific application is important for successful protein expression.

Several expression systems have been explored to produce high quantities of proteins such as *Escherichia coli*, baker's yeast, baculovirus-mediated insect cell culture, and mammalian expression systems, among others.

The bacterial system remains the major workhorse in the laboratories for the simple low-cost expression and quick generation of large quantities of recombinant protein. However, a large number of biomedical proteins suffer from poor solubility or lack of posttranslational modification associated with a prokaryotic expression hosts (Hannig and Makrides 1998; Baneyx and Mujacic 2004). Another limitation resulting from *E. coli* is its recurrent inability to produce correct disulphide bonds and the presence of endotoxins (lipopolysaccharide). Currently, the use of mutated *E. coli* strains is an alternative strategy to promote disulphide bond formation (Huang et al. 2017).

Yeast hosts, specially *Pichia pastoris* and *Saccharomyces cerevisiae*, provide an alternative to bacterial systems owing to the capacity to introduce particular post-translational modifications. However, due to technical expression requirements, yeasts are not popular in many laboratories (Boettner et al. 2002; Holz et al. 2003).

Baculovirus-insect cell expression systems represent an efficient and fast system for producing complex recombinant proteins and vaccines (Ikonomou, Schneider, and Agathos 2003; Cox 2012). Additionally, insect cells present several comparative advantages to mammalian cells, such as protein folding, modification and processing. Over the past 20 years, the baculovirus-insect cell system has become popular for the production of r-proteins, and several improvements have been implemented, including a cell culture scale-up technology, quantification methods, vector design, and commercially-available reagents. However, the main limitation of the baculovirus system is time, space consumption, and

effort required to produce a r-protein of interest. Moreover, the N-glycosylation pathway of insect cells is unable to produce the complex N-linked oligosaccharide side chains of mammals, which contain a galactose followed by a sialic acid residue (Hunt 2005; Kost, Condreay, and Jarvis 2005).

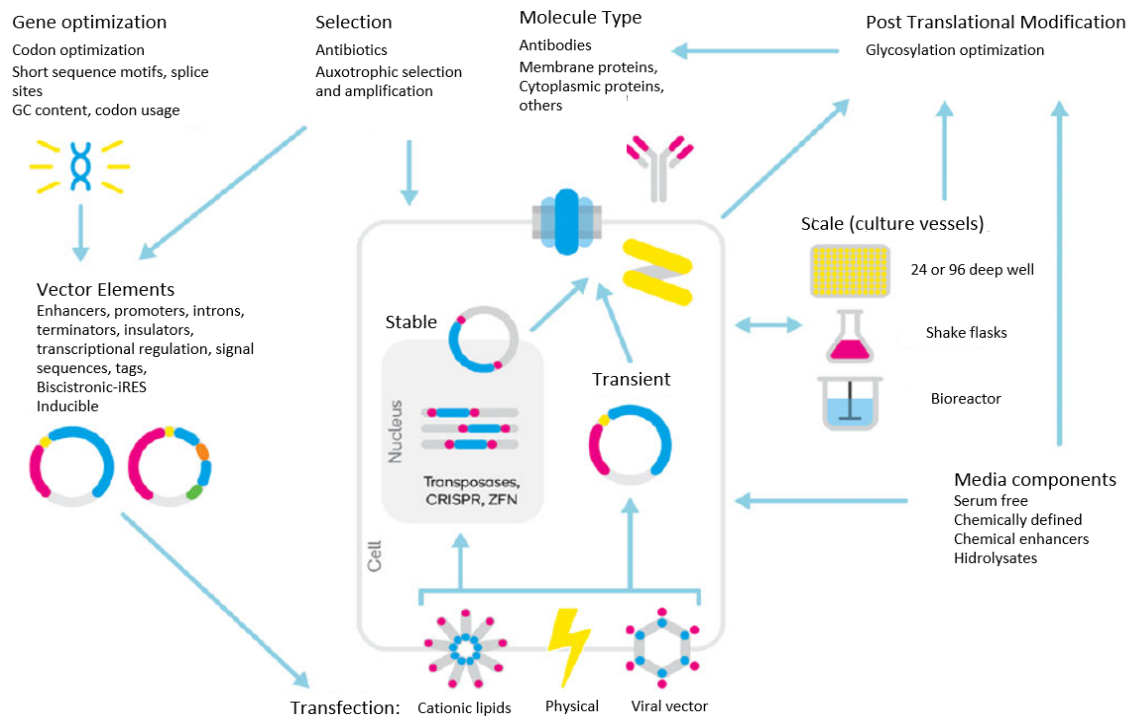
Alternatively, plant cells can be considered as an optimum system, as they are able to process eukaryotic proteins and show low production cost and a low risk for mammalian pathogenic contamination and other impurities (Pogue et al. 2010). Contrarily, plant cells may give rise to structural differences arising from different glycosylation patterns, which has important consequences for immunogenicity (Swiech, Picanço-Castro, and Covas 2012).

Furthermore, transgenic animals are an attractive option to produce low cost and high-quality therapeutic proteins. However, a high probability of product contamination with animal pathogens and the cumbersome separation of human r-proteins from their animal counterparts are major disadvantages. Compared to all these systems, the use of mammalian cells offers suitable solubility, correct folding, and completely processed r-protein (Estes and Melville 2014).

### **1.3 Development of a mammalian expression system**

The mammalian systems are gaining more importance than any other host and are used to produce approximately 70% of the commercially available therapeutic proteins so far. More than 60% of therapeutic proteins on the market are produced using mammalian cells, mainly based on the advantage to be able to produce complex glycosylated proteins that are similar to those naturally occurring in human cells, with respect to molecular structure and biochemical properties (Jianwei Zhu 2012). Additionally, in the case of mammalian cell lines and animal cell lines in general, most r-proteins can be secreted and do not require cell lysis to be extracted with subsequent protein refolding as is very often the case with bacterial hosts.

Advances in mammalian systems have focused on industrial processes, particularly the large-scale production of biotherapeutics. Despite several improvements, the selection process for the r-protein expression is not necessarily simple. Variables such as cell line, vector design, gene optimization, culture media, transfection reagents, and culture method impact the physical characteristics, yield and biologic activity of the proteins expressed, as depicted in Fig. 1.

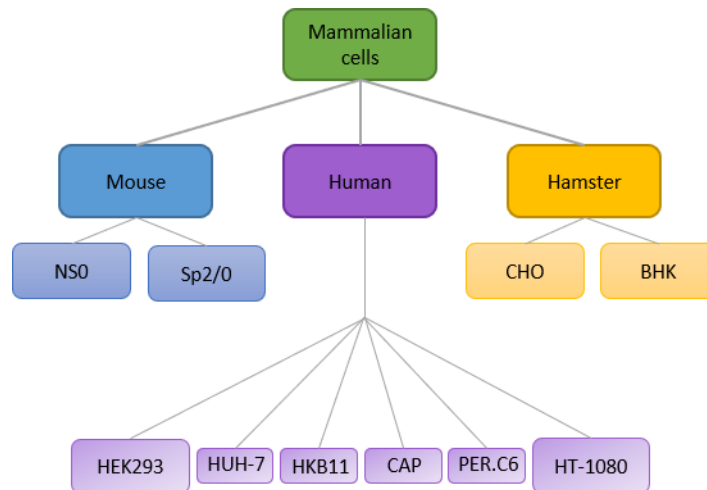


**Fig. 1. Elements to be considered for the development of a mammalian expression system:** gene optimization, vector design, molecule type to be expressed, transient or stable transfection, cell type, media selection, among others. Figure reproduced from (Hunter et al. 2019).

### 1.3.1 Cell line selection

A large number of mammalian cell lines have been explored for protein over-expression studies. The most commonly used mammalian hosts in research and industrial settings for r-protein production are the Chinese hamster ovary cells (CHO) and the human embryonic kidney (HEK) 293 cells (Swiech, Picanço-Castro, and Covas 2012; Estes and Melville 2014), although other mammalian cell types, such as murine myeloma and baby hamster kidney NS0, and SP2/0 cells have also been used (Butler and Spearman 2014). New cell lines such as PERC.C6 (D. Jones et al. 2003) from human retina, CAP and CAP-T from human amniocyte (S. Fischer et al. 2012), are reported to have remarkably high cellular productivities and generate human-type post-translational (PT) modifications (Fig. 2).

Researchers are continuously engineering new cell lines with desired features, high yields, specific PT-modifications or improved stable cell line selection. The accessibility of *omics* data has considerably improved our understanding of production cell lines, allowing improvements through cell line engineering instead of random mutagenesis (Xu et al. 2011).



**Fig. 2. Mammalian cell line classification.** Mouse, human and hamster lines. Figure reproduced from (Lalonde and Durocher 2017).

Despite the large variety of mammalian cell lines available for protein production, CHO cells are used for over 70% of therapeutic r-proteins, most of which are mabs (Jayapal, 2007). CHO and HEK-293 cells are maintained either as suspension cultures or adherent cultures, with suspension cultures showing an advantage regarding their ease to scaling up and handling. Both cell lines can generate and post-translationally modify eukaryotic proteins in functional forms at high levels. While HEK-293 cells are very easily transfectable and seem to be able to tolerate a wider variety of culture conditions during transfection, CHO cells are more difficult to transfect and present a narrower window of optimal conditions, thus requiring stable settings during scale up (C. Liu et al. 2008).

### 1.3.2 Glycosylation differences in mammalian cell lines

Glycosylation is the most complex posttranslational protein modification and is conducted by glycosyltransferases and glycosidases, which control the type and level of glycol-structures in the protein (A Varki 1998). Glycosylation may affect several crucial functions of the protein. Alterations in glycosylation patterns are associated with some human diseases, such as rheumatoid arthritis and various cancers (Gornik and Lauc 2008).

Most human proteins are glycosylated and contain one or more oligosaccharide chains (glycans) attached to the polypeptide backbone. Glycans are composed mainly of combinations of 7 monosaccharides: glucose (Glc), galactose (Gal), fucose (Fuc), mannose (Man), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc) and sialic acid or neuraminic acids (SA). Glycans may be divided into two general types, as defined by their



linkage with the protein: N-glycans, which bind to the amide nitrogen atom of asparagine (Asn); and O-glycans, which are attached to the oxygen of serine (Ser) or threonine (Thr) side chains (Bertozzi and Rabuka 2009).

Murine cells, specially CHO and BHK exhibit two important differences in the glycosylation pattern compared to human cell lines. First, glycosylation varies among and within each species and different species present differences in glycan processing, as well as in enzymatic specificities (Stanley, Raju, and Bhaumik 1996). For instance, glycosylneuraminic acid (Neu5Gc) is a sialic acid molecule, which is absent in humans because the CMAH gene is inactivated by a mutation. Second, human cell lines are not able to synthesize a terminal motif on N-glycans termed Gal $\alpha$ 1-3 Gal, so that they express antibodies against this molecule when inserted in r-proteins produce by murine cell lines (Galili 2004). The use of cultured human cells reduces the presence of immunogenic glycan structures in the r-proteins but not totally, Neu5Gc can be taken up and incorporated into secreted glycoproteins from animal products in the culture medium (Bardor et al. 2005).

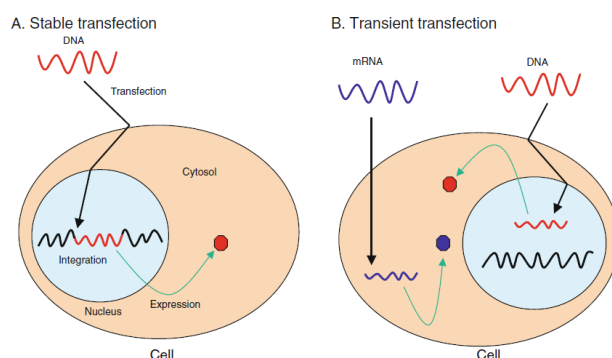
### **1.3.3 Transient vs Stable Systems**

The choice between stable and transient expression systems depends on the throughput, the quality and quantity of material needed, and the turnaround time. Both systems involve getting target DNA into cells. In transient gene expression, the vector does not integrate into the genome of the host cell, while in stable expression the DNA is inserted into the genome and maintained throughout many generations (Fig. 3).

Transient gene expression is preferred due to speed, lower cost and flexibility, thus transient transfection can be performed on a very small scale (96-well plates) but also large-volume cultures, such as performed in bioreactors (Tuveesson et al. 2008). Additionally, it is a convenient alternative when the protein may not be constitutively expressed by the cell due to toxicity issues (Nallet et al. 2009; Mignaqui et al. 2013).

For high-throughput protein expression, stable cell lines are preferred, because they provide high yields with consistent quality. However, transient expression is used as the first step to screen expression strategies in terms of construction design and molecular candidates before committing significant amounts of time and resources into stable cell line generation (Gutiérrez-Granados et al. 2018). The development of stable clones usually needs an extensive period (6-12 months) with the requirement of either a labour-intensive clonal selection process or complex and expensive laboratory equipment (Bandaranayake and Almo

2014). In contrast, transient transfection can be established in almost every cell culture laboratory, since it does not need particularly specialized equipment. Some concerns have been associated with transient gene transfection such as reproducibility, low titers or DNA and transfection reagents limiting for a large-scale application.



**Fig. 3. Schematic representation of two different transfections: (A) Stable transfection:** DNA is integrated into the host genome. **(B) Transient transfection:** Foreign DNA and RNA are delivered into nucleus and cytosol respectively, where they are translated. Figure reproduce from (Kim and Eberwine 2010).

### 1.3.4 Transfection

The insertion of the gene encoding the r-protein into a cell is required before the protein can be expressed. The choice of a transfection method depends on the amount of protein required, the cell type, and the cost to be assumed. The ideal method should have high transfection efficiency and reproducibility, and show low cell toxicity. For this discussion, a wide range of transfection procedures are available and classified as chemical, biological and physical methods (Kim and Eberwine 2010).

#### 1.3.4.1 Chemical transfection methods

Chemical transfection methods are the most widely used since their introduction in 1973 (Wurm 2004), being  $\text{CaCl}_2$  and DEA-dextran the two oldest reagents. Although they are very cost-effective, they suffer from low transfection efficiency and high toxicity to cells (Gluzman 1981). Today, cationic lipids such as TransFectin<sup>TM</sup> (Biorad), Lipofectamine (Thermofisher) or polyplexes like Polyethylenimine (PEI) are much more popular as they can be scaled up to thousands of litres. In this thesis, we optimized transient transfection in suspension Expi293F cells using 25 kDa PEI as reagent, which is inexpensive and enables to transfect a broad range of cell types, especially suspension-adapted HEK-293 and CHO cells (Gluzman 1981). Although PEI shows some disadvantages, such as low transfection

efficiency and protein yield, a systematic optimization of transfection parameters and ratios leads to successful high-level r-protein production.

#### **1.3.4.2 Physical transfection methods**

Physical transfection methods include direct injection, electroporation, biolistic particle delivery and laser-based transfection. The main disadvantage of these approaches is lack of scalability and high levels of cell death provoked by e.g. the high voltage pulses. Biolistic particles utilize gold particles conjugated with nucleic acids that are subsequently shot into cells. This method requires expensive instrumental equipment and causes cell damage.

#### **1.3.4.3 Biological transfection methods**

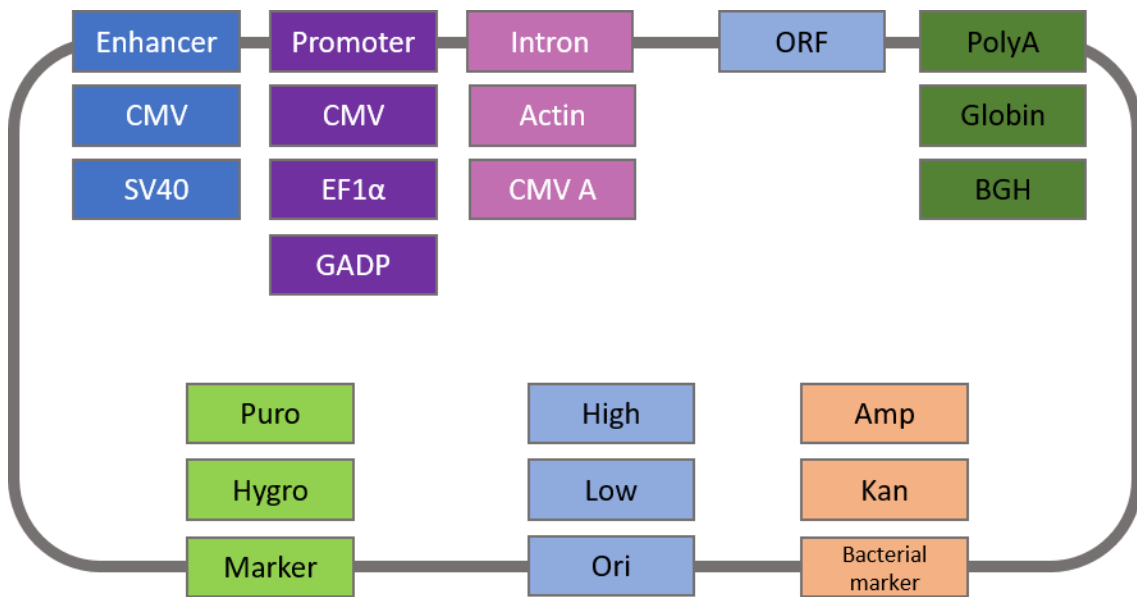
The most commonly used method in clinical research is a viral vector system which consists of a defective virus that encapsulates the gene of interest for transfer into target cells. This is the most efficient vehicle for DNA transfer. Another example is the use of the Baculovirus system, also named BacMan, which generates and amplifies virus particles in insect cells and is subsequently used for the infection of mammalian cells for target protein production. The primary disadvantage of this method is the large virus titer required to effectively infect mammalian cells, which is costly, takes a long time and is space consuming (Barsoum et al. 1997).

#### **1.3.5 Expression vector design**

The main criteria that were taken into account in choosing the plasmid backbone for the expression vector are based on the following essential elements: a promoter, multicloning site (MCS) and a polyadenylation (poly A) sequence (Kaufman 2000). Additionally, other elements are also included such as enhancers, viral amplifiers, selection markers, origins of replication, IRES elements (Fig. 4), amongst others.

Several main features have been revealed to contribute to an efficient expression vector: the plasmid should have a very high copy number in *E. coli* (for efficient DNA production), it must contain the strongest promoter available for mammalian cells, and, finally, it should be small enough to allow efficient cloning of constructs of variable length. In addition, introducing a polyA sequence at the 3' end of the RNA transcript promotes stability, efficient translation and nuclear export (Edmonds 2002). The most common mammalian terminators are SV40 Hgh, rbGlob, and BGH, which account for both termination and polyadenylation

(Gil and Proudfoot 1987). Furthermore, it has been shown that the insertion of an intron upstream from the initiation codon of the gene promote higher levels of protein expression (Lacy-Hulbert et al. 2001; Melcher, Grosch, and Hasilik 2002). To identify which combination of characteristics offers the best productivity for the gen of interest, a battery of vectors needs to be tested. While the number of potentially useful elements continues to increase, a systematic approach needs to be undertaken to optimize each element in combination with the others to maximize the vector’s properties.



**Fig. 4. Schematic design of a mammalian’s expression vector.** Figure reproduced from (Hunter et al. 2019).

### 1.3.6 Signal peptide selection

In addition to vector optimization, we must consider where the protein will be present, once expressed. The wide use of mammalian cells has been dominated by the production of secreted proteins that need a signal peptide for exports outside the cell. Intuitively, the best option of a signal peptide would be the protein’s own. However, testing a panel of commonly used signal sequences can have a dramatic effect on protein productivity, with up to fourfold larger expression levels (Knappskog et al. 2007; Kober, Zehe, and Bode 2013). Following these guidelines, the pCMV-pSport6 (Sp6) backbone was chosen (kindly provided by the Enghild Laboratory at Århus University in Denmark). Fig. 5 shows how the constructs have been generated.



methionine pools before the addition of Se-Met. Keeping the concentration of Se-Met at 60 mg/L or below is important due to its toxicity, which would lead to a lower yield of r-protein. Although some protocols report Se-Met incorporation in adherent mammalian cell culture with a 90% substitution level (Barton et al. 2006), there are no protocols available for mammalian culture in suspension.

## 1.5 Control of N-linked glycosylation

Unlike insect and fungi, mammalian cells synthesize complex N-glycans (Fig. 6), which may affect the function and ability of the r-proteins to form well diffracting crystals (Davis et al. 1993). Apart from increased protein solubility and the inherent flexibility given by sugar chains, it is likely that certain sites will only be occupied in a subpopulation of the molecules, thus introducing a source of heterogeneity in the protein sample, which may impair downstream biophysical studies.

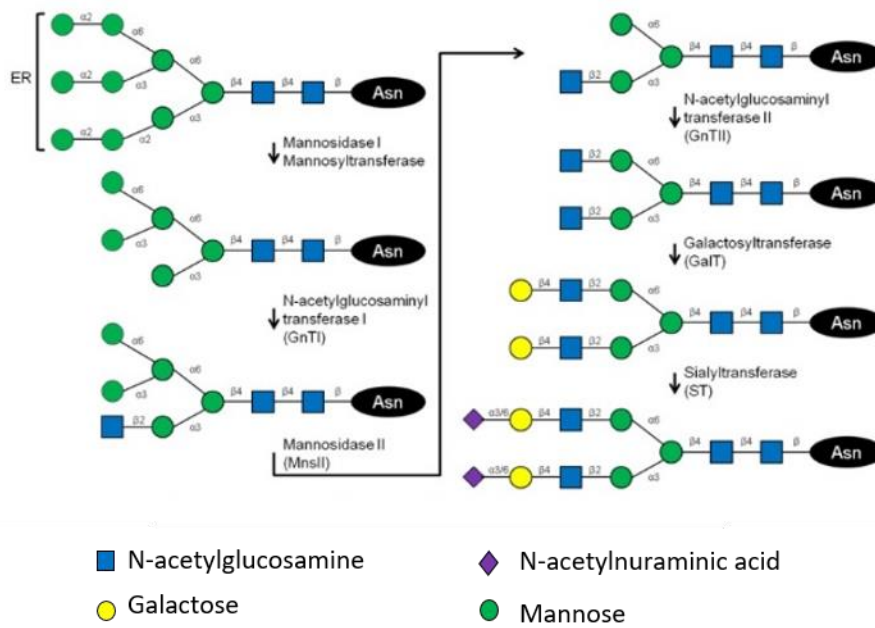
The presence of sugar is crucial for the correct folding of some r-proteins. For example, the treatment with N-glycosylation inhibitors such as tunicamycin provokes protein retention in the endoplasmic reticulum (ER) and subsequent degradation (Aricescu et al. 2006). There are currently only two approaches to tackle this problem, *viz.* the use of chemical inhibitors, which are based on manipulating N-glycosylation during glycoprotein biosynthesis by blocking the action of processing enzymes, and null mutant cell line development.

### 1.5.1 Inhibitors

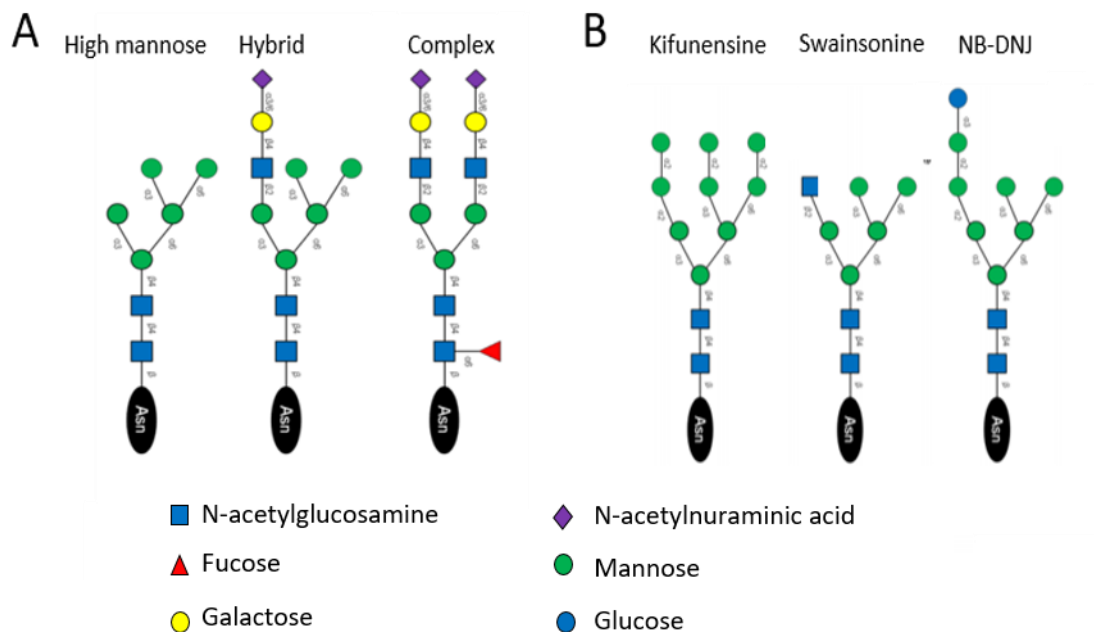
Three main inhibitors have been used in the production of glycoproteins to deal with the glycosylation pathway: N-butyldeoxynojirimycin (NB-DNJ), swainsonine and kifunensine. NB-DNJ inhibits  $\alpha$ -glucosidase, thus blocking the early stages of N-glycan processing and leading to products, which contain high-mannose or hybrid-type sugars (Fig. 7a). NB-DNJ has principally been used in combination with the Lec3.2.8.1 mutant CHO cell line. Swainsonine blocks  $\alpha$ -mannosidase II resulting in high-mannose or hybrid-type sugars while kifunensine strongly inhibits  $\alpha$ -mannosidase I activity, prompting sugars of the form  $\text{Man}_9\text{GlcNAc}_2$  (Fig. 7b; Yu et al. 2011).

In practice, kifunensine is the most widely used inhibitor in the production of glycoproteins for structural studies since the glycans generated are more homogeneous. Indeed, the structures of glycoproteins can be solved without the previous use of endoglycosidase when kifunesine is employed (for example, PDB entry 2WAH (Crispin et

al. 2009). However, often sugars need further to also trimmed with the Endo H glycosidase before crystallization studies (Bishop et al. 2009).



**Fig. 6. Schematic view of the N-glycosylation pathway in mammalian cells.** The initial precursor  $\text{Man}_9\text{GlcNAc}_2$  sugar is formed in the endoplasmic reticulum (ER) and further modifications are taken place in the Golgi apparatus. Figure adapted from (Reily et al. 2019).



**Fig. 7. Representation of the N-glycans attached to proteins.** (A) Diagrammatic representation of the major classes of N-linked sugars produced by mammalian cells (see also Fig. 6). (B) Representation of the N-glycans found when small molecule inhibitors are employed. Figure adapted from (Reily et al. 2019).

## 2. Proteolytic enzymes

---

Proteases are enzymes that catalyse the hydrolysis of the amide bond (HN-CO bond, also referred to as peptide bond) in proteins or peptide substrates. Although the terms “proteinase” and “peptidase” are used as synonyms for peptidase, it has to be mentioned that peptide bonds can also be cleaved by other mechanisms and do not necessarily involve hydrolysis (Neil David Rawlings, Barrett, and Bateman 2011). The terms proteases and peptidases address differences in the substrate length (Grassmann and Dyckerhoff, 1928) and will be referred to in this work collectively under the name of proteases.

Proteases play diverse biological roles, ranging from digestion to tissue remodelling during development and adult physiology, from signal transduction to homeostasis, proliferation and programmed cell death (apoptosis), among others. (Beynon and Bond, 2001; Neurath and Walsh 1976). Consistently, proteolytic imbalance may trigger the malfunction of individual cells or whole organisms, thereby causing certain pathological states such as inflammation, cardiovascular and neurodegenerative diseases, cancer or cell death (López-Otín and Bond 2008; Yadav et al. 2014). In this context, the mechanisms that regulate protease expression and activity must be spatially and temporally strictly regulated.

Hydrolysis of peptide bonds by proteases is not a random process. Most proteases are relatively nonspecific and may target multiple substrates, which results in massive degradation (extensive proteolysis) as observed during digestion and processing of misfolded proteins. This is the case of subtilisin, a bacterial protease, which cleaves substrates predominantly—but not uniquely—to produce assimilable peptides. Some proteases have stricter cleavage specificity and are responsible for limited proteolysis, which results in their own activation or inactivation, as well as in the modification of substrates in order to affect downstream processes, such as signalling pathways or protein-protein interactions (Cerdà-Costa and Gomis-Rüth 2014; López-Otín and Bond 2008). For instance, TEV protease recognises a specific target of ENLYFQ-S sequence for cleavage.

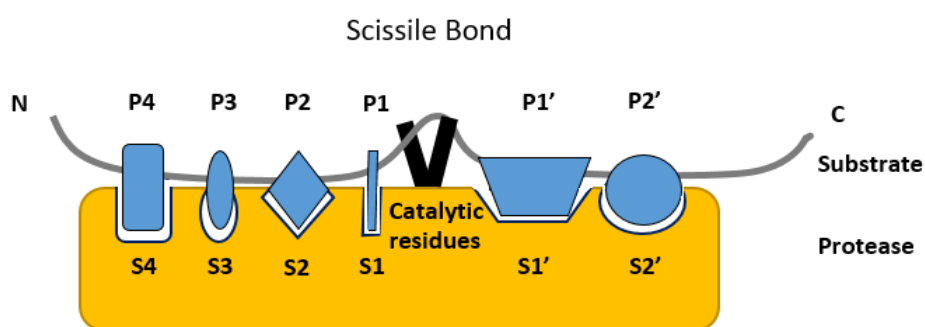
Although the simplest protease in its mature form is constituted by a catalytic domain, most of them are multidomain proteins composed by different modules or multiples copies of similar or identical domains. The catalytic domain of a protease is then linked to other domains, which may play dedicated roles in the activity regulation, substrate specificity, kinetic properties, activation abolishment, and sensitivity towards inhibitors.



## 2.1 Active site

It has been shown that the active site is generally localised in a cleft on the molecule surface between adjacent structural (sub)domains. The active site of an enzyme holds two main functions: (i) substrate binding and (ii) catalysis. The efficiency of these two actions determines the overall activity (also referred to as specificity) of the enzyme towards a particular substrate and can be greatly enhanced by exosites and allosteric sites (Drag and Salvesen 2010).

The general nomenclature of the active site cleft and their respective subsite position was postulated by Schechter and Berger in 1967. These sites are numbered starting from the cleavage site between the  $S_1$  and  $S_1'$  positions as  $S_1 \dots S_n$  towards the N-terminus of the substrate and as  $S_1' \dots S_n'$  towards the C-terminus. The positions of the substrate match the cognate sites of the enzyme and are labelled  $P_1 \dots P_n$  and  $P_1' \dots P_n'$ , respectively (Fig. 8).



**Fig. 8. Schematic representation of an enzyme-substrate complex and subsite nomenclature.** Figure adapted from (Farady and Craik 2010).

## 2.2 Classification

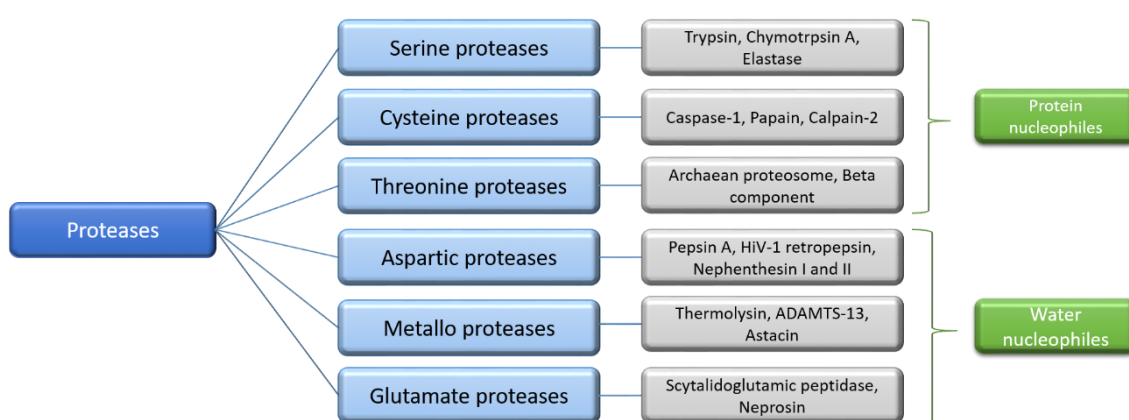
Proteases can be classified following different criteria but the most common are specificity and catalytic mechanism.

### 2.2.1 Specificity

Proteases can be divided into two groups, based on the reaction they catalyse. They are referred to as (i) exopeptidases when hydrolysis liberates amino- or carboxy-terminal amino acids or (ii) endopeptidases when cleavage occurs in an internal peptide bond. Exopeptidases are further assorted into aminopeptidases (APs) and carboxypeptidases (CPs) depending on whether they chop off N-terminal or C-terminal residues (Barett et al., 2013).

## 2.2.2 Catalytic mechanism

Proteases can also be classified based on the catalytic mechanism according to the residues contained in the active site: serine-, threonine-, cysteine-, aspartate-, metal-, and glutamate-dependent proteolytic enzymes (Fig. 9). For the first three classes, the nucleophile group that initiates the attack on the peptide bond is the hydroxyl or thiol group on the side chain of the catalytic serine/threonine and cysteine respectively (“protein nucleophiles”) (Rawlings and Barret, 2013). By contrast, the last three classes use a water molecule as a nucleophile for catalysis, which is activated either by the side chains of catalytic aspartates/ glutamates or by one or two metal ions (Fujinaga et al. 2004).



**Fig. 9. Classification of proteases based on the chemical groups involved in catalysis.** Diagram representing the six current determined classes of proteases and several representatives of each group.

This classification is still maintained in the recommendations from the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) and the International Union of Pure and Applied Chemistry (IUPAC). The main classification is the mechanism of action and substrate. Therefore, every enzyme performing a given reaction is assigned a unique EC (enzyme commission) number code prefixed by EC. Accordingly, proteases, which are a subclass within the hydrolases (EC 3 hydrolases; 3.4 subclasses), are further sorted into 19 sub-subclasses.

An overview of all proteases can be accessed as an online version on the following website <http://www.qmul.ac.uk/sbcs/iubmb/enzyme/EC3/4>. Despite being a complex system for protease classification, it opens up a recommended name for each peptidase as well as a list with alternative names that may be found in the literature referring to the same protease.

### 2.2.3 MEROPS: The peptidase database

In 1993, Rawlings and Barrett proposed a protease classification in terms of evolutionary relationship, which is currently deployed in the MEROPS database (<http://merops.sanger.ac.uk>). It consists of a sequence-based classification of proteases with homologous sequences, which are grouped into families and further into clans (Neil D Rawlings and Barrett 1993). A family is built up around a protease that has been characterized in detail at the biochemical level, which is termed as the “type example”. Families are assigned with a letter specifying the catalytic type (S, C, T, A, G, M, P, N, U or X for serine, cysteine, threonine, aspartic, glutamic, metallo, asparagine or unknown) followed by a number (Neil D Rawlings et al. 2014). There are currently over 270 protease families and 50 clans that have been described, comprising more than 900,000 sequences (Neil D Rawlings et al. 2018). The similarity in three dimensional structures, when data are available, is the best evidence to support the formation of a clan but the disposition of the catalytic residues in the polypeptide chains and the similarities in amino-acid sequence around the catalytic domain are also taken into account. Each clan has a two-letter identifier, with the first letter denoting the catalytic type of the grouped families followed by an arbitrary second capital letter.

Additionally, protein inhibitors have been included in the database since 2004. Around 2% of the genes in a genome encode protease homologues, and 1% encode protein inhibitors. However, there are many other inhibitors, including peptides and synthetic inhibitors, which are referred to as small molecule inhibitors (SMIs). Information about these SMIs is listed alphabetically and gathered in the MEROPS database. New tools in the MEROPS database include displays to present peptidase specificity when known, tables of peptidase-inhibitor interactions and dynamical alignments of representatives of each protein species at the family level.

The MEROPS database not only provides a classification system but also acts as a catalogue, which gives access to a summary page with detailed information on a single peptidase, family and clan, as well as the classification and nomenclature of each proteases. In addition, it provides a supplementary page with sequence identifiers, three-dimensional structures, literature references, among others. Henceforth, the database functions as an organizational framework around which a variety of related information is compiled (Neil D Rawlings, Barrett, and Bateman 2014).

## **2.3 Biological relevance**

Proteases are distributed across all kingdoms of living organisms and are involved in a wide variety of biological processes, such as DNA replication and transcription, cell-cycle regulation, cell proliferation, differentiation and migration, bacterial cell wall biosynthesis, viral polyprotein processing, metabolism of antibiotics, embryonic development, angiogenesis, morphogenesis, tissue remodelling, neuronal growth, alimentary uptake, bone formation, immune and inflammatory responses, homeostasis, blood coagulation and fibrinolysis, and cell death (apoptosis) (Turk 2006; López-Otín and Bond 2008). Therefore, any modification in the proteolytic activity of a protease may contribute to a range of pathologies from arthritis, fibrosis and osteoporosis to autoimmune, cardiovascular and neurological diseases. However, protease activity is not only focused on cell-internal processes, it has also an important exogenous function. Proteases can be secreted as microbial virulence factors that contribute directly or indirectly to toxicity in the host, e.g. by inactivating key proteins such as the complement system, chemo attractants, and structural components of the extracellular matrix (e.g. collagen and elastin). These strategies are required for cell invasion, proliferation, survival and colonization of pathogenic bacteria (J.-W. Wu and Chen 2011; Jusko et al. 2012; Koziel and Potempa 2013). Additionally, some proteases are also found in the venoms of poisonous predators, such as snakes, scorpions and spiders, or, contrarily, as a defence mechanism against predators, such as in scorpion fish (Carrijo et al. 2005; Fox and Serrano 2009; Rusmili et al. 2014).

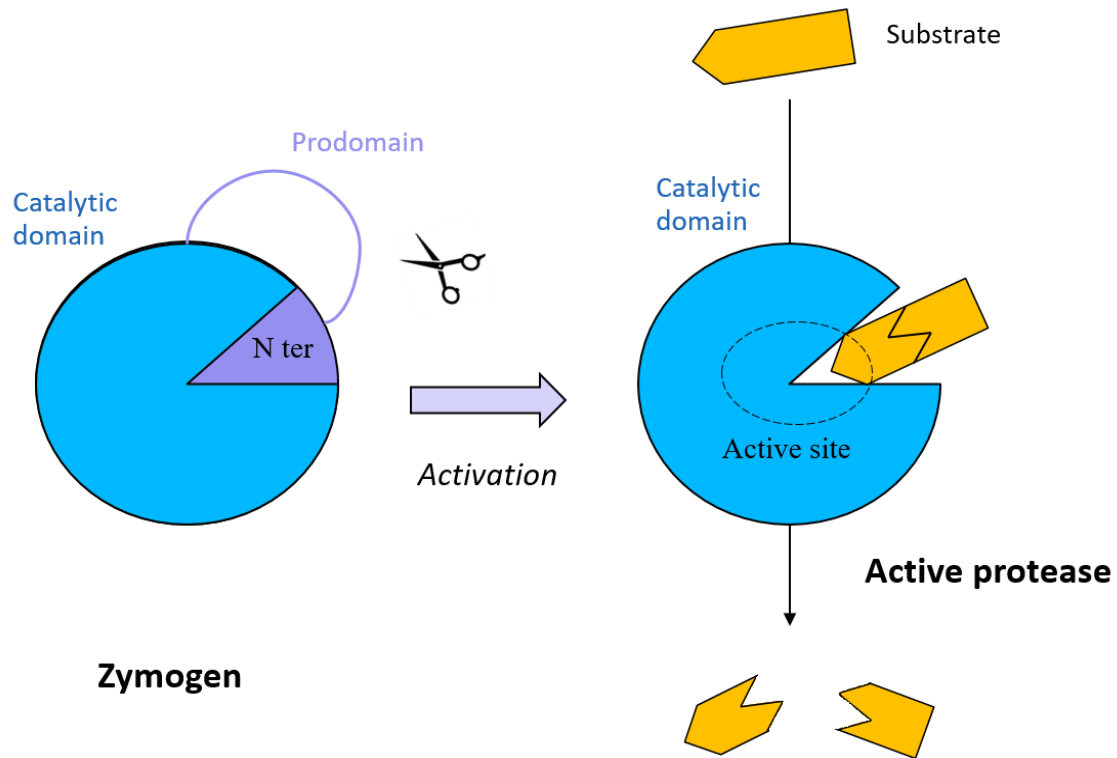
## **2.4 Proteolytic regulation**

Since proteases catalyse an essentially irreversible cleavage reaction, a wide variety of regulatory mechanisms are required to control and restrict proteolysis to act independently or in a coordinated manner in order to avoid aberrant activity.

### **2.4.1 Zymogenicity**

Zymogenicity is one of the most ubiquitous mechanisms of proteolytic regulation. It consists in the secretion of a protease as an inactive precursor or zymogen, which is activated in the right biological context. In many cases, N-terminal prosegments or prodomains (PDs) are responsible of the zymogenicity, which act by preventing access of substrates to the active site of the enzyme (Fig. 10) (Khan and James 1998). The activation is then achieved through PD removal by proteolytic cleavage, which may be either autolytic (i.e. self-cleavage) or

heterolytic (i.e. catalysed by other enzymes) and the process is denominated maturation. Additionally, some PDs act as an intramolecular chaperone and assist in the folding process and protein sorting for the downstream mature part of the enzyme (Eder and Fersht 1995; O'Donohue and Beaumont 1996; Bryan 2002).



**Fig. 10. Proteolytic activity regulation through zymogenicity.** Schematic representation of the prodomain removal from the active site through an autolytic or heterolytic cleavage (represented by scissors), which results in an active protease. Figure adapted from (Deu, Verdoes, and Bogyo 2012).

Usually, the activation may be triggered by specific environmental conditions, which determine the fate of the protease. As an example, the acidic pH of the stomach activates the digestive proenzyme pepsinogen to mature pepsin. Trypsinogen, in turn, is activated to yield trypsin by another protease, chymotrypsin. Occasionally, certain PDs are not involved in folding and/or extracellular expression of the protease. An example is ADAMTS13, a metallopeptidase that is secreted as an active enzyme to the plasma that is regulated by domains adjacent to the catalytic domain. ADAMTS13 is the only known protease that has CUB domains at the distal C-terminus, which have a negative regulatory function of enzyme activity (South et al. 2014).

### 2.4.2 Post-translational modifications

Several post-translational modifications such as acetylation, phosphorylation, glycosylation, ubiquitination and oxidation may also modulate the proteolytic activity. For instance, oxidation modifies the thiol group of a cysteine involved in chelating the catalytic zinc ion in the zymogen of some metallopeptidases, thereby promoting their activation, as found in some matrix metallopeptidases (MMPs). Some publications also have reported the effect of site-directed mutagenesis on N-glycosylation, which affects the expression of e.g. the metallopeptidase leishmanolysin (McGwire and Chang 1996). Moreover, oligomerization is another posttranslational modification, which can further influence protease activation, inhibition or substrate binding. This is found in some caspases, which require homodimer formation to whip up intramolecular activation (Pop and Salvesen 2009).

### 2.4.3 Protease Inhibitors

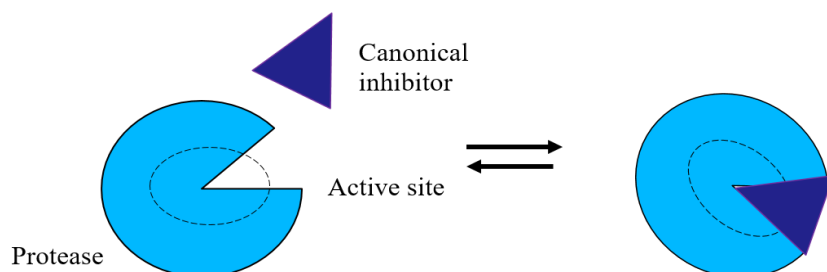
Protease inhibitors are major regulators of proteolysis by blocking access to the catalytic protease domains. Protein inhibitors may be either proteins, which are encoded in the genomes of most organisms, or small molecules resulting from the metabolism of some bacteria, fungi and plants (e.g. chymostatin, leupeptin, pepstatin A, antipain, phosphoramidon and bestatin) (Bode and Huber 2000; Neil D Rawlings 2010; Sabotič and Kos 2012).

Protease inhibitors can be classified based on their physiological role, mechanism of action, and evolutionary relationship. Inhibitors are divided into regulatory inhibitors, when they are tightly expressed and colocalise with specific target protease, and emergency inhibitors, when expressed in high levels to suppress the action of threatening proteases. Another way to classify protease inhibitors reflects their evolutionary relationship. Depending on the number of inhibitory units, proteins are referred to as single or compound inhibitors.

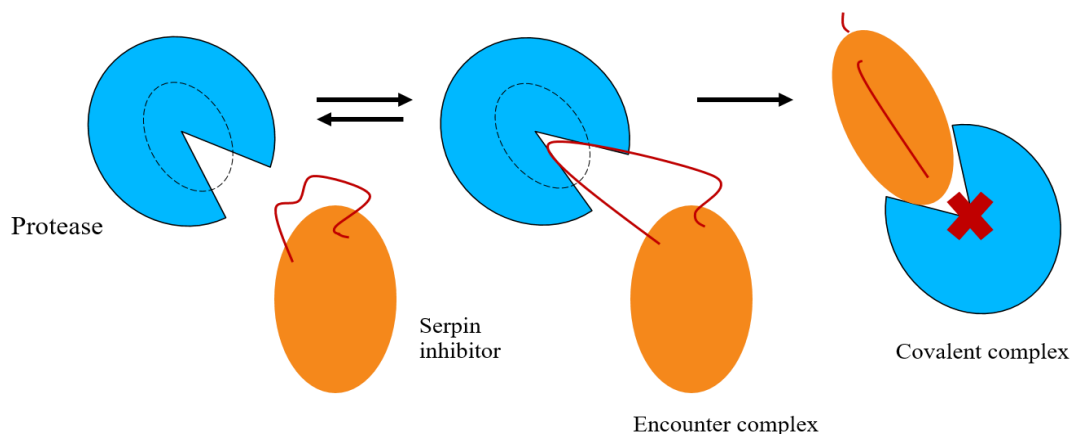
The number of inhibitor proteases identified is remarkably large, and their respective three dimensional structures are available in the Protein Data Bank, as linked to in the MEROPS database (<http://merops.sanger.ac.uk/inhibitor/>) where they are grouped in families and clans in a similar way as proteases (Neil D Rawlings, Tolle, and Barrett 2004). The mechanisms by which inhibitors modulate the proteolytic activity are diverse, ranging from interaction with catalytic residues to deadlock of the access to the catalytic domain to

irreversible entrapment of the protein (Otlewski et al. 2005; Farady and Craik 2010). Thus, inhibitors can be also divided in reversible and irreversible (Fig. 11).

### A. Reversible inhibition



### B. Irreversible inhibition



**Fig. 11. Proteolytic activity regulation through reversible and irreversible inhibitors. (A)** Schematic representation of a canonical inhibitor, which blocks the active site of the protease and prevents substrate access in a reversible manner. **(B)** Schematic representation of the serpin inhibitor, which binds the protease through a reactive loop resulting in a conformational change disrupting the active site of the protease irreversibly. Figure adapted from (Beinrohr et al. 2008).

#### 2.4.3.1 Reversible inhibitors

The vast majority of protease inhibitors are reversible and act by blocking substrate-binding sites catalytic site of the enzyme, thus preventing substrate cleavage. The inhibitor-protease complex formed can be dissociated under some circumstances (Bode et al. 1992). Hence this type of inhibitors is also termed “competitive inhibitors”. Reversible inhibitors can be classified by their mechanism of action into catalytic or non-catalytic competent inhibitors, also named canonical and noncanonical inhibitors, respectively (Farady and Craik 2010).

Canonical inhibitors follow a standard mechanism based on the insertion of a “reactive loop” into the active site mimicking the substrate of the target protease. In this context, the enzyme can hydrolyse, though very slowly, the scissile bond so that the amide bond can be eventually religated (Zakharova, Horvath, and Goldenberg 2009).

Remarkably, several inhibitors do not block catalytic residues but interact with subsites adjacent to the catalytic cleft in catalytically incompetent way. An example for this mechanism is provided by cystatins and tissue inhibitors of matrix metalloproteases (TIMPs). The former bind to papain-type cysteine proteases by inserting a wedge-like face, consisting of N-terminal residues of the inhibitor and two hairpins, into the catalytic site. In this context, the inhibitor occupies most of the active site but does not interact with the catalytic machinery (Bode and Huber 2000). TIMPs, in turn, exert inhibition via hairpin loops and N-terminal residues but also interfere with the catalytic zinc, causing subsequent displacement of the water molecule from the active site (Brew, Dinakarpanian, and Nagase 2000).

Moreover, some protease inhibitors target secondary binding sites outside the active-site cleft, the so-called exosites, which are crucial for the activity of certain proteases such as thrombin.

#### **2.4.3.2 Irreversible inhibitors**

Irreversible inhibitors are typically large molecular-weight proteins and, contrarily to the reversible inhibitors, their binding to the target protease is usually covalent, thus permanently modifying the protease and acting as “single use” inhibitor: They are also termed suicidal inhibitors (Farady and Craik 2010).

There are only two large families of trapping inhibitors, which differ in their mechanism of action. It can be site specific, as found in the family of the serine peptidases inhibitors dubbed serpins, or rely on broad-spectrum molecular-trap mechanisms, which inhibit proteases of several catalytic mechanisms by entrapping target proteases, as found for the  $\alpha_2$ -macroglobulins ( $\alpha_2$ Ms) and their relatives. Common to both types is a major conformational change that traps or disrupts the structure of the target protease.

The serpin family regulates key processes, such as coagulation, inflammation, fibrinolysis or immune response (Olson and Gettins 2011). A main structural characteristic of serpins is the “reactive centre loop” (RCL), which binds directly in a substrate-like manner to the active site of the protease. The proteolytic reaction proceeds only until formation of an acyl-enzyme complex with the N-terminal fragment of the inhibitor and release of the downstream part.



Next, the protease is pushed against the inhibitor, causing global disorder of its structure, which destroys the active site (Ye and Goldsmith 2001; Whisstock and Bottomley 2006).

By contrast, the reaction mechanism of  $\alpha_2$ M<sub>s</sub> mainly consist in conformational changes, which are triggered by protease cleavage in a highly exposed region denominated “bait region” (Sottrup-Jensen 1989). Once trapped, the proteases still hydrolyse small substrates that enter the cage (Marrero et al. 2012) but their action is limited, as protease-laden  $\alpha_2$ M is quickly cleared from the blood stream and subsequently degraded in the lysosomes (Kolodziej et al. 2002).

#### **2.4.4 Cofactor**

Cofactors ranging from ions to proteins are essential for the catalysis of certain enzymes by modulation of their final activity and stability. Ions like calcium have proven to be essential factors for several proteases (e.g. MMPs). However, the most well-known examples are in the blood coagulation pathway, where factor Va serves as a cofactor of factor Xa during the formation of the prothrombin complex, which leads to the formation of active thrombin and fibrin.

#### **2.4.5 Allosteric regulation**

Allosteric modulators (proteins or small molecules) may act outside the enzyme active site and are, thus, not involved in substrate recognition but induce a conformational change in the protease, which alter its function, either enhancing or inhibiting it. Allosteric regulators do not need to be chemically resembling the substrate as they do not compete for gaining access to the active site. For instance, glycosaminoglycans contribute to the activation of the cathepsin cysteine peptidases, by converting zymogens into better substrates. Moreover, antithrombin is a well-known allosteric inhibitor of blood coagulation proteases in addition to its serpin-type inhibitory function.

#### **2.4.6 Protein compartmentalisation and protease trafficking**

Compartmentalization and protease trafficking restrict proteolysis to specific subcellular compartments or to intracellular or extracellular localizations, respectively. This allows to localise proteases to the physiologically relevant cellular compartment in which catalytic activity occurs under optimal conditions. These conditions include calcium ions, specific pH values, and redox conditions ranging from oxidizing, as found in compartments of the

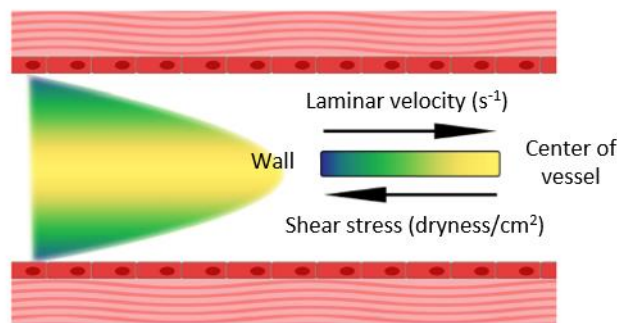
secretory pathway and the extracellular space, to reducing as in lysosomal compartments. For instance, cathepsins are involved in a variety of functions depending on their cellular locations: inside lysosomes, they contribute to protein turnover; inside the nucleus, they are involved in cell-cycle progression.

### 3. The intriguing interplay of ADAMTS13 and VWF

---

#### 3.1. Circulatory system and shear stress

Blood is a viscous fluid, which flows due to a pressure gradient generated by tangential forces at the vessel wall in the direction of the flow, the so-called “shear stress” (Hanson and Sakariassen 1998). The blood flow is laminar in a cylindrical vessel and therefore the velocity of the blood stream is highest in the middle of a blood vessel and decreases towards zero at the blood vessel wall. By contrast, the shear stress is highest at the vessel wall and in small vessels such as arterioles and capillaries (E. A. V Jones, le Noble, and Eichmann 2006) (Fig. 12).



**Fig. 12. Shear forces in vessels.** The fluid flow increases from the wall towards the centre of the vessel while the shear rate is highest at the wall.

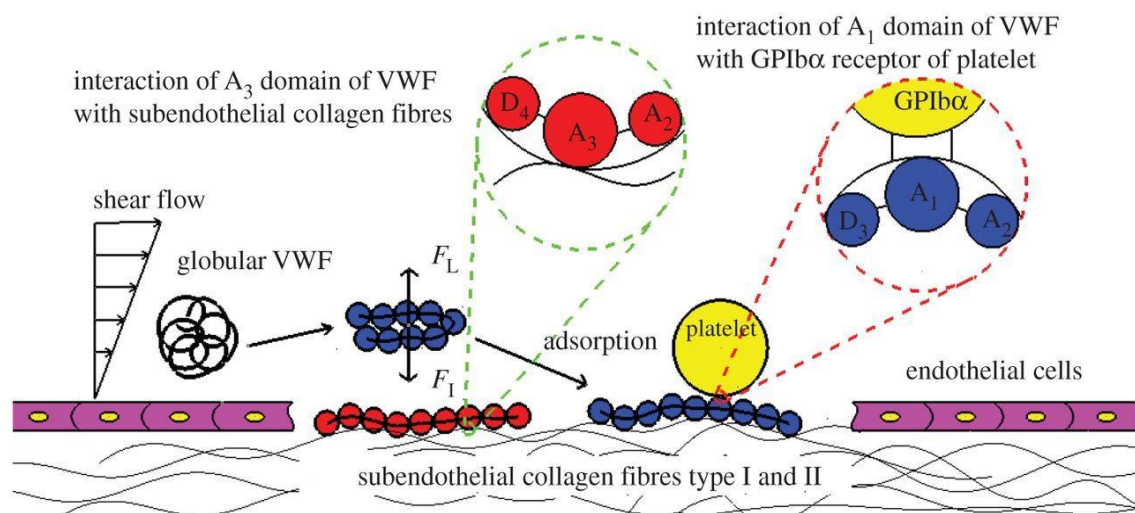
#### 3.2 Maintenance of Haemostasis

Haemostasis is a complex physiological process, which involves intricate interactions between the vessel wall, plasma proteins, and platelets to prevent blood loss after vascular injury. This process comprises three main events including initial platelet plug formation (primary haemostasis), stabilization of the platelet aggregates by a mesh of insoluble fibrin (secondary haemostasis), and limitation of the thrombus growth (fibrinolysis).

##### a. Primary haemostasis

The key players in primary haemostasis are the platelets, the multimeric glycoprotein von Willebrand factor (VWF), and collagen. Upon damage to the vascular endothelium, primary haemostasis begins with local contraction of the vasculature to reduce the blood flow at the site of damage. Circulating platelets exhibit multiple glycoproteins (Gp) on the surface. These are essential to recruit more platelets to the area around the injury in order to form a hermetic

platelet plug. VWF binds mainly to fibrillar collagen I and III using its A3 domain, while through its A1 domain it may interact partly with collagen types IV and VI. Shear forces in the blood induce conformational changes in the VWF, resulting in exhibition of the platelet glycoprotein binding site (Ib $\alpha$ ) on the A1 domain. Consequently, platelet activation also promotes platelet integrin  $\alpha$ II $\beta$ B3 interaction with fibrinogen to contribute to platelet aggregation. The ability of VWF to bind collagen, unfold and capture platelets is strongly dependent upon its multimeric size. Primary haemostasis must be regulated by the enzyme ADAMTS13 (A Disintegrin And Metalloproteinase with a Thrombospondin-type 1 motif, member 13) through proteolysis of the ultralarge forms of VWF, which have the highest clotting power, and thus preventing platelet accumulation to the site of vessel damage.



**Fig. 13. Overview of primary haemostasis.** The initially plasma and adsorbed VWF are shown as blue and red polymer chains, respectively. Each monomer (as drawn with circles) of VWF contains modular domains of D'-D3-A1-A2-A3-D4-B1-B2-B3-C1-C2-CK. The specific interaction of VWF domains with subendothelial collagen fibres type I and II as well as platelets are depicted in the zoomed circles. Figure reproduced from (Heidari et al., 2015).

## b. Secondary haemostasis

Secondary haemostasis also referred to as coagulation, is a proteolytic cascade involving sequential activation of large a number of coagulation factors, which leads to the generation of active thrombin. At the site of vessel wall damage, the tissue factor is exposed to the circulation and interacts with factor VIIa and sequentially activates factors IX and X, thus initiating the coagulation cascade. Thrombin has numerous functions including the activation of platelets and proteolysis of fibrinogen into fibrin, resulting in the consolidation of the haemostatic clot through fibrin polymerization.

### **c. Fibrinolysis**

The clot is a transient structure, which is eventually dissolved by a process named fibrinolysis once the damage to the vessel has been repaired. Plasmin is the primary “fibrinolysin” and it catalyses the degradation of the thrombus by cleaving the fibrin polymers and, consequently, counteracts pathological thrombus formation. Similar to the coagulation process, the fibrinolytic system is tightly regulated to assure balance haemostasis. A main player is  $\alpha_1$ -antiplasmin, which is the most potent regulator of plasmin activity.

### **d. The balance of haemostasis**

The balance of haemostasis plays an essential role in maintaining either procoagulant or anticoagulant mechanisms in the circulatory system extremely tightly regulated. The balance is regulated by numerous factors such as those involved in controlling platelet activation, coagulation and fibrinolysis. The absence of these factors distorts haemostasis and thereby causes either thrombotic and/or bleeding complications including arterial and inflammatory diseases, such as myocardial infarction, ischemic stroke, preeclampsia and cerebral malaria among others. In this context, the proteolytic cleavage of ultralarge VWF (UL-VWF) by ADAMTS13 is crucial for maintaining the delicate balance between haemostasis and thrombosis.

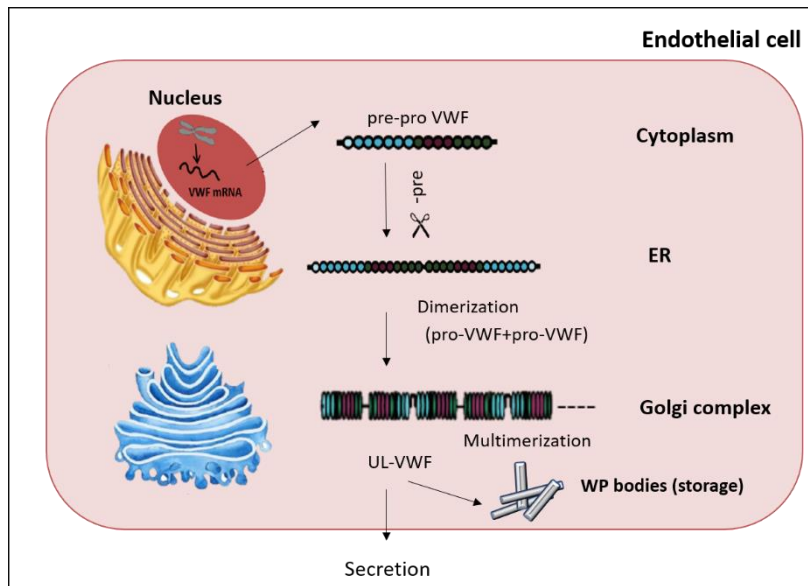
## **3.3 Von Willebrand factor (VWF)**

### **3.3.1. Biosynthesis and secretion**

VWF is produced exclusively in endothelial cells (ECs) and megakaryocytes as a 9-kb pre-pro-VWF monomer transcript that consist of 2813 amino acids (aa) with a 22-aa signal peptide, a 741-aa pro-peptide and a 2050-aa mature subunit (J E Sadler 1998).

This VWF precursor protein is subject to a series of intracellular processes including removal of the signal peptide and pro-peptide, glycosylation, and multimerization (Fig. 14). Translocation of the VWF into the ER is enabled by the proteolytic processing of the signal peptide. In the ER, pro-VWF forms dimers via disulfide bonds located at the C-terminus, in a process named “tail-to-tail” dimerization. These dimers are transported to the Golgi in which they undergo further modifications such as multimerization through an additional disulfide bond near the N-terminus of the mature subunit (“head-to-head” multimerization), N- and O-glycosylation, and proteolytic removal of the large propeptide by furin (X Zheng et al. 2001). Thereafter, UL-VWF multimers encompassing ~10,000 kDa are stored in highly

ordered tubules in the endothelium named Weibel-Palade bodies and in the  $\alpha$ -granules from platelets.



**Fig. 14. Synthesis of VWF in endothelial cells.** The VWF mRNA is translocated to the cytoplasm where it is translated to the pre-pro-VWF protein, which is further translocated to the ER via its signal peptide. The pre-peptide is cleaved in the ER, VWF dimers (protomers) are formed by a disulfide bond between two pro-VWF moieties. Further multimerization occurs in the Golgi complex. UL (Ultra large) VWF are transported to Weibel Palade bodies or secreted into the blood stream. Figure adapted from (Luo et al. 2012).

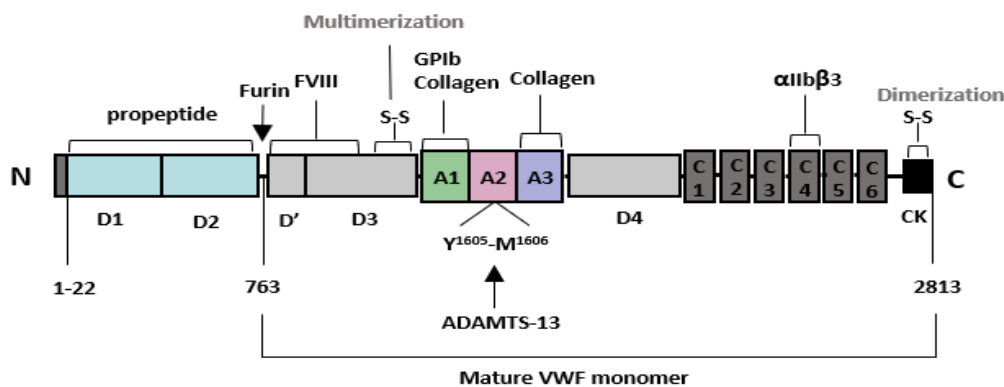
These UL-VWF multimers are constitutively released from the endothelium into the plasma (Giblin, Hewlett, and Hannah 2008). They are also secreted on demand in response to a wide range of physiologic agonists, such as thrombin, epinephrine and inflammatory cytokines (Stockschlaeder, Schneppenheim, and Budde 2014). On secretion, a proportion of VWF released from endothelial cells remains anchored to the cell surface, from which it is proteolytically cleaved and released by ADAMTS13 (Dong et al. 2002). Without this cleavage, these UL-VWF multimers would eventually form a string-like structure that would rapidly recruit platelets and leukocytes to endothelial cells causing thrombi (Bernardo et al. 2005). Therefore, the action of ADAMTS13 is essential to prevent thrombosis in the microvasculature.

### 3.3.2. Function and structure of VWF

Each VWF monomer is composed of homologous domains (Fig. 15) among which D1 and D2 play a particularly important role in *de novo* multimerization process and packaging of VWF while the mature subunit contributes to platelet adhesion and aggregation. Indeed,

several domains within the mature VWF contain functional binding or cleavage sites that contribute to specific procoagulant functions.

Several structural domains are involved in the post-translational processing of VWF and include the cysteine knot (CK) and D3 domain, which are required for dimerization and multimerization of VWF monomers respectively. Moreover, The D'/D3 domains interact with factor VIII. The A1 domain is important for binding to GPIIb $\alpha$  on platelets. It is also involved in collagen binding, although the major collagen binding site is located within the A3 domain. The A2 domain contains the cleavage site for ADAMTS13. The C4 domain, in turn, interacts with the platelet receptor GPIIb/III $\alpha$  (Schneppenheim and Budde 2011). Experimental evidence showed that either binding capacity to GPIb and GPIIb/III $\alpha$ , and platelet aggregating activity increases under high shear conditions with the degree of multimerization of the VWF (Federici et al. 1989; Fischer et al. 1996).



**Fig. 15. VWF domain structure:** consisting of a signal peptide, a propeptide and a mature VWF subunit. Carboxyl terminal CK domain and D3 domain are responsible of dimerization and multimerization. The main physiological ligands as well as the ADAMTS13 cleavage site are indicated. Figure adapted from (Kremer Hovinga et al. 2017).

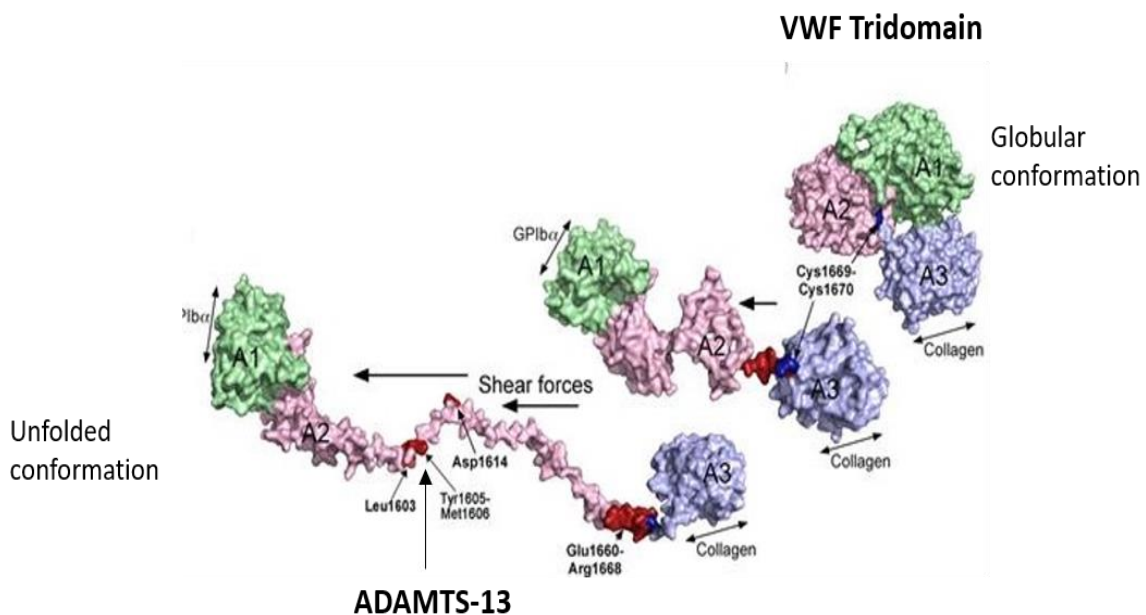
The current understanding of the organization of the VWF domains remains uncertain. Only the three A domains (A1, A2 and A3) are well characterized through crystal structures (Bienkowska et al. 1997; Emsley et al. 1998; Zhang et al. 2009). These domains are members of the VWA protein fold and family (Springer 2006) and contain a central hydrophobic  $\beta$ -sheet with 6  $\beta$ -strands, which are typically surrounded by 6 amphipathic  $\alpha$ -helices.

### 3.3.3. Regulation of size and conformation of VWF

The multimeric size of UL-VWF is regulated by ADAMTS13 proteolysis in the circulation. Certain bleeding conditions, such as von-Willebrand-Disease (VWD), and

thrombotic diseases, such as thrombotic thrombocytopenic purpura (TTP), are associated with an imbalance of the regulation of the size of VWF. As aforementioned, VWF is stored and released from platelets and endothelial cells as UL-VWF. Although UL-VWF is itself not usually detected in the blood stream, a series of lower molecular weight multimeric forms (L, I and HMW), ranging in size from 500 to 20,000 kDa, are detectable (Zhou et al. 2011). These lower molecular weight forms exist because UL-VWF is cleaved by ADAMTS13. Once secreted or proteolytically released in free circulation, VWF (L, I and HMW) multimers adopt an inactive globular conformation in order to prevent spontaneous platelet binding (Crawley et al. 2011).

In this context, the behavior of the VWF A1-A2-A3 domains is essential for VWF to function. In the circulation, the A3 domain is constitutively accessible to bind to exposed collagen. Conversely, either the glycoprotein Ib binding site within the A1 domain and the A2 exosites such as ADAMTS13 cleavage site (Y<sub>1605</sub>-M<sub>1606</sub>) remain hidden or buried. Therefore, VWF unravelling is a prerequisite for proteolysis reported to be involved not only in decoupling of the tridomain cluster (A1, A2, A3), but also in conformational changes within additionally individual domains, mainly in the VWF A2 domain (Fig. 16).



**Fig. 16. Shear-induced unfolding of VWF and cleavage through ADAMTS13.** In normal physiological conditions, VWF is in a folded globular conformation. Shear stress allows unfolding, thus exposing the A1 and A2 domains of VWF. Therefore, platelets can bind to the A1 domain while ADAMTS13 can interact and cleave VWF in its A2 domain decreasing the activity of VWF. Figure adapted from (Crawley et al. 2011)



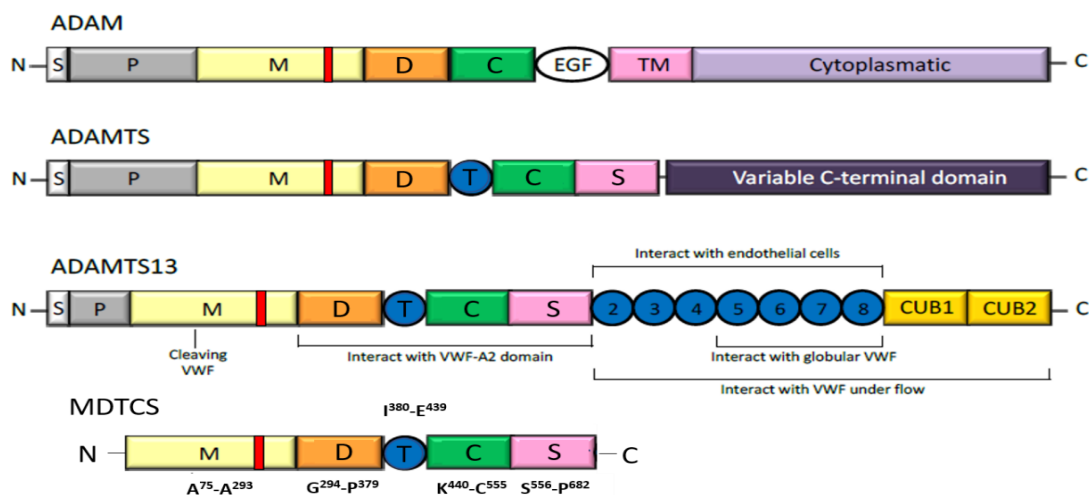
Higher shear stress is an important determinant of VWF function, not only regulating the large form VWF, but also modulating the exposure of both VWF A1 domain platelet binding sites and the VWF A2 domain ADAMTS13 binding/cleavage site(s) (Crawley et al. 2011; Denis and Lenting 2012). Additionally, platelets or the presence of FVIII enhance the tensile forces exerted on VWF, contributing to its shear-dependent unravelling. Therefore, fluid shear stress, platelets and FVIII act synergistically to promote VWF proteolysis by AD13.

### 3.4. ADAMTS13

ADAMTS13 (AD13) is recognized as a VWF-cleaving protease. Since its discovery, AD13 has generated a wave of extensive research on its structure, its mode of action, its involvement in pathology and its potential therapeutic role in cardiovascular disease.

#### 3.4.1 ADAMTS13 protease is part of a family of metalloproteases

AD13 is a metalloprotease belonging to the ADAMTS family (Levy et al. 2001). The ADAMTS family comprises 19 soluble multi-domain zinc-proteases, which share similarities with ADAM (A Disintegrin And Metalloprotease) proteases. Both protein families contain a conserved catalytic metalloprotease domain, preceded by a signal peptide and a propeptide, plus C-terminal disintegrin-like and cysteine-rich domains. ADAMTS lack the epidermal growth factor-like (EGF) and transmembrane domains present in the ADAM family.



**Fig. 17. Domain structure of the ADAM and ADAMTS protease family as well as AD13.** Signal peptide (S), propeptide (P), metalloprotease domain (M) (location of the zinc-binding motif is shown in red), disintegrin-like domain (D), cysteine-rich domain (C), EGF-like domain (EGF), transmembrane domain (TM), first thrombospondin type-I (TSP-1) repeat (T), spacer domain (S), TSP repeats 2 to 8 (2-8), two CUB domains (CUB1, CUB2). Figure adapted from (Lancellotti and De Cristofaro 2011).

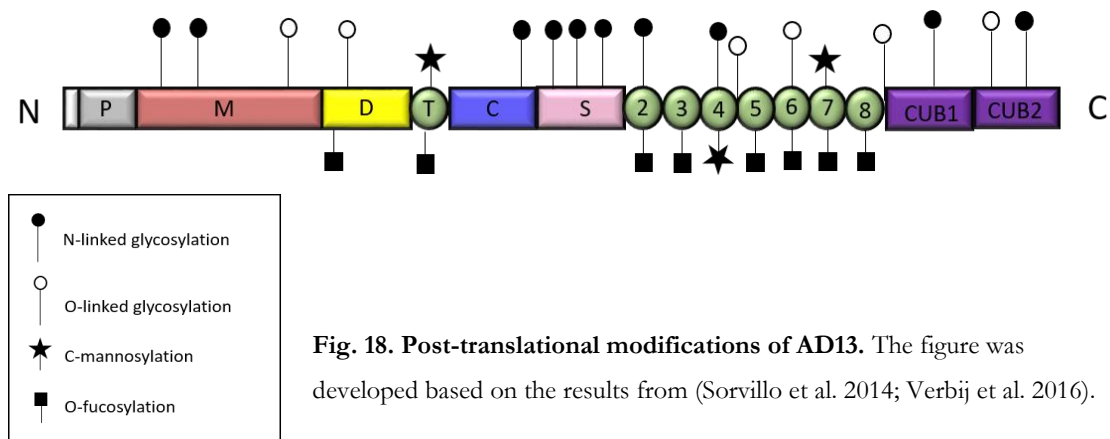
Instead; they contain thrombospondin type-1 repeat (T) and a spacer (S) domain. Additionally, the C-terminal is the most variable region within the ADAMTS family. The carboxyl terminus of AD13 contains seven more thrombospondin (TSP) repeats and two CUB domains (Fig. 17). Because of substrate specificity, but also due to the abnormally short pro-peptide and the unique C-terminal domains, AD13 seems evolutionary farthest away among the family members (X Zheng et al. 2001).

### 3.4.2 Gen, synthesis and secretion

The AD13 gene is located on chromosome 9q34, and 29 exons encode for a 5 kb mRNA transcript that results in a protein of 1427-aa (Fujikawa et al. 2001; X Zheng et al. 2001). AD13 is a glycoprotein expressed predominantly in the liver by hepatic stellate cells (Uemura et al. 2005). Nevertheless, it has also been found in platelets, cultured endothelial cells and glomerular podocytes.

AD13 circulates in the blood at a concentration of 1 $\mu$ g/ml (Soejima et al. 2006) with an estimated half-life of three days (Furlan et al. 1999). In addition, differently from other plasma proteases, AD13 is secreted into the blood as a constitutively active enzyme. For these reasons it is fundamental that AD13 is highly specific to ensure that it does not proteolyze other targets nonspecifically. Additionally, no natural inhibitors are reported to regulate AD13 function. *In vitro*, several endogenous inhibitors of MMPs, ADAMs, ADAMTS proteases have been tested with no inhibition effect on AD13 (Guo, Tsigkou, and Lee 2016).

Posttranslational modifications such as glycosylation are also important for the secretion of AD13. Glycosylation plays a role in folding, secretion, clearance and interaction with the immune system (Verbij et al. 2016). AD13 contains 10 potential N-glycosylation sites. Five out of a total of six O-linked glycans are identified as attached to a serine while just one O-linked glycan is attached to a threonine residue. There are 8 O-fucosylation sites which play an important role in either secretion or protein folding. Lastly, three C-mannosylation sites are involved also in the expression and proteolytic activity of AD13. Disrupting these glycan sites was shown to impair AD13 secretion.



**Fig. 18. Post-translational modifications of AD13.** The figure was developed based on the results from (Sorvillo et al. 2014; Verbij et al. 2016).

### 3.4.3 AD13 domain organization and structure

AD13 is translated into a 1427-aa precursor protein and has a modular, multidomain structure. From N- to C- terminus, AD13 possesses a signal peptide, a propeptide, the catalytic Metalloprotease (M) domain, a Disintegrin-like (D) domain, a Thrombospondin type 1 repeat (T), a Cysteine-rich (C) domain, a Spacer domain (S), seven more TSP domains (TSP2-8) and two CUB (Complement component C1r/C1s, sea urchin *Uegf* protein, *Bone morphogenic protein-1*) domains. The crystal structure of the entire enzyme is yet to be solved. In 2009, the crystal structure of the proximal non-catalytic domains of AD13 (D-to-S domains, DTCS) was published (Akiyama et al. 2009). Recently, the crystal structure of AD13 MDTCS was solved (Fig. 20) (Petri et al. 2019).

Functional studies reveal that AD13 contains non catalytic exosites in each of its D, C, T and S domain, which contribute to a ‘molecular zipper model’ of VWF recognition and proteolysis. Therefore, deletion of any of these ancillary domains may cause a reduction of either binding affinity and/or proteolysis of VWF.

#### 3.4.3.2 The propeptide (P<sup>30</sup>-R<sup>74</sup>)

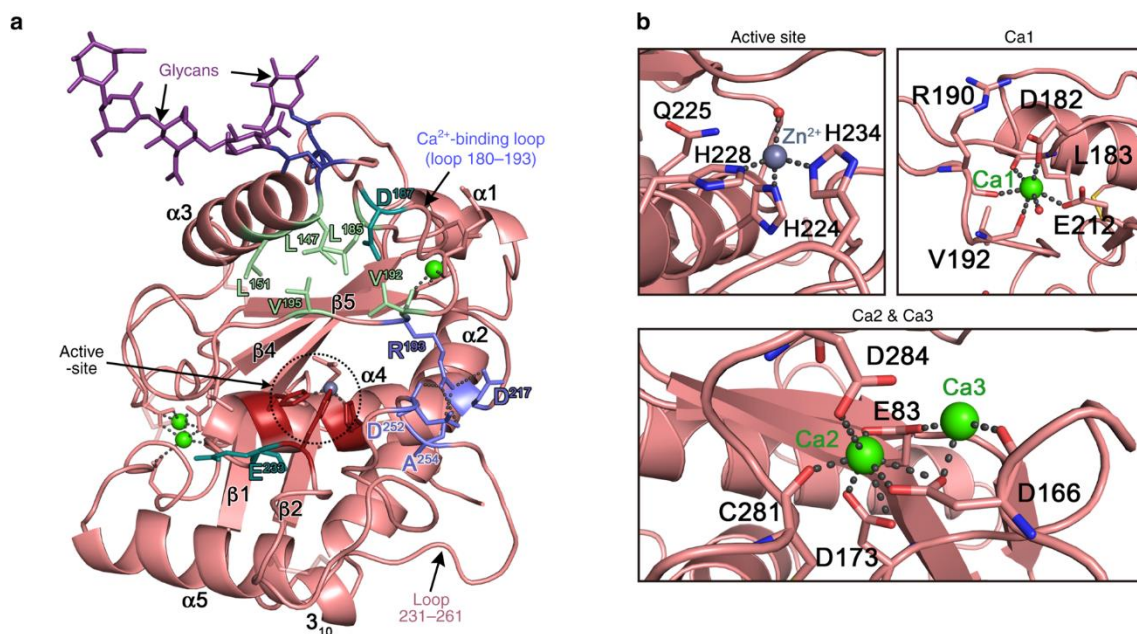
The short 41-aa propeptide is poorly conserved and is not required for secretion, folding or enzymatic activity, in contrast to other ADAMTS family members. Therefore, AD13 is constitutively active after folding in the ER and does not require propeptide removal to expose the active site (Xinglong Zheng et al. 2003).

#### 3.4.3.3 The metalloprotease domain (A<sup>75</sup>-A<sup>293</sup>)

The M domain contains the active site and several structural elements that are conserved in the reprotolysin family of proteases (HEXXHXXGXXHD; Gomis-Ruth 2009), which play an important role in the proteolytic function of AD13 to cleave the peptide bond Y<sub>1605</sub>-M<sub>1606</sub>

in the VWF A2 domain (Crawley and Scully 2013). The active site contains a  $Zn^{2+}$  ion, which is coordinated by three conserved His residues (Fig. 19b). Additionally, there is a catalytic Glu residue ( $E^{225}$  in AD13), which is indispensable in forming the active site and coordinating a water molecule for the proteolytic hydrolysis reaction to occur (Bode, Gomis-Ruth, and Stockler 1993; de Groot, Lane, and Crawley 2010).

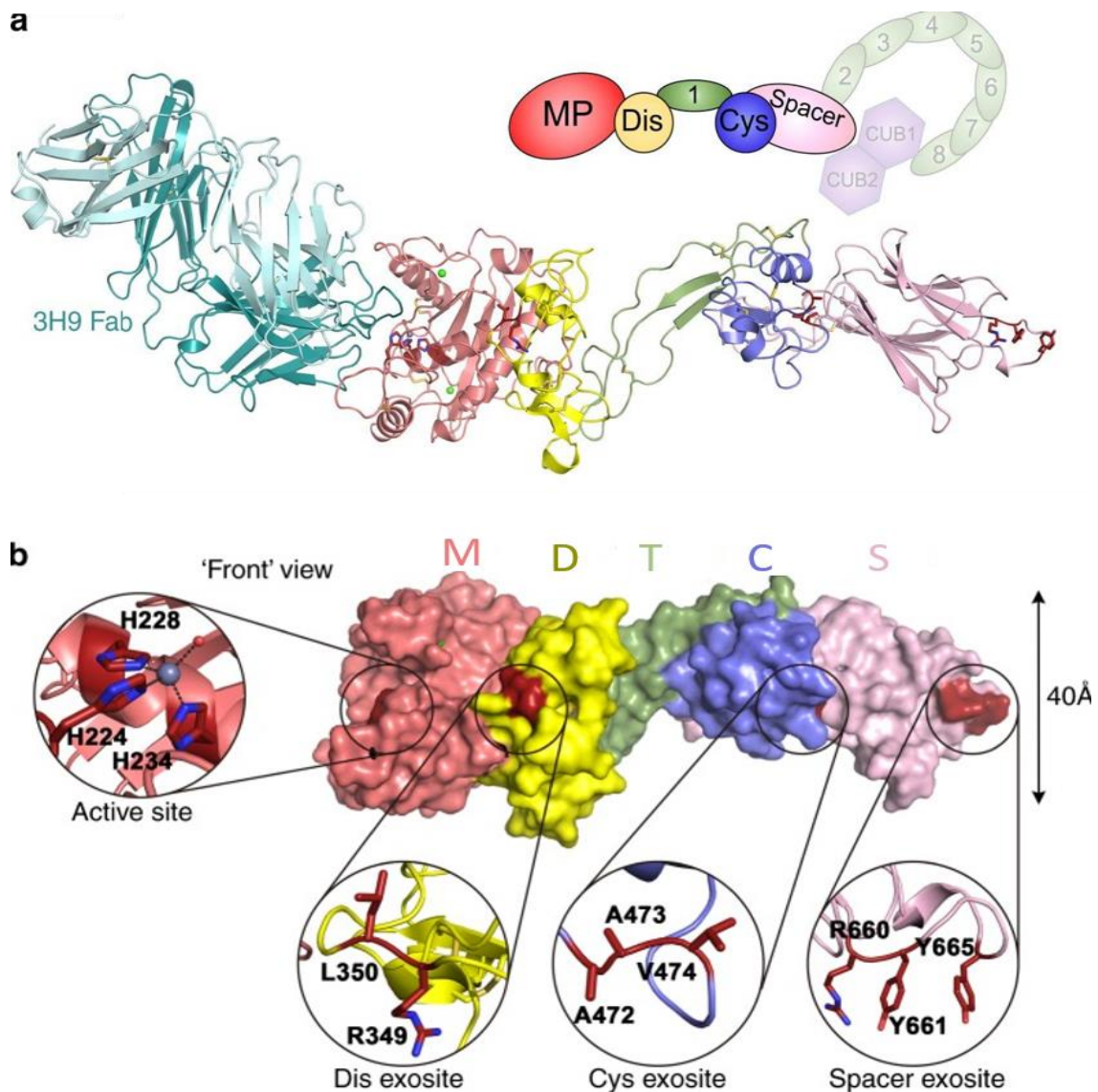
Adjacent to the  $Zn^{2+}$ -binding motif, a conserved Met residue ( $M^{249}$  in AD13) creates a tight turn ('Met-turn'), which is a structural element characteristic of this family of enzymes (Tallant et al. 2010). For optimal activity, AD13 requires divalent cations. Apart from  $Zn^{2+}$  in the active site,  $Ca^{2+}$  is also functionally relevant (Anderson, Kokame, and Sadler 2006). The first  $Ca^{2+}$  ion is coordinated by  $E^{212}$  and  $D^{182}$  and the carbonyl backbone of  $L^{183}$ ,  $R^{190}$ , and  $V^{192}$ . This functionally important  $Ca^{2+}$  binding site is located adjacent to the active-site cleft and plays an important role in maintaining the shape of the active centre.



**Fig. 19. Structure of the AD13 metalloprotease (M) domain.** (a) Crystal structure of the AD13 M domain. the active-site, loop 180–193, loop 231–263, and the metal-binding sites are marked. Two N-linked glycans are shown in purple. Surface residues  $L^{147}$ ,  $L^{151}$ ,  $L^{185}$ ,  $V^{192}$ , and  $V^{195}$  (light green) form a hydrophobic patch above the active-site that likely contributes to the S1 pocket. To the right of the active-site,  $R^{193}$ ,  $D^{217}$ ,  $D^{252}$  and  $A^{254}$  (blue) form ionic interactions with each other that bridge loops 180–193 and 231–263. (b) Four metal ions ( $Zn^{2+}$  and three  $Ca^{2+}$ ) are bound by the M domain. Figure reproduced from (Petri et al. 2019).

In AD13, the sides of the active site cleft are composed by interactions between loop 180-193 and loop 231-263 mediated by side chains of a so-called gatekeeper triad ( $R^{193}$ ,  $D^{217}$  and  $D^{252}$ ), which function to stabilize the active-site cleft. The reported crystal structure

represents a latent conformation, which is unable to accommodate the peptidic substrate (Petri et al. 2019). It has been shown that the M domain alone does not have the ability to bind nor cleave VWF specifically (Gao, Anderson, and Sadler 2008; de Groot et al. 2009), since proteolysis is dependent on multiple interactions between exosites from non-catalytic domains (DTCS) and the A2 domain of VWF (Crawley et al. 2011). However, the M domain counts with subsites that promote low affinity interactions with residues around the cleavage site of VWF. Mutations in D<sup>187</sup>, R<sup>190</sup> and R<sup>193</sup> impair proteolysis, suggesting that they are essential in contributing to the structure or function of AD13 (de Groot, Lane, and Crawley 2010).



**Fig. 20. Crystal structure of the 3H9 Fab-MDTCS (M to S) complex.** (a) Schematic representation of the MDTCS domains bound to the inhibitory 3H9 Fab fragment (PDB ID: 6QIG). (b) Surface structure of the MDTCS domains (without Fab) viewed from the front with important regions for substrate recognition. Figure reproduced from (Petri et al. 2019).

#### **3.4.3.4 The disintegrin-like domain (G<sup>294</sup>-P<sup>379</sup>)**

Although the disintegrin-like domain shows structural similarities with snake venom disintegrins, this domain does not have an RGD integrin binding sequence. Contrarily, its crystal structure revealed a similar tertiary structure to the Cys-rich domain (Petri et al. 2019). While the M domain by itself is not functional, the addition of this domain is enough to reestablish specific though inefficient VWF cleavage (Gao, Anderson, and Sadler 2008). Moreover, targeted mutagenesis of a variable region identified three residues involved in VWF proteolysis: R<sup>349</sup>, L<sup>350</sup> and V<sup>352</sup>, which interacts with D<sub>1614</sub> and A<sub>1612</sub> in VWF. Recent structural data confirm that R<sup>349</sup> is fundamental for the proteolytic processing (Akiyama et al. 2009).

#### **3.4.3.5 The thrombospondin type 1-like domain (T) (I<sup>380</sup>-E<sup>439</sup>)**

AD13 includes eight TSP domains, however only one TSP1 (T) structure is available. It is an elongated domain which shares homology with the second TSP in thrombospondin type 1. The core of this domain is formed of Trp, Arg and hydrophobic residues with two pairs of disulphide bonds. Using truncated versions of both VWF and AD13, suggests that T interacts with the Q<sub>1624</sub>-V<sub>1630</sub> region within the VWF A2 domain (Gao, Anderson, and Sadler 2008). Whether there is a functional exosite in the T, which residues are involved in the interaction with VWF remains to be concluded. Current thinking supports that T gives stability and spacing to D/C domains.

The function of TSP2-8 has been mostly investigated employing C-terminal truncation mutants of AD13. The TSP2-8 regions contain three linkers, which bring flexibility to these domains allowing AD13 to exist in both folded and extended conformations. Additionally, TSP8 was recently reported to be involved in binding for D4-CK domains in globular VWF, thus it permits the transition from a folded to an extended AD13 conformation. It will be covered in more detail in section 3.4.4.2.

#### **3.4.3.6 The cysteine-rich domain (K<sup>440</sup>-C<sup>555</sup>)**

This domain is a well-conserved and contains ten cysteine residues. Several studies showed that the cysteine rich domain is critical for VWF binding and proteolysis through hydrophobic interactions. De Groot et al identified a non-conserved region in the cysteine rich domain (G<sup>471</sup>-V<sup>474</sup>), corresponding to a V-loop, which was involved in the recognition



of hydrophobic residues in the A2 domain, such as I<sub>1642</sub>, W<sub>1644</sub>, I<sub>1649</sub>, L<sub>1650</sub>, I<sub>1651</sub> (de Groot, Lane, and Crawley 2010).

#### **3.4.3.7 The spacer domain (S<sup>556</sup>-P<sup>682</sup>)**

The spacer domain contains no cysteines and it doesn't show homology with any other known structural motif. By using truncation constructs of AD13 lacking the Spacer domain, it has shown a significantly reduction of cleavage, by losing its affinity for VWF. Therefore, the Spacer domain was essential in the N-terminal part of AD3 (MDTCS) to be able to cleave VWF. Residues R<sup>659</sup>, R<sup>660</sup>, Y<sup>661</sup> and Y<sup>665</sup> have been shown to be crucial in a AD13 binding exosite which interacts with the C-terminal region comprised in VWF A2 (E<sub>1660</sub>-R<sub>1668</sub>), being the earliest exosite binding site to be discovered (Xinglong Zheng et al. 2003; Ai et al. 2005; Majerus, Anderson, and Sadler 2005).

Recent studies have also revealed an interaction between CUB domains and the Spacer domain, which gives AD13 a compact conformation. It has been shown that the blockade of this interaction induces an extended conformation of AD13 and thus, enhances its activity up to 2 to 10 fold (Muia et al. 2014; South et al. 2014).

#### **3.4.3.8 CUB domains**

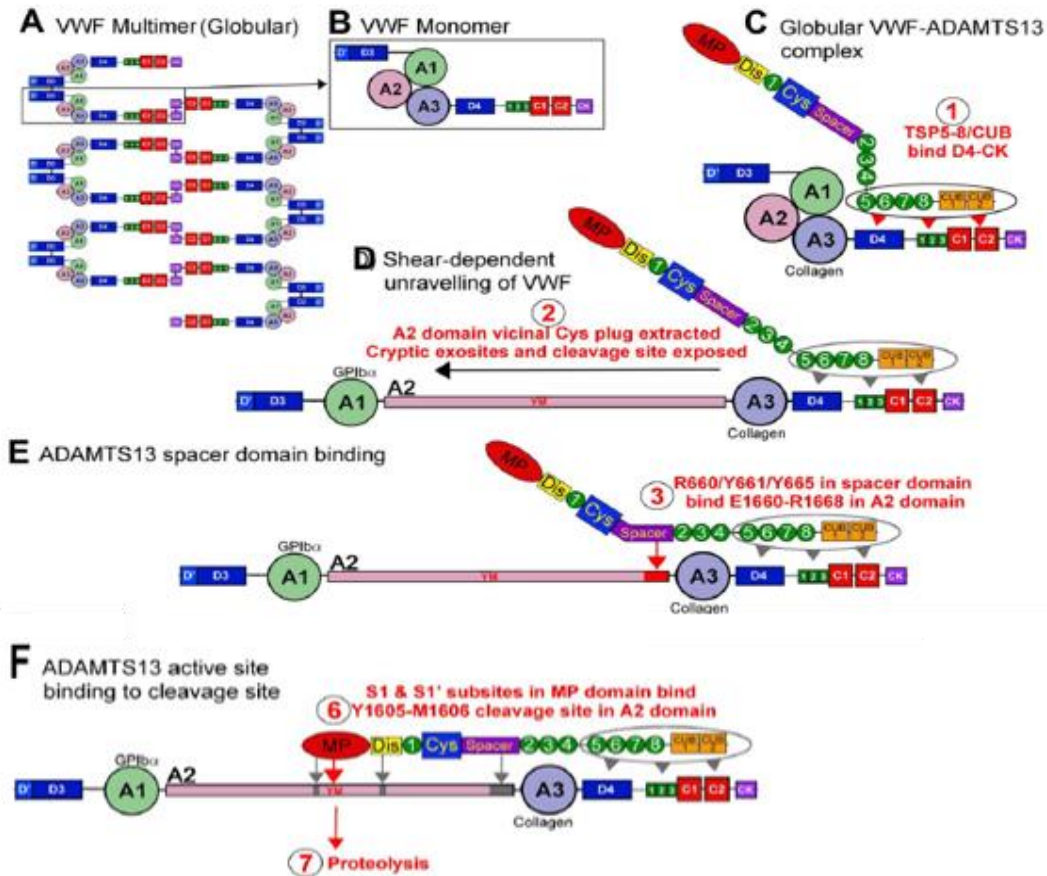
AD13 is unique among the ADAMTS family which possesses two CUB domains, located at the C-terminus of the enzyme (Porter et al. 2005). As mentioned above, AD13 CUB domains were postulated to interact with the Spacer domain allowing the folded conformation of AD13. Binding of the TSP8-CUB2 to VWF D4-CK region, was suggested to disrupt Spacer-CUB interaction leading to an active extended AD13 conformation (see section 3.4.4.2). To date, the role of the CUB domain in AD13 function remains controversial, although they may be implicated in secretion and intracellular trafficking.

### **3.4.4 Recognition and cleavage of VWF by AD13**

#### **3.4.4.1 AD13-VWF interactions**

To date, VWF is the only known substrate of AD13. The proteolytic cleavage of UL-VWF by AD13 is essential for maintaining the delicate balance between coagulation and haemostasis. Recognition of VWF by AD13 is complex and involves multiple interaction sites on VWF (Fig. 21). The main substrate specificity is largely based on the extensive interactions between non-catalytic domains (DTCS) of AD13 and VWF (Majerus, Anderson,

and Sadler 2005; Gao, Anderson, and Sadler 2008; Akiyama et al. 2009). However, the structural elements distant from the unique cleavage site (Y<sub>1605</sub>-M<sub>1606</sub>) of VWF also play an important role in the proteolysis (Zanardelli et al. 2006).



**Fig. 21. AD13-mediated cleavage of VWF.** Schematic representation of the interaction between full-length AD13 and VWF leading to VWF proteolysis. **(C-D)** Docking of the distal AD13 domains to VWF D4CK enables binding of AD13 to VWF. **(E)** Unfolding of the VWF A2 domain facilitates binding of the AD13 spacer domain to the VWF A2 domain. **(F)** High affinity binding of the spacer domain unfolding AD13 enabling the proximal AD13 domains within the DTCS region to interact with additional exosites within the VWF A2 region. Following these interactions, the catalytic site of AD13 within the metalloprotease domain is positioned to engage and cleave the Y<sub>1605</sub>-M<sub>1606</sub> scissile bond. Figure adapted from (Crawley, et al. 2011).

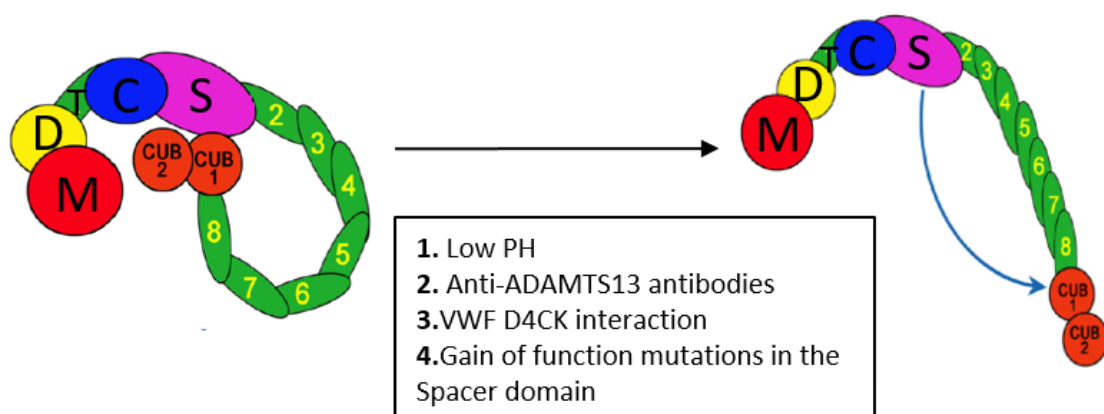
Under normal physiological conditions, VWF circulates in plasma in a globular conformation where the A2 domain is hidden in a hydrophobic core located at the C-terminal of the domain (Luken et al. 2010). Consequently, the first contact between AD13 and VWF is made with a modest affinity through their respective C-terminal exposed domains, being T5-CUB2 and D4CK respectively. Due to changes in shear stress conditions, VWF unravel, leading to expose exosites in the N-terminal region of VWF. Subsequently,



the Spacer domain of the protease recognised VWF residues E<sub>1660</sub>-R<sub>1668</sub> (Gao, Anderson, and Sadler 2008). Therefore, inducing correct positioning of the D and M domain, eventually allowing to the proteolysis of VWF between residues Y<sub>1605</sub> and M<sub>1606</sub> (Crawley et al. 2011). It is notable to mention that the VWF D<sub>1596</sub>-V<sub>1604</sub> N-terminal sequence has a structural essential determinant where VWF L<sub>1603</sub> and V<sub>1604</sub> residues became docking sites for a complementary subsite in AD13 involving L<sup>198</sup>, L<sup>232</sup>, and L<sup>274</sup> (Xiang et al. 2011). These interactions redirect the scissile bond in the catalytic cleft of AD13. At this point, additional binding interactions take place between adjacent domains, which finally trigger VWF cleavage.

#### 3.4.4.2 The conformational activation of ADAMTS13

AD13 has been described as a constitutively active enzyme. Its activity has been shown to be influenced by conformational changes due to VWF binding. Recently, AD13 has been shown to be a conformational flexible molecule which may adopt both, folded and extended conformations (Fig. 22). The folded conformation is favoured in solution and adopts an open conformation when bound to VWF. Proof of this hypothesis was provided by small angle X-ray scattering and electron microscopy which showed that AD13 is roughly folded in half whereby the C-terminal interacts with MDTCS (Muia et al. 2014; Deforche et al. 2015). The folded conformation is based on the interaction between the Spacer and CUB domains.



**Fig. 22. Conformational activation of AD13:** AD13 adopts a folded conformation, mediated by inter-domain interactions between Spacer and CUB interaction. Mutations in the Spacer domain, monoclonal antibodies, lower pH and binding of the VWF D4(-CK) domain, can abolish the interactions and results in an extension of the molecule. Figure adapted from (Petri et al. 2019).

Three linker regions comprised in TSP2-CUB are responsible for AD13 conformational flexibility (Deforche et al. 2015). A wide variety of conditions disrupt the Spacer-CUB interaction such as lower pH, mutations in the AD13 Spacer domain, several anti ADAMTS13 monoclonal antibodies and interaction with VWF D4(-CK) domain. Interestingly, this extended conformation enhances AD13 activity 2-to-10-fold, which is proposed to be due to suppressing the negative effect of the CUB domain interaction with the Spacer domain shows eventual implications in VWF binding and proteolysis (South et al. 2014; South, Freitas, and Lane 2017).

#### **3.4.4.3 Factors affecting cleavage of VWF by ADAMTS13**

AD13 is secreted as an active protease and no physiological inhibitors have been identified as of to date. Shear forces and multiple interactions of remote exosites on the protease and the substrate are finally the main regulators of AD13 activity. However, this activity might be negatively regulated by other factors such as chloride ions, haemolysis products, inflammatory cytokines, and some coagulation proteases.

Release of several inflammatory cytokines during the early stage of systemic inflammation, such as interleukin-6, may stimulate UL-VWF release and inhibit its cleavage, resulting in the accumulation of hyper-reactive multimers in plasma which induce platelet adhesion and aggregation on the endothelium (Bernardo et al. 2005). Similarly, free haemoglobin resulting from *in vitro* haemolysis, has shown to inhibit AD13 activity. In this context, AD13 inhibition results as a consequence of an inflammatory activation.

Proteases of the coagulation cascade such as thrombin and plasmin, inhibit AD13 activity by directly cleaving the protease at the site of vascular injury, contributing to thrombus formation. Other potential plasma inhibitors such as Factor H (Feng et al. 2013), a regulator of complement activation, and heparin (Nishio et al. 2004) may regulate AD13 activity through direct binding to VWF and endothelial cells.

Inorganic divalent cations, which exert a functional and structural effect on both AD13 and VWF, bind to the metalloprotease domain interfering protease activity. At high concentration,  $Zn^{2+}$  and  $Ca^{2+}$  stabilize the folded conformation of VWF avoiding its unfolding, consequently protecting it from cleavage (Zhou et al. 2011). Conversely, at concentration exceeding physiological levels,  $Mg^{2+}$  can enhance VWF cleavage under flow conditions (Dong et al. 2008).

On the other hand, AD13 activity has been shown to be positively modulated by platelets (Shim et al. 2008) and factor VIII (Cao et al. 2008; Skipwith, Cao, and Zheng 2010), proposing that their binding to specific regions of VWF may promote the conformational transitions under fluid shear stress, accelerating the rate of specific cleavage at the VWF Y<sub>1605</sub>-M<sub>1606</sub> scissile bond and thus reducing VWF multimer size.

### **3.4.5 ADAMTS13 in disease**

#### **3.4.5.1 Thrombotic thrombocytopenic purpura (TTP)**

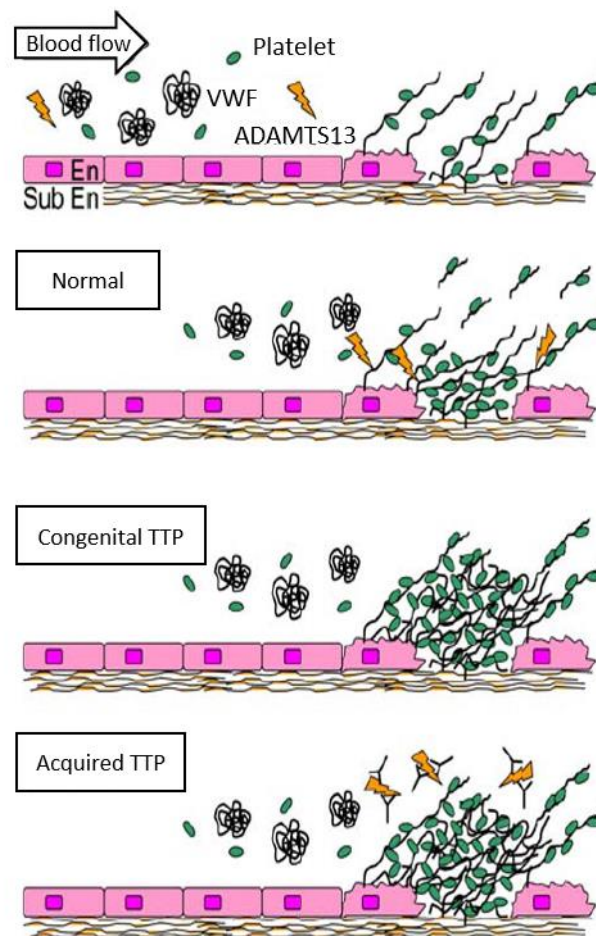
Regulation of the VWF multimer size through AD13 is crucial in maintaining hemostasis. If left unchecked, thrombosis instead of hemostasis may occur. Thrombotic thrombocytopenic purpura (TTP) is a life-threatening disease caused by a functional deficiency in AD13 that is either the result of mutations in the *ADAMTS13* gene (congenital or hereditary TTP) or the presence of anti-AD13 antibodies (acquired TTP). Severely deficient AD13 function results in the presence and persistence of highly adhesive UL-VWF multimers in circulation which can unfold and expose cryptic platelet binding sites in the absence of vascular injury. The spontaneous tethering of platelets to unfolded circulating VWF leads to the formation of circulating platelet aggregates which in turn can occlude microvessels (Fig. 23). Consequently, widespread microvascular thrombosis in various organs may lead to ischemic organ failure and ultimately death if left untreated.

TTP is a rare thrombotic disorder with an estimated annual incidence of 6 cases per million per year in the UK (Scully et al. 2008). TTP may manifest several symptoms including severe thrombocytopenia with platelet counts below 20,000 per  $\mu$ l, hemolytic anemia with schistocytosis, fever, neurological complications and renal failure (Moschcowitz et al. 1952). The importance of this disease lies not only within its incidence, but furthermore in its molecular mechanism that most likely will reveal important clinical and basic insights in the field of hematology and hemostasis research. Based on the primary mechanism causing a deficiency in AD13 activity, TTP can be clinically classified into two major forms, congenital or acquired TTP.

#### **3.4.5.2 Congenital TTP**

Congenital TTP, also named as Upshaw-Schulman syndrome, is a rare recessively inherited disease that accounts for  $\leq 5\%$  of all TTP cases (Levy et al. 2001; Lotta et al. 2010). To date, more than 140 mutations distributed throughout the *ADAMTS13* gene have been

identified, many of them impairing its secretion and activity. The majority of the reported mutations are single amino acid missense substitutions while the remaining part include insertions, deletions, nonsense mutations, and splice site mutations (Lotta et al. 2010). While heterozygous carriers have half-normal AD13 activity in plasma (40-70% of normal values) and are generally asymptomatic, individuals with homozygous or compound heterozygous mutations have a severe AD13 deficiency (<10% of normal value) and most certainly develop TTP (Lancellotti, Basso, and De Cristofaro 2013).



**Fig. 23. The pathophysiology of thrombotic thrombocytopenic purpura (TTP).** AD13 regulates the size of the platelet thrombi in the case of endothelial activation, promoting secretion of the UL-VWF, or vessel injury, by proteolysis of VWF. In TTP, AD13 activity is notably deficient, allowing the thrombus to grow in an uncontrolled way and eventually to occlude the microvessel. Congenital TTP is associated with constitutional deficiency of the protease in plasma. Complete protease absence is thought to be incompatible with life, and AD13 may be detectable at low concentrations in plasma of some patients with congenital TTP. Acquired TTP occurs as a consequence to the presence of auto-antibodies that alter either substrate recognition and binding as protease activity. Figure reproduced from (Manea and Karpman 2009).

Via site-directed mutagenesis and expression studies *in vitro* it has been demonstrated that *AD13* mutations are associated with impaired synthesis, catalytic activity, or secretion (Uchida et al. 2004; Camilleri et al. 2012). However, it is generally accepted that additional factors besides mutations of AD13 contribute to the unpredictable nature of congenital TTP. Other genetic defects, such as mutations in complement factor H gene, have been suggested (Noris et al. 2005). On the other hand, environmental/external triggers are known to influence the phenotype as patients often encounter a first bout of TTP in situations that are associated with VWF release and elevated VWF levels, such as pregnancy or infection (Fujikawa et al. 2001; Douglas et al. 2010).

#### **3.4.5.3 Acquired TTP**

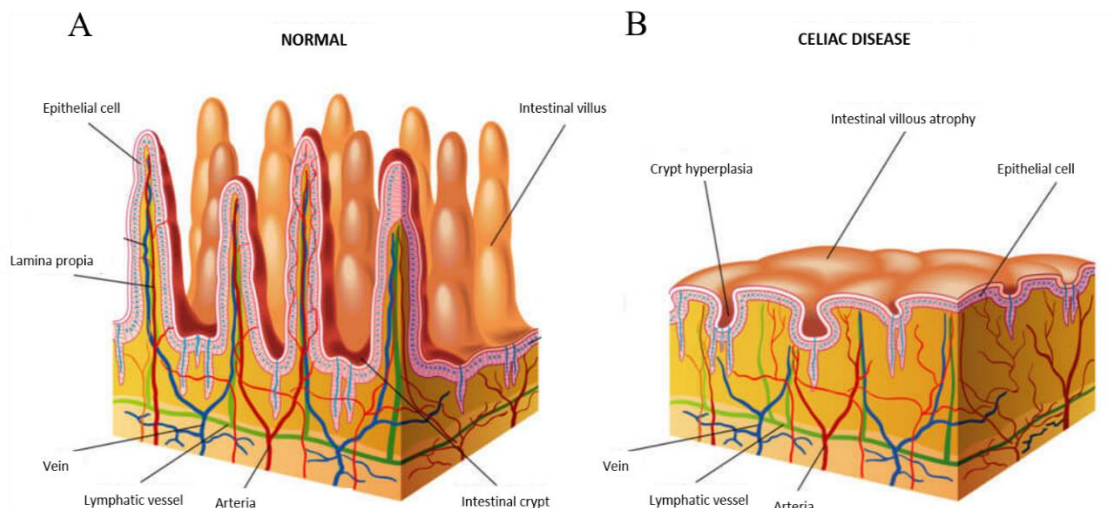
In acquired TTP, which accounts for the majority of the TTP cases (95%), a severe deficiency in AD13 function is caused by autoantibodies directed against the circulating protease (Blombery and Scully 2014). Pathogenic autoantibodies can contribute to severe ADAMTS13 deficiency (<10% of normal activity) through direct inhibition of AD13. It has been shown that the majority of these inhibitory autoantibodies primarily target antigenic sites within the spacer domain, which plays a critical role in substrate recognition (Jian et al. 2012; Thomas et al. 2015).

## 4. Prolyl-endoropeptidase, neprosin

Coeliac disease (CD) is a severe immune disorder in which genetically predisposed people can't eat gluten as it leads to damage in the small intestine. The disease is triggered by gluten-derived immunogenic peptides that are generated by the proteolytic enzymes of the stomach (e.g. pepsin) and intestine (e.g. trypsin). Enzyme-supplement therapy represents a promising approach to overcome this gluten intolerance, similar to the use of lactase tablets in people with lactose intolerance. One candidate “glutenase” (i.e. gluten degrading protease) is the prolyl endopeptidase neprosin from a carnivorous pitcher plant.

### 4.1. Coeliac disease

Gluten was identified as the main culprit of coeliac disease by Dicke in 1957, who observed that coeliac children symptomatically improved when wheat and rye were scarce during the 1944-45 famine (Kupfer and Jabri 2012; Dicke, 1957). Coeliac disease (CD) is an intestinal inflammatory disease defined as chronic small intestinal immune-mediated enteropathy precipitated by exposure to dietary gluten in individuals genetically predisposed (Ludvigsson et al. 2013) who respond to gluten intake with the presentation of villous atrophy, mucosal inflammation, crypt hyperplasia and intraepithelial lymphocyte infiltration (IELs) (Fig. 24).



**Fig. 24. Schematic representation of A) healthy and B) damaged villi.** Figure reproduced by <https://avenue360.org/blog/2019/04/25/may-is-national-coeliac-disease-awareness-month/>

### 4.1.1 Epidemiology and pathophysiology

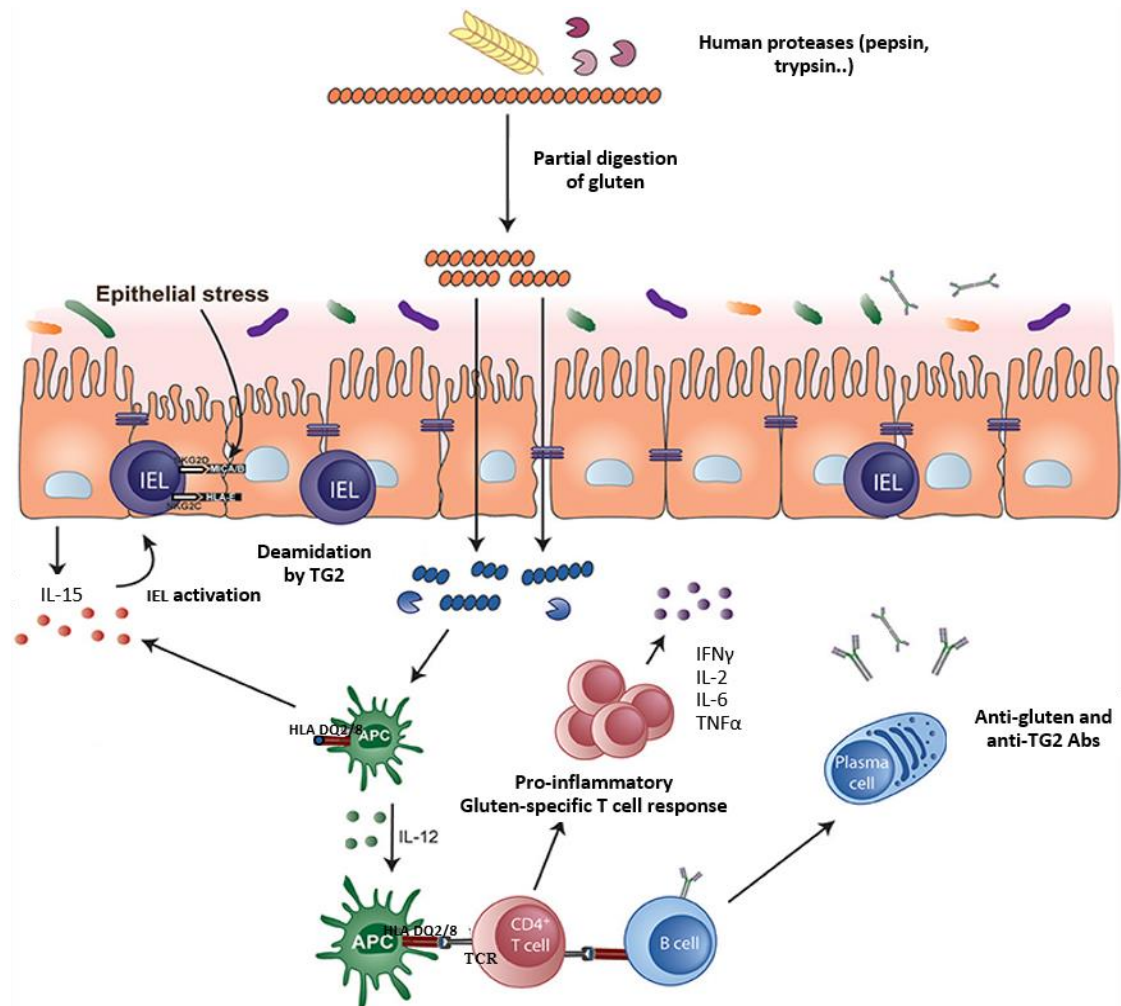
Coeliac disease affects individuals of any age, races, and ethnic groups. The overall prevalence in the Western population is approximately 1% and somewhat higher in certain European countries. The presentation of CD greatly varies from gastrointestinal and nutritional derangements to neuropsychiatric symptoms, liver diseases, infertility, among extra-intestinal manifestations (Lebwohl, Ludvigsson, and Green 2015).

Coeliac disease is induced principally by dietary gluten in genetically susceptible individuals expressing the human-leukocyte antigen (HLA)-DQ2 or DQ8 (E. Liu et al. 2014). The HLA-DQ2 haplotype frequency in patients with CD is approximately 90-95% while HLA-DQ8 is less associated with CD, with a frequency of 5-15%, indicating a strong genetic risk for the disease (Gujral, Freeman, and Thomson 2012). Moreover, population with other diseases, especially autoimmune diseases such as diabetes type I or multiple sclerosis present higher prevalence to suffer coeliac disease.

The immune response to gluten involves both the innate and adaptive immune systems. After the ingestion of gluten-containing food, gluten is partially digested by proteases of the gastrointestinal track into relatively large peptides (Koning 2012). Some of these oligopeptides pass through the intestinal epithelium and reach the lamina propria, where they may bind directly to HLA-DQ2 or HLA-DQ8 and trigger a T-cell response, which may lead to local tissue damage (W. Vader et al. 2002). Moreover, these gluten peptides may cause the release of tissue transglutaminase 2 (TG2), which can convert glutamine residues into glutamate by deamidation in a large number of gluten peptides (Molberg et al. 1998). This chemical change dramatically increases the affinity for the positively charged binding pockets of the HLA molecule on antigen presenting cells (APCs) (Schuppan, Junker, and Barisani 2009), thus amplifying the T cell response (Fig. 25).

Recognition of gluten peptides by CD4<sup>+</sup> T cells produces proinflammatory cytokines (Sollid 2002), such as interferon gamma (INF- $\gamma$ ), tumor necrosis factor-alpha (TNF- $\alpha$ ), interleukin 21 (IL-21), IL-6 and IL-15 (Nilsen et al. 1995; Abadie et al. 2011). Simultaneously, gluten can induce IL-15 production through distinct mucosal cell populations contributing to enhance the inflammatory cascade and to further activate cytotoxic intraepithelial lymphocytes (IEL), which result in direct cytotoxic epithelial damage (Jabri et al. 2000; Maiuri et al. 2003). At the same time, B-cells produce specific antibodies against gluten constituents and TG2 (mainly IgA and IgG), which also may contribute to the pathogenesis of coeliac disease (Mäki 1995; Lindfors, Mäki, and Kaukinen 2010). The circulation of these antibodies

in the bloodstream provides a possible explanation for the appearance of extraintestinal manifestations, including those in the central nervous system, such as peripheral neuropathy, migraine, and ataxia, among others (Sturgeon, Lan, and Fasano 2017).



**Fig. 25. Key steps in CD pathogenesis.** Gluten epitopes resist to GI degradation. TG2 catalyzes the deamidation of gluten peptides, which results in more efficient binding to HLA DQ molecules from APCs. APCs activate CD4<sup>+</sup> T-cells, which in turn secrete a variety of inflammatory cytokines such as IL-21 and IFN- $\gamma$  that take a part to the intestinal lesion, stimulate B cell response and promote the activation of IELs that contribute to the destruction of enterocytes expressing stress signals. Figure reproduced from (Tye-Din, Galipeau, and Agardh 2018).

## 4.2 Wheat composition: Gluten

Cereals are the most important crops worldwide, comprising a total of ~2000 million tons of grain per year; 70% of cereal widely consumed are wheat, corn and rice, which represent rich sources of starch—the basic dietary components for the growing human population (Peter R Shewry and Halford 2002). A wheat kernel contains 8-15% of protein, of which 10-



15% is albumin/globulin and 85-90% gluten (P R Shewry 2009), which appears to serve as a reserve of amino acids for the development of the seedling in the germination process. Gluten is a complex mixture composed of hundreds of related but distinct proteins, mainly prolamins (Mena and Sousa 2015). According to their solubility they are typically split into two groups: an alcohol-soluble fraction termed gliadins (monomeric) and the insoluble-glutenins (polymeric, soluble in dilute bases and acids; Fig. 26).

Similar storage proteins exist as secalin in rye, hordein in barley, and avenins in oats and are collectively referred to as “gluten” (Schalk et al. 2017). It has been revealed that gliadins play a role in the extensibility and cohesiveness of gluten, while glutenins contribute in the maintenance of the elasticity and strength of the gluten (Wieser 2007).

Today, gluten consumption in a Western diet is estimated to range from 10-20 g/day (P R Shewry 2009) and it also includes prolamins found in rice and maize. However, in order to simplify, the term gluten is commonly restricted to prolamins with the facility to trigger immunotoxicity in certain groups of patients with a gluten sensitivity spectrum.

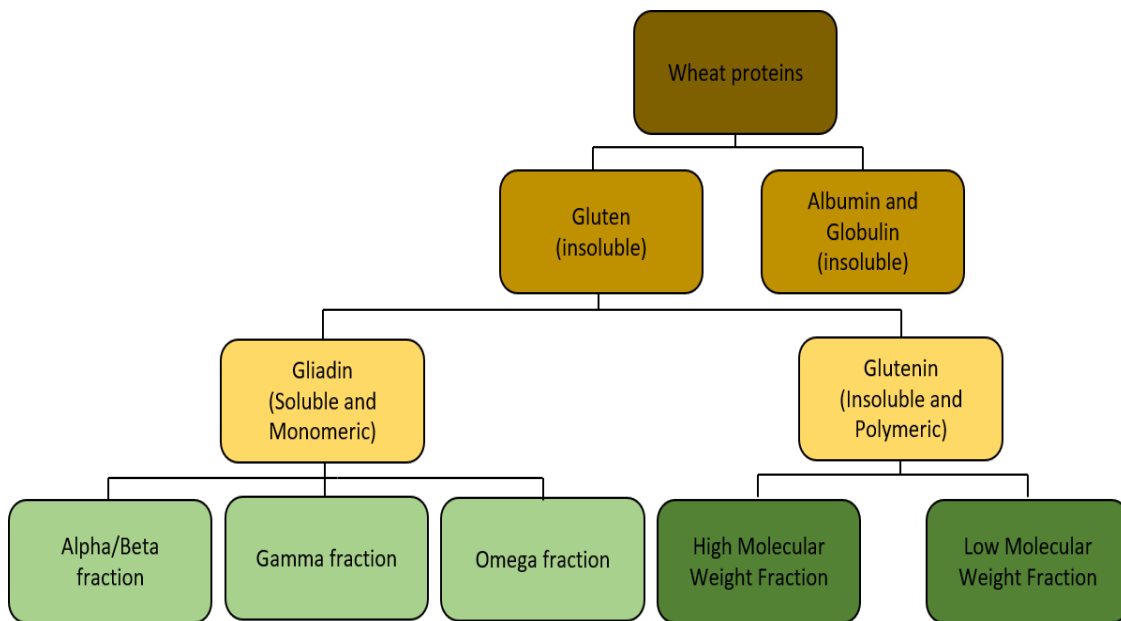


Fig. 26. Schematic representation of the wheat gluten composition.

#### 4.2.1 Gliadins

Gliadin is the major component of wheat gluten. It is composed by a single chain polypeptide linked by intramolecular disulphide bonds with an average molecular weight of 25-100 kDa. Gliadins have considerable low solubility in aqueous solution except at extreme pH (Elzoghby, Samy, and Elgindy 2012). Apart from disulphide bonds, this low solubility in

water results from the presence of hydrophobic interactions, which give rise to a folded shape and contribute to the unique properties of gluten (van den Broeck et al. 2009), including the dough quality of bread and other baked products such as pasta, cakes, pastries, and biscuits (Schalk et al. 2017).

Gliadins can be sorted into subgroups according to their sulphur content, molecular weight and may further be classified as subject to primary structures into alpha, beta, gamma, and omega ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\omega$ ) (van den Broeck et al. 2009). Gliadins are characterized by high significant amounts of glutamine (38%) and proline (20%) residues in their primary structures (Wieser 2007). The proteins and polypeptides comprise a similar structure: signal peptides for translocation into cellular compartments, a non-repetitive N-terminal region, a non-repetitive C-terminal region and a long repetitive central region which contains a repetitive glutamine and proline rich repeat unit unique to each that render them sterically inaccessible/resistant to proteolytic enzymes from the stomach, pancreas, and intestinal brush borders (Schalk et al. 2017).

### 4.3 Proteases involved in protein digestion

A wide variety of proteolytic enzymes are required to break down proteins from diet into small peptides and amino acids, since the digestive enzymes show high specificity for different types of peptide bonds (Table 1). However, the significant fraction of proline and glutamine residues in gluten proteins play a crucial role in protecting peptides against proteolytic degradation.

Enzyme	Activators	Action	Cleavage points	Products
<b>Pepsin</b>	Autoactivation	Endopeptidase	Tyr, Phe, Leu, and Asp	Large peptide and free amino acids
<b>Trypsin</b>	Enterokinase and trypsin	Endopeptidase	Arg and Lys	Oligopeptides (2-6 amino acids)
<b>Chymotrypsin</b>	Trypsin	Endopeptidase	Tyr, Trp, Phe, Met, and Leu	Oligopeptides (2-6 amino acids)
<b>Elastase</b>	Trypsin	Exopeptidase	Ala, Gly, and Ser	Oligopeptides (2-6 amino acids)
<b>CP A</b>	Trypsin	Exopeptidase	Carboxy-terminus Val, Leu, Ile and Lys	Free amino acids
<b>CP B</b>	Trypsin	Exopeptidase	Carboxyl-terminus Arg and Lys	Free amino acids
<b>Aminopeptidases</b>	Trypsin	Exopeptidase	Amino terminus	Free amino acids

**Table 1. Characteristics of gastric, intestinal, and pancreatic proteases.**

Consumed gluten proteins or polypeptides begin to be broken down in the oral cavity by proteases of bacterial origin that are naturally present in whole saliva. (Wei et al. 2016). Most protein digestion begins in the stomach under the action of pepsin (Binder et al. 2009).

In brief, pepsin is secreted as a zymogen termed pepsinogen by chief cells in the gastric mucosa. Similarly, gastric acid (HCl) is secreted by parietal cells and alters the conformation of pepsinogen, thus enabling its self-cleavage and maturation to become an active protease in the stomach. Additionally, gastric acid denatures the proteins, i.e. partly unfolds them, allowing pepsin to better access their peptide bonds. The hydrolytic activity of pepsin is maximal at pH of 1.8 to 3.5 with a high specificity for peptide linkages of aromatic residues such as (phenylalanine, tryptophan, tyrosine) and leucine (Devlin TM, 2006). Pepsin may partially digest 10-15% of dietary protein generating large peptides of 30-40 amino acids. This partial hydrolysis of gluten proteins results in the generation of high-molecular weight gluten polypeptides, also denominated as gluten immunogenic peptides (GIP), which reach the duodenum.

The main protein digestion occurs in the upper part of the small intestine through pancreatic proteases (trypsin, chymotrypsin, CP-A, CP-B, and elastase), which are secreted by the pancreatic duct along with pancreatic bicarbonate. Bicarbonate initiates the neutralization of the stomach acid, raising the pH to the optimum for the activity of the pancreatic proteases and simultaneously, leading to the irreversible inactivation of pepsin. Some oligopeptides are further hydrolyzed by brush border enzymes to tripeptides, dipeptides, and free amino acids, which are easily absorbed by the enterocytes (Ganapathy, Hay, and Kim 2012). Additionally, some enzymes in the intestinal epithelium (e.g dipeptidyl carboxypeptidase 1, dipeptidyl peptidase IV, and aminopeptidase N) are able to hydrolyze peptides with proline residues.

However, a proportion of large peptides rich in proline and glutamine are resistant to pancreatic and brush border enzymes hydrolysis (Frazer et al. 1959) triggering a toxic effect that contributes to the inflammatory processes in susceptible coeliac individuals (Hausch et al. 2002; Shan et al. 2002).

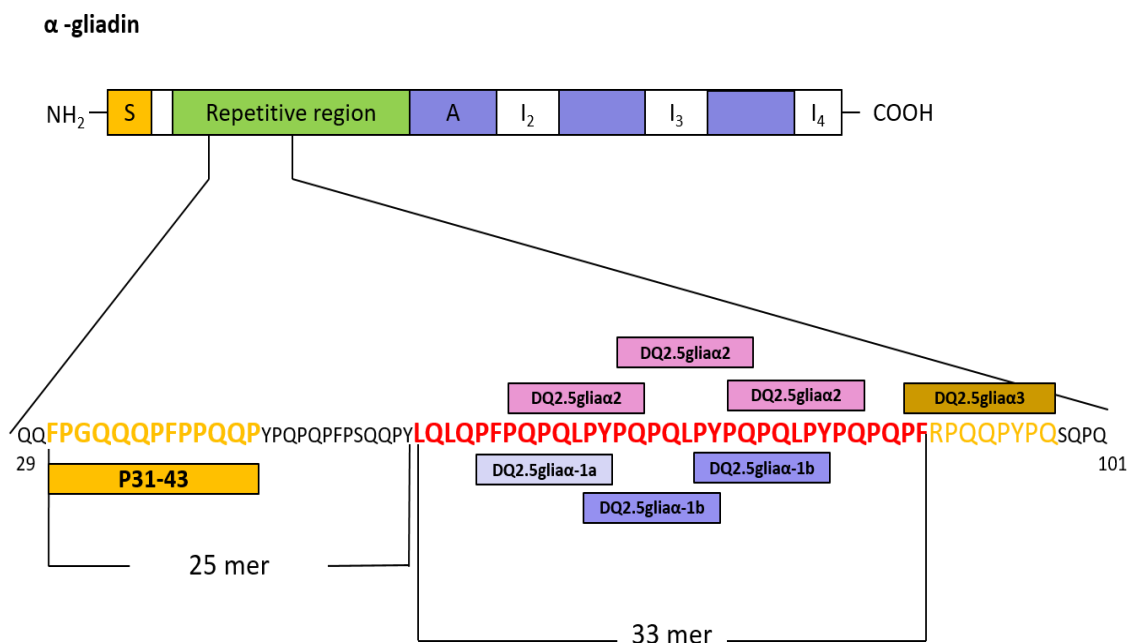
#### **4.4 Gluten immunogenic peptides (GIP)**

Hundreds of GIP have been described in wheat, rye, barley and oats and are able to trigger the activation of T-cells (Real et al. 2012). However, not all GIP are equally harmful to coeliac

disease patients, since the number of epitopes may vary between different cereals (L. W. Vader et al. 2003).

Although the complete repertoire of peptides implicated in the pathogenesis of coeliac disease persist a daunting task because of the great heterogeneity of gluten proteins (Ang et al. 2010), numerous studies have demonstrated that GIP derived from  $\alpha$ -gliadins provoke strong responses in the vast majority of patients, while responses to the other peptides are less commonly found.

The wheat  $\alpha$ -gliadin proteins contain three main GIP: the p31-43, which induces the innate immune response by inducing IL-15 production from dendritic cells and enterocytes; the 33-mer (p57-89; LQLQPFQPQLPYPQQLPYPQQLPYPQPQPF), composed of six overlapping copies of three highly stimulatory epitopes; and an additional DQ2.5-glia- $\alpha$ 3 epitope (FRPQQYPYPQ) located downstream of the 33-mer (Ozuna et al. 2015; Fig. 27).



**Fig. 27. Fragment of the  $\alpha$ -gliadin protein sequence.** Potential immunogenic peptides: p31-43 peptide, 33-mer and DQ2.5-glia- $\alpha$ 3 are shown. Figure reproduced from (Balakireva and Zamyatnin 2016).

#### 4.5 Properties of the 33-mer peptide from $\alpha$ -gliadin

The  $\alpha$ -gliadin 33-mer is considered the main immunodominant toxic peptide in coeliac patients. It is located in the N-terminal repetitive region of  $\alpha$ -gliadin and contains six overlapping copies of three different DQ2-restricted T-cell epitopes: DQ2.5-glia- $\alpha$ 1a (PFPQPQLPY), DQ2.5-glia- $\alpha$ 1b (PYPQPQLPY) and DQ2.5-glia- $\alpha$ 2 (PQPQLPYPQ) with

highly stimulatory properties (Shan et al. 2002; Fig. 27). Once the 33-mer reaches the lamina propria, it is deamidated by TG2 and plays a central role in the pathogenic cascade of CD by inducing the adaptive immune response. It has been described that the 33-mer possesses a structural setup characterized by a left-handed polyproline II helical conformation that is chosen by MHC class II ligands (Herrera et al. 2014).

Experiments *in vivo* showed that the 33-mer peptide is undigested by enzymes of the intestinal brush border. Furthermore, in a monkey model of gluten sensitivity, 33-mer can be found in the serum at the onset of disease, suggesting this peptide can invade the mucosa *in vivo* (Mazumdar et al. 2010).

#### **4.6 Novel coeliac disease therapies**

To date, no pharmacological treatment is available for gluten intolerant individuals. A life-long strict gluten-free diet is the only efficient and safe treatment available. However, complete avoidance of gluten from our daily menu can be expensive, unpleasant and complicated due to cross-contamination and/or the presence of small traces of gluten in food and medicines.

Numerous therapeutic alternatives to the gluten-free diet have been engineered to mitigate the problem related to gluten-intolerant individuals. Some of the approaches include the use of nonhuman proteases (e.g., bacteria, fungi, plants etc.) for gluten detoxification, inhibition of TG2 to prevent peptide deamidation and the development of low or zero content of immunotoxic sequences using RNAi and CRISPR/Cas9 transgenic technologies among others (Bethune et al. 2006; Gil-Humanes et al. 2014; Sánchez-León et al. 2018).

Due to growing concern regarding the use of transgenic technologies, enzyme-supplement therapy has been described as one of the most promising therapeutic approach promoting the complete digestion of cereal proteins, and thus destroying T-cell gluten epitopes, in particular (Bethune and Khosla 2012). An example of an enzyme supplement are lactase pills, administered with milk products thus, helping to prevent abdominal bloating, gas, upset stomach and diarrhea produced by these products (Suchy et al. 2012).

##### **4.5.1 Detoxification of gluten proteins with enzymatic therapy**

After ingestion, gluten immunologic peptides (with high glutamine and proline content) reach the duodenum, being poorly degraded by the enzymes present in the gastrointestinal tract. Hence, oral enzyme therapy has been proposed as an alternative of gluten free diet.

Promising enzymes are microbial prolyl endopeptidases (PEPs) which have been intensively explored as potential glutenases since they are especially effective in the hydrolysis of peptide bonds on the carboxyl side of internal proline residues in gluten-derived oligopeptides (Shan et al. 2005). A high number of proteases possessing glutenase activities were isolated from bacteria (*Flavobacterium meningosepticum* (FM-PEP), *Myxococcus xanthus* (MX-PEP) and *Sphingomonas capsulata* (SC-PEP), germinating cereals (*Triticum aestivum* L., *Hordeum vulgare*) or fungi (*Aspergillus niger* (AN-PEP), *Aspergillus oryzae*). However, several enzymes, mostly bacterial PEP, were not stable and active in the harsh environments of the stomach and/or upper small intestine even though they seem to be rapidly proteolyzed by gastric pepsin (Shan et al. 2004). Another disadvantage is the low specificity for immunogenic epitopes on gluten that results in incomplete detoxification. In order to solve this problem, a variety of strategies are being considered such as enzyme combinations with complementary specificity (e.g. Latiglutenase: combination of EP-B2 enzyme from barley and PEP from bacteria *S. capsulata*) or the development of engineered glutenases (e.g. Kuma030 from the acidophilic microbe *Alicyclobacillus* A8).

Other potential glutenases from carnivorous pitcher plants producing digestive fluid, which has displayed an impressive ability to degrade gluten into non-toxic peptides more efficiently than stomach enzymes like pepsin, are in nascent stages of investigation (Rey et al. 2016). Notably, they efficiently degraded the 33-mer peptide into non-toxic peptides (patented) and showed a high efficiency even at a substrate to enzyme ratio of 12,000:1, which translates into a low amount of protease for efficient detoxification of gluten proteins.

#### **4.6 Digestive secretion from carnivorous plants**

Carnivorous plants have developed original feeding strategies through the capture and digestion of prey, mainly insects for nitrogen uptake. They are found all around the world, especially in nutrient-poor areas. *Drosera* and *Nepenthes* are two carnivorous plant genera able to produce and excrete from their tissues a relevant amount of digestive fluid.

Plants from the genus *Nepenthes* possess highly specialized leaves named pitchers that act as pitfall-traps, also called a pitcher (Fig. 28). In the inner and basal part of the pitcher, a glandular structure which could be described as a single-stage gastrointestinal (GI) tract has been defined. It produces an acidic viscoelastic fluid intended to retain and digest caught prey (Bazile et al. 2015). The glands have two complementary functions: they produce acidic digestive fluid containing a combination of digestive enzymes, antimicrobial compounds,

acidic polysaccharides and mineral nutrients, and are further required for the assimilation of nutrients resulting from prey digestion (Owenj. et al. 1999).

Concerning the native proteases present in the digestive fluid, several studies based on genomic and proteomic research, established the enzymatic composition of the *Nepenthes* secretion. A complex mix of enzymes was defined such as aspartic, cysteine, serine and prolyl endopeptidases (Athauda et al. 2004; Stephenson and Hogan 2006; Hatano and Hamada 2012; Rottloff et al. 2016; Lee et al. 2016). Apart from these proteases, various hydrolytic enzymes including phosphatases, ribonucleases, lipases, phosphoramidasases, esterases and different chitinases involved in breaking down the prey have been identified (Eilenberg et al. 2006; Rottloff et al. 2011; Mithöfer 2011; Buch et al. 2013; Renner and Specht 2013).



**Fig. 28. *Nepenthes ventrata*: carnivorous pitcher plant:** the pitcher plant is popular and gorgeous for gardeners but is deadly for insects that fall into the trap for which it is termed. Figure reproduced from: <https://www.the-scientist.com/notebook/pitcher-plant-enzymes-digest-gluten-in-mouse-model-32109>.

#### **4.6.1 Plant aspartic proteases: nepenthesisins**

The majority of plant aspartic proteases identified so far belong to the MEROPS A1 family. Nepenthesisin I and II, are two of the most abundant and well characterized enzymes found in the digestive fluid (Takahashi et al. 2005; Hatano and Hamada 2008; Kadek et al. 2014). A noteworthy adaptation of these proteins to a harsh environment permits assimilation of prey or organic components. Additionally, these enzymes appear to be stable

over wide ranges of temperatures and pH, thus being also resistant to different classes of proteases. Owing to these characteristics, these proteases became extraordinary candidates for biotechnological and therapeutic applications. For instance, nepenthosins are considered an excellent tool for protein digestion in hydrogen/deuterium exchange mass spectrometry (HDX-MS). Although pepsin is frequently used in this technique, nepenthosins appeared to be at least 1400-fold more efficient. This is due to the high stability combined with slightly wider cleavage specificities, resulting in an improved sequence coverage and spatial resolution (Kadek, Mrazek, et al. 2014).

Finally, *Nepenthes* secretion has recently been considered as an alternative source of therapeutic enzymes for the treatment of coeliac disease. Moreover, it has been demonstrated that the fluid contains considerably-potent gluten detoxification capacity, which can be attributed to the combined action of a non-canonical aspartic protease and a novel prolyl endoprotease (Rey et al. 2016).

#### **4.6.2 A novel plant prolyl endoprotease: neprosin**

The evolution of distinct trapping mechanisms for pitcher carnivorous plants to thrive in adverse environments with limited nutrients may result in proteases with novel characteristics. The research group of Dr. Schriemer from the University of Calgary in Canada noticed recently that the fluid of the *N. ventrata* pitcher plant possessed a new cleavage specificity profile, which consists in a robust C-terminal proline activity, distinct from the action of aspartic proteases. As a result, the founding member of a novel class of prolyl endopeptidase called neprosin was identified (Lee et al. 2016). Schrader et al. (2017) produced neprosin through a low-yield recombinant protein expression system in *E. coli* and proved that it holds the potential to be employed for whole proteomic profiling and histone mapping. However, the limited yield prevented further in-depth studies.

##### **4.6.2.1 Neprosin characterization: recombinant protein**

Neprosin is a low molecular weight prolyl endopeptidase (40 kDa), which exhibits similarity based on biochemical and regulation properties with nepenthosins. Recent studies have demonstrated that neprosin is secreted as a zymogen which contains a prodomain that is cleaved and released after enzyme activation at pH 2.5. Neprosin was partially inhibited by an aspartate inhibitor called pepstatin A, and retains activity under strongly reducing conditions (Schröder et al. 2017). It shows its highest activity at pH 2.5, although it holds around 50% activity at pH 4.5. The further the pH rises, the more the enzyme starts losing



its proteolytic activity. Moreover, the activity is also influenced by temperature, being the optimal range between 37 °C and 50 °C (Lee et al. 2016).

Rey et al. generated proteolytic maps that revealed endoprotease activity with high specificity for Pro-X cleavage. Schrader et al. confirmed this cleavage site but also revealed a less frequent cleavage site behind Ala (Fig. 29). In spite of its considerably small size and in contrast with other known propyl endopeptidases, neprosin does not seem to have any restriction in substrate length. Due to these properties this enzyme might be used for a bottom-up proteomics approach for many potential applications.

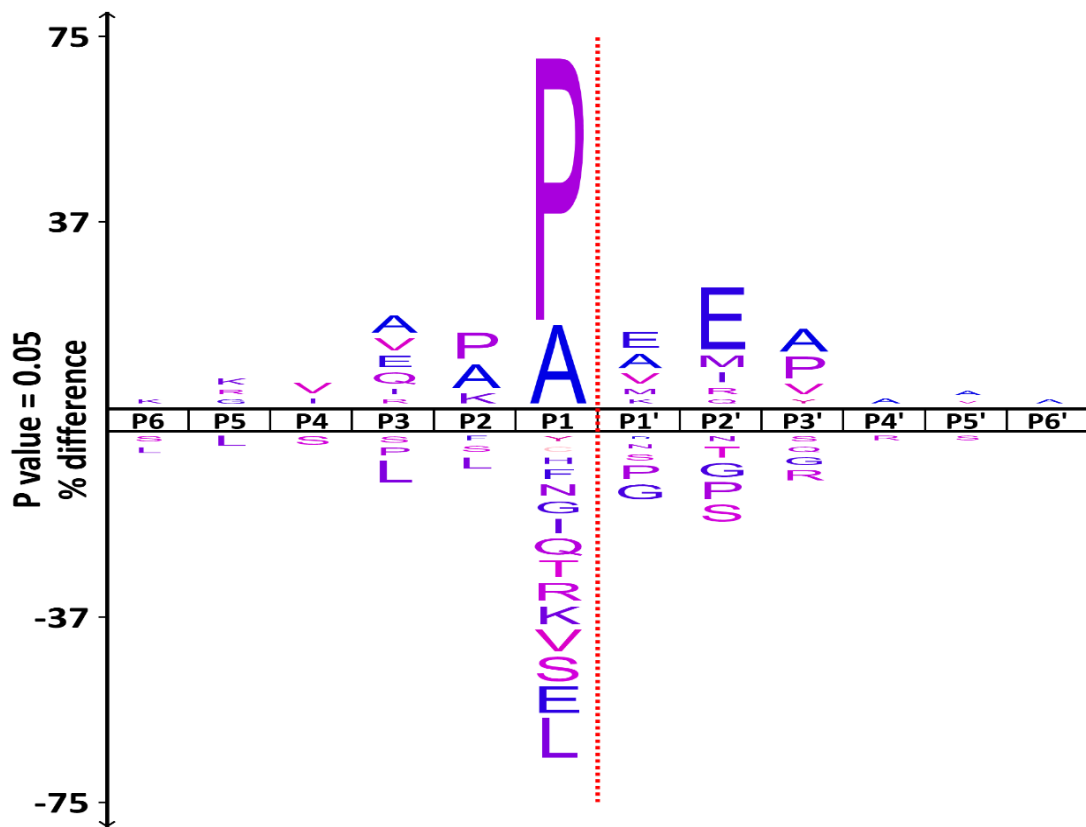


Fig. 29. Cleavage analysis of endogenous neprosin expressed as iceLogo.

#### 4.6.2.2 Biological roles of neprosin

It has been reported that neprosin may help to improve sequence coverage for whole proteome analysis since it is strongly complementary to conventional enzymes (trypsin and LysC). Additionally, the complementarity of neprosin with other enzymes is especially apparent in the analysis of histone tails. Many diseases, such as cancer and respiratory, reproductive and cardiovascular illnesses are tightly related with histone tail-regulated epigenetic mechanisms. Neprosin provides a single enzyme alternative to GluC, AspN and

trypsin for the mass spectrometry analysis of histones specially in a near-complete profiling of H3 and H4 tails (Schröder et al. 2018).

Furthermore, neprosin, nepenthesis I and II, alone or in combination are able to cleave gluten immunogenic peptides (GIP) into non-toxic peptides (Schriemer, 2014) suggesting that *Nepenthes* enzyme supplementation can be a potential tool for coeliac disease treatment. It is reported that *Nepenthes* enzymes improve the solubilisation rate of gliadin slurries required for effective digestion in the stomach, presenting synergy in combination with pepsin. An example of this synergy is provided by the combination of pepsin with nepenthesis II result in a surprising reduction of intestinal inflammation (Rey et al. 2016).

Although some glutenases have already been presented to the market, none of the available options are likely to be effective, either due to inappropriate cleavage specificity or dose restrictions (Janssen et al. 2015). One of the most promising proteases currently subjected to advanced testing is AN-PEP, a prolyl endoprotease from *Aspergillus niger*. Although it is described that AN-PEP can develop its activity within the pH constraints of the stomach, a 1000-fold higher amount than a purified mixture of *Nepenthes* enzymes was needed to achieve a comparable slurry clarification (Rey et al. 2016). Therefore, it would need to be administered in high amounts that are nearly comparable to the gluten proteins consumed (Mitea et al. 2008; Janssen et al. 2015). However, very low administration of *Nepenthes* enzymes is sufficient to enhance the solubilisation rate for gliadin slurries, which is an important condition for an effective supplementation strategy.

#### **4.6.2.3 Molecular structure of neprosin**

Neprosin is the first characterized member of what seems to be a new class of prolyl endoproteases. It is comprised of two novel DUF domains, which stands for “Domain of unknown function”, i.e. DUF 4409 and DUF239, which differ structurally and functionally from commonly known proline-cleaving enzymes (Lee et al., 2016). The DUF239 domain is well represented in plants. However, the sequence identity shared with other members of DUF239 is modest, suggesting that, this protease may have several characteristics outside the criteria of the class. Hitherto, the neprosin protein sequence has not appeared to have a homologous to any other known protein in the genomic databases.

## 5. Protein-protein interaction of TGF $\beta$ 2 with $\alpha$ -2 M

---

### 5.1. TGF $\beta$ Overview

The transforming growth factor beta (TGF $\beta$ ) signalling pathway is crucial for a variety of cell functions and was thought to emerge with the development of metazoans. TGF $\beta$  is implicated in numerous developmental processes such as the induction of epithelial to mesenchymal transition in endocardial cells that is required for correct heart development (M. Y. Wu and Hill 2009). Additionally, TGF $\beta$  plays several roles in tissue homeostasis, regulating various functions including cellular differentiation, cell-cycle arrest, extracellular matrix (ECM) production, apoptosis, and cellular migration. Partly due to its pleiotropic effects in a large number of cell-types, it is also involved in pathologies such as fibrosis and cancer. In cancer, TGF $\beta$  seems to have a dual function: on the one side, TGF $\beta$  can promote tumoral cell migration, invasion and immune evasion; on the other hand, it is a tumor-suppressor, enhancing cell-cycle arrest and apoptosis (Padua and Massagué 2009). Another important function is wound healing, allowing wound closure through the production of ECM proteins as the inhibitor of matrix metalloproteases. In contrast, in fibrotic diseases, excessive TGF $\beta$  expression and signalling provoke extensive tissue fibrosis, which jeopardises normal tissue function.

### 5.2. TGF $\beta$ cytokines

The TGF $\beta$  family is comprised by structurally and functionally related cytokines which interact with Serine/Threonine kinase receptors on the cell surface in order to mediate downstream transcriptional events through the canonical Smad signalling pathway. This family is composed by over 30 ligands including the TGF $\beta$ s, bone morphogenetic proteins (BMPs), the growth and differentiation factors (GDF) inhibins, activins nodal, and anti-Mullerian hormone proteins (Schmierer and Hill 2007).

There are three TGF $\beta$ s isoforms in mammals (TGF- $\beta$ 1, - $\beta$ 2, and - $\beta$ 3) which are observed abundantly during development and display overlapping and different temporal and spatial expression patterns. The structural homology amongst TGF $\beta$  isoforms are really high, especially at putative cleavage and integrin binding regions (Nixon, Brower-Toland, and Sandell 2000). They are ubiquitously expressed and pleiotropically engaged in the physiology of nearly all tissue and cell type activations, thus involved in a diverse array of physiological

and pathological processes, by binding to TGF $\beta$  receptors of type I, II and III (TGFR-I, -II and -III) (Travis and Sheppard 2014). Some of these effects are bi-directional. For instance a low concentration of TGF $\beta$  promotes the stimulation of mesenchymal cell proliferation while high amounts result in inhibition (Battegay et al. 1990). This emphasizes the relevance of local tissue activation in determining its cellular effects.

### 5.2.1 Processing of TGF $\beta$

TGF $\beta$  ligands are secreted as a high molecular weight latent complex requiring activation to trigger its biological activity (Gleizes et al. 1997). Three distinct types of latent TGF $\beta$  are known: the small latent complex, the large latent complex and a form that is associated with the protease inhibitor  $\alpha_2$ -macroglobulin (also associated to TGF $\beta$ 2; Lawrence 2001).

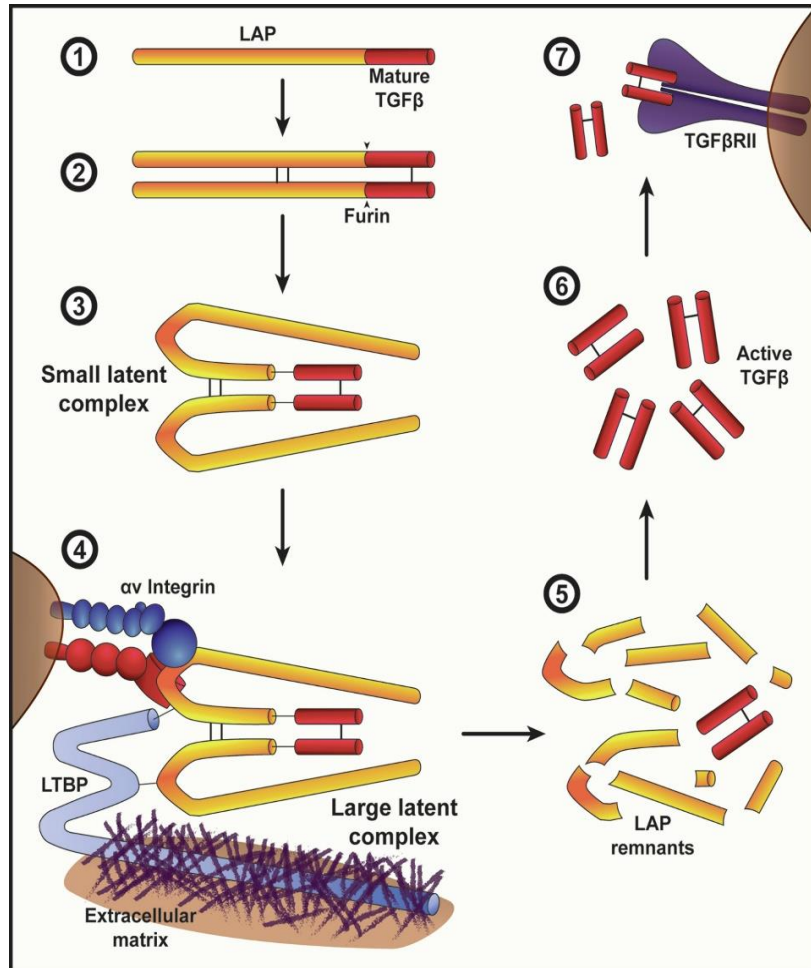
TGF $\beta$  proteins are biosynthesized as a precursor that encompasses a signal peptide and a pro-peptide, also called latency-associated peptide (LAP), and a mature peptide region at the carboxy-terminal (Wakefield et al. 1988; Fig. 30).

The first processing event occurs in the endoplasmic reticulum (E.R) through cleavage of the pre-region by endopeptidases resulting in a dimer which is linked by disulphide bonds and referred to as pro-TGF $\beta$ . Secondly, the pro-peptide is cleaved in the Golgi apparatus by a typical RXXR furin-like protease resulting in two disulphide-bonded homodimers: A N-terminal 70-85 kDa LAP and mature region also termed growth factor (GF) of 25 kDa. However, the two parts remain strongly non-covalently associated, and in dimeric state constitute the small latent complex (SLC; Miyazono, Ichijo, and Heldin 1993). This formation is required for proper folding and dimerization of the GF domain which folds concomitantly to the N-terminal prodomain (Gray and Mason 1990; Walton et al. 2009). Additionally, LAP renders GF inactive, thus preventing to bind to its receptor.

SLC can be secreted from cells or remain in the cell associated with a glycoprotein called latent TGF $\beta$  binding protein (LTBP) that binds through a single disulphide bond to LAP (Olofsson et al. 1992; Robertson et al. 2015). This complex is termed large latent complex (LLC). Although LTBP does not confer latency, it is required for efficient extracellular secretion (Miyazono et al. 1991). LTBP works as an archer latent complex for the extracellular matrix until subsequent activation. (Nunes et al. 1997).

In brief, after secretion, these complexes are targeted to fibrillin-rich microfibrils as inactive species that are covalently bound by tissue transglutaminase to the extracellular

matrix for storage. Additionally, it is also found on the surface of immune cells or in granules of platelets and mast cells (Jenkins 2008). Thus, localization and compartmentalization provide tight spatial and temporal regulation of the GFs (Rifkin 2005).



**Fig.30. Multistep process enabling the release of active TGFβ.** (1) TGFβ is biosynthesized as a propeptide comprised of latency-associated peptide (LAP) and the mature TGFβ domain. (2) Dimerization of the propeptide. Furin mediated cleavage of LAP to yield the inactive small latent complex. (3) LAP undergoes a conformational change to around the mature TGFβ while remaining non-covalently associated, blocking access to the mature cytokine. (4) The SLC can then interact with LTBP to form the secreted LLC, allowing associations with the ECM to anchor TGF outside the cell (5 and 6). Integrins are able to bind LAP at an arginine-glycine-aspartic acid (RGD) site, allowing dissociation of LAP and liberation of active TGFβ. (7) Subsequently active TGFβ unleashes the signalling in cells via TGFβRII receptor binding. Figure reproduced from (Kelly et al. 2017) .

### 5.2.2 Mechanism of activation

The synthesis and retention of the TGF $\beta$  latent complex allows tight temporal and spatial regulation of the active form of the molecule. However, TGF $\beta$  must be released from the complex to achieve its biological activity (Gleizes et al. 1997) through a process termed activation. All three TGF $\beta$  active ligands are homodimers stabilized by disulphide bonds and hydrophobic interactions (Y. Shi and Massagué 2003). Latent TGF $\beta$  can be activated by a variety of physical proceedings such as acid, heat, reactive oxygen species, and biological processes including integrin mediated activation or proteolysis. GF can be dissociated from LAP in vitro conditions, including acidic pH, in the presence of urea and SDS and by heating to high temperature (Brown et al. 1990; Lawrence 1996; Munger et al. 1997). However, activation of TGF $\beta$  in the extracellular matrix may represent an essential regulatory mechanism because the lack of regulation in the activation process may be implicating the development of numerous diseases such as the Marfan's Syndrome, Camurati-Engelmann disease, organ fibrosis, malignant and autoimmune diseases.

Proteases play a central role in TGF $\beta$  activation on a range of levels, affecting bioavailability of the molecule. Therefore, numerous proteases such as thrombin, plasmin, MMP2, MMP9 and elastase have been shown to be able to activate latent TGF $\beta$  *in vitro* (Karsdal et al. 2002; Taipale, Koli, and Keski-Oja 1992; Sato and Rifkin 1989). In addition to proteases, glycosidases have also been demonstrated to be able to activate TGF $\beta$  including calpain II, endoglycosidase F1, sialidase, cathepsin D, and PNGase F, among others (Abe, Oda, and Sato 1998; Lyons, Keski-Oja, and Moses 1988; Miyazono and Heldin 1989).

Studies using genetically modified mice, revealed that the non-proteolytic activation by thrombospondin-1 (TSP1; Crawford et al. 1998) and  $\alpha v\beta 6$  integrin (Munger et al. 1999) greatly play a part in TGF $\beta$  activation. It is well appreciated that integrins can bind to LCC to release active TGF $\beta$ , but the underlying mechanism remains unclear. However, two mechanisms of integrin mediated activation have been proposed. The first one is based on  $\alpha v\beta 5/\alpha v\beta 6$  integrins which bind to an Arg-Gly-Asp (RGD) integrin recognition motif present in the LAP domain (M. Shi et al. 2011). A tension exerted by the cytoskeleton is converted in conformational changes that allow the liberation of TGF $\beta$  and subsequently bind to its receptor (Fontana et al. 2005). The second mechanism requires the activity of proteases in order to release its active form. For example integrin  $\alpha v\beta 8$  needs the cleavage of

LAP by MT1-MMP (Mu et al. 2002). Furthermore, active TGF $\beta$  promotes the synthesis of a variety of its own activator including furin, MMPs, TSP1 and  $\alpha$ v $\beta$ 6.

### 5.3 Soluble TGF $\beta$ binding proteins

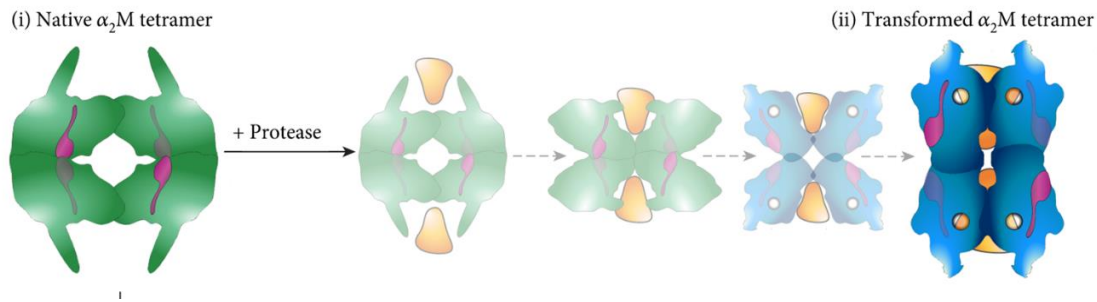
TGF $\beta$  has been described to bind to a variety of proteins, such as  $\alpha$ <sub>2</sub>-macroglobulin ( $\alpha$ <sub>2</sub>M) fibronectin, a soluble form of the TGF $\beta$  type III receptor, type IV collagen and thrombospondin and P-amyloid precursor protein among others. After release from the latent TGF $\beta$  LAP, GF may bind to  $\alpha$ <sub>2</sub>M *in vivo* preventing the binding of TGF $\beta$  to its receptors (principally type I and III).

#### 5.3.1 Alpha-2 macroglobulin ( $\alpha$ <sub>2</sub>M)

Human  $\alpha$ <sub>2</sub>M (h $\alpha$ <sub>2</sub>M) is an abundant homotetrameric plasma protease which acts as a molecular trap for a high variety of proteases (Barrett 1981). It is principally secreted by hepatocytes, but it also synthesized by lung fibroblasts, macrophages and peripheral tissues of the adrenal gland (Negoescu et al. 1994). h $\alpha$ <sub>2</sub>M regulate proteolysis in complex biological process, including signalling, blood homeostasis, tissue remodelling and defence against toxins among others (Garcia-Ferrer et al. 2017). However, h $\alpha$ <sub>2</sub>M is also found to be involved in many other functions that are not related with peptidase-binding and inhibition. Indeed, it is observed that this is related to the innate immunity proteins of host defence, especially interacting with the complement system as well as influencing the modulation of the activity of cytokines, growth factors, misfolded proteins, and lipid factors, promoting a great impact on human physiology (Armstrong and Quigley 1999).

Two variants of h $\alpha$ <sub>2</sub>M have been described: a slow mobility form capable of binding proteases (native state also named opened conformation) and a fast mobility form resulting from protease binding (induced or closed conformation) (Fig. 31).

The native form is the one circulating in blood, whereas the fast form shows an irreversible state upon protease binding (Van Leuven, Cassiman, and Van Den Berghe 1979; Barrett 1981). Subsequently, it can bind to specific receptors termed LRP1, and exhibited on many cells and thus generating a protease clearance from circulation via an internalization process.



**Fig. 31. Schematic representation of a conformational change of  $\alpha_2$ -macroglobulin.** Native  $\alpha_2$ M (shown in green) causes a conformational change due to protease-binding resulting in a transformed or induced  $\alpha_2$ M form (shown in blue). Figure reproduced from (Cater, Wilson, and Wyatt 2019).

#### 5.4 Biological interactions of $h\alpha_2$ M/TGF $\beta$

Mostly, TGF $\beta$ 1 and TGF $\beta$ 2 exert non-covalent and reversible binding to  $h\alpha_2$ M. This binding is distinct from the protease-binding mechanism since it does not induce an  $h\alpha_2$ M conformational change, meaning TGF $\beta$  on their own cannot effectuate the conversion from the native to the induced form. However, the non-covalent complex formed can be dissociated by heparin which is produced by numerous cell types, thus participating in the release of active TGF $\beta$  (McCaffrey et al. 1989).

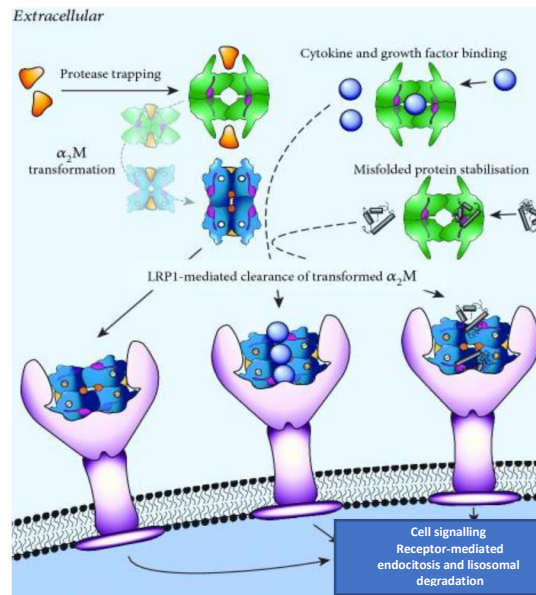
Contrarily, there are other mechanism for cytokines to bind  $h\alpha_2$ M such as covalent interaction through thiol-disulphide exchange or cytokine binding though the combination with a protease which promotes available thiol ester for reaction (Feige et al. 1996). It has been shown that TGF $\beta$ 1 and TGF $\beta$ 2 can bind to the induced form presenting similar affinity, while TGF $\beta$ 2 displays 30-fold higher affinity compared to TGF $\beta$ 1 for binding to the native form (Crookston et al. 1994).

To clarify, native  $h\alpha_2$ M may stabilized and protect cytokines from protease cleavage playing a role as a protective carrier protein (Feige et al. 1996). Therefore,  $h\alpha_2$ M is involved in the transporting of TGF $\beta$ , allowing this factor to act as an autocrine regulator of adrenocortical steroidogenic processes. Furthermore, induced  $h\alpha_2$ M is able to bind  $h\alpha_2$ M receptor/LRP enabling any associated TGF $\beta$  to be cleared via the internalization pathway (LaMarre et al. 1991) (Fig. 32).

Previous biochemical data had revealed that  $h\alpha_2$ M bind TGF $\beta$ 2 principally through the mature segment. Sequences from residues (N<sup>342</sup>ANFCAGA<sup>349</sup>) (UP entry P61812) and particularly N<sup>342</sup>, A<sup>347</sup> and A<sup>349</sup> residues play an indispensable role in determining the mature



TGF $\beta$ 2 affinity for native h $\alpha_2$ M (Webb et al. 1994). Moreover, the W<sup>354</sup> residue plays a fundamental role in the hydrophobic interaction with induced h $\alpha_2$ M (Q. Liu et al. 2001).



**Fig. 32. Schematic representation of the interaction of  $\alpha_2$ M with the LRP1 receptor.** After induction by proteases,  $\alpha_2$ M exposes the RBD domain, which is recognized by the LRP1 receptor, thus providing the clearance of proteases and all other interacting proteins that bound such as TGF $\beta$ . Figure reproduced from (Cater, Wilson, and Wyatt 2019).

# **OBJECTIVES**

---



This thesis focused on the biochemical, biophysical, functional and structural characterization of three glycoproteins, which possess an important biomedical relevance in cardiovascular disorders and celiac disease treatment.

When I started my PhD projects, the laboratory had already some basic expertise in glycoprotein production in insect cells (*Drosophila Schneider 2*) and in adherent mammalian HEK 293T cells, but yields were typically too low to allow for an in-depth characterization by structural biochemistry. Thus, my first major task was to establish and optimise in collaboration with a senior Ph.D. student of the laboratory, Laura Mariño Puertas, the recombinant expression of proteins in a suspension-based mammalian system, namely Expi293F cells, which I then utilized for my three major research projects. These focused on (i) the metallopeptidase ADAMTS13 in its complex with VWF, (ii) the prolyl-endopeptidase neprosin, and (iii), transforming growth factor  $\beta$ 2 (TGF $\beta$ 2).

(1) Since its discovery, ADAMTS13 has triggered a wave of extensive studies on its molecular structure, its mode of action, its implication in pathology and its potential therapeutic function in cardiovascular disease. We focused on the molecular mechanisms of interaction between ADAMTS13 and VWF, to finally understand the underlying working mechanism of binding and cleavage, as well as to provide a blueprint for rational drug design for therapeutic interventions. Thus, the **first project** of the present thesis (**Project 1**) aimed at two overarching goals:

- The establishment of a high yield expression system and a protein purification platform to produce sufficient protein for biochemical and biophysical studies of relevant fragments of both binding partners.
- The understanding of the interaction mechanism using a combination of biochemical, biophysical and structural techniques.
- An integrative structural analysis of conformational changes involved in ADAMTS13 function through the integrative analysis of atomic models and in-solution analysis by structural proteomics.

(2) Prolyl endopeptidases are prime pharmaceutical targets in the treatment of coeliac disease. Intriguingly, potent glutenases from carnivorous pitcher plants have been described to show a high efficiency in detoxification of gluten proteins and thus, they have a potential use in enzyme supplementation strategies. Importantly, neprosin is a scarcely described enzyme, defining a new class of prolyl endopeptidase. In addition, large scale recombinant

protein expression of neprosin has failed in the past, and only production platforms producing minute amounts are so far documented. Consequently, the **second project** of the presented thesis (**Project 2**) focused on the following goals:

- Establishment of a high-yield expression system and a robust protein purification platform to produce sufficient protein for biochemical and biophysical studies.
- Comprehensive characterization of the proteolytic mechanism of neprosin, its active site residues, and how it recognizes and cleaves the gluten-derived 33-mer immunogenic peptide. Furthermore, we aimed to identify the exact cleavage sites and analyse the resulting cleavage products as well as possible protease inhibitors.
- The identification of key structural elements important for substrate binding and proteolytic function through mutational studies and structure-function analysis.
- Evaluation and comparison of gluten-derived cleavage products (gliadin and gluten-derived 33-mer immunogenic peptide) by neprosin vs. pepsin using *ex vivo* functional assays (e.g. intestinal human cells and a coeliac mouse model).

(3) It is well described in the literature, that the activity of transforming growth factor  $\beta 2$  is modulated by  $\alpha_2$ -macroglobulin. However, the interaction between  $\alpha_2$ -macroglobulin and TGF $\beta 2$  is still ill-defined, even though three dimensional structures of both interaction partners are available. Additionally, this protein-protein interaction may also represent a good model system to understand the molecular mechanism between behind the interaction of  $\alpha_2$ -macroglobulin and its various ligand proteins. Thus, the **third project** of the present thesis (**Project 3**) had two main goals:

- The biochemical, biophysical and structural characterization of the molecular interactions between latent TGF $\beta 2$  and distinct human  $\alpha_2$ -macroglobulin variants, either produced native from human serum or recombinantly in *Drosophila* Schneider 2 and Expi293F systems.
- The determination of three-dimensional structure of TGF $\beta 2$  by X-ray crystallography.

## **RESULTS**

---



F.Xavier Gomis-Rüth  
Full Professor CSIC  
Dept. of Structural Biology, Director

## INFORME DEL DIRECTOR DE TESI

2 de febrer de 2021

Per la present vull confirmar que la Laura del Amo Maestro ha contribuït de forma excepcionalment significativa a la ciència durant la seva tesi, que ha donat lloc a la publicació dels dos articles següents:

L. del Amo-Maestro, L. Marino-Puertas, T. Goulas & **F.X. Gomis-Rüth** \* (2019). Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor  $\beta$ 2 in a new conformation. *Sci. Rep.*, **9**, 8660.

L. Marino-Puertas, L. del Amo-Maestro, M. Taulés, **F.X. Gomis-Rüth** \* & T. Goulas (2019). Recombinant production of human  $\alpha$ 2-macroglobulin variants and interaction studies with recombinant G-related  $\alpha$ 2-macroglobulin binding protein and latent transforming growth factor- $\beta$ 2. *Sci. Rep.*, **9**, 9186.

En tots dos articles, la seva contribució va ser decisiva, en un d'ells és la primera autora. A més a més, també va participar de forma molt rellevant com a segona autora en el següent article publicat, que no es discuteix a la seva tesi:

S.R. Mendes, L. del Amo-Maestro, L. Marino-Puertas, I. De Diego, T. Goulas & **F.X. Gomis-Rüth** \* (2020). Analysis of the inhibiting activity of reversion-inducing cysteine-rich protein with Kazal motifs (RECK) on matrix metalloproteinases. *Sci. Rep.*, **10**, 6317.

Així mateix, és la primera autora de dos manuscrits més que sí es presenten a la seva tesi i que es troben, respectivament, en revisió per pares a la revista *Journal of Molecular Biology* i en la fase final de la seva redacció:

L. del Amo-Maestro, A. Sagar, P. Pompach, T. Goulas, C. Scavenius, D.S. Ferrero, M. Castrillo-Briceño, M. Taulés, J.J. Enghild, P. Bernadó & **F.X. Gomis-Rüth** \* (2021). An integrative structural biology analysis of von Willebrand factor binding and processing by ADAMTS-13 in homeostasis. *J. Mol. Biol.*, en revisió.

L. del Amo-Maestro, U. Eckhard, A. Rodríguez-Banqueri, S.R. Mendes, F.J. Pérez-Cano & **F.X. Gomis-Rüth** \* (2021). Structure, function, zymogenic latency and catalytic mechanism of a plant glutamate peptidase with therapeutic potential in coeliac disease. Manuscrit en preparació.

Degut a tots aquests mèrits, no hi ha cap dubte que la doctoranda ha realitzat una feina extraordinària que supera amb escreix el que és d'esperar per a una tesi doctoral.

Ben cordialment,



F. Xavier Gomis-Rüth





The **Results** obtained along the thesis belong to the three separate projects, which are presented as a compendium of publications in subsections **Project 1, Project 2 and Project 3**. The resulting publications are:

**Project 1:**

L. del Amo-Maestro, A. Sagar, P. Pompach, T. Goulas, C. Scavenius, D.S. Ferrero, M. Castrillo-Briceno, M. Taules, J.J. Enghild, P. Bernado & F.X. Gomis-Rüth (2021). An integrative structural biology analysis of von Willebrand factor binding and processing by ADAMTS13 in homeostasis. *J. Mol. Biol.*

In project 1, I carried out the protein production and purification, biochemical binding studies as well as cleavage assays. Additionally, I determined the complex affinity and kinetic parameters. I also performed chemical cross-linking assays. However, the MS analysis of CL and HDX were carried out by different collaborators. Furthermore, the integrative analysis of atomic models and in-solution analysis by structural proteomics was implemented by Pau Bernardó.

**Project 2:**

L. del Amo-Maestro, U. Eckhard, A. Rodriguez-Banqueri, S.R. Mendes, F.J. Perez-Cano & F.X. Gomis-Rüth (2021). Structure, function, zymogenic latency and catalytic mechanism of a plant glutamate peptidase with therapeutic potential in coeliac disease.

In project 2, I performed the protein production and purification, biochemical and functional protein characterization, mutant analysis, enzymatic assays, inhibition trials among others. In turn, Xavier Gomis solved the crystal structures of both the zymogen and the auto-activated mature form of neprosin.

**Project 3:**

L. del Amo-Maestro, L. Marino-Puertas, T. Goulas & F.X. Gomis-Rüth (2019). Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor  $\beta$ 2 in a new conformation. *Sci. Rep.*, 9, 8660.

L. Marino-Puertas, L. del Amo-Maestro, M. Taules, F.X. Gomis-Rüth & T. Goulas (2019). Recombinant production of human  $\alpha$ 2-macroglobulin variants and interaction studies with recombinant G-related  $\alpha$ 2-macroglobulin binding protein and latent transforming growth factor- $\beta$ 2. *Sci. Rep.*, 9, 9186.

In project 3, I carried out the optimization of TGF $\beta$ 2 expression and purification in order to perform protein-protein interaction studies with  $\alpha_2$ -macroglobulin. I accomplished the TGF $\beta$ 2 activation experiments and I collaborated with Laura Mariño in order to perform binding studies. Furthermore, Xavier Gomis solved the protein crystal structure.

# **RESULTS: PROJECT 1**

---

*“An integrative structural biology analysis of von Willebrand factor binding and processing by ADAMTS13 in homeostasis”*



## Summary

Von Willebrand Factor (VWF), a 300 kDa plasma protein which is key to homeostasis, is cleaved at a single site by the multi-domain metallopeptidase ADAMTS13 (AD13). Importantly, VWF is the only known substrate of this peptidase, which circulates in a latent form in the blood and becomes allosterically activated upon substrate binding. Herein, we characterised the complex formation by a competent peptidase variant (AD13-MDTCS) comprising the metallopeptidase (M), disintegrin-like (D), thrombospondin (T), cysteine-rich (C), and spacer (S) domains, with a 73-residue physiologically relevant VWF-peptide, using nine complementary techniques. Pull-down assays, gel electrophoresis, and surface plasmon resonance revealed tight binding with sub-micromolar affinity. Cross-linking mass spectrometry showed that within AD13, domain D approaches M, C, and S upon substrate binding. S is positioned close to M and C, and the VWF peptide exhibits contacts with all domains. Hydrogen/deuterium exchange mass spectrometry subsequently revealed strong and weak protection for C/D and M/S, respectively. Structural analysis by multi-angle laser light scattering and small-angle X-ray scattering in solution showed that AD13 adopts a highly flexible latent conformation, and upon peptide binding, active structures that all significantly differ from the published AD13-MDTCS crystal structure. Intriguingly, the VWF peptide behaved like a self-avoiding random chain. We integrated our experimental results with computational approaches, and derived a model consisting of five distinct conformations that collectively satisfy all experimental restraints. The interaction conforms to a 'fuzzy complex' that follows a so-called 'dynamic zipper' mechanism involving numerous reversible, weak but additive interactions that result in strong binding and ultimately in substrate cleavage. Our findings comprehensively illuminate the complex biochemical interplay of the VWF:ADAMTS13 axis.



# An integrative structural biology analysis of von Willebrand factor binding and processing by ADAMTS-13 in solution.

Laura del Amo-Maestro <sup>1,#</sup>, Amin Sagar <sup>2,#</sup>, Petr Pompach <sup>3,4</sup>, Theodoros Goulas <sup>1</sup>, Carsten Scavenius <sup>5</sup>, Diego S. Ferrero <sup>6</sup>, Mariana Castrillo-Briceño <sup>1</sup>, Marta Taulés <sup>7</sup>, Jan J. Enghild <sup>5</sup>, Pau Bernadó <sup>2,\*</sup> & F. Xavier Gomis-Rüth <sup>1,\*</sup>

<sup>1</sup> Proteolysis Laboratory; Department of Structural Biology; Molecular Biology Institute of Barcelona (CSIC); Barcelona Science Park; c/Baldiri Reixac, 15-21; 08028 Barcelona (Catalonia, Spain).

<sup>2</sup> Centre de Biochimie Structurale; INSERM, CNRS and Université de Montpellier; 34090 Montpellier (France). <sup>3</sup> Institute of Microbiology of the Czech Academy of Sciences; BIOCEV; Prumyslova 595; 252 50 Vestec (Czechia). <sup>4</sup> Institute of Biotechnology of the Czech Academy of Sciences; BIOCEV; Prumyslova 595; 252 50 Vestec (Czechia). <sup>5</sup> Department of Molecular Biology and Genetics; Aarhus University; Gustav Wieds Vej 10; 8000 Aarhus C (Denmark). <sup>6</sup> Laboratory for Viruses and Large Biological Complexes; Department of Structural Biology; Molecular Biology Institute of Barcelona (CSIC); Barcelona Science Park; c/Baldiri Reixac, 15-21; 08028 Barcelona (Catalonia, Spain). <sup>7</sup> Scientific and Technological Centers (CCiTUB); University of Barcelona; Lluís Solé i Sabaris, 1-3; 08028 Barcelona (Catalonia, Spain).

# Shared first authorship.

\* Corresponding authors; e-mails: pau.bernado@cbs.cnrs.fr and xgrcri@ibmb.csic.es.

*Keywords:* protein-protein interactions; metallopeptidase; substrate cleavage; SAXS; cross-linking; H/DXMS; surface plasmon resonance; platelet aggregation; blood coagulation.

*Short title:* ADAMTS-13 and von Willebrand factor in haemostasis.

---

Von Willebrand Factor (VWF), a 300-kDa plasma protein key to homeostasis, is cleaved at a single site by multi-domain metallopeptidase ADAMTS-13. VWF is the only known substrate of this peptidase, which circulates in a latent form and becomes allosterically activated by substrate binding. Herein, we characterised the complex formed by a competent peptidase construct (AD13-MDTCS) comprising metallopeptidase (M), disintegrin-like (D), thrombospondin (T), cysteine-rich (C), and spacer (S) domains, with a 73-residue functionally relevant VWF-peptide, using nine complementary techniques. Pull-down assays, gel electrophoresis, and surface plasmon resonance revealed tight binding with sub-micromolar affinity. Cross-linking mass spectrometry with four reagents showed that, within the peptidase, domain D approaches M, C, and S. S is positioned close to M and C, and the peptide contacts all domains. Hydrogen/deuterium exchange mass spectrometry revealed strong and weak protection for C/D and M/S, respectively. Structural analysis by multi-angle laser light scattering and small-angle X-ray scattering in solution revealed that the enzyme adopted highly flexible unbound, latent structures and peptide-bound, active structures that differed from the AD13-MDTCS crystal structure. Moreover, the peptide behaved like a self-avoiding random chain. We integrated the results with computational approaches, derived a model of structures that collectively satisfied all experimental restraints, and discussed the functional implications. The interaction conforms to a ‘fuzzy complex’ that follows a ‘dynamic zipper’ mechanism involving numerous reversible, weak but additive interactions that result in strong binding and cleavage. Our findings contribute to illuminating the biochemistry of the VWF:ADAMTS-13 axis.

Von Willebrand factor (VWF) is a large, abundant multimeric plasma protein mediating blood homeostasis [1, 2]. It is synthesized by endothelial cells and megakaryocytes as a polyglycosylated 2813-residue multi-domain pre-protein with a signal peptide, a pro-region, and a mature region. The latter encompasses domains D<sup>1</sup>-D3-A1-A2-A3-D4-C1-C2-C3-C4-

C5-C6-CK [3] and is secreted into the blood as multimers ranging from dimers to ultra-large ~100-mers (ULVWF). These particles possess potent haemostatic activity and are attached to the endothelium at sites of vascular injury as VWF strings that bridge the subendothelial collagen matrix and receptor complex GPIb-IX-V on the surface of platelets to set off plug formation [2,



4]. Under homeostatic conditions, ULVWF is specifically processed to smaller multimers with lower adhesive activity by metallopeptidase (MP) ADAMTS-13 (Fig. 1a), for which VWF is the only known substrate [5-8]. Impaired ADAMTS-13 function leads to excess VWF activity, which induces thrombi due to aberrant agglutination of platelets in the absence of vascular lesions (Fig. 1a) [2, 9]. Thrombi cause ischaemic stroke, myocardial infarction, thrombotic thrombocytopenic purpura and pulmonary embolism, among other cardio- and cerebrovascular conditions, which collectively are responsible for one in four deaths worldwide [10-14]. By contrast, excessive processing of VWF leads to the bleeding disorders Heyde's syndrome and von Willebrand disease [15-19].

ADAMTS-13 is a soluble 1398-residue multi-domain protein containing a catalytic domain (domain M, residues A<sup>75</sup>-A<sup>293</sup>; numbering in superscript according to UniProt [UP] entry Q76LX8). It is preceded by a signal peptide and a pro-domain (P<sup>30</sup>-R<sup>74</sup>) that is removed by furin-type cleavage [20]. In contrast to most metallopeptidases [21], the pro-domain is not required for latency [20]; instead, the unbound enzyme is inactive and only becomes allosterically activated by substrate binding [22, 23]. This latency explains the long half-life of ADAMTS-13 of up to a week in the circulation [24]. The M domain contains the characteristic zinc-binding consensus motif of metzincin MPs (H<sup>224</sup>-E-X-X-H-X-X-G-X-X-H<sup>234</sup>) for metal binding and catalysis [25], and mutation of the general base glutamate to glutamine (E<sup>225</sup>Q) renders the enzyme incapable of cleaving VWF [26]. The M domain is followed by a disintegrin-like domain (D; G<sup>294</sup>-P<sup>379</sup>), a thrombospondin-type 1 repeat (T; I<sup>380</sup>-E<sup>439</sup>), a cysteine-rich domain (C; K<sup>440</sup>-C<sup>555</sup>), a spacer domain (S; S<sup>556</sup>-P<sup>682</sup>), a further seven T-like repeats, and two C-terminal CUB domains [23, 27, 28].

Cleavage of VWF by ADAMTS-13 occurs within domain A2 (D<sup>1498</sup>-V<sup>1665</sup>; VWF residue numbers in subscript according to UP P04275), exclusively at the Y<sup>1605</sup>-M<sup>1606</sup> bond [29]. Notably, the crystal structure of isolated A2 reveals a globular domain in which the cleavage site is not accessible [30]. Indeed, turbulent flow in the bloodstream and shear stress during VWF string-mediated binding and agglutination of platelets triggers unfolding of A2 to expose the scissile bond [26, 31, 32]. Remarkably, the interaction between A2 and ADAMTS-13 exceeds the M domain because the minimal length of a VWF-derived fragment that mimics the unfolded state

of A2 and is cleaved spans 73 residues (D<sup>1596</sup>-R<sup>1668</sup>; VWF-peptide, see [33]). Thus, exosites of ADAMTS-13 domains downstream of M must participate in substrate recognition and binding, which explains the strict substrate specificity of the enzyme [23, 28, 34]. Recognition of E<sup>1660</sup>-R<sup>1668</sup> by S, followed by binding of D<sup>1614</sup> by D through R<sup>349</sup> and of L<sup>1603</sup> by the active-site cleft of ADAMTS-13 places the scissile bond for cleaving in what has been dubbed a 'molecular zipper' model [34, 35].

Currently, structural information for ADAMTS-13 at the atomic level is available for the non-catalytic DTCS domains [36] and for a construct spanning domains MDTCS (AD13-MDTCS). The latter is in a substrate-unbound, latent conformation with an occluded active site, which was crystallised with a monoclonal antibody antigen-binding fragment (Fab) owing to the intrinsic flexibility of the molecule [23]. Moreover, structural information based on the M-domain coordinates of ADAMTS-4 and those of the aforementioned non-catalytic domains of ADAMTS-13 has been derived for various constructs in solution by small-angle X-ray scattering (SAXS), which has illuminated how the full-length protein is autoinhibited [22, 28, 37].

Herein, the inability to obtain a crystal structure of active ADAMTS-13 in complex with a VWF-derived fragment led us to perform an integrative structural biology approach. We developed high-yield recombinant production systems in human cells for highly active AD13-MDTCS and its inactive AD13-MDTCS-E<sup>225</sup>Q mutant. We comprehensively characterised its complex with VWF-derived peptides in solution by pull-down assays, gel electrophoresis (including western blotting), proteolytic assays, surface plasmon resonance (SPR), chemical cross-linking followed by mass spectrometry (MS) (CLMS), hydrogen/deuterium exchange MS (H/DXMS), multi-angle laser light scattering (MALLS) after size-exclusion chromatography (SEC), SAXS, and biocomputing. The integration of the data enabled us to propose a working model of the competent complex in solution.

## MATERIALS AND METHODS

**Expression and purification of recombinant AD13-MDTCS and AD13-MDTCS-E<sup>225</sup>Q** — A cDNA encoding AD13-MDTCS fragment A<sup>75</sup>-A<sup>685</sup> was amplified from plasmid pcDNA4/TO [38] and cloned between *Sma*I and *Not*I (Thermo Fisher Scientific) restriction sites of a modified pCMV-SPORT6

vector using three forward primers for consecutive PCR amplifications (5'-CTGGGTTCCAGGTTCACCTGGTGACGCTGCAGGCGGCATCCTA-3', 5'-TCCTGCTATGGGTA CTGCTGCTCTGGGTTCAGGTT-3', and 5'-ATGGAGACAGACACACTCCTGCTATGG GTA-3') and reverse primer 5'-GCATGCGGCCGCAAGCTTATTAG-3' (all purchased from Sigma-Aldrich). The resulting pS6-AD13-MDTCS plasmid contained the signal peptide from the V-J2-C region of a mouse Ig  $\kappa$ -chain instead of the native leader sequence, and a C-terminal human rhinovirus 3C proteinase cleavage site plus a hexahistidine (His<sub>6</sub>)-tag. Thus, the resulting protein comprised residues D+A<sup>75</sup>-A<sup>685</sup>+LEVLFQGP HHH HHH. This construct was used to generate A13-MDTCS-E<sup>225</sup>Q (plasmid pS6-AD13-MDTCS-E225Q) via a QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies), forward primer 5'-AGTCACCATGCCCATCAGATTGGGCACAG-3', and reverse primer 5'-CTGTGCCCAATCTGATGGGCAATGGTGACT-3'.

PCR amplifications were performed with Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific), plasmids were purified with the GeneJET Plasmid MaxiPrep Kit (Thermo Fisher Scientific), and constructs were verified by DNA sequencing.

Human Expi293-F cells (Thermo Fisher Scientific) adapted to FreeStyle F17 expression medium (Gibco) were grown at 37°C to a density of 3–5×10<sup>6</sup> cells/mL in a Multitron Cell Shaker Incubator (Infors HT) at 150 rpm in a humidified atmosphere with 8% CO<sub>2</sub>, and sub-cultured every 3–4 days by dilution to 0.3–0.5×10<sup>6</sup> cells/mL. Cells were then sub-cultured to 0.7×10<sup>6</sup> cells/mL in medium containing penicillin/streptomycin, 0.5 µg/mL amphotericin, 8 mM L-glutamine, and 0.2% Pluronic F-68 (Gibco). After 24 h, cells at a density of 1×10<sup>6</sup> cells/mL were transfected with 1 mg vector DNA and 3 mg linear 25 kDa polyethyleneimine (Polysciences Europe) in 20 mL Opti-MEM medium (Gibco), previously incubated at room temperature for 15–20 min, per litre of expression medium. After 3 days, the cell culture supernatant was harvested, cleared at 4°C by centrifugation at 3,500×g for 30 min, and filtered through a 0.45 µm cellulose acetate filter (Millipore).

The supernatant was then supplemented with 15 mM imidazole, incubated with nickel-nitrilotriacetic acid resin (Ni-NTA; Invitrogen) for 3–4 h, loaded onto a Poly-Prep open column (Bio-Rad) for batch affinity chromatography purification, and washed extensively with 50 mM

Tris·HCl pH 7.4, 250 mM sodium chloride, 50 mM L-arginine, 50 mM L-glutamine, 20 mM imidazole. The protein was eluted with the same buffer containing 300 mM imidazole, and fractions were pooled and desalted with the same buffer without imidazole using a PD10 column (Sigma-Aldrich). The sample was concentrated and subjected to SEC on a Superdex 200 10/300 column (GE Healthcare) attached to an ÄKTA Purifier liquid chromatography system (GE Healthcare) operated at room temperature. Protein purity and identity were assessed by 12% glycine SDS-PAGE stained with Coomassie Brilliant Blue (Thermo Fisher Scientific), as well as peptide mass fingerprinting and N-terminal sequencing at the Protein Chemistry Service and the Proteomics Facility of the Centro de Investigaciones Biológicas (Madrid, Spain), respectively. Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 2 filter devices with a 10 kDa cutoff (Sartorius Stedim Biotech). Approximate protein concentrations were determined by measuring the absorbance at 280 nm (A<sub>280</sub>) with a spectrophotometer (NanoDrop), and applying the theoretical extinction coefficient. Purified proteins were concentrated and dialysed against 10 mM HEPES pH 7.4, 150 mM sodium chloride for subsequent studies.

#### **Expression and purification of recombinant VWF-peptide and VWF-strep-peptide**

— A synthetic gene (GeneScript) encoding VWF-peptide (D<sub>1596</sub>–R<sub>1668</sub>), codon-optimised for expression in *Escherichia coli*, was amplified with forward primer 5'-ATGCCCA TGGAAGATCGTGAGCAA-3' and reverse primer 5'-GCATCTCGAGTTAACGTTGCA GAACCA-3', and introduced into the pCri-6b vector [39] between *Nco*I and *Xho*I restriction sites. The resulting plasmid (pCri6b-VWF-pep) conferred kanamycin resistance and attached an N-terminal His<sub>6</sub>-tagged glutathione S-transferase moiety plus a tobacco etch virus endopeptidase (TEV) cleavage site. An extra C-terminal Strep-tag was introduced by consecutive PCR amplifications to generate VWF-strep-peptide (plasmid pCri6b-VWF-strep-pep) using forward primer 5'-ATGCCCATGGAAGATCGTGAGCAA-3' and reverse primers 5'-TTCGAATTGTG GATGGCTCCAACCTCCACGTTGCAGAAC CAGGT-3', 5'-ACTTCCACCTCCAGAACCCTC ACCCTTTTCGAATTGTGGATGG-3', 5'-A ATTGTGGATGGCTCCATGCGCTACCTCC ACTTCCACCTCCAGAA-3', and 5'-GCATC TCGAGCTACTTTTCGAATTGTGGATGGC TCCA-3'. The plasmids were separately

transformed into *E. coli* BL21 (DE3) cells (Novagen), which were grown at 37°C in Luria Bertani medium supplemented with 30 µg/mL kanamycin. Cultures were induced at an  $A_{600}$  of ~0.8 with 0.4 mM isopropyl- $\beta$ -D-thiogalactopyranoside and incubated overnight at 20°C. Cell cultures were centrifuged at 5,000×g for 20 min at 4°C, the pellet was washed with 50 mM Tris·HCl pH 8.0, 150 mM sodium chloride, resuspended in the same buffer plus 20 mM imidazole, and supplemented with cComplete EDTA-free inhibitor cocktail tablets and DNase I (both Roche Diagnostics). Cells were lysed with a cell disruptor (Constant Systems) at 1.35 kbar, and the cell debris was removed by centrifugation at 20,000×g for 30 min at 4°C.

The supernatant containing the fusion construct of VWF-peptide was subsequently filtered and loaded onto a 5-mL Ni-NTA HisTrap HP column (GE Healthcare), and the protein was eluted with a 20–500 mM imidazole gradient in 50 mM Tris·HCl pH 8.0, 150 mM sodium chloride. For the VWF-strep-peptide, the Ni-NTA step was replaced with affinity chromatography using Strep-Tactin XT Superflow Suspension resin (IBA Life Sciences), with elution buffer 100 mM Tris·HCl pH 8.0, 150 mM sodium chloride, 50 mM biotin (VWR Life Science). Subsequently, samples were dialysed overnight at 4°C against 20 mM Tris·HCl pH 8.0, 150 mM sodium chloride, 1 mM dithiothreitol (DTT) in the presence of His<sub>6</sub>-tagged TEV at an enzyme-to-sample molar ratio of 1:100. TEV cleavage left additional residues at the N-terminus of both VWF-peptide and VWF-strep-peptide, which spanned residues GAME+D<sub>1596</sub>–R<sub>1668</sub> and GAME+D<sub>1596</sub>–R<sub>1668</sub>+GGWSHPQFEKGGGSGGGSGGWSHPQFEK, respectively. Digested samples were passed several times through Ni-NTA resin, previously equilibrated each time with 50 mM Tris·HCl pH 8.0, 150 mM sodium chloride, 20 mM imidazole, to trap His<sub>6</sub>-tagged molecules, and the flow-through containing molecules without His<sub>6</sub>-tag was collected. Samples were pooled and dialysed overnight against 20 mM Tris·HCl pH 8.0 and further purified by ion-exchange chromatography in a TSK gel DEAE-2SW column (TOSOH Bioscience) operated with a 2–50% gradient of 1 M sodium chloride in 20 mM Tris·HCl pH 8.0 applied over 40 mL. Samples were collected, pooled, concentrated by ultrafiltration, and subjected to SEC on a Superdex 75 10/300 column. The peptides were further analysed by Coomassie-stained 15% tricine SDS-PAGE.

Ultrafiltration steps were performed using Vivaspin 15 and Vivaspin 500 filter devices with a 3 kDa cutoff (Sartorius Stedim Biotech).

**Pull-down assays and western blotting analysis** — VWF-strep-peptide was incubated with Strep-Tactin resin previously equilibrated in 100 mM Tris·HCl pH 8.0, 150 mM sodium chloride for 10 min at room temperature, and then loaded onto a Poly-Prep open column (Biorad). Then, AD13-MDTCS-E<sup>225</sup>Q at 0.4 mg/mL in 50 mM Tris·HCl pH 7.4, 250 mM sodium chloride was passed several times through the column. After extensive washing with 100 mM Tris·HCl pH 8.0, 150 mM sodium chloride, the complex was eluted with the same buffer plus 50 mM biotin and analysed by 12% tricine SDS-PAGE. Gels were transferred to Hybond ECL nitrocellulose membranes (GE Healthcare) and blocked for 2 hours under gentle stirring at room temperature with 50 mL phosphate-buffered saline (PBS) plus 0.1% Tween 20 and 5% bovine serum albumin (BSA). VWF-strep-peptide was detected with the Streptavidin Peroxidase Conjugated Antibody from *Streptomyces avidinii* (Sigma-Aldrich) diluted 1:1000 in PBS plus 0.1% Tween 20 and 1% BSA. Complexes were detected using a Super Signal West Pico Chemiluminescent enhanced chemiluminescence system (Pierce). Membranes were exposed to Hyperfilm ECL films (GE Healthcare).

**SEC followed by MALLS** — Binding of VWF-strep-peptide to AD13-MDTCS-E<sup>225</sup>Q was assessed at a 10:1 molar ratio by SEC-MALLS at the joint IBMB/IRB Automated Crystallography Platform, Barcelona Science Park (Catalonia, Spain) as previously described for other proteins [40, 41]. Briefly, a Dawn Helios II apparatus (Wyatt Technologies) was used, which was coupled to a SEC Superose 6 10/300 column equilibrated in 10 mM HEPES pH 7.4, 150 mM sodium chloride at 25°C. All experiments were performed in triplicate, and the ASTRA 7 software (Wyatt Technologies) was used for data processing and analysis, for which a dn/dc value typical for proteins (0.185 mL/g) was assumed.

**Fluorogenic binding assay** — VWF-strep-peptide was labelled with an amine-reactive fluorophore (sulfosuccinimidyl-7-amino-4-methylcoumarin-3-acetate; Thermo Fisher Scientific) in 10 mM HEPES pH 7.4, 150 mM sodium chloride at tenfold molar excess of reagent for 1 h at room temperature according to the manufacturer's instructions. Thereafter, the peptide was extensively dialysed against the same

buffer to remove non-reacted dye, and mixed with AD13-MDTCS-E<sup>225</sup>Q at peptidase:peptide molar ratios of 1:5, 1:10, and 1:15. Reaction mixtures were incubated for 1 h at 37°C and subsequently analysed by 12% native PAGE. Fluorescence ( $\lambda_{\text{ex}} = 345\text{--}350$  nm and  $\lambda_{\text{em}} = 440\text{--}460$  nm) was visualised by a G:BOX F3 Gel Doc System gel reader (Syngene), and gels were thereafter stained with Coomassie.

**Proteolytic cleavage of VWF-strep-peptide by AD13-MDTCS** — AD13-MDTCS and AD13-MDTCS-E<sup>225</sup>Q in 25 mM Tris·HCl pH 7.5, 250 mM sodium chloride were preincubated in a 100  $\mu\text{L}$  reaction volume with 5 mM calcium chloride and 250  $\mu\text{M}$  zinc chloride (final concentration) for 10 min at 37°C. VWF-strep peptide in 20 mM Tris·HCl pH 7.4, 150 mM sodium chloride was subsequently added at a peptidase:peptide weight ratio of 1:10 in the absence or presence of 10 mM EDTA. Reactions were stopped at specific time points by boiling aliquots in reducing/denaturing buffer, and samples were further analysed by 15% tricine SDS-PAGE with Coomassie staining.

**SPR data acquisition, processing and kinetic data analysis** — Binding studies of AD13-MDTCS-E<sup>225</sup>Q and VWF-strep-peptide were analysed using a Biacore T200 biosensor system (GE Healthcare) at the Scientific and Technological Centres of the University of Barcelona (Catalonia, Spain). Protein Strep-Tactin (IBA LifeSciences) was immobilised on the surface of the four flow cells of a CM5 series S sensor chip (GE Healthcare) through amine coupling as described by the manufacturer. Briefly, the chip was activated by injection of 50 mM *N*-hydroxysuccinimide and 200 mM *N*-ethyl-*N'*-(dimethylaminopropyl)carbodiimide at a flow rate of 10  $\mu\text{L}/\text{min}$  for 10 min. Strep-Tactin at 50  $\mu\text{g}/\text{mL}$  in 10 mM sodium acetate pH 5.0 was injected over the sensor surface at 5  $\mu\text{L}/\text{min}$  for 10 min. Following this procedure, 3000 resonance units (RU) of Strep-Tactin were immobilized on the sensor surface. Excessive reactive groups on the sensor chip were deactivated with 1 M ethanolamine hydrochloride pH 8.5 at 10  $\mu\text{L}/\text{min}$  for 10 min. VWF-strep-peptide as ligand in running buffer (10 mM HEPES pH 7.4, 150 mM sodium chloride, 0.05% Tween 20) was bound to Strep-Tactin at 25°C at low density (20 RU) and a flow rate of 5  $\mu\text{L}/\text{min}$ . AD13-MDTCS-E<sup>225</sup>Q as the analyte was dialysed against 10 mM HEPES pH 7.5, 150 mM sodium chloride, filtered, subjected to twofold serial dilution (625 nM to 20 nM), and injected at 30  $\mu\text{L}/\text{min}$  for 120 s. The

association and dissociation phases of the complex were monitored over 90 s to determine the respective rate constants  $k_a$  and  $k_d$ . After each round of measurements, the sensor chip surface was regenerated by washing with 3 M guanidine hydrochloride at 30  $\mu\text{L}/\text{min}$  for 30 s to remove the ligand. These experiments were double-referenced so that the final signal was the measured signal minus the reference signal (flow cell without ligand) and the signal at analyte concentration zero. The overall equilibrium dissociation constant ( $K_D$ ) was determined by plotting the binding responses in the steady-state region of the sensorgrams ( $R_{\text{eq}}$ ) against analyte concentrations. The constant was also calculated from the ratio of the association and dissociation rate constants ( $K_D = k_d/k_a$ ). Sensorgrams were analysed with the BIAevaluation program (version 3.1; GE Healthcare) and fitted to a 1:1 Langmuir interaction model using global curve fitting analysis. The goodness-of-fit was assessed through the  $\chi^2$  statistical parameter and the U-value.

**CLMS data acquisition and processing** — In the first set of experiments, cross-linking of the AD13-MDTCS-E<sup>225</sup>Q:VWF-strep-peptide complex was performed either with 5 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC; Pierce) plus increasing amounts of *N*-hydroxysulfosuccinimide (NHS; Sigma; 50  $\mu\text{M}$  to 50 mM) or with bis(sulfosuccinimidyl)suberate (BS3; Pierce; 1  $\mu\text{M}$  to 1 mM) in 10 mM HEPES pH 7.4, 150 mM sodium chloride. For EDC/NHS, peptidase at 15  $\mu\text{g}/\text{mL}$  and peptide at 300  $\mu\text{g}/\text{mL}$  were incubated with NHS for 15 min at room temperature. Subsequently, EDC was added to a final concentration of 5 mM and incubated for a further 2 h at room temperature. For BS3, peptidase at 50  $\mu\text{g}/\text{mL}$  was incubated with peptide at 300  $\mu\text{g}/\text{mL}$  at 4°C for 45 min. Reactions were subsequently quenched with 2 M Tris·HCl pH 8.0 to a final concentration of 50 mM. EDC/NHS- and BS3-treated samples were assessed by 12% glycine-gradient SDS-PAGE. Reaction with BS3 was further verified by western blotting (see above). The best results were obtained with 0.5 mM BS3 after 45 min at 4°C and with 5 mM or 10 mM EDC/NHS after 2 h at room temperature, respectively.

Cross-linked proteins were prepared for MS by in-gel digestion as described previously [42]. Briefly, SDS-PAGE bands corresponding to the complex were excised from the gel and dehydrated with acetonitrile. Dried bands were incubated with a 1:1 mixture of 100 mM

ammonium bicarbonate and acetonitrile for 15 min, dried in a SpeedVac vacuum centrifuge concentrator (Savant), and digested overnight with 25 ng/mL trypsin (Promega) in 30  $\mu$ L 50 mM ammonium bicarbonate. The resulting tryptic peptides were desalted with Empore C18 SPE Disks (from 3M) packed in 10  $\mu$ L pipette tips [43]. The cross-linked complex was also digested in solution. Reaction mixtures were dried in a SpeedVac vacuum concentrator (Thermo Fisher Scientific) and denatured in 100 mM ammonium bicarbonate, 8 M urea, 10 mM DTT. Subsequently, samples were alkylated with 30 mM iodoacetamide for 30 min at room temperature in the dark, diluted to a final urea concentration of 1 M with 100 mM ammonium bicarbonate, and subsequently digested with 10  $\mu$ g trypsin overnight at 37°C. The reaction was stopped through acidification of the sample to pH <2.5 with formic acid. Tryptic peptides resulting from either digestion procedure were separated by strong cationic exchange (SCX) coupled to reversed-phase chromatography. Specifically, SCX column material (PolyLC) was dissolved in 0.1% formic acid and loaded onto a pipette-tip with an Empore C18 SPE Disk. Peptides were separated with a stepwise sodium chloride gradient (50–600 mM) and subsequently eluted with 70% 0.1% formic acid in acetonitrile for analysis. The micro-purified peptides were then subjected to LC-MS/MS analysis using an EASY nLC-1000 liquid chromatography column (Thermo Fisher Scientific) connected to an Orbitrap QE+ mass spectrometer (Thermo Fisher Scientific). Samples were injected, trapped, and desalted isocratically on a fritted precolumn (100  $\mu$ m  $\times$  2 cm inner diameter) packed in-house with RP ReproSil-Pur C18-AQ 3  $\mu$ m resin. Peptides were eluted and separated on a 15 cm analytical column of 75  $\mu$ m inner diameter packed with the same resin using a 30 min gradient (5–35%) of 0.1% formic acid in acetonitrile at a flow rate of 250 nL/min. Cross-linked peptides were identified with the MeroX software [44]. The results from three independent parallel experiments were combined.

A second set of experiments was performed with disuccinimidyl glutarate (DSG; Creative Molecules) as cross-linking agent. AD13-MDTCS-E<sup>225</sup>Q and VWF-strep-peptide (1:5 molar ratio) at 1 mg/mL in 10 mM HEPES pH 7.6, 150 mM sodium chloride was mixed with freshly prepared DSG at 50-fold molar excess and incubated for 1 h at room temperature. For quantitative studies, the peptidase:peptide complex, as well as AD13-MDTCS-E<sup>225</sup>Q alone,

were incubated with a 50-fold molar excess of light (<sup>12</sup>C) or heavy (<sup>13</sup>C) disuccinimidyl adipate (DSA; Creative Molecules). After a 30 min incubation at room temperature, AD13-MDTCS-E<sup>225</sup>Q labelled with <sup>12</sup>C or <sup>13</sup>C DSA and the complex labelled with <sup>13</sup>C or <sup>12</sup>C DSA, respectively, were mixed at a 1:1 ratio.

For DSG-mediated cross-linking, samples were separated by 4–12% Bis-Tris NuPAGE, and bands corresponding to the linked complex were excised. Cysteines were reduced with 50 mM DTT and then alkylated with 100 mM iodoacetamide in the dark. AD13-MDTCS-E<sup>225</sup>Q was in-gel deglycosylated with PNGase F (New England BioLabs) at 37°C overnight, followed by digestion with trypsin at a sample:peptidase weight ratio of 20:1, also overnight at 37°C. For DSA-mediated cross-linking, proteins in solution were reduced with 10 mM DTT, alkylated with 30 mM iodoacetamide, and digested overnight with trypsin at 37°C. Peptides were injected onto a Luna Omega 5  $\mu$ m Polar C18 100Å 20 $\times$ 0.3 mm column (Phenomenex) and desalted at 20  $\mu$ L/min for 5 min. Peptides were then separated through reversed-phase chromatography with a Luna Omega 3  $\mu$ m Polar C18 100Å 150 $\times$ 0.3 mm column (Phenomenex) at 10  $\mu$ L/min using a UHPLC 1290 Infinity II chromatography system (Agilent Technologies) with a gradient sequence of 1–10% for 1 min, 10–45% for 19 min, and 45–95% for 5 min of solvent A (0.1% formic acid, 98% acetonitrile in water) in solvent B (0.1% formic acid, 2% acetonitrile in water). The column was heated to 50°C and directly connected to a 15T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) with an electrospray ion source. Cross-linked peptides were identified with the StavroX software [45], which was set to account for fixed carbamidomethylation of cysteines and variable methionine oxidation. DSG was assumed to react with N-termini, lysines, serines, threonines, and tyrosines. The mass error threshold was set to 2 ppm, and all assigned fragments were manually curated. For quantitative studies, the Links software [46] was used for identification of cross-linked peptides. Non-overlapping isotopes of peptides cross-linked with <sup>12</sup>C and <sup>13</sup>C DSA were used to calculate isotope ratios. All experiments were performed in triplicate.

**H/DXMS data acquisition and processing**  
— AD13-MDTCS-E<sup>225</sup>Q at 300 pM in 10 mM HEPES pH 7.6, 150 mM sodium chloride, with or without a 10-fold molar excess of VWF-strep-peptide, was labelled using 10-fold dilutions in 10

mM Hepes pD 7.6 in D<sub>2</sub>O plus 150 mM sodium chloride, and incubated at room temperature for 20 s, 2 min, and 20 min. Reaction mixtures were quenched by instant acidification with 1 M glycine pH 2.3, 6 M urea, 800 mM tris(2-carboxyethyl)phosphine at a 1:1 volume ratio, and immediately flash-frozen in liquid nitrogen. Prior to analysis, samples were thawed and manually injected onto a nepenthesin-2 column. After a 3 min digestion and desalting with a Luna Omega 5  $\mu$ m Polar C18 100Å 20×0.3 mm column (Phenomenex) and an isocratic pump delivering 0.4% formic acid in water at 400  $\mu$ L/min, peptides were separated on a Luna Omega 1.6  $\mu$ m Polar C18 100Å 100×1.0 mm column (Phenomenex) with a linear 10–35% gradient of solvent C (0.4% formic acid, 95% acetonitrile in water) in solvent D (0.4% formic acid, 2% acetonitrile in water) over 12 min. In-column digestion, desalting, and separation were carried out in ice bath to reduce back-exchange. Peptides were detected using a 15T SolariX XR FT-ICR mass spectrometer operating in positive mode, and data were processed with the DataAnalysis software (version 4.2; Bruker Daltonics) and the in-house DeutEX software.

**Protein disorder analysis** — The disorder propensity of VWF-strep-peptide was assessed *in silico* using POODLE-S and POODLE-L [47], PrDOS [48], RONN [49], Spritz-L and Spritz-S [50], IUPred-L and IUPred-S [51], DISpro, and iPDA [52] through the Genesilico MetaDisorder web server [53]. In addition, the secondary structure propensity of each residue was calculated by the LS2P method [54].

**SAXS data acquisition and processing** — Samples were prepared in 10 mM HEPES pH 7.4, 150 mM sodium chloride, and scattering data were collected at beamline P12 of the Petra III storage ring of the Deutsches Elektronensynchrotron (DESY) in Hamburg (Germany) at 20°C. The isolated VWF-strep-peptide was analysed at 3 and 6 mg/mL in 40  $\mu$ l volumes. Data were collected in batch mode in 40 frames of 45 ms exposure flanked by two exposures of the buffer, and were subjected to radial averaging, background subtraction, and merging using the SASFLOW pipeline [55]. For the AD13-MDTCS-E<sup>225</sup>Q:peptide complex (1:10 molar ratio), data at 1.5–6.0 mg/mL concentration were collected using size-exclusion chromatography coupled to SAXS (SEC-SAXS) and processed using the CHROMIXS program [56]. This included selection of the frames corresponding to protein and buffer, buffer

subtraction and averaging. The SAXS profiles obtained from either batch mode or SEC-SAXS were analysed using the ATSAS suite of programs [57]. Moreover, reported SAXS data for wild-type AD13-MDTCS [28] were also reanalysed. Finally, molecular mass estimates derived from the SAXS data were calculated by Bayesian inference of values calculated using MoW [58], Size&Shape [59], V<sub>c</sub> [60], and MM<sub>Qp</sub> [61]. SAXS curves for VWF-strep-peptide at 3 mg/mL and 6 mg/mL, as well as the AD13-MDTCS-E<sup>225</sup>Q:VWF-strep-peptide complex, were deposited at the SASBDB repository {Kikhney, 2020 #5621} with codes SASDKT8, SASDKU8 and SASDKV8, respectively.

### Ensemble generation for unbound latent AD13-MDTCS and fitting to SAXS data —

The atomic coordinates of latent AD13-MDTCS (PDB 6QIG; [23]) were assessed for hinge points with HingeProt [62], and alternative conformations were generated using normal mode analysis employing an elastic network model with the elNémo server [63]. The first five nontrivial modes were used, and the range between minimal and maximal perturbation (–500, +500) was sampled in 20 steps. These structures were supplied to GAJOE within the Ensemble Optimization Methods (EOM) program [64, 65], which is part of the ATSAS suite, to obtain a sub-ensemble that collectively best described the experimental SAXS data. GAJOE employs a genetic algorithm of artificial intelligence to minimize the discrepancy ( $\chi^2$ ) between the theoretical SAXS profile of the selected sub-ensemble and the experimental one according to Equation 1:

$$\chi^2 = \frac{1}{m-1} \sum_{k=1}^m \left( \frac{I_{exp}(q) - k I_{fit}(q)}{SE_{I_{exp}(q)}} \right)^2$$

(Equation 1)

where  $I_{exp}(q)$  and  $I_{fit}(q)$  are the experimental and ensemble-averaged scattering intensities, respectively,  $SE_{I_{exp}(q)}$  are the associated experimental errors, and  $k$  is a scaling factor that is optimized during calculations.

### VWF-strep-peptide ensemble generation

— EOM was used in two steps to derive a set of conformations for VWF-strep-peptide that collectively best described the SAXS data. First, a pool of 10,000 random-coil conformations was generated for the peptide using the RanCh pool generator within the EOM program. Next, a

subset of 20 conformations from this pool was selected with GAJOE based on their collective fitting to the experimental curve using Equation 1.

**AD13-MDTCS-E<sup>225</sup>Q:VWF-strep-peptide complex ensemble generation** — A starting molecular model of active AD13-MDTCS was built for subsequent calculations. To this end, the M domain of the latent structure was remodelled with the COOT program [66] to adopt an active conformation based on active ADAMTS structures and energy-minimised with the same program. In addition, a peptide spanning L<sub>1603</sub>–T<sub>1608</sub> was modelled along the active site cleft according to templates provided by other ADAMTS and MP complexes. Finally, the rest of the AD13-MDTCS structure was kept unaltered. Based on these coordinates, an initial ensemble for the complex was generated in two steps. First, various conformations of the peptidase were generated using normal mode analysis as described above for unbound latent AD13-MDTCS. Second, for each conformation of the peptidase, 1200 conformations of the peptide were generated. For these calculations, the segment of the peptide bound in the active site was fixed while the rest of the residues upstream and downstream were modelled with RanCh. This yielded a total of 120,000 structures of the complex for which the theoretical scattering profiles were computed with CRY SOL [67]. Finally, GAJOE was run for 200 cycles to select sub-ensembles of five structures from the total ensemble of complex structures. The optimized ensemble from each cycle was analysed to determine the number of structures belonging to each of the sampled normal modes of the peptidase.

**Validation of SAXS ensembles with CLMS data** — To analyse the agreement between the SAXS-derived structures of the complex and the restraints derived from CLMS, we first filtered the ensemble of complexes to retain only the peptidase conformations selected by EOM. Next, based on the observation that the conformational distribution of VWF-strep-peptide was not modified upon binding to the peptidase, we enlarged the ensemble complex by docking 3200 peptide conformers to the peptidase. Using this enlarged ensemble, we investigated whether the C $\alpha$  atoms of the cross-linked residues were within binding distance of each other in the individual structures.

**Assembling models for the AD13-MDTCS:VWF-strep-peptide complex in solution** — Based on the SAXS-derived structure ensembles of the peptidase:peptide complex, we compiled sets of five molecular models that collectively best described the CLMS and H/DXMS data. To this end, we subjected the ensemble generated for analysing the agreement with CLMS to collective optimisation using CLMS and H/DXMS data. For each of the complex structures of the pool, the distance between all the cross-linked residues measured experimentally, and the buried solvent surface area of the protected regions of AD13-MDTCS-E<sup>225</sup>Q, were computed with python scripts using MDAnalysis [68] and the Tcl scripts using VMD [69]. Subsequently, a genetic algorithm of artificial intelligence was developed using the DEAP library [70] to collectively fit the experimental data. Briefly, this algorithm searches for an ensemble of five conformations in which all the CLMS and H/DXMS constraints are fulfilled. The objective function maximised by the genetic algorithm is provided by Equation 2:

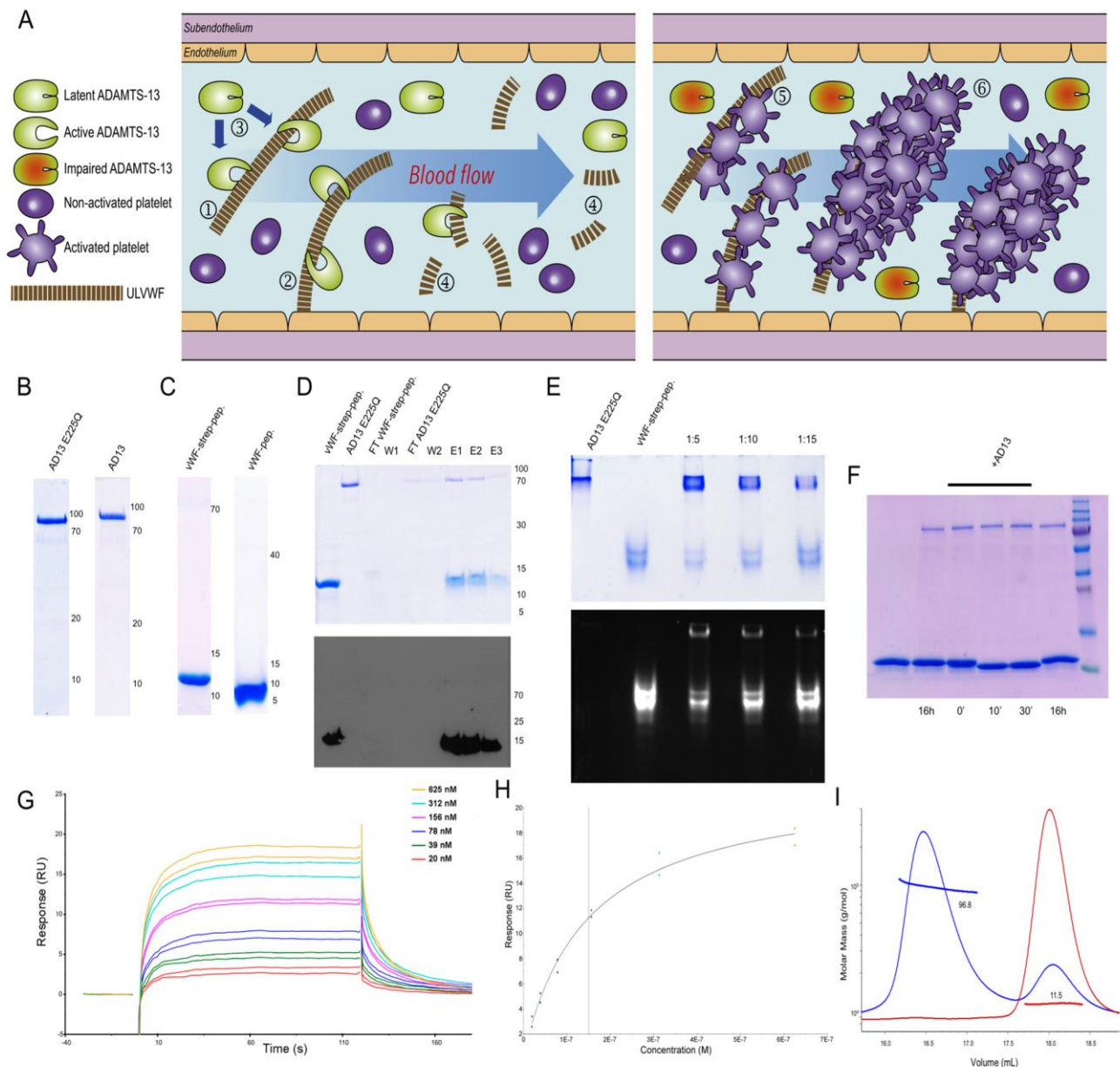
$$f = \frac{1}{n} \sum_{i=1}^n \frac{\max BS_i^{ens}}{\max BS_i^{pool}} - \frac{1}{m} \sum_{k=1}^m \frac{\min CL_k^{ens} - \min CL_k^{pool}}{\max CL_k^{pool} - \min CL_k^{pool}} \quad (\text{Equation 2})$$

where  $\max BS_i$  is the maximal buried surface area found in the selected ensemble (*ens*) or in the *pool* for site *i*,  $\min CL_k$  is the minimum distance for the cross-link *k* for the same ensembles, and *m* and *n* are the total number of experimental cross-links and protected surface areas used for selection, respectively.

## RESULTS AND DISCUSSION

**Protein production and purification, biochemical binding, and cleavage assays** — We developed efficient recombinant expression systems in human Expi293F cells, and purification schemes for AD13-MDTCS and active site point mutant AD13-MDTCS-E<sup>225</sup>Q, which were obtained in highly pure form with yields of ~1.8 and ~2.0 mg per litre of expression medium, respectively (Fig. 1b). We also produced VWF-peptide and VWF-strep-peptide, a variant with a C-terminal Strep-tag, in large quantities in a bacterial system as fusion constructs with glutathione S-transferase, which was removed after initial affinity chromatography purification (Fig. 1c). We performed pull-down assays with





**Figure 1 — Physiological function of von Willebrand factor, protein production and purification, binding studies, and biochemical assays.** (A) Scheme depicting one function of VWF in the circulation. (Left panel) Under homeostatic conditions, ULVWF multimers in solution or strings are processed by ADAMTS-13 (□ and □), which transits from a latent to an active conformation upon substrate binding (□). This creates smaller VWF multimers (□) with lower adhesive and thrombogenic activity. (Right panel) When ADAMTS-13 activity is impaired or deficient, thrombogenic ULVWFs multimers accumulate, which favours the activation of platelets (□) and augments circulating thrombi (□). (B) SDS-PAGE of pure AD13-MDTCS-E<sup>225</sup>Q (left) and AD13-MDTCS (right). (C) SDS-PAGE of pure VWF-strep-peptide (right) and VWF-peptide (left). (D) (Top panel) Pull-down assay of VWF-strep-peptide (lane 1) and AD13-MDTCS-E<sup>225</sup>Q (lane 2). The peptide was anchored to the resin until saturation (lane 3; FT, flow through) and washed (lane 4, washing step W1). The peptidase was passed through the column several times (lane 5; FT, flow through), and the resin was extensively washed (lane 6, washing step W2). The complex was then eluted with biotin-containing buffer (lanes 7–9, steps E1–E3). (Bottom panel) Western blotting of the above gel on which the VWF-strep-peptide was detected with a streptavidin antibody. (E) (Top panel) Coomassie-stained native PAGE of AD13-MDTCS-E<sup>225</sup>Q (lane 1) and fluorophore-labelled VWF-strep-peptide (lane 2), as well as incubation of both molecules at three peptidase:peptide molar ratios (lanes 3–5). (Bottom panel) The previous gel before staining visualized in a fluorescence imaging device to show fluorophore-labelled VWF-strep-peptide in the high-molecular-weight bands in complex with the peptidase. (F) VWF-strep-peptide (lane 1) was incubated with AD13-MDTCS-E<sup>225</sup>Q (lane 2), AD13-MDTCS (lanes 3–5), or AD13-MDTCS plus EDTA (lane 6) according to markers (lane 7). The peptide was cleaved after 10 min by active AD13-MDTCS but not in the other reaction mixtures. (G) SPR sensorgrams of the interaction of AD13-MDTCS-E<sup>225</sup>Q as analyte and VWF-strep-peptide as anchored ligand. Curves from a multi-cycle run at analyte concentrations between 20 nM and 625 nM are depicted, from which the kinetic equilibrium dissociation constant ( $K_D = 1.42 \times 10^{-7}$  M) was determined. (H) Plot of the steady-state response values of (G). The vertical line corresponds to the equilibrium dissociation constant ( $K_D = 1.50 \times 10^{-7}$  M). (I) SEC-MALLS chromatograms of VWF-strep-peptide (red curve) and the AD13-MDTCS-E<sup>225</sup>Q:VWF-strep-peptide complex (blue curve), which reveals an excess of peptide (right peak). The respective experimental molecular masses (in kDa) are indicated. The theoretical mass values of the peptide and the complex are 11.2 Å and ~89 Å (incl. ~10 kDa glycans), respectively.

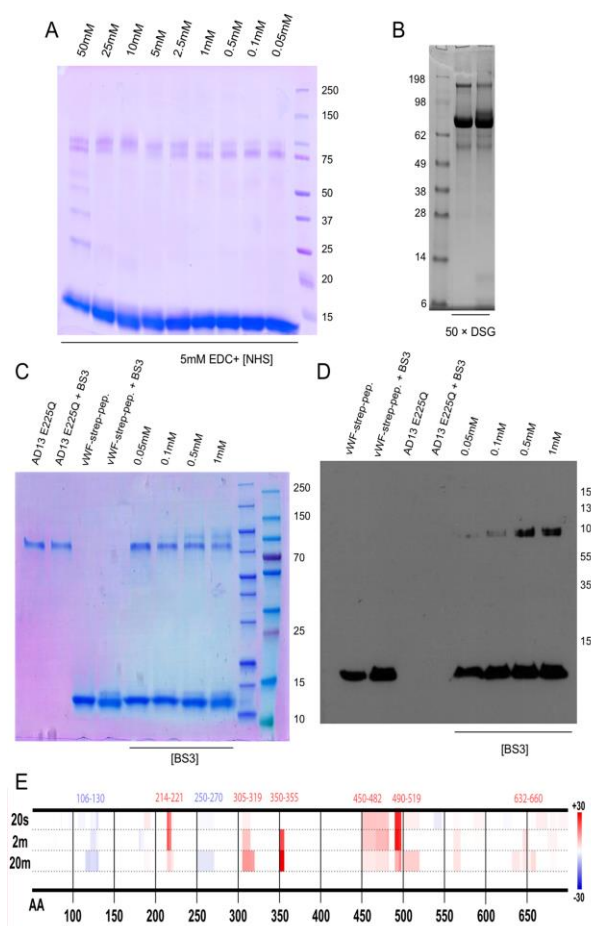


VWF-strep-peptide anchored to Strep-Tactin-resin and AD13-MDTCS-E<sup>225</sup>Q in the liquid phase, followed by SDS-PAGE and western blotting analysis. The proteins coeluted owing to a strong interaction that outcompeted binding to the resin (Fig. 1d). In addition, fluorophore labelling of the peptide and incubation with AD13-MDTCS-E<sup>225</sup>Q revealed the presence of the peptide in a fluorescent band migrating according to the mass of the peptidase in native PAGE (Fig. 1e). This result confirms that the proteins formed a complex that prevails over interactions with the gel matrix. Finally, we analysed cleavage by incubating AD13-MDTCS or AD13-MDTCS-E<sup>225</sup>Q with the VWF-strep-peptide and found that the former but not the latter completely cleaved the substrate at a single site after 10 min, which suggests that our peptidase preparation is more active than a recently published one from insect cells (see Fig. 1f in [23]). The cleavage is consistent with the reported processing at Y<sub>1605</sub>-M<sub>1606</sub> [29] and was ablated with the general metal chelator EDTA, which inactivates MPs (Fig. 1f).

**Determination of the complex affinity and kinetic parameters** — The peptidase:peptide interaction was further investigated by SPR in a multicycle, double-referenced experiment. The VWF-strep-peptide ligand was first immobilised through a Strep-tag on chip-bound Strep-Tactin. Subsequently, binding to the AD13-MDTCS-E<sup>225</sup>Q analyte was monitored at six serially diluted concentrations. The resulting sensorgrams revealed fast association and dissociation of the binding partners, which indicated high-affinity complex formation (Fig. 1g). Moreover, the response curves were fitted with an excellent goodness-of-fit ( $\chi^2 = 0.81\%$  of  $R_{\max}$ ; U-value = 5). The derived rate constants for association ( $k_a$ ) and dissociation ( $k_d$ ) were  $5.15 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  and  $7.28 \times 10^{-2} \text{ s}^{-1}$ , respectively, with an estimated dissociation half-time ( $t_{1/2} = \ln 2 / k_{\text{off}}$ ) of 9.5 s and 7.3% complex decay per s. The equilibrium dissociation constants ( $K_D$ ) from the kinetic and steady-state analysis were  $1.42 \times 10^{-7}$  and  $1.50 \times 10^{-7} \text{ M}$ , respectively (Fig. 1h), which are equivalent and signify remarkable sub-micromolar affinity. Overall, the peptide was strongly bound, which is notable for a disordered molecule of this size.

**Chemical cross-linking analysis** — The interaction was also assessed by CLMS using cross-linking agents EDC/NHS, BS3, DSG, and DSA in separate experiments. The agent DSA in its <sup>12</sup>C (light) and <sup>13</sup>C (heavy) variants was used for quantitative studies of complex formation. The

best conditions for reaction with EDC/NHS and BS3 were 5 mM EDC plus 10 mM NHS and 0.5 mM BS3, respectively (Fig. 2a,c). The latter was verified by western blotting analysis, revealing that even the highest reagent concentration did not cause significant cross-linking of the separate proteins (Fig. 2d), which excludes binding *in trans* under the assay conditions. The best conditions for DSG and DSA involved 50-fold excess of reagent (Fig. 2b).



**Figure 2** — Cross-linking conditions and hydrogen/deuterium exchange mass spectrometry heat map. **(A)** SDS-PAGE depicting the cross-linking of AD13-MDTCS-E<sup>225</sup>Q and VWF-strep-peptide after reaction with 5 mM EDC and different concentrations of NHS. **(B)** Same as (A) but with a 50-fold excess of DSG as cross-linker. *Left lane*, AD13-MDTCS-E<sup>225</sup>Q; *right lane*, AD13-MDTCS-E<sup>225</sup>Q:VWF-strep-peptide. **(C)** Same as (A) but with BS3 as cross-linker. In *lanes 2 and 4*, the concentration of BS3 was 1 mM, which did not cross-link the peptidase and peptide separately. **(D)** Similar to (C) but showing the western blotting analysis using a streptavidin antibody to detect VWF-strep-peptide. Increasing concentrations of BS3 cause more peptide to be bound by the peptidase. **(E)** Differential deuterium uptake of AD13-MDTCS-E<sup>225</sup>Q at 20 s, 2 min, and 20 min after incubation with VWF-strep-peptide. The abscissa depicts the residue numbers of the peptidase, and the ordinate shows the percentage of deuterium exchange (from +30% to -30%). Peptidase segments protected by complex formation with the peptide are shown in red. The figure is representative for three independent experiments.

MS analysis of the resulting tryptic peptides revealed 10 (EDC/NHS), 5 (BS3), 4 (DSG) and 3 (DSA) intra-peptidase connections (Suppl. Tables 1 and 2). Remarkably, K<sup>318</sup>, which is exposed on the D domain, was involved in 12 links in total with residues from the C (Y<sup>468</sup>, E<sup>492</sup>, K<sup>497</sup>, D<sup>500</sup>, and D<sup>516</sup>), M (D<sup>117</sup>, D<sup>182</sup>, and E<sup>184</sup>), and S (K<sup>559</sup>, E<sup>634</sup>, and D<sup>635</sup>) domains. Moreover, other cross-links were found between domains M and D (S<sup>272</sup>–K<sup>368</sup> and S<sup>275</sup>–K<sup>368</sup>), M and S (E<sup>184</sup>–K<sup>559</sup>), C and S (E<sup>492</sup>–K<sup>559</sup> and K<sup>497</sup>–K<sup>608</sup>), and within D (K<sup>364</sup>–K<sup>368</sup> and T<sup>358</sup>–K<sup>368</sup>), C (K<sup>441</sup>–T<sup>435</sup> and T<sup>453</sup>–K<sup>440</sup>), and S (K<sup>559</sup>–T<sup>575</sup>). Overall, these results indicate that peptide-bound active AD13-MDTCS exhibits great flexibility in solution, which brings D close to M, C, and S, as well as M and C close to S, in agreement with previous reports (see [23, 28] and references therein) and the results from H/DXMS (see below). Interestingly, while some of these cross-links are compatible with the unbound latent crystal structure in complex with an Fab fragment (PDB 6QIQ; [23]), taking into account some breathing of the hinges between vicinal domains, others require substantial structural rearrangement. This implies that the active structure in solution is significantly different or is highly flexible. Indeed, the ratios between light and heavy peptides obtained with the DSA cross-linker suggest that the peptidase adopts a more compact conformation in the presence of VWF-strep-peptide (Suppl. Fig. 1).

Regarding inter-molecular cross-links, experiments with EDC/NHS and DSG but not BS3 or DSA further revealed 11 cross-links in total between S<sup>266</sup> from M, K<sup>318</sup> from D, S<sup>388</sup> from T, K<sup>497</sup> from C, and K<sup>559</sup> from S, with VWF residues ranging Y<sub>1605</sub>–D<sub>1663</sub> (Suppl. Tables 1 and 2). Remarkably, D was found cross-linked with peptide residues that were previously shown to interact with S [73, 74]). This sheds new light on published findings that a monoclonal antibody against S exposes an epitope of another monoclonal antibody against M [75]. Finally, DSG and EDC/NHS produced a link between S<sup>561</sup> and a residue from the C-terminal Strep-tag. This supports reported biochemical evidence that S interacts with the C-terminal part of the VWF A2 domain [36, 73, 74, 76].

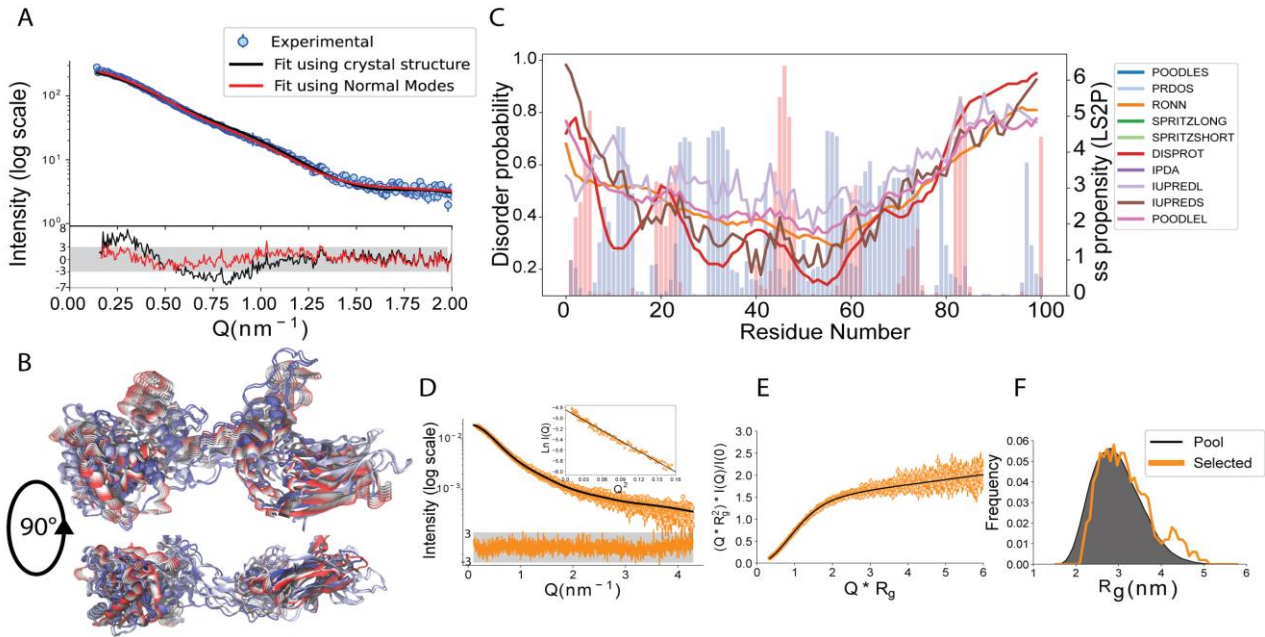
Overall, these results indicate that the peptide creeps along the surface of active AD13-MDTCS by contacting all domains.

**Surface protection studies** — The complex was further studied by H/DXMS (Fig. 2e and Suppl. Fig. 2). The resulting heatmap evinces dynamic protection of the peptidase surface

against deuterium exchange by the peptide. The strongest protection is observed in regions L<sup>350</sup>–L<sup>355</sup> and C<sup>450</sup>–K<sup>497</sup>, which would constitute the main interaction surface of the complex, with secondary interactions involving T<sup>214</sup>–T<sup>221</sup>, Y<sup>305</sup>–A<sup>319</sup> and I<sup>645</sup>–E<sup>651</sup>. These results indicate that the peptidase interacts with the peptide mainly via D and C, and secondarily with M and S, which is overall consistent with the results from CLMS (see above). Moreover, these data are in good agreement with previous reports [36, 77], including the identification of R<sup>349</sup>–L<sup>351</sup> as a peptidase exosite [35] and the shielding of C by the peptide [74].

**Latent AD13-MDTCS in solution is flexible and deviates from the crystal structure** — While reanalysing SAXS data reported previously for unbound latent AD13-MDTCS in solution [28], we found that the experimental SAXS profile did not match the theoretical SAXS profile derived from the latent crystal structure ( $\chi^2=6.0$ ). This implied substantial rearrangement of the peptidase in solution. Thus, we sampled conformations along the first five nontrivial normal modes (Suppl. Fig. 4a), which revealed bending and twisting along the T domain while the relative orientations of the M and D domains and C and S domains remained relatively constant. Consistently, three hinge residues were identified in the middle of the T domain (Suppl. Fig. 4b). We then selected sub-ensembles of conformations sampled along these normal modes that collectively fitted the SAXS data, and found that allowing for flexibility along the hinge points substantially improved the agreement between the coordinates and the experimental SAXS data ( $\chi^2=1.8$ ) (Fig. 3a,b). However, some departures from a perfect fit were observed, most probably due to the simplicity of the normal mode description of protein dynamics. Notably, these structures were mainly sampled by one nontrivial normal mode (mode 7), which largely accounts for the flexibility of the peptidase in solution (Suppl. Fig. 4c). The results demonstrate that the unbound latent peptidase is monomeric and flexible in solution, and behaves as two rigid bodies consisting of M and D, and C and S, respectively, connected by a flexible T domain that acts as a hinge.

**The VWF-strep-peptide is a self-avoiding random chain** — Next, we analysed peptide disorder by employing several bioinformatics approaches, which predicted general disorder except for the central region that has some degree of structure, based on transient secondary



**Figure 3 — Flexibility of AD13-MDTCS and VWF-strep-peptide in solution.** (A) SAXS profile of AD13-MDTCS from [28] along with the theoretical curve computed from the crystal structure (*black*) and the fit obtained using EOM with the conformations sampled along the normal modes (*red*). The residuals are presented at the bottom with the shaded region showing the range of intensities from  $-3$  to  $+3$ . (B) Two orthogonal views of the AD13-MDTCS conformations selected by EOM. (C) Disorder assessment of the peptide using different predictors. The bars show the propensity of each residue to adopt an  $\alpha$ -helical (*red*) or  $\beta$ -strand (*blue*) conformation. (D) SAXS profiles of VWF-strep-peptide at 6 mg/mL (*orange*). The black line shows the fit obtained using EOM, and the residual plot (*bottom*) reveals the difference between the experimental and the optimised curves. The Guinier plot with the linear fit is shown in the inset (*black line*). (E) Normalized Kratky plot of the profile shown in (D) with the fit using the molecular form factor approach (*black line*). (F)  $R_g$  distributions for the initial pool (*black*) and the sub-ensemble selected using EOM (*orange*).

structure elements (Fig. 3c). Moreover, assessment of the secondary structure propensity by residue further indicated that the central region is partially structured. Subsequently, we used SAXS to investigate the conformational space sampled by the peptide in solution (Suppl. Table 3). The SAXS intensity profile and Guinier plots at the highest concentration tested are shown in Fig. 3d, and the pairwise distance distribution ( $P(r)$ ) is shown in Suppl. Fig. 3. The low- $Q$  range indicates minimal aggregation, as inferred from similar radius-of-gyration ( $R_g$ ) values for different concentrations ( $3.08 \pm 0.03$  nm at 3 mg/mL and  $3.04 \pm 0.02$  nm at 6 mg/mL; Suppl. Table 3). The experimental molecular weight was calculated to be  $12.0 \pm 1.2$  kDa, which matches data from SEC-MALLS (11.5 kDa; Fig. 1i) and the theoretical molecular weight (11.2 kDa). This indicates that the protein is monomeric in solution, even at high concentration. Finally, the lack of a peak in the normalized Kratky plot further implies that the peptide is disordered (Fig. 3e).

Interestingly,  $R_g$  was higher than expected for an intrinsically disordered protein (2.86 nm) based on Flory's equation [78]. Together with the disorder prediction, this confirms some level of regular structure that results in a more elongated

conformation than that expectable for a random chain. Indeed, the Flory exponent of 0.6 calculated using the molecular form factor approach is consistent with a self-avoiding random chain rather than an intrinsically disordered protein [79]. We subsequently generated peptide conformations to select sub-ensembles of 20 conformations that collectively best described the SAXS data for the peptide using the EOM method. The results from one of the cycles are shown in Suppl. Fig. 4d, and the corresponding fit to the experimental data is shown in Fig. 3d. The peptide has a broad distribution of  $R_g$  values with an average of 3.2 nm (Fig. 3f). Again, the distribution is slightly shifted toward an extended conformation rather than a pure random coil, which further indicates partial regular structure.

#### Solution structure of the AD13-MDTCS-E225Q:VWF-strep-peptide complex

Analysis of the complex by SEC-MALLS revealed a single peak of 96.8 kDa (Fig. 1i), which is consistent with the theoretical value of  $\sim 89$  kDa. Next, we collected SAXS data for the peptidase:peptide complex in SEC-SAXS mode (Fig. 4a and Suppl. Table 3). The extracted profile of the complex corresponded to a globular

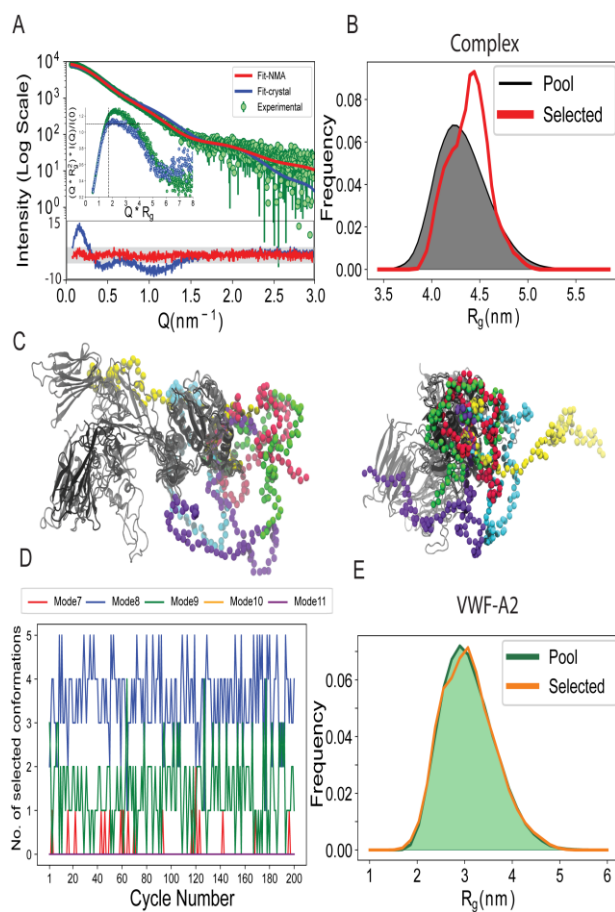


particle ( $R_g = 4.38 \pm 0.01$  nm), which is larger than the unbound peptidase ( $R_g = 3.80 \pm 0.05$  nm). Consistently, the peak of the normalized Kratky plot was shifted from 2.0 to 2.3 upon complex formation (Fig. 4a), which implies an increase in flexibility caused by the peptide. This finding was substantiated by inspection of the  $P(r)$  function of the complex, which displays fewer signatures than the peptidase alone (Suppl. Fig. 3).

We subsequently performed ensemble modelling. To provide an accurate starting molecular model for the active peptidase, the M domain from the latent crystal structure was remodelled to match active ADAMTS structures while the remaining domains were kept intact. As depicted in Fig. 4a, the ensemble generated with the rigid peptidase model based on the crystal structure did not fit the experimental SAXS profile ( $\chi^2 = 10.1$ ). This was expected given that the unbound latent peptidase in solution already diverged from the latent crystal structure (see above). We then allowed for flexibility by building distinct peptide conformations for each of the conformational ensembles generated for the isolated peptidase using normal modes (see above), resulting in 120,000 structures of the complex. Subjecting this ensemble to genetic algorithm-based optimization gave an excellent fit to the data ( $\chi^2=1.2$ ; Fig. 4a). The conformations selected in one of the cycles are shown in Fig. 4c. The  $R_g$  distribution of the selected pool was quite broad, which supports high flexibility of the complex and a slight preference for more extended structures compared with the total pool (Fig. 4b).

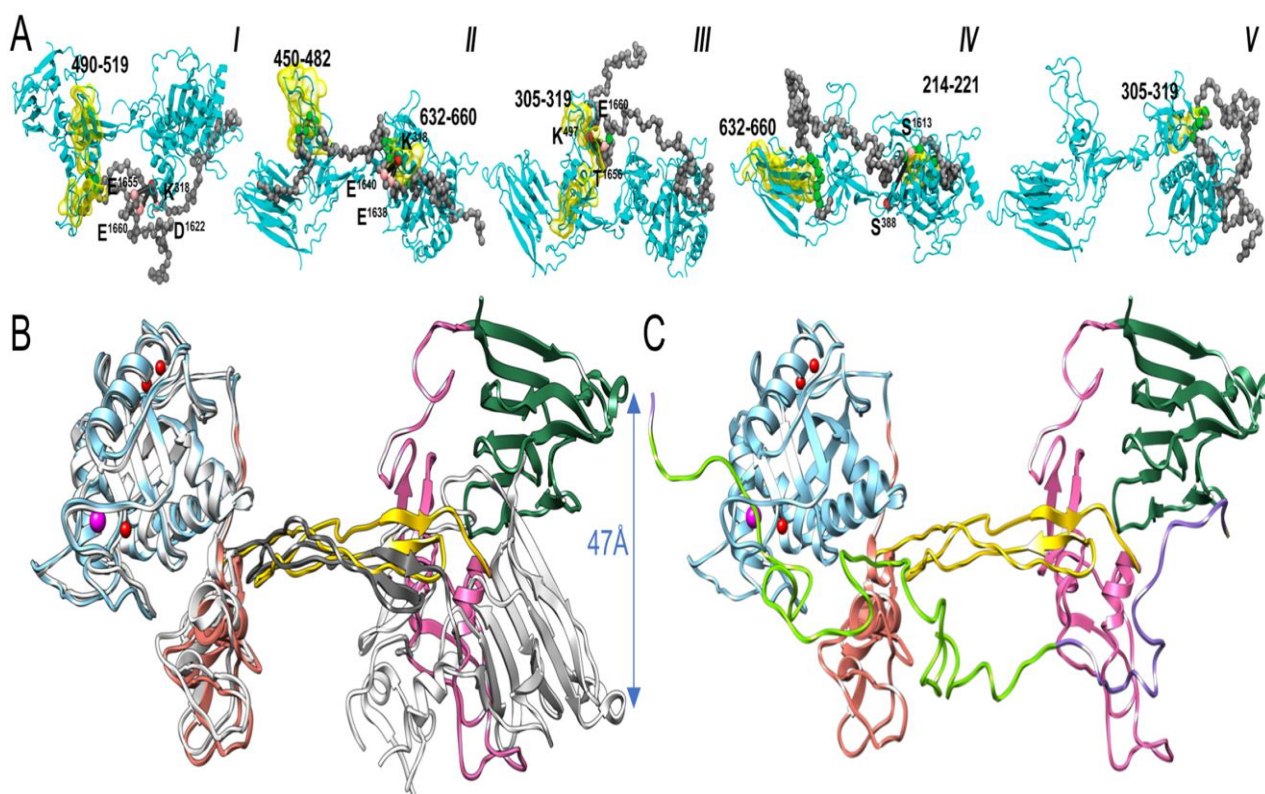
We also examined the flexibility of the peptidase in the active conformation within the complex by analysing the motional modes sampled by the individual conformations selected with the SAXS data. Interestingly, the ensembles were almost exclusively constituted from modes 8 and 9, with sporadic contributions of mode 7, suggesting that these modes alone describe the overall interdomain motions (Fig. 4d). These results differ from those obtained for the unliganded latent peptidase, where mode 7 made the largest contribution to the selected ensembles, and suggests that peptide binding changes the inherent molecular motions of the peptidase. On the other hand, the  $R_g$  distribution of the selected peptide conformations within the complex was almost identical to that of the total pool, which indicated no preferential selection of the peptide in terms of size or extension (Fig. 4e). Thus, the peptide displayed a similar degree of disorder to the unbound state upon complex formation. This

observation also indicates that the putative occurrence of multiple simultaneous binding events suggested by the ‘molecular zipper’ model [23, 34] must be transient in nature, and SAXS is not sensitive enough to detect them.



**Figure 4 — Solution structure of the AD13-MDTCS-E225Q:VWF-strep-peptide complex. (A)** SAXS intensity profile of the complex obtained from SEC-SAXS data (green) superposed with the best fit obtained by EOM using a static structure for the peptidase and 10,000 conformations for the peptide (blue curve). Further accounting for peptidase flexibility following normal mode analysis provides a much better fit to the data (red curve). The normalized Kratky plots for the peptidase and the complex are shown in the inset in which dashed lines show the expected position (1.73) and height (1.10) of the theoretical peak for a globular protein. **(B)**  $R_g$  distribution for the total pool (black curve) and the selected sub-ensemble (red curve). **(C)** Two orientations (left and right) of the EOM-selected conformations that collectively describe the experimental SAXS data. **(D)** Contribution of the individual modes to the selected sub-ensemble for 200 cycles of EOM. **(E)** Comparison of the  $R_g$  distribution of peptide conformations bound to the peptidase in the complete pool (green curve) of the complex and the sub-ensemble selected by EOM (orange curve).

**Validation of the complex structure —** We next analysed the compatibility of the intermolecular distance restraints resulting from CLMS (Suppl. Table 1) with the SAXS-derived structure ensembles. We filtered the total pool of



**Figure 5** — Integrative atomic model of the AD13-MDTCS-E<sup>225</sup>Q:VWF-strep-peptide complex. **(A)** Example of a sub-ensemble of five complex structures (I–V) selected by artificial intelligence that fulfils all the 11 inter-molecular cross-linking restraints (Suppl. Table 1). AD13-MDTCS is shown as cyan ribbons and VWF-strep-peptide as a grey string of beads, each representing a C $\alpha$  atom. Black cylinders connect the cross-linked C $\alpha$  atoms. The surfaces that are protected in H/DXMS experiments are coloured in yellow, and peptide residues that confer protection in H/DXMS experiments are shown as green spheres. These residues lie within  $\sim 5$  Å of the protected surface. In this particular ensemble, the CLMS data can be explained with three structures, while fulfilment of H/DXMS data requires two additional structures. **(B)** Ribbon-type plot of the peptidase moiety of model II of (A), which fulfils 7/11 of the experimental inter-molecular CLMS restraints (see also Suppl. Table 1), with the M, D, T, C, and S domains coloured in sky blue, salmon, gold, hot pink, and sea green, respectively. The zinc ion is coloured magenta, and calcium ions are red. The M domain of the model is superposed onto the crystal structure of latent AD13-MDTCS-E<sup>225</sup>Q (M, D, C, and S in white, T in grey). The largest distance is found between the respective Y<sup>605</sup> residues (*blue arrow*). **(C)** Ribbon plot of the peptidase:peptide complex in the orientation shown in (B). The peptide is represented as a green coil for segment D<sub>1598</sub>–R<sub>1668</sub>, and residues upstream and downstream of the construct are coloured purple.

structures and retained only those in which the peptidase is in a conformation selected by at least one of the EOM cycles. Such filtering was not required for the peptide due to its much higher flexibility (see above). The positions of the C $\alpha$  atoms of the cross-linked amino acids of the peptide and the peptidase, for which cross-linking was observed, are shown individually in Suppl. Fig. 5. The high flexibility of the peptide in the complex is evidenced by the large space explored by these individual residues, which systematically increases with the distance between the cross-linked amino acids and the docked peptide. In all cases, we found overlap between the spaces explored by the pairs of cross-linked residues. Therefore, the CLMS data are compatible with the structural model of the complex derived from SAXS data and computational methods.

**A working model for the functional AD13-MDTCS:VWF-strep-peptide complex in solution** — To generate molecular models for the competent complex, we selected peptidase conformations identified at least once by EOM from the SAXS data, and calculated peptides in random-coil conformation. Using a genetic algorithm-based selection, we acquired sub-ensembles of five conformations that agreed with all the experimental CLMS restraints, one of which is presented in Fig. 5a. Running the genetic algorithm multiple times yielded different, equally valid sub-ensembles, which indicates that there are multiple ways in which the proteins may interact. It is also important to note that these sub-ensembles alone do not fit the SAXS data but represent minimal clusters that collectively fulfil all the distance restraints.

We subsequently constructed a full-atom model for one of the conformers of this sub-ensemble (Fig. 5c). Superposition of the M domain onto that of the latent crystal structure (Fig. 5b) reveals that D is very close in both structures, with a maximal deviation of  $\sim 3$  Å. A rotation of  $\sim 10^\circ$  around P<sup>379</sup> within the linker between D and T causes a maximal displacement of  $\sim 10$  Å for S<sup>399</sup> in the distal part of the latter domain. Furthermore, the TC linker contains a hinge point, and a  $\sim 17^\circ$  rotation about A<sup>437</sup> leads to a maximal deviation of  $\sim 6$  Å at H<sup>476</sup>. Finally, only a small rearrangement is observed between C and S. Overall, these changes lead to a maximal displacement of  $\sim 47$  Å at the most distal part of S (Fig. 5b).

**Conclusions** — The molecular details of the complex formed between active AD13-MDTCS and its cognate region of the VWF A2 domain encompassing the scissile bond have proved difficult to elucidate using classical structural methods for years. Previous studies identified several binding sites on the enzyme surface that are bound by the unfolded A2 domain of VWF. Binding to these exosites, which transcend the M domain, positions the peptide within the active site cleft in the correct conformation and orientation for cleavage in what has been dubbed a ‘molecular zipper’ model [23, 34]. Herein, we set out to comprehensively analyse the peptidase:peptide interaction through an integrative structural biology approach, applying nine complementary biophysical, biochemical and biocomputational techniques.

SAXS and SEC-MALLS analyses demonstrated that the isolated proteins are highly flexible in solution. This plasticity is maintained in the complex, which therefore cannot be recapitulated by a single, rigid structure but rather conforms to a ‘fuzzy complex.’ Such complexes often link regulation to protein dynamics [80, 81]. This further implies that binding of the peptide to the peptidase, which involves specific residue pairs and protected regions identified by CLMS and H/DXMS, respectively, is transient in nature. Thus, the high binding affinity and very efficient cleavage observed in gel-electrophoresis, pull-down, SPR, and peptidolysis assays are the result of multiple low-affinity interactions and small individual energy contributions that simultaneously populate the interaction sites in a variety of ways. Counterintuitively, the complex remains highly disordered while driving the peptide in the optimal orientation for catalysis. This property is characteristic of intrinsically

disordered proteins, which comprise several consecutive polypeptide stretches that recognize multiple partners or multiple sites of the same partner [82]. This feature enables cooperative binding while preserving plasticity, which reduces the entropic cost of the interaction. Overall, the ADAMTS-13: VWF-peptide interaction revealed herein appears to conform to a ‘dynamic zipper’ mechanism, consistent with previous studies showing that modifying individual exosites of the peptidase only moderately reduces the affinity for, and cleavage of, the peptide [23, 34].

## ACKNOWLEDGMENTS

We are grateful to Roman Bonet, Laura Company, Xandra Kreplin and Joan Pous from the joint IBMB/IRB Automated Crystallography Platform and the Protein Purification Service for assistance during purification and SEC-MALLS experiments. The assistance of Aleix Tarrés-Solé and the local contacts of beamline P12 of the DESY synchrotron in Hamburg during SAXS data collection is also acknowledged. The late J. Evan Sadler and Jian Zhu, both from the Washington University School of Medicine, are thanked for providing plasmid pcDNA4/TO and original SAXS data for wild-type AD13-MDTCS, respectively. This study was supported in part by grants from Spanish, French, Danish and Catalan public and private bodies (grant/fellowship references PID2019-107725RG-I00, BES-2015-074583, ANR-10-LABX-12-01, 6108-00031B, 8022-00385B, LF18039, NNF18OC0032724, Novo Nordisk Foundation “Bio-MS”, 2017SGR3 and Fundació “La Marató de TV3” 201815). This work was also supported by EPICS-XS, project 823839, funded by the Horizon 2020 programme of the European Union. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), which are national infrastructures supported by the French National Research Agency (grants ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively). Finally, we acknowledge the Structural Mass Spectrometry unit of CIISB, an Instruct-CZ Centre, which was supported by MEYS CR (LM2018127).

## AUTHOR CONTRIBUTIONS

F.X.G.R. conceived and supervised the project; M.C.-B. and T.G. performed initial expression trials in eukaryotic systems. L.A.M. produced and purified all proteins, and performed biochemical and biophysical studies; A.S., P.P., T.G., and C.S. performed biophysical and

computational studies; M.T. supervised SPR studies; J.J.E. supervised cross-linking and MS studies; P.B. supervised SAXS studies and biocomputational calculations; and F.X.G.R. and P.B. wrote the manuscript with contributions from all authors.

## CONFLICT OF INTERESTS

The authors declare no financial or non-financial conflicts of interest with the contents of this article.

## REFERENCES

[1] Branchford BR, Monahan PE, Di Paola J. New developments in the treatment of pediatric hemophilia and bleeding disorders. *Curr Opin Pediatr.* 2013;25:23-30.

[2] de Ceunynck K, de Meyer SF, Vanhoorelbeke K. Unwinding the von Willebrand factor strings puzzle. *Blood.* 2013;121:270-7.

[3] Denorme F, Vanhoorelbeke K, de Meyer SF. von Willebrand factor and platelet glycoprotein Ib: A thromboinflammatory axis in stroke. *Front Immunol.* 2019;10:2884.

[4] Sporn LA, Marder VJ, Wagner DD. Inducible secretion of large, biologically potent von Willebrand factor multimers. *Cell.* 1986;46:185-90.

[5] Dong JF, Moake JL, Nolasco L, Bernardo A, Arceneaux W, Shrimpton CN, et al. ADAMTS-13 rapidly cleaves newly secreted ultralarge von Willebrand factor multimers on the endothelial surface under flowing conditions. *Blood.* 2002;100:4033-9.

[6] Feys HB, Anderson PJ, Vanhoorelbeke K, Majerus EM, Sadler JE. Multi-step binding of ADAMTS-13 to von Willebrand factor. *J Thromb Haemost.* 2009;7:2088-95.

[7] Chen J, Fu X, Wang Y, Ling M, McMullen B, Kulman J, et al. Oxidative modification of von Willebrand factor by neutrophil oxidants inhibits its cleavage by ADAMTS13. *Blood.* 2010;115:706-12.

[8] Ercig B, Wichapong K, Reutelingsperger CPM, Vanhoorelbeke K, Voorberg J, Nicolaes GAF. Insights into 3D structure of ADAMTS13: A stepping stone towards novel therapeutic treatment of thrombotic thrombocytopenic purpura. *Thromb Haemost.* 2018;118:28-41.

[9] Jezovnik MK, Poredos P. Idiopathic venous thrombosis is related to systemic inflammatory response and to increased levels of

circulating markers of endothelial dysfunction. *Int Angiol.* 2010;29:226-31.

[10] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380:2095-128.

[11] Crawley JT, Scully MA. Thrombotic thrombocytopenic purpura: basic pathophysiology and therapeutic strategies. *Hematology Am Soc Hematol Educ Program.* 2013;2013:292-9.

[12] Maino A, Siegerink B, Lotta LA, Crawley JT, le Cessie S, Leebeek FW, et al. Plasma ADAMTS-13 levels and the risk of myocardial infarction: an individual patient data meta-analysis. *J Thromb Haemost.* 2015;13:1396-404.

[13] Bansilal S, Castellano JM, Fuster V. Global burden of CVD: focus on secondary prevention of cardiovascular disease. *Int J Cardiol.* 2015;201 Suppl 1:S1-S7.

[14] Sadler JE. Pathophysiology of thrombotic thrombocytopenic purpura. *Blood.* 2017;130:1181-8.

[15] Sadler JE. New concepts in von Willebrand disease. *Annu Rev Med.* 2005;56:173-91.

[16] Branchford BR, Di Paola J. Making a diagnosis of VWD. *Hematology.* 2012;2012:161-7.

[17] Chen J, Hinckley JD, Haberichter S, Jacobi P, Montgomery R, Flood VH, et al. Variable content of von Willebrand factor mutant monomer drives the phenotypic variability in a family with von Willebrand disease. *Blood.* 2015;126:262-9.

[18] Ng CJ, di Paola J. von Willebrand Disease: diagnostic strategies and treatment options. *Pediatr Clin North Am.* 2018;65:527-41.

[19] Laffan M, Sathar J, Johnsen JM. von Willebrand disease: Diagnosis and treatment, treatment of women, and genomic approach to diagnosis. *Haemophilia : the official journal of the World Federation of Hemophilia.* 2020;26:in press.

[20] Majerus EM, Zheng X, Tuley EA, Sadler JE. Cleavage of the ADAMTS13 propeptide is not required for protease activity. *J Biol Chem.* 2003;278:46643-8.

[21] Arolas JL, Goulas T, Cuppari A, Gomis-Rüth FX. Multiple architectures and mechanisms of latency in metallopeptidase zymogens. *Chem Rev.* 2018;118:5581-97.

[22] Muia J, Zhu J, Greco SC, Vanhoorelbeke K, Gupta G, Westfield LA, et al. Phylogenetic and



functional analysis of ADAMTS13 identifies highly conserved domains essential for allosteric regulation. *Blood*. 2019;133:1899-908.

[23] Petri A, Kim HJ, Xu Y, de Groot R, Li C, Vandenbulcke A, et al. Crystal structure and substrate-induced activation of ADAMTS13. *Nat Commun*. 2019;10:3781.

[24] Taylor A, Vendramin C, Oosterholt S, Della Pasqua O, Scully M. Pharmacokinetics of plasma infusion in congenital thrombotic thrombocytopenic purpura. *J Thromb Haemost*. 2019;17:88-98.

[25] Cerdà-Costa N, Gomis-Rüth FX. Architecture and function of metallopeptidase catalytic domains. *Prot Sci*. 2014;23:123-44.

[26] De Ceunynck K, Rocha S, De Meyer SF, Sadler JE, Uji IH, Deckmyn H, et al. Single particle tracking of ADAMTS13 (a disintegrin and metalloprotease with thrombospondin type-1 repeats) molecules on endothelial von Willebrand factor strings. *J Biol Chem*. 2014.

[27] Takeda S, Takeya H, Iwanaga S. Snake venom metalloproteinases: Structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. *Biochim Biophys Acta*. 2012;1824:164-76.

[28] Zhu J, Muia J, Gupta G, Westfield LA, Vanhoorelbeke K, Tolia NH, et al. Exploring the "minimal" structure of a functional ADAMTS13 by mutagenesis and small-angle X-ray scattering. *Blood*. 2019;133:1909-18.

[29] Dent JA, Galbusera M, Ruggeri ZM. Heterogeneity of plasma von Willebrand factor multimers resulting from proteolysis of the constituent subunit. *J Clin Invest*. 1991;88:774-82.

[30] Zhang Q, Zhou YF, Zhang CZ, Zhang X, Lu C, Springer TA. Structural specializations of A2, a force-sensing domain in the ultralarge vascular protein von Willebrand factor. *Proc Natl Acad Sci USA*. 2009;106:9226-31.

[31] López JA, Dong JF. Shear stress and the role of high molecular weight von Willebrand factor multimers in thrombus formation. *Blood Coagul Fibrinolysis*. 2005;16 Suppl. 1:S11-S6.

[32] Bortot M, Ashworth K, Sharifi A, Walker F, Crawford NC, Neeves KB, et al. Turbulent flow promotes cleavage of VWF (von Willebrand factor) by ADAMTS13 (a disintegrin and metalloproteinase with a thrombospondin type-1 motif, member 13). *Arterioscler Thromb Vasc Biol*. 2019;39:1831-42.

[33] Kokame K, Nobe Y, Kokubo Y, Okayama A, Miyata T. FRET-S-VWF73, a first fluorogenic substrate for ADAMTS13 assay. *Br J Haematol*. 2005;129:93-100.

[34] Crawley JT, de Groot R, Xiang Y, Luken BM, Lane DA. Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor. *Blood*. 2011;118:3212-21.

[35] de Groot R, Bardhan A, Ramroop N, Lane DA, Crawley JTB. Essential role of the disintegrin-like domain in ADAMTS13 function. *Blood*. 2009;113:5609-16.

[36] Akiyama M, Takeda S, Kokame K, Takagi J, Miyata T. Crystal structures of the noncatalytic domains of ADAMTS13 reveal multiple discontinuous exosites for von Willebrand factor. *Proc Natl Acad Sci USA*. 2009;106:19274-9.

[37] Muia J, Zhu J, Gupta G, Haberichter SL, Friedman KD, Feys HB, et al. Allosteric activation of ADAMTS13 by von Willebrand factor. *Proc Natl Acad Sci USA*. 2014;111:18584-9.

[38] Anderson PJ, Kokame K, Sadler JE. Zinc and calcium ions cooperatively modulate ADAMTS13 activity. *J Biol Chem*. 2006;281:850-7.

[39] Goulas T, Cuppari A, García-Castellanos R, Snipas S, Glockshuber R, Arolas JL, et al. The pCri System: a vector collection for recombinant protein expression and purification. *PloS one*. 2014;9:e112643.

[40] Marino-Puertas L, Del Amo-Maestro L, Taules M, Gomis-Rüth FX, Goulas T. Recombinant production of human  $\alpha_2$ -macroglobulin variants and interaction studies with recombinant G-related  $\alpha_2$ -macroglobulin binding protein and latent transforming growth factor- $\beta_2$ . *Sci Rep*. 2019;9:9186.

[41] Mendes SR, Amo-Maestro LD, Marino-Puertas L, Diego I, Goulas T, Gomis-Rüth FX. Analysis of the inhibiting activity of reversion-inducing cysteine-rich protein with Kazal motifs (RECK) on matrix metalloproteinases. *Sci Rep*. 2020;10:6317.

[42] Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. 2006;1:2856-60.

[43] Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*. 2007;2:1896-906.

[44] Götze M, Pettelkau J, Fritzsche R, Ihling CH, Schäfer M, Sinz A. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *J Am Soc Mass Spectrom*. 2015;26:83-97.

[45] Götze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, et al. StavroX--a software



for analyzing crosslinked products in protein interaction studies. *J Am Soc Mass Spectrom.* 2012;23:76-87.

[46] Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, et al. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci USA.* 2000;97:5802-6.

[47] Hirose S, Shimizu K, Kanai S, Kuroda Y, Noguchi T. POODLE-L: a two-level SVM prediction system for reliably predicting long disordered regions. *Bioinformatics.* 2007;23:2046-53.

[48] Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucl Acids Res.* 2007;35:W460-W4.

[49] Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics.* 2005;21:3369-76.

[50] Vullo A, Bortolami O, Pollastri G, Tosatto SC. Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. *Nucl Acids Res.* 2006;34:W164-W8.

[51] Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005;21:3433-4.

[52] Su CT, Chen CY, Hsu CM. iPDA: integrated protein disorder analyzer. *Nucl Acids Res.* 2007;35:W465-W72.

[53] Kozłowski L, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC bioinformatics.* 2012;13:111.

[54] Estaña A, Barozet A, Mouhand A, Vaisset M, Zanon C, Fauret P, et al. Predicting secondary structure propensities in IDPs using simple statistics from three-residue fragments. *J Mol Biol.* 2020;432:in press.

[55] Franke D, Kikhney AG, Svergun DI. Automated acquisition and analysis of small angle X-ray scattering data. *Nucl Instrum Methods Phys Res A.* 2012;689:52-9.

[56] Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics.* 2018;34:1944-6.

[57] Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, et al. ATSAS 2.8: a comprehensive data analysis suite

for small-angle scattering from macromolecular solutions. *J Appl Cryst.* 2017;50:1212-25.

[58] Fischer H, de Oliveira Neto M, Napolitano HB, Polikarkpov I, Craievich AF. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J Appl Cryst.* 2010;43:101-9.

[59] Franke D, Jeffries CM, Svergun DI. Machine learning methods for X-Ray scattering data analysis from biomacromolecular solutions. *Biophys J.* 2018;114:2485-92.

[60] Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 2013;496:477-81.

[61] Hajizadeh NR, Franke D, Jeffries CM, Svergun DI. Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci Rep.* 2018;8:7204.

[62] Emekli U, Schneidman-Duhovny D, Wolfson HJ, Nussinov R, Haliloglu T. HingeProt: automated prediction of hinges in protein structures. *Proteins.* 2008;70:1219-27.

[63] Suhre K, Sanejouand YH. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucl Acids Res.* 2004;32:W610-W4.

[64] Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc.* 2007;129:5656-64.

[65] Tria G, Mertens HD, Kachala M, Svergun DI. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ.* 2015;2:207-17.

[66] Casañal A, Lohkamp B, Emsley P. Current developments in *Coot* for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* 2020;29:1069-78.

[67] Svergun DI, Barberato C, Koch MHJ. CRYSOLE - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst.* 1995;28:768-73.

[68] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem.* 2011;32:2319-27.

[69] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14:33-8.

[70] Fortin F-A, de Rainville F-M, Gardner M-A, Parizeau M, Gangé C. DEAP: Evolutionary

algorithms made easy. *J Mach Learn Res.* 2012;13:2171-5.

[71] Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr sect D.* 2019;75:861-77.

[72] Prisant MG, Williams CJ, Chen VB, Richardson JS, Richardson DC. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. *Prot Sci.* 2020;29:315-29.

[73] Gao W, Anderson PJ, Majerus EM, Tuley EA, Sadler JE. Exosite interactions contribute to tension-induced cleavage of von Willebrand factor by the antithrombotic ADAMTS13 metalloprotease. *Proc Natl Acad Sci USA.* 2006;103:19099-104.

[74] Gao W, Anderson PJ, Sadler JE. Extensive contacts between ADAMTS13 exosites and von Willebrand factor domain A2 contribute to substrate specificity. *Blood.* 2008;112:1713-9.

[75] Schelpe AS, Petri A, Roose E, Pareyn I, Deckmyn H, de Meyer SF, et al. Antibodies that conformationally activate ADAMTS13 allosterically enhance metalloprotease domain function. *Blood Adv.* 2020;4:1072-80.

[76] Pos W, Crawley JTB, Fijnheer R, Voorberg J, Lane DA, Luken BM. An

autoantibody epitope comprising residues R660, Y661, and Y665 in the ADAMTS13 spacer domain identifies a binding site for the A2 domain of VWF. *Blood.* 2010;115:1640-9.

[77] de Groot R, Lane DA, Crawley JTB. The role of the ADAMTS13 cysteine-rich domain in VWF binding and proteolysis. *Blood.* 2015;125:1968-75.

[78] Bernado P, Blackledge M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J.* 2009;97:2839-45.

[79] Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science.* 2017;358:238-41.

[80] Fuxreiter M. Fuzziness: linking regulation to protein dynamics. *Mol Biosyst.* 2012;8:168-77.

[81] Sharma R, Raduly Z, Miskei M, Fuxreiter M. Fuzzy complexes: specific binding without complete folding. *FEBS Lett.* 2015;589:2533-42.

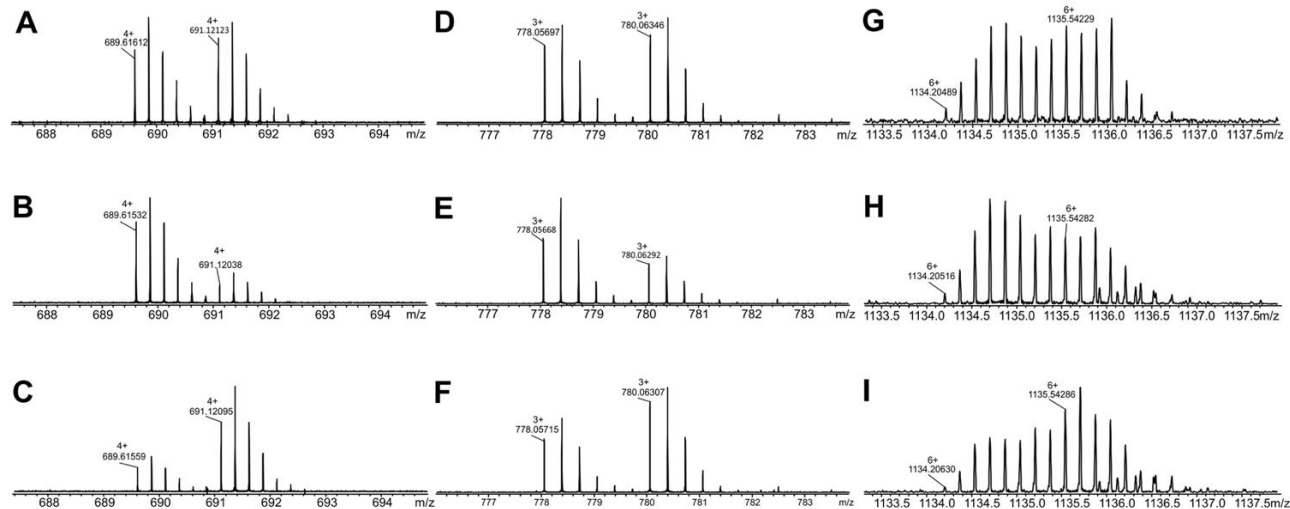
[82] Cordeiro TN, Sibille N, Germain P, Barthe P, Boulahtouf A, Allemand F, et al. Interplay of protein disorder in retinoic acid receptor heterodimer and its corepressor regulates gene expression. *Structure.* 2019;27:1270-85.



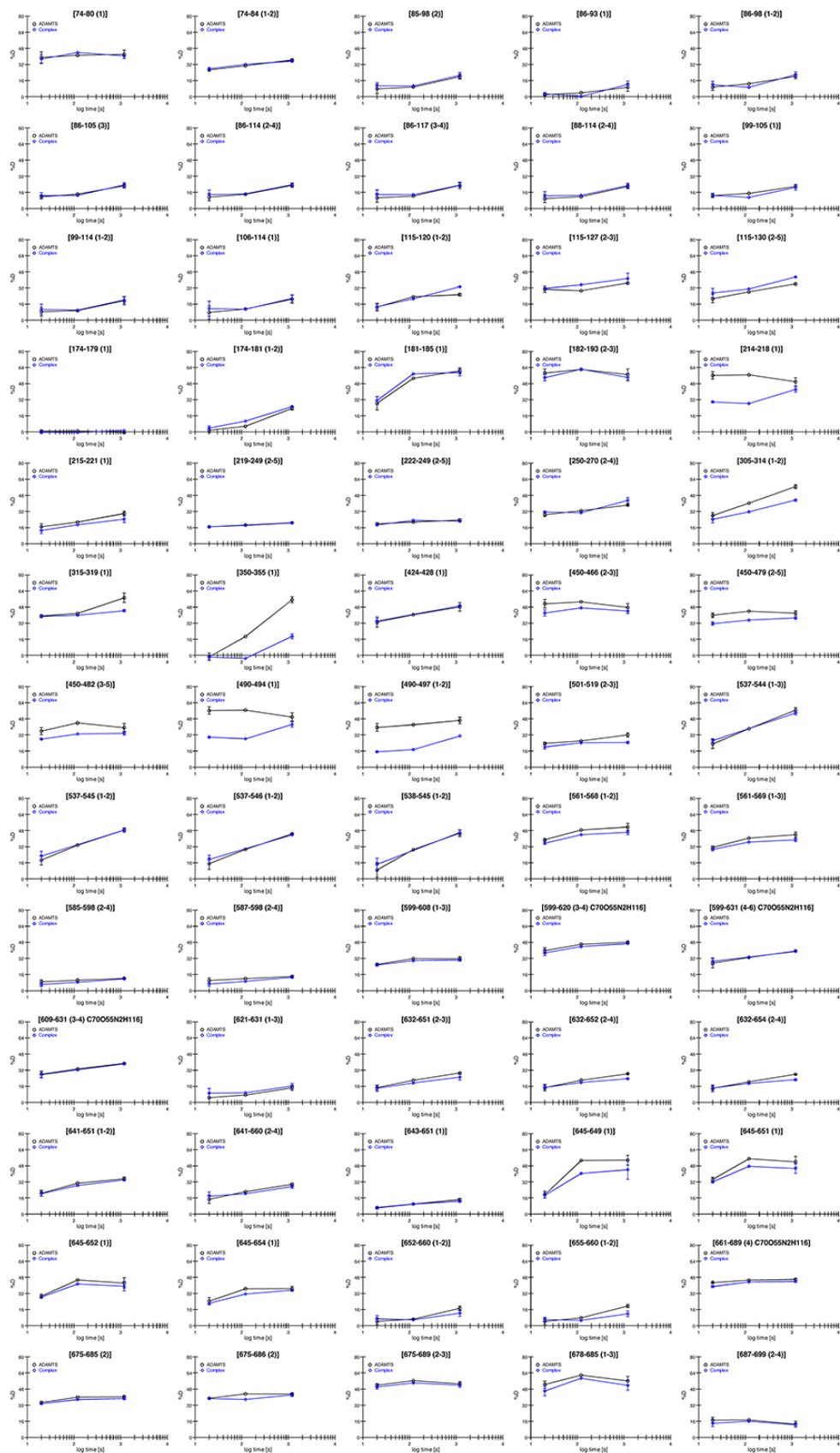
# SUPPORTING INFORMATION

## Table of Contents

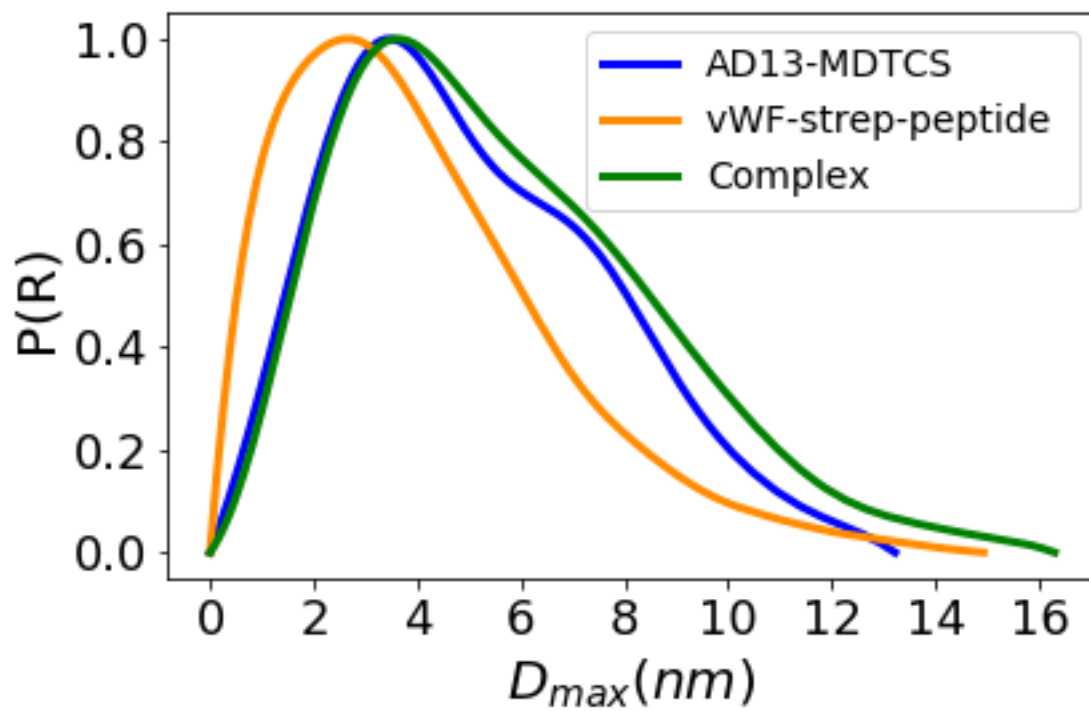
<b>1. Supplementary figures</b>	<b>1</b>
Supplementary Figure 1	1
Supplementary Figure 2	2
Supplementary Figure 3	3
Supplementary Figure 4	4
Supplementary Figure 5	5
<b>2. Supplementary tables</b>	<b>6</b>
Supplementary Table 1	6
Supplementary Table 2	7
Supplementary Table 3	8



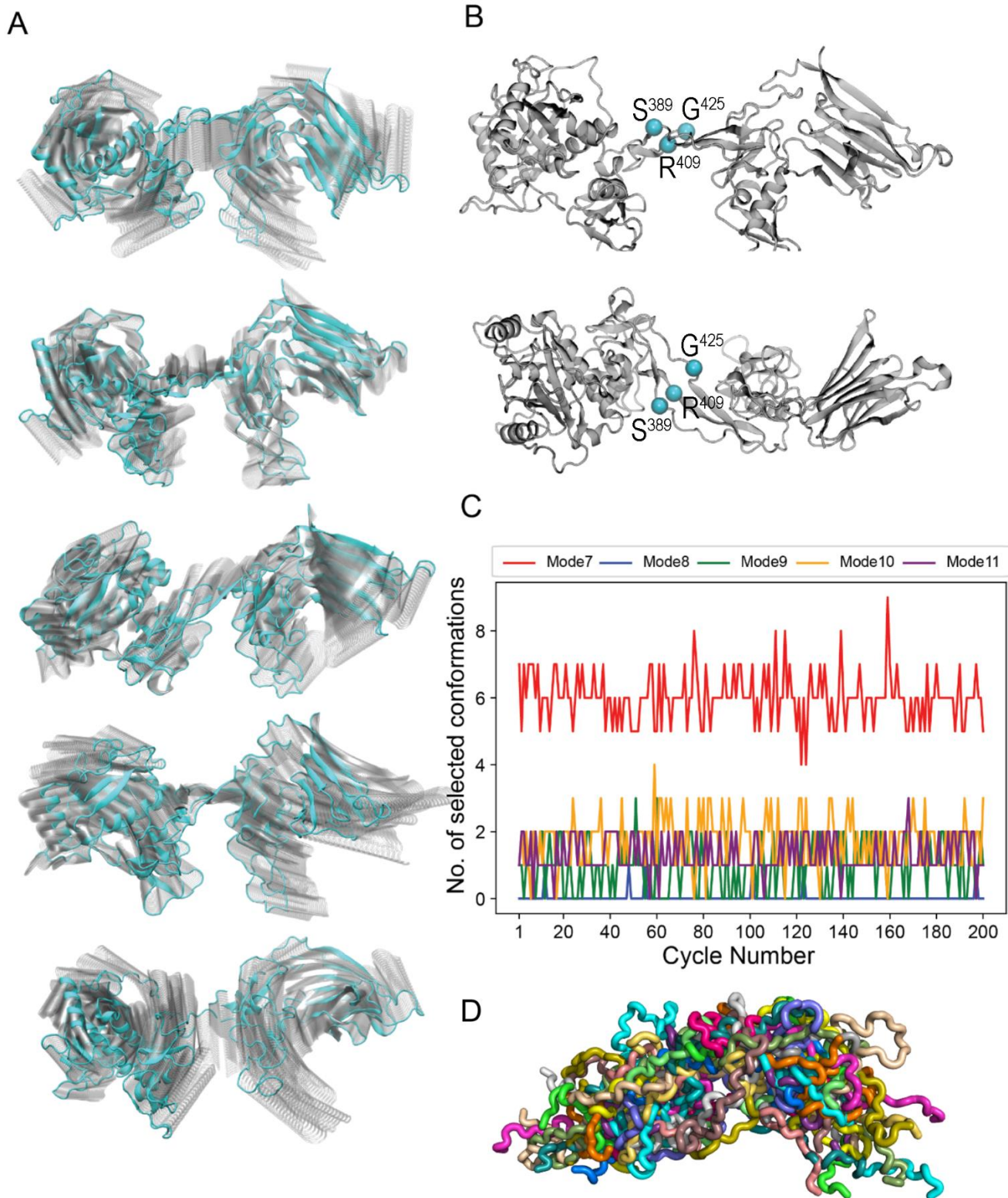
**Suppl. Fig. 1 — Original spectra from quantitative CLMS analysis of DSA-linked peptides. (A–C)** AD13-MDTCS-E<sup>225</sup>Q peptides VAFGPKAVACTFAR and AIGESFIMKR from intra-molecular cross-linking (see Suppl. Table 2) in the absence (A) or presence (B,C) of vWF-strep-peptide. In (B), the complex and AD13-MDTCS-E<sup>225</sup>Q were modified with light (<sup>12</sup>C) and heavy (<sup>13</sup>C) DSA, respectively. In (C), the complex and AD13-MDTCS-E<sup>225</sup>Q were modified with heavy (<sup>13</sup>C) and light (<sup>12</sup>C) DSA, respectively. **(D–F)** Same as (A–C) for AD13-MDTCS-E<sup>225</sup>Q peptides LLVPLLDGTECGVEK and WCSK. **(G–I)** Same as (A–C) for AD13-MDTCS-E<sup>225</sup>Q peptides SSPGGASFYHWGAAVPHSQGDALCRHMCRAIGESFIMK and VVAGKMSISPD TTYPSSLLEDGR.



**Suppl. Fig. 2** — Original data from H/DXMS. Deuterium uptake plots for AD13-MDTCs-E<sup>225</sup>Q in the presence of vWF-strep-peptide at time points 2 s, 2 min, and 20 min.

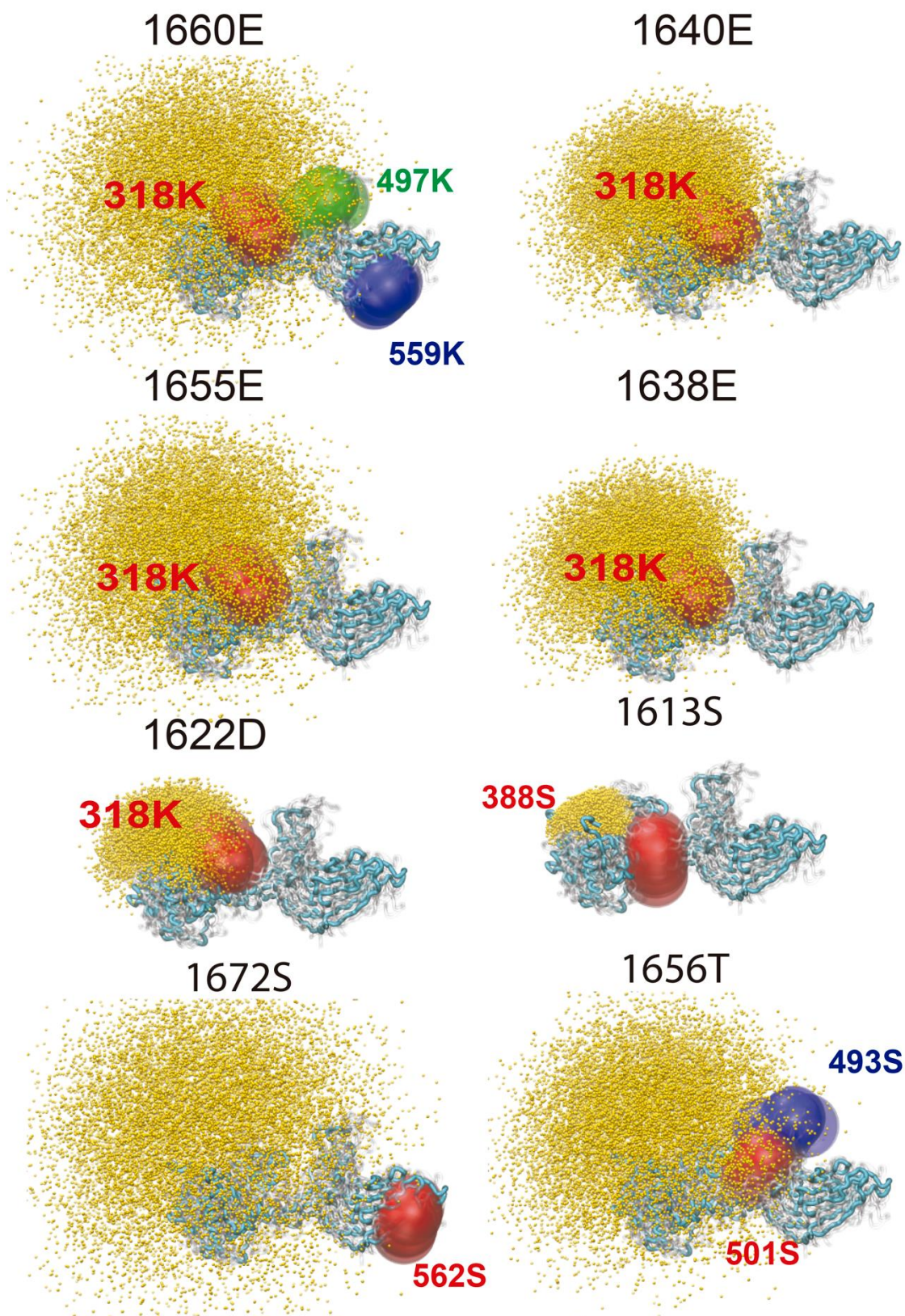


**Suppl. Fig. 3 —  $P(r)$  functions.** Plot depicting the values of  $P(r)$  for the peptidase, the peptide and the complex as a function of  $D_{max}$ .



**Suppl. Fig. 4 — Conformational studies.** (A) Top five nontrivial normal modes (mode-7 to mode-11, *top to bottom*) used to sample the conformational space of AD13-MDTCS. (B) The three derived hinge residues within the T domain are shown as cyan spheres in two orthogonal views of the latent unbound crystal structure. (C) The number of structures from each of the five normal modes selected in each of the 200 EOM cycles. (D) Example of conformations of vWF-strep-peptide selected by one of the EOM cycles.





**Suppl. Fig. 5 — Model validation through CLMS-derived restraints.** Graphical representation of the overlap between the space explored by pairs of residues observed to form cross-links (Suppl. Table 1) based on the SAXS-derived structures. Small orange spheres show the positions of C $\alpha$  atoms of the cross-linked residues of the peptide, while the large semi-transparent spheres show the positions of the corresponding residues in the distinct conformations of the peptidase. AD13-MDTCS-E<sup>225</sup>Q and vWF-strep-peptide residues are labelled in red and blue, respectively.



<b>Suppl. Table 1 — Inter- and intramolecular distance restraints derived from CLMS.</b>			
Inter-molecular		Intra-peptidase	
Residues	Max. distance (Å)	Residues	Max. distance (Å)
S <sup>266</sup> –Y <sub>1605</sub>	22	D <sup>117</sup> –K <sup>318</sup>	15
K <sup>318</sup> –D <sub>1622</sub>	15	D <sup>182</sup> –K <sup>318</sup>	15
K <sup>318</sup> –E <sub>1638</sub>	15	E <sup>184</sup> –K <sup>318</sup>	15
K <sup>318</sup> –E <sub>1640</sub>	15	E <sup>184</sup> –K <sup>559</sup>	15
K <sup>318</sup> –E <sub>1655</sub>	15	S <sup>272</sup> –K <sup>368</sup>	30
K <sup>318</sup> –E <sub>1660</sub>	15	S <sup>275</sup> –K <sup>368</sup>	30
K <sup>318</sup> –D <sub>1663</sub>	15	K <sup>318</sup> –Y <sup>468</sup>	30
S <sup>388</sup> –S <sub>1613</sub>	22	K <sup>318</sup> –E <sup>492</sup>	15
K <sup>497</sup> –T <sub>1656</sub>	22	K <sup>318</sup> –K <sup>497</sup>	22
K <sup>497</sup> –E <sub>1660</sub>	15	K <sup>318</sup> –K <sup>497</sup>	30
K <sup>559</sup> –E <sub>1660</sub>	15	K <sup>318</sup> –D <sup>500</sup>	15
		K <sup>318</sup> –D <sup>516</sup>	15
		K <sup>318</sup> –K <sup>559</sup>	30
		K <sup>318</sup> –D <sup>634</sup>	15
		K <sup>318</sup> –E <sup>634</sup> /D <sup>635</sup>	15
		T <sup>358</sup> –K <sup>368</sup>	22
		K <sup>364</sup> –K <sup>368</sup>	22
		T <sup>435</sup> –T <sup>441</sup>	22
		T <sup>453</sup> –K <sup>440</sup>	22
		E <sup>492</sup> –K <sup>559</sup>	15
		K <sup>497</sup> –K <sup>608</sup>	22
		K <sup>559</sup> –T <sup>575</sup>	22

ADAMTS-13 residues with superscript numbers according to UP Q76LX8; vWF residue numbers in subscript according to UP P04275. Maximal C $\alpha$ -C $\alpha$  distances of 15, 30, and 22 Å result from reagents EDC/NHS, BS3, and DSG or DSA, respectively [1].

[1] Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* 2014;23:747-59.

Crosslinker	[M+H] <sup>+</sup>	Error (ppm)	Sequence 1	Protein 1	From	To	Sequence 2	Protein 2	From	To	Linked residue 1	Linked residue 2
<b>DSG</b>												
<i>Intermolecular</i>												
	5926.9855	0	AGLAWSPCSR	AD13 (M-S)E <sub>225</sub> Q	258	267	EQAPNLVYMTGNPASDEIKRLPGDIQVVPIGVGNANVQELER	VWF-Strep-pep	1599	1641	S266	Y1605
	5697.8783	0.43	WSSWGPR	AD13 (M-S)E <sub>225</sub> Q	386	392	EQAPNLVYMTGNPASDEIKRLPGDIQVVPIGVGNANVQELER	VWF-Strep-pep	1598	1641	S388	S1613
	4348.316	-0.1	AIGESFIMKR	AD13 (M-S)E <sub>225</sub> Q	489	498	IGWPNAPILIQDFETLPREAPDLVLR	VWF-Strep-pep	1642	1668	K497	T1656
	5296.7495	0.59	AIGESFIMKRGDSFLDGTR	AD13 (M-S)E <sub>225</sub> Q	489	507	IGWPNAPILIQDFETLPREAPDLVLR	VWF-Strep-pep	1642	1668	K497	T1656
	3632.6334	0.22	GSFTAGR	AD13 (M-S)E <sub>225</sub> Q	560	566	GGWSPHPQFEKGGSGGGSGGWSHPQFEK	VWF-Strep-pep	1669*	1697*	S561	S1673*
<i>Intramolecular</i>												
	2531.263	-0.1	WCSKGR	AD13 (M-S)E <sub>225</sub> Q	365	370	LLVPLLDGTECGVEK	AD13 (M-S)E <sub>225</sub> Q	350	364	K368	T358
	3749.577	-0.6	TQLEFMSQQCAR	AD13 (M-S)E <sub>225</sub> Q	441	452	ACVGADLQAEMCNTQACEK	AD13 (M-S)E <sub>225</sub> Q	422	440	T441	T435
	4516.9711	-0.1	TDGQPLR	AD13 (M-S)E <sub>225</sub> Q	453	459	ACVGADLQAEMCNTQACEKTQLEFMSQQCAR	AD13 (M-S)E <sub>225</sub> Q	422	452	T453	K440
	4619.4536	0.31	KGSFTAGRAR	AD13 (M-S)E <sub>225</sub> Q	559	568	EYVTFLLVTPNLSVYIANHRPLFTHLAVR	AD13 (M-S)E <sub>225</sub> Q	569	598	K559	T575
<b>DSA</b>												
<i>Intramolecular</i>												
	2755.44	-0.1	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	AIGESFIMKR	AD13 (M-S)E <sub>225</sub> Q	489	498	K318	K497
	2332.16	0.1	LLVPLLDGTECGVEK	AD13 (M-S)E <sub>225</sub> Q	350	364	WCSK	AD13 (M-S)E <sub>225</sub> Q	365	368	K364	K368
	6800.19	0.2	YVVGKMSISPDTPYSLLEDGR	AD13 (M-S)E <sub>225</sub> Q	603	625	SSPGGASFYHWGAAPVHSGQDALCRHMCRAIGESFIMK	AD13 (M-S)E <sub>225</sub> Q	460	497	K608	K497
<b>EDC/NHS</b>												
<i>Intermolecular</i>												
	3890.06021	-3.68	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	LPGDIQVVPIGVGNANVQELER	VWF-Strep-pep	1619	1641	K318	D1622
	3890.07557	0.26	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	LPGDIQVVPIGVGNANVQELER	VWF-Strep-pep	1619	1641	K318	E1638
	3890.06873	-1.49	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	LPGDIQVVPIGVGNANVQELER	VWF-Strep-pep	1619	1641	K318	E1640
	3555.88441	-2.5	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	IGWPNAPILIQDFETLPR	VWF-Strep-pep	1642	1659	K318	E1655
	2516.34262	-0.64	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	EAPDLVLR	VWF-Strep-pep	1660	1668	K318	E1660
	2516.34172	-0.99	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	EAPDLVLR	VWF-Strep-pep	1660	1668	K318	D1663
	2173.18669	3.18	AIGESFIMKR	AD13 (M-S)E <sub>225</sub> Q	489	498	EAPDLVLR	VWF-Strep-pep	1660	1668	K497	E1660
	1844.99905	0.72	KGSFTAGR	AD13 (M-S)E <sub>225</sub> Q	559	568	EAPDLVLR	VWF-Strep-pep	1660	1668	K559	E1660
<i>Intramolecular</i>												
	2466.27553	1.8	DPSLGAQFR	AD13 (M-S)E <sub>225</sub> Q	117	125	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	D117	K318
	2651.34217	0.87	FDLELPDGNR	AD13 (M-S)E <sub>225</sub> Q	181	190	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	E184	K318
	2651.33584	-1.52	FDLELPDGNR	AD13 (M-S)E <sub>225</sub> Q	181	190	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	D182	K318
	1979.99513	0.89	FDLELPDGNR	AD13 (M-S)E <sub>225</sub> Q	181	190	KGSFTAGR	AD13 (M-S)E <sub>225</sub> Q	559	566	E184	K559
	2471.29725	1.41	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	AIGESFIMK	AD13 (M-S)E <sub>225</sub> Q	489	497	K318	E492
	2443.2145	-1.71	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	GDSFLDGTR	AD13 (M-S)E <sub>225</sub> Q	499	507	K318	D500
	3016.44693	1.05	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	EDGTLSLCVSGSCR	AD13 (M-S)E <sub>225</sub> Q	515	528	K318	D516
	2279.19913	1.15	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	VALTEDR	AD13 (M-S)E <sub>225</sub> Q	630	636	K318	D634
	2279.19064	-2.57	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	VALTEDR	AD13 (M-S)E <sub>225</sub> Q	630	636	K318	D635
	1799.94593	-0.74	AIGESFIMK	AD13 (M-S)E <sub>225</sub> Q	489	497	KGSFTAGR	AD13 (M-S)E <sub>225</sub> Q	559	566	E492	K559
<b>BS3</b>												
<i>Intramolecular</i>												
	2144.17713	0.66	RQLLSLLSAGR	AD13 (M-S)E <sub>225</sub> Q	268	278	WCSKGR	AD13 (M-S)E <sub>225</sub> Q	365	370	S272	K368
	2144.17617	0.22	RQLLSLLSAGR	AD13 (M-S)E <sub>225</sub> Q	268	278	WCSKGR	AD13 (M-S)E <sub>225</sub> Q	365	370	S275	K368
	4247.04879	3.53	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	SSPGGASFYHWGAAPVHSGQDALCR	AD13 (M-S)E <sub>225</sub> Q	460	484	K318	Y468
	2413.28848	0.07	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	SFIMKR	AD13 (M-S)E <sub>225</sub> Q	493	498	K318	K497
	2455.29041	-0.43	VAFGPKAVACTFAR	AD13 (M-S)E <sub>225</sub> Q	313	326	KGSFTAGR	AD13 (M-S)E <sub>225</sub> Q	559	566	K318	K559

\* These residues belong to the C-terminal Strep-tag.

**Suppl. Table 2 — Original data from CLMS.** The table depicts the inter- and intramolecular peptides of AD13-MDTCS-E<sup>225</sup>Q and vWF-strep-peptide cross-linked with reagents DSG, DSA, EDC/NHS or BS3 and subsequently enzymatically digested. The residue numbers are from UP Q76LX8 and UP P04275, respectively.

**Suppl. Table 3 — Data collection and scattering-derived parameters.**

<i>Instrumentation and software</i>	vWF-Strep peptide (Batch)	AD13-MDTCS-vWF complex (SEC-SAXS)
Instrument /	EMBL-P12 BioSAXS	EMBL-P12 BioSAXS
Beam geometry	Point collimation ()	Point collimation ()
Detector	Pilatus6M	Pilatus6M
Wavelength (nm)	0.124	0.124
Sample to detector distance	3.0 m	3.0 m
Exposure time	0.05s*40 frames	1s*600 frames
Measured $q$ range (nm <sup>-1</sup> )	0.02-7.32	0.02-7.32
Temperature (K)	293	293
<i>Software employed</i>		
Primary data reduction	<i>SASFLOW</i>	<i>SASFLOW</i>
Data processing	<i>PRIMUSQT</i>	<i>CHROMIXS</i>
Guinier Analysis	<i>PRIMUS QT</i>	<i>PRIMUS QT</i>
IFT Analysis	<i>GNOM 4.6</i>	<i>GNOM 4.6</i>
<i>Ab initio</i> analysis	-	-
Ensemble generation	<i>RanCh</i>	<i>Normal mode Analysis + RanCh</i>
Ensemble Fitting	<i>EOM</i>	<i>EOM</i>
Three-dimensional graphics representations	<i>PyMOL, VMD</i>	<i>PyMOL, VMD</i>

<i>Data-collection and derived parameters</i>			
	vWF-Strep peptide (Batch)		AD13-MDTCS-vWF complex (SEC-SAXS)
Concentration (mg mL <sup>-1</sup> )	3 mg/mL	6 mg/mL	6 mg/mL
Calculated monomeric ( $M_r$ ) from sequence	11.2	11.2	76.7
Structural parameters <sup>†</sup>			
$I(0)$ (AU) [from $P(r)$ ]	$0.046 \pm 3.9E-04$	$0.02 \pm 0.12e-03$	$8723 \pm 14.72$
$R_g$ (nm) [from $P(r)$ ]	$3.4 \pm 0.06$	$3.4 \pm 0.04$	$4.49 \pm 0.14$
$I(0)$ (AU) (from Guinier)	$0.045 \pm 3.1e-04$	$0.019 \pm 8.4e-05$	$8686.41 \pm 14.98$
$R_g$ (nm) (from Guinier)	$3.08 \pm 0.03$	$3.04 \pm 0.02$	$4.38 \pm 0.01$
$D_{max}$ (nm)	14	14	16.3
Molecular mass ( $M_r$ ) [Bayesian Estimate]	14.1	12.0	113.6
SASBDB Entry	SASDKT8	SASDKU8	SASDKV8

## RESULTS: PROJECT 2

---

*“Structure, function, zymogenic latency and catalytic mechanism of a plant glutamate peptidase with therapeutic potential in coeliac disease”*



## Summary

Ingestion of wheat, barley, oats and rye, may cause autoimmune responses including celiac disease, dermatitis herpetiformis and wheat allergy in gluten intolerant individuals. Gluten is a mixture of proline and glutamine-rich prolamin and glutenin molecules that render them resistant to cleavage by gastric, pancreatic and intestinal brush-border membrane peptidases. The autoimmune reactions result in the development of small-intestinal atrophy and mucosal inflammation. Treatment for gluten intolerance commonly involves a strict and lifelong gluten free diet. However, complete avoidance of gluten is a challenge due to cross-contamination and/or the presence of small traces of gluten in foods and medicines. Therefore, effective alternative treatments of gluten intolerance such as oral enzyme therapy are required. Neprosin is a glutamic protease from a carnivorous pitcher plant, commonly known as monkey cups in tropical regions. Given its high proteolytic activity at low pH and high cleavage specificity, neprosin is especially suited to detoxify gluten proteins and gluten-derivatives including wheat gliadin and its immunogenic 33-mer fragment by degrading them down to non-toxic peptides in the human stomach. Importantly, neprosin appears to be far more efficient than AN-PEP, a prolyl endoprotease from *Aspergillus niger*, which represents the currently most studied candidate to date.



# Structure, function, zymogenic latency and catalytic mechanism of a plant glutamate-type peptidase with therapeutic potential in coeliac disease

Laura del Amo-Maestro <sup>1</sup>, Ulrich Eckhard <sup>1,#</sup>, Arturo Rodríguez-Banqueri <sup>1,#</sup>, Soraia R. Mendes <sup>1,#</sup>, Tibisay Guevara <sup>1</sup>, Francisco J. Pérez-Cano <sup>2\*</sup> and F. Xavier Gomis-Rüth <sup>1,\*</sup>

<sup>1</sup> Proteolysis Laboratory; Department of Structural Biology; Molecular Biology Institute of Barcelona (CSIC); Barcelona Science Park; c/Baldiri Reixac, 15-21; 08028 Barcelona (Catalonia, Spain).

<sup>2</sup> Autoimmunity and Tolerance's Group; Department of Biochemistry and Physiology, Section of Physiology; Faculty of Pharmacy and Food Science, University of Barcelona; Av. Joan XXIII, 27-31; 08028 Barcelona (Catalonia, Spain).

\* Corresponding authors: e-mail: franciscoperez@ub.edu and xgrcri@ibmb.csic.es.

# Shared second co-authorship.

*Keywords:* human gastric digestion; coeliac disease; proline-specific endopeptidase; 33-mer peptide; gluten; gliadin; proteolytic mechanism; X-ray crystal structure; zymogen; active-site mutant.

---

Incomplete digestion of dietary gluten generates proline- and glutamine-rich toxic peptides that cause coeliac disease in predisposed individuals. The condition affects >1% of humanity by damaging the small-intestine lining, and a strict gluten-free diet is the only current treatment. Recently, therapeutic development has focused on 'glutenases' for oral enzyme therapy to cleave the toxic peptides, among which a 33-residue peptide from the gluten component  $\alpha$ -gliadin (33-mer) is predominant. Neprosin from the digestive fluid of a carnivorous pitcher plant is an efficient proline endopeptidase for proteomics and a potential glutenase. We developed a high-yield expression system for the enzyme and found it has a pH and substrate profile (Q-P\*Q-L) ideally suited to cleave both the 33-mer and gliadin. Molecular studies including crystal structures and a cohort of mutants revealed that neprosin is biosynthesised as a zymogen, which is self-activated at very acidic pH through release of a 108-residue pro-domain, which features a three-stranded antiparallel  $\beta$ -sheet and is required for proper folding and thermal stability. It laterally attaches to the 255-residue mature moiety, which atypically for peptidases is a 7+8-stranded antiparallel  $\beta$ -sandwich. The strongly curled front  $\beta$ -sheet encloses an extended active-site cleft, which runs obliquely on the concave face of the sheet and features an unprecedented pair of catalytic glutamates assisted by two tyrosines. The linker connecting both domains extends across, and thus blocks, the cleft like a substrate and contains a lysine, which is pivotal for zymogen activation following a 'pH switch' that is reminiscent of the otherwise structurally unrelated aspartate peptidases. Thus, neprosin fulfils *in vitro* criteria for a therapeutic glutenase and founds a new eukaryotic family of the poorly studied glutamate endopeptidases, with new activation and catalytic mechanisms.

## 1. INTRODUCTION

Coeliac disease (CD) is a chronic autoimmune enteropathy that affects individuals with genetic and environmental sensitisation to dietary gluten (1-4). The latter englobes crop storage proteins, dubbed prolamines, with a high content in prolines and glutamines from barley, rye, wheat, as well as triticale, the hybrid of the latter two. Prolamins causing CD include gliadin, glutenin, hordein, and secalin (2), and protracted ingestion of merely ~10 mg gluten per day suffices to cause damage in coeliacs (5). CD is a global health burden across all ages, with a worldwide serological prevalence of 1.4% (6) and an incidence increasing by 7.5% yearly (7).

The disease is caused by partially degraded gluten peptides, among which a 33-residue segment of wheat  $\alpha$ -gliadin (33-mer) is particularly pathogenic (8, 9). These peptides are recalcitrant to further cleavage by gastric, pancreatic and intestinal brush-border membrane peptidases owing to their high proline content (10). In coeliacs, they cross the intestinal epithelium and reach the lamina propria of the small-intestine mucosa, in which glutamines are deamidated by tissue transglutaminase. This enhances the affinity of the peptides for the DQ2.5/DQ2.2 or DQ8 alleles of the human leukocyte antigen (HLA) receptor on the surface of antigen-presenting cells (11, 12). These alleles are found in 90-95% and 5-10% of coeliacs,



respectively, and are required for the development of the disease (13). Receptor binding unlatches an exacerbated pro-inflammatory autoimmune T-cell mediated response, which mainly involves interleukin 15 (14) and has systemic implications including intraepithelial lymphocytosis, crypt hyperplasia, atrophy of the small-bowel villi, and mucosal inflammation. These cause chronic malabsorption of nutrients, diarrhoea, vomits, bloating, abdominal pain, and intestinal lymphomas, as well as extra-intestinal manifestations such as delayed puberty, short stature, osteoporosis, anaemia, dermatitis herpetiformis, axonal neuropathy, and cerebellar ataxia (2, 11, 15). Overall, coeliacs have a shorter life expectancy than the general population (16). Currently, there is no treatment for CD except a lifelong strict gluten-free diet (GFD), which generally restores the normal architecture of the intestinal villi (17). However, a significant fraction of coeliacs on GFD (20-37%) still suffers from intestinal symptoms (17, 18). Moreover, GFD is unbalanced and leads to nutritional deficiencies caused by reduced uptake of fiber, trace elements and vitamins (2). In addition, GFD implementation is a colossal defiance in western societies given the ubiquitous exposure to gluten-containing products, which include processed foods and medicines. Thus, there is a substantiated need to develop new therapeutic approaches against CD.

In recent years, one research line has focused on the development of endopeptidases that cleave the toxic peptides and may thus act as *bona fide* 'glutenases' for oral enzyme therapy (10, 11, 19-21), which is reminiscent of the use of lactase tablets by people with lactose intolerance (22). To be really useful for the clinic, a candidate glutanase must meet several criteria (12, 23, 24). First, it should be functional in the stomach during digestion before the gastric content exits into the duodenum and initiates the pathogenic autoimmune response (25). This requires candidates to be highly stable and potently active, as well as resistant to gastric pepsin, at extreme acidic conditions (pH ~2.5). Second, it should be catalytically very efficient in front of gliadin and the 33-mer in combination with pepsin under gastric conditions, so it can be administered at low therapeutic dosage. Dietary gluten amounts to daily ~15 g in a typical western diet (26), which adds up to proteins from meat, dairy, fish, etc. to make up 10-35% of a reference diet (27). This substrate load is a formidable challenge for any peptidase. Thirdly, a glutanase should be harmless

towards intestine structures and nutrient absorption, thus ideally inactive at neutral or slightly acidic pH values when entering the duodenum.

To date, several glutamyl and prolyl endopeptidases (PEP) from bacteria, fungi, insects and germinating cereals have been studied for therapy (2, 10, 23, 28, 29). The currently most promising candidates are a serine-type PEP from *Aspergillus niger* (AN-PEP) (11, 30-34); STAN1, which is a combination of *Aspergillus niger* aspartic-type aspergillopepsin and serine-type dipeptidyl peptidase IV from *Aspergillus oryzae* (35); latiglutenase, which is a combination of a glutamine-specific cysteine peptidase from barley and a modified serine-type prolyl oligopeptidase from *Sphingomonas capsulata* (19, 36); and KumaMax and Kuma062/TAK-062, which are synthetic enzymes obtained by computational redesign of kumamolysin, a serine endopeptidase from the acidophilic bacterium *Alicyclobacillus sendaiensis* (24, 37, 38). Despite all the efforts made, none of the candidates fulfils all the above requirements, so that results from reported and ongoing clinical trials do not yet evince either significant clinical remission in coeliacs or suitability of these enzymes to safely replace GFD (12). What is more, many self-called enzyme preparations are currently sold over the counter as dietary supplements against CD but have not proven to inactivate toxic gluten peptides, so they may actually represent a hazard for patients (11, 39).

Neprosin is a 383-residue protein, which was identified in the digestive fluid of the carnivorous pitcher plant *Nepenthes ventrata* as a PEP (40-43) and grouped within the unassigned family U74 of unknown class and mechanism in the MEROPS database ([www.ebi.ac.uk/merops](http://www.ebi.ac.uk/merops); (44)). It was assessed in the context of the protease fraction of the fluid as a potential glutenase (41). Moreover, the purified enzyme was deemed useful as a reagent for proteomics (42, 43). Here, we developed a high-yield eukaryotic production system for the recombinant neprosin zymogen, determined its activation pattern *in vitro* and assayed its thermic stability, pH profile, and general proteolytic and peptidolytic activities. We determined its inhibitory profile in front of a cohort of peptidase inhibitors and performed cleavage studies in front of gliadin and the 33-mer to assess its potential as a glutanase. Finally, we crystallised and solved the structures of the neprosin zymogen and its mature form in a product-mimicking complex. All these data

revealed the mechanism of latency, active-site architecture, catalytic mechanism and peptidase class adscription, which were validated by a cohort of mutants.

## 2. RESULTS AND DISCUSSION

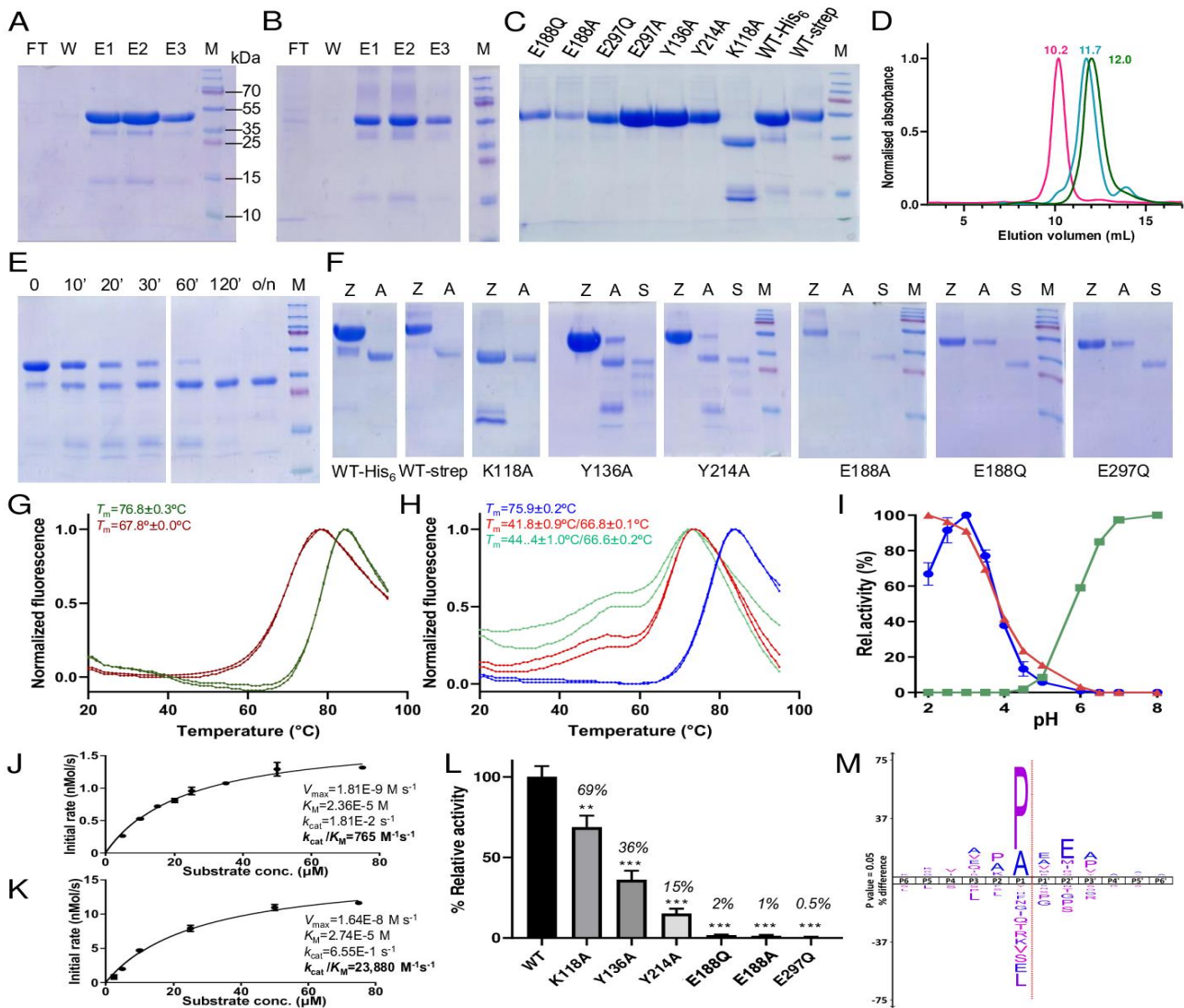
**2.1. Recombinant expression, autolytic maturation, and stability analyses** — Previous studies with neprosin were carried out mostly with purified enzyme from pitcher plant fluid as a recombinant expression system in *Escherichia coli* had only a “modest yield” of partially impure enzyme (41, 42). In fact, our efforts to reproduce this method failed to produce enough functional enzyme for further studies. We hypothesised that a eukaryotic post-translational machinery may be required for proper production and folding of neprosin. Indeed, two *N*-glycosylation sites linked to N<sup>145</sup> and N<sup>152</sup> were identified in posterior structural studies (see below). Thus, we developed an expression system based on human Expi293F cells, which yielded remarkable ~10 mg and ~8 mg of pure well-folded full-length protein with either a C-terminal hexahistidine (His<sub>6</sub>; 41 kDa) or a streptavidin (Strep; 43 kDa) tag per litre of expression medium (Fig. 1A,B,D). We observed that the protein was stable over weeks at 4°C at neutral pH and lacked proteolytic activity, which we attributed to the full-length protein acting as a latent zymogen. By contrast, both variants underwent prompt autolytic maturation at bond P<sup>128</sup>-S<sup>129</sup> (for residue numbering of neprosin in superscript, see UniProt access code [UP] C0HLV2) over time by incubation under acidic conditions to yield bands of 29 plus 11 kDa and 31 plus 11 kDa, respectively (Fig. 1D,E). These corresponded to mature neprosin variants, which were apparently equivalent to the authentic protein purified from *Nepenthes* (41), and the respective excised pro-domains. Both pro-neprosin and neprosin migrated as monomers in calibrated size-exclusion chromatography (SEC) (Fig. 1D), and differential scanning fluorimetry analysis by the thermofluor approach (45) revealed a temperature of midtransition ( $T_m$ ) of 68°C for the mature enzyme (Fig. 1G), which is remarkable for an enzyme naturally operating at ambient temperature and rather corresponds to values reported for hyperthermophilic enzymes (46). Moreover, the zymogen had a  $T_m$  that was 9°C higher (Fig. 1G), which supports a role for the pro-domain in stability and, possibly, correct folding of the full-length protein as reported for other zymogens (47,

48). Indeed, such a role was underpinned by our failure to obtain mature neprosin without the pro-domain using the same expression system (data not shown). Finally, thermofluor studies in the presence of a reducing agent revealed  $T_m$  values that were 32-34°C lower in a concentration-dependent manner and showed at least two transitions in the unfolding process (Fig. 1H). This pointed to the presence of disulphide bonds in the structure important for protein stability (see below).

**2.2. Proteolytic activity** — We assayed the pH dependence of fluorescent bovine serum albumin (BSA) cleavage by neprosin in comparison with the gastric aspartate peptidase pepsin and the pancreatic serine peptidase trypsin (Fig. 1I). We found that neprosin had a pH optimum for activity of 3, which was thus similar to pepsin (pH <2). By contrast, trypsin had optimal activity at pH 7-8, values at which both neprosin and pepsin were completely inactive. But while pepsin was irreversibly inhibited at neutral pH (49), neprosin was reversibly activated and inactivated by switching back and forth three times between pH 9.0 and 2.5, and it could be both frozen and lyophilised at pH 7.5 for storage with total recovery of activity after thawing or resuspension in buffer at acidic pH, respectively. Finally, neprosin was insensitive to digestion by pepsin at acidic pH.

To get insight into the substrate specificity of neprosin, we re-analysed published proteomics data performed mainly with purified material, which had identified the enzyme as a *bona fide* PEP (42). We identified 3001 unique cleavage sites spanning **P<sub>6</sub>-P<sub>6</sub>'** (substrate and active-site sub-site nomenclature according to (50, 51)), of which 1863 (62%) exhibited proline in **P<sub>1</sub>** (Fig. 1M). Proline was also twofold enriched over natural abundance at positions **P<sub>2</sub>** and **P<sub>3</sub>'**, while it was strongly disliked in **P<sub>1</sub>'** and **P<sub>2</sub>'**. Glutamate and methionine were threefold enriched in **P<sub>2</sub>'**, and while alanine was well accepted throughout **P<sub>6</sub>-P<sub>6</sub>'**, glycine was significantly unsuited for **P<sub>1</sub>** to **P<sub>3</sub>'**. In summary, these data revealed strong preference for substrates with a proline in **P<sub>1</sub>** and that specific positions within **P<sub>6</sub>-P<sub>6</sub>'** were unsuited for certain amino acids (Fig. 1M).

**2.3. Cleavage of gliadin and the 33-mer** — We analysed proteolysis of the main wheat gluten component gliadin by neprosin in the presence and absence of pepsin by SDS-PAGE and turbidometry, and compared the results with those for pepsin alone (Fig. 2A,B).

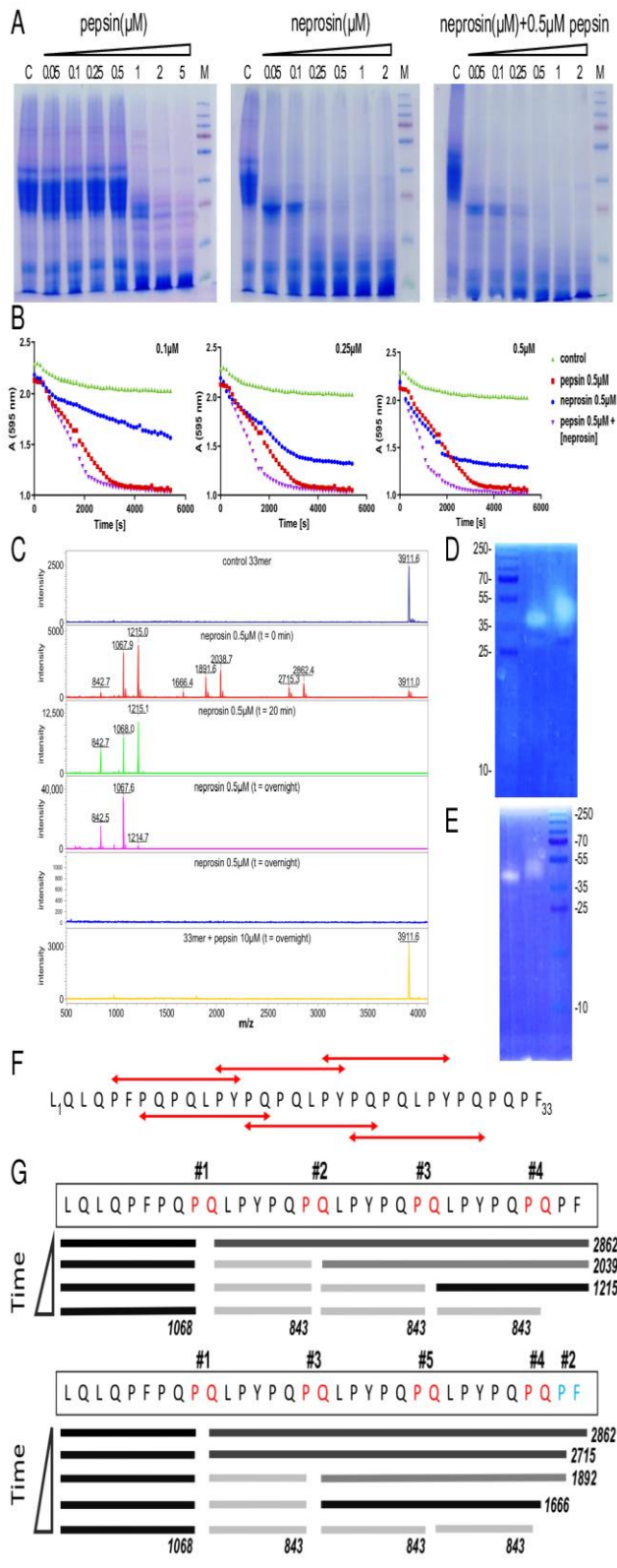


**Figure 1 — Protein purification, stability and activity.** (A) Purification of WT pro-neprosin via His<sub>6</sub>-tag or (B) Strep-tag AC. The flow through (*lane FT*), wash (*lane W*) and elution (*lanes E1-E3*) fractions were analysed by SDS-PAGE and Coomassie staining, respective lanes M correspond to the molecular mass marker. (C) Pro-neprosin mutants (K<sup>118</sup>A, Y<sup>136</sup>A, E<sup>188</sup>A, E<sup>188</sup>Q, Y<sup>214</sup>A, E<sup>297</sup>Q, and E<sup>297</sup>A) after His<sub>6</sub>-tag affinity purification compared with the WT forms of (A) and (B). (D) Profiles of size exclusion chromatography in a Superdex 75 10/300 GL column of pro-neprosin with His<sub>6</sub>-tag (*magenta*), neprosin with His<sub>6</sub>-tag (*blue*), and neprosin with Strep-tag (*green*). Each curve is labelled with the respective elution volume in mL. (E) Autolytic maturation of pro-neprosin at 37 °C over time at acidic pH. (F) Activation of pro-neprosin variants (Z) by acidic autolysis (A) or *in trans* by addition of Strep-tagged neprosin (S). Mutant K<sup>118</sup>A was obtained as a pre-activated protein after AC showing separate pro-domain and mature protein moieties (*lane Z*), which became fully activated by acidic incubation (*lane A*). Mutants Y<sup>136</sup>A, E<sup>188</sup>A, E<sup>188</sup>Q, Y<sup>214</sup>A, E<sup>297</sup>Q could only be activated *in trans*. E<sup>297</sup>A was not activated by either approach (data not shown). (G) Differential scanning fluorimetry showing curves in duplicate of the temperature dependent fluorescence variation upon thermal denaturation of neprosin (*dark red*) and pro-neprosin (*green*). The inset temperatures of midtransition (*T<sub>m</sub>*) result from the average of the inflection points of the two respective curves. (H) Same as (G) illustrating the effect of the TCEP reducing agent at 5 mM (*red*) and 10 mM (*green*) in comparison with untreated pro-neprosin (*blue*). (I) Dependence on pH of the activity of pepsin (*red*), trypsin (*green*), and neprosin (*blue*) in front of fluorescent BSA as substrate. (J, K) Kinetic analysis of neprosin-mediated cleavage of the fluorogenic (J) FS6- and (K) QPQL-peptide. The insets provide the respective values for *V<sub>max</sub>*, *k<sub>cat</sub>*, *K<sub>M</sub>*, and *k<sub>cat</sub>/K<sub>M</sub>*. (L) Peptidolytic activity of WT neprosin and mutants in front of the fluorogenic QPQL-peptide. Student's *t*-test was used to analyse the differences, a *p*-value of 0.05 was considered statistically significant (\*, *p* = 0.05; \*\*, *p* = 0.01; \*\*\*, *p* = 0.001). (M) Logo depicting the substrate preference of neprosin based on reanalysis of the data deposited by (42).

While both enzymes efficiently degraded gliadin separately at concentrations below the physiological threshold of pepsin (~5 μM; (52)), best results were obtained when both enzymes were combined. Remarkably, the optimal

concentration of neprosin added was lower than that of the gastric enzyme. In addition, neprosin degraded gliadin with a comparable efficiency to gelatin in zymography (Fig. 2D,E).

Next, we assayed peptidolysis of the 33-mer, which includes six overlapping immunogenic



**Figure 2**—Neprosin activity against molecules relevant for coeliac disease. **(A)** SDS-PAGE depicting the processing of gliadin by increasing concentrations of pepsin (*left*), neprosin (*centre*) or neprosin plus pepsin (*right*). **(B)** Curves depicting the measurements of gliadin cleavage as in **(A)** over time by turbidometry. **(C)** Mass spectra of, top to bottom, the 33-mer peptide (3911 Da); the 33-mer peptide after incubation with 0.5  $\mu\text{M}$  neprosin at 0 min, 20 min and overnight; neprosin alone; and the 33-mer peptide after overnight incubation with 10  $\mu\text{M}$  pepsin, which leaves the peptide intact. **(D)** Gliadin zymogram depicting activity of neprosin (*left lane*) and mature enzyme resulting from self-activation of pro-neprosin (*right lane*). **(E)** Same as **(D)** but showing gelatinase zymography. **(F)** Sequence of the 33-mer and extent of the six overlapping HLA-DQ2.5-binding epitopes (2, 8). The peptide corresponds to segment L<sub>76</sub>-F<sub>108</sub> of  $\alpha$ -gliadin (UP P18573). **(G)** Cleavage of the 33-mer peptide by neprosin over time proceeds according to two pathways (*top* and *bottom*).

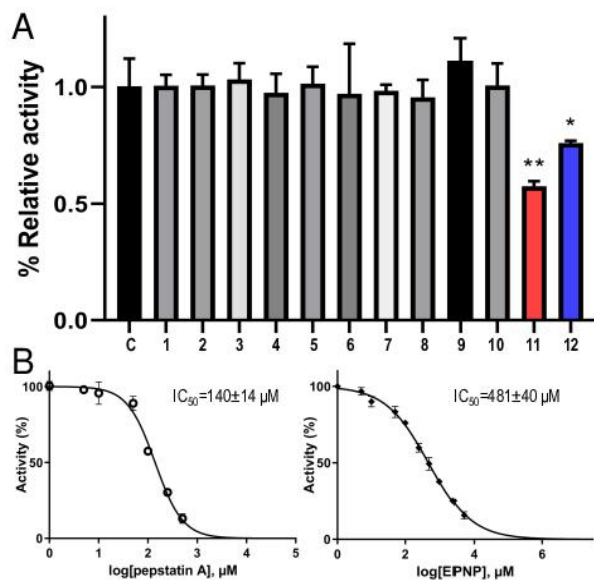
HLA-DQ2.5 T-cell epitopes with a very high content in proline and glutamine residues (8), by mass spectrometry (Fig. 2C). We found that the peptide at 250 mM was efficiently degraded by neprosin at 0.5  $\mu\text{M}$  after 20 min at pH 3. No autolytic cleavage products were detected even after overnight incubation, which corroborates the stability of the mature enzyme under acidic conditions. By contrast, pepsin failed to cleave the peptide even after overnight incubation at 20-fold higher concentration than neprosin (Fig. 2F). Analysis of the peptide cleavage fragments generated by neprosin revealed two final products, Q-L-P-Y-P-Q-P (843 Da) and L-Q-L-Q-P-F-P-Q-P (1068 Da). Monitoring cleavage over time (Fig. 2G) indicated that cleavage only occurred after five of the 13 prolines of the 33-mer, preferably at P-Q-P\*Q-L-P and always with P-Q-P in P<sub>1</sub>-P<sub>3</sub>, which qualifies the simple specificity for proline in P<sub>1</sub> derived from proteomics (see above and (42)). Overall, our results demonstrate that the 33-mer is degraded at multiple sites with the Q-P\*Q-L motif. Remarkably, a P-Q dipeptide is also found in BSA, which explains why it is a suitable substrate for neprosin (see above).

Finally, a cohort of fluorogenic peptides available in the laboratory was assayed for cleavage by neprosin. We found that a peptide containing a P-L bond ('FS6-peptide'; Mca-K-P-L-G-L-Dpa-A-R-NH<sub>2</sub>), which is a substrate of metalloproteinases (53), was cleaved with modest efficiency according to kinetic studies ( $k_{\text{cat}}/K_M = 765 \text{ M}^{-1}\text{s}^{-1}$ ; Fig. 1J). By contrast, a FS6 variant redesigned to include the neprosin cleavage site of the 33-mer, Q-P-Q-L ('QPQL-peptide'; Mca-Q-P-Q-L-Dpa-A-R-NH<sub>2</sub>), was cleaved 30-times more efficiently ( $k_{\text{cat}}/K_M = 23,880 \text{ M}^{-1}\text{s}^{-1}$ ; Fig. 1K). Thus, neprosin is a PEP with a sophisticated specificity beyond mere P-X bonds that is particularly suited to efficiently degrade the 33-mer.

**2.4. Inhibitory profile** — Given the unknown catalytic class of the enzyme, we next tested inhibition of neprosin cleavage of the QPQL-peptide by a cohort of peptidase inhibitors including the cOmplete broad-spectrum inhibitor



cocktail; the metallopeptidase inhibitors 1,10-phenanthroline, phosphoramidon, marimastat, and captopril; the serine peptidase inhibitor 4-(2-aminoethyl)-benzenesulfonyl fluoride; and the aspartate peptidase inhibitors pepstatin A, methyl-2-[(2-diazoacetyl)amino]hexanoate, and 2-[(4-nitrophenoxy)methyl]oxirane (EPNP). In addition, we followed an approach recently employed to find inhibitors of pyrroline-5-carboxylate reductase (54), whose product is proline, and assayed a cohort of proline containing/mimicking compounds including 2-acetyl-1-methylpyrrole; (*S*)-tert-butyl-2-(3-ethoxy-3-oxopropanoyl)pyrrolidine-1-carboxylate; and *N*-boc-glycylproline (Fig. 3A). We found that only pepstatin A and EPNP weakly though significantly inhibited neprosin, with derived  $IC_{50}$ -values of 140 and 480  $\mu$ M, respectively (Fig. 3B). Thus, given that pepstatin and 4-nitrophenoxy-based epoxides are inhibitors of pepsin-type aspartate endopeptidases (55, 56), which share no sequence similarity with neprosin, this pointed to an unexpected peptidase type and mechanism of catalysis.



**Figure 3 — Inhibitory studies.** (A) Effect of the tester molecules (1) 1,10-phenanthroline; (2) AEBSEF; (3) phosphoramidon; (4) marimastat; (5) cOmplete; (6) BGP; (7) captopril; (8) DAN; (9) BEOPC; (10) AMP; (11) pepstatin A; and (12) EPNP in comparison with the WT control (C). Only the latter two elicit a significant inhibition. Student's *t*-test was used to analyse the differences, a *p*-value of 0.05 was considered statistically significant (\*, *p* = 0.05; \*\*, *p* = 0.01; \*\*\*, *p* = 0.001). (B) Plot of the inhibitory activity of pepstatin A (*left*) and EPNP (*right*) in front of tester concentration with the derived  $IC_{50}$  values.

**2.5. Structure analysis of latent and mature neprosin** — To get insight into the molecular determinants of neprosin structure and function,

we crystallised pro-neprosin in an orthorhombic space group (Fig. 4A, *left* and Table 1) and found that the polypeptide was cleaved at the maturation site (P<sup>128</sup>-S<sup>129</sup>), thus crystals contained the zymogenic complex between the cleaved pro-domain and the mature moiety (Fig. 5A, *centre*). We solved the structure by single-wavelength anomalous diffraction with data collected at the lutetium peak wavelength from a derivative crystal obtained with the ‘crystallophore’ compound Lu-Xo4 (57)(Fig. 4B), which evinced substantial difference in one of the crystal cell axes compared to the native crystals (Table 1). The final refined model of the derivative complex was used to solve the native pro-neprosin structure by molecular replacement. Moreover, mature neprosin was crystallised in a monoclinic space group (Fig. 4A, *right* and Table 1). Its structure was also solved by molecular replacement, by using the coordinates of the mature part of the native zymogen.

Pro-neprosin is a compact oblong molecule of  $\sim 55 \text{ \AA}$  (*w*)  $\times \sim 45 \text{ \AA}$  (*h*)  $\times \sim 40 \text{ \AA}$  (*d*) approximate dimensions (Fig. 4C). The N-terminal pro-domain (D<sup>21</sup>-P<sup>128</sup>) is defined for A<sup>29</sup>-P<sup>121</sup> and features a globular part (A<sup>29</sup>-G<sup>112</sup>) followed by a linker (L<sup>113</sup>-P<sup>121</sup>). The former contains an antiparallel three-stranded  $\beta$ -sheet ( $\beta 2 \downarrow - \beta 1 + \beta 3 \uparrow - \beta 4 \downarrow$ ) in which the central strand is divided in two by the insertion of leftmost strand  $\beta 2$  (Fig. 4C,D,E). A long segment connects  $\beta 3$  and  $\beta 4$ , which includes short  $\alpha$ -helices  $\alpha 1$  and  $\alpha 2$ , a first of three disulphide bonds (C<sup>52</sup>-C<sup>98</sup>), and a disordered ten-residue segment in the back (Y<sup>77</sup>-N<sup>86</sup>). After  $\beta 4$ , the chain undergoes a 90°-turn and enters the linker, which runs in extended conformation along the surface of the mature enzyme moiety connecting the globular pro-domain moiety with the mature enzyme. After a second flexible segment (N<sup>122</sup>-N<sup>131</sup>), which includes the cleaved maturation site, the polypeptide enters mature neprosin (S<sup>129</sup>-A<sup>383</sup>), which is defined for T<sup>132</sup>-I<sup>382</sup>. Atypically for peptidase catalytic domains, which are generally  $\alpha/\beta$ -proteins (58), neprosin is an all- $\beta$  protein consisting of an antiparallel  $\beta$ -sandwich (Fig. 4C,D), with a seven-stranded front sheet ( $\beta 10 + \beta 12 \downarrow - \beta 9 \uparrow - \beta 8 \downarrow - \beta 7 \uparrow - \beta 16 \downarrow - \beta 5 \uparrow - \beta 20 \downarrow$ ) and an eight-stranded back sheet ( $\beta 11 \downarrow - \beta 15 \uparrow - \beta 14 \downarrow - \beta 13 \uparrow - \beta 6 \downarrow - \beta 17 + \beta 18 \uparrow - \beta 19 + \beta 22 \downarrow - \beta 21 \uparrow$ ). The front sheet is strongly curled (Fig. 4C, *right*), has its leftmost strand divided in two ( $\beta 10 + \beta 12$ ) by the insertion of  $\beta 11$  from the back sheet, and has connectivity (59) -2x, -1, -1, -1, 4x, 2x. The back sheet has two of its strands divided by a bulge

( $\beta 17+\beta 18$ ) and by the insertion of two strands ( $\beta 19+\beta 22$ ), respectively. It shows connectivity -4x, 3, -1, -1, 4x, 1, 1 and provides a scaffold for the front sheet.

The two sandwich sheets are glued by a central core of mostly hydrophobic side chains, which include unbound C<sup>334</sup> from back-sheet strand  $\beta 18$ , and are arranged in parallel, so that the respective  $\beta$ -strands are not rotated away from each other as typically found in  $\beta$ -sandwiches and  $\beta$ -barrels but roughly coincide in position and direction in a frontal view (Fig. 4C,D). Moreover, two glycosylation sites were found in the cross-over linker connecting strands  $\beta 5$  and  $\beta 6$  (L $\beta 5\beta 6$ ) from either sheet on the top and within back-strand  $\beta 6$  in the back, respectively (Fig. 4C). Finally, the front and back sheets are inter-connected by nine cross-over loops, including hairpin  $\beta 12\beta 13$ , and two further disulphides (C<sup>219</sup>-C<sup>224</sup> and C<sup>358</sup>-C<sup>379</sup>) on either side of the sandwich. All these elements contribute to a compact, coherent, and stable domain, which explains its notable pH stability and resistance towards pepsin.

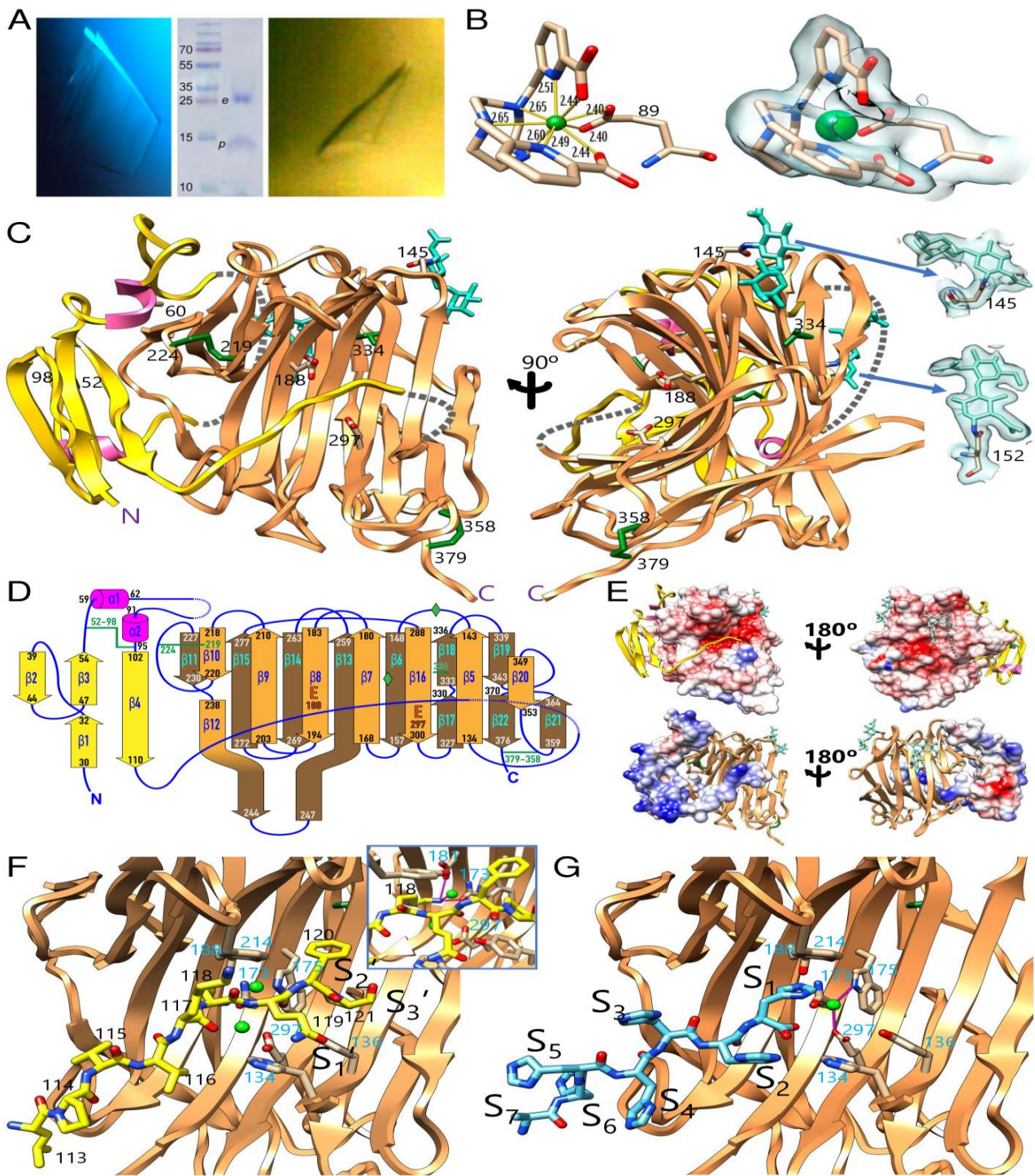
The structure of mature neprosin was further analysed (Table 1) and proved practically identical to the equivalent part of the zymogen (core *rmsd* of 0.62 Å for 250 common C $\alpha$  atoms). The only significant difference was encountered at N<sup>232</sup>-Y<sup>233</sup>, which is folded outward in the zymogen to accommodate I<sup>103</sup> at the beginning of strand  $\beta 4$  of the pro-domain. Thus, the mature moiety is essentially preformed in the neprosin zymogen as seen in most peptidases, with the notable exception of chymotrypsin-type serine peptidases (48, 60-62).

**2.6. The active site** — In the absence of previous molecular data on the function of neprosin, we hypothesised that the active-site cleft would be delineated by the pro-domain linker (Fig. 4F) as found in other zymogens (48, 60, 62). Moreover, in the structure of mature neprosin the C-terminal segment including the His<sub>6</sub>-tag ran along the surface of a symmetry mate, following a very similar chain trace and direction to the pro-domain linker, thus hypothetically mimicking a substrate or product complex (Fig. 4G). Accordingly, neprosin would possess an extended active-site cleft traversing the concave face of the sheet, obliquely to the direction of the front-sheet  $\beta$ -strands by  $\sim 55^\circ$  (Fig. 4C,E-G).

On the search for possible catalytic residues, we were inspired by the functionally, but not structurally, analogous pepsin-type aspartic peptidases, which are mainly  $\beta$ -proteins and

likewise operate at extreme acidic pH values (63). Moreover, the only neprosin inhibitors we could find were aspartate-peptidase inhibitors (see above). These enzymes employ a pair of aspartic residues bridged by a solvent for catalysis (64). Indeed, we found a conspicuous pair of glutamates (E<sup>188</sup> and E<sup>297</sup>) bridged by a solvent molecule in the substrate/product complex structure next to the C-terminus of the bound peptide (Fig. 4G). This pair was similarly arranged in the zymogen structure (Fig. 4F). We thus expressed the E<sup>188</sup>Q, E<sup>188</sup>A, E<sup>297</sup>Q, and E<sup>297</sup>A point mutants of His<sub>6</sub>-tagged pro-neprosin with comparable yields to the wild type (Fig. 1C). These variants did not autoactivate by incubation at acidic pH (Fig. 1F), so activation was performed *in trans* with catalytic amounts of mature Strep-tagged neprosin followed by final purification (Fig. 1F). Finally, the mature variants of the E<sup>188</sup>Q, E<sup>188</sup>A, and E<sup>297</sup>Q but not E<sup>297</sup>A mutants were obtained, which were completely inactive (Fig. 1L), which is in accordance with a role as a catalytic dyad. This result further entailed that the mature neprosin structure would actually mimic a product complex occupying the non-primed of the cleft at sub-sites **S**<sub>7</sub> to **S**<sub>1</sub> (Fig. 5G). As the pro-domain linker protrudes for further three residues on the right side of the cleft, these would delineate sub-sites **S**<sub>1'</sub>-**S**<sub>3'</sub> (Fig. 4F). Together with extra space in the cleft beyond **S**<sub>3'</sub>, neprosin would thus possess an extended cleft harbouring 12 sub-sites (**S**<sub>7</sub> to **S**<sub>5'</sub>). This extended cleft provides an explanation for the efficient processing of the 33-mer through recognition of the P-Q-P\*Q-L-P peptide. Moreover, we identified two further residues in the proximity of the catalytic glutamates that could potentially further participate in catalysis, Y<sup>136</sup> and Y<sup>214</sup> (Fig. 4F,G), and likewise assessed their relevance for function by mutagenesis. Both variants were produced as intact zymogens and purified with variable efficiency, and they underwent autolytic maturation (Fig. 1F). They showed significant activity spanning 36% and 15% of the wild type, respectively (Fig. 1L). Thus, we conclude that they are relevant but not critical for catalysis, possibly playing a role in the Michaelis complex or the stabilisation of the tetrahedral reaction intermediate in an 'oxyanion hole' function (65) (see below).

**2.7. Mechanism of latency** — The pro-domain is laterally attached to the left side of mature neprosin, so that its central  $\beta$ -sheet is rotated  $\sim 90^\circ$  away from the plane of the front sheet



**Figure 4 — Structure analysis of pro-neprosin.** (A) Orthorhombic crystals of pro-neprosin (left panel) contained the complex between the cleaved pro-domain (*p*) and the mature moiety (*e*) (middle panel). Mature enzyme crystals were monoclinic (right panel). (B) The structure of pro-neprosin was solved by means of a lutetium derivative obtained with the ‘crystallophore’ compound Lu-Xo4, which was found at two sites in the crystal. In one site (left panel), the Lu<sup>3+</sup> cation (green sphere) was nona-coordinated by two carboxylate oxygens plus five nitrogens from the organic scaffold and the carboxylate oxygens of protein residue E<sup>89</sup> at distances spanning 2.40–2.65 Å. The cation binding site was unambiguously defined in the final ( $2mF_{\text{obs}} - DF_{\text{calc}}$ )-type Fourier map of the derivative contoured at 1.3  $\sigma$  above threshold (right panel). (C) Ribbon-type plot of pro-neprosin in front view (left panel) and lateral view (right panel) with the pro-domain in gold with helices in magenta. The mature enzyme moiety is in orange. Disordered/cleaved segments are signalled as grey dashed lines. The two glycosylation sites at N<sup>145</sup> and N<sup>152</sup>, the seven cysteines, A<sup>60</sup>, and the two catalytic glutamates (E<sup>188</sup> and E<sup>297</sup>) are shown for their side chains and labelled. The final Fourier map around the two glycan chains is further depicted at 0.6  $\sigma$  above threshold. (D) Topology scheme of pro-neprosin with strands (labelled  $\beta$ 1– $\beta$ 22) and the two short helices ( $\alpha$ 1 and  $\alpha$ 2) as magenta rods. The terminal residues of each secondary structure element are indicated. The pro-domain has yellow strands and magenta helices, the front sheet of the mature enzyme moiety is in orange, the back sheet in brown. The seven cysteines are further indicated in green, the glycans are pinpointed as green rhombi. The catalytic glutamates are marked for reference. (E) (Top) Front view of pro-neprosin as in (C) (left) and back view (right) with the mature moiety shown for its Coulombic surface (red,  $-10$  kcal/mol·*e*; blue,  $+10$  kcal/mol·*e*) computed with Chimera (104) using standard parameters. The pI of the mature moiety is 4.3. (Bottom) Same but with the pro-domain (pI = 9.5) shown for its Coulombic surface.



(F) Close-up of (C) depicting the final segment (L<sup>113</sup>–P<sup>121</sup>) of the pro-domain defined in the final Fourier map as stick model with yellow carbons and black residue numbers running across the active-site cleft. The position of the putative S<sub>1</sub><sup>'</sup>, S<sub>2</sub><sup>'</sup> and S<sub>3</sub><sup>'</sup> sites is indicated. In addition, selected residues of the active site are depicted for their side chains with carbons in tan and numbered in blue, two solvent residues potentially relevant for catalysis are further shown (*green spheres*). The inset provides a slightly rotated close-up view to highlight the interaction (*magenta sticks*) of K<sup>118</sup> with E<sup>181</sup>, Q<sup>173</sup> and a solvent molecule. (G) Same as (F) depicting the product complex of mature neprosin, with the C-terminal tail from a symmetry mate spanning A<sup>383</sup> and the His<sub>6</sub>-tag residues as stick model with carbons in cyan occupying cleft sub-sites S<sub>7</sub> through S<sub>1</sub>. A solvent molecule potentially relevant for catalysis (*green sphere*) is bound to E<sup>297</sup> and W<sup>175</sup> (*magenta sticks*).

of the mature moiety. The interface has a calculated solvation free energy gain upon interface formation ( $\Delta^iG$ ) of -25.8 kcal/mol according to (66) and a top complex significance score (1.0), which imply that the interface is essential for complex formation. Furthermore, the interface spans 1088 Å<sup>2</sup>, with a total buried surface area of 2176 Å<sup>2</sup>, which is larger than the average value in protein-protein complexes (1910 Å<sup>2</sup>; (67)). The interface is shaped by 261 atoms from 71 residues of mature neprosin and 190 atoms from 50 residues of the pro-domain, which perform 22 hydrogen bonds, one single salt bridge (K<sup>118</sup>–E<sup>188</sup>) and one double salt bridge (E<sup>48</sup>–R<sup>204</sup>). Participating segments mainly include T<sup>132</sup>–V<sup>138</sup>, N<sup>166</sup>–S<sup>177</sup>, L<sup>184</sup>–E<sup>188</sup>, Y<sup>194</sup>–G<sup>199</sup>, Q<sup>202</sup>–W<sup>209</sup>, G<sup>213</sup>–T<sup>217</sup>, Y<sup>220</sup>–A<sup>239</sup>, D<sup>259</sup>–Y<sup>282</sup>, T<sup>291</sup>–Y<sup>299</sup>, and I<sup>352</sup> of neprosin and most of the pro-domain (D<sup>46</sup>–V<sup>50</sup>, I<sup>55</sup>–L<sup>66</sup>, H<sup>69</sup>–H<sup>93</sup>, E<sup>97</sup>–P<sup>121</sup>), which snugly embraces the mature moiety (Fig. 4E, *bottom*). Overall, this strong interaction in the zymogen explains its high stability (see above). Indeed, its importance is further highlighted by the fact that a single point mutant destabilising the interface (A<sup>60</sup>R) had a dramatic effect on folding and caused the protein to be no longer expressed (data not shown).

While mature neprosin is negatively charged (theoretical pI = 4.3; Fig. 4E, *top*), the pro-domain is positively charged (pI = 9.5; Fig. 4E, *bottom*), thus complementary to the former. It provides overall stabilisation of the zymogen through electrostatic interactions at neutral or slightly acidic pH values—the theoretical pI of pro-neprosin is 5.9—, which is consistent with intracellular biosynthesis in glands lining the pitcher operating at neutral pH. Once secreted to the digestive fluid, the low pH environment, which entails a several 1000-fold increase in proton concentration, disrupts the electrostatic interactions that keep the zymogen complex by protonating negatively charged residues, thus causing the complex to fall apart. Concretely, the S<sub>1</sub> position of the cleft is occupied by K<sup>118</sup> from the pro-domain linker in the zymogen structure, which was obtained at pH = 7.5. The residue performs a strong salt bridge with catalytic E<sup>188</sup> that would appear relevant for latency (Fig.

4F). Thus, we analysed mutant K<sup>118</sup>A, which was efficiently overexpressed but underwent partial autolytic maturation under neutral conditions at which the wild type and other mutants (see above) remained intact (Fig. 1C), and subsequent incubation at pH = 2.5 completed the autolytic maturation process. The resulting mature mutant displayed 69% activity of the wild type, possibly due to partial autolytic degradation of the zymogen or to an additional effect of the salt bridge in stability and folding. Thus, the K<sup>118</sup>–E<sup>188</sup> pair can be envisaged as a 'latency plug', which once removed at acidic pH, makes the pro-domain linker accessible for maturation cleavage following a 'pH-switch mechanism'. This sequence of events is similar to digestive aspartate peptidases (61, 68) and a lysine functionally equivalent to K<sup>118</sup> is found in pepsin and gastricsin (61). Moreover, this pH-switch mechanism also explains that prosegment interacts in a substrate-like manner with the active site but is not cleaved (compare Figs. 4F and 4G). This contrasts with most zymogens, including digestive aspartate peptidases, in which prosegments interact in non-substrate-like manners with the mature moieties as a mechanism to prevent cleavage (60-62). Finally, given that the scissile-bond position in the cleft is occupied by K<sup>118</sup>–Q<sup>119</sup> but maturation cleavage occurs at P<sup>128</sup>–S<sup>129</sup>, activation probably occurs *in trans* by a second enzyme molecule once the pro-domain linker becomes accessible for cleavage.

**2.8. Structural similarity with glutamate peptidases** — For many years, peptidases were classified into five mechanistic classes: serine, cysteine, threonine, aspartate, and metal-dependent peptidases (69). In the first three, the nucleophile attacking the scissile peptide-bond carbonyl is a class-naming catalytic residue while in the latter two, a polarised solvent molecule exerts this function (70). It was only in 2004 that the first representative of a new class, the glutamate peptidases, was structurally identified in the form of scytalidocarboxyl peptidase B from the dematiaceous fungus *Scytalidium lignicolum* (SCP-B; (71-73)). Since then, only the closely related aspergilloglutamic peptidase (~50% identical in



sequence with SCP-B) has been further analysed for its structure (74, 75) and seven others have been functionally characterised, all from fungi (76-80) except for one bacterial specimen (81). They are grouped into family G1 within the MEROPS database and further dubbed ‘pepstatin-insensitive fungal carboxypeptidase group’ (82) and ‘eqolysins’ (71).

Eqolysins are thermophilic, function under acidic conditions and are pepstatin-insensitive (81). They operate through a catalytic glutamate acting as solvent-polarising general base (E<sub>190</sub> in SCP-B, see UP P15369 for residue numbering according to the full-length protein; subtract 54 for the mature enzyme numbering commonly used in the literature). The glutamate is assisted by a glutamine (Q<sub>107</sub> in SCP-B), thus the family name *eqolysins* (71). These residues are invariant across family members and are flanked by highly similar residues (82, 83). In addition, they share a disulphide bond between cysteines 6-7 positions before the catalytic glutamine and ten positions before the catalytic glutamate, respectively (83). The SCP-B founding member is a 7+7 antiparallel  $\beta$ -sandwich, with a 20-residue signal peptide, a 34-residue pro-domain and a 206-residue enzyme moiety, which shows overall similarity with the neprosin catalytic moiety in structure similarity searches.

Superposition of the substrate-bound mature CDs of SCP-B (Protein Data Bank access code [PDB] 2IFR) and neprosin revealed 140 aligned residues with 14 gaps out of 204 SCP-B and 249 neprosin residues, which showed a core *rmsd* of 3.0 Å and a sequence identity of merely 11%. Thus, while the general architecture and strand connectivity of the two  $\beta$ -sheets is largely conserved, remarkable differences are observed in the strand-connecting loops framing the active site. These include a segment after the  $\beta$ -strand equivalent to  $\beta$ 16 in neprosin, which adopts completely different trajectories for Y<sup>299</sup>–K<sup>325</sup> (F<sub>192</sub>–P<sub>212</sub> in SCP-B) and includes a protruding, disulphide-linked, 15-residue  $\beta$ -hairpin in the fungal enzyme; L $\beta$ 7 $\beta$ 8; and L $\beta$ 8 $\beta$ 9. Above the cleft, neprosin segment A<sup>211</sup>–G<sup>237</sup>, which includes  $\beta$ 10 and  $\beta$ 11, is replaced by the partially disordered shorter segment E<sub>127</sub>–G<sub>134</sub>; this causes adjacent L $\beta$ 7 $\beta$ 8 to be folded outward when compared with SCP-B. Within the active site, the only conserved residue is the catalytic glutamate (E<sup>297</sup> in neprosin and E<sub>190</sub> in SCP-B), as well as the position of the catalytic assistant (E<sup>188</sup> in neprosin and Q<sub>107</sub> in SCP-B), which lead to completely different active-

site clefts with disparate surface shapes and substrate trajectories. Moreover, mutant Q<sub>107</sub>E, which would mimic the neprosin catalytic dyad, completely lacks activity in SCP-B (84). *Vice versa*, the E<sup>188</sup>Q neprosin is likewise completely inactive (see above). This, in turn, is consistent with different substrate specificities, which cause cleavage of F–F, L–Y, and F–Y bonds of insulin by SCP-B (84), while neprosin is a PEP ideally shaped to cleave Q–P\*L–Q (see above).

Overall, while the general architecture of neprosin agrees with the catalytic moiety of eqolysins, a much longer pro-domain (108 vs. 34 residues), the whole active-site environment, the substrate binding modus and specificity, as well as the chemical mechanism diverge. Finally, while eqolysins are restricted to fungi and bacteria, potential neprosin orthologs with ~35–40% sequence identity are widely found in plants only, including gluten-containing crops.

### 3. MATERIALS AND METHODS

**Protein production and purification** — A cDNA fragment encoding wild-type (WT) neprosin from *Nepenthes ventrata* without the signal peptide for secretion (R<sup>25</sup>–Q<sup>380</sup>; see UniProt entry [UP] C0HLV2), which is 91% identical to the orthologue from *Nepenthes alata* (UP A0A1L7NZU4), was purchased from GenScript within vector pET-28a(+) (plasmid pET-28a(+)-proNEP; see Suppl. Table 1 for vectors, plasmids and primers used). The coding fragment was cloned into vector pCMV-Sport6, which was kindly provided by Jan J. Enghild from Århus University (Denmark), between *Bst*II and *Ast*SI restriction sites to generate plasmid pS6-proNEP. This plasmid conferred resistance to ampicillin, included the signal peptide from the V-J2-C region of a mouse Ig  $\kappa$ -chain replacing the native leader sequence, and attached a C-terminal hexahistidine (His<sub>6</sub>)-tag to the protein, hereafter pro-neprosin (residues DLMV+R<sup>25</sup>–Q<sup>380</sup>+AIAHHHHHH). The plasmid was used to generate plasmid pS6-proNEP-Strep for the production of pro-neprosin-strep (residues DLMV+R<sup>25</sup>–Q<sup>380</sup>+AIAGGWSHPQFEKGGGSGGGSGGWSHPQFEK) by replacing the His<sub>6</sub>-tag with a twin Strep-tag through annealed oligonucleotide cloning. Plasmid pS6-proNEP was also used to generate plasmid pS6-NEP spanning the mature neprosin moiety (S<sup>129</sup>–Q<sup>380</sup>) plus the C-terminal His<sub>6</sub>-tag. The QuikChange Site-Directed mutagenesis kit (Stratagene) was used to generate

plasmids to produce point mutants A<sup>60R</sup>, K<sup>118A</sup>, Y<sup>136A</sup>, E<sup>188A</sup>, E<sup>188Q</sup>, Y<sup>214A</sup>, E<sup>297Q</sup>, and E<sup>297A</sup> from plasmid pS6-proNEP according to the manufacturer's instructions. PCR amplifications were performed using Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific), plasmids were purified with the GeneJET Plasmid MaxiPrep Kit (Thermo Fisher Scientific), and constructs were verified by DNA sequencing.

Proteins encoded by the pS6-proNEP, pS6-proNEP-Strep and pS6-NEP plasmids, as well as the eight point mutants, were targeted for overexpression in human Expi293-F cells (Thermo Fisher Scientific) adapted to FreeStyle F17 expression medium (Gibco). Cells were grown in a Multitron cell shaker incubator (Infors HT) at 150 rpm to a density of 3–5×10<sup>6</sup> cells/mL in a humidified atmosphere with 8% CO<sub>2</sub> at 37°C, and diluted every 3–4 d to 0.3–0.5×10<sup>6</sup> cells/mL. Thereafter, cells were sub-cultured to 0.7×10<sup>6</sup> cells/mL in medium containing penicillin/streptomycin, 0.5 µg/mL amphotericin, 8 mM L-glutamine, and 0.2% Pluronic F-68 non-ionic surfactant (Gibco). After 24 h, cells at 1.0–1.2×10<sup>6</sup> cells/mL were transfected with 1 mg plasmid DNA and 3 mg linear 25 kDa polyethyleneimine (Polysciences Europe) in 20 mL Opti-MEM medium (Gibco) previously incubated at room temperature for 15–20 min, per litre of expression medium. Cells were harvested after 3 d for protein purification.

Cell conditioned medium was cleared by centrifugation at 3500×g for 20 min at 4°C, and filtered through a 0.45 µm cellulose acetate filter (Millipore). The supernatant was then supplemented with 15 mM imidazole, incubated with nickel-nitrilotriacetic acid resin (Ni-NTA; Invitrogen) for 3–4 h at 4°C, loaded onto a PolyPrep open column (Bio-Rad) for batch affinity chromatography (AC) purification, and washed extensively with 20 mM Tris·HCl pH 7.5, 150 mM sodium chloride, 20 mM imidazole. Proteins were eluted with the same buffer containing 300 mM imidazole. For pro-neprosin-strep, the Ni-NTA step was replaced with AC using Strep-Tactin XT Superflow suspension resin (IBA Life Sciences), with 100 mM Tris·HCl pH 8.0, 150 mM sodium chloride, 50 mM D-biotin (VWR Life Science) as elution buffer. Subsequently, protein containing samples were pooled and the buffer was exchanged in a PD10 column (Sigma-Aldrich) to 20 mM Tris·HCl pH 7.5, 150 mM sodium chloride. Samples were concentrated and subjected to size-

exclusion chromatography (SEC) in a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with the same buffer. The column was attached to an ÄKTA Purifier liquid chromatography system (GE Healthcare) operated at room temperature.

Ultrafiltration was performed with Vivaspin 15 and Vivaspin 2 HY filter devices with a 5 kDa cut-off (Sartorius Stedim Biotech). Approximate protein concentrations were determined by measuring the absorbance at 280 nm ( $A_{280}$ ) with a NanoVue spectrophotometer (GE Healthcare), and applying the respective theoretical extinction coefficients ( $\epsilon_{280} = 92,625 \text{ M}^{-1}\text{cm}^{-1}$  and  $\epsilon_{280} = 82,530 \text{ M}^{-1}\text{cm}^{-1}$  for pro-neprosin and neprosin, respectively). Protein concentration was also measured using the BCA protein assay kit (Thermo Fisher Scientific) with bovine serum albumin (BSA) fraction V (Sigma-Aldrich) as standard. Protein purity was assessed by 13.5% tris-glycine SDS-PAGE stained with Coomassie Brilliant Blue R-250 (Thermo Fisher Scientific) using PageRuler Plus (10–250 kDa; Thermo Fisher Scientific) as molecular-mass marker. Protein identity was determined by peptide mass fingerprinting using a Bruker Auto Flex MALDI-TOF mass spectrometer and by N-terminal Edman sequencing in a Procise 494 protein sequencer (Applied Biosystems). The latter two assays were performed at the Protein Chemistry Service and the Proteomics Facility of the Centro de Investigaciones Biológicas (Madrid, Spain), respectively. Finally, mature WT neprosin was lyophilised in an Alpha 1-4 LD vacuum freeze dryer (Christ), stored at -20°C, and reconstituted for usage by dissolving the lyophilisate in Milli-Q water.

**Autolytic activation of pro-neprosin** — To obtain mature forms from pro-neprosin (WT or mutants) or pro-neprosin-strep by autolysis, protein samples after Ni-NTA or Strep-Tactin AC were subjected to 2 h dialysis against 20 mM Tris·HCl pH 7.5, 250 mM sodium chloride at room temperature, diluted twofold with 100 mM glycine pH 2.5, and incubated at 37°C for up to 16 h. Reactions were stopped at specific time points (0 min, 10 min, 20 min, 30 min, 1 h, 2 h, and overnight) by boiling aliquots in reducing/denaturing SDS sample buffer, and samples were analysed by 13.5% tris-glycine SDS-PAGE stained with Coomassie. Mature neprosin was subjected to buffer exchange in a PD10 column equilibrated with 20 mM Tris·HCl pH 7.5, 250 mM sodium chloride, and then to a final SEC

step in a Superdex 75 10/300 GL column with the same buffer. Protein purity and identity were assessed as aforementioned.

#### **Trans-activation of pro-neprosin mutants**

— To obtain mature neprosin point mutants from zymogens that would not autoactivate, purified pro-proteins (Y<sup>136</sup>A, E<sup>188</sup>A, E<sup>188</sup>Q, Y<sup>214</sup>A, E<sup>297</sup>Q, and E<sup>297</sup>A) were subjected to 2 h dialysis against 20 mM Tris·HCl pH 7.5, 250 mM sodium chloride at room temperature, diluted twofold with 100 mM glycine pH 2.5, and incubated with neprosin-strep at a 100:1 weight ratio overnight at 37°C. Cleaved samples were passed through a PD10 column equilibrated with 20 mM Tris·HCl pH 7.5, 1 M sodium chloride, and purified by reverse AC through incubation with Strep-Tactin XT Superflow suspension resin for 2 h at 4°C and subsequently loading onto a PolyPrep open column to remove neprosin-strep. The respective flow-through fractions containing the mature neprosin mutants were then subjected to SEC in a Superdex S75 10/300 GL column equilibrated with 20 mM Tris·HCl pH 7.5, 1 M sodium chloride and Ni-NTA AC as aforementioned for final polishing.

**Protein stability assays** — Pro-neprosin and mature neprosin were subjected to differential scanning fluorimetry in an iCycler iQ real time PCR detection system (BioRad). Samples at 0.5 mg/mL in 20 mM Tris·HCl pH 7.5, 150 mM sodium chloride (pro-neprosin) or 20 mM Tris·HCl pH 7.5, 250 mM sodium chloride (neprosin) were prepared as 25  $\mu$ L aliquots in the absence or presence of the tris(2-carboxyethyl)phosphine (TCEP; 5 mM and 10 mM) reducing agent, supplemented with 5 $\times$  SYPRO Orange protein stain (Thermo Fisher Scientific), and heated from 20°C to 95°C at 1°C/min. The temperature of midtransition ( $T_m$ ) was determined as the average of duplicate measurements of the midpoint value of the stability curve.

**Proteolytic activity and pH profile** — The fluorescent protein substrate DQ Red BSA (Thermo Fisher Scientific;  $\lambda_{EX}$  = 590 nm,  $\lambda_{EM}$  = 620 nm) at 10  $\mu$ M concentration was incubated with neprosin at 0.15  $\mu$ M in 100  $\mu$ L final volumes in triplicate experiments with buffers at pH 2–8 prepared from citric acid and dibasic sodium phosphate stock solutions. Fluorescence was monitored in an Infinite M2000 microplate fluorimeter (Tecan) overnight at 37°C. Bovine trypsin (Sigma-Aldrich) at 0.5  $\mu$ M and porcine pepsin (Fluka) at 0.5  $\mu$ M were also assayed for comparison.

#### **Cleavage studies with fluorogenic peptides and determination of kinetic parameters** —

Comparative peptidolytic activity of WT neprosin and mutants K<sup>118</sup>A, Y<sup>136</sup>A, E<sup>188</sup>A, E<sup>188</sup>Q, Y<sup>214</sup>A, and E<sup>297</sup>Q was assessed with the fluorogenic ‘QPQL-peptide’ (Mca-Q-P-Q-L-Dpa-A-R-NH<sub>2</sub>; GenScript) substrate. Reactions were carried out in 100 mM glycine pH 3.0 at 37°C in the aforementioned microplate fluorimeter at a peptidase:peptide weight ratio of 1:400, and the initial slope of the fluorescence ( $\lambda_{EX}$  = 325 nm,  $\lambda_{EM}$  = 393 nm) vs. time curve was taken as a measure of activity. Experiments were carried out in triplicate and differences were analysed for statistical significance with the *GraphPad* software (85).

In addition, the kinetic parameters of the cleavage of the QPQL-peptide, as well as of the ‘FS6-peptide’ (Mca-K-P-L-G-L-Dpa-A-R-NH<sub>2</sub>; Sigma-Aldrich), by WT neprosin were determined through reactions in the same buffer at 37°C with 0.1  $\mu$ M or 25 nM final enzyme concentration and substrate at 1–75  $\mu$ M (QPQL-peptide) or 2.5–75  $\mu$ M (FS6 peptide). The fluorescence signal, taken as a measure of cleavage product formation, was recorded over time for each substrate concentration and the respective values of the initial rate ( $v_0$ ) were derived from the slope of the linear part of the curve. Using a range of substrate concentrations and a surplus of peptidase, we measured the fluorescent signal generated after full substrate turnover and calculated the respective fluorescent units per picomole of cleaved substrate at the current instrument settings. These values were plotted against substrate concentration and data were fitted to the hyperbolic Michaelis-Menten equation ( $v = V_{max} \cdot [S] / \{K_M + [S]\}$ ) by nonlinear regression with the *GraphPad* and *SigmaPlot* (86) programs to determine the maximal velocity ( $V_{max}$ ), the Michaelis substrate affinity constant ( $K_M$ ), the turnover rate ( $k_{cat} = V_{max} / [E_{total}]$ ), and the catalytic efficiency ( $k_{cat} / K_M$ ) of the cleavage reaction. Experiments were carried out in triplicate.

**Cleavage studies of gliadin** — Wheat gliadin (Sigma-Aldrich) was prepared in 100 mM glycine pH 2.5 and sonicated to promote solubilisation. Variable concentrations of pepsin (0.05–10  $\mu$ M) from porcine gastric mucosa (Fluka), neprosin (0.05–2  $\mu$ M) or a mixture of pepsin at 0.5  $\mu$ M with neprosin (0.05–2  $\mu$ M) were used to digest 10 mg/mL slurries of gliadin in 200  $\mu$ L final reaction volumes. Reactions were monitored by

turbidometry in 96-well plates (Corning) at 37°C every 2 min for 90 min at  $\lambda = 595$  nm in a PowerWave XS microplate spectrophotometer (BioTek Instruments). Of note, plates were shaken between measurements for homogenisation of the reaction mixtures, which were quenched by boiling for 10 min and analysed by 13.5% tris-glycine SDS-PAGE.

Gliadin degradation by neprosin was further assayed by zymography using 13.5% SDS-PAGE zymograms containing either wheat gliadin (Sigma) or teleostean gelatin (Sigma), which was used as a control, at 0.1 mg/mL. Pro-neprosin was also assayed, which during the procedure became activated to the mature form. SDS-PAGE samples were prepared in the absence of  $\beta$ -mercaptoethanol, and electrophoresis was performed in an ice-filled Styrofoam box at 120 Volt. Protein renaturation was achieved by washing the zymograms twice for 30 min at room temperature with 2.5% Triton-X100 in 100 mM glycine pH 2.5, 200 mM sodium chloride. After two wash steps with the same buffer plus 0.02% Brij-35, zymograms were incubated overnight at 37 °C in the same buffer. After a quick rinse with water, zymograms were stained for 60 min with Coomassie, and excessive staining solution removed by a brief destaining step.

**Cleavage studies of the 33-mer peptide** — Neprosin cleavage of the 33-mer peptide of wheat  $\alpha$ -gliadin (purchased from GenScript), which spans sequence L-Q-L-Q-P-F-P-Q-P-Q-L-P-Y-P-Q-P-Q-L-P-Y-P-Q-P-Q-P-F (molecular mass = 3911 Da), was monitored by MALDI-TOF mass spectrometry in an AutoFLEX III mass spectrometer (Bruker Daltonik). The apparatus was equipped with a pulsed N<sub>2</sub>-laser ( $\lambda = 337$  nm) and controlled by the *flexControl* software package (Bruker Daltonik). The peptide was prepared in H<sub>2</sub>O at ~20 mg/mL and kept at -20°C until use. The cleavage reaction was carried out with substrate at ~1 mg/mL (~250 mM) in 100 mM glycine pH 3.0 at 37°C by adding neprosin (0.5  $\mu$ M final concentration) or pepsin (10  $\mu$ M). Reactions were stopped at different time points (0 min, 10 min, 20 min, 45 min, 1 h and overnight) by adding 1  $\mu$ L 2 M Tris·HCl pH 8.8 per 50  $\mu$ L of reaction mixture. Samples were subsequently diluted 1:10 with H<sub>2</sub>O, mixed at a 1:1 volume ratio with 2,5-dihydroxybenzoic acid matrix at 10 mg/mL in a solution containing 30% acetonitrile and 70% 0.1% trifluoroacetic acid, and spotted on a ground steel plate (Bruker Daltonik).

The sample was left to evaporate to dryness at room temperature, and mass spectra were acquired in positive reflectron mode at 21 kV total acceleration voltage focusing on a mass range of 200–4000 m/z. Spectra were collected by averaging 200 random laser shots and analysed using the *flexAnalysis* software (Bruker Daltonik).

**Liquid chromatography-mass spectrometry (LC-MS/MS) data analysis** — We reanalysed the cleavage specificity data of endogenous neprosin or recombinant material obtained from *Escherichia coli* as deposited at Chorus (Project ID 1262; (42)). LC-MS/MS raw files were converted to MGF-format using *ProteoWizard* (87), and data were processed using *TANDEM* (88), *Comet* (89), and *MS-GF+* ((90), as implemented in *SearchGUI* (91). Results were evaluated using *PeptideShaker* (92) at a false discovery rate of 1%. Data were nonspecifically searched for hits against the human proteome within UniProt (March 2020) using a mass tolerance of 20 ppm in both MS1 and MS2, fixed cysteine carbamidomethylation, and variable methionine oxidation. Up to 50 missed cleavages or a maximum of 5500 Da were tolerated for the parental peptide mass.

**Inhibition assays** — On the search for neprosin inhibitors, we assayed the cOmplete broad-spectrum inhibitor cocktail (Roche); the metallopeptidase inhibitors 1,10-phenanthroline, phosphoramidon, marimastat, and captopril (all from Sigma-Aldrich); the serine peptidase inhibitor 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF; Sigma-Aldrich); the aspartate peptidase inhibitors pepstatin A (Sigma-Aldrich), methyl-2-[(2-diazoacetyl)amino]hexanoate (DAN; Chemical Abstracts Service registry number [CAS] 7013-09-4; Bachem ref. 4010441), and 2-[(4-nitrophenoxy)methyl]oxirane (ENPN; CAS 5255-75-4; Apollo Scientific ref. OR26560); as well as the proline containing/mimicking compounds 2-acetyl-1-methylpyrrole (AMP; CAS 932-16-1; Sigma-Aldrich ref. 160865); (*S*)-tert-butyl-2-(3-ethoxy-3-oxopropanoyl)pyrrolidine-1-carboxylate (BEOPC; CAS 109180-95-2; Fluorochem ref. 387901); and *N*-boc-glycylproline (BGP; CAS 14296-92-5; Bachem ref. 4003703). Inhibition of the cleavage of the QPQL-peptide was assayed as aforementioned by pre-incubating neprosin at 100 nM in 100 mM glycine pH 3.0 with 100  $\mu$ M tester compound for >1 h at 37°C. The substrate was subsequently added at 10  $\mu$ M and the residual activity was monitored over 4 h as increase in fluorescence. Differences were analysed for

statistical significance with *GraphPad*. Of note, the positive control in the absence of inhibitors (100% activity) contained the same final concentration of dimethylsulfoxide as the one used for solubilisation of the inhibitors.

In addition, the half-maximal inhibitory concentration (IC<sub>50</sub>) was determined for pepstatin A and ENPN by measuring activity of 50 nM neprosin against 10 μM of substrate in the presence of inhibitor concentrations spanning 5–500 μM and 5–5000 μM, respectively, to obtain the associated inhibition curves. These curves were analysed by nonlinear regression analysis using *GraphPad*.

**Crystallization and diffraction data collection** — Crystallization conditions were screened for at the joint IBMB/IRB Automated Crystallography Platform by the sitting-drop vapor diffusion method. Reservoir solutions were prepared using a Tecan Freedom EVO robot and dispensed on 96×2-well MRC plates (Innovadyne Technologies). Crystallisation nanodrops containing 100 nL each of protein and reservoir solution were dispensed by a Phoenix/RE robot (Art Robbins). Crystallisation plates were incubated at 4°C or 20°C in Bruker steady-temperature crystal farms. Successful conditions were scaled up to the microliter range in 24-well Cryschem crystallization dishes (Hampton Research) whenever possible.

Best crystals of pro-neprosin at ~20 mg/mL in 20 mM Tris pH 7.5, 150 mM sodium chloride were obtained at 20°C with 0.1 M sodium acetate pH 4.0, 22% polyethylene glycol (PEG) 6000, 10% isopropanol as reservoir solution. Crystals were rapidly passed through cryo-buffer consisting of reservoir solution plus 15% (v/v) glycerol and flash-vitrified in liquid nitrogen for data collection. A lutetium derivative of pro-neprosin was obtained by soaking native crystals for 5 min in cryo-buffer supplemented with the ‘crystallophore’ compound Lu-Xo4 (purchased from Polyvalan) (57) at 100 mM final concentration and flash-vitrifying them without back soaking. X-ray diffraction data of native crystals were collected at 100 K on a Pilatus 6M-F pixel detector at beam line I04-1 of Diamond Light Source (Harwell, UK). Lutetium derivative data were recorded on a Pilatus 6M detector at beam line XALOC of the ALBA synchrotron (Cerdanyola, Catalonia, Spain).

Best crystals of neprosin at ~16 mg/mL in 20 mM Tris·HCl pH 7.5, 250 mM sodium chloride appeared at 4°C using 10% PEG 1000, 10% PEG 8000 as reservoir solution and were cryo-protected

with the same reservoir solution plus 15% (v/v) glycerol prior to flash-vitrification in liquid nitrogen. X-ray diffraction data were collected at 100 K at beam line ID30B of the ESRF synchrotron (Grenoble, France) using a Pilatus 6M detector.

Diffraction data were processed with programs *Xds* (93) and *Xscale*, and transformed with *Xdscov* to MTZ-format for the *Phenix* (94) and *CCP4* (95) suites of programs. All crystals contained a monomer in the crystal asymmetric unit and Table 1 provides essential statistics on data collection and processing.

**Structure solution and refinement** — The structure of pro-neprosin was solved by single-wavelength anomalous diffraction with data collected from a lutetium derivative crystal at the L<sub>III</sub>-absorption peak wavelength of 1.34 Å. The *f* and *f*' contributions were 14.1 e<sup>-</sup> and -16.8 e<sup>-</sup>, respectively, as determined by an X-ray absorption near edge scan. Diffraction data processed with separate Friedel mates served to identify two metal sites and solve the structure applying the *Autosol* protocol of the *Phenix* package (96). The resulting Fourier map was then subjected to further density modification with *wARP/ARP* (97), which produced an improved map for model building. A starting model for Lu-Xo4 was obtained by subjecting the coordinates of the metal-chelating moiety of the compound as found in its complex with Tb<sup>3+</sup> (Protein Data Bank [PDB] entry 6FRO, residue name 7MTI) to energy minimisation with the *Chimera* program (98). The resulting coordinates in PDB-format were employed together with a Lu<sup>3+</sup> ion for model building. Coordinates in MOL2-format, in turn, were used with *Grade* (<http://grade.globalphasing.org>) to generate restraints in CIF-format for model building and crystallographic refinement. Thereafter, several rounds of manual model building with the *Coot* program (99) alternated with crystallographic refinement with the *Refine* protocol of *Phenix* (100) and the *BUSTER* (101) program, which included translation/libration/screw-motion refinement, until the final model was obtained. The latter comprised pro-neprosin residues A<sup>29</sup>–A<sup>380</sup> except S<sup>76</sup>–Y<sup>85</sup> and N<sup>122</sup>–N<sup>131</sup> plus three extra C-terminal residues from the purification tag (A<sup>381</sup>–I<sup>382</sup>–A<sup>383</sup>); two Lu-Xo4 moieties at roughly half occupancy; two N-linked glycan chains totalling five sugar moieties attached to N<sup>145</sup> and N<sup>152</sup>, respectively; two acetate molecules; and 180 solvent molecules.

The structure of native pro-neprosin was solved by molecular replacement with the *Phaser* crystallographic software (102) using the protein coordinates of the lutetium-derivative crystal structure. Direct Fourier synthesis after rigid-body refinement failed because of anisomorphism between the two crystals, which mainly affected cell axis *c* (see Table 1). A unique solution was found at Eulerian angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) 193.7, 1.0, 166.5 and cell-fraction translation values (*x*, *y*, *z*) 0.003, 0.998, 0.418. This solution had a final translation function *Z*-score of 62.8 and a global log-likelihood gain after refinement of 11,705. Subsequent model building and refinement proceeded as aforementioned. The final model included residues A<sup>29</sup>–A<sup>380</sup> except Y<sup>77</sup>–N<sup>86</sup> and N<sup>122</sup>–N<sup>131</sup> plus two extra C-terminal residues from the purification tag (A<sup>381</sup>–I<sup>382</sup>), two *N*-linked glycan chains totalling four sugar moieties, as well as four acetate, one diethylene glycol, two glycerol and 259 solvent molecules.

Finally, the structure of mature neprosin was also solved by molecular replacement, using the coordinates of fragment T<sup>132</sup>–I<sup>382</sup> of native pro-neprosin. A unique solution was found at  $\alpha$ ,  $\beta$ ,  $\gamma$ , *x*, *y*, *z* values 37.8, 144.9, 33.8, 0.052, 0.470, 0.643, which had a final translation function *Z*-score of 30.5 and a global log-likelihood gain after refinement of 1756. Subsequent model building and refinement proceeded as aforementioned. The final model spanned residues T<sup>132</sup>–Q<sup>380</sup> plus the entire C-terminal tag (A<sup>381</sup>–I<sup>382</sup>–A<sup>383</sup>+H<sup>401</sup>–H<sup>406</sup>), two *N*-linked glycan chains totalling seven sugar moieties plus one triethylene glycol and 171 solvent molecules. Table 1 provides essential statistics on the final refined models, which were validated through the *wwPDB Validation Service* at <https://validate-rcsb-1.wwpdb.org/validservice> and deposited with the PDB at [www.pdb.org](http://www.pdb.org).

**Miscellaneous** — Structural superpositions and structure-based sequence alignments were calculated using the *SSM* program (103) within *Coot*. Figures were prepared with *Chimera* (104). Structure-similarity searches were performed with *Dali* (105). Protein interfaces were calculated with *PDBePISA* (66) at [www.ebi.ac.uk/pdbe/pisa](http://www.ebi.ac.uk/pdbe/pisa). The interacting surface of a complex was defined as half the sum of the buried surface areas of either molecule.

## REFERENCES

1. W. K. Dicke, J. H. van de Kamer, H. A. Weijers, Celiac disease. *Adv. Pediatr.* **9**, 277-318 (1957).
2. A. V. Balakireva, A. A. Zamyatnin, Properties of gluten intolerance: gluten structure, evolution, pathogenicity and detoxification capabilities. *Nutrients* **8** (2016).
3. K. Schalk, B. Lexhaller, P. Koehler, K. A. Scherf, Isolation and characterization of gluten protein types from wheat, rye, barley and oats for use as reference materials. *PLoS one* **12**, e0172819 (2017).
4. K. A. Scherf *et al.*, Recent progress and recommendations on celiac disease from the working group on prolamins analysis and toxicity. *Front. Nutr.* **7**, 29 (2020).
5. C. Monachesi *et al.*, Quantification of accidental gluten contamination in the diet of children with treated celiac disease. *Nutrients* **13**, 190 (2021).
6. P. Singh *et al.*, Global prevalence of celiac disease: systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **16**, 823-836 (e822) (2018).
7. J. A. King *et al.*, Incidence of celiac disease is increasing over time: a systematic review and meta-analysis. *Am. J. Gastroenterol.* **115**, 507-525 (2020).
8. S. W. Qiao *et al.*, Antigen presentation to celiac lesion-derived T cells of a 33-mer gliadin peptide naturally formed by gastrointestinal digestion. *J. Immunol.* **173**, 1757-1762 (2004).
9. D. A. van Heel, J. West, Recent advances in coeliac disease. *Gut* **55**, 1037-1046 (2006).
10. L. Shan *et al.*, Structural basis for gluten intolerance in celiac sprue. *Science* **297**, 2275-2279 (2002).
11. G. Janssen *et al.*, Ineffective degradation of immunogenic gluten epitopes by currently available digestive enzyme supplements. *PLoS one* **10**, e0128065 (2015).
12. L. Kivelä *et al.*, Current and emerging therapies for coeliac disease. *Nat. Rev. Gastroenterol. Hepatol.* **18**, in press (2021).
13. A. Díaz-Redondo, J. Miranda-Bautista, J. García-Lledó, J. P. Gisbert, L. Menchén, The potential usefulness of human leukocyte antigen typing for celiac disease screening: A systematic review and meta-analysis. *Rev. Esp. Enferm. Dig.* **107**, 423-429 (2015).

14. L. Maiuri *et al.*, IL-15 drives the specific migration of CD94<sup>+</sup> and TCR- $\gamma\delta$ <sup>+</sup> intraepithelial lymphocytes in organ cultures of treated celiac patients. *Am. J. Gastroenterol.* **96**, 150-156 (2001).
15. J. A. Tye-Din, H. J. Galipeau, D. Agardh, Celiac disease: a review of current concepts in pathogenesis, prevention, and novel therapies. *Front. Pediatr.* **6**, 350 (2018).
16. M. Tio, M. R. Cox, G. D. Eslick, Meta-analysis: coeliac disease and the risk of all-cause mortality, any malignancy and lymphoid malignancy. *Aliment. Pharmacol. Ther.* **35**, 540-551 (2012).
17. A. J. M. Daveson *et al.*, Baseline quantitative histology in therapeutics trials reveals villus atrophy in most patients with coeliac disease who appear well controlled on gluten-free diet. *GastroHep* **2**, 22-30 (2020).
18. M. El-Salhy, J. G. Hatlebakk, O. H. Gilja, T. Hausken, The relation between celiac disease, nonceliac gluten sensitivity and irritable bowel syndrome. *Nutr. J.* **14**, 92 (2015).
19. M. Siegel *et al.*, Rational design of combination enzyme therapy for celiac sprue. *Chem. Biol.* **13**, 649-658 (2006).
20. H. Wieser, P. Koehler, Detoxification of gluten by means of enzymatic treatment. *J. AOAC Int.* **95**, 356-363 (2012).
21. L. V. Savvateeva *et al.*, Glutenase and collagenase activities of wheat cysteine protease Triticain-alpha: feasibility for enzymatic therapy assays. *Int. J. Biochem. Cell Biol.* **62**, 115-124 (2015).
22. F. J. Suchy *et al.*, National Institutes of Health Consensus Development Conference: lactose intolerance and health. *Ann. Intern. Med.* **152**, 792-796 (2010).
23. M. T. Bethune, C. Khosla, Oral enzyme therapy for celiac sprue. *Methods Enzymol.* **502**, 241-271 (2012).
24. C. Wolf *et al.*, Engineering of Kuma030: a gliadin peptidase that rapidly degrades immunogenic gliadin peptides in gastric conditions. *J. Am. Chem. Soc.* **137**, 13106-13113 (2015).
25. N. E. Castillo, T. G. Theethira, D. A. Leffler, The present and the future in the diagnosis and management of celiac disease. *Gastroenterol. Rep. (Oxford)* **3**, 3-11 (2015).
26. C. Hoppe *et al.*, Intake and sources of gluten in 20- to 75-year-old Danish adults: a national dietary survey. *Eur. J. Nutr.* **56**, 107-117 (2017).
27. Institute-of-Medicine, *Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids*. (The National Academies Press, Washington, D.C., 2005), 10.17226/10490, pp. 1358.
28. I. A. Goptar *et al.*, Cysteine digestive peptidases function as post-glutamine cleaving enzymes in tenebrionid stored-product pests. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **161**, 148-154 (2012).
29. L. Cavaletti *et al.*, E40, a novel microbial protease efficiently detoxifying gluten proteins, for the dietary management of gluten intolerance. *Sci. Rep.* **9**, 13147 (2019).
30. D. Stepniak *et al.*, Highly efficient gluten degradation with a newly identified prolyl endoprotease: implications for celiac disease. *Am. J. Physiol. Gastrointest. Liver Physiol* **291**, G621-G629 (2006).
31. C. Mitea *et al.*, Efficient degradation of gluten by a prolyl endoprotease in a gastrointestinal model: implications for coeliac disease. *Gut* **57**, 25-32 (2008).
32. G. J. Tack *et al.*, Consumption of gluten with gluten-degrading enzyme by celiac patients: a pilot-study. *World J. Gastroenterol.* **19**, 5837-5847 (2013).
33. B. N. Salden *et al.*, Randomised clinical study: *Aspergillus niger*-derived enzyme digests gluten in the stomach of healthy volunteers. *Aliment. Pharmacol. Ther.* **42**, 273-285 (2015).
34. J. König, S. Holster, M. J. Bruins, R. J. Brummer, Randomized clinical trial: effective gluten degradation by *Aspergillus niger*-derived enzyme in a complex meal setting. *Sci. Rep.* **7**, 13100 (2017).
35. J. Ehren *et al.*, A food-grade enzyme preparation with modest gluten detoxification properties. *PLoS one* **4**, e6313 (2009).
36. J. Gass, M. T. Bethune, M. Siegel, A. Spencer, C. Khosla, Combination enzyme therapy for gastric digestion of dietary gluten in patients with celiac sprue. *Gastroenterology* **133**, 472-480 (2007).
37. S. R. Gordon *et al.*, Computational design of an  $\alpha$ -gliadin peptidase. *J. Am. Chem. Soc.* **134**, 20513-20520 (2012).
38. I. S. Pultz *et al.*, AGA Abstracts 1125. Kuma062 effectively digests gluten in the human stomach: results of a phase 1 study. *Gastroenterology* **158**, S-218 (2020).
39. S. Krishnareddy, K. Stier, M. Recanati, B. Leibold, P. H. Green, Commercially available

glutenases: a potential hazard in coeliac disease. *Therap. Adv. Gastroenterol.* **10**, 473-481 (2017).

40. L. Lee, Y. Zhang, B. Ozar, C. W. Sensen, D. C. Schriemer, Carnivorous nutrition in pitcher plants (*Nepenthes* spp.) via an unusual complement of endogenous enzymes. *J. Proteome Res.* **15**, 3108-3117 (2016).

41. M. Rey *et al.*, Addressing proteolytic efficiency in enzymatic degradation therapy for celiac disease. *Sci. Rep.* **6**, 30980 (2016).

42. C. U. Schröder *et al.*, Neprosin, a selective prolyl endoprotease for bottom-up proteomics and histone mapping. *Mol. Cell. Proteomics* **16**, 1162-1171 (2017).

43. C. U. Schröder, D. S. Ziemianowicz, K. Merx, D. C. Schriemer, Simultaneous proteoform analysis of histones H3 and H4 with a simplified middle-down proteomics method. *Anal. Chem.* **90**, 3083-3090 (2018).

44. N. D. Rawlings, A. Bateman, How to use the MEROPS database and website to help understand peptidase specificity. *Protein Sci.* **30**, 83-92 (2021).

45. U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, P. Nordlund, Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289-298 (2006).

46. C. Vieille, G. J. Zeikus, Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1-43 (2001).

47. J. Eder, A. R. Fersht, Pro-sequence-assisted protein folding. *Mol. Microbiol.* **16**, 609-614 (1995).

48. C. Lazure, The peptidase zymogen proregions: nature's way of preventing undesired activation and proteolysis. *Curr. Pharm. Des.* **8**, 511-531 (2002).

49. J. P. Raufman, "Pepsin." in Encyclopedia of Gastroenterology., L. R. Johnson, Ed. (Academic Press - Elsevier, London (UK), 2004), 10.1016/B0-12-386860-2/00561-X, pp. 147-148.

50. I. Schechter, A. Berger, On the size of active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162 (1967).

51. F. X. Gomis-Rüth, T. O. Botelho, W. Bode, A standard orientation for metallopeptidases. *Biochim. Biophys. Acta* **1824**, 157-163 (2012).

52. N. B. Roberts, R. Sheers, W. H. Taylor, Secretion of total pepsin and pepsin 1 in healthy

volunteers in response to pentagastrin and to insulin-induced hypoglycaemia. *Scand. J. Gastroenterol.* **42**, 555-561 (2007).

53. U. Neumann, H. Kubota, K. Frei, V. Ganu, D. Leppert, Characterization of Mca-Lys-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH<sub>2</sub>, a fluorogenic substrate with increased specificity constants for collagenases and tumor necrosis factor converting enzyme. *Anal. Biochem.* **328**, 166-173 (2004).

54. E. M. Christensen *et al.*, In crystallo screening for proline analog inhibitors of the proline cycle enzyme PYCR1. *J. Biol. Chem.* **295**, 18316-18327 (2020).

55. H. Umezawa, T. Aoyagi, H. Morishima, M. Matsuzaki, M. Hamada, Pepstatin, a new pepsin inhibitor produced by *Actinomycetes*. *J. Antibiot. (Tokyo)* **23**, 259-262 (1970).

56. A. D. Abell, D. A. Hoult, D. A. Bergman, D. P. Fairlie, Simple *cis*-epoxide-based inhibitors of HIV-1 protease. *Bioorg. Med. Chem. Lett.* **7**, 2853-2856 (1997).

57. S. Engilberge *et al.*, Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chem. Sci.* **8**, 5909-5917 (2017).

58. E. W. Stawiski, A. E. Baucom, S. C. Lohr, L. M. Gregoret, Predicting protein function from structure: Unique structural features of proteases. *Proc. Natl. Acad. Sci. USA* **97**, 3954-3958 (2000).

59. J. S. Richardson, The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34**, 167-339 (1981).

60. A. R. Khan, M. N. James, Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Prot. Sci.* **7**, 815-836 (1998).

61. A. R. Khan, N. Khazanovich-Bernstein, E. M. Bergmann, M. N. G. James, Structural aspects of activation pathways of aspartic protease zymogens and viral 3C protease precursors. *Proc. Natl. Acad. Sci. USA* **96**, 10968-10975 (1999).

62. J. L. Arolas, T. Goulas, A. Cuppari, F. X. Gomis-Rüth, Multiple architectures and mechanisms of latency in metallopeptidase zymogens. *Chem. Rev.* **118**, 5581-5597 (2018).

63. M. Fujinaga, M. M. Chernaia, N. I. Tarasova, S. C. Mosimann, M. N. G. James, Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* **4**, 960-972 (1995).



64. A. Wlodawer, A. Gutschina, M. N. G. James, "Chapter 2 – Catalytic pathways of aspartic peptidases." in Handbook of Proteolytic Enzymes, N. D. Rawlings, G. S. Salvesen, Eds. (Academic Press, Oxford, 2013), vol. 1, pp. 19-26.
65. J. D. Robertus, J. Kraut, R. A. Alden, J. J. Birktoft, Subtilisin; a stereochemical mechanism involving transition-state stabilization. *Biochemistry* **11**, 4293-4303 (1972).
66. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797 (2007).
67. J. Janin, R. P. Bahadur, P. Chakrabarti, Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133-180 (2008).
68. R. M. Herriott, Q. R. Bartz, J. H. Northrop, Transformation of swine pepsinogen into swine pepsin by chicken pepsin. *J. Gen. Physiol.* **21**, 575-582 (1938).
69. L. Boon, E. Ugarte-Berzal, J. Vandooren, G. Opdenakker, Protease propeptide structures, mechanisms of activation, and functions. *Crit. Rev. Biochem. Mol. Biol.* **55**, 111-165 (2020).
70. H. Neurath, Proteolytic enzymes, past and future. *Proc. Natl. Acad. Sci. USA* **96**, 10962-10963 (1999).
71. M. Fujinaga, M. M. Cherney, H. Oyama, K. Oda, M. N. G. James, The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proc. Natl. Acad. Sci. USA* **101**, 3364-3369 (2004).
72. B. Pillai *et al.*, Crystal structure of scytalidoglutamic peptidase with its first potent inhibitor provides insights into substrate specificity and catalysis. *J Mol Biol* **365**, 343-361 (2007).
73. M. Y. Kondo *et al.*, Studies on the catalytic mechanism of a glutamic peptidase. *J. Biol. Chem.* **285**, 21437-21445 (2010).
74. H. Sasaki *et al.*, The three-dimensional structure of aspergilloglutamic peptidase from *Aspergillus niger*. *Proc. Jpn. Acad. Ser. B - Phys. Biol. Sci.* **80**, 435-438 (2004).
75. H. Sasaki *et al.*, The crystal structure of an intermediate dimer of aspergilloglutamic peptidase that mimics the enzyme-activation product complex produced upon autoproteolysis. *J. Biochem.* **152**, 45-52 (2012).
76. P. Jara *et al.*, Cloning and characterization of the *eapB* and *eapC* genes of *Cryphonectria parasitica* encoding two new acid proteinases, and disruption of *eapC*. *Mol. Gen. Genet.* **250**, 97-105 (1996).
77. N. Poussereau, S. Creton, G. Billon-Grand, C. Rasclé, M. Fevre, Regulation of *acp1*, encoding a non-aspartyl acid protease expressed during pathogenesis of *Sclerotinia sclerotiorum*. *Microbiology* **147**, 717-726 (2001).
78. J. L. Moon, L. N. Shaw, J. A. Mayo, J. Potempa, J. Travis, Isolation and properties of extracellular proteinases of *Penicillium marneffei*. *Biol. Chem.* **387**, 985-993 (2006).
79. A. J. O'Donoghue *et al.*, Inhibition of a secreted glutamic peptidase prevents growth of the fungus *Talaromyces emersonii*. *J. Biol. Chem.* **283**, 29186-29195 (2008).
80. S. Rolland *et al.*, pH controls both transcription and post-translational processing of the protease BcACP1 in the phytopathogenic fungus *Botrytis cinerea*. *Microbiology* **155**, 2097-2105 (2009).
81. K. Jensen, P. R. Ostergaard, R. Wilting, S. F. Lassen, Identification and characterization of a bacterial glutamic peptidase. *BMC Biochem* **11**, 47 (2010).
82. N. Oda *et al.*, Nucleotide sequence of the gene encoding the precursor protein of pepstatin insensitive acid protease B, scytalidopepsin B, from *Scytalidium lignicolum*. *Biosci. Biotechnol. Biochem.* **62**, 1637-1639 (1998).
83. N. Stocchi, M. V. Revuelta, P. A. L. Castronuovo, D. M. A. Vera, A. ten Have, Molecular dynamics and structure function analysis show that substrate binding and specificity are major forces in the functional diversification of Eqolisins. *BMC bioinformatics* **19**, 338 (2018).
84. Y. Kataoka *et al.*, Catalytic residues and substrate specificity of scytalidoglutamic peptidase, the first member of the eqolisin family (G1) of peptidases. *FEBS Lett.* **579**, 2991-2994 (2005).
85. M. L. Swift, GraphPad Prism, data analysis, and scientific graphing. *J. Chem. Inf. Comput. Sci.* **37**, 411-412 (1997).
86. D. Kornbrot, Statistical software for microcomputers: SigmaPlot 2000 and SigmaStat2. *Br. J. Math. Stat. Psychol.* **53 (Pt 2)**, 335-337 (2000).
87. M. C. Chambers *et al.*, A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918-920 (2012).
88. R. Craig, R. C. Beavis, TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466-1467 (2004).

89. J. K. Eng, T. A. Jahan, M. R. Hoopmann, Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22-24 (2013).
90. S. Kim, P. A. Pevzner, MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277 (2014).
91. H. Barsnes, M. Vaudel, SearchGUI: a highly adaptable common interface for proteomics search and *de novo* engines. *J. Proteome Res.* **17**, 2552-2555 (2018).
92. M. Vaudel *et al.*, PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* **33**, 22-24 (2015).
93. W. Kabsch, XDS. *Acta Crystallogr. sect. D* **66**, 125-132 (2010).
94. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. sect. D* **66**, 213-221 (2010).
95. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. sect. D* **67**, 235-242 (2011).
96. T. C. Terwilliger *et al.*, Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. sect. D* **65**, 582-601 (2009).
97. G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171-1179 (2008).
98. E. F. Pettersen *et al.*, UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
99. A. Casañal, B. Lohkamp, P. Emsley, Current developments in *Coot* for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069-1078 (2020).
100. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. sect. D* **75**, 861-877 (2019).
101. O. S. Smart *et al.*, Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. sect. D* **68**, 368-380 (2012).
102. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
103. E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. sect. D* **60**, 2256-2268 (2004).
104. C. C. Huang, E. C. Meng, J. H. Morris, E. F. Pettersen, T. E. Ferrin, Enhancing UCSF Chimera through web services. *Nucl. Acids Res.* **42**, W478-W484 (2014).
105. L. Holm, L. M. Laakso, Dali server update. *Nucleic Acids Res.* **44**, W351-W355 (2016).
106. R. García-Castellanos *et al.*, Three-dimensional structure of MecI : Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J. Biol. Chem.* **278**, 39897-39905 (2003).
107. M. S. Weiss, Global indicators of X-ray quality. *J. Appl. Cryst.* **34**, 130-135 (2001).
108. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).
109. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D* **66**, 12-21 (2010).



**Table 1. Crystallographic data.**

<b>Dataset</b>	<b>Pro-neprosin Lu-Xo4 complex</b>	<b>Pro-neprosin Native</b>	<b>Neprosin Native</b>
Beam line (synchrotron)	XALOC (ALBA)	I04-1 (DIAMOND)	ID30B (ESRF)
Space group / protomers per a.u. <sup>a</sup>	P2 <sub>1</sub> 2 <sub>1</sub> 2 / 1	P2 <sub>1</sub> 2 <sub>1</sub> 2 / 1	P2 <sub>1</sub> / 1
Cell constants (a, b, c, in Å; β in °)	86.65, 93.35, 48.70, 90.0	86.37, 92.62, 42.69, 90.0	54.76, 49.95, 59.86, 106.57
Wavelength (Å)	1.3400	1.0050	1.0050
Measurements / unique reflections	316,335 / 25,470 <sup>h</sup>	407,612 / 32,321	84,480 / 13,077
Resolution range (Å) (outermost shell) <sup>b</sup>	63.5 – 2.05 (2.17 – 2.05)	63.2 – 1.80 (1.91 – 1.80)	52.5 – 2.35 (2.49 – 2.35)
Completeness (%) / R <sub>merge</sub> <sup>c</sup>	99.9 (99.4) / 0.182 (1.493)	99.3 (96.0) / 0.110 (1.552)	99.6 (99.2) / 0.235 (1.167) <sup>i</sup>
R <sub>meas</sub> <sup>d</sup> / CC(1/2) <sup>d</sup>	0.190 (1.564) / 0.997 (0.918)	0.115 (1.634) / 0.999 (0.710)	0.256 (1.270) <sup>i</sup> / 0.990 (0.653)
Average intensity <sup>e</sup>	9.2 (2.4)	15.3 (1.5)	5.6 (1.3)
B-Factor (Wilson) (Å <sup>2</sup> ) / Aver. multiplicity	39.8 / 12.5 (11.4)	34.4 / 12.6 (10.3)	40.3 / 6.5 (6.4)
Heavy-ion sites for phasing	2	–	–
Resolution range used for refinement (Å)	63.5 – 2.05	63.2 – 1.80	52.5 – 2.35
Reflections used (test set)	24,849 (621)	31,685 (636)	12,598 (479)
Crystallographic R <sub>factor</sub> (free R <sub>factor</sub> ) <sup>c</sup>	0.197 (0.248)	0.181 (0.228)	0.195 (0.246)
Non-H protein atoms / ionic ligands / waters / non-ionic ligands per a.u.	2604 / 2 ACT, 2 Lu-Xo4 180/1 FUC, 4 NAG	2597 / 4 ACT 259/1 FUC, 2 GOL,3 NAG,1 PEG	2011 / – / 171/1 BMA,3 MAN,3 NAG, 2PGE
R <sub>msd</sub> from target values			
bonds (Å) / angles (°)	0.008 / 0.98	0.008 / 0.96	0.008 / 1.02
Average B-factor (Å <sup>2</sup> )	48.2	35.2	39.6
All-atom contacts and geometry analysis			
Ramachandran <sup>f</sup> favoured / outliers / all analysed	321 (97.3%) / 1 (0.3%) / 330	325 (96.7%) / 2 (0.6%) / 336	246 (96.0%) / 1 (0.4%) / 256
Bond-length / bond-angle / chirality / planarity outliers <sup>f</sup>	0 / 0 / 0 / 1	0 / 0 / 0 / 1	0 / 0 / 0 / 0
Side-chain outliers <sup>f</sup> / MolProbity score	9 (3.3%) / 1.73	4 (1.4%)	7 (3.4%)
All-atom clashes / all atoms <sup>g</sup> / clashscore	16 / 5428 / 2.9	10 / 5565 / 1.8	9 / 4192 / 2.1
RSRZ outliers <sup>f</sup> / F <sub>o</sub> :F <sub>c</sub> correlation (for test set)	25 (7.5%) / 0.91 (0.88)	16 (4.8%) / 0.95 (0.92)	1 (0.4%) / 0.93 (0.90)

<sup>a</sup> Abbreviations: ACT, acetate; a.u., crystallographic asymmetric unit; BMA, β-D-mannose; FUC, α-L-fucose; MAN, α-D-mannose; NAG, β-N-acetyl-D-glucosamine; PEG, diethylene glycol; PGE, triethylene glycol. <sup>b</sup> Values in parenthesis refer to the outermost resolution shell. <sup>c</sup> For definitions, see Table 1 in (106). <sup>d</sup> For definitions, see (107, 108). <sup>e</sup> Average intensity is  $\langle I/\sigma(I) \rangle$  of unique reflections after merging according to the XDS program (93). <sup>f</sup> For the protein part only. <sup>g</sup> Including hydrogen atoms generated by MolProbity (109). <sup>h</sup> Friedel mates were kept separately for structure solution but merged for refinement. RSRZ, real-space R-value Z-score. <sup>i</sup> The overall high R<sub>meas</sub> and R<sub>merge</sub> values result from too short exposure times, which led to weak measurements.

**Supplemental Table 1. Constructs and primers for neprosin expression in mammalian cells.**

Construct name	Parental plasmid	Forward primer	Reverse primer	Protein sequence
pS6-proNEP-Strep	pS6-proNEP	<u>CGCTGGAGGTTGGAGCCATCCACAATTCGA</u> AAAGGGTGGAGGTTCTGGAGGTGGAAGTG GAGGTTGGAGCCATCCACAATTCGAAAAGT AGTAGC	<u>TCGAGCTACTACTTTTCGAATTGTGGATG</u> GCTCCAACCTCCACTCCACCTCCAGAAC CTCCACCCTTTTCGAATTGTGGATGGCTC CAACCTCCAGCGAT	DLMV+R <sup>25</sup> -Q <sup>380</sup> +AIA- GGW <sup>SHPQFEK</sup> GGGSGGGSGGW <sup>SHPQFEK</sup>
pS6-proNEP	pET-28a(+)- proNEP *	ATGCGGTGACCTAATGGTACGTAGCATTCA	GCATGCGATCGCTTGGAACCCGGACCA	DLMV+R <sup>25</sup> -Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-NEP	pS6-proNEP	ATGCGGTGACCTAAGCGCGAACACCAACCA	GCATGCGATCGCTTGGAACCCGGACCAC	DLMV+S <sup>129</sup> -Q <sup>380</sup> +AIAH <sub>6</sub>
<b>Point mutants</b>				
pS6-proNEP-A <sup>60</sup> R	pS6-proNEP	TCTACAAGCAGCCGAGGTTTCGATCACCC	GGGTGATCGAACCTCGGCTGCTTGTAGA	DLMV+R <sup>25</sup> -A <sup>60</sup> R-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-K <sup>118</sup> A	pS6-proNEP	CGGTGGTAAAGGCACAATTCCCGAACCT	AGGTTGCGGAATTGTGCTTAACCACCG	DLMV+R <sup>25</sup> -K <sup>118</sup> R-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-Y <sup>136</sup> A	pS6-proNEP	ACCAACCACCAGGCTGCGGTTATTGCGT	ACGCAATAACCGCAGCTGGTGGTTGGT	DLMV+R <sup>25</sup> -Y <sup>136</sup> R-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-E <sup>188</sup> A	pS6-proNEP	CCTGAACACCATTGCA GCGGGTTGGCAG	CTGCCAACCCGCTGCAATGGTGTTCAGG	DLMV+R <sup>25</sup> -E <sup>188</sup> A-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-E <sup>188</sup> Q	pS6-proNEP	CCTGAACACCATTCA GCGGGTTGGCAG	CTGCCAACCCGCTTGAATGGTGTTCAGG	DLMV+R <sup>25</sup> -E <sup>188</sup> Q-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-Y <sup>214</sup> A	pS6-proNEP	GACCGCGGACGGTGCTACCAGCACCG	CGGTGCTGGTAGCACCGTCCGCGGTC	DLMV+R <sup>25</sup> -Y <sup>214</sup> A-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-E <sup>297</sup> A	pS6-proNEP	GAGTGGGGTGGCGCAATCTACGATAGCA	TGCTATCGTAGATGCGCCACCCCACTC	DLMV+R <sup>25</sup> -E <sup>297</sup> A-Q <sup>380</sup> +AIA-H <sub>6</sub>
pS6-proNEP-E <sup>297</sup> Q	pS6-proNEP	GAGTGGGGTGGCGCAATCTACGATAGCA	TGCTATCGTAGATTTGCCACCCCACTC	DLMV+R <sup>25</sup> -E <sup>297</sup> Q-Q <sup>380</sup> +AIA-H <sub>6</sub>

\* A cDNA encoding pro-neprosin inserted into *Escherichia coli* plasmid pET-28a(+) was purchased from GeneScript. Restriction-site sequences are underlined, amino acid point mutations are in magenta, additional residues derived from the cloning strategy are in red, C-terminal tags are in blue.

## RESULTS: PROJECT 3

---

*“Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor  $\beta$ 2 in a new conformation”*

*“Recombinant production of human  $\alpha$ 2-macroglobulin variants and interaction studies with recombinant G-related  $\alpha$ 2-macroglobulin binding protein and latent transforming growth factor- $\beta$ 2”*




## Summary

Transforming growth factor  $\beta 2$  is a disulphide-linked dimeric cytokine that is engaged in signalling functions through binding of its cognate TGF $\beta$  receptor. To regulate this pathway, human TGF $\beta 2$  is biosynthesized as a latent 394-residue precursor (pro-TGF $\beta 2$ ) with an N-terminal latency-associated domain preceding the mature moiety. Given the importance of this cytokine and its potential in therapeutic applications, we developed and established a new high-yield human expression system for tagged pro-TGF $\beta 2$ . We further solved the X-ray crystal structure of its GF domain, which was obtained in a new crystallographic space group and in a homodimeric assembly that slightly deviates from published structures. Furthermore, biochemical, biophysical, and binding assays including surface plasmon resonance, multiple-angle laser light scattering, size-exclusion chromatography, fluorogenic labelling, gel electrophoresis and western-blot analysis were carried out to analyse the interplay of pro-TGF $\beta 2$  and human  $\alpha_2$ -macroglobulin (h $\alpha_2$ M) variants. Importantly, the latter has a great impact on human physiology and pathophysiology by participating in the inhibition of a broad spectrum of endoproteases, but also by modulating the activity of various cytokines, growth factors, hormones and other proteins.





# SCIENTIFIC REPORTS



OPEN

## Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor $\beta$ 2 in a new conformation

Laura del Amo-Maestro, Laura Marino-Puertas, Theodoros Goulas & F. Xavier Gomis-Rüth 

Transforming growth factor  $\beta$  is a disulfide-linked dimeric cytokine that occurs in three highly related isoforms (TGF $\beta$ 1–TGF $\beta$ 3) engaged in signaling functions through binding of cognate TGF $\beta$  receptors. To regulate this pathway, the cytokines are biosynthesized as inactive pro-TGF $\beta$ s with an N-terminal latency-associated protein preceding the mature moieties. Due to their pleiotropic implications in physiology and pathology, TGF $\beta$ s are privileged objects of *in vitro* studies. However, such studies have long been limited by the lack of efficient human recombinant expression systems of native, glycosylated, and homogenous proteins. Here, we developed pro-TGF $\beta$ 2 production systems based on human Expi293F cells, which yielded >2 mg of pure histidine- or Strep-tagged protein per liter of cell culture. We assayed this material biophysically and in crystallization assays and obtained a different crystal form of mature TGF $\beta$ 2, which adopted a conformation deviating from previous structures, with a distinct dimeric conformation that would require significant rearrangement for binding of TGF $\beta$  receptors. This new conformation may be reversibly adopted by a certain fraction of the mature TGF $\beta$ 2 population and represent a hitherto undescribed additional level of activity regulation of the mature growth factor once the latency-associated protein has been separated.

Transforming growth factors  $\beta$  (TGF $\beta$ s) are members of a superfamily of conserved cytokines, which also includes bone morphogenetic proteins, activins, and inhibins<sup>1–3</sup>. TGF $\beta$ s were discovered in the early 1980s<sup>4</sup>, and three close orthologs are found in mammals (TGF $\beta$ 1, TGF $\beta$ 2, and TGF $\beta$ 3<sup>5</sup>). They are ubiquitously expressed and secreted throughout the body<sup>2,6</sup>, where they are pleiotropically engaged in the physiology of nearly all tissues and cell types after activation and binding to TGF $\beta$  receptors of type I, II and III (TGFR-I, -II and -III)<sup>3,7–10</sup>. TGF $\beta$ s mainly act through the SMAD pathway, which features a family of genes similar to *Caenorhabditis* “small worm phenotype” and *Drosophila* “mothers against decapentaplegic”. This pathway entails sequential binding of TGFR-II and TGFR-I by TGF $\beta$ s to form a ternary complex that triggers phosphorylation of the latter by the former<sup>11,12</sup>. Thus, TGF $\beta$ s regulate intracellular processes including transcription, translation, microRNA biogenesis, protein synthesis, and post-translational modifications<sup>1,13</sup>, which cascade into roles in growth, proliferation, differentiation, plasticity, migration, and death of epithelial and endothelial cells, as well as lymphocytes<sup>14</sup>. TGF $\beta$ s are essential in biological events such as embryogenesis, wound healing, and immunity<sup>2</sup>, but also in pathologies such as Marfan syndrome, Parkinson’s disease, AIDS, organ fibrosis, autoimmune diseases, atherosclerosis and hypertension, asthma, diabetes, rheumatoid arthritis, encephalomyelitis, colitis, and most epithelial cancers, including prostate, breast, lung, colorectal, pancreatic, gastric, and skin cancers<sup>2,13,15–20</sup>.

Human TGF $\beta$ s are produced as latent glycosylated precursors (pro-TGF $\beta$ s), which consist of an ~250-residue N-terminal pro-domain dubbed latency-associated protein (LAP) and a C-terminal ~110-residue mature growth-factor moiety (GF). Two such precursors are joined through disulfides in the homodimeric small latent

Proteolysis Lab; Structural Biology Unit; “María-de-Maeztu” Unit of Excellence, Molecular Biology Institute of Barcelona (CSIC); Barcelona Science Park, c/Baldiri Reixac, 15-21, 08028, Barcelona, Catalonia, Spain. Laura del Amo-Maestro and Laura Marino-Puertas contributed equally. Correspondence and requests for materials should be addressed to T.G. (email: [thgcri@ibmb.csic.es](mailto:thgcri@ibmb.csic.es)) or F.X.G.-R. (email: [xgcri@ibmb.csic.es](mailto:xgcri@ibmb.csic.es))

complex. These complexes are disulfide-linked to a third protein, one of three latent TGF $\beta$  binding proteins, in the large latent complexes<sup>21–23</sup>. After secretion, these complexes are targeted to fibrillin-rich microfibrils as inactive species that are covalently bound by tissue transglutaminase to the extracellular matrix for storage. Finally, the large latent complexes are transformed into biologically active GFs by thrombospondin 1, reactive oxygen species, integrins, and/or peptidases such as furin and other related pro-protein convertases. Proteolytic cleavage by these enzymes severs the linker between GF and LAP, but the two proteins remain noncovalently associated until physically separated for function<sup>2,10,13,24,25</sup>. In this way, most TGF $\beta$  in the body is latent and sequestered within the extracellular matrix, although it is also found on the surface of immune cells or in granules of platelets and mast cells<sup>2</sup>. Thus, localization and compartmentalization ensure tight spatial and temporal regulation of the GFs<sup>22</sup>.

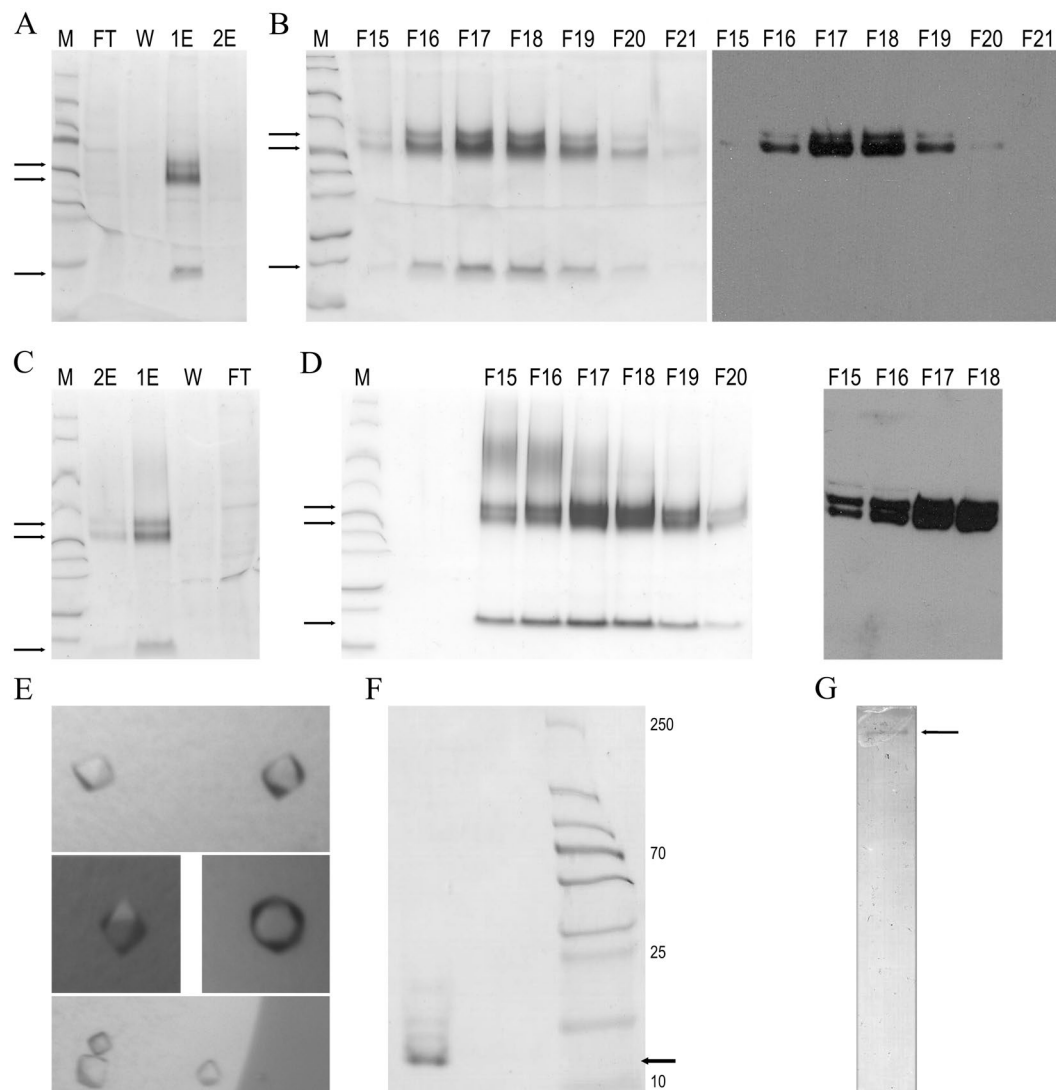
Although TGF $\beta$  GFs are very similar (70–82% sequence identity in humans) and have partially overlapping functions, they also have distinct roles. This is reflected in their differential expression during embryogenesis and specific roles in renal fibrogenesis and regulation of airway inflammation and remodeling<sup>26</sup>. In particular, human TGF $\beta$ 2 is biosynthesized as a latent 394-residue precursor (pro-TGF $\beta$ 2) arranged as a glycosylated homodimer of 91 kDa containing three glycan chains attached to each LAP. Peptidolytic activation at bond R<sup>302</sup>-A<sup>303</sup> (see UniProt [UP] entry P61812 for residue numbers of human TGF $\beta$ 2 in superscripts<sup>27,28</sup>) produces the GF, which spans 112 residues and is arranged as a disulfide-linked 25-kDa homodimer noncovalently associated with the respective LAP moieties. Given the importance of this cytokine and its potential in therapeutic applications<sup>29,30</sup>, here we developed a new high-yield human expression system for tagged pro-TGF $\beta$ 2. We report the X-ray crystal structure of its GF, which was obtained in a different crystallographic space group and deviates from current structures.

## Results and Discussion

**A new recombinant overexpression system for pro-TGF $\beta$ 2.** After its discovery, TGF $\beta$ 2 GF was initially purified from human and porcine platelets<sup>31,32</sup> and from bovine bone<sup>33</sup>. However, to obtain large amounts, recombinant expression systems are generally the option of choice as the resulting homogeneity and purity is greater than that of proteins purified from tissues or fluids. Mature human and mouse TGF $\beta$ 2 GFs were obtained in high yields (~8–10 mg/liter of cell culture) from *Escherichia coli* systems<sup>34</sup>. However, these proteins lacked glycosylated LAP, which has (glycan-mediated) functions beyond latency maintenance<sup>35–38</sup>. In addition, these systems produced insoluble protein in inclusion bodies that had to be refolded<sup>34,39</sup>. From a functional perspective, mammalian expression systems are preferred for human cytokines as they provide native environments for folding and disulfide formation in the endoplasmic reticulum, glycosylation in the Golgi, and subsequent quality control in the endoplasmic reticulum. These factors ensure that only correctly folded proteins are secreted. However, the development of such systems has proven difficult for TGF $\beta$ s<sup>29,34,40,41</sup>. Most initial trials—based on murine CHO and human HEK293 cells—outputted well below ~1 mg of pure protein per liter of cell culture, owing to low expression levels and multiple purification steps<sup>42–45</sup>. It was not until 2006 that Zou and Sun reported expression of pro-TGF $\beta$ 2 from stable CHO-lec cell lines with a significantly higher yield<sup>40</sup>. However, this system was based on stable cell lines, which generally require long selection processes for establishment, are prone to contamination, require more time for preparation and for protein expression, and have little flexibility with respect to the constructs tested. In addition, the lack of equivalence of recombinant proteins produced in CHO and HEK cells owing to differences in the glycosylation patterns between hamsters and humans has been widely documented<sup>46–48</sup>.

Here, we developed a new homologous production procedure for recombinant full-length human pro-TGF $\beta$ 2 based on transient transfection of human Expi293F cells, a HEK293 cell-line variant that was adapted to high-density suspension growth and selected for high transfection efficiency and protein expression. Such transient expression systems are flexible, the constructs can be changed and retested fast, they have comparable yields to the stable systems, and they can be much more easily shared among scientists by just sending the plasmid<sup>49</sup>. We obtained ~2.7 and ~2.3 mg of N-terminally octahistidine-tagged and Strep-tagged forms, respectively, per liter of cell culture in only three days (Fig. 1A–D). The protein was in a disulfide-linked dimeric state, and a major fraction was cleaved at the inter-domain linker due to endogenous proteolytic processing, as previously reported for this and other TGF $\beta$ s<sup>34,45</sup>. However, the LAP and the GF remained associated in size-exclusion chromatography and native polyacrylamide gel electrophoresis (PAGE). Our efforts to separate them by size-exclusion chromatography in the presence of high salt contents (1 M sodium chloride), chaotropic agents (1 M/4 M urea), detergents (0.05% sodium dodecylsulfate [SDS]), reducing agents (tris[2-carboxyethyl]phosphine or 1,4-dithiothreitol), and low-pH buffers (glycine pH 3.0) failed. We could not isolate them either by ionic exchange chromatography with 20 mM sodium acetate pH 4.0 as buffer and a sodium chloride gradient. This is consistent with previous reports on purified and recombinant pro-TGF $\beta$ 1<sup>45,50</sup> and with the very high affinity of GF and LAP for each other, with dissociation constant values in the low nanomolar range<sup>51</sup>. In contrast, other publications reported activation of TGF $\beta$ s *in vitro* but only under harsh conditions including extreme pH values, high temperature, presence of peptidases and glucosidases, and presence of SDS and urea<sup>34,40,52,53</sup>.

**Separation and crystallization of mature TGF $\beta$ 2.** During studies of the interactions between induced human  $\alpha_2$ -macroglobulin (h $\alpha_2$ M), a ~720-kDa homotetrameric pan-peptidase inhibitor purified from blood, and recombinant human pro-TGF $\beta$ 2 (Marino-Puertas, del Amo-Maestro, Taulés, Gomis-Rüth & Goulas, manuscript in preparation), a mixture of both proteins was set up for crystallization. Diffraction-grade protein crystals appeared after five days with 20% isopropanol, 0.2 M calcium chloride, and 0.1 M sodium acetate pH 4.6 as reservoir solution (Fig. 1E). Reducing SDS-PAGE (Fig. 1F) and peptide mass fingerprinting of carefully washed and dissolved crystals indicated that the crystallized species was the GF. The crystals belonged to a hitherto undescribed, tightly-packed tetragonal space group, diffracted to 2.0 Å resolution, and contained half a GF disulfide-linked dimer per asymmetric unit. Diffraction data processing statistics are provided in Table 1. These



**Figure 1.** Production, purification and crystallization of human TGF $\beta$ 2. (A) Reducing SDS-PAGE depicting N-terminally octahistidine-tagged pro-TGF $\beta$ 2 after Ni-NTA affinity purification. M, molecular mass marker; FT, flow-through; W, wash step; 1E, first elution; and 2E, second elution. Black arrows pinpoint (*top to bottom*) intact pro-TGF $\beta$ 2, LAP, and the GF in lane 1E. (B) Reducing SDS-PAGE of fractions (F15–F21) of the size-exclusion chromatography purification step (*left panel*) and Western-blot analysis of fractions F15–F21 employing an anti-histidine-tag antibody (*right panel*). (C,D), same as (A,B) for N-terminally Strep-tagged pro-TGF $\beta$ 2. In (D), an anti-Strep-tag antibody was used. (E) Representative tetragonal crystals of mature TGF $\beta$ 2 of  $\sim$ 20 microns maximal dimension. (F) Reducing SDS-PAGE of  $\sim$ 90 collected, carefully washed and dissolved diffraction-grade crystals revealing they contain mature TGF $\beta$ 2 (black arrow). (G) Gelatin zymogram of pooled and purified crystallization drop supernatant showing a band pinpointed by an arrow corresponding to the mass of human  $\alpha_2$ -macroglobulin associated with gelatinolytic activity. The original gels used for panels A–D, F and G can be found in the Supplementary Information.

crystals also appeared when pro-TGF $\beta$ 2 alone was subjected to similar crystallization conditions. In this case, the crystals diffracted to lower resolution, i.e. h $\alpha_2$ M apparently played a favorable role as an additive for crystallization. To pursue this further, we collected the supernatant from crystallization drops that had given rise to crystals, purified it by size-exclusion chromatography, and assayed a fraction that migrated according to the mass of h $\alpha_2$ M by gelatin zymography. We detected gelatinolytic activity (Fig. 1G), which points to a contaminant present in the purified h $\alpha_2$ M sample. Accordingly, we conclude that the low pH of the crystallization assay, together with the crystallization process, achieved the separation of the LAP and GF moieties that we could not obtain by chromatography (see above). This process was probably facilitated by a peptidolytic contaminant present in the purified h $\alpha_2$ M sample, as peptidases trapped within the induced tetrameric h $\alpha_2$ M cage are known to still possess activity<sup>54–56</sup>.

Dataset	Mature TGFβ2
<b>Data processing</b>	
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2
Cell constants (a and c, in Å)	55.57, 70.57
Wavelength (Å)	1.0332
No. of measurements/unique reflections	192,672/7,919
Resolution range (Å)	70.6–2.00 (2.12–2.00) <sup>a</sup>
Completeness (%)	100.0 (99.9)
R <sub>merge</sub>	0.070 (2.495)
R <sub>meas</sub> /CC <sup>1/2</sup>	0.072 (2.546)/1.000 (0.870)
Average intensity	23.4 (1.7)
B-Factor (Wilson) (Å <sup>2</sup> )/Aver. multiplicity	56.6/24.3 (24.8)
<b>Structure refinement</b>	
Resolution range used for refinement (Å)	43.7–2.00
No. of reflections used (test set)	7,511 (407)
Crystallographic R <sub>factor</sub> (free R <sub>factor</sub> )	0.217 (0.253)
No. of protein residues/atoms/solvent molecules	112/890/23
Correlation coefficient F <sub>obs</sub> -F <sub>calc</sub> with all reflections/test set	0.943/0.938
<b>Rmsd from target values</b>	
Bonds (Å)/angles (°)	0.010/1.18
Average B-factors (Å <sup>2</sup> ) (all/protein)	66.2/66.4
<b>All-atom contacts and geometry analysis<sup>b</sup></b>	
<b>Residues</b>	
in favored regions/outliers/all residues	102 (93%)/0/110
outlying rotamers/bonds/angles/chirality/planarity	4/0/0/0/0
All-atom clashscore	1.7

**Table 1.** Crystallographic data. <sup>a</sup>Data processing values in parenthesis are for the outermost resolution shell. <sup>b</sup>According to the wwPDB X-ray Structure Validation Report.

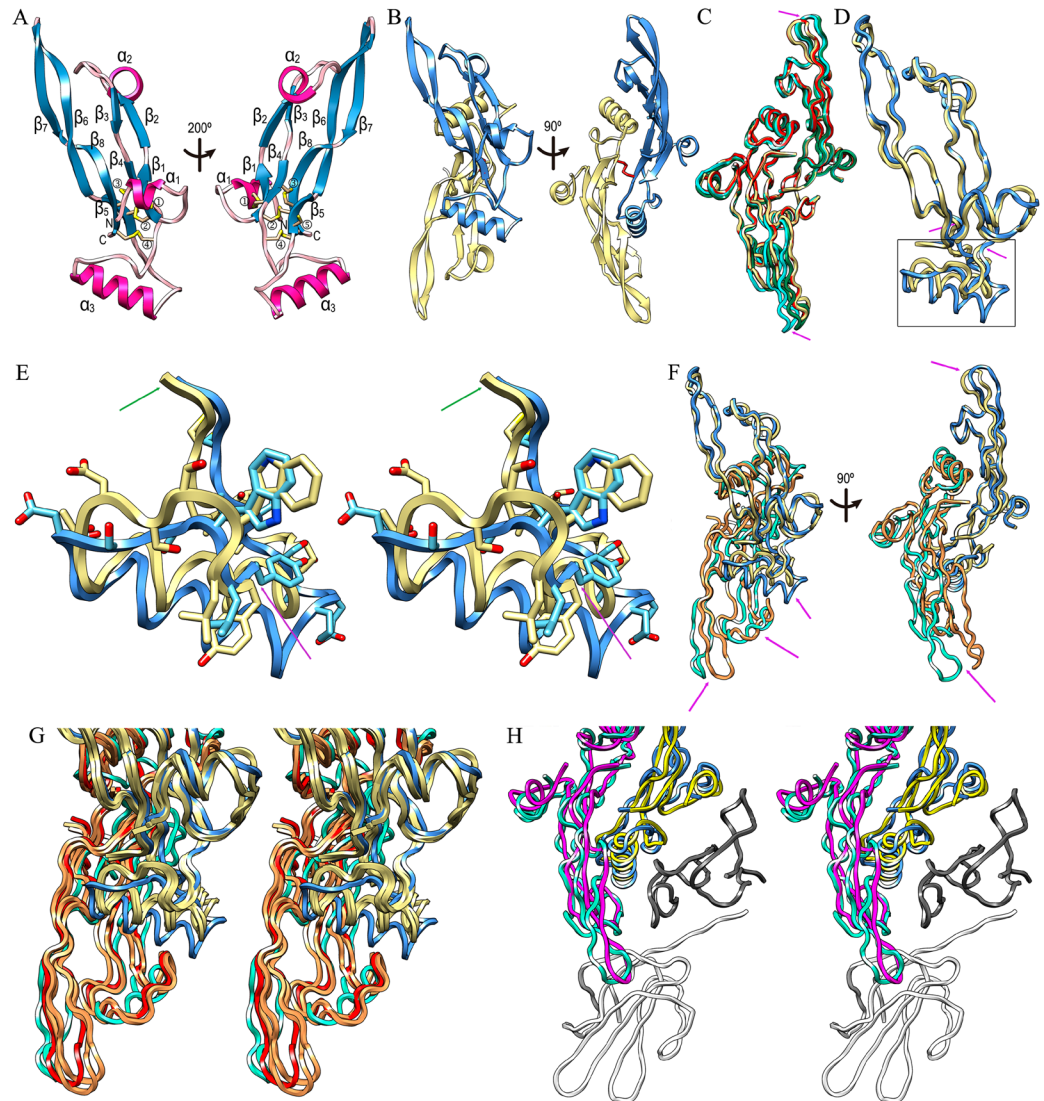
**Structure of mature TGFβ2.** The crystal structure was solved by maximum likelihood-scored molecular replacement, which gave a unique solution for one GF per asymmetric unit at final Eulerian angles and fractional cell coordinates ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ ,  $y$ , and  $z$ ) 348.4, 21.1, 119.9, 0.127, 0.696, and  $-0.899$ . The initial values for the rotation/translation function Z-scores were 6.0/14.0 and the final log-likelihood gain was 132. Taken together, these values indicated that P4<sub>1</sub>2<sub>1</sub>2 was the correct space group, and the final model after model rebuilding and refinement contained residues A<sup>303</sup>-S<sup>414</sup> and 23 solvent molecules. Segments P<sup>351</sup>-W<sup>354</sup> and N<sup>371</sup>-S<sup>377</sup> were flexible but clearly resolved in the final Fourier map for their main chains. Table 1 provides refinement and model validation statistics.

Human TGFβ2 GF is an elongated  $\alpha/\beta$ -fold molecule consisting of an N-terminal helix  $\alpha 1$  disembedding into a fourfold antiparallel sheet of simple up-and-down connectivity (Fig. 2A). Each strand is subdivided into two,  $\beta 1 + \beta 2$ ,  $\beta 3 + \beta 4$ ,  $\beta 5 + \beta 6$ , and  $\beta 7 + \beta 8$ , and the sheet is slightly curled backwards with respect to helix  $\alpha 1$ . The most exposed segment of the moiety is the tip of hairpin  $\beta 6\beta 7$ , and  $\beta$ -ribbon  $\beta 2\beta 3$  is linked by a second  $\alpha$ -helix ( $\alpha 2$ ) as part of a loop, whose conformation is mediated by the *cis* conformation of residue P<sup>338</sup>. Finally, a long loop segment connects strands  $\beta 4$  and  $\beta 5$  and contains helix  $\alpha 3$ , whose axis is roughly perpendicular to the  $\beta$ -sheet. TGFβ2 is internally crosslinked by four disulfides forming a cysteine knot (Fig. 2A) and a further intermolecular disulfide by symmetric C<sup>379</sup> residues links two crystallographic symmetry mates to yield the functional dimer. Here, helix  $\alpha 3$  of one protomer nestles into the concave face of the  $\beta$ -sheet of the other protomer (Fig. 2B). The respective tips of hairpins  $\beta 6\beta 7$ , as well as  $\beta$ -ribbons  $\beta 2\beta 3$  with connecting helices  $\alpha 2$ , are exposed for functional interactions.

**Comparison with previous TGFβ2 structures.** The structure of isolated human TGFβ2 GF was solved in 1992 by two groups simultaneously. They obtained the same trigonal crystal form with half a disulfide-linked dimer per asymmetric unit (Protein Data Bank access codes [PDB] 1TFG<sup>57</sup> and 2TGI<sup>58</sup>, see Table 2). In 2014, the complex of the GF with the Fab fragment of a neutralizing antibody was reported in an orthorhombic crystal form (PDB 4KXZ;<sup>26</sup>). These crystals contained two GF dimers per asymmetric unit, each bound to two Fab moieties. Three more structures of TGFβ2 GF were reported in 2017<sup>59</sup> (Table 2). These corresponded to engineered monomeric forms from human (PDB 5TX4; in complex with TGFR-II ectodomain) and mouse (PDB 5TX2 and 5TX6) that lacked helix  $\alpha 3$  and encompassed several point mutations.

Superposition of the GF structures of PDB 2TGI, 1TFG, and 4KXZ (Fig. 2C) revealed very similar conformations of bound and unbound protomers and dimers, despite distinct chemical and crystallographic environments, which can generally contribute to different conformations owing to crystallographic artifacts<sup>60</sup>. The *rmsd* values with respect to PDB 2TGI (considered hereafter the reference structure) upon superposition of one protomer were 0.26 Å (1TFG), 1.05 Å (4KXZ dimer AB), and 1.22 Å (4KXZ dimer DE) for 112, 111, and 111 common





**Figure 2.** TGFβ2 in a new conformation. (A) Ribbon-type plot of human TGFβ2 (*left panel*) and after vertical rotation (*right panel*). The eight β-strands β1 (residues C<sup>317</sup>-R<sup>320</sup>), β2 (L<sup>322</sup>-D<sup>325</sup>), β3 (G<sup>340</sup>-N<sup>342</sup>), β4 (F<sup>345</sup>-A<sup>347</sup>), β5 (C<sup>379</sup>-S<sup>382</sup>), β6 (D<sup>384</sup>-I<sup>394</sup>), β7 (T<sup>397</sup>-S<sup>404</sup>), and β8 (M<sup>406</sup>-S<sup>414</sup>), as well as the three helices α1 (A<sup>306</sup>-F<sup>310</sup>), α2 (F<sup>326</sup>-L<sup>330</sup>), and α3 (Q<sup>359</sup>-I<sup>370</sup>), are labeled, as are the N- and the C-terminus. The four intramolecular disulfides are depicted for their side chains and labeled ① (C<sup>309</sup>-C<sup>318</sup>), ② (C<sup>317</sup>-C<sup>380</sup>), ③ (C<sup>346</sup>-C<sup>411</sup>) and ④ (C<sup>350</sup>-C<sup>413</sup>). The cysteine engaged in a symmetric intermolecular disulfide (C<sup>379</sup>) is further labeled as ⑤. (B) Human mature TGFβ2 dimer with one protomer in the orientation of A (*left panel*) in blue and the second in pale yellow (*left*), which is related to the former through a horizontal crystallographic twofold axis. An orthogonal view is provided in the *right panel*, the intermolecular disulfide is depicted as red sticks. (C) Superposition of the Cα-traces in ribbon presentation of the dimers of previously reported structures of mature TGFβ2 (PDB 2TGI, pale yellow; PDB 1TFG, red; PDB 4KXZ dimer AB, green; and PDB 4KXZ dimer DE, aquamarine) after optimal superposition of the respective top protomers. Magenta arrows pinpoint the only points of significant deviation, i.e. the tips of respective β-ribbons β6/β7. The view is that of B (*right panel*). (D) Superposition of the protomers of PDB 2TGI in pale yellow onto the current structure (PDB 6I9J) in the view of A (*left panel*). The region of largest deviation (Y<sup>352</sup>-C<sup>380</sup>) is pinpointed by magenta arrows and framed. (E) Close-up in cross-eye stereo of the framed region of (D), with ribbon and carbons in pale yellow for PDB 2TGI and in blue/cyan for PDB 6I9J. Y<sup>352</sup> and C<sup>380</sup> are pinpointed by a magenta and a green arrow, respectively. (F) Superposition of the dimers of PDB 6I9J (top protomer in blue, bottom protomer in aquamarine) and PDB 2TGI (top protomer in pale yellow, bottom protomer in orange) in the views of (B). Owing to the different chain traces of segment Y<sup>352</sup>-C<sup>380</sup> (top magenta arrow in the *left panel*, see also [D]), substantial variations are observed at distal regions of the bottom protomers. (G) Close-up in stereo of F (*left panel*) depicting PDB 6I9J (dimer in blue/aquamarine); PDB 2TGI and 4KXZ dimer AB, both in pale yellow/orange; and human TGFβ3 as found in its complex with the ectodomains of human TGFR-I and -II (PDB 2PJY; protomers in pale yellow and red). (H) Superposition in stereo of the dimers of PDB 6I9J (dimer in blue/aquamarine) and human TGFβ3 (PDB 2PJY dimer in yellow/magenta) in complex with the ectodomains of human TGFR-I (dark grey) and -II (white) in the orientation of B (*right panel*). TGFβ2 in PDB 6I9J must rearrange to bind the receptors as performed by TGFβ3.

PDB	Resolution (Å)	Residues <sup>a</sup>	No. of residues	No. of copies in a.u.	State (Isolated/Complex)	Space group and cell constants (Å/°)	Organism	Reference	rmsd (Å) over residues <sup>f</sup>
1TFG	1.95	303–414	112	1	I	P3 <sub>2</sub> 21 a = b = 60.60, c = 75.20	Human	<sup>57</sup>	2.2/111
2TGI	1.80	303–414	112	1	I	P3 <sub>2</sub> 21 a = b = 60.60, c = 75.30	Human	<sup>58</sup>	2.3/111
4KXZ	2.83	303–414	112	4	C	P2 <sub>1</sub> 2 <sub>1</sub> 2 a = 131.20, b = 359.68 c = 64.63	Human	<sup>26</sup>	2.0/111
5TX2	1.82	303–414	93 <sup>c</sup>	2	I	C2 a = 99.46, b = 33.36 c = 54.13, β = 109.6	Mouse	<sup>59</sup>	1.3/91
5TX4	1.88	303–414	92 <sup>d</sup>	1	C	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> a = 39.02, b = 70.77 c = 77.17	Human	<sup>59</sup>	1.2/91
5TX6	2.75	303–414	93 <sup>c</sup>	3	I	P3 <sub>1</sub> 21 a = b = 81.74, c = 80.93	Mouse	<sup>59</sup>	0.8/88
5TY4	2.90 <sup>b</sup>	317–413	65 <sup>c</sup>	1	C	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> a = 41.53, b = 71.33 c = 79.51	Human	<sup>86</sup>	0.8/65
6I9J	2.00	303–314	112	1	I	P4 <sub>1</sub> 2 <sub>1</sub> 2 a = b = 55.57, c = 70.57	Human	This work	—

**Table 2.** Crystal structures of TGFβ2. <sup>a</sup>See UP entries P61812 and P27090 for human and mouse TGFβ2 sequences, respectively. <sup>b</sup>Obtained by crystal electron diffraction. <sup>c</sup>Mutant Δ354–373, K327R, R328K, L353R, A376K, C379S, L391V, I394V, K396R, T397K, and I400V. Contains an extra M at the N-terminus. <sup>d</sup>Mutant Δ354–373, K327R, R328K, L353R, A376K, C379S, L391V, I394V, K396R, T397K, and I400V. <sup>e</sup>Mutant Δ354–378, K327R, and R328K. <sup>f</sup>Computed for the common Cα atoms of a protomer with the DALI program<sup>84</sup> with respect to 6I9J.

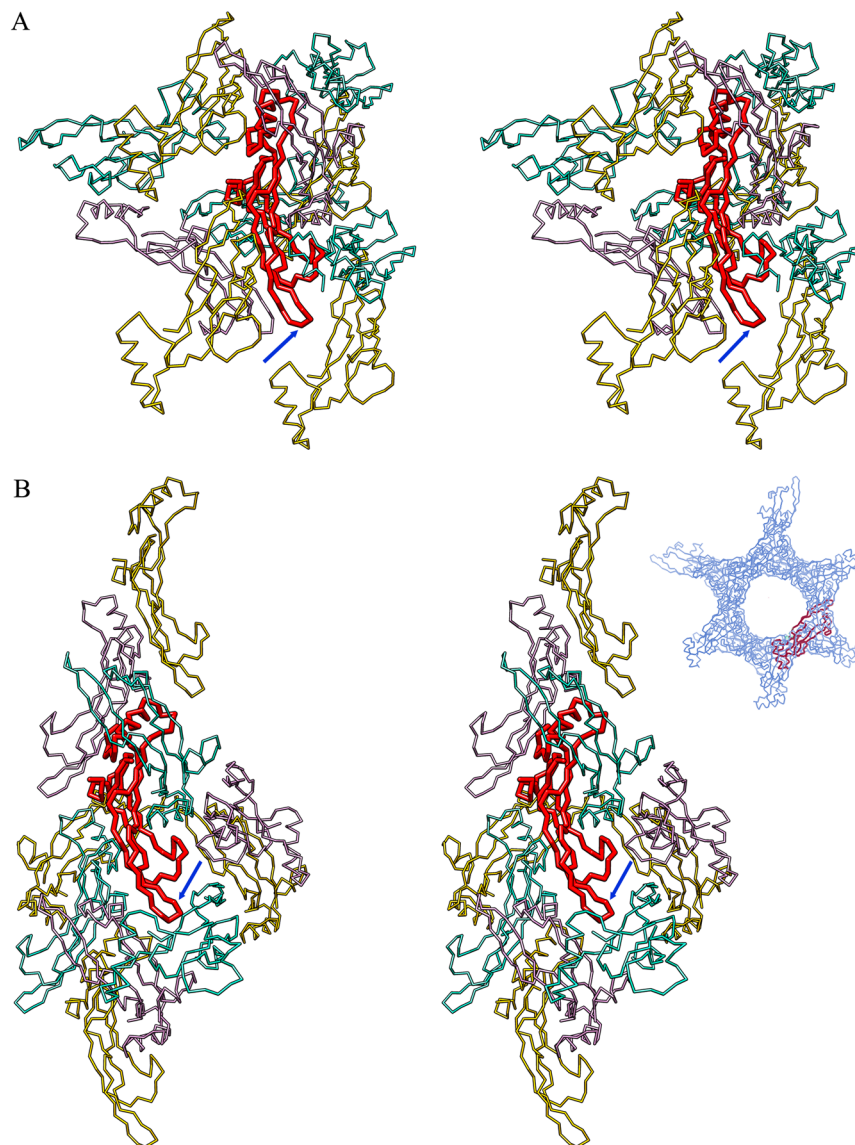
Cα atoms, respectively. Only minor variations occurred at the tips of hairpins β6β7 and at helix α2 (Fig. 2C). Interestingly, also the engineered monomeric variants kept the overall structure of the native TGFβ2 protomer, with the exception of a loop that replaced helix α3<sup>59</sup>. This close similarity was also found with the unbound human ortholog TGFβ3 (PDB 1TGJ<sup>61</sup>).

Inversely, the present GF structure (PDB 6I9J; hereafter the current structure) showed deviations from the reference structure, which are reflected by an *rmsd* of 1.90 Å for 107 common Cα atoms. The central β-sheets nicely coincide and only minor differences are found within the deviating regions of the above GF structures. However, a major rearrangement is found within segment Y<sup>352</sup>-C<sup>380</sup>, which encompasses helix α3 *plus* the flanking linkers to strands β4 and β5 (Fig. 2D,E). Owing to an outward movement of segment W<sup>354</sup>-D<sup>357</sup>, α3 is translated by maximally ~5.5 Å (at I<sup>370</sup>). This, in turn, causes downstream segment E<sup>373</sup>-C<sup>379</sup> to be folded inward, with a maximal displacement of ~6.7 Å (at A<sup>374</sup>). This displacement cascades into a slight shift (~2 Å) of C<sup>379</sup> and, thus, the intermolecular disulfide. These differences within protomers further result in variable arrangement of the dimers (Fig. 2F). The second protomer is rotated by ~10°, so that K<sup>396</sup> at the tip of hairpin β6β7 is displaced by ~9 Å. In addition, the central β-sheets overlap but are laterally shifted with respect to each other.

Finally, inspection of the respective crystal packings of the current and reference structures (Fig. 3A,B) reveals that while the latter is loosely packed in a hexagonal honeycomb-like arrangement, with large void channels of > 40 Å diameter and high solvent content (61%), the former is tightly packed (solvent content of 43%). In particular, the tip of hairpin β6β7 is not engaged in significant crystal contacts in either crystal form, which supports that both conformations are authentic and do not result from crystallization artifacts.

**Potential implications of a new dimeric arrangement.** Structural information on TGFβ binding is available for human TGFβ1 and TGFβ3, whose wild-type forms were crystallized in complexes with the ectodomains of TGFβR-I and -II (PDB 3KFD<sup>62</sup>; PDB 1KTZ<sup>63</sup>, and PDB 2PJY<sup>12</sup>). In the binary complex of TGFβ3 with TGFβR-II (PDB 1KTZ), the cytokine was observed in a unique dimeric arrangement, termed “open”, in which the respective helices α3 were completely disordered. This led the second protomer to be rotated by ~180° and the tip of hairpin β6β7 to be displaced by ~36 Å when compared with the reference structure<sup>63</sup>. The functional significance of this structure, which differs from any other mature TGFβ dimer, is not clear.

By contrast, the other two complex structures, which contained both receptors, showed that the TGFβ1 and TGFβ3 isoforms dimerize very similarly to the unbound reference structure when bound to receptors (*rmsd* of 1.43 Å and 1.61 Å for 222 and 219 common Cα atoms for PDB 3KFD and PDB 2PJY, respectively). In contrast, they differed from the current structure (Fig. 2G), which shows a *rmsd* of 2.94 Å for 216 common Cα atoms when compared with the reference. In addition, both cytokine isoforms—and by similarity probably also TGFβ2—bound their receptors in an equivalent manner. TGFβR-II was contacted by the tip of hairpin β6β7 and the linker between α2 and β3 of one protomer, and TGFβR-I was liganded by the interface between protomers created by strands β5–β8 from one protomer and α1 *plus* the region around α3 of the other protomer. In addition, the TGFβRs further contacted each other through the N-terminal extension of TGFβR-II (Fig. 2H). Notably, the binding mode of TGFβR-II was shared with the engineered monomeric variant of TGFβ2 (PDB 5TX4<sup>59</sup>). In contrast, superposition of the cytokine dimers of the current structure and the triple complexes evinced that among other structural elements hairpin β6β7 is displaced away from interacting segment S<sub>52</sub>-E<sub>55</sub> of TGFβR-II (see PDB 2PJY) by ~2.7 Å. This hairpin rearrangement also impairs interaction of the β5β8 ribbon with TGFβR-I segment I<sub>54</sub>-F<sub>60</sub>. Thus, TGFβRs could not bind the current structure in the same way they bind TGFβ1 and TGFβ3 without rearrangement.



**Figure 3.** Crystal packing. **(A)** Cross-eye stereoplot showing the crystal packing of the current TGF $\beta$ 2 structure (PDB 619J), with the protomer in the asymmetric unit in red and the surrounding symmetry mates in gold, plum and aquamarine. The tip of  $\beta$ -ribbon  $\beta$ 637 is pinpointed by a blue arrow. **(B)** Same as **(A)** showing the crystal environment of the reference TGF $\beta$ 2 structure (PDB 2TG1). The top right inset shows a view down the crystallographic threefold axis to illustrate the solvent channels. The protomer in the asymmetric unit is in red, the symmetry mates setting up the crystal lattice in blue.

**Conclusion.** We developed a recombinant human overexpression system based on Expi293F cells grown in suspension, which produces high yields of well-folded human N-terminally histidine- or Strep-tagged pro-TGF $\beta$ 2 with native post-translational modifications.

We further crystallized mature TGF $\beta$ 2 in a different crystallographic space group, which deviates mainly in the region around helix  $\alpha$ 3 from current functional TGF $\beta$ 1, TGF $\beta$ 2 and TGF $\beta$ 3 structures. Importantly, this region is the segment that shows the highest variability in sequence among TGF $\beta$ s<sup>26</sup>. Although crystal packing artifacts cannot be completely ruled out, the fact that several different crystal forms of the three TGF $\beta$  GFs had produced highly similar structures to date suggests that the new conformation may be authentic and have functional implications.

The divergent conformation of the region around helix  $\alpha$ 3 led to differences in the dimer, which could not bind cognate TGFR-I and -II in the same way as TGF $\beta$ 1 and TGF $\beta$ 3 do. Moreover, as revealed by pro-TGF $\beta$ 1 structures from pig (PDB 5VQF<sup>64</sup>) and human (PDB 5VQP<sup>65</sup> and PDB 6GFF<sup>66</sup>), this region differed substantially between latent and mature moieties owing to the presence of the respective LAPs. Hence, the GF protomers associated differently within the respective dimers. Large differences were also found between the mature form and the pro-form of the more distantly related TGF $\beta$  family member activin A (PDB 2ARV<sup>67</sup> and PDB 5HLY<sup>68</sup>). In contrast, bone morphogenetic factor 9 showed deviations in hairpin  $\beta$ 637 but kept the dimer structure (PDB



5I05<sup>69</sup> and PDB 4YCG<sup>70</sup>). Overall, we conclude that our mature structure may represent an inactive variant or be one of an ensemble of conformational states, which may still undergo an induced fit or selection fit mechanism to form a functional ternary ligand-receptor complex.

## Materials and Methods

**Protein production and purification.** The coding sequence of human pro-TGF $\beta$ 2 without its signal peptide, i.e. spanning the LAP (L<sup>21</sup>-R<sup>302</sup>) and the GF (A<sup>303</sup>-S<sup>414</sup>), was inserted in consecutive PCR steps into the Gateway pCMV-SPORT6 vector (ThermoFisher) in frame with the Kozak sequence. The signal peptide from the V-J2-C region of a mouse Ig  $\kappa$ -chain was used instead of the native leader sequence, as this was reported to be more efficient for expression<sup>40</sup>. This vector attached an N-terminal octahistidine-tag to the protein of interest. For the five PCR steps, oligonucleotides 5'-TCACCACCACCATCATCTCAGCTGTCTACCTGCAGCA-3'; 5'-GGTTCCTCCACTGGTGACCACCACCATCACCACCACCATC-3'; 5'-GGGTACTGCTGCTCTGGGTTCCAGGTTCCACTGGTGAC-3'; 5'-GACAGACACACTCCTGTCTATGGGTACTGCTGCTC-3'; and 5'-CAATCCCGGGGCCACCATGGAGACAGACACTCC-3' were used as forward primers, respectively, and 5'-CAATCTCGAGCTAGCTGCATTTGCAAGACTTTAC-3' was employed as the reverse primer. The intermediate PCR products were purified with the EZNA Cycle Pure Kit (Omega Bio-tek, USA) prior to the next step. The final PCR product was digested twice with *Sma*I and *Xho*I restriction enzymes (1 h at 37 °C and o/n at 37 °C), with an intercalated PCR product purification step. This product was ligated into pre-digested pCMV-SPORT6 by adding 2  $\mu$ L of T4 DNA Ligase (ThermoFisher) per 20  $\mu$ L of reaction (o/n at r.t.). The resulting plasmid (pS6-TGFB2-H8) was verified for its sequence (GATC Biotech) and transformed into competent *Escherichia coli* DH5 $\alpha$  cells for vector storage and production.

Prior to transfection into mammalian cells, pS6-TGFB2-H8 was produced in *E. coli* DH5 $\alpha$ , purified with the GeneJET Plasmid Maxiprep Kit (ThermoFisher) according to the manufacturer's instructions, and stored in Milli-Q water at 1 mg/mL. Expi293F cells (ThermoFisher), which had been kept in suspension in FreeStyle F17 Expression Medium (Gibco) plus 0.2% Pluronic F-68 at 150 rpm in a humidified atmosphere with 8% CO<sub>2</sub> in air at 37 °C, were transfected at a density of 1  $\times$  10<sup>6</sup> cells/mL with a mixture of 1 mg of purified pS6-TGFB2-H8 and 3 mg of linear 25-kDa polyethylenimine (Polysciences) in 20 mL of Opti-MEM Medium per liter. The mixture of the reagents was incubated for 15–20 min at room temperature and then added dropwise to the cells in 1 L-disposable Erlenmeyer flasks (Fisherbrand) after proper mixing. Cells were harvested for 72 h and then centrifuged at 2,800  $\times$  g in the JLA9.1000 rotor of an Avanti J-20XP centrifuge (Beckman Coulter) for 20 min, and the supernatant was subjected to single-step Ni-NTA affinity chromatography (washing buffer: 50 mM Tris-HCl pH 8.0, 250 mM sodium chloride, 20 mM imidazole; elution buffer: 50 mM Tris-HCl pH 8.0, 250 mM sodium chloride, 300 mM imidazole). Elutions were pooled, concentrated and subjected to size-exclusion chromatography in a Superdex 75 10/300 column (GE Healthcare) attached to an ÄKTA Purifier system (GE Healthcare) at r.t. with 20 mM Tris-HCl pH 8.0, 150 mM sodium chloride as buffer. Purified protein was routinely concentrated with Vivaspin centrifugal devices (Sartorius) with a 10-kDa cutoff, and concentrations were determined by A<sub>280</sub> in a NanoDrop Microvolume spectrophotometer (ThermoFisher). Protein identity was confirmed by peptide mass fingerprinting and purity was assessed by 10% SDS-PAGE using Tris-Glycine buffer and Coomassie Brilliant Blue or silver staining, as well as by Western-blot analysis with a histidine antibody horse radish peroxidase conjugate (His-probe Antibody H-3 HRP, Santa Cruz Biotechnology) at 1:5,000 in PBS buffer further 0.1% in Tween 20.

To produce Strep-tagged pro-TGF $\beta$ 2, plasmid pS6-TGFB2-H8 was modified by PCR to replace the histidine-tag with a twin Strep-tag with an intermediate G/S spacer (sequence WSHPPQFEKGGGSGGGSSGSAWSHPQFEK) by using forward primer 5'-GGTGGAGGTTCTGGAGGTGGAAGTGGAGGTAGCGCATGGAGCCATCCACAATTCGAAAAGCTCAGCTGTCTACCTGC-3' and reverse primer 5'-CTTTTCGAATTGTGGATGGCTCCAGTCACCAGTGGAAACCTGGAACCCAGAGCAG-3'. The resulting PCR product was purified with the EZNA Cycle Pure Kit, phosphorylated with 1  $\mu$ L of T4 polynucleotide kinase (ThermoFisher) per 20  $\mu$ L of reaction (o/n at 37 °C), and ligated as above to generate plasmid pS6-TGFB2-STREP. This plasmid was further processed and transfected into Expi293F cells for protein production as above. Cells were harvested and centrifuged as aforementioned, and the supernatant was extensively dialyzed against 100 mM Tris-HCl pH 8.0, 150 mM sodium chloride and purified by single-step affinity chromatography using Strep-Tactin XT Superflow Suspension resin (iba) according to the manufacturer's instructions (washing buffer: 100 mM Tris-HCl pH 8.0, 150 mM sodium chloride; elution buffer: 100 mM Tris-HCl pH 8.0, 150 mM sodium chloride, 50 mM biotin). Final polishing by size-exclusion chromatography, concentration, and purity assessment followed as above, except that Western-blot analysis was performed with a Streptavidin antibody horse radish peroxidase conjugate (Streptavidin-Peroxidase Antibody from *Streptomyces avidinii*; Sigma-Aldrich) at 1:1000 in PBS, supplemented with 0.1% Tween 20 and 1% BSA.

**Crystallization of mature TGF $\beta$ 2 and diffraction data collection.** Octahistidine-tagged pro-TGF $\beta$ 2, mostly cleaved in the linker between LAP and GF, at ~5 mg/mL in 10 mM Tris-HCl pH 8.0, 150 mM sodium chloride was incubated o/n at 4 °C with human induced  $\alpha_2$ -macroglobulin (h $\alpha_2$ M) at ~7 mg/mL in 10 mM Tris-HCl pH 8.0, 150 mM sodium chloride. The h $\alpha_2$ M was previously purified from blood and reacted with methylamine as reported<sup>71</sup>. The h $\alpha_2$ M/pro-TGF $\beta$ 2 reaction mixture was subjected to crystallization assays by the sitting-drop vapor diffusion method at the Automated Crystallography Platform of Barcelona Science Park. Reservoir solutions were mixed by a Tecan robot and crystallization drops of 100 nL were dispensed by a Cartesian Microsols 4000 XL (Genomic Solutions) robot or a Phoenix nanodrop robot (Art Robbins) on 96  $\times$  2-well MRC nanoplates (Innovadyne). Crystallization plates were stored at 20 °C or 4 °C in Bruker steady-temperature crystal farms, and successful conditions were scaled up to the microliter range in 24-well Cryschem crystallization dishes (Hampton Research).

Diffraction-grade crystals of ~20 microns maximal dimension were obtained at 20 °C with protein solution and 20% isopropanol, 0.2 M calcium chloride, 0.1 M sodium acetate pH 4.6 as reservoir solution from 1 or 2  $\mu$ L: 1  $\mu$ L drops. Carefully washed and pooled crystals revealed the presence of a species of ~12 kDa in SDS-PAGE, which was identified as TGF $\beta$ 2 by peptide mass fingerprinting. Crystals were cryoprotected by immersion in reservoir solution *plus* 15% glycerol and flash cryocooled in liquid nitrogen. Diffraction data were collected at 100 K on a Pilatus 6 M pixel detector (Dectris) at beam line XALOC<sup>72</sup> of the ALBA synchrotron in Cerdanyola (Catalonia, Spain). These data were processed with programs XDS<sup>73</sup> and XSCALE<sup>74</sup>, and transformed with XDSCONV to MTZ format for the CCP4 suite of programs<sup>75</sup>. Crystals belonged to space group P4<sub>1/3</sub>2<sub>1</sub>2 based on the systematic extinctions, contained one mature TGF $\beta$ 2 molecule per asymmetric unit ( $V_M = 2.14$ ; 42.6% solvent contents according to<sup>76</sup>), and diffracted to 2.0 Å resolution.

**Proteolytic assay of crystallization-drop supernatant.** Supernatant from crystallization drops that had produced crystals of mature TGF $\beta$ 2, i.e. which contained methylamine-induced h $\alpha_2$ M, was pooled and subjected to size-exclusion chromatography in a Superose6 10/300 column (GE Healthcare) at r.t. with 10 mM Tris-HCl pH 8.0, 150 mM sodium chloride as buffer. Fractions corresponding to h $\alpha_2$ M (~720 kDa) were concentrated with Vivaspin centrifugal devices (Sartorius) with a 30-kDa cutoff and employed for zymography studies. To this aim, 10% SDS-PAGE gels were prepared containing 0.2% (w/v) gelatin. Protein samples were subjected to electrophoresis at 4 °C with equal volume of SDS-PAGE sample buffer without  $\beta$ -mercaptoethanol. Gels were washed twice in 20 mM Tris-HCl pH 7.4, 150 mM sodium chloride, 10 mM calcium chloride, 2.5% (v/v) Triton X-100 for 15 min and then incubated in the same buffer without detergent for 16 h at 37 °C under gentle shaking. Gels were stained with Coomassie Brilliant Blue.

**Structure solution and refinement.** The structure of mature TGF $\beta$ 2 was solved by likelihood-scoring molecular replacement with the PHASER<sup>77</sup> program and the coordinates of a GF protomer crystallized in a different space group and unit cell (PDB 2TGI<sup>58</sup>). Subsequently, an automatic tracing step was performed with ARP/wARP<sup>78</sup>, which outputted a model that was completed through successive rounds of manual model building with the COOT program<sup>79</sup> and crystallographic refinement with the PHENIX<sup>80</sup> and BUSTER/TNT<sup>81</sup> programs. The latter included TLS refinement.

**Miscellaneous.** Structure figures were made with the CHIMERA program<sup>82</sup>. Structures were superposed with the SSM program<sup>83</sup> within COOT. A search for structural similarity against the PDB was performed with DALI<sup>84</sup>. The final model of human TGF $\beta$ 2 GF was validated with the wwPDB Validation Server (<https://www.wwpdb.org/validation>)<sup>85</sup> and is available at the PDB at <https://www.rcsb.org> (access code 6I9J).

## References

1. Massague, J. The transforming growth factor- $\beta$  family. *Annu. Rev. Cell Biol.* **6**, 597–641 (1990).
2. Jenkins, G. The role of proteases in transforming growth factor- $\beta$  activation. *Int. J. Biochem. Cell Biol.* **40**, 1068–1078 (2008).
3. Derynck, R. & Budi, E. H. Specificity, versatility, and control of TGF- $\beta$  family signaling. *Sci. Signal.* **12**, eaav5183 (2019).
4. Frolik, C. A., Dart, L. L., Meyers, C. A., Smith, D. M. & Sporn, M. B. Purification and initial characterization of a type  $\beta$  transforming growth factor from human placenta. *Proc. Natl. Acad. Sci. USA* **80**, 3676–3680 (1983).
5. Govindan, R. & Bhoola, K. D. Genealogy, expression, and cellular function of transforming growth factor- $\beta$ . *Pharmacol. Ther.* **98**, 257–265 (2003).
6. Travis, M. A. & Sheppard, D. TGF- $\beta$  activation and function in immunity. *Annu. Rev. Immunol.* **32**, 51–82 (2014).
7. Teicher, B. A. Malignant cells, directors of the malignant process: role of transforming growth factor- $\beta$ . *Cancer Metastasis Rev.* **20**, 133–143 (2001).
8. Akhurst, R. J. TGF- $\beta$  antagonists: why suppress a tumor suppressor? *J. Clin. Invest.* **109**, 1533–1536 (2002).
9. Hyytiäinen, M., Penttinen, C. & Keski-Oja, J. Latent TGF- $\beta$  binding proteins: extracellular matrix association and roles in TGF- $\beta$  activation. *Crit. Rev. Clin. Lab. Sci.* **41**, 233–264 (2004).
10. Principe, D. R. *et al.* TGF- $\beta$ : duality of function between tumor prevention and carcinogenesis. *J. Natl. Cancer Inst.* **106**, djt369 (2014).
11. Schmierer, B. & Hill, C. S. TGF $\beta$ -SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* **8**, 970–982 (2007).
12. Groppe, J. *et al.* Cooperative assembly of TGF- $\beta$  superfamily signaling complexes is mediated by two disparate mechanisms and distinct modes of receptor binding. *Mol. Cell* **29**, 157–168 (2008).
13. Li, M. O., Wan, Y. Y., Sanjabi, S., Robertson, A. K. & Flavell, R. A. Transforming growth factor- $\beta$  regulation of immune responses. *Annu. Rev. Immunol.* **24**, 99–146 (2006).
14. Ling, T. Y., Huang, Y. H., Lai, M. C., Huang, S. S. & Huang, J. S. Fatty acids modulate transforming growth factor- $\beta$  activity and plasma clearance. *FASEB J.* **17**, 1559–1561 (2003).
15. Mishra, L., Derynck, R. & Mishra, B. Transforming growth factor- $\beta$  signaling in stem cells and cancer. *Science* **310**, 68–71 (2005).
16. Massague, J. TGF $\beta$  in cancer. *Cell* **134**, 215–230 (2008).
17. Padua, D. & Massague, J. Roles of TGF $\beta$  in metastasis. *Cell Res.* **19**, 89–102 (2009).
18. Theron, A. J., Anderson, R., Rossouw, T. M. & Steel, H. C. The role of Transforming Growth Factor  $\beta$ -1 in the progression of HIV/AIDS and development of non-AIDS-defining fibrotic disorders. *Front. Immunol.* **8**, 1461 (2017).
19. Löffek, S. Transforming of the tumor microenvironment: implications for TGF- $\beta$  inhibition in the context of immune-checkpoint therapy. *J. Oncol.* **2018**, 9732939 (2018).
20. Ahmadi, A., Najafi, M., Farhood, B. & Mortezaee, K. Transforming growth factor- $\beta$  signaling: tumorigenesis and targeting for cancer therapy. *J. Cell. Physiol.* **234**, 12173–12187 (2019).
21. Annes, J. P., Munger, J. S. & Rifkin, D. B. Making sense of latent TGF $\beta$  activation. *J. Cell Sci.* **116**, 217–224 (2003).
22. Rifkin, D. B. Latent transforming growth factor-beta (TGF- $\beta$ ) binding proteins: orchestrators of TGF- $\beta$  availability. *J. Biol. Chem.* **280**, 7409–7412 (2005).
23. Hinck, A. P., Mueller, T. D. & Springer, T. A. Structural biology and evolution of the TGF- $\beta$  family. *Cold Spring Harb. Perspect. Biol.* **8**, a022103 (2016).
24. Dubois, C. M., Laprise, M. H., Blanchette, E., Gentry, L. E. & Leduc, R. Processing of transforming growth factor  $\beta$ 1 precursor by human furin convertase. *J. Biol. Chem.* **270**, 10618–10624 (1995).

25. Yang, Z. *et al.* Absence of integrin-mediated TGF $\beta$ 1 activation *in vivo* recapitulates the phenotype of TGF $\beta$ 1-null mice. *J. Cell Biol.* **176**, 787–793 (2007).
26. Moulin, A. *et al.* Structures of a pan-specific antagonist antibody complexed to different isoforms of TGF $\beta$  reveal structural plasticity of antibody-antigen interactions. *Protein Sci.* **23**, 1698–1707 (2014).
27. de Martin, R. *et al.* Complementary DNA for human glioblastoma-derived T cell suppressor factor, a novel member of the transforming growth factor- $\beta$  gene family. *EMBO J.* **6**, 3673–3677 (1987).
28. Marquardt, H., Lioubin, M. N. & Ikeda, T. Complete amino acid sequence of human transforming growth factor type  $\beta$  2. *J. Biol. Chem.* **262**, 12127–12131 (1987).
29. Wilbers, R. H. *et al.* Co-expression of the protease furin in *Nicotiana benthamiana* leads to efficient processing of latent transforming growth factor- $\beta$ 1 into a biologically active protein. *Plant Biotechnol. J.* **14**, 1695–1704 (2016).
30. Soleimani, A. *et al.* Role of TGF- $\beta$  signaling regulatory microRNAs in the pathogenesis of colorectal cancer. *J. Cell Physiol.* **234**, 14574–14580 (2019).
31. Pircher, R., Jullien, P. & Lawrence, D. A.  $\beta$ -Transforming growth factor is stored in human blood platelets as a latent high molecular weight complex. *Biochem. Biophys. Res. Commun.* **136**, 30–37 (1986).
32. Cheifetz, S. *et al.* The transforming growth factor- $\beta$  system, a complex pattern of cross-reactive ligands and receptors. *Cell* **48**, 409–415 (1987).
33. Ogawa, Y. & Seyedin, S. M. Purification of transforming growth factors  $\beta$ 1 and  $\beta$ 2 from bovine bone and cell culture assays. *Methods Enzymol.* **198**, 317–327 (1991).
34. Huang, T. & Hinck, A. P. Production, isolation, and structural analysis of ligands and receptors of the TGF- $\beta$  superfamily. *Methods Mol. Biol.* **1344**, 63–92 (2016).
35. Thomas, G. J., Hart, I. R., Speight, P. M. & Marshall, J. F. Binding of TGF- $\beta$ 1 latency-associated peptide (LAP) to  $\alpha$ v $\beta$ 6 integrin modulates behaviour of squamous carcinoma cells. *Br. J. Cancer* **87**, 859–867 (2002).
36. Nomura, K., Tada, H., Kuboki, K. & Inokuchi, T. Transforming growth factor- $\beta$ 1 latency-associated peptide and soluble  $\beta$ -glycan prevent a glucose-induced increase in fibronectin production in cultured human mesangial cells. *Nephron* **91**, 606–611 (2002).
37. Ali, N. A. *et al.* Latency associated peptide has *in vitro* and *in vivo* immune effects independent of TGF- $\beta$ 1. *PLoS ONE* **3**, e1914 (2008).
38. Lee, M. J. Heparin inhibits activation of latent transforming growth factor- $\beta$ 1. *Pharmacology* **92**, 238–244 (2013).
39. Schlunegger, M. P. *et al.* Crystallization and preliminary X-ray analysis of recombinant human transforming growth factor  $\beta$ 2. *FEBS Lett.* **303**, 91–93 (1992).
40. Zou, Z. & Sun, P. D. An improved recombinant mammalian cell expression system for human transforming growth factor- $\beta$ 2 and - $\beta$ 3 preparations. *Prot. Expr. Purif.* **50**, 9–17 (2006).
41. Boudrel, L. *et al.* Recombinant human transforming growth factor- $\beta$ 1: expression by Chinese hamster ovary cells, isolation, and characterization. *Protein Expr. Purif.* **4**, 130–140 (1993).
42. Madisen, L. *et al.* Expression and characterization of recombinant TGF- $\beta$ 2 proteins produced in mammalian cells. *DNA* **8**, 205–212 (1989).
43. Graycar, J. L. *et al.* Human transforming growth factor- $\beta$ 3: recombinant expression, purification, and biological activities in comparison with transforming growth factors- $\beta$ 1 and - $\beta$ 2. *Mol. Endocrinol.* **3**, 1977–1986 (1989).
44. Madisen, L., Lioubin, M. N., Marquardt, H. & Purchio, A. F. High-level expression of TGF- $\beta$ 2 and the TGF- $\beta$ 2(414) precursor in Chinese hamster ovary cells. *Growth Factors* **3**, 129–138 (1990).
45. Gentry, L. E., Lioubin, M. N., Purchio, A. F. & Marquardt, H. Molecular events in the processing of recombinant type 1 pre-transforming growth factor  $\beta$  to the mature polypeptide. *Mol. Cell Biol.* **8**, 4162–4168 (1988).
46. Croset, A. *et al.* Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J. Biotechnol.* **161**, 336–348 (2012).
47. van den Nieuwenhof, I. M. *et al.* Recombinant glycodelin carrying the same type of glycan structures as contraceptive glycodelin-A can be produced in human kidney 293 cells but not in Chinese hamster ovary cells. *Eur. J. Biochem.* **267**, 4753–4762 (2000).
48. Gaudry, J. P. *et al.* Purification of the extracellular domain of the membrane protein GlialCAM expressed in HEK and CHO cells and comparison of the glycosylation. *Protein Expr. Purif.* **58**, 94–102 (2008).
49. Geisse, S. & Voedisch, B. Transient expression technologies: past, present, and future. *Methods Mol. Biol.* **899**, 203–219 (2012).
50. Huber, D., Fontana, A. & Bodmer, S. Activation of human platelet-derived latent transforming growth factor- $\beta$ 1 by human glioblastoma cells. Comparison with proteolytic and glycosidic enzymes. *Biochem. J.* **277**, 165–173 (1991).
51. Miller, D. M. *et al.* Characterization of the binding of transforming growth factor- $\beta$ 1, - $\beta$ 2, and - $\beta$ 3 to recombinant  $\beta$ 1-latency-associated peptide. *Mol. Endocrinol.* **6**, 694–702 (1992).
52. Brown, P. D., Wakefield, L. M., Levinson, A. D. & Sporn, M. B. Physicochemical activation of recombinant latent transforming growth factor-beta's 1, 2, and 3. *Growth Factors* **3**, 35–43 (1990).
53. Munger, J. S. *et al.* Latent transforming growth factor- $\beta$ : structural features and mechanisms of activation. *Kidney Int.* **51**, 1376–1382 (1997).
54. Barrett, A. J. & Starkey, P. M. The interaction of  $\alpha$ <sub>2</sub>-macroglobulin with proteinases. Characteristics and specificity of the reaction, and a hypothesis concerning its molecular mechanism. *Biochem. J.* **133**, 709–724 (1973).
55. Sottrup-Jensen, L.  $\alpha$ -Macroglobulins: structure, shape, and mechanism of proteinase complex formation. *J. Biol. Chem.* **264**, 11539–11542 (1989).
56. Marrero, A. *et al.* The crystal structure of human  $\alpha$ <sub>2</sub>-macroglobulin reveals a unique molecular cage. *Angew. Chem. Int. Ed.* **51**, 3340–3344 (2012).
57. Schlunegger, M. P. & Grütter, M. G. An unusual feature revealed by the crystal structure at 2.2 Å resolution of human transforming growth factor- $\beta$ 2. *Nature* **358**, 430–434 (1992).
58. Daopin, S., Piez, K. A., Ogawa, Y. & Davies, D. R. Crystal structure of transforming growth factor- $\beta$  2: an unusual fold for the superfamily. *Science* **257**, 369–373 (1992).
59. Kim, S. K. *et al.* An engineered transforming growth factor beta (TGF- $\beta$ ) monomer that functions as a dominant negative to block TGF- $\beta$  signaling. *J. Biol. Chem.* **292**, 7173–7188 (2017).
60. Janin, J. & Rodier, F. Protein-protein interaction at crystal contacts. *Proteins* **23**, 580–587 (1995).
61. Mittl, P. R. *et al.* The crystal structure of TGF- $\beta$ 3 and comparison to TGF- $\beta$ 2: implications for receptor binding. *Protein Sci.* **5**, 1261–1271 (1996).
62. Radaev, S. *et al.* Ternary complex of transforming growth factor- $\beta$ 1 reveals isoform-specific ligand recognition and receptor recruitment in the superfamily. *J. Biol. Chem.* **285**, 14806–14814 (2010).
63. Hart, P. J. *et al.* Crystal structure of the human T $\beta$ R2 ectodomain-TGF- $\beta$ 3 complex. *Nat. Struct. Biol.* **9**, 203–208 (2002).
64. Shi, M. *et al.* Latent TGF- $\beta$  structure and activation. *Nature* **474**, 343–349 (2011).
65. Zhao, B., Xu, S., Dong, X., Lu, C. & Springer, T. A. Prodomain-growth factor swapping in the structure of pro-TGF- $\beta$ 1. *J. Biol. Chem.* **293**, 1579–1589 (2018).
66. Lienart, S. *et al.* Structural basis of latent TGF- $\beta$ 1 presentation and activation by GARP on human regulatory T cells. *Science* **362**, 952–956 (2018).
67. Harrington, A. E. *et al.* Structural basis for the inhibition of activin signalling by follistatin. *EMBO J.* **25**, 1035–1045 (2006).
68. Wang, X., Fischer, G. & Hyvonen, M. Structure and activation of pro-activin A. *Nat. Commun.* **7**, 12052 (2016).

69. Saito, T. *et al.* Structural basis of the human endoglin-BMP9 interaction: insights into BMP signaling and HHT1. *Cell Rep.* **19**, 1917–1928 (2017).
70. Mi, L.-Z. *et al.* Structure of bone morphogenetic protein 9 procomplex. *Proc. Natl. Acad. Sci. USA* **112**, 3710–3715 (2015).
71. Goulas, T., Garcia-Ferrer, I., García-Piqué, S., Sottrup-Jensen, L. & Gomis-Rüth, F. X. Crystallization and preliminary X-ray diffraction analysis of eukaryotic  $\alpha_2$ -macroglobulin family members modified by methylamine, proteases and glycosidases. *Mol. Oral Microbiol.* **29**, 354–364 (2014).
72. Juanhuix, J. *et al.* Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron Radiat.* **21**, 679–689 (2014).
73. Kabsch, W. XDS. *Acta Crystallogr. sect. D* **66**, 125–132 (2010).
74. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. sect. D* **66**, 133–144 (2010).
75. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. sect. D* **67**, 235–242 (2011).
76. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
77. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
78. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
79. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. sect. D* **66**, 486–501 (2010).
80. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. sect. D* **68**, 352–367 (2012).
81. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. sect. D* **68**, 368–380 (2012).
82. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
83. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. sect. D* **60**, 2256–2268 (2004).
84. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
85. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980–980 (2003).
86. de la Cruz, M. J. *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 (2017).

## Acknowledgements

We are grateful to Joan Pous, Roman Bonet and Xandra Kreplin from the joint IBMB/IRB Automated Crystallography Platform for assistance during crystallization experiments and to Tibisay Guevara for assistance during peptide-mass fingerprint sample preparation. We further acknowledge the help of local contacts from the ALBA synchrotron. This study was supported in part by grants from Spanish and Catalan public and private bodies (grant/fellowship references BFU2015-64487R; MDM-2014-0435; JCI-2012-13573; BES-2015-074583; BES-2013-064651; 2017SGR3; and Fundació “La Marató de TV3” 201815). The Structural Biology Unit ([www.sbu.csic.es](http://www.sbu.csic.es)) of IBMB is a “María de Maeztu” Unit of Excellence from the Spanish Ministry of Science, Innovation and Universities.

## Author Contributions

F.X.G.R. and T.H.G. conceived and supervised the work; L.d.A.M. and L.M.P. produced and purified the protein; L.M.P., L.d.A.M. and F.X.G.R. performed crystallization and structural studies; and F.X.G.R. wrote the paper with contributions from all authors.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-44943-4>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.




**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019





# SCIENTIFIC REPORTS



OPEN

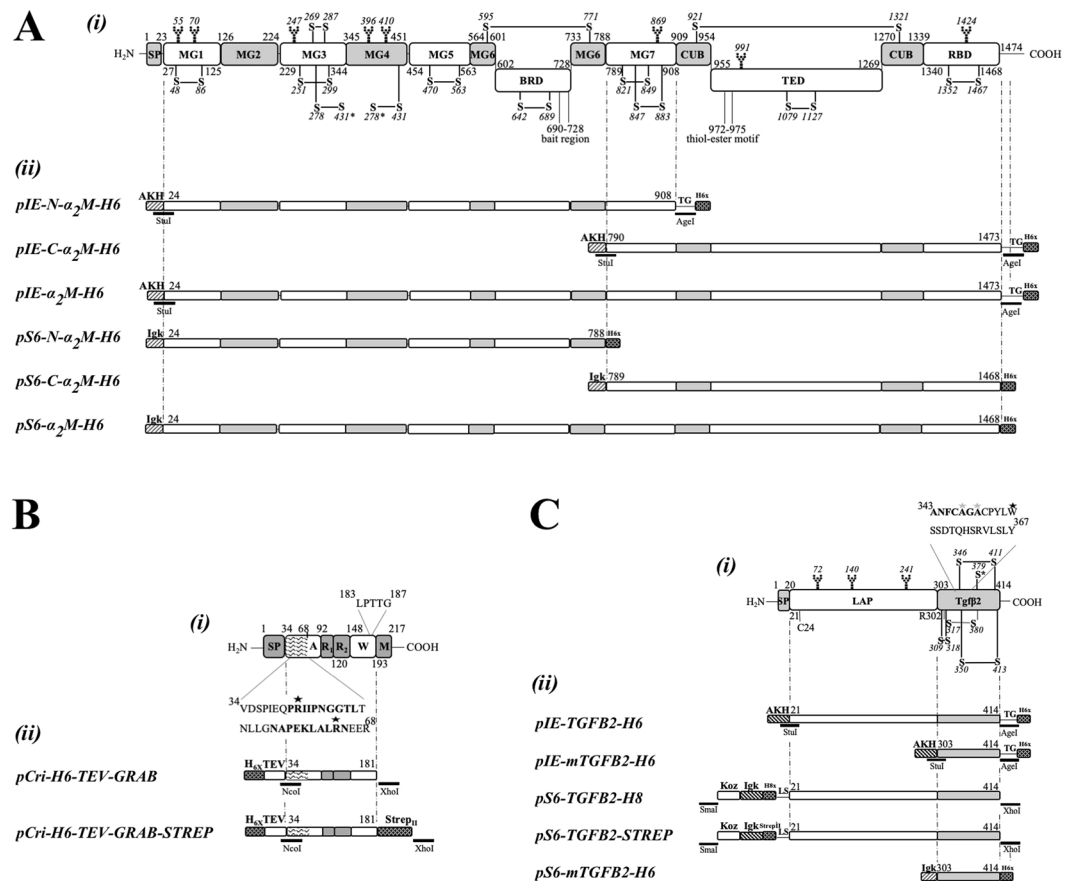
## Recombinant production of human $\alpha_2$ -macroglobulin variants and interaction studies with recombinant G-related $\alpha_2$ -macroglobulin binding protein and latent transforming growth factor- $\beta_2$

Laura Marino-Puertas<sup>1</sup>, Laura del Amo-Maestro<sup>1</sup>, Marta Taulés<sup>2</sup>, F. Xavier Gomis-Rüth  & Theodoros Goulas<sup>1</sup>

$\alpha_2$ -Macroglobulins ( $\alpha_2$ Ms) regulate peptidases, hormones and cytokines. Mediated by peptidase cleavage, they transit between native, intact forms and activated, induced forms.  $\alpha_2$ Ms have been studied over decades using authentic material from primary sources, which was limited by sample heterogeneity and contaminants. Here, we developed high-yield expression systems based on transient transfection in *Drosophila* Schneider 2 and human Expi293F cells, which produced pure human  $\alpha_2$ M ( $h\alpha_2$ M) at ~1.0 and ~0.4 mg per liter of cell culture, respectively. In both cases,  $h\alpha_2$ M was mainly found in the induced form. Shorter  $h\alpha_2$ M variants encompassing N-/C-terminal parts were also expressed and yielded pure material at ~1.6/~1.3 and ~3.2/~4.6 mg per liter of insect or mammalian cell culture, respectively. We then analyzed the binding of recombinant and authentic  $h\alpha_2$ M to recombinant latent human transforming growth factor- $\beta_2$  (pro-TGF- $\beta_2$ ) and bacterial G-related  $\alpha_2$ M binding protein (GRAB) by surface plasmon resonance, multiple-angle laser light scattering, size-exclusion chromatography, fluorogenic labelling, gel electrophoresis and Western-blot analysis. Two GRAB molecules formed stable complexes of high affinity with native and induced authentic  $h\alpha_2$ M tetramers. The shorter recombinant  $h\alpha_2$ M variants interacted after preincubation only. In contrast, pro-TGF- $\beta_2$  did not interact, probably owing to hindrance by the N-terminal latency-associated protein of the cytokine.

$\alpha_2$ -Macroglobulins ( $\alpha_2$ Ms) are large protein inhibitors, which counteract a broad spectrum of endopeptidases. To date, they have been characterized from metazoans and Gram-negative bacteria<sup>1-4</sup>. They are multi-domain molecular traps with comparable structural and biochemical properties, which present related modes of action termed “Venus flytrap” and “snap-trap” mechanisms<sup>5,6</sup>. In both cases, peptidases cut native  $\alpha_2$ M in a highly flexible bait region, which triggers a massive conformational rearrangement that induces the inhibitor and entraps the peptidase. In some family members, a second event involves a highly reactive  $\beta$ -cysteinyl- $\gamma$ -glutaminy l thioester bond, which is activated by nucleophiles such as lysines and covalently binds the prey peptidase, thus contributing to the stabilization of the enzyme:inhibitor complex. Trapped peptidases are still active but only against small substrates due to steric hindrance<sup>7</sup>. Hence,  $\alpha_2$ Ms regulate proteolysis in complex biological processes such as

<sup>1</sup>Proteolysis Laboratory, Structural Biology Unit (“Maria de Maeztu” Unit of Excellence), Molecular Biology Institute of Barcelona, Higher Scientific Research Council (CSIC), Barcelona Science Park, Helix Building; Baldiri Reixac, 15-21, 08028, Barcelona, Catalonia, Spain. <sup>2</sup>Scientific and Technological Centers (CCITUB), University of Barcelona, Lluís Solé i Sabaris, 1-3, 08028, Barcelona, Catalonia, Spain. Correspondence and requests for materials should be addressed to F.X.G.-R. (email: [xgrcri@ibmb.csic.es](mailto:xgrcri@ibmb.csic.es)) or T.G. (email: [thgcri@ibmb.csic.es](mailto:thgcri@ibmb.csic.es))



**Figure 1.** Overview of studied proteins. **(A)** Scheme depicting the domain structure of h $\alpha_2$ M (i) and the constructs studied (ii). The residue numbers correspond to UP P01023. (i) Functional regions and domains are the signal peptide (SP); macroglobulin domains 1-to-7 (MG1-MG7); the bait-region domain (BRD); the CUB domain; the thioester domain (TED); and the receptor-binding domain (RBD). Disulfide bonds are shown in black and linked cysteines are labelled. An interchain disulfide is pinpointed with an asterisk and N-linked glycosylation sites are highlighted with a sugar chain. (ii) h $\alpha_2$ M fusion proteins produced with plasmids pIEx and pCMV-Sport6. The AKH signal peptide sequence, the mouse Ig $\kappa$ -chain leader sequence, His<sub>6</sub>-tags and restriction sites are graphically represented. **(B)** Same as (A) for GRAB (UP Q7DAL7). (i) Functional regions and domains are the SP; domain A, with the binding regions of h $\alpha_2$ M hatched and in the inset; repeat regions R<sub>1</sub> and R<sub>2</sub>; the cell-wall attachment site (W), with the cell-wall anchor motif shown in magnification; and the membrane anchor (M). Critical arginine residues for h $\alpha_2$ M-binding are indicated by a star (R<sup>42</sup> and R<sup>64</sup>). (ii) GRAB fusion proteins in pCri8a with His<sub>6</sub>-tag, TEV site and Strep-tag. **(C)** Same as (A) for pro-TGF- $\beta_2$  (UP P61812). (i) Functional domains and regions are the SP; the latency associated peptide (LAP); and the mature growth factor moiety (TGF- $\beta_2$ ). Critical residues in LAP are C<sup>24</sup>, which is involved in the binding of LTBP, and R<sup>302</sup>, required for furin cleavage. Mature TGF- $\beta_2$  segment A<sup>343</sup>-Y<sup>367</sup> is involved in h $\alpha_2$ M binding, important and critical residues are indicated by a grey (A<sup>347</sup> and A<sup>349</sup>) and a black star (W<sup>354</sup>), respectively. (ii) Human TGF- $\beta_2$  fusion proteins produced with plasmids pIEx and pCMV-Sport6. The AKH signal peptide sequence, the Kozac (Koz), the mouse Ig $\kappa$ -chain leader sequence, affinity tags (His<sub>6</sub> and Strep) and restriction sites are graphically represented.

digestion, blood homeostasis, signaling, tissue remodeling and defense against toxins and other virulence factors during infection and envenomation<sup>1</sup>.

In addition to peptidase binding and inhibition,  $\alpha_2$ M<sub>s</sub> regulate several other endogenous and exogenous proteins (for a complete list, see<sup>1</sup> and references therein). Indeed, eukaryotic  $\alpha_2$ M<sub>s</sub> modify and modulate the activity of cytokines, hormones, growth factors, lipid factors and other proteins, and thus have a great impact on human physiology. A characteristic example is the interaction of human  $\alpha_2$ M (h $\alpha_2$ M), a 1,474-residue tetrameric multidomain protein (Fig. 1A), with transforming growth factors- $\beta$  (TGF- $\beta$ s), a family of ~25-kDa structurally homologous dimeric proteins (Fig. 1C). In mammals, the TGF- $\beta$  family has three members (TGF- $\beta_1$ , TGF- $\beta_2$  and TGF- $\beta_3$ ), which share 70% sequence identity and similar three-dimensional structures<sup>8</sup>. Their biological activity includes growth regulation, transcriptional activation of extracellular-matrix-related genes and chemotactic activity<sup>9,10</sup>. They are primarily regulated by the non-covalently attached N-terminal latency-associated domain (LAP)<sup>11</sup>, which acts as a pro-domain in the latent ~100-kDa pro-forms (pro-TGF- $\beta$ s). Once in circulation, LAP is removed and TGF- $\beta$  availability is regulated by h $\alpha_2$ M, which sequesters most of these cytokines

through a currently unknown mechanism<sup>10,12,13</sup>. What is known is that h $\alpha_2$ M positions E<sup>753</sup>, E<sup>737</sup> and D<sup>742</sup> within segment V<sup>723</sup>-T<sup>761</sup> (numbering according to UniProt [UP] entry P01023) are involved in TGF- $\beta_1$  binding<sup>14</sup> and that induced h $\alpha_2$ M binds the cytokine with higher affinity than the native inhibitor<sup>14</sup>.

The functional and structural properties of h $\alpha_2$ M are exploited by pathogens such as *Streptococcus pyogenes* (group A streptococci), which forms stable interactions with h $\alpha_2$ M by a surface protein, the G-related  $\alpha_2$ M-binding protein (GRAB<sup>15,16</sup>). This 23-kDa protein consists of a Gram-positive membrane anchor motif, a variable number of 28-residue repeats, and a highly-conserved N-terminal domain responsible for the interaction with h $\alpha_2$ M (Fig. 1B). By recruiting native h $\alpha_2$ M to the membrane, GRAB provides *S. pyogenes* with a mechanism to inhibit host peptidases, which protects bacterial surface structures and facilitates progressive dissemination in the infected tissue<sup>15</sup>.

These interactions have only been preliminary characterized<sup>17,18</sup>, and the mechanisms are still unknown. To shed light on them, we developed eukaryotic expression systems of h $\alpha_2$ M variants and purified the authentic protein from blood. We further used these proteins to study complex formation with GRAB and pro-TGF- $\beta_2$  by several biophysical approaches.

## Materials and Methods

**Construct preparation.** Constructs spanning fragments of the gene coding for h $\alpha_2$ M, namely full-length h $\alpha_2$ M and its N- and C-terminal parts (N-h $\alpha_2$ M and C-h $\alpha_2$ M; for details on constructs, plasmids, vectors and primers, see Table 1 and Fig. 1), and the coding sequence for GRAB from *Streptococcus pyogenes* serotype M1 (UP Q7DAL7) were amplified with primers that introduced either restriction sites for directional cloning or overhangs for restriction-free cloning. The vectors used were pCri-8a<sup>19</sup> for bacterial expression, pLex (Novagen) for expression in *Drosophila melanogaster* Schneider 2 embryonic cells (S2; Gibco), and pCMV-Sport 6 (Thermo Scientific) for expression in human Expi293F<sup>TM</sup> cells (Gibco). Polymerase chain reaction (PCR) primers and DNA modifying enzymes were purchased from Sigma-Aldrich and Thermo Scientific, respectively. PCR was performed using Phusion High Fidelity DNA polymerase (Thermo Scientific) according to the manufacturer's instructions and following a standard optimization step by thermal gradient in each reaction. Mutants were generated by a modified version of the previously described procedure<sup>20</sup>. DNA was purified with the OMEGA Biotek Purification Kit according to the manufacturer's instructions, and all constructs were verified by DNA sequencing.

**Cell-culture media.** S2 and Expi293F cells were adapted to grow in suspension in Sf-900<sup>TM</sup> II SFM culture medium (Gibco) and FreeStyle<sup>TM</sup> F17 expression medium (Gibco) with 0.2% Pluronic F-68 (Gibco) plus 8 mM L-glutamine (Gibco), respectively. Both growth media were supplemented with 0.5  $\mu$ g/mL of the antimycotic Fungizone, 100 units/mL of penicillin, and 100  $\mu$ g/mL of streptomycin sulfate (Gibco).

**Cell-culture growth.** S2 cells were cultivated in TubeSpin bioreactor tubes (TS50 for 5-to-10-mL cultures and TS600 for 100-to-200-mL cultures; Techno Plastic Products AG) as previously described<sup>21</sup>. Cells were passaged three times per week to a final density of  $4 \times 10^6$  cells/mL. The cultures were incubated at 28 °C in a shaker (Brunswick Scientific Innova) under agitation at 220 rpm.

Expi293F cells were cultivated in 125-mL or 1000-mL polycarbonate Erlenmeyer flasks (FPC0125S and FPC1000S, respectively; Tri Forest Labware) for 25-to-30-mL and 100-to-250-mL cultures, respectively. Cells were subcultured three times per week to a final density of  $0.3$ – $0.5 \times 10^6$  cells/mL and kept in suspension at 150 rpm in a Multitron Cell Shaker Incubator (Infors HT) at 37 °C in a modified atmosphere (8% CO<sub>2</sub> and 85% relative humidity). Cell densities and viability were determined by the trypan blue exclusion test<sup>22</sup>.

**Cell-culture transfection.** Linear 25-kDa polyethylenimine (PEI; Polysciences Europe GmbH) was prepared in Milli-Q water at a concentration of 1 mg/mL and pH 7.0. The solution was filter-sterilized and stored at –20 °C. Plasmid DNA was produced in *Escherichia coli* DH5 $\alpha$  cells, purified with the GeneJET Plasmid Maxiprep Kit (Thermo Scientific), and stored at –20 °C in sterile Milli-Q water at 1 mg/mL.

For transfection, S2 cells were centrifuged and resuspended in prewarmed fresh medium to a cell density of  $15 \times 10^6$  cells/mL. A mixture of 0.6  $\mu$ g DNA (see Fig. 1 and Tables 1) and 2  $\mu$ g PEI per  $1 \times 10^6$  cells and per prewarmed transfection volume was pre-incubated for 15–20 min at room temperature and then added dropwise to the cell cultures. These were further incubated for 1 hour at 28 °C and 220 rpm, subsequently diluted with prewarmed fresh medium to  $5 \times 10^6$  cells/mL and harvested after seven days for protein purification.

For mammalian cultures, Expi293F cells were transfected at a cell density of  $1 \times 10^6$  cells/mL with a mixture of 1 mg of DNA (see Table 1) and 3 mg of PEI in 20 mL of Opti-MEM Medium (Gibco) per liter of expression medium. The DNA-PEI mixture was incubated at room temperature for 15–20 min and then added dropwise to the cell cultures, which were harvested after three days for protein purification.

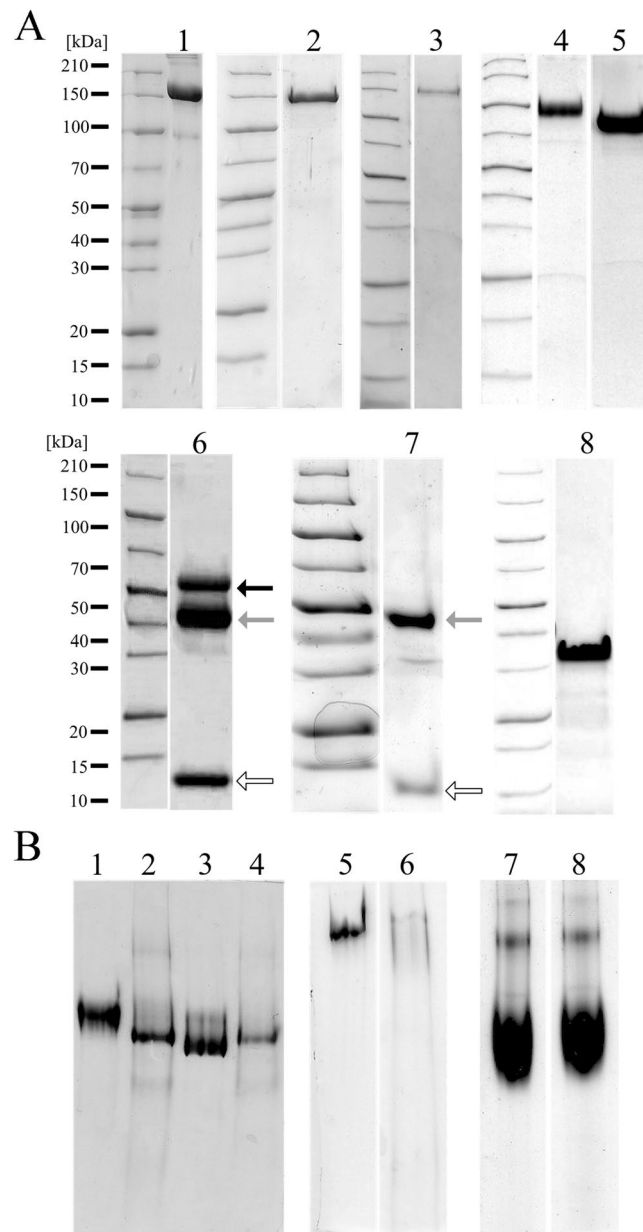
**Bacterial expression.** For the recombinant overexpression of N-terminally hexa-histidine (His<sub>6</sub>)-tagged GRAB with a tobacco-etch virus peptidase (TEV) recognition sequence, with or without an additional C-terminal Strep-tag<sup>®</sup> II tag (Strep-tag; IBA Life Sciences), plasmid pCRI8a<sup>19</sup> was transformed into *E. coli* BL21 (DE3) cells (Novagen<sup>23</sup>), and cultures were grown in lysogeny broth supplemented with 30  $\mu$ g/mL kanamycin. After initial growth at 37 °C to an  $OD_{600} \approx 0.6$ , cultures were cooled to 20 °C, and protein expression was induced with 0.4 mM isopropyl- $\beta$ -D-thiogalactopyranoside for 18–20 hours.

**Protein purification.** Protein purification steps were performed at 4 °C if not otherwise stated. For GRAB purification, bacterial cells were collected by centrifugation at  $6,000 \times g$  for 30 min, washed in buffer A (50 mM Tris-HCl, 250 mM sodium chloride, pH 7.5) and resuspended in the same buffer plus 20 mM imidazole, Complete EDTA-free Peptidase Inhibitor Cocktail Tablets and DNase I (both from Roche Diagnostics). Cells were lysed with a cell disrupter (Constant Systems) at a pressure of 1.35 kbar, cell debris was removed by centrifugation at



Plasmid name	Protein	Parental DNA	Forward-primer*	Reverse-primer*	Protein sequence**	Tags***	Comments
<i>pIE-ho<sub>2</sub>M-H6</i>	ho <sub>2</sub> M	Human c-DNA pIEx vector	CATTAGGCCTC AGTCTCTGGAAA ACCGCAGTATATG	CATTACCGGTA GGATTTC AAGATCTTTG	S <sup>24</sup> -N <sup>1473</sup> + PTG + H <sub>6x</sub>	C-t H <sub>6x</sub>	Full-length ho <sub>2</sub> M in S2 cells. The gene was inserted by directional cloning (between <i>StuI</i> and <i>AgeI</i> ) into the pIEx vector in frame with the AKH signal peptide sequence.
<i>pIE-N-ho<sub>2</sub>M-H6</i>	N-ho <sub>2</sub> M	pIE-ho <sub>2</sub> M-H6	GATCTTGAAAT CCTACCGGTCAT CATCAC	CAATACCGGT TTCAGGTT AACCAACAGAG	S <sup>24</sup> -E <sup>908</sup> + TG + H <sub>6x</sub>	C-t H <sub>6x</sub>	As above for the N-terminal half of ho <sub>2</sub> M.
<i>pIE-C-ho<sub>2</sub>M-H6</i>	C-ho <sub>2</sub> M	pIE-ho <sub>2</sub> M-H6	CAATAGGCCTCA CAGCCCTTCTT TGTGGAGCTC	CAATAGGCCT CAGCGATGA TGACGAAAG	Q <sup>790</sup> -N <sup>1473</sup> + TG + H <sub>6x</sub>	C-t H <sub>6x</sub>	As above for the C-terminal half of ho <sub>2</sub> M.
<i>pS6-ho<sub>2</sub>M-H6</i>	ho <sub>2</sub> M	pIE-ho <sub>2</sub> M-H6 pCMV-Sport 6 vector	TGGGTTCAGGTT CCACTGGTGACTCAGTC TCTGGAAAACCGCAGTAT	CGCCTAATGGTGAT GGTGAATGGTGGCTGCA AGGAGCATTGTACTCAGC	S <sup>24</sup> -S <sup>1468</sup> + H <sub>6x</sub>	C-t H <sub>6x</sub>	Full-length ho <sub>2</sub> M in Expi293F cells. The gene was inserted by restriction-free cloning into the pCMV-Sport 6 vector in frame with the Ig κ leader sequence.
<i>pS6-N-ho<sub>2</sub>M-H6</i>	N-ho <sub>2</sub> M	pIE-N-ho <sub>2</sub> M-H6 pCMV-Sport 6 vector	TGGGTTCAGGTT CCACTGGTGACTCAGTC TCTGGAAAACCGCAGTAT	CGCCTAATGGT GATGGTGAATGGTGG CTCGAGAGA GGCAGTGAAGA	S <sup>24</sup> -A <sup>788</sup> + H <sub>6x</sub>	C-t H <sub>6x</sub>	As above for the N-terminal half of ho <sub>2</sub> M.
<i>pS6-C-ho<sub>2</sub>M-H6</i>	C-ho <sub>2</sub> M	pIE-ho <sub>2</sub> M-H6 pCMV-Sport 6 vector	TGGGTTCAGGTTCCACT GGTGAATGGTGGCTGCA CCTTCTTTGTGGAGCTC	CGCCTAATGGT GGTGAATGGTGG CTGCAAGGAGC ATTGTACTCAGC	F <sup>789</sup> -S <sup>1468</sup> + H <sub>6x</sub>	C-t H <sub>6x</sub>	As above for the C-terminal half of ho <sub>2</sub> M.
<i>pCri-H6-TEV-GRAB</i>	GRAB	Synthetic DNA pCri8a vector	CAATCCATGGTT GATAGCCCGATTG	CAATCTCGAGT TAATTAACGT TCTGACGTT	GAM + V <sup>34</sup> -N <sup>181</sup>	N-t H <sub>6x</sub> + TEV	Synthetic gene of GRAB optimized for expression in <i>Escherichia coli</i> inserted into the pCri8a vector <sup>19</sup> by directional cloning between the <i>NcoI</i> and <i>XhoI</i> restriction sites.
<i>pCri-H6-TEV-GRAB-STREP</i>	GRAB	pCri-H6-TEV-GRAB	ATGCCCATGGT TGATAGCCCG	CGAATTGTGGATGGCTC CAACCTCCATTAACGT TCTGACGTTCC; CTTCCACCTCCAGA ACCTCCACCCCTTTTC GAATTGTGGATGGCTCC; GTGGATGGCTCCATGCG CTACTCCACTTCCACT CCAGAACCC; GCATCTCGAG TTACTTTT CGAATTGT GGATGGCTC CATGGC	GAM + V <sup>34</sup> -N <sup>181</sup> + G GWSHPQFEKGGG SGGSGGSAWS HPQFEK	N-t H <sub>6x</sub> + TEV- (protein)- Strep	This construct was obtained from pCri-H6-TEV-GRAB by four consecutive PCR reactions to introduce a C-terminal Strep-tag.
<i>pIE-TGFB2-H6</i>	pro-TGF-β <sub>2</sub>	Human c-DNA pIEx vector	CAATAGGCCTT GTCTACCTG CAGCACACTC	CAATACCGGTGC TGCATTTGCAAG ACTTTAC	L <sup>21</sup> -S <sup>414</sup> + TG + H <sub>6x</sub>	C-t H <sub>6x</sub>	Pro-TGF-β <sub>2</sub> in S2 cells. The gene was inserted by directional cloning (between <i>StuI</i> and <i>AgeI</i> ) into the pIEx vector in frame with AKH signal peptide sequence
<i>pIE-mTGFB2-H6</i>	TGF-β <sub>2</sub>	Human c-DNA pIEx vector	CAATAGGCCTC AGCTTTGGAT CGGCCCTATTG	CAATACCGGTG CTGCATTTGCA AGACTTTAC	A <sup>303</sup> -S <sup>414</sup> + TG + H <sub>6x</sub>	C-t H <sub>6x</sub>	As above for mature TGF-β <sub>2</sub> .
<i>pS6-TGFB2-H8</i>	pro-TGF-β <sub>2</sub>	pIE-TGFB2-H6 pCMV-Sport 6 vector	TCACCACCACATCATCTCA GCCTGTCTACCTGCAGCA; GGTTCACCTGGTGACCACC ACCATCACCACCACCATC; GGGTACTGCTGCTCTGGGTT CCAGGTTCCACTGGTGAC; GACAGACACACTCCTGCT ATGGGTACTGTGCTC; CAATCCCGGGGCCACCAT GGAGACAGACACTCC	CAATCTCGAGCT AGCTGCATTTGC AAGACTTTAC	H <sub>8x</sub> + LS + L <sup>21</sup> -S <sup>414</sup>	N-t H <sub>8x</sub>	Pro-TGF-β <sub>2</sub> in Expi293F cells. See <sup>8</sup> for details.
<i>pS6-TGFB2-STREP</i>	pro-TGF-β <sub>2</sub>	pS6-TGFB2-H8	GGTGGAGG TTCTGGAG GTGGAAGT GGAGGTAGCG CATGGAGCC ATCCACAA TTGCAA AAGCTCA GCCTGTC TACCTGC	CTTTTCGAATTGTG GATGGCTCCAGTCA CCAGTGGAAACCTGGA ACCCAGAGCAG	WSHPQFEK GGSGGGG GGSAWSHPQ FEKLS + L <sup>21</sup> -S <sup>414</sup>	N-t Strep	Pro-TGF-β <sub>2</sub> in Expi293F cells. The parental plasmid was modified by opposite primers to replace the N-terminal histidine-tag with a Strep-tag. See <sup>8</sup> for details.
<i>pS6-mTGFB2-H6</i>	TGF-β <sub>2</sub>	pIE-mTGFB2-H6 pCMV-Sport 6 vector	GTTCAGGTTCC ACTGGTGAAGC CTTTGGATG CGGCCTATTGC	CCTAATGGT GATGGTGAAT GGTGGCTGC ATTTGCAAGA CTTTACA	A <sup>303</sup> -S <sup>414</sup> + H <sub>6x</sub>	C-t H <sub>6x</sub>	Mature TGF-β <sub>2</sub> in Expi293F cells. The coding gene extracted from the parental plasmid was inserted into the pCMV-Sport 6 vector by restriction-free cloning between the Ig κ leader sequence and the C-terminal histidine-tag.

**Table 1.** Constructs, primers, plasmids and proteins. All constructs are for extracellular expression of the respective proteins. \*Restriction-site sequences and overhangs for restriction-free cloning are underlined. \*\*Peptide sequence of the expressed protein after fusion-tag removal. Amino acids derived from the construct are in bold. See also Fig. 1. \*\*\*Fused tags at the carboxy-terminus (C-t) or the amino-terminus (N-t). AKH, adipokinetic hormone; TEV, tobacco-etch virus peptidase; Ig κ, immunoglobulin κ.



**Figure 2.** Recombinant protein production and purification. **(A)** SDS-PAGE analysis of wild-type and recombinant proteins. Lanes: 1, native authentic h $\alpha_2$ M; 2, recombinant h $\alpha_2$ M from S2 cells; 3, recombinant h $\alpha_2$ M from Expi293F cells; 4, N-terminal half of h $\alpha_2$ M (N-h $\alpha_2$ M); 5, C-terminal half of h $\alpha_2$ M (C-h $\alpha_2$ M); 6, pro-TGF- $\beta_2$  produced in Expi293F cells according to<sup>8</sup>. Arrows indicate pro-TGF- $\beta_2$  (black), LAP (grey) and mature TGF- $\beta_2$  (white); 7, pro-TGF- $\beta_2$  digested by furin; 8, GRAB. **(B)** Native-PAGE analysis of wild-type and recombinant proteins. Lanes: 1 and 3, native and methylamine-induced authentic h $\alpha_2$ M; 2 and 4, native and methylamine-induced recombinant h $\alpha_2$ M from S2 cells; 5 and 6, native authentic h $\alpha_2$ M and recombinant h $\alpha_2$ M from Expi293F cells; 7 and 8, native and induced recombinant C- $\alpha_2$ M expressed from Expi293F cells.

30,000  $\times$  g for 1 hour, and the supernatant containing GRAB was kept for subsequent purification steps. For the h $\alpha_2$ M variants produced in S2 and Expi293F systems, cells were removed by centrifugation at 2,800  $\times$  g for 20 min and the supernatant was used for subsequent purification steps.

Supernatants containing the proteins of interest were incubated for 20 min (expression in insect cells) or 1 hour (expression in mammalian cells) with nickel-nitrilotriacetic acid resin (Ni-NTA; Invitrogen), which was subsequently loaded onto an open column for batch purification (Bio-Rad), washed extensively with buffer A plus 20 mM imidazole, and eluted with buffer A plus 300 mM imidazole (direct Ni-NTA). For GRAB, eluted samples were then dialyzed overnight against buffer A plus 1 mM 1,4-dithio-DL-threitol (DTT) in the presence of His<sub>6</sub>-tagged TEV at a peptidase:protein weight ratio of 1:100 and 1 mM DTT. The resulting cleavage left additional residues (glycine-alanine-methionine) at the N-terminus of the target proteins due to the cloning strategy (see Table 1). Digested samples were passed several times through Ni-NTA resin previously equilibrated with buffer A plus 20 mM imidazole to remove His<sub>6</sub>-tagged molecules and the flow-through containing untagged GRAB was collected (reverse Ni-NTA).

Protein sample	Molecular mass (kDa)
Native h $\alpha_2$ M	680.6 $\pm$ 1.8
Native h $\alpha_2$ M + GRAB	707.8 $\pm$ 3.4
Induced h $\alpha_2$ M	684.1 $\pm$ 2.7
Induced h $\alpha_2$ M + GRAB	710.6 $\pm$ 1.5
GRAB	15.5 $\pm$ 0.0
pro-TGF- $\beta_2$	105.4 $\pm$ 0.6

**Table 2.** Molecular masses determined by SEC-MALLS. Values are represented as means and standard deviations of three replicates.

In all cases, proteins eluted from direct and reverse Ni-NTA chromatographies were dialyzed overnight against buffer B (20 mM Tris-HCl, 5 mM sodium chloride, pH 7.5) and further purified by ionic-exchange chromatography (IEC) on a TSKgel DEAE-2SW column (TOSOH Bioscience) equilibrated with buffer B. A gradient of 2–30% buffer C (20 mM Tris-HCl, 1 M sodium chloride, pH 7.5) was applied over 30 mL, and samples were collected and pooled. Subsequently, each pool was concentrated by ultrafiltration and subjected to size-exclusion chromatography (SEC) in Superdex 75 10/300 (GRAB and pro-TGF- $\beta_2$ ), Superdex 200 10/300 (N-h $\alpha_2$ M and C-h $\alpha_2$ M) or Superose 6 10/300 (full-length recombinant h $\alpha_2$ M) columns (GE Healthcare Life Sciences) in buffer D (20 mM Tris-HCl, 150 mM sodium chloride, pH 7.5). Strep-tagged GRAB was purified by affinity chromatography with Streptactin<sup>®</sup>XT Superflow Suspension resin (IBA Life Sciences) and eluted with buffer E (100 mM Tris-HCl, 150 mM sodium chloride, pH 8.0) at a further 50 mM in biotin. IEC and SEC purification steps followed as above.

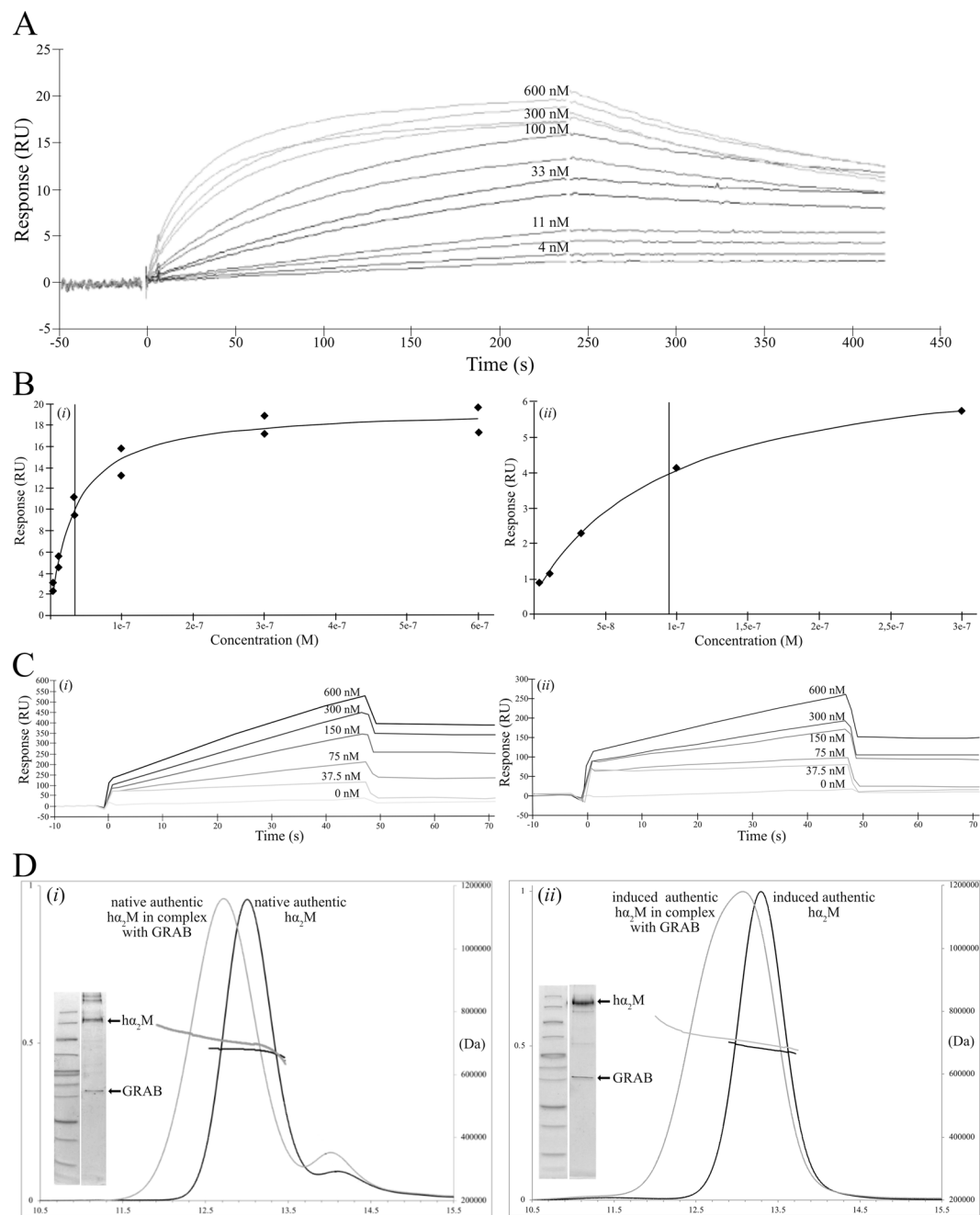
Authentic full-length h $\alpha_2$ M was isolated from blood plasma from individual donors and purified essentially as described previously<sup>17,24,25</sup>. Briefly, plasma was subjected to sequential precipitation steps with 4–12% PEG 4,000, and the final precipitate containing h $\alpha_2$ M was reconstituted in 20 mM sodium phosphate at pH 6.4. Partially purified h $\alpha_2$ M was captured with a zinc-chelating resin (G-Biosciences), washed with buffer F (50 mM sodium phosphate, 250 mM sodium chloride, pH 7.2) plus 10 mM imidazole and eluted in the same buffer plus 250 mM imidazole and 100 mM EDTA. The protein was first passed through a PD10 desalting column (GE Healthcare Life Sciences) previously equilibrated with 20 mM HEPES, pH 7.5 and then subjected to an IEC step in a Q Sepharose column (2.5  $\times$  10 cm; GE Healthcare Life Sciences), previously equilibrated with 15% buffer G (20 mM HEPES, 1 M sodium chloride, pH 7.5). A gradient of 20–30% buffer G was applied for 150 min and fractions were collected. Collected samples were dialyzed overnight against buffer H (20 mM sodium phosphate, 5 mM sodium chloride, pH 7.4) and further purified by IEC in a TSKgel DEAE-2SW column, previously equilibrated with buffer H. A gradient of 7–20% buffer I (20 mM sodium phosphate, 1 M sodium chloride, pH 7.4) was applied over 30 mL, and samples were collected and pooled. Subsequently, each pool was concentrated and subjected to a final polishing step by SEC in a Superose 6 10/300 column in buffer J (20 mM sodium phosphate, 150 mM sodium chloride, pH 7.4).

Protein identity and purity were assessed by 10–15% Tricine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE<sup>26</sup>) stained with Coomassie Brilliant Blue, peptide mass fingerprinting of tryptic protein digests, N-terminal sequencing through Edman degradation, and mass spectrometry. The latter three were carried out at the Protein Chemistry Service and the Proteomics Facilities of the Centro de Investigaciones Biológicas (Madrid, Spain). Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices of 10- to 50-kDa cut-off (Sartorius Stedim Biotech). Protein concentrations were estimated by measuring the absorbance at 280 nm in a spectrophotometer (NanoDrop) and applying the respective theoretical extinction coefficients. Concentrations were also measured by the BCA Protein Assay Kit (Thermo Scientific) with bovine serum albumin fraction V (BSA; Sigma-Aldrich) as a standard. Induced h $\alpha_2$ M was obtained by treating native h $\alpha_2$ M in buffer D with 200 mM methylamine hydrochloride for one hour at room temperature. Subsequently, the sample was dialyzed against the same buffer D.

Human pro-TGF- $\beta_2$  (UP P61812) constructs (see Table 1 and Fig. 1C) were produced in S2 and Expi293F cells and purified as reported elsewhere<sup>8</sup>. Production of mature TGF- $\beta_2$  with a C-terminal His<sub>6</sub>-tag (see Table 1) was assayed with the insect and human systems, which included harvesting periods of seven and three days, respectively. Supernatants were collected after the centrifugation at 2,800  $\times g$  for 20 min and the purification steps were, first a direct Ni-NTA in buffer A plus 20 mM imidazole for the wash step, and plus 300 mM imidazole for the elution; and finally purified by SEC with a Superdex 75 10/300 column in buffer D.

**Protein labeling.** GRAB and pro-TGF- $\beta_2$  were labelled with fluorogenic sulfo-succinimidyl-7-amino-4-methylcoumarin-3-acetate (Sulfo-NHS-AMCA; Thermo Scientific) according to the manufacturer's instructions with a 10–15 molar excess of reagent over protein in buffer J for 1 hour at room temperature. Thereafter, the proteins were extensively dialyzed against buffer J to remove non-reacted dye. To assess binding, labelled GRAB or pro-TGF- $\beta_2$  were mixed with authentic h $\alpha_2$ M (native and induced) or recombinant fragments N-h $\alpha_2$ M and C-h $\alpha_2$ M at a 4:1 molar ratio, incubated in buffer J for two hours at 37 °C, and analyzed by 10% native PAGE<sup>27</sup>. Gel fluorescence was visualized in a gel reader (G:BOX F3 Gel Doc System, Syngene) and the fluorescence was measured ( $\lambda_{ex}$  = 345–350 nm and  $\lambda_{em}$  = 440–460 nm). Negative controls (unlabeled proteins) were included in each experiment. After fluorescence detection, native gels were stained with Coomassie Brilliant Blue (Thermo Scientific) to detect the negative controls.

**Multi-angle laser light scattering.** Multi-angle laser light scattering in a Dawn Helios II apparatus (Wyatt Technologies) coupled to a SEC Superose 6 10/300 column (SEC-MALLS) equilibrated in buffer J at 25 °C was performed at the joint IBMB/IRB Crystallography Platform, Barcelona Science Park (Catalonia, Spain) to analyze binding of GRAB or pro-TGF- $\beta_2$  to native or induced authentic h $\alpha_2$ M at a molar ration of 4:1. ASTRA 7 software



**Figure 3.** Interaction of GRAB and pro-TGF- $\beta_2$  with  $h\alpha_2M$  variants. **(A,B)** Surface-plasmon resonance sensorgrams of the interaction of native or induced authentic  $h\alpha_2M$  with GRAB. Multi-cycle run for native  $h\alpha_2M$  with GRAB (A) and corresponding plot of the steady-state response (B, *i* and *ii*, for native and induced  $h\alpha_2M$ , respectively). Different  $h\alpha_2M$  concentrations were assayed to determine the rate constants that describe the kinetics and the equilibrium constants for complex strength (see also Tables 3 and 4). The vertical line in the plots of steady-state response indicates the value of the calculated equilibrium dissociation constant  $K_D$ . **(C)** Sensorgrams of the interaction of N- $h\alpha_2M$  (*i*) and C- $h\alpha_2M$  (*ii*) with GRAB. Proteins were premixed, incubated at 37°C for 1 h, injected over the chip, and the response was measured. **(D)** SEC-MALLS analysis of complex formation between GRAB and native (*left*) and induced (*right*) authentic  $h\alpha_2M$  showing the measured molecular mass distribution. Inserted figures within graphs show the SDS-PAGE analysis of the respective purified complexes.

(Wyatt Technologies) was used for data processing and analysis, for which a  $dn/dc$  value typical for proteins (0.185 mL/g) was assumed. All experiments were performed in triplicate.

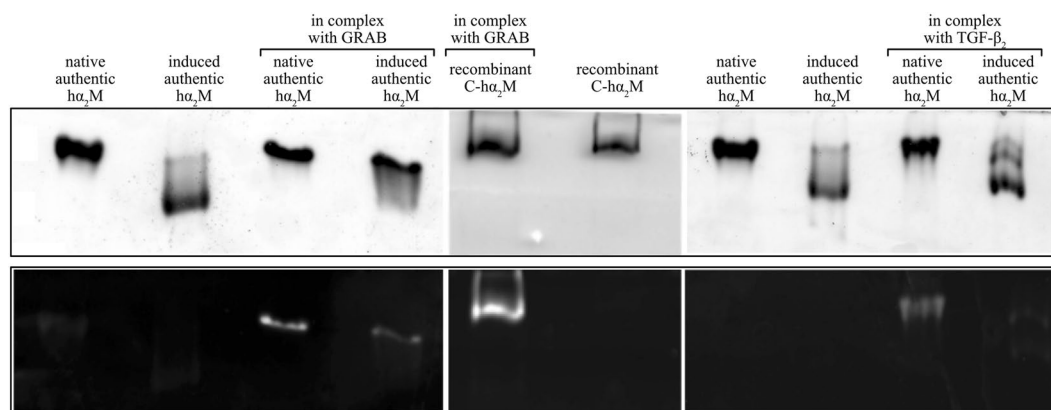
**Western blot analyses.** Protein samples were separated by 10% SDS-PAGE, transferred to Hybond ECL nitrocellulose membranes (GE Healthcare Life Sciences), and blocked for two hours under gentle stirring at room temperature with 50 mL of blocking solution (phosphate buffered saline; PBS) plus 0.1% Tween 20 and 5%

Protein sample	$k_a$ ( $M^{-1} s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$K_D$ (M)	$R_{max}$ (RU)	$\chi^2$
Native $h\alpha_2M$ + GRAB	$1.32 \times 10^{+5}$	$1.90 \times 10^{-3}$	$1.43 \times 10^{-8}$	18.51	1.23

**Table 3.** Kinetic rates and equilibrium constants of the interaction between native authentic  $h\alpha_2M$  and GRAB. Constants were calculated from the corresponding plot assuming a 1:1 interaction model (two GRAB molecules per  $h\alpha_2M$  dimer), see Fig. 3A;  $k_a$ , association rate constant;  $k_d$ , dissociation rate constant;  $K_D$ , equilibrium dissociation constant.

Protein sample	$K_D$ (M)	$R_{max}$ (RU)	$\chi^2$
Native $h\alpha_2M$ + GRAB	$3.45 \times 10^{-8}$	18.94	1.17
Induced $h\alpha_2M$ + GRAB	$9.46 \times 10^{-8}$	6.82	0.01

**Table 4.** Equilibrium constants of the interaction between native or induced authentic  $h\alpha_2M$  and GRAB. Values were derived from the corresponding plot of steady state response against concentration assuming a 1:1 model (one GRAB molecule per  $h\alpha_2M$  dimer), see Fig. 3B.



**Figure 4.** Analysis of complex formation between  $h\alpha_2M$  variants and GRAB or pro-TGF- $\beta_2$ . Complexes were separated by native-PAGE. GRAB or TGF- $\beta_2$  labelled with fluorogenic Sulfo-NHS-AMCA were visualized in a gel reader (lower panels) and then stained with Coomassie Brilliant Blue (upper panels).

BSA. His<sub>6</sub>-tagged proteins were detected by immunoblot analysis using the monoclonal His-HRP Conjugated Antibody (Santa Cruz Biotechnology) diluted 1:5,000 in PBS plus 0.1% Tween 20. Strep-tagged proteins were detected with the Streptavidin-Peroxidase Conjugated Antibody from *Streptomyces avidinii* (Sigma-Aldrich) diluted 1:1,000 in PBS plus 0.1% Tween 20 and 1% BSA. Complexes were detected using an enhanced chemiluminescence system (Super Signal West Pico Chemiluminescent; Pierce) according to the manufacturer's instructions. Membranes were exposed to Hyperfilm ECL films (GE Healthcare Life Sciences).

**Proteolytic inhibition assays.** Inhibition assays against protein substrates were performed in a microplate fluorimeter (Infinite M200, TECAN) in 200  $\mu$ L reaction volumes with the fluorescence-based EnzCheck Assay Kit containing BODIPY FL-casein ( $\lambda_{ex} = 505$  nm and  $\lambda_{em} = 513$  nm) as fluorescein conjugate (Invitrogen) at 10  $\mu$ g/mL in buffer D. Inhibition was measured after preincubation of a two-fold molar excess of authentic or recombinant  $h\alpha_2M$  with trypsin (0.25  $\mu$ g) for 15 min at room temperature. The substrate was added to the reaction mixture and the residual tryptic activity was measured over a period of two hours.

**Thiol quantification.** Detection of free sulfhydryl groups was performed with the Fluorometric Thiol Assay Kit (ab112158 assay; Abcam) following the manufacturer's instructions and using glutathione as a standard for the dose response curve. The fluorescent signal was measured in a microplate fluorimeter (Infinite M200, TECAN) at  $\lambda_{ex} = 490$  nm and  $\lambda_{em} = 520$  nm in 96-well plates containing 100  $\mu$ L reaction volumes (50  $\mu$ L of assay reaction mixture plus 50  $\mu$ L of glutathione-standard or test samples) in duplicate. Fluorescence was measured after preincubation of authentic  $h\alpha_2M$  (0.39  $\mu$ M) or C- $h\alpha_2M$  obtained from human cells (1.6  $\mu$ M), with or without treatment with methylamine for 10, 20, 30, 45 and 60 min, at room temperature.

**Surface plasmon resonance and kinetic data analysis.** The binding kinetics (association and dissociation) and affinity (complex formation at the equilibrium) of GRAB or pro-TGF- $\beta_2$  (ligands) with native authentic  $h\alpha_2M$ , induced authentic  $h\alpha_2M$ , recombinant N- $h\alpha_2M$  or recombinant C- $h\alpha_2M$  (analytes) were studied by surface plasmon resonance with a Biacore™ T200 Biosensor System (GE Healthcare Life Sciences) at the Scientific and Technological Centers of the University of Barcelona (Catalonia, Spain). To bind ligands provided with a Strep-tag, Streptactin®XT (IBA LifeSciences) was immobilized at 25 °C on the surface of the four flow cells of a sensor chip CM5 series S (GE Healthcare Life Sciences) at 3,000 response units (RU) through amine coupling, as



described previously<sup>22</sup>. Subsequently, Strep-tagged GRAB (at 9.7 nM) or pro-TGF- $\beta_2$  (at 19.0 nM) in HBNS buffer (10 mM HEPES, 150 mM sodium chloride, pH 7.4) were immobilized at low RU density on different flow cells of the chip by virtue of the strong interaction between the Strep-tag and streptactin at 5  $\mu$ L/min for 24 sec at 37 °C. To monitor association, the immobilized ligands were then exposed to the analytes at different concentrations in HBNS (4–600 nM for native and induced authentic h $\alpha_2$ M; 75–1,200 nM for N-h $\alpha_2$ M and C-h $\alpha_2$ M), which were injected at 30  $\mu$ L/min for 120–240 sec at 37 °C. Thereafter, HBNS was injected for analyte dissociation from the immobilized ligands for 90–300 sec. To dissociate bound ligands and regenerate the chip surface, 3 M guanidine hydrochloride was injected at 30  $\mu$ L/min for 30 sec after each cycle. These experiments were double referenced by keeping the first flow cell without ligand, and by an injection step at analyte concentration zero. The affinity analysis was performed by plotting binding responses in the steady-state region of the sensorgrams ( $R_{eq}$ ) against analyte concentrations to determine the overall equilibrium dissociation constant ( $K_D$ ). Sensorgrams were analyzed with the BIAEVALUATION program v. 3.0 (GE Healthcare Life Sciences) and fitted to a 1:1 Langmuir interaction model. The likelihood of fitting was assessed through the  $\chi^2$  statistical parameter<sup>28</sup>.

In a separate qualitative experiment, ligands GRAB (at 120 nM) and pro-TGF- $\beta_2$  (at 950 nM) were premixed with the analytes at different concentrations (2–150 nM for native and induced authentic h $\alpha_2$ M; 38–600 nM for N-h $\alpha_2$ M and C-h $\alpha_2$ M) and incubated for one hour at 37 °C. Subsequently, the mixtures were injected at 15  $\mu$ L/min at 37 °C according to a published multicycle method<sup>29</sup>. The binding was measured through the increase in RU after injection of the premixes and the stability of the resultant complexes through their elution with buffer HBNS at a flow rate of 30  $\mu$ L/min. Ligand solutions without analyte were used as negative controls of complex formation and the sensor surface was regenerated after each sample injection.

## Results and Discussion

**Biochemical characterization of the recombinant proteins.** Authentic h $\alpha_2$ M has been routinely isolated from blood serum, where it is found at an excess of 2–4 mg/mL but is rather heterogeneous as to conformational state, glycosylation and presence of contaminants<sup>17,18</sup>. Native recombinant h $\alpha_2$ M was obtained from immortalized myelogenous leukemia cell line K-562 but the yield was not reported<sup>14</sup>. Therefore, efforts were made here to develop a system for heterologous expression of the protein with high yield, purity and homogeneity, as well as the necessary flexibility to engineer the protein at will. Full-length h $\alpha_2$ M with a C-terminal His<sub>6</sub>-tag was expressed in S2 insect cells using a standard transfection protocol<sup>30</sup> and the signal peptide of the adipokinetic hormone (AKH) for secretion to the extracellular environment. After seven days of expression and harvesting of the supernatant, the protein was purified by affinity chromatography, IEC and SEC steps with yields of up to ~1.0 mg of pure protein per liter of expression medium (Fig. 2A). The protein migrated as a tetramer of ~690 kDa according to SEC (data not shown). Its electrophoretic mobility in native-PAGE was similar to that of induced authentic h $\alpha_2$ M (Fig. 2B), which migrates faster than the native protein<sup>31</sup>. Chemical treatment with methylamine, which mimics the transition from native to induced h $\alpha_2$ M by opening the reactive thioester bond to produce a free cysteine<sup>31</sup>, did not have any effect on protein mobility. Consistently, the protein could not inhibit trypsin activity against a fluorogenic protein substrate, even at 10-fold molar excess. We conclude that recombinant h $\alpha_2$ M produced in insect cells was in the induced form, which does not permit the physiological entrance and entrapment of attacking peptidases<sup>4</sup>, similarly to a previous report of a baculovirus expression system<sup>32</sup>. Moreover, the thioester bond was either not formed or it was opened after secretion into the extracellular environment by nucleophiles from the expression medium. Unfortunately, we could not evaluate this possibility as the composition of the commercial medium that was used is not available. However, the latter hypothesis seems more plausible given that it is reported that insects can produce thioester-containing proteins<sup>33</sup>.

We next developed a transient expression system based on Expi293F cells, which derived from the HEK293 human embryonic kidney cell line and were cultured and harvested at 37 °C for three days. The protein was furnished with the leader sequence of mouse immunoglobulin  $\kappa$  (Ig  $\kappa$ ) for secretion and produced ~0.4 mg of pure protein per liter of expression medium (Fig. 2A). The protein migrated as a tetramer in SEC and showed electrophoretic mobility in native-PAGE between native and induced authentic h $\alpha_2$ M (Fig. 2B). Consistently, its capacity to inhibit trypsin was 35% of native authentic h $\alpha_2$ M. Together, these data indicate that the recombinant protein is partly native but mainly induced. Previous studies had indicated that thioester formation is a spontaneous process triggered by the packing energy of the polypeptide chain during folding in mammals<sup>34</sup>. Therefore, the limited ability of the Expi293F system to produce native protein was attributed, as in the insect cell system above, to the expression medium rather than to the lack of crucial cell machinery for proper thioester bond formation.

Then we expressed shorter variants of h $\alpha_2$ M in the insect and mammalian systems (Fig. 2A). N-h $\alpha_2$ M spanned from macroglobulin-like (MG) domain 1 (MG1) to MG7 in the insect cell system and from MG1 to MG6 in the mammalian system. C-h $\alpha_2$ M ranged from MG7 to the C-terminal receptor binding domain in both systems (Fig. 1A and Table 1). Expression of N-h $\alpha_2$ M yielded ~1.6/~3.2 mg per liter of insect and mammalian cell culture, respectively, while the values for C-h $\alpha_2$ M were ~1.3/~4.6 mg. N-h $\alpha_2$ M formed a dimer of ~170 kDa due to the presence of an intermolecular disulfide bond (C<sup>278</sup>–C<sup>431</sup>), which is also required for dimerization of the authentic full-length protein (Fig. 1A). Consistently, the protein migrated as a monomer of ~85 kDa in the presence of reducing agents. In turn, C-h $\alpha_2$ M was monomeric (~75 kDa) and treatment with methylamine did not affect the content of free cysteines or electrophoretic mobility in native-PAGE (Fig. 2B). To follow this up, we qualitatively assayed the content of free sulfhydryl groups by a fluorometric thiol assay kit, which gave a strong fluorescent signal for both the untreated and methylamine-treated C-h $\alpha_2$ M samples. This contrasted with native full-length authentic h $\alpha_2$ M, which gave no significant signal, and was similar to methylamine-induced authentic h $\alpha_2$ M, which likewise gave a strong signal. These assays indicated that the thioester bond was opened in C-h $\alpha_2$ M as mentioned above for the full-length recombinant variant, possibly owing to a nucleophilic component of the undisclosed cell-growth medium.

The insect and mammalian systems were also assayed for expression of mature human His<sub>6</sub>-tagged TGF- $\beta_2$  (Table 1), but without noticeable yields. Therefore, full-length pro-TGF- $\beta_2$  encompassing LAP and mature TGF- $\beta_2$  (see Fig. 1C) was expressed and purified in Expi293F cells as described elsewhere<sup>8</sup>, with a final yield of ~2.7 mg and ~2.3 mg of N-terminally octahistidine-tagged and Strep-tagged forms, respectively, per liter of mammalian cell culture (Fig. 2A). The protein migrated as a dimer of ~110 kDa in SEC, which indicates that the characteristic disulfide bonds were formed between the LAP and the mature TGF- $\beta_2$  moieties. The purified protein was partially cleaved before residue A<sup>303</sup> by host peptidases. Subsequent treatment with the physiological activating endopeptidase furin produced a homogeneously cleaved species consisting of LAP associated with the mature cytokine (Fig. 2A). Under physiological conditions, TGF- $\beta_2$  maturation is a complex process that involves a cascade of events under participation of several proteins that interact with the initial complex of pro-TGF- $\beta_2$  and the latent TGF- $\beta$  binding protein (LTBP). LTBP participates as a localizer of pro-TGF- $\beta_2$  to the extracellular matrix, whereas LAP senses the changes and releases mature TGF- $\beta_2$ <sup>11</sup>. Previous studies with a Chinese hamster ovary cell expression system benefited from the sensitivity of the LAP domain towards denaturing conditions at very low pH to separate it from mature TGF- $\beta_2$ <sup>11,35</sup>. In our case, this was unsuccessful, probably due to different post-translation modifications introduced by Expi293F cells in the highly glycosylated LAP<sup>8,36</sup>.

Finally, full-length GRAB was expressed without the cell-wall anchoring region (Fig. 1B) in a bacterial system yielding ~4 mg of pure protein per liter of expression medium after affinity chromatography, IEC and SEC steps (Fig. 2A). The protein migrated as a ~55-kDa species in SEC and as a ~33-kDa species in SDS-PAGE, but the values determined by SEC-MALLS (15.5 kDa; Table 2) were closer to the theoretical mass (15.8 kDa). We attribute this abnormal migration, which was described previously<sup>11</sup>, to the highly unstructured character of the protein.

**Interaction analysis of h $\alpha_2$ M and GRAB.** Interaction of streptococci with h $\alpha_2$ M has been reported to be highly specific<sup>15,16</sup>. Group A, G and C streptococci all bind the native form, whereas only the latter interact with the induced form. This result was attributed to the types of surface proteins, which are specific for each strain. GRAB is found in group A streptococci, and we studied its interaction with native authentic h $\alpha_2$ M by surface plasmon resonance. GRAB was immobilized as a ligand through a Strep-tag on a chip with covalently bound streptavidin. In a multicycle experiment, saturation of the ligand was reached with the two highest analyte concentrations, which gave on- and off-rate kinetic constants and results from affinity analysis (Fig. 3A,B). From the sensorgrams during the sequential injections of different analyte concentrations, we observed fast association and slow dissociation of h $\alpha_2$ M from GRAB, which indicated stable complex formation. Therefore, the ligand was removed in a regeneration step to make sure that all bound h $\alpha_2$ M was eliminated between injections with different analyte concentrations. The group of curves in Fig. 3A,B were fitted to a 1:1 Langmuir interaction model. These calculations revealed a  $\chi^2$  value < 10% of  $R_{\max}$ , which is indicative of a good fit. Consistently with the sensorgrams, the association rate constant ( $k_a$ ) and the dissociation rate constant ( $k_d$ ) were  $1.32 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  and  $1.90 \times 10^{-3} \text{ s}^{-1}$ , respectively, with an estimated dissociation half-time ( $t_{1/2} = \ln 2/k_{\text{off}}$ ) of 365 sec. The equilibrium dissociation constants ( $K_D$ ) from the kinetic and affinity analysis were  $1.43 \times 10^{-8} \text{ M}$  and  $3.45 \times 10^{-8} \text{ M}$  (Tables 3 and 4), respectively, which indicates high affinity and stable complex formation. The complex was also detected by SDS-PAGE and native-PAGE employing fluorophore-labelled GRAB (Fig. 4). Finally, SEC-MALLS analysis (Fig. 3D and Table 2) showed a molecular mass difference of 27.3 kDa over free h $\alpha_2$ M, which corresponds to 1.7 molecules of GRAB. Hence, we assume that two molecules of GRAB bind one h $\alpha_2$ M tetramer.

Under a similar experimental setup, methylamine-induced authentic h $\alpha_2$ M was injected over immobilized GRAB to reach equilibrium and saturation, which enabled analysis by affinity. The affinity data permitted calculation with confidence ( $\chi^2 < 10\%$  of  $R_{\max}$ ) of the equilibrium dissociation constant ( $9.46 \times 10^{-8} \text{ M}$ ), which was three times higher than that of native h $\alpha_2$ M (Fig. 3A,B and Table 4). This is consistent with published results, which indicated that GRAB shows preference for native over protease-induced h $\alpha_2$ M<sup>16</sup>. The complex was likewise analyzed by SDS-PAGE and native-PAGE with fluorophore-labelled GRAB (Fig. 4). The results showed an increase in the molecular mass of 26.5 kDa over noncomplexed induced h $\alpha_2$ M, which is equivalent to the results for native h $\alpha_2$ M.

To map down the region of h $\alpha_2$ M engaged in GRAB binding, we repeated the above experiments with N-h $\alpha_2$ M and C-h $\alpha_2$ M. In a similar multicycle experimental setup, we could not detect any interaction. However, previous incubation of the proteins at 37 °C for one hour apparently enabled complex formation. Protein remained complexed over time after injection and washing of the chip (Fig. 3B), but in this case we could not determine the affinity constants due to the experimental setup. The complexes were subsequently evaluated in native-PAGE using fluorophore-labelled GRAB (Fig. 3D). In this case, we detected the interaction of GRAB with C-h $\alpha_2$ M but not with N-h $\alpha_2$ M.

**Interaction analysis of h $\alpha_2$ M and pro-TGF- $\beta_2$ .** Previous biochemical data had revealed that h $\alpha_2$ M binds TGF- $\beta_2$  mainly through a mature cytokine segment spanning residues A<sup>343</sup>-Y<sup>367</sup>, in which W<sup>354</sup> plays a major role<sup>10</sup>. No data have been reported on the role of LAP. However, inspection of the crystal structure of homologous pro-TGF- $\beta_1$  (see Protein Data Bank code 3RJR<sup>37</sup>) reveals that the interacting segment is partially shielded by LAP. Other studies employing a library of overlapping glutathione S-transferase fusion proteins ascribed the potential binding site for TGF- $\beta_1$  to segment V<sup>723</sup>-T<sup>761</sup> of h $\alpha_2$ M<sup>38</sup>, which was subsequently narrowed down to E<sup>737</sup>-V<sup>756</sup> employing synthetic peptides<sup>39</sup>. However, further details on the mechanism are unknown. To further shed light, we set out to characterize binding of pro-TGF- $\beta_2$  to h $\alpha_2$ M. We checked the interaction by surface plasmon resonance in multicycle experiments with immobilized pro-TGF- $\beta_2$  as ligand but could not detect complex formation. Only after analysis by native-PAGE using fluorophore-labelled pro-TGF- $\beta_2$  we observed interaction with native authentic h $\alpha_2$ M but not with the induced form or the short variants (Fig. 4). Given that the pro-TGF- $\beta_2$  sample contained a mixture of cleaved and intact protein, we assayed N-terminally His<sub>6</sub>-tagged pro-TGF- $\beta_2$  with native h $\alpha_2$ M in native-PAGE followed by Western blotting. The two proteins were not co-migrating (data not shown). Thus, we conclude that LAP prevents h $\alpha_2$ M from binding mature TGF- $\beta_2$  as suggested by structural studies on pro-TGF- $\beta_1$ .

**Conclusions.** Protein h $\alpha_2$ M is a sophisticated player to spatially and temporally restrict and regulate key physiological processes that control the distribution and activity of many proteins, including peptidases, cytokines, hormones and other physiological effectors<sup>1</sup>. Since the 1940s, several efforts have been made to understand its mechanism of action *in vivo* and *in vitro*, but they have been hampered by the unavailability of high-yield recombinant expression systems. Here, we developed insect and mammalian systems for the full-length protein and shorter fragments. The former was mainly produced in an induced state, possibly due to media components that cause induction during the time scale of expression. Thus, other media with a regulated composition will be assessed to reevaluate the recombinant systems.

The recombinant proteins plus authentic h $\alpha_2$ M were analyzed for binding with GRAB and pro-TGF- $\beta_2$ . The former tightly bound native and methylamine-induced authentic h $\alpha_2$ M, with two molecules of GRAB per h $\alpha_2$ M tetramer. The short variants, especially C-h $\alpha_2$ M, likewise complexed GRAB, but apparently through a different mechanism from the full-length forms. In contrast, full-length pro-TGF- $\beta_2$  did not complex any h $\alpha_2$ M variant, probably owing to steric hindrance by the N-terminal LAP domain.

## References

- Garcia-Ferrer, I., Marrero, A., Gomis-Rüth, F. X. & Goulas, T.  $\alpha_2$ -Macroglobulins: structure and function. *Subcell. Biochem.* **83**, 149–183 (2017).
- Goulas, T. *et al.* Structural and functional insight into pan-endopeptidase inhibition by  $\alpha_2$ -macroglobulins. *Biol. Chem.* **398**, 975–994 (2017).
- Barrett, A. J. & Starkey, P. M. The interaction of  $\alpha_2$ -macroglobulin with proteinases. *Biochem. J.* **133**, 709–724 (1973).
- Sottrup-Jensen, L., Sand, O., Kristensen, T. & Fey, G. H. The  $\alpha$ -macroglobulin bait region. Sequence diversity and localisation of cleavage sites for proteinases in five mammalian  $\alpha$ -macroglobulins. *J. Biol. Chem.* **264**, 15781–15789 (1989).
- Garcia-Ferrer, I. *et al.* Structural and functional insights into *Escherichia coli*  $\alpha_2$ -macroglobulin endopeptidase snap-trap inhibition. *Proc. Natl. Acad. Sci. USA* **112**, 8290–8295 (2015).
- Marrero, A. *et al.* The crystal structure of human  $\alpha_2$ -macroglobulin reveals a unique molecular cage. *Angew. Chem. Intl. Ed.* **51**, 3340–3344 (2012).
- Bieth, J. G., Tourbez-Perrin, M. & Pochon, F. Inhibition of  $\alpha_2$ -macroglobulin-bound trypsin by soybean trypsin inhibitor. *J. Biol. Chem.* **256**, 7954–7957 (1981).
- del Amo-Maestro, L., Marino-Puertas, L., Goulas, T. & Gomis-Rüth, F. X. Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor  $\beta_2$  in a new conformation. *Sci. Rep.* **9**, 8660 (2019).
- LaMarre, J. *et al.* An  $\alpha_2$ -macroglobulin receptor-dependent mechanism for the plasma clearance of transforming growth factor- $\beta_1$  in mice. *J. Clin. Invest.* **87**, 39–44 (1991).
- Liu, Q. *et al.* Identification of the high affinity binding site in transforming growth factor- $\beta$  involved in complex formation with  $\alpha_2$ -macroglobulin. Implications regarding the molecular mechanisms of complex formation between  $\alpha_2$ -macroglobulin and growth factors, cytokines, and hormones. *J. Biol. Chem.* **276**, 46212–46218 (2001).
- Annes, J. P., Munger, J. S. & Rifkin, D. B. Making sense of latent TGF- $\beta$  activation. *J. Cell Sci.* **116**, 217–224 (2003).
- O'Connor-McCourt, M. D. & Wakefield, L. M. Latent transforming growth factor- $\beta$  in serum. A specific complex with  $\alpha_2$ -macroglobulin. *J. Biol. Chem.* **262**, 14090–14099 (1987).
- Webb, D. J., Wen, J., Karns, L. R., Kurilla, M. G. & Gonias, S. L. Localisation of the binding site for the transforming growth factor-beta in human  $\alpha_2$ -macroglobulin to a 20-kDa peptide that also contains the bait region. *J. Biol. Chem.* **273**, 13339–13346 (1998).
- Arandjelovic, S., van Sant, C. L. & Gonias, S. L. Limited mutations in full-length tetrameric human  $\alpha_2$ -macroglobulin abrogate binding of platelet-derived growth factor-BB and transforming growth factor- $\beta_1$ . *J. Biol. Chem.* **281**, 17061–17068 (2006).
- Godehardt, A., Hammerschmidt, S., Frank, R. & Chhatwal, G. S. Binding of  $\alpha_2$ -macroglobulin to GRAB (Protein G-related  $\alpha_2$ -macroglobulin binding protein), an important virulence factor of group A streptococci, is mediated by two charged motifs in the  $\Delta A$  region. *Biochem. J.* **381**, 877–885 (2004).
- Rasmussen, M., Müller, H.-P. & Björck, L. Protein GRAB of *Streptococcus pyogenes* regulates proteolysis at the bacterial surface by binding  $\alpha_2$ -macroglobulin. *J. Biol. Chem.* **274**, 15336–15344 (1999).
- Goulas, T., Garcia-Ferrer, I., Garcia-Pique, S., Sottrup-Jensen, L. & Gomis-Rüth, F. X. Crystallization and preliminary X-ray diffraction analysis of eukaryotic  $\alpha_2$ -macroglobulin family members modified by methylamine, proteases and glycosidases. *Molec. Oral Microbiol.* **29**, 354–364 (2014).
- Andersen, G. R. *et al.* Crystallisation of proteins of the  $\alpha_2$ -macroglobulin superfamily. *Ann. N.Y. Acad. Sci.* **737**, 444–446 (1994).
- Goulas, T. *et al.* The pCri System: a vector collection for recombinant protein expression and purification. *PloS one* **9**, e112643 (2014).
- Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucl. Acids Res.* **17**, 6545–6551 (1989).
- Xie, Q. *et al.* TubeSpin bioreactor 50 for the high-density cultivation of Sf-9 insect cells in suspension. *Biotechnol. Lett.* **33**, 897–902 (2011).
- Strober, W. Trypan blue exclusion test of cell viability. *Curr. Protoc. Immunol.* **21**, A.3B.1–A.3B.2 (2001).
- Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
- Sottrup-Jensen, L., Petersen, T. E. & Magnusson, S. A thiol-ester in  $\alpha_2$ -macroglobulin cleaved during proteinase complex formation. *FEBS Lett.* **121**, 275–279 (1980).
- Imber, M. J. & Pizzo, S. V. Clearance and binding of two electrophoretic “fast” forms of human  $\alpha_2$ -macroglobulin. *J. Biol. Chem.* **256**, 8134–8139 (1981).
- Morrot, A. *et al.* Human T cell responses against the major cysteine proteinase (cruzipain) of *Trypanosoma cruzi*: role of the multifunctional  $\alpha_2$ -macroglobulin receptor in antigen presentation by monocytes. *Int. Immunol.* **9**, 825–834 (1997).
- Haider, S. R., Sharp, B. L. & Reid, H. J. A comparison of Tris-glycine and Tris-tricine buffers for the electrophoretic separation of major serum proteins. *J. Separat. Sci.* **34**, 2463–2467 (2011).
- Sinha-Datta, U., Khan, S. & Wadgaonkar, D. Label-free interaction analysis as a tool to demonstrate biosimilarity of therapeutic monoclonal antibodies. *Biosimilars* **15**, 83–91 (2015).
- Zhao, J. *et al.* Novel method for measurement of heparin anticoagulant activity using SPR. *Anal. Biochem.* **526**, 39–42 (2017).
- Shen, X., Hacker, D. L., Baldi, L. & Wurm, F. M. Virus-free transient protein production in Sf9 cells. *J. Biotechnol.* **171**, 61–70 (2014).
- Sottrup-Jensen, L.  $\alpha$ -Macroglobulin: structure, shape, and mechanism of proteinase complex formation. *J. Biol. Chem.* **264**, 11539–11542 (1989).
- van Rompaey, L. & Marynen, P. Temperature-dependent biosynthesis of thiol esters in baculovirus recombinant  $\alpha_2$ M and PZP. *Ann. N. Y. Acad. Sci.* **737**, 506–509 (1994).
- Baxter, R. H. G. *et al.* Structural basis for conserved complement factor-like function in the antimalarial protein TEP1. *Proc. Natl. Acad. Sci. USA* **104**, 11615–11620 (2007).



34. Grøn, H., Thøgersen, I. B., Enghild, J. J. & Pizzo, S. V. Structural and functional analysis of the spontaneous reformation of the thiol ester bond in human  $\alpha_2$ -macroglobulin, rat  $\alpha_1$ -inhibitor-3 and chemically modified derivatives. *Biochem. J.* **318**, 539–545 (1996).
35. Zou, Z. & Sun, P. D. An improved recombinant mammalian cell expression system for human transforming growth factor- $\beta$ . *Prot. Expr. Purif.* **50**, 9–17 (2006).
36. Crosset, A. *et al.* Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J. Biotechnol.* **61**, 336–348 (2012).
37. Shi, M. *et al.* Latent TGF- $\beta$  structure and activation. *Nature* **474**, 343–349 (2011).
38. Webb, D. J., Roadcap, D. W., Dhakephalkar, A. & Goniás, S. L. A 16-amino acid peptide from human  $\alpha_2$ -macroglobulin binds transforming growth factor- $\beta$  and platelet-derived growth factor-BB. *Prot. Sci.* **9**, 1986–1992 (2000).
39. Arandjelovic, S., Freed, T. A. & Goniás, S. L. Growth factor-binding sequence in human  $\alpha_2$ -macroglobulin targets the receptor-binding site in transforming growth factor- $\beta$ . *Biochemistry* **42**, 6121–6127 (2003).

## Acknowledgements

We are grateful to Roman Bonet, Xandra Kreplin and Joan Pous from the joint IBMB/IRB Automated Crystallography Platform and the Protein Purification Service for assistance during purification and SEC-MALLS experiments. This study was supported in part by grants from Spanish and Catalan public and private bodies (grant/fellowship references BFU2015-64487R; MDM-2014-0435; JCI-2012-13573; BES-2015-074583; BES-2013-064651; 2017SGR3; and Fundació “La Marató de TV3” 201815). The Structural Biology Unit ([www.sbu.csic.es](http://www.sbu.csic.es)) of IBMB is a “María de Maeztu” Unit of Excellence from the Spanish Ministry of Science, Innovation and Universities.

## Author Contributions

T.G. and F.X.G.R. conceived and supervised the work; L.M.P. and L.d.A.M. produced and purified the protein; L.M.P. and L.d.A.M. performed biochemical studies; M.T.M. guided the Biacore experiments; and T.G. and F.X.G.R. wrote the paper with contributions from all authors.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-45712-z>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

# **GENERAL DISCUSSION**

---



The present thesis has focused on the biochemical and biophysical/structural study of three glycoproteins, namely ADAMTS13, neprosin and TGF $\beta$ 2, and provides novel and compelling information regarding the molecular mechanism of the protein-protein interactions with their respective partner as well as on the proteolytic mechanism employed. All three proteins were successfully expressed at high levels by a mammalian protein expression system and possess an important biomedical relevance in either cardiovascular disorders or coeliac disease treatment.

The **first project** of the thesis is based on the von Willebrand Factor (VWF)/ADAMTS13 axis which is involved in a number of cardiovascular conditions and of great relevance to human health. We aimed to analyse the interaction between the metalloprotease ADAMTS13 (AD13), particularly the C-terminal truncated MDTCS variant (A<sup>75</sup>-P<sup>682</sup>; see Fig. 17 in section 3.4.1.), and its only described physiological substrate VWF, through a combination of biochemical, biophysical and computational approaches. By using the VWF73 (D<sub>1596</sub>-R<sub>1668</sub>) peptide, which contains the essential binding and cleavage sites for AD13 (Kokame et al. 2004), we expanded the current knowledge of the underlying working mechanism of substrate binding and site specific cleavage. This shall be crucial to develop novel therapeutic strategies following so-called rational drug design approaches.

The functional interactions between the exosites in the M, D, C, and S domains of AD13 and their complementary binding sites in the VWF73 are essential for efficient proteolysis to occur (Gao et al. 2006; Gao, Anderson, and Sadler 2008; de Groot et al. 2009; Majerus, Anderson, and Sadler 2005). In the absence of the D, C, S domain exosite interactions, proteolysis of VWF substrates (VWF115, VWF96, VWF73) by the M domain alone has been described to lack both specificity and efficiency (de Groot et al. 2009; Ai et al. 2005). Intriguingly, it has been demonstrated that AD13 variants lacking distal AD13 domains (T2-CUB domains) bind to the VWF fragments with higher affinity (Muia et al. 2014) than the full length protein, and show accelerated cleavage rates of peptidic substrates (Gao et al. 2006; Gao, Anderson, and Sadler 2008).

Consequently, we used AD13-MDTCS which is C-terminally truncated after the Spacer domain and an active site point mutant (E<sup>225</sup>Q) for our protein interaction assays. This mutation is localized within the zinc-binding consensus sequence (HExxH) and was introduced to change the general base glutamate essential for the proteolysis mechanism to glutamine in order to inactivate the protease.

## G.1 Functional characterization: cleavage and binding studies

Notably, it is described that the active site of AD13 is a dynamic structure; i.e. it can exist in a “latent” or “closed” state in which the active site is inaccessible, and in an “active” or “open” state in which the active site can accommodate substrate binding and facilitate cleavage (Petri et al. 2019), and for which AD13 exosite interactions with VWF are essential to shift the equilibrium towards the otherwise unfavourable active state. Therefore, the M domain is probably a highly flexible domain, which could at least in parts explain its inability to crystallize given that flexible molecules are not favourable for crystallization. Additionally, it is suggested that binding of one or more exosites result in a conformational change of M that favours the active/open state. The crystal structure of the M domain of ADAMTS4 (AD4) has been solved in the presence or absence of an active site inhibitor (Mosyak et al. 2008) indicating flexibility in the active site (Petri et al. 2019).

The first assessment of VWF-strep-peptide (Strep tag variant of VWF73) as a suitable substrate for AD13-MDTCS was performed qualitatively by conventional cleavage assays. (Figure 1F). VWF-strep-peptide was cleaved by AD13-MDTCS within 10 min. This means that VWF-strep-peptide indeed possesses all functionally important sites for proteolysis and substrate recognition. Interestingly, these results suggest that our peptidase preparation is more active than a recently published one from insect cells which only partially cleaved VWF96 and VWF115 (Petri et al. 2019). One hypothesis would be that the glycosylation pattern of our preparation is more similar to the native protein, compared to the *Drosophila* derived one. Another explanation would be that longer substrates possess additional secondary structures that impede access to the scissile bond.

In this thesis we performed several binding studies between AD13-MDTCS E<sup>225</sup>Q and VWF-strep-peptide by using different approaches. First, we performed biochemical binding tests such as pull down or fluorescent AMCA assays, which showed a stable complex that prevails over interactions with IMAC resin or with the gel matrix, respectively. Secondly, we determined the complex affinity and kinetic parameters using surface plasmon resonance (SPR). The resulting sensograms revealed fast association ( $k_a$ ) and dissociation ( $k_d$ ) of the binding partners, which indicated a transient but stable complex formation. However, the equilibrium dissociation constants ( $K_D$ ) from the kinetic and steady-state analysis were in the same order of magnitude ( $10^{-7}$ M), meaning notable sub-micromolar affinity. This stable complex can be explained, because since VWF-strep-peptide is attached to a surface by the strep-tag and due to the mechanical flow generated in SPR experiments, causing its

completely extended conformation and therefore favouring the interaction with AD13-MDTCS E<sup>225</sup>Q.

### G.1.1 Protein dynamics: structural proteomics (CLMS and HDX-MS)

The interaction was also analysed by chemical cross-linking mass spectrometry analysis (CLMS) using different crosslinking agents. K<sup>318</sup>, which is exposed in the D domain, is the major cross-linked residue found given that it was involved in a total of 12 links in the C, M and S domains. Analogous to the intramolecular cross-links, residue K<sup>318</sup> is the most abundant intermolecular cross-linking residue (see table 1 in results section). This is probably due to its central localization in an exposed and highly flexible region, and as thus it can get into vicinity of many other peptide regions.

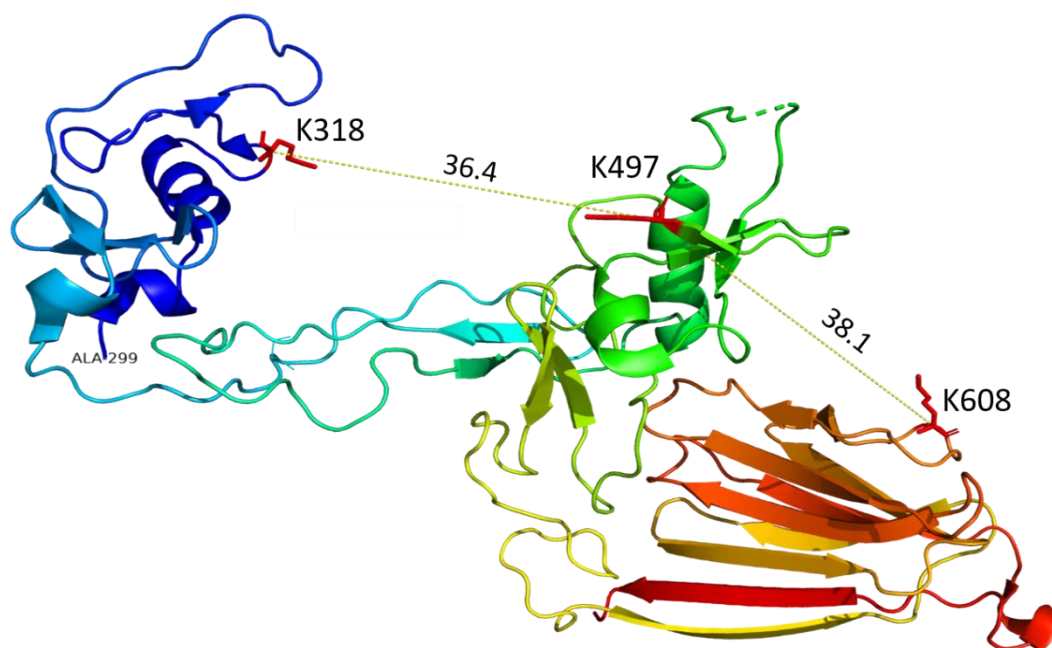
Notably, it is proposed that the D domain exosites form the most vital of all exosite interactions and not only influences the affinity of the two proteins, but also modulates the efficiency of the proteolytic event (de Groot et al. 2009; Petri et al. 2019). Intramolecular cross-linking, were found between domains, indicating that AD13-MDTCS E<sup>225</sup>Q upon substrate binding presents great flexibility in solution. This might be due to a transition from open to closed or globular conformation, bringing the individual domains closer together, being in agreement with previous reports (Petri et al. 2019; Jian Zhu et al. 2019) and HDX-MS results.

Notably, some cross-links are compatible with the crystal structure of the latent state in complex with an inhibitory antibody antigen-binding fragment (Fab; PDB 6qig; Petri et al. 2019). However, others require considerable structural rearrangement –which may reflect the active state. For instance, quantitative crosslinking (QCLMS) using disuccinimidyl adipate (DSA) as a cross linker revealed differences in the intramolecular crosslinking between K<sup>318</sup>-K<sup>497</sup>-K<sup>608</sup> in the presence and absence of the substrate (Fig. 33), indicating that the D domain is in closer proximity to the C domain upon substrate binding, and suggesting that the protease adopts a more compact or “tight” conformation in the presence of VWF-strep-peptide and thus in the active conformation. This implies that the active structure (AD13-MDTCS-E<sup>225</sup>Q bound to VWF-strep-peptide) is significantly different to the published latent form, and/or displays, even in the open conformation, a large degree of flexibility.

Consequently, we suggested that this arrangement may happen via the thrombospondin domain 1 (T), which acts as a “spacer” domain that gives flexibility to the arrangement of the N- and C-terminally attached domains.

Similarly, full length AD13 has recently been shown to be a conformationally flexible molecule, with the ability to adopt both folded and extended conformations (see figure 22 in 3.4.4.2. section) where the folded conformation depends on interactions between the Spacer and CUB domains (Muia et al. 2014; South et al. 2014; South, Freitas, and Lane 2017). It is reported that three linker regions, present amongst the T2-CUB regions (mostly composed of thrombospondin domains), convey conformational flexibility to AD13 (Deforche et al. 2015).

Moreover, it is described that the T domain contains many disulphide bonds, thus limiting the internal flexibility (but not the interdomain flexibility) which means that the hinges points would be either at the N-terminal or C-terminal side of the domain which results in the T domain to behave like a stiff arm, which can rotate because of the linker regions right after it. In this case, MDTCS can be described as an excavator where the T domain would be the arm, the main body is represented by the M+D domains, and the shovel is composed by the C+S domains, thus the T domain can bring it closer to the active site (tight conformation) when the VWF is bound, and further away (extended conformation).



**Fig. 33. Quantitative intramolecular crosslinking using DSA:** Modified model comprised of PDB 3ghm and PDB 6qig being D in dark blue, T in cyan, C in green and S in orange. Visualization of short (spacer length 7 Å) DSA cross-links. The distance between lysine residues can reach a maximum of about 22 Å. Lines represent distances between C $\alpha$ . The formation of cross-links K<sup>318</sup> - K<sup>497</sup> and K<sup>497</sup> - K<sup>608</sup> indicates that the D domain is in close proximity to the C domain compared to the crystal DTCS structure (PDB 3ghm).

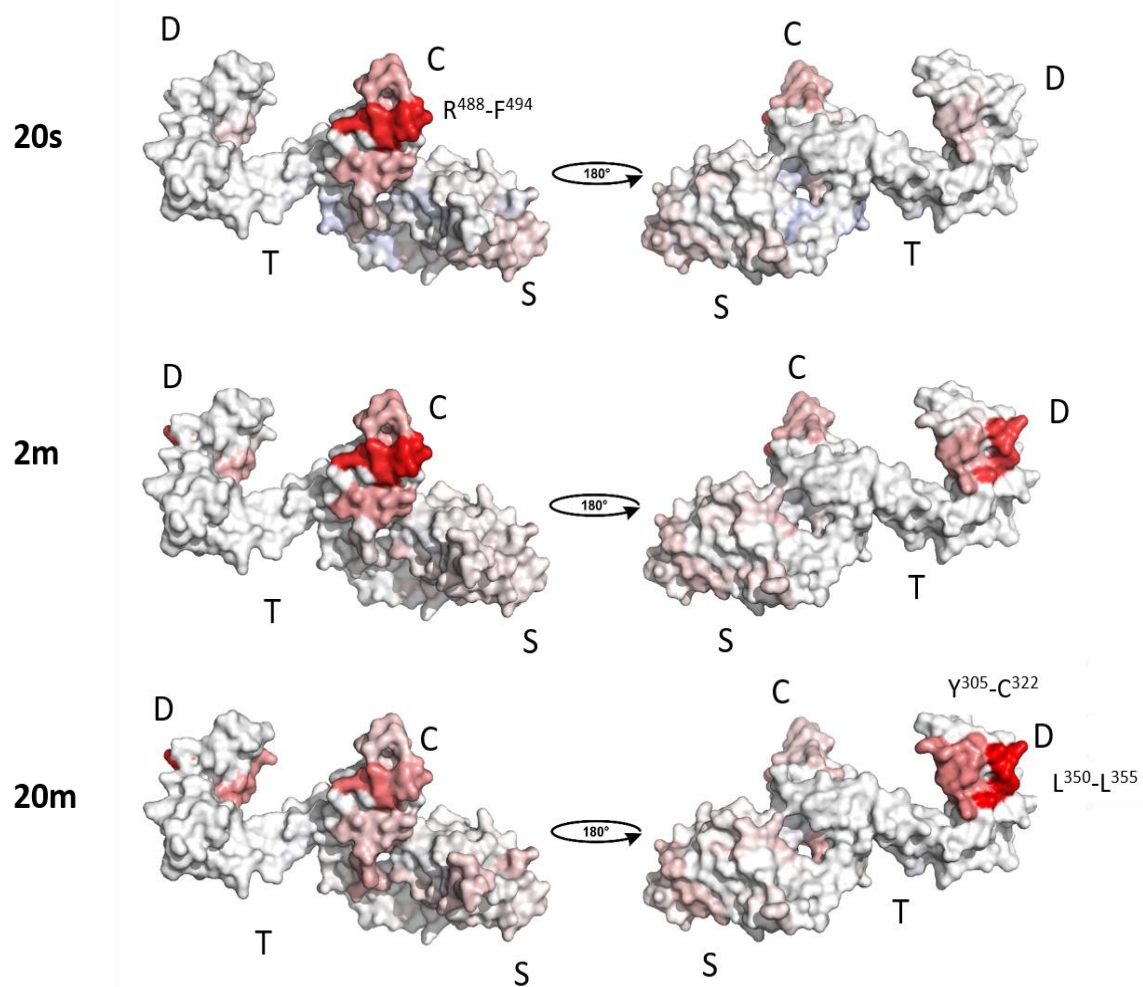
Regarding intermolecular cross-links, VWF-strep-peptide exhibits cross-linked residues along the entire AD13-MDTCS E<sup>225</sup>Q molecule, indicating the peptide slides along the surface of AD13-MDTCS E<sup>225</sup>Q by contacting all domains. This would mean that the structure gets probably more “tight” upon VWF binding, which goes hand in hand with the results observed in intramolecular crosslinking. Moreover, presuming that AD13-MDTCS E<sup>225</sup>Q is flexible, also the substrate requires to be flexible, assuming that it might “co-fold” to build the active form of the active site.

Complementary to crosslinking experiments, we performed HDX-MS studies, which disclosed protected regions AD13-MDTCS E<sup>225</sup>Q at different time points in the presence or absence of the VWF-strep-peptide. Thus, this technique allowed us to elucidate dynamic changes in surface protection, and indicate that the peptidase interacts with the peptide mainly via its C domain, and secondarily via the D domain – consistent with our results from CLMS (Fig. 34). Importantly, we have found that the region L<sup>350</sup>-D<sup>355</sup> (within D domain) shows dynamic changes in surface protection. It is in complete agreement with the previous results, where it has been reported that R<sup>349</sup>, L<sup>350</sup> and V<sup>352</sup> residues are essential for AD13 activity (Petri et al. 2019) and therefore this region may be considered as a crucial binding exosite.

In contrast, M domain shows little changes in surface protection since the coverage of the protein in HDX-MS experiments was reduced due to the AD13-MDTCS E<sup>225</sup>Q high glycosylation pattern. However, we should have expected a change of the M surface resulting from the arrangement of D, C and S domains and thereby as a result of the accommodation of the scissile peptide bond. Furthermore, it has been previously described that residues R<sup>659</sup>, R<sup>660</sup>, Y<sup>661</sup> and Y<sup>665</sup> form an exosite in the AD13 Spacer domain that binds to the VWF A2 domain and thereby promotes proteolysis (Pos et al. 2010; Jin, Skipwith, and Zheng 2010). However, our HDX-MS experiments showed a high deuteration rate in all-time points for several parts of the S domain, including the above mentioned exosite (see Suppl. Fig. 2). Thus the S domain is probably partly disordered in the present conditions, and thus no stable interactions could be identified, neither in the HDX-MS nor in the CLMS experiments.

Essentially, all remote exosites (D, C and S) make a similar contribution to the binding between VWF-strep-peptide and AD13-MDTCS E<sup>225</sup>Q. These results might be due to the induced rearrangement in the D, C and S domains by the VWF-strep-peptide binding, suggesting that the peptidase adopts a more compact conformation.





**Fig. 34. Ligand-induced conformational dynamics of AD13-MDTCS E<sup>225</sup>Q.** Colored structure of DTCS in presence of VWF-strep-peptide based on the differential deuterium uptake showing dynamic changes in surface protection. In the short time of D<sub>2</sub>O incubation (20s), an extended region containing loop R<sup>488</sup> – F<sup>494</sup> of the C domain is highly protected in the presence of VWF. With the increasing time of incubation, high protection of a region V<sup>313</sup> – C<sup>322</sup> and L<sup>350</sup> – L<sup>355</sup> of the D domain was observed. These dynamic changes in surface protection might indicate that AD13-MDTCS E<sup>225</sup>Q interacts first with the VWF via the C domain and then through the D domain.

### G.1.2 Integrative analysis: in-solution structural analysis

In order to understand the conformation and dynamics of AD13-MDTCS E<sup>225</sup>Q and the relevance for its function, we performed an integrative analysis of atomic models and in-solution analysis by structural proteomics. To complement this structural analysis, we also performed multi-angle laser light scattering (MALLS) and small-angle X-ray scattering (SAXS) in solution in presence or absence of VWF-strep-peptide.

First, we found that AD13-MDTCS E<sup>225</sup>Q exhibits flexibility in solution with TSP1 (T domain; I<sup>380</sup>-E<sup>439</sup>) acting as a hinge. To this end, we compared our experimental SAXS profile with the theoretical SAXS profile of the crystal structure (Petri et al. 2019) which depicts noticeable disagreement, but which was not unexpected. This indicates that either AD13-MDTCS has a significantly different conformation in solution or exhibits such high flexibility that a single conformation, as observed in the crystal structure, is not sufficient to explain the experimentally determined SAXS profile.

To conclude, AD13-MDTCS E<sup>225</sup>Q exist as a flexible monomer in solution with T acting as a hinge to allow different relative dynamic orientations of its individual domains. Using SAXS, we identified a crucial hinge region at the D domain facing towards the end of the thrombospondin motif 1, corresponding to S<sup>389</sup>, R<sup>409</sup> and G<sup>425</sup>. The C-alpha atoms of these residues are highlighted as cyan spheres in Suppl. fig. 4B. This of course also substantially complicates the interpretation of our CLMS experiments, as the published structure probably neither reflects the active nor latent state of AD13, and as we probably observe several three dimensional arrangements of the individual domains in our in-solution experiments.

Secondly, we analysed the AD13/VWF-strep-peptide complex by SEC-MALLS which revealed a single peak of 96.8 kDa, consisting of the theoretical value of 89 kDa, indicating the 1:1 complex stoichiometry. We also collected SAXS data for the complex in SEC-SAXS mode, which suggests that the complex corresponds to a globular particle. These findings completely align with CLMS and HDX-MS experiments.

Interestingly, we found that the selected sub-ensembles (see Suppl. Fig. 4A) completely differ from the results obtained for unligated AD13-MDTCS E<sup>225</sup>Q. This suggest that binding to the VWF-strep-peptide changes the inherent molecular motions of AD13-MDTCS-E<sup>225</sup>Q. Even though the published AD13 structure is supposedly in a latent conformation and thus inactive, the active site is actually correctly built with the catalytic zinc ion present. This further highlights the multi-layered activity regulation mechanism in AD13, which involves several exosite interactions between the proteases and its only native substrate, VWF.

Notably, it has to be considered that there are multiple ways in which the proteins may interact, thus, it is important to note that these sub-ensembles alone do not fit the SAXS data but represent minimal clusters that collectively fulfil all the distance restraints obtained from CLMS. This observation suggests that the putative presence of multiple simultaneous binding events suggested by the “molecular zipper” model (Crawley et al. 2011) must be of

transient nature, for which SAXS is a technique not sensitive enough. Moreover, AD13 displays a really sophisticated regulatory cleavage mechanism, so it is even possible that the in-solution arrangement is not even representing the situation *in vivo* as it is even more complex there (e.g. by other binding partners, the additional AD13 C-terminal domains, the multimeric VWF, etc).

AD13 is the only member of the ADAMTS family with a known role in blood plasma and with a crucial physiological role, as evidenced by a severe thrombotic phenotype associated with its inherited or acquired deficiency. All ADAMTS family members contain the MDTCS regions but differ in identity and number of C-terminal domains. Therefore, it is reasonable to predict that they share similarity of functional aspects, especially mechanism of substrate recognition.

The resolution of the crystal structure of AD13-MDTCS reveals that the active site cleft in the M domain is occluded, suggesting that the protease is in a latent conformation. Thus a rearrangement to an active state is required, in which AD13-MDTCS E<sup>225</sup>Q is capable to accommodate the substrate. Another, still possible, hypothesis is that the “closed” active site could be a consequence of the binding of the Fab rather than being the natural latent conformation of the M domain. To definitively conclude the existence of “open” and “closed” active states, a crystal structure of the open active site state is required.

Consequently, we generated first molecular models for the competent complex including all experimental CLMS restrains. Comparing our computational model with the published crystal structure, we can conclude that the M and D domains are in close proximity to each other in both structures whereas the hinge T residues permit the rotation of the molecule and cause the maximal displacement of the C and S domain (Figure 5B). Furthermore, we generate a model based on the experimental data where the VWF-strep-peptide runs through the active site cleft of the peptidase in an active conformation (Figure 5D).

A co-crystal structure of AD13-MDTCS E<sup>225</sup>Q with the VWF73 peptide would be very informative, and substrate binding may also stabilise the fold of the accessory domains as well as the M domain. Moreover, it would potentially reveal the identity of the interacting residues involved in exosite interactions, and more importantly, the arrangement of the individual domains in the active state. Complex formation and co-crystallization studies have been performed in our lab. However, we couldn't form a stable complex between the two binding partners, which is in line with our experimental insights showing low binding energy

and intensity contacts, probably due to an only transient interaction between the two proteins.

In the **second project** I focused on the characterization of a prolyl endopeptidase, which could expand the possibilities for treating coeliac disease through an enzyme supplementation strategy. For this purpose, we used a combination of biochemical, biophysical and *in vivo* approaches.

## **G.2 Structural and functional characterization *in vitro* of neprosin**

First, we developed an efficient recombinant expression system in human Expi293F cells, and established versatile purification schemes for both the zymogen (pro-neprosin) and the mature protease. Our initial attempts to express the target protein in bacterial ArcticExpress cells – as published by Schröder et al. 2017 (Schröder et al. 2017) failed to yield sufficient material for structural studies, probably since glycosylation is essential for proper protein folding and stability.

Based on the results of this thesis, all enzymatic studies performed with pro-neprosin showed no activity, confirming the inhibitory function of the N-terminal prodomain. However, an acid environment causes the activation of the pro-enzyme leading to the mature form of the protease. Thereby, a highly active protease is obtained comprising segment S<sup>129</sup>-Q<sup>380</sup>, and a molecular weight of 29 kDa, while the 11 kDa prodomain is completely degraded. We demonstrate that both omitting the prodomain from the construct, as well as the point mutation A60R, abolish expression of the protease. The A<sup>60</sup> residue is found in a prodomain hydrophobic region, hence disrupting the stability of the prodomain may cause no protein expression, indicating that the prodomain plays a crucial role in the folding and/or secretion of the protease. Moreover, stability assays demonstrated that the zymogen displays a higher  $T_m$  compared to the active neprosin, but both of them exhibited a high thermal stability, with melting points of 76.8 and 67.8 °C, respectively. We attribute this thermal and low pH stability to the effect of glycosylation and the disulphide network, as described for other proteins (Wang et al. 1996; Yoshimasu et al. 2004).

Importantly, neprosin displayed its highest enzymatic activity at highly acidic pH values. However, the pH optimum is with pH 3 slightly higher than for pepsin. Although it has been demonstrated neprosin retained activity under strongly reduced conditions (Schröder et al. 2017), TSA assays showed that neprosin appears to be less stable when treated with TCEP than untreated enzyme. However, we have never tried how this affects its activity.

Regarding nepenthesin enzymes, the presence of reducing agents causes nearly complete loss of activity (Kadek, Tretyachenko, et al. 2014). This result confirmed the presence of disulphide bonds in the molecule and proved the requirement of them for the stabilization of the protease. This behaviour might be explained by the number and location of the disulphide bonds, being 3 disulphide bonds in the case of neprosin, one located in the prodomain and two in the mature form. Furthermore, we confirmed that neprosin exhibits full proteolytic activity after being lyophilised. These results show that this enzyme has great potential as an oral enzyme supplement to degrade gluten proteins before they reach the small intestine, and thus to mitigate coeliac disease.

### **G.2.1 Gluten degradation: gliadin and 33-mer**

The gluten and gluten-like proteins from wheat, rye and barley such as gliadins possess an unusual high proline content which causes inherent resistance to proteolytic degradation by most human proteases and trigger a pro-inflammatory T-cell response in patients with coeliac disease. Although the *Aspergillus Niger* prolyl peptidase (AN-PEP) appears to be most studied candidate to date (König et al. 2017), only high doses of AN-PEP are capable of eliminating the accumulation of immunogenic peptides (Mitea et al. 2008) and thus neprosin could represent a highly probable alternative (Rey et al. 2016).

For example, our studies determined that neprosin accelerates the clarification rate of gliadin slurries, and we observed that pepsin is more efficient at gliadin digestion in the presence of neprosin suggesting a great synergy between host proteases and neprosin. It is shown that already very low neprosin levels strongly enhance the solubilisation rate for gliadin slurries, which is necessary for effective digestion in the stomach. This latter feature together with its own resistance to pepsin degradation present fundamental requirements for an effective supplementation approach, and thus neprosin is showing high potential for gluten intolerant patients.

Additionally, we confirmed the active site specificity of neprosin by re-analysing the deposited cleavage data (Schröder et al. 2017), with proline in P1 being the predominant specificity determinant, underlining the classification of neprosin as a prolyl endopeptidase. Proline was also enriched 2-fold over natural abundance in positions P2 and P3', while it is strongly disliked in P1' and P2'. This is in perfect agreement with our MALDI-based cleavage analysis of the gluten-derived immunogenic 33-mer, where the preferred cleavage site is PQP.QLP, which is present three times in the 33-mer, and gives rise to the prominent cleavage product of m/z 842.4 Da (QLPYPQP). Intriguingly, despite the high preference

and specificity for proline in P1, cleavage only occurs after 5 out of the 13 proline residues in the 33mer. This is probably due to the strong dislike of neprosin for proline in positions P1' and P2' (2.5 and 5x below natural abundance). Importantly, while neprosin readily degrades the 33-mer within minutes, pepsin fails to perform a single cut despite incubation overnight and the addition of 20x more enzyme than compared to neprosin, confirming the proteolytically-resistant properties of the immunogenic 33-mer.

It is well established that the 33-mer immunogenic peptide contains in turn six copies of immunodominant gluten peptides to which T cells responses are almost invariably identified in HLA-DQ2 positive patients (see Figure 27). Mass spectrometry is a very sensitive technique that permits highly sensitive and accurate detection and identification of degradation peptides. Our results demonstrate that neprosin is capable of removing six overlapping copies of the three stimulatory epitopes present within 33-mer fragment. However, it is reported that available enzyme supplements are only capable of removing a few N terminal aa from 26 and 33-mer gliadin fragments, leading to nine immunogenic epitopes intact (Janssen et al. 2015). These results show that neprosin is able to degrade the most toxic and immune-dominant peptide in wheat gluten, and which may cause coeliac disorders in susceptible individuals. Thus, it is highly likely that neprosin supplementation can be used to counteract the effect of gluten in gluten intolerant individuals.

In a next step, we designed a fluorogenic and crystallization peptide (QP.QL) based on the identified preferred cleavage site within the 33-mer. The kinetic parameters revealed high affinity and an extraordinary turnover rate. Similar enzymatic assays were performed with the so-called FS6 peptide. Although both substrates exhibited similar affinity values, neprosin showed a 30x higher specificity for QPQL than the FS6-peptide. Thus, we not only confirmed that neprosin is a highly active protease, but also identified the QPQL peptide as its ideal substrate for future studies.

### **G.2.2 Structural similarity with glutamate peptidases**

Most importantly, we solved the crystal structures of both the zymogen and the autoactivated mature form of neprosin. These structures revealed certain features that place neprosin in structural homology to the glutamic peptidase eqolysin such as a central  $\beta$ -sandwich composed by 7+8-stranded antiparallel  $\beta$ -sheets, a fold not previously observed for proteolytic enzymes. The crystal structure revealed a unique catalytic dyad composed by E<sup>188</sup> and E<sup>297</sup> which agrees with the structurally equivalent positioning of the eqolysin catalytic

dyad consisting of residues Q<sup>107</sup> and E<sup>190</sup>, which are highly conserved in fungal carboxyl peptidases.

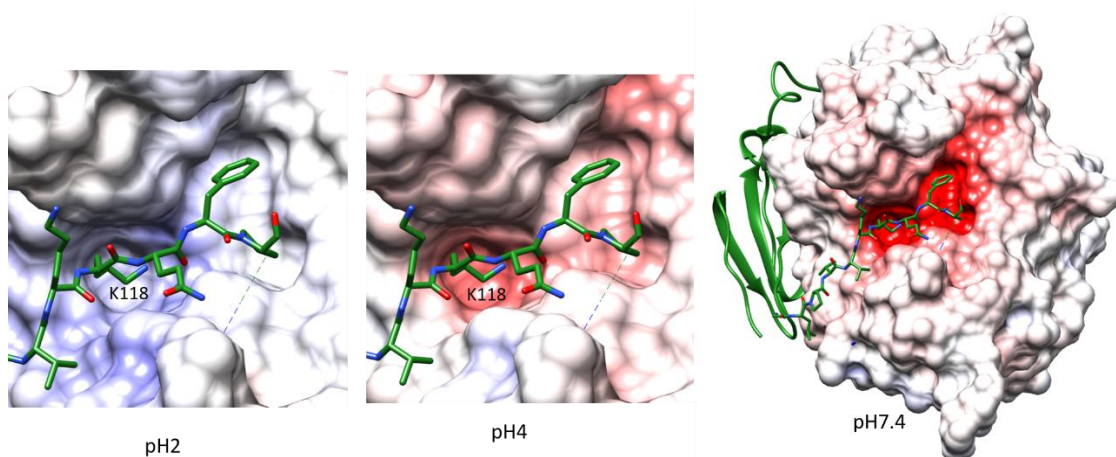
Using the DALI webserver (Citation PMID 18818215), we tried to identify structural homologs of neprosin. Even though 375 structures could be identified in the PDB90 (subset of the Protein Data Bank filtered at 90% sequence identity) with a Z-score of above 2 - typically indicating a similar fold, none of the matches had a sequence identity of above 15%. Thus no strong match could be identified. Interestingly, the Top10 hits included six sugar cleaving enzymes but also two peptidases, namely scytalidoglutamic peptidase (eqolysin; PDB ID 2ifw - Fujinaga et al. 2004) and aspergillopepsin II (APG; PDB ID 1y43 - Sasaki et al. 2004), which both belong to the MEROPS family G1 of glutamic peptidases. These two also have their catalytic residues Glu and Gln at the structurally equivalent positions to the two active site forming glutamates in neprosin. Thus, due to the low sequence similarity and the different composition of the active site, we are convinced that neprosin needs to be reassigned within the MEROPS database from U74 to form a novel glutamic peptidase family.

### **G.2.3 Mechanism of latency**

Based on our crystallographic analysis, we propose a specific auto-processing mechanism of pro-neprosin. Intriguingly, despite the structure similarity of the pro-neprosin with glutamic peptidase G1 family, the sequence of events in neprosin maturation displays a great resemblance to gastric aspartate-proteases and a lysine functionally equivalent to K<sup>118</sup> is found in pepsin and gastricsin.

Consequently, we based our hypothesis for the neprosin latency mechanism on the pH-dependent activation observed for pepsin (Richter, Tanaka, and Yada 1998) (Fig. 35). At neutral pH, the electrostatic interactions, particularly the ion pairs between the pro-segment residue K<sup>118</sup> and the two catalytic glutamates, E<sup>188</sup> and E<sup>297</sup>, maintain the zymogen in its inactive form, by stabilizing the position of the prodomain in the substrate binding cleft. This interaction along with the hydrogen bonds between Y<sup>214</sup> and E<sup>188</sup> and between Y<sup>156</sup> and E<sup>297</sup>, render the catalytic glutamates unavailable for catalysis at neutral pH. At pH values below 4, the acidic residues in the active enzyme become protonated, thus disrupting the electrostatic interactions between the prodomain and the active enzyme. Due to this repulsion of the net positive charge of K<sup>118</sup>, the active site is unblocked and the enzyme activated.

It is worth mentioning that the propeptide most probably runs along the catalytic cleft in the same direction as a substrate, given that active protein (see Figure 4 F,G in the manuscript) with its histidine tail entering a neighbouring molecule as a reference for an enzyme/substrate complex.

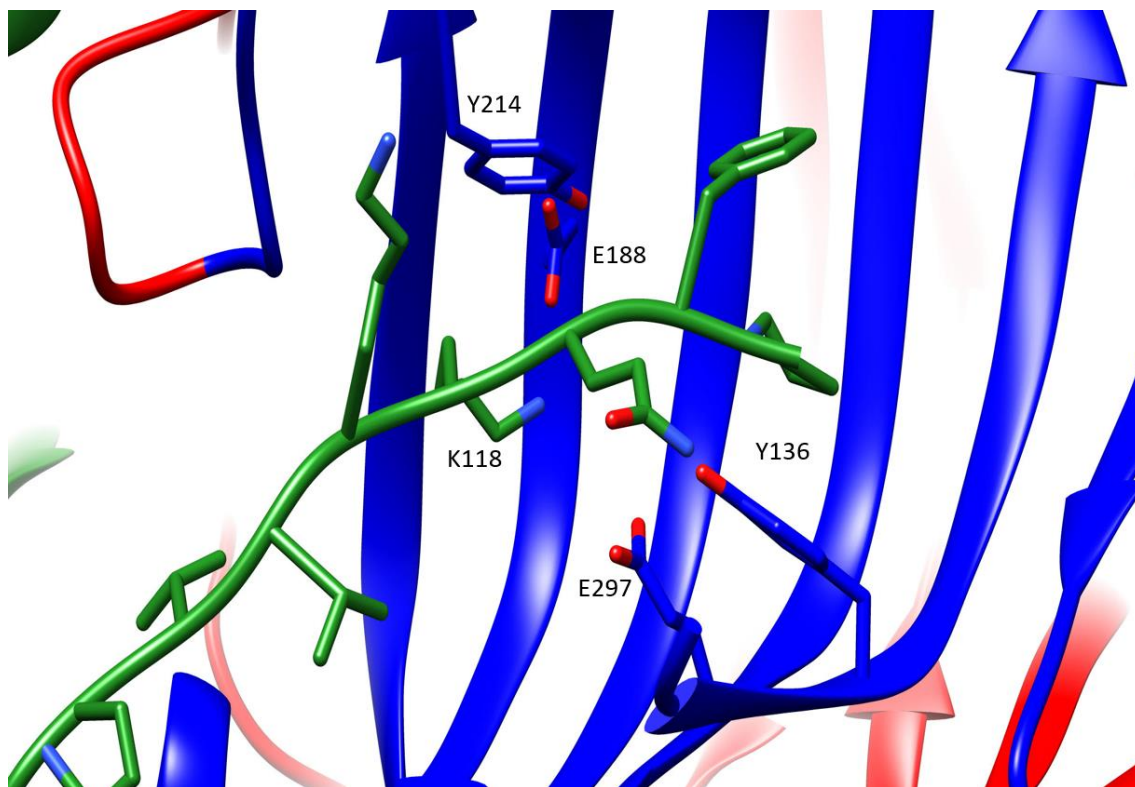


**Fig. 35. Surface electrostatic potential map of active site at different environmental pH.** The neprosin latency mechanism is based on pH dependence action. At pH 7.4, the active site cleft is charged negatively (red) while at pH 2, the active site presents a positive electrostatic potential which results in a protonation of both glutamate residues followed by a repulsion of the net positive charge giving by K<sup>118</sup> of the prodomain.

Moreover, we suggest that the major proteolytic action of activation occurs in “trans” by a second enzyme molecule once the prodomain linker becomes accessible for cleavage. Furthermore, we demonstrated the success of the autoactivation of catalytic mutants by the neprosin-strep-tag WT enzyme.

We indicated the S1 position of the cleft is occupied by K<sup>118</sup> from the prodomain linker at pH 7.5 and plays a fundamental role as a “locking” mechanism that holds the prodomain in place and prevents the proper positioning of a catalytic water molecule. This mechanism reminds of the pepsinogen structure (Uniprot P00791; PDB code: 2psg). Consequently, replacement of K<sup>118</sup> with alanine resulted in a mutant that was capable of activation at more alkaline pH values compared to that of the wild type. e.g. K<sup>118</sup>A is expressed as a homogeneous cleaved protease consisting of a mature and N-terminal prodomain (see figure 1C in the manuscript). We suggest that this mutation disrupted or altered the electrostatic interaction between K<sup>118</sup> and E<sup>188</sup>. Therefore, in the mutant, the prodomain might not be bound as tight to the active moiety, and due to this increased flexibility the zymogen would not need an acidic pH to undergo activation.





**Fig. 36.** Close up view of the pro-neprisin active site cleft and the bound prodomain in a standard orientation. Selected residues are displayed as coloured sticks and labelled. The bound V<sup>116</sup>-P<sup>121</sup> peptide (prodomain) is shown in green while the residue range after P<sup>121</sup> was not visible in the structure, thus indicating either that it was cleaved, or that it was just flexible.

#### G.2.4 Mutational analysis

In order to clarify the role of the residues which are in proximity of the catalytic dyad, we carried out mutational analysis. E<sup>188</sup>A, E<sup>188</sup>Q, E<sup>297</sup>A and E<sup>297</sup>Q mutants of pro-neprisin lost the auto-processing and the enzymatic activities of the wild type enzyme. Coupled with the results from the structural analysis of pro-neprisin, these biochemical and structural analysis confirmed that E<sup>188</sup> and E<sup>297</sup> are indeed the catalytic residues. The Y<sup>136</sup>A and Y<sup>214</sup>A mutants showed weak autocatalytic activity, suggesting the active site may be disturbed due to an impaired hydrogen bond network, and thus the enzymes are less active. In addition, they showed 36% and 15% of peptidase activity compared with that of the wild-type enzyme, respectively. Interestingly, K<sup>118</sup>A displays markedly reduced activity compared with the WT. We presume that K<sup>118</sup> in addition may have a folding function, as it was demonstrated for the prodomain itself (see above), so the absence of this residue may cause a slightly less stable variant. Another hypothesis would be that K<sup>118</sup> helps to form the active site so maybe in the

variant lacking the K<sup>118</sup>, the active site is just slightly disturbed but enough so that the distances don't work out and thus the enzyme is less active.

### **G.2.5 Inactivation/Inhibition trials**

Apparently, neprosin is inactive at pH 6 and above, however neprosin is not denatured or irreversibly inactivated with an alkaline pH (pH 9). In fact, neprosin in solutions of up to pH 9 can be reactivated upon re-acidification and shows the same affinity and efficiency for the substrate. Contrarily, eqolysin maintains its structure and activity in the pH range 2 to 7, but at pH 8 or higher it is irreversibly denatured as shown by circular dichroism (Kondo et al. 2010). According with the results, neprosin presents partial inhibition by pepstatin A and EPNP and displays resistance to DAN. All of them are typically specific inhibitors for aspartic proteases (Klein et al. 2018). Contrarily, APG and SCP-B are insensible to all aspartate inhibitors (Pillai et al. 2007; Takahashi 2013). Only SCP-B showed inhibition by EPNP. Thus, given that pepstatin A and 4-nitrophenoxy-based epoxides are inhibitors of pepsin-type aspartate endopeptidases, which share no sequence similarity with neprosin, this pointed to an unexpected peptidase type and mechanism of catalysis.

To conclude, oral supplementation with a prolyl endopeptidase could be a valuable method to eradicate the proline-rich T cell stimulatory epitopes from gluten-derived peptides. Importantly, our study showed that neprosin displays a high effectiveness in degrading gluten proteins such as gliadin and immune toxic peptides such as the infamous 33-mer at stomach-like pH conditions, thereby inactivating them before they can do harm in the small intestine of coeliac disease patients. Therefore, potential future work would be experiments in-vivo relying on the effect produced by gluten degradation products by neprosin and pepsin to intestinal cells and in a coeliac mouse model.

The major part of prolyl endoproteases proposed have the limitation of not displaying stability to the stomach pH or being cleaved by pepsin (Janssen et al. 2015). Additionally, a very low neprosin levels is required to achieve the solubilisation and great degradation of gluten proteins. Accordingly, the co-administration of neprosin with a gluten-containing meal thus may represent a feasible approach which may offers patients a viable alternative to the strict gluten-free diet, thereby tremendously improving their life quality, similar to lactase pills for lactose intolerant people. Even though the use of neprosin is not intended to fully replace a gluten-free diet, but it may act as a digestive aid protecting against the unintentional intake of gluten.

In the **third** project, we established a new high yield human expression system for the cytokine pro-TGF $\beta$ 2 given its importance and its high potential in many therapeutic applications. Furthermore, we report the X-ray crystal structure of its growth factor (GF) domain, which was obtained in a different crystallographic space group and also deviates substantially from current structures. Additionally, we examined the molecular mechanism by which pro-TGF $\beta$ 2 and h $\alpha$ <sub>2</sub>M interact since TGF $\beta$ 2 and h $\alpha$ <sub>2</sub>M can act as a growth factor and cytokine carrier, respectively and as this has been shown to have a great impact on human physiology.

We have designated a mammalian expression vector to overexpress a high amount of recombinant human pro-TGF $\beta$ 2 (UniProt ID P61812). To improve the production, a C<sup>24</sup>S mutation was created in the latency-associated peptide (LAP) region (L<sup>21</sup>-R<sup>302</sup>) and the natural leader sequence was replaced with that of mouse serum albumin. Finally, an eight histidine tag was inserted after the signal peptide to largely simplify the purification and in order to avoid interference with amino acid composition of mature TGF $\beta$ 2. The expression levels of this modified recombinant protein were at least 10 times higher than the C-terminal tagged-protein.

### **G.3 Interaction studies**

#### **G.3.1 TGF $\beta$ 2 Maturation**

TGF $\beta$ 2 is synthesized as a precursor (pro-TGF $\beta$ 2) which is partially cleaved in the Golgi apparatus at bond R<sup>302</sup>-A<sup>303</sup>. This cleavage leads to the production of mature TGF $\beta$ 2 (GF), which spans 112 residues (A<sup>303</sup>-S<sup>414</sup>) and is arranged as a disulphide-linked 25 kDa homodimer that is non-covalently associated with the LAP region and is only released upon activation (see figure 30).

The protein migrated as a dimer in SEC-MALLS (105.4) which indicates that the characteristic disulphide bonds were formed and that the LAP and GF moieties are still associated. However, our attempts to separate the two domains (e.g. by SEC in presence of high salt concentrations, chaotropic agents (urea), detergents, reducing agents and low pH buffers) failed. Importantly, the treatment with the physiological activating proprotein convertase furin and likewise with the endopeptidase trypsin, only led to the proteolysis of the full length specie into LAP and mature cytokine moiety. Although treating with proteases, we did not manage to separate the two domains since they remain associated by non-covalent interactions. Previous studies with a Chinese hamster ovary cell expression system benefited

from the sensitivity of the LAP domain towards denaturing conditions at very low pH to separate it from mature TGF $\beta$ 2 (Zou and Sun 2006). The reason why we could not manage to separate both domains is probably due to post-translation modifications introduced by Expi293F cells such a high glycosylation in LAP moiety (Croset et al. 2012).

### **G.3.2 Protein-protein interaction analysis**

Previous biochemical studies showed that h $\alpha$ <sub>2</sub>M binds predominantly to matured cytokine segment spanning residues A<sup>343</sup>-Y<sup>367</sup>, in which W<sup>354</sup> plays a major role (Q. Liu et al. 2001). However, no data have been reported on the role of LAP in the interaction. To shed further light, we set out to characterize the binding of pro-TGF $\beta$ 2 to h $\alpha$ <sub>2</sub>M through different techniques including SPR, SEC-MALLS, Sulfo-NHS-AMCA and BS<sup>3</sup> crosslinking. As a result, no interaction was detected except in Sulfo-NHS-AMCA assay with native h $\alpha$ <sub>2</sub>M. Thus, we concluded that LAP prevents h $\alpha$ <sub>2</sub>M from binding mature TGF $\beta$ 2 as suggested by structural studies on pro-TGF $\beta$ 1 which means that the interacting segment is partially shielded by LAP (Zhao et al. 2018). The fact that the interaction between both proteins did not occur was because we could not manage to activate pro-TGF $\beta$ 2.

During our crystallographic studies of interactions between induced h $\alpha$ <sub>2</sub>M and recombinant human pro-TGF $\beta$ 2, we conclude that the low pH of the crystallization condition, together with the crystallization process, actually achieved the separation of the LAP and GF moieties that we could not obtain by chromatography. We suggest that this process was probably further facilitated by a protease present in the purified h $\alpha$ <sub>2</sub>M preparation, as proteases trapped within the induced tetrameric h $\alpha$ <sub>2</sub>M cage are known to still possess proteolytic activity (Barrett and Starkey 1973; Sottrup-Jensen 1989; Marrero et al. 2012). Thus, despite using pro-TGFB2 and ha2M for crystallization, we obtained crystals containing only the TGF $\beta$  GF domain.

### **G. 3.3 Crystal structure of mature TGF $\beta$ 2**

Apart from assessing the interaction with h $\alpha$ <sub>2</sub>M, we aimed to obtain a crystal structure of pro-TGF $\beta$ 2. Based on previous studies, (Zhao et al. 2018) which reported the structure of pro-TGF $\beta$ 1 with the furin cleavage site mutated (RHRR<sup>249</sup>), we mutated R<sup>302</sup> into alanine to simulate the structure of the protein before furin cleavage. However, this mutant was not expressed suggesting that the furin cleavage site in TGF $\beta$ 2 has an impact on pro-TGF $\beta$ 2 biosynthesis or secretion.

Finally, we solved the crystal structure of mature TGF $\beta$ 2, confirming the dimeric behaviour in size exclusion chromatography. Intriguingly, our mature TGF $\beta$ 2 crystallized in a different space group, and the structure mainly deviated in the region around helix  $\alpha$ 3 from current mature TGF $\beta$ 1, TGF $\beta$ 2 and TGF $\beta$ 3 structures. Notably, this region corresponds to the segment that shows the highest sequence variability among TGF $\beta$ s (Moulin et al. 2014). The divergent conformation of this region results in a different arrangement of the dimer which in turn would not be able to bind its cognate receptors, TGFR-I and II, in the same way as suggested for TGF $\beta$ 1 and TGF $\beta$ 3. Although crystal packing artefacts can never be completely ruled out, the fact that the different crystal forms of the three TGF $\beta$  GFs resulted in highly similar structures suggests that our new conformation may be relevant in vivo and have true functional implications. Additionally, it may add a so far undescribed level of activity regulation to these cytokines. Overall, we concluded that our mature dimer structure may represent an inactive (active but “dormant”) variant or be one of an ensemble of conformational TGF $\beta$ 2 states, which may still undergo an “induced fit” or “selection fit” mechanism to form a functional ternary ligand-receptor complex.

## **CONCLUSIONS**

---



**Project 1** *“An integrative structural biology analysis of von Willebrand factor binding and processing by ADAMTS13 in homeostasis”*

- We comprehensively analysed the interplay between AD13 MDTCS E<sup>225</sup>Q and VWF-Strep-peptide through an integrative structural biology approach, using nine highly complementary techniques.
- Crosslinking mass spectrometry (CLMS) and hydrogen deuterium exchange (HDX-MS), as well as SAXS and SEC-MALLS analysis, demonstrated that the isolated proteins are highly flexible in solution. Importantly, this plasticity is maintained in the complex, which adopts a globular conformation with highly disordered segments while guiding VWF in the optimal orientation for proteolytic cleavage by AD13.
- AD13 MDTCS E<sup>225</sup>Q exists simultaneously in different, yet functionally relevant conformations. We suggest that thrombospondin domain 1 (T) acts as a crucial “spacer” segment within MDTCS which gives flexibility to the structural arrangement of the N-terminal (M+D) and C-terminal (C+S) domains.
- Our experimental insights highlight the AD13/VWF-Strep complex as a ‘fuzzy complex’ that results from multiple low-affinity interactions and small individual energy contributions, which simultaneously populate the interaction sites in a variety of ways. This implies transient interactions in nature.
- Based on our experimental data, we propose a model consisting of five molecular states that can fully explain the interaction interplay leading to the productive peptidase-substrate complex of AD13 and VWF.

**Project 2:** *“Structure, function, zymogenic latency and catalytic mechanism of a plant glutamate peptidase with therapeutic potential in coeliac disease”*

- We structurally characterized a novel prolyl endopeptidase called neprosin which forms a novel family within the glutamic peptidases due to its double Glu-motif in the active site instead of the Glu-Gln dyad found in eqolysins. The full-length zymogen was overexpressed in a mammalian expression system, purified and it was demonstrated that the protein activates itself under acidic pH conditions. The resulting mature protease spans residues S<sup>129</sup>-Q<sup>380</sup> (Uniprot ID C0HLV2) and it is the result of precise N-terminal processing removing the approx. 11 kDa prodomain.
- The functionality of recombinant neprosin was tested and proven against different substrates and inhibitors. Our data confirm that neprosin possesses high proteolytic



activity against proline rich proteins at low pH. Specifically, the hydrolysis of gluten-derived cleavage products such as the immunogenic 33-mer peptide of  $\alpha$ -gliadin was studied.

- To the best of our knowledge, neprosin is at present the only enzyme capable to remove the six overlapping copies of the three highly stimulatory T-cell epitopes present in 33-mer immunogenic peptide.
- Neprosin shares structural homology with scytailidocarboxyl peptidase B (SCP-B) and aspergiloglutamic peptidase (AGP), which both belong to the eqolysin family belonging to MEROPS family G1 of glutamic peptidases. However, we demonstrate that it also presents similar features to aspartic proteases, specifically pepsin and nephenthesins including a pH-dependent latency mechanism and partial inhibition by pepstatin A.
- The structure of pro-neprosin was solved by X-ray diffraction crystallography by using a lutetium derivative obtained by soaking and experimental SAD phasing. The crystal structure of the proform was solved to 1.80 Ångström resolution in space group  $P2_12_12$  with one molecule in the asymmetric unit (ASU). Subsequently, using molecular replacement (MR), we solved also the crystal structure of mature neprosin in space group  $P2_1$  at 2.35 Ångström resolution, and likewise with one molecule in the ASU.
- The overall structure of neprosin revealed a central “jelly roll”-like beta-sandwich consisting of 7+8-stranded antiparallel  $\beta$ -sheets, a rather uncommon fold for proteolytic enzymes. We then probed our crystal structures by site-directed mutagenesis studies and confirmed that the catalytic dyad indeed consists of a pair of glutamate residues (E<sup>188</sup> and E<sup>297</sup>), a catalytic dyad we believe is novel for peptidases.
- The crystal structures of neprosin further verified the presence of three disulfide bridges (Cys52-Cys98, Cys214-Cys224 and Cys359-Cys379) (Uniprot ID C0HLV2) and a crucial N-glycosylation pattern, which both greatly contribute to the stability of neprosin at stomach-like pH values and also under mild denaturing conditions.
- Pre-clinical studies such as *in vivo* experiments using intestinal cells or a celiac mouse model are required to ascertain if neprosin could indeed represent an effective supplement to support an adjuvant therapy or maybe even a vital alternative to a gluten-free diet in the treatment of celiac disease.

**Project 3:** *“Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor  $\beta$ 2 in a new conformation”*

- We established a mammalian protein expression system, which produced one of the highest yields of crystallization-grade pro-TGF $\beta$ 2 reported to date.
- Full-length pro-TGF $\beta$ 2 did not form a complex with any of the tested human  $\alpha$ <sub>2</sub>M variants, probably owing to steric hindrance by its N-terminal LAP domain.
- Intriguingly, we did not succeed to activate pro-TGF $\beta$ 2 using acidic pH conditions, possibly due to the Expi293F-derived pattern of glycosylation.
- Removal of the TGF $\beta$ 2 R<sup>302</sup>A furin activation site by mutation (R<sup>302</sup>A) resulted in complete abolishment of pro-TGF $\beta$ 2 expression.
- The crystal structure of homodimeric mature TGF $\beta$ 2 was solved to 2.0 Å resolution in a new space group by X-ray diffraction crystallography using molecular replacement. The analysis reveals a dimeric structure, which mainly deviates in the region around helix  $\alpha$ 3 from previously described TGF $\beta$ 1, TGF $\beta$ 2, TGF $\beta$ 3 structures. The deposited structure can be accessed at the Protein Data Bank as entry 6I9J.
- Our identified homodimeric structure of mature TGF $\beta$ 2 may represent an active but dormant conformation or one of an ensemble of possible conformational states, which may still undergo an “induced fit” or “selection fit” mechanism to form a functional ternary complex with its cognate receptor.



## **GENERAL REFERENCES**

---



## A

- Abadie, Valérie, Ludvig M. Sollid, Luis B. Barreiro, and Bana Jabri. 2011. 'Integration of Genetic and Immunological Insights into a Model of Celiac Disease Pathogenesis'. *Annual Review of Immunology* 29 (April): 493–525.
- Abe, M, N Oda, and Y Sato. 1998. 'Cell-Associated Activation of Latent Transforming Growth Factor-Beta by Calpain.' *Journal of Cellular Physiology* 174 (2): 186–93.
- Ai, Jihui, Paula Smith, Shuwei Wang, Ping Zhang, and X Long Zheng. 2005. 'The Proximal Carboxyl-Terminal Domains of ADAMTS13 Determine Substrate Specificity and Are All Required for Cleavage of von Willebrand Factor.' *The Journal of Biological Chemistry* 280 (33): 29428–34.
- Akiyama, Masashi, Soichi Takeda, Koichi Kokame, Junichi Takagi, and Toshiyuki Miyata. 2009. 'Crystal Structures of the Noncatalytic Domains of ADAMTS13 Reveal Multiple Discontinuous Exosites for von Willebrand Factor.' *Proceedings of the National Academy of Sciences of the United States of America* 106 (46): 19274–79.
- Anderson, Patricia J, Koichi Kokame, and J Evan Sadler. 2006. 'Zinc and Calcium Ions Cooperatively Modulate ADAMTS13 Activity.' *The Journal of Biological Chemistry* 281 (2): 850–57.
- Ang, Shirley, Jana Kogulanathan, Gordon A. Morris, M. Samil Kök, Peter R. Shewry, Arthur S. Tatham, Gary G. Adams, Arthur J. Rowe, and Stephen E. Harding. 2010. 'Structure and Heterogeneity of Gliadin: A Hydrodynamic Evaluation'. *European Biophysics Journal* 39 (2): 255–61.
- Aricescu, A R, R Assenberg, R M Bill, D Busso, V T Chang, S J Davis, A Dubrovsky, et al. 2006. 'Eukaryotic Expression: Developments for Structural Proteomics'. *Acta Crystallographica Section D* 62 (10): 1114–24.
- Armstrong, P B, and J P Quigley. 1999. 'Alpha2-Macroglobulin: An Evolutionarily Conserved Arm of the Innate Immune System'. *Developmental and Comparative Immunology* 23 (4–5): 375–90.
- Athauda, Senarath B P, Koji Matsumoto, Sanath Rajapakshe, Masayuki Kuribayashi, Masaki Kojima, Nobuko Kubomura-Yoshida, Akihiro Iwamatsu, Chiaki Shibata, Hideshi Inoue, and Kenji Takahashi. 2004. 'Enzymic and Structural Characterization of Nepenthesin, a Unique Member of a Novel Subfamily of Aspartic Proteinases.' *The Biochemical Journal* 381 (Pt 1): 295–306.

## B

- Balakireva, Anastasia V, and Andrey A Zamyatnin. 2016. 'Properties of Gluten Intolerance: Gluten Structure, Evolution, Pathogenicity and Detoxification Capabilities.' *Nutrients* 8 (10).
- Bandaranayake, Ashok D, and Steven C Almo. 2014. 'Recent Advances in Mammalian Protein Production.' *FEBS Letters* 588 (2): 253–60.
- Baneyx, François, and Mirna Mujacic. 2004. 'Recombinant Protein Folding and Misfolding in Escherichia Coli'. *Nature Biotechnology* 22 (11): 1399–1407.
- Bardor, Muriel, Dzung H Nguyen, Sandra Diaz, and Ajit Varki. 2005. 'Mechanism of Uptake and Incorporation of the Non-Human Sialic Acid N-Glycolylneuraminic Acid into Human Cells.' *The Journal of Biological Chemistry* 280 (6): 4228–37.

- Barrett, A J. 1981. 'Alpha 2-Macroglobulin.' *Methods in Enzymology* 80 Pt C: 737–54.
- Barrett, A J, and P M Starkey. 1973. 'The Interaction of Alpha 2-Macroglobulin with Proteinases. Characteristics and Specificity of the Reaction, and a Hypothesis Concerning Its Molecular Mechanism.' *The Biochemical Journal* 133 (4): 709–24.
- Barsoum, J, R Brown, M McKee, and F M Boyce. 1997. 'Efficient Transduction of Mammalian Cells by a Recombinant Baculovirus Having the Vesicular Stomatitis Virus G Glycoprotein.' *Human Gene Therapy* 8 (17): 2011–18.
- Barton, William A, Dorothea Tzvetkova-Robev, Hediye Erdjument-Bromage, Paul Tempst, and Dimitar B Nikolov. 2006. 'Highly Efficient Selenomethionine Labeling of Recombinant Proteins Produced in Mammalian Cells.' *Protein Science: A Publication of the Protein Society* 15 (8): 2008–13.
- Battegay, E J, E W Raines, R A Seifert, D F Bowen-Pope, and R Ross. 1990. 'TGF-Beta Induces Bimodal Proliferation of Connective Tissue Cells via Complex Control of an Autocrine PDGF Loop.' *Cell* 63 (3): 515–24.
- Bazile, Vincent, Gilles Le Moguédec, David J Marshall, and Laurence Gaume. 2015. 'Fluid Physico-Chemical Properties Influence Capture and Diet in Nepenthes Pitcher Plants.' *Annals of Botany* 115 (4): 705–16.
- Beinrohr, László, József Dobó, Péter Závodszky, and Péter Gál. 2008. 'C1, MBL-MASPs and C1-Inhibitor: Novel Approaches for Targeting Complement-Mediated Inflammation.' *Trends in Molecular Medicine* 14 (12): 511–21.
- Bernardo, A, C Ball, L Nolasco, H Choi, J L Moake, and J F Dong. 2005. 'Platelets Adhered to Endothelial Cell-Bound Ultra-Large von Willebrand Factor Strings Support Leukocyte Tethering and Rolling under High Shear Stress.' *Journal of Thrombosis and Haemostasis: JTH* 3 (3): 562–70.
- Bertozzi CR, Rabuka D. Structural Basis of Glycan Diversity. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, editors. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Chapter 2.
- Bethune, Michael T, and Chaitan Khosla. 2012. 'Oral Enzyme Therapy for Celiac Sprue.' *Methods in Enzymology* 502: 241–71.
- Bethune, Michael T, Pavel Strop, Yinyan Tang, Ludvig M Sollid, and Chaitan Khosla. 2006. 'Heterologous Expression, Purification, Refolding, and Structural-Functional Characterization of EP-B2, a Self-Activating Barley Cysteine Endoprotease.' *Chemistry & Biology* 13 (6): 637–47.
- Bienkowska, J, M Cruz, A Atiemo, R Handin, and R Liddington. 1997. 'The von Willebrand Factor A3 Domain Does Not Contain a Metal Ion-Dependent Adhesion Site Motif.' *The Journal of Biological Chemistry* 272 (40): 25162–67.
- Bishop, Benjamin, A Radu Aricescu, Karl Harlos, Chris A O'Callaghan, E Yvonne Jones, and Christian Siebold. 2009. 'Structural Insights into Hedgehog Ligand Sequestration by the Human Hedgehog-Interacting Protein HHIP.' *Nature Structural & Molecular Biology* 16 (7): 698–703.
- Blombery, Piers, and Marie Scully. 2014. 'Management of Thrombotic Thrombocytopenic Purpura: Current Perspectives.' *Journal of Blood Medicine* 5: 15–23.

- Bode, W, F X Gomis-Rüth, R Huber, R Zwillig, and W Stöcker. 1992. 'Structure of Astacin and Implications for Activation of Astacins and Zinc-Ligation of Collagenases.' *Nature* 358 (6382): 164–67.
- Bode, W, F X Gomis-Ruth, and W Stockler. 1993. 'Astacins, Serralysins, Snake Venom and Matrix Metalloproteinases Exhibit Identical Zinc-Binding Environments (HEXXHXXGXXH and Met-Turn) and Topologies and Should Be Grouped into a Common Family, the "Metzincins".' *FEBS Letters* 331 (1–2): 134–40.
- Bode, W, and R Huber. 2000. 'Structural Basis of the Endoproteinase-Protein Inhibitor Interaction.' *Biochimica et Biophysica Acta* 1477 (1–2): 241–52.
- Boettner, Mewes, Bianka Prinz, Caterina Holz, Ulf Stahl, and Christine Lang. 2002. 'High-Throughput Screening for Expression of Heterologous Proteins in the Yeast *Pichia Pastoris*' *Journal of Biotechnology* 99 (1): 51–62.
- Boohaker, R J, M W Lee, P Vishnubhotla, J M Perez, and A R Khaled. 2012. 'The Use of Therapeutic Peptides to Target and to Kill Cancer Cells.' *Current Medicinal Chemistry* 19 (22): 3794–3804.
- Brew, K, D Dinakarpanian, and H Nagase. 2000. 'Tissue Inhibitors of Metalloproteinases: Evolution, Structure and Function.' *Biochimica et Biophysica Acta* 1477 (1–2): 267–83.
- Broeck, Hetty C van den, Teun WJM van Herpen, Cees Schuit, Elma MJ Salentijn, Liesbeth Dekking, Dirk Bosch, Rob J Hamer, Marinus JM Smulders, Ludovicus JWJ Gilissen, and Ingrid M van der Meer. 2009. 'BMC Plant Biology Removing Celiac Disease-Related Gluten Proteins from Bread Wheat While Retaining Technological Properties: A Study with Chinese Spring Deletion Lines'.
- Brown, P D, L M Wakefield, A D Levinson, and M B Sporn. 1990. 'Physicochemical Activation of Recombinant Latent Transforming Growth Factor-Beta's 1, 2, and 3.' *Growth Factors (Chur, Switzerland)* 3 (1): 35–43.
- Bryan, Philip N. 2002. 'Prodomains and Protein Folding Catalysis.' *Chemical Reviews* 102 (12): 4805–16.
- Buch, Franziska, Matthias Rott, Sandy Rottloff, Christian Paetz, Ines Hilke, Michael Raessler, and Axel Mithöfer. 2013. 'Secreted Pitfall-Trap Fluid of Carnivorous Nepenthes Plants Is Unsuitable for Microbial Growth.' *Annals of Botany* 111 (3): 375–83.
- Butler, Michael, and Maureen Spearman. 2014. 'The Choice of Mammalian Cell Host and Possibilities for Glycosylation Engineering.' *Current Opinion in Biotechnology* 30 (December): 107–12.

## C

- Camilleri, R S, M Scully, M Thomas, I J Mackie, R Liesner, W J Chen, K Manns, and S J Machin. 2012. 'A Phenotype-Genotype Correlation of ADAMTS13 Mutations in Congenital Thrombotic Thrombocytopenic Purpura Patients Treated in the United Kingdom.' *Journal of Thrombosis and Haemostasis : JTH* 10 (9): 1792–1801.
- Cao, Wenjing, Sriram Krishnaswamy, Rodney M Camire, Peter J Lenting, and X Long Zheng. 2008. 'Factor VIII Accelerates Proteolytic Cleavage of von Willebrand Factor by ADAMTS13.' *Proceedings of the National Academy of Sciences of the United States of America* 105 (21): 7416–21.
- Carrijo, Linda Christian, Filipe Andrich, Maria Elena de Lima, Marta N Cordeiro, Michael



- Richardson, and Suely G Figueiredo. 2005. 'Biological Properties of the Venom from the Scorpionfish (*Scorpaena Plumieri*) and Purification of a Gelatinolytic Protease.' *Toxicon : Official Journal of the International Society on Toxinology* 45 (7): 843–50.
- Cater, Jordan H, Mark R Wilson, and Amy R Wyatt. 2019. 'Alpha-2-Macroglobulin, a Hypochlorite-Regulated Chaperone and Immune System Modulator'. Edited by Sander Bekeschus. *Oxidative Medicine and Cellular Longevity* 2019: 5410657.
- Cerdà-Costa, Núria, and Francesc Xavier Gomis-Rüth. 2014. 'Architecture and Function of Metallopeptidase Catalytic Domains.' *Protein Science : A Publication of the Protein Society* 23 (2): 123–44.
- Cox, Manon M J. 2012. 'Recombinant Protein Vaccines Produced in Insect Cells.' *Vaccine* 30 (10): 1759–66.
- Crawford, S E, V Stellmach, J E Murphy-Ullrich, S M Ribeiro, J Lawler, R O Hynes, G P Boivin, and N Bouck. 1998. 'Thrombospondin-1 Is a Major Activator of TGF-Beta1 in Vivo.' *Cell* 93 (7): 1159–70.
- Crawley, James T B, Rens de Groot, Yaozu Xiang, Brenda M Luken, and David A Lane. 2011. 'Unraveling the Scissile Bond: How ADAMTS13 Recognizes and Cleaves von Willebrand Factor.' *Blood* 118 (12): 3212–21.
- Crawley, James T B, and Marie A Scully. 2013. 'Thrombotic Thrombocytopenic Purpura: Basic Pathophysiology and Therapeutic Strategies.' *Hematology. American Society of Hematology. Education Program* 2013: 292–99.
- Crispin, Max, Thomas A Bowden, Charlotte H Coles, Karl Harlos, A Radu Aricescu, David J Harvey, David I Stuart, and E Yvonne Jones. 2009. 'Carbohydrate and Domain Architecture of an Immature Antibody Glycoform Exhibiting Enhanced Effector Functions.' *Journal of Molecular Biology* 387 (5): 1061–66.
- Crookston, K P, D J Webb, B B Wolf, and S L Gonias. 1994. 'Classification of Alpha 2-Macroglobulin-Cytokine Interactions Based on Affinity of Noncovalent Association in Solution under Apparent Equilibrium Conditions.' *The Journal of Biological Chemistry* 269 (2): 1533–40.
- Croset, Amelie, Laurence Delafosse, Jean-Philippe Gaudry, Christian Arod, Loic Glez, Christophe Losberger, Damien Begue, et al. 2012. 'Differences in the Glycosylation of Recombinant Proteins Expressed in HEK and CHO Cells.' *Journal of Biotechnology* 161 (3): 336–48.
- D**
- Davis, S J, M J Puklavec, D A Ashford, K Harlos, E Y Jones, D I Stuart, and A F Williams. 1993. 'Expression of Soluble Recombinant Glycoproteins with Predefined Glycosylation: Application to the Crystallization of the T-Cell Glycoprotein CD2.' *Protein Engineering* 6 (2): 229–32.
- Deforche, L, E Roose, A Vandebulcke, N Vandeputte, H B Feys, T A Springer, L Z Mi, et al. 2015. 'Linker Regions and Flexibility around the Metalloprotease Domain Account for Conformational Activation of ADAMTS-13.' *Journal of Thrombosis and Haemostasis : JTH* 13 (11): 2063–75.
- Denis, Cecile V, and Peter J Lenting. 2012. 'Von Willebrand Factor: At the Crossroads of Bleeding and Thrombosis.' *International Journal of Hematology* 95 (4): 353–61.

- Deu, Edgar, Martijn Verdoes, and Matthew Bogyo. 2012. 'New Approaches for Dissecting Protease Functions to Improve Probe Development and Drug Discovery.' *Nature Structural & Molecular Biology* 19 (1): 9–16.
- Dicke, W K, J H Van de Kamer, and H A Weijers. 1957. 'Celiac Disease.' *Advances in Pediatrics* 9: 277–318.
- Dong, Jing-fei, Miguel A Cruz, Khatira Aboulfatova, Cecilia Martin, Hiuwan Choi, Angela L Bergeron, Sheryl R Martini, Michael H Kroll, and Thomas A Kent. 2008. 'Magnesium Maintains Endothelial Integrity, up-Regulates Proteolysis of Ultra-Large von Willebrand Factor, and Reduces Platelet Aggregation under Flow Conditions.' *Thrombosis and Haemostasis* 99 (3): 586–93.
- Dong, Jing-fei, Joel L Moake, Leticia Nolasco, Aubrey Bernardo, Wendy Arceneaux, Corie N Shrimpton, Alicia J Schade, Larry V McIntire, Kazuo Fujikawa, and Jose A Lopez. 2002. 'ADAMTS-13 Rapidly Cleaves Newly Secreted Ultralarge von Willebrand Factor Multimers on the Endothelial Surface under Flowing Conditions.' *Blood* 100 (12): 4033–39.
- Douglas, Kenneth W, Kevin G J Pollock, David Young, Jamie Catlow, and Rachel Green. 2010. 'Infection Frequently Triggers Thrombotic Microangiopathy in Patients with Preexisting Risk Factors: A Single-Institution Experience.' *Journal of Clinical Apheresis* 25 (2): 47–53.
- Drag, Marcin, and Guy S Salvesen. 2010. 'Emerging Principles in Protease-Based Drug Discovery.' *Nature Reviews. Drug Discovery* 9 (9): 690–701.
- Durham, Timothy B, Valentine J Klimkowski, Christopher J Rito, Jothirajah Marimuthu, James L Toth, Chin Liu, Jim D Durbin, et al. 2014. 'Identification of Potent and Selective Hydrantoin Inhibitors of Aggrecanase-1 and Aggrecanase-2 That Are Efficacious in Both Chemical and Surgical Models of Osteoarthritis.' *Journal of Medicinal Chemistry* 57 (24): 10476–85.

## E

- Eder, J, and A R Fersht. 1995. 'Pro-Sequence-Assisted Protein Folding.' *Molecular Microbiology* 16 (4): 609–14.
- Edmonds, Mary. 2002. 'A History of Poly A Sequences: From Formation to Factors to Function.' *Progress in Nucleic Acid Research and Molecular Biology* 71: 285–389.
- Eilenberg, Haviva, Smadar Pnini-Cohen, Silvia Schuster, Anna Movtchan, and Aviah Zilberstein. 2006. 'Isolation and Characterization of Chitinase Genes from Pitchers of the Carnivorous Plant *Nepenthes Khasiana*.' *Journal of Experimental Botany* 57 (11): 2775–84.
- Elzoghby, Ahmed O, Wael M Samy, and Nazik A Elgindy. 2012. 'Protein-Based Nanocarriers as Promising Drug and Gene Delivery Systems.' *Journal of Controlled Release* 161 (1): 38–49.
- Emsley, J, M Cruz, R Handin, and R Liddington. 1998. 'Crystal Structure of the von Willebrand Factor A1 Domain and Implications for the Binding of Platelet Glycoprotein Ib.' *The Journal of Biological Chemistry* 273 (17): 10396–401.
- Estes, Scott, and Mark Melville. 2014. 'Mammalian Cell Line Developments in Speed and Efficiency.' *Advances in Biochemical Engineering/ Biotechnology* 139: 11–33.

## F

- Farady, Christopher J, and Charles S Craik. 2010. 'Mechanisms of Macromolecular Protease Inhibitors.' *ChemBiochem : A European Journal of Chemical Biology* 11 (17): 2341–46.
- Federici, A B, R Bader, S Pagani, M L Colibretti, L De Marco, and P M Mannucci. 1989. 'Binding of von Willebrand Factor to Glycoproteins Ib and IIb/IIIa Complex: Affinity Is Related to Multimeric Size.' *British Journal of Haematology* 73 (1): 93–99.
- Feige, J J, A Negoescu, M Keramidas, S Souchelnitskiy, and E M Chambaz. 1996. 'Alpha 2-Macroglobulin: A Binding Protein for Transforming Growth Factor-Beta and Various Cytokines.' *Hormone Research* 45 (3–5): 227–32.
- Feng, Shuju, Xiaowen Liang, Miguel A Cruz, Hangoc Vu, Zhou Zhou, Naresh Pemmaraju, Jing-Fei Dong, Michael H Kroll, and Vahid Afshar-Kharghan. 2013. 'The Interaction between Factor H and Von Willebrand Factor.' *PloS One* 8 (8): e73715. <https://doi.org/10.1371/journal.pone.0073715>.
- Fischer, B E, G Kramer, A Mitterer, L Grillberger, M Reiter, W Mundt, F Dorner, and J Eibl. 1996. 'Effect of Multimerization of Human and Recombinant von Willebrand Factor on Platelet Aggregation, Binding to Collagen and Binding of Coagulation Factor VIII.' *Thrombosis Research* 84 (1): 55–66.
- Fischer, Simon, Nadine Charara, Andrea Gerber, Jens Wolfel, Gudrun Schiedner, Bernd Voedisch, and Sabine Geisse. 2012. 'Transient Recombinant Protein Expression in a Human Amniocyte Cell Line: The CAP-T(R) Cell System.' *Biotechnology and Bioengineering* 109 (9): 2250–61.
- Fontana, Laura, Yan Chen, Petra Prijatelj, Takao Sakai, Reinhard Fässler, Lynn Y Sakai, and Daniel B Rifkin. 2005. 'Fibronectin Is Required for Integrin Alphavbeta6-Mediated Activation of Latent TGF-Beta Complexes Containing LTBP-1.' *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 19 (13): 1798–1808.
- Fox, Jay W, and Solange M T Serrano. 2009. 'Timeline of Key Events in Snake Venom Metalloproteinase Research.' *Journal of Proteomics* 72 (2): 200–209.
- FRAZER, A C, R F FLETCHER, C A ROSS, B SHAW, H G SAMMONS, and R SCHNEIDER. 1959. 'Gluten-Induced Enteropathy: The Effect of Partially Digested Gluten.' *Lancet (London, England)* 2 (7097): 252–55.
- Fujikawa, K, H Suzuki, B McMullen, and D Chung. 2001. 'Purification of Human von Willebrand Factor-Cleaving Protease and Its Identification as a New Member of the Metalloproteinase Family.' *Blood* 98 (6): 1662–66.
- Fujinaga, Masao, Maia M Cherney, Hiroshi Oyama, Kohei Oda, and Michael N G James. 2004. 'The Molecular Structure and Catalytic Mechanism of a Novel Carboxyl Peptidase from *Scytalidium lignicolum*.' *Proceedings of the National Academy of Sciences of the United States of America* 101 (10): 3364–69.
- Furlan, M, R Robles, B Morselli, P Sandoz, and B Lammle. 1999. 'Recovery and Half-Life of von Willebrand Factor-Cleaving Protease after Plasma Therapy in Patients with Thrombotic Thrombocytopenic Purpura.' *Thrombosis and Haemostasis* 81 (1): 8–13.

## G

- Galili, Uri. 2004. 'Immune Response, Accommodation, and Tolerance to Transplantation Carbohydrate Antigens.' *Transplantation* 78 (8): 1093–98.
- Ganapathy, Vaidyanathan, Joel W Hay, and Jae H Kim. 2012. 'Costs of Necrotizing Enterocolitis and Cost-Effectiveness of Exclusively Human Milk-Based Products in Feeding Extremely Premature Infants.' *Breastfeeding Medicine: The Official Journal of the Academy of Breastfeeding Medicine* 7 (1): 29–37.
- Gao, Weiqiang, Patricia J Anderson, Elaine M Majerus, Elodee A Tuley, and J Evan Sadler. 2006. 'Exosite Interactions Contribute to Tension-Induced Cleavage of von Willebrand Factor by the Antithrombotic ADAMTS13 Metalloprotease.' *Proceedings of the National Academy of Sciences of the United States of America* 103 (50): 19099–104.
- Gao, Weiqiang, Patricia J Anderson, and J Evan Sadler. 2008. 'Extensive Contacts between ADAMTS13 Exosites and von Willebrand Factor Domain A2 Contribute to Substrate Specificity.' *Blood* 112 (5): 1713–19.
- Garcia-Ferrer, Irene, Aniebrys Marrero, F Xavier Gomis-Rüth, and Theodoros Goulas. 2017. ' $\alpha(2)$ -Macroglobulins: Structure and Function.' *Sub-Cellular Biochemistry* 83: 149–83.
- Giblin, Jonathan P, Lindsay J Hewlett, and Matthew J Hannah. 2008. 'Basal Secretion of von Willebrand Factor from Human Endothelial Cells.' *Blood* 112 (4): 957–64.
- Gil-Humanes, Javier, Fernando Pistón, Rossana Altamirano-Fortoul, Ana Real, Isabel Comino, Carolina Sousa, Cristina M Rosell, and Francisco Barro. 2014. 'Reduced-Gliadin Wheat Bread: An Alternative to the Gluten-Free Diet for Consumers Suffering Gluten-Related Pathologies.' *PLoS One* 9 (3): e90898.
- Gil, A, and N J Proudfoot. 1987. 'Position-Dependent Sequence Elements Downstream of AAUAAA Are Required for Efficient Rabbit Beta-Globin mRNA 3' End Formation.' *Cell* 49 (3): 399–406.
- Gleizes, P E, J S Munger, I Nunes, J G Harpel, R Mazzieri, I Noguera, and D B Rifkin. 1997. 'TGF-Beta Latency: Biological Significance and Mechanisms of Activation.' *Stem Cells (Dayton, Ohio)* 15 (3): 190–97.
- Gluzman, Y. 1981. 'SV40-Transformed Simian Cells Support the Replication of Early SV40 Mutants.' *Cell* 23 (1): 175–82.
- Goeddel, D V, D G Kleid, F Bolivar, H L Heyneker, D G Yansura, R Crea, T Hirose, A Kraszewski, K Itakura, and A D Riggs. 1979. 'Expression in Escherichia Coli of Chemically Synthesized Genes for Human Insulin.' *Proceedings of the National Academy of Sciences of the United States of America* 76 (1): 106–10.
- Gomis-Ruth, F Xavier. 2009. 'Catalytic Domain Architecture of Metzincin Metalloproteases.' *The Journal of Biological Chemistry* 284 (23): 15353–57.
- Gornik, Olga, and Gordan Lauc. 2008. 'Glycosylation of Serum Proteins in Inflammatory Diseases.' *Disease Markers* 25 (4–5): 267–78.
- Gray, A M, and A J Mason. 1990. 'Requirement for Activin A and Transforming Growth Factor--Beta 1 pro-Regions in Homodimer Assembly.' *Science (New York, N.Y.)* 247 (4948): 1328–30.

- Groot, Rens de, Ajoy Bardhan, Nalisha Ramroop, David A Lane, and James T B Crawley. 2009. 'Essential Role of the Disintegrin-like Domain in ADAMTS13 Function.' *Blood* 113 (22): 5609–16.
- Groot, Rens de, David A Lane, and James T B Crawley. 2010. 'The ADAMTS13 Metalloprotease Domain: Roles of Subsites in Enzyme Activity and Specificity.' *Blood* 116 (16): 3064–72.
- Gujral, Naiyana, Hugh J Freeman, and Alan B R Thomson. 2012. 'Celiac Disease: Prevalence, Diagnosis, Pathogenesis and Treatment.' *World Journal of Gastroenterology* 18 (42): 6036–59.
- Guo, Cenqi, Anastasia Tsigkou, and Meng Huee Lee. 2016. 'ADAMTS13 and 15 Are Not Regulated by the Full Length and N-Terminal Domain Forms of TIMP-1, -2, -3 and -4.' *Biomedical Reports* 4 (1): 73–78.
- Gutiérrez-Granados, Sonia, Laura Cervera, Amine A Kamen, and Francesc Gòdia. 2018. 'Advancements in Mammalian Cell Transient Gene Expression (TGE) Technology for Accelerated Production of Biologics.' *Critical Reviews in Biotechnology* 38 (6): 918–40.

## H

- Hannig, G, and S C Makrides. 1998. 'Strategies for Optimizing Heterologous Protein Expression in Escherichia Coli.' *Trends in Biotechnology* 16 (2): 54–60.
- Hanson, S R, and K S Sakariassen. 1998. 'Blood Flow and Antithrombotic Drug Effects.' *American Heart Journal* 135 (5 Pt 2 Su): S132-45.
- Hatano, Naoya, and Tatsuro Hamada. 2008. 'Proteome Analysis of Pitcher Fluid of the Carnivorous Plant *Nepenthes Alata*.' *Journal of Proteome Research*. United States.
- Hatano N, Hamada T. Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. *J Proteomics*. 2012 Aug 3;75(15):4844-52.
- Hausch, Felix, Lu Shan, Nilda A Santiago, Gary M Gray, and Chaitan Khosla. 2002. 'Intestinal Digestive Resistance of Immunodominant Gliadin Peptides.' *American Journal of Physiology. Gastrointestinal and Liver Physiology* 283 (4): G996–1003.
- Herrera, María G, Fernando Zamarreño, Marcelo Costabel, Hernan Ritacco, Andreas Hütten, Norbert Sewald, and Verónica I Dodero. 2014. 'Circular Dichroism and Electron Microscopy Studies in Vitro of 33-Mer Gliadin Peptide Revealed Secondary Structure Transition and Supramolecular Organization.' *Biopolymers* 101 (1): 96–106.
- Holz, Caterina, Bianka Prinz, Natalia Bolotina, Volker Sievert, Konrad Bussow, Bernd Simon, Ulf Stahl, and Christine Lang. 2003. 'Establishing the Yeast *Saccharomyces Cerevisiae* as a System for Expression of Human Proteins on a Proteome-Scale.' *Journal of Structural and Functional Genomics* 4 (2–3): 97–108.
- Huang, Xiaofen, Xin Wang, Jun Zhang, Ningshao Xia, and Qinjian Zhao. 2017. 'Escherichia Coli-Derived Virus-like Particles in Vaccine Development.' *NPJ Vaccines* 2: 3.
- Hunt, Ian. 2005. 'From Gene to Protein: A Review of New and Enabling Technologies for Multi-Parallel Protein Expression.' *Protein Expression and Purification* 40 (1): 1–22.
- Hunter, Molly, Ping Yuan, Divya Vavilala, and Mark Fox. 2019. 'Optimization of Protein Expression in Mammalian Cells.' *Current Protocols in Protein Science* 95 (1): e77.

## I

Ikonomou, L, Y-J Schneider, and S N Agathos. 2003. 'Insect Cell Culture for Industrial Production of Recombinant Proteins.' *Applied Microbiology and Biotechnology* 62 (1): 1–20.

## J

Jabri, B, N P de Serre, C Cellier, K Evans, C Gache, C Carvalho, J F Mougnot, et al. 2000. 'Selective Expansion of Intraepithelial Lymphocytes Expressing the HLA-E-Specific Natural Killer Receptor CD94 in Celiac Disease.' *Gastroenterology* 118 (5): 867–79.

Janssen, George, Chantal Christis, Yvonne Kooy-Winkelaar, Luppó Edens, Drew Smith, Peter van Veelen, and Frits Koning. 2015. 'Ineffective Degradation of Immunogenic Gluten Epitopes by Currently Available Digestive Enzyme Supplements.' *PLoS One* 10 (6): e0128065.

Jenkins, Gisli. 2008. 'The Role of Proteases in Transforming Growth Factor- $\beta$  Activation'. *The International Journal of Biochemistry & Cell Biology* 40 (6): 1068–78.

Jian, Cui, Juan Xiao, Lingjie Gong, Christopher G Skipwith, Sheng-Yu Jin, Hau C Kwaan, and X Long Zheng. 2012. 'Gain-of-Function ADAMTS13 Variants That Are Resistant to Autoantibodies against ADAMTS13 in Patients with Acquired Thrombotic Thrombocytopenic Purpura.' *Blood* 119 (16): 3836–43.

Jin, Sheng-Yu, Christopher G Skipwith, and X Long Zheng. 2010. 'Amino Acid Residues Arg(659), Arg(660), and Tyr(661) in the Spacer Domain of ADAMTS13 Are Critical for Cleavage of von Willebrand Factor.' *Blood* 115 (11): 2300–2310.

Jones, David, Nathalie Kroos, Regina Anema, Bart van Montfort, Andre Vooy, Sven van der Kraats, Esmeralda van der Helm, et al. 2003. 'High-Level Expression of Recombinant IgG in the Human Cell Line per.C6.' *Biotechnology Progress* 19 (1): 163–68.

Jones, Elizabeth A V, Ferdinand le Noble, and Anne Eichmann. 2006. 'What Determines Blood Vessel Structure? Genetic Prespecification vs. Hemodynamics.' *Physiology (Bethesda, Md.)* 21 (December): 388–95.

Jusko, Monika, Jan Potempa, Abdulkarim Y Karim, Miroslaw Ksiazek, Kristian Riesbeck, Peter Garred, Sigrun Eick, and Anna M Blom. 2012. 'A Metalloproteinase Karilysin Present in the Majority of *Tannerella Forsythia* Isolates Inhibits All Pathways of the Complement System.' *Journal of Immunology (Baltimore, Md. : 1950)* 188 (5): 2338–49.

## K

Kadek, Alan, Hynek Mrazek, Petr Halada, Martial Rey, David C Schriemer, and Petr Man. 2014. 'Aspartic Protease Nepenthesin-1 as a Tool for Digestion in Hydrogen/Deuterium Exchange Mass Spectrometry.' *Analytical Chemistry* 86 (9): 4287–94.

Kadek, Alan, Vyacheslav Tretyachenko, Hynek Mrazek, Ljubina Ivanova, Petr Halada, Martial Rey, David C Schriemer, and Petr Man. 2014. 'Expression and Characterization of Plant Aspartic Protease Nepenthesin-1 from *Nepenthes Gracilis*.' *Protein Expression and Purification* 95 (March): 121–28.

Karsdal, Morten A, Lykke Larsen, Michael T Engsig, Henriette Lou, Mercedes Ferreras, Andre Lochter, Jean-Marie Delaissé, and Niels T Foged. 2002. 'Matrix Metalloproteinase-Dependent Activation of Latent Transforming Growth Factor-Beta Controls the Conversion of Osteoblasts into Osteocytes by Blocking Osteoblast

- Apoptosis.' *The Journal of Biological Chemistry* 277 (46): 44061–67.
- Kaufman, R J. 2000. 'Overview of Vector Design for Mammalian Gene Expression.' *Molecular Biotechnology* 16 (2): 151–60.
- Kelly, Aoife, Stephanie A Houston, Eleanor Sherwood, Joshua Casulli, and Mark A Travis. 2017. 'Regulation of Innate and Adaptive Immunity by TGFβ.' *Advances in Immunology* 134: 137–233.
- Khan, A R, and M N James. 1998. 'Molecular Mechanisms for the Conversion of Zymogens to Active Proteolytic Enzymes.' *Protein Science: A Publication of the Protein Society* 7 (4): 815–36.
- Kim, Tae Kyung, and James H Eberwine. 2010. 'Mammalian Cell Transfection: The Present and the Future.' *Analytical and Bioanalytical Chemistry* 397 (8): 3173–78.
- Klein, Theo, Ulrich Eckhard, Antoine Dufour, Nestor Solis, and Christopher M Overall. 2018. 'Proteolytic Cleavage-Mechanisms, Function, and “Omic” Approaches for a Near-Ubiquitous Posttranslational Modification.' *Chemical Reviews* 118 (3): 1137–68.
- Knappskog, Stian, Hanne Ravneberg, Christine Gjerdrum, Christiane Trosse, Beate Stern, and Ian F Pryme. 2007. 'The Level of Synthesis and Secretion of Gaussia Princeps Luciferase in Transfected CHO Cells Is Heavily Dependent on the Choice of Signal Peptide.' *Journal of Biotechnology* 128 (4): 705–15.
- Kober, Lars, Christoph Zehe, and Juergen Bode. 2013. 'Optimized Signal Peptides for the Development of High Expressing CHO Cell Lines.' *Biotechnology and Bioengineering* 110 (4): 1164–73.
- Kokame, Koichi, Masanori Matsumoto, Yoshihiro Fujimura, and Toshiyuki Miyata. 2004. 'VWF73, a Region from D1596 to R1668 of von Willebrand Factor, Provides a Minimal Substrate for ADAMTS-13.' *Blood* 103 (2): 607–12.
- Kolodziej, Steven J, Terence Wagenknecht, Dudley K Strickland, and James K Stoops. 2002. 'The Three-Dimensional Structure of the Human Alpha 2-Macroglobulin Dimer Reveals Its Structural Organization in the Tetrameric Native and Chymotrypsin Alpha 2-Macroglobulin Complexes.' *The Journal of Biological Chemistry* 277 (31): 28031–37.
- Kondo, Márcia Y, Débora N Okamoto, Jorge A N Santos, Maria A Juliano, Kohei Oda, Bindu Pillai, Michael N G James, Luiz Juliano, and Iuri E Gouvea. 2010. 'Studies on the Catalytic Mechanism of a Glutamic Peptidase.' *The Journal of Biological Chemistry* 285 (28): 21437–45.
- König, Julia, Savanne Holster, Maaïke J Bruins, and Robert J Brummer. 2017. 'Randomized Clinical Trial: Effective Gluten Degradation by Aspergillus Niger-Derived Enzyme in a Complex Meal Setting.' *Scientific Reports* 7 (1): 13100.
- Koning, Frits. 2012. 'Celiac Disease: Quantity Matters.' *Seminars in Immunopathology* 34 (4): 541–49.
- Kost, Thomas A, J Patrick Condreay, and Donald L Jarvis. 2005. 'Baculovirus as Versatile Vectors for Protein Expression in Insect and Mammalian Cells.' *Nature Biotechnology* 23 (5): 567–75.
- Koziel, Joanna, and Jan Potempa. 2013. 'Protease-Armed Bacteria in the Skin.' *Cell and Tissue Research* 351 (2): 325–37.

Kremer Hovinga, Johanna A, Paul Coppo, Bernhard Lammle, Joel L Moake, Toshiyuki Miyata, and Karen Vanhoorelbeke. 2017. 'Thrombotic Thrombocytopenic Purpura.' *Nature Reviews. Disease Primers* 3 (April): 17020.

Kupfer, Sonia S, and Bana Jabri. 2012. 'Pathophysiology of Celiac Disease.' *Gastrointestinal Endoscopy Clinics of North America* 22 (4): 639–60.

## L

Lacy-Hulbert, A, R Thomas, X P Li, C E Lilley, R S Coffin, and J Roes. 2001. 'Interruption of Coding Sequences by Heterologous Introns Can Enhance the Functional Expression of Recombinant Genes.' *Gene Therapy* 8 (8): 649–53.

Lalonde, Marie-Eve, and Yves Durocher. 2017. 'Therapeutic Glycoprotein Production in Mammalian Cells.' *Journal of Biotechnology* 251 (June): 128–40.

LaMarre, J, M A Hayes, G K Wollenberg, I Hussaini, S W Hall, and S L Gonias. 1991. 'An Alpha 2-Macroglobulin Receptor-Dependent Mechanism for the Plasma Clearance of Transforming Growth Factor-Beta 1 in Mice.' *The Journal of Clinical Investigation* 87 (1): 39–44.

Lancellotti, Stefano, Maria Basso, and Raimondo De Cristofaro. 2013. 'Proteolytic Processing of von Willebrand Factor by Adamts13 and Leukocyte Proteases.' *Mediterranean Journal of Hematology and Infectious Diseases* 5 (1): e2013058.

Lancellotti, Stefano, and Raimondo De Cristofaro. 2011. 'Structure and Proteolytic Properties of ADAMTS13, a Metalloprotease Involved in the Pathogenesis of Thrombotic Microangiopathies.' *Progress in Molecular Biology and Translational Science* 99: 105–44.

Lawrence, D. A. 2001. *Molecular and Cellular Biochemistry*.

Lawrence, D A. 1996. 'Transforming Growth Factor-Beta: A General Review.' *European Cytokine Network* 7 (3): 363–74.

Leader, Benjamin, Quentin J Baca, and David E Golan. 2008. 'Protein Therapeutics: A Summary and Pharmacological Classification.' *Nature Reviews. Drug Discovery* 7 (1): 21–39.

Lebwohl, Benjamin, Jonas F Ludvigsson, and Peter H R Green. 2015. 'Celiac Disease and Non-Celiac Gluten Sensitivity.' *BMJ (Clinical Research Ed.)* 351 (October): h4347.

Lee, Linda, Ye Zhang, Brittany Ozar, Christoph W Sensen, and David C Schriemer. 2016. 'Carnivorous Nutrition in Pitcher Plants (*Nepenthes* Spp.) via an Unusual Complement of Endogenous Enzymes.' *Journal of Proteome Research* 15 (9): 3108–17.

Leuven, F Van, J J Cassiman, and H Van Den Berghe. 1979. 'Demonstration of an Alpha2-Macroglobulin Receptor in Human Fibroblasts, Absent in Tumor-Derived Cell Lines.' *The Journal of Biological Chemistry* 254 (12): 5155–60.

Levy, G G, W C Nichols, E C Lian, T Foroud, J N McClintick, B M McGee, A Y Yang, et al. 2001. 'Mutations in a Member of the ADAMTS Gene Family Cause Thrombotic Thrombocytopenic Purpura.' *Nature* 413 (6855): 488–94.

Lindfors, Katri, Markku Mäki, and Katri Kaukinen. 2010. 'Transglutaminase 2-Targeted Autoantibodies in Celiac Disease: Pathogenetic Players in Addition to Diagnostic Tools?' *Autoimmunity Reviews. Autoimmun Rev.*



- Liu, Chaoting, Brian Dalby, Weixing Chen, Jennifer M Kilzer, and Henry C Chiou. 2008. 'Transient Transfection Factors for High-Level Recombinant Protein Production in Suspension Cultured Mammalian Cells.' *Molecular Biotechnology* 39 (2): 141–53.
- Liu, Edwin, Hye-Seung Lee, Carin A Aronsson, William A Hagopian, Sibylle Koletzko, Marian J Rewers, George S Eisenbarth, et al. 2014. 'Risk of Pediatric Celiac Disease According to HLA Haplotype and Country.' *The New England Journal of Medicine* 371 (1): 42–49.
- Liu, Q, T Y Ling, H S Shieh, F E Johnson, J S Huang, and S S Huang. 2001. 'Identification of the High Affinity Binding Site in Transforming Growth Factor-Beta Involved in Complex Formation with Alpha 2-Macroglobulin. Implications Regarding the Molecular Mechanisms of Complex Formation between Alpha 2-Macroglobulin and Growth Fa'. *The Journal of Biological Chemistry* 276 (49): 46212–18.
- López-Otín, Carlos, and Judith S Bond. 2008. 'Proteases: Multifunctional Enzymes in Life and Disease.' *The Journal of Biological Chemistry* 283 (45): 30433–37.
- Lotta, Luca A, Isabella Garagiola, Roberta Palla, Andrea Cairo, and Flora Peyvandi. 2010. 'ADAMTS13 Mutations and Polymorphisms in Congenital Thrombotic Thrombocytopenic Purpura.' *Human Mutation* 31 (1): 11–19.
- Ludvigsson, Jonas F, Daniel A Leffler, Julio C Bai, Federico Biagi, Alessio Fasano, Peter H R Green, Marios Hadjivassiliou, et al. 2013. 'The Oslo Definitions for Coeliac Disease and Related Terms.' *Gut* 62 (1): 43–52.
- Luken, Brenda M, Luke Y N Winn, Jonas Emsley, David A Lane, and James T B Crawley. 2010. 'The Importance of Vicinal Cysteines, C1669 and C1670, for von Willebrand Factor A2 Domain Function.' *Blood* 115 (23): 4910–13.
- Luo, Gui-Ping, Bing Ni, Xia Yang, and Yu-Zhang Wu. 2012. 'Von Willebrand Factor: More than a Regulator of Hemostasis and Thrombosis.' *Acta Haematologica* 128 (3): 158–69.
- Lyons, R M, J Keski-Oja, and H L Moses. 1988. 'Proteolytic Activation of Latent Transforming Growth Factor-Beta from Fibroblast-Conditioned Medium.' *The Journal of Cell Biology* 106 (5): 1659–65.

## M

- Maiuri, Luigi, Carolina Ciacci, Ida Ricciardelli, Loredana Vacca, Valeria Raia, Salvatore Auricchio, Jean Picard, Mohamed Osman, Sonia Quaratino, and Marco Londei. 2003. 'Association between Innate Response to Gliadin and Activation of Pathogenic T Cells in Coeliac Disease'. *Lancet* 362 (9377): 30–37.
- Majerus, Elaine M, Patricia J Anderson, and J Evan Sadler. 2005. 'Binding of ADAMTS13 to von Willebrand Factor.' *The Journal of Biological Chemistry* 280 (23): 21773–78.
- Mäki, Markku. 1995. '3 The Humoral Immune System in Coeliac Disease'. *Bailliere's Clinical Gastroenterology* 9 (2): 231–49.
- Manea, Minola, and Diana Karpman. 2009. 'Molecular Basis of ADAMTS13 Dysfunction in Thrombotic Thrombocytopenic Purpura.' *Pediatric Nephrology (Berlin, Germany)* 24 (3): 447–58.
- Marrero, Aniebrys, Stephane Duquerroy, Stefano Trapani, Theodoros Goulas, Tibisay Guevara, Gregers R Andersen, Jorge Navaza, Lars Sottrup-Jensen, and F Xavier Gomis-Rüth. 2012. 'The Crystal Structure of Human A2-Macroglobulin Reveals a

- Unique Molecular Cage.’ *Angewandte Chemie (International Ed. in English)* 51 (14): 3340–44.
- Mazumdar, Kaushiki, Xavier Alvarez, Juan T Borda, Jason Dufour, Edith Martin, Michael T Bethune, Chaitan Khosla, and Karol Sestak. 2010. ‘Visualization of Transepithelial Passage of the Immunogenic 33-Residue Peptide from Alpha-2 Gliadin in Gluten-Sensitive Macaques.’ *PLoS One* 5 (4): e10228.
- McCaffrey, T A, D J Falcone, C F Brayton, L A Agarwal, F G Welt, and B B Weksler. 1989. ‘Transforming Growth Factor-Beta Activity Is Potentiated by Heparin via Dissociation of the Transforming Growth Factor-Beta/Alpha 2-Macroglobulin Inactive Complex.’ *The Journal of Cell Biology* 109 (1): 441–48.
- McGwire, B S, and K P Chang. 1996. ‘Posttranslational Regulation of a Leishmania HEXXH Metalloprotease (Gp63). The Effects of Site-Specific Mutagenesis of Catalytic, Zinc Binding, N-Glycosylation, and Glycosyl Phosphatidylinositol Addition Sites on N-Terminal End Cleavage, Intracellular St’.
- Melcher, Ralph, Hans-Wilhelm Grosch, and Andrej Hasilik. 2002. ‘Plasmid Vectors with a 5’-Hybrid Intron Facilitate High-Level Glycoprotein Expression in CHO-Cells.’ *Biochimica et Biophysica Acta* 1575 (1–3): 49–53.
- Mignaqui, Ana Clara, Vanesa Ruiz, Sylvie Perret, Gilles St-Laurent, Parminder Singh Chahal, Julia Transfiguracion, Ayelen Sammarruco, Victoria Gnazzo, Yves Durocher, and Andres Wigdorovitz. 2013. ‘Transient Gene Expression in Serum-Free Suspension-Growing Mammalian Cells for the Production of Foot-and-Mouth Disease Virus Empty Capsids.’ *PLoS One* 8 (8): e72800.
- Mitea, C, R Havenaar, J Wouter Drijfhout, L Edens, L Dekking, and F Koning. 2008. ‘Efficient Degradation of Gluten by a Prolyl Endoprotease in a Gastrointestinal Model: Implications for Coeliac Disease.’ *Gut* 57 (1): 25–32.
- Mithöfer, Axel. 2011. ‘Carnivorous Pitcher Plants: Insights in an Old Topic.’ *Phytochemistry* 72 (13): 1678–82. <https://doi.org/10.1016/j.phytochem.2010.11.024>.
- Miyazono, K, and C H Heldin. 1989. ‘Role for Carbohydrate Structures in TGF-Beta 1 Latency.’ *Nature* 338 (6211): 158–60.
- Miyazono, K, H Ichijo, and C H Heldin. 1993. ‘Transforming Growth Factor-Beta: Latent Forms, Binding Proteins and Receptors.’ *Growth Factors (Chur, Switzerland)* 8 (1): 11–22.
- Miyazono, K, A Olofsson, P Colosetti, and C H Heldin. 1991. ‘A Role of the Latent TGF-Beta 1-Binding Protein in the Assembly and Secretion of TGF-Beta 1.’ *The EMBO Journal* 10 (5): 1091–1101.
- Molberg, O, S N Mcadam, R Körner, H Quarsten, C Kristiansen, L Madsen, L Fugger, et al. 1998. ‘Tissue Transglutaminase Selectively Modifies Gliadin Peptides That Are Recognized by Gut-Derived T Cells in Celiac Disease.’ *Nature Medicine* 4 (6): 713–17.
- Moschowitz, E. 1952. ‘An Acute Febrile Pleiochromic Anemia with Hyaline Thrombosis of the Terminal Arterioles and Capillaries; an Undescribed Disease.’ *The American Journal of Medicine* 13 (5): 567–69.
- Mosyak, Lidia, Katy Georgiadis, Tania Shane, Kristine Svenson, Tracy Hebert, Thomas McDonagh, Stewart Mackie, et al. 2008. ‘Crystal Structures of the Two Major Aggrecan Degrading Enzymes, ADAMTS4 and ADAMTS5.’ *Protein Science: A Publication of the*

*Protein Society* 17 (1): 16–21.

- Moulin, Aaron, Magali Mathieu, Catherine Lawrence, Russell Bigelow, Mark Levine, Christine Hamel, Jean-Piere Marquette, et al. 2014. ‘Structures of a Pan-Specific Antagonist Antibody Complexed to Different Isoforms of TGF $\beta$  Reveal Structural Plasticity of Antibody-Antigen Interactions.’ *Protein Science: A Publication of the Protein Society* 23 (12): 1698–1707.
- Mu, Dezhi, Stephanie Cambier, Lars Fjellbirkeland, Jody L Baron, John S Munger, Hisaaki Kawakatsu, Dean Sheppard, V Courtney Broaddus, and Stephen L Nishimura. 2002. ‘The Integrin Alpha(v)Beta8 Mediates Epithelial Homeostasis through MT1-MMP-Dependent Activation of TGF-Beta1.’ *The Journal of Cell Biology* 157 (3): 493–507.
- Muia, Joshua, Jian Zhu, Garima Gupta, Sandra L Haberichter, Kenneth D Friedman, Hendrik B Feys, Louis Deforche, et al. 2014. ‘Allosteric Activation of ADAMTS13 by von Willebrand Factor.’ *Proceedings of the National Academy of Sciences of the United States of America* 111 (52): 18584–89.
- Munger, J S, J G Harpel, P E Gleizes, R Mazziere, I Nunes, and D B Rifkin. 1997. ‘Latent Transforming Growth Factor-Beta: Structural Features and Mechanisms of Activation.’ *Kidney International* 51 (5): 1376–82.
- Munger, J S, X Huang, H Kawakatsu, M J Griffiths, S L Dalton, J Wu, J F Pittet, et al. 1999. ‘The Integrin Alpha v Beta 6 Binds and Activates Latent TGF Beta 1: A Mechanism for Regulating Pulmonary Inflammation and Fibrosis.’ *Cell* 96 (3): 319–28.

## N

- Nallet, Sophie, Mario Amacker, Nicole Westerfeld, Lucia Baldi, Iwo Konig, David L Hacker, Christiane Zaborosch, Rinaldo Zurbriggen, and Florian M Wurm. 2009. ‘Respiratory Syncytial Virus Subunit Vaccine Based on a Recombinant Fusion Protein Expressed Transiently in Mammalian Cells.’ *Vaccine* 27 (46): 6415–19.
- Negoescu, A, F Labat-Moleur, E Brambilla, E M Chambaz, and J J Feige. 1994. ‘Steroidogenic Adrenocortical Cells Synthesize Alpha 2-Macroglobulin in Vitro, Not in Vivo.’ *Molecular and Cellular Endocrinology* 105 (2): 155–63.
- Nettleship, Joanne E, Rene Assenberg, Jonathan M Diprose, Nahid Rahman-Huq, and Raymond J Owens. 2010. ‘Recent Advances in the Production of Proteins in Insect and Mammalian Cells for Structural Biology.’ *Journal of Structural Biology* 172 (1): 55–65.
- Neurath, H, and K A Walsh. 1976. ‘Role of Proteolytic Enzymes in Biological Regulation (a Review).’ *Proceedings of the National Academy of Sciences of the United States of America* 73 (11): 3825–32.
- Nilsen, E M, K E Lundin, P Krajci, H Scott, L M Sollid, and P Brandtzaeg. 1995. ‘Gluten Specific, HLA-DQ Restricted T Cells from Coeliac Mucosa Produce Cytokines with Th1 or Th0 Profile Dominated by Interferon Gamma.’ *Gut* 37 (6): 766–76.
- Nishio, Kenji, Patricia J Anderson, X Long Zheng, and J Evan Sadler. 2004. ‘Binding of Platelet Glycoprotein Ibalph to von Willebrand Factor Domain A1 Stimulates the Cleavage of the Adjacent Domain A2 by ADAMTS13.’ *Proceedings of the National Academy of Sciences of the United States of America* 101 (29): 10578–83.
- Nixon, A J, B D Brower-Toland, and L J Sandell. 2000. ‘Molecular Cloning of Equine Transforming Growth Factor-Beta1 Reveals Equine-Specific Amino Acid Substitutions in the Mature Peptide Sequence.’ *Journal of Molecular Endocrinology* 24 (2):

261–72.

Noris, Marina, Sara Bucchioni, Miriam Galbusera, Roberta Donadelli, Elena Bresin, Federica Castelletti, Jessica Caprioli, Simona Brioschi, Friedrich Scheiflinger, and Giuseppe Remuzzi. 2005. 'Complement Factor H Mutation in Familial Thrombotic Thrombocytopenic Purpura with ADAMTS13 Deficiency and Renal Involvement.' *Journal of the American Society of Nephrology: JASN* 16 (5): 1177–83.

Nunes, I, P E Gleizes, C N Metz, and D B Rifkin. 1997. 'Latent Transforming Growth Factor-Beta Binding Protein Domains Involved in Activation and Transglutaminase-Dependent Cross-Linking of Latent Transforming Growth Factor-Beta.' *The Journal of Cell Biology* 136 (5): 1151–63.

## O

O'Donohue, M J, and A Beaumont. 1996. 'The Roles of the Prosequence of Thermolysin in Enzyme Inhibition and Folding in Vitro.' *The Journal of Biological Chemistry* 271 (43): 26477–81.

Olofsson, A, K Miyazono, T Kanzaki, P Colosetti, U Engström, and C H Heldin. 1992. 'Transforming Growth Factor-Beta 1, -Beta 2, and -Beta 3 Secreted by a Human Glioblastoma Cell Line. Identification of Small and Different Forms of Large Latent Complexes.' *The Journal of Biological Chemistry* 267 (27): 19482–88.

Olson, Steven T, and Peter G W Gettins. 2011. 'Regulation of Proteases by Protein Inhibitors of the Serpin Superfamily.' *Progress in Molecular Biology and Translational Science* 99: 185–240.

Otlewski, Jacek, Filip Jelen, Malgorzata Zakrzewska, and Arkadiusz Oleksy. 2005. 'The Many Faces of Protease-Protein Inhibitor Interaction.' *The EMBO Journal* 24 (7): 1303–10. <https://doi.org/10.1038/sj.emboj.7600611>.

Owenjr., T Page, Kristen a Lennon, Matthew J Santo, and A M Y N Anderson. 1999. 'Pathways for Nutrient Transport in the Pitchers of the Carnivorous Plant *Nepenthes Alata*.' *Annals of Botany* 84 (4): 459–66.

Ozuna, Carmen V, Julio C M Iehisa, María J Giménez, Juan B Alvarez, Carolina Sousa, and Francisco Barro. 2015. 'Diversification of the Celiac Disease  $\alpha$ -Gliadin Complex in Wheat: A 33-Mer Peptide with Six Overlapping Epitopes, Evolved Following Polyploidization.' *The Plant Journal: For Cell and Molecular Biology* 82 (5): 794–805.

## P

Padua, David, and Joan Massagué. 2009. 'Roles of TGF $\beta$  in Metastasis.' *Cell Research* 19 (1): 89–102.

Petri, Anastasis, Hyo Jung Kim, Yaoxian Xu, Rens de Groot, Chan Li, Aline Vandenbulcke, Karen Vanhoorelbeke, Jonas Emsley, and James T B Crawley. 2019. 'Crystal Structure and Substrate-Induced Activation of ADAMTS13.' *Nature Communications* 10 (1): 3781.

Pillai, B, Maia M Cherney, Kazumi Hiraga, Katsumi Takada, Kohei Oda, and Michael N G James. 2007. 'Crystal Structure of Scytalidoglutamic Peptidase with Its First Potent Inhibitor Provides Insights into Substrate Specificity and Catalysis.' *Journal of Molecular Biology* 365 (2): 343–61.

Pogue, Gregory P, Fakhrieh Vojdani, Kenneth E Palmer, Ernie Hiatt, Steve Hume, Jim Phelps, Lori Long, et al. 2010. 'Production of Pharmaceutical-Grade Recombinant

- Aprotinin and a Monoclonal Antibody Product Using Plant-Based Transient Expression Systems.' *Plant Biotechnology Journal* 8 (5): 638–54.
- Pop, Cristina, and Guy S Salvesen. 2009. 'Human Caspases: Activation, Specificity, and Regulation.' *The Journal of Biological Chemistry* 284 (33): 21777–81.
- Porter, Sarah, Ian M Clark, Lara Kevorkian, and Dylan R Edwards. 2005. 'The ADAMTS Metalloproteinases.' *The Biochemical Journal* 386 (Pt 1): 15–27.
- Pos, Wouter, James T B Crawley, Rob Fijnheer, Jan Voorberg, David A Lane, and Brenda M Luken. 2010. 'An Autoantibody Epitope Comprising Residues R660, Y661, and Y665 in the ADAMTS13 Spacer Domain Identifies a Binding Site for the A2 Domain of VWF.' *Blood* 115 (8): 1640–49.
- R**
- Rawlings, N D, and A J Barrett. 1993. 'Evolutionary Families of Peptidases.' *The Biochemical Journal* 290 ( Pt 1 (Pt 1): 205–18.
- Rawlings, Neil D. 2010. 'Peptidase Inhibitors in the MEROPS Database.' *Biochimie* 92 (11): 1463–83.
- Rawlings, Neil D, Alan J Barrett, and Alex Bateman. 2014. 'Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates.' *Current Protocols in Bioinformatics* 48 (December): 1.25.1-33.
- Rawlings, Neil D, Alan J Barrett, Paul D Thomas, Xiaosong Huang, Alex Bateman, and Robert D Finn. 2018. 'The MEROPS Database of Proteolytic Enzymes, Their Substrates and Inhibitors in 2017 and a Comparison with Peptidases in the PANTHER Database.' *Nucleic Acids Research* 46 (D1): D624–32.
- Rawlings, Neil D, Dominic P Tolle, and Alan J Barrett. 2004. 'Evolutionary Families of Peptidase Inhibitors.' *The Biochemical Journal* 378 (Pt 3): 705–16.
- Rawlings, Neil D, Matthew Waller, Alan J Barrett, and Alex Bateman. 2014. 'MEROPS: The Database of Proteolytic Enzymes, Their Substrates and Inhibitors.' *Nucleic Acids Research* 42 (Database issue): D503-9.
- Rawlings, Neil David, Alan John Barrett, and Alex Bateman. 2011. 'Asparagine Peptide Lyases: A Seventh Catalytic Type of Proteolytic Enzymes.' *The Journal of Biological Chemistry* 286 (44): 38321–28.
- Real, Ana, Isabel Comino, Laura de Lorenzo, Francisco Merchán, Javier Gil-Humanes, María J Giménez, Miguel Ángel López-Casado, et al. 2012. 'Molecular and Immunological Characterization of Gluten Proteins Isolated from Oat Cultivars That Differ in Toxicity for Celiac Disease.' *PLoS One* 7 (12): e48365.
- Reily, Colin, Tyler J Stewart, Matthew B Renfrow, and Jan Novak. 2019. 'Glycosylation in Health and Disease.' *Nature Reviews. Nephrology* 15 (6): 346–66.
- Renner, Tanya, and Chelsea D Specht. 2013. 'Inside the Trap: Gland Morphologies, Digestive Enzymes, and the Evolution of Plant Carnivory in the Caryophyllales.' *Current Opinion in Plant Biology* 16 (4): 436–42.
- Rey, Martial, Menglin Yang, Linda Lee, Ye Zhang, Joey G Sheff, Christoph W Sensen, Hynek Mrazek, et al. 2016. 'Addressing Proteolytic Efficiency in Enzymatic Degradation Therapy for Celiac Disease.' *Scientific Reports* 6 (August): 30980.

- Richter, C, T Tanaka, and R Y Yada. 1998. 'Mechanism of Activation of the Gastric Aspartic Proteinases: Pepsinogen, Progastricsin and Prochymosin.' *The Biochemical Journal* 335 (Pt 3 (Pt 3): 481–90.
- Rifkin, Daniel B. 2005. 'Latent Transforming Growth Factor-Beta (TGF-Beta) Binding Proteins: Orchestrators of TGF-Beta Availability.' *The Journal of Biological Chemistry* 280 (9): 7409–12.
- Robertson, Ian B, Masahito Horiguchi, Lior Zilberberg, Branka Dabovic, Krassimira Hadjiolova, and Daniel B Rifkin. 2015. 'Latent TGF- $\beta$ -Binding Proteins.' *Matrix Biology: Journal of the International Society for Matrix Biology* 47 (September): 44–53.
- Rottloff, Sandy, Sissi Miguel, Flore Biteau, Estelle Nisse, Philippe Hammann, Lauriane Kuhn, Johana Chicher, et al. 2016. 'Proteome Analysis of Digestive Fluids in Nepenthes Pitchers.' *Annals of Botany* 117 (3): 479–95.
- Rottloff, Sandy, Regina Stieber, Heiko Maischak, Florian G Turini, Günther Heubl, and Axel Mithöfer. 2011. 'Functional Characterization of a Class III Acid Endochitinase from the Traps of the Carnivorous Pitcher Plant Genus, *Nepenthes*.' *Journal of Experimental Botany* 62 (13): 4639–47.
- Rusmili, Muhamad Rusdi Ahmad, Tee Ting Yee, Mohd Rais Mustafa, Wayne C Hodgson, and Iekhsan Othman. 2014. 'Proteomic Characterization and Comparison of Malaysian Bungarus Candidus and Bungarus Fasciatus Venoms.' *Journal of Proteomics* 110 (October): 129–44.
- S**
- Sabotič, Jerica, and Janko Kos. 2012. 'Microbial and Fungal Protease Inhibitors--Current and Potential Applications.' *Applied Microbiology and Biotechnology* 93 (4): 1351–75.
- Sadler, J E. 1998. 'Biochemistry and Genetics of von Willebrand Factor.' *Annual Review of Biochemistry* 67: 395–424.
- Sadler, J Evan, Joel L Moake, Toshiyuki Miyata, and James N George. 2004. 'Recent Advances in Thrombotic Thrombocytopenic Purpura.' *Hematology. American Society of Hematology. Education Program*, 407–23.
- Sánchez-León, Susana, Javier Gil-Humanes, Carmen V Ozuna, María J Giménez, Carolina Sousa, Daniel F Voytas, and Francisco Barro. 2018. 'Low-Gluten, Nontransgenic Wheat Engineered with CRISPR/Cas9.' *Plant Biotechnology Journal* 16 (4): 902–10.
- Sasaki, Hiroshi, Atsushi Nakagawa, Tomonari Muramatsu, Megumi Suganuma, Yoriko SAWANO, Masaki Kojima, Keiko Kubota, Kenji Takahashi, and Masaru Tanokura. 2004. 'The Three-Dimensional Structure of Aspergilloglutamic Peptidase from *Aspergillus Niger*.' *Proceedings of the Japan Academy, Series B* 80 (9): 435–38.
- Sato, Y, and D B Rifkin. 1989. 'Inhibition of Endothelial Cell Movement by Pericytes and Smooth Muscle Cells: Activation of a Latent Transforming Growth Factor-Beta 1-like Molecule by Plasmin during Co-Culture.' *The Journal of Cell Biology* 109 (1): 309–15.
- Savi, Chris De, Andrew Pape, John G Cumming, Atilla Ting, Peter D Smith, Jeremy N Burrows, Mark Mills, et al. 2011. 'The Design and Synthesis of Novel N-Hydroxyformamide Inhibitors of ADAM-TS4 for the Treatment of Osteoarthritis.' *Bioorganic & Medicinal Chemistry Letters* 21 (5): 1376–81.
- Schalk, Kathrin, Barbara Lexhaller, Peter Koehler, and Katharina Anne Scherf. 2017.

- Isolation and Characterization of Gluten Protein Types from Wheat, Rye, Barley and Oats for Use as Reference Materials.' *PLoS One* 12 (2): e0172819.
- Schmierer, Bernhard, and Caroline S Hill. 2007. 'TGFbeta-SMAD Signal Transduction: Molecular Specificity and Functional Flexibility.' *Nature Reviews. Molecular Cell Biology* 8 (12): 970–82.
- Schneppenheim, R, and U Budde. 2011. 'Von Willebrand Factor: The Complex Molecular Genetics of a Multidomain and Multifunctional Protein.' *Journal of Thrombosis and Haemostasis : JTH* 9 Suppl 1 (July): 209–15.
- Schröder, Christoph U, Linda Lee, Martial Rey, Vladimir Sarpe, Petr Man, Seema Sharma, Vlad Zabrouskov, Brett Larsen, and David C Schriemer. 2017. 'Neprosin, a Selective Prolyl Endoprotease for Bottom-up Proteomics and Histone Mapping.' *Molecular & Cellular Proteomics : MCP* 16 (6): 1162–71.
- Schröder, Christoph U, Daniel S Ziemianowicz, Kathleen Merx, and David C Schriemer. 2018. 'Simultaneous Proteoform Analysis of Histones H3 and H4 with a Simplified Middle-Down Proteomics Method'. *Analytical Chemistry* 90 (5): 3083–90.
- Schuppan, Detlef, Yvonne Junker, and Donatella Barisani. 2009. 'Celiac Disease: From Pathogenesis to Novel Therapies'. *Gastroenterology* 137 (6): 1912–33.
- Scully, Marie, Helen Yarranton, Ri Liesner, Jamie Cavenagh, Beverley Hunt, Sylvia Benjamin, David Bevan, Ian Mackie, and Samuel Machin. 2008. 'Regional UK TTP Registry: Correlation with Laboratory ADAMTS 13 Analysis and Clinical Features.' *British Journal of Haematology* 142 (5): 819–26.
- Shan, Lu, Thomas Marti, Ludvig M Sollid, Gary M Gray, and Chaitan Khosla. 2004. 'Comparative Biochemical Analysis of Three Bacterial Prolyl Endopeptidases: Implications for Coeliac Sprue.' *The Biochemical Journal* 383 (Pt 2): 311–18.
- Shan, Lu, Øyvind Molberg, Isabelle Parrot, Felix Hausch, Ferda Filiz, Gary M Gray, Ludvig M Sollid, and Chaitan Khosla. 2002. 'Structural Basis for Gluten Intolerance in Celiac Sprue.' *Science (New York, N.Y.)* 297 (5590): 2275–79.
- Shan, Lu, Shuo-Wang Qiao, Helene Arentz-Hansen, Øyvind Molberg, Gary M Gray, Ludvig M Sollid, and Chaitan Khosla. 2005. 'Identification and Analysis of Multivalent Proteolytically Resistant Peptides from Gluten: Implications for Celiac Sprue.' *Journal of Proteome Research* 4 (5): 1732–41.
- Shewry, P R. 2009. 'Wheat'. *Journal of Experimental Botany* 60 (6): 1537–53.
- Shewry, Peter R, and Nigel G Halford. 2002. 'Cereal Seed Storage Proteins: Structures, Properties and Role in Grain Utilization.' *Journal of Experimental Botany* 53 (370): 947–58.
- Shi, Minlong, Jianghai Zhu, Rui Wang, Xing Chen, Lizhi Mi, Thomas Walz, and Timothy A Springer. 2011. 'Latent TGF-β Structure and Activation.' *Nature* 474 (7351): 343–49.
- Shi, Yigong, and Joan Massagué. 2003. 'Mechanisms of TGF-Beta Signaling from Cell Membrane to the Nucleus.' *Cell* 113 (6): 685–700.
- Shim, Kyuhwan, Patricia J Anderson, Elodee A Tuley, Erin Wiswall, and J Evan Sadler. 2008. 'Platelet-VWF Complexes Are Preferred Substrates of ADAMTS13 under Fluid Shear Stress.' *Blood* 111 (2): 651–57.

- Skipwith, Christopher G, Wenjing Cao, and X Long Zheng. 2010. 'Factor VIII and Platelets Synergistically Accelerate Cleavage of von Willebrand Factor by ADAMTS13 under Fluid Shear Stress.' *The Journal of Biological Chemistry* 285 (37): 28596–603.
- Soejima, Kenji, Hitomi Nakamura, Masaki Hirashima, Wataru Morikawa, Chikateru Nozaki, and Tomohiro Nakagaki. 2006. 'Analysis on the Molecular Species and Concentration of Circulating ADAMTS13 in Blood.' *Journal of Biochemistry* 139 (1): 147–54.
- Sollid, Ludvig M. 2002. 'Coeliac Disease: Dissecting a Complex Inflammatory Disorder.' *Nature Reviews. Immunology* 2 (9): 647–55.
- Sorvillo, N, P H Kaijen, M Matsumoto, Y Fujimura, C van der Zwaan, F C Verbij, W Pos, R Fijnheer, J Voorberg, and A B Meijer. 2014. 'Identification of N-Linked Glycosylation and Putative O-Fucosylation, C-Mannosylation Sites in Plasma Derived ADAMTS13.' *Journal of Thrombosis and Haemostasis : JTH* 12 (5): 670–79.
- Sottrup-Jensen, L. 1989. 'Alpha-Macroglobulins: Structure, Shape, and Mechanism of Proteinase Complex Formation'. *The Journal of Biological Chemistry* 264 (20): 11539–42.
- South, Kieron, Marta O Freitas, and David A Lane. 2017. 'A Model for the Conformational Activation of the Structurally Quiescent Metalloprotease ADAMTS13 by von Willebrand Factor.' *The Journal of Biological Chemistry* 292 (14): 5760–69.
- South, Kieron, Brenda M Luken, James T B Crawley, Rebecca Phillips, Mari Thomas, Richard F Collins, Louis Deforche, Karen Vanhoorelbeke, and David A Lane. 2014. 'Conformational Activation of ADAMTS13.' *Proceedings of the National Academy of Sciences of the United States of America* 111 (52): 18578–83.
- Springer, Timothy A. 2006. 'Complement and the Multifaceted Functions of VWA and Integrin I Domains.' *Structure (London, England : 1993)* 14 (11): 1611–16.
- Stanley, P, T S Raju, and M Bhaumik. 1996. 'CHO Cells Provide Access to Novel N-Glycans and Developmentally Regulated Glycosyltransferases.' *Glycobiology* 6 (7): 695–99.
- Stephenson, Paul, and Jamie Hogan. 2006. 'Cloning and Characterization of a Ribonuclease, a Cysteine Proteinase, and an Aspartic Proteinase from Pitchers of the Carnivorous Plant *Nepenthes Ventricosa* Blanco'. *International Journal of Plant Sciences - INT J PLANT SCI* 167 (March): 239–48.
- Stockschlaeder, Marcus, Reinhard Schneppenheim, and Ulrich Budde. 2014. 'Update on von Willebrand Factor Multimers: Focus on High-Molecular-Weight Multimers and Their Role in Hemostasis.' *Blood Coagulation & Fibrinolysis: An International Journal in Haemostasis and Thrombosis* 25 (3): 206–16.
- Sturgeon, Craig, Jingtang Lan, and Alessio Fasano. 2017. 'Zonulin Transgenic Mice Show Altered Gut Permeability and Increased Morbidity/Mortality in the DSS Colitis Model.' *Annals of the New York Academy of Sciences* 1397 (1): 130–42.
- Suchy FJ, Brannon PM, Carpenter TO, Fernandez JR, Gilsanz V, Gould JB, Hall K, Hui SL, Lupton J, Mennella J, Miller NJ, Osganian SK, Sellmeyer DE, Wolf MA. National Institutes of Health Consensus Development Conference: lactose intolerance and health. *Ann Intern Med.* 2010 Jun 15;152(12):792-6.
- Swiech, Kamilla, Virginia Picanço-Castro, and Dimas Tadeu Covas. 2012. 'Human Cells: New Platform for Recombinant Therapeutic Protein Production.' *Protein Expression and Purification* 84 (1): 147–53.



## T

- Taipale, J, K Koli, and J Keski-Oja. 1992. 'Release of Transforming Growth Factor-Beta 1 from the Pericellular Matrix of Cultured Fibroblasts and Fibrosarcoma Cells by Plasmin and Thrombin.' *The Journal of Biological Chemistry* 267 (35): 25378–84.
- Takahashi, Kenji. 2013. 'Chapter 74 - Aspergilloglutamic Peptidase'. In , edited by Neil D Rawlings and Guy B T - Handbook of Proteolytic Enzymes (Third Edition) Salvesen, 307–10. Academic Press.
- Takahashi, Kenji, Senarath B P Athauda, Koji Matsumoto, Sanath Rajapakshe, Masayuki Kuribayashi, Masaki Kojima, Nobuko Kubomura-Yoshida, Akihiro Iwamatsu, Chiaki Shibata, and Hideshi Inoue. 2005. 'Nepenthesin, a Unique Member of a Novel Subfamily of Aspartic Proteinases: Enzymatic and Structural Characteristics.' *Current Protein & Peptide Science* 6 (6): 513–25.
- Tallant, Cynthia, Raquel Garcia-Castellanos, Ulrich Baumann, and F Xavier Gomis-Ruth. 2010. 'On the Relevance of the Met-Turn Methionine in Metzincins.' *The Journal of Biological Chemistry* 285 (18): 13951–57.
- Thomas, Mari R, Rens de Groot, Marie A Scully, and James T B Crawley. 2015. 'Pathogenicity of Anti-ADAMTS13 Autoantibodies in Acquired Thrombotic Thrombocytopenic Purpura.' *EBioMedicine* 2 (8): 942–52.
- Travis, Mark A, and Dean Sheppard. 2014. 'TGF- $\beta$  Activation and Function in Immunity.' *Annual Review of Immunology* 32: 51–82.
- Turk, Boris. 2006. 'Targeting Proteases: Successes, Failures and Future Prospects.' *Nature Reviews. Drug Discovery* 5 (9): 785–99.
- Turesson, Ola, Christina Uhe, Aleksei Rozkov, and Elke Lullau. 2008. 'Development of a Generic Transient Transfection Process at 100 L Scale.' *Cytotechnology* 56 (2): 123–36.
- Tye-Din, Jason A, Heather J Galipeau, and Daniel Agardh. 2018. 'Celiac Disease: A Review of Current Concepts in Pathogenesis, Prevention, and Novel Therapies.' *Frontiers in Pediatrics* 6: 350.

## U

- Uchida, Toshihiro, Hideo Wada, Minoru Mizutani, Miho Iwashita, Hiroaki Ishihara, Toshiro Shibano, Misako Suzuki, et al. 2004. 'Identification of Novel Mutations in ADAMTS13 in an Adult Patient with Congenital Thrombotic Thrombocytopenic Purpura.' *Blood* 104 (7): 2081–83.
- Uemura, Masahito, Kouko Tatsumi, Masanori Matsumoto, Masao Fujimoto, Tomomi Matsuyama, Masatoshi Ishikawa, Taka-Aki Iwamoto, et al. 2005. 'Localization of ADAMTS13 to the Stellate Cells of Human Liver.' *Blood* 106 (3): 922–24.

## V

- Vader, L Willemijn, Dariusz T Stepniak, Evelien M Bunnik, Yvonne M C Kooy, Willeke de Haan, Jan Wouter Drijfhout, Peter A Van Veelen, and Frits Koning. 2003. 'Characterization of Cereal Toxicity for Celiac Disease Patients Based on Protein Homology in Grains.' *Gastroenterology* 125 (4): 1105–13.
- Vader, Willemijn, Yvonne Kooy, Peter Van Veelen, Arnoud De Ru, Diana Harris, Willemien Benckhuijsen, Salvador Peña, Luisa Mearin, Jan Wouter Drijfhout, and Frits Koning.

2002. 'The Gluten Response in Children with Celiac Disease Is Directed toward Multiple Gliadin and Glutenin Peptides.' *Gastroenterology* 122 (7): 1729–37.
- Varki, A. 1998. 'Factors Controlling the Glycosylation Potential of the Golgi Apparatus.' *Trends in Cell Biology* 8 (1): 34–40.
- Verbij, Fabian C, Eva Stokhuijzen, Paul H P Kaijen, Floris van Alphen, Alexander B Meijer, and Jan Voorberg. 2016. 'Identification of Glycans on Plasma-Derived ADAMTS13.' *Blood* 128 (21): e51–58.
- Vlieghe, Patrick, Vincent Lisowski, Jean Martinez, and Michel Khrestchatsky. 2010. 'Synthetic Therapeutic Peptides: Science and Market.' *Drug Discovery Today* 15 (1–2): 40–56. **W**
- Wakefield, L M, D M Smith, K C Flanders, and M B Sporn. 1988. 'Latent Transforming Growth Factor-Beta from Human Platelets. A High Molecular Weight Complex Containing Precursor Sequences.' *The Journal of Biological Chemistry* 263 (16): 7646–54.
- Walton, Kelly L, Yogeshwar Makanji, Matthew C Wilce, Karen L Chan, David M Robertson, and Craig A Harrison. 2009. 'A Common Biosynthetic Pathway Governs the Dimerization and Secretion of Inhibin and Related Transforming Growth Factor Beta (TGFbeta) Ligands.' *The Journal of Biological Chemistry* 284 (14): 9311–20.
- Wang, C, M Eufemi, C Turano, and A Giartosio. 1996. 'Influence of the Carbohydrate Moiety on the Stability of Glycoproteins.' *Biochemistry* 35 (23): 7299–7307.
- Webb, D J, T L Atkins, K P Crookston, J K Burmester, S W Qian, and S L Gonias. 1994. 'Transforming Growth Factor Beta Isoform 2-Specific High Affinity Binding to Native Alpha 2-Macroglobulin. Chimeras Identify a Sequence That Determines Affinity for Native but Not Activated Alpha 2-Macroglobulin.' *The Journal of Biological Chemistry* 269 (48): 30402–6.
- Wei, Guoxian, Na Tian, Roland Siezen, Detlef Schuppan, and Eva J Helmerhorst. 2016. 'Identification of Food-Grade Subtilisins as Gluten-Degrading Enzymes to Treat Celiac Disease.' *American Journal of Physiology. Gastrointestinal and Liver Physiology* 311 (3): G571–80.
- Whisstock, James C, and Stephen P Bottomley. 2006. 'Molecular Gymnastics: Serpin Structure, Folding and Misfolding.' *Current Opinion in Structural Biology* 16 (6): 761–68.
- Wieser, Herbert. 2007. 'Chemistry of Gluten Proteins.' *Food Microbiology* 24 (2): 115–19.
- Wu, Ji-Wei, and Xiu-Lan Chen. 2011. 'Extracellular Metalloproteases from Bacteria.' *Applied Microbiology and Biotechnology* 92 (2): 253–62.
- Wu, Mary Y, and Caroline S Hill. 2009. 'Tgf-Beta Superfamily Signaling in Embryonic Development and Homeostasis.' *Developmental Cell* 16 (3): 329–43.
- Wurm, Florian M. 2004. 'Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells.' *Nature Biotechnology* 22 (11): 1393–98.
- X, Y**
- Xiang, Yaozu, Rens De Groot, James T.B. Crawley, and David A. Lane. 2011. 'Mechanism of von Willebrand Factor Scissile Bond Cleavage by a Disintegrin and Metalloproteinase with a Thrombospondin Type 1 Motif, Member 13 (ADAMTS13)'. *Proceedings of the National Academy of Sciences of the United States of America* 108 (28): 11602–7.

- Xu, Xun, Harish Nagarajan, Nathan E Lewis, Shengkai Pan, Zhiming Cai, Xin Liu, Wenbin Chen, et al. 2011. 'The Genomic Sequence of the Chinese Hamster Ovary (CHO)-K1 Cell Line.' *Nature Biotechnology* 29 (8): 735–41.
- Yadav, Lalita, Naveen Puri, Varun Rastogi, Pranali Satpute, Riyaz Ahmad, and Geetpriya Kaur. 2014. 'Matrix Metalloproteinases and Cancer - Roles in Threat and Therapy.' *Asian Pacific Journal of Cancer Prevention : APJCP* 15 (3): 1085–91.
- Ye, S, and E J Goldsmith. 2001. 'Serpins and Other Covalent Protease Inhibitors.' *Current Opinion in Structural Biology* 11 (6): 740–45.
- Yoshimasu, Mark A, Takuji Tanaka, Jong-Kun Ahn, and Rickey Y Yada. 2004. 'Effect of N-Linked Glycosylation on the Aspartic Proteinase Porcine Pepsin Expressed from *Pichia Pastoris*.' *Glycobiology* 14 (5): 417–29.
- Yu, Chao, Max Crispin, Andreas F-P Sonnen, David J Harvey, Veronica T Chang, Edward J Evans, Christopher N Scanlan, David I Stuart, Robert J C Gilbert, and Simon J Davis. 2011. 'Use of the Alpha-Mannosidase I Inhibitor Kifunensine Allows the Crystallization of Apo CTLA-4 Homodimer Produced in Long-Term Cultures of Chinese Hamster Ovary Cells.' *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 67 (Pt 7): 785–89.

## Z

- Zakharova, Elena, Martin P Horvath, and David P Goldenberg. 2009. 'Structure of a Serine Protease Poised to Resynthesize a Peptide Bond.' *Proceedings of the National Academy of Sciences of the United States of America* 106 (27): 11034–39.
- Zanardelli, Sara, James T B Crawley, Chan K N Chan Kwo Chion, Jonathan K Lam, Roger J S Preston, and David A Lane. 2006. 'ADAMTS13 Substrate Recognition of von Willebrand Factor A2 Domain.' *The Journal of Biological Chemistry* 281 (3): 1555–63.
- Zhang, Qing, Yan-Feng Zhou, Cheng-Zhong Zhang, Xiaohui Zhang, Chafen Lu, and Timothy A Springer. 2009. 'Structural Specializations of A2, a Force-Sensing Domain in the Ultralarge Vascular Protein von Willebrand Factor.' *Proceedings of the National Academy of Sciences of the United States of America* 106 (23): 9226–31.
- Zhao, Bo, Shutong Xu, Xianchi Dong, Chafen Lu, and Timothy A Springer. 2018. 'Prodomain-Growth Factor Swapping in the Structure of pro-TGF-B1.' *The Journal of Biological Chemistry* 293 (5): 1579–89.
- Zheng, X, D Chung, T K Takayama, E M Majerus, J E Sadler, and K Fujikawa. 2001. 'Structure of von Willebrand Factor-Cleaving Protease (ADAMTS13), a Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.' *The Journal of Biological Chemistry* 276 (44): 41059–63.
- Zheng, Xinglong, Kenji Nishio, Elaine M Majerus, and J Evan Sadler. 2003. 'Cleavage of von Willebrand Factor Requires the Spacer Domain of the Metalloprotease ADAMTS13.' *The Journal of Biological Chemistry* 278 (32): 30136–41.
- Zhou, Minyun, Xianchi Dong, Carsten Baldauf, Hua Chen, Yanfeng Zhou, Timothy A Springer, Xinping Luo, Chen Zhong, Frauke Grater, and Jianping Ding. 2011. 'A Novel Calcium-Binding Site of von Willebrand Factor A2 Domain Regulates Its Cleavage by ADAMTS13.' *Blood* 117 (17): 4623–31.
- Zhu, Jian, Joshua Muia, Garima Gupta, Lisa A Westfield, Karen Vanhoorelbeke, Niraj H Tolia, and J Evan Sadler. 2019. 'Exploring the "Minimal" Structure of a Functional

- ADAMTS13 by Mutagenesis and Small-Angle X-Ray Scattering.' *Blood* 133 (17): 1909–18.
- Zhu, Jianwei. 2012. 'Mammalian Cell Protein Expression for Biopharmaceutical Production.' *Biotechnology Advances* 30 (5): 1158–70.
- Zou, Zhongcheng, and Peter D Sun. 2006. 'An Improved Recombinant Mammalian Cell Expression System for Human Transforming Growth Factor-Beta2 and -Beta3 Preparations.' *Protein Expression and Purification* 50 (1): 9–17.