ADDENDUM

Page XV-XVII: Additions in 'List of Abbreviations'

ACN; Acetonitrile, HMQC; Heteronuclear multiple quantum correlation, HRMS; *High resolution mass spectrometry,* ITC; *Isothermal titration calorimetry,* LRMS; *Low resolution mass spectrometry,* PABA; p-Aminobenzoic acid, SOFAST; Band-Selective optimised flip angle short transient, TFA; *Trifluoroacetic acid.*

Page 3-4: The following acronyms have been incorporated

DHNA (dihydroneopterin aldolose), HPPK (6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase) and FPGS (folylpolyglutamate synthase).

Page 17, Figure 1.9: Addition in figure legend

"...selected HPPKs (*Staphylococcus aureus, Escherichia coli, Haemophilus influenza, Bacillus anthracis, Mycobacterium tuberculosis, Pneumocystis carinii, Saccharomyces cerevisiae, Helicobacter pylori, Francisella tularensis*). Absolutely conserved residues (orange), highly conserved residues (blue), similar residues (light blue) and..."

Page 54, line 13: delete " the spectra became considerably weaker" read " the NMR signal intensity decreased".

1

Page 58, Figure 2.15: Addition in figure legend

"The y-axis represents the amount of bound analyte in terms of RU, and the x-axis represents time after injection (s)"

Page 67: Addition in Table legend

"N-terminal sequence analysis information was obtained using the automatic sequence analyser (Waters) at CSIRO Materials Science and Engineering, Parkville."

Page 84, line 4: delete "8-mercaptoguanine" read "8-MG"

Page 9, Figure 4.5: Addition in figure legend

"The data was obtained in triplicate."

Page 115, Under Equation 5.1 now reads:

"where the coordinates of the nucleus with respect to the principal axis of the magnetic susceptibility tensor ($\Delta\chi$) are defined by r (the distance from the *e.g.* HN nucleus to the unpaired electron in the metal), θ and ϕ (the polar angles describing the metal-nucleus vector within the $\Delta\chi$ frame, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic components of the anisotropic magnetic susceptibility tensor. The orientation of the $\Delta\chi$ frame with respect to the protein frame is defined by three Euler rotations (α, β and γ) in the ZYZ convention[6]. The PCSs depend on the distance between the nucleus of interest and the paramagnetic metal ion as $1/r^3$."

Page 119 and Page 120: read "figure 5.3 as figure 5.2" and "figure 5.2 to figure 5.3".

Page 119, line 5: delete "thereby proving that only one LBT would ligate to *SaHPPK*" and read "thereby indicating that it was likely that only one LBT would ligate to *SaHPPK*"

Structure and inhibitors of *S. aureus* HPPK - a folate pathway enzyme

Submitted in total fulfillment of the requirements of the degree of

Doctor of Philosophy

By

Sandeep Chhabra

April 2012

Medicinal Chemistry

Monash Institute of Pharmaceutical Sciences (MIPS)

Monash University

MONASH University
Pharmacy and Pharmaceutical Sciences

PART A: General Declaration

Monash University
Monash Research Graduate School

Declaration for thesis based or partially based on conjointly published or unpublished work

General Declaration

In agreement with Monash University Doctorate Regulation <u>17/ Doctor of Philosophy and Master of Philosophy (MPhil) regulations</u> the following declarations are made:

I hereby declare that this thesis includes no material that has been acknowledged for the award of any other degree or diploma in any university or academic institution and, to the best of my knowledge and belief, this thesis contains no study or material previously published or written by another individual, except where due reference is made in the text of the thesis.

This thesis includes two main original papers published in peer reviewed journals and four additional papers where my contribution was essential, but was not the primary aim of this work. The core theme of the thesis is to characterise the *Sa*HPPK structure and find small molecule inhibitors for the development of novel antibacterials. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Department of Medicinal Chemistry under the supervision of Dr James D. Swarbrick, Dr Jamie S. Simpson and Dr Thomas S. Peat (CSIRO).

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of published papers my contribution to the work involved the following:

Thesis chapter	Publication title	Publication status	Nature and extent of candidate's contribution
2	Crystallisation and preliminary X-ray analysis of 6-hydroxymethyl-7,8-dihydropterin	Published	Experimental design. Conduct of laboratory work. Data analysis. Preparation of manuscripts and subsequent revisions. Total contribution 70%

	pyrophosphokinase from *Staphylococcus aureus.*		
3	Structure of *S. aureus* HPPK and the discovery of a new substrate site inhibitor.	Published	Experimental design. Conduct of laboratory work. Data analysis. Preparation of manuscripts and subsequent revisions. Total contribution 70%

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Signed:

Date: 15-08-2012

Publication Record:

Publications:

1. **Chhabra S**, Dolezal O, Collins BM , Newman J, Simpson JS, Macreadie IG, Fernley RT, Simpson JS, Peat TS, Swarbrick JD. Structure of *S. aureus* HPPK and the discovery of a new substrate site inhibitor. *PLoS One*, **2012**, 7(1): e29444.

2. **Chhabra S**, Newman J, Peat TS, Fernley RT, Caine J, Simpson JS, Swarbrick JD. Crystallisation and preliminary X-ray analysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Staphylococcus aureus*. *Acta Crystallogr* Sect F Struct Biol Cryst Commun., **2010**, *66*, 575-8.

Additional Publications during the thesis period:

1. Swarbrick JD, Ung P, **Chhabra S**, Graham B. An iminodiacetic acid based lanthanide binding tag for paramagnetic exchange NMR spectroscopy. *Angew Chem Int Ed Engl.* **2011**, *50*, 4403-4406.

2. Swarbrick JD, Shaw DJ, **Chhabra S**, Ghai R, Valkov E, Norwood SJ, Seaman MN, Collins BM. VPS29 is not an active metallo-phosphatase but is a rigid scaffold required for retromer interaction with accessory proteins. *PLoS One*, **2011**, *6* (5), e20420.

3. Swarbrick JD, Ung P, Su XC, Maleckis A, **Chhabra S**, Huber T, Otting G, Graham B. Engineering of a bis-chelator motif into a protein α-helix for rigid lanthanide binding and paramagnetic NMR spectroscopy. *Chem Commun.* **2011**, *47*, 7368-7370.

4. Graham B, Loh CT, Swarbrick JD, Ung P, Shin J, Yagi H, Jia X, **Chhabra S**, Barlow N, Pintacuda G, Huber T, Otting G. DOTA-Amide Lanthanide Tag for Reliable Generation of Pseudocontact Shifts in Protein NMR Spectra. *Bioconjug Chem.* **2011**, *22*, 2118-25.

DEDICATION

To my dearest parents, sister and late grandmother.

Acknowledgements

I want to express my gratitude to many people who helped me in this thesis.

First and foremost, I would like to thank my research supervisor Dr James Swarbrick, for being an excellent mentor and research guide. His continual passion, verging on obsession with NMR was nothing short of inspirational. He showed me how the technique can provide an answer to most questions regarding the structure and function of proteins and their interactions with small molecules at the heart of a structure based drug discovery project. I thank him for creating and designing the HPPK project and for seeing my potential and encouraging me to develop a wide skill set necessary to learn and appreciate the language of structural biology and drug discovery. Moreover, whenever we came up with an idea he was always willing 'to give it a go' no matter how insane or crazy it was at the time. I learned so much from hypothesis driven and investigative research and I thank him for this insight and to show me how to take full responsibility, particularly when experiments fail and to graciously accept criticism such as 'I told you so' from a PhD student.

I am indebted for his help with NMR assignments and experiments and with the foresight into the application of modern paramagnetic NMR using lanthanides binding tags. It wasn't possible for me to carry out this project without his endless efforts, directions and encouragement that included the freedom to carry out any of my own research ideas which ultimately helped me to grow independently as a scientist. He amply funded all the research experiments, several conferences and workshops. I am grateful for his help in sharpening my presentation and writing skills and for his constructive criticism and encouragement, in a manner typical of a Northern Englishman. Finally I thank him for personally showing me the delights, enormous benefit and social impact of late night and weekend NMR work.

I am greatly thankful to my other supervisor Dr Jamie Simpson for giving me an opportunity to pursue a PhD within MIPS, his invaluable guidance, critical

suggestions and comments, and always pushing me harder throughout this project. To this day I apologize for not keeping an up to date lab book at your continual recommendations - I now know how this impacts on writing a thesis. I am grateful to him for all the support and encouragement during my candidature.

I am indebted to Dr Tom Peat (CSIRO) for his great mentorship and enthusiasm throughout this project. Thanks for teaching me many wonderful skills and for providing the opportunity to learn state-of-the-art X-ray crystallography and particularly for being there during the graveyard shifts at the Australian Synchrotron. Your expertise and breadth of knowledge in crystallography was second to none particularly when mounting the tiny crystals of HPPK. I also thank you for facilitating additional financial support from CSIRO, for introducing me to Ollie and Janet and for sending me to the Oxford X-ray workshop in 2010.

I am deeply indebted to Dr Bim Graham, my scientific mentor and 'guru', who showed great confidence in me and made me appreciate the beauty and nuances of 'smart synthesis'. I am grateful for his enthusiastic ideas behind the design of analogues of a non-classical hit scaffold and for his invaluable help for the application of paramagnetic tags on HPPK.

I would like to thank Dr Janet Newman (C3, CSIRO) for her much appreciated efforts for the optimization of the crystallographic conditions and in showing me how to set up crystal trays. I would also like to thank Dr Olan Dolezal (CSIRO), for sharing his expertise in biophysical studies and for his help in conducting the SPR experiments and analysis of data. I am also grateful for his help in designing the mutation studies. I am thankful to Dr Brett Collins (IMB, UQ), for conducting the initial ITC experiments and his expertise in the analysis of data.

I would also like to thank Dr Ross Fernley (CSIRO) for his assistance in protein purifications and providing me with an opportunity to learn and conduct *in vitro* assays. I would like to thank Dr Joanne Caine for helping me

with protein expressions techniques. I am grateful to Dr Lesley Pearce for the assistance with the PCR and mutation experiments and also many thanks to Dr Meghan Hattarki for her help with the SPR experiments.

I am thankful to Prof. Keith Watson (WEHI) for allowing me to work at the HTS facility of WEHI (Bundoora) and the optimisation of the *in vitro* assay for plasmodium HPPK. I would like to thank Dr Ian Macreadie (RMIT) for performing the *in vivo* testing of our compounds and initial guidance and interest in folate pathway targets whilst he was at CSIRO. I am thankful to Dr Martin Scanlon for providing protein purification lab facilities. I would also like to thank Mr Nick Barlow (MIPS) for providing me with his synthetic chemistry expertise in optimising guanine chemistry. I would like to thank my colleague and close friend Mr Phuc Ung for assisting with paramagnetic tags. Mr Ivan Troitsky, my fellow lab mate and also a close friend, who has helped me with scientific writing and offered various useful tips and deserve a special vote of thanks. I would like to thank, Miss Natalie Vinh for all the chemistry and assay related discussions and to Dr. Jason Dang for NMR and Mass Spec facilities and for data analysis.

Many thanks to all my lab buddies and friends, who have made working on this project through the years so pleasant, for their help and good times: Dr Stephen Headey, Dr Kieran Rimmer, Mr Martin Williams, Dr Tony Velkov, Miss Yanni Chin, Miss Jennifer La, Miss Amelia Vom, Mr Will Nguyen, Mr Michael Lee, Miss Mansha Vazirani, Miss Michelle Miller, Miss Diana Neale and Dr Aisha Laguerre.

I thank Monash University for financial support and Monash Research Graduate School for MIPRS Scholarship during the course of my PhD. I thank CMSE (CSIRO, Parkville campus) for the top-up scholarship. I would also like to thank the C3 crystallization facility and Australian Synchrotron for data collection. Thank you to all the staff at the MIPS and CSIRO for your support in all these years.

Most importantly to Dr Pooja Sharma my friend, my girlfriend and now my wife, I thank you for constant support, encouragement and the continual belief in me. You have always been there for me and I will always be there for you. Finally, I wish to thank my parents, Smt Sheela and Shri Ramchandra, and my sister, Deepa, for their endless and selfless support and love which helped me to overcome the toughest hurdles of my life and to reach this milestone in my career.

Abstract

Folate is an essential vitamin (vitamin B9) that is required for many one-carbon transfer reactions and is a critical precursor for the biosynthesis of purines, pyrimidines, and amino acids. Folate is a dietary requirement for man and animals. However bacteria, parasites and plants can synthesize folate *de novo* using enzymes of the folate biosynthesis pathway. Thus folate biosynthesis remains a key target for antimicrobial therapy as exemplified by the sulfa drugs that date back to the 1940s, acting on the enzyme DHPS. HPPK, like other enzymes of the folate biosynthesis pathway, is absent in man and therefore is a potential drug target for the development of antimicrobial agents. Increasing rates of drug resistance to current antibiotics is a serious problem, due to the rapid adaptation of microorganisms to our chemical interventions. Because HPPK is not the target of any existing antibiotic, it is attractive for the development of new antibiotics against resistant strains.

The dearth of substrate-site inhibitors for HPPK observed in the literature reflects a combination of difficult pterin-like chemistry and a paucity of commercially available pterin-like molecules, compounded by a highly specific substrate site. This knowledge gap presents interesting research opportunities.

Herein, the first structural and biophysical data on 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (*Sa*HPPK), from the pathogen *Staphylococcus aureus* is presented. HPPK catalyses the pyrophosphoryl transfer from the cofactor (ATP) to the pterin substrate (6-hydroxymethyl-7,8-dihydropterin, HMDP). The expression, purification and protein stability was optimised to provide high yields for both screening and crystallisation experiments. A combination of fragment and *in silico* based methods was adopted to identify small molecule inhibitors of the *Sa*HPPK enzyme. Fragment screening of the Maybridge RO3 library resulted in a set of structurally diverse binders for the substrate and ATP binding sites. While the fragments bound weakly as expected (K_d in the range of ~0.5-1 mM), a ROCS *in silico* screening approach that targeted the substrate scaffold led to the discovery of 8-mercaptoguanine (8-MG) which was shown to bind with an

equilibrium dissociation constant, K_d, of ~13 µM as measured by isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR). An IC_{50} of ~41 µM for the inhibition of SaHPPK by 8-MG was determined by means of a luminescent kinase assay. In comparison to the biological substrate (HMDP), the 8-MG has no requirement for magnesium or the ATP cofactor for competitive binding to the substrate site. The 1.65 Å resolution crystal structure of the binary SaHPPK/8-MG inhibited complex showed that it binds in the pterin site and shares many of the key intermolecular interactions of the substrate. The NMR chemical shift, ^{15}N heteronuclear NOE, and ^{15}N-^{1}H residual dipolar coupling measurements, indicate that the sulfur atom of 8-MG is likely important for stabilizing and restricting motions within the L2 and L3 catalytic loops in the inhibited SaHPPK/AMPCPP/8-MG ternary complex. A high-affinity lanthanide-binding tag was used to generate pseudocontact shifts (PCS) to obtain long range structural information to provide further insight about the structure and loop position in the SaHPPK/AMPCPP/8-MG ternary complex that would otherwise be difficult to obtain using standard NMR methods. The X-ray crystal structure guided the design of 8-MG analogues which were later synthesised to explore structure–activity relationships (SAR) and to seek a chemical route for future hit expansion out of the substrate pocket and towards the ATP site

Overall, the current study describes for the first time a comprehensive analysis of the SaHPPK structure and reports the identification and characterisation of novel inhibitors and their interactions with SaHPPK all of which may provide a foundation for the development of antimicrobials that target the folate biosynthetic pathway enzyme HPPK.

Table of Contents

XII

List of Abbreviations

α	Alpha
β	Beta
CSP	Average chemical shift change (perturbation)
° C	Degrees Celsius
ACN	Acetonitrile
caMRSA	Community-Associated MRSA
DHPS	Dihydropteroate synthase
DHFR	Dihydrofolate reductase
DMSO	Dimethyl sulfoxide
*Ec*HPPK	*Escherichia coli* HPPK
FPLC	Fast protein liquid chromatography
FBS	Fragment based screen
FBDD	Fragment based drug discovery
ΔG	Change in Gibb's free energy
ΔH	Change in enthalpy
HAC	Heavy atom count
haMRSA	Hospital-acquired MRSA
HPLC	High performance liquid chromatography
HPPK	6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase
HMQC	Heteronuclear multiple quantum correlation
^{1}H-^{15}N HMQC	^{1}H-^{15}N Heteronuclear multiple quantum correlation

HRMS	High resolution mass spectrometry
HTS	High throughput Screening
ITC	Isothermal titration calorimetry
K_d	Dissociation constant
K_i	Inhibition constant
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
LBT	lanthanide binding tag
LE	Ligand efficiency
LRMS	Low resolution mass spectrometry
8-MG	8-Mercaptoguanine
MRSA	Methicillin-resistant *S. aureus* strains
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser enhancement spectroscopy
PABA	p-Aminobenzoic acid
PCSs	Pseudocontact shifts
PDB	Protein Data Bank
ppm	Parts per million
R	Gas constant
rms	Root mean square
RO3	Rule of three

RP-HPLC	Reversed-phase chromatography
rpm	Revolutions per minute
ΔS	Change in entropy
*Sa*HPPK	*Staphylococcus aureus* HPPK
SBDD	Structure based drug design
SMX	Sulfamethoxazole
SOFAST	Band-selective optimised flip angle short transient
STD	Saturation transfer difference
SPR	Surface Plasmon Resonance
3D	Three dimensional
2D	Two dimensional
T	Temperature
TFA	Trifluoroacetic acid
TMP	Trimethoprim
TSA	Thermal shift assay
vd*W*	van der *Waals*

Chapter 1

Introduction

1. Introduction

1.1 Antibiotic resistance

Antibiotic resistance is rapidly emerging as one the most significant health challenges of this century[1]. Bacterial resistance continues to increase and impacts on our treatment of bacterial infectious diseases[2]. All new antibiotics have been countered by the swift onset of resistance mechanisms[3] such that we have come full circle and are returning to conditions reminiscent of the preantibiotic era[2]. In Europe, 25,000 deaths were reported in 2007 as a result of antimicrobial resistance with an estimated cost of £1.3 billion per year (Innovative Medicines Initiative data).

Resistance is on the rise and antibiotic drug discovery is on the decline. Economic pressures combined with a low return of investment for antibiotic development mean that the pharmaceutical industry favours drugs for long term treatments of chronic disease (*e.g.* hypertension, depression, dementia, and rheumatoid arthritis) over short term curative therapies typical of antibiotics. The discovery of new antibiotics is challenging as exemplified by the fact that only six classes of antibiotics have ever been discovered[2]. Moreover, only two new systemic antibiotics have been approved by the FDA in last 30 years; the oxazolidinone linezolid (Zyvox; Pfizer Pharmaceuticals) in 2000 and the cyclic lipopeptide daptomycin (Cubicin; Cubist Pharmaceuticals) in 2003[4].

The FDA has shown its concern for an imminent health crisis by granting fast track priority to oral agents for both Gram-Positive (methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococci* (VRE), *M. tuberculosis*) and Gram-negative (*Pseudomonas aeruginosa*, *Acinetobacter baumannii, Escherichia coli*) pathogens that are devastating our health care institutions and our communities.

Most antibiotics are derived from natural products. While recent advancements in technology have helped in genome mining for underexplored microbial niches, the discovery of natural product based antibiotics is becoming rarer and rarer. The alternate, target based HTS (high-

throughput screening) campaigns, while popular, often yield disappointing results[3], which has been ascribed to the chemical incompatibility between the "drug-like" hydrophobic nature of HTS library compounds and the large, often hydrophilic properties of many known antibiotics. Nevertheless, recently, by screening existing eukaryote kinase libraries against a whole cell antibacterial assay new validated targets have been discovered along with selective broad-spectrum gram negative antibacterials[5-7]. Once a suitable target is discovered, structure-based approaches are widely accepted as a way to expedite the hit-to-lead drug discovery process[8, 9].

1.2 *Staphylococcus aureus*

Staphylococcus aureus (*Sa*) is a clinically important, opportunistic pathogen and one of the major contributors of hospital and community acquired bacterial infections. Methicillin-resistant *S. aureus* strains (MRSA, commonly known as a "superbug"), causes up to 19,000 deaths in the US alone and an estimated health care cost of $3–4 billion USD per annum[10]. MRSA strains are classified by genotypic and phenotypic characteristics. There are two main types: those originating in hospitals (nosocomial, haMRSA strains USA100 and USA200) and those in the community (caMRSA), of which the latter is almost entirely caused by the pandemic USA300 strain[11]. Infections with USA300 range from skin and soft tissue infections, to life threatening systemic infections such as bacteremia and necrotizing pneumonia. In contrast to haMRSA, caMRSA infections tend to occur in previously healthy younger patients without health care exposure[12].

MRSA infections first emerged in the UK in 1961 due to the overuse of penicillin and methicillin[13], and by the 1990s, MRSA was considered to be endemic in most areas of the world. Currently, caMRSA is more susceptible to a range of chemotherapies than the multi-drug resistant haMRSA[14]. Although resistant to tetracycline, erythromycin, clindamycin, linezolid, and in some cases vancomycin, caMRSA is largely susceptible to TMP-SMX (trimethoprim-sulfamethoxazole) combination therapy. This combination synergistically blocks the biosynthesis of folate pathway enzymes by acting on

dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS), respectively (Figure 1.1)[15, 16]. TMP-SMX resistance has emerged in haMRSA owing to a thymidine salvage pathway occurring in cases with significant necrosis associated with late stage infections[16-18]. While in principal the same escape pathway could eventuate in caMRSA strains, recent *in vitro* results show that combining TMP-SMX with the approved thymidine kinase inhibitor, 5-Iodo-2'-deoxyuridine impaired the utilisation of external thymidine up to levels greater than that found in human plasma. Synergistic bactericidal effects in *S. aureus* and broad spectrum activity against several Gram-positive and Gram-negative bacteria was observed *in vitro* offering a new potential modality to synergistic folate pathway antagonists[18].

SMX-TMP resistance in caMRSA, on the other hand, is attributed to simple mutations in the DHPS or the DHFR gene. The latter case results in a repositioning of the substrate in the active site[19], which compromises TMP based therapy. Inhibitors that either act on these mutant targets or other essential enzymes from the same pathway are two potential routes to new therapeutics for caMRSA. In conjunction with a thymidine kinase inhibitor, broad spectrum *S. aureus* therapeutics, in principle, may be potentially useful.

1.3 The folate pathway enzymes and the antibiotics TMP and SMX

Folates are essential for the growth of all living cells. The reduced form of folate, tetrahydrofolate (THF), participates in several important one-carbon transfers, critical for the biosynthesis of thymidine, glycine and methionine, and for DNA replication[20, 21]. In humans and other higher eukaryotes, dietary folate (vitamin B9) is actively transported into growing cells. Conversely, bacteria and lower eukaryotes as well as plants lack the proper transport systems for uptake across the cell membrane and therefore depend on *de novo* folate production within their own cells[22]. Several enzymes of the folate pathway (Figure 1.1), including DHPS, DHNA (dihydroneopterin aldolose) and HPPK (6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase), are absent in humans, rendering them potentially attractive as antimicrobial targets[23].

The enzyme DHPS is a target for existing sulfa drugs e.g. SMX. Similarly the DHFR enzyme is targeted by the antifolate trimethoprim (TMP). New generation analogues of TMP are still under active investigation *e.g* in 2009, Iclaprim showed better activity than TMP and was highly active against vancomycin-resistant *Enterococci* and reached stage III clinical trials[24]. Clinically, TMP-SMX is used as a convenient oral formulation with good absorption, for outpatient treatment of caMRSA associated with cellulitis, particularly in children, and bone and joint infections, meningitis, brain or spinal epidural abscesses and septic thrombosis. TMP-SMZ is also administered as salvage agents for persistent caMRSA bacteremia[25, 26]. TMP-SMZ therapy is currently administered to treat a range of other diseases including malaria and pneumocystis pneumonia (PCP).

1.4 Catalytic reactions of the folate pathway enzymes

The first step in the pathway involves GTP cyclohydrolase (GTP-CH), which catalyses the rearrangement of GTP to form the initial pterin ring structure *i.e.* 7,8-dihydroneopterin triphosphate (Figure 1.1)[27]. This is followed by enzymatic steps involving a nudix phosphatase[28] and aldolose (DHNA) to form 6-hydroxymethyl 7,8-dihydropterin (HMDP), which is a substrate for HPPK[27]. HPPK then catalyses the transfer of pyrophosphate from ATP to HMDP, resulting in 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (HMDPPP)[23]. DHPS catalyses the formation of 7,8-dihydropteroate (DHPt) by condensation of HMDPPP and *para*-aminobenzoic acid (*p*ABA)[27, 29]. Addition of the glutamate moiety to DHPt provides 7,8-dihydrofolate (DHF), and subsequent reduction by dihydrofolate reductase (DHFR) produces 5,6,7,8-tetrahydrofolate (THF)[23]. Polyglutamation by FPGS (folylpolyglutamate synthase) makes a more stable THF that remains within cells.
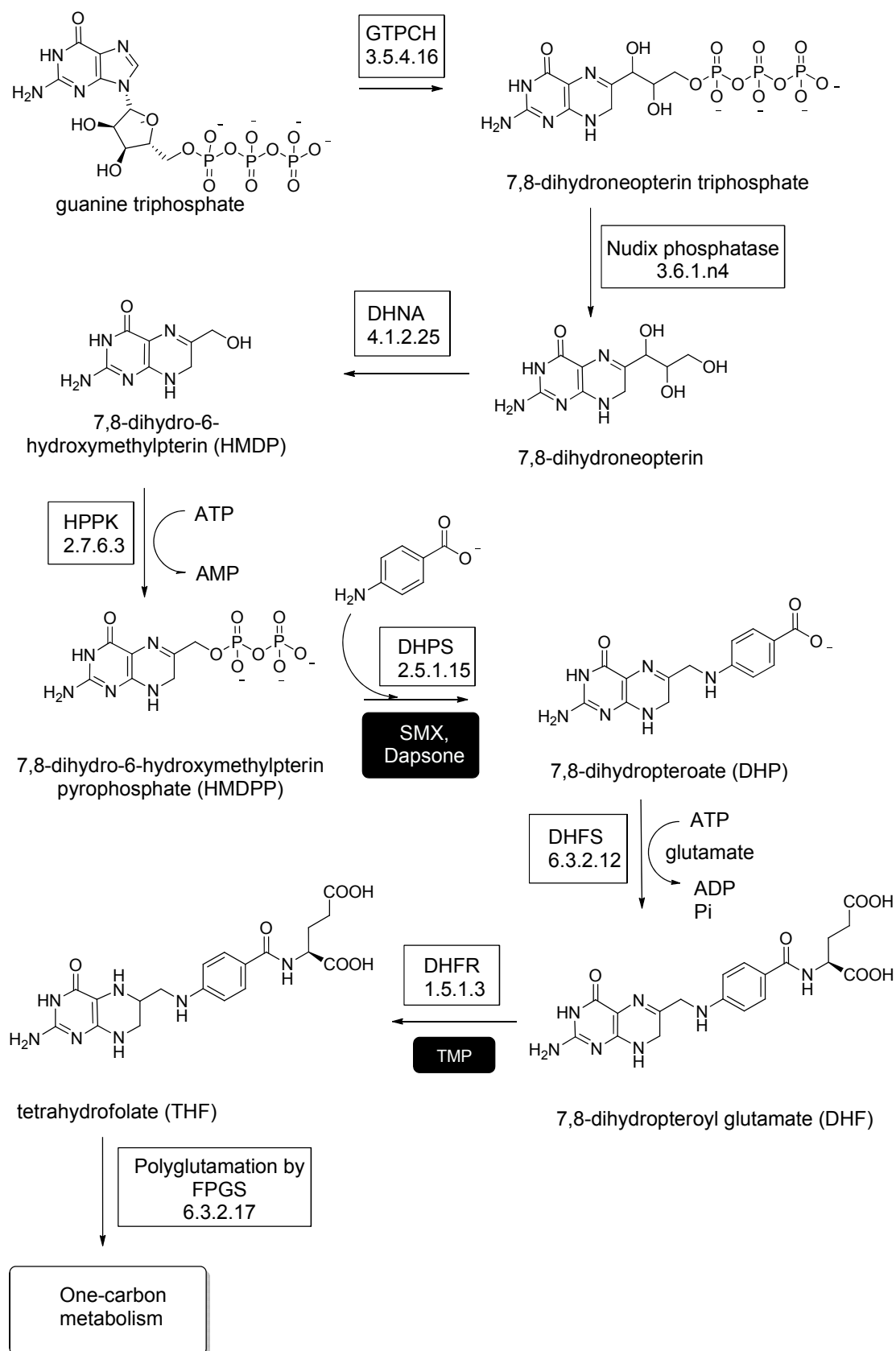
Figure1.1: The Folate Biosynthetic Pathway. Steps targeted by existing drugs are indicated in filled black boxes.

1.5 HPPK

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK; EC 2.7.6.3), the enzyme upstream and adjacent to DHPS, is not the target of any existing drug. HPPK catalyzes a Mg^{2+}-dependent pyrophosphoryl transfer from ATP to 6-hydroxymethyl-7,8-dihydropterin (HMDP) (Figure 1.2). HPPK is a small (158 residues, ~18 kDa), generally monomeric protein, which has been studied using various biophysical techniques, including X-ray crystallography and NMR spectroscopy[30]. Over 26 HPPK X-ray and NMR structures have been reported in various ligand bound and *apo* states from a variety of organisms (*Escherichia coli, Haemophilus influenzae, Saccharomyces cerevisiae, Streptococcus pneumonia, Yersinia pestis* and *Francisella tularensis*[31-36]). These data have provided atomic level information on the catalytic mechanism and protein dynamics of the reaction trajectory during catalysis[37].
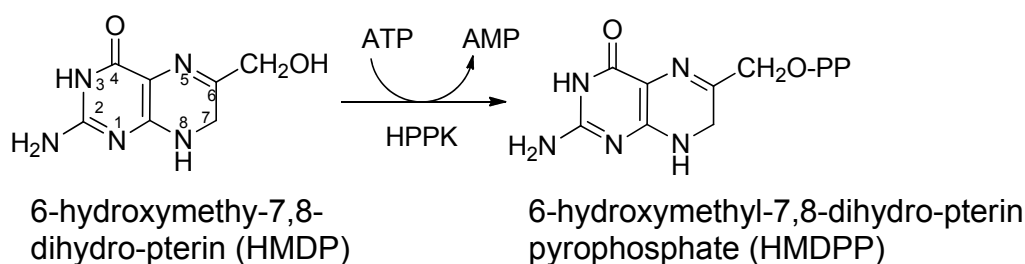
Figure 1.2: Pyrophosphoryl transfer catalysed by HPPK.

1.5.1 HPPK structure and mechanism of catalysis

Although novel, the HPPK structure is similar to a ferredoxin-like fold and predominantly comprises a six-stranded β-sheet sandwiched by two α-helices on either side (αβα sandwich fold) (Figure 1.3)[38]. Much of our understanding of the mechanism of action has been gained from X-ray structures in complex with substrates (HMDP) and the non-hydrolysable ATP analogue, AMPCPP (α,β-methylene-adenosine-5′-triphosphate).

Both structural and functional analyses of *E. coli* HPPK (*Ec*HPPK) indicate that protein dynamics are important for substrate binding[38-40], whereby large conformational changes take place in three catalytic loops: loop L1 (between α1 and β1, residues 8 to 15), loop L2 (between β2 and β3, residues 44 to 53) and loop L3 (between α2 and β4, residues 80 to 93) (Figure 1.3). These three loop regions are believed to play an important role in substrate recognition and are critical for assembling the active centre of HPPK[41]. While loop L3 undergoes the largest and most dramatic conformational change, all three loops seal the binding sites of both HMDP (substrate) and ATP (cofactor) for chemical transformation to occur[39]. Within a catalytic cycle, loop L3 opens up twice (Figure 1.4): firstly upon the binding of Mg-ATP, which is necessary for the enzyme to bring the side chains of two catalytic arginine residues (Arg82 and Arg92) into the active site[42, 43], and then upon the completion of chemical transformation, prior to the product release[37, 39].

The two ligands (HMDP and ATP) interact with two magnesium ions and a total of 26 residues, 13 of which are conserved across all known species[42]. *In vitro* kinetic studies have shown a preferred order of substrate binding. At cellular levels of magnesium, the cofactor, ATP, binds first[44]. This is consistent with X-ray crystallographic studies on the enzyme, which show that three loop regions are involved in a conformational change upon binding of ATP/AMPCPP that prepares the HMDP binding site for substrate recognition. Accordingly, in the absence of cofactor, HMDP binds weakly *in vitro* to the *apo* enzyme[45].
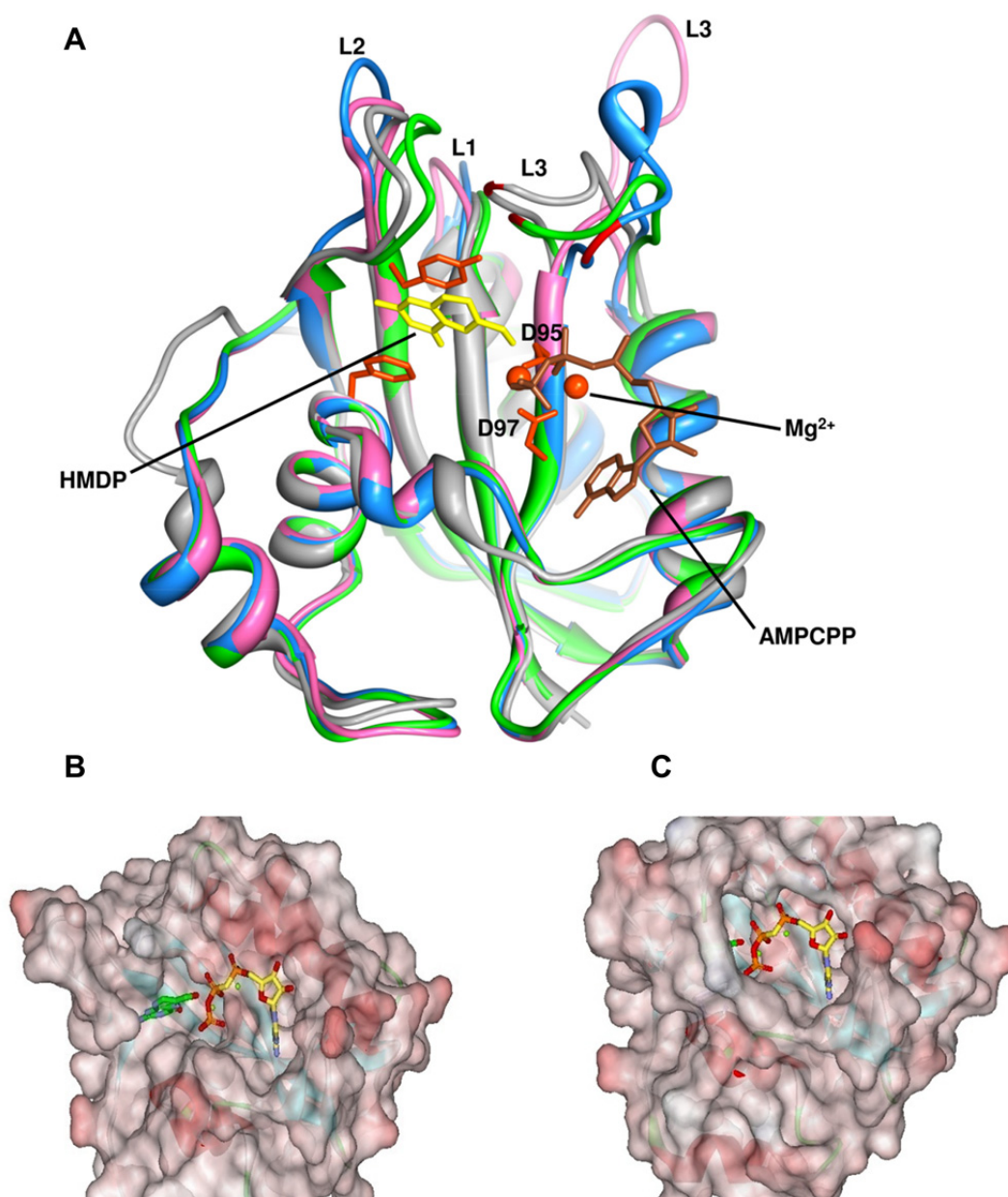
Figure 1.3: (A) Superposition of selected HPPK structures in complex with HMDP (yellow) and AMPCPP (brown), showing the prototypical HPPK fold, active site loops, the HPPK conserved ring stacking residues (orange) and magnesium ions (orange) bound by the conserved aspartate residues (D95/D97). Three *Ec*HPPK structures; 1Q0N (green), 1RAO (pink), 1HKA (blue) and HPPK from *H. influenza* 1CBK (grey) are shown. (B) Surface representation of the *Ec*HPPK *apo* enzyme (1HKA) showing HMDP (green) and AMPCPP (sticks) superimposed from a tertiary complex structure (1Q0N). Note that both the binding sites are open in the *apo* enzyme. (C) Surface representation of the *E. coli* ternary complex (1Q0N) with stick models of AMPCPP and HMDP superimposed. Note that the substrate (HMDP) binding site is closed in the ternary complex.

8
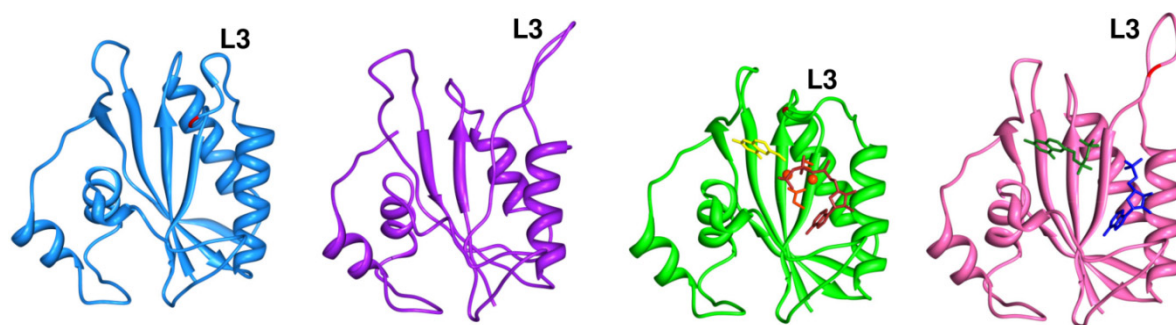
Figure 1.4: Loop L3 conformational changes during the catalysis of HPPK enzyme, apo (Blue) (1HKA), AMPCPP NMR complex (purple, without ligand in PDB) (1EQ0), HMDP+ AMPCPP complex (green) (1Q0N) and HMDPP+ AMP complex (pink) (1RAO) is shown.

A crystal structure of the ternary *Ec*HPPK complex (1Q0N) shows that the two Mg^{2+} ions in the active site play a major role in the recognition of both ATP and HMDP[45, 46]. Both Mg^{2+} ions are 6-coordinated, one bridging the α-β phosphates of AMPCPP, and the other bridging the β- and γ- phosphates[39]. Additional coordination is provided from the side chain carboxyl groups of two absolutely conserved Asp residues, Asp95 and Asp97, and the hydroxyl group of HMDP. The α- and β-phosphate groups of AMPCPP are hydrogen-bonded to the guanidinium groups of Arg92 and Arg84, and with the γ-phosphate, interact with the side chains of His115, Tyr116 and Arg121. The adenine moiety sits in a hydrophobic cleft and interacts with Ile98 and Thr112, while the ribose group is hydrogen bonded to the side chain of Arg84 and the carbonyl of Arg110 (Figure 1.5).

HMDP, on the other hand, is effectively sandwiched by two aromatic residues, Tyr53 and Phe123, and is involved in six hydrogen bonds with Thr42, Pro43, Leu45 and Asn55 (Figure 1.6), connecting positions 1, 2, 3, 4 and 8 of the substrate pterin ring (Figure 1.2) to the active site residues. Both active sites are highly selective for the substrate and cofactor. For example, the binding affinity for Mg-GTP is 260-fold less than for Mg-ATP[46].
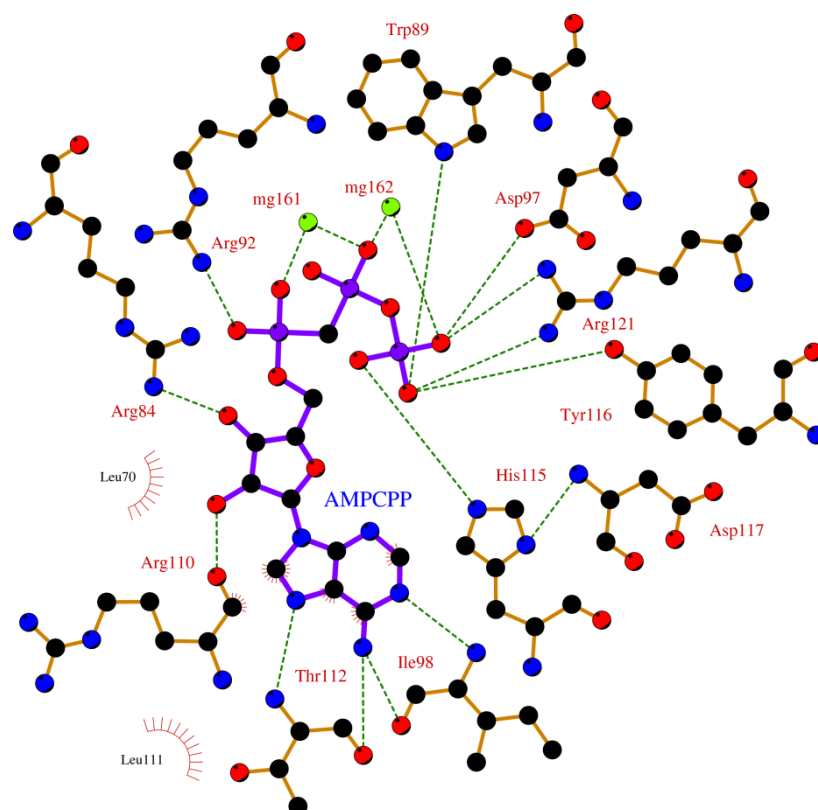
Figure 1.5: 2D Ligplot[47] schematic diagram of the cofactor subsite residues that make hydrogen bond interactions with AMPCPP.
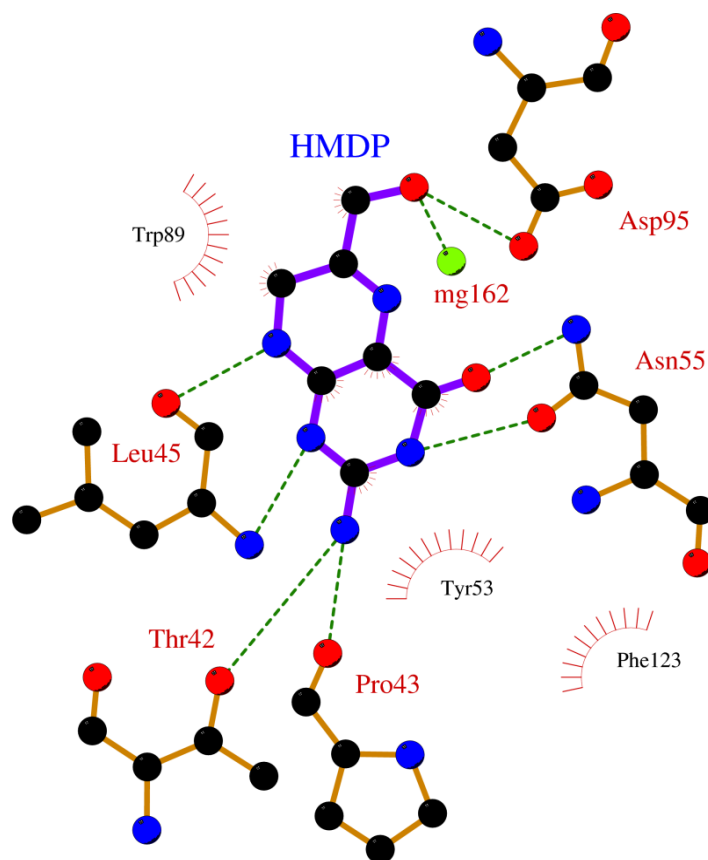
Figure 1.6: 2D Ligplot[47] schematic diagram of binding site residues making hydrogen bond interactions with HMDP.

1.5.2 Substrate binding and kinetic properties of HPPK

Binding affinities for various HPPK substrates have been reported in detail for *Ec*HPPK[44-46]. Stopped flow fluorometry experiments using the fluorescent ATP analogues 2'(3')-*O*-(*N*-methylanthraniloyl)adenosine5'-triphosphate (MANT-ATP) or 3'(2')-*O*-anthraniloyladenosine5'-triphosphate (Ant-ATP) report on the binding of ATP and AMPCPP. Competitive equilibrium displacement titration of Ant-ATP and MANT-ATP by ATP and AMPCPP gave K_d values in the range of 2.6-4.5 µM and 0.08-0.45 µM respectively[44-46]. A binding constant of 20 µM for ATP against *H. influenzae* HPPK has been determined using the intrinsic tryptophan fluorescence, which may suggest species variation[33]. In the presence of AMPCPP, HMDP binding constants have been reported in the range of 0.036-0.17 µM, based on a change in the natural fluorescence of the pterin ring. HMDP binding to the apo enzyme was too weak to be measurable. Binding kinetics for substrates of *Ec*HPPK have been determined similarly using steady-state kinetics. A summary of the binding properties of a range of substrates and analogues is presented in Table 1.1 and 1.2, respectively.

Table 1.1: Binding affinities of various substrate and product analogues to *Ec*HPPK.

Ligand Name	K_d (µM)	ΔG (kcal/mol)	Reference
ATP (with Mg^{2+})	4.5	-	[44]
ATP (with Mg^{2+})	2.6	-7.6	[45]
ATP (without Mg^{2+})	38	-6.0	[46]
ADP (with Mg^{2+})	55	-5.8	[46]
AMP (with Mg^{2+})	140	-5.3	[46]
Adenosine (with Mg^{2+})	210	-5.0	[46]
GTP (with Mg^{2+})	680	-4.3	[46]
AMPCPP (with Mg^{2+})	0.08	-9.7	[45]
AMPCPP (with Mg^{2+})	0.45	-	[44]
HMDP(with Mg^{2+}and AMPCPP)	0.036	-	[44]
HMDP(with Mg^{2+}and AMPCPP)	0.17	-9.2	[45]
HMDP (without Mg^{2+})	110	-5.5	[45]
HMDPP (with Mg^{2+})	0.2	-	[48]

1.5.3 Inhibition assays

A biochemical assay based on the monitoring of radioactively-labeled cosubstrate (α^{32}P ATP) has been developed to measure the activity of HPPK (*Ec*HPPK). This assay was successfully used to measure inhibition of HPPK by a series of bisubstrate analogues[49] (section 1.5.4), however the detection is not real-time and multiple reactions are required to determine the kinetic parameters.

An alternate HPPK/DHPS coupled assay has been developed based on monitoring the level of inorganic phosphate released after the action of a pyrophosphatase on the pyrophosphate product[44]. Similarly, this suffers from the same problem of having to stop the reaction. One of the aims of this thesis was therefore to develop and streamline a new assay suitable for high-throughput measurements.

Table 1.2: Kinetic constants (K_m) for *Ec*HPPK substrates.

Substrate Name	Method	K_m (µM)	Reference
HMDP	*steady state kinetics*	0.60 ± 0.01	44
HMDP	*radioactivity-based assay*	1.6 ± 0.4	50
ATP	*steady state kinetics*	3.4 ± 0.02	44
ATP	*radioactivity-based assay*	17 ± 3	50

1.5.4 HPPK inhibitors
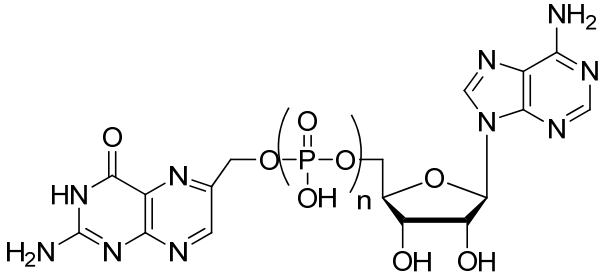
It is remarkable that despite the abundant structural information available on HPPK, only a few specific inhibitors of HPPK have been reported to date. Among these, two are based on the pterin substrate (HMDP); one featuring a geminal dimethyl substitution at the C7 position on the dihydropyrazine ring (code: 40Y67), the other, phenethyl and methyl substituted at the same position (code: 87Y76) (Figure 7)[51]. Crystal structures with these analogues bound to the *apo* enzyme have been reported (1DY3 and 1CBK), even though HMDP itself has very weak affinity for the *apo* enzyme[52]. The binding affinities of 40Y67 and 87Y76 have not been reported.
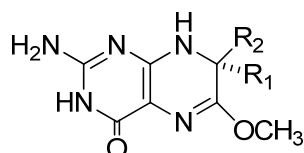
1. 40Y67, R_1 = R_2 = CH_3
2. 87Y76, R_1 = CH_3; R_2 = phenethyl

Figure 1.7: Pterin-derived HPPK inhibitors reported in literature.

A set of bisubstrate inhibitors of *Ec*HPPK has been reported[49] in which 6-hydroxymethylpterin (HMDP) is linked by 2, 3, or 4 phosphate groups to an adenosine moiety (Table 1.3). The compound with a tetraphosphate linker showed optimal inhibition. While displaying tighter binding than ATP, it bound less avidly than AMPCPP (Table 1.3). Recently, a new generation of bisubstrate analogue inhibitors was synthesised in which the pterin or 7,7-dimethyl-7,8-dihydropterin moiety is attached with a piperidine linked thioether bridge (Figure 1.8)[53]. The binding affinities (K_d) of compounds **3** and **4** (Figure 1.8) were >150 µM and 2.55 µM respectively. These data indicate that bisubstrate design to achieve higher affinity scaffolds is non-trivial.

Table 1.3: Binding affinities of bisubstrate inhibitors.

n	K_d **Value (μM)** (without Mg^{2+})	K_d **Value (μM)** (with Mg^{2+})	**IC$_{50}$ (μM)**	**Reference**
2	nd	nd	>100	[49]
3	17.4 (1.5)	4.25 (0.28)	1.27	[49]
4	1.93 (0.05)	0.47 (0.04)	0.44	[49]
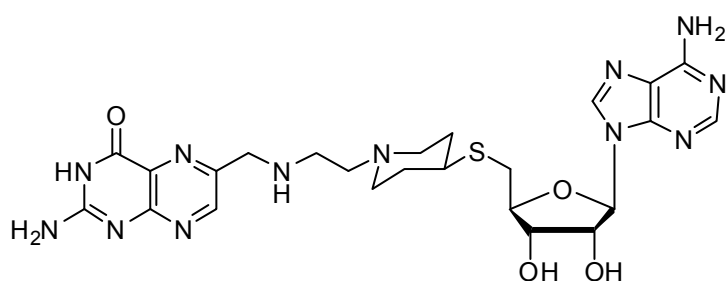
3
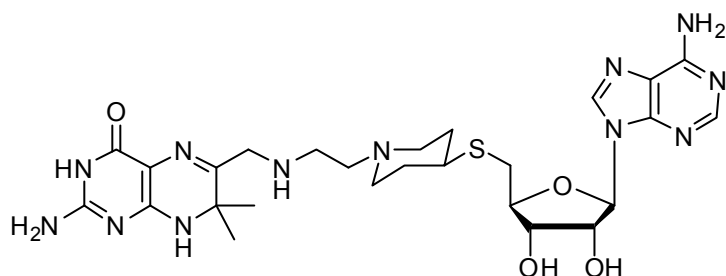
4

Figure 1.8: Bisubstrate analogue HPPK inhibitors reported in the literature.

1.5.5 HPPK structure from *S.aureus*

The structure of HPPK from *Staphylococcus aureus (Sa*HPPK*)* had not been solved, nor the functional aspects of the enzyme investigated, opening opportunities for new research. The *S. aureus* enzyme is homologous to HPPK proteins from other microorganisms. For example, it shares 34-39 % sequence identity with those sequences with deposited structures (Figure 1.9). High identity of active site residues and high structural similarity among all HPPK structures indicates that inhibitors of *Sa*HPPK should inhibit other species. In principal, this suggests that broad-spectrum inhibitors can be developed.
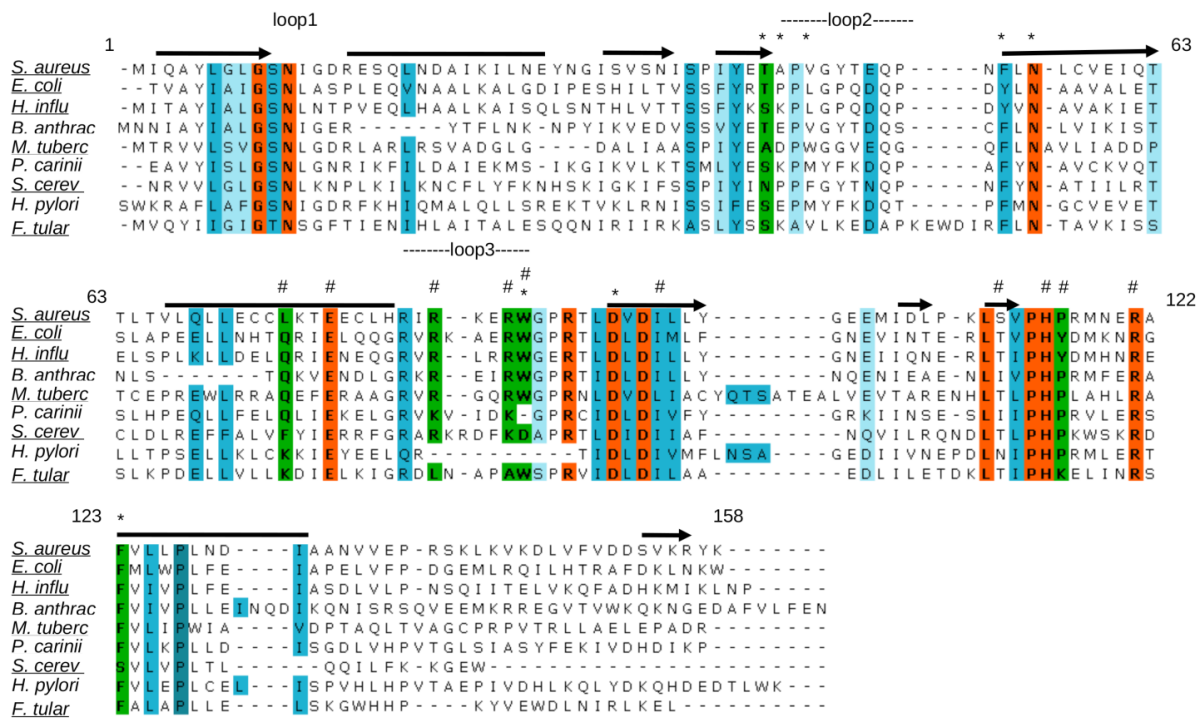
Figure 1.9: Multiple sequence alignment of selected HPPKs (*Staphylococcus aureus*, *Escherichia coli*, *Haemophilus influenza*, *Bacillus anthracis*, *Mycobacterium tuberculosis*, *Pneumocystis carinii*, *Saccharomyces cerevisiae*, *Helicobacter pylori*, *Francisella tularensis*). Absolutely conserved residues (orange), highly conserved residues (blue), similar residues (light blue) and residues that are additionally important for HPPK function and may be targeted to develop selective inhibitors are shown (green). Residues involved in HMDP (*) and ATP recognition (#) and elements of secondary structure are displayed above the alignment. Underlined sequences have had their structure determined by X-ray crystallography.

1.6 Drug discovery

1.6.1 Overview

The drug discovery process involves the identification of a small molecule (or "hit") that binds to a target and modulates the biological activity to provide therapeutic benefits. An iterative process of synthesis, characterisation, screening and assay follows to optimise potency and activity[54, 55]. Currently, the discovery of each new drug costs approximately US$1.8 billion and takes on average 10-15 years[56]. Even with several technological advances, the process of drug discovery is lengthy, expensive and exceedingly difficult, as evidenced by a low rate of new drugs to the market[57].

1.6.2 Target and hit identification strategies

A popular approach to drug discovery is to identify a relevant target with a biological rationale (as a proof-of-concept investigation), then establish its 'druggability', *i.e.* exploration of active site pockets that can accommodate drug-like molecules.

Initial hits are generated in a variety of ways depending on the level of information available. Two popular routes are High-Throughput Screening (HTS) and Fragment-Based Drug Discovery (FBDD), which are distinguished by virtue of the size of the chemical entities and the size of the library. Computational methods can aid hit identification and expedite the hit-to-lead process. *In silico* screening can also provide an independent 'rational' route to hit identification, in which compounds can be simply purchased and then tested. Hit identification can be based on either a pharmacophore model of the ligand (*i.e.* ligand-centric) or a structure of the active site, for example, X-ray data (*i.e.* receptor-centric).

1.6.3 High-throughput screening

HTS is an established high-tech approach widely used in large pharmaceutical companies for hit identification that employs extensive robotic and liquid handling automation. It is basically a process of 'brute force' screening of large libraries, often comprising over a million relatively complex drug-sized compounds (average MW ~400)[58, 59]. Recently, the focused HTS approach has gained much attention. Here, compound libraries are generated based on some chemoinformatic 'descriptors' or by designing or synthesizing target-focused compound libraries.

Identified "hits" are chemically modified using structure-activity relationships (SAR) to develop a "lead"[60]. This "hit-to-lead" process results in a series of compounds showing potent binding (low nM). During the lengthy 'translational' part of the pipeline, leads are optimised with regard to the physicochemical and pharmacokinetic properties (absorption, distribution, metabolism, and excretion–toxicity, ADMET) desirable for the development of candidate compounds for clinical evaluation.

The synergy of HTS with structural and computational methods has added focus to an otherwise random screening approach, increasing the efficiency with which high quality leads can be produced[61]. Nevertheless, HTS is notoriously susceptible to false-positive hits and several filters have been developed to identify and flag problematic compounds[62, 63]. A recent study by GlaxoSmithKline (GSK)[3] in which 67 target-based HTS campaigns were performed using 260,000-530,000 compounds for potential antimicrobial agents at a cost of $1million per HTS resulted in only 5 real lead compounds, revealing that HTS can be a very inefficient and costly approach indeed. Interestingly, in this study the low hit rate was blamed on the incompatibility between typical HTS libraries and the generally hydrophilic nature of antimicrobial targets.

1.6.4 Fragment-based drug discovery (FBDD)

FBDD represents an alternative hit identification method where a much smaller library (as few as 500 compunds) comprising low-molecular-weight compounds (typically 150–250 Da) is screened. Fragments, due to their lower molecular weight, result in more accessible chemical space being represented in a smaller number of compounds[64] than for a diverse high-throughput screening library of drug-like compounds. Additionally, FBDD allows stepwise building of ligands for more challenging targets with larger binding surface area (alleged to be undruggable)[65]. Owing to their small size, fragments bind weakly to a target (K_d ranging from 10 μM to 10 mM).

Measuring the weak binding of fragments is challenging for conventional bioassays owing to interference from the high concentrations of ligand required for significant binding at equilibrium and associated risks of aggregation and/or non-specific binding[66]. Nevertheless, high-throughput crystallography[67], nuclear magnetic resonance (NMR)[68, 69], mass spectroscopy[70], surface plasmon resonance (SPR)[71], isothermal titration calorimetry (ITC)[72] and biochemical screening[73] have proved practical and popular methods for detecting the binding of fragments.

Fragment hits can be subsequently developed into more potent and drug-like molecules by three main strategies: merging (using optimal features of two fragments), linking (joining of two weakly binding fragments), and growing (adding extra functionality) (Figure 1.10). A good fragment, while binding weakly, makes good quality interactions within the binding site, leading to a high binding efficiency per heavy atom (ligand efficiency)[74]. While it has been claimed that ligand efficiency remains relatively constant during the ideal fragment evolution process[75], it serves as more of an initial guide. Recent studies have shown a departure from linearity according to the active site 'shape'[76], *i.e.* the more open the active site, the fewer the interactions possible per heavy atom. Fragments provide optimal starting points for medicinal chemistry. Fragment libraries are frequently designed based on a 'rule of three' (RO3) rather than Lipinski's 'rule of five'[77]. In the RO3, the number of hydrogen bond donors is ≤ 3, the number of hydrogen bond acceptors is ≤ 3,

the ClogP is ≤ 3[78] and as fragments, the molecular weight is < 300 Da. An advantage of FBDD is that as the optimisation process starts early on with a low molecular weight compound in which complexity and physicochemical properties are easier to control and optimise[79] over later stage optimisation of larger HTS leads.
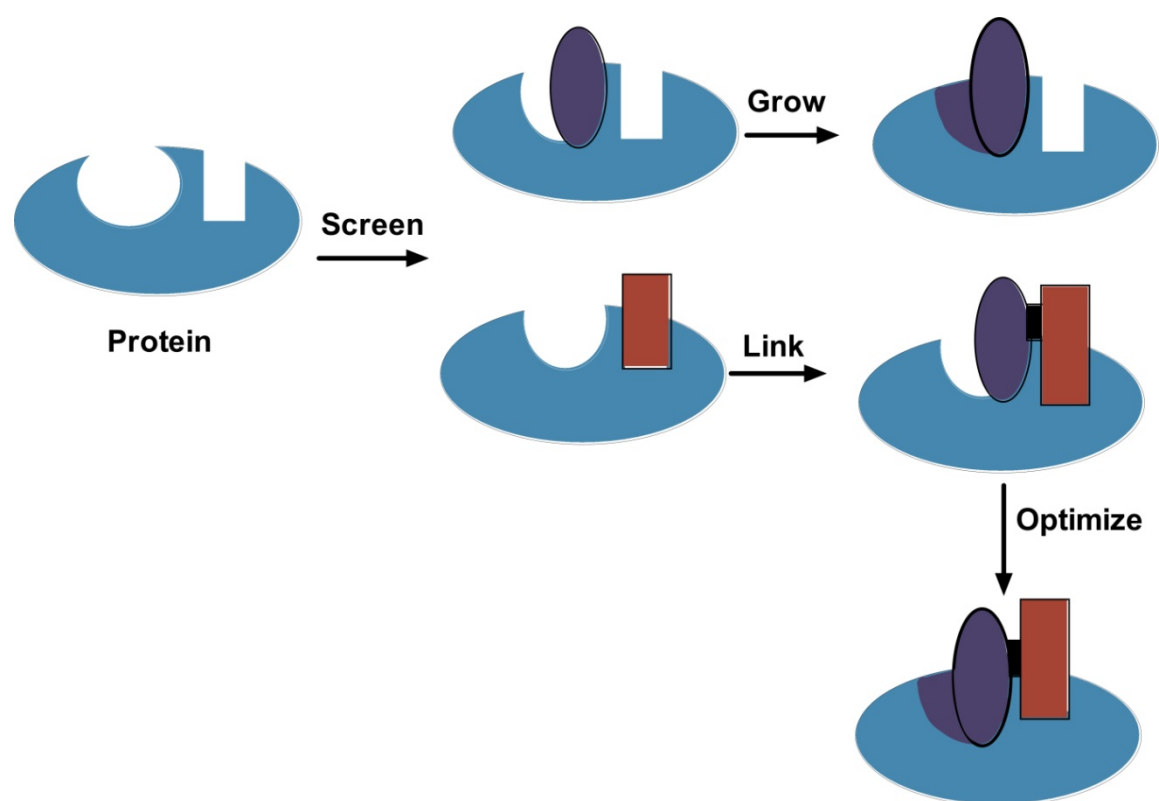
Figure 1.10: Schematic representation of the fragment-based drug discovery process. The blue ovoid and red rectangle represent fragments that bind to the target protein. These can be merged, linked or evolved into high-affinity scaffolds.

1.6.5 High-throughput virtual screening

Virtual screening (VS) is one of the most popular approaches to identify new hits using large *in silico* libraries[80]. The approach falls into two main classes: structure-based and ligand-based VS[81]. Structure-based VS, used in conjunction with the 3D-structure coordinates of the protein (usually from X-ray crystallography), includes docking, 3D pharmacophore design and *de novo* design. Owing to the poor quality of scoring functions and the enormous difficulty in predicting conformational changes upon ligand binding, comparative studies have shown that docking success is highly variable[82]. The alternate and simpler approach is ligand-centric VS, based on the molecular similarity concept, which states that similarly shaped molecules (may have different 2D structures) are more likely to have similar binding modes[83]. Ligand-based VS methods include sub-structure searching, clustering, quantitative structure-activity relationships (QSAR), pharmacophore- and 3D-shape matching techniques[84], such as ROCS[85].

1.6.6 Biophysical methods for hit validation

Once a hit is found using HTS, fragment or *in silico* methods, it is absolutely essential to validate both the binding and activity using a range of robust biophysical techniques. Detection, including quantification of binding events, can be performed using techniques such as Thermal Shift Assay (TSA)[86], Surface Plasmon Resonance (SPR)[71], Isothermal Titration Calorimetry (ITC)[87], and 1D and/or 2D NMR spectroscopy.

Detailed structural information showing the precise binding location and interactions is extremely valuable in further validation and optimisation of hit potency and selectivity. This is most commonly derived from X-ray structures, however NMR structures can play a useful role as the approximate location of binding can be rapidly obtained by 2D heteronuclear NMR spectroscopy which involves the assignment of the backbone ^{15}N and ^{13}C resonances [88]. While an NMR structure of the hit-target complex is possible for small targets (<30 kDa), NMR is generally better suited to further validate the high

resolution data content from the X-ray crystal structure. While each approach suffers from several caveats, a combined approach is very powerful for validation and to further the hit-to-lead process.

1.6.7 Hit confirmation and optimisation

Hit confirmation is an indispensable follow-up step. *In vitro* biochemical or cellular assays are generally used for hit confirmation in the drug discovery process[89]. Hits that demonstrate biological activity are used for the chemical optimisation process (hit-to-lead). The hit-to-lead process involves the integration of medicinal chemistry syntheses with design hypotheses to generate new analogs with improved potency by generating many compounds or focused libraries around the most promising hits to understand the structure-activity relationship (SAR). The addition of different substituents to the hit compound[90] reveals the chemical substituents responsible for affinity 'hot spots', and also identifies parts of the molecule that can be altered without affecting binding, *i.e.* regions that point out of the active site into the solvent. The latter may be useful sites for ADMET modifications.

The iteration of analogue syntheses, quantification and redesign using structure- based approaches enables efficient compound optimisation for lead development[73, 90].

1.7 Summary and potential for new research

The folate biosynthetic pathway presents an attractive target for developing novel antimicrobials because the enzymes from this pathway are essential for microorganisms but are absent in humans and animals. Inhibitors of dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR) enzymes in the folate biosynthetic pathway are currently used as a synergistic combination (SMX-TMP) for treating many infectious diseases. Specific mutations arising in these two enzymes are conferring resistance, rendering the future of these antibiotics uncertain. Most newly developed antibiotics are

new generation chemical modifications of the basic structures of existing drugs against the old targets and thus less effective against widespread antibiotic resistance[9, 91]. Therefore, the development of novel antimicrobial agents against new targets is urgently needed for combating the antibiotic crisis.

HPPK represents a new alternative target from the folate biosynthetic pathway. The SaHPPK enzyme has not been previously studied and may represent a new route to novel antibiotics for MRSA infections. Known inhibitors of HPPK are substrate analogs and are not suitable for development into new antimicrobials due to their poor solubility, poor stability and poor synthetic tractability. The dearth of inhibitors limits the investigation into HPPK as a worthwhile antimicrobial target.

1.8 Scope of the thesis

The work presented herein aimed to investigate for the first time the structure of SaHPPK using NMR and X-ray crystallography (with a view to understand structure-function relationships on the atomic level), and to screen, identify and biophysically characterise the interaction of small molecule inhibitors of SaHPPK function.

Figure 1.11 shows the general process that was employed to elucidate the first structure of SaHPPK in several ligand-bound states and to find new inhibitors of this medically relevant enzyme.

Chapter 2 describes the SaHPPK protein expression, buffer screening, NMR characterisation and crystallisation trials. The application and results from fragment-based screening using the Maybridge RO3 fragment library is discussed. The experimental crystallisation details were published in Acta Crystallographica Section F and directly reproduced and attached as supplementary information.

Chapter 3 has been directly reproduced from a manuscript published in PLoS One and includes additional background data of a technical nature. This

paper reports the use of *in silico* ROCS "scaffold hopping" to find a substrate site inhibitor (8-mercaptoguanine) and the ensuing binding characterisation using various biophysical techniques (TSA, ITC, SPR) and a *in vitro* biochemical assay for the determination of inhibition. A high resolution X-ray structure in complex with a novel substrate site inhibitor is presented and NMR spectroscopy is used to characterise the structure and dynamics of ligand binding.

Chapter 4 presents the SAR data generated based on the chemical synthesis of several analogues of 8-mercaptoguanine. The *Sa*HPPK X-ray structure in complex with one of the synthesised analogues will be instrumental in future lead optimisation.

Chapter 5 outlines the application of a recently developed paramagnetic lanthanide binding tag (LBT) to determine the active site loop structure of *Sa*HPPK.

The findings of this thesis are summarised and reviewed in Chapter 6 which includes future directions.
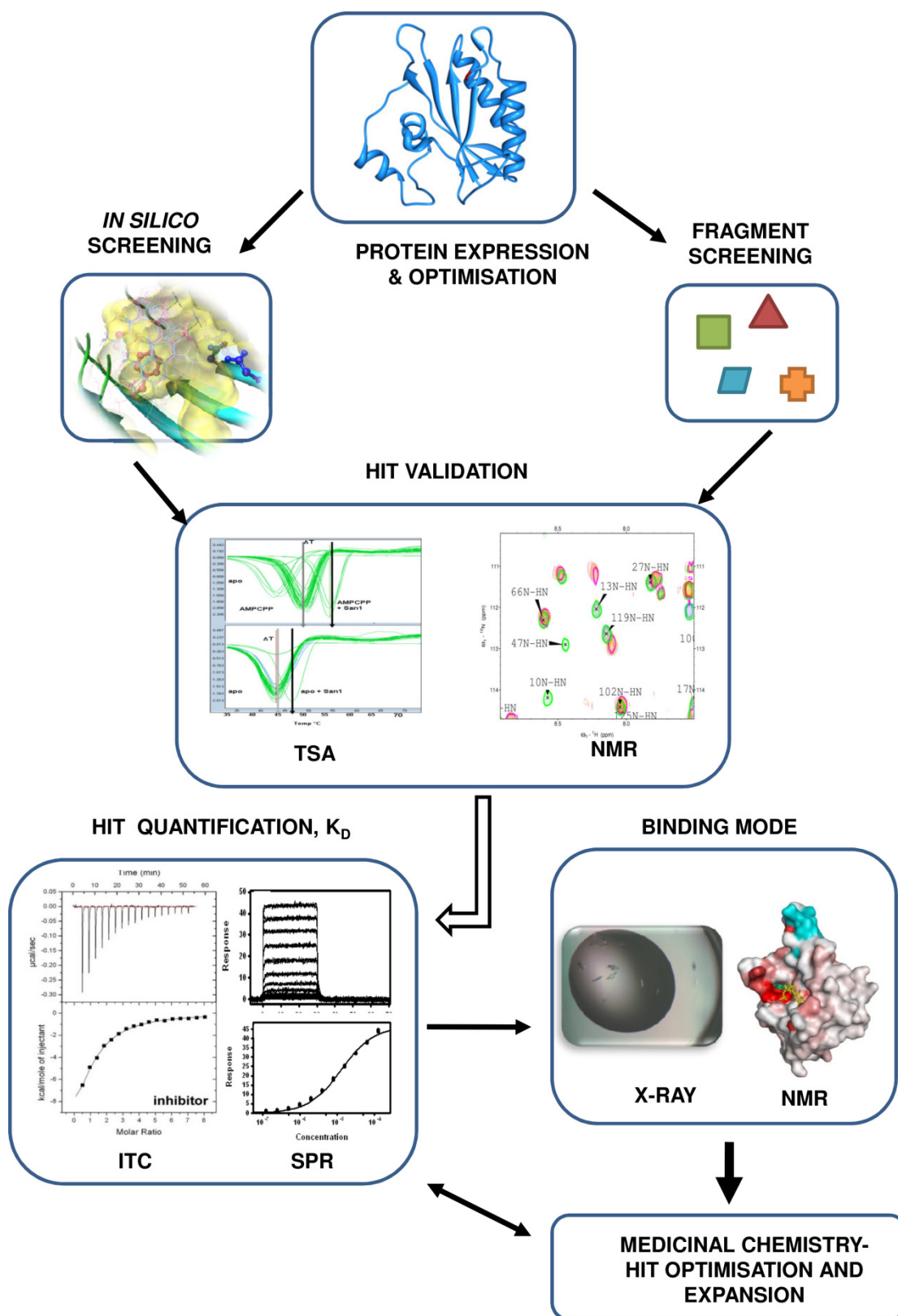
Figure 1.11: The structure-based design protocol employed in this thesis project.

1.9 References

1. Cooper, M. A.; Shlaes, D. Fix the antibiotics pipeline. *Nature* **2011,** 472, 32.

2. Taylor, P. L.; Wright, G. D. Novel approaches to discovery of antibacterial agents. *Anim Health Res Rev* **2008,** 9, 237-46.

3. Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **2007,** 6, 29-40.

4. Barrett, J. F. Can biotech deliver new antibiotics? *Curr Opin Microbiol* **2005,** 8, 498-503.

5. Bax, B. D.; Chan, P. F.; Eggleston, D. S.; Fosberry, A.; Gentry, D. R.; Gorrec, F.; Giordano, I.; Hann, M. M.; Hennessy, A.; Hibbs, M.; Huang, J.; Jones, E.; Jones, J.; Brown, K. K.; Lewis, C. J.; May, E. W.; Saunders, M. R.; Singh, O.; Spitzfaden, C. E.; Shen, C.; Shillings, A.; Theobald, A. J.; Wohlkonig, A.; Pearson, N. D.; Gwynn, M. N. Type IIA topoisomerase inhibition by a new class of antibacterial agents. *Nature* **2010,** 466, 935-40.

6. Fabbretti, A.; Gualerzi, C. O.; Brandi, L. How to cope with the quest for new antibiotics. *FEBS Lett* **2011,** 585, 1673-81.

7. Miller, J. R.; Dunham, S.; Mochalkin, I.; Banotai, C.; Bowman, M.; Buist, S.; Dunkle, B.; Hanna, D.; Harwood, H. J.; Huband, M. D.; Karnovsky, A.; Kuhn, M.; Limberakis, C.; Liu, J. Y.; Mehrens, S.; Mueller, W. T.; Narasimhan, L.; Ogden, A.; Ohren, J.; Prasad, J. V.; Shelly, J. A.; Skerlos, L.; Sulavik, M.; Thomas, V. H.; VanderRoest, S.; Wang, L.; Wang, Z.; Whitton, A.; Zhu, T.; Stover, C. K. A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore. *Proc Natl Acad Sci U S A* **2009,** 106, 1737-42.

8. Davis, A. M.; Keeling, D. J.; Steele, J.; Tomkinson, N. P.; Tinker, A. C. Components of successful lead generation. *Curr Top Med Chem* **2005,** 5, 421-39.

9. Gwynn, M. N.; Portnoy, A.; Rittenhouse, S. F.; Payne, D. J. Challenges of antibacterial discovery revisited. *Ann N Y Acad Sci* **2010,** 1213, 5-19.

10. Fischbach, M. A.; Walsh, C. T. Antibiotics for emerging pathogens. *Science* **2009,** 325, 1089-93.

11. Hidron, A. I.; Low, C. E.; Honig, E. G.; Blumberg, H. M. Emergence of community-acquired meticillin-resistant *Staphylococcus aureus* strain USA300 as a cause of necrotising community-onset pneumonia. *Lancet Infect Dis* **2009,** 9, 384-92.

12. David, M. Z.; Daum, R. S. Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. *Clin Microbiol Rev* **2010,** 23, 616-87.

27

13. Yu, Y. Y.; Wang, H.; Zhang, S. W.; Wang, B. E. Inhibition of methicillin-resistant *Staphylococcus aureus* by the compound Qingre granules. *Chin Med J (Engl)* **2010,** 123, 1017-20.

14. Naimi, T. S.; LeDell, K. H.; Como-Sabetti, K.; Borchardt, S. M.; Boxrud, D. J.; Etienne, J.; Johnson, S. K.; Vandenesch, F.; Fridkin, S.; O'Boyle, C.; Danila, R. N.; Lynfield, R. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. *Jama* **2003,** 290, 2976-84.

15. Adra, M.; Lawrence, K. R. Trimethoprim/sulfamethoxazole for treatment of severe *Staphylococcus aureus* infections. *Ann Pharmacother* **2004,** 38, 338-41.

16. Proctor, R. A. Role of folate antagonists in the treatment of methicillin-resistant *Staphylococcus aureus* infection. *Clin Infect Dis* **2008,** 46, 584-93.

17. Besier, S.; Zander, J.; Siegel, E.; Saum, S. H.; Hunfeld, K. P.; Ehrhart, A.; Brade, V.; Wichelhaus, T. A. Thymidine-dependent *Staphylococcus aureus* small-colony variants: human pathogens that are relevant not only in cases of cystic fibrosis lung disease. *J Clin Microbiol* **2008,** 46, 3829-32.

18. Zander, J.; Besier, S.; Ackermann, H.; Wichelhaus, T. A. Synergistic antimicrobial activities of folic acid antagonists and nucleoside analogs. *Antimicrob Agents Chemother* **2010,** 54, 1226-31.

19. Frey, K. M.; Liu, J.; Lombardo, M. N.; Bolstad, D. B.; Wright, D. L.; Anderson, A. C. Crystal structures of wild-type and mutant methicillin-resistant *Staphylococcus aureus* dihydrofolate reductase reveal an alternate conformation of NADPH that may be linked to trimethoprim resistance. *J Mol Biol* **2009,** 387, 1298-308.

20. Bermingham, A.; Derrick, J. P. The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *Bioessays* **2002,** 24, 637-48.

21. Swarbrick, J. D.; Iliades, P.; Simpson, J. S.; Macreadie, I. Folate biosynthesis - reappraisal of old and novel targets in the search for new antimicrobials. *Open Enzym. Inhib. J.* **2008,** 1, pp. 12-33.

22. Skold, O. Sulfonamide resistance: mechanisms and trends. *Drug Resist Updat* **2000,** 3, 155-160.

23. Derrick, J. P. The structure and mechanism of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Vitam Horm* **2008,** 79, 411-33.

24. Bourne, C. R.; Barrow, E. W.; Bunce, R. A.; Bourne, P. C.; Berlin, K. D.; Barrow, W. W. Inhibition of antibiotic-resistant *Staphylococcus aureus* by the broad-spectrum dihydrofolate reductase inhibitor RAB1. *Antimicrob Agents Chemother* **2010,** 54, 3825-33.

25. Liu, C.; Bayer, A.; Cosgrove, S. E.; Daum, R. S.; Fridkin, S. K.; Gorwitz, R. J.; Kaplan, S. L.; Karchmer, A. W.; Levine, D. P.; Murray, B. E.; M, J. R.; Talan, D. A.; Chambers, H. F. Clinical practice guidelines by the infectious diseases society of america for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children. *Clin Infect Dis* **2011,** 52, e18-55.

26. Robinson, J. L.; Salvadori, M. I. Management of community-associated methicillin-resistant *Staphylococcus aureus* skin abscesses in children. *Paediatr Child Health* **2011,** 16, 115-8.

27. Bermingham, A.; Derrick, J. P. The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *Bioessays* **2002,** 24, 637-48.

28. Klaus, S. M.; Wegkamp, A.; Sybesma, W.; Hugenholtz, J.; Gregory, J. F., 3rd; Hanson, A. D. A nudix enzyme removes pyrophosphate from dihydroneopterin triphosphate in the folate synthesis pathway of bacteria and plants. *J Biol Chem* **2005,** 280, 5274-80.

29. Babaoglu, K.; Qi, J.; Lee, R. E.; White, S. W. Crystal structure of 7,8-dihydropteroate synthase from *Bacillus anthracis*: mechanism and novel inhibitor design. *Structure* **2004,** 12, 1705-17.

30. Li, G.; Felczak, K.; Shi, G.; Yan, H. Mechanism of the conformational transitions in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy. *Biochemistry* **2006,** 45, 12573-81.

31. Blaszczyk, J.; Li, Y.; Cherry, S.; Alexandratos, J.; Wu, Y.; Shaw, G.; Tropea, J. E.; Waugh, D. S.; Yan, H.; Ji, X. Structure and activity of Yersinia pestis 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as a novel target for the development of antiplague therapeutics. *Acta Crystallogr D Biol Crystallogr* **2007,** 63, 1169-77.

32. Garcon, A.; Levy, C.; Derrick, J. P. Crystal structure of the bifunctional dihydroneopterin aldolase/6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Streptococcus pneumoniae*. *J Mol Biol* **2006,** 360, 644-53.

33. Hennig, M.; Dale, G. E.; D'Arcy, A.; Danel, F.; Fischer, S.; Gray, C. P.; Jolidon, S.; Muller, F.; Page, M. G.; Pattison, P.; Oefner, C. The structure and function of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Haemophilus influenzae*. *J Mol Biol* **1999,** 287, 211-9.

34. Lawrence, M. C.; Iliades, P.; Fernley, R. T.; Berglez, J.; Pilling, P. A.; Macreadie, I. G. The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/dihydropteroate synthase of *Saccharomyces cerevisiae*. *J Mol Biol* **2005,** 348, 655-70.

35. Xiao, B.; Shi, G.; Chen, X.; Yan, H.; Ji, X. Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. *Structure* **1999,** 7, 489-96.

36. Pemble, C. W. t.; Mehta, P. K.; Mehra, S.; Li, Z.; Nourse, A.; Lee, R. E.; White, S. W. Crystal structure of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase*dihydropteroate synthase bifunctional enzyme from *Francisella tularensis*. *PLoS One* **2010,** 5, e14165.

37. Blaszczyk, J.; Shi, G.; Li, Y.; Yan, H.; Ji, X. Reaction trajectory of pyrophosphoryl transfer catalysed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Structure* **2004,** 12, 467-75.

38. Yang, R.; Lee, M. C.; Yan, H.; Duan, Y. Loop conformation and dynamics of the *Escherichia coli* HPPK apo-enzyme and its binary complex with MgATP. *Biophys J* **2005,** 89, 95-106.

39. Blaszczyk, J.; Shi, G.; Yan, H.; Ji, X. Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 A resolution. *Structure* **2000,** 8, 1049-58.

40. Xiao, B.; Shi, G.; Gao, J.; Blaszczyk, J.; Liu, Q.; Ji, X.; Yan, H. Unusual conformational changes in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by X-ray crystallography and NMR. *J Biol Chem* **2001,** 276, 40274-81.

41. Blaszczyk, J.; Li, Y.; Wu, Y.; Shi, G.; Ji, X.; Yan, H. Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* **2004,** 43, 1469-77.

42. Blaszczyk, J.; Li, Y.; Shi, G.; Yan, H.; Ji, X. Dynamic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: crystallographic studies. *Biochemistry* **2003,** 42, 1573-80.

43. Li, Y.; Wu, Y.; Blaszczyk, J.; Ji, X.; Yan, H. Catalytic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: site-directed mutagenesis and biochemical studies. *Biochemistry* **2003,** 42, 1581-8.

44. Bermingham, A.; Bottomley, J. R.; Primrose, W. U.; Derrick, J. P. Equilibrium and kinetic studies of substrate binding to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Escherichia coli*. *J Biol Chem* **2000,** 275, 17962-7.

45. Li, Y.; Gong, Y.; Shi, G.; Blaszczyk, J.; Ji, X.; Yan, H. Chemical transformation is not rate-limiting in the reaction catalysed by *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* **2002,** 41, 8777-83.

46. Shi, G.; Gong, Y.; Savchenko, A.; Zeikus, J. G.; Xiao, B.; Ji, X.; Yan, H. Dissecting the nucleotide binding properties of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase with fluorescent 3'(2)'-o-anthraniloyladenosine 5'-triphosphate. *Biochim Biophys Acta* **2000,** 1478, 289-99.

47. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **1995,** 8, 127-34.

48. Garcon, A.; Bermingham, A.; Lian, L. Y.; Derrick, J. P. Kinetic and structural characterisation of a product complex of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Escherichia coli*. *Biochem J* **2004,** 380, 867-73.

49. Shi, G.; Blaszczyk, J.; Ji, X.; Yan, H. Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: synthesis and biochemical and crystallographic studies. *J Med Chem* **2001,** 44, 1364-71.

50. Talarico, T. L.; Dev, I. K.; Dallas, W. S.; Ferone, R.; Ray, P. H. Purification and partial characterisation of 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase and 7,8-dihydropteroate synthase from *Escherichia coli* MC4100. *J Bacteriol* **1991,** 173, 7029-32.

51. Wood, H. C. S. Specific inhibition of dihydrofolate biosynthesis- a new approach to chemotherapy. *In Chemistry and Biology of Pteridines* **1975,** , Walter de Gruyter, Berlin-New York.

52. Stammers, D. K.; Achari, A.; Somers, D. O.; Bryant, P. K.; Rosemond, J.; Scott, D. L.; Champness, J. N. 2.0 A X-ray structure of the ternary complex of 7,8-dihydro-6-hydroxymethylpterinpyrophosphokinase from *Escherichia coli* with ATP and a substrate analogue. *FEBS Lett* **1999,** 456, 49-53.

53. Shi, G.; Shaw, G.; Liang, Y. H.; Subburaman, P.; Li, Y.; Wu, Y.; Yan, H.; Ji, X. Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: New design with improved properties. *Bioorg Med Chem* **2012,** 20, 47-57.

54. Cramer, J. W.; Mattioni, B. E.; Savin, K. A. Strategies for conducting ADME studies during lead generation in the drug discovery process. *IDrugs* **2010,** 13, 857-61.

55. Gao, Q.; Yang, L.; Zhu, Y. Pharmacophore based drug design approach as a practical process in drug discovery. *Current computer-aided drug design* **2010,** 6, 37-49.

56. Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **2010,** 9, 203-14.

57. Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* **2007,** 6, 881-90.

58. Hann, M. M.; Oprea, T. I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* **2004,** 8, 255-63.

59. Schonbrun, E.; Abate, A. R.; Steinvurzel, P. E.; Weitz, D. A.; Crozier, K. B. High-throughput fluorescence detection using an integrated zone-plate array. *Lab chip* **2010,** 10, 852-6.

60. Astle, J. M.; Simpson, L. S.; Huang, Y.; Reddy, M. M.; Wilson, R.; Connell, S.; Wilson, J.; Kodadek, T. Seamless bead to microarray screening: rapid identification of the highest affinity protein ligands from large combinatorial libraries. *Chem Biol* **2010,** 17, 38-45.

61. Balakin, K. V.; Savchuk, N. P. Computational methods for analysis of high-throughput screening data. *Curr Comput Aided Drug Des* **2006, ,** 2., 1-19.

62. Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* **2010,** 53, 2719-40.

63. Feng, B. Y.; Simeonov, A.; Jadhav, A.; Babaoglu, K.; Inglese, J.; Shoichet, B. K.; Austin, C. P. A high-throughput screen for aggregation-based inhibition in a large compound library. *J Med Chem* **2007,** 50, 2385-90.

64. Fink, T.; Bruggesser, H.; Reymond, J. L. Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew Chem Int Ed Engl* **2005,** 44, 1504-8.

65. Coyne, A. G.; Scott, D. E.; Abell, C. Drugging challenging targets using fragment-based approaches. *Curr Opin Chem Biol* **2010,** 14, 299-307.

66. Hubbard, R. E. Fragment approaches in structure-based drug discovery. *J SYNCHROTRON RADIAT* **2008,** 15, 227-30.

67. Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discov Today* **2002,** 7, 522-7.

68. Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev* **2004,** 104, 3641-76.

69. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** 274, 1531-4.

70. Swayze, E. E.; Jefferson, E. A.; Sannes-Lowery, K. A.; Blyn, L. B.; Risen, L. M.; Arakawa, S.; Osgood, S. A.; Hofstadler, S. A.; Griffey, R. H. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J Med Chem* **2002,** 45, 3816-9.

71. Giannetti, A. M. From experimental design to validated hits a comprehensive walk-through of fragment lead identification using surface plasmon resonance. *Methods Enzymol* **2011,** 493, 169-218.

72. Ciulli, A.; Williams, G.; Smith, A. G.; Blundell, T. L.; Abell, C. Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods. *J Med Chem* **2006,** 49, 4992-5000.

73. Card, G. L.; Blasdel, L.; England, B. P.; Zhang, C.; Suzuki, Y.; Gillette, S.; Fong, D.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nat Biotechnol* **2005,** 23, 201-7.

74. Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* **2004,** 9, 430-1.

75. Hajduk, P. J. Fragment-based drug design: how big is too big? *J Med Chem* **2006,** 49, 6972-6.

76. Reynolds, C. H.; Tounge, B. A.; Bembenek, S. D. Ligand binding efficiency: trends, physical basis, and implications. *J Med Chem* **2008,** 51, 2432-8.

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

78. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **2003,** 8, 876-7.

79. Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **2007,** 6, 211-9.

80. Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today* **2006,** 11, 1046-53.

81. Lopez-Vallejo, F.; Caulfield, T.; Martinez-Mayorga, K.; Giulianotti, M. A.; Nefzi, A.; Houghten, R. A.; Medina-Franco, J. L. Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Comb Chem High Throughput Screen* **2011,** 14, 475-87.

82. Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* **2007,** 50, 74-82.

83. Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B. T. Molecular similarity and diversity in chemoinformatics: from theory to applications. *Mol Divers* **2006,** 10, 39-79.

84. Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. Free resources to assist structure-based virtual ligand screening experiments. *Curr Protein Pept Sci* **2007,** 8, 381-411.

85. Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J Med Chem* **2005,** 48, 1489-1495.

86. Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.;

Salemme, F. R. High-density miniaturised thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **2001,** 6, 429-440.

87. Velazquez-Campoy, A.; Ohtaka, H.; Nezami, A.; Muzammil, S.; Freire, E. Isothermal titration calorimetry. *Curr Protoc Cell Biol* **2004,** Chapter 17, Unit 17 8.

88. Wagner, G. An account of NMR in structural biology. *Nat Struct Biol* **1997,** 4 Suppl, 841-4.

89. Keseru, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discov Today* **2006,** 11, 741-8.

90. Neumann, T.; Junker, H. D.; Schmidt, K.; Sekul, R. SPR-based fragment screening: advantages and applications. *Curr Top Med Chem* **2007,** 7, 1630-42.

91. Woodford, N.; Wareham, D. W. Tackling antibiotic resistance: a dose of common antisense? *J Antimicrob Chemother* **2009,** 63, 225-9.

Chapter 2

Protein Characterisation and

Fragment Screening

Declaration for Thesis Chapter 2

Chapter 2 presents the published article (Chhabra *et al.* 2012) as supporting information. The paper is co-authored with Janet Newman, Thomas S. Peat, Ross Fernley, Joanne Caine, Jamie S. Simpson, and James D. Swarbrick*.

Declaration by candidate
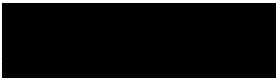
In the case of Chapter 2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Protein expression and purification, Set-up crystallisation experiments, data analysis	70

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Janet Newman[2]	X-ray experiments, intellectual input.	
Thomas S. Peat[2]	X-ray experiments, intellectual input.	
Ross Fernley[2]	Protein purification, intellectual input.	
Joanne Caine[2]	Protein expression, intellectual input.	
Jamie S. Simpson[1]	Thermal shift assay, intellectual input.	
James D.Swarbrick[1]	Construct design and Enzyme stability, intellectual input.	

Candidate's Signature	███████	Date 05-04-2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	[1]Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia.
	[2]CSIRO Division of Materials, Science and Engineering, 343 Royal Parade, Parkville, Victoria 3052, Australia.

		Date
Signature 1		04/04/2012
Signature 2		03/04/2012
Signature 3		3/04/2012
Signature 4		3/4/12
Signature 5		4/4/12
Signature 6		

2. Protein Characterisation and Fragment Screening

2.1 Introduction

Starting from a synthetic *Sa*HPPK gene that was purchased from Geneart, the optimisation of the expression and purification of recombinant *Sa*HPPK for structural studies (crystallography and NMR) is described. A range of biophysical techniques were used for characterisation of ligand binding events and *Sa*HPPK function *in vitro.*

While the *Sa*HPPK protein expressed in high yield in enriched media, the enzyme suffered from poor stability in solution. Buffer screening experiments were used to find optimised buffer conditions for the enzyme. A modified minimal-media protocol was used to express uniformly labeled ^{15}N and ^{13}C *Sa*HPPK for NMR spectroscopy.

Spectra of the ^{15}N labelled *Sa*HPPK protein showed that the enzyme was folded, which gave a high level of confidence to proceed with all further X-ray structural and fragment screening methods. Triple resonance-based NMR chemical shift assignments of the backbone resonances of *Sa*HPPK were obtained as a site-specific fingerprint to follow the amide chemical shift changes upon ligand binding and to map these onto the structure. From this, the binding site location of small ligands (substrate and cofactor) was identified by monitoring the change in the amide chemical shifts or perturbations (CSPs) in the protein NMR spectra during ligand titrations.

A fragment screen was performed using the Maybridge RO3 fragment library ('rule of three' compliant)[1], to identify structurally diverse binders for either the substrate or cofactor sites. The CSPs were used to map the location of bound fragments and their binding affinities were assessed using SPR. X-ray crystallisation experiments of the *apo* enzyme were investigated in parallel, with a view to obtaining enzyme crystals for soaking of initial hits derived either from fragment or *in silico* screening and thereby providing a route to rapidly delineate intermolecular interactions.

2.2 Results and Discussion

2.2.1 Protein stability was an initial problem

The plasmid vector (pET-28a) containg SaHPPK gene sequence with an N-terminal His_6 tag was transformed into $E.$ $coli$ cells, expressed and purified using metal affinity chromatography, giving a yield of >50 mg of purified protein per litre of media. SaHPPK was found to be unstable and aggregated, as evidenced by substantial precipitation within mins when exposed to the air under the purification conditions (50 mM Tris.HCl, 300 mM NaCl, pH 8.5). Two methods, thermal shift denaturation assay[2] and button tests[3], were used to screen for appropriate conditions to overcome initial solubility problems with SaHPPK. For NMR acquisition it was necessary to identify conditions under which samples would be stable for at least 48 h at 22 °C.

2.2.2 Thermal Shift denaturation Assay (TSA) – a convenient and robust assay

A fluorescence-based thermal shift assay was used as a general high-throughput approach to identify conditions of increased thermal stability, with the expectation that these conditions would correspond to an increase in sample longevity and would reduce the propensity of SaHPPK to aggregate. This assay measures protein unfolding, with respect to temperature, by monitoring the binding of a fluorescent dye to hydrophobic regions of proteins. The dye (Sypro Orange), used in the assay, is highly fluorescent in regions of low dielectric constants (hydrophobic regions)[4] and is quenched in an aqueous medium. Therefore, when a protein unfolds, the hydrophobic regions become exposed to which the dye binds resulting in a large increase in fluorescence[5] which is recorded. From the thermal melting curve the midpoint unfolding transition temperature (T_m), a measure of protein stability can be determined. In TSA, a higher T_m value corresponds to an increase in the thermal stability under the particular buffer conditions being tested.

A dilute (5 μM) sample of *Sa*HPPK was initially screened in different buffer solutions covering a wide pH range. Maximum stability of the protein within the pH range suitable for NMR measurements (*i.e.* generally less than pH 8-8.5), was observed at pH 8.0 (Figure 2.1). Aggregation, as judged by the poor fluorescence profile and the tendency to precipitate, was found to increase in the presence of salt (NaCl).

The effect of several additives was also explored. Notably, glycerol and sorbitol were the only additives that significantly stabilised the protein. Sorbitol is preferred over glycerol owing to its lower viscosity[6], which should result in less line broadening in NMR experiments. Sorbitol has also been used in conjunction with alignment media for the measurement of residual dipolar couplings in NMR experiments[6], which can help in characterising the *Sa*HPPK structure.
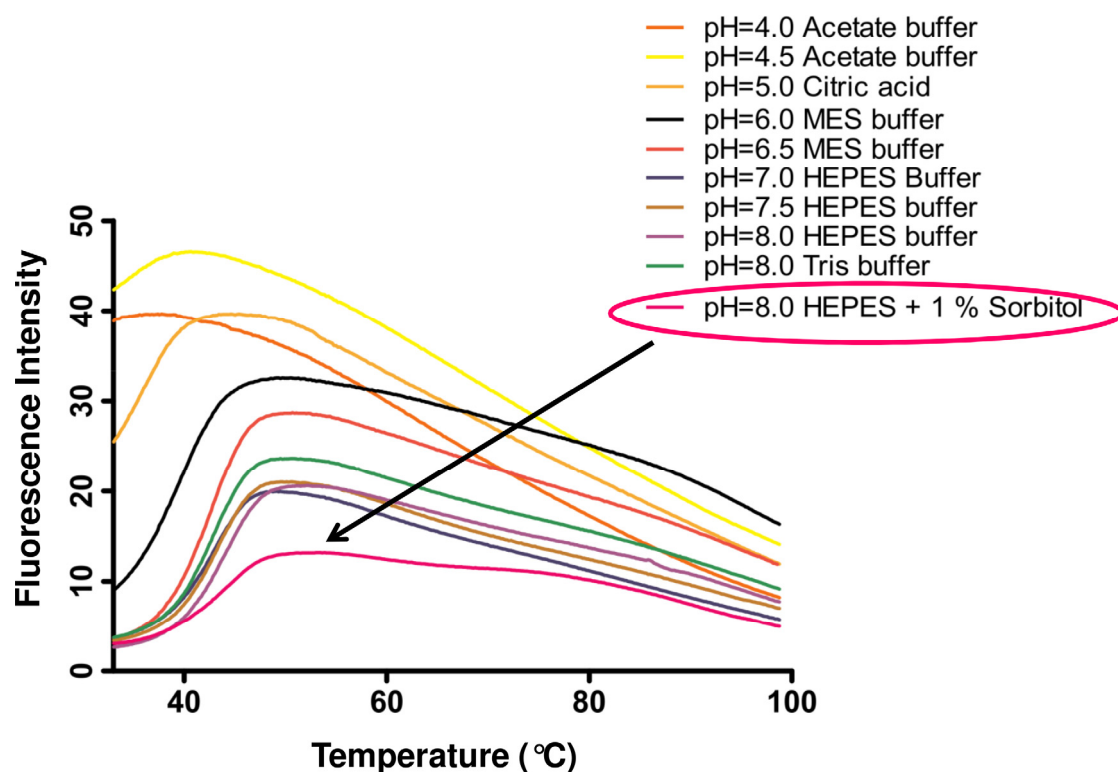
Figure 2.1: Selected examples of thermal shift profiles observed during the buffer screening for *Sa*HPPK using thermal shift assay. A higher T_m value corresponds to higher stability. Note that the variations in fluorescence intensity could be due to different quenching/behaviour of the Sypro Orange dye under different buffer conditions.

2.2.3 Microdialysis button tests

A button test[3] was used to determine the buffer conditions for increased protein longevity at higher concentrations (100-250 µM), which are typically needed for sensitivity and stability during lengthy (~48 h) 3D NMR experiments. The principal of microdialysis button screens is similar to that of X-ray crystal trials, with the difference being that the former screens for clear drops and the latter for precipitation and crystals. Details are described in the methods section 2.8.12. Optimal buffer conditions determined via both the techniques correlated well, with *Sa*HPPK most stable in a buffer comprising 50 mM HEPES (pH 8.0), 1% sorbitol and 2.0 mM DTT. All subsequent studies were performed using this buffer, except X-ray trials, for which sorbitol was omitted.

2.2.4 Thermal Shift denaturation Assay (TSA) for ligand binding

The thermal shift assay was also used to investigate the effect of ligand binding on protein stability. Ligands are known to be able to provide enhanced stability to proteins upon binding, which should produce a change in T_m value associated with unfolding of the protein[2]. The utility of the assay was examined by testing ATP and AMPCPP binding. The *Sa*HPPK enzyme showed higher thermal stability in the presence of both ATP ($\Delta T_m = 6$ °C) and AMPCPP ($\Delta T_m = 11$ °C) (Figure 2.2). This data suggest that the stability of *Sa*HPPK in solution can be further enhanced by adding ligands when running longer experiments.
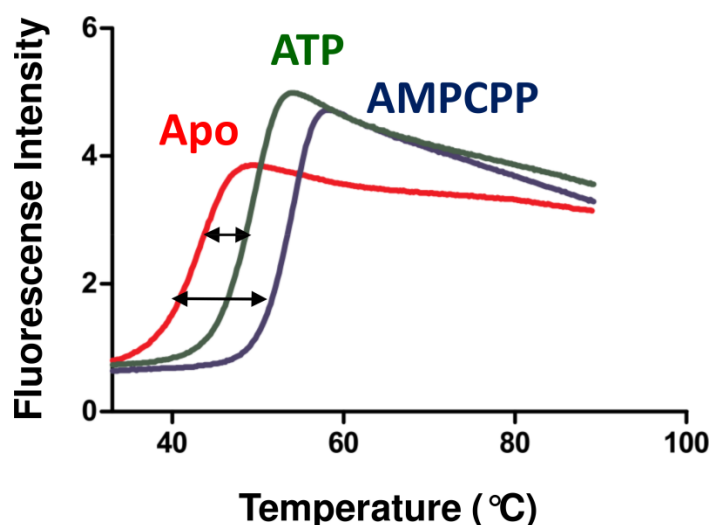
Figure 2.2: Thermal shift observed (double headed arrow) in the presence of ATP (green) and AMPCPP (blue) relative to the *apo* enzyme (red).

2.2.5 Biochemical assay

In order to investigate the activity of the purified *Sa*HPPK enzyme, an enzyme-coupled *in vitro* assay was developed to monitor substrate kinetics and enzyme inhibition. The assay is an extension of the coupled enzyme spectrophotometric assay previously reported by Fernley *et al.*[7], which involves three enzymes of the folate pathway – HPPK, DHPS and DHFR (Figure 2.3). Together, these transform the HPPK substrate, 6-hydroxymethyl-7,8-dihydropterin (HMDP), into tetrahydropteroate. The last reaction converts NADPH to NADP$^+$, which is monitored spectrophotometrically at 340 nm. DHPS and DHFR are added in excess so that the measured rate represents that of the enzyme HPPK.

Enzyme activity was thus assayed by following the decrease in absorbance at 340 nm resulting from the conversion of NADPH to NADP$^+$, and subsequently kinetic parameters for the cofactor (ATP) were obtained (K_m^{app} = 8.9 ± 3 µM) using a standard protocol (Figure 2.4). Measurement of the K_m for the substrate HMDP was attempted, however only an approximate value could be obtained (K_m^{app} ~2.5 ± 1.5 µM) as the change in absorbance of NADPH

during the initial phase of catalysis was close to the noise level (data not shown). A more sensitive assay, based on KinaseGlo[®] was thus developed (see chapter 3). An IC_{50} value of 1.64 ± 0.5 µM for the inhibition of SaHPPK by AMPCPP was determined (Figure 2.4), which is close to the value reported for EcHPPK (K_i of 0.31 µM)[8].
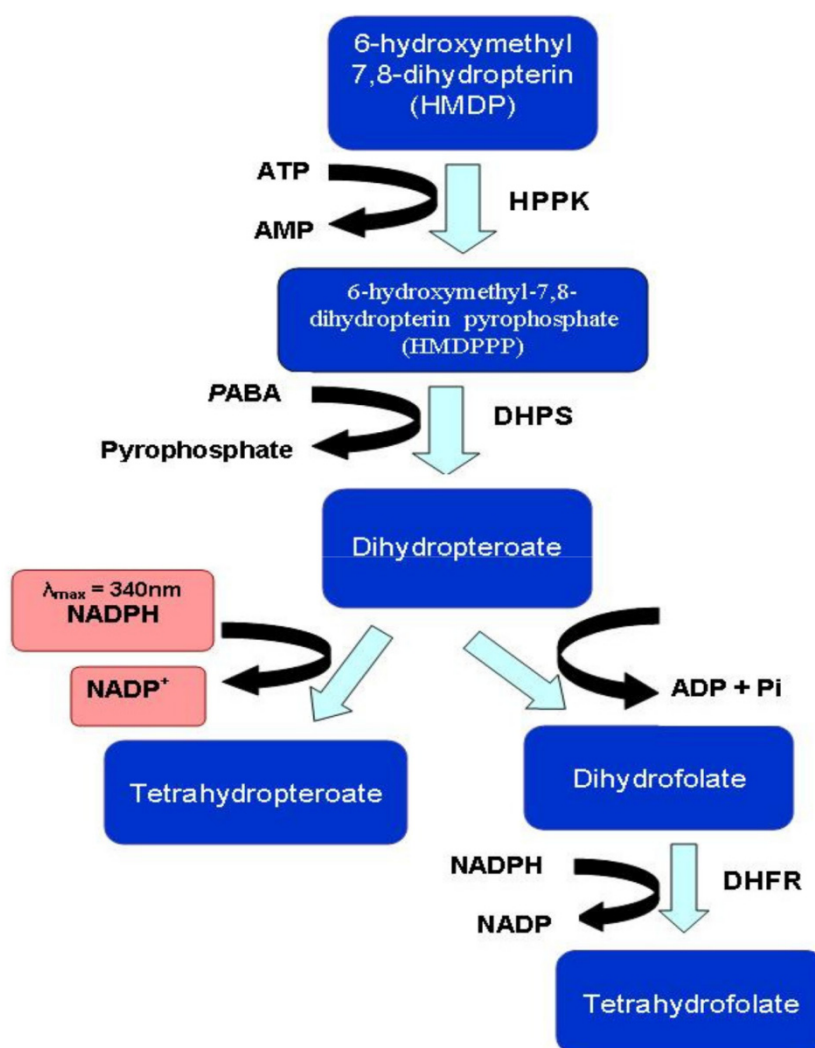
Figure 2.3: Enzyme reactions from the folate biosynthetic pathway involved in the in vitro biochemical assay.
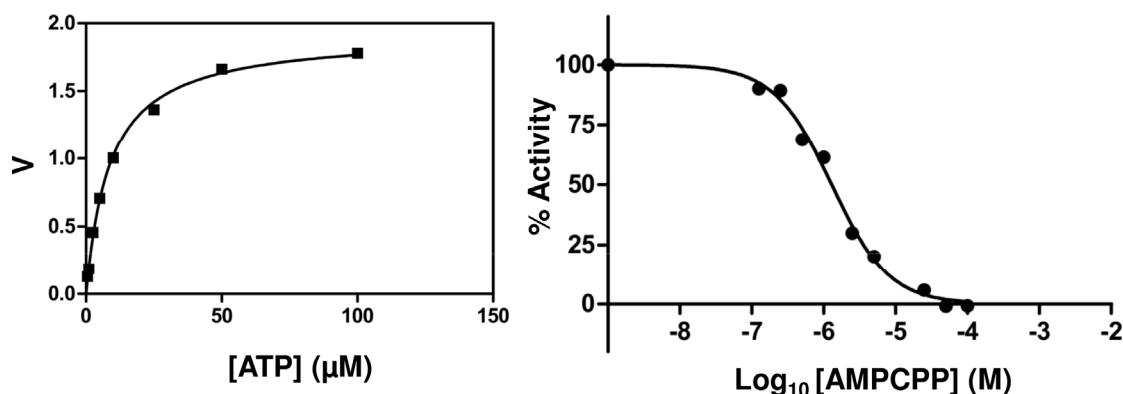
Figure 2.4: (Left) Determination of K_m^{app} for the ATP cofactor**.** The value for K_m^{app} was obtained by fitting the Michaelis-Menten equation to the experimental data using *GraphPad Prism*[9]. The kinetic parameters for ATP were determined by measuring the rate of HPPK-catalysed ATP consumption at the indicated levels of ATP concentrations and a fixed, saturating level of 100 µM HMDP substrate. The K_m^{app} value of 8.9 ± 3 µM was obtained for ATP. (Right) The dose response curve is shown for the inhibition of *Sa*HPPK by AMPCPP yielding an IC_{50} = 1.64 ± 0.5 µM.

2.3 Heteronuclear NMR spectroscopic analysis of *Sa*HPPK

2.3.1 Production of $^{13}C/^{15}N$-labeled protein for NMR experiments

Initial attempts to express isotopically-labeled ^{15}N *Sa*HPPK following the standard Marley method[10] failed, as *Sa*HPPK expressed entirely in the insoluble fraction from the cell pelet.   Several small volume (10 mL) expressions were therefore set up to screen a range of factors, including different temperatures for expression (30–37 °C) and induction (20–37°C), different concentration of IPTG (0.1–1 mM), varying induction time (5–24 hours) and changing cell lines (Rosetta, C41). All these test expressions were unsuccessful as *Sa*HPPK continued to express as inclusion bodies. Finally, the Marley method was replaced with a modified protocol from Lewis Kay's lab[11] as described in section 2.8.8.   In this way, 40–50 mg/L of soluble *Sa*HPPK were routinely obtained. While some higher order aggregates were

41

always observed after passage through the metal affinity column, a subsequent gel filtration step was implemented to remove these impurities prior to all studies. Samples were stored frozen at -80 °C or sealed under nitrogen in glass vials.

2.3.2 The ^{15}N SOFAST HMQC (band-selective optimized flip-angle short-transient heteronuclear multiple quantum coherence) spectrum of *Sa*HPPK showed a folded enzyme

^{15}N SOFAST HMQC spectra were of high quality (Figure 2.5) and the enzyme was stable for 2-3 days as judged by an unchanged spectrum when sealed under nitrogen in the NMR buffer (50 mM HEPES. pH 8.0, 1% sorbitol and 2–10 mM DTT, 10% D_2O) at a concentration <250 µM. ^{15}N SOFAST HMQC spectra of the *apo* enzyme showed a well dispersed set of resonances, consistent with a folded enzyme. However, the enzyme aggregated at high concentration (>250 µM), as judged by severe line broadening. Similar findings have been reported for the *E. coli* enzyme[12]. Line broadening was more evident when the samples were exposed to air and suggests that the aggregation involves disulfide formation through one or more of the four cysteine residues.
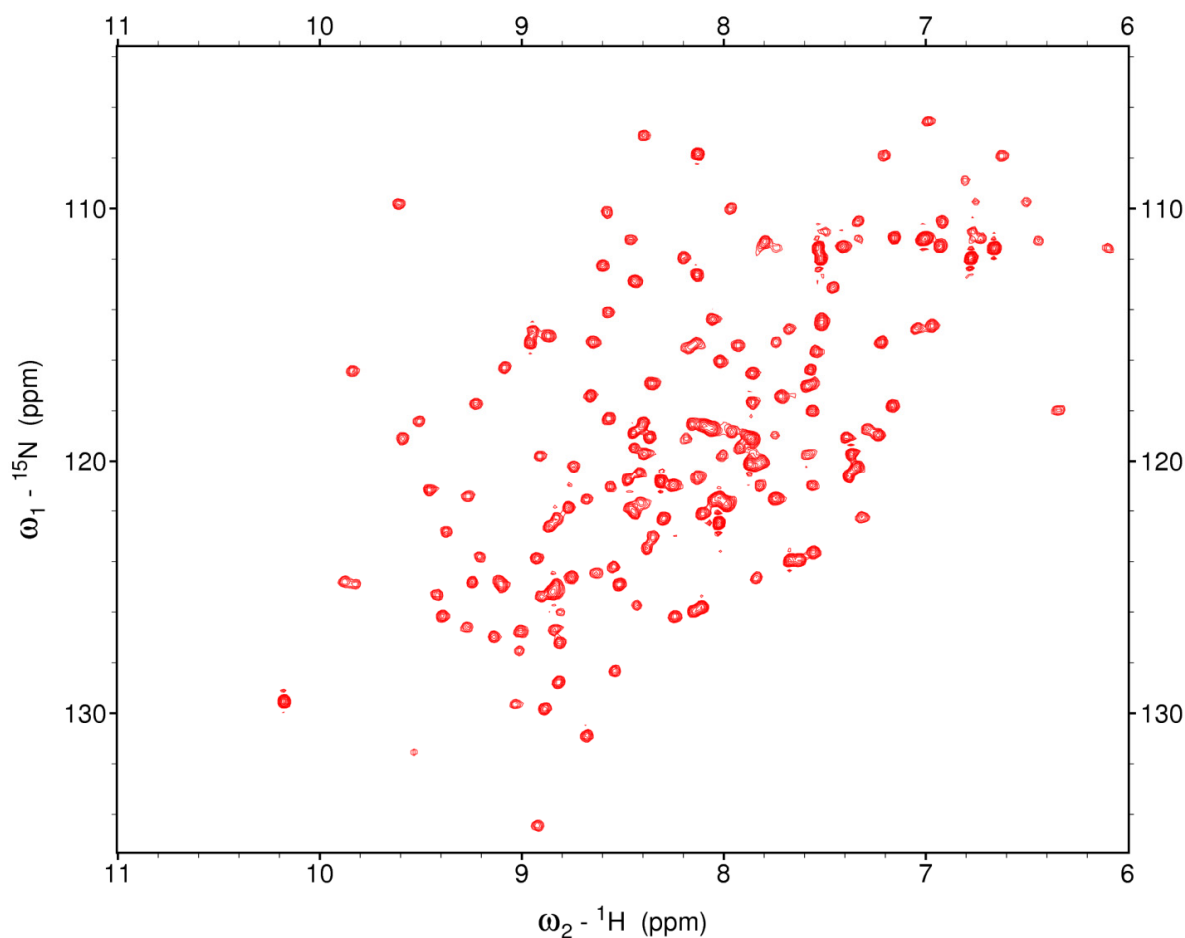
Figure 2.5: The 600 MHz ^{15}N SOFAST HMQC spectra of *apo Sa*HPPK (*ca.* 150 μM) in the presence of 10 mM MgCl$_2$ at 22 °C.

NMR titrations with substrate (HMDP), cofactor (ATP) or cofactor analogue (AMPCPP) produced large chemical shift perturbations (Figure 2.6). ATP and AMPCPP resonances were found to be in slow exchange with the *apo* resonances on the chemical shift time-scale. Therefore, to allow the mapping of these perturbations onto the structure, the protein backbone triple resonance experiments and assignments were required for each of these samples, as resonances could not be simply 'tracked' and assigned as for ligand binding in the 'fast exchange' chemical shift regime.
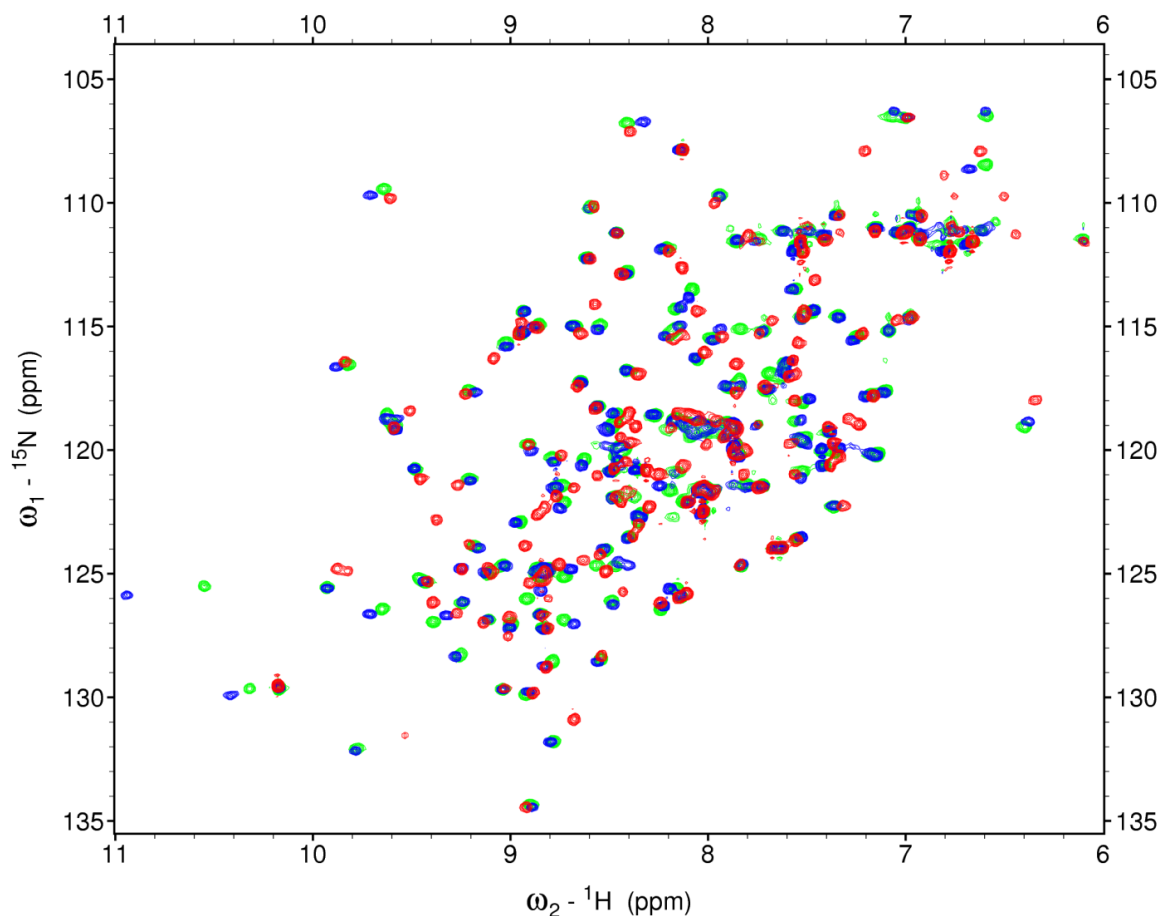
Figure 2.6: Superposition of the 600 MHz ^{15}N SOFAST HMQC spectra of *Sa*HPPK (*ca.* 150 µM) (red) in the presence of ATP (*ca.* 500 µM) (green) and AMPCPP (*ca.* 500 µM) (blue) at 22 °C.

2.3.3 Sequential backbone assignments using 3D triple resonance spectroscopy

Residue specific sequential assignments of the backbone resonances are a prerequisite to ligand binding studies. The sequential assignment gives access to a range of useful ^{15}N-H mediated NMR parameters, such as residual dipolar couplings or pseudo-contact shifts for 'modern' structural studies, as well laying the basis for the assignment of other side chain protons and NOE experiments, required for conventional structure determination. Backbone assignments further facilitate the measurement of relaxation parameters required to understand the dynamics of individual residues within a protein, including dynamic changes upon the binding of ligands.

44

A pair of complementary 3D HNCACB and CBCA(CO)NH experiments [13] were initially performed on the AMPCPP-bound sample owing to the higher thermal stability compared to the *apo* enzyme (Figure 2.2).

The HNCACB spectrum correlates the ^{13}CA and ^{13}CB chemical shift with the HN of the *i* and the HN of the *i-1* residue whereas the CBCA(CO)NH experiment correlates the HN with only the *i-1* ^{13}CA and ^{13}CB shifts. Combining the two 3D spectra therefore allows the intra- and inter-residue ^{13}CA and ^{13}CB chemical shifts to be distinguished (Figure 2.7). Correlations in the 3D spectra are thus found in the ^{13}C dimension for every HN correlation in the 2D ^{15}N HSQC experiment, giving rise to a 3D cube with axes ^{1}HN, ^{15}N and ^{13}C chemical shift. Information within this 3D cube can be presented as 2D strips (Figure 2.7) through the centre of the 2D ^{15}N HSQC peak of every amide, and show the ^{13}C resonances in the third dimension. The 2D strips can be sorted in order of sequence by matching the ^{13}CA and ^{13}CB chemical shifts as shown in Figure 2.7. Knowledge of the ^{13}CB chemical shifts in particular is very useful for identifying specific residue types[13], e.g. the threonine ^{13}CB chemical shift appears around ~70 ppm (Figure 2.7). Thus, from a connected series of 2D strips (as few as three), the identified ^{15}N, ^{1}H, ^{13}CA and ^{13}CB chemical shifts can be assigned to the atoms from residues in a given protein sequence. In this manner, the assignment of the ^{1}HN and ^{15}N peaks in the ^{15}N HSQC experiment was carried out for all non-proline atoms by virtue of the ^{13}CA and ^{13}CB chemical shifts. It was notable that many of the weaker *i-1* ^{13}CB correlations were absent (e.g. Leu68 in Figure 2.7) and some were observed in the CBCA(CO)NH spectrum. Most of the connections were thus performed using the HNCA spectrum and further verified using the HNCACB spectrum where ^{13}CB connectivities were observed.

From the assigned chemical shifts of the AMPCPP-bound sample, the unstable *apo* ^{15}N HSQC spectrum was assigned using 3D HNCA experiments in conjunction with a 3D ^{15}N NOESY experiment. The assignment of the HMDP-bound spectra followed from the *apo* spectrum, as binding mostly induced local broadening.
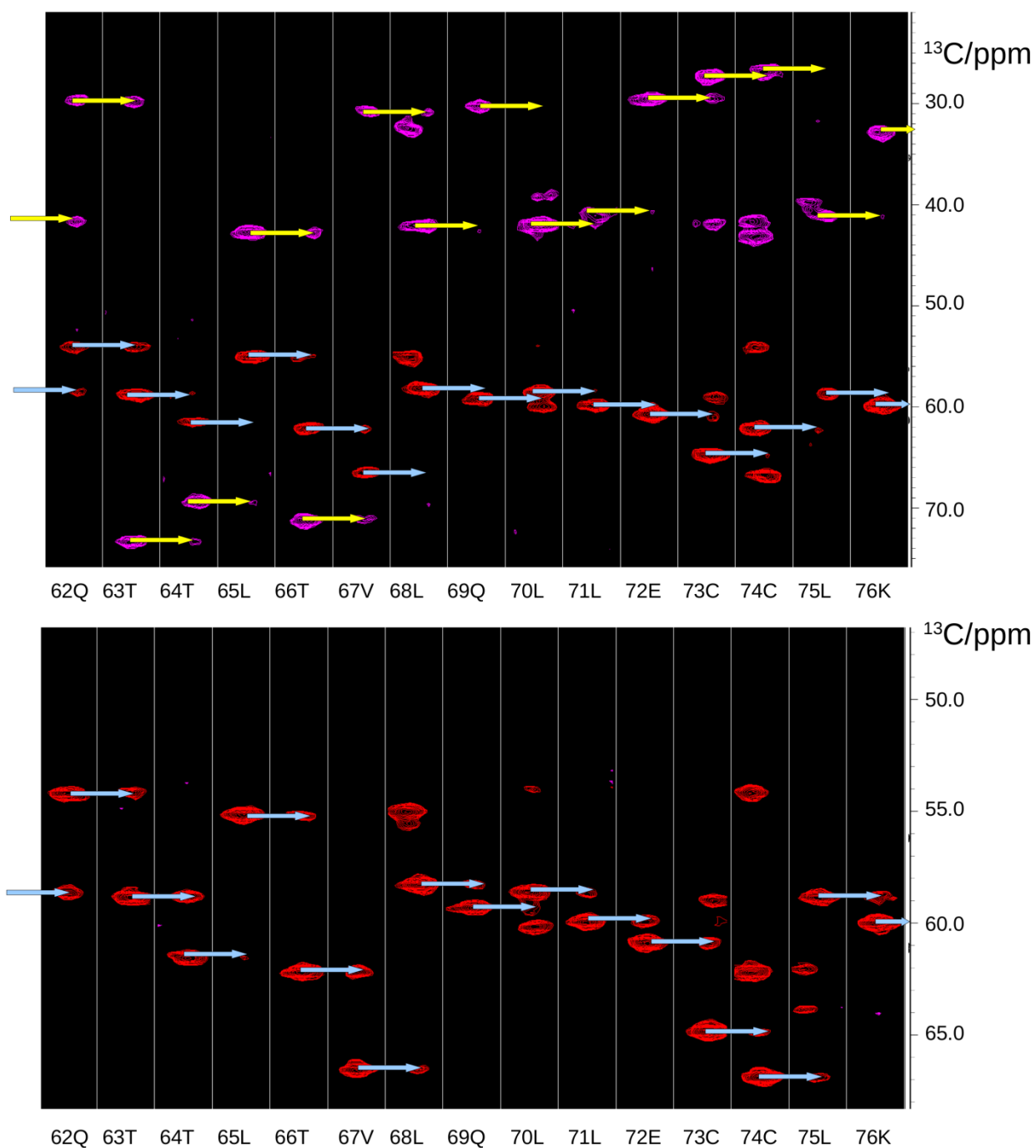
Figure 2.7: Sequential through-bond connections of *Sa*HPPK for residues Q[62]-K[76] in HNCACB (top) and HNCA strips (bottom) recorded on a ~250 μM *Sa*HPPK/AMPCPP sample, pH 8.0 at 600 MHz and a temperature of 22 °C. Preceding and intra-residue ^{15}N-^{1}HN correlations to ^{13}CA (red) and ^{13}CB (magenta) in the HNCACB spectra are shown. ^{15}N-^{1}HN correlations to ^{13}CA are show in red in the HNCA strips. Cyan and yellow arrows denote sequential connectivity between the individual residues. Note: some weaker *i-1* ^{13}CB and ^{13}CA correlations are missing (e.g. Leu68) as the HNCACB experiment was relatively insensitive.

2.3.4 Chemical shift mapping of cofactor and substrate revealed the binding sites

Chemical shift perturbations (CSPs) upon AMPCPP binding and peaks that disappeared upon HMDP (substrate) binding in the ^{15}N SOFAST HMQC spectra, were mapped onto the *Ec*HPPK structure (1RAO) (Figure 2.8) to identify those residues involved in direct binding and to identify residues susceptible to long range, indirect effects (see chapter 3 for further details).
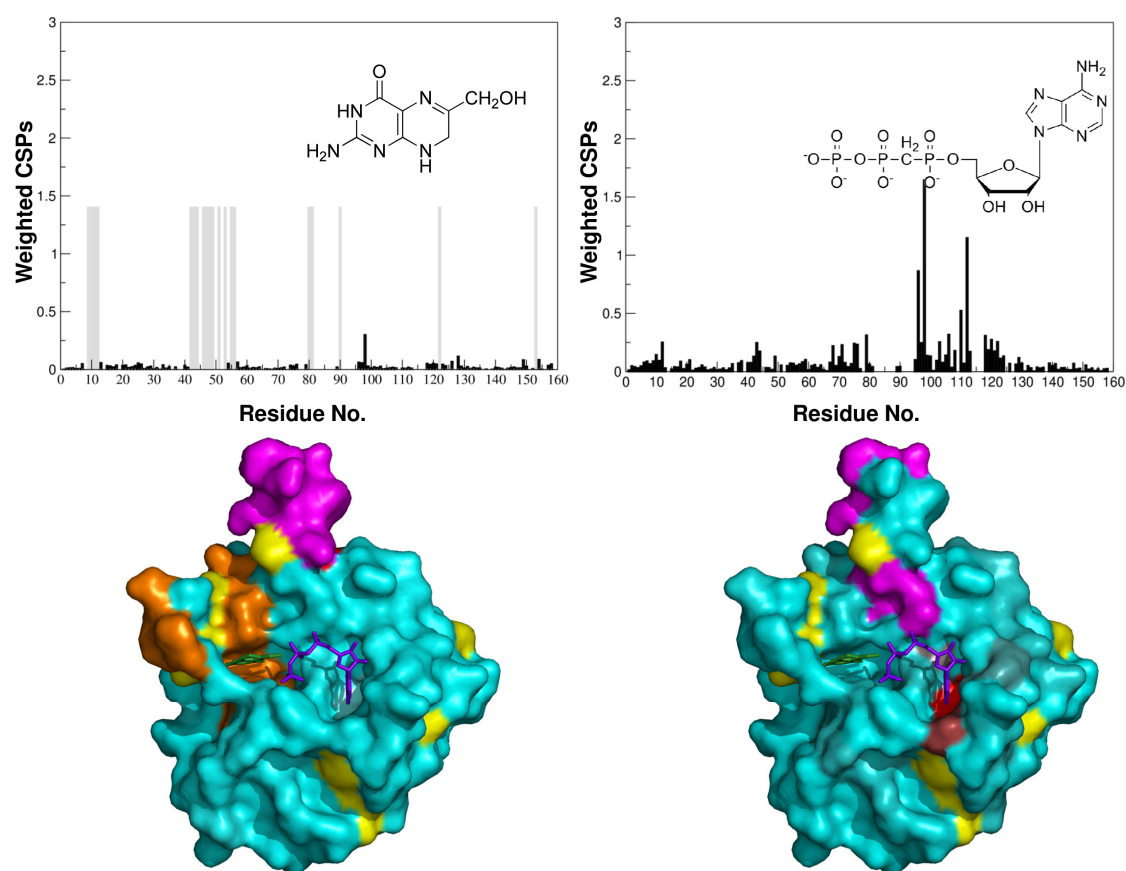
Figure 2.8: (Top) CSP data for HMDP (substrate) and AMPCPP binding to *Sa*HPPK. Grey bars indicate residues for which peaks were extensively broadened compared to the *apo* ^{15}N HSQC spectrum, and therefore not observed upon HMDP binding. Black bars indicate CSPs upon AMPCPP binding. (Bottom) CSPs mapped onto surface of *Ec*HPPK (1RAO). Residues missing in the *apo* or AMPCPP bound spectra are coloured magenta. Those additionally broadened upon HMDP binding are coloured orange. Unobserved proline residues are coloured in yellow. Residues displaying slow exchange CSPs upon binding of the AMPCPP are shaded red. HMDP (green) and AMPCPP (purple) are shown in sticks to help identifying the two binding sites.

2.4 Fragment-Based Screening (FBS)

2.4.1 Fragment screening using 1D STD NMR experiments

A fragment library was initially screened to find binders of *Sa*HPPK in the hope of discovering novel compounds with different chemical structures or scaffold to those of the cofactor and substrate.

A total of 350 compounds from the Maybridge RO3 fragment library were first screened using 1D saturation transfer difference (STD) NMR spectroscopy (Figure 2.9). About 150 compounds were identified as positive hits in the STD-NMR experiment, even though the excess ligand concentration was fairly low (~350 µM). This undesirably high hit rate (~ 40 %) could be because the active site of HPPK is relatively large (26 Å wide), with a particularly open ATP-binding pocket. Interestingly, we found a similar high hit rate during fragment screening of the adjacent enzyme (DHPS) in the pathway. Substantial active site loop motions occur in both enzymes, which may create a range of transient binding sites. Other groups such as Novartis have abandoned the STD experiment for this reason, (Wolfgang Jahnke, *person comm*) and prefer to use the alternate WaterLOGSY[14] experiments. While we did not pursue the 1D WaterLOGSY experiment, we chose to take advantage of 'Fast NMR methods' to run 2D SOFAST ^{15}N HMQC experiments.
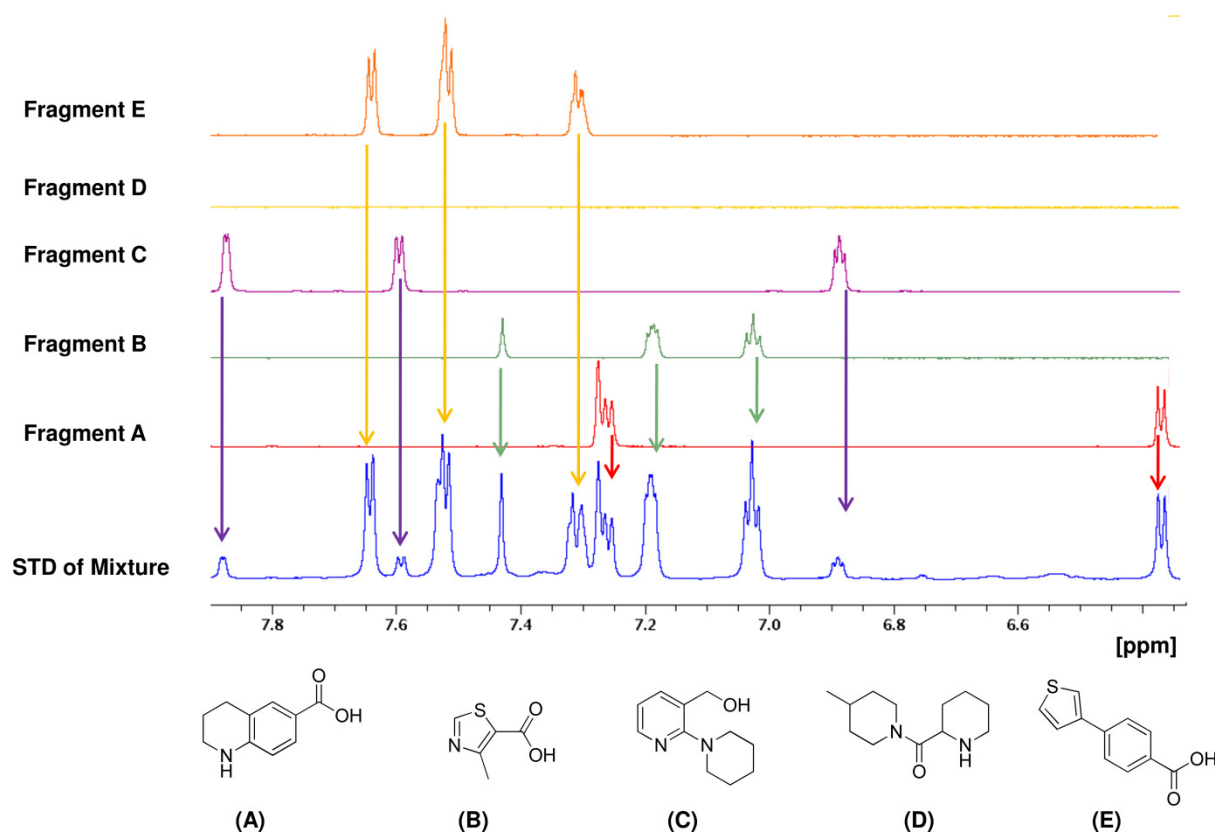
Figure 2.9: Example of a 1D STD spectrum of a mixture of fragments (each at 350 µM) in the presence of 25 µM SaHPPK. The arrows highlight fragment A, B, C and E proton signals matching those observed in the STD spectrum of the mixture, indicating those fragments which bind to SaHPPK. Note that the fragment D signals were outside the zoom region and this was not a hit.

2.4.2 Fragment screening using 2D experiments

HPPK is a small protein (~ 20 kDa including the N terminal residues of the His tag) and gave high quality 2D ^{15}N SOFAST HMQC NMR spectra. With the high hit rate from STD NMR experiments and given that ^{15}N-labeled SaHPPK expresses very well (>40 mg/L), all 750 fragments were rapidly screened initially on the cocktails of five fragments using the SOFAST 2D ^{15}N HMQC experiment which was recorded within ~12 mins for each cocktail. A total of 16 out of 150 cocktails produced CSPs when compared with the apo enzyme spectrum. Subsequently, 2D SOFAST ^{15}N HMQC experiments were re-acquired on individual fragments at a concentration of 1 mM to deconvolute

and confirm the binders. The 2D method was actually quicker, easier to analyse and far more informative than the 1D STD approach, giving both hit detection, validation and approximate site binding from chemical shift mapping.

A total of 16 confirmed hits (~ 2 % hit rate) were finally identified from 2D SOFAST ^{15}N HMQC experiments of the fragment library (Figure 2.10). Of these initial hits, fragments with thioamide substitutions on the heterocyclic ring displayed reasonable CSPs ($\Delta\delta$ > 0.2 ppm), whereas most other fragments showed very weak CSPs ($\Delta\delta$ < 0.1 ppm) and were deemed not worth pursuing at this stage. An Example of the spectrum recorded in the presence of a thioamide fragment is shown in Figures 2.11. The CSPs were mapped onto the *Ec*HPPK structure (1RAO), which revealed that most of these initial hits bind to the HMDP (substrate) site (residues E42, T43, A44, P45, V46, N55, F54 and F123), however, one fragment hit (MO00777) seems to bind specifically to the ATP site (I98, L99), also perturbing the amide resonances from metal binding residues D95 and D97 (Figure 2.12).
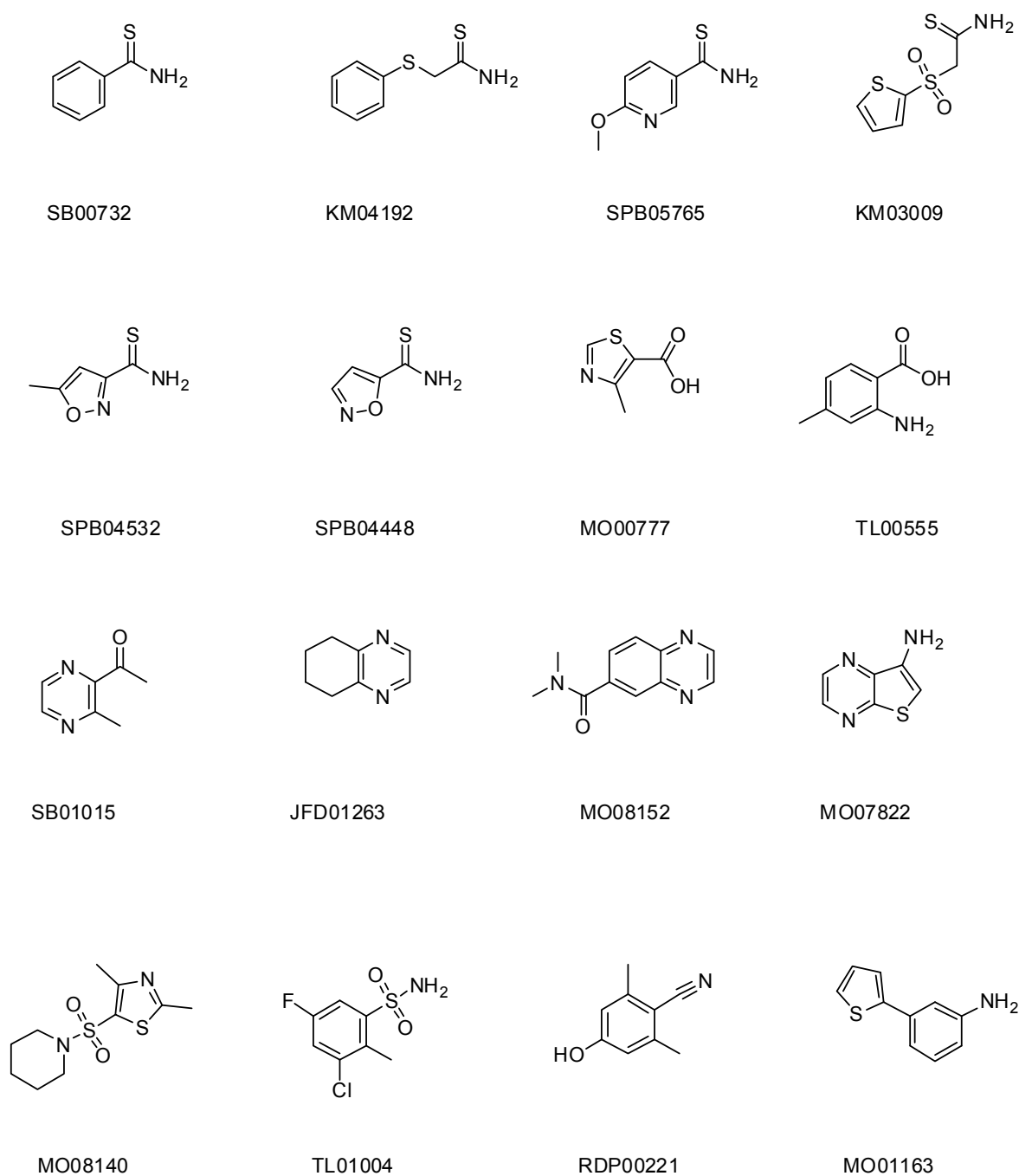
Figure 2.10: Structures of fragment hits identified in the first round of screening using 2D SOFAST ^{15}N HMQC NMR specroscopy.

51
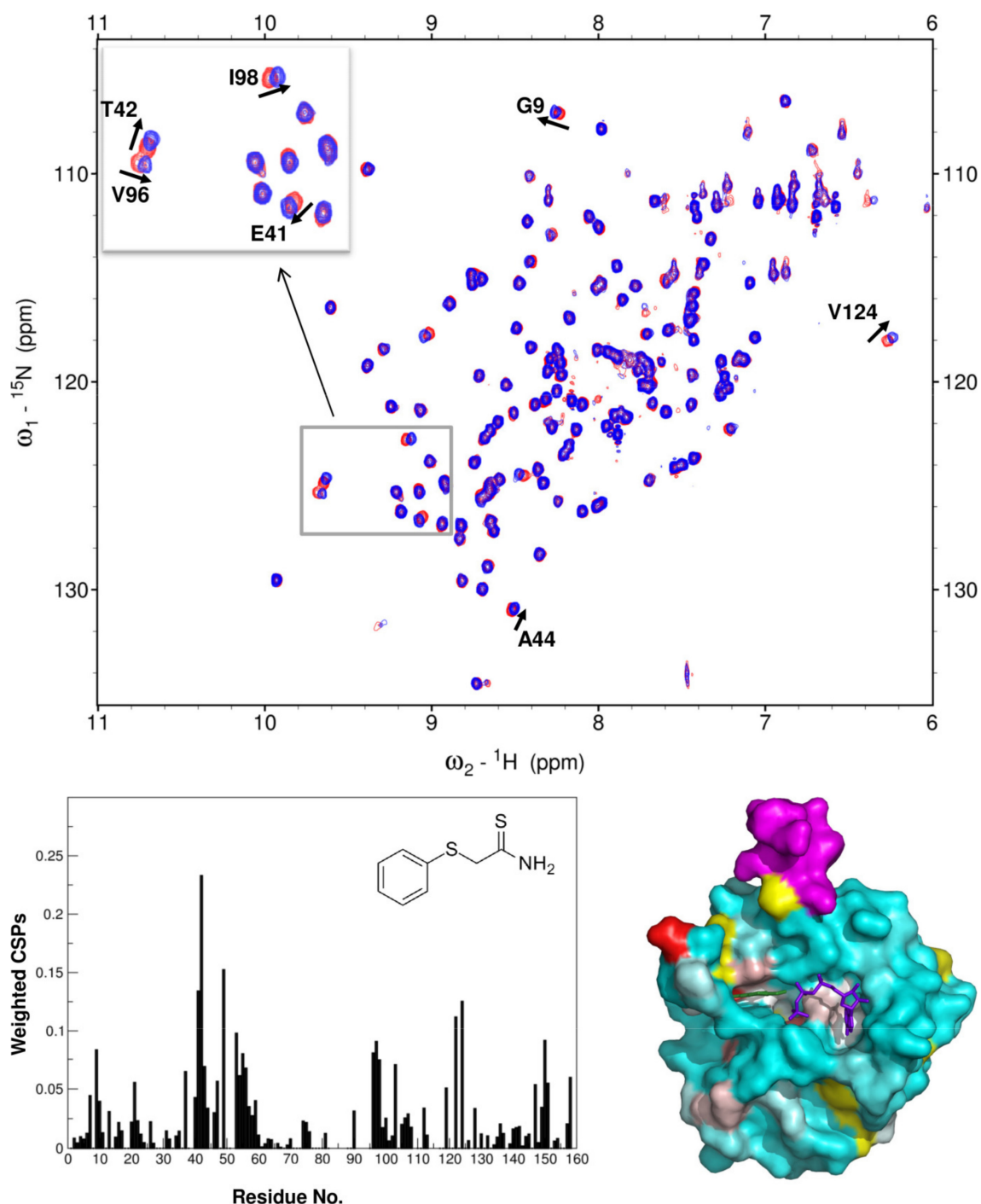
Figure 2.11: Superposition of the 600 MHz ^{15}N SOFAST HMQC spectra, (top) and CSP data of *Sa*HPPK in absence (red) and presence (blue) of 1 mM fragment hit (KM04192) (bottom). A portion of the spectrum is enlarged and boxed to emphasise perturbed resonances. The perturbed resonances were mapped onto the *Ec*HPPK (1RAO) crystal structure. HMDP (green) and AMPCPP (purple) are shown in sticks to help identify the binding sites for fragments. Residues with broadened resonances in the *apo* enzyme are coloured magenta, and unobserved proline residues are colored in yellow. Residues displaying detectable CSPs upon binding are shaded red, with the hue corresponding to the magnitude of the CSPs.
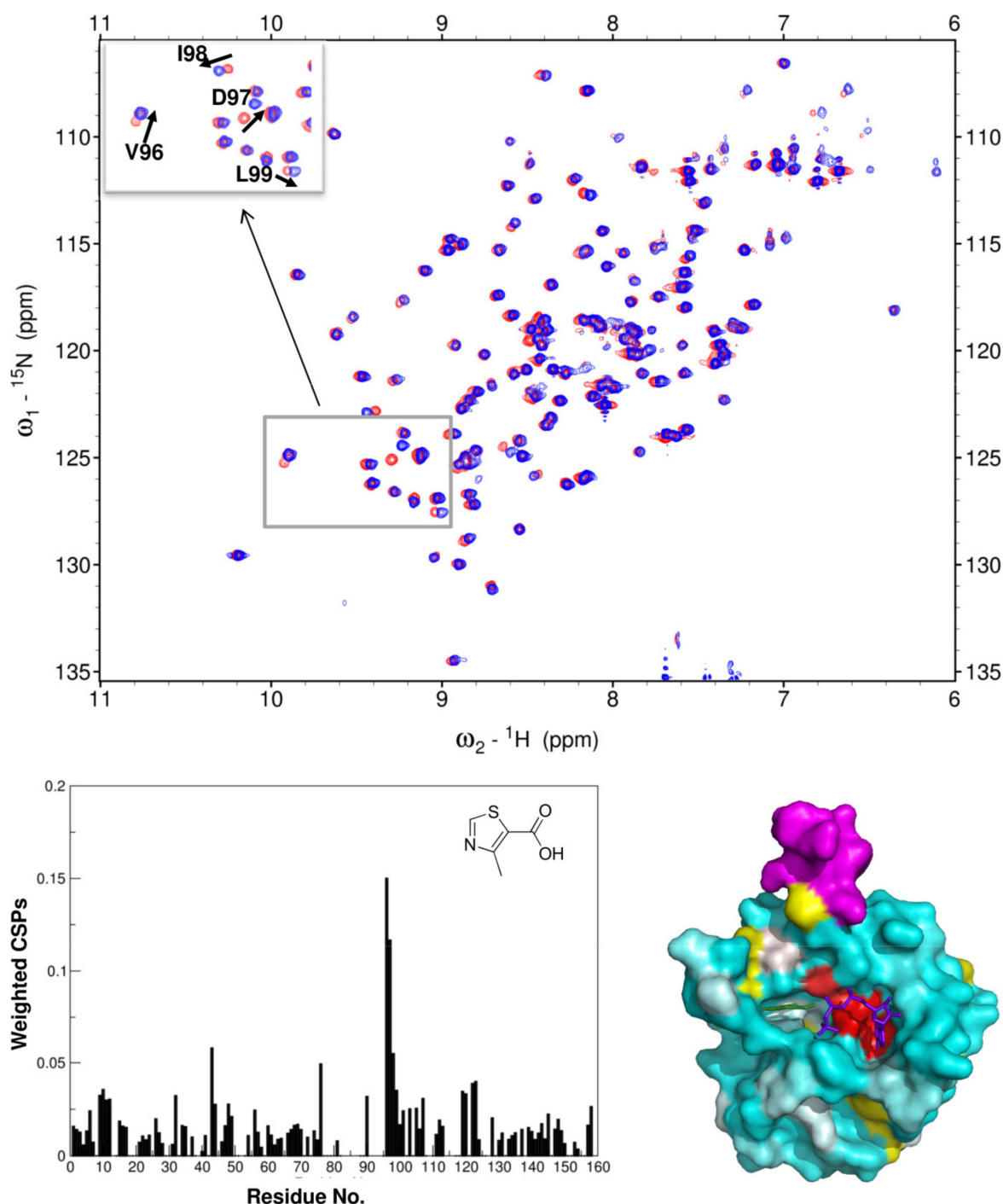
Figure 2.12: Superposition of the 600 MHz ^{15}N SOFAST HMQC spectra, (top) and CSP data of *Sa*HPPK in absence (red) and presence (blue) of 1 mM fragment hit (MO00777) (bottom). A portion of the spectrum is enlarged and boxed to emphasise perturbed resonances. The perturbed resonances were mapped onto the *Ec*HPPK (1RAO) crystal structure. HMDP (green) and AMPCPP (purple) are shown in sticks to help identify the binding sites for fragments. Residues with broadened resonances in the *apo* enzyme are coloured magenta, and unobserved proline residues are colored in yellow. Residues displaying detectable CSPs upon binding are shaded red, with the hue corresponding to the magnitude of the CSPs.

The thioamide moiety was thus defined as a simple pharmacophore and a few more analogues containing thioamide substitution were investigated. To begin with, 14 fragment analogues from commercial suppliers were selected based on a similarity search using Scifinder (https://scifinder.cas.org/scifinder/). The analogue search was restricted to compounds with MW ≤ 300 and a single substitution around the thioamide core moiety to simplify the effect of any substitutions. The selected fragments were then purchased (UkrOrgSynthesis Ltd, Ukraine) (Figure 2.13) and tested for binding to SaHPPK.

The measurement of the equilibrium binding constant (K_d) by NMR was not attempted for the fragments. While the CSPs were in fast exchange and gave binding site information (Figure 2.14), the apparent K_d determined for one of the fragments was deemed unreliable, as the NMR sample rapidly became cloudy and the NMR signal intensity decreased at higher fragment concentration (> 3 mM). This observation is typical of SaHPPK titrations owing to sample instability with handling over time. Binding affinities were instead determined by SPR experiments (section 2.5).
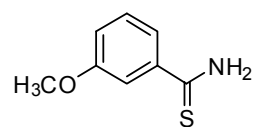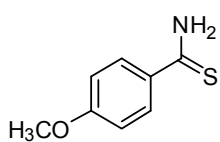
BBV- 093604   BBV-29343041   BBR- 005602   BBR- 007117
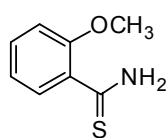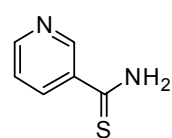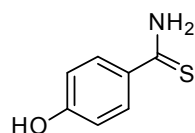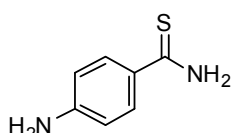
BBV-14847346   BBV-112091   BBV- 088256   BBV- 29348894
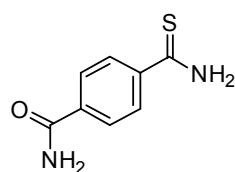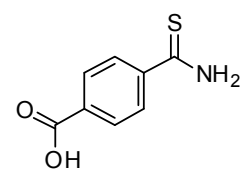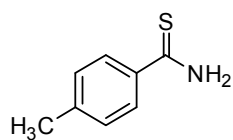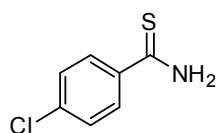
BBV- 090421   BBV- 157996   BBV- 27279788   BBV- 5094366
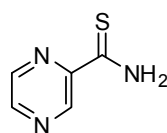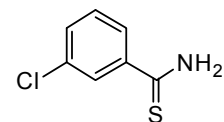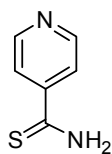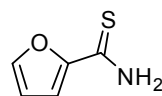
BBV- 072852   BBV- 2068048

Figure 2.13: Structures of fragments purchased based on the similarity search.
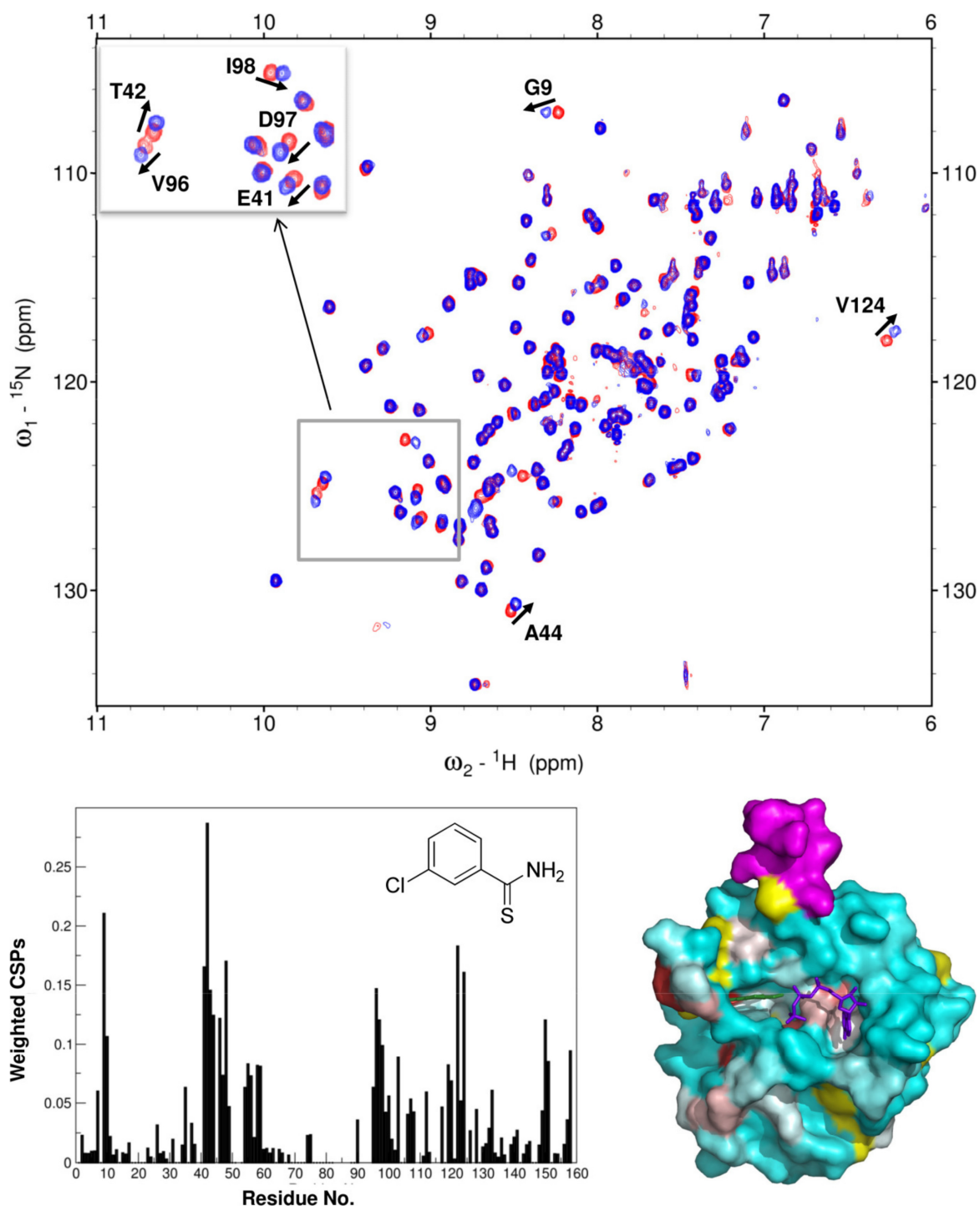
Figure 2.14: Superposition of the 600 MHz ^{15}N SOFAST HMQC spectra, (top) and CSP data of *Sa*HPPK in absence (red) and presence (blue) of 1 mM fragment hit (BBV-5094366) (bottom). A portion of the spectrum is enlarged and boxed to emphasise perturbed resonances. The perturbed resonances were mapped onto the *Ec*HPPK (1RAO) crystal structure. HMDP (green) and AMPCPP (purple) are shown in sticks to help identify the binding sites for fragments. Residues with broadened resonances in the *apo* enzyme are coloured magenta, and unobserved proline residues are colored in yellow. Residues displaying detectable CSPs upon binding are shaded red, with the hue corresponding to the magnitude of the CSPs.

## 2.5 Screening using Surface Plasmon Response (SPR)

All fragment hits and analogues were tested for their binding affinities by SPR. A minimal biotinylation approach was used to immobilize SaHPPK onto a NeutrAvidin chip surface, as described in the method sections 2.8.18 and 2.8.19. The binding affinity of the substrate, HMDP, was measured as a reference compound and also as a control to monitor the enzyme activity and integrity on the chip surface during the experiment.

The initial 16 fragments gave a very weak response ($RU_{max} < 2$) in the sensogram at a single point concentration of 100 µM, consistent with the weak CSPs in NMR. In contrast, four out of the 14 follow-up fragments showed slightly better responses ($RU_{max} > 2.5$) at 100 µM, as compared to other fragments (Figure 2.15). A titration was performed for these selected fragments to obtain binding affinity curves (Figure 2.16). While saturation could not be obtained at the highest concentration used in the experiment (256 µM), extrapolation to saturation was obtained based on the control binder (HMDP) and gave an estimate of the $K_d.$ This was found to be in the range of ~ 500 µM to 1 mM (Figure 2.16) for the four selected binders.
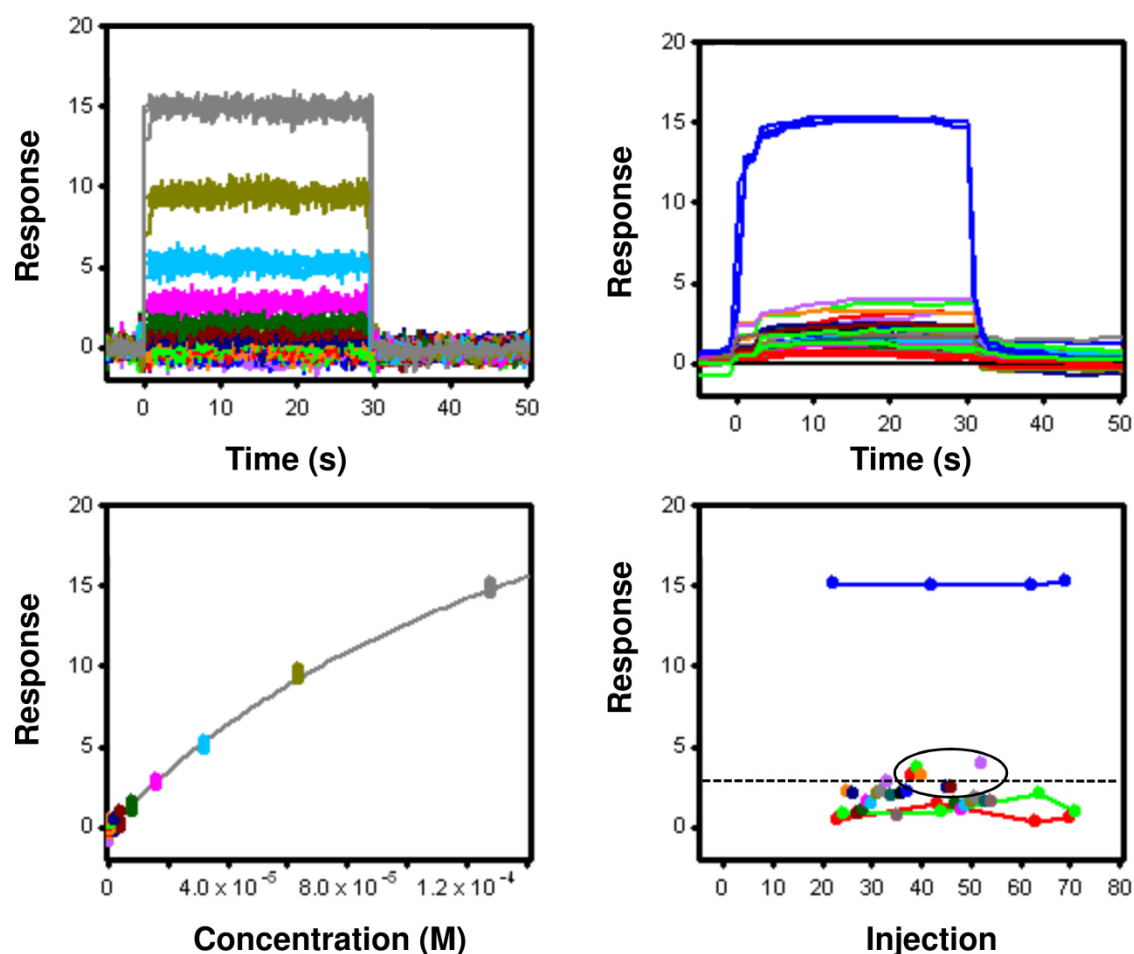
Figure 2.15: (Left top) SPR raw data and (Left bottom) steady-state response curve for HMDP, used as a control to compare the binding of fragments. (Right top) SPR raw data and (Right bottom) responses upon binding of fragments to SaHPPK. Four fragments that showed $RU_{max} > 2.5$ (above the cut-off line) are circled (Right bottom). The blue line represents the consistent response from the control compound. The y-axis represents the amount of bound analyte in terms of RU, and the x-axis represents time after injection (s).
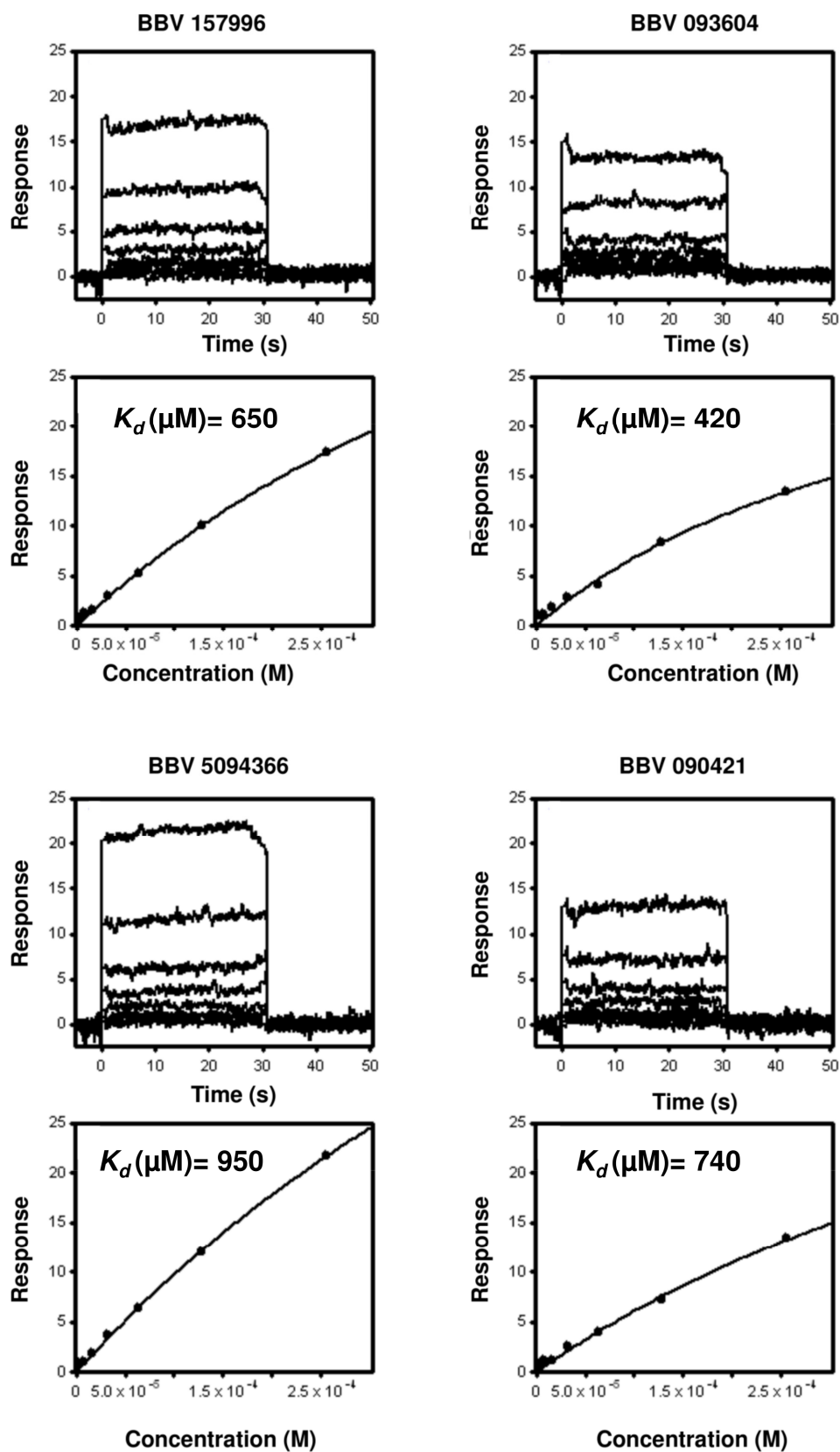
Figure 2.16: (Top) SPR raw data and (Bottom) steady-state response curves for the binding of four fragments.

2.6 X-ray crystallisation trials

A major challenge associated with FBS approaches to structure-based drug design is that owing to their small size, fragments generally bind to their target with low affinity, often beyond detection by the usual *in vitro* enzyme assays. Structural details of their binding modes are highly beneficial to facilitate compound optimisation to achieve higher affinity prior to more conventional assay-based SAR. The small size of fragments also means that computational methods are not as reliable compared to docking larger HTS compounds[15] to determine their bound conformation to the receptor. Crystallisation studies are thus an integral process along the route to understand the structure and intermolecular interactions of a bound fragment.

Despite extensive crystallisation trials, we were unable to crystallise the *apo* enzyme, which would be ideal for soaking experiments of potential fragment hits from screens. While a crystal of the AMPCPP/*Sa*HPPK complex diffracted to 3.5 Å (data not shown), no crystals of the ternary HMDP/AMPCPP/*Sa*HPPK complex were found in the experiments performed. HMDP (substrate) in solution is unstable and degrades to a weakly binding oxidised species[16] over the time period of crystallisation (~7 days). In fact, in order to crystallise *Sa*HPPK and to determine the structure to high resolution, we had to find a substrate-site binder (8-mercaptoguanine), which is more stable than HMDP.

The details of the discovery, characterisation and the structure of the bound inhibitor are described in Chapter 3. Crystals of the 8-mercaptoguanine/*Sa*HPPK complex are shown in Figure 2.17, which diffracted to 1.65 Å at the Australian Synchrotron. Crystallisation of the *Sa*HPPK/8-mercaptoguanine complex was not trivial by any means. About seventy 96 well-plates (over 14,000 droplets) were needed to grow good quality crystals. Several low quality crystals were crushed and used for subsequent seeding experiments[17] to obtain crystals of sufficient quality (Figure 2.17). It is notable that the crystals were particularly small, often multiple and flat with an aspect ratio 80×80×5 microns, which made obtaining reflections over a full 360° difficult. The MX2 microfocus beamline was

essential in order to achieve workable data and often required firing X-rays at different parts of the crystal to obtain a full dataset. Nevertheless, a low solvent content of 39.6 % meant that the crystal tolerated a high dose of X-rays without disintegrating.
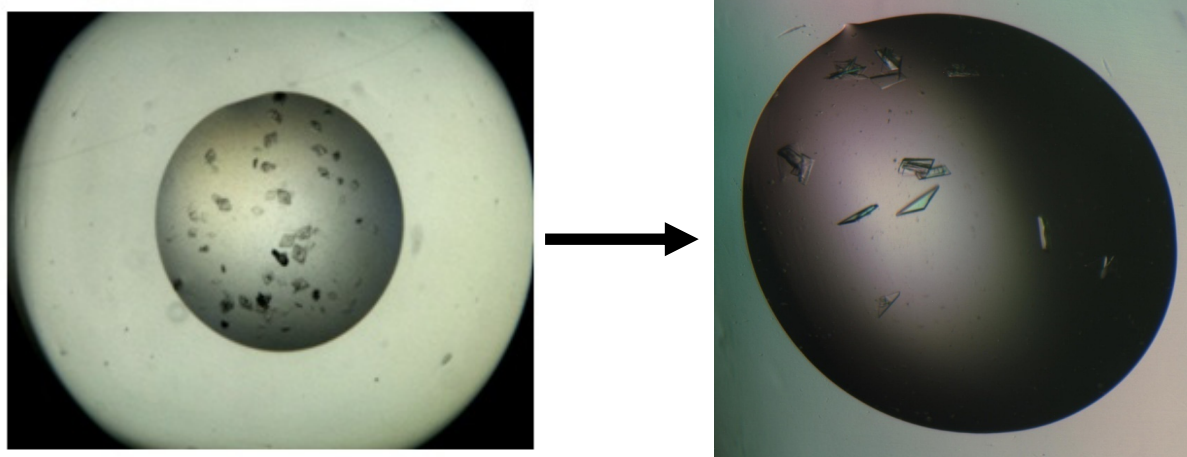
Figure 2.17: Crystals of *Sa*HPPK in complex with 8-mercaptoguanine. Initial crystal trials produced multiple stacked plate crystals (left), which were crushed and used as seed crystals to eventually obtain single crystals (right), which diffracted to 1.65 Å resolution.

2.7 Conclusions

In conclusion, the *Sa*HPPK protein was expressed and optimal conditions for prolonged sample stability were established. The NMR-based fragment screening, to our knowledge, represents the first of its type applied to HPPK, and has identified some very promising hits that were validated by CSPs to be pterin or ATP site binders in 2D SOFAST ^{15}N HMQC NMR experiments. The binding affinity of the best four thioamide fragments was in the range of 0.5–1 mM, as determined by SPR (Figure 2.18). The corresponding ligand efficiency[18, 19] is in the range of 1.7–1.8 kJ/heavy atom (above the desired cutoff value of 1.3 kJ/heavy atom) and accordingly may represent good starting points for evolution to lead-like compounds[20]. Compounds with the thioamide functionality have been reported as tuberculosis drugs[21]. As for other fragment hits, a search of the 2008.1 MDL Drug Data Report reveals that fragments MO00777 and TL00555 are embedded in dozens of entirely different compounds being investigated by around 10 drug companies for a diverse array of indications. For example, Febuxostat (Figure 2.18 B) from Teijin Pharmaceuticals is a marketed xanthine oxidase inhibitor, an enzyme that intriguingly binds purines. Given that MO00777 clearly binds closely to the metal binding residues, it may be envisaged that when combined or linked to the 8-mercaptoguanine inhibitor (see chapter 3), it may lead to a high affinity scaffold. Future directions would therefore be to investigate the competition of MO00777 with 8-mercaptoguanine and to co-crystallise these fragments, to define their precise binding location. If crystallisation was not fruitful, then the use of a modern PCS NMR approach can be undertaken[22] to help determine the structure of the bound fragment complex in a similar manner to that described for the active site loop conformation in chapter 5.

A
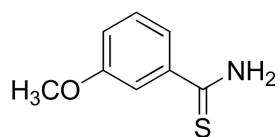
BBV- 093604
K_d = 420 μm
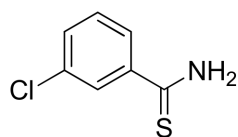
BBV- 5094366
K_d = 950 μM
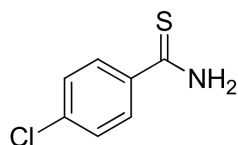
BBV-157996
K_d = 650 μM
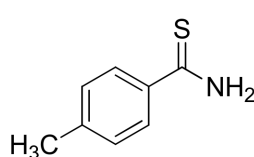
BBV- 090421
K_d = 740 μM

B
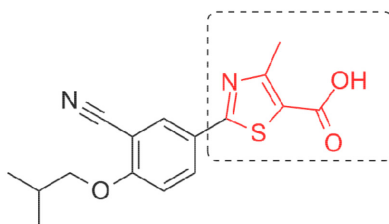
Figure 2.18: Structures of the final four fragments identified by NMR screening, together with their binding affinities, as determined by SPR. (B) The structure of Febuxostat showing the embedded core of fragment MO00777 (red).

2.8 Methods

2.8.1 *Sa*HPPK plasmid

A pET28a plasmid containing the synthesised *Sa*HPPK sequence (UniProtKB entry: Q99W87) was designed with an N-terminal hexa-His tag and a thrombin cleavage site and purchased from Geneart®.

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSHM IQAYLGLGSN IGDRESQLND AIKILNEYDG ISVSNISPI


        70         80         90        100        110        120
YETAPVGYTEQ PNFLNLCVEI QTTLTVLQLL ECCLKTEECL HRIRKERWGP RTLDVDILL


       130        140        150        160        170
YGEEMIDLPKL SVPHPRMNER AFVLIPLNDI AANVVEPRSK LKVKDLVFVD DSVKRYK
```

2.8.2 Transformation of *E. coli*

BL21 *E. coli* (λDE3) codon plus cells were transformed with pET28a plasmids using the heat shock method[23]. Competent *E. coli* (50 μL) cells were combined with plasmid DNA (10-50 ng) in a 1.5 mL centrifuge tube and incubated for 15 mins on ice. The tube was transferred to a water bath at 42 °C for 30 s then returned to ice for 2 mins. After addition of 200 μL of 2YT broth, cells were grown for 1 h with shaking at 200 rpm at 37 °C. The 250 μL culture was then plated onto 1×YT (100 μg/mL kanamycin) plate and incubated for 16 h at 37 °C.

2.8.3 Preparation of glycerol stocks

Glycerol stocks were prepared by resuspending transformed *E. coli* cells from the plate into sterile 2YT containing 15 % glycerol (this can be made by taking sterile 2YT broth and adding sterile 80 % glycerol to give a final 15 % glycerol vol:vol). The small aliquots (μL) of resuspended cells were frozen at -80 °C.

2.8.4 Expression of unlabelled *Sa*HPPK

The transformed *E. coli* BL21 (DE3) were grown overnight in 10 mL of growth media (2YT media supplemented with 100 µg/mL kanamycin antibiotic for selection). The overnight culture was then subcultured into fresh growth media (1L) (2YT + Kanamycin) and grown at 37 °C until the OD_{600} was 0.5–0.8. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final optimised concentration of 0.3 mM and expression was carried out at 28 °C for 5 h. The cultures were centrifuged at 5000 g at 4 °C for 10 mins and the supernatant was discarded. Cell pellets were stored at -80 °C until purification.

2.8.5 *Sa*HPPK purification

Cell pellets stored at -80 °C were thawed and re-suspended in 100 mL 50 mM HEPES, 5% glycerol, pH 8.0. An EDTA-free Complete™ protease inhibitor cocktail tablet (Roche) was added along with lysozyme, the latter to a final concentration of 0.2 mg/mL. After 10 mins of gentle stirring, cells were then sonicated, cell debris removed by centrifugation at 18,000 g (4 °C) for 30 mins and the supernatant filtered (0.45 µm filter). The supernatant was loaded onto a 5 mL HisTrap Ni-NTA column (Qiagen) and unbound protein was washed off with 10 mM imidazole in 50 mM HEPES-NaOH buffer, 300 mM NaCl, 5 % glycerol, 1 mM DTT, pH 8.0. The tagged protein was eluted from the HisTrap column with a 250 mM imidazole gradient in the HEPES-NaCl buffer.

2.8.6 Thrombin cleavage of the N-terminal his tagged *Sa*HPPK

The untagged protein was generated by on-column digestion carried out with thrombin to remove the N-terminal His tag. One unit of thrombin (Sigma-Aldrich) was used for each mg of tagged protein. Firstly, supernatant was loaded on to a 5 mL HisTrap column with Buffer A (50 mM Hepes, 5% glycerol, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, pH 8.0.

After loading, the column was washed with 10 column volumes (CV) of Buffer C (50 mM HEPES, 1% sorbitol, 2 mM DTT, pH 8.0) to remove salts from the column. A 5 mL solution (containing thrombin 100 U, 2 mM $CaCl_2$, 2 mM ATP, 10 mM $MgCl_2$ in 50 mM HEPES pH 8.0) was injected manually with a syringe onto the column and left overnight. The next day, cleaved protein was eluted with 10 mM imidazole in HEPES–NaCl buffer and the leftover uncleaved protein was eluted with a 250 mM imidazole gradient.

2.8.7 Size exclusion column to isolate monomer from aggregate

Both tagged and untagged versions of the protein were further purified on a Superdex 75 size-exclusion 16/60 column (GE Healthcare). The protein sample eluted from HisTrap columns was concentrated down to a volume of 5 mL and injected onto the Superdex 75 pre-equilibrated with 2 CV of Buffer C (50 mM HEPES, 1% sorbitol, 2 mM DTT, pH 8.0). After eluting with the same buffer all the fractions were analysed using a 15% SDS-PAGE gel (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) with Coomassie staining and protein containing fractions were pooled, snap-frozen and stored at -80 °C.

2.8.8 Preparation of isotopically-labelled protein for NMR spectroscopy

Isotopically-labelled samples of SaHPPK for NMR spectroscopy were prepared as per the modified protocol of Lewis Kay's lab[11].

E. coli BL21 (DE3) cells (Agilent) transformed with the plasmid were grown overnight in 3 mL of 2YT media supplemented with 100 µg/mL kanamycin for selection. The overnight culture was subcultured into 50 mL of minimal media (Table 1) that was grown to an OD_{600} of 0.5–0.7. This 50 mL culture was then centrifuged at 4000 g at 30 °C for 10 mins and the cells were gently re-suspended in 1 L of minimal media supplemented with 1.5 g of ^{15}N-ammonium chloride and/or 3 g of ^{13}C-glucose (see Table 1) and grown at 37 °C until the OD_{600} was 0.5–0.8. Isopropyl β-D-1-thiogalactopyranoside (IPTG)

was added to a final optimised concentration of 0.5 mM and expression was carried out at 20 °C for 12 h. Purification was as described above (section 2.8.5).

Table 1: Composition of minimal media

1 x M9 salts		Additives		Heat labile additives (0.22 µM filtered)	
KH_2PO_4	3 g	$FeCl_3$ (10 mg/mL)	0.2 mL	D-glucose (20g/100mL)	20 mL
Na_2HPO_4	6.78 g	1M $MgSO_4$	1mL	Thiamine (10 mg/mL)	1mL
NaCl	0.5 g	0.1M $CaCl_2$	1mL	Biotin (10 mg/mL)	1mL
$^{15}NH_4Cl$	1.5 g	$ZnCl_2$ (34 mg/mL)	0.2 mL	Kanamycin (50 mg/mL)	1 mL
MilliQ to 100 mL		to ~ 877 mL			

2.8.9 Protein concentration and molecular weight estimation

The concentration of protein samples was estimated by direct UV measurement at 280 nm using a NanoVue™ Spectrophotometer (GE healthcare). The extinction coefficient of 14440 cm^{-1} M^{-1} for HPPK was calculated from the amino acid composition using the ExPASy site (http://web.expasy.org/protparam/). The *Sa*HPPK protein identity was confirmed by SDS-PAGE and Mass spectrometry (MS) (Figure 2.19). The mass of the protein calculated from the predicted amino acid sequence with a His6-tag (20028.03 Da), compared well to the electrospray mass spectrometry (ES-MS, 20033.88 Da).

Figure 2.19: Electrospray mass spectrum of purified *Sa*HPPK protein. Inset, SDS-PAGE gel (15 %) of *Sa*HPPK purified protein fractions. Lane 1 (Thrombin cleaved fraction), lanes 2-6 uncleaved *Sa*HPPK enzyme.

2.8.10 N-Terminal sequence analysis of *Sa*HPPK

The N-terminal sequence of the purified HPPK was determined after SDS-PAGE gel and mass spectrometry verification. N-terminal sequence analysis of the protein (~100 pmol) (Table 2) gave results corresponding to the expected N-terminus for cleaved protein starting with Glu (first three non-native residues as a result of the N-terminal thrombin cleavage).

Table 2: The sequence determined from N-terminal sequence analysis

1	GLU ?*	11	LEU
2	SER	12	GLY ?*
3	HIS	13	SER
4	MET	14	ASN
5	ILE	15	ILE
6	GLN	16	GLY ?*
7	ALA	17	ASP
8	TYR	18	ARG
9	LEU	19	GLU
10	GLY ?*	20	SER

*data for these residues was not clear enough in the N-terminal sequence analysis; therefore they could be either Glu or Gly. N-terminal sequence analysis information was obtained using the automatic sequence analyser (Waters) at CSIRO Materials Science and Engineering, Parkville.

2.8.11 Buffer screening - The thermal shift denaturation assay

This fluorescence-based assay[2] was performed using an Applied Biosystems Step One Plus Thermal Cycler. Solutions of 5 µM HPPK, 10X Sypro Orange dye and different buffer solutions were added to the wells of the 96-well plate. The total well volume used was 30 µL. The plate was heated from 20–99 °C with a heating rate of 1.0 °C/min. The fluorescence intensity was measured with excitation/emission = 583/610 nm. The different buffer solutions tested were acetate (pH=4.0-4.5), citrate (pH=5.0), MES (pH=6.0-6.5), HEPES (pH=7.5-8.0), Tris (pH 8.0-8.5) containing different additives such as NaCl (50-300 mM), RE (arginine, glutamine acid) (50-100 mM), CHAPS (1-5 %), glycerol (1-10 %), sorbitol (1-10 %) and combinations of additives. T_m and

ΔT_m values were determined from the melting curves obtained. After testing several buffer conditions, the 50 mM HEPES (pH 8.0), 1% sorbitol and 2.0 mM DTT buffer condition was identified as most likely stabilising condition for *Sa*HPPK.

2.8.12 Buffer screening – The microdialysis button test

A small portion of concentrated protein (~ 200 μM) sample (10-15 µL) was added to the center well of the buttons followed by covering with a dialysis membrane and fixed in place by a rubber O-ring. The microdialysis cells were then incubated in various test buffer solutions at room temperature for up to two weeks[3]. Each day, the clarity of the samples was monitored. Most of the test buffer conditions showed extensive precipitation early on, clearly visible to the naked eye within one day into the trial. The button with the buffer condition comprising 50 mM HEPES (pH 8.0), 1% sorbitol and 2.0 mM DTT was still clear after four days.

2.8.13 TSA for ligand binding

A 96-well plate containing solutions of 5 µM HPPK, 10X Sypro Orange dye and 500 µM of test compound in buffer (50 mM HEPES (pH 8.0), 1% sorbitol, 2.0 mM DTT, 10 mM MgCl$_2$) was heated from 20–99 °C with a heating rate of 1.0 °C/min. The fluorescence intensity was measured with excitation/emission wavelengths of 583/610 nm. T_m and ΔT_m values were determined from the melting curves obtained.

2.8.14 Biochemical assay

The spectrophotometric enzyme assay was performed using a Thermo Multiskan Ascent plate reader and absorbance was recorded at 340 nm to measure the rate of oxidation of NADPH to NADP$^+$. Buffers used to make up reagent solutions were purged with nitrogen to reduce oxidation of the HMDP.

Final concentrations of reagents in the standard assay buffer for AMPCPP inhibition measurements were as follows: 100 mM Tris–HCl (pH 8.5), 10 mM MgCl$_2$, 1.0 mg/mL BSA, 20 mM β-mercaptoethanol, 5 μM ATP, 3 μM HMDP, 50 μM NADPH, 50 μM p-aminobenzoic acid (PABA), 4.0 μg DHPS and 8.0 μg DHFR. The total reaction volume used in each well was 200 μL and comprised 5 μL of test compound solution (500 μM final concentration), 10 μL of SaHPPK solution (final amount 0.4 μg) and 185 μL of assay buffer. The assay buffer was added using a multichannel pipette to initiate the reaction. The 96-well plates were incubated at 25 °C and shaken before measuring the absorbance. Each well was read a total of 20 times with 30 s intervals (total time 10 mins). The rate of decreasing absorbance at 340 nm was recorded to give the percentage of SaHPPK inhibition. Each reaction was performed in triplicate. The AMPCPP IC$_{50}$ value was calculated by fitting to a one-site competition model within GraphPad Prism.

2.8.15 NMR backbone assignments of *apo* and AMPCPP bound *Sa*HPPK

3D NMR experiments were recorded at 22 °C on a Varian Inova 600 MHz NMR spectrometer operating at a field strength of 14.1 Tesla and equipped with a Z axis gradient and a cold probe. A sample of uniformly labeled SaHPPK with ^{15}N and ^{13}C was prepared and concentrated to ~0.25 mM in NMR buffer (50 mM HEPES buffer, 90% H$_2$O, 10% D$_2$O, pH 8.0, 1% sorbitol, 10 mM DTT) and sealed under nitrogen in a standard 5 mm (O.D.) NMR tube. As the *apo* sample tended to yield broader and poorer spectra than the sample in complex with 1 mM AMPCPP and 10 mM MgSO$_4$, the assignments of the *apo* sample were facilitated by the assignments of the SaHPPK/AMPCPP complex. 2D ^{15}N SOFAST HMQC[24] experiments were acquired with both a large (1944 Hz) and a small, folded (1054 Hz) sweep width in ^{15}N (t1). The latter sweep width was incorporated into all the following 3D (^{1}H, ^{13}C, ^{15}N) experiments; 3D HNCA, 3D HNcoCA, 3D HNCACB, 3D CBCA(CO)NH and 3D ^{15}N edited NOESY. NMR data was processed using nmrPipe[25] with linear prediction in ^{13}C(t2) and mirror image linear prediction in ^{15}N(t3) and analysed using Xeasy[26] or Sparky[27]. Typically, 3D data sets were

recorded in 36–44 h. All the above experiments were acquired on the *Sa*HPPK/AMPCPP sample and the 2D ^{15}N HSQC, 3D HNCA and 3D ^{15}N NOESY experiments were recorded on the *apo* sample.

2.8.16 1D STD-NMR experiment for fragment screening

1D STD experiments[28, 29] were acquired using a Gaussian cascade to provide a 2s steady state saturation pulse centred at 0.5 ppm and 14 ppm for the saturated and reference experiment respectively. The two experiments were recorded interleaved using WATERGATE for water suppression. The total time was ~15 mins. The samples were prepared using 25 μM unlabeled *Sa*HPPK in NMR buffer (50 mM HEPES buffer, 90% H_2O, 10% D_2O, pH 8.0, 1% sorbitol, 5 mM DTT) with mixtures of 5 fragments (final concentration 350 μM each) added from 100 mM stocks in DMSO (D_6). The STD measurements were acquired on an 800 MHz Bruker Avance II spectrometer, employing a triple-resonance 5 mm TXI cryoprobe equipped with single-axis gradients and an autosampler. The STD data were processed in TOPSPIN version 3.1 (Bruker Biospin).

2.8.17 2D NMR screening of fragment libraries

2D NMR experiments were recorded at 22 °C on a Varian Inova 600 MHz NMR spectrometer equipped with a cryoprobe and Z axis gradient. Cocktails of five fragments from the Maybridge fragment library dissolved in DMSO (D_6) were added to a sample of ~70 μM *Sa*HPPK in NMR buffer to give a final individual fragment concentration of ~ 350 μM. A SOFAST ^{15}N HMQC[24] experiment was recorded using a recycle delay time of 0.3 s and 1024 and 70 complex points in the ^{1}HN (t2) and ^{15}N (t1) dimension respectively yielding a total time of ~12 mins. Follow up screens were done similarly by recording the addition of each one of five fragments from the cocktail at a concentration of 1 mM. Chemical shifts were compared to those of reference spectra of the *apo* sample with the equivalent amount of DMSO added. A DMSO titration was also performed to examine that the observed CSPs were not the result of

minor variations in DMSO concentration. Spectra were automatically processed in batch mode using in house scripts and overlaid in Sparky where they were manually analysed.

The CSPs were calculated using the following equation: $\Delta\delta = \sqrt{((\delta N^{15}\times 0.154)^2+(\delta HN^1)^2)}$ in ppm where δ refers to the chemical shift change of the resonance relative to the resonance in the *apo* SaHPPK SOFAST ^{15}N HMQC spectrum. The factor of 0.154 is used to account for the differences in the spectral width for ^{1}H and ^{15}N amide shifts which is in the ratio of ~ 0.154:1.

2.8.18 SPR - Minimal biotinylation of *Sa*HPPK

*Sa*HPPK protein was buffer-exchanged by size exclusion chromatography on a Superdex75 (10/300 GL) column to remove any oligomer impurities from the protein sample before adding biotin. The column was pre-equilibrated in 50 mM HEPES, pH 8.0, 150 mM NaCl, 0.005% (v/v) polysorbate 20, 5 mM DTT, 10 mM MgCl$_2$. The collected monomer protein peak (~30 nanomoles) was "minimally" biotinylated[30] on ice for 2 hours using an equimolar concentration of EZ-Link® Sulfo-NHS-LC-LC-Biotin (Pierce). Any unreacted biotin was removed by passing the sample again through the Superdex75 and fractions of the monomeric peak were stored frozen in 50 μL aliquots at -80 ^{o}C.

2.8.19 SPR - NeutrAvidin immobilisation and biotinylated*Sa*HPPK capture

NeutrAvidin was immobilised at 25 °C onto a CM5 sensor chip (GE Healthcare) docked in a Biacore T100 instrument (GE Healthcare) using standard amine-coupling methods and HBS-P as the running buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 0.005% (v/v) polysorbate 20). The chip surface was activated by injecting a 1:1 mixture of 50 mM EDC (*N*-ethyl-*N'*-[3-dimethylaminopropyl]carbodiimide) and 200 mM NHS (*N*-hydroxysuccinimide) across all four flow cells for 7 min at 10 μL/min. NeutrAvidin (Pierce) at 100 μg/mL dissolved in 10 mM sodium acetate, pH 5.5, was then injected for 2 min at 10 μL/min. Excess activated groups were blocked by using a 7 min

injection of 1.0 M ethanolamine, pH 8.0 at the same flow rate, resulting in immobilisation levels of approximately 13,000 resonance units (RU) in all four flow cells. The biotinylated *Sa*HPPK was captured onto a NeutrAvidin surface in capture buffer (50 mM HEPES, pH 8.0, 150 mM NaCl, 0.005% (v/v) polysorbate 20, 5 mM DTT, 10 mM $MgCl_2$) by injecting at 10 µL/min for 6 min over a single flow cell which typically resulted in immobilisation of ~10,000 RU of target protein. A minimally-biotinylated bovine carbonic anhydrase II protein was captured in a separate flow cell to provide for an unrelated negative control. The blank (typically flow-cell 1) and captured surfaces were subsequently blocked with three 1 min injections of 1 µg/mL D-biotin (Sigma) at 10 µL/min.

2.8.20 SPR - Binding experiments and data analysis

All SPR binding experiments were performed at 20 °C in SPR binding buffer (50 mM HEPES, pH 8.0, 150 mM NaCl, 0.005% (v/v) polysorbate 20, 5 mM DTT, 10 mM $MgCl_2$, 5% (v/v) DMSO). Control binder and fragments were first screened at a single concentration of 100 µM.

Selected fragments were serially diluted (either 2- or 3-fold from a top concentration of 256 µM) in SPR binding buffer keeping the final DMSO concentration at 5 % and injected for 30 s contact time at 60 µL/min and then allowed to dissociate for 60 s. To ensure complete dissociation, SPR running buffer was injected after each ligand injection to ensure complete regeneration of immobilised HPPK. Binding sensorgrams were processed using Scrubber software (version 2, BioLogic Software, Australia). Sensorgrams were first zeroed on the *y* axis and then x-aligned at the beginning of the injection. Bulk refractive index changes were removed by subtracting the reference flow cell responses. The average response of all blank injections was subtracted from all analyte injections and blank sensorgrams to remove systematic artefacts in the experimental and reference flow cells. To determine the binding affinity (K_d), responses for each analyte were fit to a 1:1 steady state affinity model available within Scrubber.

2.8.21 Crystallisation trials

Details of crystallisation experiments of *Sa*HPPK binary complex have been published in Acta Crystallogr Sect F journal. A copy of the publication is attached in the supporting information of this chapter.

2.9 References

1. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **2003,** 8, 876-7.

2. Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-density miniaturised thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **2001,** 6, 429-440.

3. Bagby, S.; Tong, K. I.; Liu, D. J.; Alattia, J. R.; Ikura, M. The button test: A small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. *J Biomol NMR* **1997,** 10, 279-282.

4. Lo, M. C.; Aulabaugh, A.; Jin, G. X.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **2004,** 332, 153-159.

5. Matulis, D.; Kranz, J. K.; Salemme, F. R.; Todd, M. J. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **2005,** 44, 5258-66.

6. Horita, D. A.; Farnsworth, D. W.; Byrd, R. A. A simple and inexpensive preparation of perdeuterated sorbitol for use as a biomacromolecule stabilisation agent in NMR studies. *J Biomol NMR* **2000,** 16, 339-42.

7. Fernley, R. T.; Iliades, P.; Macreadie, I. A rapid assay for dihydropteroate synthase activity suitable for identification of inhibitors. *Anal Biochem* **2007,** 360, 227-234.

8. Bermingham, A.; Bottomley, J. R.; Primrose, W. U.; Derrick, J. P. Equilibrium and kinetic studies of substrate binding to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Escherichia coli*. *J Biol Chem* **2000,** 275, 17962-7.

9. Motulsky, H. J.; Christopoulos, A. *Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting.*, San Diego CA, 2003.

10. Marley, J.; Lu, M.; Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **2001,** 20, 71-5.

11. Gardner, K. H.; Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1998,** 27, 357-406.

12. Lescop, E.; Lu, Z. W.; Liu, Q.; Xu, H. M.; Li, G. Y.; Xia, B.; Yan, H. G.; Jin, C. W. Dynamics of the Conformational Transitions in the Assembling of the Michaelis Complex of a Bisubstrate Enzyme: A N-15 Relaxation Study of *Escherichia coli* 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase. *Biochemistry* **2009,** 48, 302-312.

13. Cavanagh, J. Protein NMR Spectroscopy: Principles and Practice. In Academic Press, San Diego, Calif.: 1996.

14. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J Biomol NMR* **2001,** 21, 349-59.

15. Hubbard, R. E.; Chen, I.; Davis, B. Informatics and modeling challenges in fragment-based drug discovery. *Curr Opin Drug Discov Devel* **2007,** 10, 289-97.

16. Blaszczyk, J.; Li, Y.; Shi, G.; Yan, H.; Ji, X. Dynamic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: crystallographic studies. *Biochemistry* **2003,** 42, 1573-80.

17. Luft, J. R.; DeTitta, G. T. A method to produce microseed stock for use in the crystallisation of biological macromolecules. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999,** 55, 988-93.

18. Hajduk, P. J. Fragment-based drug design: how big is too big? *J Med Chem* **2006,** 49, 6972-6.

19. Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* **2004,** 9, 430-1.

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

21. Wang, F.; Langley, R.; Gulten, G.; Dover, L. G.; Besra, G. S.; Jacobs, W. R., Jr.; Sacchettini, J. C. Mechanism of thioamide drug action against tuberculosis and leprosy. *J Exp Med* **2007,** 204, 73-8.

22. Otting, G. Protein NMR using paramagnetic ions. *Annu. Rev. Biophys.* **2010,** 39, 387-405.

23. Froger, A.; Hall, J. E. Transformation of plasmid DNA into *E. coli* using the heat shock method. *J Vis Exp* **2007**, 253.

24. Schanda, P.; Kupce, E.; Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **2005,** 33, 199-211.

25. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995,** 6, 277-93.

26. Bartels, C.; Xia, T. H.; Billeter, M.; Guntert, P.; Wuthrich, K. The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. *J Biomol NMR* **1995,** 6, 1-10.

27. Goddard, T. D.; Kneller, D. G. SPARKY 3. *University of California, San Francisco*.

28. Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. NMR-based screening in drug discovery. *Q Rev Biophys* **1999,** 32, 211-40.

29. Mayer, M.; Meyer, B. Characterisation of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed.* **1999,** 38 1784-1788.

30. Papalia, G.; Myszka, D. Exploring minimal biotinylation conditions for biosensor analysis using capture chips. *Anal Biochem* **2010,** 403, 30-35.

2.10 Supporting Information

Crystallization and preliminary X-ray analysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Staphylococcus aureus*

Sandeep Chhabra,[a,b] Janet Newman,[b] Thomas S. Peat,[b] Ross T. Fernley,[b] Joanne Caine,[b] Jamie S. Simpson[a] and James D. Swarbrick[a]*

[a]Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia, and [b]CSIRO Division of Molecular and Health Technologies, 343 Royal Parade, Parkville, Victoria 3052, Australia

Correspondence e-mail: james.swarbrick@pharm.monash.edu.au

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) catalyzes the Mg^{2+}-dependent transfer of pyrophosphate from ATP to 6-hydroxymethyl-7,8-dihydropterin (HMDP), forming 6-hydroxymethyl-7,8-dihydropterin pyrophosphate, which is a critical step in the *de novo* folic acid-biosynthesis pathway. Diffraction-quality crystals of HPPK from the medically relevant species *Staphylococcus aureus* were grown in the presence of ammonium sulfate or sodium malonate and diffracted to better than 1.65 Å resolution. The crystals belonged to space group $P2_1$, with unit-cell parameters $a = 36.8$, $b = 76.6$, $c = 51.5$ Å, $\alpha = \gamma = 90.0$, $\beta = 100.2°$. The crystals contained two molecules per asymmetric unit, with a volume per protein weight (V_M) of 2.04 $Å^3 Da^{-1}$ and an estimated solvent content of 39.6%.

1. Introduction

The increase in rates of bacterial resistance to many antibiotic agents is a major health burden (Cohen, 1992, 1994; Davies, 1994; Neu, 1992; Nikaido, 1994). The number of cases of hospital- and community-acquired *Staphylococcus aureus* infections has grown exponentially together with the parallel emergence of life-threatening 'superbug' methicillin-resistant *S. aureus* (MRSA) strains (Chambers, 2005). Community-aquired strains have become resistant to tetracycline and clindamycin antibiotic treatments (Drew, 2007). In this case, trimethoprim–sulfamethoxazole (TMP–SMX) combination therapy, which synergistically blocks the biosynthesis of folate derivatives by acting on dihydrofolate reductase (DHFR) and dihydrofolate synthase (DHPS), is the main treatment option (Adra & Lawrence, 2004; Proctor, 2008). Resistance to both TMP and SMX has nevertheless emerged, placing a greater burden on vancomycin as an antibiotic of last resort. HIV-infected patients residing in care institutions are a group that is particularly at risk owing to increased exposure to TMP–SMX as an effective *Pneumocystis* treatment (Szumowski *et al.*, 2007). Ever since the introduction of sulfa drugs in the 1930s, their continuing overuse has been blamed for the growing resistance to antibiotics. This is particularly seen to be a problem in developing countries, where overuse is exacerbated by unregulated or inadequate administration. Mechanistically, this has been attributed to several mutations in the DHPS gene. Similarly, with TMP therapy mutations have been observed in the DHFR gene that result in a repositioning of the substrate in the active site (Frey *et al.*, 2009). Thus, with the recent structural data on DHPS and DHFR (Babaoglu *et al.*, 2004; Frey *et al.*, 2009) we are starting to understand the mechanism behind resistance on the atomic level, with the potential to redesign our current chemical approaches. As 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK; EC 2.7.6.3) is not the target of any existing drug, it provides an attractive new folate-pathway target for the rational design and development of novel antimicrobials and antifungals, including those that counter currently resistant isolates.

Folates are essential to all living cells for growth. Accordingly, the enzymes of this pathway (Fig. 1) have been validated as targets for antimicrobials and antifungals. The reduced form of folate, tetrahydrofolate (THF), participates in several important one-carbon transfers that are critical for the biosynthesis of thymidine, glycine

and methionine and vital for DNA replication (Schirch & Strong, 1989). The *de novo* folate-biosynthesis pathway converts 7,8-dihydroneopterin to 7,8-dihydropteroate using ATP and *para*-aminobenzoic acid (*p*ABA). It comprises three committed enzymes, of which HPPK is the second. Initially, dihydroneopterin aldolase (DHNA) catalyzes the epimerization reaction converting 7,8-dihydroneopterin into 6-hydroxymethyl-7,8-dihydropterin (HMDP). HPPK then transfers a pyrophosphate from bound ATP, resulting in 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (HMDPP). DHPS condenses HMDPP with *p*ABA, forming 7,8-dihydropteroate (DHP; Bermingham & Derrick, 2002). The enzymatic addition of glutamate is followed by reduction by DHFR, producing 5,6,7,8-tetrahydrofolate (THF). These latter two reactions are not specific to microbes, plants or protozoa. However, in the case of DHFR structural differences between the human and bacterial enzymes have meant that selective antimicrobials (such as TMP) or, in the case of the human enzyme, cancer treatments such as methotrexate have been successfully deployed.

NMR and X-ray crystal structures of HPPK have been determined from a variety of organisms including *Escherichia coli*, *Haemophilus influenzae*, *Saccharomyces cerevisiae*, *Streptococcus pneumoniae* and *Yersinia pestis* (for reviews, see Swarbrick *et al.*, 2008; Derrick, 2008). Generally speaking, the structures are well conserved and consist of a ferredoxin-like fold comprising a six-stranded β-sheet sandwiched by two α-helices on either side ($\alpha\beta\alpha$ fold). Two Mg^{2+} ions are bound within the active site and these play a major role in the recognition of both the ATP and HMDP substrates. Of the 26 HPPK X-ray and NMR structures that have been deposited, the *S. aureus* homologue shares sequence identities of 39% with the *E. coli*, 39% with the *Y. pestis*, 37% with the *H. influenzae*, 34% with the *S. pneumonia* and 37% with the *S. cerevisae* HPPK enzymes. Much of the understanding of the catalytic mechanism has been deduced from work on the *E. coli* enzyme. In this species, the active-site structure of HPPK is unusually dynamic, as demonstrated by the observation of loop changes in excess of 20 Å throughout the catalytic cycle (Lescop *et al.*, 2009; Blaszczyk *et al.*, 2004). Nevertheless, the high sequence identity of the key active-site residues combined with the high structural similarity of the ternary complexes suggests that inhibitors may have advantageous cross-reactivity potential over different species.

Here, we present a preliminary crystallographic investigation of the *S. aureus* enzyme (*Sa*HPPK) as a prelude towards understanding its catalytic mechanism and as an initial step towards the rational design of novel therapeutics.

2. Materials and methods

2.1. Expression and purification

A pET28a plasmid containing the synthesized *Sa*HPPK sequence (Geneart) was cloned with an N-terminal hexahistidine tag and a thrombin cleavage site. *E. coli* BL21 (DE3) cells transformed with the plasmid were grown overnight in 10 ml 2×YT media supplemented with 100 µg ml^{-1} kanamycin for selection. The overnight culture was then subcultured into fresh 2×YT (1 l) and grown at 310 K until the OD_{600} reached 0.5–0.8. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final optimized concentration of 0.3 mM and expression was carried out at 301 K for 5 h. The cultures were centrifuged at 5000g and 278 K for 10 min and the cells were resuspended in 100 ml 50 mM HEPES, 5% glycerol pH 8.0. An EDTA-free Complete protease-inhibitor cocktail tablet (Roche) was added together with lysozyme; the latter was added to a final concentration of 0.2 mg ml^{-1}. After 10 min, the cells were sonicated, the cell debris was removed by centrifugation at 18 000g for 30 min and the supernatant was filtered (0.45 µm filter).

The supernatant was loaded onto an Ni–NTA IMAC column (Qiagen) and unbound protein was washed off with 10 mM imidazole in 50 mM HEPES–NaOH buffer, 0.3 M NaCl pH 8.0. The tagged protein was eluted from the IMAC column with 250 mM imidazole in the HEPES–NaCl buffer; untagged protein was generated by on-column digestion carried out with thrombin to remove the N-terminal His tag. One unit of thrombin (Sigma–Aldrich) was used per milligram of tagged protein. The cleaved protein was eluted from the column with 10 mM imidazole in the HEPES–NaCl buffer. Both tagged and untagged versions of the protein were further purified on a Superdex 75 size-exclusion 16/60 column (GE Healthcare) and eluted with 50 mM HEPES, 2 mM DTT pH 8.0. Fractions were analysed using a 15% SDS–PAGE gel with Coomassie staining and protein-containing fractions were pooled and concentrated to 4.5 mg ml^{-1} using a 3 kDa molecular-weight cutoff ultrafiltration centrifugal device (Amicon). The yields of tagged and cleaved protein were typically ∼35–50 and ∼15–25 mg per litre of culture, respectively. All samples were snap-frozen and stored at 193 K.
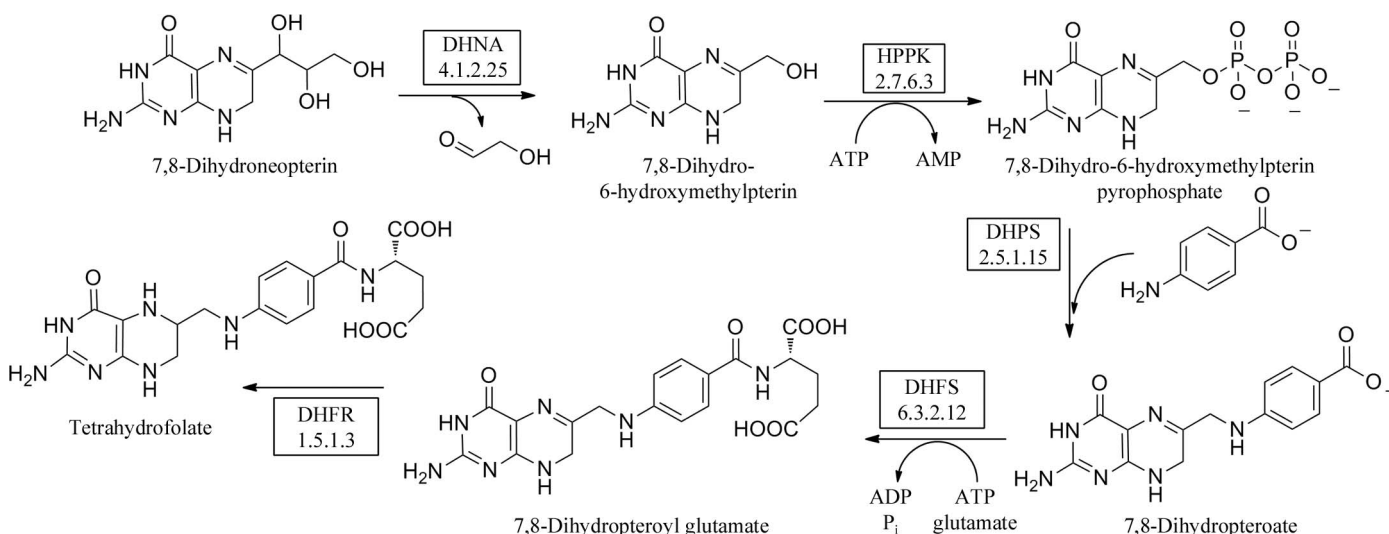
Figure 1
The folate pathway.

2.2. Crystallization

Protein-crystallization experiments (for both the tagged and untagged proteins) were performed at the CSIRO node of the Bio21 Collaborative Crystallization Centre (C3), initially using the PACT and the JCSG+ Suite commercial crystal screens (Qiagen) at 281 and 293 K. A pH *versus* salt gradient screen was also tested at both temperatures. All screens used the sitting-drop vapour-diffusion method with droplets consisting of 250 nl protein solution and 250 nl reservoir solution and a reservoir volume of 50 µl. Crystallizations were performed in SD-2 sitting-drop plates (IDEX Corporation) and were set up using a Phoenix robot (Art Robbins Industries).

Small multiple or twinned crystals were observed after a week in several ammonium sulfate and sodium malonate conditions. A sequential microseeding strategy was employed to improve the size and quality of the crystals. Seeds were created using a seed-bead protocol (Luft & DeTitta, 1999; Newman *et al.*, 2008) and were introduced into the crystallization droplets on setup using the multi-aspirate protocol of a Mosquito robot (TTP Labtech). However, even after several rounds of microseeding optimizations almost all crystals were still partially intergrown. We also tried additive screening with commercially available additive screens (including Additive Screen HT from Hampton Research and The OptiSalt Suite from Qiagen) as well as the addition of different metal ions to find crystallization conditions that would yield single crystals.

2.3. Data collection and analysis

For X-ray data collection, crystals were cryoprotected by adding 1 µl reservoir solution to the droplet and subsequently adding 1.2 µl AP/E Core 150 oil (Mobil-Exxon) to the droplet. Crystals were cryocooled in a nitrogen stream after being pulled through the AP/E oil interface. Data were collected on the microfocus (MX2) beamline of the Australian Synchrotron. 720 images of 0.5° oscillation were collected using an ADSC Quantum 315 detector. The crystal-to-detector distance was 250 mm and the exposure was 1 s for each image. The diffraction data were processed using *MOSFLM* (Leslie, 1992) and *SCALA/TRUNCATE* (Collaborative Computational Project, Number 4, 1994). A diffraction pattern of *Sa*HPPK is shown in Fig. 2.

3. Results and discussion

A problem encountered during the crystallization of *Sa*HPPK was a lack of crystal-growth reproducibility owing to batch-to-batch sample variation. This was particularly onerous as the protein would start to precipitate within 24 h of being purified. Therefore, to ensure protein reproducibility in crystallization experiments small aliquots (100 µl) of the freshly purified and concentrated protein were snap-frozen in thin-walled PCR tubes by immersion in liquid nitrogen and then stored at 193 K. The snap-frozen and rapidly thawed protein was used for both screening and optimization experiments, after side-by-side experiments showed that the frozen and thawed sample crystallized as readily and under similar conditions to the fresh protein.

Crystals of the cleaved *Sa*HPPK protein were observed in conditions containing ammonium sulfate (1.4–2.2 *M*) or sodium malonate (1–1.4 *M*) over a wide pH range. Although crystals of the tagged protein were obtained, these crystals did not diffract X-rays well and were not further pursued. Several rounds of serial microseeding were used to improve the crystal quality of the untagged protein. The crystals grown in ammonium sulfate diffracted to 2.4 Å resolution. The ammonium sulfate condition was refined by the addition of different buffers and various metals (Gd, Mn), which improved the
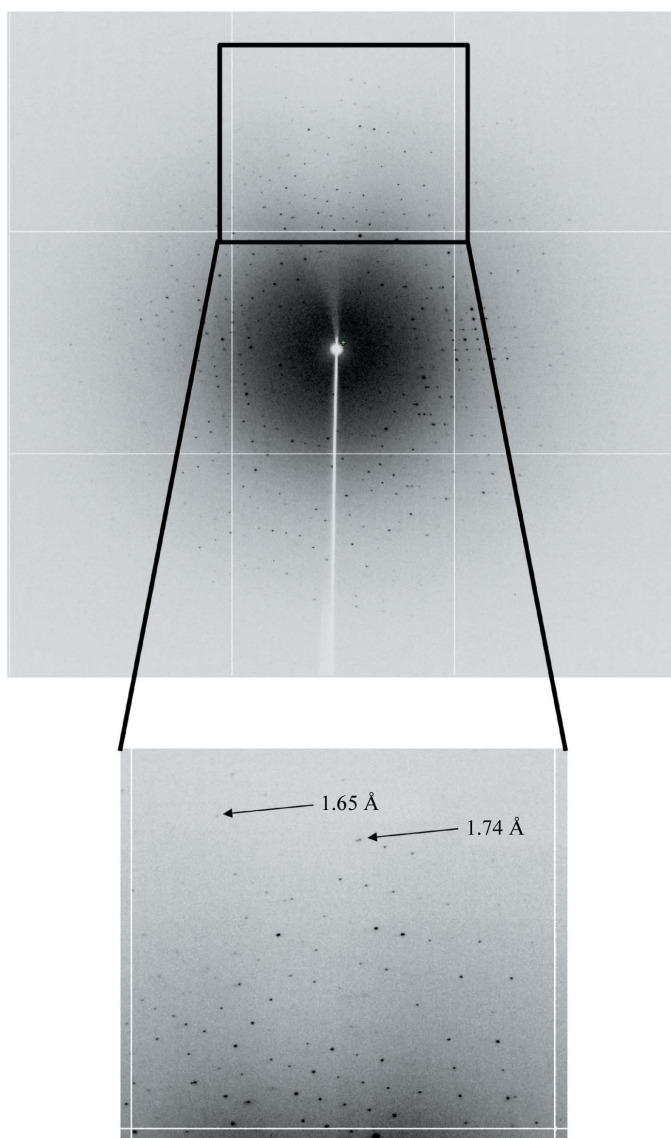
Figure 2
Diffraction image of a typical *Sa*HPPK crystal.
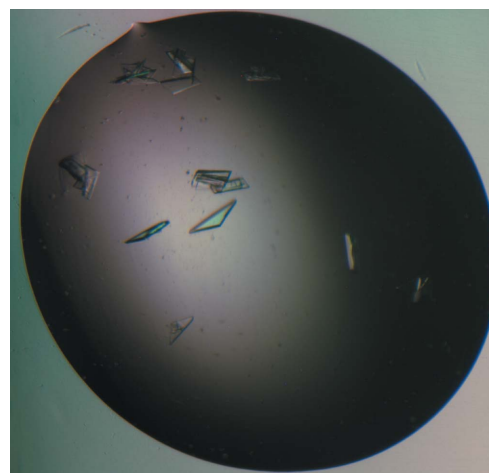
Figure 3
Crystals of *Sa*HPPK. The largest dimension of the central crystal shown is 0.143 mm.

crystallization communications

Protein	6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK)
Mass (Da)	18283
Source organism	*S. aureus*
Crystallization method	Sitting-drop vapour diffusion
Temperature (K)	293
Crystal-growth time	Approximately one week
Seeding	Microcrystals
X-ray source	MX2, Australian Synchrotron
Temperature (K)	100
Oscillation angle (°)	0.5
No. of frames used	720
Crystal-to-detector distance (mm)	250
Exposure time (s)	1
Space group	Monoclinic, $P2_1$
Unit-cell parameters (Å, °)	$a = 36.8$, $b = 76.6$, $c = 51.5$, $\alpha = \gamma = 90$, $\beta = 100.2$
No. of molecules in unit cell Z	2
Resolution range (Å)	50.70–1.65 (1.74–1.65)
No. of unique reflections	33864 (4934)
No. of observed reflections	242194
Matthews coefficient V_M (Å^3 Da^{-1})	2.04
Solvent content (%)	39.6
Completeness (%)	100 (100)
Redundancy	7.2 (7.2)
Mean $I/\sigma(I)$	18.3 (4.7)
R_{merge}† (%)	10.22 (0.447)
$R_{p.i.m.}$‡ (%)	4.1 (17.8)

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$. ‡ $R_{p.i.m.} = \sum_{hkl} [1/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity, $\langle I(hkl)\rangle$ is the average intensity of multiple observations of symmetry-related reflections and N is the redundancy.

diffraction quality of the crystals to about 1.75 Å resolution. We obtained our best diffraction (1.65 Å) from a crystal grown in 1.08 *M* sodium malonate pH 7, 0.09 *M* bis-tris pH 6.5, 0.175 *M* sodium formate, 0.01 *M* sodium acetate pH 4.6 at 293 K. Crystals similar to the one from which data were collected are shown in Fig. 3. Crystals appeared in 5 d and continued to grow to a final maximum length of ~0.140 mm over a further 15 d. In all, close to 70 96-well plates were set up (over 14 000 droplets); only around ten crystals were tested for diffraction as most of the crystals were too intergrown to be candidates for diffraction studies.

All diffraction data were processed in *MOSFLM*; in all cases the crystals were partially multiple but *MOSFLM* was able to pick out a single diffraction pattern and index these reflections despite the presence of more than one lattice. A data set was collected from a crystal grown in the malonate/formate condition described above and cryoprotected with oil, resulting in a data set which was 100% complete (see Table 1).

The space group was assigned as $P2_1$, with unit-cell parameters $a = 36.8$, $b = 76.6$, $c = 51.5$ Å, $\alpha = \gamma = 90.0$, $\beta = 100.2°$. The overall R_{merge} and $R_{p.i.m.}$ were 10.2% and 4.1%, respectively, V_M was 2.04 Å^3 Da^{-1} and the estimated solvent content was 39.6%. The Matthews coefficient (V_M; Matthews, 1968) suggested the presence of two molecules in the asymmetric unit. Molecular replacement was performed with *Phaser* (Collaborative Computational Project, Number 4, 1994) using the HPPK structure from either *H. influenzae* (PDB code 1cbk; 37% sequence identity; Hennig *et al.*, 1999) or *E. coli* (PDB code 1rao; 39% sequence identity; Blaszczyk *et al.*, 2004) as a search model. Two solutions were found, confirming the presence of two molecules in the asymmetric unit, and model building continues. Both models were tested for their ability to phase the *S. aureus* data as one was in a 'closed' (ATP site unoccupied) form and the other was in a 'open' (ATP-bound) form.

HPPK is a committed enzyme in the folate pathway of *S. aureus* and plays an integral role in the biosynthesis of tetrahydrofolate. The determination of the three-dimensional structure of *Sa*HPPK will be instrumental in the development of novel treatments for community-acquired MRSA infections.

References

Adra, M. & Lawrence, K. R. (2004). *Ann. Pharmacother.* **38**, 338–341.
Babaoglu, K., Qi, J., Lee, R. E. & White, S. W. (2004). *Structure*, **9**, 1705–1717.
Bermingham, A. & Derrick, J. P. (2002). *Bioessays*, **24**, 637–648.
Blaszczyk, J., Shi, G., Li, Y., Yan, H. & Ji, X. (2004). *Structure*, **12**, 467–475.
Chambers, H. F. (2005). *N. Engl. J. Med.* **352**, 1485–1487.
Cohen, M. L. (1992). *Science*, **257**, 1050–1055.
Cohen, M. L. (1994). *Trends Microbiol.* **2**, 422–425.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, J. (1994). *Science*, **264**, 375–382.
Derrick, J. P. (2008). *Vitam. Horm.* **79**, 411–433.
Drew, R. H. (2007). *Pharmacotherapy*, **27**, 227–249.
Frey, K. M., Liu, J., Lombardo, M. N., Bolstad, D. B., Wright, D. L. & Anderson, A. C. (2009). *J. Mol. Biol.* **387**, 1298–1308.
Hennig, M., Dale, G. E., D'Arcy, A., Danel, F., Fisher, S., Gray, C. P., Jolidon, S., Muller, F., Page, M. G., Pattison, P. & Oefner, C. (1999). *J. Mol. Biol.* **287**, 211–219.
Lescop, E., Lu, Z., Liu, Q., Xu, H., Li, G., Xia, B., Yan, H. & Jin, C. (2009). *Biochemistry*, **48**, 302–312.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Luft, J. R. & DeTitta, G. T. (1999). *Acta Cryst.* D**55**, 988–993.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Neu, H. C. (1992). *Science*, **257**, 1064–1073.
Newman, J., Pham, T. M. & Peat, T. S. (2008). *Acta Cryst.* F**64**, 991–996.
Nikaido, H. (1994). *Science*, **264**, 382–388.
Proctor, R. A. (2008). *Clin. Infect. Dis.* **46**, 584–593.
Schirch, V. & Strong, W. B. (1989). *Arch. Biochem. Biophys.* **269**, 371–380.
Swarbrick, J. D., Iliades, P., Simpson, J. S. & Macreadie, I. (2008). *Open Enzyme Inhib. J.* **1**, 12–33.
Szumowski, J. D., Cohen, D. E., Kanaya, F. & Mayer, K. H. (2007). *Antimicrob. Agents. Chemother.* **51**, 423–428.

Chapter 3

*Sa*HPPK structure and discovery

of a substrate site inhibitor

Declaration for Thesis Chapter 3

Chapter 3 presents the published article (Chhabra *et al.* 2012) co-authored with Olan Dolezal, Brett M. Collins, Janet Newman, Jamie S. Simpson, Ian G. Macreadie, Ross Fernley, Thomas S. Peat and James D. Swarbrick*.

Declaration by candidate

In the case of Chapter 3, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Protein expression, NMR and X-ray experiments, Thermal shift assay, SPR, Biochemical assays	70

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Olan Dolezal[2]	SPR experiments, intellectual input.	
Brett M. Collins[3]	ITC experiments, intellectual input.	
Janet Newman[2]	Crystallisation experiments, intellectual input.	
Jamie S. Simpson[1]	Enzyme assays, intellectual input.	
Ian G. Macreadie[4]	Assays, intellectual input.	
Ross Fernley[2]	Enzyme assays, intellectual input.	
Thomas S. Peat[2]	X-ray experiments/structure determination, intellectual input	
James D.Swarbrick[1]	Protein expression, NMR experiments/assignments, intellectual input	

Candidate's Signature	████████████	Date 05-04-2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	[1]Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia. [2]CSIRO Division of Materials, Science and Engineering, 343 Royal Parade, Parkville, Victoria 3052, Australia. [3]Institute for Molecular Bioscience, The University of Queensland, 4072 Brisbane, Australia. [4]School of Applied Sciences, RMIT University, Bundoora, Australia.

		Date
Signature 1		02/04/2012
Signature 2		
Signature 3		02/04/2012
Signature 4		04/04/2012
Signature 5		
Signature 6		02/04/2012
Signature 7		02/04/2012
Signature 8		
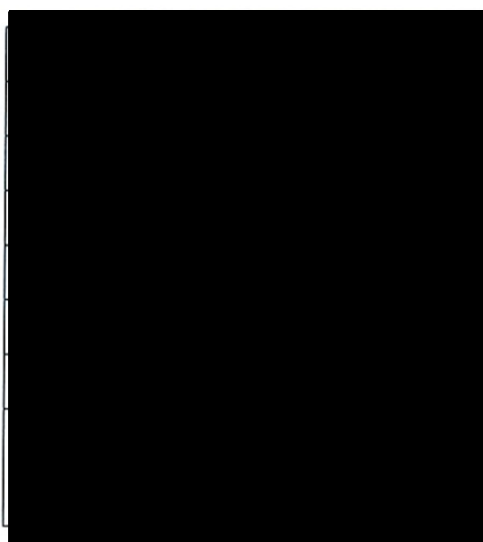
3. SaHPPK structure and discovery of a substrate site inhibitor

3.1 Introduction

The main goal of the thesis was to characterise the structure of *Sa*HPPK as an aid to the discovery of small molecules inhibitors for *Sa*HPPK. As mentioned previously (chapter 2) attempts to grow crystals of apo *Sa*HPPK were unsuccessful as were co-crystallisation trials with weakly binding fragments or the tightly binding cofactor analogue AMPCPP (K_d ~3 μM). Therefore, given the wealth of structural data on the closely related *Ec*HPPK enzyme and our extensive characterisation of *Sa*HPPK by NMR, a ligand-centric computational approach was employed to seek a substrate-like binder. Given that the pterin site is highly specific (containing six inter-molecular H-bond interactions), it was anticipated that by sharing many of these characteristic interactions, a pterin-like inhibitor would have much higher affinity than any weakly binding fragments from the Maybridge RO3 library. Furthermore we hypothesised that a more chemically stable inhibitor, as compared to the native substrate, would hopefully facilitate crystallisation studies of the complex.

To this end, this chapter reports the utility of *in silico* ROCS screening to discover a new pterin site *in vitro* inhibitor of *Sa*HPPK, 8-mercaptoguanine (8-MG), and the subsequent biophysical characterisation of binding using thermal shift, NMR, ITC and SPR. A single enzyme KinaseGlo™ assay was applied successfully for the first time to monitor inhibition of HPPK by 8-MG. While this simple assay measured the kinetics of the cofactor, it was still not suitable to determine the substrate (HMDP) kinetic parameters owing to the rapid turnover of the substrate (K_m = ~ 0.5 μM).

In parallel to crystallisation studies, the binding of the inhibitor, cofactor and substrate were characterised by NMR. The amide resonances displayed extensive line broadening that mapped to the substrate site upon inhibitor binding, characteristic of intermediate timescale exchange kinetics. When

experiments were performed in the presence of saturated ATP or AMPCPP however, slow exchange was observed, suggesting cooperativity and tighter binding in the ternary 8-MG/AMPCPP/*Sa*HPPK complex. This observation was not in agreement with multi-ligand binding studies from SPR and ITC measurements where no cooperativity between the inhibitor and cofactor was found. Slow exchange in NMR is often correlated with 'tight binding' (< low μM K_d). However, in some cases very weak binding (supra millimolar) can give rise to slow exchange as long as the combined lifetime of the on and off rates are long enough, as recently shown by the Pardi group for the binding of Theophylline to RNA (*J. Am. Chem. Soc.* (2009), 131, 5052-5053). Therefore the application of multiple techniques was required to build up a detailed picture of the binding and exchange phenomena of this ligand. The differences in exchange phenomena may be likely due to loop motional effects or an increase in the total lifetime during the hindered binding and hindered leaving of the inhibitor in the presence of AMPCPP.

The lack of non-exchangeable protons on the inhibitor suggested that NMR was not a good strategy to solve the ternary 8-MG/AMPCPP/*Sa*HPPK complex in solution. Fortunately, crystallisation trials produced good diffraction quality crystals and a high resolution structure of the binary *Sa*HPPK/8-MG complex was solved by X-ray crystallography. *Sa*HPPK crystallised with two protomers in the asymmetric unit. NMR showed sharp line widths in the 2D sofast ^{15}N-HMQC spectra inconsistent with a ~40 kDa protein indicating that *Sa*HPPK is monomeric in solution. In the crystal structure, loop L3 was 'extended out' from the active site and the cofactor site is occluded by the protomer interface in the asymmetric unit.

These observations prompted further ^{15}N relaxation (T1, T2 and NOE) and $^{1}D_{HN-N}$ residual dipolar coupling (RDC) NMR experiments to investigate the oligomeric nature, dynamics and structure of *Sa*HPPK in the 8-MG/AMPCPP/*Sa*HPPK complex in solution. Understanding the intermolecular interactions in this ternary structure in particular, would provide a wealth of data and be a handy resource to aid the chemical evolution of the inhibitor. The ^{15}N T1/T2 ratio for amides of *Sa*HPPK clearly showed that the

enzyme was a monomer in solution. While ^{15}N heteronuclear NOE data indicated that active site loops are highly mobile in the apo enzyme and when ATP is bound, in the 8-MG/AMPCPP/*Sa*HPPK complex, part of loop L3 is rigid. This observation suggested that this region (around Gly90) was likely close to the inhibitor. The increase in loop rigidity correlated with a decrease in the entropic contribution to the ΔG of binding as measured by ITC.

Amide residual dipolar coupling (RDC) values were measured to determine the alignment tensor and to then characterize the structure of the 8-MG/AMPCPP/*Sa*HPPK ternary complex in solution and to compare this with the high resolution 8-MG/*Sa*HPPK X-ray structure and to interrogate the unusual loop L3 'out' conformation observed therein.

RDC experiments provide information on the orientation of the [1]H-^{15}N vector in solution with respect to the principle axis system of the alignment tensor (Δχ) which describes how the molecule tumbles within an alignment media (a PEG/hexanol media was used in this case) in the magnetic field. The RDC derives from the orientation-dependent dipolar couplings of two proximal spins that occur in the presence of an external magnetic field (B_0) and is calculated from the axial and rhombic components for the alignment tensor.

$$RDC_{AB}^{calc} = -\frac{B_0^2 \gamma_B \gamma_A \hbar S}{120 kT \pi^2 r_{AB}^3} \left[\Delta\chi_{ax} \frac{2z_{AB}^2 - x_{AB}^2 - y_{AB}^2}{r_{AB}^2} + \frac{3}{2}\Delta\chi_{rh} \frac{x_{AB}^2 - y_{AB}^2}{r_{AB}^2} \right]$$

where γ_A and γ_B are the gyromagnetic ratios for spin A and B, $\hbar$ is the Plank constant divided by 2π, S is the order parameter describing any residual motion of the vector in the protein frame, r_{AB} is the internuclear distance, and x_{AB}, y_{AB} and z_{AB} are the coordinates of the vector AB expressed in the Δχ tensor frame of reference.

Usually in solution, dipolar interactions between the pairs of magnetic nuclei are averaged to zero by rapid, random molecular tumbling and leads to an

efficient mechanism to relaxation. When a small tunable degree of alignment, typically in the order of fraction of a percent, is imparted on a protein in solution by dissolving the protein in a suitable alignment media, the dipole-dipole couplings are observed. These RDCs are measured similarly to J couplings and can range in the order of tens of hertz. They can take both positive and negative values and can be zero when the ^{1}H-^{15}N vector points along the magic angle that makes a value of (θ = 54.7°) to the z direction of the principal axis of the alignment tensor.

Residual dipolar coupling experiments (^{1}D$_{HN}$ RDC) on the 8-MG/AMPCPP/*Sa*HPPK complex demonstrated that the RDC value for Gly90 amide in loop L3 fits closely to the ternary complex of *Ec*HPPK (1Q0N) which suggested that loop L3 is likely in a 'closed' conformation. The large chemical shift change of Gly90 further suggested that Gly90 is making a close contact with sulfur atom of the inhibitor and the high value of the ^{15}N NOE implied that it is rigid.

The details of the results of this chapter have been published in PLoS ONE Journal and publication is as follow:

3.2 PLoS ONE publication

Structure of *S. aureus* HPPK and the Discovery of a New Substrate Site Inhibitor

Sandeep Chhabra[1,2], Olan Dolezal[2], Brett M. Collins[3], Janet Newman[2], Jamie S. Simpson[1], Ian G. Macreadie[4], Ross Fernley[2], Thomas S. Peat[2], James D. Swarbrick[1]*

1 Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Australia, 2 CSIRO Division of Materials, Science and Engineering, Parkville, Australia, 3 Institute for Molecular Bioscience, The University of Queensland, Australia, 4 School of Applied Sciences, RMIT University, Bundoora, Australia

Abstract

The first structural and biophysical data on the folate biosynthesis pathway enzyme and drug target, 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (SaHPPK), from the pathogen *Staphylococcus aureus* is presented. HPPK is the second essential enzyme in the pathway catalysing the pyrophosphoryl transfer from cofactor (ATP) to the substrate (6-hydroxymethyl-7,8-dihydropterin, HMDP). In-silico screening identified 8-mercaptoguanine which was shown to bind with an equilibrium dissociation constant, K_d, of ~13 μM as measured by isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR). An IC_{50} of ~41 μM was determined by means of a luminescent kinase assay. In contrast to the biological substrate, the inhibitor has no requirement for magnesium or the ATP cofactor for competitive binding to the substrate site. The 1.65 Å resolution crystal structure of the inhibited complex showed that it binds in the pterin site and shares many of the key intermolecular interactions of the substrate. Chemical shift and ^{15}N heteronuclear NMR measurements reveal that the fast motion of the pterin-binding loop (L2) is partially dampened in the SaHPPK/HMDP/α,β-methylene adenosine 5′-triphosphate (AMPCPP) ternary complex, but the ATP loop (L3) remains mobile on the μs-ms timescale. In contrast, for the SaHPPK/8-mercaptoguanine/AMPCPP ternary complex, the loop L2 becomes rigid on the fast timescale and the L3 loop also becomes more ordered – an observation that correlates with the large entropic penalty associated with inhibitor binding as revealed by ITC. NMR data, including ^{15}N-^1H residual dipolar coupling measurements, indicate that the sulfur atom in the inhibitor is important for stabilizing and restricting important motions of the L2 and L3 catalytic loops in the inhibited ternary complex. This work describes a comprehensive analysis of a new HPPK inhibitor, and may provide a foundation for the development of novel antimicrobials targeting the folate biosynthetic pathway.

Introduction

Staphylococcus aureus is a clinically important opportunistic pathogen and one of the major contributors to hospital- and community-acquired bacterial infections. Methicillin-resistant *S. aureus* strains (MRSA, commonly referred to as the ''superbug'') cause up to 19,000 deaths annually in the US alone, and an estimated health care cost of $ 3–4 billion per annum [1]. MRSA strains are classified by genotypic and phenotypic characteristics, and are grouped into two major categories: those originating in hospitals (nosocomial, haMRSA, strains USA100 and USA200) and those in the community (caMRSA), of which the latter is almost entirely caused by the pandemic USA300 strain [2]. Infection with USA300 causes abscesses and life threatening systemic infections, such as bacteremia and necrotizing pneumonia. In contrast to haMRSA, caMRSA infections tend to occur in previously healthy younger patients without health care exposure [3]. Currently, caMRSA is more susceptible to a range of chemotherapies than the multi-drug resistant haMRSA [4]. Although resistant to tetracycline, erythromycin, clindamycin, linezolid, and in some cases vancomycin, caMRSA is largely susceptible to TMP-SMX (trimethoprim-sulfamethoxazole) combination therapy, which synergistically blocks the biosynthesis of folate derivatives by acting on dihydrofolatereductase (DHFR) and dihydropteroatesynthase (DHPS), respectively [5,6]. TMP-SMX resistance has emerged in haMRSA owing to an 'autolytic' thyamidine salvage pathway effective when polymerized DNA is released from damaged tissues [6,7,8]. TMP-SMX resistance in caMRSA is attributed to mutations in the DHFR or DHPS genes, which in the former case results in a repositioning of the substrate in the active site [9], compromising TMP-based therapy.

Classically, targets for antimicrobials are found to be essential enzymes that are unique to the micro-organism (not present in the host), and new antimicrobial drugs have been developed from molecules identified in proof-of-concept studies [10]. The folate biosynthetic pathway fits the criterion of being an attractive source of potential target enzymes, and antimicrobials against key components of this pathway are used today to treat diseases such as malaria, pneumocystis pneumonia (PCP) and caMRSA infections. Folates are essential for the growth of all living cells. The reduced form of folate, tetrahydrofolate (THF), participates in several important one-carbon transfers, critical for the biosynthesis

of thymidine, glycine and methionine, and is vital for DNA replication [11,12].

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK, EC 2.7.6.3) catalyses pyrophosphoryl transfer from ATP (cofactor) to the substrate, 6-hydroxymethyl-7,8-dihydropterin (HMDP) (Fig. 1). HPPK is the upstream and adjacent enzyme to DHPS in the folate biosynthesis pathway (Fig. 2*A*). It is not the target of any existing drug and therefore represents an attractive resource for the rational design of novel antimicrobials and antifungals to act on current TMP-SMX-resistant isolates for the treatment of caMRSA infections. HPPK is a small (158 residues, ∼18 kDa), generally monomeric protein and has been studied using various biophysical techniques, including x-ray crystallography and NMR spectroscopy. A number of x-ray and NMR structures of HPPK have been determined in various ligand-bound states (Fig. 2*B*) and from a variety of organisms: *Escherichia coli, Haemophilusinfluenzae, Saccharomyces cerevisiae, Streptococcus pneumonia, Yersinia pestis* and *Francisella tularensis* [13,14,15,16,17,18]. These data have provided atomic level information on the catalytic mechanism and protein dynamics of the reaction trajectory during catalysis [19]. Three loop regions, loops L1–3, play an important role in substrate recognition and are critical for assembling the active centre [20]. While loop L3 undergoes the largest and most dramatic conformational change during the catalytic cycle, all three loops help to seal the substrate and cofactor binding sites for the chemical transfer of a pyrophosphate from ATP to HMDP [21]. The substrate and cofactor interact with two magnesium ions and associate with a total of 26 residues in HPPK, 13 of which are conserved across all species [22]. *In vitro* kinetic studies have shown a preferred order of substrate binding. At cellular levels of magnesium, the ATP binds first, followed by HMDP [23]; in the absence of cofactor and magnesium, HMDP binds weakly *in vitro* to the *apo* enzyme [24]. Both active sites are highly selective for their ligands. For example, the affinity of *E. coli* HPPK (EcHPPK) for Mg-GTP is 260-fold less than for Mg-ATP [25]. Remarkably, only two specific pterin-site inhibitors have been reported in the literature [26]. Both are based on the pterin substrate (Fig. 1), one featuring gem-dimethyl substitution at the C7 position on the pyrimidine ring, the other a phenethyl substituent at the same position. Bisubstrate analogues of the former have been reported that display sub-micromolar affinity, which demonstrates the feasibility of developing new inhibitors based on bisubstrate-linking strategies [27].

S. aureus HPPK (SaHPPK) shares 34–39% sequence homology with HPPK enzymes from other species whose structures have been determined (Fig. 2*C*). High conservation of active site residues, and high structural similarity among all HPPK structures, suggests that HPPK inhibitors developed for one species may have advantageous cross-reactivity over many different species.

Herein, we report the first structural studies of HPPK from *S. aureus* using a combination of solution NMR and x-ray crystallographic structure determination, and the identification of a novel pterin-site inhibitor 8-mercaptoguanine (Fig. 2*D*) by *in silico* ROCS screening (Rapid Overlay of Chemical Structures) and differential scanning fluorimetry (DSF) assay. The atomic structure of SaHPPK has been determined in complex with a new pterin-site inhibitor, revealing the molecular details of inhibitor association. Binding of the inhibitor, substrate and cofactor molecules were quantified using isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR), while *in vitro* enzyme inhibition data was measured using a luciferase based luminescent assay. Detailed studies of ligand interactions using NMR highlight critical ligand-induced dynamic changes upon inhibitor, substrate and cofactor binding, which correlate with large entropic penalties to the binding thermodynamics of the inhibitor measured by ITC.

Results

Enzyme stability

Buffer conditions were screened to overcome initial solubility problems with SaHPPK using a differential scanning fluorimetry (DSF) assay [28] and button tests [29]. Final optimized buffer conditions from both techniques correlated well, with SaHPPK found to be most stable in a buffer comprised of 50 mM HEPES (pH 8.0), 1% sorbitol and 2.0–5 mM DTT.

Validation of ROCS virtual screening hits using the DSF assay

To identify potential new binders of SaHPPK, we adopted a high-throughput virtual screening approach (see methods) using ROCS (Rapid Overlay of Chemical Structures) [30]. The DSF assay used for buffer stability screening was subsequently employed as an efficient method to screen the 37 purchased ROCS compound library. DSF is a rapid, convenient and inexpensive assay to detect binding of ligands to proteins. It compares the change in the unfolding transition temperature (ΔT_m) of a protein obtained in the presence and absence of a ligand [31,32]. A ligand that binds to the protein generally causes an increase in the protein thermal melting temperature (T_m).

We tested the utility of the assay using ATP (and AMPCPP) binding. ATP and AMPCPP bind relatively tightly to EcHPPK, with dissociation constants of 2.6 and 0.45 μM, respectively [24]. The unfolding temperature (T_m) of SaHPPK increased by 6 and 11°C in the presence of saturating ATP and saturating AMPCPP, respectively, which is consistent with the previously observed tighter binding of AMPCPP to EcHPPK, as well as the results of our ITC and SPR experiments (*vide infra*). Inspired by these

Figure 1. Pyrophosphoryl transfer catalysed by HPPK.
doi:10.1371/journal.pone.0029444.g001

A

DHNPt   DHNA

ATP   [HPPK]

pABA   DHPS

Glu   DHFS

DHFR → THF

B

L2

L3

G90

L1

L3

HMDP

D95

D97

AMPCPP

C

Loop L1   Loop L2

```
            1                                                                                              *  *  *                      *    *                  63
S. aureus   - M I Q A Y L G L G S N I G D R E S Q L N D A I K I L N E Y N G I S V S N I S P I Y E T A P V G Y T E Q P - - - - - N F L N - L C V E I Q T
E. coli     - - T V A Y I A I G S N L A S P L E Q V N A A L K A L G D I P E S H I L T V S S F Y R T P P L G P Q D Q P - - - - - D Y L N - A A V A L E T
H. influ    - M I T A Y I A L G S N L N T P V E Q L H A A L K A I S Q L S N T H L V T T S S F Y K S K P L G P Q D Q P - - - - - D Y V N - A V A K I E T
B. anthrac  M N N I A Y I A L G S N I G E R - - - - - Y T F L N K - N P Y I K V E D V S S V Y E I E P V G Y T D Q S - - - - - C F L N - L V I K I S T
M. tuberc   - M T R V V L S V G S N L G D R L A R L R S V A D G L G - - - - D A L I A A S P I Y E A D P W G G V E Q G - - - - - Q F L N A V L I A D D P
P. carinii  - - E A V Y I S L G S N L G N R I K F I L D A I E K M S - I K G I K V L K T S M L Y E S K P M Y F K D Q P - - - - - A F Y N - A V C K V Q T
S. cerev    - - N R V V L G L G S N L K N P L K I L K N C F L Y F K N H S K I G K I F S S P I Y I N P P F G Y T N Q P - - - - - N F Y N - A T I I L R T
H. pylori   S W K R A F L A F G S N I G D R F K H I Q M A L Q L L S R E K T V K L R N I S S I F E S E P M Y F K D Q T - - - - - P F M N - G C V E V E T
F. tular    - M V Q Y I I G I G T N S G F T I E N I H L A I T A L E S Q Q N I R I I R K A S L Y S S K A V L K E D A P K E W D I R F L N - T A V K I S S
```

Loop L3

```
            63                    #           #               #            #  *                *           #                              #     # #          #    122
S. aureus   T L T V L Q L L E C C L K T E E C L H R I R - - K E R W G P R T L D V D I L L Y - - - - - - - G E E M I D L P - K L S V P H P R M N E R A
E. coli     S L A P E E L L N H T Q R I E L Q Q G R V R K - A E R W G P R T L D L D I M L F - - - - - - - G N E V I N T E - R L T V P H Y D M K N R G
H. influ    E L S P L K L L D E L Q R I E N E Q G R V R - - L R R W G E R T L D L D I L L F - - - - - - - G N E I I Q N E - R L T I P H Y D M H N R E
B. anthrac  N L S - - - - - - - T Q K V E N D L G R K R - - E I R W G P R T I D L D I L L Y - - - - - - - N Q E N I E A E - N L I V P H P R M F E R A
M. tuberc   T C E P R E W L R R A Q E F E R A A G R V R - - G Q R W G P R N L D V D L I A C Y Q T S A T E A L V E V T A R E N H L T L P H P L A H L R A
P. carinii  S L H P E Q L L F E L Q L I E K E L G R V K V - I D K - G P R C I D L D I V F Y - - - - - - - G R K I I N S E - S L I I P H P R V L E R S
S. cerev    C L D L R E F F A L V F Y I E R R F G R A R K R D F K D A P R T L D I D I I A F - - - - - - - N Q V I L R Q N D L T L P H P K W S K R D
H. pylori   L L T P S E L L K L C K K I E Y E E L Q R - - - - - - - - - - T I D L D I V M F L N S A - - - - G E D I I V N E P D L N I P H P R M L E R T
F. tular    S L K P D E L L V L L K D I E L K I G R D L N - A P A W S P R V I D L D I L A A - - - - - - - E D L I L E T D K L T I P H K E L I N R S
```

```
            123  *                                                             158
S. aureus   F V L L P L N D - - - - T A A N V V E P - R S K L K V K D L V F V D D S V K R Y K - - - - - -
E. coli     F M L W P L F E - - - - I A P E L V F P - D G E M L R Q I L H T R A F D K L N K W - - - - - -
H. influ    F V I V P L F E - - - - I A S D L V L P - N S Q I I T E L V K Q F A D H K M I K L N P - - - -
B. anthrac  F V L P L L E I N Q D I K Q N I S R S Q V E E M K R R E G V T V W K Q K N G E D A F V L F E N
M. tuberc   F V L I P W I A - - - - V D P T A Q L T V A G C P R P V T R L L A E L E P A D R - - - - - - -
P. carinii  F V L K P L L D - - - - I S G D L V H P V T G L S I A S Y F E K I V D H D I K P - - - - - -
S. cerev    S V L V P L T L - - - - Q Q I L F K - K G E W - - - - - - - - - - - - - - - - - - - -
H. pylori   F V L E P L C E L - - - I S P V H L H P V T A E P I V D H L K Q L Y D K Q H D E D T L W K - - -
F. tular    F A L A P L L E - - - - L S K G W H H P - - - K Y V E W D L N I R L K E L - - - - - - - - -
```

D

Figure 2. Folate pathway, HPPK structures and HPPK sequences. A, Schematic showing the folate pathway conversion of dihydroneopterin (DHNPt) into tetrahydrofolate (THF) via the five essential enzymes, DHNA, HPPK, DHPS, DHFS and DHFR. B, Superposition of selected HPPK structures in complex with HMDP (orange), 8-mercaptoguanine (elemental colouring) and AMPCPP (cyan), showing the prototypical HPPK fold, active site loops, Gly90 (red), the HPPK conserved ring stacking residues and magnesium ions (orange) bound by the conserved aspartate residues. SaHPPK (3QBC) from this study is shown (yellow). Three EcHPPK structures are shown in complex with AMPCPP and HMDP (1Q0N, green); apo (1HKA, blue); with HMDPP and AMP (1RAO, magenta) and HPPK from *H. Influenza* (1CBK, grey). C, Multiple sequence alignment of selected HPPKs. Absolutely conserved residues (orange), highly conserved residues (blue), and residues that are additionally important for HPPK function and may be targeted to develop selective inhibitors are shown (green). Residues involved in HMDP (*) and ATP recognition (#) and elements of secondary structure are displayed above the alignment. Underlined sequences have had their structure determined by x-ray crystallography. D, Structure of the SaHPPK inhibitor 8-mercaptoguanine.
doi:10.1371/journal.pone.0029444.g002

encouraging results, this rapid assay was used to screen the ROCS-generated library, using a single ligand concentration of 500 μM in 5 μM enzyme. Out of the 37 compounds tested, one compound, 8-mercaptoguanine, produced an increase in T_m of 3.4°C (as compared to *apo*), confirming binding to SaHPPK (Fig. 3A). Moreover, a similar increase in T_m was observed in the presence or absence of ATP/AMPCPP, suggesting non-competitive binding with the cofactor.

Biochemical assay

A KinaseGlo^{TM} assay was performed to test SaHPPK function and examine whether 8-mercaptoguanine inhibits SaHPPK catalysis.

While a K_m^{app} value of 10.8±2.5 μM was readily obtained for ATP, the assay is insufficiently sensitive to allow reliable determination of the relatively low K_m for HMDP, consistent with those determined for the *E. coli* enzyme (0.7–1.6 μM) using different assays [23,33]. The assay is however suitable to show inhibition and gave an apparent IC_{50} of 41±9 μM for 8-mercaptoguanine (Fig. 3B).

Thermodynamics of ligand binding by ITC

To determine the affinity and thermodynamic parameters of ligand binding to SaHPPK, we employed isothermal titration calorimetry (ITC) (Table 1; Fig. 4). Previous studies using fluorescence-based methods have reported a binding affinity (K_d)
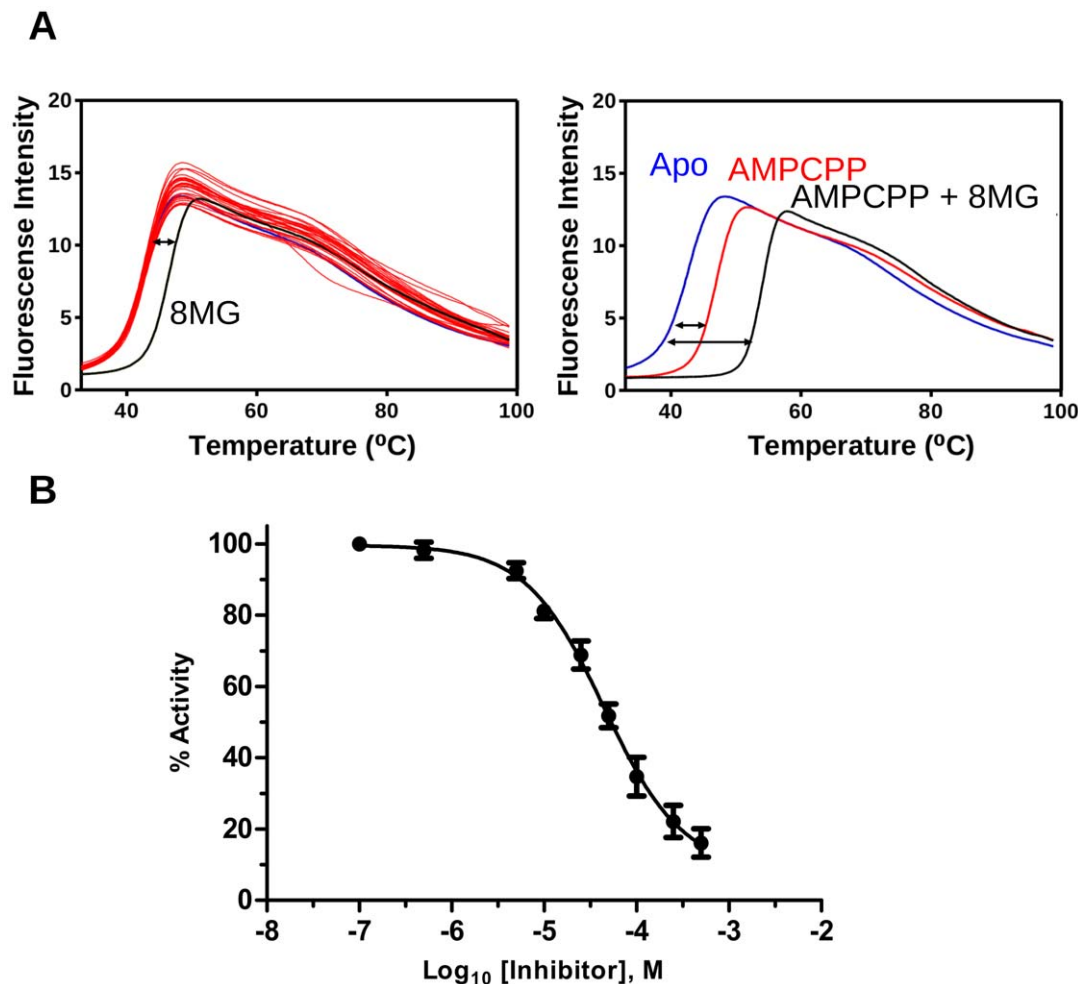
Figure 3. DSF and enzyme assays. A, Superposition of the DSF assay profiles of 37 ROCS compounds (*left*) showing the curve for 8-mercaptoguanine (8MG) (black) and thermal shift ΔT, double headed arrow) relative to the apo enzyme (blue) or relative to the AMPCPP complex (*right*). The DSF assay for AMPCPP is also shown (red). B, Dose response curves for the inhibition of SaHPPK by 8-mercaptoguanine ($IC_{50}=41$ μM). HMDP and ATP concentrations were fixed at 0.3 μM and 0.2 μM respectively.
doi:10.1371/journal.pone.0029444.g003

Table 1. Thermodynamic and kinetic parameters for the binding of ligands to SaHPPK, as determined by ITC[a] and SPR[a].

Ligand	ΔH (kCal mol^{-1})	$T\Delta S$ (kCal mol^{-1})	ΔG (kCal mol^{-1})	N	K_d (µM) ITC	K_d (µM) SPR
ATP	−6.5±0.4	−0.3±0.5	−6.2±0.1	0.98±0.11	31.0±4.5	45±2
ADP	nd[c]	nd[c]	nd[c]	nd[c]	nd[c]	760±16
AMP	nd[c]	nd[c]	nd[c]	nd[c]	nd[c]	4900±1100
AMPCPP	−6.3±1.0	1.3±1.0	−7.5±0.2	1.06±0.12	3.1±1.2	7.7±0.4
HMDP	nd[c]	nd[c]	nd[c]	nd[c]	nd[c]	100±12
HMDP+AMPCPP[b]	−10.5±1.0	−3.1±0.9	−7.4±0.2	1.23±0.15	4.0±1.2	3.6±0.3
8MG[e]	−19.6±3.4	−12.9±3.5	−6.7±0.2	1.00±0.06	12.8±3.4	10.8±0.4
8MG[e] +AMPCPP[b]	−17.2±1.4	−10.4±1.5	−6.8±0.1	0.96±0.09	11.3±2.3	nd
8MG[e] +ATP[d]	nd	nd[c]	nd[c]	nd[c]	nd[c]	8.1±1.6

[a]Values are the means ± the standard deviation for at least three experiments. See *Materials and Methods* for sample concentrations used. All ITC and SPR experiments were performed at 298 K and 293 K respectively.
[b]Concentration of AMPCPP was 1 mM in both HPPK (cell) and ligand (syringe) solutions.
[c]No data.
[d]Concentration of ATP was 5 mM.
[e]8-mercaptoguanine.
doi:10.1371/journal.pone.0029444.t001

of EcHPPK for ATP of 2.6–4.5 µM, and for AMPCPP of 0.08–0.45 µM in the presence of Mg^{2+} [23,24,25]. Binding affinities of EcHPPK for HMDP substrate vary from 0.036 to 0.17 µM in the presence of Mg^{2+} and AMPCPP [23,24].

ATP and AMPCPP were shown to bind exothermically to SaHPPK with affinities of 31 and 3.1 µM, respectively at 298 K (Table 1; Fig. S1A). In the absence of Mg^{2+}, no significant binding was observed for either compound (data not shown). These affinities are lower than those reported for EcHPPK, but we observe a very similar affinity ratio for the two compounds, with AMPCPP displaying approximately 10-fold tighter binding than ATP. The inhibitor was found to bind SaHPPK with a K_d of 12.8±3.4 µM (Fig. 4A), with AMPCPP having no effect on the affinity or thermodynamics of interaction. The interaction occurs with a large favourable binding enthalpy but with a significant entropic cost (see thermodynamic parameters in Table 1).

In saturating AMPCPP, the substrate (HMDP) bound with a K_d of 4.0±1.2 µM (Fig. 4B), which is again significantly weaker than that determined for EcHPPK by fluorescence methods [23,24]. In the presence of the inhibitor 8-mercaptoguanine, however, binding was reduced to a level that could not be detected by ITC. Therefore, it appears that the inhibitor binds competitively to the same binding site as the substrate, affecting enzyme inhibition accordingly. In line with previous studies, we observed that HMDP binding is highly dependent upon the prior incorporation of ATP, or the analogue AMPCPP. Although a weak binding signal is observed for HMDP without AMPCPP, a satisfactory fit to the data could not be obtained, which is consistent with the low affinity estimated from SPR experiments (see below).

Kinetics of ligand binding by SPR

Binding interactions of SaHPPK substrates, AMPCPP and 8-mercaptoguanine were analysed via surface plasmon resonance (SPR) using minimally biotinylated SaHPPK immobilized onto a NeutrAvidin chip surface (Table 1, Fig. 4 and Fig. S1). In this manner, we were able to obtain highly active SaHPPK surfaces not compromised by a low pH required for amine coupling method [34]. K_d values for ATP and AMPCPP were determined

to be 45±2 and 7.7±0.4 µM, respectively, in close agreement with the values determined by ITC (Table 1, Fig. S1A). The slower dissociation of AMPCPP also allowed the binding affinity to be derived by fitting to a kinetic interaction model, yielding rate and affinity parameters of $k_a = 8.8±1.3×10^4$ M^{-1} s^{-1}, $k_d = 0.5±0.1$ s^{-1} and a similar $K_d = 5.4±0.3$ µM (Fig. S1 A). Based on this result, we conclude that the higher affinity of AMPCPP compared to ATP is mainly driven by a slower dissociation rate parameter (k_d) for AMPCPP. Weak binding of ADP and AMP (Table 1, Fig. S1B), beyond the detection limit of ITC, could be estimated by steady-state fitting in SPR, extending the upper limit of detection. Binding of HMDP to SaHPPK in the absence of the cofactor was not readily detectable by ITC under the conditions used here, but could be approximated by SPR (K_d~100 µM). The binding affinity for the SaHPPK/HMDP interaction measured by SPR in the presence of AMPCPP ($K_d = 3.6±0.3$ µM) was in excellent agreement with the value obtained by ITC ($K_d = 4.0±1.2$ µM) (Table 1, Fig. 4B). Furthermore, SPR measurements confirmed that binding of 8-mercaptoguanine to SaHPPK was independent of cofactor, as estimated affinity values in the presence and absence of ATP were approximately the same (10.8±0.4 uM and 8.1±1.6 uM, respectively) (Table 1, Fig. 4A).

Backbone assignments of SaHPPK with substrate and cofactor

The chemical shift of a nucleus is sensitive to changes in its local environment and is thus a convenient probe for analysing ligand binding events and detecting conformational changes. To investigate the structure and substrate binding properties of the enzyme, we thus assigned the backbone resonances of SaHPPK under various ligand conditions using heteronuclear NMR spectroscopy and compared the change in the weighted average resultant ^{15}N and ^{1}HN chemical shift vector (chemical shift perturbations, CSPs). All ^{15}N HSQC spectra e.g. Fig. S2, showed well-dispersed sets of resonances, consistent with a folded enzyme.

No change over time was observed in the ^{15}N HSQC spectra of the SaHPPK/HMDP/AMPCPP complex under the sample conditions (see methods), at least for a period up to ~36–48 hrs
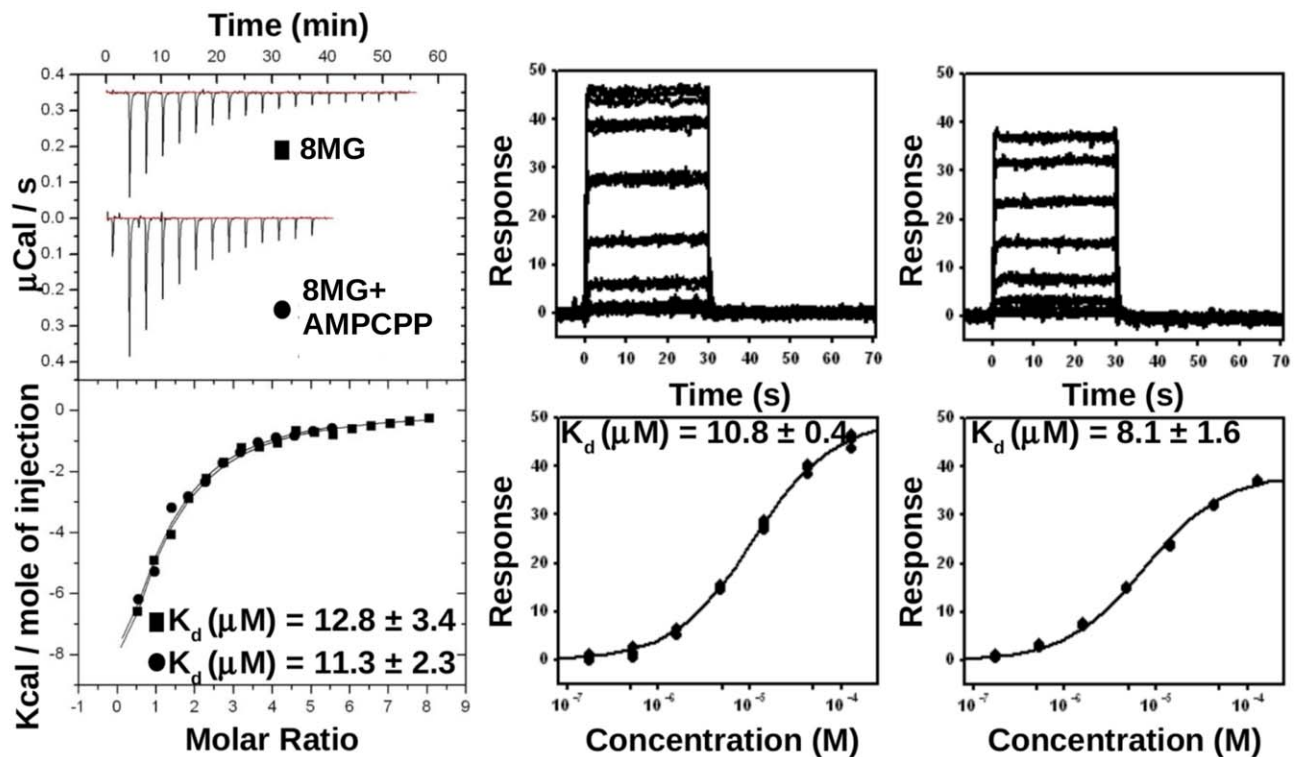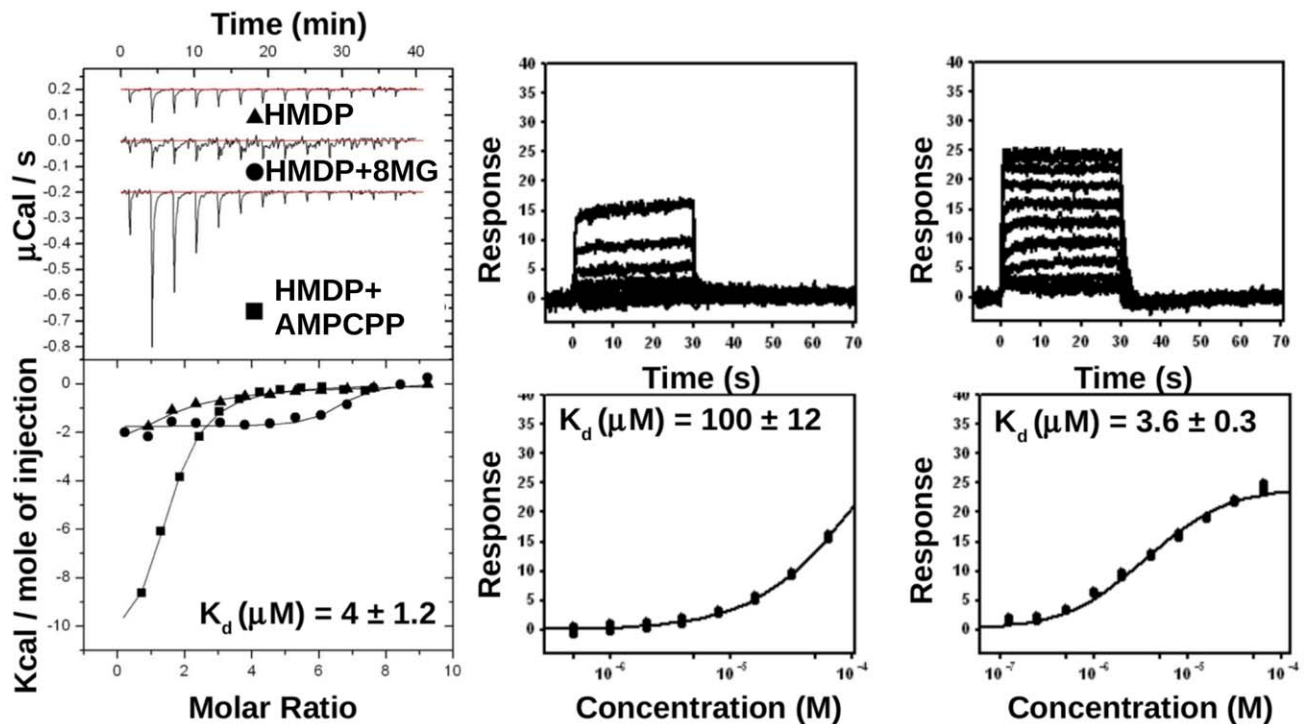
Figure 4. Binding of ligands to SaHPPK as measured by ITC and SPR. A, Raw (*top left*) and integrated ITC data (*bottom left*) for the titration of 30 μM SaHPPK with 300 μM 8-mercaptoguanine (8MG) alone (■), and 300 μM 8-mercaptoguanine in the presence of 1 mM AMPCPP (●). SPR raw data (*middle* and *right top*) and steady-state response curve (*middle* and *right bottom*) for the binding of 8-mercaptoguanine in the absence (*middle*) and presence (*right*) of 5 mM ATP. B, Raw ITC data (*top left*) and integrated normalised data (*bottom left*) for titrations of 22 μM SaHPPK with 500 μM

HMDP (▲) in the absence of 8-mercaptoguanine and in the presence of 300 µM 8-mercaptoguanine (●), and in the presence of 1 mM AMPCPP (■).
Binding of HMDP was only detected in the presence of AMPCPP. SPR raw data (*middle* and *right top*) and steady-state response curve (*middle* and *right bottom*) for the binding of HMDP in the absence or presence of saturated AMPCPP.
doi:10.1371/journal.pone.0029444.g004

(Fig. S3), which was long enough to record 3D experiments to make assignments and record relaxation experiments. Oxidation of HMDP has been reported in previous x-ray studies [22], however, while some degradation was observed in the 1D ^{1}H of isolated HMDP alone over time (data not show), spectral evidence for a bound hydroxymethylpterin oxidation product or degradation species were not observed (Fig. S4). It appears that degradation over time leads to a weakly bound product which is not competitive to the tight binding substrate. Therefore, for the first time we report NMR data on the binding and dynamics of the substrate to HPPK.

Under physiological conditions, the SaHPPK enzyme is expected to be saturated with ATP, given the binding constant of ~45 µM (Table 1, Fig. S1A). Titration of SaHPPK with ATP or AMPCPP revealed slow exchange on the NMR timescale. Substantial CSPs were observed for both cases and large perturbations were localized to the ATP site (Fig. S2). A titration of SaHPPK with fresh HMDP to saturation levels showed extensive chemical shift broadening (or disappearance) compared to the *apo* ^{15}N HSQC spectrum, and few residual CSPs. The broadening mapped to residues within loop L2 (Fig. S2B, C) and residues around the pterin subsite, including those of the β-sheet, underneath the substrate. Adding HMDP to the saturated AMPCPP complex, on the other hand, revealed slow exchange, characteristic of tighter binding. Perturbations clearly mapped to the pterin site, as expected (Fig. S2B, C) yet a few resonances in loop L2 still remained broadened.

Given that the inhibitor has no non-exchangeable ^{1}H NMR signals and standard NMR techniques would only poorly characterise the intermolecular interactions, we adopted a parallel investigation of the x-ray structure in complex with 8-mercapto-guanine.

X-ray crystallographic structure of SaHPPK in complex with 8-mercaptoguanine at 1.65 Å resolution

The x-ray crystal structure of SaHPPK in complex with the inhibitor, 8-mercaptoguanine was solved at high resolution (Fig. 5). Crystallisation conditions were as published [35]. Briefly, the SaHPPK/8-mercaptoguanine binary complex crystallised in the P2$_1$ space group, with the asymmetric unit comprising two protein molecules that contain a single bound 8-mercaptoguanine molecule per monomer and a total of 256 water molecules (Fig. 5A). Density was observed for all 158 amino acid residues of the protein, although the density for residues 85–91 was very weak in monomer B. Two (in chain A) or three (in chain B) additional non-native residues are seen as a result of the *N*-terminal thrombin cleavage site.

SaHPPK has a ferredoxin-like fold (αβα), with a central core of six β strands surrounded by four helices, typical of other monofunctional HPPK structures. The two monomers in the asymmetric unit are almost identical, with an RMSD of 0.34 Å over all 158 pairs of backbone Cα atoms. The catalytic loops L1, L2 and L3 (residues 12–14, 47–51, 82–94) have higher than average temperature factors, showing that they are likely mobile in solution, which parallels the observed broadening of residues in solution by NMR. The dimer interface encapsulates the active sites, leading to a buried surface area of 1595 Å, not including the inhibitor molecules. Eight inter-subunit hydrogen bonds are observed.

Using the SSM algorithm in Coot [36], we calculated 36% sequence identity between the structures of SaHPPK and

EcHPPK (1RAO), with a RMSD of 1.45 Å over the 148 aligned residues. While the structure of SaHPPK deviates significantly from the EcHPPK in the region leading out to the C terminus (residues 150–156) (Fig. 2B), secondary structural elements are generally very well conserved as they are across all HPPK enzymes (Fig. 2B), with major variations occurring mostly in the catalytic loops, dependant on the types of bound ligands or catalytic stage.

Comparison of 8-mercaptoguanine and HMDP binding

The crystal structure confirms that 8-mercaptoguanine binds to the pterin pocket in a similar pose to the HMDP substrate bound in EcHPPK (1Q0N) (Fig. 5B, C). Mercaptoguanine and HMDP share the same pyrimidine ring, which in both cases makes hydrogen bond contacts with several highly conserved residues; T43, A44, V46, and N56 saturate the hydrogen bond donor/acceptor sites (positions 1, 2, 3, 4, and 9) of the inhibitor (Fig. 5C). Like HMDP, the pyrimidine ring of 8-mercaptoguanine is stacked between the conserved aromatic residues, Phe54 and Phe123. A small cavity is found near the N7 position of 8-mercaptoguanine, in which a single water molecule resides, forming hydrogen bonds with the sidechain of Asp95 and the N7 of 8-mercaptoguanine. Notably, two of the interactions formed between HMDP and HPPK are absent in the case of 8-mercaptoguanine. Firstly, the imidazole ring of 8-mercaptoguanine is too far away to interact with Asp95, which forms a hydrogen bond with the hydroxymethyl group of HMDP. Secondly, a van der Waals interaction with Trp89 is missing. Loop L3 is displaced out of the active site and the Trp89-8-mercapto-guanine distance is around 25 Å, compared to ~4 Å in the HMDP/AMPCPP-bound structure of EcHPPK (pdb 1Q0N).

Comparison of loop positions

The three loops of HPPK (L1–L3) are highly dynamic in nature, changing conformation during the enzymatic cycle (Fig. 2B) [37]. Loops L1 and L2 undergo relatively minor structural changes compared to loop L3, for which the apex moves over 20 Å throughout the cycle (Fig. 2B). Loop L2 in SaHPPK caps the substrate active site and resembles closely the loops in the HMDP/AMPCPP (1Q0N) and pterin analogue-bound forms (1DY3, 1CBK). Loop L3 in SaHPPK is "extended" out from the active site and, as such, resembles most closely the loop position in the product-bound EcHPPK structure (1RAO). It is excluded from the active site by the 108–111 loop from the other monomer in the asymmetric unit (Fig. 5C, D).

Although we grew crystals routinely in the presence of 2 mM AMPCPP and 10 mM Mg^{2+}, we did not observe density for either, even when the concentrations of these species were increased to 25 and 50 mM, respectively. We therefore performed NMR measurements on the ternary SaHPPK/8-mercaptoguanine/AMPCPP complex.

Chemical shift mapping of inhibitor complexes by NMR spectroscopy

Titration of 8-mercaptoguanine into a sample of the *apo* enzyme produced a range of CSPs (Fig. 6A) and exchange regimes in the NMR spectra. 15 cross peaks broadened completely around the binding site and along the sheet (Fig. S2D) and most others exhibited slow exchange, indicative of a K_d likely in the low µM range. In contrast, when performed in the presence of saturating ATP or AMPCPP, widespread perturbations were observed in
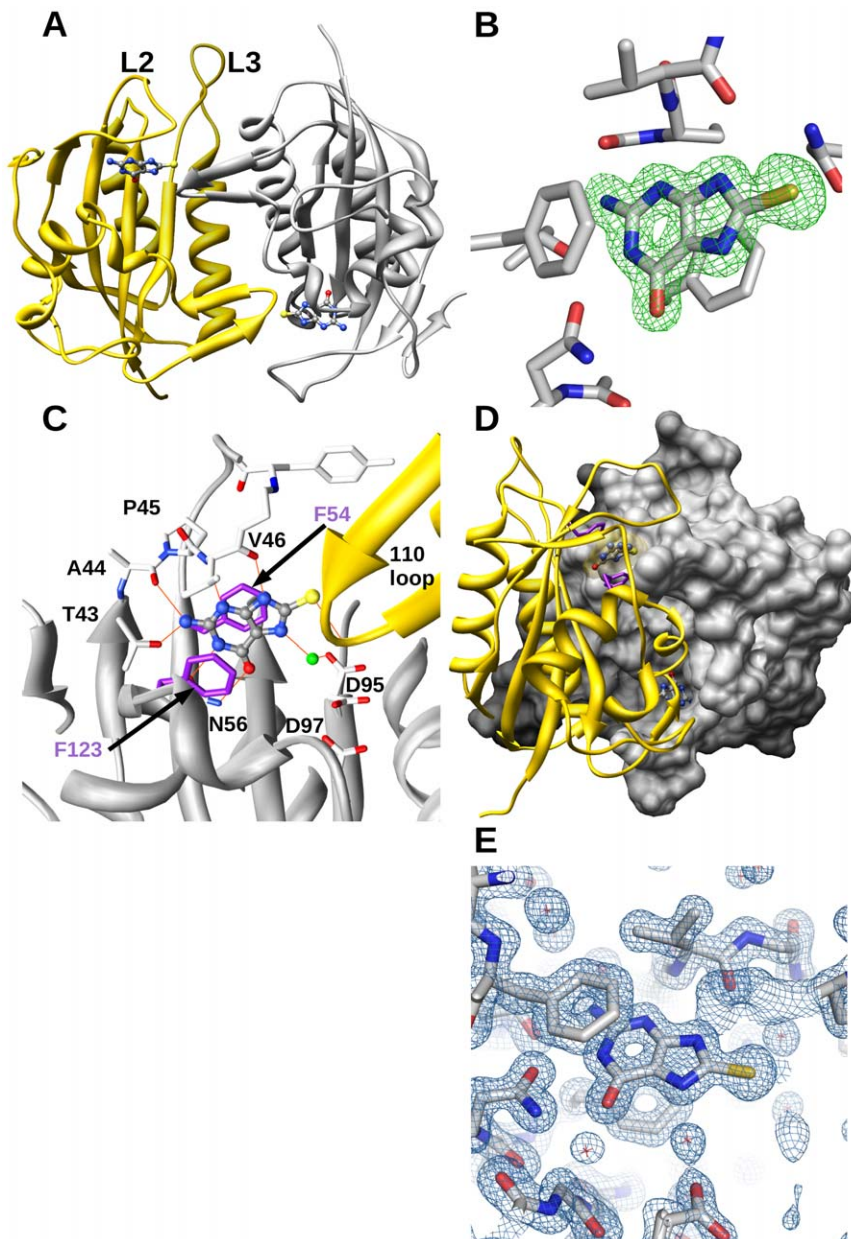
Figure 5. X-ray structure of SaHPPK with the 8-mercaptoguanine inhibitor bound. A Ribbon schematic showing the crystallographic dimer of theSaHPPK-8-mercaptoguanine binary complex. B, F_0-F_c difference density map of 8-mercaptoguanine contoured at 3.0 sigma. C, Intermolecular interactions (orange) involving 8-mercaptoguanine. Two side chain conformers for Asp95 were modelled and are shown. The second monomer is drawn in yellow ribbon. The oxygen atom of a water molecule forming hydrogen-bonds with 8-mercaptoguanine and SaHPPK is shown in green. D, Space filling representation of the dimer interface, showing 8-mercaptoguanine bound in the deep pterin pocket, sandwiched between Phe123 and Phe54 (purple), and illustrating the proximity to the 110 s loop in the other monomer. E, Final 2F_0-F_c electron density map of the pterin binding site, contoured at 2.0 sigma.
doi:10.1371/journal.pone.0029444.g005

slow exchange for all resonances, despite no change in binding affinity measured by ITC and SPR. Chemical shift perturbations clearly mapped to the respective substrate and cofactor site (Fig. S2*D*).

All ^{15}N and ^{1}HN amide chemical shifts are tabulated in Table. S1 and values for CSPs are shown in Fig. S5.

SaHPPK is a monomer in solution

The ^{15}N T1/T2 ratio for amides of SaHPPK in the presence of 8-mercaptoguanine and AMPCPP correlated well with that calculated using HYDRONMR [38] for a monomeric enzyme with a correlation time of ∼12 ns (Fig. S6). This clearly showed

that the enzyme exists as a monomer in solution at a concentration less than ∼200 µM. Line widths were similar for the SaHPPK/8-mercaptoguanine binary complex.

The linewidth and chemical shift of Gly90 amide indicates a loop ordering in the vicinity of the inhibitorfor the inhibited ternary complex

We observed that the amide of Gly90 displayed a large upfield CSP, and sharpened significantly, when comparing the spectrum of the SaHPPK/AMPCPP complex to that of the SaHPPK/8-mercaptoguanine/AMPCPP complex (Fig. 6*C*). In addition, a similar but smaller shift was observed in the SaHPPK/HMDP/AMPCPP
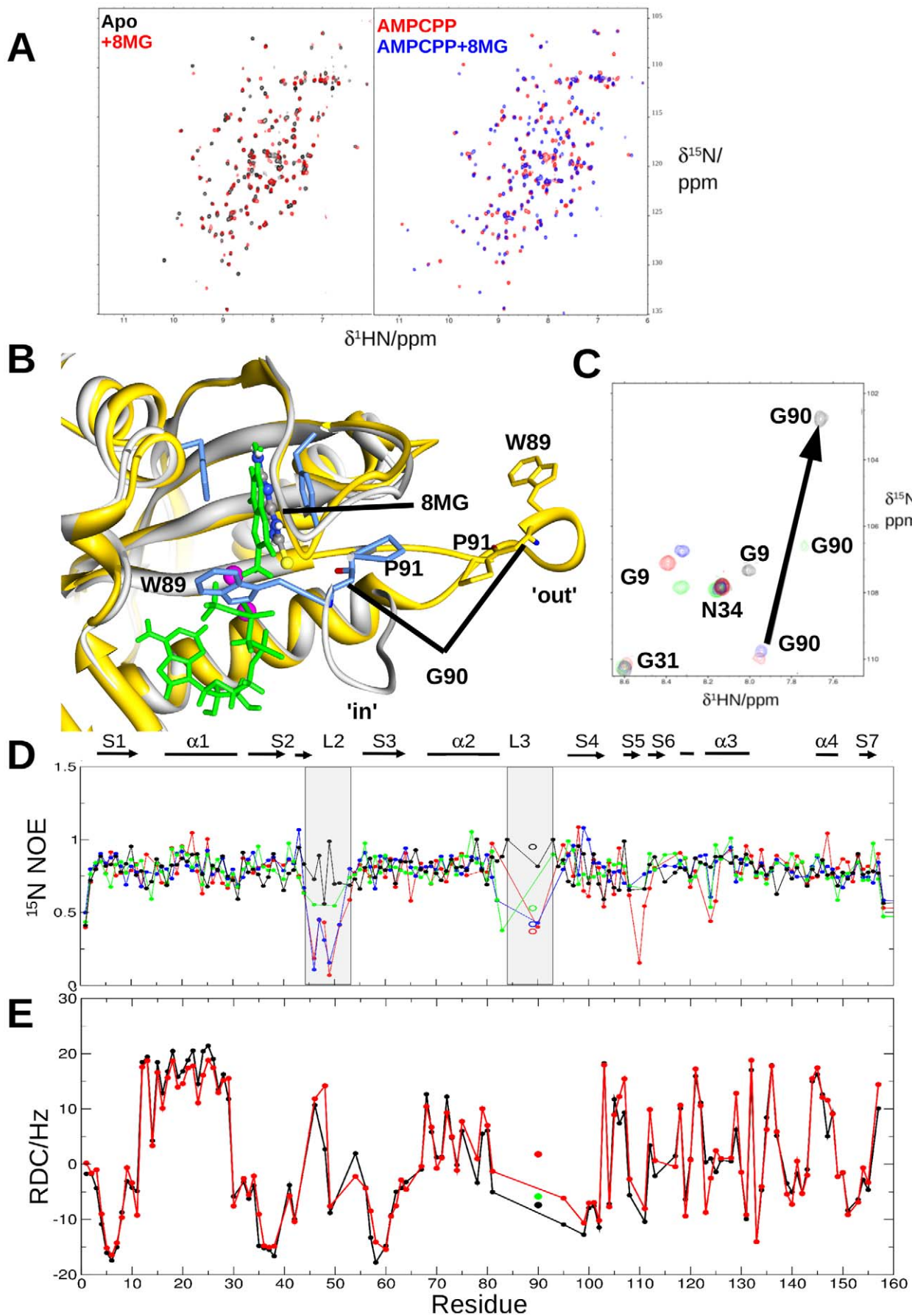
Figure 6. Binding of 8-mercaptoguanine (8 MG) to SaHPPK by NMR spectroscopy reveals an 'in' conformation of loop L3. A, Superposition of the ^{15}N HSQC spectra of ~100 µM SaHPPK with and without 0.6 mM 8-mercaptoguanine (*left*), and the ^{15}N HSQC spectra of ~100 µM SaHPPK+1 mM AMPCPP+10 mM Mg^{2+} with and without the addition of 600 µM 8-mercaptoguanine. B, Superposition of the EcHPPK/AMPCPP/HMDP (1Q0N, grey ribbon, blue sidechain, green ligands) with SaHPPK (yellow). The loop L3 'in' conformation and the loop 'out' conformation are shown for each respectively and the sidechain of Trp89 and Pro91. C, Region of the superposed ^{15}N HSQC spectra of *apo*SaHPPK (red), SaHPPK/ATP complex (blue), SaHPPK/HMDP/AMPCPP complex (green) and SaHPPK/8-mercaptoguanine/AMPCPP complex (black). The large CSP of G90 is shown. D, Superposition of the ^{15}N heteronuclear NOE values recorded at 600 MHz. The colour coding is the same as in C). Open circles are for the Hε1 of Trp89. The extent of the secondary structural elements is shown at the top, with the region corresponding to the active site loops, L2 and L3, highlighted. E, Comparison of observed (black) and calculated (red) ^{1}D$_{HN}$ RDCs for SaHPPK/8-mercaptoguanine/AMPCPP. The ^{1}D$_{HN}$ RDC for G90 from EcHPPK/HMDP/AMPCPP (1Q0N) is shown (green filled circle).
doi:10.1371/journal.pone.0029444.g006

complex and Gly90 was noticeably broadened (Fig. 6*C*). Such CSP reflects a dramatic change in the environment and a decrease in local chemical exchange effects respectively, likely associated with a decrease in the motion of loop L3 on the µs-ms timescale. Further evidence for a decrease in loop L3 dynamics is supported by more extensive assignments in loop L3 for the SaHPPK/8-mercaptoguanine/AMPCPP spectra relative to all other spectra (compare the ribbon diagrams in Fig. S2*B* with Fig. S2*D*).

^{15}N fast timescale dynamics uncovers distinct loop dampening in the inhibited ternary complex

To understand the fast timescale motion of SaHPPK in solution we recorded ^{15}N heteronuclear NOE data for the *apo* enzyme, the binary SaHPPK/ATP complex and the SaHPPK/8-mercaptoguanine/AMPCPP ternary complexes (Fig. 6*D*). Data for the 8-mercaptoguanine or HMDP binary complexes were not recorded owing to extensive broadening around the substrate site.

In all cases, the last and first residues showed relatively low ^{15}N NOE values (~0.4–0.6) and are therefore partially mobile on the ps-ns timescale. The amides of Val46–Gln51 in loop L2 were also mobile for the ATP-bound and *apo* samples. Gly90 was found to be mobile on fast timescale in *apo* and ATP-bound SaHPPK. Val124, the residue adjacent in sequence to the ring-stacking Phe123, appeared to be mobile on the ps–ns timescale (NOE~0.6) with all substrates. While Lys110 is mobile in the apo enzyme signal overlap precluded assessment of this residue in the complexes.

For the ternary SaHPPK/8-mercaptoguanine/AMPCPP complex, apart from Tyr48, the fast motions within loop L2 diminish and the loop essentially becomes rigid around the inhibitor (Fig. 6D). This is in contrast to the SaHPPK/HMDP/AMPCPP complex which remains partially mobile. While the fast motions of Gly90 and the sidechain Hε1 of Trp89 are considerably dampened in the inhibited ternary complex compared to all other complexes, there is evidence for slower underlying motion particularly in the 84–88 region as judged by resonance broadening. Finally, the fast timescale motion that was evident in Val124 (^{15}N NOE~0.5) in all other spectra also appears to dampen and this amide becomes rigid in the SaHPPK/8-mercaptoguanine/AMPCPP complex.

^{15}N Residual dipolar couplings (RDCs) support a closed loop L3 conformation for the inhibited ternary complex in solution

To understand the structure of the inhibited ternary complex in solution we recorded ^{1}D$_{HN}$ residual dipolar couplings (RDCs) of each amide. These are induced by the weak alignment of a biomolecule in solution and report the angle a given N-H vector makes within the principal axis system of the alignment tensor. Weak alignment of SaHPPK was achieved with a PEG/hexanol alignment media [39] from which RDCs up to ~20 Hz in magnitude were measured (see methods). These allowed us to derive the alignment tensor by single value decomposition within PALES [40] by fitting the measured ^{1}D$_{HN}$ ^{15}N RDCs to our x-ray structure (Fig. 6E). The striking agreement (Pearson's correlation

coefficient (R^2) = 0.97, Q factor = 24%) between the ternary SaHPPK/8-mercaptoguanine/AMPCPP complex in solution with the binary SaHPPK/8-mercaptoguanine x-ray structure shows that, overall the binary structure is a very good model for the ternary structure in solution. Similarly oriented amides within helix 1 for example (residues 14–28) lie parallel to the principal axis of the alignment tensor as evidenced by a string of fairly uniform RDCs with maximal value. Helix 2 on the other hand shows smaller RDCs and larger sinusoidal fluctuations as it is tilted away from the principal axis.

A few RDCs stand out as outliers. Interestingly, these generally map to those amides in close proximity to a ligand or metal. Residues 123–125 showed very small RDCs (0.38, 1.0, and −1.42 Hz respectively) and their calculated values (−8.7, −3.8, 4.8 Hz) result in statistically significant deviations (>~5 Hz). Low values of the RDC can result from motional averaging of the amide vector or if the NH vector points close to the magic angle with respect to the Z axis of the alignment tensor. As the ^{15}N NOE values are high for residues 123–125, the RDC data suggest small tilts of these amide vectors in solution towards the magic angle compared to those of the x-ray structure. Calculations show that rotations of 18, 36 and 22° giving small displacements of 0.3, 0.6 and 0.4 Å respectively in the amide proton coordinate can accommodate these observed RDC values.

The ^{1}D$_{HN}$ RDC for Tyr48 (2.7 Hz) deviates from the calculated (14.2 Hz) and given the ^{15}N NOE~0.55 is most likely due to motion. The measured RDC for Asp95 (−10.9 Hz) deviated markedly from that calculated from the SaHPPK/8-mercaptoguanine x-ray structure (−3.3 Hz), but agreed very well to the amide orientation calculated (−12.3 Hz) for the EcHPPK/HMDP/AMPCPP ternary structure (1Q0N) indicating that a ~30° reorientation of its H-N vector is likely in the SaHPPK/AMPCPP/8-mercaptoguanine ternary complex. Interestingly, for the Gly90 amide in loop L3, the agreement with our x-ray structure is very poor indeed and of opposite sign, deviating by over 9 Hz. However, there is a close agreement (1.5 Hz) with the EcHPPK/HMDP/AMPCPP (1Q0N) orientation (Fig. 6E).

Discussion

This work reports the discovery, binding properties and mechanism of a novel, competitive pterin site inhibitor, presented in complex with the first crystal structure of SaHPPK. The pterin site is highly specific and restricts the chemical space available for inhibitor design to structures closely resembling the pterin scaffold. Consequently, the literature is devoid of non-pterin like HPPK inhibitors [41], despite mounting structural information that has been reported over the last decade. In line with the high pterin-site specificity is the high ligand efficiency (2.3 kJ/heavy atom or K_d~13 µM over 12 heavy atoms) of 8-mercaptoguanine.

8-Mercaptoguanine has previously been reported to have biological activity. Early studies revealed some lipolytic activity [42] while in a number of cases 8-mercaptoguanine has been

shown to inhibit enzymes that normally bind purines [43,44,45,46]. Antiviral activity,without significant toxicity, was also reported in an *in vivo* mouse model [43]. Close analogues, such as 8-mercaptoguanosine, were also shown to induce interleukin-1 activity in macrophages [47]. Despite these studies, no antibacterial activity has been reported previously. Interestingly, 8-mercaptoguanine has been shown to bind to, but not inhibit, *B. anthracis* DHPS by co-crystallisation [48], which may open the possibility for a multi target inhibitor derived from this scaffold. In the present work, we did not observe growth inhibition *in vivo* by 8-mercaptoguanine in *E. coli* cell-based assays (data not shown). Given the unfavourable logP (-0.39), this is likely to be due to poor membrane permeability. This may be a disadvantage for pterin-like inhibitors in general given the hydrophilic nature and restrictive chemical space of the pterin scaffold in folate pathway enzymes. Nevertheless, while insufficient transport of a set of closely related pyrimidines as potential antifolates was implicated in their poor *in vivo* inhibition, derivatives with an additional phenyl substituent displayed sub micromolar activity *in vivo* to *T. brucei* and *L. major* [49]. The known phenethyl *in vitro* inhibitor of HPPK (pdb 1DY3) suggests that a suitably positioned phenyl group on 8-mercaptoguanine may thus be beneficial to both binding and assist cell permeability.

Given that 8-mercaptoguanine forms exactly the same number of inter-molecular hydrogen bonds as the substrate heterocyclic rings, an intriguing finding from our work is that, in the absence of cofactor, the inhibitor binds some 10-times more tightly ($K_d{\sim}12\ \mu M$) than HMDP ($K_d{\sim}100\ \mu M$ by SPR). In the HMDP/AMPCPP complex the higher affinity ($K_d{\sim}4\ \mu M$) of substrate can be rationalised by the observation of a hydrogen bond to the Mg^{2+} bound γ-phosphate of AMPCPP from the HMDP hydroxyl as seen in the EcHPPK/HMDP/AMPCPP (1Q0N) ternary complex [21,41]. We therefore hypothesised that the large increase in the ΔG of binding might be associated directly with the sulfur atom in 8-mercaptoguanine, in the absence of substantial de-solvation or structural differences. While our SaHPPK structure is a valuable resource and will support further inhibitor design strategies towards the phosphate sub-site, the 'out' position of loop L3 in our crystallographic dimer precluded us from drawing many conclusions about the nature of the interaction between the sulfur atom of the inhibitor and enzyme. Nevertheless, we have established from several different types of NMR data that residue Gly90 is in close contact with the sulfur atom of the inhibitor in solution (Fig. 6).

In the EcHPPK/HMDP/AMPCPP ternary complex (1Q0N), the tip of loop L3 is observed closed 'in' over the active site with the Trp89 sidechain Hϵ1 hydrogen bonded to the terminal phosphate. A superposition of this ternary complex with our crystal structure indicates that the sulfur atom of the inhibitor would lie ${\sim}4.6$ Å from the sidechain of Trp89 and only ${\sim}3.9$ Å from the N atom of Gly90 (Fig. 6B). The importance of this proximity is likely to account for a specific interaction to the inhibitor and thereby stabilizing loop L3 into the 'in' position. For this loop arrangement and in a protonated thiol tautomer of the inhibitor, a hydrogen bond is predicted between the thiol proton and the backbone carbonyl of Gly90. The large ^{15}N chemical shift perturbation (${\sim}7$ ppm),^{15}N spin relaxation data and NMR linewidth considerations (Fig. 6) is evidence that the Gly90 amide is in close proximity and, along with the sidechain of Trp89, becomes essentially rigid in the ternary SaHPPK/inhibitor/AMPCPP complex. A direct interaction with the inhibitor is not possible from a solvent exposed Gly90/Trp89 in a loop L3 'open' arrangement. In agreement with an important role of the sulfur atom of 8-mercaptoguanine in HPPK binding, guanine, the sulfur-free analogue of the inhibitor, displays markedly reduced affinity that was

undetectable by SPR (data not shown) and gave few CSPs in the NMR spectra (data not shown). Finally, the close agreement of the ^{15}N RDC for Gly90 to the EcHPPK ternary complex structure (1Q0N) (Fig. 6E) is evidence that G90 amide is oriented in solution as that in 1Q0N and thus further evidence of the loop 'in' conformation.

Active site motions and associated changes in loop conformations are an intrinsic requirement for enzyme activity [50] and for the structural transitions along the catalytic pathway of HPPK [51]. We propose that the thiol-mediated stabilization of loop L3 in turn stabilizes interactions to the neighbouring substrate loop L2. ITC data is in agreement with ^{15}N relaxation data and showed a significant-binding entropy penalty for 8-mercaptoguanine binding to the SaHPPK/AMPCPP complex supporting the contention that the loops L2 and L3 (particularly around Gly90) become more ordered. In contrast, the lower entropic penalty for the binding of the substrate to SaHPPK/AMPCPP is in accord with a partially mobile loop L2 and loop L3 which is likely to be functionally relevant to facilitate subtle structural changes, during pyrophosphoryl transfer. Inhibition by 8-mercaptoguanine therefore may be derived in part from impeding a catalytic role of loop L3.This notion is consistent with the observation that the loop L3 deletion in EcHPPK did not affect ATP affinity, but produced a 10^5-fold decrease in the rate constant for catalysis [20].

Intermolecular interactions to loop regions, may compromise an inhibitor as a potential long-term antimicrobial candidate. It is well documented that point mutations in loop regions have resulted in rapid sulfa drug resistance in the downstream DHPS enzyme [12,48]. Given the C8 thiol of 8-mercaptoguanine seems to make an important backbone, rather than sidechain interaction with the highly conserved Gly90 or other natively small residue substitutions (Ala or Ser) in other species, (Fig. 2), this portion of the molecule may be a beneficial component of the pharmacophore in future chemical elaboration.

Over the last decade there has been a growing interest in using RDC calculations as a powerful additional parameter for the validation and refinement of macromolecular structures [52,53]. Recently, chemical shift changes in conjunction with $^1D_{HN}$ RDC measurements have revealed ligand induced conformational changes in the active site loop in ABL kinase [54]. Here, we have established unequivocally that the SaHPPK/8-mercaptoguanine/AMPCPP ternary structure in solution is essentially the same as the x-ray structure of the binary inhibited complex (for all residues that we could measure $^1D_{HN}$ RDCs) and that loop L3 is likely 'in' (Fig. 6E). Furthermore, several NOEs to the H2 and H8 of AMPCPP were observed from amides 98, 111 and 112 in the 3D ^{15}N NOESY HSQC experiment, consistent with the pose of the adenine found in the EcHPPK structure (1Q0N) (data not shown). These data show that our structure is also a good working model of the adenine-binding pocket.

In summary, our multi-disciplinary study shows that 8-mercaptoguanine readily binds to SaHPPK with high ligand efficiency and dampens loop motions by making specific interactions to both catalytic loops. It inhibits enzyme function *in vitro* and thus presents as an important scaffold for development of SaHPPK inhibitors with increased potency and more favourable pharmacokinetic properties. Conservation of the binding site within HPPK enzymes from *E. coli*, *H. influenza*, *Y. pestis*, *S. cerevisiae* and *F. tularensis* indicates that 8-mercaptoguanine may also be active against a range of other bacterial and lower eukaryotic enzymes. Accordingly, 8-mercaptoguanine may present a novel scaffold for future broad spectrum antibiotic development in the treatment of *S. aureus* and other pathogenic infections. We are currently exploring elaboration of the 8-mercaptoguanine scaffold to this end.

Table 2. X-ray structure data processing and refinement statistics.

Spacegroup	Monoclinic, *P*21
x-ray source	MX2, Australian Synchrotron
Detector	ADSC Quantum 315
Wavelength (Å)	13000 eV (0.96Å)
Unit-cell parameters (Å, °)	a = 36.8, b = 76.6, c = 51.5 90, $\alpha = \gamma = 90.0$, $\beta = 100.2$,
Diffraction data	
Resolution range (Å)	50.70-1.65 (1.74-1.65)
No. of unique reflections	33864 (4934)
No. of observed reflections	242194
Matthews coefficient, V_M (Å^3 Da^{-1})	2.04
Solvent content (%)	39.6
Completeness (%)	100 (100)
Data redundancy	7.2 (7.2)
Mean I/σ(I)	18.3 (4.7)
R_{merge} (%)*	10.22(0.447)
$R_{p.i.m.}$ (%)#	4.1 (17.8)
Refinement (42.3–1.65 Å)	
R_{free} (%)	22.5
R_{cryst} (%)	17.9
Size of R_{free} set (%)	5
Protein native residues (dimer)	316
8-MERCAPTOGUANINE Molecules	2
Water molecules	254
RMSD from ideal values:	
Bond lengths (Å)	0.024
Bond angles (°)	2.16
Mean B factors (Å^2)	14.7
Ramachandran plot	
Residues in most favored regions (%)	91.6
Residues in additionally allowed regions (%)	8.1
Residues in generously allowed regions (%)	0.4
Residues in disallowed regions (%)	0.0

*$R_{merge} = \Sigma h \Sigma i \ |Ii(h) - <I(h)>|/\Sigma h \Sigma i Ii \ (h)$,
#$R_{pim} = \Sigma h \ [1/(N-1)]1/2 \ \Sigma i \ |Ii(h) - <I(h)>|/\Sigma h \Sigma i Ii \ (h)$.
Values in parentheses refer to the outer resolution shell (1.74-1.65Å).
Where *I* is the observed intensity, $<I>$ is the average intensity of multiple observations from symmetry-related reflections, and N is redundancy.
$R_{value} = \_jjFoj \_ jFcjj/\_jFoj$, where Fo and Fc are the observed and calculated structure factors. For R_{free} the sum is done on the test set reflections (5% of total reflections), for R_{work} on the remaining reflections.
doi:10.1371/journal.pone.0029444.t002

Methods

Crystallisation and x-ray structure determination

SaHPPK was expressed, purified and crystallised as described previously [35]. The initial phases of the binary complex were determined by molecular replacement using Phaser [55] as reported [35]. Prior to molecular-replacement calculations, the ligands (AMP, HMDPPP) and solvent molecules were omitted. Refinement was performed using *REFMAC*5 [56] and the Fourier maps *(2F₀−F_c and F₀−F_c)* were visualized in *Coot* [57]. After several rounds of manual rebuilding, 8-mercaptoguanine and water molecules were added and the model further refined to a resolution of 1.65 Å. Structure validation was conducted using *PROCHECK* [58]. Table 2 provides the statistics for the x-ray data collection and final refined model.

Preparation of isotopically-labelled protein for NMR spectroscopy

Isotopically-labelled protein samples for NMR spectroscopy were prepared as follows: *E. coli* BL21 (DE3) cells (Agilent) transformed with the plasmid were grown overnight in 3 mL of 2×YT media supplemented with 100 µg mL^{-1} kanamycin for selection. The overnight culture was subcultured into 50 mL of minimal media that was grown to an OD$_{600}$ of 0.5–0.7. This was then added to 1 L of minimal media supplemented with 1.5 g of ^{15}N ammonium chloride and 3 g of ^{13}C glucose and grown at 310 K until the OD$_{600}$ was 0.5–0.8. Isopropyl β-D-1-thiogalacto-pyranoside (IPTG) was added to a final optimised concentration of 0.5 mM and expression was carried out at 293 K for 12 hr. Purification was as reported previously [35].

NMR spectroscopy

All NMR experiments were recorded at 295 K on a Varian Inova 600 MHz NMR spectrometer equipped with a cryoprobe and Z axis gradient. Triple resonance assignments were performed on SaHPPK, firstly in the presence of 10 mM Mg^{2+} and 1 mM AMPCPP, and secondly with the further addition of 0.6 mM 8-mercaptoguanine. $^{15}N/^{13}C$-labelled SaHPPK was typically 0.15–0.25 mM in NMR buffer (90%/10% H_2O/D_2O buffer of 50 mM HEPES pH 8.0, 1% sorbitol, 10 mM DTT). Backbone assignments were obtained on these samples using the following triple resonance experiments: HNCO, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH. Assignments were further confirmed using a 3D ^{15}N-edited NOESY experiment recorded with a mixing time of 120 ms [59]. The ^{15}N ^{1}HN assignments of the following samples were derived from a pair of 3D experiments; a ^{15}N edited NOESY experiment recorded with a mixing time of 120 ms and a HNCA experiment: *apo*SaHPPK, SaHPPK in the presence of 0.6 mM 8-mercaptoguanine, and SaHPPK in the presence of 1 mM HMDP/10 mM Mg^{2+}/1 mM AMPCPP in NMR buffer. Titrations were performed by titrating ligands into 0.1 mM^{15}N-labelled SaHPPK protein samples and recording a soFast^{15}N HMQC spectrum [60].

HMDP is prone to oxidation and degradation [22] which has complicated x-ray or NMR studies to date. No change was observed in the protein spectra over the time-course of all NMR experiments (<48 hrs, Fig. S3). Samples were routinely sealed under nitrogen for all lengthy NMR experiments to slow disulfide mediated aggregation or air/light induced degradation of HMDP.

^{15}N relaxation data were recorded on ~0.15 mM ^{15}N-labelled samples of SaHPPK. ^{15}N heteronuclear NOE spectra were recorded using TROSY-type selection and with watergate suppression [61], owing to superior sensitivity compared to the sensitivity-enhanced version [62] on the Varian cryoprobe. Three seconds of weak presaturation was used to generate the desired heteronuclear NOE, and was applied on- or off-resonance at the amide proton frequency, in addition to 1 s of relaxation delay. T1 and T2 relaxation data were acquired as described [62]. The relaxation delay was sampled at 10, 30, 60, 90, 110, 200, 500, 600 ms, and 10, 30, 50, 70, 90, 110 ms for longitudinal and transverse relaxation measurements, respectively. Spectra were processed using nmrPipe [63] and analysed with XEASY [64] or SPARKY [65].Titration and relaxation 2D experiments were acquired with t_{1max} (^{15}N) = 51–62 ms and t_{2max}(^{1}H) = 142 ms.

$^{1}D_{NH}$ RDCs were measured on a ~0.1 mM ^{15}N-labelled SaHPPK sample in the presence of an anisotropic media comprising 5% (wt/vol) C12E6/hexanol [39]. RDCs were obtained by comparing coupled spectra in the presence of the orienting media against spectra in the isotropic state by recording a 2D ^{15}N IPAP-HSQC spectrum [66]. RDCs were measured using SPARKY as the isotropic (J) – aligned (J+$^{1}D_{NH}$) values recorded in the ^{15}N (t1) dimension. 114 $^{1}D_{NH}$ RDCs were fitted to the x-ray structure using single value decomposition incorporated into the "bestFit" module within PALES [40]. The $^{1}D_{NH}$ RDCs were removed for couplings derived from severely overlapping peaks in the 2D IPAP spectra and mobile residues as inferred from ^{15}N relaxation data. The error in the RDC was conservatively estimated as +/− 1 Hz, according to the ratio of the linewidth to the signal-to-noise. Final values of Da and Rhombicity were 9.6 Hz and 0.53 respectively.

Virtual screening using ROCS (Rapid Overlay of Chemical Structures)

An *in-silico* database comprised of 229,172 commercially purchasable compounds (Maybridge, Chembridge and Specs libraries) was used for the screening. Database compounds were first converted from 2D sdf format into 3D coordinates using LigPrep within Maestro [67]. A ROCS [68] run was then initiated to screen the database for potential pterin site binders using HMDP as the query molecule. The top 500 hits were ranked according to combined volume overlap (Tversky) and chemical functionality (Colour) scores. The hits were overlayed onto the crystal structure of EcHPPK bound to HMDP (1Q0N). Filtering of incompatible molecules was first assessed by manual inspection of the hits within the pterin binding pocket and compounds that showed any possible steric clashes were discarded. ROCS hits were chosen that retained key pharmacophore interactions in the pterin binding site and those that provided synthetic opportunities for scaffold optimization. A total of 44 compounds were selected and of these, 37 were available for purchase.

Differential Scanning Fluorimetry (DSF) assay for ligand binding

A 96-well plate containing solutions of 2–5 μM SaHPPK, 10× Sypro orange dye and 500 μM of test compound in buffer (50 mM HEPES (pH 8.0), 1% sorbitol, 2.0 mM DTT, 10 mM $MgCl_2$) was heated from 298 to 322 K at a rate of 1.0 K min^{-1}. The fluorescence intensity was measured with excitation/emission wavelengths of 583/610 nm. T_m and ΔT_m values were determined from the melting curves obtained.

Isothermal calorimetry (ITC)

Experiments were performed using an iTC200 instrument (MicroCal) at 298 K, with ligands titrated into solutions of SaHPPK using 18×2.2 μL or 13×3.1 μL injections. Data were fitted using Origin software to yield the thermodynamic parameters, ΔH, K_d and N (the binding stoichiometry), assuming a cell volume of 0.2 mL. These were then used to calculate the Gibb's free energy of binding, ΔG ($-RT.\ln K_a$), and entropy of binding, ΔS (using $\Delta G = \Delta H - T\Delta S$). For titrations with AMPCPP, SaHPPK and AMPCPP concentrations were typically 30 and 400 μM, respectively. For titrations with ATP, SaHPPK was typically at 70 μM and ATP at 1500 μM. For titrations with HMDP substrate, SaHPPK was typically at 22 μM and HMDP at 500 μM, with AMPCPP added to both solutions at 1 mM where indicated. For titrations with 8-mercaptoguanine, SaHPPK was typically at 11 μM and inhibitor at 300 μM. Inhibitor was prepared as a 200 mM stock solution in DMSO, and diluted into ITC buffer to a nominal concentration of 500 μM, with an equal 0.25% DMSO added to the SaHPPK sample to ensure buffer matching. Experiments were limited by the solubility of 8-mercaptoguanine, estimated to be ~300 μM from ITC experiments assuming a 1:1 binding stoichiometry.

Surface plasmon resonance (SPR)

Minimal biotinylation of SaHPPK was performed as follows: SaHPPK (~30 nanomolar) in SPR "capture buffer" (50 mM HEPES, pH8.0, 150 mMNaCl, 0.05% (v/v) polysorbate 20, 5 mM DTT, 10 mM $MgCl_2$) was incubated with an equimolar concentration of EZ-Link® Sulfo-NHS-LC-LC-Biotin (Pierce) on ice for 2 hours [69]. The biotinylated enzyme was passed through a Superdex 75 (10/300 GL) column equilibrated with SPR capture buffer to remove free biotin. NeutrAvidin (Pierce) was immobilized at 298 K onto a CM5 sensor chip docked in a Biacore T100 instrument (GE Healthcare) as described previously [70], resulting in immobilization levels of approximately 13,000 RU in all flow cells. The biotinylated SaHPPK was bound to the NeutrAvidin chip surface by injecting at 10 μL min^{-1} for 6 min over a single flow cell,

typically resulting in immobilization of ~10,000 RU of SaHPPK. A minimally-biotinylated bovine carbonic anhydrase II protein was captured in a separate flow cell to provide an unrelated negative control. The blank (typically flow cell 1) and captured surfaces were subsequently blocked with three 1 min injections of 1 μg mL^{-1} D-biotin (Sigma) at 10 μL min^{-1}.

All SPR binding experiments were performed at 293 K in SPR capture buffer with 5% (v/v) DMSO added. Small molecules (ligands) were serially diluted (either 2- or 3-fold) in SPR binding buffer and injected for 30 sec contact time at 60 μL min^{-1}, then allowed to dissociate for 60 sec. SPR running buffer was injected after each ligand injection to ensure complete regeneration of immobilized SaHPPK. Each ligand titration was performed in triplicate. Binding sensorgrams were processed, solvent-corrected and double-referenced using Scrubber software (BioLogic Software, Australia). Responses at equilibrium for each analyte were fitted to a 1:1 steady-state affinity model available within Scrubber to determine the K_d. Where the dissociation rate was sufficiently slow, binding data were fit globally to a 1:1 kinetic interaction model that included a mass transport component [71] and the K_d determined from the (k_d/k_a) ratio.

KinaseGlo^{TM} Biochemical assay

The KinaseGlo^{TM} assay kit (Promega) was used to quantify HPPK activity. In this assay, firefly luciferase utilizes the ATP remaining after HPPK catalysis to produce a luminescence signal that is directly proportional to ATP concentration; from this, the HPPK activity can be derived. The enzyme activity and optimum concentration to define kinetic parameters was determined by measuring the initial rate of ATP consumption over a range of HPPK concentrations in the presence and absence of HMDP substrate (data not shown). For kinetic measurements, an optimized HPPK concentration of 7 ng/50 μL assay volume was determined, which allowed for monitoring the first 10% of reactions turnover in a reasonable assay time period (20 min).

Measurements were performed in 96-well plates using assay buffer (100 mMTris-HCl/10 mM MgCl$_2$, pH 8.5, 0.01% (w/v) BSA, 0.01% (v/v) Tween 20 and 10 mM β-mercaptoethanol). Typically, 5 μl of test compound (dissolved in 50% DMSO) and 20 μl of enzyme were added to each well followed by 25 μl of assay buffer giving 0.3 μM pterin and 0.2 μM ATP in a total reaction volume 50 μl. After a 20 minute incubation at room temperature, the enzymatic reaction was stopped with 50 μl of KinaseGlo^{TM} reagent. Luminescence was recorded after a further 10 min using the *FLUOstar Optima* plate reader (BMG, Labtech Ltd). Kinetic data and inhibition data were fit to Michaelis-Mentenand sigmoidal dose-response equations respectively, using *GraphPad Prism*.

Accession Numbers

The coordinates and structure factors of SaHPPK in complex with 8-mercaptoguanine have been deposited in the RCSB Protein Data Bank with accession number 3QBC.

Supporting Information

Figure S1 A) ITC (left) and SPR data (right) for binding of ATP and AMPCPP to SaHPPK. Equilibrium binding constants (K$_d$) are shown. B) SPR data for the binding of ADP and AMP to SaHPPK.
(DOCX)

Figure S2 Binding of substrate, cofactor and inhibitor to SaHPPK as measured by NMR spectroscopy. A, Superposition of the ^{15}N HSQC spectra of ~100 μM SaHPPK+10 mM MgSO$_4$ with and without 1 mM AMPCPP (left), and the ^{15}N HSQC spectra of ~100 μM SaHPPK+1 mM AMPCPP with and without 1 mM HMDP (right). B, Missing amides are shown on a ribbon representation for apo HPPK (green) and those that additionally disappear in the presence of various saturating ligands (blue). C, Missing amides and CSPs mapped onto surface of HPPK. Residues with missing resonances in the apo enzyme are coloured cyan, whilst those additionally broadened are coloured magenta upon binding of ligands. Residues displaying slow exchange CSPs upon ligand binding in the AMPCPP binary and AMPCPP+8-mercaptoguanine complexes are shaded red, with the hue corresponding to the magnitude of the CSPs. The position of the AMPCPP is modeled from that in the *E. coli* HPPK (1Q0N). D, Missing amides with addition of saturating 8-mercaptoguanine are coloured magenta. CSPs for the slow exchange 8-mercaptoguanine+AMPCPP ternary complex are coloured red. Missing amides are further shown on the ribbon representation in green.
(DOCX)

Figure S3 ^{15}**N sofast HMQC spectra recorded (20 min per spectrum) over the time period shown for ~100 μM SaHPPK in complex with either 200 μM HMDP/1 mM AMPCPP (left) and 200 μM oxidized HMDP/1 mM AMPCPP (right).** Several spectral changes are observed in the oxidized HMDP/AMPCPP (right) over time but not in HMDP/AMPCPP (left).
(DOCX)

Figure S4 ^{1}**H NMR spectra of 5 mM HMDP/5 mM DTT (top) and 20 mM oxidized HMDP (bottom).** Both spectra were recorded in 50 mM Potassium phosphate buffer D$_2$O pH 7.9. The very small amount of oxidation of HMDP is just visible at ~8.6 ppm in the top spectrum. The spectra show the initial purity of the HMDP used in the NMR experiments in S3.
(DOCX)

Figure S5 CSP data for various ligands binding to SaHPPK. From top to bottom: HMDP, AMPCPP, Pterin+AMPCPP, 8-mercaptoguanine, 8-mercaptoguanine+AMPCPP. Grey regions indicate residues for which resonances were extensively/fully broadened compared to the apo ^{15}N HSQC spectrum, and therefore not observed. The CSPs were calculated using the following equation: $\Delta\delta = \sqrt{(\delta N^{15} \times 0.154)^2 + (\delta HN^1)^2}$, δ refers to the chemical shift change in ppm of the resonance relative to the apo SaHPPK ^{15}N HSQC spectrum.
(DOCX)

Figure S6 Comparison of the HYDRONMR calculated ^{15}N T1/T2 ratio (600 MHz) for a monomer SaHPPK (black) and the x-ray SaHPPK dimer (red) with the measured ^{15}N T1/T2 (blue) for SaHPPK in complex with 8-mercaptoguanine and AMPCPP.
(DOCX)

Table S1 ^{15}**N and ^{1}HN chemical shifts (ppm) for; SaHPPK, SaHPPK/AMPCPP, SaHPPK/HMDP, SaHPPK/8-mercaptoguanine/AMPCPP, SaHPPK/8-mercaptoguanine.** Chemical shifts in red are from the ^{15}Nε1 and ^{1}Hε1 resonance of Trp89.
(DOCX)

Author Contributions

Conceived and designed the experiments: SC OD TSP BMC JDS. Performed the experiments: SC OD TSP BMC IGM RF JN JDS. Analyzed the data: SC OD TSP BMC JSS JDS. Contributed reagents/materials/analysis tools: OD TSP BMC JDS. Wrote the paper: SC OD TSP BMC JSS JDS.
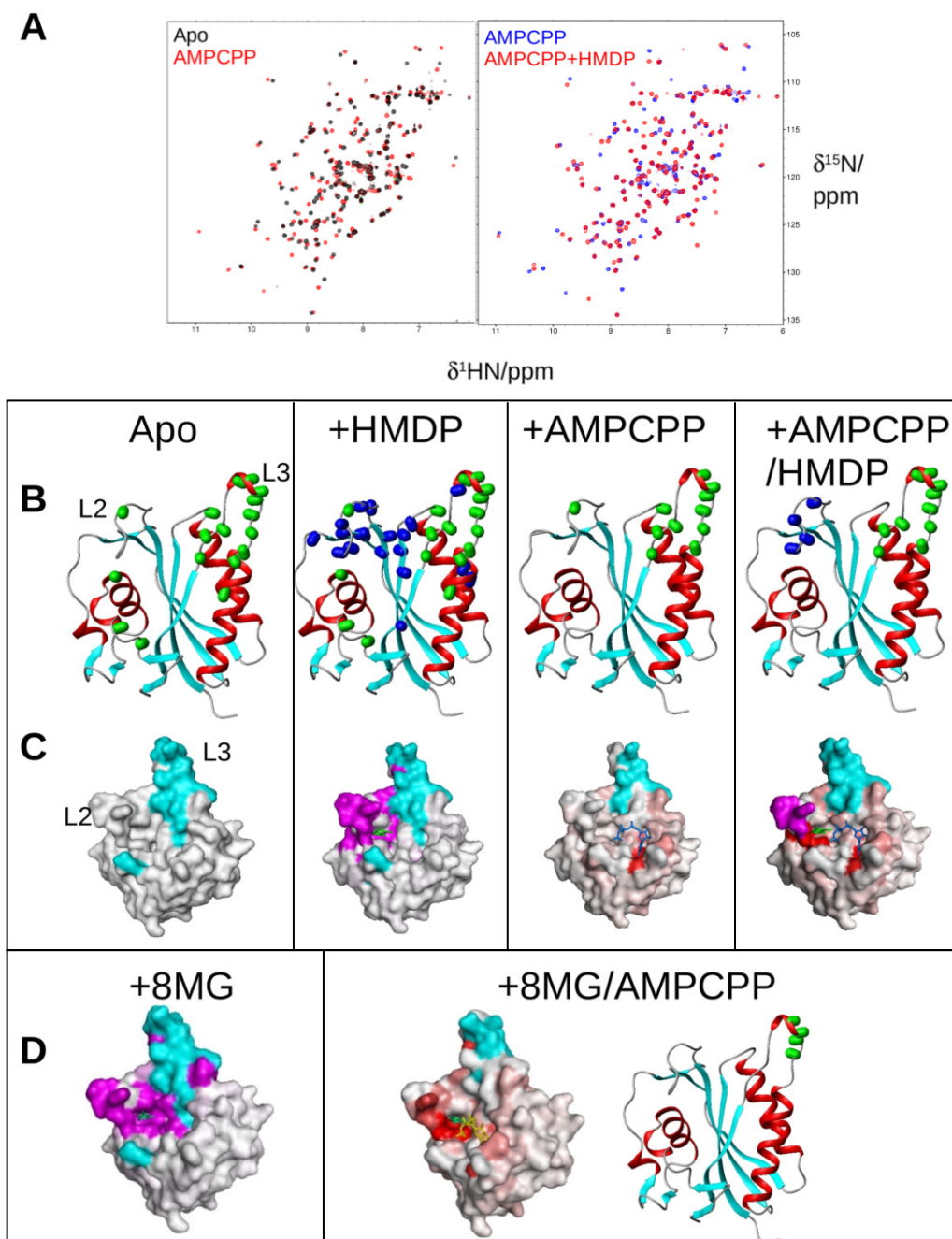
References

1. Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science 325: 1089–1093.
2. Hidron AI, Low CE, Honig EG, Blumberg HM (2009) Emergence of community-acquired meticillin-resistant Staphylococcus aureus strain USA300 as a cause of necrotising community-onset pneumonia. Lancet Infect Dis 9: 384–392.
3. David MZ, Daum RS (2010) Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. Clin Microbiol Rev 23: 616–687.
4. Naimi TS, LeDell KH, Como-Sabetti K, Borchardt SM, Boxrud DJ, et al. (2003) Comparison of community- and health care-associated methicillin-resistant Staphylococcus aureus infection. JAMA 290: 2976–2984.
5. Adra M, Lawrence KR (2004) Trimethoprim/sulfamethoxazole for treatment of severe Staphylococcus aureus infections. Ann Pharmacother 38: 338–341.
6. Proctor RA (2008) Role of folate antagonists in the treatment of methicillin-resistant Staphylococcus aureus infection. Clin Infect Dis 46: 584–593.
7. Besier S, Zander J, Siegel E, Saum SH, Hunfeld KP, et al. (2008) Thymidine-dependent Staphylococcus aureus small-colony variants: human pathogens that are relevant not only in cases of cystic fibrosis lung disease. J Clin Microbiol 46: 3829–3832.
8. Zander J, Besier S, Ackermann H, Wichelhaus TA (2010) Synergistic antimicrobial activities of folic acid antagonists and nucleoside analogs. Antimicrob Agents Chemother 54: 1226–1231.
9. Frey KM, Liu J, Lombardo MN, Bolstad DB, Wright DL, et al. (2009) Crystal structures of wild-type and mutant methicillin-resistant Staphylococcus aureus dihydrofolate reductase reveal an alternate conformation of NADPH that may be linked to trimethoprim resistance. J Mol Biol 387: 1298–1308.
10. Black MT, Hodgson J (2005) Novel target sites in bacteria for overcoming antibiotic resistance. Adv Drug Deliv Rev 57: 1528–1538.
11. Bermingham A, Derrick JP (2002) The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. Bioessays 24: 637–648.
12. Swarbrick JD, Iliades P, Simpson JS, Macreadie I (2008) Folate biosynthesis - reappraisal of old and novel targets in the search for new antimicrobials. The Open Enzyme Inhibition Journal 1: 12–33.
13. Xiao B, Shi G, Chen X, Yan H, Ji X (1999) Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. Structure 7: 489–496.
14. Hennig M, Dale GE, D'Arcy A, Danel F, Fischer S, et al. (1999) The structure and function of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from Haemophilus influenzae. J Mol Biol 287: 211–219.
15. Lawrence MC, Iliades P, Fernley RT, Berglez J, Pilling PA, et al. (2005) The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/dihydropteroate synthase of Saccharomyces cerevisiae. J Mol Biol 348: 655–670.
16. Garcon A, Levy C, Derrick JP (2006) Crystal structure of the bifunctional dihydroneopterin aldolase/6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from Streptococcus pneumoniae. J Mol Biol 360: 644–653.
17. Blaszczyk J, Li Y, Cherry S, Alexandratos J, Wu Y, et al. (2007) Structure and activity of Yersinia pestis 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as a novel target for the development of antiplague therapeutics. Acta Crystallogr D Biol Crystallogr 63: 1169–1177.
18. Pemble CW, Mehta PK, Mehra S, Li Z, Nourse A, et al. (2010) Crystal structure of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase•dihydropteroate synthase bifunctional enzyme from Francisella tularensis. PLoS One 5: e14165.
19. Blaszczyk J, Shi G, Li Y, Yan H, Ji X (2004) Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. Structure 12: 467–475.
20. Blaszczyk J, Li Y, Wu Y, Shi G, Ji X, et al. (2004) Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. Biochemistry 43: 1469–1477.
21. Blaszczyk J, Shi G, Yan H, Ji X (2000) Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 A resolution. Structure 8: 1049–1058.
22. Blaszczyk J, Li Y, Shi G, Yan H, Ji X (2003) Dynamic roles of arginine residues 82 and 92 of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: crystallographic studies. Biochemistry 42: 1573–1580.
23. Bermingham A, Bottomley JR, Primrose WU, Derrick JP (2000) Equilibrium and kinetic studies of substrate binding to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from Escherichia coli. J Biol Chem 275: 17962–17967.
24. Li Y, Gong Y, Shi G, Blaszczyk J, Ji X, et al. (2002) Chemical transformation is not rate-limiting in the reaction catalyzed by Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. Biochemistry 41: 8777–8783.
25. Shi G, Gong Y, Savchenko A, Zeikus JG, Xiao B, et al. (2000) Dissecting the nucleotide binding properties of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase with fluorescent 3′(2′)-o-anthraniloyladenosine 5′-triphosphate. Biochim Biophys Acta 1478: 289–299.
26. Stammers DK, Achari A, Somers DO, Bryant PK, Rosemond J, et al. (1999) 2.0 A X-ray structure of the ternary complex of 7,8-dihydro-6-hydroxymethylpter-inpyrophosphokinase from Escherichia coli with ATP and a substrate analogue. FEBS Lett 456: 49–53.
27. Shi G, Blaszczyk J, Ji X, Yan H (2001) Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: synthesis and biochemical and crystallographic studies. J Med Chem 44: 1364–1371.
28. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, et al. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. Journal of Biomolecular Screening 6: 429–440.
29. Bagby S, Tong KI, Liu DJ, Alattia JR, Ikura M (1997) The button test: A small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. Journal of Biomolecular Nmr 10: 279–282.
30. Rush TS, Grant JA, Mosyak L, Nicholls A (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. Journal of Medicinal Chemistry 48: 1489–1495.
31. Lo MC, Aulabaugh A, Jin GX, Cowling R, Bard J, et al. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Analytical Biochemistry 332: 153–159.
32. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2: 2212–2221.
33. Talarico TL, Dev IK, Dallas WS, Ferone R, Ray PH (1991) Purification and partial characterization of 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase and 7,8-dihydropteroate synthase from Escherichia coli MC4100. Journal of bacteriology 173: 7029–7032.
34. Huber W (2005) A new strategy for improved secondary screening and lead optimization using high-resolution SPR characterization of compound-target interactions. J Mol Recognit 18: 273–281.
35. Chhabra S, Newman J, Peat TS, Fernley RT, Caine J, et al. (2010) Crystallization and preliminary X-ray analysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from Staphylococcus aureus. Acta Crystallogr Sect F Struct Biol Cryst Commun 66: 575–578.
36. 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.
37. Lescop E, Lu ZW, Liu Q, Xu HM, Li GY, et al. (2009) Dynamics of the Conformational Transitions in the Assembling of the Michaelis Complex of a Bisubstrate Enzyme: A N-15 Relaxation Study of Escherichia coli 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase. Biochemistry 48: 302–312.
38. Garcia de la Torre J, Huertas ML, Carrasco B (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J Magn Reson 147: 138–146.
39. Ruckert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. Journal of the American Chemical Society 122: 7793–7797.
40. Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. Journal of the American Chemical Society 122: 3791–3792.
41. Derrick JP (2008) The structure and mechanism of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. Vitam Horm 79: 411–433.
42. Beavo JA, Rogers NL, Crofford OB, Hardman JG, Sutherland EW, et al. (1970) Effects of xanthine derivatives on lipolysis and on adenosine 3′,5′-monophosphate phosphodiesterase activity. Molecular pharmacology 6: 597–603.
43. Kalra S, Jena G, Tikoo K, Mukhopadhyay AK (2007) Preferential inhibition of xanthine oxidase by 2-amino-6-hydroxy-8-mercaptopurine and 2-amino-6-purine thiol. BMC biochemistry 8: 8.
44. Michael MA, Cottam HB, Smee DF, Robins RK, Kini GD (1993) Alkylpurines as immunopotentiating agents. Synthesis and antiviral activity of certain alkylguanines. Journal of medicinal chemistry 36: 3431–3436.
45. Miller RL, Ramsey GA, Krenitsky TA, Elion GB (1972) Guanine phosphoribosyltransferase from Escherichia coli, specificity and properties. Biochemistry 11: 4723–4731.
46. Stoeckler JD, Cambor C, Kuhns V, Chu SH, Parks RE, Jr. (1982) Inhibitors of purine nucleoside phosphorylase, C(8) and C(5′) substitutions. Biochemical pharmacology 31: 163–171.
47. Goodman MG, Weigle WO (1983) Activation of lymphocytes by a thiol-derivatized nucleoside: characterization of cellular parameters and responsive subpopulations. Journal of immunology 130: 551–557.
48. Hevener KE, Yun MK, Qi JJ, Kerr ID, Babaoglu K, et al. (2010) Structural Studies of Pterin-Based Inhibitors of Dihydropteroate Synthase. Journal of Medicinal Chemistry 53: 166–177.

49. Gibson CL, Huggan JK, Kennedy A, Kiefer L, Lee JH, et al. (2009) Diversity oriented syntheses of fused pyrimidines designed as potential antifolates. Organic & biomolecular chemistry 7: 1829–1842.

50. Swarbrick JD, Buyya S, Gunawardana D, Gayler KR, McLennan AG, et al. (2005) Structure and substrate-binding mechanism of human Ap4A hydrolase. J Biol Chem 280: 8471–8481.

51. Li G, Felczak K, Shi G, Yan H (2006) Mechanism of the conformational transitions in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy. Biochemistry 45: 12573–12581.

52. Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12: 1–16.

53. Lipsitz RS, Tjandra N (2004) Residual dipolar couplings in NMR structure analysis. Annu Rev Biophys Biomol Struct 33: 387–413.

54. Vajpai N, Strauss A, Fendrich G, Cowan-Jacob SW, Manley PW, et al. (2008) Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. The Journal of biological chemistry 283: 18292–18302.

55. Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr 60: 432–438.

56. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography 53: 240–255.

57. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60: 2126–2132.

58. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26: 283–291.

59. Talluri S, Wagner G (1996) An optimized 3D NOESY-HSQC. J Magn Reson B 112: 200–205.

60. Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33: 199–211.

61. Zhu G, Xia YL, Nicholson LK, Sze KH (2000) Protein dynamics measurements by TROSY-based NMR experiments. Journal of Magnetic Resonance 143: 423–426.

62. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, et al. (1994) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 Nmr Relaxation. Biochemistry 33: 5984–6003.

63. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6: 277–293.

64. Bartels C, Xia TH, Billeter M, Guntert P, Wuthrich K (1995) The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. Journal of Biomolecular Nmr 6: 1–10.

65. Goddard TD, Kneller D (2001) SPARKY 3 ed. San Francisco: University of California.

66. Ottiger M, Delaglio F, Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson 131: 373–378.

67. LigPrep (2011) Version 2.3. New York: Schrödinger, LLC.

68. ROCS (2005) Version 2.1.1. Santa FeNM: Openeye Scientific Software.

69. Papalia G, Myszka D (2010) Exploring minimal biotinylation conditions for biosensor analysis using capture chips. Analytical Biochemistry 403: 30–35.

70. Katsamba P, Carroll K, Ahlsen G, Bahna F, Vendome J, et al. (2009) Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. Proc Natl Acad Sci U S A 106: 11594–11599.

71. Myszka DG, Morton TA, Doyle ML, Chaiken IM (1997) Kinetic analysis of a protein antigen-antibody interaction limited by mass transport on an optical biosensor. Biophys Chem 64: 127–137.
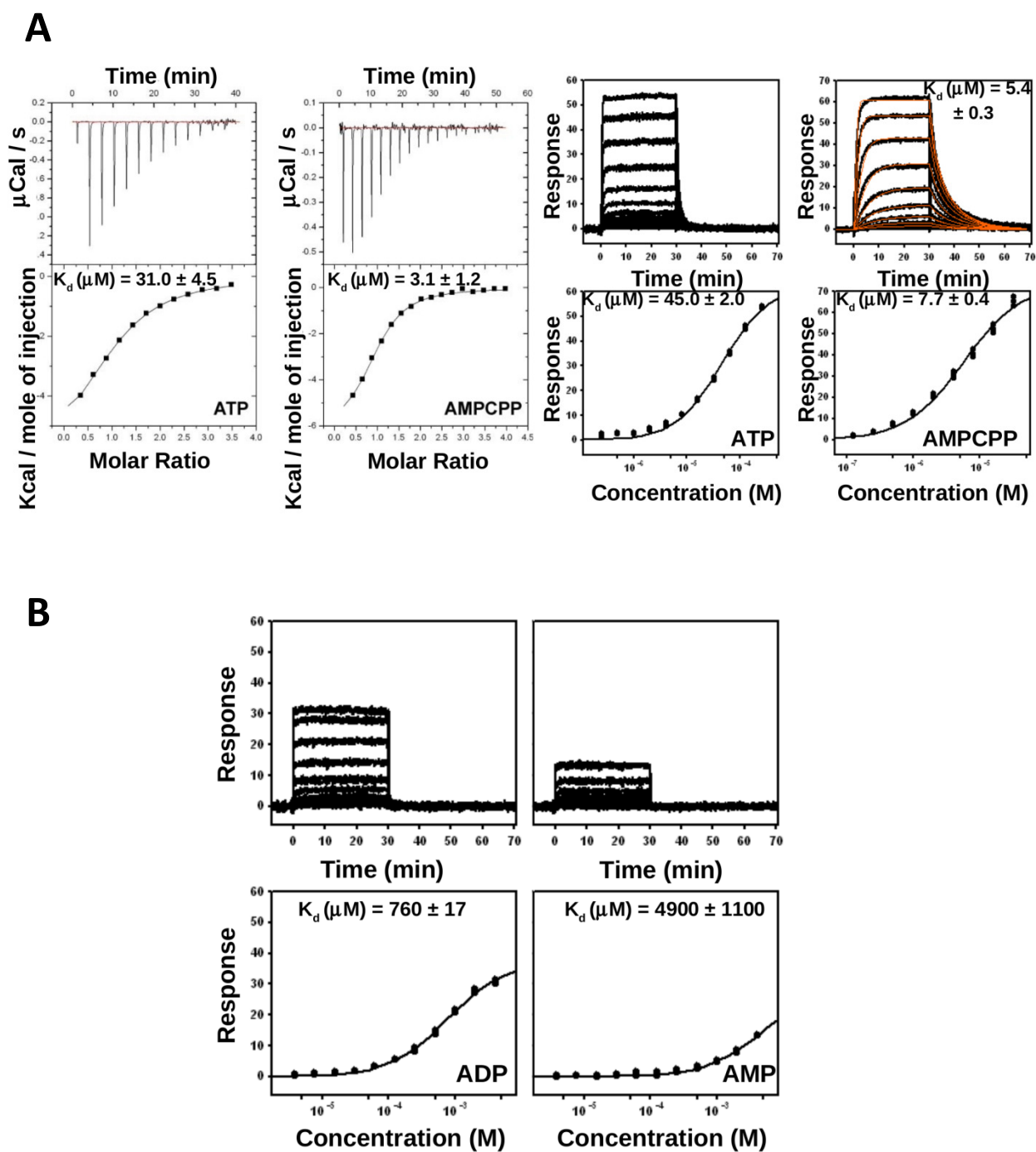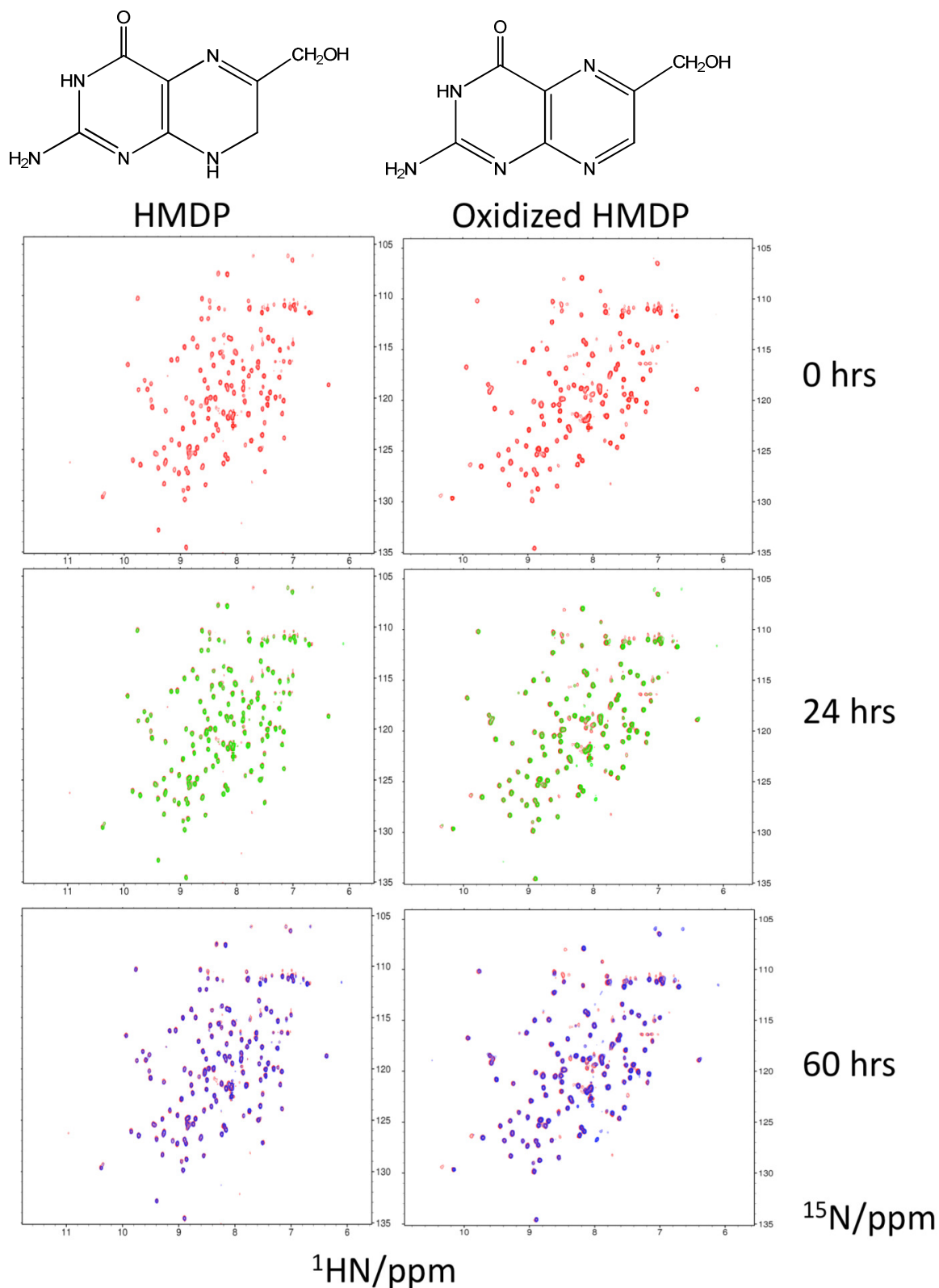
Figure S1.A) ITC and SPR data for binding of ATP and AMPCPP to SaHPPK. Equilibrium binding constants (K_d) are shown. B) SPR data for the binding of ADP and AMP to SaHPPK.

Figure S2. **Binding of substrate, cofactor and inhibitor to SaHPPK as measured by NMR spectroscopy.** A, Superposition of the ^{15}N HSQC spectra of ~100 µM SaHPPK+ 10 mM MgSO$_4$ with and without 1 mM AMPCPP (*left*), and the ^{15}N HSQC spectra of ~100 µM SaHPPK + 1mM AMPCPP with and without 1mM HMDP (*right*). B, Missing amides are shown on a ribbon representation for apo HPPK (green) and those that additionally disappear in the presence of various saturating ligands (blue). C, Missing amides and CSPs mapped onto surface of HPPK. Residues with missing resonances in the *apo* enzyme are coloured cyan, whilst those additionally broadened are coloured magenta upon binding of ligands. Residues displaying slow exchange CSPs upon ligand binding in the AMPCPP binary and AMPCPP+8-mercaptoguanine complexes are shaded red, with the hue corresponding to the magnitude of the CSPs. The position of the AMPCPP is modeled from that in the *E. coli* HPPK (1Q0N). D, Missing amides with addition of saturating 8-mercaptoguanine are coloured magenta. CSPs for the slow exchange 8-mercaptoguanine + AMPCPP ternary
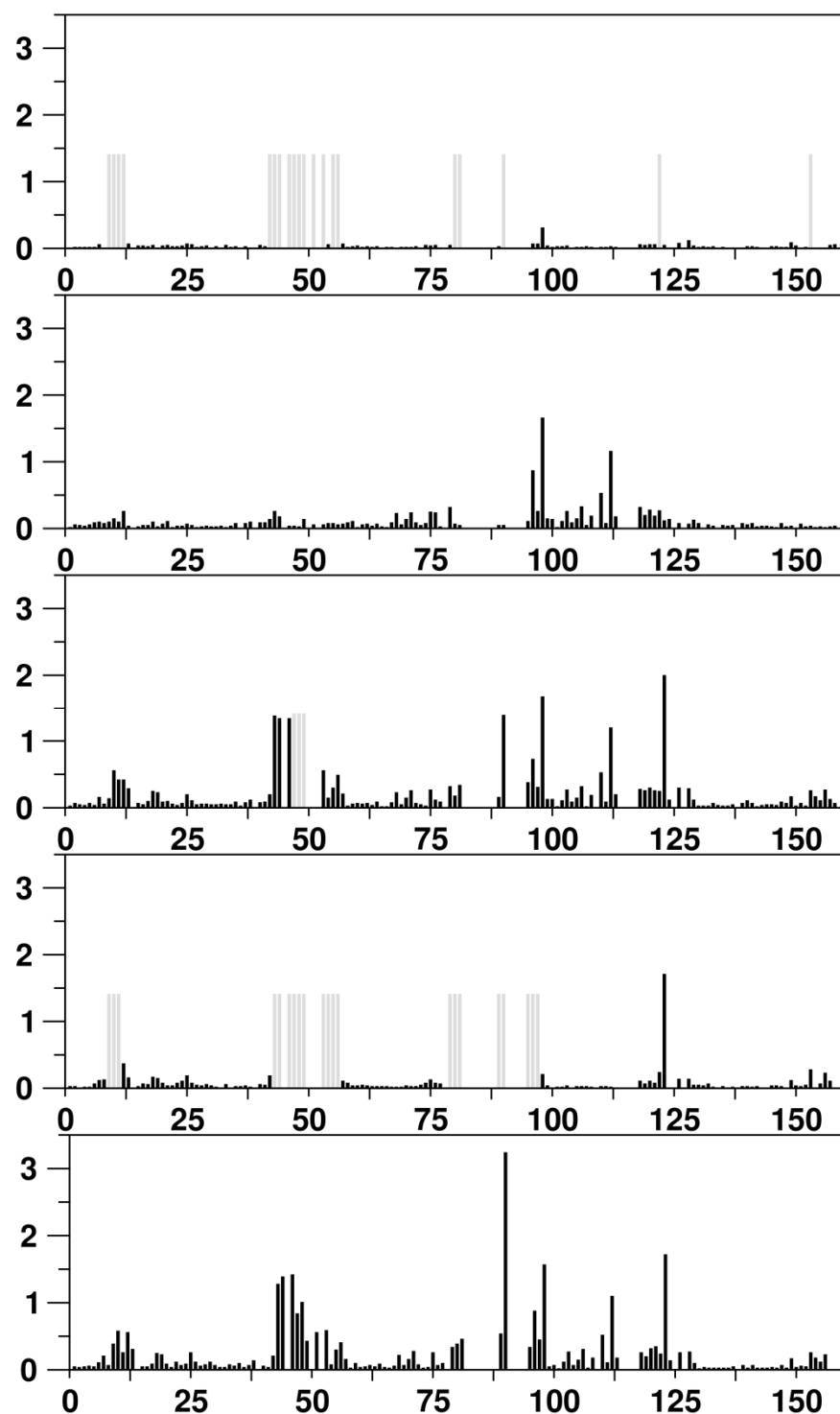
2

complex are coloured red. Missing amides are further shown on the ribbon representation in green.
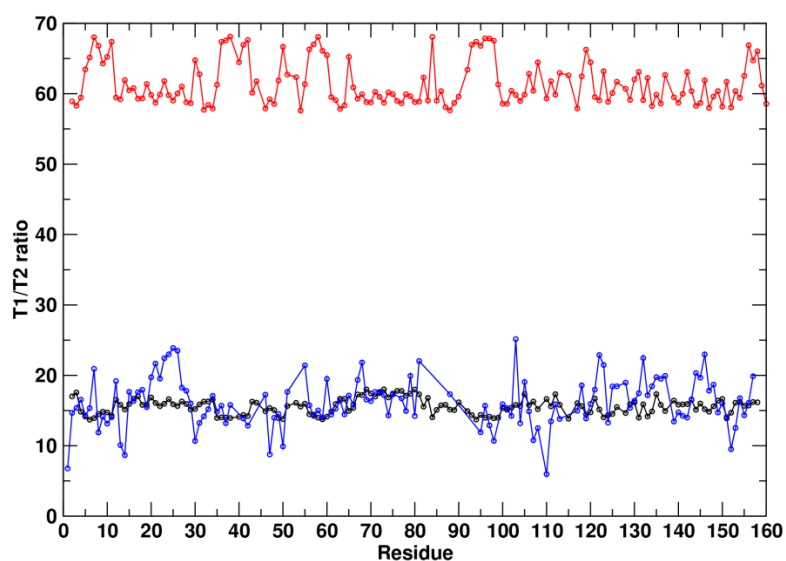
Figure S3. ^{15}N sofast HMQC spectra recorded (20 min per spectrum) over the time period shown for ~100 μM SaHPPK in complex with either 200 μM HMDP/ 1mM AMPCPP (left) and 200 μM oxidized HMDP / 1mM AMPCPP (right). Several spectral changes are observed in the oxidized HMDP/AMPCPP (right) over time but not in HMDP/AMPCPP (left).
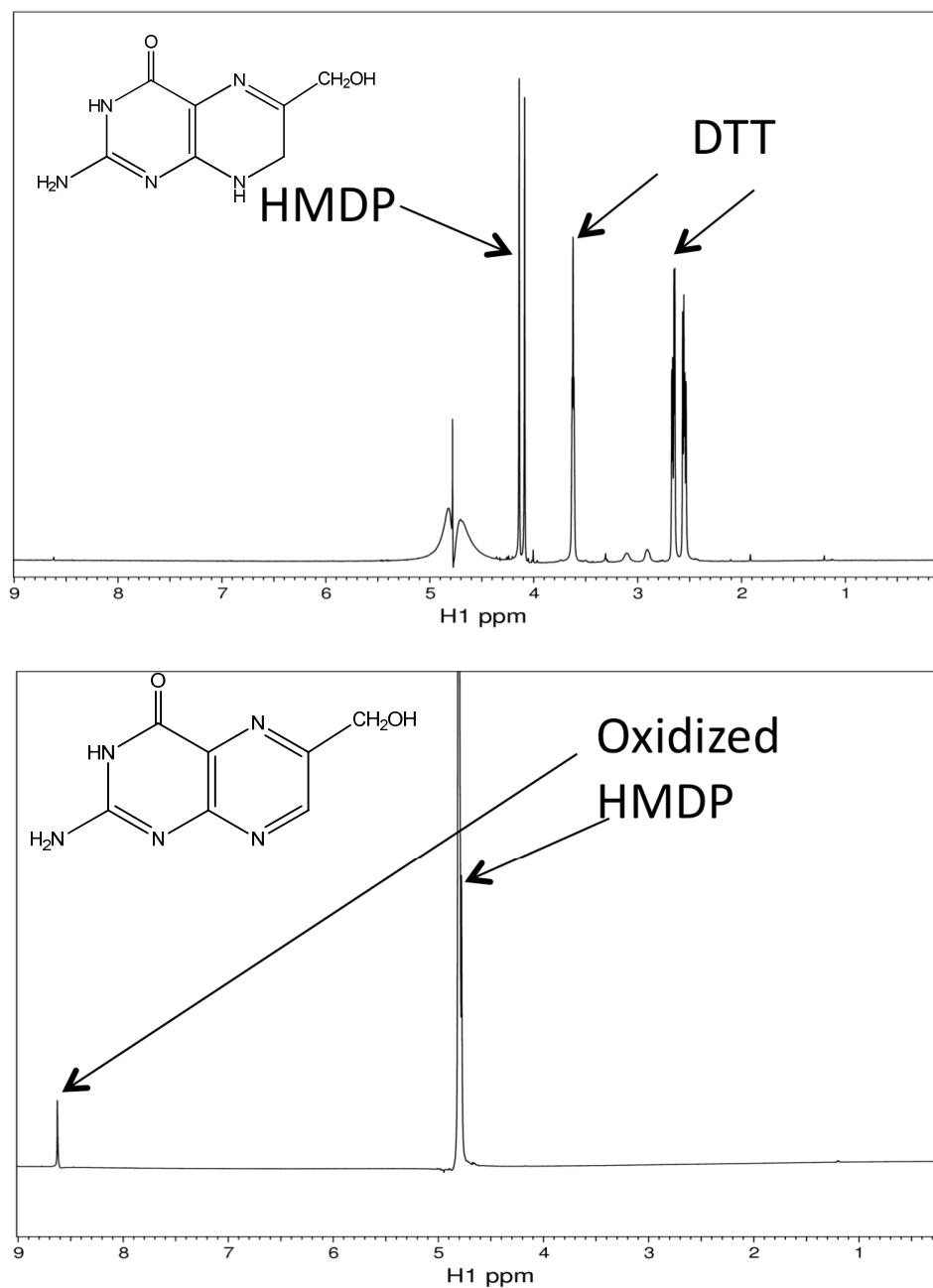
Figure S4. ^{1}H NMR spectra of 5 mM HMDP/5 mM DTT (top) and 20 mM oxidized HMDP (bottom). Both spectra were recorded in 50 mM Potassium phosphate buffer D_2O pH 7.9. The very small amount of oxidation of HMDP is just visible at ~8.6 ppm in the top spectrum. The spectra show the initial purity of the HMDP used in the NMR experiments in S3.

Figure S5. CSP data for various ligands binding to SaHPPK. From top to bottom: HMDP, AMPCPP, Pterin+AMPCPP, 8-mercaptoguanine, 8-mercaptoguanine+AMPCPP. Grey regions indicate residues for which resonances were extensively/fully broadened compared to the *apo*^{15}N HSQC spectrum, and therefore not observed. The CSPs were calculated using the following equation: $\Delta\delta = \sqrt{(\delta N^{15} \times 0.154)^2 + (\delta HN^1)^2}$ in ppm, δrefers to the chemical shift change of the resonance relative to the apoSaHPPK ^{15}N HSQC spectrum.

6

Figure S6.Comparison of the HYDRONMR calculated ^{15}N *T1/T2* ratio (600 MHz) for a monomer SaHPPK (black) and the x-raySaHPPKdimer (red)withthe measured ^{15}N *T1/T2* (blue)for SaHPPK in complex with 8-mercaptoguanine and AMPCPP.

Table S1^{15}N and ^{1}HN chemical shifts (ppm) for; SaHPPK, SaHPPK/AMPCPP, SaHPPK/HMDP, SaHPPK/8-mercaptoguanine/AMPCPP, SaHPPK/8-mercaptoguanine. Chemical shifts in red are from the ^{15}Nε1 and ^{1}Hε1 resonance of Trp89.

Res. No	Res. type	apo	apo	ampcpp	ampcpp	HMDP/ampcpp	HMDP/ampcpp	8MG/ampcpp	8MG/ampcpp	8MG	8MG
		^{15}N	^{1}HN	^{15}N	^{1}HN	^{15}N	^{1}HN	^{15}N	^{1}HN	^{15}N	^{1}HN
1	**MET**	121.66	7.97	121.74	8.00	121.75	8.00	121.73	8.00	121.52	7.97
2	**ILE**	126.77	9.00	127.19	9.00	127.25	9.01	127.13	8.99	126.71	9.01
3	**GLN**	126.18	8.25	126.28	8.22	126.34	8.23	126.23	8.21	126.26	8.26
4	**ALA**	129.84	8.89	129.77	8.90	129.85	8.91	129.76	8.90	129.95	8.87
5	**TYR**	115.28	8.65	115.00	8.68	114.96	8.69	114.98	8.66	115.33	8.66
6	**LEU**	121.38	9.26	121.24	9.20	121.24	9.24	121.03	9.16	121.58	9.31
7	**GLY**	109.85	9.62	109.68	9.70	110.29	9.76	110.55	9.78	110.58	9.59
8	**LEU**	125.26	8.89	125.63	8.84	125.44	8.84	125.16	8.77	124.62	8.76
9	**GLY**	107.11	8.39	106.77	8.32	86.86	8.33	107.33	8.01		
10	**SER**	114.01	8.57	115.12	8.55	111.58	8.20	112.28	8.06		
11	**ASN**	125.77	8.46	126.25	8.48	125.56	8.83	124.19	8.53		
12	**ILE**	118.63	7.30	117.66	7.11	116.62	7.06	116.18	6.91	116.87	7.07
13	**GLY**	111.87	8.20	111.85	8.23	111.18	8.47	111.10	8.47	111.07	8.24
14	**ASP**			122.63	8.18	121.58	8.09	121.10	8.14		
15	**ARG**	124.60	7.83	124.62	7.83	124.38	7.82	124.98	7.84	124.73	7.81
16	**GLU**	117.37	8.66	117.26	8.64	117.27	8.64	117.27	8.63	117.16	8.62
17	**SER**	114.41	7.51	114.33	7.46	114.10	7.42	113.97	7.41	114.23	7.44
18	**GLN**	119.39	7.93	119.08	7.98	119.09	8.15	119.05	8.15	119.33	8.07
19	**LEU**	117.02	7.59	116.89	7.60	117.79	7.77	117.70	7.77	117.47	7.70
20	**ASN**	116.02	8.02	116.29	8.06	116.43	8.06	116.45	8.02	116.35	8.06
21	**ASP**	122.63	8.87	122.91	8.97	122.95	8.95	122.81	8.90	122.78	8.90
22	**ALA**	122.11	8.11	122.09	8.10	121.95	8.14	121.74	8.18	122.03	8.13
23	**ILE**	117.40	7.72	117.54	7.69	117.33	7.72	117.16	7.73	117.31	7.77
24	**LYS**	119.09	7.40	119.24	7.38	119.41	7.39	119.48	7.38	119.43	7.46
25	**ILE**	119.78	7.88	119.28	7.89	119.51	8.04	119.49	8.07	119.54	8.03
26	**LEU**	120.29	8.42	120.44	8.46	120.27	8.52	120.16	8.51	120.26	8.48
27	**ASN**	111.43	7.83	111.43	7.84	111.34	7.79	111.20	7.76	111.27	7.78
28	**GLU**	116.34	7.58	116.39	7.58	116.45	7.60	116.47	7.60	116.34	7.58
29	**TYR**	120.08	7.34	119.96	7.33	119.88	7.31	119.70	7.26	119.89	7.31
30	**ASN**	124.67	8.79	124.74	8.82	124.77	8.84	124.80	8.85	124.70	8.83
31	**GLY**	110.12	8.59	110.24	8.60	110.25	8.61	110.30	8.60	110.00	8.60
32	**ILE**	121.37	7.70	121.50	7.74	121.47	7.74	121.42	7.73	121.42	7.69
33	**SER**	118.89	8.08	118.85	8.07	118.89	8.10	118.98	8.14	118.94	8.11
34	**VAL**	127.17	8.80	127.25	8.83	127.23	8.84	127.16	8.84	127.15	8.80
35	**SER**	123.86	9.21	123.95	9.16	124.06	9.15	124.01	9.13	123.81	9.19
36	**ASN**	119.09	7.87			119.16	7.88	119.10	7.87	119.03	7.87
37	**ILE**	122.32	8.30	122.61	8.36	122.61	8.35	122.54	8.32	122.51	8.30
38	**SER**	124.79	9.10	124.67	9.03	124.54	9.02	124.41	9.00	124.67	9.10
39	**PRO**										
40	**ILE**	119.71	8.40	119.94	8.46	119.94	8.46	119.91	8.43	119.79	8.44
41	**TYR**	126.62	9.27	126.12	9.23	126.15	9.24	126.43	9.27	126.32	9.24
42	**GLU**	124.80	9.88	125.63	9.93	126.01	9.84	125.96	9.82	125.94	9.83
43	**THR**	124.59	8.62	124.65	8.39	119.19	9.74	118.63	9.51		
44	**ALA**	130.98	8.68	131.82	8.80	127.16	7.49	126.88	7.40		
45	**PRO**										
46	**VAL**	120.82	8.32	120.77	8.36	125.68	9.43	124.98	9.57		
47	**GLY**	112.84	8.43	112.86	8.41			116.40	9.06		
48	**TYR**	120.73	8.15	120.84	8.17			127.23	8.01		

8

#	Res										
49	THR	116.54	7.87	117.40	7.84			116.24	8.28		
50	GLU			121.88	8.39	123.74	7.16	124.30	6.98		
51	GLN	120.14	7.87	119.82	7.86			116.58	7.78		
52	PRO										
53	ASN	118.90	8.45	119.11	8.50	115.54	8.28	115.42	8.24	115.49	8.21
54	PHE	121.58	8.68	121.50	8.76	120.74	8.63	121.07	8.69	120.90	8.70
55	LEU	115.49	8.17	114.94	8.14	114.30	8.41	113.49	8.17		
56	ASN	117.66	9.22	117.68	9.18	120.16	9.51	119.77	9.44	119.95	9.47
57	LEU	114.83	8.95	114.40	8.93	114.17	8.76	113.86	8.87	114.72	8.84
58	CYS	118.40	9.51	118.75	9.58	118.58	9.51	118.41	9.47	118.10	9.45
59	VAL	116.28	9.09	115.81	9.02	116.24	9.04	115.72	9.03	116.36	9.10
60	GLU	126.68	8.84	126.67	8.85	126.83	8.88	126.48	8.84	126.75	8.86
61	ILE	119.77	8.91	120.03	8.89	119.99	8.89	119.81	8.87	119.47	8.89
62	GLN	120.20	8.74	120.47	8.77	120.41	8.78	120.39	8.76	119.99	8.73
63	THR	119.17	9.60	119.18	9.58	119.19	9.58	119.12	9.58	119.06	9.60
64	THR	115.43	7.93	115.54	7.97	115.53	8.00	115.53	7.97	115.42	7.93
65	LEU	124.19	8.54	123.99	8.53	124.01	8.53	123.89	8.52	124.07	8.51
66	THR	112.24	8.60	112.26	8.61	112.24	8.61	112.16	8.59	112.19	8.60
67	VAL	121.84	8.77	122.36	8.75	122.33	8.75	122.14	8.74	121.82	8.77
68	LEU	118.94	7.24	120.21	7.14	120.22	7.15	120.07	7.15	118.90	7.24
69	GLN	120.99	7.57	121.09	7.52	121.08	7.53	121.05	7.51	121.00	7.55
70	LEU	122.26	8.83	121.55	8.79	121.56	8.77	121.53	8.74	122.14	8.81
71	LEU	122.08	8.44	121.44	8.25	121.38	8.23	121.25	8.20	122.09	8.42
72	GLU	117.98	7.57	117.92	7.48	117.84	7.50	117.68	7.48	117.89	7.55
73	CYS	116.91	7.58	116.76	7.61	116.94	7.61	116.99	7.57	117.08	7.54
74	CYS	123.06	8.36	122.68	8.34	123.05	8.33	123.10	8.30	123.53	8.30
75	LEU	121.06	8.57	119.80	8.43	119.90	8.39	120.07	8.36	120.88	8.45
76	LYS	119.44	8.43	120.61	8.63	120.10	8.53	119.83	8.44	119.18	8.41
77	THR	118.93	7.90	118.89	7.87	118.47	7.87	118.32	7.84	118.56	7.87
78	GLU			118.78	7.53	118.76	7.51	119.52	7.51	119.48	7.74
79	GLU	118.55	8.15	119.10	7.87	118.91	7.87	118.37	7.82		
80	CYS	118.72	8.41	118.51	8.48	119.41	8.33	121.08	8.42		
81	LEU	119.73	7.59	119.56	7.54	119.81	7.24	120.04	7.09		
82	HIS			119.24	8.08	117.14	7.85	116.28	7.84		
83	ARG					119.03	7.77	118.70	7.69		
84	ILE							130.28	8.71		
85	ARG										
86	LYS										
87	GLU										
88	ARG										
89	TRP	118.74	7.97	119.14	8.01	119.07	8.02	118.89	8.06		
89	TRP	129.54	10.18	129.60	10.17	129.64	10.34	130.39	10.68		
90	GLY	109.98	7.95	109.77	7.94	106.5	7.74	102.78	7.66		
91	PRO										
92	ARG							120.87	9.50		
93	THR					113.87	8.22	116.93	6.42		
94	LEU					115.48	9.34	114.98	8.35		
95	ASP	125.95	8.78	124.53	8.44	126.68	8.60	124.92	8.19		
96	VAL	117.97	10.26	129.93	10.41	129.16	10.33	129.94	10.40		
97	ASP	124.49	9.23	126.66	9.32	126.83	9.39	127.78	9.39		
98	ILE	122.86	9.41	125.87	10.94	126.18	10.96	126.42	10.81		

99	LEU	127.53	9.00	126.84	9.11	126.95	9.10	127.26	9.00	127.68	9.02
100	LEU	113.09	7.46	113.47	7.58	113.34	7.56	113.01	7.50	113.10	7.44
101	TYR	125.29	9.41	125.30	9.44	125.33	9.43	125.23	9.42	125.34	9.40
102	GLY	114.37	8.06	114.23	8.14	114.15	8.13	114.22	8.13	114.35	8.03
103	GLU	127.06	9.15	128.35	9.28	128.34	9.28	128.28	9.26	127.09	9.16
104	GLU	118.58	8.10	118.80	8.17	118.77	8.18	118.62	8.15	118.51	8.09
105	MET	119.02	8.37	118.55	8.28	118.49	8.28	118.41	8.27	119.03	8.36
106	ILE	126.12	9.39	126.57	9.71	126.50	9.70	126.55	9.69	126.08	9.40
107	ASP	128.72	8.82	128.73	8.83	128.76	8.85	128.85	8.86	128.78	8.84
108	LEU	124.93	8.52	124.82	8.69	124.82	8.69	124.64	8.66	124.76	8.51
109	PRO										
110	LYS	122.01	8.48	121.63	7.96	121.48	8.04	121.51	8.02	121.88	8.47
111	LEU	119.69	7.37	119.96	7.43	119.97	7.43	120.12	7.44	119.58	7.35
112	SER	120.95	8.26	124.80	9.25	124.84	9.29	124.53	9.20	121.15	8.25
113	VAL	123.86	8.92	124.69	8.82	124.87	8.83	125.08	8.85	123.87	8.93
114	PRO										
115	HIS			127.03	8.67	126.79	8.60	126.41	8.47		
116	PRO										
117	ARG			118.71	9.62	118.36	9.64	118.7	9.73		
118	MET	119.98	7.79	119.77	7.50	119.48	7.55	119.76	7.55	119.71	7.72
119	ASN	112.64	8.12	113.81	8.10	114.22	8.08	113.71	8.03	112.95	8.10
120	GLU	115.57	7.54	114.64	7.33	114.44	7.34	114.41	7.31	115.16	7.48
121	ARG	120.20	7.34	120.13	7.17	121.51	7.19	121.68	7.08	120.65	7.31
122	ALA	131.66	9.55	132.18	9.78	153.79	9.38	132.69	9.36	132.37	9.33
123	PHE	107.83	6.62	108.60	6.68	112.23	8.49	111.88	8.20		
124	VAL	118.03	6.35	118.82	6.37	118.72	6.36	118.80	6.30	118.09	6.28
125	LEU	115.01	7.94	115.01	7.94	115.41	8.02	115.35	7.88	114.93	7.84
126	LEU	115.30	8.14	115.32	8.21	115.69	8.43	115.69	8.38	115.50	8.26
127	PRO										
128	LEU	117.70	7.88	117.38	7.91	117.53	8.14	117.50	8.11	117.75	7.99
129	ASN	118.52	8.40	119.01	8.51	118.97	8.50	118.89	8.46	118.70	8.43
130	ASP	115.31	7.23	115.55	7.26	115.32	7.22	115.14	7.19	115.03	7.18
131	ILE	106.55	6.99	106.53	6.99			106.51	7.00	106.51	7.00
132	ALA	128.34	8.55	128.57	8.56	128.43	8.55	128.20	8.50	127.96	8.50
133	ALA	120.59	7.38	120.59	7.42	120.63	7.44	120.59	7.41	120.69	7.37
134	ASN	107.83	8.13	107.86	8.14	86.93	8.15	107.81	8.14	107.89	8.14
135	VAL	123.63	7.56	123.52	7.53	123.58	7.53	123.58	7.53	123.67	7.56
136	VAL	125.95	8.15	125.86	8.14	125.90	8.15	125.90	8.14	125.99	8.16
137	GLU	134.40	8.92	134.44	8.89	155.47	8.88	134.33	8.87	134.58	8.91
138	PRO										
139	ARG	121.16	9.46	120.75	9.48	120.83	9.49	120.94	9.48	121.08	9.46
140	SER	116.43	9.83	116.64	9.88	116.72	9.92	116.52	9.83	116.42	9.85
141	LYS	116.90	8.36	116.79	8.41	116.82	8.41	116.67	8.38	116.79	8.33
142	LEU	121.59	8.05	121.68	8.04	121.68	8.06	121.64	8.04	121.53	8.04
143	LYS	118.31	8.57	118.21	8.56	118.25	8.56	118.21	8.55	118.34	8.58
144	VAL	123.47	8.38	123.58	8.41	123.64	8.41	123.54	8.39	123.48	8.35
145	LYS	115.01	8.86	114.99	8.88	115.00	8.91	114.87	8.88	115.07	8.90
146	ASP	120.18	7.83	120.27	7.84	120.31	7.84	120.22	7.83	120.32	7.84
147	LEU	121.02	7.82	121.50	7.79	121.54	7.78	121.48	7.77	121.06	7.79
148	VAL	117.80	7.17	117.83	7.20	117.95	7.23	117.90	7.19	117.85	7.17
149	PHE	122.24	7.32	122.22	7.36	122.40	7.49	122.47	7.48	122.24	7.44

150	**VAL**	120.75	8.48	120.87	8.50	121.05	8.49	121.17	8.47	121.04	8.48
151	**ASP**	125.78	8.11	125.60	8.19	125.53	8.19	125.56	8.18	125.68	8.13
152	**ASP**	124.96	8.84	124.96	8.86	125.27	8.85	125.16	8.82	125.32	8.83
153	**SER**	115.30	8.96	115.23	8.93	116.27	9.15	116.19	9.12	116.33	9.17
154	**VAL**	123.93	7.62	123.97	7.63	124.95	7.57	124.97	7.53	124.83	7.60
155	**LYS**	125.20	8.84	124.98	8.82	125.35	8.73	125.29	8.70	125.29	8.76
156	**ARG**	124.89	9.10	124.94	9.12	126.35	9.26	126.03	9.23	126.08	9.23
157	**TYR**	129.70	9.04	129.70	9.03	128.97	8.94	128.81	8.94	129.02	8.97
158	**LYS**	123.86	7.68	124.01	7.65	124.09	7.73	123.88	7.73	123.54	7.70

Chapter 4

Analogue Synthesis and SAR

4. Analogue Synthesis and SAR

4.1 Introduction

Chapter 3 outlined the discovery of a novel non-competitive SaHPPK inhibitor (8-MG) using in $silico$ screening. The inhibitor displayed a binding affinity of ~13 μM in both ITC and SPR experiments and an IC_{50} of ~41 μM in the biochemical assay. X-ray crystallographic analysis showed that it binds within the pterin site and makes similar interactions as the substrate (HMDP). However, the role of the sulfur atom in the binding site could not be established due to a non-native dimer in the crystal form. NMR proved essential to study SaHPPK in solution and to clarify some of the key issues regarding the loop L3 conformation in the 8-MG/AMPCPP/SaHPPK ternary complex that failed to crystallize.

Herein, the crystal structure of the 8-MG/SaHPPK binary complex is used as a starting point to propose derivatives of 8-MG that might pick up new interactions in the binding site and surrounding regions, thereby enhancing potency and potentially improving the physiochemical properties of the ligand. A series of guanine derivatives were then purchased or synthesised and tested for activity against SaHPPK in $vitro$, in order to begin to develop an understanding of the SAR associated with inhibition of the enzyme.

SPR and ITC were used to determine the binding affinities of the analogues to quantify SAR. From the X-ray structure and 2D similarity searches, only a handful of commercially-available 8-MG analogues were identified as plausible candidates to fit within the substrate binding pocket without obvious steric clashes. It became clear from even these few derivatives that the –SH group at position C^8 is an important contributor to the free energy of binding and for in $vitro$ inhibition. New analogues were thus designed and subsequently synthesised to explore the effects of substitution at the N^9 and N^7 positions of the 8-MG scaffold.

The X-ray structure of one of the N^7 substituted analogues (**11a**) was solved in complex with SaHPPK and confirmed that extension from the N^7 position of 8-MG (towards the ATP binding site) is tolerated, affording a potential route to

developing new inhibitors, including bivalent species that simultaneously target the substrate and ATP binding sites of *Sa*HPPK.

4.2 Chemical hypotheses based on the available ligand bound structures

The following crystal structures were used to identify the key intermolecular interactions contributing to the binding affinity: the *E. coli* complex with HMDP (1Q0N), the phenethyl HMDP analogue (2-amino-6-methoxy-7-methyl-7-phenethyl-7,8-dihydropterin) (1DY3), and the *S. aureus* enzyme in complex with 8-MG (3QBC; this work). The ring-stacked pyrimidine heterocycle in all three ligands is 'tailored' for the pterin pocket, as evidenced by full saturation of all hydrogen donors and acceptors (Figure 4.1). It is therefore of limited value for chemical derivatisation aimed at improving potency. This leaves the N^9, N^7 and C^8 positions as the most viable points at which to explore the effects of substituent attachment.

Firstly, five C^8 analogues were purchased to probe the tolerance to substitution at this position. Based on the superposition of the *Sa*HPPK structure onto *Ec*HPPK (1Q0N), it was predicted that residues from loop L3 such as Gly90 or Tryp89 are in close vicinity to the sulfur atom at the C^8 position, which was supported by NMR analysis (chapter 3) (Figure 4.1 D).

Secondly, in the X-ray structure of the *Ec*HPPK/HMDP analogue (Figure 4.1 B), the phenethyl ring makes a hydrophobic edge-on interaction with Trp89 and to the Leu45 residue. It was therefore proposed that extension via the equivalent N^9 position could afford a similar interaction with Trp89 and loop L3.

Thirdly, the crystal structure of *Sa*HPPK in complex with 8-MG revealed a water-filled pocket proximal to the N^7 position. Given the hydrophilic nature of this region of the pterin pocket, it was postulated that attachment of a polar substituent to the N^7 might enhance binding affinity, potentially displacing a water molecule and picking up interactions with either Mg^{2+} ions or the nearby conserved aspartic acid residues, Asp95 and Asp97. Furthermore, by merging

the pharmacophore of 8-MG with the native substrate, it was predicted that attachment of an ethyl or propyl alcohol substituent at the N^7 position of 8-MG would position a hydroxyl moiety at the same place as the substrate. The essential C^8 sulfur substitution was retained, as 8-MG had been shown to bind in the absence of the cofactor (chapter 3).
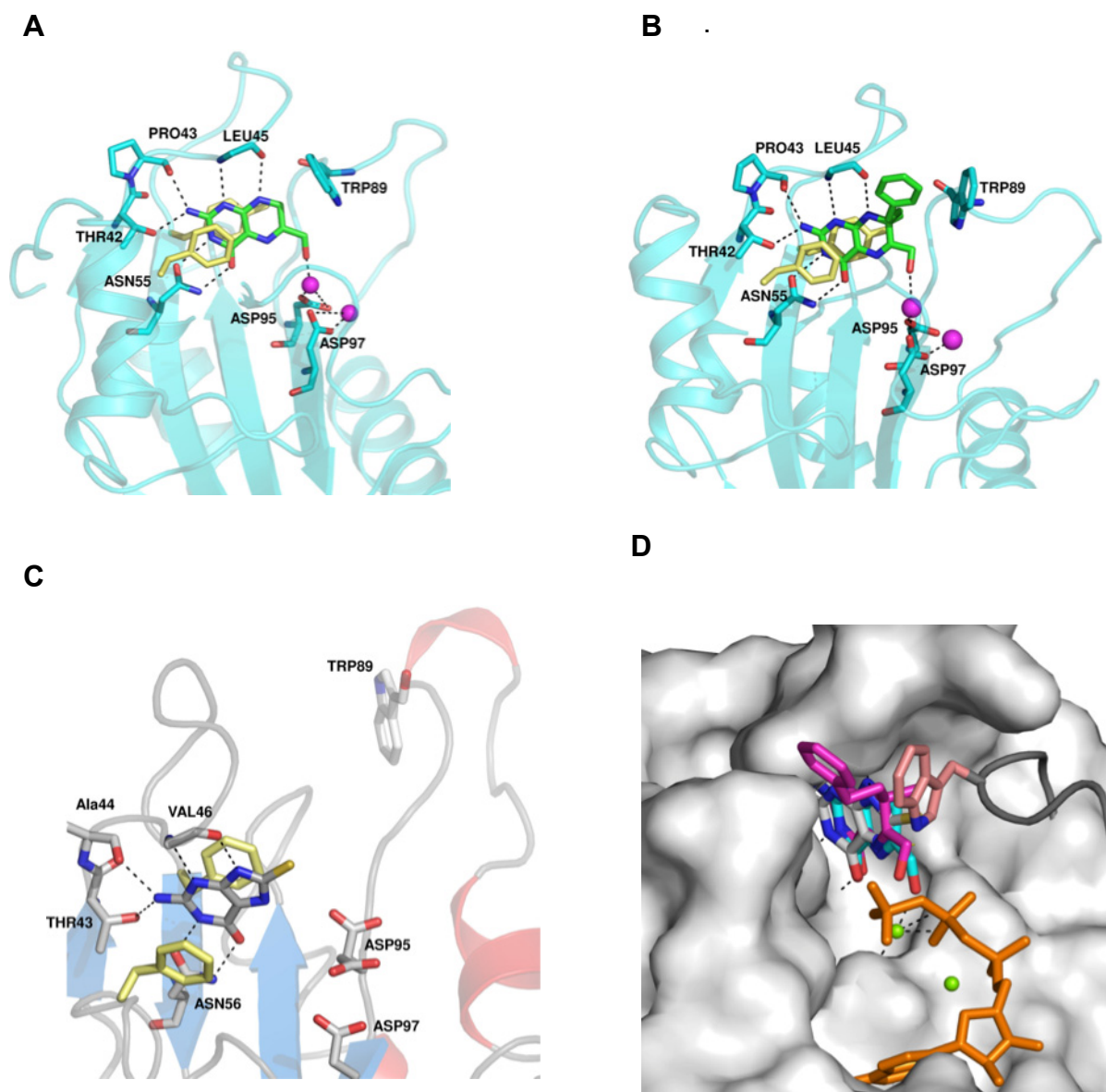
Figure 4.1: (A) HMDP (ligand carbons in green) in complex with *Ec*HPPK (1Q0N). (B) HMDP analogue (2-amino-6-methoxy-7-methyl-7-phenethyl-7,8-dihydropterin) in complex with *Ec*HPPK (1DY3). (C) 8-MG (ligand carbons in grey) in complex with *Sa*HPPK (3QBC). (D) The *Sa*HPPK and *Ec*HPPK crystal structures were aligned and superposed to compare the binding features of HMDP (cyan), HMDP phenethyl analogue (magenta) and 8-MG (grey). Surface representation of *Sa*HPPK is shown. Loop L3 of *Ec*HPPK is shown in grey. In all structures, hydrogen-bonding interactions are shown in black.

The three positions (R_1, R_2 and R_3) where varied substituents were introduced are indicated in Figure 4.2.

A

B

1
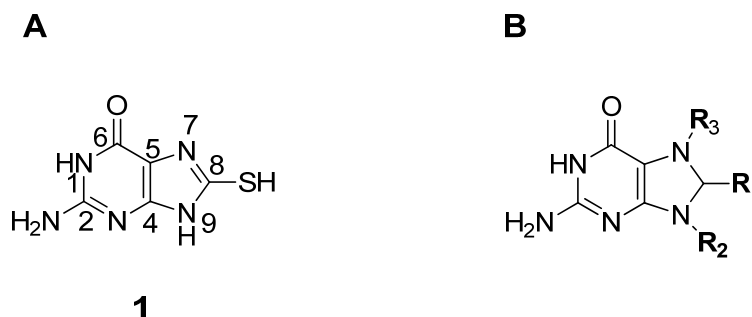
Figure 4.2: (A) Initial hit compound, 8-MG (**1**) (B) Position of R_1, R_2 and R_3 substitutions on guanine scaffold.

4.3 Design of 8-MG analogues

Analogues with different substituents at the C^8 position ($R_1 = SCH_3$, OH, CH_3, Br, morpholine) were purchased (Table 4.1, compounds **14–18**). The guanine analogues with hydrophobic N^9 substituents ($R_2 = CH_3$, C_2H_5, $CH_2C_6H_5$, $CH_2CH_2C_6H_5$), while maintaining the C^8-appended sulfur atom, were synthesised in the hope of picking up additional hydrophobic interactions (Table 4.2, compounds **5a–5d**). Chemical synthesis of 8-MG derivatives with additional substitution from the N^7 position ($R_3 = CH_2CH_2OH$, CH_2COOH, $CH_2CH_2NH_2$, $CH_2CH_2CH_2NH_2$) was also performed (Table 4.3, compounds **11a–11d**), focusing on the introduction of alcohol or amine pendants that might form hydrogen bonds to the neighbouring Asp97 or Asp95 residues. Further elaboration of these compounds to provide guanidinium derivatives was also explored (compound **12**), in the hope of forming salt bridges to the same residues.

4.4 Chemical synthesis

4.4.1 C[8]-substituted analogue

8-(Methylamino)guanosine, synthesised as described in literature[1], was hydrolysed using 1 M HCl to obtain 8-(methyamino)guanine (**13**) (Scheme 4.1) Other derivatives with C[8] substitution (**14**-**18**) were purchased from a commercial supplier (www.timtec.net/).
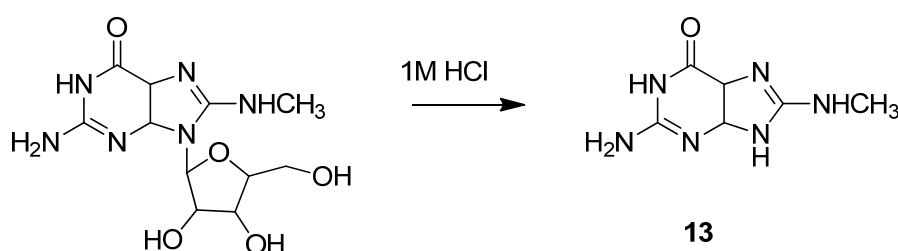
Scheme 4.1: Synthesis of 8-(methylamino)guanine.

4.4.2 N[9]-substituted isomers

The synthesis of the N[9]-substituted guanines commenced from 2-amino-6-chloropurine (**2**) as the chloro substituent is N[9]-directing[2]. Alkylation resulted in mixture of N[7]- and N[9]-alkylated isomers in 1:10 ratio respectively, that were readily separated by silica chromatography. Conversion of the 6-chloro moiety to a 6-oxo group was achieved by refluxing with 1 M HCl. The C[8] position was then brominated using NBS and subsequently replaced by sulfur by reacting with thiourea (Scheme 4.2)[3, 4]. Final compounds (**5a**–**5d**) were purified by HPLC to >95 % purity and characterised by ^{1}H and ^{13}C NMR and HR-MS analysis.
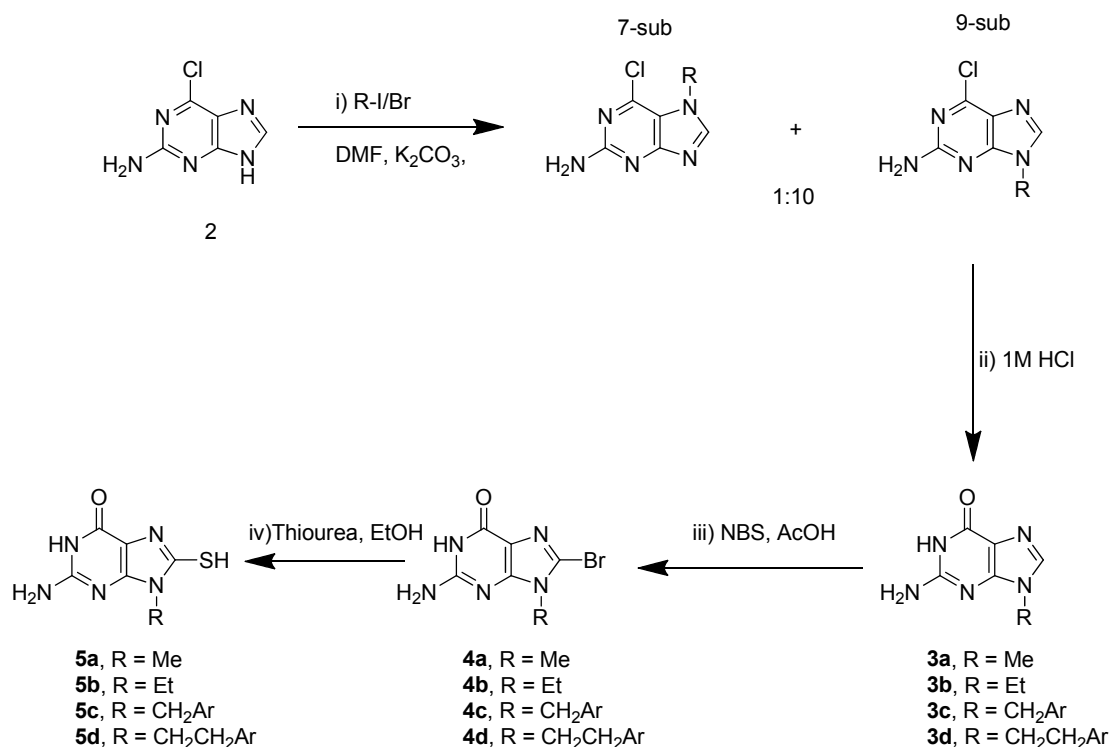
Scheme 4.2 Synthesis of N^9-subtituted 8-MG analogues.

4.4.3 N^7-substituted isomers

Compounds **7** and **8** were synthesised as described in the literature[5, 6]. Alkylation of guanines at the N^7 position is not widely reported. Benzylation of 8-bromo-N^2-acetylguanine (**8**) with benzyl bromide in the absence of base has been reported[6]. It was found that guanine **8** could be alkylated by other alkyl bromides when the reaction solution was buffered at pH 3. Alkylation reactions gave a 1:1 ratio of N^7- and N^9-alkylated products (by HPLC) from which the N^7-alkylated products (**9a–9d**) were isolated following either silica gel or reverse phase chromatography. Sulfur was appended to the 8 position using sodium thiosulfate and a catalytic quantity of aluminium-trichloride, as reported[7]. The 8-mercapto products were then deprotected under the appropriate conditions to obtain the final compounds **11a**-**11d** (Scheme 4.3). Compound **11c** was converted to the corresponding guanidine analogue through reaction with pyrazole carboxamidine (Scheme 4.4). All synthesised compounds were purified by HPLC to >95 % purity.
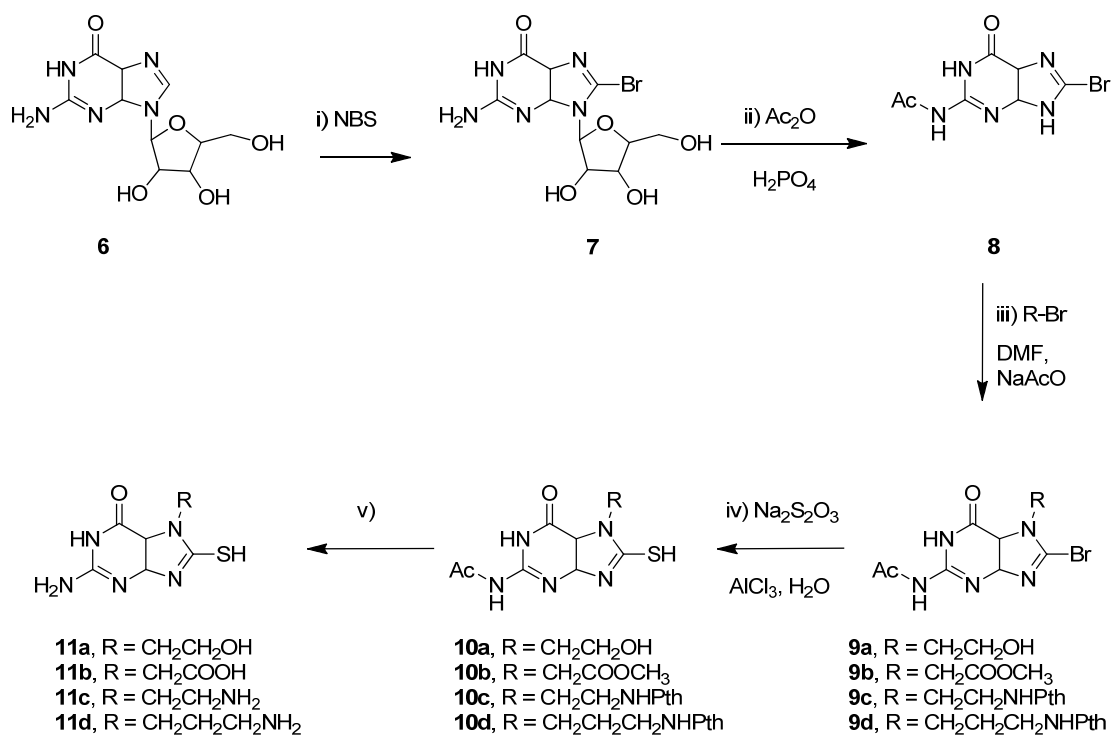
Scheme 4.3: Synthesis of N^7-subtituted 8-MG analogues.
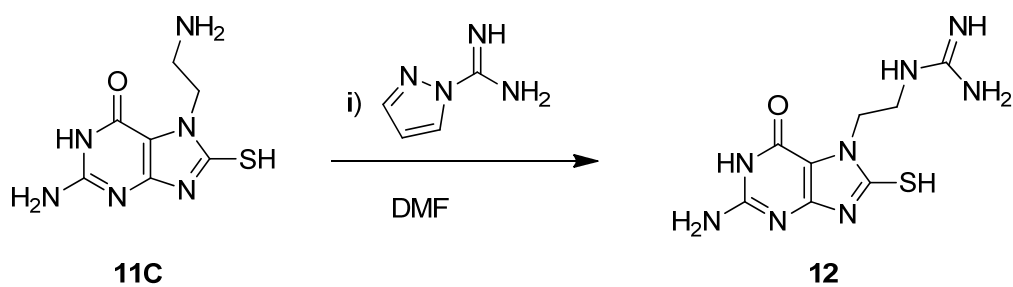
Scheme 4.4: Synthesis of 7-(2-guanidinoethyl)-8-mercaptoguanine.

4.5 Ligand binding studies

4.5.1 SPR analysis of binding

Binding of the compounds to *Sa*HPPK was quantitatively assessed using SPR to determine the dissociation constant (K_d).

Compounds (**13**–**18**) explore substitution at the C^8 position of 8-MG. SPR data indicated no binding of the –NHCH$_3$ (**13**) or –SCH$_3$ (**14**) derivatives to the enzyme (Table 4.1). Compounds **15**–**18** displayed 15-20-fold lower affinity (Figure 4.3 A, Table 4.1) than the starting hit (8-MG). These data suggest that thiol substitution at C^8 contributes significantly to binding to *Sa*HPPK. Interestingly, compound **13** displayed binding (K_d = 108 µM) when SPR experiments were performed in the presence of saturating amounts of ATP.

Analogues with hydrophobic substituents at the N^9 position of 8-MG (**5a**–**5d**) resulted in a 10-20-fold loss in affinity, indicating that this position is not favourable for modifications (Figure 4.3, Table 4.2). Compounds **11a**, **c** and **d**, however, retained affinity for the enzyme (Figure 4.4, Table 4.3), indicating that addition of substituents at the N^7 position is tolerated (although **11b**, with a carboxylate pendant, displayed no binding). Compound **11a**, with an ethyl alcohol pendant, displayed comparable affinity to 8-MG (K_d = ~12 µM), whilst the analogues with amine and guanidinum pendants (**11c**, **11d, 12**) displayed slightly weaker binding to *Sa*HPPK.

Binding of compounds **11a, 11c** and **11d** was 10-15-fold weaker under saturating ATP conditions, suggesting that binding is competitive with the cofactor.
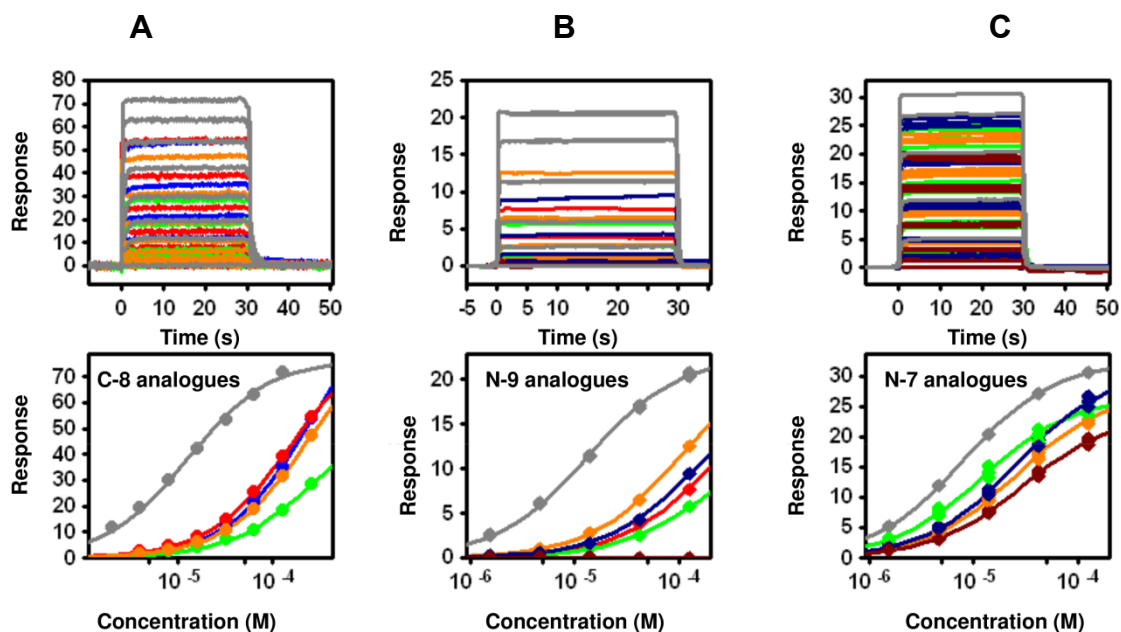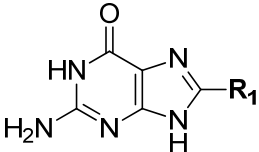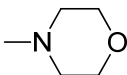
Figure 4.3: SPR raw data (*top*) and steady-state response curves (*bottom*) for the binding of C^8- (**13**-**18**), N^9- (**5a**-**5d**) and N^7- (**11a**-**11d**) substituted analogues to *Sa*HPPK. Note: For N^9- *and* N^7- substituted analogues site directed biotinylation was used whereas for C^8 *derivatives* minimal-biotinylation was used to immobilize the protein.
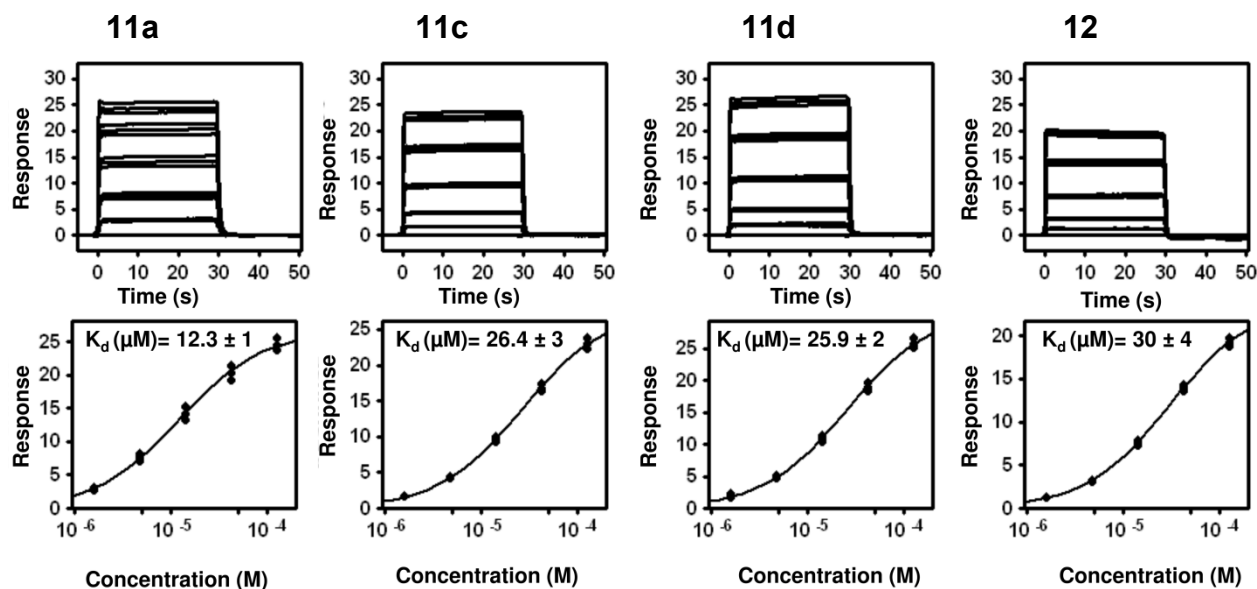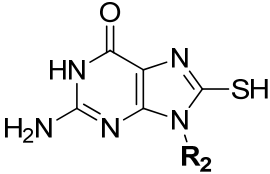
Figure 4.4: Raw data (*top*) and steady-state response curves (*bottom*) for the binding of compounds**11a, 11c**, **11d** and **12** to *Sa*HPPK.

Table 4.1: Structures of C^8-substituted guanine analogues and their *Sa*HPPK binding affinities, as determined by SPR.

C^8-substituted analogues			

Compound	R_1	$K_d{}^a$ (µM)	$K_d{}^b$ (µM)
13	NHMe	No binding	108± 5
14	SMe	No Binding	nd[c]
15	Me	159 ± 1	nd[c]
16	Br	248 ± 3	nd[c]
17	OH	257 ± 5	nd[c]
18	—N⌷O (morpholine)	246 ± 3	nd[c]
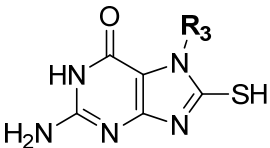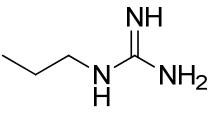
[a] K_d in the absence of ATP, [b] K_d in the presence of saturating amount of ATP, nd = No data

Table 4.2: Structures of N^9-substituted guanine analogues and their *Sa*HPPK binding affinities, as determined by SPR.

N^9-substituted analogues			

Compound	R$_2$	K_d^a(µM)	K_d^b(µM)
5a	Me	194 ± 5	410± 10
5b	Et	190 ± 10	340± 10
5c	Bn	106 ± 2	201± 6
5d	CH$_2$Bn	145 ± 4	510± 20

^{a} K_d in the absence of ATP, ^{b} K_d in the presence of saturating amount of ATP, nd = No data

Table 4.3: Structures of N^7-substituted guanine analogues and their SaHPPK binding affinities, as determined by SPR.

N^7-substituted analogues			

Compound	R₃	$K_d{}^a$(µM)	$K_d{}^b$(µM)
11a	CH₂CH₂OH	12.3 ± 1	130± 10
11b	CH₂COOH	No binding	140± 4
11c	CH₂CH₂NH₂	26.4 ± 3	160± 20
11d	CH₂CH₂CH₂NH₂	25.9 ± 2	121± 6
12		30 ± 4	No binding

^{a} K_d in the absence of ATP, ^{b} K_d in the presence of saturating amount of ATP

4.5.2 ITC analysis of binding

To confirm the ligand binding affinities, and to further determine the thermodynamic parameters of binding for selected compounds (**11a, 11c-d, 12**) isothermal titration calorimetry (ITC) experiments were performed. These analogues were found to bind SaHPPK with K_d values in the 17–35 µM range according to ITC, in excellent agreement with the values obtained by SPR. Interestingly, compound **11a** (K_d = 16.7 µM), which has a similar affinity to 8-MG (**1**) (K_d = 12.8 µM), showed a less favourable enthalpic contribution and more favourable entropic contribution towards the ΔG of binding as compared to 8-MG (Figure 4.5, Table 4.4).
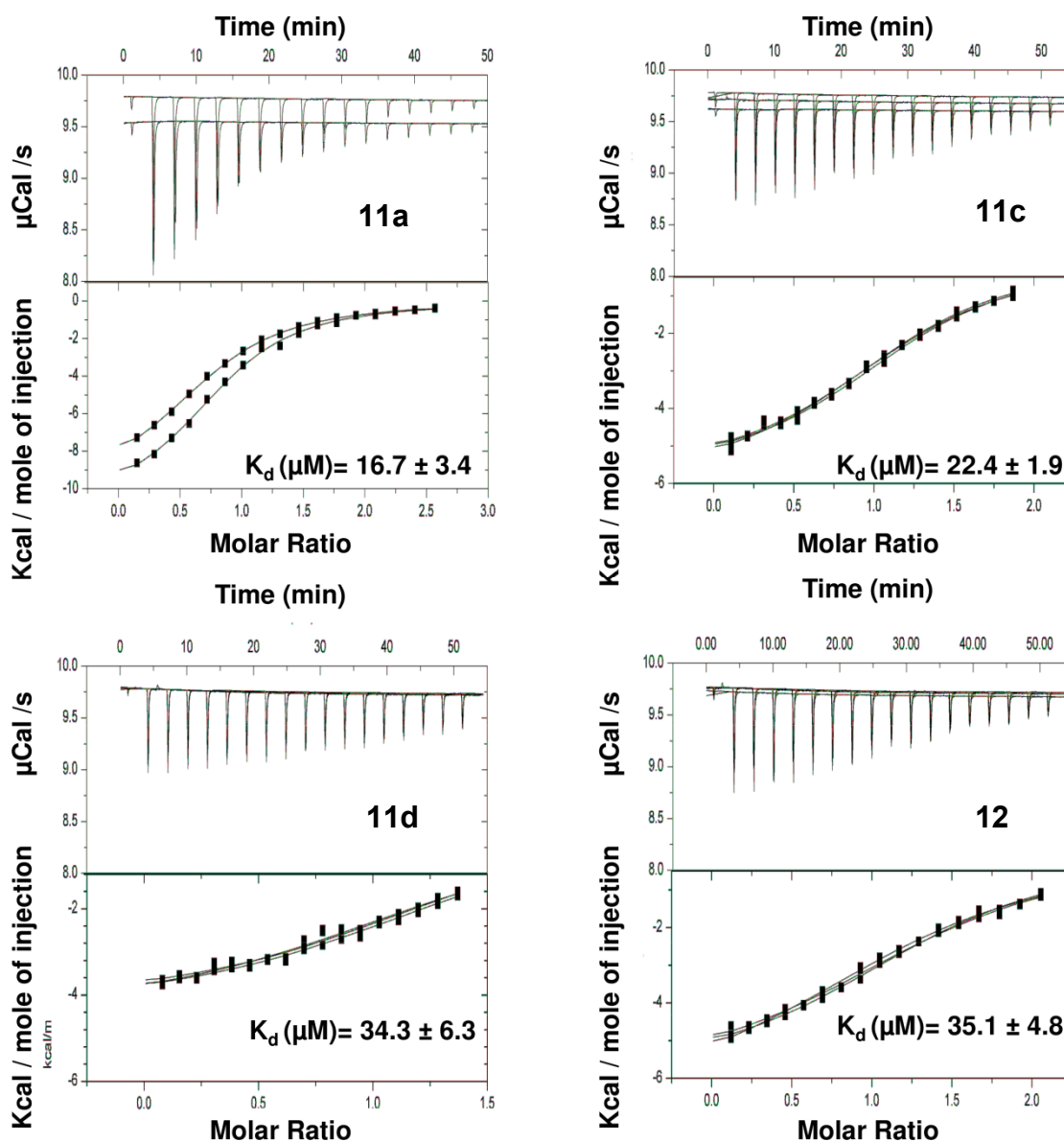
Figure 4.5: Raw data (*top*) and integrated ITC data (*bottom)* for the titration of *Sa*HPPK with compounds **11a**, **11c, 11d** and **12**. The data was obtained in triplicate.

Table 4.4: Thermodynamic and kinetic parameters for the binding of ligands to *Sa*HPPK, as determined by ITC and SPR[a]

Ligand	ΔH	$T\Delta S$	ΔG	N	K_d (μM) ITC	K_d (μM) SPR
1	-19.6±3.4	-12.9±3.5	-6.7±0.2	0.96±0.1	12.8±3.4	10.8±0.4
11a	-10.5±0.6	-4.1±0.5	-6.4±0.1	0.8±0.1	16.7±3.5	12.3±1.0
11c	-5.8±0.1	0.4±0.2	-6.2±0.1	1.2±0.02	22.4±2	26.4±3.0
11d	-6.2±0.4	-0.2±0.5	-6.0±0.1	1.3±0.05	34.3±6.2	25.9±2.0
12	-4.3±0.1	1.7±0.2	-6.0±0.1	1.4±0.02	35.1±4.8	30.0±4.0

[a] Values are the means ± the standard deviation for at least three experiments. All ITC and SPR experiments were performed at 298 K and 293 K, respectively.

4.6 X-ray structure in complex with compound 11a

Co-crystallisation experiments were performed with compounds **11a**, **11c**, **11d** and **12** to obtain the structure of the bound compounds in complex with *Sa*HPPK. However, diffraction quality crystals were only obtained for compound **11a** and a high resolution X-ray structure (1.85 Å) was solved using molecular replacement. The binary complex crystallised in the $P2_1$ space group, with a head-tail dimer in the asymmetric unit, similar to our previous structure (3QBC). Excellent quality electron density was observed, which showed the ligand bound to both monomers (Figure 4.6 A).

The crystal structure confirmed that **11a** binds to the pterin site in a similar fashion to 8-MG (3QBC) and that the hydroxyethyl pendant is directed towards the extra space available within the active site. The alcohol group appears to make two hydrogen bonds with a pair of bound water molecules (Figure 4.6 B). Presumably these compensate for the lack of hydrogen bond between the N^9 hydrogen of 8-MG and the backbone carbonyl of Val46.

A

B
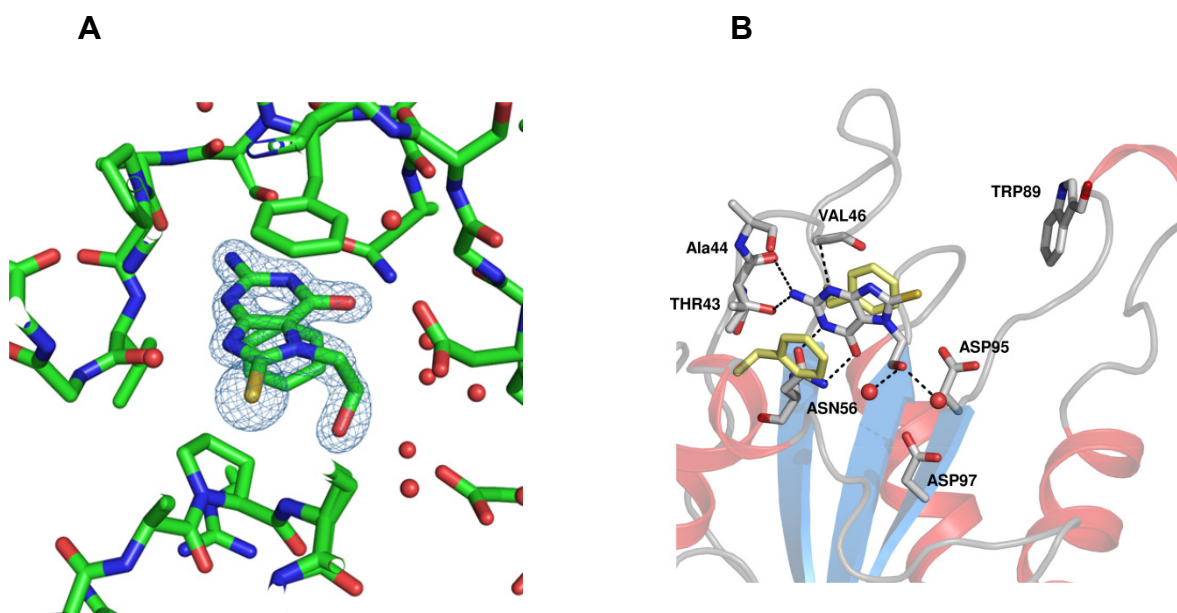
Figure 4.6: (A) $2F_O$-F_C electron density map of the pterin binding site, contoured at 2.0 sigma, showing density for 7-(2-hydroxyethyl)-8-mercaptoguanine (**11a**). (B) Intermolecular interactions (black) involving **11a** within the pterin site.

4.7 Conclusion

In conclusion, C^8, N^9, and N^7 substitutions on the guanine scaffold were explored to search chemical space with a view to increasing the potency of the initial hit compound, 8-MG, for *Sa*HPPK. The 8-MG/*Sa*HPPK crystal structure showed that the pyrimidine ring of guanine is important for recognition. Early SAR suggested that thiol substitution at C^8 position is a requisite for tight binding to *Sa*HPPK. N^9-alkylations on 8-MG were detrimental to binding, resulting in a 10-20 fold decrease in affinity. On the other hand, N^7-substituents extending into the hydrophilic pocket within the active site were generally well tolerated. The crystal structure of *Sa*HPPK in complex with **11a** showed that while substitution at the N^7 position results in a loss of one of the intermolecular hydrogen bonds formed by 8-MG (to Val46), this is compensated for by the interaction of the hydroxyethyl group with crystallographic water molecules. Further chemical evolution of the 8-MG scaffold towards the absolutely conserved metal-binding residues, Asp95 and Asp97, may yield analogues with enhanced affinities.

4.8 Experimental

4.8.1 General Experimental

All starting materials, reagents and solvents were obtained from commercial suppliers and were used without further purification unless stated otherwise. Compounds **14–18** were acquired from commercial sources (www.timtec.net/). ^{1}H/^{13}C NMR spectra were recorded with a Bruker Avance Ultrashield NMR spectrometer operating at field strength of 9.4 Tesla. Results were recorded as follows: chemical shift values are expressed in δ units acquired in the indicated solvent and at a temperature of 25°C, unless noted otherwise, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), integration and coupling constants (*J*) in Hertz. Mass spectra were acquired in the positive and negative mode using electrospray ionisation obtained on a Micromass Platform II Single Quadropole Mass Spectrometer. High Resolution Mass Spectrometry (HRMS) analyses for final compounds were recorded on a Brüker Apex-II Fourier Transform Ion Cyclotron Resonance (FTICR) Mass Spectrometer with an electrospray ion source (ESI). Melting point (mp) determination for final compounds was performed uncorrected using a Mettler Toledo MP50 melting point apparatus. Analytical reverse-phase HPLC was performed on a Waters HPLC system fitted with a Phenomenex® Luna C8 (2) 100Å column (150 mm × 4.6 mm, 5 µm) using a binary solvent system; solvent A: 0.1% TFA/H_2O; solvent B: 0.1% TFA/80%ACN/H_2O. Gradient elution was achieved using 100% solvent A to 100% solvent B over 20 min at a flow rate of 1 mL/min detecting at 254 nm. Preparative HPLC was performed with 40 minutes runs using a Waters Associates liquid chromatography system (Model 600 Controller and Waters 486 Tunable Absorbance Detector) by RP-HPLC using the binary solvent system specified above and a gradient of 0-80% solvent B in solvent A over 30 minutes at a flow rate of 7 mL/min on a Phenomenex®Axia 10µm Luna C8(2) 100 Å column (250 mm ×20 mm, 10 µm) using Empower Pro software to process absorbance signals at 254 nm.

4.8.1.2 9-Methylguanine (3a)

To a solution of 2-amino-6-chloropurine (**2**) (1.0 g, 5.9 mmol) in DMF (10 mL) was added methyliodide (367 µL, 5.89 mmol) and K_2CO_3 (815 mg, 5.89 mmol). After stirring overnight at 20 °C the solution was evaporated to dryness under reduced pressure. The crude material was redissolved in chloroform and the mixture of N^9 and N^7 isomers separated using silica gel chromatography (CHCl$_3$/MeOH). The N^9 isomer (488 mg, 2.70 mmol) was refluxed in 1 M HCl (20 mL) for 2 hours, then cooled to room temperature. The precipitate that formed was collected by filtration and dried under a stream of air to provide the title compound as a white powder (312 mg, 70%). **^{1}H NMR** (400 MHz, DMSO-d6) δ 11.97 (s, 1H), 9.21 (s, 1H), 7.50 (s, 2H), 3.71 (s, 3H). **^{13}C NMR** (101 MHz, DMSO) δ 162.3, 158.1, 157.8, 118.7, 115.8, 35.7. **LRMS** (ESI): *m/z*: 166.2 [M+H]$^+$, (100%).

4.8.1.3 9-Ethylguanine (3b)

2-Amino-6-chloropurine (**2**) (1.0 g, 5.9 mmol) was treated with ethyliodide (472 µL, 5.89 mmol) and K_2CO_3 (815 mg, 5.89 mmol). The resulting products were separated and the N^9 isomer hydrolysed as described in the preparation of **3a** to provide the title compound as a white powder (527 mg, 50%). **^{1}H NMR** (400 MHz, DMSO-d$_6$) δ 12.03 (s, 1H), 9.34 (s, 1H), 7.58 (s, 2H), 4.15 (q, *J* = 7.3 Hz, 2H), 1.43 (t, *J* = 7.3 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-d$_6$) δ 155.8, 153.1, 149.4, 136.7, 107.2, 34.0, 14.2. **LRMS** (ESI): *m/z*: 180.1 [M+H]$^+$, (100%).

4.8.1.4 9-Benzylguanine (3c)

2-Amino-6-chloropurine (**2**) (1.0 g, 5.9 mmol) was treated with benzylbromide (700 µL, 5.89 mmol) and K_2CO_3 (815 mg, 5.89 mmol). The resulting products were separated and the N^9 isomer hydrolysed as described in the preparation of **3a** to provide the title compound as a white powder (1.3 g, 90%). **^{1}H NMR** (400 MHz, DMSO-d$_6$) δ 11.79 (s, 1H), 9.18 (s, 1H), 7.41 – 7.29 (m, 7H), 5.35

(s, 2H). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 155.8, 153.8, 150.0, 137.3, 135.3, 129.0, 128.4, 127.8, 108.8, 47.5. LRMS (ESI): *m/z*: 242.2 [M+H]$^+$, (100%).

4.8.1.5 9-Phenethylguanine (3d)

2-Amino-6-chloropurine (**2**) (1.0 g, 5.9 mmol) was treated with phenethylbromide (798 µL, 5.89 mmol) and K$_2$CO$_3$ (815 mg, 5.89 mmol). The resulting products were separated and the N^9 isomer hydrolysed as described in the preparation of **3a** to provide the title compound as a white powder (1.35 g, 90%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 11.78 (s, 1H), 8.89 (s, 1H), 7.34–7.15 (m, 7H), 4.34 (t, *J* = 7.3 Hz, 2H), 3.15 (t, *J* = 7.3 Hz, 2H). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 155.5, 153.6, 149.7, 137.1, 136.9, 128.6, 128.6, 126.8, 108.4, 45.5, 34.1. LRMS (ESI): *m/z*: 256.2 [M+H]$^+$, (100%).

4.8.1.7 8-Bromo-9-methylguanine (4a)

To a suspension of **3a** (320 mg, 1.94 mmol) in glacial acetic acid (15 mL) was added *N*-bromosuccinimide (464 mg, 2.62 mmol) and the reaction was stirred overnight. The solution was poured over ice cold water and the resulting precipitate was filtered off, washed with water and methanol and dried to obtain the title compound as yellow powder (284 mg, 60%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 10.68 (s, 1H), 6.52 (s, 2H), 3.47 (s, 3H). LRMS (ESI): *m/z*: 244.1 [M+H]$^+$, (100%).

4.8.1.8 8-Bromo-9-ethylguanine (4b)

Compound **3b** (212 mg, 0.82 mmol) was treated with N-bromosuccinimide (211 mg, 1.19 mmol) in glacial acetic acid (15 mL). The resulting product was filtered, washed with water and methanol and dried to obtain the title compound as described in the preparation of **4a** as a yellow powder (122 mg, 58%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 10.66 (s, 1H), 6.58 (s, 2H), 3.96 (q, *J* = 7.2 Hz, 2H), 1.25 (t, *J* = 7.2 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 155.5,

153.8, 152.0, 120.2, 116.8, 31.4, 14.5. **LRMS** (ESI): *m/z*: 258.1 [M+H]$^+$, (100%).

4.8.1.9 8-Bromo-9-benzylguanine (4c)

Compound **3c** (455 mg, 1.89 mmol) was treated with *N*-bromosuccinimide (453 mg, 2.55 mmol) in glacial acetic acid (30 ml). The resulting product was filtered, washed with water and methanol and dried to obtain the title compound as described in the preparation of **4a** as a yellow powder (384 mg, 65%). **^{1}H NMR** (400 MHz, DMSO-d$_6$) δ 10.72 (s, 1H), 7.38–7.15 (m, 5H), 6.60 (s, 2H), 5.16 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d$_6$) δ 155.6, 154.0, 152.2, 135.8, 128.6, 127.7, 126.6, 120.8, 115.0, 45.4. **LRMS** (ESI): *m/z*: 320.1 [M+H]$^+$, (100%).

4.8.1.10 8-Bromo-9-phenethylguanine (4d)

Compound **3d** (1.17 g, 4.57 mmol) was treated with *N*-bromosuccinimide (1.1 g, 6.2 mmol) in glacial acetic acid (90 ml). The resulting product was filtered, washed with water and methanol and dried to obtain the title compound as described in the preparation of **4a** as a yellow powder (914 mg, 60%). **^{1}H NMR** (400 MHz, DMSO-d$_6$) δ 10.68 (s, 1H), 7.49–6.89 (m, 5H), 6.59 (s, 2H), 4.15 (t, *J* = 7.8 Hz, 2H), 3.00 (t, *J* = 7.8 Hz, 2H). **^{13}C NMR** (101 MHz, DMSO-d$_6$) δ 155.5, 153.8, 151.9, 137.3, 128.6, 128.4, 126.6, 120.7, 115.0, 43.9, 34.3. **LRMS** (ESI): *m/z*: 334.2 [M+H]$^+$, (100%).

4.8.1.12 8-Mercapto-9-methylguanine (5a)

A solution of **4a** (70 mg, 0.3 mmol) and thiourea (44 mg, 0.6 mmol) in EtOH (5 mL) was heated to 80 °C overnight. After cooling to room temperature, the mixture was concentrated to dryness under reduced pressure. The crude product was further purified using reverse phase chromatography (C18, 0.1% TFA in water) to provide the title compound as an off white solid (29 mg, 49%). **Mp** 259–261 °C (decomp.), **^{1}H NMR** (400 MHz, DMSO-d6) δ 7.95 (s,

2H), 2.89 (s, 3H). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 162.3, 158.1, 157.8, 118.7, 115.8, 35.7. **LRMS** (ESI): *m/z*: 198.1 [M+H]$^+$, (100%), **HRMS** (ESI): observed *m/z:* 198.0436 [M+H]$^+$; calculated *m/z:* 198.0405 [M+H]$^+$ **RP-HPLC**: t$_R$ 4.57, >95% (gradient).

4.8.1.13 8-Mercapto-9-ethylguanine (5b)

Compound **4b** (102 mg, 0.40 mmol) was refluxed with thiourea (60 mg, 0.8 mmol) in EtOH (5mL).The reaction was concentrated to dryness under reduced pressure and the crude product was further purified using reverse phase chromatography (C18, 0.1% TFA in water) to obtain the title compound as described in the preparation of **5a** as an off-white solid (51 mg, 60%). **Mp** 240-244 °C (decomp.), ^{1}**H NMR** (400 MHz, DMSO-d$_6$) δ 12.71 (s, 1H), 10.87 (s, 1H), 6.66 (s, 2H), 4.01 (q, *J* = 7.1 Hz, 2H), 1.23 (t, *J* = 7.1 Hz, 3H). ^{13}**C NMR** (101 MHz, DMSO-d$_6$) δ 155.8, 153.1, 149.4, 136.7, 107.2, 34.0, 14.2. **LRMS** (ESI): *m/z*: 212.1 [M+H]$^+$, (100%), **HRMS** (ESI): observed *m/z:* 212.0603 [M+H]$^+$; calculated *m/z:* 212.0601 [M+H]$^+$, **RP-HPLC**: t$_R$ 5.26, >98% (gradient).

4.8.1.14 8-Mercapto-9-benzylguanine (5c)

Compound **4c** (110 mg, 0.34 mmol) was refluxed with thiourea (131 mg, 1.70 mmol) in EtOH (5 mL). The reaction was concentrated to dryness under reduced pressure and the crude product was further purified using reverse phase chromatography (C18, 0.1% TFA in water) to obtain the title compound as described in the preparation of **5a** as a off white solid (65 mg, 70%). **Mp** >300 °C (decomp.), ^{1}**H NMR** (400 MHz, DMSO-d$_6$) δ 12.87 (s, 1H), 10.94 (s, 1H), 7.41–7.15 (m, 5H), 6.64 (s, 2H), 5.21 (s, 2H). ^{13}**C NMR** (101 MHz, DMSO-d$_6$) δ 164.7, 154.1, 150.8, 150.0, 136.4, 128.3, 127.3, 127.2, 103.7, 44.9. **LRMS** (ESI): *m/z*: 274.1 [M+H]$^+$, (100%), **HRMS** (ESI): observed *m/z:* 274.0770 [M+H]$^+$; calculated *m/z:* 274.757 [M+H]$^+$, **RP-HPLC**: t$_R$ 7.31, >95% (gradient).

4.8.1.15 8-Mercapto-9-phenethylguanine (5d)

Compound **4d** (500 mg, 1.50 mmol) was refluxed with thiourea (228 mg, 3.00 mmol) in EtOH (25 mL).The reaction was concentrated to dryness under reduced pressure and the crude product was further purified using reverse phase chromatography (C18, 0.1% TFA in water) to obtain the title compound as described in the preparation of **5a** as a off white solid (260 mg, 60%). **Mp** 291-295 °C (decomp.), **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.76 (s, 1H), 10.90 (s, 1H), 7.35 – 7.17 (m, 5H), 6.65 (s, 2H), 4.18 (t, J = 7.8 Hz, 2H), 2.99 (t, J = 7.8 Hz, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.1, 154.0, 150.8, 149.9, 138.0, 128.5, 128.4, 126.4, 103.6, 43.0, 33.1. **LRMS** (ESI): m/z: 288.2 (M+H, 100%), **HRMS** (ESI): observed $m/z:$ 288.0925 [M+H]$^+$; calculated $m/z:$ 288.0914 [M+H]$^+$, **RP-HPLC**: t_R 7.9, >98% (gradient).

4.8.1.16 8-Bromoguanosine (7)

Guanosine (**6**) (16.5 g, 58.3 mmol) was suspended in DMF (100 mL), NBS (12.0 g, 67.5 mmol) added and the suspension stirred overnight at room temperature. By this time all the guanosine had dissolved to a clear yellow solution. Solvent was removed under reduced pressure and 40-50 mL cold water was added to the residue. The resulting precipitate was separated by filtration and recrystallised from hot water to obtain the title compound as a white powder (14.7 g, 70%). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.82 (s, 1H), 6.50 (s, 2H), 5.69 (d, J = 6.3 Hz, 1H), 5.44 (d, J = 6.2 Hz, 1H), 5.08 (d, J = 5.1 Hz, 1H), 5.02 (q, J = 6.0 Hz, 1H), 4.92 (t, J = 6.0 Hz, 1H), 4.21 – 4.05 (m, 1H), 3.92 – 3.78 (m, 1H), 3.74 – 3.59 (m, 1H), 3.58 – 3.46 (m, 1H). **LRMS** (ESI): m/z: 362.1 [M+H]$^+$, (100%).

4.8.1.17 8-Bromo-N^2-acetylguanine (8)

A mixture of 8-bromoguanosine (**7**) (11.7 g, 32.4 mmol), Ac$_2$O (60 mL) and H$_3$PO$_4$ (0.5 mL) was stirred at 100 °C for 3 h and cooled to r.t. The resulting precipitate was separated by filtration and washed thoroughly with chloroform to yield the title compound as an off white solid (7.9 g, 90%). **^{1}H NMR** (400

MHz, DMSO-d_6) δ 2.16 (s, 3H). **LRMS** (ESI): *m/z*: 273 [M+H]$^+$, (100%), 274 [M+H]$^+$, (100%).

4.8.1.18 N^2-Acetyl-8-bromo-7-(2-hydroxyethyl)guanine (9a)

To a solution of N^2-acetyl-8-bromoguanine (**8**) (100 mg, 0.37 mmol) in DMF (1 mL) was added 2-bromoethanol (50 µL, 0.7 mmol) and DIPEA (32 µL, 0.2 mmol). The reaction was heated at 100 °C for 24 hours with periodic addition of DIPEA in order to maintain the pH between 3 and 4. The solution was dilute with water (5 mL) and purified by using reverse phase chromatography (C18, 0-4% ACN and 0.01% TFA in water) to provide the title compound as a white solid (30 mg, 26%). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 4.31 (t, *J* = 5.6 Hz, 2H), 3.72 (t, *J* = 5.6 Hz, 2H), 2.16 (s, 3H).**^{13}C NMR** (101 MHz, CDCl$_3$) δ 173.5, 156.5, 151.5, 147.4, 131.3, 113.7, 59.8, 49.4, 23.7. **LRMS** (ESI): *m/z*: 315 [M+H]$^+$, (100%), 317[M+H]$^+$, (100%).

4.8.1.19 Methyl-8-bromo-(N^2-acetylguanin-7-yl)acetate (9b)

DIPEA (1.3 mL, 7.4 mmol) and methyl bromoacetate (386 µL, 4.10 mmol) were added to a suspension of N^2-acetyl-8-bromoguanine (**8**) (1.0 g, 3.7 mmol) in DMF (5 mL) under N$_2$. After 20 h of stirring at room temperature, the solvent was removed in vacuo, and the residue was coevaporated with methanol (3 x). The crude material was redissolved in DCM and the mixture of N-9 and N-7 isomers were separated using silica gel chromatography (DCM/MeOH) to provide the title compound as a white solid (255 mg, 20%). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.95 (s, 1H), 5.02 (s, 2H), 3.70 (s, 3H), 2.16 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 173.5, 168.1, 154.7, 148.9, 147.9, 140.1, 119.6, 52.5, 44.1, 23.7. **LRMS** (ESI): *m/z*: 345.9 [M+H]$^+$, (100%), 347 [M+H]$^+$, (100%).

4.8.1.20 *N*-2- (8-bromo-*N*²-acetylguanin-7-yl)ethylphthalimide (**9c**)

DIPEA (960 µL, 5.60 mmol) and *N*-(2-bromoethyl)phthalimide (560 mg, 1.85 mmol) were added to a suspension of *N*²-acetyl-8-bromoguanine (**8**) (500 mg, 1.85 mmol) in DMF (5 mL) under N_2. The reaction was refluxed at 100 °C overnight, the solvent was removed *in vacuo*, and the residue was coevaporated with methanol (3 x). The crude mixture of N^9 and N^7 isomers was purified using silica gel chromatography (20 % petroleum benzene/80% ethylacetate/methanol) to provide the title compound as a white solid (205 mg, 25%). **¹H NMR** (400 MHz, DMSO-d_6) δ 8.00 – 7.59 (m, 4H), 4.50 (t, *J* = 7.8 Hz, 2H), 4.04 (t, *J* = 7.8 Hz, 2H), 2.15 (s, 3H). **¹³C NMR** (101 MHz, DMSO-d_6) δ 173.5, 167.3, 156.4, 151.4, 147.3, 134.5, 131.3, 130.5, 123.1, 113.8, 45.7, 37.3, 23.7. **LRMS** (ESI): *m/z*: 445 [M+H]⁺, (100%), 446 (M+H, (100%).

4.8.1.21 *N*-3- (*N*²-acetylguanin-7-yl)propylphthalimide (**9d**)

A mixture of *N*²-acetyl-8-bromoguanine (**8**) (810 mg, 2.98 mmol), N-(3-bromopropyl)phthalimide (1.10 g, 4.03 mmol), DIPEA (1.6 ml, 9.0 mmol) in DMF was refluxed at 100 °C overnight, after which the solvent was evaporated in vacuo. The crude product was diluted with water and extracted with chloroform. The pooled organic phases were dried (MgSO$_4$), filtered, and evaporated to dryness. This residue was dissolved in a minimum volume of CHCl$_3$ and purified using silica gel chromatography (CHCl$_3$/MeOH) to provide the title compound as a white solid (276 mg, 20%). **¹H NMR** (400 MHz, DMSO-d_6) δ 7.93 – 7.70 (m, 4H), 4.34 (t, *J* = 7.0 Hz, 2H), 3.63 (t, *J* = 7.0 Hz, 2H), 2.27 – 2.04 (m, 5H). **¹³C NMR** (101 MHz, DMSO-d_6) δ 173.4, 167.8, 156.3, 151.4, 147.4, 134.3, 131.6, 130.1, 122.9, 113.5, 44.7, 34.6, 28.6, 23.6. **LRMS** (ESI): *m/z*: 459 [M+H]⁺, (100%), 460 [M+H]⁺, (100%).

4.8.1.22 7-(2'-Hydroxyethyl)-8-mercaptoguanine (11a)

To a solution of N^2-acetyl-8-bromo-7-(2-hydroxyethyl)guanine (9a) (10 mg, 0.03 mmol) in water (4 mL) and acetonitrile (2 mL) was added sodium thiosulfate (10 mg, 0.1 mmol) and aluminium chloride (0.015 mmol). The solution was refluxed for 24 hours and purified by reverse phase chromatography (C18, 0.01% TFA in water) to provide the title compound as a white powder (5 mg, 69%). **Mp** >300 ^{o}C (decomp.), **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.91 (s, 2H), 6.54 (s, 5H), 4.79 (t, J = 5.7 Hz, 3H), 4.25 (t, J = 6.7 Hz, 6H), 3.64 (dd, J = 6.7, 5.7 Hz, 6H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.1, 154.1, 151.4, 149.5, 105.3, 58.5, 46.4. **LRMS** (ESI): m/z: 228.1 [M+H]$^+$, (100%), **HRMS** (ESI): observed m/z: 226.0396 [M-H]$^-$; calculated m/z: 226.0404 [M-H]$^-$, **RP-HPLC**: t_R 4.14, >98% (gradient).

4.8.1.23 (Guanin-7-yl)acetic acid (11b)

$Na_2S_2O_3.5H_2O$ (200 mg, 1.10 mmol,) was added to the suspension of methyl-8-bromo-(N^2-acetylguanin-7-yl)acetate (9b) (95.0 mg, 0.28 mmol) in water (12 mL) and acetonitrile (8mL) with a catalytic amount of $AlCl_3$ (0.02 mmol) and the reaction was refluxed for 2 days. After cooling the resulting crude product (10b) was filtered and resuspended in water/methanol/dioxane (1:0.5:2, v/v/v), and the pH of the solution was brought to 13 by adding 1M sodium hydroxide. The solution was stirred at 50 °C for 2 hours and the final compound was purified using RP-HPLC to provide the title compound as a white solid (15 mg, 20%). **Mp** 247-253 ^{o}C (decomp.), **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.99 (s, 1H), 6.62 (s, 1H), 4.87 (s, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 168.6, 165.0, 154.2, 151.4, 149.2, 105.0, 45.4. **LRMS** (ESI): m/z: 242 [M+H]$^+$, (100%), **HRMS** (ESI): observed m/z: 242.0341 [M+H]$^+$; calculated m/z: 242.0342 [M+H]$^+$, **RP-HPLC**: t_R 4.23, >95% (gradient).

4.8.1.24 7-(2'-Aminoethyl)-8-mercaptoguanine (11c)

A suspension of *N*-2-(8-bromo-*N²*-acetylguanin-7-yl)ethylphthalimide (**9c**) (160 mg, 0.36 mmol) in water (12 mL) and acetonitrile (8 mL) was mixed with $Na_2S_2O_3.5H_2O$ (447 mg, 1.80 mmol) and catalytic amount of $AlCl_3$ (0.02 mmol). The reaction was refluxed for 2 days. After cooling the mixture was concentrated to dryness under reduced pressure. The crude product, *N²*-acetyl-7-(2-(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl)-8-mercaptoguanine **10c** (50 mg, 0.13 mmol) was resuspended in methanol (1 mL) and hydrazine hydrate (12 µL, 0.36 mmol) was added. The solution was stirred for 15 hours and purified by reverse phase chromatography (C18, 0.1% TFA in water) to provide the title compound as a white solid (10 mg, 32%). **Mp** 280 – 287 $^{\circ}$C (decomp.), **¹H NMR** (400 MHz, DMSO-d_6) δ 8.22 (s, 1H), 7.85 (s, 2H), 6.78 (s, 2H), 4.41 (t, *J* = 6.1 Hz, 1H), 3.16 (t, *J* = 6.1 Hz, 1H).**¹³C NMR** (101 MHz, DMSO-d_6) δ 164.6, 154.5, 151.7, 150.1, 105.1, 42.5, 38.5. **LRMS** (ESI): *m/z*: 227.1 $[M+H]^+$, (100%), **HRMS** (ESI): observed *m/z:* 225.0564 $[M-H]^-$; calculated *m/z:* 225.0564 $[M-H]^-$, **RP-HPLC**: t_R 2.9, >98% (gradient).

4.8.1.25 7-(3'-Aminopropyl)-8-mercaptoguanine (11d)

A suspension of N-3- (*N²*-Acetylguanin-7-yl)propylphthalimide (**9d**) (150 mg, 0.33 mmol) in water (12 mL) and acetonitrile (8 mL) was mixed with $Na_2S_2O_3.5H_2O$ (400 mg, 1.60 mmol) and catalytic amount of $AlCl_3$ (0.02 mmol). The reaction was refluxed for 2 days. After cooling the mixture was concentrated to dryness under reduced pressure. The crude product, *N²*-acetyl-7-(3-(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)propyl)-8-mercaptoguanine **10d** (50 mg, 0.12 mmol) was resuspended in methanol (1 mL) and hydrazine hydrate (12 µL, 0.4 mmol, 3 eq) was added. The reaction was stirred overnight at room temperature (20 °C) and the final compound was purified by reverse phase chromatography (C18, 0.1% TFA in water) to provide the title compound as a white solid (12 mg, 44%). **Mp** 239 – 243 $^{\circ}$C (decomp.), **¹H NMR** (400 MHz, D_2O) δ 4.41 (t, *J* = 6.6 Hz, 2H), 3.08 (t, *J* = 6.6 Hz, 2H), 2.22 (t, *J* = 6.6 Hz, 2H). **¹³C NMR** (101 MHz, D_2O) δ 163.0, 154.4, 153.0, 150.1, 106.0, 42.0, 36.3, 26.5. **LRMS** (ESI): *m/z*: 241.1 $[M+H]^+$, (100%), **HRMS**

(ESI): observed *m/z:* 241.0942 [M+H]$^+$; calculated *m/z:* 241.0827 [M+H]$^+$, **RP-HPLC**: t_R 3.76, >98% (gradient).

4.8.1.26 7-(2-Guanidinoethyl)-8-mercaptoguanine (12)

A mixture of 7-(2'-aminoethyl) 8-mercaptoguanine (**11c**) (10.0 mg, 0.04 mmol) and pyrazole carboxamidine (7.0 mg, 0.1 mmol) in DMF was stirred at 50 °C for 2 days. The resulting mixture was concentrated to dryness under reduced pressure. The crude product was resuspended in water and purified using reverse phase chromatography (C18, 0.1% TFA in water) to provide the title compound as a white solid (5 mg, 45%). **Mp** 256 – 262 °C (decomp.), **^{1}H NMR** (400 MHz, DMSO-d$_6$) δ 11.18 (s, 1H), 7.64 (t, *J* = 6.2 Hz, 1H), 6.75 (s, 2H), 4.29 (t, *J* = 6.4 Hz, 2H), 3.47 (t, *J* = 6.4 Hz, 2H). **^{13}C NMR** (101 MHz, DMSO-d$_6$) δ 164.4, 156.9, 154.3, 151.5, 149.9, 104.8, 45.7, 42.8. **LRMS** (ESI): *m/z:* 269.1 [M+H]$^+$, (100%), **HRMS** (ESI): observed *m/z:* 269.0936; calculated *m/z:* 269.0928 [M+H]$^+$, **RP-HPLC**: t_R 4.23, >98% (gradient).

4.8.1.27 8-(Methylamino)guanine (18)

A solution of 8-(methylamino)guanosine (50 mg, 0.2 mmol) in 1 M HCl (10 mL) was refluxed for 2 hours and cooled to room temperature. The precipitated was collected by filtration and resuspended in water (5 mL). This mixture was made basic by drop wise addition of 1 M NaOH whereupon the precipitate dissolved. This solution was purified by reverse phase chromatography (C18, 1% TFA in water) to provide the title compound as a white solid (30 mg, quantitative). **Mp** 252 – 257 °C (decomp.), **^{1}H NMR** (400 MHz, D$_2$O) δ 2.66 (s, 3H). **^{13}C NMR** (101 MHz, D$_2$O) δ 164.2, 163.6, 162.5, 157.5, 116.2, 30.0.**LRMS** (ESI): *m/z:* 181.1 (M+H, 100%). **HRMS** (ESI): observed *m/z:* 181.0837; calculated *m/z:* 181.0832 [M+H]$^+$.

4.8.2 SPR

All SPR binding experiments were performed as described in chapter 2 and 3. The only modification was the use of site-directed biotinylation using a sulfhydryl reactive maleimide-activated biotin derivative, 1-biotinamido-4-(4'-[maleimidoethylcyclohexane]-carboxamido)butane. The maleimide-activated biotin was attached to exposed surface cysteine residues of SaHPPK and subsequently immobilised on the Biocap sensor chip (GE healthcare). All analogues were serially diluted (either 2- or 3-fold) in SPR binding buffer and injected for 30 seconds contact time at 60 µL/min and then allowed to dissociate for 60 seconds. Binding sensorgrams were processed using the Scrubber software (version 2, BioLogic Software, Australia). To determine the binding affinity (K_d), responses at equilibrium for each compound were fit to a 1:1 steady state affinity model available within Scrubber software.

4.8.3 ITC

Experiments were performed using an iTC200 instrument (MicroCal) at 298 K, with the ligands titrated into solutions of SaHPPK using 18 x 2.2 µL injections. Data were fitted using Origin software to yield the thermodynamic parameters, ΔH, K_d and N (the binding stoichiometry), assuming a cell volume of 0.2 mL. These were then used to calculate the Gibb's free energy of binding, ΔG (-RT.lnK_a), and entropy of binding, ΔS (using $\Delta G = \Delta H$ - TΔS). SaHPPK was dialysed overnight into 50 mM HEPES, 1 mM TCEP, 10 mM MgCl$_2$, pH = 8.0 buffer before running the experiment. For titrations with compounds **11a**, **11c-d** and **12**, SaHPPK was typically at 30 µM and the ligands were at 1-1.5 mM.

4.8.4 X-ray

Crystallisation experiments were performed as described in chapter 2. Briefly, co-crystallisation was set-up in the JCSG+ Suite commercial crystal screens (Qiagen) at 281 K using sitting-drop vapour-diffusion method with droplets consisting of 150 nL protein solution and 150 nL reservoir solution and a

reservoir volume of 50 µL. Crystals of the *Sa*HPPK in complex with 7-(2'hydroxyethyl)guanine (**11a**) were observed in conditions containing 240 mM sodium malonate and 20 % polyethylene glycol 3350.

The *Sa*HPPK structure (3QBC) was used to solve the initial phases of the binary complex by molecular replacement using Phaser[8]. Refinement was performed using *REFMAC*5[9] and the Fourier maps (2F_O-F_C *and* F_O-F_C) were visualised in *Coot*[10]. After several rounds of manual rebuilding, 7-(2'hydroxyethyl)guanine (**11a**) and water molecules were added and the model further refined to a resolution of 1.85 Å (R_{free} (%) = 26.4, R_{work} (%)= 20.8).

4.9 References

1. Long, R. A.; Robins, R. K.; Townsend, L. B. Purine nucleosides. XV. The synthesis of 8-aminoand 8-substituted aminopurine nucleosides. *J Org Chem* **1967,** 32, 2751-6.

2. Janssen, P.; Lewi, P.; de Jonge, M.; Koymans, L.; Daeyaert, F.; Heeres, J.; Vinkers, H. HIV replication inhibiting purine derivatives. *WO Patent 2005/028479* **2005**.

3. Sircar, J. C.; Brungardt, C. R.; Schwender, C. F. *Eur. Pat. Appl. EP 156559 A2* **1985**.

4. Michael, M. A.; Cottam, H. B.; Smee, D. F.; Robins, R. K.; Kini, G. D. Alkylpurines as immunopotentiating agents. Synthesis and antiviral activity of certain alkylguanines. *J Med Chem* **1993,** 36, 3431-6.

5. Srivastava, P. C.; Nagpal, K. Bromination of nucleosides. *Experientia* **1970,** 26, 220.

6. Baker, T. J.; Luedtke, N. W.; Tor, Y.; Goodman, M. Synthesis and anti-HIV activity of guanidinoglycosides. *J Org Chem* **2000,** 65, 9054-8.

7. Ikaunieks, M.; Madre, M. Reinvestigation of the reaction of 8-bromoguanine derivatives with sodium thiosulfate. *J. Chem. Res.* (S), **2002**, 33, 226–227.

8. Storoni, L. C.; McCoy, A. J.; Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* **2004,** 60, 432-8.

9. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **1997,** 53, 240-255.

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

Chapter 5

Paramagnetic NMR

5. Paramagnetic NMR

5.1 Introduction

Previous work (chapters 3 and 4) established that 8-MG inhibited *Sa*HPPK function *in vitro* and binds competitively to the substrate (HMDP) site. Structural details of binding were investigated by X-ray crystallography and NMR spectroscopy. NMR measurements showed that the extended loop L3 conformation in the X-ray *Sa*HPPK/8-MG binary structure was likely a poor representation for the loop L3 conformation in the *Sa*HPPK/8-MG/AMPCPP ternary complex, which unfortunately failed to crystallize. Moreover, several pieces of NMR evidence suggested that the loop L3 is probably in a ´closed´ conformation, akin to that observed in the HMDP/AMPCPP/*Ec*HPPK ternary complex (1Q0N)[1].

If a full structure determination of *Sa*HPPK/8-MG/AMPCPP were to be undertaken by NMR, it would not reveal direct ^{1}H-^{1}H interactions as delineated by the NOEs between 8-MG and the enzyme as there are no non-exchangeable protons on 8-MG. While in principle observation of ^{1}H-^{1}H interactions between the side chain residues in loop L3 and loop L2 would show that loop L3 is closed over the active site, the deficit of NOEs, combined with the difficulty in side chain assignment, (due to several broadened resonances from loop L3), would render this approach arduous, inefficient and inaccurate.

We hypothesised that the closed position of loop L3 is due to specific interactions involving the sulfur atom in 8-MG. Herein, the loop position is investigated further using a modern paramagnetic NMR approach involving the analysis of the pseudo-contact shifts (PCSs) induced by attachment of lanthanide metal ions caged within a lanthanide binding tag (LBT).

114

5.2 Paramagnetic NMR in structural biology and the pseudo contact shifts (PCSs)

Structural and dynamic studies of proteins and protein-ligand complexes by NMR spectroscopy can be enhanced by the site-specific attachment of lanthanide ions (Ln^{3+})[2, 3]. The paramagnetic effects produced by a rigidly bound lanthanide ion manifest in a range of useful NMR parameters, including pseudo-contact shifts (PCSs), paramagnetic relaxation enhancements (PREs) and residual dipolar couplings (RDCs)[2, 4]. Among these paramagnetic effects, PCSs are particularly useful and easy to measure as they manifest themselves in, for example, well resolved ^{15}N-HSQC spectra as altered chemical shifts.

A PCS corresponds to the difference in the observed chemical shift of a nucleus, $\Delta\delta^{pcs}$, in the paramagnetic spectrum compared to that in the reference diamagnetic spectrum (using La^{3+}, Y^{3+} or Lu^{3+}-loaded LBT). The observed PCS ($\Delta\delta^{PCS}$) (Eq. 5.1) of a spin is given by[5]:

$$\Delta\delta^{PCS} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos2\varphi\right] ....(5.1)$$

where the coordinates of the nucleus with respect to the principal axis of the magnetic susceptibility tensor ($\Delta\chi$) are defined by r (the distance from the *e.g.* HN nucleus to the unpaired electron in the metal), θ and ϕ (the polar angles describing the metal-nucleus vector within the $\Delta\chi$ frame, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic components of the anisotropic magnetic susceptibility tensor. The orientation of the $\Delta\chi$ frame with respect to the protein frame is defined by three Euler rotations (α, β and γ) in the ZYZ convention[6]. The PCSs depend on the distance between the nucleus of interest and the electron of the paramagnetic metal ion as $1/r^3$.

The $\Delta\chi$ tensor can be conveniently represented by an isosurface that pictorially shows the surface in 3D space associated with a given PCS value (Figure 5.1). Depending on the paramagnetic strength of the lanthanide ion, the nuclear spins close to the metal ion (<15 Å) experience substantial broadening due to PRE and are not observed in regular ^{15}N HSQC experiments. Heteronuclear ZZ exchange methods can be used to probe

nuclei within this region[7, 8] or the use of a weakly paramagnetic lanthanide, e.g. Ce^{3+} with less PRE broadening.

Figure 5.1: Isosurface representation of the $\Delta\chi$ tensor showing the position of the bound lanthanide (green) in a LBT ligated by a disulfide bond to the A28/C28 mutant of human Ubiquitin. Positive (red lobes) and negative (blue lobes) PCS isosurfaces trace the surface of space associated with PCS values with magnitudes of 0.2 and 1.5 ppm for a $\Delta\chi_{ax} = 32\times10^{-32}m^3$ (a typical value for a rigidly bound Dy^{3+} LBT). The invisible broadened PRE zone is shown schematically as a yellow sphere within which amides are not detected in regular ^{15}N HSQC spectra for Dy^{3+} LBTs.
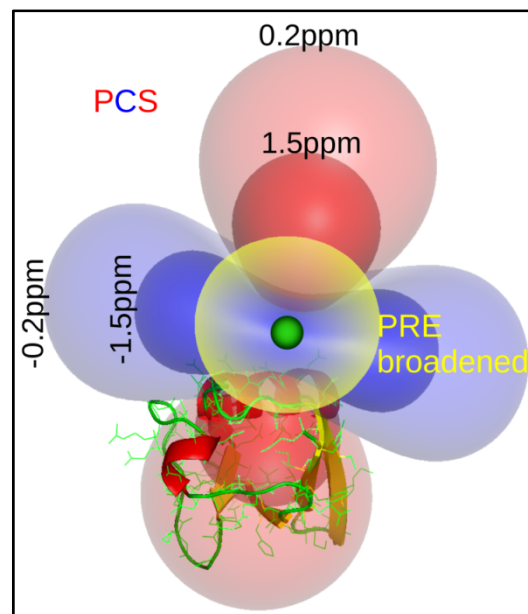
The metal coordinate and magnitude of $\Delta\chi$ can be determined from a known structure and the measurement of at least 8 PCSs. Thereafter, any PCS can be converted into spatial information. The advantage of PCS approaches to structural problems is that a simple chemical shift can be accurately, precisely and easily measured, and without additional assignments, compared to other parameters (NOE intensities, or small J couplings). NOEs, the principal parameter of NMR structure determination, are limited to interactions <5 Å, whilst PCSs can be observed over a long range (up to 50 Å) for strongly paramagnetic ions. This is due to an r^{-3} dependency (rather than r^{-6}, as for NOEs) (Eq. 5.1). Similarly, RDCs such as $^1D_{HN}$ report the H-N vector orientation with respect to the alignment tensor, and while they are not limited by distance, they lose the useful distance information as the H-N distance is fixed.

A deciding factor (and often the bottleneck for practical PCS analysis) is that the LBT has to be held relatively rigid with respect to the protein surface,

otherwise movement of the tag will move the negative and positive PCS lobes of the isosurface over the nucleus leading to a diminished observable PCS for that nucleus.

The lanthanide series of metals are uniquely suited for PCS measurements, having a range of inherent paramagnetic strengths and associated $\Delta\chi$ values. For example, dysprosium (Dy^{3+}) will induce large PCSs (large $\Delta\chi$) and cerium (Ce^{3+}) very small PCSs (small $\Delta\chi$). This is useful for assignments of the paramagnetic spectra. As long as the metal is held in the same geometry and the tag in the same average position relative to the protein, then the PCS for any given nucleus for different lanthanides will all lie on a straight line in the 2D NMR spectra[9]. Therefore, this provides a route to facilitate the assignment of the paramagnetic spectra, which otherwise is generally very difficult with a single metal.

5.3 Most proteins do not bind lanthanides natively – the need for LBTs

While some proteins can bind lanthanide metals (e.g. some metalloproteins)[10], these are few and far between. In order to take advantage of the induced paramagnetic effects of a bound lanthanide ion, the metal ion has to be coordinated to a chemical pendant or LBT and covalently attached to the protein (most often via a disulfide bond to a solvent exposed free cysteine residue). Depending on the chemical structure and the number of coordinating sites (up to 9 for a lanthanide) metal binding can be relatively weak (3-4 coordination, K_d ~low μM, giving rise to labile ions in solution), or extremely tight (8-9 coordination, K_d ~pM).

Recently, two new LBTs have been designed and developed in house based on iminodiacetic acid (IDA)[8] and nitrilotriacetic acid (NTA)[11]. These produced extraordinarily large PCSs in ^{15}N-HSQC spectra of ubiquitin, but were unsuitable for use with SaHPPK owing to the large excess of metal (Mg^{2+}) required for ATP or AMPCPP binding, which competed with the relatively labile lanthanide ions. Therefore, a new cyclen-based LBT, C1 (Figure 5.2),

which fully encapsulates a lanthanide ion, protecting it from exchange with the large excess of Mg^{2+}, was developed in collaboration with Dr. Bim Graham, Prof. G. Otting at ANU and Prof. G. Pintacuda at CERN[9]. By virtue of three chiral arms and large phenyl groups within its structure, the tag assumes one conformation, so that only a single set of signals is observed in the NMR spectra, corresponding to a single LBT-protein diastereomer. This tag is activated with a thiopyridine leaving group, and owing to the extremely tight binding of lanthanides, was specifically designed for structural investigations of kinases in the presence of a large excess of magnesium ions. It has been tested with three proteins so far, namely, Ubiquitin A28C, ArgN[9] and DEN NS2B-NS3pro[12]. With a range of LBTs being developed to date, PCS methods are beginning to emerge as a powerful structural technique to rapidly determine protein-ligand[13] and protein-protein[10, 14] interactions, and to refine protein structures[15].

5.4 Using the PCS data to determine the $\Delta\chi$ tensor for structure analysis

At least three different Cyclen-Ln^{3+} LBTs to the same cysteine and ^{15}N HSQC spectra are recorded for each of two paramagnetic Cyclen-Ln^{3+} tagged protein samples and the shifts compared with those in the Cyclen-Y^{3+} tagged protein spectrum as a suitable diamagnetic reference tag. PCSs are measured in the higher resolution ^{1}HN dimension rather than in the ^{15}N dimension by subtracting the chemical shift (δ) of the diamagnetic amide from that of the paramagnetic amide. Several PCS are then fitted to the structure of the protein (e.g. from X-ray or NMR) to derive the magnitude and orientation (Euler angles α, β an γ) of the $\Delta\chi$ tensor and the metal coordinate within the software program Numbat[6].

For example, using these data, a measured PCS from several amides in a loop, not used to derive the $\Delta\chi$ tensor can be compared to those calculated for any proposed loop model to validate the structure of that loop.

5.5 Suitability of *Sa*HPPK for labeling with a LBT

Addition of DTNB ("Ellman's reagent") to *Sa*HPPK produced a bright yellow colour, indicating the presence of at least one free thiol suitable for ligation. A ^{15}N HSQC spectrum (Figure 5.3) showed localised CSPs consistent with only one reactive cysteine (Cys80), thereby indicating that it was likely that only one LBT would ligate to *Sa*HPPK. This finding is consistent with the X-ray structure, which indicates that out of the four cysteines (Cys58, 73, 73, 80), only Cys80 is solvent exposed.
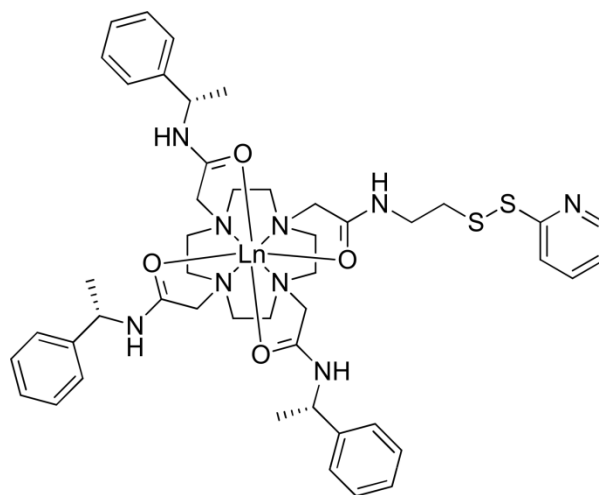
Figure 5.2: The cyclen-based octadentate C1 LBT used in the present study. The tag features three methylbenzylacetamide pendants with *S* stereochemistry and one acetamide pendant with a pyridine dithiol group to facilitate coupling to an exposed cysteine residue. It is likely that a water molecule also coordinates the bound lanthanide ion.

119

5.6 Tagging of *Sa*HPPK with the C1 LBT.

A uniformly ^{15}N-labelled sample of *Sa*HPPK was expressed and purified as described in chapter 2. The *Sa*HPPK was reduced by the addition of 10 mM DTT for 15 minutes, with excess DTT subsequently removed using a PD 10 column (GE Healthcare) equilibrated with freshly degassed NMR buffer (50 mM HEPES, pH 8.0). A 5-fold excess of the C1 tag (Figure 5.3)[9] was added to a ~80 μM sample of ^{15}N labelled *Sa*HPPK and mixed for ~15 minutes at room temperature with a pipette. A small amount of protein precipitate produced during the course of mixing was removed by centrifugation. Excess tag was then removed by passage through a PD 10 column and the sample concentrated using a Millipore ultrafiltration device to a final concentration of 80 μM. The tagging reaction yielded 80-90% C1-tagged *Sa*HPPK as judged by NMR. All spectra were recorded on the C1-tagged *Sa*HPPK/8-MG/AMPCPP ternary complex.
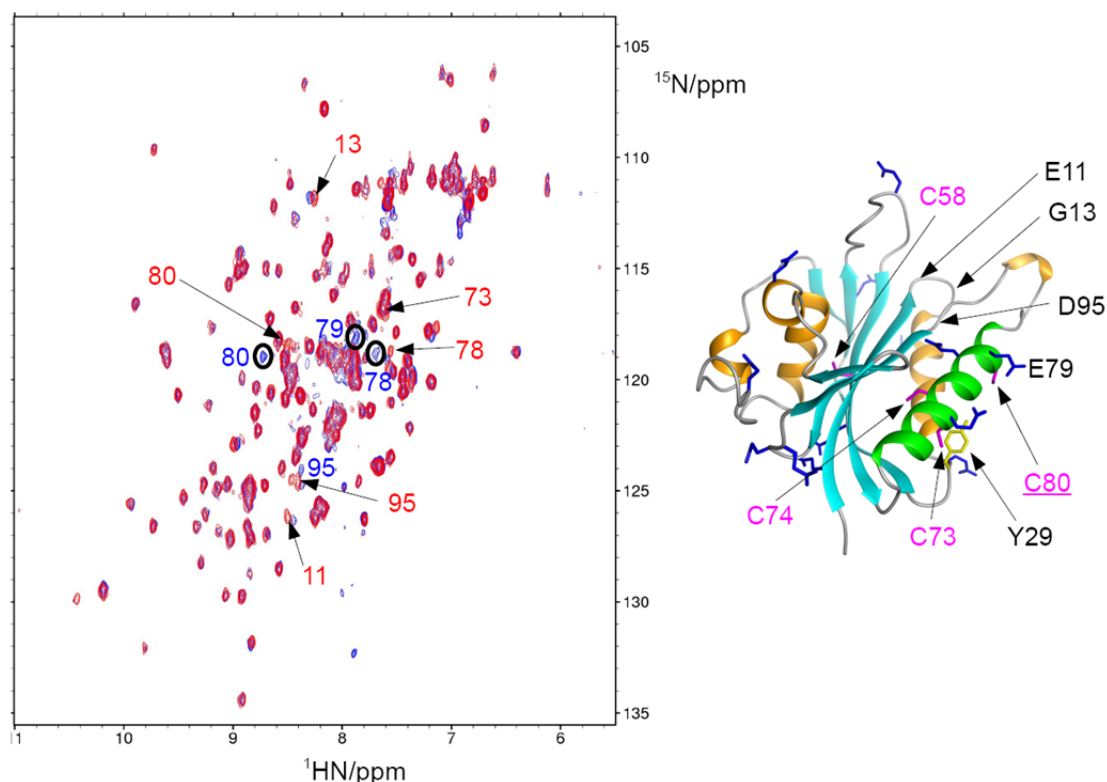
Figure 5.3: ^{15}N sofast HSQC spectra of ~80 μM apo *Sa*HPPK (red) and *Sa*HPPK after reaction with 300 μM DTNB for 3 h (blue). The structure of *Sa*HPPK is shown on the right, with the four cysteines indicated. Some tentative assignments are shown for the *Sa*HPPK/DTNB spectrum (blue).

5.7 Assignment of the PCS data and calculation of the $\Delta\chi$ tensor

The assignment of the Tm^{3+}- and Dy^{3+}-loaded C1 LBT Cys80-ligated *Sa*HPPK/8-MG/AMPCPP ^{15}N HSQC spectra was facilitated by an overlay of the spectra with that of the Y^{3+}-loaded form (serving as a diamagnetic reference) (Figure 5.4). The assignments of the three well resolved amides (Gly90, Gly7 and Gly9) illustrate that the PCSs in the Dy^{3+} spectrum are approximately 1.5 times that in the Tm^{3+} spectrum. Several PCSs from well resolved amides were used to calculate the approximate magnitude and orientation of the $\Delta\chi$ tensor from the Dy^{3+}-loaded C1 LBT bound to the Cys80 of *Sa*HPPK (pdb 3QBC). An iterative procedure was used to manually include more and more PCS values into the calculation until all PCS data from the assigned amides were included in the least-squares optimisation of the $\Delta\chi$ tensor using Numbat[6]. The same procedure was used for the *Sa*HPPK/8-MG/AMPCPP ^{15}N HSQC spectrum from the Tm^{3+}-loaded LBT. Finally, all 8 parameters (the metal x, y, and z coordinate, the magnitude of the axial and rhombic components of each $\Delta\chi$ tensor for Tm^{3+} and Dy^{3+}, and the three Euler angles, α, β and γ) were simultaneously optimised with respect to both the Dy^{3+} and Tm^{3+} PCS data. The PCS values from the several observed amide resonances in loop L3 were not included in the calculation. The final magnitude, orientation and optimised metal positions are shown in Table 5.1. Excellent convergence of the metal position and the orientation of the two $\Delta\chi$ tensors (as judged by very similar Euler angles) was observed for the Dy^{3+}- and Tm^{3+}-loaded C1 LBTs.

**Table 5.1:** Tensor parameters (in UTR convention) for the paramagnetic C1 LBTs bound to Cys80 in *Sa*HPPK.

	Dy^{3+}	Tm^{3+}
$\Delta\chi_{Ax}$ (x10^{-32}m^3)	22.5 ± 0.8	-14.8 ± 0.4
$\Delta\chi_{Rh}$ (x10^{-32}m^3)	9.5 ± 0.3	-8.1 ± 0.5
α (°)	115.3	116.7
β (°)	116.7	117.8
γ (°)	27.9	33.2
Metal coord (Å)	17.0 ± 0.1, 2.4 ± 0.1, -10.0 ± 0.2	
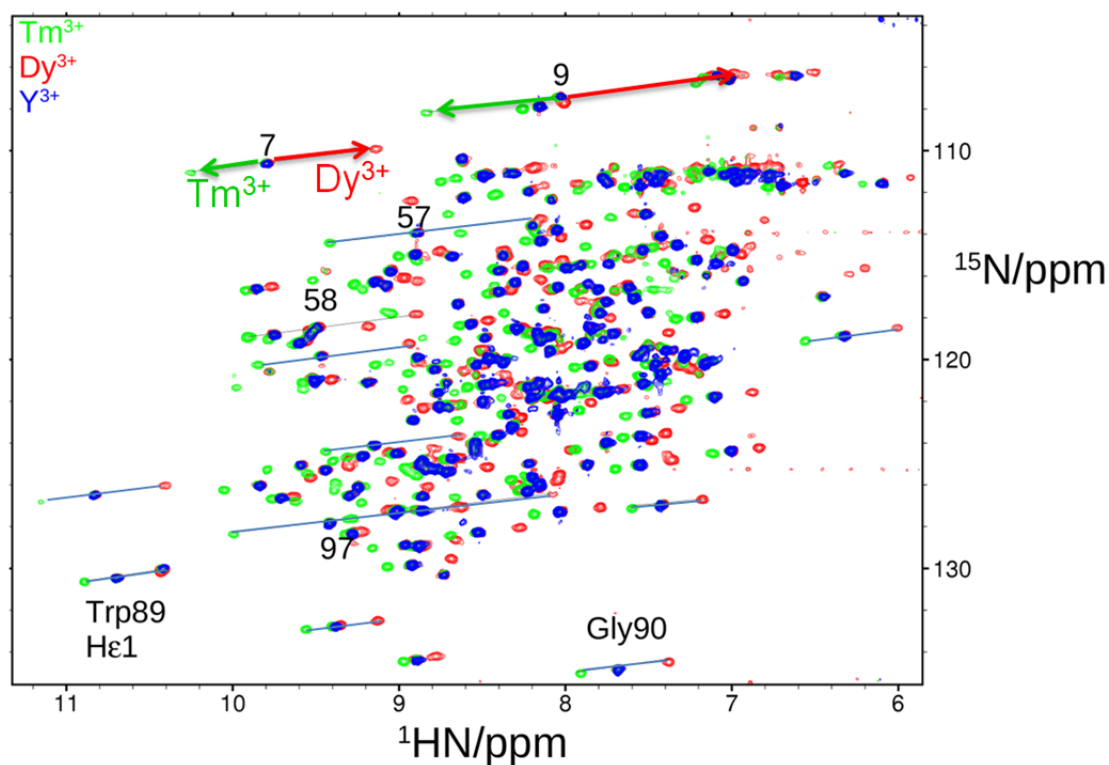
Figure 5.4: Superposition of ^{15}N HSQC spectra of ~80 µM lanthanide ion-loaded, C1-tagged ^{15}N *Sa*HPPK in the presence of 10 mM MgCl$_2$, 1 mM AMPCPP and 500 µM 8-MG dissolved in 50 mM HEPES pH 8.0 10% D$_2$O 90% H$_2$O.

5.8 An excellent fit between the measured and calculated PCS data

Using the X-ray structure of *Sa*HPPK/8-MG (pdb 3QBC), a striking agreement is observed between the measured and calculated PCS values for the Dy^{3+} and Tm^{3+} C1-tagged *Sa*HPPK. This shows that the *Sa*HPPK/8-MG binary X-ray structure is a very good representation of the *Sa*HPPK/8-MG/AMPCPP ternary structure for all amides used in the calculation of *Sa*HPPK (except those in loop L3, not included in the calculation). It is notable that PCSs from amides within the substrate loop L2 show an excellent fit, and therefore the position of L2 has not changed in the *Sa*HPPK/8-MG/AMPCPP ternary structure compared to that observed in the X-ray structure. The L3 loop position can be scrutinised by examination of the three PCS from Gly90 HN, Trp89 Hε1 and Val83 HN. For the solvent exposed conformation of the X-ray structure, the correlation between observed and calculated PCSs for these resonances is very poor for both Dy^{3+}- and Tm^{3+}-loaded C1-tagged *Sa*HPPK (magenta in Figure 5.5). When the L3 loop position was modeled on that observed in the 'closed' *Ec*HPPK/HMDP/AMPCPP ternary structure (1Q0N), the correlation coefficient of the PCSs for these amides (blue in Figure 5.5) was very good, providing strong evidence that the L3 loop is clamped down over the active site.
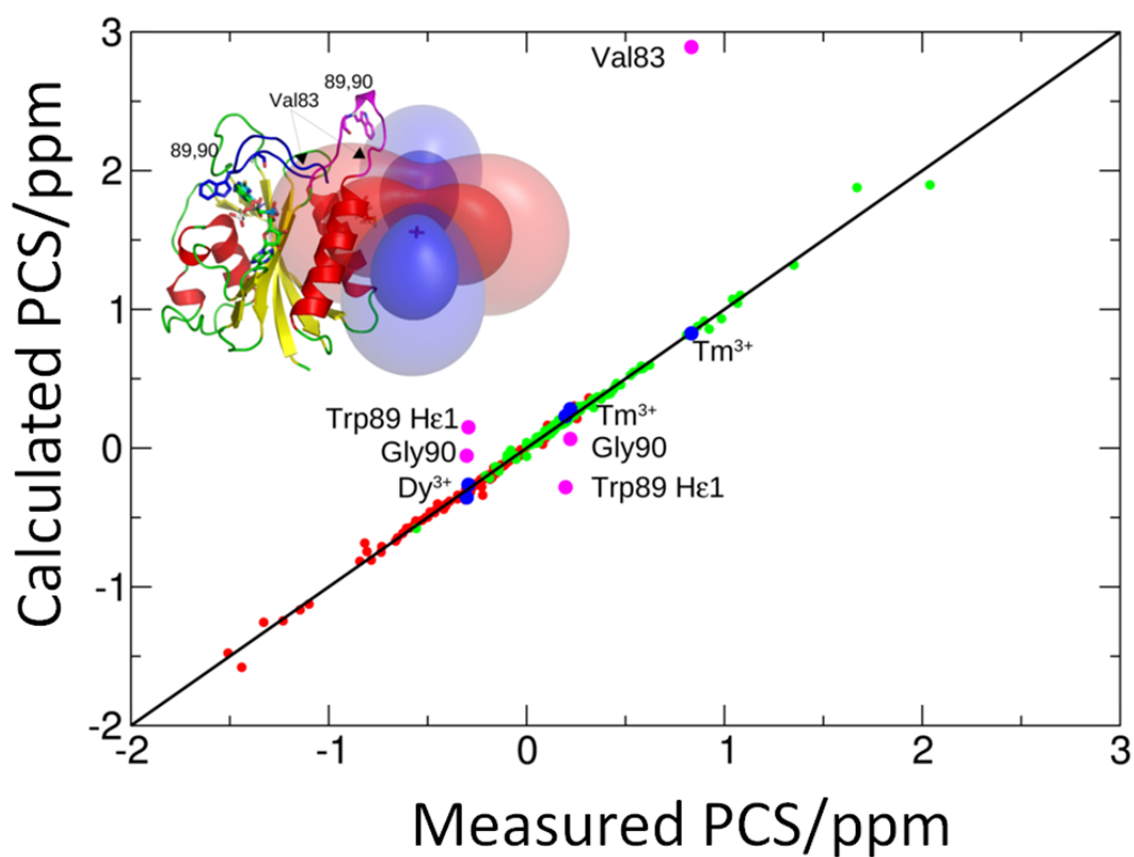
Figure 5.5: Correlation between measured and calculated PCS for *Sa*HPPK ligated to Tm^{3+} C1 (green) and Dy^{3+} C1 (red). Five PCS from Gly90 HN, Trp89 Hε1 and Val83 HN are correlated for the closed loop (blue, PDB 1Q0N) and the open loop (magenta, PDB 3QBC) conformations. Both loops are shown and the isosurface for the Dy^{3+} C1 tagged *Sa*HPPK contoured at 2.9 ppm (inner lobes) and 0.7 (outer lobes), where the red is negative and blue positive surfaces of PCS 3D space. The position of the lanthanide ion is shown by a small cross. The sidechain of Trp89 is also shown.

5.9 Conclusions

Initial PCS data shows that the L3 loop conformation in the *Sa*HPPK/8MG/AMPCPP ternary structure is very similar to the closed conformation found in the *Ec*HPPK/HMDP/AMPCPP complex (1Q0N)[1]. Further data has been acquired and analysed for the other enantiomer of the C1 tag, including the Tb^{3+} bound metal (by Dr James Swarbrick) and from the measurement of a total of 512 PCS data points is in agreement with these initial findings. When the closed loop conformation of *Ec*HPPK/HMDP/AMPCPP is modeled, the sulfur atom of 8-MG lies in close proximity to the carbonyl oxygen atoms of Gly90 and Trp89 (3.7 and 3.0 Å, respectively). If protonated, the SH--OC (Gly90) distance would be 2.3 Å and the geometry suitable for hydrogen bonding. The PCS model is therefore in part suggestive that the role of sulfur may be to facilitate a hydrogen bond interaction between the SH to the carbonyl of Gly90 (Figure 5.6). While the distance to the Trp89 carbonyl is shorter, the geometry is not suitable for hydrogen bond formation, as the sulfur lies underneath and behind the carbonyl oxygen of Trp89. This geometry may instead support a S•••C(π) interaction to the C of the carbonyl, as reported in metal thiolate complexes[16]. It may also indicate the presence of a hyperconjugative σC(2)-S•••πCO interaction[16]. Tentative evidence for these interactions is provided by the fact that the shift for the ^{15}N of Gly90 is unusually large (~7 ppm) in the *Sa*HPPK/8-MG/AMPCPP complex compared to that in the *Sa*HPPK/AMPCPP complex.

In the future, tagging of the *Sa*HPPK protein with both the C2 and C1 LBTs could be used to help locate the position and determine the bound structure of fragments from our screening program that are otherwise difficult to determine from standard chemical shift mapping (chapter 2), in a manner similar to that recently demonstrate by Saio *et al*[17].
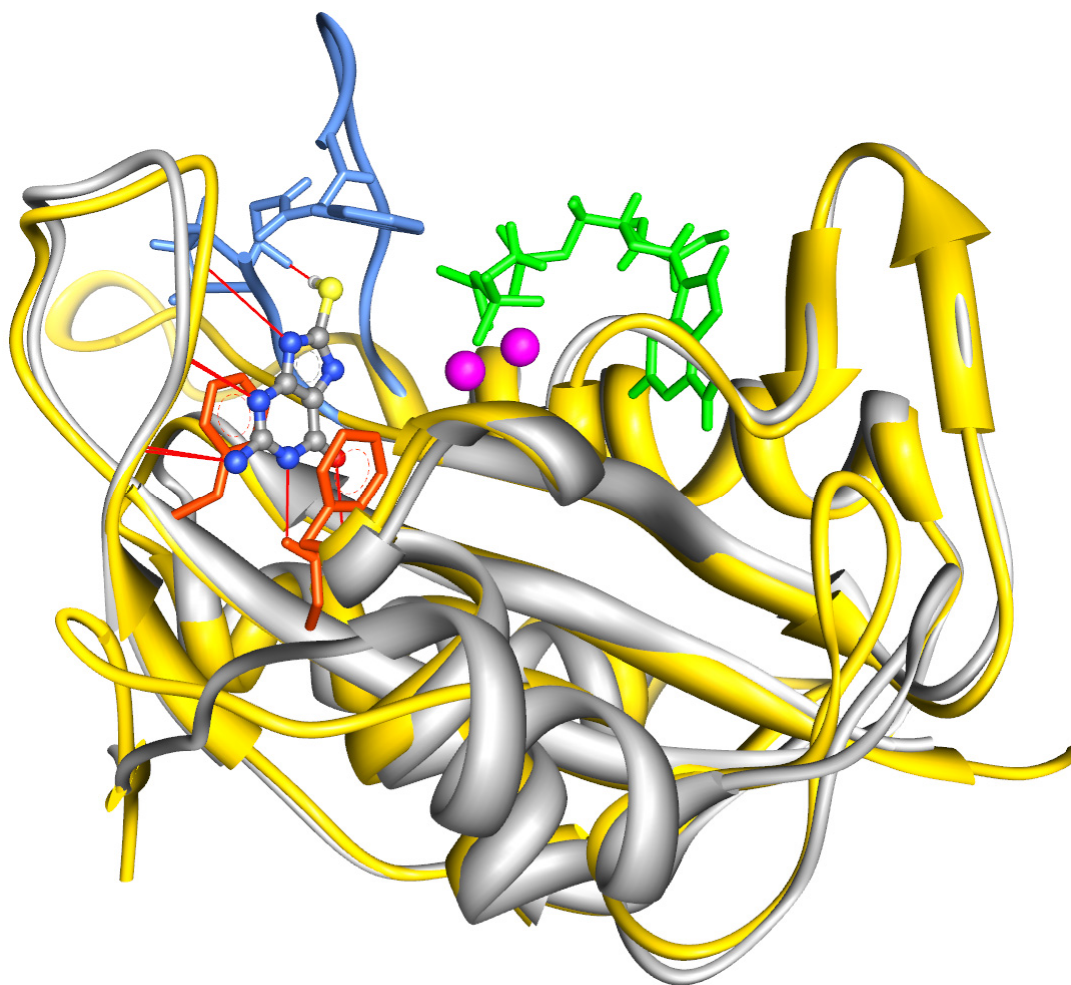
Figure 5.6: PCS verified model of *Sa*HPPK/8-MG/AMPCPP showing the closed L3 loop conformation (blue) of *Ec*HPPK/AMDP/AMPCPP (grey, pdb 1Q0N) and the bound AMPCPP (green) and two magnesium ions (magenta). A hydrogen bond may exist between a SH of 8-MG and the carbonyl of Gly90 as shown.

5.10 References

1. Blaszczyk, J.; Shi, G.; Yan, H.; Ji, X. Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 A resolution. *Structure* **2000,** 8, 1049-58.

2. Otting, G. Protein NMR using paramagnetic ions. *Annu Rev Biophys* **2010,** 39, 387-405.

3. Su, X. C.; Huber, T.; Dixon, N. E.; Otting, G. Site-specific labelling of proteins with a rigid lanthanide-binding tag. *Chembiochem* **2006,** 7, 1599-604.

4. Keizers, P. H.; Ubbink, M. Paramagnetic tagging for protein structure and dynamics analysis. *Prog Nucl Magn Reson Spectrosc* **2011,** 58, 88-96.

5. Bertini, I.; Luchinat, C.; Parigi, G. Magnetic susceptibility in paramagnetic NMR *Prog Nucl Magn Reson Spectros.* **2002,** 40, 249– 273.

6. Schmitz, C.; Stanton-Cook, M. J.; Su, X. C.; Otting, G.; Huber, T. Numbat: an interactive software tool for fitting Deltachi-tensors to molecular coordinates using pseudocontact shifts. *J Biomol NMR* **2008,** 41, 179-89.

7. John, M.; Park, A. Y.; Dixon, N. E.; Otting, G. NMR detection of protein 15N spins near paramagnetic lanthanide ions. *J Am Chem Soc* **2007,** 129, 462-3.

8. Swarbrick, J. D.; Ung, P.; Chhabra, S.; Graham, B. An iminodiacetic acid based lanthanide binding tag for paramagnetic exchange NMR spectroscopy. *Angew Chem Int Ed Engl* **2011,** 50, 4403-6.

9. Graham, B.; Loh, C. T.; Swarbrick, J. D.; Ung, P.; Shin, J.; Yagi, H.; Jia, X.; Chhabra, S.; Barlow, N.; Pintacuda, G.; Huber, T.; Otting, G. DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconjug Chem* **2011,** 22, 2118-25.

10. Pintacuda, G.; Park, A. Y.; Keniry, M. A.; Dixon, N. E.; Otting, G. Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein-protein complexes. *J Am Chem Soc* **2006,** 128, 3696-702.

11. Swarbrick, J. D.; Ung, P.; Su, X. C.; Maleckis, A.; Chhabra, S.; Huber, T.; Otting, G.; Graham, B. Engineering of a bis-chelator motif into a protein alpha-helix for rigid lanthanide binding and paramagnetic NMR spectroscopy. *Chem Commun (Camb)* **2011,** 47, 7368-70.

12. de la Cruz, L.; Nguyen, T. H.; Ozawa, K.; Shin, J.; Graham, B.; Huber, T.; Otting, G. Binding of low molecular weight inhibitors promotes large conformational changes in the dengue virus NS2B-NS3 protease: fold analysis by pseudocontact shifts. *J Am Chem Soc* **2011,** 133, 19205-15.

13. Zhuang, T.; Lee, H. S.; Imperiali, B.; Prestegard, J. H. Structure determination of a Galectin-3-carbohydrate complex using paramagnetism-based NMR constraints. *Protein Sci* **2008,** 17, 1220-31.

14. Keizers, P. H.; Mersinli, B.; Reinle, W.; Donauer, J.; Hiruma, Y.; Hannemann, F.; Overhand, M.; Bernhardt, R.; Ubbink, M. A solution model of the complex formed by adrenodoxin and adrenodoxin reductase determined by paramagnetic NMR spectroscopy. *Biochemistry* **2010,** 49, 6846-55.

15. Simon, B.; Madl, T.; Mackereth, C. D.; Nilges, M.; Sattler, M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew Chem Int Ed Engl* **2010,** 49, 1967-70.

16. Wladislaw , B.; Viertler , H.; Olivato , P. R.; Calegão , I. C. C.; Pardini, V. L.; Rittner, R. Interaction between the carbonyl group and a sulphur atom. Part 9. The relationship between conformation and ground- and excited-state interactions in some α-sulphur-substituted cycloalkanones. *J Chem Soc, Perkin Trans. 2* **1980**, 453-456.

17. Saio, T.; Ogura, K.; Shimizu, K.; Yokochi, M.; Burke, T. R., Jr.; Inagaki, F. An NMR strategy for fragment-based ligand screening utilizing a paramagnetic lanthanide probe. *J Biomol NMR* **2011,** 51, 395-408.

Chapter 6

Summary and Future Directions

6. Summary and Future Directions

In the present study we have thoroughly characterised for the first time the *Sa*HPPK enzyme and its interaction with a range of ligands. Our goal in part was to address the deficit in inhibitor studies of HPPK. With the discovery of novel small molecule binders using a combination of both ROCS *in silico* and fragment screening approaches we have clearly met our main goal. We have further contributed to the growing structural data on the HPPK family of enzymes in general by solving the first high resolution X-ray crystal structures of *Sa*HPPK and in complex with the novel substrate site inhibitor, 8-mercaptoguanine and an analogue. A range of biophysical methods such as TSA, NMR, SPR and ITC were used to characterise the structure and dynamics of ligand binding and to validate and rank the hits from our screening initiatives.
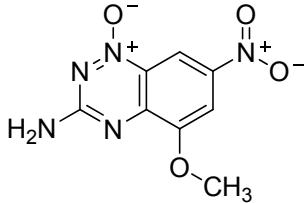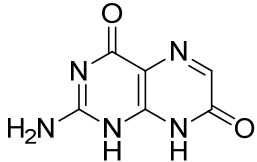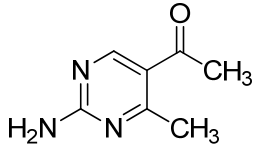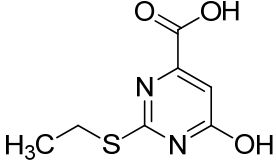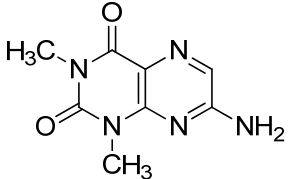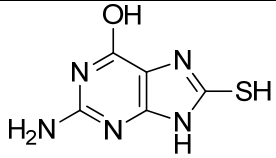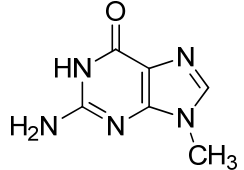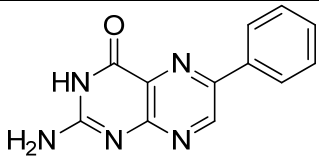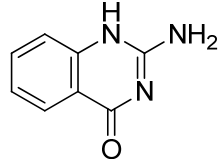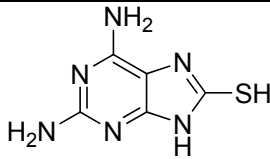
During our investigations, we were confronted with some substantial challenges to surmount, ranging from simple sample stability and longevity to a failure to crystallise the apo enzyme. In order to further our structural studies to crystallise the enzyme we had to find a non-hydrolysable pterin-site inhibitor (8-MG) that was non-competitive to the cofactor. While we solved the crystal structure of the *Sa*HPPK/8-MG binary complex we were unable to obtain crystals for the *Sa*HPPK/AMPCPP/8-MG ternary complex and moreover, the active site loop L3 was found curiously to be in an 'open' conformation in the crystal form. Therefore, to investigate the active site loop conformations, and in part to validate the crystal structure, we used NMR spectroscopy to study the ternary complex in solution. A detailed study of protein backbone dynamics was undertaken using ^{15}N heteronuclear NMR in different ligand bound complexes. We also took advantage of recent applications of modern techniques such as residual dipolar couplings (chapter 3) and novel paramagnetic NMR (chapter 5) to scrutinise the conformation of active site loops upon ligand binding events. The combination of chemical shift, ^{15}N heteronuclear NOE, RDC and paramagnetic PCS NMR experiments revealed that closed loops L3 and L2, and the rigidification of part of loop L3 and most of the loop L2 is involved in the mechanism of *Sa*HPPK inhibition by 8-mercaptoguanine. Furthermore, the synthesis of several analogues of 8-mercaptoguanine helped us to elucidate structure-activity-relationships around the 8-

MG scaffold. The analogues and the SAR revealed key 'hot spots' critical for binding and a route out from the highly specific pterin-site as the only way to improve the potency of future compounds, vital for lead expansion. Of the identified fragment hits some appeared to bind at the substrate and ATP active sites by simple chemical shift titrations. Most had good ligand efficiency and thus offer options for evolution to more potent compounds with more classic Lipinski like properties. A future area would be to use our new PCS NMR methods to find the binding site and bound conformation of these weak binding fragments.

In summary, this detailed structural and biophysical study has opened-up a range of possibilities for the rational design and detailed characterisation of new inhibitors of the *Sa*HPPK enzyme based on the 8-MG scaffold and from new fragment hits. Ongoing efforts to develop more effective multi-drug therapies to combat both TMP-SMZ sensitive and many other resistant strains of *S. aureus* could benefit from the current structural and SAR study of *Sa*HPPK.
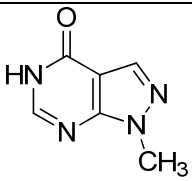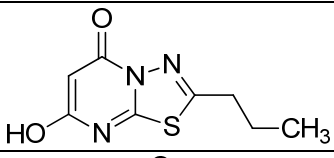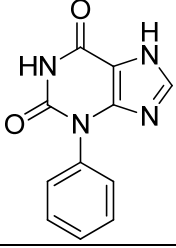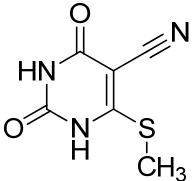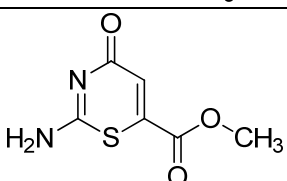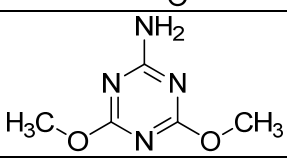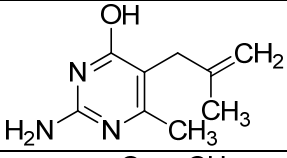
7. Appendix 1 – ROCS Library

No.	Compound Code	Chemical Name	Structure
1	AB-323/138874 14	2-amino-1-methyl-4(*1H*)-pteridinone	
2	RH 00268	4-ethyl-2-mercapto-6-oxo-1,6-dihydropyrimidine-5-carbonitrile	
3	SB 01255	5-nitropyrimidine-4,6-diol	
4	AF-399/369800 49	5,6-dimethyl-2-sulfanylthieno[2,3-d]pyrimidin-4(*3H*)-one	
5	AC-907/341280 46	6-amino-5-[(4-chlorobenzylidene)amino]-2,4-pyrimidinediol	
6	AF-886/314110 29	1-(2-amino-4-methyl-1,3-thiazol-5-yl)ethanone	
7	SEW 00445	ethyl 5-oxo-3-thioxo-2,3,4,5-tetrahydro-1,2,4-triazine-6-carboxylate	
8	7728261	6-amino-5-phenyl-2-thioxo-2,3-dihydropyrimidin-4(*1H*)-one	
9	AB-323/250483 13	3-[(2-amino-5-nitro-6-oxo-1,6-dihydro-4-pyrimidinyl)(2-cyanoethyl)amino]propanenitrile	

10	AK-968/40605465	5-methoxy-7-nitro-1,2,4-benzotriazin-3-amine 1-oxide	
11	59945	Isoxanthopterin	
12	CD 10099	1-(2-amino-4-methylpyrimidin-5-yl)ethan-1-one	
13	RF 03588	2-(ethylsulfanyl)-6-hydroxy-4-pyrimidinecarboxylic acid	
14	AB-323/13887536	7-amino-1,3-dimethyl-2,4(*1H,3H*)-pteridinedione	
15	R238635	2-amino-8-mercapto-9H-purin-6-ol	
16	67074	9-methylguanine	
17	AB-323/13887419	2-amino-6-phenyl-4(*3H*)-pteridinone	
18	BTB 05782	2-aminoquinazolin-4(*1H*)-one	
19	AG-690/11086033	2,6-diamino-9H-purin-8-yl hydrosulfide	

20	AB-323/13887365	2-amino-6,7-dimethyl-5-(3-pyridinylcarbonyl)-5,6,7,8-tetrahydro-4(*3H*)-pteridinone	
21	L136654	2-amino(1,3,5)triazino(1,2-A)benzimidazol-4(*3H*)-one	
22	KM 02704	5-(isopropylsulfonyl)pyrimidine-2,4-diamine	
23	RF 02684	4-(tert-butyl)-6-methoxy-1,3,5-triazin-2-amine	
24	AR-471/43264772	2-amino-8-(methylsulfanyl)-1,9-dihydro-6H-purin-6-one	
25	6298986	4*H*-1,3,5-Triazino[2,1-*b*]benzoxazole-4-thione, 2-amino	
26	RJF 00185	2-amino-6-(trifluoromethyl)pyrimidin-4-ol	
27	RJC 03295	8-amino-1,3-dimethyl-3,7-dihydro-1H-purine-2,6-dione	
28	AC 10402	6-(methylsulfanyl)-7H-purin-2-ylamine	
29	5546844	2-amino-4,7,7-trimethyl-7,8-dihydroquinazolin-5(*6H*)-one	

133

30	GK 01643	1-methyl-1,5-dihydro-4H-pyrazolo[3,4-d]pyrimidin-4-one	
31	AG-690/36920042	2-propyl-5H-[1,3,4]thiadiazolo[3,2-a]pyrimidine-5,7(6H)-dione	
32	9015888	3-phenyl-1H-purine-2,6(3H,7H)-dione	
33	SPB 05754	6-(methylthio)-2,4-dioxo-1,2,3,4-tetrahydropyrimidine-5-carbonitrile	
34	RF 02299	methyl 2-amino-4-oxo-4H-1,3-thiazine-6-carboxylate	
35	RF 03347	4,6-dimethoxy-1,3,5-triazin-2-amine	
36	RJF 00938	2-amino-6-methyl-5-(2-methylallyl)pyrimidin-4-ol	
37	AB-323/13887082	N-benzyl-N-(2,4-diamino-6-oxo-1,6-dihydro-5-pyrimidinyl)acetamide	