



**Andreia Catarina Fernandes Pinto Fernandes**

Licenciada em Biologia Aplicada

**Structural and Functional  
Characterization of a New Bacterial  
Target Against Tuberculosis:  
The Phosphatase PtpA**

Dissertação para obtenção do Grau de Mestre em  
Bioquímica

Orientador: Doutora Teresa Santos Silva, Professora  
Auxiliar, FCT-NOVA

Co-orientador: Doutora Maria Manuel Marques,  
Professora Auxiliar, FCT-NOVA



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Setembro de 2019**





**Andreia Catarina Fernandes Pinto Fernandes**

Licenciada em Biologia Aplicada

**Structural and Functional  
Characterization of a New Bacterial  
Target Against Tuberculosis:  
The Phosphatase PtpA**

Dissertação para obtenção do Grau de Mestre em  
Bioquímica

Orientador: Doutora Teresa Santos Silva, Professora  
Auxiliar, FCT-NOVA

Co-orientador: Doutora Maria Manuel Marques,  
Professora Auxiliar, FCT-NOVA



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

Setembro de 2019



“Structural and Functional Characterization of a New Bacterial Target Against Tuberculosis:  
The Phosphatase PtpA”

Copyright © em nome de Andreia Catarina Fernandes Pinto Fernandes, da Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



## Acknowledgements

Um ano de tanto! Há um momento em que fazer uma retrospeção é fundamental para pararmos e percebermos que todos os frutos bonitos que começamos a colher são nascidos das sementes que plantamos com esforço e dedicação ao longo de um (longo) ano. Foram dias inteiros de aprendizagem, de desassossego, de gargalhadas, de frustrações, de persistência, de novas amizades, e sobretudo, de ser absolutamente feliz no caminho que escolhi.

Gosto de acreditar que tenho nas minhas escolhas os melhores professores de vida, aos quais quero agradecer e dedicar esta secção:

Em primeiro lugar, quero agradecer à minha Família que sempre me apoiou incondicionalmente e que possibilitou a realização desta tese. Em especial aos meus pais, Maria e Paulo, por todo o apoio prestado ao longo destes anos fora de casa. Sem vocês não teria chegado onde cheguei. À minha pestinha, o meu irmão Diogo. Mano, prometo que irei mais vezes a casa para te chatear, para te ajudar com os trabalhos de casa e para o resto que precisares. Aos meus avós Maria e Armindo, os segundos pilares da minha vida. “Vovós”, embora não percebam bem o que eu faço, um agradecimento não chega por todo o vosso carinho e apoio. Nunca conseguirei retribuir o tanto que fizeram e que continuam a fazer por mim. Crescer tem sido mais difícil longe de vocês.

Agradeço à Professora Maria João Romão pela oportunidade de realizar esta dissertação no seu fantástico grupo que é o XTAL.

À minha orientadora Teresa. Professora, um infindável obrigado por todos os momentos de orientação, preocupação e ensinamentos. Agradeço ainda pelas novas experiências de vida, pelas conversas fora do ambiente de trabalho, pelo seu notório positivismo perante a vida. Não poderia ter escolhido melhor orientadora e melhor companheira da minha primeira viagem de avião! Muito obrigada por tudo!

À minha co-orientadora, professora Maria Manuel Marques pela sua simpatia e disponibilidade para esclarecer todas as minhas dúvidas. Foi um gosto explorar esta nova área consigo. Obrigada!

Ao Hérnan Terenzi um sincero agradecimento pela oportunidade de colaboração e pela sua prontidão em responder aos meus pedidos. Obrigada!

Ao Marino, a minha “bússola orientadora”. Ponto prévio - um obrigada nunca será suficiente para te agradecer. Foste sem dúvida um elemento fundamental neste processo. Desde o meu revirar de olhos à tua incessante pergunta matinal “Quais os planos para hoje?”, aos fins de dia tardios no lab a discutir as minhas várias dúvidas existenciais, às conversas sobre futebol, à tua ilimitada paciência em ajudar-me na bancada. Devo-te muito do que sei hoje. És a prova de resiliência

em pessoa e espero que um dia consigas realmente alcançar o que queres, porque o mereces (Que lamechice!). Dúvidas? Sabes que muitas. Amanhã às 8h?

Ao Francisco Leisico, por todos os momentos os quais abdicaste do teu trabalho para me transmitir toda a tua sabedoria e me servires o teu prestável apoio. Foste incansável e um pilar essencial durante o meu percurso. Desejo-te toda a sorte do mundo!

Jayaraman Muthukumaran, I will be forever grateful for your availability and patience throughout my learning process. You taught me a lot about computational work. Thanks for all the moments we shared in X-ray room. धन्यवाद

Filipa Engrola, a minha parceira do laboratório. Obrigada por todos os momentos que partilhamos juntas, pelo trabalho de equipa, pelas conversas, pelas pausas na varanda do 4º piso, pelos almoços na teresa gato. Foste uma agradável surpresa e espero que continuemos a manter esta amizade!

Ao João, o meu “vizinho da frente”. A vida no lab seria muito mais enfadonha sem as nossas desavenças (saudáveis, não é Pacote?). Apesar de todas estas questões sócio-temporais (créditos ao Marino), quero agradecer-te pelas ajudas que não hesitaste em ceder-me. Obrigada, rezingão!

Ao resto do grupo XTAL - Márcia pela sua extraordinária capacidade de coordenar um laboratório inteiro e mesmo assim ainda ter tempo para as mais desconcertantes dúvidas. É a nossa Super-Mulher; À Diana Gesto, pelas tuas dicas e pelos momentos zen das nossas idas ao ginásio; À Patrícia, à Clara, à Raquel, à Viviana, à Filipa Trovão, à Benedita, à Diana Ribeiro, ao Cristiano, à Ana Luísa Carvalho e à Angelina Palma pelos bons momentos que partilhamos no cantinho do café, pelos almoços e pelo *escape room*. Foram essenciais para a minha integração no grupo! Muito obrigada!

E por último, mas não menos importante (de todo!), ao Duarte, o meu companheiro de vida. São sempre muitas as palavras que guardo para ti. E que saibas o quanto mais forte e melhor sou porque tu existes na minha vida. Agradeço-te por tudo o que fazes por mim, sobretudo pelo sossego que transmites ao meu coração-furacão e pelo teu infindável apoio nas minhas fases menos boas. Que continuemos nesta subida a ganhar fôlego juntos e o resto fica para nós.

*“Anything is possible if you’ve got enough nerve”*

*- J.K. Rowling*



# Abstract

Tuberculosis (TB) is one of the top causes of death remaining a major public health problem worldwide. *Mycobacterium tuberculosis* is the agent of TB, infecting the human respiratory tract. Its remarkable pathogenicity hinges upon the ability to challenge the immune system of the host by secreting phosphatases into macrophages. Among them, Protein Tyrosine Phosphatase A (PtpA) plays a key role on the infection process, preventing the phagosome-lysosome fusion and promoting the inhibition of phagosome acidification. Thus, PtpA becomes a promising target for the development of new anti-TB drugs.

The aim of this work is to contribute to find new structure-based drug design approaches against TB, studying the inhibitory properties of three different families of compounds towards PtpA – chalcones, thiosemicarbazones and azaindoles.

The protein was overexpressed in *E. coli* – final yield of 20 mg protein/ liter of culture – and successfully purified using affinity chromatography. To provide new insights into the binding mode of the studied compounds, molecular docking studies were performed suggesting thiosemicarbazones as non-competitive inhibitors and the chalcones and azaindoles with a preferential active site binding.

The protein was also biophysically characterized. The oligomeric state was confirmed by SEC, proving that PtpA is a monomer in solution. The protein stability was assessed through TSA revealing that, with 10% glycerol, PtpA resists to the effects of 10% DMSO. TSA was also used to find a suitable protein storage condition (-80°C) and to confirm PEG400 as an alternative solvent for the inhibitors. In addition, distinct biophysical approaches – TSA, MST and urea-gel electrophoresis – were implemented to detect protein-ligand interactions but definitive evidence were not obtained.

Ligand-free and co-crystallization assays were extensively explored and several crystals were tested at the ESRF, Diamond and MAX IV. Two crystal structures were obtained: a co-crystallization PtpA-Lap11 structure at 3.6 Å resolution and a soaking PtpA-C33 structure at 2.8 Å resolution. Despite the low/medium resolution obtained, both structures reveal the potential binding of the inhibitors with suspicious density blobs near His120B for Lap11 and at the active site for C33. The ligands were preliminarily modelled but further refinement cycles are required to elucidate the respective binding.

**Keywords:** Tuberculosis, *Mycobacterium tuberculosis*, Protein Tyrosine Phosphatase A (PtpA), Chalcones and Thiosemicarbazones, Azaindoles, *In silico* studies, Biophysical and Structural Characterization



## Resumo

A tuberculose (TB) é uma das principais causas de morte permanecendo um problema global de saúde pública. A bactéria *Mycobacterium tuberculosis* é responsável pela doença afetando o sistema respiratório através da sua notável capacidade de desafiar o sistema imunológico do hospedeiro, secretando fosfatases para os macrófagos. A proteína tirosina fosfatase A (PtpA) desempenha um papel crucial na infecção, impedindo a fusão do fagossoma com o lisossoma e consequente acidificação do fagossoma. Assim, a PtpA constitui um alvo promissor para o desenvolvimento de novos fármacos anti-tuberculose.

Este trabalho pretende explorar novas abordagens no desenvolvimento de terapias contra a tuberculose, estudando os mecanismos de inibição da PtpA desencadeados por três diferentes famílias de compostos – chalconas, tiosemicarbazonas e azaindoles.

A proteína foi sobre-expressa em *E. coli* (rendimento final de 20 mg/ litro de cultura) e purificada por cromatografia de afinidade. O modo de ligação dos compostos em estudo à proteína foi elucidado por cálculos de *docking* molecular concluindo-se que as tiosemicarbazonas apresentaram um comportamento alostérico enquanto as chalconas e os azaindoles se ligaram preferencialmente ao centro ativo.

A caracterização biofísica da proteína confirmou o seu estado oligomérico (monômero em solução) por cromatografia de exclusão molecular. A estabilidade foi avaliada por TSA, concluindo-se que a presença de 10% de glicerol confere resistência aos efeitos de 10% de DMSO. Experiências adicionais de TSA identificaram a melhor condição de armazenamento da proteína (-80°C) e confirmaram o PEG400 como um solvente alternativo para os inibidores. Paralelamente, diferentes abordagens biofísicas – TSA, MST e eletroforese em gel de ureia – foram implementadas para detetar interações proteína-ligando, mas não foram obtidas evidências definitivas de interação.

Múltiplos ensaios de cristalização originaram vários cristais testados em sincrotrões distintos – ESRF, DLS e MAX IV. Obtiveram-se duas estruturas, resultantes de ensaios de co-cristalização (PtpA-Lap11, resolução de 3.6 Å) e de *soaking* (PtpA-C33, resolução de 2.8 Å). Apesar das resoluções obtidas, identificaram-se *blobs* de densidade nas duas estruturas que poderão corresponder aos ligandos – perto do resíduo His120B (Lap11) e no centro ativo (C33). A modelação preliminar dos ligandos foi testada, sendo necessários ciclos de refinamento adicionais para esclarecer a eventual ligação.

**Palavras-chave:** Tuberculose, *Mycobacterium tuberculosis*, Proteína Tirosina Fosfatase A (PtpA), Chalconas e Tiosemicarbazonas, Azaindoles, Estudos *In silico*, Caracterização Biofísica e Estrutural



# Table of Contents

<b>Acknowledgements</b> .....	<b>VII</b>
<b>Abstract</b> .....	<b>IX</b>
<b>Resumo</b> .....	<b>XI</b>
<b>Table of Contents</b> .....	<b>XIII</b>
<b>List of Figures</b> .....	<b>XVII</b>
<b>List of Tables</b> .....	<b>XXI</b>
<b>List of Abbreviations and Symbols</b> .....	<b>XXIII</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1 Tuberculosis Epidemic .....	3
1.1.1 From Prehistory to the Present .....	3
1.1.2 Transmission, Risk Factors and Symptoms.....	4
1.2 <i>Mycobacterium tuberculosis</i> .....	5
1.2.1 Pathogen Characterization and Cell Wall Structure.....	5
1.2.2 Pathogenesis: Macrophage Manipulation and Phagosome Hijacking .....	5
1.3 Protein Tyrosine Phosphatase A.....	5
1.3.1 PtpA Mechanism of Action .....	5
1.3.2 PtpA Structural Characterization .....	6
1.3.3 The Catalytic Mechanism .....	8
1.4 Treatments and Major Challenges .....	9
1.4.1 Current Treatments: Antibiotics and Vaccine.....	9
1.4.2 HIV Coinfection and Drug Resistance.....	9
1.5 Targeting PtpA for TB Drug Development - Inhibitors.....	9
1.6 Drug Design.....	12
1.7 Biomolecular Crystallography.....	13
1.7.1 From X-rays to Structure .....	13
1.7.2 Protein Crystallization .....	14
1.7.3 Data Collection from Crystals .....	16
1.7.4 The “Phase Problem” and Structure Solution .....	18
1.7.5 Model Building, Refinement and Validation .....	19
1.7.6 Synchrotron Radiation .....	20
<b>2. Objectives</b> .....	<b>21</b>
<b>3. Materials and Methods</b> .....	<b>25</b>
3.1 PtpA Heterologous Expression .....	27
3.2 PtpA Purification .....	27
3.3 <i>In Silico</i> Characterization Studies.....	28

3.3.1 Molecular Docking .....	28
3.3.2 Molecular Dynamics Simulation .....	29
3.4 Biophysical Characterization Studies .....	29
3.4.1 Size Exclusion Chromatography .....	29
3.4.2 Thermal Shift Assay .....	30
3.4.3 Microscale Thermophoresis .....	30
3.4.4 Urea-Polyacrylamide Gel Electrophoresis .....	31
3.5 Structural Characterization Studies .....	32
3.5.1 Crystallization Assays .....	32
3.5.1.1 Ligand-free PtpA Crystallizations .....	32
3.5.1.2 Co-crystallizations Experiments .....	33
3.5.1.3 Microseeding Experiments .....	35
3.5.2 X-ray Crystallography .....	35
3.5.2.1 X-ray Diffraction Experiment and Data Collection .....	35
3.5.2.2 Structure Determination and Refinement .....	35
<b>4. Results and Discussion .....</b>	<b>37</b>
4.1 PtpA Heterologous Expression .....	39
4.2 PtpA Purification .....	39
4.3 <i>In Silico</i> Characterization Studies .....	42
4.3.1 Molecular Docking .....	42
4.3.1.1 Chalcones and Thiosemicarbazones .....	42
4.3.1.2 Azaindoles .....	45
4.3.2 Molecular Dynamics Simulation .....	47
4.4 Biophysical Characterization Studies .....	49
4.4.1 Size Exclusion Chromatography .....	49
4.4.2 Thermal Shift Assay .....	51
4.4.2.1 Protein Stability .....	51
4.4.2.2 Protein-Ligand Interactions .....	55
4.4.3 Microscale Thermophoresis .....	59
4.4.3.1 Protein Labeling <i>via</i> His-tag .....	59
4.4.3.2 Binding Check Assay .....	59
4.4.3.3 Binding Affinity Assay .....	62
4.4.4 Urea-Polyacrylamide Gel Electrophoresis .....	64
4.5 Structural Characterization Studies .....	65
4.5.1 Crystallization Assays .....	65
4.5.1.1 Ligand-free and Co-crystallization Experiments .....	65
4.5.1.2 Microseeding Experiments .....	68
4.5.2 Structure Determination and Analysis .....	70
4.5.2.1 Data Collection and Structure Solution .....	70
4.5.2.2 Refinement of PtpA-Lap11 and PtpA-C33 Structures .....	74

4.5.3 Characterization of the Structures.....	75
4.5.3.1 PtpA-Lap11 Structure .....	75
4.5.3.2 PtpA-C33 Structure .....	78
<b>5. Conclusions and Future Perspectives .....</b>	<b>83</b>
<b>6. Bibliography .....</b>	<b>91</b>
<b>7. Appendix.....</b>	<b>97</b>
7.1 Appendix 1.....	99
7.2 Appendix 2.....	100
7.3 Appendix 3.....	101
7.4 Appendix 4.....	102
7.5 Appendix 5.....	104
7.6 Appendix 6.....	108
7.7 Appendix 7.....	109





# List of Figures

<b>Figure 1.1</b> – TB burden countries worldwide and estimated incidence rates in 2017. South Africa and south-east Asia countries are the most affected by TB incidence. ....	3
<b>Figure 1.2</b> – Representation of the main types of TB conditions. In latent TB, <i>Mycobacterium tuberculosis</i> is inhaled reaching the first-line defense cells, the alveolar macrophages. The infected macrophages recruit other immune cells to attack forming a granuloma, where the pathogen is contained in a latency state. The active TB infection occurs when infection overcomes the immune response. Consequently, the granuloma fails to contain the infection resulting in bacteria dissemination and re-entry into respiratory tract to onward transmission. ....	4
<b>Figure 1.3</b> – PtpA mechanism of action during <i>Mycobacterium tuberculosis</i> infection. (A) PtpA is secreted into the macrophage cytosol by Mtb. (B) PtpA protein binds to subunit H of V-ATPase complex localizing and dephosphorylating its substrate, the VPS33B protein. (C) Membrane fusion machinery is blocked and V-ATPase pump is excluded from mycobacterial phagosome. ....	6
<b>Figure 1.4</b> – (A) Crystal structure of Mtb PtpA at 1.9 Å resolution (PDB:1U2P). (B) Overall structure of the active site which comprises the PTP loop (residues 11-18) and a group of hydrophobic residues that form the wall of the active site cavity. The structure exhibits a chlorine ion in the phosphate binding site. Figure performed using PyMOL software. ....	7
<b>Figure 1.5</b> – (A) Superposition of Mtb PtpA (green, PDB:1U2P) and the human LMW-PTPA (purple, PDB: 5PNT) (B) Electrostatic potential surface of PtpA (C) Electrostatic potential surface of HCPTPA. ....	7
<b>Figure 1.6</b> – The catalytic mechanism of PTPases. ....	8
<b>Figure 1.7</b> – General chemical structures of a chalcone and a thiosemicarbazone. ....	10
<b>Figure 1.8</b> – General chemical structure of an azaindole drawn in ChemDraw. ....	10
<b>Figure 1.9</b> – Chemical structures of the three families of compounds – chalcones and thiosemicarbazones with inhibitory effect to PtpA, and the library of available azaindoles compounds (1, 15 and 16 are synthesized). Structures drawn in ChemDraw. ....	10
<b>Figure 1.10</b> – Representation of the main characterization studies and techniques of protein-ligand interactions used in many drug design projects. ....	12
<b>Figure 1.11</b> – Schematic representation of the main steps involved in structural protein determination. Prior to start the crystallization trials, it is necessary to express and purify the protein of interest using different molecular biology and biochemistry procedures. After obtaining suitable crystals for the x-ray diffraction experiment, the electron density maps are obtained which are used to build the initial model. This preliminary model is then refined several times and validated leading to the final model which is ready to be deposited in the PDB. ....	13
<b>Figure 1.12</b> – Schematic representation of the unit cell in the crystal lattice. The unit cell constants: the three vectors - a, b and c and the three angles - $\alpha$ (between b and c), $\beta$ (between a and c) and $\gamma$ (between a and b) ....	14
<b>Figure 1.13</b> – Illustrative representation of the assembly of unit cells in a three-dimensional crystal lattice. By symmetry operations, the asymmetric unit generates the unit cell, which in turn, through translation operations, generates the crystal lattice, which is repeated in an orderly manner in three-dimensional space. ....	14
<b>Figure 1.14</b> – Protein crystallization diagram. In diagram are represented two distinct regions: undersaturation and supersaturation. The supersaturation zone is divided in metastable zone, nucleation zone and precipitation zone. The soluble proteins molecules will move from the undersaturation zone into nucleation zone where nuclei are formed, which will slowly and orderly grow to form crystals in the metastable zone. ....	15
<b>Figure 1.15</b> – Vapor diffusion method and the hanging and sitting-drop techniques. ....	16

**Figure 1.16** – Bragg's law geometric representation (left) and equation (right). The Bragg's law can be graphically interpreted allowing the understanding of an X-ray experiment as the reflection on a set of imaginary planes in the crystal.....17

**Figure 1.17** – Schematic representation of the X-ray diffraction experiment. An intense X-ray beam is directed at the protein crystal (a three-dimensional array of repeated and ordered units). The X-rays are diffracted by the atom's electrons and are recorded in a detector.....17

**Figure 1.18** – (A) European synchrotrons facilities. (B) Schematic representation of a synchrotron structure. (C) Illustration of a beamline, a physical and dynamic space supplied by equipment that brings X-ray beam to the sample in study.....20

**Figure 4.1** – On the left, IMAC chromatogram resulting from PtpA purification. Legend: A – Elution peaks of proteins with no affinity to the nickel column; B - Elution peak of proteins with low affinity to the nickel column (step of 12% B); C - PtpA elution peak resulting in the increase of the ionic strength by a gradient of buffer B (12-100%B). On the right, the close-up view of PtpA elution peak and the corresponding protein collected fractions (1 mL).....40

**Figure 4.2** – SDS-PAGE (10%) with fractions from the expression and purification steps of PtpA.....40

**Figure 4.3** – PtpA desalting chromatogram. The blue peaks correspond to the protein elution (two injections were applied) and the red peaks represent the imidazole and salt elution from the desalting columns.....41

**Figure 4.4** – Schematic representation of the thiosemicarbazones overlapped, obtained by blind docking using AutoDock calculations, and their respective substituent groups in R<sub>1</sub> and R<sub>2</sub>. Figure prepared in PyMOL.....43

**Figure 4.5** – Allosteric binding mode of the inhibitor Lap11 to PtpA obtained by blind docking. (A) Lap11 binding pose on the protein electrostatic surface. (B) Cartoon protein representation (green) with the hydrophobic residues interacting with Lap11 colored in red (Arg108, Met109, Phe113, Pro115, Gly118, Thr119, Ala121, Gly148 and Asp151). H-bonded residues (Ser112, Asp114 and Ser117) are shown in green sticks and hydrogen bonds illustrated in black dashed lines. Figures prepared in PyMOL.....44

**Figure 4.6** – Best blind docking-based pose for the Lap04 compound within the adjacent site of PtpA. (A) Protein electrostatic surface (calculated by APBS in PyMOL) showing the R<sub>1</sub> and R<sub>2</sub> groups in differently charged regions. (B) LigPlot schematic diagram of PtpA-Lap04 interactions. The residues circled in red represent the hydrophobic interactions. Residues Ser112 and Ser117 are the main residues responsible for H-bonds in this complex (green dashed lines).....44

**Figure 4.7** – Representation of the blind docking-based binding mode of the chalcone C33 within the catalytic site of PtpA. (A) Electrostatic surface of PtpA (calculated by APBS in PyMOL) with C33. (B) Close-up view of the interactions between the inhibitor and the protein. The PTP loop residues Cys 11, Thr12 and Arg17 (dark grey, sticks) are involved in ligand binding through hydrogen bonds. Colour code: the protein structure is represented as green cartoon, the PTP loop is shown in dark grey and the hydrophobic residues are depicted in red for an easily interpretation; the ligand is shown in light grey sticks and the hydrogen bonds in black dashed lines. The figures were prepared in PyMOL program.....45

**Figure 4.8** – Ligplot diagrams showing the protein-ligand interactions of top three azaindoles based on the energy score (A)- Azaindole 4; (B) - Azaindole 6n; (C) - Azaindole 12. Non-ligand residues involved in hydrophobic contacts are depicted in red and the formation of hydrogen bonds is shown by a green dashed line.....46

**Figure 4.9** – Predicted binding mode of azaindole 4 within the PtpA active site by blind docking. (A) Azaindole 4 pose into the active site pocket, a predominantly positively charged cavity. (B) The catalytic residues (sticks, dark grey) Thr12, Asn14, Ile15 and Cys16 establishing hydrogen bonds (dashed black lines) with the oxygen atoms of the ligand (stick, light grey). Surface charge distribution of PtpA generated by PyMOL APBS tools.....47

**Figure 4.10** – RMSD results of unbound PtpA (black lines) and PtpA-Azaindole 4 complex (red lines) through the MD simulations.....48

**Figure 4.11** – (A) SEC chromatogram obtained for PtpA in the absence of DTT. (B) SDS-PAGE gel analysis of the different fractions obtained from the SEC (fractions were identified according to the numeration of the chromatogram).....50

<b>Figure 4.12</b> – (A) SEC chromatogram obtained for PtpA in the presence of DTT. (B) SDS-PAGE gel analysis of the different fractions obtained from the SEC (fractions were identified according to the numeration of the chromatogram). Numbers 13-17 correspond to the collected fractions during the SEC run.....	50
<b>Figure 4.13</b> – Effect of glycerol and DMSO on PtpA thermal stability. The melting curve of 10% DMSO presents a lower $T_m$ reflecting that protein stability is noticeably affected. By the contrary, the presence of 10% glycerol significantly stabilizes the protein even when a similar concentration of DMSO is used.....	52
<b>Figure 4.14</b> – Representation of the melting curves from different storage conditions. Frozen PtpA stability is ensured (dark blue curve, with a higher melting temperature comparatively to protein stored at 4°C (light blue curve).....	53
<b>Figure 4.15</b> – Effect of 10% (A), 15% (B) and 20% (C) of PEG 400 (blue curves) and DMSO (orange curves) organic solvents concentrations in PtpA thermal stability. Protein stability is ensured in presence of PEG 400, in which positive melting shifts occurred towards the control (grey curves). By the contrary, protein stability is affected by the presence of DMSO as reflected by the negative-shifted curves.....	54
<b>Figure 4.16</b> – Melting curves of PtpA incubated with C33 compound at different concentrations. The data shows that C33 has a minimal increasing impact in PtpA thermostability with no significant increases in the melting temperatures values.....	56
<b>Figure 4.17</b> – Melting curves of PtpA incubated with Lap11 and Lap04 thiosemicarbazones at different concentrations. (A) Compared to the control (red curve), Lap11 does not display significant shifts in the $T_m$ values of PtpA. (B) TSA curves in the presence of Lap04 in which is a positive shift was detectable in the highest compound concentrations (blue and green curves).....	57
<b>Figure 4.18</b> – Melting curves of PtpA incubated with Azaindole 1 (A) and Azaindole 15 (B) at different concentrations. Both compounds do not affect the PtpA thermostability as significant $T_m$ shifts were not observed.....	58
<b>Figure 4.19</b> – C33 and Lap11 binding check results of MST experiment. (A) MST trace curves of C33 ligand incubated 20 minutes with 25 nM PtpA. (B) Amplitude response of C33 curves. (C) MST trace curves of Lap11 ligand incubation with PtpA (25 nM) for 20 minutes. (D) Amplitude response of Lap11 curves.....	62
<b>Figure 4.20</b> – Urea-polyacrylamide gel electrophoresis of ligand-free and incubated PtpA.....	65
<b>Figure 4.21</b> – Representation of some PtpA crystals obtained in co-crystallization trials. Needle-shaped PtpA crystals were obtained from (A) PtpA-C33 from 10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5; (B) PtpA-C37 from 10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5 (C) PtpA-Lap11 based on 20% PEG 8K, 0.1 M Phosphate Citrate 4.2, 0.2 M NaCl.....	66
<b>Figure 4.22</b> – PtpA-JE02 crystals obtained by co-crystallization, using the crystallization condition 50% PEG 200, 0.1 M Phosphate 6.2, 0.2 M NaCl. The diffraction experiment revealed that the crystal analyzed was salt rather than protein.....	68
<b>Figure 4.23</b> – Needle crystals obtained from 10% PEG 8K, 8% ethylene glycol and 0.1 M HEPES 7.5 crystallization condition to perform the microseeding experiment.....	68
<b>Figure 4.24</b> – (A) Resulting PtpA-azaindole4 crystals from microseeding. (B), (C) Ligand-free protein crystals obtained by micro-seeding.....	69
<b>Figure 4.25</b> – Diffraction pattern of PtpA-Lap11 crystal (left) and PtpA-C33 crystal (right). Crystals diffracted up to 3.6 Å (PtpA-Lap11) and 2.8 Å (PtpA-C33) at beamlines I04 (Diamond) and ID23-2 (ESRF), respectively.....	70
<b>Figure 4.26</b> – Representation of PtpA-Lap11 crystal structure. A disulfide bond is noticeable established between Cys53 of molecule A (represented in blue cartoon) and Cys53 of molecule B (green, cartoon). Picture has been prepared using PyMOL.....	76

**Figure 4.27** – Crystal packing of PtpA-Lap11 structure. The distances between the positioned molecules in the crystal packing are indicative of a high solvent content. In red are represented the original asymmetric unit molecules. Figure prepared in PyMOL.....76

**Figure 4.28** – Structural representation of the electron density blob near to the protein allosteric site in the PtpA-Lap11 crystal structure. The blob was found only in one of the two molecules of the asymmetric unit (molecule B) next to the His120 residue. In blue is represented the  $2F_o-2F_c$  electron density map contoured at  $1\sigma$  and in green the  $F_o-F_c$  density map contoured at  $3\sigma$ . Figure prepared in PyMOL.....78

**Figure 4.29** – Representation of the PtpA-C33 crystal structure with visible density blobs in the active site region. The disulfide bond between Cys53 from both molecules is also depicted. Electron density maps  $2F_o-F_c$  (blue) and  $F_o-F_c$  (green) – are contoured at 1 and 3  $\sigma$ , respectively. The two asymmetric unit molecules A and B are represented as blue and green cartoons. Figure prepared in PyMOL.....79

**Figure 4.30** – Representation of the structures of the two fragments generated from C33 ligand.....79

**Figure 4.31** – Representation of the first modelling attempt with fragment 1 in the density blob at the active site in molecule A (**A**) and molecule B (**B**). The electronic density map  $2F_o-F_c$  (blue) is contoured at  $1\sigma$ . Ligand color code: ligand in stick in which carbon atoms in grey and oxygen atoms in red. Protein color code: protein in blue/green cartoon, active site residues in sticks, in which carbon atoms in blue/green, nitrogen in blue and sulphur in yellow. Figure prepared in PyMOL.....80

**Figure 4.32** – Representation of the fragment 1 modelled in the density blob at the active site in molecule A (**A**) and molecule B (**B**). Active site residues, namely Arg17, Ile15 in molecule A and Cys11 and Arg17 in molecule B are involved in hydrogen bonds with the methoxy groups of the ligand. The electronic density map  $2F_o-F_c$  (blue) is contoured at  $1\sigma$ . Ligand color code: ligand in stick in which carbon atoms in grey and oxygen atoms in red. Protein color code: protein in blue/green cartoon, active site residues in sticks, in which carbon atoms in blue/green, nitrogen in blue and sulphur in yellow. Figure prepared in PyMOL.....81

**Figure 4.33** – Representation of the second fragment of C33 compound in molecule A (**A**) and molecule (**B**) of the asymmetric unit. The fragment was modelled in the density blob and positive peaks are observable nearby the carbonyl group of the fragment in both molecules. The electronic density maps  $2F_o-F_c$  (blue) and  $F_o-F_c$  (green) are contoured at 1 and 3  $\sigma$ . Ligand color code: ligand in stick in which carbon atoms in grey and oxygen atoms in red. Protein color code: protein in blue and green cartoon, active site residues in sticks, in which carbon atoms in blue/green, nitrogen in blue and sulfur in yellow. Figure prepared in PyMOL.....82

**Figure 7.1** – Crystal systems and allowed space groups for protein molecules. Four types of unit cell: Primitive (P); Centered in the side (C), in the body (I), or in the face (F). The spheres in the Bravais Lattice represent the lattice points.....100

**Figure 7.2** – Plasmid pET28a map and cloning region provided by Professor Hernan Terenzi.....101

**Figure 7.3** – Ligplot diagrams of the main interactions established by the azaindole's family of compounds with PtpA protein.....104

**Figure 7.4** – TSA curves for C37 and R06 compounds.....107

**Figure 7.5** – Binding Check MST traces of (**A**) C37 (**B**) R06 (**C**) JE02 (**D**) Lap04 (**E**) Azaindole 1 (**F**) Azaindole 15 and (**G**) Azaindole 16.....108

# List of Tables

<b>Table 3.1</b> – Optimization trials tested for the crystallization condition of ligand-free PtpA, performed at 4°C and 20°C using 700 µL of precipitant solution in the reservoir.....	32
<b>Table 3.2</b> – Summary of the crystallization screens used for PtpA crystallization at different concentrations and temperatures.....	32
<b>Table 3.3</b> – Co-crystallization assays conditions with chalcones (C33, C37, R6 and JE02), thiosemicarbazones (Lap04 and Lap11) and azaindoles (1,15 and 16) by sitting-drop techniques.....	34
<b>Table 4.1</b> – Docking results of PtpA and chalcones and thiosemicarbazones. The values were predicted by AutoDock Vina and AutoDock.....	43
<b>Table 4.2</b> – AutoDock Vina and AutoDock docking results of the best three hits from the azaindoles' library with PtpA.....	46
<b>Table 4.3</b> – RMSD and potential energy analysis of the MD simulations of the complex under study.....	49
<b>Table 4.4</b> – Melting temperatures (°C) obtained from TSA experiments to assess the effect of glycerol and DMSO on PtpA stability.....	52
<b>Table 4.5</b> – $T_m$ values from two different storage conditions.....	53
<b>Table 4.6</b> – Melting temperatures (°C) obtained for PEG400 and DMSO experiments at 10% and 15% concentrations.....	55
<b>Table 4.7</b> – $T_m$ and $\Delta T_m$ values obtained for C33, Lap11, Lap04, Azaindole 1 and Azaindole 15 compounds. $\Delta T_m$ values were calculated considering the protein control in which no ligand was added to the protein.....	58
<b>Table 4.8</b> – MST binding check assay results regarding the complex (protein-ligand) with the respective signal-to-noise ratio and amplitude parameters.....	61
<b>Table 4.9</b> – MST binding affinity results for PtpA incubated with Lap 11 compound for 30 minutes (room temperature) and overnight (at 4°C, 20°C, and 37°C) with the respective signal-to-noise ratio and amplitude parameters.....	63
<b>Table 4.10</b> – Sum up of the crystallization conditions and diffraction results of PtpA co-crystallizations. The conditions were provided from the JCSG <sup>+</sup> screen at 20°C.....	67
<b>Table 4.11</b> – Crystallization conditions obtained from micro-seeding experiments.....	69
<b>Table 4.12</b> – Data collection statistics from the diffraction data corresponding to PtpA incubated with Lap11 ligand (left) and to PtpA incubated with C33 ligand (right). The values in parentheses correspond to the highest resolution shell.....	71
<b>Table 4.13</b> – Z-score and LLG values from rotation and translation functions obtained using the Phaser program for PtpA-Lap11 and PtpA-C33 datasets.....	73
<b>Table 4.14</b> – Refinement statistics for PtpA-Lap11 and PtpA-C33 structures using Low Resolution Pipeline (LORESTR) from CCP4 suite of programs.....	75
<b>Table 7.1</b> – $IC_{50}$ and $K_i$ values of chalcones and thiosemicarbazones compounds.....	99
<b>Table 7.2</b> - JCSG <sup>+</sup> Crystallization Screen (Molecular Dimensions).....	102
<b>Table 7.3</b> – AutoDock Vina and AutoDock docking results of the three families of compounds under study.....	106



## Abbreviations and Symbols

$\alpha$  – Wave phase

$\varepsilon$  – Molar Extinction Coefficient

$\lambda$  – Wavelength

$\Delta G$  – Binding free energy

$\rho(x,y,z)$  – Electron density

$I_{hkl}$  – Intensity of reflections

$\sigma I_{hkl}$  – Intensity associated error

$\theta$  – Angle between the incident wave and the crystal plane

$d_{hkl}$  – Distance between the planes in the crystal lattice

Å – Angstrom

BCG – Bacilli Calmette-Guérin

CCP4 – Collaborative Computational Project Number 4

CHES – Cyclohexylaminoethanesulfonic acid

COOT – Crystallographic Object-Oriented Toolkit

Cryo-EM – Cryo-Electron Microscopy

DLS – Diamond Light Source

DMSO – Dimethyl sulfoxide

DTT – 1,4-Dithiothreitol

*E.coli* – *Escherichia coli*

EDTA – Ethylenediamine Tetraacetic Acid

ESRF – European Synchrotron Radiation Facility

$F_{hkl}$  – Structure Factor

$F_{calc}$  – Calculated Structure Factor

$F_{obs}$  – Observed Structure Factor

FT – Fourier Transformer

$h,k,l$  – Miller Indices

HEPES – 4-(2 hydroxyethyl) 1 piperazineethanesulfonic acid

HIV – Human Immunodeficiency Virus

HMLW PTPA – Human Low Molecular Weight Protein Tyrosine Phosphatase A

HMW PTPs – High Molecular Weight Protein Tyrosine Phosphatases

HPLC – High-performance Liquid Chromatography

IMAC – Immobilized Metal Affinity Chromatography

IPTG – Isopropyl- $\beta$ -D-1-thiogalactopyranoside

ITC – Isothermal Titration Calorimetry

KDa – Kilodalton

$K_d$  – Binding Affinity

$K_i$  – Inhibition Constant

LB – Luria-Bertani  
 LLG – Log Likelihood Gain  
 LMWPTP – Low Molecular Weight Protein Tyrosine Phosphatase  
 OD – Optical Density  
 MAD – Multiple Wavelength Anomalous Dispersion  
 MD – Molecular Dynamics  
 MDR – Multi drug-resistant  
 MIR – Multiple Isomorphous Replacement  
 MR – Molecular Replacement  
 MST – Microscale Thermophoresis  
*Mtb* – *Mycobacterium tuberculosis*  
 NaCl – Sodium Chloride  
 NMR – Nuclear Magnetic Resonance  
 PDB – Protein Data Bank  
 PEG – Polyethylene glycol  
 PMSF – Phenylmethylsulphonyl fluoride  
 PtpA – Protein Tyrosine Phosphatase A  
 PtpB – Protein Tyrosine Phosphatase B  
 PTP loop – Protein Tyrosine Phosphate-binding Loop  
 PTPs – Protein Tyrosine Phosphatases  
 RMSD – Root Mean Square Deviation  
 Rpm – Rotations *per* minute  
 SEC – Size Exclusion Chromatography  
 SapM – Secreted Acid Phosphatase M  
 SAXS – Small Angle X-ray Scattering  
 SAD – Single wavelength Anomalous Dispersion  
 SDS-PAGE – Sodium dodecyl sulfate–polyacrylamide gel electrophoresis  
 SPR – Surface Plasmon Resonance  
 STD-NMR – Saturation Transfer Difference Nuclear Magnetic Resonance  
 TB – Tuberculosis  
 T<sub>m</sub> – Melting Temperature  
 TMAO – Trimethylamine N-oxide  
 Tris – Tris(hydroxymethyl)aminomethane  
 TSA – Thermal Shift Assay  
 V-ATPase – Vacuolar-type H<sup>+</sup>-ATPase  
 V<sub>M</sub> – Mathews Coefficient  
 VPS33B – Vacuolar Protein Sorting 33B  
 WHO – World Health Organization  
 XDR – Extensively drug-resistant



# CHAPTER 1

---

## **I**NTRODUCTION



## 1.1 Tuberculosis Epidemic

### 1.1.1 From Prehistory to the Present

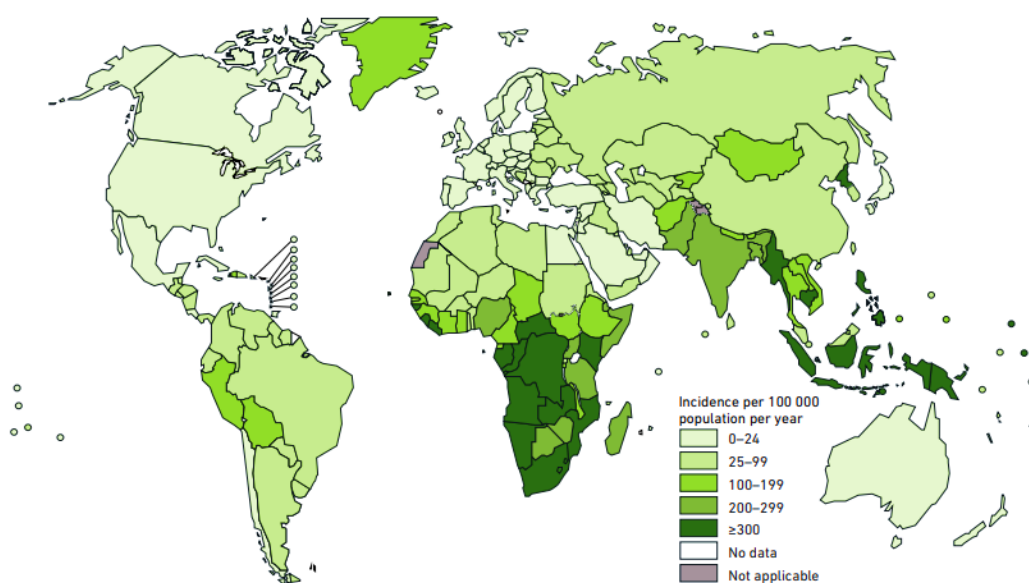
Tuberculosis (TB) is one of the biggest killers that plagued mankind causing disease, deformities and death since times immemorial. Traces of this ancient scourge were found among several mummies in Egypt civilizations and in European skeletons. In the Middle Ages, the epidemic expanded across the world and became a leading cause of mortality in the following centuries.<sup>1-4</sup>

Pulmonary tuberculosis is still the most common form of the disease and the term “*consumption*” was assigned meaning a “wasting condition accompanied by severe respiratory symptoms, sputum production and death – a body status consumed by the disease”.<sup>5,6</sup>

This mysterious disease reached epidemic proportions between the 18<sup>th</sup> and 19<sup>th</sup> centuries and became better understood when, in 1882, the German microbiologist Robert Koch demystified the cause of TB with his finding of *Mycobacterium tuberculosis* (*Mtb*) as the responsible agent.<sup>7,8</sup> This discovery was a scientific hallmark and together with other factors, such as the socioeconomics advances, the improved nutrition conditions and the social welfare, offered hope for the eradication of TB.<sup>1,5,8</sup>

Nowadays, TB remains a public health problem worldwide, particularly in developing countries (Figure 1.1) and, accordingly to the estimates of the Global Tuberculosis Report by World Health Organization, 10 million people were infected resulting in approximately 1.6 million deaths in 2018.<sup>9</sup>

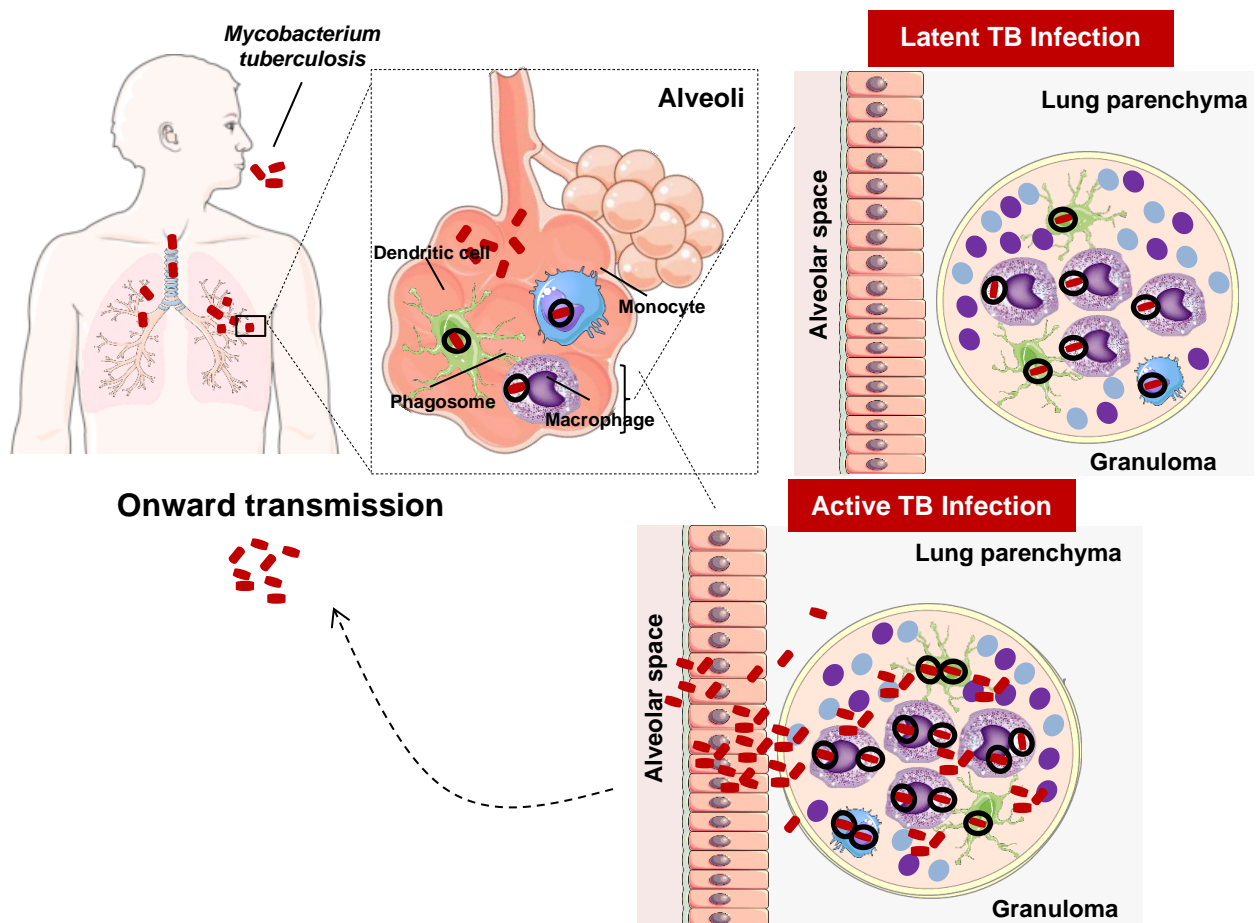
According to *Direção Geral de Saúde*, (DGS) report, Portugal had the higher percentage of reduction of TB cases when compared to the other European countries, in 2018. In the same year, the incidence rate was 15.4 cases per 100,000 habitants, with a total of 1,703 reported cases.<sup>10</sup>



**Figure 1.1** - TB burden countries worldwide and estimated incidence rates in 2017. South Africa and south-east Asia countries are the most affected by TB incidence. Map illustration adapted.<sup>9</sup>

### 1.1.2 Transmission, Risk Factors and Symptoms

This infectious disease is spread when people with pulmonary tuberculosis illness expel the bacteria through the air and consequently, people nearby may breathe and become infected.<sup>11</sup> When inhaled, the bacteria settles in the lungs and begins to multiply. However, not everyone infected with TB becomes sick and two TB-related conditions exist – latent and active TB infection (Figure 1.2).<sup>11,12,13</sup> In latent TB infection, the pathogen is contained within granuloma and remains inactive for a lifetime without causing manifestations of the disease.<sup>12,13,14</sup> In an active TB state, the pathogen becomes active and the infection is not controlled by the immune system of the host. People feel sick and reveal a set of symptoms, spreading the bacteria to others.<sup>11,12</sup>



**Figure 1.2** - Representation of the main types of TB conditions. In latent TB, *Mycobacterium tuberculosis* is inhaled reaching the first-line defense cells, the alveolar macrophages. The infected macrophages recruit other immune cells to attack, forming a granuloma, where the pathogen is contained in a latency state. The active TB infection occurs when infection overcomes the immune response. Consequently, the granuloma fails to contain the infection resulting in bacteria dissemination and re-entry into respiratory tract to onward transmission. Figure adapted.<sup>12</sup>

Among known risk factors, the probability of developing active TB is much higher among HIV infected individuals but is also attributable to other determinants such as undernutrition, type 2 diabetes mellitus, alcohol consumption and smoking habits. Furthermore, TB patients present a dynamic spectrum of symptoms including fever, persistent cough, fatigue, lack of appetite, weight loss and, in some cases, hemoptysis.<sup>1,5,6,15</sup>

## **1.2 *Mycobacterium tuberculosis***

### **1.2.1 Pathogen Characterization and Cell Wall Structure**

*Mtb*, commonly known as the “tubercle bacillus” in 1882 owing to Koch discovery, is the causative agent of TB. This nonsporulating mycobacterium is an obligate intracellular pathogen with a rod-shape, with no flagellum nor capsule.<sup>15</sup> *Mtb* classification as a Gram-positive or Gram-negative bacterium is a controversial issue, since the bacteria contains a thick layer of peptidoglycan which is structurally similar to Gram-positive, but also shares similar properties with Gram-negative, containing an outer layer of lipids.<sup>16,17</sup> The uncommon mycobacterial cell wall architecture is based on a peptidoglycan core attached to mycolic acids and its lipidic environment is an important feature responsible for resistance to drugs, which is essential for *Mtb* pathogenicity and virulence.<sup>18,19</sup>

### **1.2.2 Pathogenesis: Macrophage Manipulation and Phagosome Hijacking**

The *Mtb* remarkable capacity to cause the disease hinges upon a successful ability to challenge the immune system of the macrophage cell. The *Mtb* mechanisms to establish infection also rely in modulations of an array of signalling pathways, which consequently shut down critical processes, promoting its own survival.<sup>20,21,22</sup>

Once inhaled, the pathogen invades the pulmonary alveoli and infects the human alveolar macrophages, the first line of defense against microbe invasion. In a normal defense response mechanism, these macrophages engulf foreign bodies into phagosomes. These phagosomes are involved in membrane-fusion processes in endosomal pathways, allowing them to acquire antimicrobial properties by the direct recruitment of the vacuolar H<sup>+</sup>-ATPase pump (V-ATPase).<sup>22,23</sup> This pump transports protons across the membrane, creating a profoundly acidic lumen, which prompts the inhibition of bacterial invasion and enhances the beginning of the immune response.<sup>24</sup>

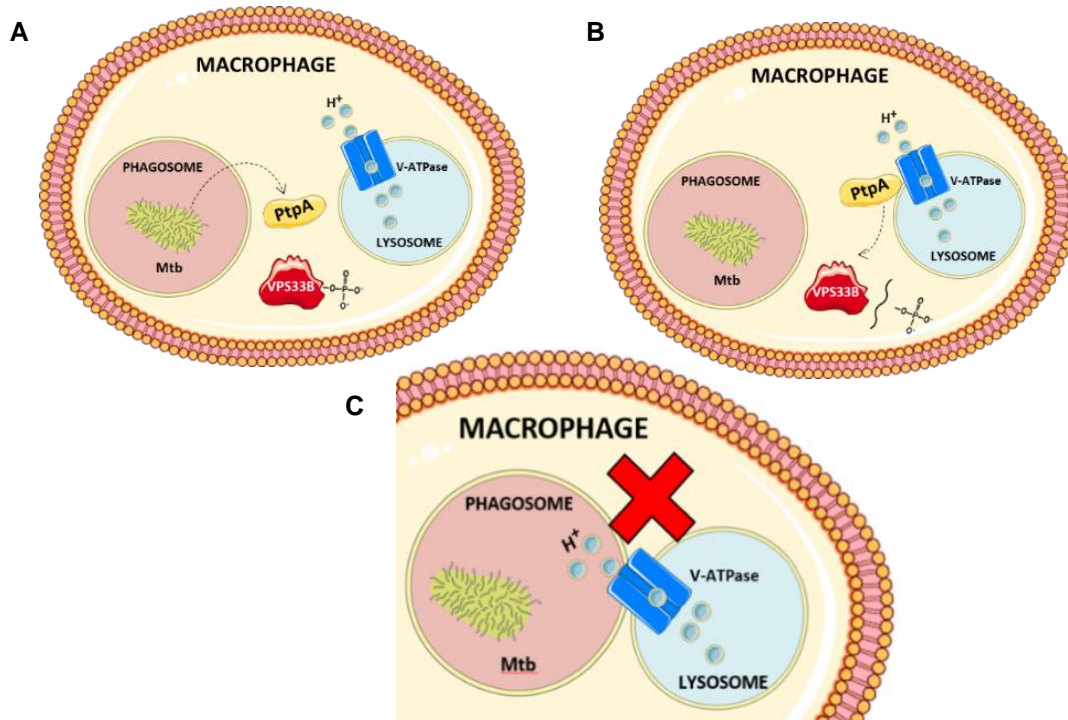
One of the strategies developed by *Mtb* to remain alive within the human host during macrophage infection is the secretion of phosphatases. In particular, the manipulation of the host phosphorylation pathways by protein tyrosine phosphatases (PTPs) is a crucial event in infection and pathogenesis of *Mtb*.<sup>25,26</sup> These PTPs play a key role in regulating physiological cellular processes such as metabolism, transcription, growth, migration and immune response.<sup>27, 28,29</sup>

## **1.3 Protein Tyrosine Phosphatase A**

### **1.3.1 PtpA Mechanism of Action**

Among the *Mtb*-secreted proteins, three distinct phosphatases – PtpA (Protein tyrosine phosphatase A), PtpB (Protein tyrosine phosphatase B) and SapM (Secreted acid phosphatase M) – are essential for *Mtb* pathogenesis. PtpA protein is the one that most contributes to its survival and pathogenesis and is secreted to the macrophage cytosol of the host.

PtpA binds to the subunit H of V-ATPase, in order to localize and dephosphorylate the substrate – the Vacuolar Protein Sorting 33B (VPS33B), which is involved in endocytic membrane fusion machineries (Figure 1.3 A, B). These key components of the endocytic pathway disruption mechanisms displayed by PtpA prevent the phagosome-lysosome membranes fusion and blocks the V-ATPase recruitment to the phagosome membrane, which consequently inhibits the phagosome acidification and maturation, enabling *Mtb* survival (Figure 1.3 C) <sup>26,30,31,32</sup>



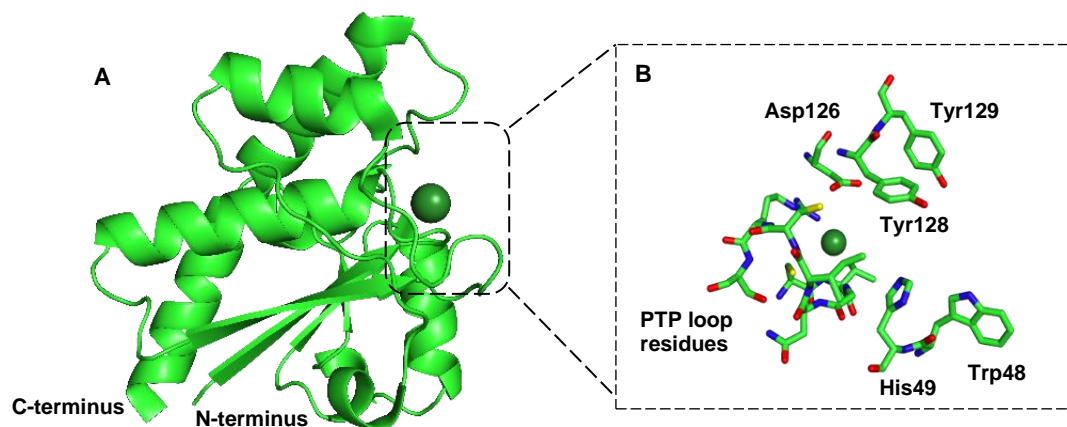
**Figure 1.3** - PtpA mechanism of action during *Mycobacterium tuberculosis* infection. **(A)** PtpA is secreted into the macrophage cytosol by *Mtb*. **(B)** The protein binds to subunit H of V-ATPase complex, localizing and dephosphorylating its substrate, the VPS33B protein. **(C)** Membrane fusion machinery is blocked and V-ATPase pump is excluded from mycobacterial phagosome.

### 1.3.2 PtpA Structural Characterization

PTPs reverse the effect of protein tyrosine kinases by removing the phosphate group from phosphorylated tyrosine residues in proteins, governing the balance of phosphorylation levels in important cellular signalling processes.<sup>27,28</sup>

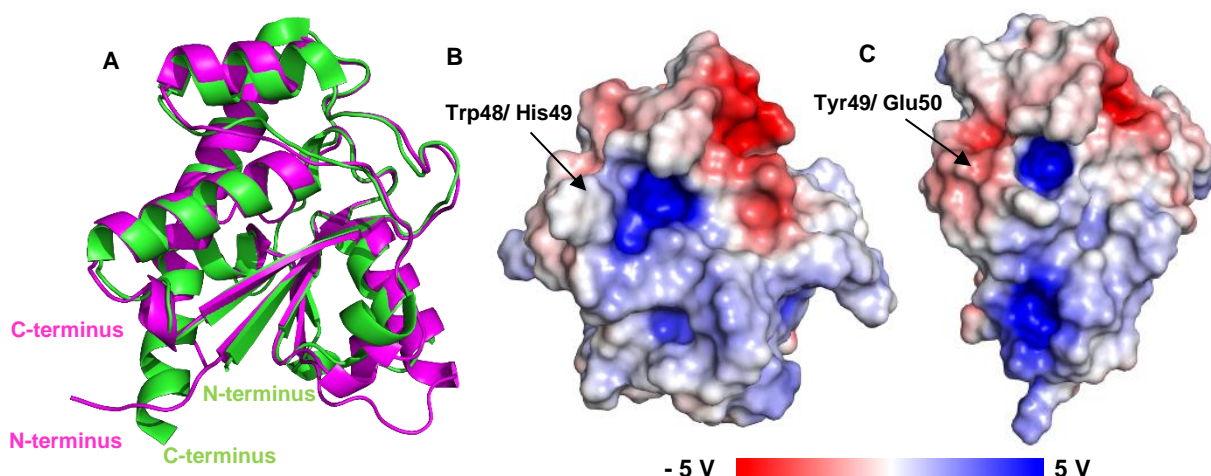
Generally, PTPs are divided into three main families: the high-molecular-weight, the dual-specificity and the low-molecular-weight family, respectively. Among these phosphatases families, PtpA is a low-molecular-weight protein tyrosine phosphatase (LMWPTP) with a molecular weight of 20 kDa.<sup>27,33,34</sup> The LMWPTPs display no sequence homology and no three-dimensional folding similarities with the high-molecular-weight phosphatases (HMWPTPs). However, they share a common catalytic mechanism due to a conserved active-site sequence motif C(X)<sub>5</sub>R(S/T) among all tyrosine phosphatases members. This active site motif, called protein tyrosine phosphate-binding loop (PTP loop), lies between the C-terminus of the first  $\beta$ -strand and the N-terminus of the first  $\alpha$ -helix.<sup>26,34,35</sup>

Two crystal structures of PtpA are deposited at 1.9 Å (PDB entry:1U2P) and 2.5 Å resolution (PDB entry:1U2Q). The protein has a single domain characterized by a central parallel four-stranded  $\beta$  sheets flanked by  $\alpha$  helices (Figure 1.4 A). The active site comprises the PTP loop (residues 11 to 18) and Trp48, Tyr128, Tyr129 residues, that form a hydrophobic crevice (Figure 1.4 B).<sup>26,34</sup>



**Figure 1.4 - (A)** Crystal structure of *Mtb* PtpA at 1.9 Å resolution (PDB:1U2P). **(B)** Overall structure of the active site which comprises the PTP loop (residues 11-18) and a group of hydrophobic residues that form the wall of the active site cavity. The structure exhibits a chlorine ion in the phosphate binding site. Figure performed using PyMOL software.

PtpA exhibits 38% sequence homology with the human low molecular weight protein tyrosine phosphatase A (HLMW-PTPA), mostly referred as HCPTPA (PDB entry: 5PNT), and the overall structural comparison reveals high structural similarities between them, with a root mean square deviation (RMSD) of 0.77 Å for the superposition of the C- $\alpha$  atoms (Figure 1.5 A). However, the pronounced surface charge distribution differences are perceptible by the analysis of the electrostatic surface, shown in Figures 1.5 B and 1.5 C. The substitution of Trp48 and His49 residues in *Mtb* PtpA by Tyr49 and Glu50 in HCPTPA creates a negatively charged patch near the active site, which appears to be a structural peculiarity important to consider in drug design approaches.<sup>26,34</sup>

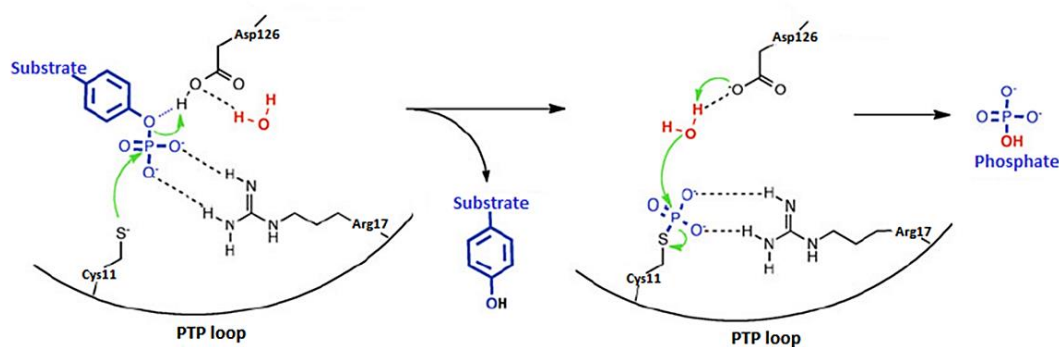


**Figure 1.5 - (A)** Superposition of *Mtb* PtpA (green, PDB:1U2P) and the human LMW-PTPA (purple, PDB: 5PNT) **(B)** Electrostatic potential surface of PtpA **(C)** Electrostatic potential surface of HCPTPA. Regions of different surface charge are depicted, in which the residues Tyr49 and Glu50 create a negative charge patch in the opening area of active site. Cartoon diagram and electrostatic potential surfaces performed by PyMOL software.



### 1.3.3 The Catalytic Mechanism

The commonly accepted catalytic mechanism of PtpA, represented in Figure 1.6, is based on the catalytic Cys11 residue that acts as the nucleophile and attacks the phosphate group bound to the tyrosine of the substrate protein. Simultaneously, the Asp126 residue donates a proton to the dephosphorylated tyrosine. This results in the formation of a phospho-cysteine intermediate that is immediately hydrolyzed by the same aspartic acid, which abstracts a proton from a water molecule. This induces the release of phosphate, restoring the catalytic Cys11. The Arg17 plays a key role in phosphorylated substrates binding, contributing with two hydrogen bonds with the phosphate oxygen atoms.<sup>27,34,36</sup>



**Figure 1.6** - The mechanism of the dephosphorylation reaction catalyzed by PTPases. The Cys11 residue of the PTP loop functions as the catalytic nucleophile. Release of the dephosphorylated substrate occurs following the transfer of a proton from an adjacent residue, Asp126. This residue acts as a base that activates a water molecule responsible for the hydrolysis of the cysteinyl-phosphate thioester bond, leading to the release of inorganic phosphate and restoring the Cys11. Adapted.<sup>36</sup>



## 1.4 Treatments and Major Challenges

### 1.4.1 Current Treatments: Antibiotics and Vaccine

In 1921, revolutionary developments by Albert Calmette and Camille Guérin led to the introduction of Bacilli Calmette-Guérin (BCG) vaccine and, years later, other scientific efforts contributed to the discovery and implementation of the first antibiotics. By the 1980s, the number of cases dropped, which led to consider TB disease as a forgotten disease all over the world.<sup>5,7,12</sup>

Currently, a six-month regimen for tuberculosis treatment comprises four drugs: isoniazid, rifampicin, pyrazinamide and ethambutol. These treatment regimen has several limitations such as lack of effectiveness, poor patient adherence, long-term toxicity, and high costs. In infants and children under 5 years old, the BCG vaccine protects from *Mtb* infection and control studies showed its effectiveness to be about 80%. In contrast, the efficiency in adolescents and adults remains uncertain.<sup>37,38</sup>

### 1.4.2 HIV Coinfection and Drug Resistance

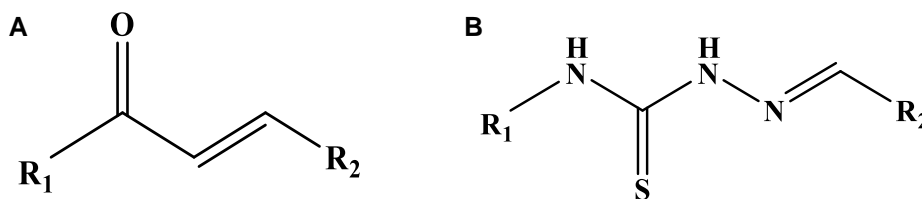
The attempt to control the disease globally is now confronted with two critical epidemiologic problems, which led to TB resumption. Firstly, the human immunodeficiency virus (HIV) comorbidity dramatically increase the susceptibility to the disease due to the weaker immune system of the victims, which makes them prone to develop TB. Secondly, the emergence of multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB) strains is a major concern that compromises the control of the disease, challenging the diagnosis and therapy.<sup>39,40</sup>

## 1.5 Targeting PtpA for TB Drug Development - Inhibitors

Since PtpA mechanism is crucial for pathogen survival within macrophages, there is a growing interest in finding compounds capable of inhibiting the action of this phosphatase. Selectivity is the major challenge for the development of inhibitors for this target, since the active site is highly conserved across the phosphatase family and structural features must be explored. Nevertheless, there is an evident lack of structural and biophysical studies to elucidate PtpA-inhibitors interactions.<sup>26</sup>

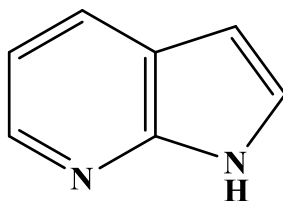
Extensive studies have been carried out in this research field and chalcones and thiosemicarbazones emerged as potential PtpA inhibitors that could be used in TB therapy. Chalcones are intermediate compounds in the biosynthesis of flavonoids and their  $\alpha$ ,  $\beta$ -unsaturated ketone system is a promising scaffold for PtpA inhibition. Those compounds have two groups that specifically contribute to the inhibition mechanism: the methoxyl group, which contributes through hydrogen bonds with catalytic residues; and the naphthyl group that establishes hydrophobic interactions with active site residues, promoting the inhibitor orientation and stabilization (Figure 1.7 A).<sup>41-44</sup> Furthermore, thiosemicarbazones (Figure 1.7 B), which are classified as imine derivatives, are also reported as an important class of non-competitive

inhibitors. The naphthyl moiety in R<sub>1</sub> and R<sub>2</sub> electron-withdrawing substituents enhance the inhibitory activity. These inhibitory properties display crucial insights in TB drug discovery.<sup>45</sup>



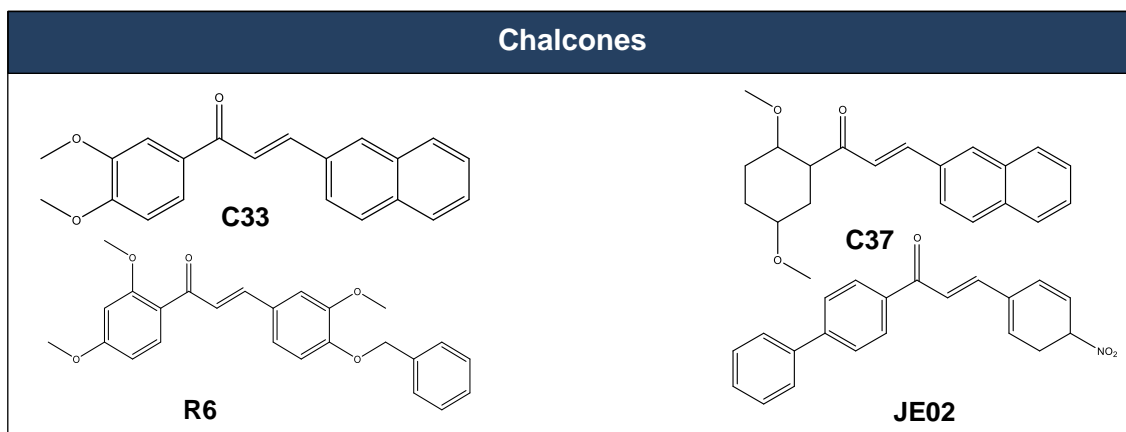
**Figure 1.7** - General chemical structures of a chalcone (A) and a thiosemicarbazone (B). Drawn in ChemDraw.

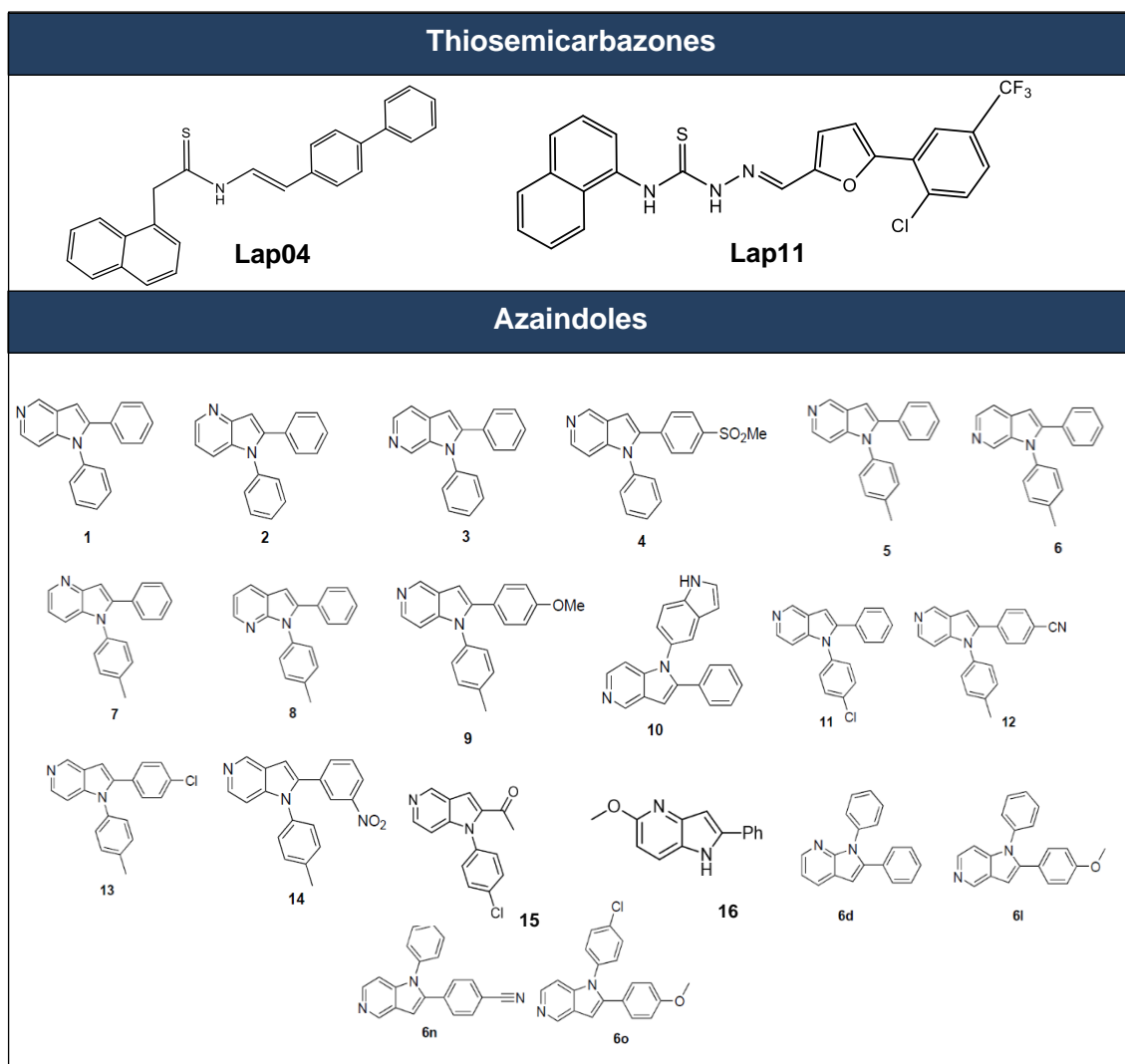
More recently, *in silico* studies with other protein tyrosine phosphatases, such as human PTPs, have identified azaindoles as suitable inhibitor candidates, prompting the exploration of these new class of compounds towards PtpA.<sup>46</sup> The privileged structure of azaindole (Figure 1.18), which is the combination of two heterocycles (the electron-rich pyrrole ring and the electron-lack pyridine ring), has attracted the interest of the scientific community for its physicochemical properties, as well as pharmacological properties, which have great potential in the area of medicinal chemistry.<sup>47,48</sup>



**Figure 1.8** - General chemical structure of an azaindole drawn in ChemDraw.

In this dissertation, the three families of compounds were studied and in Figure 1.9 are represented the respective chemical structures. Chalcones (C33, C37, R6 and JE02) and thiosemicarbazones (Lap04 and Lap11) were provided by Professor H ernan Terenzi (UFSC, Brazil) and supplementary information from published data regarding these compounds is available in Appendix 7.1. The azaindole-based library was supplied by Professor Maria Manuel Marques (LAQV, FCT-NOVA), from which 1, 15 and 16 compounds had been previously synthesized.



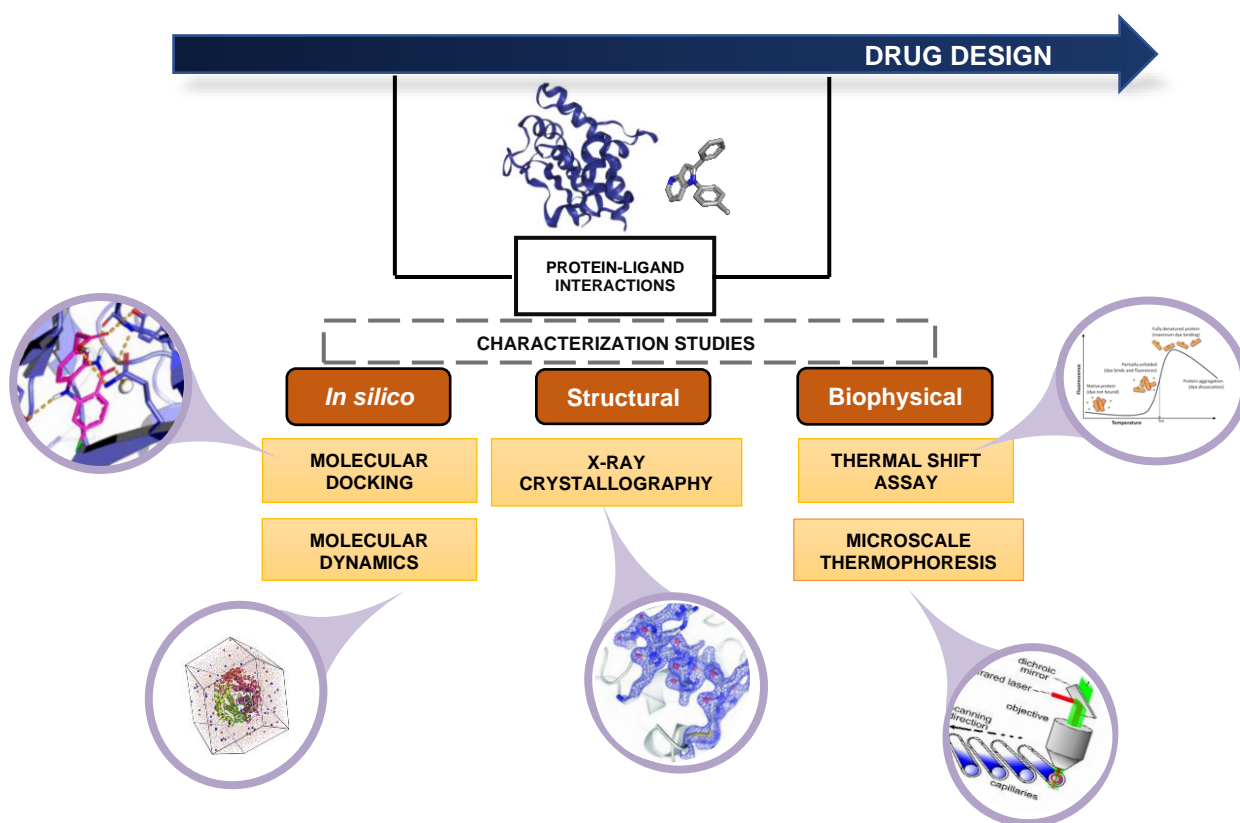


**Figure 1.9** - Chemical structures of the three families of compounds – chalcones and thiosemicarbazones with inhibitory effect to PtpA, and the library of available azaindoles compounds (1, 15 and 16 are synthesized). Structures drawn in ChemDraw.

## 1.6 Drug Design

Nowadays, drug design research is undergoing an exponential interest, allowing plentiful opportunities to develop several disease treatments. In drug discovery process, the biological targets are the most important motivation factors, since their characteristics and properties dictate the development of potential molecules. The understanding between receptor-drug interactions is an outstanding challenge in which is required a detailed knowledge of the backdrop of this association.<sup>49,50,51</sup>

In this thesis, a panel of *in silico*, biophysical and structural approaches were performed to systematically characterize the protein-ligand interactions (Figure 1.10). The structural information from X-ray crystallography continues to be a central and valuable source of breakthroughs in drug design, representing the main focus of this work.



**Figure 1.10** - Representation of the main characterization studies and techniques of protein-ligand interactions used in many drug design projects.

As aforementioned, crystallography technique undoubtedly has a key importance since provides structural information of receptor-ligand interactions and will be discussed in detail in the next chapter.

## 1.7 Biomolecular Crystallography

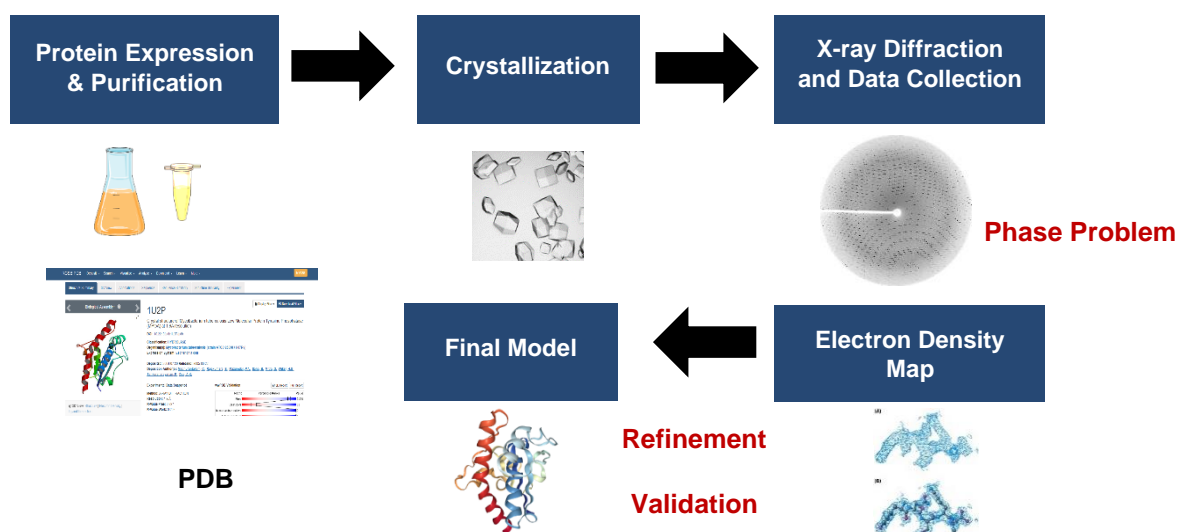
Proteins are essential biomolecules involved in several important cell events and, to a better understanding of the biological processes and their function, it is critical to study the proteins' structure. This three-dimensional structure information is a key point that will elucidate the function and dynamics of the biomolecules. In order to characterize and determine the structure of these bio-macromolecules, different techniques are applied for this purpose, namely Nuclear Magnetic Resonance (NMR), cryo-Electron Microscopy (cryo-EM) and X-ray crystallography.<sup>52</sup>

This last technique is still the most used for the structure determination of molecules and is strongly applied in protein-ligand interactions studies for drug design approaches.<sup>50</sup>

### 1.7.1 From X-rays to Structure

In the late of the 19<sup>th</sup> century, the scientist Wilhelm Röntgen bestowed to the scientific world his revolutionary discovery of the X-rays. The X-rays are high-energy electromagnetic radiation with a wavelength from 0.1 to 100Å, which is in the same range as interatomic bond distances (approximately 1Å). This type of radiation can be produced by shooting a metal target, usually copper (Cu) or molybdenum (Mo), with electrons produced by a heated filament and accelerated by an electric field.<sup>53,54</sup>

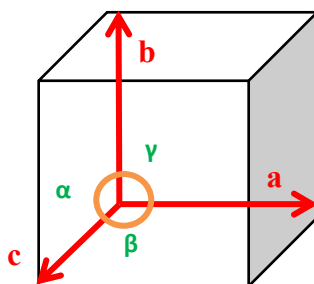
The principle of X-ray crystallography is based on the formation of crystals of biological macromolecules, so the determination of a three-dimensional structure through it is extremely dependent on the existence and on the internal order of monocrystals. Once properly organized in the crystal, these molecules are capable to diffract the X-rays producing a diffraction pattern, that can be used to calculate electron density maps. Figure 1.11 summarizes the most important steps in X-ray protein crystallography.<sup>53,54,55</sup>



**Figure 1.11** - Schematic representation of the main steps involved in structural protein determination. Prior to start the crystallization trials, it is necessary to express and purify the protein of interest using different molecular biology and biochemistry procedures. After obtaining suitable crystals for the x-ray diffraction experiment, the electron density maps are obtained which are used to build the initial model. This preliminary model is then refined several times and validated leading to the final model which is ready to be deposited in the PDB.

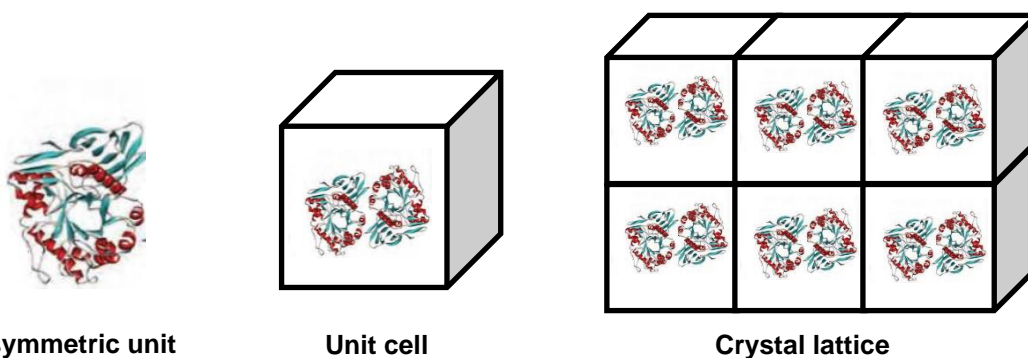
## 1.7.2 Protein Crystallization

A crystal can be defined as an orderly repetition in the three-dimensional space of identical building blocks, called unit cells. By definition, a unit cell is the smallest repeating unit that generates a crystal when translated into the 3 directions of space. Each unit cell is characterized by six parameters defined as unit cell constants: the lengths of the three axes ( $a$ ,  $b$  and  $c$ ) and the angles formed between them ( $\alpha$ ,  $\beta$  and  $\gamma$ ) as illustrated in Figure 1.12.<sup>54,56,57,58</sup>



**Figure 1.12** - Schematic representation of the unit cell in the crystal lattice. The unit cell constants: the three vectors -  $a$ ,  $b$  and  $c$  and the three angles -  $\alpha$  (between  $b$  and  $c$ ),  $\beta$  (between  $a$  and  $c$ ) and  $\gamma$  (between  $a$  and  $b$ ).

The unit cell consists of one or more asymmetric units that are defined as the smallest unit that generates a unit cell by applying symmetry operations. The crystal lattice is defined, the unit cell repeats. In Figure 1.13 is exemplified the relation between asymmetric unit, unit cell and crystal lattice.<sup>54</sup>

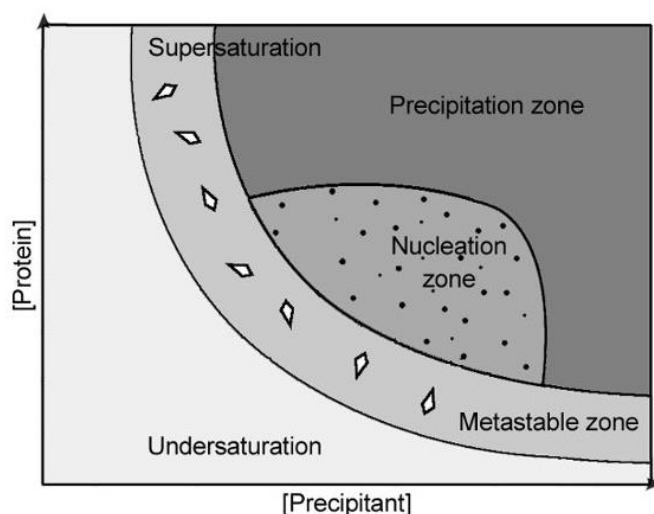


**Figure 1.13** - Illustrative representation of the assembly of unit cells in a three-dimensional crystal lattice. By symmetry operations, the asymmetric unit generates the unit cell, which in turn, through translation operations, generates the crystal lattice, which is repeated in an orderly manner in the three-dimensional space. Figure adapted.<sup>54</sup>

According to the geometry of the unit cell, 7 crystal systems can be generated. According to the lattice points that form the unit cell, there are 4 different types of unit cells (P, C, I and F), which combined with the 7 systems, lead to the 14 Bravais lattice (Appendix 7.2). These are the basis of 230 different space groups generated by the allowed rotation and translation symmetry operations of the asymmetric unit. However, as proteins are inherently chiral, and neither mirror planes nor inversion centers are allowed, and the number of possible space groups for these macromolecules is reduced to 65.

The limiting step in crystallography is obtaining suitable protein crystals for the diffraction experiment. In fact, it is quite difficult to predict the ideal crystallization condition that ensures the crystals formation. Consequently, many crystallization screening experiments with different precipitating agents or pH values are usually carried out at different temperatures and using different crystallization methods. The crystallization step is often the more time-consuming and without good diffracting crystals is very difficult to solve the protein structure.<sup>54</sup>

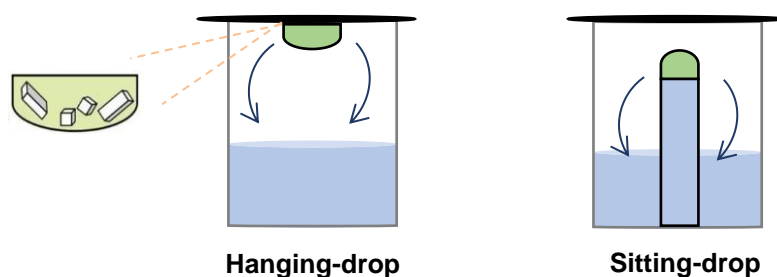
Different methodologies are used in protein crystallization, but the principle that needs to be fulfilled is to decrease the protein solubility to achieve a supersaturated state of the protein. The crystallization diagram (Figure 1.14) is useful for the better understanding of the crystallization process.



**Figure 1.14** - Protein crystallization diagram. In diagram are represented two distinct regions: undersaturation and supersaturation. The supersaturation zone is divided in metastable zone, nucleation zone and precipitation zone. The soluble proteins molecules will move from the undersaturation zone into nucleation zone where nuclei are formed, which will slowly and orderly grow to form crystals in the metastable zone. Adapted.<sup>54</sup>

One widely used crystallization method is the vapor diffusion, in which a protein/precipitant mixture is allowed to equilibrate over a closed reservoir that contains precipitant solution at higher concentrations. The siliconized glass slide covers the reservoir and the protein/precipitant concentrations in the drop slowly increase by water transfer to the more concentrated solution in the reservoir. The protein concentration in the drop increases once the system reaches equilibrium, which may lead to the formation of a protein crystal.

In the vapor diffusion method, the crystallization setup can be performed by two different techniques – Hanging and Sitting drop, as depicted in Figure 1.15. The protein/precipitant solution can be placed in the siliconized cover slip – the Hanging-drop technique – or in a support inside the reservoir – the Sitting-drop technique.<sup>54,56</sup>



**Figure 1.15** - Vapor diffusion method and the hanging and sitting-drop techniques, respectively.

As previously referred, getting suitable protein crystals for the diffraction experiment is a time-consuming step. Many factors influence the crystallization process, such as protein purity, the presence of ligands, temperature variation and pressure conditions. Other important factors that influence the success of the protein crystallization are chemical factors: pH, precipitant type, protein and precipitant concentration. Furthermore, first trials testing crystallization screenings also determine the proper crystallization conditions.<sup>57,58</sup>

An innovative step and great improvement in X-ray crystallography happened with the introduction of the first crystallization robots, enabling to test a great variety of crystallization conditions with a small amount of protein, becoming a high throughput process. However, the main problems in using the crystallization robots is that the crystals are frequently quite small, sometimes not suitable for the diffraction experiment, and the scale-up can be a difficult process. Once a protein crystal is obtained, the next step is the diffraction experiment, which consists on focusing an X-ray beam in the crystal.<sup>57,58</sup>

### 1.7.3 Data Collection from Crystals

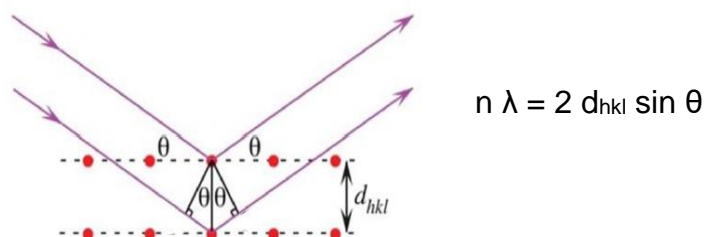
Protein crystals have a high solvent content (20-80%) which is reflected in their fragile nature. Some procedures are adopted before the diffraction experiment and data collection. The newly crystals can be transferred to a harvesting buffer (similar to the precipitant solution but with higher precipitant concentration). Posteriorly they should be transferred to a suitable cryoprotectant solution, usually prepared by adding 20-30% (m/v) of glycerol or using other cryoprotecting agents as paratone. These procedures prevent crystal degradation and avoid ice formation.<sup>59</sup>

After the successful crystallization process and the procedures mentioned above, the diffraction experiment is carried out by focusing an X-ray beam on the crystal. The X-rays interact with the electrons of the protein molecules and then are scattered by these electrons originating a diffraction pattern with different reflections (spots), derived from the crystal's internal organization.<sup>53,54</sup>

The phenomenon of X-ray diffraction was explained by Lawrence L. Bragg, who formulated a physical model to explain conditions where diffraction was observed. During the diffraction experiment, X-rays penetrate the crystals and reflections are extinguished by interference. Considering  $\lambda$  as the wavelength of the X-rays,  $d$  as the perpendicular distance

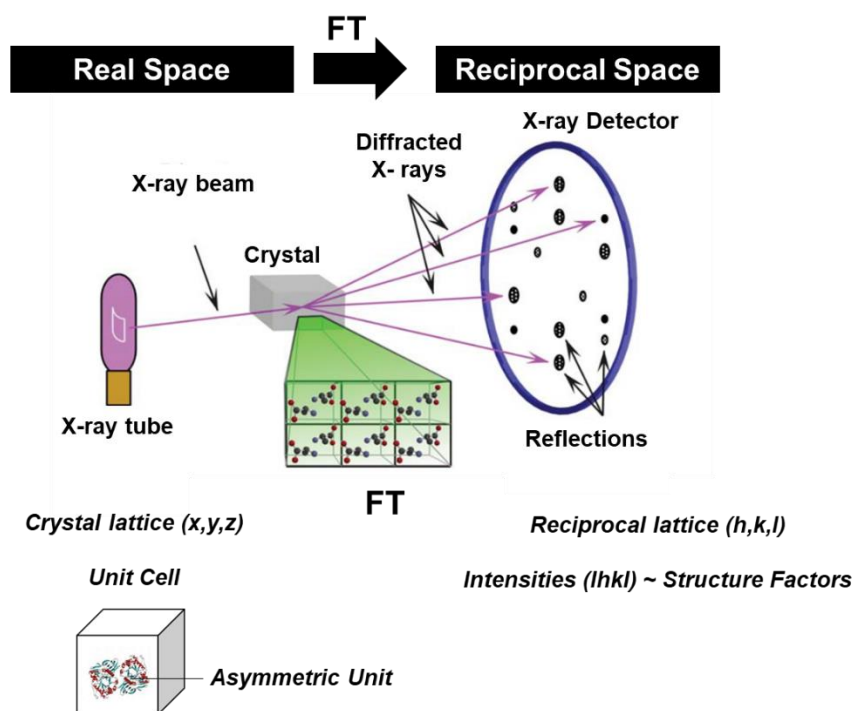


between the successive planes,  $\theta$  as the angle between the planes and the incident or reflected X-ray, and  $n$  as an integer, then reinforcement of waves occurs only when equation in Figure 1.16 is obeyed - the Bragg's Law. If the scattered waves from X-rays are in phase, they interfere constructively, and Bragg's law is obeyed leading to a diffraction pattern. However, if the waves are out of phase, they are cancelled out by destructive interference, and there is no diffraction (Figure 1.16).<sup>52,53,54,</sup>



**Figure 1.16** - Bragg's law geometric representation (left) and equation (right). The Bragg's law can be graphically interpreted allowing the understanding of an X-ray experiment as the reflection on a set of imaginary planes in the crystal. Adapted.<sup>54</sup>

Each reflection can be treated like a wave, and its phase and intensity contain the necessary information to determine the molecular structure of the protein. There is an inverse relationship between the space of unit cells in the crystal (crystal lattice) and the space of reflections in the detector (reciprocal lattice). In the crystal (real space), the position of an atom in the unit cell can be defined by a set of spatial cartesian coordinates  $(x,y,z)$ , while in the diffraction pattern (reciprocal space), a reflection position is given by the Miller indices  $(h,k,l)$ , as schematically represented in Figure 1.17.<sup>52</sup>



**Figure 1.17** - Schematic representation of the X-ray diffraction experiment. An intense X-ray beam is directed at the protein crystal (a three-dimensional array of repeated and ordered units). The X-rays are diffracted by the atom's electrons and are recorded in a detector.

The structure factor ( $F_{hkl}$ ) is a function that describes the X-ray wave responsible for a given reflection in the diffraction pattern being the sum of the diffraction of all the electrons in the unit cell. Applying an inverse Fourier transform (FT), the electron density can be defined. This operation allows commuting from the reciprocal space (structure factor) into the real space (electron density), as the FT describes the relation between an object and its inverse.

$$F_{hkl} = \int_V \rho(x, y, z) e^{[2\pi i(hx+ky+lz)]} dV$$

$\begin{array}{c} \downarrow \text{FT}^{-1} \quad \uparrow \text{FT} \end{array}$

$$\rho(x, y, z) = \sum_{h,k,l} |F_{hkl}| \cdot e^{-2\pi i(hx+ky+lz-\alpha hkl)}$$

Thus, as aforesaid, from the structure factors  $F_{h,k,l}$  it is possible to calculate an electron density map ( $\rho$ ), which can be used to determine atomic coordinates ( $x, y, z$ ) of all atoms in the real space. By the analysis of the previously referred formula, the electron density can be calculated by knowing two parameters: the amplitude of  $F_{h,k,l}$  ( $|F_{hkl}|$ ) and the phases of the waves that reach the detector with constructive interference. The amplitudes  $|F_{hkl}|$  are experimentally determined, proportional to the square root of the intensity of the reflections, however, there is no information available regarding the phase angle ( $\alpha hkl$ ) that reaches the detector, which would enable us to calculate the electron density map directly from the equation, leading to the so-called “phase problem” in crystallography.<sup>56,62</sup>

#### 1.7.4 The “Phase Problem” and Structure Solution

To overcome the “phase problem”, different methods have been designed in protein crystallography to obtain the missing phases.

Regarding the increasing number of structures available in the PDB nowadays, the most common method used to solve the “phase problem” is Molecular Replacement (MR). On the other hand, in cases where there are no homologous structures, the inclusion of heavy atoms and anomalous dispersion, allows the determination of the phases through the conventional methods named Multiple Isomorphous Replacement (MIR) and Multiple/Single Anomalous Dispersion (MAD/SAD), respectively.<sup>53,54,56</sup>

To estimate the initial phases in MR, a known model structure is used, usually with 30% or higher sequence identity to the protein of interest. This methodology requires finding the position of the protein of interest in the asymmetric unit, based on rotation and translation functions, using the search model. In the rotation step, the search model is properly oriented in the unit cell, which is achieved by calculating and superimposing Patterson maps of both search model and protein. Once appropriately oriented, the model is positioned in the unit cell by comparing the structure factors of both model and protein.<sup>53,54,61</sup>

The two crystal structures described in this dissertation were solved using molecular replacement with a similar protein as search model.

### 1.7.5 Model Building, Refinement and Validation

Once the “phase problem” is solved, an electron density map can be calculated, and model building and refinement are the next steps prior to obtain a final model.

The electron density map can be optimized by building the model and calculating new improved phases through new FT calculations. This is an interactive process leading to an improvement of the phases and consequently of the electron density maps, allowing to obtain a model that conveniently explains the experimental data.<sup>54,52</sup> The initial model present errors that are eliminated during refinement process in which differences between the observed  $|F_{obs}|$  and calculated  $|F_{calc}|$  amplitudes should be minimized.

As the word refinement refers to the adjustment of model parameters, some refinement quality parameters were established.<sup>54,52,62</sup> One of the most important parameters is the R-factor which quantify the agreement between the constructed model and the experimental data, according with the following formula:

$$R = \frac{\sum |F_{obs}| - |F_{calc}|}{\sum |F_{obs}|}$$

R-factor values should be as small as possible to indicate the agreement between the calculated and observed amplitudes. High R-factor values indicate significant differences and, most likely, a correct solution has not been found. With a proper model refinement, the R-factor progressively decreases, and it is accepted that values around/below 20% point for a global satisfactory protein model.<sup>62</sup>

The  $R_{free}$  parameter is also important and used to measure the agreement between the experimental data and the final model. It corresponds to a little fraction of data (5-10% of the reflections) which is not used for the refinement process. This gives an unbiased agreement measure. R-factor and  $R_{free}$  should have the same behaviour, meaning that a decrease of R-factor must be followed by a decrease of  $R_{free}$ , otherwise errors are being introduced in the model. Therefore, for medium/high resolution data, it is expected that the difference between R-factor and  $R_{free}$  should not exceed 5%.<sup>52,54,55</sup>

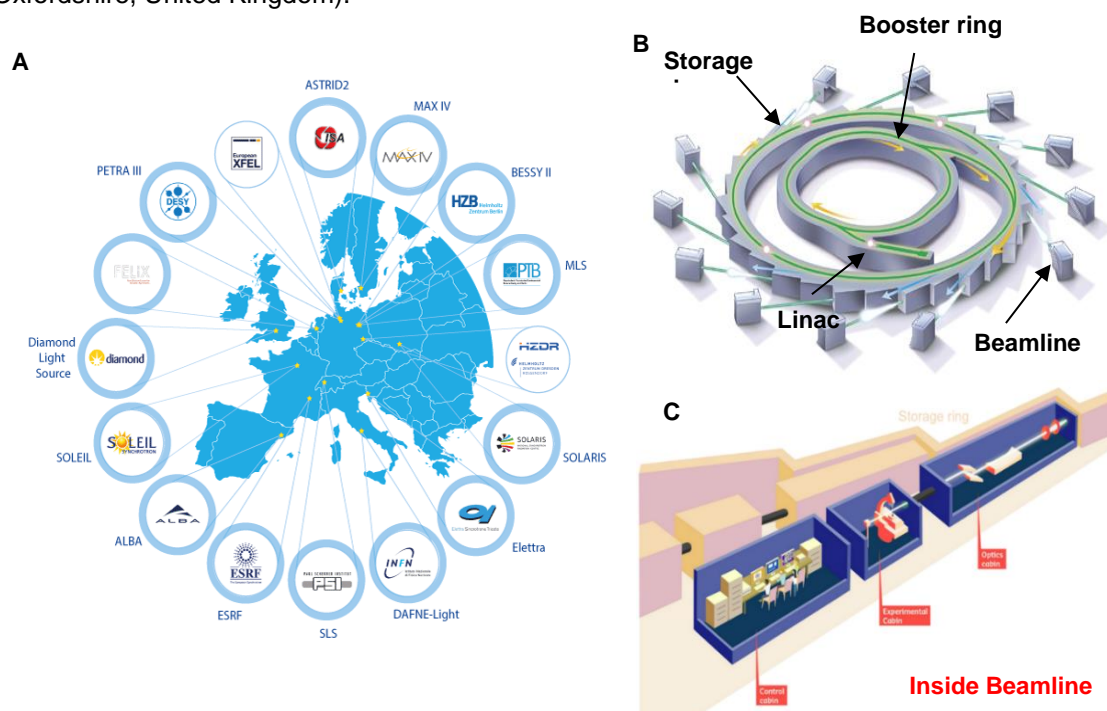
Both R-factor and  $R_{free}$  parameters uniquely describe the global errors found in the model, not reflecting putative local errors. The B factor is used for such purpose and is a parameter related to the thermal vibration of an atom around its position. If an atom is not properly located into the electron density, the B factor value is high. Additionally, atoms located on the protein surface and in areas with greater mobility usually have higher B-factor values than those found inside it.

Once the electron density is greatly explained by the model, the refinement process is completed. The validation proceeds, using several criteria to evaluate the results of structural analysis and then the final model can be deposited in the Protein Data Bank (PDB) database.<sup>54,62</sup>

### 1.7.6 Synchrotron Radiation

The introduction of synchrotron radiation facilities was a remarkable and revolutionary step in X-ray crystallography. They are widely spread worldwide, with several facilities installed in Europe (Figure 1.18 A). As represented in Figure 1.18 B, a synchrotron works as a storage ring, where electrons move in a circle, almost at the speed of light, through magnetic fields, generating intense X-rays. These resulting X-rays are emitted as thin beams, each directed toward a beamline next to the accelerator, and are used for diffraction experiments (Figure 1.18 C). Using synchrotron radiation is advantageous when compared to *in house* sources, since it produces highly intense and bright beam, the wavelength is tuneable, the data collection process is shortened to a few minutes and particularly important, the data quality is enormously better.

The crystal structures described in this dissertation were obtained at the European Synchrotron Facility (ESRF, Grenoble, France) and at the Diamond Light source (DLS, Oxfordshire, United Kingdom).<sup>55,63</sup>



**Figure 1.18** – (A) European synchrotrons and X-Ray Free-Electron Laser facilities. (B) Schematic representation of a synchrotron structure. (C) Illustration of a beamline, a physical and dynamic space supplied by equipment that brings X-ray beam to the sample in study. Adapted.<sup>63</sup>

# CHAPTER 2

---

## **O**BJECTIVES



Since the protein tyrosine phosphatase A inhibition attenuates the growth of *Mtb*, this phosphatase becomes a suitable biological target for the development of new anti-tuberculosis drugs. Therefore, the central focus of this dissertation is to study the interaction of PtpA with a set of selected compounds with putative inhibitory activity: chalcones, thiosemicarbazones and azaindoles. In order to systematically characterize the protein-inhibitor interactions, an integrated panel of computational, structural and biophysical drug design approaches will be used contributing to find new efficient and selective inhibitors towards PtpA as viable anti-TB drugs.

Within the proposed goals, it is firstly necessary to overexpress PtpA in *Escherichia coli* system and purify the protein using affinity chromatography, based on a well-established protocol, to obtain large amounts of pure protein for further structural and biophysical studies.

From a computational standpoint, molecular docking calculations will be applied to obtain structural information regarding the binding mode of a set of specific and selective compounds reported in literature – chalcones (C33, C37, R6 and JE02) and thiosemicarbazones (Lap04 and Lap11). In parallel, the exploration of a new azaindole-based family (20 compounds library) will be carried out in order to evaluate the binding mode towards PtpA protein and to identify possible alternative binding sites. Molecular dynamics simulation will be made with the best azaindole candidate from molecular docking, to provide a detailed information about the numerous dynamic processes that occur in a biological system, which will allow, in this case, the study of the protein and complex behaviour, and its properties over time.

The three families of compounds will be also used to evaluate their putative interactions with PtpA by a large range of biophysical techniques namely Thermal Shift Assay (TSA) and Microscale Thermophoresis (MST). Urea polyacrylamide gel electrophoresis will be performed to detect the potential conformational changes of the protein upon ligand interaction and Size Exclusion Chromatography (SEC) will be used to confirm the protein oligomeric state.

Finally, a structural approach will be followed to confirm the obtained computational results. Thus, X-ray crystallography will be used to characterize, at the atomic level, the interactions of PtpA with the mentioned inhibitors. Several crystallization assays, using several commercial screens, will be conducted to optimize previously found crystallization conditions and to find new ones in order to improve the resolution of diffraction data. Different and innovative approaches to obtain suitable crystals will also be explored such as co-crystallization and micro-seeding experiments. The crystals will be analysed using Synchrotron radiation and the structures will be determined by Molecular Replacement. Different crystallographic software will be used for model building and refinement steps, in order to elucidate the binding mechanism of the inhibitors to PtpA.

This multidisciplinary drug design project allows to deal with several techniques at the interface of structural and molecular biology, with the goal to solve a relevant and worldwide biological problem: the tuberculosis disease.





# CHAPTER 3

---

## **M**MATERIALS & **M**METHODS



The experimental expression and purification procedures were based on the protocols provided by the collaboration with Professor H ernan Terenzi from Federal University of Santa Catarina, Brazil, where the molecular biology component of this project was carried out, particularly the strain and plasmid choices. Media composition and the growth conditions were optimized at the Protein Crystallography group, previously to this work.

### 3.1 PtpA heterologous expression

In order to express the recombinant PtpA protein, the plasmid pET28a (Appendix 7.3) was transformed into *Escherichia coli* BL21(DE3) cells. These cells, harboring the recombinant plasmid, were inoculated in 50 mL Luria-Bertani (LB) medium containing 50  g/mL of kanamycin (NZYtech) and incubated overnight at 37 C and 200 rpm (Orbital Shaker-Incubator ES-20, Grant Bio). Afterwards, 12 mL of the cells culture were transferred to 2 L Erlenmeyer flasks containing 500 mL LB medium supplemented with 50  g/mL kanamycin, which were incubated at 37 C and 180 rpm for 2 hours (Shaker Gallenkamp). Protein expression was induced when the optical density at 600 nm (OD<sub>600</sub>) values were between 0.6 and 0.8 with 500  M IPTG (Isopropyl- -D-1-thiogalactopyranoside) (NZYtech) and then incubated during 4 hours at 30 C and 150 rpm (Shaker Gallenkamp). The cells were recovered by centrifugation in an Avanti J-26 XPI centrifuge (Beckman Coulter, JA-10 rotor) at 10000xg, 6 C for 20 minutes. The supernatant was discarded and the obtained pellets were stored at -80 C.

### 3.2 PtpA Purification

The harvested cells were resuspended in lysis buffer A (20 mM Tris-HCl 8.0, 10 mM Imidazole, 500 mM NaCl, 5 mM DTT) containing 40  g/mL of protease inhibitor PMSF (phenylmethylsulphonyl fluoride, Fluka). The cells were then disrupted by sonication (UP100H MS7, Hielscher Ultrasonics) during 10 cycles of 1 minute (0.5 cycle and 80% amplitude) with 1 minute of homogenization between each cycle. The resulting extract was centrifuged for 30 minutes (Centrifuge 5804 R, Eppendorf) at 12000xg and 6 C to separate the soluble and insoluble fractions and the supernatant (soluble fraction) was recovered for the subsequent purification steps.

Since PtpA carries a hexa-histidine tag at the N-terminus, the protein purification was based on Immobilized Metal Ion Chromatography (IMAC) using a 5 mL Ni<sup>2+</sup> column (HisTrap<sup>TM</sup> HP, GE-Healthcare Bio-Sciences) equilibrated with buffer A. This column was coupled to a HPLC AKTA Start apparatus (GE-Healthcare) and the UNICORN Start 1.0 program was used to monitor the parameters of the purification process. After equilibrating the system with buffer A and with the elution buffer B (20 mM Tris-HCl 8.0, 500 mM Imidazole, 500mM NaCl, 5 mM DTT), the soluble fraction was loaded into the nickel column. The column was washed with 20 mL of buffer A removing the proteins with no affinity to the Ni<sup>2+</sup> at 2 mL/min flow. The protein of interest was eluted by a gradient from 12 to 100% using the elution buffer B and 1 mL protein fractions were collected.

The purity of the protein fractions was assessed by a 10% SDS-PAGE, in which 20  $\mu\text{L}$  of mixture of sample (15  $\mu\text{L}$ ) and loading buffer (5  $\mu\text{L}$ ) were loaded, previously heated at 100°C for 5 minutes. After a good separation of the bands, the gel was stained with Coomassie Blue (Fluka).

The fractions containing purified recombinant protein were pooled together followed by a desalting step in which the buffer of the sample was exchanged into buffer H (100 mM HEPES 7.0, 150 mM NaCl, 5 mM EDTA, 5 mM DTT and 10% Glycerol) For that purpose, three coupled HiTrap™ (GE Healthcare Bio-Sciences) desalting columns, previously equilibrated with the buffer H, were used. Then, the protein was concentrated by ultrafiltration for 10 minutes cycles at 3500xg and 6°C, using a 10 kDa pore membrane (Centricon Vivaspin® Turbo 15, Sartorius) previously equilibrated with buffer H. The protein concentration was calculated using the Plate Reader SpectraMax190 apparatus, considering a theoretical molar extinction coefficient ( $\epsilon_{280\text{nm}}$ ) of 15470  $\text{M}^{-1} \text{cm}^{-1}$ .

### **3.3 *In Silico* Characterization Studies**

#### **3.3.1 Molecular Docking**

Initially, two indispensable procedures must be performed: the protein and ligand preparation steps. Firstly, the crystal structure of PtpA was retrieved from Protein Data Bank (PDB entry code: 1U2P) for the protein preparation. Afterwards, using the Molecular Graphics Laboratory (MGL) Tools program <sup>64</sup>, several procedures were performed to achieve a successful protein preparation including the addition of polar hydrogens, the non-polar hydrogens merging, the assignment of Kollman charges and the conversion into PDBQT format file. Afterwards, the ligand molecules were drawn in ChemDraw version 18.0 program and the CORINA Classic online version software was used to get the three-dimensional structure of the inhibitors in a MOL format file. Posteriorly, this MOL format file of the ligand was submitted to the next step: the ligand preparation in MGL Tools. This step was achieved by adding Gasteiger charges, adding polar hydrogens and merging non-polar hydrogens and was saved in the PDBQT format file.

Once protein and ligand preparation were completed, the docking calculations were performed using two programs: AutoDock Vina 1.1.2 and AutoDock 4.2.<sup>65,66</sup> In AutoDock Vina, a configuration file was created, composed by receptor (protein) and ligand information, box coordinates and 100 binding modes settled (to increase the accuracy in the search). This configuration file possesses the necessary information and consists in the input file that is used to run the AutoDock Vina program. Afterwards, the program originated 20 binding modes of the ligand and each binding mode was analyzed individually. The choice of the best binding mode lied in the most negative binding free energy ( $\Delta G$ ) value.

In order to perform the AutoDock calculations, some crucial preliminary steps must be performed. Using Autogrid (a running tool from AutoDock), the receptor grid map (cubic box) with X=,112 Y=116, Z=116 number of points, grid spacing of 0.375 and grid centered at X=8.1, Y=6.47, Z=22.02 was generated. From the results, the best protein-ligand complex was selected based

on three parameters analysis: the estimated binding free energies ( $\Delta G$ ), in which the most negative value was considered the best, the estimated inhibition constant ( $K_i$ ) and the highest number of populations. All docking results were analyzed and compared using MGL Tools, PyMOL 2.3.2<sup>67</sup> and LigPlot+ 1.5.1<sup>68</sup> softwares.

### 3.3.2 Molecular Dynamics Simulation

Molecular Dynamics (MD) simulations were performed using Gromacs software version 5.05<sup>69</sup>. The three-dimensional structure of physiological PtpA was modelled based on the PDB template (PDB:1U2P).

The ligand-free PtpA MD simulation was performed using OPLS-AA all atom force field<sup>70</sup> and the cubic box parameters were defined as default. 9123 water molecules were generated, and 4 sodium ions were added to balance the global net charge. After the energy minimization, temperature and pressure equilibration of the systems, the MD simulations were carried out for 30 ns, at physiological temperature (37°C) and atmospheric pressure (1.0 bar).

The MD simulations of the complex PtpA-azaindole4 were performed using the GROMOS96 54a7<sup>71</sup> force field. The topology of azaindole 4 was generated using the ProdrG server. After solvent generation of 8966 waters, the system global net charges must be zero, so electrostatic balancing by 3 sodium ions replacement was required.

The results were analyzed through several built-in functions from the Gromacs software in which is included the root mean square deviation (rmsd, function: g\_rms).

## 3.4 Biophysical Characterization Studies

### 3.4.1 Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) was performed to determine and characterize the oligomeric state of PtpA protein in solution using a HPLC system (Prominence, Shimadzu) coupled with a Superdex-75 column 10/300 (GE Healthcare Life Sciences, 17517401). The Lab Solutions program was used to set-up the experience parameters.

The experience was performed with two different H buffers, only differing by the presence and absence of the reducing agent DTT (1,4- Dithiothreitol). The system was first washed with Milli-Q water and equilibrated with the respective buffer H with (100 mM HEPES 7.0, 150 mM NaCl, 5 mM EDTA, 5 mM DTT, 10% Glycerol) and without DTT (100 mM HEPES 7.0, 150 mM NaCl, 5 mM EDTA, 10% Glycerol) at 0.7 mL/minute flow rate. The protein was diluted to 6.4 mg/mL (buffer H without DTT) and to 10mg/mL (buffer H with DTT) being injected in the HPLC system with a flow rate set to 0.7 mL/minute. In the absence of DTT, 0.5 mL protein fractions were collected between 13.6 mL and 16.5 mL of elution volume. In the presence of DTT, similar protein fractions were collected between 12.6 mL and 16.1 mL.

The collected fractions were analyzed by denaturing polyacrylamide gel electrophoresis (10% SDS-PAGE) as described above.

### 3.4.2 Thermal Shift Assay

Thermal Shift Assay (TSA) experiments were firstly performed to assess the PtpA stability in three different approaches: 1) upon adding 10% DMSO (Dimethyl sulfoxide) and 10% glycerol; 2) testing two storage conditions (4°C and -80°C, overnight) and 3) investigate PEG 400 as an alternative solvent to DMSO testing 10%, 15% and 20% PEG/DMSO.

The reaction total volume was 20 µL: 17 µL of sample and 3 µL of Dye 8x, (Applied Biosystems, ThermoFisher Scientific), with a final concentration of the protein of 20 µM. Each measurement was executed in triplicates and the controls were also prepared (DMSO/glycerol and protein without dye; DMSO/Glycerol with dye 8x; PEG400/DMSO and protein without dye; PEG400/DMSO with dye 8x).

Additionally, TSA was also used to characterize the interactions between PtpA and the inhibitors. A first approach consisted on using increasing concentrations of ligand (C33, C37, R6, JE02, Lap04, Lap11, Azaindole 1, Azaindole 15 and Azaindole 16) solubilized in 100% DMSO: 1:1, 1:2, 1:5, 1:10, 1:50, 1:100, 1:250, 1:500 protein-ligand millimolar ratios (protein at 20 µM concentration). Afterwards, for Lap04 compound, a wide range of protein-ligand proportions was tested – namely 1:1, 1:2, 1:5, 1:10, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1:90, 1:100, 1:125, 1:150, 1:175, 1:200, 1:250 and 1:500 mM – with three protein concentrations (2 µM, 10 µM and 20 µM). Once again, each measurement was executed in triplicate and the respective controls (protein without dye; ligand with dye and ligand without dye) were also prepared.

The above-mentioned experiments were centrifuged for 1 minute at 1000 rpm (Centrifuge 5804 R, Eppendorf) and performed using MicroAmp® fast 96-well reaction plates (Applied Biosystems, ThermoFisher Scientific) in 2 minutes cycles of 1% increments between 25°C and 95°C in a StepOnePlus Real-Time PCR System (Applied Biosystems, ThermoFisher Scientific). Data processing and analysis were performed using Microsoft Office Excel.

### 3.4.3 Microscale Thermophoresis

Microscale Thermophoresis (MST) experiments comprised three steps: protein labelling *via* His tag, binding check and binding affinity.

First, the protein was labeled with a fluorescent dye (RED-tris-NTA, NanoTemper Technologies), which attaches to the His tag, enabling to detect and to analyze. For this purpose, a 100 µL solution with a final concentration of 100 nM of dye was prepared from a stock solution (5 µM) in buffer H (previously described). To this sample volume, 100 µL of protein at 200 nM concentration was added. The protein (100 nM final concentration) and dye (50 nM final concentration) mixture sample was incubated for 30 minutes and centrifuged (Centrifuge Velocity 13µ, Dynamica) at 4°C, 15000xg for 10 minutes.

The next procedure, referred as binding check, starts by preparing two samples individually. In the first one, 25 µL PtpA (100 nM) was mixed with 25 µL 10% DMSO diluted in buffer H (22.5 µL buffer and 2.5 µL 100% DMSO) for the target-only sample. In the second, 25 µL of 100 nM PtpA was mixed with 25 µL of ligand in buffer H (22.5 µL buffer and 2.5 µL ligand

dissolved in 100% DMSO) for the complex sample. Posteriorly, each sample was divided into four capillaries (Monolith NT.115 Capillary – NanoTemper Technologies) which were placed in the positions 1-4 (target-only sample) and 5-8 (complex sample) of the device tray. The measure was executed in a Monolith NT.115 device (NanoTemper Technologies).

Considering the binding check results, the binding affinity experiment was performed with Lap11 ligand. The compound was dissolved in 100% DMSO and further serially diluted in buffer H ensuring a constant DMSO concentration through the dilution series to protect the target from high DMSO concentrations. This binding affinity comprises three dilution steps each one using 16 capillaries.

In the first dilution step, 10  $\mu$ L of 2 mM Lap11 (dissolved in 100% DMSO) were pipetted into tube 1 and 7  $\mu$ L of 100% DMSO were pipetted into tubes 2-16. Afterwards, a volume of 7  $\mu$ L of the compound from tube 1 was transferred to tube 2 mixing carefully by gently pipetting up and down. The serial dilution was continued up to the final step where 5  $\mu$ L from tube 15 were transferred to tube 16 and 5  $\mu$ L from tube 16 were discarded to get an equal volume in all samples (Lap11 concentration ranging from 0.1 mM to 3.05 nM). Noteworthy, it is extremely important to maintain an equal number of up and down pipetting procedure in the tubes preparation.

Posteriorly, in the second step, 16 new tubes were prepared with 54  $\mu$ L assay buffer H and 6  $\mu$ L were transferred from the first dilution step to the corresponding buffer tube. The resulting DMSO concentration was 4% while a 1:10 dilution is observed for the ligand.

In the third and last step, 16 new tubes were prepared, each one containing 10  $\mu$ L of labeled target (PtpA). Then, 10  $\mu$ L of diluted Lap11 from the second step were added to the respective protein-containing tubes. The 16 final serially diluted protein-ligand samples were transferred to 16 capillaries positioned in 1 to 16 positions in the device tray and the measurement was started in the Monolith NT.115 device.

### **3.4.4 Urea-Polyacrylamide Gel Electrophoresis**

Urea-polyacrylamide gel electrophoresis was performed using a Novex 6% Tris-TBE urea minigel in a XCell SureLock™ Mini-Cell from Invitrogen. Running buffer was diluted 1:5 from 89 mM Tris base and 89 mM boric acid while the sample buffer used contained 45 mM Tris base and 45 mM boric acid, 6% Ficoll Type 400, 3.5 M urea and 0.005% bromophenol blue.

PtpA was incubated with the ligands (C33, C37, R6, Lap04 and Lap11) for 2 hours in a 1:2 protein-ligand molar ratio (protein at 20 mg/mL). Afterwards, the samples for the urea gel were prepared adding 5  $\mu$ L of protein-ligand solution and 5  $\mu$ L of sample buffer. Three control samples of ligand-free PtpA at 20 mg/mL, 5 mg/mL and 2.5 mg/mL were also prepared. Finally, the samples were loaded into the urea gel which was subjected to 180 V and 40 mA for 3 hours. The protein bands were visualized by staining with Coomassie blue.

## 3.5 Structural Characterization Studies

### 3.5.1 Crystallization Assays

#### 3.5.1.1 Ligand-free PtpA Crystallization Assays

A crystallization condition had been previously obtained in the Protein crystallography laboratory. For that, an initial screening experiment was performed with ligand-free PtpA (16.7 mg/mL) using a 96-well plate (Molecular Dimensions) and a crystallization robot (Oryx8, Douglas Instruments) with the commercial screen JCSG<sup>+</sup> (Molecular Dimensions, MD1-37) (Appendix 7.4) at 4°C with 1  $\mu$ L drops (0.67  $\mu$ L protein and 0.33  $\mu$ L precipitating agent). The determined crystallization condition was then further optimized in scale-up trials – 24-well plate (XLR, Molecular Dimensions) – by Hanging-drop technique. Different approaches were used including the use of different protein-precipitant drop proportions, different protein and precipitant concentrations and the variation of pH and temperature (4°C and 20°C) values. The best protein crystals were found in the condition based on 6 % PEG 6K e 0,1 M HEPES 6.5, 15 days after the experimental setup at 4°C in 4  $\mu$ L drops (3  $\mu$ L protein and 1  $\mu$ L precipitant agent) with 700  $\mu$ L precipitant agent in the reservoir.<sup>72</sup>

Nonetheless, these first optimization trials failed in finding a reproducible condition. Therefore, I started new optimization experiments through the variation of precipitant type and its respective concentration in scaled-up conditions with 3  $\mu$ L protein and 1  $\mu$ L precipitant solution hanging drops, at 4°C and 20°C (Table 3.1).

**Table 3.1** - Optimization trials tested for the crystallization condition of ligand-free PtpA, performed at 4°C and 20°C using 700  $\mu$ L of precipitant solution in the reservoir.

Protein Concentration (mg/mL)	PEG 1.5K (%)	PEG 3K (%)	PEG 4K (%)	PEG 6K (%)	HEPES (M)	pH	Protein:Precipitant drop proportion ( $\mu$ L)
24	4	4	4	4	0.1	6.5	3:1
	6	6	6	6			
	8	8	8	8			
	10	10	10	10			
	12	12	12	12			
	14	14	14	14			

Additionally, since the original crystals were obtained with 6% PEG 6K, further crystallization optimizations were attempted to understand if the precipitant agent (PEG 6K) concentration (4-14%) and different protein:precipitant drop proportions (1:1; 2:1; 3:1) influenced the formation of PtpA crystals.

In parallel, new crystallization conditions were searched, testing several available screens namely JCSG<sup>+</sup>, JBS 5,6,7,8 (JenaBioScience), PEG ION (Hampton Research) and an *in-house* sparse matrix screen (80!).<sup>73</sup> These trials, represented in Table 3.2, were carried out at 4°C and 20°C with different protein concentrations, using the crystallization robot Oryx8 (Douglas



Instruments) on 96-well crystallization plates (drops of 0.67  $\mu$ L of protein and 0.33  $\mu$ L of precipitant solution).

**Table 3.2** - Summary of the crystallization screens used for PtpA crystallization at different concentrations and temperatures.

Crystallization Screen	Protein Concentration (mg/mL)	Temperature (°C)
JCSG <sup>+</sup>	20	4; 20
JBS 5,6,7,8 80!	17	4; 20
PEG ION	17	4; 20
	16	20

### 3.5.1.2 Co-crystallization Experiments

Co-crystallization trials were attempted using a 1:2 mM protein-ligand ratio with 10% DMSO. PtpA was incubated with the compounds (chalcones, thiosemicarbazones and azaindoles) for different periods of time – overnight (16 hours), 2 hours and 10 minutes – at 4°C. Following the incubation time, the protein-ligand sample was centrifuged (centrifuge Velocity 13 $\mu$ , Dynamica) at 12000 rpm for 10 minutes. For the co-crystallization trials, the JCSG<sup>+</sup> screen was selected for the robot crystallization procedure based on the sitting-drop technique, through 1 $\mu$ L drops (0.67  $\mu$ L protein+ligand incubation and 0.33  $\mu$ L precipitating agent). The usual two temperatures (4°C and 20°C) were tested, however, due to the limited amount of available compounds, the co-crystallization with azaindoles 1, 15 and 16 were solely performed at 20°C. The Table 3.3 summarizes the details of the co-crystallization trials with the three families of compounds previously described.

**Table 3.3** - Co-crystallization assays conditions with chalcones (C33, C37, R6 and JE02), thiosemicarbazones (Lap04 and Lap11) and azaindoles (1, 15 and 16) by sitting-drop technique.

Crystallization Screen	Ligand	Incubation time	Protein Concentration (mg/mL)	Temperature (°C)
JCSG <sup>+</sup>	C33	16 h	19	4; 20
		2 h		
		10 min		
	C37	16 h	20	4; 20
		2 h		
		10 min		
	R6	16 h	16	4; 20
		2 h		
		10 min		
	JE02	16 h	16	4; 20
		2 h		
		10 min		
	Lap04	16 h	19	4; 20
		2 h		
		10 min		
	Lap11	16 h	19	4; 20
		2 h		
		10 min		
	Azaindole 1	16 h	19	20
		2 h		
		10 min		
	Azaindole 15	16 h	19	20
		2 h		
		10 min		
	Azaindole 16	16 h	19	20
		2 h		
		10 min		

Based on the most promising co-crystallization conditions (i) 10% PEG 6K, 0.1 M Bicine 9, and (ii) 24% PEG 1.5K, 20% Glycerol, further optimization trials were performed to improve the diffraction quality of the crystals. In these optimization efforts maintaining the mentioned protein-ligand ratio, small variations in the initial conditions were done by testing: 1) different protein concentrations (19 mg/mL and 9.5 mg/mL), 2) different precipitant agent concentrations (18-30% PEG 1.5K; 6-16 % PEG 6K), 3) different buffer pH values (Bicine 9, 8.5, 8, 7.6) and 4) the addition of several additives (CaCl<sub>2</sub>, NaCl, MgCl<sub>2</sub>, Li<sub>2</sub>SO<sub>4</sub>).

### 3.5.1.3 Microseeding Experiments

The crystallization condition based on 10% PEG 8K, 8% ethylene glycol and 0.1 M HEPES 7.5 from the co-crystallization trial with Lap11 was selected to provide the seeds for microseeding experiments.

Firstly, the crystals were transferred from the original crystallization drop to a harvesting buffer (15% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5) drop (50  $\mu$ L). This stabilizing condition with the crystals was then pipetted an eppendorf tube containing a seed bead (Microseed beads<sup>TM</sup>, Molecular Dimensions). The crystals were smashed into crystalline particles by vortexing and were centrifuged at 12000 rpm for 5 minutes. Serial dilution solutions (1:10 and 1:100) were prepared from the seed stock, providing different amounts of nuclei, and 10 aliquots of 20  $\mu$ L were stored at -80°C.

For the microseeding assay setup, the commercial screen JCSG<sup>+</sup> (Appendix 7.4) and the seed dilution solutions 1:10 and 1:100 were tested in 96-well plates at 20°C. One ligand of each family was selected –C33, Lap11 and azaindole 16 – and two methods of co-crystallization were tried. In the first approach, the protein was incubated with the ligand for 2 hours in 1:2 mM ratio (10% DMSO) and 1.1  $\mu$ L volume drops were executed through the crystallization robot Oryx8 (0.67  $\mu$ L protein-ligand sample, 0.10  $\mu$ L seed solution 1:10/ 1:100 and 0.33  $\mu$ L reservoir precipitant solution). The second method was based on an additive experiment in which the ligand was used as an additive to the drop (0.67  $\mu$ L protein, 0.10  $\mu$ L seed solution 1:10/1:100, 0.10  $\mu$ L ligand and 0.33  $\mu$ L reservoir precipitant solution).

Moreover, both seed solutions (1:10 and 1:100) were also tested in similar nano-crystallization trials with ligand-free PtpA at 16 mg/mL (drops with 0.67  $\mu$ L protein, 0.10  $\mu$ L seed solution 1:10/1:100 and 0.33  $\mu$ L precipitant solution).

## 3.5.2 X-ray Crystallography

### 3.5.2.1 X-ray Diffraction Experiment and Data Collection

The best obtained crystals were flash frozen with paratone oil or the respective harvesting buffer supplemented with 30% glycerol as a cryo-protectant. The crystals were exposed to synchrotron X-ray radiation at beamlines I04 (Diamond Light Source (DLS) - Oxfordshire, United Kingdom), ID23-2, ID30A-3 (European Synchrotron Radiation Facility (ESRF) – Grenoble, France) and BioMax (MAX IV – Lund, Sweden) and completed datasets were collected.

### 3.5.2.2 Structure Determination and Refinement

The datasets were processed using several programs. Initially, the measured reflections and respective intensities were integrated with the program XDS<sup>74</sup> and scaled with Aimless<sup>75</sup> from the CCP4<sup>76</sup> (Collaborative Computational Project Number 4) suite of programs which also gives some quality data diffraction statistics.

The phase problem was solved by Molecular Replacement (using the PDB entry 1U2P as template) with Phaser<sup>77</sup> (from CCP4) and the refinement was performed using the Low-Resolution Pipeline (LORESTR)<sup>78</sup> followed by visualization and adjustment of the model and electron density map in Coot (Crystallographic Object-Oriented Toolkit)<sup>79</sup>. The figures were performed using the visualization PyMOL 2.3.2 software.<sup>67</sup>

All these procedures correspond to the Lap11 co-crystallization data set and the same was followed to reprocess the data of C33 soaking experiment, previously obtained in the Protein Crystallography laboratory.<sup>72</sup>

# CHAPTER 4

---

## **R**ESULTS & **D**ISCUSSION



## 4.1 PtpA Expression

The transformation step, that consisted in the insertion of the plasmid pET28a encoding the 20 kDa PtpA into *E. coli*, BL21(DE3) strain competent cells, was successfully achieved. The referred plasmid pET28a used in transformation was previously isolated and obtained in the laboratory using the MiniPrep-NZYTech kit.

The bacteria cells with the incorporated plasmid were used in the protein heterologous expression. In this system, the expression of the recombinant protein of interest is induced by addition of IPTG in the culture medium. IPTG is a molecular mimic of lactose, being therefore used to induce the protein expression when the coding gene is under the control of the *lac* operator. The OD<sub>600nm</sub> values in 0.6-0.8 range are good indicators of the beginning of the exponential phase, which is usually the ideal time for the inducer addition. The cultures were grown for 4 hours at 30°C, which corresponds to the best conditions to produce the recombinant protein (expression optimization tests were done prior to this work).

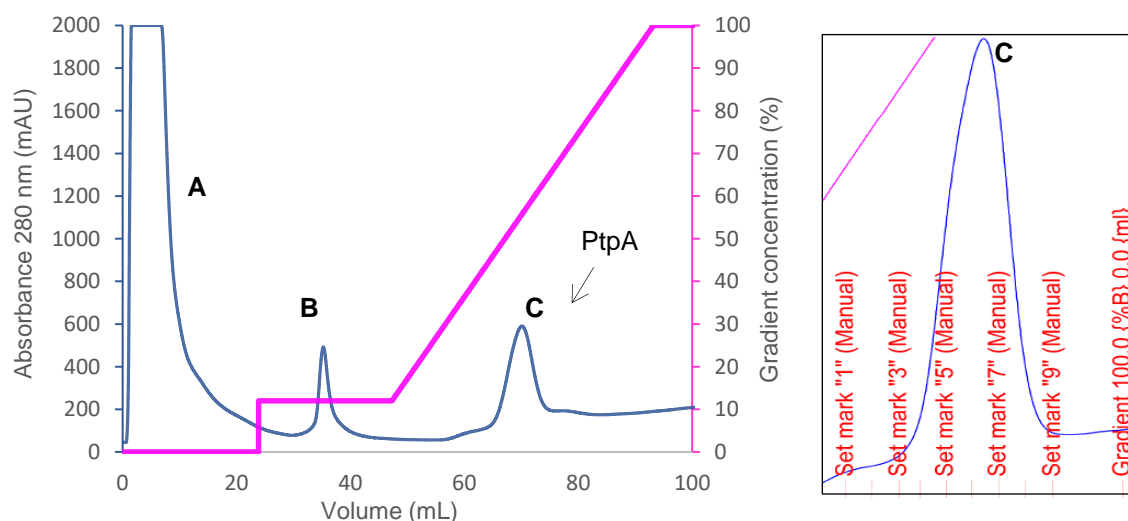
## 4.2 PtpA Purification

The pellets obtained in the expression step were lysed using an ultrasonic cell disruptor, in which ultrasound generated acoustic cavitation and shear forces break the cell membrane of *E. coli*, consequently leading to cellular disintegration. Then, the resulting homogenate was centrifuged, being possible to separate the soluble fraction, which contains proteins, from the insoluble cellular material that forms deposited sediment (pellet). The supernatant (soluble fraction) was collected and submitted to the purification step.

The step of PtpA protein purification comprised an Immobilized Metal Affinity Chromatography (IMAC) using a HisTrap™ column coupled to the AKTA Start system. The IMAC method is based on a reversible interaction between the protein of interest and a specific ligand coupled to the chromatographic matrix. In this case, the hexahistidine tag localized at the N-terminal of PtpA is retained on the column by immobilized nickel ions in the form of nickel nitriloacetic acid (Ni-NTA). This way, the recombinant protein is separated from *E. coli* BL21 endogenous proteins due to high affinity of the histidine tag to the nickel. Since the chemical structure of imidazole is very similar to the histidine structure, the elution of proteins and contaminants with low affinity to the column becomes possible.

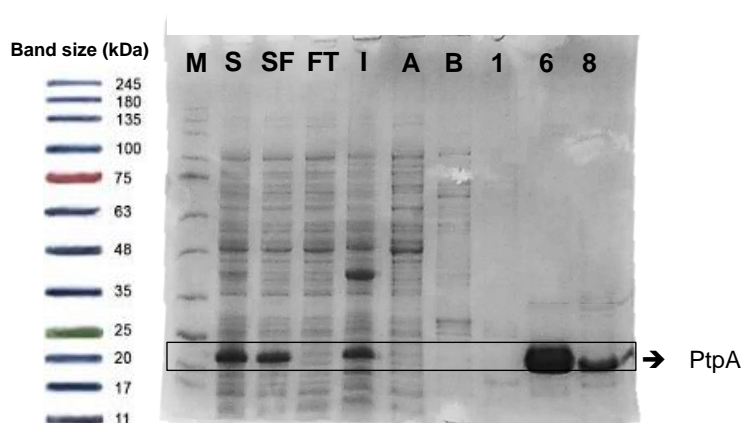
By the analysis of the chromatogram in Figure 4.1, as expected, the soluble fraction contains a high content of proteins that do not show affinity for nickel which are eluted using buffer A (20 mM Tris-HCl pH 8.0, 10 mM Imidazole, 500 mM NaCl, 5 mM DTT), corresponding to peak A. The target protein as well as some other contaminants bind to the column. Such contaminants that exhibit low affinity towards Ni<sup>2+</sup> are eluted by introducing a step

of 12% buffer B (which has an equal composition with buffer A, but with 500mM imidazole), corresponding to peak B. Finally, a gradient of buffer B (12-100%) was created and, the protein was eluted between 45% and 60% (peak C) approximately, corresponding to 225 and 300 mM imidazole concentration, respectively.



**Figure 4.1** – On the left, IMAC chromatogram resulting from PtpA purification. Legend: **A** – Elution peaks of proteins with no affinity to the nickel column; **B** – Elution peak of proteins with low affinity to the nickel column (step of 12% B); **C** – PtpA elution peak resulting in the increase of the ionic strength by a gradient of buffer B (12-100%B). On the right, the close-up view of PtpA elution peak and the corresponding protein collected fractions (1 mL).

The fractions from expression and purification were examined by SDS-PAGE and the results are shown in Figure 4.2.



**Figure 4.2** – SDS-PAGE (10%) with fractions from the expression and purification steps of PtpA. Gel legend: **M**- NZYTech protein marker II; **S**- Sonicated; **SF**- Soluble fraction; **FT**- Flow Through; **I**- Insoluble fraction; **A, B**- peaks A and B of the IMAC chromatogram; **1,6,8**- Respective fractions of the PtpA elution peak (peak C).

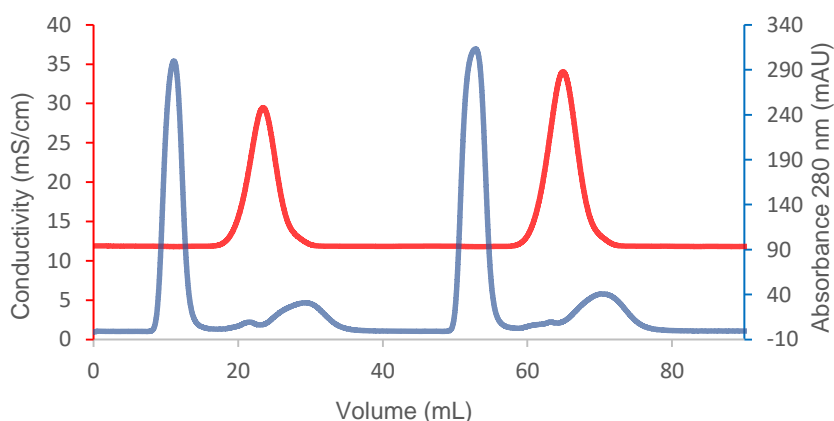
Analyzing the gel, it is possible to verify the presence of an intense band (20 kDa) corresponding to PtpA in the sonicated well (S), which reflects the success of the expression step.



The same band is also observable in the soluble fraction gel lane (SF). Since these samples were obtained prior to the purification step, other bands of smaller and higher molecular weights are visible, corresponding to contaminants. In addition, the loaded sample which corresponds to flow-through (FT) does not show the presence of a band corresponding to PtpA protein, suggesting that the protein was successfully retained in the column, when loaded into the IMAC column in purification. The same is applied to A and B gel wells, which correspond to the peaks A and B of the chromatogram, indicating the success of the IMAC method. Additionally, the well “1” corresponds to a “shoulder” feature at the beginning of peak C, with “shoulder” features, indicating that PtpA is not being eluted in this portion. In wells “6” and “8” it is possible to verify bands corresponding to PtpA protein, due to the existence of bands in 20 kDa. In “6”, which corresponds to a middle-peak fraction, a very intense band is present. Thus, PtpA protein exhibits a satisfactory degree of purity by IMAC and no further purification steps were performed.

Following the purification procedures, it is necessary to remove imidazole from the PtpA buffer, since it destabilizes the protein intramolecular interactions. The pure fractions from the SDS-PAGE gel were pooled together and loaded on three HiTrap™ desalting columns coupled to the AKTA Start HPLC with the main purpose of buffer exchange to buffer H (100 mM HEPES pH 7.0, 150 mM NaCl, 5 mM DTT, 5 mM EDTA).

After loading the sample, the protein is eluted in the buffer of interest, as represented in the blue peaks of the chromatogram in Figure 4.3. Posteriorly, the salt and imidazole pass through the column with the concomitant change of conductivity (red peaks in chromatogram) and absorbance at 280 nm.



**Figure 4.3** – PtpA desalting chromatogram. The blue peaks correspond to the protein elution (two injections were applied) and the red peaks represent the imidazole and salt elution from the desalting columns.

After protein expression and purification, a reasonable yield of approximately 20 mg of soluble protein *per liter* of culture was achieved. The protein was stable during the concentration reaching concentrations up to approximately 24 mg/mL to be used in subsequent structural and biophysical assays.

## 4.3 *In silico* studies

### 4.3.1 Molecular Docking

#### 4.3.1.1 Chalcones and Thiosemicarbazones

In 2005, Waldmann and co-workers, the pioneers who suggested the first PtpA inhibitors, reported that structural scaffolds of natural compounds provide features that can act as biologically relevant elements in proteins' binding.<sup>80</sup>

An example of natural compounds are the chalcones, which emerged as agents with chemotherapeutic potential against *Mtb* and as potent phosphatase inhibitors. This family of compounds was explored by Chiaradia and collaborators who tested the effects of chalcones in the PtpA inhibition by kinetics studies, molecular modeling, selectivity and cell-based assays, classifying them as competitive inhibitors. More recently, Terenzi *et al* reported, for the first time, thiosemicarbazones as a new potential class of inhibitors toward PtpA. The compounds were biologically evaluated, displaying reasonable inhibition and specificity. Kinetic studies showed a non-competitive mode of inhibition for Lap04 and Lap11 compounds that present a naphthyl moiety, a structural feature important for inhibition potency.<sup>43,44,45</sup>

Using molecular docking, we wanted to investigate and propose the binding mode of both competitive and non-competitive compounds to the protein. This computational method emerges as a crucial step in understanding the mode of interaction of these potent and selective inhibitors with PtpA protein. The first binding prediction from these compounds was achieved using the AutoDock Vina program. However, this program encompasses several simplifications in order to allow faster computational calculations (two orders of magnitude speed-up when compared with AutoDock 4.0 program). Therefore, to achieve more realistic and rigorous results, AutoDock 4.0 program was also used to predict the binding mode.

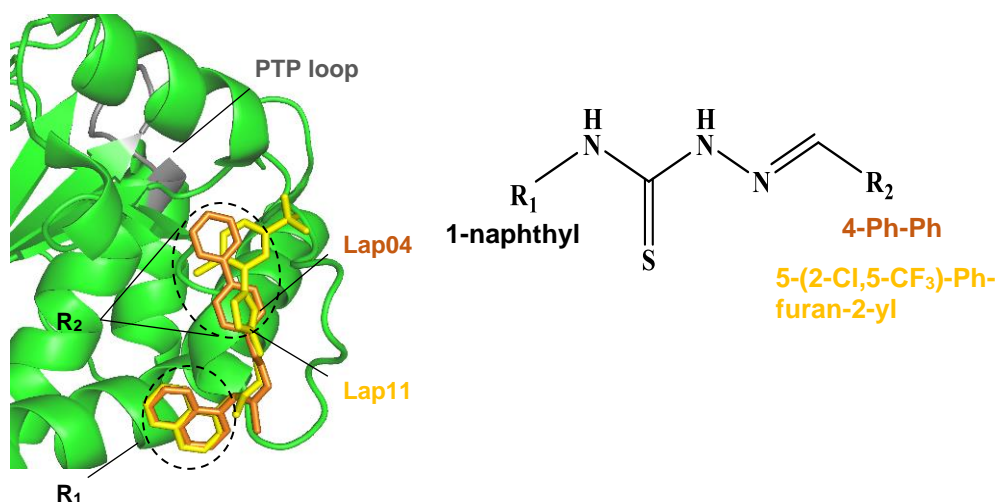
Blind docking, a method in which the ligand is docked to the whole protein surface without any prior knowledge of the target pocket, was performed for each compound in order to identify the location of the binding site and eventually corroborate the modelling studies reported in literature by Terenzi *et al*.

Table 4.1 shows the best results obtained for each tested compound - binding free energies ( $\Delta G$ ) and inhibition constants ( $K_i$ ), using the docking programs previously referred.

**Table 4.1** - Docking results of PtpA and chalcones and thiosemicarbazones. The values were predicted by AutoDock Vina and AutoDock.

Compound	AutoDock Vina		AutoDock	
	$\Delta G$ (kcal/mol)	$K_i$ ( $\mu M$ )	$\Delta G$ (kcal/mol)	$K_i$ ( $\mu M$ )
C33	- 6.66	13.12	- 7.66	2.43
C37	- 6.71	12.06	- 7.64	2.51
R6	- 6.40	20.35	- 7.16	5.64
JE02	- 7.20	5.28	- 7.41	3.70
Lap04	- 7.50	3.18	- 9.21	0.17
Lap11	- 7.91	1.62	- 8.88	0.31

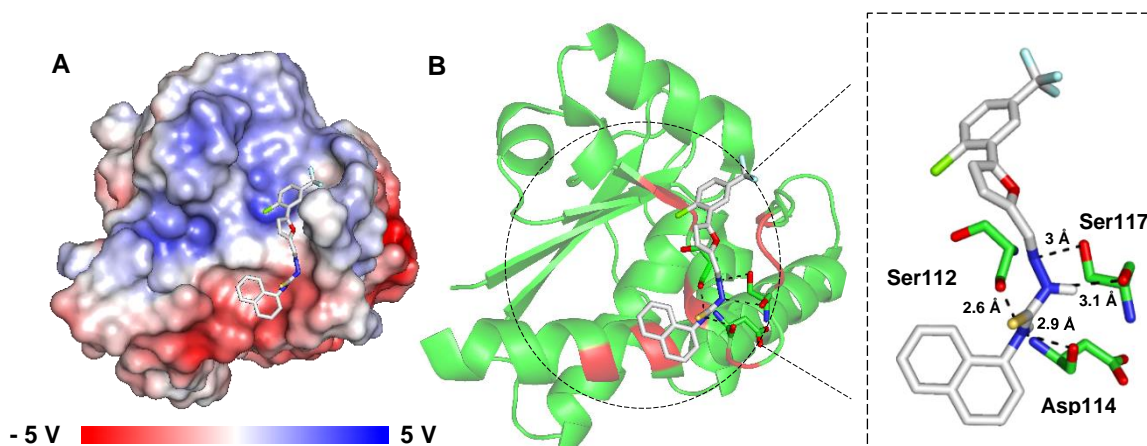
Generally, the best AutoDock Vina and AutoDock scores are displayed by thiosemicarbazones Lap04 ( $\Delta G$ : - 9.21 kcal/mol) and Lap11 ( $\Delta G$ : - 8.88 kcal/mol). In fact, the inhibitory potency of these compounds, exhibiting nanomolar/micromolar values is noticeable higher when compared with the ones obtained for the chalcones. As represented in Figure 4.4, both inhibitors share a common naphthyl moiety in  $R_1$  group which is described by Terenzi *et al* as an extremely important group for the inhibitory activity against PtpA.<sup>45</sup>



**Figure 4.4** – Schematic representation of the thiosemicarbazones overlapped, obtained by blind docking using AutoDock calculations, and their respective substituent groups in  $R_1$  and  $R_2$ . Figure prepared in PyMOL.

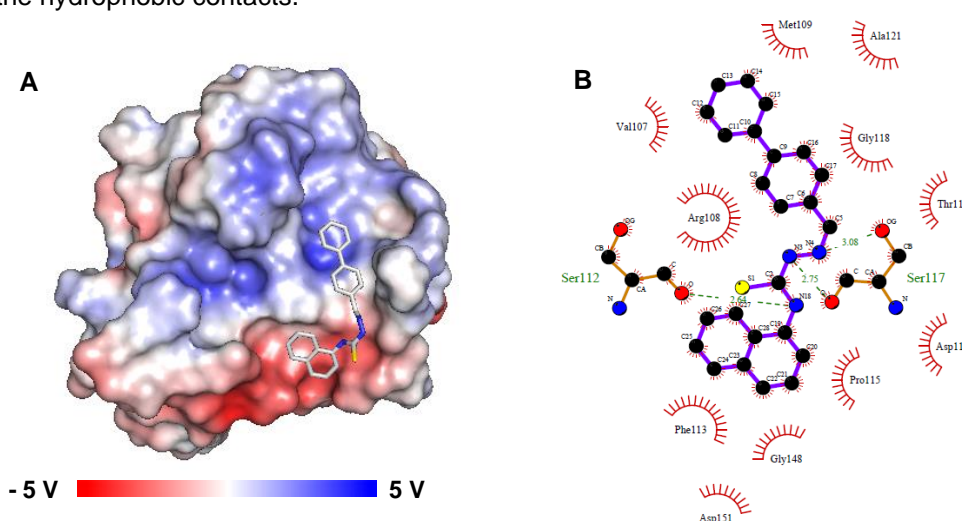
The predicted docking-based binding mode of Lap11 compound indicates that the inhibitor adopts an extended conformation within an adjacent site of the protein which does not correspond to the catalytic site. In detail, the naphthyl group ( $R_1$ ) is docked in a hydrophobic groove, negatively charged, delimited by the residues Arg108, Phe113, Pro115, Gly148 and Asp151. The second group is docked in a positive charge cavity and the last phenyl ring, containing two electron-withdrawing groups (Cl and F), provides a strengthened pose in this

region (Figure 4.5 A). The protein residues Ser112, Asp114 and Ser117 establish hydrogen bonds with Lap11, playing a key role in orientating the inhibitor within the binding site (Figure 4.5 B).



**Figure 4.5** - Allosteric binding mode of the inhibitor Lap11 to PtpA, obtained by blind docking. **(A)** Lap11 binding pose on the protein electrostatic surface calculated by APBS in PyMOL. **(B)** Cartoon protein representation (green) with the hydrophobic residues interacting with Lap11 colored in red (Arg108, Met109, Phe113, Pro115, Gly118, Thr119, Ala121, Gly148 and Asp151). H-bonded residues (Ser112, Asp114 and Ser117) are shown in green sticks and hydrogen bonds illustrated in black dashed lines. Figures prepared in PyMOL.

Based on the docking scores mentioned in Table 4.1, Lap04 also shows a preferential binding in the same allosteric site predicted from Lap11 inhibitor. As depicted in Figure 4.6 A, the naphthyl moiety is docked in a cavity surrounded by hydrophobic residues namely Phe113, Pro115, Asp151 and Gly148. In addition, Ser112 and Ser117 are the main residues establishing hydrogen bonds with the nitrogen atoms of ligand, as observable in Figure 4.6 B. In addition, the nonpolar part of the compound ( $R_2$  group) fits in a positive and neutral charge groove establishing hydrophobic interactions, which significantly contributes to the complex stabilization. In this region, Val107, Arg108, Met109, Gly118, Thr119 and Ala121 are the main residues responsible for the hydrophobic contacts.

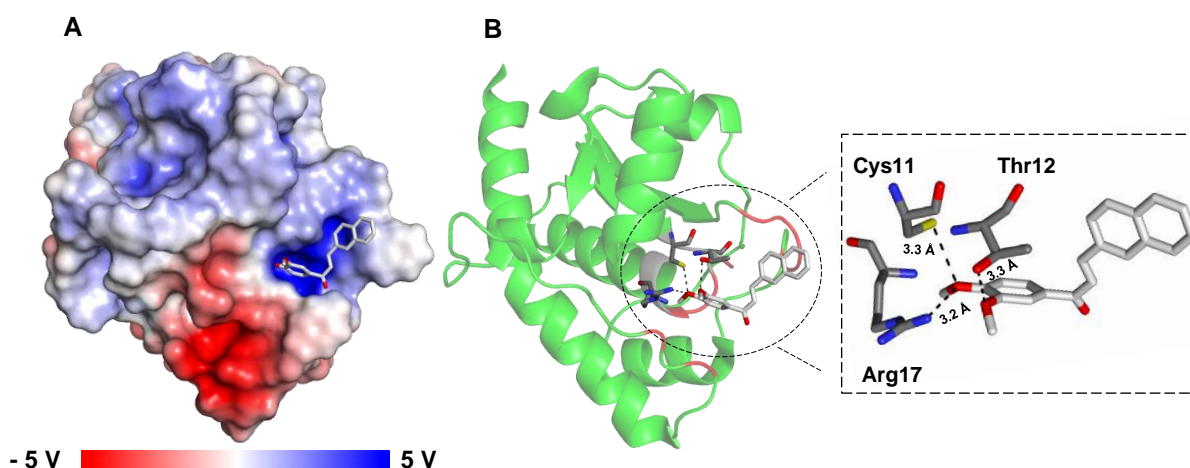


**Figure 4.6** – Best blind docking-based pose for the Lap04 compound within the adjacent site of PtpA. **(A)** Protein electrostatic surface (calculated by APBS in PyMOL) showing the  $R_1$  and  $R_2$  groups in differently charged regions. **(B)** LigPlot schematic diagram of PtpA-Lap04 interactions. The residues circled in red represent the hydrophobic interactions. Residues Ser112 and Ser117 are the main residues responsible for H-bonds in this complex (green dashed lines).

Considering the chalcones family, the best protein-ligand complex pose was selected for the four compounds based on the estimated free energy of binding and inhibition constant values which are above-mentioned in Table 4.1.

Among the analyzed chalcones, C33 ligand exhibits the best docking score ( $\Delta G$ : - 7.66 kcal/mol by AutoDock). As expected, this competitive inhibitor is found at the enzyme active site (Figure 4.7 A and B) with the methoxy groups of the first ring oriented near the PtpA active site pocket, positioned in a favourable orientation to share three hydrogen bonds with the PTP loop residues Cys11, Thr12 and Arg17. One of the oxygen atoms of the methoxy groups shares two hydrogen bonds with Cys11 and Arg17 and the oxygen of the other methoxy substituent group shares with Thr12 residue. The residues Gly13, Ile15, Cys16, Gly46, Asn47, Trp48, His49, Asp126 and Tyr128 undergo hydrophobic interactions.

On the other hand, the 2-naphthyl group (last two rings) establishes  $\pi$ -stacking interactions with the Trp48 residue. These aromatic ring interactions and the H-bonds reveal to be particularly important in orientating the inhibitor within the active site, being crucial in the complex stabilization.



**Figure 4.7** – Representation of the blind docking-based binding mode of the chalcone C33 within the catalytic site of PtpA. **(A)** Electrostatic surface of PtpA (calculated by APBS in PyMOL) with C33. **(B)** Close-up view of the interactions between the inhibitor and the protein. The PTP loop residues Cys 11, Thr12 and Arg17 (dark grey, sticks) are involved in ligand binding through hydrogen bonds. Colour code: the protein structure is represented as green cartoon, the PTP loop is shown in dark grey and the hydrophobic residues are depicted in red for an easily interpretation; the ligand is shown in light grey sticks and the hydrogen bonds in black dashed lines. The figures were prepared in PyMOL program.

The AutoDock Vina results regarding the six compounds as well as the AutoDock binding mode results of C37, R6 and JE02 ligands are represented in Appendix 7.5.

#### 4.3.1.2 Azaindoles

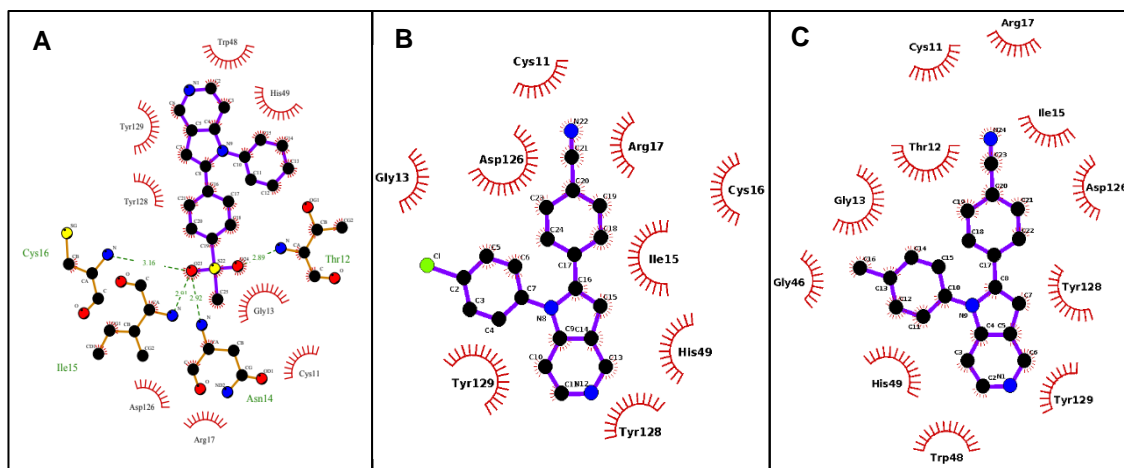
A collaboration established with Professor Maria Manuel Marques (Organic Synthesis Laboratory at LAQV-REQUIMTE, FCT-NOVA) allowed exploring azaindole derivatives as putative PtpA inhibitors. So far, no structural data on the binding mode of these molecules have

been reported in the literature, existing a limited understanding of their mechanism of action to guide subsequent critical developments in order to improve specificity and selectivity.

The azaindoles' library was docked to PtpA protein using AutoDock Vina and AutoDock programs for blind docking experiments. The best pose of each compound as well as the respective docking scores ( $\Delta G$  values) can be consulted in Appendix 7.5. Since there are twenty compounds, the top three were selected in order to facilitate the results presentation and discussion. The predicted binding free energies and inhibition constants are presented in Table 4.2. and the best pose of the three hits are depicted in Figure 4.8.

**Table 4.2** - AutoDock Vina and AutoDock docking results of the best three hits from the azaindoles' library with PtpA.

Compound	AutoDock Vina		AutoDock	
	$\Delta G$ (kcal/mol)	$K_i$ ( $\mu M$ )	$\Delta G$ (kcal/mol)	$K_i$ ( $\mu M$ )
<b>4</b>	- 6.40	20.35	- 8.25	0.89
<b>6n</b>	- 6.71	12.06	- 7.69	2.31
<b>12</b>	- 6.82	10.02	- 7.75	2.08



**Figure 4.8** – Ligplot diagrams showing the protein-ligand interactions of top three azaindoles based on the energy score (A)- Azaindole 4; (B) - Azaindole 6n; (C) - Azaindole 12. Non-ligand residues involved in hydrophobic contacts are depicted in red and the formation of hydrogen bonds is shown by a green dashed line.

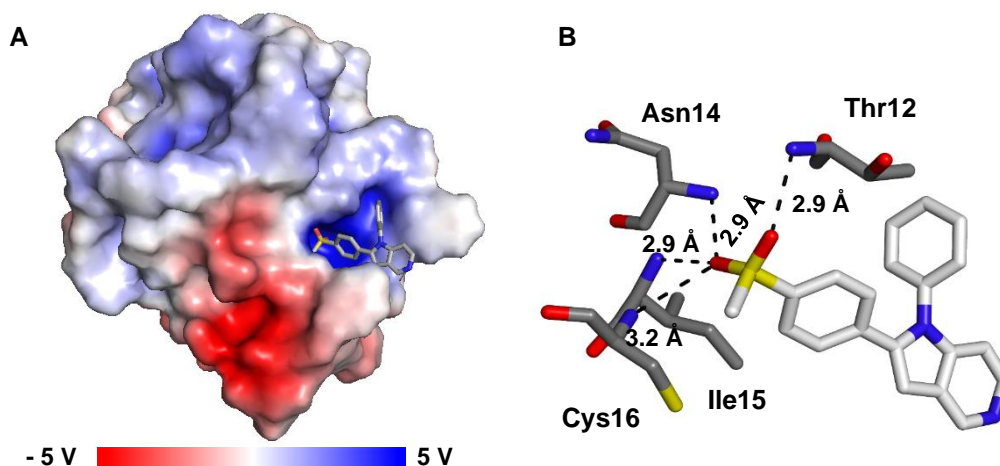
As observable in Figure 4.8, this family of compounds has a characteristic Y-shape that accommodates nicely in the PtpA active site. Considering the azaindoles 6n and 12, the results suggest a binding pose within the active site of PtpA. As represented in the Ligplot diagrams, only hydrophobic contacts are established, depicted as red circled residues in Figure 4.8.

In the case of azaindole 6n (Figure 4.8 B), the cyclohexane carbonitrile group is positioned inwards, creating hydrophobic contacts with the active site residues Cys11 and Arg17. Consequently, the pyrrole and pyridine rings are more exposed to the outside of the cavity. The same occurs with compound 12, which is oriented in such a position that allows the last two rings to participate in  $\pi$ -stacking contacts with Trp48 residue (Figure 4.8 C). Both ligands do not stablish



hydrogen bonds with PtpA protein, suggesting hydrophobic interactions as decisive contacts in ligand orientation.

The best hit corresponds to azaindole 4 which has a considerably lower binding free energy ( $\Delta G$ : -8.25 kcal/mol) and an inhibition constant of 0.89  $\mu\text{M}$  (Table 4.2). The Y-shape geometry nicely sits into the positively charged active site pocket as depicted in Figure 4.9 A.



**Figure 4.9** – Predicted binding mode of azaindole 4 within the PtpA active site by blind docking. **(A)** Azaindole 4 pose into the active site pocket, a predominantly positively charged cavity. **(B)** The catalytic residues (sticks, dark grey) Thr12, Asn14, Ile15 and Cys16 establishing hydrogen bonds (dashed black lines) with the oxygen atoms of the ligand (stick, light grey). Surface charge distribution of PtpA generated by PyMOL APBS tools.

In more detail, Figure 4.9 B shows the Y-shaped compound establishing H-bonds – absent in the previous discussed azaindoles – between the oxygen atoms of the sulphonyl group and the nitrogen atoms of the protein residues Thr12, Asn14, Ile15 and Cys16, which can explain the observed higher affinity of azaindole 4 toward PtpA. Moreover, the pyridine and pyrrole fused rings undergo  $\pi$ - $\pi$  interactions with the side chain of the non-polar aromatic Trp48 residue. Along with this, a few hydrophobic interacting residues were also observed namely Cys11, Gly13, Arg17, His48, Asp126, Tyr128 and Tyr129 residues (see in Appendix 7.5).

Additionally, it should be noted that the predicted binding free energy of this azaindole is very similar to the one determined for the interaction between PtpA and the thiosemicarbazone Lap11, the most potent inhibitor reported in literature. The elucidation of the molecular interactions, namely of azaindole 4, will bring significant insights on the subject making possible the identification of a new class of compounds with great potential to further development as PtpA inhibitors and anti-TB drugs.

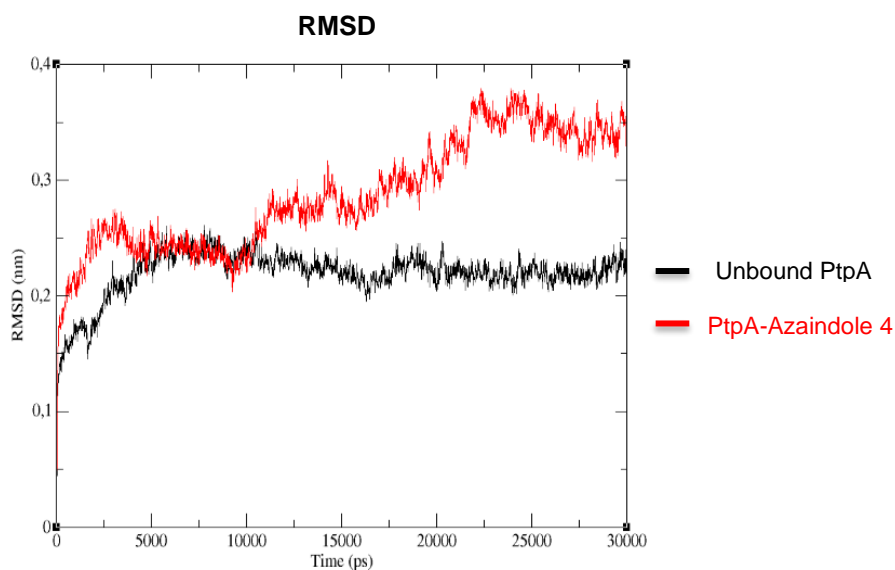
### 4.3.2 Molecular Dynamics Simulation

Currently, molecular dynamics (MD) is a widely used tool applied to investigate the structure, flexibility and thermodynamic properties of biological molecules and their complexes (protein-protein and protein-ligand).

Since the properties of a biomolecule have a profound effect on its function, the understanding of these characteristics retains particular importance. Molecular dynamics simulation allows the study of these features, analysing numerous dynamic processes that occur in biological systems and providing a detailed information about the individual and collective movement of particles of these systems in a given timeframe.

In this project, MD was performed aiming to provide structural information regarding the PtpA behaviour in solution and its interactions with the best hit from the azaindole family. Among the information obtained, binding site accessibility and ligand-residue specific interactions can be retrieved. Thus, after visual inspection of the simulation, it was noticeable that azaindole 4 interacts with the active site of the protein throughout all the simulation evidencing a structural disarrangements.

In addition, the RMSD analysis is also an important aspect to consider in a molecular dynamics experiment. In general, for an equal set of identical atoms present in two different conformations, RMSD provides a measure of the difference in position of these two sets of atoms. The higher the RMSD value, greater is the difference between the two structures. In this case, the complex PtpA-Azaindole 4 shows higher RMSD values during the 30 ns time period than the unbound protein (Figure 4.10). In fact, the presence of the compound azaindole 4 enhances the displacement of the protein atoms which was expected since the ligand accommodation at the binding site implies a structural rearrangement of the protein residues.



**Figure 4.10** – RMSD results of unbound PtpA (black lines) and PtpA-Azaindole 4 complex (red lines) through the MD simulations.

The complex has a higher potential energy value when compared with the unbound PtpA (Table 4.3) which indicates that the ligand binding is destabilizing the protein. This is corroborated by the higher RMSD values and by visual examination of the simulation, in which some secondary structure elements were found to be disordered.



**Table 4.3** - RMSD and potential energy analysis of the MD simulations of the complex PtpA-Azaindole 4.

	<b>RMSD (nm)</b>	<b>Potential Energy (kJ/mol)</b>
Unbound PtpA	0.213	-463486
PtpA-Azaindole 4	0.287	-394935

Since this preliminary analysis revealed to be useful to characterize the dynamic behaviour of PtpA- Azaindole 4, it is intended to extend its application to other compounds from different families in order to get similar insights.

## 4.4 Biophysical Characterization Studies

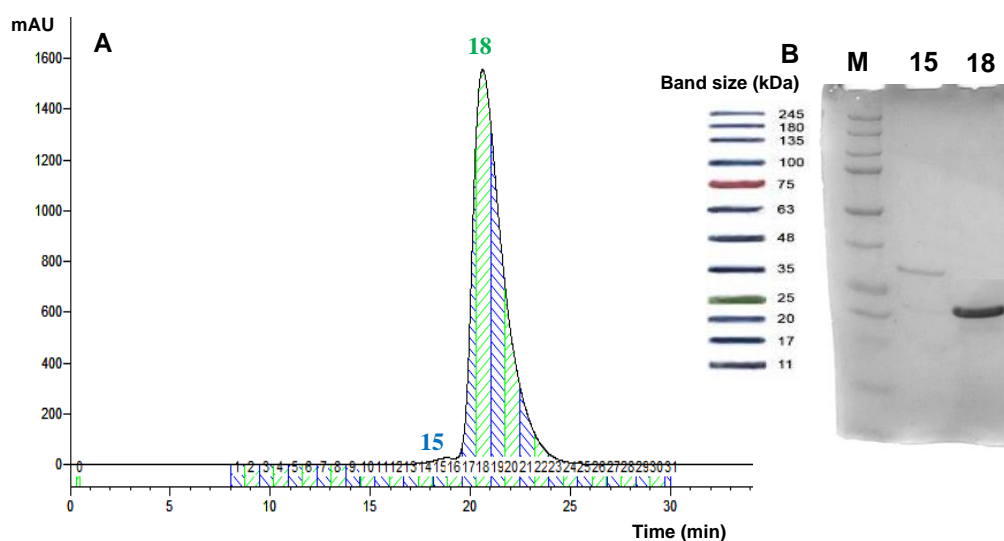
### 4.4.1 Size Exclusion Chromatography

The purpose of the Size Exclusion Chromatography (SEC) experiment was to study the oligomeric state of the protein in solution since a dimer – formed through a disulfide bridge - is observed in the crystallographic model (as further discussed in section 4.5.3). In addition, this third purification step could be advantageous for the crystallization process since a high degree of purity is many times critical to successfully obtain good diffracting crystals.

As represented the Figures 4.11 A and 4.12 A , a SEC experiment originates an elution profile, illustrating the variation on UV absorbance of the sample components which are eluted from the column according to their apparent size. Therefore, this technique allows an effective separation of proteins with different molecular weights: the elution of the smallest molecules is more delayed than the largest, which are eluted first from the column.<sup>83</sup>

In order to achieve the proposed objective of this experiment, two separated assays were performed using a Superdex S75 column, varying the composition of buffer H - presence and absence of the reducing agent DTT). Thus, two desalting procedures were carried out prior to the SEC run, using two different final H buffers: the first containing 100 mM HEPES 7.0, 150 mM NaCl, 5 mM EDTA, 10% Glycerol (Buffer H); and the second, with the same constituents plus 5 mM DTT (Buffer H\_DTT).

The SEC chromatogram obtained for the PtpA sample in Buffer H (Figure 4.11 A) shows a single peak eluted from 19 to 23 minutes which corresponds to a molecular weight around 20 kDa. Therefore, the SEC experiment clearly suggests a homogeneous sample, with the presence of the protein as a monomer in solution.

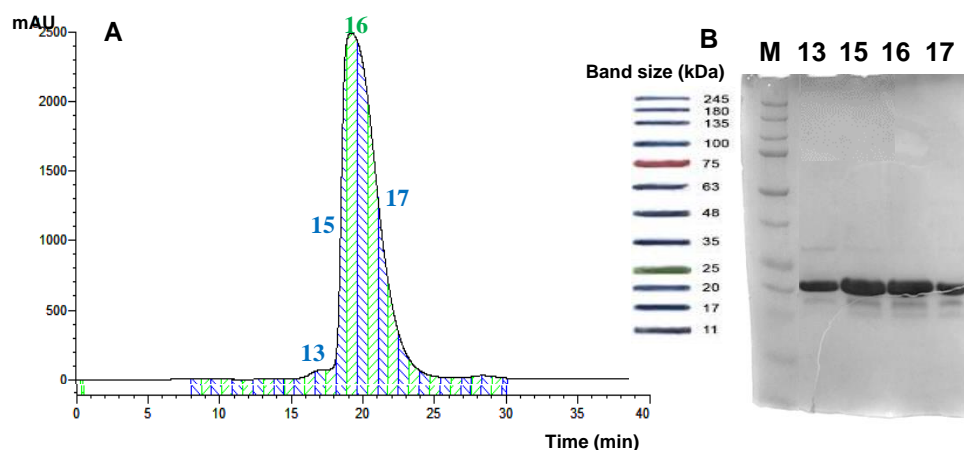


**Figure 4.11** – (A) SEC chromatogram obtained for PtpA in the absence of DTT. (B) SDS-PAGE gel analysis of the different fractions obtained from the SEC (fractions were identified according to the numeration of the chromatogram).

The fraction 18 of the chromatogram was analysed in SDS-PAGE gel to confirm that the peak corresponds to PtpA. Additionally, the observed small “peak-shoulder” in the beginning of the elution peak was also investigated.

By the gel analysis of fraction 18, it was possible to verify a very intense band of approximately 20 kDa, corresponding to the expected monomeric form of the protein. On the other hand, fraction 15, corresponding to the suspicious bump, does not present any band around 20kDa suggesting that it corresponds to a contaminant (Figure 4.11 B).

A second SEC experiment was performed using the protein with Buffer H\_DTT (Figure 4.12 A). Once again, the resulting chromatogram shows a similar elution profile behaviour with a single peak eluted from 18 to 23 minutes, approximately, corresponding to a molecular weight of 20 kDa, approximately. Therefore, the results also indicate that, even in the presence of DTT, PtpA remains as a monomer in solution.



**Figure 4.12** – (A) SEC chromatogram obtained for PtpA in the presence of DTT. (B) SDS-PAGE gel analysis of the different fractions obtained from the SEC (fractions were identified according to the numeration of the chromatogram). Numbers 13-17 correspond to the collected fractions during the SEC run.

Analyzing the results obtained by SDS-PAGE electrophoresis, the fractions 13, 15, 16 and 17 showed a band corresponding to PtpA protein (20 kDa). On the other hand, by the gel inspection, it was possible to observe lower bands much narrower than PtpA band likely corresponding to proteolysis process.

Therefore, the pure fractions (13 to 19 in Figure 4.12 B) were pooled and concentrated to a protein concentration of 16.5 mg/mL (DTT presence) and 15 mg/mL (DTT absence), respectively. Ligand-free crystallizations attempts were performed with these protein samples, but no crystals were obtained.

As major conclusion, the SEC results evidence that, regardless the reducing agent presence, PtpA presents a monomeric oligomerization state in solution. This feature is in agreement with the previously obtained Small Angle X-ray Scattering (SAXS) results supporting the hypothesis by which the disulfide bond, existing in the discussed X-ray structures, corresponds effectively to a crystallographic artefact.<sup>72</sup>

## **4.4.2 Thermal Shift Assay**

### **4.4.2.1 Protein Stability**

Thermal Shift Assay (TSA) provides relevant biophysical information regarding the protein stability in a wide set of different conditions. The technique is based on the principle that the structural stability and the resistance to denaturation of a protein can be increased due to mutations, different buffer conditions, or to the interaction of ligands - by establishing non-covalent interactions with several protein residues. In this case, protein denaturation is promoted through the increase of temperature, which eventually will cause internal hydrophobic residues to be exposed and activate the fluorescence of a dye present in solution. The fluorescence intensity increases with protein denaturation which is followed in a determined range of temperatures (usually from 25°C to 95°C). A TSA experiment provides the melting temperature value ( $T_m$ ) of the protein, corresponding to the temperature at which half of the protein molecules are denatured.<sup>84,58,86</sup>

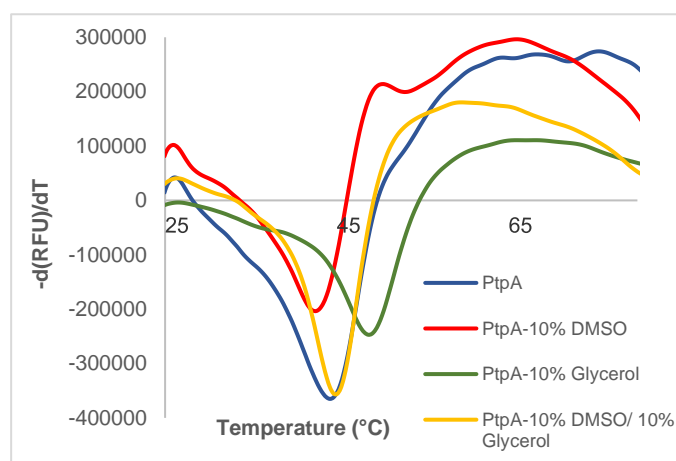
TSA experiments were conducted with the objective of characterizing the PtpA thermal stability, and to detect potential protein-ligand interactions. The protein thermal stability was evaluated considering three different scenarios.

In the first scenario, the effect of DMSO in the protein stability was investigated. Since the ligands under study are only soluble in organic solvents, it is crucial to determine their potential destabilization effects on the protein, and a 10% DMSO concentration was selected to be tested.

An initial TSA experiment was tried using a PtpA sample diluted in buffer H without glycerol (blue melting curve, Figure 4.13) and a  $T_m$  value of 44.60°C was obtained (Table 4.4). A similar experiment was performed by adding 10% DMSO to the protein (red melting curve, Figure 4.13) revealing a negative  $T_m$  shift ( $\Delta T_m$ : -1.99 °C) when compared with the control value (Table 4.4). This result indicates that the used percentage of DMSO possibly induces slight structural changes towards a more disordered conformation reflected by its lower thermostability.

Nevertheless, it should be highlighted that the presence of glycerol in the final protein buffer (buffer H) was previously determined to significantly stabilize PtpA.<sup>72</sup> In this sense, an additional TSA experiment confirmed that 10% glycerol (green melting curve, Figure 4.13) causes a positive  $T_m$  shift of 4.05°C (Table 4.4).

Considering that the buffer H contains 10% glycerol, its presence should be taken into account to evaluate the effects of DMSO on the protein. A final TSA experiment was performed using simultaneously both 10% DMSO and 10% glycerol (yellow melting curve, Figure 4.13) resulting in a  $T_m$  value of 45.01°C (Table 4.4). This value is very similar to the one found for the first sample (44.60°C), corroborating the remarkable importance of glycerol in PtpA thermal stability, since its presence reduces the negative destabilization effect of DMSO on PtpA. Therefore, based on the TSA results, the use of 10% DMSO appears to be a reasonable compromise to maintain the protein stable in further protein-ligand studies.



**Figure 4.13** - Effect of glycerol and DMSO on PtpA thermal stability. The melting curve of 10% DMSO presents a lower  $T_m$  reflecting that protein stability is noticeably affected. By the contrary, the presence of 10% glycerol significantly stabilizes the protein even when a similar concentration of DMSO is used.

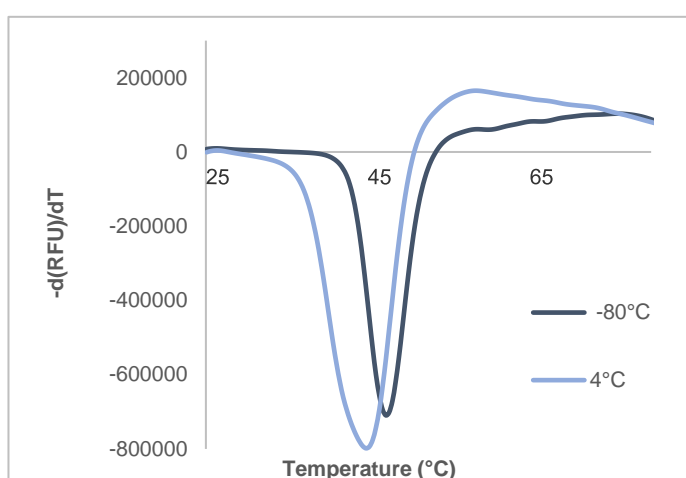
**Table 4.4** – Melting temperatures (°C) obtained from TSA experiments to assess the effect of glycerol and DMSO on PtpA stability.

Melting Curve	$T_m$ (°C)	$\Delta T_m$ (°C)
PtpA	44.60	-
PtpA-10% DMSO	42.61	-1.99
PtpA-10% Glycerol	48.65	+4.05
PtpA-10% DMSO/Glycerol	45.01	+0.41

In the second scenario, the effects of two storage conditions on the protein quality were tested. It is known that long-term storage or repeated freeze-thaw cycles can cause loss of functional properties and conformational changes in proteins.

In the case of PtpA, the protein revealed to be unstable when stored overnight at 4°C, and an alternative storage condition at -80°C was proposed. Thus, a TSA experiment was carried out to assess the PtpA thermostability when stored at 4°C and -80°C, evaluating the effect of freeze-thaw cycles on its stability and functional properties.

In Figure 4.14 are represented the melting curves and in Table 4.5 are depicted the respective  $T_m$  values regarding the two different storage conditions. Compared with the protein stored at 4°C (light blue melting curve) with a  $T_m$  of 44.81°C, a different melting curve profile was found for the protein stored under freezing conditions (dark blue melting curve). Hence, it was possible to conclude that protein remains very stable while frozen and after being thawed, showing a higher melting temperature value (47.64 °C). This experiment supports the storage at -80°C ensuring protein stability for further biophysical and structural studies.



**Figure 4.14** – Representation of the melting curves from different storage conditions. Frozen PtpA stability is ensured (dark blue curve) with a higher melting temperature comparatively to protein stored at 4°C (light blue curve).

**Table 4.5** -  $T_m$  (°C) values from two different storage conditions.

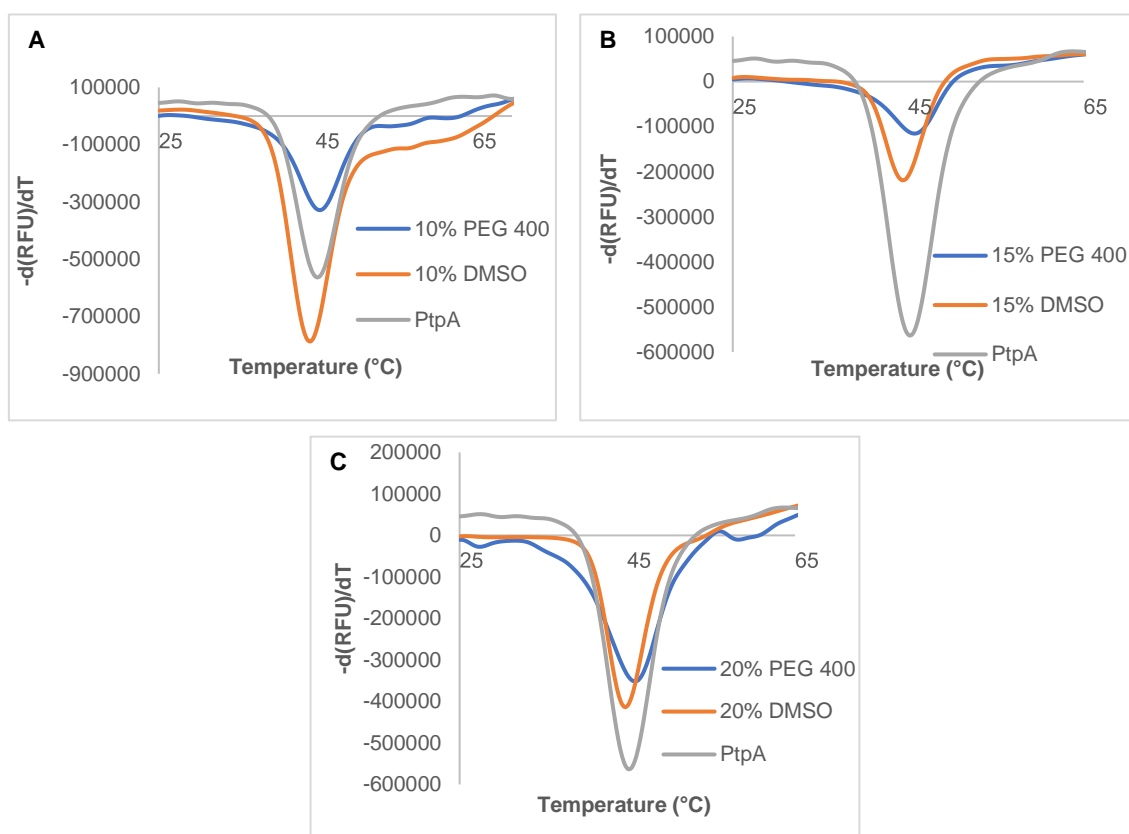
Melting Curve	$T_m$ (°C)
PtpA 4°C	44.81
PtpA - 80°C	47.64

In the third and last scenario, suitable alternative ligand solvents were investigated. During this dissertation project, several challenges and difficulties were faced in using DMSO as the organic solvent to solubilize the ligands under study. DMSO is known to be a harsh solvent, a feature which, as explained, affects the protein stability. Since a compromise must be reached to enable the protein stability and the solubilization of the compounds, testing other solvents seemed to be an important aspect to be explored, eventually enabling higher ligand concentrations in further protein-ligand interaction studies without jeopardizing the protein stability.

The well-known PEG 400, a member of the polyethylene glycol family, has been widely used as a solvent for drug delivery, with multiple applications in the pharmaceutical industry. According to literature, PEG 400 was used, as an alternative to DMSO, to solubilize ellagic acid. In fact, the compound solubility was much higher in PEG 400 (0.6 mL to dissolve 5 mg) rather using DMSO (2 mL to dissolve 5 mg), which prompt us to consider it as an alternative in further studies.<sup>87</sup> Firstly, aiming to infer about the effect of PEG 400 on PtpA and the respective  $T_m$  values

comparing to DMSO, a TSA experiment was planned testing three distinct concentrations – 10%, 15% and 20%.

Analyzing the results represented in Table 4.6, the protein  $T_m$  values are higher in the presence of PEG 400 rather than DMSO. The values of  $\Delta T_m$  (+1.71°C, +1.31°C and +1.36°C for each tested PEG 400 concentration) are reflected by the positive shifts (blue curves, Figure 4.15) obtained when compared with the protein control sample (grey curves, Figure 4.15) confirming that PtpA stability is ensured. Noteworthy, even with a concentration of 20% PEG 400 (Figure 4.15, C), PtpA demonstrated to be very stable, possessing a  $T_m$  (45.96°C) higher than the presented by the control (44.60°C). On the other hand, confirming the previous assays, it was shown that DMSO affects the protein (orange curves, Figure 4.15) as reflected by the decreasing shifts of the curves when compared to PEG 400 and the control. As expected, the destabilization effects on the protein are more pronounced at higher DMSO concentrations of 15% and 20% (Figure 4.15 B, C).



**Figure 4.15** – Effect of 10% (A), 15% (B) and 20% (C) of PEG 400 (blue curves) and DMSO (orange curves) concentrations in PtpA thermal stability. Protein stability is ensured in presence of PEG 400, in which positive melting shifts occurred towards the control (grey curves). By the contrary, protein stability is affected by the presence of DMSO as reflected by the negative-shifted curves.

**Table 4.6** – Melting temperatures (°C) obtained for PEG400 and DMSO experiments at 10% and 15% concentrations.

Melting Curve	T <sub>m</sub> (°C)	ΔT <sub>m</sub> (°C)
PtpA	44.60	-
PtpA-10% PEG 400	46.31	+1.71
PtpA-10% DMSO	44.31	-0.29
PtpA-15% PEG 400	45.91	+1.31
PtpA-15% DMSO	43.91	-0.69
PtpA-20% PEG 400	45.96	+1.36
PtpA-20% DMSO	44.17	-0.43

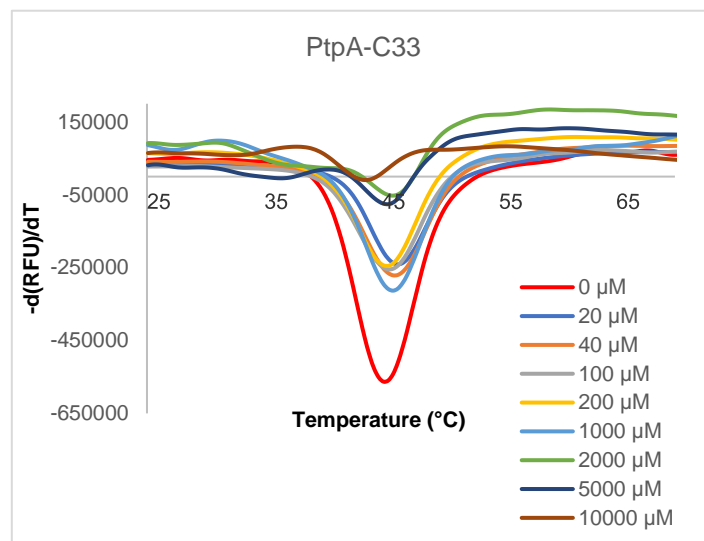
Thus, PEG 400 has proved to be an attractive candidate for new approaches prompting to new experiments to overcome the restrictions and the difficulties faced with DMSO. Further studies should be explored testing PEG 400 as the organic solvent to solubilize the studied ligands.

#### 4.4.2.2 PtpA-ligand Interactions

As aforesaid, TSA experiments were also performed aiming to provide insights on the identification of protein-ligand interactions. Although its lower sensibility when compared to other techniques, TSA possesses great advantages such as requiring less amounts of protein, allowing a faster experimental preparation and execution, and the possibility of testing many different conditions in a single assay. It is assumed that the existence of binding of the respective inhibitor candidates to PtpA would affect the protein stability, usually stabilizing it, increasing its resistance to thermal denaturation. Thus, in a PtpA-ligand interaction, a shift in protein melting temperature (T<sub>m</sub>) towards higher values should occur indicating the mentioned increase in stability.

The described approach was attempted to detect the binding of the ligands under study (C33, C37, R06, Lap04, Lap11, Azaindole 1 and Azaindole 15) toward PtpA, testing increasing concentrations of ligand - protein:ligand ratios of 1:1; 1:2; 1:5; 1:10; 1:50; 1:100; 1:250; 1:500 - with a constant protein concentration (20 μM).

Observing the melting curves profile (Figure 4.16), C33 inhibitor candidate did not display a significant shift in the protein T<sub>m</sub> (Table 4.7) when compared to the control sample (protein without any ligand). A similar behaviour was observed in the melting curves of the other tested chalcones compounds (C37 and R06) which do not show significant variations in the melting temperature values (Appendix 7.6). Considering the presented results, TSA does not suggest the existence of interactions between the analyzed chalcones and PtpA, although such interactions cannot be discarded.



**Figure 4.16** – Melting curves of PtpA incubated with C33 compound at different concentrations. The data shows that C33 has a minimal increasing impact in PtpA thermostability with no significant increases in the melting temperatures values.

Regarding the thiosemicarbazones family, Lap11 and Lap04 compounds were submitted to an identical approach aiming to detect an interaction with the protein. By the analysis of TSA melting curves obtained with Lap11, it was possible to observe small positive shifts in  $T_m$  values (Figure 4.17, A). However, these variations (values represented in Table 4.7) do not exceed 2°C, failing to be considered as an evidence of interaction. As before, this implicates that it is not possible to prove Lap11 binding to PtpA through this experiment.

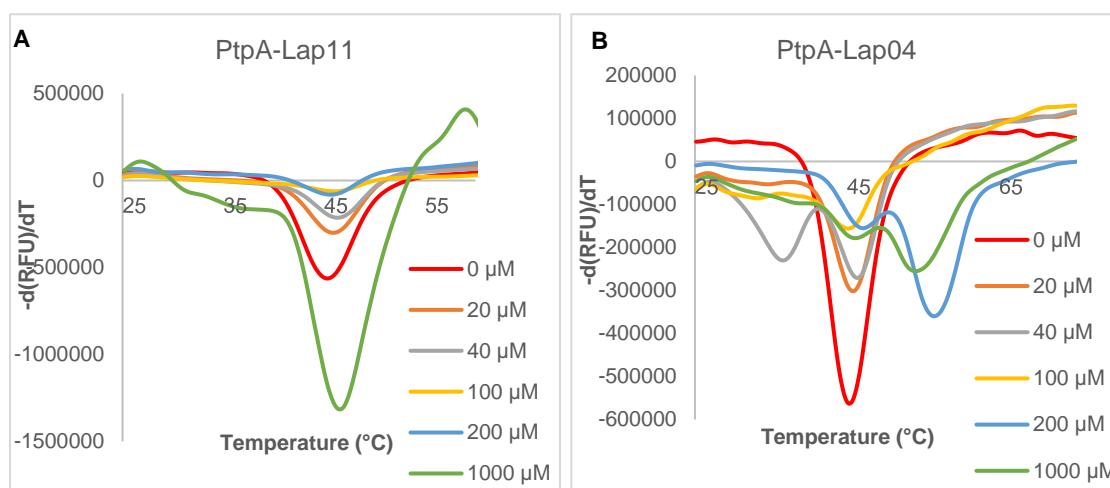
On the other hand, it was possible to confirm a significant  $T_m$  shift in the presence of Lap04 compound with  $\Delta T_m$  of 11.53 °C and 10.13°C for higher concentrations of compound (200 μM and 1000 μM) as indicated in Table 4.7. Thus, Lap04 raised attention among the ligands under study presenting the highest shift in the protein  $T_m$  compared to the control sample (*i.e.* the protein without any ligand incubation, red curve, Figure 4.17, B). Moreover, a particular attention was given to this thiosemicarbazone, since the theoretical calculations performed as part of the computational work developed in this dissertation, identified Lap04 as the candidate with the best docking scores.

To validate these findings, providing a more accurate analysis, different control assays were performed using the compound at the mentioned concentrations. The results corresponding to the lower concentrations show a melting curve with a suspicious  $T_m$  value in the range of 54°C. Some doubts have been raised since an identical value was observable when higher concentrations of Lap04 (200 μM and 1000 μM) were incubated with 20 μM of PtpA. This led to speculate that the detected  $\Delta T_m$  values in the PtpA-Lap04 trials resulted solely from the contribution of the compound. In addition, the melting curves (green and blue curves, Figure 4.17, B) corresponding to the highest compound concentrations exhibit a small bump with  $T_m$  values



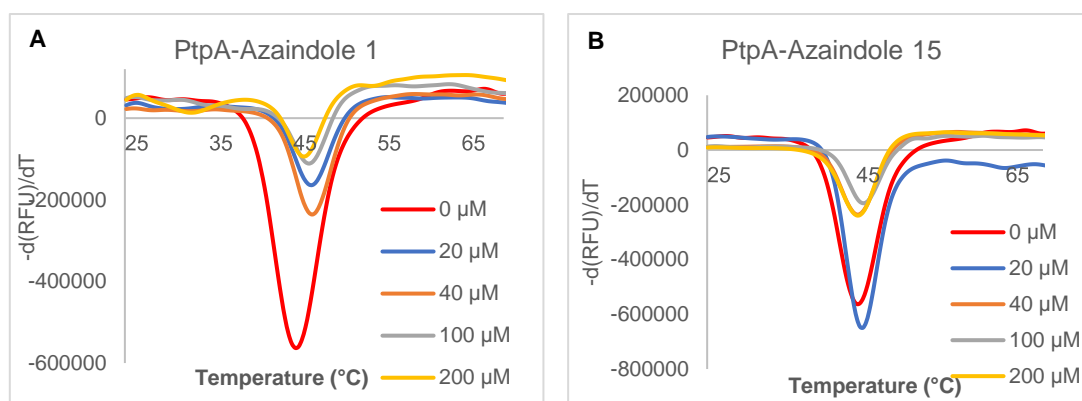
very similar to the protein control (approximately 45°C). Therefore, a hypothesis could be formulated by which, considering the significant excess, Lap04 largely contributes for the observable highest  $T_m$  values and melting curve shifts. By the same hypothesis, the much lower protein concentration, when compared to the compound, origins the small inflexion curve at 45°C (Figure 4.17, B ).

In addition, it is also important to note that the PtpA-Lap04 samples with higher concentrations of compound (2000  $\mu$ M, 5000  $\mu$ M and 10000  $\mu$ M) show an evident precipitate formation as confirmed by visual inspection of the thermofluor plate. Since these melting curves do not exhibit a well-defined melting temperature, the curves were not included in Figure 4.17, B. Such occurrence was due to the low solubility of thiosemicarbazones at the used 10% DMSO concentration highlighting the importance to find suitable alternative solvents as previously discussed.



**Figure 4.17** – Melting curves of PtpA incubated with Lap11 and Lap04 thiosemicarbazones at different concentrations. **(A)** Compared to the control (red curve), Lap11 does not display significant shifts in the  $T_m$  values of PtpA. **(B)** TSA curves in the presence of Lap04 in which is a positive shift was detectable in the highest compound concentrations (blue and green curves).

A final TSA experiment was carried out to observe possible binding of two synthesized azaindoles (1 and 15) compounds to PtpA protein. Despite the unpromising calculations from docking, these two compounds were preliminary used in order to optimize the experimental conditions required to other azaindoles in future experiments. As depicted in Figure 4.18 and Table 4.7, small variations in  $T_m$  were detected even if, considering the instrument sensitivity, apparently not significant enough to prove the ligand binding. These results were expected, due to the aforementioned low binding affinity predicted by the molecular docking, and new TSA experiments with the most promising azaindoles are required.



**Figure 4.18** – Melting curves of PtpA incubated with Azaindole 1 (**A**) and Azaindole 15 (**B**) at different concentrations. Both compounds do not affect the PtpA thermostability as significant  $T_m$  shifts were not observed.

**Table 4.7** –  $T_m$  and  $\Delta T_m$  values obtained for C33, Lap11, Lap04, Azaindole 1 and Azaindole 15 compounds.  $\Delta T_m$  values were calculated considering the protein control in which no ligand was added to the protein.

Compound	Concentration ( $\mu$ M)	$T_m$ (°C)	$\Delta T_m$ (°C)
C33	0	45.16	-
	20	46.37	+1.21
	40	45.97	+0.81
	100	45.57	+0.41
	200	45.57	+0.41
	1000	45.97	+0.81
	2000	45.97	+0.81
Lap11	0	45.16	-
	20	46.02	+0.86
	40	46.42	+1.26
	100	46.02	+0.86
	200	45.61	+0.45
	1000	46.42	+1.26
Lap04	0	45.16	-
	20	45.47	+0.31
	40	45.87	+0.71
	100	46.68	+1.52
	200	55.72	+11.53
	1000	54.29	+10.13
Azaindole 1	0	45.16	-
	20	47.26	+2.10
	40	47.29	+2.13
	100	46.85	+1.89
	200	46.46	+1.3
Azaindole 15	0	45.16	-
	20	45.27	+0.11
	40	45.48	+0.32
	100	45.88	+0.72
	200	46.29	+1.13

Taking into account the discussed results regarding the protein-ligand interaction studies, TSA experiments were not insightful to reveal if these compounds interact with PtpA. This apparent lack of interaction was not expected, especially for the chalcones and thiosemicarbazones families, as these compounds are reported as good inhibitors of PtpA. Since TSA sensitivity may not be the best to accurately detect a binding interaction, we decided to use a more sensitive technique – Microscale Thermophoresis.

#### **4.4.3 Microscale Thermophoresis**

Microscale thermophoresis (MST) allows for quantitative analysis of protein interactions. The technique is based on thermophoresis, defined as the direct motion of molecules in temperature gradients. Since thermophoresis is influenced by binding-induced changes of various molecular properties, a binding event in turn results in a quantifiable fluorescence variation. In addition, MST benefits from very low sample consumption and short measurement times, which makes it a widely applicable approach to quantify interactions.<sup>88,89</sup>

##### **4.4.3.1 Protein Labelling *via* His-tag**

One of the labelling strategies in MST assays is the His-tag labelling approach in which His-tagged biomolecules can be labelled site-specifically with a tris-NTA-conjugated NT fluorophore. Therefore, the tris-NTA moieties conjugated with NT-647 dye binds to the protein histidine-tag representing the easiest and fastest methodology to properly label proteins for interaction studies. The fluorescence signal of the target (protein) needs to be within 200-2000 fluorescence counts range to be considered as an efficient labelling. Also importantly, adsorption to the capillary walls should be prevented since it can alter the results in the next MST trials.

The PtpA labelling was efficiently achieved with an average of 1067 fluorescence counts which is within the required values previously mentioned. No adsorption of the fluorescent target to the capillary walls was detected, ensuring the sample quality for the next experiments. Importantly, the indicated labelling confirmation was performed with the protein at 50 nM (MST Led – 20%). The very robust fluorescence detected prompts to further experiments with PtpA at 25 nM (MST Led – 40%), which allows testing higher protein-ligand ratios.

##### **4.4.3.2 Binding Check Assay**

The binding check assay consists on an experiment in which the signal of the target (protein) and the complex (protein-ligand) are compared to give a qualitative result whether there is an interaction. Since the main purpose of MST is to detect potential protein-ligand interactions, it is recommended to firstly perform a binding check experiment. In addition, this step also allows assessing different quality parameters including sample aggregation, surface adsorption, fluorescence intensity, photobleaching and homogeneity.

Two major parameters provided by the binding check experiment should be analyzed: the MST traces profiles and the signal-to-noise ratio values.

The analysis of the MST traces behaviour is crucial in a binding check experiment. These traces show the change in fluorescence intensity of all capillaries over time, while a temperature increase is applied by the Infra-red laser, reflecting the MST signal of the fluorescent target molecule (protein) and the complex (protein-ligand). The difference between the MST signal of the target and the complex is given by the response amplitude. Secondly, the signal-to-noise ratio parameter is used to evaluate the quality of the binding data and it is defined as the response amplitude divided by the noise of the measurement. More than 5 is desirable while more than 12 reflects an excellent interaction detection.

In the present study, the ligands C33, C37, R6, JE02, Lap04, Lap11 and the azaindoles 1, 15, 16 were submitted to a binding check experiment with PtpA and the results are represented in Table 4.8.

Several approaches were implemented to enhance the success of the experiment including different protein-ligand incubation times – 0 and 20 minutes, at room temperature - and decreasing the protein concentration to 25 nM to ensure a higher amount of compound available to interact with the protein.

**Table 4.8** – MST binding check assay results regarding the complex (protein-ligand) with the respective signal-to-noise ratio and amplitude parameters.

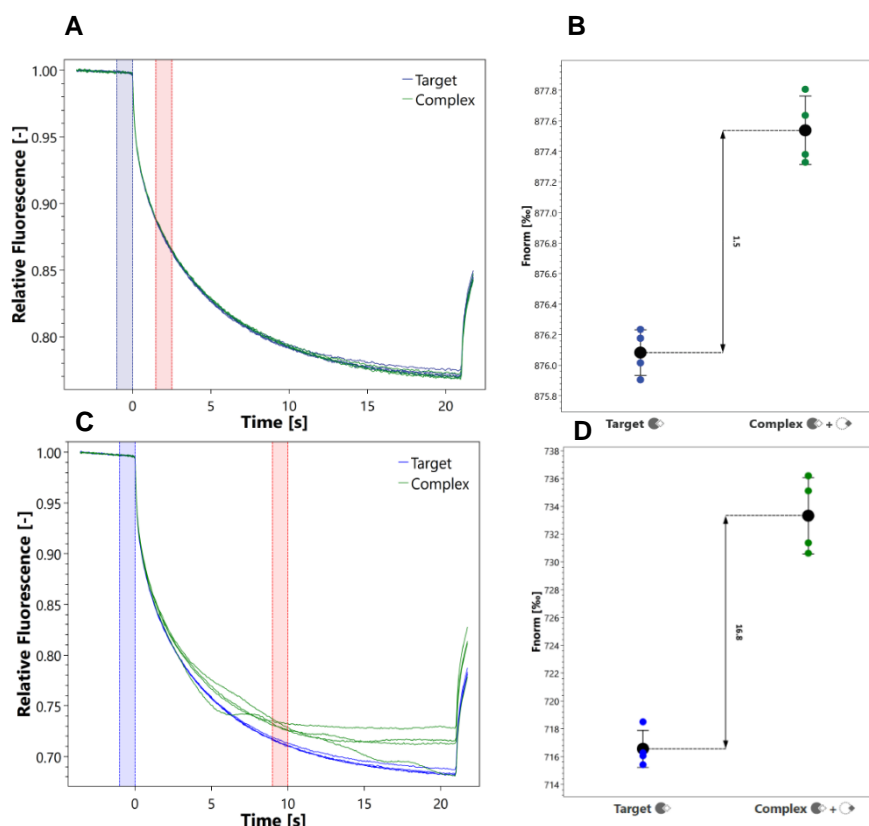
Experimental Conditions	Signal-to-Noise Ratio	Response Amplitude	Binding Result
50 nM PtpA + 0.13 mM C33	1.5	0.9	No binding detected
25 nM PtpA + 0.13 mM C33	3.0	1.2	No binding detected
25 nM PtpA + 0.13 mM C33 20 minutes incubation	7.8	1.5	No binding detected
25 nM PtpA + 0.15 mM C37 20 minutes incubation	3.6	1.5	No binding detected
25 nM PtpA + 0.12 mM R6 20 minutes incubation	0.5	0.1	No binding detected
25 nM PtpA + 0.037 mM JE02 20 minutes incubation	0.8	1.4	No binding detected
50 nM PtpA + 0.13 mM Lap04 20 minutes incubation	2.0	0.5	No binding detected
25 nM PtpA + 0.13 mM Lap04 20 minutes incubation	1.9	0.9	No binding detected
25 nM PtpA + 0.1 mM Lap11 20 minutes incubation	8.2	16.8	Possible binding detected
25 nM PtpA + 0.18 mM Azaindole 1 20 minutes incubation	1.0	0.4	No binding detected
25 nM PtpA + 0.22 mM Azaindole 15 20 minutes incubation	2.3	0.6	No binding detected
25 nM PtpA + 0.23 mM Azaindole 16 20 minutes incubation	4.2	2.9	No binding detected

Among the experiments performed with all the referred compounds, C33 and Lap11 assays (both with 20 minutes incubation time) revealed signal-to-noise ratio values of 7.8 and 8.2, respectively, which are above the ideal threshold (Figures 4.19 A and C). However, in the case of C33, the response amplitude value of the MST traces is 1.5 which is too low to conclude binding (Figure 4.19 B). By the contrary, Lap11 traces show differences between the target curves (blue) and the complex curves (green) with a response amplitude value of 16.8 (Figure 4.19 D).

For the other compounds, unpromising signal-to-noise ratio and response amplitudes values were obtained which could reflect a weak binding from the ligands to the protein (Table 4.8 and Appendix 7.7).

Due to the insufficient detected signal-to-noise ratio of these compounds, one adopted strategy was to change the MST power medium into MST power high. The higher the MST power,

the larger is the temperature increase, directly influencing the signal. Despite this strategy, the results were not improved and a possible hypothesis is that the high MST power can cause destabilization of the complex sample leading to aberrant MST traces.



**Figure 4.19** – C33 and Lap11 binding check results of MST experiment. **(A)** MST trace curves of C33 ligand incubated 20 minutes with 25 nM PtpA. **(B)** Amplitude response of C33 curves. **(C)** MST trace curves of Lap11 ligand incubation with PtpA (25 nM) for 20 minutes. **(D)** Amplitude response of Lap11 curves.

#### 4.4.3.3 Binding Affinity Assay

Typically, as aforesaid, the binding check experiment is completed in order to determine whether an interaction can be observed. If the results reveal an interaction between the target and the ligand, the final MST step – the binding affinity assay – can be achieved in order to quantify the affinity of the complex interactions (protein-ligand).

For this purpose, the MST signal of a fluorescent target is measured in the presence of diverse concentrations of ligand. More specifically, a serial dilution of ligand concentration is prepared and a constant amount of the target is added to each point of the dilution series. Measuring the MST signal for all these points (ligand concentrations), it is possible to calculate the  $K_d$  (binding affinity) which is defined as the ligand concentration at which 50% of the target is bound.

Since the binding check experiments only suggest a possible interaction with Lap11 ligand, this thiosemicarbazone was selected to perform a binding affinity experiment. Initially, the sample was placed in the capillaries and the MST measurement of the incubated complex was

performed immediately, but unsuccessful signal-to-noise ratio and response amplitude values were obtained (Table 4.9). A similar experiment was tried in which the complex was incubated for 30 minutes at room temperature prior to the measurement in order to favor the protein-ligand binding. Nevertheless, suitable signal-to-noise ratio and response amplitude values were not also obtained (Table 4.9).

Considering the lack of results of the described approaches, an alternative attempt was designed considering an important parameter that can greatly influence the binding process, either favouring or disfavouring it: temperature. Thus, protein-ligand samples were incubated overnight (17 hours) at three different temperatures: 4°C, 20°C and 37°C. The signal-to-noise ratio obtained for these assays was too low and the response amplitude was too small compared to the noise. Consequently, these attempts also failed to get optimal signal-to-noise ratio values, as summarized in Table 4.9, and proper results were not achieved preventing the determination of  $K_d$  based on the dose-response curve.

**Table 4.9** – MST binding affinity results for PtpA incubated with Lap 11 compound for 30 minutes (room temperature) and overnight (at 4°C, 20°C, and 37°C) with the respective signal-to-noise ratio and amplitude parameters.

Experimental Conditions	Incubation conditions	Signal-to-Noise Ratio	Response Amplitude	Result
25 nM PtpA + 0.1 mM Lap11	Instantly, room temperature	3.4	2.9	No binding
	30 minutes, room temperature	3.0	5.4	No binding
	17 hours, 4°C	0.4	1.2	No binding
	17 hours, 20°C	0.9	0.8	No binding
	17 hours, 37°C	2.1	0.8	No binding

Therefore, it was not possible to conclude on the interactions established by any of compounds under study with PtpA. It is plausible to speculate that the compounds are weakly interacting with the protein, and consequently this interaction is not detectable using MST technique. On the other hand, the low solubility of compounds and the use of DMSO should be considered as the major reasons that could also explain the unpromising results. Due to this limitation, the ligands' concentrations used in the experiments possibly were not enough to ensure a detectable interaction between ligands and protein.

#### 4.4.4 Urea Gel Electrophoresis

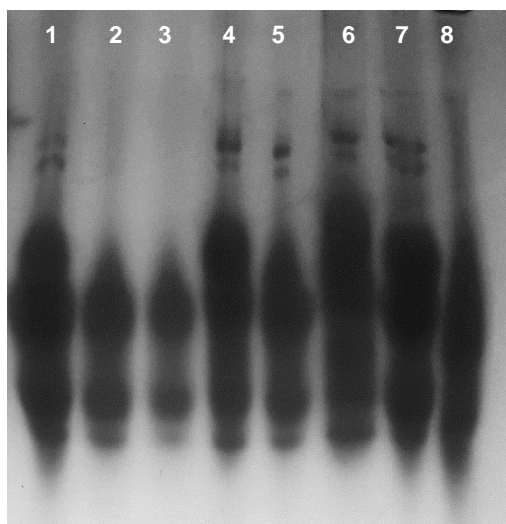
Urea gel electrophoresis is a biophysical technique which is based on the use of a polyacrylamide gel electrophoresis containing a high concentration (6 M) of urea. Urea acts as a denaturing agent and proteins will have different mobility profiles in the gel according to their conformation: compact forms are more resistant to the denaturation presenting a more pronounced migration in the gel. Therefore, this method is to be useful to detect potential protein-ligand interactions by comparing the ligand-free and the ligand-bound electrophoretic profiles especially in the cases where protein-ligand complexes assume a more closed conformation. In the case of PtpA, some kind of conformational change could occur upon ligand interaction, especially for the non-competitive inhibitor thiosemicarbazone Lap 11.<sup>90,91</sup>

In this case, urea gel electrophoresis was used to assess the binding of both chalcones and thiosemicarbazones compounds to PtpA. For that, the ligands C33, C37, R06, Lap04 and Lap11 were incubated for 2 hours with the protein in a 1:2 protein-ligand ratio. Simultaneously, the experiment also included testing three different PtpA ligand-free concentration samples – 20 mg/mL, 5 mg/mL and 2.5 mg/mL – in order to determine the ideal amount of protein to be loaded into the gel. Figure 4.20 shows the results obtained after a 3 hours run with both PtpA ligand-free and PtpA-ligand samples previously mentioned.

Focusing on the PtpA ligand-free samples (lanes 1-3), as expected, pronounced differences on the intensity of the bands are visible due to the different amounts of PtpA. In fact, the gel lane (1), corresponding to the protein at 20 mg/mL, the band is much more intense than the others but presents a smear that hampers the respective analysis. Therefore, future urea gel electrophoresis experiments should be tried with less concentrated protein samples ranging from 5 to 10 mg/mL. Nevertheless, the three used protein concentrations present a similar electrophoretic profile which supports the comparison with the PtpA-ligand samples prepared with a 20 mg/mL concentration. It was found that the mobility pattern of the electrophoretic bands corresponding to the ligand incubation (lanes 4-8) is closely related to the one observed for the ligand-free PtpA samples (lanes 1-3) suggesting that, although the ligands might bind to PtpA, such binding does not induce any relevant conformational change, namely a more compact and closed form. These results confirm the previous SAXS experiments by which a similar conformation was obtained between the ligand-free and the C33-incubated PtpA samples.<sup>72</sup>

The urea gel electrophoresis experiment corroborates the previous described TSA trials that also failed to conclude protein-ligand interactions. However, as properly discussed, the existence of protein-ligand interactions cannot be discarded. Additional urea gel electrophoresis experiments varying some parameters should be tried to further elucidate this question namely the protein concentration and the ligand incubation conditions (time and temperature).





**Figure 4.20** - Urea-polyacrylamide gel electrophoresis of ligand-free and incubated PtpA. Gel wells: 1, 2, 3 – Ligand-free PtpA at 20, 5 and 2.5 mg/mL, respectively; 4 – PtpA-C33 incubation; 5- PtpA-C37 incubation; 6- PtpA-R06 incubation; 7- PtpA-Lap04 incubation; 8- PtpA-Lap11 incubation. In all the incubated PtpA samples, the protein was at 20 mg/mL (1mM) and ligands at 2 mM.

## 4.5 Structural Characterization Studies

### 4.5.1 Crystallization Assays

#### 4.5.1.1 Ligand-free and Co-crystallization Experiments

The first ligand-free PtpA crystallization approach consisted in reproducing the previously determined condition<sup>72</sup> - 6% PEG 6K, 0.1 M HEPES 6.5 - using the vapor diffusion method, with 4  $\mu$ L hanging drops (3  $\mu$ L of protein and 1  $\mu$ L of precipitant solution) and 700  $\mu$ L of the precipitant solution in the well. In addition, the condition was further optimized manipulating several parameters including: 1) different drop proportions (1:1 and 2:1); 2) varying the precipitating agent concentration (4%-14% PEG 6K); and 3) different related precipitating agents (PEG 1.5K, PEG 3K and PEG 4K). The experiments were performed in 24-well plates with a protein concentration of 24 mg/mL.

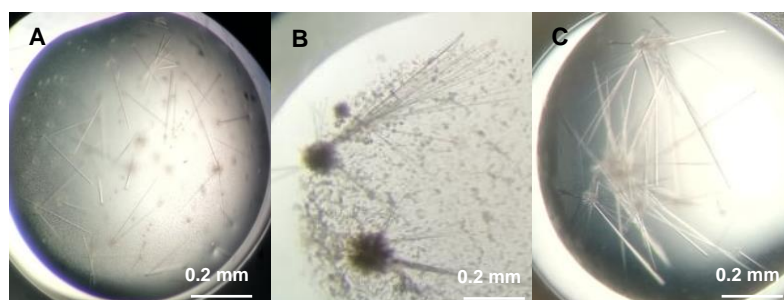
These attempts failed to produce any crystals and since the crystallization is a complex process, in which minimal changes can influence crystals appearance, initial screening assays – using both commercial and home-made crystallization screens JCSG<sup>+</sup>, JBS 1,2,3,4, JBS 5,6,7,8, PEG ION and 80! were conducted to find new suitable and reproducible crystallization hits, to be further optimized if needed. At this stage, the main goal is to test the maximum number of conditions as possible with the lowest protein amount. Such goal was successfully achieved by using a nanocrystallization robot system to prepare 96-well crystallization plates (70  $\mu$ L of protein for each assay). All the screens were tested using the vapour diffusion method and sitting-drop technique at two distinct temperatures (4°C and 20°C).

Once again, no crystal formation was observed in any of the described tests regardless the crystallization screen used. Since the finding of a suitable protein crystallization condition was

not fulfilled, subsequent soaking experiments were not possible preventing the characterization of the interactions between the protein and the compounds under study.

As an alternative to soaking and to increase the chances of protein-ligand crystals formation, co-crystallization experiments were then performed. The protein was incubated with the ligands (chalcones, thiosemicarbazones and three synthesized azaindoles) in a 1:2 mM protein-ligand ratio using three different incubation times: long (overnight – 17 hours), intermediate (2 hours) and short (10 minutes) periods. Additionally, the screen JCSG<sup>+</sup> (Appendix 7.4) was selected and to understand the influence of temperature on the crystallization process, the trials were attempted at 4°C and 20°C. Under these experiments, long thin needles were obtained after one week at 20°C, as depicted in Figure 4.21.

The obtained crystals were stabilized in a harvesting buffer - a similar solution to the precipitating solution but with a slightly higher concentration of the precipitating agent - which prevents the crystal from dissolving during the respective handling process. Then, the crystals were transferred to a suitable cryo-protectant solution (achieved by the addition of 30% glycerol to the stabilizing solution or using paratone oil) to avoid ice formation within the crystal lattice which interferes with the interpretation of the diffraction results.



**Figure 4.21** – Representation of some PtpA crystals obtained in co-crystallization trials. Needle-shaped PtpA crystals were obtained from **(A)** PtpA-C33 from 10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5; **(B)** PtpA-C37 from 10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5 **(C)** PtpA-Lap11 based on 20% PEG 8K, 0.1 M Phosphate Citrate 4.2, 0.2 M NaCl.

The crystals were flash-frozen in liquid nitrogen to minimize their degradation during the diffraction experiment and analyzed by synchrotron radiation. The Table 4.10 represents the crystallization conditions in which was possible to observe crystal formation as well as the synchrotron beamlines and the respective diffraction results.

The analysis of results reveal that unsuitable diffraction results were obtained from the C33 and C37 co-crystallization trials. The best resulting crystals exhibited a very poor diffraction while some others did not diffract at all.

On the other hand, two of the tested conditions in the co-crystallization trial with Lap11 diffracted up to 3.9 Å and 4 Å resolution. These results prompt to further extensive trials in order to improve and optimize the needle-shaped crystals morphology and mainly their diffraction quality (detailed description in section 3.5.1.2 of Materials and Methods).

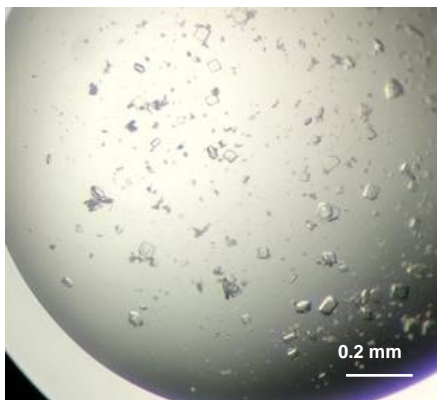
A slight resolution improvement has been achieved (3.6 Å), varying the precipitating agent concentration, even with the thin needle shape prevalence. However, the obtained

resolution value is still unsatisfactory for detailing protein-ligand interaction, compelling us to continue the optimization work.

**Table 4.10** - Sum up of the crystallization conditions and diffraction results of PtpA co-crystallizations. The conditions were provided from the JCSG+ screen at 20°C.

Compound	Incubation Time	Crystallization Condition	Beamline (Synchrotron)	Diffraction Result
<b>C33</b>	17 hours	20% PEG 6K, 0.1 M Bicine 9.0	BioMAX (MaxIV)	No diffraction
		10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5		~ 6 Å
	2 hours	10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5		No diffraction
<b>C37</b>	17 hours	20% PEG 6K, 0.1 M Bicine 9.0	BioMAX (MaxIV)	~ 5 Å
		10% PEG 8K, 0.1 M Imd 8.0		~ 8 Å
		10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5		No diffraction
	2 hours	10% PEG 8K, 8% ethylene glycol, 0.1 M HEPES 7.5		No diffraction
<b>JE02</b>	10 minutes	50% PEG 400, 0.1M Act 4.5, 0.2 M Li <sub>2</sub> SO <sub>4</sub>	BioMAX (MaxIV)	Salt
		50% PEG 200, 0.1 M Phosphate 6.2, 0.2 M NaCl		Salt
<b>Lap11</b>	2 hours	10% PEG 6K, 0.1 M Bicine 9	ID30A-3 (ESRF)	3.9 Å
		10% Isopropanol, 0.1 M HEPES 7.5, 0.2 M NaCl		No diffraction
		24% PEG 1.5K, 20% Glycerol	I04 (Diamond)	4 Å
		12% PEG 6K, 0.1 M Bicine 9		3.6 Å

Considering the PtpA-JE02 co-crystallization experiment, the resulting crystals looked morphological different, appearing as very small crystals with irregular shapes as shown in Figure 4.22. However, the diffraction results revealed that the crystals were salt. In fact, the presence of acetate and phosphate buffers in the crystallization conditions is known to increase the probability of salt crystals formation.

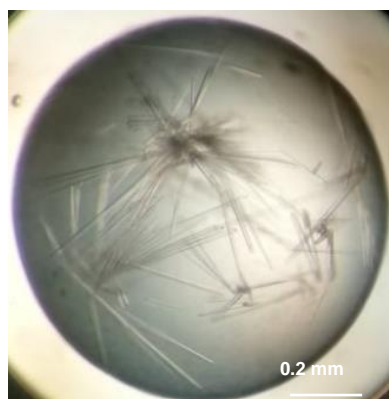


**Figure 4.22** – PtA-JE02 crystals obtained by co-crystallization, using the crystallization condition 50% PEG 200, 0.1 M Phosphate 6.2, 0.2 M NaCl. The diffraction experiment revealed that the crystal analyzed was salt rather than protein.

Therefore, is essential to invest time in more crystallization trials, using other screens to determine the best conditions to obtain quality crystals.

#### 4.5.1.2 Microseeding Experiments

As aforementioned in section 3.5.1.3, the originated crystals from the 10% PEG 8K, 8% ethylene glycol and 0.1 M HEPES 7.5 crystallization condition were selected to prepare a seed stock to be used in microseeding trials (Figure 4.23). The microseeding method is expected to improve the quality of diffraction promoting the nucleation and consequent crystal growth by adding small nuclei previously formed.

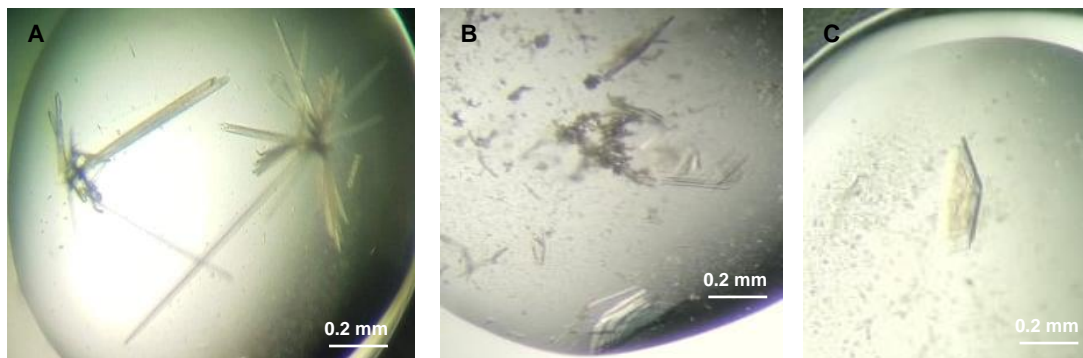


**Figure 4.23** – Needle crystals obtained from 10% PEG 8K, 8% ethylene glycol and 0.1 M HEPES 7.5 crystallization condition to perform the microseeding experiment.

One compound of each family was selected to proceed – the chalcone C33, the thiosemicarbazone Lap11 and the azaindole 16. Moreover, microseeding was also used in new ligand-free protein trials aiming to improve the chances of PtA crystals appearance, not achieved by other methods.

Crystals were obtained using the azaindole 16 during 2 hours of incubation with protein (1:2 mM protein-ligand ratio) and 1:10 seeds proportion. However, no significant differences were

verified either in their morphology (needle shaped crystals, Figure 4.24 A) or diffraction (Table 4.11). Nevertheless, microseeding allows to obtain ligand-free protein crystals with different sizes and morphologies which had never been achieved in the previously mentioned efforts (Figure 4.24 B, C).



**Figure 4.24** – (A) Resulting PtpA-azaindole 16 crystals from microseeding. (B), (C) Ligand-free protein crystals obtained by microseeding.

In Table 4.11 are represented the most promising conditions in which was possible to observe the growth of crystals using microseeding.

**Table 4.11** – Crystallization conditions and diffraction results obtained from micro-seeding experiments.

Seeding	Crystallization Condition	Diffraction Result
Ligand-free PtpA + 1:10 seeding	1.6 M $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.1 M MES 6.5	$\sim 8 \text{ \AA}$
	20% PEG 2K, 0.1 M Tris 8.5, 0.2 M TMAO	$\sim 6 \text{ \AA}$
PtpA+Azaindole16	20% PEG8K, 0.1 M CHES 9.5	$\sim 5 \text{ \AA}$
(2h incubation) + 1:10 seeding	10% PEG 8K, 8% Ethylene glycol, 0.1 M HEPES 7.5	Not measured

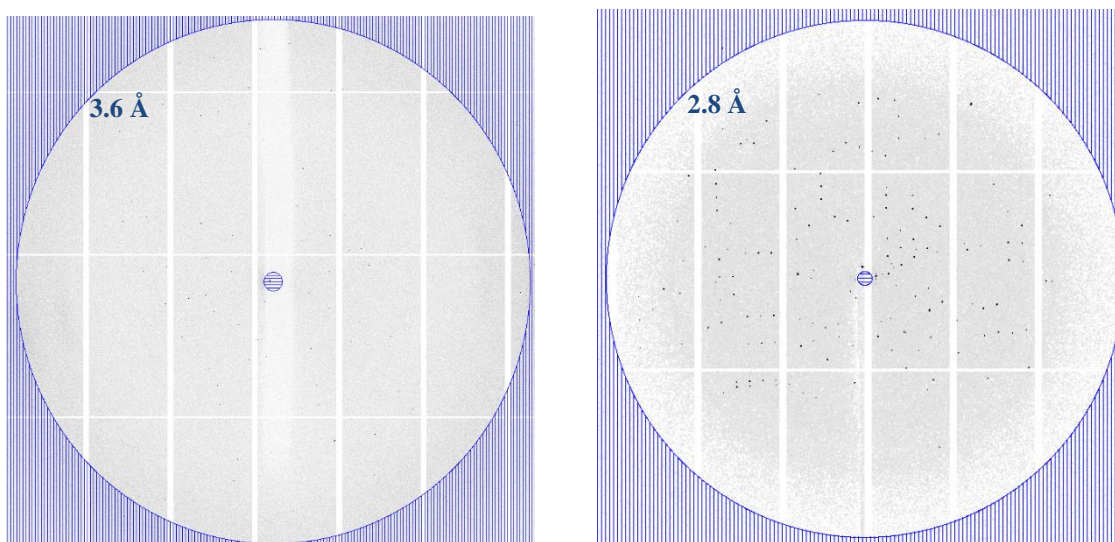
As mentioned, no improvements in diffraction quality were achieved for both incubated and ligand-free crystals. However, obtaining the first ligand-free PtpA crystals should be considered a very positive result. Thus, these conditions are promising candidates to further optimization trials, aiming to achieve better diffracting crystals.

## 4.5.2 Structure Determination and Analysis

In the next sections, a detailed structural characterization of two structures will be presented and discussed. The structures were obtained from crystallization experiments with compounds belonging to two different families – thiosemicarbazone Lap11 and chalcone C33. The dataset from C33 compound was previously obtained in the Protein crystallography Laboratory and were available for the processing step.

### 4.5.2.1 Data Collection and Structure Solution

The recorded reflections in the diffraction pattern (Figure 4.25) need to be processed in order to determine the Miller indices ( $hkl$ , coordinates in the reciprocal space) with their intensities ( $I_{hkl}$ ) and the associated error ( $\sigma_{hkl}$ ).



**Figure 4.25** – Diffraction pattern of PtpA-Lap11 crystal (left) and PtpA-C33 crystal (right). Crystals diffracted up to 3.6 Å (PtpA-Lap11) and 2.8 Å (PtpA-C33) at beamlines I04 (Diamond) and ID23-2 (ESRF), respectively.

The collected reflections were indexed with XDS program originating an hkl file with the referred parameters. Moreover, the parameters that characterize the unit cell (constants  $a$ ,  $b$ ,  $c$  and angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ) were also obtained, suggesting that the crystal belongs to the hexagonal crystal system (Table 4.12).

In order to obtain an accurate space group determination, Aimless from the CCP4 suite of programs, was used. This tool analyzes the position and the intensity of the reflections from the hkl file based on systematic absences to suggest the most probable space group that reflect the crystal symmetry. In this case, the program indicated that both crystals belong to the primitive space group P6<sub>1</sub>. Additionally, Aimless program also scales the data in which the relative intensities of all the reflections are analysed and settled in a same scale ensuring a proper analysis by the next structure resolution steps. These different intensities may be related to some factors including variations in the radiation beam intensity during the data collection, problems

with the crystal assembly or crystal damage throughout the experiment due to synchrotron radiation. Importantly, Aimless also provides a set of statistics of the data collection, which is represented in Table 4.12. Furthermore, 5% of the reflections were randomly selected in this step: such reflections will not be used for the resolution and refinement of the structure, but only will guide the calculation of the  $R_{\text{free}}$  parameter.

**Table 4.12** – Data collection statistics from the diffraction data corresponding to PtpA incubated with Lap11 ligand (left) and to PtpA incubated with C33 ligand (right). The values in parentheses correspond to the highest resolution shell.

Data Collection	PtpA-Lap11	PtpA-C33
X-ray Source	Diamond (I04)	ESRF (ID23-2)
Wavelength (Å)	0.9795	0.8730
Unit Cell Parameters (Å, °)	a=b=162, c=53.8 $\alpha=\beta=90$ , $\gamma=120$	a=b=163.5, c=53.7 $\alpha=\beta=90$ , $\gamma=120$
Space Group	P6 <sub>1</sub>	P6 <sub>1</sub>
Resolution (Å)	46.77-3.60 (3.94-3.60)	47.20-2.80 (2.95-2.80)
$R_{\text{pim}}$	0.062 (0.688)	0.050 (0.548)
$\langle I/\sigma(I) \rangle$	10.7 (1.4)	12.5 (1.8)
CC <sub>1/2</sub>	0.99 (0.72)	0.99 (0.76)
Total number of observed reflections	198319 (48403)	413307 (61269)
Total number of unique reflections	9598 (2256)	20549 (2959)
Mosaicity (°)	0.16	0.07
Completeness (%)	100 (99.9)	100 (99.9)
Multiplicity	20.7 (21.5)	20.1 (20.7)
Molecules per ASU	4	4
Mathews Coefficient, $V_M$ (Å <sup>3</sup> /Da)	2.54	2.58
Solvent Content (%)	51.63	52.42

$$R_{\text{pim}} = \frac{\sum_{\text{hkl}} [1/(N-1)]^{1/2} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}, \text{ where } N \text{ is the multiplicity measured.}$$

$$\text{Multiplicity} = \frac{\text{Total number of observed reflections}}{\text{Total number of unique reflections}} \quad \text{Solvent Content} = 1 - \frac{1.23}{V_M} \quad V_M = \frac{\text{Unit Cell Volume}}{\text{Unit Cell Molecular Weight}}$$

The information presented in the previous table, allows to draw some conclusions on data quality. Firstly, the signal to noise ratio parameter ( $\langle I/\sigma(I) \rangle$ ) indicates how many times the intensity of reflections is higher than noise. In this case, the overall intensity of reflections is 10 and 12 times higher than the associated noise in PtpA-Lap11 and PtpA-C33 data, respectively (and the values for the highest resolution shell are 1.4 and 1.8).

Another important parameter indicative of data quality is the correlation coefficient (CC<sub>1/2</sub>) that also provides information regarding the signal noise ratio. As the name hints, the measured reflections are randomly divided into two halves followed by the calculation of the mean intensities

of each data subset. For the highest resolution shell, this parameter must be above 0.5 which is observable for both datasets.

The multiplicity parameter expresses how many times a given reflection was measured, achieved by knowing the total number of reflections and the number of unique reflections in the dataset. Thus, data with high multiplicity present greater confidence allowing a better calculation of the reflection intensities, as well as their associated error, which is verified in the crystals in study (21.5 and 20.7 for PtpA-Lap11 and PtpA-C33, respectively).

Completeness of the data is another important quality parameter, which indicates the percentage of measured reflections compared with the total number of possible reflections at a given resolution and should be as close as possible to 100%. In both cases, the value is 99.9%, which is very close to 100%, indicating that the data collection strategy was well executed and complete datasets were obtained.

Mosaicity illustrates the internal organization level of the crystal by representing the degree of disorder between the crystal unit cells in the three-dimensional space. The low mosaicity values exhibited by the two structures, namely  $0.16^\circ$  and  $0.07^\circ$  for PtpA-Lap11 and PtpA-C33, respectively, are indicative of well-ordered crystals.

The Matthews coefficient, designated by  $V_M$ , allows crystal content analysis by relating the unit cell volume with the molecular weight of the protein. From this parameter, the solvent content present in the crystal is calculated and, knowing the molecular weight of the protein, it is also possible to predict the number of molecules present in the asymmetric unit, which is essential for the structure determination.  $V_M$  calculations suggest that both analysed crystals present a solvent content of approximately 50% and 4 molecules in the asymmetric unit.

$R_{pim}$  is one more parameter used to assess the experimental data quality reflecting the agreement between the equivalent reflections. This value should be in 5-10% range which is verified in both structure cases (6.2% and 5.0% for PtpA-Lap11 and PtpA-C33, respectively).

At this stage of data processing, the amplitude value of each reflection was calculated which is used for solving the structure, along with the phase angle. However, the phase information is lost during the data collection, leading to the so-called “phase problem” in crystallography. Then, it is necessary to calculate the phase angle of each reflection in order to obtain the necessary structure factors which allow the electron density map calculation. Molecular Replacement (MR) method, mentioned in section 1.7.4, was used to overcome the phase problem, since a PtpA model is deposited in the PDB (code: 1U2P). Although the space group ( $P2_12_12_1$ ) of the deposited 1.9 Å structure is different, it belongs to the same organism (*Mtb*), making it the most suitable model for this study. The chloride ion present in active site and the water molecules of the deposited model were removed ensuring no bias from these structure elements.

The phase determination and calculation of the electron density was performed by Phaser program. Considering the number of molecules in the asymmetric unit proposed by the Matthews coefficient, Phaser tries to correctly orient and position the model in the calculated electron



density, using the experimental amplitudes and calculating Patterson maps, through rotation and translation operations. In general, the rotation operation finds the orientation of the model in the asymmetric unit and the translation operation finds, for that specific orientation, the molecule in the asymmetric unit certifying that the model fits in the asymmetric unit.

The results from Phaser were analysed, focusing special attention to the statistics generated by rotation and translation operations: Log Likelihood Gain (LLG) and Z-score parameters. On one hand, the LLG parameter quantifies how much better the experimental data can be explained using the model in question rather than using a random distribution of the same atoms. On the other hand, the Z-score indicates how many standard deviations the found solution is above the average and is used to gauge about the structure solving outcome. These value must be greater than 4 for the rotation function (RFZ) and 8 for the translation function (TFZ) to be considered a correct solution. These values were analysed confirming simultaneously if the proposed model was properly explained by the electron density in COOT program.

Surprisingly, despite the Mathews coefficient suggestion - four molecules in the asymmetric unit - only two molecules were found by Phaser. In fact, other two additional molecules were identified, but both were largely misplaced presenting significant portions with no corresponding electron density and originating a doubtful crystal packing.

Interestingly, Phaser suggests the same space group than Aimless –  $P6_1$  – which, considering the previous observations, appears to be incompatible with the presence of four molecules in the asymmetric unit. This prompts to consider if the space group was correctly determined and an additional structure solving experiment was attempted by selecting the option “All Alternative Space Groups” in Phaser. Once again, similar results were obtained suggesting the  $P6_1$  space group and the presence of two molecules in the asymmetric unit - 75.82% (Lap11) and 76.21% (C33) of solvent content. The manual inspection of the resulting electron density maps in COOT confirmed such hypothesis, with no unexplained electron density. Similarly, the values represented in Table 4.13 – translation values higher than 8 and rotation values of 4.6 and 4.3 – are considered quite reasonable.

**Table 4.13** – Z-score and LLG values from rotation and translation functions obtained using the Phaser program for PtpA-Lap11 and PtpA-C33 datasets.

Parameters	PtpA-Lap11	PtpA-C33
Z-score (rotation)	4.6	4.3
Z-score (translation)	22.8	14.6
LLG	468	285

Two files from Phaser are created – a new mtz file which contains information about the electron density and a pdb file containing the information about the position of all the atoms present in the model (x,y,z real space coordinates).

#### 4.5.2.2 Refinement of PtpA-Lap11 and PtpA-C33 Structures

Once the initial electron density map is obtained, the respective refinement must be carried out. The refinement step aims to minimize the differences between the observed ( $F_{\text{obs}}$ ) and calculated ( $F_{\text{calc}}$ ) structure factors, allowing the progressive adjustment of the model to the electron density.

An initial refinement cycle was performed using the Low Resolution Pipeline (LORESTR) from CCP4 for the two structures in study. After this first refinement, both models and the respective electron density maps ( $2F_o - F_c$  and  $F_o - F_c$ ) were examined in COOT and it was possible to observe an electron density blob near the active site in both structures. In the PtpA-Lap11 structure, an extra density blob close to His120 residue in molecule B was also identified (further discussed in the next subchapter).

Both structures present acceptable R-factor and  $R_{\text{free}}$  values for an initial refinement. Next steps of the refinement process included the fitting of the protein residues into the electron density as well as the addition of water (4 molecules in the PtpA-Lap11 and 44 molecules in the PtpA-C33 structure) and glycerol molecules (3 molecules in the PtpA-C33 structure). Additionally, depending on the structure, the ligand Lap11 was modelled in the density blob near His120B while two fragments of C33 were modelled at the active site. A detailed discussion of this subject can be found in the next subchapter.

Both structures are still under refinement, however a preliminary analysis of the process is possible (Table 4.14). Regarding PtpA-Lap11 structure, R-factor and  $R_{\text{free}}$  values did not show a significant improvement through the refinement steps which can be explained by the difficulties faced during the model building due to the poor structure resolution. On the other hand, in PtpA-C33 model, it was possible to observe a gradual decrease of the R-factor to a final value of 20.8%. Similarly,  $R_{\text{free}}$  also progressively decreases to 24.4% through the refinement process indicating a correct model building.

The validation of the structures was done by model geometry analysis, an information given by the RMSD values and the Ramachandran plot. The RMSD values for bond length and bond angle are within the expected range (less than 0.02 Å and 3°, respectively), which corroborate the geometry validation of both the structures.

Looking at the Ramachandran plot, representing the analysis of the conformational dihedral angles of the main chain, it can be confirmed that the vast majority of residues (above 97% for both structures) are in favoured regions.

**Table 4.14** - Refinement statistics for PtpA-Lap11 and PtpA-C33 structures using Low Resolution Pipeline (LORESTR) from CCP4 suite of programs.

Parameter	Final Refinement	Final Refinement
	PtpA-Lap11	PtpA-C33
R-factor (%)	23.9	20.8
R <sub>free</sub> (%)	26.0	24.4
RMSD bond length (Å)	0.012	0.018
RMSD bond angle (°)	1.91	2.89
Ramachandran Plot (%)		
Residues in favoured regions	97.73	97.05
Residues in allowed regions	2.27	2.95
Residues in disallowed regions	0	0

R-factor =  $\frac{\sum ||F_{\text{calc}}| - |F_{\text{obs}}||}{\sum |F_{\text{obs}}|} \times 100$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure factor amplitudes.

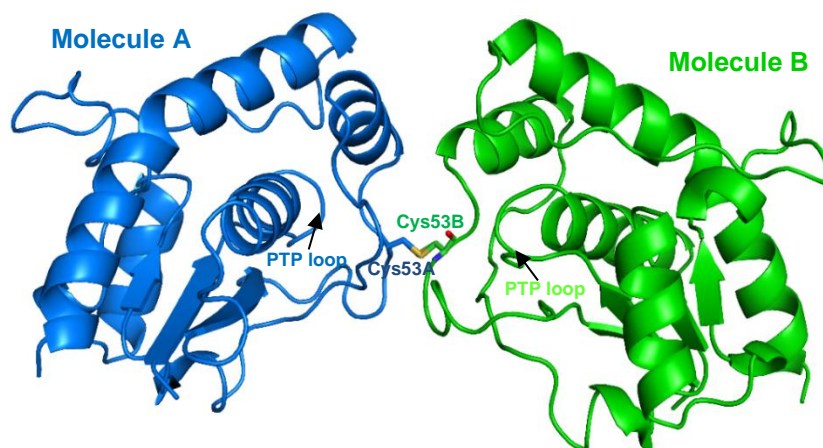
R<sub>free</sub> is calculated for a randomly chosen 5% of the reflections for each dataset.

## 4.5.3 Characterization of the Structures

### 4.5.3.1 PtpA-Lap11 Structure

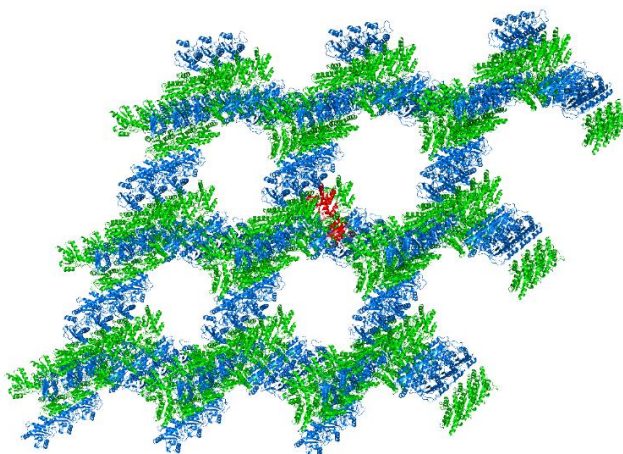
Analyzing the final structure of PtpA-Lap11, it is possible to observe two molecules, each one composed by a central four-stranded parallel  $\beta$ -sheet flanked by  $\alpha$  helices on both sides, very similar to the model deposited in the PDB (Figure 4.26). Both structures share an RMSD value at the C $\alpha$  positions of 0.35 Å. Furthermore, the typical PTP loop lies at the C-terminus of the first  $\beta$ -sheet and the N-terminus of the first  $\alpha$ -helix.

In the electron density map, it was also possible to observe the existence of a crystallographic dimer linked by a disulfide bond between the Cys53 residues from the neighbouring PtpA molecules of the crystal lattice (Figure 4.26). The formation of such disulfide bond is unexpected since a reducing agent (DTT) was used during the purification protocol as well as in the protein final buffer. Moreover, as properly discussed in Chapter 4.4, biophysical studies suggest that PtpA is a monomer in solution. The disulfide bond is not present in the structure reported in literature and it is assumed to be a crystallographic artefact, but more detailed studies are required to draw a clearer picture of its possible biological function.



**Figure 4.26** – Representation of PtpA-Lap11 crystal structure. A disulfide bond is noticeable established between Cys53 of molecule A (represented in blue cartoon) and Cys53 of molecule B (green, cartoon). Picture has been prepared using PyMOL.

As previously indicated, the determination of the correct space group was not straightforward. The suggested  $P6_1$  space group with two molecules in the asymmetric unit led to a crystal packing with large hexagonal-shaped empty cavities ( $> 100 \text{ \AA}$ ) as represented in Figure 4.27. Such crystal organization is compatible with the calculated electron density and could explain the low resolution obtained for the PtpA-Lap11 crystal structure, since it indicates a high crystal solvent content (around 75.82% as predicted by the Matthews coefficient for 2 molecules in the asymmetric unit). This hypothesis corroborates the fragile nature of protein crystals observed during handling and freezing processes.



**Figure 4.27** – Crystal packing of PtpA-Lap11 structure. The distances between the positioned molecules in the crystal packing are indicative of a high solvent content. In red are represented the original asymmetric unit molecules. Figure prepared in PyMOL.

After a careful inspection of the electron density map, a density blob was clearly visible, in both  $2F_o - F_c$  and  $F_o - F_c$  maps, at the active site near to the PTP loop. Considering the low resolution of the crystal structure ( $3.6 \text{ \AA}$ ), at the current refinement stage, it was not possible

elucidate the nature of this density blob. However, some hypothesis could be formulated regarding its existence. On one hand, significant disordered and ambiguous parts of the density were found and many difficulties were faced in the main chain residues tracing as well as the respective side chains adjustment (e.g. Arg17) in the density map. Thus, the interpretation of the electron density was a challenging task and it becomes unclear if the presence of the blobs results from an uncorrected model building (e.g. existing different rotamers from those proposed).

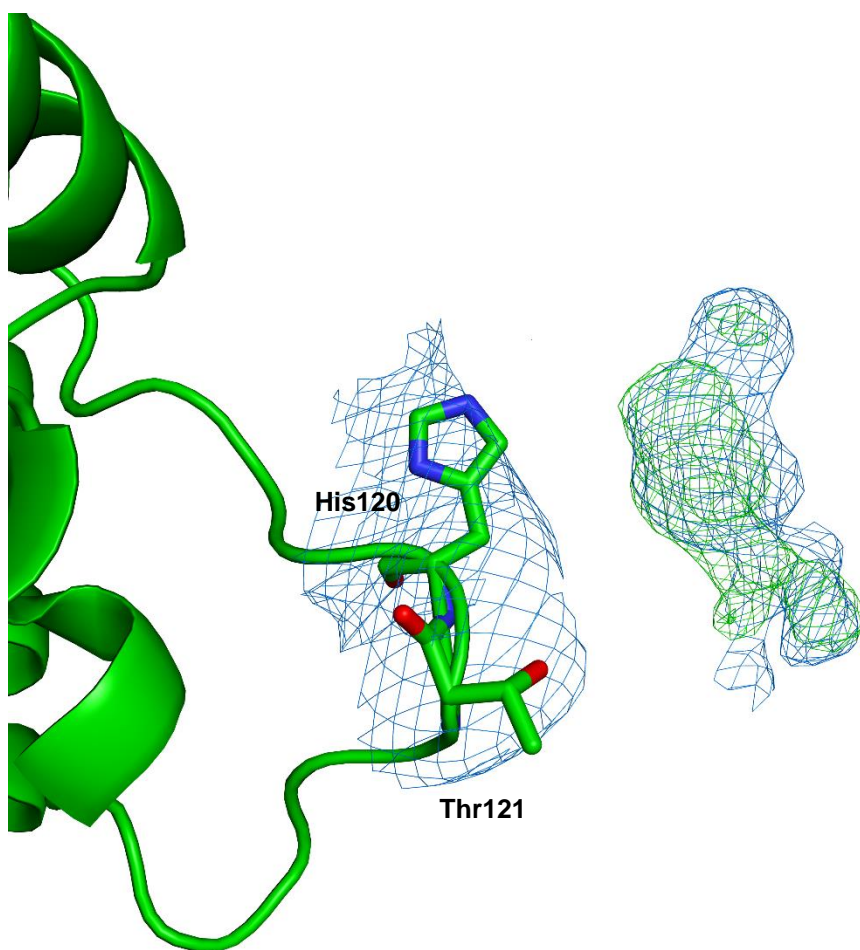
On the other hand, it was hypothesized that the densities correspond to components present in the protein buffer (e.g. HEPES) or originated from the crystallization experiment (PEG and glycerol are suitable candidates). Given the Lap11 allosteric binding behaviour reported in literature and suggested by molecular docking, the presence of the ligand in this region is not expected. Nevertheless, it was modelled, but unsuccessfully, even for small ligand's fragments.

Once again, the achieved low resolution does not allow a clear clarification of the blobs nature and future work is required relying on the optimization of the crystallization conditions in order to get new and suitable crystals for a more detailed comprehension of the subject.

By a continuous examination of the model and electron density map in COOT program, an interesting feature was observed. The PtpA-Lap11 crystal structure shows an extra density blob near the His120 residue in molecule B. Considering the molecular docking calculations with Lap11, previously discussed in section 4.3.1, the results revealed that Lap11 is an inhibitor able to bind to an adjacent site in which His120 residue is included. (Figure 4.28).

In order to elucidate the mentioned hypothesis, the ligand structure was modelled in this region. A refinement cycle was performed and the result was analyzed in COOT revealing that electron density is missing in some portions of the ligand. Moreover, the B factors of the ligand were also analyzed revealing to be very high ( $252.9 \text{ \AA}^2$ ) when compared to the protein values ( $154 \text{ \AA}^2$ ). These results, along with the initial difficulties to fit the ligand into the density blob, were not conclusive remaining the doubts about the Lap11 binding.

Undoubtedly, more refinement cycles are necessary to provide conclusive insights about this density. Equally important, obtaining good diffraction crystals reveals to be crucial to prove the existence of this density in high resolution structures to have a clearer and stronger evidence of the eventual ligand presence.

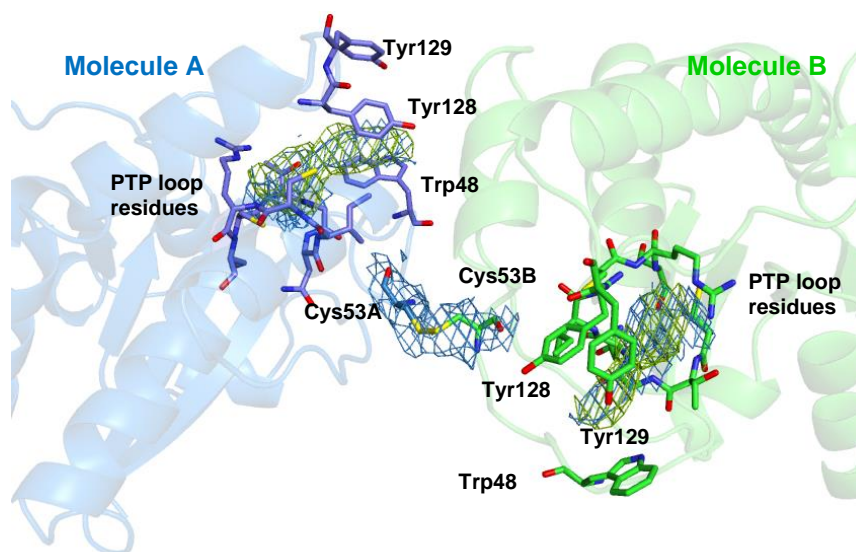


**Figure 4.28** –Structural representation of the electron density blob near to the protein allosteric site in the PtpA-Lap11 crystal structure. The blob was found only in one of the two molecules of the asymmetric unit (molecule B) next to the His120 residue. In blue is represented the  $2F_o-2F_c$  electron density map contoured at  $1\sigma$  and in green the  $F_o-F_c$  density map contoured at  $3\sigma$ . Figure prepared in PyMOL.

#### 4.5.3.2 PtpA-C33 Structure

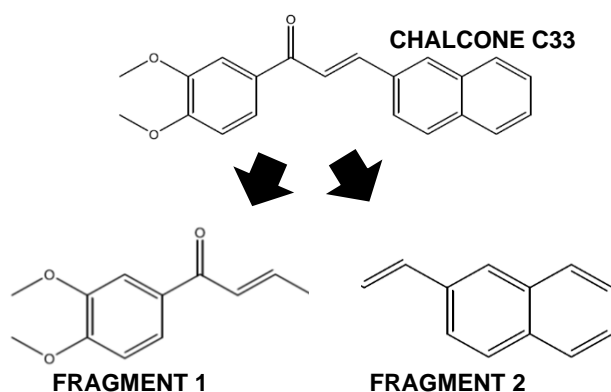
The C33 crystal structure was obtained using the crystallization condition based on 10% PEG 6K, 0.1 M HEPES 7.5. The compound was soaked at 20 mM into PtpA crystals (PtpA concentration: 17 mg/mL) and the soaking experiment was prolonged for 4 hours. Remarkable differences in crystals morphology were not observed, when compared to the crystals originated by co-crystallization trials. Additionally, a better resolution value was achieved and the structural similarities with Lap11 structure were also evident - same secondary structure elements, crystal packing and the disulfide bond were confirmed in both structures (Figure 4.29).

In both electron density maps  $2F_o-F_c$  and  $F_o-F_c$ , it was possible to observe a suspicious blob near the active site pocket in both molecules of the asymmetric unit. After an initial refinement, a careful and detailed examination in this protein region reveals that the sidechains of the residues were correctly assigned which, as explained, was a problematic issue in PtpA-Lap11 structure. This led to consider that this blob could correspond to the ligand C33 or a possible fragment.



**Figure 4.29** – Representation of the PtpA-C33 crystal structure with visible density blobs in the active site region. The disulfide bond between Cys53 from both molecules is also depicted. Electron density maps  $2F_o - F_c$  (blue) and  $F_o - F_c$  (green) are contoured at 1 and 3  $\sigma$ , respectively. The two asymmetric unit molecules A and B are represented as blue and green cartoons. Figure prepared in PyMOL.

In order to ascertain the question, the ligand was modelled in the respective density blob. However, considerable challenges were faced in this first ligand modelling attempt, due to the large size of the ligand and a significant part of the molecule (last two rings) miss any electron density. The unsuccessful modeling for the whole ligand led to adopt another strategy - considering the size and shape of the blobs in both molecules, two fragments of the ligand were proposed, as schematically represented in Figure 4.30.

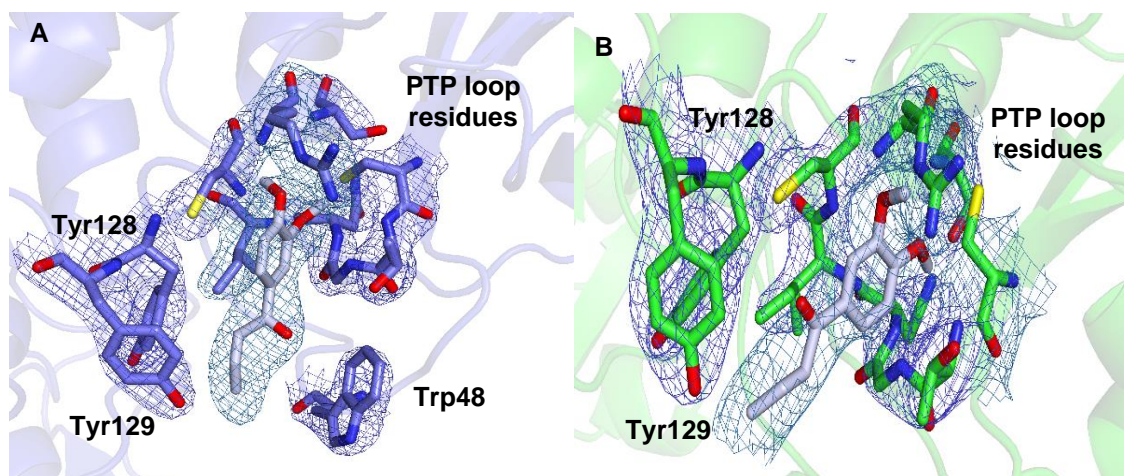


**Figure 4.30** – Representation of the structures of the two fragments generated from C33 ligand.

These two C33 fragments were modelled in the density blob and different refinement steps were conducted to test the possible fitting of each one in the electron density map. Firstly, the fragment 1 was modelled in both molecule A and molecule B with the o-methoxy and m-



methoxy groups directed into the deeper cavity, pointing towards the active site residues (Figure 4.31).



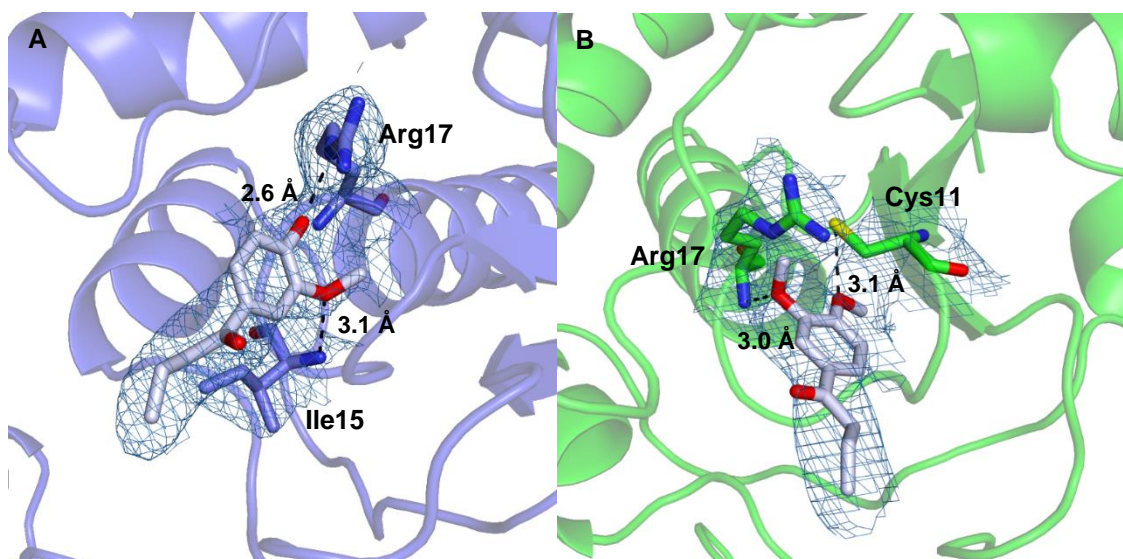
**Figure 4.31** – Representation of the first modeling attempt with fragment 1 in the density blob at the active site in molecule A (**A**) and molecule B (**B**). The electronic density map  $2F_o - F_c$  (blue) is contoured at  $1\sigma$ . Ligand color code: ligand in stick in which carbon atoms in grey and oxygen atoms in red. Protein color code: protein in blue/green cartoon, active site residues in sticks, in which carbon atoms in blue/green, nitrogen in blue and sulphur in yellow. Figure prepared in PyMOL.

It should be highlighted that the selected orientation of the fragment 1 at the active site represents the most probable orientation, considering the theoretical molecular docking binding studies. In detail, this orientation allows the dimethoxy substituted A-ring to establish H-bond interactions with the active site residues, as represented in Figure 4.32. Beyond that, the surrounding PTP loop residues (residues 11-18) play a crucial role in ligand stabilization. The fragment was modelled with an occupancy value of 1 in both molecules A and B. This occupancy value indicates that the atoms of the fragment are positioned in the coordinates indicated by the model in the totality of the crystal molecules.

In addition to the occupancy value, it is essential to examine another important parameter: the B factor. Protein chains present equivalent B factor values of  $75.4\text{ \AA}^2$  and  $74.2\text{ \AA}^2$  for molecule A and B, respectively. On the other hand, the average calculated B factor value for fragment 1 is  $59.9\text{ \AA}^2$  (molecule A) and  $76.6\text{ \AA}^2$  (molecule B). Besides, it should be also taken into consideration the B factor values of the surrounding environment of the fragment corresponding to the active center:  $52.6\text{ \AA}^2$  and  $69.3\text{ \AA}^2$  for molecule A and B, respectively. The analysis of the mentioned values supports the modelled hypothesis, namely because the B factor values of the fragment are quite similar to the respective neighbouring atoms in both molecules.

Although the results obtained for the fragment 1 orientation in the active site blob are convincing, further refinement cycles would be useful to the subject elucidation. If confirmed, it gives strong evidences that must be considered and validated in higher resolution structures.

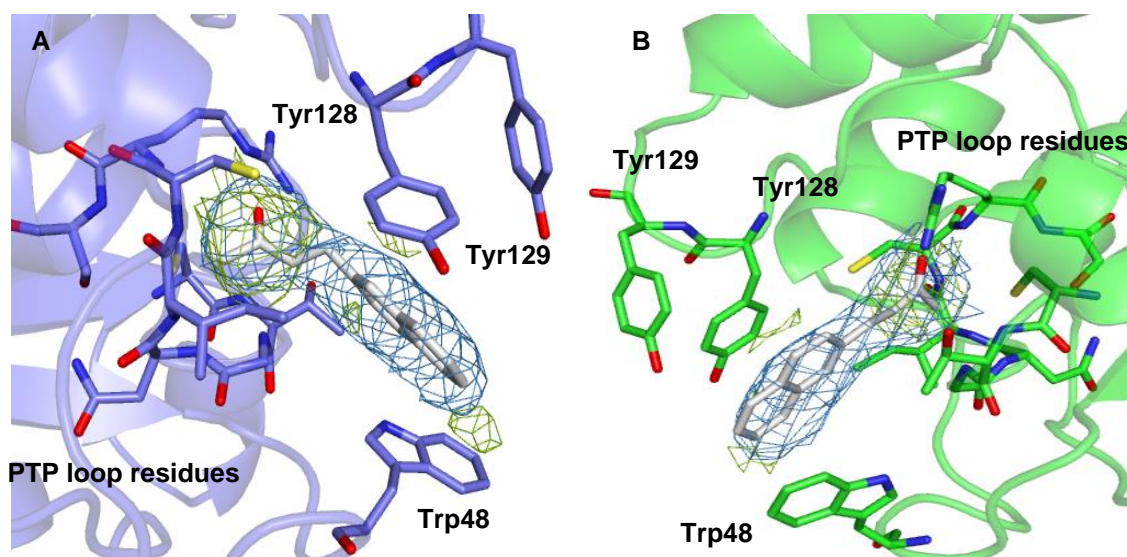




**Figure 4.32** – Representation of the fragment 1 modelled in the density blob at the active site in molecule A (**A**) and molecule B (**B**). Active site residues, namely Arg17, Ile15 in molecule A and Cys11 and Arg17 in molecule B are involved in hydrogen bonds with the methoxy groups of the ligand. The electronic density map  $2F_o - F_c$  (blue) is contoured at  $1\sigma$ . Ligand color code: ligand in stick in which carbon atoms in grey and oxygen atoms in red. Protein color code: protein in blue/green cartoon, active site residues in sticks, in which carbon atoms in blue/green, nitrogen in blue and sulphur in yellow. Figure prepared in PyMOL.

As aforementioned, an alternative approach was attempted by which the fragment 2 of C33 (including the last two rings previously reported in Figure 4.30) was modelled into the electron density blob for the two molecules of the asymmetric unit.

After a refinement cycle, a particular attention was given to the B factor values corresponding to the ligand fragment in both molecules and to the respective surrounding environment residues. Compared with the B factor values of the active site residues ( $54.5 \text{ \AA}^2$  in molecule A and  $66.9 \text{ \AA}^2$  in molecule B), the B factor values of the fragment revealed to be very high –  $69 \text{ \AA}^2$  (molecule A) and  $97 \text{ \AA}^2$  (molecule B) – not favoring a correct modelling of the fragment. Besides this, the visualization of  $2F_o - F_c$  and  $F_o - F_c$  electron density maps in COOT program allowed to confirm an advent of a positive density blob near the carbonyl group of the fragment (Figure 4.33). Besides, the fragment does not established interactions with the protein residues. This indicates that fragment 2 does not seem to be the most appropriated, as expected.



**Figure 4.33** – Representation of the fragment 2 of C33 compound in molecule A (**A**) and molecule (**B**) of the asymmetric unit. The fragment was modelled in the density blob and positive peaks are observable nearby the carbonyl group of the fragment in both molecules. The electronic density maps  $2F_o - F_c$  (blue) and  $F_o - F_c$  (green) are contoured at 1 and 3  $\sigma$ . Ligand color code: ligand in stick in which carbon atoms in grey and oxygen atoms in red. Protein color code: protein in blue and green cartoon, active site residues in sticks, in which carbon atoms in blue/green, nitrogen in blue and sulfur in yellow. Figure prepared in PyMOL.

# CHAPTER 5

---

## **C**ONCLUSIONS & **F**UTURE **P**ERSPECTIVES



Tuberculosis, the global leading infectious killer, remains as a public health problem causing 10 million new cases per year and a pool of two billion infected individuals. The outcome of *Mycobacterium tuberculosis* infection depends on the struggle between the host antimicrobial defense arsenal and the sophisticated bacterial countermeasures. Its great ability to challenge the immune system and survive in the infected host makes *Mtb* the most successful human pathogen.

Although the worldwide incidence rate of TB per head seems to be decreasing, much remains to be discovered to effectively decrease its occurrence and eradication. The current knowledge is not commensurate with this ambitious aims – the existing drugs and vaccine reveal to be insufficient – being necessary to establish renewed efforts in TB prevention and treatment. Moreover, multidrug resistance strains to traditional antibiotics have caused a resurgence of the infection. Thus, considering the current scenario, alternative viable approaches against TB are required.

Attention has been given to the interactions between the pathogen and the human host and understanding the underlying molecular mechanisms involved in the infection process is crucial. The pathogen depends of numerous effectors to promote its survival in the hostile host environment and a significant research interest has been shifted to the secreted protein phosphatases, as alternative therapeutic targets to treat TB. In fact, *Mtb* after being engulfed by human macrophages, secretes a cocktail of proteins able to modulate several signalling pathways, attenuating the host immune response. Protein tyrosine phosphatase A (PtpA) is one of those proteins particularly investigated and implied in the pathogenicity of *Mtb*, due to its capacity to dephosphorylate key host proteins, preventing the phagosome acidification and the phagosome-lysosome fusion.

Therefore, PtpA is considered a very promising drug target to fight the disease. One of the major challenges in the development of new inhibitors is selectivity, since the new conceived molecules must exclusively inhibit PtpA and not interact with human phosphatases. The family of chalcones and derivatives has been particularly investigated showing promising *in vitro* results, but there is still an evident lack of structural and biophysical reports to elucidate the PtpA-inhibitors interaction mechanisms.

In this dissertation, an integrated *in silico*, biophysical and structural strategy was implemented to systematically characterize the protein-ligand interactions contributing to the future design of new, efficient and selective inhibitors towards PtpA as viable anti-TB drugs.

The first step of the indicated strategy consisted in protein expression and purification procedures which were successfully achieved with a yield of 20 mg of protein per liter of culture. The purity of PtpA was quite high using affinity chromatography procedures and the protein could be concentrated up to 20 mg/mL. The production of pure protein enabled the application of several techniques, from a biophysical and structural point of view, aiming to characterize the protein with promising inhibitor candidates.

Molecular docking studies were performed in order to clarify the binding mode – namely the binding sites – of chalcones and thiosemicarbazones compounds which were reported as

PtpA selective inhibitors by Hernán Terenzi and co-workers. Results from these theoretical investigations showed that thiosemicarbazones compounds preferably bind to an allosteric site in the protein structure. The residues involved in this possible allosteric binding are common to the hot spots previously reported by Terenzi *et al.* On the other hand, docking calculations revealed that chalcones occupy the active site pocket, interacting with PTP-loop residues.

Among the two families, the thiosemicarbazone Lap04 compound revealed the best docking scores, showing high binding affinities toward PtpA ( $\Delta G$ : -9.21 kcal/mol). The naphthyl group emerged as a key moiety in the complex orientation, establishing electrostatic interactions with PtpA, while the R<sub>1</sub> group mainly contributed for the complex stabilization through hydrophobic interactions. In the case of Lap11, the compound bears a phenyl ring substituted with two electron-withdrawing groups, which revealed to be crucial members for interacting with the allosteric region. In addition, Lap11 displayed hydrogen bonds with protein residues, which play a pivotal role in ligand orientation.

In parallel, this computational method allowed to explore a new family of compounds as putative PtpA inhibitors – azaindoles. Since no data is reported in literature, docking calculations are a relevant source to a better understanding concerning the binding mechanism towards the protein. A library composed by twenty azaindoles was screened using AutoDock Vina and AutoDock softwares, suggesting compound Azaindole 4 as a possible promising hit ( $\Delta G$ : -8.25 kcal/mol) sitting at the active site through polar and hydrophobic interactions. Its characteristic Y-shaped allows the compound to be nicely accommodated in the active site region.

As general conclusion, the results herein discussed provided valuable insights towards a better understanding of the binding mode of this class of inhibitors. In addition, it should be noted that the synthesis of azaindole 4 is worthy in order to experimentally validate the docking results, using a similar experimental rational to the one described. The behaviour of the complex PtpA-Azaindole 4 was also evaluated by MD, suggesting possible disarrangements in protein structure, mainly in the active site upon ligand interaction.

From a biophysical standpoint, several methodologies were used to characterize different features of the protein and the possible protein-ligand interactions. Firstly, TSA experiments were conducted to assess protein thermostability under the presence of DMSO, since this agent is required to solubilize the ligands. The results suggested that 10% DMSO started to jeopardize the protein stability. Nevertheless, it was also observable that the protein's T<sub>m</sub> diluted in a buffer with the concomitant presence of 10% glycerol and 10% DMSO was inclusively slightly higher than the one observed in the absence of both molecules. This reveals that glycerol drastically reduces the negative destabilization effect of DMSO on the protein, allowing its use as ligand solvent in moderate concentrations.

Secondly, TSA was crucial to determine the best storage conditions of the protein guaranteeing its quality for further experimental studies. According to the results, PtpA presents a 2.83°C lower T<sub>m</sub> value when stored at 4°C if compared with a -80°C stored sample suggesting that the protein stability is not affected by freeze-thaw cycles.

Finally, due to the previously mentioned struggles caused by DMSO, alternative organic solvents were considered, namely PEG 400. Herein, TSA was again a helpful technique showing a positive shift of the protein  $T_m$  value when incubated with higher concentrations of PEG400 rather than the one obtained with similar concentrations of DMSO. Therefore, as future work, testing this attractive organic solvent candidate to solubilize the ligands under study is strongly recommended.

Regarding the detection of protein-ligand binding, TSA experiments were not conclusive to validate the existence of these interactions. A visible  $T_m$  shift was noticeable in Lap04 experiment, however it corresponded to the compound and not to the PtpA-Lap04 interaction. Nonetheless, such interactions cannot be discarded due to the low sensitivity of the technique.

Alternatively, additional biophysical characterization studies were performed through two other techniques: MST experiments, and urea-polyacrylamide gel electrophoresis. PtpA was successfully labelled to be used in the MST experiments with compounds from the three families. The binding check results did not show observable shifts in the MST traces corresponding to the target and to the complex, corroborating the previous TSA results and compromising the next binding affinity step. Nevertheless, an exception was found with Lap11 which was properly tested in a binding affinity assay, although not allowing to obtain affinity curves and the respective  $K_d$  value. Further experimental conditions must be reformulated including different protein-ligand proportions, which could be achieved using PEG 400 as an alternative organic solvent.

Urea-polyacrylamide gel electrophoresis showed that the protein-ligand samples under study had similar electrophoretic mobility profiles when compared to the ligand-free protein, possibly due to the binding mechanism does not induce any conformational change to a more compacted one.

Although the used biophysical methods did not unequivocally prove protein-ligand interactions, the obtained results should not be considered as an evidence of their inexistence. For example, low binding affinities of the inhibitors towards PtpA can hamper the detection of any interaction on the used conditions. Therefore, in addition to the proposed modifications of such experimental conditions, other distinct techniques are planned to be included in the study namely STD-NMR (Saturation Transfer Difference Nuclear Magnetic Resonance), ITC (Isothermal Titration Calorimetry) and SPR (Surface Plasmon Resonance).

From the structural perspective, extensive ligand-free PtpA crystallization trials were performed in order to reproduce a crystallization condition previously obtained in the laboratory. Due to the difficulties faced on this task, an alternative approach was followed aiming to determine new suitable crystallization conditions using several initial screens with different precipitants, concentrations, pH values and temperatures. However, regardless the used conditions, all these considerable efforts failed to produce suitable crystals.

To overcome the unsuccessful crystallization trials attempted with ligand-free protein, co-crystallization experiments emerged as a promising alternative method and different new conditions were found with C33, C37, R6, JE02 and Lap11 compounds. However, using synchrotron radiation, most of the obtained crystals revealed, if any, a very poor diffraction. Salt

crystals were also obtained, in particular the ones resulting from the co-crystallization with the JE02 compound.

Focusing on the Lap11 co-crystallization results, two crystals diffracted at low resolution values up to 3.9 Å and 4 Å. After multiple optimization attempts, a slight resolution improvement was achieved in the crystallization condition – containing 12% PEG 6K, 0.1 M Bicine 9 - and a complete 3.6 Å dataset was collected. Despite the obtained resolution, the structure was solved and two suspicious density blobs were found at the active site region and near the His120B residue which, according to the docking results, is within the determined allosteric site.

At the current stage of refinement, there are still many open and important questions directly related to the blobs nature since the electron density was difficult to interpret and for model building. To properly answer some of these speculations, one of the first planned steps is to improve the crystals quality. Consequently, further optimizations of the co-crystallization conditions are required in order to obtain well-ordered crystals diffracting at higher resolution. The use of additive screens and as well as the use of specific ionic liquids with the same purpose are planned, enabling further investigations of the protein-ligand interactions at the atomic level. With the coordinates retrieved by the docking results, the experimental analysis of the protein site-specific regions - where these compounds are expected to bind - would be of great interest.

A second structure was obtained using a previously collected dataset corresponding to a PtpA-C33 crystal obtained by soaking. The crystal structure was processed and solved at 2.8 Å resolution revealing a much more interpretable electron density than the one found for the PtpA-Lap11 structure. The  $2F_o - F_c$  and  $F_o - F_c$  maps showed a suspicious blob near the active site pocket, in both molecules of the assymetric unit. This finding let to speculate the binding of the ligand to the protein as suggested by the docking calculations. Taking into consideration the shape of the blob, two C33 fragments were modelled in this extra electron density. In this case, fragment 1 – with a similar orientation suggested by the docking – was considered the most probable as corroborated by the analysis of B factor values and by the visual inspection of the electron density map after the respective refinement.

Unexpectedely, both PtpA structures presented a common feature related to the presence of a disulfide bond between Cys53A and Cys53B. This finding was considered a crystallographic dimer rather than a biological dimer since a complementary SEC experiment – with and without DTT, revealed a single elution peak corresponding to the monomeric form of the protein.

As previously stated, obtaining good-diffracting crystals is a crucial step to further studies on PtpA-inhibitor interactions. In the pursuit for new crystallization methodologies, microseeding was used to enhance the chances of crystal formation. Although the poor resolution observed, the first ligand-free protein and PtpA-Azaindole 16 crystals were obtained, a progression never achieved before. Considering these remarkable positive results, future work should rely on different optimization experiments namely using the referred methods. In addition, subsequent soaking experiments would be carried out, aiming to obtain better structural data that allow a more accurate characterization of protein-ligand interactions.



In conclusion, although some of the proposed objectives were successfully achieved and the project contributed to advance knowledge in this challenging area of research, some questions are still unanswered and should be pursued. This dissertation allowed me to thoroughly explore the technique of X-ray crystallography, from the crystallization process to the analysis of the structure. The lack of good diffracting crystals, the low success of co-crystallization trials and the limited access to synchrotrons are just some representative examples of how challenging X-ray crystallography technique is. Nevertheless, the importance of X-ray crystallography in the drug design field is undeniable. Its combination with some other complementary biophysical methodologies, implemented in this project, revealed to be extremely useful allowing the elucidation of several molecular details which, in turn, contributed to propose some hypothesis to be confirmed in the future.

As concluding remarks, new findings, together with the already achieved results, will be essential to open avenues in the future use of these classes of molecules as viable and safe therapeutic agents helping to improve the life conditions of a significant percentage of the population affected by TB.



# CHAPTER 6

---

## **B**IBLIOGRAPHY

1. McMillen, C., *Discovering Tuberculosis – A global history; 1900 to the present*, Yale University Press, 2016.
2. Daniel, T.M., Tuberculosis, *Respiratory Medicine*, 100, pp. 1862-1870, 2006.
3. Zimmerman, M.R., Pulmonary and osseous tuberculosis in an Egyptian mummy, *Bulletin of the New York Academy of Medicine*, 55, pp. 604-608, 1979.
4. Nerlich, A. G., Haas, C. J., Zink, A., Szeimies, U., Hagedorn, H. G., Molecular evidence for tuberculosis in an ancient Egyptian mummy, *Lancet*, 350, pp. 1404-1405, 1997.
5. Grosset, J.H., Chaisson, R.E., *Handbook of Tuberculosis*, Springer Nature, 47, 2006.
6. Houk, H., Alert, H., Tuberculosis, *Marshall Cavendish Benchmark*, 2010.
7. Fogel, N., Tuberculosis: A disease without boundaries, *Tuberculosis*, 95, pp. 527-531, 2015.
8. Lakhtakia, R., The Legacy of Robert Koch: Surmise, search, substantiate, 14, pp. 37-41, 2014.
9. World Health Organization. *Global Tuberculosis Report 2019* (WHO, 2019)
10. <https://www.dgs.pt/portal-da-estatistica-da-saude/diretorio-de-informacao/diretorio-de-informacao/por-serie-963780-pdf.aspx?v=11736b14-73e6-4b34-a8e8-d22502108547> (Accessed at 18<sup>th</sup> June 2019)
11. <https://www.cdc.gov/tb/> (Accessed at 9<sup>th</sup> July 2019)
12. Pai, M., Behr, A.M., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C.C., Ginsberg, A., Swaminathan, S., Spigelman, M., Getahun, H., Menzies, D., Raviglione, M., Tuberculosis, *Nature Reviews Disease Primers*, 2, pp. 16076-16081, 2016.
13. Barry, E.C., Boshoff, I.H., Dartois, V., Dick, T., Ehrt, S., Flynn, J., Schnappinger, D., Wilkinson J.R., Young, D., The spectrum of latent tuberculosis: rethinking the biology and intervention strategies, *Nature Reviews Microbiology*, 7, pp. 845-855, 2009.
14. Esmail, H., Barry, C. E.3rd, Young, D., Wilkinson, R. J., The ongoing challenge of latent tuberculosis, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130437, 2014.
15. Ducati, R.G., Ruffino-Neto, A., Basso, L.A., Santos, D.S., The resumption of consumption – A review on tuberculosis, *Memórias do Instituto Oswaldo Cruz*, 101, pp. 697- 714, 2006.
16. Fu, L.M., Fu-Liu, C.S., Is *Mycobacterium tuberculosis* a closer relative to Gram-positive or Gram-negative bacterial pathogens ?, *Tuberculosis*, 82, pp. 85-90, 2002.
17. Hett, E.C., Rubin, E. J., Bacterial Growth and Cell Division: A Mycobacterial Perspective, *Microbiology and Molecular Biology Reviews*, 72, pp. 126-156, 2008.
18. Brennan, P.J., Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*, *Tuberculosis*, 83, pp. 91-97, 2003.
19. Gomez, J.E., McKinney, J.D., *M. tuberculosis* persistence, latency, and drug tolerance, *Tuberculosis*, 84, pp. 29-44, 2004.
20. Stutz, M.D., Clark, M.P., Doerflinger, M., Pellegrini, M., *Mycobacterium tuberculosis*: Rewiring host cell signaling to promote infection, *Journal of Leukocyte Biology*, 103, pp. 259-268, 2018.
21. Koul, A., Herget, T., Klebl, B., Ullrich, A., Interplay between mycobacteria and host signalling pathways, *Nature Reviews Microbiology*, 2, pp.189-202, 2004.
22. Upadhyay, S., Mittal, E., Philips, J.A., Tuberculosis and the art of macrophage manipulation, *Pathogens and Disease*, 76, pp. 1-12, 2018.
23. Tjelle, T.E., Lovdal, T., Berg, T., Phagosome dynamics and function, *BioEssays*, 22, pp. 255-263, 2000.

24. Sun-Wada, G.H., Tabata, H., Kawamura, N., Aoyama, M., Wada, Y., Direct recruitment of H<sup>+</sup>-ATPase from lysosomes for phagosomal acidification, *Journal of Cell Science*, 122, pp. 2504-2513, 2009.
25. Wong, D., Chao, J.D., Av-Gay, Y., *Mycobacterium tuberculosis*-secreted phosphatases: From pathogenesis to targets for TB drug development, *Trends in Microbiology*, 21, pp. 100-109, 2013.
26. Silva, A.G., Taberner, L., New strategies in fighting TB : targeting *Mycobacterium tuberculosis*-secreted phosphatases MtpA & MtpB, *Future Medicinal Chemistry*, 2, pp.1325-1337, 2010.
27. MacCari, R., Ottanà, R., Low molecular weight phosphotyrosine protein phosphatases as emerging targets for the design of novel therapeutic agents, *Journal of Medicinal Chemistry*, 55, pp. 2-22, 2012.
28. Ardito, F., Giuliani, M., Perrone, D., Troiano, G., Muzio, L., The crucial of protein phosphorylation in cell signalling and its use as targeted therapy, *International Journal of Molecular Medicine*, 40, pp. 271-280, 2017.
29. Charbonneau, H., Tonks, N.K., 1002 protein phosphatases?, *Annual Review of Cell and Developmental Biology*, 8, pp. 463-493, 1992.
30. Wang, J., Ge, P., Qiang, L., Tian, F., Zhao, D., Chai, Q., Zhu, M., Zhou, R., Meng, G., Iwakura, Y., Gao, G., Liu, C.H., The mycobacterial phosphatase PtpA regulates the expression of host genes and promotes cell proliferation, *Nature Communications*, 8, pp. 1-16, 2017.
31. Bach, H., Papavinasasundaram, K.G., Wong, D., Hmama, Z., Av-Gay, Y., *Mycobacterium tuberculosis* Virulence Is Mediated by PtpA Dephosphorylation of Human Vacuolar Protein Sorting 33B, *Cell Host and Microbe*, 3, pp. 316-322, 2008.
32. Wong, D., Bach, H., Sun, J., Hmama, Z., Av-Gay, Y., *Mycobacterium tuberculosis* protein tyrosine phosphatase (PtpA) excludes host vacuolar-H<sup>+</sup>-ATPase to inhibit phagosome acidification, *Proceedings of the National Academy of Sciences of the United States of America*, 108, pp. 19371-19376, 2011.
33. Burke, T.R., Zhang, Z., Protein–Tyrosine Phosphatases: Structure, Mechanism, and Inhibitor Discovery, *Biopolymers (Peptide Science)*, 47, pp. 225-241, 1998.
34. Madhurantakam, C., Rajakumara, E., Mazumdar, P.A., Saha, B., Mitra, D., Wiker, H.G., Sankaranarayanan, R., Kumar, A., Crystal Structure of Low-Molecular-Weight Protein Tyrosine Phosphatase from *Mycobacterium tuberculosis* at 1.9-Å Resolution, *Journal of Bacteriology*, 187, pp. 2175-2181, 2005.
35. Lazo, J.S., McQueeney, K.E., Sharlow, E.R., New Approaches to Difficult Drug Targets: The Phosphatase Story, *SLAS Discovery*, 22, pp.1071-1083, 2017.
36. Tautz, L., Critton, D.A., Grotegut, S., Protein Tyrosine Phosphatases: Structure, Function, and Implication in Human Disease, *Methods in Molecular Biology*, 1053, pp. 179-221, 2013.
37. Kaufmann, S.H.E., Evans, T.G., Hanekom, W.A., Tuberculosis vaccines: time for a global strategy, *Science Translational Medicine*, 7, pp.1-3, 2015.
38. Tiberi, S., Du Plessis, N., Walzl, G., Vjecha, J.M., Rao, M., Ntoumi, F., Mfinanga, S., Kapata, N., Mwaba, P., McHugh, T., Ippolito, G., Migliori, G.B., Maeurer, M.J., Zumla, A., Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies, *The Lancet Infectious Diseases*, 18, pp. 183-198, 2018.
39. Maitre, T., Aubry, A., Jarlier, V., Robert, J., Veziris, N., Bernard, C., Sougakoff, W., Brossier, F., Cambau, E., Mougari, F., Raskine, L., Multidrug and extensively drug-resistant tuberculosis, *Médecine et Maladies Infectieuses*, 47, pp. 3-10, 2017.

40. McBryde, E.S., Meehan, M.T., Doan, T.N., Ragonnet, R., Marais, B.J., Guernier, V., Trauere, J.M., The risk of global epidemic replacement with drug-resistant *Mycobacterium tuberculosis* strains, *International Journal of Infectious Diseases*, 56, pp. 14-20, 2017.
41. Zhang, Z., Drugging the undruggable: Therapeutic potential of targeting protein tyrosine phosphatases, *Accounts of Chemical Research*, 50, pp. 122-129, 2017.
42. Chiaradia, L.D., Mascarello, A., Purificação, M., Vernal, J., Cordeiro, M.N., Zenteno, M.E., Villarino, A., Nunes, R.J., Yunes, R.A., Terenzi, H., Synthetic chalcones as efficient inhibitors of *Mycobacterium tuberculosis* protein tyrosine phosphatase PtpA, *Bioorganic & Medicinal Chemistry*, 18, pp. 6227-6230, 2008.
43. Chiaradia, L.D., Martins, P.G., Cordeiro, M.N., Guido, R.V., Ecco, G., Andricopulo, A.D., Yunes, R.A., Vernal, J., Nunes, R.J., Terenzi, H., Synthesis, Biological Evaluation, And Molecular Modeling of Chalcone Derivatives As Potent Inhibitors of *Mycobacterium tuberculosis* Protein Tyrosine Phosphatases (PtpA and PtpB), *Journal of Medicinal Chemistry*, 55, pp. 390-402, 2012.
44. Mascarello, A., Chiaradia, L.D., Vernal, J., Villarino, A., Guido, R.V., Perizzolo, P., Poirier, V., Wong, D., Martins, P.G., Nunes, R.J., Yunes, R.A., Andricopulo, A.D., Av-Gay, Y., Terenzi, H., Inhibition of *Mycobacterium tuberculosis* tyrosine phosphatase PtpA by synthetic chalcones: Kinetics, molecular modeling, toxicity and effect on growth, *Bioorganic & Medicinal Chemistry*, 18, pp. 3783-3789, 2010.
45. Sens, L., Souza, A.C., Pacheco, L.A., Menegatti, A.C., Mori, M., Mascarello, A., Nunes, R.J., Terenzi, H., Synthetic thiosemicarbazones as a new class of *Mycobacterium tuberculosis* protein tyrosine phosphatase A inhibitors, *Bioorganic & Medicinal Chemistry*, 26, pp. 5742-5750, 2018.
46. Homan, K.T., Balasubramaniam, D., Zabell, A.P., Wiest, O., Helquist, P., Stauffacher, C.V., Identification of novel inhibitors for a low molecular weight tyrosine phosphatase via virtual screening, *Bioorganic & Medicinal Chemistry*, 18, pp. 5449-5456, 2010.
47. Twine, S.M., Murphy, L., Phillips, R.S., Callis, P., Cash, M.T., Szabo, A.G., The Photophysical properties of 6-Azaindole, *The Journal of Physical Chemistry B*, 107, pp. 637-645, 2003.
48. Mérour, J-Y., Buron, F., Plé, K., Bonnet, P., Routier, S., The azaindole framework in the design of kinase inhibitors, *Molecules*, 19, pp. 19935-19979, 2014.
49. Liota, E., Spyrou, G., Vassilatis, D. K., Cournia, Z., Structure-Based Virtual Screening for Drug Discovery : Principles , Applications and Recent Advances, *Current Topics in Medicinal Chemistry*, 14, pp. 1923-1928, 2014.
50. Renaud, J-P., Chung, C-W., Danielson, U.H., Egner, U., Hennig, M., Hubbard, R.E., Nar, H., Biophysics in drug discovery: impact, challenges and opportunities, *Nature Reviews Drug Discovery*, 15, pp. 679-698, 2016.
51. Barnash, K.D., James, L.I., Frye, S.V., Target class drug discovery, *Nature Chemical Biology*, 13, pp. 1053-1056, 2017.
52. Krishnan, V., Rupp, B., Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy, *Encyclopedia of Life Sciences*, John Wiley & Sons Ltd, 2006
53. Rupp, B., Biomolecular Crystallography – Principles, Practice and Application to Structural Biology, *Garland Science Taylor & Francis Group*, 1<sup>st</sup> Edition, 2010.
54. Rhodes, G., Crystallography Made Crystal Clear – A Guide for Users of Macromolecular Models, *Academic Press*, 3<sup>rd</sup> Edition, 2006.

55. Carvalho, A.L., Trincão, J., Romão, M.J., X-ray Crystallography in Drug Discovery in Ligand-Macromolecular Interactions in Drug Discovery – Methods and Protocols edited by Ana Cecília A. Roque, *Human Press*, 1<sup>st</sup> Edition, pp. 31-56, 2010.
56. Romão, M.J., Cristalografia de Proteínas: metodologias e aplicações em Bioquímica, *Boletim de Biotecnologia*, 53, pp. 18-36, 1996.
57. Li, M., Chang, W., Protein crystallization, *Photosynthesis Research*, 102, pp. 223-229, 2009.
58. McPherson, A., Gavira, J.A., Introduction to protein crystallization, *Acta Crystallographica F. Structural Biology Communications*, 70, pp. 2-20, 2014.
59. Pusey, M.L., Liu Z-J., Tempel, W., Praissman, J., Lin, D., Wang, B-C., Gavira, J.A., Ng, J-D., Life in the fast lane for protein crystallization and X-ray crystallography, *Progress in Biophysics and Molecular Biology*, 88, pp. 359-386, 2005.
60. Lamb, A.L., Kappock, T.J., Silvaggi, N.R., You are lost without a map: Navigating the sea of protein structures, *Biochimica et Biophysica Acta*, 1854, pp. 258-268, 2015.
61. Garry, T.L., Introduction to phasing, *Acta Crystallographica Section D. Biological Crystallography*, 66, pp. 325-338, 2010.
62. Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M., Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures, *FEBS Journal*, 275, pp. 1-21, 2008.
63. <http://esrf.eu/about/synchrotron-science/synchrotron> (Accessed at 27<sup>th</sup> July 2019)
64. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility, *Journal of Computational Chemistry*, 30, pp. 2785–2791, 2009.
65. Vina Trott, O., Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry*, 31, pp. 455–461, 2010.
66. Huey, R.; Morris, G.M.; Olson, A.J.; Goodsell, D.S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation, *Journal of Computational Chemistry*, 28, pp. 1145–1152, 2007.
67. DeLano, W., Pymol: An open-source molecular graphics tool, *CCP4 Newsletter On Protein Crystallography*, pp. 44-52, 2002.
68. Laskowski, R.A.; Swindells, M.B. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery, *Journal of Chemical Information and Modeling*, 51, pp. 2778-2786, 2011.
69. Spoel, D.V., Lindahl, E., Hess, B., Groenhof, G., GROMACS: Fast, Flexible, and Free, *Journal of Computational Chemistry*, 26, pp. 1701-1718, 2005.
70. <http://www.pomeslab.com/files/lipidCombinationRules> (Accessed 8<sup>th</sup> July 2019)
71. Schmid, N., Eichenberger, A.P., Choutko, A., Riniker, S., Winger, M., Mark, A.E., Van Gunsteren, W.F., Definition and testing of the GROMOS force-field versions 54A7 and 54B, *European Biophysics Journal*, 40, pp. 843–856, 2011.
72. Gonçalves, J., Caracterização estrutural e funcional da interação entre inibidores e PtpA: um possível alvo contra a tuberculose, Master Thesis, FCT-UNL, 2018.
73. Jancarik, J., Kim, S.H., Sparse matrix sampling: A screening method for crystallization of proteins, *Journal of Applied Crystallography*, 24, pp. 409-411, 1991.
74. Kabsch, W., XDS, *Acta Crystallographica Section D. Biological Crystallography*, 66, pp. 125-132, 2010.
75. Evans, P.R., Scaling and assessment of data quality, *Acta Crystallographica Section D. Biological Crystallography*, 62, pp. 72-82, 2006.

76. Dodson, E.J., Winn, M., Ralph, A., Collaborative Computational Project, number 4: providing programs for protein crystallography, *Methods in Enzymology*, 277, pp. 620-633, 1997.
77. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., Phaser crystallographic software, *Journal of Applied Crystallography*, 40, pp. 658-674, 2007.
78. Kovalevskiy, O., Nicholls, R.A., Murshudov, G.N., Automated refinement of macromolecular structures at low resolution using prior information, *Acta Crystallographica Section D. Structural Biology*, 72, pp. 1149-1161, 2016.
79. Emsley, P., Cowtan, K., Model-building tools for molecular graphics, *Acta Crystallographica Section D. Biological Crystallography*, 60, pp. 2126-2132, 2004.
80. Manger, M., Scheck, M., Prinz, H., Kries, J.P., Langer, T., Saxena, K., Schwalbe, H., Furstner, A., Rademann, J., Waldmann, H., Discovery of Mycobacterium tuberculosis Protein Tyrosine Phosphatase A (MtpA) Inhibitors Based on Natural Products and a Fragment-Based Approach, *ChemBioChem*, 6, pp. 1749-1753, 2005.
81. Mackerel, A., Empirical Force Fields for Biological Macromolecules: Overview and Issues, *Journal of Computational Chemistry*, 25, pp. 1584-1604, 2004.
82. González, M.A., Force fields and molecular dynamics simulations, *EDP Sciences*, 12, pp. 169-200, 2011.
83. Foltá-Stogniew, E., Oligomeric States of Proteins Determined by Size-Exclusion Chromatography Coupled With Light Scattering, Absorbance, and Refractive Index Detectors, *Methods in Molecular Biology*, 328, pp. 97-112, 2006.
84. Lo, M.C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., Ellestad, G., Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery, *Analytical Biochemistry*, 332, pp. 153-159, 2004.
85. Boivin, S., Kozak, S., Meijers, R., Optimization of protein purification and characterization using Thermofluor screens, *Protein Expression and Purification*, pp. 192-206, 2013.
86. Huynh, K., Partch, C.L., Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay, *Current Protocols in Protein Science*, 79, pp. 1-14, 2015.
87. Bala, I., Bhardwaj, V., Hariharan, S., Kumar, M.N., Analytical methods for assay of ellagic acid and its solubility studies, *Journal of Pharmaceutical and Biomedical Analysis*, 40, pp. 206-210, 2006.
88. Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., Breitsprecher, D., Microscale Thermophoresis: Interaction analysis and beyond, *Journal of Molecular Structure*, 1077, pp. 101-113, 2014.
89. Seidel, S.A., Dijkman, P.M., Lea, W.A., Bogaart, G.V., Willemsen, M.J., Lazic, A., Joseph, J.S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F.A., Ladbury, J.E., Schreiber, G., Watts, A., Braun, D., Duhr, S., *Methods*, 59, pp. 301-315, 2013.
90. Summer, H., Grämer, R., Dröge, P., Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE), *Journal of Visualized Experiments*, 32, pp. 1-3, 2009.
91. Santos, M.F.A., Structural and Functional Studies on the Reactivity of CORMs with Plasma Proteins, Doctoral Thesis, FCT-UNL, 2016.



# CHAPTER 7

---

## A PPENDIX

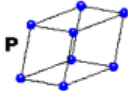
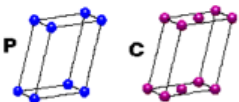

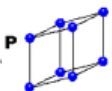
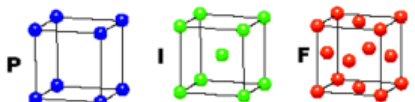
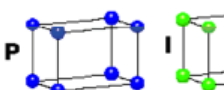
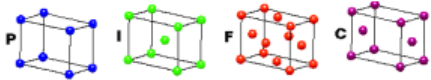


## Appendix 7.1

**Table 7.1** -  $IC_{50}$  and  $K_i$  values of chalcones and thiosemicarbazones compounds.<sup>43,44,45</sup>

Compound	Molecular Weight (g/mol)	$IC_{50}$ ( $\mu M$ )	$K_i$ ( $\mu M$ )
C33	381.13	$23.1 \pm 1.6$	$4.9 \pm 1.0$
C37	318.13	$8.4 \pm 0.9$	$5.4 \pm 1.4$
R06	404.13	$15.1 \pm 4.2$	$12 \pm 1.04$
JE02	329.11	$3.05 \pm 0.06$	Not published
Lap04	381.13	$7.4 \pm 2.8$	$5.6 \pm 1.0$
Lap11	473.06	$2.9 \pm 1.7$	$1.7 \pm 0.3$

## Appendix 7.2

Crystal system	Cell length	Cell angles	Protein Space Groups	Bravais Lattice
<i>Triclinic</i>	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$	P1	
<i>Monoclinic</i>	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ, \beta \neq 90^\circ$	P2, P2 <sub>1</sub> C2	
<i>Hexagonal</i>	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$	P6, P6 <sub>3</sub> , P6 <sub>6</sub> , P6 <sub>3</sub> , P6 <sub>2</sub> , P6 <sub>1</sub> , P622, P6 <sub>1</sub> 22, P6 <sub>3</sub> 22, P6 <sub>2</sub> 22, P6 <sub>3</sub> 22, P6 <sub>4</sub> 22	
<i>Trigonal</i>			P <sub>3</sub> , P3 <sub>1</sub> , P3 <sub>2</sub> P312, P321, P3 <sub>1</sub> 12 P3 <sub>1</sub> 21, P3 <sub>2</sub> 12, P3 <sub>2</sub> 21	
<i>Cubic</i>	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$	P23, P2 <sub>1</sub> 3, P432, P4 <sub>1</sub> 32, P4 <sub>2</sub> 32, P4 <sub>3</sub> 32 I23, I2 <sub>1</sub> 3, I432, I4 <sub>1</sub> 32 F23, F432, F4 <sub>1</sub> 32	
<i>Tetragonal</i>	$a = b \neq c$		P4, P4 <sub>1</sub> , P4 <sub>2</sub> , P4 <sub>3</sub> , P477, P477, P477, P477, P4 <sub>1</sub> 2 <sub>1</sub> 2, P4 <sub>2</sub> 22, P4 <sub>2</sub> 2 <sub>1</sub> 2, P4 <sub>3</sub> 2 <sub>1</sub> 2, P4 <sub>3</sub> 22 I4, I4 <sub>1</sub> , I4 <sub>1</sub> 22, I422	
<i>Orthorhombic</i>	$a \neq b \neq c$		P222, P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> , P2 <sub>1</sub> 2 <sub>1</sub> 2, P222 <sub>1</sub> I222, I2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> F222, C222, C222 <sub>1</sub>	

**Figure 7.1** - Crystal systems and allowed space groups for protein molecules. Four types of unit cell: Primitive (P); Centered in the side (C), in the body (I), or in the face (F). The spheres in the Bravais Lattice represent the lattice points.

Appendix 7.3

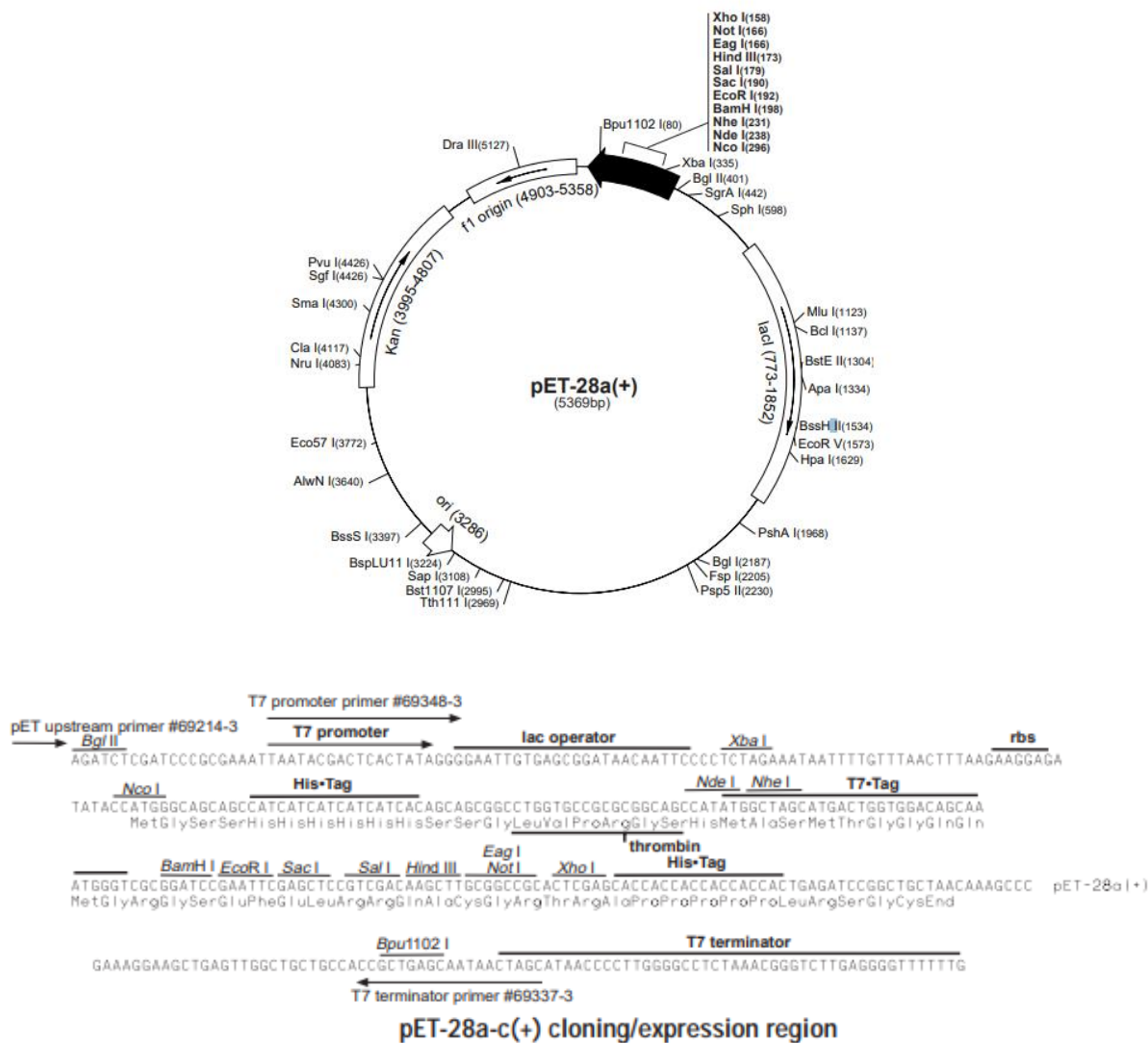


Figure 7.2 - Plasmid pET28a map and cloning region provided by Professor H ernan Terenzi.

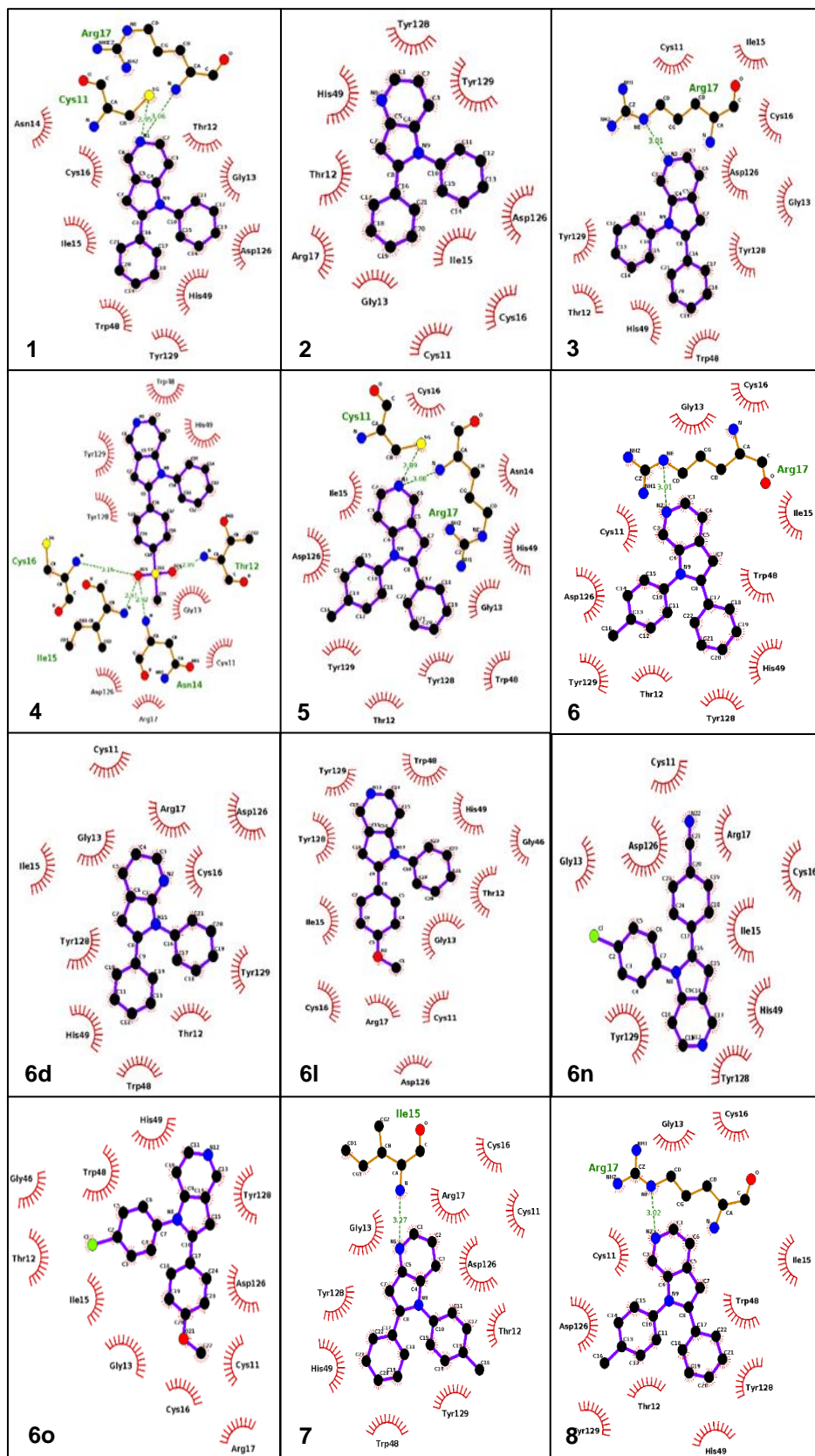
## Appendix 7.4

**Table 7.2 - JCSG<sup>+</sup> Crystallization Screen (Molecular Dimensions).**

1	0.2 M Lithium sulfate, 0.1 M Sodium acetate 4.5, 50 % PEG 400
2	0.1 M Sodium citrate 5.5, 20 % PEG 3000
3	0.2 M Ammonium citrate dibasic, 20 % PEG 3350
4	0.02 M Calcium chloride dihydrate, 0.1 M Sodium acetate 4.6, 30 % MPD
5	0.2 M Magnesium formate dihydrate, 20 % PEG 3350
6	0.2 M Lithium sulfate 0.1 M Phosphate/citrate 4.2, 20 % PEG 1000
7	0.1 M CHES 9.5, 20 % PEG 8000
8	0.2 M Ammonium formate, 20 % PEG 3350
9	0.2 M Ammonium chloride, 20 % PEG 3350
10	0.2 M Potassium formate, 20 % PEG 3350
11	0.2 M Ammonium phosphate monobasic, 0.1 M Tris 8.5, 50 % MPD
12	0.2 M Potassium nitrate, 20 % PEG 3350
13	0.8 M Ammonium sulfate, 0.1 M Citrate 4.0
14	0.2 M Sodium thiocyanate, 20 % PEG 3350
15	0.1 M BICINE 9.0, 20 % PEG 6000
16	0.1 M HEPES 7.5, 10 % PEG 8000, 8 % Ethylene glycol
17	0.1 M Sodium cacodylate 6.5, 40 % MPD, 5 % PEG 8000
18	0.1 M Phosphate/citrate 4.2, 40 % Ethanol
19	0.1 M Sodium acetate 4.6, 8 % PEG 4000
20	0.2 M Magnesium chloride hexahydrate, 0.1 M Tris 7.0, 10 % PEG 8000
21	0.1 M Citrate 5.0, 20 % PEG 6000
22	0.2 M Magnesium chloride hexahydrate, 0.1 M Sodium cacodylate 6.5, 50 % PEG 200
23	1.6 M Sodium citrate tribasic dihydrate pH 6.5
24	0.2 M Potassium citrate tribasic monohydrate, 20 % PEG 3350
25	0.2 M Sodium chloride, 0.1 M Phosphate/citrate 4.2, 20 % PEG 8000
26	1.0 M Lithium chloride, 0.1 M Citrate 4.0, 20 % PEG 6000
27	0.2 M Ammonium nitrate, 20 % PEG 3350
28	0.1 M HEPES 7.0, 10 % PEG 6000
29	0.8 M Sodium phosphate monobasic monohydrate, 0.1 M Sodium HEPES 7.5, 0.80 M Potassium phosphate monobasic
30	0.1 M Phosphate/citrate 4.2, 40 % PEG 300
31	0.2 M Zinc acetate dihydrate, 0.1 M Sodium acetate 4.5, 10 % PEG 3000
32	0.1 M Tris 8.5, 20 % Ethanol
33	0.1 M Sodium/potassium phosphate 6.2, 25 % 1,2-Propandiol, 10 % Glycerol
34	0.1 M BICINE 9.0, 10 % PEG 20,000
35	2.0 M Ammonium sulfate, 0.1 M Sodium acetate 4.6
36	10 % PEG 1000, 10 % PEG 8000
37	24 % PEG 1500, 20 % Glycerol
38	0.2 M Magnesium chloride hexahydrate, 0.1 M Sodium HEPES 7.5, 30 % PEG 400
39	0.2 M Sodium chloride, 0.1 M Sodium/potassium phosphate 6.2, 50 % PEG 200
40	0.2 M Lithium sulfate, 0.1 M Sodium acetate 4.5, 30 % PEG 8000
41	0.1 M HEPES 7.5, 70 % MPD
42	0.2 M Magnesium chloride hexahydrate, 0.1 M Tris 8.5, 20 % PEG 8000
43	0.2 M Lithium sulfate, 0.1 M Tris 8.5, 40 % PEG 400
44	0.1 M Tris 8.0, 40 % MPD
45	0.17 M Ammonium sulfate, 25.5 % PEG 4000, 15 % Glycerol
46	0.2 M Calcium acetate hydrate, 0.1 M Sodium cacodylate 6.5, 40 % PEG 300
47	0.14 M Calcium chloride dihydrate, 0.07 M Sodium acetate 4.6, 14 % 2-Propanol, 30 % Glycerol
48	0.04 M Potassium phosphate monobasic, 16 % PEG 8000 20 % Glycerol
49	1.0 M Sodium citrate tribasic dihydrate, 0.1 M Sodium cacodylate 6.5
50	2.0 M Ammonium sulfate, 0.1 M Sodium cacodylate 6.5
51	0.2 M Sodium chloride, 0.1 M HEPES 7.5, 10 % 2-Propanol
52	1.26 M Ammonium sulfate, 0.1 M Tris 8.5, 0.2 M Lithium sulfate
53	0.1 M CAPS 10.5 40 % MPD
54	0.2 M Zinc acetate dihydrate, 0.1 M Imidazole 8.0, 20 % PEG 3000
55	0.2 M Zinc acetate dihydrate, 0.1 M Sodium cacodylate 6.5, 10 % 2-Propanol

56	1.0 M Ammonium phosphate dibasic, 0.1 M Sodium acetate 4.5
57	1.6 M Magnesium sulfate heptahydrate, 0.1 M MES 6.5
58	0.1 M BICINE 9.0, 10 % PEG 6000
59	0.16 M Calcium acetate hydrate, 0.08 M Sodium cacodylate 6.5, 14.4 % PEG 8000
60	0.1 M Imidazole 8.0, 10 % PEG 8000
61	0.05 M Cesium chloride, 0.1 M MES 6.5, 30 % Jeffamine® M-600
62	3.2 M Ammonium sulfate, 0.1 M Citrate 5.0
63	0.1 M Tris 8.0 20 % v/v MPD 2-16 2-17
64	0.1 M HEPES 7.5, 20 % Jeffamine® M-600
65	0.2 M Magnesium chloride hexahydrate, 0.1 M Tris 8.5, 50 % Ethylene glycol
66	0.1 M BICINE 9.0, 10 % MPD
67	0.8 M Succinic acid pH 7.0
68	2.1 M DL-Malic acid pH 7.0
69	2.4 M Sodium malonate dibasic monohydrate pH 7.0
70	1.1 M Sodium malonate dibasic monohydrate, 0.1 M HEPES 7.0, 0.5 % Jeffamine® ED-2003
71	1.0 M Succinic acid, 0.1 M HEPES 7.0, 1 % PEG 2000 MME
72	0.1 M HEPES 7.0, 30 % Jeffamine® M-600
73	0.1 M HEPES 7.0, 30 % Jeffamine® ED-2003
74	0.02 M Magnesium chloride hexahydrate, 0.1 M HEPES 7.5, 22 % Poly(acrylic acid sodium salt) 5100
75	0.01 M Cobalt(II) chloride hexahydrate, 0.1 M Tris 8.5, 20 % Polyvinylpyrrolidone
76	0.2 M TMAO, 0.1 M Tris 8.5, 20 % PEG 2000 MME
77	0.005 M Cobalt(II) chloride hexahydrate, 0.1 M HEPES 7.5, 12 % PEG 3350, 0.005 M Cadmium chloride hemi(pentahydrate), 0.005 M Magnesium chloride hexahydrate, 0.005 M Nickel(II) chloride hexahydrate
78	0.2 M Sodium malonate dibasic monohydrate, 20 % PEG 3350
79	0.1 M Succinic acid, 15 % PEG 3350
80	0.15 M DL-Malic acid, 20 % PEG 3350
81	0.1 M Potassium thiocyanate, 30 % PEG 2000 MME
82	0.15 M Potassium bromide, 30 % PEG 2000 MME
83	2.0 M Ammonium sulfate, 0.1 M BIS-Tris 5.5
84	3.0 M Sodium chloride, 0.1 M BIS-Tris 5.5
85	0.3 M Magnesium formate dihydrate, 0.1 M BIS-Tris 5.5
86	1.0 M Ammonium sulfate, 0.1 M BIS-Tris 5.5, 1 % PEG 3350
87	0.1 M BIS-Tris 5.5, 25 % PEG 3350
88	0.2 M Calcium chloride dihydrate, 0.1 M BIS-Tris 5.5, 45 % MPD
89	0.2 M Ammonium acetate, 0.1 M BIS-Tris 5.5, 45 % MPD
90	0.1 M Ammonium acetate, 0.1 M BIS-Tris 5.5, 17 % PEG 10,000
91	0.2 M Ammonium sulfate, 0.1 M BIS-Tris 5.5, 25 % PEG 3350
92	0.2 M Sodium chloride, 0.1 M BIS-Tris 5.5, 25 % PEG 3350
93	0.2 M Lithium sulfate, 0.1 M BIS-Tris 5.5, 25 % PEG 3350
94	0.2 M Ammonium acetate, 0.1 M BIS-Tris 5.5, 25 % PEG 3350
95	0.2 M Magnesium chloride hexahydrate, 0.1 M BIS-Tris 5.5, 25 % PEG 3350
96	0.2 M Ammonium acetate, 0.1 M HEPES 7.5, 45 % MPD

## Appendix 7.5



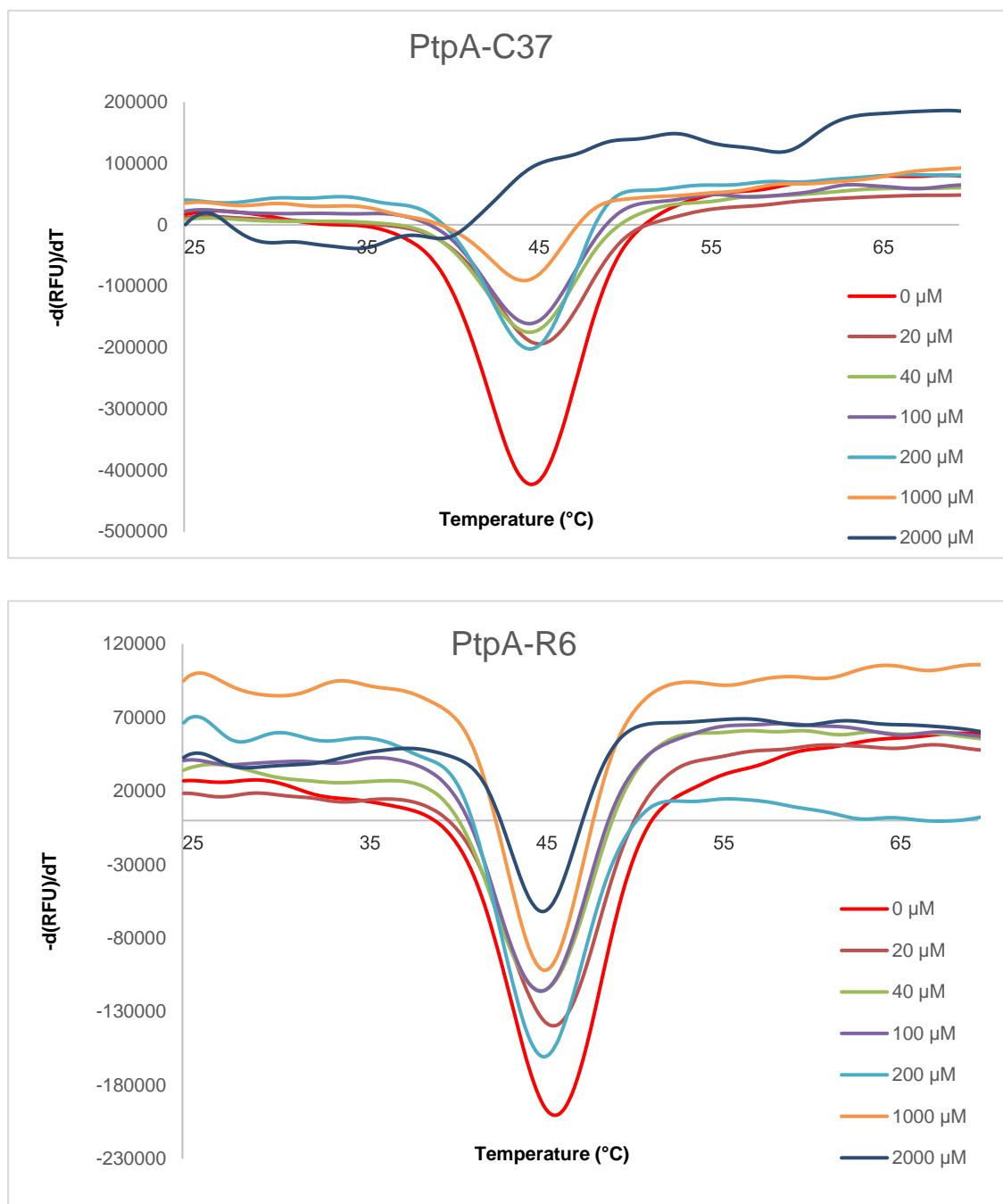




**Table 7.3** - AutoDock Vina and AutoDock docking results of the three families of compounds under study.

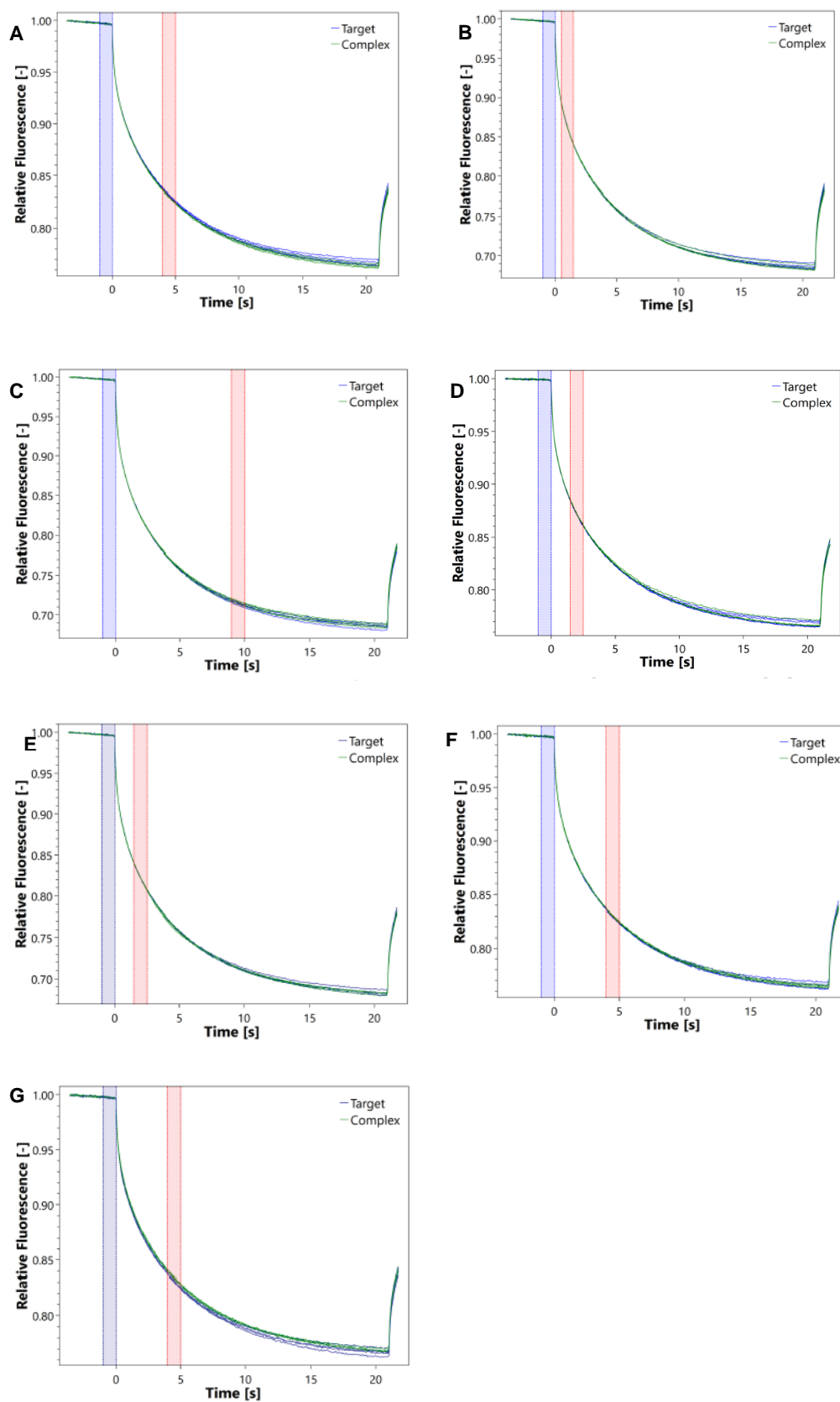
Compound	Docking Scores ( $\Delta G$ : kcal/mol)	
	AutoDock Vina	AutoDock
1	-6.4	-7.17
2	-6.5	-7.14
3	-6.5	-7.17
4	-6.4	-8.25
5	-6.2	-7.25
6	-6.2	-7.21
6d	-6.2	-7.19
6l	-6.4	-7.42
6n	-6.7	-7.69
6o	-6.4	-7.39
7	-6.3	-7.21
8	-6.3	-7.16
9	-6.3	-7.43
10	-6.2	-7.60
11	-6.2	-7.27
12	-6.8	-7.75
13	-6.4	-7.53
14	-6.8	-7.39
15	-6.7	-6.96
16	-6.3	-7.21
C33	-6.6	-7.66
C37	-6.7	-7.64
R6	-6.4	-7.16
JE02	-7.2	-7.41
Lap04	-7.9	-9.21
Lap11	-7.5	-8.88

## Appendix 7.6



**Figure 7.4** – TSA curves for chalcones C37 and R6. No significant  $T_m$  shifts were observed when compared to the control (red curve).

## Appendix 7.7



**Figure 7.5** – Binding Check MST traces of (A) C37 (B) R06 (C) JE02 (D) Lap04 (E) Azaindole 1 (F) Azaindole 15 and (G) Azaindole 16