Kent Academic Repository

Dodina, Maria (2020) *Biophysical and biochemical characterizations of parasitic and human PDEs as a potential drug targets.* **Doctor of Philosophy (PhD) thesis, University of Kent,.**

Downloaded from

The version of record is available from

This document version
UNSPECIFIED

DOI for this version

Licence for this version
UNSPECIFIED

Additional information

Versions of research works

Biophysical and Biochemical Characterizations of Parasitic and Human PDEs as a Potential Drug Targets

Student: Maria Dodina

PhD Supervisor: David Brown

University of Kent
School of Biosciences
April 2020

Contents

Acknowledgment and Dedication

I would like to dedicate this thesis to my family who supported me through all my studies in UK and motivated me to achieve my goals.

I would also like to thank my friends Kristina and Maxim for believing in me and cheering me up when I needed support.

I would like to acknowledge all lab members at the University of Kent for sharing with me their knowledge and helping me with using their equipment.

I would like to acknowledge Susi Schroeder for supporting me through all my PhD studies where she has showed me many techniques and provided me with the all necessary information and materials. Specific thanks are going to my PhD supervisor David Brown who helped me develop professionally and mentally.

I would like to thank the PDE4NPD consortium where members from different countries showed me their support and especially colleagues at the Free University of Amsterdam (VU) with who I worked in close collaboration.

Abbreviations

ACs – Adenyl Cyclases
BPH – benign prostatic hyperplasia
cAMP – Cyclic Adenosine Monophosphate
cGMP – cyclic guanosine – monophosphate
CAD – coronary artery disease
CD – Circular Dichroism
cspA – cold shock protein A
DMSO – Dimethyl sulfoxide
eNOS – Endothelial nitric oxide synthase
ED – erectile dysfunction
EDC – 1-ethyl-3-(3-(dimethylamino)propyl) carbodiimide hydrochloride
E. coli – Escherichia coli
FID – free interface diffusion method
GCS – Gianotti Crosti Syndrome
GCS – Glasgow Coma Scale
GST – Glutathione S-transferase
IDA – Iminodiacetic acid
IEC – Exchange Chromatography
IBMX – 3-isobutyl-1-methylxanthine
IPTG – Isopropyl β-d-1-thiogalactopyranoside
IMAC – Immobilized Metal – Affinity Chromatography
LB broth – Luria – Bertani broth
L. infantum – Leishmaniasis Infantum
MAPK – mitogen – activated protein kinase
NTDs – Neglected Tropical Diseases
NMR – Nuclear Magnetic Resonance
NHS – N-hydroxysuccinimide
NTA – Nitrilotriacetic acid
OD – optimal density
PDE – Phosphodiesterase
PKA – protein kinase A
PAH – pulmonary arterial hypertension
P – pocket – parasitic pocket
PCR – Polymerase Chain Reaction
RNAi – RNA interference
SIF – stumpy induction factor
SDS – sodium dodecyl sulphate
SPR – Surface Plasmon Resonance
SOC – super optimal broth

SLIC – Sequence Ligation Independent Cloning

SEC – Size Exclusion Chromatography WHO – World Health Organization

SAXS – Small Angle Scattering

T. cruzi – Trypanosoma cruzi

T. brucei – Trypanosoma brucei

TB – Terrific Broth

TBS – Tween Buffer

VSG – variant surface glycoproteins

VD–SD – vapour diffusion siting drop

VD–HD – vapour diffusion hanging drop

CHAPTER 1

Introduction

1.1 Kinetoplastid Parasites

Kinetoplastid parasites are a family of Trypanosomatids that cause a number of diseases worldwide; Leishmaniasis, Chagas disease and African Sleeping sickness are collectively termed as neglected parasitic diseases (NTDs). In the current study the main focus will be on 2 main infections: Leishmaniasis - caused by *Leishmania* species and African sleeping sickness - caused by *T. brucei*, Figure 1.1.

There are 700,000 to 1 million cases of Leishmaniasis that are registered, with 26,000 to 65,000 deaths annually, whereas sleeping sickness causes 10,000 reported cases annually and Chagas disease is reported to have 6 to 7 million people being infected worldwide *(1).*
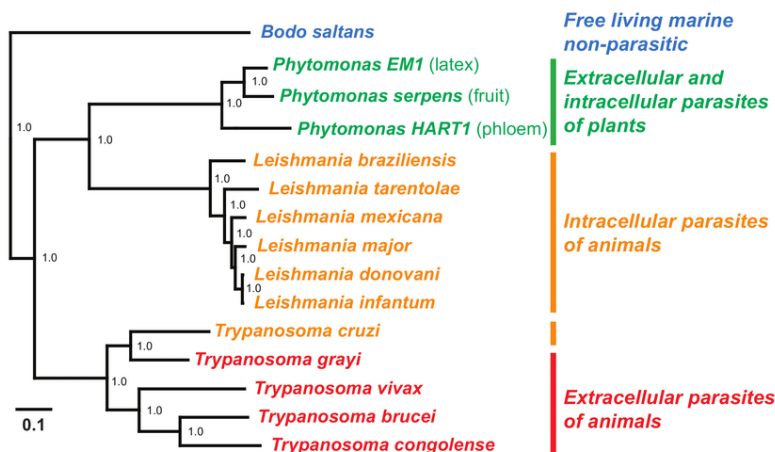
Figure 1.1: Phylogenetic tree of Kinetoplastid parasites. *(2)*

The main countries that are at risk are 3rd world countries with tropical and sub-tropical regions where the health system is not well developed.

Kinetoplastid organisms are characterized by having single flagellum and containing a particular organelle called kinetoplast.

1.2 Leishmaniasis

Leishmaniasis is a parasitic disease that is a widely distributed and causes endemic waves of outbreak in 100 countries in the regions of Asia, Africa, Latin America and even Southern Europe, Figure 1.2. More than 300 million people are at risk to develop this health condition.
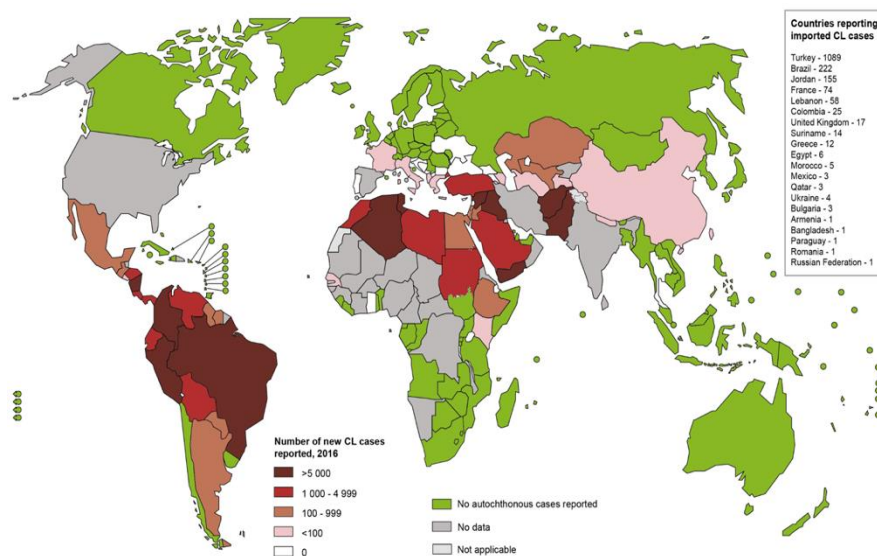
Figure 1.2: Leishmaniasis distribution across the world, (World Health Organisation) (WHO)

Leishmaniasis is a tropical health condition that mainly predominates in countries with poorly developed health system. Also, factors such as climate and environment can have a significant effect on disease transmission and occurrence. Infection is transmitted by female phlebotomine sandflies.

There is a significant discussion on the treatment of Leishmaniasis and other tropical diseases that can be used in order to control its transmission as well as the development of a new cases worldwide. Nowadays, there is an increased risk of infection as the modern world is more dynamic in

2

terms of tourism and immigration, and global warming could also be a factor to the disease being more widespread and being able to be transmitted to other European countries (World Health Organization). There are 3 types of Leishmaniasis, Figure 1.3: visceral – which is the most dangerous condition with fever, splenomegaly, weight loss, hypertrophy of liver and spleen; disfiguring mucosal – causes skin, mouth and nose ulcers and cutaneous type which is known of causing only skin ulcers.

Figure 1.3: 3 types of Leishmaniasis. A) visceral (WHO), B) disfiguring mucosal *(3)* C) cutaneous *(4)*

Treatments for Leishmaniasis include use of pentavalent antimonial (meglumine antimoniate & sodium stibogluconate), paromomycin,

sitamaquine, amphotericin B, miltesfosine and pentamidine drugs, Figure 1.4.
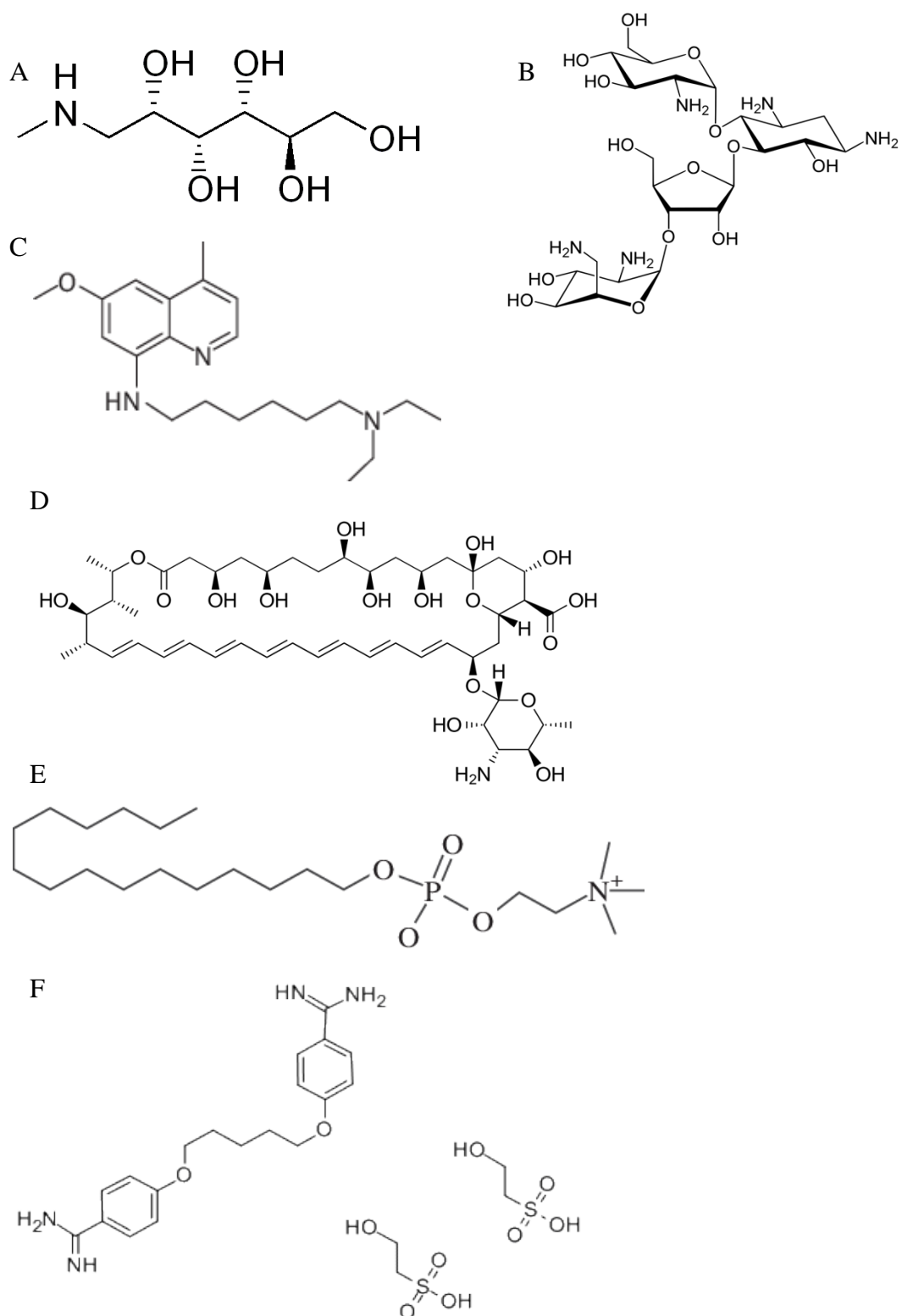
Figure 1.4: Current drug candidates for treating leishmaniasis. A) meglumine antimoniate, B) paromomycin, C) sitamaquine, D) amphotericin B, E) miltesfosine, F) pentamidine. ChemicalBook.com

However, there are certain drawbacks of current treatments, such as: drug's potency against infection, effectiveness on late stages, rising resistance and its cost. Historically, a Dihydroartemisinin (artesunate) derivative was the therapy recommended by WHO (World Health Organization) and is also often used to treat *Plasmodium falciparum* malaria in tropical and subtropical countries *(5).*

1.3 Mechanism of host cell invasion in Leishmaniasis

Transmission occurs by sand fly where parasitic amastigotes are transmitted during blood feeding of fly from the infected animal. In the gut of sand fly, amastigotes are transformed into infective metacyclic promastigotes *(6)*. After that, promastigotes migrate into the pool of host blood during sand fly feeding *(7)*. When sand fly bites the host, it stimulates immune system response of the host where neutrophils migrate to the site of infection to engulf promastigotes, followed by migration of cells away from the site of infection *(8)*. Promastigotes survive in a parasitic specific vacuole where they are transformed into amastigotes by dividing and breaking the cell. Amastigotes that were released after cell burst start infecting phagocytes causing infection. The last step in the cycle characterized by sand fly is taking up infected phagocytes during blood feeding, Figure 1.5.

Leishmaniasis
(Leishmania spp.)

Sandfly Stages

Human Stages

1. Sandfly takes a blood meal (injects promastigote stage into the skin)
2. Promastigotes are phagocytized by macrophages
3. Promastigotes transform into amastigotes inside macrophages
4. Amastigotes multiply in cells (including macrophages) of various tissues
5. Sandfly takes a blood meal (ingests macrophages infected with amastigotes)
6. Ingestion of parasitized cell
7. Amastigotes transform into promastigote stage in midgut
8. Divide in midgut and migrate to proboscis

i = Infective Stage
d = Diagnostic Stage

Figure 1.5: Schematic representation of Leishmania life cycle. *(9)*

1.4 Trypanosomatids

Another class of kinetoplastid parasites is trypanosomatids where T. cruzi is causing American sleeping sickness (Chagas disease) and T. brucei that is causing African sleeping sickness disease. These 2 infections are also health conditions that are lack of the effective drug treatment, where only 4 – 5 drugs are available on the market. There are 2 stages in the sleeping sickness: initial and late stages. Neurological conditions such as tremor, incontinence and GCS can be observed at both stages of disease where somnolence is enhanced at later stage, Table 1.1.

Neurological Sign	% Early	% Late
Gait Ataxia	63	52
Tremors	60	69
incontinence	20	15
neuropathy	29	31
somnolence	57	55
gcs<15	14	14
gcs<12	0	10

Table 1.1: 2 stages of sleeping sickness disease. Neurological conditions during 2 stages. GCS – Glasgow Coma Scale, **(10)**

Germanin (suramin) and Pentam (pentamidine) are the only 2 drugs that can inhibit further infection development as they are acting on the initial stage before more damage of the CNS (central nervous system) can occur.
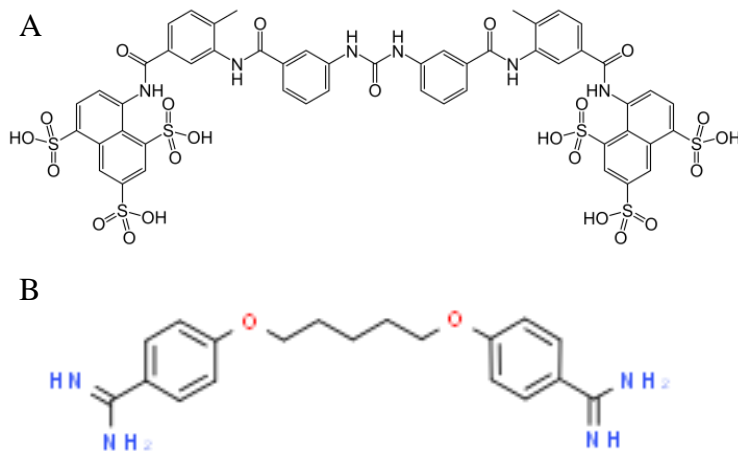
Figure 1.6: Drugs that are used during initial stage of infectious, A) Germanin, B) Pentam. **(10)**

When the condition has reached its final stage a very toxic drug is prescribed – Arsobal (arsenical melarsoprol), where side effects can lead to mortality in 5 to 10% cases. Unfortunately, as in case of Leishmaniasis there is no drug treatment that will act at both early and late stages, with less serious side effects. The Trypanosoma genus causes disease among

domestic animals as well as humans. The main transmission route is by the blood – sucking tsetse fly between animals and humans. If left untreated it can be fatal. Chagas disease is caused by T. cruzi where transmission happens from the animal to the host by bedbugs that have fed on contaminated faeces. The disease might not show any obvious symptoms however some patients can suffer from the physiological alteration in the heart as well as in the intestines and oesophagus *(11)*.

1.5 Mechanism of the host cell invasion in T. brucei

In the case of the T. brucei parasite, it is transmitted by the tsetse fly, *Glossina spp*, Figure 1.7. Trypanosomes develop in the gut of the fly and then transfer towards the salivary glands before infecting a new host organism. When the parasite is in the new host it is in its slender variation form, in the bloodstream, before its further transformation into a stumpy form *(12)*. In the bloodstream Trypanosoma can survive due to alteration of the host cell immune response by mimicking its antigens, like variant surface glycoproteins (VSG) that are connected with glycosylphosphatidylinositol (GPI). When T. brucei is transferred into its procyclic form the VSG is replaced by procyclin *(13)*. Parasites then move through the salivary gland of the fly where epimastigotes forms and attaches to the wall of the glands. Trypanosomes start dividing, increasing its number to be released into the gland of lumen before infecting a new host. At different stages of the trypanosome life cycle there are morphological changes where kinetoplast has different cellular localization. When trypanosome is in the bloodstream the kinetoplast locates closer to the end of cell; during its procyclic stage it is in the middle, between the cell nucleus and the end of the cell; during the epimastigote stage it is closer to the cell nucleus. However, it is still unclear why such morphological changes occur during each trypanosome

life cycle where the whole life cycle of trypanosome is very regulated *(14)*. Procyclin coat of the trypanosome forms after 2 hours of migration followed by kinetoplast migration where DNA starts its synthesis for 6 to 12 hours. Later, this stage is followed by continuation of proliferation stage where cells develop its mitochondrial properties. There is still some debate regarding the trypanosome life cycle in the mammalian bloodstream where most experiments have been performed in strains that have been modified in the laboratory. However, there is a known factor of changing slender into a stumpy form under the influence of SIF (stumpy induction factor) *(15)*. Also, this process is believed to be regulated by the cAMP signalling pathway where Phosphodiesterase (PDE) plays a crucial role.
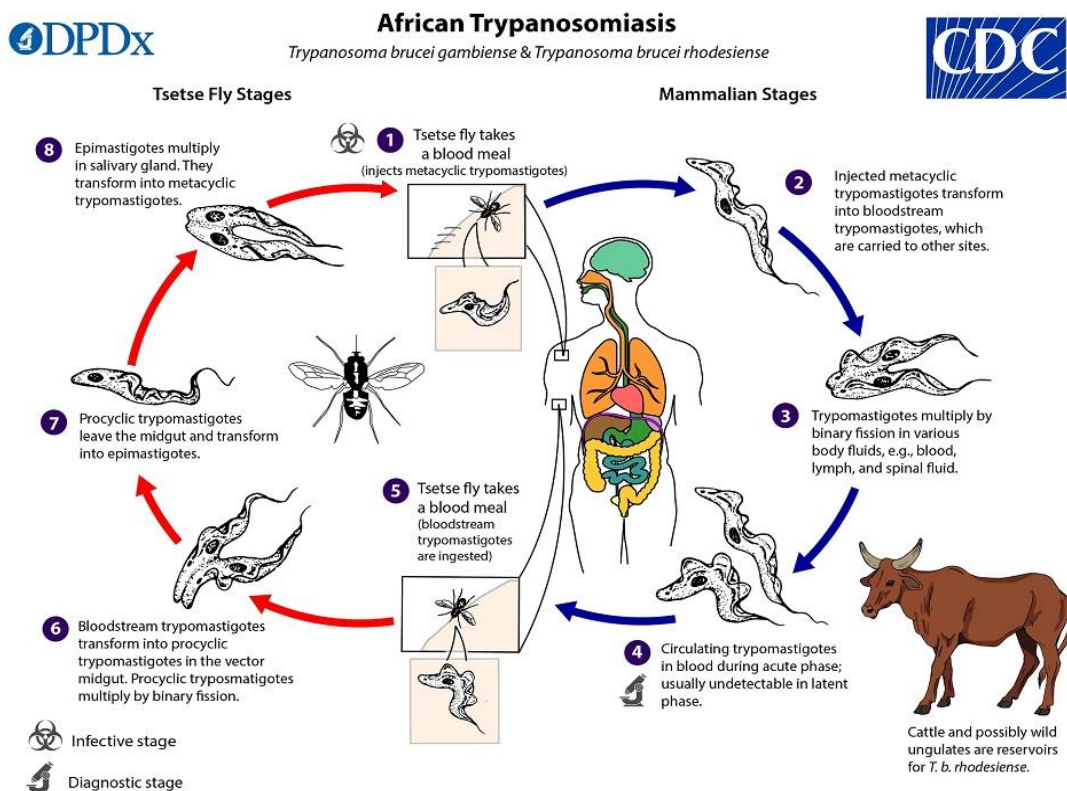
Figure 1.7: Schematic representation of T. brucei life cycle. (Centres for Disease Control and Prevention)

1.6 Cyclic Adenosine Monophosphate (cAMP) pathway in kinetoplastids

Second messengers are intracellular molecules that respond to any changes within the cell that are caused by extracellular molecules first messenger stimuli. Secondary messengers are important in cell signalling as they control processes such as cellular differentiation, proliferation and cell apoptosis. Kinetoplastids have cAMP as 2nd messenger that plays a vital role in parasitic survival, Figure 1.8. It has been shown that the genome of *T. brucei* has more than 70 genes and pseudogenes for encoding adenyl cyclases (ACs). The role of cAMP has been investigated with respect to parasite life cycle and what role it plays in infection development *(16)*.
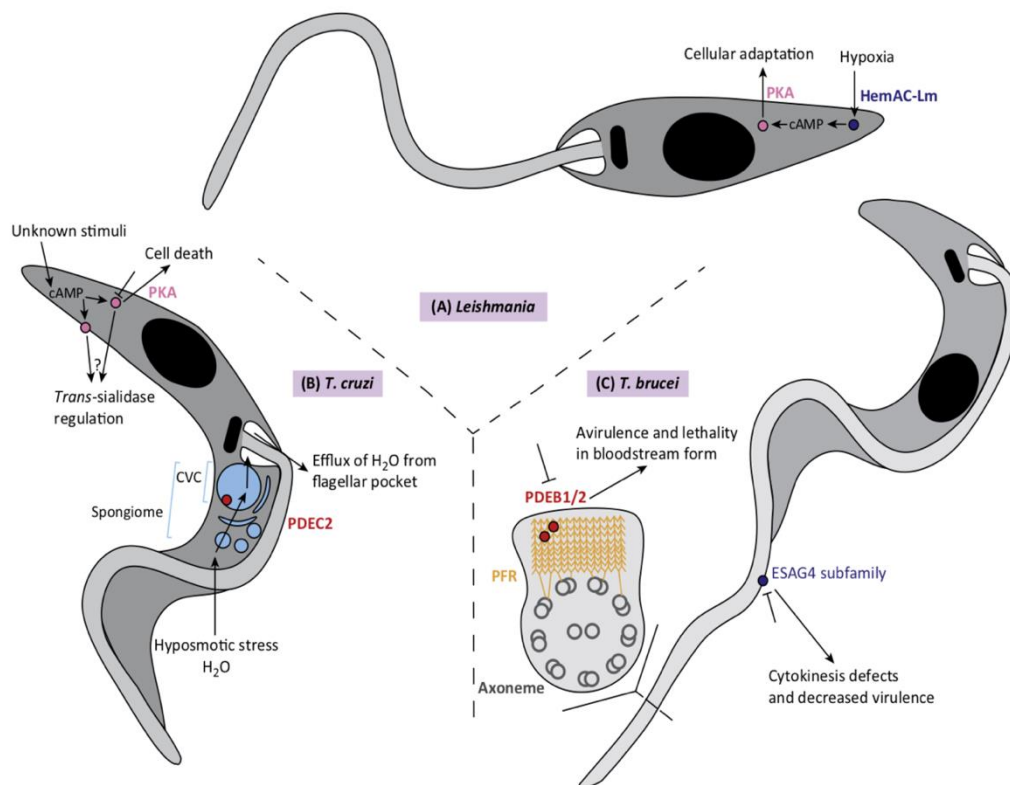
Figure 1.8: Schematic representation of secondary messenger signaling (cAMP) in kinetoplastid parasites. *(17)*

1.7 Proteins of cAMP pathway as a potential drug target

One of the cAMP regulatory targets is PKA (protein kinase A), that controls phosphorylation of substrate proteins, this in turn acts on ion channels and can further activate other enzymes. Studies on Leishmania showed that parasite has a critical dependency upon cAMP. The knockdown of the Hem-AC gene (globin-coupled heme containing adenylate cyclase) from the *L. major* led to cell death due to elevated level of 2nd messenger in the cell system as this protein directly stimulates adenylate cyclase activity. T. cruzi also showed a dependency upon cAMP and the knockdown of the same gene led to a fatal effect *(18)*.

In 2013 Gould et al investigated in more detail the downstream effectors in the T. brucei and found genes that were directly related to the cAMP signalling such as *CARP1-CARP4*. These genes were believed to be a part of cAMP signalling cascade and were important in regulation of secondary messenger activity, where its knockdown led to cell death. However, it is still unclear what exact function they perform in the signalling cascade. *CARP2* and *CARP4* were observed in flagellar proteome *(19, 20, 21)*. It was suggested by many findings that the flagellum is a very important organelle for parasitic survival and can potentially act as a drug target for inhibition of infection *(22).* Another kinase, MAPK (mitogen – activated protein kinase) is also present in kinetoplastid and performs the function as a chemoattractant in response to second messenger signalling. As in mammals it has also been proven that MAPK is important in interactions with PKA and PDE functioning. There are still debates regarding how vital it is for parasitic survival and it has also been suggested that it plays an important role in host – vector transfer *(23)*.

1.8 Cyclic nucleotide phosphodiesterases PDEs as a drug target

One of the biggest discoveries regarding possible new targets in kinetoplastids was targeting cyclic nucleotide phosphodiesterases (PDEs) in order to inhibit parasitic infection. PDEs are a class of enzyme that catalyse secondary messengers, such as 5', 3' – cyclic adenosine (cAMP) and guanosine – monophosphate (cGMP). These 2nd messengers are hydrolysed by PDE enzyme, where cAMP converts into 5' – AMP and cGMP into 5' – GMP. cAMP and cGMP are very important in controlling signal transfer where it's amplitude, duration and localization, play a vital role in the cell signalling and proliferation. There are 11 known PDE families in humans and according to variations in posttranslational modifications there are approximately 100 PDE polypeptides with different localizations and functioning *(24)*. The 3D structures of catalytic domain of human PDEs are highly similar, however the amino acid similarity among them is only 30 – 40%, Figure 1.9 *(25)*.
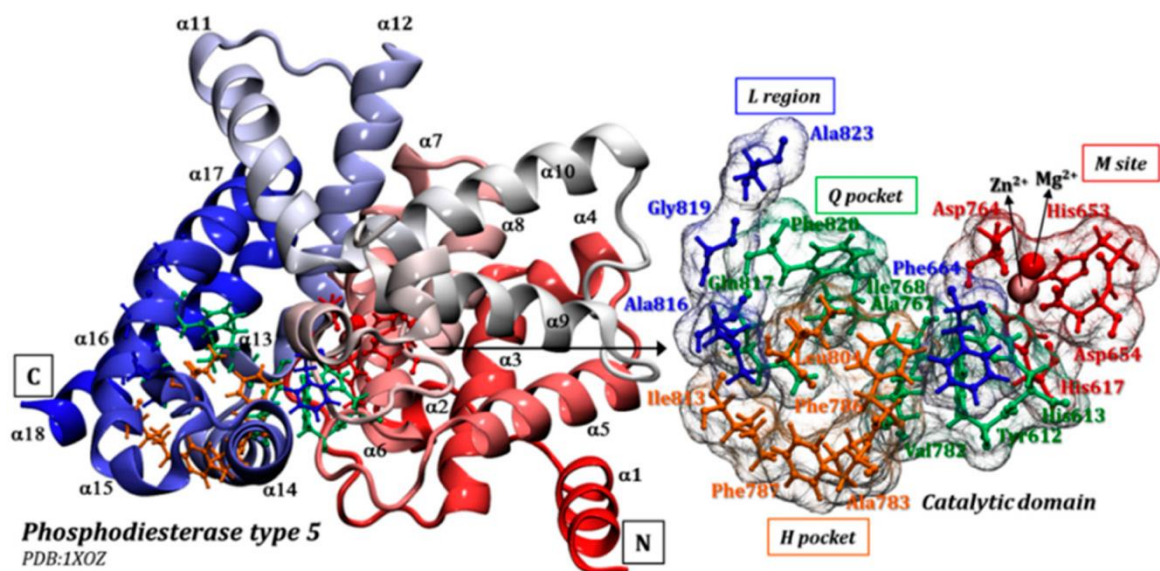
Figure 1.9: 3D structure of hPDE5 enzyme with highlighted amino acid residues of catalytic domain. *(103)*

12
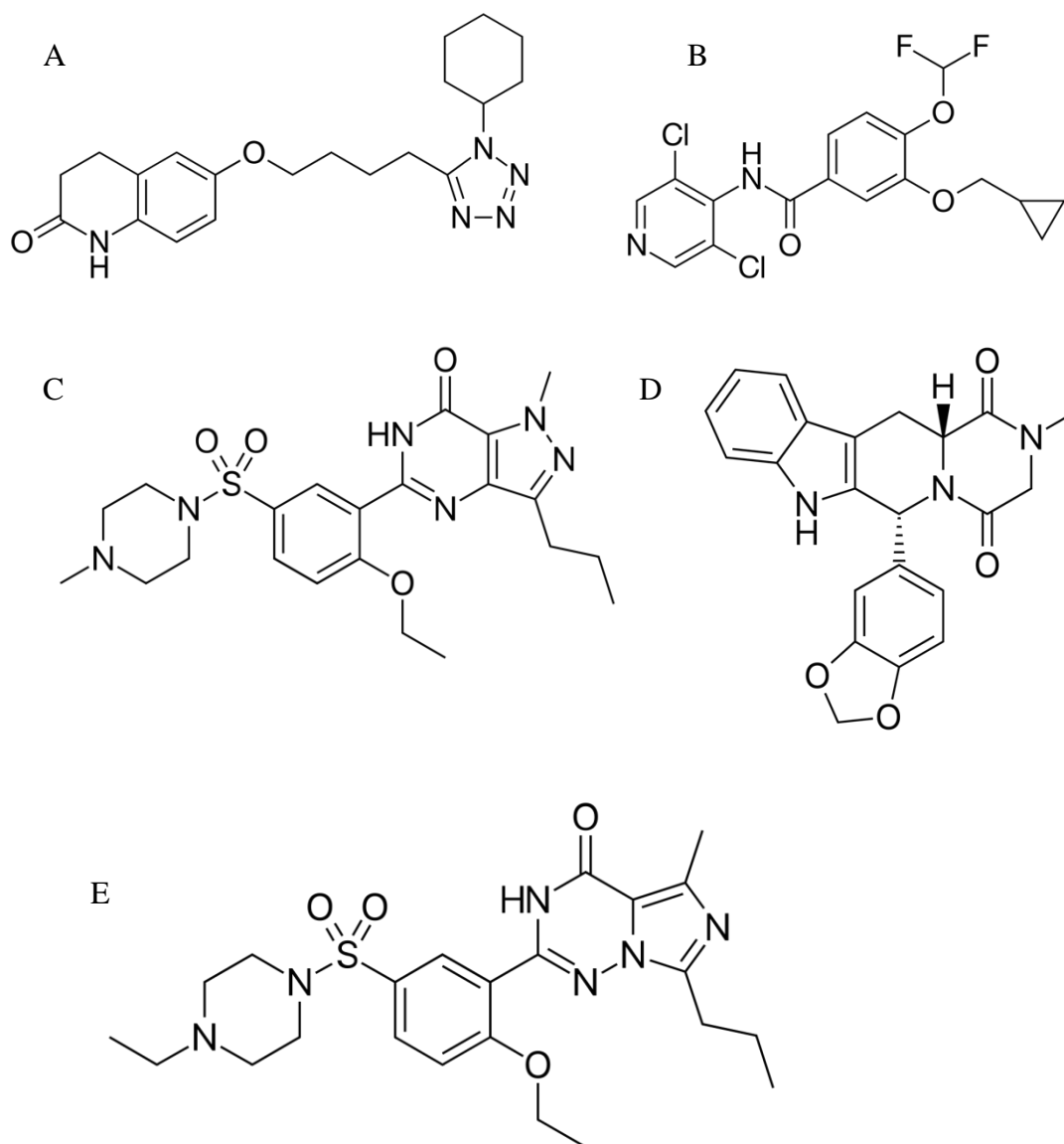
Figure 1.10: PDE – targeting drugs

The PDE enzymes are important for breaking phosphodiester bond and are significant focus of clinical research. PDEs are involved in different roles and can be found in healthy and/or in leukemic lymphocytes performing various functions *(26)*. PDEs have been implicated as potential therapeutic targets in health conditions such as: hypertension, depression, heart diseases, asthma, schizophrenia and infections that are caused by parasites. The most well-known drugs targeting PDE inhibition are: Pletal (cilostazol) acting on PDE3, Daxas (roflumilast) acting on PDE4, Viagra

(sildenafil) as well as Cialis (tadalafil) and Levitra (vardenafil) acting on
PDE5, Figure 1.10.
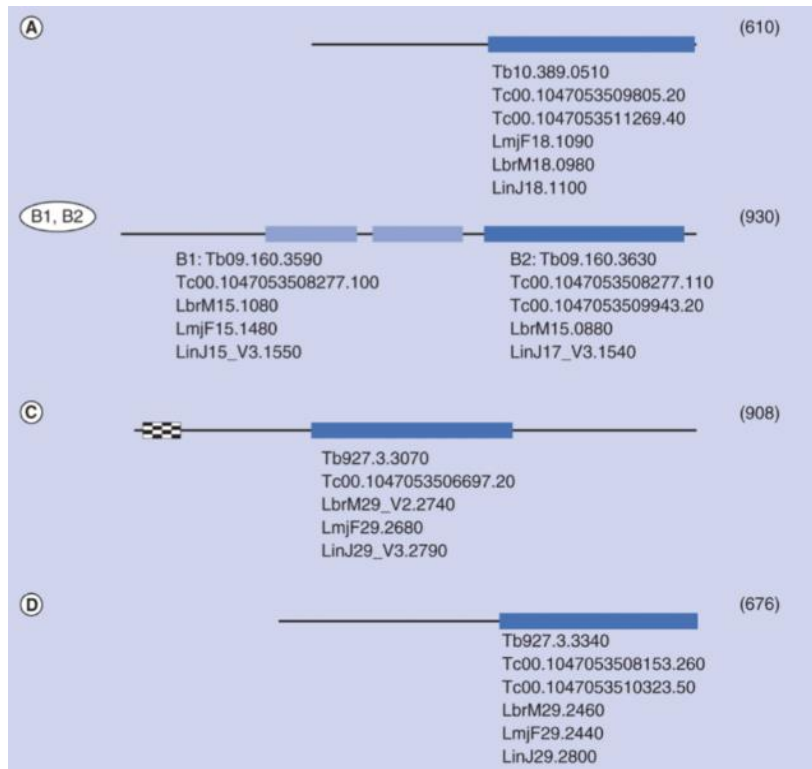
Figure 1.11: encoded parasitic PDEs (PDEA, PDEB,
PDEC and PDED). The blue boxes represent catalytic
domains. *(18)*

Kinetoplastid parasites contain several PDE class I genes which codes for
4 genetically different class I PDEs, Figure 1.11. However, among the
kinetoplasts species genes are highly conserved, Figure 1.11 *(27)*. There
are 4 distinct PDEs in parasites: PDEA, PDEB (codes for highly conserved
PDEB1 and PDEB2, with different subcellular location) *(28, 29, 30, 31)*,
PDEC (dual-substrate PDE with FYVE domain on N site) *(32, 33)* and
PDED that has only been identified as a DNA sequence. There were a
number of studies that confirmed PDEs as an essential protein for parasitic
life cycle and causing infection. These studies included those that showed
that PDEB1 and PDEB2 knock down altered the normal cell proliferation

and produced multiflagellated cells that eventually went to lyse stage. One possible explanation could be the localization of PDEB1 and PDEB2 in the paraflagellar rod, which is vital for motility and parasite life cycle in the host organism *(34)*.

T. cruzi PDEC was considered to be a potential drug target among the PDEs in that kinetoplast. PDEC has a FYVE domain (4 cysteine rich proteins that bind to phosphatidylinositol 3-phosphate) that is followed by 2 coiled – coil regions. It has been shown by Schoijet et al that this domain is vital for cell survival as well as its mobility in general. PDEC acts in cell osmosis and is triggered when the hypo osmotic stress releases active metabolites that are essential for cells secret water through pocket in the flagellum *(35)*. Therefore, inhibition of the parasitic PDEC will eventually lead to the cell mortality. Another RNAi study showed that genes *PDEB1* and *PDEB2* in T. brucei are essential for parasitic life cycle *(36)*. The experiment has been performed on RNAi (RNA interference) where its construct is induced by tetracycline and was incorporated into the T. brucei genome. The wild type strain of trypanosome is resistant to tetracycline but when trypanosome genome was altered, PDEB1 and PDEB2 activity was inhibited by RNAi. Experiments showed that when there was an absence of the tetracycline inducer in the system, tetracycline – repressor system was bound to the DNA operator and inhibited the expression of double stranded interfering RNA. These enabled trypanosomes to proliferate and cause infection in the host organism. Studies showed that if animals were given tetracycline prior to the infection, the tetracycline repressor complex was leaving the DNA operator to promote expression of interfering RNA to alter the PDE functioning. Inhibition of the PDE directly resulted on secondary messenger activity where increased cAMP level led to cell death *(37)*.

1.9 Human PDE5 as a drug target

PDE5 is a human enzyme that catalyses secondary messenger cGMP into its metabolite GMP. This action follows by calcium and inositol triphosphate to stimulate calcium release from the intracellular stores. This process regulates signal transduction in cellular environment. Inhibition of the PDE5 results in the high concentration of cGMP in the cell where calcium level would be decreased proportionally to cGMP resulting in relaxation of smooth muscles, Figure 1.12.

PDE5 can be found in different tissues of human body while the most predominant tissues are corpus cavernosum and retina. The 1st PDE5 inhibitor was Sildenafil that was prescribed to patients with erectile
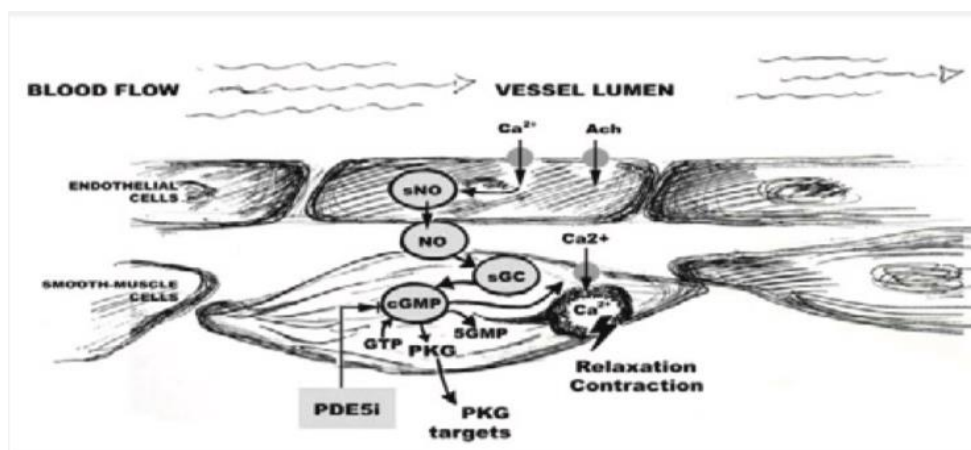
Figure 1.12: the mechanism of action of PDE5, catalysis of cGMP, *(38)*

dysfunction and pulmonary arterial hypertension *(39)*. PDE5 is also important for other systems, where it controls significant processes: vascular relaxation and improvement of pulmonary hypertension, improvement of cardiomyocyte stiffness and contractility, increasing eNOS activity where insulin signalling is improved in endothelial cells, improving insulin sensitivity in muscles and increasing aromatase activity, adipogenesis and insulin sensitivity.

Sildenafil was first produced in 1980 and introduced into the market in 1998 by Pfizer, where the primary aim was to treat arterial hypertension

and angina pectoris. It was suggested that inhibiting PDE5 would result in coronary vasodilation which was significant for patients with CAD (coronary artery disease). Early studies reported penile erection as a side effect, this unexpected observation became the primary condition for which the drug was developed.

	Sildenafil	Tadalafil	Vardenafil	Avanafil
Available doses (mg)	25-50-100	5-10-20	2.5-5-10-20	50-100-200
Effectiveness	60 min	15-30 min	30 min	15-30 min
Lasting (hours)	4-8 h	24-36 h	2-8 h	1-6 h
Food interaction	High - fat	Not significant	High - fat	Not significant
Alcohol interaction	None	None	None	None
Active metabolites	N-demethylation	None	Demethylation	Methylation, glucuronidation
Cytochrome P isoenzymes	3A4, 2C9, 2C19 and 2D6	3A4	3A4, 3A5, 2C9	3A4, 2C
Excretion	77 – 88% feces: minor in urine	70 % feces 30% urine	91 – 95 % feces minor in urine	62 % in feces 21 % in urine

Table 1.2: Pharmacokinetics and pharmacodynamics of PDE5 inhibitors

After sildenafil, 3 other drugs were introduced into the market, Table 1.2. In 2003 Vardenafil and Tadalafil were prescribing to patients with ED disfunction. In 2010, Vardenafil (Staxyn) was used in the form of ODT (oral disintegrating tablet) and was proved to be easier in its administration to the patient. In 2008, Eli Lilly got an FDA approval to administer Tadalfil once a day only, followed by its use to treat BPH (benign prostatic

hyperplasia) in 2011. In 2012, Avanafil was introduced into the market where drug has an advantage of its fast action. Nowadays, Sildenafil and Tadalafil are also in use to treat PAH (pulmonary arterial hypertension) using such names as Revatio and Adcirca *(40)*.

1.10 X–ray structures of T. brucei, Leishmania and T. cruzi PDEs

Structural and sequence comparisons have shown that parasitic PDEs are very similar to human PDEs as well as being conserved among themselves. TbrPDEB1 and LmjPDEB1 were the first parasitic PDEs for which X–ray crystallographic structures were determined.

In 2007 the crystal structure of LmjPDEB1 was solved, where the catalytic domain was comprised of residues 582-940, and the residues 597-931 were visible in solved apo crystal structure *(41)*. LmjPDEB1 structure has 16 α-helices however no β-strands, Figure 1.10. PDE catalytic domain have 2 divalent metal ions, located at the bottom of the active site. Zn interacts with His685/His722/Asp835 as well as with 2 molecules of water. Mg in its turn interacts with Asp722 as well as with 5 other water molecules. Figure 1.13 shows that human PDE and parasitic PDEs are very similar in terms of metal ions localization.

The overlay of LmjPDEB1 crystal structure with the human PDE crystal structure showed root-mean-square deviation value of approximately 1.3-2.0 Å for all α - Carbon atoms of PDEB1 structure, PDE2A apo structure, PDE3B in complex with IBMX (3-isobutyl-1-methylxanthine, PDE4D2 in complex with
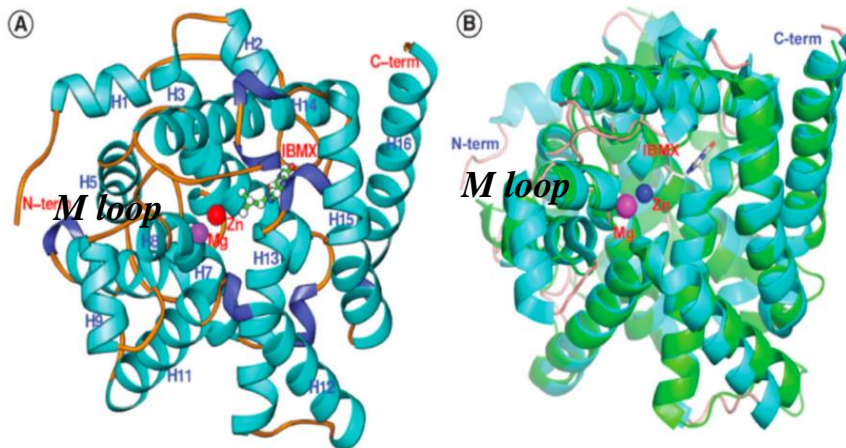
Figure 1.13: X-ray structure of L. major PDEB1. A) diagram of the L. major PDEB1 catalytic domain. B) overlapping of parasitic PDEB1 with human PDE4(green). *(94)* The pdb code was absent for current structures from the review paper.

IBMX, PDE5A1 in complex with IBMX, PDE7A1 with IBMX, PDE7A1 with IBMX, PDE9A2 with IBMX and finally PDE10A2 in complex with secondary messenger cAMP. Superposition of parasitic PDEs with human PDEs reveals high similarity in protein 3D structure *(42)*.

Studies showed that H9 helix and H loop of LmjPDE residues 729-754 as well as M loop of residues 858-882 had a shift in its localization by 3 Å for carbon α atom from human PDE. Therefore, it can be concluded that since H/M loops were playing a significant role in protein ligand interaction this led to further discovery of selective antiparasitic inhibitors *(43, 44)*.

1.11 Studies on parasitic P – Pocket in kinetoplastid parasites

Structural studies of parasitic PDEs have identified a specific pocket in the parasitic PDE that is located next to the active site and is very close to the actual ligand binding site *(45)* termed the P – pocket, Figure 1.14.
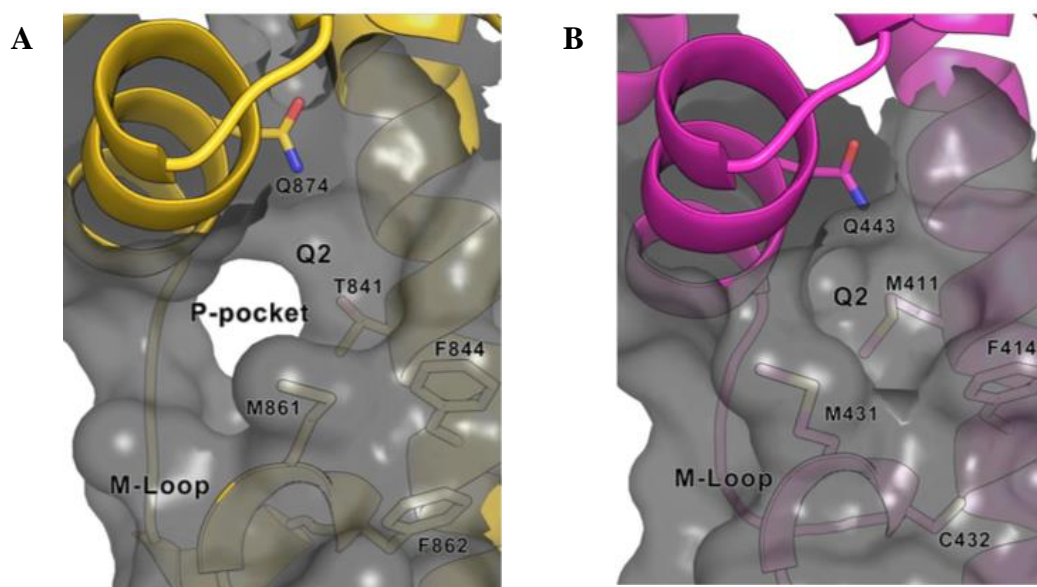
Figure 1.14: Schematic representation of catalytic domains in A) T. brucei PDEB1, representing p pocket and B) active site of human PDE4 where there was no gap in the surface and only Q2 is present. *(46)*

The P – pocket was observed in LmjPDEB1 and also when the TbrPDEB1 3D structure was solved, indicating that this unique feature may be common to all kinetoplastid parasite PDEs *(47)*. Since the primary aim of this thesis was to help develop new anti-parasitic drug candidates, it was also proposed that this unique P – pocket feature can be used as a primary ligand interaction target in order to make compounds selective for parasite over human PDEs.

Figure 1.15 represents P – pocket binding site, where the open cavity was found between H14, H15 helixes and M loop. Conserved residues were: Gln874, Val840 and Phe877 (hydrophobic clamp), Phe844 and Phe880 (shown as sticks). Hydrophobic clamp is formed between 2 hydrophobic residues, such as Val840 and Phe877 and is considered as a usual binding

site for PDE inhibitors. The P – pocket residues were: Ala837, Thr841, Tyr845, Asn867, Met868, Glu869 and Leu870 (represented as lines), Figure 1.14 *(76)*.
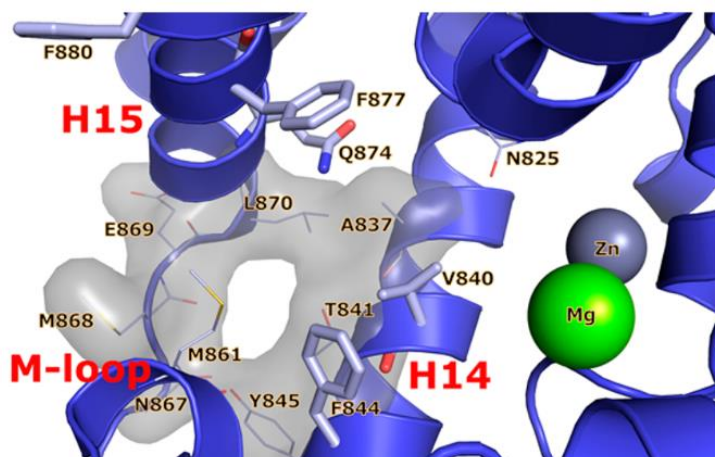
Figure 1.15: Schematic representation of catalytic domains in T. brucei PDEB1, representing p pocket: its residues and 2 metals Mg and Zn, (PDB code 4I15) *(76)*

1.12 Ligand – Protein Interactions

There are a variety of ligand – protein interactions that play an important role in the binding mode, such interactions are: hydrophobic interactions, hydrogen bonds, π- stacking, salt bridges. Amide stacking and cation-π interactions, Figure 1.16, *(49)*.

Data from 2015 suggested that among data that was analysed there were more hydrophobic interactions in high efficiency ligands *(49).* The amount of hydrogen bonds in such highly efficient ligands were less. One possible explanation could be, that when ligands are optimized in the drug design process it is simpler to increase potency through non-directional hydrophobic interactions over improving binding energy by specific, directional hydrogen bonding. Fragments bind to the protein with higher ligand efficiency (LE = (ΔG)/N) *(32)*, than many larger ligands with more polar interactions in fragment – protein complexes on average with

doubled amount of hydrogen bonds per atom and higher proportion of electrostatic interactions.

Since fragments are inherently weak binders they are tested at high concentrations where it is important, they have high polarity to ensure to have high solubility. In comparison to the final drug compounds, fragments are also more flexible in terms of their binding poses in order to allow optimal polar interactions as usually affinity increases are gained by introduction hydrophobic interactions.
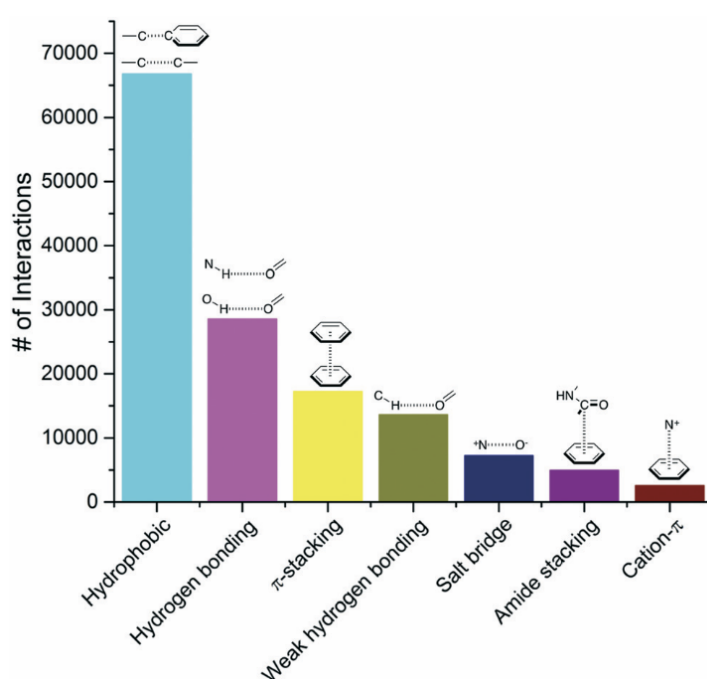
Figure 1.16: Bar chart representing how often each type of interaction occurs among protein-ligand complexes from PDB, ***(49)***

1.12.1 Introduction of Hydrophobic Forces

Hydrophobic forces are characterized by interactions of carbon and carbon, halogen or sulphur atoms. There are 5 different types of hydrophobic forces, where the interaction between the receptor aliphatic carbon and aromatic carbon of ligand were the most predominating type of hydrophobic forces. This shows that aromatic rings are important for ligand binding mode, where benzene ring is the most common used ring

(50, 51). Amino acids that are the most common in formation of hydrophobic interactions are: Leucine, Valine, Isoleucine and Alanine. One of the common features is the formation of hydrophobic interactions between aliphatic or aromatic carbon in the protein with chlorine or fluorine atoms in the ligand and between sulphur atom of Methionine with aromatic carbon atom in the ligand. It was also proved by Valley et al in 2012 that hydrophobic interaction between Methionine and aromatic carbon atom of the ligand yields additional energy of 1 – 1.5 kj *(52).* A buried methyl of a ligand in a hydrophobic region of a protein is calculated to provide approximately 0.7 kj which can translate to an increase of binding affinity by as much as 3.2 – fold *(53).* By occupying the hydrophobic clamp, methyl group displaces water molecules that are in the ligand binding site hence potentially increases the ligand potency through an increase in entropy of the system *(54).*

1.12.2 Hydrogen Bonds

The most common hydrogen interactions are between N – H and O, followed by O – H and O, and N – H and N, Figure 1.17.
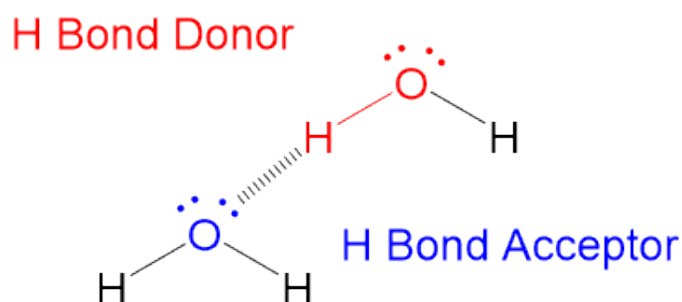
Figure 1.17: Hydrogen Bonds, (studyorgo.com)

It has been shown by de Freitas that proteins were mostly acting as H – donors. Arginine is more likely to form H – bonds then the Lysine due to

the fact that it has 3 N – atoms in its structure of a sidechain. In O – H and O bond formation, compound was a donor where the acceptor more often was the Aspartic acid where H – bonds were charged, whereas in case of Asparagine, Glycine and Glutamine there were neutral bonds formation. Water molecules in the protein structure act as a very important regulator for H – bonds formation. The usual separation of atoms between each other during H – bond formation is 3.0 Å. The free energy of H – bond interactions is -1.5 – 1.5 kj to – 4.7 – 1.5 kj. One of the most important factor that influence H – bond interaction strength is the H – interaction environment i.e. if it is solvent exposed, or shielded in a hydrophobic environment *(55).* In some experiments, it has been shown that by addition of polar atoms into the ligand optimal H bonding interactions resulted in enhanced binding affinity of ligand by more than 500-fold *(56).*

1.12.3 π - stacking interactions

The most common example of a π - interaction between ligand and proteins is exemplified by the interaction of the delocalized aromatic ring of Phenylalanine an aromatic ring of a ligand. A similar interaction is observed by residues Tyrosine in 36.8% cases (data was studied from Protein Data Bank), followed by Tryptophan in 8.7% cases and Histidine in 5.1% cases. Such π - stacking interactions play a crucial role in the drug design approach as they provide significant binding energy. However, it has been shown that decreasing the number of aromatic rings in the ligand structure leads to improvement of other physiochemical properties, such as solubility *(57).*

1.12.4 Weak Hydrogen bonds

Current type of interactions is characterized by the formation of bond between C – H and O, where C atom is aromatic and where O atom from proteins act as electrons acceptors. Such amino acids as Glycine, Aspartic and Glutamic acids were found the most often to be acceptors where Leucine was acting as a donor in most of the cases. The distance between atoms was found to be 3.4 Å *(49)*. Interactions between C – H and O are half less in magnitude strength in comparison with NH and O=C bonds. It has been shown that such interactions are very important for molecules recognition as well as protein folding and how stable it is *(58, 59, 60).*

1.12.5 Salt Interactions

Salt bridges arise by the formation of bond between positively charged N atom of the protein structure with a negatively charged O coming from the ligand or vice versa. It has been shown that Arginine amino acid played a role of cation in most of the salt bridges interactions *(61).* The analysis of such interactions revealed that there was higher density of O with negative charge around the ω N atom than with the secondary amine ε N. As motioned before in Section 4.2.1.4, as with hydrogen bonds the strength and binding energy of salt bridges is dependent on the local environment. Current paper didn't mention approximate bond energy.

1.12.6 Amide π - stacking interactions

Such interactions are characterized by the amide group of the amino acid backbone π - stacking with the aromatic ring of a ligand or the amide of a ligand staking with an aromatic amino acid residue. There are 2 types of interactions here, such as face-to-face and edge-to-face. Glycine and Tryptophan are examples of where this can occur within a protein structure. It was suggested by Duan et al in 1999 that π - stacking

interactions are important for protein structures as well as being vital for drug – protein binding mode *(62)*. Current paper didn't mention approximate bond energy.

1.12.7 Cation π - interactions

These interactions are characterized by the formation of bond between N positively charged atom of the protein and aromatic ring of the ligand where the interaction is electrostatic due to the presence of a negatively charged electron cloud *(63)*. Arginine is one of the most often amino acid involved in such interactions. Cation π - interactions are important for the protein stability and functionality *(64)*.

1.12.8 Halogen Interactions

Current interactions are characterized by the formation of bond between C – X and Y, whereas X represents Cl, Br and I atoms; while Y represents such atoms as O, N, S, Figure 1.18.
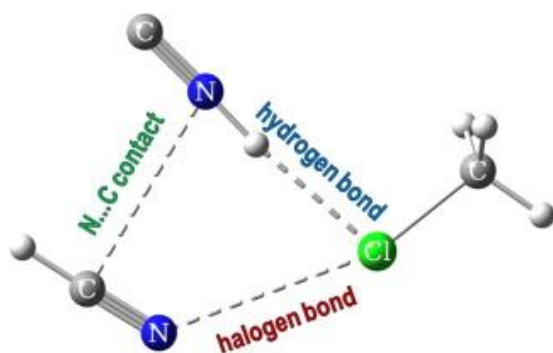
Figure 1.18: Halogen Interactions, *(65)*

The protein side chain or its backbone provides Y atom for interaction while X atom is coming from the ligand. This type of interaction arises when there is σ-hole from any of the halogen atom and a nucleophile which acts as an acceptor *(66, 67, 68)*. Even though that fluorine is the halogen it doesnt form halogen bonding due to the fact that it has higher electronegativity and is less polarizable then 3 previosly mentioned

halogens *(69).* Among 3 halogens that form halogen interactions Cl atom is the most observed, followed by bromine and then iodine *(70, 71).* The value of the angle in halogen bond formation is 156° and is characterized as a linear arrangement, where the O atom is acting as an acceptor more often than any other atoms. The main reason why halogens are introduced into the ligand structure is to improve membrane permeability and metabolic stability of the compound.

1.13 Review of previous Medicinal Chemistry design of selective parasitic PDEs inhibitors over human PDEs.

This section describes two recent papers focused on the design of novel parasitic PDE inhibitors *(73)*.

Since NPD – 001 was the most potent PDE inhibitor it was tested against
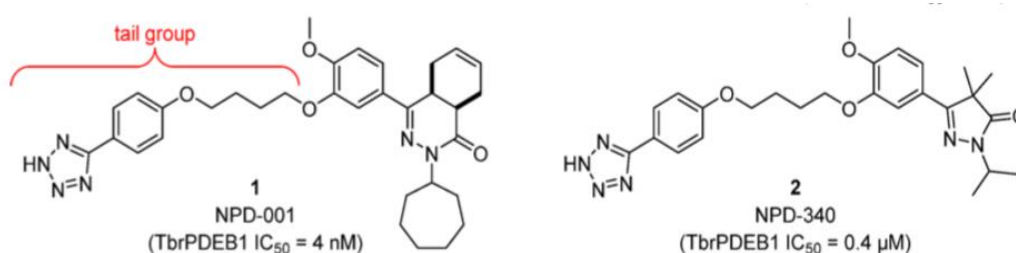
Figure 1.19: hPDE4 and TbrPDEB1 inhibitors: NPD-001 and NPD-340, *(73)*

Compd	R	TbrPDEB1[a]	hPDE4B1[b]	IC50 TbrPDEB1/IC50 hPDE4B1
5	CH₃	6.5	9.2	501
1	(tetrazole)—O—(CH₂)₄-	8.4	9.2	6
12	(methyltetrazole)—O—(CH₂)₄-	8.5	9.3	6
13	(tetrazole)—O—(CH₂)₄-	8.0	9.2	16
14	(imidazole)—O—(CH₂)₄-	7.9	9.7	63
15	(purine)—N—(CH₂)₄-	7.8	8.8	10
17	HO₂C—O—(CH₂)₄-	6.5	9.5	1000

a −log IC₅₀ on TbrPDEB1 (*n* = 2, ±0.2).
b −log IC₅₀ on human PDE4B1 (*n* = 2, ±0.2).

Table 1.3: analogs of series 1. Potency values against hPDE4. a- logIC50 on TbrB1 (n=2, +- 0.2), b- logIC50 on human PDE4B1 (n=2, +-0.2). *(72)*

other families of PDEs, where the tail group of it was used, Figure 1.19. Among many human PDEs where the inhibitor only showed some kind of interaction with high drug concentration, human PDE4 was inhibited at the same potency as parasitic PDEs. Therefore, it was assumed that hPDE4 was the main target that would be used for comparison of inhibitor selectivity. NPD – 001 series compounds were potent against human PDE4 whereas no selectivity was observed between the parasitic PDE and the human PDE. Hence, the next step was to synthesize new tool compounds where phenylpyridazinones were tested against hPDE4, Table 1.3.

Table 1.3 represents phenylpyridazinones compounds that were tested to identify its potency against human and parasitic PDEs. Analysis revealed that there was 100 – fold variation in the inhibition of potency against TbrPDEB1 whilst in human PDE4 the variation was only 8 – fold. Comparing compound 5 and 1 showed the significance of tetrazole presence in the chemical structure where 80 – fold difference in affinity against the parasitic TbrPDEB1 was observed. The ratio between IC_{50} of parasitic and human PDEs was calculated for different compounds, Table 1.3, highlighting the differences from 6 – 1000 hence pointing out that selectivity between human and parasitic PDEs is possible. As a result, it can be concluded that the optimization of parasitic PDEs inhibitors is not possible with this series of compounds and further experiments would be required, *(72)*.

Another inhibitor analogue, NPD – 340 that also has an effect on parasitic T. brucei PDEB1 was studied *(73)*. However, the main disadvantage was that it had a higher potency against human PDE4 than towards parasitic PDEs. Blazer et al performed a docking analysis where dialkoxyphenyl moiety from both NPD – 001 and NPD – 340, Figure 1.16, was linked to Gly874 of parasitic PDE by forming hydrogen bond and appeared to be found in the hydrophobic clamp of the active site.

It was shown that these compounds were interacting with PDE utilizing their tetrazole group. NPD – 001 was found in binding pocket of active site in TbrPDEB1 with the dialkoxyphenyl interacting with Val840 and Phe877. In comparison to the previous study, Blaazer et al research revealed another binding feature such as that the side chain of the inhibitor was out of the P – pocket, while in the previous study it was suggested to have interaction with Phe880 *(73)*. Analysis of the binding region revealed that hydrophobic clamp was the most favourable anchor region for ligands allowing flexible side chains orientated towards the P – pocket. Hence it was concluded that synthesizing derivatives with different alkyl linker would eliminate such problem as the selectivity. Another finding was made regarding the observation of the ligand binding into the hydrophobic clamp by mainly aromatic groups, whereas hetero-aryl groups were orientated towards parasitic P-pocket in TbrPDEB1. Therefore, it was concluded by Blazer et al that replacing flexible linker of NPD – 001 of the dialkoxyphenyl ring will enable the inhibitor to be more flexible in order to be directed towards the P – pocket.

A set of experiments were performed where 3 – 6 biphenyl derivatives of NPD-001, Table 1.4, were used in crystallization with TbrPDEB1 and human PDE4, Table 1.4. Derivative 3 and 4 showed higher selectivity towards parasitic PDEB1.
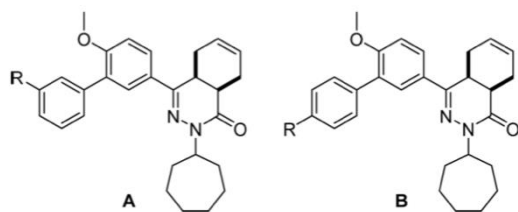
	Scaffold[b]	R	TbrPDEB1 pK_i	hPDE4B1 pK_i	Selectivity[c]
roflumilast			5.5	9.4	0.0001
1 (NPD-001)			8.4	9.2	0.16
3 (NPD-740)	A	O‖ —OH	5.1	4.9	2
4 (NPD-733)	B	O‖ —OH	5.8	5.1	5
5 (NPD-584)	B	O‖ —NH₂	6.8	6.8	1
6 (NPD-744)	B	—CN	6.3	7.5	0.1

Table 1.4: Synthesised biphenyl derivatives with introduction of biphenyl linker into NPD-001 at position 3 *(73)*

Analysis of bindings of compounds NPD-001 and NPD-340 (pdb code absent) revealed that hydrophobic clamp was interacting with the linker where it was directing towards such aromatic residues as Phe880 in TbrPDEB1 *(73).* Therefore, it was decided to synthesize analogues with the substitution of flexible region at position 3 with the introduction of the phenyl ring that will perform a function of biphenyl system and will be vectoring towards P – pocket.

In conclusion, parasitic PDEs and human PDEs as much as they were similar, they were still structurally different hence resulting in different binding mode and ligand potency.

Screening has showed differences in selectivity between human and parasitic PDEs. The potency was increased nearly 10 – fold in TbrPDEB1 when carboxamide part was introduced in the para – position of inhibitor 5

as compared to the meta substituted acid, Table 1.5. The next step was to synthesize derivatives based on the scaffold of inhibitor – 5 in order to enhance its inhibitor selectivity between human and parasitic PDEs. By the addition of the methoxyethyl group (compound 7) the selectivity was improved by 7 – fold while the potency was unchanged. Derivatives 8 and 9 showed increase in potency against parasitic PDEB1 with Ki values of 100 nM and 99 nM. The inhibitor – 8 was also more selective towards TbrPDEB1 then to human PDE4B1 by 10 – fold, Table 1.5. Derivative – 9, Table 1.5, was more selective against TbrPDEB1 by 19 – fold and it was concluded that glycinamide inhibitors were more in favour to be used against parasitic PDEs. Derivatives 10 – 13, Table 1.5, were synthesized with the addition of N – methyl, N, N – dimethyl and N – isopropyl side chains on the P – pocket tail. However, such changes decreased the inhibition activity of the compounds. Compounds 12 and 13 were alkylated as a result hydrophobicity of inhibitors was changed whereas potency and selectivity were decreased. Compound – 14 was less selective and showed to have less inhibitory activity. Derivatives 16 and 17 of compound – 8 were extended by the addition of hydroxyethyl and methoxyethyl groups respectively which gives increase in the selectivity by 8 – fold.
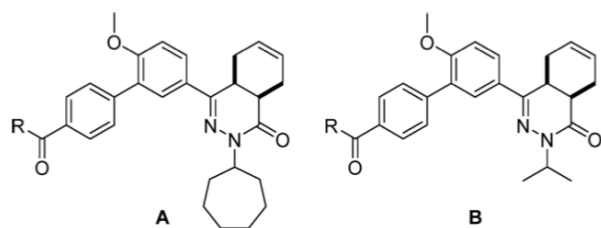
	Scaffold[b]	R	TbrPDEB1 pK_i	hPDE4B1 pK_i	Selectivity[c]
7 (NPD-734)	A		6.5	5.7	7
8 (NPD-008)	A		7.0	6.0	10
8a (NPD-949)	A (4aS,8aR)		5.4	4.3	13
8b (NPD-1373)	A (4aR,8aS)		7.4	6.2	16
9 (NPD-039)	B		7.0	5.7	19
10 (NPD-935)	A		6.4	5.6	7
11 (NPD-936)	A		6.4	5.6	6
12 (NPD-939)	A		6.1	5.5	4
13 (NPD-942)	A		5.7	5.2	4
14 (NPD-800)	A		6.3	6.0	2
15 (NPD-801)	A		6.1	5.4	5
16 (NPD-937)	A		6.4	5.5	7
17 (NPD-878)	A		6.3	5.4	8

Table 1.5: derivatives 7-17, studies on SAR in parasitic pocket of TbrPDEB1 *(73)*.

It was further suggested by Blazer et al, to perform crystallization of the TbrPDEB1 with derivatives 8 and 9. High – resolution structures were obtained at resolutions of 1.8 Å and 2.0 Å respectively, Figure 1.20. The crystal packing was very similar to the TbrPDEB1 with NPD – 001.
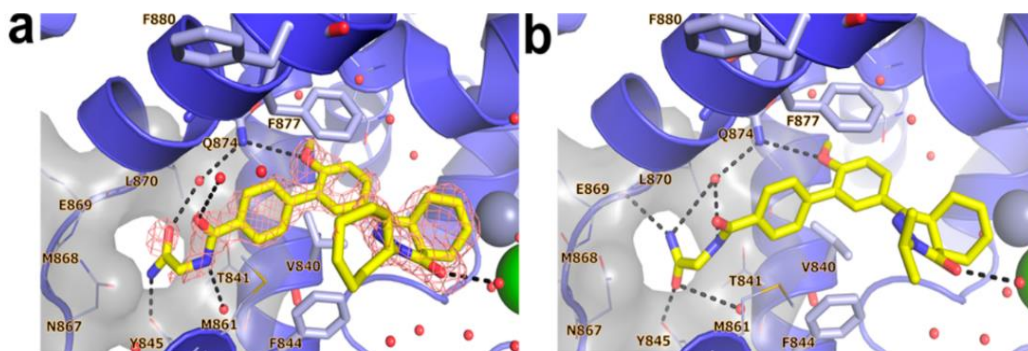
Figure 1.20: T. brucei PDEB1 crystal structure with A) derivative 8 B) derivative 9, *(73)*

Figure 1.20 represented 2 crystal structures where compounds 8 and 9 are bound. According to the ligand protein interaction Gly874 interacted with both ligands where ligands were bound to the hydrophobic site by biphenyl tail of methoxyphenyl ring of the compounds. Both compounds were directed towards the P – pocket where the inhibitor 8 was bound to the protein via interaction of glycinamide tail with Tyr845 and forming 3 – Hydrogen interactions between water molecules and Thr841, Met861, Leu870, Gly873 and Gly874.
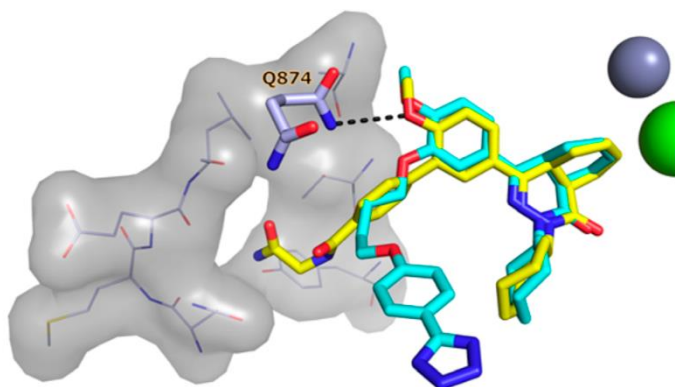
Figure 1.21: comparison of crystal structure T. brPDEB1 with NPD-001 and crystal structure with compound 8.

Next suggestion was to align these 2 inhibitors to see the differences in binding mode and their orientation toward the P – pocket.

Figure 1.21 represented 2 aligned crystals structures where inhibitor – 8 (yellow) showed a tail direction towards the P – pocket. Derivative – 9 (blue) showed different structural confirmation where hydrogen bonds were formed with Tyr845, Glu869, Thr841, Met861, Leu870 and Gly874. Another important characteristic that was studied by Blazer et al was the B factor of the P – pocket where the data showed that parasitic pocket was very flexible in comparison to the rest of the protein structure. Unfortunately the exact data was not mentioned in the paper. Glycinamide tail of compounds 8 and 9 showed electron density that was weaker in comparison to the rest of the molecule what suggested that inhibitor was more flexible in the P – pocket. These findings suggested that further experiments were important where the main focus would be orientated on the modification of the aliphatic heterocyclic tail aiming to enhance its flexibility in the P – pocket.

Table 1.6 showed newly synthesized derivatives where inhibitor 18 and 19 showed less selectivity between the parasitic PDE and human PDE4B1 with decreased inhibitory activity. Derivative 20 in comparison to inhibitors 18 and 19 was 3 times more selective towards the parasitic PDEB1 then to human PDE4B1. Compound 21 with (R)-pyrrolidin-3-ol addition to the compound structure was more selective (x 9 times) towards targeting the P – pocket as well as having higher potency level. Derivative 22 showed 15-fold selectivity for TbrPDEB1, where derivative 24 was 10-fold more selective fro parasitic PDEB1. However, compound – 22 was chosen as a preferable compound for analysing its interaction with TbrPDEB1 in the crystal form. The data came at 1.8 Å where 3-hydroxypyrrolidine tail was directed towards the P – pocket by forming

hydrogen bonds with Met861 and Asp867, Figure 1.22. The overall mode of binding was similar to other inhibitors.
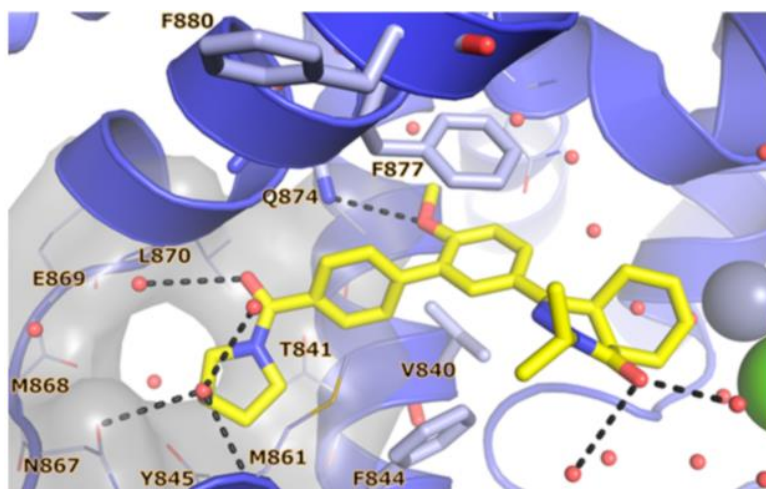
Figure 1.22: crystal structure TbrPDEB1 with compound 22 *(73)*
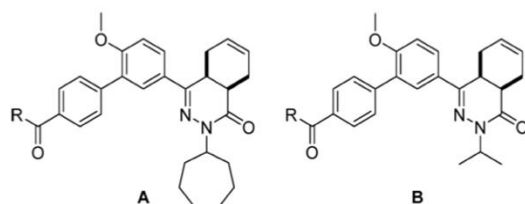
	Scaffold[b]	R	TbrPDEB1 pK_i	hPDE4B1 pK_i	Selectivity[c]
18 (NPD-060)	A		6.2	6.2	1
19 (NPD-062)	A		5.8	5.6	1.4
20 (NPD-887)	A		6.0	5.5	3
21 (NPD-746)	A		6.6	5.7	9
22 (NPD-038)	B		6.7	5.5	15
23 (NPD-802)	A		6.3	5.2	13
24 (NPD-885)	B		6.2	5.0	18

Table 1.6: analysis of aliphatic heterocycles potency and selectivity in T. brucei PDEB1

Further analysis revealed that compound – 8 showed increase of intracellular cAMP level in the bloodstream of trypanosomes at 100nm. Compounds 8 and 9 were considered to be as the most promising inhibitors from all the series of compounds that showed inhibitory activity against TbrPDEB1. However, further optimizations were required in order to consider these compounds as potential drug candidates against kinetoplastid parasites.

1.14 IBMX-PDE interaction

The ligand IBMX is a known nonspecific PDE inhibitor against human PDEs however it didn't show the same level of potency against parasites. In LmjPDEB1, where the IC_{50} was only 580 µM ligand – protein interactions showed an H bond between O6 of the IBMX xanthine ring and Nε2 of amino acid Glu887 of LmjPDEB1, Figure 1.23. IBMX binding mode in parasitic PDEs was very similar to binding mode of ligand in human PDE, however IBMX position and its localization was slightly different to humans.

The key glutamine was observed in all kinetoplastid PDEs with one exception. T. cruzi PDEC and T. brucei PDEC, where it was replaced by Serine residue.
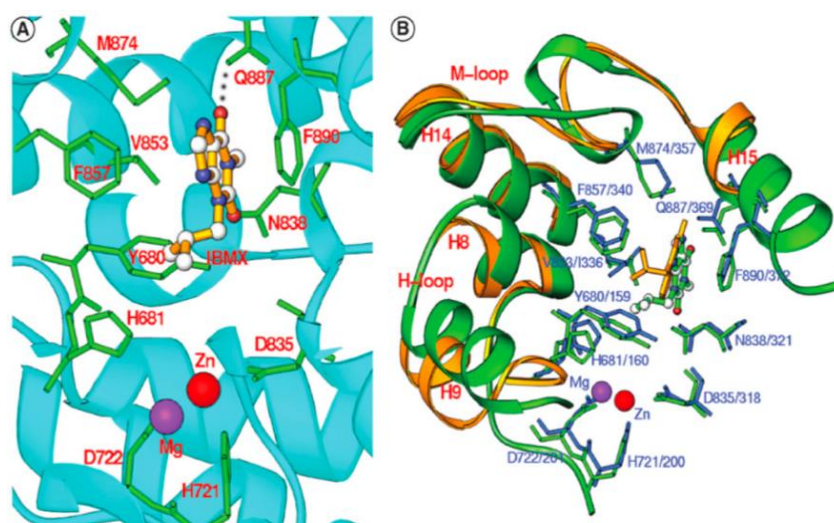
Figure 1.23: A) interaction of IBMX with L. major PDEB1 at active site. B) overlapping of parasitic (green) and human(gold/blue) PDEs and its interaction with ligand. *(73)*

1.15 The Main Aim of Current Thesis

In current thesis the main aim was to analyse parasitic PDEs from Leishmania species and T. brucei as well as to study human PDE5 drug – binding mode in more details. Since nowadays there are still no effective treatments to treat patients with current infectious this question is still under investigation whether we can develop selective inhibitors that will cause less side effects and would be more potent against parasitic PDEs. In the 3rd chapter the aim would be to express, purify and analyse parasitic PDED where no preliminary data was obtained before and only DNA sequence was available at that time. Therefore, it was important to find out if these proteins can also be considered as a potential drug target that play a vital role for parasite survival. The 4th chapter would be orientated on drug design approach where hPDE5 protein was used as a target. The aim was to develop a Fluoro fragment library and where strong binders would be identified not only by 1 technique but by several techniques, such as: Xchem, NMR, SPR and manual crystallization. Another aim in current chapter was also to try and use Xchem platform which is one of the latest development by Diamond Synchrotron. This is a high throughput method where high number

of crystals can be shot and analysed. The 5th and the 6th chapters were about TbrPDEB1 as a drug target where different biophysical techniques were used with 2 libraries, such as DSI-Poised and Chembl. Previous results showed that Xchem didn't identify many fragments as binders while other techniques did hence, one of the idea was to improve and develop general Xchem protocol that can be used for different proteins. Also, in a collaboration with VU university it was possible to synthesise new derivatives that would be more selective and possibly potent against TbrPDEB1.

CHAPTER 2

Materials and Methods

2.1 Introduction

Investigating protein structure, and its function with biophysical and biochemical techniques requires a significant amount of high quality protein that can be produced by recombinant over – expression followed by a series of purification steps, where the protein of interest is isolated from the mixture of other host proteins expressed in bacterial, insect or mammalian cells. The choice of the different methods used to express and purify proteins, is dependent on the nature and properties of protein.

2.2 Materials and Methods

In this chapter, the experimental techniques used in this thesis will be discussed as well as materials and suppliers that have been used.

2.2.1 Materials

2.2.1.1 Chemicals

Most chemicals, DEAE-sephacel and antibiotics were purchased from Sigma – Aldrich Ltd. Other materials were purchased from manufacturers as follows: IPTG and ampicillin from Melford Laboratories Ltd; disposable and empty PD10 columns from Amersham Biosciences; MiniElute^{TM} Gel Extraction Kit, QIAprep$^®$ Spin Miniprep Kit, and QIAquick$^®$ Gel Extraction Kit; restriction enzymes were purchased from Promega and from New England Biolabs, Inc; cloning vectors from Novagen; tryptone, yeast,

bacterial agar from Oxoid Ltd, QuickChange® II Site – Directed Mutagenesis Kit from Agilent Technologies; Roche FastStart High Fidelity PCR System from Roche Diagnostics GmbH.

2.2.1.2 Bacterial strains

Bacterial strains were purchased from Novagen, Invitrogen or Promega. Strains (or plasmids) were purchased or provided by Jane Munday (University of Glasgow) and Dr Susanne Schroeder (University of Kent).

Strains	*Genotype and/or phenotype*	*Description*	*Reference*
T7 Express E. coli	fhuA2lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet^S)2[dcm]R(zgb-210::Tn10--Tet^S) endA1Δ(mcrC-mrr)114::IS10	High efficiency competent cells suitable for T7 protein expression. There are no proteases Lon and OmpT, resistant to phage T1(fhuA2)	New England BioLabs® Inc.
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int:(lacI::PlacUV5::T7 gene1) i21 Δnin5*	Suitable for T7 expression, deficient in protease, free of animal products	New England BioLabs® Inc.
BL21(DE3) pLysS	F⁻ *ompT hsdS*$_B$ (r$_B$⁻, m$_B$⁻) *gal dcm* (DE3) pLysS (Cm^R)	Controlled by *lac UV5* promoter, deficient in proteases Ion and OmpT	Promega
Rosetta™ 2 (DE3)	F- *ompT hsdS*B(rB- mB-) *gal dcm* (DE3) pRARE2 (CamR)	Derivative of BL21, suitable for expression of proteins with rare codons, supply tRNA for 7 codons (AGA,AGG, AUA,CUA, GGGA, CCC and CGG)	Novagen
Rosetta™2(DE3) pLysS	F- *ompT hsdS*B (rB- mB-) *gal dcm* (DE3) pLysSRARE2 (CamR)	Express T7 lysozyme to suppress basal expression of T7 RNA polymerase	Novagen

Table 2.1: Bacterial Strains

Name	Plasmid	Description	Reference or source
pET15b	pET15bTEV_L. infantum_PDED	Promoters: AmpR, lac1 HisTag-N-terminus Ampicillin resistance	Novagen
	pET15bTEV_L. donovani_PDED		
pET28a	pET28a_Tbrucei_PDED	HisTag-N/C– terminus Kanamycin resistance	Novagen
	pET28a_L. infantum_PDEB1		
	pET28a_L. donovani_PDEB1		
pColdI	pColdI_L. infantum_PDED	Promoter: cold-shock protein A Induction at low temperature Lac operator Ampicillin resistance	Takara
	pColdI_L. donovani_PDED		
pColdTF	pColdTF_L. infantum_PDED	Chaperone TF (trigger factor)	Takara
	pColdTF_L. donovani_PDED		
pOPINF	pOPINF_L. infantum_PDED	Mammalian/bacterial/insect expression His-Tag N-terminus Ampicillin resistance	Protein Production UK
pOPINJ	pOPINJ_L. infantum_PDED	GST – fusion tag	
pOPINM	pOPINM_L. infantum_PDED	MBP – fusion tag	
pOPINS3C	pOPINS3C_L. infantum_PDED	SUMO – fusion tag	
pGKJE8	dnaK-dnaJ-grpE-groES-groEL		Takara
pGro7	groES-groEL		Takara
pKJE7	dnaK-dnaJ-grpE		Takara
pGTF2	groES-groEL-tig		Takara
pTf16	tig		Takara

Table 2.2: Plasmids that were used during this research.

2.2.1.3 Primers

A list of all primers used for amplification in this study can be found in Table 2.3. All primers were obtained from Invitrogen Life Technologies.

Protein	Forward Primer	Reverse Primer
L. infantum_PDEA_cd	AGCCATATGAGCCGAGCTGCGATATG	CCCCGGATCCCTACGAGTCGTCGTGGTTG
L. infantum_PDEB1_cd_1	AGCCATAATTCCGGCGTTGGCGAGAACATCATG	CCCCGGATCCTTAAACAATCG

L. infantum_PDEB1B2_cd_2	AGCGCTAGCGAAAACCTGTACTTCCAAGCGGTGACGCCGGAGGAGCGTGAG	CCTGGATCCTTAAACAATCGAGGATCG
L. infantum_PDEB1B2_cd_3	AGCGCTAGCGAAAACCTGTACTTCCAAGCGGAGCCGATGGATGCAGCT	CCTGGATCCTTAAACAATCGAGGATCG
L.infantum_PDED_fl	AAGTTCTGTTTCAGGGCCCGATGGGCAGCGGAAATACGAAGC	ATGGTCTAGAAAGCTTTAGTCCCTCACCTCCTCCACG
L.infantum_PDED_cd_1	GTATTTTCAGAGCCATATGGAGGCGTTCGAGATCTTC	CGGATAACAATTCCCCGGATCCCTTCTGCG
L.infantum_PDED_cd_2	GTATTTTCAGAGCCATATGGACCGCTGGGACTACGACACC	CAATTCCCCGGATCCGTCCTCTACCGACTC
L.infantum_PDED_cd_3	GTATTTTCAGAGCCATATGGACGACGCCGTACTGCG	CTTTGTTAGCAGCCGGATCCTTACTTCTGCGCCTCCTCC
L.infantum_PDED_cd_B	GTATTTTCAGAGCCATATGTGGGACTACGAC	CCTGGATCCTTACTGCGCCTCCTCCGTTTTG
L.infantum_PDED_cd_C	GTATTTTCAGAGCCATATGGGCATCGACAGGTGGGAC	CCTGGATCCTTAGGCCTCCTCGGTCTTCC
L.infantum_PDED_cd_JM_1	AAGTTCTGTTTCAGGGCCCGGAGGGCGTTCGAGATC	ATGGTCTAGAAAGCTTTACTTCTGCGCCTCC
L.infantum_PDED_cd_JM_2	AAGTTCTGTTTCAGGGCCCGGACCGCTGGGAC	ATGGTCTAGAAAGCTTTAGTCCTCTACCGACTCGG
L.infantum_PDED_cd_JM_3	AAGTTCTGTTTCAGGGCCCGGACGACGCCGTACTGC	ATGGTCTAGAAAGCTTTAGTCCTCTACCGACTC
PDEB1B2_cd_B	AGCGCTAGCGAAAACCTGTACTTCCAAGCGGTGACGCCGGAGGAGCGTGAG	CCTGGATCCTTAAACAATCGAGGATCG

PDEB1B2_cd_C	AGCGCTAGCGAAAACCTGTACTTCCAAGCGGAGCCGATGGATGCAGCT	AGCGCTAGCGAAAACCTGTACTTCCAAGCGCAGCGCAACAGCATTACG
L. donovani_PDED_fl	AAGTTCTTGTTTCAGGGCCCGATGGGCAGCGGCAACACC	ATGGTCTAGAAAGCTTTAGTCCCTCACCTCCTCCACG
L. donovani_PDED_cd	GTATTTTCAGAGCCATATGGAGGCGTTCGAGATCTTC	GCTGGATCCTTACTCCTGCGCCTCCTCCG
L. infantum_PDEC_cd	AGCCATATGCTAGGAGAGCTGCGTG	CCCCGGATCCTTATTGCTCCACCGCCTTCTCCTG

Table 2.3: List of primers that were used during current study

2.2.1.4 Commercial screens for DNA work

- PCR: Fast Start High Fidelity PCR System

- Kits: QIAprep Spin Miniprep Kit, Thermo Scientific GeneJet Gel extraction kit

- Restriction enzymes: Promega

- Agarose gel: Melford agarose high purity, SYBR safe DNA gel stain 10,000 x concentration in DMSO

2.2.1.5 Media and Solutions for Bacterial Work

Luria – Bertani (LB) broth:	Tryptone	10 g
	Yeast extract	5 g
	NaCl	5 g

Made up to 1 L with dH$_2$O and autoclaved.

Super LB broth:	Tryptone	32 g
	Yeast extract	20 g
	NaCl	5 g

Made up to 1 L with dH$_2$O and autoclaved.

Luria – Bertani agar:

15 g of Bacterial agar were added to 1 L LB broth before autoclaving.

The following, filter – sterilised supplements, were added prior to use:

1 M MgCl$_2$	12.5 mL
1 M MgSO$_4$	12.5 mL
20 % (w/v) glucose	20 mL

All solutions were made up separately and autoclaved before being mixed (filter sterilised), except antibiotics which were added post autoclaving.

Terrific Broth (TB):

Tryptone	12 g
Yeast extract	24 g
Glycerol	5 g

Additive	Stock Concentration	Final Concentration
MgSO4	1M in dH2O	2mM
Sucrose	1M in dH2O	25mM
Glucose	1M in dH2O	25mM
Glycerol	1M in dH2O	10%

Table 2.4: Additives List

Antibiotic	Stock Concentration	Final Concentration
Ampicillin	100 mg mL^{-1} in dH$_2$O	100 μg mL^{-1}
Kanamycin	50 mg mL^{-1} in dH$_2$O	50 μg mL^{-1}
Chloramphenicol	34 mg mL^{-1} in dH$_2$O	20 μg mL^{-1}

Table 2.5: Antibiotics List

2.2.1.6 Media and Solutions for DNA work

TE Buffer:

Tris – HCl, pH 8.0   10   mM

EDTA, pH 8.0   1   mM

6 x DNA Loading Buffer:

Bromophenol blue (w/v)   0.25%

Glycerol (v/v)   50%

TE Buffer (v/v)   50%

Molecular size marker Precision Plus Protein ™ Standards, Bio – Rad:

2.2.1.7. Solutions for Protein Purification

2.2.1.7.1 PDED catalytic domain protein purification

Standard Buffer for small scale expression:

Tris – HCl, pH 7.5	20 mM
NaCl	100 mM
Imidazole	10 mM
DTT	1 mM

- IMAC purification, Ni – NTA

Resuspension buffer:

Tris – HCl, pH 7.5	20 mM
NaCl	500 mM
Imidazole	20 mM
DTT	1 mM
Sucrose	25 mM
Glycerol	10 %
MgCl2	2 mM

Added protease cocktail (Roche)

Wash Buffer:

Tris – HCl, pH 7.5	20 mM
NaCl	500 mM
Imidazole	50 mM
DTT	1 mM

Elution Buffer:

Tris – HCl, pH 7.5	20 mM
NaCl	500 mM

DTT	1 mM
Sucrose	25 mM
Glycerol	20 %
MgCl2	2 mM

- *IEC purification*

IEC buffers: Buffer A, Elution buffer of IMAC purification

Buffer B, Elution buffer of IMAC purification + 1M NaCl

- *SEC purification*

SEC buffer: Elution buffer of IMAC purification

2.2.1.7.2 Solutions for T. brucei PDEB1 purification

- IMAC purification

Resuspension Buffer:

Tris – HCl, pH 7.5	20 mM
NaCl	200 mM
Imidazole	10 mM
B-mercaptoethanol	2 mM

protease inhibitor cocktail tablet x2 (Roche)

Buffer A: Resuspension Buffer deficient in protease inhibitor cocktail

Buffer B: Resuspension Buffer deficient in protease inhibitor, 300mM imidazole

Thrombin Cleavage Buffer:

Tris – HCl, pH 7.5	20 mM
NaCl	100 mM
Imidazole	10 mM
Glycerol	5%
B-mercaptoethanol	2 mM

- *IEC purification*

Buffer A:
$\qquad$ Tris – HCl, pH 7.5   20 mM

  NaCl   100 mM

  Imidazole   10 mM

  Glycerol   5%

  B-mercaptoethanol   2 mM

Buffer B: Buffer A + 1M NaCl

Storage Buffer: $MgCl_2$ 1 mM

  Tris-HCl, pH 7.5   20 mM

  NaCl   500 mM

  Glycerol   5%

  B-mercaptoethanol   2 mM

2.2.1.7.3 Solutions for PDE5 protein purification

Lysis Buffer: Tris – HCl, pH 7.5   50 mM

  NaCl   100 mM

  DTT   1 mM

  free protease inhibitor (tablet)

Blue Sepharose buffer A: Bis – Tris, pH 6.4   50 mM

  NaCl   50 mM

  DTT   1 mM

| Blue Sepharose buffer A, wash: | Buffer A+ 1M NaCl |
| Elution Buffer: | Buffer A + 20 mM cGMP |

SEC buffer:	Bis-Tris, pH 6.4	50 mM
	NaCl	500 mM
	DTT	1 mM

2.2.1.7.4 Solutions for protein acrylamide gels

| 10x Running Buffer: | Tris – HCl | 30g L^{-1} |
| | Glycine | 144 g L^{-1} |

2x SDS Sample Buffer:	0.5 M Tris – HCl, pH 6.8	6.0 mL
	Glycerol	4.8 mL
	SDS 10 % (w/v)	9.6 mL
	Bromophenol blue 0.05 % (w/v)	1.2 mL
	dH$_2$O	24.0 mL

Add 14 µL β-Mercaptoethanol per mL 2x SDS-Buffer.

| Fixing Solution: | Ethanol | 300 mL |
| | Acetic acid | 100 mL |

Made up to 1L with dH$_2$O.

Coomassie blue stain:	Trichloracetic acid (100 %)	250mL
	Coomassie blue R250	0.60 g
	SDS	0.10 g

Tris – HCl	0.25	g
Glycine	0.15	g

Made up to 500 mL

2.2.1.8 Composition of SDS – PAGE

SDS GELS						
Running gels	**8 %**	**10 %**	**12.5 %**	**15 %**	**Stacking gel**	**5 %**
dH$_2$O (mL)	5.7	4.7	3.4	2.2	dH$_2$O (mL)	3.4
30 % Acrylamide (mL) Acryl/Bis^{TM} 29:1	4	5	6.3	7.5	30 % Acrylamide (mL) Acryl/Bis^{TM} 29:1	1.5
1.5 M Tris-HCl, pH 8.8 (mL)	3.8	3.8	3.8	3.8	0.5 M Tris-HCl, pH 6.8 (mL)	1.9
10 % (w/v) SDS (mL)	1.5	1.5	1.5	1.5	10 % SDS(w/v) (mL)	0.75
10 % (w/v) APS (mL)	0.15	0.15	0.15	0.15	10 % (w/v) APS (mL)	0.075
TEMED (mL)	0.01	0.01	0.01	0.01	TEMED (mL)	0.01

Table 2.6: SDS gel Composition

2.2.1.9 Solutions for Western blot

Bio-Rad western blot kit was used, where blocking buffer was provided.

Transfer buffer: TBS buffer (50 mM Tris – Cl, pH 7.6; 150 mM NaCl)

Antibody: anti HisTag, Sigma-Aldrich

Substrate:

Stock:	Methanol	3 ml
	4 – chloronaphthol	3 mg

Working solution: 2.5 ml of substrate dissolved in TBS to a final volume of 25 ml, H$_2$O$_2$ (30%) 9 μL

2.2.1.10 Crystallization

There were 3 types of crystallization experiment used in this thesis: **Broad screening** using a TTP Labtech Mosquito nanoliter dispensing robot and commercial sparse matrix screens, where a variety of different conditions are explored in order to find the most suitable environment for the protein to form a crystal lattice. After, **manual optimizations** are performed, where conditions identified from the previous step are explored to optimize crystals using manual pipetting in microliter crystallization drops in 24 or 48 well Linbro plates by handing drop vapor diffusion, whereas **XChem ready crystallization** – is the specific robotic nanoliter optimization of conditions to produce large numbers of drops for the latest high through crystallography methods in low volume drops in SwissCI plates.

2.2.1.10.1 Commercial Crystallization Kits

MacroSol™ MD1 – 22

Morpheus® MD1 – 46

3D Structure Screen MD1 – 13

Stura FootPrint Screen MD1-20

SG1™ Screen HT-96 MD1-89

PACT *premier*™ MD1-29

Wizard Classic

JCSG-*plus*™ HT-96 MD1-40

2.2.1.10.2 Manual Crystallization, hanging drop method

Chemicals were used from Sigma-Aldrich, Melford and Fisher chemical

Plates: Greiner Bio-One ComboPlate 24-Well protein crystallization plate for hanging drop application.

2.2.1.10.3 XChem Plates

Plate type: Swiss CI XTAL (SD-3) 3 well midi

2.2.1.11 NMR Solutions

A Maybridge library of a selection of 589 fluorine containing fragment compounds was designed by the University of Kent and synthesized in Maybridge. 339 compounds had the Fluorine atom was attached to the aromatic ring, 3 compounds had Fluorine in a CF2 group and 247 compounds where Fluorine was in CF3 group. Fragments were shipped in powder form and then diluted in a d6DMSO to a concentration of 100μM. When mixtures were prepared the final ligand concentration in mixtures was 0.1 μM. For Fluorine fragment screening 1D NMR experiment was performed where 2 fragment mixtures were made, where mixture 1 was a control that didn't contain any protein and the second mixture was with 5μM of protein. In the final NMR screening set, there were 28 mixtures with 8-23 fragments in each mixture. In total 4 mg of PDE5 protein was used. The protein was in buffer that it was purified during SEC, Section 2.2.1.6.3. The NMR instrument that was used: 600 MHz five channel Bruker Avance III spectrometer with a cryoprobe which can observe a wide range of biologically relevant nuclei at high sensitivity and a 24 place sample changer. The spectrometer is used for structural biology, molecular dynamics measurements, drug binding and discovery campaigns, metabolomics and the analysis of small molecules.

2.2.1.12 SPR Solutions

All SPR experiments were performed with a Biacore T200 surface Plasmon resonance biosensor instrument (GE healthcare). Proteins were immobilized on CM5 series S sensor chips. Consumables were obtained from GE

Healthcare. The neutravidin immobilization was run on HBS-N at a flow speed of 10 µL/min at 25 °C. The matrix of the sensor chip was activated by injecting, on all flow channels, a mixture of 0.1 M *N*-hydroxysuccinimide (NHS) and 0.4M 1-ethyl-3-(3-(dimethylamino)propyl) carbodiimide hydrochloride (EDC) at a flow rate of 10 µM for 420 seconds. Subsequently, neutravidin (0.30 mg/mL) in a 10 mM NaAc solution (pH 5.0) was injected for 120 seconds. Unreacted activated groups of the dextran matrix were deactivated by injection of ethanolamine. HCl (1 M) for 420 s. Bovine Carbonic Anhydrase was purchased from Sigma-Aldrich. The proteins were buffer-exchanged using a Amicon Ultra 0.5 mL centrifugal filter (Merck). EZ-Link Sulfo-NHS-Biotin (Thermo Fisher Scientific) was diluted with the protein. Biotinylated proteins were desalted using 0.5 mL Zeba Spin desalting columns (Thermo Scientific).

2.2.1.13 Activity Assay

PDELight^{TM} HTS cAMP Phosphodiesterase Assay Kit, 500 Test

Lonza stimulation buffer (Reaction buffer):

- 50 mM Hepes
- 100 mM NaCl
- 5 mM $MgCl_2$
- 100 nM $ZnCl_2$
- 5% Glycerol
- Further pH adjusted with NaOH (10 M) to pH 7.8

2.2.2 Methods

2.2.2.1 Microbiological methods

2.2.2.1.1 Sterilization

Before any media or solutions can be used with cells, they were autoclaved to sterilize them. Autoclaving was performed at 121 °C for 15 minutes and 1 bar pressure. Temperature sensitive substances, such as antibiotics were added post autoclaving and then solutions were filtered through sterile filters (0.22uM).

2.2.2.1.2 Bacteria's Storage

When cells were transformed with the plasmid of interest it was important to be able to prepare bacteria for long-term freezer storage. The bacterial culture was prepared whereby inoculation with 1:1 ratio with glycerol at final glycerol concentration 50%. The culture was then stored at – 80 °C.

2.2.2.1.3 Bacteria Plates

Bacteria stored at –80 °C were thawed at RT for 10 – 15 minutes used where bacteria were streaked onto appropriate plate containing antibiotic. Agar plates were incubated overnight at 37 °C.

2.2.2.1.4 Liquid Cultures

Bacteria were inoculated in liquid culture where the appropriate antibiotic was added into the corresponding media. Aerobic cultures were shaking at ~ 200 rpm in a 1L flasks at 37 °C.

2.2.2.1.5 E. coli Competent Cells

Single bacteria colonies were inoculated into 100 ml of LB broth and grown overnight at 37 °C shaking at ~ 200 rpm to OD_{600} of 0.3. Next, the bacterial culture was cooled on ice for 15 minutes followed by centrifuging at 850 x g at 4 °C. The supernatant was discarded, and the

pellet was resuspended in solution containing 0.1M CaCl$_2$ and stored on ice for 30 min followed by centrifugation. The supernatant was discarded, and pellet was resuspended in 0.25 ml of 0.1M CaCl2 with 25% of glycerol (v/v). Cells were aliquoted into Eppendorf tubes of volume 50 µL and stored at – 80 °C.

2.2.2.1.6 Transformation of E. coli

The 50 µL aliquot of competent cells were thawed on ice for 10 minutes followed by addition of 1 µL of plasmid DNA. After, it was incubated on ice for 20 minutes. The mixture was then heat-shocked at 42 °C for 42 seconds. The cells were then cooled on ice for 2 minutes followed by the addition of 300 µL SOC broth and grown for 1 hour at 37 °C shaking at ~ 200 rpm. After cells were plated onto agar plate containing required antibiotic.

2.2.2.2 Protein Construct Design

DNA sequences were analysed by various software to identify site of catalytic domain boundaries in the full-length gene sequence.
Blast searches using blasttp algorithm (protein – protein BLAST) were used to compare sequence of interest with known sequences in the ensemble database and 3D structures in the protein data bank. The Pfam software was used to analyze sequences and show possible domain boundaries. The goal was to identify the best possible construct, including purification tags, that will potentially give the best levels of expression as well as the highest protein yield as possible with limited flexible regions to aid crystallization. There were certain additional factors that need to be considered, in particular the beginning and end of the construct, where the hydrophobic acids have to be avoided as well as

not terminating the construct in the middle of secondary structural elements.

2.2.2.3 Gene Amplification

2.2.2.3.1. Polymerase Chain Reaction (PCR)

PCR is a technique that used in genetics where the main aim is to amplify a gene of interest from the genome. The currently used technique was first introduced by Kary Mullis in 1983. The method is based on thermal cycling where DNA of interest can be amplified between 0.1 and even up to 40 kbp. The following figure 2.1

represents PCR cycle that will be explained later in more detail.
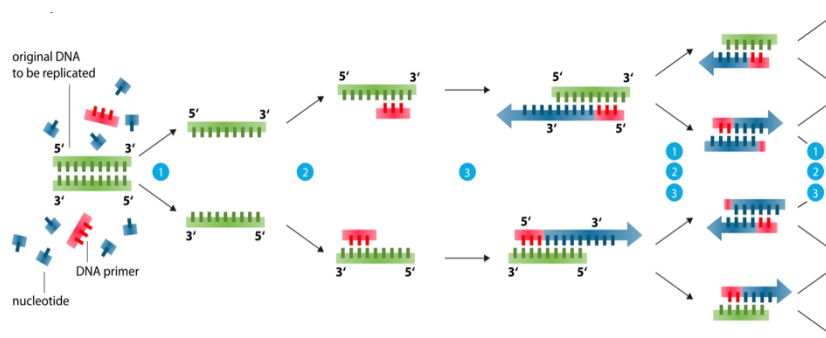
Figure 2.1: schematic representation of PCR, (Wikipedia)

There are several steps in thermal cycle to amplify DNA, Table 2.8:

• initializing – the reaction where reaction is heated;

• denaturation – where DNA disrupts (strands separation) by high temperature;

• annealing – temperature decreased down to 50-65 °C when primers attach to each end of the DNA segment;

• elongation – main function is performed by DNA polymerase where new amplified DNA strand is formed;

• hold – the reaction temperature decreased down to 4 °C.

	μL
ddH₂O	37.5 -32.5
10 x PCR buffer (containing 18 mM MgCl₂)	5
DMSO	0-5
5 mM dNTPs each	2
10 μM 5' primer	2
10 μM 3' primer	2
DNA template	1
***Taq* polymerase (5 U μL⁻¹)**	0.5

Table 2.7: PCR reaction composition

Step	*Temp*	*Time*	*Cycles*	*Function*
1	96 °C	2 min	*1*	Initial Denaturation of the chromosomal DNA
2	95 °C	30 secs	*35*	Denaturation of the amplified DNA
3	55-58 °C	30 secs		Annealing of the primers
4	72 °C	1 min per 1,000 bp		Elongation
5	72 °C	5 min	*1*	Final elongation
Stop	4 °C	HOLD		

Table 2.8: PCR reaction temperatures

2.2.2.3.2. Gel Electrophoresis

After the PCR reaction, the DNA product was loaded into agarose gel for analysis.

2.2.2.3.3 Agarose Gel

The agarose gel was prepared at concentration 1% (w/v) where agarose has been dissolved in 1xTAE buffer and an intercalating agent such as ethidium bromide with concentration being in solution of $\mu g\ mL^{-1}$. PCR products were mixed with loading dye of a final concentration being 20% (v/v) and loaded into agarose gel. Gel electrophoresis was run at 80V-100V for 1 hour.

2.2.2.3.4 Analysis of Agarose Gel

The agarose gel was analysed using UV light where the fluorescent dye (ethidium bromide) absorbs UV radiation of 320 nm and re-emits at 590 nm in a red-orange wavelength range. The dye is an intercalating agent that helps visualise the DNA fragments under UV trans illuminator. Gel pictures were taken through a red filter.

2.2.2.4 Isolation and Preparation of DNA Fragments for Cloning

2.2.2.4.1 Obtaining Product from Agarose Gel

Each DNA fragment has a size and DNA markers enable choice of the required fragment that can be cut from the gel by using a scalpel. The next step was purification of DNA where QIAquick® Gel Extraction Kit (Qiagen) was used. Protocol provided with the kit was followed.

2.2.2.4.2 Preparation for DNA Cloning and Ligation

Reaction Compositions	Amount (μL)
Insert	2.5
vector	1.5
2 x Rapid Ligation buffer (Promega)	5
T4 DNA Ligase (Promega 3 U μL^{-1})	1

Table 2.9: Ligation reaction composition

In order to perform a ligation of vector and DNA of interest, restriction digest have to be performed by same enzymes where complementary ends can be obtained. Ligation reaction has been performed at room temperature. The protocol is as follows:

Recognition Sequence	Restriction Enzyme
A/AGCTT	HindIII HindIII-HF®
A/GATCT	BglII
CA/TATG	NdeI
CCC/GGG	SmaI
CCGC/GG	SacII
CCTNN/NNNAGG	EcoNI
G/CTAGC	NheI NheI-HF®
G/GATCC	BamHI BamHI-HF®
TTAAT/TAA	PacI
T/CTAGA	XbaI

Table 2.10: Restriction Enzymes List, (ProMega, BioLabs New England)

2.2.2.4.3 Plasmid Isolation

DH10B™ *E. coli* cells were transformed with plasmid of interest. After 12 hours of growth plasmid was isolated by using QIAprep® miniprep kit (Qiagen). The preparation was performed following miniprep kit handbook.

2.2.2.4.4 Accessing Plasmid Integrity, Restriction Digest

In order to check plasmid integrity and ensure there was no mutation in gene of interest, it was important to perform restriction digests where enzymes cut the sequence at specific sites. The reaction was prepared according to Promega information and incubated at 37 °C for 2 hours. The digest protocol can be found below:

	Single digest	Double digest
ddH$_2$O	4 μL	4 μL
Restriction enzyme 1	1 μL	0.5 μL
Restriction enzyme 2	-	0.5 μL
10 x Buffer	1 μL	1 μL
Plasmid DNA	4 μL	4 μL

Table 2.11: Digest Protocol

2.2.2.5 Cloning

2.2.2.5.1 Cloning into *pET E. coli* T7 Expression Vectors

Vectors and amplified DNA fragments were digested at specific restriction sites where specific sequences have been introduced into 5' and 3' primers. After the gene of interest has been purified from the agarose gel and the vector has been digested, a ligation reaction was performed, Section 2.2.2.3.2. Competent *E.coli* DH10b cells were

transformed with ligated plasmid and plated onto LB agar plates containing appropriate antibiotic such as: ampicillin or kanamycin. Colonies were grown overnight after which 1 colony was inoculated into the liquid broth to amplify plasmid of interest. The next day, the restriction digest was performed followed by sequencing the plasmid by using Genewiz service to verify the DNA sequence. pET15b was modified in order to introduce TEV cleavage site by ThermoFisher Scientific (Lifetech). The gene of interest can be tagged either on N-terminus or C-terminus with a 6 – histidine residues tag to aid identification and purification of any expressed protein.

2.2.2.5.2 Cloning into pOPIN vector suite

The pOPIN vector suite allows expression of the gene of interest in multiple hosts: bacteria, insect and mammalian cells. As well as providing versatile vectors for expression recombinant proteins it also provides vectors with fusion tags, such as SUMO, GST, MBP and etc., that enhances protein folding. There were 4 vectors that were used: pOPINF, pOPINM, pOPINJ and pOPINS3C. DNA of interest was directly cloned into the digested vectors by Sequence Ligation Independent Cloning (SLIC). The protocol can be found below:

	Stock concentration	Volume added	Final concentration
Linearized vector (e.g., 5 kb)	100 ng/μl	1 μl	10 ng/μl
Insert 1 (PCR product, e.g., 1 kb)	40 ng/μl	1 μl	4 ng/μl
Insert 2 (PCR product, e.g., 1 kb)	40 ng/μl	1 μl	4 ng/μl
10X BSA		1 μl	1X
10X NEB Buffer 2		1 μl	1X

Table 2.12: SLIC protocol reagents, *(77)*

2.2.2.5.3 Cloning into pCold expression vectors: pColdI and pColdTF

This expression system uses a cold shock expression method. There are several different vectors: pColdI and pColdTF. PCR product of catalytic domain was cloned into a digested vector with appropriate restriction enzymes where complementary ends were formed. The ligation method was followed as given in section *2.2.1.4.1.*

2.2.2.6 Protein Expression and Purification

2.2.2.6.1 Production of recombinant protein in E. coli

E. coli cells were transformed with the recombinant plasmid of interest and transformed onto a plate containing the appropriate antibiotic. The following day, 1 colony was inoculated into 20 ml starter culture and left growing overnight at 37 °C, shaking at ~ 200 rpm. The next day, the starter culture was used for inoculation 1L 2YT or TB broth with antibiotic and grown at 37 °C shaking at ~ 200 rpm until the OD_{600} reached 0.6-0.7. When the culture reached the required OD_{600} value the cells were chilled on ice for 30 minutes after that an induction agent IPTG was added, varying in concentrations from 0.2 mM to 0.5 mM to induce protein expression. In studies with co-expressing chaperones 2 other inducers were added, such as arabinose and tetracycline. Cells were expressing the protein of interest from 12 hours to 48 hours at different temperatures: from 19 °C to 21 °C. After, cells were centrifuged at 3500 x g for 20 minutes at 12 °C. The supernatant was discarded, and pellet was resuspended in appropriate buffer in 1:10 (w/v) ratio.

2.2.2.6.2 Cell Lysis

2.2.2.6.2.1 Sonication

Harvested cells were disrupted by sonication (Sonics Vibracell Ultrasonic processor) for 6 times 45 seconds each round and followed by 30 seconds break after each round. After, cells were centrifuged at 35,000 x g for 25 minutes at 4 °C.

2.2.2.6.2.2 Cell Disruption

Harvested cells were resuspended in appropriate buffer and passed through a cell disrupter (Stansted Fluid) that was operating in range of 12,000 – 14,000 psi. The sample was then centrifuged at 35,000 x g for 30 minutes at 4 °C and the supernatant retained for gel analysis and protein purification.

2.2.2.6.3 Protein Purification

2.2.2.6.3.1 Immobilized Metal – Affinity Chromatography (IMAC)

The main principal of protein separation from a mixture of other proteins. There are several types of affinity chromatography: affinity chromatography, IMAC and tagged protein purification (e.g. Glutathione: GST fusion approach). The overall process can be characterized as competitive binding where the folded protein of interest has a freely available additional tag that allows it to bind to a specific resin while contaminant proteins flow through the column without any interaction with the resin. The bound protein of interest is then displaced from the column by eluting with an appropriate ligand. In the current study metal affinity chromatography was used where a HexaHis-Tag has been fused with the recombinant protein on its N/C-terminus in order to bind to

divalent nickel metal in resin. Ni-NTA resin uses nitrilotriacetic acid (NTA) as well as tetradentate chelating ligand where NTA binds to Ni ions by forming 4 interactions. His-Tag contains polyhistidine residues that has a strong binding affinity to metal ions. Imidazole ring of histidine forms bonds by its electron donor groups with metal ion. As a result, the protein of interest will interact with immobilised metals while the contaminants are flowing through the column. To elute protein competitive binder is used, such as imidazole which at high concentration substitute protein and protein of interest elutes or changing of pH of the column buffer *(74).*

IMAC can give 95% purity if there is a lot of protein of interest in mixture. It has also been shown that IMAC could be used when different expression systems have been used to express a protein of interest such as: *Escherichia coli,* mammalian cells and baculovirus infected insect cells.

2.2.2.6.3.1.1 Fusion His-Tag with Recombinant protein

In some experiments, it is possible to use longer tags with higher number of His residues if purification method required more vigorous washing before eluting the recombinant protein. The main reason is to improve the purity and final protein yield. The protein constructs chemistry and its region of interest dictates where the tag has to be placed, it can be either on C terminus or N terminus, as folding can inhibit protein metal interaction where His – Tag will be buried inside the protein so it won't be recognised by divalent metal. Sometimes protein tags can interfere with the recombinant protein functioning and folding, therefore it is advisable to use short purification tags in order to overcome such problem. His – Tag was incorporated into the expression vector by

polymerase chain reaction (PCR) as it has been performed in current study as well as directed mutagenesis.

2.2.2.6.3.1.2 **IMAC resin**

Initially iminodiacetic acid (IDA) was used as a matrix that interacts with divalent metals by 3 sites. However, there was a main disadvantage that made using this matrix as less preferable. Metal ions have a weak interaction with the resin hence resulting in decrease in purity as well as protein yield. The latest development of new matrix with using nickel nitriloacetic acid and cobalt carboxymethyl aspartate that are linked to resin. In comparison to previous matrix this 2 interacts with divalent metals by 4 sites while leaving the other 2 available for recombinant protein to attach by its His-Tag, Figure 2.2.

a. Nickel-nitriloacetic acid (Ni^{+2}-NTA)

b. Cobalt-carboxylmethylaspartate (Co^{+2}-CMA)
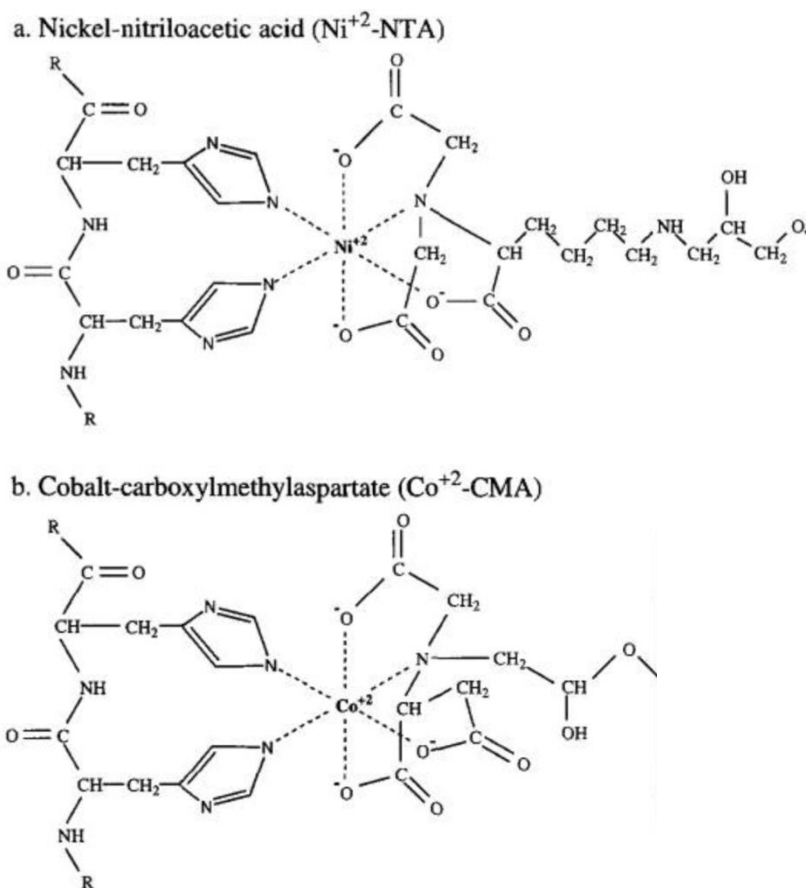
Figure 2.2: Formation of bonds between His-Tag and divalent metal ions. A) Ni-NTA *(75)* B) Co-CMA *(76)*

One of the main advantages of using such a robust system is that it can tolerate a wide variety of conditions where different buffers can be used as well as the presence of detergents. Another benefit of using IMAC is the fact that resin can be reused many times as long as it is cleaned and regenerated properly so the efficiency of protein purification won't be affected. The binding efficiency of Ni-NTA is 5-10 mg/ml of recombinant protein with affinity being Kd=10^{-13} M whereas Co-CMA has lower binding affinity however higher level of specific binding resulting in purer elute.

2.2.2.6.3.1.3 Protein Purification Process

There are two types of purification methods that can be performed by using IMAC technique: purification under native and denatured conditions. Native conditions are used when imidazole in elution buffer as a competitive binder where active proteins are eluted. If protein doesn't experience any misfolding and aggregation this method can be used. However, in the case of purifying protein that exhibit some difficulties in folding and forming aggregates it is advisable to use denatured conditions. However, proteolytic and phosphatase enzymes functioning can be affected *(77).* Later on, the misfolded proteins can be refolded by using dialysis although this is often a difficult and low yielding procedure *(78).* In current study, native gel was used.

2.2.2.6.3.1.4 IMAC purification process

There are two ways how IMAC can be performed, either by batch incubation or prepacked column. In current study, prepacked column was used in all experiment's protein purification. Since the resin has a high affinity to His-Tag there is a minimum amount of resin required to successfully purify protein. In current experiment 5 ml column were used. Such chemical additives as sodium chloride and imidazole can be used in binding buffer to reduce nonspecific bindings of contaminant proteins that don't have His-Tag incorporated into their protein structure. When the lysate has been loaded into the column containing the protein of interest, further step of wash can be performed where imidazole concentration can be increased while not disrupting the interaction between divalent metal ions and His-Tag of protein. The last step is the elution where 2 different methods are used: lower pH down to 4.5-5.3 and using competitive binder with high concentration so the interaction

between the tag and the metal will be broken and protein of interest will be eluted. In the current study, high concentration of imidazole was used as lowering pH can damage the protein and our main focus was to work on a completely novel target and protein that hasn't been expressed and purified before. It is also important to maintain a buffer condition where no chelating agents can interfere with metal ions as the divalent ions will be stripped out of the column resulting in decrease of protein purity *(79).*

2.2.2.6.3.2 Protein Purification According to its Charge and Hydrophobicity

There 2 types of purification: hydrophobic based qualities purification and ion exchange chromatography (IEC). The first method is based upon protein separation according to its hydrophobicity where hydrophobic interaction chromatography media contains both hydrophobic and hydrophilic regions. Contaminant proteins are aggregated during the purification and as a result increases the purity of protein of interest. The second method that was used in the current study is IEC where proteins are separated according to its ionic charge. This method has been introduced in 1960 whereas prepacked columns are in use where resin is positively charged to separate negatively charged protein of interest which are also called anions. While negative charged resin is used to separate positively charged proteins that are also called cations. The technique is highly popular since it is useful for purification not only of proteins but also small peptides and nucleic acids as long as the molecule has a net ionic charge. IEC can still be used even if there is a very small difference in charge between proteins in mixture. Another advantage of this technique is a high loading capacity.

2.2.2.6.3.2.1 Principle of Separation

The main principle of protein separation is based on isoelectric focusing of protein where isoelectric point (pI) plays a crucial role in separation protein of interest from the contaminants. pI is defined as the pH when the protein of interest does not have any net charge. pH > pL identifies as the condition when protein has net negative charge and pH < pI when protein has a net positive charge. The protein charge is directly depending upon specific groups in protein structure that are responsible for net charge of the protein while also having different pKa numbers which are acid ionization constants. pH plays a crucial role where proteins that contain in its structure basic and acidic groups will change their charge with the change of pH. Such proteins called amphoteric. Titration curve shows how the charge of protein relates to the pH of the environment, Figure 2.3. From the curve in this Figure it can be seen that change of pH directly correlates with change in charge of protein. Increase of pH value leads to decrease in net charge to more anionic form.

This relationship between pH and net charge of protein enabling IEC to be very useful technique to elute protein of interest. As well binding according to protein charge other interactions can take an action, such as van der Waals interactions and small nonpolar bonds.

2.2.2.6.3.2.2 IEC Purification

At the beginning the prepacked column has to be equilibrated in the same buffer as your sample is to equilibrate the stationary phase. In the current study sodium chloride was used with 100mM concentration as a starting buffer. It is very crucial for protein to be exactly in the same buffer with 100mM salt to ensure that the protein will bind and not flow through the column with other contaminates.
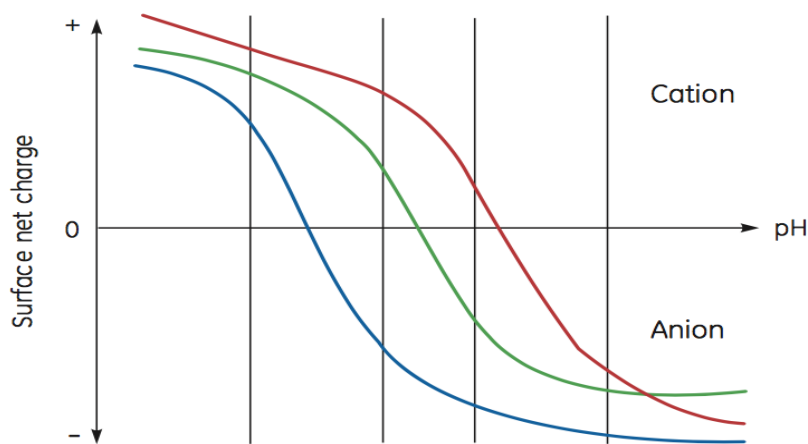
Figure 2.3: Titration curve. Figure represents 3 theoretical protein titration curves where each curve shows how the net surface charge of the protein is changing due to pH change.

After the protein mixture is loaded into the column where the protein of interest will bind to the column and contaminants will flow through. As it has been mentioned earlier opposite charged molecules will bind to the column media thus increasing the amount of recombinant protein of interest in the column. The 3rd step is washing of column with higher concentration of salt (1 M) in order to remove non-specific binding which will ultimately result is purer elution. The very last step and the most important in the whole process is the elution itself where the ionic strength will be increased by increasing the concentration of sodium chloride where the sodium ions will act as some competitive binders with protein of interest leading to disruption of interaction between protein and media resulting in protein of interest being eluted. There are two types of elution: step elution where buffer with high concentration of salt is applied directly into the column and gradient elution where salt concentration is increasing gradually resulting in purer sample as by knowing the ionic strength of your protein we can predict the approximate percentage of salt that is required to elute exactly that protein. After the purification column has to be washed and regenerated

so the next protein run can be performed. The schematic representation of the whole process can be seen in the following diagram, Figure 2.4.
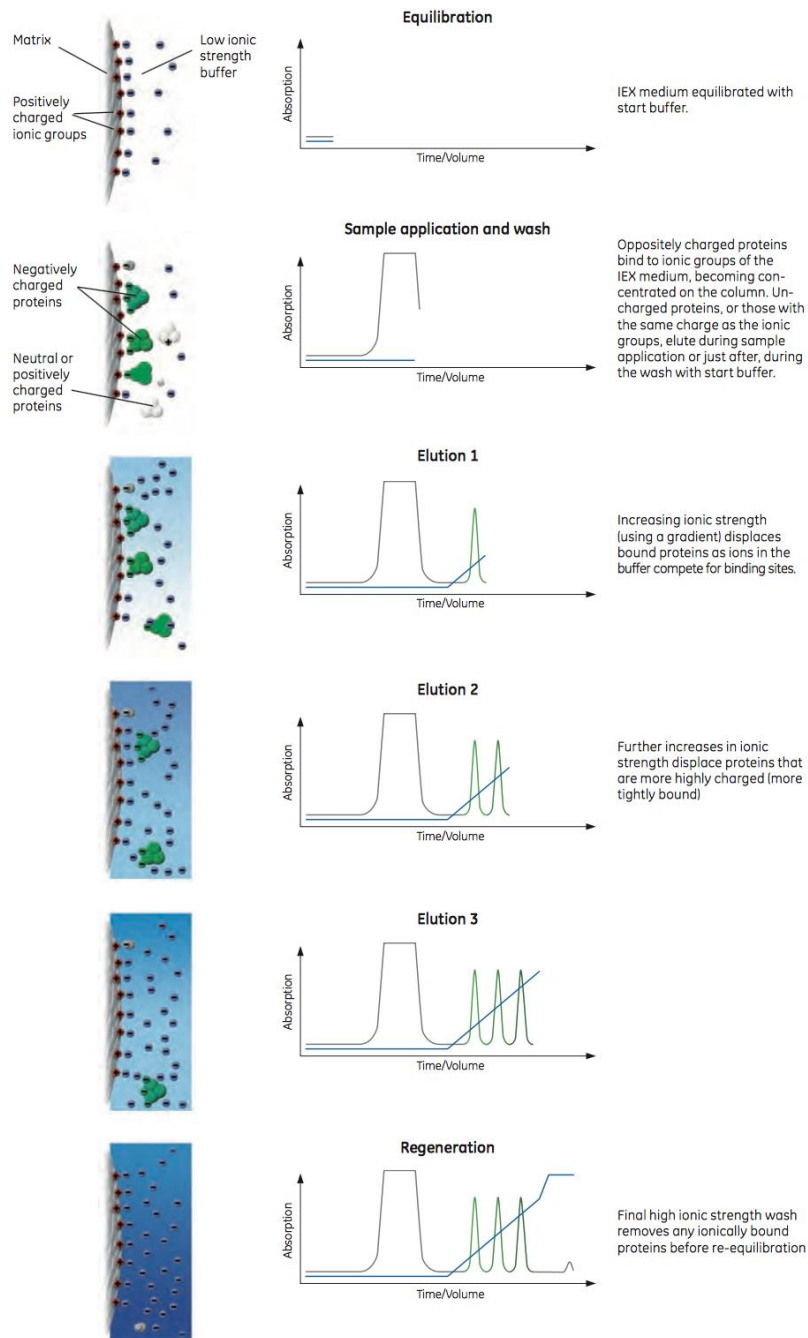
Figure 2.4: Schematic representation of anion exchanger, GE Healthcare manual

2.2.2.6.3.2.3 Factors Affecting IEC Purification

I. Resolution – tells us how well the protein is separated from other contaminants in the mixture. This can be visualised by how the peaks are separated on the chromatogram. The description for resolution is the

width between to elution peaks in contrast to the distance from point A to point B of each peak base.

II. Efficiency – is described as how well a prepacked column can separate recombinant protein of interest from other contaminants that are present in the mixture. Efficiency directly correlates with column packing where any alteration is column zones will change its efficiency. The main reason that leads to loss of efficiency, is when a column is continuously flushed with mixtures full of molecules like proteins or peptides, so it is very important to clean and regenerate columns properly after each use. Another possible reason for reduced efficiency is the quality of column packing where uneven distribution of matrix can lead to poorer protein purity.

III. Selectivity – is described as how selective the column matrix is to the protein of interest. Many factors influence it such as environment pH, how strong protein is bonded with matrix, and buffers that are used to elute protein. pH plays a crucial role in selectivity since it dictates the differences in protein charge and other contaminants. As it was mentioned above, step elution is used when a high concentration of salt is directly applied into the column to elute the protein. This method is used if the protein PI is known so the pH and ionic strength you expect it to elute at can be estimated. Gradient elution is useful when the target has not been purified before, to help establish the elution conditions that are required for that particular protein. The principle of the method is that the pH and salt concentration is slowly increasing so the molecules with weakest ionic strength interactions will elute, first followed by other proteins that have stronger interactions with the media. This method is a useful additional step if IMAC chromatography did not yield high purity of sample.

2.2.2.6.3.2.4 IEC media

The column is packed with a special matrix that enable purification of protein according to its net charge. The properties of IEC media allow them to have a charge opposite to theprotein of interest and form ionic interactions. There are many pores, to enable high binding affinity and a large surface area, for the required protein to bind. The inherent physical strength of media itself makes it as a safe component to use in purification, as it keeps a constant volume (no compression under pressure) thus the elution step won't be disrupted. Also, the process is fast and with the constant flow rate. This can minimise errors and hence improve the accuracy of the whole experiment. There are several types of media that can be used in packing the column. In the current study the column was prepacked with sepharose beads where agarose forms interactions with each other at various cross linking to enable efficient purification. The first stage in IEC is when the column is equilibrated at a particular pH and ionic strength to enable protein of interest to bound. The exchanger groups are linked to oppositely charged ions, such as chloride or sodium. This stage is followed by injecting the sample where protein of interest replaces the charged ions and binds to the resin. This is followed by washing the column with the starting buffer where unbound contaminants can be eluted form the column. The final step is the elution of the desired protein where buffer composition is slightly changed in order to replace protein with opposite charged ions hence eluting protein. This happens by increasing ionic strength of buffer by increasing the concentration of salt or by changing pH of buffer.

2.2.2.6.3.3 Size Exclusion Chromatography (SEC)

A third method that has been used extensively during this thesis is separation of proteins according to its size and shape where prepacked column has porous gel that according to protein masses will elute proteins at different time. This technique has been firstly proposed by Synge and Tiselius *(80)*, where scientists recognised the pattern when smaller molecule have an ability to enter pores while the proteins with larger molecular weight tend to pass through quicker. The difference between SEC and IE or IMAC is that the chemical environment doesn't play any role in SEC. SEC is similar to the previously described purification approaches with the first step being column equilibration, followed by injection of the protein of interest which is then eluted by using the same buffer as the one used for equilibration. The key property of column that is that the matrix is formed of spherical beads. Large molecules will elute first as they cannot enter the pores of matrix due to their size. This is the reason that protein aggregates can easily be purified out from the sample. Smaller folded protein molecules will elute last as they can enter pores, so it takes longer for lower molecular proteins to travel down the column. SEC is a useful technique for purification of recombinant protein of interest, analysis of protein aggregation and also buffer exchange. Usually SEC is the last step in protein purification as this method yields a protein of high purity and removes aggregates.

Separation - before performing any biophysical studies, such as X-ray crystallography or NMR, it is important to see that protein is in homogenous state and there are no contaminants and aggregates that will interfere with the experiment.

Buffer exchange – different proteins require different buffer environment in order to be used further studies or for storage purposes. The "gentlest" way to buffer exchange the protein of interest is to use SEC column.

2.2.2.6.3.3.1 SEC Optimization

There are certain factors that can alter protein purification. Recombinant proteins can form ionic bonds with columns mobile phase that ultimately can change elution results due to shift in retention time, peak being not able to distinguish between aggregate and target protein *(81, 82).* Also, there are certain physical factors that can influence protein elution, such as column volume, flow rate to load the column with the protein mixture and its volume.

2.2.2.7 Protein Detection and its further Characterization
2.2.2.7.1 Western Blot

Protein samples were run on standard SDP – PAGE. Gels were then placed between transfer membrane and paper. An electric field is used to transfer protein from the SDS – PAGE onto the membrane. When the electric field transfer was done the membrane was placed into a tray where it was incubated with milk powder for 2 hours at 4 °C. After incubation, the membrane was washed 2 x 10 mins with TBS – Tween buffer to remove any milk solution left. The next step was to incubate the membrane with the primary antibody in dilution of 1:2000 in 10 ml of TBS buffer. The membrane was left shaking for 1 hour at Room Temperature. After incubation, the membrane was washed with TBS – Tween buffer to remove all antibodies. Simultaneously the substrate was prepared where 6 ml of methanol was added to 4 – chloronaphthol from the freezer. This mixture was then added to TBS – Tween where hydrogen peroxide (H_2O_2) 25 µl was added into the solution as the last additive. The substrate was added and was agitated on a rocker until stained bands would appear (2-10 mins).

Further Analysis of sample – a number of analysis techniques can be used to assess the final protein sample e.g. mass spectrometry, multangular light scattering, Circular Dichroism (CD) and small angle scattering (SAXS) which can reveal more details on secondary structure, post translational modifications, shape and aggregation

2.2.2.7.2 Enzymatic Activity Assay

PDEs in combination with adenylate cyclase controls cAMP second messenger signalling. The main function of PDE is to hydrolyse cAMP to adenosine monophosphate (AMP). PDELight™ is a commercial assay which uses a bioluminescent method to detect hydrolysis of cAMP. A detection (AMP-DR) reagent is used to identify the breakdown of cAMP into AMP. AMP-DR directly converts AMP into ATP. The luciferase enzyme uses light that is emitted from ATP and luciferin, Figure 2.5.

$$cAMP \xrightarrow{\text{PDE}} AMP \xrightarrow{\text{AMP-DR}} ATP + Luciferin + O_2 \xrightarrow{\text{Luciferase}} Oxyluciferin + AMP + PPi + CO_2 + LIGHT$$
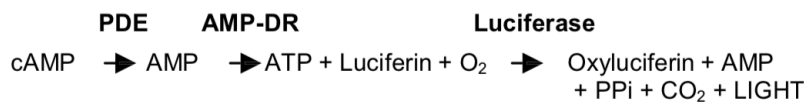
Figure 2.5: Schematic representation of cAMP hydrolysis, Lonza manual

Light intensity is directly correlated with the amount of AMP being formed. Therefore, according to the light intensity it is possible to link it to the PDE concentration and activity in the reaction.

The following protocol was provided by the University of VU.

I.   Storage of PDEs and preparation of reagents

PDEs   Thawed aliquots on ice were diluted with Lonza Stimulation Buffer (S.B.) (50mM Hepes, 100mM NaCl, 5mM $MgCl_2$, 0.1mM $ZnCl_2$, 5% glycerol, volume adjusted with MiliQ, pH 7.8) to a

suitable working concentration (section 2). Kept on ice. 2.5µl/well required.

cAMP

Thawed 10mM aliquot(s) were diluted with Stimulation Buffer to required concentration. Kept on ice, to allow to reach R.T. 15mins before use. 5µl/well required.

AMP control

Thawed 10mM aliquot(s) were diluted with Stimulation Buffer to required concentration. Kept on ice, allowing to reach R.T. 15mins before use. 5µl/well required.

Detection Reagent

In the dark, sufficient number of aliquots were thawed and diluted to 80% with Reconstitution Buffer. 5µl/well required. Kept in the dark on ice, allowing to reach R.T. 15mins before use.

II.   Determining a suitable working PDE concentration.

Prepared PDE dilution series in S.B.: 1:10, 1:100, 1:1000, 1: 10,000 and 1: 100,000.

In triple (reaction volume 15µl):

2.5µl of diluted PDE

Spin plate

+

2.5µl S.B. containing 6% DMSO

Spin plate, cover and incubate at room temperature for20min

<div align="center">

+

5µl Detection Reagent (diluted to 80% with R.B.)

Spin plate

+

5µl cAMP at 10µM (cAMP is $^1/_3$ final volume so prepare 30µM)

Note time '0' and Spin plate

</div>

Controls:

1. AMP - 5µl of cAMP were replaced with 5µl AMP (at the same concentration i.e. 10µM)

2. No PDE - 2.5µl of PDE were replaced dilution 2.5µl of S.B.

2.2.2.7.3 X – ray Crystallization

2.2.2.7.3.1 Introduction

Crystallization by definition as a separation biophysical technique where solid phase is separated from a mother liquor and where crystals occur with an ordered internal arrangement of molecules, ions or atoms *(83).* There are 3 steps in crystallization process: nucleation where crystalline condition is appearing, growth and growth cessation. Different particles crystallize at different speed and some such as protein can be difficult to crystallize. There are two types of forces that directly influence crystallization: atomic and intermolecular forces. Types of crystallization can vary where several methods are existing in order to reduce solubility: cooling, antisolvent additions, evaporation and reaction (precipitation). For a protein to precipitate the equilibrium has to be transformed into a supersaturated state where the amount of energy that was obtained during entering that state can be released during formation of precipitation, Figure 2.6.
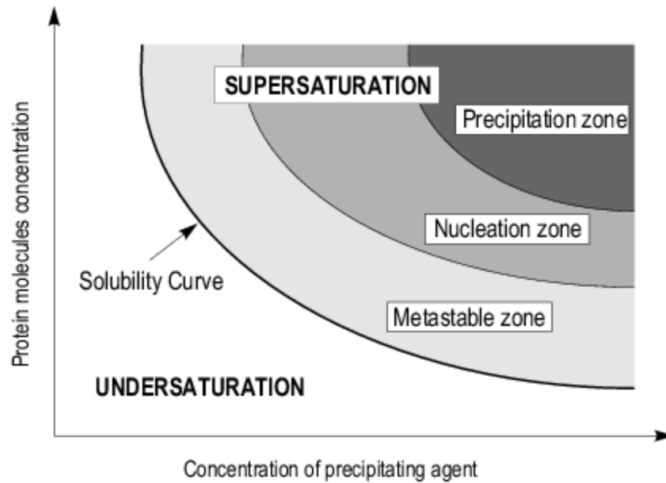
Figure 2.6: Crystallization phase diagram. *(86)*

There are certain techniques that can be used to transfer solution into a supersaturated state, such as: batch, vapour diffusion and liquid – liquid diffusion. Vapour diffusion is the most common technique in use where sitting drop (VD – SD) or hanging drop (VD – HD) techniques are in use. This technique is characterized by mixing protein with the precipitant in 1 drop and that is equilibrated against a reservoir mixture that is double in concentration of the precipitant mixture. During equilibrium water, leaves the protein drop into the reservoir by using air – gap separations and leading to a gradual increase in concentration of protein – precipitant mixture thus resulting in supersaturated state.

The batch crystallization method is described as an under – oil method where protein of interest is mixed with the precipitant mixture and is covered by a paraffin oil *(84).* One of the main disadvantages is the problem of the sample drying out due to the fact that plastic is permeable to water molecules. By using capillaries and/or membranes excess distribution mixture can be reduced hence resulting in decrease of mass transport. This method is called free interface diffusion method (FID) where the protein and the precipitant are in state of equilibrium before protein crystallization *(85).* Another technique that is also used is known

79

as batch crystallization which is the counter – diffusion method (CD) where the system is moved towards supersaturation instead of equilibration *(87).* In current study VP – SD method was used.

2.2.2.7.3.2 Crystallisation Screening

To obtain a crystal structure of a novel target it is normal to first use commercial sparse matrix screens that contain a wide variety of different conditions that are known to have previously given crystals. There are a variety of screens obtained from Molecular Dimension where used on the Mosquito robot. The main advantages of using the Mosquito are low protein consumption, pipetting accuracy and speed. The crystal plate is then stored at different temperatures for the best suitable environment for the protein to crystallize. Usually it takes at least 24 hours for nucleation to appear, however it all depends upon the nature of protein and crystallisation conditions where some proteins can strongly precipitate and yield no crystals and others simply remain in solution. To analyse crystallisation drops a stereo light microscope is used. Different phases can be observed. The main difficulty in choosing the most suitable condition is the identification of conditions that after manual optimization will produce crystals suitable for X-ray diffraction if there are no obvious crystals in the initial trials. After analysis of the crystallisation drops, results can be plotted in a phase diagram with different zones according to protein solubility. There are 9 scores whereas each represents protein crystallization by analysing protein's drop quality.

Score	Protein State	Description
0	Protein wasn't precipitated	- Drop remained clear for 2 weeks - Concentration of protein had to be increased
1	There were non-protein particles	- Bacterial growth - Fibres from clothes
2	Small precipitation	- Protein precipitates
3	Strong precipitation	- Unfavorable conditions - Decrease protein concentration -
4	Gelatinous precipitation	- White/transparent - Conditions must be optimized
5	Phase separation	- Droplets in drop with oil skin - Temperature is important
6	Spherulites formation	- Transparent clusters - Decrease protein concentration - Seeding application
7	1D crystal formation	- Needles formation - Optimization required
8	2D crystal formation	- Optimization: seeding, addition of additives
9	3D crystal formation	- Crystal > 0.5mm

		- Check for salts crystals

Table 2.13: Crystallization phases diagram, Molecular Dimensions Manual

$$[ppt]_{drop} = \frac{[ppt]_{reservoir}}{2}$$
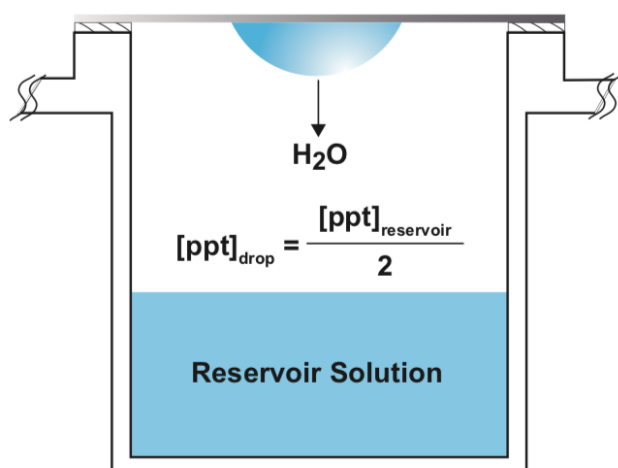
Figure 2.7: Hanging drop, Vapour Diffusion Method. Hampton Research, solutions for crystal growth manual

2.2.2.7.3.3 Manual Screening

In the current study the hanging drop vapour diffusion method was used where protein is mixed with precipitant solution and the mixture is placed in vapour equilibration where each drop was placed with a reservoir containing the same precipitant. 500 μl of reservoir was used where the drop concentration was lower where the volume was 2 μl or 4 μl in protein to precipitate ratio 1:2, 2:2 or 2:1. Equilibration occurred by water leaving the protein drop and diffusing back to the reservoir. This result in increase of supersaturation where the concentration of both protein and precipitant increases leading to reaching the equilibration.

2.2.2.7.3.4 XChem

At Diamond Synchrotron, there is a new X – ray development that enables users to perform high through put crystallography experiments where up to 1000 compounds can be screened individually in less than a week. The entire process includes soaking of crystals, harvesting and data collection. Data analysis can be performed on site or remotely through a NoMachine interface. Another advantage of using XChem platform is data analysis pipeline where PanDDA map analysis allows users to identify even very weak binders such as fragment compounds bound to the target protein.

The 1st step in preparation of an XChem experiment is to purify protein of high purity where protein is in homogenous state in order to produce high quality protein crystals that will sustain soaking with high concentration fragment long duration soaks.

The 2nd step after protein was purified is to crystalize it in PSI plates that are suitable for ECHO – dispensing in Diamond Synchrotron. Protein plated by mosquito using specifically designed protocol for that.

The 3rd step was to choose library of interest, where 2 libraries were used: DSI – Poised (provided by Diamond), Fluoro – fragment library (provided by Maybridge) and also a selection of fragments from ChEMBL.

The 4th step was to perform soaking using the ECHO dispensing system that was located on site.

The 5th step was to harvest crystals by using microscope to visualize it and loops for fishing it.

The 6th step was data collection where crystals were screened individually automatically.

The 7th step is analysis of collected diffraction data where XChem explorer was used. There are several stages in the analysis, which are:

loading data sets, reprocessing data, running DIMPLE to solve the protein structure using its apo model that was provided by the user, creating the ligand restraints, finding hits (PanDDA), PanDDA analysis, PanDDA inspect and final step was model fitting by using Coot.

The 8th step is solving and refining the structures by using CCP4 and its modelling in Coot.

2.2.2.7.4 Nuclear Magnetic Resonance

Each fragment was Quality Controlled (QC) before its run with the protein of interest in order to identify that it would not degrade or become contaminated during the actual run. Fragments with a singular fluorine should exhibit 1 fluorine in a 19F CPMG experiment. Fragments with 2 fluorine's in 2 different positions should display 2 fluorine peaks in a 19F CPMG experiment. If the spectra showed any other additional small peaks it was considered as a contaminant or compound degradation, hence that fragment was abandoned from further analysis.

2.2.2.7.5 Surface Plasmon Resonance

All experiments were performed with a Biacore T200 surface Plasmon resonance biosensor instrument (GE healthcare). Proteins were immobilized on CM5 series S sensor chips. All solutions were freshly prepared, degassed, and filtered. The neutravidin immobilization was run on HBS-N at a flow speed of 10 µL/min at 25 °C. The matrix of the sensor chip was activated at a flow rate of 10 µM for 420 seconds. Subsequently, neutravidin (0.30 mg/mL) in a 10 mM NaAc solution (pH 5.0) was injected for 120 seconds. Unreacted activated groups of the dextran matrix were deactivated by injection of ethanolamine. HCl (1 M) for 420 s.

The proteins were buffer-exchanged and diluted to 0.5 mg/mL and injected on the flow channels until 3000 RU in the same buffer at 15 °C. Biocytin (0.05 mg/mL) was injected for 120 seconds on all flow channels at 15 °C. All compounds were dissolved in DMSO and diluted to 100 µM in 0.5 M Tris-HCl (pH 8.0), 150 mM NaCl, 4 mM $MgCl_2.6H_2O$, 100 nM $ZnCl_2$, 5% glycerol (v/v), 2 mM 2-Mercaptoethanol, 0.005% Tween-20 (v/v), 2% DMSO (v/v). A 6-point concentration range of each compound was prepared in a dilution series and measured in multicycle experiments. Compounds were injected for 60 seconds and their dissociation was monitored for 300 seconds. All titrations were run at 25 °C at a flow speed of 50 µL/min. Data analysis were performed with BIAcore evaluation software 2.0. Signals were subtracted from reference surface and blanc injections. DMSO correction was performed. The affinity was determined by fitting a Langmuir binding equation to steady state binding signals at different concentrations.

5.5.1 SPR sensograms representation

The most common method that is used in SPR analysis is the use of Langmuir model where the experiment is characterized by 1:1 interaction of ligand and a protein. There are 2 main states in the reaction: association constant (k_a, $M^{-1}s^{-1}$) and the dissociation constant (k_d, s^{-1}). Equation 1 represents the process.

$$A+B \xrightleftharpoons[k_d]{} AB \qquad\qquad \textit{Equation 1}$$
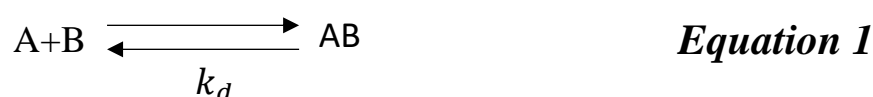
The equilibrium constant - K_D (M) is calculated using 2 kinetic constants:

$$- K_D = k_d/k_a$$

In order to relate the interaction between ligand and the protein of interest on the SPR sensograms, specific mathematical equations are used, Figure 4.34.

$$R_t = \frac{R_{max}\,[A]}{K_D + [A]}\left[1 - e^{-(k_a\,[A]\,+\,k_d)t}\right]$$

Equation 2

$$R_t = R_0 e^{-k_d t}$$
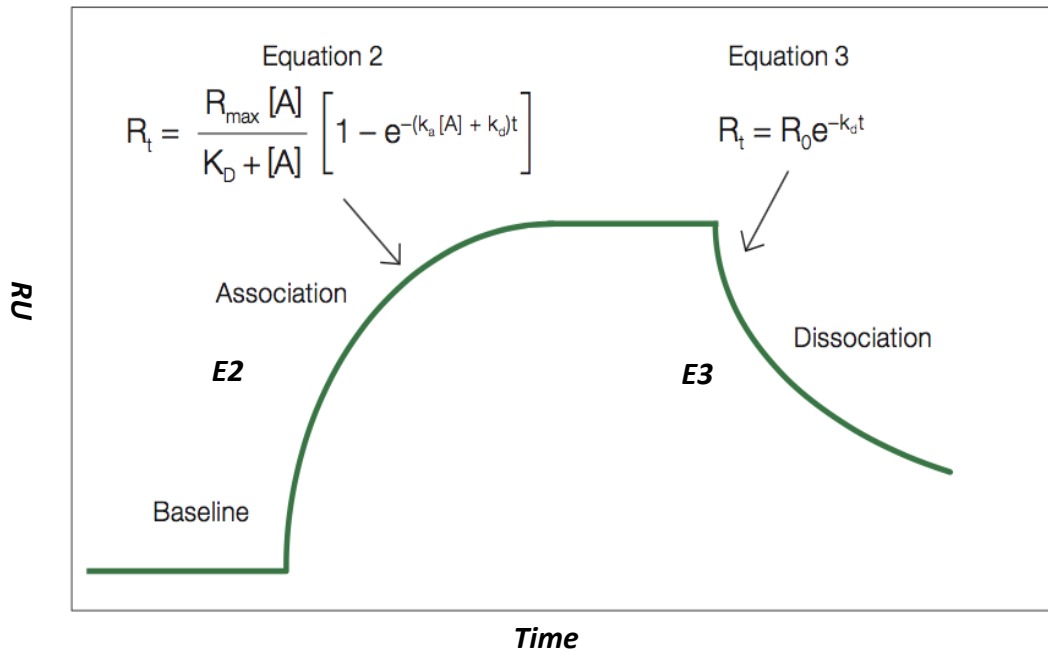
Equation 3

Figure 5.34: Reference sensogram of SPR experiment representing 3 states: baseline, association and dissociation. E2 – equation 2, E3 – equation 3. (BioRed manual)

In the Figure 5.34, Equation 2 represents the association constant while Equation 3 represents dissociation constant. The baseline state is followed by association state where the molecules of the protein or ligand of interest are flowing over the ligand surface and the rate of the complex formation is measured. Therefore, Equation 2 forms a derivative equation of such process:

$$R_t = \frac{R_{max}[A]}{K_D+[A]}\,[1 - e^{-(k_a[A]+k_a)t}]$$
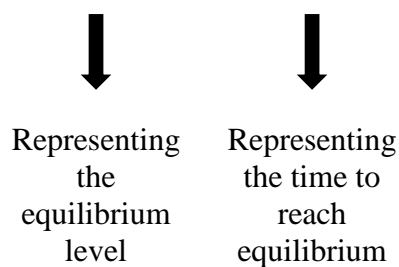
Representing the equilibrium level   Representing the time to reach equilibrium   *Equation 4*

Equation 3 represents the dissociation phase where the amount of protein or ligand of interest in the flow cell is reduced by introducing running buffer into the system instead of the analyte sample. This equation represents how quickly the system has reached a particular point during the experiment as a dissociation phase.

When the data has been obtained there are certain rules that should be followed in order to present the results. To calculate K_D value the following equation would be used:

$$R_{eq} = \frac{R_{max[A]}}{K_{D+[A]}}$$ *Equation 5*

The equation 5 represents the equilibrium state during the experiment where the complex is in steady state.

A For $[A] \gg K_D \longrightarrow R_{eq} = R_{max}$

B $R_{eq} = \dfrac{R_{max}\,[A]}{K_D + [A]}$

For $[A] \ll K_D \longrightarrow R_{eq} = \dfrac{R_{max}\,[A]}{K_D}$
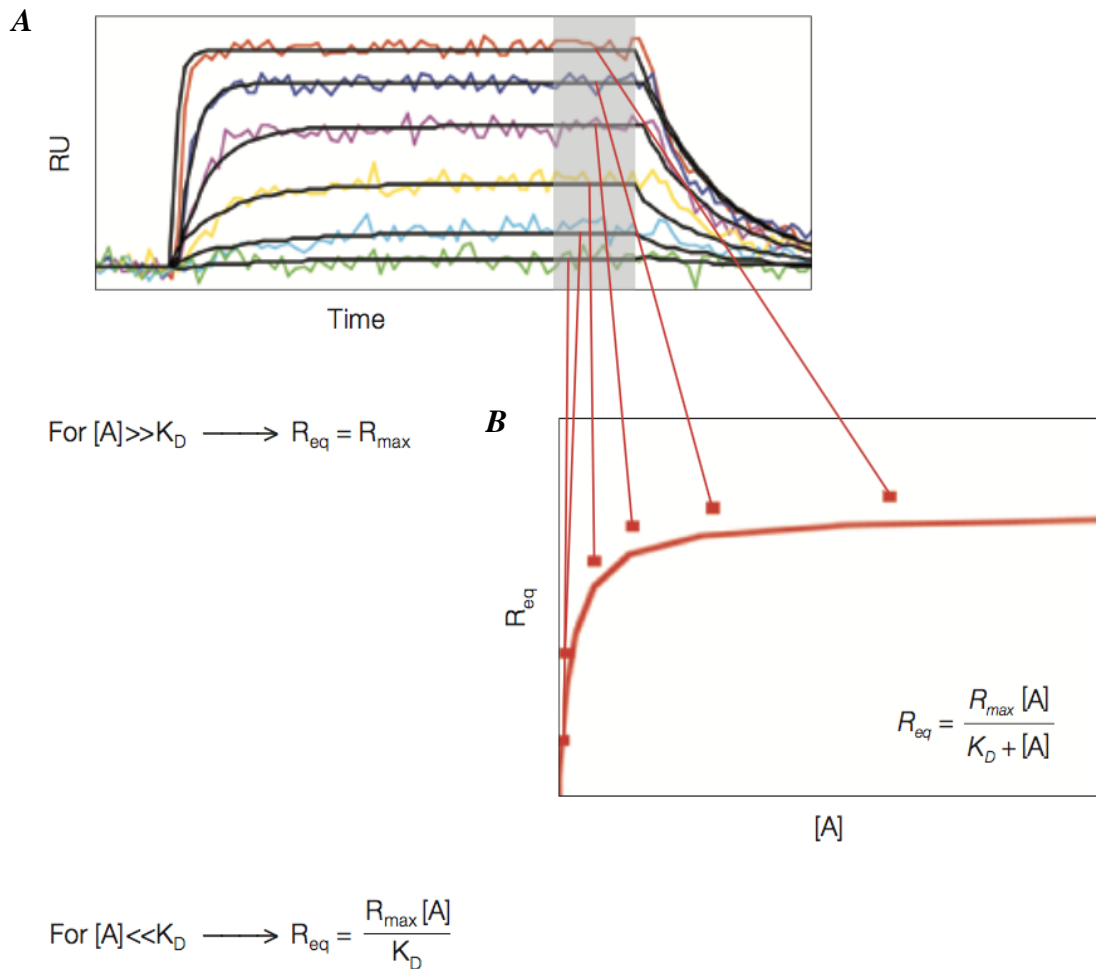
Figure 5.35: SPR sensogram of experiment representing running protein at different concentrations to identify equilibrium constant, *A)* measurement of R_{eq} over a different protein concentrations, *B)* values are plotted on a single curve (BioRad manual)

The equilibrium response (R_{eq}) is measured at different protein/ligand concentrations followed by its plotting onto the graph, Figure 5.35. R_{eq} is proportional to the analyte concentration at its low concentartions but when the concentration increases the value is approaching its theoretial maximum of R_{eq} (Protein interaction analysis, Guide to SPR Data Analysis on the ProteinON XRP36 System, BioRad manual).

CHAPTER 3

Expression and Purification of Parasitic PDEDs from

L. infantum, L. donovani and *T. brucei*

3.1 Introduction

The results presented in the current chapter describe the study of recombinant protein production of PDED from *L. infantum, L. donovani* and *T. brucei*. At the beginning of the current study only the DNA sequence was available with no experimental data on the protein encoded. Figure 3.1 represents *PDED* full length sequence alignment between the 3 species stated above.
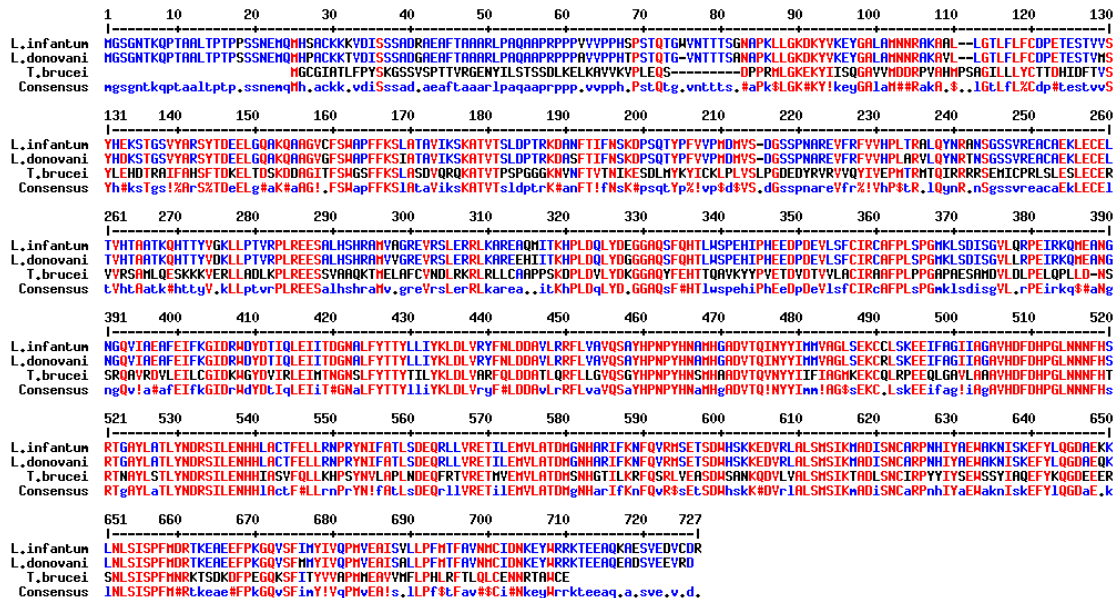
Figure 3.1: Multiple protein sequence alignment of PDED from *L. infantum, L. donovani and T. brucei,* MultAlin software was used, (http://multalin.toulouse.inra.fr/multalin/)

Some preliminary work has been reported on PDEB1 from L. major *(88)*, where a catalytic domain crystallization construct in Leishmania has been analysed, however, it was important to study other parasitic PDEs as potential drug targets and the least well studied were PDEDs. L. infantum and L. donovani species cause visceral Leishmaniasis that can be fatal or cause serious medical conditions especially when there is a co-infection with HIV (World Health Organization, March 2019). It is hoped that further experimental data on PDED could help validate it as a drug target and could be used for the designing of a new or modified tool compound that may be effective for treatment of Leishmaniasis worldwide. According to the blast analysis, using blast.ncbi.nlm.nih.gov software, the Leishmania species showed 99% sequence similarity amongst the PDED's while within the Trypanosoma, the sequence similarity was only 49.28%. Cloning was initially pursued on 2 strains of Leishmania species namely: L. infantum and L. donovani. Software such as Pfam (collection of protein families and sequences alignments) and Phyre (software for protein 3D structure recognition) (https://pfam.xfam.org) were used to identify the most favourable constructs for protein expression, purification and its further characterization.

3.2 Results

3.2.1 Leishmania PDED

Genomic DNA of *L. infantum PDED* was provided by the University of Antwerp where the full length of *PDED* from strain *JPCM5* was amplified using the PCR technique, Section 2.2.2.2.1. There are 724 amino acids in the full-length sequence. As mentioned, the Pfam software was used to define the catalytic domain location in the *PDED* full length sequence. According to Pfam the catalytic domain was located between

amino acids 461 and 695, which was also confirmed by visually comparing its sequence with other known PDE catalytic domain structures. In addition to this approach for construct design the Phyre (Protein Homology/AnalogY Recognition Engine) software was used to analyse a possible 3D protein structure based on already deposited structures of other proteins. This software was useful for analysing protein secondary structure where any disordered regions were identified. After identifying the location and sequence of catalytic domain a number of constructs were chosen to clone and express the protein. The catalytic domain constructs were designed with varying N and C termini. The first, and very important, step in structural biology is the identification of the appropriate construct that will produce high protein yield in a recombinant expression system which may be easily purified to produce homogeneous protein that then may result in a highly ordered crystal that diffracts well. Structure Based Drug Design (SBDD) approach requires the development of a robust crystal system to analyse possible binding modes between ligands and target protein. The final yield of purified protein can be influenced significantly by construct design, where tags, secondary structure and composition of hydrophobic and hydrophilic amino acids were taken into consideration at the beginning and end of the sequence as they could alter PDED expression levels, its solubility and stability during the purification *(89).*

The first construct that was designed for PDED started at amino acid 394 which was Glutamic acid and ended at the residue 714 which was a Lysine. Lysine and Glutamic acid were both charged hydrophilic amino acids also capable of formation of salt bridges. It had been proven previously by Boël G., et al that reduction of the use of rare codons (codon optimization) could play an important role in level of protein

expression *(90)*. The 3 constructs of catalytic domain of L. infantum PDED were designed, Figure 3.2.

3.2.2 Cloning

The catalytic domain was cloned using primers (Section 2.2.1.3, Table 2.3) into a designed pET15b vector with the inclusion of a TEV cleavage site, for later removal of the purification tag Figure 3.3.
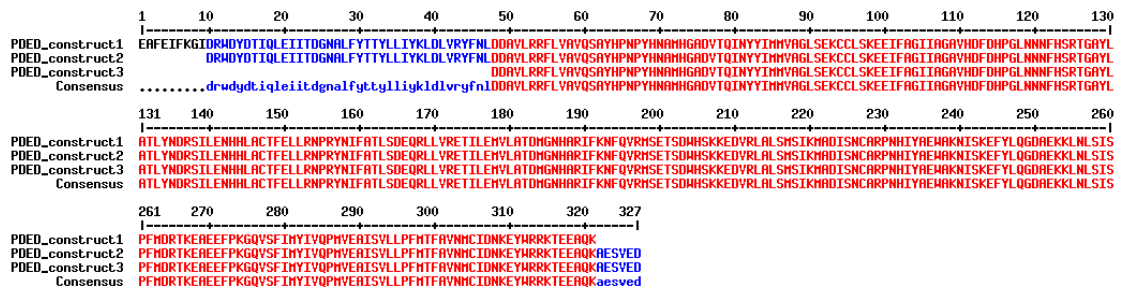
Figure 3.2: Protein Sequence Alignment. 3 designed constructs of PDED catalytic domain from L. infantum. MultAlin software was used (http://multalin.toulouse.inra.fr/multalin/)
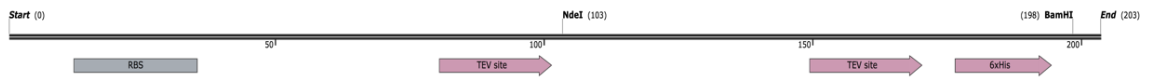
Figure 3.3: designed TEV cassette for insertion into pET15b native vector, SnapGene software was used to design current construct
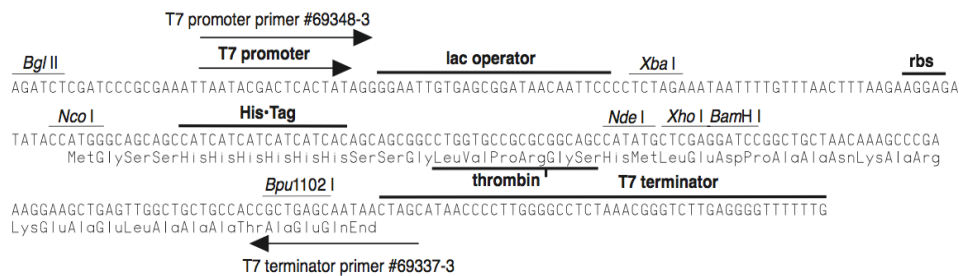
Figure 3.4: native pET15b plasmid, (www.addgene.org)

For PDED construct insertion, 2 restriction enzymes were used: BamH1(*G'GATCC*) and NdeI (*CA'TATG*), Section 2.2.2.4.2.

Figure 3.4 showed the DNA sequence of native vector pET15b that was used as a scaffold for designing modified version with TEV cleavage site, Figure 3.5.

The main advantage of using this cassette is that TEV site is located on N and C terminus where the gene could be fused to a HisTag on either or both termini. PDED catalytic domain was then cloned into pET15bTEV modified vector. Details of the vector and vector containing PDED catalytic domain can be found further, Figures 3.5, 3.6.
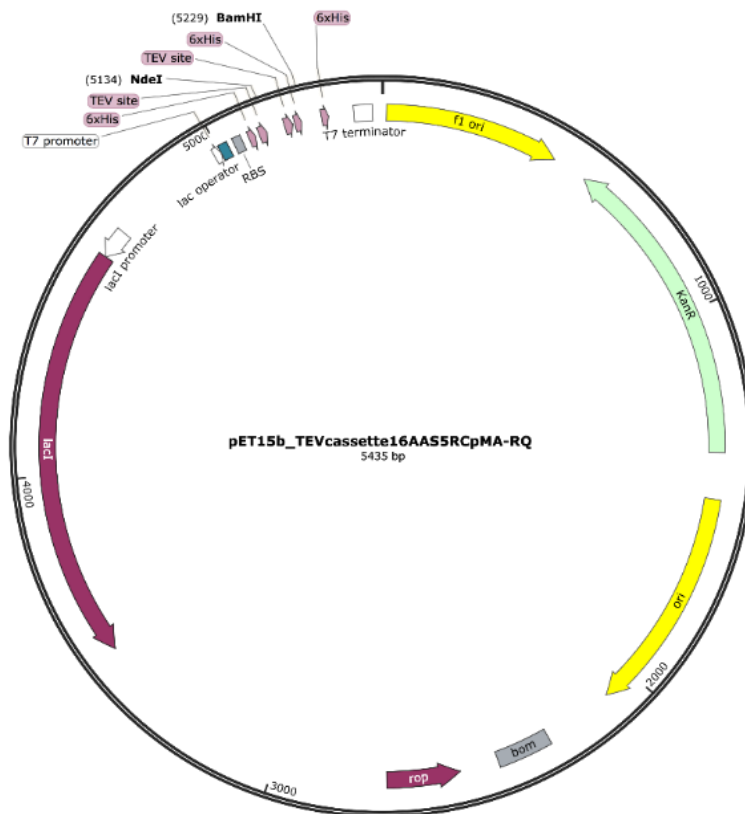
Figure 3.5: schematic representation of TEV cassette cloned into pET15b vector, Vector map was designed using SnapGene software
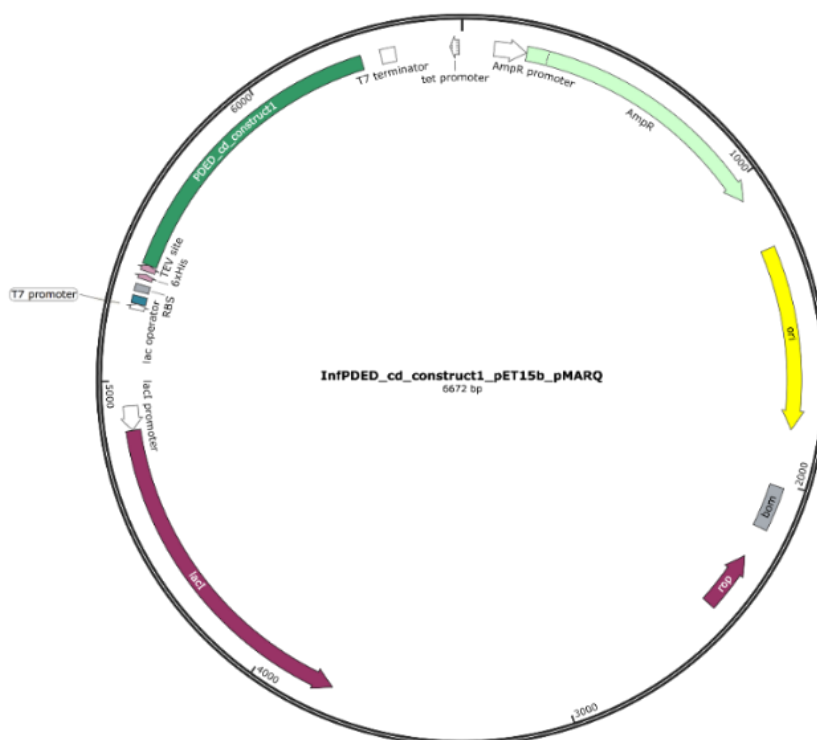
Figure 3.6: schematic representation of pET15b vector with inserted cassette containing TEV cleavage site and PDED catalytic domain. Vector map was designed using SnapGene software.

Amplification and cloning were performed by PCR and ligation methods, Sections 2.2.2.3 - 2.2.2.5. After cloning it was important to check the plasmid integrity, since any deletions or mismatch could result in different amino acids being translated and as a result different protein being expressed, so the plasmids were sent for sequencing.

3.2.3 PDED expression

Once the oligonucleotide sequences were confirmed, the next step was to analyse the level of PDED expression and its solubility. Initially a small-scale expression and purification method was performed where standard conditions were used, Section *2.2.1.7.1.* In that particular experiment *E. coli* T7 express cells were used.
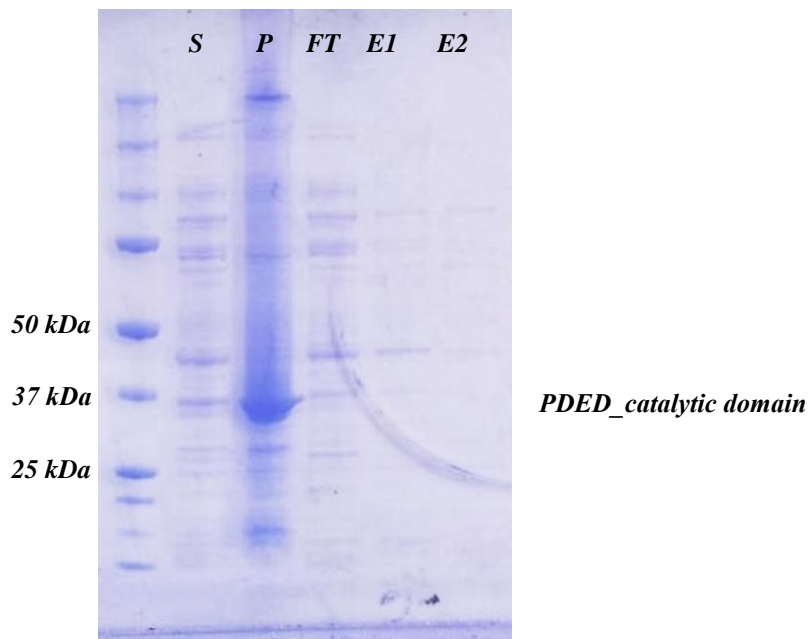
Figure 3.7: SDS-PAGE of Ni-NTA small scale expression of L. infantum PDED catalytic domain. The expected molecular weight - 37. 3 kDa.
Abbreviation: S (supernatant), P (pellet), FT (flow through), E (elution)

As shown in Figure 3.7, catalytic domain of PDED was expressed, however according to the SDS-PAGE it could only be detected in the pellet fraction. The main reason could be protein insolubility due to misfolding. Due to such poor soluble expression levels, additional new constructs were studied, Figure 3.8.
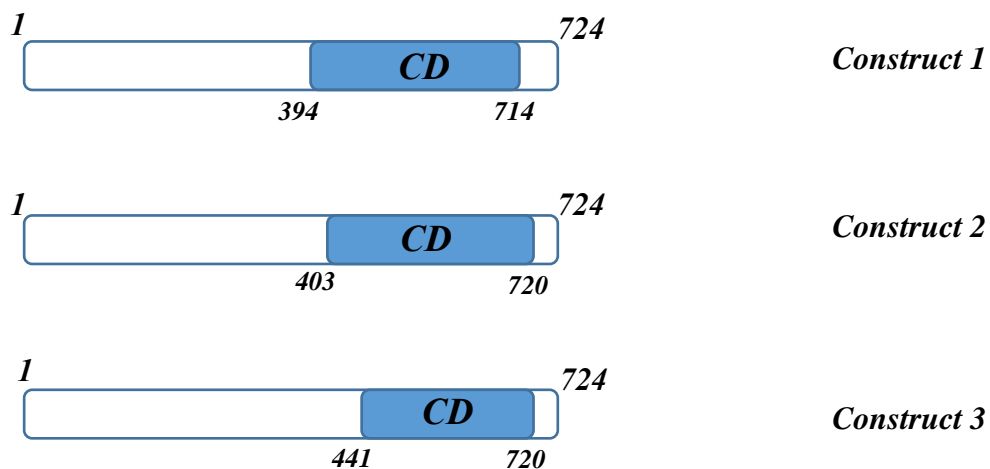
Figure 3.8: schematic representation of 3 designed constructs of PDED catalytic domain

Construct 2 was designed shorter then construct 1 where the start of the 2nd construct was shorter by 10 amino acids, Figure 3.8. However, the level of expression of this construct was similar to the expression level of 1st construct where PDED was only observed in the pellet fraction, meaning it was still insoluble. Construct 3 was designed to be shorter by 47 amino acids than the 1st construct. Again, similar expression levels were observed where no soluble material was obtained in the supernatant. Simultaneously, other lab members tried to express PDEDs from 2 other parasitic organisms, such as T. brucei and T. cruzi where multiple constructs of *PDED* catalytic domain were designed and no soluble protein was observed using an *E. coli* expression system. A variety of constructs with N and C terminal truncations, as well as some targeted surface entropy mutations, were performed but no soluble protein was obtained in all cases. As a result, it was concluded that unfortunately although using different length PDED constructs from 3 different species, the only PDED protein observed was in the pellet fraction with no improvement in protein solubility. Hence, the next step was to change the pET expression vector system and to use pOPIN vector suite.

3.2.4 pOPIN vector suite

An advantage of using the pOPIN expression system is that it is a versatile system that could be used with bacteria *(E. coli)*, insect (baculo) and mammalian cells (HEK293). The pOPIN expression system is a direct cloning system using the SLIC method, Section 2.2.2.5.2. The pOPIN system has multiple vectors with different solubilizing tags that can be used: SUMO, MBP, GST tags and etc. The main reason behind using the pOPIN vector suit was to improve protein solubility and hence develop an expression system that would show a significant difference in expression of protein through utilization of the different fusions or tags.

The amplified full length and catalytic domain of PDED were cloned directly into a range of vectors, where the PDED constructs were fused to different fusion tags. Similar expression and purification conditions were used at small-scale to assess protein solubility and identify if any differences with previous expression system could be determined. SDS-PAGE gels showed no signs of soluble material in supernatant fraction in all constructs. Therefore, more sensitive western blot technique was used, Section 2.2.2.7.1, Figure 3.9.
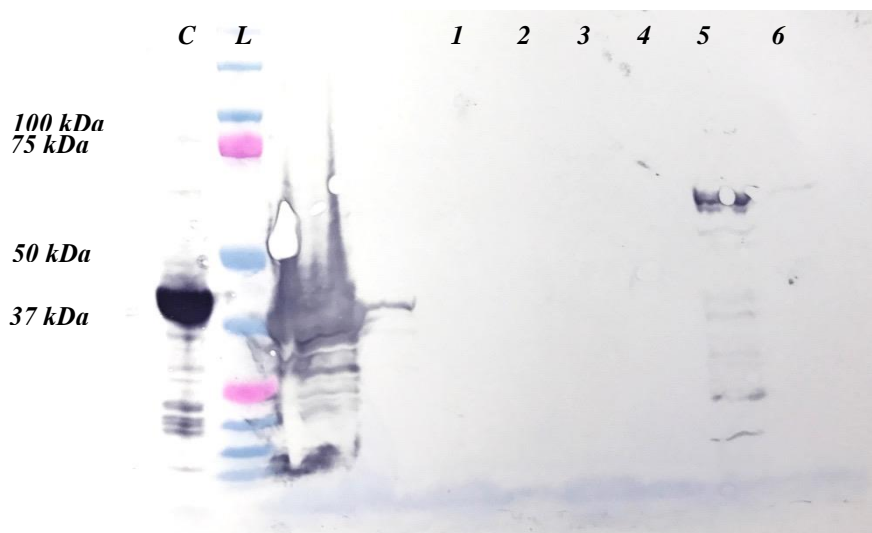
Figure 3.9: Western Blot analysis of small scale expression, Ni-NTA of catalytic domain of PDED from L. infantum using pOPIN vectors: pOPINJ, pOPINM and pOPINS3C

Key:
C – control
L – marker ladder
1 – pOPINJ, pellet
2 – supernatant
3 – pOPINM, pellet
4 – supernatant
5 – pOPINS3C, pellet
6 – supernatant

As seen in Figure 3.9, only fraction 5 which was the pellet of PDED protein expressed in pOPINS3C had a band of the expected molecular weight of PDED_HisTag_SUMO construct at 51.4 kDa. However, the

pellet fraction also showed that there was degradation where multiple size bands could be visualized. Therefore, it was concluded that PDED solubility wasn't improved using the new vector suit with fusion tags. This led to further exploration of alternative expression systems, such as the Cold Shock System.

3.2.5 pCold Expression System

The pCold expression system, which exploits the cold shock expression method was used for the expression of the catalytic domain of PDED since previous reports, *(91)*, showed that cold shock system was suitable for expressing of difficult proteins that were prone to misfolding. The Cold shock system contains cold shock protein A that acts as a promoter in the protein expression, enhancing higher levels of protein expression. It was hoped that the use of this system would improve protein folding where the expression occurs more slowly at lower temperatures hence protein folding machinery might allow more soluble protein being produced. The lower temperature (15 °C) may also reduce the expression of the contaminants and reduce protease activity that could negatively affect protein yield. The cold shock expression system also allowed labelling with different tags and/or fusions, e.g. with chaperone Trigger Factor. Two vectors from pCold system were used: pColdI and pColdTF, where protein of interest was fused with solubilizing chaperone such as Trigger Factor (TF) in pColdTF vector. *L. infantum PDED* was cloned into 2 vectors: pColdI and pColdTF, Section 2.2.2.4.
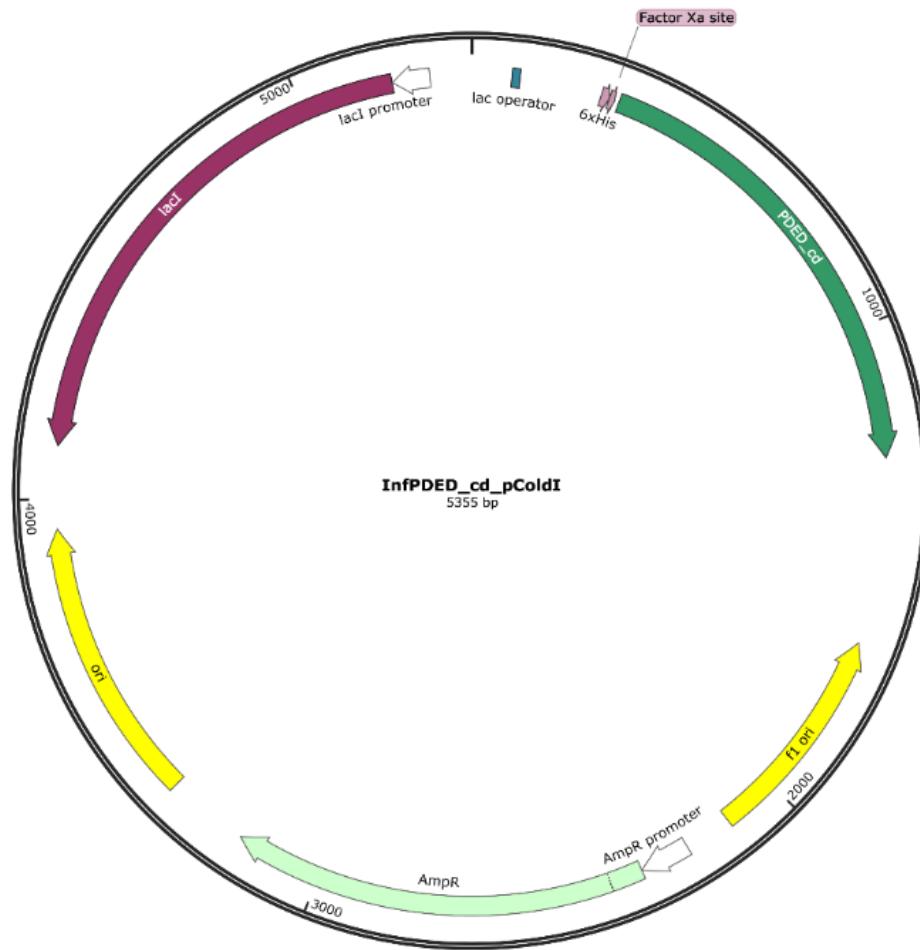
Figure 3.10: schematic representation of cloned PDED catalytic domain gene of L. infantum into pColdI vector. Vector diagram was designed using SnapGene software.
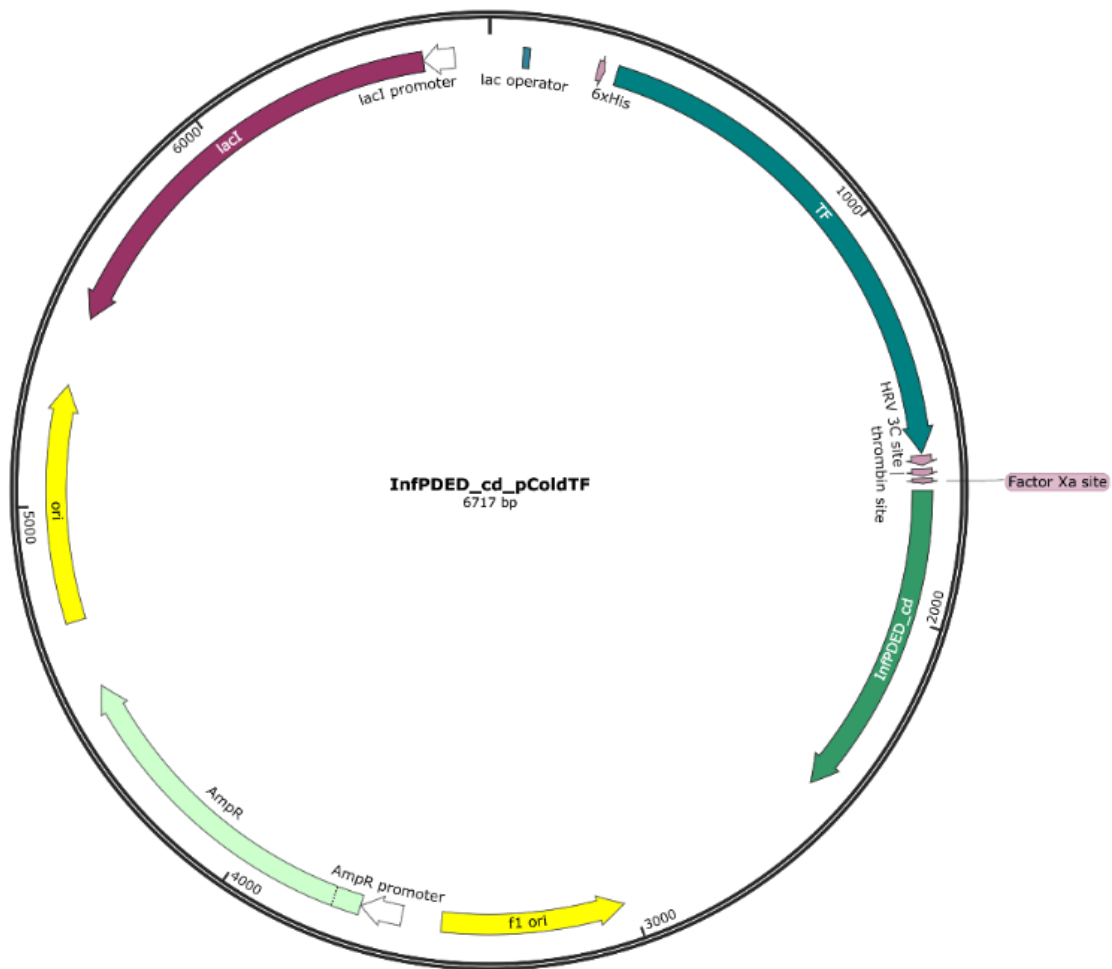
Figure 3.11: schematic representation of cloned PDED catalytic domain gene of L. infantum into pColdTF vector, Vector construct was designed using SnapGene software

Trigger Factor (TF) is a prokaryotic ribosome – associated protein of 48 kDa acts as a chaperone to enhance co-translational modifications of PDED in the hope it will improve its folding by reducing formation of Inclusion Bodies (IBs).

For current study construct_1 of *L. infantum* was used, Figure 3.8.
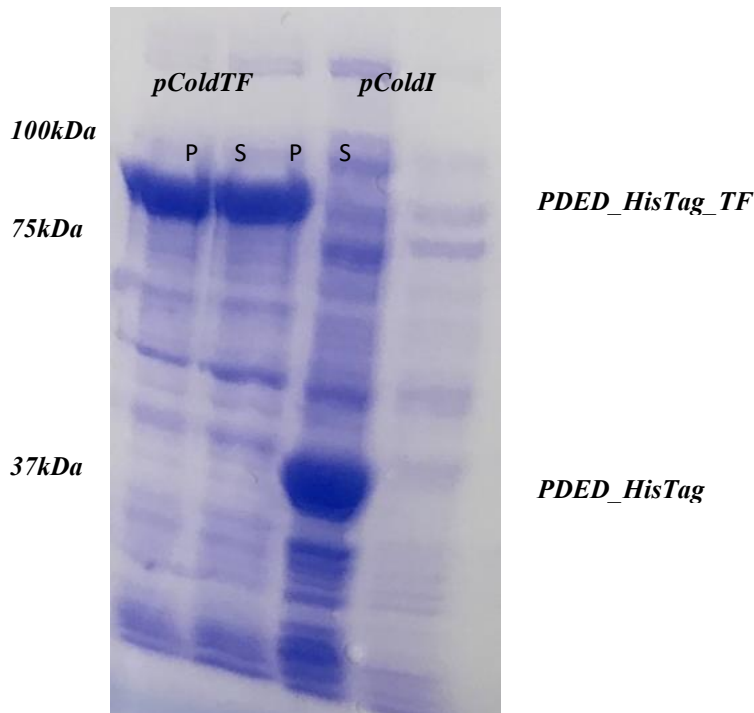
Figure 3.12: SDS-PAGE of Ni-NTA small scale expression of L. infantum PDED catalytic domain in pColdI and pColdTF vectors.
 Key: 1 – PDED in pColdTF pellet
    2 - PDED in pColdTF supernatant
    3 - PDED in pColdI pellet
    4 - PDED in pColdI supernatant

Figure 3.12 shows Ni-NTA purification of a small-scale expression where standard *E. coli* T7 expressed cells were used with standard expression conditions for this system, Section 2.2.1.7.1. Figure 3.12, highlights that use of the pColdTF expression vector yielded reduced insoluble material in the pellet fraction with expected molecular weight of TF_HisTag_PDED construct being 87.4 kDa, and some soluble material in a supernatant fraction. Since this significant improvement in protein solubility was observed it was decided to proceed studies on *L. infantum PDED* construct_1 with the additional use of the Takara chaperones plasmid set, which could further improve soluble protein yields.

3.2.6 Takara chaperone plasmid set

To further improve protein solubility Takara chaperone plasmid set would be used where PDED would be co-expressed with 5 different plasmids, Table 3.1.

No	Plasmid	Chaperone	Promoter	Inducer	Resistant Marker
1	pG-KJE8	*dnaK-dnaJ-grpE-groES-groEL*	araB Pzt-1	L-Arabinose Tetracycline	Chloramphenicol
2	pGro7	*groES-groEL*	araB	L-Arabinose	Chlora-mphenicol
3	pKJE7	*dnaK-dnaJ-grpE*	araB	L-Arabinose	Chloramphenicol
4	pG-Tf2	*groES-groEL-tig*	Pzt-1	Tetracycline	Chloramphenicol
5	pTf16	*tig*	araB	L-Arabinose	Chloramphenicol

Table 3.1: Takara chaperone plasmid set, Takara product manual

Table 3.1 represented characteristics of chaperone plasmids where each plasmid was carrying chloramphenicol resistance genes. Therefore, when competent cells were prepared (see Section 2.2.2.1.5) containing different chaperone plasmids it was possible to identify if the transformation was successful. Chaperones have *araB* and *Pzt-1* promoter's genes therefore chaperones and protein of interest could be expressed by using different induction agents, Table 3.1.
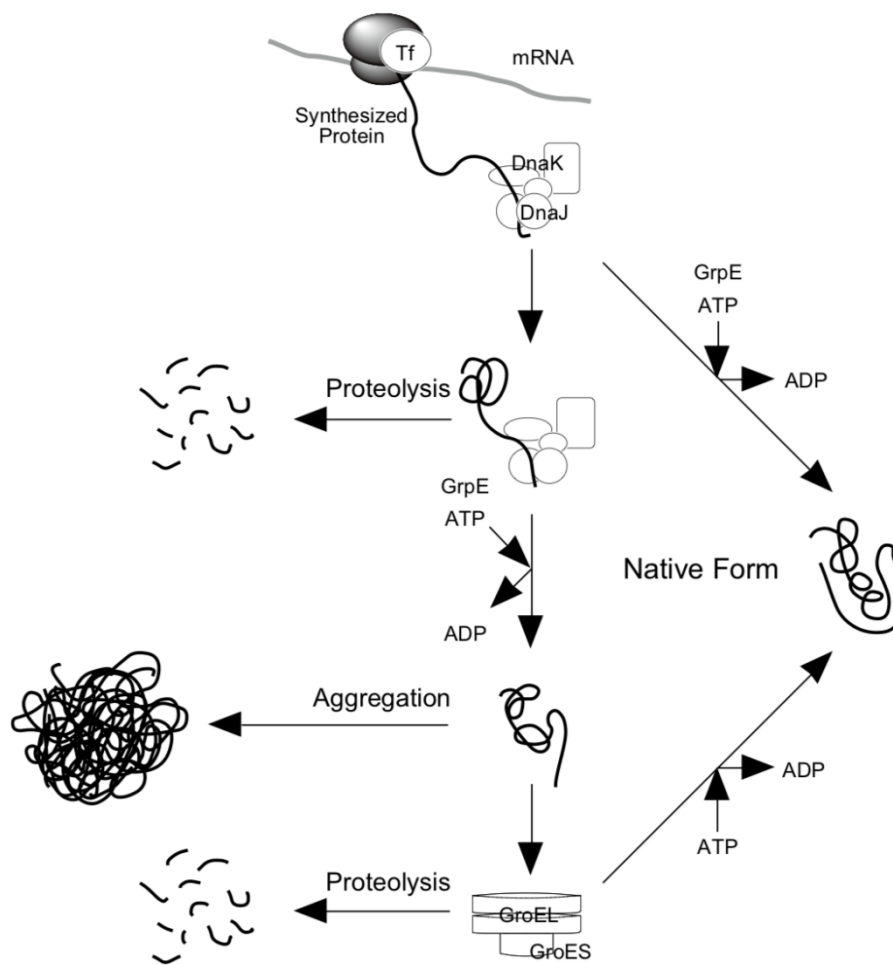
Figure 3.13: Schematic diagram of chaperones functioning during protein folding, (Takara chaperon plasmid manual)

Small scale expression (20 ml) was used to assess soluble expression levels, it was important to maintain the same number of cells in order to keep the results accurate and reliable for comparison with the 5 different chaperon plasmids. Co-expression of PDED were performed and western blot analysis was used. Primary antibody anti – HisTag was used in western blot analysis to identify the best co-expression system for PDED, Figure 3.14. TcrPDEB1, a protein with high level of soluble expression was used as a control.
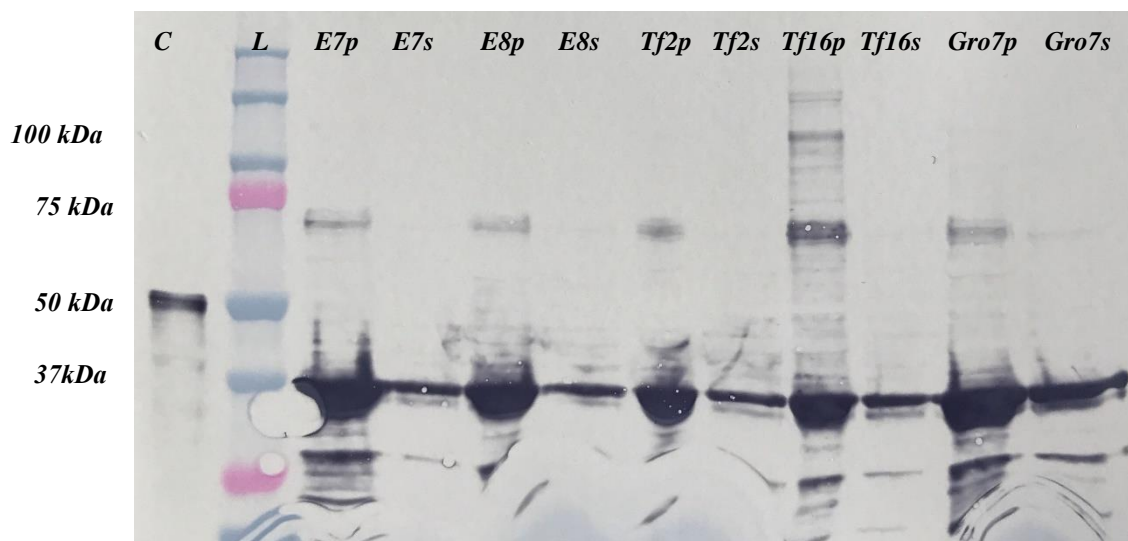
Figure 3.14: Western blot of a small-scale co-expression of PDED catalytic domain in pColdI vector with 5 chaperone plasmids, where C – control and L – ladder marker, E – elution fraction

Key:
C – control
L – ladder
E7 – pKJE7
E8 – pG-KJE8
Tf2 – pG-Tf2
Tf16 – pTf16
Gro7 – pGro7
S – supernatant
P – pellet
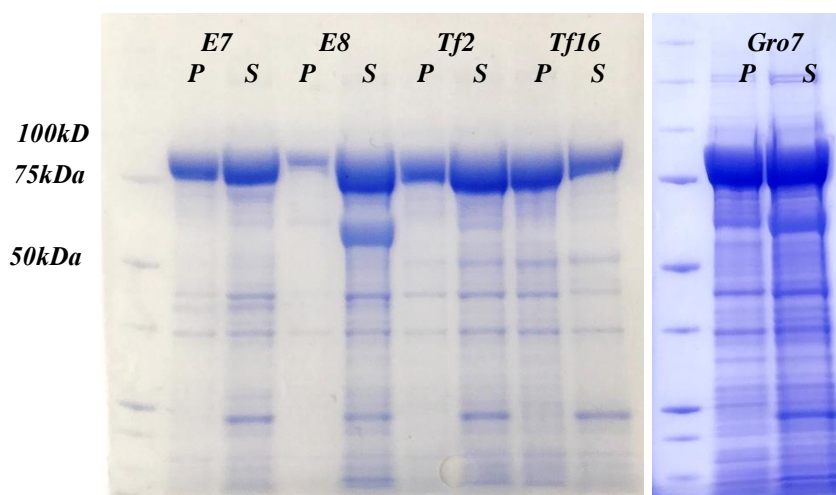
Figure 3.15: SDS-PAGE of NI-NTA purification method of PDED (cloned in pColdTF) co-expression with Takara chaperon plasmid.

Key:
E7 – pKJE7   E – elution fraction
E8 – pG-KJE8
Tf2 – pG-Tf2
Tf16 – pTf16
Gro7 – pGro7
P – pellet
S – supernatant

According to the western blot analysis, Figure 3.14, and SDS-PAGE as a measure of solubility of PDED in pColdTF, Figure 3.15, the highest level of PDED expression was observed with co-expression with the plasmid pGro7. Chaperone plasmid pGro7 contained 2 chaperones GroES and GroEL that were cytoplasmic chaperones. GroEL is a prokaryotic protein that is found in prokaryotic cytosol where the main function is to hold and fold protein with cooperating factors GroES, Hsp10 and Cpn10 and being an ATP consumer. GroES is a sHTP type of chaperone and is found in the cytosol with the main function being holding and preventing protein aggregation during stress response. Both GroES and GroEL are ATP dependent *(92)*. After the preferred expression system was identified, it was important to purify protein of high quality and high yield, in order to perform biochemical, biophysical and structural analysis, such as crystallization. Therefore, the next stage was to scale up and express larger quantity of cells and further optimization of expression and purification conditions including different fermentation strategies with different additives.

3.2.7 Optimization of Protein Expression and Purification

The 1L co-expression of PDED with chaperone plasmid was induced under the same conditions as small scale culture, Section 2.2.1.7.1, in order to see if the expression level would be consistent, but also with additives to investigate optimization of protein yield and quality. Different concentrations of the inducing agent arabinose were tested from the range of 1mg/ml to 5 mg/ml where *Gro-ES* and *Gro-EL* genes of chaperones were induced. There were 2 pCold expressing vectors used: pColdI_PDED_cd and pColdTF_PDED_cd. After comparing results of expression from both vectors using 5 different concentrations of arabinose it was concluded that pColdI need higher concentration of inducer then

pColdTF vector, as using 4 mg/ml of arabinose in pColdTF co-expression led to decrease in amount of protein solubility. The optimum concentration for pColdTF was found to be 1.5 mg/ml while for pColdI the optimum was suggested to be 4 mg/ml.

Class of additive	Example	Concentration range	Target
Salts	NaCl, KCl	100-500mM	Maintaining ionic strength
Detergents	Triton x-100	0.1%-1%	Solubilizing
Glycerol		5%-20%	Stabilization
Sugars	Glucose, Sucrose, Sorbitol	25mM	Stabilization and Solubilizing
Metal Chelators	EDTA	1mM	Decrease damage with oxygen
Reducing agents	DTT, beta mercaptoethanol	1-4mM	Decrease damage with oxygen
Ligands	IBMX		Stabilization
Metal ions	Mg, Zn	1-10mM	Stabilization and Solubilizing

Table 3.2: Additives used in buffers optimization, *(93)*

Table 3.2 detailed the list of additives that were used in different variations and at different concentrations to improve protein expression yields. To compare additive's effects on PDED expression, small scale culture (20 ml) was used with equal number of cells. The best results were obtained with buffer composition such as: 20mM Tris pH 7.5, 500mM NaCl, 1mM DTT, 25mM sucrose, 10% (v/v) glycerol and 2mM

MgCl2. This buffer showed the most promising results in terms of PDED expression and amount of soluble protein being produced in the solution, Figures 3.16 and 3.17. Same buffer was used for 1L bacteria culture expression.
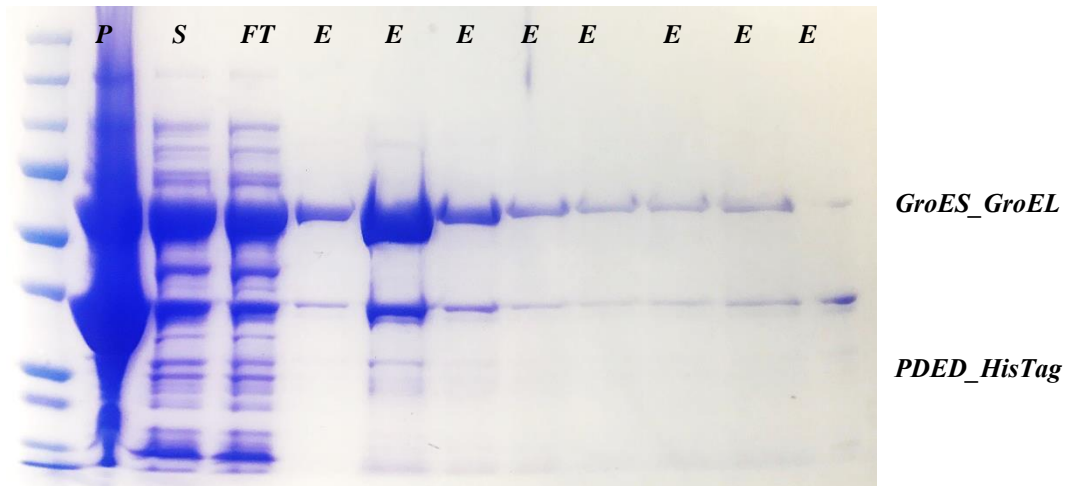
Figure 3.16: IMAC purification of PDED catalytic domain, pColdI expression vector, co-expression with pGro7. Molecular Weights: PDED – 37.2 kDa, GroEs/GroEl – 70kDa
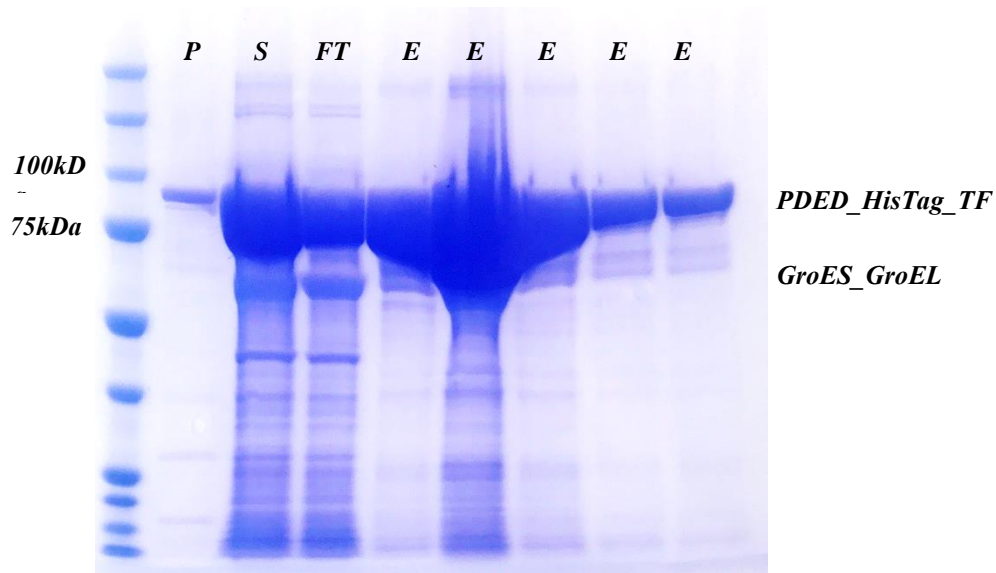
Figure 3.17: IMAC purification of PDED catalytic domain, pColdTF expression vector, co-expression with pGro7. Molecular Weights: PDED_TF_HisTag – 87.4 kDa, GroEs/GroEl – 70kDa

After finding an appropriate purification conditions, where soluble PDED catalytic domain was obtained it was important to analyse protein quality and activity by biophysical and biochemical techniques, such as: protein crystallography and enzymatic activity assay.

3.2.8 Enzymatic Activity Assay and Further Optimizations

A key experiment was to assess PDED enzyme activity after 1st step of IMAC purification where elution fractions were used Figure 3.16 and 3.17.
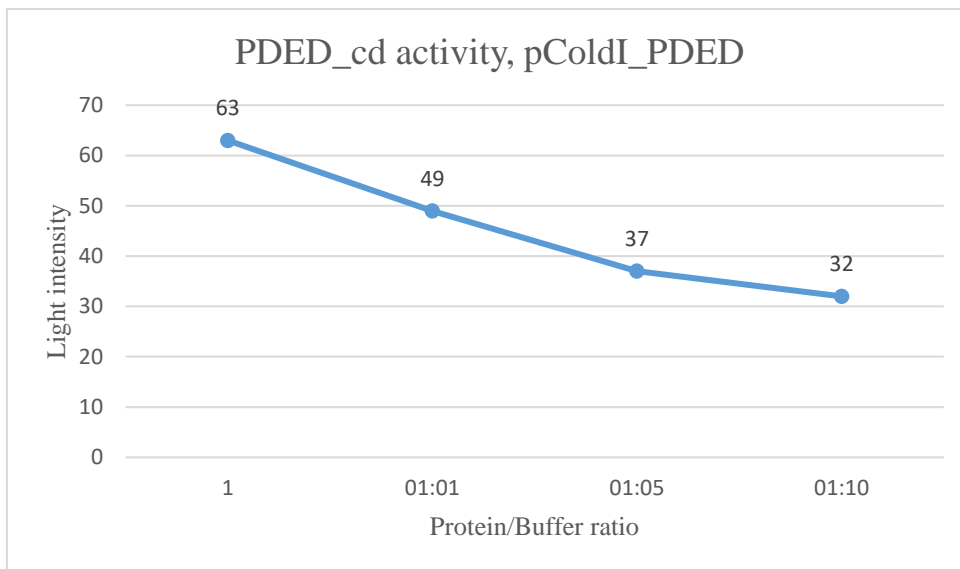
Figure 3.18: enzymatic activity assay of PDED that's was expressed in pColdI vector. Experiment was performed by using Lonza assay kit.
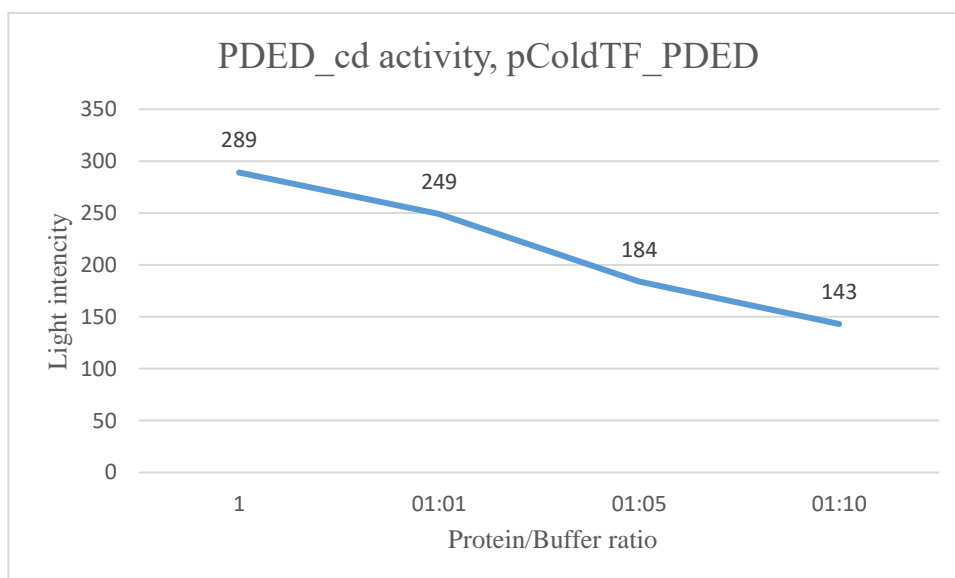
Figure 3.19: enzymatic activity assay of PDED that's was expressed in pColdTF vector. Experiment was performed by using Lonza assay kit.

According to Figure 3.19, PDED that was co-expressed in pColdTF vector with GroEL and GroES chaperones had 4.5 times higher activity level in comparison to pColdI protein. PDED that was expressed in pColdTF vector was fused to trigger factor, which may prevent the protein degradation resulting in a more stable and active PDE to catalyse cAMP hydrolysis. Based on the activity assay results it was decided to continue to progress scale up and purification of PDED fused with the trigger factor.

The next step was to optimize purification conditions to maximize protein yield of correctly folded material. During each step of purification, i.e. IMAC, IEC and SEC, buffer conditions were varied in order to see which component would enhance solubility of protein and its expression level. Such techniques as IEC and SEC would be used in order to improve protein sample purity. The main reason behind was that only highly pure protein material can be used in crystallography. The gel filtration chromatography (SEC) was used as the last step in protein purification where protein would be analysed according to its size and how homogenous the sample was. Before proceeding towards analytical purification, it was necessary to cleave PDED from the HisTag_TF complex. Thrombin cleavage was performed overnight at 4 °C and followed by the 2nd run of IMAC purification where uncleaved PDED protein would be separated from HisTag_TF complex and other contaminants. In order to access protein quality after 2nd Ni-NTA purification, western blot analysis was performed, Figure 3.20.
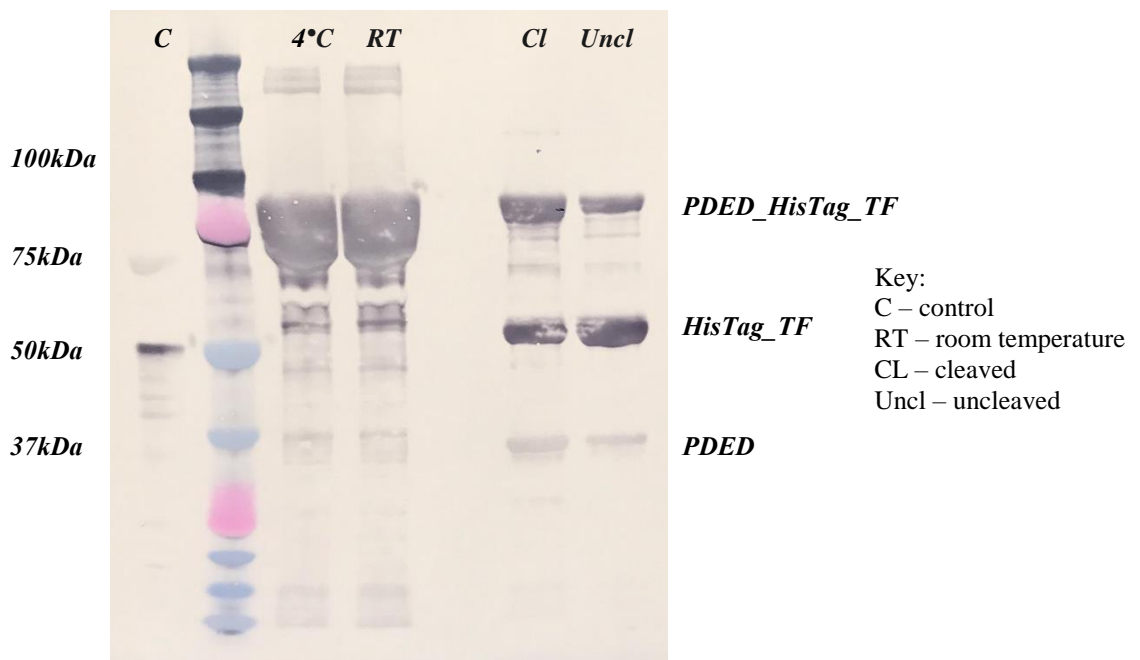
Figure 3.20: Western blot analysis of PDED after thrombin cleavage. Purification was performed using affinity NI-NTA column.

The Figure 3.20 showed PDED state in the solution where 3 different cleavage conditions were analysed: incubation with thrombin at room temperature and 4°C for 2 hours and incubation of PDED with thrombin overnight at 4°C. According to the Figure above 3.20, samples incubated at room temperature and 4 °C showed protein degradation as different molecular weights fragments could be visualized from the western blot. The flow through sample that was supposed to contain only cleaved sample also contained uncleaved PDED that was still fused with TF_HisTag complex. This experiment showed that thrombin cleavage was not efficient and PDED couldn't easily be separated from uncleaved sample by this single step. One possible reason could be protein misfolding where aggregated parts of PDED were attached to soluble cleaved PDED, due to its high concentration in a protein sample, hence preventing cleaved protein separation.

3.2.9 Size Exclusion Chromatography

Since the protein separation from HisTag_TF complex caused issues it was decided to try and separate PDED by directly using SEC chromatography. Two separate samples of secondary Ni-NTA purification, Figure 3.20, would be used on SEC. The chromatography diagram of SEC at absorbance of A280 nm would show peaks according to protein molecular weight where PDED could be eluted separately from other contaminants.
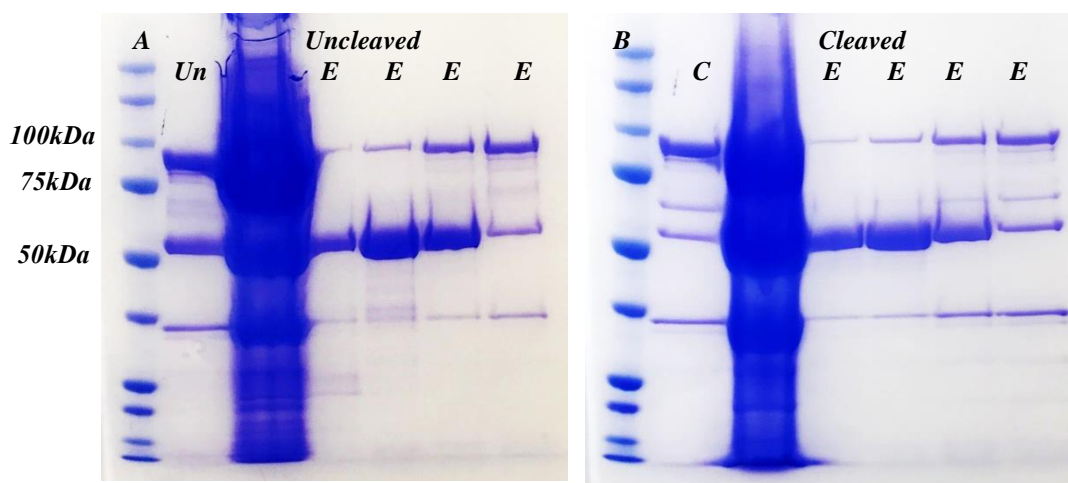
Figure 3.21: SDS-PAGE of elutions after SEC. 2 separate runs: A – uncleaved PDED from HisTag after 2nd run of IMAC and B – cleaved PDED after 2nd run of IMAC

This approach was also used to access how homogenous protein was and if this PDED construct was suitable for crystallization trials. According to the Figure 3.21 both runs A and B showed poor separation of protein from its aggregates and HisTag_TF complex. Both runs contained the protein of interest; however, aggregates were the predominant species in the sample. Also, after a short storage time, when sample was accessed on western blot there was more degradation, suggesting that aggregates parts in sample further promotes PDED degradation. Therefore, it was concluded that misfolding of PDED made this protein unsuitable for further analysis using crystallography. After investigation of several rounds of gel filtration it

was still not possible to obtain highly pure PDED catalytic domain and the concentration of the cleaved sample decreased as protein tended to degrade after each purification run.

3.2.10 N-terminal Region of PDED

According to the previous results where the main focus was on catalytic domain of PDED, analysis of protein expression and its behaviour in solution revealed that PDED was insoluble and a number of optimizations were required, such as: cloning, expression and purification in order to solubilize protein and express it. Since PDEB contained 2 domains upstream of the catalytic domain (Gaf-A and Gaf-B) it was suggested that PDED could potentially contain upstream domains that influenced its solubility and potentially regulate its PDED activity *(94).* Therefore, it was decided to express and check enzymatic activity of N-terminal region of PDED to investigate if there are any unknown domains that could potentially have cyclic nucleotide phosphodiesterase activity. The sequence of N-terminal region was amplified by using PCR, Section 2.2.2.3. and cloned into modified pET15b_TEV vector as described in Section 2.2.2.5, Figure 3.22.
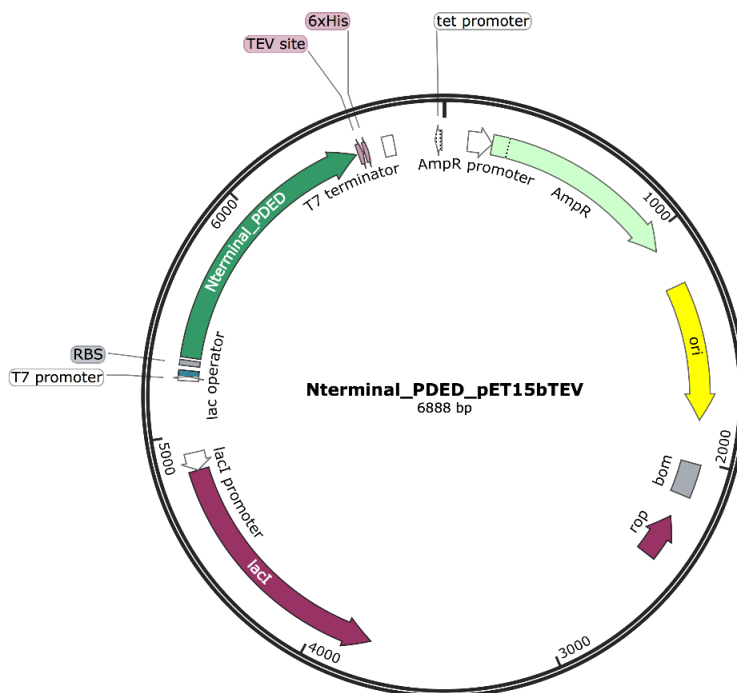
Figure 3.22: N-terminus of PDED L. infantum cloned into pET15bTEV. Vector construct was designed using SnapGene software.

The N-terminal region was expressed in Rosetta 2 (DE3) competent cells that were already transformed with chaperon plasmid pGro7. The choice of this chaperon plasmid was due to the fact that this system yielded the highest level of soluble protein material with co-expression with chaperones GroES and GroEL, Section 3.2.6 for the catalytic domain of PDED. Preliminary attempts at expression of the N-terminal construct showed that half of the expressed material was in the pellet fraction suggesting that protein was misfolded similar to the catalytic domain PDED, Section 3.2.3.

Although not fully purified the supernatant fraction was assessed for PDE enzymatic activity assay but no PDE activity was observed, suggesting that only catalytic domain of PDED has cyclic nucleotide phosphodiesterase activity.

3.2.11 L. infantum PDED, full length

The full-length gene sequence of L. infantum PDED was cloned into pOPINF vector and expressed in *E. coli* cells, Section 2.2.2.5. Analysis of the sequence showed that there were a number of rare codons, Table 3.3.

Amino Acid	*Rare Codon*	*Frequency of Occurrence in PDED full length*
Arginine	CGA	6
	CGG	10
	AGG	1
Glycine	GGA	4
	GGG	4
Isoleucine	AUA	2
Leucine	CUA	3
Proline	CCC	9
Threonine	ACG	22

Table 3.3: Sequence analysis of rare codons occurrence in L. infantum PDED full length

PDED was analysed in terms of the occurrence of rare codons. From 725 codons in the full-length sequence of *L. infantum PDED* sequence, 61 were rare codons. Therefore, it was suggested to use a bacterial strain that was particular in use for expressing proteins with rare codons. Rosetta™ 2 (DE3) host strains (DE3) were a derivative of BL21 and were used to express proteins with higher numbers of rare codons. Figure 3.22 showed an SDS-PAGE where full length of *L. infantum PDED* was expressed using small scale expression method, Ni-NTA.
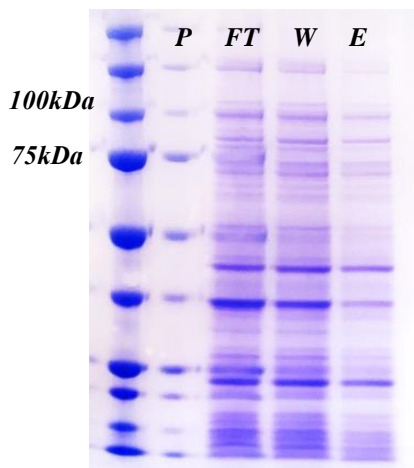
Figure 3.23: SDS-PAGE of IMAC purification of L. infantum full length PDED. Expression in Rosetta 2 (DE3) cells. P – pellet fraction, FT – flow through, W – wash fraction, E – elution fraction.

The expected molecular weight of full length PDED was 85.3 kDa. According to the Figure 3.23, SDS-PAGE showed multiple bands around expected size however it wasn't possible to determine if it was L. infantum PDED full length. However, despite the low signal, it was decided to continue with the purification to optimize purified yield. The same expression conditions of PDED catalytic domain were used for expression and solubilizing full length PDED, Section 3.2.7.
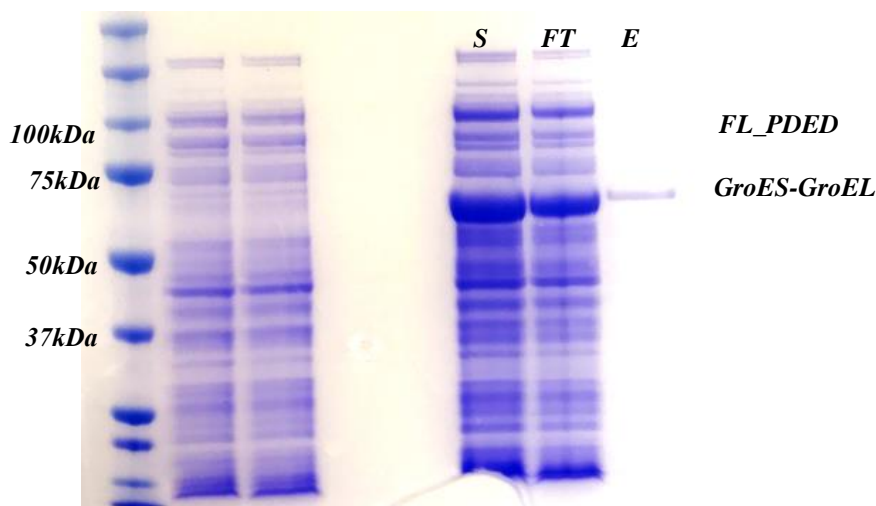
Figure 3.24: SDS-PAGE of IMAC purification of full length PDED. Expression in Rosetta 2 (DE3) cells with pGro7 chaperon plasmid

The expected molecular weights of proteins to be expressed were 70 kDa of GroES fused with GroEL and 85.2 kDa of full length PDED. According to the figure above there was good expression of the GroES - GroEL construct, whereas full length expression couldn't be detected. Hence, the next step was to optimize growth conditions. Magnesium was used as an additional additive in the 2YT media during bacteria growth and further protein expression. Magnesium acts as cofactor for many enzymes and helps proteins to fold correctly, as well as it could also be found in cell walls. Magnesium Sulphate was also believed to act as an additive that suppress selective pressure on bacteria cells during proteins expression, hence promoting expression of high molecular weight proteins *(95)*. The Rosetta 2 (DE3) cells, used in current experiment, were transformed with pGro7 plasmid for co-expression with PDED full length cloned into pOPINF vector, Section 2.2.2.4 and 2.2.2.5. Bacteria were growing in 2YT media and 2 different combinations of additives were used: 1st media - 25mM sucrose and 2mM $MgSO_4$; 2nd media - 12.5mM glucose and 2mM $MgSO_4$. Cells were induced at OD_{600} 0.6-0.7, where the cells could sustain a good level of protein expression. After induction, the cells were harvested. The 1L of cell culture with sucrose and magnesium gave 18 grams of cells, while the same amount of cell culture with magnesium and glucose gave 14 grams of cells. The protein purification buffers used were the same as used in the case of PDED catalytic domain, Section 2.2.1.7, but with the addition of 100mM of $ZnCl_2$ and 50mM Arginine. As PDEs are known to have 2 metal ions bound in the active site of the catalytic domain, the addition of Mg and/or Zn can improve protein solubility and help minimize misfolding of protein. Since the expressed protein was fused with HisTag, the IMAC purification method was used where elution fraction gave 15.8 mg of total protein where with full length L. infantum PDED protein was with the mixture of other host protein contaminants.

The purification of full length was performed similar to previous purification (Section 3.2.7.) but with an increase of the concentration of imidazole in the wash buffer in order to remove more contaminants. Figure 3.25 shows the SDS-PAGE with cleaved protein.
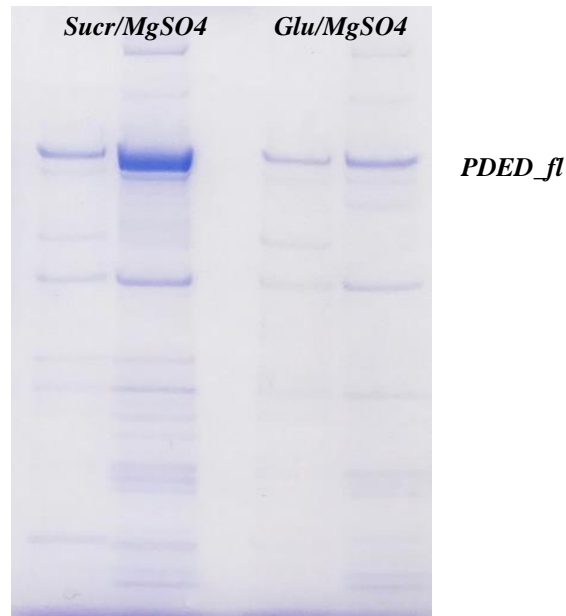
Figure 3.25: SDS-PAGE of IMAC, His-Tag cleavage of PDED full length

Figure 3.25 showed efficient tag cleavage and highlights this expression media was more preferable for full length PDED. The clear band on the gel, is at the expected molecular weight of full length PDED at 85.3 kDa. The sucrose media showed the highest level of protein solubility as compared to the glucose media.

3.2.12 *Leishmania Donovani PDED*

L. donovani PDED was the 2nd target protein where no preliminary data was available at that time. The PDED was amplified by PCR from the strain *MHOM/IN/1983/AG8* that was provided by the University of Antwerp.

Analysis of full-length sequence was performed by using Pfam and Phyre software where catalytic domain constructs were designed. The catalytic

domain was cloned into pET15bTEV modified vector, Section 2.2.2.4,
Figure 3.26.

Small scale expression conditions were used as with L. infantum PDED
catalytic domain, Section 2.2.1.7.1. The sequence alignment of *L. infantum*
and *L. donovani* PDED revealed that PDED sequence was highly
conserved between the 2 species, Figure 3.26.
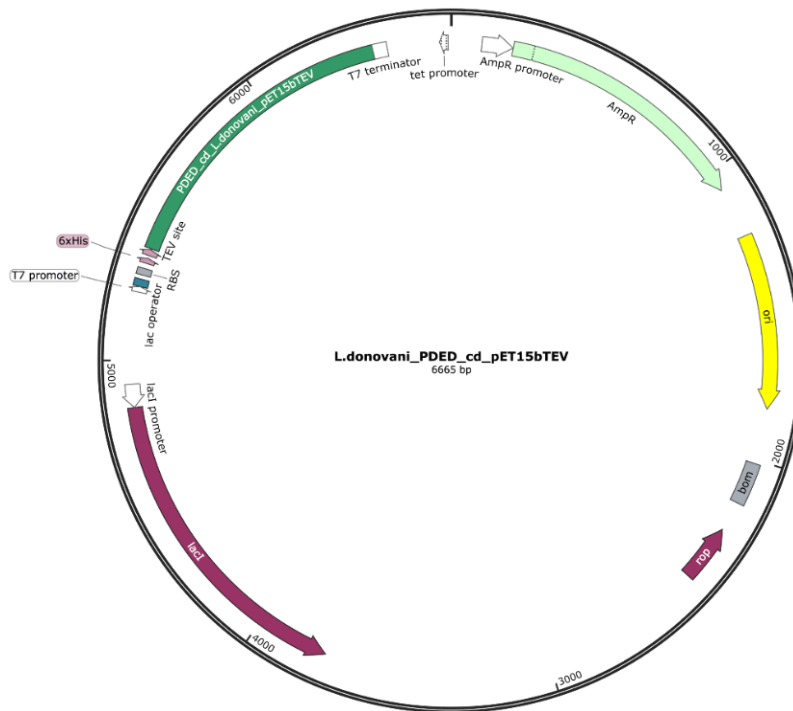
Figure 3.26: L. donovani PDED catalytic domain cloned into
pET15bTEV. Vector construct was designed by using SnapGene software.
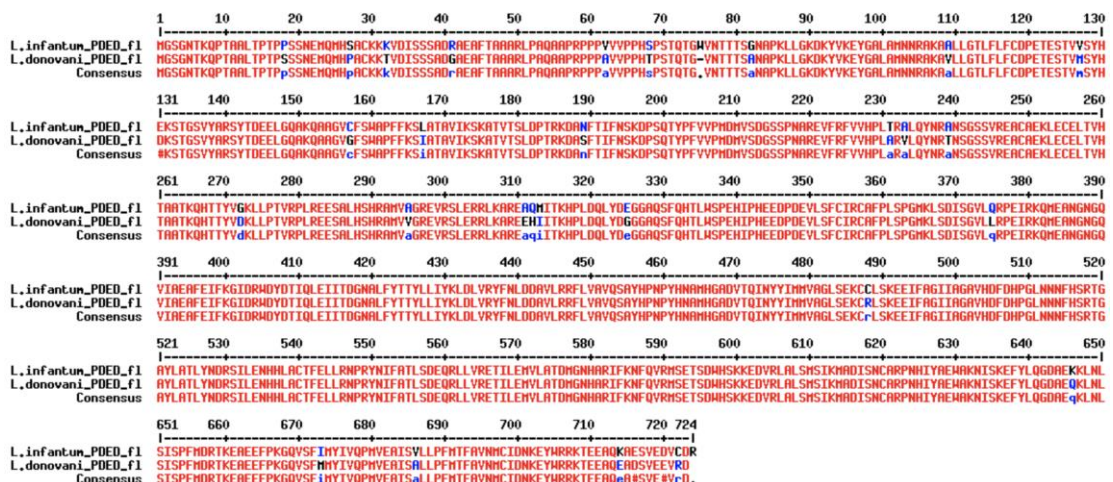
Figure 3.27: protein sequence alignment of PDED full lengths between L.
infantum and L. donovani. MultALign software was used.

The Figure 3.27 shows the protein sequence alignment between L. infantum and L. donovani full length PDED protein.
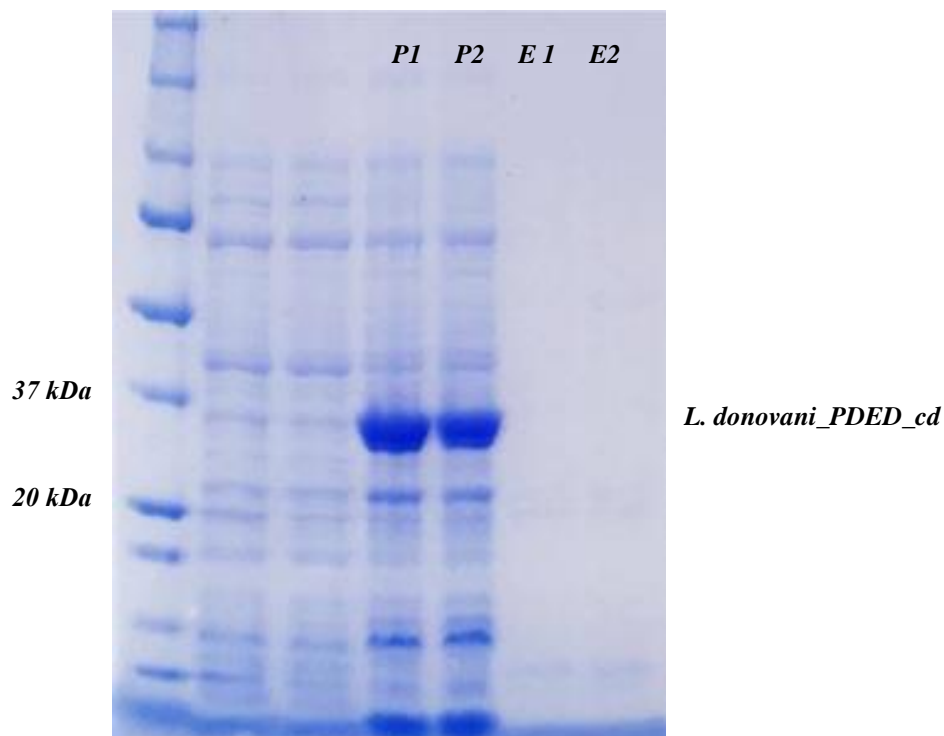
Figure 3.28: NI-NTA purification. SDS-PAGE of small scale expression of L. donovani PDED catalytic domain. P – pellet, E – elution, Molecular Weight: PDED_HisTag complex was 39.6 kDA

The following results showed that catalytic domain of PDED in L. donovani was insoluble in solution and overexpressed protein was only observed in a pellet fraction. *L. donovani PDED* catalytic domain was cloned into pCold vector suite where 2 vectors were used: pColdI and pColdTF. The same expression and purification conditions were used as in Section 3.2.7, where L. infantum was expressed. Unfortunately, L. donovani expression didn't reveal any differences in soluble yield when compared to L. infantum and the protein was also prone to aggregation as soon as it was cleaved from the trigger factor. Therefore, it was decided not to proceed further with the analysis of L. donovani PDED since the protein showed a similar pattern of poor expression and solubility. However, it was interesting to see if different species that cause another

parasitological disease would show any differences in protein behaviour since PDED proteins sequences are more diverse. The next study was to analyse protein expression of T. brucei PDED.

3.2.13 *Trypanosoma brucei PDED, catalytic domain*

The next protein target to analyse was a PDED from an alternative kinetoplastid, such as Trypanosoma brucei, the kinetoplastid that causes African trypanosomiasis. The first step was to see how similar catalytic domains of PDED are between Leishmania and Trypanosoma. Figure 3.29 shows the protein sequence alignment of PDED catalytic domains where 2 sequences appear genetically distinct from each other with less than 60% identity. A similar broad approach to that performed in Section 3.2.1.5, was carried out with T. brucei PDED co-expression with 5 chaperones plasmid sets and analysed using standard small-scale expression conditions.
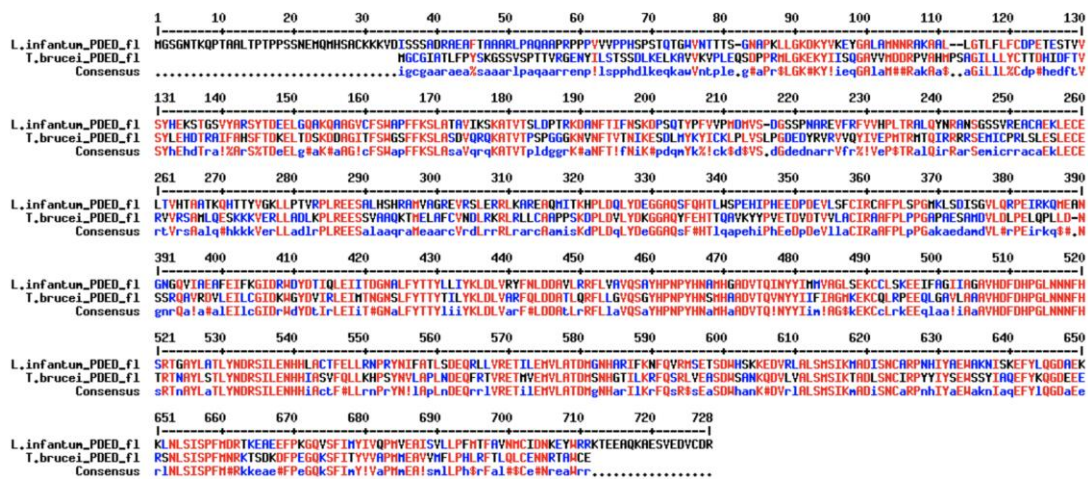
Figure 3.29 protein sequence alignment of PDED catalytic domains from L. infantum an T. brucei.
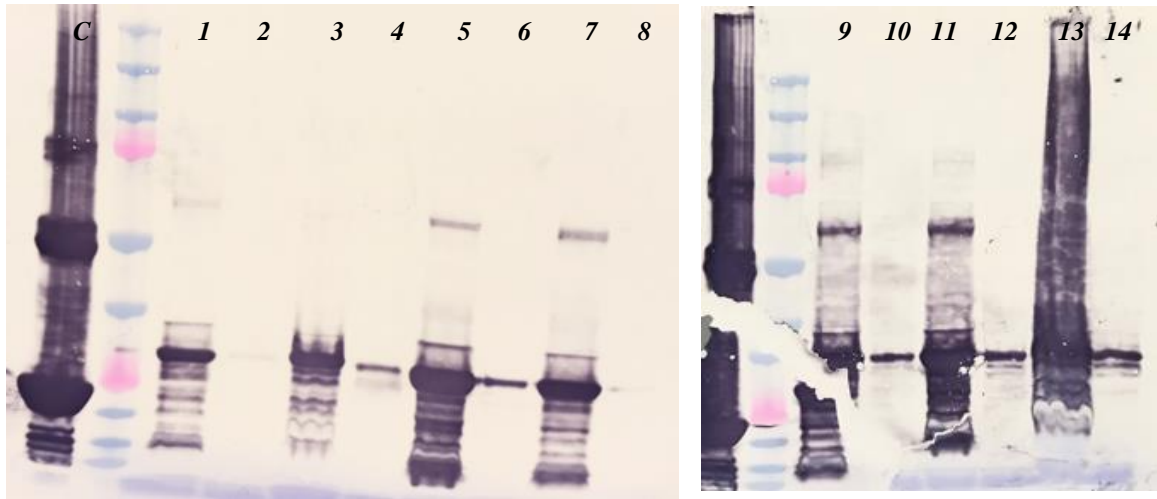
Figure 3.30: NI-NTA small scale expression. SDS-PAGE small scale co-expression of catalytic domain PDED from T. brucei

Key:
1.   pG - Tf2, TB media, ethanol, pellet

2.   supernatant

3.   pG-KJE8, TB media, ethanol, pellet

4.   supernatant

5.   pG - Tf22YT, sucrose, pellet

6.   supernatant

7.   pG - Tf22YT, ethanol, pellet

8.   supernatant

9.   pG-KJE8, TB media, ethanol, pellet

10.   supernatant

11.   pG-KJE8, 2YT media, sucrose, pellet

12.   supernatant

13 - pG-KJE8, TB media, sucrose, pellet
14 – supernatant

Amplification and cloning steps were initially performed by Dr. Abhi Singh, where a construct of comprising residues 412-670 showed the best expression level. This construct was cloned into pET28a vector with HexaHisTag at the N-terminus. The pET28a_PDED_Tbr construct was co-expressed with 5 chaperones plasmids. Figure 3.30. The most soluble

protein was obtained with its co-expression with chaperon plasmids: pG-KJE8 and pG - Tf2.

Small scale expression of T. brucei PDED catalytic domain revealed that the most preferable conditions for solubilizing of PDED protein was its co-expression with chaperon plasmid pG-KJE8 in TB media in addition of 25mM sucrose where sample 12 showed the highest amount of soluble protein with the least degradation pattern in comparison to other samples, Figure 3.30. The next step was to scale up expression in 2L of bacteria culture, where the same growth conditions were used. Attempts to purify T. brucei PDED catalytic domain, with a series of purification steps, such as: IMAC, IEC and SEC chromatography was planned. However, during the 1st step of purification (IMAC) PDED was found in a pellet fraction only indicating that misfolding mechanism was predominated leading to insoluble protein being produced.

3.3 Conclusion

PDED was a potential novel drug target that previously had not been characterized and expressed. Although PDEs were known to be an important enzyme that plays a crucial role in parasitic survival, since the late 1980's, there was little information regarding the role of parasitic PDED and hence the focus of this study. As previously mentioned, kinetoplastid parasites such as T. cruzi, T. brucei and Leishmania species have highly conserved sequences among each class of PDE: PDEA, PDEB, PDEC and PDED *(96).* Some preliminary work in amplification and cloning was done in T. cruzi and T. brucei where different constructs were designed with various mutations being introduced. The main issue with the initial PDED expression was its solubility in solution and the proteins only appeared in a pellet fraction and not in the supernatant hence further purification and characterization was not possible. As mentioned

earlier, preliminary work was done on parasitic PDEDs by Dr. Abhi Singh (T. brucei) and by Dr. Susanne Schroeder (T. cruzi) where multiple protein constructs yielded no soluble protein.

In current study the primary focus was on Leishmania species, namely L. infantum and L. donovani.

This chapter describes the first expression and purification of an active soluble parasitic PDED catalytic domain construct from L. Infantum. This protein could be expressed and purified and was active when it was fused to the trigger factor, however once cleavage was performed there was a significant loss in the activity due to protein degradation.

Full length PDED was also expressed and purified however the protein was not sufficiently stable to perform enzyme assays.

These results demonstrate that PDED protein folding machinery required significant optimization, including addition of fusion tags and chaperones as well as extensive investigation of expression and purification including additives. The approach used for L. Infantum also yielded low levels of soluble protein for L. donovani PDED however the protein suffered from the same instability issues after removal of tags. Further experiments utilizing similar approaches on T. brucei PDED yielded substantial improvements in soluble expression levels in a chaperone dependent manner but once again proteins were liable to degradation.

In summary novel proteins were cloned, expressed, purified and characterized. New protein expression approaches were developed, as standard approaches only yielded insoluble proteins. This approach allowed the protein to be further characterized by using enzymatic activity assay.

Chapter IV

Developing and testing a Fluoro fragment library for use by NMR, X-ray crystallography and SPR, exemplified on human PDE5.

4.1 Introduction

This chapter describes the generation and initial analysis of a new fragment library of fluorinated fragments comparing hit rates using different biophysical screens.

4.1.1. Development of a Fluorine Labelled Fragment Library

Substitution of hydrogen atoms by fluorine in drug compounds is a widely used approach in drug discovery, which can improve drug potency as well as its pharmacokinetic properties and metabolic stability. Fluorine containing compounds and fragments can also be exploited by fluorine NMR. The much simpler fluorine spectrum, compared with proton spectra, allow mixtures (or cocktails) of larger numbers of fragments to be screened, as compared to proton ligand observed NMR, thus reducing the time to analyse hundreds of potential fragments. We report here on the results of screening a library that was co-developed between the groups of Dave Brown at the University of Kent and Charles River, and Maybridge. The library selection was filtered from 5227 compounds of all fluorinated compounds in the Maybridge collection. The physicochemical property thresholds were used *(97)* ensuring certain criteria were met:

$$140 < MW < 300,$$

$$logP \leq 3,$$

$$rotatable\ bonds \leq 4,$$

$$rings \leq 4,$$

$$HBA/HBD \geq 3$$

Overall 954 compounds were selected. Then fingerprints were calculated using MOE syl script written by Andrew Henry, Chemical Computing Group. The algorithm was based on that was of Vulpetti *(98)*. The results were as follows: 317 clusters at 85% similarity and 768 clusters at 90% similarity. This was followed by removal of potentially mutagenic compounds, resulting in 590 compounds for further analysis.
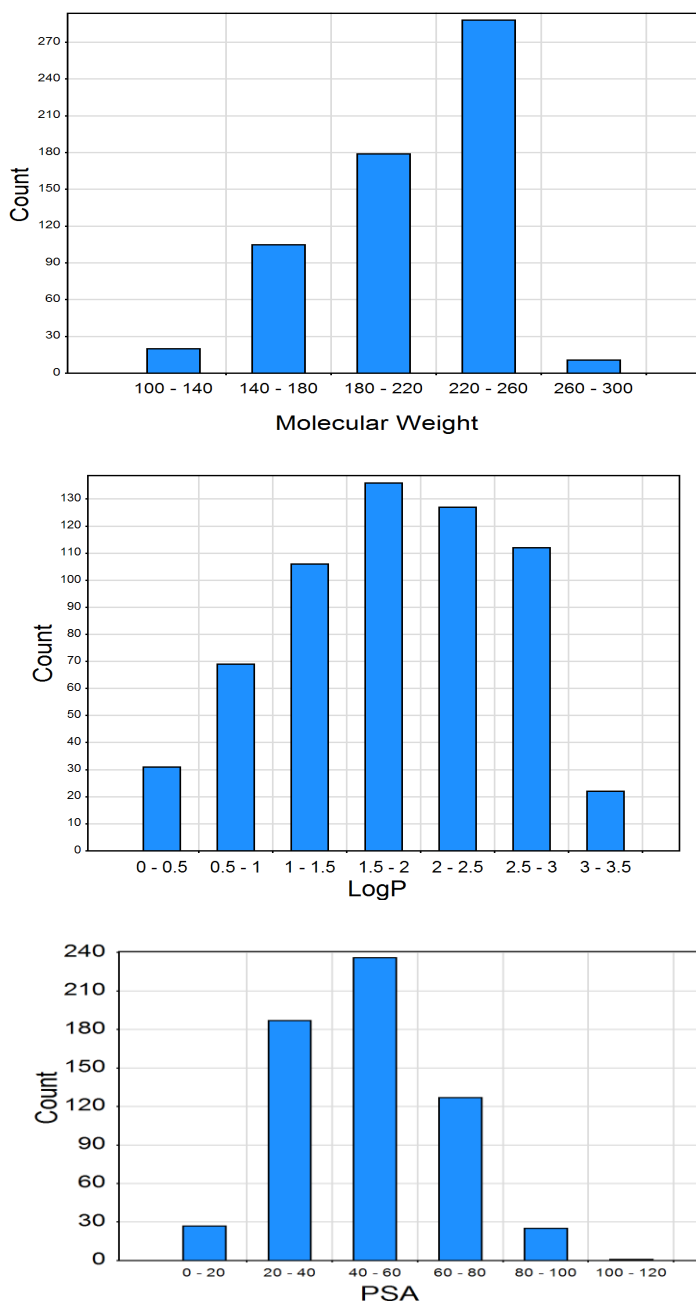
Figure 4.1: Fluorinated Library Properties, Maybridge

Figure 4.1 shows the properties of the final fluoro library selected after solubility and purity tests were checked using NMR spectra recorded for all individual compounds (work performed by Denisa Hoxha at CRL in Dave Brown's lab). The majority of compounds showed good solubility (compounds that showed poor solubility in DMSO at 100 mM were

excluded), Figure 4.2. The final fluorine file contained 421 compounds that were used in NMR, SPR and XChem experiments.
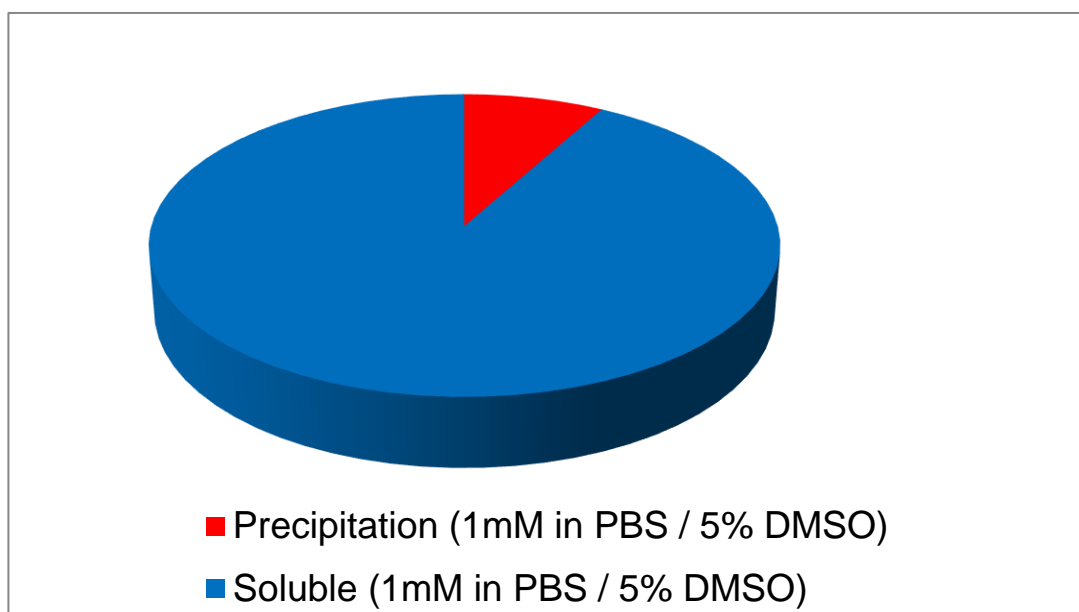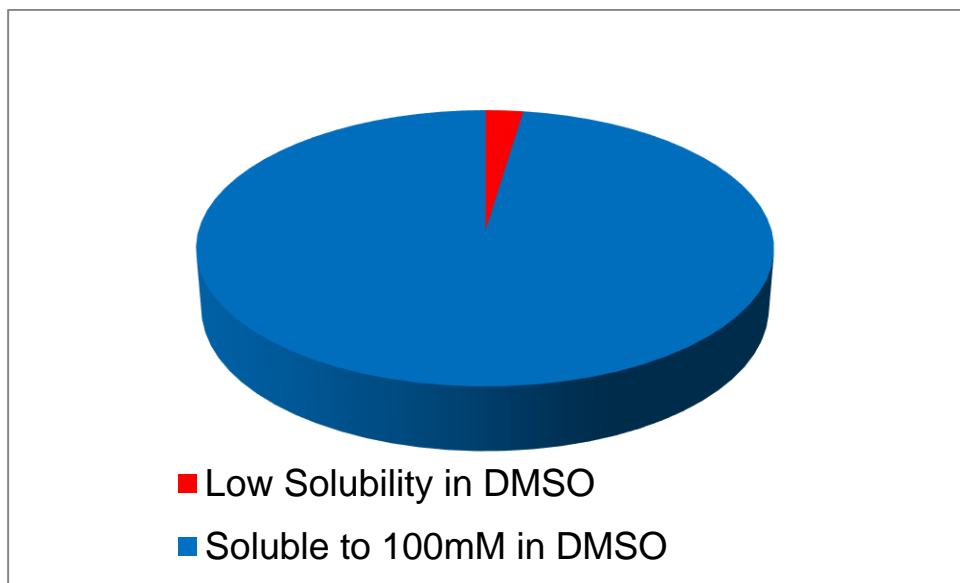
Figure 4.2: Fluoro fragments solubility properties

Figure above 4.2 showed the proportion of precipitated compounds towards soluble compounds. As it was important to exclude fragments prone to precipitation from the final library list. When the library was compiled, it was important to test it on a well-behaved target on which fragment screening had previously been performed and for which a

suitable soakable crystal systems for X-ray crystallographic follow up was already in place. There were two proteins selected that were used in SPR screening in collaboration with Steve Irving at Charles River, namely human PDE5 (catalytic domain) and human bromodomain - containing protein 4 (BRD4 - BD1). The fluoro library was screened by SPR at 100 μM against biotinylated proteins that were captured on a CM5 SPR chip derivatised with streptavidin. The following figure summarise the results for 2 proteins, Figure 4.3. The line that can be observed from the Figure 4.3 was an arbitrary line that was used to define an Rmax value where
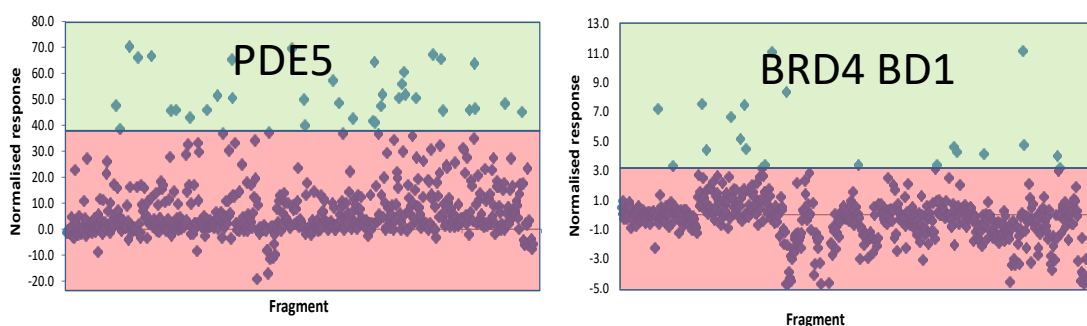
Figure 4.3: SPR screening of Fluoro fragment library using human PDE5 catalytic domain and BRD4 BD1 domain

"good binders" were separated from the "poorer binders".

According to the Figure 4.3, fragments from fluoro library were well behaved in the system and the majority of binders showed 'square wave' sensograms which is indicative of fast on/off binders. Fragments which exhibited non-specific binding effects were rejected from the final list of hits. The hit cut off for PDE5 was set at a normalised response of 38 RU for PDE5 as a relatively large number of hits were obtained (giving a hit rate of 5.9 %), while a lower hit rate (2.9 %) was observed for BRD4 – BD1 with an RU cut off of 3.2. The study also allows comparison of the hits and hit rates of the different biophysical techniques but also further analysis of the binding mode of fragments against human PDE5 which is of interest as these may be generic PDE binders.

4.2 Results

4.2.1 NMR Data

NMR fragment screens can be a very time-consuming process. In the current experiment the fluoro fragment library was used where large number of compounds would be screened in each mixture, in order to minimise the running time and consumables. Each mixture contained between 8 to 23 compounds. There were NMR constraints that had to be followed when composing the compound mixtures: maximum spectral window 30 ppm, ideally a minimum of 0.5 ppm between fragments, fragments to be mixed in a final volume of 650 μl in D_2O and final ligand concentration in mixtures was 0. 1 μM (stock concentration of 100 μM). After careful examination of each fragment and preparing mixtures, 28 fragment cocktails were prepared. The example of mixture 14 is illustrated in Figure 4.4.

There were 2 samples made of each mixture where 1 was without protein and used as a control and 2nd one with the PDE5 protein. The protein concentration was 5 μM in the 650 μl mixture, and the total protein usage for the entire screen was 4 mg.
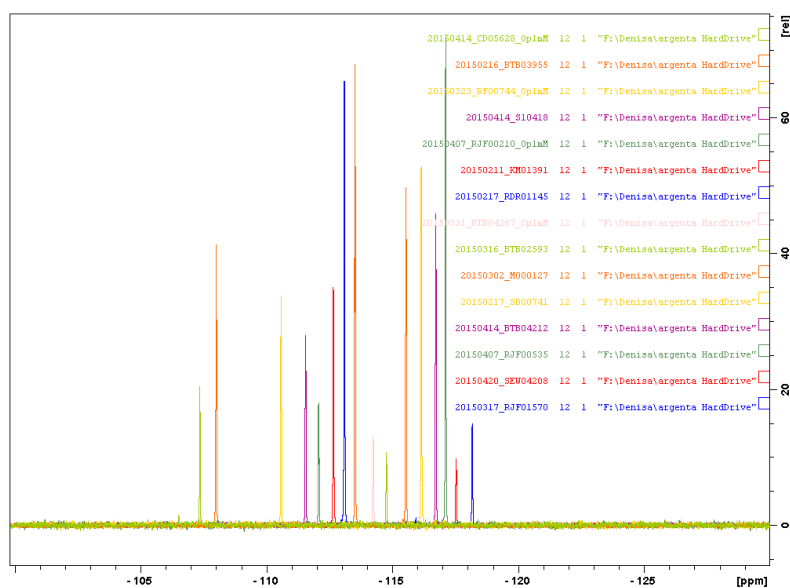
Figure 4.4: NMR Mixture 14
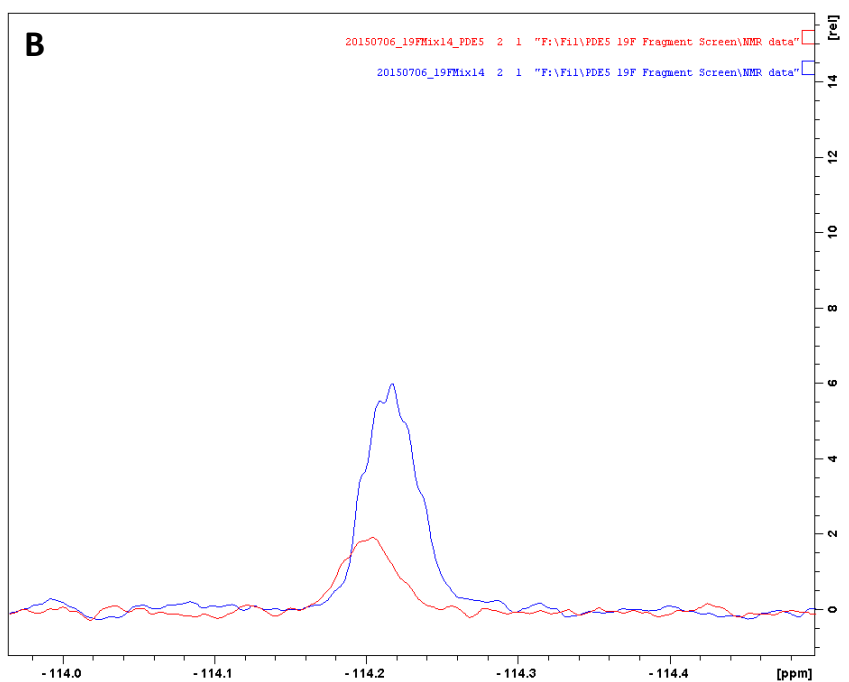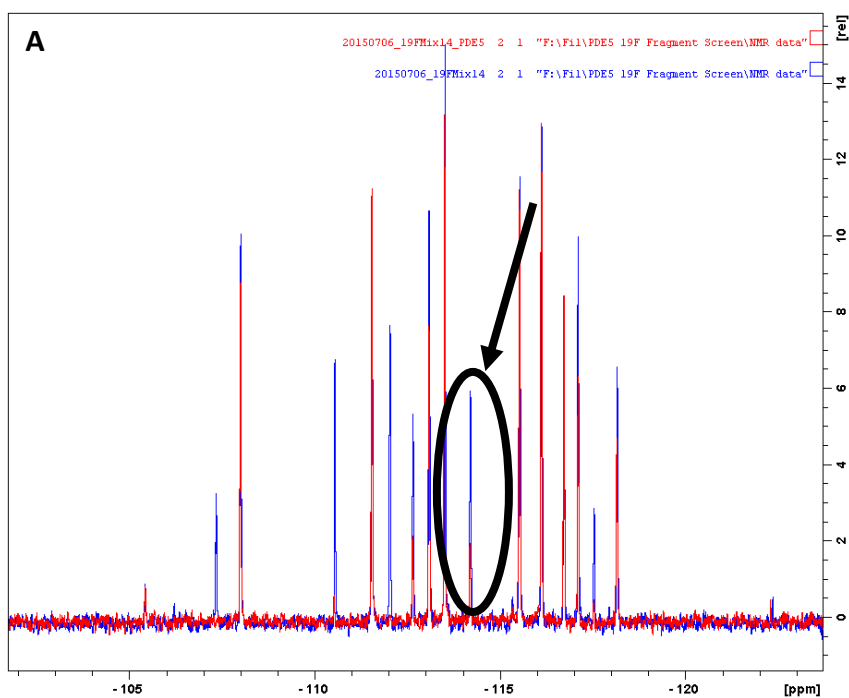
Figure 4.5: 19F CPMG experiment of mixture 14. A) overlay of mixture with and without PDE5 protein, B) drop in signal of binding. Key: Blue Peak – mixture without the protein, Red Peak – mixture with the protein

Each fragment mixture was run through the NMR 19F CPMG (spectra where Fluorine atom is used to identify a binding event) experiment where

the total run time was 72 hours. To identify if there was binding of one or more fragments in a mixture, the peaks of two different NMR runs compared with the spectrum of mixture of fragments would be overlaid onto the spectrum of the equivalent mixture in the presence of the protein (hPDE5) to detect if there was any drop in the peak height, Figure 4.5 A and B.

Figure 4.5 A represents the spectrum of mixture 14. The overlaid spectrum shows the peak intensity drop, where red peak is the run of mixture 14 with the hPDE5 and blue peak is the mixture without the protein, Figure 4.5 A. Figure 4.5 B represents one of the peaks corresponding to one fragment from the whole mixture. It can be observed that there was a drop in the signal which means there was a binding between hPDE5 and the fragment. To identify the intensity of peak drop TopSpin software was used where integral values were calculated by using chemical shift's values from NMR runs. A > 50% drop was considered as a strong binder and < 50% as a weak binder. In the current example, the identified fragment bound to hPDE5 had a peak drop of 70.71% suggesting it was a strong binder.

Out of 421 fragments that were run in mixtures by NMR, 101 fragments showed a peak intensity drop of > 95%. These fragments were classified as strong binders, however to further test this and study the interactions in more details X-ray crystallography was required.

34 fragments were selected for further validation by X–ray crystallography, on the basis that those fragments were identified as strong binders by either NMR and/or SPR screens and some classified as such by both methods, Table 4.1.

Fragment	NMR	SPR	
	19F CPMG Chemical Shift	Kd (uM)	% of max Response
AW01202	-115.4 ppm		
BTB00588	-117.7 ppm	4.2	64.4
HTS04341	-117 ppm	9.2	60.6
BTB05106	-119.5 ppm	89	45.2
BTB05938	-108.8 ppm		
HTS05996	-109.5 ppm		
MO01209	-113.5 ppm		
HTS03409	-61.7 ppm		
CD04945	-116.2 ppm		
SP01339	-122.3 ppm		
SEW05128	-62.5 ppm		
RF00744	-110.6 ppm		
CD08996	-115.4 ppm		
MO07029	-119.9 ppm		
S11211	-112.6 ppm		
RJF00210	-112.1 ppm	60	45.9
HAN00244	-70.5 ppm		
S08246	-114.5 ppm	12.7	50.7
HTS10254	-109.7 ppm	46	50.7
CD07436	-107.7 ppm		

KM09455	-64.4 ppm		
RF04864	-73.6 ppm		
CD11422	-119.7 ppm		
KM02354	-117.6 ppm		
BTB07539	-121.7 ppm		
SPB06580	-118.9 ppm	12	57.4
RDR03354	-59.9 ppm	32.5	48.7
BTB02754	-105.7 ppm	60	46.6
HTS04838	-113.3 ppm	69	67.4
SEW05363		ND	45.8
BTB07539	-121.7 ppm		
KM06103	-121.4 ppm		

Table 4.1: 32 selected hPDE5 Fluoro fragments by NMR and SPR screens for manual crystallization

22 fragments out of 34 were identified as strong binders in NMR but not in SPR, 2 out of 34 fragments were selected for being strong binders in SPR but not strong binders by NMR, 10 were strong binders in both NMR and SPR screens. There were 11 compounds that were tested in both SPR and NMR screens, Table 4.1, and were further validated by manual crystallisation.

4.2.2 Correlation between NMR and SPR Data

The highest SPR response value was observed with fragment CD07988, where the NMR screen showed that it was a weak binder and hence wasn't chosen for manual data collection. The lowest value for SPR run was

obtained for fragment CC53013. That fragment wasn't run on NMR due to the reason that it didn't pass quality control (QC) test where each fragment was run individually and showed more then 1 peak meaning of possible contamination. The second lowest SPR data was obtained with fragment SPB02475 where NMR run also confirmed it was a weak binder. TG00013 fragment showed the lowest Kd value. This fragment was run on NMR in mixture 22 where the chemical shift identified it as a strong binder. Another fragment that was also picked up by SPR and NMR screening was fragment SPB06580. RJF00210 fragment was also identified by SPR and NMR screens as a strong binder. Next fragment that was identified as a strong binder by NMR was RF04864 with SPR run couldn't determine its Kd value. RDR03354 was identified as a strong binder with NMR and SPR screens. HTS04838 was run in a mixture 18 where NMR chemical shift and SPR Kd values showed that it was a strong binder. Next fragment that was determined as a hit was HTS04341 with both screens, however NMR identified it as a weak binder. CD07436 was run on NMR and SPR, whereas SPR Kd value was not determined. Fragment CD04945 was identified as a PDE5 hit on NMR run but the SPR screen did identify it as a binder but no Kd value could be determine. BTB06033 was identified in both screens. Therefore, this compound was classified as a strong binder according to the data obtained from the two screens. BTB05106 was classed into a group of strong binders were both values from NMR and SPR suggested it. Next fragment that was used in both screens was fragment BTB02754 where SPR data was as well as NMR chemical shift. Fragment BTB00588 was identified as a strong binder with NMR and SPR runs. Current subchapter compared data, Table 4.2, that was obtained from SPR and NMR screens where one method determined compound as a stronger binder while the other one classified it as a weak binder. Those

fragments that showed different results were excluded from the manual
crystallization.

BTB00588	4.2 uM	50%
CD07988	66.2 %, 340 uM	41%
CC53013	22.9%	X
SPB02475	37.2%	38.14%.
TG00013	0.9 uM	50%
SPB06580	12 uM	50 %
RJF00210	60 uM	50 %
RF04864	X	50%
RDR03354	32.5 uM	59.9 ppm
HTS04838	69 uM	50%
HTS04341	9.2 uM	117 ppm
CD07436	X	107.7 ppm
CD04945	X	50%
BTB06033	178 uM	50%
BTB05106	89 uM	50%
BTB02754	60 uM	105.7 ppm

Table 4.2: comparison of SPR and NMR data

4.2.3 Manual Soaking

There were 34 fragments that were selected for manual X-ray soaking,
where only 32 diffraction data was collected as other 2 soaks didn't
survive the shipment to Diamond Synchrotron. Initial human PDE5
crystals were provided by Colin Robinson from Charles River company.
New trays were created with similar reservoir conditions to the crystal
trays. Fragments were created as singletons and were added to 50 µl of
reservoir solution (-IPA) to a final concentration of 1 mM (stock =100 mM
therefore 0.5 µl of fragment was added) to make a ratio of 1:3. Then, 1
crystal was transformed to 1 µl of the fragment solution and left soaking
for 1-2 days. After, crystals were transformed to a cryoprotectant (30%
glycerol) and transferred to pucks. Out of 34 compounds 32 fragments
from Table 4.1 gave diffraction data. The crystallographic data was further
refined by other lab members.
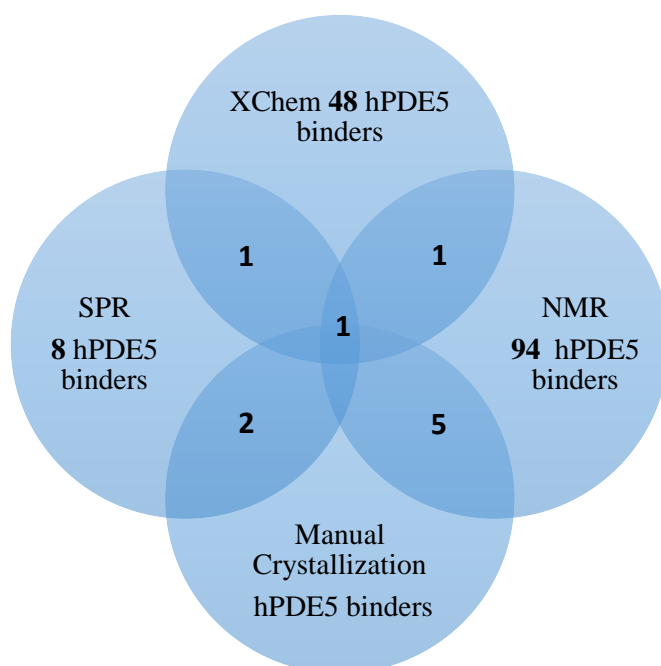
Figure 4.6: Summary representation of number of fragments that were validated by each biophysical method

PDE5 X-ray structures were solved by other lab members hence the 3D structures were not included in current chapter.

4.2.4 XChem Data

XChem (https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html) is a platform developed at the Diamond Light Source Synchrotron for high throughput, automated crystallography. Users to first identify suitable conditions (DMSO tolerance, ligand soaking time and ligand concentration) and then screen a large number of fragments once a suitable soaking crystal system has been established. Software developed for the platform, XChem explorer, combined with use of PanDDA maps also allows the identification of very week binding modes with as little as 10% occupancy that are unlikely to be identified by NMR or SPR.

The fluoro file were also analysed crystallographically by performing XChem on the library where just 376 compounds were selected to align with the 384 well low volume echo dispensing plates. The main aim was to further validate the final list of strong PDE5 binders.

Out of 376 compounds only117 soaks gave high quality diffraction data, representing 31 % of all fragments. 51 fragments were identified as binders through the XChem pipeline. Out of all fragments found in XChem only 2 were also identified by either SPR and/or NMR techniques and only 1 fragment – HTS04341 was identified by all 4 techniques, Figure 4.6. A clear observation is that XCHEM seemed to find a number of weaker hits (which were not listed in NMR or SPR experiments as these hit lists were focussed on strong binders) but failed to detect nearly all strong binders previously identified. A possible reason is crystal lattice disruption by more soluble potent binders at high soak concentrations and long soak times.

4.3 Conclusion

The chapter describes the development and testing of a fluoro – fragment library, with the aim of validating the library and examining if fragments would be detected by more than one biophysical technique. The compounds for the fragment library were provided by Maybridge and the initial step was to check fragments suitability for inclusion. It was very important to identify if fragments were soluble enough and testing there was no precipitation since encountering such problems during the actual experiment would reduce experimental accuracy. The initial library was comprised of 590 fragments and was cut down to 421 fragments when each fragment was individually analysed on NMR spectra. Since most fragments contained one Fluoro atom in its structure then 1 peak should be visualized on NMR spectra only. Obviously, there would be additional peaks for the two and three fluorine containing examples. If additional peaks would be obtained it would indicate that there were contaminations. The final list was made of 421 fragments that were used for compiling mixtures. Since only 1, 2 or 3 peaks would be observed from each

fragment it was suggested to combine 8 – 23 fragments in one mixture without spectral overlap, in order to minimise the NMR running time as well as protein consumption. After spectra analysis using TopSpin software 101 fragments were classified as strong binders with more than 50 % drop in signal when hPDE5 was bound to Fluoro fragment. It was then important to see if the same fragments would be validated by SPR, XChem and manual crystallization. By comparing data that was obtained after performing SPR, XChem and manual crystallization, only 1 fragment was identified by all 4 biophysical methods. XChem gave the least positive data as only 2 fragments were identified as hPDE5 binders out of 376 compounds that were loaded onto XChem pipeline. One of the main reasons could be crystal lattice disrupted by more soluble potent binders at high soak concentrations and long soak times. Even though, that hPDE5 crystal system is believed to be a robust system for fragment screening, issues were encountered during the XChem, suggesting that experimental optimisations have to be done in order to make XChem more suitable for high throughput data collection with this protein target and fragment file.

Chapter V

Biophysical Analysis of TbrPDEB1 using Biophysical techniques: XChem and SPR

5.1 Introduction

TbrPDEB1 is a validated drug target *(99)*, where inhibitors of this enzyme lead to an increase of intracellular levels of the secondary messenger cAMP, causing the cell death. It has been shown by Seebeck et al that a dual knockdown of PDEB1 and PDEB2 led to altered cell division in parasite, resulting in the cell death and hence elimination of infection *(18).* Previous studies used human PDE inhibitors as a starting point for developing a new generation of antiparasitic drugs that could be potent and selective parasitic PDEs inhibitors *(100)*. The most advanced PDE inhibitors that have been designed for antiparasitic treatment were piclimast derivatives developed by the group of Leurs at Vrije University of Amsterdam (VU) *(46)*. However, a major challenge that still remains is to improve selectivity towards parasitic PDEs over human PDE isoforms. A key strategy would be to target the unique P – pocket found in parasitic PDEs that is absent in human PDEs.

The aim discussed in the current chapter is to identify and analyse TbrPDEB1 ligand binding using 2 structural biology techniques, namely SPR and X – ray crystallography.

5.2 Crystallography with XChem using the DSI – Poised fragment library

There were 466 fragments that were used during the XChem experiment where individual TbrPDEB1 crystals were soaked with a single fragment through use of the Labcyte Echo acoustic nanoliter dispenser. Then, there was a manual inspection performed by me to see how many crystals survived. During the process 20% of crystals were lost, presumably through lattice disruption by the fragment dispensed resulting in them dissolving or severely cracking; or mounting failed due to poor or reduced mechanical robustness. After mounting and cryo protection the crystals were cryo cooled to 100 degrees K and stored in unipucks prior to X-ray data collection on the I04-1 beamline at DIAMOND. Data sets were collected following automatic loop centring using a standard protocol of 1800, 0.1degree oscillation images with 0.1 sec exposure. During data collection, some crystals were clearly too small in size to yield high resolution data sets and some suffered radiation damage leading to failure of the automatic data processing with the XIA2 pipeline. All together there were 466 fragment soaks and 50 soaks with simple DMSO as controls. 314 crystals were mounted, cooled and exposed to the X-ray beam. XIA2 and DIMPLE analysis showed that out of 314 data sets, 21 data sets failed to process. The final number of diffraction data that were subjected for PanDDA inspection was 293. Out of 293 samples, 75 were rejected through manual inspection of the processing results by the fact that samples were probably non – isomorphous or diffraction limits were too low. The highest resolution was 1.7 Å and the lowest utilised was 3.5 Å. From the remaining 218 data sets, 136 samples were classified as interesting data with evidence of fragment binding manually observed. Further, more detailed manual inspection of the PanDDA maps resulted in

61 high confidence hits and 10 medium confidence hits. PanDDa is the Pan-Dataset Density Analysis that enables users to investigate the bound compound in details with less or no noise. This work was performed on the DIAMOND computational cluster. These positive data were then downloaded to our local Linux machine and further analysed using the CCP4 software suite. In some cases, the map was too weak to unequivocally place a ligand hence that data was omitted from further study. Follow up work was focussed on 20 data sets where each structure was refined and where the binding mode could be clearly observed. There were certain parameters that were controlled and discussed before the actual experiment based on initial XChem screening. The soaking time remained as in the test run with the DSI Poised library subset, which were: 2 hours and 4 hours. Since compounds were diluted in a DMSO it was also important to be consistent with that as DMSO could also potentially destroy the crystal lattice. From the screening experiment run where TbrPDEB1 crystals were soaked with a small subset of the DSI – Poised library the statistics was as following:

- 20% DMSO with 3 to 4 hours of soaking gave more mounting consistency (less crystal damage)
- 40% DMSO with 3 to 4 hours of soaking time
- 25% DMSO with 3 to 4 hours of soaking time
- 20% DMSO give total cryo protection as well

According to the statistics 20 % DMSO was the most preferable concentration for crystal to survive as well providing a cryo protection. Therefore, it was decided to use 20% of DMSO as a target concentration in soaking.

It was previously reported that TbrPDEB1 forms a unique open cavity (P – pocket) between helix 14, helix 15 and the M – loop. That cavity was also observed in structures of LmjPDEB1 and TcrPDEC *(41)*. As previously

mentioned since the P – pocket is a unique feature to parasitic PDEs it could be used as a selectivity feature for designing new drug candidates against parasitic PDEs. The binding site for TbrPDEB1 consists of the following residues: Gln874, hydrophobic clamp - Val840 and Phe877 and aromatic residues - Phe844 and Phe880. The P – pocket binding site is characterised by composition of such residues as Ala837, Thr841, Tyr845, Asn867, Met868, Glu869 and Leu870, Figure 1.15 *(46)*.

Therefore, in the current experiment, fragments will be subdivided into 2 main groups: binders in hydrophobic clamp and "selective" fragments that bind in or towards the P – pocket.

5.3 Ligand – Protein Binding Mode

TBrB1 – Ligand 0074 Crystal Structure
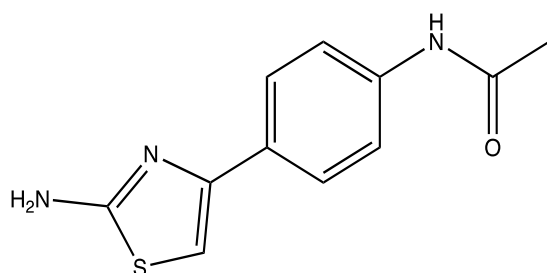
Figure 5.1: Ligand 0074, N-[4-(2-amino-1,3-thiazol-4-yl) phenyl] acetamide
Characteristics: $C_{11}H_{11}N_3O_5$
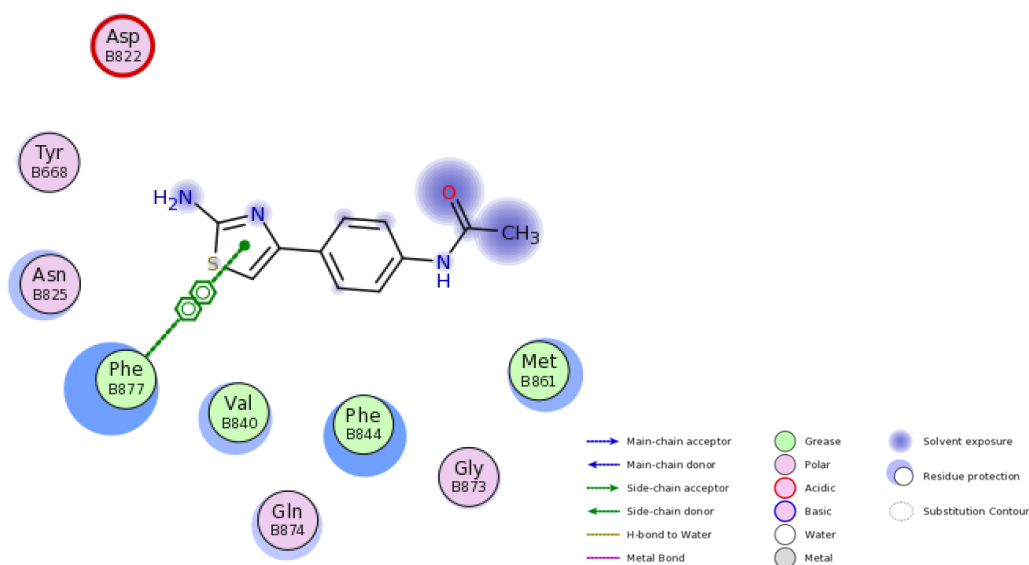
Figure 5.2: FLEV plot produced by COOT of the binding site of chain B, first binding site.
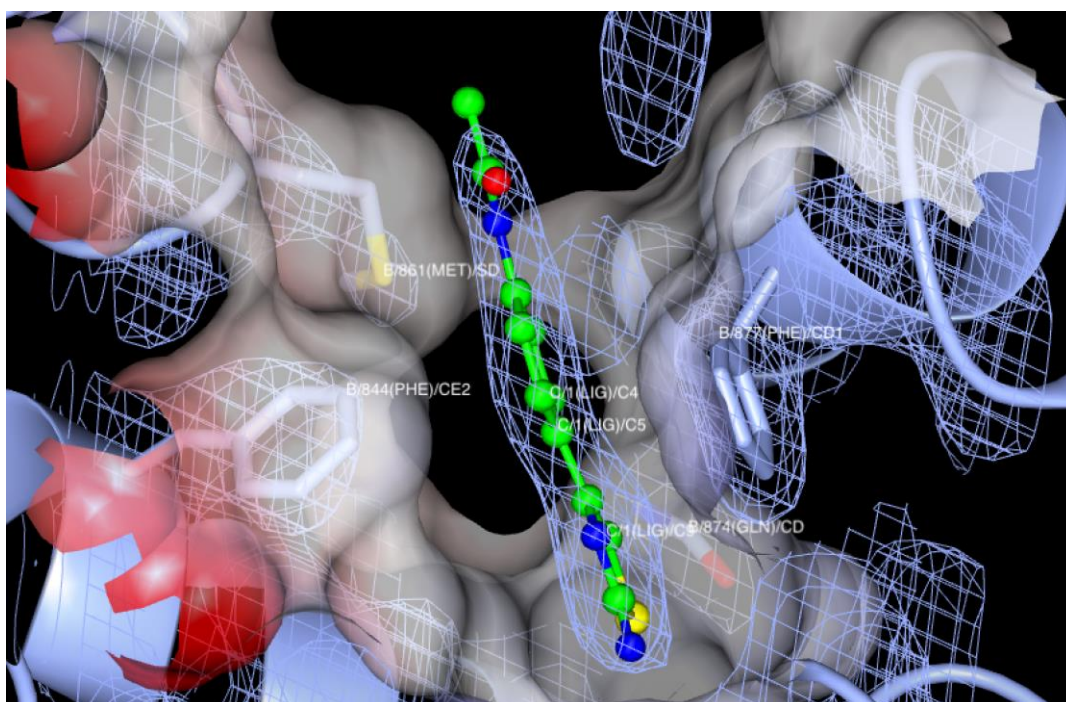
Figure 5.3: Surface representation of the ligand binding site with the ball and stick representation of TbrB1 – 0074 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ, chain B, first binding site in the hydrophobic clamp

The 0074 fragment data set was selected for further refinement where the fragment was found in 3 binding sites: 2 in chain B and 1 in chain A.

Figure 5.2 represents ligand binding site environment where 9 amino acids surround the 0074 ligand. Chain B has 2 binding sites, in the first case the ligand was bound to a hydrophobic clamp between residues Phe877 and Phe844, Figure 5.3. The ring structure of the 0074 compound forms hydrophobic interactions with Phe877. Other protein residues such as: Asp822, Tyr668, Asn825, Val840, Phe844, Gly873 and Met861 also in close proximity to the ligand. The benzene ring of 0074 offers substitution vectors towards P – pocket but further computational analysis is required in order to model a possible analogue that will grow towards the P – pocket to increase its selectivity.
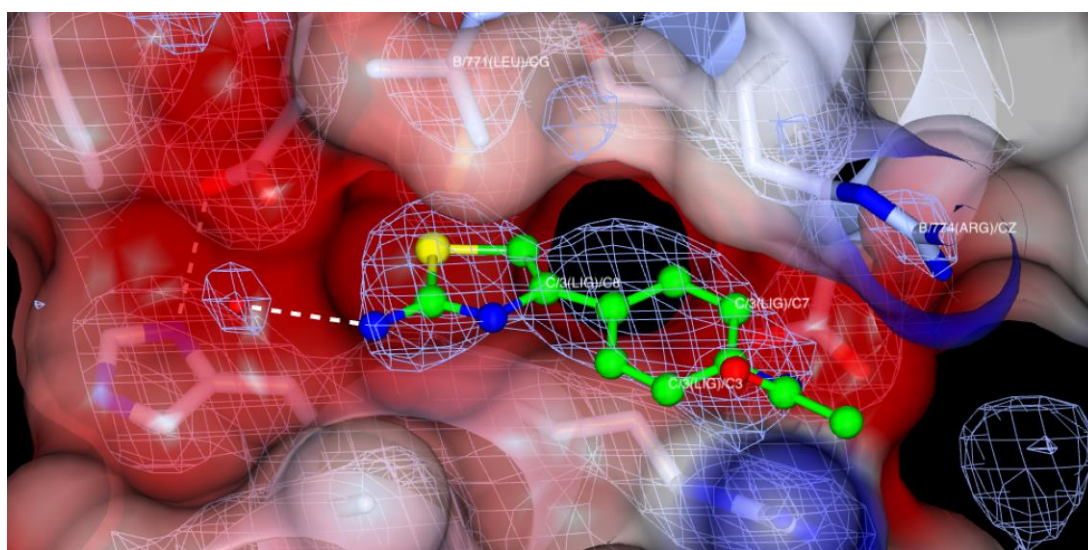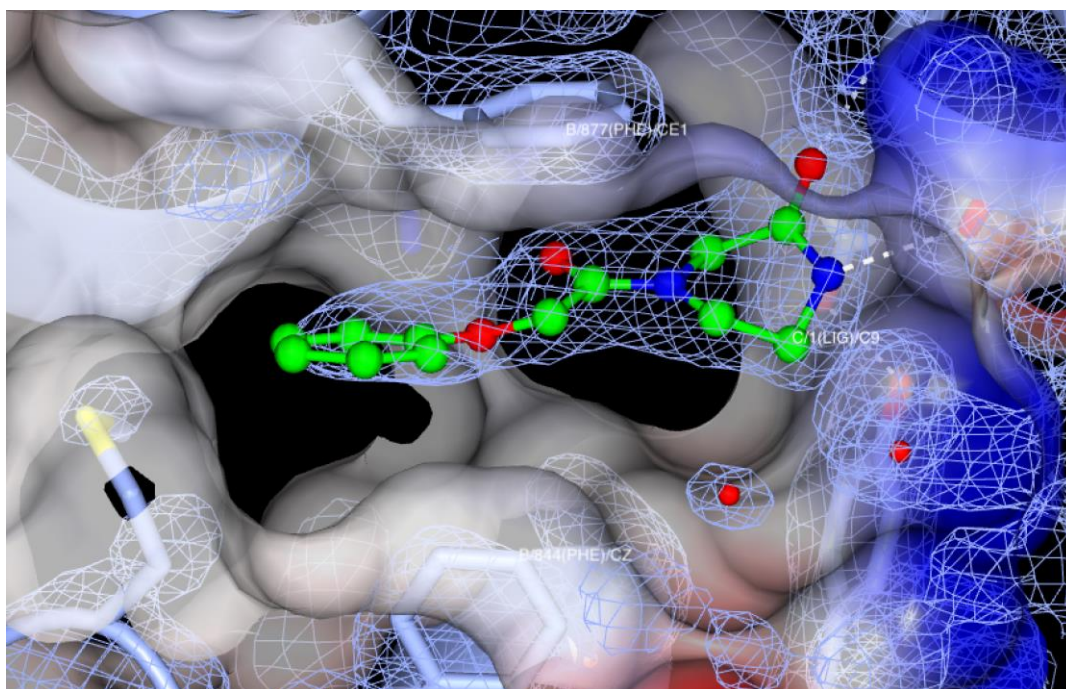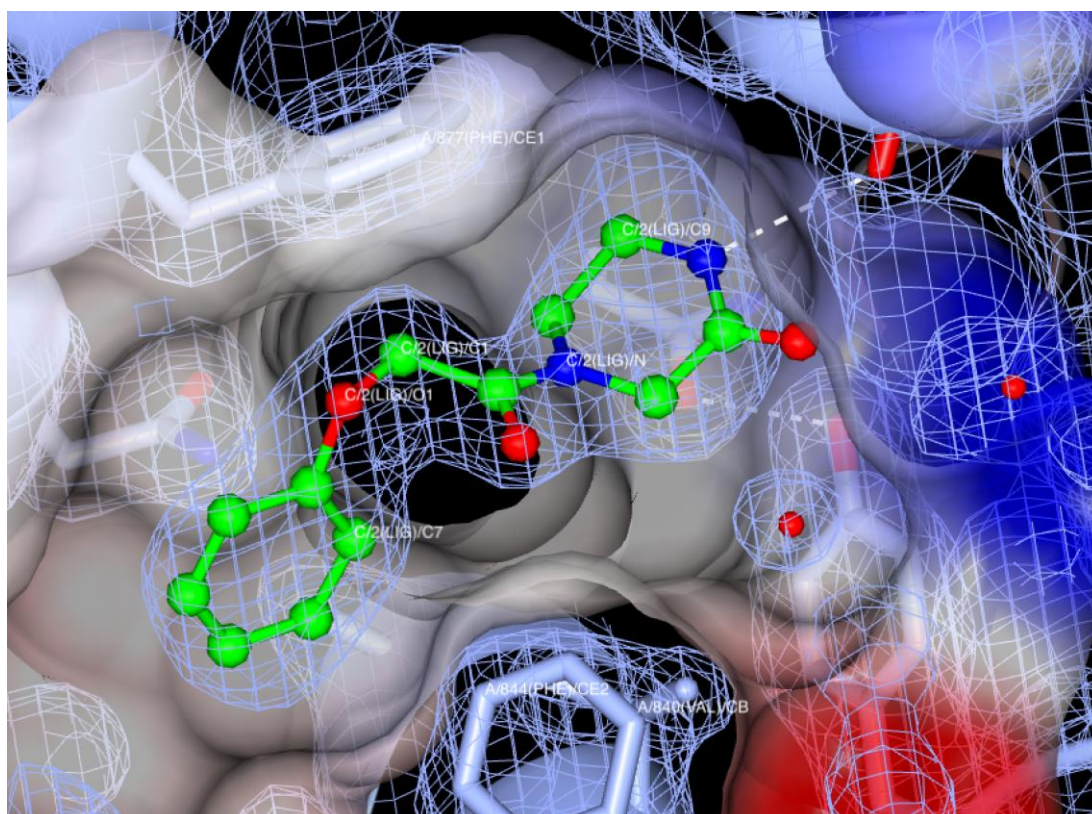

Figure 5.4: Surface representation of the ligand binding site with the ball and stick representation of 0074 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ, TbrPDEB1, second binding site, B chain

The second binding site was at chain B that could also be considered as novel, Figure 5.4.
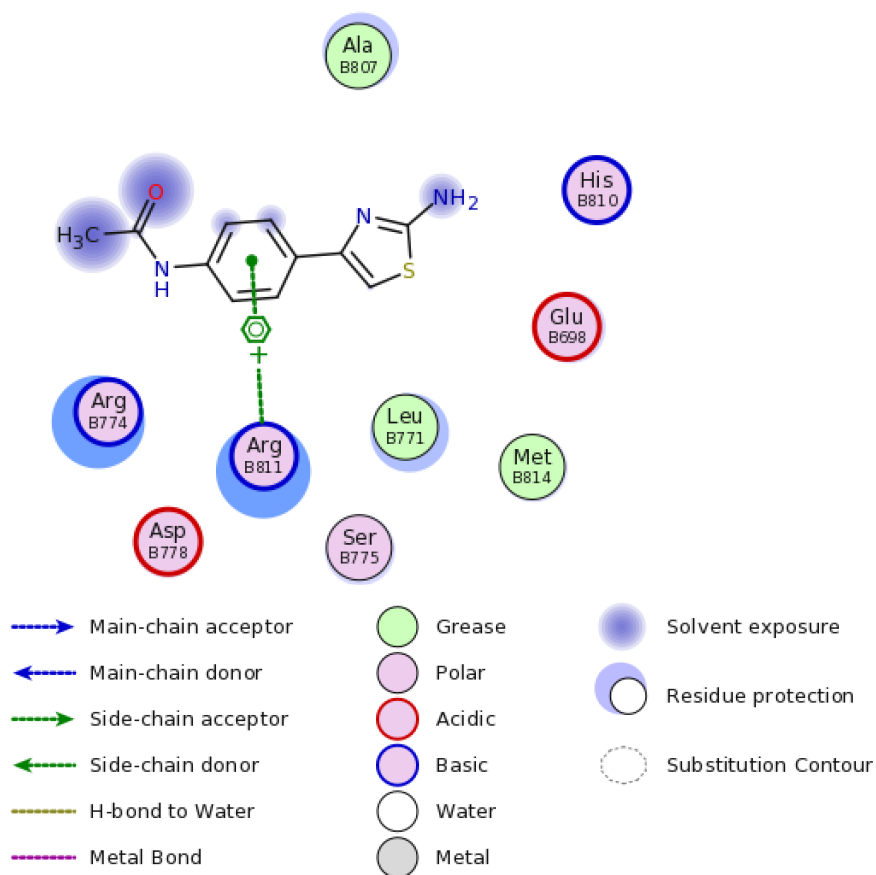
Figure 5.5: FLEV plot produced by COOT of the binding site, chain B, second binding site

Figure 5.5 represents the second binding site environment where Arg811 forms an interaction with the benzene ring of the 0074 ligand. In this binding mode, the ligand is surrounded by amino acids: Ala807, Arg774, Arg811, Asp778, Ser775, Leu771, Met814, Glu698 and His810. The benzene ring forms hydrophobic interaction with the Arg811 side chain and the positively charged guanidine group of the Arg811 positioned over the carbonyl moiety of 0074. On the other side of the ligand there is a hydrogen bond between a water molecule and the NH_2 group of the 4-methylthiazol-2-amine ring structure of 0074 fragment, Figure 5.4. The third binding site appeared in chain A, is also in the hydrophobic clamp Figure 5.6.
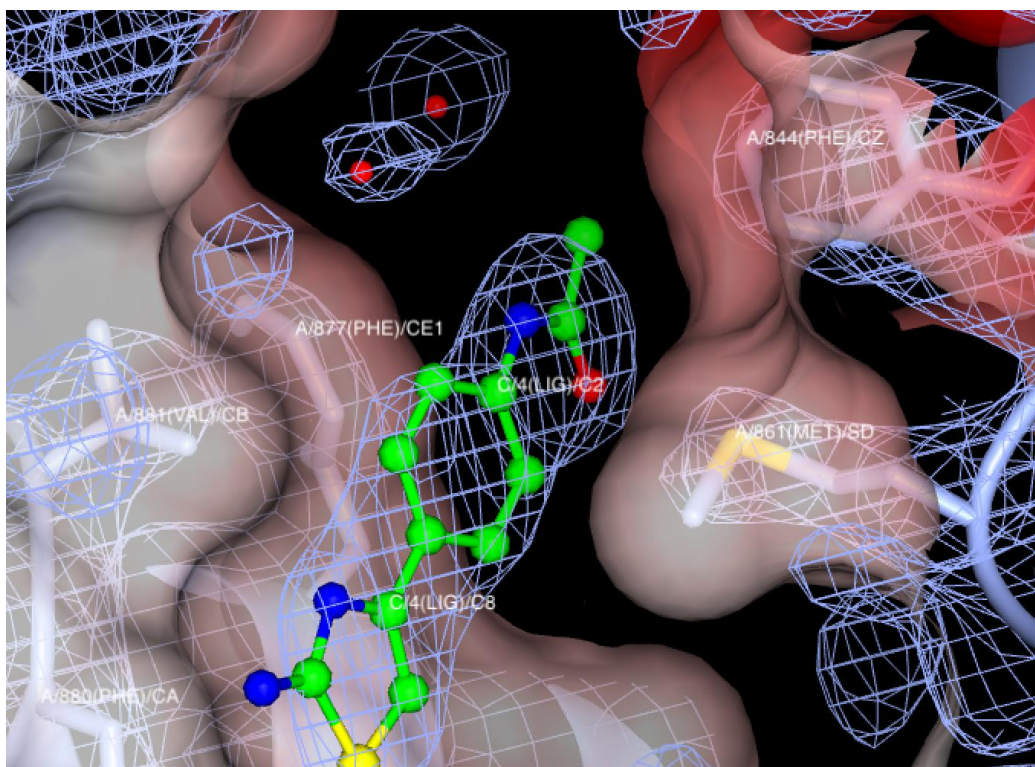
Figure 5.6 Surface representation of the ligand binding site with the ball and stick representation of 0074 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ, T. brucei PDEB1, chain A hydrophobic clamp
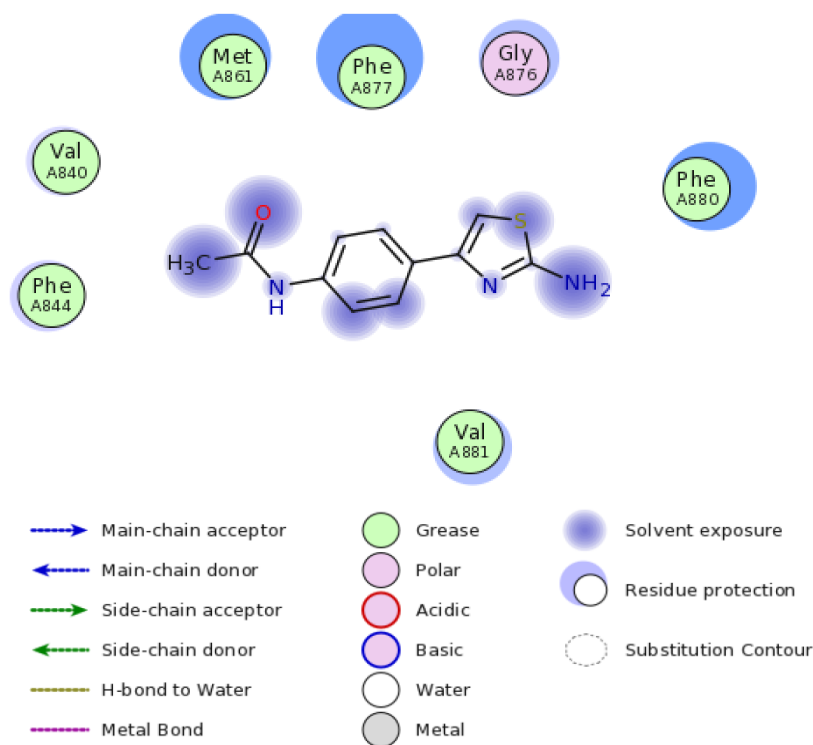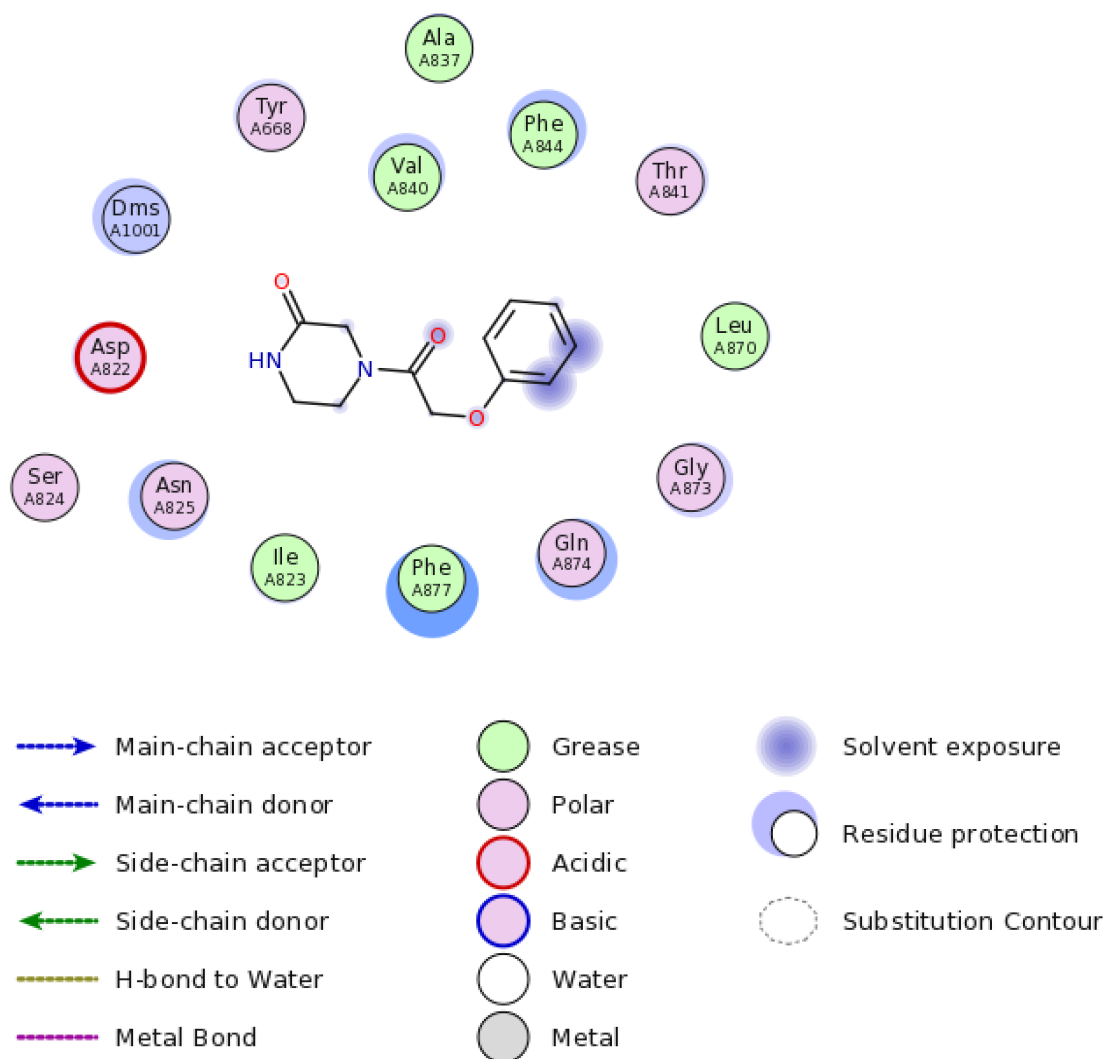

Figure 5.7: FLEV plot produced by COOT of the binding site of chain A, 3rd binding site

Figure 5.6 represents the 0074 ligand binding site at chain A, and Figure 5.7 shows the amino acids that surround the ligand.

The fragment was bound to the characteristic binding site likewise in chain B. The benzene ring is located near Phe877 ring forming strong hydrophobic interactions. It can also be seen in the Figure 5.6, there were also 2 well-ordered water molecules.

Out of 3 binding events, 2 were found occupying hydrophobic clamp that was formed by Val840 and Phe877 and the distal aromatic residues such as Phe844 and Phe880.

TBrB1 – Ligand 0209 Crystal structure

The next fragment data set of interest, from a novel binding mode point of view was 0209, Figure 5.8, where 2 binding sites were observed, in chain A and chain B respectively, Figure 5.10 and 5.11.
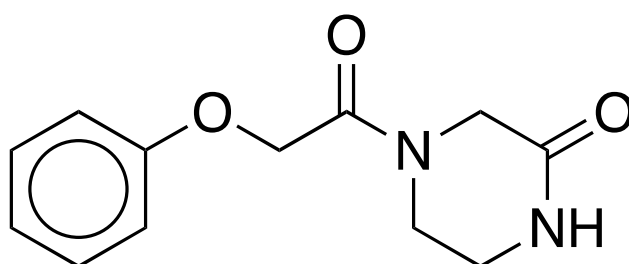
Figure 5.8: Fragment 0209, 4-(2-phenoxyacetyl) piperazin-2-one
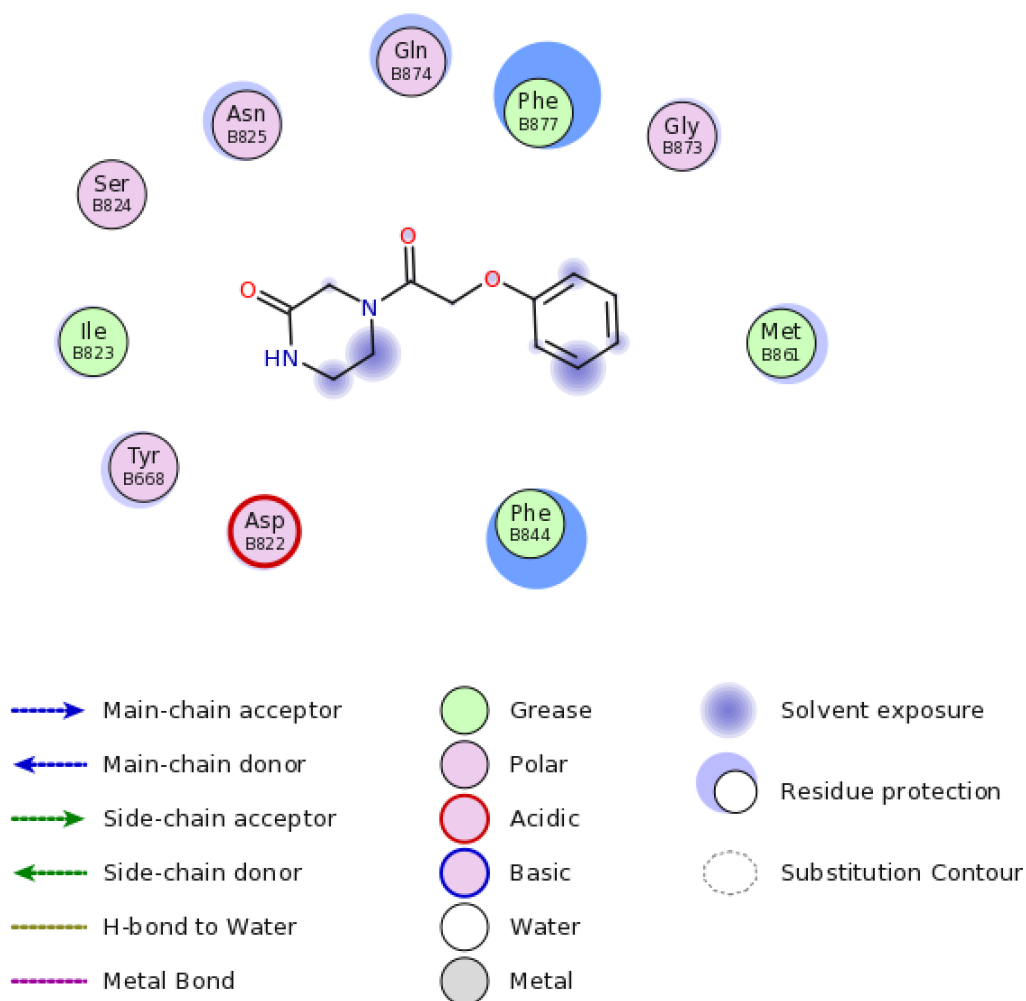
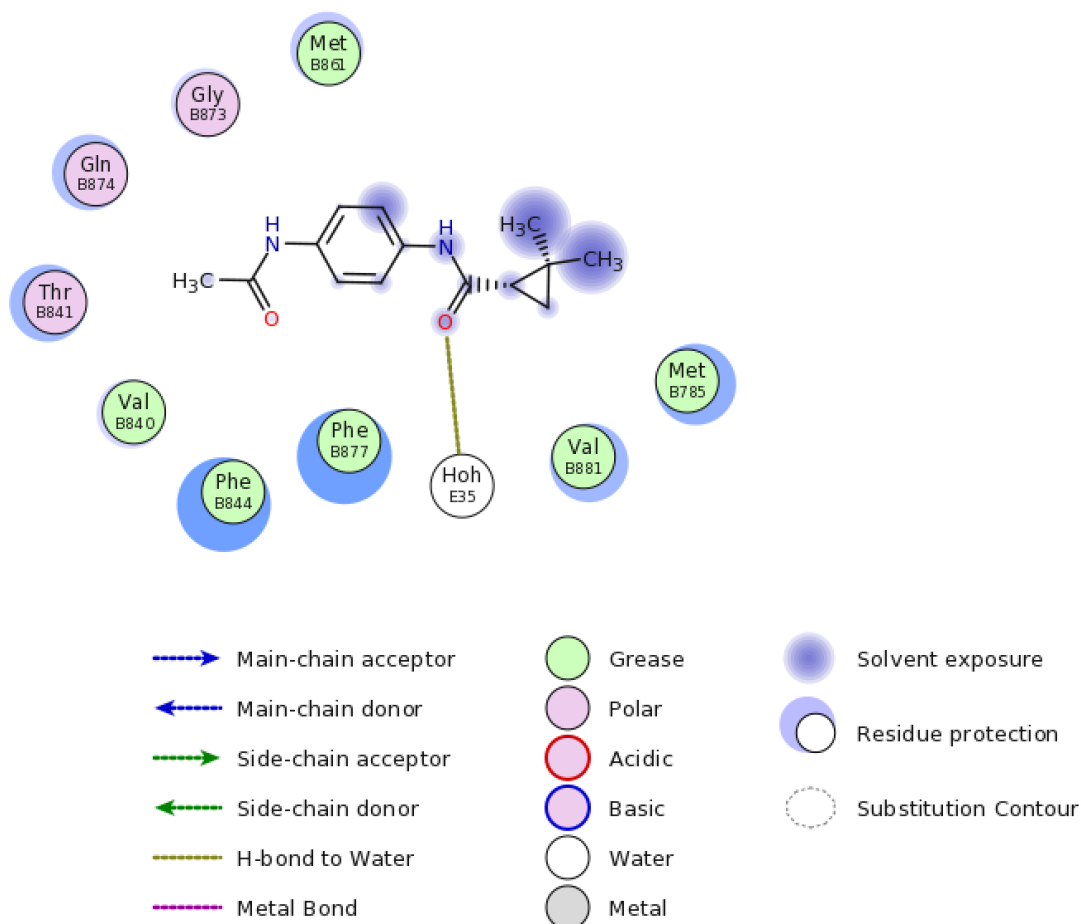***Figure 5.9:*** FLEV plot produced by COOT of the binding site of chain B, first binding site

Figure 5.9 represents the binding environment of ligand 0209 in molecule B using Coot's FLEV tool to highlight ligand protein interactions. The fragment is in the hydrophobic clamp which is a standard binding site. The benzene ring of the ligand was directed towards the P – pocket allowing the potential to utilise this vector for the design of new ligands to synthesise that may directly interact with the P – pocket residues. There were also additional interactions between the protein and the ligand that forms hydrogen interaction. The second interaction was shown to be between the hydrogen atom of benzene ring and oxygen atom of Asp822 forming a CHO Hydrogen bond *(101)*.

Figure 5.10: Surface representation of the ligand binding site with the ball and stick representation of 0209 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ , T. brucei PDEB1, Chain B hydrophobic clamp

Figure 5.11: Surface representation of the ligand binding site with the ball and stick representation of 0209 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ , T. brucei PDEB1, Chain A hydrophobic clamp

There is also a water molecule next to the Phe844 residue which also

improves stability and binding interactions between the ligand and the protein.

Figure 5.11 represents the binding mode in chain A where the fragment was located in the hydrophobic clamp and was also orientated towards the residues Met861 and Met868.

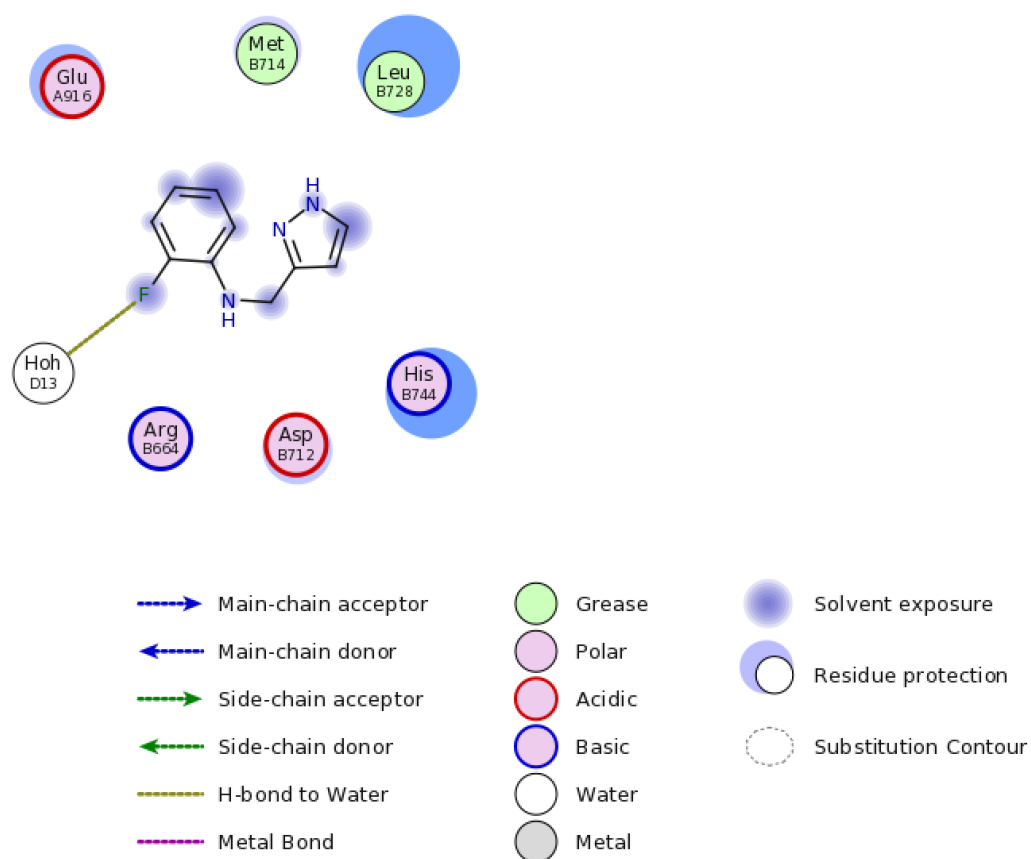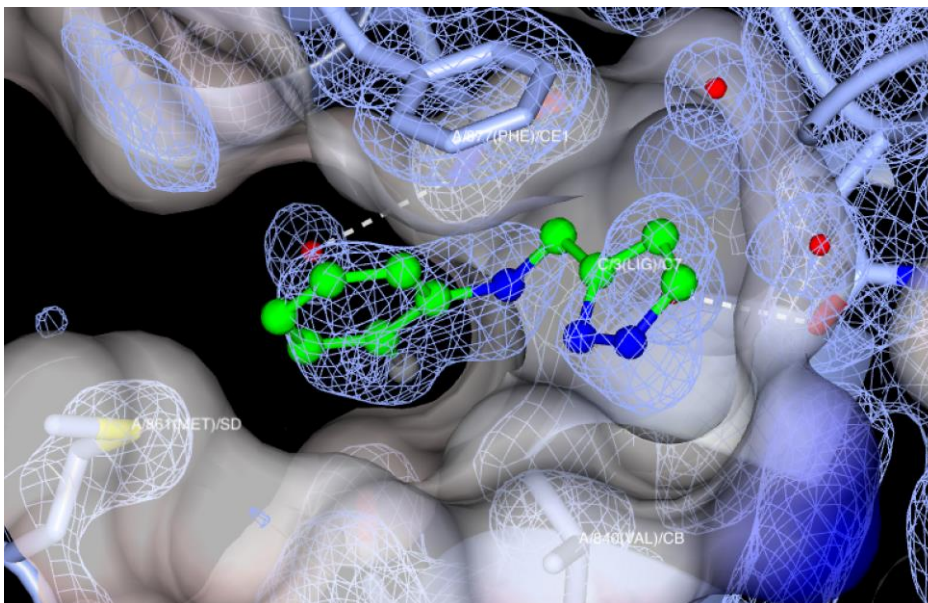*Figure 5.12:* FLEV plot produced by COOT of the binding site of chain A, second binding site

Figure 5.12 represents the ligand 0209 binding environment in chain A where ligand was found at its normal binding site in the hydrophobic clamp.

TBrB1 – Ligand 0218 Crystal structure

The next fragment that was analysed was 0218, Figure 5.13, which was located in the hydrophobic pocket and close to the residues such as Met861 and Met868 that composed P – pocket.
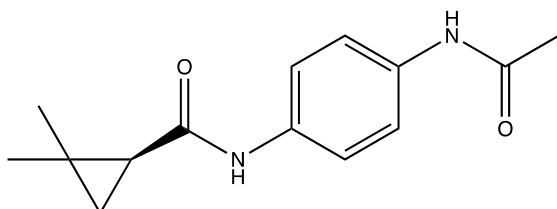
Figure 5.13: Fragment 0218, N1 –[4-(acetylamino) phenyl]-2,2-dimethylcyclopropane-1-carboxamide

Figure 5.14: FLEV plot produced by COOT of the binding site of chain B, first binding site
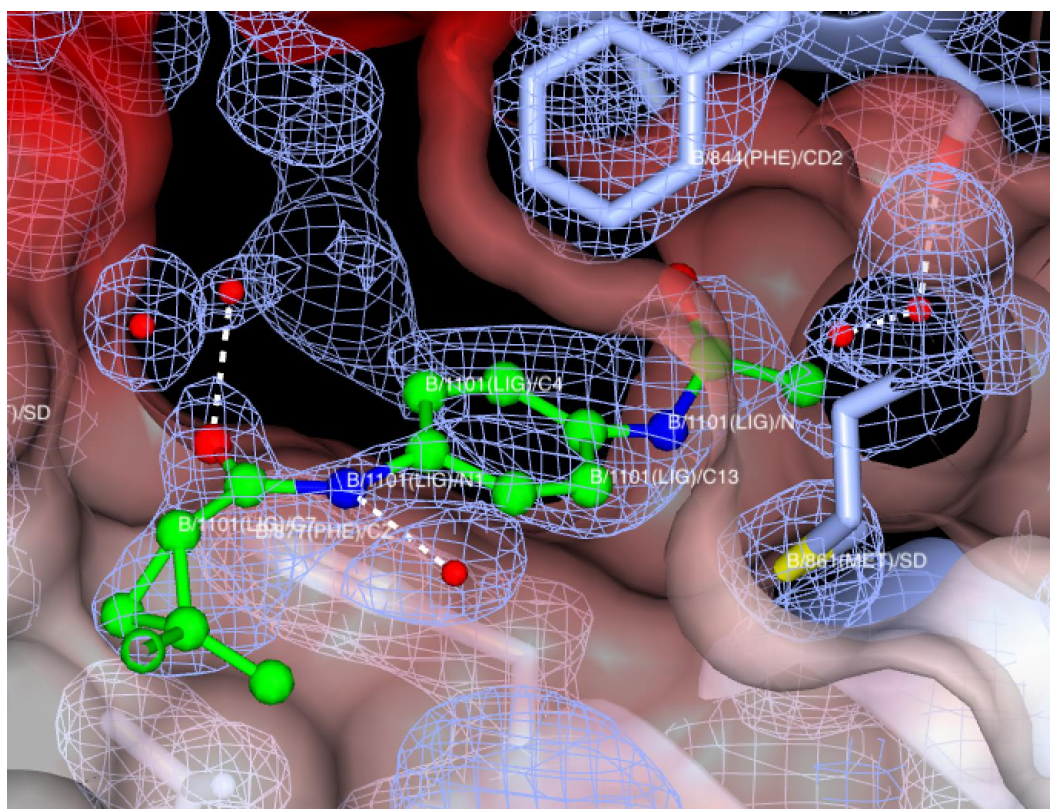
Figure 5.15: Surface representation of the ligand binding site with the ball and stick representation of 0218 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ, T. brucei PDEB1, Chain B hydrophobic clamp

As it can be seen in Figure 5.15, the fragment was located in the hydrophobic clamp of residues Val840, Phe844 and Phe877. Figure 5.14 also indicates additional surrounding amino acids: Gly873, Gly874, Thr841, Val881 and Met785. In the binding site, there were 5 well – ordered water molecules where 2 out of the 5, formed direct hydrogen bonds with the fragment and 2 others formed hydrogen bonds to the side chain of the Thr841. These interactions are considered to contribute significantly to the binding affinity. Unlike other fragments 0218 has only 1 binding site, in chain B which is the most preferable chain for inhibitors binding in this crystal system due to slight hydrophobic shielding by packing residues.

TBrB1 – Ligand 0268 Crystal structure

The next fragment that was analysed was 0268, Figure 5.16, where chain B showed a novel binding site, whereas chain A was bound to the fragment at the usual binding site.
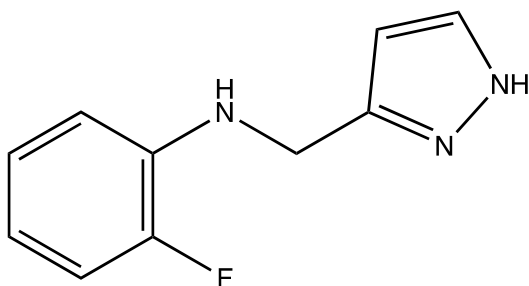
Figure 5.16: Fragment 0268, 2 – fluoro-N-[(1H-pyrazol-3-yl) methyl] aniline

Figure 5.17: FLEV plot produced by COOT of the binding site of chain A, first binding site

Figure 5.18: Surface representation of the ligand binding site with the ball and stick representation of 0268 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ , T. brucei PDEB1, Chain A hydrophobic clamp
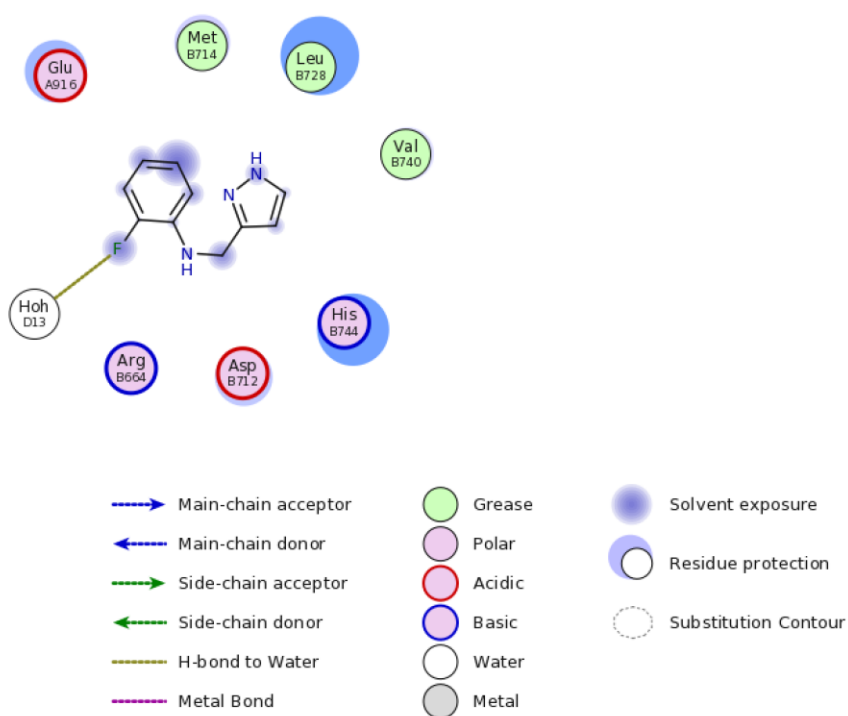
Figure 5.19: FLEV plot produced by COOT of the binding site of chain B, second binding site

Figure 5.18 represents the binding mode between TbrPDEB1 catalytic domain and 0268, where the fragment occupied the hydrophobic clamp in

chain A. The fragment was in close contact with such protein residues as: Phe877, Met861, Thr841, Gly874, Val840, Ala937, Tyr668, Asp825, Iso823 and Asp822, Figure 5.17.
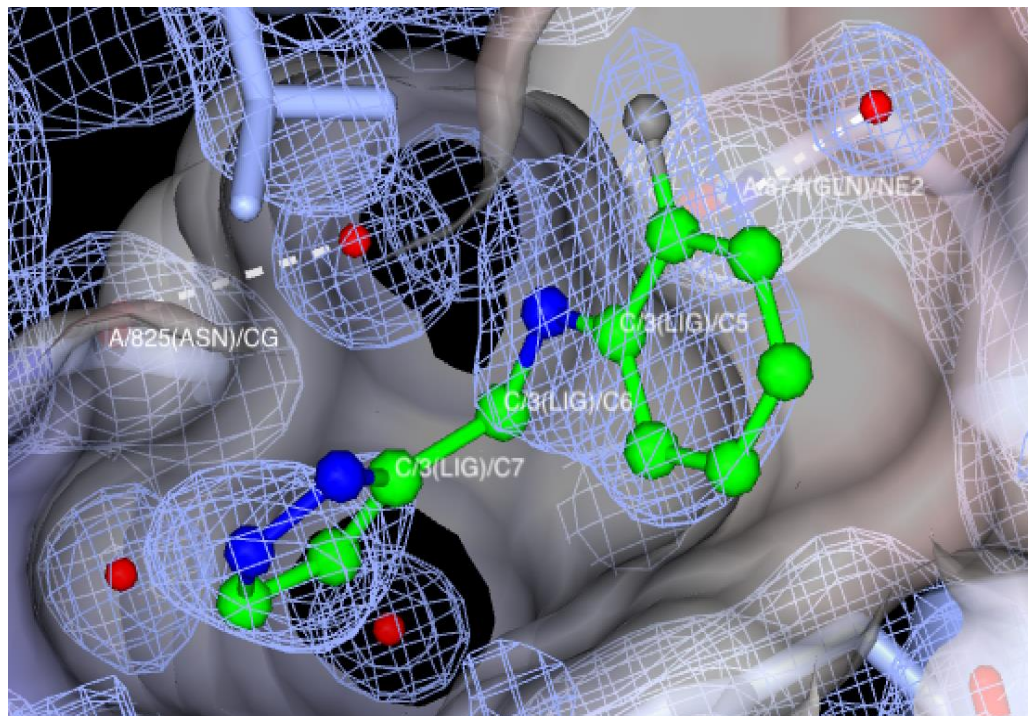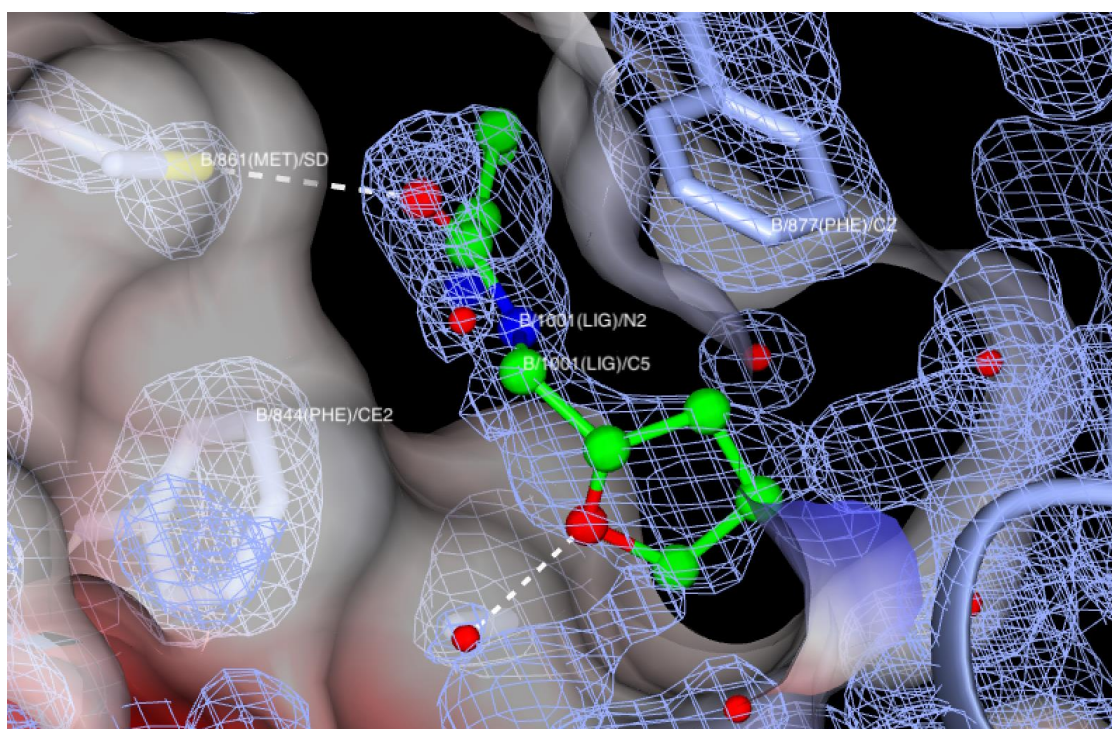

Figure 5.20: Surface representation of the ligand binding site with the ball and stick representation of 0268 fragment with $|2Fo\text{-}Fc|\alpha_{calc}$ map at a counter level of 1.2 σ , T. brucei PDEB1, Chain B novel binding site

The second binding event was in chain B where novel site was observed, Figure 5.20.

The ligand was surrounded by amino acids, such as: Leu728, Met714, Arg664, Glu916, Asp712, His744, Leu741, Val740 and Tyr721, Figure 5.19. There were also other hydrogen bonds between water molecules and the side chains of protein residues: Glu916, Leu915 and Arg918. Such hydrogen interactions improve binding and protein stability. This binding site has not previously been observed and as such was considered a novel site that required further investigation.

TBrB1 – Ligand 0269 Crystal structure

The next fragment that was analysed was 0269, Figure 4.21, where ligand was bound in 2 sites.
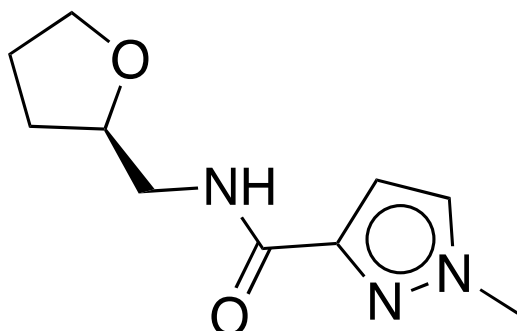
Figure 5.21: Fragment TbrB1 – 0269, (R)-1-methyl-N-((tetrahydrofuran-2-yl) methyl)-4H-1-parazole-3-carboxamide
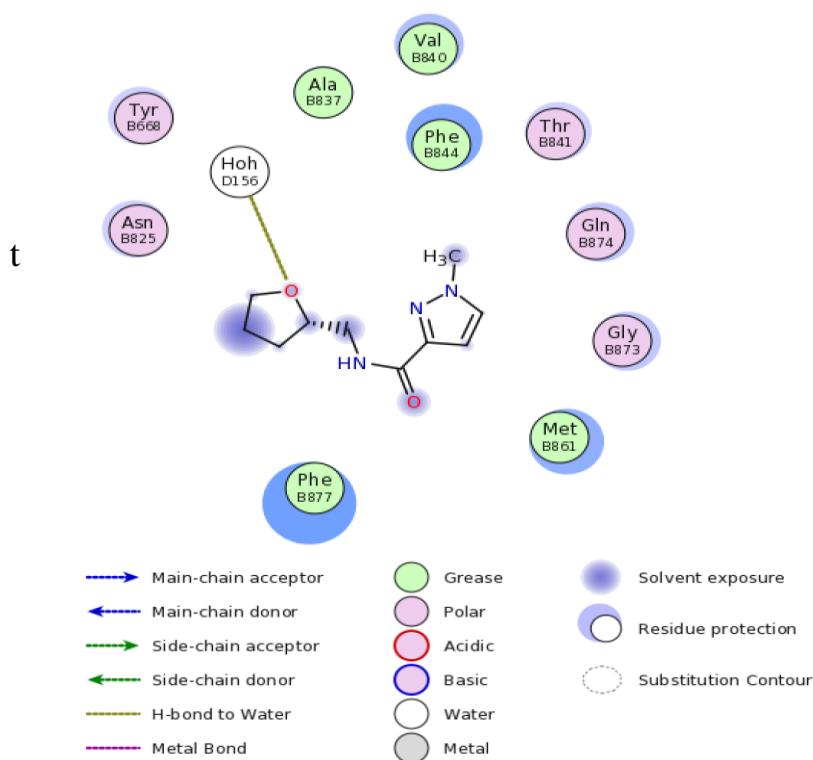
Figure 5.22: FLEV plot produced by COOT of the binding site, binding site of chain B, 1ˢᵗ binding site

The binding site was formed by amino acids: Tyr668, Asp825, Phe877, Met861, Gly873, Glu874, Thr841, Phe844, Val840 and Ala837, Figure

5.22. The fragment 0269 formed hydrogen bond between its carbonyl oxygen atom and the side chain of Met861 residue that forming a P – pocket cavity. It is also forming hydrogen bond between one of the nitrogen atom of the imidazole ring and water molecule that was present in the binding site, Figure 5.23.

There was also a formation of a hydrogen bond between a water molecule and the protein residue Thr841. Chain A, Figure 5.24 shows that ligand was bound in the hydrophobic clamp however some additional density was present that suggested that fragment could also have an alternative pose.

Figure 5.23: Surface representation of the ligand binding site with the ball and stick representation of 0269 fragment with |2Fo-Fc| α_{calc} map at a counter level of 1.2 σ , T. brucei PDEB1, Chain B hydrophobic clamp
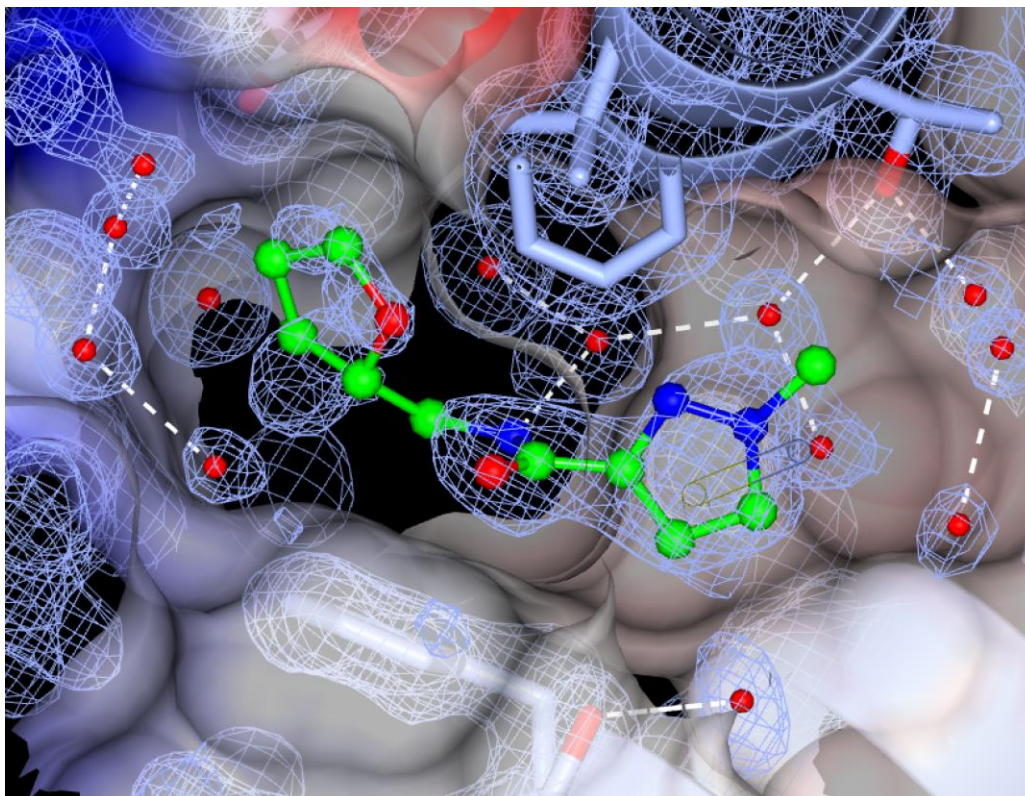
Figure 5.24: Surface representation of the ligand binding site with the ball and stick representation of 0269 fragment with $|2Fo\text{-}Fc|\alpha_{calc}$ map at a counter level of 1.2 σ, T. brucei PDEB1, Chain A hydrophobic clamp

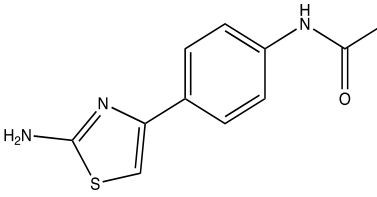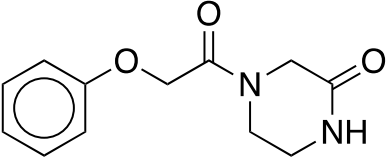Figure 5.24 represents the interactions formed between the fragment and the protein. The fragment 0269 formed a hydrogen bond interaction with one of the water molecule. An observed glycerol from the cryoprotectant and also formed a hydrogen bond with the water molecule as well, where it was present as a cry - protectant. There are also hydrogen bonds formed between the water molecules in the binding site.

TBrB1 – Ligand 0248 Crystal structure
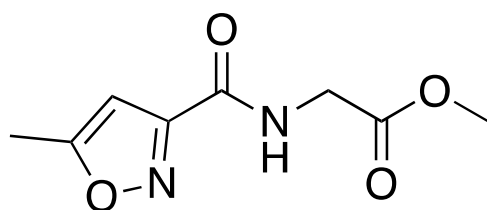
Figure 5.25: Fragment 0248, [(5-Methyl-isoxazole-3-carbonyl)-amino]-acetic acid methyl ester methyl 2-{[(5-methyllisoxazol-3-yl) carbonyl] amino} acetate
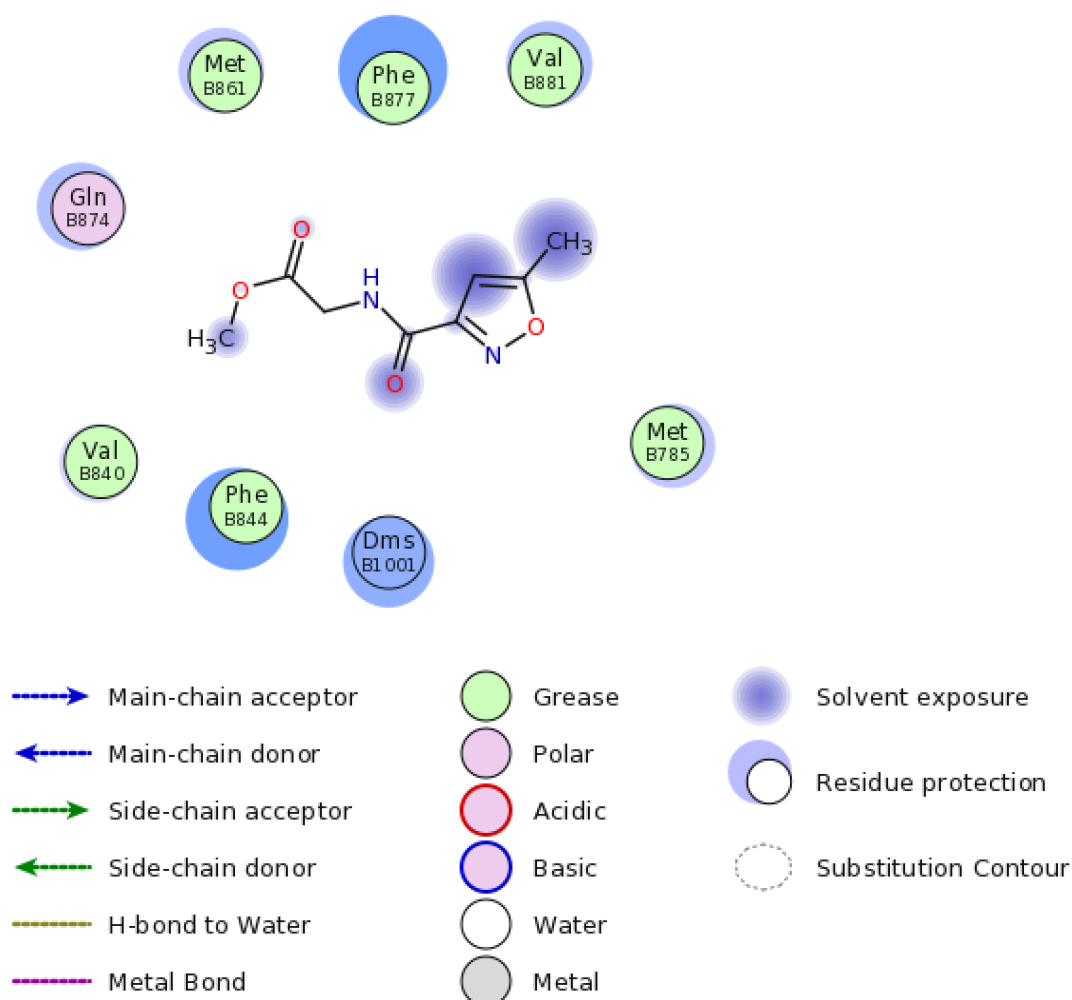
Figure 5.26: FLEV plot produced by COOT of the binding site in chain B, 1st binding site

The last fragment that was analysed was 0248, Figure 5.25, where only one binding mode was observed in chain B, in the hydrophobic clamp.

Fragment 0248 was surrounded by the amino acids: Phe877, Met861, Gly874, Val840, Phe844, Met785 and Val881, Figure 5.26.
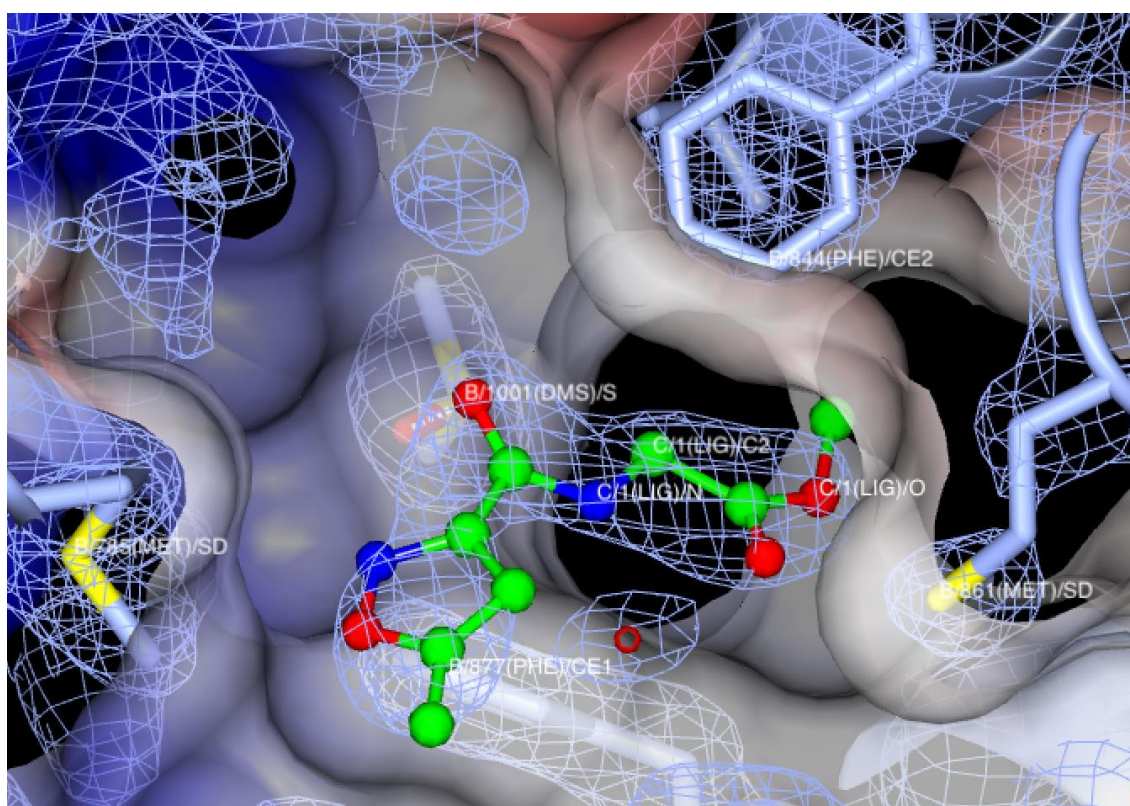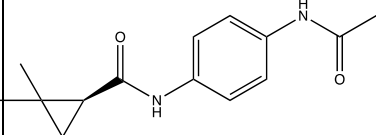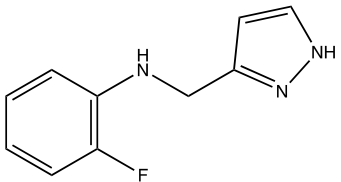

Figure 5.27: Surface representation of the ligand binding site with the ball and stick representation of 0248 fragment with |2Fo-Fc|α_{calc} map at a counter level of 1.2 σ , T. brucei PDEB1, Chain B hydrophobic clamp

From the Figure 5.27 it can be seen that fragment was bound in the commonly observed binding site, where the hydrophobic clamp is thought to provide the majority of the protein ligand interaction energy. The fragment also formed an interaction with a DMSO molecule that was used for fragments solubilisation.

After the analysis of the most interesting compounds, it was important to see which ones offered growth vectors towards the P – pocket to increase their potency and potential selectivity towards parasitic PDEs over human PDEs.

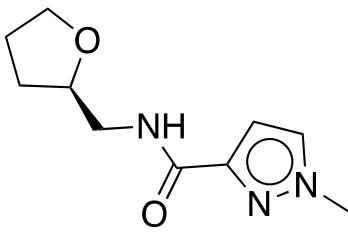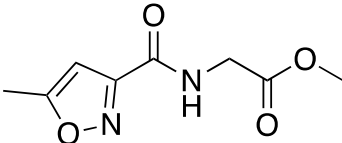| Fragment	Structure	Binding mode
TbrB1-0074		3 binding sites: chain A – hydrophobic clamp chain B – hydrophobic clamp and novel binding site (Arg811)
TbrB1-0209		2 binding sites: chain A – hydrophobic clamp chain B – hydrophobic clamp
TbrB1-0218		1 binding site: chain B – hydrophobic clamp
TbrB1-0268		2 binding sites: chain A – hydrophobic clamp chain B – novel binding site (Leu718, Arg664, Glu916)
TbrB1-0269		2 binding sites:

		chain A – hydrophobic clamp chain B – hydrophobic clamp
TbrB1-0248		1 binding site: chain B – hydrophobic clamp

Table 5.1: summarised table of TbrB1 fragments

5.4. Computational Analysis of Fragments

In a collaboration with Lorena Zara from the University of VU a computational analysis of 6 chosen ligands with vectors towards the P – pocket was performed.

The fragment 0218 was a cyclopropane that was directed towards the P – pocket and there was an aromatic interaction with residues Phe877 and Phe844, which formed a hydrophobic clamp, Figure 5.28. The idea behind the current fragment is to see whether it could be grown into the P – pocket where certain analysis was performed.
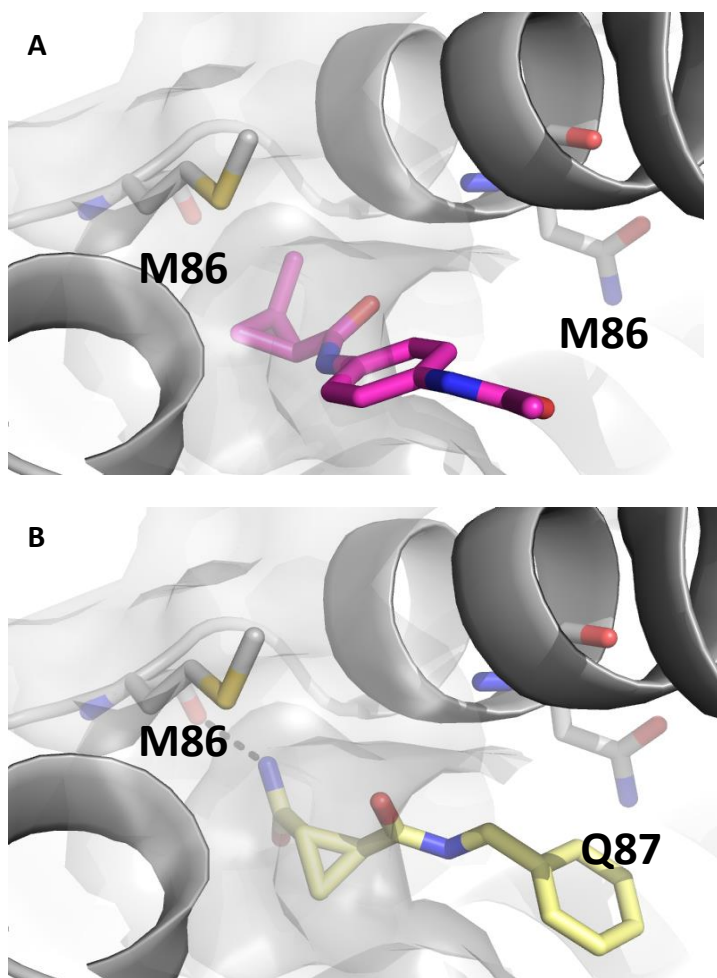
Figure 5.28: Fragment 0218. **A)** Current Ligand protein binding mode **B)** Newly synthesised derivative of 0218 fragment (Figure 5.29)
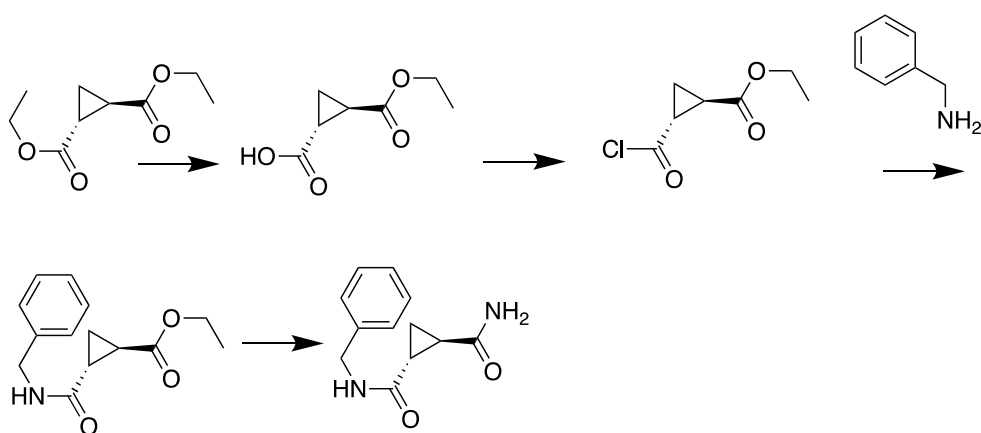

Figure 5.29: Synthesis of fragment 0218 towards P – pocket for interaction with Met868

Met868 residue was one of the residues that compose the P – pocket.
It was suggested to use synthesis route above, Figure 5.29, to grow
derivatives that would be vectoring towards the P – pocket. Computational
enumeration of such ligands followed by a search of the Zinc database
(https://enamine.net) highlighted in the table 5.2 below.

TbrBr1 - 0218
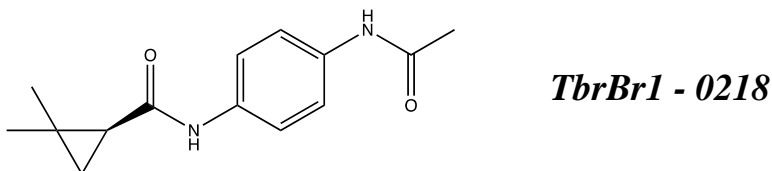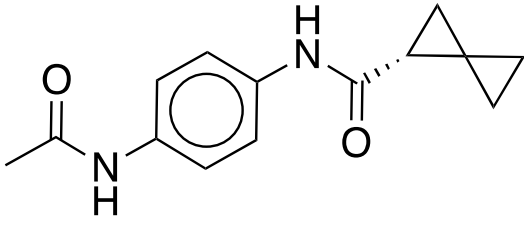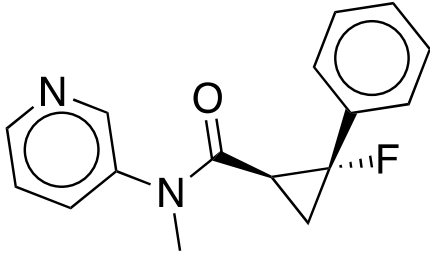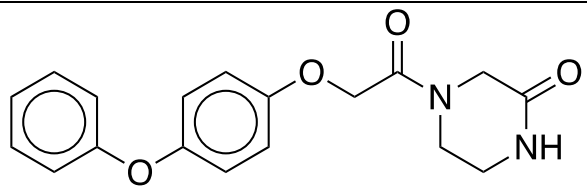
Fragment	Zinc Name
	ZINC643298333
	ZINC334722503
	ZINC1437151

Table 5.2: 0218 fragments derivatives

The fragment 0209 has ligand bound in 2 chains: chain A and B, where in
both chains fragment was interacting with the hydrophobic clamp.
Docking analysis showed that fragment could be grown towards the P –
pocket to interact with its residues, Figure 5.30.
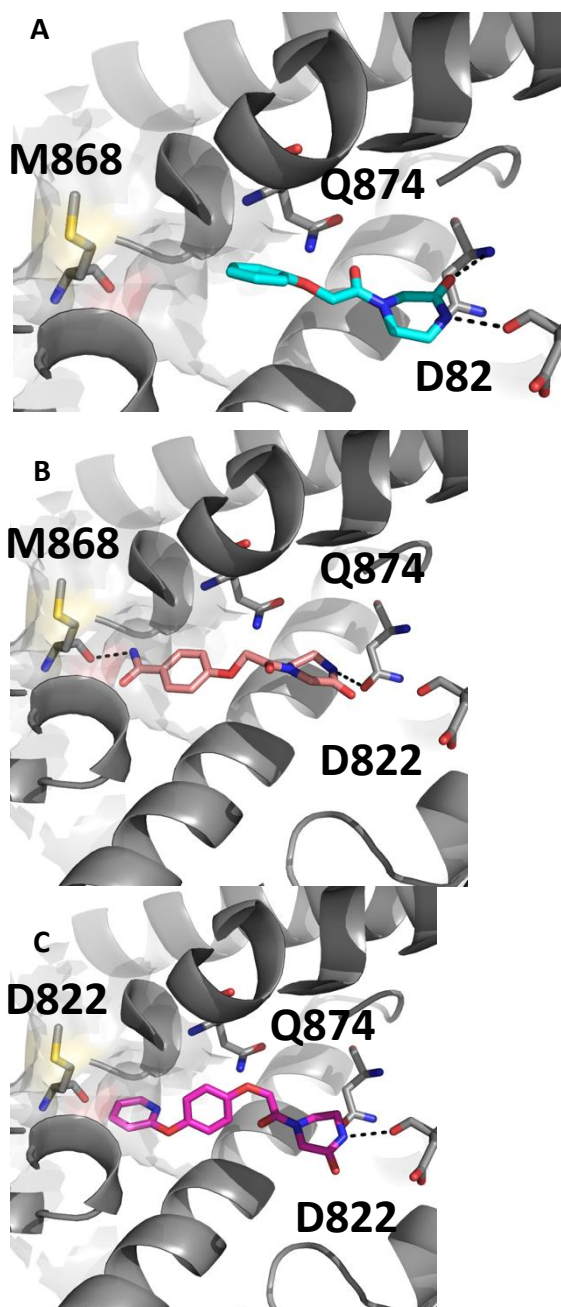
Figure 5.30: **A**) Ligand 0209, protein binding mode **B, C**) Docking pose

Fragment	*Zinc Name*
	ZINC12212809

Table 5.3: 0209 fragments derivatives

The next fragment that was subjected to computational analysis was 0074. The docking models of the Zinc molecules can be observed in the Figure 5.31 B and C.
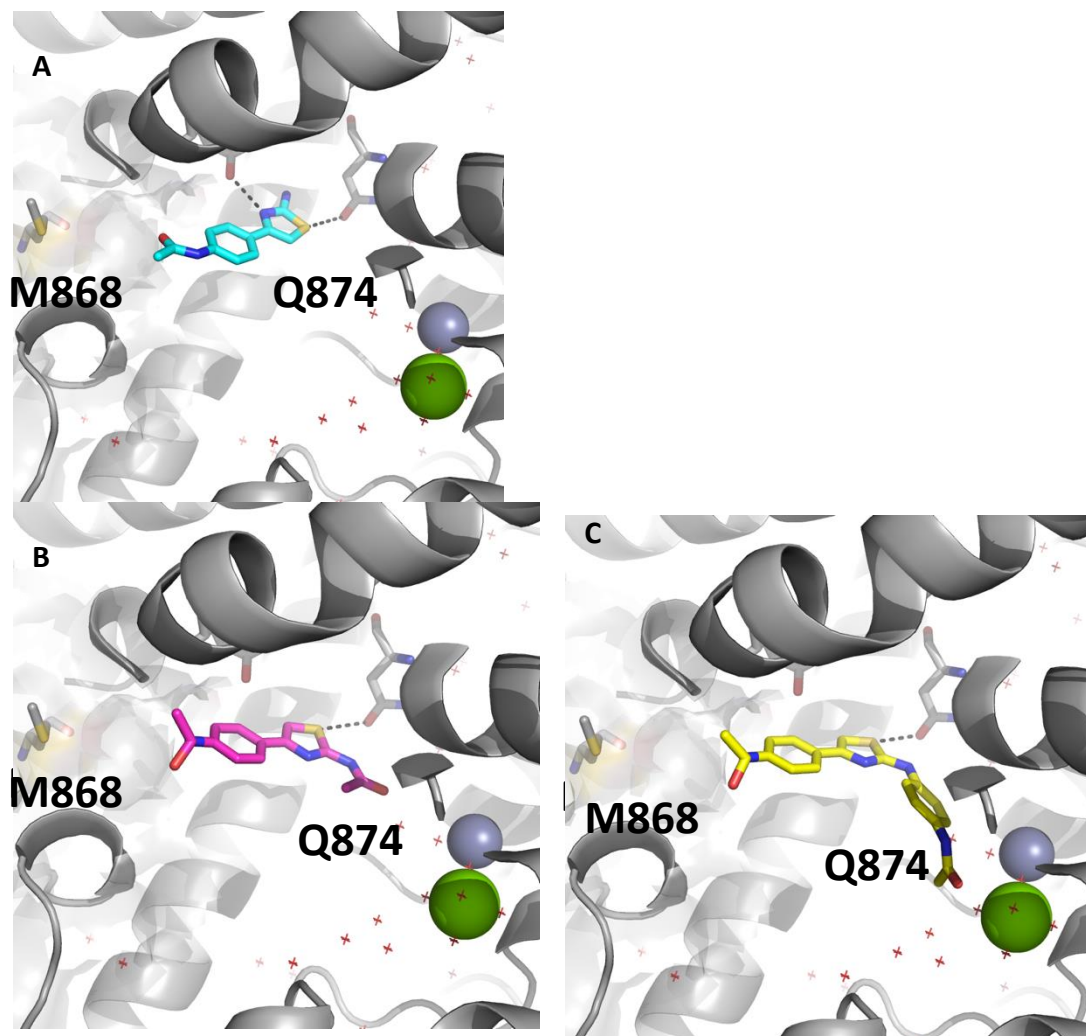
Figure 5.31: **A)** Ligand 0074 protein binding mode **B)** ZINC3156385 docking **C)** ZINC813384 docking

Fragment	Zinc Name
	ZINC3156385
	ZINC813384

Table 5.4: 0074 fragments derivatives

The second approach that was discussed in collaboration with the University of VU was the possibility to merge observed fragments structures from XChem in order to identify new derivatives that will be selective towards the P – pocket. Thus, is exemplified in two cases with the merging of the fragments 0074 and 0209 with the fragment 0269

    I.   Fragments 0074 and 0269 gave a final compound:

0074   **0269**

Figure 5.32: Merged compound of fragments 0074 and 0269

II. Fragments 0209 and 0269 gave a final compound:

0209

0269

Figure 5.33: Merged compound of fragments 0209 and 0269

Synthesis and studies of current fragments would be the next step in current study.
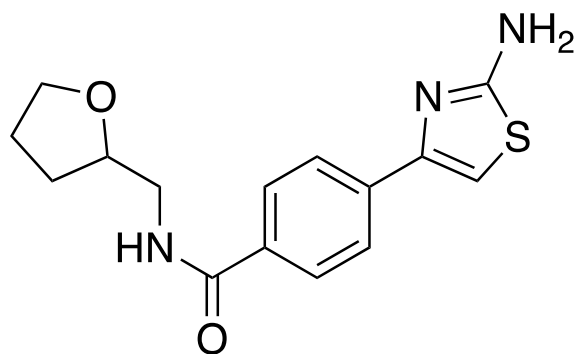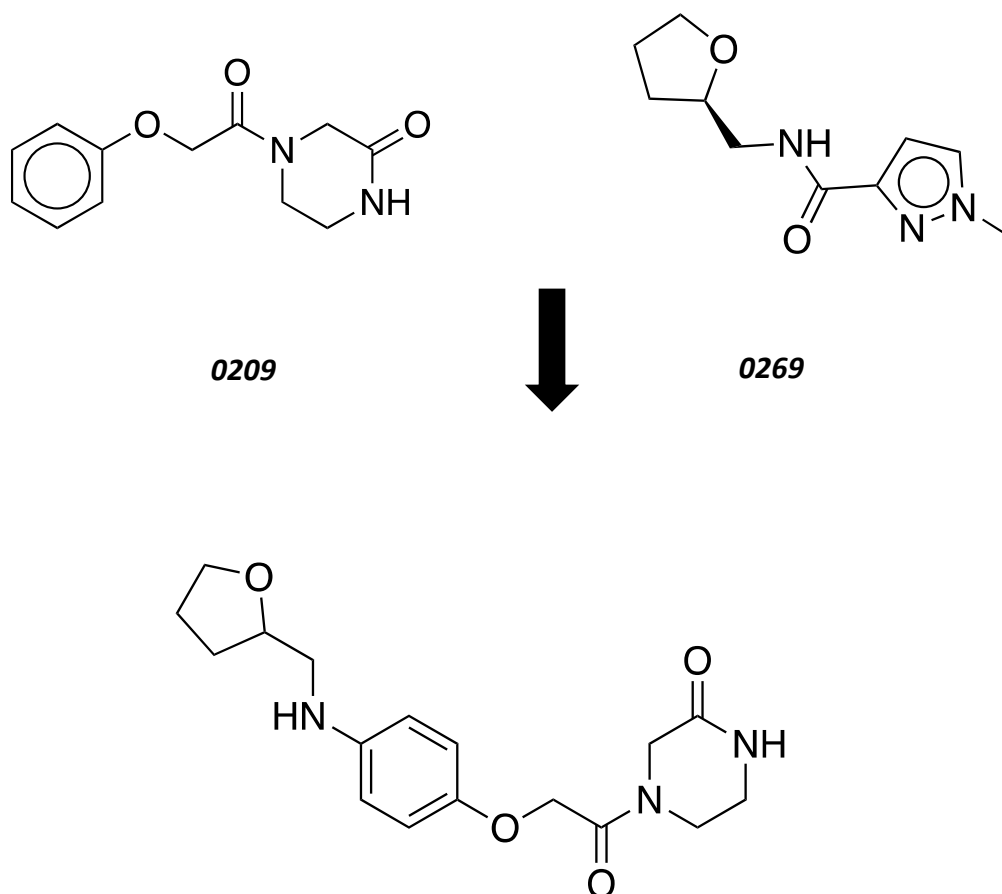
5.5 SPR data of fragments selected from the XChem run

The six fragments that were considered suitable for generation of derivatives directed towards P – pocket that were used in SPR analysis. There were 5 compounds that were run on SPR that were used during XChem run, where it was important to see whether these compounds could also show any binding affinity towards parasitic PDEB1 and if there is any difference with human PDE4 binding affinity that was run in parallel.

5.5.2 SPR data

As it was mentioned earlier 6 compounds were analaysed by SPR on a Biacore T200: 0218, 0412, 0269, 0209 and 0074, tested at 1 – 30 uM concentration ranges. There were 3 immobilised proteins that current ligands were tested against: TbrPDEB1, TbrPDEB2 and hPDE4.

0218	*Proteins*		
	TbrPDEB1	*TbrPDEB2*	*hPDE4*
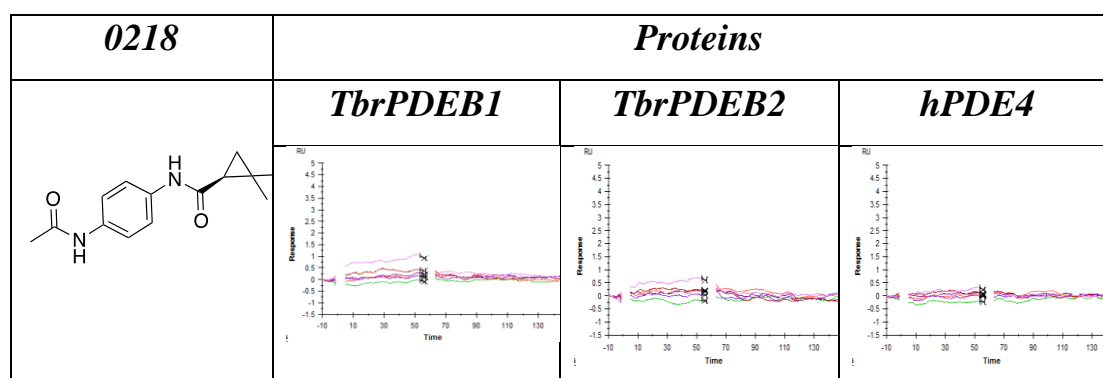
Table 5.5: SPR data spectra analysis of 3 proteins run with XChem hit 0218 ligand

Table 5.5 represents SPR sensorgram of response of the 3 different proteins with 0218. It can be observed that there was no detectable binding in all 3 cases. There was no association state where the complex was forming and hence there was no equilibrium state. Therefore, it can be concluded that current fragment was too weak and it couldn't be detected under these conditions by SPR.

0412	Proteins		
	TbrPDEB1	TbrPDEB2	hPDE4

Table 5.6: SPR data spectra analysis of 3 proteins run with XChem hit 0412 ligand

Table 5.6 represents SPR binding analysis of 0412 against the three different proteins. Once again according to the results in Table 5.6 it can be seen that there was measurable binding in all 3 SPR runs.
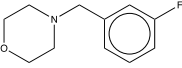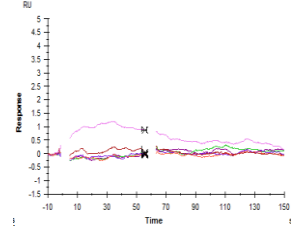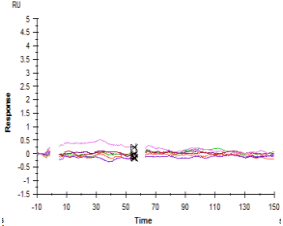
0269	Proteins		
	TbrPDEB1	TbrPDEB2	hPDE4

Table 5.7: SPR data spectra analysis of 3 proteins run with XChem hit 0269 ligand

According to the Table 5.7, it can be observed that there was a slight difference to Tables 5.5 and 5.6 and there was a slight increase in the signal of SPR experiment with parasitic proteins and its highest concentration (30 uM) but not an interpretable sensorgram to calculate any binding affinity, while there were no spectra change in hPDE4 SPR spectrogram.
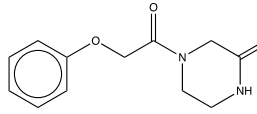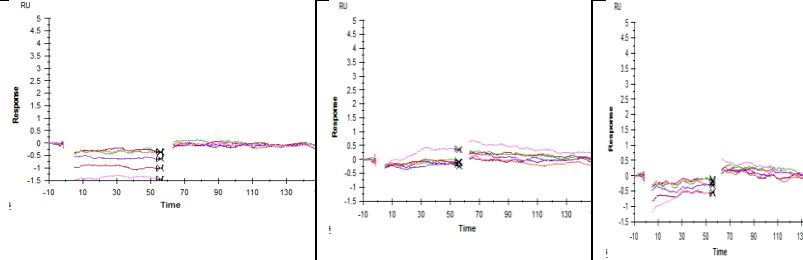
0209	Proteins		
	TbrPDEB1	TbrPDEB2	hPDE4

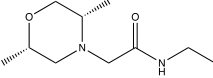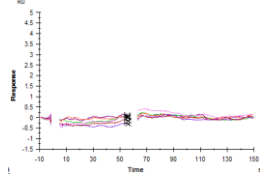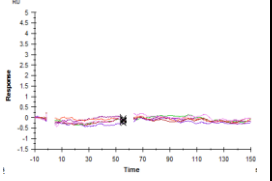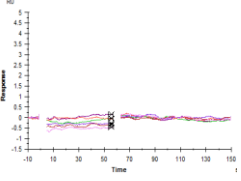*Table 5.8:* SPR data spectra analysis of 3 proteins run with XChem hit 0209 ligand

Table 5.8 represents SPR binding sensorgram where 0209 ligand was used against 3 proteins as in previous experiments. However, no binding was observed in all 3 cases.

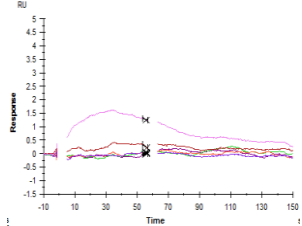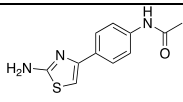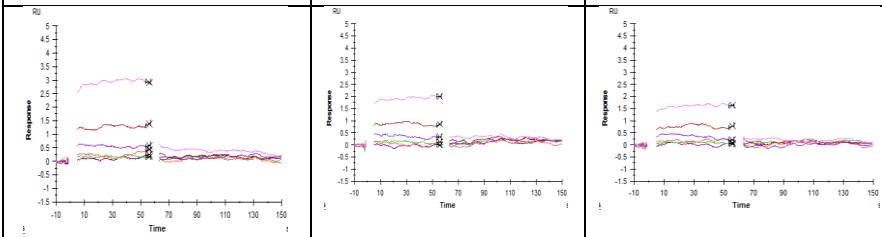| 0074	Proteins		
	TbrPDEB1	TbrPDEB2	hPDE4

Table 5.9: SPR data spectra analysis of 3 proteins run with XChem hit 0074 ligand

According to the Table 5.9 there was a signal at high concentration of ligand, and it was clear that parasitic PDEs had stronger signal then hPDE4. However, such results could be due to nonspecific binding and no formal binding affinity could be deduced from such data.

Among all 5 compounds only 0074 showed a binding signal at high concentrations but this may not be specific binding. However, the previous computational analysis, Section 5.4, showed that it may be possible that derivatives from 0074 can be further grown towards the P – pocket in

order to improve its affinity and potentially selectivity. To analyse this ZINC813384, Table 5.4, was run on SPR and the following results were obtained.

ZINC813384	Proteins		
	TbrPDEB1	TbrPDEB2	hPDE4
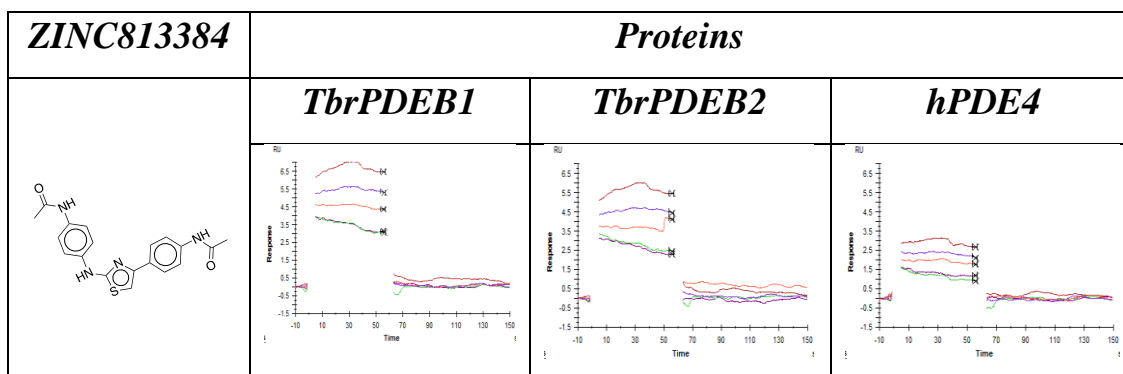
Figure 5.10: SPR data spectra analysis of 3 proteins run with XChem hit 0074 derivative ZINC813384 ligand.

According to the Table 5.10 a binding signal can be observed in all 3 instances, at different concentrations. Again, the data quality does not allow determination of a KD but there seems to be a qualitative indication of an increase binding affinity. Further experiments and enumeration of similar compounds grown in this direction would be required to confirm the hypothesis. However, protein was prone to precipitation hence it is required to re-run the SPR screen.

The main outcome of SPR experiment was: TbrB1 – 0074 showed the most promising result where parasitic protein had higher signal then human PDE4 enzyme, as well as the derivative of 0074 that showed higher signal with parasitic PDEB1, Table 5.11.

Fragment	SPR response against TbrB1
TbrB1 - 0218	x
TbrB1 - 0412	x
TbrB1 - 0269	x
TbrB1 - 0209	x

TbrB1 - 0074	√
TbrB1 – 0074 (derivative ZINC813384)	√

Table 5.11: Summary of SPR screen, 2 fragments showed stronger signal against TbrB1 enzyme.

5.6 Conclusion

The XChem platform allows the high-throughput screening of ligand-soaked crystals to access the binding mode of the high number of compounds. One of the biggest advantage of using such a screening method was that even 10% binding occupancy could be found, however as it can be seen from the SPR it may prove difficult to measure and rank such fragments on the basis of affinity using SPR.

It was very important to optimize the system and compare it with other biophysical methods that let the users to analyse ligand-protein binding mode. In current experiment, TbrPDEB1 protein was used as a target that was tested using DSI - Poised library that was provided by the Diamond Synchrotron. 466 compounds were used during the XChem run but only 293 data sets were successfully diffracted and detected ligand binding modes were manually analysed for 20 complexes. Each data set has multiple binding events. The primary aim was to find the novel interesting fragments, whether the ligand was bound in the normal binding site which was a hydrophobic clamp or ligand was more orientated towards the selective P – pocket. As the P – pocket is a distinct feature that is only found in parasitic PDEs, fragments that had moieties or substitution vectors directed towards P – pocket would be more desirable as it would increase the binding selectivity between human and parasitic PDEs. Out of 293 analysed diffraction data sets, 20 ligands – protein structures were analysed manually. The most interesting ones were chosen to be 6

fragments where more complete refinement was performed, and detailed analysis was performed. Out of 6 fragments, 5 were bound in both chains, whereas fragment 0248 was only found in chain B in the hydrophobic clamp. Such fragments as 0074 and 0268 showed novel binding sites at chain B, whereas fragment 0074 had 3 binding sites, 2 in chain B and 1 in chain A.

Such findings led to the next step where in collaboration with the VU University, computational analysis of 5 - the most interesting compounds was performed in order to see what fragments were the most promising and interesting from the chemical and biological point of view. Identifying the most promising compounds enabled the experiment to proceed further by performing SPR direct binding analysis where 5 fragments were used together with 1 derivative of the fragment 0074. The experiment was performed with 3 different proteins, such as: TbrPDEB1, TbrPDEB2 and human PDE4. Out of 5 runs only compound 0074 showed association and dissociation curves where the signal was observed, however there was also a high probability of unspecific binding.

The fact that previous experiments by SPR screening had found fragments with measurable affinities raises another question: why were the higher affinity fragments not found in the XChem screen? One explanation could be that at the very high concentration of the fragment soaks used in this experiment, the more potent fragments disrupted the crystal lattice, hence preventing the detection of some highly potent ligands while leaving the weaker binding fragments that couldn't be detected on other biophysical techniques.

Therefore, the next step would be to test selected inhibitors that have been identified previously by SPR as manual X-ray soaks, in order to identify if there in any ligand's concentration dependency of crystal lattice quality.

Chapter VI

Biophysical Analysis of TbrPDEB1 using PDE-like Fragment's from ChEMBL by XChem and SPR

6.1 Introduction

The previous XChem screen (Chapter V) showed that overall from 466 attempted soaks 290 data sets were accepted by PanDDA and 31 positive data events were identified indicative of bound ligands. Since the XChem platform is a new development in using X-ray crystallography for fragment binding, it was important to see if the approach or the system could be optimised in order to minimise false negative experimental errors through improving its detection of potential high affinity binders that may disrupt the crystal lattice at high concentrations. Refining the current technique could have significance to industrial experiments as well as in academic studies.

The aim was to test the hypothesis that using the "standard" high ligand concentrations of soaks used in XChem fragment screening could result in a number of false negatives due to sample damage. This would be tested by using different fragment soaking concentrations as well as using different soaking times.

6.2 ChEMBL PDE-like Fragment's

A selection of fragments from the ChEMBL datasets with known PDE activity (performed by IOTA, PDE4NPD collaborators) were chosen to be tested on TbrPDEB1 catalytic domain where 31 fragments were selected, purchased and used for manual soaking, using different soaking concentrations and soaking times, Table 6.1.

It was hoped these variables would have allowed us to see if high affinity fragments may destroy the crystal lattice and identify the more appropriate concentrations and soak times for standard XChem protocols to ensure the highest affinity fragments are detected. It was important to perform manual soaking before proceeding to an XChem data collection as it would be easier to regulate concentration and time variables as compared to varying diffusion rates that could occur with echo dispensing used in the XChem process.

There were 5 conditions that were used during manual crystallization: ligand concentration of 40 mM and soaking for 24 hours, 30 mM ligand concentration for 24 hours of soaking and 40 mM ligand concentration for 48 hours soaking.

Ligand	Ligand ID	30mM/24hs	40mM/24hs	40mM/48hs	50mM/72hs	50mM/96hs
1	ChEMBL31877	1.67	x	1.39	X	X
2	ChEMBL88119	1.99	1.63	1.47	X	1.59
3	ChEMBL131164	2.26	1.99	X	X	X
4	ChEMBL124706	1.88	3.58	X	X	X
5	ChEMBL760	X	X	X	X	X
6	ChEMBL279898	1.86	X	1.56	X	X
7	ChEMBL189	X	1.78	1.87	X	1.49
8	ChEMBL58355	X	2.94	X	X	1.93
9	ChEMBL131181	1.83	1.71	1.83	X	X
10	ChEMBL330581	X	1.8	2.14	X	1.75
11	ChEMBL100112	X	1.97	1.52	X	1.55
12	ChEMBL1779264	X	X	1.48	X	X
13	ChEMBL1779265	X	X	X	X	X

14	ChEMBL619	X	X	X	X	X
15	ChEMBL255611	X	1.64	2.11	X	1.56
16	ChEMBL119506	X	1.76	1.84	1.89	X
17	ChEMBL118193	X	1.76	1.79	X	X
18	ChEMBL372055	X	1.64	1.19	X	1.75
19	ChEMBL451589	X	X	1.85	X	X
20	ChEMBL255421	X	X	1.65	X	X
21	ChEMBL190	X	X	X	1.75	X
22	ChEMBL2172707	X	X	1.79	X	X
23	ChEMBL3622905	X	X	1.47	X	X
24	ChEMBL45854	X	X	1.69	X	1.47
25	ChEMBL66732	X	X	1.52	X	X
26	ChEMBL270636	X	X	1.44	X	X
27	ChEMBL1779267	X	X	1.27	1.48	X
28	ChEMBL270635	X	X	1.48	X	X
29	ChEMBL484928	X	X	1.52	X	X
30	ChEMBL255130	X	X	1.38	X	X
31	ChEMBL255821	X	X	X	X	X

Table 6.1: Table 5.1 ChEMBL's ID and conditions that were used for each ligand together with the diffraction limits noted in Å data that was obtained (X denotes no processable data set obtained). The best hits were further analysed and refined using the CCP4 software suite.

6.3 Ligand – Protein Binding Mode Analysis

After running the XChem analysis 11 ligands were identified that gave the most promising, of those 5 ligand data sets were chosen for complete structural analysis. The current section represents data that was obtained after XChem run.

Ligand 1, ChEMBL31877, Figure 6.1, was found in chain B in hydrophobic clamp that was formed by residues Phe877 and Val840, Figure 6.2.
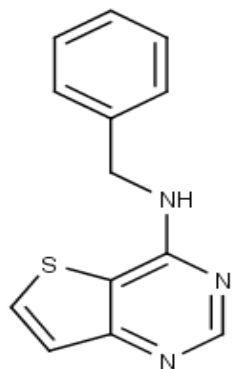
Figure 6.1: Ligand 1: ChEMBL31877, N-benzylthieno[3,2-d] pyrimidin-4-
amine

Figure 6.2: crystal structure of TbrPDEB1 catalytic domain and Ligand 1:
ChEMBL31877, chain B, hydrophobic clamp
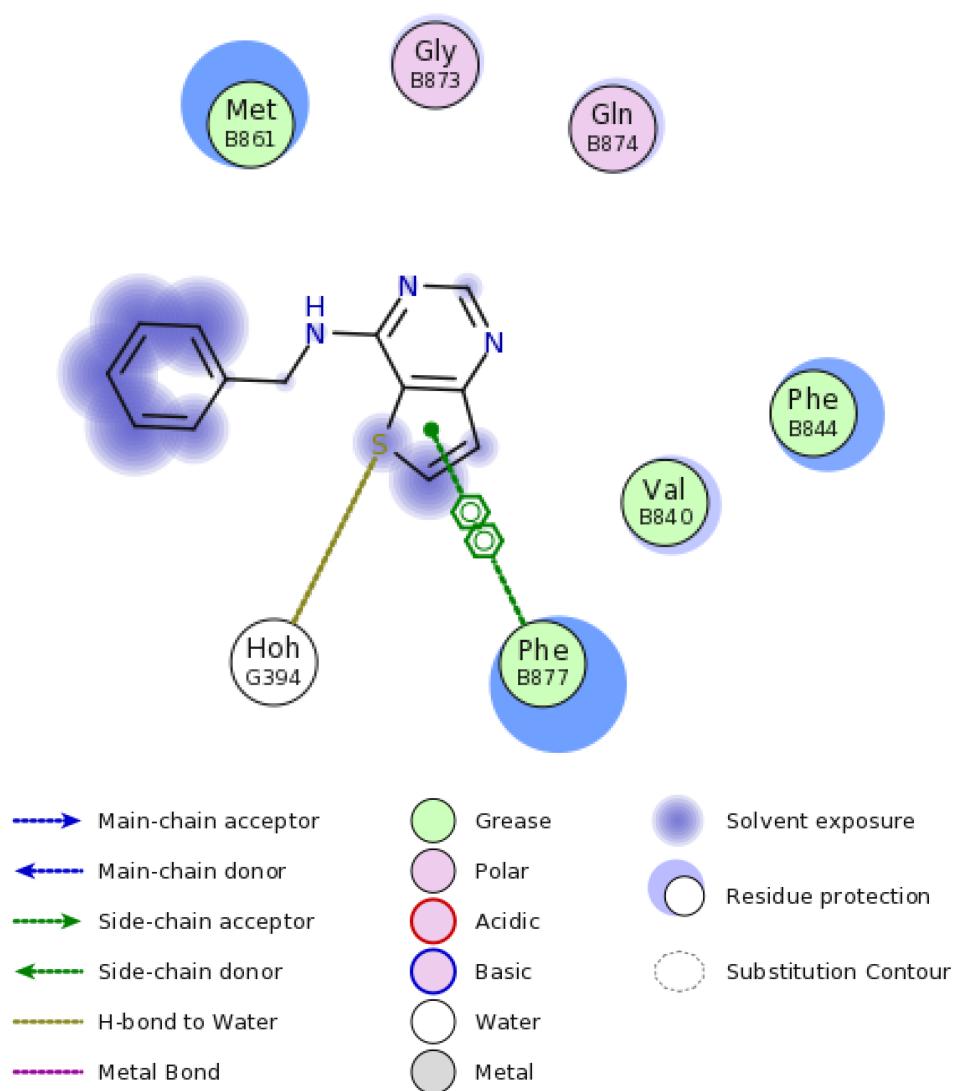
Figure 6.3: FLEV plot produced by COOT of the binding site of chain B, first binding site, Ligand 1: ChEMBL31877

The binding mode shows the ligand surrounded by amino acids: Met861, Gly873, Glu874, Phe844, Val840 and Pheny877, Figure 6.3. ChEMBL31877 formed 2 major types of interactions: hydrophobic interaction with Phe877 with which the ligand PI stacks and a sulphur containing hydrogen bond with one of the water molecule that was present in the binding site. An interesting feature of current binding was that it was close to the P – pocket residue Met861 and therefore computational analysis could reveal that further chemical synthesis could identify new

possible derivatives that could be more selective against TbrPDEB1 than human PDEs.

Ligand 3: ChEMBL131164, Figure 6.4, was only bound at 1 site at chain B occupying the hydrophobic clamp between residues Phe877 and Val840.
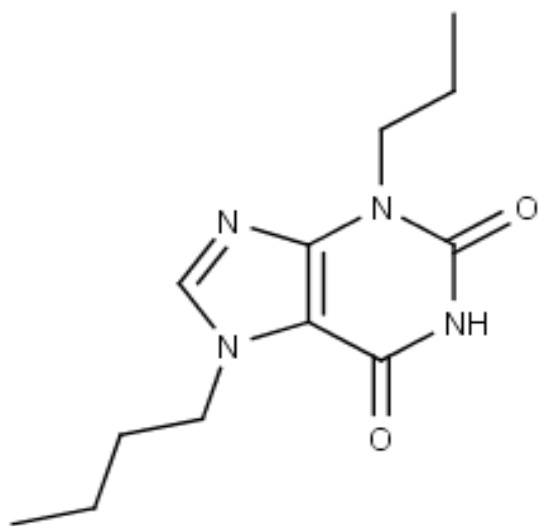
Figure 6.4: Ligand: ChEMBL131164, 7 – butyl-3-propyl-2,3,6,7-tetrahydro-1H-purine-2,6-dione

According to the Figures 6.5 and 6.6 there was a formation of Hydrogen bond between one of the oxygen atom of the benzene ring with the protein side chain of Met861 residue that forms a part of the P – pocket.
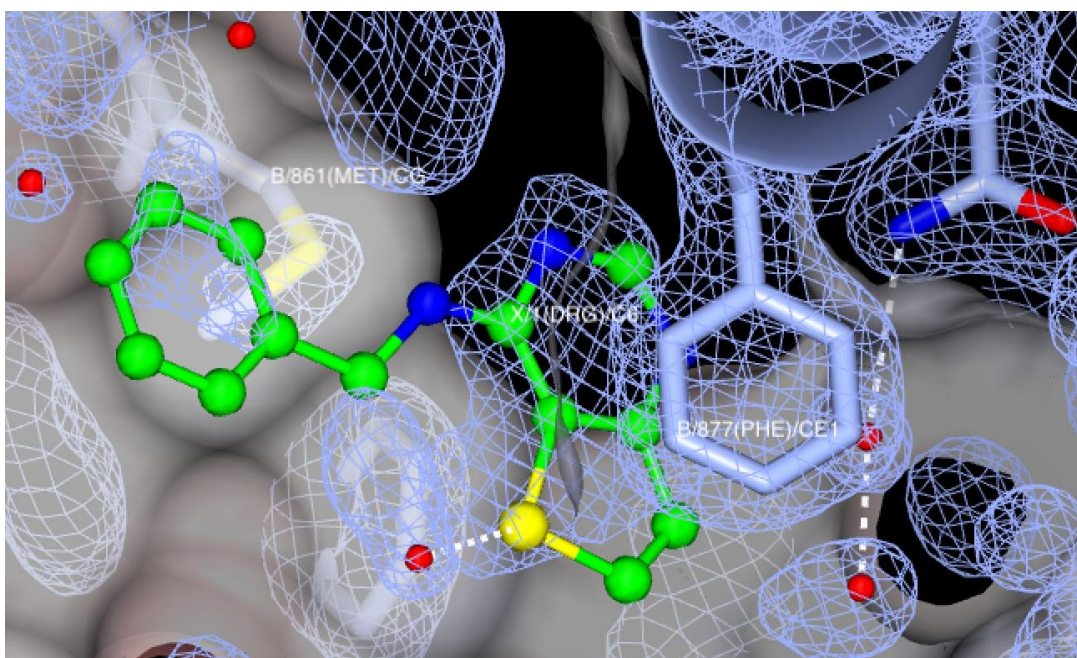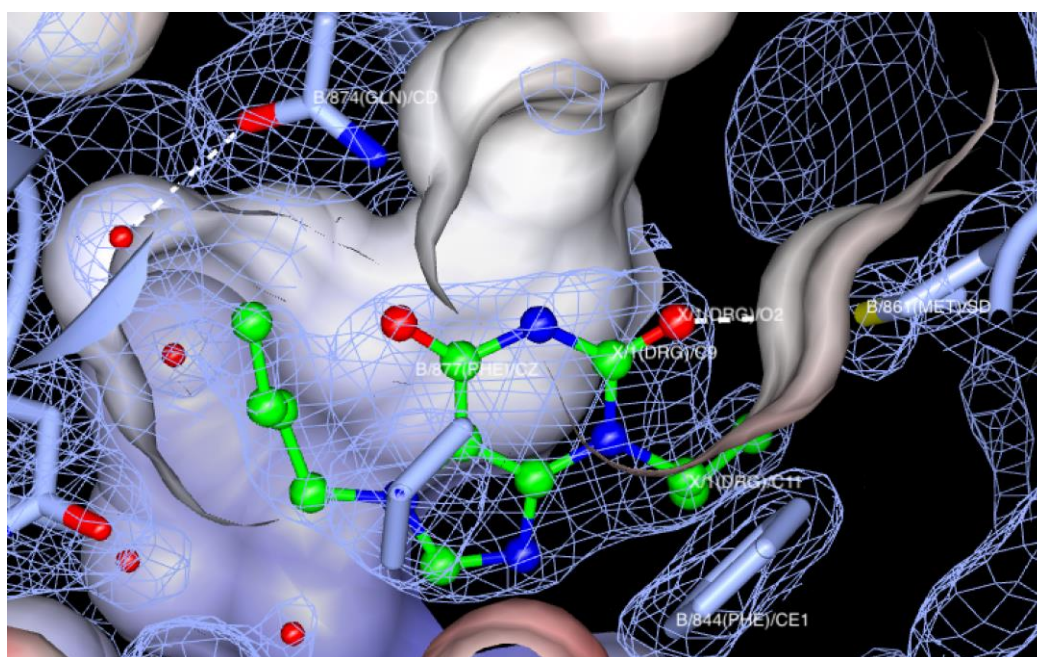
Figure 6.5: crystal structure of T. brucei PDEB1 catalytic domain and Ligand 3: ChEMBL131164, chain B, hydrophobic clamp
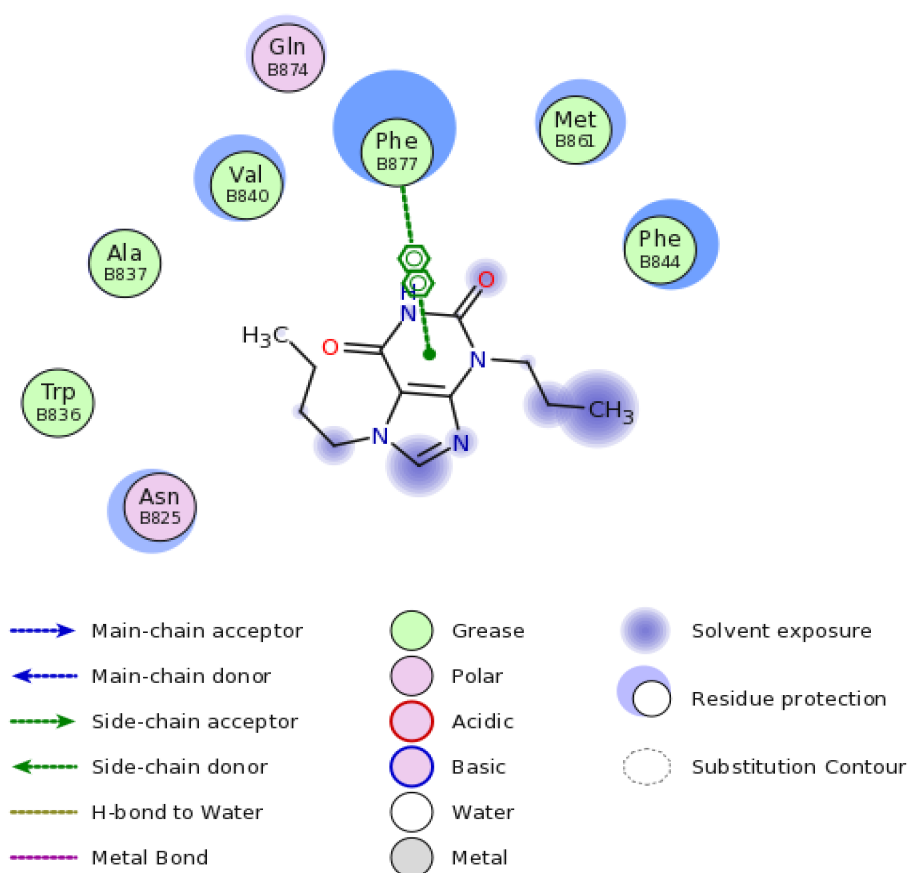
Figure 6.6: FLEV plot produced by COOT of the binding site of chain B, Ligand 3: ChEMBL131164

Figure 6.5 shows Hydrogen bond formation between one of the water molecule and Glu874. Such interaction improved binding affinity of Ligand into the hydrophobic clamp. Figure 6.6 represents ligand binding site environment that was build using Coot's software. According to the current figure there was interaction between Phe877 residue and an aromatic ring of the ligand.

The third ligand that was used for ligand binding mode analyses was **Ligand 22: ChEMBL2172707**, Figure 6.7.
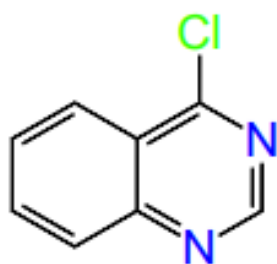
Figure 6.7: Ligand 22: ChEMBL2172707, 4-chloroquinazoline

Figure 6.8: crystal structure of T. brucei PDEB1 catalytic domain and Ligand 22: ChEMBL2172707, chain B, hydrophobic clamp
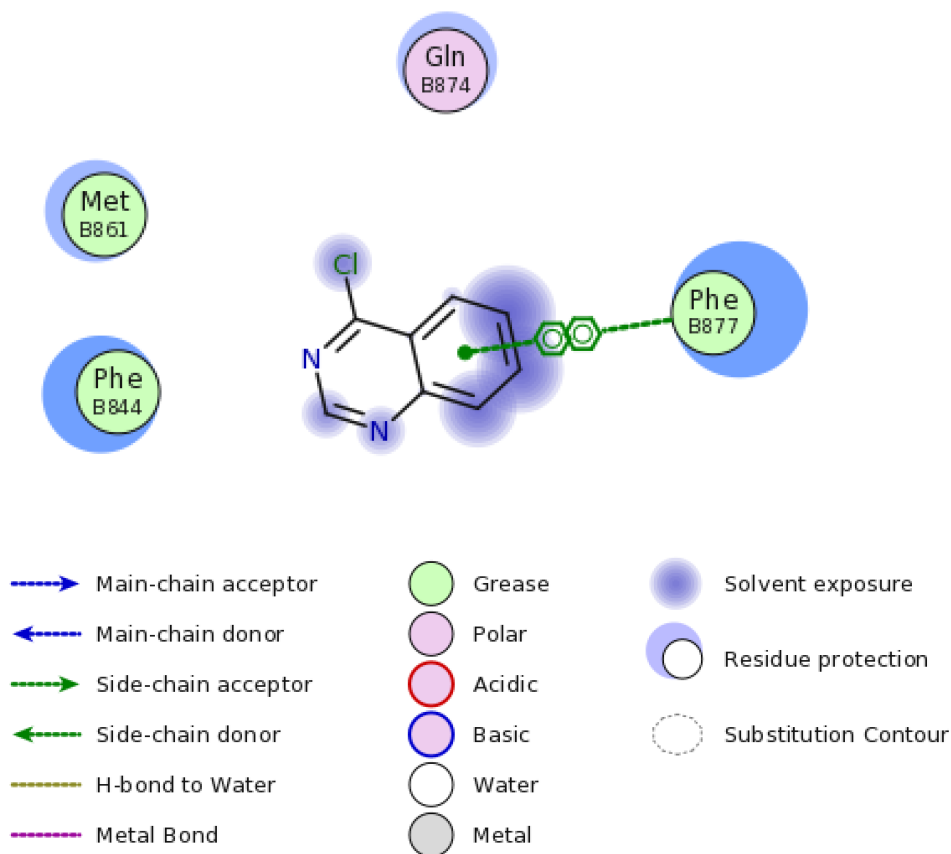
Figure 6.9: FLEV plot produced by COOT of the binding site of chain B, first binding site Ligand 22: ChEMBL2172707

Ligand 22: ChEMBL2172707 was flanked by residues including Gly874, Met861, Phe844 and Phe877, Figure 6.9. The benzene ring formed a face-to-face π - stacking interactions with the aromatic ring of Phe877.

Ligand 12: ChEMBL1779264, Figure 6.10. The ligand was once again bound in the hydrophobic clamp of the catalytic domain, Figure 6.11.
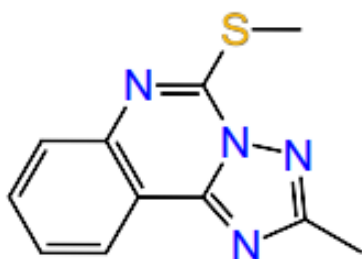
Figure 6.10: Ligand 12: ChEMBL1779264, 2-methyl-5-(methylsulfanyl)-[1,2,4] triazol[1,5-c]quinazoline

The binding mode shows the Ligand ChEMBL1779264 surrounded by amino acids: Val840, Phe877, Tyr668, Met785, Met861, Gly873, Phe844 and Gly874, Figure 6.11.

ChEMBL1779264 formed 2 hydrogen bond interactions with 2 water molecules as well as π - stacking interactions between Phe877 residue of the protein and the aromatic triazene ring of the ligand. ChEMBL1779264 is also located close to the hydrophobic protein residues such as Met861 and Met868 which are the residues of the P – pocket. Therefore, the current ligand could be used as a starting point for synthesis of derivatives that would be directed towards the P – pocket more in order to be more selective towards the parasitic PDEs rather than human PDEs.
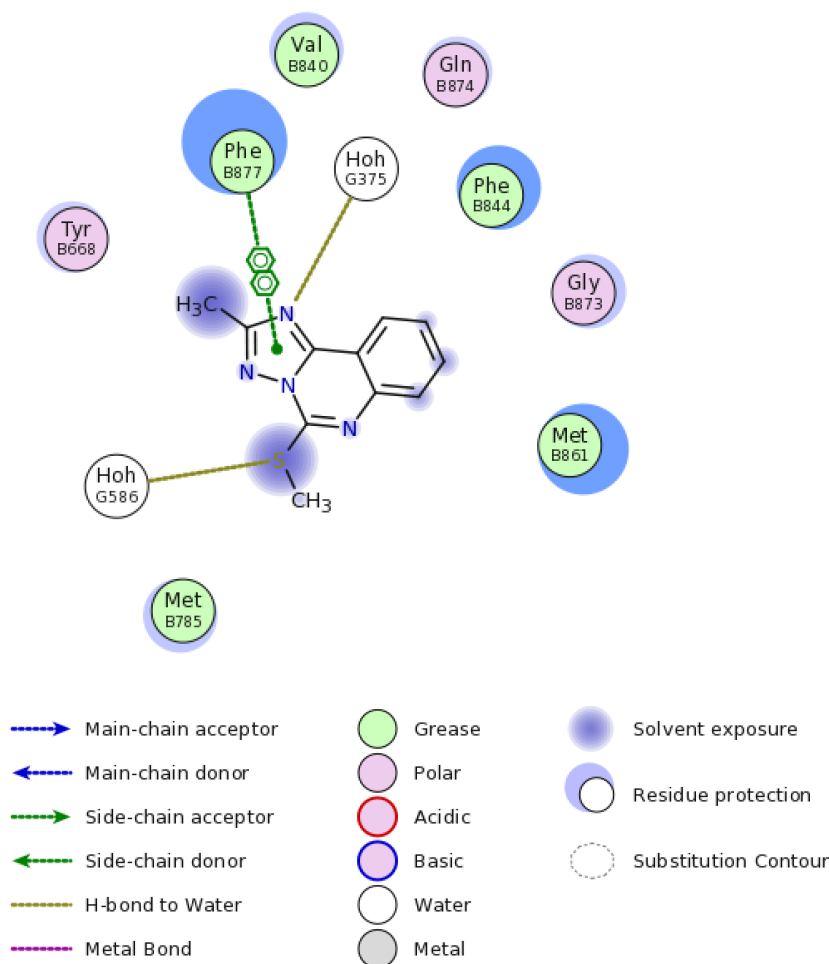
Figure 6.11: FLEV plot produced by COOT of the binding site of chain B, first binding site Ligand 12: ChEMBL1779264
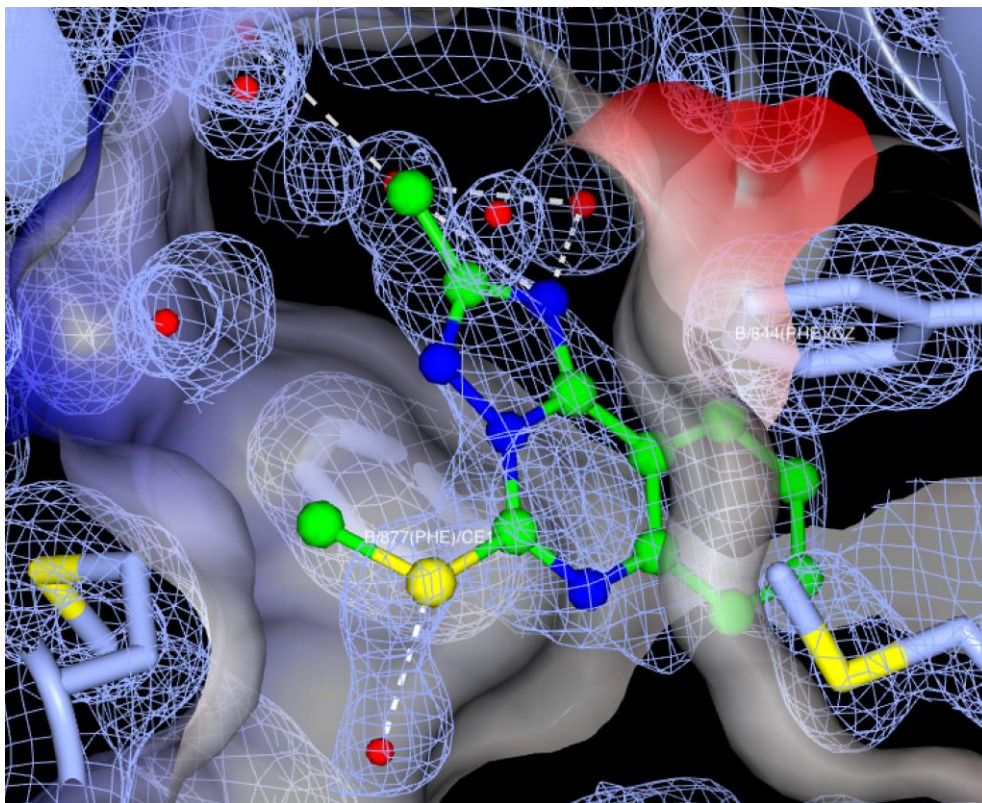
Figure 6.12: crystal structure of T. brucei PDEB1 catalytic domain and Ligand ChEMBL1779264, chain B, hydrophobic clamp

The last ligand that would be refined further was

Ligand 9: ChEMBL131181, Figure 6.13.
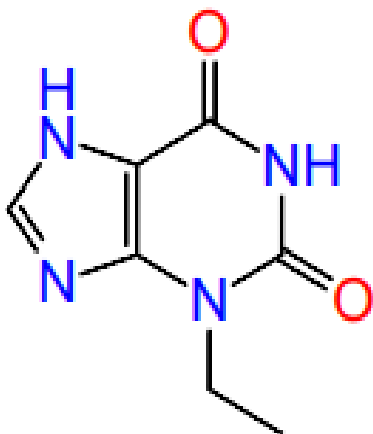
Figure 6.13: Ligand 9: ChEMBL131181, 3-ethyl-2,3,6,7-tetrahydro-1H-purine-2,6-dione

Ligand 9: ChEMBL131181 was found bound in the hydrophobic clamp of chain B and was surrounded by residues: Tyr668, Phe877, Phe844, Val840, Gly874, Val826, Asp825 and Iso823, Figure 6.14 and Figure 6.15.
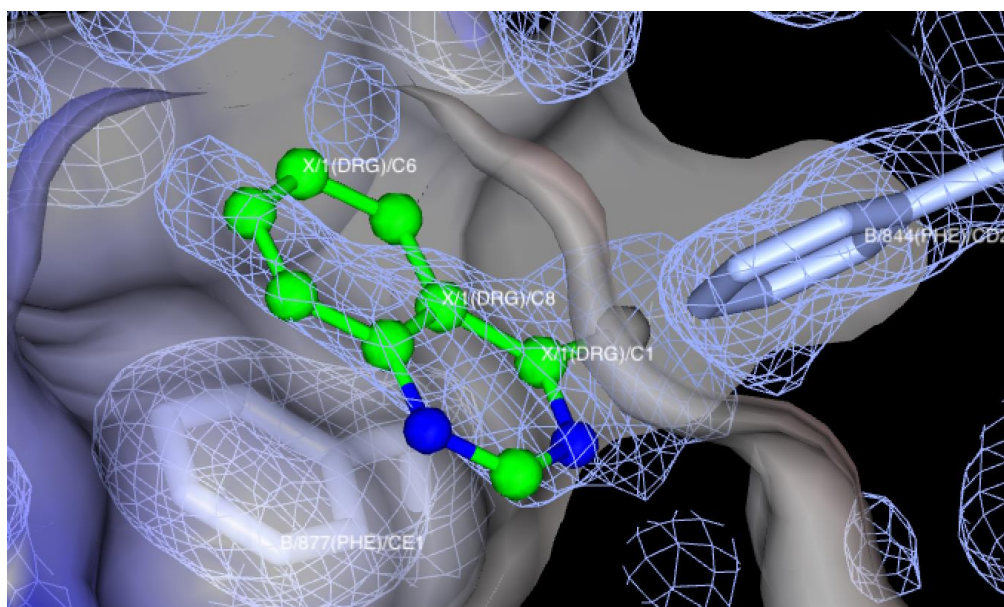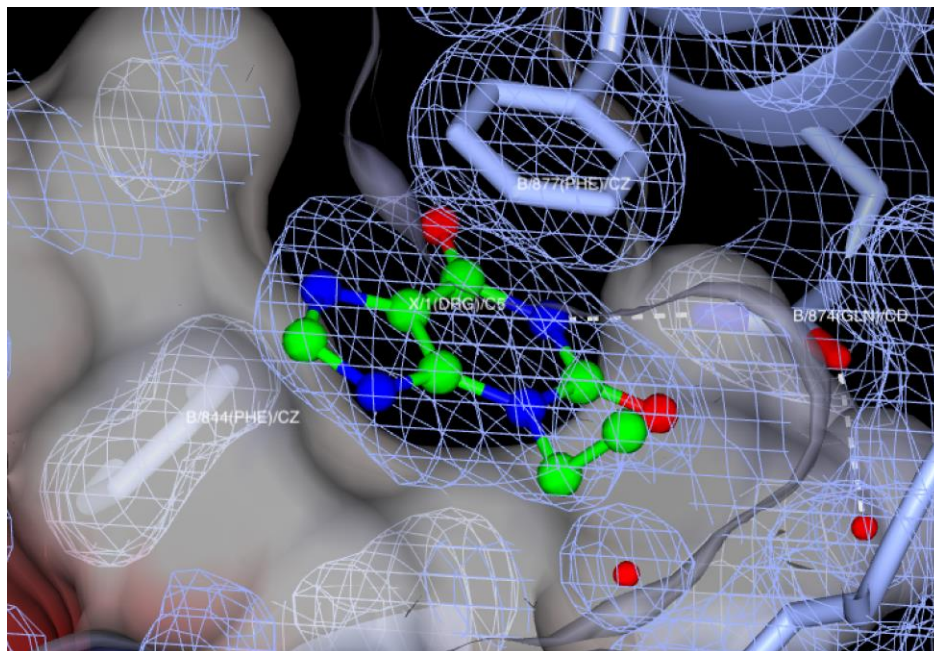

Figure 6.14: crystal structure of T. brucei PDEB1 catalytic domain and Ligand 9: ChEMBL131181, chain B, hydrophobic clamp
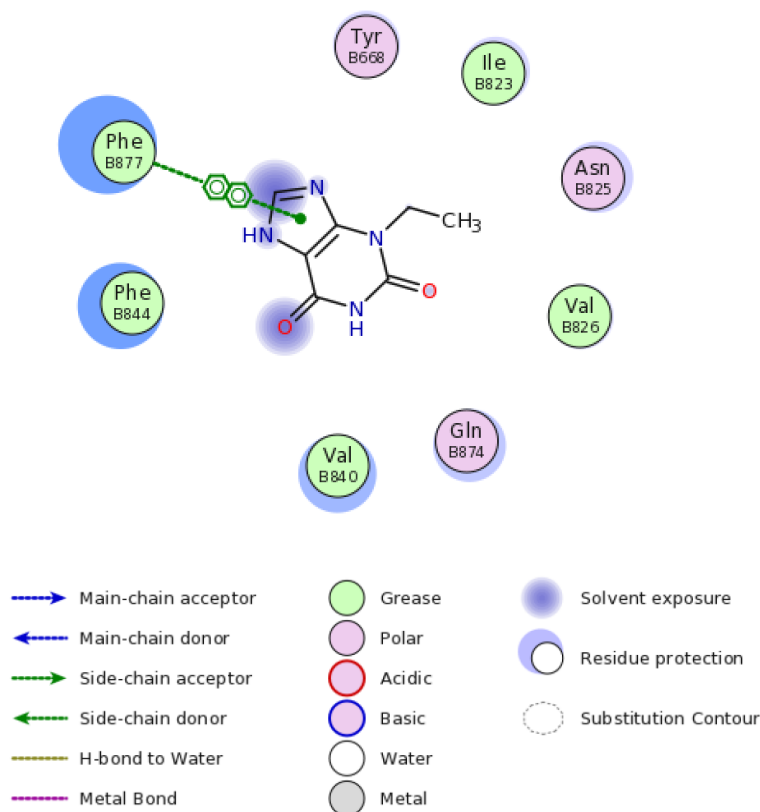
Figure 6.15: FLEV plot produced by COOT of the binding site of chain B, first binding site, Ligand ChEMBL131181

There are 2 key interactions between the ligand and the protein: one is the formation of π - stacking interactions between Phe877 and the aromatic indazole ring of the ligand, and another is the hydrogen bond formed between the conserved Gln874 residue and the NH atom of the ligand. Such interactions stabilise ligand – protein binding mode. Ligand ChEMBL131181was also bound in chain A and it was similarly positioned in the hydrophobic clamp, Figure 6.16.
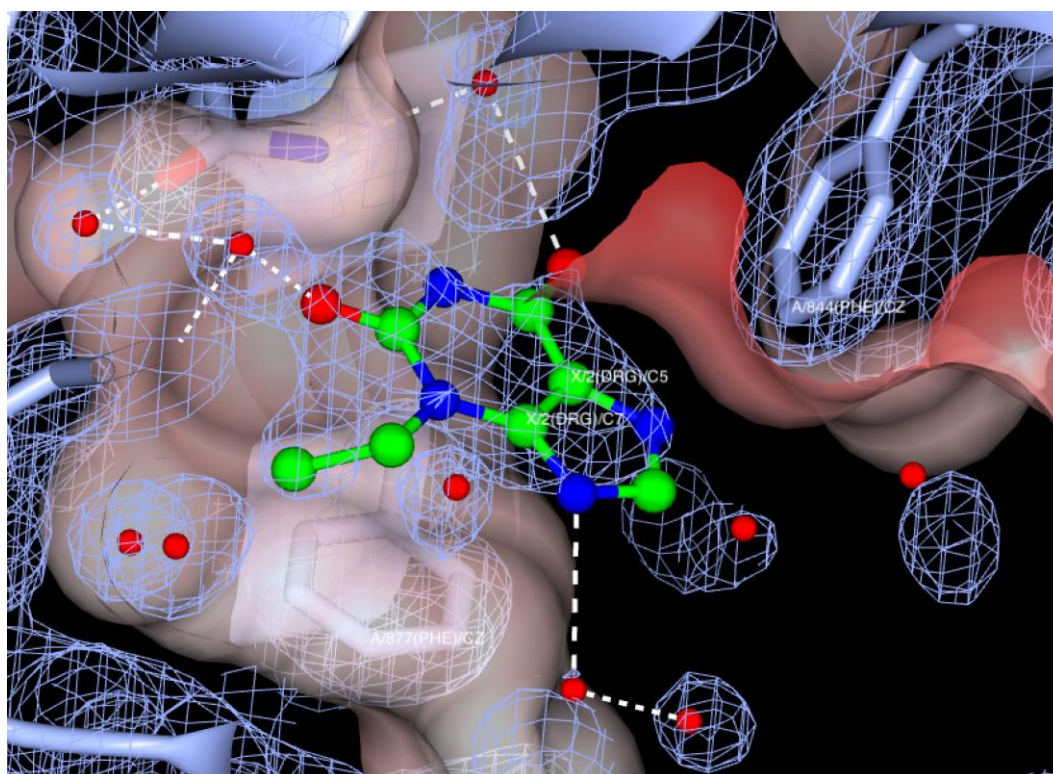
Figure 6.16: crystal structure of T. brucei PDEB1 catalytic domain and Ligand ChEMBL131181, chain A, hydrophobic clamp

In this binding mode ligand 9: ChEMBL131181 was surrounded by such amino acids as: Phe877, Phe844, Met861, Gly874 and Asp825, Figure 6.17. Ligand ChEMBL131181 formed 3 hydrogen bonds with 3 water molecules and a π - stacking interactions between the indazole ring of ligand and aromatic side chain of the protein residue Phe877. In comparison to the chain B, where only 1 H bond interaction was present,

chain A also had water molecules that formed 3 H bonds with the ligand, one of them spanning to Gln874.
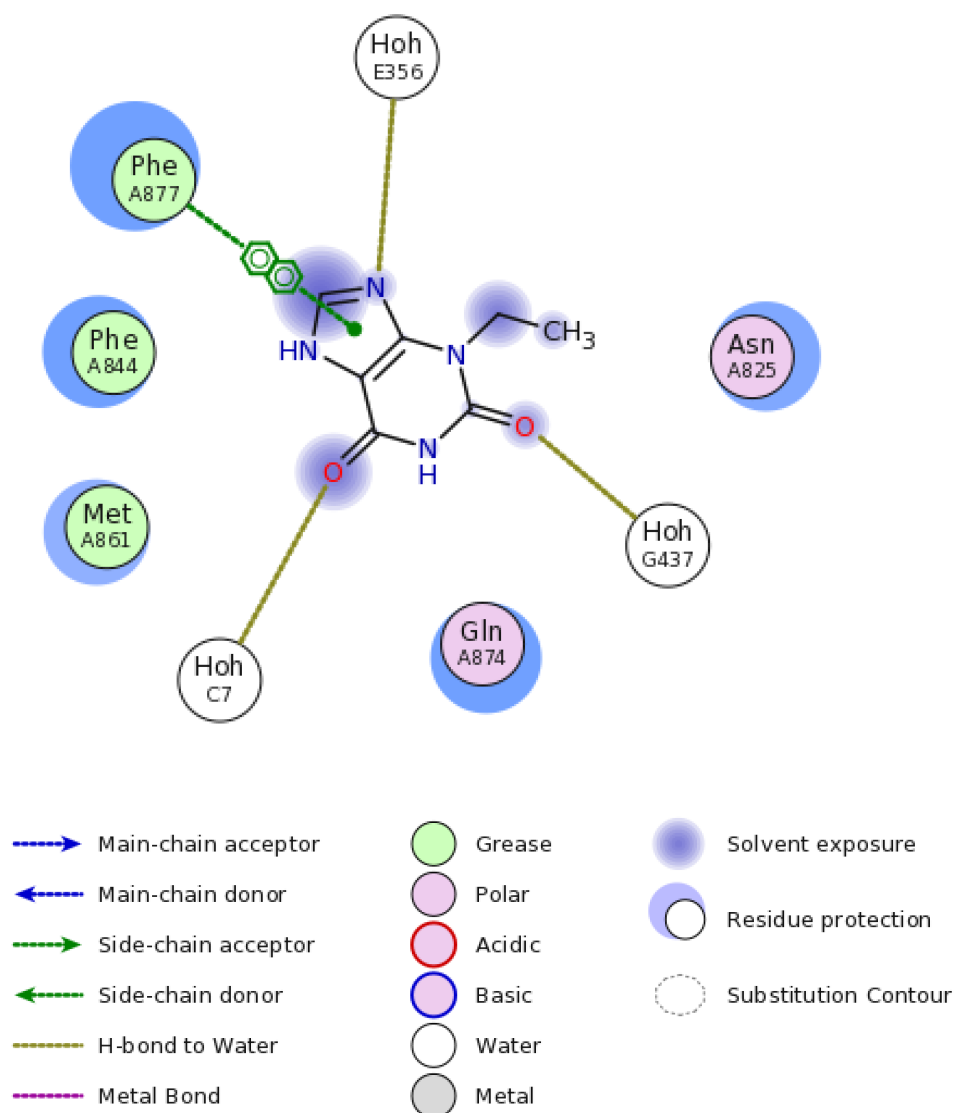
Figure 6.17: FLEV plot produced by COOT of the binding site of chain A, second binding site Ligand ChEMBL131181

Fragment	Structure	Binding Mode
Ligand 1, **ChEMBL31877**		1 binding site: chain B – hydrophobic clamp
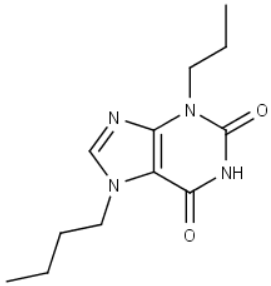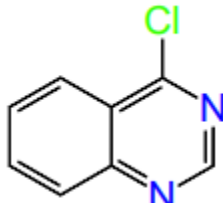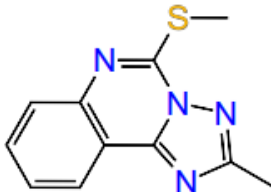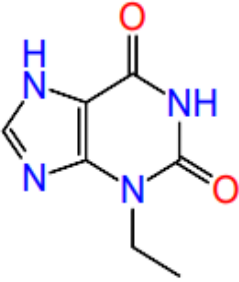
Ligand 3: **ChEMBL131164**		1 binding site: chain B – hydrophobic clamp
Ligand 22: **ChEMBL2172707**		1 binding site: chain B – hydrophobic clamp
Ligand 12: **ChEMBL1779264**		1 binding site: chain B – hydrophobic clamp
Ligand 9: **ChEMBL131181**		2 binding sites: chain A – hydrophobic clamp chain B – hydrophobic clamp

Table 6.2: summary of solved structures

6.4 XChem of low concentration soaks

The aim of this experiment was to further identify optimal concentrations for 3 chosen ligands that were analysed in Section 6.3, where duration of soaking and ligands concentration would be varied.

In the previous section 5.3, 5 compounds were analysed, and its ligand binding mode determined when the experiment was performed as manual soaking. These compounds were then chosen for an XChem run where 10 points of concentration gradient were used. It was decided to use 2 time periods to soak fragments with the protein: 2 hours and 4 hours. Therefore, by using 10 concentrations, 2 soaking time, with 5 compounds there were 200 samples that would require using 12.5 pucks and 7 hours of beamline. To accurately prepare the fragments dilutions a mosquito liquid dispenser was used to dispense into an echo compatible 384LDV source plate. The final planned concentration of 250 mM was not achievable due to the 20% DMSO soak plus 1M stock compound limit, so a 9-point titration was designed. The final concentration used were: 1 mM, 3 mM, 5 mM, 10 mM, 20 mM, 50 mM, 100 mM, 150mM and 200 mM. There were 2 different protein crystallisation batches that were used in current experiment. Batch 1 contained small fragile crystals where 80% of crystals didn't show any diffraction data. Batch 2 contained good size crystals where 150 um loops were used. At 4 hours of soaking fragments x0143 (Ligand ChEMBL31877) and x0193 (Ligand ChEMBL131181) suffered and the crystal lattice was damaged. There were 3 ligands that showed best diffraction data, where the binding event could be observed in high concentration soaks as well. Unfortunately, due to issues with PanDDA maps it was hard to identify the differences in binding intensity between different concentrations hence at that time it was not possible to plot an accurate dose-dependent curve. These 3 ligands were chosen as interesting targets for docking analysis (Ligands 1, 3 and 9), and SPR screening.

6.5 SPR Data Validation

The XChem experiment revealed that the best binders were ChEMBL31877, ChEMBL131164 and ChEMBL131181 that were all

xanthine derivatives and where their chemical structures were closely related to IBMX ligand. Human PDE4 and TbrPDEB1 catalytic domains were immobilised for SPR. The samples purity was accessed on Liquid chromatography – mass spectrometry (LC-MS) that separated molecules and analysed masses of proteins. The maximum concentration of ligands used was 100 µM. The results were as following.

- **_Ligand_** ChEMBL131181
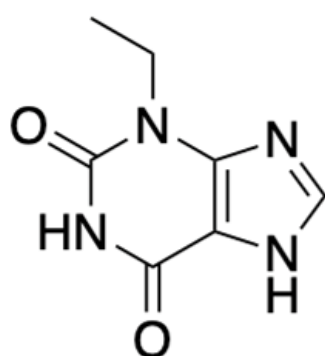
Figure 6.18: 3-ethyl-2,3,6,7-tetrahydro-1H-purine-2,6-dione

Protein	T. brucei PDEB1	hPDE4
SPR data		

Table 6.3: SPR data of ChEMBL131181, 2 target proteins: T. brucei PDEB1 and hPDE4

Table 6.3 represented data of protein – ligand binding analysis, where there was no confirmed binding in either of the proteins as no association and dissociation phases can be observed.

- *__Ligand__* ChEMBL31877
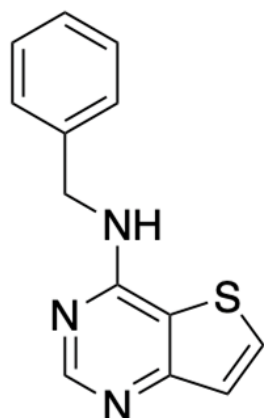
Figure 6.19: N-benzylthieno[3,2-d]pyrimidin-4-amine

Protein	T. brucei PDEB1	hPDE4
SPR data		

Table 6.4: SPR data of ChEMBL31877, 2 target proteins: T. brucei PDEB1 and hPDE4

According to the data presented in Table 6.4 there was evidence of specific binding between the ligand and human PDE4 while the sensorgram for TbrPDEB1inferred unspecific ligand binding.

The last ligand that was measured on SPR was ChEMBL131181 where the results were as follows, Table 6.5.
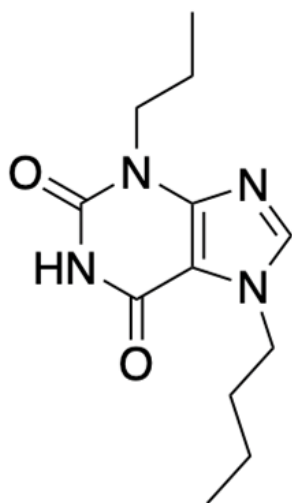
Figure 6.20: 7 – butyl-3-propyl-2,3,6,7-tetrahydro-1H-purine-2,6-dione,

Protein	T. brucei PDEB1	hPDE4
SPR data		

Table 6.5: SPR data of ChEMBL131164, 2 target proteins: T. brucei PDEB1 and hPDE4

According to the data presented in Table 6.5, there was unspecific binding between the ligand and TbrPDEB1 while hPDE4 didn't show any binding and no association and dissociation phases were observed.

6.6 Conclusion

31 compounds were screened manually, of these 30 showed binding at one or more concentrations or soak time.

Manual inspection of these 30 binding events were analysed and revealed that only 5 compounds were interesting from the drug discovery

perspective point. Hence, these compounds were chosen to be further investigated by using XChem platform to explore optimal concentrations as well as different soaking time, such as 2 hours and 4 hours. Unfortunately, the XChem processing and PanDDA analysis failed for the final XChem experiment and it was not possible to build a dose-dependency curve for calculation of a binding. Out of 5 compounds 3 compounds were further analysed by the SPR run, to see if any measurable binding affinity could be obtained and if any selectivity was present between human and parasitic PDE.

However, it was clear from the manual soaking experiments for ChEMBL131181and ChEMBL131164 that there was an indicative dose response and negative effect on the crystals for high concentration longer duration soaks supporting the hypothesis.

There were 2 compounds out of 3 that showed SPR response: ChEMBL31877 and ChEMBL131164 where 1st fragment showed specific binding with human PDE4 protein while the 2nd one showed unspecific binding with the parasitic PDEB1. Hence, current results couldn't be used further for drug binding mode validation and current experiment should be optimized and rerun again in future.

The focus of this experiment was to optimise an XChem protocol based on high affinity fragments. The protocol could be used by either Academic or Industrial projects to ensure that high affinity fragments could be crystallographically detected. Some crystal systems can't withstand high concentration soaks or soaking time therefore destruction of crystal lattice will affect the effectiveness of the experiment. Unfortunately, due to PanDDA issues the failed to determine binding affinities. Therefore, the next step would be to repeat the experiment where PanDDA software issues will be resolved.

Chapter VII

Summary and Discussion

7.1 Novel Expression and Purification of Parasitic PDEDs from L. infantum, L. donovani and T. brucei

Previous experiments that were performed by other lab members were focused on the expression of PDED from T. brucei and T. cruzi. There was no preliminary data about PDED protein expression and purification from all 3 species. Previous experiments on T. cruzi using different expression systems, such as: bacteria, insect and mammalian cells didn't show any differences in soluble protein expression, either full length or catalytic domain protein constructs remained insoluble. Expression studies focused initially on the T. cruzi species. Small scale expression of similar constructs for L. infantum and L. donovani showed that protein was insoluble as it was only observed in a pellet fraction. Since the previous work had shown that different expression systems didn't appear to improve PDED solubility it was decided to perform mutations where hydrophobic amino acids were mutated into polar amino acids in order to make the protein more soluble. Different constructs were designed, and a variety of mutations introduced. Unfortunately, no positive outcome was observed, and the protein still remained insoluble. The next approach was to clone the construct into a pOPIN vectors where solubilisation tags were linked to the construct. According to western blot analysis, TcrPDED was observed in the supernatant fraction, however the signal was so weak that

it was obvious that protein was still predominantly in its insoluble form. TbrPDED catalytic domain was also studied and revealed the same pattern of protein behaviour as it was only observed in the pellet fraction. Fusing the target proteins with solubilizing tags by use of the pOPIN vectors resulted in a small amount of soluble expression. This small amount of soluble material was sent to the university VU to test the catalytic activity of the protein a cAMP assay. Unfortunately, no activity was observed probably due to the fact that any expressed protein had degraded.

As T. brucei and T. cruzi are more closely related species than Leishmania it was hoped that studying Leishmania PDEDs would overcome the potential issues with the solubility. L. infantum and L. donovani were used as targets where catalytic domains and full-length proteins were amplified and cloned into a standard pET 15 vector. Standard expression conditions were used, and no soluble protein was obtained. It was clear that all PDED proteins were prone to misfolding and aggregation. However, it was decided to pursue similar optimizations method as attempted with T. cruzi where different constructs were designed and cloned into pET vector and pOPIN vector suite. Western blot experiment showed similar patterns of expression where a signal was only observed in a pellet fraction. There were three designed constructs that were cloned into a pET15b vector and 4 vectors of pOPIN vector suite, such as: pOPINF, pOPINM, pOPINJ and pOPINS3C. By using pOPIN vectors it was possible then to use other expression systems like insect cells and mammalian cells. There was no difference in expression level between the three constructs as well as between using different vectors where solubilisation tags didn't reveal any improvement in expression level. Following further analysis of additional expression systems, pCold was identified as a more suitable for expressing of insoluble proteins. The system has different vectors where each of them

contain cold – shock Protein A *(cspA)*, which acted as a promoter for expressing recombinant proteins. Induction of protein's expression was performed at low temperature – 15 °C. Low temperature induction was used to suppress the expression of E. coli proteins and inhibit protease activity that negatively result on recombinant protein expression. There were 2 vectors that were used: pColdI and pColdTF. The main difference between these 2 vectors was the addition of a trigger factor (TF) into the pColdTF vector that was fused with PDED catalytic domain. There was a significant difference in protein expression comparing to the protein being cloned into pET15b vector and pOPIN vectors. The pColdTF vector gave more soluble protein in a supernatant fraction. Trigger factor was acting as a chaperone to improve protein solubility. As there was an improvement in soluble protein yields with the fused chaperone it was decided to purchase Takara plasmid chaperone set where 5 plasmids contained a combination of different chaperones. The co-expression was performed using small scale culture in order to see which plasmid with chaperones enhance protein solubility the most. pGro7 plasmid that contained 2 chaperones GroES and GroEL was used in the following experiments as the most soluble material was obtained with the co-expression with that plasmid. GroEs and GroEL chaperones work together to control the folding of proteins from unfolded or partially folded state. It has been proven that some recombinant proteins that can't be expressed in E. coli cells were finally successfully purified by co – expressing it with the GroES and GroEL complex *(102)*. Main advantage of using such chaperone complex is the ability of it to refold proteins from its denatured state as well as if protein was partially folded only. Previous experiments performed on Rubisco (ribulose bisphosphate carboxylase – oxygenase was denatured prior its incubation with the complex and ATP in order to try to renature it) using GroES and GroEL showed that 80% of protein could be renatured.

There was a substantial difference in protein expression between pColdI and pColdTF vectors where in the 2nd place PDED was also fused with the trigger factor. So initially protein was promoted by trigger factor whereas chaperon complex GroES/GroEL improved and speed up the folding reaction. The inside cavity of GroEL is made by hydrophilic side chains that speeds up the folding rate. When PDED protein was expressed by itself the folding reaction was inhibited by protein aggregation with misfolded parts that were observed using SDS-PAGE and Western blot analysis.

Takara chaperone plasmids were used to identify the most preferable chaperone complex for TbrPDED solubilisation. In these experiments, another Takara plasmid showed the best result, pG-KJE8 which carries 5 chaperones where 2 of them were GroES/ GroEL complex. Co-expression of protein in the presence of chaperones showed that it was possible to solubilize it for further purification. However, larger scale expression of TbrPDED didn't yield enough of protein material and also the protein was shown to degrade more quickly than in the case of Leishmania PDED, hence it was not possible to perform cAMP enzyme activity assays.

In contrast, LinfPDED that was initially cloned into pColdTF vector and was fused with TF was used in Lonza assay and clear enzyme activity was detected, but again the fused PDED was still susceptible to protein degradation. Attempts to reduce degradation by removal of the trigger factor from LinfPDED, and a second Ni-NTA purification run, showed it was not possible to separate cleaved material from HisTag-PDED. The final outcome of the study showed that expression and purification systems have been developed for PDED catalytic domain protein, using different cloning and expression approaches, its optimizations and further

purification did not yield robust, stable protein for downstream biophysical analysis.

Although low levels of soluble PDED catalytic domain were produced, it showed enzymatic activity against secondary messenger cAMP which proved that it was a true phosphodiesterase. Unfortunately, it was still clear that PDED was prone to aggregation. Therefore, further studies would be required to investigate the factors that can be used in order to maintain protein structure intact in solution so further biophysical analysis can be performed, such as NMR, X-ray and SPR.

7.2 Biophysical Analysis of Human PDE5 using NMR, XChem and SPR

IN a wider study of fragment screening methods and to find additional PDE fragments and as part of the development of Fluoro fragment library human PDE5 was used. 589 fluorine containing compounds were provided by Maybridge and analysed in NMR, SPR and XChem experiments. The first screen was SPR where the strongest binders according to SPR sensograms were determined. It was important to run SPR first as being less time consuming and resources consuming process among all other 2. There were 2 proteins used in SPR analysis: human PDE5 and human BRD4-BDE1 domain to help determine if the fragments were specific to the target and also because the BRD4 data could be used for further studies. Table 4.2 represented SPR data of the strongest hits that were observed with human PDE5 where percentage response and Kd values were determined.

As this was a fluorine fragment file 19F CPMG was used where spectra should show only 1 peak per Fluorine atom being present in the fragment. After spectral analysis, the list of suitable well-behaved fragments was

decreased to 421 compounds. Out of 28 mixtures 2 mixtures failed and excluded from further experimental analysis. TopSpin software was used for analysing and comparing each mixture with the protein and mixture without the protein that was used as a control. To classify fragment as a strong or a weak binder, integrals of peak drops were calculated using the software where the final number was given as a percentage in peak drop. More than 50% in signal decrease was considered as a strong binder to PDE5 and hence less then 50% as a weak binder. Analysis of each mixture revealed that there were 101 fragments that were considered as strong binders with drop in signal around 95%. These fragments were further selected for manual crystallization since current fragment were positive hits in SPR and NMR screens. Then it was important to analyse the drug binding mode by analysing its diffraction data. 34 fragments were selected for X – ray where 22 fragments were classified as strong binders in NMR screen but not in SPR run, 2 fragments out of 34 were strong binders in SPR screen but weak binders in NMR run, 10 fragments were confirmed as strong binders in both NMR and SPR screens. Since differences between strength of binding was observed X–ray would reveal more details regarding ligand – protein binding mode. Out of 34 crystals 32 diffraction data sets were collected and was analysed manually. X – ray showed 13 strong binders that were also confirmed by NMR screen, 11 fragments that were confirmed by X-ray, NMR and SPR, 3 fragments were confirmed by XChem, NMR and manual X – ray but 2 of them were missed in SPR run. The final result was that out of 421 compounds only 1 fragment HTS04341 have been confirmed by all 4 techniques as a strong binder.

One of the main concerns that was raised was why only 3 hits have been identified by XChem run while 32 crystals were successfully detected with

manual soaking and where 13 fragments showed strong binding mode with PDE5 according to SPR and NMR screens.

This issue could due to the same problem that has been encountered in previous XChem experiment with TbrPDEB1.

Since XChem is as a relatively new platform there is need of further protocol optimization to avoid false negatives. A second important technical issue that has to be resolved is PanDDA maps analysis where it will be possible to design dose-dependence curve hence enabling to develop a generic protocol that can be used with different targets.

7.3 Biophysical Analysis of T. brucei PDEB1 using XChem platform and SPR screen

As suitable yields of a parasitic PDED could not be obtained focus switched to developing additional ligands for the validated TbrPDEB1, as TbrPDEB1 was previously expressed, purified and crystallised and the protocols were known. This system could be used to develop tool compounds that would be a starting point for developing selective fragments for further testing against other parasitic PDEs. A second important point was to find the binders that would be classified as strong binders by more than one biophysical technique. Since TbrPDEB1 was a stable crystal system it was chosen to be used in XChem. The DSI – Poised library was provided by Diamond Synchrotron where 466 fragments in the library. During mounting it was identified that almost 20% of crystals soaked were destroyed. There were 335 crystal data sets loaded into a run. After X-ray data collection the XChem explorer software was used where several rounds of analysis were performed, such as: DIMPLE and PanDDA inspect where restraints were developed, and even low binding affinity sites were found. PanDDA inspect was performed

manually and 290 samples were investigated in terms of binding sites and how many events were observed in each data set. During the analysis, the user can decide if the event is interesting and if the binding site looks confident. Out of 215 accepted data by PanDDA 131 samples were classified as interesting, with 61 hits being confident. In order to refine structures data sets were manually processed with CCP4 i2. After careful analysis 20 hits were identified as interesting binders and were refined. Next step was to analyse the novelty of fragment protein binding from chemical point of view. As it has been mentioned earlier, one of the main criteria in current study was to find compounds that would be selective between human PDEs and parasitic PDEs. The distinct feature of parasitic PDEs was that it has parasitic cavity called P – pocket that wasn't found in human proteins. Therefore, it was important to investigate if any of those interesting hits would be suitable for further synthetic chemistry follow up in order to grow it towards the P-pocket. Analysis focused on 6 compounds that were chosen for computational studies in collaboration with the Free University of Amsterdam (VU).

Although XChem gives atomic resolution information on the fragment binding interactions the affinity of the interaction is unknown and current computational methods are not accurate enough to calculate and rank the affinity of fragments. So, the next step was to perform SPR on the 6 compounds. To check the experiment accuracy as well as to see how selective the fragments were 3 proteins were used in the SPR study: TbrPDEB1 and PDEB2, and human PDE4. All 3 proteins were PDE catalytic domains only. According to the SPR sensograms, no strong binders were observed. The only fragment that showed any dose response was the derivative ZINC813384 of the compound 0074 where association and dissociation curves were observed against all 3 immobilised proteins

and it was slightly stronger in TbrPDEB1 and PDEB2 runs than in human PDE4.

The final statistics data were as follows: out of 466 compounds less than half of the library (215) was processed through PanDDA inspect and of 20 fragments classified as interesting hits 6 were screened on SPR. Only 1 fragment showed a response and was a confirmed binder in both biophysical methods.

Out of all crystals that were used for XChem run 20% of soaks were not mounted due to crystal lattice destruction, while the other 30% did not give full high resolution data sets. As a result, only 50% of library was crystallographically analysed.

Crystal's mounting was performed manually, so some crystals were soaking for 2 hours and some for 4 hours. Therefore, it can be concluded that differences in soaking time could result in lattice destruction as some crystals were exposed to fragments longer. Fragments concentrations and soaking time could be 2 variables that can be tested in order to identify any dependence on crystal lattice quality. Although the highest resolution was 1.7 A which was a workable data where lots of details in protein – ligand binding could be observed.

7.4 Biophysical Analysis of TbrPDEB1 using XChem platform and SPR screen, Low concentration soaks

For the detailed ligand soak study TbrPDEB1 catalytic domain was chosen again and the source of PDE fragments was a subset of the ChEMBL library. Manual crystallization was initially used to help optimize and identify a general protocol that could be applied to other protein targets in

order to maximize the selection of the highest affinity fragments in the output of XChem.

31 PDE active fragments were selected from ChEMBL data. Each fragment was soaked individually with TbrPDEB1 catalytic domain where 5 different soak conditions were used: 40mM/24 hours, 40mM/48 hours, 30mM/24 hours, 50mM/72 hours and 50mM/96 hours. The longer high concentration soaking conditions was soaking for 72 hours and 96 hours with 50mM ligand concentration, clearly had a deleterious effect on the crystal quality. However, at 96 hours of soaking of 50mM there were still some fragments soaks that showed diffraction to high resolution (1.5 Å).

It is clear from the soaking experiments that not only DMSO has an effect on the crystal lattice but the different small fragment compounds themselves have differential effects on crystal packing.

Based on the examination of data that was obtained after manual data collection it was decided to use 5 fragments for more detailed multiple concentration soaks in XChem analysis to try and identify the concentration at which the high affinity fragments could be detected by creating a dose response curve.

From the manual soaks Ligand ChEMBL31877gave the highest resolution for 48 hours of soaking with ligand concentration 40mM, higher concentration couldn't be tolerated by the crystal lattice. Ligand ChEMBL31877 showed 2 different poses in manual soaks therefore that ligand was considered as an interesting target.

Ligand ChEMBL131164 had lower resolution of 1.99 being the best one among 3 soaks of that ligand where the most preferable soaking condition was 40mM ligand for 24 hours. When the crystal was visualized under the

microscope cracks were observed. Also, it could be seen that with the increase of ligand concentration and soak time the resolution of the diffraction was reducing.

Ligand ChEMBL131181 was also only observed at concentrations of 40mM and 30mM for 24 hours and 48 hours suggesting that longer time of soaking and higher concentration led to crystal lattice disruption. The Ligand – protein binding analysis for ligand ChEMBL131181 revealed that there were different ligand conformations in a binding site with high ligand occupancy suggesting that it may be a strong binder.

Ligand ChEMBL1779264 was only observed at 40mM concentration for 48 hours concentration with the resolution being 1.4. Refinement showed that it was a good binder. The last ligand was 22 where the only diffraction data was obtained at 40mM soaks for 48 hours with the resolution being 1.79. It was interesting to see if these last 2 ligands that gave diffraction data only at 1 condition will show any differences in XChem and if PanDDA will be able to detect the differences in ligand occupancy.

A dose – dependence curve could help determine a more generic optimized protocol where most of the library would be used and hence as much as possible diffraction data would be obtained. Perhaps indicating that XChem should be run at both low and high concentrations fragment soaks.

There were 2 soaking times 2 hours and 4 hours with the concentration gradients being used: 0.1 mM, 0.5 mM, 1mM, 5mM, 10mM, 20mM, 50mM, 100mM, and 200mM. According to the data there was no direction correlation between crystals lattice quality and correlation as some crystals failed to be mounted even at lowest concentration such as 1mM and were fine at 200mM. Clearly there must have been variation in the original

crystal or there was variation in the local concentration or diffusion rates with the small crystallization drops used in the XChem process. Therefore, we couldn't determine the direct dependency upon soaks concentration.

When the diffraction data for all 5 ligands was analyzed using XChem explorer software there was no obvious differences in the maps and ligand occupancy between the lowest and the highest concentrations.

Comparison of the chemical structures for each of the 5 fragments showed that they all are similar to IBMX, a known general PDE inhibitor including LmjPDEB1.

One issue was that PanDDA has failed at first place to identify the reference map that was the same as it was used in the first run of XChem

Unfortunately, due to PanDDA issues, it was not possible to plot dose-dependence curve.

Ligands ChEMBL131181, ChEMBL31877 and ChEMBL131164 were analyzed and were screened on SPR. SPR run revealed that 2 out of 3 showed binding either between human PDE4 or/and TbrPDEB1. As in previous screen with DSI-Poised library fragments human PDE4 was used as a control. Ligand ChEMBL31877 showed specific binding with human PDE4 and unspecific with TbrPDEB1. Ligand ChEMBL131164 showed binding with TbrPDEB1 which was unspecific, and trays were looking similar to Ligand ChEMBL31877 where human PDE4 didn't show any binding mode with Ligand ChEMBL31877. Unfortunately, there were no resources available with the collaborator to follow up on the hybrid compound design based on these data.

Chapter VIII

References

1. World Health Organization. Research priorities for Chagas 1316 disease, human African trypanosomiasis and leishmaniasis. WHO 1317 Technical Report Series 2012, 975:1−100.

2. Jaskowska E., Butler C., Preston G., Kelly S. Phytomonas: Trypanosomatids adapted to plant environments. PLOS Pathogens 2015; 11 (1):1-17.

3. Walton B. C., Valverde L. Racial differences in espundia. Ann. Trop. Med. Parasitol. 1979; 73 (1): 23-9.

4. Stark C. G. Leishmaniasis Crisis Presentation. Medscape. 2019.

5. Apt W. Current and developing therapeutic agents in the treatment of Chagas disease. Drug Des. Devel. Ther. 2010; 4:243–253.

6. Rogers M. E., Chance M. L., Bates P. A. The role of promastigote secretory gel in the origin and transmission of the infective stage of Leishmania mexicana by the sand-fly Lutzomyia longipalpis. Parasitology. 2002; 124:495–507.

7. Gossage S. M., Rogers M. E., Bates P. A. Two separate growth phases during the development of Leishmania in sand flies: implications for understanding the life cycle. Int. J. Parasitol. 2003; 33:1027–1034.

8. Sacks D. L., Perkins P. V. Development of infective stage Leishmania promastigotes within phlebotomine sand flies. Am. J. Trop. Med. Hyg. 1985; 34:456–459.

9. Hailu A., Mudawi Musa A., Royce C., Wasunna M. Visceral Leishmaniasis: new health tools are needed. PLOS Medicine. 2005; 2(7): e211.

10. Lescure FX, Le Loup G, Freilij H, et al. Chagas disease: changes in knowledge and management. Lancet Infect. Dis. 2010; 10(8):556–570.

11. MacLean L., Reiber H., Kennedy P., M. Stenberg J. Stage progression and neurological symptoms in Trypanosoma brucei rhodesience sleeping sickness: role of the CNS inflammatory response. PLOS Neglected Tropical Diseases 2012; 6(10): e1857.

12. Matthews K. R., Ellis J. R., Paterou A. Molecular regulation of the life cycle of African trypanosomes. Trends Parasitol. 2004; 20:40–47.

13. Roditi I., Liniger M. Dressed for success: the surface coats of insect-borne protozoan parasites. Trends Microbiol. 2002; 10:128–134.

14. Matthews K. R. Developments in differentiation of *T. brucei*. Parasitol. Today. 1999; 15:76–80.

15. Vassella E., Reuner B., Yutzy B., Boshart M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J. Cell Sci. 1997; 110:2661–2671.

16. Shalaby, T. et al. (2001) The regulatory subunit of a cGMP-regulated protein kinase A of Trypanosoma brucei. Eur. J. Biochem. 268, 6197–6206.

17. Gould M.K., de Koning, H.P. Cyclic-nucleotide signaling in protozoa. FEMS Microbiol. Rev. 2001; 35: 515–541.

18. Oberholzer M. et al. The Trypanosoma brucei cAMP phosphodiesterases TbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. FASEB J. 2007; 21:720–731.

19. D'Angelo M.A., et al. A novel calcium-stimulated adenylyl cyclase from Trypanosoma cruzi, which interacts with the structural flagellar protein paraflagellar rod. J. Biol. Chem. 2002; 277: 35025–35034.

20. Paindavoine, P. et al. A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases

located on the flagellum of Trypanosoma brucei. Mol. Cell. Biol. 1992; 12: 1218–1225.

21. Huang, H. Signal transduction in Trypanosoma cruzi. Adv. Parasitol. 2001; 75, 325–344.

22. Yamaguchi, T. et al. Sorafenib inhibits cAMP-dependent ERK activation, cell proliferation, and in vitro cyst growth of human ADPKD cyst epithelial cells. Am. J. Physiol. Renal Physiol. 2010; F944–F951.

23. Awuoche, E.O. Tsetse fly saliva: could it be useful in fly infection when feeding in chronically aparasitemic mammalian hosts. Open Vet. 2012; J.2,95.

24. Aandahl, E. M., Abi-Gerges, A., Beavo, J. A., et al. Cyclic Nucleotide Phosphodiesterases. In: Beavo, JA.; Francis, S.; Houslay, M., editors. Health and Disease. CRC Press; Boca Raton. 2007.

25. Beavo JA, Houslay MD, Francis SH. Cyclic nucleotide phosphodiesterase superfamily. In: Beavo JA, Francis S, Houslay M, editors. Cyclic Nucleotide Phosphodiesterases in Health and Disease. CRC Press. 2007; Chapter 1.

26. Weiss, B., Winchurch, R.A. Analyses of cyclic nucleotide phosphodiesterases in lymphocytes from normal and aged leukemic mice. Cancer Res. 1978; 38:1274-1280.

27. Wentzinger, L.; Seebeck, T. Protozoal phosphodiesterases. In: Beavo, JA.; Francis, S.; Houslay, M., editors. Cyclic Nucleotide Phosphodiesterases in Health and Disease. CRC Press. 2007. Chapter 14.

28. Luginbuehl E., Ryter D., Schranz-Zumkehr J., Oberholzer M., Kunz S., Seebeck T. The N-terminus of phosphodiesterase TbrPDEB1 of *Trypanosoma brucei* contains the signal for integration into the flagellar skeleton. Eukaryotic Cell. 2010; 9:1466–1475.

29. D'Angelo M. A., Sanguineti S., Reece J. M., Birnbaumer L., Torres H. N., Flawiá M. M. Identification, characterization and subcellular localization of TcPDE1, a novel cAMP-specific phosphodiesterase from *Trypanosoma cruzi*. Biochem. J. 2004; 378(1):63–72.

30. Rascón A., Soderling S. H., Schaefer J. B., Beavo J. A. Cloning and characterization of a cAMP-specific phosphodiesterase (TbPDE2B) from *Trypanosoma brucei*. Proc. Natl Acad. Sci. USA. 2002; 99(7):4714–4719.

31. Zoraghi R., Seebeck T. The cAMP-specific phosphodiesterase TbPDE2C is an essential enzyme in bloodstream form *Trypanosoma brucei*. Proc. Natl Acad. Sci. USA. 2002; 99:4343–4348.

32. Kunz S., Oberholzer M., Seebeck T. A FYVE-containing unusual cyclic nucleotide phosphodiesterase from *Trypanosoma brucei*. FEBS J. 2005; 272:6412–6422.

33. Schoijet A. C., Miranda K., Medeiros L. C. Defining the role of a FYVE domain in the localization and activity of a cAMP phosphodiesterase implicated in osmoregulation in *Trypanosoma cruzi*. Mol. Microbiol. 2010; 79(1):50–62.

34. Broadhead, R. et al. Flagellar motility is required for the viability of the bloodstream trypanosome. Nature. 2006; 440: 224–227.

35. Docampo, R. et al. The role of acidocalcisomes in the stress response of Trypanosoma cruzi. Adv. Parasitol. 2011; 75: 307–324.

36. Zoraghi R., Seebeck T. The cAMP-specific phosphodiesterase TbPDE2C is an essential enzyme in bloodstream form *Trypanosoma brucei*. Proc. Natl Acad. Sci. USA. 2002; 99:4343–4348.

37. Oberholzer M., Marti G., Baresic M., Kunz S., Hemphill A., Seebeck T. The *Trypanosoma brucei* cAMP phosphodiesterasesTbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. FASEB J. 2007; 21:720–731.

38. Ciulla M. M., Vivona P. PDE5 Inhibitors, Erectile Dysfunction and beyond: How, Sometimes, Indications are the Consequences of Marketing Strategies and/or Serendipity. SciMedCentral. 2017; 2(1): 1005.

39. Galie N., Ghofrani H. A., Torbicki A., Barst R. J., Rubin L. J., Badesch D., et al. Sildenafil use in pulmonary arterial hypertension (SUPER)study group. Sildenafil citrate therapy for pulmonary arterial hypertension. N. Engl. J. Med. 2005; 353.

40. A. Huang S., D. Lie J. Phosphodiesterase – 5 (PDE5) inhibitors in the management of erectile dysfunction. P T. 2013; 38(7): 407, 414-419.

41. Wang H, Yan Z, Geng J, Kunz S, Seebeck T, Ke H. Crystal structure of the Leishmania major phosphodiesterase LmjPDEB1 and insight into the design of the parasite-selective inhibitors. Mol. Microbiol. 2007; 66(4):1029–1038.

42. Huai Q., Liu Y., Francis S. H., Corbin J. D., Ke H. Crystal structures of phosphodiesterases 4 and 5 in complex with inhibitor IBMX suggest a conformation determinant of inhibitor selectivity. J. Biol. Chem. 2004; 279(13):13095–13101.

43. Wang H., Liu Y., Huai Q., et al. Multiple conformations of phosphodiesterase-5: implications for enzyme function and drug development. J. Biol. Chem. 2006; 281(30):21469–21479.

44. Ke H., Wang H. Crystal structures of phosphodiesterases and implications on substrate specificity and inhibitor selectivity. Curr. Top. Med. Chem. 2007; 7(4):391–403.

45. Oberholzer M., Marti G., Baresic M., Kunz S., Hemphill A., Seebeck T. The *Trypanosoma brucei* cAMP phosphodiesterasesTbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. FASEB J. 2007; 21:720–731.

46. Jansen C., Wang H., Kooistra A. J., de Graaf C., Orrling K. M., Tenor H., Seebeck T., Bailey D., de Esch I. J., Ke H., Leurs R. Discovery of novel Trypanosoma brucei phosphodiesterase B1 inhibitors by virtual screening against the unliganded TbrPDEB1 crystal structure. J. Med. Chem. 2013; 56(5): 2087-2096.

47. de Koning, H.P. et al. Pharmacological validation of Trypanosoma brucei phosphodiesterases as novel drug targets. J. Infect. Dis. 2012; 206, 229–237.

48. Mohamed R., Degac J., Helms V. Composition of overlapping protein-protein and protein-ligand interfaces. PLoS One. 2015; 10(10): 1–18.

49. de Freitas R. F., Schapira M. A systematic analysis of aromatic protein-ligand interactions in the PDB. Medchemcomm. 2017; 8(10): 1970-1981.

50. Ritchie T. J., Macdonald S. J. F. Physiochemical descriptors of aromatic character and their use in drug discovery. J. Med. Chem. 2014; 57(17): 7206–7215.

51. Taylor D., MacCoss M., Lawson A. D. G. Rings in Drugs. J. Med. Chem. 2014; 57(14): 5845–5859.

52. Valley C. C., Cembran A., Perlmutter J. D., Lewis A. K., Labello N. P.

, Gao J., Sachs J. N. J. Biol. Chem., 2012, 287,   34979–34991.

53. Davis A. M., Teague S. J., Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. Angew. Chem., Int. Ed. 1999; 38(6): 736–749.

54. Leung C. S., Leung S. S., Tirado-Rives J. and Jorgensen W. L. Methyl effects on protein-ligand binding. J. Med. Chem. 2012; 55(9): 4489–4500.

55. Shoichet B. K. Prediction of a protein's affinity for a ligand has been improved through fundamental physical modeling. Nat. Biotechnol. 2007; 25: 1109–1110.

56. Muley L., Baum B., Smolinski M., Freindorf M., Heine A., Klebe G., Hangauer D. G. Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: synthesis, modelling, and molecular dynamics simulations of a congeneric series of thrombin inhibitors. Med Chem. 2010; 53(5): 2126-35.

57. Lovering F., Bikker J., Humblet C. Escape from flatland: increasing saturation as an approach to improving clinical success. J. Med. Chem. 2009; 52 (21): 6752-6756.

58. Horowitz S., Trievel R. C. Carbon – oxygen hydrogen bonding in biological structure and function. J. Bio. Chem. 2012; 287(50):41576-82.

59. Musah R. A., Jemsen G. M., Rosenfeld R. J., McRee D. E., Goodin D. B., Bunte S. W. Variation in strength of an unconventional C – H to O Hydrogen bond in an engineered protein cavity. J. Am. Chem. Soc. 1997; 119(38): 9083-9084.

60. Pierce A. C., Sandretto K. L., Bernis G. W. Kinase inhibitors and the case for CH…O hydrogen bonds in protein-ligand binding. Proteins 2002; 49(4): 567-76.

61. Nagy P. I., Erhardt P. W. Theoretical studies of salt-bridge formation by amino acid side chains in low and medium polarity environments. J. Phys. Chem. B. 2010; 114(49): 16436-42.

62. Duan G., Smith V. H., Weaver D. F. An ab initio and data mining study on aromatic amide interactions. Chemical Physics Letter. 1999; 310(3-4): 323-332.

63. Mahadevi A. S., Sastry G. N. Cation - π interaction: its role and relevance in chemistry, biology, and material science. 2013; 113(3): 2100-38.

64. Dougherty D. A. The cation -π interaction. Acc. Chem. Res. 2013; 46(4): 885-93.

65. Domagata M., Matczak P., Palusiak M. Halogen bond, hydrogen bond and N-C interaction – On interrelation among these three noncovalent interactions. Science Direct. 2012; 998: 26-33.

66. Clark T., Hennemann M., Murray J. S., Politzer P. Halogen bonding: the sigma-hole. Proceedings of "modelling intercations in biomolecules II". J. Mol. Model. 2007; 13(2): 291-6.

67. Politzer P., Murray J. S., Clark T. Halogen bonding and other σ-hole interactions: a perspective. Phys. Chem. Chem. Phys. 2013; 15(27): 11145-11588.

68. Zhang Q., Xu Z., Shi J., Zhu W. Underestimated halogen bonds forming with protein backbone in protein data bank. J. Chem. Model. 2017; 57(7): 1529-1534.

69. Hardegger L. A., Kuhn B., Spinnler B., Anselm L., Ecabert R., Stihle M., Gsell B., Thoma R., Diez J., Benz J., Plancher J. M. Hartmann G., Banner D. W., Haap W., Diederich F. Systematic investigation of halogen bonding in protein-ligand interactions. Angew. Chem. Int. Ed. Engl. 2011; 50(1): 314-8.

70. Xu Z., Yang Z., Liu Y., Lu Y, Chen K., Zhu W. Halogen bond: its role beyond drug-target binding affinity for drug discovery and development. J. Chem. Inf. Model. 2014; 54(1): 69-78.

71. Lu Y., Shi T., Wang Y., Yang H., Yan X., Luo X., Jiang H., Zhu W. Halogen bonding a novel interaction for rational drug design. J. Med. Chem. 2009. 52 (9): 2854-2862.

72. Veerman, J., van den Bergh T., Orrling, K. M., Jansen, C., Cos, P., Maes, L., Chatelain, E., Ioset, J. R., Edink, E. E., Tenor, H., Seebeck, T., de Esch, I. J. P., Leurs, R., Sterk, G. J. Synthesis and evaluation of analogues of the phenylpyridazinone NPD-001 as potent trypanosomal TbrPDEB1 phosphodiesterase inhibitors and in vitro trypanocidals. Bioorg. Med. Chem. 2016, 24, 1573−1581.

73. Blaazer A. R., Singh A. K., de Heuvel E., Edink E., Orrling K. M., Veerman J. J. N., van der Bergh T., Jansen C., Balasubramaniam E., et al. Targeting a subpocket in Trypanosoma brucei Phosphodiesterase B1 (TbrPDEB1) Enables the Structure – Based Discovery of Selective Inhibitors with Trypanocidal Activity. J Med Chem. 2017; 61(9): 3870-3888.

74. Jeong J., Yim H., Ryu J., Lee H. S., Lee J., Seen D., Kang S. G. One-step sequence and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Amer. Soc. for Micr. 2012; 78 (15): 5440 – 5443.

75. Porath J. Immobilized metal ion affinity chromatography. Protein Expr Purif. 1992; 3(4): 2630281.

76. Crowe J., Masone B. S., Ribbe J. One-step purification of recombinant proteins with the 6xHis tag and Ni-NTA resin. Mol Biotechn. 1995; 4(3): 247-258.

77. DePalma A. Keeping Tabs on Polyhistidine Tags. GEN Tech. Trends in Biotech. 2016; 36(5).

78. Janknecht R., de Martynoff G., Lou J., Hipskind R. A., Nordheim A., Stunnenberg H. G. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc. Natl. Acad. Sci. 1991; 88(20): 8972-8976.

79. Wingfield P. T. Overview of the Purification of Recombinant Proteins. Curr. Protoc. Protein Sci. 2015; 80: 1-50.

80. Bornhorst J. A., Falke J. Reprint of: Purification of proteins using polyhistidine affinity tags. Meth. in Enz. 2000; 326:245-254.

81. Synge R. L. M., Tiselius A. Fractionation of Hydrolysis Products of Amylose by Electrokinetic Ultrafiltration in an Agar-Agar Jelly. Biochemical Journal. 1950; 24(2):41-42.

82. O'Callaghan D. M., Donnelly W. J., Slattery H. M., Mulvihill D. M. Non-Size Exclusion Effects During Gel Permeation Chromatography of Milk Protein Hydrolysates on an Fplc Superose 12 Column. Journal of Liquid Chromatography. 1995;18(8):1543–1562.

83. Yu C.-M., Mun S., Wang N.-H.L. Phenomena of Insulin Peak Fronting in Size Exclusion Chromatography and Strategies to Reduce Fronting. Journal of Chromatography A. 2008; 1192(1):121–129.

84. Cai T., Zgu L., Yu L. Crystallization of organic glasses: effects of polymer additives on bulk and surface crystal growth in amorphous nifedipine. Pharm Res. 2011; 28(10): 2458-2466.

85. Chayen N. E., Saridakis E. Protein crystallization: from purified protein to diffraction – quality crystal. Nat. Methods. 2008; 5(2): 147-153.

86. Li L., Ismagilov R. F. Protein crystallization using microfluidic technologies based on valves, droplets, and SlipChip. Ann. Rev. Biophys. 2010; 39: 139-158.

87. Garcia-Ruiz J. M. Counter diffusion methods for macromolecular crystallization. Methods Enzymol. 2003; 368: 130-154.

88. De Koning H. P., Gould M. K., Sterk G. J., Tenor H., Kunz S., Luginbuehl E., Seebeck T. Pharmacological validation of Trypanosoma brucei phosphodiesterases as novel drug targets. J. Infect. Dis.2012; 206: 229–237.

89. Trevino S. R., Scholtz J. M, Pace C.N. Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favourably than the other hydrophilic amino acids in RNase Sa. J. Mol. Biol. 2007; 366(2):449-460.

90. Boël G., Letso R., Neely H., Price W. N., Wong K. H., Su M., Luff J., Valecha M., Everett J. K., et al. Codon influence on protein expression in E.coli correlates with mRNA levels. Nature. 2016; 529(7586): 358-363.

91. Luis, B.G. Cloning and initial characterization of a family A DNA polymerase from Entamoeba histolytica: A putative mitochondrial DNA Polymerase. FASEB J. 2007; 21: A1039.

92. Mamipour M., Yousefi M., Hasanzadeh M. An overview on molecular chaperones enhancing solubility of expressed recombinant proteins with correct folding. Int. J. Biol. Macromol. 2007; 102: 367-375.

93. Bondos S., Bicknell A. Detection and prevention of protein aggregation before, during, and after purification. Analytical Biochemistry. 2003; 316: 223-231.

94. Jansen C., Wang H., J. Kooistra A., de Graaf C., Orrling K., Tenor H., Seeback T., Bailey D., J. P. de Esch I., Ke H., Leurs R. Discovery of novel Trypanosoma brucei phosphodiesterase B1 inhibitors by virtual screening

against the unliganded TbrPDEB1 crystal structure. J. Med. Chem. 2013; 56(5): 2087-2096.

95. Webb M. The influence of Magnesium on cell division. 1949; 418-424.

96. Bland N. D., Wang C., Tallman C., Gustafson A. E., Wang Z., Ashton T. D., Ochiana S. O., McAllister G., Cotter K., Fang A. P., Gechijian L., Garceau N., Gangurde R., Ortenberg R., Ondrechen M. J., Campbell R. K., Pollastri M. P. Pharmacological Validation of *Trypanosoma brucei* Phosphodiesterases B1 and B2 as Druggable Targets for African Sleeping Sickness. J. Med. Chem. 2011; 54:8188–8194.

97. Hamilton M. A., Russo R. C., Thurston R. V. Trimmed Spearman – Karber method for estimating median lethal concentrations in toxicity bioassays. Environmental Science and Technology. 1977; 11(7): 714 – 719.

98. Vulpetti A., Hommel U, Landrum G.m Lewis R., Dalvit C. Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of Fluorine. J. Am. Chem. Soc. 2009; 131(36): 12949-12959.

99. Alberti S., Bohse K., Arndt V. The Co-chaperone HspBPl Inhibits the CHIP ubiquitin ligase and stimulates the maturation of the cystic fibrosis transmembrane conductance regulator. Mol. BioI. Cell 2004; 15(9):4003-4010.

101. Cuervo A. M., Stefanis L., Fredenburg R. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. Science 2004; 305(5688) :1292-1295.

102. Cuervo A. M. Autophagy: In sickness and in health. Trends Cell BioI 2004; 14(2):70-77.

103. Oliveira I. P., Lescano C. H. In Silico Mapping of Essential Residues in the Catalytic Domain of PDE5 Responsible for Stabilization of its Commercial Inhibitors. Sci. Pharm. 2019, 87(4), 29.

Appendix 1

Table 1 Data collection and refinement statistics for TbrB1 catalytic domain crystals

	TbrB1 - 0074 (*1*)	TbrB1 - 0209 (*2*)	TbrB1 - 0218 (*4*)	TbrB1 - 0268 (*5*)	TbrB1 - 0269 (*6*)	TbrB1 - 0248 (*7*)
Data collection						
Space group	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1
Molecule/a.s.u	2	2	2	2	2	2
Cell dimensions□□						
a, b, c (Å)	113.79, 115.32, 68.92	114.030, 114.890, 68.390	114.597, 114.858, 68.534	113.339, 115.066, 69.011	115, 114.75, 68.35	115.12, 114.61, 68.27
α, β, γ (°)	90, 108.26, 90	90, 108.51, 90	90, 108.34, 90	90, 108.113, 90	90, 108.16, 90	90, 108.15, 90
Resolution (Å)	78.852-2.396 (2.51-2.42)	78.741-1.891 (1.95-1.91)	65.05-3.09 (1.92-1.88)	49.93-2.21 (1.67-1.64)	79.13-1.84 (1. 70-1.67)	79.13-2.25 (2.25-2.18)
CC 1/2	99.6 (0.56)	99.6 (41.8)	99.5 (53.7)	99.5 (35.9|)	99.7 (25.9|)	97.3 (56.8)
$I / \sigma I$	8.3 (1.2)	8.3 (1.0)	5.2 (1.3)	8.2 (1.6)	7.7 (0.3)	6.5 (1.2)
Completeness (%)	99.7 (99.9)	99.4 (99.1)	97.7 (87)	99.1 (93.9)	99.8 (98.3)	99.4 (99.8)
Refinement						
Resolution (Å)	2.42	1.91	1.88	1.64	1.67	2.18
No. reflections	32164 (1615)	64213 (3157)	66743 (3303)	102052 (5025)	96028 (4780)	43620 (2165)
R_{work} / R_{free}	0.2/0.263	0.202/0.252	0.222/0.268	0.202/0.238	0.201/0.238	0.194/0.251
No. atoms						
Protein	5260	5260	5260	5260	5254	5260
Ligand/ion	78/4	72/4	50/4	39/4	49/4	52/4
Water	286	312	386	415	445	403
B-factors						
Protein	63.91	41.09	37.42	39.92	28.21	37.39
Ligand/ion	88.19/43.12	56.4/27.45	56.41/21	59.19/26.22	47.16/15.81	56.97/25.86
Water	53.14	41.65	39.81	44.84	34.83	36.31
R.m.s. deviations						
Bond lengths (Å)	0.0065	0.0157	0.0095	0.0147	0.0149	0.0143
Bond angles (°)	1.413	1.94	1.639	1.78	1.82	1.84

Data were collected from one crystal in each case. All structures will be deposited with the PDB on acceptance of the thesis