



UNIVERSITAT DE  
BARCELONA

## Nucleic acid and small molecule recognition by Smad transcription factors and co-factors

Błażej Mikołaj Bagiński



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

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

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





UNIVERSITAT DE  
BARCELONA

# Nucleic acid and small molecule recognition by Smad transcription factors and co-factors

Universitat de Barcelona  
Facultat de Biologia  
Programa de Doctorado en Biomedicina  
101117 Metabolisme, senyalització metabòlica i patologies associades

Work presented in this thesis was carried out in the  
Institute for Research in Biomedicine (IRB Barcelona),  
Parc Científic de Barcelona,



INSTITUTE  
FOR RESEARCH  
IN BIOMEDICINE



Barcelona Institute of  
Science and Technology

Memòria presentada per Błażej Mikołaj Bagiński per optar al grau de  
doctor per la Universitat de Barcelona.

DIRECTOR  
Dr Maria J. Macias

CO-DIRECTOR  
Dr Radosław Pluta

BŁAŻEJ M  
BĄGIŃSKI

DOCTORANDO  
Błażej Bagiński

TUTOR  
Dr Antonio Zorzano

Barcelona, September 2022



## Declarations

All of the presented results are the author's original work and contribution unless otherwise acknowledged. Some parts of presented work have already been published as cited (September 2022).

The work described in this thesis was conducted in Structural Characterization of Macromolecular Assemblies laboratory, directed by the ICREA research prof. Dr Maria J. Macias. This laboratory is a part of the Mechanisms of Disease unit of the Institute for Research in Biomedicine IRB Barcelona.

## Funding

This work was carried out between December of 2018 and September of 2022 and was co-funded by the Horizon 2020 Maria Skłodowska-Curie Actions "cofund BIST predoctoral programme" (PREBIST 754558; December 2018 – December 2022). This grant has also covered multiple courses, workshops and professional travels (incl. samples transportation).

This work was supported by the Spanish MINECO program (BFU2017-82675-P, to Maria J.Macias). FEBS travel grant has covered the registration fee and accommodation during the course PC19-044 Biomolecules in Action II. Access to the ALBA synchrotron was covered by BAG proposals (2018092972, 2019093780, 2020094472 and 2021095308). Access to the European Synchrotron Radiation Facility in Grenoble was covered by the MX-2284 and MX-2356 BAG Barcelona.



**Barcelona Institute of  
Science and Technology**



## Acknowledgements

One of the most important things I learned during this 4-year long journey, is how dispersed, interdisciplinary and complex modern scientific projects are. To obtain even a 'simple' protein structure I needed to coordinate a dozen human and a few robotic hands. With this in mind, I would like to acknowledge and thank certain people without which I wouldn't be able to reach this point.

Chciałbym podziękować moim warszawskim współpracownikom, od których miałem szczęście się uczyć i zaszczyt pracować, w szczególności dr Ewie Skowronek, dr Lucynie Budźko, dr Piotrowi Małeckiemu, dr Dorocie Niedziałek, oraz dr Romanowi Szczepanowskiemu i dr hab. Krzysztofowi Skowronkowi z Międzynarodowego Instytutu Biologii Molekularnej i Komórkowej. Serdeczne podziękowania również tym, którzy przyczynili się do samej możliwości rozpoczęcia przeze mnie tych studiów doktoranckich – osobom z Instytutu Biochemii i Biofizyki PAN (w szczególności prof. dr hab. Ewie Kuli-Świeżewskiej), oraz dr Arkadiuszowi Ciesielskiemu wraz z prof. dr hab. Michałem Cyrańskim z Wydziału Chemii UW. Bez Państwa ta podróż nie mogłaby się rozpocząć.

Quería agradecer a todos los compañeros/as durante esta etapa en el IRB Barcelona. De mi laboratorio, a Dr. Eric Aragón y Lidia Ruiz por su inestimable ayuda en la expresión y purificación de proteínas, Dr. Pau Martín por sus consejos y IT support y Carles Torner por todo el tiempo que me dedico en el laboratorio. También Dr Radosławowi Plucie za wspólne planowanie eksperymentów i porady. Finalmente, a la Prof. Maria J. Macias, por su tiempo, ayuda, consejos y organización de todos los procesos en el laboratorio.

Además, quería dar mis gracias a Dr. Joan Pous de Plataforma Automatizada de Cristalografía (PAC), por todos los años de trabajo que me permitieron aprender muchísima experiencia con robots y a la Dr. Lorena Gonzalez por todo el tiempo dedicado en relación a mi aprendizaje de biología celular y su proyecto de investigación, donde he tenido el placer de participar. A los miembros de mi comisión de doctorado, por su tiempo, ayudas y consejos - Prof. Antonio Zorzano y Dr. Isabelle Brun-Heath (incluyendo sus reactivos) del IRB Barcelona y Dr Roeland Boer, Beamline scientist en sincrotrón ALBA.

Vorrei anche ringraziare i miei amici italiani: Luca, Martina, Margherita e Veronica per il vostro sostegno (im)morale durante tutti questi anni.





## Abstract

Tumour progression largely depends on the signalling networks that direct cell viability, growth and dissemination (metastasis). Transforming growth factor beta (TGF- $\beta$ ) and Bone Morphogenetic Protein (BMP) are cytokines that induce a plethora of physiological functions in all vertebrates, by activating or repressing various signal transducers. One of these is the Smad family of transcription factor proteins, which play an essential role in early embryogenesis, development, cell immunity, homeostasis, tissue repair and many other essential processes during human life. Mutations in the TGF- $\beta$  signalling components (including Smads) inactivate the cell's tumour suppression functions, facilitating the survival of cancer cells.

Upon activation by the receptor, Smads form active homo- and heterotrimeric complexes to interact with DNA sequences (promoters, enhancers) proteins (other transcription factors, co-activators or co-repressors), and also with RNAs. Smad-RNA complexes were reported in the scientific and medical literature; however, this research was largely carried out on a cellular level and the exact details of these interactions remain poorly understood. I have been working on the characterization of Smad-RNA complex formation. I have also determined the structure of a C2 domain present in NEDD4-L, one of the ubiquitin ligases that degrade Smad proteins

In a different line of research, which is aimed at defining new targets for drug screening, I determined the structure of (among others) the protein Deadhead, belonging to the lethal(3)malignant brain tumour signature genes and responsible for the development and sex determination in *D. melanogaster* and several structures of a p38 $\alpha$  (MAPK14) kinase, bound to compounds that regulate its function in cells. This work will pave the way for the future optimisation of these compounds to improve their pharmacological properties.

# Table of Contents

<b>Declarations</b> .....	<b>3</b>
<b>Funding</b> .....	<b>3</b>
<b>Acknowledgements</b> .....	<b>5</b>
<b>Abstract</b> .....	<b>7</b>
<b>Table of Contents</b> .....	<b>8</b>
<b>1. Introduction</b> .....	<b>11</b>
<b>1.1 Drosophila germline-specific thioredoxins</b> .....	<b>12</b>
<b>1.2 Novel, non-canonical p38<math>\alpha</math> inhibitors</b> .....	<b>13</b>
<b>1.3 Smads as transcription factors</b> .....	<b>14</b>
1.3.1 Smad-RNA interactions.....	16
<b>1.4 NEDD4-like C2 domain: mutant and ligands</b> .....	<b>18</b>
<b>2. Objectives of this thesis</b> .....	<b>21</b>
<b>3. Materials and Methods</b> .....	<b>23</b>
<b>3.1 Protein production</b> .....	<b>24</b>
3.1.1 Construct preparation, cloning and culturing .....	24
3.1.2 Protein purification.....	25
3.1.3 Experimental procedure .....	26
<b>3.2 RNA constructs design</b> .....	<b>27</b>
<b>3.3 RNA production</b> .....	<b>29</b>
3.3.1 Generation of the RNA template.....	29
3.3.2 <i>In vitro</i> transcription .....	30
3.3.3 RNA refolding procedure .....	31
<b>3.4 Electrophoretic mobility shift assays</b> .....	<b>32</b>
<b>3.5 Gel filtration</b> .....	<b>32</b>
<b>3.6 Small Angle X-ray Scattering</b> .....	<b>33</b>
3.6.1 Experimental procedures.....	35
<b>3.7 Crystallography</b> .....	<b>36</b>
3.7.1 Sample preparation.....	36

3.7.2	Crystallization.....	37
3.7.3	Diffraction data acquisition.....	40
3.7.4	Obtaining electron density map.....	42
3.7.5	Model building and refinement .....	43
3.7.6	Experimental - crystallization and data collection.....	45
3.7.7	Experimental - data analysis, structure determination .....	47
<b>3.8</b>	<b>In silico procedures.....</b>	<b>48</b>
3.8.1	SimRNA.....	48
3.8.2	NPDock.....	48
<b>4.</b>	<b><i>Results and Discussion.....</i></b>	<b>51</b>
<b>4.1</b>	<b>Structure of the thioredoxin Deadhead.....</b>	<b>53</b>
4.1.1	Biophysical properties of DhD and TrxT.....	53
4.1.2	Structures of DhD and TrxT.....	54
4.1.3	DhD surface charge and DNA binding.....	56
<b>4.2</b>	<b>Novel, non-canonical p38<math>\alpha</math> kinase inhibitors.....</b>	<b>58</b>
4.2.1	Three structures reveal ligand interactions .....	58
4.2.2	Compounds occupy active centre/hinge area.....	60
<b>4.3</b>	<b>Smad-RNA complexes.....</b>	<b>64</b>
4.3.1	Determination of binding characteristics.....	64
4.3.2	Complex characterisation .....	68
<b>4.4</b>	<b>Human NEDD4-like C2 domains.....</b>	<b>73</b>
4.4.1	New structure of the WT N4L C2 domain .....	73
4.4.2	F91I structure, apo and ligands-bound.....	76
<b>5.</b>	<b><i>Conclusions.....</i></b>	<b>80</b>
	<b><i>Bibliography.....</i></b>	<b>82</b>
	<b><i>Appendix.....</i></b>	<b>93</b>
<b>A.1</b>	<b>Supplementary data .....</b>	<b>94</b>
<b>A.2</b>	<b>List of publications.....</b>	<b>101</b>
<b>A.3</b>	<b>List of deposited crystal structures .....</b>	<b>102</b>
<b>A.4</b>	<b>Curriculum Vitae .....</b>	<b>104</b>



## 1. Introduction

To successfully realise my objectives, I partook in two additional projects.

*Drosophila* thioredoxin provided me with skills necessary for obtaining protein structures, while the p38 $\alpha$  for protein-ligand crystallography.

## 1.1 *Drosophila* germline-specific thioredoxins

Extracellular proteins are mostly in an oxidized form, however, the interior of the cell is a reducing environment, which favours thiol groups over, for example, disulphide bridge formation. Thioredoxins (Trx) are a class of small proteins, present in virtually all living organisms, in all cellular compartments, that perform redox reactions. This redox control is essential in maintaining cellular homeostasis, protecting proteins from oxidative aggregation, helping cope with environmental stresses (such as reactive oxygen species), regulating apoptosis and is a required cofactor for many metabolic processes (e.g. as an electron donor for ribonucleotide reductase in DNA synthesis) (Collet and Messens, 2010).

*Drosophila* displays three distinct stages of the post-embryonic life cycle: larva, pupa and the adult fly. The first stage of its life, from fertilisation to egg hatching, lasts for 21~22 hours and among others, determines the sex of the developed fly (Gilbert, 2000). During this work, we determined structures of two germline-specific thioredoxin enzymes are responsible for this transition - Deadhead (DhD) and thioredoxin T (TrxT). Both are small, ~12 kDa monomeric proteins, with the core structure containing three  $\alpha$ -helices and four  $\beta$ -strands, although the typical thioredoxin arrangement contains four  $\alpha$ -helices and five  $\beta$ -strands, with the active centre CXXC motif (CGPC for DhD). The reduction of both cysteines from S2 to (SH)<sub>2</sub> requires NADPH and is catalysed by thioredoxin reductase (Svensson et al., 2003).

Both proteins are involved in the oocyte-to-embryo transition, with DhD being female (expressed in oocytes) and TrxT male germline-specific. The former was proven to be required for early embryogenesis and timing of the protamine-to-histones exchange. This activity is not recognised by another *Drosophila* thioredoxin, Thioredoxin 2, present in both sexes (Petrova et al., 2018). Furthermore, both TrxT and DhD belong to the *lethal(3)malignant brain tumour* signature genes (Rossi et al., 2017) and form part of the survival network of genes mediating DNA damage response (Ravi et al., 2009). Lastly, DhD was proven to interact and/or modulate ribosomal/RNA-binding proteins and ribonucleoprotein in embryos and oocytes (Petrova et al., 2018). Our work has focused on providing additional structural information on these two thioredoxins, which have been overlooked until now.

## 1.2 Novel, non-canonical p38 $\alpha$ inhibitors

p38 $\alpha$  (also known as MAPK14) is a member of the mitogen-activated protein kinase family (MAPK). This is a ubiquitously expressed enzyme that plays a key role in cellular homeostasis and tissue organisation. Amongst others, its substrates include transcription regulators (Whitmarsh, 2010) and cell cycle regulators (Lemaire et al., 2006), forming a reply for the environmental stress conditions or DNA damage, as well as tumour suppressor p53 (Perfettini et al., 2005).

Its structure is similar to other kinases and consist of 360 amino acids, folded into two globular domains, linked by a flexible hinge region that contains the canonical ATP-binding pocket. The N-terminal domain is composed of  $\beta$ -sheets, whereas the C-terminal domain is mostly of  $\alpha$ -helices (Wang et al., 1997a).

p38 $\alpha$  is activated through phosphorylation by MAPK2 in the canonical pathway or by autophosphorylation (Canovas and Nebreda, 2021). A non-canonical pathway relies on binding of a TAB1 co-factor which facilitates this process (Ge et al., 2002). This alternative mechanism of activation has been associated with cardiomyocyte death in myocardial ischemia-reperfusion injury (IRI), a condition that leads to heart failure (De Nicola et al., 2018).

Attempts to limit the activity of this enzyme resulted in the generation of a substantial number of small molecule inhibitors, divided into two different categories. Type I inhibitors, such as the SB203580 bind directly within the active centre-hinge region and compete with the natural ligand (Young et al., 1997). Type II inhibitors, such as the BIRB796 (Pargellis et al., 2002) stabilise the inactive conformation of the protein, severely limiting the outside activation. Unfortunately, these compounds have shown a limited efficacy in clinical trials and none has progressed to phase III (Haller et al., 2020). To overcome these issues two lines of research emerged: binding to p38 $\alpha$  outside of the canonical ATP-binding site to stabilise inactive conformation (Comess et al., 2011) and the use of cell-permeable peptides reducing phosphorylation (Wang et al., 2013).

In this work we investigated a novel series of compounds, blocking autophosphorylation but only slightly reducing the p38 $\alpha$  activity (Nebreda et al., 2020).

### 1.3 Smads as transcription factors

Tumour progression largely depends on the signalling networks that direct cell viability, growth and dissemination (metastasis). Transforming growth factor beta (TGF- $\beta$ ) is a cytokine that induces a plethora of physiological functions in all vertebrates. Therefore, understanding how TGF- $\beta$  acts both in health and in disease is key to revealing the aspects of the signalling cascade that are susceptible to targeting with new therapeutic and diagnostic tools. Mutations in several of the TGF- $\beta$  signalling core components inactivate the tumour suppressor function of the signalling network, and facilitate the survival of cancer cells (Massagué, 2012), (Massagué and Obenauf, 2016), (Macias et al., 2015).

One of the best-studied networks activated by TGF- $\beta$  involves a family of structurally similar and highly conserved signal transducers, named SMADs (Sma- and Mad-related proteins). Smad proteins play essential roles in embryo development, differentiated tissue homeostasis, and immunity responses (Massagué, 1998). The SMAD family is composed of eight members, divided into three distinct sub-types: receptor-activated SMADs (R-Smads, generally Smad2 and 3 for TGF- $\beta$  pathway and Smad1, 5 and 8 for BMP pathway), inhibitory SMADs (I-Smads, Smad 6 and 7) and a single common SMADs (Co-Smad or Smad4). A schematic of the mechanism of action is presented in Figure 1-1. Upon ligand binding formation of fully functional receptor occurs (1) and the receptor kinase can be activated (2). R-Smads upon activation (3) form heterotrimers with one protomer of Smad4 (Co-Smad) and subsequently translocate into nucleus where they bind with various cofactors, recognising selected genes (4). This occurs by recognising and binding specific sequences, called Smad Binding Elements (Jonk et al., 1998). These complexes can be downregulated by the I-Smads (5). Further along the activation route, Smad trimer forms a complex with Trim33, a histone markers binder (6). Finally, the entire assembly can bind additional cofactors (such as FoxH1) to activate regulated genes (7). At the end of the process (8) active Smads can either be deactivated by a phosphatase to be available for the subsequent cycles, or targeted by ubiquitin ligases (such as NEDD4-L) to be degraded (Macias et al., 2015).



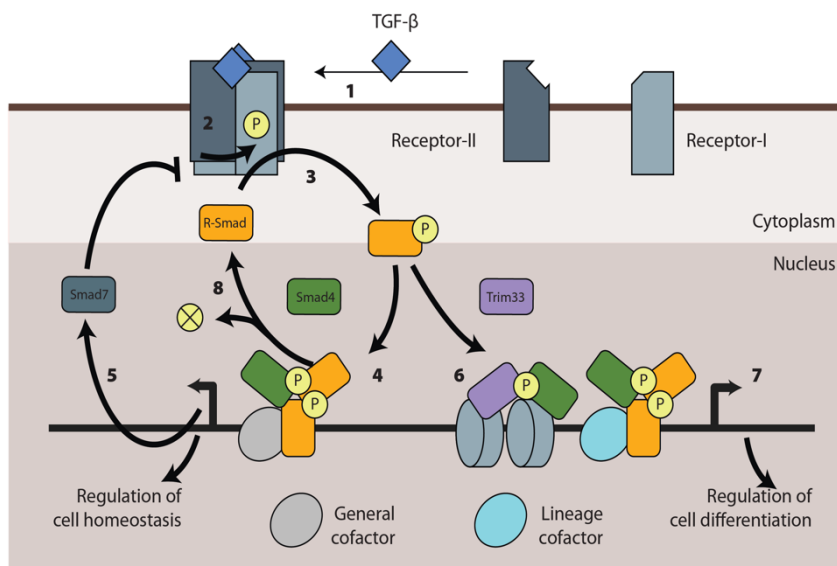


Figure 1-1 Smad mechanism of complex formation and transcription regulation. Another variant of Figure 1A from (Macias et al., 2015) (licence no 5387741078814).

Smads are ~400-500 amino acids long. They fold into two globular regions, called MH1 and MH2 at the N-terminus and the C-terminus, respectively. These two domains are connected by a linker region whose sequence is less conserved than of the domains and often acts as a regulatory hub (Alarcón et al., 2009; Aragón et al., 2011). The MH1 domain (present in R- and Co- Smads) is responsible for DNA binding, while the MH2 domain (present in R-, Co- and I- Smads) acts as a hub for protein-protein interactions with other Smads and transcriptional co-activators and co-repressors (Shi and Massagué, 2003), (Macias et al., 2015).

Structures of several key TGF-β signalling cascade components, such as the TGF-β hormone, the receptor, and some structures of Smad domains are determined. The latter include complexes of MH1 domain with the DNA motif identified in the past, and with the new motifs recently characterized in our laboratory (Martin-Malpartida et al., 2017), (Aragón et al., 2019). Several complexes of the MH2 domains (Miyazono et al., 2018), (Wu et al., 2001) and of the linker bound to co-factors and ubiquitin ligases (also NEDD4-like) (Aragón et al., 2011), (Aragón et al., 2012), (Gao et al., 2009), and reviewed in (Macias et al., 2015) are also available.

### 1.3.1 Smad-RNA interactions

Smad-RNA interactions have been reported in scientific and medical literature, including both indirect and direct Smad protein-RNA binding. Indirectly, a plethora of microRNAs is reported to influence the TGF- $\beta$  pathway leading to atherosclerosis, an example being Smad3 targeting the promoter of miR-4286 (Li et al., 2022). Contrarily, Smad7 is being targeted by miR-92a (Zhang et al., 2019).

The structural aspects of the direct Smad-RNA binding remain poorly understood - for example, Smads interactions with RNA in microRNA maturation and lncRNA-mediated gene expression control remains an open question in the field. On one hand, Smads were shown to bind *in vivo* to CAGAC sequence in pri-miRNAs and the lncRNA GAS5 (Tang et al., 2017). This CAGAC motif and also 5GC motifs are located in the DNA major groove and are the major DNA binding sites of Smad proteins genome-wide (Davis et al., 2010). It is worth noting that there are substantial differences between the structure of the DNA helix (B-form) and that of the RNA helix (A-form), which is characterized by a narrower and deeper major groove that would preclude the standard Smad-DNA recognition mode. Indeed, a study revealed that Smad3 binds poorly to ssRNA and dsRNA *in vitro*, regardless of the presence or absence of known Smad binding sequences (Dickey and Pyle, 2017). However, also that Smad3 binds with apparent high affinity to RNA sequences containing large internal loops or bulges. This affinity is comparable to values obtained for CAGAC and 5GC dsDNA motifs, suggesting a biological role for RNA binding by Smad3.

RNA molecules themselves are characterised by a vast diversity of the created secondary structures. They are a result of non-standard annealing of the complementary strands (Figure 1-2), but can also be formed by a non-canonical base pair. Such pairs deviate from standard Watson-Crick A:U/C:G by either 180° rotation of one of the bases in the Hoogsteen base pairing (Sharma et al., 2010) or nitrogen base swap, as is the case in G•A pair (Olson et al., 2019), U•U (Sheng et al., 2013) or others. These interactions can happen along all three edges of the nitrogen base. Furthermore, an internal bulge forms when such weaker interaction or a complete mismatch occurs within the dsRNA strand.

A better understanding of the sequence and structural features required for RNA binding by Smads would improve our understanding of miRNA biology, lncRNAs involvement in TGF- $\beta$ /Smad gene expression regulation, protein-RNA recognition in general, and the potential for Smads to interact with other RNA ligands in the cell.

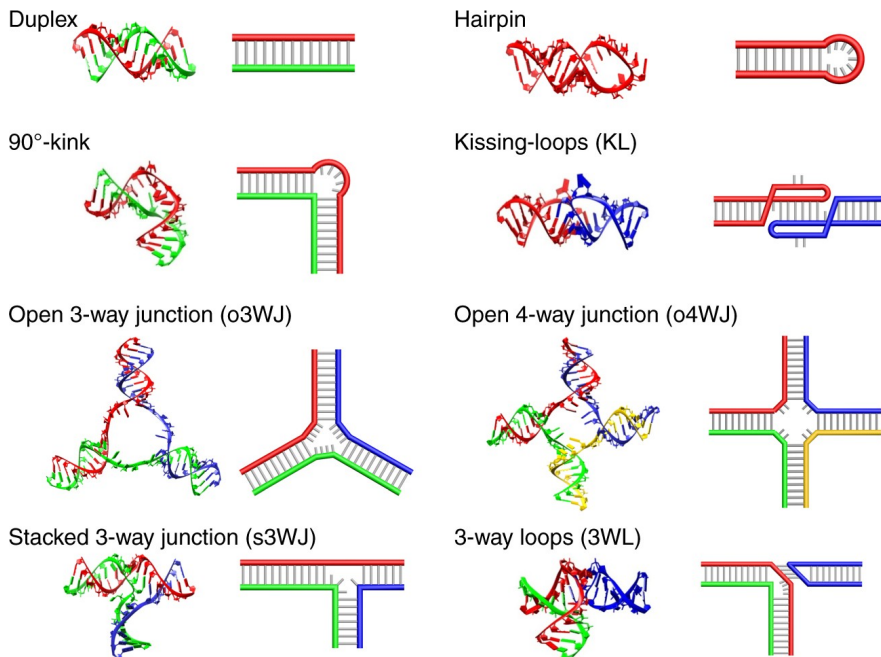


Figure 1-2 RNA exhibit a substantial variety of motifs, aided by canonical and non-canonical base pairing. Source: Figure1 of (Li et al., 2018), Creative Commons Attribution 4.0 license.

## 1.4 NEDD4-like C2 domain: mutant and ligands

Ubiquitylation is one of the major processes controlling biological signalling. One of its main mechanisms of action is marking for degradation of misfolded, damaged or unnecessary proteins by covalently linking (one or more) protein tag (ubiquitin), usually to the protein N-terminus, but also to the amino group of the Lysine sidechain. This results in directing the entire complex for proteasome degradation. Ubiquitin is a small, 76 residues, ~8,6 kDa protein (Swatek and Komander, 2016).

NEDD4-L (neural precursor cell expressed, developmentally down-regulated 4-like, N4L) is a ubiquitin ligase of the NEDD4 family. It consists of an N-terminal C2 domain (binding phospholipids), 4 WW domains (responsible for protein-protein interactions) and the HECT domain (ligase domain) (Escobedo et al., 2014). Functionally, the C2 domain requires calcium to operate (Nalefski and Falke, 1996).

The protein has a clinical significance in Parkinson's disease (Conway et al., 2022), prostate, breast, bladder and pancreatic cancer (Chen and Matesic, 2007), or as a regulator of renal epithelial sodium channel (in lung and kidney). In Liddle's syndrome the channel protein, due to the mutation, can no longer be recognised by the N4L and degraded, leading to excess sodium and water resorption (Enslow et al., 2019).

N4L is a co-factor of Smad regulation pathway. It is structurally similar to the Smurfs (Smad ubiquitin regulatory factors 1 and 2), a previously identified ubiquitin ligases for R-Smads and TGF- $\beta$  superfamily receptors. It was proven not only that N4L acts analogically (inducing Smad degradation), but also binds Smads both indirectly (Smad2 and 3 via the Ski-related novel protein N complex) and directly (Smad7) (Kuratomi et al., 2005), (Tang et al., 2011).

A mutation in the N4L C2 domain, F91I, was identified in a genetic screening of a patients suffering from periventricular nodular heterotopia. This type of epilepsy is caused by improper migration of the neurons during the early development of the fetal brain (Aghakhani et al., 2005). Other mutations of N4L are also reported to result in polymicrogyria, cleft palate and syndactyly, albeit in the HECT and WW domains (Stouffs et al., 2020).





## 2. Objectives of this thesis

### **1. Discerning differences between DhD and TrxT proteins**

By describing these specific germline proteins, we aim to increase the understanding of this model organism and create a possibility to design new pest and disease control vectors.

### **2. Verifying the mechanism of action of novel p38 $\alpha$ inhibitors**

Dysregulation of p38 $\alpha$  signalling has been linked to diseases including inflammation, immune disorders, and cancer. By describing this novel, non-canonical inhibitors, we aim to lay foundation for their future optimisation.

### **3. Smad-RNA complex characterisation**

By verifying and characterising the Smad-RNA binding we aim to aid to the discussion about non-canonical Smad binding partners and potential implications for human health and disease.

### **4. Characterisation of Smad co-factors**

We aim to help better explain Smad pathway regulation and influence by expanding the structural knowledge about other protein co-factors.





## **3. Materials and Methods**

### 3.1 Protein production

Proteins are expressed according to the regulation processes happening naturally in all living cells. However, it is possible to hijack some of these processes to force living cells to produce large amounts of the protein of interest. Because of their speed of growth and cost-effectiveness, *Escherichia coli* bacteria are the preferred host production organism (if suitable for the particular protein).

In the case of the investigated proteins, it was possible to apply a standard laboratory protocol of protein production (of DNA design, plasmid preparation, *E. coli* transfection and protein purification) to obtain all proteins used for studies described in this thesis. Briefly, this project is divided into several stages (Figure 3-1).

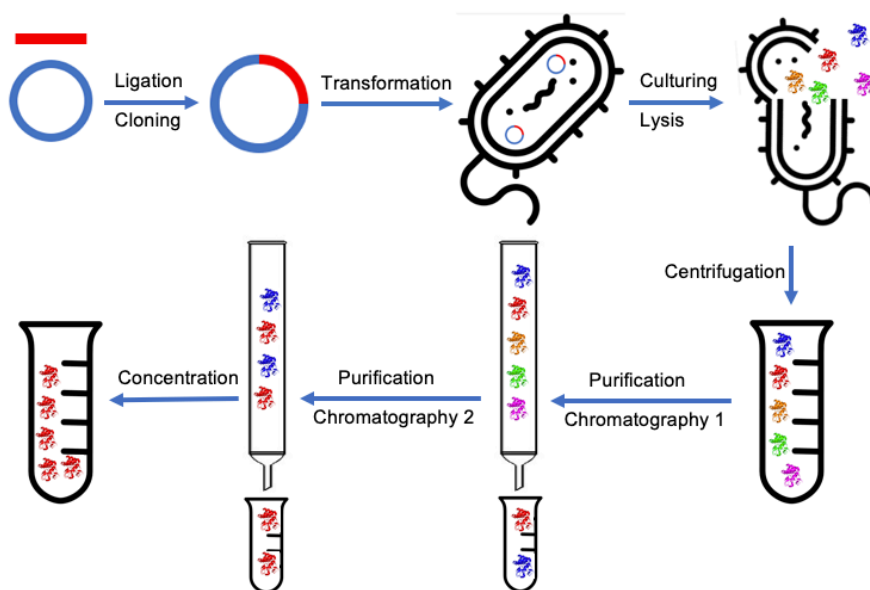


Figure 3-1 Recombinant protein purification cycle in the *E. coli* host organism.

#### 3.1.1 Construct preparation, cloning and culturing

First, an artificial gene is prepared. Its nucleotide sequence can be directly taken from a database (such as the GenBank) or preferably, by taking an amino acid sequence of the protein (or a fragment of the protein) and translating it to the nucleotide sequence. In both cases, it is advised to optimize the sequence for efficient use by the production organism by replacing some tri-letter codons with the identical meaning, but more

abundant within the production organism cell, to take advantage of the varying levels of tRNA due to the degeneracy of the genetic code.

This information has to be then delivered, in understandable format, to the host organism. This is usually done by inserting a designed sequence into the plasmid expression vector - a circular piece of DNA containing a regulatory sequence (promoter) and an antibiotic resistance gene (used in the selection process), with the most popular example being pET containing a T7 promoter. Various enzymes are used in this process, such as restriction enzymes cutting dsDNA in a precise location and ligases joining DNA phosphate-deoxyribose backbone.

The verified plasmid can be then inserted (transfected) into the selected host organism, usually, an *E. coli* strain specialised for protein expression. This is usually done through varying environment temperature, inducing stress response in bacterial cells and creating pores in their cell membrane, allowing the plasmid DNA to be uptaken into the cell (heat-shock method). Once successful, cells can be grown in the antibiotic-containing media. Upon reaching the desired cell count, the promotor governing the production of the protein of interest can be activated, beginning the biological synthesis.

Section based on (Rosano and Ceccarelli, 2014).

### 3.1.2 Protein purification

In the vast majority of the cases, expressed proteins stay within bacterial cells, necessitating cell lysis as the first step of the purification process, followed by the centrifugation or filtration of the cellular debris.

Chromatography procedures are extensively used in protein purification, with affinity chromatography being usually the first cleaning step. As the protein constructs are artificially designed, it is possible to add a purification tag at the C-terminus tail which would selectively bind to the resin in the column (such as a his-tag or SUMO tag). For structural biology studies, such tags should be removed after the procedure via selective protease enzymatic digestion. Proteins can subsequently be purified further using other types of chromatographies, such as an ion exchange (selectively binding proteins based on their surface charge) and gel filtration (mechanical separation by size). Reversed-phase chromatography has limited use with protein.

Section based on (Rosano and Ceccarelli, 2014).

### 3.1.3 Experimental procedure

Smad2 $\beta$  isoform lacking the E3 insert - MH1 domain (S2-E3) was purified as described in (Aragón et al., 2019), Smad3 (P84022) Pro10-Pro136 (MH1 domain; S3) was purified analogically. In short, N-terminal His-tagged fusion protein followed by a TEV cleavage site in pETM11 plasmids were expressed in *E. coli* BL21(DE3) at 37°C until OD600 of 0,6~0,8 and induced with IPTG (20°C ON). Purification from the supernatant of centrifuged lysate was carried out using affinity (HiTrap Chelating HP column) in NGC Quest 10 Plus Chromatography Systems (BIO-RAD), using 40 mM Tris-HCl pH 8,0, 400 mM NaCl, 40 mM imidazole, 2 mM TCEP and 0,1% Tween. The elution buffer contained 400 mM imidazole and no Tween. Proteins were then digested with TEV protease (4°C ON) during dialysis to the buffer containing 20 mM Tris-HCl pH 7.2, 100 mM NaCl and 2 mM TCEP. The same buffer was used in size exclusion chromatography (HiLoad Superdex 75 16/600 prepgrade column), this was also the final buffer. In the case of Smad3 MH1, the second affinity chromatography step was performed due to the incomplete TEV digestion. It was performed in the benchtop gravity column setup using the same resin (but fresh, uncontaminated) and buffers as in the first chromatography.

The quality of obtained protein (purity and homogeneity) was assessed using standard denaturing protein gels (6% acrylamide resolving gel in a standard buffer containing SDS), in addition to the gel filtration itself. Additionally, the RNase contamination was assessed for each batch of protein by its incubation with reference RNA at 37°C for at least 30 minutes before performing nucleic acid denaturing gel (6/12/15% 19:1 acrylamide:bisacrylamide in a standard TBE buffer + 6M urea at 20 W/gel).

One of the batches of Smad2-E3 and Smad3, as well as Deadhead (Freier et al., 2021) and NEDD C2 WT and mutant, were kindly provided by Lidia Ruiz and Dr Eric Aragon.

### 3.2 RNA constructs design

The design process of this short, structured molecule was based on the substitutions and truncations of the nucleotides, guided by bioinformatic analysis with secondary structure predictions offered by the ViennaRNA Web Services package (Gruber et al., 2008), particularly RNAfold WebServer (Hofacker et al., 2002). This type of software tries to predict a secondary fold of the molecule by trying to minimise the estimated free energy of the entire sequence (following the laws of thermodynamics in minimising enthalpy). Additionally, a heatmap representation of predicted base pair probability (red being low, green being high) is superimposed on the drawings, allowing for visual guidance on the predicted hydrogen bond presence and strength.

Furthermore, designs were verified using the RNAcofold WebServer (Bernhart et al., 2006). This software assesses the probability of dsRNA formation (self-complementarity), which would be undesired in these experiments as it will not only prevent the folding of designed RNA hairpins but also create an undesired heterogenous population during the subsequent experiments (in the case of SAXS even making their results unreliable).

Investigated RNA constructs were designed based on miR-21 (Davis et al., 2010) and a family of constructs derived from domain four of a group IIC intron by Dickey and Pyle (Dickey and Pyle, 2017), especially construct 17e, and are presented in Figure 3-2.

For comparison purposes (as a more complicated system), a modified aptamer part (sensing azaaromatic compounds) of the yjdF riboswitch (Rfam family RF01764, part of sequence AJ938182.1 of *Staphylococcus aureus* RF122) was selected. Similarly to the tRNA, it contains three hairpins and one asymmetric stem linked by 4-way junction. Computer simulations indicate its tertiary structure being similar to the L of tRNA.

The sequences of all described constructs are listed in Supplementary Table 1.

## Materials and Methods

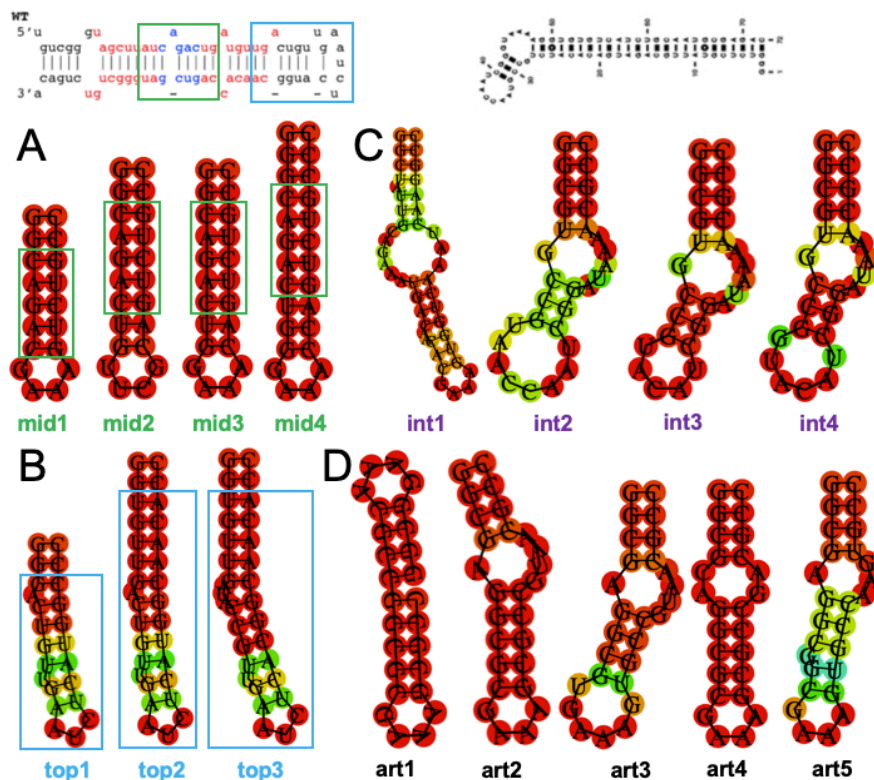


Figure 3-2 Designed RNA molecules that progressed to the wet laboratory stage. Groups A and B base on miR-21 (Davis et al., 2010), group C on the group IIC intron derived construct 17E (Dickey and Pyle, 2017), and group D is composed of artificially created constructs.

4 main groups of various length RNA hairpins were designed. Groups A and B (Figure 3-2) were based on miR-21. Group A consist of constructs investigating sequence binding, containing putative Smad-RNA binding sequence in their stems of various lengths ('mid' for middle, Figure 3-2A).

Group B aim to verify the recognition of the apical loop of this hairpin (the most prominent structural feature), presenting fragments of microRNAs apical loop sequence ('top', Figure 3-2B).

The third group was based on the family of constructs derived from domain four of a group IIC intron ('int' for intron, Figure 3-2C). Two binding spots are presented in its EMSA profile (Dickey and Pyle, 2017), Figure 4), which we presumed to be the activity of the apical loop and the asymmetric bulge, not of dsRNA steam. As such, a truncated construct was created (int1), as well as its derivatives either stabilizing or destabilizing the connection between these unpaired regions (int2-4).

Finally, the fourth group ('art' for artificial, Figure 3-2D) consist of artificial molecules exhibiting diverse structural features, such as art1, a dumbbell-shaped construct with double GNRA loop GAAA, and art2 and its derivatives, featuring one of the same loops and one asymmetric bulge.

### 3.3 RNA production

Designed short RNAs were chemically synthesized by biomers.net GmbH, upon reception resuspended in 10mM Tris-HCl pH 7.2. Inhouse enzymatic production was also tested, but for such short constructs, the yields were unsatisfactory.

#### 3.3.1 Generation of the RNA template

On the contrary, yjF aptamer was produced in-house via the *in vitro* transcription (IVT). In this process, the single-stranded RNA is synthesised on the matrix of DNA, with T7 RNA polymerase enzyme performing the reaction in the artificially created environment. Due to the preferences, product properties and final product homogeneity required for structural biology use, RNA production protocol was slightly modified from the usually used in the laboratories (e.g. as described in (Beckert and Masquida, 2011)).

The given RNA sequence was led by the standard class III T7 promotor sequence (Supplementary Table 1), which enforced change to the first two nucleotides of the construct, due to their influence on the enzyme yield. These were replaced with GG (and thus the second and third nucleotides from the end were replaced by CC). Furthermore, to stabilize the construct's stem the first three nucleotides were removed, and a further 6 (positions 89-95) were turned into complementary to the beginning of the construct.

DNA template was produced via the PCR assembly process, for which all DNA primers were generated with the help of the Primerize Web Server (Tian et al., 2015). This algorithm provides the sequences of primers which are short and partially overlapping (Figure 3-3, sequences are presented in Supplementary Table 1). The advantage of this approach is that it allows benefitting from a cost-effective synthesis of shorter DNA primers, which after assembly yield longer dsDNA IVT templates in a shorter timeframe than the classical plasmid approach.

## Materials and Methods

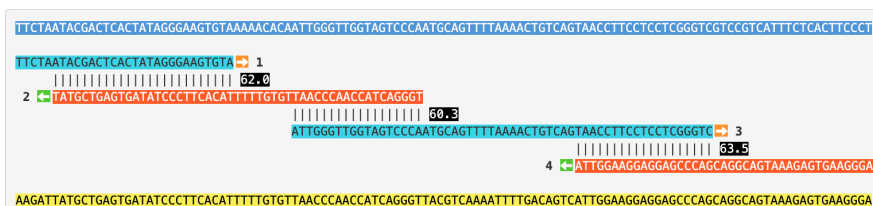


Figure 3-3 An output of the Primerize calculations – set of 4 partially complementary primers (1F, 2R, 3F, 4R) of similar melting temperature (highlighted in black), that were assembled into the dsDNA product (blue and yellow).

Finally, during placing the order at the manufacturer (biomers.net) the second nucleotide of the 4R primer (a penultimate nucleotide of the transcript) was replaced with a chemically modified nucleotide, containing methoxy moieties at the ribose C2 position (C2'-OMe-G), as described in (Kao et al., 1999). This alteration ensured the homogeneity of the 3' end of the final RNA transcript.

### 3.3.2 *In vitro* transcription

Primers were assembled with the PCR reaction consisting of 140  $\mu$ l of the 10x WALK buffer (Supplementary Table 3), 112  $\mu$ l of 2,5 mM dNTPs mix, 28  $\mu$ l of the WALK DNA polymerase and 56  $\mu$ l of each of 4 primers, with primers 2R and 3F diluted 100x (as the internal primers are not constantly consumed in the pace of the reaction) of the total volume of 1400  $\mu$ l divided into 50  $\mu$ l fractions. Initial denaturation was carried out at the 94°C for 180s and the final elongation at 72°C for 600s, with 35 cycles in-between consisting of 15s denaturation at 94°C, 30s of annealing at 60°C and 60s of elongation at 72°C each.

Resulting DNA template was precipitated overnight by adding sodium acetate to the final concentration of 0,3 M and 2,5 volumes of 96% EtOH, incubating overnight at -20°C, centrifuging for 30min, washing with fresh 70% EtOH, centrifuging again and evaporating. It was than resuspended in 10 mM Tris-HCl pH 7.4.

Large scale purification was set up at as three reactions of 1 ml, composed of 100  $\mu$ l of IVT buffer (Supplementary Table 3), 12  $\mu$ l of MgCl<sub>2</sub>, 160  $\mu$ l of 25 mM NTP mix, 20  $\mu$ l of 50U RNase inhibitors, 50U of T7 RNA polymerase 200 U/ $\mu$ l and 60  $\mu$ g of template dsDNA (based on (Kieft and Batey, 2004)). It was subsequently separated into 100  $\mu$ l aliquots and incubated in 40°C overnight.



Resulting reaction was purified using the large-scale denaturing urea-PAGE method. All aliquots were combined, mixed 1:1 with formamide and denatured at 95°C for 3 minutes. They were loaded onto the 3 mm thick gel (10% 19:1 acrylamide:bisacrylamide with 6 M urea and 1x TBE buffer), resolved for 2,5 hours at 29 W (with metal plate attached) and imaged with UV shadowing (265 nm and the fluorescence screen) to facilitate extraction of the band of interest with scalpel. Gel was then crushed through a standard 5ml syringe, eluted into the 10 mM Tris-HCl pH 7.4 overnight and concentrated.

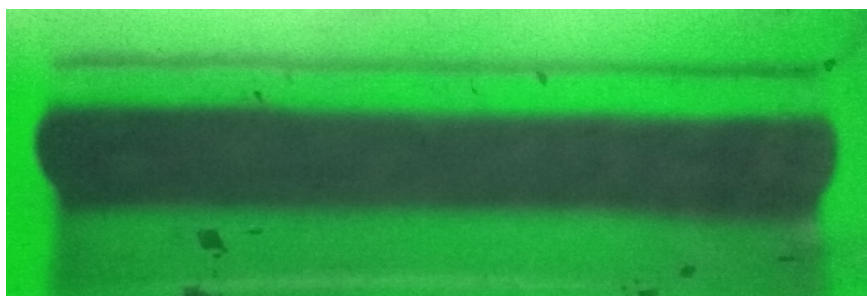


Figure 3-4 UV shadowing visualization of denaturing electrophoresis used to purify IVT reaction. Top band: DNA template, bottom band: RNA product.

### 3.3.3 RNA refolding procedure

Optimisations and verifications of refolding procedures (buffer additives, temperature, time) were carried out for every RNA construct to minimise dsRNA (instead of hairpin) being used in the experiments, which resulted in rejection of some of the designed constructs whose hairpin-dsRNA population heterogeneity could be neither eliminated nor stabilised (data not presented). For most of the constructs, the procedure involved heavily diluting the RNA material, placing a containing 1,5ml plastic tube(s) in a dry heating block of 95°C for 2~3 minutes and snap cooling it by placing it into the metal dry bath block held on ice. The sample state was then verified using standard non-denaturing polyacrylamide gel electrophoresis (12% 19:1 acrylamide:bisacrylamide run at 100V in 4 °C for ~1h with standard Tris-Glycine buffer, 50% formamide loading dye and stained with ethidium bromide) and concentrated (Amicon® Ultra Centrifugal Filters 3kDa).

### 3.4 Electrophoretic mobility shift assays

Electrophoretic mobility shift assay is a variety of gel electrophoresis focused on visualising changes in sample's mobility in relation to the changed weight, physical volume and surface charge at a given pH. All these properties of biological macromolecules are changing upon binding, especially in the case of protein-nucleic acid interactions, which allows for visual assessment of this process (Hellman and Fried, 2007). A major advantage of this technique is the possibility to vary the ratios and concentration of sample components, which greatly extends provided information (allowing for e.g. binding strength and ratio assessment).

The homogeneity of refolded RNA hairpins was assured before the experiment (native polyacrylamide gels described in section 3.3), while proteins were freshly defrosted from 50 $\mu$ l aliquots. Binding buffer composition is provided in Supplementary Table 3, none of the buffers contained EDTA. Fixed amounts of RNA (4,8 $\mu$ M, 400~700 $\mu$ g/reaction) were mixed with protein at indicated molar ratios.

Binding reaction was assembled on ice of 1 $\mu$ l of 10x stock of this buffer, 2 $\mu$ l 50% glycerol and appropriate quantities of RNA, protein and mQ H<sub>2</sub>O, with protein always being added last. The incubation period (on ice) lasted for minimum 20 minutes, with subsequent reaction loading on 1,5mm, 8% 19:1 acrylamide:bisacrylamide native Tris-Glycine gel, resolved in 4°C for ~1 hour with 90V. A positive control (usually a three lanes of reference RNA + no/1:1/1:2 protein) was always included, as well as a reference lane with no protein added.

### 3.5 Gel filtration

Gel filtration is a technique in which the components of the mixture are separated based on their differing molecular weight, size and/or shape, with the use of inert columns (with a bead made of dextran, cellulose, agarose, silica or others). As a general rule, the bigger and less spherical the molecule, the faster it's eluted from the column (Stellwagen, 2009). This allows for the analysis of mixtures, especially beneficial for monitoring binding and macromolecular complex formation properties.

Experiments were performed using Superdex 75 increase 10/300 glass column in the NGC Quest 10 Plus Chromatography Systems (BIO-RAD), at room temperature, with the TEV digestion buffer as the running buffer (Supplementary Table 3). Injected sample volume was below 200 $\mu$ l and the flow was below 0,7 ml/min.

### 3.6 Small Angle X-ray Scattering

Small-Angle X-ray Scattering is a physical technique which uses scattered X-ray radiation (typically below 3 degrees) to quantify the nanoscale properties of an object (Jacques and Trehwella, 2010). A variety of this technique, Wide-angle X-ray Scattering and Small-Angle Neutron Scattering, were not used in the presented work.

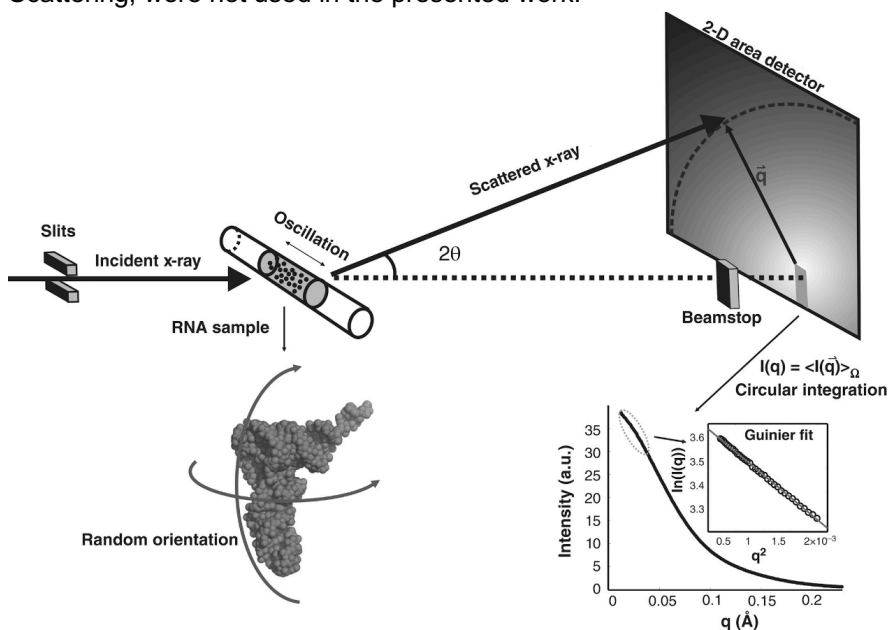


Figure 3-5 A schematic of the SAXS experiment. The flowing macromolecular sample is illuminated with X-ray radiation and the scattering intensities are recorded. Figure 1 of (Chen and Pollack, 2016), licence no 5393060950530.

Investigated sample (in Biological SAXS usually a quartz capillary with a liquid sample) is placed within the monochromatic X-ray beam, and the intensity of scattered radiation is recorded as a function of the distance from the centre of the beam, or a deflection angle (Figure 3-5). The registered radiation intensity is generally of low resolution, but provides high-precision information (Figure 3-6) about the molecule's weight, its shape, hydration shell, maximum dimension  $D_{max}$  and radius of gyration  $R_g$  (average distance to the centre of the density of the molecule against the scattering length density, provided sample maintains its monodispersity (Kikhney and Svergun, 2015)).

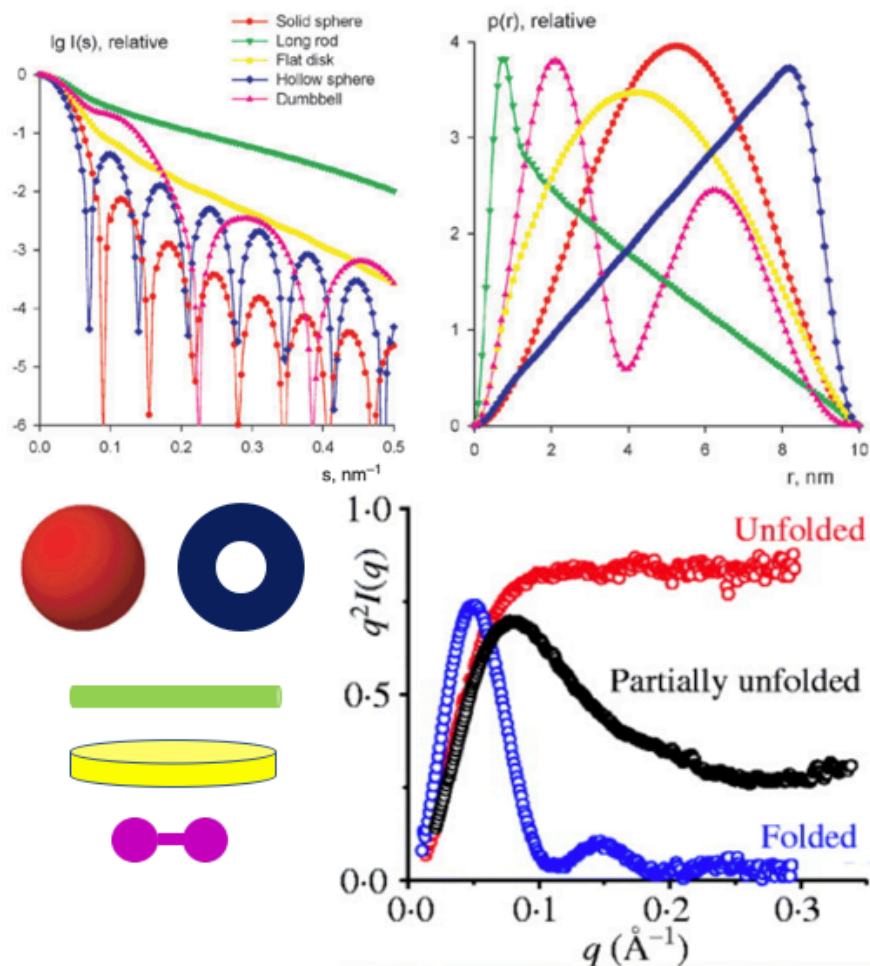


Figure 3-6 Morphological information about the structure of the biomolecule provided by SAXS. An example of scattering, and the  $p(r)$  function for 5 example macromolecular morphologies. Modified from Figure 5 (Svergun and Koch, 2003), licence 1270898-1. Bottom right panel: Kratky plot provides information about the flexibility of the sample. Modified from Figure 24 (Putnam et al., 2007).

In comparison to Cryogenic electron microscopy or crystallography, the experiment does not yield atomic coordinates, however, it maintains an aqueous environment, and is much better suited for investigating larger-sized macromolecules than NMR. These properties make BioSAXS an established method in the biological sciences where it is particularly useful in studying proteins and their complexes (Kikhney and Svergun, 2015).

The main limitations of this technique for protein research are the sample radiation damage and registered signal-to-noise ratio. In the former, the

high energies generated by synchrotrons can cause bond breakages and free radical formation within the sample. Free radical absorbers (such as reducing agents ex. DTT), lowered X-ray dose and/or protein concentration can all be used to prolong the life of the sample (Kuwamoto et al., 2004). In the latter case, the problem is fundamental as the difference in atom's electron densities between nitrogen, carbon and oxygen of proteins and oxygen of water-based buffering solution is extremely small. As a consequence, higher energies are required, but new generations of detectors are optimized to minimise the contribution of background scattering (Mertens and Svergun, 2010).

### 3.6.1 Experimental procedures

SAXS data (RNA, protein, p-R complex in protein concentrations 0,5/1/2/4 mg/ml and RNA 49% molar ratio) was acquired on Beamline 29 (BM29) at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). During the sample preparation step RNA was refolded, protein and RNA buffers were exchanged (Amicon® Ultra 0.5 mL Centrifugal Filters), while final samples were centrifuged for 10 min at 10000 g before data acquisition. Experiments on BM29 were collected at 10 °C, from samples of 45 µL, flowing through 1mm quartz capillary. 10 consecutive frames were taken with 1s exposure time each at 12.5 keV, with 100% transmission (Pilatus 2 M detector).

The matching buffer (XBL) of each sample was collected before and after sample acquisition and subtracted from sample scattering to account for buffer contribution to scattering. Image conversion to the 1D profile, scaling, buffer subtraction and radiation damage assessment was done using the in-house software pipeline available at BM29 (Pernot et al., 2013). Further processing was done by the ATSAS software suite (Manalastas-Cantos et al., 2021). Finally, a 3D model fitting to SAXS data was performed using CRY SOL software (Svergun et al., 1995). This software creates the expected SAXS scattering curve for the given structure file and statistically compares it to the experimental data.

### 3.7 Crystallography

Macromolecular Crystallography (MX, also referred to as Protein Crystallography) is arguably the most powerful method for determining the atomic three-dimensional structures of biological molecules, as it allows for the structural determination of such macromolecules with atomic resolution (Figure 3-7 presents an overview of the experiment). It forms the fundamentals of rational drug design and is vital for linking structure with function, investigating protein folding and relating structural information such as evolutionary relationships.

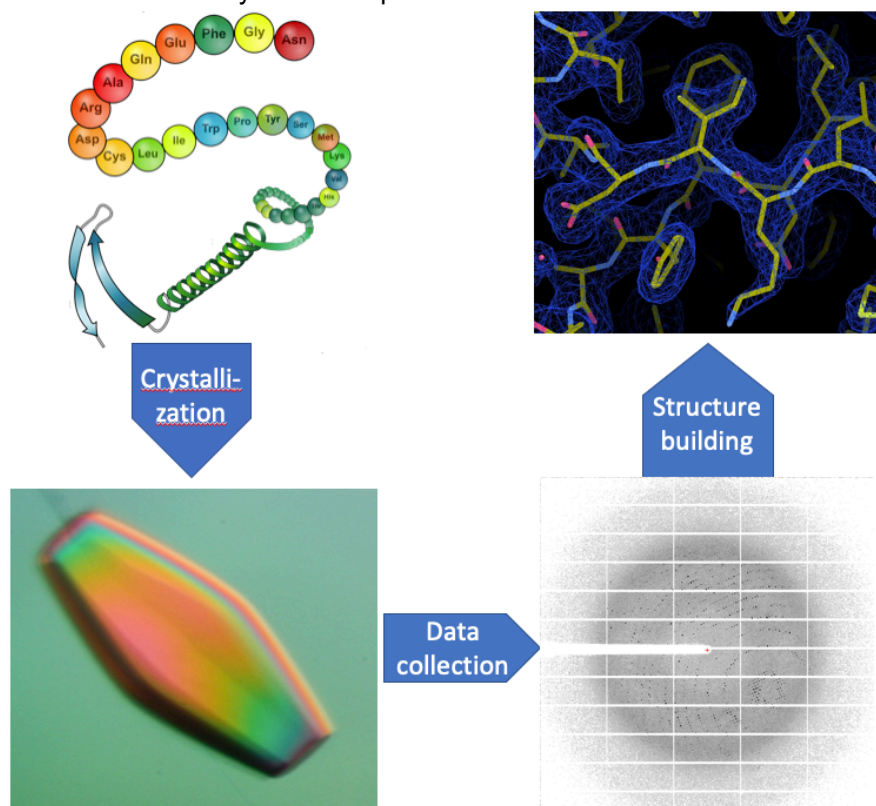


Figure 3-7 An overview of the protein crystallography experiment, aiming to obtain (solve) the structure from liquid protein sample. Top-left picture: seekpng.com.

#### 3.7.1 Sample preparation

For crystallography, the sample is suspended in the aqueous buffer, at the concentration within the solubility of the protein, but allowing for achieving a supersaturated solution (usually at 2~50 mg/mL) (Dessau and Modis, 2011). Additional precipitating agents may be introduced at this stage, as well as additional compounds, e.g. small molecules (so-called co-crystallization) or a particular refolding procedure (for nucleic acids).

### 3.7.2 Crystallization

Protein molecules can crystallize under controlled and regulated conditions, and the aim of the crystallization procedure is to obtain well diffracting crystals, which is a repeatable 3D arrangement of molecules forming a defined 3-dimensional crystal lattice (Figure 3-8). Such an arrangement immobilizes molecules and amplifies scattered X-ray radiation, making it intense enough to be registered by modern radiation detectors. Crystallization of protein/nucleic acid samples can take anywhere from minutes to weeks.

The unit cell is the smallest, periodically repeating in the crystal's x,y and z axis assembly, most often in the shape of cubes or prisms (e.g. rectangular or penta-hexa-octagonal), with defined angles  $\alpha$ ,  $\beta$ , and  $\gamma$  and varying side lengths. There is a total of 230 possible space groups, but for biological macromolecules, only some of them are possible due to the molecule's inherent chirality.

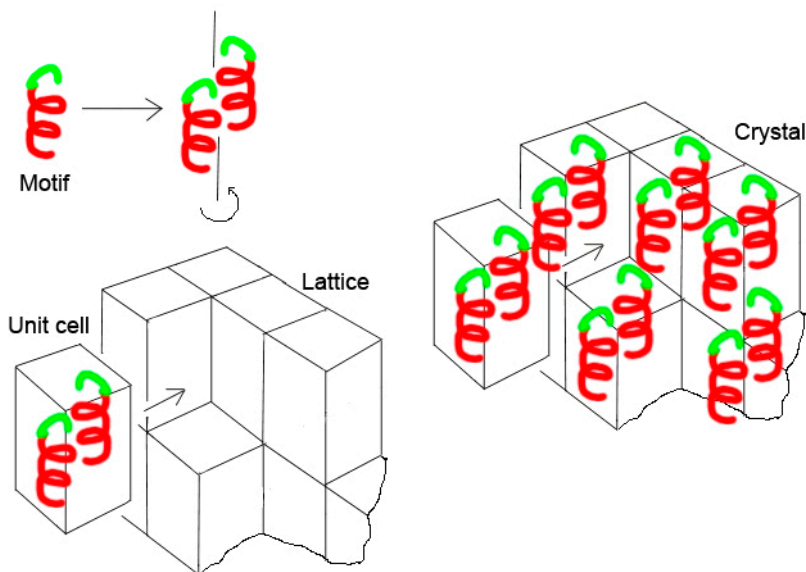


Figure 3-8 A visualisation of the crystal self-assembly process. Proteins favour crystal formation over supersaturated liquid solution. A unit cell is the smallest periodically repeating part of a crystal. There are multiple types of unit cells. Source: "Crystallography-Cristalografia" website [xtal.iqfr.csic.es](http://xtal.iqfr.csic.es).

## Materials and Methods

To greatly facilitate this process, various chemical precipitants are added. Their nature and mechanism of action divide them into 5 flexible categories:

- Salts (organic and inorganic), which provide ions stabilising proteins and helping them overcome surface charges repulsion (ex. cations facilitating nucleic acid crystallization), with the most common being ammonium sulphate,
- Buffers maintain the pH of the environment which affects the protein's surface charges distribution (such as Tris, HEPES or arsenic-based cacodylate),
- High molecular weight polymers (such as polyethene glycol average molecular weight of 3350 Da, or their monomethyl esters) and low molecular weight polymers/monomers/compounds (such as polyethene glycol average molecular weight of 400 Da, ethylene glycol, glycerol), which introduce molecular crowding; 'virtually' increasing sample's concentration by occupying part of the available buffer volume,
- Volatile organic solvents (such as ethanol, isopropanol, acetone, but also formate and acetate) that evaporate more readily than water changing their concentrations quicker, but can also stabilize less water-soluble parts of proteins,
- Additional, sample-tailored additives such as small molecules (ex. drugs) known or perceived to stabilise macromolecule of interest, detergents (in case of membrane proteins), protease inhibitors, metal chelating agents, heavy metal salts used to circumvent the phase problem or any others (based on (McPherson and Gavira, 2014)).

There are multiple ways in which crystallization of the biological macromolecule can be induced, with the most frequently used technique being a sitting drop vapour diffusion experiment.

In this technique protein sample is mixed (in varying ratios) with a precipitant solution (usually a total volume of 150~500 nl), in an enclosed cell containing a space for the said drop (or multiple said drops) and a reservoir (usually 40~100  $\mu$ l) filled with the same precipitate solution. This results in the presence of two liquid mediums within one enclosed compartment: a large volume of pure precipitate solution, and an orders of magnitude smaller volume of precipitate diluted with a protein sample (the drop). The difference in a concentration of both solutions causes water from the drop to evaporate into the reservoir, gradually increasing sample concentration (Figure 3-9, left; based on (Dessau and Modis, 2011)). It is also the easiest type of crystallization experiment to automatize with the



use of modern liquid-handling pipetting robots, which is of essence considering the possible number of possible precipitate combinations. Furthermore, robotics allows for significant minimisation of the volume of the precious protein sample, although structural biology still requires its significant amounts. A variety of this technique, using the hanging drop, is also presented in Figure 3-9.

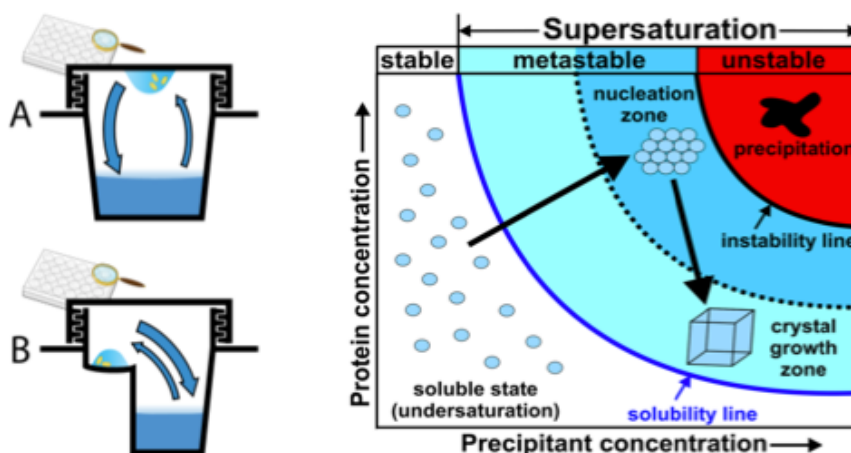


Figure 3-9 Left: Two main types of protein crystallization experiments, based on the exchange of vapours in an enclosed compartment. A: hanging drop variety, B: sitting drop variety. Part of Wikimedia Commons user's Adenosine graphic. Right: phase diagram of protein crystallization. Figure 6 of (Bijelic and Rompel, 2018).

Such Crystallization event occurs in two major steps and both require a supersaturated solution (Figure 3-9, right). The first one, nucleation, is the formation of the crystal nuclei, a “starting point” of the emerging crystal. This process is not yet well understood, with some theories being spontaneously forming partially ordered (paracrystalline) intermediates or statistical fluctuations in solution properties.

The second step, crystal growth, happens when the macromolecules are exceeding their solubility limit (supersaturation) are removed from the solution in an organized way, by being forced to transition to the organized solid phase, expanding the crystalline lattice (McPherson and Gavira, 2014). Unfortunately, another possibility of restoring solubility equilibrium is amorphous precipitation, which usually terminates the experiment.

Interestingly, it may happen that the conditions favourable for nucleation and crystal growth are not the same (or not existing in the same sample drop). In such cases, macroseeding and microseeding procedures can be used as a part of the crystal optimisation steps. In their case crystal (respectively micrometre dimensions, or invisible in the visible light shards

## Materials and Methods

of the crystals) are transferred from the “nucleation” condition to the crystal growth promoting condition.

In one rare case, the author managed to crystallize an RNA molecule without the addition of precipitant in the sample drop, as long as precipitant containing isopropyl alcohol was present in the reservoir.

### 3.7.3 Diffraction data acquisition

The x-rays deflected (“scattered”) by the atoms in equivalent positions in the crystal lattice concentrate into sharp signal spots, of which most can be registered (crystal diffraction pattern). The macromolecular structure can be determined by analysis of the intensities and positions of the diffraction spots.

For this interaction to occur the wavelength needs to be within the diffraction limit of an observed object. With protein atoms having an approximate radius and being separated by distances of the order of  $\sim 1\text{\AA}$  (0,1 nm), respective light spectra fall within the X-ray band of electromagnetic radiation. Nowadays, the diffraction images are usually obtained at the specialised crystallography beamlines, using the synchrotron generated X-rays and captured using CCD detectors, due to their power and automation achieving a high speed of the data acquisition (of minutes per crystal).

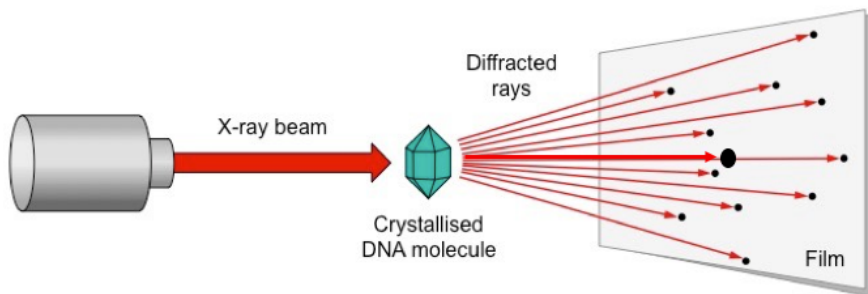


Figure 3-10 A schematic of an X-ray experiment. Modified from [ib.bioninja.com.au](http://ib.bioninja.com.au).

The details of a diffraction experiment are shown in Figure 3-10– some fraction of a coherent, monochromatic X-ray beam is being diffracted when interacting with the electrons of the (protein) atoms constructing molecules within the crystal, creating a pattern that gets registered on the detector. Non-interacting radiation is blocked by a beamstop to avoid detector damage. Individual diffraction images containing hundreds of reflections (diffraction spots) are recorded during small rotation increments of the crystal and combined into a diffraction data set. Four examples of such images are presented in Figure 3-11.

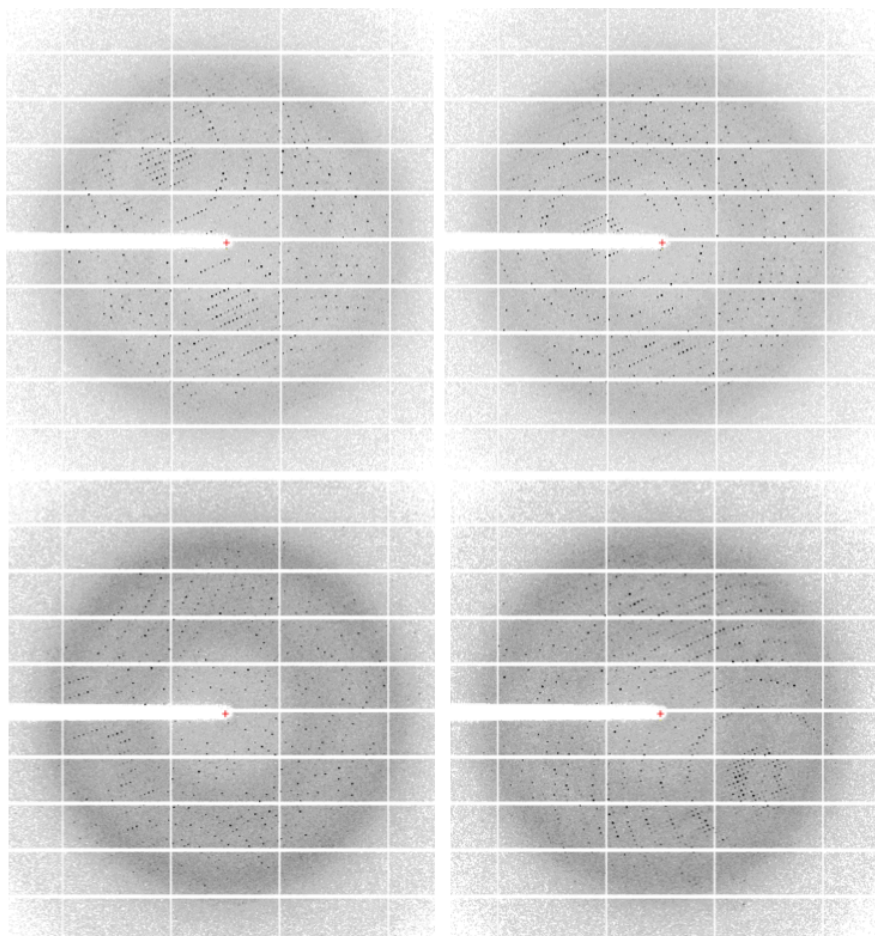


Figure 3-11 Four of the 1200 diffraction images of the Deadhead protein.

One of the most important parameters is the resolution of obtained images, preferably within 1~2,5 Å, which permits further atom placement with high certainty (carbon-carbon bond is ~1,5 Å, but the resolution of 3 Å is generally sufficient to detect enough information for estimating amino acids location). As the first step of data analysis, diffraction spots allow for determination of the unit cell, its dimensions and space group (the larger the cell the more spots visible for each area). The resolution of captured images can be adjusted by physically moving the detector to optimise the use of its surface (as light is being diffracted in the shape of the cone). Unfortunately, the higher the resolution is, the weaker the intensity and quality of the crystal's registered diffraction. In some cases, multiple rounds of crystal growth optimisation are needed to obtain a sufficiently diffracting crystal. Notably, for the majority of the crystals, it is not required to collect

## Materials and Methods

the entire 360° of images, because the internal crystal symmetry of certain space groups forms periodical signal pattern (paragraph based on (Smyth and Martin, 2000)).

Crystals also have to be stable enough in the X-ray beam to allow the measurement of a complete diffraction data set from a single crystal. In this regard, flash-freezing of protein crystals under proper conditions at cryogenic temperatures has significantly reduced radiation damage problems. However, this approach usually requires the transfer of the crystal to a solution containing cryoprotectant additives (or their addition), usually an analogical precipitate solution with increased concentration of glycerol, polyethene glycol of average molecular weight of 400 or 3350 Da, high salt concentrations or others. Their role is to prevent water from forming ice crystals, which would introduce their additional diffraction (sample vitrification, not freezing occurs).

### 3.7.4 Obtaining electron density map

X-rays interact with the electron cloud of a molecule's atoms, thus the atoms themselves are not directly observed. As a result, the raw output of the data processing of the crystallography experiment is the electron density map (Figure 3-12), into which the known sequence of the crystallized protein is fitted in subsequent stages.

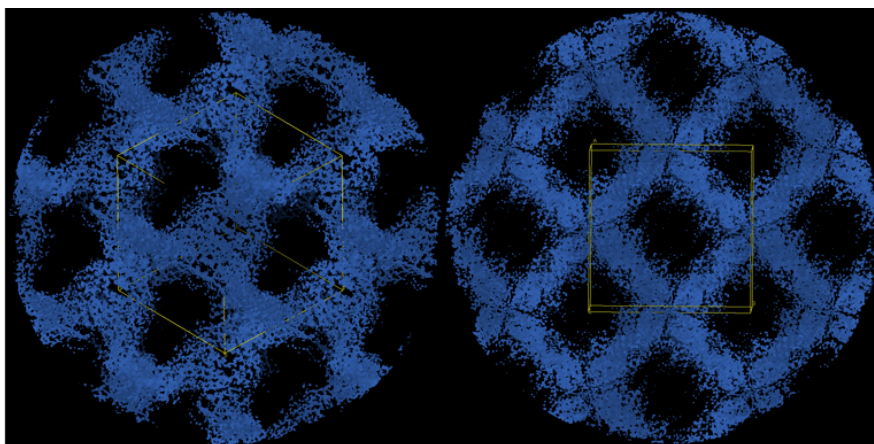


Figure 3-12 Electron density image of the crystal of the Deadhead protein. Asymmetric unit in yellow (tetragonal  $P4_32_12$ ).

However, due to the physical properties of the light (constructive and destructive interference), part of the information is lost before being recorded by the detector and needs to be mathematically recovered. This

can be solved by recalculating measured intensities of the diffraction spots with the estimates for the relative phases of the diffracted rays (estimating these phases is the phase problem in crystallography). There are three most often used ways of solving it:

- direct method – in the case of small molecules, the data is generally of such a high resolution and the number of atoms in the unit cell so manageable that it is possible to ‘brute force’ the problem by computing possible solutions in a reasonable time (even less than an hour),
- Multiwavelength Anomalous Dispersion (MAD) – every chemical element has a certain amount of distinct ‘absorption edges’, a wavelength at which crossing scattering of radiation gets radically altered (in phase and amplitude). By introducing atoms not normally present in the macromolecular crystals (by using selenomethionine during the protein production, or by soaking crystals with heavy metal salts) it is possible to scan around their absorption edges, obtaining additional phase information.
- molecular replacement – if the structure (or a computational model) of a sufficiently similar structure of another protein is available, it can be inserted in various positions and rotations into the unit cell. The structure factors can then be calculated for each position and compared with the experimental set. In case of a close match, the electron density map of the protein of interest can be calculated using the estimates generated from that orientation and translation. This was the method used to solve the structures presented in this thesis (Cowtan, 2001).

### 3.7.5 Model building and refinement

In the last step of the crystallography experiment, the calculated electron density map is filled with the model of atoms of the protein sequence (Figure 3-13). The model building is carried out using programs that graphically display the electron density and allow placement and manipulation of protein backbone markers and residues. Various electron density fitting and geometry refinement tools, as well as automated model building programs greatly accelerate the process. The latter (e.g. used during the molecular replacement) often provide a quite reasonable starting model.

Interestingly, some atom positions can be used to refine the parameters of the map, forming an iterative cycle of model building and refinement. There also exist statistical methods helpful in preventing overfitting the data.

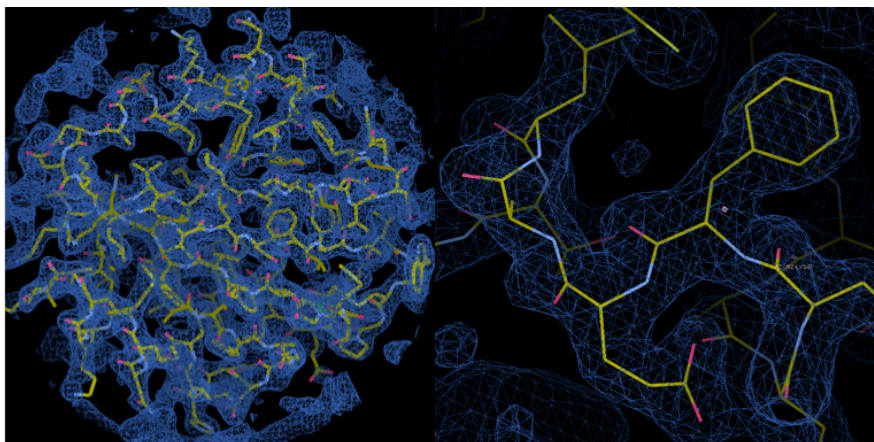


Figure 3-13 Model building step – filling electron densities with atoms and bonds formed by amino acids. Left: starting point provided by the software. Right: closeup.

Parts of the structure may lack appropriate signal quality or resolution, resulting in them being invisible in the final protein structure. These usually correspond to the flexible, or terminal parts of the protein which were not immobilized in the same position across various unit cells. In such a situation, computer modelling may prove helpful.

Lastly, in the structure refinement (solving) process polder maps are a useful tool. It is a type of an omit map, where a mask is applied to a specific region of the density map and bulk solvent is excluded. This allows for the visualisation of weak signals and/or extending the visible ones, but has to be used carefully to avoid introducing bias into the structure (Liebschner et al., 2017).

### 3.7.6 Experimental - crystallization and data collection

High-throughput crystallization screening and optimization experiments were performed at the Automated Crystallography Platform (PAC) of IBMB-CSIC and IRB. Screening and first crystal optimisation steps were usually performed with Phoenix protein dispenser from ARI, and further optimisation experiments with Oryx8 from Douglas Instruments. Solutions preparation and transfers were usually performed with Tecan Evo 100.

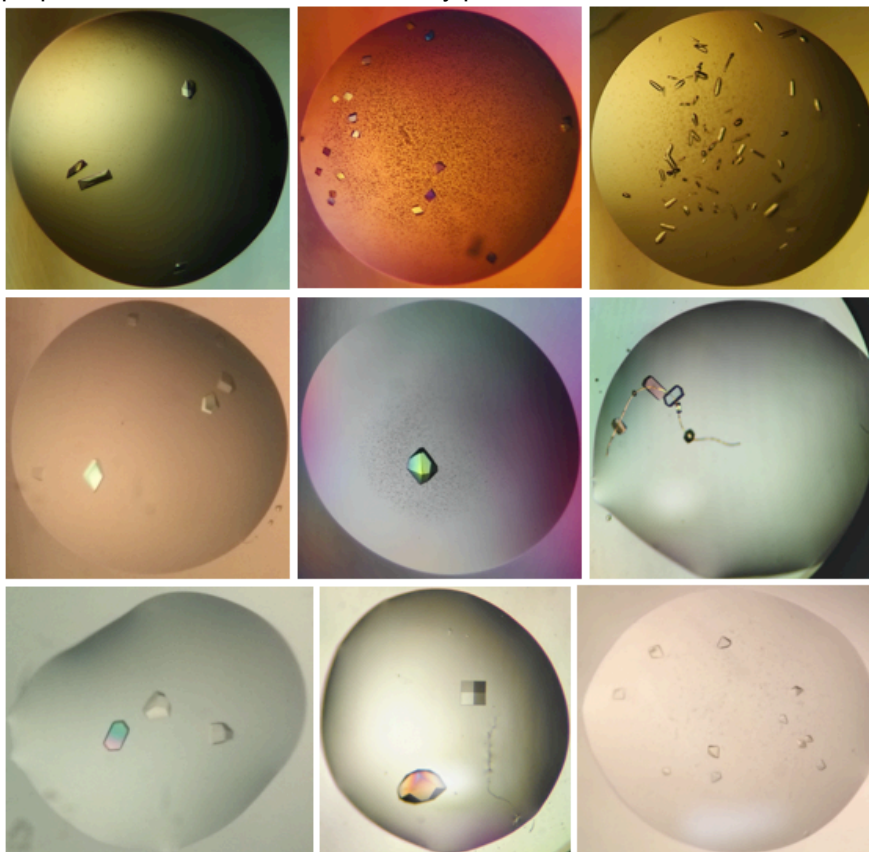


Figure 3-14 Examples of the morphology of author's crystals of various proteins obtained in this work.

Crystallization was usually carried out in sitting-drop vapour diffusion experiments (Figure 3-14) using three-drop 96-well plates by mixing protein and reservoir solution in 2:1, 1:1 and 1:2 ratios with final volumes of 300 nL. Crystals were then monitored with the help of Bruker AXS Crystal Farm Imaging Systems (both 4°C and 20°C).

Table 3-1 presents compositions of final samples, precipitate and cryoprotectant buffers used during flash freezing in standard polyamide

## Materials and Methods

resin loops. Almost all of p38 $\alpha$  samples were crystallized by Dr Joan Pous with the help of Dr Lorena Gonzalez.

Table 3-1 Crystallization details - sample preparation, growth (including protein:reservoir solution ratio) and cryoprotection parameters.

Protein	Sample	Crystallization	p:r	°C	Cryoprotection
Deadhead (DhD) (Freier et al., 2021) <b>6ZMU</b>	10 mg/ml; 20 mM Tris pH 7.2, 100 mM NaCl, 5 mM TCEP	0.1 M sodium citrate pH 5.0, 3.2 M ammonium sulfate, 1:2 protein ratio	2:1	20	0.1 M sodium citrate pH 5.0, 3.6 M ammonium sulphate
WT p38 $\alpha$ with SB203580 and CAS 2094667-81-7 <b>7PV3</b>	10-15 mg/ml, compounds NC-p38i, 1,5x, SB203580 1 mM, ATP $\gamma$ S 500 $\mu$ M; in 20 mM Tris pH 7.5, 100 mM NaCl, 10 mM MgCl <sub>2</sub> , and 5 mM TCEP as described in (Lee et al., 2006)	0.1M HEPES pH 7 + 15%w/v PEG 4000	1:1	4 (15d) + 20 (2d)	10% glycerol
WT p38 $\alpha$ with SB203580 and CAS 2094667-81-7 <b>7Q1A</b>		25 %w/v PEG 3350			10% glycerol
p38 $\alpha$ C162S with SB20358 and CAS 2094667-81-7 <b>7Z6I</b>		0.087 M Ca Acetate, 0.1M MES pH 6.0, 7.1% PEG 550MME		50% glycerol + 25% PEG 3,35k	
p38 $\alpha$ C162S with ATP $\gamma$ S and CAS 2094667-81-7 <b>7Z9T</b>	10 mM MgCl <sub>2</sub> , and 5 mM TCEP as described in (Lee et al., 2006)	27.5% PEG3350, 0.1M BIS-TRIS pH 6.8	1:1	20	[no adicional]
p38 $\alpha$ C162S with CAS2094511-69-8 <b>7PVU</b>		27.0% PEG3350, 0.1M BIS-TRIS pH 6.8			[no adicional]
N4L C2 WT	Protein 7-9 g/l in 20 mM Tris pH 7.2, 250 mM NaCl, 2 mM DTT,	22% PEG smear medium, 0,2 M NaKPO <sub>4</sub> , 10% Glycerol, 0,1 M HEPES pH 7,5	2:1	20	1:1 reservoir : 50% PEG 3350
N4L C2 WT with phosphoserine (not visible)	1,2 mM CaCl <sub>2</sub> , ligand 2x	22,5% PEG smear high, 0,05 M MgCl <sub>2</sub> , 0,05 M Na <sub>3</sub> -citrate, 0,1			1:1 reservoir : 50% PEG 3350



	protein (molar ratio) where applicable	M bis-tris- propane pH 7,8			
N4L C2 F91I		20% PEG 6000, 0,2 M LCl, 10% ethylene glycol			1:1 reservoir : ethylene glycol
N4L C2 F91I with phosphoserine		20% PEG 3350, 0,2 M Na- malonate, 0,2 M Na-acetate			1:1 reservoir : ethylene glycol
N4L C2 F91I with inosito- 1,4,5- triphosphate		30% PEG 8000, 0,2 M Na- acetate, 0,1 M Na-cacodylate pH 6,5			1:1 reservoir : 50% PEG 3350

The X-ray data were collected at 100 K using a PILATUS 6M detector on BL13-XALOC at the ALBA Synchrotron Light Source, Barcelona (Juanhuix et al., 2014) by myself and/or Dr Radoslaw Pluta. In the case of p38 $\alpha$  crystals, data collection was almost always performed by Dr Joan Pous. Detailed statistics of collected data are provided in Supplementary Table 2.

### 3.7.7 Experimental - data analysis, structure determination

Data reduction and processing were either carried out using the autoPROC pipeline (Vonrhein et al., 2011), containing the STARANISO module (Vonrhein et al., 2018), deployed either at the Alba synchrotron or locally. Phenix software suite (Liebschner et al., 2019) was used for the majority of the process, while CCP4 (Winn et al., 2011) installed locally was used for some file conversions. Structures were solved by molecular replacement with the Phaser (McCoy et al., 2007), and refinement was carried out using Phenix.refine (Afonine et al., 2012), manually with Coot (Emsley et al., 2010), aided by the use of the PDB-REDO (Joosten et al., 2014). The detailed statistics of the refinement procedure and structure files are presented in Supplementary Table 2.

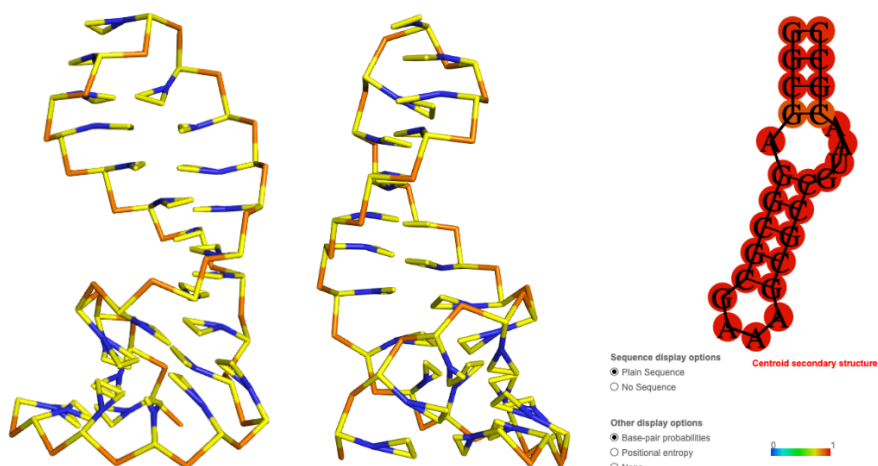
Additionally, polder maps were used to verify the modelling of atoms within electron densities of weaker signal intensity, and the bound compounds. They were calculated using phenix.polder (Liebschner et al., 2017) without searched atoms in place and exclusion sphere radius within 5-10 Å, to minimize bias. The correctness of masks was also verified. Figures were generated with Coot and PyMOL (Open-Source PyMOL, property of Schrodinger, LLC).

## 3.8 In silico procedures

To help analyse the obtained experimental information, especially using visualisation techniques, two bioinformatics tools were implemented in the pace of this work.

### 3.8.1 SimRNA

SimRNA (Boniecki et al., 2016) is an algorithm that can fold RNA molecules *de novo*, using only provided sequence information, via the coarse grained representation of the nitrogen bases, replica exchange of parallel calculations and statistical methods to cluster and estimate biological probability of the results.



Both the webserver available at <https://genesilico.pl/SimRNAweb/> and run locally, was used with the default settings to create models of RNA hairpins designed in the pace of this work. In the case of such short, rigid, uncomplicated molecules (single hairpin variations) obtained models (lowest free energy cluster) were considered equally to structures.

### 3.8.2 NPDock

NPDock (Tuszynska and Bujnicki, 2011), a software designed to predict possible complexes of protein-nucleic acids by performing docking, scoring of possible poses, clustering of the best-scored models and refinement of the highest scored solutions.

The webserver (Tuszynska et al., 2015), available at <http://genesilico.pl/NPDock>, was used to predict the protein-RNA binding site and visualise the assembled complex. It was run with almost default settings, additionally providing only the protein's interface information (amino acids verified crystallographically to bind DNA in the canonical Smad-DNA complexes). Three best results were then assessed and combined.



## **4.Results and Discussion**



## 4.1 Structure of the thioredoxin Deadhead

Structures of the germline-specific Deadhead and thioredoxin T proteins from *Drosophila melanogaster* reveal unique features among thioredoxins, International Union of Crystallography Journal in March 2021 (Freier et al., 2021). My contribution to the manuscript consisted of crystallizing and solving the Deadhead (DhD) structure (PDB ID: 6ZMU) and of verification of unspecific DNA binding by the positively charged DhD protein ((Freier et al., 2021) Figure 5c).

This project was a continuation of Dr Regina Freier's in our laboratory, in the course of which the structure of *D. melanogaster* Thioredoxin T (TrxT) protein was solved (PDB ID: 6Z7O). It is presented in this chapter solely for comparison purposes.

### 4.1.1 Biophysical properties of DhD and TrxT

Sequence comparison revealed that DhD and TrxT are present exclusively in Schizophora (a section of the true flies), which may make them a promising target for pest and disease vector control.

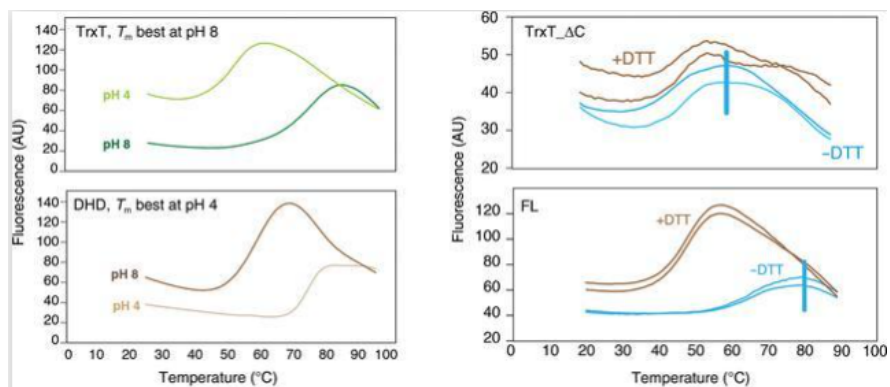


Figure 4-1 Left column: observed increase in fluorescence is related to the unfolding and denaturation of both tightly packed proteins. A clear preference for DhD for pH 4.0 and TrxT for pH 8.0 can be observed. Right column: the stability of full-length TrxT is susceptible to the DTT addition, suggesting an additional disulphide bridge outside of the protein's core. Part of the Figure 2 of (Freier et al., 2021).

An intriguing difference in properties of these recombinant proteins was observed in stability assays (thermal shift assay), where it was found that TrxT is thermally stable in alkaline buffers (pH 8.0) with a  $T_m$  of  $>70^\circ\text{C}$ , whereas acidic buffers (pH 4.0) are necessary for DhD to reach analogical stability (Figure 4-1, left column).

## Results and Discussion

Furthermore, the reducing environment assay (also fluorescence based) revealed that this environment had no influence on the DhD stability, while the  $T_m$  of TrxT was lowered by approx.  $20^\circ\text{C}$ , which suggests the presence of an additional stabilising disulphide bridge outside of the protein's core, within the flexible end part not visible in the crystal structure. To verify this hypothesis, a C-terminal truncation of TrxT was cloned and purified. Although expressed similarly to the extended construct, it showed a melting temperature of  $20^\circ\text{C}$  lower than that of the WT (Figure 4-1, right column). This thermofluorometric analysis was performed by other co-authors of the manuscript (L. Ruiz).

### 4.1.2 Structures of DhD and TrxT

Both DhD and TrxT structures were determined by X-ray crystallography. Interestingly, TrxT crystals were obtained only in the presence of the reducing agent TCEP, while the DhD protein required non-reducing growth conditions.

The asymmetric unit of DhD contains 4 monomers whereas that of TrxT contains only one (Figure 4-2 row A), however, in the case of TrxT, two neighbouring molecules are engaged in a local symmetry stabilized through the coordination of the zinc atom (located between Asp65 and Glu69 of TrxT A and His105 and Glu88 of TrxT B).

Although most thioredoxin structures fold into four  $\alpha$ -helices and five  $\beta$ -sheets (Collet and Messens, 2010), both DhD and TrxT present only four  $\beta$ -strands,  $\beta_2$ – $\beta_5$  (Figure 4-2 row B). The first " $\beta_1$ ", corresponding to the N-terminal region, is not defined as a  $\beta$  strand as it is stabilized by only a single hydrogen bond with its neighbouring  $\beta_3$ . The remaining  $\beta$  strands (2,4,5) run antiparallel, except  $\beta_3$  which is parallel to  $\beta_2$ . This core of the protein is surrounded by four  $\alpha$ -helices (Figure 4-2 row C, left). The active centre (Cys-Gly-Pro-Cys) is located at their interphase, at the C-terminal of  $\beta_2$  strand and N-terminal of distorted helix  $\alpha_2$ . In the TrxT structure, both C31 and C34 of the active site are oxidised, stabilizing the protein's tail end which corresponds to the NMR data.

Interestingly, probably solely because of the buffer conditions, one of the four monomers within DhD's ASU (chain D) also shows oxidation, while chains A, B and C present the reduced form (Figure 4-2 row C, middle and right).



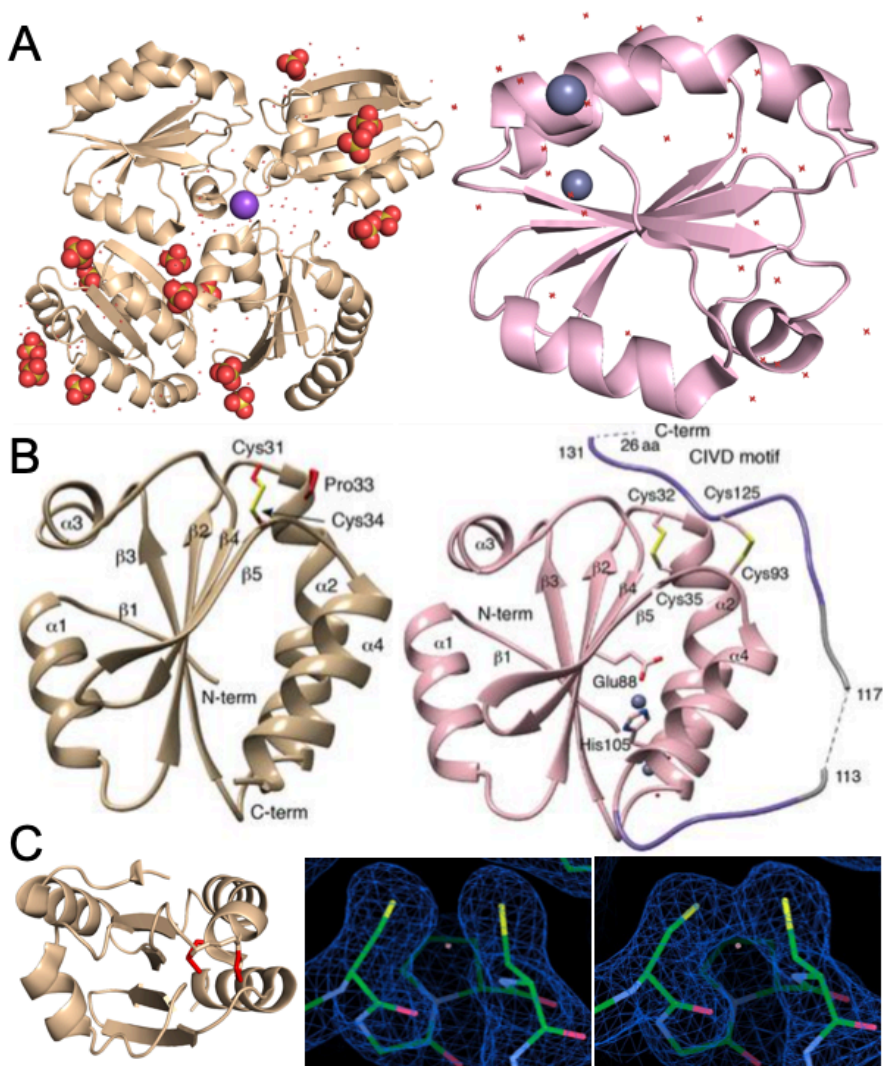


Figure 4-2 DhD (yellow) and TrxT (pink) structures overview. A: Asymmetric units consist of 4 proteins (DhD) and one (TrxT). B: Notable features of both proteins. Part of the (Freier et al., 2021) Figure 3. C: Details of the DhD structure. Left: 'bird's-eye view' showing 4  $\alpha$ -helices surrounding core  $\beta$ -sheets (active centre in red). Middle and right: electron density maps of C31 and C34 of chains B and D respectively showing reduced and oxidised forms within one crystal. P33 is visible in the background.

### 4.1.3 DhD surface charge and DNA binding

TrxT is a typical example of eukaryotic thioredoxins, having negatively charged patches on their surface. However, *D. melanogaster* DhD presents an unusual arrangement of positively charged ones (Figure 4-3A).

Usually, such charge distribution indicates possible interactions with phospholipids (biological membranes) or nucleic acids. Indeed, it was reported that DhD binding partners include (amongst others) ribosomal and RNA-binding proteins (Petrova et al., 2018). A comparison of sequences with other *Drosophila* species was performed (other co-authors) and it was found that these patches of positive charges are indeed being conserved.

To further investigate whether and how this surface charge distribution would facilitate nucleic acid-DhD binding, I performed an EMSA binding assay using the fluorescently labelled DNA (7,5 nM) as a probe (Figure 4-3B). I compared DhD with Lysozyme, a positively charged known electrostatic DNA binder, with BSA ballast and with Smad3 MH1 domain, which selectively binds used DNA construct.

The addition of BSA proved no different for the DNA in all tested concentrations. Smad3, performing also the function of the positive control, was exhibiting a clear, strong, single preferential binding spot on a target DNA. On the contrary, Lysozyme addition resulted in aggregates (unspecific binding) in the low-affinity range, needing ~250 concentration equivalents to bind around half of the sample. As expected, DhD exhibited analogical behaviour, however, the binding was much stronger, showing noticeable interaction visible with as low as 40 molar equivalents. Moreover, these interactions were planned to be investigated with NMR, but attempts proved unsuccessful due to said aggregation.

In conclusion, during this project, two structures of *D. melanogaster* thioredoxins were published, and the differences in the stability, properties and DNA binding were verified. This data will be helpful in understanding the mechanisms that drive sex determination in flies. Given the potential of using *Drosophila* as a model system for studying ageing, anticancer and Parkinson's therapy models (Gonzalez, 2013), we expect that these structures can reveal how the redox system is affected in many diseases. Finally, due to their restricted presence in *Schizophora* (a section of the true flies), these structures can help in the design of small-molecular binders to modulate native redox homeostasis, thereby providing new applications for the control of plagues that cause human diseases and/or bring about economic losses by damaging crop production (Freier et al., 2021).

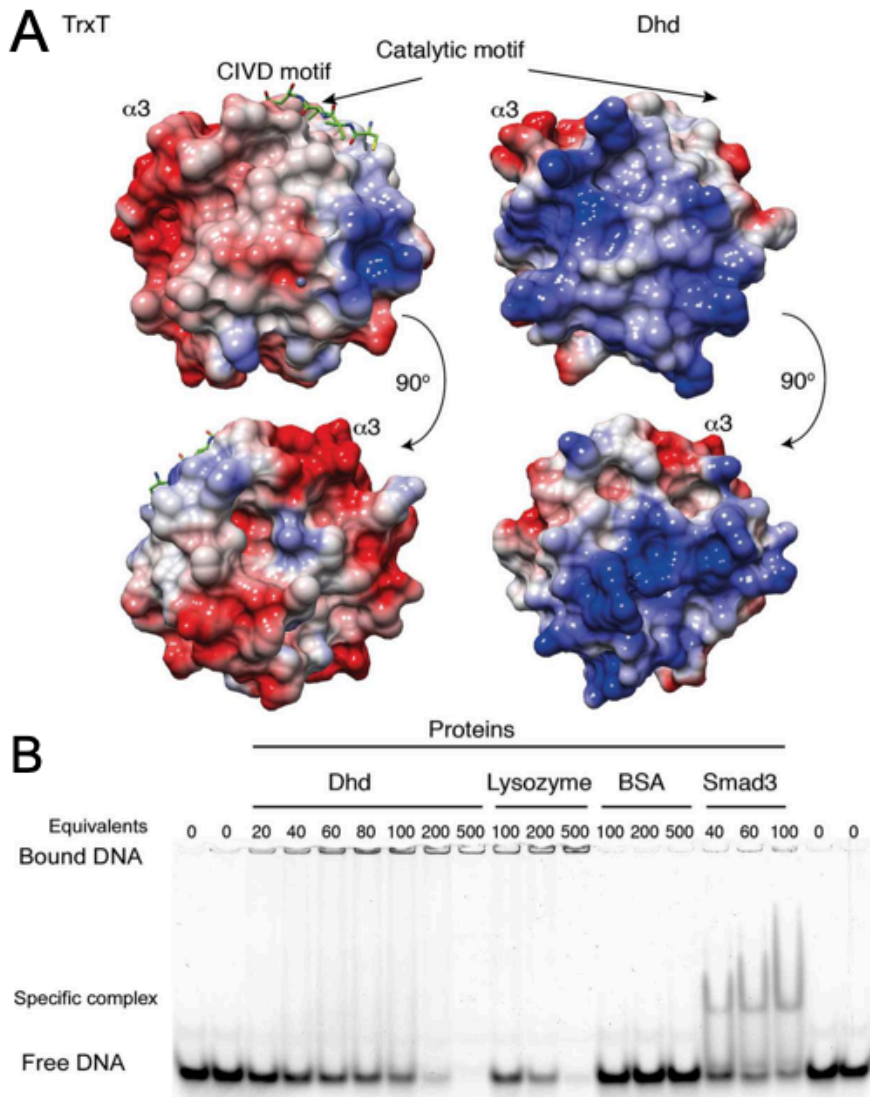


Figure 4-3 A: Surface charge distribution of TrxT and DhD proteins (blue – positive charge, red – negative charge). B: EMSA experiment showing strong, but unspecific DhD binding to the fluorescently labelled DNA (7,5 nM) . Fragment of (Freier et al., 2021) Figure 5.

## 4.2 Novel, non-canonical p38 $\alpha$ kinase inhibitors

### *Structural basis of p38 $\alpha$ autophosphorylation inhibitors that target the non-canonical activation pathway*

González L., Díaz L., Pous J., **Baginski B.**, Scarpa M., Duran A., Igea A., Martín P., Ruiz L., Esguerra M., Collizzi F., Mayor-Ruiz C., Biondi R. M., Soliva R., Macias M.J., Orozco M., Nebreda A. R. In submission, 2022.

This project was a collaboration with IRB's Signalling and Cell Cycle Laboratory, headed by prof. Angel R. Nebreda. During this project, we have obtained 3 structures of the *Mus musculus* p38 $\alpha$  kinase, containing or not two different ligands (CAS 2094667-81-7, further called L37 or 2094511-69-8, further called L38; patent WO 2020/120576 A1) and one canonical inhibitor (SB20358, further called SB): **7Z6I**, **7Z9T**, **7PVU**. Furthermore, one of them contains ATP $\gamma$ S (a stable ATP analogue).

### 4.2.1 Three structures reveal ligand interactions

p38 $\alpha$  is a monomeric protein, composed of around 350-360 amino acids, and its unphosphorylated (Wang et al., 1997b) and SB203580-bound ((Wang et al., 1998), PDB ID 1A9U) structures have been determined. Like other members of the MAP kinase family of proteins, it is composed of two domains (lobes) separated by a deep groove (Wilson et al., 1996), creating a potential binding pocket that was targeted by the described compounds. The N-terminal lobe has a binding pocket for ATP, and the tail end domain performs various functions (such as phosphorylation and magnesium binding). Although two of the presented structures, 7Z9T and 7PVU, contain two molecules in the asymmetric unit, there is no evidence of p38 $\alpha$  being anything other than a monomer in solution (Meng et al., 2002).

The asymmetric units of all five ligand-containing structures are presented in Figure 4-4B. In the case of two proteins within ASU, the chain A designation was assigned to the protein chain exhibiting a stronger ligand signal. The superposition of all the chains A (Figure 4-4A) reveals that globally, besides the variability in the 170-184 loop due to its flexibility, and the amount of resolvable amino acids, all 3 structures are not significantly different.

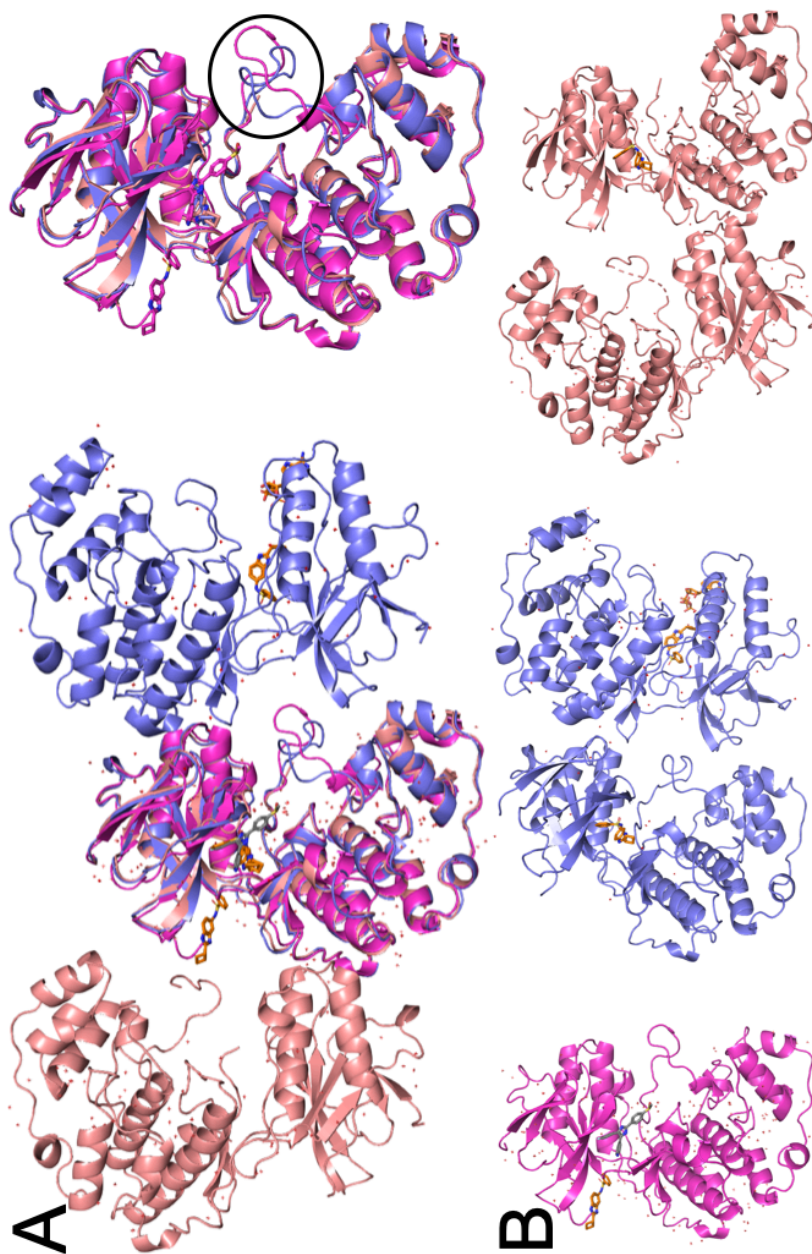


Figure 4-4 Three p38 $\alpha$  crystallography structures containing either of the two ligands. 7Z6I in magenta, 7Z9T in purple and 7PVU in peach. A: superposition of chains A, with flexible loop barring the entrance to the ATP binding site in black circle. B: Asymmetric units of each structures, with ligand marked in orange and canonical SB inhibitor in grey.

### 4.2.2 Compounds occupy active centre/hinge area

SB is a canonical inhibitor of the p38 $\alpha$ . The molecule consists of three rings, one with p-fluoride and one with p-sulfonamide attached to the central imidazole. Our ligands are composed of a central benzodiazole, linking together cyclobutene and benzenesulfonamide (all presented in Figure 4-5A). Of note is a single atom difference between the two ligands (fluoride on the o position of the terminal benzene ring).

One of the structures, 7Z6I, has the SB inhibitor within its active centre (Figure 4-5B). The interacting amino acids are located in the hinge area of the protein, between the two domains, starting from Tyr35 and Val38 initiating the sequence of four (out of five) antiparallel  $\beta$ -sheets (the core of the protein, N-terminal domain). The middle one of these sheets (third) houses Ala51 and Lys53, while the fourth one is Leu104, which directly interacts with the SB's fluoride atom (at  $\sim 3$  Å). Similarly to the Ala51 and Lys53, Leu108 at the "bottom" of the pocket possibly interacts with the SB due to the close distance. Met109 is a known, conserved amino acid which analogically to presented structures, forms a hydrogen bond with its natural ATP substrate (Wang et al., 1998) and Ser154 and Asp168 are located opposite to the five antiparallel  $\beta$ -sheets.

The structures that include two with compounds L37 and one with L38 reveal that ligands predominantly occupy the active centre cavity (region below the G-rich loop also known as P-loop), except for 7Z6I (Figure 4-5C, left), where ligand L37 is located "behind" it, in the hinge region (interacting with Lys79, Gly85, His107 and Leu353 close to C-terminal and bridging to Val183 of protein's symmetry mate). Chains A of structures 7Z9T and 7PVU (Figure 4-5D) presents both compounds L37 and L38 (respectively) in virtually the same position within the active centre (superposition in Figure 4-5E), forming three hydrogen bonds with the residues Lys53 and Asp168 and indicating that fluoride atom does not hinder ligand's interactions. In particular, the 2-fluoro-benzenesulfonamide group of compound L38 occupied the hydrophobic pocket formed by the Thr106 gatekeeper residue. Interestingly, in the 7Z9T structure, chain B (Figure 4-5C, right) we observed ligand still within the same location but adopting a in a different orientation, closer to the Asp168 and Arg67, but still within the same location, suggesting a potential dual binding mode. However, the ligand's signal was also weak in this case, suggesting a lower affinity of higher mobility in this orientation in comparison to the orientation presented in chain A. Also, in this structure, a coordinated ATP $\gamma$ S is visibly interacting with Gln60 and His64. The above data suggest that NC-p38i can bind to

the ATP pocket of p38 $\alpha$  in a conformation-dependent manner. In the presence of a potent ATP-competitive ligand such as SB203580, which binds with nanomolar affinity, NC-p38i can also bind to p38 $\alpha$  together with the ATP-competitor.

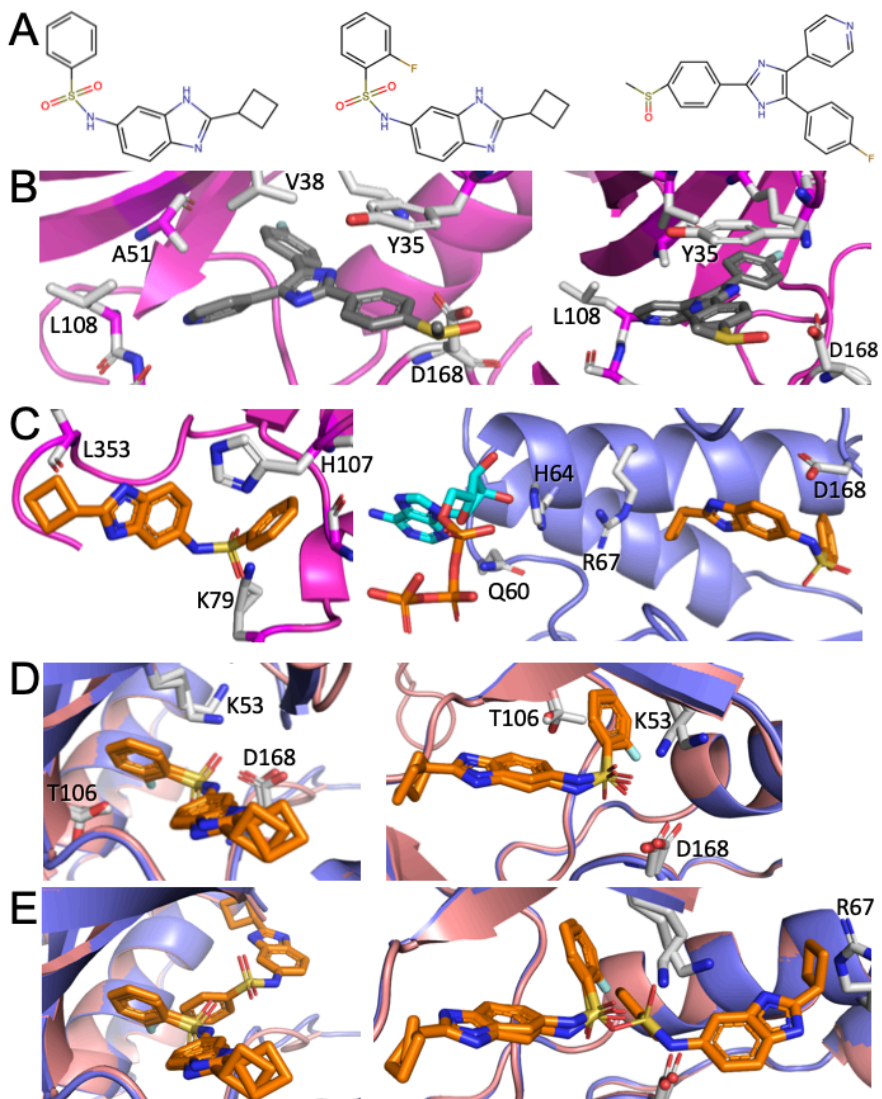


Figure 4-5 A: structural formulas of the described compounds. L37 on the left, L38 in the middle and canonical inhibitor SB on the right. B: SB in the active centre of the 7Z6I. C left: L37 in the hinge of 7Z6I. C right: 7Z9T chain B presenting L37 and ATP $\gamma$ S. D: superposition of the chains A of the 7Z9T and 7PVU containing L37 and L38 respectively, showing virtually the same position. E: as D, but also 7Z9T chain B ligand overlaid on top, showing presumed alternate orientation possible within the binding pocket. The occupancy of this ligand is zero.

## Results and Discussion

To ensure the proper positioning of the compounds, I used polder maps to enhance the weaker intensity signals to detect the orientation of L37, (Figure 4-6A). and using this approach, I could build the loop 170-182 (not visible in other structures), which also confirmed that the signals assigned to the ligand do not belong to any amino acid. In the 7PVU structure (ligand L38), the final electron density in chain A presents the compound, but I still used a polder map during refinement to facilitate the definition of its orientation (Figure 4-6B). In this case, no compound was identified in chain B.

7Z9T chain A exhibited a verifiable ligand L37 signal (Figure 4-6C, left) and solvent flattening was used to link the densities for its unequivocal assignment (Figure 4-6C, right). Unfortunately, its chain B possessed only a fragmentary map, but with the aid of solvent flattening I rebuilt enough density to orient the compound (Figure 4-6D). Because of that, I set its occupancy to zero. In the case of the ATP $\gamma$ S polder map unequivocally showed its presence (Figure 4-6E).

In summary, we have determined three ligand-containing p38 kinase structures, which draw a clear picture of the ligand's positions and may prove beneficial for subsequent rounds of *in silico* ligand design and optimisation.



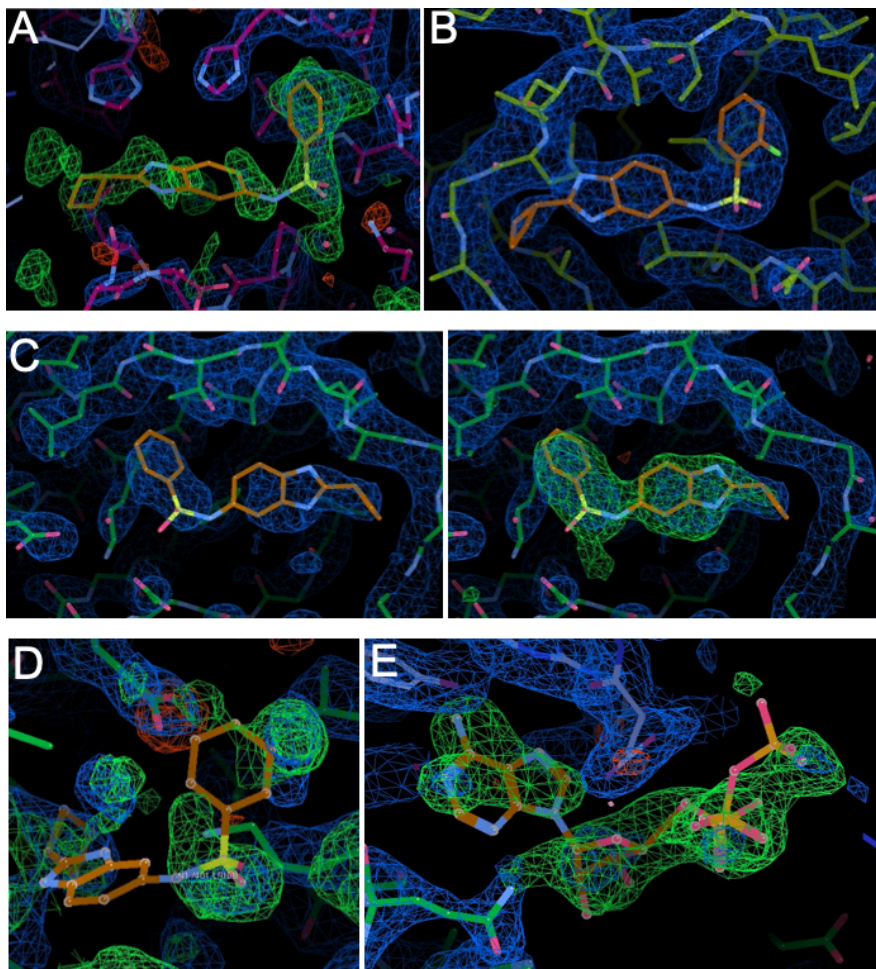


Figure 4-6 Electron density (blue) and polder (green) maps of the ligands (orange). A: Structure 7Z6l, where polder maps were used to ensure proper orientation of L37. B: 7PVU chain A – ligand L38 signal was strong enough for its unequivocal placement, but I used solvent flattening during structure refinement. C: 7Z6l chain A, where I used polder map to link densities ensuring proper ligand L37 orientation. D: the same structure, chain B, where the second L37 was oriented thanks to the polder maps, but due to the interrupted electron density its occupancy was set to zero. E: polder map clearly showing the location of coordinated ATP $\gamma$ S.

## 4.3 Smad-RNA complexes

Proteins were prepared by Lidia Ruiz and Dr Eric Aragón or helped me to purify these proteins. The RNAs we designed with Dr Radosław Pluta, Dr Marc Malfois and especially Dr Cristián Huck Iriart (Alba NCD-SWEET Beamline) helped with SAXS analysis.

### 4.3.1 Determination of binding characteristics

Initially, I performed a series of EMSA experiments to assess the Smad-RNA complex formation. From the initial trial of Smad MH1 domains, we selected the SMAD2 $\beta$  isoform and Smad3 (described in (Aragón et al., 2019), S2-E3) because they gave the most stable and reproducible complexes with RNAs. Of all the designed RNAs (presented in Figure 3-2), we selected those forming a stable hairpin for further EMSA experiments (Subsequent page: Figure 4-7). Constructs top1 and mid4 were also included, as consistently forming stable hairpin-dsRNA equilibrium.

Molecules designed around the putative Smad binding site (Subsequent page: Figure 4-7A) are all bound by the protein, although a preference for shorter constructs can be observed. Interestingly, for a heterogeneous population of mid4 a binding to both hairpin and dsRNA can be observed with no clear preference.

Second group, presenting the miR-21 apical loop (Subsequent page: Figure 4-7B), barely exhibits complex formation at 1:1 protein:RNA ratios, but analogically to the first group 1:2 complexes are fully formed. Interestingly, top1 presents strong binding, but smear suggests the instability of the resulting complex.

Of the third group of constructs (based on (Dickey and Pyle, 2017)) only one RNA was folding stably (int4, Subsequent page: Figure 4-7C), indicating binding of two proteins, corresponding to two structural features predicted.

Finally, constructs art1 and art2, members of the last group of artificial designs (Subsequent page: Figure 4-7D), do not present the “intermediary” binding of 1:1 stoichiometry, instead both of their structural features are saturated with proteins with equal speed and preference. I can speculate that once the first protein is bound, the movement speed of the complex drops and for the protein it is energetically more favourable to saturate the second binding site rather than to bind another, more mobile RNA molecule. Distribution of the hydration shell may be also altered, as no sterical

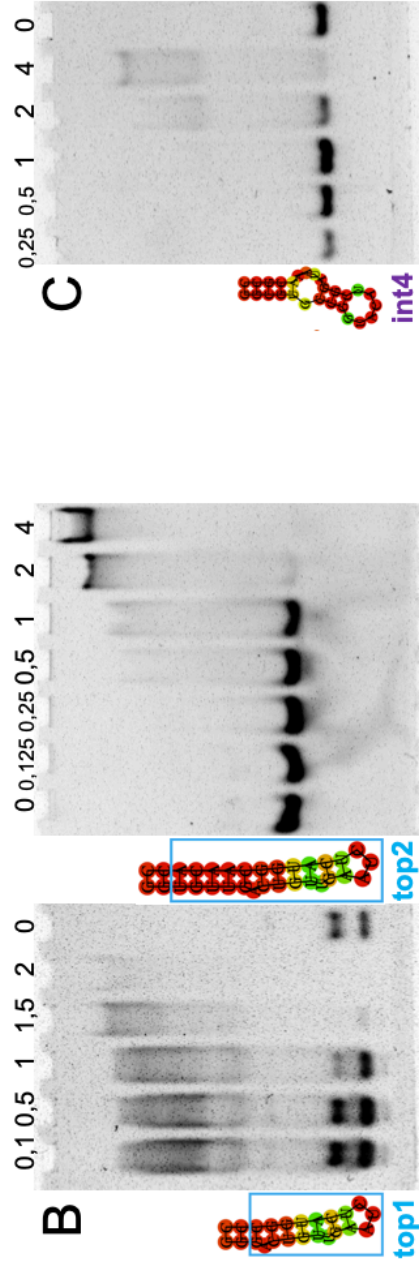
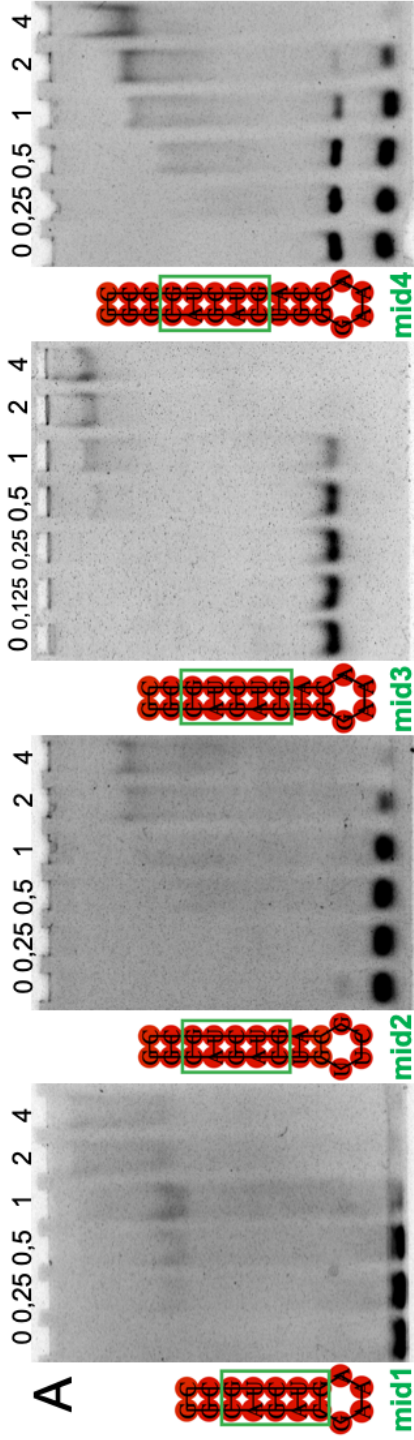
rearrangements should be needed for protein's physical access to any of the sites. Art2 derivatives (art3 and art5), prove that the relaxation of the stability of the asymmetric bulge-apical loop had a negative effect on the protein binding. Finally, with the removal of the art2's asymmetry of its bulge (construct art4), the preference for one of the binding sites can again be observed (statistically the apical loop is expected to be bound first).

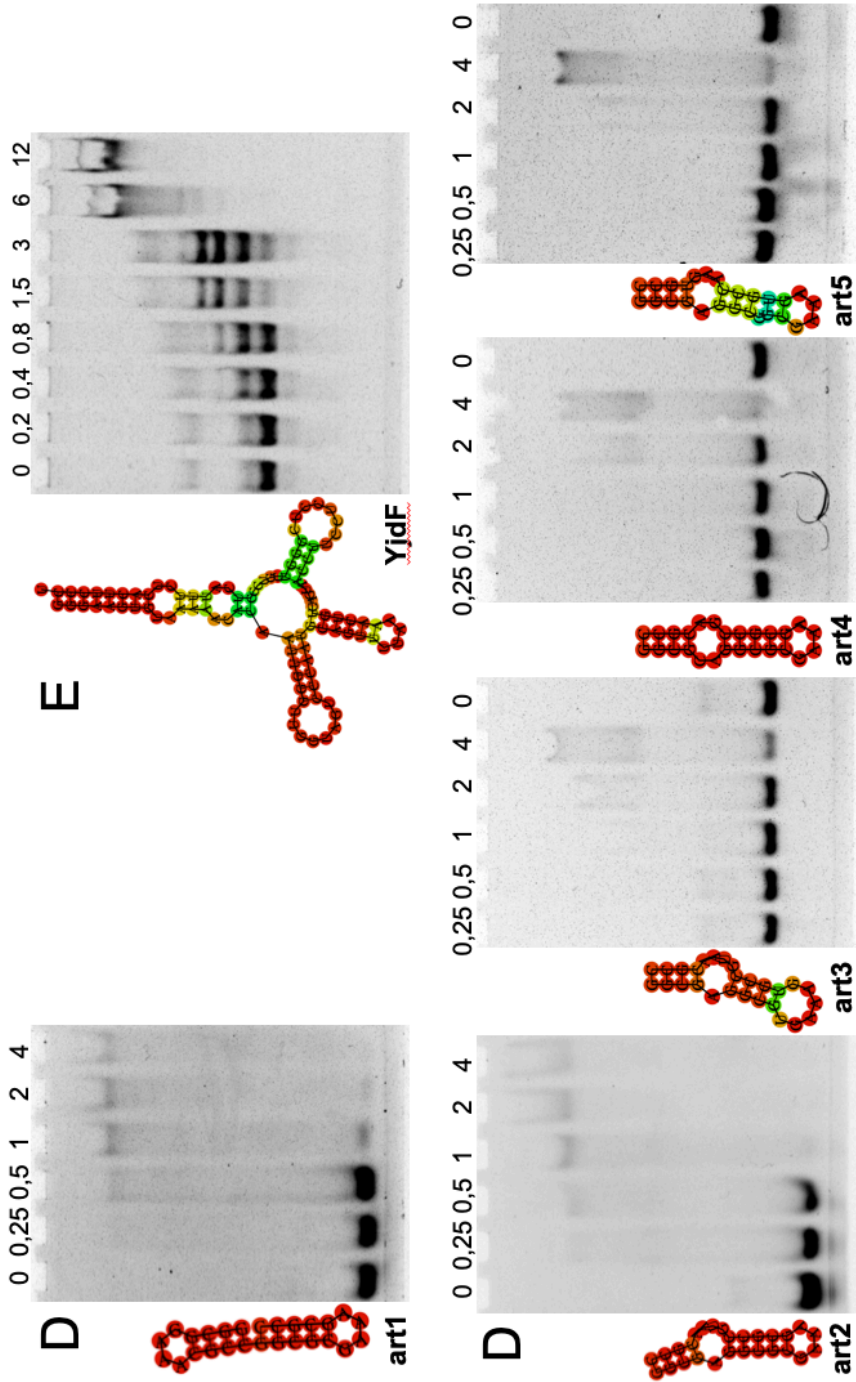
Moreover, a comparison of the results of the EMSA experiments of the int4 and art2 RNAs, sharing fold, presents virtually the same outcome despite their varying sequences.

In all, all of the presented hairpins exhibit strong binding by the S2-E3, with a preferred 2 protein to 1 RNA stoichiometry and no discernible impact of their sequence on that process. Furthermore, constructs with a singular structural feature (a loop or bulge) exhibit a single binding preference, and those of multiple reflect that number or possible binding combinations (especially visible in the case of YjdF riboswitch aptamer, possessing three loops and a stem; Subsequent page: Figure 4-7E). The visible exception is the art5 construct, but the apical loop was probably in an equilibrium state between one big loop or two smaller ones, as the stabilizing base-paired region was impaired in this construct.

Subsequent page: Figure 4-7. Electrophoretic Mobility Shift Assays of the selected constructs are presented in Figure 3-2. The amount of RNA is constant, the numbers above images indicate the protein molar ratio (X protein:1 RNA).

Groups A and B are based on miR-21 (Davis et al., 2010), group C on the construct 17E derived from group IIC intron (Dickey and Pyle, 2017), group D composed of artificially created constructs, E marks YjdF bacterial riboswitch.





### 4.3.2 Complex characterisation

To extend the information about the complex formation, and to discern between the 1:1 and 1:2 complex stoichiometry preferences, I performed a gel filtration experiment. Art2 and S3 were selected based on the stability and repeatability of their interactions (Figure 4-8). Assuming lack of the preference of Smad-RNA complex formation, for both 1:1 and 1:2 samples a single elution peak would be expected (with minimal or no unbound fractions detectable). However, this is the case only for 1:2 complex, and for the ~1:1 ratio, two distinct peaks can be observed, clearly indicating the possibility of binding of two MH1s for every single RNA hairpin. Furthermore, I observed that most of the protein:RNA samples started to precipitate (sample turned milky, but not flaking) briefly after mixing, if not mixed in 2:1 p:R ratio (isolated components were stable).

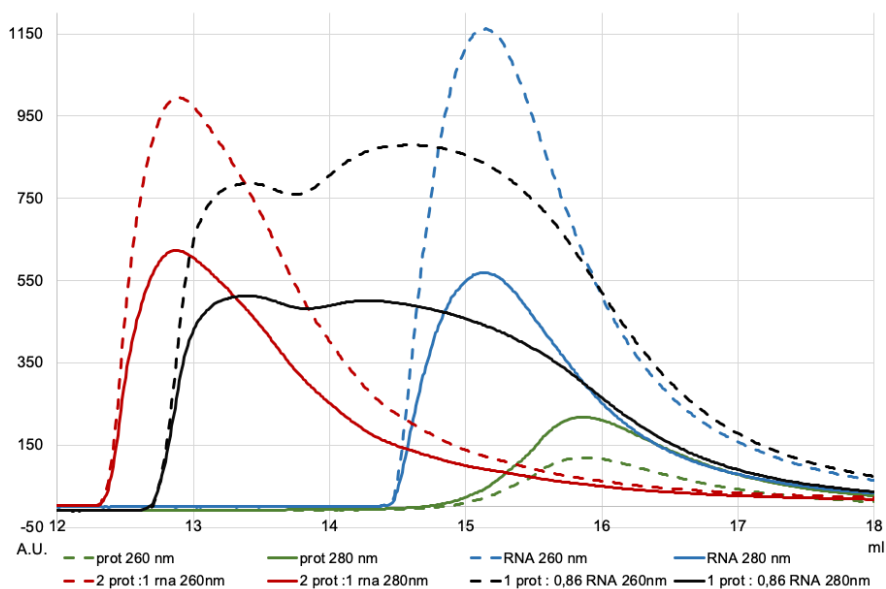


Figure 4-8 Gel filtration experiment of protein (Green), RNA (blue) and their 2:1 (red) and 1:0,86 (black, partially precipitated sample estimation) complexes.

With this data I attempted a complex reconstruction, first via the bioinformatics analysis of S2-E3:art2, using the NPDock webserver. The top scoring (lowest energy) models of three clusters of simulation trajectories (Figure 4-9A) show three predicted different possible Smad-RNA assemblies, mediated analogously as in the case of the protein's canonical DNA binding site (interacting loop coloured in black). Said protein-DNA structure (PDB ID:6H3R) is also shown as a reference.

Furthermore, the superposition of two *in silico* models presents the possibility of binding two proteins on the same RNA hairpin (Figure 4-9B).

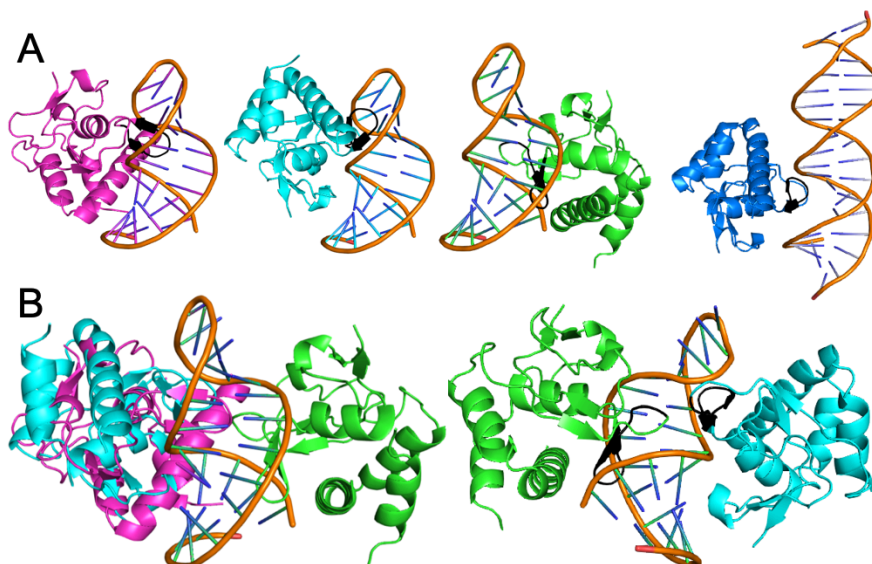


Figure 4-9 A: NPdock predictions, representatives of top 3 clusters of S2-E3:art2 models. Last image: structure of S2-E3 binding major groove of its canonical DNA (PDB 6H3R). B left: superposition of top clusters of *in silico* predictions. B right: two of the representatives suggested a model of 2 protein:1 RNA complex.

To validate the *in-silico* predictions, I performed SAXS measurements for the same protein and RNA, both as separate samples and as a mixed complex, in the series of dilutions. For the RNA and protein alone, the data was acceptable and did not exhibit concentration dependence. Unfortunately, the data of the p:R complex exhibited a high level of noise, which made buffer subtraction unreliable (Figure 4-10 top left). Further data analysis revealed that the protein is very sensitive to radiation damage and prone to aggregation, as was the complex. As such, these results should be considered preliminary.

The  $P(r)$  distribution (Figure 4-10, top middle), as expected presents elongation of the RNA molecule and sphericity of the protein. It also shows sphericity of the complex, suggesting the protein binding at the sides, analogically to the DNA (see PDB ID: 6H3R), rather than at the “ends” of the RNA molecule (as in this case, the complex would appear highly elongate). Kratky plot (Figure 4-10, top right) shows the flexibility of the complex and relative compactness of the protein.

## Results and Discussion

The weight analysis of the obtained data (Figure 4-10, lower half) is pointing toward the 1:2 RNA:protein ratio. Also, the Porod volume, Molecular weight and volume of correlation are close to the expected values, given the quality of the data. However, this is under the assumption that the mass of the protein is that of a dimer, which could have been due to a different protein production batch or radiation-induced sample changes. Size and Shape analysis deviates significantly from other complex molecular mass estimates, but this may be because software assumes a spherical object, as we can see the highest estimation error with pure RNA sample.

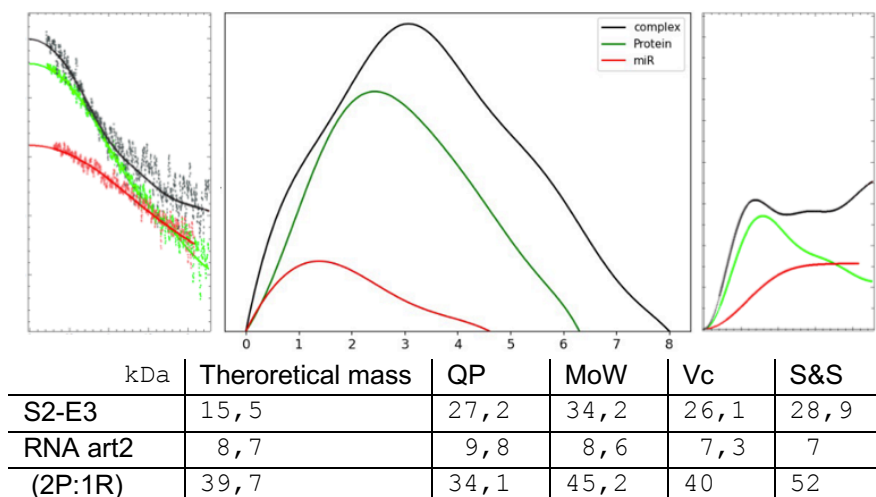


Figure 4-10 SAXS data analysis in collaboration with Dr Cristián Huck Iriart. RNA in red, protein in green, complex in black. Top row left: experimental data. Middle:  $P(r)$  distribution function showing elongation for RNA sample and sphericity for the remaining two. Right: Kratky plot showing the flexibility of the protein and RNA (partially unfolded) and complex as being highly flexible. Bottom row: – mass estimation analysis. Porod volume, molecular weight and volume of correlation analysis point towards 2 protein:1 RNA complex stoichiometry (assuming protein mass is that of a dimer), with shape and size deviating from the expected weight.

To further investigate the complex geometry, I attempted to compare manually created dummy S2-E3:5gcB complexes to the SAXS experimental data. Complex models and the respective  $\chi^2$  test score estimate of expected experimental data are presented in Figure 4-11.



Presented models were sorted according to their increasing  $\chi^2$  test scores. Interestingly, the model based on the NPDock prediction did not score the highest, instead, the more elongated, spheroid shape is expected to be a closer representation of the in-solution complex. It was also verified that the data does not show a single protein bound to the RNA. However, SAXS data analysis revealed the high flexibility of the complex, suggesting dynamics within the system.

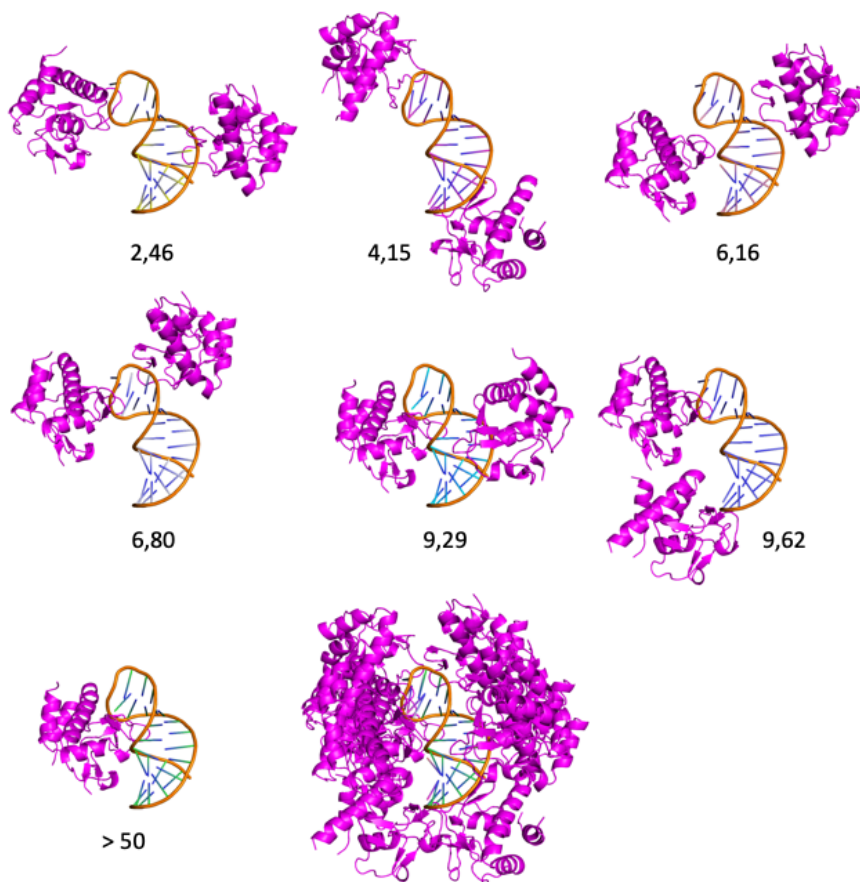


Figure 4-11 Manually created S2-E3:5gcB complex models and the respective  $\chi^2$  test values of their expected SAXS scattering curves compared to obtained experimental data, sorted increasingly. Centre: analogue of the 2:1 NPDock prediction ( $\chi^2$  of 9,29), and one of its clusters representatives as penultimate. The ultimate figure superimposes all investigated areas of the hairpin.

## Results and Discussion

Presented SAXS data should be treated as preliminary, as it exhibits aggregation and elevated noise level characteristics. Nevertheless, it allows for initial characterisation of the structure of the pp:R complex. Furthermore, care needs to be taken when analysing data as most of the bioSAXS software is written with proteins in mind, with phosphorus atoms providing higher contrast to solvent than carbon and nitrogen.

Concluding, I have investigated the Smad-RNA binding mechanism, and have proven that for the RNA hairpin model, both artificial and a fragment of natural molecules, S2-E3 MH1 domain does not exhibit any sequence-binding preference. The complex formation follows predicted structural features, apical loops (GNRA as well as WT ones) and internal kinks and bulges. Furthermore, I have also characterised the binding ratio of 2 proteins for each RNA molecule, which is analogical to the Smad-DNA binding (PDB 6TCE, 5NM9).

Given how the RNA molecules are crucial for cellular processes, coupled with how abundant, yet tightly controlled are Smad proteins, these results might help implement further experiments to clarify how transcription factors-RNA binding takes place and the role of long noncoding RNAs (lncRNAs).

## 4.4 Human NEDD4-like C2 domains

During this project I have obtained 5 structures of the C2 domain of the NEDD4-like family of E3 ubiquitin ligases:

WT1, of the wild type (initial molecular replacement by Dr Radoslaw Pluta)

WT2, of the wild type, crystallized in presence of phosphoserine ligand

Mut1, F91I mutant

Mut2, F91I mutant with phosphoserine ligand

Mut3, F91I mutant with inosito-3-phosphate ligand.

Structures are in final refinement and deposition process, and as such should be treated as preliminary, and the accompanying publication is in preparation (September 2021). Structure 2NSQ (Walker et al., 2006) of the wild-type sequence (apo form) is presented in this work solely for comparison purposes.

### 4.4.1 New structure of the WT N4L C2 domain

The entire family of E3 HECT ubiquitin ligases consist of 9 members, with one of them being the neural precursor cell expressed developmentally downregulated gene 4-like (NEDD4L or NEDD4-2). The role of this protein is as a final, third step of the ubiquitination process, targeting proteins for degradation. Proven binders include Smad2,3,7 (Kuratomi et al., 2005) and 4 (Morén et al., 2005), making N4L the regulator of TGF- $\beta$  and BMP signalling pathways.

The NEDD4L protein consists of an N-terminal C2 domain (binding  $\text{Ca}^{2+}$  using it to mediate interactions with phospholipids), 4 WW domains (protein-protein interaction – substrate and regulatory protein binding) and the C-terminal HECT domain (ubiquitin ligation). C2 domain has approximately 150 amino acids, folded into two sets of four antiparallel  $\beta$ -sheets, joined with flexible linking loops. The interaction centre is located “underneath” this  $\beta$ -sheets arrangement, opposite the N-terminal start of the chain.

The structure of the wild-type NEDD-4L C2 domain has been previously determined (PDB ID: 2NSQ). However, our structure besides marginally better resolution (1,75 Å vs 1,85 Å) presents more visible amino acids and differences in some of their orientations, thereby extending the available information that might contribute to improving the present knowledge of how this domain folds and interacts with ligands and cofactors.

## Results and Discussion

In our structure WT1 (Figure 4-12A) there are trace signals corresponding to (presumably) first amino acids Ala2-Gly6 of the protein, however, despite the series of polder maps I tried, I was not able to place the residues unequivocally. Based on the local environment, we believe is that some signals are visible because they were partially immobilised, interacting with the crystal symmetry mates, rather than the inherent structural property of the protein itself.

Residues coordinating the  $\text{Ca}^{2+}$  ion (Figure 4-12B) are well defined in our case, although in the 2NSQ structure, the authors placed water in its place. The first loop, Asp36-Gly39, is traceable in our structure WT1, but not in 2NSQ (Figure 4-12C). This loop contains the Asp36, which together with Ala40 coordinate the calcium ion. Across the calcium-binding site, on the other side of the ion, loop Glu96-Leu104 presents a slight shift of approximately 1-2 Å, bringing it closer to the central ion, and containing the Asp95-Glu96-Asn97 directly interacting with the  $\text{Ca}^{2+}$ . On the opposite end of the protein, two apical loops Asp119-Thr121 (+ the Tyr126) and Arg136-Arg141 present a slightly different than 2NSQ conformation, but this flexibility can be expected in such a region.

As previously described in (Plant et al., 1997), the C2 domain binds phosphate-containing lipids, notably phosphatidylserine and phosphatidylinositol. I undertook attempts of co-crystallization of this protein, however, to avoid possible solubility problems I used phosphoserine (PS) and inositol-3-phosphate (IP3), as they lack the hydrocarbon fatty acid group. Unfortunately, I did not succeed in obtaining WT structure with a visible ligand signal. However, structure WT2, crystallized in the presence of phosphoserine (Figure 4-12A and B), shows a significant difference within the  $\text{Ca}^{2+}$  binding region in the differing orientation of the Asp34-Ser41 loop, which is oriented closer to the  $\text{Ca}^{2+}$  site (Figure 4-12D). I suspect this is an influence of the ligand, as it would be difficult to explain this movement with the influence of symmetry mates or another coordinated ion in this region of the crystal. Biochemically, the only difference in the sample preparation between the WT1 and WT2 structures is the presence of the PS ligand. Other structural differences with WT1 include the shifts of more flexible parts of the protein, notably residues Gly4-Pro8, Ser13-Ser19, or Tyr126-Thr127. Further crystallography trials of this sample are ongoing.

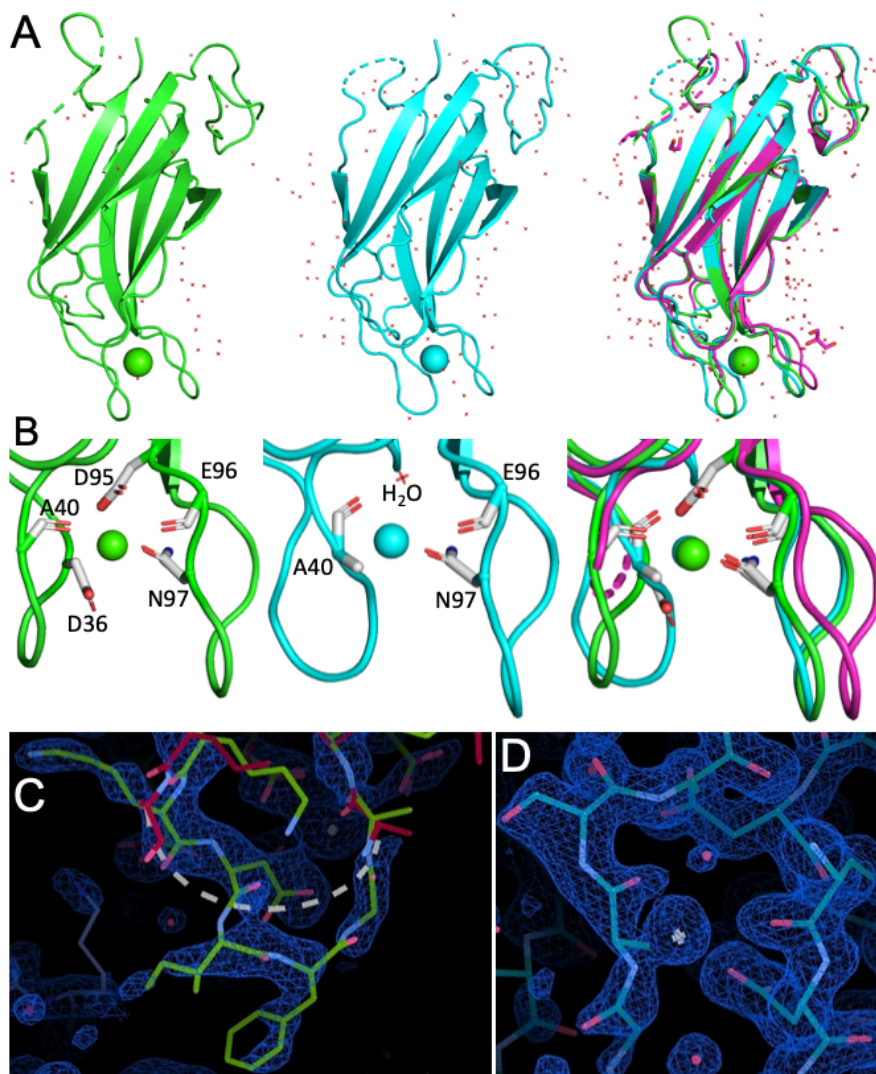


Figure 4-12 Structures of the wild-type NEDD4-like C2 domains. WT1 in green, WT2 in cyan and 2NSQ (Walker et al., 2006) in magenta. A, left and centre: asymmetric units. Right: superposition of all three. B: Coordination of the central calcium ion. Right: superposition showing a difference in loop orientations. C: electron density map of loop Asp36-Gly39, not present in 2NSQ. D: Ca<sup>2+</sup> in proximity to Ala40.

### 4.4.2 F91I structure, apo and ligands-bound

Various mutations in the N4L protein are reported (Stouffs et al., 2020) or (Pecimonova et al., 2021) as causes of periventricular nodular heterotopia. F91I is a point mutation located within the C2 domain's core (Figure 4-13A) and can be described as deletion of two atoms from the benzene ring, leaving it open.

Structurally (besides the mutation itself), the most significant difference between the mutant and WT protein lies within the calcium-binding centre. In all three of the mutant structures, both loops surrounding the ion (Lys35-Ser41 and Asp95-Asp103) are further apart than the wild types, with the structure WT1 having the Phe38-Leu99 distance of 9,4 Å in comparison to the 14,5 Å of the mutant (Figure 4-13B). Further differences include the presence of the second calcium ion, in the vicinity of Asn97, Thr100 and the Asp102, which corresponds with molecular docking predictions (Guillén et al., 2013). Finally, despite the high resolution of this structure (at 1 Å only the density of the main C $\alpha$  chain is visible), for the Lys35-Ala40 region.

Structure Mut2, contain the phosphoserine ligand in the calcium-binding site (Figure 4-13C, left), with its phosphate part coordinated by the Ca<sup>2+</sup>Asp42, and the Glu96 within the Asp95-Asp103 loop. The seine element of the ligand interacts with the nitrogen atoms of the Arg98, Leu99 and Thr100, as well as its terminal oxygen atom. I used polder map during the refinement steps, which allowed for its certain placement despite its electron density not being continuous (Figure 4-13E, left). Interestingly, the coverage of the entire protein (and amino acids side chains) with its electron density map is even fuller than in the case of Mut1, despite the slightly lower overall resolution.

The fifth structure of this project, Mut3, contains the inosito-1,3,4-triphosphate bound outside of the calcium loops, next to the residues Leu115, Met150 and coordinated waters (Figure 4-13C, right). Interestingly, it is not located in the vicinity of the calcium site (the positions of both ligands relative to the protein are presented in Figure 4-13D). Similarly to before, I used polder map to confirm the ligand's presence and orientation (Figure 4-13E, right).

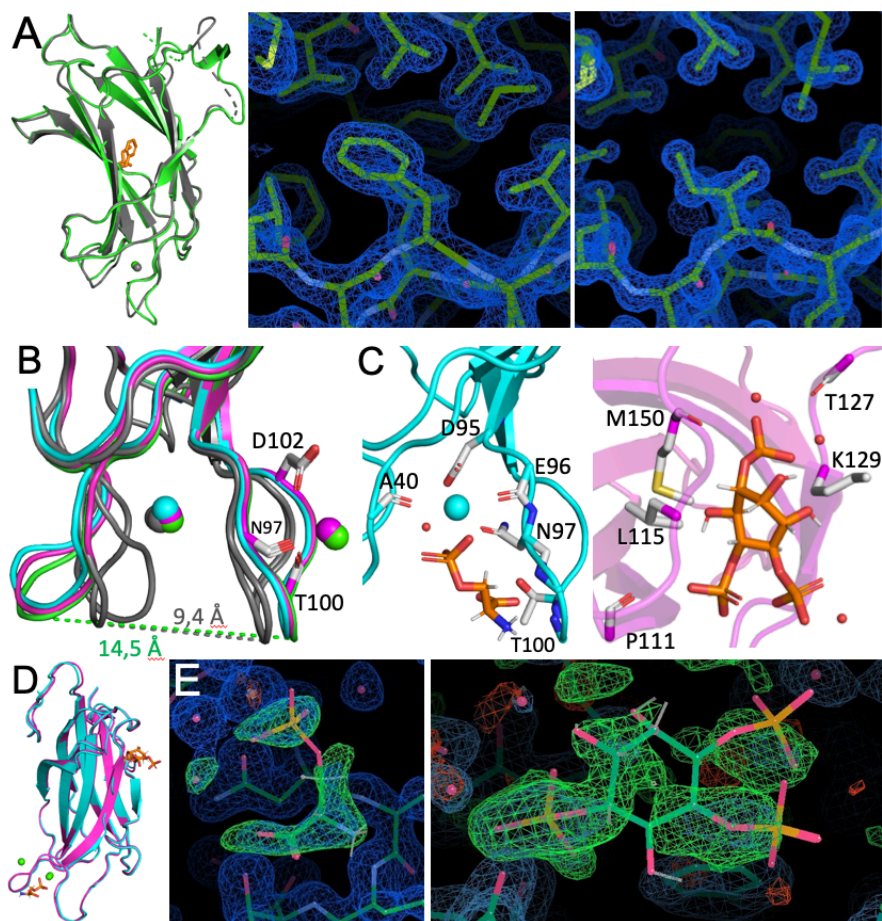


Figure 4-13 F91I mutant structures: apo Mut1 (green), Mut2 containing phosphoserine ligand (cyan), and Mut3 containing inosito-1,3,4-triphosphate ligand (magenta). A left: alignment of N4L C2 WT1 (grey) and F91I (Mut1) presenting the location of the mutation. Middle and right: electron density maps of the 91st amino acid of these structures, respectively. B: active centres of three mutant structures compared to the WT apo form. C: Phosphoserine and I3P ligands (orange) and their respective interactions. D: location of ligands (orange) relative to the proteins. E: polder maps extend electron density information and confirm the positions of the ligands.

## Results and Discussion

In conclusion, these five structures of both wild type and mutants, with two different ligands presented in this chapter expand the available structural knowledge about the interactions of these domains. They illustrate how the differences between wild type and mutant can correlate with the functional defects observed in diseases associated with these mutations.

We expected IP3 to also be found in the phosphoserine location, in the vicinity of the  $\text{Ca}^{2+}$  interaction site. Such predictions have already been made for the Smurf1 C2 domain (Scott et al., 2020), where molecular docking within the active centre predicted two possible conformations (K28, R32, K85, K88 and K89 correspond to the  $\text{Ca}^{2+}$  binding site). However, a molecular docking experiment predicted ligand's binding corresponding to the presented here "at the side" of the C2A domain of rabphilin 3A (Guillén et al., 2013). Of note are both  $\text{Ca}^{2+}$  ions in their simulations (as in presented Mut1 and 3), described as "*Ca<sup>2+</sup> increases the affinity of the C2A and C2B domains of rabphilin 3A to bind PI(4,5)P2-containing membranes*".

Furthermore, one of the biochemical assays performed in parallel with the crystallography trials of the point mutation variant of the C2 domain was the thermofluorometric stability assay (performed by L. Ruiz and L. Villarreal Fernández). Presented in Figure 4-14, it revealed the lower stability of the mutant with respect to the WT. This difference is also observed in the presence or absence of additional calcium, from 48,2°C to 46°C with no additional ions, from 53°C to 50,8°C with 10 mM  $\text{Ca}^{2+}$  and from 54°C to 52,2°C with 20 mM  $\text{Ca}^{2+}$ . Although the overall stability difference is barely 1~3°C, it can be considered significant.

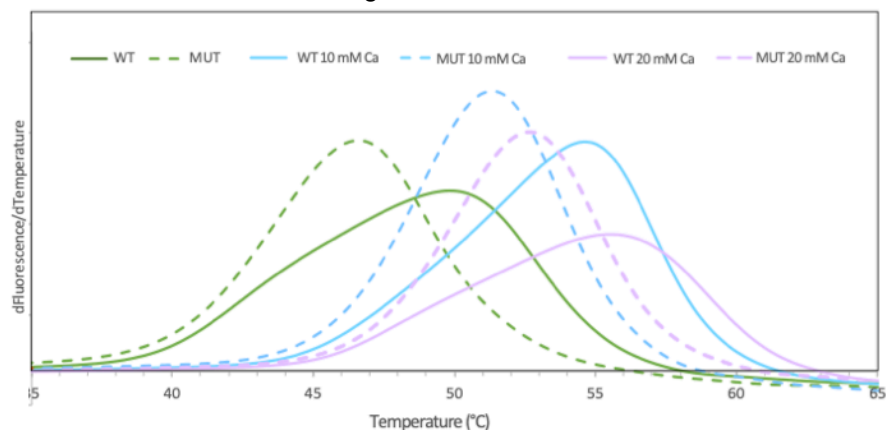


Figure 4-14 Thermofluorometric analysis of the N4L C2 WT and F91 mutant, presenting the lowering of the protein's stability both with and without extra  $\text{Ca}^{2+}$  supplementation. A modified Figure 4.3. from the Lidia Villarreal Fernández MSc thesis (Universitat de Barcelona Master's degree in Biomedicine, academic year 2020-2021).





## 5. Conclusions

This thesis work presents four different chapters dedicated to studying four biologically relevant questions using X-ray crystallography. The projects have in common the search for new targets for drug design as well as characterising the effects of disease associated point mutations in the protein structures.

### 1. *D. melanogaster* germline-specific protein Deadhead

The new Trxs structures revealed two main differences of these specific germline proteins with respect to other previously characterised. Whereas the Trx surfaces, including TrxT, are negatively charged, Dhd has extended positively charged patches. By displaying a positively charged distribution (and not negative as in most Trxs), Dhd proteins would speed up the selection of specific redox targets during initial encounter complexes in vivo. Among these redox reactions are the reduction of intermolecular disulfide bonds in *Drosophila* protamine oligomers to facilitate their eviction from DNA.

Our results may have applications for the design of inhibitory molecules to reduce and control fly plagues by selecting the germline Trx proteins as targets. These plagues, such as those of black fly species that spread diseases such as river blindness in Africa and the Americas (World Health Organization), have an impact on human health. Others negatively affect the economies of many countries worldwide due to losses in fruit and vegetable production. In this context, the charge distribution of Dhd should drive the selection of molecular binders which differ from those preferentially selected by Trx-2 (present in many insects) and TrxT counterparts.

### 2. Novel, non-canonical p38 $\alpha$ ligands

p38 $\alpha$  is a versatile protein kinase that can control numerous processes that play an important role in cellular responses to stress. Dysregulation of p38 $\alpha$  signalling has been linked to several diseases, including inflammation, immune disorders, and cancer. These effects suggest that targeting p38 $\alpha$  could be therapeutically beneficial in treating some serious diseases. We have described a new class of compounds that effectively block p38 $\alpha$  autophosphorylation but have little effect on the activity of the canonical pathway. X-ray crystallography studies indicate that these molecules can

interact with p38 $\alpha$  in different regions in and around the active site. Our collaborators are testing these new p38 $\alpha$  inhibitors to reduce cardiomyocyte death in cellular models of ischemia-reperfusion.

### **3. Characterisation of Smad-RNA complexes**

In this project I used two biochemical and one computational techniques to characterise Smad binding to RNA. Although the project will require further development in the future, it defines the initial steps to help understand how Smad-RNA interactions occur and their possible consequences in human health and disease.

### **4. Novel structures of NEDD4-L mutant and its ligands**

Switching from a non-active to an active conformation in response to intracellular changes such as an increase in cytoplasmic calcium levels is an important and common mechanism of regulation among the E3 ligases of the Nedd4L family.

A nonsense point mutation in the C2 domain has been reported to cause a neurological disease called periventricular nodular heterotopia 7, but so far no interpretation of the mutation has been made, either from a structural or functional point of view. After determining these structures, we have detected allosteric effects in the region involved in the coordination properties of the metal, but not at the site of the mutation. I have also validated the available *in silico* predictions by obtaining structures with ligands. Based on these and previous results from our group, we suggest that high levels of cytoplasmic calcium may not only act as a switch in the conformation of Nedd4L ligase by modulating its activation and inactivation through the interaction of calcium with the C2 domain, but may also act as a chemical chaperone and stabilize the folded state by increasing the melting temperature its denaturation point.

## Bibliography

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* 68, 352–367. <https://doi.org/10.1107/S0907444912001308>.

Aghakhani, Y., Kinay, D., Gotman, J., Soualmi, L., Andermann, F., Olivier, A., and Dubeau, F. (2005). The role of periventricular nodular heterotopia in epileptogenesis. *Brain* 128, 641–651. <https://doi.org/10.1093/brain/awh388>.

Alarcón, C., Zaromytidou, A.-I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., et al. (2009). Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* 139, 757–769. <https://doi.org/10.1016/j.cell.2009.09.035>.

Aragón, E., Goerner, N., Zaromytidou, A.-I., Xi, Q., Escobedo, A., Massagué, J., and Macias, M.J. (2011). A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev.* 25, 1275–1288. <https://doi.org/10.1101/gad.2060811>.

Aragón, E., Goerner, N., Xi, Q., Gomes, T., Gao, S., Massagué, J., and Macias, M.J. (2012). Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF-β Pathways. *Structure* 20, 1726–1736. <https://doi.org/10.1016/j.str.2012.07.014>.

Aragón, E., Wang, Q., Zou, Y., Morgani, S.M., Ruiz, L., Kaczmarek, Z., Su, J., Torner, C., Tian, L., Hu, J., et al. (2019). Structural basis for distinct roles of SMAD2 and SMAD3 in FOXH1 pioneer-directed TGF-β signaling. *Genes Dev.* 33, 1506–1524. <https://doi.org/10.1101/gad.330837.119>.

Beckert, B., and Masquida, B. (2011). Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.* 703, 29–41. [https://doi.org/10.1007/978-1-59745-248-9\\_3](https://doi.org/10.1007/978-1-59745-248-9_3).

Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2006). Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.* 1, 3. <https://doi.org/10.1186/1748-7188-1-3>.

Bijelic, A., and Rempel, A. (2018). Polyoxometalates: more than a phasing tool in protein crystallography. *ChemTexts* 4, 10. <https://doi.org/10.1007/s40828-018-0064-1>.

Boniecki, M.J., Lach, G., Dawson, W.K., Tomala, K., Lukasz, P., Soltysinski, T., Rother, K.M., and Bujnicki, J.M. (2016). SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* 44, e63. <https://doi.org/10.1093/nar/gkv1479>.

- Canovas, B., and Nebreda, A.R. (2021). Diversity and versatility of p38 kinase signalling in health and disease. *Nat. Rev. Mol. Cell Biol.* 22, 346–366. <https://doi.org/10.1038/s41580-020-00322-w>.
- Chen, C., and Matesic, L.E. (2007). The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer Metastasis Rev.* 26, 587–604. <https://doi.org/10.1007/s10555-007-9091-x>.
- Chen, Y., and Pollack, L. (2016). SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip. Rev. RNA* 7, 512–526. <https://doi.org/10.1002/wrna.1349>.
- Collet, J.-F., and Messens, J. (2010). Structure, function, and mechanism of thioredoxin proteins. *Antioxid. Redox Signal.* 13, 1205–1216. <https://doi.org/10.1089/ars.2010.3114>.
- Comess, K.M., Sun, C., Abad-Zapatero, C., Goedken, E.R., Gum, R.J., Borhani, D.W., Argiriadi, M., Groebe, D.R., Jia, Y., Clampitt, J.E., et al. (2011). Discovery and characterization of non-ATP site inhibitors of the mitogen activated protein (MAP) kinases. *ACS Chem. Biol.* 6, 234–244. <https://doi.org/10.1021/cb1002619>.
- Conway, J.A., Kinsman, G., and Kramer, E.R. (2022). The Role of NEDD4 E3 Ubiquitin-Protein Ligases in Parkinson's Disease. *Genes (Basel)* 13. <https://doi.org/10.3390/genes13030513>.
- Cowtan, K. (2001). Phase Problem in X-ray Crystallography, and Its Solution. In ELS, John Wiley & Sons, Ltd, ed. (Wiley), p.
- Davis, B.N., Hilyard, A.C., Nguyen, P.H., Lagna, G., and Hata, A. (2010). Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Mol. Cell* 39, 373–384. <https://doi.org/10.1016/j.molcel.2010.07.011>.
- Dessau, M.A., and Modis, Y. (2011). Protein crystallization for X-ray crystallography. *J. Vis. Exp.* <https://doi.org/10.3791/2285>.
- De Nicola, G.F., Bassi, R., Nichols, C., Fernandez-Caggiano, M., Golforoush, P.A., Thapa, D., Anderson, R., Martin, E.D., Verma, S., Kleinjung, J., et al. (2018). The TAB1-p38 $\alpha$  complex aggravates myocardial injury and can be targeted by small molecules. *JCI Insight* .
- Dickey, T.H., and Pyle, A.M. (2017). The SMAD3 transcription factor binds complex RNA structures with high affinity. *Nucleic Acids Res.* 45, 11980–11988. <https://doi.org/10.1093/nar/gkx846>.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. <https://doi.org/10.1107/S0907444910007493>.
- Enslow, B.T., Stockand, J.D., and Berman, J.M. (2019). Liddle's syndrome mechanisms, diagnosis and management. *Integr. Blood Press. Control* 12, 13–22. <https://doi.org/10.2147/IBPC.S188869>.

Escobedo, A., Gomes, T., Aragón, E., Martín-Malpartida, P., Ruiz, L., and Macias, M.J. (2014). Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L. *Structure* 22, 1446–1457. <https://doi.org/10.1016/j.str.2014.08.016>.

Freier, R., Aragón, E., Bagiński, B., Pluta, R., Martín-Malpartida, P., Ruiz, L., Condeminas, M., Gonzalez, C., and Macias, M.J. (2021). Structures of the germline-specific Deadhead and thioredoxin T proteins from *Drosophila melanogaster* reveal unique features among thioredoxins. *IUCrJ* 8, 281–294. <https://doi.org/10.1107/S2052252521000221>.

Gao, S., Alarcón, C., Sapkota, G., Rahman, S., Chen, P.-Y., Goerner, N., Macias, M.J., Erdjument-Bromage, H., Tempst, P., and Massagué, J. (2009). Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol. Cell* 36, 457–468. <https://doi.org/10.1016/j.molcel.2009.09.043>.

Ge, B., Gram, H., Di Padova, F., Huang, B., New, L., Ulevitch, R.J., Luo, Y., and Han, J. (2002). MAPKK-independent activation of p38alpha mediated by TAB1-dependent autophosphorylation of p38alpha. *Science* 295, 1291–1294. <https://doi.org/10.1126/science.1067289>.

Gilbert, S.F. (2000). *Early Drosophila Development*. .

Gonzalez, C. (2013). *Drosophila melanogaster*: a model and a tool to investigate malignancy and identify new therapeutics. *Nat. Rev. Cancer* 13, 172–183. <https://doi.org/10.1038/nrc3461>.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* 36, W70-4. <https://doi.org/10.1093/nar/gkn188>.

Guillén, J., Ferrer-Orta, C., Buxaderas, M., Pérez-Sánchez, D., Guerrero-Valero, M., Luengo-Gil, G., Pous, J., Guerra, P., Gómez-Fernández, J.C., Verdaguer, N., et al. (2013). Structural insights into the Ca<sup>2+</sup> and PI(4,5)P<sub>2</sub> binding modes of the C2 domains of rabphilin 3A and synaptotagmin 1. *Proc Natl Acad Sci USA* 110, 20503–20508. <https://doi.org/10.1073/pnas.1316179110>.

Haller, V., Nahidino, P., Forster, M., and Laufer, S.A. (2020). An updated patent review of p38 MAP kinase inhibitors (2014-2019). *Expert Opin. Ther. Pat.* 30, 453–466. <https://doi.org/10.1080/13543776.2020.1749263>.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* 2, 1849–1861. <https://doi.org/10.1038/nprot.2007.249>.

Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319, 1059–1066. [https://doi.org/10.1016/S0022-2836\(02\)00308-X](https://doi.org/10.1016/S0022-2836(02)00308-X).

Jacques, D.A., and Trehwella, J. (2010). Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein Sci.* 19, 642–657. <https://doi.org/10.1002/pro.351>.

Jonk, L.J., Itoh, S., Heldin, C.H., ten Dijke, P., and Kruijer, W. (1998). Identification and functional characterization of a Smad binding element (SBE) in the JunB promoter that acts as a transforming growth factor-beta, activin, and bone morphogenetic protein-inducible enhancer. *J. Biol. Chem.* *273*, 21145–21152. <https://doi.org/10.1074/jbc.273.33.21145>.

Joosten, R.P., Long, F., Murshudov, G.N., and Perrakis, A. (2014). The PDB REDO server for macromolecular structure model optimization. *IUCrJ* *1*, 213–220. <https://doi.org/10.1107/S2052252514009324>.

Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S., and Benach, J. (2014). Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron Radiat.* *21*, 679–689. <https://doi.org/10.1107/S160057751400825X>.

Kao, C., Zheng, M., and Rüdiger, S. (1999). A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* *5*, 1268–1272. <https://doi.org/10.1017/s1355838299991033>.

Kieft, J.S., and Batey, R.T. (2004). A general method for rapid and nondenaturing purification of RNAs. *RNA* *10*, 988–995. <https://doi.org/10.1261/rna.7040604>.

Kikhney, A.G., and Svergun, D.I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* *589*, 2570–2577. <https://doi.org/10.1016/j.febslet.2015.08.027>.

Kuratomi, G., Komuro, A., Goto, K., Shinozaki, M., Miyazawa, K., Miyazono, K., and Imamura, T. (2005). NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF-beta (transforming growth factor-beta) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem. J.* *386*, 461–470. <https://doi.org/10.1042/BJ20040738>.

Kuwamoto, S., Akiyama, S., and Fujisawa, T. (2004). Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. *J. Synchrotron Radiat.* *11*, 462–468. <https://doi.org/10.1107/S0909049504019272>.

Lee, S.-J., Zhou, T., and Goldsmith, E.J. (2006). Crystallization of MAP kinases. *Methods* *40*, 224–233. <https://doi.org/10.1016/j.ymeth.2006.05.003>.

Lemaire, M., Froment, C., Boutros, R., Mondesert, O., Nebreda, A.R., Monsarrat, B., and Ducommun, B. (2006). CDC25B phosphorylation by p38 and MK-2. *Cell Cycle* *5*, 1649–1653. <https://doi.org/10.4161/cc.5.15.3006>.

Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., and Adams, P.D. (2017). Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D Struct. Biol.* *73*, 148–157. <https://doi.org/10.1107/S2059798316018210>.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* *75*, 861–877. <https://doi.org/10.1107/S2059798319011471>.

Li, M., Zheng, M., Wu, S., Tian, C., Liu, D., Weizmann, Y., Jiang, W., Wang, G., and Mao, C. (2018). In vivo production of RNA nanostructures via programmed folding of single-stranded RNAs. *Nat. Commun.* *9*, 2196. <https://doi.org/10.1038/s41467-018-04652-4>.

Li, X., Wang, J., Wu, C., Lu, X., and Huang, J. (2022). MicroRNAs involved in the TGF- $\beta$  signaling pathway in atherosclerosis. *Biomed. Pharmacother.* *146*, 112499. <https://doi.org/10.1016/j.biopha.2021.112499>.

Macias, M.J., Martin-Malpartida, P., and Massagué, J. (2015). Structural determinants of Smad function in TGF- $\beta$  signaling. *Trends Biochem. Sci.* *40*, 296–308. <https://doi.org/10.1016/j.tibs.2015.03.012>.

Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., et al. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* *54*, 343–355. <https://doi.org/10.1107/S1600576720013412>.

Martin-Malpartida, P., Batet, M., Kaczmarska, Z., Freier, R., Gomes, T., Aragón, E., Zou, Y., Wang, Q., Xi, Q., Ruiz, L., et al. (2017). Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. *Nat. Commun.* *8*, 2070. <https://doi.org/10.1038/s41467-017-02054-6>.

Massagué, J. (1998). TGF-beta signal transduction. *Annu. Rev. Biochem.* *67*, 753–791. <https://doi.org/10.1146/annurev.biochem.67.1.753>.

Massagué, J. (2012). TGF $\beta$  signalling in context. *Nat. Rev. Mol. Cell Biol.* *13*, 616–630. <https://doi.org/10.1038/nrm3434>.

Massagué, J., and Obenauf, A.C. (2016). Metastatic colonization by circulating tumour cells. *Nature* *529*, 298–306. <https://doi.org/10.1038/nature17038>.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* *40*, 658–674. <https://doi.org/10.1107/S0021889807021206>.

McPherson, A., and Gavira, J.A. (2014). Introduction to protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.* *70*, 2–20. <https://doi.org/10.1107/S2053230X13033141>.

Meng, W., Swenson, L.L., Fitzgibbon, M.J., Hayakawa, K., Ter Haar, E., Behrens, A.E., Fulghum, J.R., and Lippke, J.A. (2002). Structure of mitogen-activated protein kinase-activated protein (MAPKAP) kinase 2 suggests a bifunctional switch that couples kinase activation with nuclear



export. *J. Biol. Chem.* **277**, 37401–37405.  
<https://doi.org/10.1074/jbc.C200418200>.

Mertens, H.D.T., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128–141. <https://doi.org/10.1016/j.jsb.2010.06.012>.

Miyazono, K.-I., Ohno, Y., Wada, H., Ito, T., Fukatsu, Y., Kurisaki, A., Asashima, M., and Tanokura, M. (2018). Structural basis for receptor-regulated SMAD recognition by MAN1. *Nucleic Acids Res.* **46**, 12139–12153. <https://doi.org/10.1093/nar/gky925>.

Morén, A., Imamura, T., Miyazono, K., Heldin, C.-H., and Moustakas, A. (2005). Degradation of the tumor suppressor Smad4 by WW and HECT domain ubiquitin ligases. *J. Biol. Chem.* **280**, 22115–22123. <https://doi.org/10.1074/jbc.M414027200>.

Nalefski, E.A., and Falke, J.J. (1996). The C2 domain calcium-binding motif: structural and functional diversity. *Protein Sci.* **5**, 2375–2390. <https://doi.org/10.1002/pro.5560051201>.

Nebreda, Á., Fernandez, A., Velasco, L., Orozco, M., Soliva, R., Diaz, L., and Soler, D. (2020). Patent p38 $\alpha$  AUTOPHOSPHORYLATION INHIBITORS (Barcelona).

Olson, W.K., Li, S., Kaukonen, T., Colasanti, A.V., Xin, Y., and Lu, X.-J. (2019). Effects of noncanonical base pairing on RNA folding: structural context and spatial arrangements of G·A pairs. *Biochemistry* **58**, 2474–2487. <https://doi.org/10.1021/acs.biochem.9b00122>.

Pargellis, C., Tong, L., Churchill, L., Cirillo, P.F., Gilmore, T., Graham, A.G., Grob, P.M., Hickey, E.R., Moss, N., Pav, S., et al. (2002). Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **9**, 268–272. <https://doi.org/10.1038/nsb770>.

Pecimonova, M., Radvanszky, J., Smolak, D., Budis, J., Lichvar, M., Kristinova, D., Rozova, I., Turna, J., and Szemes, T. (2021). Admixed phenotype of NEDD4L associated periventricular nodular heterotopia: A case report. *Medicine (Baltimore)* **100**, e26136. <https://doi.org/10.1097/MD.00000000000026136>.

Perfettini, J.-L., Castedo, M., Nardacci, R., Ciccocanti, F., Boya, P., Roumier, T., Larochette, N., Piacentini, M., and Kroemer, G. (2005). Essential role of p53 phosphorylation by p38 MAPK in apoptosis induction by the HIV-1 envelope. *J. Exp. Med.* **201**, 279–289. <https://doi.org/10.1084/jem.20041502>.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20**, 660–664. <https://doi.org/10.1107/S0909049513010431>.

Petrova, B., Liu, K., Tian, C., Kitaoka, M., Freinkman, E., Yang, J., and Orr-Weaver, T.L. (2018). Dynamic redox balance directs the oocyte-to-embryo transition via developmentally controlled reactive cysteine changes. *Proc Natl Acad Sci USA* 115, E7978–E7986. <https://doi.org/10.1073/pnas.1807918115>.

Plant, P.J., Yeger, H., Staub, O., Howard, P., and Rotin, D. (1997). The C2 domain of the ubiquitin protein ligase Nedd4 mediates Ca<sup>2+</sup>-dependent plasma membrane localization. *J. Biol. Chem.* 272, 32329–32336. <https://doi.org/10.1074/jbc.272.51.32329>.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40, 191–285. <https://doi.org/10.1017/S0033583507004635>.

Ravi, D., Wiles, A.M., Bhavani, S., Ruan, J., Leder, P., and Bishop, A.J.R. (2009). A network of conserved damage survival pathways revealed by a genomic RNAi screen. *PLoS Genet.* 5, e1000527. <https://doi.org/10.1371/journal.pgen.1000527>.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5, 172. <https://doi.org/10.3389/fmicb.2014.00172>.

Rossi, F., Molnar, C., Hashiyama, K., Heinen, J.P., Pampalona, J., Llamazares, S., Reina, J., Hashiyama, T., Rai, M., Pollarolo, G., et al. (2017). An in vivo genetic screen in *Drosophila* identifies the orthologue of human cancer/testis gene SPO11 among a network of targets to inhibit lethal(3)malignant brain tumour growth. *Open Biol.* 7. <https://doi.org/10.1098/rsob.170156>.

Scott, J.L., Frick, C.T., Johnson, K.A., Liu, H., Yong, S.S., Varney, A.G., Wiest, O., and Stahelin, R.V. (2020). Molecular analysis of membrane targeting by the C2 domain of the E3 ubiquitin ligase smurf1. *Biomolecules* 10. <https://doi.org/10.3390/biom10020229>.

Sharma, P., Chawla, M., Sharma, S., and Mitra, A. (2010). On the role of Hoogsteen:Hoogsteen interactions in RNA: ab initio investigations of structures and energies. *RNA* 16, 942–957. <https://doi.org/10.1261/rna.1919010>.

Sheng, J., Gan, J., Soares, A.S., Salon, J., and Huang, Z. (2013). Structural insights of non-canonical U\*U pair and Hoogsteen interaction probed with Se atom. *Nucleic Acids Res.* 41, 10476–10487. <https://doi.org/10.1093/nar/gkt799>.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 113, 685–700. [https://doi.org/10.1016/s0092-8674\(03\)00432-x](https://doi.org/10.1016/s0092-8674(03)00432-x).

Smyth, M.S., and Martin, J.H. (2000). x ray crystallography. *MP, Mol Pathol* 53, 8–14. <https://doi.org/10.1136/mp.53.1.8>.

Stellwagen, E. (2009). Chapter 23 gel filtration. In Guide to protein purification, 2nd edition (Elsevier), pp.373–385.

Stouffs, K., Verloo, P., Brock, S., Régal, L., Beysen, D., Ceulemans, B., Jansen, A.C., and Meuwissen, M.E.C. (2020). Recurrent NEDD4L variant in periventricular nodular heterotopia, polymicrogyria and syndactyly. *Front. Genet.* 11, 26. <https://doi.org/10.3389/fgene.2020.00026>.

Svensson, M.J., Chen, J.D., Pirrotta, V., and Larsson, J. (2003). The ThioredoxinT and deadhead gene pair encode testis- and ovary-specific thioredoxins in *Drosophila melanogaster*. *Chromosoma* 112, 133–143. <https://doi.org/10.1007/s00412-003-0253-5>.

Svergun, D.I., and Koch, M.H.J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* 66, 1735–1782. <https://doi.org/10.1088/0034-4885/66/10/R05>.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* 28, 768–773. <https://doi.org/10.1107/S0021889895007047>.

Swatek, K.N., and Komander, D. (2016). Ubiquitin modifications. *Cell Res.* 26, 399–422. <https://doi.org/10.1038/cr.2016.39>.

Tang, L.-Y., Yamashita, M., Coussens, N.P., Tang, Y., Wang, X., Li, C., Deng, C.-X., Cheng, S.Y., and Zhang, Y.E. (2011). Ablation of Smurf2 reveals an inhibition in TGF- $\beta$  signalling through multiple mono-ubiquitination of Smad3. *EMBO J.* 30, 4777–4789. <https://doi.org/10.1038/emboj.2011.393>.

Tang, R., Zhang, G., Wang, Y.-C., Mei, X., and Chen, S.-Y. (2017). The long non-coding RNA GAS5 regulates transforming growth factor  $\beta$  (TGF- $\beta$ )-induced smooth muscle cell differentiation via RNA Smad-binding elements. *J. Biol. Chem.* 292, 14270–14278. <https://doi.org/10.1074/jbc.M117.790030>.

Tian, S., Yesselman, J.D., Cordero, P., and Das, R. (2015). Primerize: automated primer assembly for transcribing non-coding RNA domains. *Nucleic Acids Res.* 43, W522-6. <https://doi.org/10.1093/nar/gkv538>.

Tuszynska, I., and Bujnicki, J.M. (2011). DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking. *BMC Bioinformatics* 12, 348. <https://doi.org/10.1186/1471-2105-12-348>.

Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., and Bujnicki, J.M. (2015). NPdock: a web server for protein-nucleic acid docking. *Nucleic Acids Res.* 43, W425-30. <https://doi.org/10.1093/nar/gkv493>.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* 67, 293–302. <https://doi.org/10.1107/S0907444911007773>.

Vonrhein, C., Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., and Bricogne, G. (2018). Advances in automated data analysis and processing within *autoPROC*, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using *STARANISO*. *Acta Crystallogr. A Found. Adv.* 74, a360–a360. <https://doi.org/10.1107/S010876731809640X>.

Walker, J.R., Avvakumov, G.V., Xue, S., Butler-Cole, C., Finerty Jr., P.J., Weigelt, J., Sundstrom, M., Arrowsmith, C.H., Edwards, A.M., Bochkarev, A., et al. (2006). RCSB PDB - 2NSQ: Crystal structure of the C2 domain of the human E3 ubiquitin-protein ligase NEDD4-like protein <https://www.rcsb.org/structure/2nsq>.

Wang, Q., Feng, J., Wang, J., Zhang, X., Zhang, D., Zhu, T., Wang, W., Wang, X., Jin, J., Cao, J., et al. (2013). Disruption of TAB1/p38 $\alpha$  interaction using a cell-permeable peptide limits myocardial ischemia/reperfusion injury. *Mol. Ther.* 21, 1668–1677. <https://doi.org/10.1038/mt.2013.90>.

Wang, X.S., Diener, K., Manthey, C.L., Wang, S., Rosenzweig, B., Bray, J., Delaney, J., Cole, C.N., Chan-Hui, P.Y., Mantlo, N., et al. (1997a). Molecular cloning and characterization of a novel p38 mitogen-activated protein kinase. *J. Biol. Chem.* 272, 23668–23674. <https://doi.org/10.1074/jbc.272.38.23668>.

Wang, Z., Harkins, P.C., Ulevitch, R.J., Han, J., Cobb, M.H., and Goldsmith, E.J. (1997b). The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution. *Proc Natl Acad Sci USA* 94, 2327–2332. <https://doi.org/10.1073/pnas.94.6.2327>.

Wang, Z., Canagarajah, B.J., Boehm, J.C., Kassisà, S., Cobb, M.H., Young, P.R., Abdel-Meguid, S., Adams, J.L., and Goldsmith, E.J. (1998). Structural basis of inhibitor selectivity in MAP kinases. *Structure* 6, 1117–1128. [https://doi.org/10.1016/s0969-2126\(98\)00113-0](https://doi.org/10.1016/s0969-2126(98)00113-0).

Whitmarsh, A.J. (2010). A central role for p38 MAPK in the early transcriptional response to stress. *BMC Biol.* 8, 47. <https://doi.org/10.1186/1741-7007-8-47>.

Wilson, K.P., Fitzgibbon, M.J., Caron, P.R., Griffith, J.P., Chen, W., McCaffrey, P.G., Chambers, S.P., and Su, M.S. (1996). Crystal structure of p38 mitogen-activated protein kinase. *J. Biol. Chem.* 271, 27696–27700. <https://doi.org/10.1074/jbc.271.44.27696>.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242. <https://doi.org/10.1107/S0907444910045749>.

Wu, J.-W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D.J., Kyin, S., Muir, T.W., Fairman, R., et al. (2001). Crystal structure of a phosphorylated smad2. *Molecular Cell* 8, 1277–1289. [https://doi.org/10.1016/S1097-2765\(01\)00421-X](https://doi.org/10.1016/S1097-2765(01)00421-X).

Young, P.R., McLaughlin, M.M., Kumar, S., Kassis, S., Doyle, M.L., McNulty, D., Gallagher, T.F., Fisher, S., McDonnell, P.C., Carr, S.A., et al. (1997). Pyridinyl imidazole inhibitors of p38 mitogen-activated protein kinase bind in the ATP site. *J. Biol. Chem.* 272, 12116–12121. <https://doi.org/10.1074/jbc.272.18.12116>.

Zhang, Z., Jiang, H., Li, X., Chen, X., and Huang, Y. (2019). MiR-92a regulates brown adipocytes differentiation, mitochondrial oxidative respiration, and heat generation by targeting SMAD7. *J. Cell. Biochem.* <https://doi.org/10.1002/jcb.29539>.



## **Appendix**





5gcb2	GGCGAGGCGUGAAAGUGCCGUAACGCC (((.((((.....)))).....)))
5gcb6	GGCGCAGGCGCGAAAGCGCCGACGCC (((.((((.....)))).....)))
5gcb7	GGCGAGGCGGCGAAAGUGCCAAGUGCC (((.((((.....)))).....)))
yjdF (RF122 S. aureus)	AUAUAGAAGUGUAAAAACACAAUUGGGUUGGUAGUCCCAAUGCAGUUUUAAAACUGUCAGUAACCUUC UCCUCGGGUCGUCCGUCAUUUCUAAUAAAACU
yjdF_IVT	GGGAAGUGUAAAAACACAAUUGGGUUGGUAGUCCCAAUGCAGUUUUAAAACUGUCAGUAACCUUC CUCCUCGGGUCGUCCGUCAUUUCACUUCCCU (((((((.((((.....))))))(((.....)))))).....((((.....)))).....))..
Class III T7 promotor	TTCTAATACGACTCACTATA
yjdF primer 1F	TTCTAATACGACTCACTATAGGGAAGTGTA
yjdF primer 2R	TGGGACTACCAACCCAATTGTGTTTTTACACTTCCTATAGTGAGTCGTAT
yjdF primer 3F	ATTGGGTTGGTAGTCCCAATGCAGTTTTTAAACTGTCAGTAACTTCCTCCTCGGGTC
yjdF primer 4R	AGGGAAGTGAGAAATGACGGACGACCCGAGGAGGAAGGTTA

Supplementary Table 2 Data collection and refinement statistics of presented structures.

Protein	Deadhead	p38 $\alpha$ C162S			N4L C2 WT		N4L C2 F91I		
Entry	6ZMU	7Z6I	7Z9T	7PVU	(WT1)	(WT2)	(Mut1)	(Mut2)	(Mut3)
Molec. repl.	1XWA	4LOO	4LOO	4LOO	2NSQ	WT1			
Resolution range	77.31–1.95 (2.05–1.97)	77.55–2.25 (2.37–2.25)	79.70–2.60 (2.74–2.60)	77.69–1.82 (1.92–1.82)	46,48–1,75	44,66–1,19	46,81–1,07	46,79–1,25	46,64–1,43
Space group	P4 <sub>3</sub> 2 <sub>1</sub> 2	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub>	P 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Mosaicity		0.32°	0.29°	0.35°					
a, b, c (Å)	111.83, 111.83, 107.00	65.14, 74.57, 77.55	67.14, 68.07, 79.98	66.92, 67.97, 79.69	33.389 59.477 74.511	30.988 57.15 71.556	32.962 60.045 74.731	32.848 60.022	32.845 59.744
$\alpha$ , $\beta$ , $\gamma$ (°)	90	90	$\beta=94.78$	$\beta=94.75$	90	90	90	90	90
Multiplicity	10.8 (9.9)	5.6 (5.6)	3.0 (3.1)	5.8 (5.8)					
Completeness (%)	90.0 (38.3)	100 (100)	99.2 (99.0)	100 (100)					
$\langle I/\sigma(I) \rangle$	18.1 (1.5)	16.5 (9.2)	9.2 (3.5)	16.0 (4.4)					
Wilson B factor (Å <sup>2</sup> )	45	29	44	41					
Rmerge	0.07 (1.50)	0.092 (0.451)	0.086 (0.465)	0.057 (0.453)					
Rmeas	0.07 (1.57)								
CC1/2	1.00 (0.62)								
Rwork/Rfree	0.19/0.22	0,21/0,23	0,21/0,27	0,21/0,25	~0,25	~0,21	~0,16	~0,17	~0,25
R.m.s.d., bonds (Å)	0.86	0,0075	0,0038	0,01					

R.m.s.d., angles (°)	0.99	1,42	1,39	1,18					
Ramachandran favored (%)	97.8	91,3	86,0	92,9					
Ramachandran allowed (%)	2.2	4,6	7,4	5,3					
Ramachandran outliers (%)	0	4,1	6,6	1,8					
Clashscore	6.95								
Average B factor (Å <sup>2</sup> )		33	60	70					
Overall	39.07								
Macromolecules	37.61								
Ligands	97.88								
Solvent	44.83								

Deposition of these five structures not finalized for September 2022.

Supplementary Table 3 Buffer compositions

TBE	0,45 M Tris base, 0,45 M H <sub>3</sub> BO <sub>3</sub> , 0,01M EDTA
RNA denaturing loading dye	Sample mixed 1:1 with formamide
TG	0,2 M Tris base, 2M Glycine
Protein gel cathode buffer	0.1M Tris, 1M Tricine, 0.1% SDS pH 8.25
Protein gel anode buffer	0.2M Tris pH 9.0
Protein gel loading buffer	50 mM Tris 8 pH, 12% glycerol, 4% SDS, 0.01% Coomassie blue G-250, 2% 2-Mercaptoethanol
Affinity chromatography binding buffer	40 mM Tris-HCl pH 8,0, 400 mM NaCl, 40 mM imidazole, 2 mM TCEP and 0,1% Tween
Affinity chromatography elution buffer	40 mM Tris-HCl pH 8,0, 400 mM NaCl, 400 mM imidazole, 2 mM TCEP
TEV buffer	20 mM Tris-HCl pH 7.2, 100 mM NaCl and 1 mM TCEP
WALK polymerase 10x PCR reaction buffer (as per manufacturer)	200 mM Tris-HCl pH 8,5, 100 mM KCl, 100 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 20 mM MgSO <sub>4</sub> , 1% Triton X-100
IVT buffer (based on (Kieft and Batey, 2004))	40 mM Tris-HCl pH 8.0, 10 mM DTT, 2 mM spermidine, 0,1% Triton X-100
EMSA binding buffer	25mM Tris-HCl pH 7.4, 115mM KCl, 1mM MgCl <sub>2</sub> , 1mM DTT, 0,1g/l BSA
XBL buffer	20 mM Tris pH7.5, 100 mM NaCl, 2 mM TCEP

Supplementary Table 4 Providers of commercially-sourced reagents and supplies.

<b>Component</b>	<b>Manufacturer</b>
<b>Protein purification</b>	
pET plasmid	IRB's Protein Expression facility
Primers	Sigma
Recombinase A	New England Biolabs
Bacteria strain	IRB's Protein Expression facility
Lysozyme	Sigma
DNAse	Merck
Acrylamide:bisacrylamide 19:1	Bio-Rad
Precision Plus dual color standard	Bio-Rad
<b>Chomatography columns</b>	
HisTrap™ HP 1/5 ml	Cytiva
HiLoad 16/600 Superdex™ 75 pg	GE Healthcare
Superdex™ 75 Increase 10/300 gl	Cytiva
<b>RNA production</b>	
Commercially sourced RNAs	Biomers.net
Primerize primers	Sigma-Aldrich
WALK DNA polymerase	A&A Biotechnology
Deoxyribonucleotides	NZYtech
GeneRuler 1kb DNA Ladder	Thermo Scientific
T7 RNA polymerase	A&A Biotechnology
BSA	Sigma
Rybonucleotides	A&A Biotechnology
RNAse inhibitors	EURx
PRECISION RNA mass marker 10-100nt	FUTURE synthesis
<b>Crystallography</b>	
Crystal screens	Respective manufacturers
Conditions and chemicals	Molecular Dimensions



## A.2 List of publications

Three more manuscripts in preparation (September 2022). I am a co-first author in one of them.

*Structural basis of p38 $\alpha$  autophosphorylation inhibitors that target the non-canonical activation pathway*

González L., Díaz L., Pous J., **Baginski B.**, Scarpa M., Duran A., Igea A., Martín P., Ruiz L., Esguerra M., Collizzi F., Mayor-Ruiz C., Biondi R. M., Soliva R., Macias M.J., Orozco M., Nebreda A. R.

In submission, 2022

*Molecular basis for DNA recognition by the maternal pioneer transcription factor FoxH1*

Pluta R, Aragon E, Prescott N, **Baginski B.**, Ruiz L, Flood J, Martin-Malpartida P, Massague J, David Y, Macías M

In review, 2022

*MODOMICS: an operational guide to the use of the RNA modification pathways database.*

Boccaletto, P., and **Bagiński, B.**

Methods Mol. Biol. 2284, 481–505, 2021

This publication is related to the author's work at previous laboratory.

*Structures of the germline-specific Deadhead and thioredoxin T proteins from Drosophila melanogaster reveal unique features among thioredoxins.*

Freier, R., Aragón, E., **Bagiński, B.**, Pluta, R., Martin-Malpartida, P., Ruiz, L., Condeminas, M., Gonzalez, C., and Macias, M.J.

IUCrJ 8, 281–294, 2021

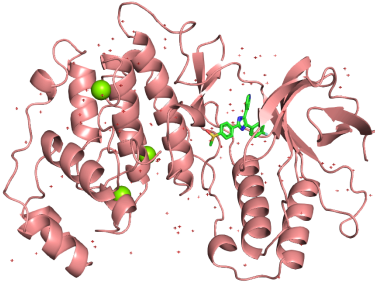
*Unveiling the dimer/monomer propensities of Smad MH1-DNA complexes.*

Ruiz, L., Kaczmarek, Z., Gomes, T., Aragon, E., Torner, C., Freier, R., **Baginski, B.**, Martin-Malpartida, P., de Martin Garrido, N., Marquez, J.A., et al. (2021).

Comput. Struct. Biotechnol. J. 19, 632–646, 2021

### A.3 List of deposited crystal structures

For the end of September 2022. Further six structures awaiting deposition.

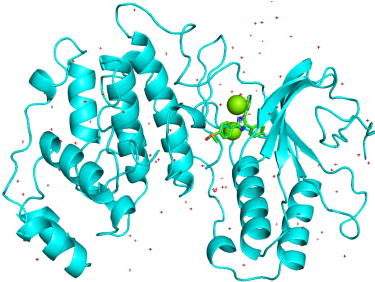


8ACM

*Crystal structure of WT p38alpha*

Pous, J., **Baginski, B.**, Gonzalez, L., Macias, M.J., Nebreda, A.R.

08.07.2022, hold for publication

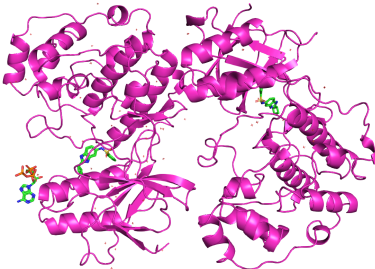


8ACO

*Crystal structure of WT p38alpha*

Pous, J., **Baginski, B.**, Gonzalez, L., Macias, M.J., Nebreda, A.R.

08.07.2022, hold for publication



7Z9T

*Crystal structure of p38alpha C162S in complex with ATPγS and CAS 2094667-81-7 (in catalytic site, Y35 out), P 1 21 1*

**Baginski, B.**, Pous, J., Gonzalez, L., Macias, M.J., Nebreda, A.R.

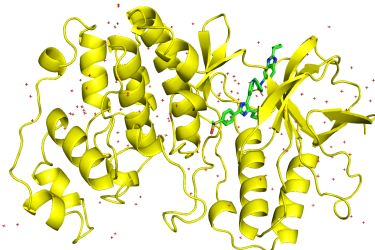
2022-03-21, hold for publication

7Z6I

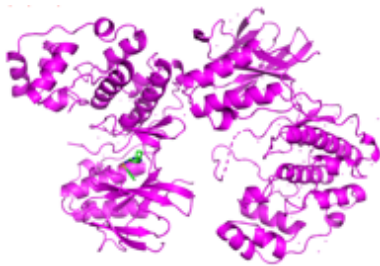
*Crystal structure of p38alpha C162S in complex with SB20358 and CAS 2094667-81-7 (behind catalytic site; Y35 in), P 21 21 21*

**Baginski, B.**, Pous, J., Gonzalez, L., Macias, M.J., Nebreda, A.R.

2022-03-11, hold for publication



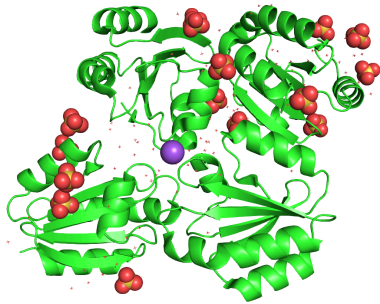




7PVU

*Crystal structure of p38alpha C162S in complex with CAS2094511-69-8, P 1 21 1*

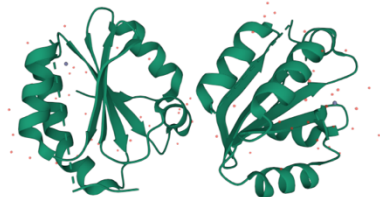
**Baginski, B.**, Pous, J., Gonzalez, L., Macias, M.J., Nebreda, A.R.  
2021-10-05, hold for publication



6ZMU

*Crystal structure of the germline-specific thioredoxin protein Deadhead (Thioredoxin-1) from Drosophila melanogaster, P43212*

**Baginski, B.**, Pluta, R., Macias, M.J.  
2020-07-03



6Z7O

*Crystal structure of Thioredoxin T from Drosophila melanogaster*  
Freier, R., Aragon, E., **Baginski, B.**, Pluta, R., Martin-Malpartida, P., Torner, C., Gonzalez, C., Macias, M.  
2020-05-31

## A.4 Curriculum Vitae

### Education

- 2018-2022**      **PhD in Biomedicine**  
**University of Barcelona; Faculty of Biology**
- 2015-2016**      **Master's degree in Biotechnology (with Merit)**  
**University of Glasgow;**  
**College of Medical, Veterinary and Life Sciences**  
Serine recombinase- role of the central E-helix in reaction complex formation.
- 2011-2015**      **Bachelor's degree in Biotechnology**  
**University of Wroclaw; Faculty of Biotechnology**  
Overview of the methods of aptamer selection for biotechnology applications.

### Research experience

- 2018-2022**      **PhD in Biomedicine**  
**Institute for Research in Biomedicine (IRB Barcelona)**  
**Structural Characterization of Macromolecular Assemblies**  
Assessing small molecules for their anti-cancer mechanism of action, and deciphering the connections between cell signalling and gene expression using high resolution structural biology (mostly crystallography). Investigating how these processes are regulated/misregulated and their implications in human diseases.  
Protein-small molecules/DNA/RNA interactions.
- Feb 2017- Nov 2018**      **Research Technician**  
**International Institute of Molecular and Cell Biology in Warsaw (IIMCB)**  
**Laboratory of Bioinformatics and Protein Engineering**  
Bacterial riboswitch research (one of the antibiotic resistance mechanisms). Structural biology approach to RNA (x-ray crystallography, SHAPE, SAXS), aided by in silico prediction and design.  
Design of RNA molecules, production of high purity material for crystallography experiments. Data collection at various synchrotrons across the Europe.
- Feb-Jun 2015**      **Voluntary practice**  
**Wroclaw University of Environmental and Life Sciences**  
**Faculty of Food Science**  
Operation, maintenance and construction details of bioreactors; *Yarrowia lipolytica* cell culturing. Research on proteolytic activity of selected *Bacillus cereus* strains.
- Aug 2013**      **Practice in analytical laboratory**

### **Beskid Oncology Centre– City Hospital in Bielsko-Biala**

Research and microscopic observations of infant's blood and of blood of patients in different leukaemia stages.

Under supervision: assessment of morphological, chemical and atomic composition of patient's full blood and plasma. Assessing patient's antigen response to viral infections. Assistance during patient's blood samples acceptance, handling blood diagnostic stations and patients evidence database.

### **March-July 2013 Voluntary practice**

#### **University of Wroclaw, Laboratory of Nuclear Proteins**

Learning how to single-handedly work with standard laboratory equipment. Mainly work related to nucleic acids.

### **Additional training**

#### **Sep-Dec 2021 From Scientists to Innovators for Industry**

EIT Health course for scientists facilitating industrial collaborations.

I performed primary research (empathic interviews with PD patients, neurologist and patient organization), organized meetings with experienced mentors (Boehringer Ingelheim, Philips, MADoPA) and was responsible for the technical part of project (including the Android app).

**S2C2 CryoEM CCP-EM Modeling Workshop** - cryogenic electron microscopy structure solving course at the Stanford's SLAC National Accelerator Laboratory (USA). 10-13 Nov 2020

**CCP-EM Spring Symposium** – cryogenic electron microscopy structure solving workshop. 28.04-01.05.2020

**From Academia to Project Management and From Academia to Industry**, Oct 2019

**Peer2Scale Creativity Workshop: Experience a full cycle of innovation in healthcare**, Oct 2019

**Best Practices and Co-creation for RRI, and Lean Start-up Research**, Sep 2019

**FEBS Practical Course: Biomolecules in Action II** - techniques for sample preparation and quality assessment in DESY, Hamburg. 23–28 June 2019

**EMBO Practical Course: Solution scattering from biological macromolecules** - Small Angle X-ray Scattering course at EMBL Hamburg. 19-26.11.2018

**CCP4/Ben-Gurion University workshop** – X-ray crystallography protein structure determination course in Beersheba, Israel. 18-23.02.2018

**Crystallography A, Laboratory** – theoretical course of crystallography for students. Faculty of Chemistry, Warsaw University, winter semester 2017.

**Polish Atomic Agency** - registered radiation worker (category B; highest)

Multiple poster presentations and talk at the international conferences. Basic programming - Bash (Linux scripts), Kotlin (Android Studio), HTML.

## Voluntary activities

### **2020,2021,2022 Crazy about Biomedicine tutor**

I delivered lectures and laboratory workshops in drug design crystallography during this semester-long workshop, co-organized with Fundació Catalunya La Pedrera. Aimed for high-school students to help them guide their future careers. I also performed interviews.

### **2021 IRB's Artist in residence - Dr Jo Milne**

An enriching opportunity to exchange ideas, reflections, and learn lessons in the differences in how we (scientists) and artists can look at the same picture of the world around us.

### **Jul/Aug 2021 Treball de recerca mentor**

Mentoring 2 students during their school research project (protein crystallography).

### **Semester II 2020 Mentor at Teach&Learn programme (Grupa Maspex)**

Helping high school pupils in understanding biology and chemistry.

### **Apr-Jun 2020 Orfeu Programme volunteer – COVID19 diagnostic in Catalonia**

Orfeu programme begun with the start of COVID19 pandemic, as a temporary mean to increase healthcare system capacity for Covid-19 mass testing.

I proposed and tested the use of one of the robots I had experience with, to significantly increase efficiency of one of the kit preparation process stages. As a volunteer in the RNA extraction team, I was using robot to transfer patient's samples and begin the viral RNA extraction process in class 2 bio-safety laboratory.

