DOCTOR OF PHILOSOPHY

Structural examination of trypanosomatid tubulin-binding cofactors and pteridine reductase 1 inhibition

Barrack, Keri L.

Award date:
2013

Awarding institution:
University of Dundee

DOCTOR OF PHILOSOPHY

Structural examination of trypanosomatid tubulin-binding cofactors and pteridine reductase 1 inhibition

Keri L. Barrack

2013

University of Dundee

A thesis submitted for the degree of Doctor of Philosophy

Structural examination of trypanosomatid tubulin-binding cofactors and pteridine reductase 1 inhibition

Keri L. Barrack

Supervisor:

Professor William N. Hunter

College of Life Sciences

University of Dundee

December 2013

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are a number of people without whom this thesis would not have been possible. Firstly, and most importantly, I would like to thank Bill Hunter. In appointing me as a research technician fresh from my undergraduate studies in a rather different subject, he took an initial risk and has kept faith in me throughout. His supervision has been instrumental to my progress and will always be remembered.

Thanks are next due to Paul Fyfe. Not only did he help immensely with all protein purification and crystallography related issues but has been an extremely patient office-mate and friend. Our regular discussions have always been valuable, whether work or life related.

Many more members of the Hunter group, past and present, are also owed thanks. Alice Dawson, Sharon Shepherd, Rachel Morgan, Thomas Eadsforth, Scott Cameron, Jen Fleming and Lindsay Tulloch are only a few of the names I need to mention (those I have missed, I thank you too!). I am grateful to everyone who introduced me to a new technique, supported my constant learning or made my time more enjoyable. Collaborators in Strathclyde, Glasgow and here in Dundee and funding agencies likewise receive my thanks. Neighbours in AHF and MAJF groups and support staff have also played their silent roles in these studies for which I greatly appreciate.

Finally, I am extremely grateful to my family and friends outside the world of science. Their love and support, whether they know it or not, has been necessary to keep me on track. Thank you.

DECLARATION

I hereby declare that the following thesis is based on the results of investigations conducted by myself, and that this thesis is of my own composition. This thesis has not, in whole or part, been previously presented for a higher degree. Work other than my own is clearly indicated in the text by reference to the relevant researchers or publications. I would like to thank those people who allowed me to refer to their unpublished observations and data, which are individually acknowledged where they occur.

Keri L. Barrack

The work presented in this thesis is the work of the candidate Keri L. Barrack. Conditions of the relevant Ordnance and Regulations have been fulfilled.

Professor William N. Hunter

SUMMARY

Trypanosomatid parasites are the causative agents of neglected tropical diseases for which current therapies are inadequate. As primitive eukaryotic organisms, they also represent a useful model system to investigate fundamental cellular biology while studies of potential drug targets endeavour to develop new drug molecules. Aspects of both of these areas are explored in this thesis.

Microtubules are polymers of tubulin and are essential in eukaryotes for cell division, motility and maintenance of cell morphology. Five tubulin-binding cofactors (TBC, named A-E) are proteins implicated in the folding, polymerisation and processing of tubulin, the major component of the trypanosomatid cytoskeleton. At the initiation of this study, there was no structural information available for any trypanosomatid TBC. We therefore sought to investigate these proteins by X-ray crystallography and assess their potential tubulin-interaction capabilities to support the current functional model. The crystal structure of tubulin-binding cofactor A (TBCA) from *Leishmania major* is presented, determined using diffraction data to 1.9 Å resolution. Prior to tubulin polymerisation, TBCA forms a complex with β-tubulin in the pathway to αβ-tubulin heterodimerisation. It maintains a soluble pool of β-tubulin and can prevent premature polymerisation. This is a short helical protein, similar in structure to published homologues. The similarities and some distinct local features that may impact on β-tubulin binding are discussed. In particular, the surface properties of a prominent bend in the helix bundle represents an area that may be capable of interacting with its tubulin partner.

Tubulin-binding cofactor C (TBCC) is implicated in stimulating the hydrolysis of GTP bound to β-tubulin prior to release of the assembly-competent αβ-tubulin heterodimer from a supercomplex between TBCC, TBCD, TBCE and both tubulin subunits. Full-length recombinant *Trypanosoma brucei* and *Leishmania major* tubulin-binding TBCC were degraded and crystallisation could not be achieved. However, crystals of a truncated TBCC construct were obtained. Despite efforts to optimise crystallisation and diffraction data, the structure was not solved for inclusion in this thesis. Instead, homologous structures were analysed and a potential tubulin interaction site is suggested based on the proposed GTPase-stimulating activity of TBCC and the similarity with the human protein, Retinitis Pigmentosa 2 (RP2), predicted to contain a domain with similar fold. Progress towards the soluble recombinant expression of the other cofactors also lays the foundation for future investigations into trypanosomatid TBC structure and function.

Pteridine reductase 1 (PTR1), an enzyme unique to trypanosomatids, is the subject of Part II of this thesis. PTR1 is a broad-spectrum NADPH-dependent reductase, catalysing the two-stage reduction of biopterin to dihydrobiopterin and tetrahydrobiopterin and that of folate to dihydrofolate and tetrahydrofolate. As such, it can provide a bypass mechanism for the reduction of folates, reducing the therapeutic action of traditional antifolate molecules in these organisms. Inhibition of PTR1 is therefore desirable from a drug discovery viewpoint. The crystal structure of *Leishmania donovani* PTR1 was determined using data extending to 2.5 Å resolution with a view to generating ligand-complex structures and providing a model for inhibitor design. This structure was found to contain a disordered active site, with several loop regions not modelled or relocated. A sulfate molecule from the crystallisation mixture

binds in the cofactor phosphate binding-site and the sequential binding of cofactor before substrate or inhibitor can not occur. Although this crystal form was considered unsuitable for further studies, it provides the only structure of PTR1 in the absence of cofactor.

With an established crystallisation protocol, *Trypanosoma brucei* PTR1 then forms the basis of a collaborative investigation of over 100 novel potential inhibitory molecules. Kinetic evaluation, isothermal titration calorimetry (ITC) and co-crystallisation were applied to generate ligand-binding profiles of pyrrolopyrimidine derivatives. Several interesting binding features were identified from the 24 ligand complex structures obtained. These include the discovery of two covalent inhibitors, confirming the reactivity of a non-conserved active site cysteine, and molecules that are able to bind simultaneously at two locations within the active site pocket, exploiting hydrogen-bonding interactions with key catalytic and other nearby residues. The thermodynamic binding profiles of seven inhibitors also provide insight into the enthalpic and entropic contributions to ligand binding. We assessed the suitability of ITC for this system and while a high attrition rate was observed, chemical substitutions were able to enhance the binding entropy. These studies have strengthened our understanding of the structure-activity relationship between PTR1 and inhibitors, offering opportunities to develop new molecules that focus on increasing the potency generated by favourable enthalpy alongside improving the drug-like properties.

LIST OF ABBREVIATIONS

Standard SI abbreviations are used throughout except Ångstrom (Å), representing 10^{-10} m and calorie (cal) representing 4.184 J. Amino acids are abbreviated using the standard three letter code or single letter identifier and atoms abbreviated to their elemental symbol.

ADMET	absorption, distribution, metabolism, excretion, toxicity
ADP	adenosine diphosphate
AI	autoinduction
$AmSO_4$	ammonium sulfate
Arl	adenosine diphosphate ribosylation factor-like
ASU	asymmetric unit
At	*Arabidopsis thaliana*
CAP-Gly	cytoskeleton-associated protein-glycine-rich
CCD	charged couple device
CD	circular dichroism
CCP4	collaborative computational project number 4
CCT	chaperonin containing T-comlex polypeptide (TCP)-1
cLogP	calculated octanol/water partition coefficient
***C*-terminus**	carboxyl-terminus
DHB	dihydrobiopterin
DHFR	dihydrofolate reductase
DHFR-TS	dihydrofolate reductase-thymidylate synthase
DLS	Diamond Light Source
DMPK	drug metabolism and pharmacokinetics
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNaseI	deoxyribonuclease I
dNTP	deoxyribonucleotide triphosphate
DPI	diffraction precision index
dTMP	2'-deoxythymidine-5'-monophosphate
dUMP	2'-deoxyuridine-5'-monophosphate
DSF	differential scanning fluorimetry
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediamine tetraacetic acid
ESRF	European Synchrotron Radiation Facility
FDA	Food and Drug Administration

GAP	guanosine triphosphatase-activating protein
gDNA	genomic deoxyribonucleic acid
GDP	guanosine diphosphate
GeneDB	Gene Data Bank
GST	glutathione-*S*-transferase
GTP	guanosine-5'-triphosphate
GTPase	guanosine triphosphatase
HA	heavy atom
HAT	human African trypanosomiasis
HEAT	**H**untingtin, **e**longation factor 3, protein phosphatase 2**A**, yeast kinase **T**OR1
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HRP	horseradish peroxidase
Hs	*Homo sapiens*
IC_{50}	inhibitory concentration (50 %)
ID	identity
IPTG	isopropyl thio-β-D-galactoside
ITC	isothermal titration calorimetry
JCSG	Joint Centre for Structural Genomics
K_a	association constant
K_d	dissociation constant
K_i	inhibition constant
K_m	Michaelis-Menten constant
LB	lysogeny broth
Ld	*Leishmania donovani*
LE	ligand efficiency
Lm	*Leishmania major*
MAD	multiple-wavelength anomalous dispersion
MALDI-TOF	matrix-assisted laser desorption/ionisation time-of-flight
MAP	microtubule associated protein
MBP	maltose-binding protein
MES	2-(*N*-morpholino)ethanesulfonic acid
MIR	multiple isomorphous replacement
MIRAS	multiple isomorphous replacement with anomalous scattering
MR	molecular replacement
MRC	Medical Reseach Council
MS	mass spectrometry
MME	monomethyl ether
MIDAS	modern intelligent dynamic alternative screen
MPD	2-methyl-2,4-pentanediol
MTX	methotrexate
NADPH	nicotinamide adenine dinucleotide phosphate (oxidised)

$NADP^+$	nicotinamide adenine dinucleotide phosphate (reduced)
NMR	nuclear magnetic resonance
***N*-terminus**	amino-terminus
OD	optical density
***p*ABA**	*p*-aminobenzoyl
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PGA	poly-γ-glutamic acid
***p*I**	isoelectric point
PIPES	1,4-piperazinediethanesulfonic acid
PTR1	pteridine reductase 1
PVDF	polyvinylidene difluoride
RNA	ribonucleic acid
RP2	retinitis pigmentosa 2
r.m.s.d	root mean square deviation
SAD	single-wavelength anomalous dispersion
SAR	structure-activity relationship
Sc	*Saccharomyces cerevisiae*
SDG	Strathclyde Dundee Glasgow (compound series)
SDR	short-chain dehydrogenase-reductase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	selenomethionine
SIR	single isomorphous replacement
SIRAS	single isomorphous replacement with anomalous scattering
SSM	secondary structure matching
TAQ	2,4,6-triaminoquinazoline
Tb	*Trypanosoma brucei*
TBC	tubulin-binding cofactor
Tc	*Trypanosoma cruzi*
TEV	tobacco etch virus
T_m	melting temperature
TMQ	trimetrexate
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
WHO	World Health Organisation

PREFACE

This thesis contains research in two project areas. Both topics are based on proteins from species of trypanosomatid parasites. **Part I** describes work towards providing structural knowledge of tubulin-binding cofactor proteins involved in the formation and dynamics of microtubules. **Part II** provides structural, thermodynamic and kinetic details of a series of inhibitors of the enzyme, pteridine reductase 1. Although somewhat distinct, some experimental methodologies cross between both subjects and each part should therefore not be read in complete isolation. A brief introduction to trypanosomatid parasites and crystallographic methods are included in **Part I Chapter 1** alongside an overview of microtubule organisation and tubulin-binding cofactors. Pteridine reductase 1 is introduced in **Part II Chapter 4**.

Part of this work has been published, cited below, and additional manuscripts are in preparation.

Dawson, A., Tulloch, L.B., Barrack, K.L., Hunter, W.N., (2010). High-resolution structures of *Trypanosoma brucei* pteridine reductase ligand complexes inform on the placement of new molecular entities in the active site of a potential drug target. *Acta Crystallogr D Biol Crystallogr* **66**, 1334–1340.

Barrack, K.L., Tulloch, L.B., Burke, L.A., Fyfe, P.K., Hunter, W.N., (2011). Structure of recombinant *Leishmania donovani* pteridine reductase reveals a disordered active site. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 33–37.

PART I

Tubulin-binding cofactors

1. Introduction

1.1. Trypanosomatid parasites

Leishmania and *Trypanosoma* are digenetic genera of the Trypanosomatidae family from the order Kinetoplastida (Bush *et al*., 2001). These primitive eukaryotic protozoan parasites are responsible for human diseases that include Leishmaniasis and African sleeping sickness, also known as human African trypanosomiasis (HAT), and Chagas disease. The flagellated trypanosomatids are potentially valuable organisms to study processes involved in the folding of tubulin and the organisation of microtubules as these form the main component of the trypanosomatid cytoskeleton. In particular, *Leishmania major* and *Trypanosoma brucei* are well characterised biochemically and genetically and have been selected as models for the studies presented in this thesis.

1.1.1. Leishmaniasis

Leishmaniasis is transmitted by the bite of a female sandfly (Desjeux, 2004). Two forms of Leishmaniases include cutaneous and visceral infections, caused by *L. major* and *L. donovani*, respectively, which are just two of more than 20 species of *Leishmania* parasites (Desjeux, 1996). When an infected sandfly takes a blood meal from a human host, the infective promastigote form of the parasite are injected into the the host. Inside host macrophages, phagocytosed promastigotes then transform into amastigotes, multiplying in various host tissues. The insect stage of the life cycle continues when infected macrophages are ingested during further sandfly blood meals and the amastigotes transform into promastigotes and proliferate in the insect midgut before migration to the proboscis to be transmitted to a new host (Centres for Disease Control and Prevention, 2013).

Leishmaniasis presents a serious health risk in 88 countries (Desjeux, 1996) and it is estimated that approximately 12 million people are currently infected. The most serious form, visceral Leishmaniasis, has an incidence of approximately 500,000 new cases per year (Desjeux, 2001). Visceral Leishmaniasis is fatal if untreated as the infection, which presents as ulcerative lesions weeks or months following the infective insect bite, can disseminate through internal organs including the liver, spleen, bone marrow and distant lymph nodes (Murray *et al*., 2005). The more common cutaneous infection (approximately 1.5 million cases annually; Desjeux, 2001), where lesions remain localised to the skin, is less serious but can lead to disabling manifestations if untreated. Such estimates may be under-representations of the true epidemiology due to incomplete reporting of disease incidence and insufficient diagnostic capabilities (Desjeux, 2004).

1.1.2. Trypanosomiasis

The tsetse fly is the insect vector responsible for transmission of HAT, which is prevalent in 36 countries in sub-Saharan Africa. Two *T. brucei* parasite subspecies, *T. b. rhodesiense* and *T. b. gambiense*, are causative agents of acute and chronic infections in humans, respectively. Acute infections present within weeks of the initial injection of the trypomastigote form of the parasite into the host while symptoms of chronic infection can emerge several years later. The parasites enter the host bloodstream via the lymphatic system. The bloodstream trypomastigotes multiply and travel throughout the body where they can be injested by further tsetse flies taking a blood meal from an infected individual. Within the insect vector, the parasites undergo further transformation into procyclic trypomastigotes in the midgut of the fly, epimastigotes following departure from the midgut and finally in the salivary gland of the tsetse fly,

the cells proliferate and mature into metacyclic trypomastigotes, ready for injection into another mammalian host (Vickerman, 1985; McKean, 2003; Centres for Disease Control and Prevention, 2012). The chronic form of HAT involves two stages. The first stage presents with non-specific flu-like symptoms including fever, headache and joint pain. Cardiac and kidney disease can also develop, amongst other serious complications, if untreated. The second stage manifests when the parasite migrates across the blood-brain barrier, resulting in neurological symptoms such as confusion, depleted coordination, sensory disturbances and disruption to the normal sleeping cycle. Ultimately, progressive mental deterioration can then lead to coma and death.

With over 70 million people estimated to be at risk of contracting HAT (Simarro *et al*., 2012), the majority of reported cases are of the chronic form (Simarro *et al*., 2011) but both are fatal if left untreated (Fairlamb, 2003; Simarro *et al*., 2008). The subspecies, *T. b. brucei* (named from this point simply as *T. brucei*) is often used for laboratory studies as it presents a lesser risk of human infection (Gibson, 2012).

Chagas disease, sometimes known as American trypanosomiasis, is spread to humans via the infected faeces of a triatomine bug (often known as the 'kissing bug') through ingestion of contaminated foods, blood transfusions or from mother to foetus during pregnancy. Like HAT, Chagas disease also occurs with acute and chronic phases. Acute infection often results in swelling around the site of transmission and other mild symptoms including fever, headache and muscle pain, but can also lead to severe inflammation of heart or brain tissue. The chronic phase follows the acute phase with many people remaining asymptomatic. While parasitaemia may be low or undectectable during this stage, localisation of the parasites to cardiac, neurological or

gastrointestinal tissues can lead to the development of abnormalities in these organs, particularly serious in immunocompromised individuals (Prata, 2001; Nunes *et al*., 2013). It is estimated that approximately 8 million people in at least 21 countries are infected with the causative organism, *Trypanosoma cruzi* (WHO, 2013a). The disease presents a serious medical and socioeconomic burden to endemic regions of Latin America (Dias *et al*., 2002). While proteins from *T. cruzi* are not directly under study within the scope of this thesis, it is possible that information gathered may also be extended to this related parasite.

1.1.3. Current therapies

Examples of the treatment options available for Leishmaniasis include pentavelent antimonials, miltefosine, the first choice oral treatment, and amphotericin B, an agent that attacks the cell wall. Antimonial treatments are invasive and can result in toxic side-effects (Berman, 1997). Miltefosine has teratogenic properties and a poor pharmacokinetic profile, which may contribute to increased drug resistance while amphotericin B is also able to attack the host cell membrane, and has itself been the cause of fatalities (Reithinger *et al*., 2007). Drugs approved for the treatment of HAT include suramin, pentamidine, eflornithine, melarsoprol and nifurtimox (Fairlamb, 2003) but these are also prone to resistance or adverse side-effects. It is therefore clear that, with high cost, high toxicity, difficult administration and increasing drug resistance as major contributors to the poor overall efficacy (Fairlamb, 2003; Croft *et al*., 2006), current therapies for Leishmaniasis and HAT, as well as Chagas disease, are inadequate and new drugs are sought. The World Health Organisation considers these amongst the 17 neglected tropical, or ‘orphan’, diseases (WHO, 2013b; Fairlamb, 2003) and a greater understanding of trypanosomatid biology, with a particular focus on microtubule

biogenesis here, may aid in the discovery of improved treatments.

1.2. Microtubules

Microtubules are dynamic polymers of αβ-tubulin heterodimers that play an important role in many diverse eukaryotic cellular processes. The heterodimers are arranged in protofilaments in a head-to-tail manner with a plus end, where the β-tubulin subunit is exposed, and a minus end. Multiple individual polar protofilaments, usually thirteen, are organised to form a hollow microtubule (**Figure 1.1**). Each tubulin subunit binds a guanine nucleotide. For α-tubulin, guanosine-5'-triphosphate (GTP) is permanently bound and is inaccessible due to its location at the dimer interface while the nucleotide bound to β-tubulin can be hydrolysed to GDP and/or exchanged. Microtubule growth via the addition of dimer units typically occurs at the microtubule plus end (Margolis and Wilson, 1998; Howard and Hyman, 2009) and is affected by the status of the β-tubulin guanine nucleotide. Polymerisation takes place at the GTP-β-tubulin microtubule tip and subsequent hydrolysis ensures the microtubule is comprised mainly of GDP-tubulin with only a 'GTP-cap' (Howard and Hyman, 2003). Exposed GDP-β-tubulin at the plus end leads to microtubule depolymerisation, or catastrophe (Drechsel and Kirschner, 1994; Howard and Hyman, 2009). Cytoskeleton structure, cell motility and cell division (Steinborn *et al*., 2002) are dependent on the correct architecture and dynamic behaviour of these cylindrical polymers, with microtubule growth and catastrophe supported by numerous microtubule-associated proteins (MAPs). The tubulin subunits within microtubules can be the target of various post-translational modifications, including polyglutamylation, polyglycylation and tyrosination of the

carboxyl terminus, which influence the behaviour and affinity of these MAPs, modulating the activity of the microtubules concerned (Bonnet *et al*., 2001).

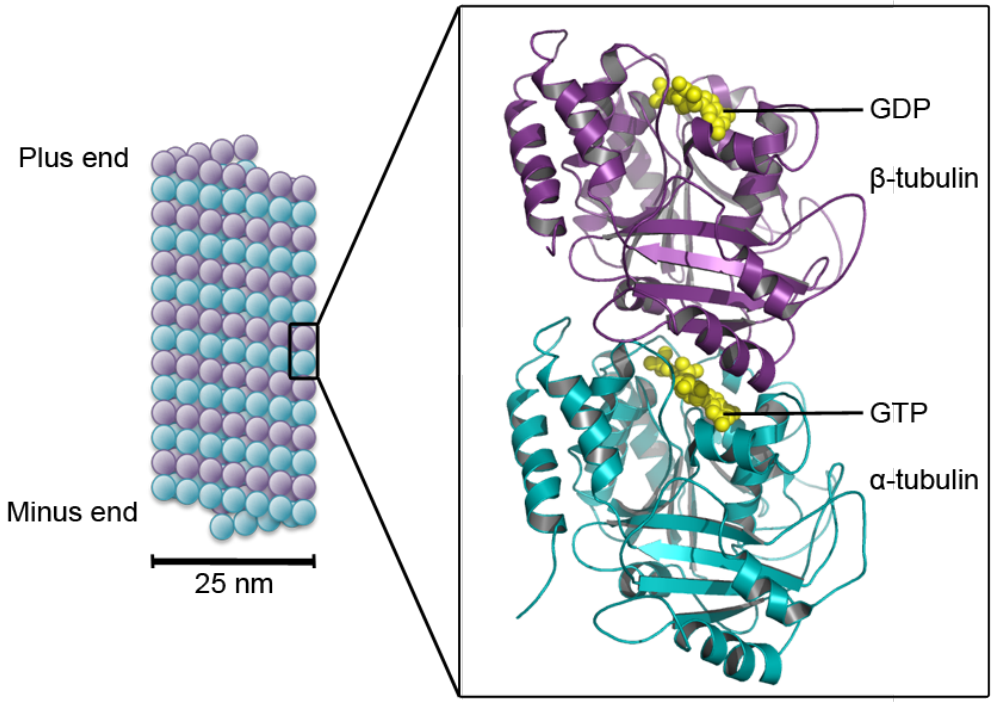


Figure 1.1. A typical microtubule
A schematic representation of a microtubule (left) with tubulin subunits depicted as cyan and purple circles. α- and β-tubulin are also shown in cartoon form (PDB 1tub; Nogales *et al*., 1998; right) with bound nucleotides represented by yellow spheres.

1.2.1. Trypanosomatid microtubules

In trypanosomatids, microtubules are abundant in the flagellum (Gallow and Anderton, 1983; Seebeck *et al*., 1983), facilitating cell movement (Ralston *et al*., 2009), while a complex network of subpellicular microtubules beneath the cell membrane maintains the overall parasite shape (Gallo and Precigout, 1988; Gull, 1999). The precise arrangement of microtubules is dependent on their function. For example, flagellar microtubules extend from the basal body along the trypanosome flagellar attachment zone and adopt a 9+2 arrangement in the axoneme that is typical of eukaryotic flagella. Two single microtubules are in the centre of a larger tubule formation consisting of nine fused-doublet microtubules, one of which contains thirteen protofilaments while the

fused component is an incomplete ten-protofilament microtubule (Ralston and Hill, 2008; Ralston *et al*., 2009). The flagella is required for pathogenicity as well as motility, attaching the parasite to the salivary gland of the insect vector prior to the infection of a new host and also contributing to the evasion of the host immune system (Engstler *et al.*, 2007; Ralston and Hill, 2008). Downregulation of α- and β-tubulin gene expression in high-throughput *T. brucei* RNA interference studies resulted in significant loss of cell fitness (Alsford *et al*., 2011). Anti-trypanosome tubulin antibodies prevent *in vitro* cell growth (Lubega *et al*., 2002) and tubulin polymerisation is the target of studies that seek to identify new therapeutic agents for diseases caused by trypanosomatids (Morgan *et al*., 2008), highlighting the importance of tubulin and microtubules in these organisms.

1.3. Tubulin-binding cofactors

Complex protein folding mechanisms exist in many systems to ensure correct protein structure and function. To form the thirteen protofilaments that constitute the typical microtubule, the folding of the α- and β-tubulin subunits must be carefully regulated to ensure their correct tertiary structure and to prevent spontaneous aggregation or premature polymerisation (Lundin *et al*., 2010). Improper tubulin folding can have a detrimental effect on many cellular functions. The tubulin-binding cofactors (TBC) were first identified through their ability to associate with individual tubulin subunits and implicated in the folding and heterodimerisation process (Fontalba *et al*., 1993; Gao *et al*., 1993; Melki *et al*, 1996; Tian *et al*., 1997; Lopez-Fanarraga *et al*., 2001).

Tubulin polypeptides are first processed by prefoldin (Vainberg *et al*., 1998) and the cytosolic chaperonin containing T-complex polypeptide 1, CCT (Valpuesta *et al*., 2002). The route to tubulin dimerisation is not fully validated but a working model suggests that at least five TBCs, named A-E, capture the *quasi*-native tubulin monomers when released from CCT (Gao *et al*. 1992; Melki *et al*., 1993; Tian *et al*., 1995). A pathway of transfer of the individual subunits follows (**Figure 1.2**). α- and β-tubulin are first acquired by tubulin-binding cofactor B (TBCB, Tian *et al*., 1997) and TBCA (Gao *et al*., 1993; 1994) and transferred to TBCs E and D, respectively. Tubulin subunits are brought together in a supercomplex along with cofactors C, D and E (Tian *et al*., 1997). Hydrolysis of GTP by β-tubulin, influenced by TBCC, then enables the release of the αβ-tubulin heterodimer that, following nucleotide exchange, can proceed to polymerisation (Fontalba *et al*., 1993; Tian *et al*., 1996; Kirik *et al*., 2002; Lundin *et al*., 2010). It should be noted that the tubulin-binding cofactors can be named differently throughout the literature whereby the terms tubulin-binding and tubulin-folding as well as cofactor and chaperone are often used interchangeably. For clarity here, all are named tubulin-binding cofactors (TBC) followed by the respective A-E identifier, except in cases where a homologous protein is universally given a unique notation.

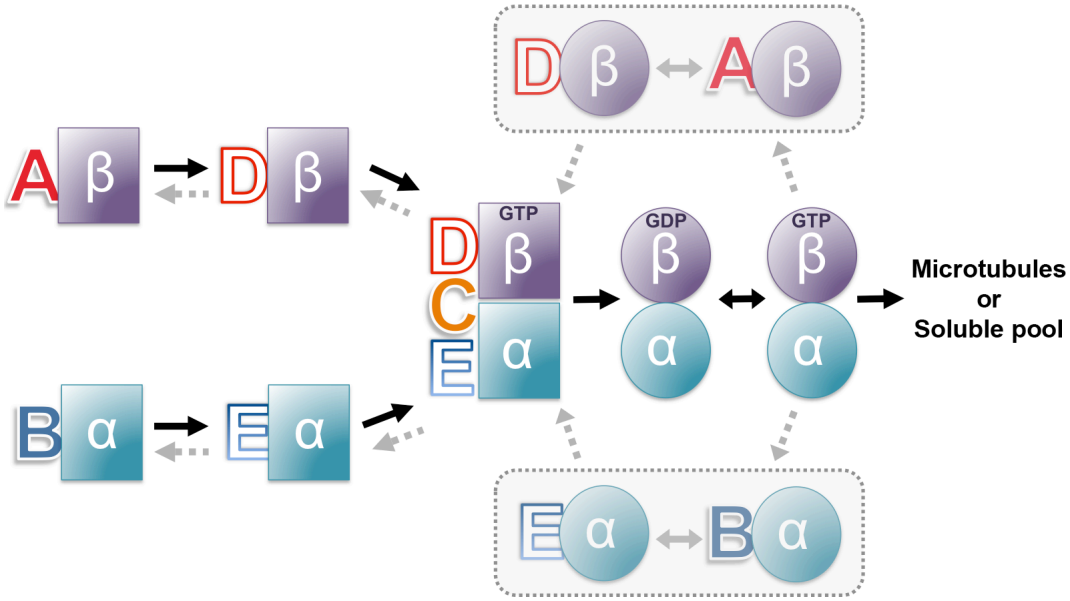


Figure 1.2 Schematic model of tubulin dimerisation
A simplified schematic representation depicting the potential involvement of TBCs A-E in the pathway to the dimerisation of α- and β-tubulin (based on models from Tian *et al*., 1997; Bhamidipati *et al*., 2000; Lopez-Fanarraga *et al*., 2001; Grynberg *et al*., 2003; Lundin *et al*., 2010). *Quasi*-native α- and β-tubulin subunits following release from CCT are depicted as cyan and purple squares while assembly-competent subunits are shown as circles of the same colour. TBCs are represented by the appropriate A-E identifier. Grey or dashed arrows indicate potential pathways that are not fully understood.

In addition to their participation in the post-translational folding of tubulin prior to heterodimerisation and polymerisation (Gao *et al*., 1994), there is growing evidence that the TBCs also contribute to the dissociation of microtubules with a greater involvement in microtubule dynamics (Martín *et al*., 2000; Bhamidipati *et al*., 2000; Kortazar *et al*., 2007). Microtubule dynamics, the elongation or shrinkage of most microtubules (Desai and Mitchison, 1997), are influenced by many MAPs. The TBCs are thought to be distinct from other MAPs that only bind to polymerised microtubules but interactions with fully folded native tubulin have been reported (Bhamidipati *et al*., 2000; Kortazar *et al*., 2007). TBCD is able to disrupt the αβ-tubulin heterodimer to sequester GTP-β-tubulin. Cells overexpressing TBCD result in a substantial loss of detectable microtubules (Martín *et al*., 2000) or cell death unless rescued by β-tubulin (Hirata *et al*., 1998) or prevented by an additional modulator, Arl2 (ADP ribosylation factor-like

2; Bhamidipati *et al*., 2000). Similarly, disruption to the genes encoding TBCs B and E in fission yeast is lethal (Hirata *et al*., 1998). These two cofactors may form a ternary complex with α-tubulin following dissociation of the αβ-tubulin heterodimer while free β-tubulin is captured by TBCA (Kortazar *et al*., 2007). Additionally, TBCA can prevent spontaneous over-polymerisation through the maintenance of a reservoir of excess soluble β-tubulin (Lopez-Fanarraga *et al*., 2001) and microtubule assembly can be compromised if TBCA levels are altered (Archer *et al*., 1995). Tubulin monomers acquired through the dissociation process can then be recycled through subsequent dimerisation and incorporated into growing microtubules or these pools may act as an intermediate in the route towards degradation (Fanarraga *et al*., 1999).

The complex functions exhibited by the TBCs as outlined above are only beginning to be dissected. There is a lack of detailed evidence to substantiate the tubulin dimerisation pathway shown in **Figure 1.2** and analysis of the three-dimensional structures of the TBCs can support such validation efforts. Secondary structure predictions and domain analysis suggest that each TBC possesses its own distinct structural features (Grynberg *et al*., 2003). Crystal structures are available for three TBCA homologues (PDB 1qsd, Steinbacher, 1999; 1h7c, Guash *et al*., 2002; 3mxz, Lu *et al*., 2010) and nuclear magnetic resonance (NMR) studies afford insight into fragments of TBCB (1t0y, Lytle *et al*., 2004; 1v6e, Zhao *et al*., unpublished; 1whg, Saito *et al*., unpublished; 2kj6, Mani *et al*., unpublished; 2kjr, Ramelot *et al*., unpublished), TBCC (2l3l, Garcia-Mayoral *et al*., 2011; 2yuh, Saito *et al*., unpublished) and TBCE (1wjn, Sato *et al*., unpublished). Several cofactors contain motifs implicated in protein-protein interactions. In particular, TBCs B and E are believed to contain both ubiquitin-like and cytoskeleton-associated protein-glycine-rich (CAP-Gly) domains.

CAP-Gly modules are found in a number of other MAPs, important for microtubule recognition and binding (Weisbrich *et al*., 2007). TBCE also contains a leucine-rich linker segment and TBCD is a large protein with a HEAT repeating motif, both of which are involved in mediating interactions with other proteins (Grynberg *et al*., 2003). TBCA and TBCC consist of compact helical regions while TBCC is extended by an additional β-stranded domain, sharing similarity with the human protein, retinitis pigmentosa 2 (RP2).

At the beginning of the studies presented in this thesis, no structure of a protist TBC was known. It is desirable to understand how the distinct structural folds of each TBC impacts on function and to compare trypanosomatid TBC structures with those of other organisms to build a more complete model of their complex functional capabilities.

1.4. Crystallisation and structure determination

Crystallographic methods are the primary source of structural information presented in this thesis. A crystal structure can provide detailed atomic information and inform on protein functionality, which is inherently linked to overall and local structural features. Interaction potential can be analysed and protein-protein or protein-ligand complexes can provide molecular details of an interaction captured in crystalline form. Numerous texts are currently available to aid the study and application of protein crystallography. Notably, Rupp (2009) provides a comprehensive review of biomolecular crystallography and only an outline of selected experimental methodologies utilised in both **Part I** and **Part II** is presented here.

1.4.1. Crystallisation

Protein crystals are comprised of individual molecules repeating in a uniform manner in a crystal lattice. Prior to crystallisation, protein samples should usually be homogeneous and in the purest form possible to ensure that no impurities impede on the ordered packing of protein molecules. Crystal growth takes place in the metastable region of the crystallisation phase diagram (Asherie, 2004; **Figure 1.3**) where the concentrations of protein and the mixture of chemicals added to induce crystallisation (precipitant) are optimal. These concentrations and the composition of the precipitant mixture are unknown and cannot be accurately estimated. Rather, proteins are purified to homogeneity and subjected to screening procedures to search for suitable crystallisation-inducing agents. Before crystal growth can occur, a nucleation event must first begin the process. This can be spontaneous through first entering the labile zone but the state must return to the metastable region (Stura *et al*., 1994) as overnucleation can hinder the formation of single crystals suitable for diffraction studies. Alternatively, a seed can be introduced to a metastable mixture to initiate crystallisation (Bergfors, 2003) such as a fragment of a previous nucleated crystal, a whole crystal to increase the crystal size or via an exogenous or artificial seed (Chayen *et al*., 2001; D'Arcy *et al*., 2003a).

The main methods of crystallisation utilised in the studies presented in this thesis are sitting drop and hanging drop vapour diffusion. Sitting drops consist of a protein-precipitant mixture 'sitting' on a ledge adjacent to each of the 96-wells of a screening plate and are typically used for initial robotic screening of potential crystallisation conditions. Hanging drops for the optimisation or reproduction of known crystallisation conditions are used in a 24-well format where a larger volume drop can be placed on a

siliconised coverslide, inverted over a reservoir containing approximately 1 mL precipitant mixture and sealed. Additionally, some microbatch and under-oil crystallisation experiments were attempted. While vapour diffusion methods allow the equilibration of the concentrated reservoir and less concentrated protein-precipitant drops through net transfer of water, microbatch procedures capture the state within the phase diagram immediately following preparation (Chayen *et al*., 1990). Using oils of different composition, such as silicone oil in place of paraffin oil, can allow the investigator to manipulate the rate of vapour diffusion (Chayen, 1997a; 1997b; 1999; D'Arcy *et al*., 2004).

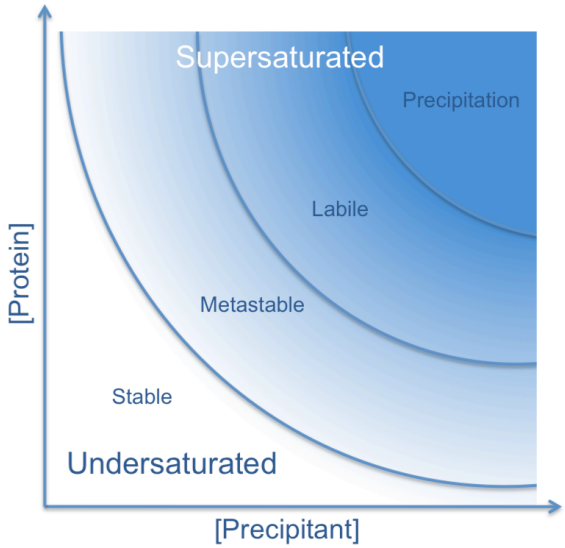


Figure 1.3. Crystallisation phase diagram
A generic crystallisation phase diagram (based on Asherie, 2004). Crystal growth occurs in the metastable zone following a nucleation event. Increased protein and precipitant concentration leads to greater supersaturation but beyond the labile region, protein precipitation or aggregation occurs. Proteins remain in solution in the undersaturated, stable zone.

1.4.2. Crystal structure determination

Following the successful growth of suitable crystals of a target protein, X-ray diffraction data are obtained. Monochromatic X-rays are directed at a single crystal suspended within a cryoprotective mixture, mounted on a goniometer and cooled in a stream of gaseous nitrogen. Electrons within the ordered crystal lattice diffract the

incident X-rays, a phenomenon first discovered by von Laue in 1912 (von Laue, 1915; Schmahl and Steurer, 2012). A series of spots or reflections are detected on a diffraction image when Bragg's law (Bragg, 1913; Bragg and Bragg, 1913) is satisfied, given as

$$\boldsymbol{n\lambda = 2d\sin\theta}$$

where n is any integer, λ is the X-ray wavelength, d is the spacing between lattice planes and θ is the incident angle of the wave. Multiple diffraction images are collected as the crystal is rotated. The diffraction experiment can then provide two of the three necessary parameters to obtain the three-dimensional structure of the protein. The X-ray wavelength used in the experiment is known and the amplitude of the wave can be derived from the intensity of the reflections. However, the phase of the wave cannot be measured directly (Hauptman, 1991). This is known as the crystallographic 'phase problem' and is the major hurdle common to solving any macromolecular crystal structure (Taylor, 2003).

Ab initio methods for phase estimation can be employed in simple cases with atomic resolution diffraction, such as that of small molecule crystals. However, this approach has limited application to macromolecular structure solution although improved large-scale computing clusters are providing increased capabilities for such methods (for example, *ARCIMBOLDO*; Rodríguez *et al*, 2012). Otherwise, methods of determining the phase include molecular replacement (MR), single or multiple wavelength anomalous dispersion (SAD or MAD) and single or multiple isomorphous replacement (SIR or MIR) (Taylor, 2003). A combined use of isomorphous and anomalous scattering methods are also possible (SIRAS or MIRAS).

MR can be employed when a structure of a related protein is available (Rossmann and Blow, 1962). The model must share a structurally similar fold as the target protein which is typically indicated by a sequence identity of at least 25 % (Taylor, 2003), although there will undoubtedly be exceptions. Patterson maps can be calculated without knowledge of the phases for both the target protein using the amplitudes of the diffraction data and for the homologous structure based on the known atomic coordinates. The Patterson function, originally introduced by Patterson (1934; 1935), can be described as

$$\boldsymbol{P(u,v,w)} = \frac{\mathbf{1}}{\mathbf{V}} \sum_{\boldsymbol{hkl}} \boldsymbol{|F(hkl)|^2} \mathbf{e}^{\,\boldsymbol{-2\pi i\,(hu+kv+lw)}}$$

where V is the volume of the unit cell, (*u,v,w*) is a point on the Patterson map and |*F*(*hkl*)| represents the amplitudes for a set of indices, *hkl*. It can be applied to obtain information relating to the interatomic distances within the structure. A successful solution requires the superimposition of the search model Patterson function onto that of the target (Rossmann and Blow, 1962), first identifying the correct orientation by rotation and subsequently the translated positional component.

Isomorphous crystals can be soaked with heavy atoms and intensity differences for the derivatives used to solve the phase problem (Perutz, 1956). Diffraction datasets are collected from a native crystal and one (SIR) or more (MIR) heavy atom derivatives and the differences in measured reflection intensities can allow the heavy atom locations to be derived by the analysis of Patterson maps.

Anomalous dispersion methods, SAD and MAD, exploit wavelength-dependent properties of certain atoms within a native or derivative crystal (Hendrickson *et al*.,

1985). In particular, sulfur atoms of cysteine and methionine residues can provide sufficient anomalous signal at a Cu $K\alpha$ home source (Doutch *et al*., 2012) or crystals grown using protein that incorporates seleno-methionine in place of methionine (Hendrickson *et al*., 1990) can be used to identify the selenium atoms at synchrotron beamlines by utilising a shorter wavelength. Many other elements can be used in this manner, including iodine or other heavy atoms soaked or co-crystallised with the target protein, and each display a unique absorption curve (Merritt, 2012). These profiles provide the elemental absorption edge, when the X-ray photons of a particular wavelength are absorbed by the atom to promote an electron from an inner shell. X-rays are then scattered with an altered phase For native data, Friedel's law denotes that the intensities of a reflection, (*h,k,l*), are equal to that of the symmetry-related reflection, (*-h,–k,–l*), but this does not hold in the presence of an anomalous scatterer (Taylor, 2010). The resultant difference, the anomalous or Bijvoet difference, can then be used to derive the positions of the anomalous scattering atoms, producing a substructure that can be used to provide phases for the rest of the structural model (Taylor, 2010). The anomalous differences are apparent at all wavelengths but are greatest at the absorption edge. Data collected at more than one wavelength during a MAD experiment maximises the effect, where the absorption peak, inflection and remote X-ray energies are typically selected (Burla *et al*., 2004). Accurate measurements are required as the differences can be relatively small and highly redundant data are usually necessary to ensure this accuracy. Providing radiation damage is not significant, SAD or MAD can be performed using a single crystal and in the absence of a structurally similar model suitable for MR, the popularity of SAD, typically at or close to the absorptive peak wavelength, to solve the phase problem is increasing (Liu *et al*., 2012).

Once known, phases can be used to provide an electron density map (Hauptman, 1991). Phases can be improved through density modification and cycles of model building by interpretation of the electron density maps and refinement procedures. More structural information provides improved phases, which can then present more interpretable features. One way to monitor progress of the cyclic model fitting process is the calculation of residual factors or *R*-factors. A subset of data, typically around 5 % of the total, are excluded from the calculations of the working *R*-factor, R_{work}, so that this R_{free} should remain unbiased (Brünger, 1992). Overall, the *R*-factors reflect how well the structural model represents the experimental data and can act as a guide to the investigator, or an observer, as to the quality of the structural model. R_{work} and R_{free} should generally remain within approximately 5 % of each other. In a perfect structural model, an R_{factor} of zero would result but in macromolecular crystallography, this figure is often much higher (Wilson, 1950) due to errors in the data and model and is usually related to the diffraction data resolution.

1.5. Aims

Trypanosomatids represent a valuable eukaryotic model organism and in particular, studies of the biology of the flagellum and the architecture of the trypanosome cytoskeleton are areas that may offer wider scientific relevance. Despite the importance of microtubules in these organisms, there is currently little biochemical and structural information on tubulin folding and microtubule formation in trypanosomatid parasites. At the beginning of the studies presented in this thesis, no structure of a protist TBC was known. We aim to generate recombinant gene expression systems for the TBCs from the genetically tractable species, *T. brucei* and *L. major*. Isolation and

crystallisation of these proteins will then form the basis of structural examination using X-ray crystallographic methods. Structural features that are conserved between organisms can contribute to a fundamental understanding of eukaryotic biology while differences between the TBCs from different organisms can provide insight into specialist features of trypanosomatid parasites. Ultimately, advancement in the biological understanding of the trypanosomatids can provide a stronger basis for drug development.

PART I

Tubulin-binding cofactors

2. Materials and methods

2.1. Materials

2.1.1. General reagents

All chemical reagents were purchased from Sigma-Aldrich, VWR International (BDH Prolabo and AnalaR Normapur branded products) or Formedium unless otherwise indicated. Restriction endonucleases were purchased from New England Biolabs UK. Oligonucleotides were custom synthesised by Thermo Fisher Scientific. HyperLadder I DNA standard molecular weight marker was purchased from Bioline Reagents Ltd.

Materials for routine SDS-PAGE were purchased from both Life Technologies and Bio-Rad including pre-cast Novex NuPAGE Bis-Tris gels, Mini-PROTEAN TGX Stain-Free gels and corresponding running buffers. Protein standard markers were also from either Life Technologies (SeeBlue Plus2 Pre-Stained Standard) or Bio-Rad (Precision Plus Protein All Blue and Unstained standards). Coomassie-based Instant*Blue* protein stain was obtained from Expedeon.

2.1.2. Bacteria and media

All strains of *Escherichia coli* were purchased from Stratagene unless included within a molecular biology cloning kit, such as Life Technologies' TOP10 competent cells. Lysogeny Broth (LB; Bertani, 1951) and Autoinduction (AI) liquid media (Studier, 2005) were prepared by the College of Life Sciences Media Services (University of Dundee). LB-agar plates were also provided by this service.

2.2. Molecular biology

2.2.1. Tubulin-binding cofactors

Genes predicted to encode TBCs A-E were identified in *T. brucei* and *L. major* strains 927 and Friedlin, respectively, from annotations available in *GeneDB* (Logan-Klumpler *et al*., 2012). Oligonucleotide primers were designed to amplify the full-length open reading frames from genomic DNA (gDNA) by polymerase chain reaction (PCR). Additional nucleotides were engineered at both the 5′ and 3′ ends based on known restriction endonuclease recognition sequences to allow subcloning into the expression vector, pET15b-TEV (modified from pET15b, Novagen). Under the control of the T7 promoter, this vector produces protein with an *N*-terminal hexa-histidine tag that is cleavable by tobacco etch virus (TEV) protease. A full list of primers is given in **Table 2.1**. PCR was carried out using KOD Hot Start DNA polymerase (Life Technologies) according to the manufacturer's guidelines. Typically, a 50 μL reaction mixture was composed of 0.3 μM sense and anti-sense primers, 0.2 mM each dNTP, 1.5 mM $MgSO_4$, 0-2.5 % (v/v) DMSO, 100 ng gDNA template and 1 U polymerase in associated buffer. Thermal cycling conditions are summarised in **Table 2.2**.

Target Protein	GeneDB ID	PCR Primers (5′ to 3′)	Restriction enzyme
*Tb*TBCA	Tb11.01.7825	CATATGTCGACAAGTGAGGGCAATGC GGATCCTTATTGTTTAGACAAAGTTAGCTGTGC AC	*Nde*I *BamH*I
*Tb*TBCB	Tb10.61.2930	---	---
*Tb*TBCC	Tb11.01.1240	CATATGGAGGAAAGGTTTCTTAGAACG GGATCCTCATACCGCAGCAGCATCC	*Nde*I *BamH*I
*Tb*TBCD	Tb927.8.6200	ATTAATATGGATGGTGAACAATTAGAAACAGAT CCCCTCACG AGATCTTCAGTACCCTGCCTCATGTACGAGATG TAGGTATG	*Ase*I *Bgl*II
*Tb*TBCE	Tb927.3.2680	CTCGAGATGGCGAGTCCGGAAGTCAG CTCGAGTCAACTCCGAAGGGATGTGTC	*Xho*I *XhoI*
*Lm*TBCA	LmjF32.2970	CATATGATGTCTGATTCTACTGAATCCACC GGATCCTTACGAAACTGCCGCTTGGCCGTC	*Nde*I *BamH*I
*Lm*TBCB	LmjF18.0460	CATATGTCTATTGTAAAGGTAATGATC GGATCCTCAGTATTCTTGAGGAGGAAAG	*Nde*I *BamH*I
*Lm*TBCC	LmjF36.3160	CATATGGAGGCGAAGTTCCTCAAGCTGC GGATCCTCAATGGTCTGCGGCAACGGTGGAG	*Nde*I *BamH*I
*Lm*TBCD	LmjF24.2020	CATATGCCTTCACCTCACATGGAGAAGGTGCCA ATG GGATCCTCAGTAGCCGGTCTCCTGCACGAGCGA CTTG	*Nde*I *BamH*I
*Lm*TBCE	LmjF03.0770	CATATGCCGTCCCCGTCGTCATCGACG GGATCCTTACCGCAACGAGGCGTCCTCGACG	*Nde*I *BamH*I
*Tb*α-tubulin	Tb927.1.2340	CATATGCGTGAGGCTATCTGC GGATCCCTAGTACTCCTCCACATCC	*Nde*I *BamH*I
*Tb*β-*t*ubulin	Tb927.1.2370	CATATGCGCGAAATCGTCTGCGTTCAG CTCGAGCTAGTATTGCTCCTCCTCGTCG	*Nde*I *Xho*I
*Lm*α-tubulin	LmjF13.0280	CATATGATGCGTGAGGCTATCTGC GGATCCTTAGTACTCCTCGACGTCC	*Nde*I *BamH*I
*Lm*β-tubulin	LmjF33.0794	CATATGCGTGAGATCGTTTCCTGC AGATCTCTAGTAGGCCTCCTCtTCCTC	*Nde*I *Bgl*II

Table 2.1. TBC and tubulin PCR primers
All primers used for PCR of TBCs and tubulins from *T. brucei and L. major* in conjunction with the *GeneDB* accession ID and restriction enzymes used for subcloning. Underlined nucleotides are enzyme recognition sites and in lower case are any silent mutations introduced to avoid primer self-complementarity. *Tb*TBCB was the subject of studies by colleagues Rachel Morgan and Jennifer Fleming and is included in this list for completeness only.

Blunt-ended PCR products were ligated into pCR-Blunt II-TOPO (Life Technologies) and transformed to TOP10 cells. The product was digested with appropriate restriction endonucleases, creating cohesive ends and allowing insertion into pET15bTEV using T4 DNA ligase (New England Biolabs). Final plasmid DNA were transformed to

selected competent *E.coli* cells via the 42°C heat-shock technique (Maniatis *et al*., 1982; Sambrook *et al*., 1989).

PCR Stage	Temperature (°C)	Time (s)	
Polymerase activation	95	120	
Denaturation	95	20	Repeated for 30 cycles
Annealing	Primer T_m	10	
Extension	70	15-20 kb^{-1}	
Final Extension	70	300-600	
Storage	4	∞	

Table 2.2. PCR parameters
Summary of PCR parameters used to amplify TBC and tubulin gene sequences from gDNA. Primer melting temperatures (T_m) were estimated using *OligoCalc* (Kibbe, 2007).

2.2.2. Tubulin

Based on work published by Giles *et al*. (2009) and *GeneDB* annotated sequences (Logan-Klumpler *et al*., 2012), genes encoding *T. brucei* and *L. major* α- and β-tubulin were cloned from gDNA following the same protocols as the TBCs. Details are included in **Tables 2.1** and **2.2**. Final gene sequences were inserted into expression vectors pET15b-TEV, pET15b-MBP-TEV (modified from pET15b-TEV) and pGEX-6P-1-TEV (modified from pGEX-6P-1, GE Healthcare). The latter two vectors contained sequences encoding maltose-binding protein (MBP) and glutathione-*S*-transferase (GST) fused to the *N*-terminus of the protein of interest, both cleavable by TEV protease. The MBP fusion construct was based on the pET15b-TEV vector while the GST fusion construct maintained its original pGEX-6P-1 backbone with the precision protease recognition site replaced with that of TEV protease.

2.2.3. Truncated TBC construct design

All full-length annotated sequences were examined using several bioinformatic approaches including the secondary structure prediction software *PSIPRED* (Jones,

1999), the disorder predictor *GlobPlot* (Linding *et al*., 2003) and the crystallisability prediction tool *XtalPred* (Slabinski *et al*., 2007). Accompanied by manual inspection of multiple sequence alignments (*MUSCLE*; Edgar, 2004), these programs assisted in the design of a number of truncated and globular domain constructs. For most constructs, PCR using the corresponding full-length plasmid as template DNA was followed by cloning procedures as discussed previously. Selected constructs were generated using Genscript's CloneEZ system where primers contained additional bases at the 5'-end corresponding to the target vector sequence. PCR products were inserted directly into linearised vector using the CloneEZ recombination enzyme. All truncated TBC constructs are shown in **Table 2.3**.

2.2.4. Site-directed mutagenesis

Conservative leucine-methionine mutations were introduced to the *Lm*TBCC *C*-terminal domain construct referred to from this point as *Lm*TBCC152 (*Lm*TBCC residues 152-335). Primers to generate L215M and L223M mutants were designed with the aid of Agilent Technologies' *QuikChange primer design tool* (Novoradovsky *et al*., 2005) and *OligoCalc* (Kibbe, 2007) (**Table 2.3**). PCR was carried out using PFU Ultra HF polymerase (2.5 U, Stratagene). Also included in the PCR mixture were 125 ng sense and antisense primers, 0.2 mM each dNTP and 5 ng non-mutated plasmid template. 16 cycles of 95°C (30 s), 55°C (1 min) and 68°C (1 min kb^{-1}) were followed by treatment with *dpn*I (10 U) and the final sample transformed to XL1 Blue cells. This process was repeated to generate a third construct containing both mutations.

Template	Construct	PCR Primers (5′ to 3′)	Restriction enzyme
*Tb*TBCA	5-128	CATATGGAGGGCAATGCTGCTAATCGC	*Nde*I
		GGATCCTTACGAAACTGCCGCTTGGCCGTC	*BamH*I
	19-128	CATATGGCTGAAGATCCTTTTGTAAAGGCTC	*Nde*I
		GGATCCTTACGAAACTGCCGCTTGGCCGTC	*BamH*I
*Tb*TBCC	1-107	CATATGGAGGAAAGGTTTCTTAGAACG	*Nde*I
		GGATCCTCAGGATCGGGGCCCCTG	*BamH*I
	1-112	CATATGGAGGAAAGGTTTCTTAGAACG	*Nde*I
		GGATCCTCAAGAGGAAAACTTGAAGGATCGG	*BamH*I
	138-308	CATATGGAACCGGAACTTCCTGCAG	*Nde*I
		GGATCCTCATACCGCAGCAGCATCC	*BamH*I
	149-308	CATATGGCACGTGACAGAACGCTATG	*Nde*I
		GGATCCTCATACCGCAGCAGCATCC	*BamH*I
*Tb*TBCE	1-208	CTCGAGATGGCGAGTCCGGAAGTCAG	*Xho*I
		CTCGAGTCAATTCTTGTTAGCTTCGTGAAGGATG	*Xho*I
	44-208	CTCGAGCTTGTTGGGATGGAGAAAACTCG	*Xho*I
		CTCGAGTCAATTCTTGTTAGCTTCGTGAAGGATG	*Xho*I
	294-531	CATATGTCCCTCGTGGATTACGCGTATC	*Nde*I
		CTCGAGTCAACTCCGAAGGGATGTGTC	*Xho*I
	314-531	CATATGACCATCACAGATGCATGCACACTG	*Nde*I
		CTCGAGTCAACTCCGAAGGGATGTGTC	*Xho*I
	436-531	CATATGGCAAGTCACGATGGTACCATGC	*Nde*I
		CTCGAGTCAACTCCGAAGGGATGTGTC	*Xho*I
*Lm*TBCA	20-125	CATATGGCGCCGAACGAAAAGACGCTG	*Nde*I
		GGATCCTTACGAAACTGCCGCTTGGCCGTC	*BamH*I
*Lm*TBCB	1-138 *	GTATTTTCAGGGCCATATGTCTATTGTAAAG	*Nde*I
		GTTAGCAGCCGGATCCTTACACTCCAATGCCAG	*BamH*I
	1-151 *	GTATTTTCAGGGCCATATGTCTATTGTAAAG	*Nde*I
		GTTAGCAGCCGGATCCTTACGCCTCCGTCTTG	*BamH*I
	1-165 *	GTATTTTCAGGGCCATATGTCTATTGTAAAG	*Nde*I
		GTTAGCAGCCGGATCCTTAGCCGGGTTGGCAAC	*BamH*I
	1-222 *	GTATTTTCAGGGCCATATGTCTATTGTAAAG	*Nde*I
		GTTAGCAGCCGGATCCTTATACCTGATTAG	*BamH*I
*Lm*TBCC	1-103	CATATGGAGGCGAAGTTCCTCAAGCTGC	*Nde*I
		GGATCCTCACGCTGCGCCGCTGCAC	*BamH*I
	1-111 *	GTATTTTCAGGGCCATATGGAGGCGAAGTTCCTC	*Nde*I
		GTTAGCAGCCGGATCCTCACGCCTTGAGGCGTG	*BamH*I
	152-335 *	GTATTTTCAGGGCCATATGAGCAGCACCAACG	*Nde*I
		GTTAGCAGCCGGATCCTCAATGGTCTGCGGCAAC	*BamH*I
	164-335 *	GTATTTTCAGGGCCATATGTCCACGGGCACG	*Nde*I
		GTTAGCAGCCGGATCCTCAATGGTCTGCGGCAAC	*BamH*I
*Lm*TBCC 152-335	L215M	GCGTGCCATCAGCTGCGAaTGAAGGGCTG	---
		CAGCCCTTCAtTCGCAGCTGATGGCACGC	---
	L223M	GCTGCACAAACCTCGATaTgTACGTGTGGTGCGCG	---
		CGCGCACCACACGTAcAtATCGAGGTTTGTGCAGC	---

Table 2.3. TBC truncated constructs and corresponding PCR primers
All completed truncated TBC clones are listed. Constructs are named according to the full-length template amino acid numbering. Restriction enzyme recognition sites are underlined and mutated nucleotides are in lower case. Constructs marked with * were cloned using the CloneEZ system (Genscript).

2.2.5. DNA purification and sequencing

DNA integrity was monitored throughout all stages of cloning by gel electrophoresis in 1% agarose gels stained with ethidium bromide (approximately 0.2 μg mL^{-1}). High purity plasmid DNA was prepared using the QIAprep spin miniprep kit (Qiagen) and digested DNA fragments excised from agarose gels were purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. An extinction coefficient of 0.02 μg mL^{-1} cm^{-1} at 260 nm was used to estimate DNA concentration. Sample quality was further assessed by ensuring an $A_{260/280}$ of approximately 1.8 (values significantly lower than 1.8 may be indicative of protein contamination). All final DNA constructs were verified by DNA sequencing (DNA Sequencing Service, University of Dundee). Where required, additional internal primers were designed to collect sequence data along the entire length of the gene of interest.

2.3. Gene expression and protein purification

2.3.1. Recombinant protein production

Plasmids containing the target genes were transformed into *E.coli* BL21 (DE3) Gold or BL21 (DE3) pLysS cells for expression. For proteins produced in LB, each 1 L media containing 50 mg L^{-1} carbenicillin (and 25 mg L^{-1} chloramphenicol if required) was inoculated with 15 mL starter culture and grown in 2 L Erlenmeyer flasks at 37°C in an Infors standard shaking incubator (200 rpm) until the optical density (OD) at 600 nm reached 0.6-0.8. Cultures were cooled, expression induced with 0.5-1 mM isopropyl thio-β-D-galactoside (IPTG) and growth continued at 18-22°C for 16 h. For expression using AI media, a freshly transformed cell colony was added directly to the broth and allowed to grow for 72 h at room temperature with agitation at 200 rpm. *Lm*TBCD was

expressed using ArcticExpress (DE3) cells in LB. Cultures were grown at 30°C to an OD_{600} of 1-2 then induced with 0.5 mM IPTG and incubated at 12.5°C for 16 h. All cells were harvested by centrifugation at 4000 rcf for 30 min at 4°C and cell pellets resuspended in appropriate lysis buffer for storage at -20°C.

2.3.2. Protein purification

Quantification of TBCA constructs was carried out via the measurement of absorbance at 595 nm in Coomassie (Bradford) protein assay reagent (Thermo Fisher Scientific) using bovine serum albumin as a reference (based on Bradford, 1976). Protein presence was monitored during purification by both A_{280} and A_{215}. All other proteins were quantified by the measurement of absorbance at 280 nm and concentration estimated using theoretical extinction coefficients (*ProtParam*; Gasteiger *et al*., 2005). Size exclusion chromatography columns (GE Healthcare) were all previously calibrated using molecular mass standards from Bio-Rad (thyroglobulin, 670 kDa; γ-globulin, 158 kDa; ovalbumin, 44 kDa; myoglobin, 17 kDa; vitamin B_{12}, 1.35 kDa) and were equilibrated in corresponding running buffer prior to use. All immobilized metal affinity columns were pre-loaded with $NiCl_2$ and washed with relevant buffer solution. ÄKTA Explorer, Purifier and Prime systems (GE Healthcare) were used throughout and all buffers were filter sterilised (0.2 μm) and degassed prior to use. Standard purification buffers are provided in **Table 2.4**.

Buffer Name	Buffer
Lysis	50 mM Tris-HCl pH 7.5, 250 mM NaCl, 20 mM imidazole
Elution	50 mM Tris-HCl pH 7.5, 250 mM NaCl, 800 mM imidazole
Gel Filtration	50 mM Tris-HCl pH 7.5, 250 mM NaCl

Table 2.4. Standard purification buffers
The composition of standard purification buffers. Alteration to these standard buffers are detailed where appropriate.

Purification of all recombinant native TBCs followed the same general procedure. Frozen cells were thawed and resuspended in additional lysis buffer supplemented with an EDTA-free protease inhibitor cocktail tablet (Roche) and approximately 10 μg mL^{-1} deoxyribonuclease I (DNaseI). Cells were lysed using a French pressure cell press at 16 kpsi and lysate was clarified by centrifugation at 37,500 rcf for 30 min at 4°C. Soluble supernatant was filtered (0.2 μm) and loaded onto a 5 mL HisTrap HP column (GE Healthcare) for an initial affinity chromatography capture step. A gradient of 20-800 mM imidazole was applied. Eluate containing the over-expressed his-tagged protein of interest was treated with TEV protease at 22-30°C for 1-3 h. Dialysis to remove excess imidazole was followed by reverse affinity chromatography. Cleaved protein no longer bound to the column and was therefore separated from any histidine-rich contaminants, the protease and the cleaved peptide during this step. Purification was finalised by size exclusion chromatography using a Superdex 75 or Superdex 200 26/60 gel filtration column. Proteins were then dialysed and concentrated using centrifugal concentrator filter units with an appropriate molecular weight cutoff (Millipore or Sartorius). All proteins were stored at 4°C and further experiments performed within one week of purification. The mass of all final samples was verified by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF, University of Dundee Proteomics Facility) and purity assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.3.3. Selenomethionine protein preparation

Selenomethionine-labelled *Lm*TBCA and *Lm*TBCC152 mutant constructs (L223M and L215M/L223M double mutant) were prepared following an established protocol (Hall *et al*., 1999). *E.coli* B834 (DE3) cells, which are auxotrophic for methionine, were transformed with the appropriate expression plasmid and cultured in minimal media (Molecular Dimensions). This media, containing all amino acids except L-methionine, was supplemented with 40 mg L^{-1} selenomethionine (Molecular Dimensions) and 50 mg L^{-1} carbenicillin. Bacteria were cultivated at 37°C to mid-log phase. Following IPTG induction, incubation at 22°C was continued for a further 16 h. Cells were then harvested, lysed and the selenomethionyl-proteins were purified as described for native TBCs. Selenomethionine (SeMet) incorporation was confirmed by MALDI-TOF analysis.

2.3.4. TEV protease preparation

Catalytically active recombinant TEV protease was overproduced (pRK793 expression system; Kapust *et al*., 2001) and purified by similar protocols to those described above. Briefly, 1 mM IPTG was used to induce expression in BL21 (DE3) Codon Plus RIL cells cultured in LB. The histidine-tagged enzyme was purified by a single affinity chromatography step. Purified TEV protease was flash cooled in liquid nitrogen in 1 mg aliquots (typically 2-5 mg mL^{-1}) and stored at -80°C in 50 mM sodium phosphate buffer pH 8.0 containing 150 mM NaCl, approximately 250 mM imidazole, 10 % (v/v) glycerol, 1 mM dithiothreitol (DTT) and 1 mM EDTA. 1 mg TEV protease was used to cleave up to 20 mg tagged target protein.

2.3.5. Small scale test expression

The recombinant expression of TBCs B, D, E and all tubulin constructs was first tested on a small-scale. Typically, cells (BL21 (DE3) GOLD, BL21 (DE3) pLysS or Rosetta (DE3) pLysS) containing the relevant expression plasmid were grown in 15-20 mL liquid media (LB, AI or NZCYM (Blattner *et al.*, 1977)) at 37°C or ambient temperature. Expression was induced with 0.1-1 mM IPTG for cultures not grown in AI media. After induction, incubation temperature and duration were also varied. Cells were harvested by centrifugation and lysed chemically using BugBuster Mastermix (Novagen) according to the manufacturer's protocol. Soluble and insoluble cell lysate were then analysed by SDS-PAGE and immunoblot if required.

2.3.6. Western blot analysis

A Western blot was carried out to confirm the presence of target protein, if not apparent by SDS-PAGE. Following separation by SDS-PAGE, proteins were transferred to a polyvinylidene difluoride (PVDF) membrane using a Trans-Blot Turbo transfer pack and system (Bio-Rad). The membrane was then blocked using 5 % (w/v) dried skimmed milk followed by sequential applications of primary and secondary antibodies. Mouse monoclonal antibodies to 6XHis-tag or GST (1-3 μg, Abcam) were used as primary antibodies and the secondary antibody was anti-mouse IgG (raised in goat) conjugated to horseradish peroxidase (HRP) (0.25-0.7 μg, Sigma-Aldrich). Dilutions and wash steps were performed in phosphate-buffered saline (PBS) supplemented with 0.05 % (v/v) polysorbate 20. Pierce ECL Plus Western blotting substrate from Thermo Fisher Scientific was used to detect the secondary antibody HRP. Lumi-film chemiluminescent detection film (Roche) was then exposed to the membrane (1-30

minutes) and developed using a Compact X4 automatic X-ray film processor (Xograph Healthcare).

2.3.7. Protein identification

To confirm the identity of recombinant TBC and tubulin proteins and to investigate signs of degradation or proteolysis of full-length *Lm*TBCA, *Tb*TBCC and *Lm*TBCC, samples were subjected to SDS-PAGE and the protein bands visualised using Coomassie-based stain. Appropriate gel bands were excised and subjected to trypsin digestion. Peptides were identified by reverse phase liquid chromatography on a 4000 QTRAP LC/MS/MS system (AB SCIEX) and data were analysed using the *MASCOT* search engine (Matrix Science; Proteomics Facility, University of Dundee).

2.3.8. LmTBCC152 size exclusion chromatography

Approximately 0.4-0.5 mg freshly purified (stored 1-2 days at 4°C) *Lm*TBCC152 was passed through a Superdex 75 10/300 size exclusion column pre-equilibrated with either standard gel filtration buffer (**Table 2.4**) or 50 mM MES, 250 mM NaCl pH 6.5 to mimic the buffer composition during crystallisation. A sample of protein prepared previously and stored at 4°C for 30 days was passed through the same column in standard gel filtration buffer only. Two further experiments were carried out using monomeric and multimeric species immediately after isolation from a Superdex 75 26/60 column.

2.3.9. Buffer extraction screen

Full-length *Lm*TBCD and *Lm*TBCE were cultured as described (**2.3.1**). Cells were distributed equally in a 24-well round-bottomed block (Qiagen), harvested and frozen at -20°C. Approximately 0.15 g cell paste per well was resuspended in 0.75 mL of solution containing BugBuster protein extraction reagent (Novagen), DNaseI and one of 24 buffers originally developed for a protein solubility screen (Fyfe, 2008; **Table 2.5**). Standard lysis buffer was used as a control sample. As for small-scale expression tests, following centrifugation, soluble lysates were assessed by SDS-PAGE and Western blot.

Position	Buffer (0.1M)	pH	Position	Buffer (0.1M)	pH
A1	Glycine	3	**C1**	PIPES	7
A2	Citric acid	3.2	**C2**	MOPS	7
A3	Sodium acetate	4	**C3**	Sodium/Potassium phosphate	7
A4	Citric acid	4.5	**C4**	HEPES	7.5
A5	Sodium acetate	5	**C5**	Tris	7.5
A6	Sodium citrate	5.5	**C6**	(H)EPPS	8
B1	Sodium/Potassium phosphate	6	**D1**	Imidazole	8
B2	Bis-tris	6	**D2**	Borate (from acid)	8.5
B3	MES	6.2	**D3**	Tris	8.5
B4	ADA	6.5	**D4**	Bicine	9
B5	Bis-tris propane	6.5	**D5**	CHES	9.5
B6	Sodium cacodylate	6.5	**D6**	CAPS	10

Table 2.5. Buffer extraction screen
A list of the buffers used in a 24-well protein extraction screen.

2.4. Differential scanning fluorimetry

2.4.1. Thermal stability buffer screen

Differential scanning fluorimetry (DSF) was used to examine the potential stabilising effects of different buffer conditions. A thermal stability buffer screen was carried out according to methods similar to those described by Niesen *et al.* (2007). Protein was prepared to a final concentration of 5 μM in 10 mM HEPES, 150 mM NaCl pH 7.5 supplemented with SYPRO orange fluorescent dye (Life Technologies). 23 buffers,

prepared at 4x stock concentration, were added to the protein-dye mixture to a final 1x working concentration in a total volume of 40 μL. Each reaction was carried out in duplicate or triplicate in a thin-walled 96-well ABgene PCR plate (Thermo Fisher Scientific) sealed with a transparent adhesive PCR seal (4titude). Fluoresence readings were measured (excitation and emission wavelengths of 492 and 610 nm, respectively) using an Mx3005P QPCR system (Stratagene) and the incubation temperature was increased from 25°C to 95°C at 1°C min^{-1}. Data were analysed to calculate protein melting temperatures (T_m) using the QPCR system software, MxPro and GraphPad Prism (GraphPad Software).

2.4.2. GTP-binding screen

*Lm*TBCC152 was prepared in 20 mM Tris-HCl pH 7.5, 100 mM NaCl plus SYPRO orange dye as for the DSF buffer screening assay described above (**2.4.1**). The addition of different buffers as potential stabilising agents was replaced with 0-100 μM GTP and the same program executed.

2.5. Isothermal titration calorimetry

The ability of *Lm*TBCC152 to bind GTP was tested via isothermal titration calorimetry (ITC). In brief, 0.2 mM GTP was titrated against 20 μM *Lm*TBC152 in 50 mM PIPES, 100 mM NaCl pH 7 at 30°C. Further experimental details of this technique are provided in **Part II** (**5.5**).

2.6. Circular dichroism

Circular dichroism (CD) was performed on full-length *Tb*TBCA and *Tb*TBCC. Proteins were prepared in compatible buffer (20 mM Tris-HCl pH 7.5, 20 mM NaCl; Kelly *et al.*, 2005) and experiments carried out by Sharon Kelly (Protein Characterisation Facility, University of Glasgow). Far UV spectra were recorded on a JASCO J-810 spectropolarimeter in 0.02 cm pathlength quartz cuvettes with a bandwidth of 1 nm at 50 nm min^{-1} and a 0.5 s response time. The results of 5 scans were averaged and secondary structure estimates obtained using the CONTIN algorithm (Provencher and Glöckner, 1981; van Stokkum *et al.*, 1990) available on the *DICHROWEB* server (Lobley *et al.*, 2002; Whitmore and Wallace, 2004).

2.7. Crystallisation

2.7.1. Crystallisation screening

Sparse-matrix and focussed crystallisation screens were purchased from Molecular Dimensions (JCSG+, PGA, MIDAS and Morpheus) and Qiagen (Classics, PEGs, MPD and $AmSO_4$). These suites are each composed of 96 unique pre-mixed conditions designed to exploit known crystallisation agents whilst covering a wide chemical space, resulting in the highest possible crystallisation hit rates. 60 μL of each solution was dispensed into the reservoirs of MRC 96-well polystyrene sitting-drop plates (Molecular Dimensions) using a JANUS Automated Workstation (PerkinElmer). A Phoenix DT crystallisation robot (Art Robbins Instruments and Rigaku) was used to dispense two 0.1-0.2 μL drops of reservoir solution mixed with 0.1-0.2 μL protein at desired concentration (typically 5-10 mg mL^{-1} in the first instance). Crystallisation plates were incubated at room temperature and regular images taken by an automated Minstrel DT

imaging system (Rigaku). Images were inspected using the associated *CrystalTrak* software to identify potential crystallisation hits. All proteins of satisfactory purity were subjected to this screening process and were filtered (0.1 μm) immediately prior to use. Any potential screening hits were subsequently scaled up and further optimised to 24-well VDX hanging drop plates with 1 mL reservoir volume and 22 mm siliconised glass cover slides (Hampton Research).

2.7.2. LmTBCA crystallisation

Initial crystal hits of the shortened construct *Lm*TBCA20-125 were obtained and subsequently reproduced using full length *Lm*TBCA. Following several rounds of optimisation, crystallisation was achieved at 18°C using the hanging-drop vapour-diffusion method. Native untagged protein at 4 mg mL^{-1} in 50 mM Tris-HCl pH 7.5, 100 mM NaCl was mixed with an equal volume of reservoir solution (0.2 M $(NH_4)_2HPO_4$ and 1.6 M $(NH_4)_2SO_4$) and crystals appeared within 1-3 days. SeMet-*Lm*TBCA crystals were grown using the same technique with protein at 3.8 mg mL^{-1} and reservoirs containing 0.3 M LiCl and 1.0 M $(NH_4)_2SO_4$.

2.7.3. LmTBCC152 crystallisation

2.7.3.1. Screening hits

Crystals of poor morphology were first produced in a number of conditions of the PEGS suite using native *Lm*TBCC152 at 7.3 mg mL^{-1} in 20 mM Tris-HCl pH 7.5, 100 mM NaCl. Extensive attempts to optimise these hits were made, of which a selection are detailed below.

2.7.3.2. General optimisation

Initial hits were reproduced in both sitting-drop and hanging-drop formats. General optimisation of these hits included varying factors such as protein and precipitant concentration via the actual solution concentrations or through the protein:precipitant ratio within each drop. Crystallisation buffer type and pH was varied and protein sample buffer was altered according to the results of a DSF buffer screen. The effect of filtering protein samples was examined (Chayen, 2009) as well as the overall incubation temperature (4 and 18°C).

2.7.3.3. Reductive methylation

Reductive methylation of *Lm*TBCC152 lysine residues was carried out according to methods described by Kim *et al*. (2008), based on earlier studies by Rypniewski *et al*. (1993) and Rayment (1997). This method selectively modifies the amino group of lysine residues with the substitution of a terminal methyl group, potentially altering the protein surface and crystallisation properties. In brief, *Lm*TBCC152 at 5 mg mL^{-1} in 50 mM HEPES pH 7.5, 100 mM NaCl was treated with multiple additions of dimethylamine-borane complex and formaldehyde at 4°C. The reaction was quenched with 1 M Tris-HCl pH 7.5 and the protein solution extensively dialysed against the original buffer supplemented with 1 mM DTT. Precipitated material was removed by centrifugation and a sample of the final product was sent for MALDI-TOF analysis to assess the degree of methylation.

2.7.3.4. Seeding

Microseeding techniques were employed, where crystal fragments or sub-microscopic crystals were used as nucleants in fresh crystal drops (Bergfors, 2003). Large clusters

of crystalline needles grown by methods described above (***2.6.3.1***) were transferred to a sample of mother liquor in a 1.5 mL microcentrifuge tube. In the presence of a small nylon ball (4-5 mm diameter) such as those used in a traditional French pressure cell press (or alternatively, a seed bead purchased from Hampton Research), the crystal mixture was vigorously agitated using a vortex mixer for 10-60 s (Luft and DeTitta, 1999). Crystallisation drops were prepared as previously with reservoirs containing a lower precipitant concentration. For example, 20 % (w/v) polyethylene glycol monomethyl ether (PEG 2000 MME) compared to 25 % (w/v) but a gradient of values was always screened. Seed stock was then introduced to the drop and the tray left undisturbed for a minimum of 2 days. Various tools were used to dispense seed mixture including a standard nylon loop, a natural fibre seeding device (Hampton Research) and even a human eyelash (kindly supplied by Paul K. Fyfe).

For second and third rounds of experiments, seed stocks were obtained from crystals grown as a result of initial seeding. Seed crystals were always collected immediately prior to use. Crystals of different size and appearance were chosen as seeds. The level of dilution in reservoir solution and time of agitation were also varied. Additionally, seeding tools were passed through a drop containing crystal clusters and directly streaked through a fresh drop without any dilution or pulverisation (Stura and Wilson, 1991).

Other optimisation methods were applied in combination with microseeding efforts. Sitting-drops were prepared to screen around the hit condition and 0.05 μL seed mixture added using the Phoenix crystallisation robot or drops were streak-seeded. Commercial

sparse-matrix and additive screens were repeated with the addition of seed mixture and experiments under oil were also performed in the presence of seed material.

2.7.3.5. Crystallisation under oil

Standard hanging-drops were prepared with a layer of oil applied over the reservoir solution as a method of controlling the rate of vapour diffusion (Chayen, 1997a). A mixture of paraffin and silicone oils (Molecular Dimensions) were used in different ratios and barrier depths (Chayen, 1997b). Experiments similar to the microbatch under oil technique (Chayen *et al*., 1990) were performed following protocols used by D'Arcy *et al.* (2003b, 2004). A 1:1 mixture of paraffin and silicone oils (6 mL) was dispensed over the 72 wells of a polystyrene Terasaki plate (Greiner Bio-One). 1 μL each of protein and crystallisation solution were then pipetted under the oil layer and incubated at 18°C. Crystallisation solutions were chosen to screen the hits obtained from vapour-diffusion experiments and an additional sparse-matrix screen (Qiagen Classics) was also prepared in this manner.

2.7.3.6. Additives

Further crystal optimisation efforts involved the addition of small chemical components to screen for improved crystal morphology. This comprised performing a commercially available additive screen (Hampton Research). Manually prepared sitting drops containing equal volumes of *Lm*TBCC152 (7 and 5 mg mL^{-1}) and known crystallisation solution (0.1 M MES pH 6.5, 30 % (w/v) PEG 2000 MME) were supplemented with 96 different additives and crystal growth monitored. Several other chemicals such as dioxane and glycerol were added to the crystallisation mixture (5-15 % (v/v)) while 1-2 mM GTP and/or DTT were added to the protein sample in a bid to enhance

crystallisation. The incorporation of iodide ions to aid experimental crystal phasing power was attempted by adding NaI or KI to reservoir solutions (2-10 mM) or as a partial replacement for NaCl in the protein sample buffer (10-100 mM).

2.8. X-ray data collection and processing

2.8.1. X-ray sources

These studies were carried out using a number of in-house and synchrotron X-ray sources, details of which are provided in **Table 2.6**.

Source	Location	Details	Detector
A	In-house	Rigaku MicroMax-007 Rotating Anode Fixed wavelength (1.5418 Å)	Rigaku R-AXIS IV^{++} Image-plate
B	In-house	Rigaku MicroMax-007 HF Rotating Anode Fixed wavelength (1.5418 Å)	Rigaku R-AXIS IV^{++} Image-plate
C	In-house	Rigaku MicroMax-007 HFM Rotating Anode Fixed wavelength (1.5418 Å)	Rigaku Saturn 944 HG CCD
D	ESRF ID23-2	Synchrotron radiation Fixed wavelength (0.8726 Å)	Rayonix MarMosaic 225 CCD
E	DLS I03	Synchrotron radiation Tuneable wavelength (0.6-2.48 Å)	DECTRIS PILATUS 6M-F
F	DLS I04	Synchrotron radiation Tuneable wavelength (0.62-2.25 Å)	ADSC QUANTUM 315r CCD

Table 2.6. X-ray sources
Details of all X-ray sources used. Also included are any sources used for studies discussed in **Part II**. DLS, Diamond Light Source, Oxfordshire, UK; ESRF, European Synchrotron Radiation Facility, Grenoble, France; CCD, Charge Coupled Device.

2.8.2. X-ray diffraction screening

All crystals were screened in-house prior to any data collection. Crystals were first harvested in a nylon loop (Hampton Research) and cryoprotected. *Lm*TBCA crystals were passed through a cryoprotectant solution containing 0.2 M $(NH_4)_2HPO_4$, 1.6 M $(NH_4)_2SO_4$, 25 % (v/v) glycerol and placed directly in a stream of gaseous nitrogen. *Lm*TBCC152 crystals were cryoprotected by 40 % PEG 400 (v/v) or a fresh preparation

of mother liquor containing 20 % (v/v) glycerol and flash cooled in liquid nitrogen. A number of native *Lm*TBCC152 crystals were tested using a cryoprotectant mixture supplemented with 0.2-0.5 M NaI.

*Lm*TBCA crystals were tested using sources A and B (**Table 2.6**) while *Lm*TBCC152 crystals were tested using source C. Crystals mounted on a goniostat and cryogenically preserved in a stream of gaseous nitrogen were then exposed to X-rays. Two diffraction images were typically collected at phi angles of 0° and 90° (sources A and B) or 70° (source C) with a rotation of 0.5° per image. Images were inspected, the diffraction quality and resolution assessed and the highest standard crystals progressed for further data collection.

2.8.3. LmTBCA structure solution

SeMet and native *Lm*TBCA data were collected at source D (**Table 2.6**) at a fixed wavelength of 0.8726 Å to approximately 2.3 Å and 1.9 Å resolution, respectively. SeMet data were indexed and integrated using *XDS* (Kabsch, 2010). The symmetry was assessed and intensity measurements scaled by *POINTLESS* and *SCALA* (Evans, 2006) in the *CCP4* crystallographic software suite (Collaborative Computational Project, Number 4, 1994; Winn *et al*., 2011). Selenium sites were identified and density modification performed using *PHENIX* (Adams *et al.*, 2010). Native data were also processed using *XDS* (Kabsch, 2010) followed by *POINTLESS* and *SCALA* (Evans, 2006). *Phaser* (McCoy *et al.*, 2007) was used to perform molecular replacement based on the initial SeMet model. The electron density maps and the structural model were inspected and manipulated in *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010) and

alternated with multiple rounds of refinement performed with *REFMAC5* (Murshudov *et al*., 2011). Solvent molecules were placed based on difference density maps, chemical environment and were consistent with the crystallisation mixture composition. Overall model quality was assessed using *MolProbity* (Chen *et al.*, 2010) and validation tools within *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010).

2.8.4. LmTBCC152

Multiple native *Lm*TBCC152 diffraction datasets were collected using X-ray sources C and E (**Table 2.6**) at various wavelengths (0.9763 Å, 1.5418 Å and 1.7700 Å). Data were collected from two crystals of SeMet-*Lm*TBCC152-L223M at an experimentally derived optimum wavelength of 0.9793 Å (source E, **Table 2.6**). Data collection strategies were followed according to suggestions from *MOSFLM* (Leslie, 2006; Battye *et al.*, 2011), *EDNA* (Diamond Light Source) and *d*TREK* (Rigaku).

Data were indexed and integrated where possible using *XDS* (Kabsch, 2010) or *MOSFLM* (Leslie, 2006; Battye *et al.*, 2011), as stand-alone packages or via the *xia2* pipeline (Winter, 2010), or *HKL-3000* (Otwinowski and Minor, 1997). Scaling was performed using *AIMLESS* (Evans and Murshudov, 2013). The *SHELX* program suite (Sheldrick, 2008; 2010; Grune, 2008; Pape and Schneider, 2004) and *Phaser* (McCoy *et al.*, 2007) were used in attempts to identify selenium or sulfur atom positions using SeMet or native data, respectively. *Phaser* (McCoy *et al.*, 2007), *MOLREP* (Vagin and Teplyakov, 1997) and *PHENIX* (Adams *et al.*, 2010) were used to search for a solution by molecular replacement. Search models were prepared using *CHAINSAW* (Stein, 2008) or by the *PHYRE2* homology model web server (Kelley and Sternberg, 2009). Additionally, any models were inspected in *Coot* (Emsley and Cowtan, 2004; Emsley *et*

al., 2010) and superimposed with published homologues via the secondary-structure matching tool (Krissnel and Henrick, 2004) to guide further model adjustments.

2.8.5. Structural analysis

All structural figures were prepared with *PyMOL* (DeLano, 2002) and annotated with Adobe Creative Suite 5. Other programs or webservers used in structural analyses are cited where appropriate. *Lm*TBCA coordinates and structure factors were deposited in the Protein Data Bank (Bernstein *et al*., 1977) with accession code 4cqi.

PART I

Tubulin-binding cofactors

3. Results and discussion

3.1. Tubulin-binding cofactor A

3.1.1. Protein production

*Tb*TBCA and *Lm*TBCA are short polypeptides composed of 128 and 125 amino acid residues, respectively, with an approximate mass of 14 kDa. Full-length and truncated constructs were generated and soluble protein overproduced in *E. coli*. *Tb*TBCA cloned from gDNA contained an alanine at position 61 in place of the threonine suggested by the annotated database sequence (Logan-Klumpler *et al*., 2012). This may be an artefact introduced during cloning or due to a gDNA strain variant. All truncated *Tb*TBCA constructs also contained this difference as the full-length gene was used as template for amplification. Both TBCAs are tryptophan-deficient and it was necessary to monitor the purification by A_{215} in addition to the standard A_{280}. Two peaks corresponding to full-length *Tb*TBCA were observed on size exclusion chromatography. Elution volumes suggested protein species of approximately 43.8 kDa and 26.9 kDa. The oligomeric state of these species cannot be accurately attributed since the experimental and expected masses are quite different. If taken as accurate, the larger species corresponds to a trimeric solution while the smaller 26.9 kDa mass suggests an intermediate species of between one and two protein chains. Although numerically closer to a dimeric mass, the predicted non-globular shape of TBCA could contribute to a non-standard elution of a monomeric sample. *Tb*TBCA circular dichroism experiments confirmed the presence of a highly α-helical protein solution (**Figure 3.1A**), in agreement with known structural details of TBCA folding as a compact, non-globular molecule. A similar elution profile was seen for *Lm*TBCA and discrepancies in oligomeric state are also reported for other TBCA homologues (Steinbacher, 1999; Guasch *et al.*, 2002; Lu *et al.*, 2010). A trimeric TBCA has not

been reported in the literature, only monomer and dimer (Steinbacher, 1999), and a 1:1 stoichiometric relationship with β-tubulin has been demonstrated using *Hs*TBCA (Llosa *et al*., 1996). It is therefore believed that *Tb* and *Lm*TBCA are mainly monomeric in solution under the conditions discussed here with two or more monomers also interacting to produce a larger unknown oligomer. Both species were progressed to crystallisation experiments.

An additional unusual feature was seen during the purification of *Lm*TBCA. The presence of a double band on SDS-PAGE remained visible after His-tag cleavage and was separated following size exclusion chromatography (**Figure 3.1B**). A mass difference of approximately 1 kDa differentiated the two species, determined by MALDI-TOF mass spectrometry. However, the sample with smaller mass was seen to oligomerise while the sample of expected mass behaved as described above. The exact cause of this remains unclear. Degradation of the *N*-terminus was not indicated, as both species appeared to possess cleavable His-tags. Protein ID studies proved to be inconclusive. As stated previously, all protein species were kept separate and progressed for additional study where possible. SeMet-*Lm*TBCA did not display this unusual behaviour. Monomeric SeMet-*Lm*TBCA was purified for crystallisation and MALDI-TOF mass spectrometry confirmed the full incorporation of selenomethionine in place of the five native methionine residues (**Figure 3.1C**).

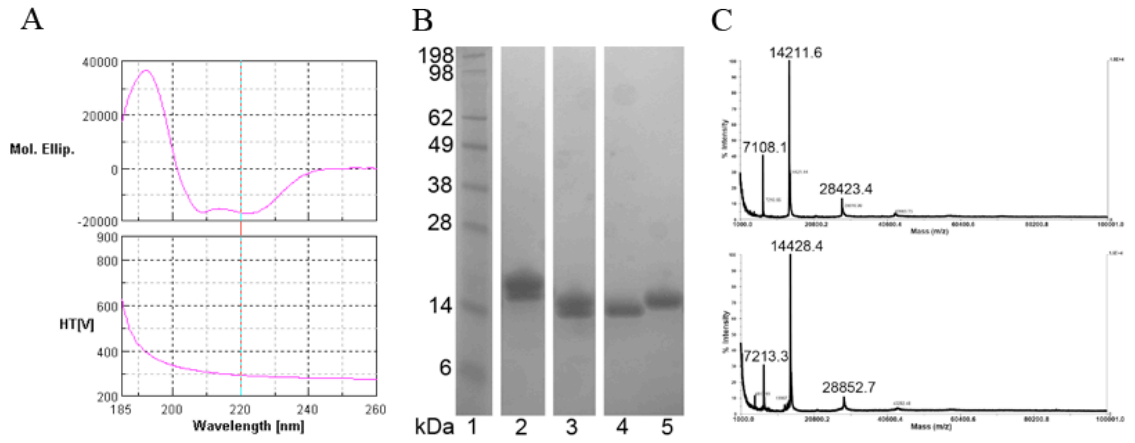


Figure 3.1. TBCA purification
(A) *Tb*TBCA circular dichroism far UV spectra. **(B)** *Lm*TBCA SDS-PAGE. Lane 1, molecular weight standards; lane 2, His-tagged *Lm*TBCA; lane 3, *Lm*TBCA following His-tag cleavage; lane 4, *Lm*TBCA multimeric species; lane 5, *Lm*TBCA monomeric species. **(C)** MALDI-TOF spectra indicating a mass difference of 218.6 Da between native *Lm*TBCA (upper panel) and SeMet-*Lm*TBCA (lower panel). This agrees with a theoretical mass difference of 234.5 Da for the incorporation of five selenomethionine residues.

3.1.2. LmTBCA crystallisation and structure solution

Several promising *Lm*TBCA crystallisation conditions were identified in initial sparse-matrix sitting-drop vapour-diffusion screens. Only the monomeric solution of expected full-length mass produced crystal hits. The predominant precipitating agent was $(NH_4)_2SO_4$ and diffraction quality native and SeMet crystals were grown under conditions described in **2.7.2**. Maximum native crystal size was achieved after approximately 3-5 days with dimensions of up to 1.0 x 0.2 x 0.2 mm. Fewer, smaller crystals were obtained of SeMet-*Lm*TBCA (**Figure 3.2A**).

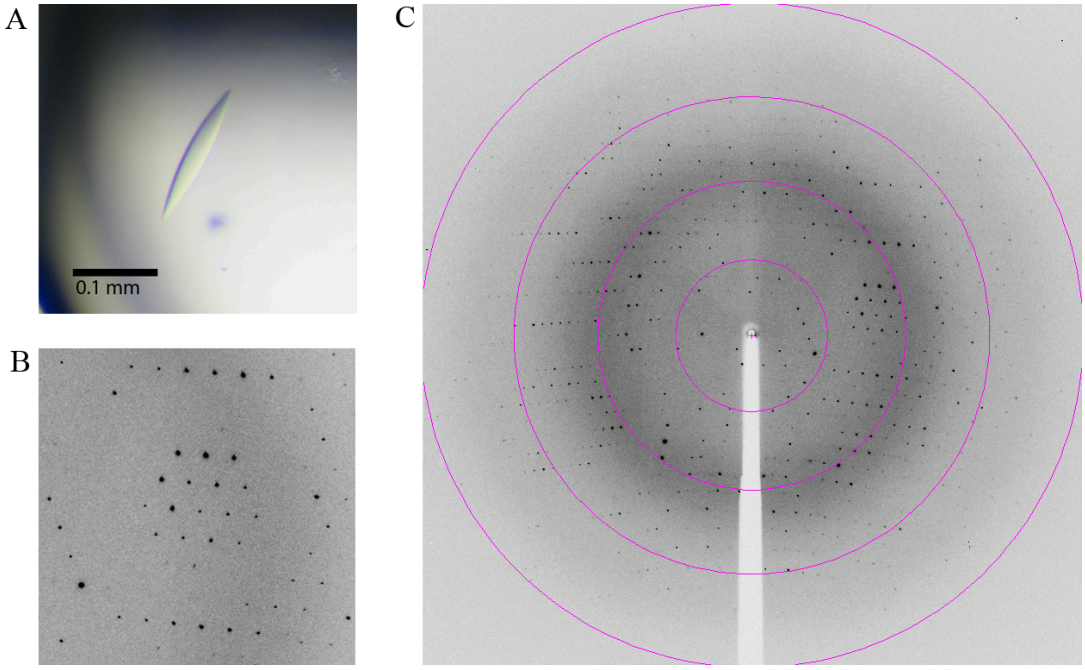


Figure 3.2. *Lm*TBCA crystal and diffraction
(A) A single SeMet-*Lm*TBCA crystal. **(B)** and **(C)** show X-ray diffraction images from a native *Lm*TBCA crystal. The outermost magenta resolution ring represents 1.8Å resolution.

Data collection and processing details are provided in **2.8.3**. Molecular replacement using homologue structures as a search model were unsuccessful. Instead, phases were calculated experimentally by SAD through the production of a SeMet derivative. Native and SeMet diffraction data extended to 1.9 Å and 2.3 Å, respectively, and an example diffraction image is shown in **Figure 3.2B-C**. Crystals were isomorphous in space group $P3_121$ with unit cell parameters of $a = b = 76.8$ Å, $c = 39.4$-39.5 Å, $\alpha = \beta = 90°$ and $\gamma = 120°$. A Matthews coefficient (V_M) of 2.36 $Å^3$ Da^{-1} and a predicted solvent content of 50 % (Matthews, 1968; Collaborative Computational Project, Number 4, 1994) indicated the presence of one *Lm*TBCA molecule per asymmetric unit. Two consecutive SeMet residues, equivalent to Met77 and Met78 in native *Lm*TBCA, were useful in validating selenium atom positions through inspection of initial electron density maps. Detailed crystallographic refinement statistics are presented in **Table 3.1**.

	SeMet *Lm*TBCA	Native *Lm*TBCA
Resolution range (Å)	39.5-2.3 (2.4-2.3)	39.4-1.9 (2.0-1.9)
Space Group	$P3_121$	$P3_121$
Unit cell parameters	$a=b=$ 76.8, $c=$ 39.5 Å $\alpha=\beta=$ 90, $\gamma=$120°	$a=b=$ 76.8, $c=$ 39.4 Å $\alpha=\beta=$ 90, $\gamma=$120°
Wavelength (Å)	0.8726	0.8726
No. Reflections	73555 (10683)	130402 (19204)
No. Unique Reflections	6205 (878)	10821 (1549)
R_{merge} [a] (%)	13.8 (61.3)	7.3 (44.8)
R_{pim} [b] (%)	5.9 (26.4)	2.2 (13.2)
Completeness (%)	100 (100)	100 (100)
Mean *I/σ(I)*	19.5 (6.0)	24.9 (6.8)
Redundancy	11.9 (12.2)	12.1 (12.4)
Wilson *B* factor (Å^2)	35.3	20.9
R_{work} [c] (%)	-	18.5
R_{free} [d] (%)	-	22.8
R.m.s.d bonds (Å)	-	0.0134
R.m.s.d angles (°)	-	1.586
Total protein residues	-	107
Total protein atoms	-	855
Average protein *B* factor (Å^2)	-	27.3
DPI [e] (Å)	-	0.138
Ramachandran plot:		
Favoured (%)	-	96.19
Allowed (%)	-	2.86
Outliers (%)	-	0.95
Additional groups:		
Solvent (No./Average *B* (Å^2))	-	90/35.0
Sulfate (No./Average *B* (Å^2))	-	1/33.4
Glycerol (No./Average *B* (Å^2))	-	1/39.6

Table 3.1. *Lm*TBCA data collection and refinement statistics.
Values in parentheses correspond to the highest resolution shell. [a.] $R_{merge} = \sum h \sum i ||(h,i) - \langle I(h) \rangle| \sum h \sum i\ I(h,i)$; where $I(h,i)$ is the intensity of the ith measurement of reflection h and $\langle I(h) \rangle$ is the mean value of $I(h,i)$ for all i measurements. [b.] R_{pim}, precision-indicating merging R-factor, is R_{merge} adjusted by a factor of $\sqrt{(1/n-1)}$ where n is the number of times a given reflection is observed. [c.] $R_{work} = \sum hkl||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor amplitude and the F_c is the structure-factor amplitude calculated from the model. [d.] R_{free} is calculated with a subset of data that are excluded from refinement calculations (5 %) using the same method as for R_{merge}. [e.] DPI, diffraction-component precision index (Cruickshank, 1999).

3.1.3. LmTBCA structure

*Lm*TBCA is composed of three α-helices, named here as α1, α2 and α3 progressing from the *N*- to *C*-terminus. The helices are in an antiparallel arrangement linked by two short loops at opposite ends. The elongated bundle has a maximum length of

approximately 59 Å and is 23 Å at its widest. α1 and α2 have similar lengths of up to 56 Å and 52 Å, respectively. α3 is less than half the length of the full monomer at only 21 Å. The model contains residues Glu19 to Ser125. The first 18 residues could not be modelled due to the absence of corresponding electron density. Ten full helical turns make up each of the first two helices and α3 has only four turns. α1 extends from residue Asn22 to Ser60, α2 from Asp63 to Lys99 and α3 from Glu108 to Ala122. The loop linker segments are poorly ordered with respect to the well-organised helices. A bend in α1 and α2 of approximately 20-25° gives an overall curved appearance to the molecule. At the same level, both helices appear to twist, allowing for the incorporation of α3 (**Figure 3.3A**).

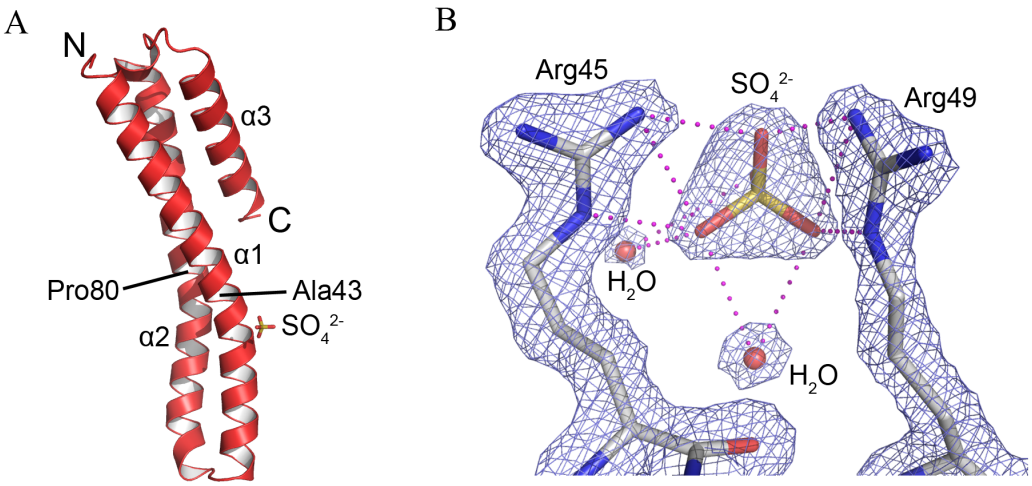


Figure 3.3. Sulfate bound to the surface of *Lm*TBCA
(A) Overall view of a single *Lm*TBCA molecule, represented as red ribbons. Ala43 and Pro80 are labelled as approximate positions of the bends in α1 and α2. Sulfate is bound to the surface of α1. **(B)** A closer view of the surface sulfate coordinating with Arg45 and Arg49 is shown as gold (S) and red (O) sticks. The arginine residues are also coloured by element (C, grey; N, blue) and two water molecules are depicted as red spheres. Blue mesh represents electron density ($2F_o$-F_c contoured at 2 σ) and magenta dotted lines indicate potential hydrogen bonds.

The surface sulfate seen in **Figures 3.3A and 3.3B** was modelled in a large unoccupied tetrahedral mass of electron density located between two arginine residues (Arg45 and Arg49). Crystals were grown in the presence of ammonium sulfate and the hydrogen

bonding pattern seen is suggestive of a tightly bound molecule. This coordination is at the level of the bends in α1 and α2 and may be a consequence of the surface curvature or it is possible that the interaction played a role in inducing the helix direction change. Such a feature of crystallisation may also be indicative of a natural interaction site, where a similar entity may bind at this position *in vivo*. A glycerol molecule, used as the cryoprotectant, is also bound to the protein, linking Glu41 on α1 and Asp119 on α3 (not shown).

Despite the growth of crystals using the monomeric sample obtained from size exclusion chromatography and the presence of one molecule in the asymmetric unit, evidence of a covalent dimer is apparent. A single disulfide link is observed between Cys58 at the *C*-terminal end of α1 and the same residue of a symmetry-related *Lm*TBCA molecule, generated using the crystallographic symmetry operation - x, - x + y, - z + 1/3. Cα atoms are separated by a distance of 7.6 Å and sulfur atoms are 2.2 Å apart, an accepted disulfide bond length (**Figure 3.4**; Richardson, 1981). Electron density agrees with the presence of a covalent bond at this position. Attempts were made to corroborate this observation by subjecting dissolved crystals to SDS-PAGE and MALDI-TOF MS. The reduced environment of SDS-PAGE sample preparation should denature a protein sample into its monomeric form. Here, a shift in mass was seen, suggestive of a strong linkage between monomers with only partial denaturation. However, due to the location of the bond at the tip of an elongated cylindrical structure, the extreme lengthening of the paired molecules resulted in an unusual gel profile (not shown). It was therefore difficult to confirm an accurate molecular mass of the crystallised species. Unfortunately, mass spectrometry analysis did not provide any additional knowledge on this matter nor on the presence of the initial 18 amino acids.

Residues immediately surrounding Cys58 display higher *B* factors compared to the rest of the molecule due to the poorly ordered side chain atoms. Mean *B* factor of Asp55 to Pro66 is 56.0 $Å^2$ versus 27.3 $Å^2$ for the protein as a whole. This is indicative of disorder or flexibility with the link at Cys58 stabilising the placement of the α1-α2 loop. A cysteine at this position is not conserved throughout TBCAs suggesting it is not a critical functional residue. Nevertheless, the formation of this Cys58-Cys58 disulfide bond may have aided in the crystallisation of *Lm*TBCA. Purification and crystallisation solutions were not supplemented with reducing agents. It cannot be ruled out that the linked species was present during purification in low quantity although crystals were only obtained using sample originally characterised as a monomer.

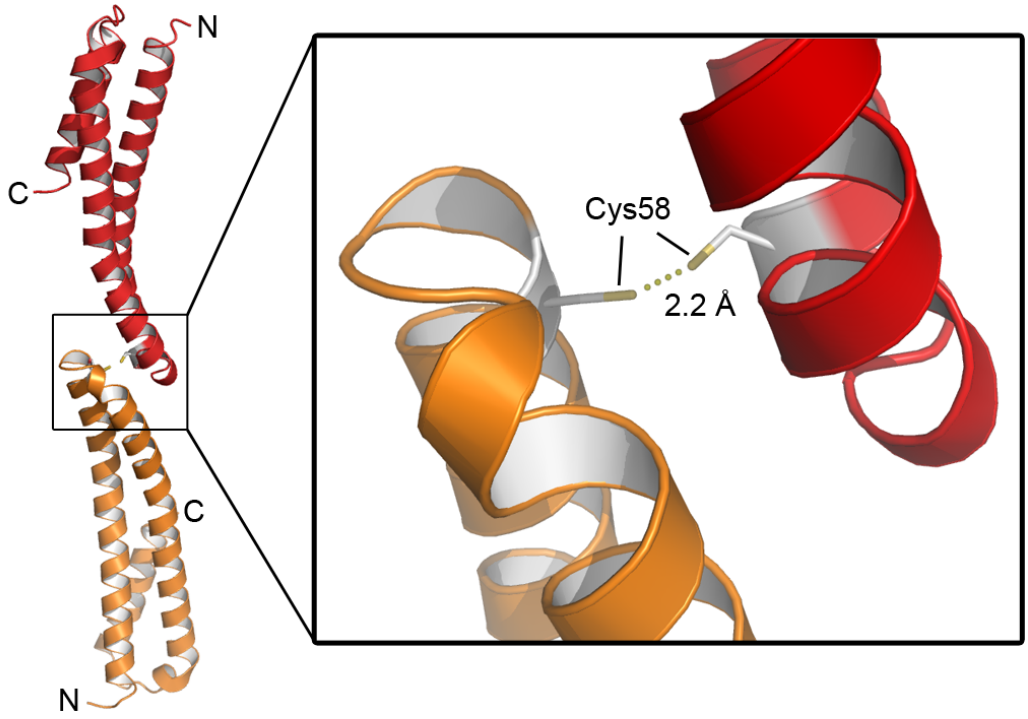


Figure 3.4. *Lm*TBCA disulfide bond
*Lm*TBCA monomer (red) and a symmetry related molecule (orange) are linked by a disulfide bond between Cys58 residues with side chain atoms shown as sticks (C, grey; S gold).

A network of hydrogen bonds contributes to maintaining the overall protein shape. The individual helices contain a classical 4_{13} hydrogen bonding arrangement while inter-

helical hydrogen bonds align the helices with respect to each other. The bend in α2 mentioned above is caused by the absence of a backbone hydrogen bond donor from Pro80. This lack of typical α-helical hydrogen bond between Pro80 and Gln76, replaced with only a weak hydrogen bond between Pro80 Cδ and Gln76 O, disrupts the standard α-helix organisation (**Figure 3.5**). A number of hydrogen bonds close to Pro80 appear stretched but remain within an accepted length range. For example, the distance from the amide group of Val79 to the carbonyl of Ala75 is 3.5 Å while other neighbouring hydrogen bonds are between approximately 2.9 and 3.2 Å. Notable links between α1 and α2 are distributed along the length of the molecule, including salt bridges Lys29-Asp92 and Asp39-Arg85 and a hydrogen bond at Glu50-Gln68. There is, however, little direct association between α3 and the two longer helices.

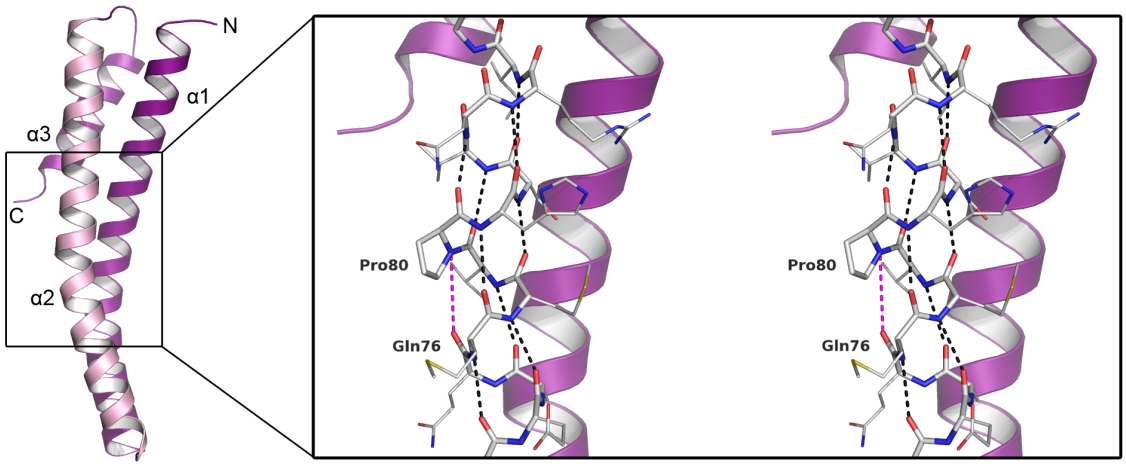


Figure 3.5. Backbone hydrogen bonds of *Lm*TBCA α2
The area highlighted shows a stereo view of helices α1 and α3 as purple ribbons. α2 is depicted as pale pink ribbon or as sticks coloured by element (C, grey; O, red; N, blue; S, gold). Black dashed lines represent standard α-helical hydrogen bonds. Pro80 disrupts this bonding pattern allowing the helix to bend. The magenta dashed line between Pro80 and Gln76, which does not represent a hydrogen bond, has a length of 4.20 Å.

3.1.4. TBCA structure comparison and potential for tubulin interaction

Overall, published TBCA structures (*Arabadopsis thaliana* TBCA (*At*TBCA) PDB 3mxz, Lu *et al*., 2010; *Homo sapiens* TBCA (*Hs*TBCA) PDB 1h7c, Guasch *et al*., 2002; *Saccaromyces cerevisiae* Rbl2p (*Sc*Rbl2p) PDB 1qsd, Steinbacher, 1999) adopt the same general conformation as *Lm*TBCA. Three helices of approximately equal length are connected by short loops. However, the surface curvature is not identical. A proline at the same position as Pro80 is strictly conserved in these species but the resultant bend is most pronounced in *Lm*TBCA. The distortion of α1 is also unique to *Lm*TBCA. *Sc*Rbl2p helices α1 and α2 are almost linear and it is the crossover of the two lengths that mimics an overall curved appearance. The apparent effect of the proline residue in this case is counteracted by an increased hydrogen bond length at what would have become the concave helix surface, producing a less prominent direction change. *Hs*TBCA has an unusual conformation with α2 kinked in the opposite direction to that seen in *Lm*TBCA. The conserved proline then simply returns the helix to its original course. Of the published models, *At*TBCA appears the most structurally similar to *Lm*TBCA with a root mean square deviation (r.m.s.d.) of 1.8 Å when 92 Cα atoms are aligned. Sequence similarity is also the greatest at about 27 % compared to 26 % and 19 % for *Sc*Rbl2p and *Hs*TBCA, respectively.

Despite a visual resemblance to *Lm*TBCA, the crystal structure of *Lm*TBCA could not be determined via molecular replacement approaches based on any of the published homologous structural models. Elongated, exclusively α-helical proteins have proved challenging subjects for molecular replacement in previous studies (Sundaramoorthy *et al*., 2008) including attempts to use *Sc*Rbl2p as a model to solve the structure of

*Hs*TBCA (Guasch *et al*., 2002). The inherent difficulties can be attributed to ambiguities in crystallographic rotation and translation, particularly in the direction of the helices and the rotation about the helical axes.

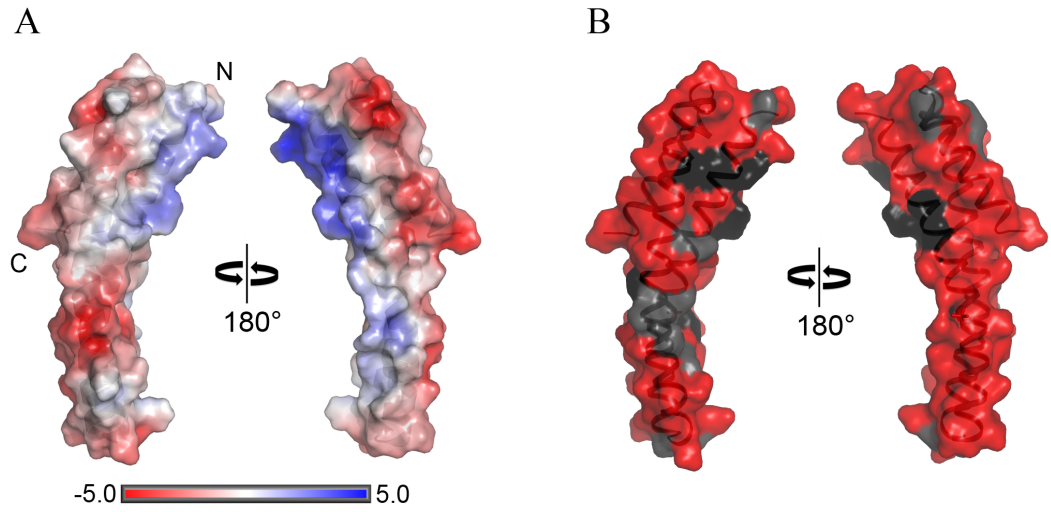


Figure 3.6. *Lm*TBCA electrostatic potential
(A) A van der Waals surface representation of *Lm*TBCA coloured according to electrostatic protein contact potential (-5 kTe^{-1}, red to 5 kTe^{-1}, blue) created using *PDB2PQR* (Dolinsky *et al*., 2004) and *APBS* (Baker *et al*., 2001). **(B)** The semi-transparent surface of *Lm*TBCA coloured by sequence identity to known homologous structures (*Hs*TBCA, *Sc*Rbl2p and *At*TBCA). Amino acids only present in *Lm*TBCA are coloured red. Increasing similarity is represented by a darkening grey scale with residues identical in all four species shown in black.

*Lm*TBCA α1 and α2 each consist of an abundance of hydrophobic amino acids with side chains mainly directed towards the core of the tri-helical bundle. Surface electrostatic potential mapping indicates areas of localised polarity distributed over the protein. The concave exterior, when represented as a van der Waals surface projection, is significantly more positively charged in comparison with the rest of the molecule (**Figure 3.6A**) which has an overall calculated *p*I of 5.2 (*ProtParam*; Gasteiger *et al*., 2005). The concentration of positive charge is attributed to the solvent-facing arginine and lysine residues of α1. It is also the area of greatest sequence identity with structural homologues (**Figure 3.6B**). A helical wheel schematic shows the distribution of amino acids along α1 and 80 % of positive residues are localised along the surface described

(**Figure 3.7**). The TBCA binding partner, β-tubulin, is a highly conserved polypeptide (Sullivan and Cleveland, 1986) and is largely negatively charged with an overall pI of 4.6-4.7 (*ProtParam*; Gasteiger *et al*., 2005). Additionally, both α- and β-tubulin *C*-terminal tails are negatively charged and are known to bind to other MAPs and cationic molecules (Cross *et al*., 1991; Lefevre *et al*., 2011). The concave surface of *Lm*TBCA might therefore present a favourable site for β-tubulin interaction.

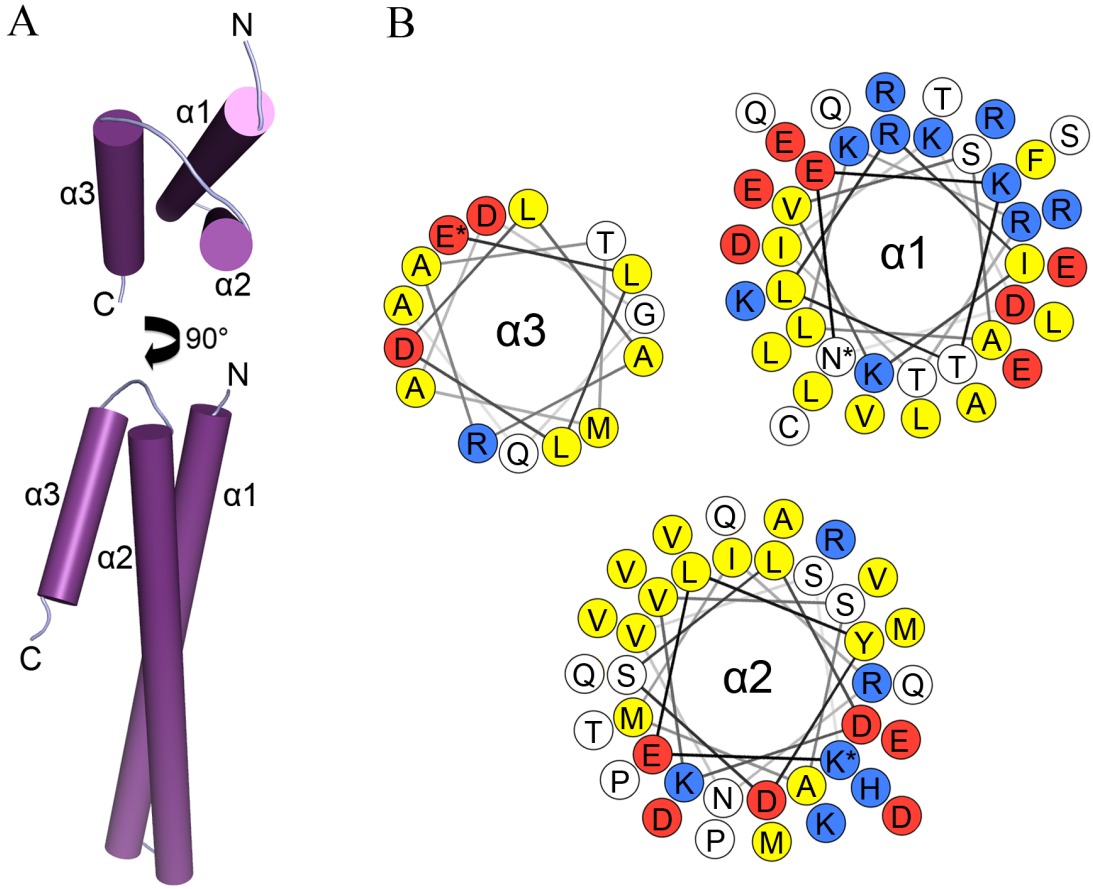


Figure 3.7. Arrangement of *Lm*TBCA helices
(A) *Lm*TBCA helices α1, α2 and α3 are shown as violet cylinders linked by pale blue loops viewed from above (top) and following rotation by 90° (bottom). **(B)** A helical wheel projection showing the approximate location of the backbone of all helix residues. α1 and α3 amino acids are arranged from *N*- to *C*-terminus in a clockwise orientation beginning at Asn22 and Glu108, respectively (marked with *). α2 reads clockwise starting at Lys99 from the *C*- to *N*-terminus. Amino acids with positively charged side-chains are represented as blue circles and negatively charged residues are coloured red. Hydrophobic side-chains are yellow while all others are uncoloured. The three helices are arranged in a similar orientation to **(A)**. It should be noted that due to the curvature of α1 and α2, the distribution depicted above is for guidance only and does not infer specific inter-helical interactions. The helical wheel diagrams are adaptations of those produced using the tool by Armstrong and Zidovetzki (2009), available at http://rzlab.ucr.edu/scripts/wheel/wheel.cgi.

The most apparent variation in TBCA protein sequences is the *N*-terminal extension observed in trypanosomatids (**Figure 3.8A**). This region has negligible predicted secondary structure and the electron density of *Lm*TBCA provides no evidence of an ordered structure. It is possible that the additional 18 residues may therefore be of little functional value and the *Lm*TBCA crystal structure represents a functional unit, directly comparable to homologues lacking the extended *N*-terminus.

*Lm*TBCA lacks a significant hydrophobic surface site characteristic of proteins involved in interactions with an unfolded or partially folded partner such as those seen in GroEL of the archetypal bacterial protein folding system (Fenton *et al*., 1994) and members of the hsp70 (heat-shock protein of 70 kDa mass) molecular chaperone family (Flynn *et al*., 1991). This is in agreement with the hypothesis that β-tubulin is already in a folded state when initially presented to the tubulin-binding cofactors and the cofactors do not contribute to tubulin folding. Indeed it has been shown that TBCA does not recognise denatured β-tubulin (Archer *et al*., 1998). Due to the shortage of clusters of exposed hydrophobic residues on *Lm*TBCA, it could be argued that the binding event occurs at a site different from its polymerisation partner. When β-tubulin sequences from the same organisms as those with published TBCA structures are aligned, sequences are between 70 % (*Sc*β-tubulin) and 85 % (*Hs* and *At*β-tubulin) identical to *Lm*β-tubulin. This similarity between polypeptides strongly suggests that the location of binding events involving β-tubulin is conserved. Interactions with the globular surface should then require an equivalent level of homology between partner molecule binding sites. Since the overall sequence identity between TBCAs is much lower than that of β-tubulin, localised regions of high conservation offer the greatest interaction potential. β-tubulin

sequences diverge mainly at the *C*-terminal tail, a site implicated in many other microtubule-related interactions (for example, Cross *et al*., 1991 and Lefevre *et al*., 2011) and, without detailed structural knowledge, cannot be ruled out as the site of interaction with TBCA.

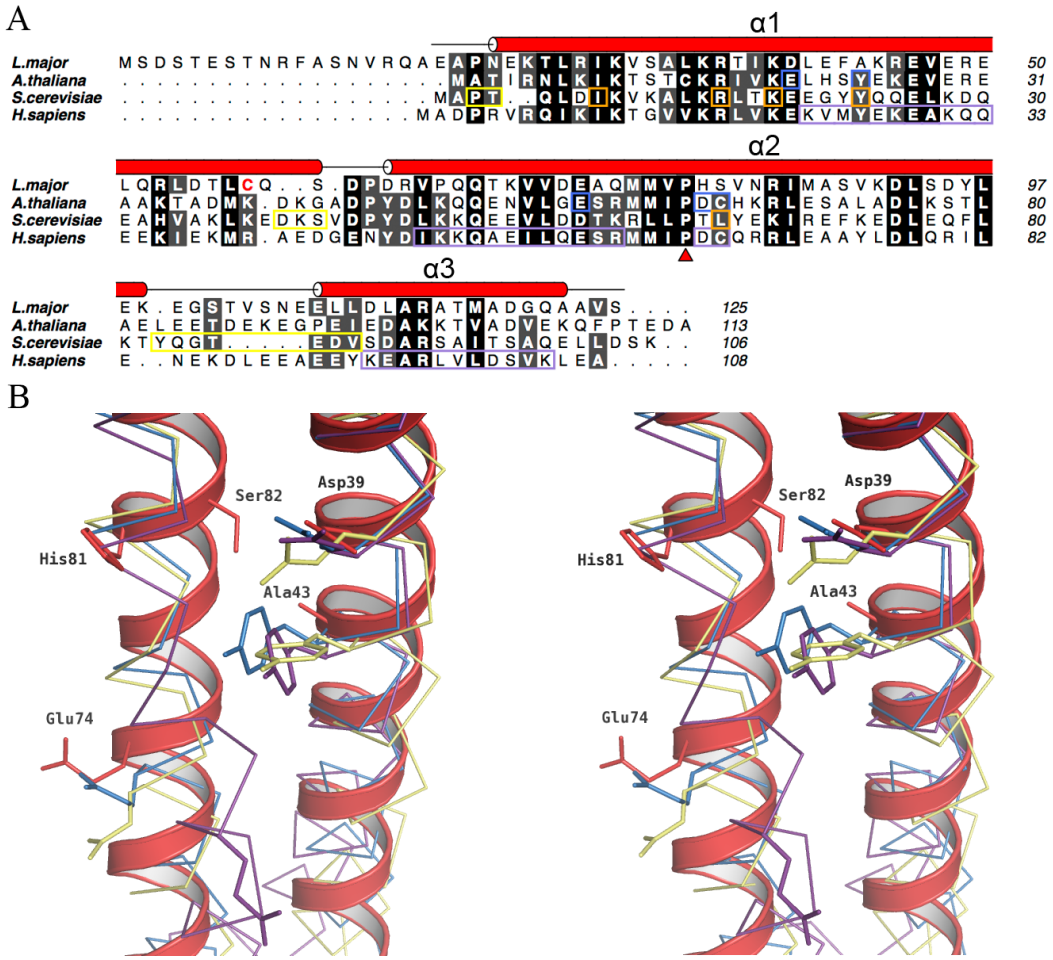


Figure 3.8. TBCA sequence and structure comparison
(A) Structure-based sequence alignment of *Lm*TBCA and published homologues *A. thaliana* TBCA (3mxz), *S. cerevisiae* Rbl2p (1qsd) and *H. sapiens* TBCA (1h7c). Sequence similarity is indicated by white text on a grey-scaled background. Residues highlighted in coloured boxes (blue, yellow and purple) are implicated in binding β-tubulin. Amino acids in orange boxes are thought to affect β-tubulin binding but are also located at the *Sc*Rbl2p homodimer interface. Cys58 is shown as red text and the location of the conserved proline (Pro80) is marked with a red triangle. Sequences were aligned with *MUSCLE* (Edgar, 2004) and the figure prepared using *ALINE* (Bond and Schüttelkopf, 2009). **(B)** Stereoview image of *Lm*TBCA helices α1 and α2 (red ribbon) with Cα backbone traces of *At*TBCA (blue), *Sc*Rbl2p (yellow) and *Hs*TBCA (purple). Selected residues are shown as sticks of the same colours, labelled according to *Lm*TBCA sequence and numbering. Residues at the positions of Asp39, Ala43 and Glu74 are critical for β-tubulin binding in *A.thaliana*. His81 and Ser82 are also thought to play a functional role. In this view, *Hs*TBCA Glu59 (aligns with Glu74 labelled above) appears distant from its sister residues but is located at the same position on α2. Alignment was performed using SSM based on Cα atoms of complete PDB chains.

Several studies have presented theories of how TBCA interacts with β-tubulin but there remains no consensus on the true model, if indeed there is a single mode of action. Peptide mapping and competition experiments suggest β-tubulin interacts with broad regions on all three helices of *Hs*TBCA (Guasch *et al*., 2002). α3 was not considered essential but binding activity was diminished when it was removed. Two specific amino acid mutations appeared to influence binding, D66E and C67S. *Hs*TBCA Asp66 and Cys67 correspond to His81 and Ser82 in *Lm*TBCA. The mutation to glutamic acid at the position of His81 reduced binding activity but the mutation to serine at the position of *Lm*TBCA Ser82 increased the level of *Hs*TBCA-β-tubulin complex detected. These residues are not strictly conserved but their effect on function suggests a role in the binding event. Immediately preceding His81 and Ser82 is the highly conserved Pro80 that facilitates the distortion of α2 seen in all known TBCA structures. Although not directly involved in molecular contacts with β-tubulin, its conservation suggests the importance of the cofactor's overall shape.

*At*TBCA mutagenesis and co-immunoprecipitation studies revealed residues Glu20, Tyr24 and Glu57 were critical for binding β-tubulin (Lu *et al*., 2010). According to sequence and structural alignments, these correspond to Asp39, Ala43 and Glu74 in *Lm*TBCA while only Glu57 differs in *Sc*Rbl2p, replaced conservatively by Asp57, and all three residues remain identical in *Hs*TBCA. Mutation of each of these individually to alanine in *At*TBCA resulted in no detection of a TBCA-β-tubulin complex. Located on the *C*-terminal half of α1 and *N*-terminal half of α2, this is the region of greatest variation between these four structures when the whole molecule is overlaid using secondary structure matching procedures (Krissinel and Henrick, 2004). If a plane were

to dissect the protein at the level of the *C*-terminus, mean Cα deviations are 0.8 Å more in the lower portion than in the upper, when viewed in the orientation of **Figure 3.3A**. The negatively charged side chain atoms of Asp39 and Glu74 are exposed on the surface of the helix bundle (**Figure 3.8B**). The hydrophobic Ala43, however, lies more buried and the Cβ atom extends towards α2 unlike the large side-group of *At*TBCA Tyr24 which projects into solvent on the same surface as Glu20. The ability of a single Tyr-Ala mutation to eliminate binding does indicate a critical functional role but this is contradicted by the presence of alanine in native *Lm*TBCA. Perhaps *Lm*TBCA displays lower affinity for β-tubulin as a result or there may be additional or alternative contributions made by amino acids elsewhere. For example, the distinct curvature of *Lm*TBCA could present residues along the entire length of the helices, including the positively charged region of α1 discussed, towards the binding partner.

Conflicting with the potential binding pattern described, computational docking onto the surface of β-tubulin suggests that the homodimeric *Sc*Rbl2p interacts via the short loops rather than the helices (You *et al*., 2004). Helices α1 and α2 instead form the dimer interface and a number of the residues discussed above make contacts with or are buried by the second molecule (Steinbacher, 1999). Both monomer subunits are predicted to form links with β-tubulin so perhaps these residues are equally important for function in *Sc*Rbl2p where a dimer must first be formed. The same study identifies the site of β-tubulin most likely to participate as an area that overlaps with the footprint of the αβ-tubulin dimerisation site (You *et al*., 2004). This corresponds with the ability of a TBCA-β-tubulin complex to prohibit premature αβ-tubulin dimer formation since α-tubulin and the cofactor cannot bind to β-tubulin simultaneously (Llosa *et al*., 1996;

Fanarraga *et al*., 1999; Abruzzi *et al*., 2002). Alternatively, interaction at an allosteric site may require TBCA to induce conformational change to the β-tubulin dimerisation surface but there is currently no data to support this.

Of the known TBCA structures, only *Sc*Rbl2p is reported to form a homodimeric species. Although the dimer interface is notably hydrophilic (Steinbacher, 1999), the arrangement of the same few hydrophobic residues forming inter-molecular contacts is absent in *Lm*TBCA, *At*TBCA and *Hs*TBCA. *Lm*TBCA displayed some oligomerisation during purification but was primarily monomeric in solution. In the crystal structure, contacts between molecules are distributed around the surface and not clustered to resemble a dimerisation interface. The covalent disulfide link made by Cys58 is completely different to the configuration of the *Sc*Rbl2p dimer. A 1:1 stoichiometry of TBCA:β-tubulin is also described elsewhere (Llosa *et al*., 1996). It is therefore unlikely that *Lm*TBCA adopts the same functional conformation reported for *Sc*Rbl2p but the potential for this to occur cannot be dismissed without further evidence.

Several protein structures possess a fold similar to *Lm*TBCA. As expected, lead results from a search for structural relatives using *PDBeFold* (Krissinel and Henrick, 2004) and the *Dali* server (Holm and Rosenström, 2010) were *At*TBCA, *Sc*Rbl2p and *Hs*TBCA. Details of these and selected subsequent structural neighbours are provided in **Table 3.2**. Inspection of a number of molecules found by *PDBeFold* suggests that they were identified by the incidence of one or more helices of equivalent length to *Lm*TBCA, sometimes termed 'α-ten modules' due to the presence of ten α-helical turns. A precise functional relationship, if any, is yet to be established but the ability to form protein-protein complexes appears to be a common theme. Interestingly, this approach

highlighted a number of distant structural neighbours that contain a three-helix bundle motif with the function of acting as molecular chaperones. Indeed, two such proteins are actinin, a molecule known to participate in binding to actin, the primary component of cytoskeletal microfilaments (Sjöblom *et al*., 2008) and spectrin, also related to the cytoskeleton in eukaryotes (Bennet and Baines, 2001). Subunits of the prefoldin protein-folding complex were also listed. This fold is therefore acknowledged as a binding partner of important cellular elements in systems both outside of and analogous to microtubule dynamics. However, there remains no single conclusive binding pattern.

PDB	Description	No. aligned residues	RMSD (Å)	*Z*-score	Publication
3mxz	*At*TBCA	103 (97)	2.7 (2.04)	13.5 (5.3)	Lu *et al*., 2010
1qsd	*Sc*Rbl2p	97 (96)	2.2 (2.76)	12.7 (4.1)	Steinbacher, 1999
1h7c	*Hs*TBCA	101 (90)	3.6 (2.32)	12.4 (4.2)	Guasch *et al*., 2002
1s94	Syntaxin Habc domain	(86)	(2.27)	(3.7)	Bracher and Weissenhorn, 2004
4i0x	ESAT-6-like protein	(73)	(2.91)	(2.7)	N/A
3k6c	Uncharacterised	(65)	(2.39)	(2.8)	N/A
1g73	SMAC	106	3.6	10.1	Wu *et al*., 2000
1oxz	ARF binding protein	84	2.3	9.3	Zhu *et al*., 2003
3fb2	*Hs*α-spectrin	95	3.2	8.6	N/A
1u5p	*Gg*α-spectrin	95	3.5	8.4	Kusunoki *et al*., 2004
3edv	*Hs*β-spectrin	95	3.6	8.3	Davis *et al*., 2009
2zqm	Prefoldin (β)	85	3.3	8	Kida *et al*., 2008
2zdi	Prefoldin (β)	83	2.7	8	Ohtaki *et al*., 2008
1fxk	Prefoldin	85	2.5	8	Siegert *et al*., 2000
1wlx	*Hs*α-actinin4	97	4.5	5.8	N/A
1hci	*Hs*α-actinin2	96	3.8	4.6	Ylänne *et al*., 2001

Table 3.2. *Lm*TBCA structure similarity search results
Selected results from structure similarity searches based on *Lm*TBCA chain A. Results from *PDBeFold* (Krissinel and Henrick, 2004) are shown in parantheses, all others were obtained from the *Dali*-server (Holm and Rosenström, 2010). *Z*-scores are a measure of the statistical significance of each match and are defined in the associated publications. A *Z*-score provided by *Dali* of less than 2 is considered spurious. Sequence identity of the examples shown was calculated as between 29 and 3 %. SMAC, Second mitochondira-derived activator of caspases; Habc, Helix a, b and c; ESAT, Early secreted antigenic target protein; ARF, Adenosine diphosphate-ribosylation factor; *Gg*, *Gallus gallus*; N/A, no associated publication.

3.2. Tubulin-binding cofactor C

3.2.1. Protein production

Full-length *Tb*TBCC and *Lm*TBCC were recombinantly expressed and purified in soluble form. The isolated monomeric samples, as suggested by size exclusion chromatography, were subject to crystallisation screening but no suitable conditions were identified. *Tb*TBCC circular dichroism analysis indicated the presence of correctly folded protein and DSF studies resulted in melting curves also suggestive of a folded species. No buffers in the DSF thermal stability buffer screen provided a significant positive shift in protein T_m. Maximum T_m values were 53.2°C for *Tb*TBCC and 50.8°C for *Lm*TBCC, an increase of only 0.7°C and 1.2°C from standard buffers, respectively. Nevertheless, further crystallisation experiments were prepared following exchange to those buffers showing the highest T_m and no changes to crystallisability were observed.

3.2.2. Degradation and truncation

Within 7 days of purification, SDS-PAGE and MALDI-TOF analyses indicated protein products at a molecular mass smaller than expected for full-length TBCC (both *Tb* and *Lm*). These products were not apparent during or immediately following purification and were believed to be evidence of degradation or proteolytic activity. Incubation of mass-verified full-length *Tb*TBCC with chymotrypsin resulted in a similar distribution of protein fragments (**Figure 3.9A**). Peptide identification following further proteolysis by trypsin of the two unknown protein bands matched those present in *Tb*TBCC (**Figure 3.9B**), corresponding to an *N*-terminal fragment of approximately 12 kDa and a larger 20 kDa *C*-terminal region. The mass of these fragments also coincides with

secondary structure and globularity predictions that at least two distinct protein domains may be present.

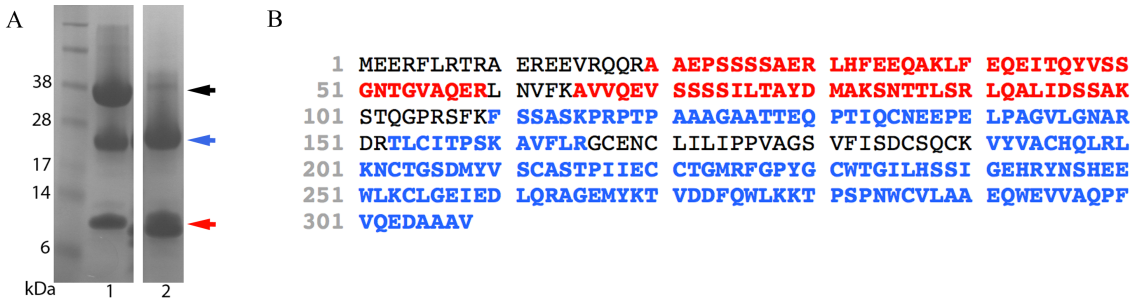


Figure 3.9. *Tb*TBCC proteolysis
(A) SDS-PAGE of *Tb*TBCC. Lane 1 shows untreated *Tb*TBCC 2 days after initial purification. Lane 2 is *Tb*TBCC after addition of chymotrypsin. The approximate location of full-length *Tb*TBCC (black arrow), degradation product I (blue arrow) and degradation product II (red arrow) are indicated. **(B)** Mass spectrometry protein ID results using samples I and II from SDS-PAGE. Matched peptide fragments of degradation products I and II are shown as blue and red text, respectively.

As there was no success in obtaining crystals of full length TBCC, this degradation event was exploited in order to isolate a more stable protein construct that was potentially more likely to crystallise. Treatment with chymotrypsin had already been shown to result in products similar in size to natural degradation and was subsequently used to accelerate the process. The addition of chymotrypsin directly to TBCC prior to crystallisation screening produced no crystals. Size exclusion chromatography following limited proteolysis was used to separate the stable protein domains. It was possible to isolate only the larger *C*-terminal portion in this way. However, crystal growth remained elusive.

Several expression constructs were generated (details in **Table 2.3**), designed with the aid of proteolysis experiments in combination with several sequence-based prediction tools (described in **2.2.3**). *Tb*TBCC constructs 1-107, 1-112 and 138-308 and *Lm*TBCC construct 152-335 were purified and screened for suitable crystallisation conditions.

However, other truncated constructs produced insufficient soluble material or were of unsatisfactory purity to attempt crystallisation.

3.2.3. LmTBCC152 crystallisation

*Lm*TBCC152 was the only truncated construct from *L. major* that resulted in soluble recombinant material purified to an acceptable level. A high yield of approximately 10-20 mg final purified product was obtained from 1 L *E. coli* cell culture. Initial crystal hits were observed in a screen utilising polyethylene glycol (PEG) of different mass as the precipitating agent. Solutions of 0.1 M MES pH 6.5 containing 25 % PEG 2000 MME, PEG 4000 or PEG 6000 mixed with *Lm*TBCC152 at 7.3 mg mL^{-1} in 20 mM Tris-HCl pH 7.5, 100 mM NaCl produced crystalline material as shown in **Figure 3.10.**

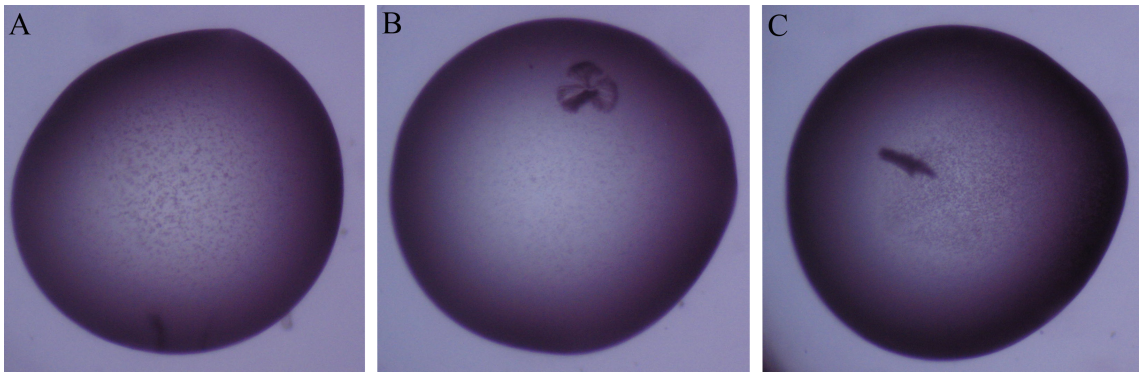


Figure 3.10. Initial *Lm*TBCC152 crystals
Crystals observed in 0.2 μL sitting-drops (drop diameter approximately 0.5-1 mm). Crystallisation solutions contained **(A)** 0.1 M MES pH 6.5, 25 % (w/v) PEG 2000 MME, **(B)** 0.1 M MES pH 6.5, 25 % (w/v) PEG 4000 and **(C)** 0.1 M MES pH 6.5, 25 % (w/v) PEG 6000.

Due to crystal morphology and size, optimisation was required before X-ray diffraction experiments were attempted. A grid-screen around the original hit conditions in larger volume hanging-drop format produced crystals of similar morphology. These large clusters of many fine needle crystals (**Figure 3.11A-B**) were found to break into smaller fragments when disturbed. In order to slow the speed of crystal nucleation and

potentially form fewer individual crystals, both protein and PEG precipitant concentrations were reduced (Blow *et al*., 1994). However, this resulted in drops that either contained similar crystalline material or remained clear. Experiments prepared under an oil layer to slow vapour-diffusion (Chayen, 1999) did not improve crystal morphology. Microseeding to introduce fragments of the sub-optimal initial crystals as nucleants in otherwise clear drops (Bergfors, 2003) did show signs of improved crystal morphology in our hands (**Figure 3.11C-F**). The process involved a large degree of trial and error with little logical explanation for the variation in quality of crystal growth. The majority of crystals were obtained following this strategy through extensive experiment repetition and grid screening (Cudney *et al*., 1994).

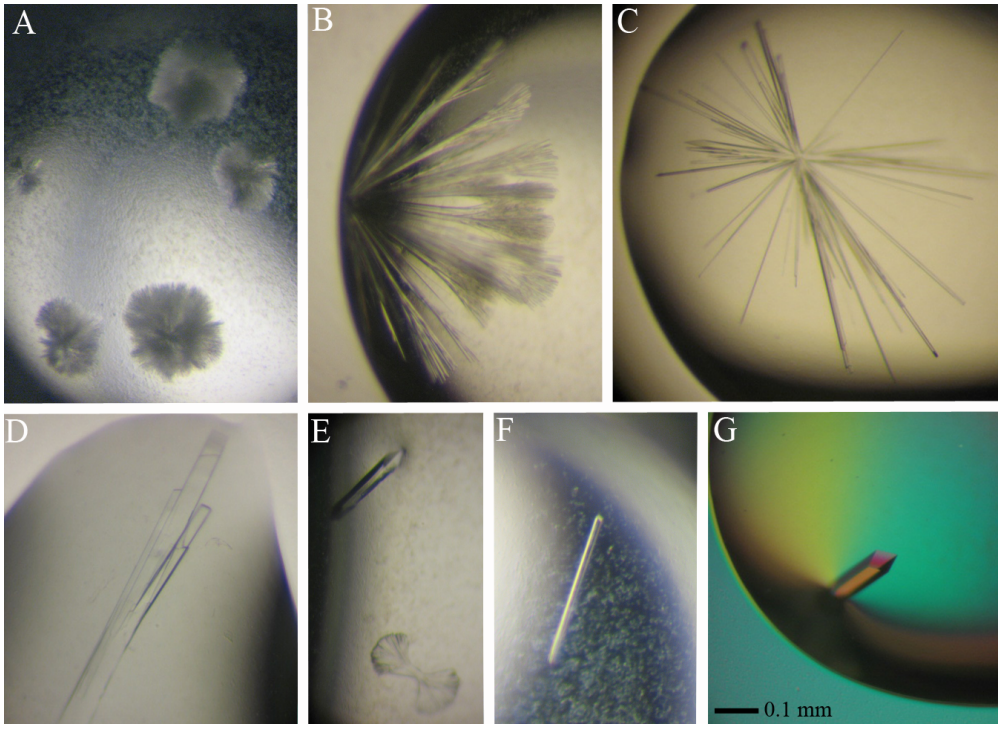


Figure 3.11. *Lm*TBCC152 crystal optimisation
A selection of crystals of *Lm*TBCC152. Panels **(A)** and **(B)** show crystals obtained without seeding. **(C)-(G)** all show crystals after one or more rounds of microseeding. **(G)** was photographed under polarised light. Crystal sizes range from between approximately 0.1 to 0.5 mm along the longest edge.

Other methods of optimisation that were attempted included reductive methylation of surface lysine residues (Walter *et al*., 2006; Kim *et al*., 2008), which proved unhelpful in this case. A large amount of protein precipitated during the reaction. A sample of the soluble material was analysed by MALDI-TOF but no peaks were observed despite positive evidence from SDS-PAGE. Crystallisation was not carried out using the low yield of potentially methylated *Lm*TBCC152.

KI or NaI were included in crystallisation conditions to prevent disruption to already-formed crystals when soaked with a high concentration of iodide ions to be utilised for experimental phase determination. However, the presence of iodide appeared detrimental to crystal growth and crystallisation drops contained a greater amount of precipitated material. Crystals cryoprotected in a solution supplemented with NaI were observed to split or partially dissolve prior to cooling. Alternative additives were found to show little or no positive effect on crystallisation. Samples containing DTT and/or GTP crystallised with the same morphology. ITC and DSF were used to assess the potential interaction of *Lm*TBCC152 with GTP but provided no evidence of a binding event. ITC was performed under a limited set of conditions and a non-significant T_m change observed in DSF does not necessarily equate to a lack of association (Niesen *et al*., 2007). Therefore, the ability of these additives to directly associate with *Lm*TBCC152 and the level of incorporation within the crystal lattice is still unknown at this stage.

One hypothesis for the problematic crystal growth was the possible presence of a non-homogenous protein solution. Although the sample collected immediately following size exclusion chromatography was believed to be monomeric *Lm*TBCC152, it is

possible that oligomerisation was dynamic and a dimer was able to form under certain conditions, such as in the crystallisation mixture or simply over a prolonged incubation time. Samples prepared under various conditions were therefore analysed via size exclusion chromatography according to **2.3.8** (**Figure 3.12**). A small peak at an elution volume suggestive of dimer formation was present for all samples. Long-term storage and exchange to the buffer used in crystallisation (MES pH 6.5) also produced aggregated material. It is possible that this may have contributed negatively to crystallisation. However, the predominant species remains monomeric and crystallisation experiments were prepared immediately following purification.

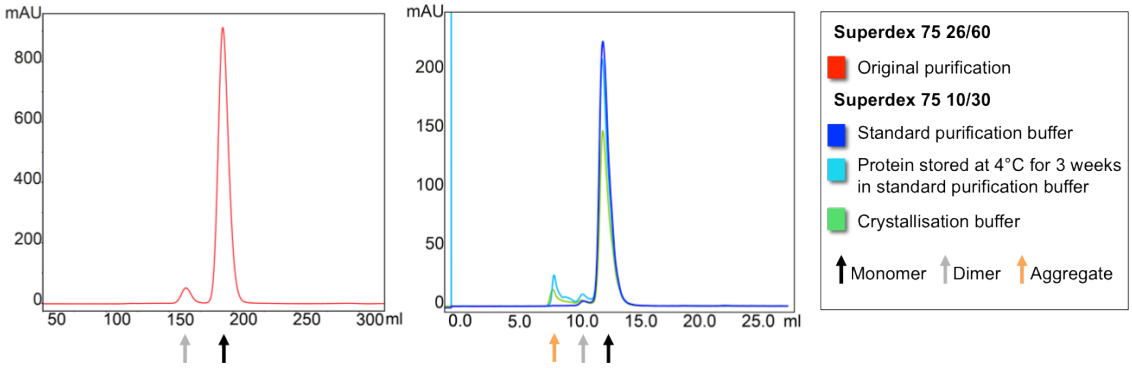


Figure 3.12. *Lm*TBCC152 oligomerisation analysis
Analytical size exclusion chromatography of *Lm*TBCC152 under different buffer conditions and sample storage. The left panel shows the original purification chromatogram. The right panel shows three chromatograms overlaid, each representing a different experiment as detailed in the key provided.

SeMet-*Lm*TBCC152 was produced as described in **2.3.3**. *Lm*TBCC152 contains only one methionine residue, excluding Met1. In order to improve possible experimental phasing procedures downstream, nucleotides encoding leucine residues were conservatively mutated to produce 1-2 additional methionine residues in the protein product. Initial gene constructs were prepared by site-directed mutagenesis (**2.2.4**). Amino acids targeted for mutation were identified based on a multiple sequence alignment of TBCC homologues. Both L215 and L223 are relatively well conserved between species and are indeed found as methionine at the equivalent positions in

Aspergillus flavus and *T. brucei* homologues, respectively (**Figure 3.13A**). Both the L215M-L223M double mutant and the L223M single mutant SeMet proteins were crystallised (**Figure 3.13B-C**). The L215M-L223M crystals were small and poorly formed and only the L223M crystals were subject to X-ray diffraction.

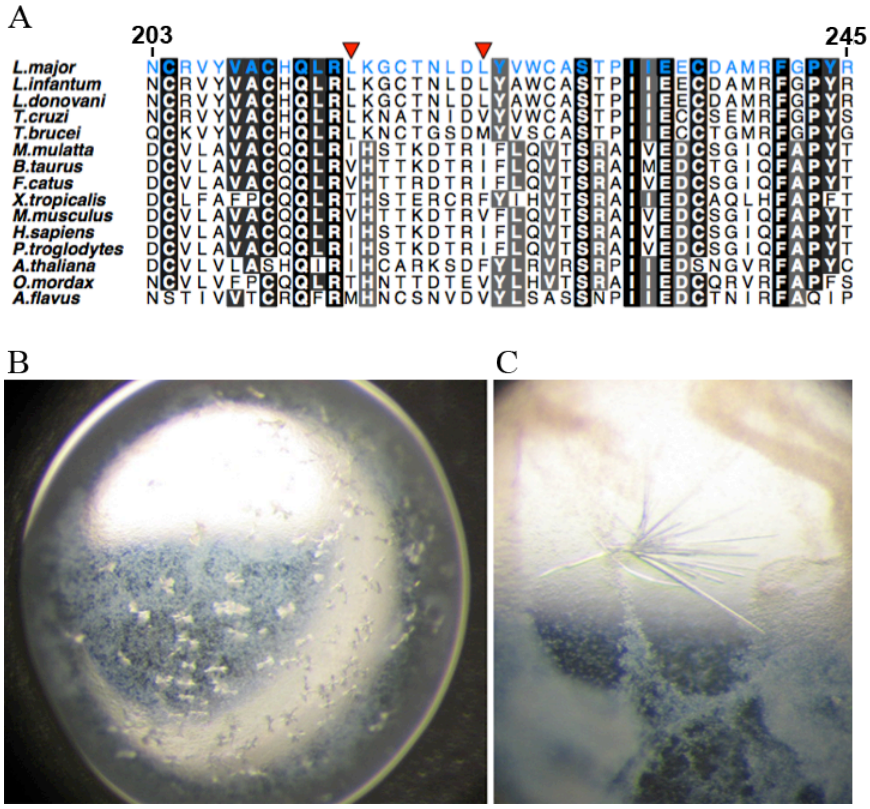


Figure 3.13 SeMet-*Lm*TBCC152 sequence mutation and crystallisation
(A) An excerpt of a TBCC multiple sequence alignment prepared using *BLAST* (Altschul *et al*., 1997) and *ALINE* (Bond and Schüttelkopf, 2009). *Lm*TBCC is shown as blue text from residues 203-245. Boxes on a darkening greyscale highlight sequence identity. The positions of Leu215 and Leu223 are indicated by red triangles. These residues were mutated to Met for the preparation of a SeMet derivative. **(B)** L215M-L223M double mutant crystals and **(C)** L223M single mutant crystals.

3.2.4. LmTBCC152 data collection and processing

Owing to the difficulties in obtaining single diffraction-quality crystals, over 160 native and SeMet *Lm*TBCC152 crystals were tested. A total of 7 datasets were collected and processed as described in **2.8.4**. Data from different crystals were indexed either in monoclinic space group $P2_1$ or orthorhombic $P2_12_12_1$. Corresponding unit cell parameters and other representative data processing statistics are listed in **Table 3.3**.

	Native A	Native B	SeMet C	SeMet D
Resolution range (Å)	46.20-2.3 (2.38-2.30)	44.56-2.10 (2.16-2.10)	28.23-2.88 (3.06-2.88)	28.08-3.40 (3.67-3.40)
Wavelength (Å)	1.5418	1.5418	0.9793	0.9793
Space Group	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Unit cell parameters				
***a* (Å)**	36.98	50.78	37.38	50.25
***b* (Å)**	92.4	67.48	93.05	66.37
***c* (Å)**	47.74	92.87	47.6	92.96
β (°)	108.3	90.0	108.2	90.0
Solvent content [a] (%)				
1 molecule per ASU	68	68	68	68
2 molecules per ASU	35	37	36	35
No. Reflections	260105 (6026)	134827 (10463)	15595 (1014)	43592 (8307)
No. Unique Reflections	12796 (865)	19293 (1558)	5419 (491)	4594 (895)
R_{merge} [b] (%)	14.1 (59.1)	23.8 (72.8)	18.6 (40.9)	19.1 (38.6)
Completeness (%)	94.6 (66.0)	100.0 (100.0)	77.2 (43.6)	99.2 (96.7)
Mean $I/\sigma(I)$	20.4 (2.9)	5.8 (2.5)	7.6 (2.8)	10.2 (7.0)
Redundancy	20.3 (7.0)	7.0 (6.7)	2.9 (2.1)	9.5 (9.3)

Table 3.3. *Lm*TBCC152 data processing statistics

Example data processing statistics from two native and two SeMet *Lm*TBCC152 crystals, indexed in two possible space groups identified. [a] Solvent content was estimated according to Matthews (1968) using a calculated molecular mass of 20476 Da in all cases. [b] R_{merge} is as defined in **Table 3.1**. Values in parentheses correspond to the highest resolution shell. ASU, asymmetric unit.

Due to the disparity in symmetry classification, it is possible that the data have been indexed in the incorrect space group or translational pseudosymmetry may be present. The high R_{merge} values shown in **Table 3.3** suggest that these data are of low quality, even where high redundancy is also a contributory factor. Growth of single crystals was extremely difficult and when achieved, the crystals were often split or cracked (see **Figure 3.11D** as an example). There may therefore be physical imperfections present within the crystal lattice that are not visible through the microscope. Radiation may also have led to additional crystal damage, particularly as *Lm*TBCC152 has a high overall content of sulfur-containing amino acids (12 cysteine and 2 methionine). These are known to be susceptible to radiation-induced structural changes (Burmeister, 2000) owing to their relatively large absorption coefficient (Garman, 2003). Radiation damage-induced phasing (RIP; Ravelli *et al*., 2003) could not be applied as any 'after'

data (data obtained after X-ray burn) were unusable.

Attempts to solve the structure by MR, sulfur-SAD or via the combined MR-SAD approach using highly redundant data collected in-house, processed in space group $P2_1$, were not successful. Nor was a solution obtained using data merged from multiple crystals. Similar experiments performed using native orthorhombic data were equally unproductive. MR search models included an NMR model of *Hs*TBCC *C*-terminal domain (PDB 2yuh, Saito *et al*., unpublished) and the *N*-terminal domain of *Hs*RP2 (PDB 2bx6, Kühnel *et al*. (2006); 3bh6 and 3bh7, Veltel *et al*. (2008)). Aligned regions of these proteins share only approximately 20-25 % sequence identity with *Lm*TBCC152. Tests were performed using the entire aligned region, following truncation to remove potential flexible regions or with side chain atoms removed as a polyalanine model. Molecular replacement is believed to have proved unsuccessful for similar reasons as for *Lm*TBCA (**3.1.4**). *Lm*TBCA is predominantly α-helical while *Lm*TBCC152 structure predictions and homologues identified based on sequence suggest that it is composed of β-strands in a helical arrangement, a β-helix (Yoder *et al*., 1993).

SeMet data collected at the absorption peak wavelength, as indicated by an energy scan (**Figure 3.14A**), appeared to provide sufficient anomalous signal (not shown) but several software packages for phase determination by SAD failed to identify the locations of selenium atoms. As mentioned, the crystals also appeared highly sensitive to synchrotron radiation. Diffraction visibly deteriorated over the experiment duration to such an extent that additional data could not be collected at the inflection or remote wavelengths for the application of MAD phasing. An increasing R_{merge} and *B* factor,

decreasing reflection intensity as well as discolouration of the cryo-cooled sample by the X-ray beam (**Figure 3.14B**) may also be indicative of radiation effects (Garman, 2010).

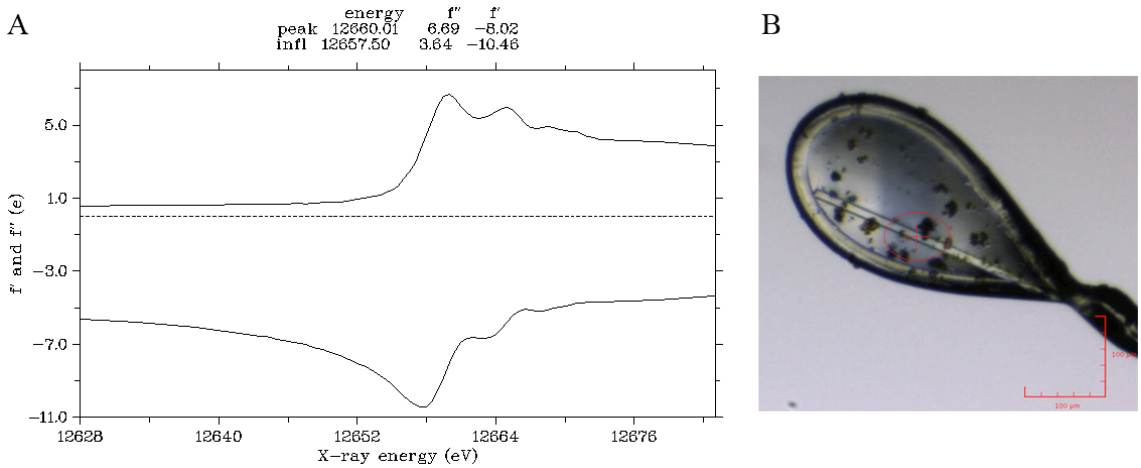


Figure 3.14. SeMet-*Lm*TBCC152 energy scan and radiation effects
(A) Energy scan of a SeMet-*Lm*TBCC152 (L223M) crystal showing the selenium peak and inflection X-ray energies and associated anomalous scattering factors, f' and f''. This figure was generated by *Chooch* (Evans and Pettifer, 2001). **(B)** A native *Lm*TBCC152 crystal in the X-ray beam showing a vertical path of radiation-induced discolouration.

3.2.5. TBCC sequence and homologue-based structure analysis

Based on the protein sequence, *Lm*TBCC152 is most similar to other known or putative TBCCs and is also predicted to share structural similarity with the protein RP2. This protein is encoded by the retinitis pigmentosa 2 gene in *H. sapiens* and is implicated in diseases causing retinal degeneration. Consisting of two structural domains, the *N*-terminal region is homologous to the *C*-terminal domain of TBCC (Bartolini *et al*., 2002). RP2 has been shown to function as a GTPase-activating protein (GAP) for Arl3 (ADP-ribosylation factor-like 3) (Kühnel *et al*., 2006; Veltel *et al*., 2008). TBCC has also been proposed to possess GAP activity in the hydrolysis of GTP bound to β-tubulin prior to release of the αβ-tubulin heterodimer from the TBC-tubulin supercomplex (Tian *et al*., 1999). Crystal structures of RP2 in complex with its Arl3 partner and GTP analogues can therefore help to inform on the location of the nucleotide site and the

potential interaction of TBCC with β-tubulin. A solution structure of the *Hs*TBCC *C*-terminal domain (PDB 2yuh) has no associated publication and together with the RP2 homologue structures (PDB 2bx6, Kühnel *et al*., 2006; PDB 3bh6, 3bh7, Veltel *et al*., 2008) comprise the limited structural details available for the *C*-terminal domain of TBCC. The *N*-terminal region of *Hs*TBCC has also recently been determined by NMR (PDB 2l3l, Garcia-Mayoral *et al*., 2011). An alignment of the full-length *Lm*TBCC and *Hs*TBCC sequences with the homologous domain of *Hs*RP2 is shown in **Figure 3.15A**. The regions of *Hs*TBCC with known structure are highlighted. There are several regions where sequence is maintained between these proteins. In particular, as also evident in the section of a larger alignment shown in **Figure 3.13A**, the arginine at position 214 and neighbouring residues are strictly conserved throughout species in TBCC proteins and RP2. This arginine has been shown to severely affect the rate of Arl3 GTP hydrolysis when mutated in RP2 (Veltel *et al*., 2008) and is also key in the function of other GAPs (Scheffzek *et al*., 1997). Other amino acids important for catalytic function include Gln212, Glu234 and Phe290 (numbered according to *Lm*TBCC sequence). This strict conservation between species strongly suggests conservation of function and further emphasises the role of TBCC as a GTPase-activating protein. *Hs*TBCC has indeed been shown to partially replace RP2 functionality where the reduced activity was justified by a lower affinity for Arl3 as a protein partner (Veltel *et al*., 2008). Studies have also demonstrated that TBCC and TBCD are together required for the hydrolysis of GTP by β-tubulin and stimulate release the αβ-tubulin heterodimer (Tian *et al*., 1997). TBCD does not appear to display a motif or fold common to any known GAPs, suggesting it may be TBCC adopting the more classical GAP role.

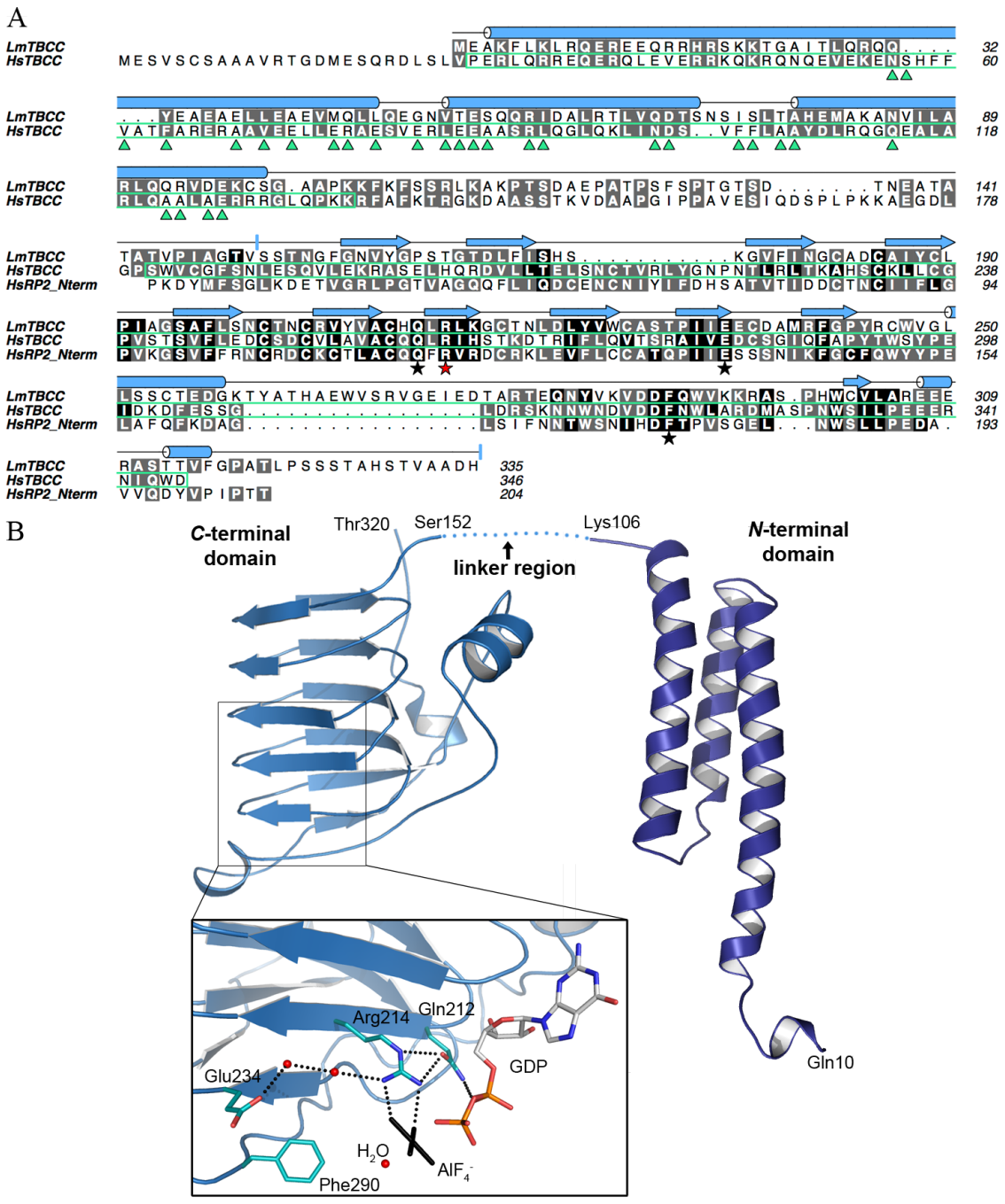


Figure 3.15. *Lm*TBCC homologue sequence alignment and structural model
(A) Full-length *Lm*TBCC aligned with *Hs*TBCC and the *N*-terminal domain of *Hs*RP2 (residues 38-204). Sequences were aligned using *MUSCLE* (Edgar, 2004) and annotated using *ALINE* (Bond and Schüttelkopf, 2009). The secondary structure of *Lm*TBCC predicted by *PSIPRED* (Jones, 1999) is shown as blue cylinders (α-helices) and arrows (β-strands) and the crystallised portion is indicated by blue markers. Regions of *Hs*TBCC with NMR structures available are surrounded by a green border. Green triangles highlight residues perturbed by tubulin and peptides in interaction studies (Garcia-Mayoral *et al*., 2011). Selected important amino acids for *Hs*RP2 GAP activity are shown by black stars while the red star indicates the strictly conserved catalytic arginine (Veltel *et al*., 2008). **(B)** Structural models of the *Lm*TBCC *N*- and *C*-terminal domains generated by *PHYRE2* (Kelley *et al*., 2009). A zoomed view also shows the important catalytic residues of *Hs*RP2 (PDB 3bh7) as cyan sticks, labelled according to *Lm*TBCC, and their interaction with GDP and AlF_4^- when complexed with Arl3 (not shown).

Structural homology models of the two individual *Lm*TBCC domains were prepared using the *PHYRE2* webserver (Kelley and Sternberg, 2009). Displayed in cartoon form in **Figure 3.15B**, the *N*-terminal domain is based on the same domain of *Hs*TBCC (approximately 70 % model confidence with 86 % sequence coverage) while the *C*-terminal domain is based on the homologous region of *Hs*RP2 (100 % model confidence with 78 % coverage). *Hs*RP2 complexed with Arl3 contains a molecule of GDP residing at the interface between the two proteins. The catalytic arginine and other important residues described above are shown in **Figure 3.15B** when superimposed on the structural model of the *Lm*TBCC *C*-terminal domain. Additional residues of Arl3, including Gln71 (not shown), act to position the water molecule required for hydrolysis and further contacts are made between the protein surfaces. This could therefore explain the difficulties in detecting a binding event between *Lm*TBCC and GTP as the partner protein may be necessary and the nucleotide does not bind, or has a low affinity, to TBCC alone. Nor does there appear a suitable pocket on the putative model of *Lm*TBCC152 for the nucleotide to bind.

Secondary structure predicted from sequence alone (*PSIPRED*; Jones, 1999) shown in **Figure 3.15A** presents some inconsistencies with that of the models. For example, the initial α-helix of *Lm*TBCC is predicted to span from Ala3 to Leu48. However, the model depicts a break in this helix, also described in *Hs*TBCC where at least the first 30 amino acids are highly flexible and have no fixed orientation. These residues are among those that Garcia-Mayoral and coworkers (2011) also found were perturbed during NMR experiments in the presence of a β-tubulin peptide or the αβ-tubulin heterodimer. The broad range of residues, noted by green triangles in **Figure 3.15A**, suggests that a large surface area of the domain interacts with tubulin and/or a large

conformational change takes place. The *N*-terminal tail may therefore become ordered upon complex formation as seen in other unstable microtubule binding proteins (Al-Bassam *et al*., 2002).

Although the predicted α-helical configuration of the *Lm*TBCC *N*-terminal region could be described as visually similar to that of *Lm*TBCA (**3.1**), it is difficult to align the sequences or structures accurately. Pairwise sequence alignement by *MUSCLE* (Edgar, 2004) suggests an identity of approximately 17 % but caution should be exercised in this analysis. Alignment of the 'maintained' amino acids structurally is problematic due to the differences in α-helix lengths. Perhaps the helical region performs a similar role to TBCA but equally it may be required to interact with another component of the supercomplex and based on these limited comparisons, it is not possible to infer specific common functional capabilities.

The link between the TBCC *N*- and *C*-terminal domains is predicted to be disordered, suggesting that the connection is flexible. A similar feature is seen in *Tb*TBCB where the two domains were also modelled separately (Fleming *et al*., 2013). It is also clear that TBCC is composed of two distinct units. Preliminary small-scale pull-down trials using his-tagged *N*-terminal domain (*Lm*TBCC1-111 or *Tb*TBCC1-107) and untagged *Lm*TBCC152 or *Tb*TBCC138 did not identify an interaction (data not shown). Indeed, the propensity of full-length TBCC to cleave without the intentional addition of a protease enzyme is indicative of the individual nature of the domains.

There is currently no knowledge of how the cofactors are arranged in the supercomplex with α- and β-tubulin or details of their molecular interactions. Due to the proposed

function of TBCC as a GTPase-activator, catalysing β-tubulin GTP hydrolysis, we can postulate that the *C*-terminal domain of *Lm*TBCC interacts with β-tubulin through the surface where the strictly conserved arginine finger resides. The equivalent region of RP2 is revealed as the interface between RP2 and Arl3 so this part of the conserved protein fold is known to be receptive to protein-protein interactions and has a complementary surface charge distribution. Complex structures of RP2 containing GDP can therefore inform on the possible position of the β-tubulin nucleotide with respect to *Lm*TBCC. A crude superimposition of the GDP molecules in the RP2-Arl3-GDP complex and in β-tubulin (PDB 1tub (Nogales *et al*., 1998) and a more recent PDB 4ihj (Prota *et al*., 2013) were both examined) results in a series of structural clashes, particularly with the β-tubulin α-helices α5, α11 and α12 (Nogales *et al*., 1998). Using secondary structure matching procedures (Krissinel and Henrick, 2004) to align β-tubulin with Arl3 reveals a potential association (**Figure 3.16**). 107 Cα atoms were aligned between Arl3 and β-tubulin, mainly localised to the initial 242 β-tubulin amino acids. Overall sequence identity was poor at approximately 7 % but the aligned secondary structure appears to agree. GDP of β-tubulin is more buried than when bound to Arl3 where it is exposed on the surface and readily susceptible to the action of RP2. In this representation, the nucleotide guanine is rotated by up to 90° and the superimposed side chain of Arg214 is no longer within acceptable contact distance. The functional residues may therefore reposition to access the required catalytic orientation (Bourne, 1997). As a simplified, approximate model, exact contact details cannot be reliably extracted and more complex molecular modelling strategies could provide a more accurate prediction in the absence of experimental data. However, this is the first structural representation proposed for the interaction between TBCC and β-

tubulin. This representation includes only the *Lm*TBCC *C*-terminal domain but there remains potential for placement of the *N*-terminal region that agrees with proposed interactions with residues located at the β-tubulin *C*-terminus (Garcia-Mayoral *et al*., 2011).

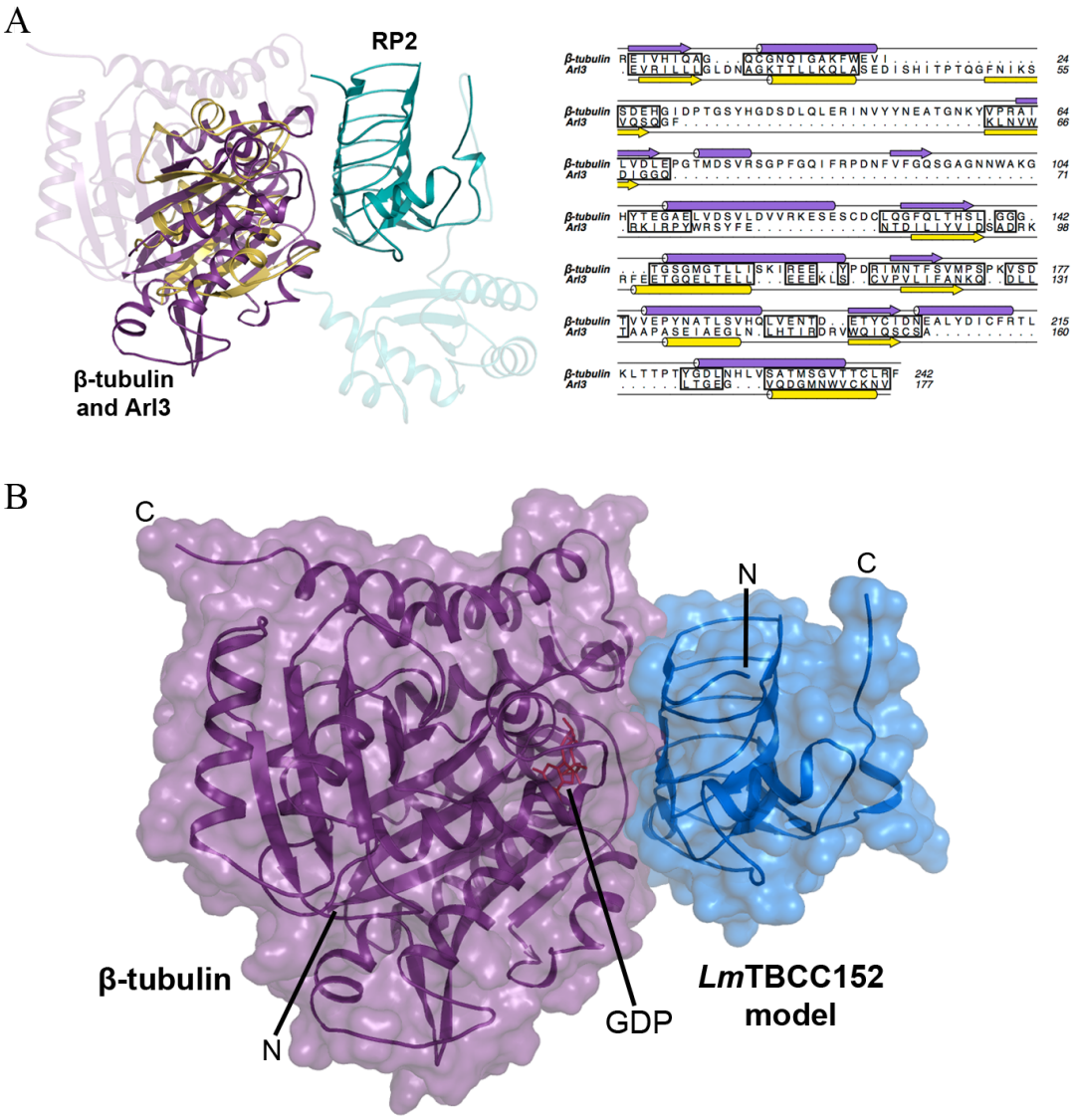


Figure 3.16. Potential configuration of *Lm*TBCC *C*-terminal domain and β-tubulin
(A) *Hs*Arl3 (yellow; PDB 3bh7 Veltel *et al*., 2008) with *Hs*β-tubulin (purple; PDB 4ihj, Prota *et al*., 2013) residues 2-242 superimposed by secondary structure matching. The alignment (right) shows the secondary structure of the highlighted regions of β-tubulin and Arl3 more clearly. Boxed portions of sequence indicate the 107 structurally aligned residues (*SSM*; Krissinel and Henrick, 2004). *Hs*RP2 from the RP2-Arl3 complex is shown in cyan with the *N*-terminal domain highlighted. **(B)** A semi-transparent van der Waals surface representation of the potential configuration of the *Lm*TBCC152 model (blue) and β-tubulin (purple). β-tubulin was aligned with Arl3 of the RP2-Arl3 complex as shown in **(A)** and the model of *Lm*TBCC was generated by *PHYRE2* (Kelley and Sternberg, 2009). β-tubulin GDP is indicated at the interface as red sticks and β-tubulin is in the same orientation in both **(A)** and **(B)**.

3.2.6. Addendum: LmTBCC152 structure solution

During the preparation of this thesis, the expression system for native and SeMet-*Lm*TBCC152 (L223M), crystallisation conditions and information retrieved from previous data collection experiments were passed to Paul K. Fyfe and a student, Alex Finney. New data collected at Diamond light source were processed using *XDSGUI* (Kabsch, 2010) and four selenium sites successfully identified by the *PHENIX* (Adams *et al*., 2010) pipeline. The majority of the model was automatically built using data extending to 2.0 Å resolution. The solution was obtained using monoclinic data indexed in space group $P2_1$ with unit cell dimensions of a = 37.64, b = 93.24, c = 48.28 Å and the angle, β = 108.4°. Two molecules were present in the asymmetric unit corresponding to a crystal solvent content of approximately 35 % (Matthews, 1968). Data were obtained from a single crystal, performed using a helical collection strategy to minimise radiation damage. Similar data processing problems were initially encountered as described above and due to 'crystal slippage' caused by the intentional collection strategy.

As expected, the TBCC *C*-terminal domain primarily adopts a β-helix formation. **Figure 3.17A** shows the overall structure of *Lm*TBCC152. It has an r.m.s.d. of 2.4 Å and 2.0 Å when aligned with 131 Cα atoms of the *Hs*TBCC *C*-terminal solution structure (PDB 2yuh) and 121 Cα atoms of the *Hs*RP2 *N*-terminal domain (PDB 3bh7), respectively. The homology model described above (**3.2.5**) also displayed an r.m.s.d. of approximately 2.0 Å when 120 Cα atoms are aligned and the superimposed structures are shown in **Figure 3.17B**. Structure is highly maintained over the β-helix (coloured blue) with larger deviations apparent in the flanking loop regions (red and orange). This

may have contributed to the difficulties obtaining the correct solution by molecular replacement using a search model containing this segment while the β-helix alone was not sufficient.

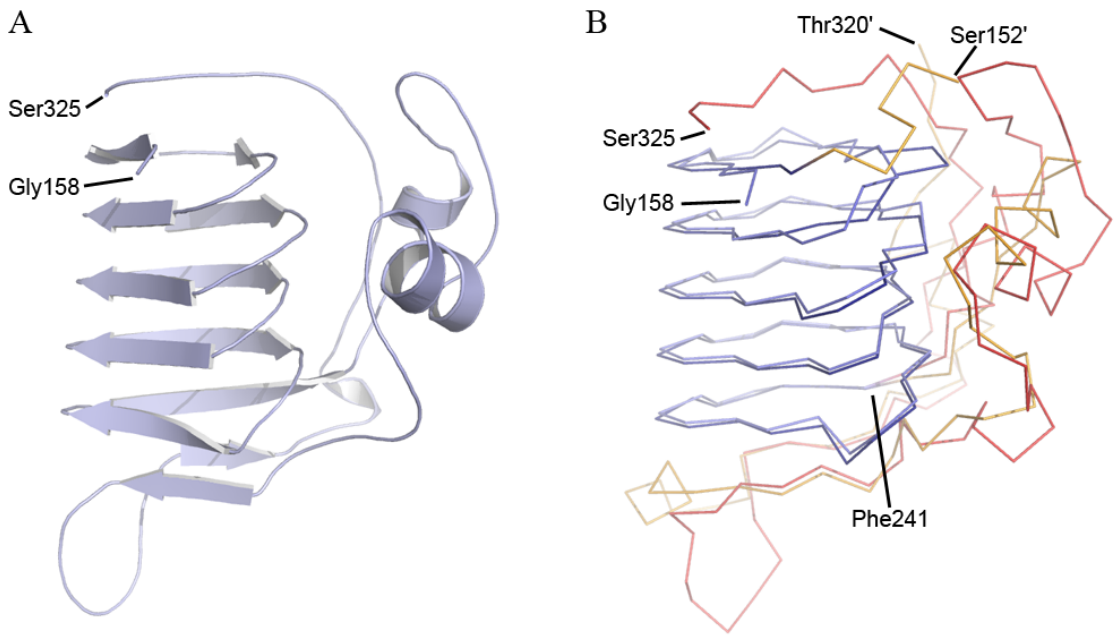


Figure 3.17 *Lm*TBCC152 overall structure
(A) Cartoon diagram of the crystal structure of *Lm*TBCC152. The *N*- and *C*-terminal residues of the model, Gly158 and Ser325, are labelled. **(B)** The *Lm*TBCC152 structure overlaid with the *PHYRE2* homology model. *Lm*TBCC152 is coloured pale blue and red while the homology model is coloured dark blue and orange. The blue regions of both models consist of Gly158 to Phe241. *N*- and *C*-terminal residues of the homology model are labelled Ser152' and Thr320', respectively.

3.3. Tubulin-binding cofactors B, D and E

3.3.1. TBCB

Crystal structures of the *Tb*TBCB *N*-terminal ubiquitin-like (Ubl) domain and *C*-terminal CAP-Gly domain were recently determined by colleagues (Fleming *et al.*, 2013; PDB accession codes 4b6w and 4b6m, respectively). *Lm*TBCB and *Tb*TBCB protein sequences are approximately 60 % identical. *Lm*TBCB was therefore over-expressed as full-length and extended domain constructs in attempts to bridge the link and identify the relative orientations of the individual domains. Despite the production

of soluble protein material, no crystals were obtained to progress the structural knowledge of TBCB.

3.3.2. TBCD

TBCD is the largest of the cofactors with a mass of approximately 150 kDa. Genes encoding TBCD were cloned into expression vectors from *T. brucei* and *L. major* gDNA. The final *Lm*TBCD translated sequence was identical to that annotated in databases (see **Table 2.1** for database details). However, a number of nucleotides were replaced in *Tb*TBCB. This may be a result of *T.brucei* strain variation but it is also possible that mutations were introduced experimentally. Initial small-scale test expressions showed little obvious soluble protein material of the expected molecular mass. A buffer extraction screen was carried out to assess whether different lysis buffering systems improved the soluble yield of protein following gene expression under identical conditions. Western blot analysis indicated the presence of small amounts of His-tagged sample and several buffers appeared to increase the level detected (**Figure 3.18A**). Protein ID was validated as *Lm*TBCD with 56 % peptide coverage when searched against the *L.major* sequence database (**Figure 3.18B**). 1 L cells were cultured following the same protocol as the small-scale test and progressed to standard purification methods using one buffer identified (0.1 M borate pH 8.5) throughout purification. Protein of expected molecular mass was then visible by SDS-PAGE. However, additional studies including size exclusion chromatography and DSF suggested the sample was highly aggregated or improperly folded and so, unsuitable for crystallography at this stage.

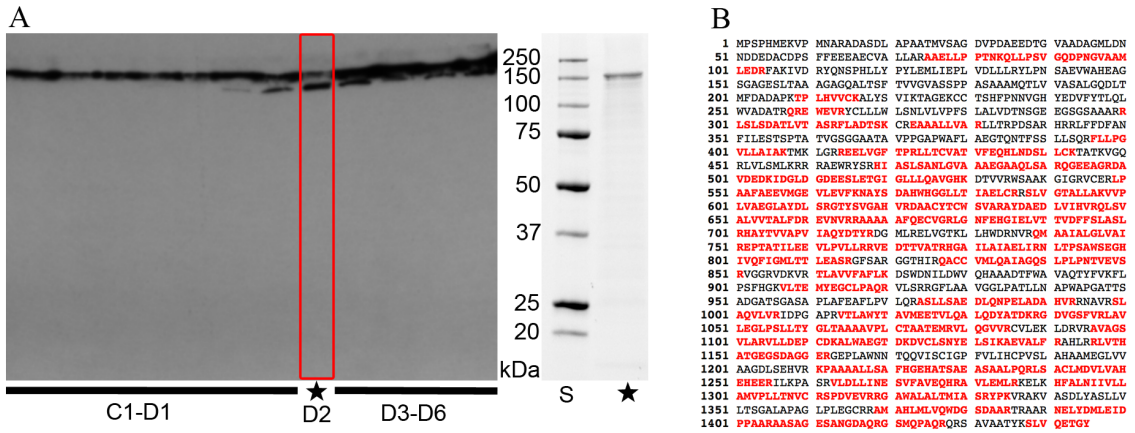


Figure 3.18 *Lm*TBCD purification

(A) Western blot of *Lm*TBCD buffer extraction screen and SDS-PAGE of *Lm*TBCD following affinity chromatography. Only soluble samples using buffers C1-D6 are shown (listed in **Table 2.5**). Buffer D2, indicated by a black star in the lane highlighted by a red box, consisted of 0.1 M borate pH 8.5. The horizontal band along the top of the image is believed to be aggregated *Lm*TBCD material. On SDS-PAGE, the lane indicated by a black star contains soluble *Lm*TBCD following capture by affinity chromatography. Lane S contains molecular weight standards that apply to both the Western blot and SDS-PAGE images. **(B)** Protein ID results from an excised SDS-PAGE gel slice. Sequence given in red text indicates peptides matched to *Lm*TBCD with an overall coverage of 56 %.

3.3.3. TBCE

All gene expression studies carried out to produce *T. brucei* and *L. major* TBCE have resulted in insoluble material. Full-length and truncated domain constructs have produced no measurable soluble protein using a bacterial expression system. TBCE is predicted to consist of two domains similar to TBCB with the *N*- and *C*-terminal domains reversed. Despite the ability to produce soluble recombinant TBCB, there remain some unknown features preventing the correct production of TBCE in *E. coli*. Protein refolding experiments performed by others have also shown no success and TBCE may only be obtained through isolation from a native source or gene expression using a eukaryotic cell host (Kortazar *et al* ., 2006).

3.4. Tubulin

According to a number of reports, soluble tubulins have been produced recombinantly in *E.coli* (Oxberry *et al*., 2001; MacDonald *et al*., 2001; 2003; 2004; Fennell *et al*., 2006; Giles *et al*., 2009). Published expression protocols to produce α- and β-tubulin from *T. brucei* were therefore followed (MacDonald *et al*., 2003; Giles *et al*., 2009) and additional fusion-tag constructs generated. In all cases, protein product was observed at the correct molecular mass but no soluble material was detected (**Figure 3.19**). The advantage of obtaining soluble tubulin recombinantly versus a native source is that the α- and β-tubulin monomers could be produced individually in the absence of other MAPs. Structural or interaction studies may then have been possible to interrogate TBC function. However, despite several positive reports, the procedure has not been successful in our hands. Tubulin undergoes a series of post-translational modifications (Janke and Kneussel, 2010) and indeed a complex folding process *in vivo* and it should not be possible to accurately reproduce these using a bacterial expression system. As such, recombinant tubulins are prone to aggregation or accumulation in inclusion bodies (MacDonald *et al*., 2003).

A small sample of porcine assembly-competent tubulin was obtained from Rachel E. Morgan (originally purchased from Cytoskeleton, Inc). An initial crude assessment was made of the ability of the sample to form polymerised microtubules in the absence of any TBCs. Microtubule concentration is proportional to absorbance at 340 nm so when monitored over time (up to 1 h), the formation of microtubules through the polymerisation of αβ-tubulin heterodimers can be assessed (Shelanski *et al*., 1973; Lee and Timasheff, 1977). The addition of ligands that may disrupt the polymerisation

reaction will therefore alter the standard absorbance curve. However, further development of the assay would be required in order to pursue this as one avenue for studying the effects of TBCs on tubulin polymerisation and may represent some promising future work.

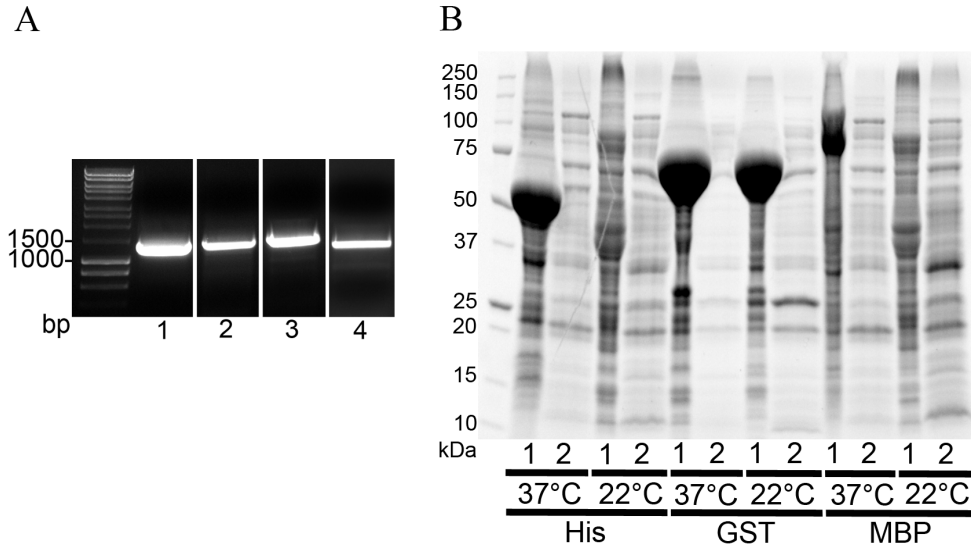


Figure 3.19 Tubulin PCR and test expression
(A) PCR products of α- and β-tubulin from *T. brucei* and *L. major* gDNA. Lane 1, *Tb*β-tubulin; lane 2, *Lm*β-tubulin; lane 3, *Tb*α-tubulin; lane 4, *Lm*α-tubulin. Length is indicated in nucleotide base pairs (bp). **(B)** *Lm*β-tubulin test expression of His-, GST- and MBP-tagged constructs, cultured in LB at the temperatures indicated. Lanes numbered 1 contain the insoluble pellet fraction following chemical lysis and lanes numbered 2 contain soluble samples. Large bands are observed in several insoluble lanes at the expected mass of the corresponding tagged β-tubulin product.

3.5. Part I summary and concluding remarks

The structure of recombinant TBCA from *L. major* is presented. This monomeric protein has a compact architecture consisting of three α-helices arranged in an anti-parallel bundle. A proline residue mid way along α2 distorts one of the longest helices, giving the molecule a curved surface. When compared with structural homologues, this helical bend is found to be more prominent in *Lm*TBCA and is uniquely accompanied by a similar bend in α1 where there is no equivalent proline. The resultant concave face is described as convex in related proteins. With the function of capturing excess free β-

tubulin in microtubule dynamics, potential binding sites are discussed. Amino acids conserved between species that have been described as functionally important elsewhere may also be implicated in binding in *Lm*TBCA. However, some contrasting areas of sequence are apparent in *Lm*TBCA. For example, a tyrosine in *At*TBCA is reported to be vital for β-tubulin binding through studies involving a Tyr-Ala mutation. Alanine is native at this position in *Lm*TBCA, raising questions regarding the implied functional properties of the residue. Meanwhile, a cluster of exposed positively charged amino acids on the concave surface of *Lm*TBCA could favourably bind to the highly negative β-tubulin. Although we have not performed functional experiments using *Lm*TBCA to confirm binding analogous to any of those already reported, accurate details of the interaction would be best identified by the elucidation of a β-tubulin-TBCA complex crystal structure.

Tubulin-binding cofactor C from *T. brucei* or *L. major* is a two-domain protein that when produced recombinantly degrades into two separate parts, each of which appeared relatively stable in isolation. Although the full-length protein and the *N*-terminal domain could not be crystallised, crystals of the *C*-terminal domain of *Lm*TBCC, composed of residues 152-335, have been generated and X-ray diffraction data collected from native crystals and a selenomethionine-derivative. Unfortunately, the crystallographic phase problem could not be solved using these data.

The TBCC *C*-terminal domain construct, *Lm*TBCC152, is predicted to adopt a similar conformation to a homologue protein, *Hs*RP2. An arrangement of at least ten β-strands form an overall β-helix with an additional, more flexible region neighbouring the helical

surface. A solution structure of the equivalent domain of *Hs*TBCC consists of multiple positions for these residues. A potential site of interaction with β-tubulin as a GTPase activator is also discussed. It has since been shown through the successful solution of the *Lm*TBCC152 structure by a colleague using crystals of SeMet-*Lm*TBCC152 (L223M) reproduced following the procedures outlined in this thesis that the protein is indeed organised as predicted. Detailed examination of the newly established crystal structure is required and assessment of the GTP hydrolysing properties of *Lm*TBCC is an exciting opportunity for future work.

Structural details of *T. brucei* TBCB have recently been described by colleagues (Fleming *et al*., 2013). Cofactors D and E, however, remain elusive and additional efforts to obtain soluble protein are necessary for progression in this area but these studies lay the foundations for such research.

Overall, **Part I** of this thesis provides reproducible gene expression protocols and an initial examination of the structural details of trypanosomatid TBCs. Extending from these results will allow the molecular interactions between tubulin subunits and cofactors to be further explored.

PART II

Pteridine reductase 1

4. Introduction

4.1. Pteridine reductase 1

Part II of this thesis presents studies on the enzyme, pteridine reductase 1 (PTR1; EC 1.5.1.33). An overview of enzyme function and inhibition is provided as well as an introduction to isothermal titration calorimetry, a biophysical technique employed alongside the structural and kinetic evaluation of potential PTR1 ligands.

4.1.1. Overview of function

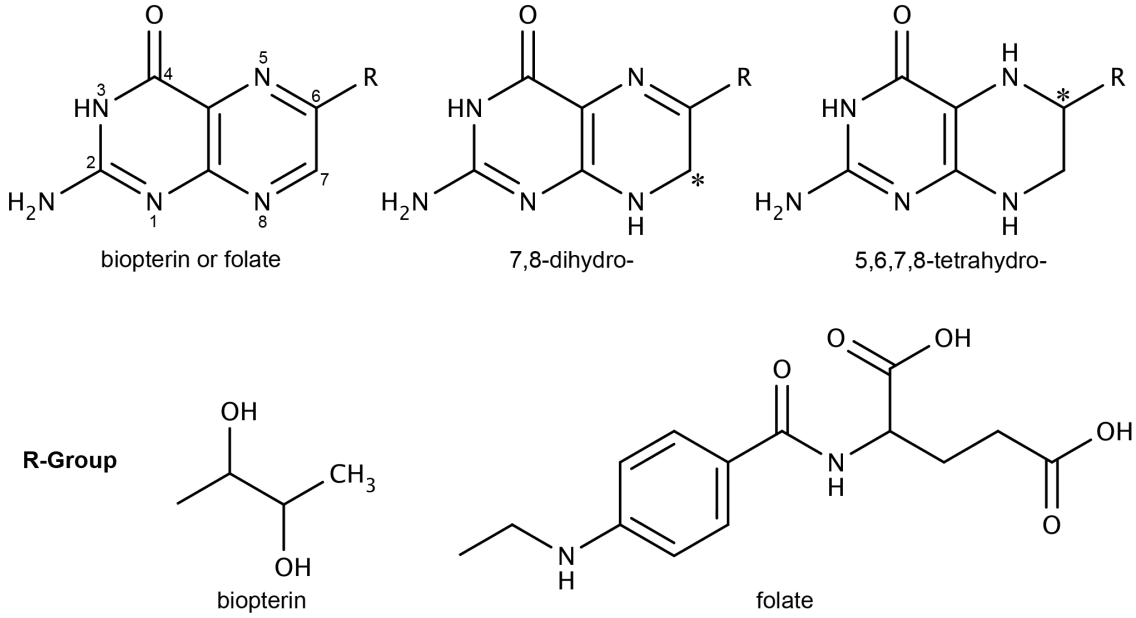


Figure 4.1. PTR1 substrates and products
PTR1 catalyses the two-stage reduction of biopterin or folate to the 7,8-dihydro- and 5,6,7,8-tetrahydro-forms. All reductions are accompanied by NADPH oxidation to $NADP^+$, providing a hydride to the carbon atoms at positions 7 and 6, marked with *. All chemical structures throughout this thesis were prepared using *MarvinSketch* (ChemAxon).

PTR1 is a member of the NADPH-dependent short-chain dehydrogenase-reductase superfamily (SDR) and is unique to trypanosomatid parasites. The primary function of PTR1 is catalysis of the two-stage reduction of biopterin to 5,6,7,8-tetrahydrobiopterin via the 7,8-dihydrobiopterin intermediate (Bello *et al*., 1994; Nare *et al*., 1997a). Pterins are pteridine derivatives; a pyrimidine and pyrazine fused ring system with

additional amino and keto groups at positions 2 and 4, respectively (**Figure 4.1**). Trypanosomatid parasites are pterin auxotrophs, employing a pathway that includes PTR1 in the salvage of pterins from the host. The function of reduced pterins in trypanosomatids is not fully characterised but they are required for parasite growth (Bello *et al*., 1994) and implicated in metacyclogenesis (Cunningham *et al*., 2001), the production of nitric oxide and resistance to oxidative stress (Nare *et al*., 1997b).

In addition to the reduction of unconjugated pterins (biopterin), PTR1 also displays the ability to reduce conjugated pterins (folate) to the dihydro- and active tetra-hydro forms (Bello *et al*., 1994; Nare *et al*., 1997a; **Figure 4.1**). Tetrahydrofolate is normally produced via the reduction of folate and dihydrofolate by the enzyme, dihydrofolate reductase (DHFR; EC 1.5.1.3). DHFR is present in humans with the folate substrate obtained through diet (vitamin B_9) but PTR1 is absent and tetrahydrobiopterin is sourced via the biosynthesis from GTP. Reduced folates are utilised most notably as cofactors in the biosynthesis of 2'-deoxythymidine-5'-monophosphate (dTMP) from 2'-deoxyuridine-5'-monophosphate (dUMP) via the enzyme thymidylate synthase (TS; EC 2.1.1.45). In trypanosomatids, DHFR exists as a bifunctional enzyme fused with thymidylate synthase, DHFR-TS.

The catalytic mechanism of PTR1 is similar to that of other SDRs and has been described elsewhere (Gourley *et al*., 2001; **Figure 4.2**). The first reduction step involves donation of a hydride from the NADPH cofactor to substrate at C7, generating $NADP^+$. A protonated aspartic acid residue does not directly interact with the substrate but donates a hydrogen bond to a strictly conserved active site tyrosine that provides a proton to substrate N8 and is stabilised by a nearby lysine residue during the transition,

which also helps to orient the cofactor. The second reduction step occurs at positions 5 and 6 of the dihydro- substrate and hydride is transferred from the nicotinamide C4 to substrate C6. An activated water molecule replaces tyrosine as the proton donor to N5 while the cofactor pyrophosphate and an ideally placed arginine residue assist in the temporary enolisation of the substrate keto group.

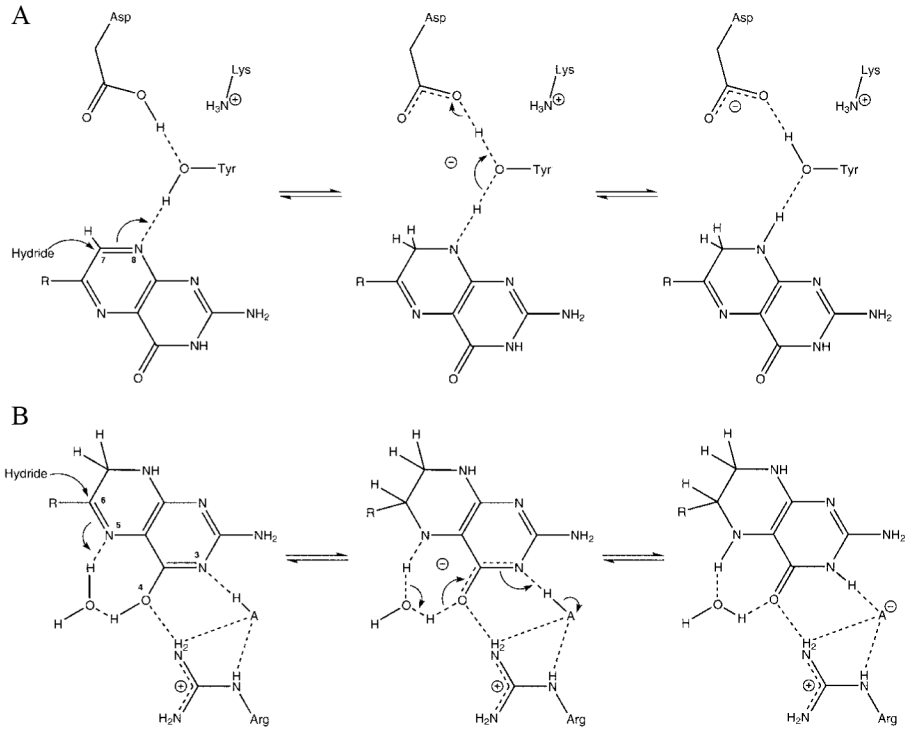


Figure 4.2. PTR1 mechanism
The PTR1 mechanism, modified from Gourley *et al*. (2001) with permission, showing the first **(A)** and second **(B)** reduction steps to produce the dihydro- and tetrahydro- forms of biopterin or folate. R represents the same groups provided in **Figure 4.1** and the cofactor phosphate is shown as AH. In *Trypanosoma brucei* PTR1, amino acids depicted correspond to Arg14, Asp161, Tyr174 and Lys178.

4.1.2. Pteridine reductase 1 inhibition

Molecules targeting folate synthesis and metabolism are in clinical use in the treatment of bacterial infections, certain cancers as well as malaria. In particular, DHFR is the target of such drugs (Blakley, 1995; Kompis *et al*., 2005), affecting DNA synthesis by limiting the supply of dTMP. The presence of PTR1 in trypanosomatid parasites provides a bypass mechanism (Hardy *et al*., 1997; Nare *et al*., 1997a) for the reduction

of folates when DHFR is inhibited by traditional antifolates and is therefore detrimental to their effectiveness in these organisms (Nare *et al*., 1997a; Luba *et al*., 1998). In *Leishmania* parasites, PTR1 is the only enzyme known to reduce biopterin and has been shown to be essential for cell growth *in vitro* by gene knockout studies where growth was successfully rescued by the provision of reduced pterins (Bello *et al*., 1994; Nare *et al*., 1997b). Additionally, the loss of *T. brucei* cell viability and virulence resulted from knockdown studies by RNA interference in the bloodstream form of the parasite (Sienkiewicz *et al*., 2010). PTR1 therefore represents an interesting target for the development of potential therapeutic agents in the fight against diseases caused by trypanosomatid parasites such as human African trypanosomiasis, Chagas disease and Leishmaniasis (Ong *et al*., 2011). A single molecule that can target both PTR1 and trypanosomatid DHFR-TS or an inhibitor specific to PTR1 for use in combination with known antifolates are both desirable prospects (Nare *et al*., 1997b). Or indeed, there may be potential for effective therapies to target this candidate alone.

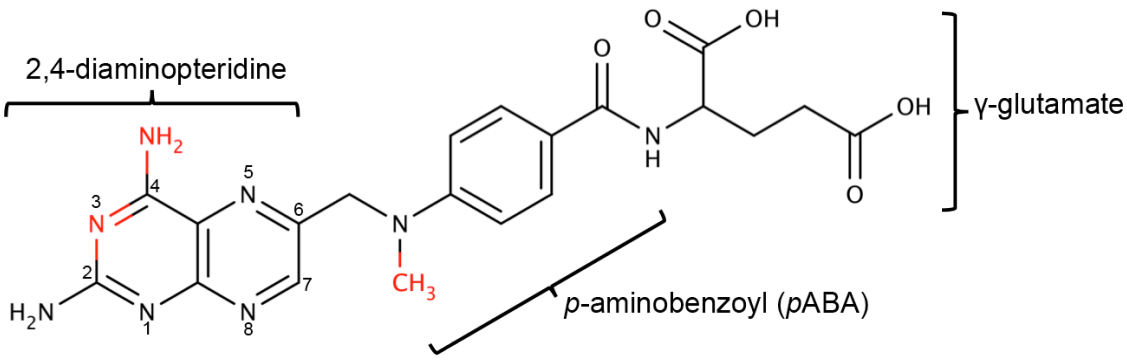


Figure 4.3. Methotrexate chemical structure
Methotrexate (4-amino-N^{10}-methyl-pteroylglutamic acid; MTX) has a highly similar structure to folate and the differences are highlighted red. See **Figure 4.1** for the chemical structures of folate and reduced folates.

The archetypal antifolate, methotrexate (MTX) is shown in **Figure 4.3**. The chemical structure of MTX is highly similar to that of folate, differing at position 3 and 4 of the pteridine moiety and with an additional methyl group in the aminobenzyl region (shown

in red in **Figure 4.3**). While able to inhibit PTR1, MTX is less effective against this enzyme than it is against DHFR-TS (IC_{50} 1.1 μM against *Lm*PTR1 and 0.005 μM against *Lm*DHFR-TS; Hardy *et al*., 1997) and resistance can be generated by *Leishmania* parasites (Hardy *et al*., 1997; Nare *et al*., 1997a). *L. major* parasites displayed a high sensitivity to MTX in PTR1 knockout studies (Bello *et al*., 1994) and upregulation of PTR1 expression is one possible mechanism of MTX resistance, emphasising that current antifolate drugs are inadequate for the treatment of diseases caused by trypanosomatid parasites.

4.1.3. Structural knowledge

Crystal structures of PTR1 from *L. major* (*Lm*PTR1) and *T. brucei* (*Tb*PTR1) have been published using diffraction data extending to 1.75 Å and 2.2 Å resolution, respectively (Gourley *et al*., 2001; Dawson *et al*., 2006). Additional insight is provided by ternary complex crystal structures containing cofactor and substrate or inhibitor of *Lm*PTR1 (Gourley *et al*., 2001; McLuskey *et al*., 2004; Schüttelkopf *et al*., 2005; Cavazzuti *et al*., 2008) and *Tb*PTR1 (Mpamhanga *et al*., 2009; Tulloch *et al*., 2010; Spinks *et al*., 2011). Collectively, these structures have afforded an understanding of cofactor and substrate binding, informed a potential mechanism of action and presented possible modes of inhibition.

Sequence similarity between PTR1 from different trypanosomatid species is high. For example, *Tb*PTR1 and *Lm*PTR1 share approximately 50 % identical amino acids and are over 80 % similar when conservative substitutions are considered. This is reflected by a high conservation of overall protein structure, described in more detail in **Chapter 6**, but there are several notable active site differences with some inhibitors displaying

different affinities for the orthologous enzymes, including MTX (K_i 0.152 μM and 0.039 μM for *Tb*PTR1 and *Lm*PTR1, respectively; Dawson *et al*., 2006). This includes a tryptophan-histidine (*Tb*Trp221-*Lm*His241) and cysteine-leucine exchange (*Tb*Cys168-*Lm*Leu188) and differences in a loop known to line the substrate-binding region of the active site (Dawson *et al*., 2006). These sequence disparities may be sufficient to demand inhibitors with specificity for a single orthologue and, coupled with the potential of ligands to adopt distinctive orientations within the PTR1 active site, it is essential to characterise new inhibitors structurally to confirm the mode of binding. Furthermore, molecules have yet to be produced with sufficient inhibitory effect against a trypanosomatid PTR1 together with the appropriate physicochemical properties and efficacy to allow progression to later stages of drug development, underlining the need for additional rational design efforts.

4.2. Isothermal titration calorimetry

4.2.1. Practical overview

Virtually all molecular binding events are accompanied by an exchange of heat (Freyer and Lewis, 2008). Isothermal titration calorimetry (ITC) is a label and immobilisation-free technique that can be used to characterise the thermodynamics of those binding interactions. Modern calorimeters are highly sensitive, detecting heat changes as little as 0.1 μcal (0.4 μJ) and the change in heat is directly proportional to the amount of a particular reaction that has occurred (Wiseman *et al*., 1989).

The studies presented here have been carried out using a power compensation isothermal titration calorimeter, containing two identical cells maintained at the same

temperature throughout the course of the experiment. In the case of a ligand:protein titration, multiple successive injections of a set volume of ligand are added to protein held within the experimental cell. When a binding event occurs, the change in heat is sensed and power (typically recorded in μcal sec^{-1}) is adjusted accordingly to maintain the constant temperature with respect to the unmodified reference cell (Leavitt and Freire, 2001). In an exothermic reaction where heat is evolved, power is consequently reduced and results in a downward peak on the raw data panel (**Figure 4.4A**). Large heat changes are expected on initial injections when all protein binding sites are available for complex formation. Subsequent injections result in smaller heat changes as fewer titrant molecules are able to bind to a protein partner. Finally, when saturation occurs, only the effects of dilution and mixing are observed. A single continuous injection is indeed possible but not employed during this project. Two drawbacks of ITC are the relatively low throughput and high reagent consumption levels although instruments with greater automation and reduced sample volume capabilities are becoming available to allow application of the technique earlier in the drug discovery pipeline (Weber and Salemme, 2003).

4.2.2. Thermodynamic profile

A thermodynamic profile of an interaction can be obtained from a single ITC experiment (Wiseman *et al*., 1989). This includes the direct measurement of the association (or binding) constant, K_a, the stoichiometry of the interaction, N, and the change in enthalpy caused by the binding event, ΔH. Indeed, calorimetry is the only method able to measure enthalpic changes directly (Freire *et al*., 1990; Ladbury, 2010). With knowledge of the universal gas constant, R, and the temperature, T, calculation of

the dissociation constant, K_d, the change in entropy, ΔS, and the observed free energy of binding, ΔG can be accomplished using **Equations 4.1** and **4.2**.

$$\Delta G = \Delta H - T\Delta S = -RT\ln K_a \qquad \textbf{(Equation 4.1)}$$

$$K_d = 1/K_a \qquad \textbf{(Equation 4.2)}$$

Enthalpy is the change in energy as heat and is positive for endothermic reactions or negative in exothermic reactions, reflecting both the breakage and formation of non-covalent bonds. Entropy is traditionally described as a measure of the disorder or randomness in a system and is the subject of the second law of thermodynamics that states the universe tends towards maximum disorder. Increased disorder is reflected by a greater ΔS value (often reported as the term from **Equation 4.1**, -TΔS) and appears driven by desolvation events associated with hydrophobic interactions and loss of conformational freedom on complex formation, both of ligand and protein. The free energy is an overall measure of the spontaneity of a reaction and, as **Equation 4.1** states, is dependent on the enthalpic and entropic events as are the affinity descriptors, K_a and K_d.

In order to accurately determine these thermodynamic properties from one experiment, the final curve (isotherm) produced must be the optimum shape. This is signified by the unitless parameter, *c*, defined as the product of the association constant and the total number of binding sites (**Figure 4.4B**; Wiseman *et al*., 1989; Pierce *et al*., 1999; Broecker *et al*., 2011). In a tight-binding system, low macromolecule concentrations are necessary or the curve may be too sharp, producing errors in the derived K_a. Conversely, for a weaker binding interaction, the full sigmoidal curve may not be

observed and the experimental ΔH will not reflect the accurate value, nor can the entropic contribution to the overall free energy be extracted. A window of approximately 1-1000 represents an ideal *c* value (Wiseman *et al*., 1989) while a more conservative upper limit of 500 (Turnbull, 2011) or an optimum of approximately 40 (Broecker *et al*., 2011) are also recommended. It is therefore imperative that experimental parameters are appropriate and in some cases, a compromise between optimum theoretical sample concentrations and achieving measurable signal must be met (Biswas and Tsodikov, 2010).

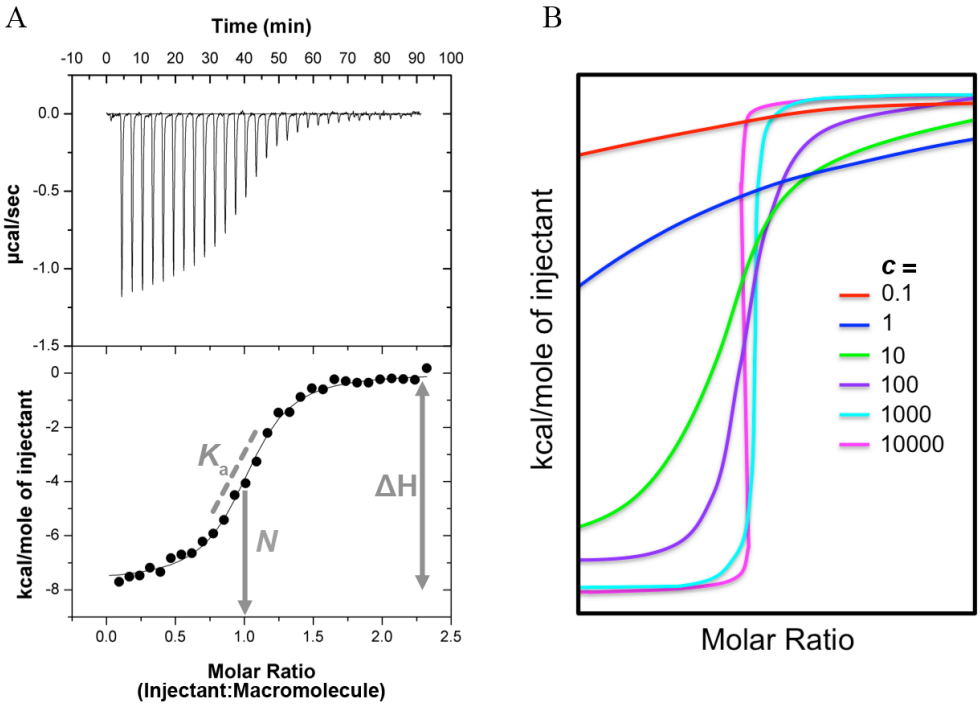


Figure 4.4. The ITC isotherm

(A) An experimental ITC isotherm. The upper panel displays the raw data. Each peak represents a single injection of titrant. The lower panel shows the integrated peaks to which the best-fitting curve is modeled. K_a, *N* and ΔH are measured directly. **(B)** Schematic to emphasise the shape of the binding isotherm determined by the unitless parameter, *c*. This figure is a modified version of that in Wiseman *et al*., 1989.

4.2.3. ITC and crystallography

Drug discovery programs have traditionally focussed on improving the affinity of inhibitors (Leavitt and Freire, 2001). Other important properties of an inhibitor

molecule must also be considered (Ladbury *et al*., 2010). This includes the mechanism of inhibition, pharmacokinetic profile as well as production concerns such as simplicity and cost. Contributing to an appreciation of such properties, structural information can be utilised to build a structure-activity relationship (SAR) and guide chemical optimisation (Congreve *et al*., 2005; Hunter, 2009) while an experimentally derived thermodynamic profile can provide insight into a binding system that predictions, structural models or measurements of potency alone do not (Salemme *et al*., 1997; Weber and Salemme, 2003; Chaires, 2008). Ligands with identical affinities can have drastically different enthalpic and entropic contributing components (Cooper *et al*., 2001; Li *et al*., 2008). While interpretation may not be trivial (Cooper *et al*., 2001; Chaires, 2008), it is essential to consider the thermodynamic parameters and to relate them to high resolution structures to understand any correlation between binding energetics and structure (Leavitt and Freire, 2001; Ladbury, 2010). As such, an approach exploiting multiple biophysical techniques is important to heighten the overall insight of a ligand-binding event and aid in the design of effective inhibitors.

4.3. Aims

With a combined use of crystallography, a kinetic inhibition assay and ITC, we aim to characterise the atomic details of PTR1-ligand interactions and identify molecular features that have the greatest contribution to affinity. While structural data are available for PTR1 in a number of ligand complexes, previous experiences demonstrated that crystals of the *L.major* enzyme suffered from poor reproducibility, typically produced only medium resolution X-ray diffraction and were often of a form not optimal for inhibitor studies (W.N.Hunter, personal communication). We therefore

looked to the *L. donovani* and *T. brucei* enzymes to provide an improved model for inhibitor studies. *Tb*PTR1 then provides the basis for work undertaken as part of a collaboration to examine a series of novel potential inhibitory compounds with a view to establishing SAR for the series, to identify new interaction features and contribute to future inhibitor design efforts. The use of thermodynamic binding profiles has not yet been utilised in a wide inhibitor study for this enzyme and we set out to assess the value of the technique and to generate ligand binding profiles to guide inhibitor optimisation. PTR1 therefore not only represents a useful target to investigate inhibitors from a drug design perspective but also an interesting model to examine protein-ligand interactions.

PART II

Pteridine reductase 1

5. Materials and methods

5.1. General materials and methods

All general materials and methods are as provided in **Part I Chapter 2**. In this chapter, any protocol modifications or additional details specific to studies involving the enzyme pteridine reductase 1 (PTR1) are described.

5.2. PTR1 ligands

PTR1 cofactors $NADP^+$ and NADPH were purchased from Melford and all pterin substrates were obtained from Schircks Laboratories. Non-commercial potential PTR1 ligand molecules were synthesised by Judith Huggan and Abedawn Khalaf (University of Strathclyde, Glasgow) or obtained from the Center for Organic and Medicinal Chemistry, Research Triangle International (North Carolina, USA). Ligands were prepared in 100 % (v/v) dimethyl sulfoxide (DMSO) to 100 mM based on dry weight. The insolubility of any compound at this concentration was established following the application of sonication and heat (42°C). Solubilised compounds were stored in glass vials at -20°C and 10-fold serial dilutions (in DMSO) made as working stocks.

5.3. Gene expression and protein purification

5.3.1. Leishmania donovani PTR1

A plasmid containing the gene encoding PTR1 from *L. donovani*, originally cloned from gDNA, was obtained from W.N.Hunter. The gene of interest was subcloned into pET15b-TEV using restriction endonucleases *Nde*I and *BamH*I to allow the hexa-histidine tag of the protein product to be cleavable by TEV protease. BL21 (DE3) GOLD cells transformed with this plasmid were cultured in LB as described in **2.3.1**.

Purification was also similar to that described in **2.3.2** using buffers containing 50 mM Tris-HCl pH 7.7, 200 mM KCl and 20-800 mM imidazole. No size exclusion chromatography step was performed. Cleaved *Ld*PTR1 was exchanged to a solution of 20 mM Tris-HCl pH 7.7 and concentrated to 5-10 mg mL^{-1}.

5.3.2. Trypanosoma brucei PTR1

The *T. brucei* PTR1 expression plasmid, pET15b-*Tb*PTR1 was obtained from W.N.Hunter (Dawson *et al*., 2006). Recombinant expression was carried out as for *Ld*PTR1 above. *Tb*PTR1 was then purified by a single affinity chromatography step in standard buffers (**Table 2.4**). Tagged protein was exchanged using a PD-10 desalting column (GE Healthcare) to 20 mM Tris-HCl pH 7.5 for crystallisation or 50 mM Tris-HCl pH 7.5, 250 mM NaCl, 20 % (v/v) glycerol for kinetic experiments. Following concentration to 15-20 mg mL^{-1} in a centrifugal concentrator with 100 kDa molecular mass limit, aliquots of *Tb*PTR1 were flash cooled in liquid nitrogen and stored at -80°C until required. For ITC, *Tb*PTR1 was freshly prepared and stored at 4°C in 20 mM sodium citrate pH 3.7 for a maximum of 7 days.

5.4. Spectrophotometric assay

Inhibition of *Tb*PTR1 activity was carried out according to an established spectrophotometric assay (Dawson *et al.*, 2006). In brief, 1 mL samples containing 30 μg *Tb*PTR1 (0.96 μM) , 20 μM dihydrobiopterin and 0-1 mM compound of interest in 20 mM sodium citrate pH 3.7 were warmed to 30°C in acrylic cuvettes (Sarstedt). Each reaction was started by the addition of 100 μM NADPH. The decrease in absorbance at 340 nm was monitored for 120 s using a Shimadzu UV-2450 spectrophotometer

coupled to *UVProbe* (Shimadzu). Dihydrobiopterin stocks were prepared in 0.1 M NaOH to 10 mM and stored at -20°C. Fresh NADPH was dissolved in water to 10 mM immediately prior to each set of approximately 50 measurements. DMSO to a final concentration of 1 % (v/v) was present throughout. All potential inhibitors were initially assessed in duplicate at 10 μM and 50 μM. Compounds showing inhibition of greater than 60-70 % at 50 μM were assayed in triplicate at a range of concentration points (9-12) to produce a full dose-response curve. Mean PTR1 inhibition (%) was plotted against the log compound concentration and a sigmoidal curve fit using four-parameter non-linear regression in *SigmaPlot* (Systat Software). IC_{50} values were extracted and K_i values calculated using the Cheng-Prusoff equation for competitive inhibition where $K_i = IC_{50} / (1 + [S]/K_m)$ (Cheng and Prusoff, 1973). Note that this assumes stoichiometry of 1:1 and that all inhibitor binding reactions are reversible. Substrate K_m was used as published (10.9 μM; Dawson *et al.*, 2006).

5.5. Isothermal titration calorimetry

ITC experiments were all carried out using a VP-ITC system (MicroCal, GE Healthcare) controlled by the software, *Origin5* (OriginLab). 20-100μM *Tb*PTR1 was prepared in the 1.5 mL sample cell and titrated with 0.1-1 mM compound. Compounds were prepared from 100 mM stocks, diluted to the desired concentration using buffer identical to that of the protein sample. Both samples in 20 mM sodium citrate pH 3.7 were supplemented with 1 mM $NADP^+$ and DMSO to a final concentration of 5 % (v/v). An initial injection of 2 μL to remove any residual air from the syringe was followed by 29-39 injections of 6-8 μL titrant at 3 minute intervals (0.5 μL s^{-1} injection rate). The injection needle also acted as a stirring paddle, mixing the sample cell

contents at 300 rpm. Heat evolved due to the addition of compound was measured as the change in power required to maintain the sample cell at 30°C with respect to the reference cell. The baseline was determined automatically and manually adjusted where necessary. Raw data peaks were integrated and iteratively fit to the OneSite curve model. No values were fixed in the first instance. The association constant (K_a), stoichiometry (N) and change in enthalpy (ΔH) were extracted from the best-fitting curve and the entropic change (ΔS) calculated. For a number of compounds, the ligand concentration input values were retrospectively adjusted to reflect a 1:1 stoichiometry. Control titrations were performed when possible in the absence of protein by injecting compound into buffer alone and the resultant integrated heats subtracted from the main experiment data. A standard titration of 30 injections of 0.5 mM MTX (8 μL) into 40 μM *Tb*PTR1 was performed to verify the activity and quantitation of each fresh protein preparation. Each experiment was repeated 2-3 times and all analyses were carried out using *Origin5*.

5.6. *Ld*PTR1 crystallisation and structure solution

Crystals of *Ld*PTR1 were grown by hanging-drop vapour diffusion at 18°C. Crystallisation drops contained an equal volume of 5 mg mL^{-1} *Ld*PTR1 in 20 mM Tris-HCl pH 7.7, 1 mM $NADP^+$, 20 mM DTT and 1 mM MTX mixed with 0.1 M MES pH 6.5, 1.6 M $(NH_4)_2SO_4$ and 10 % (v/v) dioxane.

A nylon loop was used to harvest suitable crystals that were then passed through a cryoprotectant solution containing 40 % (v/v) PEG 400 and cooled to -173°C in a stream of nitrogen gas. Crystals were screened for diffraction in-house using X-ray

source A (**Table 2.6**). The best diffracting crystal was manually annealed (the cryostream was physically blocked for 5-10 s) and diffraction data collected at source F (**Table 2.6**).

Diffraction data were integrated using *MOSFLM* (Leslie, 2006) and scaled in *SCALA* (Evans, 2006). Molecular replacement by *Phaser* (McCoy *et al.*, 2007) was performed using a single monomeric chain of the published *L. major* PTR1 structure as the search model (PDB 1e7w; Gourley *et al*., 2001). *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010) was used to manipulate the structural model and in the placement of solvent molecules. Refinement was carried out using *REFMAC5* (Murshudov *et al*., 2011).

5.7. *Tb*PTR1-ligand co-crystallisation and structure solution

*Tb*PTR1 was crystallised in the presence of cofactor and ligand of interest following published methods (Tulloch *et al*., 2010). A mixture of 4-6 mg mL^{-1} *Tb*PTR1, 1 mM $NADP^+$, 20 mM DTT and 1 mM compound was incubated on ice for 1 h prior to crystallisation. The solution was centrifuged to remove any insoluble material. 1-2 μL protein-ligand solution was then mixed with 1 μL crystallisation solution containing 1.7-2.7 M NaOAC and 20-50 μM sodium citrate pH 4.5-5.0. Drops were suspended on siliconised glass coverslips above reservoirs containing the latter solution at 18°C.

Crystals grown in an excess of 2.6 M NaOAC were placed in a nylon loop and flash-cooled directly in liquid nitrogen. Otherwise, a solution of 3 M NaOAC was used to cryoprotect the *Tb*PTR1-ligand co-crystal prior to cooling to -173°C. Diffraction data were collected in-house or at Diamond light source (sources A, B, C, E and F, **Table**

2.6). Data were collected to 1.7-2.4 Å resolution and typically covered approximately 180 degrees of rotation. X-ray images were integrated using *XDS* (Kabsch, 2010) or *MOSFLM* (Leslie, 2006; Battye *et al.*, 2011) and scaled in *SCALA* (Evans, 2006) or *AIMLESS* (Evans and Murshudov, 2013). Molecular replacement was performed by *MOLREP* (Vagin and Teplyakov, 1997) or data were refined directly against the *Tb*PTR1 tetrameric starting model using *REFMAC5* (PDB 2c7v; Dawson *et al*., 2006; Murshudov *et al*., 2011). Electron and difference density map inspection and model manipulation with *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010) was combined with multiple further refinement cycles. Non-crystallographic symmetry restraints were not applied and geometry restraint weightings were manually adjusted in later refinement calculations. The subset of data used to calculate R_{free} was maintained in all structure analyses. Cofactor and solvent molecules were added to the model in *Coot* (Emsley and Cowtan, 2004; Emsley *et al.*, 2010) from the associated monomer library (Vagin *et al.*, 2004). Simple amino acid modifications were incorporated in the same manner. Novel ligands were drawn using *JME Molecule Editor* (Ertl, 2010) and coordinates generated using *PRODRG* (Schüttelkopf and van Aalten, 2004). Geometry restraints were obtained from *PRODRG* (Schüttelkopf and van Aalten, 2004) or *eLBOW* (Moriarty *et al.*, 2009) and each ligand placed according to difference density Fourier maps. For compounds SDG 33 and SDG 65, two conformations were modelled and refined, each with occupancy set to 0.5.

It should be noted that a script known colloquially as *BigRedButton*, written by Paul K. Fyfe, was developed using feedback from usage with these data and proved especially useful for initial data and model assessment. This script called on *xia2* (Winter, 2009) to process data with *XDS* (Kabsch, 2010) then *MOLREP* (Vagin and Teplyakov, 1997)

and *REFMAC5* (Murshudov *et al*., 2011), producing interpretable maps and an initial model directly from X-ray images. The *Tb*PTR1 active site was visualized and ligand presence quickly evaluated before proceeding with further data collection or more comprehensive data processing procedures.

Structural analysis was carried out using programs described in **2.8.5** and additional software are cited directly where applicable.

PART II

Pteridine reductase 1

6. Results and discussion

6.1. *Leishmania donovani* pteridine reductase 1

6.1.1. Protein production, crystallisation and structure solution

Recombinant *Ld*PTR1 was produced as described in **5.3**. Following initial crystallisation screening in sitting-drop format, optimised crystals formed by hanging-drop vapour diffusion within 2-3 days at 18°C (**5.6**). Crystals grew only in the presence of cofactor, $NADP^+$, and inhibitor, MTX. Maximum crystal size was approximately 0.1 x 0.1 x 0.05 mm. Attempts to crystallise *apo*-*Ld*PTR1 under these conditions, or to find new suitable conditions, were unsuccessful. Initial X-ray diffraction was improved by annealing the crystal. Further details of data collection and subsequent structure refinement statistics are provided in section **5.6** and **Table 6.1**. Crystals display space group $C222_1$ with unit cell lengths of a = 107.51 Å, b = 126.44 Å and c = 87.51 Å. A Matthews coefficient (V_M) of 2.49 $Å^2$ Da^{-1} (Matthews, 1968) corresponded to a crystal solvent content of approximately 50 % and the presence of two *Ld*PTR1 subunits per asymmetric unit. The structure of *Ld*PTR1, assigned PDB accession 2xox, was determined to 2.5 Å resolution by molecular replacement using a published *Lm*PTR1 structure as the search model (PDB 1e7w; Gourley *et al*., 2001).

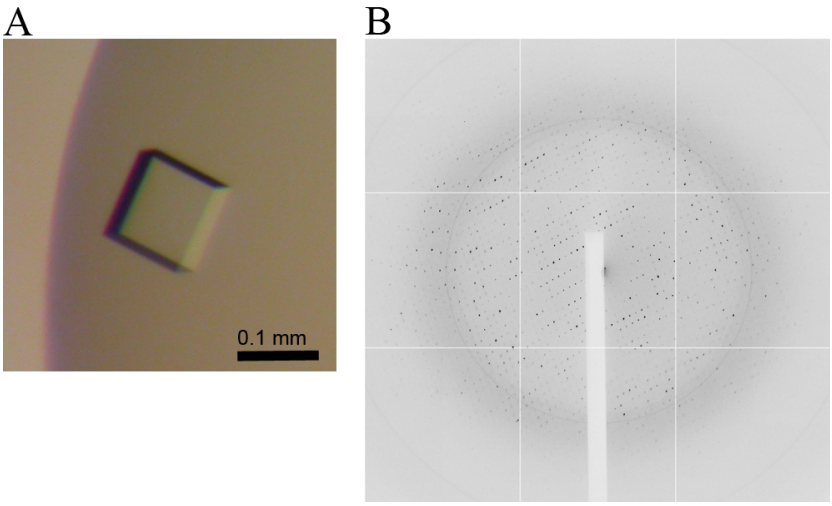


Figure 6.1. *Ld*PTR1 crystal and diffraction
(A) *Ld*PTR1 crystal. **(B)** An example diffraction image (2.5 Å resolution at the image edge).

Resolution range (Å)	32.0-2.5 (2.6-2.5)
Space Group	$C222_1$
Unit cell parameters	a = 107.51, b = 126.44, c = 87.51 Å $\alpha = \beta = \gamma = 90°$
Wavelength (Å)	0.973
No. Reflections	144562 (21139)
No. Unique Reflections	21006 (3022)
R_{merge} [a] (%)	9.9 (42.1)
R_{pim} [b] (%)	4.1 (17.1)
Completeness (%)	99.9 (100.0)
Mean $I/\sigma(I)$	11.0 (3.4)
Redundancy	6.9 (7.0)
Wilson B factor (Å^2)	47.7
R_{work} [c] (%)	22.8
R_{free} [d] (%)	28.5
R.m.s.d bonds (Å)	0.019
R.m.s.d angles (°)	1.766
Total protein residues	458
Total protein atoms	3201
Average protein B factor (Å^2)	43.4
DPI [e] (Å)	0.396
Ramachandran plot:	
Favoured (%)	95.2
Allowed (%)	4.6
Outliers (%)	0.2
Additional groups:	
Solvent (No./Average B (Å^2))	24 / 38.0
Sulfate (No./Average B (Å^2))	2 / 53.4

Table 6.1. *Ld*PTR1 data collection and refinement statistics
Values in parentheses correspond to the highest resolution shell (2.6–2.5 Å). [a.] $R_{merge} = \sum h \sum i |(h,i) - \langle I(h) \rangle \sum h \sum i\, I(h,i)$; where $I(h,i)$ is the intensity of the ith measurement of reflection h and $\langle I(h) \rangle$ is the mean value of $I(h,i)$ for all i measurements. [b.] R_{pim}, precision-indicating merging R-factor, is R_{merge} adjusted by a factor of $\sqrt{(1/n-1)}$ where n is the number of times a given reflection is observed. [c.] $R_{work} = \sum hkl||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor amplitude and the F_c is the structure-factor amplitude calculated from the model. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. [e.] DPI, diffraction-component precision index (Cruickshank, 1999).

6.1.2. LdPTR1 overall structure

This crystal form of *Ld*PTR1 has two subunits in the asymmetric unit (A and B, **Figure 6.2A**). These subunits are highly conserved structurally with an r.m.s.d. of 0.51 Å when 212 Cα atoms are overlaid. Unless otherwise indicated, *Ld*PTR1 subunit A was used for all further structural analysis. PTR1 is a tetrameric enzyme and this is generated by

the symmetry operation –x, y, -z + ½ (a 2_1 screw axis parallel to *c*). Each monomer presents as a classical Rossmann fold (Rao and Rossmann, 1973; Gourley *et al*., 2001). Seven parallel β-strands are sandwiched between two sets of three α-helices (**Figure 6.2B**). Several loops have residues that could not be modelled, located at β3-α3, β4-α4 and β6-α6 in addition to extreme *N*- and *C*- terminal residues.

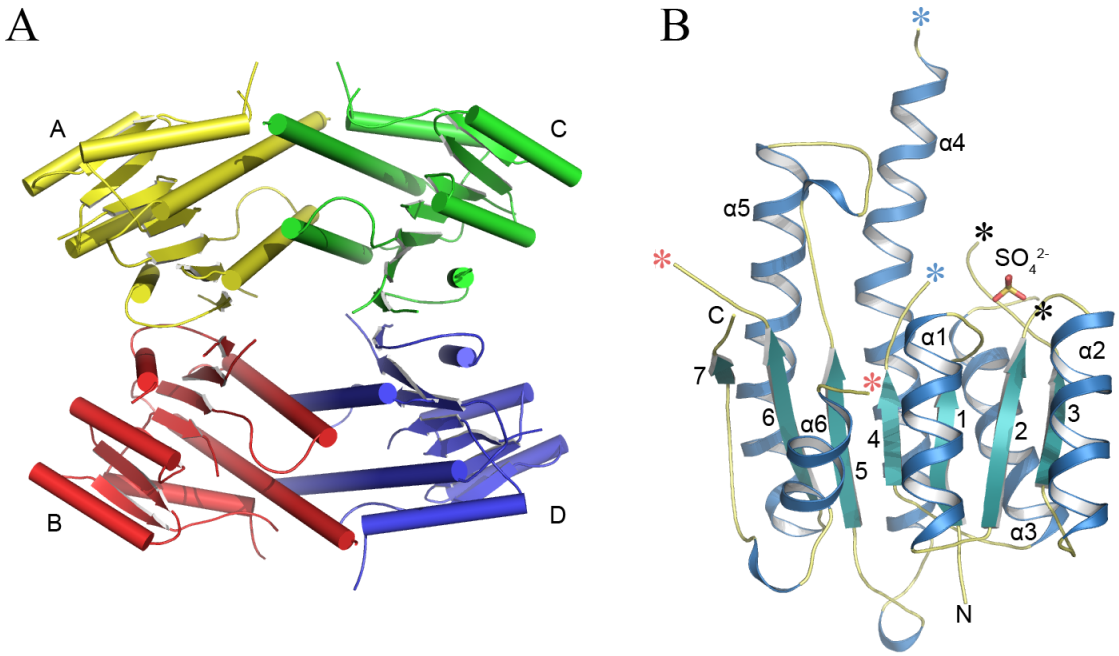


Figure 6.2. *Ld*PTR1 overall structure
(A) A cartoon depiction of the *Ld*PTR1 tetramer. Subunits A and B alone comprise the crystallographic asymmetric unit. Coloured according to chain, β-strands are shown as arrows and α-helices as cylinders. **(B)** An *Ld*PTR1 monomer. β-strands are shown as cyan arrows and are labelled 1-7. Helices α1-α6 are labelled (blue ribbons) while potential 3_{10} helices in loops α4-β5, β5-α5 and α6-β7 are unlabelled. Breaks in the yellow coloured loops are signified by black, blue and red coloured asterisks. A sulfate ion is shown as sticks in the active site (S, yellow; O, red). Secondary structure features were assessed using *STRIDE* (Frishman and Argos, 1995; Heinig and Frishman, 2004).

6.1.3. A disordered LdPTR1 active site

Amino acids that create the PTR1 active site are strictly conserved between *Ld*PTR1 and *Lm*PTR1 (Gourley *et al*., 2001). Sequence comparisons also show that these residues are highly similar to those involved in binding the cofactor, substrate and product in *Tb*PTR1 (Dawson *et al*., 2006). In contrast to *Lm*PTR1 and *Tb*PTR1,

however, the *Ld*PTR1 structural model presented displays a poorly ordered active site region. The electron density does not indicate the presence of $NADP^+$ or MTX, despite their presence in excess during crystallisation. MTX is a known inhibitor of PTR1 with a K_i of 39 ± 19 nM and 152 ± 16 nM against the *L. major* and *T. brucei* enzymes, respectively (Dawson *et al.*, 2006).

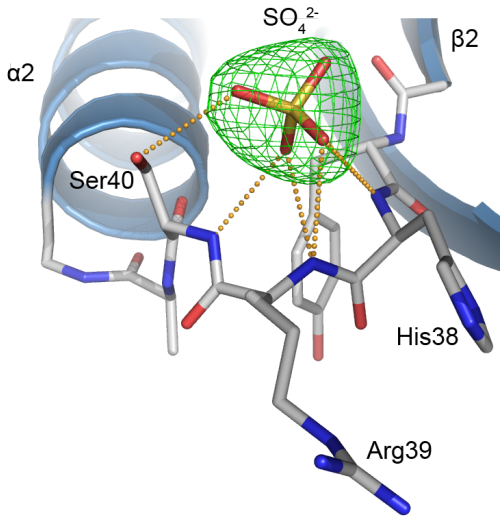


Figure 6.3. Sulfate bound in the *Ld*PTR1 active site
A sulfate ion (yellow and red sticks) is bound to *Ld*PTR1, making multiple contacts with conserved β2-α2 loop residues, His38, Arg39 and Ser40, shown as sticks coloured by element (C, grey; N, blue; O, red). Potential hydrogen bonds are indicated as orange dashed lines with lengths of 2.7-3.6 Å. The F_o-F_c difference density omit map, calculated by omitting sulfate from the final model, is shown as green mesh contoured at 3.5 σ,.

A sulfate ion was modelled into a tetrahedral mass of electron density close to the short β2-α2 loop (**Figure 6.3**). The density did not extend beyond the spatial radius of the single sulfate molecule. In *Lm*PTR1 and *Tb*PTR1, the cofactor adenine 2'-phosphate group is located in the same polar cavity that is formed by the β2-α2 loop and the turn connecting β1 and α1. The sulfate forms hydrogen bonds with three main chain amides, His38, Arg39 and Ser40, and the hydroxyl side group of Ser40. These three residues are identical in *Lm*PTR1 and form similar contacts with the cofactor phosphate. Owing to the high concentration of ammonium sulfate (1.6 M) required to obtain these

*Ld*PTR1 crystals, sulfate appears to have displaced any $NADP^+$ that may otherwise have been bound in solution.

During the catalytic cycle, NADPH must first bind, followed by substrate. Product and oxidised cofactor then dissociate sequentially. The same mechanism of binding is described for inhibitors (Luba *et al*., 1998; Gourley *et al*., 2001). The location of the sulfate in this *Ld*PTR1 model restricts the access of NADP(H) and thus, the ability for an additional ligand molecule to join the binary complex. Critical residues Phe113 and Arg17 are amongst those found disordered in *Ld*PTR1. Phe113, in the missing β4-α4 loop, is normally involved in stabilising the ligand position within the catalytic site. It is required to form π-stacking interactions along with the cofactor nicotinamide and ligand (Gourley *et al*., 2001; Dawson *et al*., 2006). The disordered side chain of Arg17 affects the binding of the cofactor pyrophosphate (McLuskey *et al*., 2004). Other key catalytic residues are Asp181, Tyr194 and Lys198. Tyr194 is oriented similar to other published PTR1 structures. However, Asp181 is in the relocated β5-α5 loop while weak electron density suggests the placement of the Lys198 side chain is unsuitable to form stabilising interactions previously observed with either the cofactor ribose or neighbouring residues Ser111 and Asn147 (Gourley *et al*., 2001; Schüttelkopf *et al*., 2005).

6.1.4. LdPTR1 and LmPTR1 structure comparison

Overall, *Ld*PTR1 and the *Lm*PTR1 model used for molecular replacement share over 90 % sequence identity. This is translated to an r.m.s.d of 1.3 Å when 202 Cα atoms are aligned in *Coot* (Emsley and Cowtan, 2004; **Figure 6.4**). There are, however, several

important differences between the crystal structures. Residues 227-254 remain unmodelled in this structure of *Ld*PTR1, which includes the substrate-binding loop (Tulloch *et al*., 2010). Additionally, a short α-helix between β6 and α6 is also unaccounted for in this region. In *Lm*PTR1, a total of seven α-helices are present, with residues 235-244 forming α6. Indeed, secondary structure prediction by the *PSIPRED* server (Buchan *et al*., 2010) suggests that residues 235-243 should form an α-helix in *Ld*PTR1.

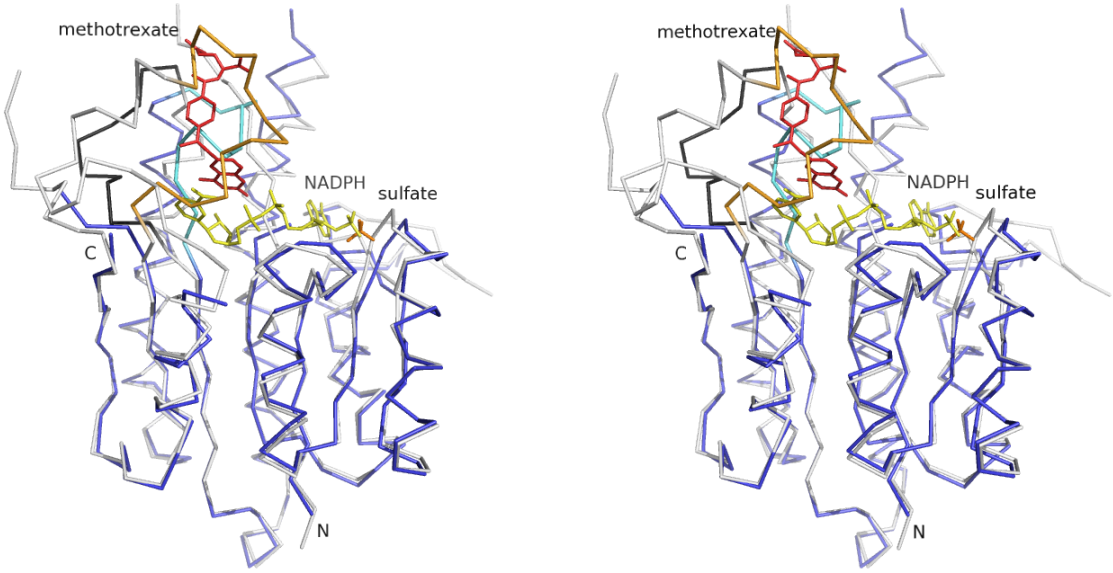


Figure 6.4 *Ld*PTR1 and *Lm*PTR1 backbone comparison
A stereo diagram of *Ld*PTR1 (blue) and *Lm*PTR1 (grey) backbone traces overlaid. The sulfate bound to *Ld*PTR1 is shown as orange sticks while NADPH and methotrexate bound to *Lm*PTR1 are coloured yellow and red, respectively. The substrate-binding loop of *Lm*PTR1, absent in this structure of *Ld*PTR1, is highlighted orange. The β5-α5 loop is repositioned in *Ld*PTR1 (highlighted in cyan) with respect to the *Lm*PTR1 loop (black).

As mentioned briefly, the loop connecting β5 and α5 displays a large change in position between *Lm*PTR1 and *Ld*PTR1. Distances of up to 16 Å are observed for residues topologically equivalent to *Ld*PTR1 Val180-Gly190. The repositioned loop overlaps with both the cofactor nicotinamide and the pteridine-like moiety of MTX. A binary complex of *Lm*PTR1 with NADPH (PDB 2bfo; Schüttelkopf *et al*., 2005) displays the same overall conformation as the ternary complex used in comparisons. The same

loops are also relocated with respect to *Tb*PTR1 complexed with $NADP^+$ and MTX (PDB 2c7v; Dawson *et al*., 2006) which shares approximately 53 % amino acids to *Ld*PTR1. It is possible that the placement of this loop in the active site prevented complex formation. Equally, the lack of ligand may have allowed for an increase in flexibility in the region and the loop is able to adopt a different position within the vacant active site. Further inhibitor studies were unfortunately not possible based on this crystal form.

6.2. *Trypanosoma brucei* pteridine reductase 1

6.2.1. Compound details

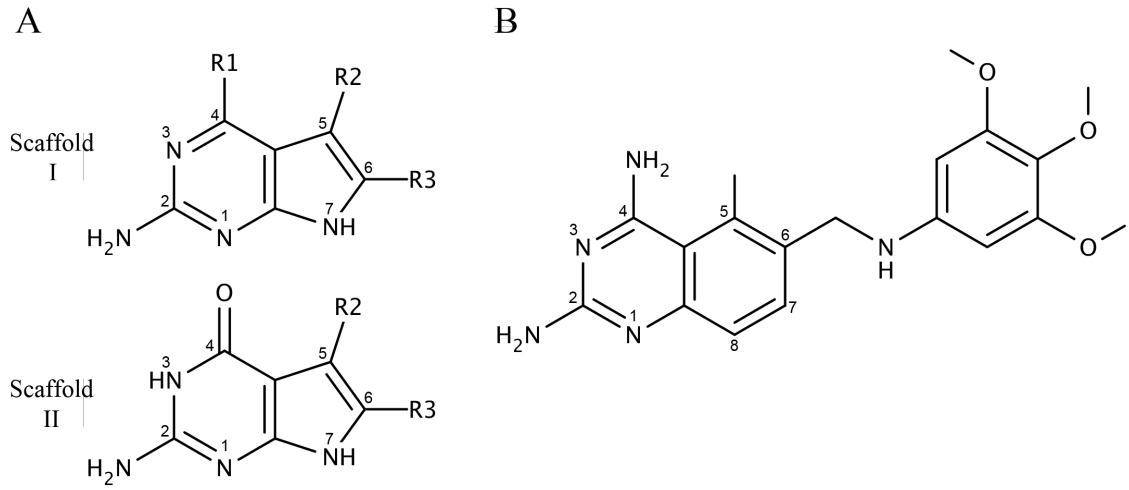


Figure 6.5. Compound scaffolds
(A) The majority of compounds tested were built around two pyrrolopyrimidine derivatives with substitutions made at the R1, R2 and R3 groups (positions 4, 5 and 6). **(B)** Chemical structure of trimetrexate, 5-methyl-6-[(3,4,5-trimethoxyphenyl)aminomethyl]quinazoline-2,4-diamine.

Compounds were synthesised by collaborators at the University of Strathclyde (named SDG; **S**trathclyde **D**undee **G**lasgow) and the Research Triangle Institute (trimetrexate). Based on a scaffold previously known to make important interactions with key amino acids (Tulloch *et al*, 2010), a variety of chemical substitutions were made at three

positions of two pyrrolopyrimidine scaffolds (**Figure 6.5A**). A number of the assessed compounds fall outwith this framework but all of those detailed structurally are derived from the two core scaffolds. A single exception is the known DHFR inhibitor, trimetrexate (TMQ; **Figure 6.5B**). SDG compound structures are provided in **Appendix A**.

6.2.2. Spectrophotometric assay

All SDG compounds soluble to 100 mM in 100 % (v/v) DMSO were screened in duplicate at two concentration points, 10 μM and 50 μM, against 30 μg mL^{-1} *Tb*PTR1 (0.96 μM). 30 compounds that failed to dissolve adequately at the desired concentration were immediately rejected from further experiments and a total of 102 compounds were assessed. Inhibition was calculated as a percentage where 0 % inhibition was measured in the absence of inhibitor. Background NADPH oxidation was subtracted from all measurements.

As described in **5.4**, compounds displaying at least 60-70 % inhibition of *Tb*PTR1 at 50 μM were analysed across a range of concentrations, typically from 0.025 μM to 100 μM. K_i values are provided for 54 compounds in **Table 6.2** and two representative dose-response curves are shown in **Figure 6.6**. All reported K_i values were obtained under the experimental conditions described in **5.4**, based on no prior knowledge of structural binding modes and assume reversible competitive inhibition with a stoichiometry of 1:1. The K_i of MTX was measured as 0.17 μM in agreement with the published value of 0.15 μM (Dawson *et al*., 2006). Kinetic results are discussed in more detail in **6.2.6** and **Table 6.2** is referred to throughout this chapter.

Compound	R1	R2	R3	K_i (μM)
		--- Scaffold I ---		
SDG 4	NH_2	CN	Br	3.32 ± 0.13
SDG 5 [a]	NH_2	CN	C_6H_4COH	0.20 ± 0.01
SDG 7	$NHCH_3$	H	$CO_2C_2H_5$	9.88 ± 0.10
SDG 10	$NHCH_2C_6H_{11}$	CN	H	1.62 ± 0.02
SDG 23	NC_4H_8S	CN	H	8.75 ± 0.16
SDG 32	NC_4H_8	H	H	4.17 ± 0.07
SDG 33	NC_4H_8S	H	H	8.64 ± 0.09
SDG 53	NC_4H_8	CCC_6H_5	H	0.19 ± 0.01
SDG 65	NH_2	$C_6H_4CH_3$	H	0.32 ± 0.01
SDG 67	NH_2	C_6H_5	H	0.40 ± 0.01
SDG 68	NH_2	C_6H_4F	H	0.48 ± 0.01
SDG 69	NH_2	CN	H	4.87 ± 0.06
SDG 72	NC_4H_8	CN	H	0.80 ± 0.02
SDG 73 [a]	NC_4H_8	CN	C_6H_4COH	0.20 ± 0.01
SDG 74 [b]	NH_2	CN	$CHCHC_6H_5$	0.16 ± 0.01
SDG 76	NH_2	CN	CCC_6H_5	0.24 ± 0.01
SDG 77 [b]	NH_2	CN	$CHCHC_6H_5$	0.34 ± 0.01
SDG 80	NH_2	$CH_2CH_2C_6H_5$	H	0.26 ± 0.01
SDG 82	NH_2	CN	$CHCHC_6H_4CH_3$	0.27 ± 0.01
SDG 84	NH_2	CN	$CH_2CH_2C_6H_5$	0.35 ± 0.01
SDG 100	NH_2	C_6H_5	C_6H_5	0.59 ± 0.01
SDG 106	NHC_6H_{11}	C_6H_4F	H	0.56 ± 0.01
SDG 107	NHC_6H_{11}	C_6H_5	C_6H_5	0.20 ± 0.01
SDG 112	NH_2	C_6H_5	C_6H_4F	0.24 ± 0.01
SDG 114	NH_2	C_6H_4F	C_6H_4F	0.30 ± 0.01
SDG 120	$N(CH_3)CH_3$	C_6H_5	C_6H_5	0.29 ± 0.01
SDG 122	NH_2	$C_6H_4OCH_3$	C_6H_4F	0.58 ± 0.01
SDG 123	$N(CH_3)CH_3$	$C_6H_4OCH_3$	C_6H_4F	0.30 ± 0.01
SDG 127	NH_2	C_6H_5	C_6H_4Br	0.14 ± 0.01
SDG 130	NH_2	C_6H_5	$C_6H_4CH_2CH(CH_3)CH_3$	0.58 ± 0.01
SDG 132	NH_2	C_6H_5	$C_6H_4SO_2CH_3$	1.28 ± 0.02
SDG 134	NH_2	C_6H_5Cl	C_6H_5F	0.29 ± 0.01
		--- Scaffold II ---		
SDG 57	-	CH_3	H	7.33 ± 0.13
SDG 60	-	$C_6H_4CH_3$	H	1.21 ± 0.03
SDG 61	-	C_6H_4F	H	1.26 ± 0.01
SDG 62	-	C_6H_5	H	1.18 ± 0.01
SDG 70	-	$CH_2NHC_6H_4OCH_3$	H	7.08 ± 0.06
SDG 71	-	$CH_2NHC_6H_4Cl$	H	2.53 ± 0.06
SDG 75	-	CN	$CH_2CH_2C_6H_5$	0.12 ± 0.01
SDG 81	-	$CH_2CH_2C_6H_5$	H	0.27 ± 0.01
SDG 99	-	C_6H_5	C_6H_5	1.17 ± 0.02
SDG 102	-	CH_3	C_6H_5	1.06 ± 0.02
SDG 111	-	C_6H_5	C_6H_4F	0.51 ± 0.01
SDG 113	-	C_6H_4F	C_6H_4F	0.76 ± 0.02
SDG 115	-	C_6H_4Cl	C_6H_4Cl	0.25 ± 0.01
SDG 126	-	C_6H_5	C_6H_4Br	0.23 ± 0.01
SDG 128	-	$CH_2CH_2C_6H_5$	C_6H_5	0.95 ± 0.02
SDG 133	-	C_6H_4Cl	C_6H_4F	0.47 ± 0.01

Table 6.2. (Part 1)

Compound	R1	R2	R3	K_i (μM)
		--- Scaffold I* ---		
SDG 54 [c]	NC_4H_8	H	H	4.23 ± 0.06
SDG 85 [d]	NH_2	$CO_2C_2H_5$	C_6H_5	0.77 ± 0.02
SDG 88 [d]	NH_2	$CO_2C_2H_5$	$C_6H_5SCH_3$	1.08 ± 0.04
SDG 89 [d]	NH_2	$CONHC_6H_4Cl$	H	1.09 ± 0.02
SDG 91 [d]	NH_2	$CONHC_6H_4C_2H_5$	H	0.97 ± 0.02
SDG 93 [d]	NH_2	$CONHC_6H_4SCH_3$	H	0.68 ± 0.01

Table 6.2. *Tb*PTR1 inhibition

SDG compounds are categorised by scaffold and listed according to name. Crystal structures were obtained of *Tb*PTR1 containing compounds highlighted in grey rows. Scaffold I* consists of compounds that contain an additional modification to scaffold I, distinct from the R-group substitutions. [a.] SDG 5 and SDG 73 were subsequently discovered to bind covalently to *Tb*PTR1 under the conditions used for crystallisation and values here were calculated assuming reversible inhibition. [b.] SDG 77 is the (*Z*)-isomer of SDG 74 (*E*). [c.] $NHOCF_3$ replaces position 2 NH_2 of scaffold I. [d.] O replaces position 7 NH of scaffold I.

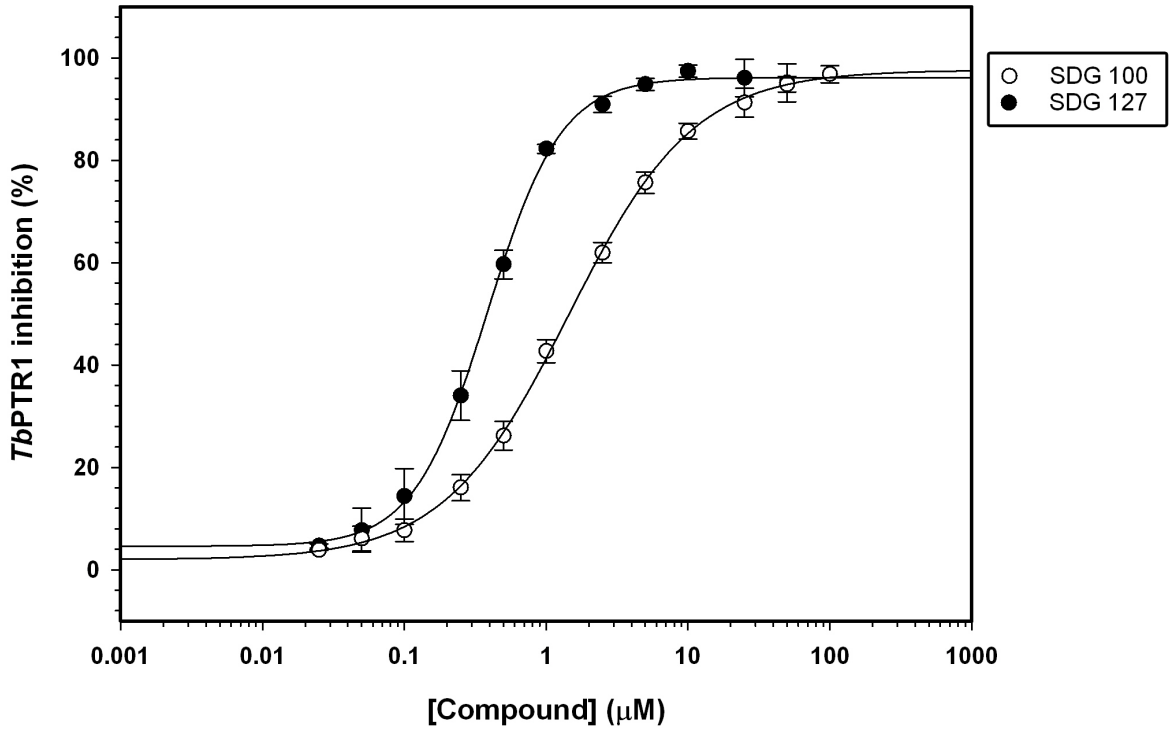


Figure 6.6. *Tb*PTR1 inhibition

Inhibition of *Tb*PTR1 by SDG 100 (open circles) and SDG 127 (closed circles). Each data point represents the mean of three replicates and error bars denote the standard deviation of the same three measurements.

6.2.3. Ligand co-crystallisation and structure determination

Recombinant *Tb*PTR1 was obtained in high yield following the protocol described in **5.3.2**. Crystallisation experiments were prepared (**5.7**) in order to co-crystallise the histidine-tagged protein in the presence of the oxidised cofactor, $NADP^+$, and

compound of interest. Compounds that displayed *Tb*PTR1 inhibition greater than approximately 60 % at 50 μM were considered for crystallisation. A cut-off value was not strictly imposed and individual compound structure was also taken into account. For example, a single molecule was commonly selected to represent a group of highly similar compounds. Crystals grew to their maximum size within 1-3 days (**Figure 6.7**).

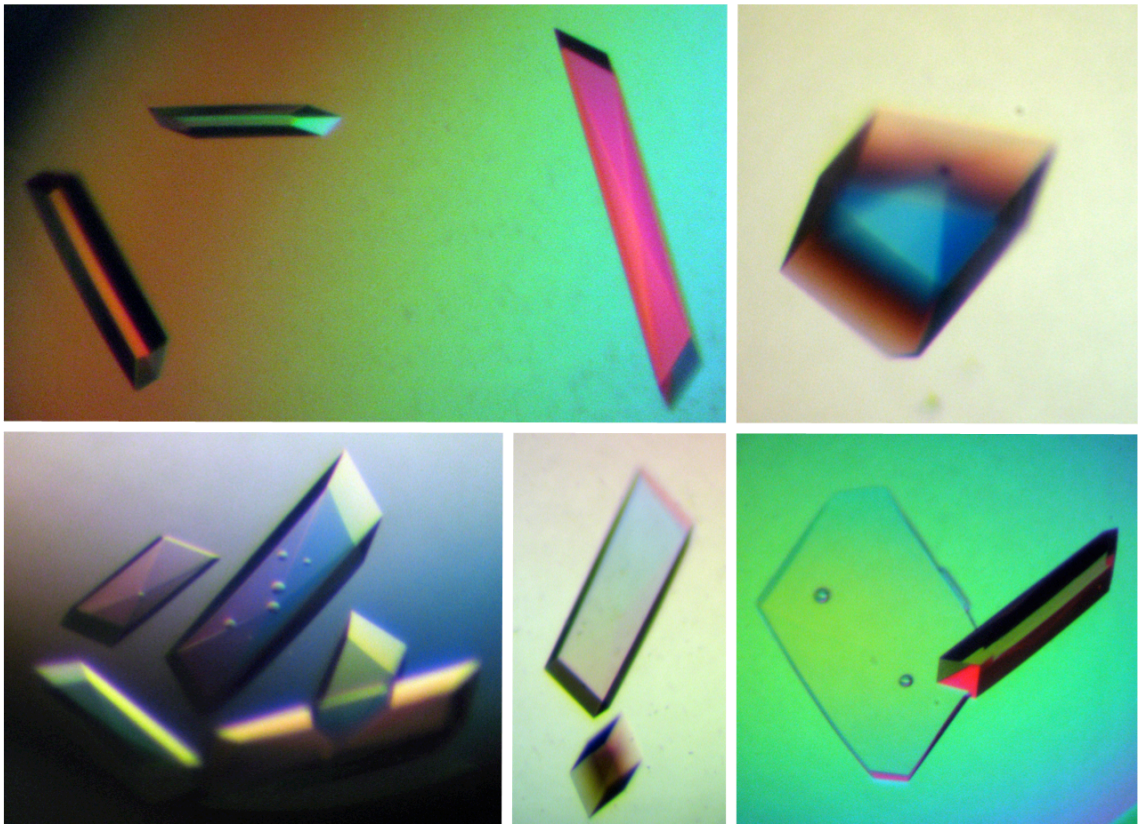


Figure 6.7. *Tb*PTR1 ligand co-crystals
A selection of *Tb*PTR1 crystals viewed under polarised light. These crystals were all grown in the presence of $NADP^+$ and selected inhibitors and range in size up to approximately 0.1-0.5 mm along the longest dimension.

Diffraction data were collected to 1.7-2.4 Å resolution from 32 *Tb*PTR1-inhibitor co-crystals. Of these, 23 novel ternary complex structures have been successfully determined. The remaining 9 datasets gave poor quality electron density into which ligands could not be satisfactorily modelled or questionable novelty and ambiguities in the chemical structure were observed. In particular, electron density for SDG 77 was inconsistent with the expected structure and SDG 54 appeared identical to that of SDG

32, perhaps as a product of degradation. Additionally, the high-resolution crystal structure of TMQ bound to *Tb*PTR1 was solved using data extending to 1.3 Å resolution alongside that of pemetrexed, cryomazine and the 2,4-diamino pyrimidine derivative, PY848, which were determined by Alice Dawson and Lindsay B. Tulloch.

All crystals were isomorphous in monoclinic space group $P2_1$. Four molecules were present in the asymmetric unit and together these comprise the functional PTR1 tetramer. Each subunit was treated individually and non-crystallographic symmetry restraints were not applied during refinement. For the high-resolution TMQ complex, anisotropic thermal parameters were included in refinement calculations. Data collection and refinement statistics are provided in **Table 6.3**. Structure factors and coordinates are deposited in the Protein Data Bank (Bernstein *et al*., 1977). Selected features observed in these structures are discussed.

6.2.4. TbPTR1-inhibitor complex structural features

6.2.4.1. Overall structure and active site organisation

All *Tb*PTR1 models adopt the same conformation as described (Dawson *et al*., 2006). Briefly, the Rossmann-fold repeat is displayed with a seven-stranded β-sheet flanked on either side by three α-helices. A seventh α-helix is present, connected to the substrate-binding loop (α6-β6). A number of loop regions could not be modelled in their entirety due to a lack of sufficient electron density. The missing residues vary between structures and in some cases between subunits of the same tetramer but are generally found in the β4-α4 and α4-β5 loops. All published PTR1 structures also appear poorly ordered at the β4-α4 loop, indicating an inherent flexibility in the region.

Compound		SDG 5	SDG 23	SDG 32	SDG 33	SDG 53	SDG 57	SDG 60	SDG 65
PDB ID		4cl8	4cld	4cle	4clh	4clo	4clr	4clx	4cm1
Resolution Range (Å)		44.71-2.20 (2.32-2.20)	19.74-1.77 (1.87-1.77)	38.26-1.80 (1.90-1.80)	41.72-1.85 (1.95-1.85)	19.72-1.88 (1.98-1.88)	29.20-1.75 (1.84-1.75)	28.97-1.85 (1.95-1.85)	45.44-1.90 (2.00-1.90)
Spacegroup		$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters	***a* (Å)**	74.4	74.8	74.1	74.7	74.1	74.4	74.1	74.7
	***b* (Å)**	89.4	90.1	88.5	89.7	89.2	90.8	89.8	90.9
	***c* (Å)**	82.2	82.8	84.2	82.8	84.4	84.4	84.2	82.7
	***β* (°)**	115.5	115.8	115.3	115.8	115.3	115.3	115.7	115.7
Wavelength (Å)		1.5418	1.5418	1.5418	1.5418	1.5418	0.9795	0.9795	1.5418
No. Reflections		160238 (22916)	438368 (49733)	186446 (20597)	191101 (5790)	348586 (38207)	746440 (93244)	346818 (50978)	261207 (27306)
No. Unique Reflections		49452 (7198)	87724 (10834)	87484 (11287)	67566 (3426)	75665 (9499)	100986 (14024)	84575 (12340)	77185 (10322)
R_{merge} [a] (%)		19.6 (66.1)	6.8 (42.8)	4.5 (29.0)	4.8 (31.6)	5.1 (21.0)	8.0 (55.8)	7.5 (45.7)	10.3 (36.8)
R_{pim} [b] (%)		12.4 (42.2)	3.4 (21.9)	3.8 (25.0)	3.2 (29.2)	2.6 (11.6)	3.2 (23.2)	4.1 (25.3)	6.6 (27.0)
Completeness (%)		100.0 (100.0)	92.1 (78.0)	96.3 (85.9)	80.5 (28.2)	94.0 (81.1)	99.0 (94.8)	99.8 (100.0)	98.5 (90.8)
Mean $I/\sigma(I)$		7.4 (3.0)	17.7 (3.7)	10.1 (2.3)	13.8 (2.3)	20.5 (6.6)	17.9 (3.4)	16.2 (3.6)	11.0 (2.9)
Redundancy		3.2 (3.2)	5.0 (4.6)	2.1 (1.8)	2.8 (1.7)	4.6 (4.0)	7.4 (6.6)	4.1 (4.1)	3.4 (2.6)
Wilson *B* factor (Å^2)		21.0	11.7	20.1	21.5	13.2	11.3	11.5	13.0
R_{work} [c] (%)		14.87	14.38	15.16	14.41	15.68	13.72	14.38	18.43
R_{free} [d] (%)		20.70	18.31	19.27	19.34	20.35	16.92	18.13	23.33
R.m.s.d. bonds (Å)		0.013	0.012	0.014	0.014	0.015	0.013	0.014	0.013
R.m.s.d. angles (°)		1.787	1.603	1.672	1.689	1.714	1.609	1.714	1.721
DPI [e] (Å)		0.26	0.11	0.12	0.16	0.15	0.09	0.12	0.17
Ramachandran plot:									
Favoured (%)		96.7	96.3	96.5	96.8	96.4	96.7	96.8	96.9
Allowed (%)		3.3	3.7	3.5	3.1	3.6	3.3	3.2	3.1
Outliers (%)		0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Total number of atoms		8389	8696	8818	8696	8632	8681	8677	8691
Average *B* for all atoms (Å^2)		19.9	18.1	22.1	22.1	20.1	18.5	19.2	13.6
Number of protein residues	**Subunit A**	251	251	252	252	251	250	251	251
	Subunit B	251	250	250	251	252	250	250	249
	Subunit C	248	251	251	251	251	249	250	250
	Subunit D	251	251	250	250	249	250	251	250
Average *B* (Å^2)	**Subunit A**	19.3	17.1	22.3	21.0	22.6	16.9	18.5	12.3
	Subunit B	18.8	16.4	20.1	20.2	19.4	18.2	17.7	11.7
	Subunit C	20.1	17.2	20.8	22.7	18.0	17.7	18.8	13.2
	Subunit D	20.5	16.8	20.4	22.1	17.6	16.8	18.6	13.6
Additional groups (No./Average *B* (Å^2)):									
$NADP^+$		4 / 14.8	4 / 13.0	4 / 16.1	4 / 16.4	4 / 14.4	4 / 12.61	4 / 14.3	4 / 8.0
Inhibitor		4 / 25.8 *	4 / 33.2	6 / 27.1	8 / 24.3 *[2]	4 / 23.6	4 / 12.00	4 / 16.1	4 / 11.4 *[2]
H_2O		606 / 24.7	883 / 28.1	923 / 32.0	718 / 29.6	779 / 28.6	930 / 28.86	763 / 28.3	886 / 22.7
Acetate		4 / 36.3	2 / 21.6	-	2 / 22.5	-	-	1 / 18.1	2 / 16.1
DTT		-	-	-	-	-	-	-	4 / 30.9 *
DTD		-	-	2 / 54.6	-	-	-	-	-
Glycerol		-	-	-	-	1 / 39.4	-	-	-

Table 6.3. (Part 1)

Compound		SDG 67	SDG 68	SDG 73	SDG 75	SDG 80	SDG 82	SDG 99	SDG 100
PDB ID		4cm3	4cm4	4cm5	4cm6	4cm7	4cm8	4cm9	4cma
Resolution Range (Å)		53.65-1.95 (1.99-1.95)	19.76-1.81 (1.90-1.81)	29.42-1.99 (2.09-1.99)	29.81-1.70 (1.79-1.70)	40.82-1.90 (2.00-1.90)	15.55-1.90 (2.00-1.90)	43.82-1.90 (2.00-1.90)	44.22-1.85 (1.95-1.85)
Spacegroup		$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters	***a* (Å)**	74.1	74.7	74.0	74.2	74.6	74.6	74.0	74.0
	***b* (Å)**	89.8	89.7	88.3	89.4	90.0	90.0	87.6	88.4
	***c* (Å)**	82.4	82.7	84.4	84.2	82.4	82.6	84.1	84.2
	***β* (°)**	115.5	115.7	115.4	115.3	115.5	115.6	115.8	115.8
Wavelength (Å)		1.5418	1.5418	1.0440	0.9796	1.5418	1.5418	1.5418	1.5418
No. Reflections		257516 (13907)	421593 (45403)	439009 (52117)	374306 (31434)	275972 (30604)	260866 (20413)	270693 (29486)	283493 (23089)
No. Unique Reflections		70625 (4441)	83224 (10114)	66126 (9011)	104658 (12005)	71727 (8584)	71435 (9025)	72202 (9073)	77144 (8047)
R_{merge} [a] (%)		6.4 (37.1)	4.4 (19.0)	12.6 (69.0)	7.3 (39.2)	8.5 (37.0)	6.3 (17.2)	7.0 (40.4)	5.9 (33.1)
R_{pim} [b] (%)		5.8 (33.2)	2.2 (9.7)	5.3 (29.8)	4.5 (27.2)	5.0 (22.6)	3.6 (14.1)	4.2 (25.8)	3.6 (22.6)
Completeness (%)		99.5 (97.5)	92.6 (77.1)	97.7 (91.5)	95.8 (75.4)	92.8 (76.1)	92.3 (80.0)	95.0 (81.9)	92.9 (66.9)
Mean $I/\sigma(I)$		11.4 (2.6)	24.7 (8.2)	10.3 (2.6)	10.8 (2.3)	17.3 (5.1)	13.9 (4.8)	16.3 (3.2)	19.2 (3.5)
Redundancy		3.6 (3.1)	5.1 (4.5)	6.6 (5.8)	3.6 (2.6)	3.8 (3.6)	3.7 (2.3)	3.7 (3.2)	3.7 (2.9)
Wilson *B* factor (Å^2)		19.8	13.8	20.3	11.8	8.1	9.7	16.0	10.5
R_{work} [c] (%)		20.00	14.63	15.84	17.35	13.48	14.15	15.15	14.59
R_{free} [d] (%)		26.05	18.25	21.23	20.72	17.86	19.18	20.24	19.06
R.m.s.d. bonds (Å)		0.015	0.013	0.015	0.013	0.013	0.014	0.014	0.014
R.m.s.d. angles (°)		1.825	1.674	1.840	1.644	1.632	1.72	1.771	1.729
DPI [e] (Å)		0.20	0.12	0.17	0.11	0.14	0.15	0.15	0.13
Ramachandran plot:									
Favoured (%)		95.7	96.8	96.1	96.9	95.8	96.9	96.4	96.2
Allowed (%)		4.3	3.2	3.9	3.1	4.2	3.1	3.6	3.6
Outliers (%)		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
Total number of atoms		8296	8522	8212	8525	8650	8958	8407	8564
Average *B* for all atoms (Å^2)		26.3	19.6	30.3	19.8	16.2	15.7	23.2	19.7
Number of protein residues	**Subunit A**	251	251	250	250	252	250	249	250
	Subunit B	250	251	250	251	252	251	249	250
	Subunit C	249	251	249	251	251	249	249	250
	Subunit D	250	250	249	249	250	250	249	250
Average *B* (Å^2)	**Subunit A**	24.6	17.0	30.5	19.5	15.3	13.8	21.9	17.8
	Subunit B	23.4	16.9	29.3	19.0	14.3	13.3	21.0	17.5
	Subunit C	29.2	20.7	29.6	18.9	15.9	14.4	24.0	21.2
	Subunit D	27.5	20.7	30.8	18.6	16.5	14.6	23.8	19.6
Additional groups (No./Average *B* (Å^2)):									
$NADP^+$		4 / 19.1	4 / 13.2	4 / 24.1	4 / 14.7	4 / 10.2	4 / 9.8	4 / 17.0	4 / 13.8
Inhibitor		4 / 25.8	4 / 21.3	4 / 34.4 *	4 / 23.8	4 / 16.9	4 / 27.0	4 / 23.7	4 / 20.4
H_2O		521 / 30.1	718 / 29.0	472 / 36.0	769 / 28.4	774 / 24.8	1041 / 26.7	623 / 30.8	695 / 28.2
Acetate		1 / 24.5	2 / 26.0	-	1 / 23.6	4 / 26.8	2 / 19.0	1 / 31.4	6 / 29.5
DTT		4 / 51.5 *	-	-	-	-	2 / 67.2	3 / 50.1 *	-
DTD		-	-	-	-	-	-	-	-
Glycerol		-	1 / 46.5	-	-	-	2 / 53.7	-	-

Table 6.3. (Part 2)

Compound		SDG 106	SDG 107	SDG 120	SDG 122	SDG 126	SDG 127	SDG 128	TMQ
PDB ID		4cmb	4cmc	4cme	4cmg	4cmi	4cmj	4cmk	2x9v
Resolution Range (Å)		37.80-2.00 (2.05-2.00)	38.57-1.85 (1.95-1.85)	44.21-1.85 (1.95-1.85)	44.83-2.00 (2.11-2.00)	57.24-1.90 (2.00-1.90)	13.94-2.20 (2.27-2.20)	45.14-2.00 (2.11-2.00)	31.05-1.30 (1.37-1.30)
Spacegroup		$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters	***a* (Å)**	74.1	74.2	74.1	73.9	74.5	74.3	74.5	74.6
	***b* (Å)**	89.7	89.5	88.4	89.7	90.3	90.2	90.3	90.7
	***c* (Å)**	84.4	84.4	84.2	84.1	82.1	82.2	82.3	82.5
	***β* (°)**	115.4	115.6	115.7	116.0	115.6	115.7	115.5	115.7
Wavelength (Å)		1.5418	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418	0.9790
No. Reflections		218536 (15718)	254066 (20891)	285451 (23541)	247789 (35017)	271269 (29756)	153737 (11807)	246243 (34842)	1337387
No. Unique Reflections		66870 (4932)	83009 (10936)	76479 (7699)	63882 (9041)	73160 (9197)	46799 (3980)	66280 (9614)	239044
R_{merge} [a] (%)		8.0 (37.5)	5.7 (24.4)	7.2 (36.2)	17.0 (61.2)	8.8 (35.8)	14.8 (67.9)	20.0 (76.5)	8.7 (40.1)
R_{pim} [b] (%)		7.4 (34.0)	3.7 (19.9)	4.3 (23.8)	10.0 (35.9)	5.3 (23.1)	13.2 (56.7)	12.0 (46.6)	4.0 (18.3)
Completeness (%)		99.4 (99.2)	97.9 (88.8)	91.9 (63.9)	96.1 (93.6)	95.0 (82.0)	94.5 (88.3)	99.6 (99.4)	98.8 (98.6)
Mean $I/\sigma(I)$		12.9 (3.4)	14.4 (3.5)	17.8 (3.6)	9.8 (3.0)	10.6 (3.2)	5.1 (2.3)	7.4 (2.1)	15.9 (4.6)
Redundancy		3.3 (3.2)	3.1 (1.9)	3.7 (3.1)	3.9 (3.9)	3.7 (3.2)	3.3 (3.0)	3.7 (3.6)	5.6 (5.5)
Wilson *B* factor (Å^2)		17.4	11.9	11.8	12.7	13.6	13.3	13.0	10.0
R_{work} [c] (%)		17.24	18.35	15.05	16.06	16.02	20.51	17.31	12.10
R_{free} [d] (%)		22.33	23.01	19.36	21.27	21.01	28.48	23.32	14.80
R.m.s.d. bonds (Å)		0.015	0.014	0.014	0.015	0.014	0.014	0.016	0.010
R.m.s.d. angles (°)		1.743	1.720	1.714	1.804	1.746	1.832	1.919	1.405
DPI [e] (Å)		0.19	0.15	0.14	0.18	0.16	0.39	0.19	0.04
Ramachandran plot:									
Favoured (%)		96.6	96.0	96.6	96.3	96.8	94.0	96.5	96.5
Allowed (%)		3.2	4.0	3.4	3.7	3.1	5.7	3.5	3.5
Outliers (%)		0.2	0.0	0.0	0.0	0.1	0.3	0.0	0.0
Total number of atoms		8567	8592	8590	8279	8524	8112	8332	9148
Average *B* for all atoms (Å^2)		19.0	20.9	16.8	17.9	19.8	17.2	19.3	15.3
Number of protein residues	**Subunit A**	251	251	249	250	250	251	250	250
	Subunit B	250	249	249	251	251	251	251	251
	Subunit C	249	249	250	250	250	251	251	249
	Subunit D	249	251	249	250	250	249	250	250
Average *B* (Å^2)	**Subunit A**	18.4	21.7	15.0	17.3	18.2	16.1	18.3	11.9
	Subunit B	17.8	19.9	15.3	17.8	18.8	15.7	18.6	11.8
	Subunit C	18.4	19.1	17.9	18.6	19.7	19.1	19.6	12.8
	Subunit D	19.0	20.4	16.7	17.4	20.9	18.9	20.1	12.8
Additional groups (No./Average *B* (Å^2)):									
$NADP^+$		4 / 12.6	4 / 14.2	4 / 11.2	4 / 12.5	4 / 14.6	4 / 11.5	4 / 14.3	4 / 8.8
Inhibitor		4 / 20.7	4 / 27.5	4 / 16.4	4 / 19.0	4 / 18.5	4 / 21.1	4 / 22.5	4 / 15.5
H_2O		767 / 26.8	764 / 28.3	734 / 24.8	517 / 22.0	698 / 25.9	355 / 14.6	569 / 22.4	1202 / 34.9
Acetate		1 / 16.2	-	-	2 / 19.3	2 / 21.1	-	1 / 21.9	6 / 19.7
DTT		-	-	-	-	-	-	-	-
DTD		-	-	-	-	-	-	-	-
Glycerol		-	-	-	-	-	-	-	-

Table 6.3. Data collection and refinement statistics for 24 *Tb*PTR1-inhibitor complex crystal structures

Values in parentheses correspond to the highest resolution shell. All statistical definitions are as provided for **Table 6.1.** *B* factor values marked with * were calculated for the Cys168 to which the ligand is covalently bonded and were also included in the overall subunit calculations. *[2] SDG 33 and SDG 65 were present in multiple conformations and the value provided is the average of all conformers.

A single subunit is shown in **Figure 6.8**, depicting the secondary structure and location of the substrate and inhibitor binding site. The tetramer is similar to that shown for *Ld*PTR1 in **Figure 6.2A**. R.m.s.d between all four chains is consistent, at approximately 0.29 ± 0.09 Å when 248-252 Cα atoms are aligned (individual values were calculated using *SSM* (Krissinel and Henrick, 2004) in *Coot* (Emsley and Cowtan, 2004; Emsley *et al*., 2010)). Due to this high level of structural conservation, analysis has been carried out using subunit A unless otherwise specified.

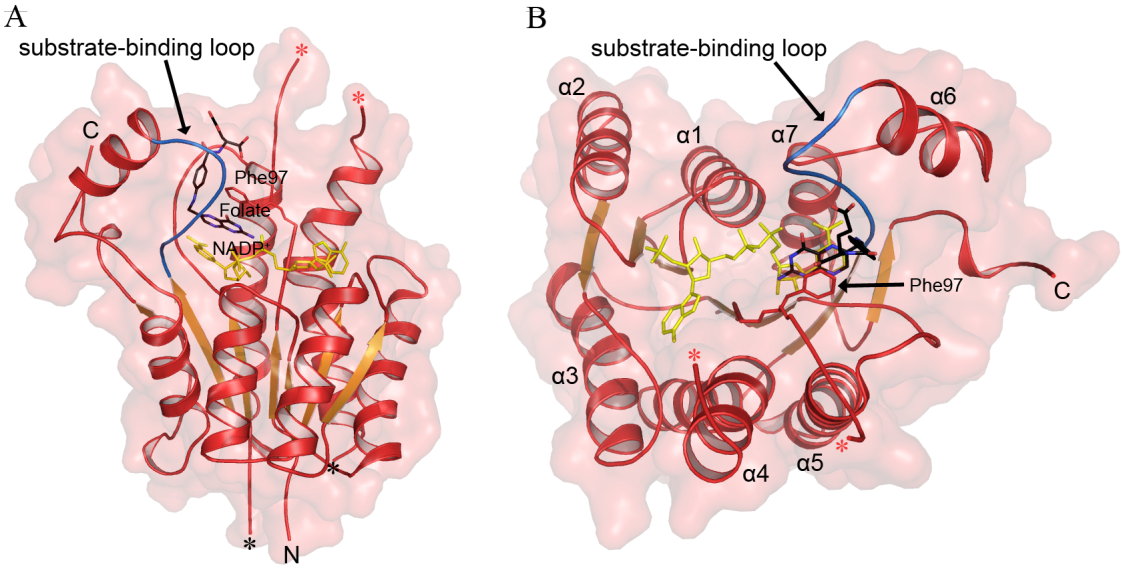


Figure 6.8. *Tb*PTR1 monomer architecture
(A) A monomer of *Tb*PTR1 shown as red and orange ribbons with a semi-transparent van der Waals surface and **(B)** rotated to show the active site from above**.** This figure was prepared using *Tb*PTR1 from the SDG 60 ternary complex but shows $NADP^+$ and folate from PDB 3bmc (Tulloch *et al*., 2010) in the catalytic site. Gaps in two loops are linked by coloured asterisks (*). The substrate-binding loop between β6 and α6 is coloured blue, cofactor yellow and folate black (with blue N and red O). The side-chain of Phe97 is displayed as red sticks to show the π-stacking of the $NADP^+$, folate and phenylalanine ring systems. **(B)** depicts the approximate orientation of the enzyme used to prepare subsequent figures.

In all instances, $NADP^+$ is bound within the active site, creating the right environment for interactions with inhibitor molecules (Dawson *et al*., 2006). Key hydrogen bonds contributing to position the cofactor in *Tb*PTR1 include those between the adenine moiety and the side chain carboxyl of Asp62; Ser95 amide and the adenine ribose O; the adenine 2' phosphate and His35, Asn36 and Ser37 amides and Ser37 hydroxyl groups

located on β2-α2; both α- and β-phosphates and Arg14 through the side chain amide and amino atoms; the nicotinamide ribose and the amino group of Lys178; Ser207 amide and Leu208 carbonyl groups with the nicotinamide N7 (not shown). While the adenine moiety lies more buried in the elongated cleft, the extended cofactor forms the floor of the catalytic site, creating additional positions for substrate or inhibitor interaction, particularly at the nicotinamide ribose and phosphate groups (**Figure 6.9**). The association of cofactor with members of the SDR family are well characterised (Duax *et al*., 2003) and the sequential binding of ofactor then substrate in PTR1 is conserved (Luba *et al*., 1998).

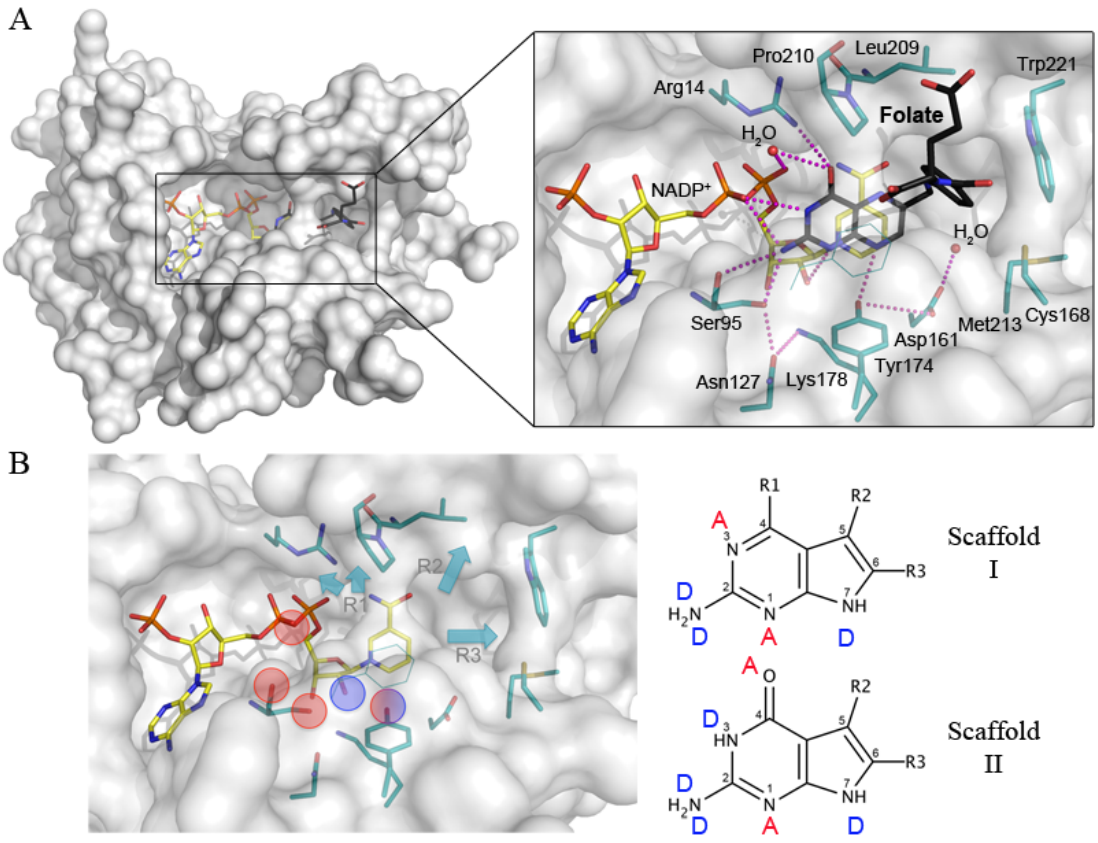


Figure 6.9. Organisation of the *Tb*PTR1 active site
(A) A surface representation of *Tb*PTR1 with cofactor and substrate, folate, bound (PDB 3bmc) showing key residues that create the active site pocket. Potential hydrogen bonds are depicted as magenta dotted lines and all atoms are coloured according to atom type: O, red; N, blue; S, gold; P, orange; C, yellow ($NADP^+$), cyan (*Tb*PTR1) or black (folate). Phe97 is not labelled. **(B)** The active site with folate removed and key hydrogen donor or acceptor groups circled blue or red, respectively. Scaffolds I and II are shown opposite and possible hydrogen donor or acceptor groups are designated D or A, respectively. Arrows on the schematic also indicate the intended direction of R1, R2 and R3 substitutions.

The vacant active site of *Tb*PTR1 is approximately 30 x 22 x 15 Å (Dawson *et al*., 2006) before $NADP^+$ is bound. Both cofactor and folate are shown in the active site in **Figure 6.9A** (PDB 3bmc; Tulloch *et al*., 2010). MTX and other molecules designed to prevent PTR1 activity are competitive inhibitors so it is necessary to generate interactions equivalent or surplus to those made with substrate. Amino acids important for positioning substrate or MTX include Ser95, which donates a hydrogen bond to Asn127 and is able to accept hydrogen bonds from a ligand at both the main chain carbonyl and side chain hydroxyl groups. Tyr174 is able to share hydrogen bonds with nitrogen-based groups of the pterin moiety and is also positioned close to Asp161 and Lys178, together comprising what is often considered the catalytic triad. The *p*-aminobenzoyl (*p*ABA) and γ-glutamate tail of both folate and MTX make no additional hydrogen bonds with *Tb*PTR1 and are directed out of the active site cavity. The region of the active site at the opposite side to the cofactor (at the right of most figures presented here, indicated in **Figure 6.9B**) contains sub-pockets that are not exploited by either substrate or MTX, representing an ideal area for inhibitor development. Scaffolds I and II (**Figure 6.5**) were previously shown to successfully generate key contacts with cofactor and catalytic residues (Tulloch *et al*., 2010) and were therefore subject to further modification with new interactions identified. **Figure 6.9B** shows the *Tb*PTR1 active site with folate removed and highlights several amino acids and cofactor groups that these scaffolds were intended to share hydrogen bonds with. Hydrogen bond donor or acceptor groups were assigned based on previous knowledge of active site interactions and most of the designated groups of scaffolds I and II are complementary to those of the active site if binding occurs in the orientation shown and no other protonation events occur. Regions of the active site to be explored by chemical substitutions at R1, R2 and R3 are also indicated.

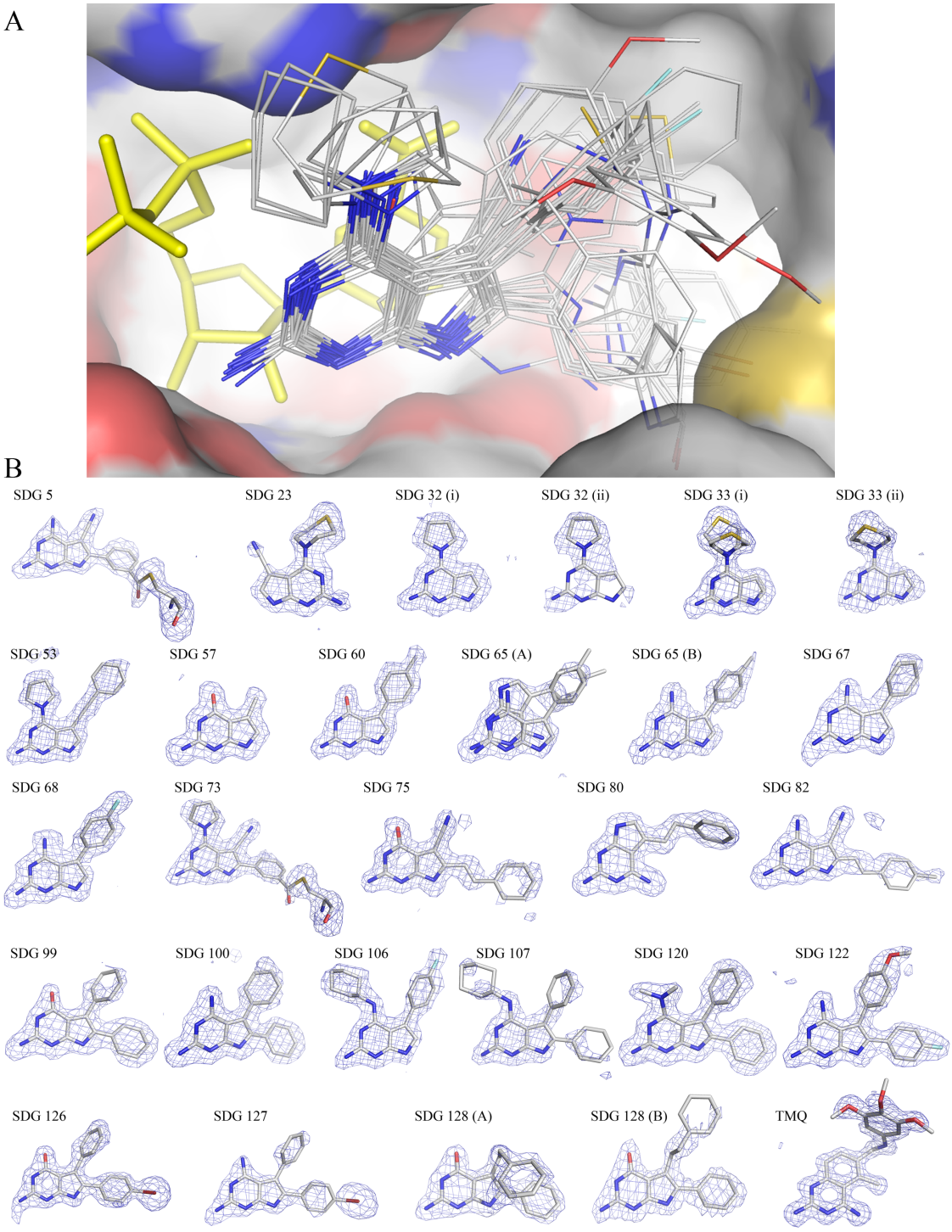


Figure 6.10. *Tb*PTR1 active site containing NADP$^+$ and 24 inhibitors
(A) To display the range of inhibitors assessed structurally, all 24 ligands are overlaid on a surface representation of the *Tb*PTR1 active site based on the ternary complex with SDG 60. Five amino acids have been removed for clarity (Phe97, Pro210, Ala212, Met213 and Glu217). NADP$^+$ is shown as yellow sticks. All other atoms are coloured according to element (C, grey; N, blue; O, red; S, gold; F, pale blue; Br, brown). **(B)** Difference density omit maps of all inhibitors, represented by blue mesh. F_o-F_c maps were calculated with the molecule removed from the final model and contoured at 3 σ (TMQ at 4 σ). (i) and (ii) indicate the primary and secondary molecules observed in the active site while (A) and (B) indicate molecules bound to subunits A or B. SDG 5 and SDG 73 are shown linked to Cys168.

When all 24 compounds with crystal structures are superimposed in the *Tb*PTR1 active site (**Figure 6.10A**), it is clear that the core scaffold position appears well conserved and the widely dispersed appearance of the R-group substituents demonstrates the range of molecules assessed. Corresponding omit maps are also provided for each compound (**Figure 6.10B**). In most cases, the inhibitor shows the same binding pose in all four subunits in the asymmetric unit and it is only necessary to describe a single example. However, we have observed ligands that adopt two orientations at partial occupancy and these will be described alongside other features of ligand binding.

6.2.4.2. Ligand orientation

Substrate molecules all bind in the same orientation, shown by the *Tb*PTR1 structure in a complex with folate (PDB 3bmc; Tulloch *et al*., 2010; **Figure 6.9**) and by *Lm*PTR1 structures complexed with biopterin, dihydro- and tetra-hydrobiopterin (PDB 2bf7, 1e92, 2bfp; Gourley *et al*., 2001; Schüttelkopf *et al*., 2005). The core pterin moiety lies coplanar with the nicotinamide and the side chain of Phe97 in a π-π stack with the 2-amino group donating hydrogen bonds to Ser95. The substrate 4-carbonyl is then directed towards Pro210 of the substrate-binding loop. MTX and a number of other inhibitor molecules adopt a different orientation to maximise hydrogen bonding capabilities with the 4-amino group positioned to donate a hydrogen bond to Tyr174, taking the place of substrate N8. Of the 24 ligand-complex structures here, only SDG 80 and TMQ adopt the MTX-like orientation. The pyrrolo-pyrimidine core of SDG 80 aligns in the same way as the pteridine of MTX. The ethylphenyl group then extends towards Trp221, deeper into the hydrophobic cavity than the MTX *p*ABA group which is directed to solvent. The partner of SDG 80 based on scaffold II binds according to the substrate orientation (PDB 3jqb, Tulloch *et al*., 2010) and is shown schematically in

Figure 6.11. This published ligand is SDG 81 and the binding orientation was confirmed[1]. The change in orientation replicates that seen between folate and MTX where a similar chemical substitution is made. There does not appear to be any conformational changes imposed on active site amino acids to accommodate either SDG 80 or SDG 81 and of all other SDG models studied, the NH_2:N:NH pattern of scaffold I and II positions 2:1:7 appears to favour the substrate-binding orientation.

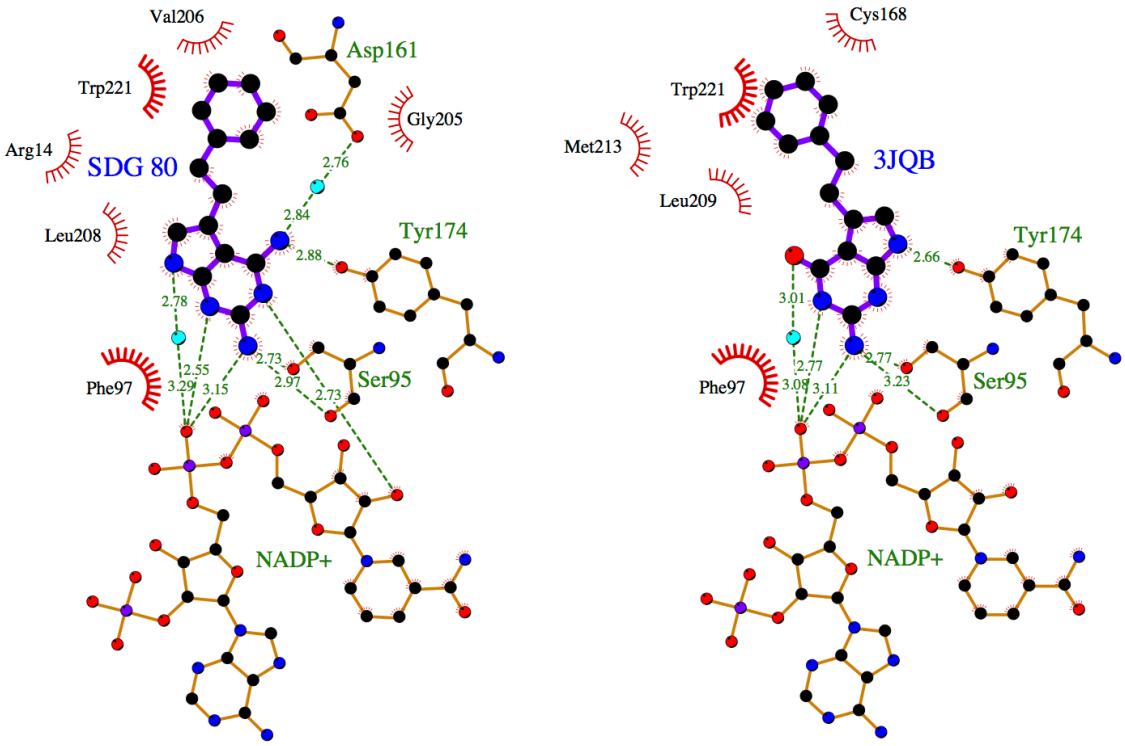


Figure 6.11. SDG 80 and SDG 81 (3jqb) adopt different orientations
Schematic depicting SDG 80 bound in the MTX-like orientation (left) while the ligand from PDB 3jqb (SDG 81) adopts the substrate-like orientation (right). Ligands are shown as large circles coloured by residue (C, black; N, blue; O red) with purple bonds. Nearby residues and $NADP^+$ have orange bonds. Water molecules (cyan circles) have been filtered and are only shown if at least two contacts are made. Potential hydrogen bonds are depicted as green dashed lines and hydrophobic interactions represented by red curves. This figure was prepared using $LigPlot^+$ (Wallace *et al*., 1995) and all contacts were calculated automatically by the associated *HBPLUS* (McDonald and Thornton, 1994).

[1]*The crystal structure of SDG 81, published by Tulloch et al. (2010), was inadvertently re-determined. Further details are therefore not presented.*

TMQ has a quinazoline core and lacks a nitrogen at position 8 (equivalent to N7 in pyrrolopyrimidines) to favourably interact with Tyr174 if it were bound in the substrate orientation. In the MTX orientation, the 4-amino group is optimal to provide this contact. All previous ligands based on scaffold II also adopt the substrate-like orientation (Tulloch *et al*., 2010). When all R-groups were represented by H, relatively poor inhibition of both *Tb*PTR1 and *Lm*PTR1 resulted, indicating that additional substitutions are required.

SDG 65, with an amino group at R1 of scaffold I, presents as a dual conformer adopting the MTX-like orientation with half occupancy in subunit A. It also appears to bind in the substrate-like orientation at half occupancy while it has been modelled solely in this orientation in all other subunits. It is possible that the highly similar SDG 67 and SDG 68 are also able to bind in this manner but electron density suggested a preference for the substrate-like binding mode. McLuskey *et al*. (2004) found a similar feature in the *L. major* ternary complex with 2,4,6-triaminoquinazoline (TAQ) where multiple orientations were displayed by the inhibitor. Again, this may indicate that additional substitutions are necessary to drive a single binding mode. For example, addition of bulky groups at the R3 position may be one way to ensure the substrate-like orientation is adopted as large clashes would be expected to prohibit binding in any other orientation.

Interestingly, SDG 23 binds in an unexpected orientation that is neither substrate nor MTX-like. Rather, the pyrrolopyrimidine ring system is flipped by 180° from the traditional substrate-like position (**Figure 6.12**). The 2-amino group of SDG 23 is directed towards Asp161, sharing a hydrogen bond with a side chain OD atom. Tyr174

can still make contact with the N and NH at positions 1 and 7 but the entire core is more distant from the cofactor. Only a weak water-mediated hydrogen bond connects the inhibitor to the $NADP^+$ phosphate and Ser95 (mean NH-H_2O distance is 3.6 Å) while a hydrogen bond remains with the nicotinamide ribose (mean distance 3.1 Å). A water molecule in place of the amino group that forms hydrogen bonds with the cofactor α-phosphate and Ser95 has previously been seen in the *L. major* enzyme complexed with trimethoprim, a DHFR inhibitor containing a pyrimidine-2,4-diamine core (Schüttelkopf *et al*., 2005).

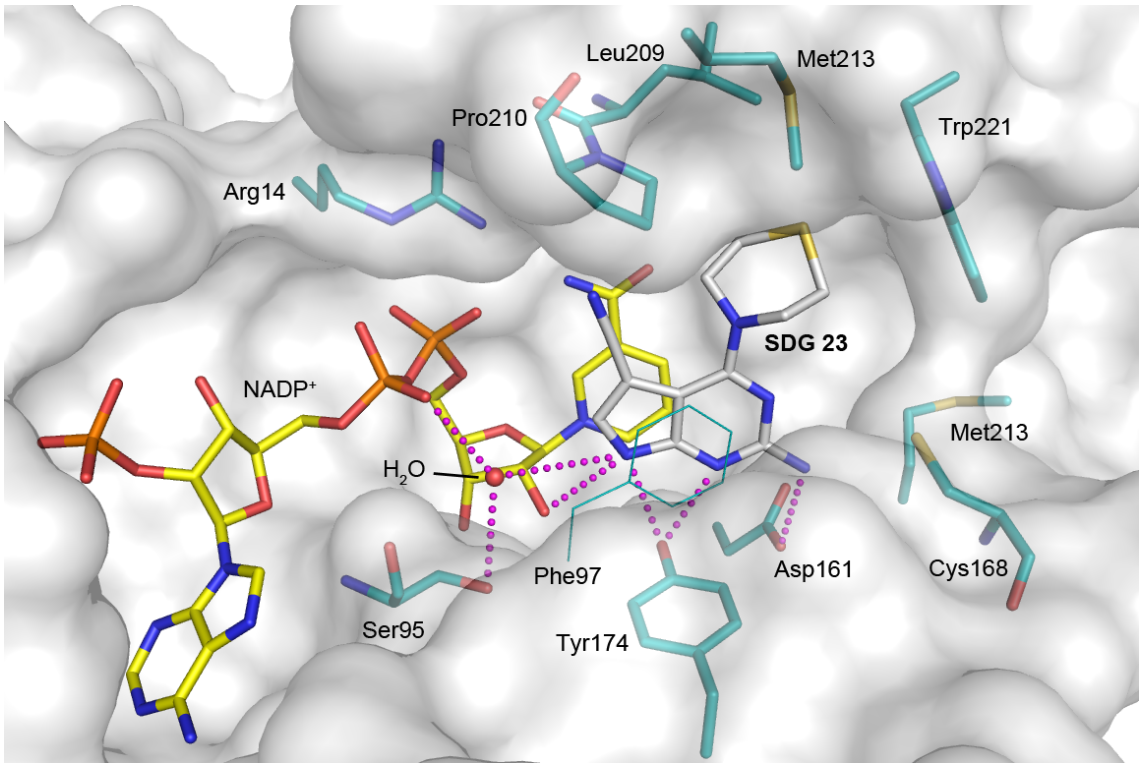


Figure 6.12. SDG 23 binds in an unfavourable orientation
*Tb*PTR1 is shown as a semi-transparent van der Waals surface. Phe97 was removed prior to surface generation and is shown as thin lines here for clarity. Selected active site residues, $NADP^+$ and SDG 23 are shown as sticks, coloured by atom using the same colour scheme as **Figure 6.9** except SDG 23 C atoms are shown here in grey. Potential hydrogen bonds shared with SDG 23 are shown in magenta and the water molecule with hydrogen bonds to Ser95 and cofactor is shown as a red sphere.

The alternative orientation of SDG 23 was likely brought about by the large thiomorpholine group at position 4, clashing with residues of the substrate-binding loop,

including Pro210. However, SDG 33 contains the same group and adopts the substrate-like orientation inducing movement of those residues. This therefore implies that it is the combination of R1 thiomorpholine and R2 carbonitrile in SDG 23 that prohibits the intended substrate-like interactions while not containing the ideal organisation to exploit the MTX orientation. Compounds that bind in this manner, making only water-mediated hydrogen bonds with the cofactor phosphate and Ser95, are typically weak inhibitors, including SDG 23 (K_i 8.75 μM). Understanding the features that cause this suboptimal placement is important for improving inhibitor design. The core framework employed here has never been structurally assessed with an R1 substitution larger than an amino group. It has been demonstrated that while the active site can accommodate a thiomorpholine (SDG 33), pyrrolidine (SDG 32, SDG 53 and SDG 73), N^4,N^4-dimethyl (SDG 120) or an N^4-cyclohexane group (SDG 106 and SDG 107) and bind in substrate-like orientation, the coincident R2 constituent can influence the ultimate binding mode. In particular, the inability to accommodate a large thiomorpholine substitution at the R1 position in combination with an R2 carbonitrile in order for substrate-like binding is highlighted.

6.2.4.3. The substrate-binding loop

The substrate-binding loop is flexible, particularly between residues 205-213. It has already been shown in *Ld*PTR1 that the absence of ligand in the catalytic site likely affects the placement of this loop and in *Tb*PTR1, we observe that this loop shows some degree of disorder even where both cofactor and inhibitor appear bound in an ordered manner. In most cases, only the most prevalent conformation of the loop is modelled. However, in some complexes, there are likely two (or indeed more) possible positions of this loop. Care should be taken in the analysis of such placement, as there are likely

multiple conformers where the loop has not been adequately stabilised. Despite this, it is clear that scaffold I compounds with R1 substituents other than NH_2 result in relocation of this loop. For example, SDG 53 illustrates that a pyrrolidine group at R1 coupled with an extended inflexible hydrophobic R2 group directed towards Trp221 forces the loop and part of β6 to be displaced (**Figure 6.13**).

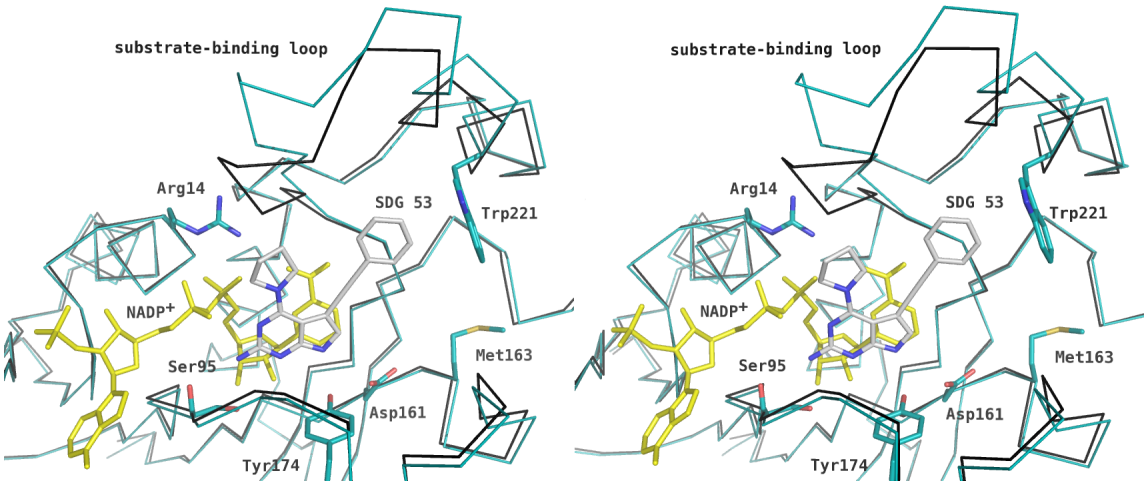


Figure 6.13. Substrate-binding loop displaced by SDG 53
A stereo image demonstrating the different positions of the substrate-binding loop between *Tb*PTR1 complexed with SDG 53 (cyan ribbon) and folate (PDB 3bmc, black). $NADP^+$ is displayed as yellow sticks while SDG 53 and selected active site residues are coloured by atom (C, cyan or grey; N, blue, O, red; S, gold).

6.2.4.4. Subunit variation and conformational changes

While non-crystallographic symmetry restraints were not imposed during structure refinement (**5.7**) and the *Tb*PTR1 tetramer chains are typically well conserved structurally (**6.2.4.1**), there are some local variations observed. This is mainly restricted to the placement of flexible amino acid side-chains, particularly involving exposed surface residues or disordered loops. Active site residues generally adopt a fixed conformation but have been shown to adjust position to accommodate certain ligands. For example, the substrate-binding loop often appears disordered (described in **6.2.4.3**) and Trp221 on the border of the active site has been displaced by inhibitors

(Mphamhanga *et al*., 2009). Here, SDG 128 causes an unusual repositioning of Phe97 (**Figure 6.14**). As described, Phe97 usually forms a stacked arrangement with the cofactor nicotinamide and the substrate pterin moiety (**Figure 6.8**). In subunits A, C and D of the SDG 128 ternary complex, the R2 phenethyl group is able to fold back, almost creating a second stack over the R3 phenyl group. Phe97, whilst still aligned over the core scaffold, is displaced by up to 2.5 Å to overlay only the 6-membered pyrimidine ring and no longer appears coplanar with both cofactor and ligand. Subunit C displays a less prominent relocation than A and D while in subunit B, Phe97 remains in the same position as all other PTR1 complexes. The R2 phenethyl in subunit B is instead extended towards the hydrophobic pocket formed by Leu209, Met213 and Trp221. The extended conformation is also accompanied by a small adjustment of the substrate-binding loop.

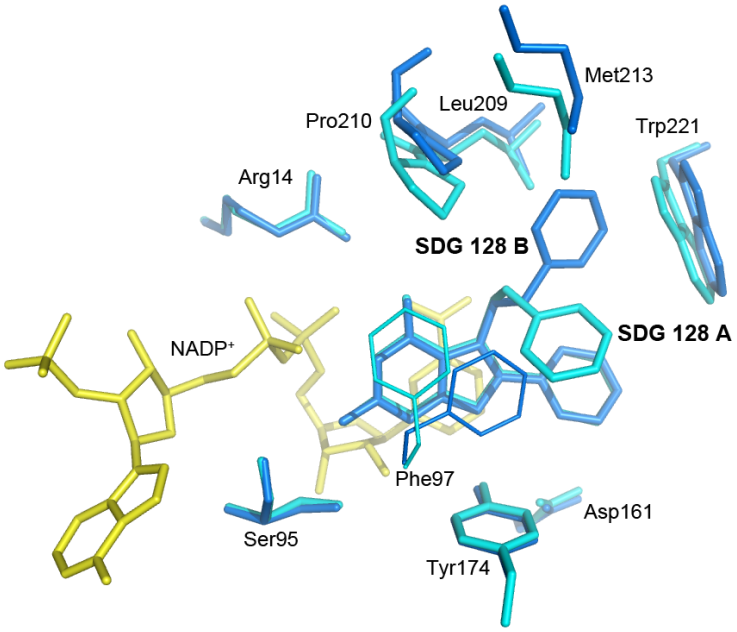


Figure 6.14. Active site Phe97 displaced by SDG 128
In three out of four subunits, Phe97 is displaced by a phenethyl substitution at R2. $NADP^+$ is shown as yellow sticks. Subunit A active site residues and inhibitor are coloured pale blue while the equivalent residues and ligand of subunit B are coloured dark blue. Phe97 is shown in both cases as thin lines of the same colours.

6.2.4.5. Two inhibitors binding in a single active site

In two structures, there is evidence of multiple inhibitor molecules binding in the same active site. SDG 33 shows well defined electron density for two molecules per subunit and the same applies to SDG 32 in subunits A and D. Both SDG 32 and SDG 33 are low molecular weight compounds with mass of 203 and 235 Da, respectively. As such, they could be classified within the category of 'fragments' (Carr *et al*., 2005). Both SDG 32 and SDG 33 are weak inhibitors of *Tb*PTR1 in the low micromolar range (4.17 μM and 8.64 μM, respectively; **Table 6.2**). However, as fragments, their potency is desirable as potential starting molecules (Rees *et al*., 2004). The mode of binding of the second molecule can therefore be particularly useful in the design of further compounds expanding on the core template. The distinct poses adopted by SDG 33 are shown in **Figure 6.15**. The first, or primary, pose assumes the classic substrate-like orientation. The R1 thiomorpholine group is present in two conformations, each of half occupancy, in subunits A and B. In one conformer, the group is directed towards Arg14, potentially acting as a weak hydrogen bond acceptor while the other conformer, as in subunits C and D, is modelled with the sulfur aimed out of the cavity. The secondary pose adopted by SDG 33 displays a more ordered placement than that of SDG 32 (see omit maps in **Figure 6.10B**) and lies almost perpendicular to the primary molecule in what is usually seen as a cavity containing water molecules or in some cases, DTT. Separated by less than 4 Å, the angle between the ring planes of both molecules is approximately 90° and is indicative of favourable edge-face π-interactions while the molecule at the secondary site can also interact with the side chain of Phe171 in the same manner. The 2-amino group of the secondary SDG 33 lies within favourable hydrogen bonding distance to both the backbone carbonyl of Gly205 and the carboxylate of Asp161. In structures of both the folate and MTX-bound enzyme, this position is occupied by an ordered water

molecule. Indeed, during the first reduction step in the enzyme mechanism (**Figure 4.2**), Asp161 likely acquires a proton from solvent that is passed on to substrate by Tyr174. Here, the side chain of Asp161 also shares a hydrogen bond with SDG 33 N1 while a water molecule hydrogen bonds to the 7-NH group. The ordered R1 thiomorpholine group extends to facilitate hydrophobic interactions with residues in the β6-α6 substrate-binding loop and does not cause any significant conformational change to nearby amino acids.

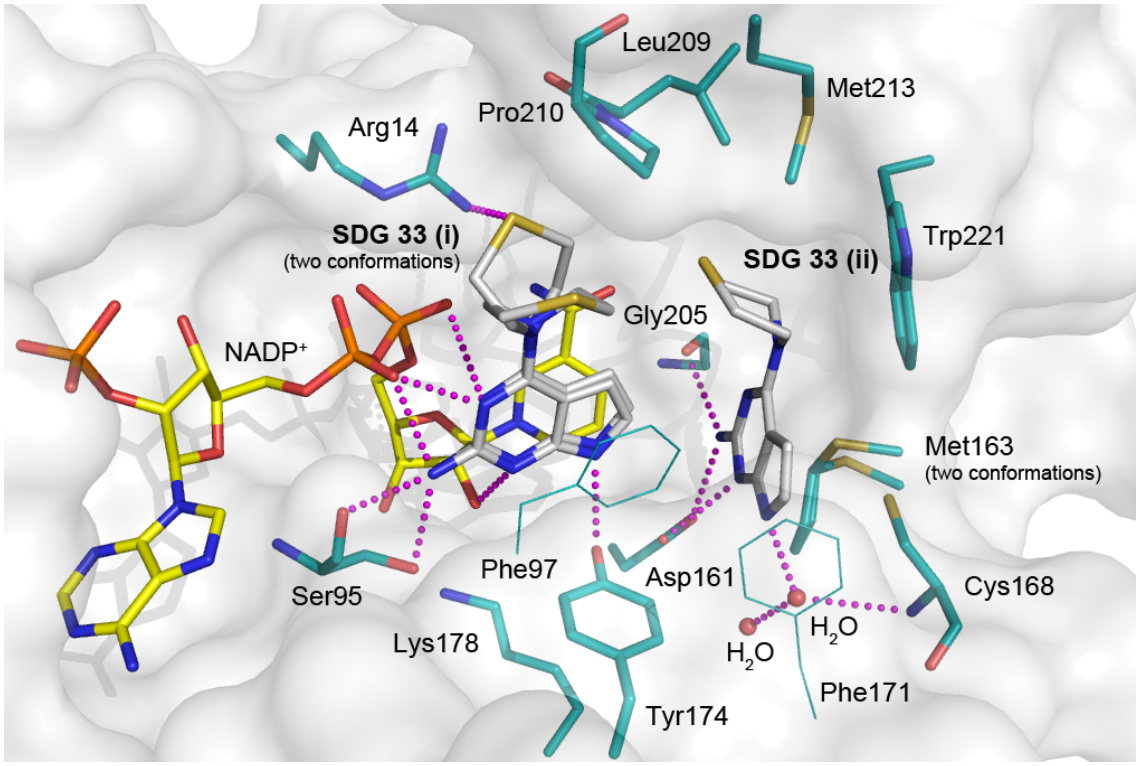


Figure 6.15. Two SDG 33 molecules bound in the *Tb*PTR1 active site
The *Tb*PTR1 active site is represented as described in **Figure 6.12**. The primary SDG 33 molecule is labelled (i) and is present in two conformations. The secondary SDG 33 molecule (ii) is shown sharing hydrogen bonds with Asp161, Gly205 and a water molecule. Also shown is Phe171 as thin cyan lines, forming a border to the secondary binding site.

This secondary binding location has been occupied by an inhibitor molecule on two previous occasions (Mpamhanga *et al*., 2009). However, only in one subunit were two inhibitor molecules bound simultaneously, displaying only partial occupancy. The main site molecule also did not form all of the well-characterised key hydrogen bonds in

either the standard substrate or MTX-like orientations. The ability to bind simultaneously in both the substrate-like orientation and at this secondary site is therefore displayed for the first time by SDG 33 and to some extent, SDG 32. It is possible that the second molecules are permitted to bind due to the high compound:enzyme ratio present in the crystallisation mixture (at least five-fold excess). However, compounds binding solely in the second site have been shown by Mpamhanga *et al*. (2005) to inhibit *Tb*PTR1 activity (apparent K_i 0.4 μM, measured under different assay conditions). Therefore, binding in this orientation may indeed contribute to inhibition and kinetic values provided for SDG 33 may be inappropriate. This perhaps also contributes to the difference between the apparent potency of SDG 33 and SDG 32, where the second molecule is less ordered. Nevertheless, other fragment-sized molecules crystallised with PTR1 generally do not display the same ability to bind at multiple locations indicating that the properties of SDG 33 are indeed favourable and demonstrates a desirable area to optimise a single molecule capable of exploiting the otherwise vacant secondary site in combination with the primary catalytic region.

6.2.4.6. Covalent interactions and modifications

A cysteine residue, Cys168, at an opening to the catalytic site is often susceptible to covalent modifications. Oxidation to sulfenic acid is regularly observed (in at least one subunit of complexes with SDG 53, 57, 68, 82, 99, 120 and 126) and it has been found to form covalent links with DTT (in at least three subunits of complexes with SDG 65, 67 and 99) or cacodylate (Dawson *et al*., 2006). A unique observation presented is the covalent attachment of two inhibitors to *Tb*PTR1 via Cys168 (**Figure 6.16A**). SDG 5 and SDG 73 both contain a 3-formylphenyl group attached to scaffold I at position 5 (R3). The pyrrolopyrimidine core binds in the substrate orientation as described for

other molecules in the series. The R3 phenyl ring is rotated by approximately 30° with respect to the main scaffold in order to present the reactive formyl group to contact Cys168 where, following nucleophilic attack by the deprotonated thiol, they are linked by a covalent bond.

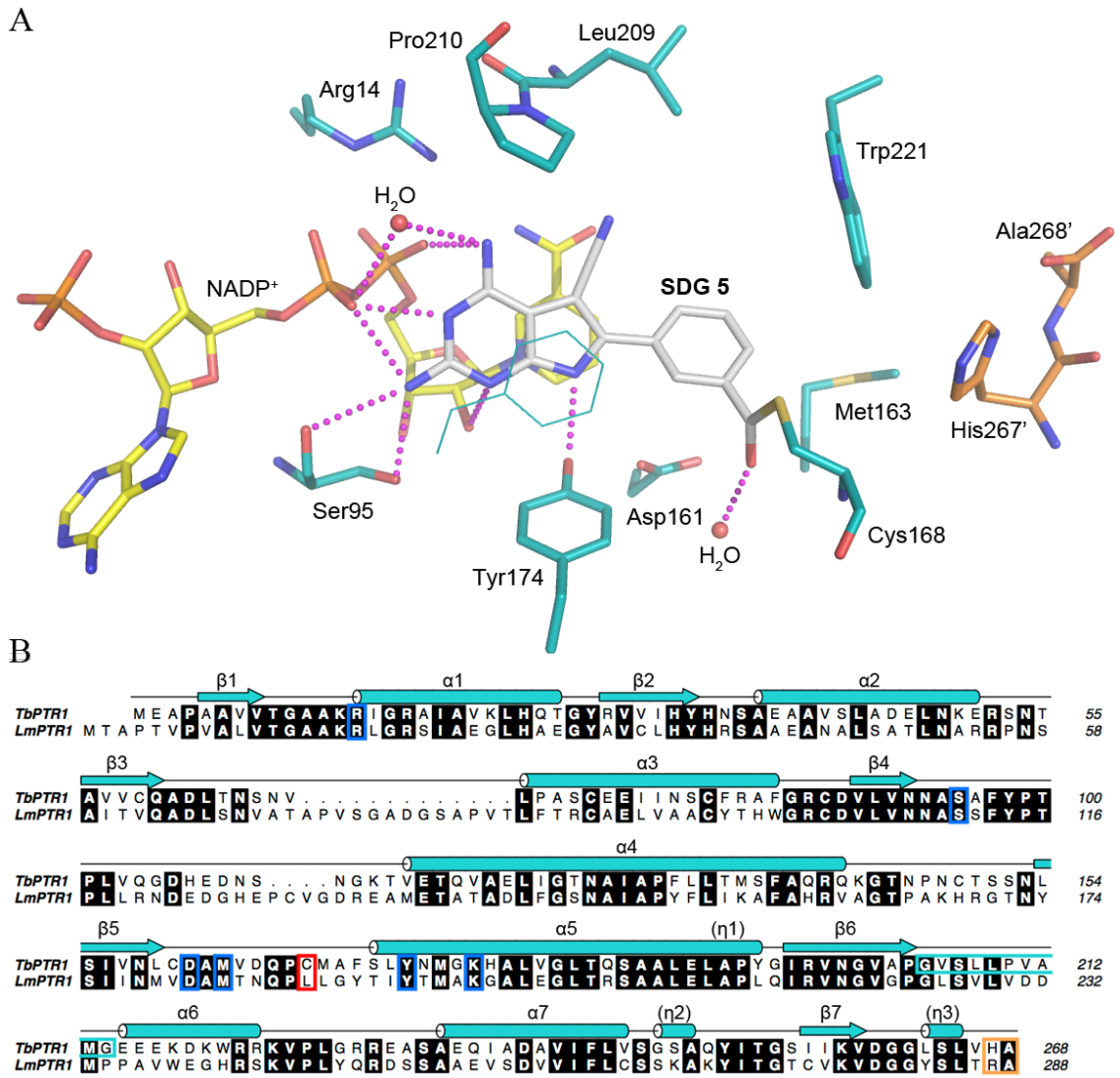


Figure 6.16. SDG 5 is a covalent inhibitor of *Tb*PTR1

(A) SDG 5 is shown covalently attached to Cys168, surrounded by selected active site residues and $NADP^+$. All atoms are coloured as previously except His267′ and Ala268′ from subunit D, represented by orange C atoms. These two residues are shown to highlight the proximity for their potential involvement in the covalent linkage mechanism. **(B)** A sequence alignment of *Tb*PTR1 and *Lm*PTR1. Conserved amino acids are shown as white text on a black background. Selected conserved active site residues are highlighted by blue boxes while the *Tb*PTR1 β6-α6 substrate binding loop is surrounded by a teal box. The non-conserved Cys168 and partially conserved *C*-terminal residues are highlighted red and orange, respectively. *Tb*PTR1 secondary structure according to Dawson *et al*. (2006) is shown as teal cylinders (α-helices or η-turns) and arrows (β-strands).

In drug development, there is a reluctance to utilise molecules that bind covalently to their target. Non-specific binding leading to toxicity is the major concern and as such, compounds that risk generating high toxicity are often avoided at the early stages of the development process. Covalent inhibitors do not automatically fail to meet the criteria of traditional drug-like models such as Lipinski's 'rule of five' (Lipinski *et al*., 2001). Many commonly used drugs are in fact covalent modifiers including aspirin (Roth *et al*., 1975; Tóth *et al*., 2013) and β-lactam antibiotics (Sainsbury *et al*., 2011). Indeed, it is estimated that almost one third of enzymes targeted by marketed drugs (21 out of 71 enzymes targeted by 317 FDA approved drugs in 2005, USA) undergo a covalent modification (Robertson, 2005), but rarely are they designed with this intended mechanism of action (Singh *et al*., 2011). An advantage of purposely designing covalent dugs includes the potential lower dose requirement. If highly specific, toxicity can then also be lower than anticipated.

The potential reactivity of *Tb*PTR1 Cys168 was previously postulated by Dawson *et al*. (2006) to be enhanced by the proximity of His267 and the carboxyl terminus of Ala268 from a neighbouring subunit (subunit D), an arrangement similar to that seen in cysteine proteases (Tyndall *et al*., 2005) where the basic histidine side chain is able to deprotonate the thiol group. Cys168 is not conserved between PTR1 of different species (**Figure 6.16B**). The equivalent residue in *Lm*PTR1 is Leu184 so compounds SDG 5 and SDG 73 may result in poorer inhibition of the *L. major* enzyme where the functional groups would clash. SDG 5 and SDG 73 are the first covalent inhibitors of recombinant *Tb*PTR1 described. Therefore, if covalent modification were the intended route for PTR1 inhibitor development, these structures demonstrate that Cys168 of *Tb*PTR1 can be exploited. Alternatively, this reactive group tethered to Cys168 could

be utilised in the development of more potent inhibitors similar to an approach employed in fragment-based drug discovery (Erlanson *et al*. 2004).

6.2.5. Isothermal titration calorimetry

ITC was attempted with the majority of compounds in **Table 6.2**. Suitable isotherms were achieved for seven SDG inhibitors. The major problem encountered causing this attrition rate involved compound solubility. The same buffer system as the inhibition assay was used and although optimisation may have allowed measurement of an increased number of samples, the derived thermal parameters would not have been comparable (Pierce *et al*., 1999; Olsson *et al*., 2008). Additionally, some titrations resulted in low heat changes with noisy background dilution effects (such as SDG 23 and SDG 53).

Two example thermodynamic profiles are provided in **Figure 6.17**. Data obtained from ITC are listed in **Table 6.4** and shown graphically in **Figure 6.18**. All ITC experiments were performed using fresh *Tb*PTR1 preparations and based on the measured stoichiometry of a control titration using MTX, the quantification of active enzyme was confirmed to be within an acceptable range ($N = 1 \pm 0.1$). Consequently, the only variable causing large changes in N should be the preparation of ligand. Excluding SDG 32 and 33, all crystal structures presented demonstrate 1:1 stoichiometry and have been analysed on that basis. Ligand concentrations were therefore retrospectively adjusted to normalise the variation in experimental N values and produce a value of $N = 1$. Both experimental and adjusted data are provided in **Table 6.4** and all further analyses have been carried out using the adjusted values. Data for SDG 32 and SDG 33 titrated against *Tb*PTR1 are not included here. The N values obtained were

approximately 0.7 and 0.6, respectively, and the crystal structures of these inhibitors show evidence of two molecules in the active site (**6.2.4.5**) so it is not appropriate to assume $N = 1$. As mentioned, two molecules binding in the active site may be a crystallographic artefact but equally, the particularly low stoichiometric value is suggestive of an unusual interaction, introducing additional complications to interpretation. ITC experiments where two ligand molecules bind to a single protein monomer should in theory return $N = 2$ and the isotherm show signs of two separate thermal events where a two-site binding model could be applied in the curve-fitting procedure. Here, perhaps the primary and secondary molecules' affinity for *Tb*PTR1 do not differ to a large enough extent to distinguish the interactions.

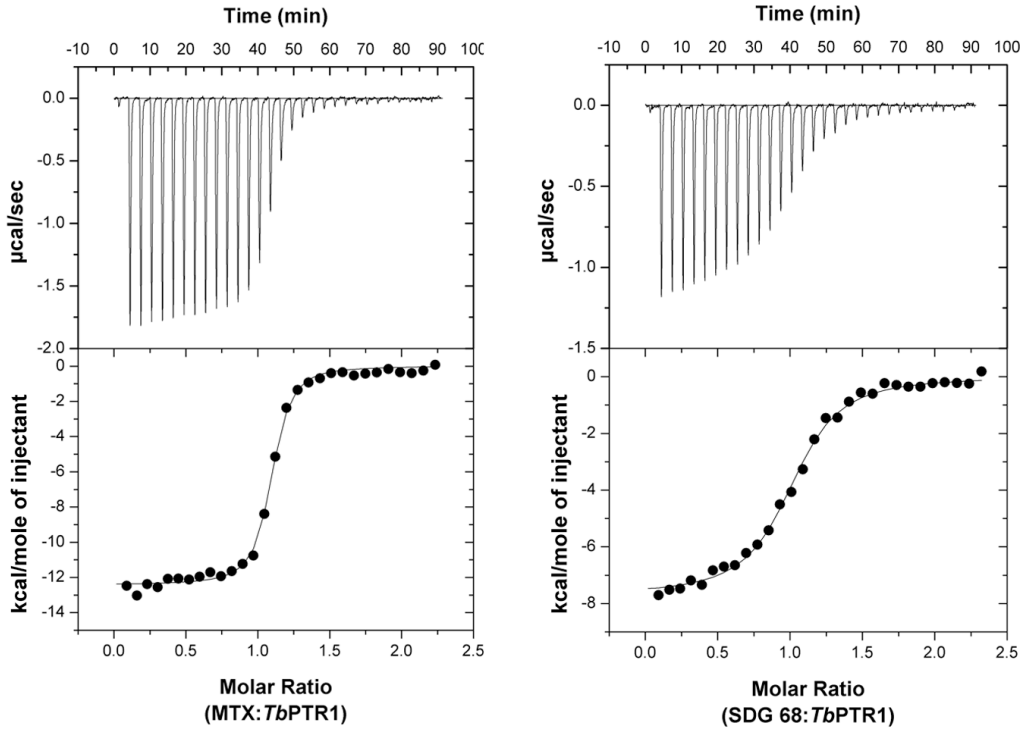


Figure 6.17. *Tb*PTR1 ITC isotherms
Representative thermodynamic profiles of MTX (left) and SDG 68 (right) titrated against *Tb*PTR1-$NADP^+$. The upper panels show the raw data and the lower panels show the curve of best fit following peak integration.

All isotherms displayed an acceptable *c* value of between 5 and 250 (**Table 6.4**). The optimum *c* for accurate data extraction can be considered to be approximately 40

(Broecker *et al*., 2011) while others suggest a far wider range of 1-1000 (Wiseman *et al*., 1989). In the case of MTX, with the highest *c* of approximately 230, performing the titration using a lower inhibitor concentration to produce a lower *c* may also have resulted in lower and perhaps immeasurable heat exchange. While, for isotherms with *c* < 10, the availability of ligand and solubility at higher concentrations were prohibitive.

		N	K_d (μM)	ΔG (kJ mol^{-1})	ΔH (kJ mol^{-1})	-TΔS (kJ mol^{-1})	*c*	LE (kcal mol^{-1} HA^{-1})	K_i (μM)
	Biopterin	0.71 ± 0.02	3.88 ± 0.32	-31.41 ± 0.21	-66.11 ± 1.60	34.69 ± 1.53	7.4	0.44	-
		1.00 ± 0.01	5.48 ± 0.29	-30.53 ± 0.15	-46.97 ± 1.15	16.44 ± 1.28	7.4	0.43	
	DHB	0.82 ± 0.07	5.35 ± 0.96	-30.66 ± 0.42	-63.24 ± 2.93	32.58 ± 2.51	6.2	0.43	-
		1.00 ± 0.01	6.22 ± 0.29	-30.36 ± 0.21	-50.10 ± 1.75	19.73 ± 1.94	6.5	0.43	
	MTX	1.08 ± 0.02	0.19 ± 0.04	-39.04 ± 0.50	-51.67 ± 1.47	12.64 ± 1.93	231.3	0.28	0.17
		1.00 ± 0.01	0.18 ± 0.04	-39.17 ± 0.51	-55.40 ± 2.29	16.23 ± 2.76	230.8	0.28	
Scaffold I	**SDG 57**	1.33 ± 0.01	1.95 ± 0.23	-33.15	-29.71 ± 0.38	-3.44	41.0	0.66	7.33
		1.00 ± 0.01	1.46 ± 0.17	-33.88	-39.62 ± 0.50	5.74	41.2	0.67	
	SDG 60	1.10 ± 0.03	0.61 ± 0.01	-36.05 ± 0.02	-35.02 ± 0.82	-1.03 ± 0.84	53.8	0.48	1.21
		1.01 ± 0.00	0.56 ± 0.00	-36.31 ± 0.00	-38.70 ± 1.20	2.40 ± 1.20	54.5	0.48	
Scaffold II	**SDG 65**	1.07 ± 0.03	0.38 ± 0.02	-37.23 ± 0.16	-31.09 ± 0.64	-6.14 ± 0.53	111.6	0.49	0.32
		1.00 ± 0.03	0.36 ± 0.03	-37.38 ± 0.21	-33.09 ± 1.12	-4.29 ± 0.91	111.3	0.50	
	SDG 67	1.45 ± 0.04	1.10 ± 0.11	-34.60 ± 0.25	-25.53 ± 0.94	-9.07 ± 1.14	53.2	0.49	0.40
		1.01 ± 0.01	0.76 ± 0.06	-35.51 ± 0.19	-36.73 ± 1.92	1.21 ± 2.11	53.2	0.50	
	SDG 68	1.14 ± 0.02	1.18 ± 0.21	-34.42 ± 0.48	-27.55 ± 0.35	-6.87 ± 0.82	39.6	0.46	0.48
		1.00 ± 0.01	1.04 ± 0.17	-34.75 ± 0.45	-31.40 ± 0.85	-3.35 ± 1.29	39.5	0.46	
	SDG 106	1.33 ± 0.04	4.12 ± 1.28	-31.31 ± 0.79	-23.18 ± 1.14	-8.13 ± 1.93	13.5	0.31	0.56
		1.00 ± 0.00	3.13 ± 1.04	-32.01 ± 0.85	-30.61 ± 0.76	-1.40 ± 1.62	13.5	0.32	
	SDG 132	0.90 ± 0.02	4.27 ± 0.76	-31.18 ± 0.47	-25.89 ± 0.79	-5.30 ± 1.20	8.7	0.28	1.28
		0.98 ± 0.01	4.69 ± 0.82	-30.94 ± 0.47	-23.72 ± 1.02	-7.23 ± 1.38	8.6	0.27	

Table 6.4. ITC thermodynamic data

Thermodynamic parameters obtained from ITC of seven SDG inhibitors (two scaffold I and five scaffold II, as indicated), MTX and substrates, biopterin and dihydrobiopterin (DHB). All values are the average of 2-3 independent experiments and corresponding standard deviations are shown. SDG 57 was a single experiment only and reported errors are those produced by curve-fitting procedures and not provided for the entropic value. Experimental values are reported in white rows, adjusted values are shaded grey. Ligand efficiency per heavy atom (HA) was calculated using ΔG in kcal mol^{-1} (1 J = 4.184 cal). K_i values from **Table 6.2** are provided for reference.

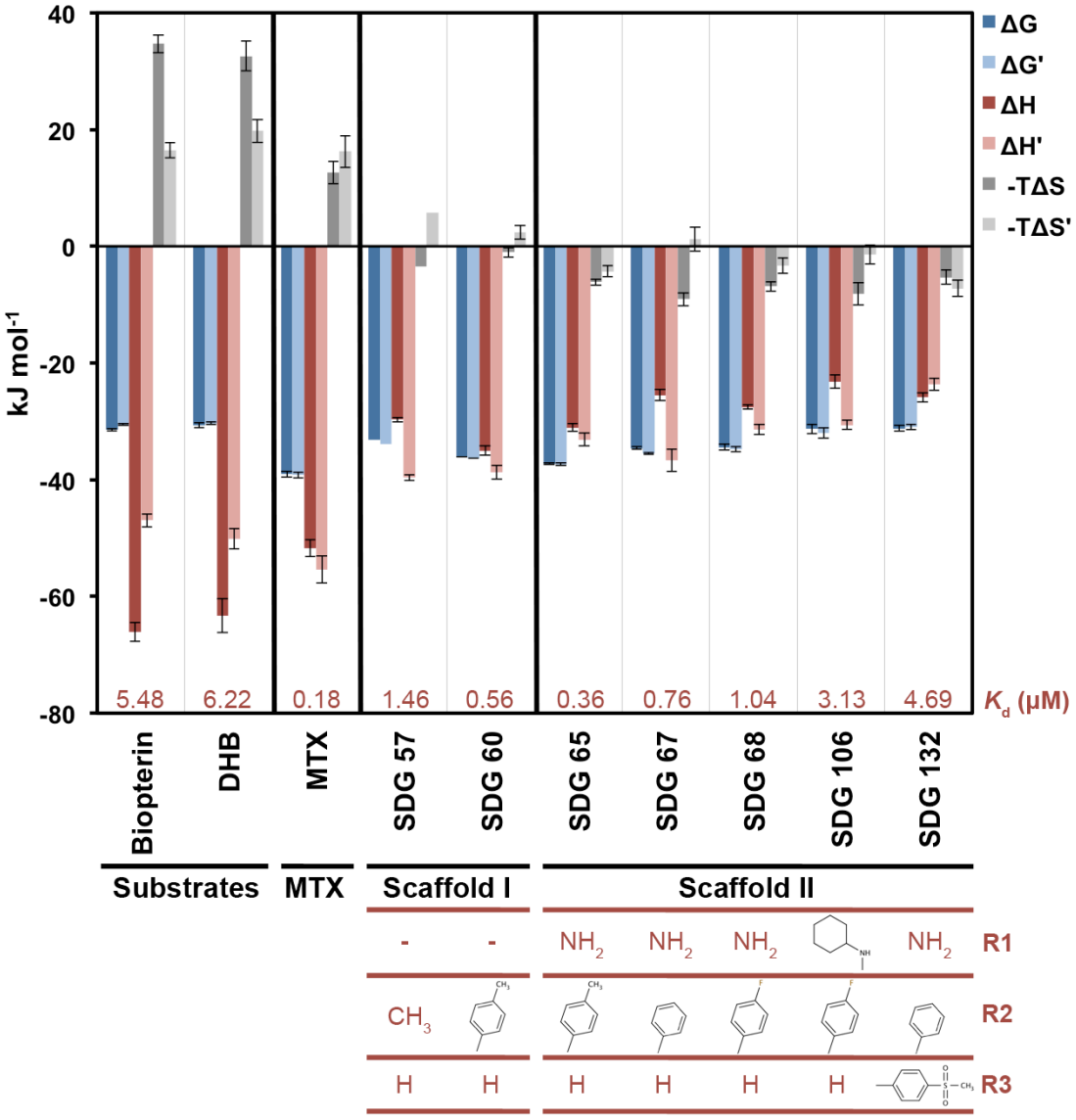


Figure 6.18. *Tb*PTR1 ligand thermodynamics
Thermodynamic data extracted from ITC isotherms, converted to the SI units, kJ mol^{-1}. All values are the average of 2-3 independent experiments and corresponding standard deviations are shown. SDG 57 was a single experiment only and reported errors are those produced by curve-fitting procedures and not provided for the entropic value. ΔG, ΔH and -TΔS are experimental values while ΔG′, ΔH′ and -TΔS′ indicate the adjusted values of Gibbs' free energy, enthalpy and entropy, respectively. K_d values from **Table 6.4** and R-group details of the SDG compounds are also provided.

Overall, a bimodal distribution is observed where ligand binding produces either a positive or negative change in entropy coupled with a negative free energy of binding and change in enthalpy. In terms of favourable binding, greater negative values of ΔG, ΔH and -TΔS are often considered most desirable (Holdgate, 2001). The enthalpic and

entropic changes of all ligands in **Figure 6.18** and **Table 6.4** are widely distributed (largest ΔΔH and Δ-TΔS are 31.68 and 26.96 kJ mol^{-1}, respectively) whereas ΔG fluctuates relatively little (largest ΔΔG 8.81 kJ mol^{-1}). This is due to high enthalpy-entropy compensation resulting from the opposing ΔH and -TΔS values of some ligands. The relationships between the thermodynamic parameters obtained from all ITC experiments are shown in **Figure 6.19**. The expected logarithmic relationship between ΔG and K_d is clearly demonstrated by the ligands, which range in affinity from 0.18 to 6.22 μM (**Figure 6.19A**). However, the scattered distribution of ΔH and -TΔS values is apparent (**Figure 6.19B-C**) and is not evident from experiments measuring potency alone. Additionally, while ΔG remains relatively stable with respect to the wide-ranging ΔH, the entropic penalty of ligands with favourable ΔH is shown in **Figure 6.19D** compared with the lower, more favourable, entropic contribution produced by ligands binding with poorer enthalpy. These relationships further highlight the apparent effect of entropy-enthalpy compensation in this set of ligands.

The compensatory effect may be a consequence of the hydrophobic properties of the SDG compound R-group substitutions compared with those of substrate or MTX. Hydrophobic interactions between two species tend to generate a favourable increase in the disorder of a system as water molecules arranged in an ordered manner around the individual hydrophobic groups are distributed into bulk solvent on binding (Ladbury and Chowdhry, 1996; Bronowska, 2011). The ordered intramolecular bonds between solvent first need to be broken and can contribute to the poor simultaneous enthalpic change. In contrast, biopterin, DHB and MTX all display relatively large positive -TΔS values. The polar groups attached to the pterin moiety are not surrounded by the same

ordered system of water molecules as a hydrophobic group and so, desolvation on binding is not entropically driven. In general, the net strength of the bonds formed to produce a favourable enthalpic contribution can also be accompanied by a greater loss of conformational freedom of both ligand and protein, giving a poor change in entropy (Leavitt and Freire, 2001). This compensatory effect can be difficult to overcome (Cooper *et al*., 2001). Improvements to entropy or enthalpy alone can compromise the highly favourable counterpart as demonstrated by the SDG compounds, which display a more favourable change in entropy but a reduced enthalpy when compared with substrates or MTX.

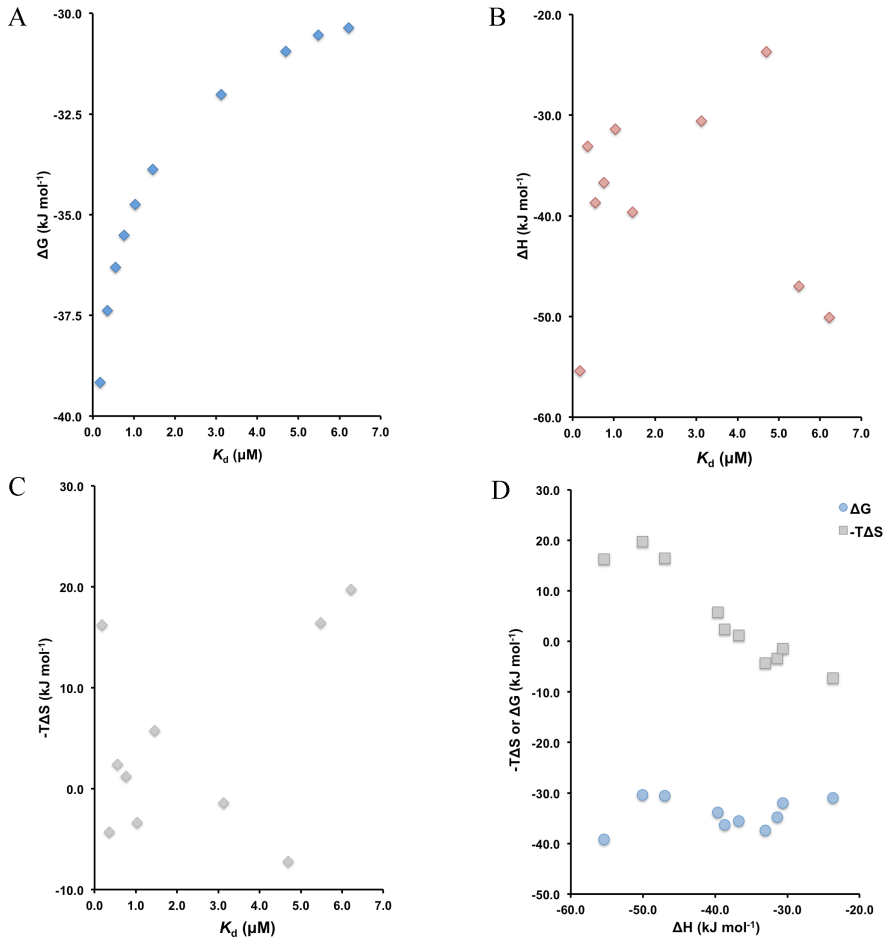


Figure 6.19. ITC thermodynamic relationships
Relationships between thermodynamic properties of ten ligands titrated against *Tb*PTR1. **(A)** – **(C)** show K_d *versus* ΔG (blue), ΔH (red) and -TΔS (grey). **(D)** shows the relationship between ΔH and both ΔG (blue circles) and -TΔS (grey squares).

6.2.5.1. Ligand efficiency

Ligand efficiency (LE) is the average contribution to the overall ΔG by each non-hydrogen atom (or 'heavy atom', HA) and was calculated using the experimental free energy of ligand binding based on the idea of Kuntz *et al*. (1999) where LE = - ΔG/HA (**Table 6.4**). ΔG is directly related to K_a (thus, K_d) via the relationship $\Delta G = -RT\ln K_a$ where R is the universal gas constant (8.314 J K^{-1} mol^{-1} or 1.987 cal K^{-1} mol^{-1}) and T is the temperature (K). As such, the LE calculated using K_d as suggested by Hopkins *et al*. (2004) is identical. In the absence of K_d or ΔG data for the majority of SDG compounds, the K_i or IC_{50} can therefore be used in their place (Hopkins *et al*., 2004) and are provided later. Comparisons of the experimentally derived LE of the 7 SDG inhibitors are indeed in close agreement with the alternative values.

SDG 57 is the smallest inhibitor in the series with only an additional methyl group at R2 of scaffold II and at 0.67 kcal mol^{-1} HA^{-1} (2.80 kJ mol^{-1} HA^{-1}), LE is higher than that of the other inhibitors tested, biopterin and DHB. The crystal structure of SDG 57 bound to *Tb*PTR1 confirms that although the overall affinity is relatively low (K_d 1.46 μM), the majority of atoms participate in the key contacts with active site residues. Addition of more complex R-groups must then also favourably interact with active site residues to maintain the high LE. However, the lower LE values of much larger ligands like MTX and SDG 132 suggest that this is not always the case and highlights the need for optimisation.

6.2.5.2. The effect of minor ligand changes

In order to assess the effect of different chemical substitutions on the thermodynamics of binding, it is important to first understand whether the observed differences are in fact significant. Holdgate (2001) suggests $\Delta\Delta H$ or $\Delta\Delta G$ of ≥ 4 kJ mol^{-1} or a change in $-T\Delta S$ of ≥ 8 kJ mol^{-1} is significant. While differences to this extent are observed, it is mainly between SDG ligands that also display the largest structural differences such as SDG 57 and SDG 132. Interpreting such differences is fraught with difficulties as there are many factors contributing to the observed changes. Small chemical substitutions therefore allow for the most reliable thermodynamic interpretations. While the introduction of a minor structural difference may not produce such enhanced thermodynamic effects, overall changes can still be used to gauge potential trends. SDG 65, 67 and 68 are all very similar compounds, differing by the addition or type of a single heavy atom. To aid in these comparisons, the seven SDG compounds studied by ITC are shown in **Figure 6.20**. **Table 6.5** provides values for the change in thermodynamic parameters of SDG 65 and SDG 68 with respect to SDG 67.

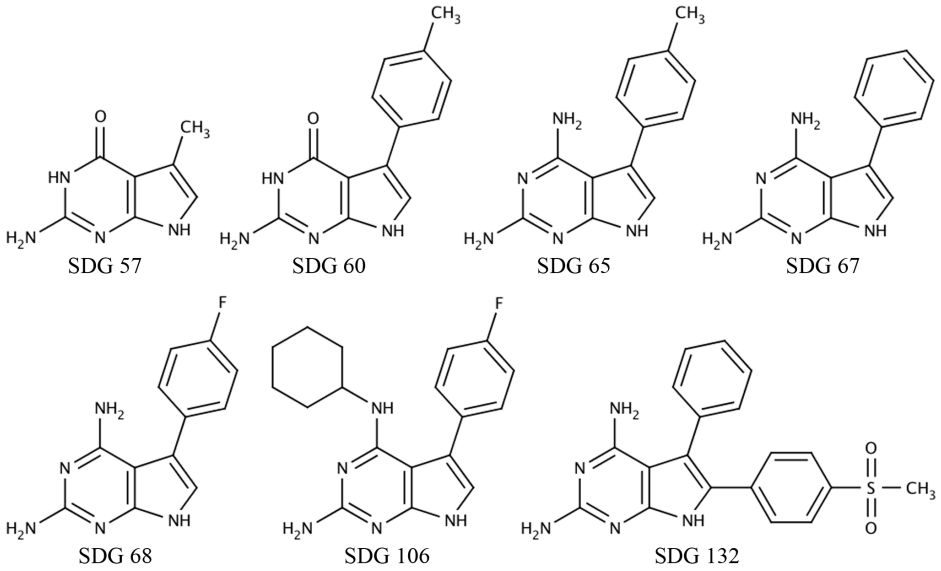


Figure 6.20. SDG compounds studied by ITC
The chemical structures of SDG 57, SDG 60, SDG 65, SDG 67, SDG 68, SDG 106 and SDG 132. SDG 57 and SDG 60 are scaffold II compounds, all others are scaffold I and atom numbering is as shown in **Figure 6.5**.

	K_d	ΔG	ΔH	-TΔS	LE
SDG 67	0.76	-35.51	-36.73	1.21	0.50
	ΔK_d	**ΔΔG**	**ΔΔH**	**Δ -TΔS**	**ΔLE**
SDG 65	-0.40	-1.87	3.64	-5.50	0.00
SDG 68	0.28	0.76	5.33	-4.56	-0.04

Table 6.5. Thermodynamic comparisons: SDG 67, 65, and 68.
Thermodynamic parameters of SDG 67 are taken from **Table 6.4**. All units are as provided previously, omitted here for simplicity. The observed change in these values are provided for SDG 65 and SDG 68, highlighted cyan if the change is more favourable or red if the parameter becomes less favourable with respect to SDG 67.

SDG 65 and SDG 68 contain methyl and fluoride groups, respectively, extending from the R2 phenyl group of SDG 67. LE values of 0.50 and 0.46 kcal mol^{-1} HA^{-1} coupled with the direct comparison of affinity (K_d = 0.36 and 1.04 μM, respectively) suggest that the methyl substitution is more favourable. Indeed, the ΔG and -TΔS values of SDG 65 are improvements to those of SDG 67. However, the more favourable entropic contribution of both SDG 65 and SDG 68 is accompanied by an enthalpic penalty. Both compounds display greater hydrophobicity and a larger surface area than the phenyl group alone, and may account for these effects. The increased length given by the methyl and fluoride R2 substitutions allow for an increased propensity to form van der Waals interactions with Trp221. While the compensatory differences produced by SDG 68 are detrimental to affinity, SDG 65 has the highest overall affinity for *Tb*PTR1 of these three compounds with no effect on ligand efficiency.

	K_d	ΔG	ΔH	-TΔS	LE
SDG 60	0.56	-36.31	-38.70	2.40	0.48
	ΔK_d	**ΔΔG**	**ΔΔH**	**Δ -TΔS**	**ΔLE**
SDG 65	-0.20	-1.07	5.61	-6.69	0.02

Table 6.6. Thermodynamic comparisons: SDG 60 and 65.
Thermodynamic parameters of SDG 60 are taken from **Table 6.4**. All units are as provided previously, omitted here for simplicity. The observed change in these values are provided for SDG 65, highlighted cyan if the change is more favourable or red if the parameter becomes less favourable with respect to SDG 60.

A pairwise comparison of only SDG 65 and SDG 60 (**Table 6.6**) provides an interesting observation. These compounds are partners belonging to scaffold I and II, respectively.

While exchanging the carbonyl of SDG 60 to the R1 amino group of SDG 65 results in an overall increase in affinity through improvements to -TΔS and ΔG, there is a significant loss in favourable enthalpy. There is little change in the hydrophobicity, suggested by cLogP values of 1.9 and 2.1 (calculated using *Molinspiration*, Cheminformatics, 2013) to account for such changes. When the inhibitors are analysed based on the same binding orientation, it is possible that the amino group of SDG 65 forms weaker hydrogen bonds with nearby water molecules or the carbonyl of SDG 60 can accept a hydrogen bond donated by Arg14. A 'moderate' hydrogen bond has been estimated to have a dissociation energy in the range of 17-62 kJ mol^{-1} (Steiner, 2002). But, ΔH reflects the net change of all non-covalent bonds and it is still extremely difficult and inappropriate to assign changes in ΔH to the breakage or formation of a single hydrogen bond, particularly when the strength of each bond is effectively unknown. Alternatively, this pattern of thermodynamic changes can signify an alteration in binding mode (Holdgate, 2001). While the crystal structures show both inhibitors adopt the substrate binding orientation, there is evidence that SDG 65 can also assume the MTX-like orientation in one subunit. The only obvious difference between poses is the loss of an ordered water molecule. This may account for the loss in enthalpy while entropy is gained from the water returning to bulk solvent. However, precise details of specific water molecules are difficult to assess even with guidance of crystallographic models (Davis *et al*., 2003).

	K_d	ΔG	ΔH	-TΔS	LE
SDG 68	1.04	-34.75	-31.40	-3.35	0.46
	ΔK_d	**ΔΔG**	**ΔΔH**	**Δ -TΔS**	**ΔLE**
SDG 106	2.09	2.74	0.79	1.95	-0.14

Table 6.7. Thermodynamic comparisons: SDG 68 and 106.
Thermodynamic parameters of SDG 68 are taken from **Table 6.4**. All units are as provided previously, omitted here for simplicity. The observed change in these values are provided for SDG 106, highlighted cyan if the change is more favourable or red if the parameter becomes less favourable with respect to SDG 68.

SDG 68 and SDG 106 (**Table 6.7**) both contain the same R2 group and comparisons can inform on the effect of the large N^4-cyclohexyl group extending from R1. It is clear that not only is affinity compromised but, while perhaps not considered significant, there are no thermodynamic advantages to the addition of this group. Structural data are available for both inhibitors and, as discussed in **6.2.4.2** and **6.2.4.3**, although the *Tb*PTR1 active site can accommodate such a substitution at R1, binding can induce some conformational changes, particularly the substrate-binding loop. It is difficult to pinpoint a single cause for the observed thermodynamic effects as the substitution is larger than a single atom but we can speculate that this conformational change is costly and may indeed be the source of the reduced affinity. A similar relationship is seen between SDG 67 and SDG 132, where all parameters are significantly less favourable when the complex R3 group is added. However, limited conclusions can be drawn as other substitutions at this position may be thermodynamically favourable.

	K_d	ΔG	ΔH	-TΔS	LE
SDG 57	1.46	-33.88	-39.62	5.74	0.67
	ΔK_d	**ΔΔG**	**ΔΔH**	**Δ -TΔS**	**ΔLE**
SDG 60	-0.90	-2.43	0.92	-3.34	-0.19

Table 6.8. Thermodynamic comparisons: SDG 57 and 60.
Thermodynamic parameters of SDG 57 are taken from **Table 6.4**. All units are as provided previously, omitted here for simplicity. The observed change in these values are provided for SDG 60, highlighted cyan if the change is more favourable or red if the parameter becomes less favourable with respect to SDG 57.

Comparing scaffold II compounds SDG 57 and SDG 60 **(Table 6.8**) gives a pattern analogous to that of SDG 67 and SDG 65 (**Table 6.5**). The addition of a more hydrophobic group at R2 produces favourable entropic changes while a small detrimental change in enthalpy results. Again, the net effect is an improvement in affinity but in this case, the additional group is larger than a methyl attachment and ligand efficiency is compromised.

Overall, even with both thermodynamic and structural data for six SDG compounds (no crystal structure is available for SDG 132), it remains extremely difficult to understand the precise underlying cause of the thermodynamic events. In particular, the contributions made by water must be considered and these are not always fully apparent in crystal structures. These studies have made efforts to characterise some of the observed effects produced by small changes to ligand structure and show that the entropic change produced on ligand binding has been improved. Although the SDG compounds are not enthalpically optimised compared to substrates and methotrexate, the improved entropic contributions do reduce the effect of enthalpy-entropy compensation and the resultant free energy values are comparable. Improving ΔS may be considered relatively easy compared with that of ΔH by the addition of bulky or hydrophobic groups (Ladbury *et al*., 2010; Ferenczy and Keserű, 2012) but solubility can be compromised if this strategy is overused. Optimisation of enthalpy has been shown to directly lead to improved inhibitors in other systems, such as HIV protease inhibitors (Ohtaka *et al*., 2004; Chaires, 2008). Here, future improvements should therefore focus on enhancing the enthalpy of binding by these scaffolds. It is more difficult to design and synthesise compounds that possess the ideal bond lengths and properties to make specific hydrogen bonds to improve enthalpy and these are often obtained fortuitously (Ladbury *et al*., 2010). One prospect is to perhaps mimic the apparent enthalpically favourable position 6 group of biopterin and DHB substrates in addition to the entropically favourable hydrophobic R3 phenyl system utilised in these SDG inhibitors. Or, to exploit the hydrogen bonding capabilities of Asp161 and Gly205 as shown by SDG 33 may improve enthalpy whilst a hydrophobic R2 substitution can generate favourable entropy.

6.2.6. Overall analysis

With one exception, all molecules in **Table 6.2** obey the guidelines often used to predict oral bioavailability, Lipinski's (2001) 'rule of five' (≤ 500 Da, ≤ 5 hydrogen bond donors, ≤ 10 hydrogen bond acceptors, cLogP ≤ 5). Containing substitutions at all three R-groups of scaffold I, SDG 107 is one of the best inhibitors of the series (K_i 0.2 μM) but violates the guidelines with a cLogP of 5.4. Changes in enthalpy, entropy and free energy of those studied by ITC are also within the range typical of drug-like molecules (-80 to 20 kJ mol^{-1}, -60 to 40 kJ mol^{-1} and -50 to -30 kJ mol^{-1}, respectively; Holdgate, 2001). The value of efficiency indices such as LE are being recognised and implemented within drug discovery strategies to guide the optimisation of a chemical series alongside conventional parameters such as potency (Abad-Zapatero, 2007). Here, LE has been monitored and while penalties were paid by some compounds, the overall LE was maintained above approximately 0.3 kcal mol^{-1} HA^{-1} even with relatively large phenyl substitutions (**Figure 6.21A**). A molecule containing 25 non-hydrogen atoms with 1 nM potency (K_i) would have an LE of approximately 0.5 kcal mol^{-1} HA^{-1} (Abad-Zapatero, 2007) and although there is still room for improvement, some SDG compounds boast an LE of greater than 0.5 (**Figure 6.21A**).

cLogP, the calculated octanol-water partition coefficient, values for each compound were predicted based on contributions from individual chemical groups using *Molinspiration* (Cheminformatics, 2013) and indicates the compound lipophilicity The same method for cLogP calculation is employed by the *ZINC* compound database (Irwin *et al*., 2012), widely used in virtual screening approaches. While there appears to be little correlation between K_d and cLogP (**Figure 6.21B**), this is based on a very low sample number from a set of compounds that displayed poor solubility. Indeed, the low

solubility of those not suitable for ITC was perhaps dominated by their higher lipophilicity. K_i values also appear widely distributed with respect to cLogP but the compounds with the poorest potency of this series are also the least lipophilic. Conversely, all compounds with a cLogP of greater than 2.5 also have an apparent K_i of less than 1.5 μM, suggesting that hydrophobic substitutions improve overall potency. This is further confirmed by the structural observations showing hydrophobic R2 and R3 groups extending from the catalytic centre to make favourable hydrophobic contacts with peripheral active site residues.

Of all compounds studied, those belonging to scaffold I with an R1-amino group typically displayed greater potency than the equivalent scaffold II compound where all other substitutions were identical. The trend between K_i and ΔG of scaffold I compounds appears to resemble that between K_d and ΔG (**Figure 6.21C**). The two scaffold I compounds, SDG 57 and SDG 60, appear as outliers. It may be inappropriate to draw conclusions from only two data points but at this stage, we cannot completely rule out that these two scaffold II compounds were to some extent themselves reduced by *Tb*PTR1 during the spectrophotometric assay. If true, these compounds are able to block substrate binding but NADPH oxidation can still occur through hydride donation to the pyrrolopyrimidine C5 and an artificially low inhibition rate is observed. ITC was performed in the presence of oxidised cofactor, $NADP^+$, and only the effect of titrated ligand binding is measured. Many scaffold II compounds displayed reasonable *Tb*PTR1 inhibition but the possibility of inhibitor reduction may suggest that scaffold I compounds are more valuable for further study.

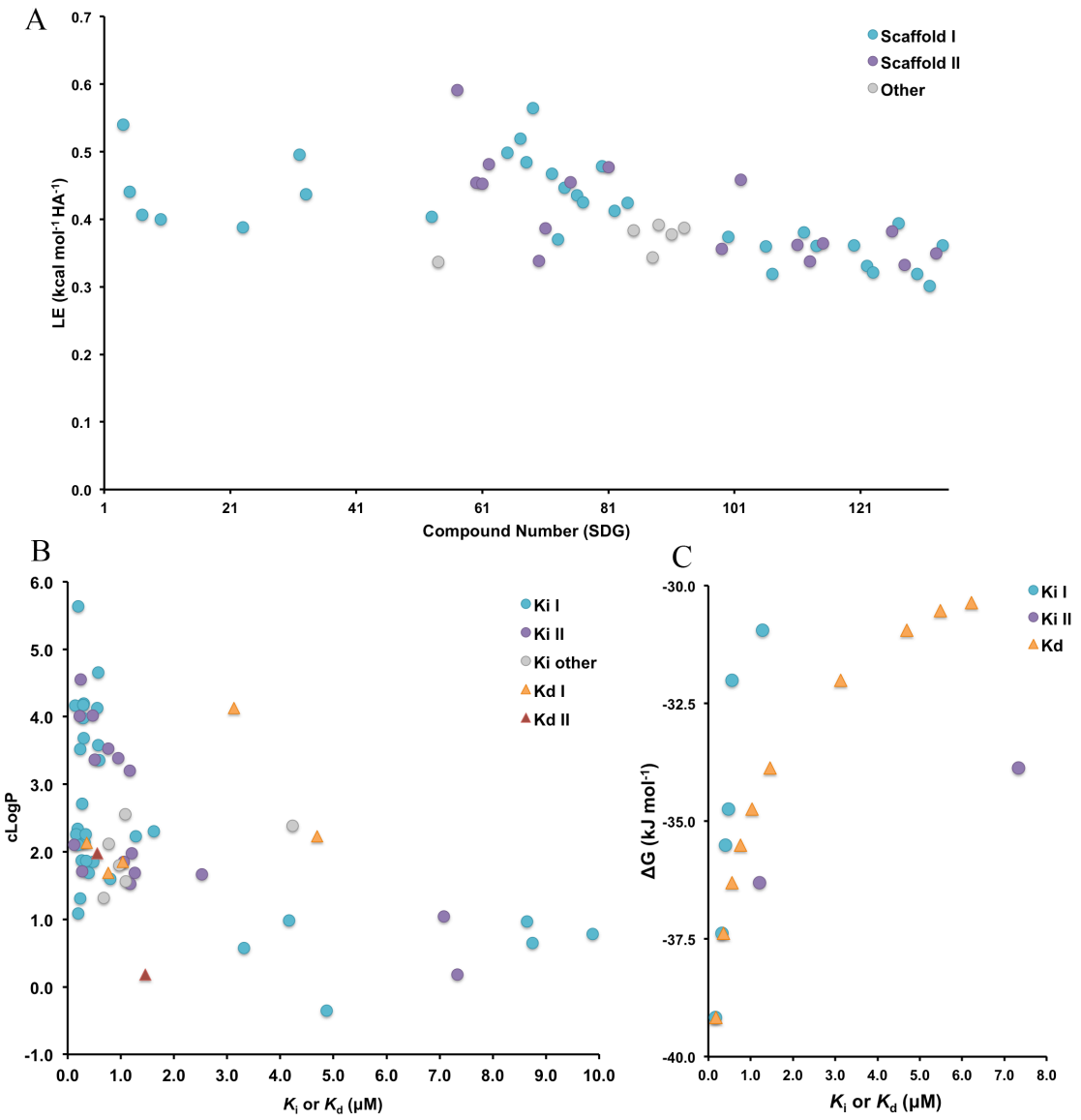


Figure 6.21. Ligand efficiency, LogP and kinetic relationships
(A) Ligand efficiency according to SDG compound number and coloured according to scaffold (scaffold I, green; scaffold II, purple; other, grey circles). **(B)** Calculated LogP values (cLogP) are plotted against the K_i values of all compounds in **Table 6.2** (coloured as in **(A)**) and the K_d of ligands in **Table 6.4** (scaffold I, orange; scaffold II red triangles). **(C)** The relationship between ΔG and K_i (scaffold I and MTX, green circles; scaffold II, purple circles) or K_d (orange triangles, also shown in **Figure 6.19**).

The scaffolds alone are weak inhibitors of PTR1 (Tulloch *et al*., 2010) and the addition of only a methyl group at R2 also results in relatively poor inhibition (SDG 57). An additional phenyl substituent at either R2 (SDG 62) or R3 (SDG 102) immediately improves the inhibitory effect. Additional chemical modifications were therefore introduced to instigate and identify further improvements.

Substitutions larger than NH_2 can be made at R1 with an improvement in potency (for example, SDG 122 to SDG 123 and SDG 100 to SDG 120 or SDG 107) but can force the ligand into a suboptimal position that cannot form key hydrogen bonds directly with cofactor or Ser95, resulting in weak affinity and poor heat exchange on ITC (SDG 23).

The use of halogen substituents such as a 4-bromophenyl group at R3 produced one of the highest inhibition rates of both scaffolds (SDG 126 and SDG 127). Addition of fluorine at the same position had a lower effect (SDG 111 and 112) but still improved the potency over the non-halogenated compounds (SDG 99 and 100). Combining the R3 4-fluorophenyl with the same group at R2 (SDG 114) had a small detrimental effect where replacement with a 3-chlorophenyl (SDG 133 and SDG 134) partially regains the potency. Additional groups beyond the R2 phenyl generally showed no overall advantage over the ring substitution alone (including SDG 111 to SDG 113, SDG 112 to SDG 114, SDG 67 to SDG 68) except in the case of SDG 65 where a methyl addition did enhance affinity through an improvement in entropy although the scaffold II equivalent behaved marginally better without this addition (SDG 62 compared to SDG 60). While the R2 4-fluorophenyl of SDG 68 reduced overall affinity, the entropic contribution was more favourable than the phenyl alone so should not be ruled out. The use of 4-chlorophenyl groups at R2 and R3 (SDG 115) resulted in greater inhibition than the equivalent 4-fluoro substituents (SDG 113) but it is not known which chloro-group had the greatest effect. Conventionally, halogen components are utilised to improve ADMET properties (absorption, distribution, metabolism, excretion and toxicity) to ultimately aid in producing a drug-like molecule with good bioavailability (Bronowska, 2011). More recent applications suggest that halogenated ligands can inform structure based design studies through optimisation of halogen bonds (Lu *et al*.,

2009; Xu *et al*., 2011) that form based on the anisotropic distribution of charge surrounding the atom (Kolář *et al*., 2013), analogous to hydrogen bonds (Auffinger *et al*., 2004). While it is possible that the presence of non-covalent attraction between the bromo-group of SDG 126 or SDG 127 and His267' or Cys168 carbonyl may support the observed inhibitory improvements, it is likely primarily attributed to the ability of the large bromine atom to fill the cavity lined by Tryp221, Cys168 and Met213.

While extension from a phenyl group at R2 may not be favourable in most cases other than SDG 65, lengthening the distance of the phenyl from the core scaffold by either an ethyl (SDG 80) or an ethynyl (SDG 53) extension did appear to improve inhibition. The crystal structures show that for SDG 53, inhibition was enhanced through interactions with the hydrophobic pocket bordered by Leu209, Met213 and Trp221 that is better reached by the longer extension. SDG 80, however, adopts the MTX-like orientation and the ethylphenyl group does not fill the pocket to the same degree, possibly contributing to the slightly lower observed potency. SDG 53 contains a pyrrolidine substitution at R1, preventing the MTX-like orientation through clashes with Tyr174 and Asp161 while SDG 80 would share a hydrogen bond with Tyr174 in either orientation. A combined R1-pyrrolidine and R2-phenethyl has not yet been investigated.

Compounds with a carbonitrile substituent at R2 generally inhibited *Tb*PTR1 well. While this group alone only slightly improved inhibition from a methyl (SDG 69), additions at R3 made significant improvements. In the spectrophotometric assay, SDG 74 and SDG 75 appeared to be the best performing compounds of scaffold I and II, respectively, where the simultaneous R3-styryl partially fills the pocket generated by

Met163, Cys168, Trp221 and also bordered by His267' from subunit D. Similar to the R2 phenyl substituent, it appears that a longer extension of the phenyl group is beneficial at R3. However, the potency drops marginally when either a rigid phenethynyl (SDG 76) or a flexible phenethyl group (SDG 84) replaces the styryl.

When the two main scaffolds are not strictly obeyed, such as replacement of the scaffold I 7-NH with O (SDG 85, SDG 88, SDG 89, SDG 91, SDG 93), inhibition did not improve beyond approximately 0.7 μM (K_i) but extensive R-groups were not examined. Structural data are not available to examine the binding of these compounds due to difficulties generating diffraction-quality crystals but we can speculate that the MTX-like pose would be adopted. The O in place of the scaffold I 7-NH may assume the binding position of the substrate 4-carbonyl while the R1-amino group is situated to donate a hydrogen bond to Tyr174.

To conclude, these studies have shown that hydrophobic groups extending to interact with Trp221 not only improve potency but do so by enhancing the favourable entropic contribution to binding. SDG 32 and SDG 33 highlight that a second site where hydrogen bond donors close to Asp161 and Gly205, replacing an ordered water molecule, can be utilised in a single molecule that simultaneously fulfils the well-characterised primary binding features. This strategy in combination with a hydrophobic extension may then improve the enthalpic change on binding which is compromised when entropy alone is optimised. Indeed, multiple substitutions on the framework employed here can be advantageous, particularly with extended phenyl groups and there remains potential for substitutions at all three R-groups with early attempts appearing promising (SDG 107 or SDG 123, no structure available). The

physical properties of such compounds must be considered carefully as large hydrophobic groups with high cLogP values are likely to further reduce solubility and can lead to poor bioavailability and toxicity (Hughes *et al*., 2008) but the use of halogenated substituents can be utilised to reduce the effects as well as gain potency.

6.2.7. Additional studies

The majority of the work presented forms part of a wider project to improve inhibitors targeting PTR1 in the search for new therapies against HAT and Leishmaniasis. A brief summary of additional studies and some comparisons with results of biochemical investigations are provided.

6.2.7.1. Compound synthesis

As noted in **5.2**, chemical synthesis of all SDG compounds was performed by the laboratories of Professor Colin Suckling and Dr Colin Gibson at the department of Pure and Applied Chemistry, University of Strathclyde (Glasgow, UK).

6.2.7.2. Biological activity

Experiments to assess the effect of SDG compounds on the activity of *T. brucei brucei* parasites *in vitro* were performed by Professor Michael Barrett's laboratory at the Institute of Infection, Immunity and Inflamation, University of Glasgow (UK). Compounds that demonstrated activity against both recombinant *Tb*PTR1 and *T. b. brucei* cells were progressed for study against HEK (human embryonic kidney) and *L. mexicana* amastigote cell lines. In some cases, compounds displayed inhibition of recombinant *Tb*PTR1 but were inactive against the parasite, suggestive of problems in cell uptake. For example, SDG 65 gave a K_i of 0.32 μM and K_d of 0.36 μM against

*Tb*PTR1 but resulted in an IC_{50} of 170 μM when evaluated against *T. b. brucei* cells. Conversely, compounds that appeared effective against *T. b. brucei* but showed little inhibition of *Tb*PTR1 were perhaps acting in a non-specific manner. Compounds were tested using both folate-rich HMI9 and folate-deficient CMM growth media with similar results, indicative of a non-competitive relationship with folate unlike the DHFR inhibitor, MTX.

Some molecules that have been shown to target PTR1 with submicromolar K_i values were found to display high nanomolar or low micromolar trypanocidal activity whilst not producing significant HEK cell toxicity. IC_{50} values of less than 1 μM against *T. b. brucei* with at least 50-fold greater IC_{50} against HEK cells were considered significant. For example, SDG 112 displayed an IC_{50} against *T. b. brucei* of 0.321 μM (HMI9) or 0.082 μM (CMM) compared to 49.190 μM against HEK cells. As a reference, MTX IC_{50} against *T. b. brucei* was 3.656 μM in HMI9 media or 0.011 μM in folate deficient media, a change of > 300-fold. Most compounds assessed also showed greater activity against *T. b. brucei* than *L. mexicana* amastigotes where activity was typically poorer than that of MTX and amphotericin B controls (IC_{50} approximately 7.4 μM and 0.2 μM, respectively). Overall, SDG 99, 100, 111-114, 122, 124, 127, 128, 133 and 134 were found to be most effective with IC_{50} values against *T. b. brucei* lower than that of MTX in folate-rich media, from 0.265 μM (SDG 122) to 2.247 μM (SDG 100) as shown in **Table 6.9**. While not restricted to a single scaffold, all of these compounds contain phenyl substitutions at both R2 and R3, shown structurally to generate hydrophobic interactions in two active site sub-pockets. Indeed, SDG 127 was the most potent in the

series against *Tb*PTR1 and, while other scaffold I compounds were toxic *in vivo*, both SDG 127 and SDG 122 were able to significantly reduce parasitaemia.

Compound	K_i (μM)	IC_{50} (μM)			
		***T.b.brucei* CMM media**	***T.b.brucei* HMI9 media**	**HEK**	***L.mexicana***
SDG 112	0.24	0.321	0.082	49.190	19.460
SDG 122	0.58	0.265	0.083	39.140	20.450
SDG 128	0.95	0.396	0.135	33.180	9.140
SDG 114	0.30	0.594	0.149	47.340	21.180
SDG 134	0.29	0.392	0.185	34.593	24.400
SDG 127	0.14	0.970	0.248	39.630	34.090
SDG 99	1.17	0.640	0.407	Not available	7.652
SDG 100	0.59	2.247	0.583	62.883	>100
SDG 111	0.51	0.738	0.614	>200	>100
SDG 113	0.76	1.390	0.739	160.600	>100
SDG 133	0.47	1.405	0.767	57.700	>100

Table 6.9. Biological activity
The most effective compounds that were found to display IC_{50} values of less than 1 μM against *T. b. brucei* and low toxicity, where IC_{50} values against HEK cells were at least 50-fold greater than those against *T. b. brucei*. Compounds are listed in order of effectiveness when all parameters are considered. Scaffold I compounds are in white rows and scaffold II compounds are highlighted grey. Biochemical K_i values against recombinant *Tb*PTR1 and *L. mexicana* IC_{50} values are also shown. Error values are omitted for clarity. K_d values were not achieved for these compounds by ITC.

6.2.7.3. PTR1 selectivity

A number of compounds were analysed against *Tc*DHFR activity by the laboratory of Professor Debasish Chattopadhyay at the University of Alabama (Birmingham, Alabama, USA). SDG 5, 23, 32, 53, 57, 60 and 68 were tested and displayed little or no DHFR inhibition indicating specificity for PTR1. Previous results of collaborators have also shown that pyrrolopyrimidines do not inhibit *Hs*DHFR.

6.2.7.4. Pharmacokinetics

Selected compounds were assessed by the Drug Metabolism and Pharmacokinetics group (DMPK) of Dr. Kevin Read (Biological Chemistry and Drug Discovery, University of Dundee, UK). Of those tested, it was found that toxicity levels were above ideal and prevented effective assessment using a HAT mouse model. Together with relatively modest nanomolar *in vitro* potency and the low fraction of unbound compound in plasma protein binding studies, high dosage levels were necessary but this amplified the toxic effects. Despite this, SDG 99, SDG 112 and SDG 122 displayed acceptable exposure levels on initial dosage and for SDG 112 in particular, the exposure was maintained by a longer half-life. Overall, DMPK studies provided proof-of-concept that compounds of this series do have potential if further optimised to reduce toxicity.

6.3. Part II summary and concluding remarks

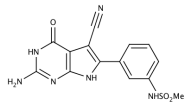
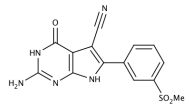
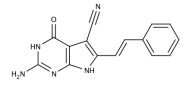
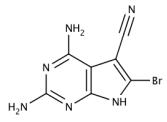
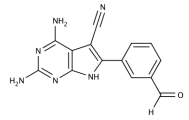
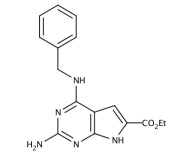
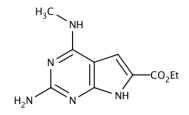
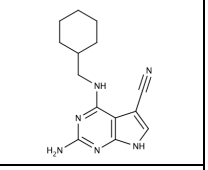
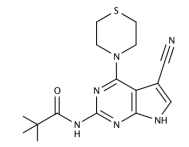
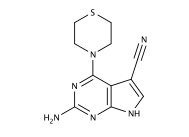
The crystal structure of PTR1 from *L. donovani* has been determined using diffraction data to 2.5 Å resolution. Unfortunately, this crystal form of *Ld*PTR1 is not amenable for further ligand interaction studies. A sulfate obtained from the crystallisation mixture blocks the cofactor binding site and due to the sequential ordered mechanism of PTR1, neither substrate or inhibitor can then form a complex. As no fully apo-PTR1 structures have been reported, this model of *Ld*PTR1 confirms the structural importance of the NADP(H) cofactor. It also provides new information on the mobile nature of the β5-α5 loop as well as the important substrate binding loop and other residues known to be critical for function.

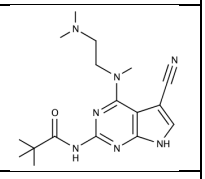
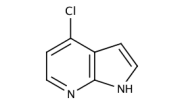
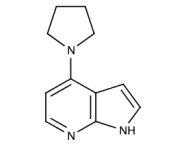
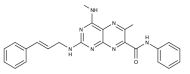
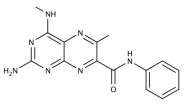
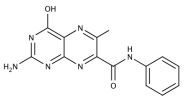
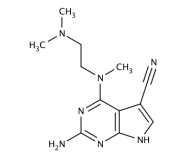
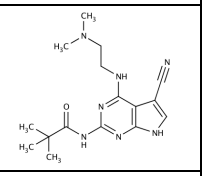
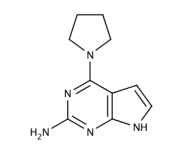
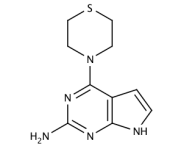
Recombinant *T. brucei* PTR1 was used to assess the inhibition of over 100 novel small molecules, 23 of which were co-crystallised in a ternary complex with *Tb*PTR1 and oxidised cofactor, $NADP^+$. The thermodynamics of ligand binding was also characterised for seven inhibitors and compared with that of substrate and an archetypal dihydrofolate reductase inhibitor, methotrexate, while an additional high resolution crystal structure of PTR1 complexed with another known antifolate, trimetrexate, was also determined. Examination of these structures has revealed a number of interesting new features. Most adopt a binding mode similar to substrate and allowed the development of larger substitutions at positions 5 and 6 of a pyrrolopyrimidine core. Previously unexplored position 4 substitutions were also shown to be accommodated through movement of the flexible substrate-binding loop. Two compounds attached covalently to an active site cysteine and two compounds were observed to bind at multiple locations in the active site. While a similar binding mode has been shown previously, these molecules were able to simultaneously make important contacts with cofactor in substrate-like orientation as well as at the secondary site, providing guidance for future developments.

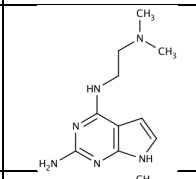
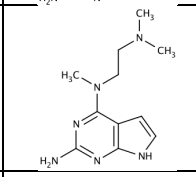
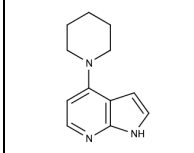
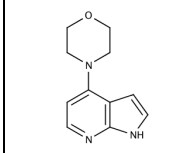
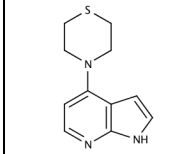
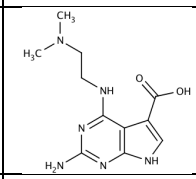
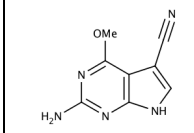
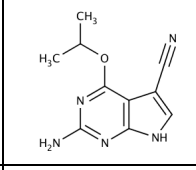
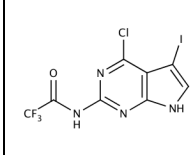
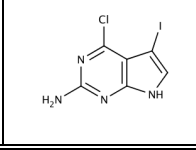
Valuable data can be obtained from ITC but this is strongly dependent on the behaviour of both protein and compound under the selected experimental conditions. Here, many compounds could not be tested due to their relatively low aqueous solubility. Based on the seven SDG inhibitors for which acceptable experimental data were obtained, some difficulties in data interpretation were also encountered. Crystallographic information was important to validate ligand:protein stoichiometry where it was necessary to account for errors that might have been introduced in sample preparation. While structural data cannot always identify the precise cause in changes to enthalpy and

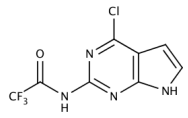
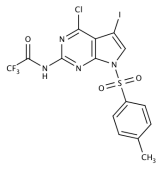
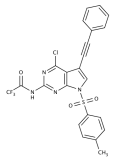
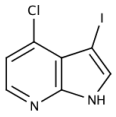
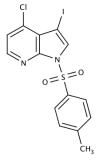
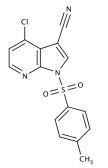
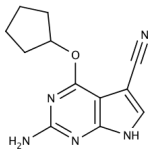
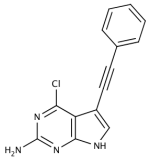
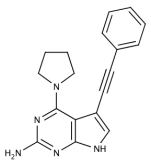
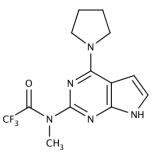
entropy as it can be considered a ‘snapshot’ of a dynamic system, it remains extremely important to employ a strategy where multiple biophysical techniques can be used complement one another and allow for the most reliable overall analysis. ITC has highlighted the improvements made to the interaction entropy as well as the importance of further optimising the enthalpic contribution. Together with structural and kinetic data, the ligand binding profiles and SAR produced by these studies forms a solid platform for future advances targeting the inhibition of trypanosomatid PTR1.

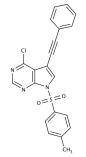
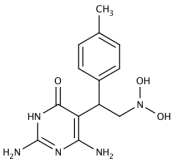
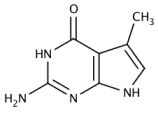
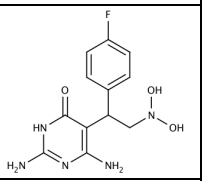
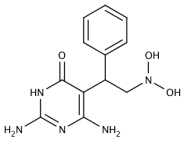
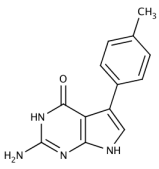
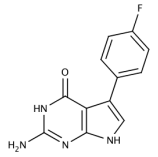
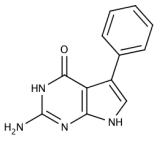
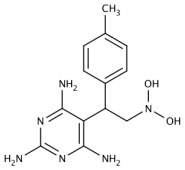
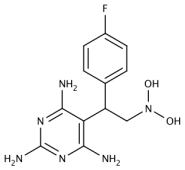
APPENDIX A

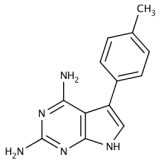
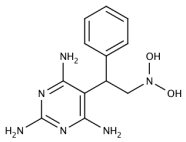
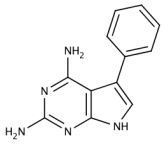
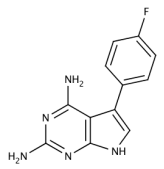
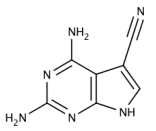
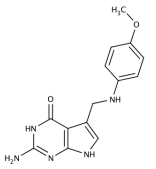
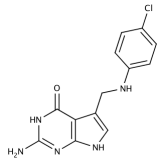
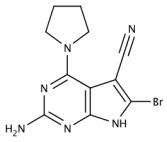
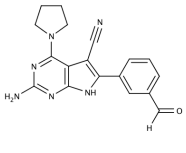
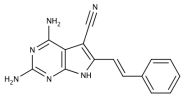
Name	Structure
SDG 1 (I)	
SDG 2 (I)	
SDG 3 (I)	
SDG 4 (K)	
SDG 5 (K, C)	
SDG 6 (I)	
SDG 7 (K)	
SDG 10 (K)	
SDG 22	
SDG 23 (K, C)	

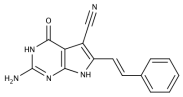
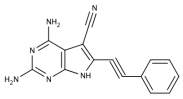
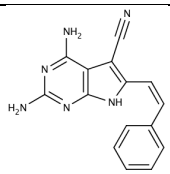
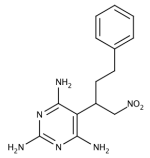
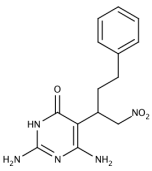
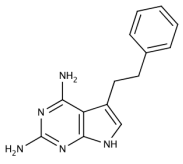
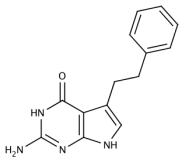
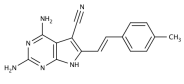
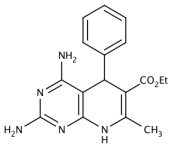
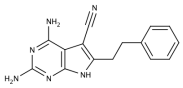
Name	Structure
SDG 24 (I)	
SDG 25	
SDG 26	
SDG 27	
SDG 28	
SDG 29 (I)	
SDG 30	
SDG 31	
SDG 32 (K, C)	
SDG 33 (K, C)	

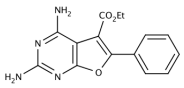
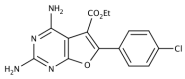
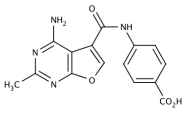
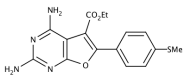
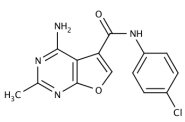
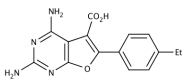
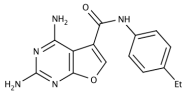
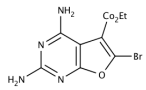
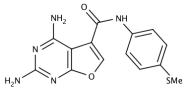
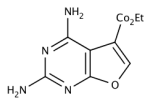
Name	Structure
SDG 34	
SDG 35	
SDG 36	
SDG 37	
SDG 38	
SDG 39	
SDG 41	
SDG 42	
SDG 43	
SDG 44	

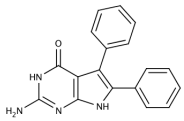
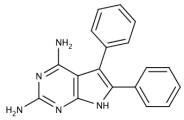
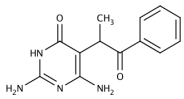
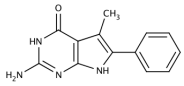
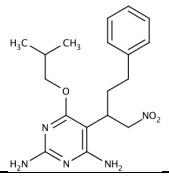
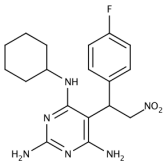
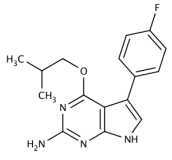
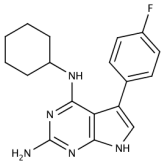
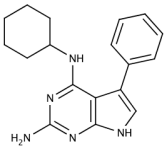
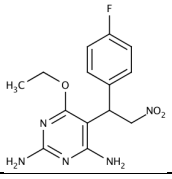
Name	Structure
SDG 45	
SDG 46	
SDG 47 (I)	
SDG 48	
SDG 49 (I)	
SDG 50	
SDG 51	
SDG 52	
SDG 53 (K, C)	
SDG 54 (K)	

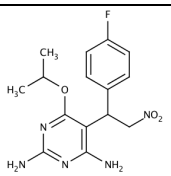
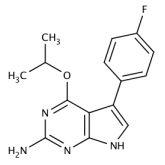
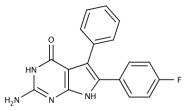
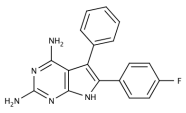
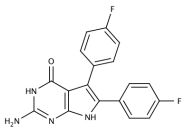
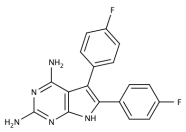
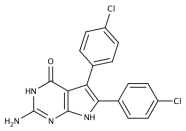
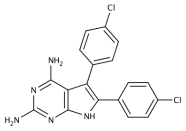
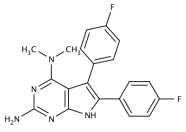
Name	Structure
SDG 55 (I)	
SDG 56	
SDG 57 (K, C)	
SDG 58	
SDG 59	
SDG 60 (K, C)	
SDG 61 (K)	
SDG 62 (K)	
SDG 63 (I)	
SDG 64	

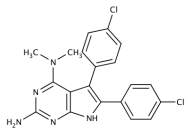
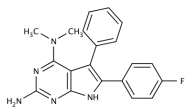
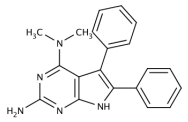
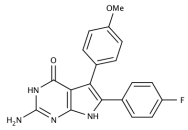
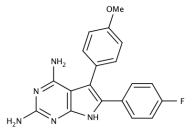
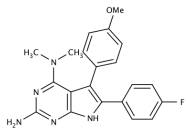
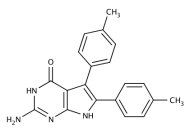
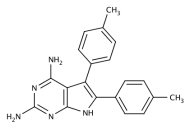
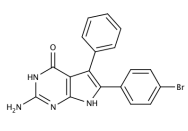
Name	Structure
SDG 65 (K, C)	
SDG 66 (I)	
SDG 67 (K, C)	
SDG 68 (K, C)	
SDG 69 (K)	
SDG 70 (K)	
SDG 71 (K)	
SDG 72 (K)	
SDG 73 (K, C)	
SDG 74 (K)	

Name	Structure
SDG 75 (K, C)	
SDG 76 (K)	
SDG 77 (K)	
SDG 78	
SDG 79	
SDG 80 (K, C)	
SDG 81 (K)	
SDG 82 (K, C)	
SDG 83	
SDG 84 (K)	

Name	Structure
SDG 85 (K)	
SDG 86	
SDG 87	
SDG 88 (K)	
SDG 89 (K)	
SDG 90	
SDG 91 (K)	
SDG 92	
SDG 93 (K)	
SDG 94	

Name	Structure
SDG 99 (K, C)	
SDG 100 (K, C)	
SDG 101	
SDG 102 (K)	
SDG 103	
SDG 104	
SDG 105	
SDG 106 (K, C)	
SDG 107 (K, C)	
SDG 108	

Name	Structure
SDG 109	
SDG 110	
SDG 111 (K)	
SDG 112 (K)	
SDG 113 (K)	
SDG 114 (K)	
SDG 115 (K)	
SDG 116 (I)	
SDG 117 (I)	

Name	Structure
SDG 118 (I)	
SDG 119 (I)	
SDG 120 (K, C)	
SDG 121 (I)	
SDG 122 (K, C)	
SDG 123 (K)	
SDG 124	
SDG 125 (I)	
SDG 126 (K, C)	

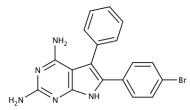
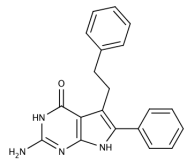
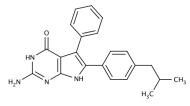
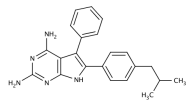
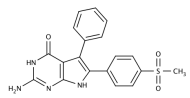
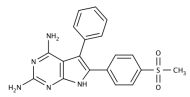
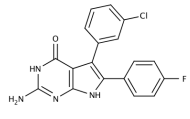
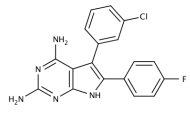
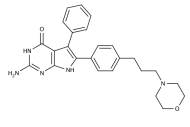
Name	Structure
SDG 127 (K, C)	
SDG 128 (K, C)	
SDG 129	
SDG 130 (K)	
SDG 131 (I)	
SDG 132 (K)	
SDG 133 (K)	
SDG 134 (K)	
SDG 135	

Table A.1. SDG compounds

SDG compounds are listed with according to name. Compounds insoluble at the required concentration are marked **(I)**. Compounds listed in **Table 6.2** with associated kinetic data are marked **(K)** and those with crystal structures determined are marked **(C)**. All other compounds were not progressed beyond screening at two concentration points. The structures of twelve insoluble and three soluble compounds assessed are unavailable and not listed here.

APPENDIX B

Figures of all 24 *Tb*PTR1-ligand complex structures are shown individually along with selected active site residues and solvent. Subunit A was used to prepare all figures where inhibitors adopt the same conformation in all tetramer chains. For ligands that displayed significantly different conformations between chains, a second figure is provided. Atoms are all coloured accordingly: C, yellow ($NADP^+$), cyan (*Tb*PTR1) or grey (inhibitor); O, red; N, blue; S, gold; P, orange; F, pale blue; Br, brown. Phe97 is shown as thin lines and is unlabelled. Water molecules within 3.5 Å of the inhibitor are depicted as red spheres.

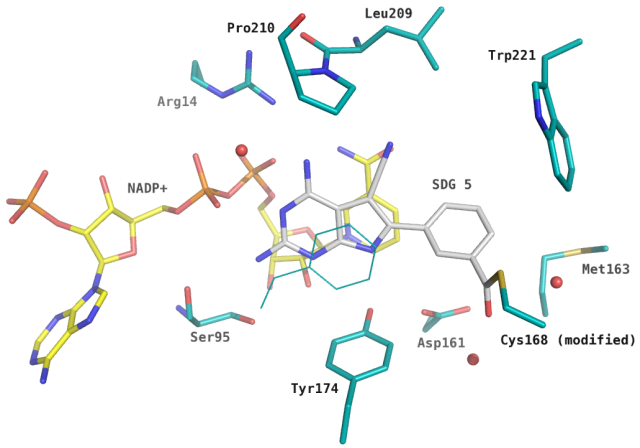


Figure B.1. SDG 5
2,4-diamino-6-(3-formylphenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile

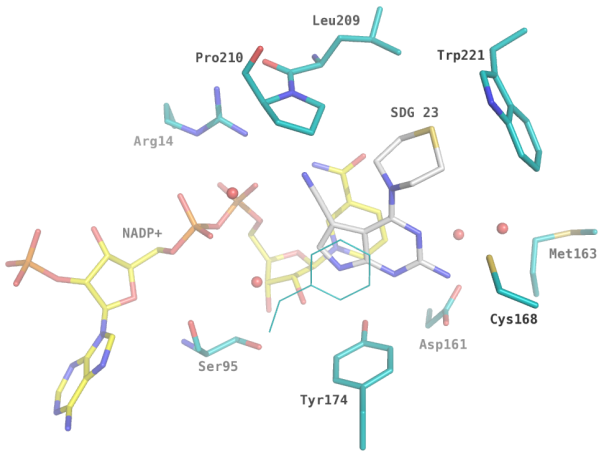


Figure B.2. SDG 23
2-amino-4-thiomorpholino-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile

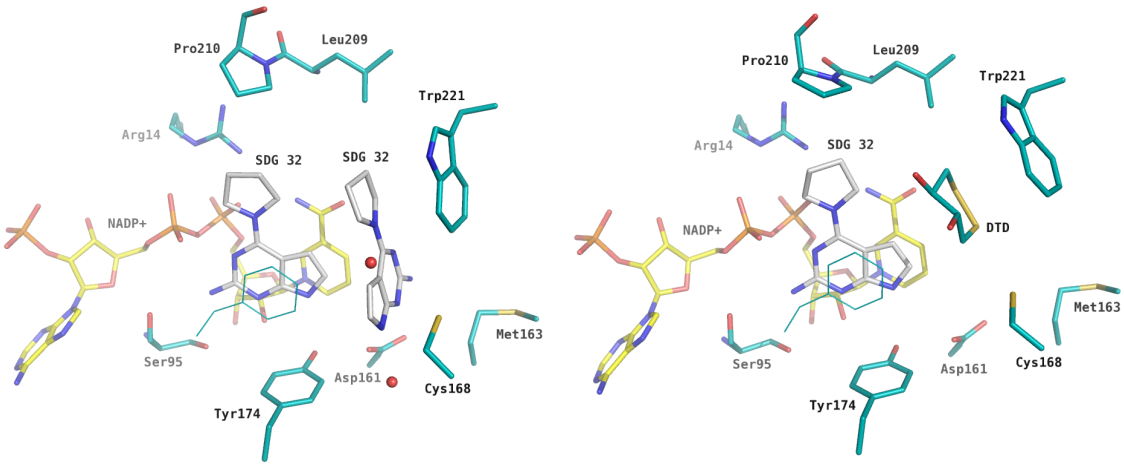


Figure B.3. SDG 32
Subunit A, left; subunit B, right.
4-(pyrrolidin-1-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-2-amine

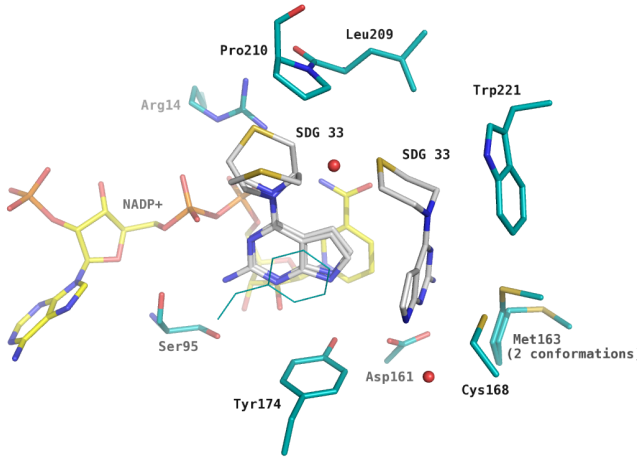


Figure B.4. SDG 33
4-thiomorpholino-7*H*-pyrrolo[2,3-*d*]pyrimidin-2-amine

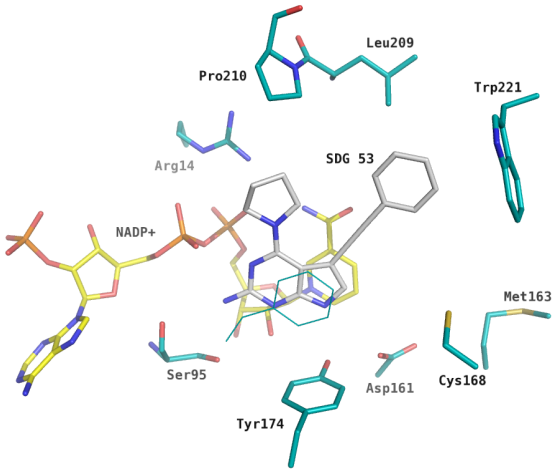


Figure B.5. SDG 53
5-(phenylethynyl)-4-(pyrrolidin-1-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-2-amine

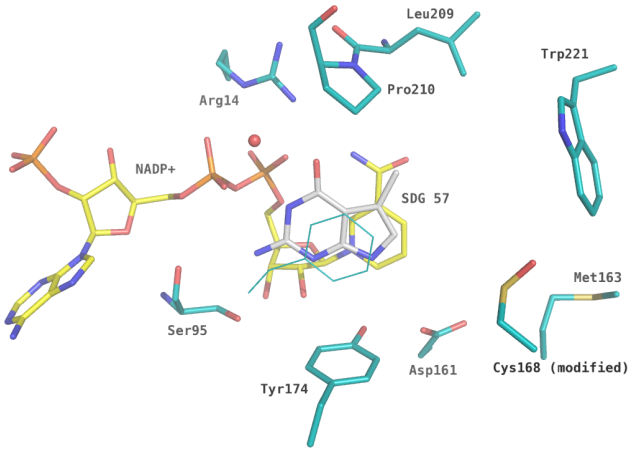


Figure B.6. SDG 57
2-amino-5-methyl-3*H*-pyrrolo[2,3-*d*]pyrimidin-4(7*H*)-one

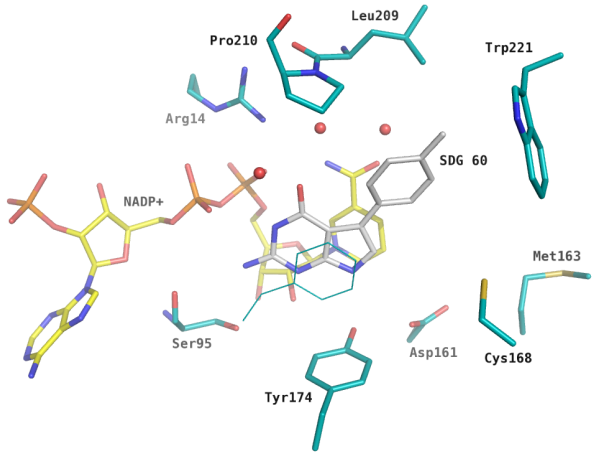


Figure B.7. SDG 60
2-amino-5-(*p*-tolyl)-3*H*-pyrrolo[2,3-*d*]pyrimidin-4(7*H*)-one

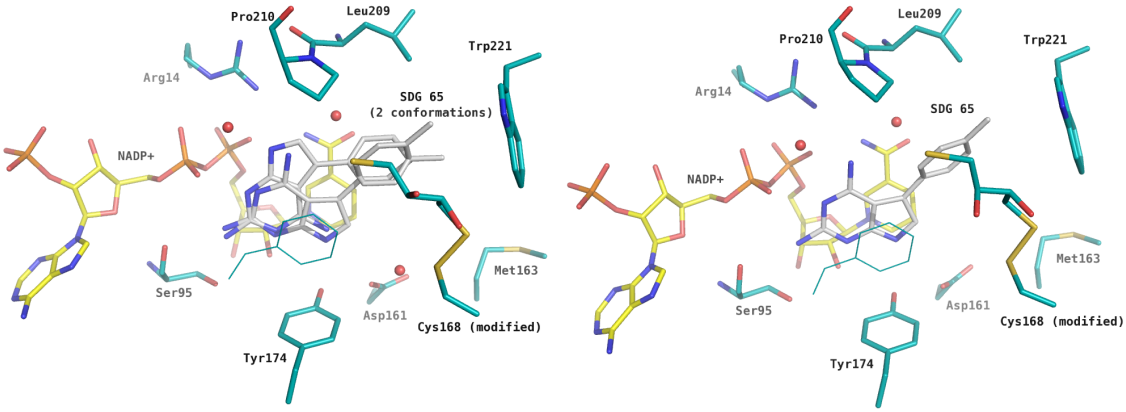


Figure B.8. SDG 65
Subunit A, left; subunit B, right.
5-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

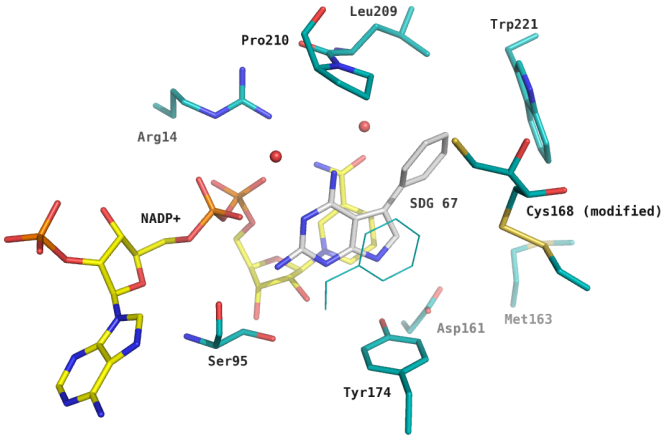


Figure B.9. SDG 67
5-phenyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

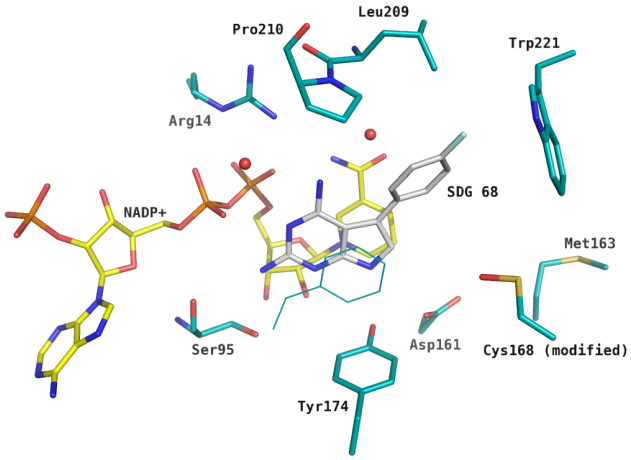


Figure B.10. SDG 68
5-(4-fluorophenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

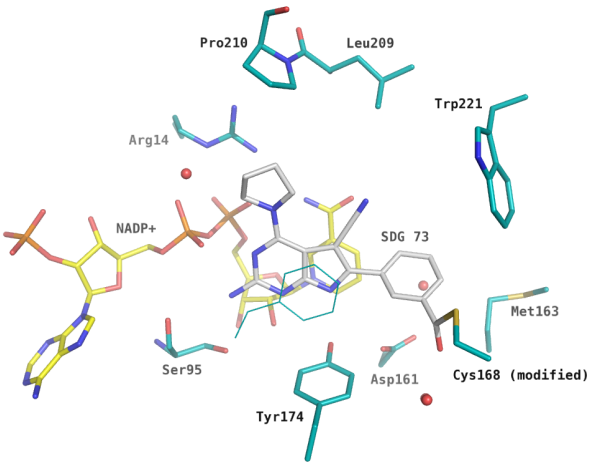


Figure B.11. SDG 73
2-amino-6-(3-formylphenyl)-4-(pyrrolidin-1-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile

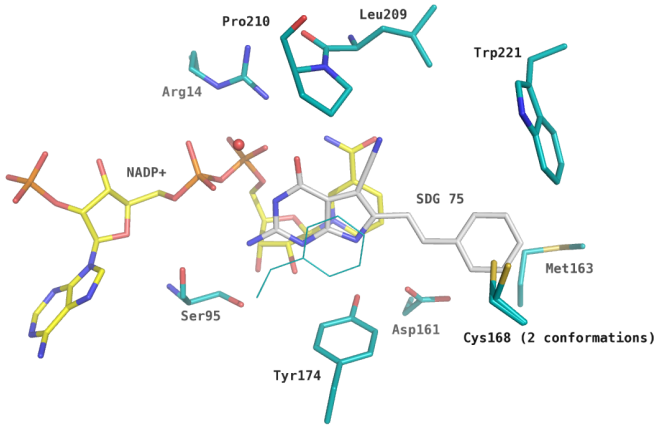


Figure B.12. SDG 75
(*E*)-2-amino-4-oxo-6-styryl-4,7-dihydro-3*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile

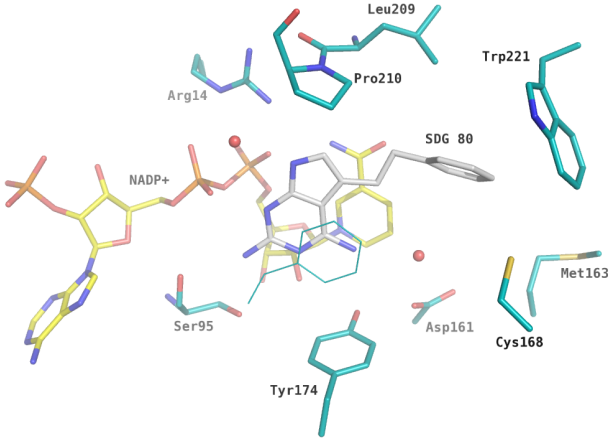


Figure B.13. SDG 80
5-phenethyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

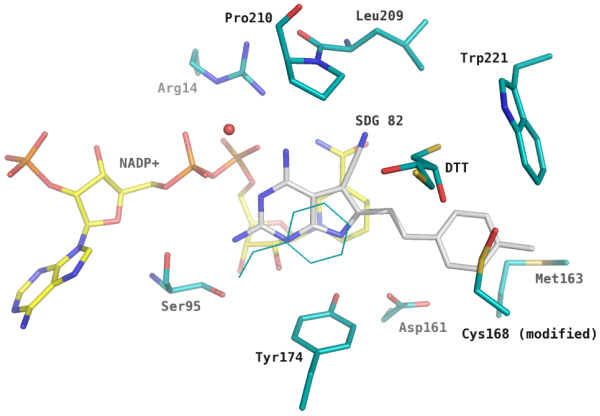


Figure B.14. SDG 82
(*E*)-2,4-diamino-6-(4-methylstyryl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile

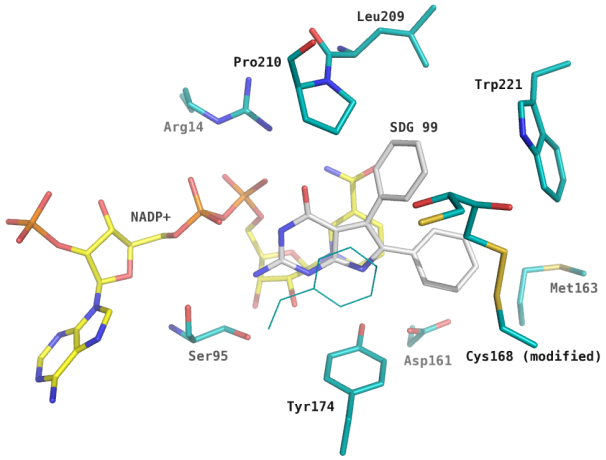


Figure B.15. SDG 99
2-amino-5,6-diphenyl-3*H*-pyrrolo[2,3-*d*]pyrimidin-4(7*H*)-one

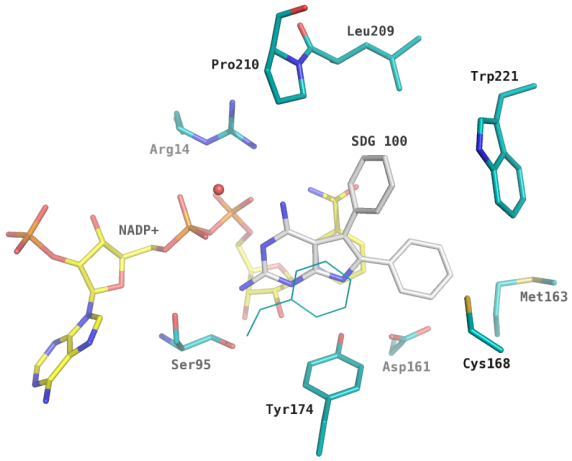


Figure B.16. SDG 100
5,6-diphenyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

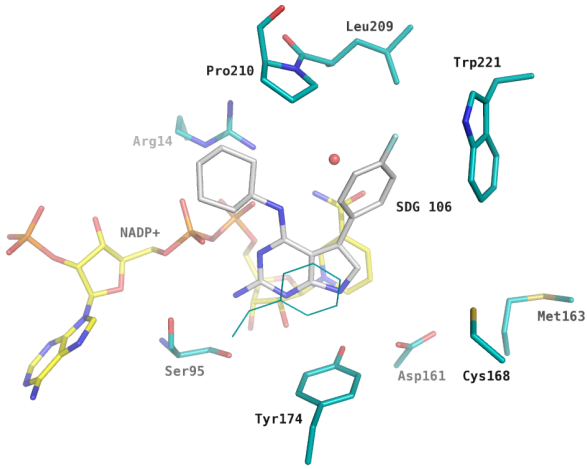


Figure B.17. SDG 106
N^4-cyclohexyl-5-(4-fluorophenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

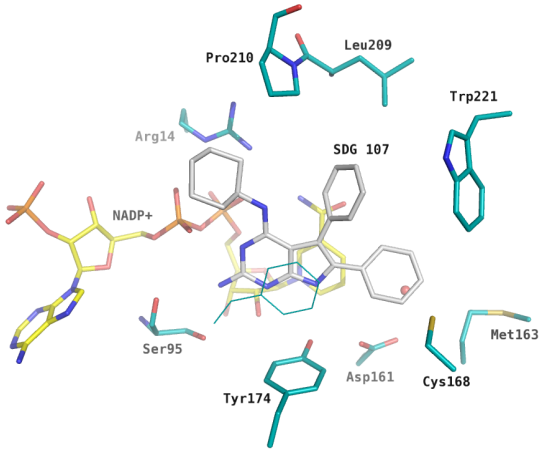


Figure B.18. SDG 107
N^4-cyclohexyl-5,6-diphenyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

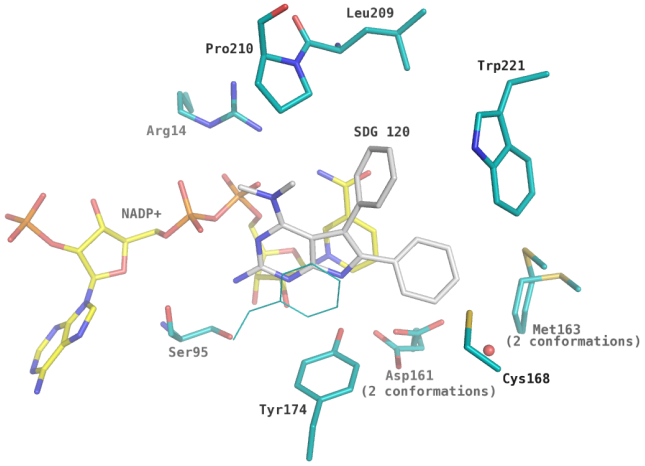


Figure B.19. SDG 120
N^4,N^4-dimethyl-5,6-diphenyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

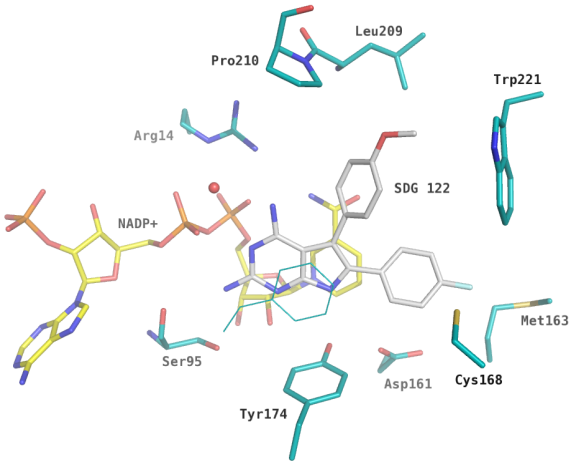


Figure B.20. SDG 122
6-(4-fluorophenyl)-5-(4-methoxyphenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

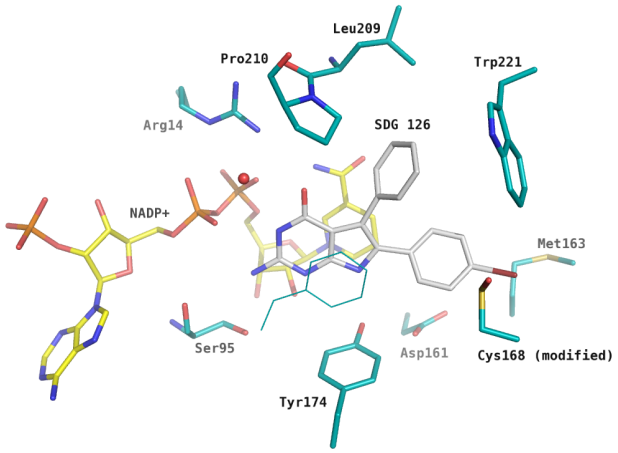


Figure B.21. SDG 126
2-amino-6-(4-bromophenyl)-5-phenyl-3*H*-pyrrolo[2,3-*d*]pyrimidin-4(7*H*)-one

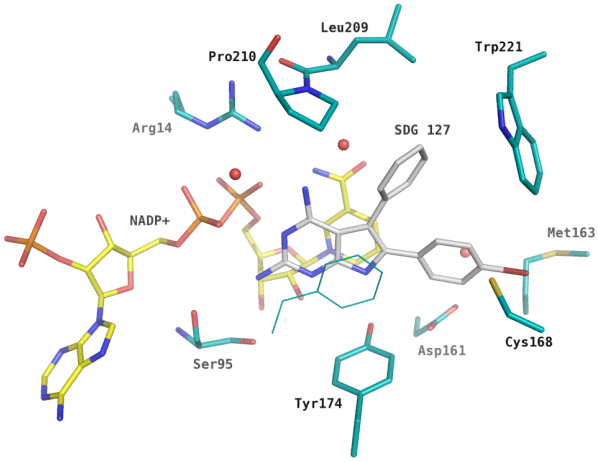


Figure B.22. SDG 127
6-(4-bromophenyl)-5-phenyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-2,4-diamine

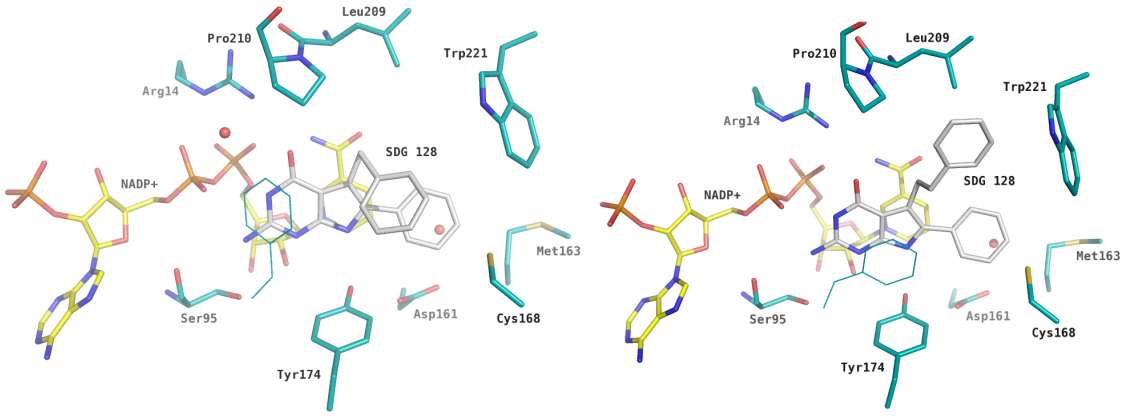


Figure B.23. SDG 128
2-amino-5-phenethyl-6-phenyl-3*H*-pyrrolo[2,3-*d*]pyrimidin-4(7*H*)-one

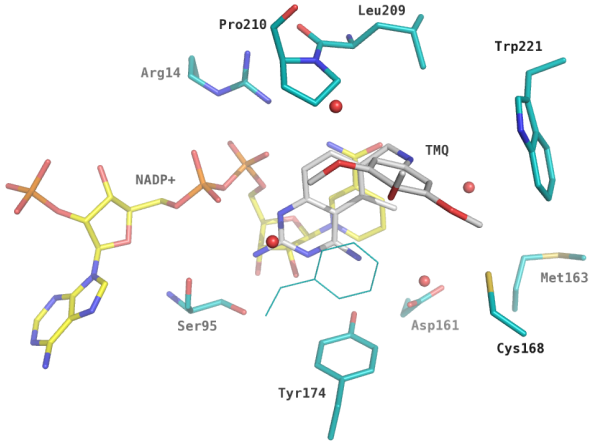


Figure B.24. TMQ
5-methyl-6-[(3,4,5-trimethoxyphenyl)aminomethyl]quinazoline-2,4-diamine

REFERENCES

Abad-Zapatero, C., 2007. Ligand efficiency indices for effective drug discovery. *Expert Opin. Drug Discov*. **2**, 469–488.

Abruzzi, K.C., Smith, A., Chen, W., Solomon, F., 2002. Protection from free β-tubulin by the β-tubulin binding protein Rbl2p. *Mol Cell Biol* **22**, 138–147.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.

Al-Bassam, J., 2002. MAP2 and tau bind longitudinally along the outer ridges of microtubule protofilaments. *J Cell Biol* **157**, 1187–1196.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., Horn, D., 2011. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Research* **21**, 915–924.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Archer, J.E., Magendantz, M., Vega, L.R., Solomon, F., 1998. Formation and function of the Rbl2p-β-tubulin complex. *Mol Cell Biol* **18**, 1757–1762.

Archer, J.E., Vega, L.R., Solomon, F., 1995. Rbl2p, a yeast protein that binds to β-tubulin and participates in microtubule function in vivo. *Cell* **82**, 425–434.

Armstrong, D., Zidovetzki, R., 2009. Helical Wheel Projections. University of California, Riverside. Available: http://rzlab.ucr.edu/scripts/wheel/wheel.cgi.

Asherie, N., 2004. Protein crystallization and phase diagrams. *Methods* **34**, 266–272.

Auffinger, P., Hays, F.A., Westhof, E., Ho, P.S., 2004. Halogen bonds in biological molecules. *Proc Natl Acad Sci USA* **101**, 16789–16794.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**, 10037–10041.

Bartolini, F., Bhamidipati, A., Thomas, S., Schwahn, U., Lewis, S., Cowan, N., 2002. Functional overlap between retinitis pigmentosa 2 protein and the tubulin-specific chaperone cofactor C. *Journal of Biological Chemistry* **277**, 14629–14634.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271–281.

Bello, A.R., Nare, B., Freedman, D., Hardy, L., Beverley, S.M., 1994. PTR1: a reductase mediating salvage of oxidized pteridines and methotrexate resistance in the protozoan parasite *Leishmania major*. *Proc Natl Acad Sci USA* **91**, 11442–11446.

Bennett, V., Baines, A.J., 2001. Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol. Rev*. **81**, 1353–1392.

Bergfors, T., 2003. Seeds to crystals. *J Struct Biol* **142**, 66–76.

Berman, J.D., 1997. Human leishmaniasis: clinical, diagnostic, and chemotherapeutic developments in the last 10 years. *Clin. Infect. Dis*. **24**, 684–703.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M., 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of molecular biology* **112**, 535–542.

Bertani, G., 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol*. **62**, 293–300.

Bhamidipati, A., Lewis, S., Cowan, N., 2000. ADP ribosylation factor-like protein 2 (Arl2) regulates the interaction of tubulin-folding cofactor D with native tubulin. *Journal of Cell Biology* **149**, 1087.

Biswas, T., Tsodikov, O.V., 2010. An easy-to-use tool for planning and modeling a calorimetric titration. *Anal Biochem* **406**, 91–93.

Blakley, R.L., 1995. Eukaryotic dihydrofolate reductase. *Adv. Enzymol. Relat. Areas Mol. Biol*. **70**, 23–102.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., Smithies, O., 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196**, 161–169.

Blow, D.M., Chayen, N.E., Lloyd, L.F., Saridakis, E., 1994. Control of nucleation of protein crystals. *Protein Sci* **3**, 1638–1643.

Bond, C.S., Schüttelkopf, A.W., 2009. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr* **65**, 510–512.

Bonnet, C., Boucher, D., Lazereg, S., Pedrotti, B., Islam, K., Denoulet, P., Larcher, J.C., 2001. Differential binding regulation of microtubule-associated proteins MAP1A, MAP1B, and MAP2 by tubulin polyglutamylation. *Journal of Biological Chemistry* **276**, 12839–12848.

Bourne, H.R., 1997. G proteins. The arginine finger strikes again. *Nature* **389**, 673–674.

Bracher, A., Weissenhorn, W., 2004. Crystal structure of the Habc domain of neuronal syntaxin from the squid Loligo pealei reveals conformational plasticity at its C-terminus. *BMC Struct. Biol*. **4**, 6-14.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.

Bragg, W.H., Bragg, W.L., 1913. The reflection of X-rays by crystals. *Proceedings of the Royal Society of London A* **88**, 428-438.

Bragg, W.L., 1913. The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society* **17**, 43–57.

Broecker, J., Vargas, C., Keller, S., 2011. Revisiting the optimal *c* value for isothermal titration calorimetry. *Anal Biochem* **418**, 307–309.

Bronowska, A.K., 2011. Thermodynamics of ligand-protein interactions: implications for molecular design, in: Moreno-Pirajan, J.C. (Ed.), Thermodynamics - interaction studies - solids, liquids and gases. *InTech*, 1–48.

Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.

Buchan, D.W.A., Ward, S.M., Lobley, A.E., Nugent, T.C.O., Bryson, K., Jones, D.T., 2010. Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38**, W563–W568.

Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Moustiakimov, M., Polidori, G., Siliqi, D., 2004. MAD phasing: choosing the most informative wavelength combination. *Acta Crystallogr D Biol Crystallogr* **60**, 1683–1686.

Burmeister, W.P., 2000. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr D Biol Crystallogr* **56**, 328–341.

Bush, A.O., Fernández, J.C., Esch, G.W., Seed, J.R., 2001. Parasitism: The diversity and ecology of animal parasites, Cambridge: Cambridge University Press.

Carr, R.A.E., Congreve, M., Murray, C.W., Rees, D.C., 2005. Fragment-based lead discovery: leads by design. *Drug discovery today* **10**, 987–992.

Cavazzuti, A., Paglietti, G., Hunter, W.N., Gamarro, F., Piras, S., Loriga, M., Allecca, S., Corona, P., McLuskey, K., Tulloch, L., Gibellini, F., Ferrari, S., Costi, M.P., 2008. Discovery of potent pteridine reductase inhibitors to guide antiparasite drug development. *Proc Natl Acad Sci USA* **105**, 1448–1453.

Centers for Disease Control and Prevention, 2012. Parasites - African Trypanosomiasis. [Updated August 2012] Available: http://www.cdc.gov/parasites/sleepingsickness/biology.html

Centers for Disease Control and Prevention, 2013. Parasites – Leishmaniasis. [Updated January 2013] Available: http://www.cdc.gov/parasites/leishmaniasis/biology.html

Chaires, J.B., 2008. Calorimetry and thermodynamics in drug design. *Annu. Rev. Biophys*. **37**, 135–151.

Chayen, N., 1997a. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *Journal of Applied Crystallography* **30**, 198–202.

Chayen, N., 1997b. The role of oil in macromolecular crystallization. *Structure* **5**, 1269–1274.

Chayen, N.E., 1999. Crystallization with oils: a new dimension in macromolecular crystal growth. *Journal of Crystal Growth* **196**, 434–441.

Chayen, N.E., 2009. Rigorous filtration for protein crystallization. *J. Appl. Cryst* **42**, 743-744.

Chayen, N.E., Saridakis, E., El-Bahar, R., Nemirovsky, Y., 2001. Porous silicon: an effective nucleation-inducing material for protein crystallization. *Journal of molecular biology* **312**, 591–595.

Chayen, N.E., Shaw Stewart, P.D., Maeder, D.L., Blow, D.M., 1990. An automated system for micro-batch protein crystallization and screening. *J Appl Cryst* **23**, 297-302.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21.

Cheng, Y., Prusoff, W.H., 1973. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol*. **22**, 3099–3108.

Collaborative Computational Project, N.4., 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763.

Congreve, M., Murray, C.W., Blundell, T.L., 2005. Structural biology and drug discovery. *Drug discovery today* **10**, 895–907.

Cooper, A., Johnson, C.M., Lakey, J.H., Nöllmann, M., 2001. Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem*. **93**, 215–230.

Croft, S.L., Sundar, S., Fairlamb, A.H., 2006. Drug resistance in leishmaniasis. *Clin. Microbiol. Rev*. **19**, 111–126.

Cross, D., Dominguez, J., Maccioni, R.B., Avila, J., 1991. MAP-1 and MAP-2 binding sites at the C-terminus of β-tubulin. Studies with synthetic tubulin peptides. *Biochemistry* **30**, 4362–4366.

Cruickshank, D.W., 1999. Remarks about protein structure precision. *Acta Crystallogr D Biol Crystallogr* **55**, 583–601.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., McPherson, A., 1994. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr* **50**, 414–423.

Cunningham, M.L., Titus, R.G., Turco, S.J., Beverley, S.M., 2001. Regulation of differentiation to the infective stage of the protozoan parasite *Leishmania major* by tetrahydrobiopterin. *Science* **292**, 285–287.

D'Arcy, A., Mac Sweeney, A., Haber, A., 2003a. Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta Crystallogr D Biol Crystallogr* **59**, 1343–1346.

D'Arcy, A., Mac Sweeney, A., Haber, A., 2004. Modified microbatch and seeding in protein crystallization experiments. *J Synchrotron Radiat* **11**, 24–26.

D'Arcy, A., Mac Sweeney, A., Stihle, M., Haber, A., 2003b. The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Acta Crystallogr D Biol Crystallogr* **59**, 396–399.

Davis, A., Teague, S., Kleywegt, G., 2003. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed* **42**, 2718–2736.

Davis, L., Abdi, K., Machius, M., Brautigam, C., Tomchick, D.R., Bennett, V., Michaely, P., 2009. Localization and structure of the ankyrin-binding site on beta2-spectrin. *Journal of Biological Chemistry* **284**, 6982–6987.

Dawson, A., Gibellini, F., Sienkiewicz, N., Tulloch, L.B., Fyfe, P.K., McLuskey, K., Fairlamb, A.H., Hunter, W.N., 2006. Structure and reactivity of *Trypanosoma brucei* pteridine reductase: inhibition by the archetypal antifolate methotrexate. *Mol Microbiol* **61**, 1457–1468.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System. http://www.pymol.org.

Desai, A., Mitchison, T.J., 1997. Microtubule polymerization dynamics. *Annu. Rev. Cell Dev. Biol*. **13**, 83–117.

Desjeux, P., 1996. Leishmaniasis. Public health aspects and control. *Clin. Dermatol*. **14**, 417–423.

Desjeux, P., 2001. The increase in risk factors for leishmaniasis worldwide. *Trans R Soc Trop Med Hyg* **95**, 239–243.

Desjeux, P., 2004. Leishmaniasis: current situation and new perspectives. *Comp. Immunol. Microbiol. Infect. Dis*. **27**, 305–318.

Dias, J.C.P., Silveira, A.C., Schofield, C.J., 2002. The impact of Chagas disease control in Latin America: a review. *Mem. Inst. Oswaldo Cruz* **97**, 603–612.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665–7.

Doutch, J., Hough, M.A., Hasnain, S.S., Strange, R.W., 2012. Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. *J. Synchrotron Rad* **19**, 19-29.

Drechsel, D.N., Kirschner, M.W., 1994. The minimum GTP cap required to stabilize microtubules. *Curr Biol* **4**, 1053–1061.

Duax, W.L., Pletnev, V., Addlagatta, A., Bruenn, J., Weeks, C.M., 2003. Rational proteomics I. Fingerprint identification and cofactor specificity in the short-chain oxidoreductase (SCOR) enzyme family. *Proteins* **53**, 931–943.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., Overath, P., 2007. Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131**, 505–515.

Erlanson, D.A., Wells, J.A., Braisted, A.C., 2004. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* **33**, 199–223.

Ertl, P., 2010. Molecular structure input on the web. *J Cheminform* **2**, 1.

Evans, G., Pettifer, R.F., 2001. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of Applied Crystallography* **34**, 82–86.

Evans, P., 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72–82.

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214.

Fairlamb, A.H., 2003. Chemotherapy of human African trypanosomiasis: current and future prospects. *Trends Parasitol* **19**, 488–494.

Fanarraga, M.L., Párraga, M., Aloria, K., del Mazo, J., Avila, J., Zabala, J.C., 1999. Regulated expression of p14 (cofactor A) during spermatogenesis. *Cell Motil Cytoskeleton* **43**, 243–254.

Fennell, B.J., Naughton, J.A., Dempsey, E., Bell, A., 2006. Cellular and molecular actions of dinitroaniline and phosphorothioamidate herbicides on *Plasmodium falciparum*: tubulin as a specific antimalarial target. *Mol Biochem Parasitol* **145**, 226–238.

Fenton, W.A., Kashi, Y., Furtak, K., Horwich, A.L., 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**, 614–619.

Ferenczy, G.G., Keserű, G.M., 2012. Thermodynamics of fragment binding. *J Chem Inf Model* **52**, 1039–1045.

Fleming, J.R., Morgan, R.E., Fyfe, P.K., Kelly, S.M., Hunter, W.N., 2013. The architecture of *Trypanosoma brucei* tubulin-binding cofactor B and implications for function. *FEBS J* **280**, 3270–3280.

Flynn, G.C., Pohl, J., Flocco, M.T., Rothman, J.E., 1991. Peptide-binding specificity of the molecular chaperone BiP. *Nature* **353**, 726–730.

Fontalba, A., Paciucci, R., Avila, J., Zabala, J.C., 1993. Incorporation of tubulin subunits into dimers requires GTP hydrolysis. *J Cell Sci* **106**, 627–632.

Freire, E., Mayorga, O.L., Straume, M., 1990. Isothermal titration calorimetry. *Analytical Chemistry* **62**, 950–959.

Freyer, M.W., Lewis, E.A., 2008. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol* **84**, 79–113.

Frishman, D., Argos, P., 1995. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579.

Fyfe, P.K., 2008. Protein Solubilisation. International Patent WO2008/015419.

Gallo, J.-M., Anderton, B., 1983. A subpopulation of trypanosome microtubules recognised by a monoclonal antibody to tubulin. *EMBO J* **2**, 479–483.

Gallo, J.M., Precigout, E., 1988. Tubulin expression in trypanosomes. *Biol. Cell* **64**, 137–143.

Gao, Y., Melki, R., Walden, P.D., Lewis, S.A., Ampe, C., Rommelaere, H., Vandekerckhove, J., Cowan, N.J., 1994. A novel cochaperonin that modulates the ATPase activity of cytoplasmic chaperonin. *J Cell Biol* **125**, 989–996.

Gao, Y., Thomas, J.O., Chow, R.L., Lee, G.H., Cowan, N.J., 1992. A cytoplasmic chaperonin that catalyzes β-actin folding. *Cell* **69**, 1043–1050.

Gao, Y., Vainberg, I.E., Chow, R.L., Cowan, N.J., 1993. Two cofactors and cytoplasmic chaperonin are required for the folding of α- and β-tubulin. *Mol Cell Biol* **13**, 2478–2485.

Garcia-Mayoral, M.F., Castaño, R., Fanarraga, M.L., Zabala, J.C., Rico, M., Bruix, M., 2011. The solution structure of the N-Tterminal domain of human tubulin binding cofactor C reveals a platform for tubulin interaction. *PLoS ONE* **6**, e25912.

Garman, E., 2003. “Cool” crystals: macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol* **13**, 545–551.

Garman, E.F., 2010. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr* **66**, 339–351.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook* 571-607.

Gibson, W., 2012. The origins of the trypanosome genome strains *Trypanosoma brucei brucei* TREU 927, *T. b. gambiense* DAL 972, *T. vivax* Y486 and *T. congolense* IL3000. *Parasit Vectors* **5**, 71.

Giles, N.L., Armson, A., Reid, S.A., 2009. Characterization of trifluralin binding with recombinant tubulin from *Trypanosoma brucei*. *Parasitol. Res.* **104**, 893–903.

Gourley, D.G., Schüttelkopf, A.W., Leonard, G.A., Luba, J., Hardy, L.W., Beverley, S.M., Hunter, W.N., 2001. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. *Nat Struct Biol* **8**, 521–525.

Grune, T., 2008. *mtz2sca* and *mtz2hkl*: facilitated transition from *CCP4* to the *SHELX* program suite. *J. Appl. Cryst*. **41**, 217-218.

Grynberg, M., Jaroszewski, L., Godzik, A., 2003. Domain analysis of the tubulin cofactor system: a model for tubulin folding and dimerization. *BMC bioinformatics* **4**, 46.

Guasch, A., Aloria, K., Pérez, R., Avila, J., Zabala, J.C., Coll, M., 2002. Three-dimensional structure of human tubulin chaperone cofactor A. *Journal of molecular biology* **318**, 1139–1149.

Gull, K., 1999. The cytoskeleton of trypanosomatid parasites. *Annu Rev Microbiol* **53**, 629–655.

Hall, D.R., Gourley, D.G., Leonard, G.A., Duke, E.M., Anderson, L.A., Boxer, D.H., Hunter, W.N., 1999. The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli*: a novel combination of domain folds. *EMBO J* **18**, 1435–1446.

Hardy, L., Matthews, W., Nare, B., Beverley, S., 1997. Biochemical and Genetic Tests for Inhibitors of *Leishmania* Pteridine Pathways. *Exp Parasitol* **87**, 158–170.

Hauptman, H.A., 1991. The phase problem of x-ray crystallography. *Rep. Prog. Phys.* **54**, 1427-1454.

Heinig, M., Frishman, D., 2004. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* **32**, W500–2.

Hendrickson, W.A., Horton, J.R., LeMaster, D.M., 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**, 1665–1672.

Hendrickson, W.A., Smith, J.L., Sheriff, S., 1985. Direct phase determination based on anomalous scattering. *Methods in Enzymology* **115**, 41–55.

Hirata, D., Masuda, H., Eddison, M., Toda, T., 1998. Essential role of tubulin-folding cofactor D in microtubule assembly and its association with microtubules in fission yeast. *EMBO J* **17**, 658–666.

Holdgate, G.A., 2001. Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *BioTechniques* **31**, 164–84.

Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–9.

Hopkins, A.L., Groom, C.R., Alex, A., 2004. Ligand efficiency: a useful metric for lead selection. *Drug discovery today* **9**, 430–431.

Howard, J., Hyman, A.A., 2003. Dynamics and mechanics of the microtubule plus end. *Nature* **422**, 753–758.

Howard, J., Hyman, A.A., 2009. Growth, fluctuation and switching at microtubule plus ends. *Nat Rev Mol Cell Biol* **10**, 569–574.

Hughes, J.D., Blagg, J., Price, D.A., Bailey, S., Decrescenzo, G.A., Devraj, R.V., Ellsworth, E., Fobian, Y.M., Gibbs, M.E., Gilles, R.W., Greene, N., Huang, E., Krieger-Burke, T., Loesel, J., Wager, T., Whiteley, L., Zhang, Y., 2008. Physiochemical drug properties associated with *in vivo* toxicological outcomes. *Bioorg. Med. Chem. Lett*. **18**, 4872–4875.

Hunter, W.N., 2009. Structure-based ligand design and the promise held for antiprotozoan drug discovery. *Journal of Biological Chemistry* **284**, 11749–11753.

Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S., Coleman, R.G., 2012. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* **52**, 1757–1768.

Janke, C., Kneussel, M., 2010. Tubulin post-translational modifications: encoding functions on the neuronal microtubule cytoskeleton. *Trends in Neurosciences* **33**, 362–372.

Johnson, K.A., Goody, R.S., 2011. The original Michaelis constant: translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264–8269.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* **292**, 195–202.

Kabsch, W., 2010. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.

Kapust, R.B., Tözsér, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., Waugh, D.S., 2001. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng*. **14**, 993–1000.

Kelley, L.A., Sternberg, M.J.E., 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.

Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta* **1751**, 119–139.

Kibbe, W.A., 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43–6.

Kida, H., Sugano, Y., Iizuka, R., Fujihashi, M., Yohda, M., Miki, K., 2008. Structural and molecular characterization of the prefoldin β subunit from Thermococcus strain KS-1. *Journal of molecular biology* **383**, 465–474.

Kim, Y., Quartey, P., Li, H., Volkart, L., Hatzos, C., Chang, C., Nocek, B., Cuff, M., Osipiuk, J., Tan, K., Fan, Y., Bigelow, L., Maltseva, N., Wu, R., Borovilos, M., Duggan, E., Zhou, M., Binkowski, T.A., Zhang, R.-G., Joachimiak, A., 2008. Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat Methods* **5**, 853–854.

Kirik, V., Mathur, J., Grini, P.E., Klinkhammer, I., Adler, K., Bechtold, N., Herzog, M., Bonneville, J.-M., Hülskamp, M., 2002. Functional analysis of the tubulin-folding cofactor C in *Arabidopsis thaliana*. *Current Biology* **12**, 1519–1523.

Kolář, M., Hobza, P., Bronowska, A.K., 2013. Plugging the explicit σ-holes in molecular docking. *Chem. Commun*. **49**, 981.

Kompis, I.M., Islam, K., Then, R.L., 2005. DNA and RNA synthesis: antifolates. *Chem. Rev*. **105**, 593–620.

Kortazar, D., Carranza, G., Bellido, J., Villegas, J.C., Fanarraga, M.L., Zabala, J.C., 2006. Native tubulin-folding cofactor E purified from baculovirus-infected Sf9 cells dissociates tubulin dimers. *Protein Expr Purif* **49**, 196–202.

Kortazar, D., Fanarraga, M.L., Carranza, G., Bellido, J., Villegas, J.C., Avila, J., Zabala, J.C., 2007. Role of cofactors B (TBCB) and E (TBCE) in tubulin heterodimer dissociation. *Exp Cell Res* **313**, 425–436.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256–2268.

Kuntz, I.D., Chen, K., Sharp, K.A., Kollman, P.A., 1999. The maximal affinity of ligands. *Proc Natl Acad Sci USA* **96**, 9997–10002.

Kusunoki, H., Minasov, G., Macdonald, R.I., Mondragón, A., 2004. Independent movement, dimerization and stability of tandem repeats of chicken brain α-spectrin. *Journal of molecular biology* **344**, 495–511.

Kühnel, K., Veltel, S., Schlichting, I., Wittinghofer, A., 2006. Crystal structure of the human retinitis pigmentosa 2 protein and its interaction with Arl3. *Structure* **14**, 367–378.

Ladbury, J.E., 2010. Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochem Soc Trans* **38**, 888.

Ladbury, J.E., Chowdhry, B.Z., 1996. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chemistry & Biology* **3**, 791–801.

Ladbury, J.E., Klebe, G., Freire, E., 2010. Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov* **9**, 23–27.

Leavitt, S., Freire, E., 2001. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* **11**, 560–566.

Lee, J.C., Timasheff, S.N., 1977. In vitro reconstitution of calf brain microtubules: effects of solution variables. *Biochemistry* **16**, 1754–1764.

Lefevre, J., Chernov, K.G., Joshi, V., Delga, S., Toma, F., Pastre, D., Curmi, P.A., Savarin, P., 2011. The C terminus of tubulin, a versatile partner for cationic molecules: binding of TAU, bolyamines, and calcium. *J Biol Chem* **286**, 3065–3078.

Leslie, A.G.W., 2006. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48–57.

Li, L., Dantzer, J.J., Nowacki, J., O'Callaghan, B.J., Meroueh, S.O., 2008. PDBcal: a comprehensive dataset for receptor-ligand interactions with three-dimensional structures and binding thermodynamics from isothermal titration calorimetry. *Chemical Biology & Drug Design* **71**, 529–532.

Linding, R., Russell, R.B., Neduva, V., Gibson, T.J., 2003. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**, 3701–3708.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev*. **46**, 3–26.

Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F., Hendrickson, W.A., 2012. Structures from anomalous diffraction of native biological macromolecules. *Science* **336**, 1033–1037.

Llosa, M., Aloria, K., Campo, R., Padilla, R., Avila, J., Sánchez-Pulido, L., Zabala, J.C., 1996. The β-tubulin monomer release factor (p14) has homology with a region of the DnaJ protein. *FEBS Lett* **397**, 283–289.

Lobley, A., Whitmore, L., Wallace, B.A., 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **18**, 211–212.

Logan-Klumpler, F.J., De Silva, N., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J.A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., Phan, I., Farris, C., Mitra, S., Ramasamy, G., Wang, H., Tivey, A., Jackson, A., Houston, R., Parkhill, J., Holden, M., Harb, O.S., Brunk, B.P., Myler, P.J., Roos, D., Carrington, M., Smith, D.F., Hertz-Fowler, C., Berriman, M., 2012. GeneDB--an annotation database for pathogens. *Nucleic Acids Res* **40**, D98–D108.

Lopez-Fanarraga, M., Avila, J., Guasch, A., Coll, M., Zabala, J.C., 2001. Review: postchaperonin tubulin folding cofactors and their role in microtubule dynamics. *J Struct Biol* **135**, 219–229.

Lu, L., Nan, J., Mi, W., Li, L.-F., Wei, C.-H., Su, X.-D., Li, Y., 2010. Crystal structure of tubulin folding cofactor A from *Arabidopsis thaliana* and its β-tubulin binding characterization. *FEBS Lett* **584**, 3533–3539.

Lu, Y., Shi, T., Wang, Y., Yang, H., Yan, X., Luo, X., Jiang, H., Zhu, W., 2009. Halogen Bonding - A Novel Interaction for Rational Drug Design? *J Med Chem* **52**, 2854–2862.

Luba, J., Nare, B., Liang, P.H., Anderson, K.S., Beverley, S.M., Hardy, L.W., 1998. *Leishmania major* pteridine reductase 1 belongs to the short chain dehydrogenase family: stereochemical and kinetic evidence. *Biochemistry* **37**, 4093–4104.

Lubega, G.W., Ochola, D.O.K., Prichard, R.K., 2002. *Trypanosoma brucei*: anti-tubulin antibodies specifically inhibit trypanosome growth in culture. *Exp Parasitol* **102**, 134–142.

Luft, J.R., DeTitta, G.T., 1999. A method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* **55**, 988–993.

Lundin, V.F., Leroux, M.R., Stirling, P.C., 2010. Quality control of cytoskeletal proteins and human disease. *Trends Biochem Sci* **35**, 288–297.

Lytle, B.L., Peterson, F.C., Qiu, S.-H., Luo, M., Zhao, Q., Markley, J.L., Volkman, B.F., 2004. Solution structure of a ubiquitin-like domain from tubulin-binding cofactor B. *J Biol Chem* **279**, 46787–46793.

MacDonald, L.M., Armson, A., Thompson, A.R.C., Reynoldson, J.A., 2004. Characterisation of benzimidazole binding with recombinant tubulin from *Giardia duodenalis*, *Encephalitozoon intestinalis*, and *Cryptosporidium parvum*. *Mol Biochem Parasitol* **138**, 89–96.

MacDonald, L.M., Armson, A., Thompson, R.C., Reynoldson, J.A., 2001. Expression of *Giardia duodenalis* β-tubulin as a soluble protein in *Escherichia coli*. *Protein Expr Purif* **22**, 25–30.

MacDonald, L.M., Armson, A., Thompson, R.C.A., Reynoldson, J.A., 2003. Characterization of factors favoring the expression of soluble protozoan tubulin proteins in *Escherichia coli*. *Protein Expr Purif* **29**, 117–122.

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Margolis, R.L., Wilson, L., 1998. Microtubule treadmilling: what goes around comes around. *Bioessays* **20**, 830–836.

Martín, L., Fanarraga, M.L., Aloria, K., Zabala, J.C., 2000. Tubulin folding cofactor D is a microtubule destabilizing protein. *FEBS Lett* **470**, 93–95.

Matthews, B.W., 1968. Solvent content of protein crystals. *Journal of molecular biology* **33**, 491–497.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658–674.

McDonald, I.K., Thornton, J.M., 1994. Satisfying hydrogen bonding potential in proteins. *Journal of molecular biology* **238**, 777–793.

McKean, P.G., 2003. Coordination of cell cycle and cytokinesis in *Trypanosoma brucei*. *Curr. Opin. Microbiol*. **6**, 600–607.

McLuskey, K., Gibellini, F., Carvalho, P., Avery, M.A., Hunter, W.N., 2004. Inhibition of *Leishmania major* pteridine reductase by 2,4,6-triaminoquinazoline: structure of the NADPH ternary complex. *Acta Crystallogr D Biol Crystallogr* **60**, 1780–1785.

Melki, R., Rommelaere, H., Leguy, R., Vandekerckhove, J., Ampe, C., 1996. Cofactor A is a molecular chaperone required for β-tubulin folding: functional and structural characterization. *Biochemistry* **35**, 10422–10435.

Melki, R., Vainberg, I.E., Chow, R.L., Cowan, N.J., 1993. Chaperonin-mediated folding of vertebrate actin-related protein and γ-tubulin. *J Cell Biol* **122**, 1301–1310.

Merritt, E.A., 2012. X-ray anomalous scattering. [Updated 2012] Available: http://skuld.bmsc.washington.edu/scatter/

Morgan, R.E., Ahn, S., Nzimiro, S., Fotie, J., Phelps, M.A., Cotrill, J., Yakovich, A.J., Sackett, D.L., Dalton, J.T., Werbovetz, K.A., 2008. Inhibitors of tubulin assembly identified through screening a compound library. *Chemical Biology & Drug Design* **72**, 513–524.

Moriarty, N.W., Grosse-Kunstleve, R.W., Adams, P.D., 2009. Electronic ligand Builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074–1080.

Mpamhanga, C.P., Spinks, D., Tulloch, L.B., Shanks, E.J., Robinson, D.A., Collie, I.T., Fairlamb, A.H., Wyatt, P.G., Frearson, J.A., Hunter, W.N., Gilbert, I.H., Brenk, R., 2009. One scaffold, three binding modes: novel and selective pteridine reductase 1 inhibitors derived from fragment hits discovered by virtual screening. *J Med Chem* **52**, 4454–4465.

Murray, H.W., Berman, J.D., Davies, C.R., Saravia, N.G., 2005. Advances in leishmaniasis. *Lancet* **366**, 1561–1577.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355–367.

Nare, B., Hardy, L.W., Beverley, S.M., 1997a. The roles of pteridine reductase 1 and dihydrofolate reductase-thymidylate synthase in pteridine metabolism in the protozoan parasite *Leishmania major*. *J Biol Chem* **272**, 13883–13891.

Nare, B., Luba, J., Hardy, L.W., Beverley, S., 1997b. New approaches to *Leishmania* chemotherapy: pteridine reductase 1 (PTR1) as a target and modulator of antifolate sensitivity. *Parasitology* **114** S101–10.

Niesen, F.H., Berglund, H., Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212–2221.

Nogales, E., Wolf, S.G., Downing, K.H., 1998. Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**, 199–203.

Novoradovsky, A., Zhang, V., Ghosh, M., Hogrefe, H., Sorge, J.A., Gaasterland, T., 2005. Computational principles of primer design for site directed mutagenesis. *Technical Proceedings of 2005 NSTI Nanotechnology Conference and Trade Show* **1**, 532–535.

Nunes, M.C.P., Dones, W., Morillo, C.A., Encina, J.J., Ribeiro, A.L., Council on Chagas Disease of the Interamerican Society of Cardiology, 2013. Chagas disease: an overview of clinical and epidemiological aspects. *J. Am. Coll. Cardiol*. **62**, 767–776.

Ohtaka, H., Muzammil, S., Schön, A., Velazquez-Campoy, A., Vega, S., Freire, E., 2004. Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets. *Int. J. Biochem. Cell Biol*. **36**, 1787–1799.

Ohtaki, A., Kida, H., Miyata, Y., Ide, N., Yonezawa, A., Arakawa, T., Iizuka, R., Noguchi, K., Kita, A., Odaka, M., Miki, K., Yohda, M., 2008. Structure and molecular dynamics simulation of archaeal prefoldin: the molecular mechanism for binding and recognition of nonnative substrate proteins. *Journal of molecular biology* **376**, 1130–1141.

Olsson, T.S.G., Williams, M.A., Pitt, W.R., Ladbury, J.E., 2008. The thermodynamics of protein–ligand interaction and solvation: insights for ligand design. *Journal of molecular biology* **384**, 1002–1017.

Ong, H.B., Sienkiewicz, N., Wyllie, S., Fairlamb, A.H., 2011. Dissecting the metabolic roles of pteridine reductase 1 in *Trypanosoma brucei* and *Leishmania major*. *J Biol Chem* **286**, 10429–10438.

Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 1–20.

Oxberry, M.E., Geary, T.G., Winterrowd, C.A., Prichard, R.K., 2001. Individual expression of recombinant α- and β-tubulin from *Haemonchus contortus*: polymerization and drug effects. *Protein Expr Purif* **21**, 30–39.

Pape, T., Schneider, T.R., 2004. *HKL2MAP*: a graphical user interface for macromolecular phasing with *SHELX* programs. *J. Appl. Cryst*. **37**, 843-844.

Patterson, A.L., 1934. A Fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review* **46**, 372-376.

Patterson, A.L., 1935. A direct method for the determination of the components of interatomic distances in crystals. *Z. Kristallogr*. **90**, 517–542.

Perutz, M.F., 1956. Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Cryst.* **9**, 867-873.

Pierce, M.M., Raman, C.S., Nall, B.T., 1999. Isothermal titration calorimetry of protein-protein interactions. *Methods* **19**, 213–221.

Prata, A., 2001. Clinical and epidemiological aspects of Chagas disease. *Lancet Infect Dis* **1**, 92–100.

Prota, A.E., Magiera, M.M., Kuijpers, M., Bargsten, K., Frey, D., Wieser, M., Jaussi, R., Hoogenraad, C.C., Kammerer, R.A., Janke, C., Steinmetz, M.O., 2013. Structural basis of tubulin tyrosination by tubulin tyrosine ligase. *J Cell Biol* **200**, 259–270.

Provencher, S.W., Glöckner, J., 1981. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37.

Ralston, K.S., Hill, K.L., 2008. The flagellum of *Trypanosoma brucei*: new tricks from an old dog. *Int. J. Parasitol*. **38**, 869–884.

Ralston, K.S., Kabututu, Z.P., Melehani, J.H., Oberholzer, M., Hill, K.L., 2009. The *Trypanosoma brucei* flagellum: moving parasites in new directions. *Annu Rev Microbiol* **63**, 335–362.

Rao, S.T., Rossmann, M.G., 1973. Comparison of super-secondary structures in proteins. *Journal of molecular biology* **76**, 241–256.

Ravelli, R.B.G., Leiros, H.-K.S., Pan, B., Caffrey, M., McSweeney, S., 2003. Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* **11**, 217–224.

Rayment, I., 1997. Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods in Enzymology* **276**, 171–179.

Rees, D.C., Congreve, M., Murray, C.W., Carr, R., 2004. Fragment-based lead discovery. *Nat Rev Drug Discov* **3**, 660–672.

Reithinger, R., Dujardin, J.-C., Louzir, H., Pirmez, C., Alexander, B., Brooker, S., 2007. Cutaneous leishmaniasis. *Lancet Infect Dis* **7**, 581–596.

Richardson, J.S., 1981. The anatomy & taxonomy of protein structure. *Adv Protein Chem* **34**, 167–339.

Robertson, J.G., 2005. Mechanistic basis of enzyme-targeted drugs. *Biochemistry* **44**, 5561–5571.

Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I.M., Potratz, M., Sheldrick, G.M., Usón, I., 2012. Practical structure solution with ARCIMBOLDO. *Acta Crystallogr D Biol Crystallogr* **68**, 336–343.

Rossmann, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst. **15**, 24-31.

Roth, G.J., Stanford, N., Majerus, P.W., 1975. Acetylation of prostaglandin synthase by aspirin. *Proc Natl Acad Sci USA* **72**, 3073–3076.

Rupp, B., 2009. Biomolecular crystallography: principles, practice, and application to structural biology. Garland Science.

Rypniewski, W.R., Holden, H.M., Rayment, I., 1993. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8 Å resolution. *Biochemistry* **32**, 9851–9858.

Sainsbury, S., Bird, L., Rao, V., Shepherd, S.M., Stuart, D.I., Hunter, W.N., Owens, R.J., Ren, J., 2011. Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. *Journal of molecular biology* **405**, 173–184.

Salemme, F.R., Spurlino, J., Bone, R., 1997. Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. *Structure* **5**, 319–324.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: A laboratory manual, 2nd Edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F., Wittinghofer, A., 1997. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338.

Schmahl, W.W., Steurer, W., 2012. Laue centennial. *Acta Crystallogr A Found Crystallogr* **68**, 1–2.

Schüttelkopf, A.W., Hardy, L.W., Beverley, S.M., Hunter, W.N., 2005. Structures of *Leishmania major* pteridine reductase complexes reveal the active site features important for ligand binding and to guide inhibitor design. *Journal of molecular biology* **352**, 105–116.

Schüttelkopf, A.W., van Aalten, D.M.F., 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355–1363.

Seebeck, T., Whittaker, P.A., Imboden, M.A., Hardman, N., Braun, R., 1983. Tubulin genes of *Trypanosoma brucei*: a tightly clustered family of alternating genes. *Proc Natl Acad Sci USA* **80**, 4634–4638.

Shelanski, M.L., Gaskin, F., Cantor, C.R., 1973. Microtubule assembly in the absence of added nucleotides. *Proc Natl Acad Sci USA* **70**, 765–768.

Sheldrick, G.M., 2008. A short history of SHELX. *Acta Crystallogr A Found Crystallogr* **64**, 112–122.

Sheldrick, G.M., 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479–485.

Siegert, R., Leroux, M.R., Scheufler, C., Hartl, F.U., Moarefi, I., 2000. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* **103**, 621–632.

Sienkiewicz, N., Ong, H.B., Fairlamb, A.H., 2010. *Trypanosoma brucei* pteridine reductase 1 is essential for survival in vitro and for virulence in mice. *Mol Microbiol* **77**, 658–671.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Ruiz-Postigo, J.A., Fèvre, E.M., Mattioli, R.C., Jannin, J.G., 2012. Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl Trop Dis* **6**, e1859.

Simarro, P.P., Diarra, A., Ruiz Postigo, J.A., Franco, J.R., Jannin, J.G., 2011. The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward. *PLoS Negl Trop Dis* **5**, e1007.

Simarro, P.P., Jannin, J., Cattand, P., 2008. Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med* **5**, e55.

Singh, J., Petter, R.C., Baillie, T.A., Whitty, A., 2011. The resurgence of covalent drugs. *Nat Rev Drug Discov* **10**, 307–317.

Sjöblom, B., Salmazo, A., Djinović-Carugo, K., 2008. α-Actinin structure and regulation. *Cellular and molecular life sciences* **65**, 2688–2701.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I.A., Lesley, S.A., Godzik, A., 2007. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405.

Spinks, D., Ong, H.B., Mpamhanga, C.P., Shanks, E.J., Robinson, D.A., Collie, I.T., Read, K.D., Frearson, J.A., Wyatt, P.G., Brenk, R., Fairlamb, A.H., Gilbert, I.H., 2011. Design, synthesis and biological evaluation of novel inhibitors of *Trypanosoma brucei* pteridine reductase 1. *ChemMedChem* **6**, 302–308.

Stein, N., 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst* . **41**, 641-643.

Steinbacher, S., 1999. Crystal structure of the post-chaperonin β-tubulin binding cofactor Rbl2p. *Nat Struct Biol* **6**, 1029–1032.

Steinborn, K., Maulbetsch, C., Priester, B., Trautmann, S., Pacher, T., Geiges, B., Küttner, F., Lepiniec, L., Stierhof, Y.-D., Schwarz, H., Jürgens, G., Mayer, U., 2002. The *Arabidopsis* PILZ group genes encode tubulin-folding cofactor orthologs required for cell division but not cell growth. *Genes Dev*. **16**, 959–971.

Steiner, T., 2002. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed*. **41**, 48–76.

Studier, F.W., 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207–234.

Stura, E.A., Satterthwait, A.C., Calvo, J.C., Kaslow, D.C., Wilson, I.A., 1994. Reverse screening. *Acta Crystallogr D Biol Crystallogr* **50**, 448–455.

Stura, E.A., Wilson, I.A., 1991. Applications of the streak seeding technique in protein crystallization. *Journal of Crystal Growth* **110**, 270–282.

Sullivan, K.F., Cleveland, D.W., 1986. Identification of conserved isotype-defining variable region sequences for four vertebrate β tubulin polypeptide classes. *Proc Natl Acad Sci USA* **83**, 4327–4331.

Sundaramoorthy, R., Fyfe, P.K., Hunter, W.N., 2008. Structure of *Staphylococcus aureus* EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. *Journal of molecular biology* **383**, 603–614.

Taylor, G., 2003. The phase problem. A*cta Crystallogr D Biol Crystallogr* **59**, 1881–1890.

Taylor, G.L., 2010. Introduction to phasing. *Acta Crystallogr D Biol Crystallogr* **66**, 325–338.

Tian, G., Bhamidipati, A., Cowan, N.J., Lewis, S.A., 1999. Tubulin folding cofactors as GTPase-activating proteins: GTP hydrolysis and the assembly of the α/β-tubulin heterodimer. *J Biol Chem* **274**, 24054–24058.

Tian, G., Huang, Y., Rommelaere, H., Vandekerckhove, J., Ampe, C., Cowan, N.J., 1996. Pathway leading to correctly folded β-tubulin. *Cell* **86**, 287–296.

Tian, G., Lewis, S.A., Feierbach, B., Stearns, T., Rommelaere, H., Ampe, C., Cowan, N.J., 1997. Tubulin subunits exist in an activated conformational state generated and maintained by protein cofactors. *J Cell Biol* **138**, 821–832.

Tian, G., Vainberg, I.E., Tap, W.D., Lewis, S.A., Cowan, N.J., 1995. Quasi-native chaperonin-bound intermediates in facilitated protein folding. *J Biol Chem* **270**, 23910–23913.

Tóth, L., Muszbek, L., Komáromi, I., 2013. Mechanism of the irreversible inhibition of human cyclooxygenase-1 by aspirin as predicted by QM/MM calculations. *J. Mol. Graph. Model*. **40**, 99–109.

Tulloch, L.B., Martini, V.P., Iulek, J., Huggan, J.K., Lee, J.H., Gibson, C.L., Smith, T.K., Suckling, C.J., Hunter, W.N., 2010. Structure-based design of pteridine reductase inhibitors targeting African sleeping sickness and the leishmaniases. *J Med Chem* **53**, 221–229.

Turnbull, W.B., 2011. Divided We Fall? Studying low-affinity fragments of ligands by ITC *GE Healthcare*.

Tyndall, J.D.A., Nall, T., Fairlie, D.P., 2005. Proteases universally recognize beta strands in their active sites. *Chem. Rev.* **105**, 973–999.

Vagin, A., Teplyakov, A., 1997. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N., 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184–2195.

Vainberg, I.E., Lewis, S.A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H.L., Cowan, N.J., 1998. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**, 863–873.

Valpuesta, J.M., Martín-Benito, J., Gómez-Puertas, P., Carrascosa, J.L., Willison, K.R., 2002. Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. *FEBS Lett* **529**, 11–16.

van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R., Groen, F.C., 1990. Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem* **191**, 110–118.

Veltel, S., Gasper, R., Eisenacher, E., Wittinghofer, A., 2008. The retinitis pigmentosa 2 gene product is a GTPase-activating protein for Arf-like 3. *Nat Struct Mol Biol* **15**, 373–380.

Vickerman, K., 1985. Developmental cycles and biology of pathogenic trypanosomes. *Br. Med. Bull.* **41**, 105–114.

von Laue, M., 1915. Concerning the detection of X-ray interferences. *Nobel Lecture.* 1–9.

Wallace, A.C., Laskowski, R.A., Thornton, J.M., 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. **8**, 127–134.

Walter, T.S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., Grimes, J.M., 2006. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617–1622.

Weber, P.C., Salemme, F.R., 2003. Applications of calorimetric methods to drug discovery and the study of protein interactions. *Curr Opin Struct Biol* **13**, 115–121.

Weisbrich, A., Honnappa, S., Jaussi, R., Okhrimenko, O., Frey, D., Jelesarov, I., Akhmanova, A., Steinmetz, M.O., 2007. Structure-function relationship of CAP-Gly domains. *Nat Struct Mol Biol* **14**, 959–967.

Whitmore, L., Wallace, B.A., 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668–73.

WHO, 2013a. Chagas disease (American trypanosomiasis) [Updated March 2013] Available: http://www.who.int/mediacentre/factsheets/fs340/en/index.html

WHO, 2013b. Neglected Tropical Diseases [Updated November 2013] Available: http://www.who.int/neglected_diseases

Wilson, A.J.C., 1950. Largest likely values for the reliablity index. *Acta Cryst* **3**, 397-398.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.

Winter, G., 2010. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst*. **43**, 186-190.

Wiseman, T., Williston, S., Brandts, J.F., Lin, L.N., 1989. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**, 131–137.

Wu, G., Chai, J., Suber, T.L., Wu, J.W., Du, C., Wang, X., Shi, Y., 2000. Structural basis of IAP recognition by Smac/DIABLO. *Nature* **408**, 1008–1012.

Xu, Z., Liu, Z., Chen, T., Chen, T., Wang, Z., Tian, G., Shi, J., Wang, X., Lu, Y., Yan, X., Wang, G., Jiang, H., Chen, K., Wang, S., Xu, Y., Shen, J., Zhu, W., 2011. Utilization of Halogen Bond in Lead Optimization: A case study of rational design of potent phosphodiesterase type 5 (PDE5) Inhibitors. *J Med Chem* **54**, 5607–5611.

Ylänne, J., Scheffzek, K., Young, P., Saraste, M., 2001. Crystal structure of the α-actinin rod reveals an extensive torsional twist. *Structure* **9**, 597–604.

Yoder, M.D., Lietzke, S.E., Jurnak, F., 1993. Unusual structural features in the parallel β-helix in pectate lyases. *Structure* **1**, 241–251.

You, L., Gillilan, R., Huffaker, T.C., 2004. Model for the yeast cofactor A-β-tubulin complex based on computational docking and mutagensis. *Journal of molecular biology* **341**, 1343–1354.

Zhu, G., Zhai, P., He, X., Terzyan, S., Zhang, R., Joachimiak, A., Tang, J., Zhang, X.C., 2003. Crystal structure of the human GGA1 GAT domain *Biochemistry* **42**, 6392-6399.