The Power of Fragments: FBLD approach to investigate protein structures

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie der

Philipps-Universität Marburg

vorgelegt

von

Francesca Magari

aus **Silandro (Schlanders)**

Marburg/Lahn **2020**

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr.Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von April 2016 bis Januar 2020 durchgeführt.

Abstract

Fragments are small chemical entities with the extraordinary advantage of belonging to a wide variety of functional groups able to explore various chemical spaces, penetrate small protein pockets and allow the mapping of various protein binding sites. Nowadays, fragment-based lead discovery (FBLD) is the method of choice to screen fragment libraries and plays an increasingly important role in the drug development process. It becomes therefore particularly important to have a general-purpose fragment library ready to use. In our group a 96-fragment library was validated on eight protein targets, which was designed in a joint project between the HZB MX-group at BESSY II (AG Weiss) and the Institute of Pharmaceutical Chemistry, University of Marburg (AG Klebe) (Part I). Among the eight validated proteins, there are Thermolysin (TLN, Part II) and *Trypanosoma cruzi* Farnesyl Pyrophosphate Synthase (*T. cruzi* FPPS, Part III) which were the main focus of the present study. The 96-fragment library was therefore used as primary X-ray crystallographic screening method to search for novel chemical scaffolds. However, a new soaking protocol had to be established prior to the fragment screening. In fact, unlike the typical TLN inhibitor, fragments usually have insufficient affinity to displace on their own the autoproteolitically product Val-Lys dipeptide which occupies the TLN active site. The protein crystals were therefore incubated in an isopropanol solution able to displace the dipeptide and subsequentially screened against the fragment library resulting in seven crystallographic fragment hits. Unlike TLN which is usually used as a model protein, *T. cruzi* FPPS is a key enzyme involved in the mevalonate pathway and is essential for sterol production. *T. cruzi* and *T. brucei* parasites cause *Chagas disease* (CD) and human African trypanosomiasis (HAT) respectively and inhibition of FPPS in these species is therefore a valid target to fight these two tropical diseases. The present work focused first on optimizing the stability of protein samples in solution using various techniques such as thermal shift assay (TSA) and light scattering (LS). A crystallization and soaking protocol were subsequently established prior to performing a crystallographic fragment screening of

the above-mentioned 96 fragment library. In addition, the role of the magnesium ion as cofactor of the enzyme was also elucidated, either alone or in complex with isopentenyl pyrophosphate (IPP) or bisphosphonates (BPs). Based on the discovery by *Jahnke et al.* of a new allosteric pocket in *human* FPPS, it was assumed that a similar pocket was also present in the Trypanosoma species which was therefore the main target-site of the present work. The fragment screening project resulted in three crystallographic hits, one of which binding the allosteric pocket while the other two in a new pocket whose function is still unknown. They can serve as good candidates for the design of a new series of lead compounds.

In addition to the 96-fragment library, other series of compounds of fragment-size were screened against Endothiapepsin (EP, Part IV) and different *human* Carbonic Anhydrase (*h*CA, Part V) isoforms. In particular, a series of tetrazole compounds were crystallographically screened against EP in order to validate a novel pipeline for the rapid development of novel aspartyl protease inhibitors using an anchoring approach. The hydrazide moiety was here chosen as fragment-anchor able to bind the catalytic dyad in the EP active site. Moreover, in addition to X-ray crystallography which was the dominant screening method used in the present thesis, a series of para-substituted benzenesulfonamide were screened against different *h*CAs isoforms using Surface Plasmon Resonance (SPR). According to *Gaspari et al.*, the kinetics of binding of *h*CAII depend on the nature of the substituent that decorate the benzenesulfonamide moiety. In particular, the association rate k_{on} becames faster by increasing the hydrophobic nature of the para-alkyl chain due to a pre-binding event. This finding was also confirmed in the present study. However, it seems that more than one features contribute to the trends in k_{on} and k_{off}. Furthermore, in contrast with *h*CAII, *h*CAXII showed a faster dissociation rate by increasing the length of the para-alkyl chain, probably due to the higher surface hydrophilicity. The nature of the surface around the active site of *h*CAs plays therefore an important role in the kinetics of binding.

In conclusion, fragments are powerful molecules able to explore and discover new protein pockets, to evolve into more potent lead compounds and often they are used as starting point in several screening methods in the field of FBLD.

Zusammenfassung

Fragmente sind kleine organische Moluküle mit einer großen zahl funktioneller Gruppen, die in der Lage sind, verschiedene chemische Räume zu erforschen, in kleine Proteintaschen einzudringen und Proteinbindungsstellen zu finden. Heutzutage ist die Fragment-basierte Leitstrukturentwicklung (FBLD) die Methode der Wahl, um Bindungsstelle in Proteienen zu suchen und spielt eine immer wichtigere Rolle in der Medikamentenentwicklung. Daher ist es notwendig, eine universell einsetzbare Fragmentbibliothek zu haben. In unserer Gruppe wurde eine 96-Fragment-Bibliothek an acht Protein-Targets validiert, die in einem Gemeinschaftsprojekt zwischen der HZB MX-Gruppe am BESSY II (AG Weiss) und dem Institut für Pharmazeutische Chemie der Universität Marburg (AG Klebe) entwickelt wurde (Part I). Unter den acht validierten Proteinen sind Thermolysin (TLN, Part II) und Trypanosoma cruzi Farnesyl Pyrophosphate Synthase (T. cruzi FPPS, Part III), die im Mittelpunkt der vorliegenden Studie standen. Die 96-Fragment-Bibliothek wurde daher als primäre röntgenkristallographische Screeningmethode zur Suche nach neuartigen chemischen Gerüsten eingesetzt. Allerdings musste vor dem Fragmentscreening ein neues Soaking-Protokoll erstellt werden. Tatsächlich haben die Fragmente im Gegensatz zum typischen TLN-Inhibitor in der Regel keine ausreichende Affinität, um das Autoproteolyseprodukt Val-Lys-Dipeptid, das das aktive Zentrum von TLN besetzt, allein zu verdrängen. Die Proteinkristalle wurden daher in einer Isopropanol-Lösung inkubiert, die in der Lage ist, das Dipeptid zu verdrängen, und anschließend gegen die Fragmentbibliothek gescreent(to screen?), was zu sieben kristallographischen Fragmenthits führte. Im Gegensatz zu TLN, das üblicherweise als Modellprotein verwendet wird, ist *T. cruzi* FPPS ein Schlüsselenzym im Mevalonatweg beteiligt und für die Sterolproduktion unerlässlich ist. *T. cruzi-* und *T. brucei-*Parasiten verursachen die Chagas-Krankheit (CD) bzw. die humane afrikanische Trypanosomiasis (HAT). Die Hemmung von FPPS bei diesen Spezies ist daher ein gültiges Ziel zur Bekämpfung dieser beiden Tropenkrankheiten. Die vorliegende Arbeit

konzentrierte sich zunächst auf die Optimierung der Stabilität von Proteinproben in Lösung unter Verwendung verschiedener Techniken wie dem Thermal shift-Assay (TSA) und der Lichtstreuung (LS). Anschließend wurde ein Kristallisations- und Soaking-Protokoll erstellt, bevor ein kristallographisches Fragmentscreening der oben erwähnten 96-Fragment-Bibliothek durchgeführt wurde. Darüber hinaus wurde auch die Rolle des Magnesium-Ions als Kofaktor des Enzyms aufgeklärt, entweder allein oder im Komplex mit Isopentenylpyrophosphat (IPP) oder Bisphosphonaten (BPs). Basierend auf der Entdeckung einer neuen allosterischen Tasche in humanem FPPS durch *Jahnke et al.* wurde angenommen, dass eine ähnliche Tasche auch in der Spezies Trypanosoma vorhanden ist, die daher Gegenstand der vorliegenden Arbeit war. Das Fragment-Screening-Projekt ergab drei kristallographische Hits, von denen einer in die allosterische Tasche bindet, während die beiden anderen in einer neuen Tasche binden, deren Funktion noch unbekannt ist. Sie können als gute Kandidaten für das Design einer neuen Serie von Leitstrukturen dienen.

Neben der 96-Fragment-Bibliothek wurden weitere Reihen von Verbindungen mit Fragment-Größe gegen Endothiapepsin (EP, Part IV) und verschiedene humane Carboanhydrase-Isoformen (*h*CA, Part V) gescreent. Insbesondere wurde eine Reihe von Tetrazol-Verbindungen kristallographisch gegen EP gescreent, um eine neue Pipeline für die schnelle Entwicklung neuer Aspartyl-Protease-Inhibitoren mit einem Verankerungsansatz zu validieren. Die Hydrazideinheit wurde hier als Fragment-Verankerung gewählt, die in der Lage ist, an die katalytische Dyade im aktiven Zentrum des EP zu binden. Zusätzlich zur Röntgenkristallographie, die in der vorliegenden Arbeit die dominierende Screeningmethode war, wurde eine Reihe von para-substituiertem Benzolsulfonamiden gegen verschiedene *h*CAs Isoformen mittels Oberflächenplasmonenresonanz (SPR) gescreent. Nach *Gaspari et al.* hängt die Assoziationskinetik der Bindung von *h*CAII von der Art des Substituenten ab, den das Benzolsulfonamidgerüst trägt. Speziell wird die Assoziationsrate k_{on} durch die Bildung eines hydrophoben Kontakts bestimmt, dessen Bildung mit der vergrößerten Oberfläche der Liganden begüntstigt wird. Dieser Befund wurde auch in der vorliegenden Studie

bestätigt. Es scheint jedoch, dass mehr als ein Merkmal zu den Trends bei k_{on} und k_{off} beiträgt. Darüber hinaus zeigte hCAXII im Gegensatz zu hCAII eine schnellere Dissoziationsrate durch die Verlängerung der para-Alkylkette, wahrscheinlich aufgrund der höheren Oberflächenhydrophilie. Die Beschaffenheit der Oberfläche um das aktive Zentrum von hCAs spielt daher eine wichtige Rolle für die Bindungskinetik.

Zusammenfassend lässt sich sagen, dass Fragmente mächtige Moleküle sind, die in der Lage sind, neue Proteintaschen zu erforschen und zu entdecken, um zu potenteren Leitverbindungen entwickelt zu werden, und sie werden oft als Ausgangspunkt für verschiedene Screening-Methoden im FBLD verwendet.

Table of contents

List of abbreviations

Å	Angstrom (1 Å = 10^{-10} m)
ACT	Acetate ion
ADME	Absorption Distribution Metabolism Excretion
ALE	Alendronate
BBB	Blood-brain barrier
BCT	Bicarbonate ion
BDC	background density correction factor
BF	Bloodstream form
BNZ	Benznidazole
BPs	Bisphosphonates
CCC	Chronic chagasic cardiomyopathy
CD	Chagas disease
CNS	Central nervous system
DLS	dynamic light scattering
DMAPP	dimethylallyl diphosphate
DMSO	Dimethyl sulfoxide
DNDi	Drug for Neglected Diseases initiative
DSF	Differential scanning fluorimetry
e-	Electron

ECG	Electrocardiogram
EI-MS	Electron ionization mass spectrometry
ELISA	Enzyme-linked immunosorbent assay
EP	endothiapepsin
ES	enzyme-substrate complex
FARM	First Aspartate rich motif
FBLD	fragment-based lead discovery
FDPS	Farnesyl diphosphate synthase
FEXI	Fexinidazole
FPP	farnesyl pyrophosphate
FPPS	Farnesyl pyrophosphate synthase
GOL	glycerol
GPP	geranyl pyrophosphate
HAT	Human African Trypanosomiasis
Hepes	2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid
HPLC	high-performance liquid chromatography
HTS	high throughput screening
ITC	isothermal titration calorimetry
IPA	isopropanol
IM	Intramuscular
IPP	Isopentenyl diphosphate

IV	Intravenous
K_a	association constant at equilibrium
k_a	association rate constant
K_d	dissociation constant at equilibrium
k_d	dissociation rate constant
kDa	kilodalton
k_{off}	dissociation rate constant
k_{on}	association rate constant
LS	Light scattering
MALS	Multiangle light-scattering
MD	molecular dynamics
MDG medium	Metal, aspartic acid, glucose media
MES	2-(N-morpholino) ethanesulfonic acid
Mn	Mumber average molecular weight
modAI	modified auto-inducing medium
MOE	Molecular Operating Enviroment
MPD	methyl-2,4-pentanediol
MS-ESI	mass spectroscopy-electron spray ionization
MSCS	multiple-solvent crystal structure
Mw	Molecular weight
MX	macromolecular X-ray crystallography

m/z	mass divided by charge number (MS)
N-BPs	N-Bisphosphonates
Nbs	Nanobodies
NFZ	Nifurtimox
NTDs	Neglected tropical diseases
NTR	Nitro-reductase enzyme
OD_{600}	Pptical density, measured at a wavelength of 600 nm.
PanDDA	Pan-Dataset Density Analysis
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PF	Procyclic form
POS	Posaconazole
pO_2	partial pressure of oxygen
PPi	Pyrophosphate
R_h	Hydrodynamic radius
R_g	Radius of gyration
RI	Refractive index
RIS	Risedronate
RMSD	root-mean-square deviation

rpm	Revolutions per minute
RT	Room-temperature
SAD	Single-wavelength anomalous dispersion
SARM	Second Aspartate rich motif
σ_i	Map uncertainty for dataset i
SEC	Size-exclusion chromatography
SOC medium	Super Optimal broth with Catabolite repression medium
SPR	surface plasmon resonance
SQS	squalene synthase
SS	Short stumpy form
T. brucei	*Trypanosoma brucei*
T. cruzi	*Trypanosoma cruzi*
T_m	Melting temperature
TSA	Thermal shift assay
Tsal	Tsetse saliva
TLN	thermolysin
TRIS	tris(hydroxymethyl)aminomethane
UV	Ultraviolet
vdW	van der Waals
VSG	Variable surface glycoproteins
W	water molecule

WHO	World Health Organization
ZOL	Zoledronate
24-SMT	Δ^{24}-sterol methyltransferase

Part I: An Introduction to the Fragment Library

Exhaustive X-ray Crystallographic Screening of a Hit-Enriched 96 Fragment Library

1. Introduction

Since the Food and Drug Administration (FDA) approved in 2011 Vemurafenib (trade name Zelboraf) as the first drug designed using a fragment-based screening approach, more and more research groups and even drug research in industry, strongly devoted to the classic High-Throughput Screening (HTS) approach as prime source for candidates in lead discovery, consider small chemical entities called fragments increasingly favorable as starting points for drug development [1], [2], [3].

A fragment is defined as a chemical entity of small size and low molecular weight that usually follows "The Rule of Three" (Ro3), i.e. having a maximum of three H-bond donors/acceptors, clogP ≤ 3, less than 20 non-hydrogen atoms (MW ≤ 300 Da), not more than three rotatable bonds and PSA(Polar Surface Area) ≤ 60Å^2 [4],[5], [6].

Fragments have the extraordinary advantage of being small chemical entities with a large variety of functional groups able to explore various chemical spaces, penetrate small protein pockets and allow the mapping of various protein binding sites [5], [7], [8].

Due to their low number of atoms they can establish comparably few interactions but of high quality, which makes the establishment of unfavorable and geometrically compromised interactions with the target protein less likely. Instead only the favorable ones will result in successful binding [3]. On the other hand, fragments are often characterized by weak and low overall affinity [3]. This aspect makes it challenging to detect them. This is not crucial in the first step of screening as fragments only provide a starting point to further explore the chemical space and they are versatile starting point for follow-up lead optimization. This consists of growing putative ligands into the binding pocket starting from initial fragment hit as well as linking or merging two or more fragments in proximity together [3]. This step improves the binding properties and helps to morph into a lead compound.

In the past, HTS was the method of choice to screen compound libraries of potential hits. This method requires test compounds with a molecular weight of 300-500 Da, much larger than fragments [3], [9]. Often this comes with the disadvantage of restricting the room for optimization of the pharmacodynamic and pharmacokinetic properties of the discovered hit [10]. Each optimization step will inevitably lead to an increase in size of the starting compound, leading to problems of bioavailability.

Increasingly, fragment-based lead discovery (FBLD) matures to be the method of choice to discover first hits and fragment screening libraries play an increasingly important role in the early phase of the drug development process. In particular, protein X-ray crystallography is the dominant technique to determine the three-dimensional structures of macromolecules, and applied in fragment screening, it directly detects and characterizes the binding modes of the weakly binding fragments providing valuable starting point for lead optimization. In contrast, hits from HTS need elaborate testing to verify their binding and to validate their influence of the target protein, e.g. by occupying its active site.

Considering the importance exhibited by these small molecules for initial drug screening, it is of fundamental importance to have a fragment library ready to use, at best at early stage a general-purpose library, carefully designed and thoroughly validated on a broad scope of proteins belonging to different classes. Such a library has been assembled by us and is commercially available from Jena Bioscience. It was developed in a joint project between the HZB MX-group at BESSY II (AG Weiss) and the Institute of Pharmaceutical Chemistry, University of Marburg (AG Klebe). It consists of 96 entries selected for chemical diversity and has demonstrated a mean a hit rate of almost 9 % across a sample of eight validated proteins.

We chose X-ray crystallography as a primary screening technique to validate our 96 fragment library against the eight different targets without further use of any other biophysical method (such as NMR, TSA, HCS, SPR) to prefilter putative candidates for crystallography. In fact, as reported by *Schiebel et al.* screening strategies using a sequential cascade of biophysical methods prior to X-ray crystallography can be rather misleading as

they might miss an important fraction of fragments finally resulting in a crystal structure [11].

Furthermore, thanks to third-generation synchrotron sources and the latest methodological improvements in automated crystal mounting systems, data collection and processing, it has become increasingly feasible to screen entire libraries of fragments in very short time.

The bottleneck of X-ray crystallography is finding optimal conditions for crystallization and soaking [6]. The optimization of crystals suitable for fragment screening, depending on the protein target, can be elaborate. To detect weakly binding fragments by a sensitive technique such as X-ray crystallography, it is necessary to have good diffracting crystals (< 2.5 Å resolution) in order to clearly identify them in the electron density [10]. Therefore, we selected for this study eight target proteins with known and well-established crystallization and soaking conditions ready for screening against our 96 fragment library.

The goal of the present work is to validate our 96 fragment library against the different proteins to demonstrate its versatility and achievable hit rates for fragments screening campaigns.

1.1.   The 96 fragments library

The design of the present library was made starting from the most promising crystallographic hits found in previous fragment screening studies. These studies involved an in-house 361 fragment library developed by the working group of Prof. Klebe at the Institute of Pharmaceutical Chemistry (University of Marburg) and the 96-fragment library developed by the HZB MX-group at BESSY II (AG Weiss). Both fragment libraries were validated against the aspartic protease Endothiapepsin (EP) using X-ray crystallography and for both a hit -rate of about 10% was found. In particular, 71 fragment hits were detected in the first study while nine fragments were detected in the second study [11], [12].

The assembly of the novel library started with a close examination of the results obtained with the previous screenings, followed by a careful selection aimed at the design of a sub library covering a large portion of chemical space with high chemical diversity.

The library represents a wide range of chemical classes including aromatic and cyclic compounds decorated with functional groups frequently found in drug molecules, Xanthines, amino acid-like compounds and natural compounds. The selection also followed the Ro3, although not strictly, apart from compound stability, good water and DMSO solubility. The library was stored at -20 °C as 1 M fragment stock solution in DMSO. The selected fragments are all of good purity and commercially available at low price, providing wide screening of the chemical space, high hit rate and cost-effectiveness in one go.

Overall the 96 fragments were assembled by 52 entries extracted from the 361-fragment library, 34 compounds from the HZB screening set, which were part of the PDBeChem database and includes six hits found in EP and 28 hits in complex with other protein deposited in the Protein Data Bank (PDB). In addition, seven natural product-like fragments come from the FRGx collection of AnalytiCon and the remaining three compounds were individually retrieved from the PDB and considered for their uniqueness to complement our selection. When several fragments with a similar chemical structure were found as hits from the 361 library, only one representative fragment per chemical class was considered to avoid undesired redundancy and preserve chemical diversity.

Our library, also known as "Frag2Xtal Screen", is available at "Jena Bioscience" and it is "ready-to-use" for soaking experiments. To prepare the crystals, soaking was preferred over co-crystallization as the second technique requires tedious adjustments of the crystallization conditions which will depend on the fragments under investigation. With regard to the screening of the entire library, soaking was performed always using the same conditions for a given target protein independent of the fragment under examination. This accelerates the entire process towards high throughput screening.

The Frag2Xtal Screen consists of 96 fragments dispersed into wells of an MRC 3 crystallization plate. In particular, a 50 nM fragment solution solubilized in DMSO was added and subsequentially dried in two wells while the third one was kept free for further optimization or cryoprotection. Soaking experiments were performed as follows: (i) adding 30 µL of crystallization buffer to the reservoir, (ii) adding 0.5 µL of crystallization buffer onto the dried fragments and (iii) placing 1-2 protein crystals per drop. After sealing the plate, the crystals are incubated for 1-48 h depending on their soaking time and after the crystals were mounted onto cryoloops they are flash frozen.

1.2. Validation of the 96 fragment library

To demonstrate the broad scope of our general purpose fragment library validation against one single target protein would not be sufficient. A larger screening including different target proteins of different classes or biochemical pathways, with different disease background and requiring diverse soaking and crystallization conditions is needed.

We crystallographically screened our library against eight diverse targets: Protein Kinase A (PKA), tRNA-guanine transglycosylase (TGT), Carbonic Anhydrase II (CAII), Thrombin (THR), 17-β-hydroxysteroid-dehydrogenase 14 (HSD-14), Thermolysin (TLN), Farnesyl Pyrophosphate Synthase (FPPS) and Trypsin (TRY).

These enzymes play important roles in drug discovery either as targets or related surrogates involved in multiple diseases such as Alzheimer's disease (EP, FPPS), cancer (EP, PKA, CAII and FPPS), AIDS (EP), malaria (EP), hypertension (EP and CAII), inflammation processes (PKA), glaucoma (CA), obesity (CAII), stroke (THR), myocardial infarction (THR), deep-vein thrombosis (THR), pulmonary Embolism (THR) and even in bacterial diseases such as shigellosis (TGT) or trypanosomiasis as Chagas Disease (FPPS). In addition, EP, TLN, CAII and Trypsin are often used as surrogates for the protease or hydrolyze families with proteins of unknown three-dimensional structures involved in pathogenic diseases.
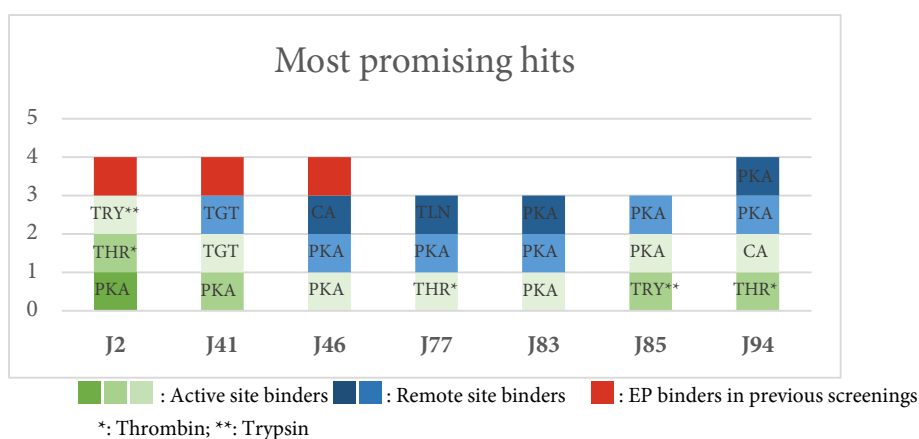
2. Results and Discussion

As mentioned above, a fair part of the library was assembled considering hits discovered with EP. Accordingly for this protein the highest rate of success was obtained with 62 hits corresponding to a hit -rate of 64.58 %. Among the eight additional targets, PKA exhibited a particularly high hit rate with 30 bound fragments (31.25 %), followed by nine hits for Thrombin (9.37 %), eight for CA (8.33 %), seven for Thermolysin (7.29 %), six for TGT (6.26 %), four for Trypsin (4.17 %), three for FPPS (3.12 %) and two for 14-β-HSD (2.08 %). On average, the screening showed an 8.62% hit-rate among the eight target proteins under investigation, excluding EP. Overall 131 complex structures with an average resolution of 1.38 Å were refined and deposited in the PDB. For four targets we subsequently performed an analysis with PanDDA and retrieved another nine fragment structures, which definitely shows that an even higher hit rate can be achieved with our fragment collection. In particular, for Thermolysin and FPPS which were the main focus of the present thesis, PanDDA analysis revealed five and two further fragment hits, respectively. (Chapter II, Section 3.2. and Chapter III, Section 3.4.). It is also remarkable to see that 80 of the 96 fragments showed up in a crystal structure. Considering that 31 fragments were taken from entries in the PDB, overall only six fragments of the total 96 collection did not appear yet in a crystallographic complex.

Out of 96 entries, seven fragments occurred most frequently, thus resulting in the most promising hits. **J2** (for chemical formulae of all entries see Table, Appendix 4) is the fragment that binds to most target proteins (PKA, Thrombin, Trypsin and EP). Interestingly, in all of these targets **J2** binds in the respective catalytic centers. It therefore seems to be the most versatile and suitable fragment to start with for designing an active-site ligand. Possibly its rigid scaffold with partly aromatic character along with polar facilities to experience H-bonds as donor or acceptor provides this fragment with its exceptional but also promiscuous binding capability. The amino acid-like fragment **J77**, D-

Arginine, binds to the catalytic site e.g. of Thrombin, a remote site in PKA, as well as in a sub-pocket next to the catalytic center of Thermolysin. **J94** also bind to three different proteins. In particular CA, where the sulfonamide group of **J94** coordinates the catalytic Zn^{2+} ion. Additionally, it shows up in the active site of Thrombin and in two remote sites of PKA. **J41** binds twice to TGT, either in the active and in a remote site. In addition, the fragment occupies the catalytic site of PKA and also EP. **J46** and **J85** bind both to PKA twice, in the active and in a remote site. Moreover, **J46** binds in a remote cavity of CA and was also found in the EP active site. Furthermore, **J85** is found as active-site hit in Trypsin. The fragment that binds most frequently in three copies to the same protein is **J83** which is found in PKA once in the active site and twice in remote sites.

Table 2.1. Most promising fragments hits

Active site binders : green and light green; Remote site binders : dark blue and blue; EP binders in previous screenings : red
*: Thrombin; **: Trypsin

2.1   Comparison of binding modes

It is interesting to see that a given fragment shows either similar or deviating binding modes and interaction patterns with the investigated proteins. **J2** establishes direct H-bonding interactions with the residues in the ATP-binding site of PKA and the catalytic dyad of EP, whereas in Thrombin and Trypsin interstitial water molecules participate in the binding. Remarkably, in all cases Asp residues participate in contacts with **J2**. As shown in Figure 2.1., the exocyclic amino group of **J2** is within H-bonding distance (3.1 Å and 3.0 Å) to the backbone carbonyl oxygen of Glu170 and the carboxylate group of Asp184 in PKA, while it binds with the same distances to the Asp residues of the catalytic dyad (Asp35 and

Asp219) in EP. Thrombin and Trypsin have similar structures and are involved in similar interactions with **J2**. In Thrombin, the exocyclic nitrogen atoms of **J2** bind to the backbone carbonyl oxygen of Trp250 and the terminal carboxylic acid group of Asp222 *via* two water molecules (3.0 Å and 3.5 Å) but it also interacts with the latter water molecule directly (2.9 Å). In Trypsin, the primary amino group of **J2** serves either as a donor or an acceptor of hydrogen bonds by interacting with the side chain hydroxyl group of Ser190 and the carboxylate group of Asp189 either directly or *via* an interstitial water molecule. The endocyclic nitrogen of **J2** forms in PKA hydrogen bonds as a donor in its protonated form with the carboxylate group of Glu127 and in Trypsin where it interacts with the backbone carbonyl oxygen of Gly219. In addition, it binds most likely in the non-protonated form to EP accepting an H-bond from the OH group of the Thr222 side chain while it binds in the protonated form to Thrombin establishing an H-bond with the terminal carboxyl group of Asp222.
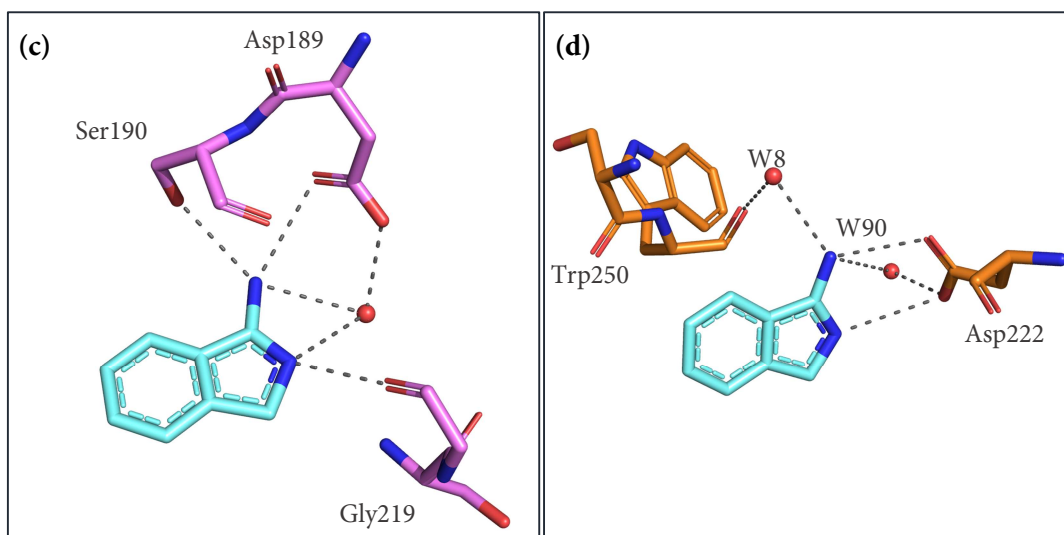

(continued on next page)

10
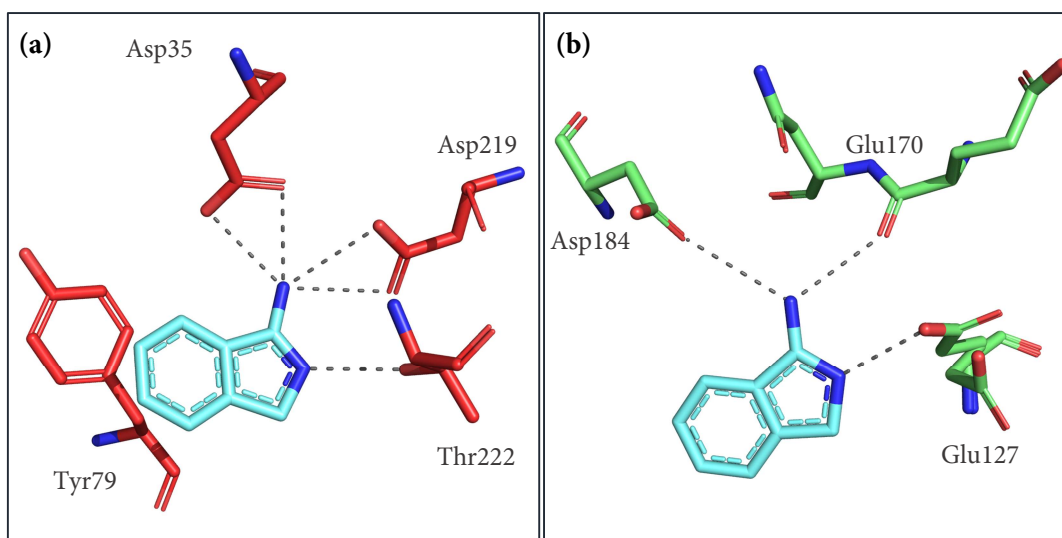
Figure 2.1. Comparison of binding modes of **J2** (aquamarine stick) in complex with **(a)** EP (red stick), **(b)** PKA (green stick), **(c)** Trypsin (violet stick) and **(d)** Thrombin (orange stick). Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

The amino acid-like fragment **J77**, D-Arginine, interacts in the substrate recognition site of PKA in two conformations which refine to an occupancy of 65 % and 53 %. Furthermore, it binds in a remote cavity of Thrombin and next to the S1' and S2' sub-pockets of TLN. The guanidinium group of the fragment establishes electrostatic interactions with the carboxylate group of Glu230 in the active site of PKA and with the carboxylate group of Asp222 in the remote site of Thrombin, whereas in TLN it is the carboxylate part of **J77**, which is involved in charged interactions with the positively charged guanidinium moiety of Arg203 and imidazole ring of His231 side chain. In addition, the negatively charged carboxylate group of the fragment interacts with two water molecules in TLN (W658 and W683) and as many in PKA, where the carboxylate group of **J77** is in double conformation, each of which interact with two glutamine residues *via* an interstitial water molecule. In particular, the higher populated conformer interacts with the carboxylate group of Glu203 side chain through W62 while the less populated one with the Glu170 side chain *via* W110. Furthermore, both conformers interact directly with the guanidinium group of Arg133 side chain. Interestingly, W62 mediates an intramolecular interaction between the negatively charged carboxylate group and the positively charged guanidinium moiety of **J77**. Two water molecules are also found in Thrombin, where, in contrast to TLN and PKA, it

11

interacts neither with the terminal guanidino nor with the carboxylate group of **J77** but with the NH nitrogen atom (3.3 Å and 2.7 Å) and the primary amino group (3.3 Å and 3.0 Å) next to the carboxylic portion of the fragment. Similarly to PKA, in Thrombin, the carboxylate part of the ligand is within hydrogen bonding distance to the hydroxyl function (2.6 Å) and nitrogen backbone of Ser228 (3.3 Å).
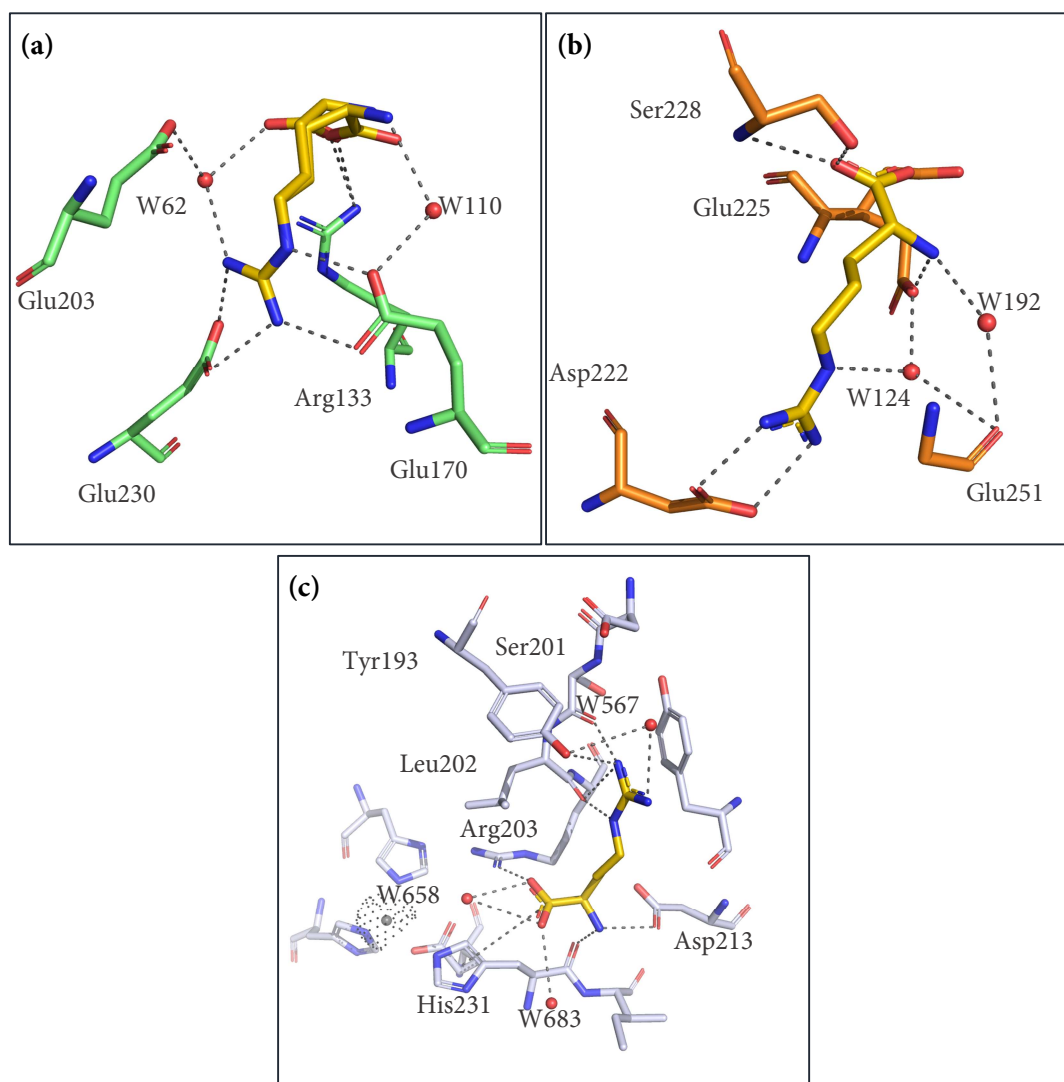
Figure 2.2. Comparison of binding modes of **J77** (yellow stick) in complex with **(a)** PKA (green stick), **(b)** Thrombin (orange stick) and **(c)** TLN (silver stick). Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

The sulfonamide **J94** displays different binding modes in PKA, CA and Thrombin as seen in Figure 2.3. In particular, the sulfonamide group of the fragment coordinates the catalytic Zn^{2+} ion in CA whereas in Thrombin and two remote sites of PKA, **J94** dominantly interacts

with the protein *via* water molecules. **J94** binds to PKA twice with a refined occupancy of 74 % in both cases. At one binding site, **J94** interacts with the protein only indirectly *via* two water molecules which bridge the sulfonamide moiety with the guanidinium group of the Arg45 side chain, the carboxylate function of Glu334 and the backbone carbonyl oxygen of Ile335, as seen in Figure 2.3. (c). A similar binding pattern is found in Thrombin, where the sulfonamide moiety interacts with the backbone carbonyl oxygen of Cys224 (3.0 Å and 3.2 Å) and the backbone nitrogen of Gly251 (3.0 Å and 3.1 Å) mediated through W156 and W56, respectively. In addition, the nitrogen atom and one oxygen atom of the sulfonamide bind directly *via* H-bonding to the backbone carbonyl oxygen of Gly251 (3.3 Å) and the backbone nitrogen atom of Glu225 (3.1 Å). In the second binding site of PKA, the nitrogen atom of the sulfonamide group is involved in a hydrogen bond with the backbone oxygen of Thr299 (3.2 Å) while one oxygen atom forms a direct H-bond with the hydroxyl group of Thr300 (3.2 Å) and the second oxygen atom binds to the nitrogen backbone of Asp301 *via* an interstitial water molecule. In contrast, the nitrogen atom of the sulfonamide moiety in CA is not in contact with any water molecule. In particular, it coordinates the Zn^{2+} ion hold in position by His94, His96 and His119 and it is in H-bonding distance to the hydroxyl function of the Thr199 side chain (2.8 Å). Finally, one of the two oxygen atoms of the sulfonamide is involved in a hydrogen bond with the nitrogen backbone of Thr199 (2.9 Å).
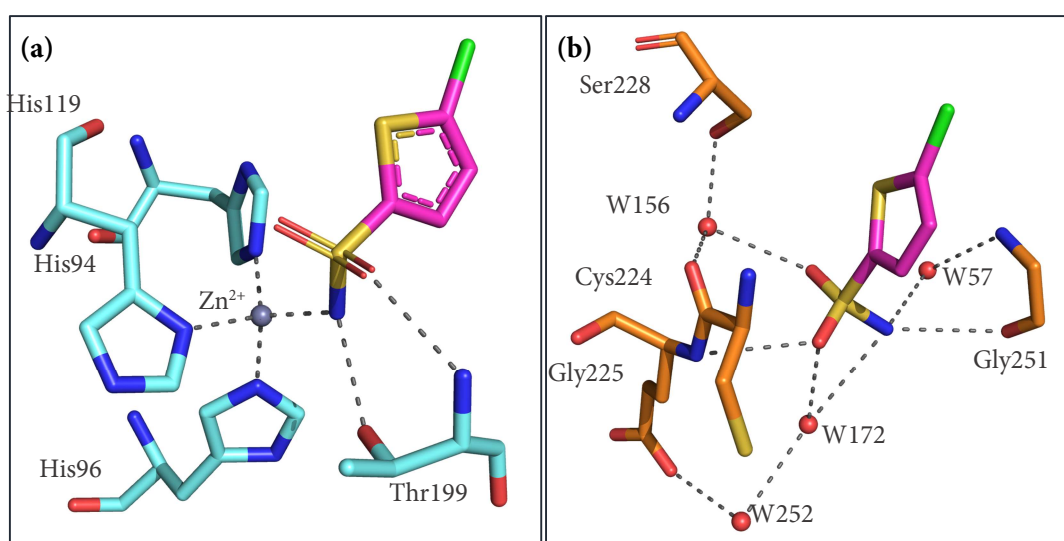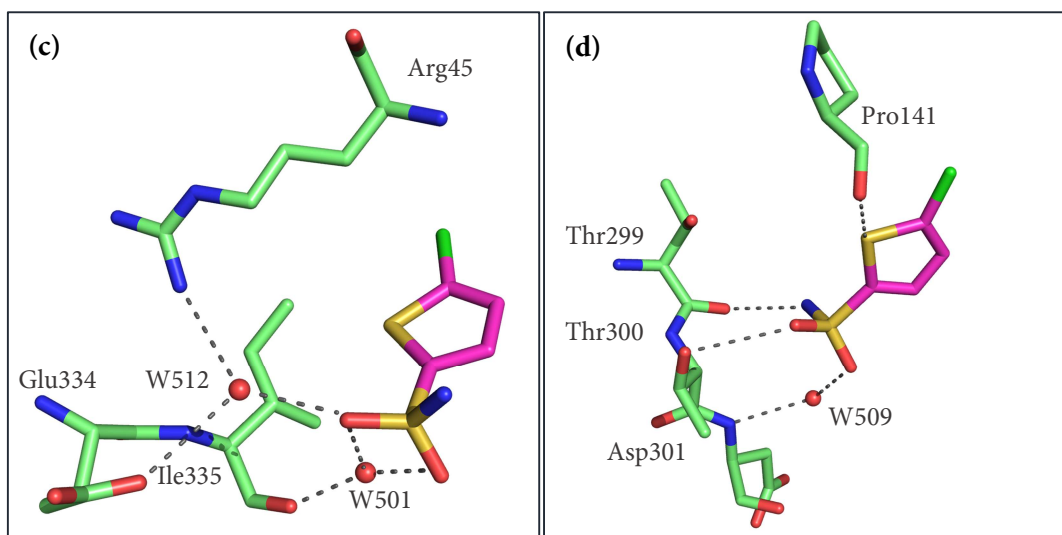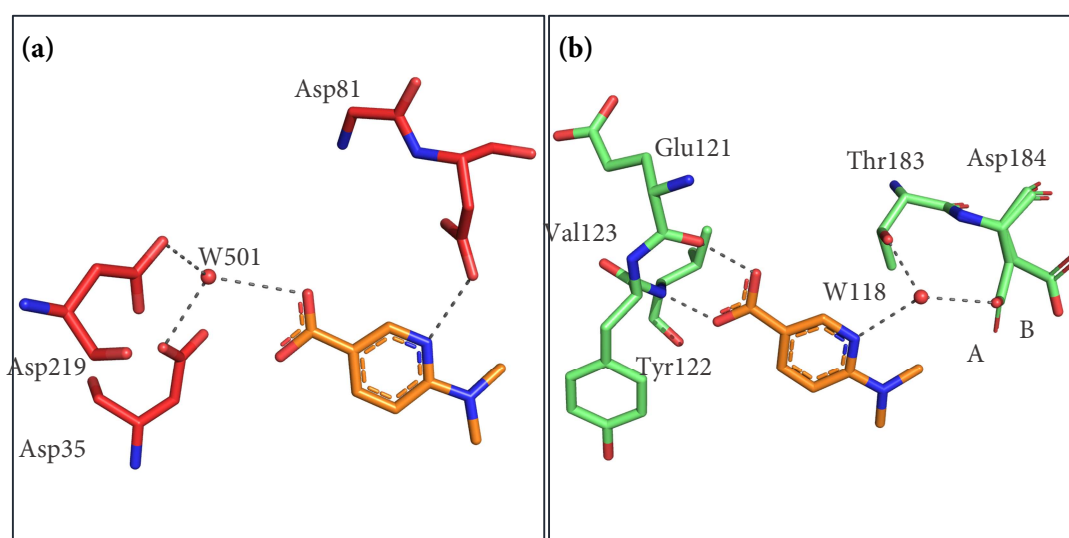
Figure 2.3. Comparison of binding modes of **J94** (magenta stick) in complex with **(a)** CA (aquamarine stick), **(b)** Thrombin (orange stick) and **(c)** PKA (green stick) in the ATP-binding site and **(d)** in a remote site. Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

J41 showed instead a rather conserved binding mode especially in PKA and the two binding sites of TGT. As shown in Figure 2.4., in all cases, the carboxylate group of the fragment formed only direct interactions with the surrounding protein residue whereas it interacts with the catalytic dyad in EP mediated by the catalytic water molecule W501. In particular, the carboxylic acid group of **J41** interacts with the hinge backbone atoms in the ATP-bind site of PKA establishing H-bonds with both, the backbone carbonyl oxygen of Glu121 (2.6 Å) and the backbone nitrogen of Val123 (2.8 Å). This pattern requires the acidic group in protonated form. Similarly, in the catalytic site of TGT, the acidic group is involved in hydrogen bonds with the backbone NH of Gly230 (2.8 Å), the terminal amino group of Gln203 (2.9 Å) and the terminal carboxylate group of Asp156 (2.5 Å). Again, this interaction pattern requires the acidic group of **J41** in protonated state. In a remote site of TGT, the carboxylic acid portion of the fragment is involved in direct H-bonds with the backbone nitrogen (2.9 Å) and the terminal hydroxyl group of Met134 (3.5 Å). In addition, it interacts with one nitrogen atom of His133 side chain (2.7 Å). This pattern does not uniquely define the protonation state of the acidic group and it depends on the protonation state of the imidazole moiety of His133. Notably, the catalytic dyad (Asp36 and Asp219) of

EP which usually interacts directly or *via* the catalytic W501 with ligand donor functionality, it likely also requires **J41** to interact in protonated state at its acidic group. Obviously the electron pushing substituent at the central pyridine moiety help the fragment to be present with an uncharged acidic group. The nitrogen atom of the pyridine moiety of the fragment also interacts either, directly as with the terminal carboxylic group of Asp81 in EP or indirectly through an interstitial water molecule as observed for PKA and at the remote site of TGT. In particular, the pyridine nitrogen in PKA is involved in H-bonds with the hydroxyl function of Thr183 (2.9 and 2.6) and it also forces the carboxylate group of Asp184 to adapt a second orientation. In the remote cavity of TGT, the pyridine nitrogen interacts water-mediated with the nitrogen atom of Ala80 while it establishes no interactions in the active site. In all three proteins for which **J41** was found to be an hit, the terminal tertiary amino group does not establish any kind of interactions whereas the aromatic portion of the ligand is mainly involved in hydrophobic contacts, as the π-π stacking interaction, shown in Figure 2.4 (c), in which it stacks against the Tyr106 in the active site of TGT. Remarkably in the site of EP the pyridine nitrogen requires to be protonated whereas in the water-mediated situations it could either bind as donor or acceptor depending on the adopted protonation state.
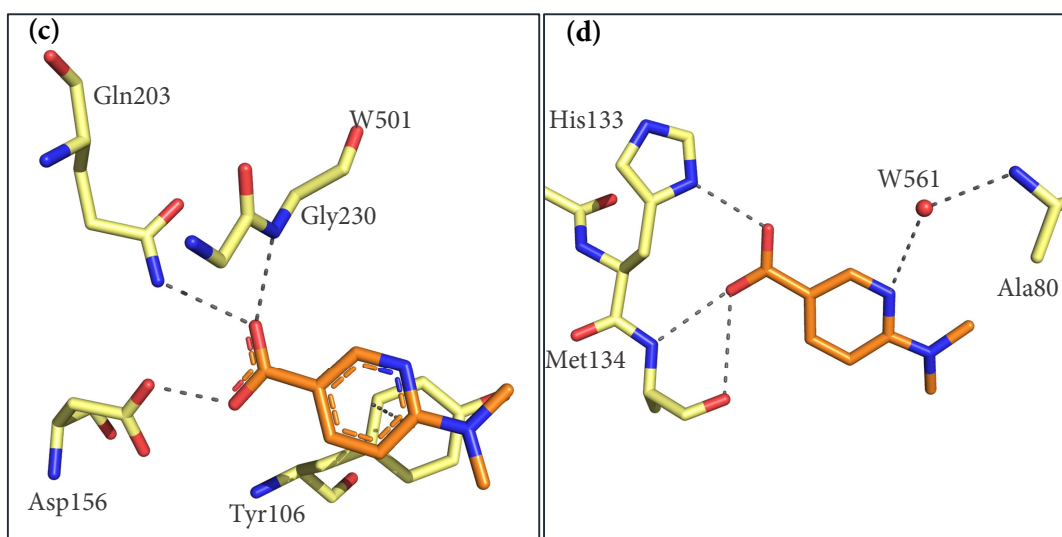

(continued on next page)

Figure 2.4. Comparison of binding modes of **J41** (orange stick) in complex with **(a)** EP (red stick), **(b)** PKA (green stick) and **(c)** TGT (yellow stick) in the active site and **(d)** in a remote site. Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

J46 occupies two sites in PKA, the ATP-binding site and a remote site, and binds in a remote cavity in CA. In addition, it was found to bind to the catalytic site of EP. It is interesting to note that, in CA and in the remote site of PKA, **J46** is accommodated in the protein interface between two symmetry-related molecules. The amide carbonyl oxygen of **J46** establishes direct interactions only in the remote cavity of PKA where it interacts with the nitrogen atom of the Lys21 side chain (2.8 Å). In all other cases, it interacts with a single water molecule as in EP or it is involved in a water-mediated hydrogen bond as with the nitrogen backbone of Val123 in the ATP-binding site. Interestingly, W523 connects the carbonyl oxygen with the primary amine of **J46**, thus bridging between both the functional groups of the fragment. The nitrogen atom of the amide moiety is mostly involved in direct interactions with the oxygen backbone of His260' and the hydroxyl function of the Thr183 side chain in PKA. It also establishes a direct H-bond with the oxygen atom of the carboxylate group of Asp81 in EP or in CA, it binds water-mediated to the acidic group of Glu69. The primary amino group on the opposite site of **J46** is involved in direct hydrogen bonds in all three proteins. In addition, it establishes a contact mediated through the lytic water (W501) with the catalytic dyad of EP. In CA, the primary amine of **J46** is involved in an H-bond with the carboxylate group of Asp72 whereas in the active site of PKA it is within

16

H-bonding distance to the backbone carbonyl oxygen of Glu170 (2.7 Å) and Asn171 (3.0 Å). At the remote site, the primary amine of the ligand binds to PKA through a direct H-bond with the backbone C=O of Ser259'.
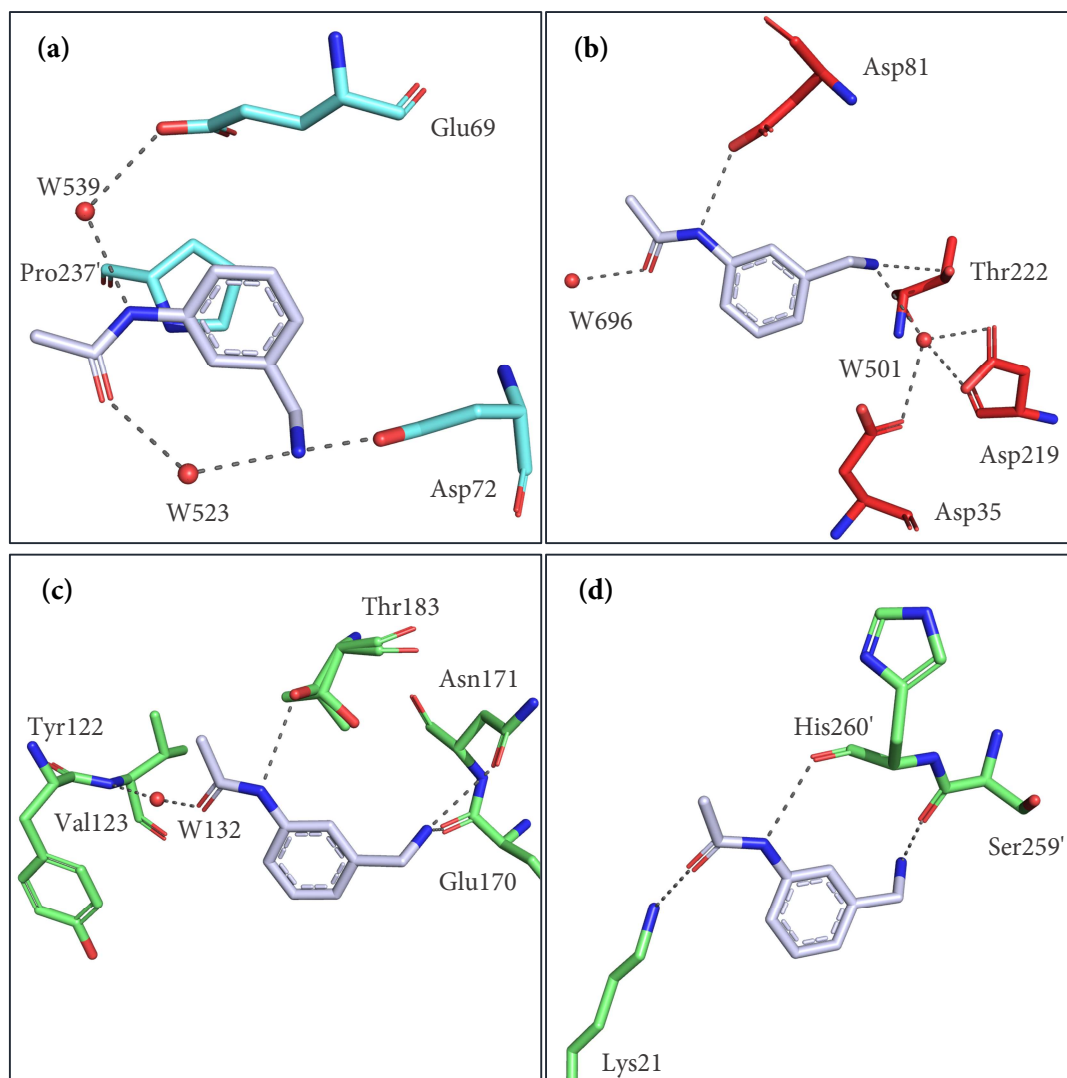
Figure 2.5. Comparison of binding modes of **J46** (silver stick) in complex with **(a)** CA (aquamarine stick), **(b)** EP (red stick) and **(c)** PKA (green stick) in the ATP-binding site and **(d)** in a remote site. Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

At the ATP-binding site, **J85** interacts in a similar fashion as **J41** with the hinge residues. The amide moiety is involved in H-bonds as donor and acceptor with the backbone C=O of Glu121 and the backbone NH of Val123. With a similar pattern in Trypsin, the nitrogen atom of the amide group of **J46** forms an H-bond with the backbone carbonyl oxygen of Gly219 (3.1 Å) while the oxygen atom of the amide moiety interacts with the backbone NH

of Gln192 (3.3 Å). At the meta portion of the molecule, the primary amino group establishes a single direct interaction with the hydroxyl group of Thr183 in PKA while in Trypsin it interacts directly with the hydroxyl group of Ser190 side chain and *via* an interstitial water molecule with the carboxylate group of the Asp189. **J85** accommodates on the PKA surface establishing direct interaction with the surrounding residues. In contrast to what was previously discussed at the ATP-binding site, the terminal amino group of the fragment interacted with the backbone carbonyl oxygen of Phe43 (3.0 Å) and Leu40 (3.2 Å). On the other site of the molecule, the nitrogen atom of the amide interact *via* an H-bond with the backbone C=O of Il35 while its oxygen atom is within H-bonding distance to one nitrogen atom of the guanidinium moiety of the Arg45 side chain (2.9 Å).
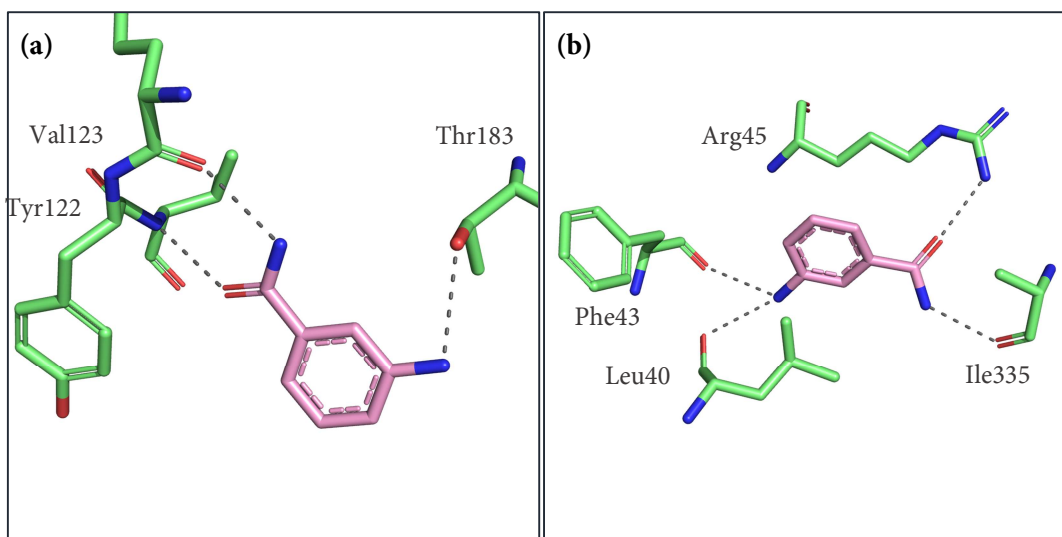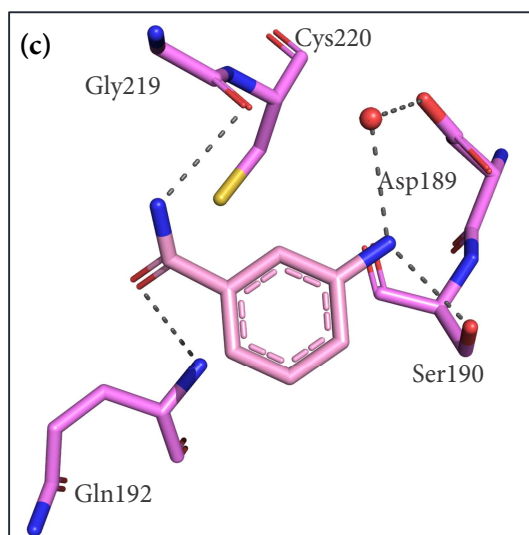
Figure 2.6. Comparison of binding modes of **J85** (pink stick) in complex with **(a)** PKA (green stick) in the ATP-binding site and **(b)** in a remote site and **(c)** Trypsin (violet stick). Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

The fragment **J83** binds in three copies to the same protein. It is found three times in PKA, one of them binds in the ATP-binding site while the other two are at remote sites. Similarly to **J41**, **J46** and **J85**, one portion of the fragment is oriented towards the hinge region whereas the other part is oriented towards the entrance of the catalytic pocket. In the ATP-binding site, the oxygen atom of the oxazole ring interacts *via* an H-bond with the backbone NH of Val123 while the exocyclic amino group interacts most likely in its uncharged form (pKa=4.6) with the surrounding protein residues through two interstitial water molecules. In particular, one water molecule bridges the exocyclic nitrogen with the carboxylate group of Asp184 (2.9 Å and 3.2 Å) while the other one with the backbone C=O of Glu170 (3.1 Å and 3.0 Å). In the remote cavity, one copy of the fragment establishes a direct H-bond with the indole nitrogen atom of the Trp30 side chain (3.1 Å) as well as a water-mediated H-bond with the amidino nitrogen group of Arg93 (3.2 Å and 2.9 Å). The other copy of **J83** instead, does not establish any direct interactions with the protein residues but rather interacts *via* three water molecules which are rearranged in a triangular pattern around the amino group of the fragment.
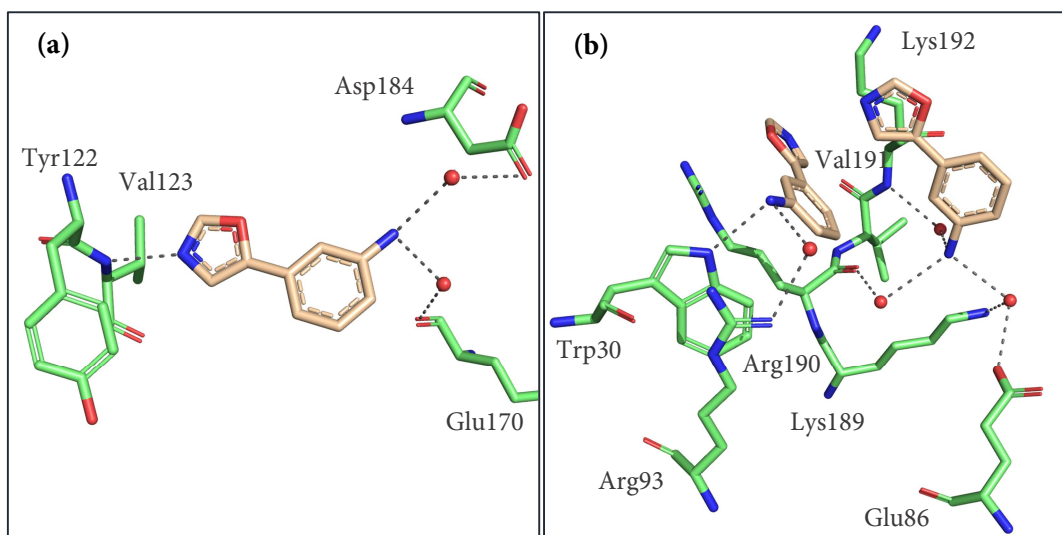
Figure 2.7. Comparison of binding modes of **J83** (wheat stick) in complex with PKA (green stick) in **(a)** the ATP-binding site and **(b)** in a remote site. Water molecules are displayed as red spheres. Protein-fragment interactions are depicted as grey dashed lines.

2.2   What is next? Fragment growing and merging

The discovery of a fragment hit in electron density maps is only the first step in the long process of optimizations that lead to the search for a promising lead candidate, which could then develop into a future drug compound. Fragment growing, linking and merging are the mostly used strategies to enhance affinity of weakly bound fragments.

As mentioned in the previous section, two fragments, **J6** and **J15,** were found to bind to 17β-HSD14. These fragments did not show any significant inhibitory activity against the target protein, even at 250 μM fragment concentrations [13]. However, compound **4** which was derived from a growing strategy applied on **J15,** showed the highest affinity among the series of optimized compounds with a K_i of 1.9 μM [13]. In addition, compound **9,** which originated from the linking of **J6** with **J15,** exhibited 67 % of inhibitory activity against 17β-HSD14 and a K_i of 17.6±0.5 [13]. These two optimization strategies thus led to an increased affinity compared to the starting fragments. X-ray crystal structures were also obtained for the optimized compounds.

Compound	R	COOH position
2	H	3
3	H	4
4	OH	3
5	OH	4
6		3
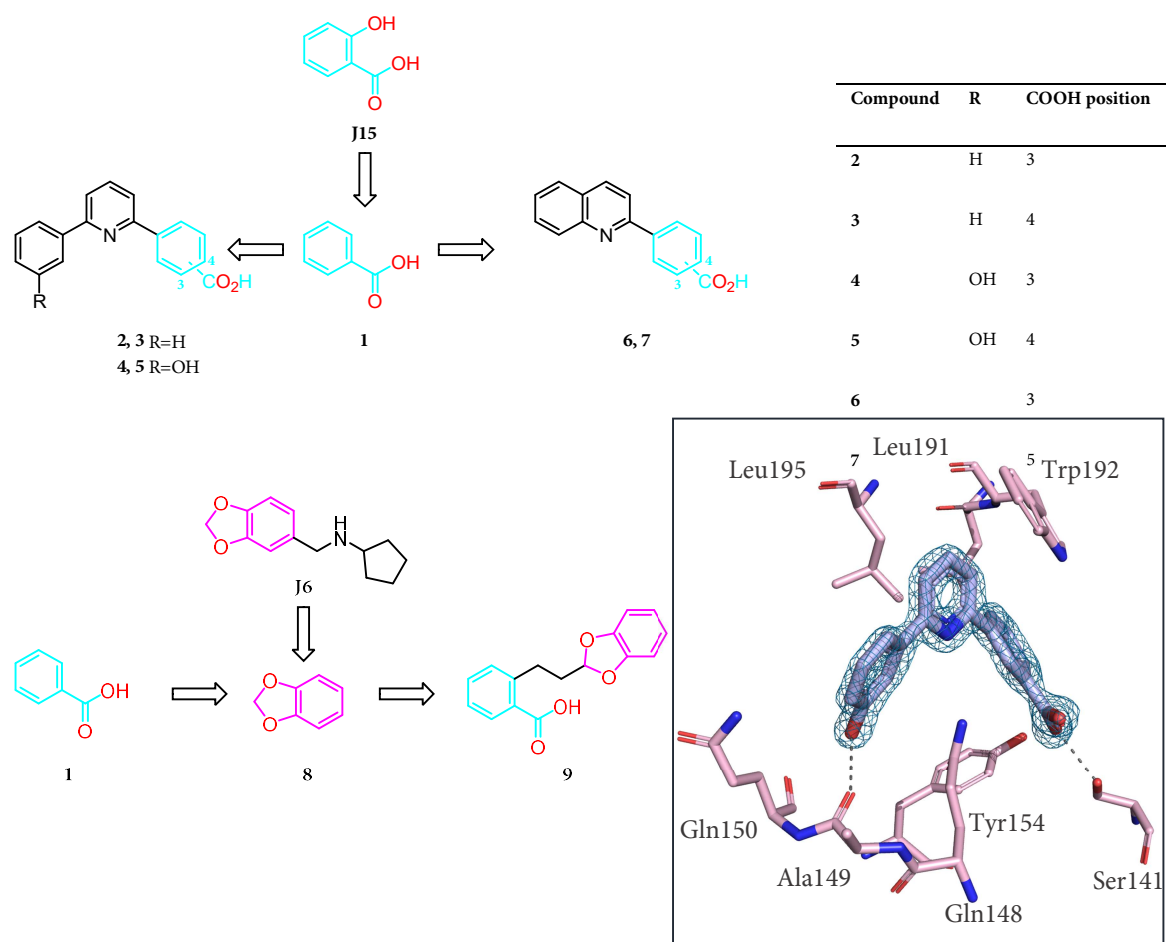
Figure 2.8. In the upper left, growing of **J15** into the optimized compound **2**, **3**, **4**, **5**, **6** and **7**. In the upper right, table showing the nature of the substituent R and the position of R and COOH attached to the benzene ring. In the lower left, merging of **J6** and **J15** into the optimized compound **9**. In the lower right, binding mode of **4** (dark violet) in complex with 17β-HSD14. 2mFo-DFc maps are contoured at 1 σ and represented as a blue mesh.

21
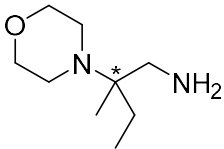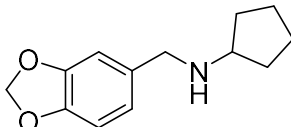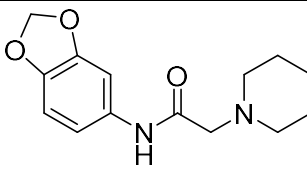
3. Conclusion
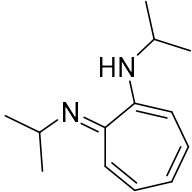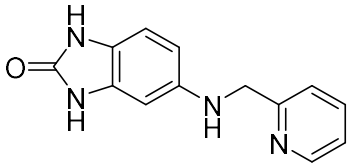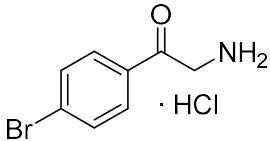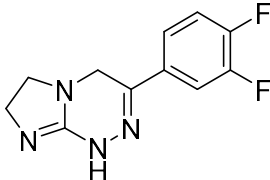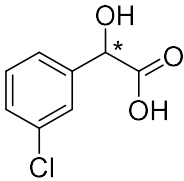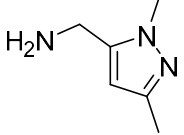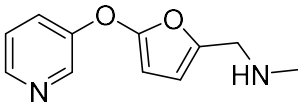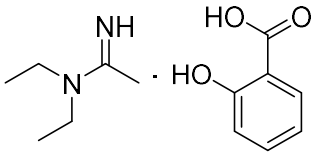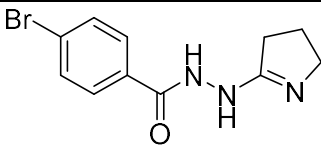
The present fragment library, although consisting of only 96 entities, proved to be very effective with a hit rate of almost 9 % as an average among the eight-screened proteins. The library was also successful in identifying versatile and promising starting points for the future development of a lead compound characterized by chemical diversity. In fact, in addition to the already-known warheads, the 96-library identified new groups, which addressed the catalytic pockets of the target proteins under investigation.
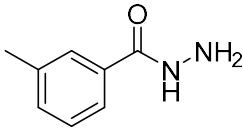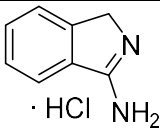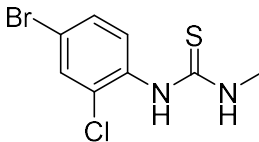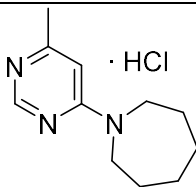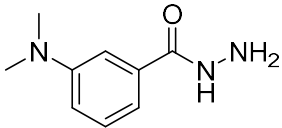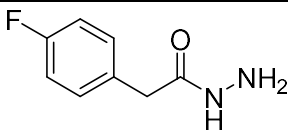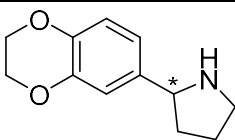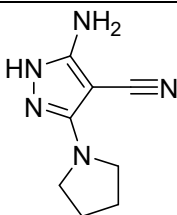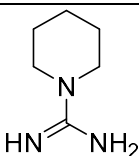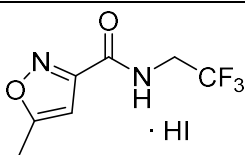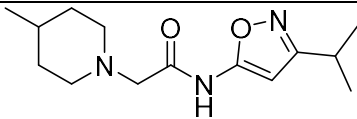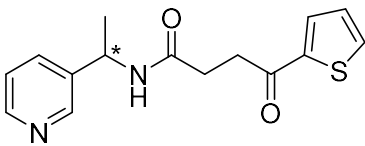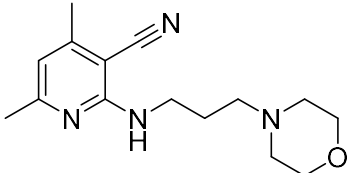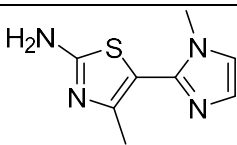
Even proteins as EP, PKA, CA and TLN whose 3D structure is known and studied for a quite some time, were of great interest as they revealed new chemical scaffolds which were not found in the PDB. Phenol, pyridine, isoquinoline, indazole, azaindole, diazole, triazole and oxadiazole are few examples of functional groups which usually bind to the ATP-binding site of PKA. In addition to those, carboxy-ester (**J32**), amidine (**J33**), carboxylate group (**J41**), nitro group (**J65**) and oxazole ring (**J83**) were found by screening the present library. Furthermore, the catalytic dyad (Asp36 and Asp219) of EP usually interacts directly or *via* the catalytic W501 with donor functionalities as primary and secondary amines, hydrazide and piperidine. Interestingly, as previously reported [14], it also interacts with acceptor functionality as the carboxylic acid of **J41**. Moreover, new chemical entities coordinated the catalytic Zn^{2+} ion of CA and TLN. Some examples included hydrazide (**J1** and **J17**) in CA and carboxamide (**J96**) in TLN.
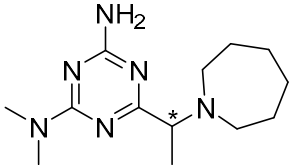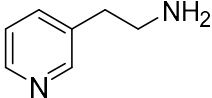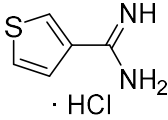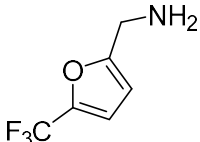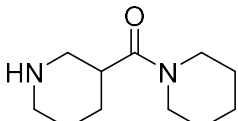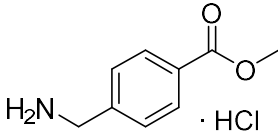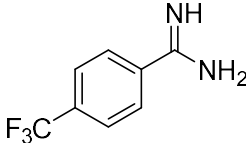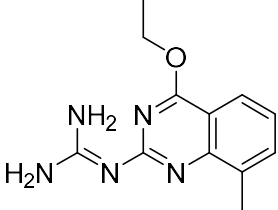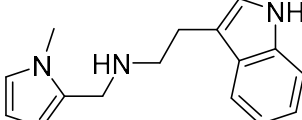
Considering the successful hit-rate and the possibility of increasing the affinity of the starting compounds by fragment growing, linking or merging strategies, the present 96-fragment library could be a promising crystallographic primary screening for the discovery of novel compounds characterized by versatility, innovation and chemical diversity.
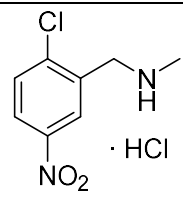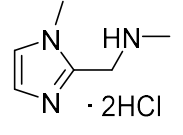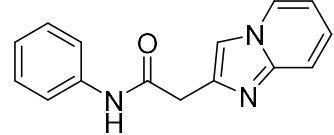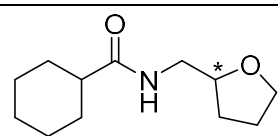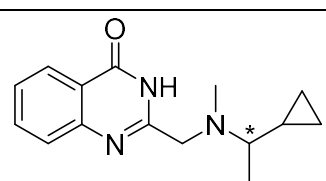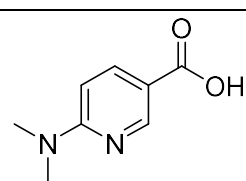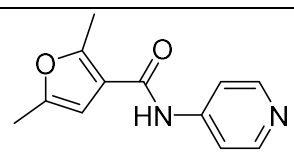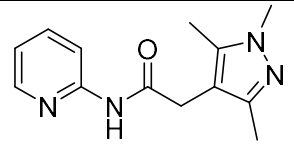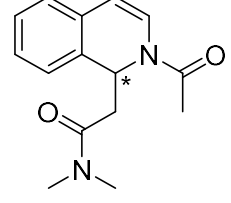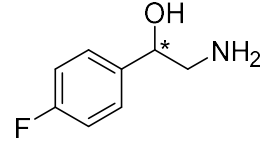
4. Appendix
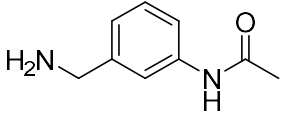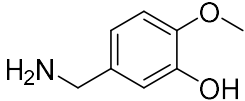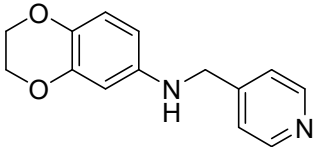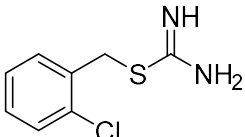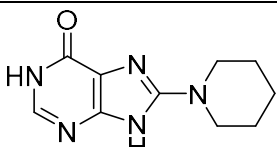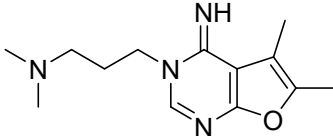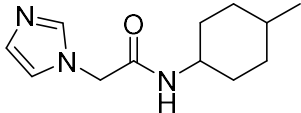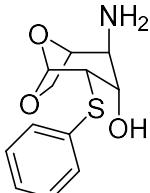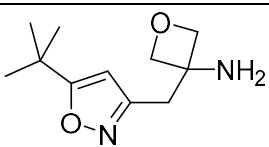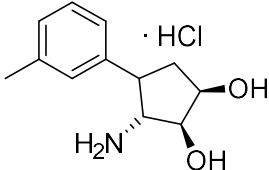
Jena ID	Structure	MW (g/mol)	Target Protein	Resolution (Å)	PDB IDs
1		150.18	CA	1.68	6RM1
			EP	1.59	4Y36
2		168.62	PKA	1.58	5N36
			Thrombin	1.45	n. d.
			Trypsin	1.11	n. d.
			EP	1.25	4Y3E
3		279.58	EP	1.75	4Y4U
4		227.74	EP	1.55	4Y3P
5		172.27	EP	1.58	4Y3W
6		219.28	14HSD	2.25	6GTU
			PKA	1.59	5N37
			EP	1.48	4Y3D
7		262.31	PKA	1.53	5N1E
			EP	1.12	4Y3Z
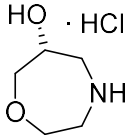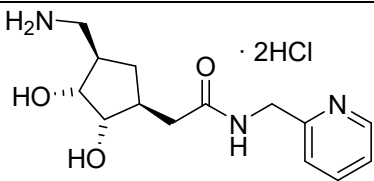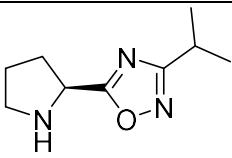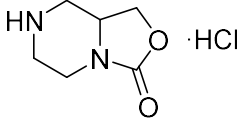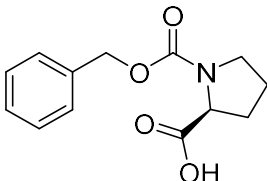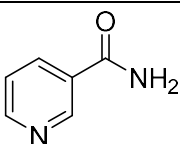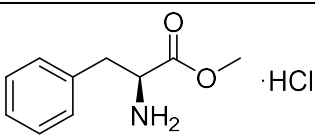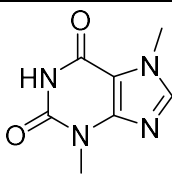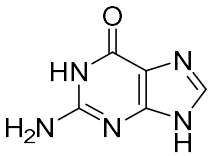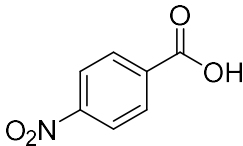
8		204.32	EP	1.48	4Y43
9		240.27	EP	1.67	4Y4X
10		250.52	Thrombin EP	1.40 1.60	n. d. 4Y51
11		236.23	CA EP	0.95 1.62	6S9Z 4Y53
12		186.59	EP	1.61	4Y54
13		125.18	TLN CA EP	1.48 1.02 1.63	6SBK 6SAC 4Y56
14		204.23	TGT PKA EP	1.45 1.61 1.49	6SFO 5N1D 4Y57
15		252.31	14HSD CA EP	2.47 1.08 1.48	6EMM 5M78 4Y4Z
16		282.14	EP	1.20	4Y39
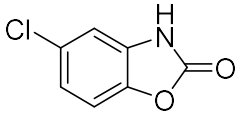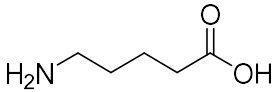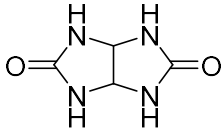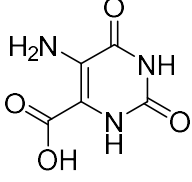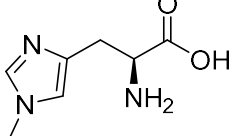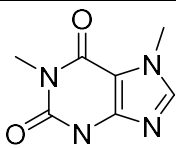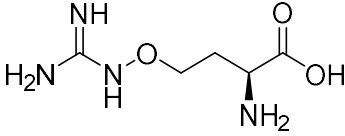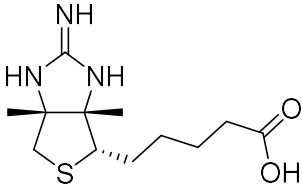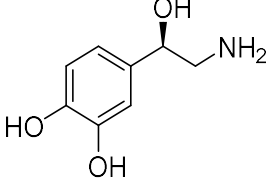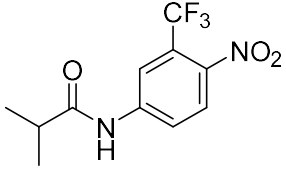
17		179.22	CA	1.22	6RMP
			EP	1.32	4Y50
18		168.17	Thrombin	1.17	n. d.
			EP	1.55	4Y3M
19		205.26	PKA	1.45	5N39
			EP	1.0	4Y41
20		177.21	PKA	1.50	5N7P
			EP	1.30	4Y4T
21		127.19	Trypsin	1.19	n. d
			EP	1.45	4Y4W
22		336.05	TLN	1.53	6SC0
			EP	1.30	4Y4E
23		265.36	EP	1.24	5J25
24		288.37	EP	1.19	4Y47
25		274.37	EP	1.24	4Y44
26		194.26	PKA	1.40	5N3A
			EP	1.25	4Y3X
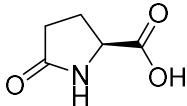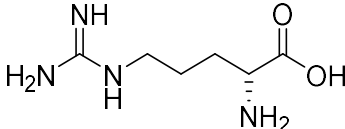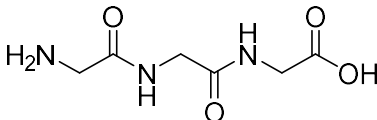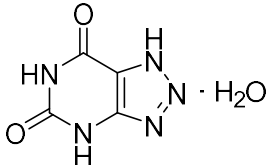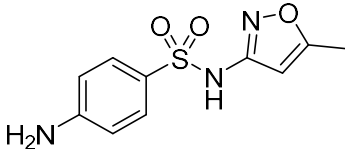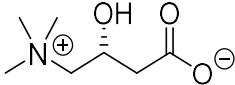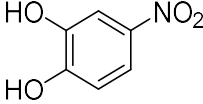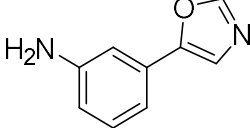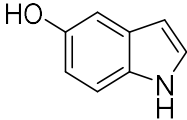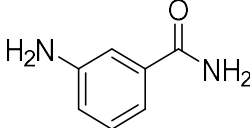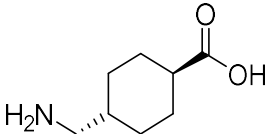
27		264.38	EP	1.17	4Y3A
28		122.17	TLN	1.30	6SB9
			PKA	1.64	5N3B
			EP	1.23	4Y3H
29		162.64	Trypsin	1.25	n.d.
			PKA	1.77	5N3C
			EP	1.16	4Y3L
30		165.12	EP	1.45	4Y5A
31		196.29	EP	1.42	4Y3T
32		201.65	PKA	1.18	5N1H
			EP	1.07	4YCK
33		188.15	PKA	1.77	5N3D
			EP	1.13	4YCT
34		245.29	EP	1.70	4YCY
35		253.35	EP	1.25	4YD3
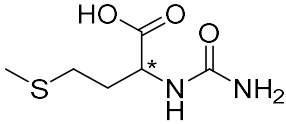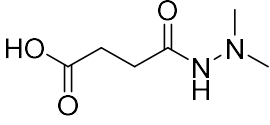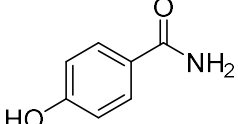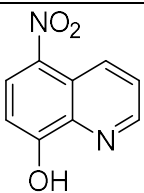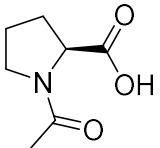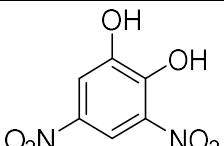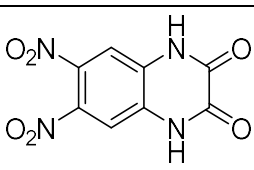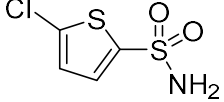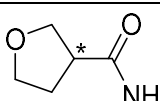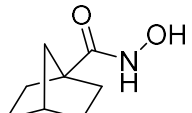
36		237.08	EP	1.21	4YD5
37		125.18	EP	1.30	4YD6
38		251.29	EP	1.42	4YD7
39		211.31	EP	1.17	4Y58
40		257.34	EP	1.24	4Y5B
41		166.18	TGT PKA EP	1.38 1.53 1.10	5SW3 5N3E 4Y3S
42		216.24	PKA EP	1.49 1.19	5N1L 4Y5C
43		244.30	EP	1.47	4Y5G
44		258.32	EP	1.34	4Y3N
45		155.17	PKA EP	1.80 1.44	5N1K 4Y5K

46		164.21	CA	1.10	6SAS
			PKA	1.68	5N3F
			EP	1.13	4Y3G
47		153.18	PKA	1.44	5N1M
			EP	1.28	4Y4A
48		242.28	EP	1.06	4Y45
49		200.68	EP	1.13	4Y3R
50		219.25			
51		248.33	FPPS	1.62	6SF8
			EP	1.28	4Y5M
52		221.30	EP	1.23	4Y5P
53		253.32	EP	1.20	
54		210.28	Thrombin	1.42	n. d.
			EP	1.75	n. d.
55		243.73	EP	1.45	n. d.

56		153.61	PKA EP	1.16 1.70	5N3G
57		279.34	EP	1.50	n.d.
58		181.24			
59		178.62	CA	0.95	6SAY
60		249.27			
61		122.13	PKA	1.36	5N3H
62		215.68	TLN PKA EP	1.82 1.14 1.10	6SC3 5N3I 4Y38
63		180.17			
64		151.13	TGT	1.12	5N6F
65		167.12	PKA	1.12	5N3J

66		169.56			
67		117.15			
68		142.12			
69		171.11			
70		169.18	EP	1.31	4Y3J
71		194.19	FPPS EP	1.59 1.27	6SF9 4Y4D
72		176.18	TGT PKA EP	1.36 1.33 1.44	5UTI 5N3K 4Y4G
73		271.38	EP	1.20	4ZEA
74		169.18	Thrombin PKA EP	1.43 1.38 1.03	n. d 5N3L 4Y4J
75		276.22			

76		129.12			
77		174.20	TLN Thrombin PKA	1.41 1.40 1.23	6SCK n. d. 5N3M
78		189.17			
79		171.12	TGT	1.55	5UTJ
80		253.28	Thrombin	1.48	n. d.
81		161.2	PKA	1.22	5N3N
82		155.11	FPPS	1.51	6SDN
83		160.18	PKA	1.32	5N3O
84		133.15	PKA	1.58	5N3P
85		136.15	Trypsin PKA	1.36 1.31	n. d 5N3Q
86		157.21	TGT	1.42	5V3C

87		192.23	PKA	1.36	5N3R
88		160.17	TLN	1.42	6SCU
89		137.14	Thrombin PKA	1.49 1.14	n. d. 5N3S
90		190.16			
91		190.16			
92		200.11			
93		252.14			
94		197.65	CA Thrombin PKA	1.09 1.45 1.21	6SB7 n. d. 5N3T
95		115.13			
96		155.20	TLN	1.56	6SC1

n. d: not deposited in the PDB yet.

5. References

[1]   G. Bollag *et al.*, "Vemurafenib: The first drug approved for BRAF-mutant cancer," *Nat. Rev. Drug Discov.*, vol. 11, no. 11, pp. 873–886, 2012.

[2]   C. Zhang *et al.*, "Design and pharmacology of a highly specific dual FMS and KIT kinase inhibitor," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 14, pp. 5689–5694, 2013.

[3]   D. E. Scott *et al.*, "Fragment-based approaches in drug discovery and chemical biology," *Biochemistry*, vol. 51, no. 25, pp. 4990–5003, 2012.

[4]   H. Köster *et al.*, "A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes," *J. Med. Chem.*, vol. 54, no. 22, pp. 7784–7796, 2011.

[5]   F. Giordanetto *et al.*, "Fragment Hits: What do They Look Like and How do They Bind?," *J. Med. Chem.*, vol. 62, no. 7, pp. 3381–3394, 2019.

[6]   M. J. Hartshorn *et al.*, "Fragment-based lead discovery using X-ray crystallography," *J. Med. Chem.*, vol. 48, no. 2, pp. 403–413, 2005.

[7]   J. Schiebel *et al.*, "One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists," *ChemMedChem*, vol. 10, no. 9, pp. 1511–1521, 2015.

[8]   R. . E. Hubbard, "Part I The Concept of Fragment-based Drug Discovery," *Fragm. Drug Discov. Lessons Outlook*, pp. 1–36, 2016.

[9]   C. W. Murray *at al.*, "Experiences in fragment-based drug discovery," *Trends Pharmacol. Sci.*, vol. 33, no. 5, pp. 224–232, 2012.

[10]   J. Schiebel *et al.*, "High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits," *Structure*, vol. 24, no. 8, pp. 1398–1409, 2016.

[11]   J. Schiebel *et al.*, "Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study," *ACS Chem. Biol.*, vol. 11, no. 6, pp. 1693–1701, 2016.

[12]   F. U. Huschmann *et al.*, "Structures of endothiapepsin-fragment complexes from crystallographic fragment screening using a novel, diverse and affordable 96-compound fragment library," *Acta Crystallogr. Sect. FStructural Biol. Commun.*, vol. 72, pp. 346–355, 2016.

[13]   N. Bertoletti, "Structural Characterization of 17 β -Hydroxysteroid Dehydrogenase Type 14 and Inhibitor Optimization Using Crystallography and Computational Techniques," *PhD Thesis*, 2017.

[14]   N. Radeva, "From Active-Site Mapping to Lead Discovery using Fragment-based Approaches on the Aspartic protease Endothiapepsin Dissertation," *PhD Thesis*, 2016.

Part II: Thermolysin (TLN)

Crystallographic screening of an inhouse fragment library

1. Introduction

Thermolysin (TLN; EC 3.4.24.27) is a thermostable zinc metalloproteinase secreted by the bacterium *Bacillus thermoproteolyticus*. TLN and TLN-like proteins are used by bacteria to break down exogenous proteins for nutrition and as virulence factors aiding in host colonization and tissue degradation. TLN belongs to the endopeptidase subfamily and its activity consists of the hydrolysis of amide peptide bonds along the peptidic backbone at the N-terminal end next to large hydrophobic amino acids such as Leu, Ile or Phe.

1.1. Thermolysin mechanism of reaction

The generally accepted mechanism for TLN-mediated hydrolysis proceeds *via* the two-step reaction depicted in Scheme 1.1 [1].
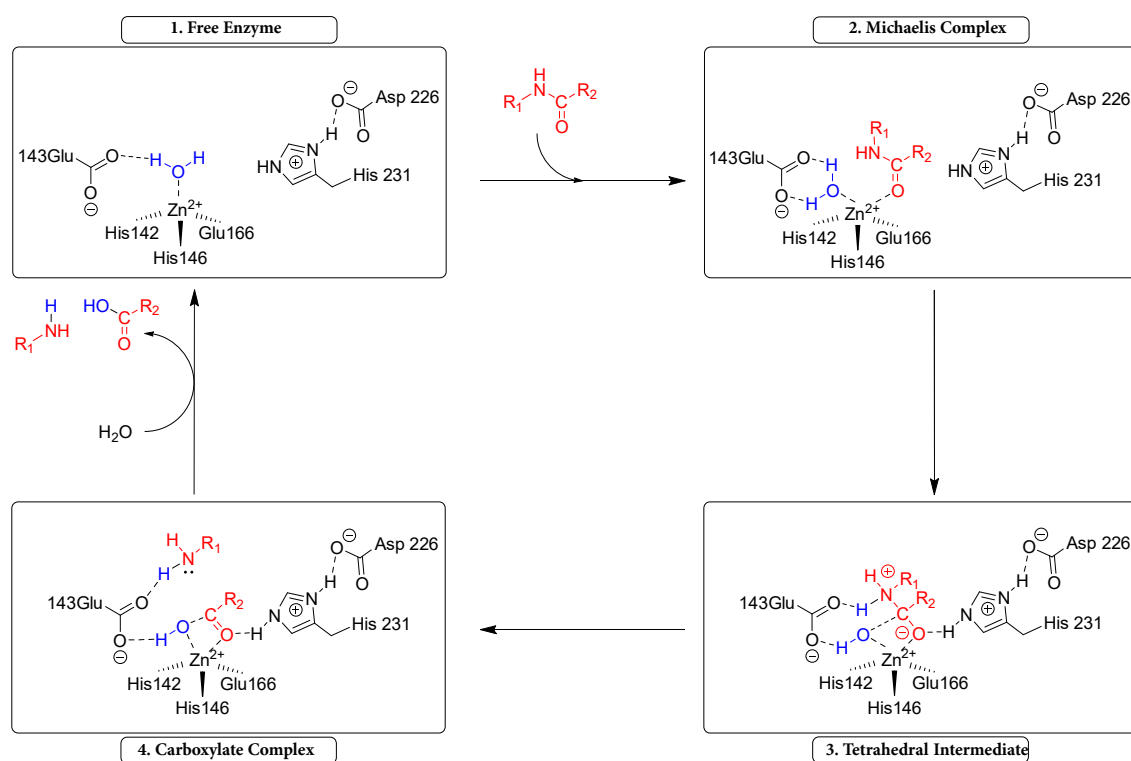
Scheme 1.1. Experimentally proposed reaction cycle for TLN. Picture adapted from[1].

In the active site, the catalytic Zn^{2+} ion is tetrahedrally coordinated with three residues belonging to the HEXXH motif (His 142, His 146 and Glu 166) which is typical for metalloproteinases and the catalytic water molecule (Scheme 1.1., 1-Free Enzyme).

An incoming peptide substrate expands the coordination geometry at the Zn^{2+} ion and shift the catalytic water molecule towards Glu143. In the reaction process, the Zn^{2+} ion associates with the carbonyl oxygen of the substrate forming the enzyme-substrate complex (ES), also known as ES-Michaelis complex (Scheme 1.1., 2- Michaelis complex). The mutual influence of the glutamate side chain and the metal ion strongly enhances the nucleophilicity of the water oxygen atom, promoting an attack on the carbonyl carbon atom of the substrate. The proton accepted by Glu143 is immediately shuttled to the nitrogen atom of the former substrate peptide bond. As a result, a tetrahedral geminal diolate intermediate is formed, in which the Zn^{2+} ion is pentacoordinated and the peptide carbon is tetrahedrally coordinated (Scheme 1.1., 3-Tetrahedal Intermediate). The nucleophilic attack by water on the substrate going from the ES-Michaelis complex to the geminal diolate is the rate-limiting step for the hydrolysis.

The negatively charged carboxylate motif of Asp226 hydrogen-bonded to the protonated side chain of His231 helps to stabilize the positive charge required for catalysis.

The carbon-nitrogen (C-N) bond of the former substrate peptide bond in the geminal diolate intermediate can be considered highly polarized and therefore the intermediate complex breaks relatively easy apart.

In the next step, a direct cleavage of the peptide C-N bond takes place, with the amine product proposed to be released in its protonated form, in accordance with the neutral pH conditions. Again, Glu143 serves here as a proton shuttle, abstracting the second proton from the water oxygen atom and delivering this proton to the amine nitrogen atom (Scheme 1.1., 4-Enzyme-Carboxylate Complex). The catalytic cycle is closed by the carboxyl product release, probably mediated by the incoming water molecule uptake[1].

The catalytic activity of TLN is dependent on both, temperature and pH. The maximum catalytic activity has been measured to be at pH 7.2 and 40 °C. As a result of the thermostability provided by the protein-bound calcium ions, no significant loss of activity or conformational stability can be recorded below 70 °C[2].

1.2.   Thermolysin structure

The first three-dimensional structure of Thermolysin had already been solved in 1972[3]. It consists of 316 amino-acid residues organized into two domains which are separated by the active site. The N-terminal domain (green) consists of several β-strands and one long α-helix. The C-terminal domain (red) is rich in α-helices (Figure 1.1.). The active site of the protease is located in the cleft which is formed by the two domains. A central helix (blue) connects the two domains. The central α-helix contains the typical HEXXH sequence coordinating the Zn^{2+} ion (black dots). Furthermore, Thermolysin contains four Calcium ions important for the thermostability and for the correct folding of the protease (Figure 1.1.).
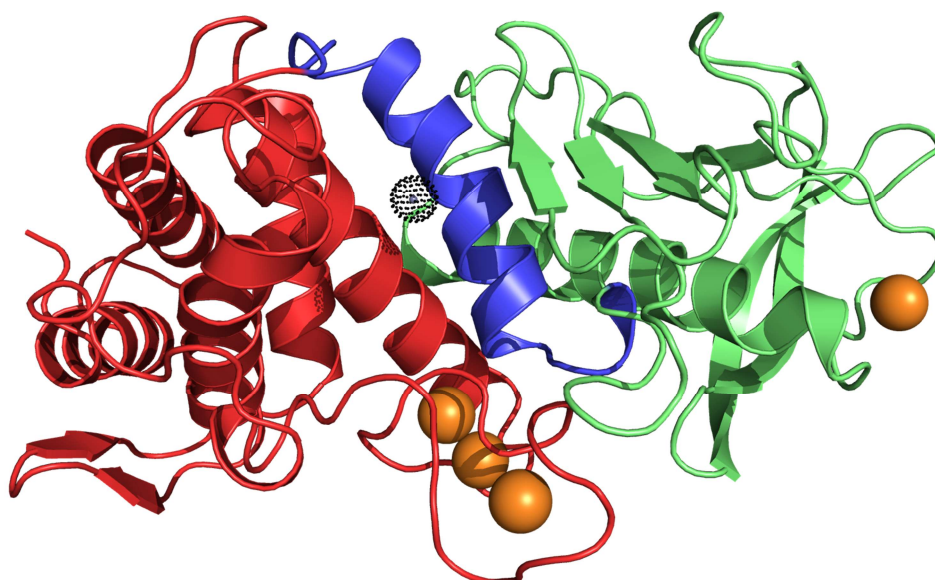
Figure 1.1. Cartoon representation of Thermolysin secondary structure (PDB 6SBK). The central helix (blue) connects the N-terminal domain rich in β-strands (green) and the C-terminal domain rich in α-helices (red). The center of the catalytic Zn^{2+} ion is represented as a grey sphere while the van der Waals radius is depicted as black dots. Four Ca^{2+} ions important for protein stability are represented as orange spheres.

The TLN active site is divided into three subpockets:

- The S1' pocket is a deep and well-defined cavity surrounded by Leu133, Val139, Ile188 and Arg203. It is important for substrate recognition and accommodates hydrophobic amino acid side chains. The guanidinium group of Arg203 can form hydrogen bonds with the carbonyl backbone of the peptide substrate.

- The S2' pocket is a shallow, open bowl-shaped cavity which is surrounded by Leu202, Phe130, Asn111, Asn112 and Ala113 and is rather hydrophobic and unspecific. Ligands can form H-bonds with Asn112 and Ala113. It is more easily accessible by water molecules compared to the nearby S1' pocket.

- The S1 pocket is a large and unspecific pocket and more solvent exposed than the former ones. It is well accessible to water molecules and it is delimited by Phe114, Trp115 and Tyr157.

A closer look at the active site shows that the crystal structure of the putatively free Thermolysin has a Val-Lys dipeptide bound occupying the region between the S1' and S2' pockets (Figure 1.2.) There is no crystal structure with empty pockets in the Protein Data Bank (PDB). This peptide is a product of the autoproteolysis of Thermolysin in concentrated solutions cleaved from the C-terminus of the protein and it is inevitably present during the crystallization process.
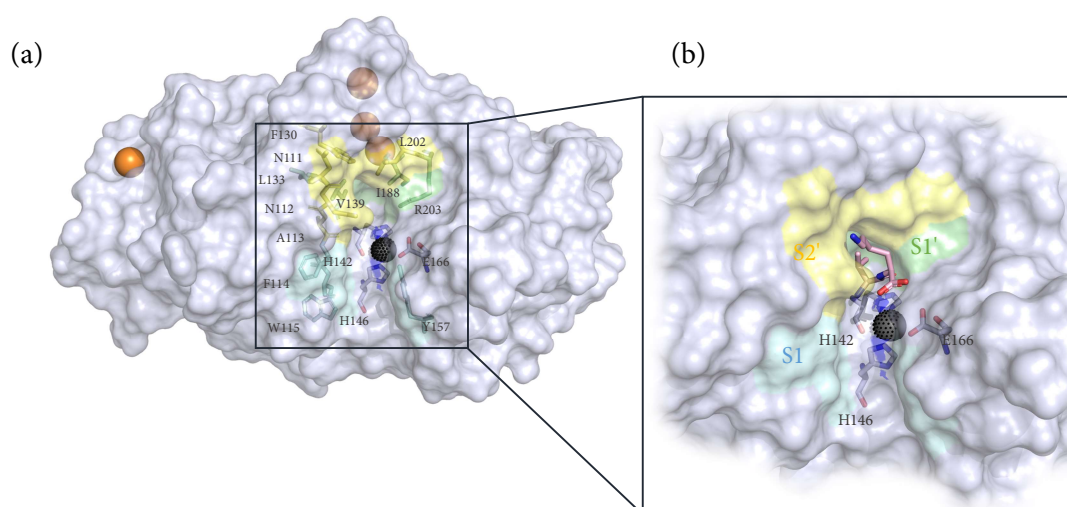
Figure 1.2. (a) Surface representation of Thermolysin. The active site is divided into three sub-pockets represented as a colored surface and as a stick representation. In particular, the specific S1' pocket is depicted in green, the hydrophobic S2' in yellow and the solvent exposed S1 pocket in cyan. The catalytic Zn^{2+} ion (black sphere) is tetrahedrally coordinated with the HEXXH motif (His142, His146 and Glu166) shown as a blue-white stick model with color-coded heteroatoms (nitrogen atom in blue and oxygen atom in red). The four Ca^{2+} ions are represented as orange spheres. **(b)** Close-up view of the active site hosting the autocleavage product Val-Lys dipeptide represented as a stick model and colored in pink with color-coded heteroatoms. The same color codes, dashes, sphere, sticks and surface representations will be applied to all following figures.

Thermolysin itself is not important for any pathogenic processes in humans. However, it is a well studied model system for metalloproteinases with unknown three-dimensional structures which are involved in pathogenic processes, for example the angiotensin converting enzyme (ACE) implicated in hypertension [4], the human endothelin converting-enzyme (ECE-1) associated with human breast cancer [5]and heart failure as well as the recent work on β-amyloid peptide against the human neutral endopeptidase (Neprilysin) involved in Alzheimer's disease[6].

1.3.   Aim of the project and state of the art

The present work aimed at finding the optimal soaking conditions suited for a crystallographic screening of an in-house general purpose 96 fragment library. Until now many studies have been done on TLN in complexes with inhibitors. This is confirmed by the great abundance of literature on the subject, including the 250 deposited PDB structures. Most of the previous studies aimed at designing and optimizing substrate and transition state analogs. They are usually peptide-like molecules able to complex the catalytic Zn^{2+} ion and decrease the catalytic activity of TLN and TLN-like proteins by mimicking the transition state of the catalytic reaction. This is the case of phosphoramidates, a class of phosphorous-containing amino acids and peptide-like compounds that mimic the tetrahedral intermediate of the TLN hydrolysis mechanism[7]. A series of such compound is described by *Krimmer et al.*[8]. Here, the phosphonamidate moiety is the central core of the molecular scaffold and interacts with the catalytic Zn^{2+} ion while the carboxybenzyl group interacts with the solvent-exposed S1 pocket and the leucine side chain occupies the deep and specific S1' pocket. Different hydrophobic substituents (R1 and R2) were synthesized to analyze the water network patterns around the S2' pocket (Figure 1.3.).
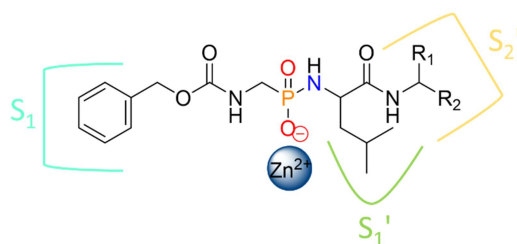
Figure 1.3. Typical molecular scaffold of TLN inhibitors. The phosphonamidate moiety coordinates the Zn^{2+} ion and is decorated with hydrophobic substituents indicated as R1 and R2. The different subpockets, in which portions of the ligand are accommodated, are represented in a schematic fashion.

In other studies[9], the different subpockets and their accessibility to water molecules were investigated. Furthermore, the selectivity and potency of already known scaffolds were optimized.

However, several studies have been carried out with small molecules of fragment size. One of the first studies was carried out in the early 1990s where solvent molecules (water, acetonitrile, acetone, isopropanol) were used as probe molecules to investigate their occupation in protein pockets[10]. This study uses the idea that small organic molecules can diffuse into protein crystals through water channels in the solid material. By placing the protein crystal in a solution containing organic molecules, they can diffuse through the channels into the protein pockets. This is the fundamental principle of soaking experiments. Later, also in a study by *Behnen et al.* small highly soluble probe molecules, such as aniline, urea, N-methyl urea, 2-bromoacetate, 1,2-propanediol, nitrous oxide, benzamidine, and phenol, were soaked into crystals of various proteins including Thermolysin to map their binding pockets and to detect hot spots of binding with respect to hydrophobic and hydrophilic properties[11].

The idea of the present study is to use a crystallographic FBLD approach to screen a general purpose 96-fragment library (see Part I) to map and investigate TLN pockets and find initial fragment hits for further development into lead compounds.

.

2. How to remove the Val-Lys dipeptide from the active site?

2.1 The use of Isopropanol

Performing an X-ray crystallographic fragment screening requires three preconditions: (i) Availability of a fragment library, (ii) preparation of protein crystals that support soaking experiments and (iii) diffract to resolution below 2.5 Å to detect binding of low affinity fragments.

Since an in-house fragment library is ready for use (see Part I), the first step was to find the optimal soaking condition that Thermolysin crystals can support.

Putative Thermolysin apo crystals have the obstacle that they inevitably host the Val-Lys autoproteolysis product occupying the catalytic pocket. In fact, during the rather concentrated crystallization conditions one molecule of TLN cleaves off the Val-Lys C-terminus of another molecule. Therefore, it was necessary to find a strategy to first displace this one-digit micromolar binding dipeptide and then expose TLN crystals to soaking experiments with the probe fragments. In previous studies it was not necessary to apply such a procedure as sufficiently potent ligands, capable of displacing the dipeptide and binding instead in the catalytic site were used. Fragments may not have the required affinity to displace this dipeptide on their own and therefore it becames necessary to establish a new soaking protocol.

It is known from literature that isopropanol (IPA) can be used as a small organic molecule to probe the Thermolysin binding site but also, as it is applied in very high concentration, it is capable of removing Val-Lys from the active site. A previous study of *A. C. English et al.*[12] showed the multiple binding of IPA molecules by increasing its concentration in soaking experiments. In particular, they used a multiple-solvent crystal structure determination (MSCS) approach to identify the position and the orientation of IPA in the TLN crystal structure. They soaked TLN crystals into 2, 5, 10, 80, 90, and 100 % isopropanol solution achieving a resolution better than 2.2 Å. Crystals remained in space group $P6_122$

but the cell dimensions shrank, such that in 100% isopropanol each cell dimension is reduced by approximately 1 Å compared to the IPA-free crystals, probably caused by an overall loss of water from the interstitial regions. Isopropanol variably affected the crystals by introducing fine cracks and in general led to an increase in mosaicity and decrease in maximum resolution at which data could be collected. However, it was demonstrated that 5 % of IPA was sufficient to remove the dipeptide from the active site and observe a single IPA molecule in the S1' pocket[12]. Nevertheless, IPA only fully occupies all the active site subpockets at very high concentration (Figure 2.1.).
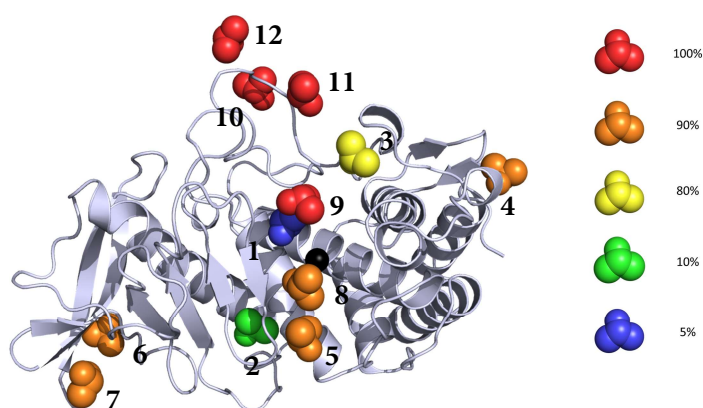
Figure 2.1. Cartoon representation of Thermolysin (PDB 8TLI) showing the binding sites of isopropanol. Molecules of isopropanol are shown in a sphere representation and are colored according to the order of binding with increasing isopropanol concentration, from blue to red (see legend on the right). The binding of isopropanol to TLN is additive. The IPA molecules are numbered according to the order in which they were observed to bind. The active site Zn^{2+} ion is shown as black sphere. Picture adapted from[12]

After the protein crystallization and before the soaking, it is therefore necessary to introduce an incubation step of TLN in isopropanol. After that, it is essential to wash the isopropanol off from the binding sites again. This is achieved by transferring the isopropanol to isopropanol-free buffer.

It is also important to consider the processes of incubation and soaking together. In fact, the conditions used in one process influence the other. In order to find the best conditions, two major screenings were performed. Particularly, the following aspects were taken into account:

1. Optimal isopropanol concentration in the incubation buffer for complete removal of the Val-Lys dipeptide;
2. Influence of IPA on mechanical integrity and diffraction power of TLN crystals;
3. Influence of incubation and soaking time on mechanical integrity and diffraction power;
4. Selection of buffers components to prevent severe osmotic shock effects on the crystals.

In the first screening, known conditions reported in the literature were used and then specifically adapted according to the crystals' capability to tolerate these. The crystals underwent three steps: incubation with IPA, soaking and cryoprotection prior to freezing in liquid nitrogen. In the second screening, the best condition found in the first one was further optimized, and the number of steps was reduced by incorporating cryoprotective agents in the soaking buffer.

The key idea is indeed to incubate the crystals with IPA to remove the autoproteolysis product and then transfer them into an isopropanol-free solution to wash the isopropanol out and ensure that it does not occupy the catalytic site and affect the binding of potential fragments.

The buffers composition used for the first and second screening are reported in Appendix 6.1.

2.2 The first optimization screening

The soaking conditions reported in the literature all include Tris·HCl pH 7.5 as buffering agent. *Krimmer et al.*[9] used 100 mM Tris·HCl pH 7.5 and 2 mM CaCl$_2$. In *Behnen et al.*[11] 30 % IPA and 100 mM Tris·HCl was used while in *English et al.*[12] 10 mM Tris·HCl pH 7.5, 3.2 mM calcium acetate and 7 % DMSO.

The buffer composition of *Krimmer et al.* was the starting point for finding optimal incubation and soaking conditions. More precisely for the incubation buffer 30 % IPA was added to remove the Val-Lys dipeptide from the catalytic site. The addition of isopropanol

was compensated by a slight increase of the buffer agent concentration. In fact, IPA (pKa=16.5) is slightly less acidic than water (pKa=15.7) and this could modify the buffering capacity of Tris. Therefore, 120 mM Tris·HCl was used instead of 100 mM. The first incubation (**Ia**, Figure 1.5) condition tested was composed of 120 mM Tris·HCl pH 7.5, 2 mM CaCl$_2$ + 30 % IPA. The same condition without IPA (**A**, Figure 1.5.) was used as soaking buffer.

Another strategy was to use a buffer with high PEG concentration (**B**, figure 1.5.) that was supposed to be suitable for soaking experiments against various protein crystals, as applied by Kan Fu. This buffer was used for soaking and cryoprotection whereas the incubation buffer consisted of **B** with 30 % of isopropanol.

The third strategy consisted of **Ia** as incubation and **B** as soaking buffer. In all three strategies after soaking the crystals were transferred for a few seconds in cryoprotection buffer (**C**, Figure 1.5.) prior to rapid freezing.

The first parameter under investigation was the incubation time. The crystals were placed in the incubation solution for 12, 24, 48 and 72 hours showing no substantial differences in diffraction quality and regular shape in comparison to untreated crystals. In all three strategies, incubation time was therefore not crucial for the crystal's survival.

However, crystals from different incubation buffers behaved differently in the soaking buffers, depending on composition and duration of incubation.
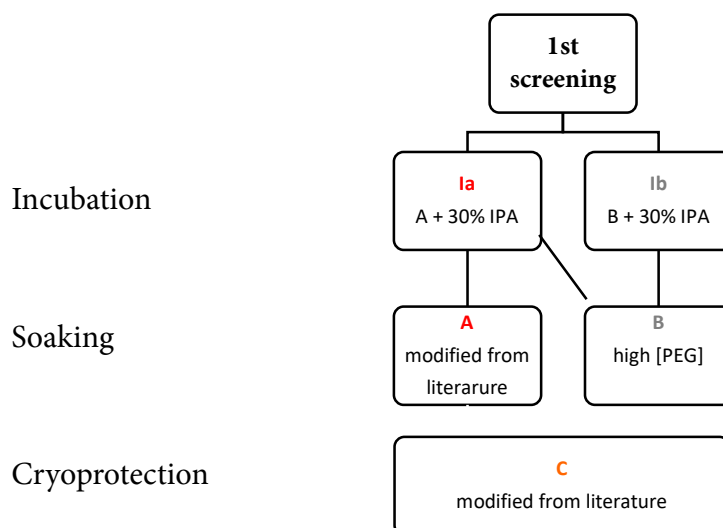
Figure 2.2. Schematic representation of the three strategies applied in the first optimization screening. They are summarized as follows: The first strategy is Ia-A-C: the second one is Ib-B-C and the final one is Ia-B-C.

Crystals incubated in buffer **Ia** and soaked in buffer **B**, broke after a few seconds on a very regular basis. In particular, they were fragmented into many slices along their entire length. It is likely that crystals were affected by osmotic shock caused by the sudden change in the solute concentration around the crystal, which resulted in a rapid modulation of the exposure and exchange of water across its surface. In fact, by transferring crystals from a PEG-free solution to another one with high concentration of PEGs, which makes the environment dense and viscous, could cause a loss of water from inside the crystal causing it to break into small slices.

To avoid osmotic shock, the very same composition of the incubation and soaking buffer was then used, apart from the addition of IPA in the incubation phase. Crystals incubated in **Ia** were then transferred into soaking buffer **A** while the ones incubated in **Ib** were moved into soaking buffer **B**.

However, the crystals in the high PEG condition did not survive in this case either. As soon as they were transferred from the solution with IPA to the one without, they immediately dissolved, first creating a halo around the surface of the crystal and then dissolving them almost completely or greatly reducing their size within a few tens of seconds. It was then

assumed that the conditions were too drastic and that the crystal needed to adapt slowly to the absence of IPA. From solution **Ib** with 30 % IPA the crystals were then transferred into a solution with 20 %, then 15 %, 10 % and finally 5 % IPA before being transferred into buffer **B**, i.e. in the soaking buffer with no IPA. Unfortunately, the crystals did not survive in this scenario either. As a control test, the crystals were transferred from the crystallization solution directly to the soaking solution **B**, thus avoiding the incubation step with IPA. Half of the crystals were still intact, but the resolution limit was too low (2.5 Å-2.6 Å) for the planned fragment screening project. Moreover, without IPA, the Val-Lys dipeptide would not have been removed from the catalytic site. It was therefore decided not to use such a condition at high PEG concentrations for further optimization.

The strategy of using the same buffer to avoid osmotic stress of crystals showed different outcomes in the case of transfer from **Ia** to **A**. In this case, the incubation time in IPA was crucial for the success of the soaking step. Crystals survived up to 72 hours of incubation but once transferred into the soaking buffer they showed a different behavior. Crystals incubated up to 24 hours survived in buffer **A** for about 1-2 hours. Those incubated for the longest time (24-72 hour) were broken or cracked after about one minute. The crystals were then incubated in **Ia** for 24 h, transferred to **A** for about an hour and subsequently immersed in the cryobuffer for a few seconds before freezing. These crystals showed X-ray diffraction although not yet ideal for detecting fragments. The reduced resolution (2.4 Å-2.5 Å) could result from the many transfers to which the crystals were subjected. In each transfer step, the crystals could have been increasingly damaged for reasons of mechanical stress. It is also important to note that the crystals immersed in IPA moved continuously in the solution drop which impeded their fishing and thus potentially increased the chance of inflicting damage on them.

Nevertheless, the last condition was the starting point for a second optimization screening.

2.3   The second optimization screening

In the second optimization screening, the best condition resulting from the previous one was further improved and also the number of steps was reduced from three to two by incorporating cryoprotective agents in the soaking buffer to minimize the chances of crystal cracking due to too many transfers. As a consequence, both osmotic and mechanical stress were reduced.
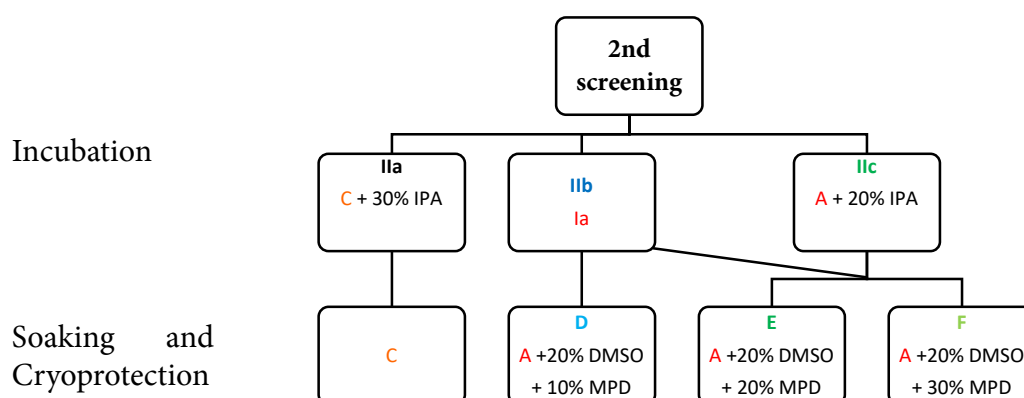
Figure 2.3. Schematic representation of the strategies applied in the second optimization screening. They most promising one is the IIc-E strategy which was used for the crystallographic screening of an inhouse 96 fragment library.

The first strategy was to use the cryoprotection buffer **C** as a soaking buffer. The addition of 30 % of IPA to **C** resulted in the incubation buffer **IIa**. However, transfer from **IIa** to **C** resulted in disintegration after 20 seconds.

In the second strategy, the same incubation buffer selected in the first screening was used again. In the soaking buffer 30 % of IPA, which is included in **Ia**, was replaced by an equal volume (30 %) of the cryoprotectants DMSO (20 %) and MPD (10 %). After an incubation period of 1-24 hours in **Ia** the crystals survived in the soaking buffer **D** for up to three hours. This was used for further optimization.

MPD is a cryoprotecting agent that is usually used at concentrations between 15-40%. Although the presence of 20% DMSO, 10% MPD may not be sufficient. Two further soaking buffers were tested, indicated as **E** and **F** in Figure 1.6. which contained 20 % and

30 % MPD respectively. The crystals were incubated in **Ia** for 1-24 hours and then transferred to **E** or **F**. Crystals extracted from the well containing buffer **E** showed a resolution of 2.2 Å while those from the well containing buffer **F** diffracted only until 2.6 Å.

Up to this point, only the soaking buffer was optimized while the incubation buffer always contained 30 % of IPA, as reported in the literature. Alternatively, the low resolution in the X-ray diffraction experiments can also results from the excessively high concentration of IPA in the incubation buffer. It is well known that isopropanol itself may damage the crystals, but is a lower concentration sufficient to fully displace the Val-Lys dipeptide?

Furthermore, the soaking solutions tested up to that point did not contain any fragments or ligands to prevent the ligands themself from interfering with the optimization process. Organic molecules can have deviating influence on protein crystals: Some stabilize the crystals; others make them more fragile and prone to cracking. The use of buffers alone made the optimization processes objective and easier to compare.

Phenol is a well-known fragment-like active site binder of TLN and was then used to both validate the most promising conditions (buffer **E**) and check if also 20 % of IPA were sufficient to displace the Val-Lys dipeptide from the active site.

Subsequent to growth, Thermolysin crystals were transferred to a solution consisting of **A** including 20 % IPA, (**IIc**, Figure 1.6.). This condition is identical to **Ia** except for the lower concentration of isopropanol. After an incubation period of about 16 hours (overnight), the crystals were then transferred to buffer **E** containing 5 % w/v phenol for about 2.5 hours. The good resolution of the dataset (1.2 Å) allowed to identify the phenol bound to the catalytic pocket in place of the Val-Lys dipeptide. In addition, no interfering IPA molecules were observed in the catalytic site. This demonstrates that even 20 % IPA were sufficient to remove the autoproteolysis product (Figure 2.4.).
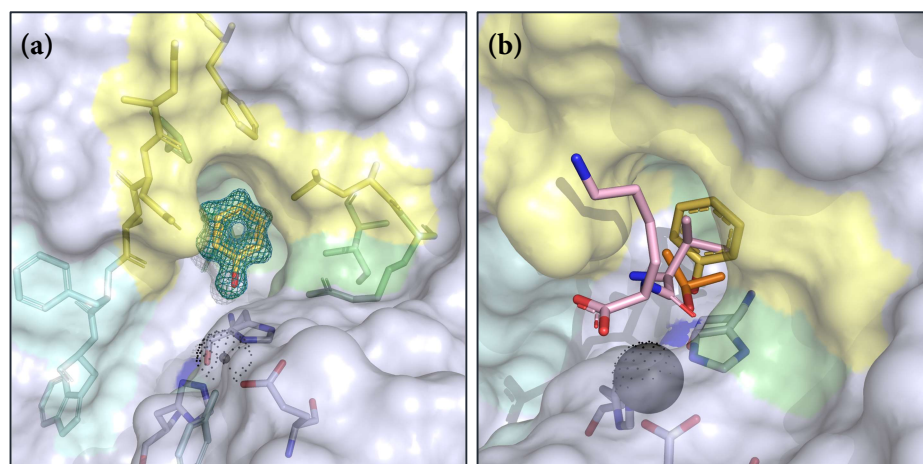
Figure 2.4.(a) Electron density map showing the phenol molecule bounded to the catalytic pocket. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh). **(b)** Superposition of Val-Lys dipeptide (pink sticks), IPA (orange sticks), and phenol (gold stick). The IPA and phenol molecules occupy the same position as the Val residue into the S1' pockets.

Since the last incubation and soaking protocol was validated and a good resolution had been achieved, it was subsequently applied to carry out a fragment screening project

.

3. X-ray crystallographic fragments screening

As reported in Section 1.2., the aim of the project was to perform a crystallographic fragment screening of an inhouse 96-fragments library, which was validated against several targets to demonstrate its suitability for general purposes (see Part I.). Thermolysin is among seven target proteins against which the 96 fragment library was screened in the Klebe group laboratory.

The entire library was successfully screened against TLN crystals and high quality datasets for each fragment were collected with resolutions between 1.2 Å and 2.5 Å and a completeness of 99-100% [Appendix_Crystal.Tables]. TLN crystals remained in the hexagonal space group P6$_1$22 (178) but the *c-axis* length was reduced by about 1 Å compared to the one usually reported in the literature[9]. This may be due to soaking experiments or, more likely, to the incubation step with isopropanol. As discussed earlier (Section 2.1.), crystals exposed to IPA do shrink in a way that each cell dimension is reduced by approximately 1 Å. Under the here applied conditions, however, only the *c-axis* was affected by IPA reducing its length from 130 Å to 129 Å.

The high-resolution datasets allowed the identification of seven fragments out of 96 in the electron density yielding a hit rate of 7 %. This hit-rate was further increased using the PanDDA software (Section 3.2.) with which five additional fragments could be identified. Overall, 12 hits were found, resulting in a hit rate of 12.5 % in total.

3.1   Binding mode of fragments hits

The initial seven fragments hits were identified in the difference electron density by using the automated refinement pipeline described by *Schiebel et al.*[13]. Out of these, four are active-site binders while the remaining three bind in remote cavities. In particular, **J13**, **J88**

and **J96** coordinate the catalytic Zn^{2+} ion with their oxygen atoms while the aliphatic moiety of **J22** and **J96** deeply penetrates into the S1' pocket, which usually hosts the valine

side chain of the Val-Lys dipeptide. In the other structures, the S1' pocket is occupied by solvent molecules such as TRIS (TLN-**J77**) or DMSO (TLN-**J13**, TLN-**J28** and TLN-**J62**). In the cavity immediately above and between the pocket S1' and S2' binds **J77**, a region usually occupied only by water molecules. The carboxylate moiety of **J13** and **J88** coordinate the Zn^{2+} ion while the respectively pyrazole and hydrazine moieties are oriented towards the mostly solvent-exposed S1 pocket. **J13** binds a second time in a small surface-exposed pocket next to Asn159. Two other fragments bind outside the catalytic site: **J28** lies on the surface while **J62** sits in the crystal packing contacting to adjacent Thermolysin molecules. Bound IPA molecules from the incubation step were found in only two structures, i.e. TLN-**J22** and TLN-**J13**. They are very far from the active site and in none of the 12 sites previously identified by *English et al.*[12].
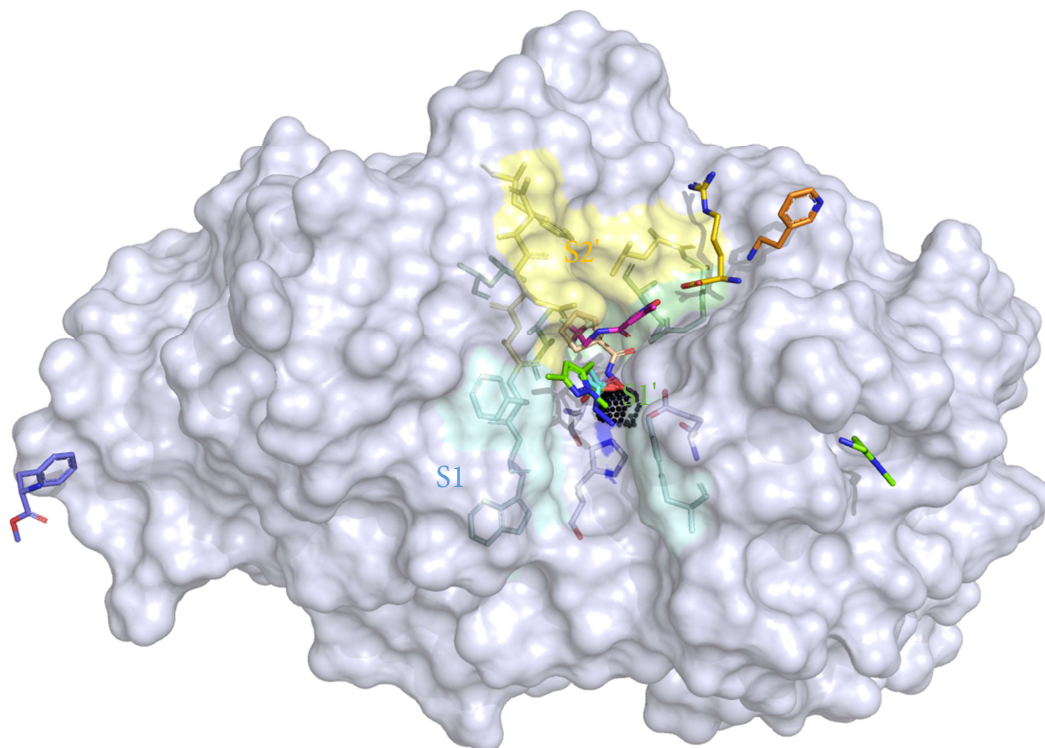
Figure 3.1. Surface representation of TLN with the seven fragments bounding the catalytic pocket and remote cavities. The fragments hits are represented as a stick and each of them is depicted with a different color. **J13** is in green, **J22** in magenta, **J28** in orange, **J62** in violet, **J77** in yellow, **J88** in light blue and **J96** in wheat.

3.1.1. Thermolysin in complex with the primary amine J13 (TLN-J13)

J13 binds to Thermolysin at two different sites: in a solvent-exposed surface pocket and in the catalytic site. This pocket is located below and immediately behind the S1 pocket (Figure 3.2.). The opening of this pocket is rather wide and consists of hydrophilic residues (Asn159, Glu160, Tyr217, Arg220, Thr222, Thr224 and Asn233) while in its inner part it becomes narrower, deeper and surrounded by hydrophobic residues as Ile236 and Phe281 (Figure 3.2.(b)).
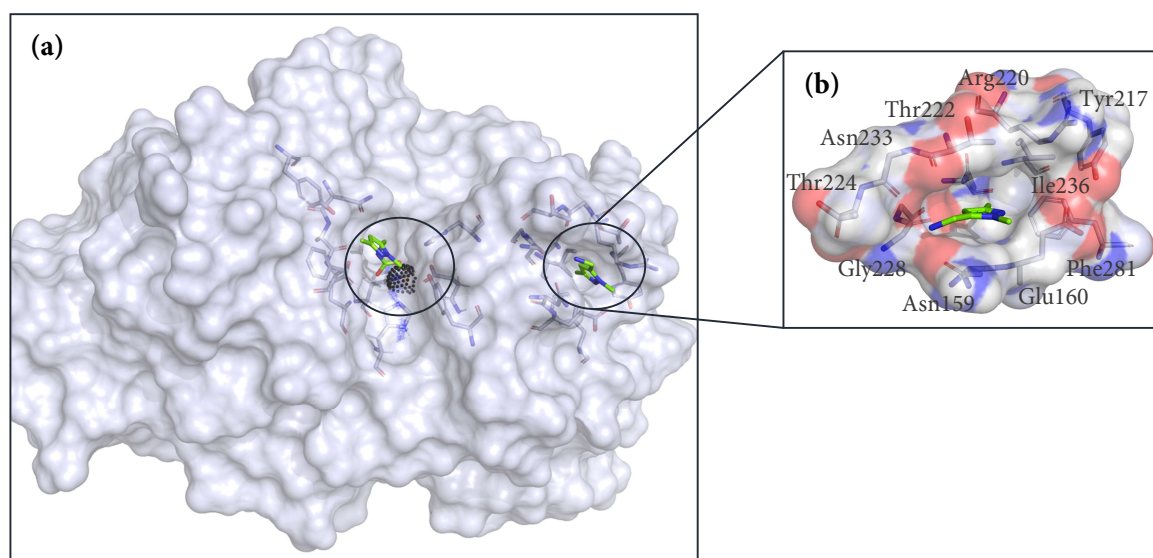
Figure 3.2.(a) Surface representation of TLN with the two binding sites encircled in blue. In the active site the first copy of **J13** coordinates the Zn^{2+}ion while the second **J13** accommodates in a remote cavity. The residues involved in the interaction are also represented as a stick model. **(b)** Close-up view of the remote cavity represented as a stick model and as a surface showing the polar entry pocket and the inner hydrophobic one.

The nitrogen atom of the primary amine is in H-bonding distance of the backbone oxygen atoms of Thr222 (3.5 Å) and Gly223 (3.5 Å) as well as of the oxygen atom OD1 of Asn159 (2.8 Å). The interaction of nitrogen atom N2 with Glu160 is mediated by two water molecules. W693 interacts with the nitrogen backbone (3.1 Å) while W540 with the oxygen atom OE1 (2.7 Å) (Figure 3.3.).
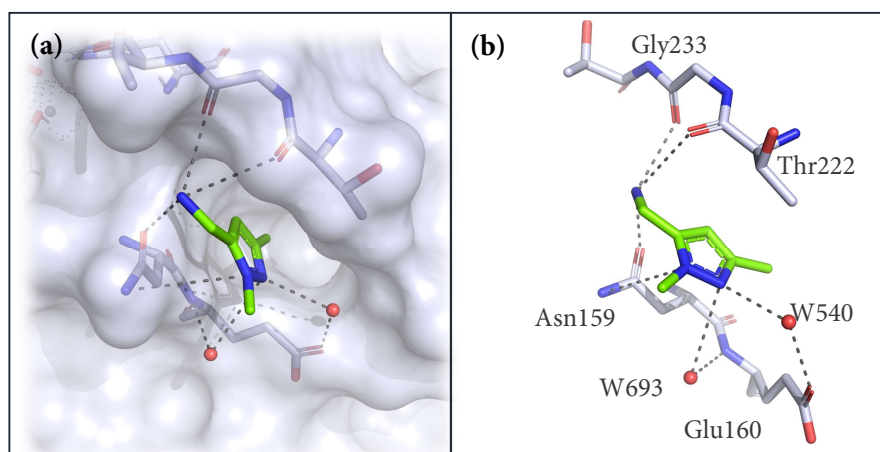
Figure 3.3. (a) Surface representation and **(b)**binding mode of **J13** depicted as a green stick.

In the active site, the primary amine seems to have reacted with some molecule present in the crystallization conditions. In fact, the 2mFo-DFc electron density map shows additional density both before and after building the whole fragment (Figure 3.4.(**a**) and (**b**)). Initially, two water molecules were built in the residual density. In the apo structure, the Zn^{2+} ion is usually pentacoordinated with the HEXXH motif (His 142, His 146 and Glu 166) and two water molecules. However, there is still some residual density that would suggest a missing atom linking the two water molecules (Figure 3.4.(**c**)).
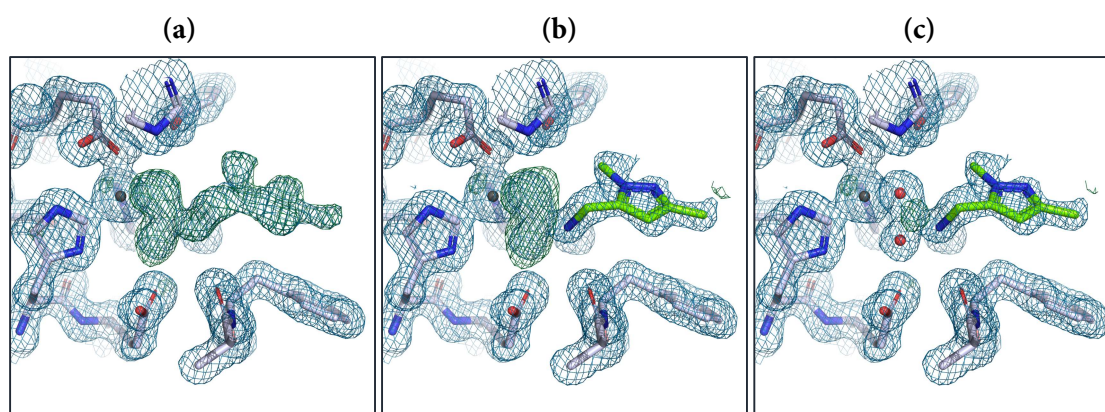
Figure 3.4. 2mFo-DFc electron density map before **(a)** and after **(b)** building **J13** in the residual density. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density) and at –3σ (red mesh, negative density). **(c)** Two water molecules (red crosses) were built but after refinement there is still positive residual density indicating a not yet assigned atom.

A CO_2 molecule, usually present in aqueous buffer and in equilibrium with the carbonate ion was built into the unexplained residual density which was then removed (Figure 3.5. (**a**)). As a result, the 2mFo-DFc density would justified placement and refinement of three missing atoms which cannot, however, be attributed to the CO_2 molecule alone as the density does not indicate a linear arrangement. In fact, the central carbon atom of the supposedly triagonal planar moiety and the **J13** nitrogen atom are in covalent bond distance and there is a continuous electron density that suggests either two molecules in two different however overlapping configuration with partial occupancy or the formation of a new covalent product. A bicarbonate ion (BCT) (Figure 3.5. (**b**)) and an acetate ion (ACT) (Figure 3.5. (**c**)), probably derived from sodium and calcium acetate present in concentrations of about 30 % in the lyophilized Thermolysin, were assigned to the density and compared. As indicated by the previous continuous density, both, the presence of an acetate and bicarbonate ion seem to coexist simultaneously with **J13** as they both refine to an occupancy of about 85 %. However, in the overlapping area between the nitrogen atom of the fragment and the bicarbonate hydroxyl function negative density is indicated (Figure 3.5. (**b**)) while the assigned acetate ion causes the primary amine to move out of the electron density (Figure 3.5. (**c**)). This indicates that only one of these groups can be built into the density with a covalent attachment to the fragment.

| (a) | (b) | (c) |
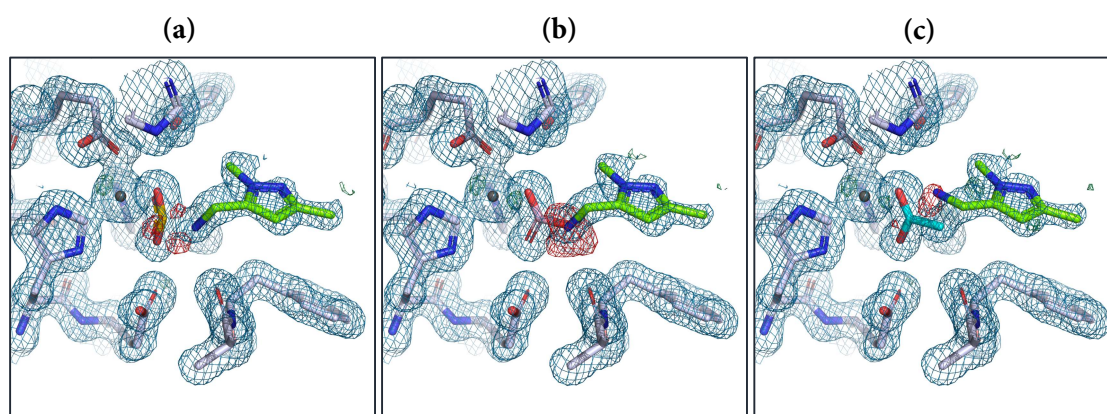
Figure 3.5. (a) mFo-DFc Electron density map after building a CO_2 molecule next to the Zn^{2+} ion. Comparison between BCT **(b)** and ACT **(c)** assigned to the electron density. Negative density **(b)** remains in the overlapping area between the nitrogen atom and the BCT hydroxyl function. ACT **(c)** causes the primary amine of **J13** to move out of the electron density. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density) and at –3σ (red mesh, negative density).

A reaction between the bicarbonate or acetate ion with the primary amine of **J13** was then assumed. This results in a carbamate and an amide derivative of J13, named in the following **J13a** and **J13b**, respectively (Figure 3.6.). After refinement, both molecules fit into the electron density with neither positive nor negative residual difference electron density (2mFo-DFc), with occupancies refine to 91 % and 85 % respectively. In terms of interactions and binding modes they are quite similar except for the unfavorable interaction between the acetate methyl group and the carbonyl backbone of Glu166 (red dots) which are too close whereas a typical hydrogen-bonding distance (2.6 Å) is suggested (Figure 3.6. (**b**)).
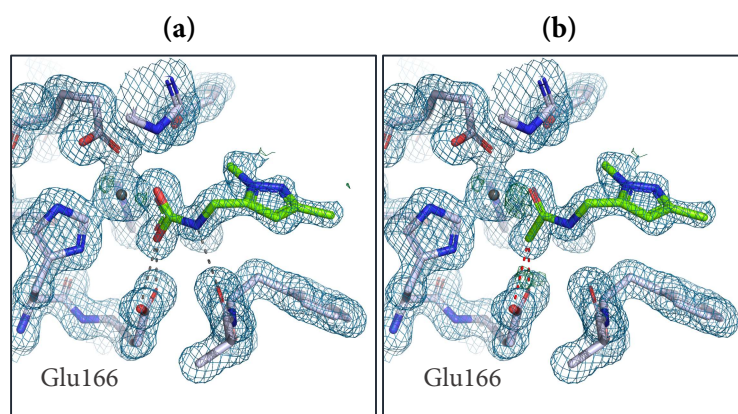
Figure 3.6. mFo-DFc Electron density map of **J13a (a)** and **J13b (b)**. The methyl group of ACT molecule is involved in an unfavorable interaction with carbonyl backbone of Glu166 (red dots), whereas a typical hydrogen-bonding distance (2.6 Å) is formed (grey dots) and not a C-O van der Waals distance. Favorable and unfavorable interactions are depicted as grey and red dots respectively. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density) and at –3σ (red mesh, negative density).

As it is difficult to differentiate **J13a** from **J13b** on the basis of the electron density alone, ESI-MS experiments were then carried out. The following aspects were then addressed:

(1)- Does **J13** starting material contain other species that can justify the observed density?

(2)- Is the 1M **J13** stock solution in DMSO pure?

(3)- Is the compound in the soaking buffer pure?

(4)- Is CO_2 reacting with the compound?

(5)- Does the Zn ion playing a role in **J13a** or **J13b** formation?

As indicated by the MS-ESI spectra the original substance is pure both in powder and in DMSO stock solution and there is no substantial difference even with or without presence of CO_2 in the soaking condition. To create a saturated CO_2 solution a small amount of dry ice was added to the soaking condition. In all cases, there are four dominant species: the peak at m/z of 126 refers to J13+1H while those at m/z of 109, 97 and 56 can be assigned to **J13** fragmentation as confirmed by the CFM-ID spectral prediction software (http://cfmid.wishartlab.com/predict). No peaks at 170 and 168 m/z referring to **J13a** and **J13b** +1H respectively, have been identified (Figure 3.7.).
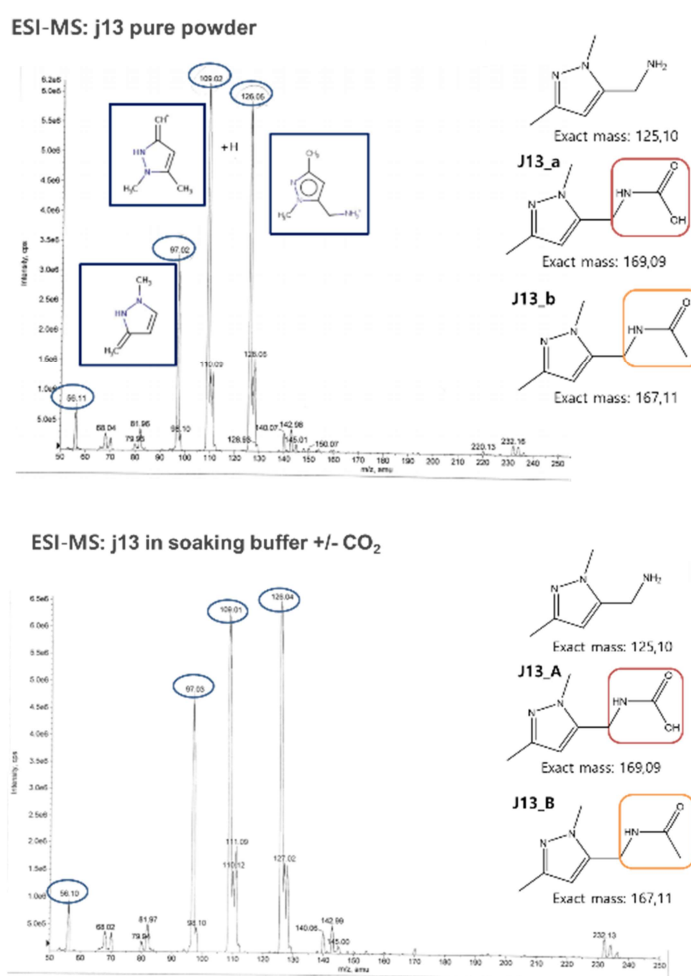
Figure 3.7. MS-ESI spectra of **J13** as pure starting material (top) and **J13** in the soaking solution with and without addition of CO_2. In the two spectra there are four dominant species: the peak at m/z of 126 refers to **J13**+1H while those at m/z of 109, 97 and 56 can be assigned to **J13** fragmentation. No peaks at 170 and 168 m/z referring to **J13a** and **J13b** +1H, respectively, have been identified.

As mentioned in Section 3.1.2. the Zn^{2+} ion in Thermolysin can easily coordinate a bicarbonate ion. Accordingly, it appears very reasonable to assign **J13a** instead of **J13b** to the electron density. No unfavorable interactions would then occur, and it is also likely to assume that **J13a** is the product of the back reaction catalyzed by Thermolysin. Although carbamates are not stable in aqueous solution due to their fast decarboxylation, it is possible that in crystals particularly under cryogenic conditions they are the most favored form. This may also be promoted by the Zn^{2+} ion located in the catalytic site and would explain why **J13a** was only found in the catalytic pocket and not in the soaking buffer as

described above. In addition, Zn^{2+} ion has a positive charge, which depletes the O-H bonds of a water molecule that is coordinated in the apo-structure. It creates a strongly polarized environment that would facilitate the nucleophilic attack of the primary amine onto the CO_2 dissolved in water. As reported by *Paul et al*[14], the electrophilic nature of the carbon atom in CO_2 makes it susceptible to nucleophilic attack by various nitrogen, oxygen, and carbon donors. Primary and secondary amines can act as a nucleophile (*Lewis* base) by direct attack on free CO_2 to form a zwitterion, which rapidly rearranges to the carbamic acid via intramolecular proton transfer. The resulting carbamate–bicarbonate equilibrium depends on many parameters, such as the amine's basicity, nucleophilicity, the local concentration, and the pH and solution temperature. The highest CO_2 loading capacity of primary amines was achieved at the lowest temperature studied and this could lead to the idea that cryogenic conditions encouraged carbamate formation as such derivatives are usually unstable at RT in aqueous solution. The following mechanism is therefore proposed for the reaction of CO_2 with primary (or secondary) amine in aqueous solution[14] (Scheme 3.1.).
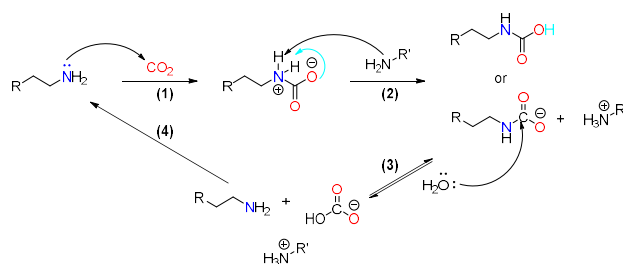
Scheme 3.1. Nucleophilic attack on free CO_2 with the formation of zwitterion/carbamic acid (**1**) followed by the deprotonation by another equivalent of basic amine to form a carbamic acid or a carbamate (2) which partially or completely hydrolysis into (bi)carbonate. Picture adapted from [14].

For the multiple arguments mentioned above, **J13a** was then built into the electron density. In all investigated structures, the Zn^{2+} ion atom is tetrahedrally coordinated with the HEXXH motif meaning that the NE2 atoms of His142 and His146 are involved in coordinative bonds of 2.0 Å and 2.1 Å, respectively, as well as OE2 of the Glu166 by 2.0 Å. In this specific case, the fourth atom coordinating the Zn^{2+} ion is the carbonyl oxygen atom of **J13a** through an 2.1 Å interaction. This oxygen is also involved in hydrogen bonds with the hydroxyl group of Tyr157 (3.4 Å) and with the N2 of His231 (2.9 Å). The O1 of **J13a** is within H-bonding distance both with the carbonyl group (2.5 Å) and the hydroxyl group (3.3 Å) of Glu143 side chain. Finally, the carbamide nitrogen is involved in a hydrogen bond with the backbone oxygen of Ala113 (Figure 3.8.).
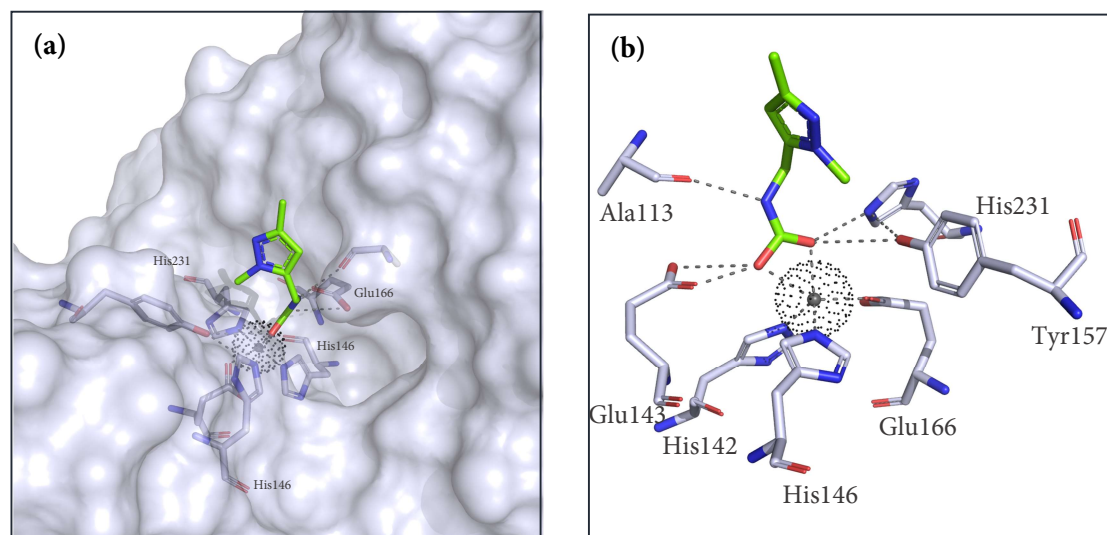
Figure 3.8. (a) Surface representation and **(b)** binding mode of **J13a** represented as a green stick model. The catalytic Zn^{2+} ion (black sphere) is tetrahedrally coordinated with the HEXXH motif (His142, His146 and Glu166) and the carbonyl oxygen atom of **J13a**. The catalytic Zn^{2+} ion is represented as a grey sphere while the van der Waals radius is depicted as black dots.

3.1.2.　Thermolysin in complex with J22 (TLN-J22)

The catalytically active Zn^{2+} ion is tetrahedrally coordinated by the HEXXH motif as described above (Section 3.1.2.). Here the fourth atom involved in the coordination of Zn^{2+} ion is given by the O1 oxygen atom of a bicarbonate (BCT) molecule that has to be deprotonated. O1 is also involved in an H-bond (2.8 Å) with the N2 atom of the side chain of His231. The O2 atom of BCT is in H-bonding distance with both, the OE1 (2.9 Å) and OE2 atoms (2.5 Å) of the Glu143 side chain. Finally, the BCT O3 atom interacts via hydrogen bonds with both the carbonyl backbone of Ala113 (3.2 Å) as well as with the N1 atom of **J22** (3.5 Å) (Figure 3.9.). The BCT binding mode is very similar to the carbamide head group of **J13a**.
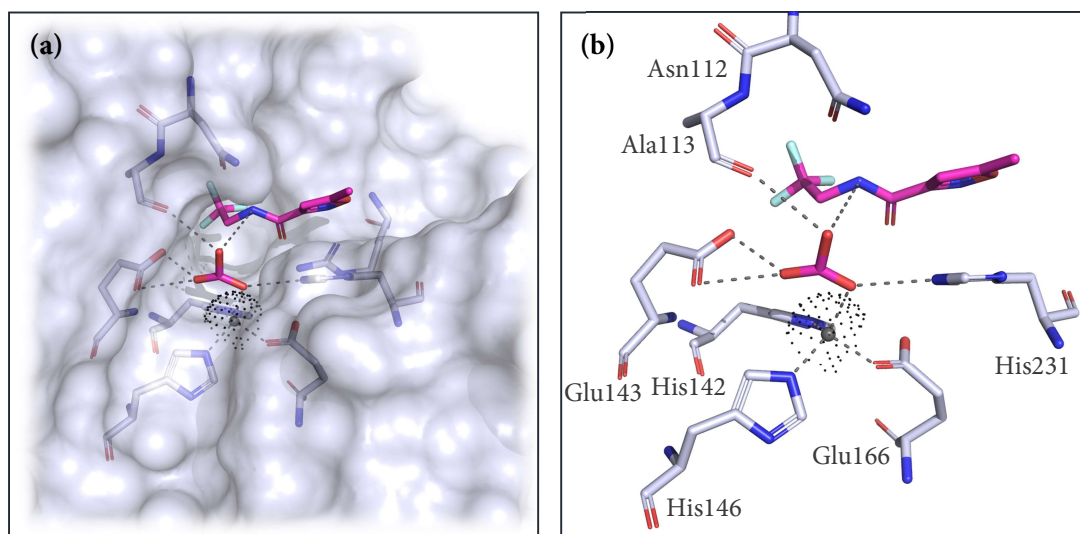
Figure 3.9. **(a)** Surface representation and **(b)** binding mode of BCT represented as a magenta stick model. The catalytic Zn^{2+} ion (black sphere) is tetrahedrally coordinated with the HEXXH motif (His142, His146 and Glu166) and with the O1 oxygen atom of BCT. The catalytic Zn^{2+} ion is represented as a grey sphere while the van der Waals radius is depicted as black dots. In the back, **J22** is represented as a magenta stick.

Alternatively to BCT, an acetate ion (ACT) was also placed in the electron density. However, the methyl group of ACT would be too close (3.2 Å) to the oxygen backbone of Ala113, causing a clash.as the observed distance is a typical hydrogen bond. In addition, the root mean squared Z score (RMSZ) value, which indicates the difference between the observed and assumed value and is expected to be between 0 and 1, for bond lengths is about 8 for ACT while zero for BCT. |Z|>2 for bond lengths is again very high (100%) for ACT and zero for BCT. Based on the favorable interactions described above and these statistics, BCT appears to be the better option and was therefore built into the electron density here as well as in other structures discussed later (Figure 3.9.).

As shown in Figure 3.1., **J22** binds into the catalytic site. It is interesting to observe that the trifluoromethyl group is located deeply in the S1' pocket, occupying the space that is usually occupied by the Val side chain of the Val-Lys dipeptide (Figure 3.10. **(a)**).
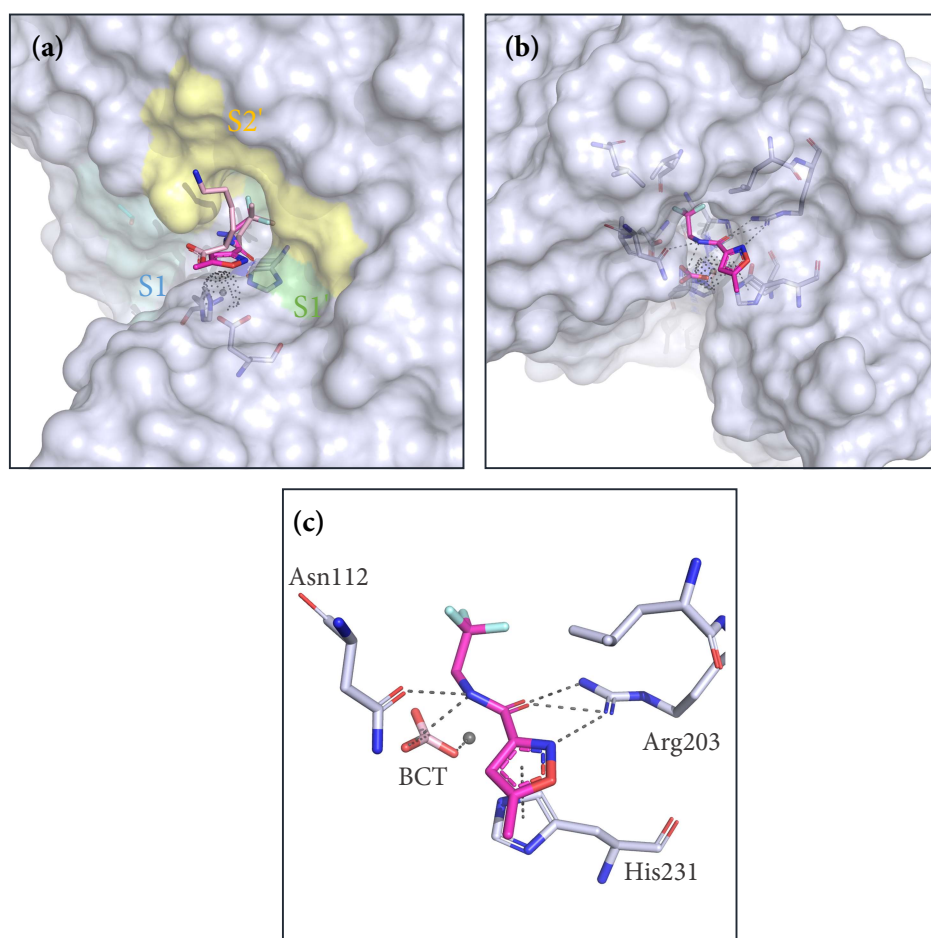
Figure 3.10. **(a)** The trifluoromethyl group of **J22** occupies the space that is usually occupied by the Val side chain of the Val-Lys dipeptide in the S1' pocket. **(b)** Surface representation and **(c)** binding mode of **J22** represented as a magenta stick model.

According to the convention assigned by *Cramer et al.*[14] the Asn112 residue is in "closed conformation" and interacts *via* a hydrogen bond (2.9 Å) with the nitrogen atom of the amide group of **J22**. The same atom also interacts with the hydroxyl function of BCT as described previously.

The amide oxygen atom, on the other hand, is within H-bonding distance (2.9 Å and 3.0 Å) of the terminal nitrogen atoms of the Arg203 side chain. The isoxazole ring interacts *via* π-π interaction with the imidazole group of His231, but it is also involved in an H-bond with the Arg203 side chain (2.9 Å) (Figure 3.10. (**c**)).

3.1.3. Thermolysin in complex with J28 (TLN-J28)

J28 sits on the surface and interacts with the protein through hydrogen bonds either directly or *via* water molecules. In both cases, it interacts with the side chain carboxylate group of the Asp213. In particular, the primary amine of **J28** interacts through hydrogen bonds with the oxygen atom OD1 (2.6 Å) and *via* water W575 with the oxygen atom OD2 (2.7 Å) of Asp213. **J28** also interacts with water W709 which is present in double conformation and also binds to water W575 (Figure 3.11.).
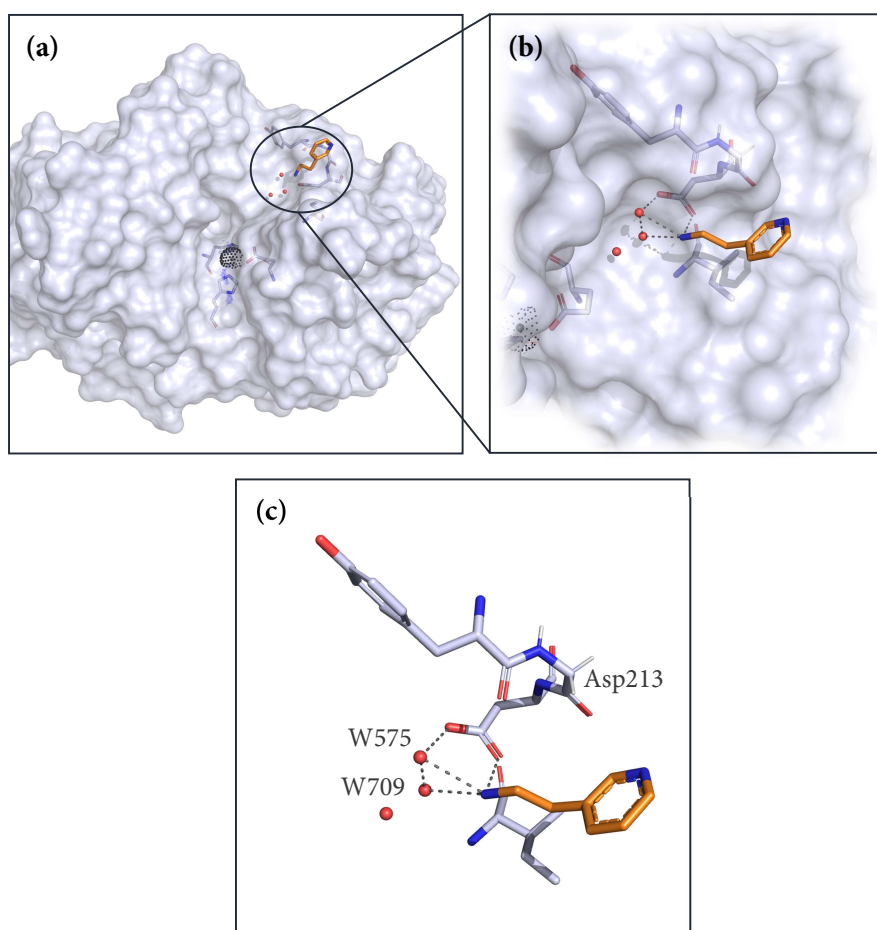
Figure 3.11. **(a)** Surface representation of TLN with the bound **J28** represented as an orange stick model. **(b)** Close-up view of the surface and **(c)** binding mode of **J28**.

3.1.4. Thermolysin in complex with J62 (TLN-J62)

J62 sits in the crystallographic packing interface and interacts with nearby symmetry mate (x-y, -y, -z) residues. In particular, the nitrogen atom of **J68** is involved in two hydrogen bonds, one of which with the backbone oxygen atom of Tyr211 (2.7 Å) and the other with the carbonyl oxygen atom of the Asp213 side chain (2.9 Å). The catalytic Zn^{2+} ion is coordinated by a bicarbonate ion in the same fashion as described in Section 3.1.2 (Figure 3.12.).
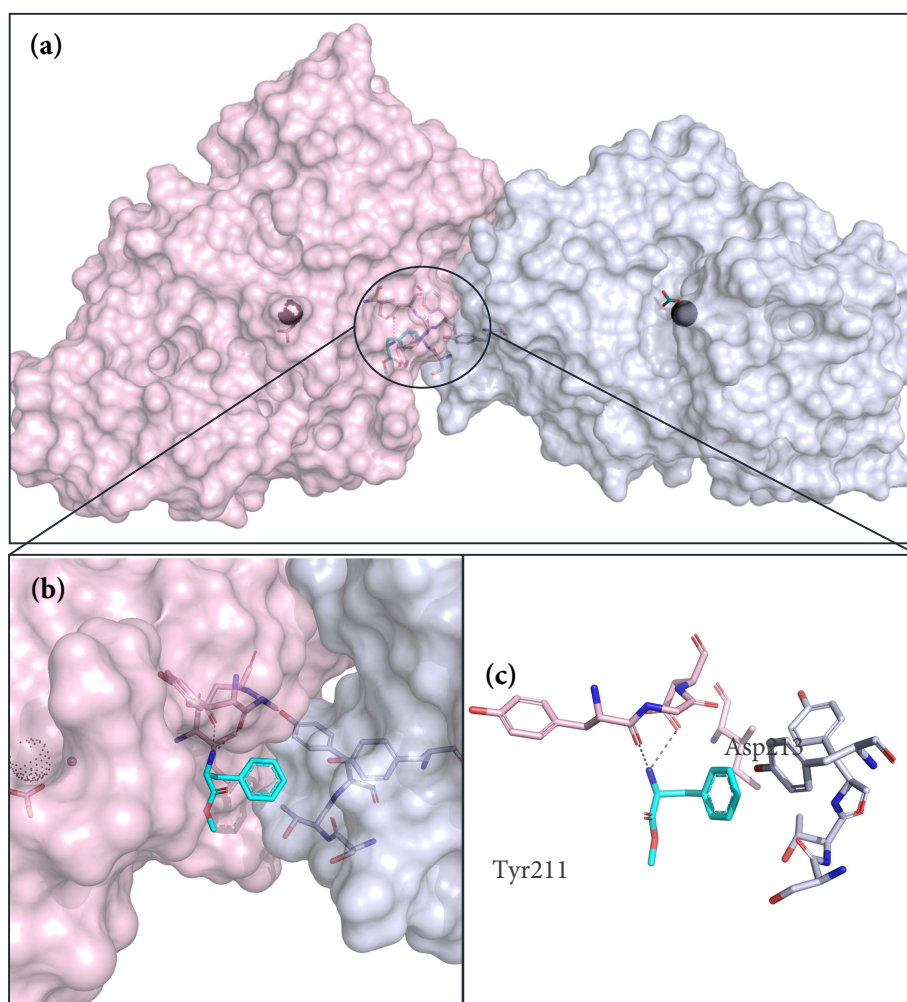
Figure 3.12. **(a)** Surface representation of TLN with the bounded **J62** sitting between the two crystallographic symmetry mates. The (x-y, -y, -z) mate is depicted in pink. **(b)** Close-up view of the surface and **(c)** binding mode of **J62**.

3.1.5. Thermolysin in complex with J77 (TLN-J77)

J77, D-Arginine, occupies a fairly large cavity located near the catalytic site slightly above and between the S1' and S2' pockets. The guanidinium moiety is involved in hydrogen bonds with the backbone carbonyl oxygen of Ser201 (2.7 Å) and Leu202 (3.5 Å). It interacts directly with the hydroxyl group of Tyr193 and indirectly through an interstitial water molecule (W567), thus establishing a hydrogen-bonding network which ranges from 3.1 Å to 3.5 Å. In addition, the nitrogen atom NE binds to the backbone oxygen atom of Leu202 (2.8 Å).
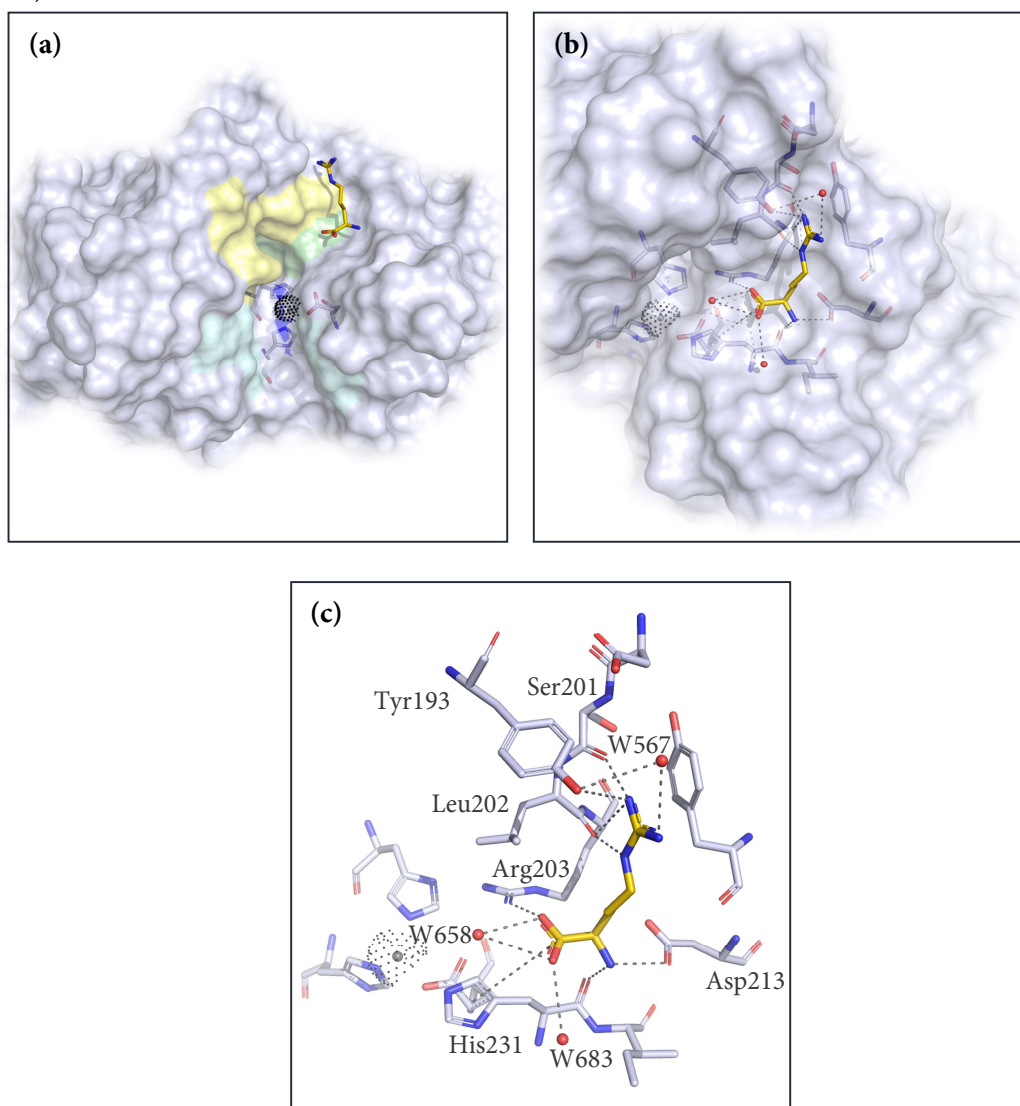
Figure 3.13. **(a)** **J77** accommodates near the catalytic site. **(b)** Surface representation and **(c)** binding mode of **J77** represented as a yellow stick model.

The amine nitrogen atom is within H-bonding distance of both the backbone carbonyl oxygen atom of His231 (3.0 Å) and the carbonyl oxygen atom of the Asp213 side chain (2.8 Å). The carboxylate function is involved in a hydrogen bonding network with waters molecules W658 (2.9 Å) and W683 (3.3 Å). In addition, the carboxylate group is most likely negatively charged and involved in a salt bridge with the positively charged Arg203 side chain (Figure 3.13.).

3.1.6.   Thermolysin in complex with J88 (TLN-J88)

J88 coordinates the Zn^{2+} ion *via* its carboxylate group in the same way as it was previously reported for **J13** (Section 3.1.1.) while the other portion is oriented towards the highly solvent exposed S1 pocket. In this pocket, the fragment creates some disorder in the main chain of the protein, in particular close to Tyr157. In consequence, the electron density was initially difficult to interpret. **J88** has a dimethyl amino terminal group which is not entirely visible in the electron density and would even clash with the Tyr157 side chain which adapts two alternative conformations. It would be reasonable to suppose that **J88** only binds when Tyr157 is present in conformation B. However, **J88** refine to an occupancy of 80% and this would not match the electron density that is suggested by a group occupancy refinement of the fragment and of the Tyr157 in the conformation B. In fact, as the occupancy of conformation A is the predominant one, this residue would clash with the fragment.

In addition, the DSX software (Material and Method, Section 6.3.2.) confirmed that the methylamine nitrogen is at an unfavorable distance to the positions of the atoms CZ and CE1 of the Tyr157 in conformation A. Possibly, the fragment binds first quite strongly to the Zn^{2+} ion *via* its carboxylic group, leaving the remaining part of the fragment very flexible. Nevertheless, the arrangement in the pocket, particularly an enhanced mobility of the terminal portion of the Tyr157, resulting in double conformation is induces and the refinement leads to the above-described occupancies (Figure 3.14.). This hypothesis is also confirmed by the significant increase in B-factors assigned along the molecule, which are

approximately 9 Å² next to the Zn²⁺ ion and reaches 25 Å² near the terminal dimethylamine nitrogen (Figure 3.15.).

(a)   **(b)**   **(c)**
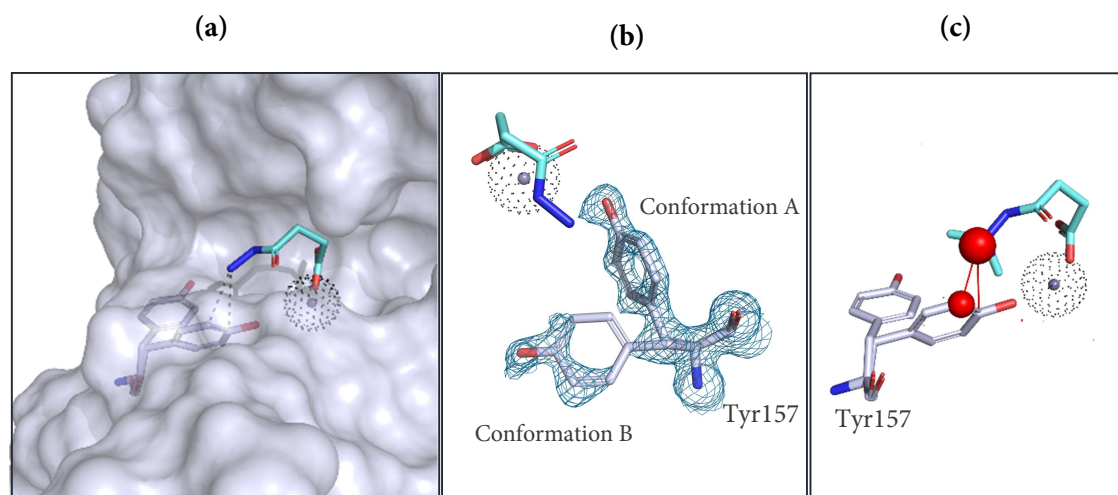
Figure 3.14. **(a)** Short distance interaction between **J88** and Tyr157 in conformation A. **(b)** 2mFo–DFc electron density map of Tyr157 in conformation A and B. 2mfo-DFc map is displayed at a contour level of 1σ (blue mesh). **(c)** Unfavorable interaction (red stick) and position (red spheres) of CZ and CE atoms of Tyr 157 side chain with **J88** calculated with DSX software.

The MOE software (Material and Method, Section 6.3.1.) was then used to perform an energy minimization of the fragment and Tyr157 under investigation. The energy minimization was performed both with and without the terminal methyl groups. After minimization, the carbonyl group and the two nitrogen atoms of **J88** moved upwards by 0.8, 0.6 and 0.4 Å respectively (Figure 3.15 (**c**)). This causes the fragment to move away from Tyr157. In fact, prior to the energy minimization the distances between the N1 atom of **J88** and the CZ and CE1 atoms of Tyr157 were 3.0 Å and 2.9 Å respectively. After the minimization, the distances increased to 3.6 Å and 3.3 Å respectively (Figure 3.15. (**a**) and (**b**)). This may indicate that the fragment's binding pose assigned by the refinement of the crystal structure is affected by the disorder and only leads locally to rather inaccurate values.
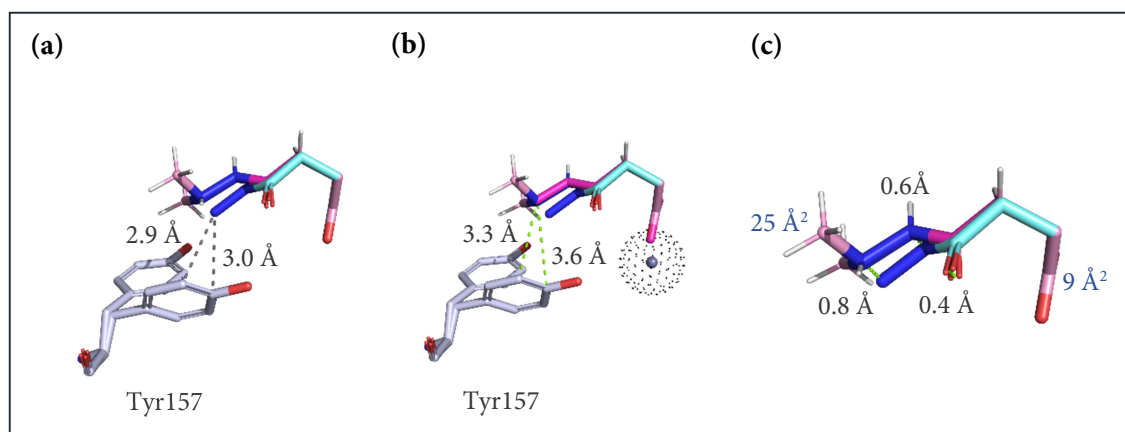
Figure 3.15. Distance between the CZ and CE atoms of Tyr 157 side chain with terminal nitrogen atom of **J88** before **(a)** and after **(b)** energy minimization with MOE. **(c)** After minimization, the carbonyl group and the N1 and N atoms of **J88** moved upwards by 0.8, 0.6 and 0.4 Å respectively. The B-factors increase along the molecule from 9 Å^2 next to the Zn^{2+} ion to 25 Å^2 near the terminal di-methylamine nitrogen atom.

Also as shown in Figure 3.17, **J88** is stabilized due the many interactions of the carboxylate while the hydrazide portion in more loosely bound. In particular, in addition to the coordination of the Zn^{2+} ion, J88's oxygen atom O1 is within hydrogen-bonding distance (2.8 Å) of His231's nitrogen N2 atom while oxygen atom O2 is involved in H-bonds (2.7 Å and 2.9 Å) with Glu143's carboxylate. The negatively charged carboxylate of the fragment is further stabilized *via* salt bridges to the adjacent His231, His142 and His146. The terminal carbonyl group of **J88** binds indirectly to the side chains of His231 and Asn112 *via* two water molecules (W585 and W710), thus forming a local hydrogen-bonding network. The water molecules W659 and W527 connect the N1 nitrogen atom with the carboxylate group and interacts with the OH group of Tyr157 in conformation B.
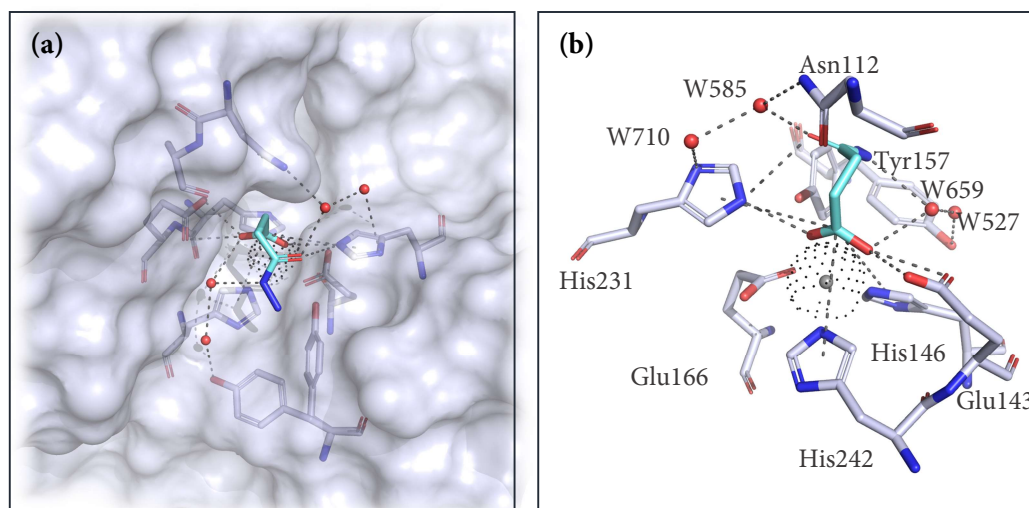
Figure 3.16. (a) Surface representation and **(b)**binding mode of **J88** depicted as a cyan stick model.

3.1.7. Thermolysin in complex with J96 (TLN-J96)

J96 deeply penetrates into the S1' pocket, which usually hosts the valine residue that is part of the Val-Lys dipeptide. The hydrophobic norbornane moiety of the ligand is indeed located exactly where the valine side chain of the Val-Lys dipeptide is usually placed and is involved in hydrophobic interactions with the residues of Val139, Glu143 and Leu202. The Zn^{2+} ion is tetrahedrally coordinated by HEXXH motif as described above (Section 3.1.1.) and the fourth atom, involved in the coordination, is contributed by the OH group of the fragment. This hydroxyl group is also within H-bonding distance of the N2 atom of His142 (3.3 Å) and His231 (2.7 Å), with the OED atoms of Glu143 (3.2 Å) and Glu166 (3.4 Å) as well as with the hydroxyl group of Tyr157 in conformation A (3.4 Å). The nitrogen atom of **J96** is involved in a direct hydrogen bond with the hydroxyl oxygen atom of Glu134 (2.7 Å) and indirectly mediated by a water molecule (W525) with the backbone oxygen of Ala113 (2.8 Å) and the carbonyl oxygen atom of the Arg112 side chain. The carbonyl oxygen atom of the ligand is within H-bonding distance of the nitrogen atoms of the Arg203 side chain (3.0 Å and 3.3 Å) as well as with the N2 nitrogen atom of His231 (3.0 Å) (Figure 3.17.).

Figure 3.17. (a) The hydrophobic norbornan scaffold of **J96** occupies the space that is usually accomodated by the Val side chain of the Val-Lys dipeptide in the S1' pocket. **(b)** Surface representation and **(c)** binding mode of **J96** represented as a wheat stick model.

3.2   PanDDA analysis enhanced hit rate

In a fragment screening project, it is not always ideal to use the same condition to screen all the fragments under investigation as promising candidates could be discarded. The fragments differ in affinity, solubility and potency and therefore some may take longer or required higher concentrations than others to diffuse into the crystals and reach the binding site of the target protein. In soaking experiments, the same fragment concentration and incubation time are usually used to screen an entire compound library. In the ideal case, each compound should indeed be treated according to its physicochemical properties and kinetic profile. However, this would lead to a tedious, time consuming and expensive fragment screening campaign.

It was therefore decided to screen the whole library applying the same conditions for all entries and to collect 96 datasets. They were then also inspected by means of the PanDDA (Pan-Dataset Density Analysis) method. PanDDA is a powerful approach to find fragments binding with low occupancy which are difficult to detect with the conventional refinement pipelines. More details of the present method are provided in the Materials and Methods section (Section 5.5).

All 96 datasets were successfully processed, and none of them was rejected. PanDDA analysis identified 26 interesting datasets where it appeared likely to find a bound fragment and the 33 event maps were distributed across seven protein sites (Figure 3.18).
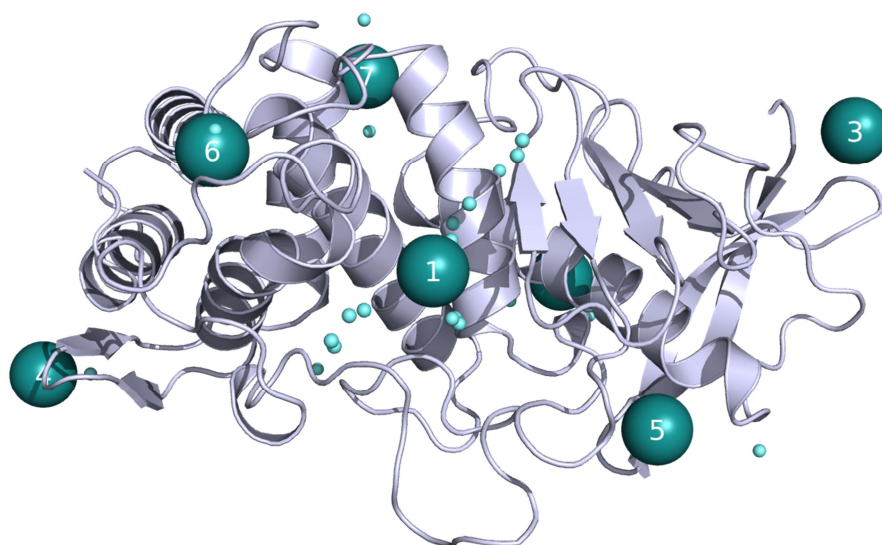
Figure 3.18. Cartoon representation of the Thermolysin secondary structure. The protein is divided into seven sites depicted as teal spheres and numbered in chronological order of identification. The small aquamarine spheres represent the position of the 26 plausible hits identified as interesting datasets by PanDDA analysis.

Among the 26 relevant datasets, seven corresponded to the hits already found by the conventional pipeline described in Section 3.1, while the remaining 19 were subjected to manual inspection resulting in the identification of five additionl hits in the event maps. Out of these, **J44** and **J35** bind in the active site between the S1' and S2' subpockets (Figure 3.18 and Figure 3.19).

In particular, the terminal di-methylcarboxamide moiety of **J44** penetrates into the S1' pocket, which usually hosts the valine side chain of the Val-Lys dipeptide while **J44** is largely accommodated on the surface. **J20**, **J89** and **J91** bind in remote cavities which correspond to sites 5, 3 and 2 respectively (Figure 3.19).
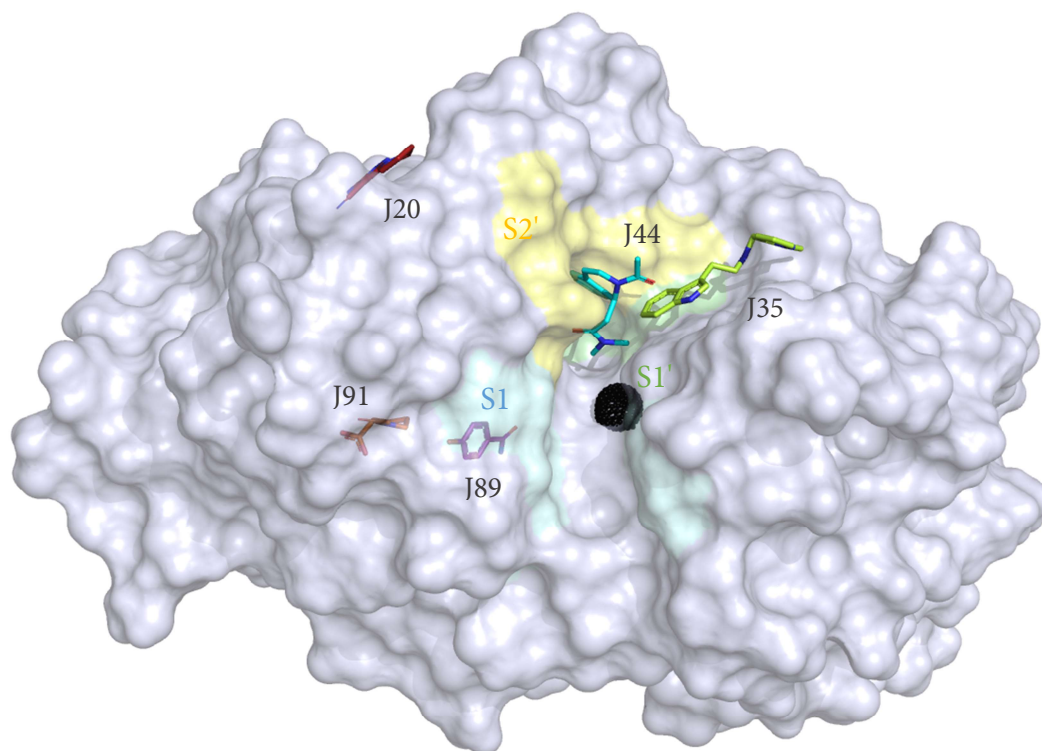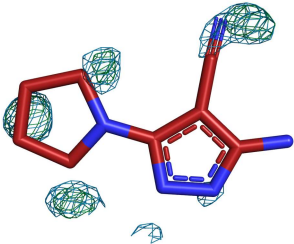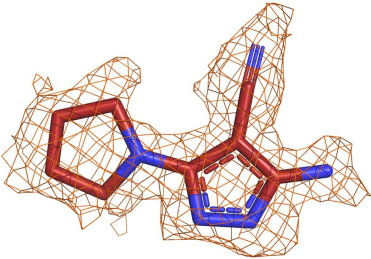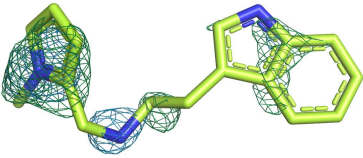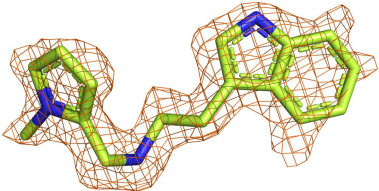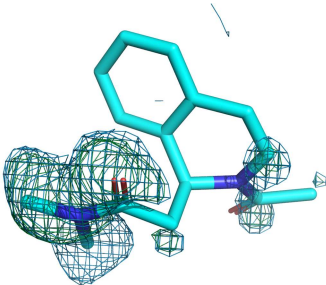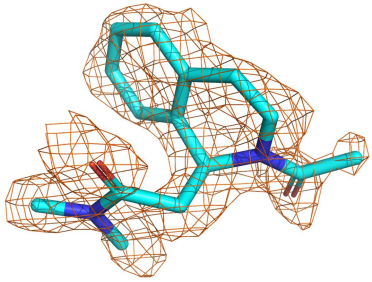
Figure 3.19. Surface representation of Thermolysin colored in whiteblue. The S1' pocket is depicted in green, the hydrophobic S2' pocket in yellow and the solvent exposed S1 pocket in cyan. The five PanDDA hits are displayed as stick models. In the active site, **J35** is depicted in lemon and **J44** in cyan. In remote sites, **J20** is colored in darkred, **J89** in magenta and **J91** in yellow.

As a result of the use of PanDDA, the hit rate in the fragment screening campaign was almost doubled, from 7 % to 12.5 %.

It is therefore always advisable, after manual inspection of the individual datasets, to search for additional hits with PanDDA, which is able to generate event maps where it is likely to find weakly bound fragments. However, they might be difficult to detect by the conventional difference maps. As shown in Figure 3.21, it would have been unreasonable to build the PanDDA suggested hits into the conventional maps since they appear insufficient to be properly interpreted. Instead they suggest a pattern more similar to water or other solvent molecules rather than to represent a contiguous density that speaks for fragments under investigation. On the other hand, the event map is clear and easy to interpret, and the fragment can then be assigned with confidence.

	Conventional maps	**PanDDA event maps**
J20 Resolution= 1.40 Å Occupancy= 0.48 RMSD = 0.106 Å σ_i = 0.137 BDC = 0.89		
J35 Resolution= 1.70 Å Occupancy= 0.72 RMSD = 0.132 Å σ_i = 0.191 BDC = 0.74		
J44 Resolution= 1.40 Å Occupancy= 0.62 RMSD = 0.100 Å σ_i = 0.174 BDC = 0.81		
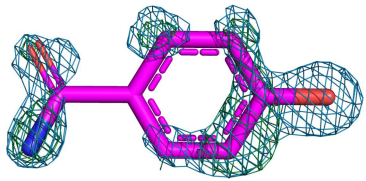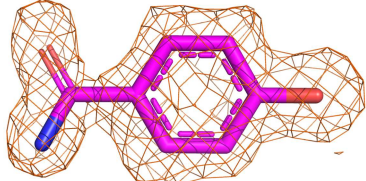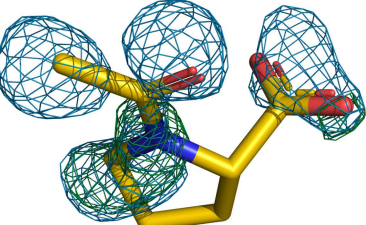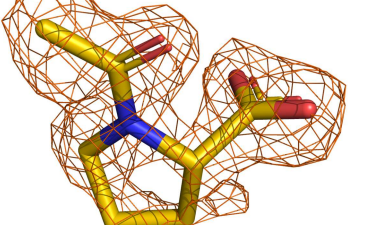
J89		
Resolution= 1.38 Å Occupancy= 0.66 RMSD = 0.128 Å σ_i = 0.147 BDC = 0.83		
J91		
Resolution= 1.54 Å Occupancy= 0.57 RMSD = 0.092 Å σ_i = 0.191 BDC = 0.83		

Figure 3.20. Comparison between conventional electron density maps and PanDDA event maps. PanDDA hits are represented as a stick model and colored according to Figure 3.20. Conventional maps derived from the final step of the autorefine pipeline described by *Schiebel et al.*, after removing hydrogen atoms and water molecules. In the central column, conventional 2mFo-DFc maps are contoured at 1 σ and represented as a blue mesh while mFo-DFc maps are contoured at 3 σ and represented as green mesh. Conventional maps did not allow to identify the fragment in the electron density. In the right column, PanDDA event maps are contoured at 2 σ and represented as orange mesh. In the left column, the fragment ID is marked in bold and statistical information such as resolution, occupancy, RMSD, σ_i and BDC of PanDDA event map are given. RMSD: root-mean-squared deviation of the atomic coordinates, it compares the positions of all atoms built into event density with respect to their positions in the selected reference apo-structure and it should be below 1 Å; σ_i : map uncertainty for the dataset *i*. It indicates the deviation of the dataset *i* with respect to the mean map which is also known as background; BDC: background density correction factor. The best signal for the event map calculation is at BDC = 0.80.

However, considering the different nature of conventional electron density maps and the PanDDA event map, this method was only used as a tool to identify fragments which justify further investigation. Subsequent steps could include prolongation of the incubation or exposure time and/or increase of the fragment concentration, in order to obtain a new dataset which verifies that the fragment is detectable also by conventional maps. This

approach avoids the risk of starting a hit-to-lead optimization with fragments which are very weakly bound to the protein target or which have only minor occupancy.

Nevertheless, PanDDA is an excellent tool to find potential hits that would have been completely overlooked by manual inspection of conventional maps.

3.2.1 Thermolysin in complex with J20 (TLN-J20)

J20 binds in a solvent-exposed surface pocket where it is mostly involved in hydrophobic interaction with surrounding residues such as Val13, Pro69, Phe130, Ile131 and Pro132. The primary amine of **J20** is involved in a weak hydrogen bond with the backbone oxygen atom of Ser65(3.0 Å) (Figure 3.21.).
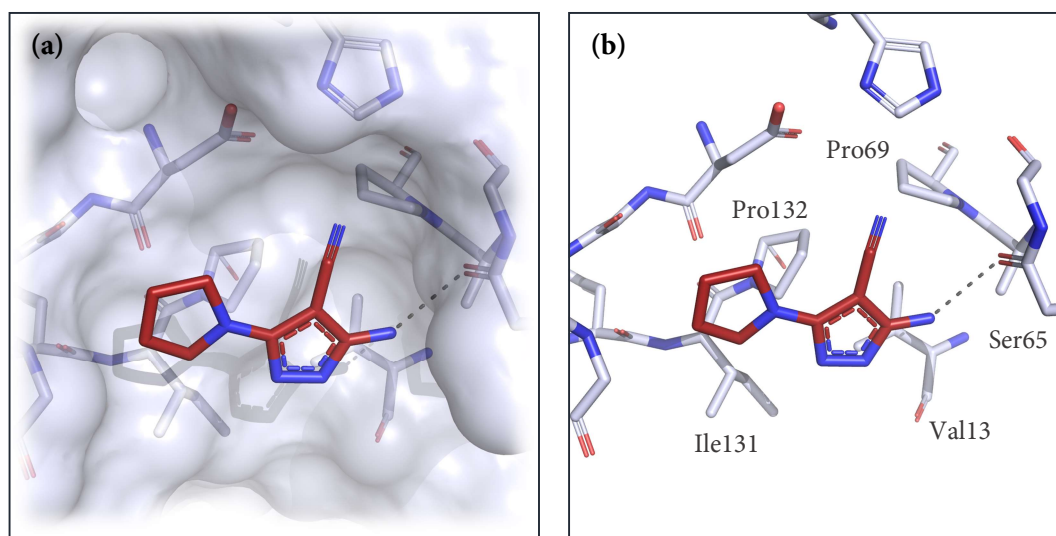
Figure 3.21 (**a**) Surface representation and (**b**) binding mode of **J20** represented as a dark red stick model.

3.2.2 Thermolysin in complex with J35 (TLN-J35)

J35 sits on the surface between the S1' and S2' sub-pockets. The aliphatic amine of the fragment is within hydrogen bonding distance to the oxygen atoms of the Asp213 side chain (2.8 Å and 3.5 Å). On the other side of the molecule, the nitrogen atom of the indole moiety is involved in a H-bond with the carbonyl oxygen atom of the Asp226 side chain (3.5 Å) while the aromatic six-membered ring interacts via π-π stacking with the

imidazole moiety of His231 in a fashion similar to the interaction between the side chain

of His231 and **J22** (Figure 3.22, for **J22** see Figure 3.10.).
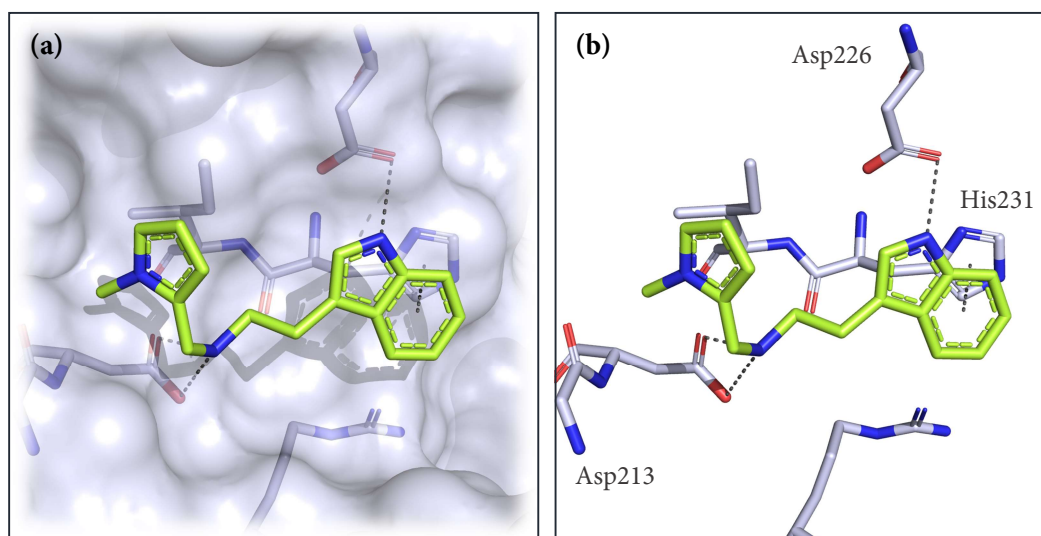
Figure 3.22.(**a**) Surface representation and (**b**) binding mode of **J35** represented as a lime stick model.

3.2.3   Thermolysin in complex with J44 (TLN-J44)

Similarly to **J22** and **J96**, the dihydroisoquinoline moiety of **J44** deeply penetrates into

the S1' pocket which usually hosts the valine side chain of the Val-Lys dipeptide. Here, it

is involved in hydrophobic interactions with the Phe130, Val139, Ile188 and Leu202 side

chain. The oxygen atom of Asn112 side chain is in "close conformation" and interacts *via*

hydrogen bonding with the dimethylacetamide moiety of the fragment (3.1 Å). On the

opposing side of the molecule, the terminlacetyl group is at H-bonding distance with the

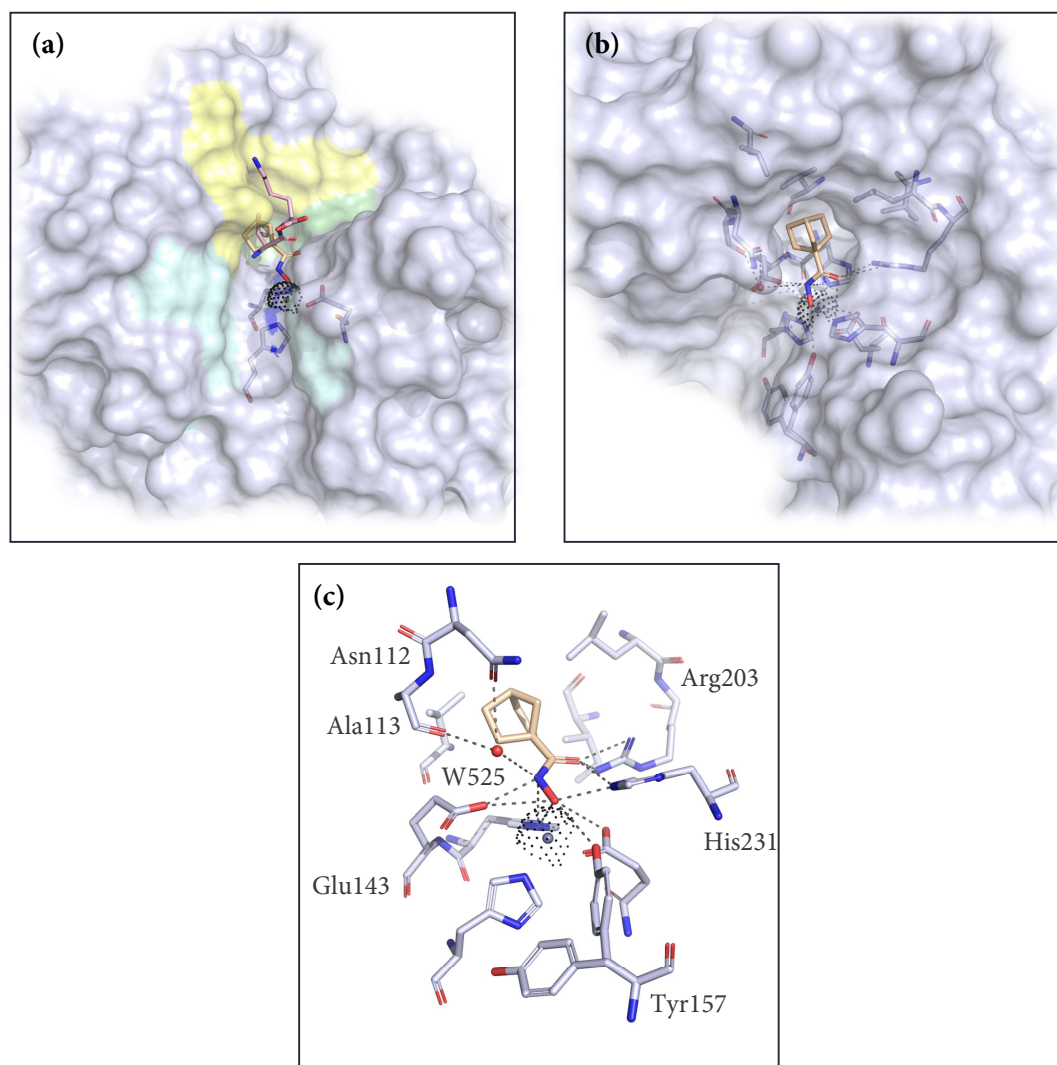water molecule W124 (3.1 Å) (Figure 3.23.).
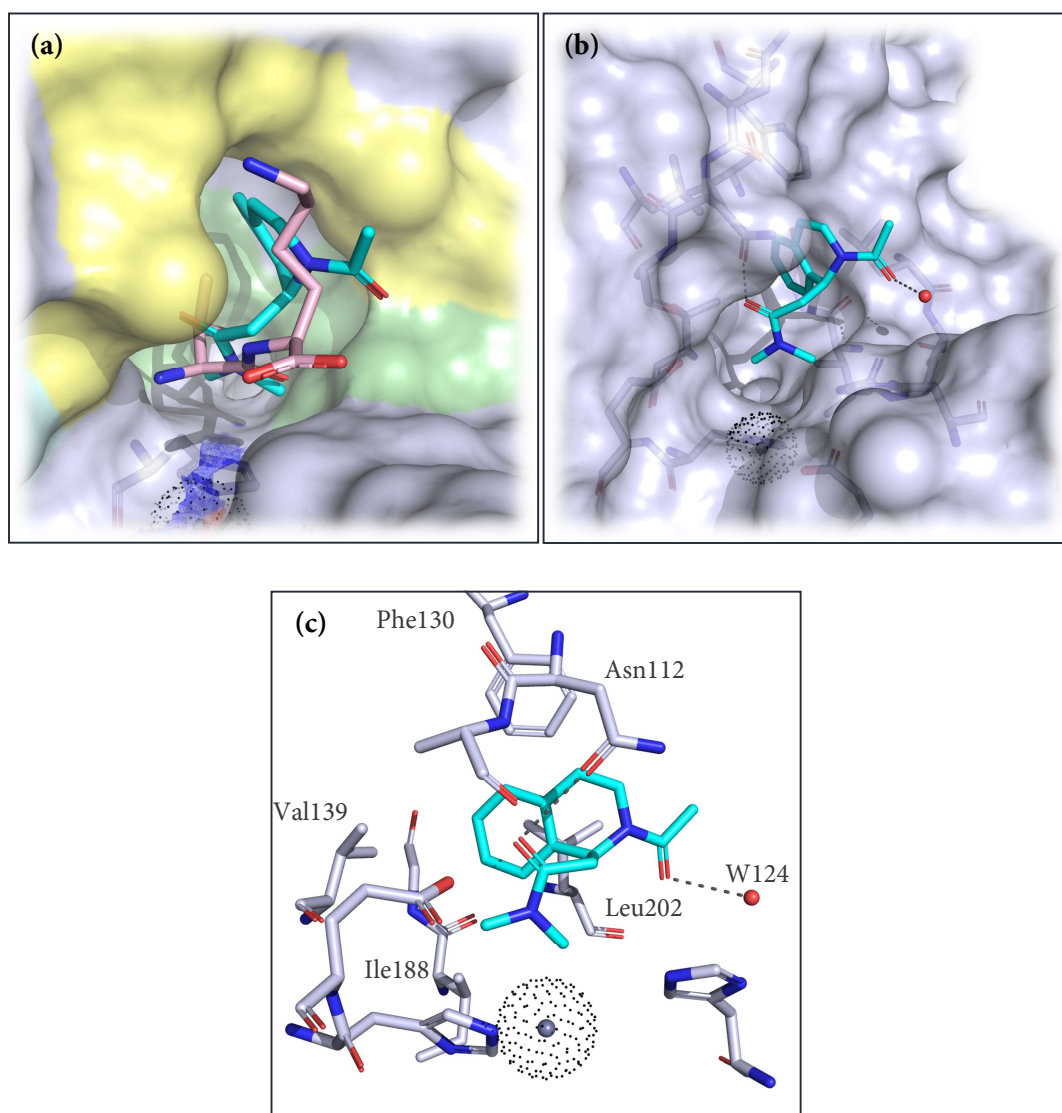
Figure 3.23. (**a**) The hydrophobic dihydroisoquinolin scaffold of **J44** occupies the space that is usually accommodated by the Val side chain of the Val-Lys dipeptide in the S1' pocket. (**b**) Surface representation and (**c**) binding mode of **J44** represented as a cyan stick model.

3.2.4   Thermolysin in complex with J89 (TLN-J89)

J89 sits on the surface where it is involved in two H-bonds. In particular, one carboxylate oxygen of the Asp82 side chain interacts with the phenolic OH group of the fragment (2.6 Å) while the carbonyl backbone of Tyr179 interacts with the oxygen atom of the amide group of **J89** (3.0 Å). The phenol ring of Tyr179 side chain is also involved in a π-π interaction with the central aromatic ring of the fragment (Figure 3.24.).
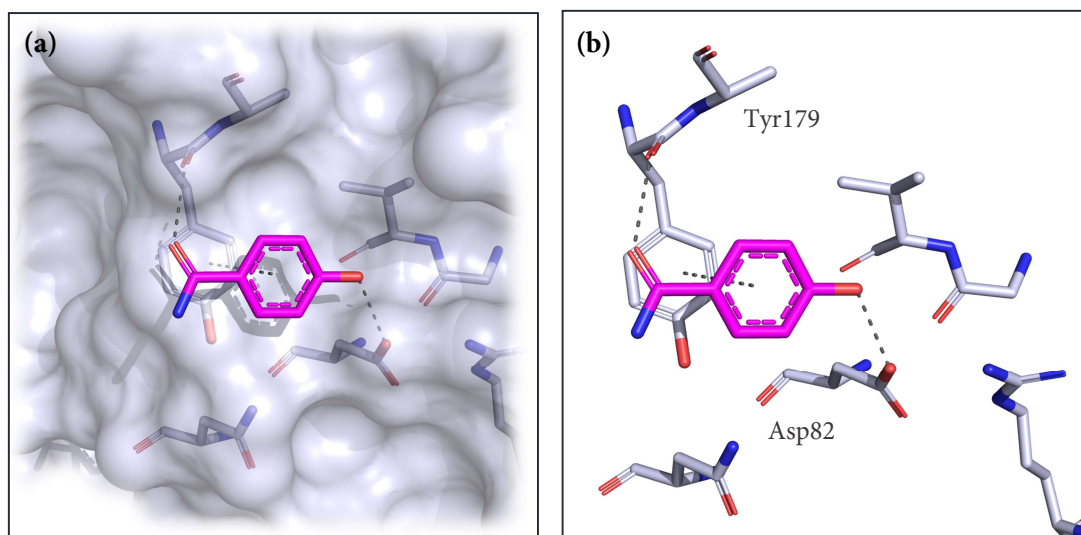
Figure 3.24.(a) Surface representation and **(b)** binding mode of **J89** represented as a magenta stick model.

3.2.5   Thermolysin in complex with J91 (TLN-J91)

J91, L-proline, accommodates in a solvent-exposed surface pocket which is located on the back of the catalytic site. One carboxylate oxygen of the fragment is within H-bonding distance with the guanidinium moiety of the Arg35 side chain (2.7 Å and 3.4 Å) while the second acidic oxygen atom of **J91** is involved in an H-bond with the carbonyl group of the Asn21 side chain (3.5 Å). Furthermore, the primary amine of Arg35 side chain interacts with the nitrogen atom of L-proline ring (3.2 Å) (Figure 3.25).
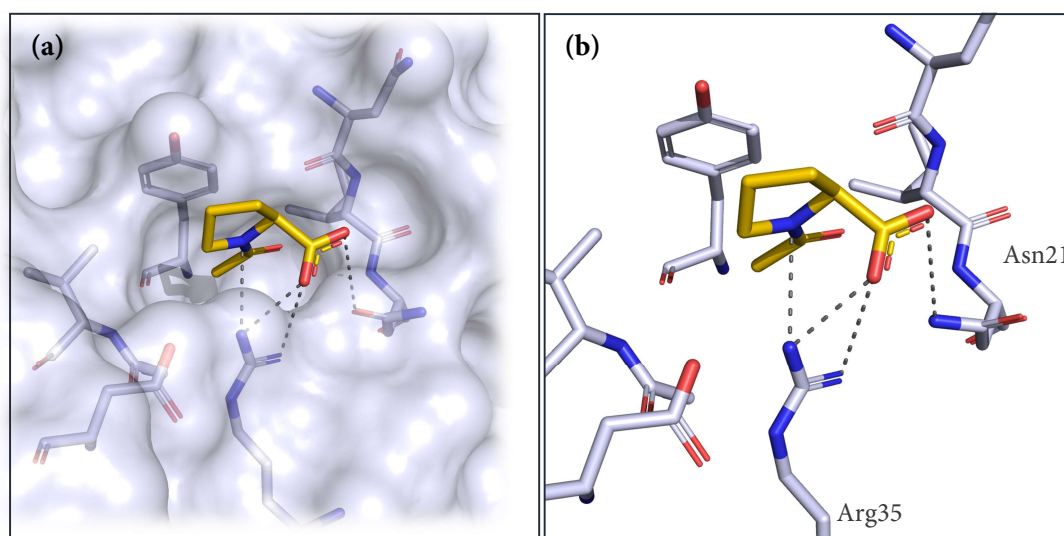
Figure 3.25. (a) Surface representation and **(b)** binding mode of **J91** represented as a gold stick model.

4. Conclusion and Outlook

In the present work, a new soaking protocol suited for a crystallographic screening of an in-house 96 fragment library was established. Unlike the phosphonamidate molecular scaffold, typical for TLN inhibitors, fragments usually have insufficient affinity to displace on their own the autoproteolytically produced Val-Lys dipeptide which occupies the TLN active site. After crystal growth, Thermolysin crystals were therefore transferred to a solution consisting of 20 % IPA and incubated overnight to displace the bound dipeptide. Crystals were then incubated in IPA-free soaking solution for about three hours which contained 100 mM fragment concentration in the final drop. The entire library was successfully screened against TLN crystals and high-quality datasets with each exposed fragment were collected allowing identification of seven fragments out of 96 in the electron density yielding a hit rate of 7 %. The hit-rate was further augmented using the PanDDA software as a tool to identify weakly bound fragments which might require further investigation. PanDDA analysis allowed five additional fragments to be identified in the event maps. Thus, 12 hits were found, resulting in a hit rate of 12,5 % in total.

It is interesting to note that **J22**, **J44** and **J96** occupy the catalytic site and penetrate deeply into the S1' pocket, which usually accommodates the valine side chain of the Val-Lys dipeptide. They have different scaffolds than those commonly found for Thermolysin inhibitors. For exemple, **J96** coordinates the Zn^{2+} ion and could be a promising candidate for the design of novel lead compounds characterized by innovation and chemical diversity.

The fragment screening campaign has also revealed new remote pockets as the small one near the catalytic site occupied by **J13**. It is still unknown whether addressing this pocket by a ligand plays a role in modulating, potentiating or inhibiting the catalytic activity of the enzyme. However, it may be the starting point for a future study.

Comparing and analyzing the different binding modes, it is interesting to note that PanDDA hits (Section 3.2.) establish an average less interactions with TLN than the

fragments found using the autorefine pipeline (Section 3.1.). This is probably due to their weak interactions and their lower occupancy. Possibly, more care is needed while evolving these hits in a follow-up lead development campaign.

Although thermolysin is not directly involved in any human disease, it is becoming increasingly relevant in studies of Alzheimer's disease. In particular, recent studies conducted by *Leite J.P. et al.*[6] revealed the identification of the clearance mechanism of $A\beta_{1-40}$ by Neprilysin (NEP). Although it is the most extensively studied peptidase associated with $A\beta$ clearance, no crystallographic data are available for the NEP-$A\beta$ peptides complex. Instead, the authors used TLN since it shares the same $A\beta$ peptide recognition mechanism with NEP to elucidate its cleavage site suggesting that the C-terminal product could inhibit the enzymatic activity due to its slow dissociation from the catalytic site. These studies have highlighted the importance of considering model systems such as TLN in the study of serious diseases such as Alzheimer's. With first insights how to inhibit NEP, a fragment screening campaign could be a promising strategy for the discovery and design of promising candidates against this disease as well as to elucidate further mechanisms of action.

5. Material and Methods

5.1. Crystal preparation, incubation in IPA and soaking

Thermolysin was purchased from Calbiochem (EMD Biosciences) as lyophilized powder with a purity of 62.9 %. For crystallization experiments, no further purification steps are required. TLN crystals were obtained using the sitting drop vapor diffusion method at 19 °C. TLN powder was suspended in 50 µL of pure DMSO and mixed with the same volume of crystallization buffer, composed of 3.8 M CsCl and 100 mM Tris·HCl pH 7.5, to obtain a 4 mM protein solution. The solution was further mixed by pipetting gently up and down. After centrifugation for 3 minutes at 13000 rpm at RT, 1 µL of clear yellowish supernatant was transferred into each well of a crystallization plate (24-well XRL Plate from Molecular Dimension) and the reservoir wells were filled with 1 mL of demineralized water. The crystallization plates were properly sealed using a clear tape (1.88 inch wide Crystal Clear Sealing Tape from Hampton Research).

TLN crystals appeared after 24 hours and stopped growing after five days. The mature TLN crystals were used for soaking experiments. As reported in Section 2.3., TLN crystals (max. six crystals per well) were incubated in a solution containing 120 mM Tris·HCl pH 7.50, 2 mM $CaCl_2$ and 20 % isopropanol for 3-24 hours. The crystals were then transferred into a new well (max. three crystals per well) consisting of 4.5 µL of soaking buffer (120 mM Tris·HCl pH 7.50, 2 mM $CaCl_2$, 20 % DMSO and 20 % MPD) and 0.5 µL of 1 M fragment stock solution in DMSO resulting in a 100 mM concentration in the final drop. Depending on the fragment under examination, the soaking time ranged from a few minutes to a maximum of 3 hours. Crystals were then fished and directly flash-frozen in liquid nitrogen.

5.2. Data collection and processing

Diffraction data for **J13**, **J22** and **J62** were collected on beamline XRD1 at ELETTRA Synchrotron Light Source in Basovizza (Trieste, Italy). Datasets were collected on a Dectris

Pilatus 2M pixel detector at a wavelength of 1.00000 Å at 100 K. Diffraction data for **J28**, **J77**, **J88** and **J96** were collected on beamline BL14.1 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin. Datasets were collected on a Dectris Pilatus 6M pixel detector at a wavelength of 0.91841 Å at 100 K[15]. XDS and XDSAPP were used for indexing, integration and scaling of the datasets[16][17].

5.3. Structure determination and refinement

The identification of possible hits was first performed using the autorefine pipeline described by *Schiebel et al.* [13]. Whenever regions of interest were identified, the refinement was subsequently carried out manually.

The structures were solved by molecular replacement using Phaser[18]. Model refinement (xyz coordinates, individual *B*-factors, occupancies) was done with Phenix.refine using the version 1.15.2-3472 [18] and model building into electron density maps (2mFo–DFc and mFo–DFc) with Coot (0.8.9)[19][20].

A randomly chosen subset of 5 % of the reflections was excluded from the refinement and used for the calculation of $R_{(free)}$. As a first refinement step, Cartesian simulated annealing was performed (default settings). *B*-factors for all model atoms (except for hydrogen atoms) were refined anisotropically for resolutions < 1.42 Å, TLS for resolution 1.43 Å < 1.60 Å and isotropically for resolution < 1.60 Å. Hydrogen atoms (riding model) were added to the amino acids with Phenix.refine. Alternative conformations of amino acid side chains and ligand moieties were assigned if an occupancy of at least 20 % was obtained after refinement. The program *Chemoinfo* was used for SMILES code generation (www.cheminfo.org/flavor/malaria/Utilities/SMILES_generator_checker.html), ligand molecules and restraints were created with the Grade Web Server (http://grade.globalphasing.org).

5.4. Accession Code

Atomic coordinates and experimental details for the described crystal structures were deposited with the PDB and will be released upon publication under the accession codes

6SBK (TLN-**J13**), 6SC0 (TLN-**J22**), 6SB9 (TLN-**J28**), 6SC3 (TLN-**J62**), 6SCK (TLN-**J77**), 6SCU (TLN-**J88**) and 6SC1 (TLN-**J96**).

5.5.   PanDDA method

PanDDA (Pan-Dataset Density Analysis) method was developed by *Pearce, N. M. et al.* [21] and aims at finding weekly bound fragments not observable in conventional electron density map. It was extensively used in crystallographic fragment screening campaign.

The input to a PanDDA analysis is a series of refined crystallographic datasets of the same crystal system and in the best case the same solvent and buffer components. Each dataset consists of a refined structure from the eleventh step of the autorefinement pipeline described in *Schiebel at al.*[13] and associated diffraction data including 2mFo–DFc structure factors. The only difference between the datasets is the presence of different fragments soaked into the protein crystal.

It is important to note that the conventional electron density maps (2mFo–DFc and mFo-DFc) describe the contents of the unit cells averaged over the whole crystal lattice and not the contents of a single unit cell (online dictionary of crystallography). Furthermore, small molecules often bind to only a fraction of the proteins in the crystal lattice. It results that the content of the individual unit cells may be different from each other and in each of them minor deviating events may occur (such as small conformational changes or weakly bound fragments and small molecules from the buffer solution) which are difficult to detect considering the average alone.

Furthermore, in a crystal structure a major and minor populated state my coexist. Conventional maps represent the major state very well but usually obscure minor events that are visible only in the minor populated state. In crystals where a state of interest (usually bound ligand) remains at sub-unitary occupancy, the total electron density in that region is an average of this bound state and the complementary ground state, e.g., the state of the uncomplxed protein. PanDDA analysis is able to accentuate the minor state after

background correction. 30 or more datasets independently measured and belonging to the same crystal system are needed to obtain an accurate representation of the ground state density (also called background or unbonded state).

The input protein structures, and their maps are first aligned and then analyzed to calculate an averaged ground-state map for subtraction. Using multiple maps allows a Z-score measure to be calculated at each point in every data set, that reflects how significantly the data set deviates from the ensemble of data sets at that point. Since the Z-score is small in the individual ground state density (meaning no substantial deviation from the mean ground state) but it deviates significantly from the mean map in the bound state map, it allows the identification of regions of interest in the individual data set maps where it is likely to find the bound ligand. Once an interesting region has been identified in a particular dataset by a high Z-score value, the observed electron density at the site is a superposition of the bound and unbound states of the crystal. When the bound state (also called changed state) of interest contains a weakly bound ligand which is present only in the minor state of the crystal, the signal is usually obscured by the superimposed ground state which is indeed the major state of the crystal. Subtracting from the individual dataset the mean ground state density, suitably weighted by an appropriate background density correction factor (BDC), it removes the crystallographic superposition between major unbounded and minor bounded state and creates a partial difference map called an event map, which is usually fully interpretable (eq. 1). The event map approximates the electron density for the bound state corresponding to the ligand-bound fraction of the unit cells in the crystal [21][22].

[event map] = [dataset map] – BDC X [ground state map]   (eq. 1)
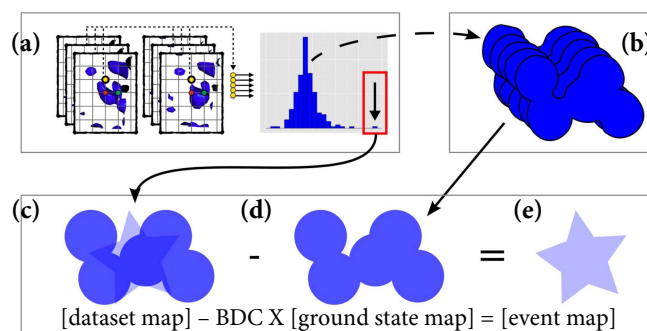
Figure 4.1. (a) Structure alignment (left) and identification of background (Gaussian curve) and event (arrow) that deviate from the mean ground state and where it is likely to find a bound ligand. (b) Calculation of average ground-state density based on the multiple data sets. (c) Identification of region of interest in which the density is a superposition of changed-state (star in the background) and ground-state (in blue) densities; the changed state is obscured by the superimposed ground state. (d) Subtracting the main ground state calculated in (b) to (c), the event map is calculated. It is an estimation of the changed state. Image adapted from [21].

The optimal Background Density Correction factor (BDC) is essential for extracting the best signal. It is determined algorithmically and objectively by maximising the difference between the local correlation and the global correlation of a background-corrected map to the ground state. In other words, BDC maximizes the contrast between the ground state map and the change state map resulting in an event map density that approximate only the changed configuration of the site, including protein backbone and side-chain conformations as well as a ligand binding event.
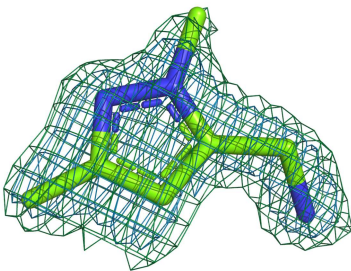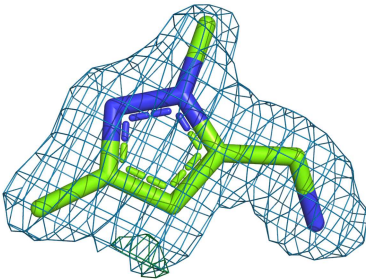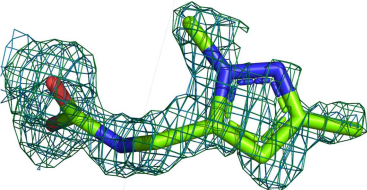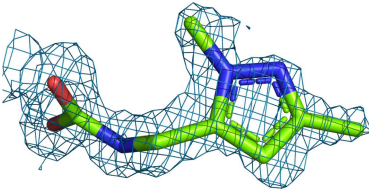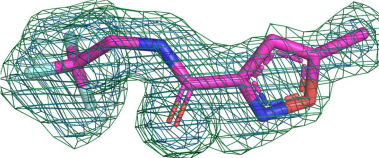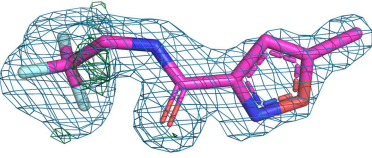
It has been calculated that the optimal value of BDC is 0.80. At this value, the optimal signal for the event map is generated. BDC = 0.0 corresponds to the observed density (2mFo–DFc maps) while BDC=1.0 to the difference mFo–DFc map[21].
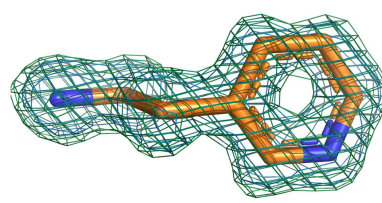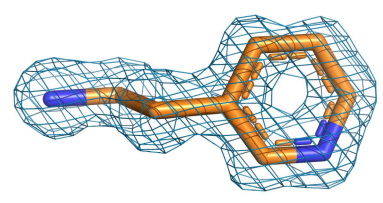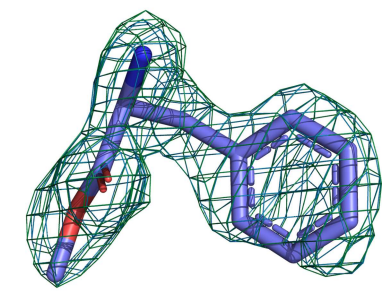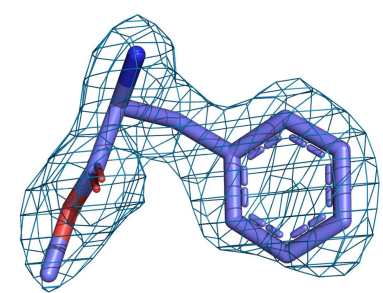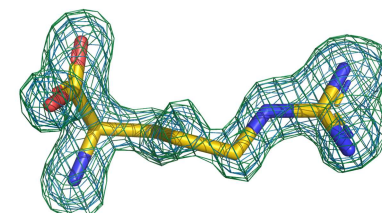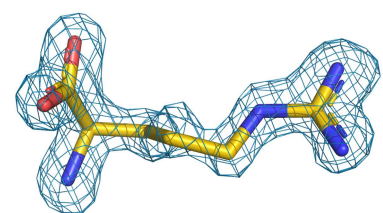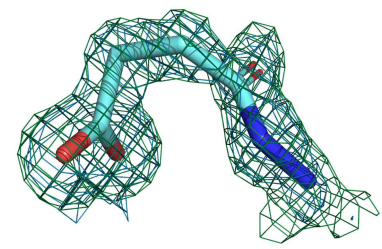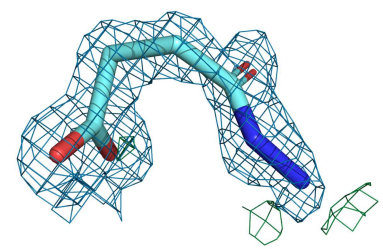
6. Appendix

6.1. Buffers composition used in the optimization screenings

First optimization screening		Second optimization screening	
Ia	30 % IPA 2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5	**IIa**	30 % IPA 10 mM Ca $(CH_3COO)_2$ 10 mM Tris·HCl pH 7.5 5 % DMSO 20 % MPD
Ib	30 % IPA 35% PEG 3350 25 % PEG 400 300 m M NaCl 400 mM Tris pH 7.5	**IIb**	30 % IPA 2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5
A	2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5	**IIc**	20 % IPA 2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5
B	35% PEG 3350 25 % PEG 400 300 m M NaCl 400 mM Tris pH 7.5	**D**	2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5 20 % DMSO 10 % MPD
C	10 mM Ca $(CH_3COO)_2$ 10 mM Tris·HCl pH 7.5 5 % DMSO 20 % MPD	**E**	2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5 20 % DMSO 20 % MPD
		F	2 mM $CaCl_2$ 120 mM Tris·HCl pH 7.5 20 % DMSO 30 % MPD

6.2. Electron density maps of the refined hits

	Omit maps	Refined maps
J13 Resolution= 1.48 Å Occupancy= 1.00 PDB ID= 6SBK		
J13a Resolution= 1.48 Å Occupancy= 0.92 PDB ID= 6SBK		
J22 Resolution= 1.53 Å Occupancy= 1.00 PDB ID= 6SC0		

J28 Resolution= 1.30 Å Occupancy= 1.00 PDB ID= 6SB9		
J62 Resolution= 1.82 Å Occupancy= 1.00 PDB ID= 6SC3		
J77 Resolution= 1.41 Å Occupancy= 1.00 PDB ID= 6SCK		
J88 Resolution= 1.42 Å Occupancy= 0.80 PDB ID= 6SCU		

J96

Resolution= 1.56 Å

Occupancy= 1.00

PDB ID= 6SC1
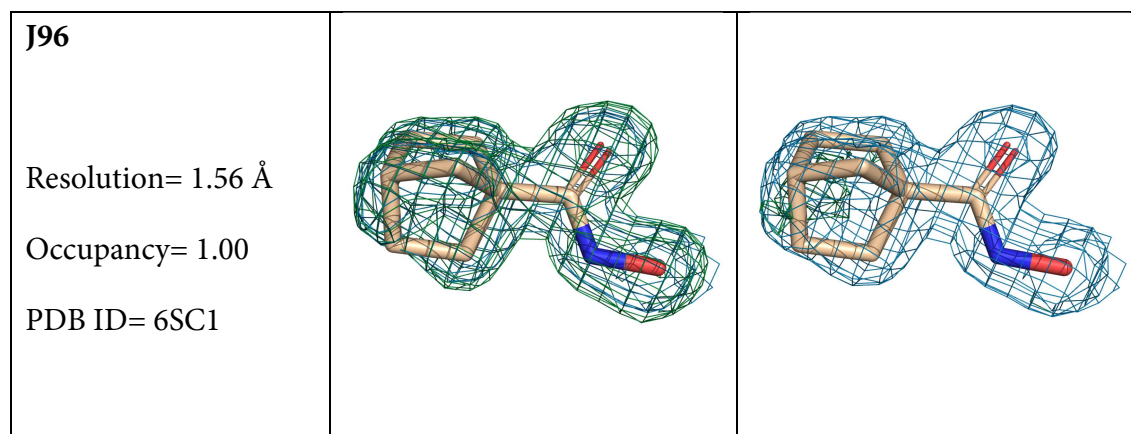
Figure 3.21. Comparison between omit maps and fully refined electron density maps that justify the fragments built into the electron density. Thermolysin hits are represented as a stick model and colored according to Figure 3.1. In the central column, the omit maps clearly reveal the presence of the fragments in the electron density. In the right column, final and fully refined electron density maps. 2mFo-DFc maps are contoured at 1 σ and represented as a blue mesh while mFo-DFc maps are contoured at 3 σ and represented as green mesh.

6.3. Data collection and refinement statistics [a]

PDB code	6SBK	6SC0	6SB9	6SC3
	J13-TLN	**J22**-TLN	**J28**-TLN	**J62**-TLN
(A) Data collection and processing				
X-ray source	ELETTRA, XRD1	ELETTRA, XRD1	BESSY II, 14.1	ELETTRA, XRD1
Wavelength (Å)	1.00000	1.00000	0.91840	1.00000
space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
unit cell parameters a=b, c (Å)	92.56, 129.27	92.57, 129.11	92.83, 129.13	92.74, 129.01
α=β, γ (°)	90, 120	90, 120	90, 120	90, 120
Matthews coefficient [b] ($Å^3$ /Da)	2.3	2.3	2.3	2.3
solvent content [b] (%)	47.1	47.1	47.4	47.2
(B) Diffraction data				
resolution range (Å)	46.28-1.48	46.28-1.53	46.41-1.30	46.38-1.82
unique reflections	54619 (8612)	49092 (7698)	80826 (12774)	29838 (4713)
Redundancy	7.6 (7.3)	6.5 (6.5)	9.8 (9.7)	5.3 (5.2)
$R(I)_{sym}$ (%)	6.1 (47.8)	5.0 (46.4)	5.8 (49.4)	8.0 (47.1)
Wilson B factor ($Å^2$)	13.8	16.3	12.1	18.9
Completeness (%)	99.2 (98.3)	98.6 (97.2)	99.8 (99.3)	99.5 (99.6)
CC (1/2) (%) [e]	99.9 (91.1)	99.9 (92.3)	99.9 (92.0)	99.8 (90.1)
<I/σ(I)>	20.24 (3.99)	21.17 (4.11)	21.18 (4.07)	14.72 (3.48)
(C) Refinement				
resolution range (Å)	46.28-1.48	46.29-1.53	46.42-1.30	46.38-1.82
reflections used in refinement	54619	49089	80820	29835
final R value for all reflections (work/free) (%)	15.0/16.9	14.5/16.6	12.7/15.1	16.4/20.5
protein residues	316	316	316	316
water molecules	270	263	331	189
ligand atoms	9	12	14	9	13
other ligands atoms (Zn | Ca | GOL | BCT DMSO | IPA | TRIS)	1 | 4 |-|- 12 | 4 | 8	1 | 4 |-|4 4 | 4 | 8	1 | 4 |-|- 20 |-| -	1 | 4 |-| 4 4 |-| -
RMSD from ideality: bond lengths (Å)	0.004	0.008	0.006	0.008
RMSD from ideality: bond angles (°)	0.750	0.943	0.914	0.889
Ramachandran plot [c]				
residues in most favored regions (%)	89.3	88.1	88.5	89.3
residues in additionally allowed regions (%)	9.6	10.7	10.4	9.6
residues in generously allowed regions (%)	0.7	0.7	0.7	0.7
residues in disallowed regions (%)	0.4	0.4	0.4	0.4
Mean B factor protein ($Å^2$) [d]	15.5	18.7	14.2	22.0
Mean B factor ligand ($Å^2$) [d]	24.5 | 31.7	32.5	18.9	27.0
Mean B factor water molecules ($Å^2$) [d]	26.6	30.9	30.2	29.0
Mean B factor other ligands (Zn |Ca | GOL | BCT DMSO | IPA | TRIS)	11.9 | 12.5 |-|- 24.1 | 34.0| 28.3	14.7 | 15.2 |-| 17.7 24.2 | 27.2 | 23.9	12.1 | 12.0 |-|- 25.4 |-|-	17.8 | 19.2 |-| 24.8 36.4 |-|-

[a]Values in parenthesis describe the highest resolution shell. [b] Calculated with MATTPROB [23]. [c] Calculated with PROCHECK [24]. [d] Mean B factors were calculated with MOLEMAN. [e] R(I)sym = (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i))).

PDB code	6SCK	6SCU	6SC1
	J77-TLN	J88-TLN	J96-TLN
(A) Data collection and processing			
X-ray source	BESSY II, 14.1	BESSY II, 14.1	BESSY II, 14.1
Wavelength (Å)	0.91840	0.91840	0.91840
space group	$P6_122$	$P6_122$	$P6_122$
unit cell parameters a=b, c (Å)	92.79, 129.45	92.31, 129.62	92.80, 128.90
$\alpha=\beta$, γ (°)	90, 120	90, 120	90, 120
Matthews coefficient [b] (Å^3/Da)	2.3	2.3	2.3
solvent content [b] (%)	47.4	47.0	47.3
(B) Diffraction data			
resolution range (Å)	46.40-1.41	46.15-1.42	46.40-1.56
unique reflections	63850 (10112)	61953 (9812)	47287 (7457)
Redundancy	13.0 (13.0)	13.0 (12.9)	12.9 (12.9)
R(I)$_{sym}$ (%)	7.6 (49.7)	8.0 (49.5)	6.7 (49.4)
Wilson B factor (Å^2)	13.4	13.3	16.7
Completeness (%)	100.0 (99.8)	100.0 (99.9)	99.9 (99.4)
CC (1/2) (%) [e]	99.9 (94.3)	99.9 (94.5)	99.9 (94.0)
$<I/\sigma(I)>$	20.07 (4.49)	18.89 (4.49)	23.95 (4.83)
(C) Refinement			
resolution range (Å)	46.39-1.41	46.15-1.42	50.00-1.56
reflections used in refinement	63843	61946	47281
final R value for all reflections (work/free) (%)	12.9/15.5	13.2/16.2	15.1/17.5
protein residues	316	316	316
water molecules	296	254	218
ligand atoms	12	9	11
other ligands atoms (Zn | Ca | GOL | BCT DMSO | IPA | TRIS)	1 | 4 |-|- 24 |-| 8	1 | 4 |-|- 4 |-|-	1 | 4 | 24 |- 12 | 4 |-
RMSD from ideality: bond lengths (Å)	0.007	0.007	0.008
RMSD from ideality: bond angles (°)	0.916	0.882	0.930
Ramachandran plot [c]			
residues in most favored regions (%)	88.9	90.0	89.3
residues in additionally allowed regions (%)	10.4	8.9	9.6
residues in generously allowed regions (%)	0.4	0.7	0.7
residues in disallowed regions (%)	0.4	0.4	0.4
Mean B factor protein (Å^2) [d]	15.2	15.3	19.6
Mean B factor ligand (Å^2) [d]	24.7	15.8	24.9
Mean B factor water molecules (Å^2) [d]	30.3	30.9	29.1
Mean B factor other ligands (Zn | Ca | GOL | BCT DMSO | IPA | TRIS)	15.1 | 13.0 |-|- 29.4 |-| 13.9	10.5 | 12.8 |-|- 25.7 |-|-	16.3 | 16.1 | 42.8 |- 34.7 | 27.7 |-

[a]Values in parenthesis describe the highest resolution shell. [b] Calculated with MATTPROB [23]. [c] Calculated with PROCHEK [24]. [d] Mean B factors were calculated with MOLEMAN. [e] R(I)sym = (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i))).

6.4.   Additional software used

6.4.1.   MOE for energy minimization calculation

Molecular Operating Environment (MOE, version 2019.0101) was used for energy minimization of the crystallographic complex TLN-J88 (6SCU) (Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, QC, Canada).

6.4.2.   DSX for per-atom score contributions calculation

The scoring function DSX (version 0.9) was used to calculate the per-atom score contributions. The putative ligand-protein complex geometry is automatically displayed as a PyMOL session. Favorably interacting atoms are shown as blue spheres whereas unfavorable interactions as red spheres. The sizes of the spheres correspond to the values of the contributing per-atom scores, e.g. a large blue sphere denotes a more attractive interaction than a smaller one, a large red sphere stands for a more repulsive interaction than a smaller one [25].

The input file for DSX is the docking pose in format *.mol2* which was converted from *.sdf* file using fconv[26].

Docking was performed using HYBRID (version 3.3.0.3, OpenEye Scientific Software)[27]. Prior to docking, the receptor preparation was performed using the program pdb2receptor (OEDocking version 3.0.1. OpenEye Scientific Software, Santa Fe, NM, http://www.eyesopen.com). The output file is in *.oeb.gz* format. Ten conformers of J88 were generated using OMEGA (version 3.1.0.3, OpenEye Scientific Software) [28].

For docking:

- The input file *.oeb.gz* contained the receptor under investigation;
- The input file *.mol2* is the fragment **j88**.

The *.sdf* output file contained the fragment **J88** docked into Thermolysin and it was used for DSX calculation.

7. References

[1]   V. Pelmenschikov *et al.*, "A theoretical study of the mechanism for peptide hydrolysis by thermolysin," *J. Biol. Inorg. Chem.*, vol. 7, no. 3, pp. 284–298, 2002.

[2]   S. Kunugi *et al.*, "pH and Temperature Dependences of Thermolysin Catalysis: Catalytic Role of Zinc-Coordinated Water," *Eur. J. Biochem.*, vol. 124, no. 1, pp. 157–163, 1982.

[3]   P. M. Colman *et al.*, "The structure of thermolysin: An electron density map at 2.3 Å resolution," *J. Mol. Biol.*, vol. 70, no. 3, pp. 701–724, 1972.

[4]   O. A. Adekoya *et al.*, "The thermolysin family (M4) of enzymes: Therapeutic and biotechnological potential," *Chem. Biol. Drug Des.*, vol. 73, no. 1, pp. 7–16, 2009.

[5]   M. Smollich *et al.*, "On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer," *Breast Cancer Res. Treat.*, vol. 106, no. 3, pp. 361–369, 2007.

[6]   J. P. Leite *et al.*, "Alzheimer's Aβ 1-40 peptide degradation by thermolysin: evidence of inhibition by a C-terminal Aβ product," *FEBS Lett.*, vol. 593, no. 1, pp. 128–137, 2019.

[7]   C. M. Kam *et al.*, "Inhibition of Thermolysin and Carboxypeptidase A by Phosphoramidates," *Biochemistry*, vol. 18, no. 14, pp. 3032–3038, 1979.

[8]   S. G. Krimmer *et al.*, "Methyl, Ethyl, Propyl, Butyl: Futile But Not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors," *ChemMedChem*, vol. 9, no. 4, pp. 833–846, Apr. 2014.

[9]   S. G. Krimmer *et al.*, "Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands," *J. Med. Chem.*, vol. 59, no. 23, pp. 10530–10548, 2016.

[10]   C. Mattos *et al.*, "Locating and Characterizing Binding Sites on Proteins," *Nat. Biotechnol.*, vol. 14, no. 5, pp. 595–599, 1996.

[11]   J. Behnen *et al.*, "Experimental and Computational Active Site Mapping as a Starting Point to Fragment-Based Lead Discovery," *ChemMedChem*, vol. 7, no. 2, pp. 248–261, 2012.

[12]   A. C. English *et al.*, "Locating interaction sites on proteins: The crystal structure of thermolysin soaked in 2% to 100% isopropanol," *Proteins Struct. Funct. Genet.*, vol. 37, no. 4, pp. 628–640, 1999.

[13]   J. Schiebel *et al.*, "High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits.," *Structure*, vol. 24, no. 8, pp. 1398–409, Aug. 2016.

[14]   P. V. Kortunov *et al.*, "CO2 Reaction Mechanisms with Hindered Alkanolamines: Control and Promotion of Reaction Pathways," *Energy and Fuels*, vol. 30, no. 2, pp. 1223–1236, 2016.

[15]   U. Mueller *et al.*, "The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives," *Eur. Phys. J. Plus*, vol. 130, no. 7, 2015.

[16]   W. Kabsch *et al.*, "XDS," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 125–132, 2010.

[17]   K. M. Sparta *et al.*, "XDSAPP2.0," *J. Appl. Crystallogr.*, vol. 49, no. 3, pp. 1085–1092, Jun. 2016.

[18]   A. J. McCoy *et al.*, "Phaser crystallographic software ," *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007.

[19]   P. Emsley *et al.*, "Coot: Model-building tools for molecular graphics," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12 I, pp. 2126–2132, 2004.

[20]   E. Krissinel *et al.*, "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12 I, pp. 2256–2268, 2004.

[21]   N. M. Pearce *et al.*, "A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density," *Nat. Commun.*, vol. 8, pp. 24–29, 2017.

[22]   N. M. Pearce *et al.*, "Partial-occupancy binders identified by the Pan-Dataset Density Analysis method offer new chemical opportunities and reveal cryptic binding sites," *Struct. Dyn.*, vol. 4, no. 3, 2017.

[23]   K. A. Kantardjieff *et al.*, "Proceedings of 10th International Conference on Formal Methods in Computer-Aided Design, FMCAD 2010, Lugano, Switzerland, October 20-23," *Fmcad*, pp. 1865–1871, 2010.

[24]   R. A. Laskowski *et al.*, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, 1993.

[25]   G. Neudert and G. Klebe, "DSX: A knowledge-based scoring function for the assessment of protein-ligand complexes," *J. Chem. Inf. Model.*, vol. 51, no. 10, pp. 2731–2745, 2011.

[26]   G. Neudert and G. Klebe, "fconv: Format conversion, manipulation and feature computation of molecular data," *Bioinformatics*, vol. 27, no. 7, pp. 1021–1022, 2011.

[27]   M. McGann *et al.*, "FRED and HYBRID docking performance on standardized datasets," *J. Comput. Aided. Mol. Des.*, vol. 26, no. 8, pp. 897–906, 2012.

[28]   P. C. D. Hawkins *et al.*, "Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database," *J. Chem. Inf. Model.*, vol. 50, no. 4, pp. 572–584, 2010.

Part III: Farnesyl pyrophosphate synthase (FPPS)

Understanding protein stability by analyzing its properties both in solution and in crystalline form to carry out a crystallographic screening of an inhouse fragment library.

1. Introduction

1.1. Neglected disease

According to the World Health Organization (WHO), neglected tropical diseases (NTDs) are a diverse group of diseases that prevail in tropical and subtropical conditions present in 149 countries. They affect more than one billion people, approximately one-sixth of the world's population, especially communities living in poverty, without adequate sanitation and in close contact with infectious vectors, domestic animals and livestock (www.who.int; Retrieved 11 October 2019)).

They are called "neglected," because they do not receive as much attention as other diseases such as HIV or malaria by the pharmaceutical industry or organizations administering the research funds (www.who.int; Retrieved 11 October 2019).

People affected by these neglected and severe diseases do not attend school, are unable to work and do not have sufficient medical care, thus reducing their quality of life and opportunities for success. NTDs can thus reinforce the cycle of poverty among the world's disadvantaged populations. However, globalization, tourism and increasing migration flows contributed in spreading NTDs also among the emerging market economies and wealthy countries, including North America, Europe and other industrialized countries which are part of the G20. It is therefore a worldwide problem. This new mapping of the global NTDs burden was named "blue-marble health", referring to the photograph of Earth taken by the crew of spacecraft Apollo 17 in 1972, and is meant to foster a global dialogue, becoming the symbol of the need for global health cooperation, promoting awareness among G-20 countries and encouraging international cooperation and support for lower income populations[1], [2], [3].
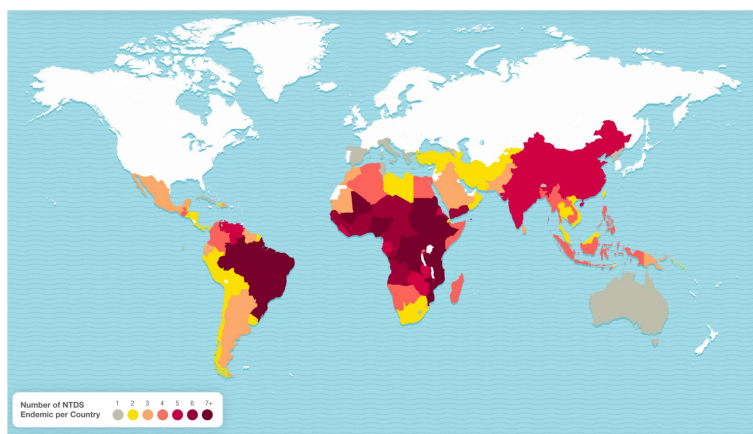
Figure 1.1. Map displaying the countries in which NTDs are endemic based on 2009-2010 data and international borders. Figure taken from [3].

A breakthrough in efforts to combat neglected tropical diseases dates back to 2007, when WHO convened the first meeting of global partners resulting in a shared commitment to support its strategies, goals and targets. In 2010 the first WHO report on NTDs was released to demonstrate the successes achieved since 2007 but also to plan future goals such as poverty reduction, universal primary education and improved maternal health. A historic moment in the fight against NTDs occurred in 2010 when the WHO Strategic and Technical Advisory Group for Neglected Tropical Diseases and partners launched a roadmap which contains the targets and milestones for the prevention, control, elimination and eradication of NTDs in the period 2012-2020. Objectives include preventive chemotherapy which aims at optimizing the large-scale use of safe and single-dose medicines, intensification of disease management to improve cases detection, preventing mortality, reducing morbidity and interrupting transmission**.** Humans, through their behavior, culture or food supply, have become incorporated in the transmission cycle of pathogens. These infections have therefore a zoonotic component that makes them more difficult to combat (www.who.int; Retrieved 11 October 2019) .[4]

Oher resolutions aim thus at improving veterinary public health through an integrated human and animal health approach and at providing drinking water, sanitation and hygiene for the 900 million people who do not have access to safe drinking water and for

the 2.5 billion people who live without adequate sanitation. These initiatives would thus reduce the transmission and propagation of NTDs ensuring better control. In the roadmap, WHO also defined 17 diseases as neglected including dengue, rabies, leprosy, Chagas disease, human African trypanosomiasis and leishmaniasis. (www.who.int; Retrieved 11 October 2019).

On 30 January 2012, the London Declaration for NTDs gathered academic and research partners, multinational pharmaceutical companies, world banks, United States government and non- governmental organizations and donors. Members renewed their commitment in helping the WHO to achieve its goals by controlling or eliminating ten NTDs by the end of this decade. The London Declaration also promoted "next-generation treatments and interventions" and ensured multi-drug therapies and availability, vector control using insecticidal spraying and the access to clean water, basic sanitation, health education and stronger health systems in endemic areas as well as helping neighboring countries to prevent its propagation [2].

In view of the progress and challenges achieved by the WHO Roadmap on Neglected Tropical Diseases 2012-2020, a further World Health Assembly initiative was held in Geneva on 09 April 2019 and in line with the WHO's 13th General Work Program 2019-2023 (GPW13). It concerns the expansion of access to prevention, diagnosis, treatment and care interventions with the goal of achieving universal health coverage by 2030.

Among the neglected diseases identified by the WHO is trypanosomiasis which include the human African and American trypanosomiasis.

The main focus of the present work is to investigate the farnesyl pyrophosphate synthase (FPPS), a target proteins involved in trypanosomiasis.

1.2.   Trypanosomiasis

The term trypanosomiasis refers to several diseases in vertebrates caused by parasitic protozoan trypanosomes which belong to the order Kinetoplastea and to the family of the hemoflagellate Trypanosomatidae. The name derived from the Greek *trypano-* (borer) and *soma* (body) because of their corkscrew-like motion. The majority of trypanosomes are heteroxenous and therefore need more than one obligatory host to complete their life cycle and are usually transmitted *via* an infected vector (www.who.int, Retrieved 11 October 2019) .

The trypanosomes which are pathogenic in humans are the species *Trypanosoma brucei*, especially the subspecies *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense*, and *Trypanosoma cruzi*. The first causes the human African trypanosomiasis (HAT) also known as sleeping sickness. The second one is responsible of the American trypanosomiasis better known as Chagas disease (CD). Both carry the parasite through a transmission vector which is the Tsetse fly (species *Glossina*) in the sleeping sickness while the insect responsible for Chagas disease is the so-called kissing or vampire bug (family *Reduviidae* and subfamily *Triatominae)*.

1.2.1   Distribution

Tsetse flies exist almost exclusively in Africa and therefore sleeping sickness is endemic in this continent where nearly 300 separate active foci are recognized with about 60 million people at risk in 36 out of 52 of Africa's countries [5]. In the last 10 years, over 70% of reported cases occurred in the Democratic Republic of the Congo while between 10 and 100 new cases were declared among Angola, Central African Republic, Chad, Congo, Gabon, Guinea, Malawi and South Sudan in 2018 (WHO, https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african- Retrieved 11 October 2019).
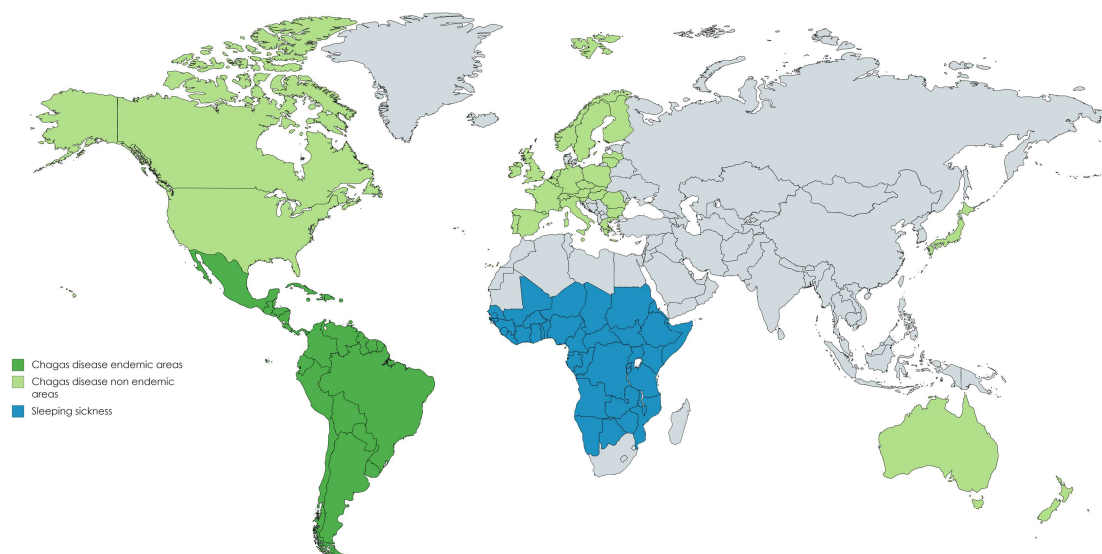
Figure 1.2. Map displaying the distribution of Sleeping sickness (blue) and Chagas disease (green) in endemic countries. Due to increasing imigration, CD is spreading to non-endemic countries (light green) such as U.S., Canada, Europe, Japan, Australia and New Zealand. Statistical data were taken from [6], [7] and [8]. Figure adapted from [9] and created using mapchart.net (https://mapchart.net/).

Chagas disease is endemic in 21 continental Latin American countries especially in Bolivia (6.8% prevalence), Argentina (4.1 %), El Salvador (3.4 %), Honduras (3.1 %), Paraguay (2.5 %), Guatemala (2 %), Ecuador (1.7 %), French Guyana, Guyana, and Surinam (1.2 %), Venezuela (1.2 %), Nicaragua (1.1 %), Brazil (1 %), and Mexico (1 %) [10]. In 2016, the WHO estimated that a total of 6-8 million people are infected with *T. cruzi* with about 12,000 deaths each year attributed to Chagas disease (www.who.net; Retrieved 11 October 2019). In recent decades, many efforts were made to reduce the incidence of CDs in Latin America but meanwhile a dramatic increase in CD cases was recorded in non-endemic countries. In fact, due to the constant migration of individuals from endemic areas CD is becoming a severe concern in non-endemic countries such as the USA, Canada, Europe, Australia and Japan. In particular, in 2005, out of 22 million people coming from Latin American countries and living in the U.S, 300,167 were infected with *T. cruzi* while 3.5 % of immigrants living in Canada were estimated to be infected in 2006. There is a large number of immigrants from the CD's endemic countries also in Europe. There are about 5 million Latin Americans living mainly in Spain, Italy, France, United Kingdom, and

Switzerland. In 2009, there were about 123,078 immigrants and 68,318 infected cases recorded with the great majority (ca. 48 million) in Spain. Moreover, 4000 cases of *T. cruzi*-infected residents were estimated in Japan while 1,928 and 82 were the reported cases in Australia and New Zealand, respectively [6], [7], [8] and [9] (Figure 1.2.).

In 110 years from its discovery and characterization in Latin America by *Carlos Ribeiro Justiniano Chagas*, CD reached a global distribution. Since 1909 when the first case was diagnosed in Berenice, a 2-year-old child, CD has become a great concern in both endemic and non-endemic countries [8]. There is still much to be done to completely eradicate the disease, but active collaborations between countries around the world and the search for new drugs is certainly essential.

1.2.2   Transmission and Lifecycle

As described in Section 1.2., HAT and CD are transmitted by vectors which are Tsetse fly and triatomine bugs, respectively. Being heteroxenic parasites, the life cycles of *T. brucei* and *T. cruzi* involve two hosts: the insect vector and the mammalian cells. In both hosts occur several differentiation stages which are essential for the progression of the infection.

In the life cycle of *T. cruzi* there are alternating stages in which the parasite is present in the blood-sucking triatome insect and others in humans. *T. cruzi* has the capability to infect any cell, mainly macrophages, fibroblasts and epithelial cells. During its life cycle (Figure 1.3.), the parasite evolves through three main forms: amastigotes (proliferative form found inside the human cells), epimastigotes (proliferative form found in the intestine of the kissing bug), and metacyclic trypomastigotes (the infective form)[8]. In the peripheral bloodstream of the mammalian host, *T. cruzi* exists in the form of trypomastigotes (1). During its blood meal, the insect vector ingests the trypomastigotes which go to the stomach where they are differentiated into spheromastigotes (2) and subsequently into epimastigotes (3) which reach the midgut. Epimastigotes attach to the epithelial intestinal cells through their long flagella. They are the replicative form (4) and differentiate into the

infective metacyclic trypomastigotes (5). Afterwards, they detach from the intestinal epithelia and migrate to insect rectum to be excreted with highly infective feces (6) [9].

Usually the infection of mammalians takes place through direct inoculation of these forms through the ocular mucosa or the lesioned skin during insect blood meal. Other transmission mechanism are blood transfusion, organ transplantation or vertical transmission from mother to child through the placenta. Although oral transmission was usually associated with infected animals, nowadays it is recognized to be an important form of transmission also in humans *via* contaminated food or beverages [8].

Once in the human host, the metacyclic trypomastigotes invade the cells at the inoculation site (e.g., fibroblasts, macrophages, and epithelial cells) and through recognition, adhesion and internalization processes involving both parasite and human cells, start the intracellular life cycle of *T. cruzi* (7). This cycle involves several steps like the formation of an endocytic vacuole known as the parasitophorous vacuole (8), the differentiation of the long and thin trypomastigote in the rounded amastigote forms (also called intracellular spheromastigotes) (9) and the simultaneous lysis of the parasitophorous vacuole membrane by some enzymes secreted by the parasite (10). Here, the amastigotes are in contact with the host cell organelles and start to divide and replicate (11)[9], [11]. Finally, the amastigotes turn into trypomastigotes (12), which burst out of the cell, enter the bloodstream (13), infect tissues (14) and spread the infection in various organs. The diagnostic stage occurs with the trypomastigotes's detection in the bloodstream.
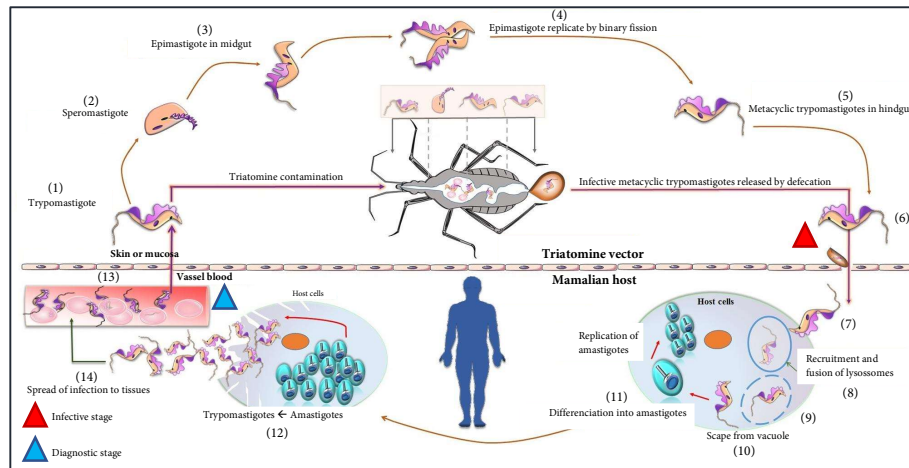
Figure 1.3. Lifecycle of *T. cruzi* parasite. Triatomine bug stage (1-6) and human stage (7-14). Figure adapted from [12].

The major risk factor in developing HAT is the Tsetse fly and is directly linked to the bite exposure level in a specific area. Although the primary transmission way is the bite of an infected Tsetse fly, the disease is also transmitted from mother to child as the parasite can cross the placenta, through sexual contact and accidental infection by contaminated needles. The people more exposed to HAT are the one living the rural area where the Tsetse fly is mostly found [13], [14] (ww.who.net; Retrieved 12 October 2019).

In contrast to the trigenic *T. cruzi* life cycle, *T. brucei* has a digenetic life cycle, alternating between the blood/tissues of the mammalian host and alimentary tract of the Tsetse fly vector [13].

During a blood meal, an infected Tsetse-tropanosome fly bites a human host, thus initiating the parasite lifecycle within the mammalian host. Metacyclic trypomastigotes are inoculated in the host dermis together with saliva components (such as Tsal) which are involved in the modulation of the host's early immune response (1) (Figure 1.4.). The parasites through the host lymphatic system reach the bloodstream where their first differentiation takes place. In fact, metacyclic trypoamastigotes express a heterogeneous metacyclic variable surface glycoproteins (VSG) coat and turn into the long slender (LS) form (2). LS trypoamastigotes are the dividing and proliferating forms in the bloodstream (BF) which express a unique bloodstream form VSG. This allows them to survive in the

highly oxygenated and glucose-rich blood of the mammalian host. Upon rapid division of the BFs, the parasitemia peak is reached and the LS forms differentiate into the non-dividing short stumpy (SS) forms (3) which are suitable to survive inside the Tsetse fly vector[13].

Upon the next blood meal, the parasite lifecycle within the Tsetse fly host starts. The SS forms are ingested by the Tsetse fly and in its midgut differentiate into the procyclic forms (PF) (4) as their coat is being replaced by the procyclin to survive in the proline-rich and low oxygenated environment of the host. In order to adapt to the different environment in the digestive tract and in the salivary glands of the fly, the parasite goes through several differentiation stages. The PFs migrate to the proventriculus and express a bloodstream alanine rich protein coat differentiating in the epimastigote forms (5). Subsequently, they migrate to the salivary glands where they differentiate into metacyclic forms (6) that are ready to complete their life cycle[13].
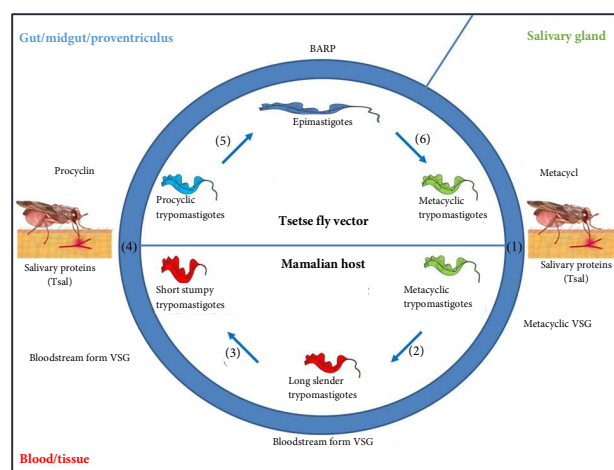
Figure 1.4. Lifecycle of the *T. brucei* parasite. Human stage (1-4) and tsetse fly stage (4-6). Figure adapted from [13].

1.2.3   Symptoms and Diagnosis

CD and HAT have a complex pathophysiology and their symptoms vary from person to person with different clinical manifestations.

CD may develop the following three clinical phases: The acute phase, the asymptomatic or indeterminate phase and the chronic phase[9].

The acute phase is usually asymptomatic. The most fortunate people show generic symptoms such as prolonged fever, headache, fatigue, body aches, muscle pain, rash, loss of appetite, diarrhea, nausea and vomiting but also rarely symptoms as lymphadenitis, hepatomegaly and splenomegaly within 60 days from the infection. In case of vectoral transmission, the infected individual may present clinical signs at the entry portal resulting from the inoculation of *T. cruzi*. When the entry occurred through the skin, they develop a Chagoma which is a nodular or ulcerative skin lesion. If infection occurred through the periorbital mucosa the characteristic Romaña sign is developed, which is a unilateral painless periorbital swelling.

After four to eight weeks the disease enters the chronic phase with simultaneous cessation of the clinical manifestations. In about 60-70 % of the people the chronic infection remains clinically silent for life. This is the most common asymptomatic or indeterminate form of CD. Clinical manifestations may occur in 30-40 % of cases even after 10-30 years. They mostly affect the heart and the gastrointestinal tract especially the esophagus, the stomach, the small bowel and the colon [8], [9].

The most severe clinical manifestation of CD is the Chronic chagasic cardiomyopathy (CCC). It is present in 1/3 of infected individuals and it is characterized by diffuse myocarditis, followed by the progressive replacement of muscle tissue with fibrous tissue and lastly the heart failure which is the late stage of clinical progression[8].

In the acute phase, CD can be diagnosed by parasitological tests detecting the trypomastigotes circulating in the bloodstreams while in the chronic phase by at least two serological tests such as immunofluorescence and enzyme-linked immunosorbent assay (ELISA) to detect the anti-*T. cruzi* IgG antibodies. In addition, routine laboratory tests as electrocardiogram (ECG) and chest radiography should be performed at any stage of the disease [8] (www.who.net; Retrieved 11 October 2019).

Differently from the traditional acute or chronic classification present in CD, the stages of HAT are named accordingly to the clinical progression of the disease[9].

The first stage or haemolymphatic phase begins when the metacyclic trypomastigotes enter the body and reach the bloodstream and the lymphatic system. Here, an inflammatory reaction starts, causing swelling of the skin and enlargement of the lymph nodes in the neck. This stage involves non-specific symptoms such as fever, headaches, joint pains and itching.

The second stage, also called meningoencephalitic phase, begins when the trypanosome parasites cross the blood-brain barrier (BBB) and infect the central nervous system including the brain. This leads to serious neurological disorders as confusion, poor coordination, difficulties with speech, paralysis, progressive mental deterioration as well as sleep cycle disruptions in which the patient sleeps through the day and suffers from insomnia at night. This disorder inspired the common name of the disease, i.e., sleeping sickness [9], [14].

The diagnosis of HAT must be done as early as possible to prevent the disease to progres in the meningoencephalitic phase. It involves two screenings: (i) Checking for clinical signs such as swollen lymph nodes in the neck and extended daytime sleeping and (ii) direct detection of parasites in blood or cerebrospinal fluid by microscopic examination [14](www.who.net; Retrieved 12 October 2019). However, currently, there are no effective and reliable methods available to diagnose HAT. In fact, the host antibody response is rather severe, and antibodies remain in circulation in the bloodstream for long periods of time. This prevents discrimination between active and treated infections based on antibody detection. In addition, methods for detecting antigen-based trypanosomes that would bypass the antibody-based detection system are not available yet[13].

A recent discovery [13] involves the use of nanobodies (Nbs) obtained from llama's which are immunized with factors derived from (i) Tsetse flies such as the procyclin coat isolated form procyclic forms or (ii) the mammalian host, in particular from the parasites

glycoprotein present in the blood stream such as the variable surface glycoprotein (VSG) to detect the tsetse fly bite exposure.

An exhaustive and active screening of the population at risk is necessary both in CD and HAT to identify patients at an early stage and reduce transmission. In fact, infected individuals may die before they can be diagnosed and treated.

1.2.4   Pharmacological treatment and control strategies

Some drugs for the treatment of CDs and HATs are currently available. However, due to their low pharmacokinetic profile and severe side effects, it is necessary to find new drugs [5].

Nifurtimox (**1**, NFX, Figure 1.5.) and Benznidazole (**2**, BNZ, Figure 1.5.), two nitro-heteroaromatic compounds introduced in the 1960s and 1970s, respectively, are the current therapeutic options for the treatment of CD. Importantly, both NFZ (nitrofuran) and BNZ (nitroimidazole) have a nitro group which is responsible of their pharmacological action. Like other nitro-compounds, their mechanism of action is to generate radical species (ROS) which can damage the parasite's DNA or cellular organelles. Under anaerobic conditions, the nitro group is reduced by nitro-reductase enzymes (NTR) to create reactive nitro radical species which react with the nucleic acids of the parasite causing a significant breakdown of DNA and thus inhibiting DNA and proteins synthesis. They also inhibit a parasite-specific antioxidant defense enzyme called trypanothione reductase [15], [16], [17].

1   **2**
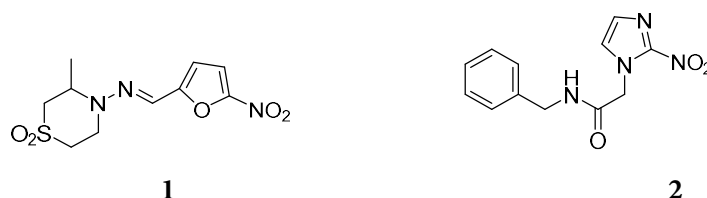
Figure 1.5. (1) Nifurtimox (NFX) and **(2)** Benznidazole (BNZ) for the treatment of CD.

Both drugs act against the acute phase, reducing the severity of the symptoms and the duration of the parasitemia but they are not effective against the chronic phase[9], [5]. They both have trypanocidal activity, mainly against circulating trypomastigotes and they are better tolerated by young children compared to adults. It was demonstrated that BNZ is safer and more efficacious than NFZ and it is thereby the first line treatment[9]. However, they are rapidly metabolized by the CYP450 system, so their administration requires high and prolonged doses[17]. The treatment consists of oral BNZ administration of 5–10 mg/kg per day in two doses for 30–60 days or 8–10 mg/kg of NFZ in three doses for 30–120 days [5], [15], [16].

As a result of their long-term treatment and the high dose required for their therapeutic action, they cause several side effects mainly in the gastrointestinal tract such as vomiting, abdominal pain, anorexia, weight loss, headache, weakness, loss of memory, irritability, convulsions, sleep disturbances, tremor, polyneuritis, hypersensitive skin reactions and furthermore they are potentially genotoxic, carcinogenic and mutagenic [15], [17], [18].

Thus new strategies and formulations are needed to bypass the side effects of NFX and BNZ in the treatment of CD. Among the possible strategies are drug repurposing and prodrug strategies [18]. Drug repurposing, also known as drug repositioning, relies on finding new applications for marketed therapeutics. The antifungals Fexinidazole (**9**, FEXI, Figure 1.8.) and Posaconazole (**3**, POS, Figure 1.6.) are two repositioned drugs that reached clinical trials. These antifungal drugs inhibit the sterol 14a-demethylase (CYP51), an enzyme involved in sterol biosynthesis and essential component for the stability and fluidity of cell membranes. In fact, protozoa and fungi use ergosterol while mammals use cholesterol, thus making trypanosomes particularly sensitive to sterol synthesis inhibition [19]. As reported by *Molina et al.* POS demonstrated significant activity both *in vitro* and *in vivo* but in the clinical stage did not show sufficient trypanocidal activity in humans. However, it was suggested to combined with BNZ [18], [19]. *Torrico et al.* identified the water-soluble ravuconazole prodrug E1224 (**4**, Figure 1.6.) which could be used in combination with NFX or BNZ [18], [20].

Recently, vaccines to stop or at least delay CD progression as well as to prevent congenital transmission of *T. cruzi.* were also under investigation[21].
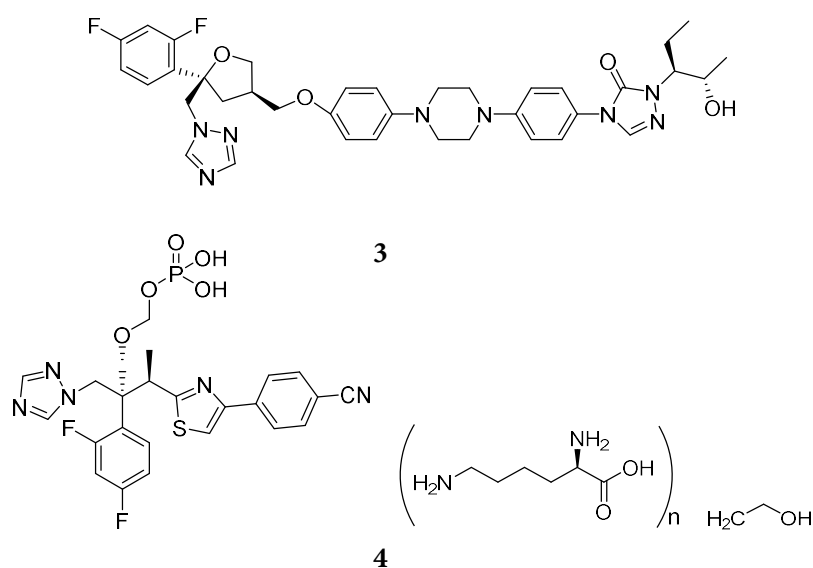
Figure 1.6. (3) Posaconazole and **(4)** prodrug E1224.

Many more drugs are available for the treatment of HAT. In particular, the drugs Pentamidine (**5**, Figure 1.7.), Melarsoprol (**6**, Figure 1.7.), Eflornithine (**7**, Figure 1.7.), Suraminx (**8**, Figure 1.7.) and NFX (**1**, Figure 1.5.) are prescribed both according to the infectious trypanosome sub-species and to the disease stage.

For HAT, caused by *T. b. gambiense*, first stage treatment involves intramuscular (IM) or intravenous (IV) administration of Pentamidine with 4 mg/kg/day by slow infusion (ca. 2 h) for 7 days. The most frequent side effect is pain at the injection site as well as hypoglycemia and hypotension. In the last 50 years the only available drug used in the second stage treatment is Melarsoprol (IV administration of 2.2 mg/kg/day for ten days). Lately, it was replaced with Eflornithine monotherapy (400 mg/kg/day) as the first-line treatment. However, to be effective, it required 56 IV infusions, each lasting 30 minutes, over at least 14 days [9]. *Priotto et al.* proposed then a combined therapy with Nifurtimox-Eflornithine (NECT) which show the same efficacy rate as the Eflornithine in monotherapy but it is easier to administer (IV infusion of 400 mg/kg of Eflornithine every 12 h for seven

day and OS administration of 15 mg/kg/day of NFX every eight hours for ten days) and potentially protective against the emergence of resistant parasites [9], [22].

The only treatment available for HAT caused by *T. b. rhodesiense* is Suramin (4–5 mg/kg the first day, then 20 mg/kg IV weekly for five weeks) for the first stage while Melarsoprol (2.2 mg/kg/day IV for ten days or three series of 3.6 mg/kg/day IV for three days spaced by intervals of seven days) for the second stage[9].
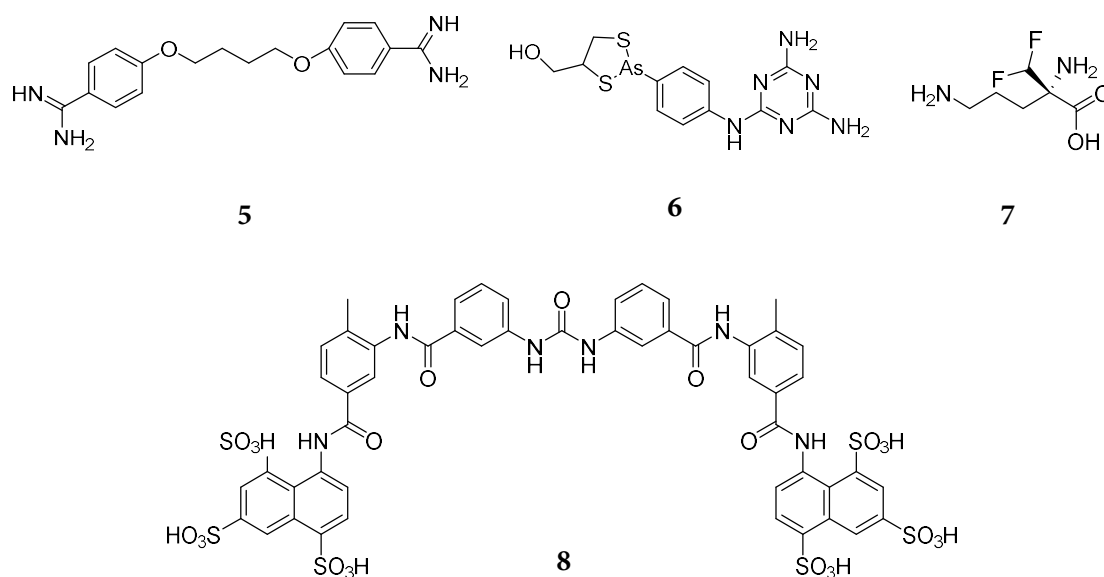
Figure 1.7. (5) Pentamidine, **(6)** Melarsoprol, **(7)** Eflornithine and **(8)** Suraminx for the treatment of HAT.

The pharmacological treatment of HAT also includes some repositioned drugs such as the Fexinidazole (**9**, FEXI, Figure 1.8.). It is a 5-nitroimidazole which inhibits the DNA synthesis and was developed by the Drugs for Neglected Diseases initiative (DNDi), in collaboration with Sanofi, for the oral treatment of HAT as well as CD. Currently, FEXI is in stage II/III of clinical trials but since November 2018 it is used for the treatment of both first and second stage in adults and children aged ≥ 6 years and weighted ≥ 20 kg[23].

Another drug in clinical trials is the boron-containing small molecule SCYX-7158 (**10**, Figure 1.8.). It was demonstrated to be potent in trypanocidal assays *in vitro* and has good physicochemical and ADME properties, i.e. it is orally available, metabolically stable and CNS permeable. It is able to cross the BBB to reach therapeutically-relevant concentrations

in the brain and cerebrospinal fluid of rodents. In a murine model it showed significant activity by oral administration, including cure of CNS[24]. It is currently in phase II/II of clinical trials (www.dndi.org/2016/clinical-trials/clinical-trials-hat/; Retrieved 17/10/2019).
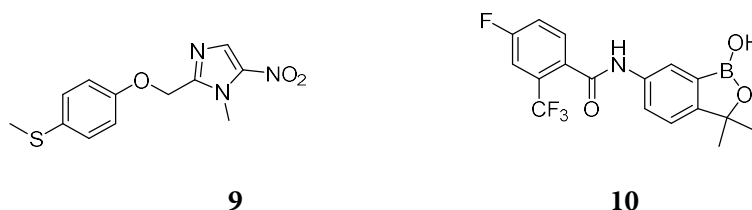
Figure 1.8. (9) Fexinidazole and **(10)** SCYX-7158 for the treatment of HAT.

1.3.   FPPS as an interesting target protein

1.3.1.   The sterol pathway

As briefly reported in Section 1.2.4. sterols are an important class of organic molecules that are essential for the cell membrane integrity, stability, permeability and fluidity. In humans, they are vital precursors for cholesterol, which is a pioneer in steroids, bile acids, lipoproteins, and vitamin D biosynthesis, as well as ubiquinone, dolichol and isopentenyladenine production. They are also involved in the post-translational prenylation of small GTP-binding proteins including the farnesylated H-Ras, K-Ras, N-Ras and the geranylgeranylated Rho, Rac and Cdc42 proteins [25]. These small proteins are responsible of several process such as cellular signal transduction, intracellular vesicular traffic and they also mediate protein-protein interaction (PPI).

Whereas mammals produce cholesterol, fungi and protozoa produce ergosterol[19]. Ergosterol differs from cholesterol by the presence of a 24-methyl group and Δ^7 and Δ^{22} double bonds. The enzymes responsible for introducing the methyl group and the double bond Δ^{22} are only present in the parasite sterol biosynthesis pathway but are missing in humans[26].
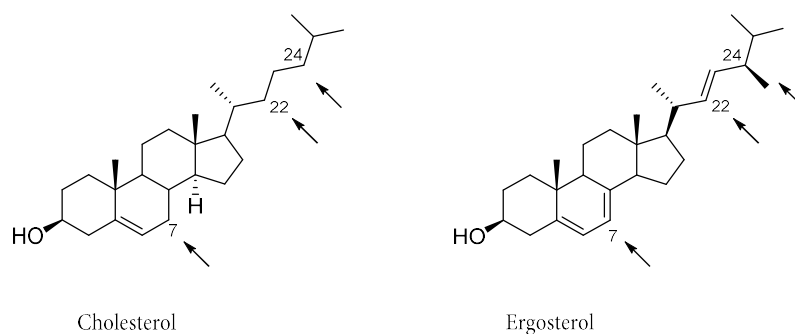
Figure 1.9. Chemical structure of Cholesterol and Ergosterol. The arrows represent their differences in position 7, 22 and 24.

Similar to humans, *T. brucei* is able to use both endogenous and exogenous sources for the sterol biosynthesis while. *T. cruzi* is mainly dependent on *de novo* production[19], [27]. It was previously discussed how *T. brucei* is able to modify its metabolism and adapt to different environments (Section 1.2.2.). In particular, in the mammalian host bloodstream the parasites exist in the long slender form (B-LS) and short stumpy form (B-SS) which are the pre-adapted form to life in the Tsetse fly midgut. Here, B-SS differentiates in the replicative procyclic forms (PFs)[28]. Already in 1995, *Coppens et al.* discovered that the circulating bloodstream forms contained cholesterol, which comes from the mammalian host. The parasite is indeed capable of incorporating cholesterol from the surrounding environment through the endocytosis receptor mediated uptake of low-density lipoprotein (LDL), lysosomal digestion and cholesterol release in the parasite cytosol[29].

However, the main source of sterols for the parasites is the endogenous production which occurs *via* the mevalonate, isoprenoid and sterol biosynthesis pathways.

Trypanosoma species, especially *T. cruzi*, are therefore particularly sensitive to sterol synthesis inhibition. It is not surprising that the sterol biosynthesis is one of the major pathways being inhibited by chemotherapy to combat *T. cruzi* and *T. brucei* parasites. As well as ergosterol, 24-alkylsterols are also important for the parasite's survival[27].

The carbon skeleton of sterol molecules originates from acetyl-CoA, which is the starting substrate of the mevalonate pathway (Figure 1.10.). The first reaction is catalyzed by the Acetoacetyl-CoA thiolase which is a SCP2-thiolase-like protein (SLP) present in the Trypanosoma parasite but absent in humans [30]. SLP catalyzes the condensation of two acetyl-CoA units to form first acetoacetyl-CoA, followed by the addition of a third unit to form 3-hydroxy-3- methylglutaryl-CoA (HMG-CoA) catalyzed by the HMG-CoA synthase. HMG-CoA is then reduced by the HMG-CoA reductase to give mevalonic acid [31].

The mevalonate pathway, which in Trypanosomatides mainly occurs in the mitochondrial compartment while in humans in the cytosol, is followed by the isoprenoid pathway (Figure 1.10.). The mevalonate undergoes a double ATP-dipendent phosphorylation step catalyzed by kinases to generate the mevalonate diphosphate which is subsequently decarboxylated to form the isopentenyl diphosphate (IPP). The isomerization of IPP by isopentenyl diphosphate isomerase produces dimethylallyl diphosphate (DMAPP). The head-to-tail condensation of IPP and DMAPP generates first geranyl pyrophosphate (GPP) and then farnesyl pyrophosphate (FPP) in a two-step reaction catalyzed by the farnesyl pyrophosphate synthase (FPPS). FPPS is located in an important branching point across several biosynthetic pathways. In fact, it can be either the precursor for ubiquinones, dolichols and heme A production as well as the substrate for farnesylated and geranylgeranylated proteins or it can also be the starting point for the sterol biosynthesis pathway[31].
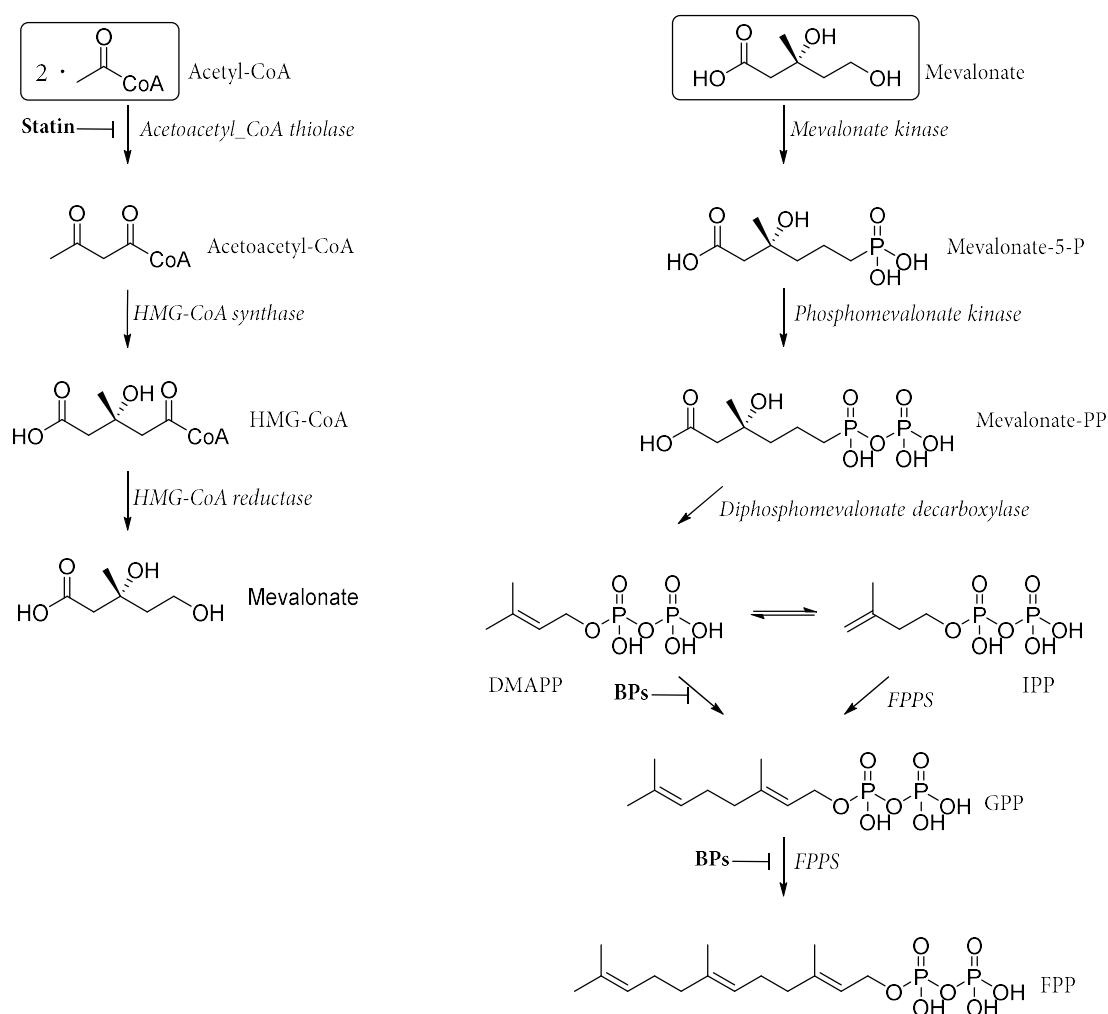
Figure 1.10. Mevalonate (left) and isoprenoid (right) pathways. Drugs inhibiting a specific enzyme are indicated as "–|"

The first committed step in sterol biosynthesis involves two consecutive condensations of two FPP molecules catalyzed by the squalene synthase (SQS) to produce squalene (Figure 1.11.). Subsequently, squalene is converted in 2,3-oxidosqualene (or squalene epoxide) by the enzyme squalene epoxidase. The 2,3-oxidosqualene is the first cyclic form of sterols and undergoes a subsequent cyclization catalyzed by the 2,2-oxidosqualene cyclase to form lanosterol in one of the most complex reactions in the entire sterol pathway. The sterol biosynthesis continues with several sequential transformations to form cholesterol in mammals and ergosterol in trypanosomatides and fungi. Some of these reactions include the demethylation of the ring system at the C4 and C14 positions. In *T. brucei* parasite, lanosterol is first converted in 31-norlanosterol by a three step reaction which leads to the

removal of the methyl group at the C4 position followed by the C14 demethylation catalyzed by the lanosterol-14α-methyl demethylase (also known as C14α-demethylase) which belongs to the cytochrome P-450 enzyme complex. In *T. cruzi* parasite, the demethylation takes place first at C14 and then at C4 position.

These reactions lead to zimosterol production which is then transformed into fecosterol by the Δ^{24}-sterol methyltransferase (also known as 24-SMT) that is present in plants, fungi, and trypanosomatides but absent in mammalian sterols. Subsequently, fecosterol is converted into episterol by the Δ^7-Δ^8 isomerase[31]. The zimosterol is the branching point for the biosynthesis of several compounds in *T. brucei* while episterol plays the same role in the *T. cruzi* parasite. At this point, the pathway is divided into two branches, one of which leads to the ergosterol biosynthesis in *T. cruzi* epiamastigotes while the other one leads to 24-sterol production in *T. cruzi* amastigotes which due to a lack of Δ^5 desaturase activity are not able to synthesize $\Delta^{5,7}$ sterols[27].

Several drugs inhibit the biosynthetic pathway of ergosterol and cholesterol in humans and trypanosomes respectively. A few cases include: (i) Statins like Atorvastatin which inhibit the HMG-CoA reductase and are used to treat hyperlipidemia in humans; (ii) Bisphosphonates (BPs) and N-Bisphosphonate (N-BPs) as Risedronate, which inhibit FPPS and are used to treat osteoporosis, different bone resorption diseases and even cancer; (iii) Quinuclidine as BPQ-OH which inhibits the squalene synthase (SQS); (iv) Allylamine as Terbinafin which inhibits the squalene epoxidase and is used especially as antifungal drug; (v) Azole as Posaconazole which inhibits the C14α-demethylase and is also used in fungal infections as well as in combination with BNZ to treat trypanosomiasis; (vi) Azasterol as 22,26-azasterol, which inhibits the 24-SMT that is not present in mammalian cells, and is therefore selectively targets *Trypanosoma species*[31].

Among the many enzymes involved in this pathway, FPPS is certainly one of the key enzymes as it is involved not only in ergosterol and 24-sterol production but is also the substrate for the synthesis of ubiquinones, dolichols and heme A as well as is implicated in

the post-translational prenylation of small GTP-binding proteins belonging to the Ras and

Rho families.
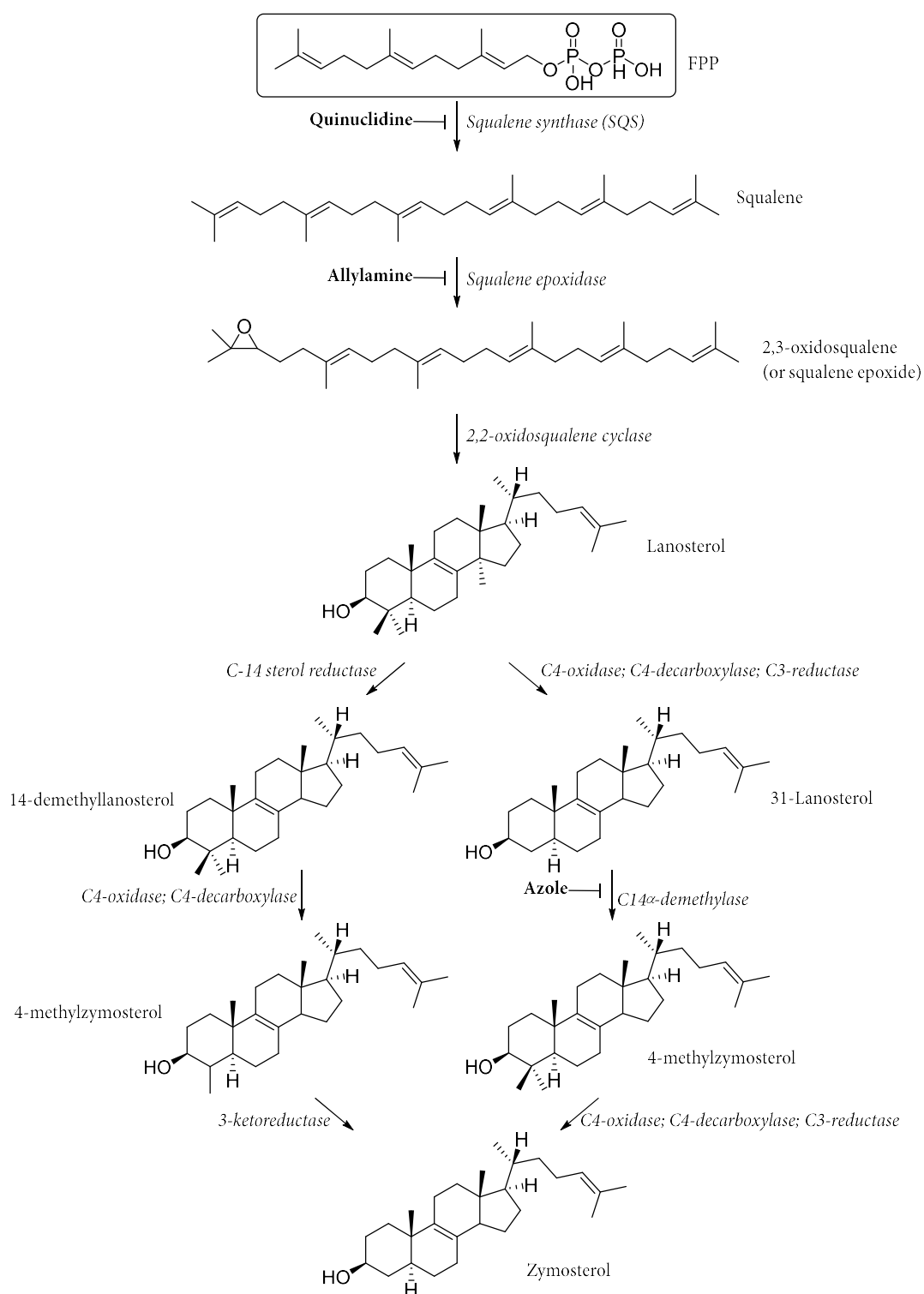
Figure 1.11. Sterol biosynthesis pathway from FPPS to Zymosterol. Drugs inhibiting specific enzymes are indicated by
"⊣"
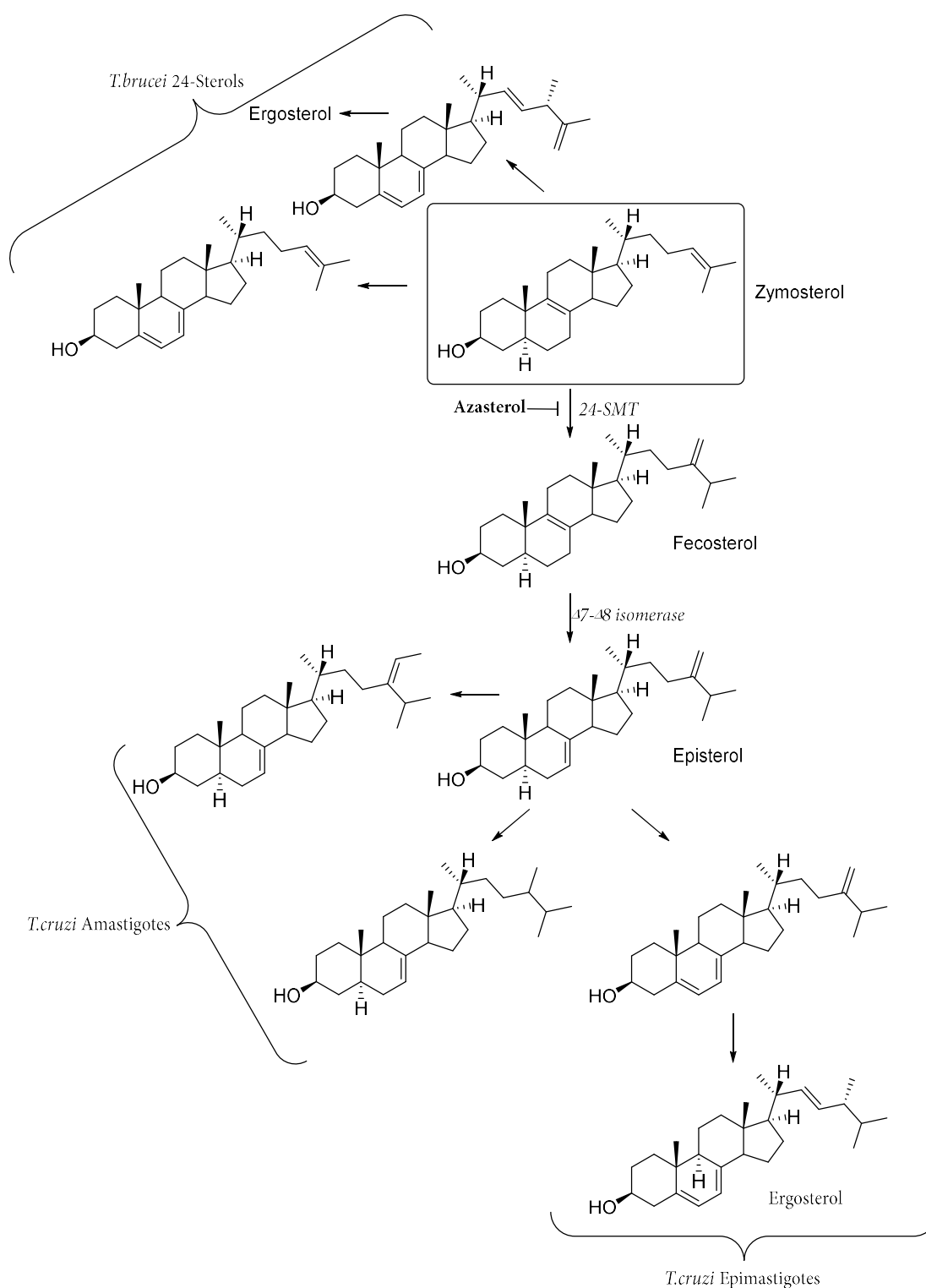
Figure 1.12. Sterol biosynthesis pathway from Zymosterol to Ergosterol. Drugs inhibiting specific enzymes are indicated by "—|". Zimosterol is the branching point for the biosynthesis of several sterols in *T. brucei* while the Episterol plays the same role in *T. cruzi* parasite. Interestingly, *T. cruzi* epimastigotes synthesize ergosterol while *T. cruzi* amastigotes are not able to synthesize $\Delta^{5,7}$ sterols because they lack Δ^5 desaturase activity.

1.3.2.   Structure topology of FPPS species

Farnesyl pyrophosphate synthase (FPPS), also known as Farnesyl diphosphate synthase (FDPS) (EC 2.5.1.10) is an enzyme which belongs to the class of transferases. It is a functional homodimer with a monomeric molecular weight in the range of 32-44 kDa depending on the species [32]. In particular, in the present work the focus was on the *human*, *T. cruzi* and *T. brucei* species in which each monomer weights 40, 41 and 42 kDa, respectively. The comparison of the amino acid sequences of several FPPS species revealed the presence of seven conserved regions (I–VII) and a characteristic 11-residues insertion between amino acids 179 –189 in *T. cruzi* and 184–194 in *T. brucei*, which are unique in the *Trypanosoma* species (Figure 1.13. yellow loop). Furthermore, *T. brucei* FPPS has an additional loop consisting of six amino acids (residues 66-71) which are missing in *T. cruzi*[32],[33].

FPPS is characterized by two Asp-rich motifs, which are important for the catalytic activity and are indicated by a DDXXD motif where "D" stands for an Asp and "X" any amino acid. These motifs are FARM (first Aspartate rich motif) and SARM (second Asp-rich motif) which are located in the II and VI domain respectively (Figure 1.13., red helices). FARM and SARM are about 12 Å apart and are located on the opposite sides of the central catalytic cavity with their carboxylic groups oriented inwardly into the cavity. The total length of the IPP final product is regulated by two amino acids which are at the fourth and fifth position upstream of the FARM and which were therefore designated as the chain length determination Region (CLD). These residues are aromatic amino acids which provide the necessary steric bulk to prevent the alkyl IPP group to exceed a particular length. They are located near the bottom of the DMAPP site and there is usually a conserved Tyr in the *Trypanosoma* species present (HY is the CLD in the *Tc*FPPS) (Figure 1.13., blue stick representation). SARM is crucial for enzymatic catalysis and together with FARM it is involved in the conversion of IPP in GPP and FPP as well as in the products release. The protozoal FPPSs also have, upstream of SARM, the conserved FQ residues, which keep DMAPP or GPP in the proper conformation required for their IPP condensation.

Furthermore, the I, V and VII domains contain conserved Lys residues which participate in substrate binding[32], [34].
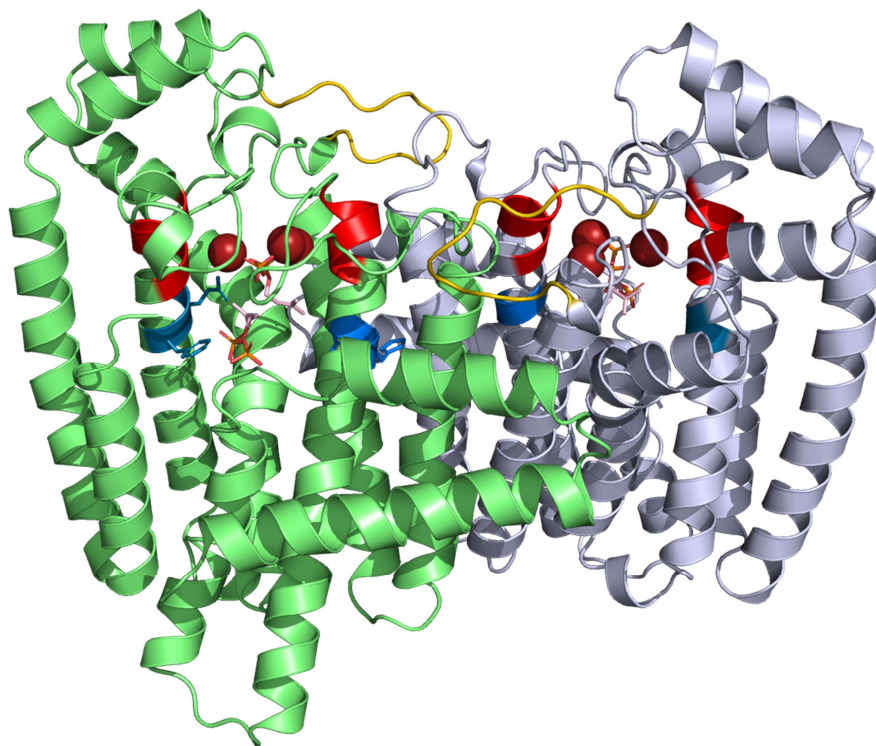
Figure 1.13. Cartoon representation of *T. cruzi* FPPS secondary structure. The two monomers which form the homodimer are depicted in green and lightblue. The natural substrates DMAPP and IPP are shown as pink sticks. In the active site, DMAPP coordinates three Mg^{2+} ions in the allylic pocket while IPP sits in the homoallylic pocket. The characteristic trypanosomal loop, which consist of 11 residues, is colored in yellow. The conserved Asp-rich FARM and SARM motifs are depicted in red and located opposite to the active site. HY, which is part of CLD, is represented as both blue helix and stick model and is located at the bottom of the allylic site.

In accordance to other species, the protozoic FPPSs are α-helical proteins. In particular, the overall structure consists of 13 α-helices divided into three regions. The first region includes the antiparallel helices A and B which are oriented orthogonally to the regions two and three. The second region consist of helices C, D and E, which are arranged antiparallel to each other along with a part of the helix F which connects the second with the third region. The FARM is located at the C-terminus of helix D while the SARM is located at the C-terminal end of helix H which belongs to the third region. Apart from helix H, the third

region comprises also the helices G, I and J, which are again arranged antiparallelly to each

other, and the three small helices α1, α2, and α3[32].
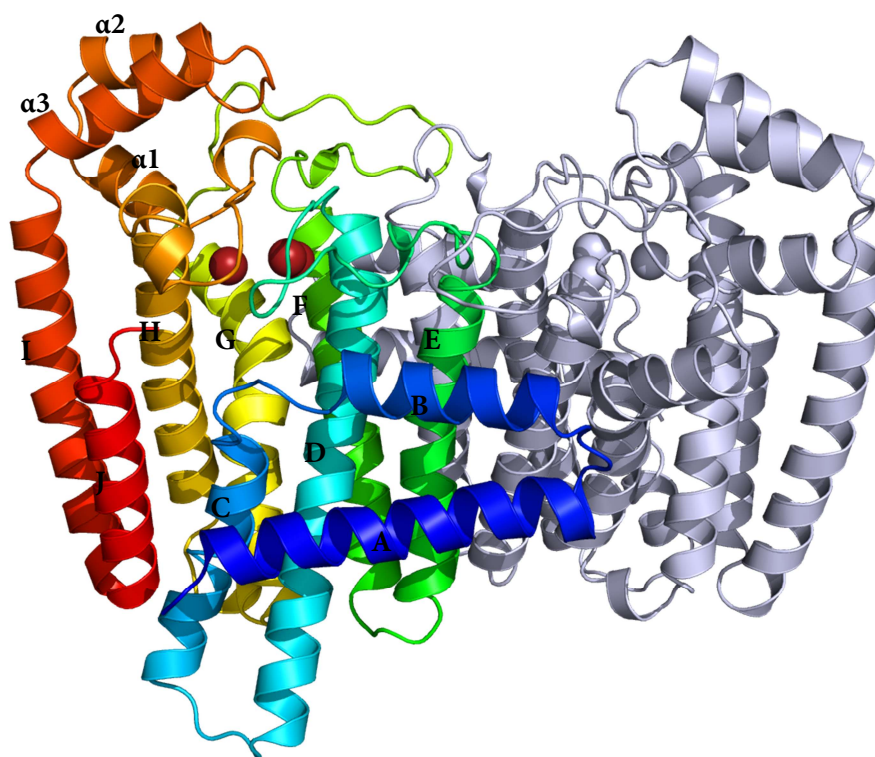
Figure 1.14. Cartoon representation of *T. cruzi* FPPS depicted with a different color accordingly to each of the 13 α-helices which form the secondary structure. One monomer is colored in a rainbow, the symmetry related one in lightblue.

The catalytic pocket of FPPS is further divided into two sub-pockets: (i) the homoallylic

pocket which hosts the IPP and (ii) the allylic one in which the DMAPP is octahedrally

coordinated with three Mg^{2+} ions[33]. In the bound form, the pyrophosphate moiety of the

DMAPP coordinates three Mg^{2+} ions which are the cofactor of the enzyme. The Mg^{2+} ions

are further coordinated by Asp98, Asp102 from the FARM and Asp250 from the SARM.

Additional water molecules complete the octahedral coordination sphere of the metal ion.

Moreover, the Mg^{2+} ions are essential in the condensation mechanism of FPPS as they

polarize the O-C bond of the DMAPP substrate and thereby promote the attack and C-C

bond formation of IPP[32].

More precisely, the FPPS mechanism of action consists of a three-step reaction indicated as ionization-condensation-elimination during which the enzyme undergoes a series of conformational changes that allow both substrate bonding and products release (Scheme 1.1) [35], [36], [37]. (a) Unliganded human FPPS adapts an open conformation. Upon DMAPP binding, the two Asp-rich DDXXD motifs approach and the H–I and D–E loops form a gate which closes the entrance to the allylic binding site, thus forcing the enzyme to adapt a "partially closed" conformation. In the allylic pocket, the polarization of the O-C bond of DMAPP triggers the ionization step of DMAPP due to the coordination with the Mg^{2+} ions together with the hydrogen-bonding network with the side chain of basic amino acids such as Lys and Arg, which results in the formation of a dimethylallyl geranyl carbocation (Scheme 1.1.). (b) In the second step of the reaction, also known as condensation step, once the IPP occupies the homoallylic pocket, the C-terminal tail of the enzyme, which consists of the basic amino acids Lys350, Arg351, Arg352 and Lys353, clamps down to close the IPP binding site in a "fully closed" conformation. The closed conformation is stabilized by electrostatic interactions between the terminal amino group of Lys57, and the C-terminal carboxylate group. It is catalytically active since it arranges IPP and DMAPP in the correct orientation for the subsequent condensation reaction. Here, the double bond of IPP attacks the C1 atom of the carbocation forming a tertiary carbocation intermediate which is stabilized by electrostatic interactions both with the phosphate group coordinated by Mg^{2+} ions and by the oxygen atoms of Lys200 and the side chain of Thr201. (c) In the final step, the intermediate undergoes elimination of the C2 proton to form the GPP product. During catalysis, the release of pyrophosphate (PP_i) from the DMAPP reduces the strength of the interactions with the DDXXD motifs. This destabilizes the closed conformation and leads to a conformational change toward the open conformation. The open conformation supports chain elongation and allows the translocation of GPP to the allylic pocket. There the binding of another IPP molecule is accomplished in order to initiate a new catalytic cycle that allows the formation of the final FPP product. [35], [37], [38].
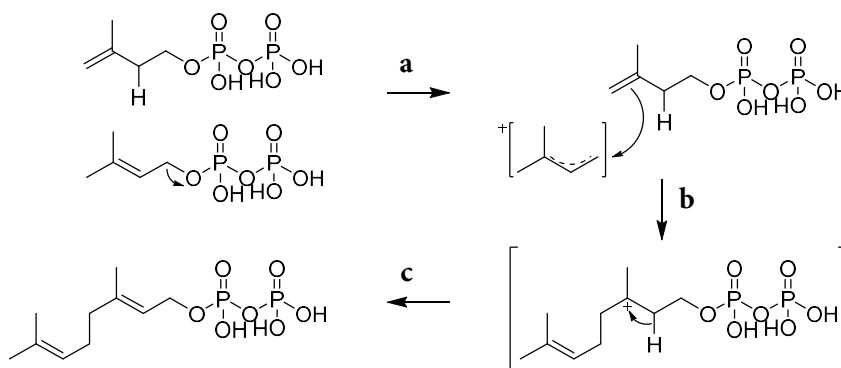
Scheme 1.1. FPPS mechanism of action consisting of **(a)** ionization, **(b)** condensation and **(c)** elimination.

1.3.3. How to inhibit FPPS activity?

1.3.3.1. Bisphosphonate as a potent active site inhibitor

As reported in section 1.3.2., BPs are a class of compounds which inhibit FPPS. They are currently used in clinical practice to treat a wide range of bone related diseases in humans. These include Paget's disease, hypercalcemia, postmenopausal and steroid-induced osteoporosis and osteolysis associated with multiple myeloma and metastatic cancers [35]. Additionally, BPs are involved in the activation of human $\gamma\delta$ T-cells, which kill tumor cells, as well as they help to reduce tumor cell invasion, migration and angiogenesis[39]. Recent studies also demonstrated that FPPS is over-expressed in the thoracic aorta collected from diabetic and hypertensive animal models and mice with over-expression of FPPS showed cardiac hypertrophy, fibrosis and heart failure. In all these cases, the administration of BPs provided cardiovascular beneficial effects both *in vitro* and *in vivo*[40]. Furthermore, they alleviate atherosclerosis and are under investigation for neurogenerative diseases like Alzheimer's disease, although their mechanism of action has not been elucidated yet [25], [41].

BPs are pyrophosphate (PP$_i$) analogues, in which the oxygen atom of the central P–O–P bond has been replaced by a carbon atom resulting in a P–C–P group which resists a potential enzymatic hydrolysis[42]. The proposed mechanism of action consists of mimicking the intermediate transition state. Replacing one or both hydrogen atoms linked

to the central carbon atom with various substituents (R_1 and R_2) resulted in different classes of BPs that are commonly classified as first, second and third generation.
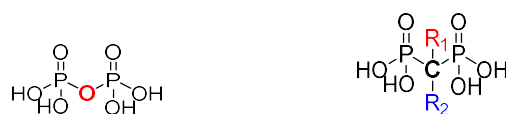
Figure 1.15. PPi molecule (left) and on basic structure of analogous BPs (right). The central oxygen atom of PPi (red) was substituted by a carbon atom (black) carrying two substituents R_1 (red) and R_2 (blue) in the BPs structure. R_1 is usually an OH group while R_2 ranges from CL to more complex nitrogen-containing groups. When R_2 has one or more nitrogen atoms, the BPs are named N-BPs.

This classification is based on their chemical structure, pharmacodynamic properties, mechanism of action as well as potency, activity and toxicity[42]. In terms of their chemical structure, R_1 is usually a hydroxyl (OH) group which has a high affinity for calcium phosphate crystals, and is thus involved in the tight binding with the calcium present in the bone mineral. R_1 can also be a hydrogen (H) or a chlorine (Cl) atom. The R_2 ranges from Cl to more complex nitrogen (N)-containing organic sidechains and is responsible for the potency of the different classes of BPs [42], [43].

BPs of the first generation such as Clodronate and Etidronate are characterized by a short alkyl chain in R_2. Neither do they contain nitrogen atoms, nor do they inhibit directly the FPPS. In fact, it was suggested that they are metabolized into non-hydrolysable ATP analogs which are accumulated in osteoclasts or macrophages thus affecting their function[44].

The nitrogen-containing bisphosphonates (N-BPs) are classified based on their R_2 in aliphatic or cyclic and aromatic compounds. The aliphatic compounds contain a terminal amino group and they belong to the second generation BPs like Pamidronate and Alendronate. The third generation features cyclic or aromatic side chain, e.g. Risedronate, Ibandronate and Zoledronate and are the most potent class of BPs and are therefore the first-choice treatment for all kind of bone-related disease. The suggested mechanism of action of N-BPs includes their deposition on the bone surface in the resorption *lacunae*

followed by their incorporation into osteoclasts which leads to their inactivation and apoptosis as well as the activation of the osteoblasts with an overall positive "bone balance" and increase in bone mass[42]. N-BPs directly inhibit FPPS and thus block the prenylation processes of small GTPases proteins which are essential for the osteoclasts function[35].
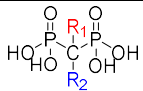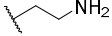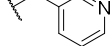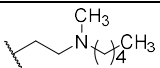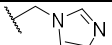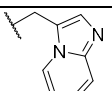
After oral administration, BPs are absorbed in the intestinal tract to about 1-10 % while the remaining 90-99 % pass into the bloodstream, where they have a half-life of 1-15 hours. Then they are conveyed into the kidneys and excreted in the urine. Only 20-50 % of absorbed BPs reach the bone tissue while the remaining part is also excreted. Due to the high affinity of BPs for bone tissue, their half-life is much longer than the one in the blood stream. This is the reason for their therapeutic effect and the need of daily or weekly oral administration for several years. The discovery of third-generation BPs has improved the quality of life of patients with regard to both, the side effects caused by intestinal absorption and the route of administration. Nowadays, BPs are intra-venously administered once every three months or even annually. In addition, it was demonstrated that they have a superior therapeutic effect compared to oral administration[42].

In addition to their effects on bone tissue and on the above-mentioned implication in heart and neurodegenerative diseases, BPs have also an antiparasitic activity *in vitro* and *in vivo* against *T. cruzi* and *T. brucei* FPPS[35], [45], [46]. In fact, it was demonstrated that FPPS is essential for parasite viability and therefore it is a validated molecular target for drug development[34].

In 2006 *Gabelli et al.* obtained the crystal structures of both Alendronate (ALE) and Risedronate (RIS) in complex with the *T. cruzi* FPPS[33]. Three years later, *Huang et al.* investigated other N-BPs carrying in R_2 position an imidazole (Zoledronate, ZOL or 91B), indole (Minodronate or 261A) or pyridine moieties with *meta* substituted fluorine (461A), alkyl chains (476A) or aromatic rings (300B)[45]. In 2012, *Aripirala et al.* investigated linear 2-alkylaminoethyl-1,1-BPs which exhibited IC_{50} values in the nanomolar range against *T. cruzi* proliferation[47]. In 2006 *Mao et al.* performed studies on *T. brucei* obtaining the crystal structures of FPPS in complex with ALE and RIS as well as with several other

compounds substituted in R_2 position with pyridinium, imidazolium, alkylammonium and amidinium moieties and investigated also their protonation state by solid-state ^{13}C and ^{15}N NMR[46]. Comparison with *human* FPPS showed a similar binding mode with either *T. cruzi.* or *T. brucei* species[46]. Two years later, in the same working group, a potent analog of Pamidronate, with R_2 = phenylalkyl (BPH-210) was investigated which showed low micromolar activity against T. *brucei* FPPS as well as cell growth inhibition[48]. In 2009, *Zhang et al.* discovered dual inhibitors of FPPS and GPPS belonging to the pyridinium BPs class (BPH-714, BPH-721 and BPH-722) which inhibit in *human* FPPS tumor cell growth both *in vitro* and *in vivo*, but also bind in a similar way to the *T. brucei* FPPS[39]. Furthermore, an analogous compound of RIS was synthesized by *Liu et al.*, showing an IC_{50} = 230 nM which is better than the value calculated in a cell growth inhibit for RIS (IC_{50} = 300nM)[49]. In 2015, *Yang et al.* screened a library of 925 potential *T. brucei* inhibitors which are mainly analogs of ZOL, resulting in a low micromolar activity both *in vitro* and *in vivo* but showing no substantial correlation between FPPS inhibition and *T. brucei* cell growth inhibition[50].

Table 1.1. Commonly used N-BPs.

	R_1	R_2
Etidronate	—OH	—CH$_3$
Clodronate	—Cl	—Cl
Pamidronate	—OH	—NH$_2$
Alendronate	—OH	—NH$_2$
Risedronate	—OH	(pyridinyl)
Ibandronate	—OH	(N(CH$_3$)(CH$_2$)$_4$CH$_3$)
Zoledronate	—OH	(imidazolyl)
Minodronate	—OH	(imidazopyridinyl)

300B		
461A		
BPH-210		
476A		
BPH-721		
BPH-722		
BPH-714		

Compared to the conventional drugs which usually have a certain degree of lipophilicity, bisphosphonates are highly charged and water-soluble molecules. They can enter cells following a pinocytotic mechanism such as osteoclasts and macrophages. However, due to their low lipophilicity and their previously mentioned pharmacokinetic problems, N-BPs are not able to penetrate into tumor cells and other non-endocytic cells. Several attempts to increase their lipophilicity by appropriate R_2 substituents, reduced affinity for bone tissue and replacement of the bisphosphonate group culminated in poorly active drugs. The inhibition mechanism of N-BPs on FPPS is singular. Most enzymes usually follow a competitive inhibition mechanism, i.e. upon inhibitor-enzyme (I-E) complex formation, the increased concentrations and accumulation of natural substrates enable displacement of the inhibitor, thereby leading to the recovery of enzymatic catalytic activity. In contrast, once the N-BP binds to FPPS in its open conformation by mimicking the DMAPP, it adapts a partially closed conformation and allows an IPP molecule to bind to the homoallylic site. Upon IPP binding, the enzyme assumes a fully closed conformation that is further stabilized by IPP. The N-BP-IPP-FPPS complex is extremely stable and does not allow accumulated substrates to displace N-BPs due to the closed catalytic pockets and thus inhibits catalytic activity [37].

1.3.3.2.   Overcoming the BPs pharmacokinetic limitations by an allosteric pocket

Since the bisphosphonate motif seemed to be essential for FPPS inhibition, the discovery of a new allosteric pocket pioneered the design of novel non-bisphosphonate drugs having a completely new chemical scaffold. This discovery dates back to 2010 when *Jahnke et al.* identified a new pocket by fragment-based drug discovery using both NMR and X-ray crystallography. Such a pocket was defined as allosteric based on its preference for lipophilic ligands with a negatively charged substituent which would mimic its downstream metabolites although its role or function had not been elucidated.

In the same work, *Jahnke et al.* developed two compounds (**11** and **12**, Figure 1.16) which are the first reported non-bisphosphonate FPPS inhibitors with IC_{50} values of 200 and 80 nM respectively, values which are similar to those of the most potent N-BPs, the ZOL (IC_{50}= 130 nM). This newly discovered pocket is located near the C terminus of the enzyme, adjacent to the IPP binding site and it is enclosed by the helices C, G, H, the C-terminal of helix J and comprises few residues of the B-C and H-I loops, the N-terminus of helix A and also the C-terminus tail. In particular, the base of the pocket is rather hydrophobic and surrounded by Phe206, Phe239, Leu344, Ile348 and Tyr10 side chains, while the upper part presents both positively charged (Lys57, Arg60 and Lys347) and polar (Asn59 and Thr63) residues [51].
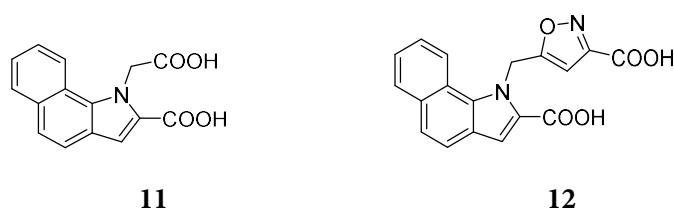
Figure 1.16. Compounds **11** and **12** are potent allosteric non-bisphosphonate inhibitors of human FPPS.
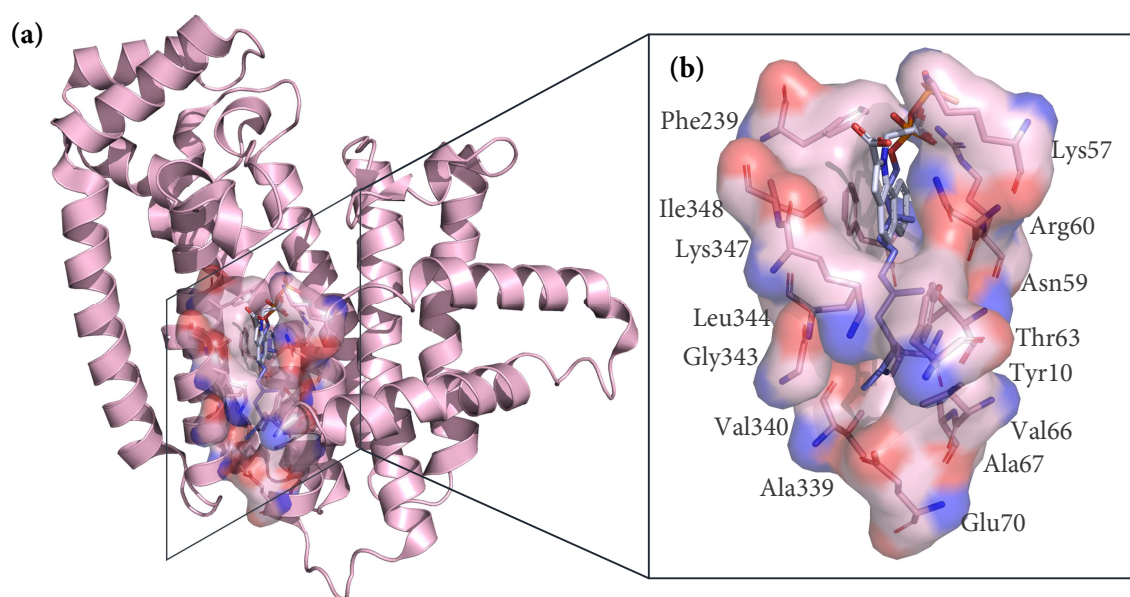
Figure 1.17. (a) Cartoon representation of *human* FPPS. Compounds **11** (grey stick) and **FPP** (blue stick) binding in the allosteric pocket. **(b)** Close-up view of allosteric pocket of *human* FPPS represented as a stick model and surface.

However, the evidence that this new pocket is really of allosteric nature only occurred in 2017 when *Park et al.* discovered that FPPS is inhibited by FPP which is its physiological product and binds to the same pocket identified by *Jahnke et al.* a few years earlier[52]. In fact, allostery is the process by which proteins transmit the effect of binding at one site to another functional site, which is often distant from the first one but allows for the regulation of the protein's activity[53]. The authors determined the binding mode of the FPP-FPPS complex by X-ray crystallography as well as the thermodynamic parameters by ITC both in presence and in absence of Mg^{2+} ions. Interestingly, the binding of DMAPP and GPP strongly depends on the presence of Mg^{2+} ions while FPP binds FPPS even in the absence of the cofactor with a K_d of 5–6 μM. It was then suggested that, under physiological conditions (millimolar concentrations of Mg^{2+} ions), DMAPP and GPP bind preferentially to the active site, and FPP to the allosteric site[52].

As described in Section 1.3.2., in the unbound state, FPPS adopts an open conformation which is, after GPP formation, ready to bind a new DMAPP molecule and starts another catalytic cycle. However, when FPP binds to its allosteric pocket first, the enzyme is locked in its open conformation and unable to proceed to the next catalytic cycle, thus promoting

the enzyme's inactive state. The FPP binding induces a protein conformational change (especially around Tyr10) that expands the allosteric pocket and reshapes its surface to accommodate the long hydrocarbon tail of FPP[53].

It was then generally accepted to use the term "allosteric" to characterize this new pocket since FPP binding induces an enzyme conformational change which interferes with DMAPP binding at the distal allylic site[52].

This mechanism of enzymatic inhibition by its own products allows enzymes to accomplish a rapid regulatory feedback process (compared to feedback from downstream metabolites) without compromising their catalytic efficiency but only diminishing the product release and/or backward reaction[52].

1.4.   Aim of the Project

N-BPs are active-site binders and are currently used in clinical practice for the treatment of bone related diseases in humans, but they also seem to have great potential as anticancer drugs and in neurodegenerative diseases [35],[39],[40],[41],[52]. However, due to their low pharmacokinetic properties, there is a constant search for new chemical entities with new scaffolds and/or different mechanisms of action.

The discovery by *Jahnke et al.* regarding a new allosteric pocket and powerful non-bisphosphonate inhibitors in *human* FPPS allowed the development of novel classes of compounds which have a different mechanism of action, i.e. not mimicking natural substrates but regulating the catalytic activity by binding in the allosteric pocket [51].

Since the FPPS structure presents highly conserved domains among the different species, it was assumed that a similar allosteric pocket would also be present in the Trypanosoma parasites. This new pocket was therefore investigated in the present study. The main aim of the project was to use a crystallographic fragment-based lead discovery (FBLD) approach to screen the general purpose in-house 96-fragment library (see Part I) to map and

investigate FPPS species, eventually focusing on putative allosteric fragment hits for the subsequent optimization into lead compounds. For the matter of time, the crystallographic screening was mainly focused on *T. cruzi* FPPS.

A precondition for a successful crystallographic screening is a pure, stable and monodisperse sample. In addition, protein crystals need to tolerate a sufficiently high DMSO (or other organic solvents to enhance solubility) and fragment concentration in the soaking conditions. The protein sample stability and buffer constituents therefore needed to be investigated.

Furthermore, the role of magnesium ions as cofactor of the enzyme as well as its binding mode in the unbounded crystal structure are not well elucidated. In fact, several PDB structures were deposited with FPPS in complex with BPs and non-BP compounds in the presence of magnesium ions as cofactor. However, no crystal structure of FPPS binding magnesium ions alone has been deposited.

Before embarking on a fragment screening project, it is therefore essential to have a stable protein material and a clear understanding of the structure of the holoenzyme in order to compare it with potentially bound fragments.

2. The role of Magnesium and pH in protein activity and stability

2.1   The importance of the enzyme's cofactor

As mentioned in Section 1.3.2., FPPS is an enzyme that needs a cofactor to perform its catalytic activity. Cofactors are portion of a protein not assembled from amino acids which can be divided into inorganic ions or organic complexes which are also called coenzymes [54]. The role of cofactors is to help and assist enzymes in their chemical reactions. They are catalysts, i.e. molecules capable of accelerating the rate at which chemical reactions occur[55]. In the absence of the cofactor, enzymes are usually inactive and incapable of carrying out any biochemical transformation. In this state they are called apoenzymes while the enzyme together with its cofactor is instead called holoenzyme [55]. In the present study the holoenzyme is the FPPS- Mg^{2+} ion complex. Magnesium is the most abundant metal ion in living cells, followed by Zn (Zn^{2+}), Ca (Ca^{2+}), Fe (Fe^{2+}), Mn (Mn^{2+}), Cd (Cd^{2+}), Cu (Cu^{2+}), Co (Co^{2+}) and Ni (Ni^{2+}) [56],[57]. It catalyzes over 300 enzymatic reactions and is involved in many metabolic processes such as glycolysis, *Krebs* cycle, β-oxidation, helps to create, repair and protect DNA, and regulates the concentration and the transport across cell membranes of important ions such as sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}). Furthermore, it is involved in creating cellular energy. ATP is indeed often called Mg-ATP since it requires Mg^{2+} ions to be biologically active[58],[59]. In the case of FPPS, it is important for the catalyzed chain elongation reaction. It is known from literature that FPPS requires a divalent metal ion for its activity. However, whereas Mn^{2+} could replace the Mg^{2+} ions although the enzyme activity is reduced by about 1/3, Zn^{2+} and Ca^{2+} ions completely inhibit FPPS activity [34]. As a result, they cannot replace Mg^{2+} in the same way as Mn^{2+} ions do, therefore this effect needs to be taken into account. In addition, Mg^{2+} ions bind weakly to protein and enzymes with a commonly found $K_a \leq 10^5$ M^{-1} and its binding site is less specific than that of other divalent ions such as Zn^{2+} and Ca^{2+}. This can be explained by the very high Mg^{2+} ion concentration (0.5-1mM) inside the cell which is

reflected by a less conserved composition of magnesium binding sites compared to those of other ions[57].

Cofactors play therefore an essential role in enzyme function and their removal affects enzyme activity and stability.

In fact, as will be discussed in more detail in Section 3.2., *T. cruzi* FPPS crystals were initially very unstable. They grew within two to three days but afterwards they rapidly changed shape, formed cracks and resolution decreased fast. After a week, the crystals decayed even further by gradually losing their regularly shape and they did not show X-ray diffraction. The few crystals with regular shape, which did not indicate changes of their habitus, did not diffract any longer. In addition, after its purification and freezing in liquid nitrogen for storage, the protein sample had to be used as soon as possible. In fact, even after a few months, the thawed protein did no longer crystallize at all. It became therefore necessary to optimize the buffer conditions to enhance FPPS stability.

It was assumed that the protein crystals were not stable because the protein solution in which they were stored was not optimal. In order to optimize all the protein species under investigation, a systematic approach was used which included a thermal shift assay (TSA) and light scattering (LS).

2.2   The thermal shift assay (TSA)

Thermal Shift Assay (TSA), also known as Differential Scanning Fluorimetry (DSF) is a valid method to investigate protein stability and to screen several buffers in order to identify under which condition the protein is increasingly stable. In fact, it has been demonstrated that the overall quality of the protein sample is strongly correlated with stability and homogeneity which increases structural order, decreases protein flexibility, lowers the tendency of aggregation and improves the success of crystallization [60].

In a typical TSA experiment the protein solubilized in the buffers under investigation is mixed with a hydrophobic fluorescence dye (SYPRO® Orange) which does not fluoresce in aqueous solution but does so in non-polar environments. In the ideal case, no fluorescence is detected at room temperature because the protein is correctly folded, and no hydrophobic areas are exposed to the environment. As the temperature increases, the protein starts to unfold or "melt" and exposes its hydrophobic parts on the outer surface. As a result, the dye binds to these hydrophobic areas causing an increase in fluorescence which is detected by the instrument. The system is then able to estimate the melting temperature T_m from the inflection point of the melting curve which corresponds to the temperature at which hydrophobic areas of the protein become exposed and the protein starts unfolding [60]. A shift of the melting curve to higher temperature indicates an increase in T_m and therefore in protein stability, whereas a shift to lower temperatures indicates destabilization. More details of the present method are provided in the Material and Methods section (Section 6.2.).

In order to find the most promising buffer composition a systematic approach was used. Only one parameter was changed at a time, including Tris, Hepes, salt (NaCl) and Mg^{2+} ions concentration in the form of $MgCl_2$. In particular, Tris and Hepes concentrations were set for the whole screening at 25 mM because this is the usual concentration reported for FPPS buffer in the literature.

The screening mainly focused on the effect of pH and Mg^{2+} ions concentration on protein stability. Each buffer was thus evaluated at different pH values, i.e. 6.5, 7.0, 7.5 and 8.0, and the $MgCl_2$ concentration was checked at 2, 3 and 5 mM as well as in absence of Mg^{2+} ions. In contrast, the NaCl concentration was screened in bigger steps, i.e. 25, 50, 75, 100, 150 and 200 mM.

As a result, it was possible to screen 96 different conditions each for Tris and Hepes buffer according to the Scheme shown in the Figure 2.1.

		1	2	3	4	5	6	7	8	9	10	11	12
	[MgCl₂] mM	0	2	3	5	0	2	3	5	0	2	3	5
A	pH: 6.5												
B	pH: 7.0			25mM NaCl				50mM NaCl				75mM NaCl	
C	pH: 7.5												
D	pH: 8.0												
E	pH: 6.5												
F	pH: 7.0			100mM NaCl				150mM NaCl				200mM NaCl	
G	pH: 7.5												
H	pH: 8.0												

Figure 2.1. 96-well plate set-up for TSA screening. The plate was divided into six matrices each of which had a different salt concentration ranging from 25 to 200 mM NaCl. Each matrix was further divided into four rows corresponding to different pH values (6.5, 7.0, 7.5 and 8.0) and four columns corresponding to different MgCl₂ concentrations. Since the buffer component (Tris or Hepes) did not differ, each well had a distinct combination of pH, NaCl and MgCl₂ concentration.

It is interesting to note that protein stability increases as MgCl₂ concentration increases and pH decreases in almost all cases. Furthermore, the samples are more stable in Tris compared to Hepes buffer and at salt concentrations between 25 and 75 mM NaCl.

In Figure 2.2 the melting curves are shown at one salt concentration for each protein species under investigation which were 25 mM NaCl for *T. brucei* and 75m M for *T. cruzi* and human FPPS (Figure 2.2.). Those salt concentrations revealed the highest stability. In fact, at higher salt concentrations (100, 150 and 200 mM NaCl), a shift of the curves to lower temperature was observed. As shown in Figure 2.2 (right), the melting curves referring to the buffers at pH 6.5 (orange) and 7.0 (green) show a significant shift to higher temperature compared to those at pH 8.0 (blue), resulting in a increase of T_m by at least two degrees. However, there was no big difference between pH 6.5, 7.0 and 7.5 as the T_m for these conditions differ only by about half a degree.

On the other hand, a comparison of the melting curves against different MgCl₂ concentrations (Figure2.2., left), showed that a comparably low concentration (2 mM, yellow melting curve) is already sufficient to stabilize the protein causing an increase in T_m compared to the buffers in the absence of Mg^{2+} ions (grey). In addition, the melting curve is further shifted to the right at 3 mM (orange) and 5 mM (dark red) MgCl₂.

Melting Curve Plot/Derivative Plot
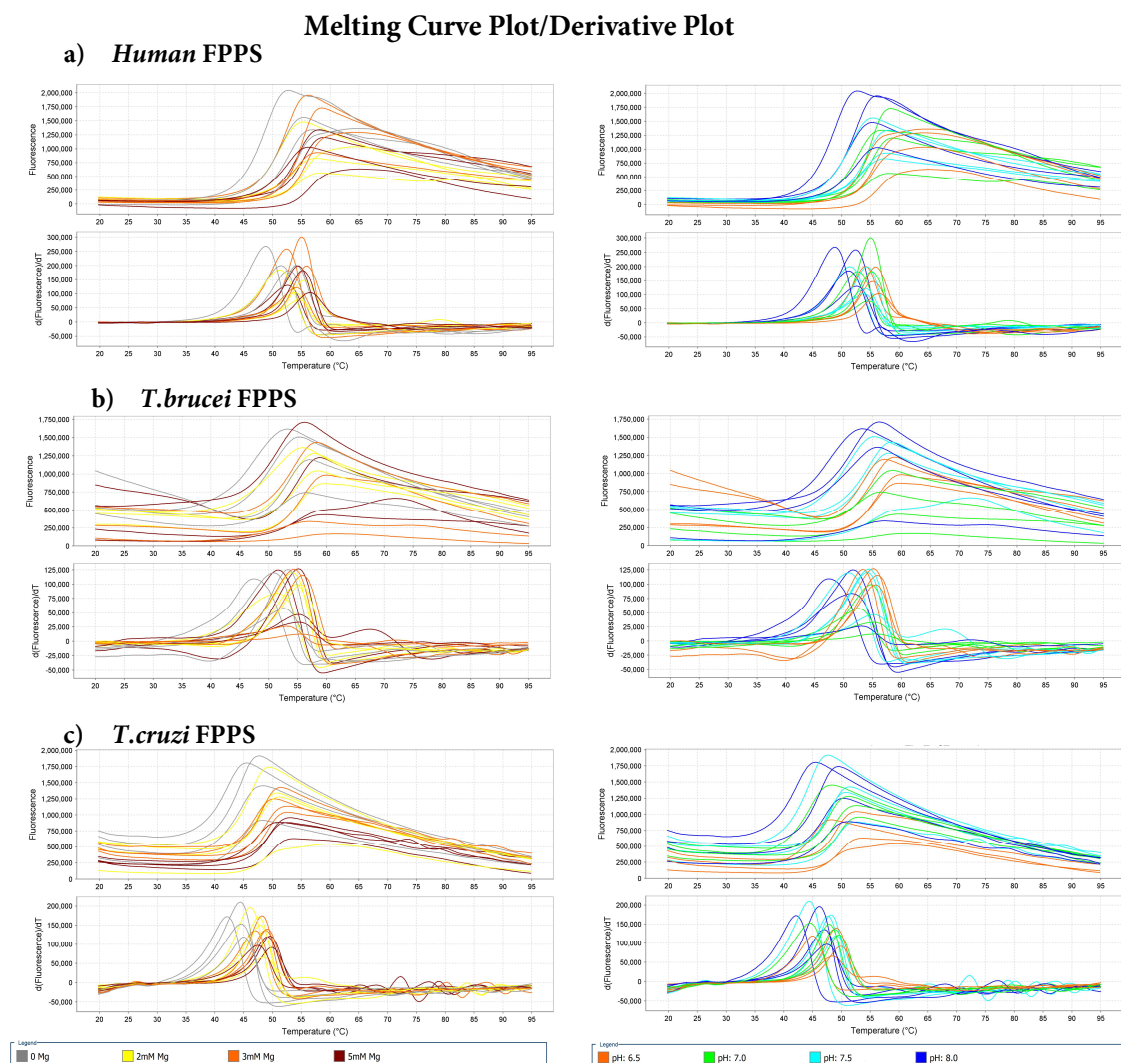
Figure 2.2. Melting curve and derivative plot of (a) *human* (b)*T. brucei* and (c) *T. cruzi* FPPS at a 75, 25 and 75 mM NaCl, respectively. On the left, comparison of buffer components based on magnesium concentration and on the right under different pH value. Interestingly, the protein stability increases by increasing magnesium concentration and lowering the pH.

A comparison of the stability of the protein species under the optimized buffer conditions along with the initial storage conditions shows a distinct increase in T_m from 4 °C to 9 °C (Figure 2.3). It is worth noting that initially all three proteins were stored in the same buffer consisting of Tris pH 8.0 and 200 mM NaCl. As shown in Figure 2.3, the T_m of *human* FPPS increased from 48.7 °C to 55.04 °C with a Δ_{Tm} of 6.34 °C, T_m of T. *brucei* from 46.30 °C to 55.92 °C (Δ_{Tm}=9.62) and T_m of *T. cruzi* from 45.27 °C to 49.94 °C (Δ_{Tm}=4.67 °C).

Melting Curve Plot/Derivative Plot
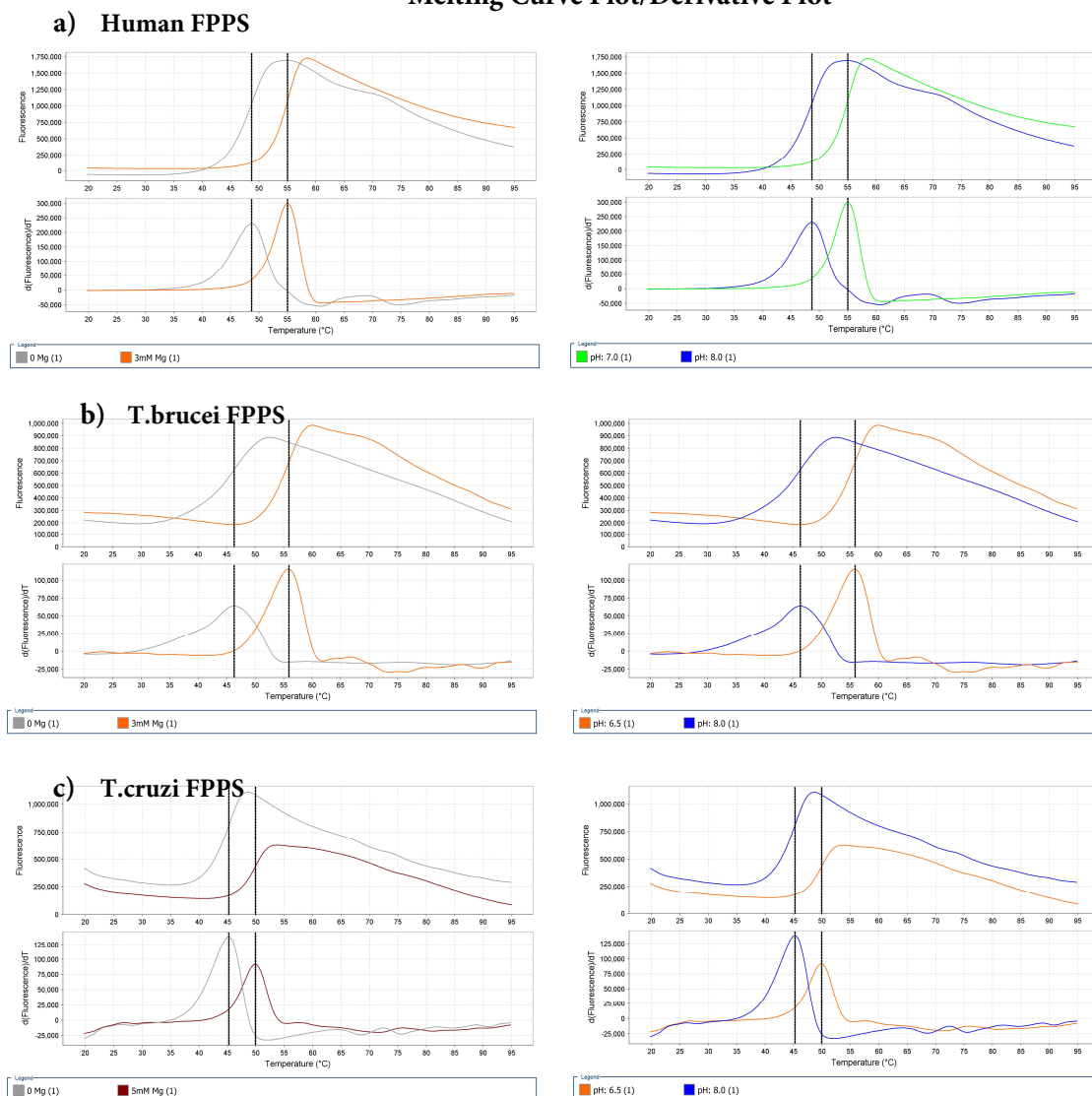
Figure 2.3. Comparison of melting curve and derivative plot of **(a)** *human* **(b)** *T. brucei* and **(c)** *T. cruzi* FPPS in the original storage buffer and in the optimized one. The investigated species were stabilized by 6.34 (a), 9.62 (b) and 4.67 (c) degrees, respectively.

The original storage buffer was then replaced by the optimized, distinctly different conditions, according to TSA results. *Human* FPPS was then stored in 25 mM Tris, pH 7.0, 75 mM NaCl, 3 mM and 2 mM TCEP while *T. brucei* FPPS in 25 mM Tris, pH 6.5, 25 mM NaCl, 3 mM $MgCl_2$ and 2 mM TCEP and *T.cruzi* FPPS in 25 mM Tris, pH 6.5, 75 mM NaCl, 5 mM $MgCl_2$ and 2 mM TCEP.

143

These selected buffer conditions were further evaluated by Light Scattering (LS) to assess their polydispersity. In fact, a protein can be stable in a certain buffer, but the composition of the sample is not necessarily monodisperse. However, a pure, stable and monodisperse sample is essential for a successful crystallization.

2.3   SEC-MALS experiments

Size-exclusion chromatography (SEC), coupled with molecular weight-sensitive detectors, such as refractive index (RI) and multiangle light-scattering (MALS) detectors, is the most widely used technique for determining and studying the solution properties of protein samples. MALS detectors measure the molar mass (weight average) based on the radius of gyration (R_g) and they are also able to detect and quantify the population of various oligomeric states in the protein sample[61]. Adding a dynamic-light scattering (DLS) detector allows for rapid identification of aggregates and also measures the protein size based on the hydrodynamic radius (R_h). This combination of molecular weight and dimensions is used to obtain important information about size, shape and conformation of a protein. In addition, refractometers are used to measure the concentration of any macromolecule regardless of chromophores. Coupling to a UV instrument enables the differentiation of a protein sample from a non-protein sample in the case of small impurities. More details of the present method are provided in the Material and Methods section (Section 6.3.).

The advantage of such devices is that all these parameters are measured simultaneously in the same flow cell and measurement volume.

As shown in Figure 2.4, all the protein samples show similar behavior. They are homogeneous and monodispersed as they have a polydispersity (Molecular weight (Mw)/number average Mw (Mn)) equal to 1 and present no aggregates. They are dimers in solution with $Mw = Mn = 7.777 \times 10^4$ g/mol for the *human* and 8.371×10^4 g/mol and 8.228×10^4 g/mol for *T. brucei* and *T. cruzi* FPPS, respectively. The R_h is also fairly similar

among the species, i.e. 3.747 nm for the *human*, 3.706 nm for the *T. brucei* and 3.675 nm for the *T. cruzi* FPPS.

This demonstrates that the protein samples are both stable and monodisperse in the TSA-optimized buffers and can be stored under these conditions.
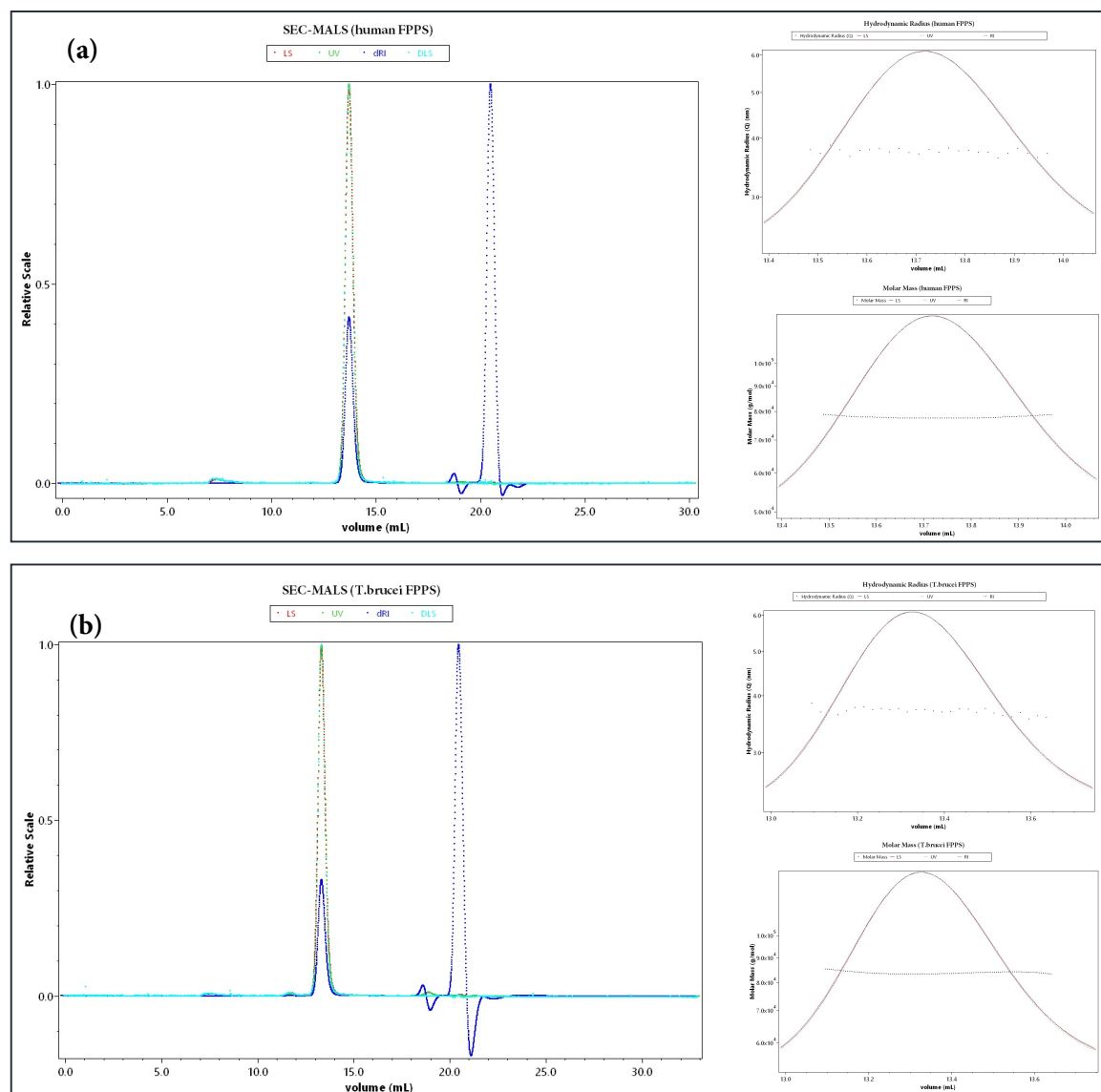
(continued on next page)
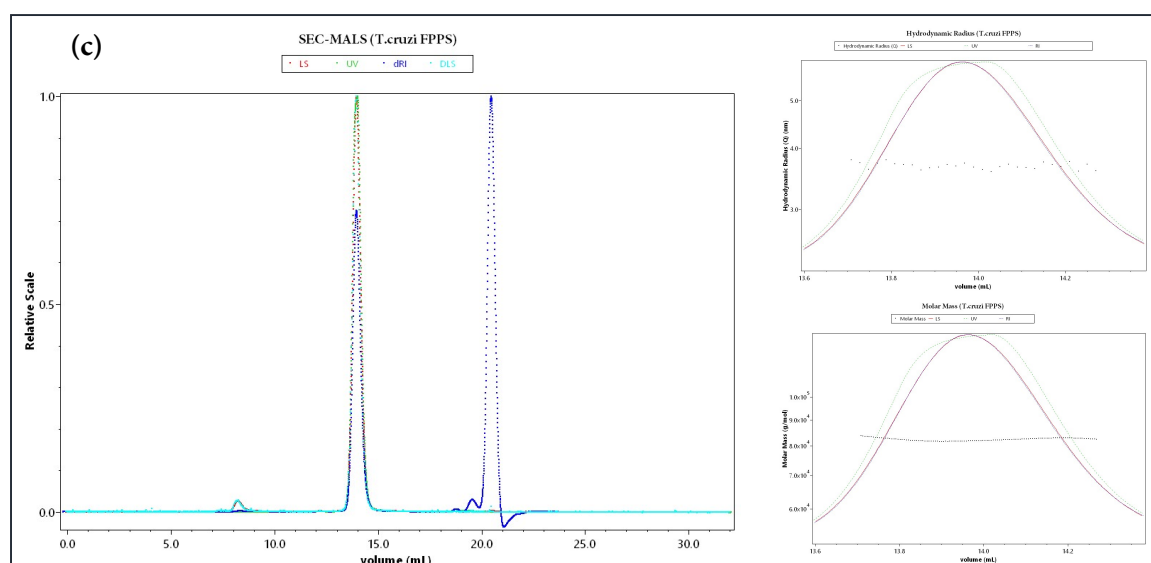
Figure 2.4. On the left side of each box, a comparison of SEC-MALS experiments of **(a)** *human* **(b)** *T. brucei* and **(c)** *T. cruzi* FPPS in the optimized TSA buffers. All species under investigation show a similar behavior. They all present a sharp peak around 14 mL of elution volume in which the curves referring to LS, UV, dRI and DLS perfectly overlap. The peak on the right (20 mL) represents a salt pick which is always present in SEC-MALS experiments. The small peak on the left (8 mL) is an impurity probably due to small dust particles in the samples, demonstrating the sensitivity of the method. This latter peak does not refer to protein aggregates as there is no UV absorption. On the right side of each box, *Gaussian* distribution based on the hydrodynamic radius (upper panel) and molar mass (lower panel) of **(a)** *human* **(b)** *T. brucei* and **(c)** *T. cruzi* FPPS in the TSA-optimized buffers. The horizontal dashed black lines indicate that the samples are homogeneous over the entire elution volume. Enlarged graphs are reported in Appendix 7.1.2.

In addition, the role played by magnesium ions in protein stability and conformational change was evaluated by MD simulations.

2.4   MD Simulations

Molecular dynamics (MD) simulations allows to study complex dynamic processes that occur in biological systems but that are difficult to examine experimentally such as protein conformational changes including the transformation mechanism from an opened to a closed state. This method also permits to differentiate the stability of conformations in terms of energetically favorable or unfavorable. The first step to compare two or more systems is record changes along the MD trajectory in term of mechanical descriptors. In

the case of FPPS these descriptors allowed to describe the collective intermolecular interactions of the enzyme in the bound/unbound state.

In fact, as discussed in Section 1., the catalytic mechanism of FPPS involves a conformational transformation between open, partially closed and fully closed state. In the unbound state, the enzyme adopts an open conformation while the binding of IPP and DMAPP or a BPs entails a fully closed conformation. In particular, the distance between the structural FARM and SARM motifs was used as mechanical descriptor for evaluating the enzyme's transformation pathway. The open conformation refers to a FARM-SARM distance greater than 20 Å, the partially closed to a distance between 19-20 Å whereas the fully closed conformation is characterized by distances below 19 Å.

The protein behavior in presence or absence of Mg^{2+} ions, evaluated by MD simulations (over 500 nanoseconds), were performed for all species under investigation. As reported in Figure 2.5., *human*, *T. cruzi* and *T. brucei* FPPS show similar trends. In presence of Mg^{2+} ions, *T. brucei* FPPS has a FARM-SARM distance distribution of 20.5 % in the closed (Figure 2.5, red area **(a)**), 25.4 % in the partially closed (Figure 2.5, orange area **(a)**) and 53.96 % in the open conformation (Figure 2.5, green area **(a)**) while in absence of Mg^{2+} ions the distance distributions come to 10.18 % in closed (Figure 2.5, red area **(b)**), 11.39 % in partially closed (Figure 2.5, orange area **(b)**) and 78.42 % in open state (Figure 2.5, area **(b)**), respectively.

The *human* FPPS also showed a different behavior with and without Mg^{2+} ions. The open conformation was adapted of about 49 % of the time with Mg^{2+} ions (Figure 2.5, green area **(c)**) and 57 % of the time without Mg^{2+} ions (Figure 2.5, green area **(d)**) while the fully closed conformation was formed in 27 % (Figure 2.5, red area **(c)**)and 18 % (Figure 2.5, red area **(d)**) of the time along the trajectory, respectively.

Comparison of histograms of FARM-SARM distances with/without Mg²⁺
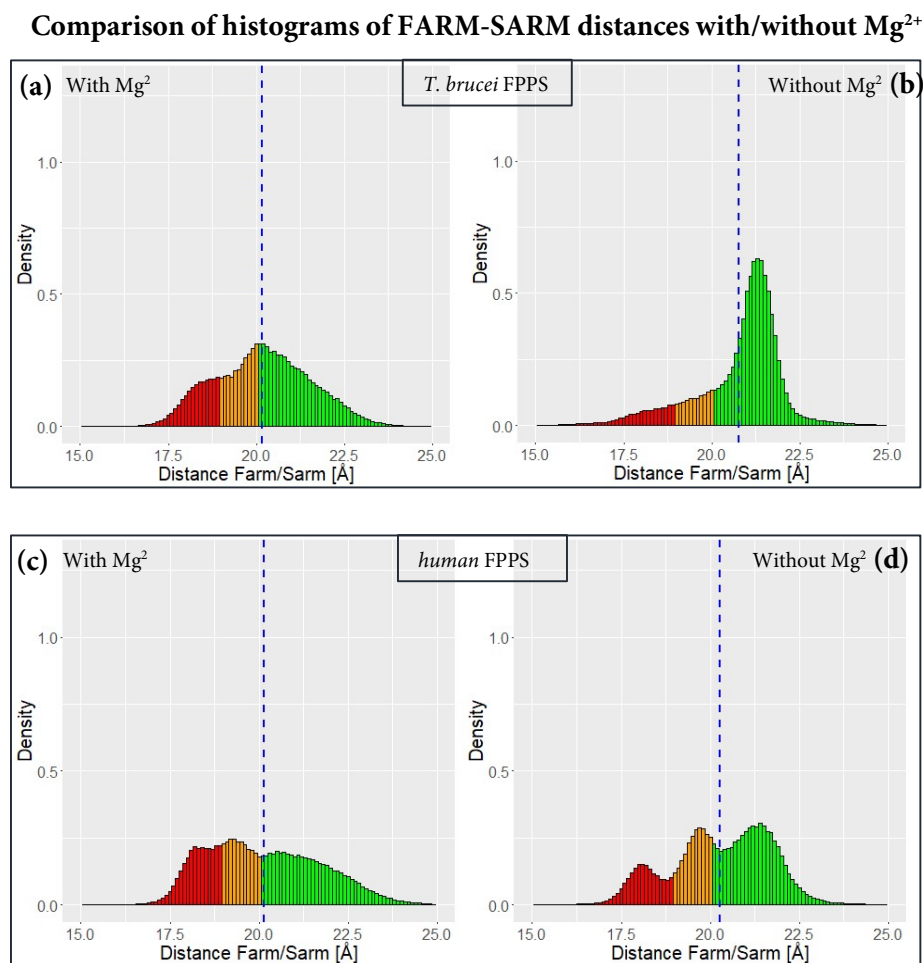
Figure 2.5. Comparison of histograms of FARM-SARM distances between **(a, b)** *T. brucei* and **(c, d)** *human* FPPS in presence **(a, c)** and absence **(b, d)** of Mg²⁺ ions. The red area refers to the fully closed, the orange to the partially closed and the green area to the open conformation of the enzyme.

In order to obtain a more conclusive insight into the simulation data, a principal component analysis (PCA) of the so-called variance-covariance matrix of the complete trajectories investigated was also performed. This matrix describes how much each atom moves during the trajectory and how much each atom moves in conjunction to other atoms. By using these data as input file, principal components have then been extracted, which correspond to complex movements of all atoms together for a given MD trajectory. Out of ten, two components explained nearly 60 % of the variance of the movements for a given MD trajectory which are indicated as "Mode1" and "Mode2". Table 2.1 shows how much each mode is represented in terms of percentage for both, *human* FPPS in presence

and in absence of Mg^{2+} ions. As reported in the first two columns, "Mode1" and "Mode2" can already explain 57 % of the complete movement of the protein in the MD trajectory. It is important to underline that the modes for *human* FPPS in presence and in absence of Mg^{2+} ions were computed for the respective trajectory. As a result, for example, *human* "Mode1" in presence of Mg^{2+} ions is different from *human* "Mode1" in absence of Mg^{2+} ions.

Table 2.1. Modes in the MD trajectory of *human* FPPS with and without Mg^{2+} ions represented in %.

Modes	1*	2*	3	4	5	6	7	8	9	10
human FPPS with Mg^{2+} (%)	**39.12**	**18.10**	10.28	8.42	7.25	4.87	3.49	3.29	2.86	2.33
human FPPS without Mg^{2+} (%)	**32.85**	**25.66**	13.92	7.22	5.45	4.36	3.22	2.90	2.50	1.93

*: "Mode1" and "Mode2" explain 57 % of the complete movement of the protein in the MD trajectory.

The visual representation of the two most represented movements, "Mode1" and "Mode2", are shown in Figure 2.6. Each atom of the protein backbone is represented as a vector. The direction of the arrows indicates the direction of the movements while the length of the arrows indicates how much each atom moves. Rigid parts of the protein backbone are shown in red while the more movement the more the color turns into blue (Figure 2.6.).

With Mg^{2+} Mode1   Without Mg^{2+}

39.12 %   32.85 %
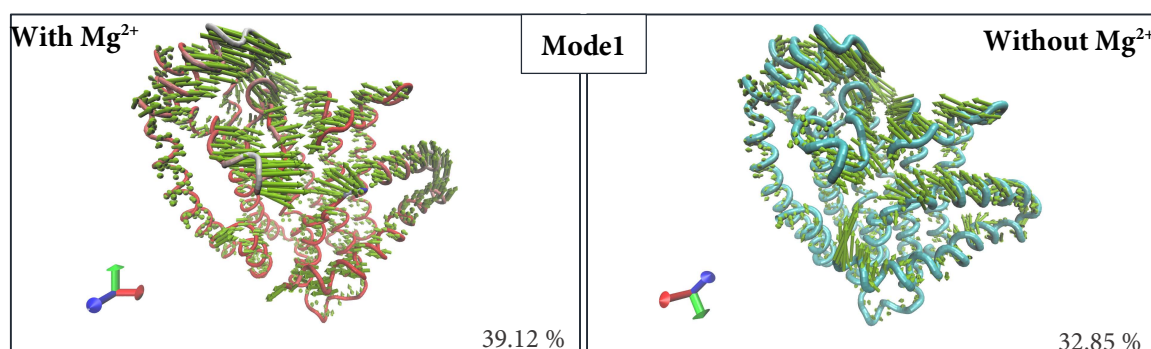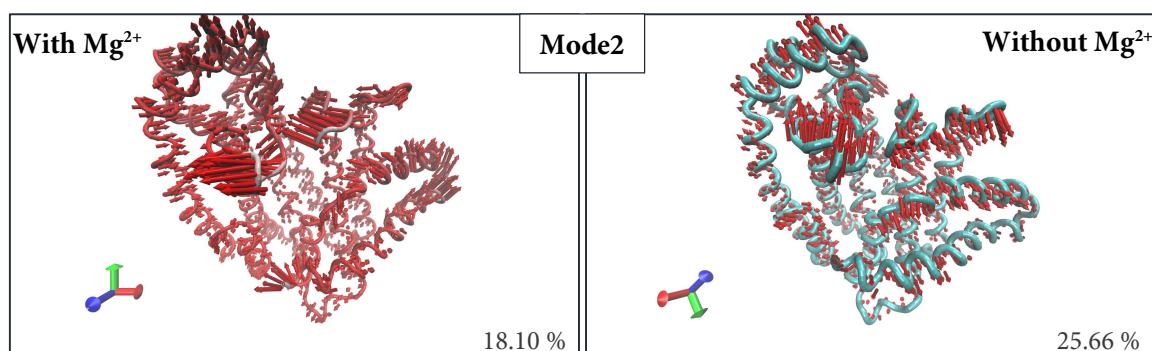
Figure 2.6. Visual representation of PCAs "Mode1" and "Mode2" of *human* FPPS with and without Mg^{2+} ions. The combination of "Mode1" and "Mode2" explain 57 % of the complete movement of the protein in the MD trajectory. The direction of the arrows indicates the direction of the movements while the length of the arrows indicates how much each atom of the protein backbone moves.

The correlation between "Mode1" and "Mode2" and how they interplay with each other were also investigated. In Figure 2.7., *human* FPPS "Mode1" is plotted against "Mode2". As shown in the 2d-histograms (Figure 2.7.), the combination between "Mode1" and "Mode2" samples different regions in their motions, with and without Mg^{2+} ions. These different motions are probably related to the open and closed states.

In presence of Mg^{2+} ions (Figure 2.7. **(a)**) two regions of high density are displayed. The extremum at larger Mode2 values represents the open state while the one at lower Mode2 values, next to the first one, represents the closed state. In absence of Mg^{2+} ions (Figure 2.7. **(b)),** a slightly different extremum at larger Mode2 values representing the open state is also displayed. However, the extremum at lower Mode2 values is missing in absence of Mg^{2+} ions. Additional spots are instead displayed in absence of Mg^{2+} ions which may refer to additional transition states (Figure 2.7. **(b),** dashed arrows).
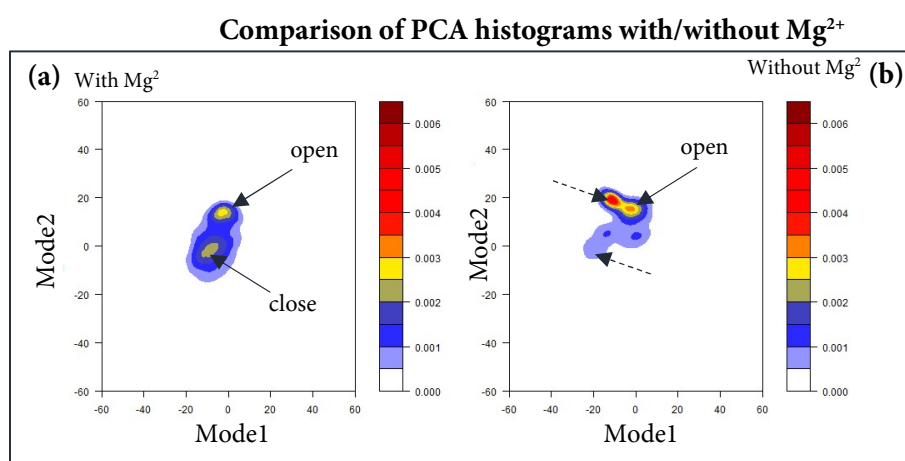
Figure 2.7. Comparison of PCA 2d-histograms of *human* FPPS with **(a)** and without **(b)** Mg^{2+} ions. In both histograms, there is an upper hot spot indicating the open state. In **(a)** there is also a spot indicating the closed state which is missing in **(b)**. The dashed arrows in **(b)** indicate two additional states which are present only in absence of Mg^{2+} ions.

Using two different way to evaluate the data, i.e. the histograms referring to FARM-SARM distance and the PCA analyzing the principal directions of molecular motion, the MD simulations suggested the same finding: in absence of Mg^{2+} ions the open conformation of the enzyme occurs more frequently compared to the situation when Mg^{2+} ions are present. This indicates that the partially closed and closed conformation are energetically less favorable in absence of Mg^{2+} ions.

In conclusion, using different techniques such as TSA, SEC-MALS and MD simulations, the same results were obtained regarding the importance of Mg^{2+} ions for the stability and the correct distribution of conformational states of the protein species under investigation.

Despite the importance of Mg^{2+} ions highlighted in Section 2.1 and confirmed in Section 2.2., 2.3. and 2.4., no crystal structure of the holoenzyme has yet been deposited in the PDB. Instead, *human* (2F7M), *T. cruzi* (1YHK) and *T. brucei* (4RYP) FPPS apo-structures in the presence of phosphate or sulphate ions picked up from the crystallization conditions were deposited. Furthermore, crystal structures hosting Mg $^{2+}$ ions have been determined only with BPs in complex with IPP or DMAPP, but no crystal structures are available with both natural substrates simultaneously binding to the allylic (DMAPP) and homoallylic (IPP)

sub-pockets. In addition, the deposited PDB structures usually contain three Mg^{2+} ions in the catalytic pocket and they are coordinated to the phosphate groups of the BPs. Surprisingly, also a crystal structure of *T. brucei* FPPS in complex with BPs and coordinating four Mg^{2+} ions (4RXC) has been deposited. Based on this finding, it was questioned, whether Mg^{2+} ions are commonly present in the catalytic pocket or only in the presence of bound BPs and additionally, whether the enzyme's topology is affected by the presence or absence of the Mg^{2+} ions cofactors.

With the replacement of the storage buffers and the development of a new soaking protocol, which will be discussed in section 3.2.3., the protein material crystallized even half a year after the purification and the protein crystals did no longer decay after a week.

This new procedure enabled the performance of a successful fragment screening project and resulted in the first holo-FPPS structure as well as the first structure with both the natural substrate occupying the allylic and homoallylic pocket.

2.5   The first holo-FPPS structure

As the Mg^{2+} ion (0.76 Å) has a similar diffraction power and an ionic radius only slightly smaller than that of a water molecule (1.38 Å), both species are difficult to distinguish solely based on the electron density. Actually, one possible way to identify Mg^{2+} is the detection of the typical octahedral geometry around Mg^{2+} ion coordinated by six water molecules at a distance of about 2.1 Å from the central metal ion [62]. An alternative and immediate way to recognize metal ions in X-ray crystal structure is to use the electron density properties of certain atoms to scatter X-rays differently depending on their atomic number Z and energy of the incident X-rays, thus generating an anomalous scattering. However, not all atoms have a strong enough anomalous signal. It is a typical property of heavier atoms and some transition metal such as Mn and Zn. Unfortunately, Mg does not have this property. In order to obtain a conclusive holo-FPPS structure, Mg^{2+} was therefore replaced by Mn^{2+} which exhibits the same charge and a similar atomic radius (0.86 Å) as Mg^{2+} but also

produces a fair anomalous signal. An apo-crystal of *T. cruzi* FPPS was soaked in 10 mM MnCl$_2$ obtaining a data set of 1.9 Å resolution from a single-wavelength anomalous dispersion (SAD) experiment. The anomalous signal revealed the presence of two Mn^{2+} ions in the catalytic site. One of them (Mn^{2+}a) has the typical octahedral coordination geometry while the other (Mn^{2+}b) is not strongly bound to the protein as it is involved in only one single interaction with the Asp102 side chain (2.3 Å) (Figure 2.8. **(a)** and **(b)**). Mn^{2+}a is octahedrally coordinated by the Asp250 side chain (2.08 Å) and completes the hexa-coordination with five water molecules W506 (2.20 Å), W579 (1.96 Å), W592 (2.12 Å), W600 (2.07 Å) and W648 (2.20 Å) (Figure 2.8. **(c)**). The Mn-O bond distances range from 1.96 Å to 2.20 Å , which are similar to that reported in literature [62].
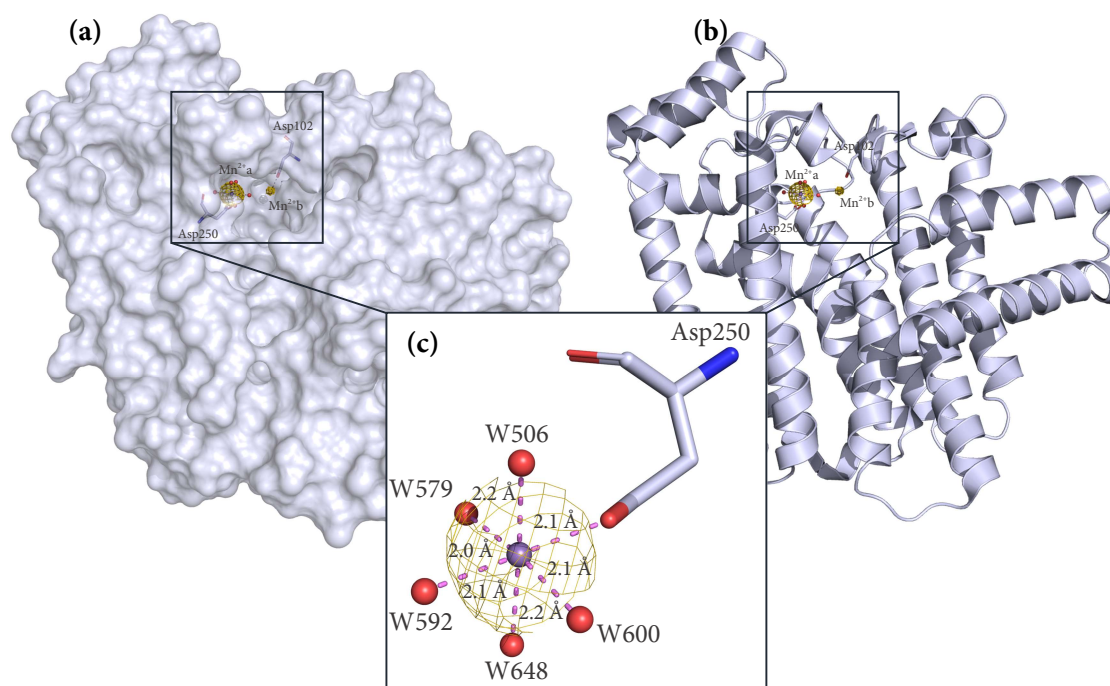
Figure 2.8. (a) Surface and **(b)** cartoon representation of the *T. cruzi* FPPS holo-structure (PDB 6SFA). The center of the two Mn^{2+} ions (Mn^{2+}a and Mn^{2+}b) is represented as a violet sphere while the anomalous electron density maps are depicted as a yellow mesh displayed at a contour level of 4σ. **(c)** Close-up view of the octahedral coordination of Mn^{2+}a with the Asp250 side chain and five water molecules.

2.6   The first FPPS structure with both allylic and homoallylic pockets occupied by the natural substrate IPP

Using the same approach as in the holo structure, i.e. replacing the Mg^{2+} ions with the Mn^{2+} ones, the first *T. cruzi* FPPS structure hosting the natural substrate in both, the allylic and homo-allylic sub-pockets was obtained. Initially, several attempts were made to soak DMAPP and IPP together, but the crystals immediately dissolved or formed cracks, probably due to the high reactivity of the two substrates. However, since the natural substrates differ only in their double bond position, which would not be distinguishable under sufficiently high resolution in the electron density, the IPP was then soaked into an FPPS crystal at a concentration able to occupy both, the homo-allylic and the allyl site. An apo-crystal of *T. cruzi* FPPS was then soaked with 10 mM IPP solubilized in $MnCl_2$. SAD experiments were carried out to identify either the number of Mn^{2+} ions and to assign the exact location of the IPP phosphate groups in case the resolution did not allow to fully interpret the electron density. In fact, phosphorus (P) as well as Mn generate an anomalous signal. Seven peaks were assigned to anomalous signals, three of which were attributed to the Mn^{2+} ions ($Mn^{2+}a'$, $Mn^{2+}b'$ and $Mn^{2+}c'$) and four to the P atoms that are part of the phosphate groups of the two IPP molecules (Figure 2.11. **(b)**). Unlike the holo-structure, three Mn^{2+} ions are therefore present in the tertiary complex FPPS-IPP-Mn^{2+}. This would suggest that the enzyme coordinates two Mn^{2+} ions in the catalytic site and an additional one is bound to the IPP (or BPs) and incorporated upon binding into the active site.
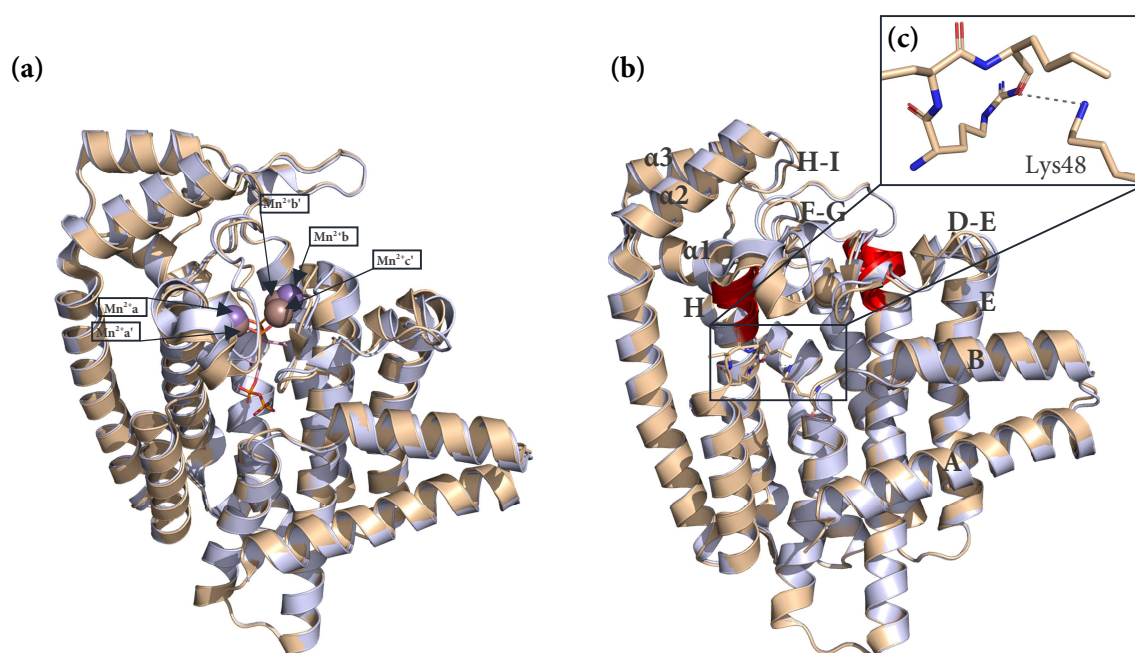
Figure 2.9. Superposition of the holo-structure in opened conformation (whiteblue, PDB 6SFA) and FPPS-IPP-Mn^{2+}complex in closed conformation (wheat, PDB 6T7N). **(a)** Mn^{2+} ions are represented as whiteblue (Mn^{2+}a, Mn^{2+}b) and wheat (Mn^{2+}a', Mn^{2+}b' and Mn^{2+}c') sphere. The two IPP molecules occupying the allylic and homoallylic subpockets are represented as pink stick. **(b)** Helices A, B, E, H, α1, α2 and α3 and loops D-E, F-G, H-I are involved in the conformational transformation. FARM and SARM motifs are depicted in red. **(c)** Close-up view of the C-terminal tail of the FPPS-IPP-Mn^{2+}complex which clamps down towards the homoallylic pocket is represented as stick model. The primary amine of the Lys48 side chain is involved in an electrostatic interaction with the C-terminal carboxylate group which further stabilizes the closed conformation.

The superposition of the holo-structure with the FPPS-IPP-Mn^{2+} complex reveals that the former is in an open while the latter in a closed conformation. In fact, the FARM and SARM motifs of the tertiary complex come closer together reducing the volume of the catalytic pockets. The conformational transformation also involves the helices A, B, E, H, α1, α2 and α3 and the loops D-E, F-G, H-I which move towards the catalytic site. It is interesting to note that in the holo structure it was not possible to build residues 360-362, which form the C-terminal tail, into the electron density, probably due to their high residual mobility. In the FPPS-IPP-Mn^{2+} complex instead, these residues clamp down towards the homoallylic pocket and the formed closed conformation is further stabilized by the electrostatic interactions between the terminal amino group of Lys48 and the C-terminal carboxylate group (Figure 2.9. **(c)**). Notably, Mn^{2+}a overlaps with Mn^{2+}a' whereas Mn^{2+}b only partially overlaps with Mn^{2+}b' while Mn^{2+}c' is located between Mn^{2+}a' and Mn^{2+}b' (Figure 2.9. **(a)**).

A closer look to the catalytic site reveals that IPP binds to both the allylic sub-pocket, usually occupied by DMAPP, and the homoallylic sub-pockets (Figure 2.10.).
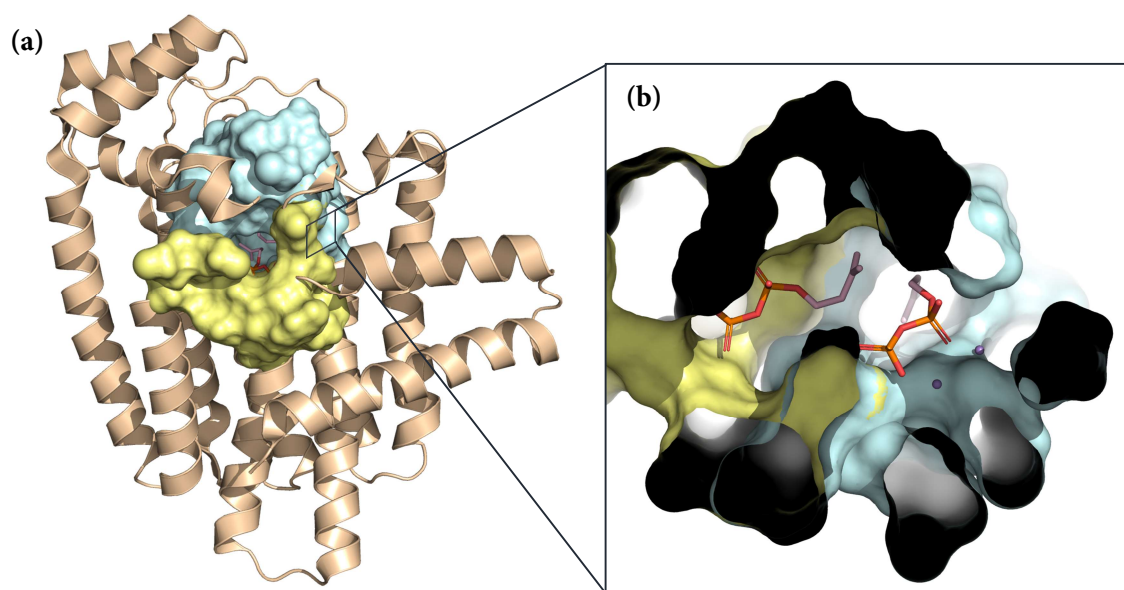

Figure 2.10. (a) Cartoon representation of FPPS-IPP-Mn^{2+}complex (PDB 6T7N). Surface representation of the allylic and homoallylic sub-pockets depicted in cyan and yellow, respectively. **(b)** Close up view of the solvent excluded surface showing the IPP molecules in the active site. Mn^{2+}ions are represented as violet spheres.

The phosphate groups of the allylic IPP are surrounded by a hydrogen bonding network, Mn^{2+} ions and basic amino acid residues (Figure 2.11.).

In particular, the oxygen atoms of the terminal P=O (2.0 Å) and the central OH (2.5 Å) groups of IPP coordinate Mn^{2+}a' together with W49 (2.2 Å), W97 (2.2 Å), W216 (2.3 Å) and the side chain of Asp250 (2.0 Å). Mn^{2+}b' , similarly to Mn^{2+}a', is octahedrally coordinated by the oxygen atoms of the carbonyl of Asp98 (2.0 Å) and Asp102 (2.1 Å) side chain, W21 (2.0 Å) and W6 (2.1 Å) and with the oxygen atoms of the terminal OH (2.0 Å) and the central P=O (2.3 Å) groups of IPP. Mn^{2+}c' also features an octahedral geometry. It is coordinated by the three water molecules W213 (1.9 Å), W214 (2.1 Å) and W75 (2.2 Å), the side chain of Asp102 (2.2 Å), the side chain of Asp98 (2.1 Å) and with the oxygen atoms of the central P=O group of IPP (2.0 Å) (Figure 2.11. **(a)**). In addition to coordinating the Mn^{2+} ions, the negatively charged phosphate groups are further stabilized by ionic interactions with the positively charged side chains of basic amino acids Arg107, Lys207

and Lys264 (Figure 2.11. **(c)**). Moreover, a dense water network bridges between the phosphate groups and the side chains of amino acids Ser104, Arg107, Gln167, Lys 207, Asp254 and Lys273, which form the allylic pocket (Figure 2.11. **(d)**). The terminal allylic moiety of IPP is accommodated in this pocket establishing hydrophobic interactions with the methyl and methylene groups of the side chains of Asp98, Thr163, Gln167 and Lys207 (Figure 2.11. **(c)**).
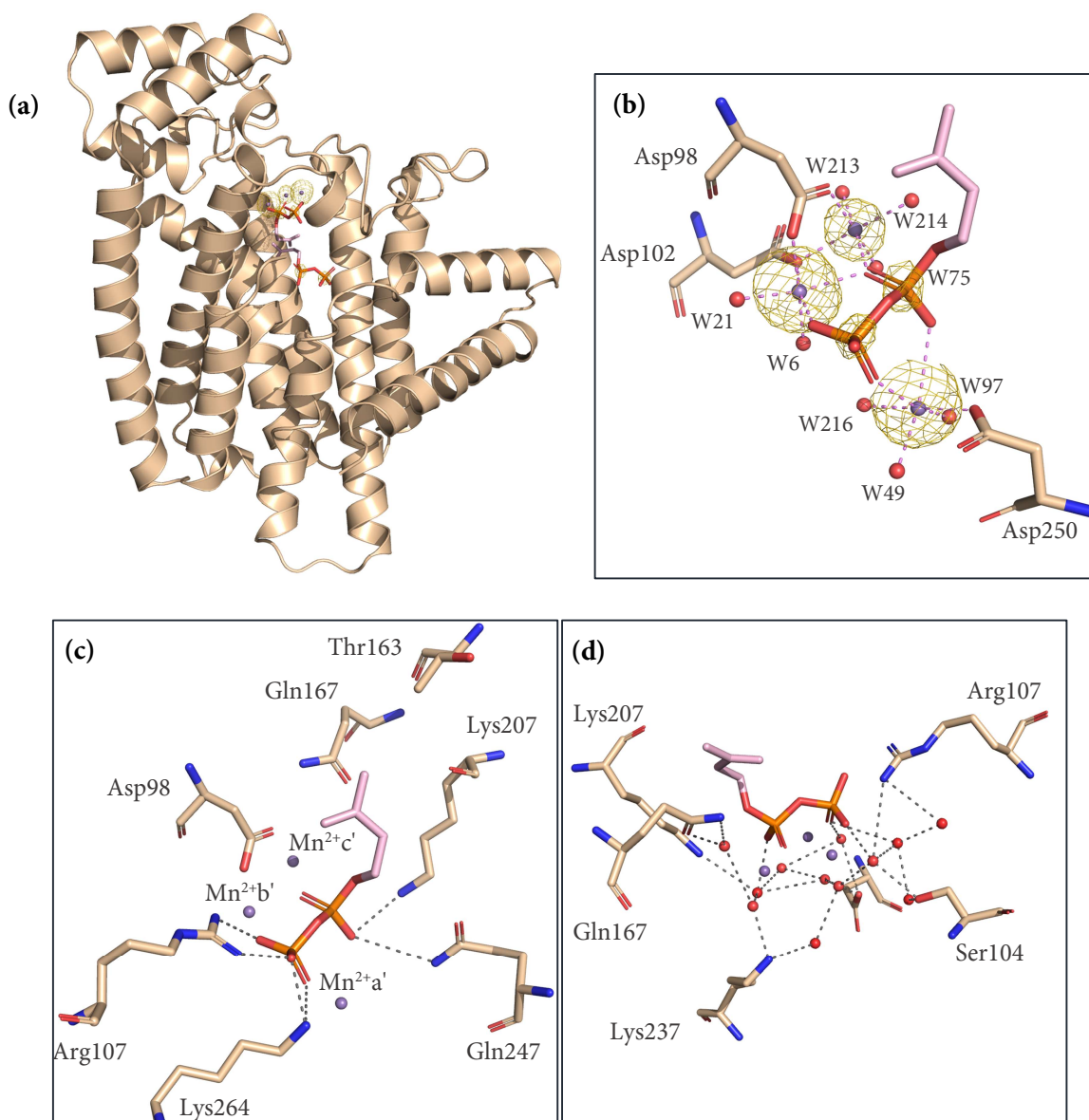
Figure 2.11. (a) Cartoon representation of FPPS-IPP-Mn^{2+}complex (PDB 6T7N) binding the IPP molecules (pink sticks) and Mn^{2+} ions (violet sphere) in the active site. The anomalous electron density peaks caused by the P and Mn atoms are depicted as a yellow mesh and displayed at a contour level of 4σ. Binding mode of IPP (pink sticks) in the allylic site, **(b)** the coordination with the three Mn^{2+} ions, **(c)** the electrostatic interactions with the surrounding basic amino acids and **(d)** the water network which further stabilizes the bound IPP.

In the homoallylic site, on the other hand, there are no Mn^{2+} ions and the negatively charged phosphate groups of IPP are stabilized by electrostatic interactions and *via* hydrogen bonds both directly and mediated by interstitial water molecules with the many basic amino acids forming the homoallylic pocket. In particular, the oxygen atoms of the terminal OH groups of IPP are at hydrogen-bonding distance with the NH nitrogens of the Arg51 side chain and the terminal nitrogen atom of Gln91 (Figure 2.12 **(b)**). The water molecules W564 and W586 bridge the backbone nitrogen atom of Tyr49 with the terminal hydroxy functions of the phosphate group (Figure 2.12 **(b)**) which are further stabilized by ionic interactions with the positively charged amino groups of Lys48 and Arg108 side chain (Figure 2.12 **(a)**). Lys48 is a key amino acid. It interacts with the terminal phosphate group of IPP directly with the backbone nitrogen and *via* W605 with the primary amine (Figure 2.12 **(b)**) but also stabilizes the central phosphate group through charged interactions together with the positively charged guanidinium moiety of Arg51 (Figure 2.12 **(a)**). Furthermore, the hydroxy function of the central phosphate moiety of IPP is involved in hydrogen bonds with the guanidinium group of Arg360 directly and the primary amino group of Lys264 *via* W521 (Figure 2.12 **(b)**). Finally, the allylic moiety of IPP is involved in hydrophobic interactions with the terminal methyl group of the side chain of Leu95 and with the methylene group of the side chain of Asp250 (Figure 2.12 **(a)**).
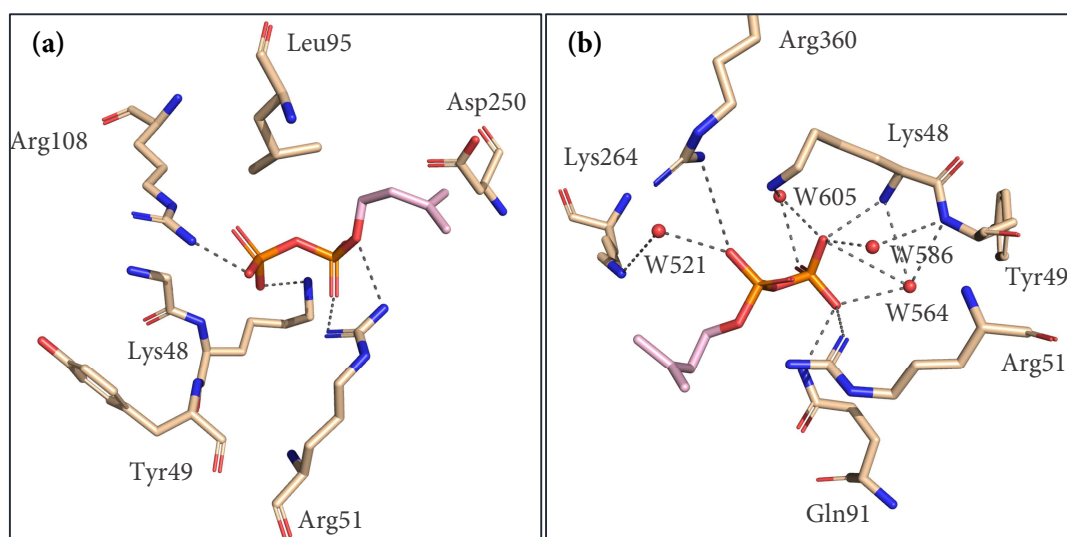
Figure 2.12. Binding mode of IPP in the homoallylic sub-pocket. **(a)** Electrostatic interactions between the phosphate groups of IPP and the side chains of basic amino acids Lys48, Tyr49, Arg51 and Arg108. The allyl moiety of IPP is involved in hydrophobic interactions with methyl and methylene groups of the side chains of Leu95 and Asp250. **(b)** Hydrogen bonding network between the phosphate group and the surrounding residues both directly and mediated by water molecules.

3. X-ray crystallographic fragments screening

As mentioned in Section 1.4. the aim of the project was to use a crystallographic fragment-based lead discovery approach to screen a general purpose in-house 96-fragment library.

Circa thousands of crystallization condition for *T. cruzi*, *T. brucei* and *human* FPPS were screened at the MarXtal laboratory in Marburg, resulting in few good starting points. Furthermore, Joy Petrick (EU-ITN AEGIS, ESR 12) and Lena Muenzker (EU-ITN FRAGNET) shared their crystallization condition for *T. cruzi* and *T. brucei*, respectively. The crystallization for *T. brucei* from Lena Muenzker was not reproducible and was not follow any further. Additionally, several attempts of optimizing the initial MarXtal hits resulted in salt crystals or in a diffraction not suitable to perform a fragment screening project. The crystallization condition of *T. cruzi* from Joy Petrick, on the other hand, was reproducible after an initial optimization (referred to as "seeding condition" from now on) and used, together with a condition found by further optimization reported later, for the screening of the 96-fragment library. *Human* FPPS also showed suitable starting crystals with the MarXtal screening, which underwent several optimization cycles that will be discussed in section 3.1.

In the first cycle of crystallization, a total of seven screens (JCSG Core I, JCSG Core II, JCSG Core III, JCSG Core IV, PACT, JCSG+ and Classics) were performed at the MarXtal laboratory for all three species under investigation. Each of the seven screen consists of a 96-well plate containing as many conditions which were tested at two different protein concentrations (10 mg/mL and 16 mg/mL). After buffer exchange with the optimized TSA buffers (Section 2.2.), the second cycle of crystallization, which consists of 12 screens (JCSG Core I, JCSG Core II, JCSG Core III, JCSG Core IV, PACT, JCSG+, Classics, Classics lite, Cryos, AmSO$_4$, Anions and Morpheus) were carried out. The most promising hits of each crystallization cycle were optimized manually in a 24-well plate.

All tested crystallization conditions are reported in Appendix 6.2.

3.1   The crystallization of *human* FPPS

3.1.1.   The first crystallization cycle

Among the hundreds of tested conditions, one resulted in producing a protein single crystal (JCSG Core I, H11) while in another ones (JCSG Core IV, H10) small crystals appeared after few days in the crystallization drop. These two starting conditions did not show any major improvements obtained by a matrix screenings based on varying the concentration of proteins and buffer components. However, the condition reported in the literature by *Jahnke et al.* was successful reproduced [51].

3.1.2.   The second crystallization cycle

After replacing the storage buffer with the TSA-optimized one, a second crystallization cycle was carried out at MarXtal laboratory resulting in four good starting points, two of which (Cryos C3 and D10) were selected and further optimized in a 24-well plate.

The first optimization cycle consisted of a screening around the initial conditions by varying the concentration of the buffer components. The most promising conditions were then selected and subjected to a second optimization cycle that involved a pH screening or the use of micro seeding techniques. This cycle of selection of the most promising conditions and optimizations was carried out to obtain crystals with a resolution suitable for a fragment screening project, i.e. < 2.5 Å [63].

The matrix screens around Cryos C3 and D10 are reported in the Table 3.1. below.

Table 3.1. Matrix screen around Cryos C3.

		1	2	3	4	5	6
	$(NH_4)_2HPO_4$ (M)	1.4	1.5	1.6	1.7	1.8	1.9
	0.08 M Tris						
A	pH:7.0						
B	pH:7.5						
C	pH:8.0						
D	pH:8.5						

In all:
0.08 M TRIS
25 % Glycerol

The dark pink box refers to the Cryos C3 starting conditions.

No significant improvement from the starting conditions (pink box) was observed in the first optimization cycle of Cryos C3.

On the other hand, as shown in Table 3.2., the optimization of Cryos D10 consisted of two consecutive cycles. The concentration of TRIS was kept constant for the entire screening. The optimized crystals diffracted up to 2.2 Å.

First optimization cycle (left):

KH_2PO_4 (M)		1	2	3	4	5	6
		0.4	0.5	0.6	0.7	0.8	0.9
	NaH_2PO_4 (M)						
A	0.5				■		
B	0.6			■			
C	0.7	■					
D	0.8		■				

Second optimization cycle (right):

0.08 M HEPES pH values		1	2	3	4	5	6
		6.0	6.5	7.0	7.5	8.0	8.5
A	0.7 M KH_2PO_4						
A4	0.5 M NaH_2PO_4				■		
B	0.4 M KH_2PO_4						
C1	0.7 M NaH_2PO_4				■		
C	0.5 M KH_2PO_4						
D2	0.8 M NaH_2PO_4				■		

Table 3.2. Matrix screen around Cryos D10. First (left) and second (right) optimization cycles. The dark pink box refers to the starting Cryos D10 condition while the light pink ones refer to the selected and further optimized conditions.

A similar approach was used to optimize the crystallization conditions of *T. cruzi* FPPS (Section 3.2.).

3.2 The crystallization of *T. cruzi* FPPS

3.2.1 The first crystallization cycle

Nine promising conditions resulted from the first screening (JCSG Core I B6, JCSG Core I G5, JCSG Core II F12, JCSG Core II G8, JCSG Core III G4, CSG Core III C6, JCSG Core IV H10 and JCSG+ C5). They all underwent a matrix screen optimization in a similar manner as previously discussed for the human FPPS (Section 3.1., Table 3.2.) by varying the concentration of the buffer components around the initial conditions. The most promising hits were subjected to a second optimization cycle. The crystals that grew too fast were crystallized again at 10 °C and 4 °C, while the small crystals were subjected to micro seeding techniques, such as the Classic E9 condition which was crystallized using the

streak seeding technique. However, as shown in Figure 3.1., although the crystals showed a regular and well-defined habitus, the diffraction limit felt close to 2.5 Å.
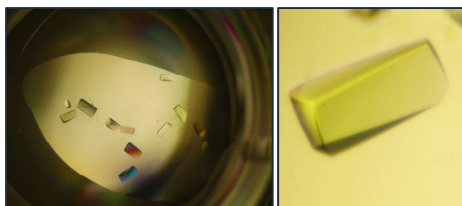

Figure 3.1. On the left, *T. cruzi* FPPS crystals after matrix screening and micro seeding (Classic E9). On the right, close-up view of a single crystal.

The crystallization condition that led to the most promising crystals was JCSG Core II G8. In particular, for the matrix screen, the pH values were increased by increments of 0.5 units in a range from 4.5 to 5.5 while the $(NH_4)_2SO_4$ concentration ranged from 0.12 M to 0.18 M at 0.02 M increments. As shown in Figure 3.2., crystals increased their dimensions as the pH value and $(NH_4)_2SO_4$ concentration increased. As a result of this optimization, crystals diffracted up to 2.3 Å.
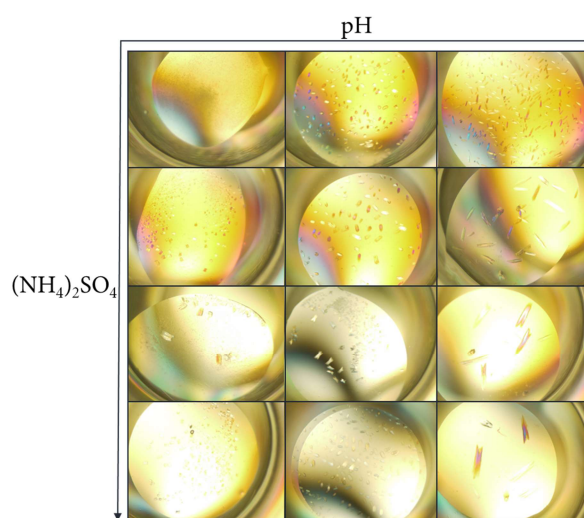

Figure 3.2. Matrix screen of *T. cruzi* FPPS crystals around the initial JCSG Core II G8 condition. Matrix screen showing the increase in crystal dimensions with increasing pH value and $(NH_4)_2SO_4$ concentrations.

Another promising crystallization condition was the seeding condition which requires the use of a seed stock solution resulting from crystals previously broken into small pieces and

diluted to 1:10, 1:100 and 1:1000 with the reservoir buffer. Interestingly, the crystals increased in size as the dilution level increased (Figure 3.3.) but they did not show a substantial difference in X-ray diffraction quality (2.3 Å).
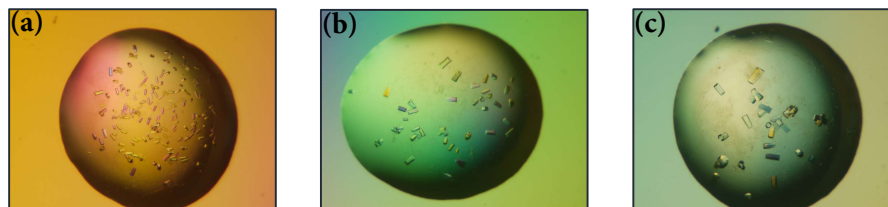

Figure 3.3. Increase in crystals size with increasing dilution of the seed stock from **(a)** 1:10 to **(b)** 1:100 and **(c)** 1:1000.

Both the seeding and optimized JCSG Core II G8 crystals grew in two to three days (Figure 3.4. **(a)**), but afterwards they rapidly changed shape, formed cracks, and resolution decreased fast (Figure 3.4. **(b)**). After a week they decayed even further by turning into irregularly shaped crystals which showed no X-ray diffraction anymore (Figure 3.4. **(c)**). The few crystals with a regular shape, which lasted a week, did not diffract any longer.

Figure 3.4. Decay of *T. cruzi* FPPS apo-crystals after **(a)** three **(b)** five and **(c)** seven days from their growth.

This decay phenomenon was even faster when the crystals were soaked in 100 mM fragment solution initially prepared from a 1 M DMSO stock solution. At the beginning, it was assumed that DMSO or the high fragment concentration would damage the crystals. Therefore, the fragment concentration and soaking time were reduced and the DMSO concentration that the crystals could tolerate was also tested. Subsequently, it was concluded that the conditions were too drastic, and the crystals were damaged due to osmotic shock. Crystals were then transferred and incubated in progressively increasing fragment concentrations to adapt to the soaking. However, these gentler conditions did not

work either. Finally, it was observed that the crystals were also decaying as long as they were in the apo-state. In addition, the protein material turned out to be no longer usable after about three months from the purification because it did not crystallize any longer, but instead, a precipitate appeared in the droplet, which increased over time. It was then assumed that the protein crystals were not stable because the protein solution in which they were stored was not ideal. The storage conditions were then optimized as described in Section 2.2. and 2.3. This resulted in both more stable crystals and more durable protein material, which crystallized up to six months after purification.

3.2.2 The second crystallization cycle

The protein in the TSA-optimized buffer was also used to perform an additional crystallization screen at the MarXtal laboratory resulting in three good starting conditions (JCSG+ C5, Anions E9 and Anions F9), which were further optimized in a 24-well plate.

Notably, these conditions differed only in the buffer component and the applied pH value. The first condition (JCSG+ C5) consisted of Hepes pH:7.5, the second (Anions E9) MES pH:6.5 and the third one (Anions F9) Tris pH:8.5. The crystals appeared after one day in the first two conditions and within two to three days in TRIS buffer.

The conditions were all reproducible and the crystals showed similar and regular shapes (Figure 3.5.).
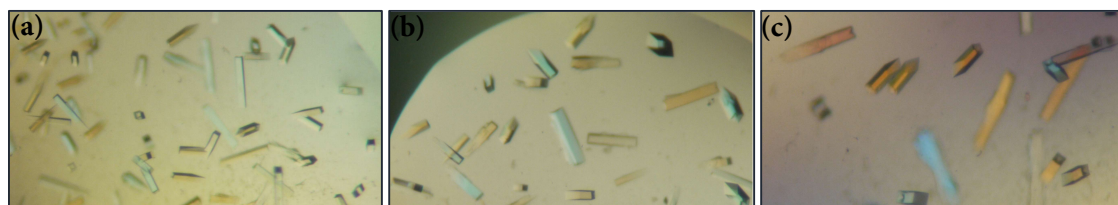

Figure 3.5. Reproduction of *T. cruzi* FPPS apo-crystals in **(a)** Hepes **(b)** MES and **(c)** TRIS buffer in a 24-well plate.

Table 3.3. Matrix screen around JCSG+ C5C. First (left) and second (right) optimization cycles.

		1	2	3	4	5	6
KH_2PO_4 (M)		0.6	0.7	0.8	0.9	1.0	1.1
	NaH_2PO_4 (M)						
A	0.6			▨			
B	0.7						
C	0.8	▨		▩	▨		
D	0.9						

		1	2	3	4	5	6
KH_2PO_4 (M)		0.4	0.5	0.6	0.7	0.8	0.9
	HEPES						
A	pH:6.5						
B	pH:7.0						
C	pH:7.5			▨		▩	▨
D	pH:8.0						

The green box refers to the starting JCSG+ C5C condition while the light blue ones refer to the selected and further optimized conditions.

After two optimization cycles (Table 3.3.), the crystals were tested in-house on a MAR Scanner 345 mm Image Plate detector at a wavelength of 1.542 Å. The ones in Hepes buffer showed X-ray diffraction until 2.8 Å at the synchrotron source. The poor resolution did not allow these crystals to be used for a fragment screening project. Further optimization cycles were required.

Nevertheless, it is interesting to note that a new crystal form was obtained. In fact, *T. cruzi* FPPS usually crystallizes in a hexagonal space group and it is characterized by a long c-axis of about 400 Å. This makes the data collection challenging as a compromise between a successful data collection and good resolution needs always to be found. In fact, high distances in real space corresponds to short distances in reciprocal space which may lead to difficulties in indexing as overlapping reflections can results. In consequence of the above-mentioned new conditions, *T. cruzi* FPPS also crystallized in the tetragonal space group I422 (97) with a c-axis reduced by approximately 100 Å compared to the hexagonal one.

3.2.3   The soaking protocol

Once stable apo-crystals were obtained, a soaking protocol had to be set up. As the crystals appeared to be sensitive to both mechanical and osmotic stress, the fragment concentration in the final drop was reduced and the number of transfers before their freezing were minimized.

First, the fragment solubilized in pure DMSO was diluted with the crystallization buffer to obtain a pre-mix with a final fragment concentration of 250 mM. Then, the required volume of the pre-mix to obtain a soaking concentration of 50 mM was pipetted onto the edge of the crystallization drop. The soaking solution could then slowly diffuse from the edge towards the entire drop, thus reducing both, the mechanical stress due to the transfer of the crystal into a new soaking drop, and the osmotic shock as the crystals could gently adapt to the fragment solution added to the initial crystallization condition (Figure 3.6, **(b)**).

The classical soaking protocol involves the transfer of the crystal from the crystallization to a soaking drop, prior to flash-freezing in liquid nitrogen (Figure 3.6, **(a)**). If the soaking solution does not contain sufficient amount of cryoprotectant, an additional transfer is necessary.
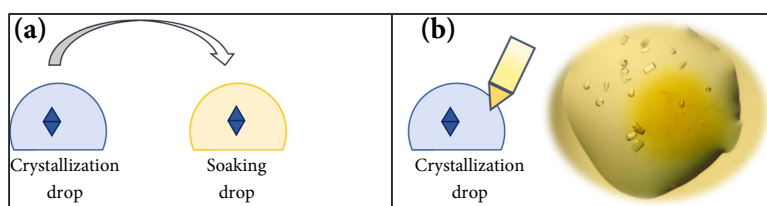

Figure 3.6. (a) Conventional and **(b)** new soaking protocol. **(a)** After their growth, crystals are usually transferred from the crystallization (blue) to the soaking (yellow) drop. **(b)** The fragment solubilized in soaking buffer (yellow) is pipetted directly in the crystallization drop (blue). On the right, the image shows the diffusion of the fragment solution into the initial crystallization drop.

Although the soaking conditions contained 20 % glycerol (GOL), it was not sufficient to cryoprotect the crystals. A screening to find the most suitable cryoprotectant was therefore carried out. It consisted of mixing the crystallization buffer with a concentration ranging from 15 % to 30 % of cryoprotectants like GOL, ethylene glycol (EG), MPD, DMSO and different ratio of two cryoprotectants such as MPD/DMSO and GOL/DMSO were also tested.

3.2.3.1   Xtunnel for the assessment of protein soakability

The evaluation of the soakability of the three species was performed using a script for the visualization of solvent tunnels in PyMOL called Xtunnel, developed by *Dr. Alexander Metz.*

In particular, it was possible to visualize and analyze the protein crystal packings as well as accessible channels through which fragments can travel to the protein binding pockets.

As shown in Figure 3.7., the protein of interest is located in the central unit cell of a 3x3 unit cell packing and is represented in green. A sphere around the central protein molecule is cutted out. In the generated image, tunnels are shown as transparent areas whereas blocked regions are represented in red. Any protein molecules from the packing with at least one atom within 5 Å from the central protein molecule of interest are displayed in yellow. All additional protein molecules with at least one atom within 5 Å from the central unit cell of the 3x3 unit cell packing are depicted in orange.

The manual inspection showed that *human*, *T. brucei* and *T. cruzi* FPPS crystallize all in the soakable crystal forms and exhibit sufficiently-sized channels (lowest diameter of the channel around 5 Å) suitable for soaking experiments.
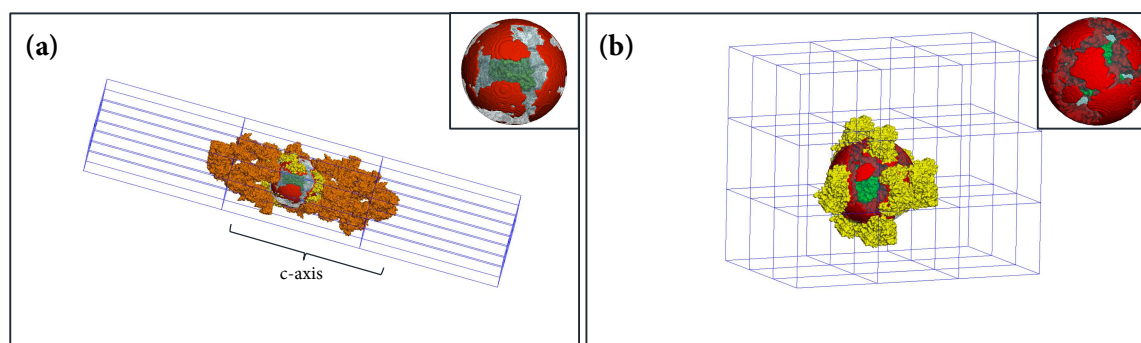

Figure 3.7. Protein crystal packing of **(a)** *T. cruzi* and **(b)** *T. brucei* FPPS generated with Xtunnel. On top right, close-up view of protein of interest depicted in green, accessible area in light blue and blocked area in red **(a)** The long c-axis of about 400 Å is indicated by a curly brackets.

The soakability of the crystallized protein was also evaluated experimentally. In fact, the aforementioned crystallization and soaking method were validated by obtaining the

T. cruzi FPPS holo-structure as well as the FPPS-IPP-Mn^{2+} tertiary complex (Section 2.5. and 2.6.) and they were subsequently applied to the fragment screening project.

3.3   The screening of the 96-fragment library

The entire 96-fragment library was successfully screened against *T. cruzi* FPPS obtaining high quality datasets with resolutions between 1.4 Å and 2.7 Å and a completeness of 99-100 % in data collection (Appendix, crystallographic table). Three fragment hits were identified in the electron density map. Out of these, two (**J51** and **J71**) are located in a remote pocket whose function is still unknown while the other one (**J82**) occupies the allosteric site.

3.3.1   The allosteric pocket

As described in Section 1.4., the project aimed at screening a general purpose 96-fragment library to map and investigate the different FPPS species, preferentially focusing on putative fragments hits found in the allosteric pocket. The latter pocket is well known in the human protein (Section 1.3.3.2.) while it has only been hypothesized for *T. brucei* and *T. cruzi* FPPS.

As previously discussed, the allosteric pocket in *human* FPPS is located near the C terminus of the enzyme, adjacent to the IPP binding site (Figure 3.8 **(a)**). In particular, the base of the pocket is rather hydrophobic (Phe206, Phe239, Leu344, Ile348 and Tyr10) while the upper part is composed by both positively charged (Lys57, Arg60 and Lys347) and polar (Asn59 and Thr63) residues.
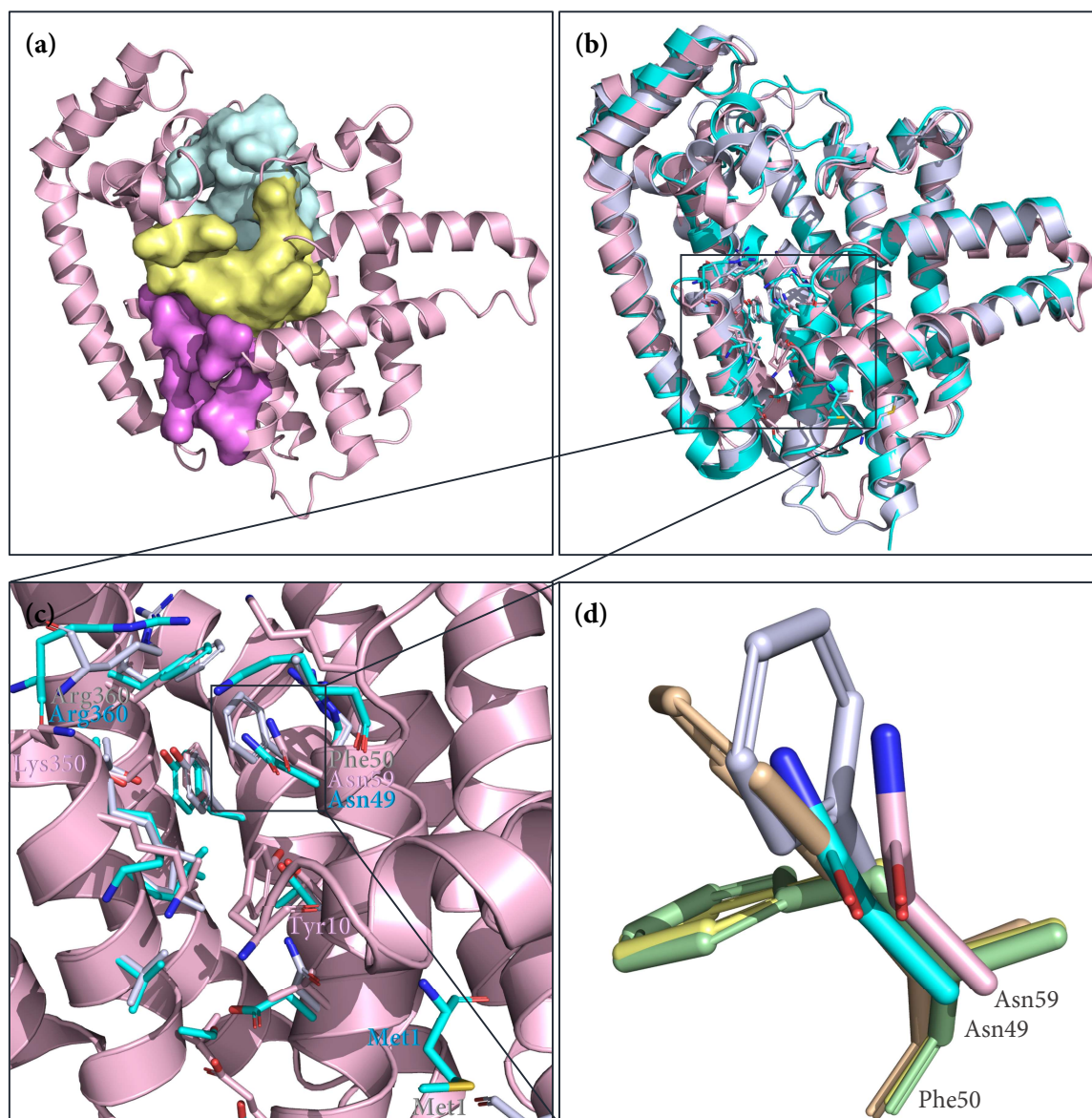
Figure 3.8. (a) Cartoon representation of *human* FPPS depicted in pink. The allylic (light blue), homoallylic (yellow) and allosteric (violet) pockets are represented as a surface. **(b)** Superposition of *human* (pink), *T. brucei* (cyan) and *T. cruzi* (whiteblue) FPPS as Cartoon representation. Amino acid residues which are part of the allosteric pocket are depicted as a stick model. **(c)** Close-up view of the allosteric pocket, only the backbone of the human enzyme displayed. Key amino acids which differ among the species (color-coding as above) are labeled next to the corresponding residue. **(d)** Close-up view of Phe50 of *T. cruzi* FPPS. It adapts three major conformations: (i) fully closed (green and yellow stick) in the apo-structure; (ii) partially open (wheat stick) in complex with natural substrate and (iii) fully open (lightblue stick) when a fragment binds the allosteric pockets. The Phe50 torsion angle ranges from 143.4 ° in apo-structure to 179.4 ° in the bound state. *Human* and *T. brucei* FPPS have an Asn residue instead. This may makes their pockets larger than the *T. cruzi* FPPS one.

The superposition of *human*, *T. brucei* and *T. cruzi* FPPS crystal structures shows that their allosteric pockets are quite similar, but a closer inspection reveals some differences in key residues (Figure 3.8. **(b)** and **(c)**, (Table 3.4.)). In particular, the *human* Tyr10 frontally closes the allosteric pocket where *T. brucei* and *T. cruzi* FPPS have the Met1 residue, which

is far remote from the pocket (Figure 3.8. **(c)**). An opposite situation occurs in the upper part of the allosteric pocket. In fact, Arg360 in *T. brucei* and *T. cruzi* are located at the top of helix J, at the entrance and between the homoallylic and the allosteric pockets and seal the latter from the top. The corresponding Lys350 in the *human* isoform, moves instead away from the pocket, allowing a direct connection between the homoallylic and allosteric pockets (Figure 3.8. **(c)**). In front of helix J, the helix C has, in the upper part, Phe50, which is a key residue that acts as a gate between the homoallylic and allosteric pocket in the *T. cruzi* FPPS. In fact, this residue creates a larger steric barrier than the corresponding Asn59 and Asn49 in *human* and *T. brucei* FPPS. In addition, Phe50 adopts three major conformations (i) fully closed in apo-structure, i.e. when the enzyme adapts an open conformation (ii) partially open, when the enzyme binds the IPPs and adapts a fully closed conformation (iii) fully open, when a fragment binds the allosteric pocket. In fact, Phe50 moves upwards by about 36 ° (torsion angle), thus maintaining the open state of the pocket (Figure 3.8. **(d)**).

Table 3.4. Major differences in residues which form the allosteric pocket among the FPPS species.

human	**Tyr10[a]**	Asn59	Val66	Glu70	Phe206	Lys350
T. brucei	Met1	Asn49	Asp56	Ser60	Tyr218	**Arg360[a]**
T. cruzi	Met1	**Phe50[a]**	Asn57	Gly61	Tyr213	**Arg360[a]**

[a] refers to most important residues

3.3.1.1   The MD simulations

MD simulations revealed differences between the allosteric pockets found in the different species. In particular, the conformation adopted by the catalytic pocket when a hypothetical inhibitor binds the allosteric pocket and the C/J distance of the allosteric pocket were investigated.

As reported in Section 2.4., the *human* apo FPPS has a FARM-SARM distance distribution showing the protein to 49 % of the time in the open conformation. It was computed that, in the presence of an allosteric inhibitor, e.g. compound **1**, developed by *Jahnke et al.* [51], the enzyme even remains to 59 % of the time in open conformation. A similar behavior was

simulated for the Trypanosoma species. This result may suggest that an allosteric inhibitor may help to stabilize particularly the open conformation. To confirm the above-mentioned result, an umbrella sampling of FARM-SARM distances was also performed. This study revealed that the low energy state shifts to bigger distances when an allosteric inhibitor is bound to this pocket compare to the unbound form (Figure 3.9.). This suggests that the open conformation is energetically stabilized once an allosteric inhibitor is bound.
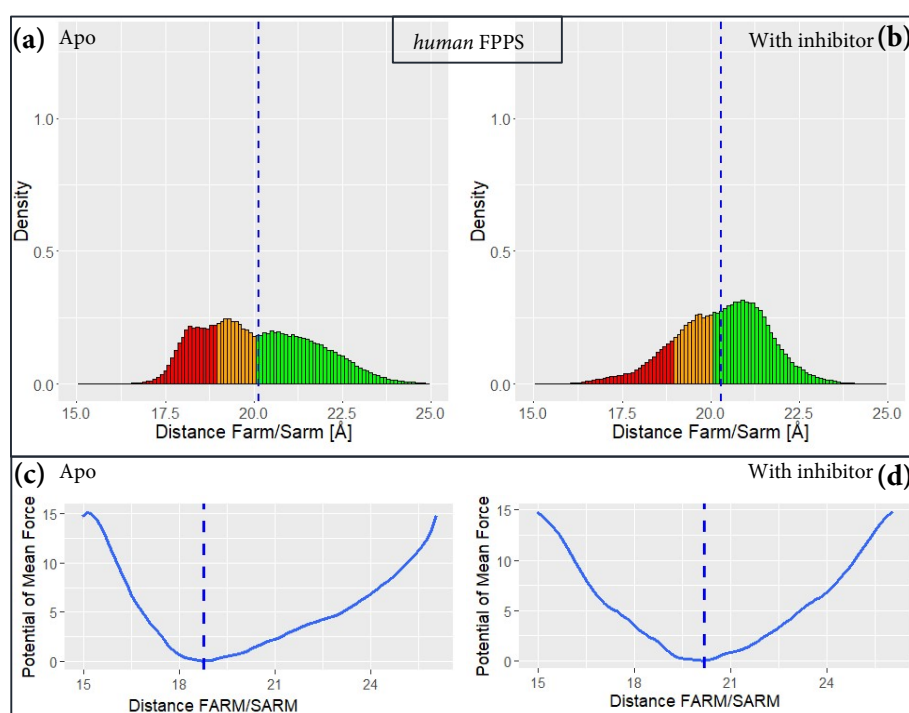
Figure 3.9. Comparison of histograms **(a, b)** and the potential of mean force **(c, d)** of the FARM-SARM distances between the *human* FPPS **(a, c)** apo and **(b, d)** allosteric inhibitor bound state using an umbrella sampling during the simulation. In the histograms, the red area refers to the fully closed, the orange to the partially closed and the green area to the open conformation of the enzyme. In the umbrella sampling, the lower energy minimum shifts to bigger distances when an allosteric inhibitor binds to the protein. Both methods show that an allosteric inhibitor stabilizes the open conformation of the enzyme.

The distance between the helices C and J (C/J distance from now on) was chosen as descriptor to compare the behavior of the allosteric pocket in the different species. Interestingly, *human* and *T. brucei* FPPS experienced a similar C-J distance of 11.6 Å. In the case of *T cruzi* the upper limit is 11 Å and even lower in the apo-structure. This indicates

that the allosteric pocket of the latter species is smaller compared to the other two species (Figure 3.10.). This difference is probably due to the special role found for Phe50, which adopts multiple conformations and may have a key role as gating residue.
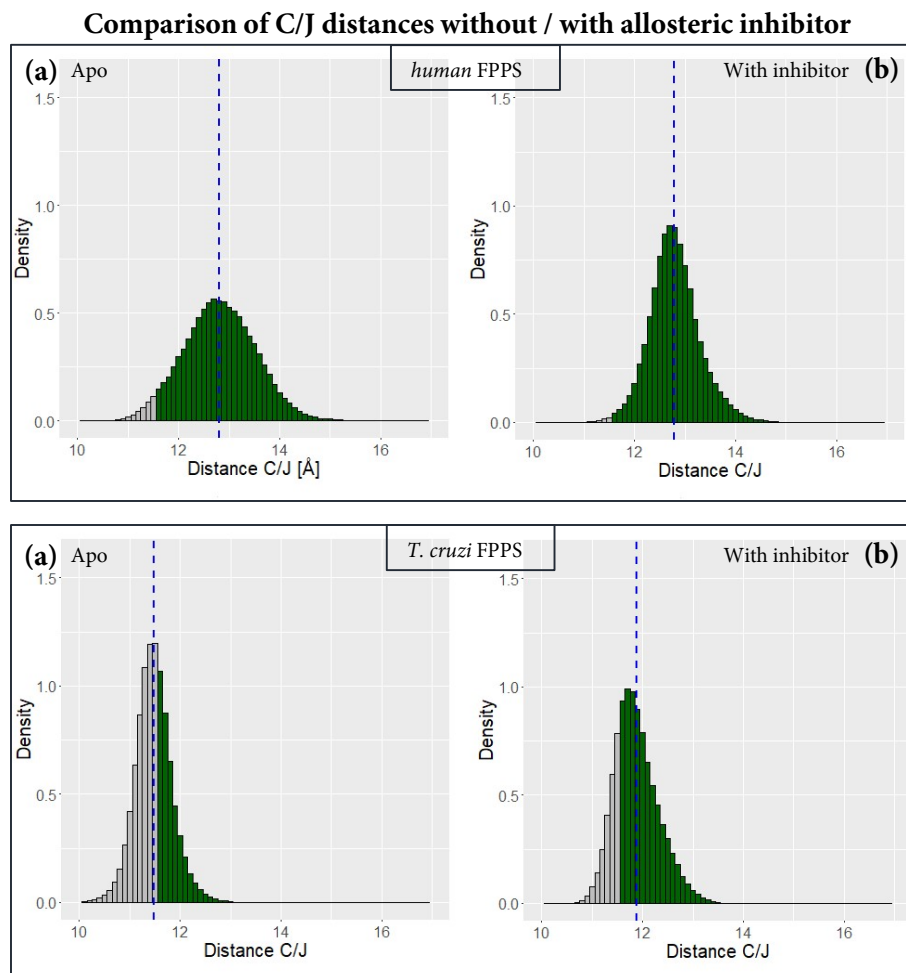

Figure 3.10. Comparison of histograms of C/J distances between **(a, b)** *human* and **(c, d)** *T. cruzi* FPPS in **(a, c)** apo-structure and **(b, d)** with a bound allosteric ligand. The gray area refers to the possible C/J distances without any bound ligands while the green area refers to the distances possible to adopt and giving space to accommodate an allosteric ligand.

3.3.1.2   Binding mode of T. cruzi FPPS in complex with J82

The fragment screening revealed one fragment hit in the allosteric pocket which is a nitrobenzene, substituted in *meta* and *para* position with hydroxyl groups. **J82** penetrates into the allosteric pocket where it is sandwiched in between the Phe50 side chain, which adopts the fully open conformation, and the Tyr213 side chain with which the aromatic ring of the fragment is involved in hydrophobic interactions such as aliphatic CH-π

interaction and π-π stacking. Furthermore, the nitro group of **J82** is an electron withdrawing group (EWG) which interacts eletrostaticlly *via* π-hole interaction with the electron donating groups (EDG) of Tyr213 and Thr357 side chains. It has been reported that nitrobenzene compounds are often involved in π-hole interactions with protein side chains [64]. This interaction seems to be responsible for the superior binding affinity or inhibition of a nitro aromatic ligand compared to its non-nitro analogue [64].

On the opposing side of the molecule, both hydroxyl functions are within H-bonding distance with the oxygen atom of the Thr54 side chain (2.3 Å and 3.2 Å). In addition, the hydroxyl group in *para* position of **J82** is involved in a hydrogen bond with the oxygen atom of the carbonyl backbone of Thr212 (2.9 Å).
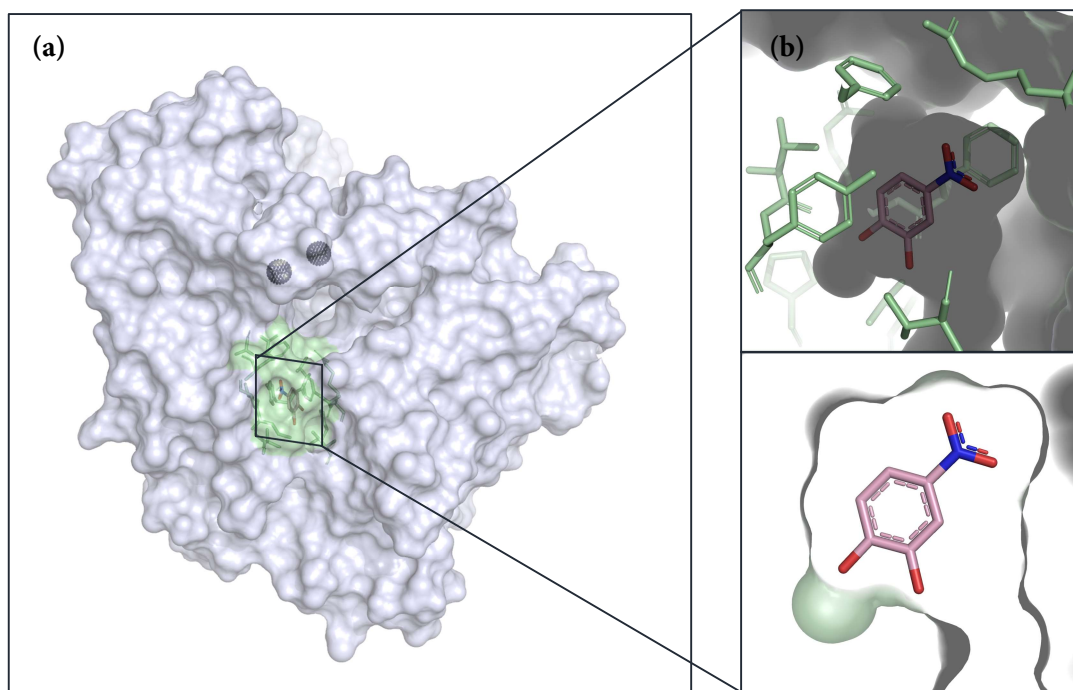
(figure continues on the next page)
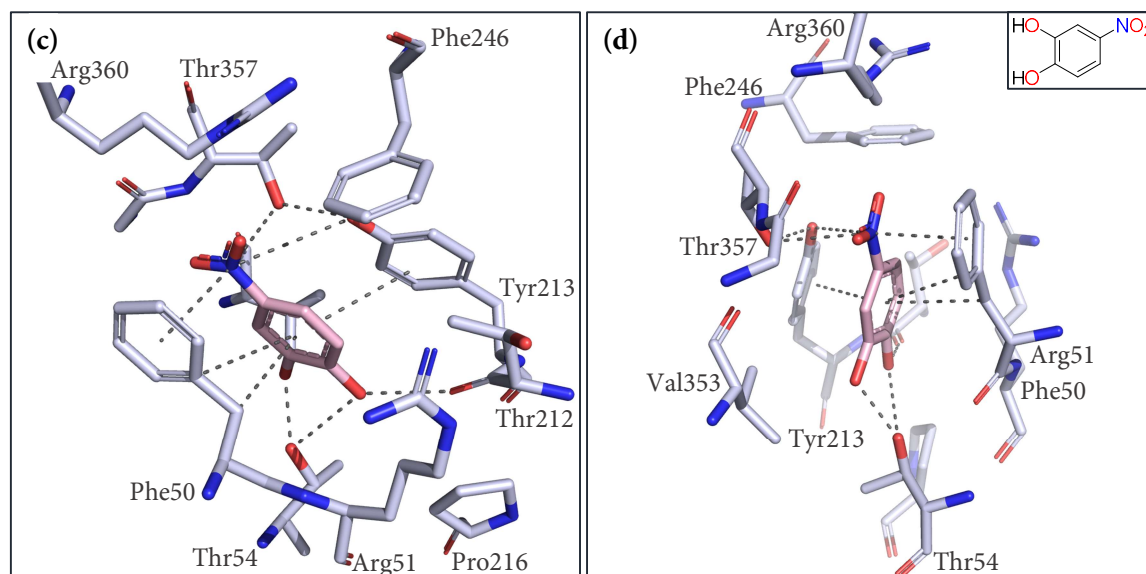
Figure 3.11. (a) Surface representation of *T. cruzi* FPPS. Compound **J82** (pink stick) binding in the allosteric pocket (green stick and surface). **(b)** Close-up view of allosteric pocket of *T. cruzi* FPPS represented as a stick model and surface with the bound **J82** (pink stick). **(c, d)** Binding mode of **J82** with the surrounding residues represented as stick model.

Since **J82** was particularly interesting considering the aim of the project, its binding mode was further evaluated by the DSX scoring function and subsequentially by soaking *T. cruzi* FPPS crystals in different analogous of **J82** such as nitrobenzene, pyrocatechol, 3-nitrophenol and 4-nitrophenol.

3.3.1.2.1   DSX computer scoring

The docking program HYBRID (version 3.3.0.3, OpenEye Scientific Software) was used to reproduce the crystal structure with **J82** and a good overlapping between the crystal structure (pink) and the docked (magenta) pose is indicated in Figure 3.11., **(a)**. **J82** is allocated in the allosteric pocket at favorable distance and position with respect to the surrounding amino acid side chain. Application of the DSX scoring function shows by graphical means the quality of the adopted binding mode. The blue spheres in Figure 3.11. **(b)**, represent the per-atom DSX score contributions whereby the size of the spheres is an indication for the strength of the attractive interactions. The large blue spheres denote a more attractive interaction than smaller spheres. With respect to interactions, the OH

group in para position of **J82** is involved in a favorable interaction with the oxygen atom of the Thr54 side chain (blue line in Figure 3.11, **(b)**). More details of the present method are provided in the Material and Methods section (Section 6.8.).
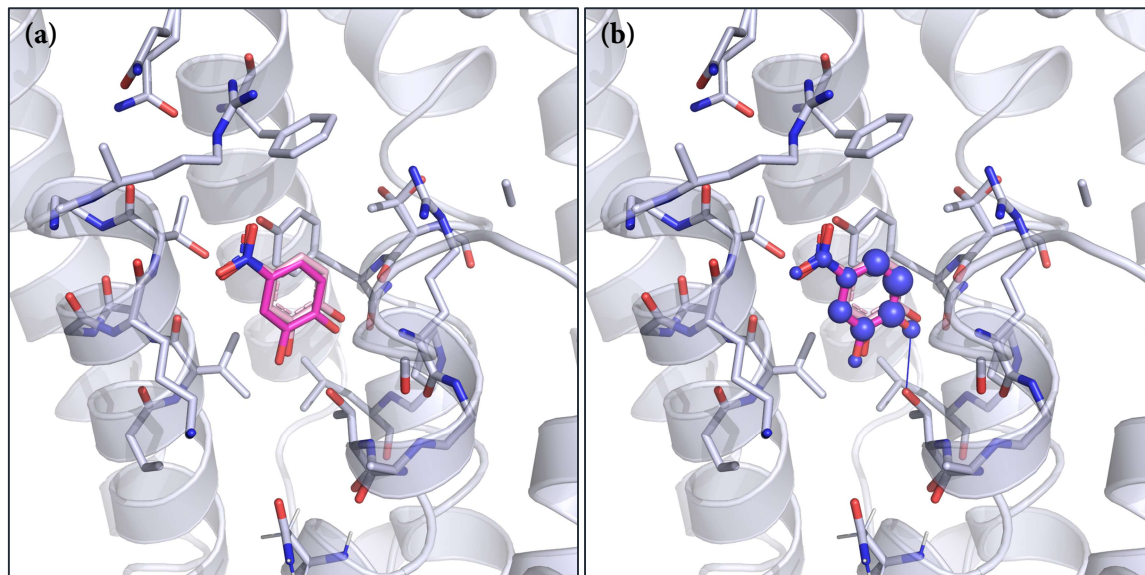

Figure 3.11. Cartoon representation of T. cruzi FPPS. The allosteric pocket is highlighted as whiteblue stick model. **(a)** Superposition of the crystal structure (pink stick) and the docking (magenta stick) pose of **J82** in the allosteric pocket. **(b)** DSX scoring function shows a favorable distance (blue line) and a sphere-based scoring of the positions where ligands atoms have been found (blue sphere) of **J82**.

3.3.1.2.2   A series of J82 derivatives

J82 derivatives (Figure 3.12.) were soaked into crystals of *T. cruzi* FPPS to investigate the individual role played by each substituent of the benzene ring. Interestingly, pyrocatechol and nitrobenzene did not bind, whereas 3-nitrophenol and 4-nitrophenol showed a similar binding mode as **J82** (Figure 3.12, 3.13). This rational approach showed that at least one hydroxyl function and a nitro group must be present at the same time for the molecule to bind with this binding mode into the allosteric pocket. Perhaps, the hydroxyl group, being an electron donor group, increases the delocalization of π-electrons from the benzene ring and thus enhances the electron attracting effect of the nitro group and stabilizes the molecule by increasing the number of possible resonance structures.

The nitrobenzene moiety of 3-nitrophenol and 4-nitrophenol interacts in a similar manner as previously discussed for **J82** (Section 3.3.1.2.).
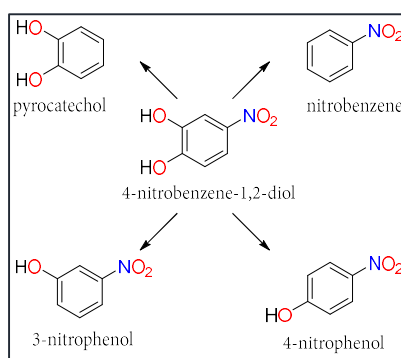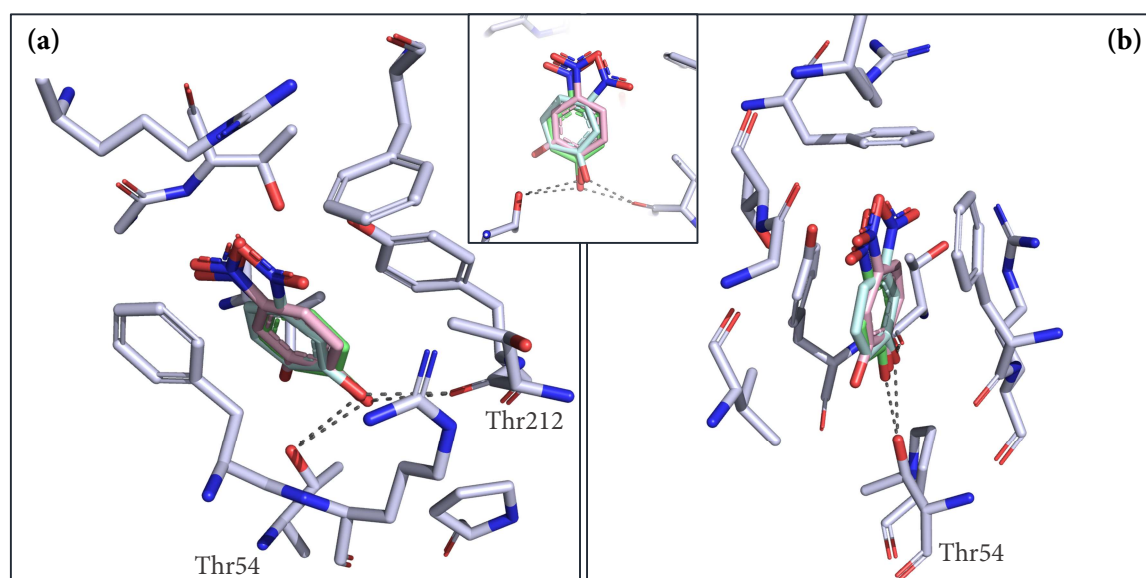
Figure 3.12. Chemical structure of **J82** derivatives.

Figure 3.13. Binding modes of **J82** derivatives show a good overlap between **J82** (pink stick) and 4-nitrophenol (green stick) whereas the 3-nitrophenol (light blue stick) reoriented the OH group towards Thr212, causing the nitro group to move compared to the position assumed in **J82** and 4-nitrophenol. In the middle, close-up view of the hydrogen bonds between the fragment hits and the oxygen atom of the Thr54 side chain and the carbonyl backbone of Thr212.

However, while 4-nitrophenol overlaps almost perfectly with **J82** and establishes the same interactions that the hydroxyl group in *para* position of **J82** makes with Thr54 and Thr212 (2.8 Å), the 3-nitrophenol instead reorients the OH group towards the oxygen atom of the carbonyl backbone of Thr212 with which it establishes an hydrogen-bond interaction (2.7 Å). Like **J82**, the two derivatives are also within H-bonding distance to the hydroxyl group of the Thr54 side chain (Figure 3.13).

Perhaps it was due to the structural arrangement of the 3-nitrophenol, that it was not possible, as for the 4-nitrophenol, to build the fragment in the electron density immediately after the MR but further refinement steps were needed. In addition, the 3-nitrophenol

exhibits an occupancy of 77 % and its electron density map is weaker compared to the *para* substituted analogue which also fully occupies the allosteric pocket (Appendix, 6.3.).

3.3.1.2.3   J82 follow-up

Exploring the chemical space around **J82** and using the Frag4Lead website to find larger analogs, two small follow-up compounds, 2-methoxy-5-nitrophenol and 2-methoxy-4-nitrophenol, were also soaked into *T. cruzi* FPPS crystals. Interestingly, the first bound to the allosteric pocket while the second one did not. This is an indication that small changes can be critical for the success of the binding event. It is known that the effect of the substituents in *para* position of a nitrobenzene is greater than the one in *meta* position. Among several electron donating groups, the methoxy function provides a higher total effect compared to the hydroxyl group [65]. Considering both the higher priority of the *para* position and the higher electron donating properties of the methoxy group, the 2-methoxy-5-nitrophenol may be more favorable compared to its analogue and thus preferentially binds to the protein.

As shown in Figure 3.14, the 2-methoxy-5-nitrophenol binds into the allosteric pocket in two configurations with an occupancy of 68 % and 32 %. In particular, the nitrobenzoic moiety shows a similar binding mode as **J82** and its derivatives, while both the methoxy and hydroxyl function interact *via* a hydrogen bond with the oxygen atom of the Thr54 side chain, but not with the carbonyl backbone oxygen of Thr212, in contrast to **J82** as its derivatives.
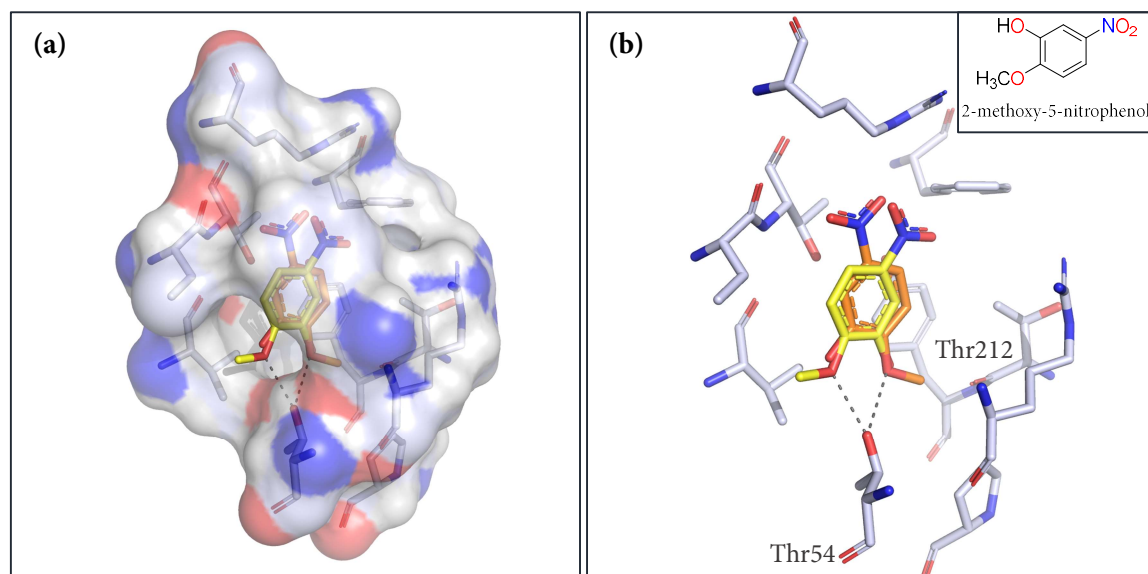
Figure 3.14. (a) Surface and stick representation of the allosteric pocket of T. *cruzi* FPPS binding the 2-methoxy-5-nitrophenol in two alternative configuration with an occupancy of 68 % (orange stick) and 32 % (yellow stick). **(b)** Binding mode of 2-methoxy-5-nitrophenol in the allosteric pocket represented as whiteblue stick. The compound is involved in a hydrogen bond with the oxygen atom of Thr54 side chain but does not interact with the carbonyl backbone of Thr212.

3.3.2   The discovery of an unknown pocket

The fragment screening identified also an additional pocket, located between helices H, I and α3. It is a rather hydrophobic pocket, surrounded by Val252, Phe256, Trp276, Leu277, Tyr312, Ser313, Leu217, Gln318 and Phe320. The two fragment hits, **J51** and **J71** (caffeine), have a similar chemical structure and showed similar binding modes (Figure 3.15.).
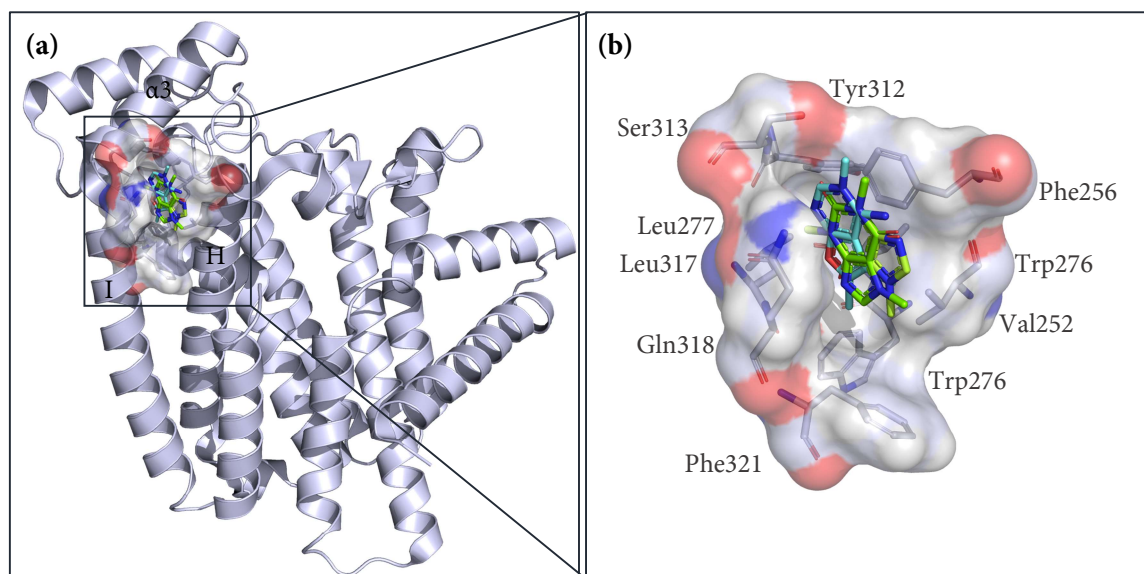
Figure 3.15. (a) Cartoon representation of *T. cruzi* FPPS. The new pocket is located between helices H, I and α3 and is represented as a surface and stick model. The fragments **J51** (aquamarine stick) and **J71** (green and lime stick) show a similar binding mode. **(b)** Close-up view of **J51** and **J71** binding the new pocket depicted as surface and stick model.

In particular, the pyrimidine ring of **J51** and the purine rings of **J71** are involved in π-π interactions with the Phe256 side chain. Further hydrophobic interactions involved the methyl and methylene groups of Val252 and Leu277 side chain with the fused ring of the fragment hits (Figure 3.16.).

The nitrogen atom of the Gln318 side chain interacts *via* W707 with the imine group of **J51** (3.1 Å and 3.3 Å) (Figure 3.16. **(a)**).

J71, caffeine, adopts two configurations with an occupancy of 60 % and 40 %. The six-membered ring of the two orientations perfectly overlaps, while the five-membered ring is rotated by about 180 ° against each other. In addition, the O11 oxygen atom of the carbonyl group of **J71** (3.2 Å) is within H-bonding distance with the oxygen atom of the Ser313 side chain (3.3 Å) (Figure 3.16. **(b)**).
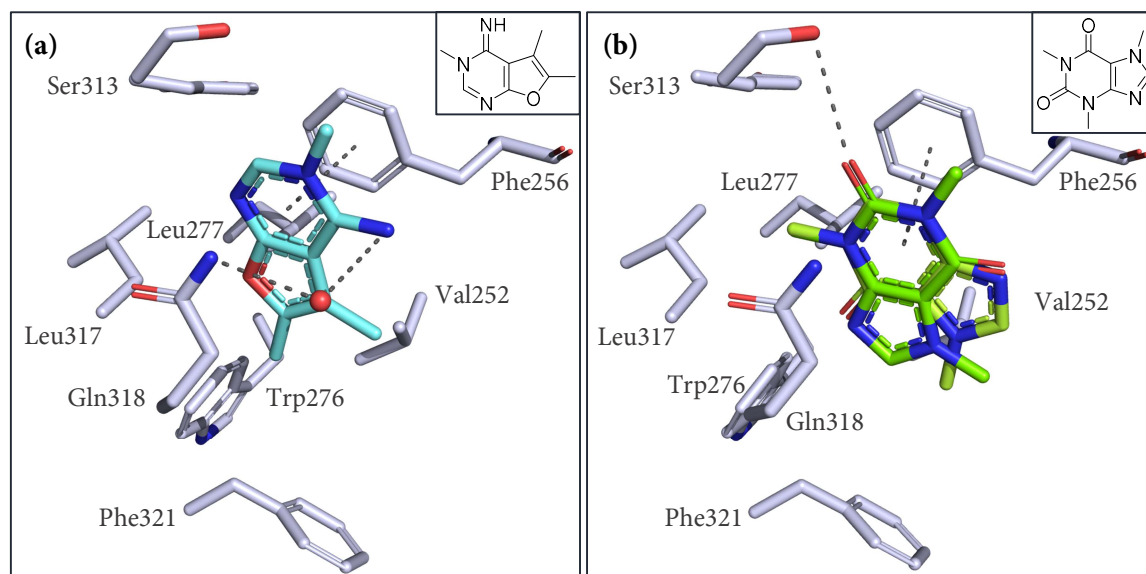
Figure 3.16. Binding mode of **(a)** **J51** and **(b)** **J71** represented as stick model. **J71** binds in double conformation with an occupancy of 60 % (green sticks) and 40 % (lime sticks).

3.4   PanDDA analysis and additional hits

As already discussed in Part II (TLN), Section 3.2, PanDDA is a powerful approach to find fragments binding with low occupancy which are difficult to detect with the conventional maps. As input files datasets were used belonging to both, the 96-fragment library and the derivatives of the follow-up compounds of **J82**. PanDDA analysis identified 44 event maps distributed across 10 protein sites, 33 of which were denoted as interesting datasets where it appeared likely to find a bound fragment (Figure 3.17). Among the 33 interesting datasets, PanDDA detected all hits which were already found by applying the conventional pipeline and discussed in the previous sections (Section 3.3.). After manual inspection, three additional hits were found as relevant in the event maps. They correspond to an occupancy of 71-80 % (Appendix, Section 7.4.). Out of these, **J29** accommodates in the protein interface proximity between the two monomers, enclosed by helices F, G and H. **J59** bind between the H-α1 loop and the F-G loop which is part of the characteristic 11-residues insertion between amino acids 179-189 which is unique in the *Trypanosoma* species. In addition, the Pyrocatechol which is a surrogate of **J82** did not bind in the

allosteric pocket (Section 3.3.1.2.2.). However, PanDDA analysis revealed that it binds instead in a small shallow pocket on the top of helix D and enclosed by the D-E loop.
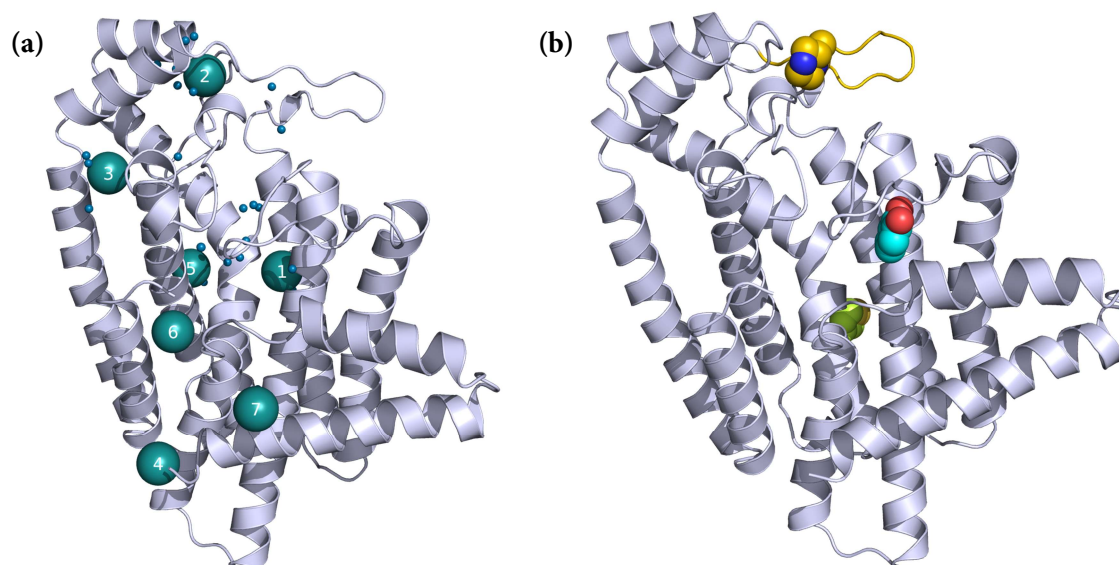
Figure 3.17. Cartoon representation of the *T. cruzi* FPPS secondary structure. (a)The protein is divided into seven sites depicted as teal spheres and numbered in chronological order of identification. The small aquamarine spheres represent the position of the 33 plausible hits identified as interesting datasets by PanDDA analysis. (b) PanDDA hits are represented as lime (**J29**), yellow (**J59**) and cyan (pyrocatechol) spheres. The characteristic *Trypanosomal* 11-residues insertion loop between amino acids 179-189 is colored in yellow.

3.4.1 *T. cruzi* FPPS in complex with J29

J29 accommodates in a surface-exposed hydrophobic pocket, next to the protein interface between the two monomers, where the thiophene ring is mainly involved in hydrophobic interaction with the methyl and methylene groups of Leu155, Lys158, Val159, Leu214 and Val218 side chain. The nitrogen atoms of the amidino function of **J29** is involved in a long-range charge-assisted interaction with the oxygen atoms of the carboxylate group of Glu236 side chain (Figure 3.18.).
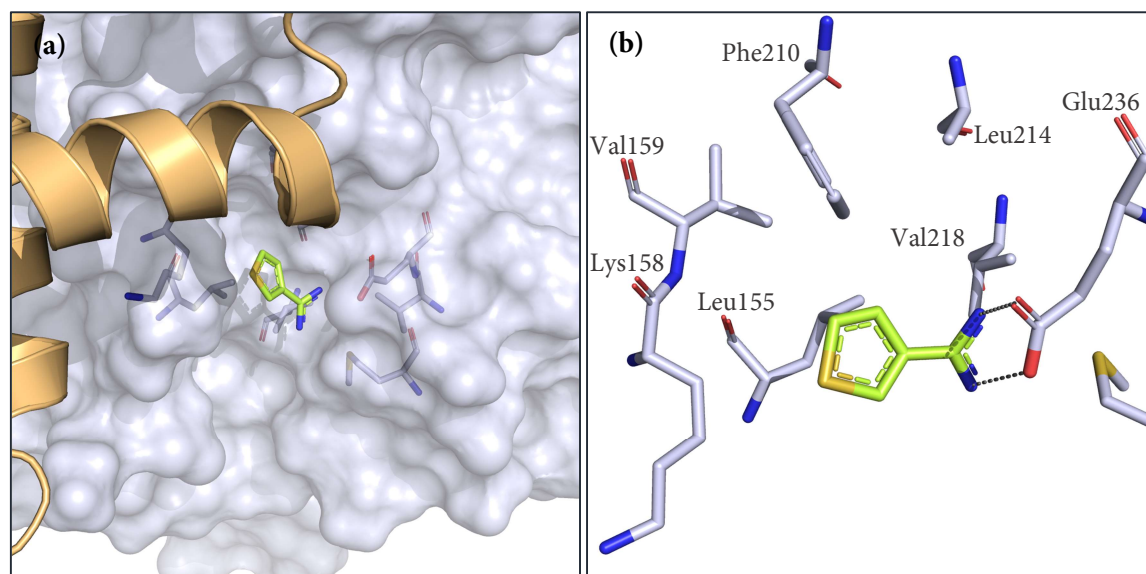
Figure 3.18. (a) Surface representation and **(b)** binding mode of **J29** represented as lime stick model. The corresponding monomer which is part of the homodimer is represented as light orange cartoon.

3.4.2   *T. cruzi* FPPS in complex with J59

J59 binds between the H-α1 loop and the F-G loop which is part of the characteristic *Trypanosomal* 11-residues insertion between amino acids 179-189. In particular, the NH nitrogen atom of **J59** is within hydrogen-bonding distance to the carbonyl oxygen of the backbone of Glu270 and Gly298 (3.0 Å and 2.8 Å). In addition, the carbonyl group of **J59** is involved in a hydrogen bond with the backbone carbonyl oxygen of Pro187 *via* an interstitial water molecule (Figure 3.19).
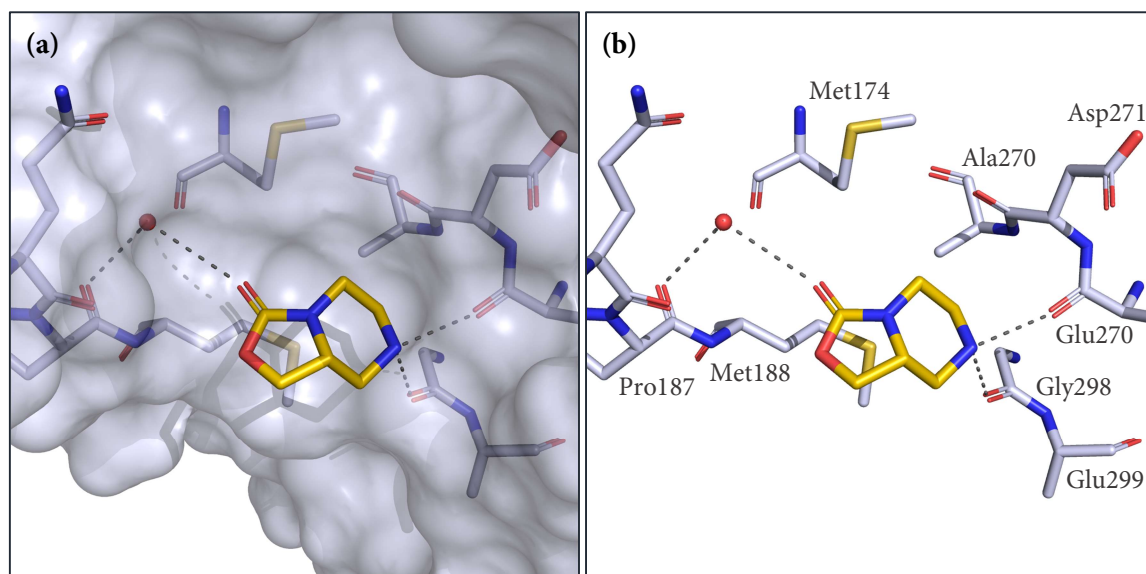
Figure 3.19. (a) Surface representation and **(b)** binding mode of **J29** represented as yellow stick model.

3.4.3   *T. cruzi* FPPS in complex with Pyrocathecol

Pyrocatechol accommodates in a small shallow pocket on the top of helix D and is enclosed by the D-E loop. Its benzene ring is involved in hydrophobic interactions with the methyl and methylene groups of Thr43, Ala92, Leu95, Val96, Arg107and Arg108 side chain. Both oxygen atoms of the fragment are within hydrogen-bonding distance with the backbone amino group of Trp113 (3.1 Å and 3.0 Å). In addition, one oxygen atom is further stabilized by a hydrogen bond with the carboxylate group of Asp99 side chain (2.6 Å) while the other one is involved in H-bonding with the carbonyl backbone oxygen of Pro111 (3.4 Å) (Figure 3.20).
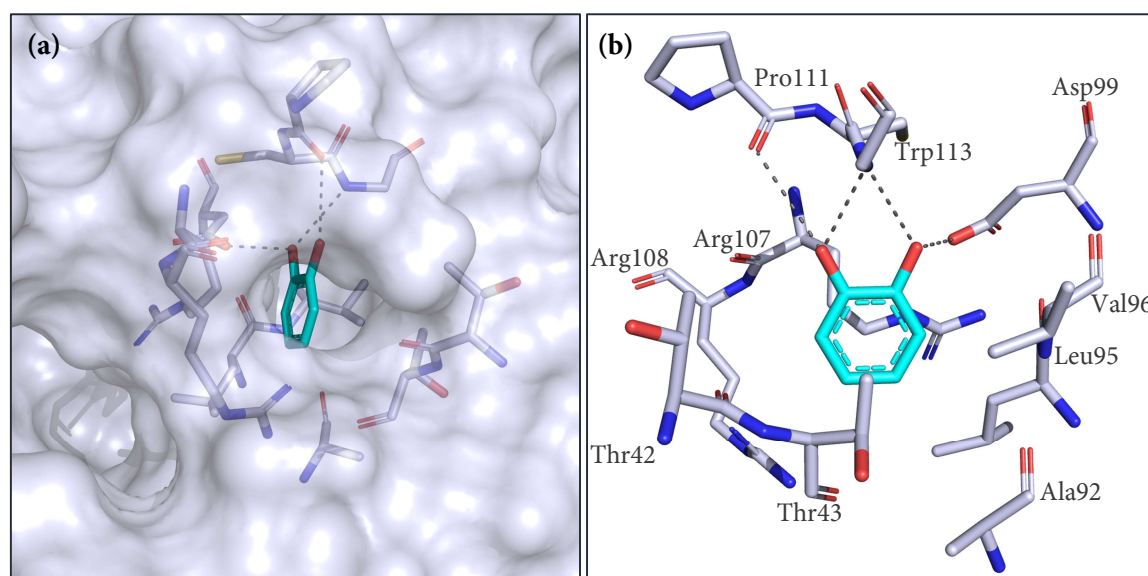
Figure 3.20. (a) Surface representation and **(b)** binding mode of Pyrocatechol represented as cyan stick model.

4. Characterization of FPPS species with Mass Spectroscopy (MS)

According to the literature, all FPPS species studied in this thesis should be stable homodimers. However, while a protocol for the in-house measurement of thermodynamic parameters of *human* FPPS in complex with inhibitors could be established by isothermal titration calorimetry (ITC), the stability of FPPS as a homodimer was being questioned, according to the shape of the recorded thermograms. In fact, the major difficulty was to find the correct model for fitting the recorded data. Usually, if a more complex model is assumed, the more parameters need to be fitted. Usually a better fitting model will result, due to the larger number of adjustable parameters. This could potentially lead to overfitting the parameters and overestimate the isotherm. In fact, the models that produced a better fitting in the current case were rather complex models that assumed a conformational change, a dissociation or association of the two monomers before the inhibitor was bound or more events seemed to occur in parallel at the same time (Figure 4.1.).
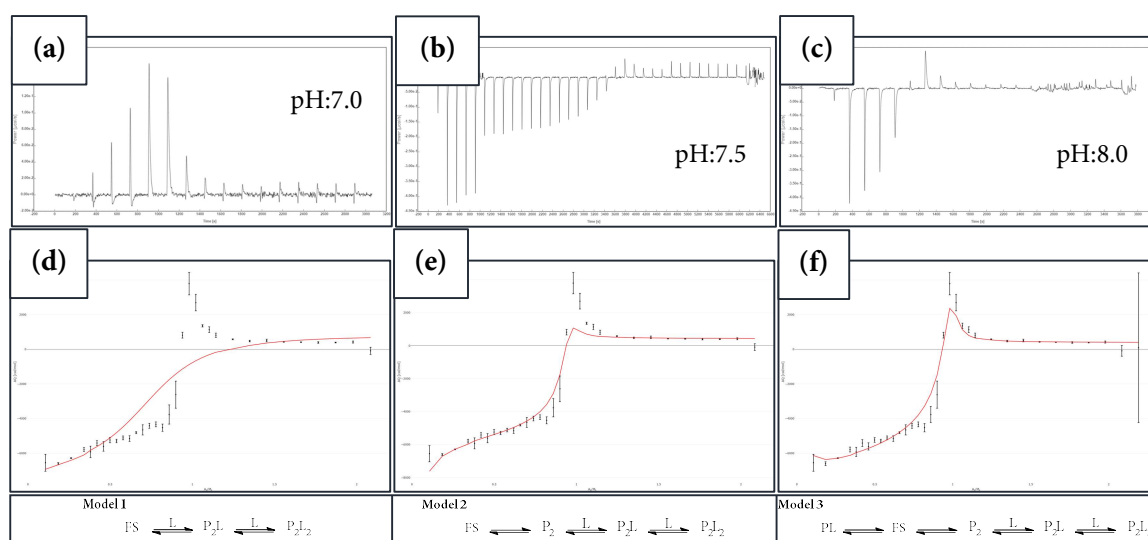
Figure 4.1. ITC data of processed isotherm of titration experiment of RIS into *human* FPPS. performed at **(a)** pH: 7.0, **(b)** pH: 7.5 and **(c)** pH: 8.0. Notably, the titration is pH dependent. The isotherm showed an endothermic profile at pH: 7.0 **(a)** while an exothermic followed by an endothermic profile at pH: 7.5 **(b)** and 8.0 **(c)**. The following models were assigned to the titration experiments of RIS into *human* FPPS applying different models performed at pH: 7.5. **(d)** Fitted isotherm by applying Model 1 which involves the binding of two subsequent RIS molecules (L) to the two monomers (P) that form the homodimer (P_2). **(e)** Fitted isotherm by applying Model 2 which involves the dimer formation first (P_2), followed by the binding of two subsequent RIS molecules (L) that form the homodimer (P_2). **(f)** Fitted isotherm by applying Model 3

which involves Model 2 and the concurrent dimer dissociation in the monomer-RIS complex (PL). FS: free specie; L: ligand (RIS); P: FPPS protein as a monomer; P_2: FPPS protein as a dimer; PL: FPPS monomer in complex with one molecule of RIS. P_2L: FPPS dimer with a bound RIS molecule in one of the two available active sites (one active site per monomer). P_2L_2: FPPS dimer with two RIS molecule occupying both active sites (one for each monomer).

In particular, it was questioned whether FPPS is really a stable homodimer or instead in equilibrium with its monomeric form. Accordingly, Mass Spectroscopy (MS from now on) experiments were performed to elucidate the behavior of the enzyme and characterize it more deeply.

4.1   Evaluation of FPPS apo-structures

The study was focused on investigating a putative monomer/dimer (M/D from now on) equilibrium mechanism by measuring: (i) the M/D ratio; (ii) the concentration dependency of the M/D ratio; (iii) kinetic exchange experiments. $^{13}C^{15}N$-labeled and unlabeled *human*, *T. cruzi* and *T. brucei* FPPS were the species under investigation.

Initially, as a control experiments, denaturing MS was performed to calculate the exact mass of the protein species, to check whether they are really homodimers in which the two monomers are connected by weak interactions, or whether they are instead linked by more stable and covalent interactions which include in MS not only disulfide bonds but also any kind of strong no covalent interactions.

The amount of 125 µM of protein was solubilized in a denaturating solvent containing a mixture of acetonitrile/water/acetic acid in the ratio 50/50/1 at pH 2.6. The high acidity of the solution increases the number of charges in the system which, due to the strong Coulomb forces, break non-covalent interactions. 20 µL of sample were then injected in the MS system and measured at a final concentration of 10 µM. Under denaturing condition, all three FPPS species were present in a monomeric form, indicating that the two monomers really form the homodimer *via* non-covalent interactions. Subsequently, native nano ESI-MS experiments were carried out, using both low-resolution (LCT) and high-

resolution (Synapt G2) instruments. Under these conditions the folding of the protein remined intact. The protein samples were solubilized in Caesium iodide (CsI) solution at a final concentration of 10 µM and the experiments were performed at different cone voltages (80 V, 100 V, 120 V and 200 V). The higher the applied voltage, the higher the energy is in the system and the better are the separation and resolution of the peaks. In contrast, an applied low voltage results in broad, overlapping and not well-resolved peaks from which it is difficult to calculate the exact mass. On the other hand, an excessively high voltage could lead to an artificial rupture of the dimer in the two monomers. It was therefore necessary to find a compromise between a voltage sufficiently high to ensure high resolution leading to a precise calculation of the exact mass but low enough that no breaking of the weak interactions occurred forming the dimer. All three species under investigation showed a similar behavior. At all applied voltages, the most abundant species in the MS spectrum was the dimer (around 5000 m/z). At lower voltages, a small fraction of a tetramer appeared, probably due to the gas phase conditions, while at higher voltages, the monomer was also present (Figure 4.2.).
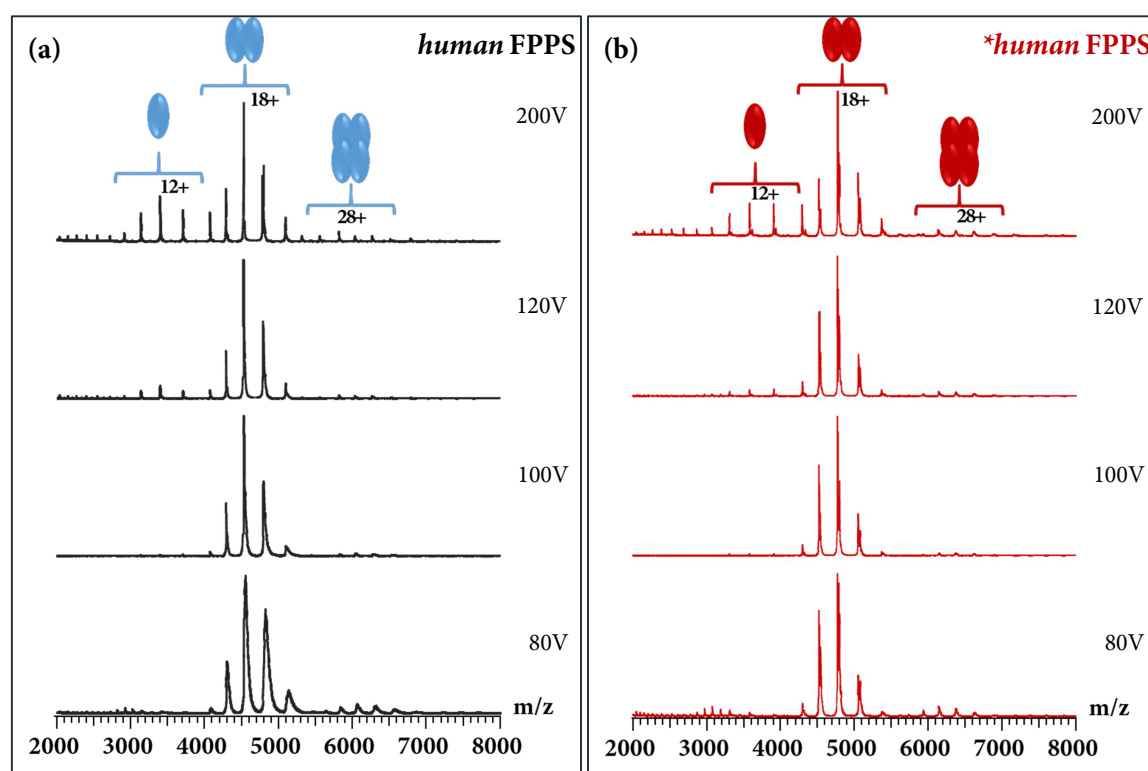
Figure 4.2. Native nano ESI-MS spectra of **(a)** unlabeled and **(b)** $^{13}C^{15}N$-labeled *human* FPPS recorded at different cone voltages. The dimer (5000 m/z) is the predominant species. * : $^{13}C^{15}N$-labled protein.

The MS spectra for the *T. cruzi* and *T. brucei* species and the exact calculated masses for all species under investigation are reported in the Appendix (Section 7.4.1. and 7.4.2.).

Subsequently, a putative concentration dependency of the M/D ratio was also evaluated. In particular, the protein species were tested at 1 µM, 5 µM, 10 µM and 20 µM and at different cone voltages (100 V, 150 V, 200 V and 250 V). The dimer was the predominant species among all three species under investigation at all tested protein concentrations and voltages. Interestingly, it was observed that, as the protein concentration increased, the intensity of the peak corresponding to the tetramer increased while at lower protein concentrations it almost disappeared. It therefore appeared likely to be a complex which is formed concentration-dependent only in the gas phase at high protein concentrations.

Furthermore, kinetic exchange experiments were performed by mixing $^{13}C^{15}N$-labeled and unlabeled protein in a ratio of 1:1 to investigate if the homodimer subunits exchange after a certain period of time.

The protein mixtures were therefore incubated for either 20-30 minutes, two hours and 24 hours. As shown in Figure 4.3. and despite their structural similarity, *T. cruzi* and *T. brucei* FPPS showed a distinct behavior. No exchange between $^{13}C^{15}N$-labeled and unlabeled *T. cruzi* FPPS was observed in the dimer region, even after 24 hours of reaction time whereas the subunits of *T. brucei* FPPS did exchange after 24 hours resulting in a new species, the heterodimer whose peak was found at 4800 m/z, exactly between the one corresponding to the fully $^{13}C^{15}N$-labeled (4900 m/z) and unlabeled (4700 m/z) protein species.
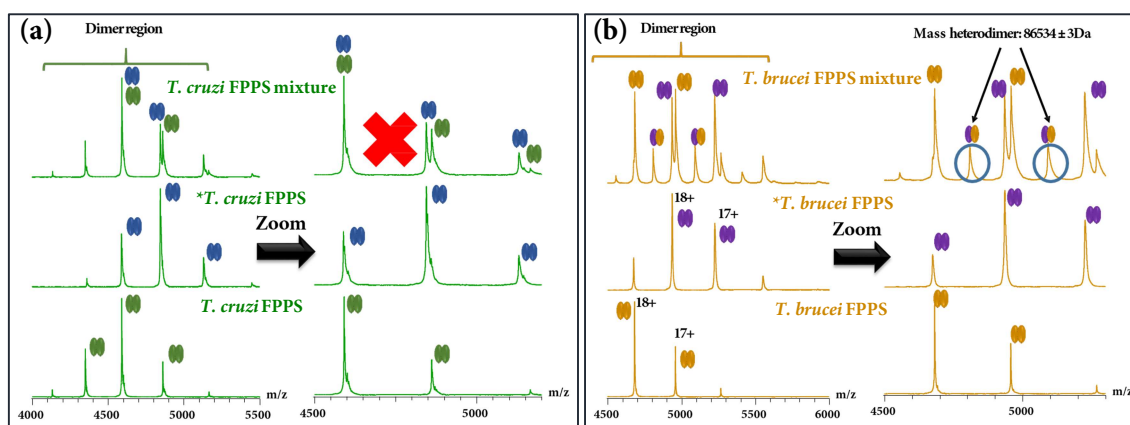
Figure 4.3. Kinetic exchange experiments of mixed unlabeled and $^{13}C^{15}N$-labeled **(a)** *T. cruzi* and **(b)** *T. brucei* FPPS. In the dimer region, the peak referring to the heterodimer is encircled in blue. ⋆ ⋆ : $^{13}C^{15}N$-labeled protein.

The *human* FPPS showed a similar behavior as *T. cruzi* FPPS. The MS spectra are reported in the Appendix (Section 7.4.3.)

Considering the different species' behavior in kinetic studies, the protein conformational change was also evaluated by Ion Mobility Spectroscopy (IMS) with which it is possible to assess their Collision Cross Section (CCS) and Collision Induced Unfolding (CIU).

The CCS represents the effective area of the protein ion colliding with the molecule of buffer in the gas phase. Usually, the higher the CCS, the higher is the number of collisions between protein and buffer molecule in the gas phase [66]. The CCS of a stable and compact dimer, characterized by strong intramolecular interactions, is therefore smaller than that of an extended, flexible and loosely packed dimer, which, by exposing more surface area, increases the chances of colliding with the buffer molecule in the gas phase. In a similar manner, CCS may increase or decrease due to protein conformational changes. In Figure 4.4., the CCS of all species under investigation is plotted against the charged state of the ions. However, no global conformational change was indicated. The differences in CCS were only due to the different exact mass of the FPPS species. In fact, the exact mass of $^{13}C^{15}N$-*T. brucei* FPPS is the highest while the one of unlabeled *human* FPPS is the lowest among all species under investigation.
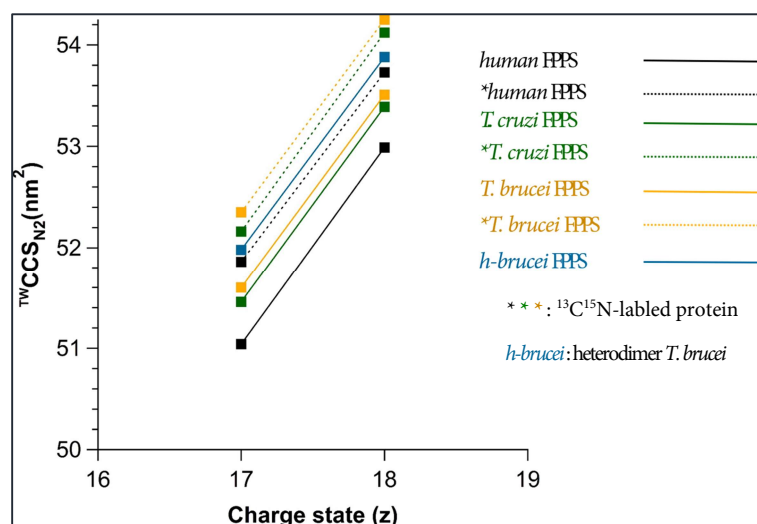
Figure 4.4. CCS as a function of the charge state of the protein ions. Differences in CCS only depend on different exact masses of the species under investigation. No global conformational change was detected. $*\,*\,*$: $^{13}C^{15}N$-labled protein. *h-brucei*: heterodimer *T. brucei* FPPS.

The protein dynamic and stability in the gas phase was also evaluated by CIU experiments. The native structure is usually kept at low charged state and it corresponds to the ground state in Figure 4.5. The ions are then activated through energetic collision with gas atoms such as Argon in order to increase their internal energy content and induce their unfolding [67]. In fact, by increasing the collision energy the proteins traverse different transition states that lead to different states of unfolding. All proteins under investigation showed similar CIU fingerprints. They started unfolding around 50 V, then traversed two or three transition states with the last one representing highest degree of unfolding. The only difference among the species is that *human* and *T. brucei* FPPS passed through more transition states during the unfolding mechanism compared to *T. cruzi* FPPS (Figure 4.5.).
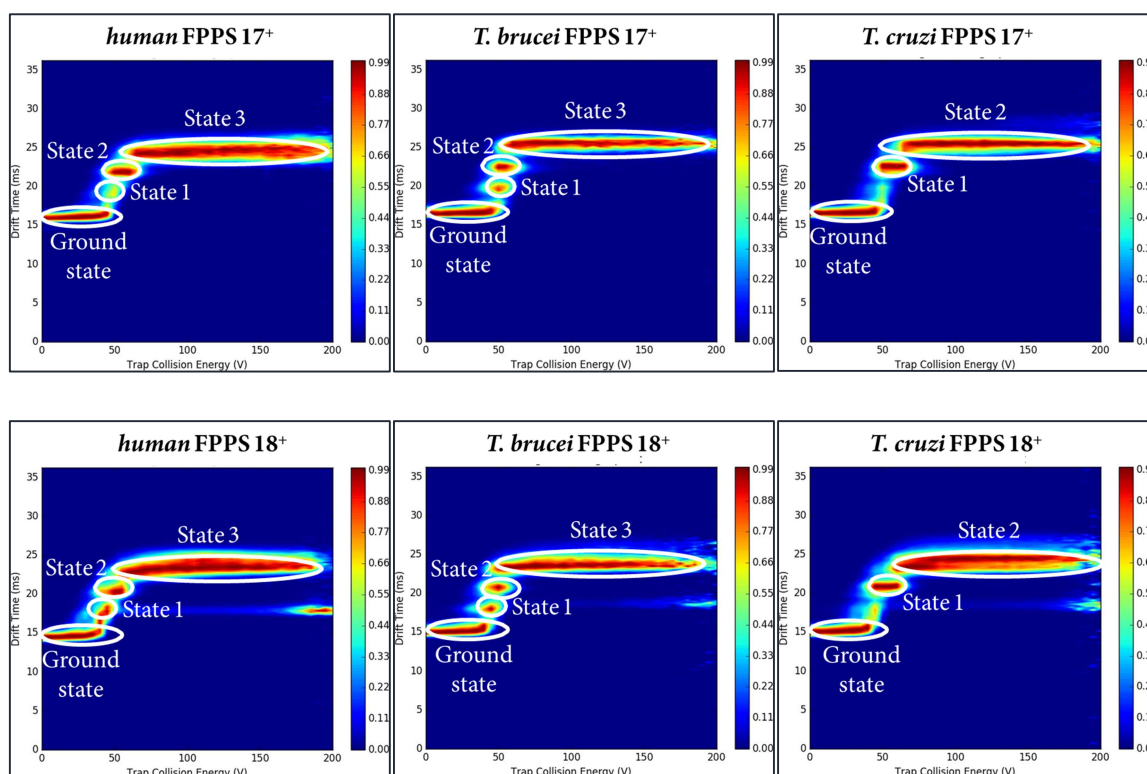
Figure 4.5. CIU fingerprints of *human*, *T. brucei* and *T. cruzi* FPPS at two different charged states (17⁺ and 18⁺). *Human* and *T. brucei* FPPS passed through three transition states, whereas *T. cruzi* FPPS only traversed two.

4.2   FPPS in complex with Risedronate

The above-described experiments revealed that, with the exception of *T. brucei* FPPS, *human* and *T. cruzi* FPPS are stable dimers in the gas phase. However, to better elucidate the unusual behavior of *human* FPPS in complex with RIS during ITC experiments (Figure 4.1.), the FPPS-RIS complex stability was further investigated by MS. In fact, since the protein alone did not show any conformational transition, it was then questioned whether the ligand itself may stabilizes the dimer or rather promote the dimer dissociation.

Kinetic studies on the three protein under investigation were then performed in order to elucidate how the binding occur. In particular, since ITC experiment may indicate that more than one binding event may occur at the same time, it was also questioned whether two molecules of RIS bind at the same time to each monomer or if they bind one after the other, even in combination with a conformational change.

The amount of 5 µM of protein was then incubated for 15 minutes with RIS in ratios of 1:0.5, 1:1, 1:2, 1:3, 1:4, 1:5, 1:10 and 1:20. In particular, the 1:0.5 ratio assumes that one molecule of RIS binds to one FPPS monomer whereas the 1:1 ratio assumes that one molecule of RIS occupy each of the two active sites present in homodimer. The 1:10 and 1:20 ratios are therefore measured in excess of ligand. As displayed in Figure 4.6., the 1:10 ratio already showed an unspecific binding in which between five to eight molecules of RIS accommodated the complex with *human*, *T. cruzi* and *T. brucei* FPPS.

Legend: ● 84254 ± 8 Da 1 RIS   ● 8549 ± 6 Da   2 RIS   ● 84851 ± 8 Da 3 RIS   ● 85159 ± 6 Da 4 RIS   ● 85427 ± 10 Da 4 RIS   ● 85748 ± 9 Da 5 RIS
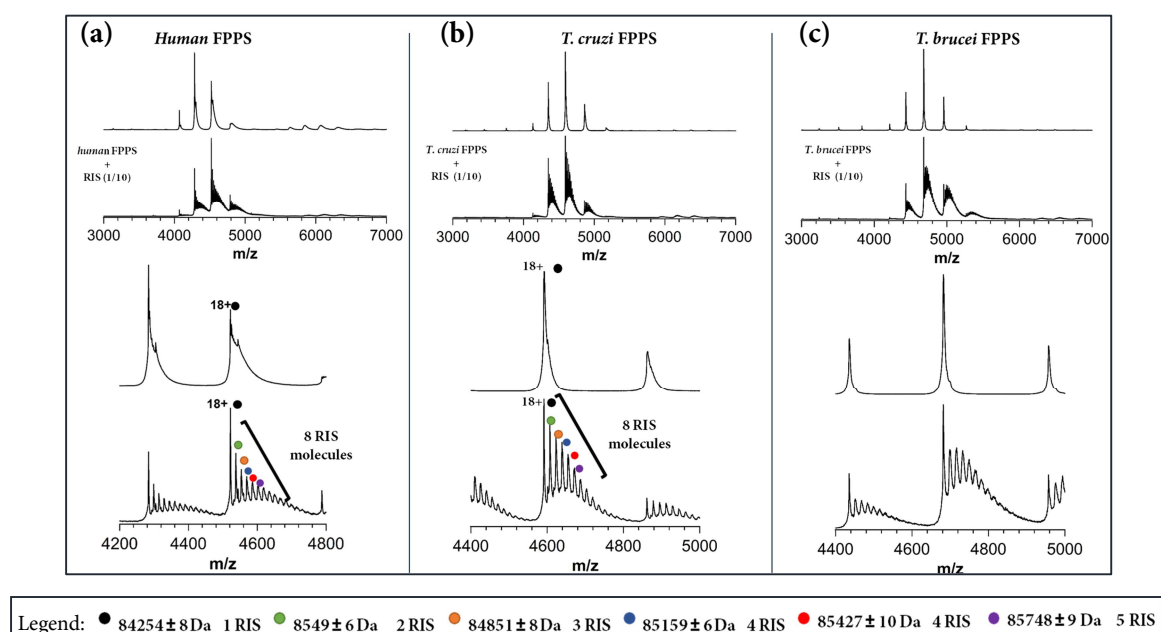
Figure 4.6. In the upper part, native MS of **(a)** *human*, **(b)** *T. cruzi* and **(c)** *T. brucei* FPPS in complex with RIS at a 1:10 ratio. In the lower part, zoom-in of the FPPS-RIS complex at a charged state of 18+. The colored cycles indicated FPPS in complex with different RIS molecules.

It is important to note that these experiments were performed without Mg^{2+} ions as they could clog or damage the MS instruments. However, as previously reported in Section 2., Mg^{2+} ions are of fundamental importance for the catalytical mechanism of the FPPS species. In addition, it is known from literature that RIS binds FPPS only in presence of Mg^{2+} ions by complexing them in a similar manner as the DMAPP/IPP does in the allylic pocket (Section 2.6.). It was therefore decided to add 0.5 µM of magnesium acetate to the 1:5 ratio. As shown in Figure 4.7., the mass spectra are completely different depending on the presence of absence of Mg^{2+} ions. In particular, the spectrum is much more homogeneous

in all three species under investigation in presence of Mg^{2+} ions. In addition, only one population is present in the MS spectra whereas in absence of Mg^{2+} ions there are up to five or six RIS molecules binding to the FPPS homodimer. Another important difference is that the apo-FPPS species is not observed in presence of Mg^{2+} ions while it is present in the absence of Mg^{2+} ions. The presence of Mg^{2+} ions probably shifted the equilibrium in the direction of the FPPS-RIS complex formation.
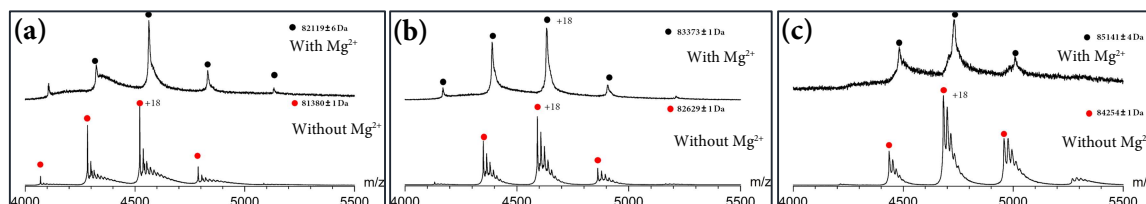
Figure 4.7. Native MS of **(a)** *human*, **(b)** *T. cruzi* and **(c)** *T. brucei* FPPS in complex with RIS at a 1:5 ratio. Interestingly, the MS shows different behavior in presence (black) and in absence of Mg^{2+} ions. Only one population is observed in presence of Mg^{2+} ions while several ones are observed in absence of Mg^{2+} ions.

At the moment, no data are available concerning the 1:0.5, 1:1, 1:2, 1:3 and 1:4 ratios.

Thereby, it is possible to conclude that the presence of Mg^{2+} ions modify the complexation mechanism of FPPS species with RIS. The metal ion can act as a catalyst or it can even be directly involved in the complexation reaction. However, the system under investigation revealed to be quite complex and no evidence of the exact number of RIS molecules and Mg^{2+} ions appeared in the MS spectra.

MS experiments were performed in close collaboration with Dr. Hernandez-Alba at the LSMBO in Strasbourg and were part of a joint EU-ITN project with Joy Petrick (EU-ITN AEGIS, ESR 12) and Lena Muenzker (EU-ITN FRAGNET) from NIBR.

5. Conclusions and Outlook

FPPS is a key enzyme involved in the mevalonate pathway and is essential for sterol production. *T. cruzi* and *T. brucei* parasites cause CD and HAT respectively, and inhibition of FPPS in these species is lethal and therefore a valid target to fight these tropical diseases. BPs are known active-site binders but due to their inadequate pharmacokinetic properties, new chemical entities with new scaffolds and/or different mechanism of action are needed. The discovery of an allosteric pocket in *human* FPPS by *Jahnke et al.* [51] pioneered the search for new molecules in this respect.

The present work provided fundamentals knowledge for a future drug-discovery project by systematically analyzing samples of proteins to study their properties either in solution, in the crystalline state as well as in the gas phase using different techniques.

In particular, the current study was initially focused on optimizing the stability of protein samples in solution, an essential precondition for crystallization and subsequent soaking experiments in the framework of a fragment screening project, by TSA and LS. The results revealed under which conditions *human*, *T. brucei* and *T. cruzi* FPPS are more stable and what are the ideal buffer conditions applicable for storage. In addition, these studies have confirmed the importance of cofactors and particularly magnesium ions for the stability of the protein samples.

Moreover, the protein species' soakability was evaluated both experimentally and by the use of a computational packing analysis using Xtunnel.

Subsequently, the stability of the protein crystals and their tolerance towards DMSO to support sufficiently high concentrations of fragment solutions for soaking experiments, were investigated. This study allowed to set up a crystallization and soaking protocol to successfully screen the 96-fragment library against *T. cruzi* FPPS. The fragment screening project resulted in three crystallographic hits, one of which binds to the allosteric pocket. The other two occupy another pocket whose function is still unknown. **J82**, the allosteric

binder, was particularly interesting considering the aim of the project (Section 1.4.). Its binding mode was further investigated by soaking *T. cruzi* FPPS crystals with different derivatives as well as with a small number of follow-up compounds in order to investigate, which interactions were essential to bind to the allosteric pocket. This definitely confirms J82 as a valid hit and not as a singleton as frequent binder.

Moreover, three additional hits were found by a PanDDA analysis. In particular, **J59** which binds the characteristic Trypanosomal 11-residues insertion loop could be a promising starting point for the development of selective compounds against *T. cruzi* FPPS and possibly to interfere with Chagas disease.

Furthermore, the role of magnesium ions as cofactor of the enzyme was elucidated, which led to the first *T. cruzi* FPPS holo-structure as well as the first FPPS-IPP-Mn^{2+} tertiary complex.

Finally, the protein species were also characterized by MS to investigate their stability, putative conformational changes and behavior in the gas phase.

In the present study, the protein species were characterized in detail with different methods and their behavior was elucidated both in solution and gas phase. For the matter of time, only small fragments were soaked in *T. cruzi* FPPS crystals. However, the discovered hits could serve as promising candidates for the design of a new series of lead compounds.

6. Materials and Methods

6.1 Protein expression and purification

The protein material was initially expressed and purified during a three-months secondment at the collaboration site "Novartis institute for Biomedical Research (NIBR)" in Basel. Additional protein material for subsequent studies was provided by both, Joy Petrick (EU-ITN AEGIS, ESR 12) and Lena Muenzker (EU-ITN FRAGNET). In addition, a new protocol for in-house purification was set up by modifying and adapting the protocol provided by industry to our in-house equipment obtaining a similar yield.

6.1.1 Protein expression

For protein expression, the competent cells BL21(DE3) strain of *E. coli* and the plasmids pLAF2285, pLAF2286 and pLAF2287 for *human*, *T. cruzi* and *T. brucei* FPPS, respectively, were used. The plasmids consisted of a T7 promoter, N-terminal 6His-Tag, a PreScission/HRV 3C cleavage site and a resistance vector against Kanamycin.

The protein expression protocol comprises four steps: (i) transformation; (ii) overnight culture; (iii) fermentation and (iv) harvesting.

(i) Solid and cold LB medium, previously prepared and autoclaved, was heated up in the microwave for about two minutes. Once cooled down, 200 μL of Kanamycin stock solution (50 mg/mL) was added to 200 mL of liquid LB medium for a final concentration of 50 μg/mL. The LB medium-Kanamycin mix was then poured into 100 x 15mm agar plates. It was let to cool down until the agar solidified. In the meantime, a vial containing 50 μL of competent cells was placed on ice. Then, 1 μL of plasmid was added to the cells and mixed gently by tapping. The vial was incubated on ice for 20 - 30 minutes. The cells were then heat-shocked by incubating the vial for one minute in a 42 °C water bath and sequentially quickly placed on ice for about five minutes. After the addition of 400 μL of Super Optimal

broth with Catabolite repression medium (SOC medium), the vial was placed in a shaking incubator at 37 °C for one hour at 500 rpm (Thermomixer comfort 2 mL, EPPENDORF). 150 µL of the transformation reaction solution was then placed in the center of an agar plate and spread all over with a cell spreader. The agar plate was then inverted to avoid condensation and incubated (incubator INCU-Line, VWR) at 37 °C overnight.

(ii) The next day, some colonies appeared on the agar plate, one of which was picked using a Microstreaker® and transferred into a Falcon tube (DB falcon 14mL, Falcon™ Round-Bottom Polypropylene Tubes) which contained 2 mL of LB medium and 2 µL of Kanamycin. This pre-culture tube was then incubated at 37 °C for about six hours in a shaking incubator at 200 rpm (Shaker X, KUHNER).The pre-culture was then transferred into a Corning Erlenmeyer flask containing 40 mL of metal, aspartic acid, glucose media (MDG medium), which was placed in the shaking incubator and cells were grown overnight.

(iii)The next day, the OD_{600} of the pre-culture (diluted 1:10 with Milli Q water) was checked and, at a value around 7, was transferred into the fermenter which contained 1.5 L of modified auto-inducing medium (modAI medium). The fermenter was set to 37 °C, pH 7.0, 900 rpm, 80 % of pO_2 and 2.5 nL/min of air flow. As soon as OD_{600} reached 10, the temperature was reduced to 18 °C and stirred overnight.

(iv) The next day, the OD_{600} was checked and an equal volume of culture was collected in two sterilized plastic bags for centrifugation (WHIRL-PAK, NASCO). The culture was then centrifuged for 30 minutes at 3500 rpm, 4 °C (Sorvall RC3BP). Alternatively, centrifugation bottles (1000 mL, NALGENE) can also be used and centrifugated for 20 minutes at 6000 rpm, 4 °C (Sorvall Lynx 6000, THERMO SCIENTIFIC). After the removal of the supernatant, the cell pellet was collected in a plastic bag and stored at -80 °C.

6.1.2   Protein purification

The pellet was resuspended in 500 mL of lysing buffer (LB from now on) consisting of 300 mM NaCl, 50 mM Tris, 5 mM Imidazole, 10 %(v/v) GOL and 0.5 mM TCEP. 10 μL of Benzonase (Benzonase® Nuclease, 5KU, Sigma-Aldrich/Merck) and five cOmplete™ Protease Inhibitor Cocktail EDTA-free Tablets (Roche, Germany) were added to the solution. The cells were then homogenized using a polytron disperser (Sonicator), lysed using an EmulsiFlex (EmulsiFlex-C5™, AVESTIN, Mannheim, Germany) and then centrifugated at 10000 rpm at 4 °C for 60 minutes (Avanti™ J25, BECKMAN COULTER™). The supernatant was then collected and filtrated with a 0.45 μm filter (Ultrafree®-MC filters, pore size 0.45 μm, Merck).

For protein purification, 100 mL of supernatant was loaded onto a 5 mL Ni-NTA column (HisTrap FF, GE Healthcare Life Science, Freiburg, Germany) and eluted (1 mL/min, 0.5 mPa) with a buffer consisting of 150 mM NaCl, 50 mM Tris, 200 mM Imidazole, 10 %(v/v) GOL and 0.5 mM TCEP. The protein elution was performed by applying a gradient of Imidazole from 0 mM to 200 mM in one hour. The eluted protein solution was then transferred in a tubular membrane (Regenerated Cellulose Tubular Membrane MWCO: 12000-14000, Roth™) with 150 μL of PreScission/HRV 3C protease and dialyzed overnight at 4 °C against 4 L of buffer (300 mM NaCl, 50 mM Tris, 10 %(v/v) GOL and 0.5 mM TCEP) both to cleave off the N-6His-Tag and remove the imidazole. To remove the protease as well as putative unspecific proteins, the sample was loaded on another 5 mL Ni-NTA column and the flow through, containing the targeted protein, was collected and concentrated by centrifugation (Heraeus® Multifuge® 3SR, The Thermo Scientific) to a volume of 10 mL using a Vivaspin concentrator with 30kDa cut-off (Vivaspin®, 20 mL, MWCO 30000, Sartorius Stedim Biotech GmbH). As additional purification step, a size-exclusion chromatography was performed. The protein was loaded onto a SEC column (Superdex 200, GE Healthcare Life Science, Freiburg, Germany) and eluted with a running buffer consisting of 200 mM NaCl, 50 mM Tris pH 8, 2 mM TCEP or in TSA-optimized

buffer (Section 2.2.). The protein solution was concentrated (Heraeus® Multifuge® 3SR, The Thermo Scientific) to 6 mg/mL, flash frozen in liquid nitrogen and stored at -80 °C.

6.2 TSA

TSA experiments were performed on a real-time PCR system (QuantStudio™ 3, 12K Flex system) using QuantStudio™ Design & Analysis software (version 1.4.2.). The typical workflow consists of sample preparation, temperature scan run and data analysis.

The assay was performed in a 96-well PCR plate with a total volume of 20 μL. Each well consists of 2.0 μL of protein sample at a concentration between 3 mg/mL and 5 mg/mL, 0.3 μL of 1:10 dilution of dye (SYPRO® Orange) in water and 17.7 μL of buffer with different formulation.

The experimental procedure consists of three step: (i) In the first step, the protein sample was subjected at a heating rate of 1.6 °C/s until a temperature of 20 °C was reached and kept constant for one minute. (ii) In the second, also called dissociation step, the temperature was increased by 0.05 °C/s until a temperature of 95 °C was reached and kept constant for one minute. (iii) In the final step, the temperature was decreased by 1.6 °C/s until 20 °C and kept constant for one minute. The wavelength of the fluorescence scan for excitation and emission were set to the spectroscopic maxima of SYPRO® Orange (472 nm and 570 nm, respectively).

The melting curves were analysed using Protein Thermal Shift Software (version 1.3) from Thermo Fisher Scientific. In a typical melting curve, the fluorescence intensity is plotted as a function of temperature. The resulting curve is sigmoidal in ideal case and can be used to estimate the melting temperature T_m of the protein of interest, which corresponds to the time axis at the inflection point (Figure 5.1., upper pannel).

As shown in Figure 5.1., no fluorescence was detected at RT and until 40-45 °C because the protein was correcly folded and the hydrophobic regions are oriented towards the inner

core of the protein. As the temperature increases, the protein begins to unfold and exposes its hydrophobic parts towards the surface to which the dye then binds and modify the spectroscopic properties. After reaching the maximal fluorescence, a decay in fluorescence intensity is observed, which is due to thermal motions of the fluorescent dye, protein aggregation and dye dissociation. The numerical derivatives of the melting curves are also calculated to better visualize the rate of change (Figure 5.1., lower pannel). In this case, T_m corresponds to the maximal peak height.
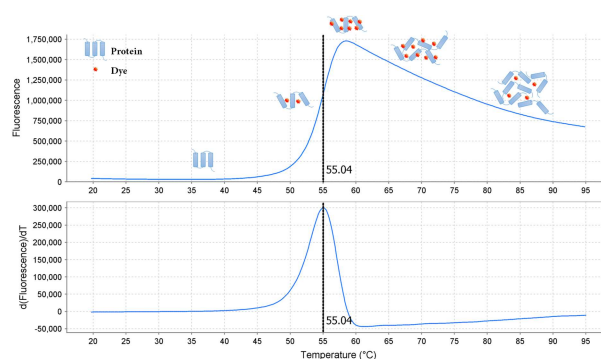
Figure 5.1. TSA melting curve plot (upper pannel) and derivative plot (lower pannel) showing the T_m calculated as midpoint of the inflection curve and peak height, respectively (black vertical dashed line). The T_m corresponds to the temperature at which the hydrophobic areas of the protein are exposed to the environment and the dye binds. After the maximal fluorescence is reached, the protein aggregates and the dye dissociates, causing a decrease in fluorescence (image after 58 °C).

6.3   SEC-MALS

SEC-MALS experiments were performed on miniDAWN™ TREOS® system from Wyatt Technology at the Freie Universität Berlin in the working group of Prof. Markus C. Wahl, under the supervision of Dr. Jan Wollenhaupt.

Prior to the experiments, the purity of the samples was checked in-house using the analytical column Superdex® 200 Increase 10/300 GL (Bed volume= 24mL) purchased from GE Healthcare. The protein concentrations ranged between 3 mg/mL and 5 mg/mL with 250 μL injection volume and an elution flow rate of 0.6 mL/min. The chromatograms are shown in the Appendix 6.1.1.

The MALS device was coupled with an HPLC (UV Instrument: G1314F). The loaded protein concentration was 5mg/mL for *human* and *T. brucei* FPPS and 6 mg/mL for *T. cruzi* FPPS. The injection volume was 60 µL at flow rate of 0.6 mL/min for all species under investigation. From the HPLC, the samples passed through the static light scattering instrument (MALS, 879-TS, wavelength: 660.0 nm) and the generated signal was detected by three detectors with a scattering angle of 43.6°, 90.0° and 136.4° degrees. For dynamic light scattering experiments (DLS, model: Wyatt QELS+) the signal was detected at 90.0° degrees. Finally, the sample was conveyed to a refractometer (RI Instrument: ERC RefractoMax520) and data collected.

The data were processed and analyzed using ASTRA® software (version 6.1.4.25) from Wyatt Technology. Only one sharp peak was identified proving that the protein samples were homogeneously monodispersed in the TSA-optimized buffer. The processed results are reported in the Table below (Table 5.1.).

Table 5.1. Parameters determined by SEC-MALS experiments-

		human FPPS	*T. brucei* FPPS	*T. cruzi* FPPS
Calculated Mass (µg)		214.05	226.49	308.71
Molar mass (g/mol)	M_n	7.777×10^4 (±0.069%)	8.371×10^4 (±0.066%)	8.228×10^4 (±0.097%)
	M_w	7.778×10^4 (±0.069%)	8.371×10^4 (±0.066%)	8.228×10^4 (±0.096%)
Polydispersity (M_w/M_n)		1.000 (±0.098%)	1.000 (±0.093%)	1.000 (±0.137%)
Radius of gyration R_g (nm)		2.6 (±31.4%)	3.2 (±1.9%)	2.8 (±44.6%)
Hydrodynamic radius R_h (nm)		3.747 (±1.973%)	3.706 (±1.981%)	3.675 (±1.904%)

6.4   Crystal preparation and soaking

T. cruzi FPPS crystals were obtained using the sitting drop vapor diffusion method at 19 °C. Two crystallization contition were used: (i) the optimized JCSG Core II G8 and (ii) the seeding conditions(Section 3.2.1.). (i) For the first crystallization condition, 1.2 µL of 6 mg/mL protein solution (0.2 M NaCl, 0.05 M Tris pH 8.0, 0.002 M TCEP) were mixed in the drop well with the same volume of crystallization buffer (0.18 M $(NH_4)_2SO_4$, 0.08 M

NaOAc pH 5.50, 20 % PEG 4K, 20 % Glycerol). The reservoir wells were filled with 500 μL of crystallization buffer and properly sealed using a clear tape. *T. cruzi* FPPS crystals appeared after two days and stopped growing after three days. (ii) In the second crystallization condition, the crystals were obtained by microseeding technique. In particular, 1.2 μL of 12 mg/mL protein solution (0.025 M NaCl, 0.010 M Tris pH 7.40, 0.002 M TCEP) was mixed with 0.8 μL of crystallization buffer (0.004 M $ZnSO_4$, 0.008 M MES pH 6.50, 12 % PEG MME 550, 12 % Glycerol) and with 0.4 μL of seed-stock solution. The reservoir consisted of 500 μL of crystallization buffer. The seed stock solution was obtained from crystals previously broken into small pieces and diluted to 1:10, 1:100 and 1:1000 with the reservoir buffer. *T. cruzi* FPPS crystals for seeding were obtained by mixing 1 μL of 6 mg/mL protein solution (0.2 M NaCl, 0.05 M Tris pH 8.0, 0.002 M TCEP) with the same volume of crystallization buffer (0.16 M $(NH_4)_2SO_4$, 0.08 M NaOAc pH 5.00, 20 % PEG 4K, 20 % Glycerol). *T. cruzi* FPPS crystals appeared whitin five days and stopped growing after a week.

Prior crystallization, the protein solution was filtrated and sterilised using a 1.5 mL tube (Ultrafree®-MC filters, pore sizes 0.45 μm, Merck).

The mature *T. cruzi* FPPS crystals were used for soaking experiments. As reported in Section 3.2.3., the 1 M fragment stock solution solubilized in DMSO was diluted with the crystallization buffer to obtain a pre-mix with a final concentration of 250 mM. Then, a calculated volume of pre-mix was pipetted onto the rim of the crystallization drop to obtain a final fragment concentration of 50 mM. Depending on the tolerance of the crystals to the fragments under examination, the soaking time ranged from a few minutes to a maximum of 48 hours. Crystals were then transferred for a few seconds in a drop well consisting of 5 μL of crystallization buffer containing 20 % GOL prior to flash-freeze them in liquid nitrogen.

The crystallization and soaking were performed in a 24-well crystallization plate (24-well XRL Plate fromMolecular Dimension) and sealed using a clear tape (1.88 inch wide Crystal Clear Sealing Tape from Hampton Research).

6.5   Data collection and processing

Diffraction data for 3-nitrophenol, 4-nitrophenol, 2-methoxy-5-nitrophenol, **J51**, **J71** and apo-structure were collected on beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). Datasets were collected on a Dectris Pilatus 6M pixel detector at a wavelength of 0.9763 Å at 100 K. Diffraction data for **J82** were collected on beamline BL14.1 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin. The dataset was collected on a Dectris Pilatus 6M pixel detector at a wavelength of 0.91841 Å at 100 K [68]. XDS and XDSAPP were used for indexing, integration and scaling of the datasets [69],[70].

6.6   Structure determination and refinement

The identification of possible hits was first performed using the autorefine pipeline described by *Schiebel et al.*[63]. Whenever regions of interest were identified, the refinement was subsequently carried out manually.

The structures were solved by molecular replacement using Phaser [71]. Model refinement (xyz coordinates, individual *B*-factors, occupancies) was done with Phenix.refine using the version 1.15.2-3472 and model building into electron density maps (2mFo–DFc and mFo–DFc) with Coot (0.8.9) [72], [73]

A randomly chosen subset of 5 % of the reflections was excluded from the refinement and used for the calculation of R$_{(free)}$. As a first refinement step, Cartesian simulated annealing was performed (default settings). *B*-factors for all model atoms (except for hydrogen atoms) were refined with TLS refine. Hydrogen atoms (riding model) were added to the amino acids with Phenix.refine. Alternative conformations of amino acid side chains and ligand moieties were assigned if an occupancy of at least 20 % was obtained after refinement. The program *Chemoinfo* was used for SMILES code generation (www.cheminfo.org/flavor/malaria/Utilities/SMILES_generator_checker.html

), ligand molecules and restraints were created with the Grade Web Server (http://grade.globalphasing.org).

6.7 Experimental phasing by the SAD (single-wavelength anomalous dispersion) method

SAD experiments were performed on beamline BL14.1 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin. Datasets were collected on a Dectris Pilatus 6M pixel detector at a wavelength of 1.893174 Å which correspond to the absorption edge peak of Manganese.

As reported in Section 2.5. and 2.6., the holo-structure and the FPPS-IPP complex was obtained collecting a dataset from manganese-derivatized crystals with which SAD experiments were carried out and experimental phases were obtained.

The significant anomalous signal $[d''/sig(d'')_{holo} = 0.73\text{--}6.80; d''/sig(d'')_{FPP\text{-}IPP} = 0.68\text{--}3.59]$ allowed the determination by SHELXD ($CC_{max} = 42.5$ for holo structure; $CC_{max} = 44.9$ for FPP-IPP complex) of 2 out of 2 possible Mn sites for the holo-structure and 6 out of 6 Mn atoms sites for the FPP-IPP complex. SAD phasing was carried out using SHELXE [74]. An initial electron-density map was calculated and a model (one single chain; 425 residues;) was generated *via* ARP/wARP [75] as implemented in the CCP4 package [71]. Further model building and refinement were carried out using Coot [73] and PHENIX [72] as previously described in Section 5.6.

6.8 Accession code

Atomic coordinates and experimental details for the described crystal structures were deposited with the PDB and will be released upon publication under the assession codes 6SDN (**J82**), 6SDO (3-nitrophenol), 6SDP (4-nitrophenol), 6SDQ (2-methoxy-5-

nitrophenol), 6SE2 (apo-structure), 6SF8 (**J51**), 6SF9 (**J71**), 6SFA (holo-structure) and 6T7N (IPP).

6.9 Additional software used: DSX

The scoring function DSX (version 0.9) was used to calculate the per-atom score contributions. The putative ligand-protein complex geometry is automatically displayed as a PyMOL session [76].

The input file for DSX is the docking pose in format *.mol2* which was converted from *.sdf* file using fconv [77].

Docking was performed using HYBRID (version 3.3.0.3, OpenEye Scientific Software) [78].

Prior to docking, the receptor preparation was performed using the program pdb2receptor (OEDocking version 3.0.1. OpenEye Scientific Software, Santa Fe, NM, http://www.eyesopen.com). The output file is in *.oeb.gz* format. Ten conformers of **J82** were generated using OMEGA (version 3.1.0.3, OpenEye Scientific Software) [79].

For docking:

- The input file .oeb.gz contained the receptor under investigation;
- The input file .mol2 is the fragment **j82**.

The .sdf output file contained the fragment **J82** docked into *T. cruzi* FPPS and it was used for DSX calculation.

7. Appendix

7.1 SEC-MALS experiments

7.1.1 Chromatograms (Superdex® 200 Increase 10/300 GL)

Human FPPS

T. brucei FPPS
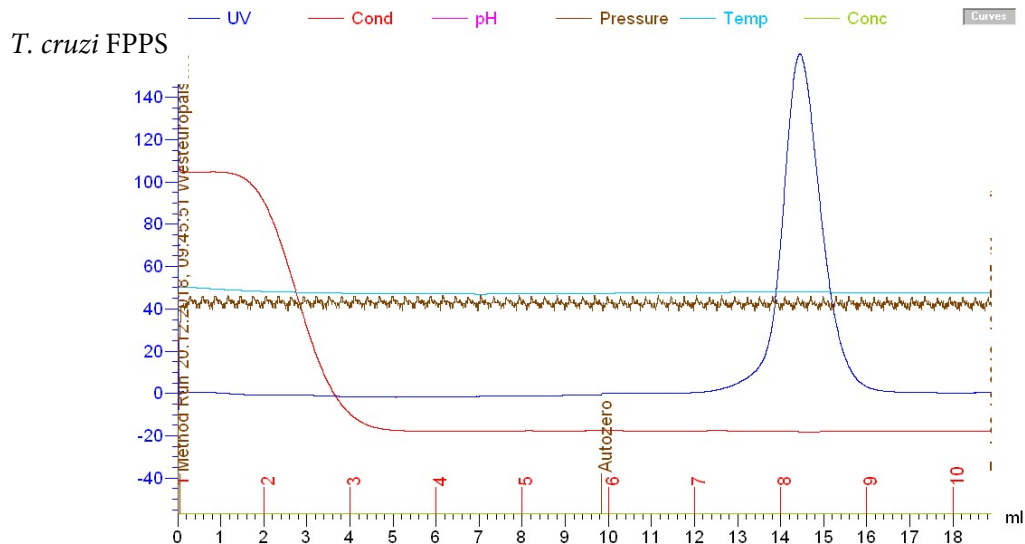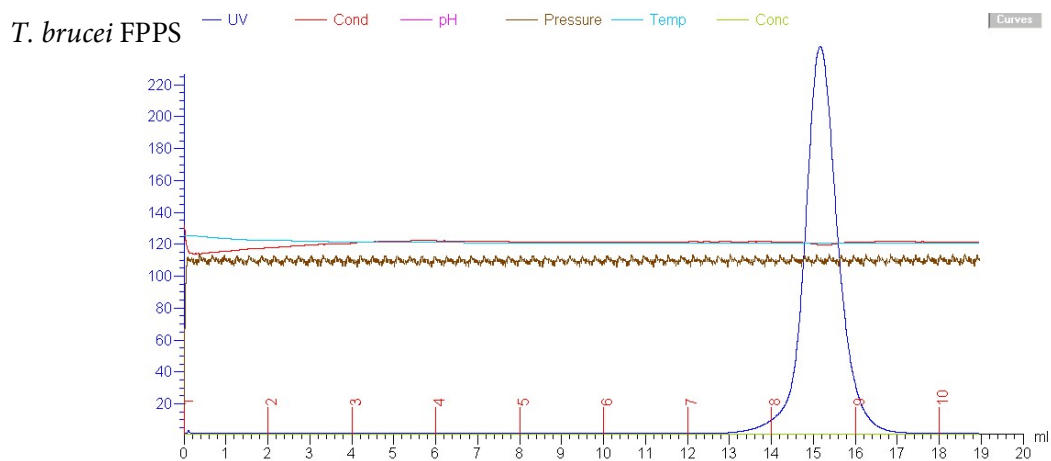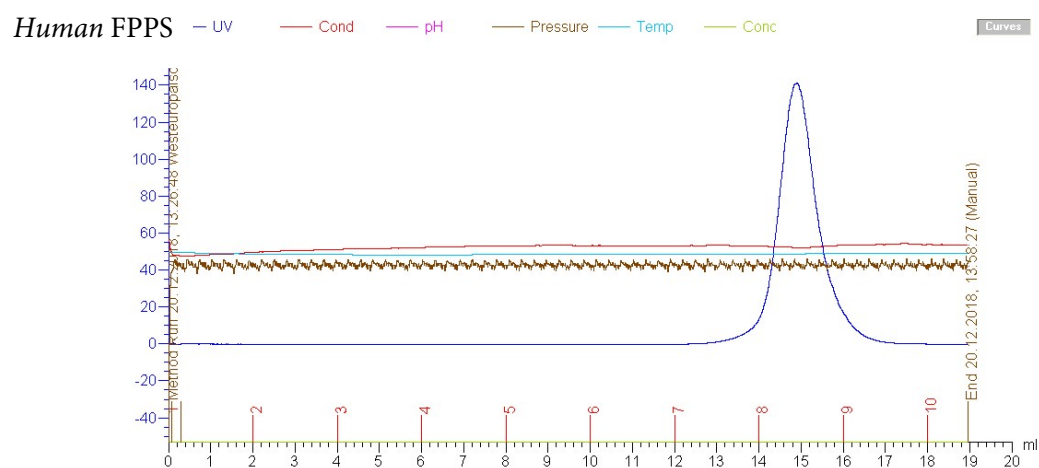
T. cruzi FPPS

7.1.2   SEC-MALS (miniDAWN™ TREOS® system)

Human FPPS
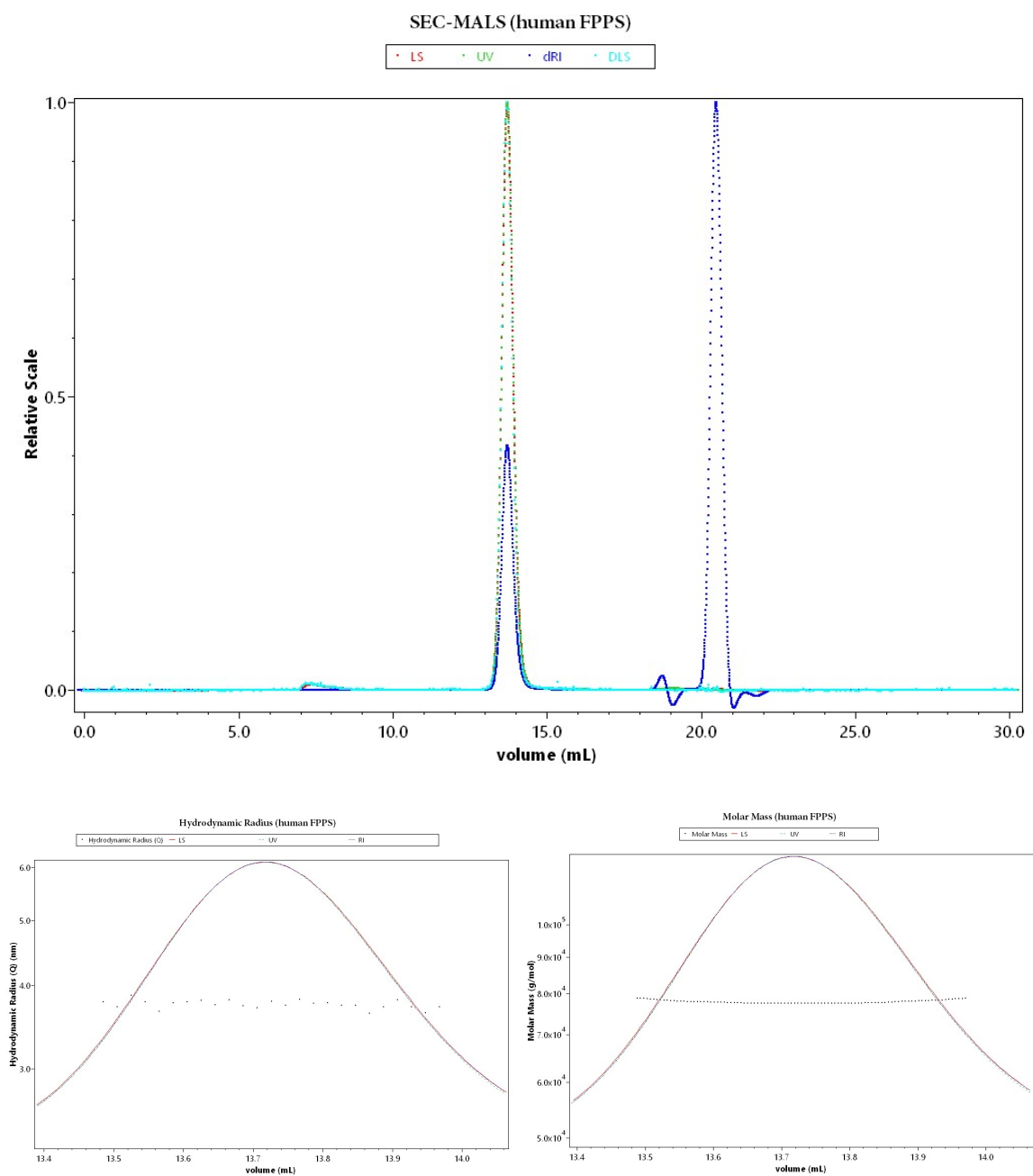
T. brucei FPPS
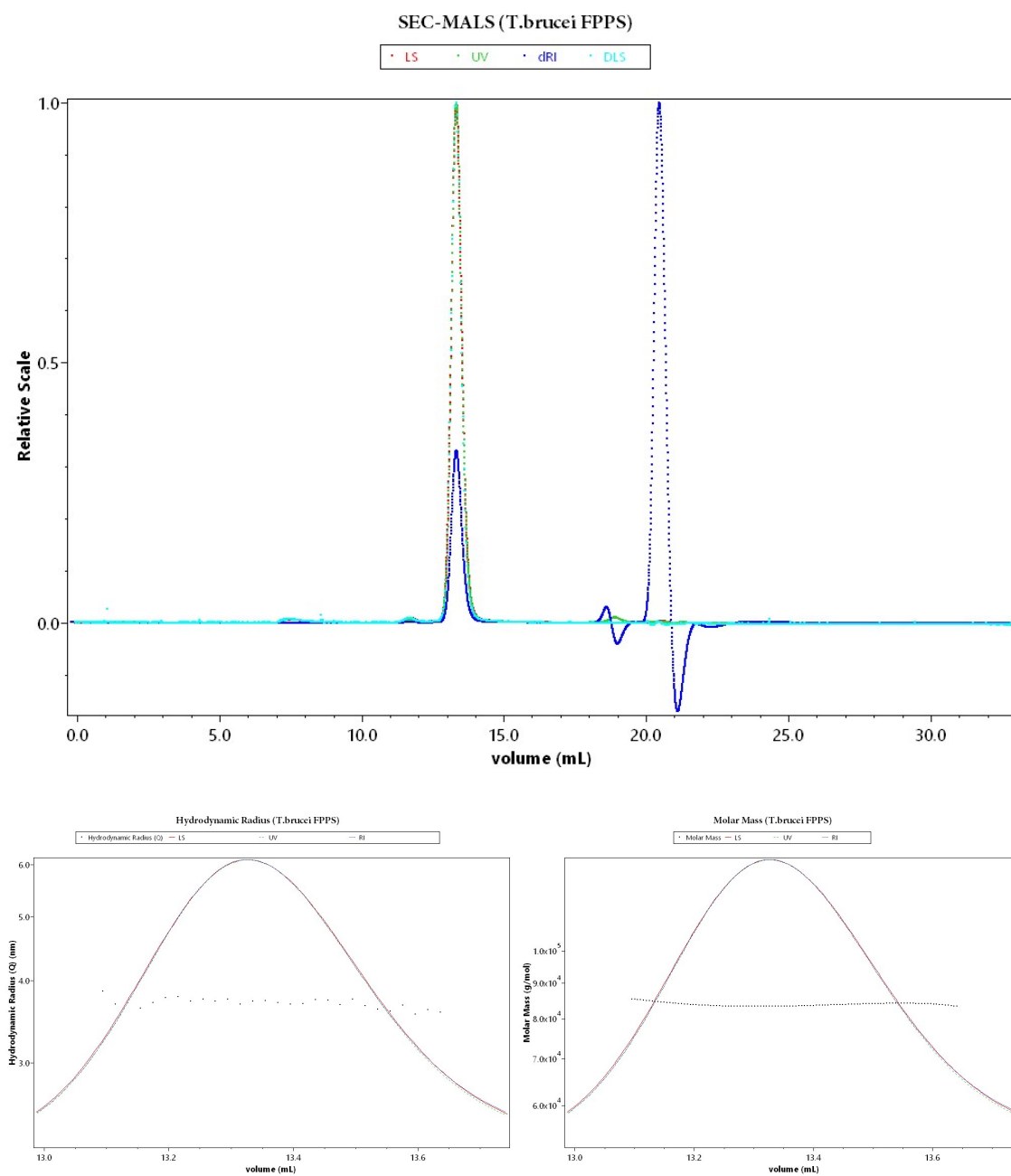
T. cruzi FPPS
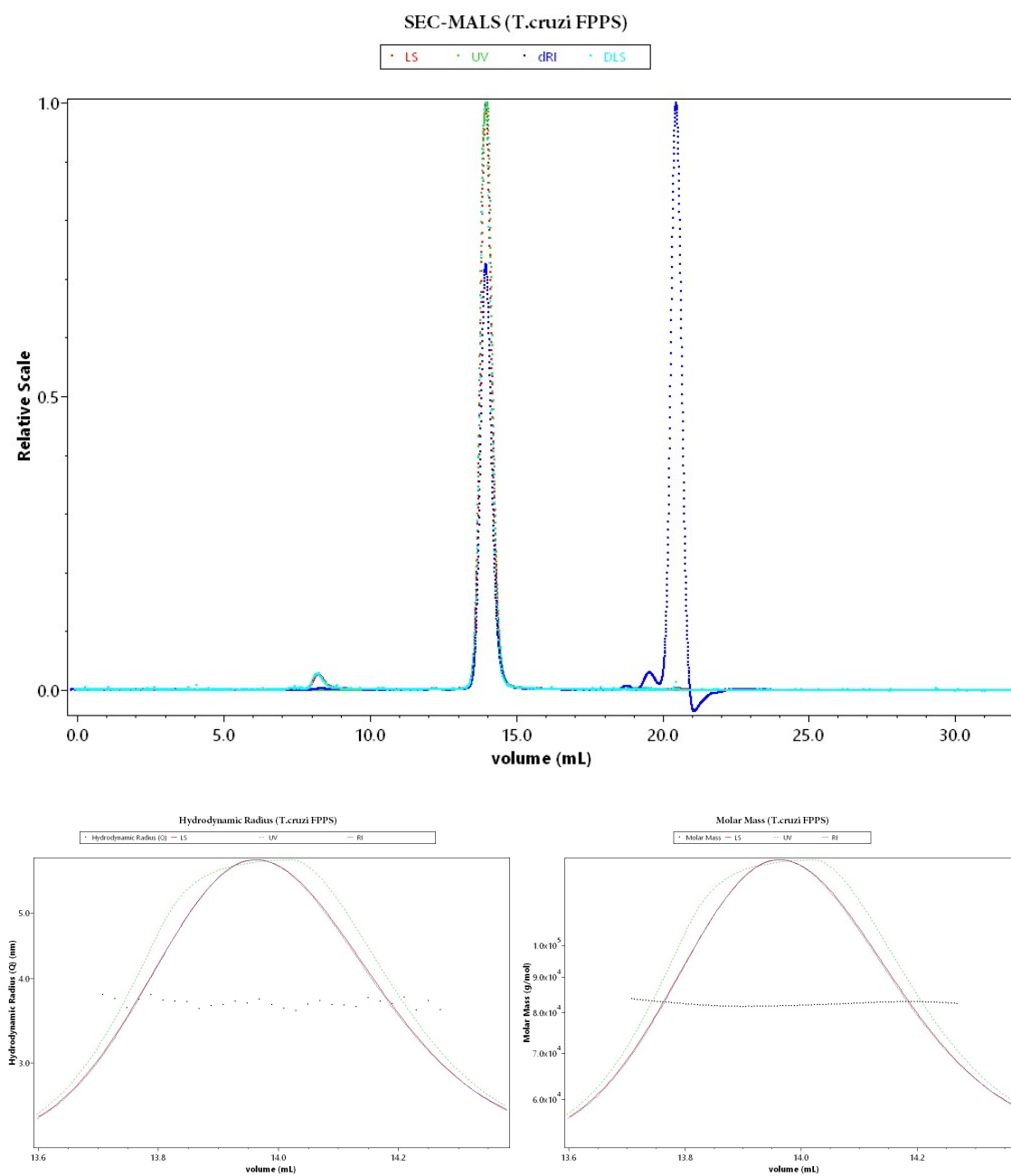
7.2 Crystallization conditions

7.2.1 *Human* FPPS

	Screen	Protein conc.	Crystallization buffer composition
			First Optimization cycle[a]
1	JCSG Core I, H11	16 mg/mL	1 M LiCl, 0.10 M tri-sodium citrate, 10 % (w/v) PEG 6K, pH:4.0
2	JCSG Core IV, H10	16 mg/mL	1.6 M NaH_2PO_4, 0.4 M K_2HPO_4, 0.1 M sodium phosphate citrate, pH: 4.2
3	From literatuere[51]	16 mg/mL	1.2 M NaH_2PO_4 /K_2HPO_4, 25 % (v/v) glycerol, pH 5.3
			Second Optimization cycle[b]
4	Cryos, C3	14 mg/mL	1.6 M $(NH_4)_2HPO_4$, 0.08 Tris, 20 %(v/v) glycerol, pH: 8.5
5	Cryos, D10	14 mg/mL	0.6 M KH_2PO_4, 0.6 M NaH_2PO_4, 25 % (v/v) glycerol, pH: 7.5

[a] protein in the original storage buffer (0.05 M Tris, 0.2 M NaCl, pH 8.0). [b] protein in TSA buffer.

7.2.2 *T. Cruzi* FPPS
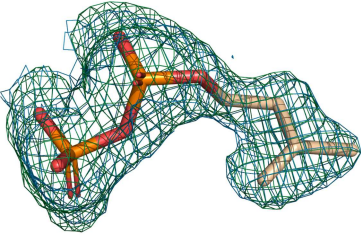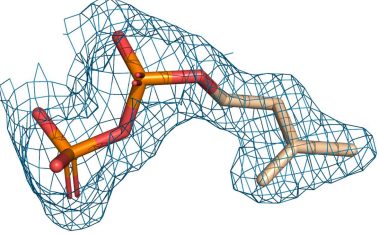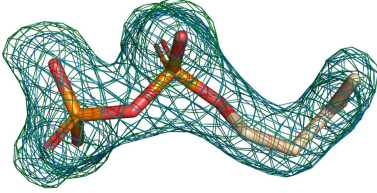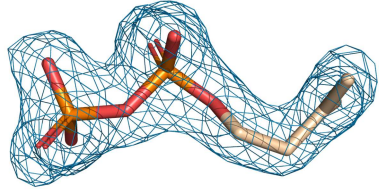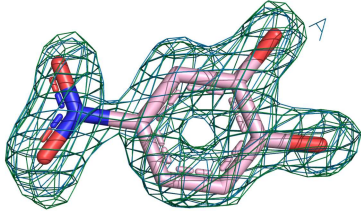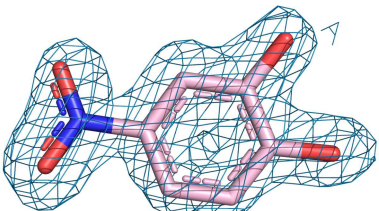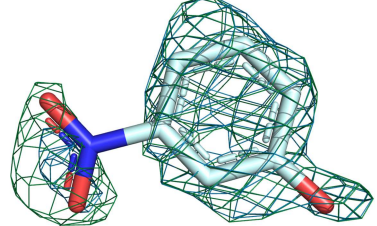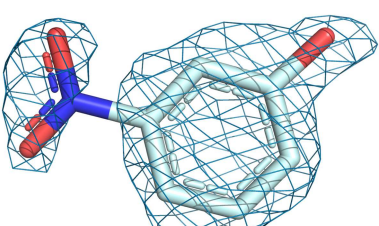
	Screen	Protein conc.	Crystallization buffer composition
			First Optimization cycle[a]
1	JCSG Core I, B6	16 mg/mL	0.8 M NaH_2PO_4, 0.1 M Hepes, 0.8 M KH_2PO_4, pH:7.5
2	JCSG Core I, G5	16 mg/mL	0.10 M tri-sodium citrate, 10 %(w/v) PEG 6K, pH: 5.0
3	JCSG Core II, F12	10 mg/mL	0.1 M sodium phosphate citrate, 5 % (w/v) PEG 3K, 25 % (v/v) 1.2 propanediol, 10 % (v/v) glycerol, pH: 4.2
4	JCSG Core II, G8	10 mg/mL	0.16 M $(NH_4)_2SO_4$, 0.08 M NaOAc, 20 % (w/v) PEG 4K, 20 % (v/v) glycerol, pH:4.6
5	JCSG Core III, G4	10 mg/mL	0.1 M sodium phosphate citrate, 40 % (w/v) PEG 600, pH:4.2
6	JCSG Core IV, H10	16 mg/mL	1.6 M NaH_2PO_4, 0.4 M K_2HPO_4, 0.1 M sodium phosphate citrate, pH: 4.2
7	JCSG+, C5	16 mg/mL	0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 0.1 M Hepes, pH: 4.9
8	Classic, E9	16 mg/mL	1.0 M Li_2SO_4 ,0.10 M tri-sodium citrate, 0.5 M $(NH_4)_2SO_4$, pH:5.6
9	Seeding condition[c]	6.8 mg/mL	0.16 M $(NH_4)_2SO_4$, 0.08 M NaOAc, 20 % (w/v) PEG 4K, 20 % (v/v) glycerol, pH:5.0 (seed stock condition)
		12 mg/mL	0.009 M ZnSO4, 0.08 M MES, 19.42 % PEG MME 550, 15% Glycerol, pH 6.5
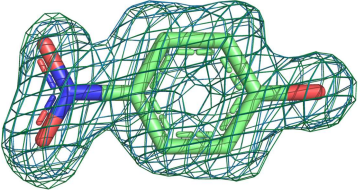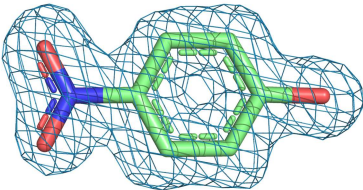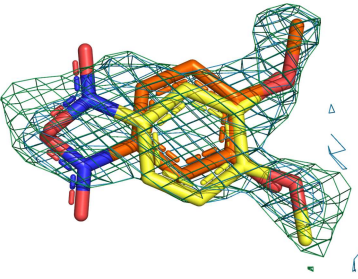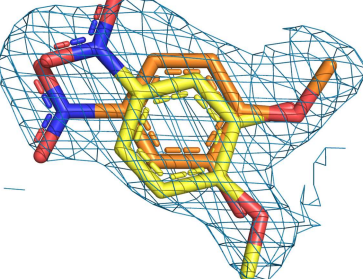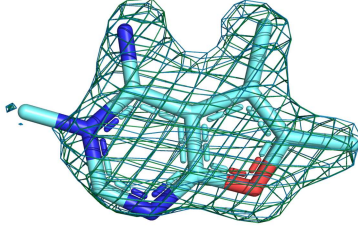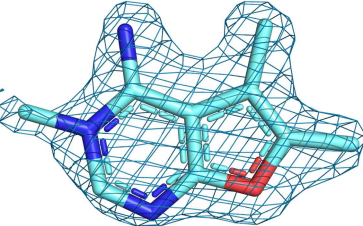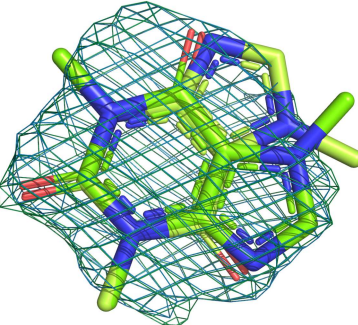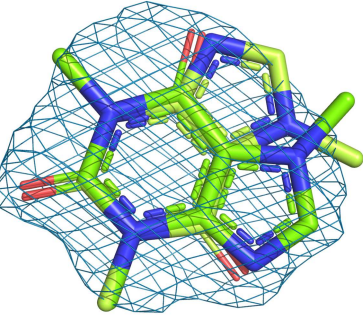
			Second Optimization cycle[b]
10	JCSG+, C5	12 mg/mL	0.8 M KH_2PO_4, 0.8 M NaH_2PO_4, 0.1 M Hepes, pH:7.5
11	Anions, E9	12 mg/mL	0.9 M KH_2PO_4, 0.9 M NaH_2PO_4, 0.1 M Mes, pH:6.5
12	Anions, F9	12 mg/mL	0.9 M KH_2PO_4, 0.9 M NaH_2PO_4, 0.1 M Tris, pH:8.5

[a] protein in the original storage buffer (0.05 M Tris, 0.2 M NaCl, pH 8.0). [b] protein in TSA buffer. [c] condition provided from Joy Petrick (ESR 12).
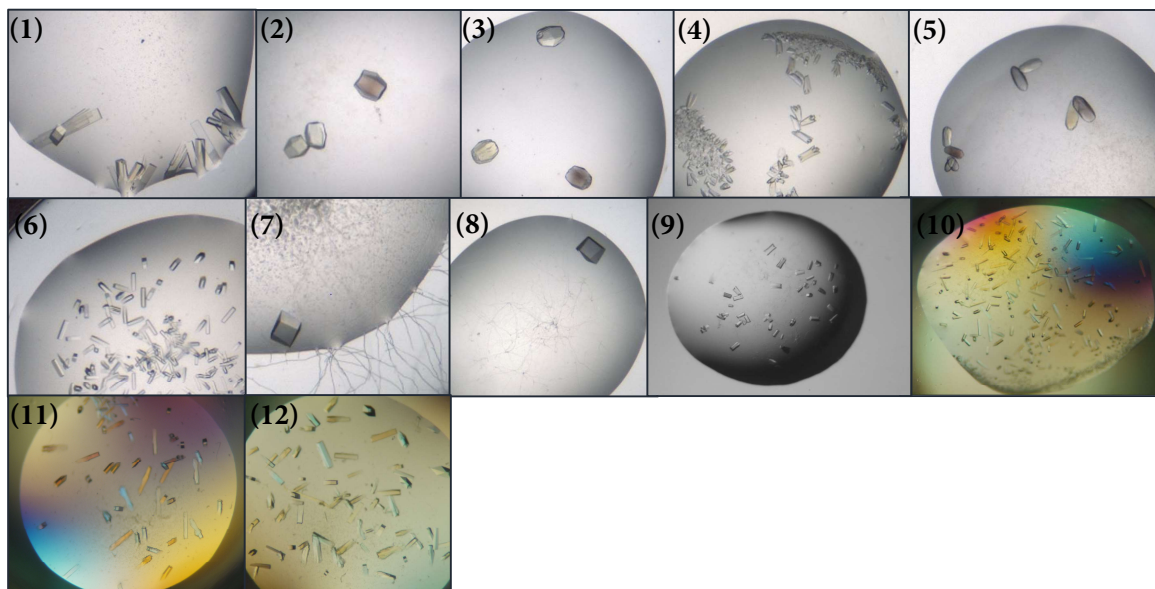
7.3   Electron density maps of the refined hits

	Omit maps	Refined maps
IPP (allyl) Resolution= 1.90 Å Occupancy= 1.00 PDB ID= 6T7N		
IPP (homoallyl) Resolution= 1.90 Å Occupancy= 1.00 PDB ID= 6T7N		
J82 Resolution= 1.51 Å Occupancy= 1.00 PDB ID= 6SDN		
3-nitrophenol Resolution= 1.68 Å Occupancy= 0.77 PDB ID= 6SDO		

4-nitrophenol Resolution= 1.5 Å Occupancy= 1.00 PDB ID= 6SDP		
2m5n Resolution= 1.68 Å Occupancy= 0.64/0.36 PDB ID= 6SDQ		
J51 Resolution= 1.62 Å Occupancy= 1.00 PDB ID= 6SF8		
J71 Resolution= 1.59 Å Occupancy= 0.60/0.40 PDB ID= 6SF9		

7.4 Data collection and refinement statistics[a]

PDB code	6SFA	6T7N	6SE2
	holo-FPPS	**IPP**-FPPS	**apo**-FPPS
(A) Data collection and processing			
X-ray source	BESSY II, 14.1	BESSY II, 14.1	PETRA III, P11
Wavelength (Å)	1.89317	1.89317	0.97620
space group	$P6_122$	$P6_122$	$P6_122$
unit cell parameters a=b, c (Å)	57.75, 397.56	58.45, 391.79	57.94, 397.48
$\alpha=\beta$, γ (°)	90, 120	90, 120	90, 120
Matthews coefficient [b] ($Å^3$/Da)	2.4	2.3	2.4
solvent content [b] (%)	47.9	47.5	47.7
(B) Diffraction data			
resolution range (Å)	49.70-1.90	49.00-1.90	50.00-1.45
unique reflections	57642 (8564)	58433 (8979)	71946 (11290)
Redundancy	16.8 (7.0)	16.3 (5.1)	20.5 (20.1)
$R(I)_{sym}$ (%)	7.4 (50.0)	4.8 (43.0)	5.3 (47.5)
Wilson B factor ($Å^2$)	30.2	30.6	18.9
Completeness (%)	98.4 (90.3)	98.9 (93.5)	100.0 (99.9)
CC (1/2) (%) [e]	99.6 (87.5)	100.0 (87.3)	99.9 (96.1)
$\langle I/\sigma(I)\rangle$	24.24 (2.57)	35.46 (2.69)	32.83 (6.42)
(C) Refinement			
resolution range (Å)	49.62-1.90	49.01-1.90	59.78-1.45
reflections used in refinement	57637	58425	71943
final R value for all reflections (work/free) (%)	16.7/20.2	17.4/20.8	17.0/18.7
protein residues	358	362	360
water molecules	174	168	279
ligand atoms	-	14 | 14	-
other ligands atoms (Mn | Zn | SO4)	2 |-| 5	3 |-|-|	- | 2 | 5
RMSD from ideality: bond lengths (Å)	0.010	0.010	0.008
RMSD from ideality: bond angles (°)	0.970	0.944	0.913
Ramachandran plot [c]			
residues in most favored regions (%)	93.6	94.0	93.6
residues in additionally allowed regions (%)	6.1	5.7	6.1
residues in generously allowed regions (%)	0.0	0.0	0.0
residues in disallowed regions (%)	0.3	0.3	0.3
Mean B factor protein ($Å^2$) [d]	33.7	32.8	24.7
Mean B factor ligand ($Å^2$) [d]	-	30.3 | 26.5	-
Mean B factor water molecules ($Å^2$) [d]	39.5	35.9	31.9
Mean B factor other ligands (Mn | Zn | SO4)	42.1 |-| 42.7	25.6 |-|-|	- | 23.1 | 22.9
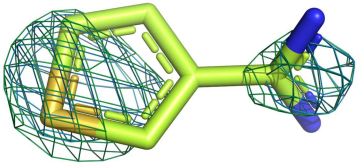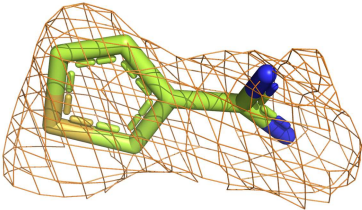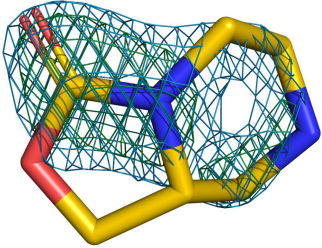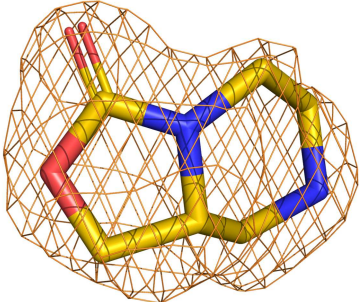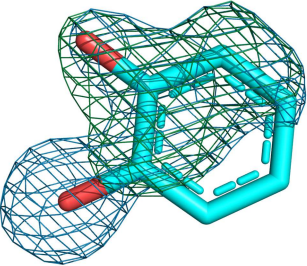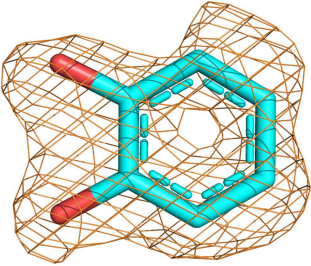
[a]Values in parenthesis describe the highest resolution shell. [b] Calculated with MATTPROB [80]. [c] Calculated with PROCHECK [81]. [d] Mean B factors were calculated with MOLEMAN. [e] $R(I)sym = (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i)))$.
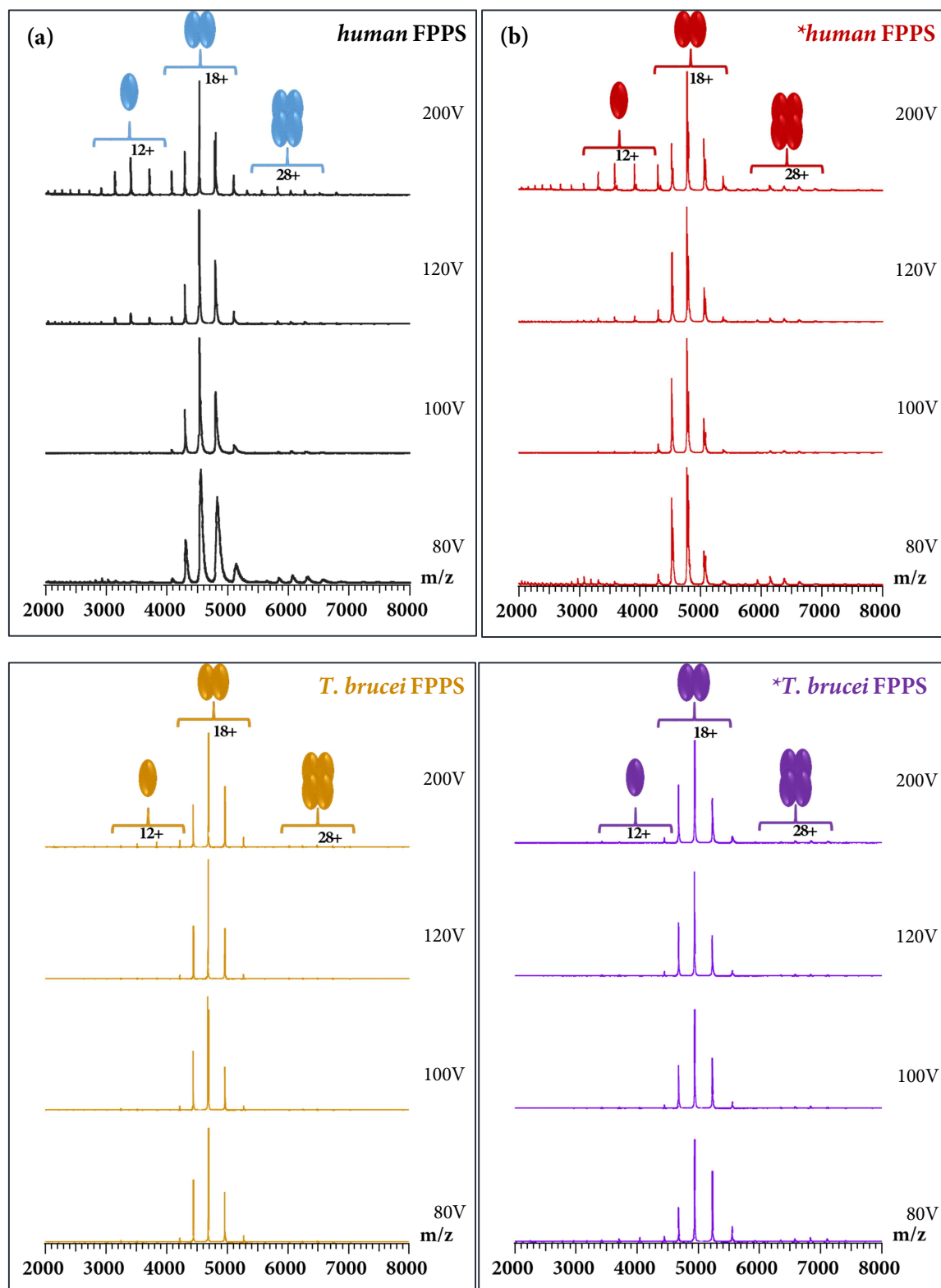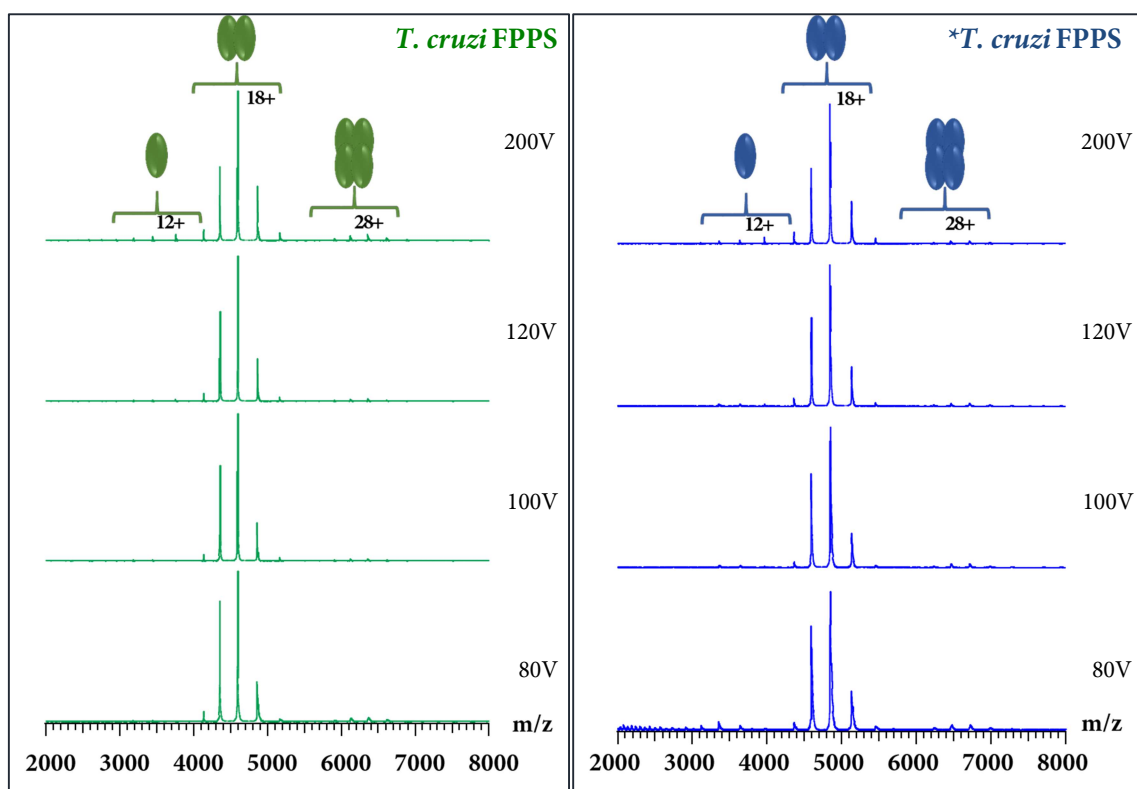
PDB code	6SF8	6SF9	6SDN
	J51-FPPS	**J71**-FPPS	**J82**-FPPS
(A) Data collection and processing			
X-ray source	PETRA III, P13	PETRA III, P13	BESSY II, 14.1
Wavelength (Å)	0.97625	0.97625	0.91841
space group	$P6_122$	$P6_122$	$P6_122$
unit cell parameters a=b, c (Å)	57.90, 398.29	57.99, 398.18	58.03, 397.70
α=β, γ (°)	90, 120	90, 120	90, 120
Matthews coefficient [b] (Å^3 /Da)	2.4	2.4	2.4
solvent content [b] (%)	47.6	47.8	47.8
(B) Diffraction data			
resolution range (Å)	48.62-1.62	48.70-1.59	48.72-1.51
unique reflections	52291 (8151)	54312 (8015)	64357 (10121)
Redundancy	17.4 (7.6)	15.2 (4.9)	19.0 (11.0)
R(I)$_{sym}$ (%)	5.0 (48.3)	8.5 (49.3)	10.7 (49.4)
Wilson B factor (Å^2)	22.2	22.6	15.0
Completeness (%)	99.9 (99.3)	98.8 (92.7)	99.9 (99.5)
CC (1/2) (%) [e]	100.0 (91.6)	99.8 (69.6)	99.6 (89.9)
<I/σ(I)>	35.16 (3.47)	19.20 (2.04)	17.55 (3.13)
(C) Refinement			
resolution range (Å)	48.63-1.62	48.69-1.59	42.48-1.51
reflections used in refinement	52289	54306	64355
final R value for all reflections (work/free) (%)	17.3/19.2	18.3/20.8	16.2/19.0
protein residues	361	360	360
water molecules	283	193	313
ligand atoms	13	28	11
other ligands atoms (Zn | SO4)	-| 5	- | 5	2 | 5
RMSD from ideality: bond lengths (Å)	0.008	0.009	0.008
RMSD from ideality: bond angles (°)	0.925	0.923	0.885
Ramachandran plot [c]			
residues in most favored regions (%)	93.0	93.3	93.6
residues in additionally allowed regions (%)	6.7	6.4	6.1
residues in generously allowed regions (%)	0.0	0.0	0.0
residues in disallowed regions (%)	0.3	0.3	0.3
Mean B factor protein (Å^2) [d]	24.1	27.8	18.7
Mean B factor ligand (Å^2) [d]	34.4	48.0	33.6
Mean B factor water molecules (Å^2) [d]	33.1	35.0	28.2
Mean B factor other ligands (Zn | SO4)	- | 27.8	- | 29.9	17.7 | 16.5

[a]Values in parenthesis describe the highest resolution shell. [b] Calculated with MATTPROB [80]. [c] Calculated with PROCHECK [81]. [d] Mean B factors were calculated with MOLEMAN. [e] R(I)sym = (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i))).

PDB code	6SDO	6SDP	6SDQ
	3-nitrophenol -FPPS	4-nitrophenol -FPPS	2-methoxy-5-nitrophenol -FPPS
(A) Data collection and processing			
X-ray source	PETRA III, P13	PETRA III, P13	PETRA III, P13
Wavelength (Å)	0.97625	0.97625	0.97625
space group	$P6_122$	$P6_122$	$P6_122$
unit cell parameters a=b, c (Å)	57.82, 398.25	57.86, 397.73	57.92, 398.08
α=β, γ (°)	90, 120	90, 120	90, 120
Matthews coefficient [b] (Å^3/Da)	2.4	2.3	2.4
solvent content [b] (%)	47.6	47.3	47.6
(B) Diffraction data			
resolution range (Å)	50.00-1.68	50.00-1.45	50.00-1.68
unique reflections	46230 (6734)	71867 (11311)	46477 (6916)
Redundancy	19.7 (14.7)	18.9 (17.6)	17.1 (11.8)
$R(I)_{sym}$ (%)	4.8 (27.0)	6.4 (48.5)	6.4 (23.4)
Wilson B factor (Å^2)	25.1	18.2	22.7
Completeness (%)	98.7 (91.6)	100.0 (100.0)	99.0 (93.8)
CC (1/2) (%) [e]	99.8 (98.5)	99.8 (94.6)	99.7 (98.0)
$<I/\sigma(I)>$	37.80 (6.43)	26.50 (4.54)	26.85 (6.10)
(C) Refinement			
resolution range (Å)	50.08-1.68	49.72-1.45	56.92-1.68
reflections used in refinement	46222	71864	46472
final R value for all reflections (work/free) (%)	17.0/19.0	16.7/18.8	16.6/18.3
protein residues	360	362	361
water molecules	181	261	224
ligand atoms	10	10	24
other ligands atoms (Zn | SO4)	2	5	2	5	2	5
RMSD from ideality: bond lengths (Å)	0.009	0.008	0.009
RMSD from ideality: bond angles (°)	0.950	0.920	0.938
Ramachandran plot [c]			
residues in most favored regions (%)	93.6	94.0	93.3
residues in additionally allowed regions (%)	6.1	5.7	6.4
residues in generously allowed regions (%)	0.0	0.0	0.0
residues in disallowed regions (%)	0.3	0.3	0.3
Mean B factor protein (Å^2) [d]	28.9	21.8	24.9
Mean B factor ligand (Å^2) [d]	37.9	19.8	37.9
Mean B factor water molecules (Å^2) [d]	36.3	31.7	32.6
Mean B factor other ligands (Zn | SO4)	27.3	25.4	24.9	19.0	28.5	22.7

[a]Values in parenthesis describe the highest resolution shell. [b] Calculated with MATTPROB [80]. [c] Calculated with PROCHECK [81]. [d] Mean B factors were calculated with MOLEMAN. [e] R(I)sym = (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i))).

7.5   PanDDA hits

	Conventional maps	**PanDDA event maps**
J29 Resolution= 2.10 Å Occupancy= 0.71 RMSD = 0.351 Å σ_i = 0.343 BDC = 0.76		
J59 Resolution= 1.56 Å Occupancy= 0.80 RMSD = 0.307 Å σ_i = 0.257 BDC = 0.68		
Pyrocathecol Resolution= 1.68 Å Occupancy= 0.77 RMSD = 0.181 Å σ_i = 0.231 BDC = 0.88		

7.6 MS Spectra

7.6.1 Native MS

$* * *$: $^{13}C^{15}N$-labled protein

7.6.2 Calculated Exact Mass

Protein Species	Monomer (MW)	Dimer (MW)
human FPPPS	40687 ± 1 Da	81402 ± 8 Da
$^{13}C^{15}N$-*human* FPPS	42944 ± 4 Da	85901 ± 12 Da
T. cruzi FPPS	41315 ± 1 Da	82629 ± 1 Da
$^{13}C^{15}N$- *T. cruzi* FPPS	43597 ± 1 Da	87194 ± 1 Da
T. brucei FPPS	42126 ± 1 Da	84253 ± 2 Da
$^{13}C^{15}N$- *T. brucei* FPPS	44402 ± 2 Da	88812 ± 7 Da
T. brucei FPPS-heterodimer	86534 ± 3 Da	

7.6.3   Kinetic exchange experiments

★ ★ ★ : $^{13}C^{15}N$-labled protein
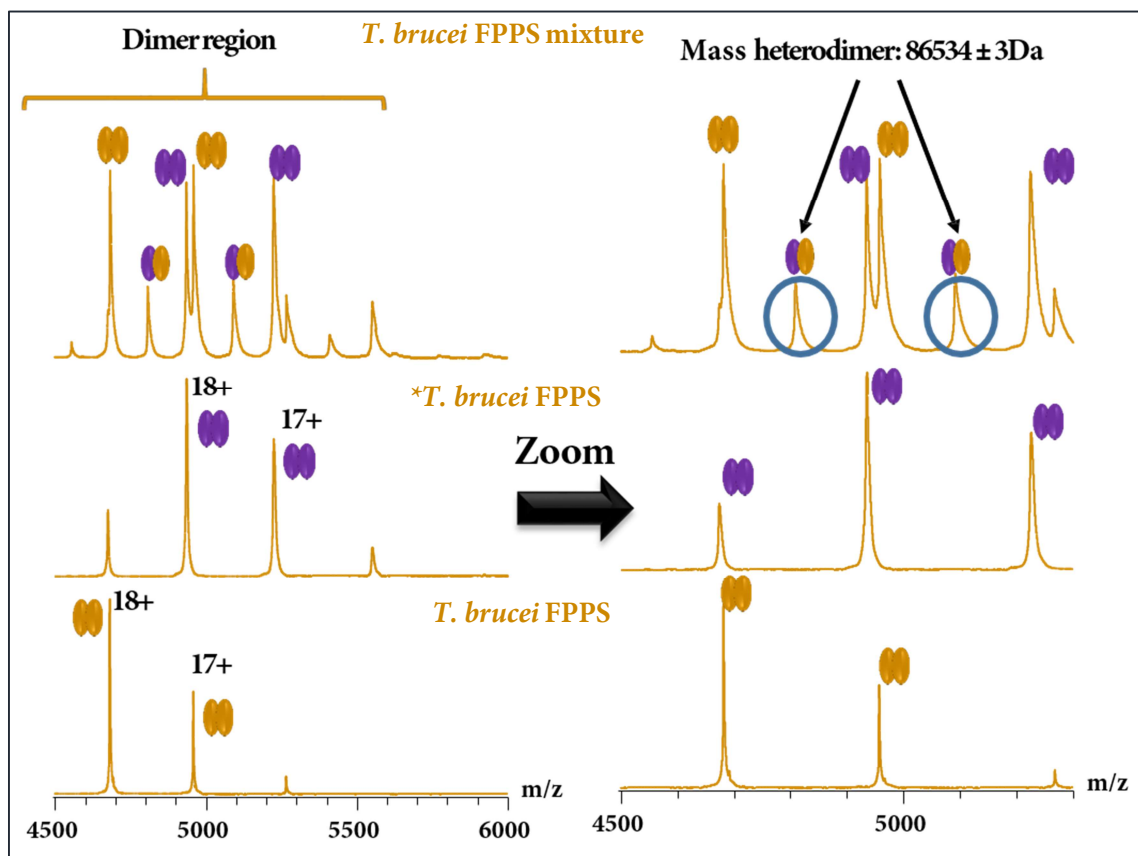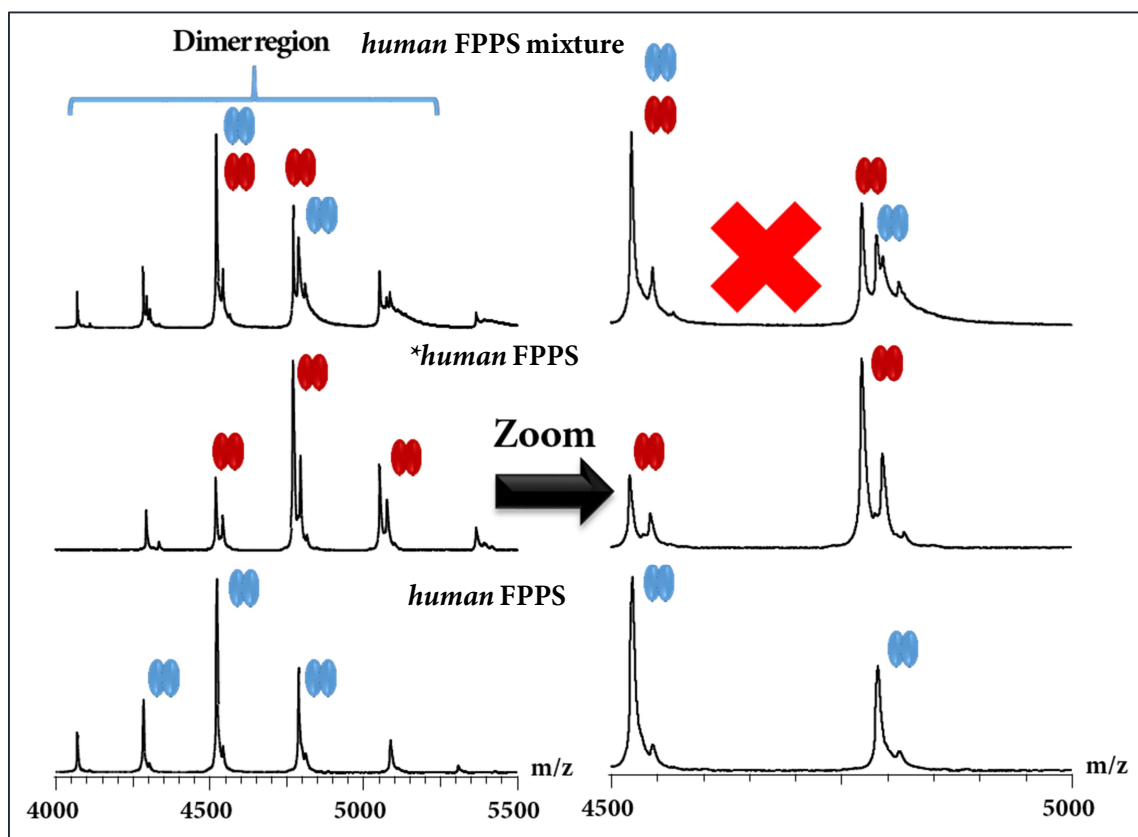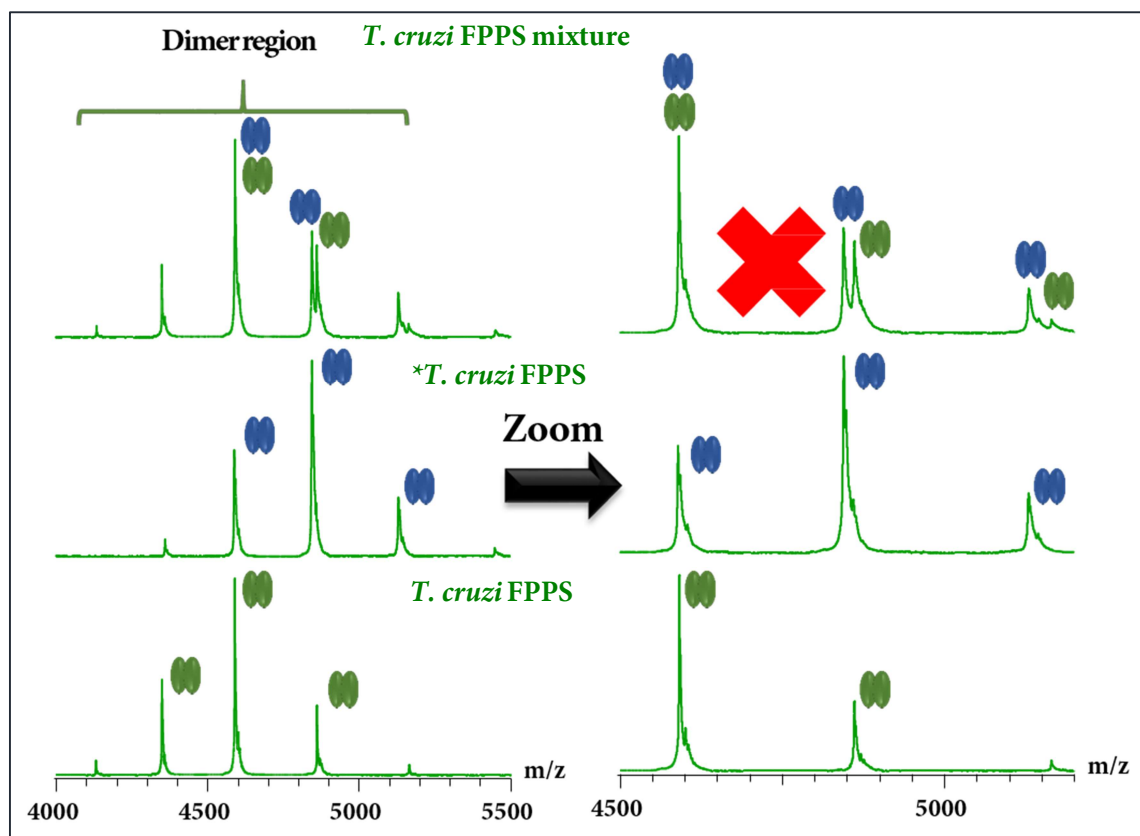
8. References

[1] P. J. Hotez, "The disease next door," *Foreign Policy*, 2013.

[2] P. J. Hotez *et al.*, "Blue Marble Health and the Global Burden of Disease Study 2013," *PLoS Negl. Trop. Dis.*, vol. 10, no. 10, pp. 6–11, 2016.

[3] M. Parker *et al.*, "Neglected Tropical Diseases in Biosocial Perspective." *J. Biosoc. Sci* , vol. 48, no. S1, pp. 1-15, 2016.

[4] WHO, "Accelerating Work To Overcome the Global Impact of Neglected Tropical," *Accel. Work to overcome neglected Trop. Dis. a roadmap implementation.*, 2012.

[5] M. P. Barrett *et al.*, "The trypanosomiases," *Lancet*, vol. 362, no. 9394, pp. 1469–1480, 2003.

[6] C. Bern *et al.*, "An Estimate of the Burden of Chagas Disease in the United States," *Clin. Infect. Dis.*, vol. 49, no. 5, pp. e52–e54, 2009.

[7] L. Basile *et al.*, "Chagas disease in european countries: The challenge of a surveillance system," *Eurosurveillance*, vol. 16, no. 37, p. 3, 2011.

[8] K. C. F. Lidani *et al.*, "Chagas Disease: From Discovery to a Worldwide Health Problem," *Front. Public Heal.*, vol. 7, no. July, pp. 1–13, 2019.

[9] L. Bolívar Zaruma, *Proteins and Proteomics of Leishmania and Trypanosoma*, vol. 74. Dordrecht: Springer Netherlands, 2014.

[10] M. R. Alexander, "What is the global prevalence of hypertension (high blood pressure)?," *Medscape*, 2019.

[11] W. De Souza *et al.*, "Review on Trypanosoma cruzi: Host cell interaction," *Int. J. Cell Biol.*, vol. 2010, 2010.

[12]   K. C. F. Lidani *et al.*, "The complement system: A prey of Trypanosoma cruzi," *Front. Microbiol.*, vol. 8, no. APR, pp. 1–14, 2017.

[13]   B. Stijlemans *et al.*, "Nanobodies As tools to Understand, diagnose, and treat African trypanosomiasis," *Front. Immunol.*, vol. 8, no. JUN, pp. 1–12, 2017.

[14]   Your genome org, "What is African sleeping sickness?,", *Facts*, (https://www.yourgenome.org/facts/what-is-african-sleeping-sickness), 2016.

[15]   F. Barreira *et al.*, "New perspectives on the therapeutic use of Benznidazol," *Dis. Clin. Res. Platf. - Newsl. number 6*, no. May, pp. 8–10, 2016.

[16]   S. Patterson and S. Wyllie, "Nitro drugs for the treatment of trypanosomatid diseases: Past, present, and future prospects," *Trends Parasitol.*, vol. 30, no. 6, pp. 289–298, 2014.

[17]   S. Pund and A. Joshi, *Nanoarchitectures for Neglected Tropical Protozoal Diseases: Challenges and State of the Art*. Elsevier Inc., 2017.

[18]   A. Catarina and C. Silva, "Chagas Disease Treatment and Rational Drug Discovery : A Challenge That Remains," vol. 10, no. August, pp. 1–6, 2019.

[19]   I. Molina *et al.*, "The use of posaconazole against Chagas disease," *Curr. Opin. Infect. Dis.*, vol. 28, no. 5, pp. 397–407, 2015.

[20]   F. Torrico *et al.*, "Articles Treatment of adult chronic indeterminate Chagas disease with benznidazole and three E1224 dosing regimens : a proof-of-concept , randomised , placebo-controlled trial," vol. 3099, no. 17, 2018.

[21]   E. Dumonteil *et al.*, "A therapeutic preconceptional vaccine against Chagas disease: A novel indication that could reduce congenital transmission and accelerate vaccine development," *PLoS Negl. Trop. Dis.*, vol. 13, no. 1, pp. 31–35, 2019.

[22]   G. Priotto *et al.*, "Nifurtimox-efl ornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis : a multicentre , randomised , phase III , non-inferiority trial," *Lancet*, vol. 374, no. 9683, pp. 56–64, 2009.

[23]   E. D. Deeks, "Fexinidazole : First Global Approval," *Drugs*, no. 0123456789, pp. 1–6, 2019.

[24]   R. T. Jacobs *et al.*, "SCYX-7158 , an Orally-Active Benzoxaborole for the Treatment of Stage 2 Human African Trypanosomiasis," vol. 5, no. 6, 2011.

[25]   J. Park *et al.*, "Human isoprenoid synthase enzymes as therapeutic targets," *Front. Chem.*, vol. 2, no. JUL, pp. 1–21, 2014.

[26]   R. Docampo and S. N. J. Moreno, *Biochemistry of Trypanosoma cruzi*, Second Edi. Elsevier Inc., 2017.

[27]   R. O. Cosentino and F. Agüero, "Genetic profiling of the isoprenoid and sterol biosynthesis pathway genes of Trypanosoma cruzi," *PLoS One*, vol. 9, no. 5, 2014.

[28]   T. K. Smith *et al.*, "Metabolic reprogramming during the Trypanosoma brucei life cycle [version 2 ; referees : 4 approved] Referee Status :," *F1000Research*, vol. 6, no. May, pp. 1–12, 2017.

[29]   I. Coppens and P. J. Courtoy, "Exogenous and endogenous sources of sterols in the culture-adapted procyclic trypomastigotes of Trypanosoma brucei," *Mol. Biochem. Parasitol.*, vol. 73, no. 1–2, pp. 179–188, 1995.

[30]   R. K. Harijan *et al.*, "The SCP2-thiolase-like protein (SLP) of Trypanosoma brucei is an enzyme involved in lipid metabolism," *Proteins Struct. Funct. Bioinforma.*, vol. 84, no. 8, pp. 1075–1096, 2016.

[31]   W. de Souza and J. C. F. Rodrigues, "Sterol Biosynthesis Pathway as Target for Anti-trypanosomatid Drugs," *Interdiscip. Perspect. Infect. Dis.*, vol. 2009, pp. 1–19, 2009.

[32]   A. Srivastava *et al.*, "Structural Analysis of Farnesyl Pyrophosphate Synthase from Parasitic Protozoa, a Potential Chemotherapeutic Target," *Infect. Disord. - Drug Targets*, vol. 8, no. 1, pp. 16–30, 2012.

[33]   S. B. Gabelli *et al.*, "Structure and mechanism of the farnesyl diphosphate synthase from Trypanosoma cruzi: implications for drug design.," *Proteins*, vol. 62, no. 1, pp. 80–88, 2006.

[34]   M. K. Dhar *et al.*, "Farnesyl pyrophosphate synthase: A key enzyme in isoprenoid biosynthetic pathway and potential molecular target for drug development," *N. Biotechnol.*, vol. 30, no. 2, pp. 114–123, 2013.

[35]   K. L. Kavanagh *et al.*, "The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs," vol. 103, no. 20, 2006.

[36]   D. J. Hosfield *et al.*, "Structural Basis for Bisphosphonate-mediated Inhibition of Isoprenoid Biosynthesis," *J. Biol. Chem.*, vol. 279, no. 10, pp. 8526–8529, 2004.

[37]   J. Rondeau *et al.*, "Structural Basis for the Exceptional in vivo Efficacy of Bisphosphonate Drugs," pp. 267–273, 2006.

[38]   D. W. Christianson *et al.*, "Structural and Chemical Biology of Terpenoid Cyclases," *Chem. Rev.*, vol. 117, no. 17, pp. 11570–11648, 2017.

[39]   Y. Zhang *et al.*, "Lipophilic bisphosphonates as dual farnesyl/geranylgeranyl diphosphate synthase inhibitors: An X-ray and nmr investigation," *J. Am. Chem. Soc.*, vol. 131, no. 14, pp. 5153–5162, 2009.

[40]   G. Iacobellis *et al.*, "Treating the Bone to Protect the Heart: Potential Newer Mechanisms and Targets," *Am. J. Med. Sci.*, vol. 357, no. 6, pp. 451–452, 2019.

[41]   S. Cremers *et al.*, "Pharmacology of bisphosphonates," *Br. J. Clin. Pharmacol.*, vol. 85, no. 6, pp. 1052–1062, 2019.

[42]	R. Bartl and C. Bartl, "The Osteoporosis Manual," *Osteoporos. Man.*, 2019.

[43]	G. Hampson and I. Fogelman, "Clinical role of bisphosphonate therapy," *Int. J. Womens. Health*, vol. 4, no. 1, pp. 455–469, 2012.

[44]	J. C. Frith *et al.*, "The Molecular Mechanism of Action of the Antiresorptive and Antiinflammatory Drug Clodronate: Evidence for the Formation in Vivo of a Metabolite That Inhibits Bone Resorption and Causes Osteoclast and Macrophage Apoptosis," *Arthritis Rheum.*, vol. 44, no. 9, pp. 2201–2210, 2001.

[45]	C. Huang *et al.*, "Binding of nitrogen-containing bisphosphonates (N-BPs) to the Trypanosoma cruzi farnesyl diphosphate synthase homodimer," no. September, pp. 888–899, 2009.

[46]	J. Mao *et al.*, "Solid-state NMR, crystallographic, and computational investigation of bisphosphonates and farnesyl diphosphate synthase-bisphosphonate complexes," *J. Am. Chem. Soc.*, vol. 128, no. 45, pp. 14485–14497, 2006.

[47]	S. Aripirala *et al.*, "Design, synthesis, calorimetry, and crystallographic analysis of 2-alkylaminoethyl-1,1-bisphosphonates as inhibitors of trypanosoma cruzi farnesyl diphosphate synthase," *J. Med. Chem.*, vol. 55, no. 14, pp. 6445–6454, 2012.

[48]	R. Cao *et al.*, "Structures of a potent phenylalkyl bisphosphonate inhibitor bound to farnesyl and geranylgeranyl diphosphate synthases," pp. 431–439, 2008.

[49]	Y. L. Liu *et al.*, "Farnesyl diphosphate synthase inhibitors with unique ligand-binding geometries," *ACS Med. Chem. Lett.*, vol. 6, no. 3, pp. 349–354, 2015.

[50]	G. Yang *et al.*, "In Vitro and in Vivo investigation of the inhibition of Trypanosoma brucei cell growth by lipophilic bisphosphonates," *Antimicrob. Agents Chemother.*, vol. 59, no. 12, pp. 7530–7539, 2015.

[51]   W. Jahnke *et al.*, "Allosteric non-bisphosphonate FPPS inhibitors identified by fragment-based discovery.," *Nat. Chem. Biol.*, vol. 6, no. 9, pp. 660–666, 2010.

[52]   J. Park *et al.*, "Human farnesyl pyrophosphate synthase is allosterically inhibited by its own product," *Nat. Commun.*, vol. 8, p. 14132, 2017.

[53]   B. Stijlemans *et al.*, "Nanobodies As Tools to Understand, Diagnose, and Treat African Trypanosomiasis," *Front. Immunol.*, vol. 8, no. JUN, pp. 1–12, Jun. 2017.

[54]   O. H. Hashim and N. A. Adnan, "Coenzyme, cofactor and prosthetic group - Ambiguous biochemical jargon," *Biochem. Educ.*, vol. 22, no. 2, pp. 93–94, 1994.

[55]   A. Augustyn *et al.* "Enzymes," *Encyclopædia Britannica, inc.*, (https://www.britannica.com/science/enzyme), 2019.

[56]   I. Dokmanić, M. *et al.* , "Metals in proteins: Correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 64, no. 3, pp. 257–263, 2008.

[57]   D. Piovesan *et al.*, "The human ' magnesome ' : detecting magnesium binding sites on human proteins," *BMC Bioinformatics*, vol. 13, no. Suppl 14, p. S10, 2012.

[58]   K. Pasternak *et al.*, "Review Paper Biochemistry of Magnesium," *J. Elem.*, vol. 15, no. 3, pp. 601–616, 2010.

[59]   "How Magnesium Works & Why It Is Important." *Center for Magnesium Education & Research* (https://www.magnesiumeducation.com/how-magnesium-works-why-it-is-important), 2019

[60]   L. Reinhard *et al.*, "Optimization of protein buffer cocktails using Thermofluor," *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 69, no. 2, pp. 209–214, 2013.

[61]   M. P. Tarazona and E. Saiz, "Combination of SEC/MALS experimental procedures and theoretical analysis for studying the solution properties of macromolecules," *J. Biochem. Biophys. Methods*, vol. 56, no. 1–3, pp. 95–116, 2003.

[62]   I. Persson *et al.*, "Hydrated metal ions in aqueous solution: How regular are their structures?," *Pure Appl. Chem.*, vol. 82, no. 10, pp. 1901–1917, 2010.

[63]   J. Schiebel *et al.*, "High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits.," *Structure*, vol. 24, no. 8, pp. 1398–409, Aug. 2016.

[64]   A. Bauzá *et al.*, "π-Hole Interactions Involving Nitro Aromatic Ligands in Protein Structures," *Chem. – A Eur. J.*, pp. 13436–13443, 2019.

[65]   H. Szatylowicz *et al.*, "Classical and reverse substituent effects in meta- and para-substituted nitrobenzene derivatives," *Struct. Chem.*, vol. 28, no. 4, pp. 1125–1132, 2017.

[66]   V. D'Atri *et al.*, "Adding a new separation dimension to MS and LC–MS: What is the utility of ion mobility spectrometry?," *J. Sep. Sci.*, vol. 41, no. 1, pp. 20–67, 2018.

[67]   S. M. Dixit *et al.*, "Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future," *Curr Opin Chem Biol.* ; vol. 42, pp. 93–100, 2018.

[68]   U. Mueller *et al.*, "The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives," *Eur. Phys. J. Plus*, vol. 130, no. 7, 2015.

[69]   W. Kabsch *et al.*, "XDS," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 125–132, 2010.

[70]   K. M. Sparta *et al.*, "XDSAPP2.0," *J. Appl. Crystallogr.*, vol. 49, no. 3, pp. 1085–1092, 2016.

[71]   A. J. McCoy *et al.*, "Phaser crystallographic software ," *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007.

[72]   P. D. Adams *et al.*, "PHENIX: A comprehensive Python-based system for macromolecular structure solution," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 213–221, 2010.

[73]   P. Emsley and K. Cowtan., "Coot: Model-building tools for molecular graphics," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12 I, pp. 2126–2132, 2004.

[74]   G. M. Sheldrick *et al.*, "A short history of SHELX," *Acta Crystallogr. Sect. A Found. Crystallogr.*, vol. 64, no. 1, pp. 112–122, 2008.

[75]   G. Langer *et al.*, "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7," *Nat. Protoc.*, vol. 3, no. 7, pp. 1171–1179, 2008.

[76]   G. Neudert and G. Klebe, "DSX: A knowledge-based scoring function for the assessment of protein-ligand complexes," *J. Chem. Inf. Model.*, vol. 51, no. 10, pp. 2731–2745, 2011.

[77]   G. Neudert and G. Klebe, "fconv: Format conversion, manipulation and feature computation of molecular data," *Bioinformatics*, vol. 27, no. 7, pp. 1021–1022, 2011.

[78]   M. McGann *et al.*, "FRED and HYBRID docking performance on standardized datasets," *J. Comput. Aided. Mol. Des.*, vol. 26, no. 8, pp. 897–906, 2012.

[79]   P. C. D. Hawkins *et al.*, "Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database," *J. Chem. Inf. Model.*, vol. 50, no. 4, pp. 572–584, 2010.

[80]   K. A. Kantardjieff *et al.*, "Proceedings of 10th International Conference on Formal Methods in Computer-Aided Design, FMCAD 2010, Lugano, Switzerland, October 20-23," *Fmcad*, pp. 1865–1871, 2010.

[81]   R. A. Laskowski *et al.*, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, 1993.

Part IV: Endothiapepsin (EP)

Rapid Discovery of Novel Aspartyl Protease Inhibitors Using An Anchoring Approach

1. Introduction

1.1. Endothiapepsin as a model protein

Endothiapepsin (EP, EC 3.4.23.22). is a monomeric enzyme of 330 amino acids and a molecular weight of 33.8 kDa derived from the fungus *Endothia parasitica* [1]. EP belongs to the aspartic protease family which were the first type of enzymes known to humanity and, the prototype porcine pepsin, was the second enzyme to be crystallized and the first of which a diffraction pattern was obtained by X-ray crystallography[2].

This group of enzymes are wildly present in humans, plants, fungi and retroviruses and comprise homologous enzymes including pepsin, chymosin, renin, cathepsin D and also EP[1] [2]. Sources of aspartic proteases are mainly found in acidic environments within cells and tissues such as the stomach of mammals, yeasts and fungi where they exhibit their optimal catalytic activities at acidic pH, between 2 and 5 [2], [3], [4].

The aspartic protease group of enzymes consist of two structurally similar domains, each of which contributes with one aspartic acid to the catalytic dyad (Asp35 and Asp219) which form the active site [1]. The binding pocket is located in a large cleft which is formed between the two domains and can accommodate large substrate peptides consisting of seven to eight amino acid residues [2]. The catalytic site is delimited in the upper part by the so-called flap region which is the most flexible part of the enzyme that opens the catalytic pocket to allow the substrate binding and close it upon binding [5].

The crystal structure of EP revealed nine pockets in the catalytic site [5]. In particular, the S1', S2', S1 and S2 pockets are the most relevant ones considering the present study. The S1' pocket is the most hydrophobic one and is surrounded by Ile300, Ile302 and Ile304 while the S2' by Ile77, Leu133 and Phe194. The S1 pocket is amphiphilic and surrounded by Tyr79, Phe116 and Leu125 while the S2 pocket is rather hydrophilic and surrounded by Thr222, Thr223 and Tyr226.

As all member of the family, EP catalyzes the hydrolysis of substrate peptide bond, preferably between hydrophobic residues. The catalytic activity is due to Asp35 and Asp219 which have a pKa value ~ 5 and ~ 2 respectively [4]. Considering the optimal catalytic activity at acidic condition, Asp35 is deprotonated while Asp219 is protonated. The crystal structure of native aspartic protease revealed a catalytic water molecule within hydrogen-bonding distance of all oxygen atoms of the aspartate side chain which form the catalytic dyad [1]. The water molecule is partially displaced upon substrate binding and is extremely polarized by the negatively charged Asp35 side chain which promotes a nucleophilic attack by the water molecule on the scissile bond carbonyl group of the substrate [1](Figure 1.1. (1)). The resulting geminal-diol tetrahedral intermediate is only shortly stabilized *via* hydrogen bonds by Asp219 (Figure 1.1. (2)) as the scissile C-N bond of the substrate is rapidly cleaved to produce the final product, a carboxylic acid and a primary amine [1](Figure 1.1. (3)).
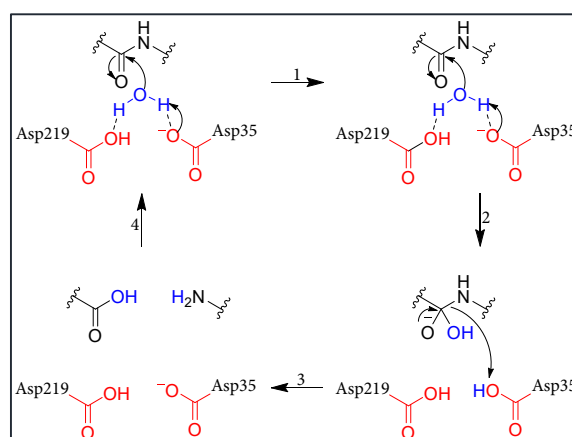
Figure 1.1. Catalytic mechanism of aspartic protease. Figure adapted from [1].

Although EP is not a drug target *per se*, it has received considerable attention as relevant surrogate in drug discovery programs due to its sequence similarity and folding architecture with related drug target. Some examples include renin for the design of antihypertensive drugs, beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) for the studies concerning Alzeimer's disease, cathepsin D for its implication in cancer, HIV1 protease as the causative factor for the development of AIDS, plasmepsin and serum

amyloid P component (SAP) for the design of antimalarial and antimycotics drugs, respectively [2], [3], [5].

Due to its structural similarity to the above-mentioned targets, EP is often used as a model protein. In addition, it is easily extractable from Suparen®, its crystallization conditions are well known, and the crystals diffract X-rays at high resolutions enabling to obtain valid information on the binding mode of a wide range of compounds.

1.2.   Aim of the project

The present work is part of a recently submitted article in collaboration with Markella Konstantinidou (EU ITN AEGIS, ESR13) from the Drug Design group of Prof. A. Dömling (University of Groningen), the Department of Drug Design and Optimization of Prof. A.K. Hirsch (Helmholtz..) and C.J. Camacho from the Department of Computational and Systems Biology (University of Pittsburgh).

The main aim of the project was the introduction and development of a pipeline for the rapid discovery of novel aspartyl protease inhibitors using an anchoring approach. The validity of the hypothesis and method was validated by several soaked crystal structure and the inhibitory potency of the molecules was finally accessed by an enzyme assay.

The contribution to this paper of the author of this thesis concerned the crystallographic part, which will be discussed in detail in Section 2.2.

2. Rapid Discovery of Novel Aspartyl Protease Inhibitors Using An Anchoring Approach

2.1   An overview of the novel pipeline

The discovery and development of a novel drug is highly time, resources and investment-intensive undertaking with very low rate of success if compared with other industrial processes. The bottlenecks of an early stage drug discovery project are often time-consuming and expensive high-throughput screening and hit to lead optimization process which may require a lot of time and resources. it becomes therefore necessary to have a pipeline for the rapid and effective screening of a series of compounds in the shortest possible time.

In Prof. A. Dömling group a specialized pharmacophore search technology which is named AnchorQuery was recently introduced [6], [7]. The technology is based upon a >30 million database of virtual compounds which can be easily synthesized through one-step multi-component reaction (MCR) chemistry and contains an *anchor* motif that is bioisosteric to an amino acid residue. An *anchor* is defined as an amino-acid side chain in the interface of protein-protein interaction (PPI) which is contributing above average to its energetics, for example a side-chain that buries a large amount of surface area at the core of the binding interface [8]. Anchors are usually part of energetic hot spots [9]. The AnchorQuery technology was validated and proven by the discovery of novel and bioactive scaffolds for p53/MDM2 or PDK1 systems [10], [11]. However, the current limitation is its sole use in discovering molecules which mimic amino acids side chain in the contest of PPIs.

The present work was then focused on combining the concept of an *anchor* with one-step MCR chemistry to expand its use to the wide area of fragment-based drug discovery. A new pipeline of tools was therefore proposed and validated for the discovery of novel Asp-protease inhibitors using EP as a model protein.

The starting point was to find a suitable *anchor* which would bind to both aspartic acid residues of the catalytic dyad. It was then considered to use a hydrazine moiety because of its simple structure enabling a broad range of organic chemistry and its ability to interact with two carboxylic acids. Furthermore, it mimics the catalytic water molecule present in the active site of the native aspartic protease (Section 1.1.). In addition, under acidic condition, the hydrazine moiety is protonated, either at the exo- or endo-nitrogen, thus forming ionic interactions with the negatively charged carboxylic acids. Then a scaffold was designed which is either accessible through MCR and has a hydrazine motif as warhead. The scaffold is easily synthesized *via* two-steps reaction and its chemical structure is reported in Figure 2.1.**(a)**. Initially a small library of 17 derivatives was screened and five entries showed inhibitory activity in the enzymatic assay (Figure 2.1. (b)).
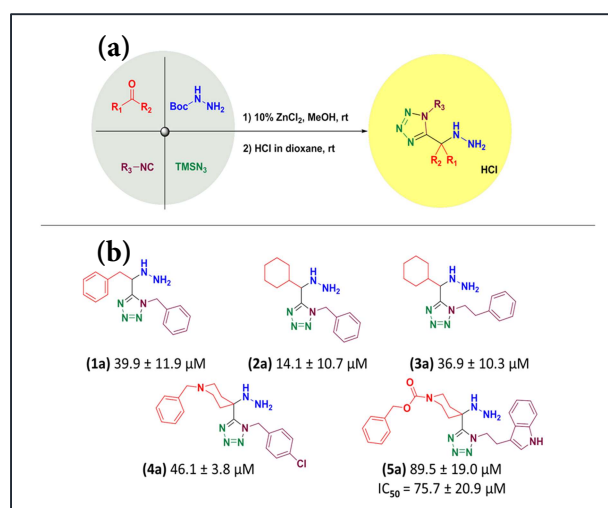
Figure 2.1. (a) Design of a MCR scaffold containing the hydrazine motif. **(b)** Chemical structure, inhibitory concentration at 200 μM and IC_{50} values of hits from the first screening.

Subsequentially, the initial scaffold was further optimized using the hydrazine moiety as anchoring fragment. A protocol for tailor-made virtual library screening was then developed. In the first step, a virtual library was created starting from commercially available starting materials and able to cover a wide range of chemical space to ensure novelty of the designed compounds. Then, for the 3D anchoring of the hydrazine fragments, different protonation states and orientations between the catalytic Asp35 and

Asp219 were considered and helped to filter putative fragments for the designed library. Finally, a pharmacophore was generated, including the hydrazine moiety and the tetrazole ring. In total, 12 optimized hits were selected (Figure 2.2.), synthesized and evaluated with both, a fluorescence-based assay and X-ray crystallography.
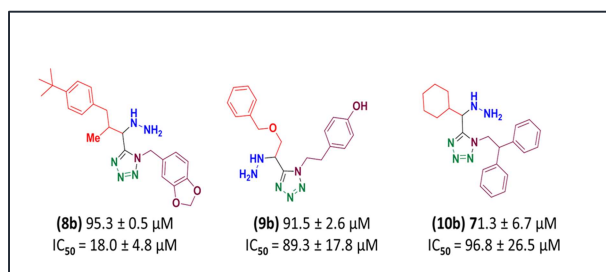
Figure 2.2. Chemical structure, inhibitory concentration at 200 µM and IC$_{50}$ values of the most promising compounds from the second screening.

The determination of 3D structural geometries are key to understand the binding mode of the active compounds and to validate the present approach. The comparison between the crystal structures of **3a**, **3b** and **8b** and the different docking poses showed a perfect overlap of the pharmacophore while differences were mainly observed in the conformation of the cyclohexyl group of **3a** (Figure 2.1.).

2.2   Experimental session

Compound **3a** is the first crystallographic hit in the series. Starting from this compound, **8b** and **3b** were designed in the second screening. All ligands share the same central 5-(hydrazinylmethyl)-1H-tetrazole moiety, but they differentiate in the substituents attached to it. The central moiety interacts in the binding pocket in all ligands in a similar way: the carboxylic groups of the catalytic dyad (Asp35 and Asp219) interact through hydrogen bonds with the primary amine of the hydrazine (2.7-3.4 Å) while N3 and N4 nitrogen atoms of the tetrazole ring is within H-bonding distance with the backbone nitrogen of Gly80 (2.9 Å and 3.2 Å). In addition, N3 is involved in further binding to the backbone nitrogen

of Asp81 (3.2 Å) (Figure 2.3. and Figure 2.4.). Compound **3a** binds in two orientations and **3a_a** is the most favorable one (Figure 2.4. **(b)**).
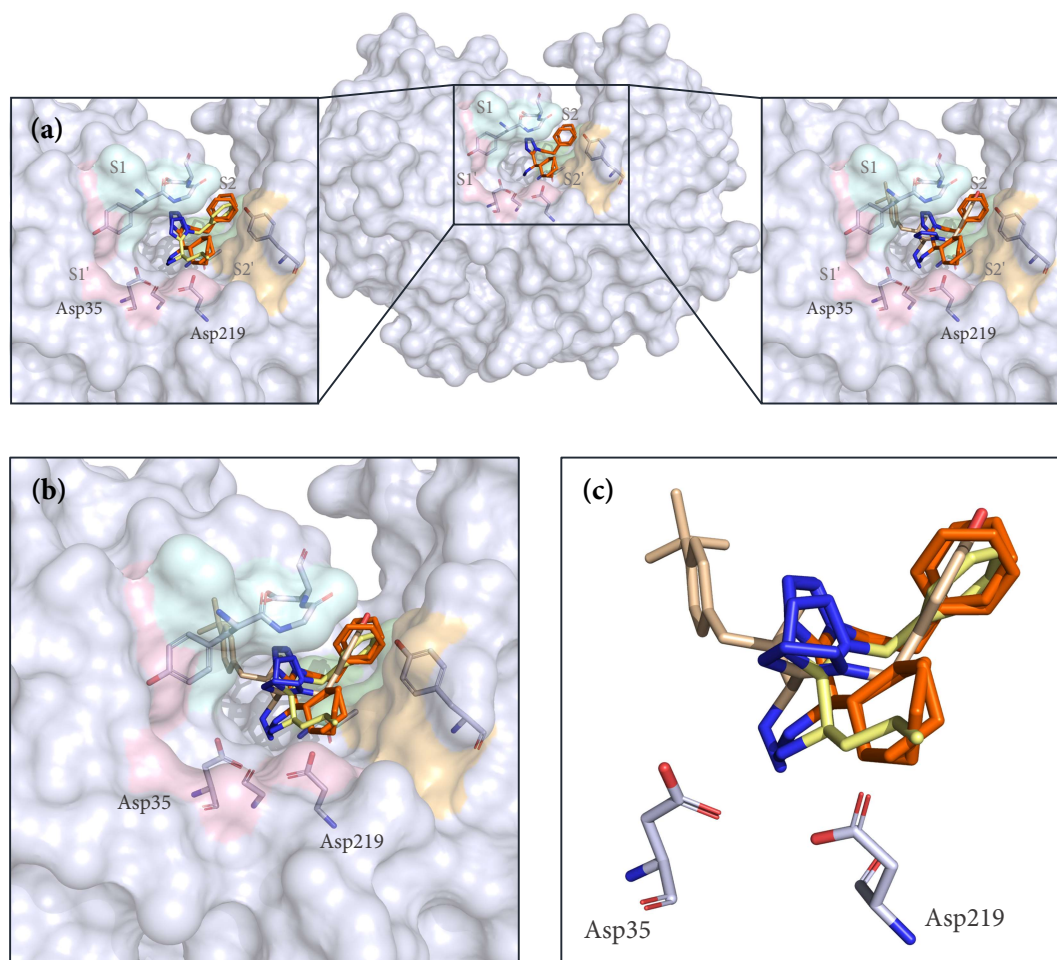
Figure 2.3. a) Surface representation of EP colored in light grey with the first crystallographic hit **3a** colored in orange as a stick model. The hydrophobic S1' pocket is colored in yellow, the S2' pocket in pink and the hydrophilic S2 pocket in green. The S1 pocket is located behind the flap region (Gly80-Asp81). The catalytic dyad (Asp35 and Asp219) is represented by stick models. Oxygen atoms are colored in red and nitrogen atoms in blue. On the right, close-up view of the accommodation into the binding pocket of **8b** colored in sand and represented as a stick model superposed with **3a**. Residues involved in interactions with the ligands are shown in grey as stick models and they are labeled using the one-letter code. H-bonds are depicted as grey dashed lines. A single water molecule is represented as a red sphere. On the left, close-up view of the accommodation into the binding pocket of **3b** colored in yellow and represented as a stick models superposed with **3a**. **b, c)** Superimposition of the binding poses of **3a**, **8b** and **3b**. **8b** and **3b** show a different binding mode with rotated orientation of the central tetrazole ring by about 90 degree. **b)** Surface accommodation of the superposed ligand in the S1', S2' and S2 pockets. **c)** Superposition of the three ligands represented as stick models to better show the difference of about 90 degrees of their central tetrazole ring. The same color codes, dashes, sphere, sticks and surface representations apply to all figures.

The superposition of **3a_a** and **3b** shows that their substituents occupy also the same sub-pockets: The benzyl ring of both ligands binds in the mostly hydrophobic S1' pocket while the cyclohexyl moiety of **3a_a** sits between the S1' and S2' pockets, as does the butyl moiety of **3b** (Figure 2.3., Figure 2.4. and Figure 2.6.).
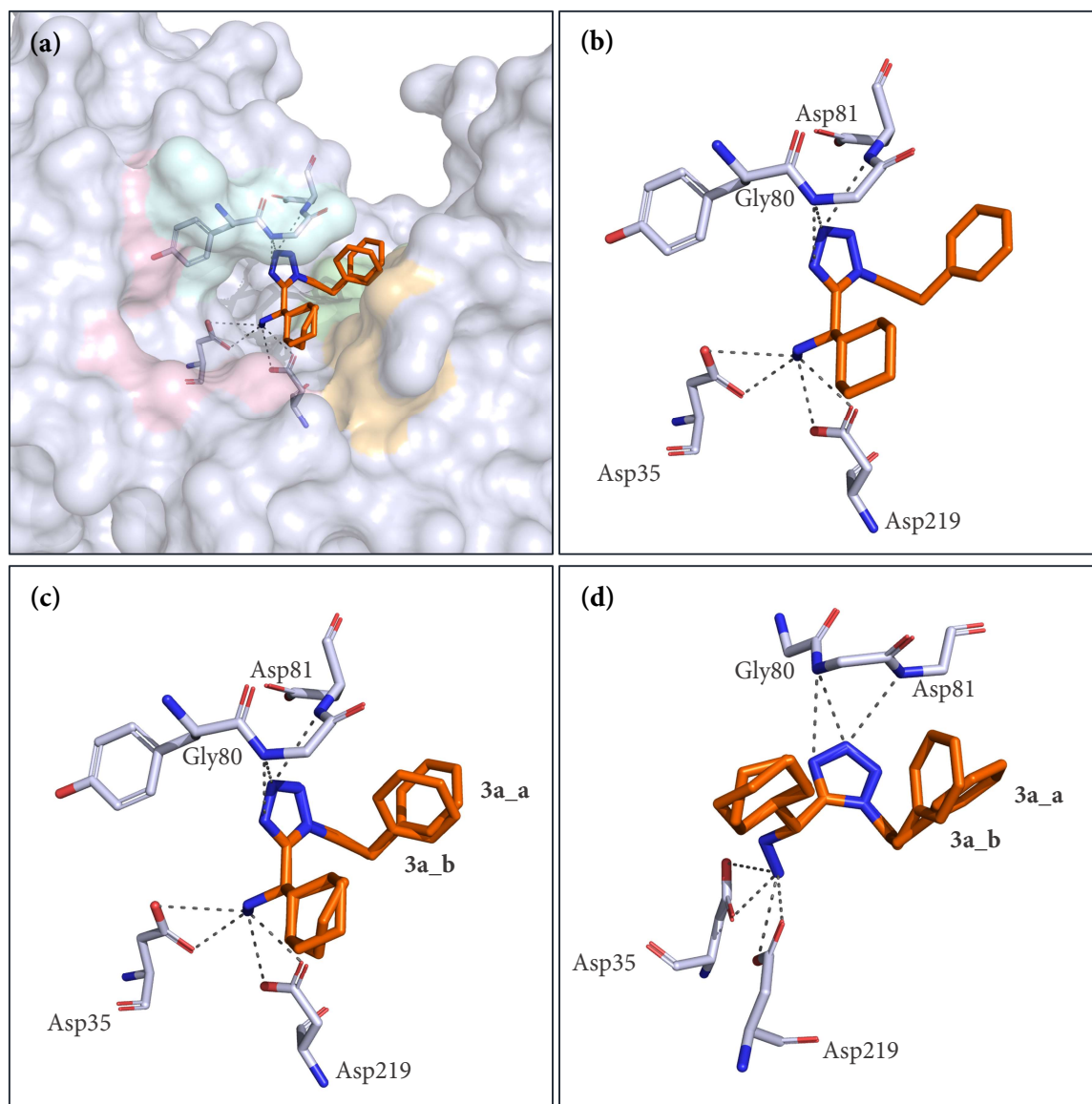
Figure 2.4. *a*) Surface representation of **3a** accommodation into the binding pocket. The ligand adopts two conformation named **3a_a** and **3a_b**. **b)** Stick representation and interaction of the most favorable conformation **3a_a c, d)** Stick representation of the binding mode and ligand interactions are shown in two orientations to visualize all the interaction. Front **c)** and back **d)** view of **3a** binding mode. **c)** The carboxylic groups of the catalytic dyad (Asp35 and Asp219) interact through hydrogen bonds with the hydrazine primary amine. **d)** N3 and N4 of the tetrazole ring forms further hydrogen bonds with the backbone nitrogen of Gly80 and Asp81.

Compound **8b** forms direct interactions with the catalytic dyad (2.7-3.0 Å) but also binds indirectly to the carbonyl group of Gly37 (2.7 Å) mediated through W543 (2.9 Å). N3 of the tetrazole ring is involved in a hydrogen bond with the nitrogen backbone of Gly80 (302 Å) which is part of the flap region. The benzodioxolic motif penetrates into the hydrophilic S2 pocket where it is involved in a hydrogen bond with the OH group of Tyr226 (3.5 Å). At the other end of the molecule, the tert-butylbenzene ring penetrates deeply into the S1 pocket, closely behind the flap region, in the direction of Phe116 and Leu125 (Figure 2.3. and Figure 2.5).
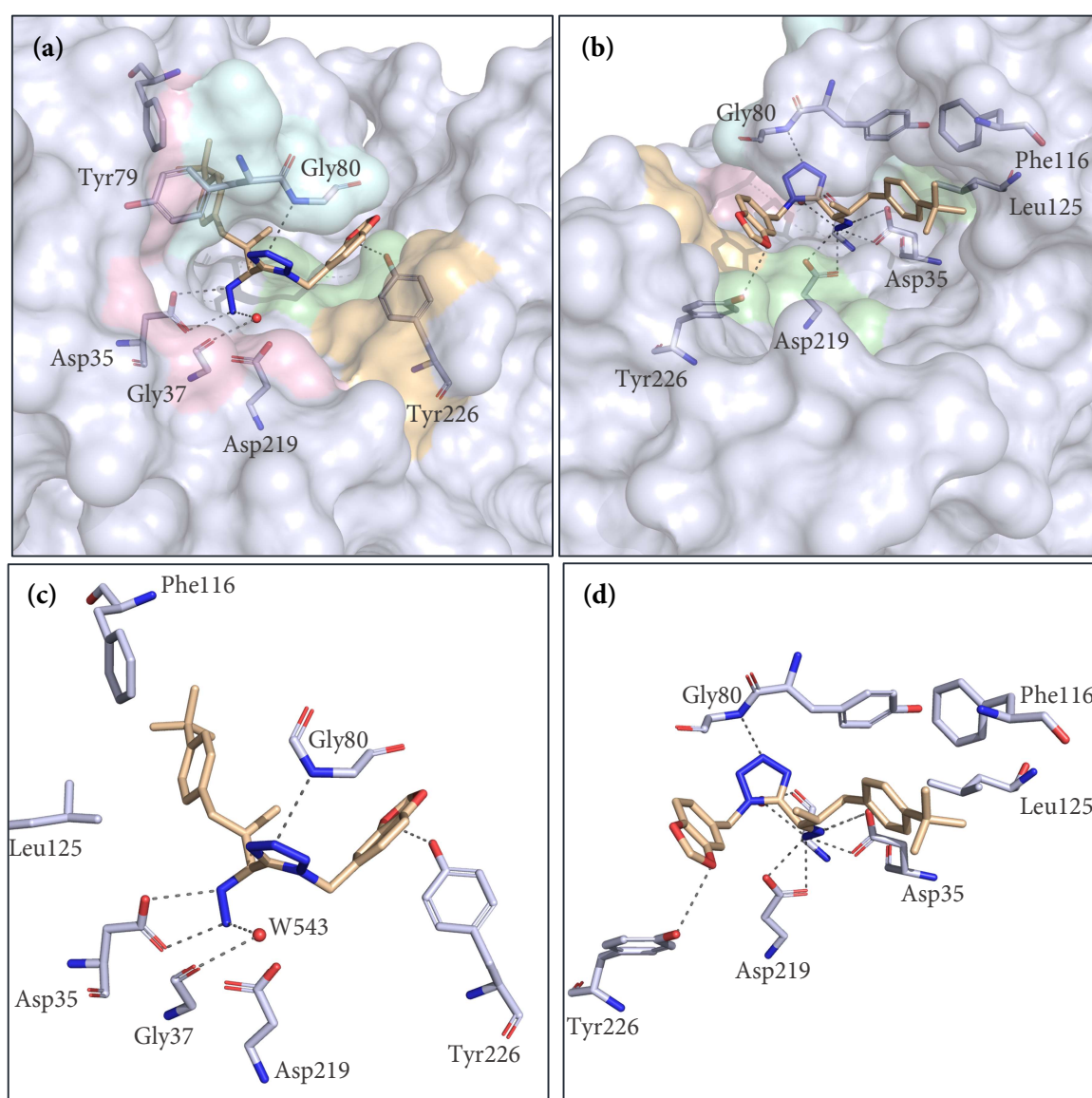
Figure 2.5. *a*) Surface representation of **8b** accommodation into the binding pocket visualized from the front and **b)** from the back to better show the whole surface involve in the binding. **c, d)** Strick representation of the binding mode and ligand interactions are shown in two flipped orientation to visualize all the interaction.

The tetrazole ring is oriented differently in **8b** and **3b**. As a mutual superposition of both ligands shows, there is a difference in orientation of 90 degrees orizontally between the two ligands. In addition, they also differ in the quality of the observed electron density. While for **8b**, the electron density was completely visible after molecular replacement, for **3b**, the obtained density was more difficult to interpret because it was incomplete for the ligand. **3b** was built step by step into the electron density with only slight improvements between consecutive refinement steps. In addition, there is a residual of positive and negative difference electron density even in the final step of the refinement (Appendix 6.1.). This could be due to a possible double conformation as well as to an incomplete occupancy of **3b** at the active site. However, after a few cycles of refinement the occupancy for a second conformation remained at only 14%. Nevertheless, it is interesting to note that **3b** creates some disorder of the protein main chain next to the flap region. Because of this, there is an interruption of the chain between residues Tyr79 and Asp81 and it was not possible to build Gly80 into the electron density. This disorder may explain the remaining residual positive and negative electron density that is observed in the active site (Figure 2.6. and Appendix 6.1.).
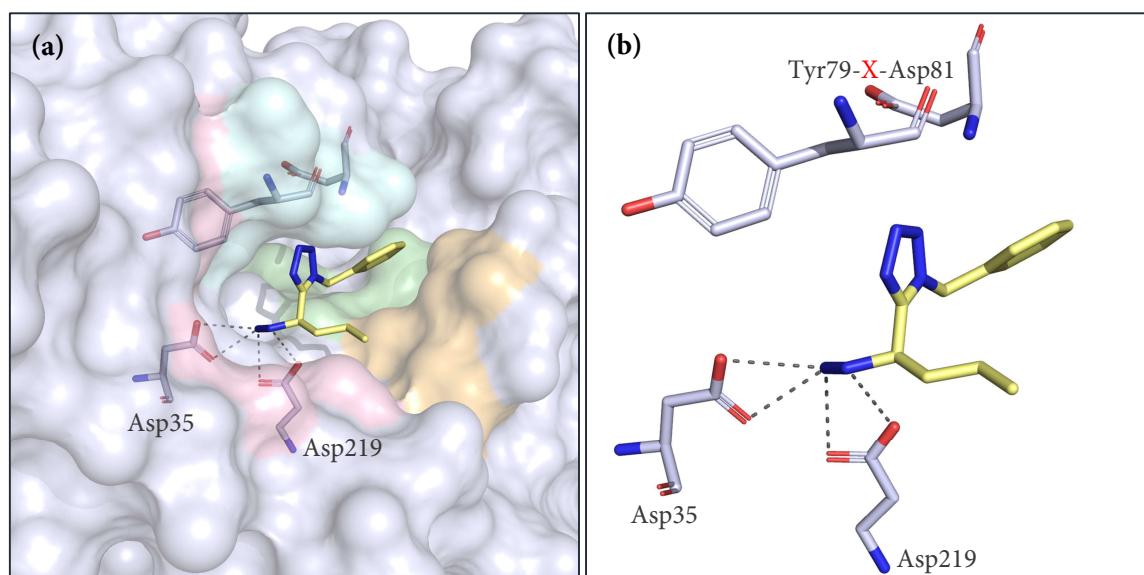
Figure 2.6. *a*) Surface representation of **3b** accommodated in the binding pocket. **b)** Stick representation of the binding mode and ligand interactions with the catalytic dyad (Asp35 and Asp219). In the flap region, interruption of the chain between residues Tyr79 and Asp81. The Gly80 missing residue is represented as an "X".

3. Interpreting unusual electron density

Although EP is a well-known protein, simple to crystallize and expected to be straightforward, under certain circumstances it may show unusual behavior. In particular, some compounds of the second round of synthesis, showed a discrepancy between the soaked fragment and the crystal structure resulting in complex and difficult to interpret electron density maps.

The two most problematic cases, compounds **12** and **22** were then further investigated, using ESI-MS experiments and polder maps to find an explanation for the unexpected electron density in the protease-bound state.

3.1   EP in complex with compound 22

Compound **22** binds EP in the S6 pocket where the fragment showed a refined occupancy of 70 %. The whole fragment was built into the electron density maps. However, one of the two methylene groups which connects the phenylethyl moiety with the tetrazole ring showed neither positive nor negative residual difference electron density. Probably, this may be explained by the high flexibility of the ethyl chain (Figure 3.2.). In addition, further positive electron density was found next to the catalytic dyad. However, there was no residual density which could be assigned to the tetrazole ring. It was therefore assumed that the ring was first opened and then broken into several fragments which are responsible for the finally observed residual electronic density, but which were difficult to assign and fully interpreted as one unique product.
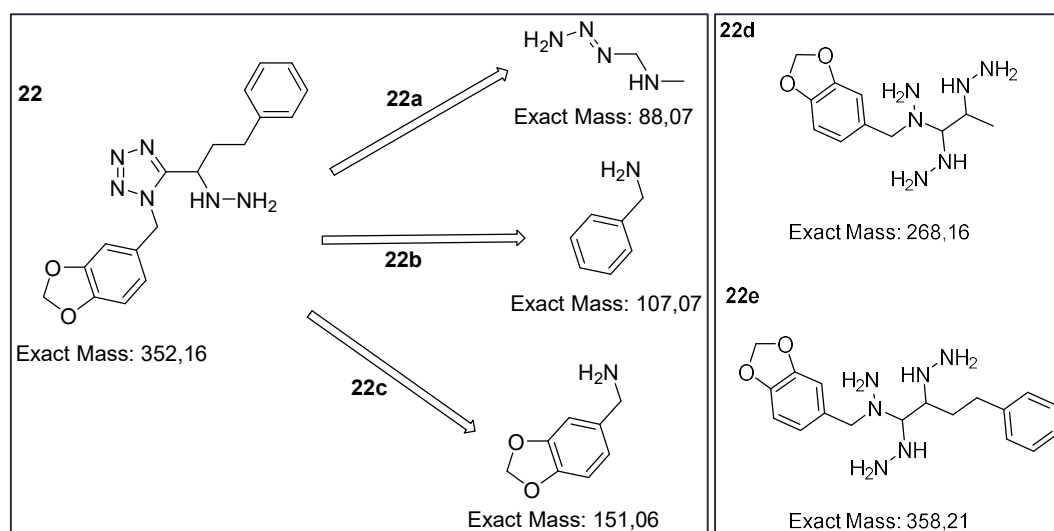
Figure 3.1. Chemical structure and exact mass of the soaked compound **22** and the hypothesized fragments **22a**, **22b**, **22c** **22d** and **22e** which were assigned to the electron density maps next to the catalytic dyad, after several cycle of polder maps and refinement.

Initially, the fragments **22a** and **22b** (Figure 3.2.) were built into the electron density maps. After refinement, two positive electron density picks appeared next to the phenyl ring, probably referred to the two oxygen atoms of the 1,3-benzodioxolic ring (**22c**, Figure 3.1. and **(c)**, Figure 3.2.). Further residual positive electron density also appeared next to **22a** which was still difficult to assign with confidence to any atom of compound **22**. However, no electron density could still be assigned to the putatively present tetrazole ring.
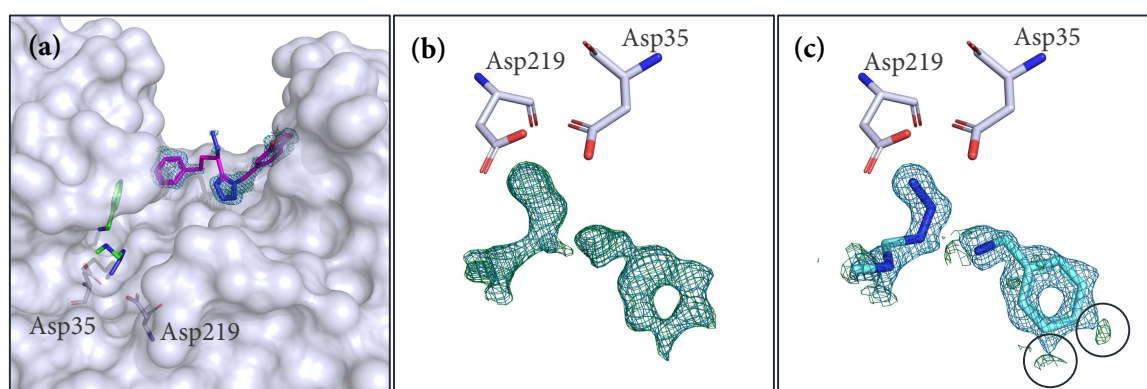
Figure 3.2. Surface representation of EP with the bound compound **22** (magenta stick) in the S6 pocket. Next to the catalytic dyad (Asp35 and Asp219), the hypnotized fragments **22a** and **22b** (green stick) based on the electron density maps. Electron density map before **(b)** and after **(c)** building **22a** and **22b** in the residual density (aquamarine stick). The two positive electron density picks next to the benzyl ring (dark grey circles), probably referred to the two oxygen atoms of the 1,3-benzodioxolic ring The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density).

246

Subsequentially, a polder map was generated with Phenix [12] and additional residual density appeared (Figure 3.3. (a), (b)). Polder maps are indeed widely used in crystallography to better identify a small region of interest. In particular, it consists of excluding the bulk solvent around the omitted region of interest which is usually around the supposed ligand in order to better visualize weak densities which can be obscured by the bulk solvent. In fact, the residual positive electron density which was initially assigned to **22c** became more interpretable and the 1,3-benzodioxolic ring was then built into the electron density and refined. However, the electron density maps after refinement cannot fully justified a fully intact 1,3-benzodioxolic ring (Figure 3.3. (c)).
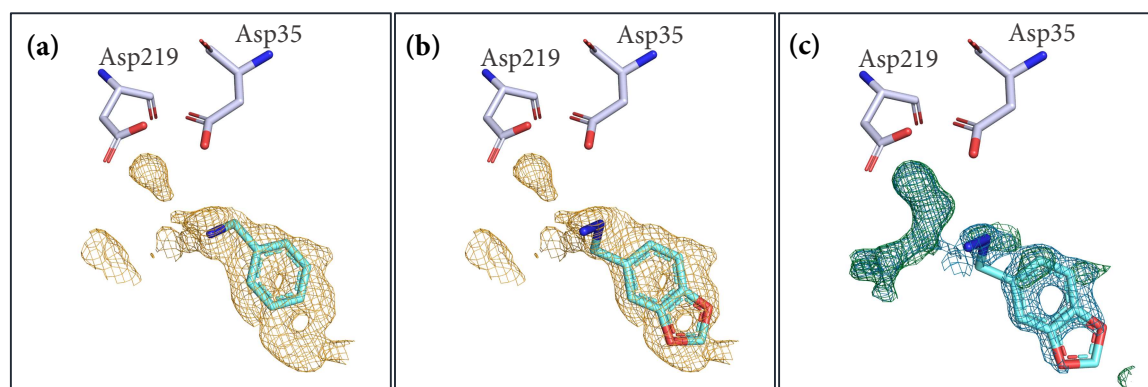
Figure 3.3. (a) Polder maps around **22b**. **(b)** hypnotized fragment **22c** based on the polder map. **(c)** Electron density maps after refinement of **22c**. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density). The polder map is displayed at a contour level of 2σ (gold mesh).

The entire compound 22 was built in the electron density maps after several cycles of polder maps and refinement by which more and more of a conclusive residual electron density appeared.

In particular, the atomic coordinates of the refined structure are shown in Figure 3.3.**(c)**. They were then used as input for the generation of a second polder map where additional residual density appeared. Based on the obtained residual density and assuming a possible rupture of the tetrazole ring, fragment **22d** (Figure 3.4. (b)) was built into the electronic density and further refined (Figure 3.4. (c)). As shown in Figure 3.4. (c), additional positive

electron density appeared at the end of the long aliphatic chain. The atomic coordinates of the assigned and subsequently refined structure is shown in Figure 3.4. (c) and it was used as input file for the next generation of an additional polder map.
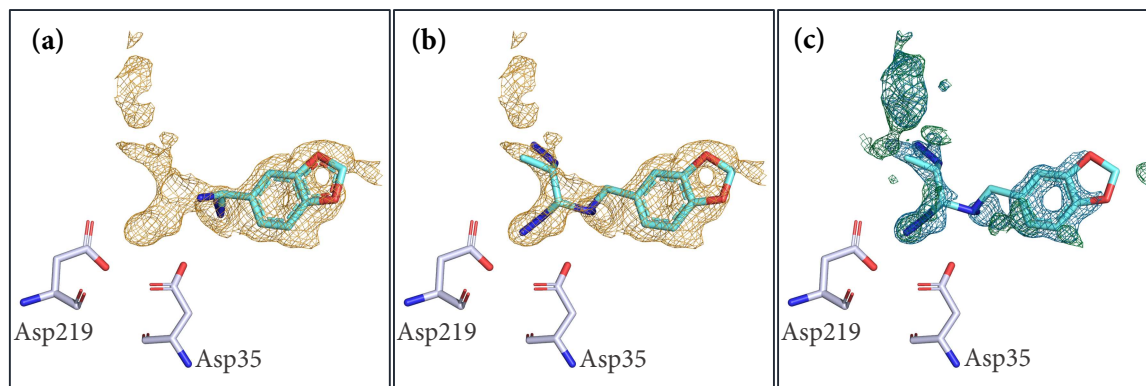
Figure 3.4. (a) Polder maps around **22c**. **(b)** hypnotized fragment **22d** based on the polder map. **(c)** Electron density maps after refinement of **22d**. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density). The polder map is displayed at a contour level of 2σ (gold mesh).

The third polder maps revealed additional positive electron density which can be assigned to the terminal phenyl ring in likely a distribution over two conformations (Figure 3.5. **(b)**, **(c)**).
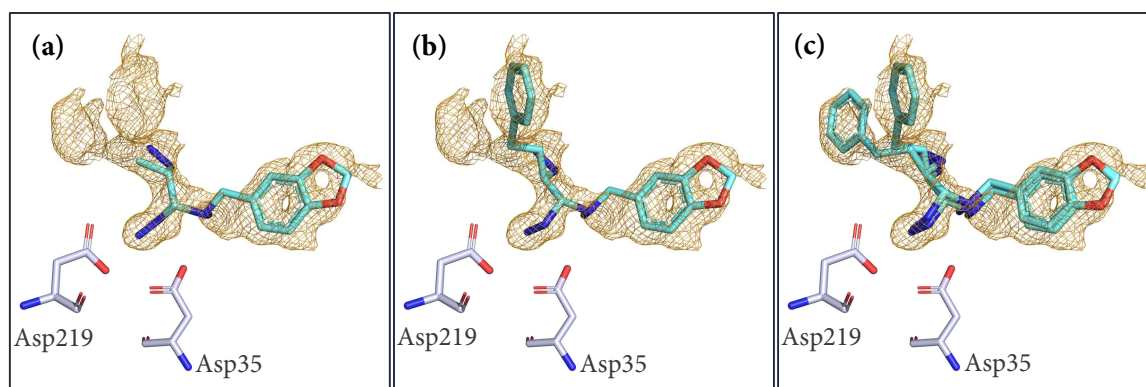
Figure 3.5. (a) Polder maps around **22d**. **(b)** Hypnotized fragment **22e** with opened tetrazole ring based on the polder map in a single or **(c)** double configuration. The polder maps are displayed at a contour level of 2σ (gold mesh).

As shown in Figure 3.6., the hypothesized compound was entirely built in the electronic density using several cycles of polder maps and refinement. However, the electronic density

after the last refinement cycle still does not fully justify to built the entire compound without assumption of a pronounced disorder.
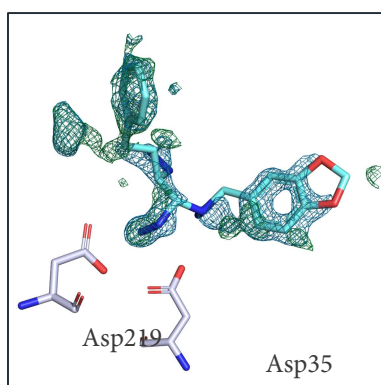
Figure 3.6. The hypnotized compound **22e** with the opened tetrazole ring refined in a single configuration. The residual positive density next to the benzyl ring may indicate a second configuration. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density).

The described strategy of iterative alternating cycles of polder maps, coordinate assignment and refinement were performed to assess how far it was possible to proceed with the interpretation of electron density maps and to hopefully detect a fully closed tetrazole ring. However, neither positive nor negative residual difference electron density could justify to build the tetrazole ring into the density.

Despite the generation of multiple polder maps, it was still difficult to interpret the observed electron density maps. ESI-MS experiments were therefore carried out and the following aspects were addressed:

(1)- Does **22** product material used for the crystallization trials contain other molecular species that can explain the observed density?

(2)- Was the compound used for the soaking trial pure?

(3)- Was the acidity present in the protein environment (pH: 4.6) sufficient to break or degrade the expected product **22**?

(4)- Was the protease activity responsible to break or degrade any starting material of pure **22**?

As indicated by the MS-ESI spectra reported in Figure 3.7. the original powder of the substance is pure **(a)** and there is no substantial difference even in presence **(c)** or absence **(b)** of HCl in the soaking condition. In all cases there are two dominant species: the peak at 135 and 149 m/z can be assigned to **22** fragmentations confirmed by the CFM-OD spectral prediction software (http://cfmid.wishartlab.com/predict). The peak at 353 m/z referring to **22**+H is present when the compound is solubilized in the soaking buffer while it is absent in the MS spectra which refers to the pure powder substance indicating that the compound is completely fragmented in the most stable specie at 135 m/z.

Subsequently, it was checked whether the presence and the catalytic activity of the enzyme breaks down or degrades compound **22**. An equivalent of EP was therefore incubated overnight in 1 mL of **22** solubilized in the soaking buffer with the addition of HCl. The next day, the protein was separated from the compound using a 3 mL (3000 Da membrane cut-off) Vivaspin and the flow through was collected and analyzed with ESI-MS while the upper part was discard in order to avoid spectra which are difficult to interpret due to the high complexity of protein sample. As shown in Figure 3.7. **(d)**, two new species appeared in the ESI-MS spectra: the peak at 182 m/z may refers to **22** fragmentation while the peak at 115 m/z could not be assigned to any species derived from the fragmentation of compound **22**.
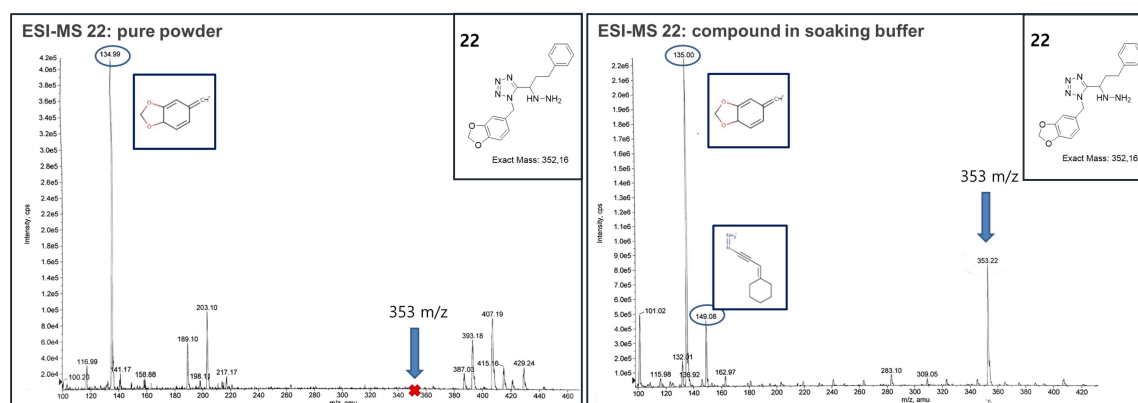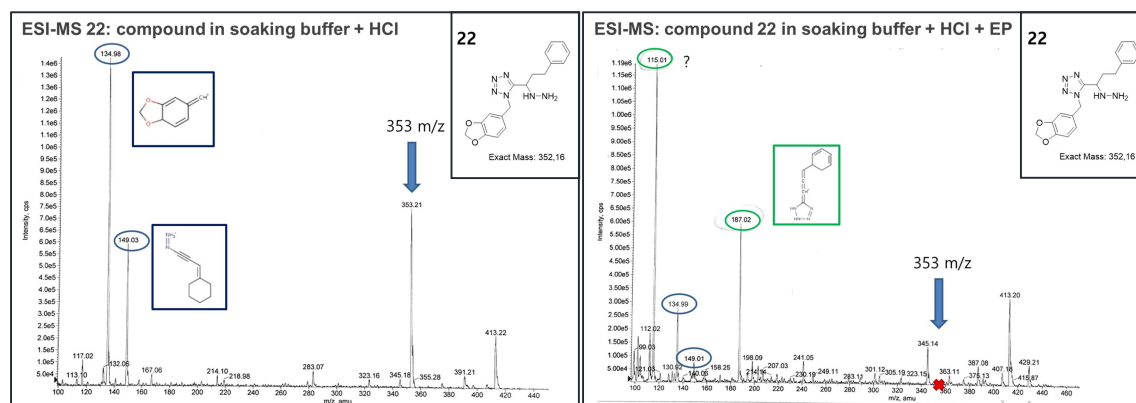
Figure 3.7. ESI-MS spectra of compound **22** **(a)** pure powder, **(b)** in soaking condition, **(c)** in soaking condition with addition of one equivalent of HCl, **(c)** after ON incubation with one equivalent of EP. The species derived from **22** fragmentation are encircled in blue and their chemical structures are represented in blue boxes. The additional species, probably derived from EP catalytic activity, are encircled in green.

It is known from the literature that azidoazomethines can undergo a concerted cyclization reaction to tetrazoles and *vice versa* as back reaction according to the Scheme depicted in Figure 3.8. The direction of this equilibrium depends on the nature of the substituents in the tetrazole ring. Electron-withdrawing groups (EWG) on the azomethine nitrogen favor the open form, whereas hydrogen and electron pushing substituents favor the tetrazole moiety [13].

Assuming the opening of the tetrazole ring as indicated in **4** (Figure 3.8.) and assigning to actual groups R and R' found as corresponding substituents in compound **22**, it is possible to notice that the substituent R has an exact mass of 150 Da while R' of 164 Da. In addition, supposing that R' would form an adduct with sodium ions, it would have an exact mass of 164+23=187 Da, which corresponds to one of the two additional peaks of the spectrum in Figure 3.7. (d) encircled in green.
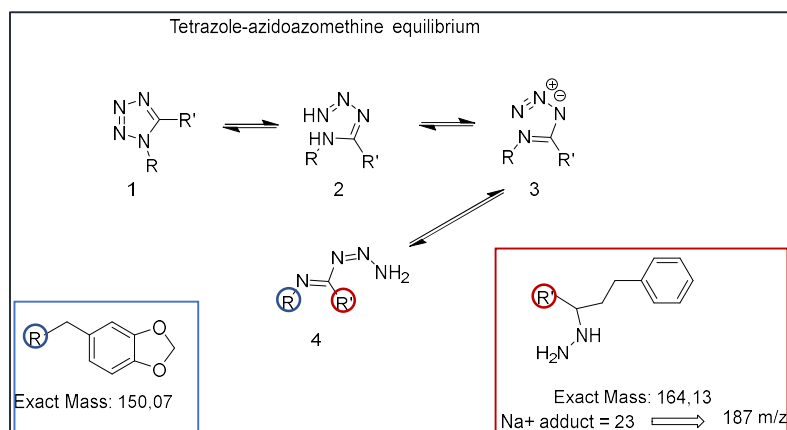
Figure 3.8. Tetrazole-azidoazomethine equilibrium showing the opening of the tetrazole ring. R and R' are the two substituents of compound **22**. Assuming that R' formed an adduct with sodium, the species has a 187 m/z which correspond to one of the two additional species found after ON incubation of 22 with EP (Figure 3.7. **(d)**, Figure 3.10).

Based on these findings the following mechanism of fragmentation of compound **22** was supposed (Figure 3.9.) in order to explain, admittedly in a purely speculative manner, why the additional peaks at 115 m/z in Figure 3.7. (d) encircle in green could have occurred.
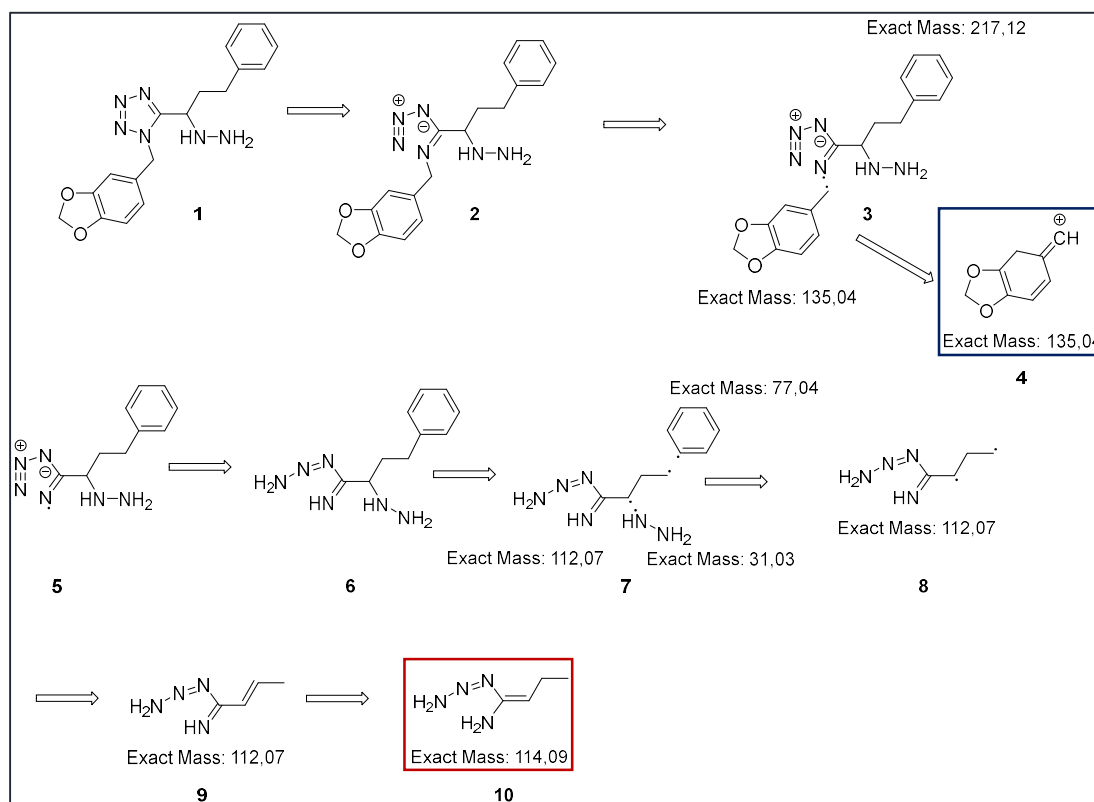
Figure 3.9. Supposed mechanism of fragmentation of compound **22**. The encircle species in blue which has an exact mass of 135, is always present in ESI-MS spectra (Figure 3.7.) and is the most stable species. The encircle species in red which has an exact mass of 114 may correspond to the additional species fond after ON incubation of **22** with EP (Figure 3.7. **(d)**, Figure 3.10).

After the opening of the ring according to the azidoazomethine-tetrazole equilibrium, the product **3** (Figure 3.9.) might be fragmented by forming the species **4** at 135 m/z, which is always present in the spectra shown in Figure 3.7. because it is likely the most stable species, and **5** which had an exact mass of 217 Da. The latter might undergo successive fragmentations until it formed the species **10** which had an exact mass of 114. Assuming that this species existed in the form of **10**+H, it would have 115 m/z which corresponds to the highest peak in Figure 3.7 (d) encircle in green.

Assuming that these fragmentations may be acceptable, the two additional peaks encircled in green and in red in Figure 3.10. could be justified, although in a purely speculative way. This phenomenon is probably due to the proteolytic activity of the enzyme which resulted in more cleavage products for some compounds compared to others. Until now, there is no clear and conclusive idea of what really happened and the chemistry described above is a possible hypothesis which cannot, however, justified the additional electron density found in the crystal structure next to the catalytic center.
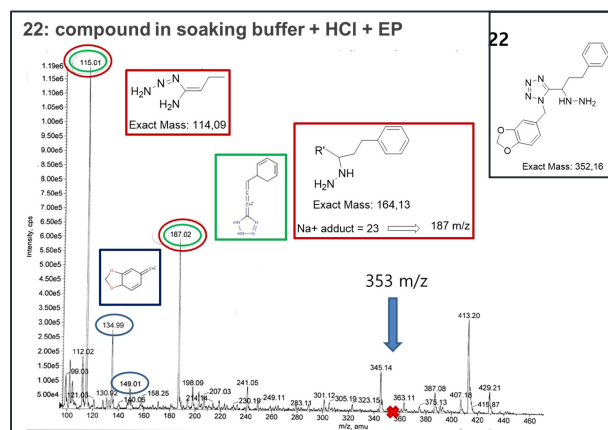
Figure 3.10. Supposed fragmentation species of ESI-MS spectra referred to compound **22** after overnight incubation with EP (Figure 3.7. (**d**)). The encircle species in blue are always present in ESI-MS spectra (Figure 3.7.) while the encircled species in green are only present after incubation with EP. The species at 187 m/z may derived either from **22** fragmentation (green) or R' which formed an adduct with sodium (red). The encircle species in red are the supposed fragments of compound **22** described in the fragmentation mechanism in Figure 3.9.

3.2   EP in complex with compound 12

The high resolution of the complex with the compound **12** (1.12 Å) revealed a clearer and better interpretable electron density maps, compared to that of compound **22**. It seems to support the idea of an opening of the tetrazole ring which was hypothesized in the previous case (Section 3.2.).

In fact, also in this case, there was a discrepancy between the soaked compound **12** and what the residual electron density map may indicated. According to the electron density, compound **12a** and **12b** were then hypothesized and build into the electron density (Figure 3.11.).
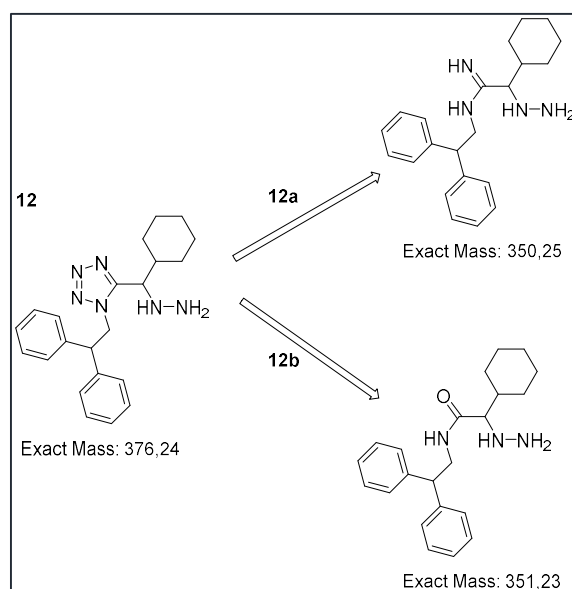
Figure 3.11. Chemical structure of the soaked compound **12** and the hypnotized fragments **12a** and **12b**, supposedly derived from the opening of the tetrazole ring.

In Figure 3.12. is displayed as 2mFo–DFc electron density maps before building compound **12** into the electron density **(a)** and after refinement **(b)**. As already suggested by the positive electron density in Figure 3.12 **(a)**, the tetrazole ring is missing and a linear chain appeared instead. In addition, the putative geometry of the tetrazole ring would be incorrect and, after refinement, compound **12** moved also out of the electron density maps **(b)**.
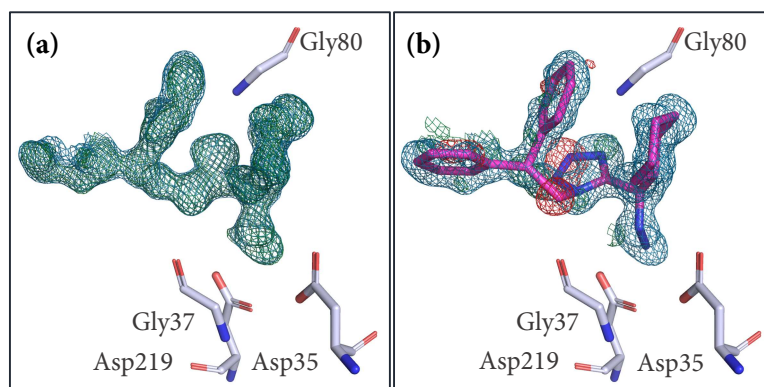
Figure 3.12. Electron density map before **(a)** and after **(b)** building compound **12** in the electron density maps. There is no indication of the tetrazole ring in the electron density maps. After refinement, negative electron density appeared and the compound moved out of the electron density in correspondence of the tetrazole ring The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density) and at –3σ (red mesh, negative density).

Compound **12b** and **12b** were then build into the electron density and refined. Both compounds fitted in the electron density very nicely and their positive residual density indicated also a possible double configuration of one of the two benzyl ring. In addition, the nitrogen atoms of the acetamidine (**12a**) and amide (**12b**) groups are involved in a hydrogen bond with the backbone oxygen of Gly37 (Figure 3.13.). Although the acetamidine group would result directly from the opening of the tetrazole ring by elimination of N_2, it is less stable in solution than the amide group of **12b**. The latter would also be involved in an additional hydrogen bond between the oxygen atom of the amide moiety and the nitrogen backbone of Gly80 (Figure 3.13. **(b)**).
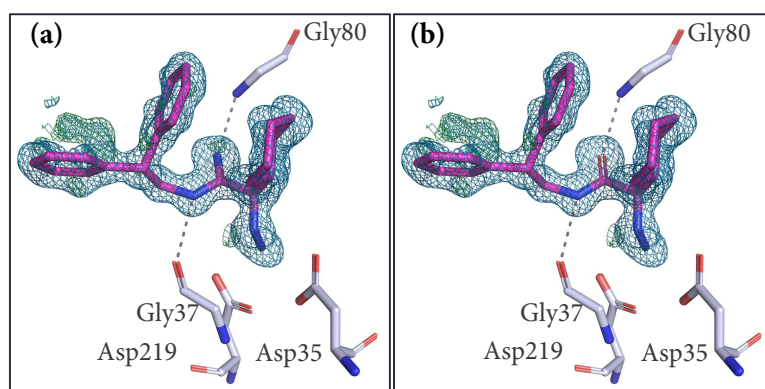
Figure 3.13. Electron density map of **(a)** **12a** and **(b)** **12b** built into the electron density maps. Both compounds fitted perfectly in the electron density. They are involved in a hydrogen bond with the backbone oxygen of Gly37. **12b** is also involved in an additional hydrogen bond between the oxygen atom of the amide moiety and the nitrogen backbone of Gly80. The residual positive density may indicate two conformations adopted by the benzyl group. The 2mFo–DFc electron density map is displayed at a contour level of 1σ (blue mesh), and the mFo–DFc electron density map at +3σ (green mesh, positive density) and at –3σ (red mesh, negative density).

However, ESI-MS experiments showed that neither of the two hypothesized species was contained in the sample. Therefore, it was not possible to fully interpret the observed electron density and to build either **12a** or **12b** with confidence into the electron density maps.

4. Conclusion and Outlook

In the present work a novel pipeline was introduce for the rapid development of novel aspartyl protease inhibitors having a hydrazine moiety as central *anchor*. In particular, the classic concept of *anchor* , which was originally developed to find bioisosteres of an amino acids side chain deeply buried in the interface of a PPI, was expanded and applied as binding motif in the active site of an enzyme with the idea to make it available for fragment-based drug discovery. In the established protocol, the choice of the fragment-anchor is the most important step and has to include a sequence of atoms which significantly contribute to the binding interaction between the designed ligands and the protein of interest. In addition, not only MCR scaffolds but a broader chemistry could be applied. By using commercially available starting material and less efforts in follow-up synthesis, it could be possible to prioritize docking hits as potential candidates for chemical synthesis instead of applying strategies involving a time-demanding and random trial component.

Interestingly, compound **8b** showed the lowest IC_{50} among the two series and it was the only one that was completely visible in the electron density after molecular replacement, while for **3a** and **3b**, the obtained density was more difficult to interpret because of the incompletely defined density for the ligand. Nevertheless, the comparison between the crystal structures and the different docking poses showed a perfect overlap of the pharmacophore while differences were only observed in the orientation of the substituent attached to the tetrazole ring.

Although compounds **12** and **22** showed IC_{50} values comparable to those of **3a** and **3b**, it was not possible to build them in electronic density straightforwardly because there was no residual density which could be assigned to the tetrazole ring. It was therefore assumed that the ring would first open and then break into several fragments which appeared in the residual electronic density, but which were difficult to assign and fully interpret by both, ESI-MS and the generation of several polder maps.

It is important not to underestimate the functional properties and the vaste opening of the binding site in a protein such as EP which has been well studied and characterized for a long time. However, under certain circumstances, EP may show unexpected behavior that is difficult to explain, prticulary if ligands with limited stability are studied.

5. Material and Method

5.1 Crystal Preparation and Soaking

Endothiapepsin was obtained from Suparen 600® from DSM Food Specialities (DSM Food Specialties, AX Delft, Netherlands). After buffer exchange with 0.1 M sodium acetate at pH 4.6 using a Vivaspin 20 with a molecular weight cutoff of 10 kDa, the purified protein was concentrated to 5 mg/mL suitable for crystallization.

EP crystals appeared after 2 days using the sitting drop vapor diffusion method at 19 °C. The crystallization drop consisted of an equal volume of reservoir and protein solution. The reservoir is composed of 0.1 M NH4Ac, 0.1 M NaAc, and 24–30% (w/v) PEG 4000 at pH of 4.6. After some minutes of equilibration, the streak seeding method was used to enhance crystal quality. Ligands **3a**, **3b** and **8b** were dissolved at 1M stock concentration in DMSO and soaked for 28 h into the uncomplexed protein crystals at a final concentration of 90 mM (9% (v/v) DMSO). The soaking drops consisted of 70 mM NH4Ac and 70 mM NaAc, 16–20% (w/v) PEG 4000, and 19–23% (v/v) glycerol at pH 4.6. After soaking, the crystals were quickly flash-frozen and ready for data collection.

5.2 Data collection, processing, structure determination and refinement

Diffraction data for compound **3a** have been collected on Bruker Iμs Microfocus operated inhouse. Datasets were collected on a MAR Scanner 345 mm Image Plate detector at a wavelength of 1.542 Å. Diffraction data for **8b** and **3b** have been collected on BL14.1 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin. Datasets were collected on a Dectris Pilatus 6M pixel detector at a wavelength of 0.91841 Å. XDS and XDSAPP were used for indexing, integration and scaling of the datasets. The structures were determined by molecular replacement using Phaser.[7] Model refinement (xyz coordinates, individual B factors, occupancies) was done with Phenix.refine (version

259

1.15.2-3472) and model building into electron density maps (2mFo–DFc and mFo–DFc) with Coot (0.8.9). The resolution cut-off for the refinement of **8b** was set to 1.15 Å due to the high R-factor between successive shells. A randomly chosen subset of 5% of the reflections was excluded from the refinement and used for the calculation of R(free). As a first refinement step, Cartesian simulated annealing was performed (default settings). B factors for all model atoms (except for hydrogen atoms) were refined anisotropically. Hydrogen atoms (riding model) were added to the amino acids with Phenix.refine. Alternative conformations of amino acid side chains and ligand moieties were assigned to the electron density if an occupancy of at least 20% was obtained after refinement. Chemoinfo
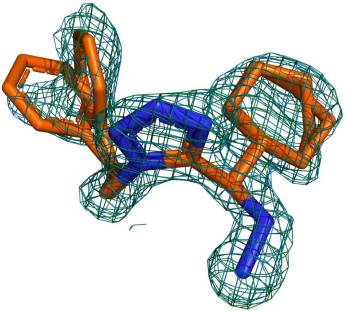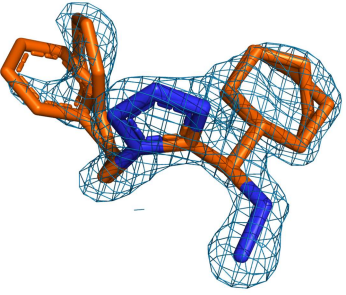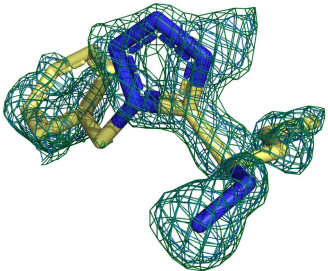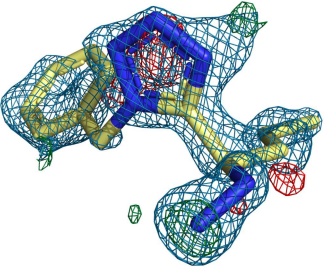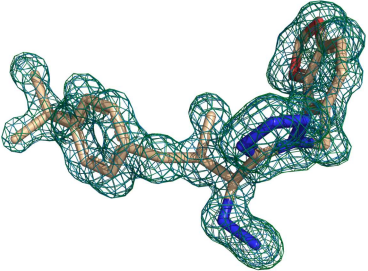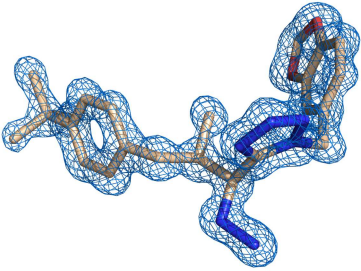
[http://www.cheminfo.org/flavor/malaria/Utilities/SMILES_generator___checker.html] was used for SMILE generation, ligand molecules and restraints were created with the Grade Web Server. [http://grade.globalphasing.org]

5.3   Accession codes

Atomic coordinates and experimental details for the crystal structures of **3a**, **8b** and **3b** (PDB codes 6SCV, 6RON and 6RSV) will be released upon publication.

6. Appendix

6.1   Electron density maps of the refined hits

	Omit maps	**Refined maps**
3a Resolution= 1.70 Å Occupancy= 0.53/0.47 PDB ID= 6SCV		
3b Resolution= 1.10 Å Occupancy= 1.00 PDB ID= 6RSV		
8b Resolution= 1.13 Å Occupancy= 1.00 PDB ID= 6RON		

6.2   Data collection and refinement statistics [a]

PDB code	6SCV	6RON	6RSV
	3a-EP	**8b**-EP	**3b**-EP
(A) Data collection and processing			
X-ray source	Bruker Iμs Microfocus	BESSY II, 14.1	BESSY II, 14.1
Wavelength (Å)	1.54200	0.91840	0.91840
space group	$P2_1$	$P2_1$	$P2_1$
unit cell parameters a, b, c (Å)	45.28, 73.04, 52.88	45.76, 73.35, 53.17	45.40, 73.27, 52.95,
β (°)	109.38	109.65	109.75
Matthews coefficient [b] ($Å^3$/Da)	2.4	2.5	2.4
solvent content [b] (%)	49.6	50.6	49.9
(B) Diffraction data			
resolution range (Å)	42.72-1.70	43.10-1.13	42.7-1.10
unique reflections	35726 (5629)	122038 (18907)	129009 (20112)
Redundancy	4.0 (3.8)	3.7 (3.4)	3.6 (3.4)
$R(I)_{sym}$ (%)	3.8 (22.4)	4.3 (48.3)	2.9 (34.3)
Wilson B factor ($Å^2$)	13.9	10.3	14.5
Completeness (%)	99.3 (97.7)	97.7 (93.9)	97.7 (94.4)
CC (1/2) (%) [e]	99.9 (94.7)	99.8 (86.9)	100 (88.9)
$<I/\sigma(I)>$	24.8 (6.3)	15.8 (2.6)	21.2 (3.2)
(C) Refinement			
resolution range (Å)	39.58-1.70	43.10-1.15	42.7-1.10
reflections used in refinement	35700	115124	128984
final R value for all reflections (work/free) (%)	14.0/16.4	13.4/15.3	12.6/14.4
protein residues	330	330	329
water molecules	274	330	312
ligand atoms	44	31	18
other ligands atoms (GOL | DMSO)	6	6 |4	6 |4
RMSD from ideality: bond lengths (Å)	0.005	0.008	0.010
RMSD from ideality: bond angles (°)	0.8	1.0	1.2
Ramachandran plot [c]			
residues in most favored regions (%)	93.9	93.5	93.8
residues in additionally allowed regions (%)	6.1	6.5	6.2
residues in generously allowed regions (%)	0.0	0.0	0.0
residues in disallowed regions (%)	0.0	0.0	0.0
Mean B factor protein ($Å^2$) [d]	15.4	12.4	13.2
Mean B factor ligand ($Å^2$) [d]	35.5	16.3	26.6
Mean B factor water molecules ($Å^2$) [d]	24.8	27.8	28.4
Mean B factor other ligands (GOL | DMSO)	24.0	16.4 | 30.5	13.7 | 35.5

[a] Values in parenthesis describe the highest resolution shell. [b] Calculated with MATTPROB[14]. [c] Calculated with PROCHECK[15].
[d] Mean B factors were calculated with MOLEMAN. [e] R(I)sym = (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i)))

7. References

[1] L. Coates *et al.*, "The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction," *J. Am. Chem. Soc.*, vol. 130, no. 23, pp. 7235–7237, 2008.

[2] P. B. Szecsi *et al.*, "The aspartic proteases: Preface," *Scand. J. Clin. Lab. Investig. Suppl.*, vol. 52, no. 210, p. 2, 1992.

[3] B. M. Dunn *et al.*, "Overview of Pepsin-like Aspartic Peptidases," *Curr. Protoc. Protein Sci.*, vol. 25, no. 1, pp. 21.3.1-21.3.6, 2001.

[4] L. Pearl and T. Blundell, "The active site of aspartic proteinases," *FEBS Lett.*, vol. 174, no. 1, pp. 96–101, 1984.

[5] N. Radeva, "From Active-Site Mapping to Lead Discovery using Fragment-based Approaches on the Aspartic protease Endothiapepsin Dissertation," *PhD Thesis*, 2016.

[6] D. R. Koes *et al.*, "AnchorQuery: Rapid online virtual screening for small-molecule protein–protein interaction inhibitors," *Protein Sci.*, vol. 27, no. 1, pp. 229–232, 2018.

[7] D. Koes *et al.*, "Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists," *PLoS One*, vol. 7, no. 3, 2012.

[8] D. Rajamani *et al.*, "Anchor residues in protein-protein interactions," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 31, pp. 11287–11292, 2004.

[9] T. Clackson and J. A. Wellst, "A Hot Spot of Binding Energy in a Hormone-Receptor Interface Author (s): Tim Clackson and James A . Wells Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2886248 JSTOR is a not-for-profit servi," vol. 267, no.

5196, pp. 383–386, 2016.

[10]   E. Surmiak *et al.*, "Rational design and synthesis of 1,5-disubstituted tetrazoles as potent inhibitors of the MDM2-p53 interaction," *Eur. J. Med. Chem.*, vol. 126, pp. 384–407, 2017.

[11]   P. E. Bunney *et al.*, "Orexin activation counteracts decreases in nonexercise activity thermogenesis (NEAT) caused by high-fat diet," *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, Jul. 2017.

[12]   D. Liebschner *et al.*, "Polder maps: Improving OMIT maps by excluding bulk solvent," *Acta Crystallogr. Sect. D Struct. Biol.*, 2017.

[13]   E. Cubero *et al.*, "Theoretical Study of Azido - Tetrazole Isomerism : Effect of Solvent and Substituents and Mechanism of Isomerization," vol. 1, no. 6, pp. 4723–4731, 1998.

[14]   K. A. Kantardjieff *et al.*, "Proceedings of 10th International Conference on Formal Methods in Computer-Aided Design, FMCAD 2010, Lugano, Switzerland, October 20-23," *Fmcad*, pp. 1865–1871, 2010.

[15]   R. A. Laskowski *et al.*, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, 1993.

Part V: Carbonic Anhydrase (CA)

Kinetic Evaluation of Sulfonamide Series Binding CA Isoforms

1. Introduction

1.1. Carbonic Anhydrase isoforms: an overview

1.1.1. Distribution and structure topology

Carbonic anhydrases (CAs, EC 4.2.1.1) are a family of metalloenzymes which are widly present in mammals (α-class), plants and some prokaryotes (β-class), archaebacteria (γ-class) and diatoms (δ-class). The human (*h*) α-CAs comprise 15 isoforms, 12 of which are catalytic (hCAs I-IV, VA, VB, VI, VII, IX, XII-XIV) and three of which are non-catalytic (hCAs VIII, X and XI) due to the lack of one or more histidine residues that coordinate the Zn^{2+} ion in the active site [1].

Although these isoforms are structurally very similar, they differ in terms of catalytic activity, tissue localization and distribution (Table 1.1.) [1]. In particular, the present study focused on six *h*CA isoforms which are CAI, CAII, CAVII, CAIX, CAXII and CAXIII. hCAs I and II are present in the red blood cells (RBCs) and in the gastrointestinal (GI) tract. They are involved in the efficient transportation of CO_2 among cells and tissues. *h*CAII is also present in eyes, osteoclasts, kidneys, lungs, testes and brain, and is involved in bone resorption, respiration and pH regulation. Furthermore, together with hCAIII and hCAVII, hCAII is supposed to play an antioxidant role in cells which have high rates of oxidation, e.g. hepatocytes, adipose tissues and skeletal muscle fibers[1], [2], [3]. hCA VII seems to be involved in the cerebrospinal fluid production in the central nervous system (CNS). hCA XIII is expressed in the reproductive organs and is supposed to play a role in pH regulation and ensuring proper fertilization. The transmembrane isoforms IX and XII are mainly present in the GI tract. In addition, they are found to be over-expressed in epithelial tumours such as cervix, prostate, kidneys, lungs and breast. In particular, they seem to facilitate adaption of tumours to a hypoxic microenvironment and also promote the growth of metastases.

Table 1.1. *h*CAs isoforms under investigation in the present study

Isoforms	K_{cat} (s^{-1})	Localizzation		Strucrure
		Sub-cellular	Tissue/Organ	
*h*CA I	2.0×10^5	Cytosol	RBCs, GI tract	Monomeric
*h*CA II	1.4×10^6	Cytosol	RBCs, GI tract, eyes, Osteoclasts, kidneys, lungs, testes, brain	Monomeric
*h*CA VII	9.5×10^5	Cytosol	CNS	Monomeric
*h*CA IX	3.8×10^5	Transmembrane	Tumours, GI mucosa	Dimeric
*h*CA XII	4.2×10^5	Transmembrane	Renal, intestinal, reproductive epithelia, eye, tumours	Dimeric
*h*CA XIII	1.5×10^5	Cytosol	Kidneys, brain, lungs, gut, reproductive tract	Monomeric

With respect to the *h*CA topology, except for the transmembrane *h*CA IX and XII isoforms which have a dimeric structure, the other cytosolic *h*CA isoforms exist in monomeric forms. The secondary structure is composed of seven right-handed α-helices and a twisted β-sheet formed by 10 β-strands, two of which are parallel and eight anti-parallel.

1.1.2. Mechanism of action

HCA II is the mostly studied isoform. Its active site consists of a cone-shaped cavity divided into a hydrophobic (Val121, Val143, Leu198, Val207 and Trp209) and hydrophilic (Tyr7, Asn62, His64, Asn67, Thr199 and Thr200) region. At the base of the cavity lies the Zn^{2+} ion which is tetrahedrally coordinated by three conserved histidine residues (His94, His96 and His119). In the X-ray crystal structure of native CA, the fourth atom coordinating the Zn^{2+} ion is given by the oxygen atom of a catalytic water molecule (or hydroxide ion, Figure 1.1.). Although the catalytic pocket is highly conserved among the different isoforms, the area surrounding the catalytic cone exhibits a certain variability in terms of hydrophobicity and hydrophilicity[1].

According to the literature, CAs catalyze the reversible hydration of carbon dioxide (CO_2) and dehydration of a bicarbonate (HCO_3^-) molecule **(1)** in a two-step reaction[1], [4]. In the first step, a CO_2 molecule enters the catalytic site and is subjected to a nucleophilic

attack by the hydroxyl group coordinating the Zn^{2+} ion (Zn-OH^-) to form the zinc-bound

bicarbonate (Zn-HCO_3^-). Subsequentially, the HCO_3^- ion is displaced by a water molecule

(Zn-H_2O) and diffuses into the bulk solvent **(2)**. In the second step, the proton transfer

from the Zn-H_2O to the bulk solvent regenerates the Zn-OH^- complex **(3)**. More precisely,

the transfer of the proton occurs *via* His64 which is also known as "proton shuttle".

Interestingly, this key residue exists in two conformations, referred to as inward and

outward conformations which are 7.5 Å and 12.0 Å away from the Zn^{2+} ion, respectively. It

has been suggested that His64 acts as a shuttle, accepting the proton *via* a network of

ordered water molecules from the active site and subsequently releasing the proton into the

bulk solvent.

General reaction	$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$	(1)
First step	E-Zn-$OH^- + CO_2 \rightleftharpoons E$-$Zn$-$HCO_3^- \rightleftharpoons E$-$Zn$-$H_2O + HCO_3^-$	(2)
Second step	E-Zn-$H_2O + B \rightleftharpoons E$-$Zn$-$OH^- + BH^+$	(3)
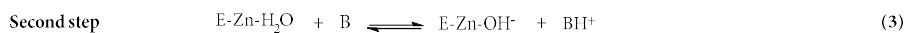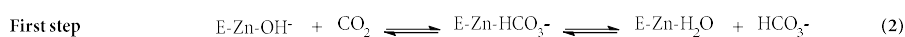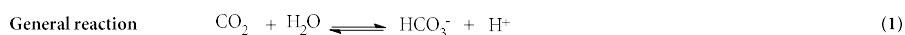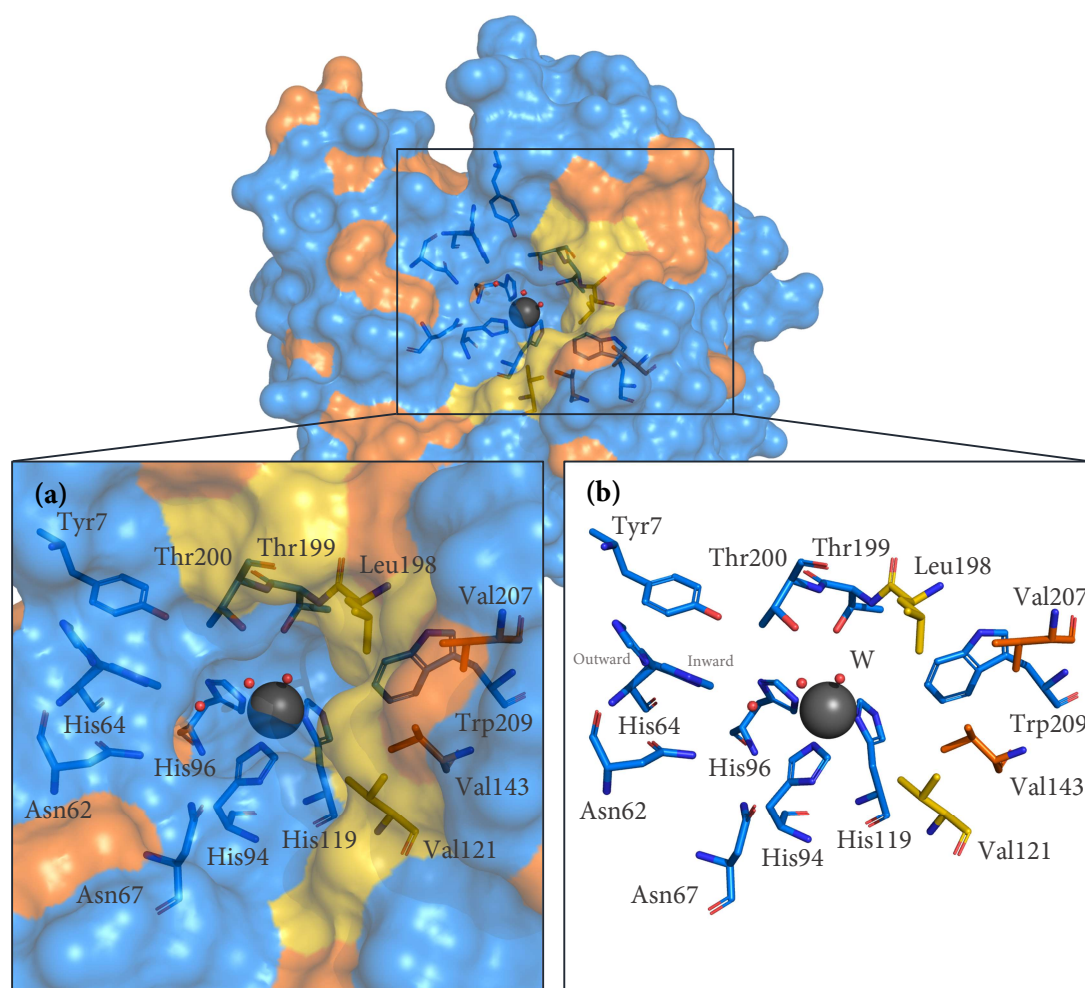
Figure 1.1. Surface representation of *h*CA II. The protein surface is colored in orange for hydrophobic (Val, Ile, Leu, Phe, Met, Ala, Gly, and Pro) and in blue for charged and polar (Arg, Asp, Asn, Glu, Gln, His, Lys, Ser, Thr, Tyr, Trp, and Cys) residues. The HP domain (Ile91, Val121, Phe131, Val135, Leu198, Pro201 and Pro202) is represented in yellow. Close-up view of the active site represented as **(a)** Surface and **(b)** stick model. The catalytic Zn^{2+} ion (gray sphere) is tetrahedrally coordinated by His94, His96, His119 and the catalytic water molecule W (red sphere). Two additional water molecules connect the Zn^{2+} ion to His6, the so-called proton shuttle. His64 adopts an inward or outward configuration.

1.2.   An overview of SPR

SPR is a powerful technique to measure biomolecular interactions in real-time and in a label-free environment. This technique allows for the identification of the binding of two or more interactants to each other, to determine their binding affinity and to measure the association and dissociation rates. One of the interactants, usually the protein under investigation, is immobilized on the sensor surface, which is, according to SPR terminology, referred to as "ligand (B)", as it is attached or better immobilized on the surface chip. The

other interactants correspond to the compounds to screen which are called analytes (A). They are solubilized in the so-called "running buffer", are free in solution and passed over the surface chip. The kinetic constants k_{on} and k_{off} are constants describing the rate of formation or dissociation of the protein-ligand complex (or ligand-analyte complex according to SPR terminology).

$$A + B \xrightleftharpoons[K_{off}]{K_{on}} AB$$

$k_{on} = k_a$ → Association rate constant

$k_{off} = k_b$ → Dissociation rate constant

Equilibrium is reached when $[A]\,[B]\,k_{on} = [AB]\,k_{off}$

Equilibrium dissociation constant $K_d = \dfrac{k_{off}}{k_{on}} = \dfrac{[A]\,[B]}{[AB]}$

Analyte (A) = Compounds
Ligand (B) = Protein immobilised
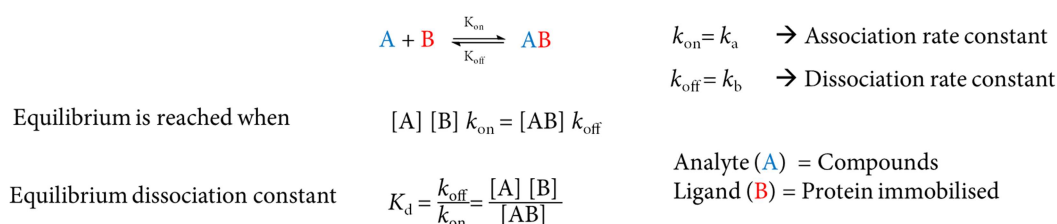
In a typical 1:1 interaction model between analyte A and ligand B, the concentration of complex on the surface [AB] is given in resonance units (RU) by the response above the baseline. The concentration of free ligand on the surface [B] is not measured directly. However, the total ligand concentration can be expressed in RU as the maximum analyte binding capacity R_{max} (Figure 1.2.). The concentration of free ligand can be then calculated by R_{max}-R. The association and dissociation are measured in arbitrary units (RU) and displayed in a graph called sensorgram.

Steady State

$$R_{eq} = \frac{C \cdot R_{max}}{C + K_D}$$

$$R_{max} = R_{ligand} \cdot \left(\frac{Mr_{analyte}}{Mr_{ligand}}\right) V_{ligand}$$

Dissociation

$$\frac{dR}{dt} = -K_D \cdot R$$

Association

$$\frac{d[LA]}{dt} = k_{on} \cdot [L] \cdot [A] - k_{off} \cdot [LA]$$
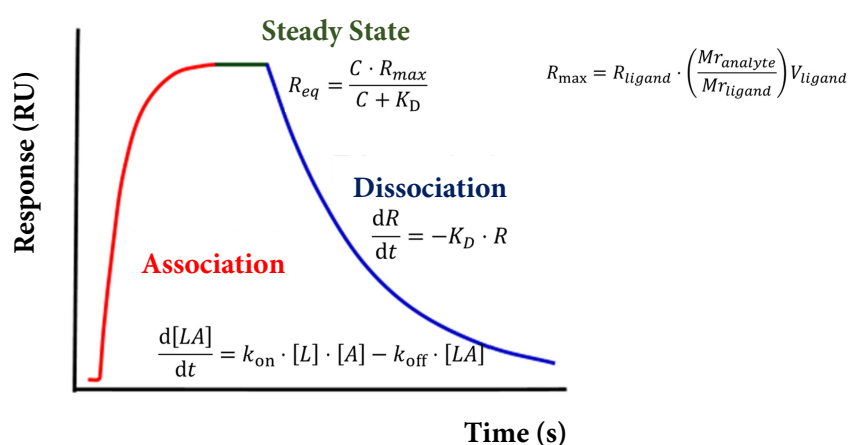
Figure 1.2. Typical SPR sensorgram displaying the association, steady state and dissociation phase.

In particular, the interaction kinetics can be divided in three phases:

(1) Association: the analyte A binds to the protein B (red part of the curve, Figure 1.2.);
(2) Steady state: the amount of A binding to B is equal to the amount of A dissociating from B (green part of the curve, Figure 1.2.);
(3) Dissociation: the analyte A dissociates from B (blue part of the curve, Figure 1.2.).

Each of these three phases contains information about the rate of association and dissociation (k_{on} and k_{off}) and the binding affinity (K_d), according to the equations shown in Figure 1.2.

1.3.   Aim of the project and state of the art

The most representative class of compounds that inhibit CAs are benzenesulfonamides. They are widely studied, potent compounds which inhibit the enzyme by displacing the catalytic water molecule and coordinating the catalytic Zn^{2+} ion.

According to literature, the mechanism of sulfonamide binding to hCA II has not been fully understood and elucidated yet. Based on experimental data on the pH dependence of the association process, two possible mechanism of actions have been suggested: (i) Considering that sulfonamides are usually weak acids, the neutral form may coordinate the Zn^{2+} ion by displacing the hydroxyl ion (the catalytic water molecule has indeed a pK_a: 6.8 and is likely presents as OH^-) or (ii) a deprotonated and in consequence charged sulfonamide displaces the catalytic water molecule bound to the Zn^{2+} ion[5].

In addition, a more complicated and multistep process has also been proposed. In fact, it was shown that the kinetics of binding depend on the nature of the substituents that decorate the benzensulfonamide moiety. In particular, it was observed that the k_{on} of the para-substituted n-alkyl benzenesulfonamides increases with the length of a para-alkyl chain [5]. This discovery was attributed to a pre-binding event between the hydrophobic part of the ligand (alkyl chain) and the so-called "second hydrophobic binding site" at the entrance of the catalytic cone which consists of several hydrophobic residues to form a

solvent-exposed hydrophobic patch (HP) [5]. According to *Gaspari et al.* [5], the largest HP domain was found for the propyl-para-substituted benzenesulfonamide, which includes Ile91, Val121, Phe131, Val135, Leu198, Pro201 and Pro202 (Figure 2.1.). In the same study, the authors proposed, that sulfonamide inhibitors first bind to the HP domain *via* the hydrophobic substituent in para position, causing the inhibitor to adopt the so-called "F (flipped) conformation". Subsequently, the inhibitor would then flip over, adopting the "S (straight) conformation" pointing towards the inner part of the catalytic pocket and orienting the sulfonamide polar head towards the catalytic zinc ion, ready to establish a coordination bond. It is during this conformational transition from F to S conformation that the ligand may change its protonation state. In particular, it was suggested that it enters the catalytic pocket in its neutral form and as soon as the distance between the nitrogen atoms of the sulfonamide's polar head is within 2.7 Å - 2.5 Å of the catalytic Zn^{2+} ion, the inhibitor then turns into its anionic form and finally coordinates the Zn^{2+} ion [5].

In addition, kinetic studies by SPR revealed, that the association rate k_{on} increases with increasing hydrophobic nature of the substituent in the para position of the benzenesulfonamide due to the above discussed pre-binding event. The dissociation rate k_{off}, on the other hand, appeared to be related to the interaction between the ligand and the solvation water molecule [5].

In the present work, four series of para-substituted sulfonamides, synthesized by *Khang Ngo* and *Hans-Dieter Gerber*, were screened by SPR to further investigate the suggested dependency between k_{on} and hydrophobicity upon binding to *h*CA II. Furthermore, the k_{on} and k_{off} measured by SPR were also compared with the ones measured by *Steffen Glöckner*, who established a protocol to obtain kinetic information by *kin*ITC [6]. The association and dissociation rates were deeply investigated by measuring several compounds which differ by a methylene group or the attachment at the central benzene moiety *via* carbon or oxygen. This systematic approach is intended to elucidate, in how far small structural changes can affect kinetic data.
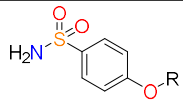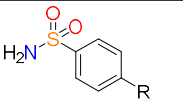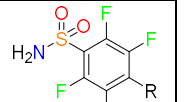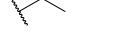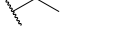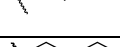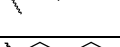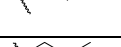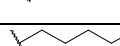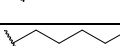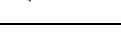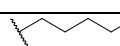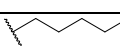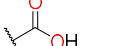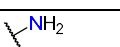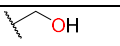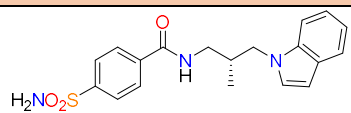
Moreover, one series of compounds was also screened against different hCAs isoforms including CAI, CAII, CAVII, CAIX, CAXII and CAXIII. In fact, the hCAs isoforms differ in their HP domain and hydrophobicity, makes it interesting to investigate and compare, in how far the structural differences among the isoforms affect the k_{on} and k_{off} of the selected series of compounds.

SPR experiments were performed in collaboration with Giulia Opassi (EU-ITN AEGIS, ESR 6) during a three-months secondment at the collaboration site "Faculty of Science and Technology, Deptartment of Chemistry - BMC" in Uppsala (Sweden).

2. Experimental part

As reported in Section 1.3., the aim of the project was to screen by SPR four series of para-substituted benzenesulfonamides to compare kinetic rate constants and binding affinities between different series and against different hCA isoforms. (Figure 2.1.). (i) The first series consisted of seven n-alkyl-ether compounds with $1 \leq n \leq 6$. In addition, the primary alcohol -OH is also included in the series. (ii) The second series consisted of a heterogeneous group of compounds consisting of apolar n-alkyl-substituents with $1 \leq n \leq 7$, as well as polar groups such as amino group, carboxylic acid and ethyl alcohol. (iii) The third series consisted of benzenesulfonamides decorated in ortho and meta positions by four fluorine atoms and carrying in para position various substituents such as n-alkyl chains with $1 \leq n \leq 3$ or an additional fifth fluorine atom. (iv) The fourth series consisted of a single compound substituted in para position with an amide group carrying an alkyl-indole moiety (Table 2.1.).

Table 2.1. Compounds under investigation divided into four series

		I Series		II Series		III Series
Structure						
Names	SOH		S0		S0FH	
	O1		S1		S0F	
	O2		S2		S1F	
	O3		S3		S2F	
	O4		S4		S3F	
	O5		S5			
	O6		S6			
			S7			
			4CBS			
			SULF			
			S2OH			

IV^ Series	
Structure **CAH7**	

All four series of compounds were screened against *h*CAII while the second series was also screened against several *h*CAs isoforms such as CAI, CAVII, CAIX, CAXII and CAXIII (Figure 2.1.).
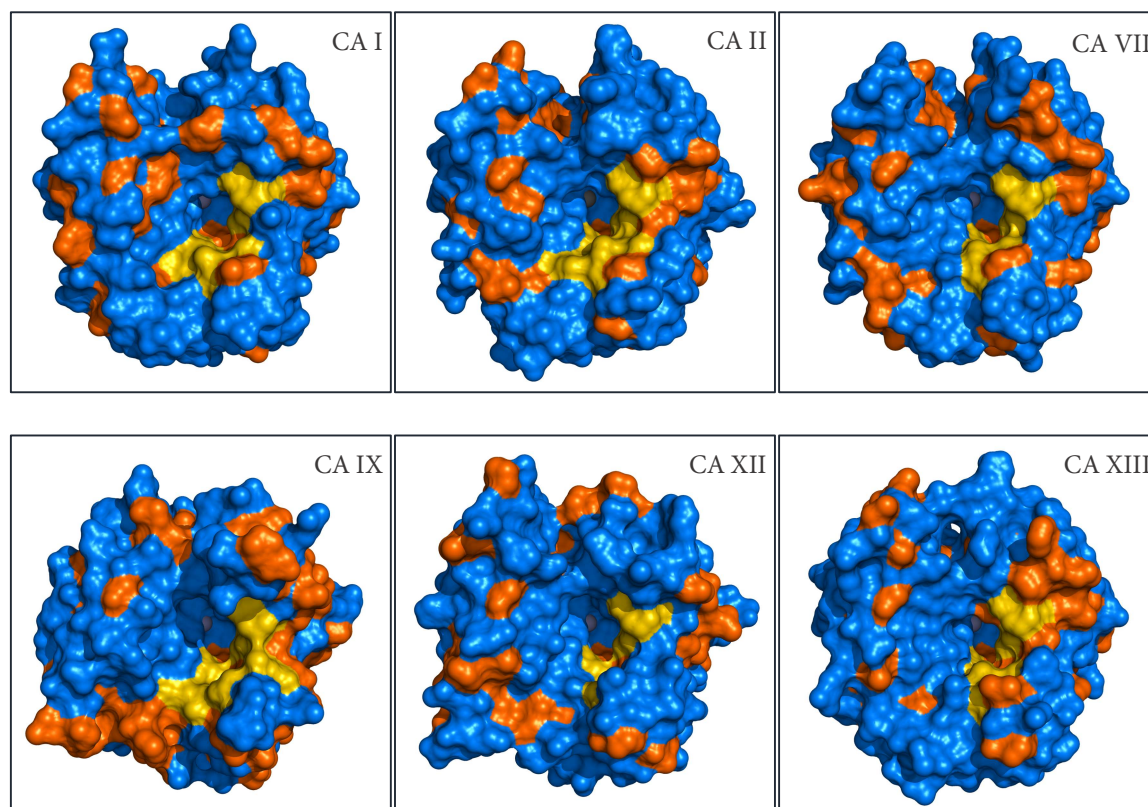
Figure 2.1. Comparison of surface representations of hCA isoforms. The protein surfaces are colored in orange for hydrophobic (Val, Ile, Leu, Phe, Met, Ala, Gly, and Pro) and in blue for charged and polar (Arg, Asp, Asn, Glu, Gln, His, Lys, Ser, Thr, Tyr, Trp, and Cys) residues. The HP domain (Ile91, Val121, Phe131, Val135, Leu198, Pro201 and Pro202 for *h*CA II) is represented in yellow.

2.1   The *h*CAII

Initially, the protocol set by *Gaspari et al.* was reproduced using the Biacore 3000 device to compare the obtained kinetic data. In particular, the compounds **1**, **2**, **3, 4** and **5** from *Gaspari et al.* correspond to compounds **S0**, **S1**, **S2, S3** and **S2OH** in the present work which were used as reference compounds. As shown in Figure 2.2., due to the low resolution of the device, it was not possible to extract any kinetic data. In fact, as the Biacore 3000 has a noise level of 1 RU, no washing step was performed between the measurements of two consecutive compounds and therefore the device is not suitable to perform a small-molecules screening project. In fact, without a washing step, a portion of the first compound measured could still be bonded to the chip surface where the protein is immobilized. As a result, during the next run, there is the risk of measuring the second

compound combined with part of the first one that remained attached on the surface chip. This phenomenon was particularly pronounced during measurements of compounds of the second series, which showed a significant tendency to stick to the sensor chip. This phenomenon is called "carry over", and in such case it is probably due to the hydrophobic nature of the *n*-alkyl chain in para position of the benzenesulfonamides.
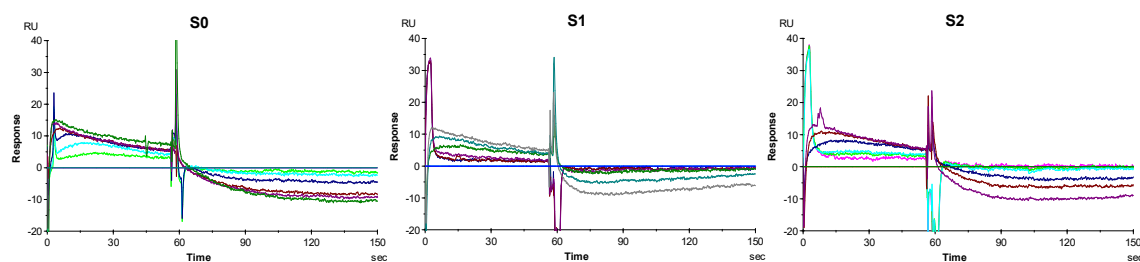
Figure 2.2. Comparison of SPR sensorgrams recorded with the equilibrium analysis method. The shape of the curves did not allow to obtain kinetic information.

Therefore, it was necessary to change strategy. A Biacore T200 device was then used. This instrument is more powerful (the noise level is 0.1 RU) than the Biacore 3000, allowing a washing step between two consecutive measurements using the running buffer containing 50 % DMSO and is therefore suitable for a fragment screening project. However, despite the washing step and the higher resolution, the compounds still showed a tendency to stick to the sensor chip. It was then decided to replace the classic "equilibrium analysis" with the "single cycle kinetics" method.

In a typical "equilibrium analysis" the injection of the analyte needs to be long enough to reach steady state and the analyte is injected stepwise at increased concentrations. After the injection at maximum concentration, the process is reversed, and several lower concentrations are injected in order to demonstrate that the system is reversible. The "single cycle kinetics" is instead useful when the interactions between the ligand and the analyte are difficult to regenerate, as in the case of the above-mentioned compounds of the second series, which hardly detached from the sensor chip. In this method, the analytes are injected stepwise from low to high concentration. In contrast to the "equilibrium analysis", the dissociation time is kept short between the consecutive concentration steps, not letting

the compounds to completely dissociate, while it is long at the end of the analysis to allow the analyte to dissociate. Single cycle kinetic was therefore the method of choice. The carry over phenomenon was minimized and interpretable curves, from which kinetic information could be extracted, were then obtained. To facilitate the comparison and better visualize the trend of the different classes of compounds, k_{on} was plotted against k_{off} in the so-called "Interaction Kinetic Graph (IKG)".

Initially, the protocol was validated by measuring **S0**, **S1**, **S2, S3** and **S2OH** as reference ligands and comparing them with the previously determined value from *Gaspari et al.* As shown in Figure 2.3., the reference compounds (red squares) are quite comparable with the measured ones (blue triangles), except for **S3** and **S2OH** which show differences particularly in the k_{off}. In general, the interpretation of the data was neither straightforward nor conclusive. In fact, several factors seemed to be involved in the association and dissociation processes.
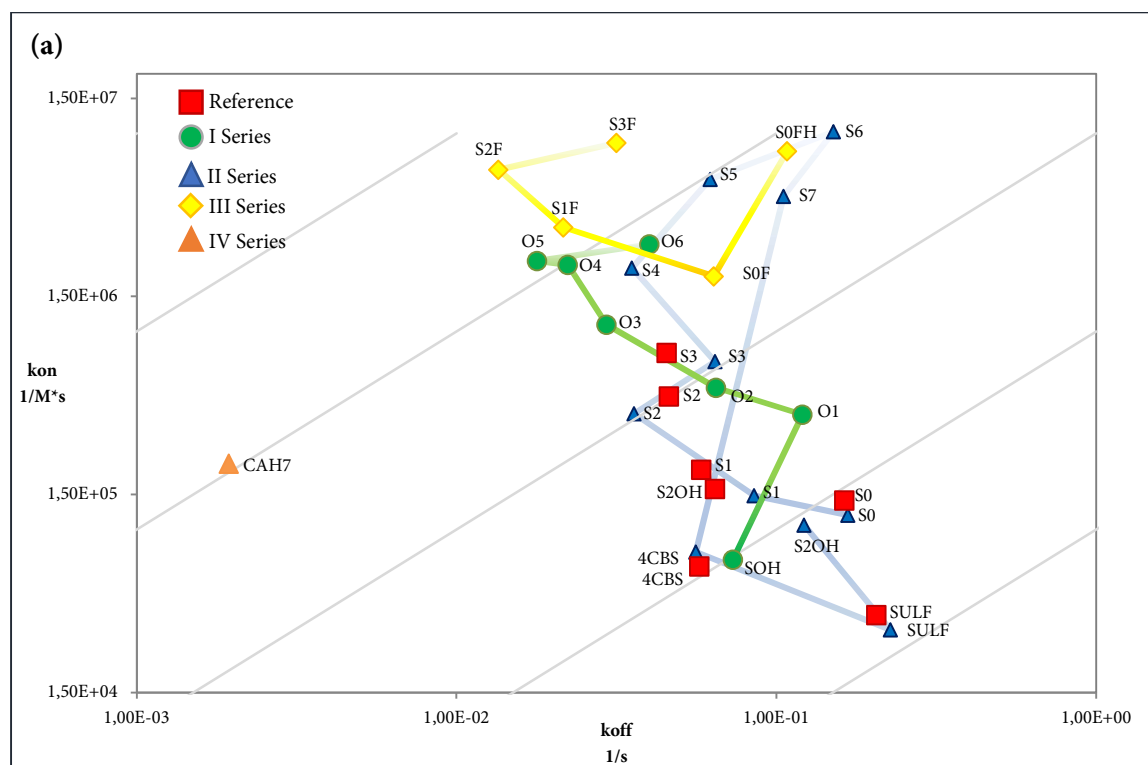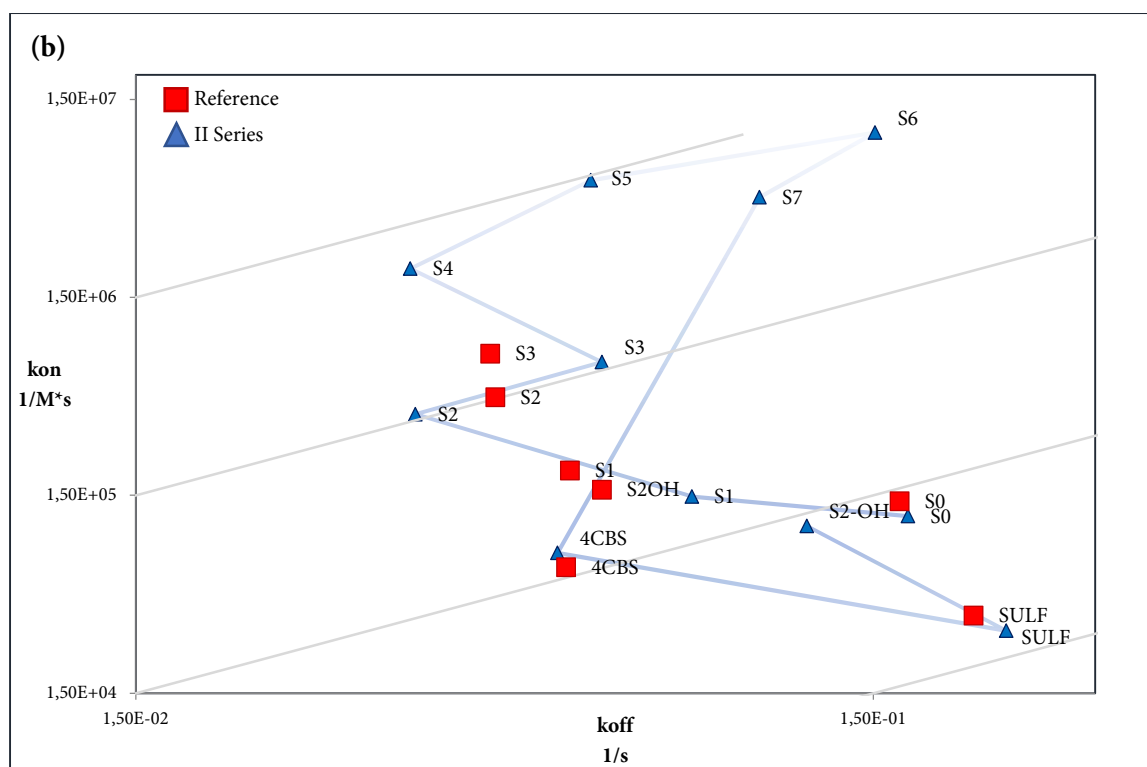
Figure 2.3. (a) Interaction kinetic graph displaying the trends of the four series of compounds. The kinetic parameters k_{on} and k_{off} were obtained by the single cycle kinetic method. **(b)** Zoom-in of interaction kinetic graph displaying the trends of the reference and the second series of compounds for clarity. The kinetic parameters k_{on} and k_{off} were obtained by the single cycle kinetic method.

However, it was possible to detect a certain trend in the first (green) and second (blue) series of compounds.

As displayed in Figure 2.3. **(b)**, in the *n*-alkyl-series (blue) the k_{on} mostly increased with the length of the alkyl chain from **S0** (R=H) to **S6** (R=6*n*). At the same time, the k_{off} decreased progressively, except for **S3** and **S5**, which exhibited faster dissociation rates. This could be due to both, mobility and preferred conformation adopted by the alkyl chain [6]. In fact, crystallographic data indicated that **S5** orients the terminal methyl group toward the inside of the pocket by adopting a gauche conformation, whereas **S2**, **S4** and **S6** are facing and interacting with the hydrophobic surface [6]. The terminal methyl group of **S3**, on the other hand, protrudes out of the catalytic pocket, not pointing towards the protein surface and neither establishing strong hydrophobic interactions with the surrounding residues [6]. In the absence of any hydrophobic interaction with the protein surface, it is reasonable to

assume that **S3** and **S5** may therefore dissociate faster. SPR data seem to indicate a maximum limit in the length of the alkyl chain able to interact with the hydrophobic protein surface. In fact, **S7** showed a slower association and dissociation rate than **S6**. The reason could be due to the pre-binding event where the terminal methyl group does not interact with the hydrophobic surface while, once bound to the catalytic site, it could be involved in additional hydrophobic interactions that would slow down its dissociation rate.

In the ether's series (I Series, Table 2.1. and green circles in Figure 2.3. **(a)** and 2.4. **(a)**), it is possible to see a certain trend in the association and dissociation rate. In fact, as in the previously discussed alkyl series, the k_{on} mainly increases as the chain length increases while the k_{off} decreases on a regular basis. Also in this case, there seems to be a maximum limit concerning the length of the ethereal chain that may influences the association and dissociation rate. In fact, k_{on} increases and k_{off} decreases gradually from **O1** to **O5** whereas **O6** shows an opposite trend, comparable to **S5** in the previous case.
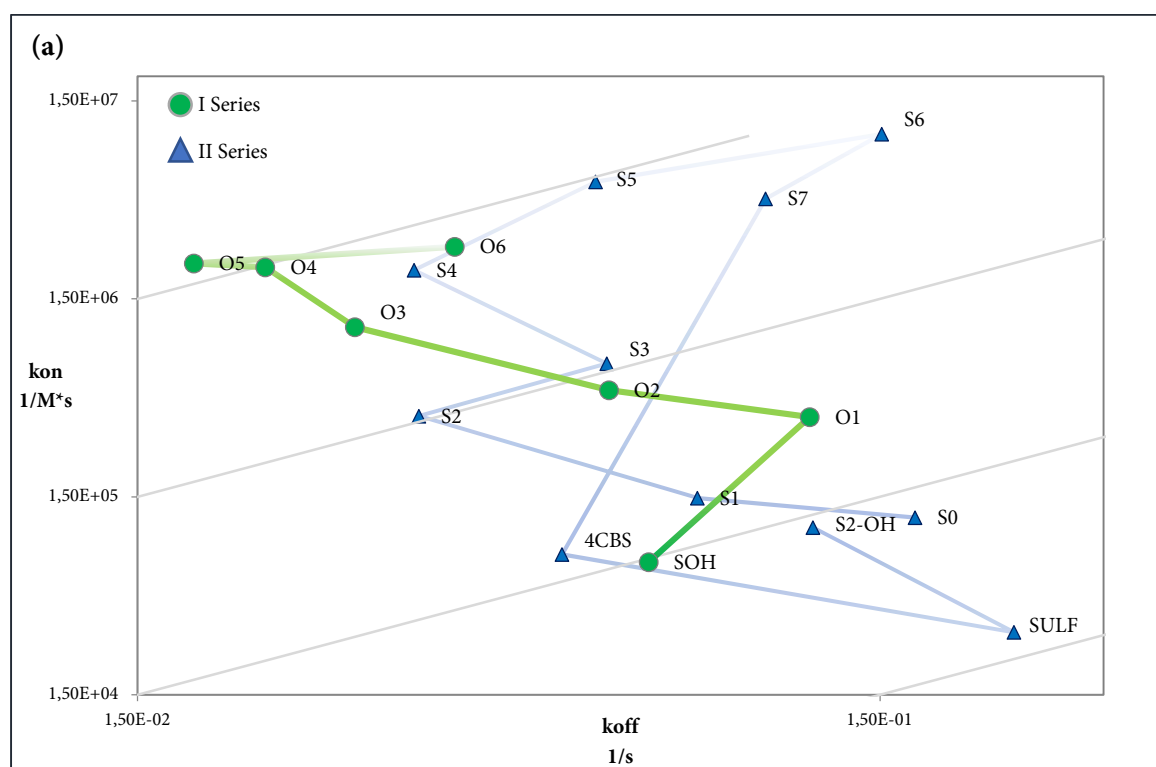
(continued on next page)
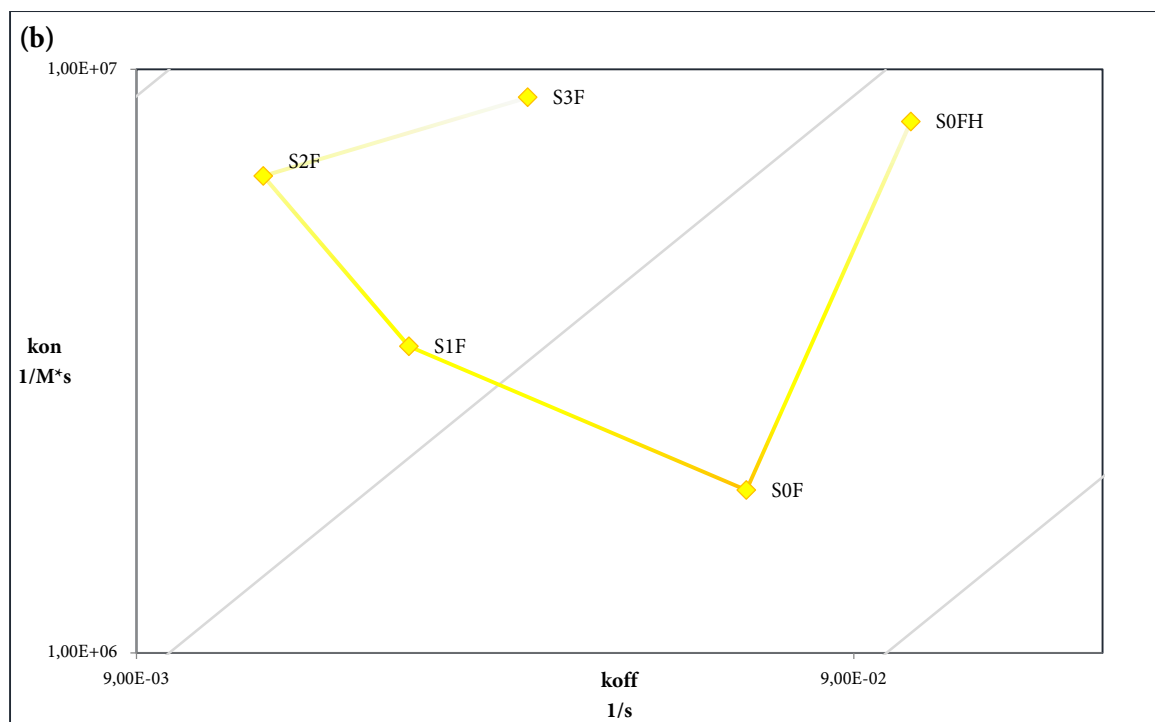
Figure 2.4. Zoom-in of interaction kinetic graph reported in Figure 2.3. displaying the trends of **(a)** the first (green), second (blue) and **(b)** third (yellow) series of compounds for clarity. The kinetic parameters k_{on} and k_{off} were obtained by the single cycle kinetic method.

It is interesting to note that the fluorinated compounds of the third series (III Series, Table 2.1. and yellow rhombus in Figure 2.3. **(a)** and 2.4. **(b)**) exhibit higher affinity against CAII (Appendix 5.1.).and the fastest association across all the measured series of compounds, possibly due to the more hydrophobic nature of the fluorination or a shift in pKa values. As in the aforementioned cases, k_{on} increased and k_{off} decreased gradually from **S0F** to **S2F**, whereas **S3F** shows an opposite trend.

CAH7, belonging to the fourth series (IV Series, Table 2.1. and orange triangle in Figure 2.3. **(a)**), showed an out-of-range behavior compared to the previous series. In fact, the k_{off} of the above discussed compounds differed only by one order of magnitude while k_{on} covered a wider range from 10^4 to 10^7 s^{-1}. In this case, k_{on} falls into the expected range while k_{off} showed the lowest value among all series (Figure 2.5.). This extremely slow dissociation could be due to the interaction of the indole ring with an additional other hydrophobic patch. The indole moiety is therefore also called "secondary recognition elements".

Moreover, it is important to note that the model which better fitted the sensorgram data was that of a "two state reaction". Unlike the classic 1:1 binding (Section 1.2.), the two-state retraction model assumes a conformational change that forms and stabilizes the complex AB*. The dissociation of AB* in A+B can only occur by passing through the intermediate state AB. Since it is not possible to detect conformational changes with SPR, to decide wheater experimental data are better fitted with a two states reaction model it means that a conformational change may take place and should be further investigated with other techniques. Interestingly, the compound shows also an unusual ITC titration thermogram indicating an unusual binding event possibly overlaid by a slow conformational transition of the formed complex.

$$A + B \; \underset{k_{off1}}{\overset{k_{on1}}{\rightleftharpoons}} \; AB \; \underset{k_{off2}}{\overset{k_{on2}}{\rightleftharpoons}} \; AB^{\star}$$
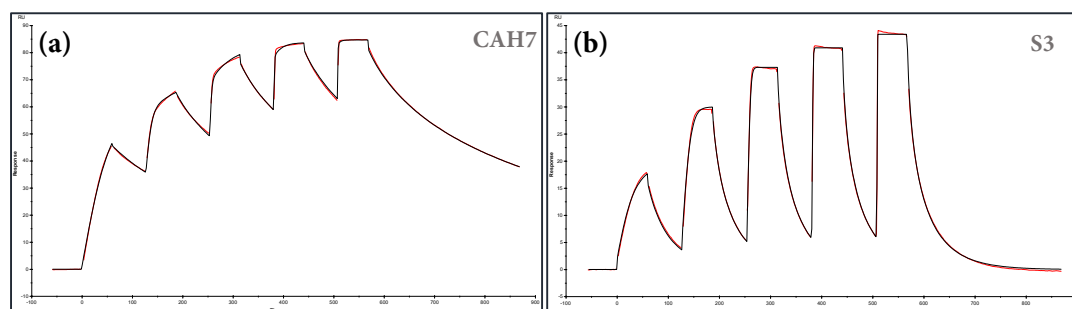
Figure 2.5. Comparison between the sensorgrams of **(a)** CAH7 and **(b)** S3 displaying the extremely slow dissociation of the former. At the end of the last injection **S3** dissociates completely returning to the baseline, while **CAH7** apparently does not dissociate and shows a nearly irreversible behavior.

2.2   The hCA isoforms

The n-alkyl series (II series in Table 2.1.) was also screened against several hCA isoforms such as CAI, CAVII, CAIX, CAXII and CAXIII (Figure 2.2.). For all six isoforms under investigation, the binding affinity was obtained (Appendix 5). In addition, for the hCAVII and hCAXII kinetic data could be extracted from the sensorgrams.

Comparing hCAII with hCAVII and hCAXII in terms of surface lipophilicity/hydrophilicity (Figure 2.6.), the first two isoforms look very similar while the

latter is more hydrophilic. In particular, comparing the residues that form the HP domain two hydrophobic residues (Ile9 and Val135) were repleced with two polar residues (Thr88 and Ser133) (Table 2.2.). As a result, k_{on} and k_{off} showed a different trend than the one described for hCAII (Section 2.1.).
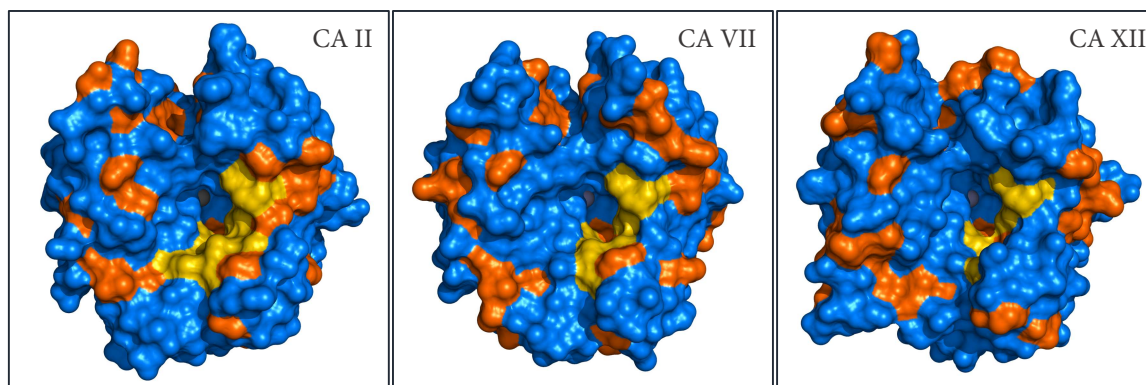
Figure 2.6. Comparison of surface representation of hCAs II, VII and XII isoforms. The protein surfaces are colored in orange for hydrophobic (Val, Ile, Leu, Phe, Met, Ala, Gly, and Pro) and in blue for charged and polar (Arg, Asp, Asn, Glu, Gln, His, Lys, Ser, Thr, Tyr, Trp, and Cys) residues. The HP domain (Ile91, Val121, Phe131, Val135, Leu198, Pro201 and Pro202 for hCA II) is represented in yellow.

Table 2.2. Comparison of residues that form the HP domain among hCAII, hCAVII and hCAXII

h**CAII**	Ile91	Val121	Phe131	Val135	Leu198	Pro201	Pro202
h**CAVII**	Leu92	Val123	Phe133	Ala137	Leu200	Pro203	Pro204
h**CAXIII**	**Thr88***	Val119	Ala129	**Ser133***	Leu197	Pro200	Pro201

*: Polar residues present only in hCAXII

Firstly, analyzing the interaction kinetic graph reported in Figure 2.7., k_{on} and k_{off} covered a smaller range than for hCAII and they are shifted toward slower on and slower off-rates compared to hCAII. A closer look showed that for hCAXII the association rate increased from S1 to S6 in a similar manner as previously discussed for hCAII. However, in contrast with this latter one, the dissociation rate became faster, probably due to the higher surface hydrophilicity.

Although very similar to the hCAII, it was not possible to explain the kinetic data of hCAVII, which did not show a well-defined trend in terms of both k_{on} and k_{off}.
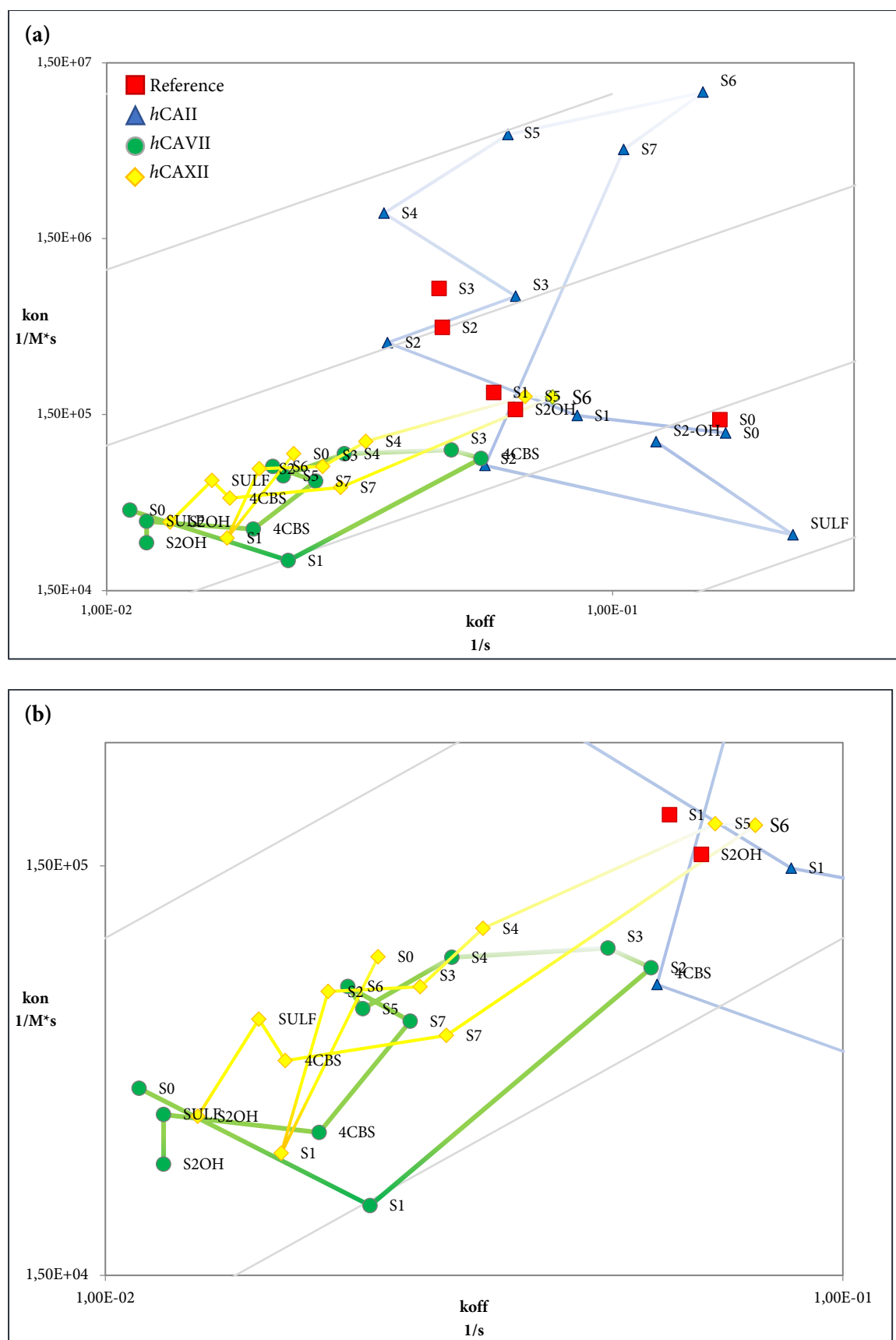
Figure 2.7. (a) Interaction kinetic graph displaying the trends of the *n*-alkyl-series against CAII (blue), CAVII (green) and CAXII (yellow). **(b)** Zoom-in of **(a)**. The kinetic parameters k_{on} and k_{off} were obtained by the single cycle kinetic method.

3. Conclusion and Outlook

In the present work, four series of para-substituted benzenesulfonamides were screened by SPR, obtaining binding affinity and kinetic data for all compounds under investigation on hCAII. Compounds previously investigated by *Gaspari et al.* were used as reference compounds to validate the method. In particular, the traditional "equilibrium analysis" was replaced by the single cycle kinetic method due to the high carry-over phenomenon. This was probably caused by the extensive hydrophobicity of the alkyl chains which stick to the sensor chip surface and lead to sensorgrams difficult to interpret. As already reported by *Gaspari et al.*, the association rate k_{on} increases with the ligand's hydrophobicity, becoming faster as the n-alkyl chain length increases.

In particular, it was noticed that k_{on} reaches a maximum value depending on the length of the alkyl chain and afterward it decreases or remains almost unchanged. In the first series, for example, the k_{on} increases on a regular basis from **O1** to **O4** while **O5** and **O6** show similar values. Also, the second series showed a similar trend. The association rate regularly increased from **S1** to **S6** while it unexpectedly decreases for **S7**. Moreover, the compounds carrying a hydroxyl (**S2OH**), amine (**SULF**), carboxylic groups (**4CBS**) or even a sole hydrogen atom showed a slower association rate among the four series, confirming thereby that hydrophobicity at the terminal end opposite of the sulfonamide group plays an essential role in the gradation of k_{on}.

Concerning the dissociation rate k_{off}, it usually decreases with the chain length with the exception of **S3** and **S5**, probably because of non-optimal interaction with the hydrophobic protein surface.

CAH7 showed an out of range behaviour, displaying the slowest dissociation rate across all series of compounds. This is probably due to the interaction of the indole ring with an additional hydrophobic surface patch. In addition, it seems that a conformational change is accompanying ligand binding.

Moreover, one series of compounds was also screened against different hCAs isoforms including CAI, CAVII, CAIX, CAXII and CAXIII. For all isoforms under investigation, the binding affinity was obtained and for two of them, hCAVII and hCAXII, also kinetic data. Interestingly, both the association rate of hCAXII increases in a similar manner as previously discussed for hCAII while the dissociation rate showed an opposite trend, thus becoming faster. This is probably due to the replacement of two hydrophobic with two polar residues.

In conclusion, the interpretation of the kinetic data is rather complex, and it seems that several features contribute to the trends in k_{on} and k_{off} such as hydrophobicity and pKa values of the compounds, but also chain length important for the adopted preferential geometry of the ligand upon bonding and during establishing first surface contacts between the protein and the hydrophobic or hydrophilic portion of the ligand.

4. Materials and method

4.1   Proteins and ligands

Carbonic anhydrases I, II, VII, IX, XII and XIII were kindly provided by the Department of Biothermodynamics and Drug Design, Institute of Biotechnology, Vilnius University (Lithuania).

The benzenesulfonamide inhibitors were kindly synthesized by *Khang Ngo* and *Hans-Dieter Gerber*.

4.2   SPR

The experiments were performed on a Biacore T200 surface plasmon resonance biosensor instrument (GE Healthcare, Uppsala, Sweden) at the Department of Chemistry - BMC, Uppsala University (Sweden).

For immobilization, the amine coupling method was used. The protein samples (10 mM sodium acetate pH: 5.0) was immobilized on the sensor chip series S Sensor Chip CM5. The carboxymethyl dextran matrix of the sensor chip was activated with 0.1 M N-hydroxysuccinimide (NHS) and 0.4 M 1-ethyl-3-(3-(dimethylamino) propyl) carbodiimidehydrochloride (EDC) at a flow rate of 10 µL/min for 10 minutes. Subsequently, activated groups of the dextran matrix which did not bind the protein of interest were deactivated injecting 1 M ethanolamine hydrochloride (pH 8.5) for 10 min. This protocol lead to an immobilization level of about 5000 RU.
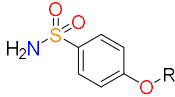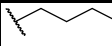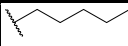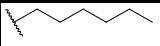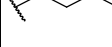
The immobilization was performed at 75 µg/mL for *h*CAI, 100 µg/mL for *h*CAII, 25 µg/mL for hCAVII and *h*CAXII, 200 µg/mL for hCAIX and *h*CAXIII.

The compounds were measured using the single cycle kinetic method. PBST (10 mM phosphate, 140 mM NaCl, 2.7 mM KCl, pH 7.4, 0.05 % Tween® 20) with 2% DMSO was

used as running buffer. Compounds were stored in 10 mM DMSO stock solutions and measured at five different concentrations (5 μM-20 nM) in PBST buffer. The compounds were injected at a flow rate of 50 μL/min for 120 s. Then, the dissociation was monitored for 300 s. A washing step cycle using the running buffer with 50 % DMSO was performed between consecutive measurement. The data were evaluated with BIAevaluation v.3.0 software (GE Healthcare, Uppsala, Sweden). The kinetic parameters k_{on} and k_{off} were obtained using both a 1:1 interaction model and two state reaction.

5. Appendix
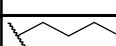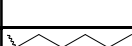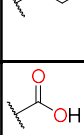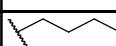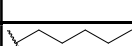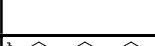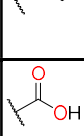
5.1 Kinetic and affinity values of hCAII

		Compound	pH	k_{on} (M^{-1}s^{-1})	k_{off} (s^{-1})	K_D (nM)
I Series		SOH	7.4	$7,04 \cdot 10^4$	$7,32 \cdot 10^{-2}$	$1,04 \cdot 10^{-6}$
		O1	7.4	$3,80 \cdot 10^5$	$1,21 \cdot 10^{-1}$	$3,18 \cdot 10^{-7}$
		O2	7.4	$5,20 \cdot 10^5$	$6,47 \cdot 10^{-2}$	$1,25 \cdot 10^{-7}$
		O3	7.4	$1,08 \cdot 10^6$	$2,94 \cdot 10^{-2}$	$2,72 \cdot 10^{-8}$
		O4	7.4	$2,17 \cdot 10^6$	$2,23 \cdot 10^{-2}$	$1,03 \cdot 10^{-8}$
		O5	7.4	$2,27 \cdot 10^6$	$1,79 \cdot 10^{-2}$	$7,89 \cdot 10^{-9}$
		O6	7.4	$2,75 \cdot 10^6$	$4,01 \cdot 10^{-2}$	$1,46 \cdot 10^{-8}$
II Series		S0	7.4	$1,18 \cdot 10^5$	$1,67 \cdot 10^{-1}$	$1,42 \cdot 10^{-6}$
		S1	7.4	$1,48 \cdot 10^5$	$8,52 \cdot 10^{-2}$	$5,75 \cdot 10^{-7}$
		S2	7.4	$3,85 \cdot 10^5$	$3,59 \cdot 10^{-2}$	$9,33 \cdot 10^{-8}$
		S3	7.4	$7,06 \cdot 10^5$	$6,43 \cdot 10^{-2}$	$9,11 \cdot 10^{-8}$
		S4	7.4	$2,09 \cdot 10^6$	$3,54 \cdot 10^{-2}$	$1,69 \cdot 10^{-8}$
		S5	7.4	$5,86 \cdot 10^6$	$6,21 \cdot 10^{-2}$	$1,06 \cdot 10^{-8}$
		S6	7.4	$1,02 \cdot 10^7$	$1,51 \cdot 10^{-1}$	$1,48 \cdot 10^{-8}$
		S7	7.4	$4,81 \cdot 10^6$	$1,05 \cdot 10^{-1}$	$2,19 \cdot 10^{-8}$
		4CBS	7.4	$7,70 \cdot 10^4$	$5,60 \cdot 10^{-2}$	$7,27 \cdot 10^{-7}$
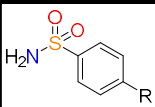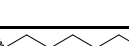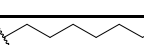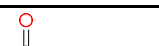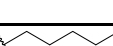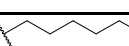
	R group	Compound	pH	k_{on} (M^{-1}s^{-1})	k_{off} (s^{-1})	K_D
	(NH$_2$)	SULF	7.4	$3,11 \cdot 10^4$	$2,27 \cdot 10^{-1}$	$7,31 \cdot 10^{-6}$
	(OH)	S2OH	7.4	$1,05 \cdot 10^5$	$1,22 \cdot 10^{-1}$	$1,16 \cdot 10^{-6}$
II Series_reference	(H)	S0	7.4	$1,40 \cdot 10^5$	$1,63 \cdot 10^{-1}$	$1,15 \cdot 10^{-6}$
		S1	7.4	$2,00 \cdot 10^5$	$5,82 \cdot 10^{-2}$	$2,93 \cdot 10^{-7}$
		S2	7.4	$4,70 \cdot 10^5$	$4,61 \cdot 10^{-2}$	$9,83 \cdot 10^{-8}$
		S3	7.4	$7,80 \cdot 10^5$	$4,54 \cdot 10^{-2}$	$5,78 \cdot 10^{-8}$
	(OH)	S2OH	7.5	$1,60E \cdot 10^5$	$6,43 \cdot 10^{-2}$	$3,92 \cdot 10^{-7}$
III Series	(H)	S0FH	7.4	$8,13 \cdot 10^6$	$1,08 \cdot 10^{-1}$	$1,33 \cdot 10^{-8}$
	(F)	S0F	7.4	$1,90 \cdot 10^6$	$6,37 \cdot 10^{-2}$	$3,35 \cdot 10^{-8}$
		S1F	7.4	$3,35 \cdot 10^6$	$2,16 \cdot 10^{-2}$	$6,45 \cdot 10^{-9}$
		S2F	7.4	$6,56 \cdot 10^6$	$1,35 \cdot 10^{-2}$	$1,28 \cdot 10^{-9}$
		S3F	7.4	$8,95 \cdot 10^6$	$3,16 \cdot 10^{-2}$	$1,53 \cdot 10^{-9}$
IV Series		CAH7	7.4	$2,15 \cdot 10^5$	$1,94E \cdot 10^{-3}$	$9,19 \cdot 10^{-9}$

5.2 Kinetic and affinity values of hCAVII and XII

	R group	Compound	pH	k_{on} (M^{-1}s^{-1})	k_{off} (s^{-1})	K_D (nM)
CA VII	(H)	S0	7.4	$4,29 \cdot 10^4$	$1.112 \cdot 10^{-2}$	$2.59 \cdot 10^{-7}$
		S1	7.4	$2,22 \cdot 10^4$	$2.285 \cdot 10^{-2}$	$1.03 \cdot 10^{-6}$
		S2	7.4	$8,45 \cdot 10^4$	$5.497 \cdot 10^{-2}$	$6.50 \cdot 10^{-7}$

Structure	Name	pH			
(propyl chain)	S3	7.4	$9{,}45\cdot10^4$	$4.802\cdot10^{-2}$	$5.08\cdot10^{-7}$
(butyl chain)	S4	7.4	$8{,}97\cdot10^4$	$2.951\cdot10^{-2}$	$3.29\cdot10^{-7}$
(pentyl chain)	S5	7.4	$6{,}72\cdot10^4$	$2.235\cdot10^{-2}$	$3.32\cdot10^{-7}$
(hexyl chain)	S6	7.4	$7{,}61\cdot10^4$	$2.132\cdot10^{-2}$	$2.80\cdot10^{-7}$
(heptyl chain)	S7	7.4	$6{,}27\cdot10^4$	$2.591\cdot10^{-2}$	$4.14\cdot10^{-7}$
(carboxylic acid)	4CBS	7.4	$3{,}35\cdot10^4$	$1.95\cdot10^{-2}$	$5.82\cdot10^{-7}$
(NH_2)	SULF	7.4	$3{,}71\cdot10^4$	$1.2\cdot10^{-2}$	$3.24\cdot10^{-7}$
(OH)	S2OH	7.4	$2{,}80\cdot10^4$	$1.2\cdot10^{-2}$	$4.28\cdot10^{-7}$
CA XII (H)	S0	7.4	$8{,}99\cdot10^4$	$2.344\cdot10^{-2}$	$2.61\cdot10^{-7}$
(methyl)	S1	7.4	$2{,}98\cdot10^4$	$1.732\cdot10^{-2}$	$5.81\cdot10^{-7}$
(ethyl)	S2	7.4	$7{,}40\cdot10^4$	$2.005\cdot10^{-2}$	$6.11\cdot10^{-7}$
(propyl)	S3	7.4	$7{,}60\cdot10^4$	$2.673\cdot10^{-2}$	$3.52\cdot10^{-7}$
(butyl)	S4	7.4	$1{,}06\cdot10^5$	$3.252\cdot10^{-2}$	$3.08\cdot10^{-7}$
(pentyl)	S5	7.4	$1{,}90\cdot10^5$	$6.715\cdot10^{-2}$	$3.53\cdot10^{-7}$
(hexyl)	S6	7.4	$1{,}88\cdot10^5$	$7.611\cdot10^{-2}$	$4.04\cdot10^{-7}$
(heptyl)	S7	7.4	$5{,}78\cdot10^4$	$2.901\cdot10^{-2}$	$5.02\cdot10^{-7}$
(carboxylic acid)	4CBS	7.4	$5{,}02\cdot10^4$	$1.754\cdot10^{-2}$	$3.49\cdot10^{-7}$
(NH_2)	SULF	7.4	$6{,}33\cdot10^4$	$1.615\cdot10^{-2}$	$2.55\cdot10^{-7}$
(OH)	S2OH	7.4	$3{,}67\cdot10^4$	$1.335\cdot10^{-2}$	$3.64\cdot10^{-7}$

5.3 Affinity values of hCAI, IX and XIII

	R group	Compound	pH	K_D (nM)
CA I	—H	S0	7.4	$9.34 \cdot 10^{-8}$
	(methyl)	S1	7.4	$1.09 \cdot 10^{-7}$
	(ethyl)	S2	7.4	$1.55 \cdot 10^{-8}$
	(propyl)	S3	7.4	$1.12 \cdot 10^{-7}$
	(butyl)	S4	7.4	$3.01 \cdot 10^{-7}$
	(pentyl)	S5	7.4	$4.83 \cdot 10^{-5}$
	(hexyl)	S6	7.4	n.d.
	(heptyl)	S7	7.4	$9.34 \cdot 10^{-8}$
	—CH$_2$COOH	4CBS	7.4	$2.03 \cdot 10^{-5}$
	—CH$_2$NH$_2$	SULF	7.4	$3.73 \cdot 10^{-6}$
	—CH$_2$OH	S2OH	7.4	$6.06 \cdot 10^{-6}$
CA IX	—H	S0	7.4	$2.08 \cdot 10^{-6}$
	(methyl)	S1	7.4	$1.71 \cdot 10^{-6}$
	(ethyl)	S2	7.4	$1.54 \cdot 10^{-7}$
	(propyl)	S3	7.4	$8.88 \cdot 10^{-8}$
	(butyl)	S4	7.4	$1.02 \cdot 10^{-7}$
	(pentyl)	S5	7.4	$3.62 \cdot 10^{-6}$
	(hexyl)	S6	7.4	$3.61 \cdot 10^{-7}$
	(heptyl)	S7	7.4	$9.22 \cdot 10^{-7}$

	![structure]	4CBS	7.4	$2.04 \cdot 10^{-6}$
	![structure]	SULF	7.4	$2.02 \cdot 10^{-5}$
	![structure]	S2OH	7.4	$2.94 \cdot 10^{-6}$
CA XIII	![structure]	S0	7.4	$1.84 \cdot 10^{-6}$
	![structure]	S1	7.4	$8.46 \cdot 10^{-6}$
	![structure]	S2	7.4	$3.16 \cdot 10^{-6}$
	![structure]	S3	7.4	$5.17 \cdot 10^{-7}$
	![structure]	S4	7.4	$1.35 \cdot 10^{-6}$
	![structure]	S5	7.4	$1.38 \cdot 10^{-5}$
	![structure]	S6	7.4	$7.29 \cdot 10^{-5}$
	![structure]	S7	7.4	$2.36 \cdot 10^{-6}$
	![structure]	4CBS	7.4	$2.12 \cdot 10^{-5}$
	![structure]	SULF	7.4	n.d.
	![structure]	S2OH	7.4	$7.70 \cdot 10^{-6}$

6. References

[1]   M. Aggarwal *et al.*, "Structural annotation of human carbonic anhydrases," *J. Enzyme Inhib. Med. Chem.*, vol. 28, no. 2, pp. 267–277, 2012.

[2]   D. Ü. Payaz *et al.*, "Synthesis carbonic anhydrase enzyme inhibition and antioxidant activity of novel benzothiazole derivatives incorporating glycine, methionine, alanine, and phenylalanine moieties," *J. Enzyme Inhib. Med. Chem.*, vol. 34, no. 1, pp. 343–349, 2019.

[3]   A. Di Fiore *et al.*, "Protective role of carbonic anhydrases III and VII in cellular defense mechanisms upon redox unbalance," *Oxid. Med. Cell. Longev.*, vol. 2018, 2018.

[4]   V. M. Krishnamurthyjee *et al.*, "Carbonic Anhydrase as a Model for Biophysical and Physical- Organic Studies of Proteins and Protein–Ligand Binding," *Chem Rev*., vol. 108, pp. 946–1051, 2008.

[5]   R. Gaspari *et al.*, "Kinetic and Structural Insights into the Mechanism of Binding of Sulfonamides to Human Carbonic Anhydrase by Computational and Experimental Studies," *J. Med. Chem.*, vol. 59, no. 9, pp. 4245–4256, 2016.

[6]   S. Glöckner *et al.*, "Conformational changes in alkyl chains determine the thermodynamic and kinetic binding profiles of Carbonic Anhydrase inhibitors and reveal effects of improved shape complementarity," *ACS, Chem. Biol.*, submitted.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my mentor and supervisor **Prof. Gerhard Klebe** for the continuous support during my Ph.D study, for his patience, motivation, enthusiasm and immense knowledge. His guidance, both scientific and personal, helped me through all these years away from home. This work would not have been possible without his supervision.

I am very thankful to **Prof. Andreas Heine** for introducing me to macromolecular X-ray crystallography, for sharing his vast knowledge with me and for the great amount of time he invested in me. Thanks for the many insightful discussions and for always having an open door for questions and valuable suggestion.

I would like to thank **Prof. Peter Kolb** for taking the time for me, for helping me to believe more in myself and for his support and countless valuable suggestions.

I would like to thank **Dr. Wolfgang Jahnke** from the Novartis Institutes for Biomedical Research (Basel) for being my industrial supervisor, reviewing my thesis and taking the time to come (for the third time) to Marburg.

I am very grateful to **Prof. Moritz Bünemann** and **Prof. Frank Runkel** for taking their time to evaluate my PhD thesis and be part of my PhD thesis committee.

I want to thank my friends and colleagues from the **AG Klebe** and **AG Kolb** for contributing to the nice working environment. In particular, I would like to thank all my former (**Dr. Stefan Krimmer**, **Dr. Jonathan Cramer** and **Dr. Christof Siefker**) and current office mates. I thank **Dr. Marko Kljajic** and **Jud Mohammed Badran** for the friendly and happy atmosphere in the 03A15 as well as for the very nice time together. Thanks to **Dr. Stefan Merkl** to always believe in me and for our various chats in so many fields. Many thanks to my travel companion and Mediterranean sister **Dr. Engi Hassaan** for all the good time we shared and spent together both at work and in our spare time. In addition to the

the University of Groningen (the Netherlands.). In addition, I would like to thank **Dr. Sarah Cianferani** and **Dr. Oscar Hernandez-Alba** from the Laboratory of Bioorganic Mass spectrometry (LSMBO, Strasbourg, France) and **Prof. Dr. Markus Wahl** from the Freie Universität Berlin (Germany).

I am very grateful to the MX-team at BESSY II (Helmholtz-Zentrum Berlin, Germany) for their enthusiasm, advices during data collection and sharing their valuable knowledge with me. I would particularly like to acknowledge the help and support of **Dr. Manfred Weiss**, **Dr. Christian Feiler**, **Dr. Martin Gerlach** and **Dr. Jan Wollenhaupt**. Every time I went to BESSY, I always learned something new from each of them.

I would like to thank the Synchrotron radiation facilities ELETTRA (Basovizza, Italy) and DESY/EMBL (Hamburg, Germany) for their help and support during data collection.

Thanks to my dear friends **Jessica Desogus** and **Yodi** for the memorable time we spent together, their support and their contagiously good humor.

I would like to thank my best friends **Michele** and **Marta** for always being with me despite the distance, for their loyalty and sincere friendship. I want to thank them for being part of my life.

A special thank goes to **Steffen Glöckner** for always being there for me, to protect me in so many ways, to encourage me in hard times but also to celebrate our successes together. I want to thank him for his loyalty, his support, his immense love and being part of my life.

Finally, a heartfelt thanks to my **family** for their constant encouragement and support during my studies, for giving me so much strength and for teaching me never to give up. I thank them for always being at my side to celebrate the good times but also to comfort me during the most difficult moments.

Erklärung

Ich versichere, dass ich meine Dissertation

„The Power of Fragments: FBLD approach to investigate protein structures"

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den..........................

....

(Francesca Magari)