
Biophysical and structural studies of novel F₄₂₀-dependent
oxidoreductases in *Mycobacterium tuberculosis*



National Institutes
of Health

Ellene H. Mashalidis

Newnham College

University of Cambridge

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

October 2012

Declaration

I, Ellene H. Mashalidis, declare that this thesis, 'Biophysical and structural studies of novel F₄₂₀-dependent oxidoreductases in *Mycobacterium tuberculosis*' and the work presented therein are my own.

The work described in this thesis was carried out by me at the Department of Chemistry, University of Cambridge and at the Tuberculosis Research Section, National Institutes of Allergy and Infectious Disease, National Institutes of Health under the joint supervision of Professor Chris Abell and Dr. Clifton Barry, III, between October 2008 and September 2012.

This dissertation is the product of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

The work described is original except where indicated by reference, and has not been submitted for any other degree at this or any other university.

This thesis contains fewer than 60,000 words.

Signed _____

Date October 1st, 2012 _____

Περὶ δὲ τῆς ἀπορίας τῆς εἰρημένης περὶ τε τοὺς ὀρισμοὺς καὶ περὶ τοὺς ἀριθμοὺς, τί αἴτιον τοῦ ἓν εἶναι; πάντων γὰρ ὅσα πλείω μέρη ἔχει καὶ μὴ ἔστιν οἷον σωρὸς τὸ πᾶν ἀλλ' ἔστι τι τὸ ὅλον παρὰ τὰ μέρη, ἔστι τι αἴτιον.

With respect both to definitions and to numbers, what is the cause of their unity? In the case of all things which have several parts and in which the totality is not as it were a mere heap, the whole is something besides the parts.

Aristotle, *Metaphysics* 1045a.11

Acknowledgements

I would like to thank *Professor Chris Abell* and *Dr. Clifton Barry* for supervision, advice, funding, and for the privilege of carrying out research in their laboratories.

I thank the *National Institutes of Health-Oxford-Cambridge Scholars Program* for my PhD studentship and for giving me the opportunity to engage in an international collaboration.

I have deep appreciation for *Dr. David Garboczi*, *Dr. Apostolos Gittis*, and *Dr. Kavita Singh* at the Structural Biology Section, National Institutes of Allergy and Infectious Diseases for teaching me the basics of protein crystallography and for inspiring a sense of scientific curiosity in me.

I thank *Dr. Tathagata Mukherjee* at the Tuberculosis Research Section, National Institutes of Allergy and Infectious Diseases for his invaluable training and for teaching me to be confident and competent in the laboratory. I especially thank *Dr. Helena Boshoff* for her support and guidance and *Dr. Alessio Ciulli* for useful discussions and advice.

I also thank collaborators *Dr. Andreas Bender* for his help with chemical informatics, *Dr. Dijana Matak-Vinkovic* for her help with mass spectrometry, and *Dr. Marko Hyvönen* and *Dr. May Marsh* for their help with protein purification strategies and for introducing me to protein crystallography.

Very special thank you goes to my parents, *Dr. and Mrs. Efstathios and Paraskevi Mashalidis*, for their unwavering love and support. I became interested in the world of scientific inquiry because of my upbringing and I would have not pursued this path had it not been for the sense of curiosity, compassion, and work ethic they instilled in me.

I would also like to thank my dear friends *Peter O'Neill*, *Dr. Sarah Winfield*, *Dr. Paweł Sledź*, *Christopher Stubbs*, *Dr. Chiara Valenzano*, *Dr. Duncan Scott*, *Dr. Stefen Lang*, *Dr. Anthony Coyne*, *Dr. Rajavel Srinivasan*, *Dr. Christina Spry*, *Jennifer Boland*, *Rhian Holvey*, *Dr. Nigel Howard*, *Sean Hudson*, *Dr. Sharon Wong*, and *Lisa Goldfeder*. My graduate school experience was enriched by their support and friendship.

Publications

The following work by the author has been published, or is in preparation for publication, as part of, or as mentioned and referenced, in this thesis:

- (1) **Mashalidis EH**, Lang S, Sledz P, Abell C. A Three-Stage Biophysical Screening Cascade for Fragment-Based Drug Discovery. *Nat. Protoc. In press* (2013). (Chapter 1)
- (2) **Mashalidis EH**, Mukherjee T, Sledz P, Matak-Vinković D, Boshoff H, Abell C & Barry CE 3rd. Rv2607 from *Mycobacterium tuberculosis* is a pyridoxine 5'-phosphate oxidase with unusual substrate specificity. *PLoS ONE* 6, e27643 (2011). (Chapter 3)
- (3) **Mashalidis EH**, Gittis A, Abell C, Tomczak, A, Barry CE 3rd & Garboczi DN. Molecular insights into the binding of coenzyme F₄₂₀ to Rv1155 from *Mycobacterium tuberculosis*. *In preparation*. (2013). (Chapter 4)
- (4) **Mashalidis EH**, Hudson SA, Bender A, Barry CE 3rd & Abell C. An Application of Fragment-Based Approaches to Probe the Function of an Unknown Protein in *Mycobacterium tuberculosis*. *In preparation*. (2013). (Chapter 5)

Abstract

The *Mycobacterium tuberculosis* (*Mtb*) genome encodes hundreds of proteins with unknown or putative function based on sequence similarity to characterised proteins. Incomplete knowledge of *Mtb* functional genomics undermines efforts to understand the complex biological repertoire of this deadly human pathogen. While bioinformatics provides useful clues about gene function, biochemical characterisation of gene products remains essential to accurate annotation, particularly in cases where functional redundancy is suggested based on the presence of multiple homologous proteins. The *Mtb* genome includes 7 genes which encode conserved hypothetical proteins with unknown function that are annotated as “pyridoxine 5'-phosphate (PNPOx)-like” proteins based on structural similarity: Rv2607, Rv2991, Rv1155, Rv2074, Rv3369c, Rv1875, Rv0121c. PNPOx is a flavin mononucleotide (FMN)-dependent enzyme involved in the biosynthesis of the crucial cofactor, pyridoxal 5'-phosphate (PLP). The work presented in this thesis shows that while Rv2607 exhibits canonical PNPOx activity, Rv1155 and Rv2991 do not have affinity for FMN as do other members of the PNPOx class. They instead bind tightly to the unusual flavin coenzyme F₄₂₀ by isothermal calorimetry (ITC), saturation transfer difference (STD) NMR, and X-ray crystallography. Although recent studies have drawn attention to F₄₂₀ due to its role in activating the promising anti-tuberculosis drug PA-824, little is known about the function of this coenzyme in *Mtb* metabolism. In order to identify the substrate recognition properties of Rv1155 and Rv2991 respectively, a novel application of fragment-based approaches was employed. A library of fragments was designed based on known ligands to flavoenzymes and screened for binding to Rv1155 and Rv2991 in the presence and absence of F₄₂₀ using STD NMR and differential scanning fluorimetry. Rv1155 and Rv2991 show affinity for distinct classes of fragments. Using a computational method known as “virtual fragment linking,” fragment binding data was used to predict and test potential substrates. This approach to probing structure-function relationships may serve as a generalisable method for exploring functional diversity within a structural class of proteins.

Contents

Chapter 1: Introduction

1.1 Tuberculosis is a pressing global health problem	1
1.2 Approaches to TB drug discovery	3
1.2.1 Whole-cell screening	4
1.2.2 Target-based screening and rational drug design	6
1.2.2.1 High-throughput screening for rational drug design	7
1.2.2.2 Fragment-based approaches	8
1.2.2.3 Shortcomings of rational drug design	13
1.2.3 WCS and rational design as complementary approaches	14
1.3 Functional annotation of the <i>Mtb</i> genome	15
1.4 A novel application of fragment-based methods for natural substrate prediction	16

Chapter 2: Analysis of pyridoxine 5'-phosphate oxidase-like proteins in *M. tuberculosis* using sequence and structure-based comparisons

2.1 Background	18
2.2 Analysis of PNPOx-like proteins using comparative genomics	20
2.2.1 Protein sequence alignments	20
2.2.2 Gene clusters and gene synteny analysis	22
2.3 Structural Analysis of PNPOx-like proteins	24

Chapter 3: Biophysical and enzymatic characterisation of Rv2607 as a pyridoxine 5'-phosphate oxidase

3.1 Background	34
3.2 Protein expression and purification	38
3.3 Analysis of Rv2607 using mass spectrometry	38
3.4 Kinetic studies of Rv2607	41
3.4.1 LC-MS assay to detect Rv2607-catalysed PLP formation	42
3.4.2 Enzymatic reaction time course using ¹ H NMR spectroscopy	43
3.4.3 Michaelis-Menten kinetics for Rv2607 and PNP	44

3.5 Proposed mechanism	45
3.6 Conclusions	45

Chapter 4: Identification of Rv2991 and Rv1155 as novel F₄₂₀ binding proteins

4.1 Introduction	46
4.1.1 F ₄₂₀ -dependent activation of PA-824, a TB drug currently in clinical trials	47
4.1.2 F ₄₂₀ is a low redox potential flavin coenzyme and obligate hydride donor	47
4.1.3 F ₄₂₀ -dependent enzymes in <i>Mtb</i>	50
4.2 Objective and approach	51
4.3 Expression and purification of Rv1155 and Rv2991	52
4.4 Purification of F ₄₂₀	52
4.5 Evidence for Rv1155-F ₄₂₀ and Rv2991-F ₄₂₀ binding interactions	53
4.5.1 Rv1155 and Rv2991 bind to F ₄₂₀ as determined by biophysical methods	53
4.5.2 X-ray crystallographic studies with Rv1155 and Rv2991	59
4.6 Trial enzymatic reactions with Rv1155 and Rv2991	65
4.6.1 Rv1155 and Rv2991 do not exhibit PNPOx activity	63
4.6.2 Rv1155 and Rv2991 do not recognize aflatoxin as a substrate	64
4.7 Conclusions	66

Chapter 5: Design, development, and screening of a targeted flavoprotein fragment library

5.1 Background	67
5.2 Objective and rationale	68
5.3 Methodological approach	69
5.3.1 Fragments as chemical tools	69
5.4 F ₄₂₀ -dependent enzymes Rv1155 and Rv2991 as model CHPs	71
5.5 Library design and development	73
5.6 Fragment screen of Rv1155 and Rv2991	75
5.7 Virtual fragment linking results and preliminary enzyme assays	78
5.8 Conclusions and future work	82

Chapter 6: Experimental

6.1 Protein overexpression and purification	84
6.1.1 Purification of Rv2607	84
6.1.2 Purification of Rv1155 and Rv2991	85
6.2 Biophysical characterisation of Rv2607, Rv1155, and Rv2991 by MS	87
6.2.1 Confirmation of molecular weight	87
6.2.2 MS analysis of oligomerisation state and cofactor stoichiometry	87
6.2.3 Analysis of co-purified cofactor from Rv2607	88
6.3 Enzymatic characterisation of Rv2607	89
6.3.1 Synthesis of PNP	89
6.3.2 LC-MS assay for Rv2607-dependent PLP formation	90
6.3.3 NMR time-course assay of Rv2607-dependent PLP formation	91
6.3.4 Steady state kinetic characterisation of Rv2607	92
6.4 Test enzymatic reaction with Rv1155 and Rv2991	93
6.5 Purification of F ₄₂₀	93
6.5.1 Isolation and purification of F ₄₂₀ from <i>M. smegmatis</i>	93

6.5.2 Enzymatic reduction of F ₄₂₀	95
6.6 Protein crystallography	99
6.7 Isothermal titration calorimetry	96
6.8 Optimisation of DSF experiments	100
6.9 Fragment library design, development, and screening	101
6.9.1 Fragment library design	101
6.9.2 Fragment library construction	101
6.9.3 NMR screen of fragment library	102
6.9.4 DSF screen of fragment library	102
6.10 Rv2991 substrate prediction	103
6.10.1 Virtual fragment linking	103
6.10.2 Test enzymatic assays with Rv2991	103
Appendix A	106
Appendix B	108
Appendix C	118

Abbreviations

Abbreviation	Name
BLAST	basic local alignment search tool
pBLAST	basic local alignment search tool for protein sequences
BTZ	benzothiazinone
β -Me	beta mercaptoethanol
CHP	conserved hypothetical protein
CD	circular dichroism
CPMG	Carr-Purcell-Meiboom-Gill
Dali	distance alignment matrix method
Ddn	deazaflavin-dependent nitroreductase
DMSO	dimethyl sulfoxide
DNPH	2,4-dinitrophenylhydrazine
DSF	differential scanning fluorimetry
DTT	dithiothreitol
DXP	deoxyxyulose 5'-phosphate
FAD	flavin adenine dinucleotide
FBDD	fragment-based drug discovery

FGD	F ₄₂₀ -dependent glucose 6-phosphate dehydrogenase
FMN	flavin mononucleotide
GST	glutathione S-transferase
His	polyhistidine tag
HTS	high-throughput screening
HPLC	high-pressure liquid chromatography
IC ₅₀	half maximal inhibitory concentration
iNOS	nitric oxide synthase
IPTG	isopropyl thiogalactoside
ITC	isothermal titration calorimetry
KEGG	Kyoto Encyclopedia for Genes and Genomes
LB	Luria-Bertani
LC	liquid chromatography
LLM	luciferase-like monooxygenase
MBP	maltose binding protein
MDR-TB	multi-drug resistant tuberculosis
MIC	minimum inhibitory concentration
MS	mass spectrometry
<i>Mtb</i>	<i>Mycobacterium tuberculosis</i>
MW	molecular weight
NAD	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
nESI-MS	nano electrospray ionisation mass spectrometry
NOE	nuclear Overhauser effect
NMR	nuclear magnetic resonance
PBS	phosphate buffer saline
PDB	Protein Databank
PEG	polyethylene glycol

PLP	pyridoxal 5'-phosphate
PMP	pyridoxamine 5'-phosphate
PNP	pyridoxine 5'-phosphate
PNPOx	pyridoxine 5'-phosphate oxidase
RMSD	root mean square deviation
RNI	reactive nitrogen intermediates
SAR	structure activity relationship
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPR	surface plasmon resonance
SPE	solid phase extraction
STD	saturated transfer difference
TRIS-HCl	tris(hydroxymethyl)aminomethane hydrochloric acid
TB	tuberculosis
TB-SGC	Tuberculosis Structural Genomics Consortium
TEV	tobacco etch virus
TFA	trifluoroacetic acid
T _m	melting temperature
TSP-d ₄	2,2,3,3-D ₄ -3-(trimethylsilyl)propionic acid
VFL	virtual fragment linking
WaterLOGSY	water-ligand observed by gradient spectroscopy
WCS	whole-cell screening
WHO	World Health Organization
WT	wild type
XDR-TB	extensively drug resistant tuberculosis

Figures and Tables

Figure 1.1. Chemical structures of first-line TB drugs.	2
Figure 1.2. Chemical structures of anti-TB compounds discovered using WCS.	5
Figure 1.3. Summary of the fragment elaboration approach to discover Vemurafenib, the first fragment-based drug to be approved by the FDA.	9
Figure 1.4. Three stage fragment screening cascade for fragment selection.	11
Figure 1.5. A new strategy for TB drug design	14
Figure 2.1. Schematic of PNPOx-catalysed PLP formation	19
Table 2.1. Summary of NCBI gene product annotations for PNPOx-like proteins in selected organisms where PNPOx activity is considered to play a role in PLP salvage	21
Table 2.2. Pairwise percent identities of PNPOx-like proteins	22
Figure 2.2. Gene cluster comparisons of PNPOx-like proteins in <i>Mtb</i> and <i>M. smegmatis</i>	23
Figure 2.3. Structural superimposition of <i>E. coli</i> PNPOx and <i>Mtb</i> PNPOx-like proteins (protomers)	25

Table 2.3. Summary of Dali results for PNPOx-like proteins Rv2607, Rv1155, Rv2991, and Rv2074	26
Figure 2.4. Sequence alignment of <i>E. coli</i> PNPOx, <i>H. sapiens</i> PNPOx, and <i>Mtb</i> PNPOx-like proteins	27
Figure 2.5. Structure of the <i>E. coli</i> PNPOx active site	28
Figure 2.6. Structural superimpositions of <i>E. coli</i> PNPOx active site with predicted active sites in <i>Mtb</i> PNPOx-like proteins	30
Figure 2.7. Residues in Rv2074, Rv2991, and Rv1155 that could engage in hydrogen bonding with FMN	31
Figure 2.8. Structural superimposition of Rv1155-FMN and Rv1155-PLP	32
Figure 3.1. PNPOx-catalysed PLP formation in DXP-dependent and DXP-independent PLP biosynthetic pathways	37
Figure 3.2. Identification of cofactor bound to Rv2607 by nESI-MS	39
Figure 3.3. Mass spectrum of intact Rv2607 with co-purified FMN analysed by nESI-MS	40
Figure 3.4. MS/MS spectrum of the Rv2607 homodimer	41
Figure 3.5. Reverse-phase HPLC analysis of the PNPOx activity of Rv2607	42
Figure 3.6. ¹ H NMR analysis of the conversion of PNP into PLP with time	43
Figure 3.7. Michaelis-Menten plot for Rv2607 with PNP as a substrate	44
Figure 3.8. A schematic of the proposed mechanism for the oxidation of PNP by Rv2607	45
Figure 4.1 A comparison of the nitroimidazole mechanism of action and the NO release that occurs as a result of the innate immune response	48
Figure 4.2. Structures of F ₄₂₀ , flavins, and NADP ⁺ and their associated chemical reduction potentials	49
Figure 4.3. Predicted F ₄₂₀ -binding enzymes in <i>Mtb</i> using phylogenetic profiling	51
Figure 4.4. Cofactor recognition of Rv1155 and Rv2991 assessed by DSF	54
Figure 4.5. F ₄₂₀ binds to Rv1155 and Rv2991 by STD NMR	56
Figure 4.6. ITC binding isotherms for F ₄₂₀ titrated into Rv1155 and Rv2991	57

Figure 4.7. Nano-ESI mass spectra of intact Rv2991 and Rv1155 in the presence of F ₄₂₀	58
Figure 4.8. Crystal structure of Rv1155 bound to F ₄₂₀ at 2.8 Å resolution	60
Figure 4.9. Electrostatic potential surface representation of F ₄₂₀ binding site in Rv1155	62
Figure 4.10. Structural alignment of Ddn, Rv1155, and Rv2991	63
Figure 4.11. Test enzymatic reactions conducted with Rv1155 and Rv2991	65
Figure 5.1. A schematic describing the 'fragment-to-substrate' workflow	71
Figure 5.2. Structural comparison of F ₄₂₀ -dependent Rv1155 and Rv2991	73
Figure 5.3. STD NMR fragment hits for Rv2991	76
Figure 5.4. STD NMR fragment hits for Rv1155	78
Figure 5.5. Schematic demonstrating the VFL approach	79
Figure 5.6. Rv2991-F ₄₂₀ H ₂ enzyme assay for predicted substrates	81
Table 6.1. Electrospray ionisation MS analysis Rv2607, Rv1155, and Rv2991 proteins used in the biophysical and enzymatic assays	88
Figure 6.1. Reverse-phase HPLC analysis of control reactions between PLP, PNP, or PMP, and DNPH	91
Figure 6.2. F ₄₂₀ concentration found in <i>M. smegmatis</i> cells transformed with FbiABC overexpression vector compared to WT	94
Figure 6.3. Enzymatic reduction of F ₄₂₀ by FGD-1	96
Figure 6.4. Chromatogram and SDS-PAGE analysis of Rv1155-F ₄₂₀ co-purification by SEC	97
Table 6.2. Crystallographic conditions that produced crystals for Rv1155-F ₄₂₀	98
Figure 6.5. Optimisation of apoenzyme Rv2991 crystals	99
Table 6.3. Select VFL hits that were tested for enzymatic activity with Rv2991 and F ₄₂₀ H ₂	104
Table A.1. Crystallographic refinement statistics for Rv1155-F ₄₂₀ datasets 1-3	106
Table A.2. Crystallographic refinement statistics for Rv1155-F ₄₂₀ datasets 4-5	107

Table C.1. Melting temperatures of Rv2991 and Rv1155 in the presence of fragments from the targeted flavoenzyme library	118
Figure C.1. STD NMR spectra for fragment hits against Rv2991 from the targeted flavoenzyme fragment library	119
Figure C.2. STD NMR spectra for fragment hits against Rv1155 from the targeted flavoenzyme fragment library	120
Table C.2. Fragments from the Maybridge library that increase the T_m of Rv1155 by at least twice the standard deviation of the DMSO-only control on each screening plate	120
Figure C.3. STD NMR spectra for fragment hits against Rv1155 from the Maybridge Library	121

Chapter 1

Introduction

1.1 Tuberculosis is a pressing global health problem

Tuberculosis (TB) has afflicted humans for thousands of years. The earliest evidence of TB infection was discovered in the preserved remains of a mother and child who lived in the Eastern Mediterranean approximately 9000 years ago¹. Today, the global burden of TB is enormous. The causative bacterium, *Mycobacterium tuberculosis* (*Mtb*), is amongst the most ubiquitous human pathogens known, with an estimated one third of the world population carrying the bacterium². For reasons that are poorly understood, only around 5-10% of people who are infected with the bacterium develop active disease³. The most recent World Health Organization (WHO) *Global Tuberculosis Report* revealed

that in 2011 there were 8.7 million new cases of TB and 1.4 million deaths from TB². One in four deaths amongst HIV-positive individuals is caused by TB⁴.

The current drug regimen for TB treatment is decades old and falls short of controlling this global epidemic⁵. Drug-susceptible TB is currently treated over the course of a minimum of 6 months with the use of isoniazid, rifampicin, pyrazinamide and ethambutol for the first 2 months, followed by isoniazid and a rifampicin treatment for the remaining 4 months⁵. This treatment regimen is problematic primarily because it is too long and often results in patient non-compliance, which leads to drug-resistant disease⁶. Further, some first-line TB drugs are incompatible with certain treatments for other conditions, such as antiretrovirals for HIV⁷ and some anti-diabetic medications⁸.

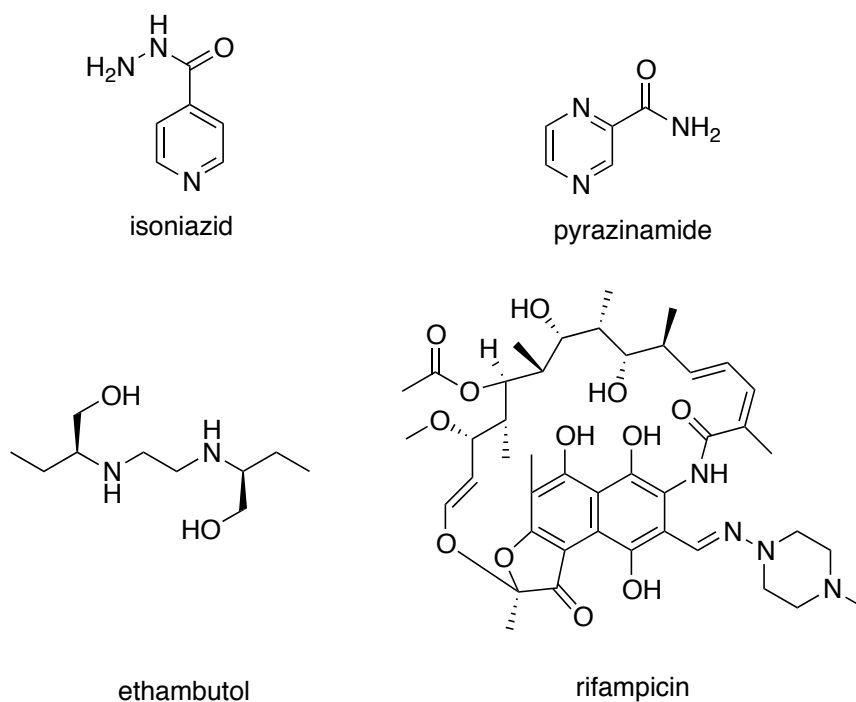


Figure 1.1. Chemical structures of first-line TB drugs.

The emergence of drug-resistant strains of *Mtb* is perhaps the most significant threat to the control of TB worldwide⁹. These strains are classified by the extent of their drug resistance as either multidrug-resistant TB (MDR-TB) or

extensively drug-resistant TB (XDR-TB). MDR-TB is characterised by resistance to one or more of the first-line drugs against TB, including isoniazid and rifampicin, which are the most potent drugs to treat TB currently available¹⁰. XDR-TB is resistant to isoniazid and rifampicin as well as to any of the fluoroquinolones (for example, ofloxacin or moxifloxacin) and to at least one of three intravenously administered second-line drugs (amikacin, capreomycin or kanamycin)¹¹. The mortality rate among XDR-TB patients is around 70%⁵. Recently, patients with TB that is resistant to all known drugs have been reported¹².

The widespread prevalence of *Mtb* and drug-resistant strains underscores the urgent need for new antibiotics to treat TB¹³. The emergence of this global health threat has driven a recent increase in research activity to identify and validate new TB drug targets¹⁴. Despite renewed interest in developing new TB treatments, realisation of this goal has met with difficulty. There are two major challenges: 1) understanding biological processes that are vulnerable to chemotherapeutic intervention and 2) designing and implementing strategies to develop lead compounds that target these processes¹⁵.

1.2 Approaches to TB drug discovery

For much of the 20th century, antibiotics were largely discovered from microbial natural products¹⁶. The first antibiotic agent against TB was streptomycin⁶, isolated from *Streptomyces griseus*. By the late 1990s, the widespread availability of genome sequencing technologies sparked an interest in identifying new targets for antibiotic design¹⁷. The genome of *Mtb* was sequenced in 1998 and gene products were identified primarily based on sequence similarity with known proteins^{18,19}.

Candidate drug targets were initially selected by identifying genes required for the growth and survival of *Mtb*. 'Genetic essentiality' was ascertained via a series of experiments^{20,21} that measured the phenotypic effect of a disrupted gene of interest. Potential targets were taken forward if they had no

close homologues in humans. Once a target was identified and validated with genetic experiments, it was expressed, purified, and screened against large compound libraries to identify inhibitors of target activity. This process is referred to as the genome-based, 'rational' approach to TB drug discovery. Concurrently, an alternative approach to TB drug discovery was developed that entails screening compound libraries against cultures of *Mtb* cells. Using this method, a hit compound is one effective in killing *Mtb* cells, as quantified by a minimum inhibitory concentration (MIC) value²². Although whole-cell screening has produced far more hits with anti-tubercular activity, each approach offers certain advantages and limitations, which are described in the following.

1.2.1 Whole cell screening

Whole cell screening (WCS) has been successful in producing hits that subsequently can be developed into clinical candidates for TB. Important anti-TB compounds derived from active compounds identified by WCS include the benzothiazinones (BTZs)^{23,24}, PA-824²⁵, SQ109²⁶, and TMC207 or bedaquiline²⁷ (trade name Sirturo), which recently became the first drug in four decades to be approved by the United States Federal Food and Drug Administration (FDA) for the treatment of TB²⁸.

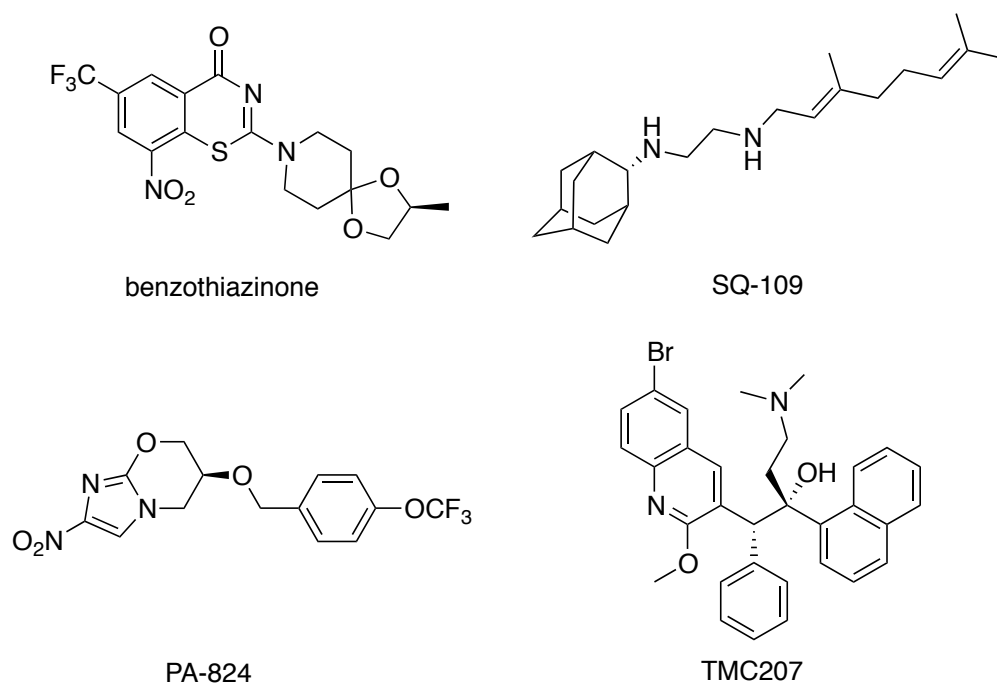


Figure 1.2. Chemical structures of anti-TB compounds discovered using WCS.

WCS is typically conducted in a high-throughput format that monitors the effect of the compound on *Mtb* growth and survival. Compounds can be screened against the bacterium in a microplate under a wide variety of environmental conditions, including those that mimic various disease states. Bacterial viability can be rapidly accessed with the use of spectroscopic techniques that detect colorimetric indicator dyes²⁹, ATP levels^{30,31,32} or recombinant GFP levels³³. Compounds with low MIC are further investigated to determine the effect of dose on cell viability and to rule out non-specific cytotoxicity. The mechanism of action for promising active compounds is determined by sequencing the genome of spontaneous drug-resistant mutants^{25,24}, affinity purification³⁴, or comparative transcriptional fingerprinting for drugs with a known killing mechanism³⁵. This approach often reveals new, compound-induced bactericidal mechanisms and targets that are not necessarily predictable by genome analysis.

Although WCS screening has been fruitful in identifying compounds that inhibit *Mtb* growth, it also has several limitations and weaknesses. For example, growth inhibition can occur by several mechanisms and follow-up investigation of hit compounds sometimes reveals that they are generally cytotoxic³⁶. Further, target identification for a WCS hit is a lengthy and costly process that can be difficult to carry out successfully in practice, particularly if drug-resistant mutant strains prove to be particularly challenging to isolate³⁷. Optimisation of a WCS-derived hit without knowledge of its target is an inefficient process that precludes hypothesis-driven analysis of structure-activity relationships. By contrast, target-based screening and rational drug design, discussed in the following section, can rapidly advance a hit to a lead compound. Although WCS offers the advantage of identifying targets vulnerable to inhibition, subsequent efforts to optimise lead compounds are hampered by a lack of information about the target.

1.2.2 Target-based screening and rational drug design

At the turn of the 21st century, the availability of technology to facilitate high-throughput crystallography changed the landscape of drug design and heralded the era of rational, structure-based drug design³⁸. It was thought that three-dimensional structures of proteins would bridge the gap between genomic information and medical breakthroughs³⁹. Target-based drug design generally involves screening compound libraries against the target protein using biophysical or enzymatic assays. X-ray crystal structures of the hit compounds in complex with the target protein are then obtained and used to design new compounds with enhanced potency⁴⁰. Several structural genomics initiatives were established in the last decade to encourage rational drug design efforts. The Tuberculosis Structural Genomics Consortium (TB-SGC) has produced over 250 unique open reading frame structures since its inception in 2000⁴¹. One advantage of rational drug design programmes for bacterial proteins (over eukaryotic proteins) is that they are often readily expressed using recombinant

systems and can be purified in large quantities required for crystallisation and screening experiments¹⁷; however, there are many ways in which bacterial systems are challenging for lead compound development, as described in section 1.2.3. Rational drug design was particularly emphasised in the early phases of TB drug discovery and although it has been relatively unsuccessful in producing lead compounds, it still plays an important role today's TB drug design efforts.

1.2.2.1 High-throughput screening for rational drug design

Target-based high-throughput screening (HTS) involves screening a purified macromolecule against large libraries containing on the order of 10^4 - 10^6 compounds. HTS typically includes assays that measure either *in vitro* protein activity or binding affinity^{22,37}. Many HTS programmes are initiated with *a priori* knowledge of the target structure. 'Hit' compounds, *i.e.* those that meet the imposed standards for activity inhibition or binding, are co-crystallised with the target to elucidate the binding mode.

In vitro protein activity assays monitor rates of product formation in the presence and absence of a test compound by a spectrophotometric technique. For compounds that inhibit the protein activity relative to the control, half the maximal inhibitory concentration (IC_{50}) is typically measured, which is obtained from a dose-response curve. Inhibitors for a number of TB protein targets have been discovered using this approach, including InhA (target of known drug isoniazid⁴²), PknA and PknB (essential serine/threonine kinases^{43,44,45}), MenB (an enzyme in the menaquinone biosynthetic pathway⁴⁶), and PanC (pantothenate synthetase⁴⁷).

HTS programmes are sometimes conducted with binding assays rather than activity assays, although this approach is relatively less common in TB drug design efforts. It involves one or more methods that provide rapid, qualitative binding results. Compounds binding to the target protein are typically further

characterised to determine the binding affinity and binding mode. The techniques used to collect such data are described in the following section on fragment-based approaches to rational drug design. HTS programmes for *Mtb* isocitrate lyase (GlaxoSmithKline and Novartis Institute for Tropical Diseases) and InhA (GlaxoSmithKline and Global Alliance for TB Drug Development) have been reviewed³⁷, although the results have not been published. In each case, very hydrophobic active sites may have selected for compounds with undesirable physiochemical properties³⁷.

While a number of novel inhibitors of *Mtb* proteins have been identified by target-based HTS, none have emerged as clinical candidates. This is partly due to the limitations of genome-based, rational drug design for anti-TB drug design in general, as discussed in section 1.2.1.3. However, the HTS approach to rational drug design introduces additional challenges to lead compound discovery. HTS requires access to large very compound libraries which are costly to develop and challenging to maintain⁴⁸. Although HTS produces more tightly binding initial hits, increasing their potency and simultaneously improving their biophysical profile can often be difficult⁴⁹. These challenges can be addressed by fragment-based approaches to rational drug design.

1.2.2.2 *Fragment-based approaches*

Fragment-based drug discovery (FBDD) is gaining traction in the pharmaceutical industry as a complementary or alternative approach to HTS. The first fragment-based drug, Vemurafenib, was approved by the FDA in 2011 for the treatment of metastatic melanoma⁵⁰ and there are several fragment-based drug candidates currently in clinical trials⁵¹ (please refer to Figure 1.3 for a summary of the fragment-based approach that led to the discovery of Vemurafenib). 'Fragments' are low molecular weight molecules (<300 Da) with low chemical complexity that interact weakly, yet specifically, with a macromolecular target⁵². Information about fragment binding modes can be used to link, alter, or grow

bound fragments so as to improve affinity for the target. FBDD can be applied to a wide range of targets, including enzymes^{53,54,55,56,57}, nucleic acids⁵⁸, and protein-protein interactions^{59,60}. Fragments have been used to derive lead antibiotics against bacterial targets^{53,54,61,62,63,64,65}. The FBDD approach is an appealing alternative to HTS technology because enhancing the potency of a fragment hit is a stepwise process that allows for the systematic incorporation of ligand binding interactions and parallel fine-tuning of its physicochemical properties.

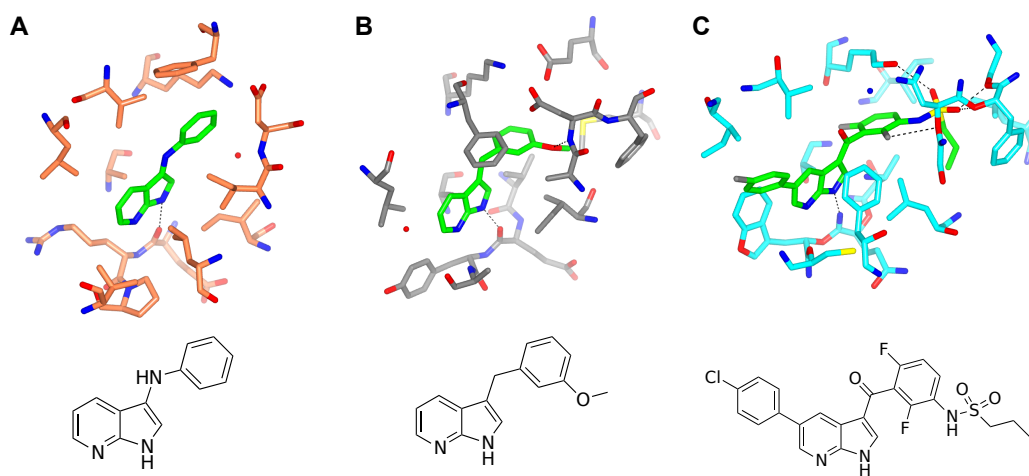


Figure 1.3. Summary of the fragment elaboration approach to discover Vemurafenib, the first fragment-based drug to be approved by the FDA. Vemurafenib targets oncogenic B-raf kinase, which is implicated in malignant melanoma. (A) The structure of fragment 3-aminophenyl-7-azaindole shown bound to Pim-1 kinase ($IC_{50} > 200 \mu M$). (B) The structure of an elaborated fragment, 3-(3-methoxybenzyl)-7-azaindole, bound to the kinase domain of fibroblast growth factor receptor 1 ($IC_{50} \approx 1.9 \mu M$). (C) The structure of Vemurafenib bound to B-raf kinase ($IC_{50} = 31 nM$)⁶⁶. Superimposition of the structures in A and B, revealed that the azaindole scaffold can target the hinge region in kinases and that modifying substitution sites on the scaffold can optimize potency and selectivity for B-raf kinase⁶⁷. Screening various substituted compounds against the oncogenic B-raf kinase led to the discovery that the difluoro-phenylsulfonamide motif has enhanced selectivity for the oncogenic B-raf⁶⁷.

The relative weakness of fragment-target interactions requires sensitive biophysical methods to detect binding. Techniques such as ligand-observed nuclear magnetic resonance (NMR)⁶⁸, differential scanning fluorimetry (DSF)⁶⁹, high-throughput X-ray protein crystallography^{70,71}, surface plasmon resonance (SPR)⁷², and mass spectrometry (MS)⁷³ are used to screen fragment libraries.

These screening technologies have been used in FBDD programmes in various combinations and manners, with no two screening programs completely alike. There is no “universal” way to conduct a fragment screen, partly because this process is dependent on both the nature of the target and the availability of screening resources.

In our laboratory, we have demonstrated that many targets are tractable in FBDD programs using a three-stage process consisting of *preliminary screening*, *validation*, and *characterisation* stages, as described in Figure 1.1. The preliminary fragment screening is conducted with DSF. The validation phase consists of confirming the preliminary screen hits by ligand-observed NMR methods. In the final phase, the validated fragments are further characterised by approaches, such as isothermal titration calorimetry (ITC) and X-ray crystallography. Low throughput, low cost methods are used to screen and validate the majority of the fragment library followed by lower-throughput, information-rich experiments to characterise relatively few binding fragments in detail. Each of these methods is described below.

DSF (also known as thermal shift or thermal unfolding) is a spectroscopic technique used to measure the temperature-dependent unfolding of a protein. This is achieved by monitoring the fluorescence of a dye that preferentially binds to hydrophobic protein surfaces exposed upon unfolding^{74,75,76}. The dependence of fluorescence on temperature is described by a sigmoidal curve and the inflection point of the transition defines the melting temperature of the protein (denoted as T_m). A change in T_m in the presence of a ligand is termed the “thermal shift” and is used to distinguish binding ligands ($\Delta T_m > 0$) from non-binders ($\Delta T_m \leq 0$)⁷⁷. A number of experimental parameters must be optimised to conduct a fragment screen using DSF. These include dye concentration, protein concentration, buffer type and ionic strength, ligand concentration, temperature range, and DMSO concentration. A number of dyes are commonly used in DSF⁷⁷, but SYPRO® Orange is preferred for fragment screening due to its high

excitation wavelength (500 nm), which reduces the risk of compound interference⁶⁹.

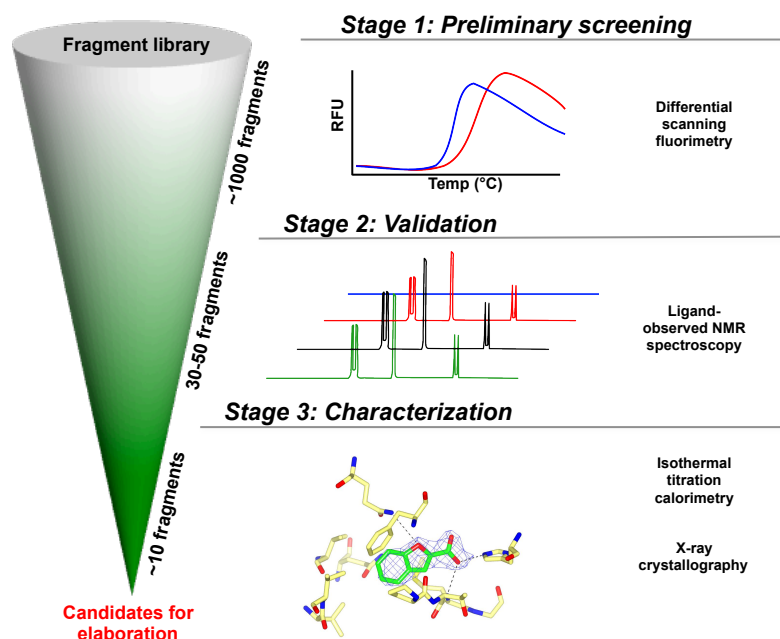


Figure 1.4. Three stage fragment screening cascade for fragment selection. The cascade consists of preliminary fragment screening, hit validation, and binding characterisation. Higher throughput techniques are used at an early stage of the process, and more complicated lower-throughput approaches are used for characterisation.

NMR spectroscopy is perhaps one of the most trusted fragment screening techniques, as it was the first reported tool to screen fragments⁷⁸ and is commonly used today⁷⁹. Ligand-observed NMR experiments are most widely applied to fragment screening because they do not require protein structural information and are not restricted by protein size⁸⁰. Examples of ligand-observed NMR techniques include saturation transfer difference (STD)⁸¹, water-ligand observed by gradient spectroscopy (WaterLOGSY)⁸², and Carr-Purcell-Meiboom-Gill (CPMG) sequence^{83,84}.

STD involves selective perturbation of protein-specific methyl proton resonances (typically 0.5 ppm to 1 ppm), which rapidly diffuses throughout the protein⁸¹. This results in a transfer of magnetisation from the protein to any

transiently bound ligands via the nuclear Overhauser effect (NOE)⁶⁸. If the fragment binds to the target protein, the build-up of NOE that is transferred to the ligand results in enhanced signal corresponding to the resonances of that ligand in the STD spectrum. A number of factors affect the signal strength in an STD experiment, including protein size, duration of on resonance irradiation, the frequency of irradiation, the dissociation constant of the ligand, the ligand/protein ratio, and the field of the spectrometer⁸⁵. As such, it is useful to vary parameters such as ligand and protein concentration, the frequency of irradiation, and the duration of irradiation. A control experiment in the absence of protein is performed to confirm that the observed change is protein-dependent.

In the WaterLOGSY experiment, magnetisation is transferred to the ligand from the bulk water, mainly through water molecules bound in or adjacent to the ligand binding site^{86,82}. If the fragment is bound, the negative NOEs will dominate the signal. No change in the sign of the WaterLOGSY signal indicates that the fragment is unbound. CPMG is used to detect fragment binding by relying on differences in the relaxation rates of fragment NMR signals between the bound and free states⁸⁷. In the CPMG experiment, a time delay of 100–400 ms is used before acquisition to filter out the observed signals for rapidly relaxing ligands^{83,84}. Small molecules relax slowly giving rise to sharp peaks, while proteins relax quickly and tend to show broad NMR signals. Relaxation-edited NMR experiments exploit such differences in relaxation properties between proteins and small molecules⁶⁸. A signal decrease is expected in the presence of protein if the fragment binds the protein.

Fragments that are confirmed to bind to the protein by NMR are typically characterised by ITC to obtain thermodynamic information about the binding interaction. ITC directly measures the heat associated with forming or breaking non-covalent bonds (ΔH_{obs}) and the stoichiometry of binding (n). This information can be used to determine the K_D . With the equilibrium binding constant and $\Delta H^{\circ}_{\text{obs}}$, ΔS° and ΔG° can be calculated using the relationships in the following equation.

$$RT\ln K_D = \Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

Full thermodynamic characterisation of binding fragments is very useful for downstream inhibitor design and optimisation.

Fragment-based approaches have been used to discover inhibitors for proteins in *Mtb*, including of pantothenate synthetase (PtS)^{88,89} (best K_D = 860 nM), the protein tyrosine phosphatases PtpA⁶⁴ and PtpB⁶⁵ (best K_i = 1.4 μ M for PtpA, K_i = 0.22 μ M for PtpB), and cytochrome P450⁵⁴ (best K_D = 28 μ M). However, thus far no fragment-derived compounds have been reported with a low MIC for *Mtb*. While inactive against wild type *Mtb*, the aforementioned PtS inhibitors were shown to inhibit the growth of conditional mutant strains that underexpress pantothenate synthetase⁹⁰.

1.2.2.3 Shortcomings of rational drug design

Genome-derived, target-based antibiotic design efforts have failed to produce clinical candidates to date¹⁷. Inhibitors developed by rational methods have been ineffective in measurably reducing *Mtb* growth and survival. A problem may be that some inhibitors fail to reach their target either because they cannot penetrate the thick mycobacterial cell wall³⁶ or because of efflux mechanisms⁹¹. There is insufficient understanding of the physiochemical basis for cell-permeability. Even if the inhibitor overcomes the hurdle of cell entry, bactericidal mechanisms may be too slow. For example, rationally-designed, potent inhibitors of *Mtb* peptide deformylase were taken up by cells⁹², but the kinetics of growth inhibition were so slow that only moderate efficacy was observed in mouse models^{36,36}.

Another potential pitfall for target-based inhibitor design efforts is the effect of compensatory metabolic mechanisms. While a potent inhibitor can effectively inhibit one enzyme in a metabolic pathway, alternative pathways can circumvent the metabolic effects of this inhibition. For example, nicotinamide adenine dinucleotide (NAD) salvage pathways in *Mtb* can compensate for inhibition of the *de novo* NAD biosynthetic pathway⁹³, which is essential for *Mtb*

survival and was thought to be a good drug target⁹⁴. Some compensatory metabolic pathways can be difficult to predict from the *Mtb* genome, particularly because about one third of the *Mtb* genome encodes proteins of unknown or putative function¹⁹.

1.2.3 WCS and rational design as complementary approaches

WCS and target-based approaches were previously conducted in isolation, which limited TB drug discovery efforts by their respective weaknesses. While WCS is an effective method of identifying targets in *Mtb* vulnerable to chemical intervention, there is no efficient way to enhance the potency of a WCS hit. Rational, target-based approaches can rapidly drive up the potency of hit compounds. New strategies aim to integrate these two approaches to efficiently arrive at lead compounds³⁶. WCS and genome sequencing is used to identify target and discover active, cell-permeable chemical scaffolds, while rational methods are applied to enhance the binding affinity and introduce favourable physiochemical properties. Protein structural information bridges the gap between these two phases. The process is summarised in a schematic shown in Figure 1.2.

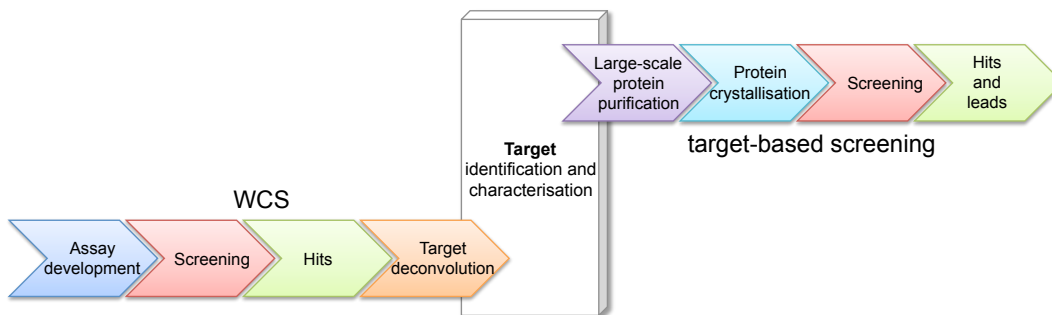


Figure 1.5. A new strategy for TB drug design. The schematic represents the workflow for combined WCS (left) and for target-based screening (right). WCS is applied to identify active, cell-permeable compounds and a target vulnerable to inhibition. Bioinformatic analysis can be used as a starting point for generating hypotheses about the function of the target. The target must be cloned into an expression vector, expressed, and purified for basic biochemical characterisation. A protocol for protein purification is developed on a scale that is appropriate for crystallisation trials and screening experiments. Screening is conducted via ligand-binding or enzyme assays. Hits and leads are developed and potency is increased based on protein structural information. The figure is

adapted from Terstappen, G.C. *et al. Nat. Rev. Drug Discov.* 2007.³⁴

This new approach for TB drug discovery is currently being applied to the BTZ series of compounds, which target DprE1, a decaprenylphosphoryl arabinofuranose²⁴. DprE1 was identified by whole genome sequencing of BTZ-resistant mutants²⁴, it was functionally characterised⁹⁵, and crystal structures of DprE1 complexed with active compounds were recently reported^{96, 97}. This structural information serves as foundation for future inhibitor design studies. Central to this approach is identifying and characterising the protein target, a task which is often non-trivial, as there is a serious dearth of functional genetic information for *Mtb*. This challenge to TB drug discovery is further discussed in the following section, as well as in Chapter 5.

1.3 Functional annotation of the *Mtb* genome

The whole genome sequencing of H37Rv¹⁸, a laboratory reference strain, was an important advance in the understanding of *Mtb* biology. However, at the time of its release, only 40% of H37Rv gene products could be putatively annotated with computational methods that assign function based on sequence similarity to proteins in other organisms. Four years after the sequence of the H37Rv genome was published, it was re-annotated using updated genomic and functional data for sequence comparisons, which identified new coding sequences and adjusted predicted protein lengths¹⁹. This study found that of the 3995 predicted proteins, 2052 could be annotated with a putative function (51%), although only around 150 of these proteins were characterised experimentally. Further, 272 proteins were annotated as “proteins with unknown function” without homologues in other organisms. A staggering 1051 proteins were annotated as ‘conserved hypothetical proteins’ (CHPs), which are proteins with homologues in multiple phylogenetic lineages, none of which have a known function. An online public database, TubercuList⁹⁸ (<http://tuberculist.epfl.ch/>), provides current and curated information on the H37Rv genome that is revised as experimental studies yield

new functional information. The *Mtb* proteome includes 1032 CHPs⁹⁸ today, despite intensive efforts to close this knowledge gap.

1.4 A novel application of fragment-based methods for natural substrate prediction

A major objective of this doctoral work is to develop an experimental approach to rapidly generate and test hypotheses of CHP function in *Mtb* using high-throughput biophysical screening technologies. The ‘biological function’ of a protein can be difficult to define, as there are several means of describing the functional properties of proteins. Herein, CHP function is defined in terms of its molecular recognition properties, which can reveal important information about substrate specificity. ‘Molecular recognition’ describes chemical phenomena guided by specific, non-covalent interactions⁹⁹. Molecular recognition between proteins and small molecules in the cell is crucial to many important biological processes¹⁰⁰, and it is therefore key to understanding protein function.

This thesis aims to explore the applicability of fragment-based approaches (described in Section 1.2.2.2) to interrogate CHP function, rather than to discover drug-like molecules. Fragments may be effective chemical tools to investigate enzyme substrate specificity. They are able to make efficient, ‘high-quality’ binding interactions because they contribute significantly to favourable binding energy, even though they are weak binders by virtue of their size¹⁰¹. It has been shown that fragment-enzyme interactions can recapitulate the binding of natural substrates¹⁰². Fragments can identify the ‘hot spots’ of a protein, which contribute most to the binding energy of a ‘fragmented’ natural ligand¹⁰³. Further, fragments have been discovered that mimic structural features of natural ligands^{88,104}. This approach is further described in Chapter 5.

This novel application of fragment-based methods, along with structural and enzymatic approaches, is applied to a family of enzymes annotated in the

Mtb genome as pyridoxine 5'-phosphate oxidase (PNPOx)-like proteins with unknown function. This thesis describes the characterisation of three enzymes within this class: Rv2607, Rv1155, and Rv2991. Bioinformatic studies of PNPOx-like proteins are described in Chapter 2. This analysis led to the hypothesis that Rv2607 is a true PNPOx, while Rv1155 and Rv2991 are structurally-related to, but functionally distinct from, other known PNPOxs. Biophysical and enzymatic characterisation of Rv2607 is described in Chapter 3 and that of Rv1155 and Rv2991 is discussed in Chapter 4. Chapter 5 describes work done to date to screen Rv1155 and Rv2991 with a targeted fragment library to identify reactions catalysed by these enzymes.

Chapter 2

Analysis of pyridoxine 5'-phosphate oxidase-like proteins in *M. tuberculosis* using sequence- and structure-based comparisons

2.1 Background

The *Mtb* genome encodes a class of related CHPs annotated as pyridoxine 5'-phosphate oxidase (PNPOx)-like proteins with unknown function. This class includes Rv2607, Rv1155, Rv2991, Rv2074, Rv0121c, Rv1875, and Rv3369. They belong to superfamily *cI00381* in the National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD). This superfamily is composed of hundreds of PNPOxs and other flavin mononucleotide (FMN) binding proteins. The seven *Mtb* proteins are related to each other by a shared domain called 'pyridoxine 5'-phosphate oxidase domain' (PF01243), as determined by the Pfam algorithm¹⁰⁵. Although PLP is essential for *Mtb* growth

and virulence *in vivo*¹⁰⁶, the PNPOx-like proteins are non-essential for *Mtb* survival according to Himar1-based transposon mutagenesis experiments in H37Rv^{21,105}.

In organisms as disparate as *E. coli* and *H. sapiens*, there is only one gene for PNPOx^{106,107}. It is an FMN-dependent enzyme encoded by the gene *pdxH* that oxidises either pyridoxine 5'-phosphate (**1**), a primary alcohol, or pyridoxamine 5'-phosphate (**3**), a primary amine, into pyridoxal 5'-phosphate (**2**)¹⁰⁸ (Figure 2.1). Pyridoxal 5'-phosphate (PLP), the active form of vitamin B₆, is one of the most ubiquitous and versatile cofactors in the cell¹⁰⁹. For detailed information about this reaction and its relation to other biosynthetic pathways, please refer to Chapter 3, Figure 3.1.

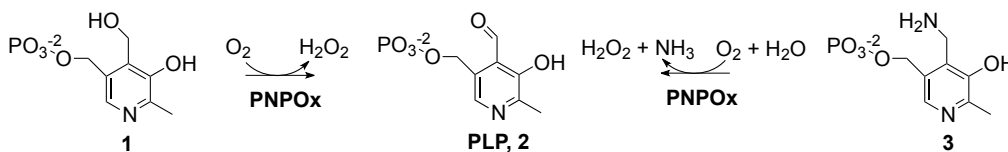


Figure 2.1. Schematic of PNPOx-catalysed PLP formation. PNPOx catalyses the production of PLP, the bioactive form of vitamin B₆. PNPOx is capable of oxidising either a primary amine or a primary alcohol to an aldehyde in the presence of its cofactor, FMN. Molecular oxygen is the terminal electron acceptor of this reaction.

Although this enzyme catalyses the same reaction in both *E. coli* and in humans, it serves a different metabolic function in each organism. In *E. coli*, PNPOx participates in *de novo* PLP biosynthesis¹⁰⁸. Humans lack the biosynthetic machinery to produce PLP *de novo* and PNPOx therefore functions to salvage PLP¹⁰⁷. There are two classes of organisms capable of *de novo* PLP biosynthesis: those that produce PLP via the deoxyxylulose (DXP)-dependent pathway and those that synthesise PLP via the DXP-independent pathway¹¹⁰. In the DXP-dependent pathway, PNPOx directly participates in PLP formation by catalysing the final step of the PLP biosynthesis¹¹¹. Few organisms have the DXP-dependent pathway, including *E. coli* and some γ -proteobacteria. The majority of PLP-synthesising organisms (most bacteria, archaea, fungi, and

plants) use the DXP-independent pathway, in which PLP is synthesised by a single multienzyme complex called PLP synthase¹¹². More in-depth information on DXP-dependent and independent pathways is provided in Chapter 3 (Figure 3.1). Organisms that synthesise PLP via PLP synthase are thought to use PNPOx for PLP salvage. However, bioinformatic analysis described in this thesis shows that some PLP-biosynthesising organisms with genes encoding PLP synthase lack genes for PNPOx (Table 2.1). This observation suggests that in organisms with the DXP-independent pathway, PNPOx is not required to maintain adequate PLP levels in the cell.

The genome of *Mtb* includes homologues of PLP synthase and PNPOx. However, the *Mtb* genome is unusual amongst organisms using the DXP-independent pathway because it has seven genes encoding copies of PNPOx homologues. Because PNPOx activity is not required for PLP production in organisms expressing PLP synthase, the number of PNPOx homologues in *Mtb* is particularly surprising and unexpected. This suggests that these PNPOx-like proteins in *Mtb* may serve diverse, yet unknown functions not related to PLP metabolism. This chapter describes efforts to generate hypotheses on alternate functions of PNPOx-like proteins in *Mtb* by mining available genomic, proteomic, and structural information.

2.2 Analysis of PNPOx-like proteins using comparative genomics

2.2.1. Protein sequence alignments

The NCBI protein basic local alignment sequence tool (BLASTp) was used to search for homologues of the *Mtb* pfam01243 PNPOx-like proteins in other organisms. *Mycobacterium smegmatis*, a close relative of *Mtb*, encodes homologues to all of the PNPOx-like proteins found in *Mtb*, except Rv3369 (Table 2.1). These proteins are found in other mycobacteria, including *M. ulcerans*, *M. avium*, and *M. gilvum*. The vastly reduced genome of *M. leprae*,

which is incapable of surviving outside of the host organism, encodes homologues of Rv1155 and Rv2607 only¹¹³. Another PLP-synthesising organism, *Streptomyces coelicolor*, has homologues of four of the seven *Mtb* PNPOx-like proteins. These homologues are annotated as PNPOx-like CHPs with unknown function in the genomes examined.

<i>Mtb</i>	<i>M. smegmatis</i>	% ID	<i>M. leprae</i>	% ID	<i>S. coelicolor</i>	% ID
Rv2607	MSMEG_5675	69	ML2131	67	SCO4387	44
Rv1155	MSMEG_5170	80	ML1508	88	SCO5312	57
Rv2991	MSMEG_0048	73	—	—	SCO1357	38
Rv2074	MSMEG_3880	85	—	—	SCO0838	53
Rv0121c	MSMEG_6526	59	—	—	—	—
Rv1875	MSMEG_6576	35	—	—	—	—
Rv3369	—	—	—	—	—	—

Table 2.1. Summary of NCBI gene product annotations for PNPOx-like proteins in selected organisms where PNPOx activity is considered to play a role in PLP salvage. Homologues were assigned using BLASTp¹¹⁴ for sequence alignments with E-values less than 10^{-6} . Percent identities (% ID) between the *Mtb* PNPOx-like genes and their respective homologues in *M. smegmatis*, *M. leprae*, and *S. coelicolor* are shown to the right of the gene annotation. Percent identity has been calculated by dividing the number of identical residues by the length of the alignment using pBLAST ALIGN (www.blast.ncbi.nih.gov).

Sequence alignments of PNPOx-like proteins in *Mtb* reveal relatively low sequence similarity within the group (Table 2.2). Among the PNPOx-like family members that share significant sequence similarity (a pair having an E value of less than $1e-05$), the sequence identity is around 30%. There is no statistically-significant sequence similarity between Rv2607 and any of the other PNPOx-like proteins.

	Rv2607	Rv1155	Rv2991	Rv2074	Rv0121c	Rv1875	Rv3369
Rv2607	100	–	–	–	–	–	–
Rv1155	–	100	33	24	31	30	27
Rv2991	–	33	100	24	28	–	–
Rv2074	–	24	24	100	–	–	–
Rv0121c	–	31	28	–	100	–	–
Rv1875	–	30	–	–	–	100	–
Rv3369	–	27	–	–	–	–	100

Table 2.2. Pairwise percent identities of PNPOx-like proteins. Percent identity has been calculated by dividing the number of identical residues by the length of the alignment using pBLAST ALIGN (www.blast.ncbi.nih.gov). Where the percent identity is replaced by a dash, the E value was too high to align the two sequences (E value > 1e-05).

2.2.2 Gene clusters and gene synteny analysis

Gene clusters and gene synteny were analysed for clues on PNPOx-like protein function. A gene cluster is a set of genes that encodes proteins that participate in the production of the same, similar, or related products¹¹⁵. Gene clusters are created by gene duplication and divergence. The duplicated genes evolve and acquire new but related function, and maintain physical proximity on the chromosome¹¹⁵. Comparative gene synteny, as applied here, is an approach to generating hypotheses about gene function, which stipulates that two or more gene clusters may have derived from a single, ancestral genomic region that is present in other organisms. Evolutionarily-related organisms have similar blocks of genes, perhaps in similar relative locations in the genome¹¹⁶.

The genetic regions surrounding PNPOx-like genes were examined in *Mtb* and *M. smegmatis*. Schematics of each gene region in *Mtb* and the analogous region in *M. smegmatis* are shown in Figure 2.2. Gene *rv2607* is upstream of *rv2606c* and *rv2604c*, homologues of PLP biosynthetic genes, *pdx1* and *pdx2* (Figure 2.2A). Pdx1 and Pdx2 (twelve molecules of each) associate to form the PLP synthase complex. The spatial proximity of *rv2607* to *pdx1/pdx2* in the *Mtb* genome suggests that Rv2607 may be involved in PLP metabolism.

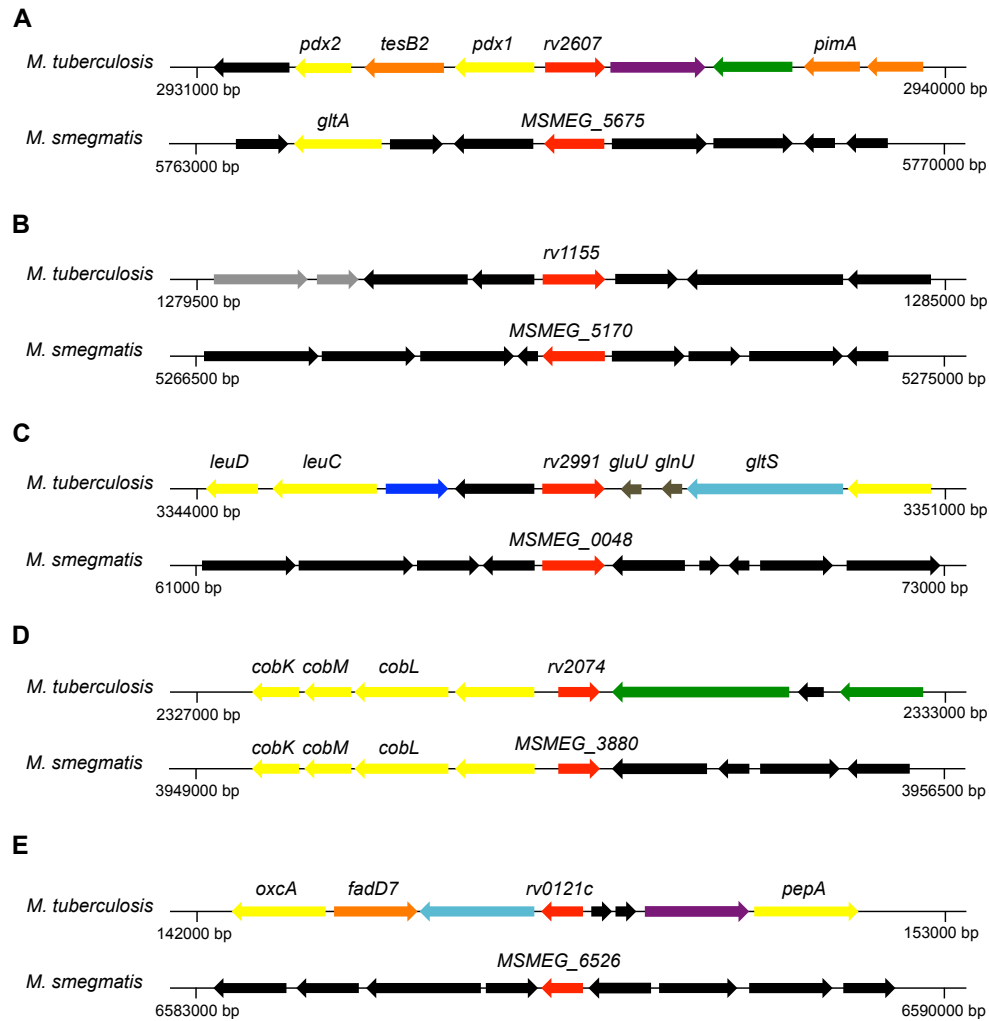


Figure 2.2. Gene cluster comparisons of PNPOx-like proteins in *Mtb* and *M. smegmatis*. In each panel, genes are coloured according to their function: PNPOx-like proteins in *Mtb* and homologues in *M. smegmatis* (red), intermediary metabolism (yellow), lipid metabolism (orange), information pathways (cyan), RNA (brown), regulation (grey), PE/PPE family of proteins (purple), membrane proteins (green), and unknown proteins (black). Gene region information was obtained from the databases TubercuList⁹⁸ v2.6 (<http://tuberculist.epfl.ch/index.html>) and SmegaList v1.0 (<http://mycobrowser.epfl.ch/smegmalist.html>). Gene regions surrounding *Mtb* PNPOx-like proteins Rv2607 (A), Rv1155 (B), Rv2991 (C), Rv2074 (D), and Rv0121c (E) and the respective homologues in the genome of *M. smegmatis*. In the *Mtb* genome, the gene encoding Rv2607 is adjacent to *pdx1* and *pdx2*, proteins that form the PLP synthase. Other annotated gene products near the PNPOx-like genes include: *tesB2* (acyl-CoA thioesterase II), *pimA* (alpha-mannosyltransferase), *gltA* (citrate synthase), *leuC/D* (3-isopropylmalate dehydratase), *gluU* (tRNA-glutamate), *glnU* (tRNA-glutamine), *gltS* (glutamyl-tRNA synthase), *cobK* (precorrin-6X reductase), *cobM* (precorrin-3 methylase), *cobL* (precorrin-6Y C(5,15)-methyltransferase), *oxaA* (oxalyl-CoA decarboxylase), *fadD7* (fatty-acid-CoA ligase), and *pepA* (serine protease).

M. smegmatis homologues of *pdx1* (MSMEG_2937), *pdx2* (MSMEG_2939), and *rv2607* (MSMEG_5675) are not found in the same position in the genome, relative to the *Mtb* proteins. In both the genomes of *Mtb* and *M. smegmatis*, the gene encoding Rv2074 is upstream of a series of enzymes involved in cobalamin biosynthesis (*cobK*, *cobM*, and *cobL*). The gene regions surrounding Rv1155, Rv2991, and Rv0121c in *Mtb* or *M. smegmatis* do not provide much useful information, as many are also CHPs or proteins of unknown function.

2.3 Structural analysis of PNPOx-like proteins

X-ray structures of unliganded Rv2607¹¹⁷, Rv2991, Rv1155¹¹⁸, and Rv2074¹¹⁹ have been reported. Each protein structure was annotated as a PNPOx-like protein with unknown function in the Protein Databank (PDB). The structures of the canonical PNPOxs of *E. coli*¹²⁰ and human¹²¹ in complex with their ligands, PLP and FMN, have been crystallised and determined to 1.8 Å and 2.0 Å resolution, respectively. PNPOx proteins consist of two domains: a large and a small domain (Figure 2.3). The large domain consists of an anti-parallel beta barrel with Greek key topology, capped on each end by an alpha helix. The small domain is primarily alpha helical. Each protein varies in architecture of the alpha helices that comprise the small domain. In *E.coli* and human PNPOx, the small domain forms the PLP binding site, while the large domain binds FMN. The biologically-active form of PNPOx is the homodimer¹⁰⁸.



Figure 2.3. Structural superimposition of *E. coli* PNPOx and *Mtb* PNPOx-like proteins (protomers). The alignment was performed using the MultiProt Server (<http://bioinfo3d.cs.tau.ac.il/MultiProt/>) with the *E. coli* PNPOx¹⁰⁷ (cyan; PDB ID: 1G77), Rv2607¹¹⁷ (red; PDB ID: 2A2J), Rv1155¹¹⁸ (green; PDB ID: 1W9A), Rv2991 (yellow; PDB ID: 1RFE), and Rv2074¹¹⁹ (purple; PDB ID: 2ASF). The image was rendered using Chimera¹²².

PNPOx-like proteins Rv1155, Rv2074, Rv2991, and Rv2607 share common structural elements. Superimposing all PNPOx-like proteins yields a low C-alpha root-mean-square deviation (RMSD) of 1.43 – 1.71 Å (number of residues: Rv1155, 140; Rv2074, 124; Rv2991, 153; Rv2607, 202). The coordinates of each of these proteins were compared with those of all other structures deposited in the PDB using the distance alignment matrix method (Dali)¹²³(Table 2.2). This technique is the three-dimensional equivalent of a sequence alignment. The algorithm uses a sum-of-pairs method to compare intramolecular distances¹²³. The top result structure shown in Table 2.2 (second column) has the lowest RMSD and best Z score relative to the query molecule (first column) among all other structures deposited in the PDB.

One clear structural difference among the proteins in question is that Rv1155, Rv2991, and Rv2074 contain a similar small domain, while those of Rv2607 (red) and the *E. coli* PNPOx (cyan) are better aligned. This difference is also reflected in the Dali results. Structurally, it appears that Rv1155, Rv2991, and Rv2074 are more similar to each other than they are to either Rv2607 or the *E. coli* PNPOx.

<i>Mtb</i> PNPOx	Top Dali result	RMSD (Å)	No. residues	Z score	% ID
Rv2607	<i>E. coli</i> PNPOx	1.3	199	25.0	41
Rv1155	Rv2074	2.6	125	13.0	24
Rv2991	Rv2074	1.9	125	15.0	24
	<i>B. subtilis</i> BH0577	1.9	126	15.8	19
Rv2074	Rv2991	1.9	160	15.0	24

Table 2.3. Summary of Dali results for PNPOx-like proteins Rv2607 (PDB ID: 2A2J), Rv1155 (PDB ID: 1W9A), Rv2991 (PDB ID: 1RFE), and Rv2074 (PDB ID: 2ASF). The top Dali result has the lowest backbone RMSD relative to the query molecule of all structures deposited in the PDB. The Z score above 2 indicates that the structures have significant structural similarities. All results were obtained using the Dali Server DaliLite¹²³ v.3 (http://ekhidna.biocenter.helsinki.fi/dali_server/start).

There is high sequence similarity amongst PNPOxs across many organisms, particularly in the region of the protein that forms the binding site at the dimer interface. Five sequence motifs are conserved in PNPOxs of various species: -RIVLLK- (residues 67-72 in *E. coli*), -SRKA- (residues 87-90), -QS- (146-147), -EFWQG- (residues 189-193), and -RLHDR- (residues 197-201). Figure 2.4 A shows sequence alignment of two known PNPOxs (*E. coli* and human) and the *Mtb* PNPOx-like proteins. In Figure 2.4, the conserved sequences are coloured and correspond to the residues shown in the structure of *E. coli* PNPOx, which demonstrates that these motifs are situated around the PLP and FMN binding sites. Of the *Mtb* PNPOx-like proteins, only Rv2607 shares the canonical PNPOx sequence motifs; all other PNPOx-like proteins in *Mtb* lack these conserved sequences.

A

```

E. coli   TAMVVATVDEHGQPVQRIVLLKHYDEKG-----MVFYTNLGSRKAHQIENNPRVSLLPFW----- 105
Human    NAMCLATCTRDGKPSARMLLLKGFKDG-----FRPFTNFESRKGKELDSNPFASLVFYW----- 133
Rv2607   NAMVLATVA-DGKPVTRSVLCKILDESG-----VAFFTSYTSKAGQEQLAVTFYASATFPW----- 114
Rv1155   SIGVLATIKHDGRPQLSNVQYHF-DPRKLL-----I-----QVSIABPRAKTRNLRDRPRASILVDA-DDGW--SY 79
Rv2991   RTGTLATIGPDGQPHLTAMWYAV-IDGE-----IWLETKAKSQKAVNLRDRPRVSLFLED-GDTYDTRLRG 86
Rv2074   HLAMLTLRADNSPHVAVGTFP-DPKTHI-----A-----RVITTGGSQKAVNADRSGLAVLSQV--DGA--RW 81
Rv1875   GLAVVSTVRADGTVQASLVNVGL-LPHPVSGE-----PSLGFTTYGKVKLGNLRARFQLAVTFR--NGW--QW 78
Rv3369   LYGWLTTVARSGQVPRLVWFYF-DGTD-----LTVYSMPQAQVAHITAHFPQVSLNLDL----- 78
Rv0121c  PVARLATSTPDGTPHLVPPVFAL-GARRPAEATGADVITYAVDAKRKTQRLRRRLANLEHNPRASVLVDSYADDWTQLWW 92

E. coli   HTLERQVMVICKAERLSTLEVMVKYFHSRPRDSQIGAMVSKQSRRISARGILESKFLE---LQKQFQ---QGE-VPL-PS 176
Human    EPLNRQVRVEGPKVKKLPEEEAECYFHSRPKSSQIGAVVSHQSVIPDREYLRRKNEE---LEQLYQ---DQE-VPK-PK 204
Rv2607   YLGRQAHVQGPVSKVSTEEIPTYSMRPRGALGAWASQSRPVGSRALDNLAE---VTRRFAD--QDQ-IPV-PP 186
Rv1155   AVAEGTAQLTPPA-----AAP-DDDTV-----EALIALYRN---IAGEHSDWDDYR-QAMVTD 127
Rv2991   VSEFQVAEIVEEP-----EA-----LH-----R-----V-G-VSVWER---YTGPHYTDECKPMVDQM-MN 130
Rv2074   LSLEGRAAVNSDI-----DA-----V-----RD-----AELRYAQR-----YR-TPRPNP 115
Rv1875   ATVEGRAQLVGPD-----DP-----RPW-----LVDG-----ERLRLLLREVFTAAGTHDDWEYD-RVMAQE 131
Rv3369   DNGGAGIIVGGTAAVVATDVDCRDDAPY-----WAKYRE---DAAKFGL-TEA-IAA-YS 128
Rv0121c  VRADGVAIHRDG-----EV-----M-----R---AAYRL---LRAKYAQ--YQS-VPL-NG 129

E. coli   FWGGFRVSLQIEFWQGEHLRLHDRFLYQR-----ENDAWKIDRLAP 218
Human    SWGGYVLYPQVMEFWQQTNRHLHDIRVFRRLPTGDSPLGPMTHRGEDWLYERLAP 261
Rv2607   GWGGYRIAPEIVEFWQRENRMHNRIRVAN-----G---RRLERLQP 224
Rv1155   RRVLLT--LPISHVYGLPP---GM-----R 147
Rv2991   KRVGVRIVARTRRSWDHRKLGPHMSVGG-----STAP 163
Rv2074   RRVVIE--VQIERVLGSAD---LLDR-----A 137
Rv1875   QRAVVL--ITPTRIYSN-----G 147
Rv3369   TR--LK--ITPTRVWTTPT-----G 144
Rv0121c  P--VIA--IAVQRWA-SWH-----A 144

```

B

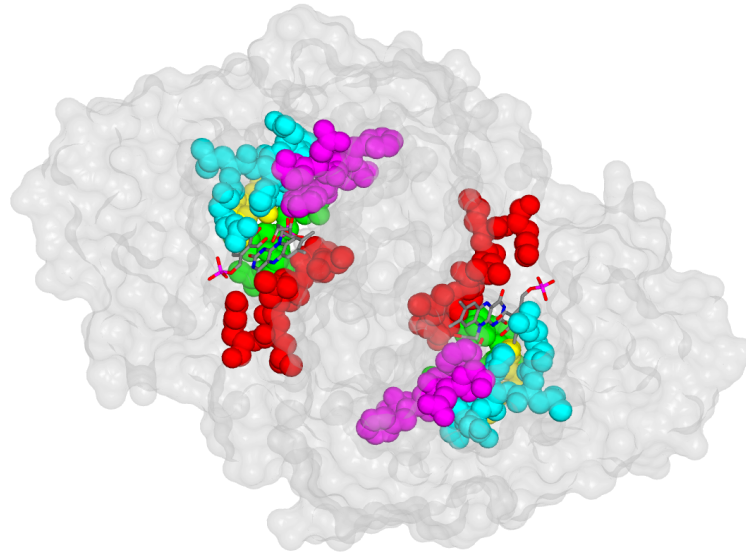


Figure 2.4. Sequence alignment of *E. coli* PNPOx, *H. sapiens* PNPOx, and *Mtb* PNPOx-like proteins. (A) Sequence alignment of *E. coli* PNPOx, human PNPOx, and PNPOx-like proteins from *Mtb* (Rv2607, Rv1155, Rv2991, Rv2074, Rv1875, Rv3369, and Rv0121c) with highlighted PNPOx sequence motifs. (B) Surface representation of *E. coli* PNPOx (PDB ID: 1G77) bound to PLP and FMN (C atoms, grey) with residues comprising conserved PNPOx sequence motifs represented as spheres: -RIVLLK- (red), -SRKA- (green), -QS- (yellow), -EFWQG- (purple), -RLHDR- (cyan). The alignment was performed with Toffee¹²⁴ (<http://tcoffee.org.cat/apps/tcoffee/do:regular>) and the structure was created with molecular graphics software, CCP4mg¹²⁵.

A detailed analysis of the residues that form the binding pocket of each PNPOx-like protein in *Mtb* is crucial for formulating hypotheses about enzymatic activity. Figure 2.5 shows the *E. coli* PNPOx active site with bound PLP and FMN¹²⁰. The FMN is engaged in a hydrogen-bonding network, making interactions with residues Arg 67 (A), Ile 68 (A), Leu 70 (A), Tyr 82 (A), Thr 83 (A), Arg 88 (B), Lys 89 (B), Gln 111 (B), Gln 146 (B), Ser 147 (B), Arg 201 (B). The PLP also makes extensive contacts with both the protein monomers. Stacking interactions between the PLP and FMN stabilise the binding of both ligands.

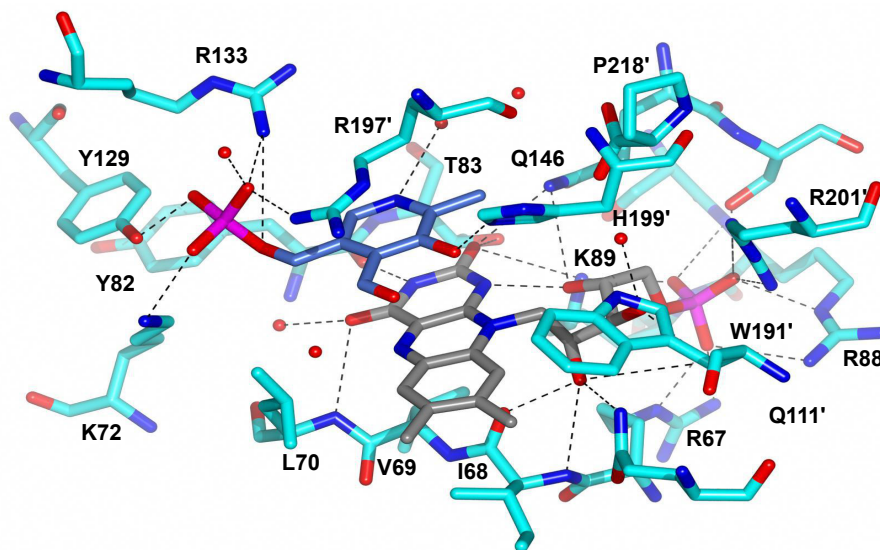


Figure 2.5. Structure of the *E. coli* PNPOx active site. The substrate is PLP (C atoms, blue) and FMN is in grey (PDB ID: 1G77)¹⁰⁸. Hydrogen bonds are represented as dashed black lines. Residues from a second protomer are denoted with an apostrophe. The active site is formed at the dimer interface. The image was created using CCP4mg¹²⁵.

The respective active sites of Rv2607, Rv2074, Rv2991, and Rv1155 are found in the small domain, as predicted by CASTp¹²⁶, a server that identifies structural pockets and cavities on protein surfaces (<http://stsfw.bioengr.uic.edu/castp/>). The *E. coli* PNPOx exhibits similar domain arrangement (Figure 2.3), with a validated active site situated in corresponding

position. Figure 2.6 shows structural alignments of the *E. coli* PNPOx with Rv2607 (A), Rv2074 (B), Rv2991 (C), and Rv1155 (D), respectively. All structures of the *Mtb* PNPOx-like proteins are unliganded, while the *E. coli* PNPOx structure includes both PLP and FMN in the active site. Residues in the *E. coli* protein that interact with PLP and FMN have been compared to spatially analogous residues in the *Mtb* PNPOx-like proteins. Polar contacts were analysed between the ligands (PLP and FMN) and any surrounding atoms, excluding the solvent and including both the *E. coli* and *Mtb* proteins. Residues in analogous positions relative to PLP or FMN in the *E. coli* and *Mtb* proteins are defined as having a 'conserved interaction' with that ligand.

Overall, the Rv2607 binding site structure is very similar to that of the *E. coli* PNPOx. Almost all the ligand-protein interactions seen in the *E. coli* PNPOx are conserved in Rv2607 (Figure 2.6A). There is also a high degree of residue type conservation, with some exception in the residues that form backbone interactions with the FMN, e.g. Ser77 and Phe91 in Rv2607. Pédelacq *et al.* report observing some density in the FMN binding site of Rv2607, but could not definitively place the cofactor¹¹⁷. It appears that Rv2607 is an unliganded PNPOx structure, as Gln155 and Ser 156 (*E. coli* Gln146 and Ser147) are disengaged from interaction with the phosphate moiety of FMN and Tyr138 (*E. coli* PNPOx Tyr129) assumes a conformation that prevents it from forming a hydrogen bond with the phosphate of PLP (Figure 2.6 A).

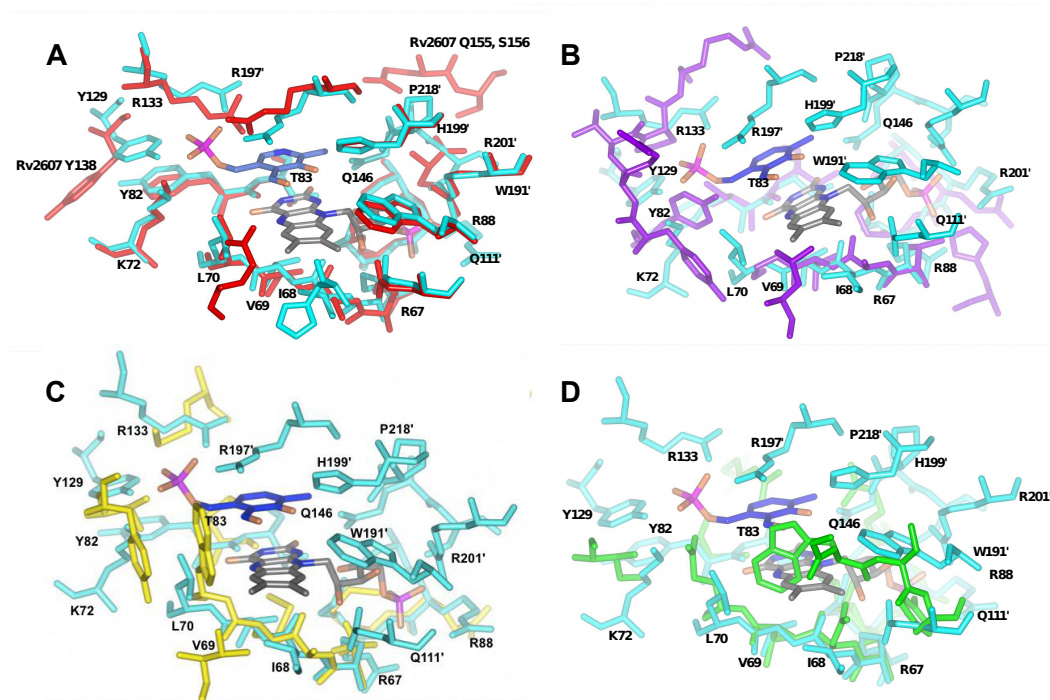


Figure 2.6. Structural superimpositions of *E. coli* PNPOx active site with predicted active sites in *Mtb* PNPOx-like proteins. Structural alignment of *E. coli* PNPOx, cyan, PDB ID: 1G77, and (A) Rv2607, red, PDB ID: 2A2J; (B) Rv2074, purple, PDB ID: 2ASF; (C) Rv2991, yellow, PDB ID: 1RFE; (D) Rv1155, green, PDB ID: 1W9A. Labeled residues correspond to the *E. coli* PNPOx structure unless otherwise noted. Residues shown in each structure are 4.0 Å away from PLP (C atoms, blue) and FMN (C atoms, grey). PLP and FMN shown are in the *E. coli* PNPOx structure active site only. The images were rendered using CCP4mg¹²⁵.

The structural overlays of *E. coli* PNPOx with Rv2704 (Figure 2.6B), Rv2991 (Figure 2.6 C) and Rv1155 (Figure 2.6 D) reveal that there is little evidence that PLP binding could take place in the *Mtb* proteins. In Rv2074, only one of the five residues involved in the PLP-PNPOx interaction is conserved: Tyr104 (Tyr82 in *E. coli* PNPOx). In Rv2991, no residues interacting with PLP in *E. coli* PNPOx are conserved. Only one residue, Gln37, seems to be spatially positioned within the Rv1155 active site to interact with PLP.

Some structural evidence suggests that Rv2074, Rv2991, and Rv1155 could interact with their predicted cofactor, FMN. A structural superimposition of Rv2074 and the *E. coli* PNPOx reveals that Rv2074 lacks many of the residues that interact with FMN in the *E. coli* protein. However, Lys61 in Rv2074 (Lys89 in

E. coli PNPOx) could make interactions with the ribityl moiety of FMN and Thr55 (Thr83 in *E. coli* PNPOx) could hydrogen bond to the pyrimidine portion of the isoalloxazine ring system of FMN, as the analogous residues do in the *E. coli* PNPOx active site. Further, several main chain interactions with the cofactor are conserved in Rv2074. These backbone-FMN interactions are also conserved in the predicted Rv2991 active site, along with Lys61 and Thr55, which are analogous residues to Lys89 and Thr83 in the *E. coli* structure. These conserved FMN binding interactions are also observed in the Rv1155 active site, which is comprised of residues available to engage in main chain interactions with FMN (as in Rv2074 and Rv1155). Lys 57 in Rv1155 is found in the analogous position as *E. coli* Lys89, Rv2074 Lys61, and Rv2991 Lys61. Figure 2.7 shows all the conserved residues in Rv2074, Rv2991, and Rv1155 that could engage in hydrogen bonding interactions with FMN.

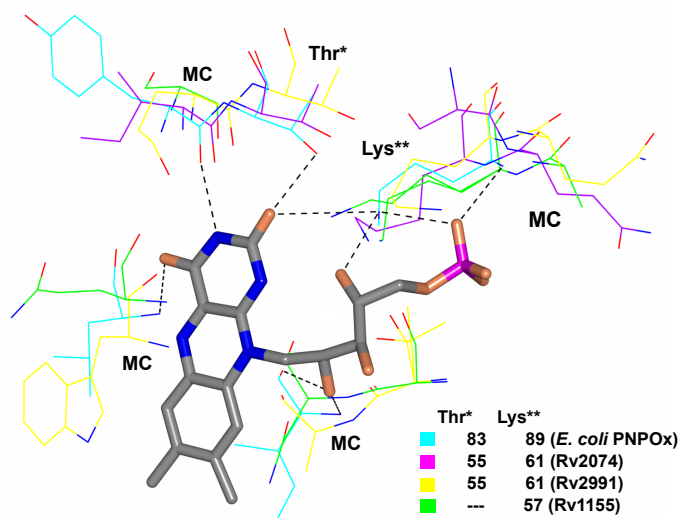


Figure 2.7. Residues in Rv2074, Rv2991, and Rv1155 that could engage in hydrogen bonding interactions with FMN. Superimposition of conserved residues interacting with FMN in *E. coli* PNPOx (cyan, PDB ID: 1G77), Rv2074 (purple, PDB ID: 2ASF), Rv2991 (yellow, PDB ID: 1RFE), and Rv1155 (green, PDB ID: 1W9A). FMN is present in the *E. coli* PNPOx structure only. Hydrogen bonds are represented with dashed lines between FMN and *E. coli* PNPOx. Conserved residues Thr and Lys in Rv2074, Rv2991, and Rv1155 could anchor the ribityl and pyrimidine of FMN, as they do in *E. coli* PNPOx. Several spatially analogous residues in Rv2074, Rv1155, and Rv2991 that could interact with FMN are not conserved, but the cofactor could make contacts with the main chain of each protein, as is seen in the *E. coli* PNPOx (denoted as 'MC'). The image was rendered using CCP4mg¹²⁵.

Despite that there are few interactions between Rv1155 and PLP, Biswal *et al.* have reported structures of Rv1155 with bound PLP and FMN in separate crystals¹²⁷. The structural overlay of Rv1155-FMN and Rv1155-PLP (Figure 2.8) indicates that the binding modes of FMN and PLP in Rv1155 are very different with respect to the *E. coli* PNPOx active site (Figure 2.5). The Rv1155-FMN/Rv1155-PLP overlay shows that simultaneous binding of FMN and PLP seems improbable in Rv1155, as FMN and PLP are clashing and are not properly positioned for enzymatic reaction. The *E. coli* PNPOx binding site, on the other hand, is arranged for electron transfer from the substrate to the FMN. These findings suggest that the Rv1155-PLP structure may not be reflective of the biologically-relevant substrates of Rv1155. Rv1155-FMN may be biologically relevant, although FMN in Rv1155 has a different binding mode than it does in the *E. coli* PNPOx and other related proteins.

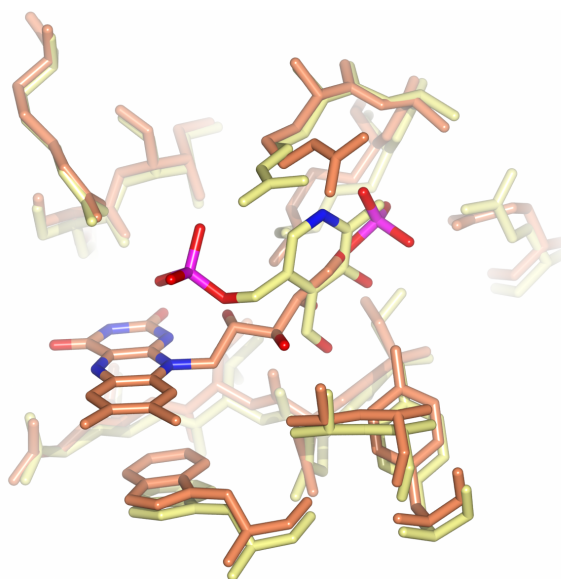


Figure 2.8. Structural superimposition of Rv1155-FMN and Rv1155-PLP. Rv1155-FMN (PDB ID; 1Y30) is shown in orange and Rv1155-PLP (PDB ID: 2AQ6) is shown in yellow¹²⁷. The spatial orientation of FMN and PLP in the proposed Rv1155 active site indicates that these ligand-bound structures are not biologically relevant.

Structural analysis of each PNPOx-like protein reveals that the active sites of Rv1155, Rv2991, and Rv2074 differ significantly from those of known PNPOxs.

Generally, the *Mtb* PNPOx-like proteins have fewer conserved interactions with FMN and PLP. These observations suggest that Rv1155, Rv2991, and Rv2074 may not bind PLP or FMN at all. Rv2607, however, seems to have the highest structural and sequence similarity with the *E. coli* PNPOx. Rv2607 is also the most highly conserved in this protein family across other mycobacterial species, including in the reduced genome of *M. leprae*, and it has the highest homology to the PNPOxs of *E. coli*, *Streptomyces coelicolor* and *Saccharomyces cerevisiae*. Further, *rv2607* is in close proximity to genes encoding PLP synthase in the *Mtb* genome. It is, therefore, the most logical candidate for the canonical PNPOx of *Mtb*. The PNPOx activity of Rv2607 is further explored in Chapter 3.

Chapter 3

Biophysical and enzymatic characterisation of Rv2607 as a pyridoxine 5'-phosphate oxidase

3.1 Background

The term 'vitamin B6' collectively refers to pyridoxine 5'-phosphate (PNP) (1), pyridoxal 5'-phosphate (PLP) (2), pyridoxamine 5'-phosphate (PMP) (3), and their respective non-phosphorylated forms¹²⁸. The bioactive form of vitamin B₆, PLP, acts as a cofactor to over 140 known enzymatic reactions referenced in the Enzyme Commission (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>), which corresponds to about 4% of all classified enzymatic activities. It has been described as one of the most versatile cofactors in nature, as it participates in transamination, decarboxylation, racemisation, C α -C β bond cleavage, and α,β -elimination reactions¹⁰⁹.

A number of enzymes involved in vitamin B₆ metabolic pathways in *Mtb* remain uncharacterised, despite the important role this vitamin plays in the cellular metabolism of this pathogen. PLP biosynthesis is required for *Mtb* survival and virulence in vivo and it is a predicted cofactor for a number of enzymes essential to *Mtb* growth¹⁰⁶. Therefore, vitamin B6 biosynthesis and salvage may be attractive target pathways for the development of novel compounds with anti-TB activity.

In both prokaryotic and eukaryotic organisms, PNPOx is an FMN-dependent enzyme encoded by the gene *pdxH* that produces PLP by oxidising either PNP or PMP¹⁰⁸. In *E. coli* and some γ -proteobacteria, PNPOx participates in the deoxyxylulose 5'-phosphate (DXP)-dependent PLP biosynthetic pathway¹¹¹. The DXP-dependent pathway consists of 7 enzymes: GapA, PdxB, PdxF, PdxA, DXPS, PdxJ and PNPOx (Figure 3.1, left). Together, PdxA and PdxJ catalyse the formation of PNP (**1**) from DXP and 1-hydroxy-3-aminopropane phosphate. *E. coli* PNPOx then oxidises PNP into the bioactive product, PLP. PNPOx in *E. coli* can also recognise PMP as a substrate, although this reaction is not part of the *de novo* PLP biosynthetic pathway and is a vitamin B₆ salvage reaction.

Most prokaryotes and plants biosynthesise PLP via an alternative route called the DXP-independent pathway¹¹² (Figure 3.1, right). The DXP-independent pathway includes PLP synthase, a macromolecular complex consisting of enzymes encoded by the genes *pdx1* and *pdx2*¹¹⁰. This complex is capable of carrying out *de novo* PLP biosynthesis directly. Pdx2 is a glutaminase, which produces ammonia by hydrolysing glutamine, and Pdx1 is a synthase, which condenses ammonia, ribulose 5-phosphate, and glyceraldehyde-3-phosphate. X-ray crystal structures of PLP synthase^{129,130} reveal twelve Pdx1 units are arranged in two hexameric rings to which twelve Pdx2 subunits attach. The complex is described as having a 'cogwheel-like' configuration, in which Pdx1 forms the core of the wheel and the individual Pdx2 units are the cogs¹¹⁰. PLP synthase has been identified as a possible drug target for malaria^{131,132}.

The *Mtb* genome encodes homologues of proteins involved in the DXP-independent PLP biosynthetic pathway, *pxd1* (Rv2606c) and *pxd2* (Rv2604c), but lacks the genes for enzymes in the DXP-dependent pathway¹⁸ (please refer to Chapter 2, Figure 2.2 A for a schematic of the region surrounding Rv2607 in the *Mtb* genome). It has been demonstrated that *Mtb* synthesises PLP via the DXP-independent pathway using PLP synthase and that disruption of the *pxd1* gene generates a vitamin B₆ auxotrophic *Mtb* mutant¹⁰⁶. Organisms with genes that encode both PLP synthase and PNPOx probably only use the oxidase to salvage PLP after it participates as a cofactor in enzymatic reactions¹⁰⁷. It is unclear whether *Mtb* can acquire PLP or other forms of vitamin B₆ from the host¹⁰⁶.

Of the seven PNPOx-like proteins in *Mtb*, Rv2607 is the most logical candidate for the canonical PNPOx (please refer to Chapter 2 for a detailed analysis). Rv2607 has highest sequence similarity to known PNPOxs (Table 2.1). The predicted active site of Rv2607 is similar to that of PNPOx homologues in other organisms (Figure 2.6 A). Residues known to be important for PLP binding in the *E. coli* and human PNPOx are nearly all conserved in Rv2607 (Figure 2.4 A). FMN could not be definitively placed in the *Mtb* PNPOx active site because the electron density for the FMN cofactor is weak in the Rv2607 crystal structure. However, residues known to interact with FMN in the *E. coli* and human PNPOx enzymes are conserved in Rv2607. Gene expression for *rv2607* has been detected in over 300 different experimental conditions in vitro¹³³ (www.tbdb.org/expressionHistory.shtml) as well as in macrophages¹³⁴. Expression of this gene is upregulated by stresses such as exposure of *Mtb* to certain inhibitors of translation and by nitroimidazoles, such as PA-824¹³³.

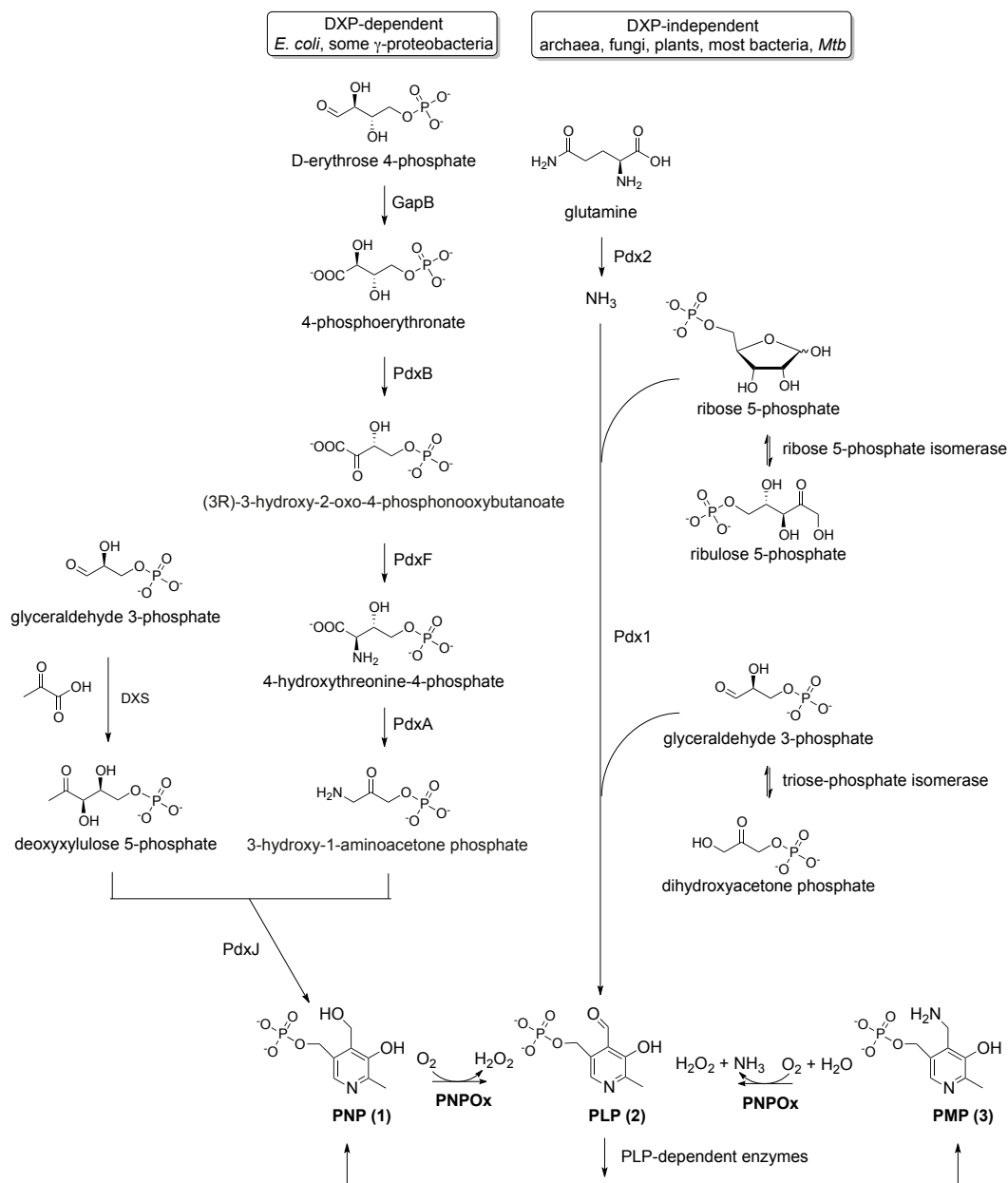


Figure 3.1. PNPOx-catalysed PLP formation in DXP-dependent and DXP-independent PLP biosynthetic pathways. In *E. coli*, PNPOx catalyses the last step in the DXP-dependent PLP biosynthetic pathway¹¹¹. Most organisms capable of PLP biosynthesis produce PLP via PLP synthase, a macromolecular complex consisting of Pdx1 and Pdx2¹¹⁰. Organisms with genes that encode both PLP synthase and PNPOx likely use PNPOx to salvage PLP from PNP and PMP¹⁰⁷, which are produced by enzymes that use PLP as a cofactor. These two distinct PLP biosynthetic pathways have not yet been found to coexist in the same organism¹³⁵.

3.2 Protein expression and purification of Rv2607

Rv2607 was obtained as an aminoterminal 6x-histidine (His_{6x}) tag fusion construct. A one-step nickel affinity purification procedure was employed to isolate Rv2607 and it was freshly purified before each assay. The proper folding of Rv2607 was demonstrated using circular dichroism (CD) spectroscopy. In its native state, Rv2607 gives a CD signal that is expected from proteins with mixed anti-parallel β -barrel and α -helical secondary structure. This CD signal disappears when the protein is denatured with 6 M guanidium hydrochloride for 3 hrs at 25 °C.

SDS-PAGE analysis showed that the protein was purified to >95% homogeneity and the yield of the purified protein was 140 mg per liter of cells, as determined using the Bradford test¹³⁶, calibrated using bovine serum albumin (BSA) standards. Two major protein bands were observed by SDS-PAGE, one corresponding to the Rv2607 monomer (~27.5 kDa) and the other corresponding to a homodimer (~55 kDa). The dimeric form of the protein appeared to be stable under reducing and denaturing conditions in SDS-PAGE (3% β -mercaptoethanol, 0.1 % SDS). To confirm the identity of the band corresponding to 55 kDa, it was analysed by MS, which is described in the following section.

3.3 Analysis of Rv2607 using mass spectrometry

Purified Rv2607 was bright yellow in colour, suggesting the presence of a co-purified cofactor. The yellow cofactor bound tightly to Rv2607 as purified and was unambiguously identified as FMN (m/z 457.2 Da) by nanoflow electrospray ionisation mass spectrometry (nESI-MS) after it was isolated from the heat-denatured protein (Figure 3.2).

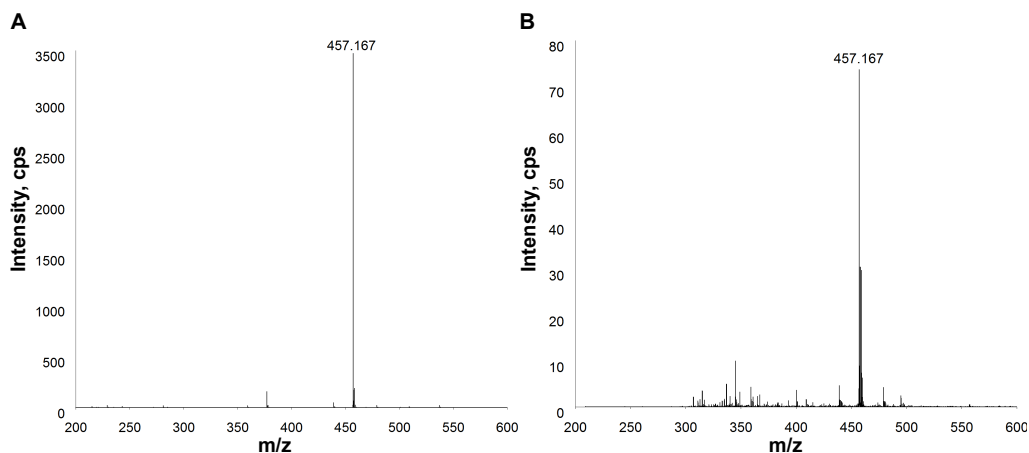


Figure 3.2. Identification of cofactor bound to Rv2607 by nESI-MS. Cofactor was extracted from Rv2607 by heat denaturation, followed by filtration with a 10K Da molecular weight (MW) cutoff filter. FMN has a molecular weight of 456.2 Da. (A) Mass spectrum of commercial standard of FMN. (B) Mass spectrum of cofactor isolated from heat-denatured Rv2607.

The intact Rv2607-FMN complex was analysed by nESI-MS, which revealed it exists in a predominantly dimeric oligomerisation state (Figure 3.3), which is consistent with other known PNPOxs¹⁰⁸. Although two FMN binding sites exist at the homodimer interface of Rv2607, only one FMN molecule per dimer was present in the protein as purified (Figure 3.3). A 2:1 ratio of PNPOx protomer to FMN has been reported for other PNPOxs^{137,138,139}.

In order to investigate the stability of the dimer, the charge state +14 of Rv2607 was isolated in a tandem MS experiment (MS/MS) and submitted to collision-induced dissociation (CID). The MS/MS spectrum (Figure 3.4) shows that dissociation begins to occur at accelerating voltages of 150 V (trap) and 110 V (transfer). High accelerating voltages were required to induce dissociation of this very stable Rv2607 homodimer. Small populations of both monomers are visible at high and low m/z respectively and several additional peaks in the spectrum indicate that such high voltages also cause fragmentation of the protein, which has been previously reported for a different system¹⁴⁰. This result is consistent with reducing SDS-PAGE analysis of the protein, which reveals a stable dimer. The crystal structure of Rv2607 indicates that the two protomers

are very tightly wrapped around each other, as the dimer interface consists of about 20% of the total protomer surface¹¹⁷. This work was done in collaboration with Dr. Dijana Matak-Vinković (Department of Chemistry, University of Cambridge).

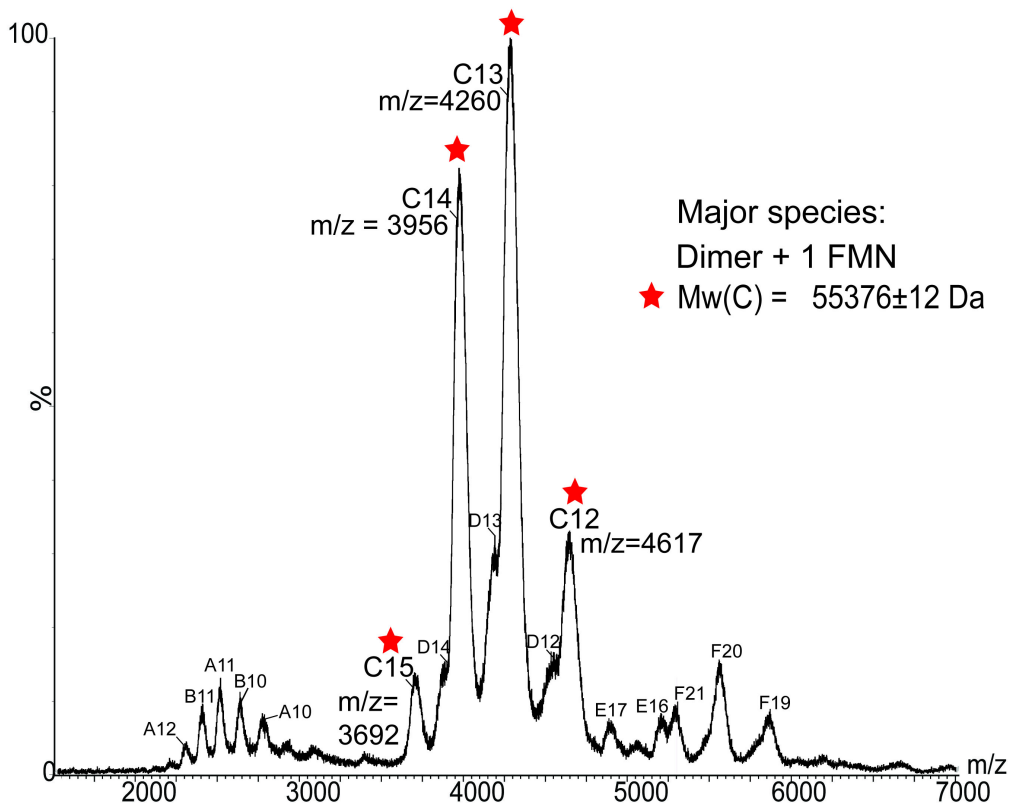


Figure 3.3. Mass spectrum of intact Rv2607 with co-purified FMN analysed by nESI-MS. The major species, C, has peaks with charges ranging from 15+ to 12+ in the 3500-4500 m/z range that correspond to a molecular weight of 55376 ± 12 . This experimental mass is consistent with the calculated mass for the complex of dimeric Rv2607 with one molecule of FMN bound (55442 Da), which is derived from the amino acid sequence of the 6x-histidine tagged Rv2607 (54986 Da; 27493 Da per protomer) and the molecular weight of FMN (456 Da). Monomeric tagged Rv2607 and truncated Rv2607 monomer (missing 10-11 residues), are present in solution and correspond to molecular weights of A (27450 ± 52 Da) and B (26215 ± 54 Da) respectively. A minor dimeric species comprised of full length and truncated monomer is represented as species D. Higher molecular weight oligomers (trimers E and tetramers F with molecular weights of 110.6 and 82.7 kDa, respectively) are electro-spray-induced non-specific association of monomers and dimers.

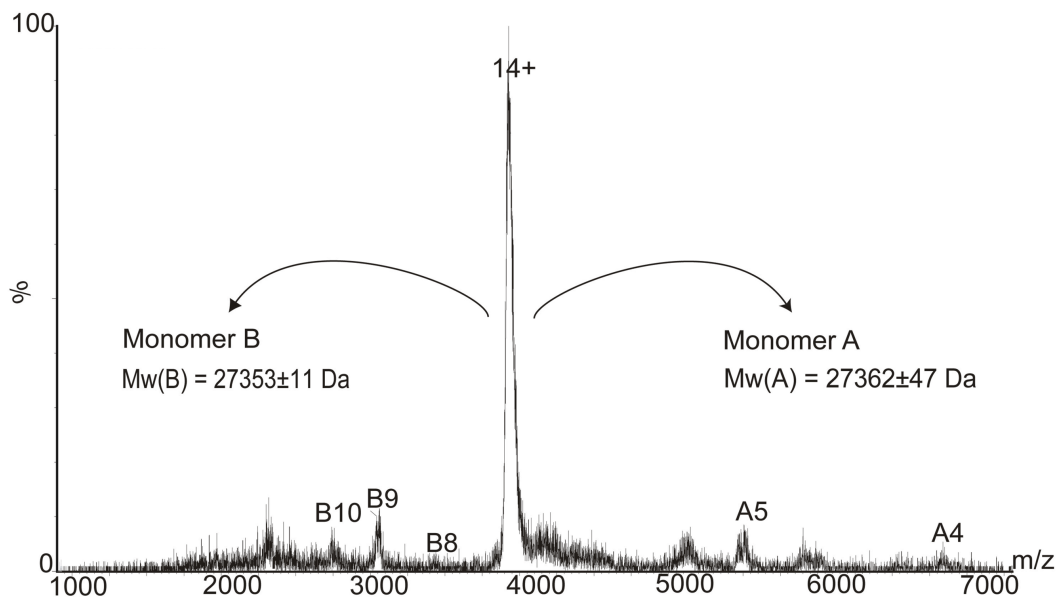


Figure 3.4. MS/MS spectrum of the Rv2607 homodimer. The isolated precursor ion, charge state 14+ of the dimeric Rv2607 complex was submitted to CID with accelerating voltages 150 V (Trap) and 110 V (Transfer). Low populations of charge-state series A and B corresponding to monomers with measured masses 27362 Da (monomer, high m/z region) and 27353 Da (low m/z region) begins to occur at these high voltages. High accelerating voltages also caused fragmentation of the protein, indicated by a few unassigned peaks at the low and high m/z region.

The 55 kDa protein band was unambiguously identified as Rv2607 by isolating the protein by gel filtration and subjecting it to proteolytic digestion. The digested fragments were analysed by liquid chromatography (LC) and MS/MS, assigned amino acid sequence, and queried with the SEQUEST algorithm¹⁴¹. This analysis indicates that the proteolysed fragments belong to the parent molecule, Rv2607.

3.4 Kinetic studies of Rv2607

3.4.1 LC-MS assay to detect Rv2607-catalysed PLP formation

Rv2607-mediated catalysis of PLP formation was examined by LC-MS (Figure 3.5). To resolve the product, PLP, from the substrates, PNP and PMP, PLP was derivatised with 2,4-dinitrophenylhydrazine (DNPH) to form a PLP-DNP

hydrazine. It has been previously reported that PLP readily forms a stable hydrazone with phenylhydrazine¹⁴². DNPH reactivity with PLP, PNP, and PMP was tested (refer to Chapter 6, Figure 6.1 for these control reactions). DNP hydrazones are not formed with either PNP or PMP because they lack a reactive aldehyde.

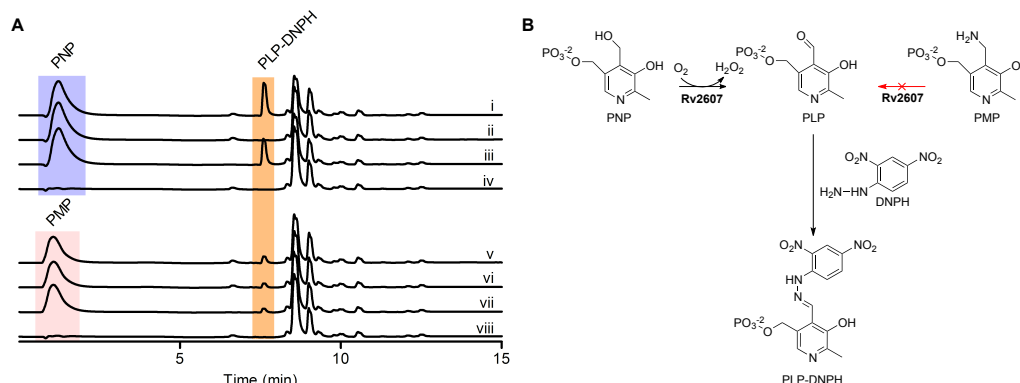


Figure 3.5. Reverse-phase HPLC analysis of the PNPOx activity of Rv2607. (A) HPLC chromatograms (270 nm) of reaction mixtures containing PNP (i-iv) or PMP (v-viii), FMN, and Rv2607, and the associated control reactions. All reactions were carried out in 25 mM potassium phosphate buffer (pH 7.8), incubated at 25 °C for 3 hrs, and quenched with DNPH (0.7 mM final concentration). Where present, the reaction components were at the following concentrations: 1 mM PNP or PMP, 10 μ M FMN, and 10 μ M Rv2607. Reaction mixtures contained: (i) PNP, FMN, and Rv2607, (ii) PNP and FMN (enzyme negative control), (iii) PNP and Rv2607 (no added FMN), (iv) FMN and Rv2607 (substrate negative control), (v) PMP, FMN, and Rv2607, (vi) PMP and FMN (enzyme negative control), (vii) PMP and Rv2607 (no added FMN), (viii) FMN and Rv2607 (substrate negative control). The peak to the right of PLP-DNPH is DNPH, which has a retention time of 8.6 min. (B) Schematic representing assay to identify Rv2607-catalysed PLP formation with DNPH derivitisation.

A peak corresponding to the chemically-synthesised PLP-DNP hydrazone (retention time = 7.60 min) was detected in the enzymatic reaction of Rv2607 with PNP. Electrospray ionisation in the positive mode confirmed it to be the DNP hydrazone of PLP ($m/z = 428$). No such peak was observed in the enzyme and substrate (PNP) negative controls, however, a peak was observed in the absence of added FMN since the enzyme co-purified with FMN. The PLP-DNP hydrazone was not detected when PMP was tested as a substrate (Figure 3.5). To eliminate the possibility that the PLP formation observed was due to the presence of contaminating endogenous *E.coli* PNPOx, *E.coli* BI21(DE3) cell

lysates were tested for PNPOx activity and it was shown that PLP can only be detected when Rv2607 is overexpressed. PNPOx activity assays were also performed using an inactive PNPOx-like His_{6x}-tagged protein expressed and purified in the same manner as was Rv2607. No PLP production was detected. These observations confirm that the observed PNP turnover is solely due to Rv2607. Further, *E. coli* PNPOx demonstrates PMP turnover¹⁰⁸, while Rv2607 does not.

3.4.2 Enzymatic reaction time course using ¹H NMR spectroscopy

Rv2607-catalysed formation of PLP was observed directly by ¹H NMR spectroscopy in the presence of Rv2607 and PNP over time (Figure 3.6). Increasing signals corresponding to the C2-¹H and the aldehyde proton of PLP were observed with time with a concurrent decrease in signal from the C2-¹H of PNP. Under these experimental conditions, there was 75% conversion of the substrate after 18 hours. Spiking the reaction mixture with a synthetic standard of PLP confirmed that the observed chemical shifts in the reaction mixture correspond to PLP (Figure 3.6).

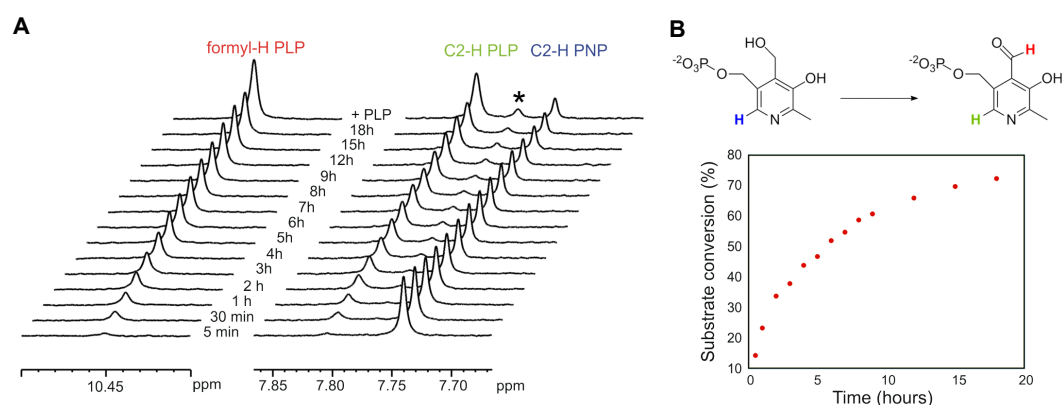


Figure 3.6. ¹H NMR analysis of the conversion of PNP into PLP with time. The enzymatic reaction (1 mM PNP, 28 μM enzyme, 10% D₂O in 25 mM potassium phosphate buffer, pH 7.8) was incubated for 18 hrs at 25 °C in the spectrometer. Left; stacked ¹H NMR spectra recorded at various time points and after the addition of authentic PLP. The starred peak corresponds to PLP hydrate. Right; a plot of percent substrate conversion versus time. Substrate conversion was determined by comparing integrals of the C2-¹H signals associated with PLP (7.80 ppm) to that of PNP (7.75 ppm).

3.4.3 Michaelis-Menten kinetics for Rv2607 and PNP

Rv2607-catalysed PNP oxidation exhibited Michaelis-Menten kinetics with a k_{cat} and K_{M} of 0.01 s^{-1} and $360 \text{ }\mu\text{M}$ respectively (Figure 3.7). The reported k_{cat} and K_{M} for the *E. coli* enzyme¹⁰⁸ and PNP are $0.2\text{-}0.8 \text{ s}^{-1}$ and $2 \text{ }\mu\text{M}$, respectively and the k_{cat} and K_{M} for the human enzyme¹²¹ with PNP are 0.19 s^{-1} and $1.8 \text{ }\mu\text{M}$, respectively. While both the *E. coli* and human PNPOx recognize PMP as a substrate, significant PLP formation was not seen when PMP was tested as a substrate for Rv2607. To our knowledge, Rv2607 is the first example of a PNPOx that does not oxidise PMP to PLP. The biological implication of the kinetic differences between Rv2607 and other PNPOxs is not yet known. It may be that Rv2607 exhibits a relatively low catalytic efficiency because it is not directly involved in the biosynthesis of PLP (as is the *E. coli* PNPOx) and unlike the human PNPOx, it is operating in a cellular environment which possesses the machinery for de novo PLP biosynthesis.

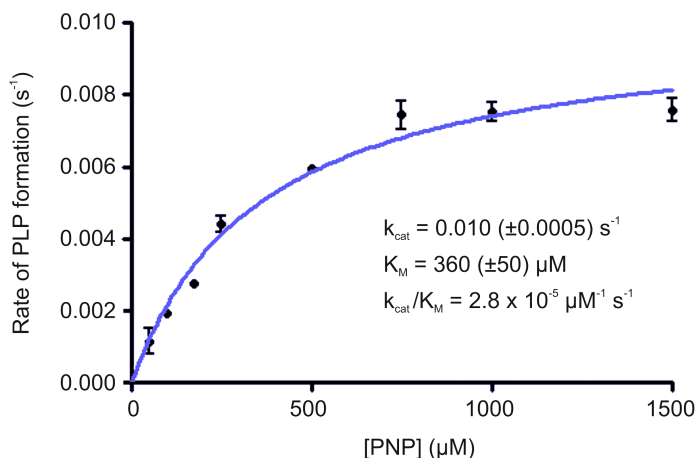


Figure 3.7. Michaelis-Menten plot for Rv2607 with PNP as a substrate. The rate of PLP formation was monitored spectrophotometrically ($\lambda_{\text{max}} = 388 \text{ nm}$, $\epsilon = 4900 \text{ cm}^{-1}\text{M}^{-1}$) for various concentrations of PNP. All solutions were made in 100 mM potassium phosphate buffer, pH 7.8.

3.5 Proposed mechanism

The mechanism of PNP oxidation has not been determined with certainty, although it has been proposed that direct hydride transfer may occur from C4' of the substrate to N5 of FMN¹⁰⁷. This would result in resonance stabilisation of the oxidised product. We propose that for Rv2607, a hydride is transferred from PNP to FMN and oxidised FMN is regenerated with molecular oxygen.

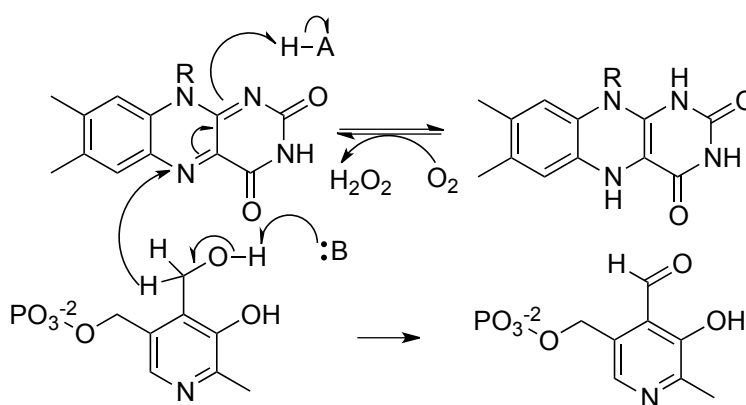


Figure 3.8. A schematic of the proposed mechanism for the oxidation of PNP by Rv2607. A hydride from the substrate, PNP, is transferred to FMN bound to Rv2607. In the Rv2607 active site, His209 could act as a base to deprotonate PNP and Glu59 could act as an acidic group to protonate N1 of FMN. The oxidised cofactor is regenerated in the presence of molecular oxygen.

3.6 Conclusions

Although the *Mtb* protein encoded by Rv2607 was annotated as a member of the PNPOx family based on sequence alignment, biochemical data to support this classification had not been previously reported. PNPOx isolated from prokaryotic and eukaryotic sources (pig¹³⁷ and sheep¹³⁸ brain, recombinant human¹²¹, recombinant *E. coli*¹⁴³ and rabbit liver¹⁴⁴) are reported to catalyse the oxidation of PNP as well as PMP to form PLP. Rv2607 catalyses the oxidation of PNP, but not PMP, in the presence of FMN, which makes it unique within this family of enzymes.

Chapter 4

Identification of Rv2991 and Rv1155 as novel F₄₂₀ binding proteins

4.1 Introduction

Rv2991 and Rv1155 are annotated in the *Mtb* genome as CHPs that are structurally related to the PLP-producing enzyme PNPOx. Bioinformatic analysis indicates that they may be dependent on flavin cofactor FMN (please refer to Chapter 2 for further details). This thesis chapter investigates the cofactor preference of Rv1155 and Rv2991 using biophysical techniques and provides evidence that Rv2991 and Rv1155 do not bind FMN, as predicted, but instead bind to an unusual flavin found in mycobacteria called F₄₂₀.

4.1.1 F_{420} -dependent activation of PA-824, a TB drug currently in clinical trials

Current drug design efforts target biological features that exist in *Mtb*, but that do not exist in human cells. For example, isoniazid¹⁴⁵ and ethambutol¹⁴⁶ are first-line TB drugs that target metabolic pathways unique to mycobacteria. They inhibit proteins involved in the biosynthesis of the mycolic acids, which comprise part of the unique mycobacterial cell wall. This strategy is thought to be effective because it decreases the likelihood of off-target effects in the patient.

Currently in clinical trials is a promising TB candidate drug, PA-824, that requires activation by the mycobacterial nitroreductase Ddn (gene name: *rv3547*)^{25,147}. Ddn is dependent on the flavin coenzyme F_{420} (7,8-didesmethyl-8-hydroxy 5-deazaflavin), which is prevalent in mycobacteria, but is relatively rare in nature and absent in humans. Ddn facilitates hydride transfer from the reduced coenzyme, $F_{420}H_2$, to PA-824, resulting in the release of nitrous oxide, which is quickly converted to nitric oxide (NO) in the cell (Figure 4.1)¹⁴⁷. It has been demonstrated that this intracellular release of reactive nitrogen intermediates (RNI) leads to cell death by inhibiting multiple enzyme and nucleic acid targets within *Mtb*¹⁴⁸. PA-824 is effective for persistent, latent, active, and drug resistant infections. Interestingly, it is well documented that RNI are used as a natural defense system by activated host macrophages against *Mtb*¹⁴⁹. Therefore, PA-824 is the first antibiotic that functions by mimicking the chemical defence strategies of innate immunity¹⁵⁰. The effect of PA-824-mediated RNI release on *Mtb* viability is magnified in part because Ddn- F_{420} , a unique mycobacterial enzyme-coenzyme system, is required for denitration.

4.1.2 F_{420} is a low redox potential flavin coenzyme and obligate hydride donor

Coenzyme F_{420} was first observed in mycobacteria in the early 1960s^{151,152} although it was not chemically identified until 1972 when it was isolated from methanogenic Archaea¹⁵³. It is now known that F_{420} and its precursor FO are present in a variety of organisms across several phylogenetic lineages, including

archaea, bacteria and eukaryotes. F_{420} is named after its intense absorbance at 420 nm, which is dependent on the reduction-oxidation (redox) state of the molecule.

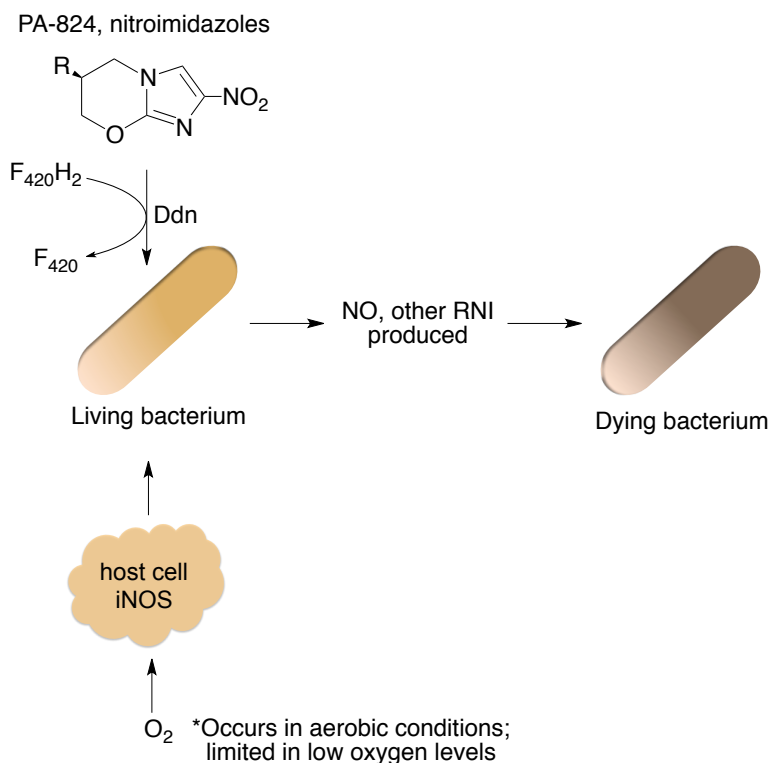


Figure 4.1 A comparison of the nitroimidazole mechanism of action and the NO release that occurs as a result of the innate immune response. In aerobic conditions, which favour replicative bacteria, the host cell can produce RNI via the inducible nitric oxide synthase (iNOS); however, iNOS activity is significantly reduced in anaerobic conditions (which favours latent, persistent *Mtb*). Nitroimidazoles are capable of producing NO and other RNI in both aerobic and anaerobic conditions. Figure adapted from Nathan, C. *Science*. 2008¹⁵⁰.

F_{420} shares some structural similarities to more common flavins, such as riboflavin, FMN and flavin adenine dinucleotide (FAD) (Figure 4.2). These include the isoalloxazine ring core, ribityl moiety, and phosphate. However, the tricyclic ring system in F_{420} differs from that in FMN and FAD in the following respects: 1) F_{420} lacks the central nitrogen at position 5 found in other common flavins, which is replaced by a carbon atom. For this reason, the F_{420} ring system is referred to as a *deazaisoalloxazine* ring system, and its reactivity is restricted to hydride transfer at this site¹⁵⁴. 2) F_{420} has a hydroxyl group at the C8 position, while

riboflavin, FMN, and FAD are methylated at C7 and C8. Extending beyond the phosphate moiety, F_{420} has a lactyl group and polyglutamate tail with γ -glutamyl linkages. In *Mtb*, most F_{420} molecules have a polyglutamate tail that is five to six glutamate residues in length¹⁵⁵.

Although coenzyme F_{420} and other common flavins are structurally similar, they differ significantly in their redox properties. For example, F_{420} is an obligate hydride donor, while FMN, FAD, and riboflavin can participate in one and two-electron redox reactions^{145,156}. In this sense, F_{420} is more similar to the conventional hydride carrier, NAD(P)^+ . However, F_{420} has a lower midpoint potential for hydride transfer (-380 mV) than NAD(P)^+ (-320 mV)^{146,157}. It has been suggested that the low reduction potential of F_{420} may play a role in the non-replicating persistence of *Mtb*, as F_{420} could facilitate redox reactions in anaerobic environments^{25,157}. It is also thought that F_{420} -dependent enzymes may be involved in protection against oxidative stress and nitrosative damage^{147,158,147,159,148,160}. This is consistent with evidence which shows *Mtb* mutant stains defective in F_{420} biosynthesis are hypersensitive to RNI^{148,149}.

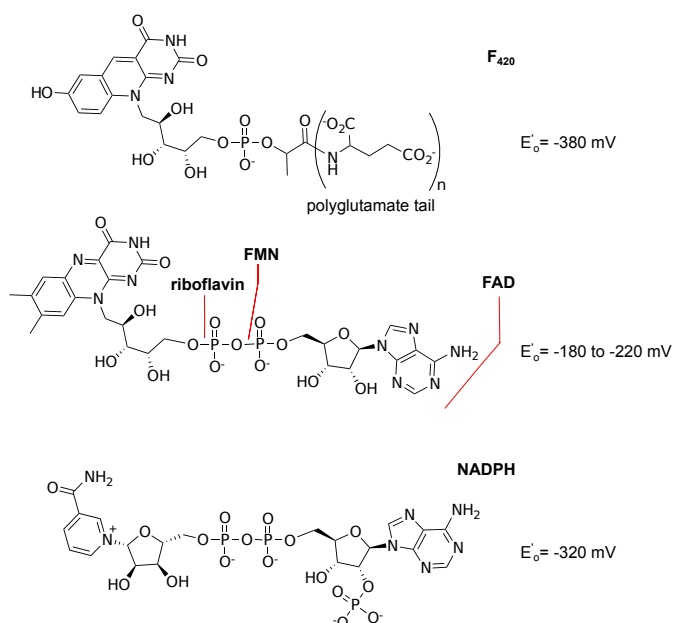


Figure 4.2. Structures of F_{420} , flavins, and NADP⁺ and their associated chemical reduction potentials^{150,157,150,161}.

F_{420} -catalysed enzymatic reactions have been studied in a range of microbes, but particularly in methanogens, where they play an integral role in carbon fixation reactions. Enzymes Mer and Mtd use coenzyme F_{420} to catalyse the reduction of N^5 , N^{10} -methylene-tetrahydromethanopterin (methylene- H_4 MPT) and methenyl- H_4 -tetrahydromethanopterin (methenyl- H_4 MPT) respectively in the reduction of carbon dioxide to methane^{162,163}. It has also been demonstrated that deazaflavin-dependent oxidoreductases in *Streptomyces* participate in the biosynthesis of antibiotics tetracycline¹⁶⁴ and tomaymycin¹⁶⁵, for which the coenzyme facilitates the enzymatic redox transformations of ketones and imines, respectively. F_{420} -dependent enzymes have been shown to catalyse aromatic nitro reduction in the degradation of picric acid and 2,4-dinitrophenol¹⁶⁶. F_{420} has been implicated in the degradation of mycotoxins in *Mycobacterium smegmatis*, which involves the reduction of carbon–carbon double bonds¹⁶⁷.

4.1.3 F_{420} -dependent enzymes in *Mtb*

Little is known about the enzymes in *Mtb* that require F_{420} as a coenzyme and the reactions they catalyse. One that has been well studied in *Mtb* and related mycobacteria is the F_{420} -dependent glucose 6-phosphate dehydrogenase (FGD)^{168,152,169,153,170}. The *Mtb* genome encodes two FGDs (*fgd1*, *fgd2*). This enzyme catalyses the oxidation of glucose 6-phosphate, producing reduced $F_{420}H_2$, which is required for PA-824 activation. Interestingly, the *Mtb* genome encodes two glucose 6-phosphate dehydrogenases that are dependent on NADPH as well.

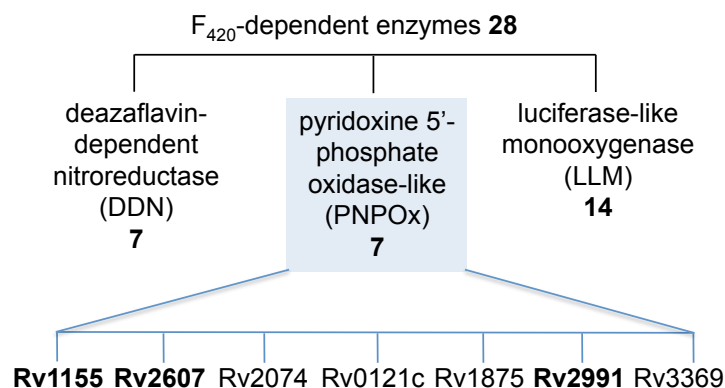


Figure 4.3. Predicted F_{420} -binding enzymes in *Mtb* using phylogenetic profiling. There are 28 predicted F_{420} -dependent enzymes, 7 of which belong to the DDN class, 14 to the LLM class, and 7 to the PNPOx class. This figure was created based on the computational study from Selengut *et al. J. Bacteriol.* 2010²⁶.

With the exception of FGD and Ddn, no other F_{420} -dependent proteins in *Mtb* have been experimentally identified and characterised in *Mtb*. A recent comparative genomics study suggests that there are at least 28 unique F_{420} -binding enzymes in *Mtb* that fall into three classes: the deazaflavin-dependent nitroreductase (DDN), the luciferase-like monooxygenase (LLM), and the PNPOx families, which all include homologues in other organisms that are known to bind either FMN or FAD^{154,171} (Figure 4.3). The family members are related to one another by global fold similarity. These enzymes were linked with F_{420} metabolism in part by identifying orthologues in multiple species that contain F_{420} biosynthetic genes. Conserved residues important to F_{420} binding were identified in each enzyme class^{155,171}.

4.2 Objective and approach

The aim of this work is to identify and characterise the cofactor dependence of predicted flavoproteins Rv1155 and Rv2991. There are two facets to this objective: 1) to demonstrate a cofactor-protein binding interaction 2) to observe cofactor-mediated enzyme catalysis. The cofactor preference of Rv1155 and Rv2991 was investigated using biophysical techniques, which is described in

Section 4.4 of this chapter. Test enzymatic reactions for Rv1155 and Rv2991 are reported in Section 4.5.

4.3 Expression and purification of Rv1155 and Rv2991

Rv1155 was best expressed as a maltose binding protein (MBP)-His_{6x} fusion protein. Other fusion constructs of Rv1155 tested for expression in BL21(DE3) *E.coli* cells included glutathione S-transferase (GST) and His_{6x} tags. The Rv1155-His_{6x} fusion protein was insoluble. While the GST-Rv1155 fusion protein expressed soluble protein, the yield was about 7 mg per 1 L of cell culture, which was half the yield for His_{6x}-MBP-Rv1155. Rv2991 was expressed and purified as a His_{6x} fusion protein. The MBP-His_{6x} and His_{6x} tags were efficiently cleaved from the Rv1155 and Rv2991 gene products respectively with tobacco etch virus (TEV) protease. After a 2 hr incubation at room temperature (25 °C), 95% cleavage of each protein was observed with a molar ratio of 1:100 TEV to fusion protein. The tag, uncleaved fusion protein, and TEV-His_{6x} were removed from the reaction mixture with a second nickel affinity column. Each protein was further purified by gel filtration and stored in 100 mM Tris-HCl, pH 7.5, 50 mM NaCl. The final protein solution was >95% pure as assessed by SDS-PAGE. The mass of untagged Rv1155 was determined to be 16227.0 Da by MS, which is the calculated molecular weight of Rv1155 without the first methionine and with an added glycine. The mass of Rv2991 is 18034.0 Da by MS, which corresponds to the molecular weight of Rv2991 without the first methionine or the terminal proline and with an added glycine at the N-terminus. Unlike Rv2607, recombinant Rv1155 and Rv2991 isolated from *E. coli* cells were colourless proteins and there was no evidence to suggest that either Rv1155 or Rv2991 co-purified with a small molecule.

4.4 Purification of F₄₂₀

F₄₂₀ is not available commercially available and must be isolated from an organism with the biosynthetic machinery to produce the coenzyme.

M. smegmatis cell cultures have been previously used for F₄₂₀ purification¹⁷² and a recent protocol was developed to increase F₄₂₀ production by transforming *M. smegmatis* cells with an overexpression vector containing *fbiA*, *fbiB*, and *fbiC*¹⁷³, genes involved in F₄₂₀ biosynthesis. *M. smegmatis* cells transformed with the vector containing *fbiABC* have higher F₄₂₀ production levels (~5.5x) than do WT cells under the conditions tested (See Chapter 6, Figure 6.2). The method for F₄₂₀ purification described in this thesis was developed by modifying and optimising existing methods^{172,173} and is described in detail in Chapter 6, section 6.5. From 800 g *M. smegmatis* +*fbiABC* cells, 102 mg pure F₄₂₀ was obtained.

4.5 Evidence for Rv1155-F₄₂₀ and Rv2991-F₄₂₀ binding interactions

4.5.1 Rv1155 and Rv2991 bind to F₄₂₀ as determined by biophysical methods

A crystal structure of Rv1155 bound to FMN has been determined¹²⁷; however, there is disagreement over the binding affinity of FMN to Rv1155^{127,118}. Although Rv2991 is annotated as a probable FMN-dependent protein in the *Mtb* genome, no experimental binding data has been published to confirm the Rv2991-FMN binding interaction. A series of redox cofactors were assessed for binding using DSF, including FMN, flavin adenine dinucleotide (FAD), and nicotinamide adenine dinucleotide phosphate (NADPH), and F₄₂₀ (Figure 4.4). Both Rv1155 and Rv2991 show a strong preference for F₄₂₀. A significant thermal shift was observed for each protein in the presence of F₄₂₀. With the addition of F₄₂₀ (50 µM), the T_m of Rv1155 was increased by 5.3 ± 0.4 °C, while that of Rv2991 was increased by 6.4 ± 0.4 °C. The T_m of either Rv1155 or Rv2991 was not significantly increased in the presence of FMN, FAD, or NADPH (Figure 4.4 C). Some stabilisation in the T_m of Rv2991 was observed with FMN (2.2 ± 0.4 °C), but a binding interaction between FMN and Rv2991 could not be detected by ITC. Further, Rv2991, unlike FMN-dependent Rv2607, does not co-purify with FMN when expressed recombinantly in *E. coli* cells.

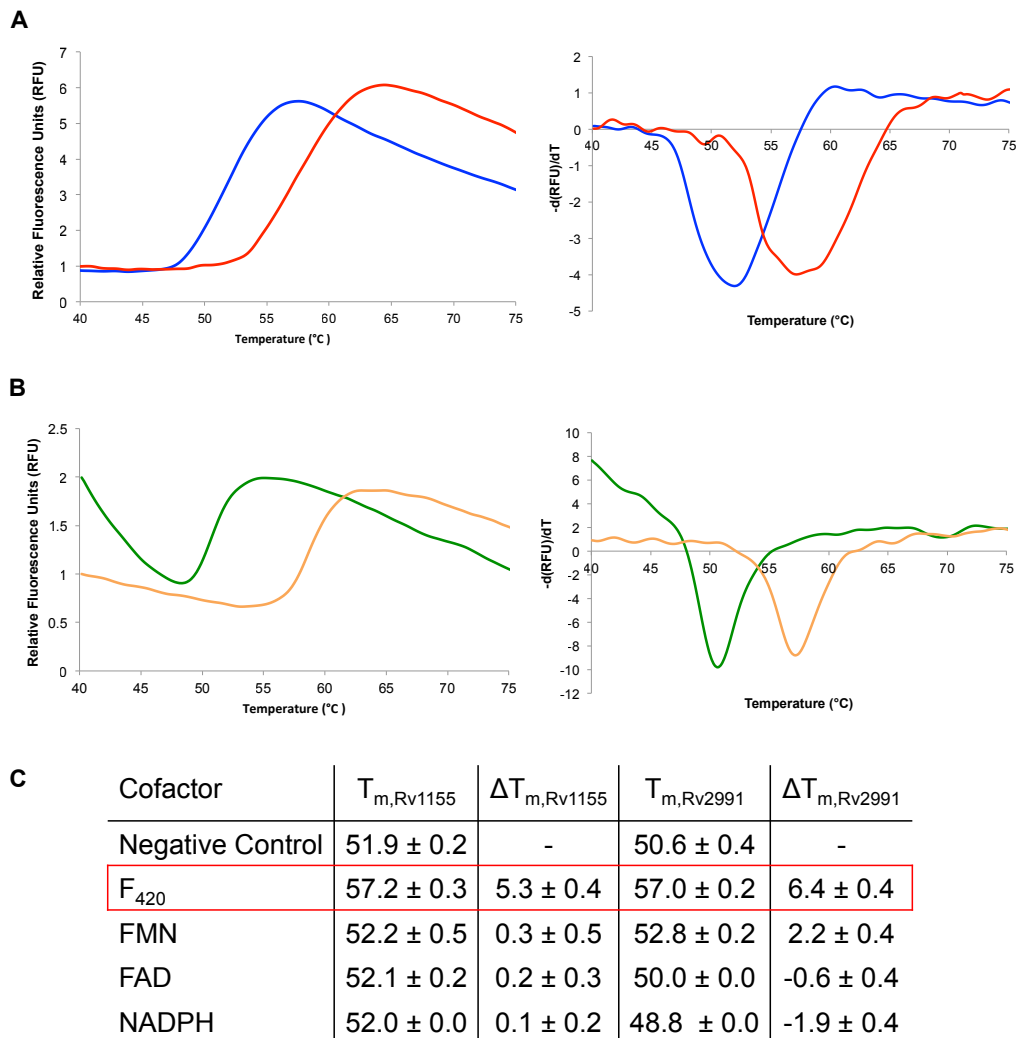


Figure 4.4. Cofactor recognition of Rv1155 and Rv2991 assessed by DSF. (A) Melting curves of Rv1155 (left) in the presence (red) and absence (blue) of F_{420} and the associated negative derivatives with respect to temperature (right). (B) Melting curves of Rv2991 (left) in the presence (orange) and absence (green) of F_{420} and the associated negative derivatives with respect to temperature (right). (C) Table of T_m values for Rv1155 and Rv2991 in the presence of F_{420} , FMN, FAD, or NADPH. The change in T_m in the presence of a test cofactor (ΔT_m) is relative to the cofactor negative control. Fluorescence measurements (λ_{ex} : 587 nm, λ_{em} : 607 nm) were made over a temperature range of 30 - 80°C. All Rv1155 samples (100 μ L total volume) contained the following: 100 mM Tris-HCl pH 8.0, 100 mM NaCl, 2.5x SYPRO® Orange, and 5 μ M Rv1155, with or without 50 μ M cofactor. All Rv2991 samples (100 μ L total volume) contained the following: 20 mM potassium phosphate buffer pH 7.4, 20 mM NaCl, 20 mM DTT, 1x SYPRO® Orange, and 5 μ M Rv2991, with or without 50 μ M cofactor.

Preliminary DSF trials with apoenzyme Rv2991 revealed very high baseline signal, i.e. high fluorescence at temperatures lower than the T_m . This high initial fluorescence can be indicative of protein aggregation or denaturation. The presence of Rv2991 aggregates was later confirmed by dynamic light scattering (DLS). Extensive DSF buffer optimisation studies were carried out to eliminate the aggregates (discussed in Chapter 6). Buffer conditions that produced the lowest initial fluorescence contained 20 mM dithiothreitol (DTT), which prevents cross-linking via exposed cysteines on the Rv2991 surface (2 cysteines per Rv2991 protomer). The baseline DSF signal for Rv2991 is lower in the presence of F_{420} , but not with other cofactors tested. Some remaining initial fluorescence may be attributed to aggregation caused by an unstructured carboxyterminal 'tail' in Rv2991 comprised of 14 residues. In order to investigate the contribution of this unstructured region to aggregation, a mutant of Rv2991 was created that lacks the final 14 residues at C-terminus of Rv2991 (-KLGLPHMSVGGSTA-). This truncated mutant was unstable and precipitated upon isolation, indicating that the unstructured tail is important to Rv2991 stability.

F_{420} binding to Rv1155 and Rv2991 was confirmed by STD NMR.(Figure 4.5). The observed aromatic resonances of the deazaisoalloxazine protons are consistent with those reported in the literature¹⁵⁵ and are clearly seen in the difference spectra for both the Rv1155- and Rv2991-containing samples.

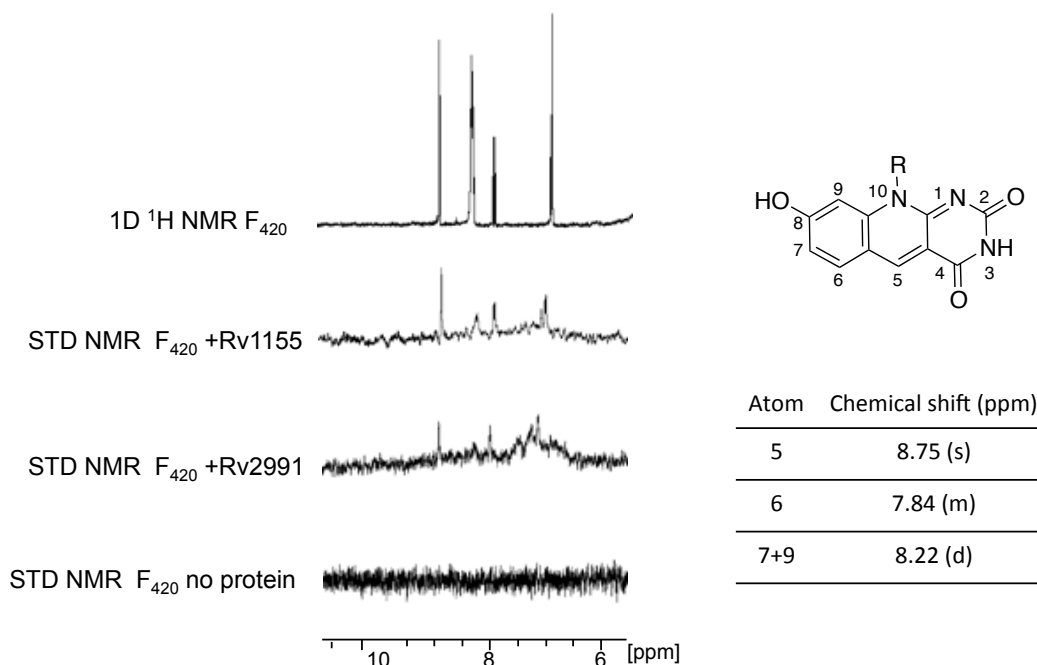


Figure 4.5. F_{420} binds to Rv1155 and Rv2991 by STD NMR. The aromatic region of the 1D ^1H NMR spectrum for F_{420} (top) includes signals due to protons in the deazaalloxazine ring system (top right). The STD NMR spectra indicate that F_{420} is binding in both Rv1155 and Rv2991. Each sample contains 25 μM Rv1155 or Rv2991 (or the equivalent volume of buffer for the protein negative control) and 500 μM F_{420} in 100 mM potassium phosphate buffer, pH 7.8, 50 mM NaCl, 10% D_2O . The STD signal was not observed in the absence of protein.

The Rv2991- F_{420} and Rv1155- F_{420} binding interactions were further characterised by ITC (Figure 4.6). The K_D s for F_{420} binding to each enzyme are in the low micromolar range, which is consistent with the reported binding constants for other F_{420} -dependent proteins^{170,169,174}. ITC also indicates that Rv1155 and Rv2991 each bind F_{420} in a 1:1 molar ratio. This observation is confirmed for each enzyme-coenzyme complex by nESI-MS, which showed that Rv1155 and Rv2991 predominantly exist as homodimers that bind two molecules of F_{420} (Figure 4.7). One dominant species is observed by nESI-MS for Rv2991- F_{420} , which is a homodimer bound to two molecules of F_{420} -6, a F_{420} species with six glutamates (Figure 4.7 A). The spectrum also reveals that a relatively smaller proportion of sample is unliganded monomer. The mass spectrum for Rv1155- F_{420} reveals a heterogeneous population of enzyme-coenzyme complexes, with

evidence for dimeric Rv1155 binding one or two molecules of F_{420} -5 (F_{420} with five glutamates), one or two molecules of F_{420} -6, both F_{420} -5 and F_{420} -6, and no F_{420} (apoenzyme). The most prevalent species in this mixture, however, was homodimeric Rv1155 bound to two molecules of F_{420} -6. Differences in the mass spectra for Rv2991 and Rv1155 could be attributed to the molar ratio of protein to F_{420} present in each sample (1:5 for Rv2991 and 1:1 for Rv1155). This is further discussed in the following on X-ray crystallographic studies with Rv1155 and Rv2991.

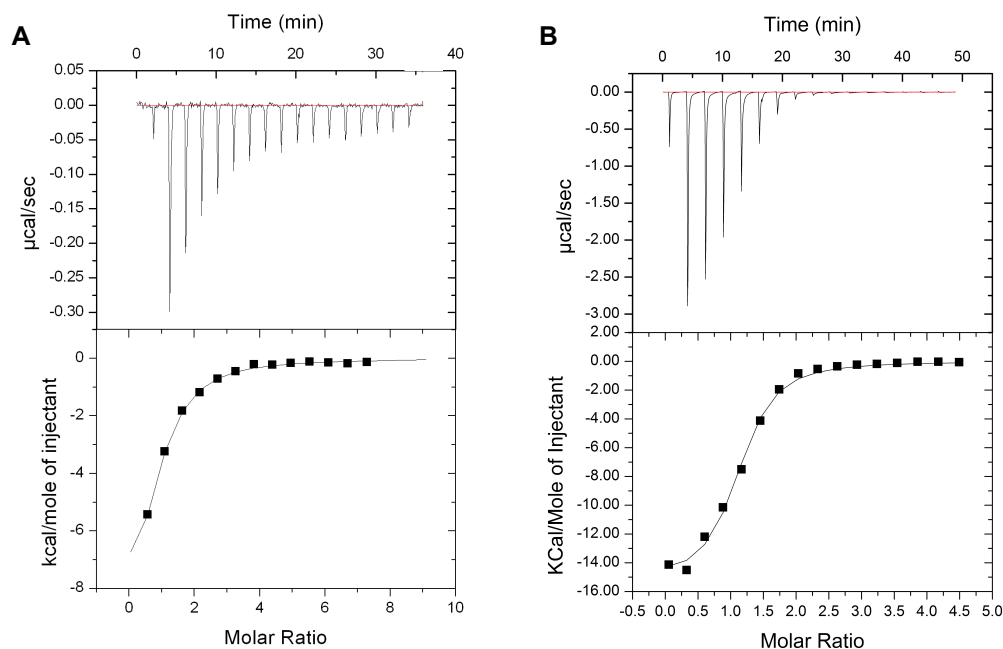


Figure 4.6. ITC binding isotherms for F_{420} titrated into Rv1155 and Rv2991. A solution of (A) F_{420} (275 μM) was injected into Rv1155 (10 μM), and (B) F_{420} (1150 μM) was injected into Rv2991 (85.5 μM). The ligand and protein were dialysed into the same buffer, 100 mM potassium phosphate, pH 7.4, 20 mM NaCl. The enthalpy change for each injection as a function of the molar ratio of total injected F_{420} to Rv1155 (A) or Rv2991 (B) is shown (lower panel, squares). A one-site binding model fitting was used to the K_D for Rv1155 (3 μM) or Rv2991 (4 μM).

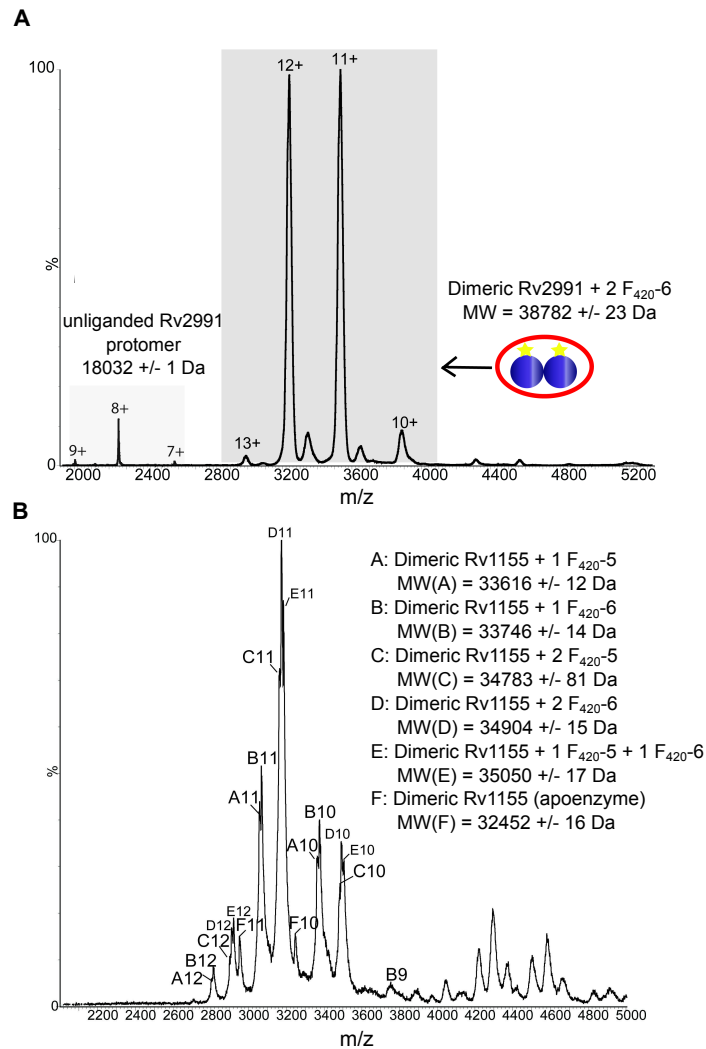


Figure 4.7. Nano-ESI mass spectra of intact Rv2991 and Rv1155 in the presence of F₄₂₀. (A) Mass spectrum of Rv2991. The spectrum clearly shows the predominant species has peaks with charges ranging from +13 to +10 in the 2800-4000 m/z range that correspond to a MW of 38782 ± 23 Da. This experimental mass is consistent with the calculated mass for the complex of dimeric Rv2991 with two molecules of F₄₂₀₋₆ bound (38644 Da), which is derived from the amino acid sequence of Rv2991 excluding the aminoterminal methionine and the carboxyterminal proline and including one additional glycine residue retained from the TEV cleavage (36064 Da; 18032 Da per protomer) and the MW of F₄₂₀₋₆ (1290 Da). A minor species in the sample is unliganded monomeric Rv2991, corresponding to a mass of 18032 Da. The concentration of Rv2991 used to generate the spectrum is 12 μM, in the presence of 5x molar excess of F₄₂₀. (B) Mass spectrum of Rv1155. The sample is heterogeneous, with a mixture of dimeric Rv1155 bound to one or two molecules of F₄₂₀₋₅ or F₄₂₀₋₆. Unliganded dimeric Rv1155 also exists in this mixture. The calculated mass of dimeric Rv1155 excluding the aminoterminal methionine and including an added glycine is 32453 Da, which is consistent with the MW of species F (apoenzyme Rv1155). The MW of F₄₂₀₋₅ is 1159 Da. The concentration of Rv1155 used to generate the spectrum is 10 μM with approximately 1:1 molar ratio of Rv1155 to F₄₂₀.

The biophysical data presented here unambiguously demonstrate that F_{420} binds to both Rv1155 and Rv2991. Rv1155 and Rv2991 should therefore be annotated as F_{420} -dependent enzymes, a classification which deviates from their current status in *Mtb* genome databases as FMN-dependent PNPOx-like proteins.

4.5.2 X-ray crystallographic studies with Rv1155 and Rv2991

Five X-ray datasets were measured from yellow crystals of the Rv1155- F_{420} complex, which diffracted X-rays from 1.9 Å to 2.8 Å resolution. Each crystal structure was similar, but varied slightly in ways that provided details about the binding mode of F_{420} in Rv1155. For statistics on the datasets collected, please refer to Appendix A. All Rv1155- F_{420} crystals were in the space group $P2_12_12_1$ and had unit cell dimensions of $a = 54.04$ Å, $b = 65.76$ Å, $c = 76.59$ Å, while Rv1155 apoenzyme crystallised in the space group $P2_1$ and had unit cell dimensions $a = 46.95$ Å, $b = 55.12$ Å, $c = 55.23$ Å. The strongest electron density for F_{420} was observed in the lowest resolution dataset. There is clear electron density in each structure to guide the unambiguous placement of two glutamate residues, the phosphate, the lactyl moiety, and the ribityl moiety of F_{420} . The additional three to four glutamate residues of the F_{420} polyglutamate chain are likely unstructured in the solvent channels of the crystal. The negative charges of the phosphate group and the polyglutamate tail of F_{420} are accommodated by a tunnel of positively-charged residues in Rv1155 (Figures 4.8 and 4.9), an interaction which anchors the coenzyme to the protein.

Just adjacent to the ring system binding site is a large, unoccupied and uncharged cleft (Figure 4.9), which may be the binding site of potential substrates. Crystal structures of F_{420} binding proteins, and of other flavoproteins in which the flavin directly engages in redox chemistry with the substrate, indicate that any potential substrate must be in close physical proximity to the ring system. In F_{420} -binding proteins, hydride transfer is achieved via transfer from the

central carbon of the ring system. Therefore, substrates must bind near the ring system in order to facilitate redox reactions.

Although there is conclusive evidence for the placement of the coenzyme in the protein, the five crystal structures obtained are all ambiguous about the face of the ring system, that is, whether the pyrimidine portion of the deazaalloxazine is positioned away from or toward Lys 57. However, the F_{420} electron density in the 2.8 Å resolution dataset provides evidence for the vector of the dezaialloxazine ring (Figure 4.8). In this dataset, density extending beyond the ribityl moiety to the central nitrogen of the ring system is observed. This vector for the dezaialloxazine ring system places it close to the predicted substrate binding site, which likely requires a hinge motion of 1-2 Å to facilitate enzymatic catalysis.

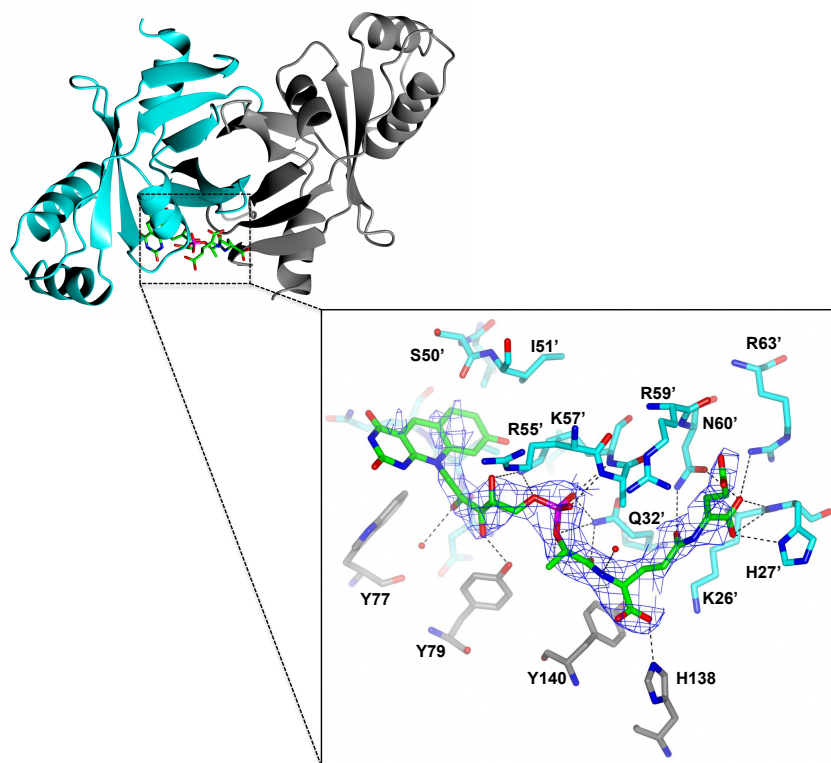


Figure 4.8. Crystal structure of Rv1155 bound to F_{420} at 2.8 Å resolution. Rv1155 is a homodimer with one molecule of F_{420} bound. Residues marked with apostrophe belong to chain B (cyan) and unmarked residues belong to chain A (grey). Electron density for F_{420} is contoured at 1.0σ level. The image was rendered using molecular graphics software CCP4mg¹²⁵.

The same sample that was used to obtain Rv1155-F₄₂₀ crystals was also analysed by nESI-MS, as previously discussed (Figure 4.7 B). The mass spectrum indicates that several species of Rv1155-F₄₂₀, although the predominant complex in the mixture is Rv1155 bound to two molecules of F₄₂₀-6. However, only one F₄₂₀ molecule per dimer is observed in the crystal structure. It may be that the concentration of F₄₂₀ (1:1 molar ratio of Rv1155 to F₄₂₀) was not sufficient to populate all the binding sites (please refer to Chapter 6 for details on protein purification). Another explanation is that the apoenzyme Rv1155 crystal formed in the drop and F₄₂₀ subsequently soaked into the crystal. This interpretation is supported by the observation that unstructured glutamates after the first two residues in the polyglutamate chain of F₄₂₀ interact with the unoccupied polyglutamate binding site of a symmetry-related Rv1155 homodimer. The strong electron density ($\sigma = 1$) for one glutamate residue in the F₄₂₀ binding site of the symmetry related molecule indicates that the unstructured glutamates of F₄₂₀ bound in one unit cell may be preventing the binding of a second F₄₂₀.

Tryptophan 77 of Rv1155 may be involved in positioning the deazaisoalloxazine ring system of F₄₂₀. Molecular dynamics simulations predicted that Trp 77 would engage in a π -stacking interaction with the ring system; however, the structure shown here does not support this model. Nonetheless, the variation in electron density for Trp 77 among all datasets collected may provide some insight into the role of this residue at the active site. In the unoccupied active site, Trp 77 (chain B) is disordered in all datasets. In the F₄₂₀-occupied active site, there is clear electron density for Trp 77. This observation suggests that the presence of F₄₂₀ orders Trp 77.

Missing electron density due to disorder in the crystal at key, biochemically-relevant regions of these molecules makes it difficult to understand the active site of this enzyme with a high level of mechanistic detail. Further co-crystallization experiments may produce a more useful crystal form for this

protein. However, these structures provide strong evidence for the F_{420} binding site on this protein and a starting point for identifying the substrate binding site.

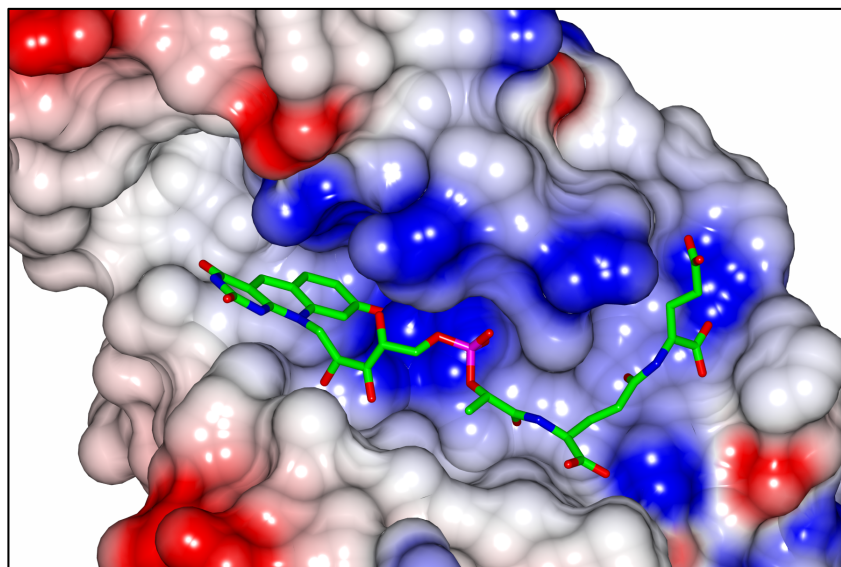


Figure 4.9. Electrostatic potential surface representation of F_{420} binding site in Rv1155. Positively charged residues (at neutral pH) accommodate the phosphate and polyglutamate of F_{420} . The proposed substrate binding site is adjacent to the deazaisoalloxazine ring system of F_{420} . Molecular graphics representation of the molecule was rendered using CCP4mg¹²⁵.

Co-crystallisation screens with Rv2991- F_{420} did not produce diffraction-quality crystals of the complex. These screens are described in Chapter 6 of this thesis. However, crystals of apoenzyme Rv2991 could be reproduced with known crystallisation conditions (Pedro Alzari *et al.*, unpublished results). This indicates that the isolated protein is properly folded and suitable for crystallisation trials. Soaking experiments and co-crystallisation trials are currently underway to obtain the F_{420} -bound Rv2991 structure.

A structural superimposition of Rv2991, Rv1155, and F_{420} -dependent Ddn reveals that they share a high degree global fold similarity (Figure 4.10) even though the sequence identity of these three enzymes is only around 20%. Ddn oligomerises into a pentamer, while Rv2991 and Rv1155 are dimers. The architecture of the small domain for Ddn is markedly different than that of either

Rv2991 or Rv1155 These observations suggest that Rv2991 and Rv1155 should be in a class of F_{420} -binding proteins distinct from Ddn, an observation which is supported by studies of homologues in *M. smegmatis*¹⁷⁵ and by computational analyses¹⁷¹.

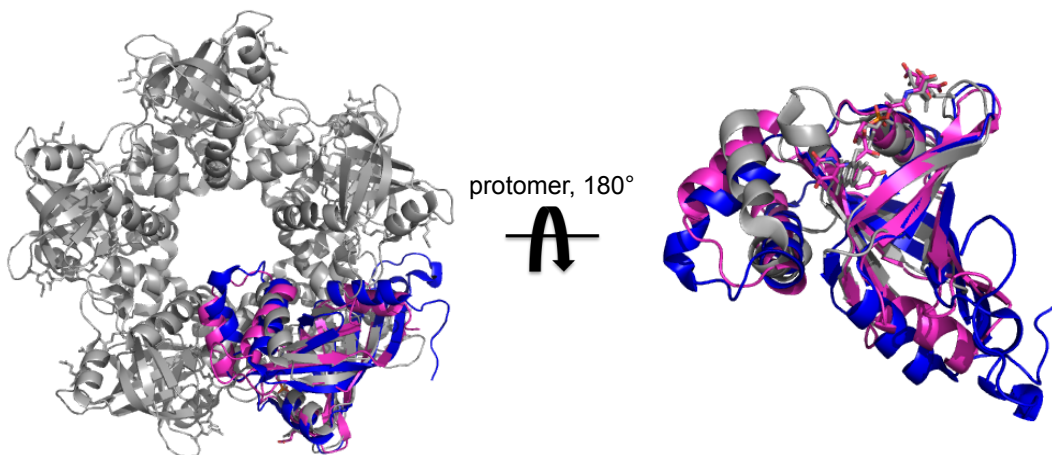


Figure 4.10. Structural alignment of Ddn- F_{420} , Rv1155- F_{420} , and Rv2991. The alignment was performed using the MultiProt Server (<http://bioinfo3d.cs.tau.ac.il/MultiProt/>) with Ddn¹⁷⁴ (grey, PDB ID: 3R5W), Rv1155- F_{420} (magenta), and Rv2991 (blue, PDB ID: 1RFE). C-alpha RMSD of Rv2991 and Ddn is 1.51 and of Rv1155 and Ddn is 1.28. The C-alpha RMSD of Rv2991 and Rv1155 is 1.45. (Number of residues: Rv1155, 141; Rv2991, 151; Ddn 109).

4.6 Trial enzymatic reactions with Rv1155 and Rv2991

4.6.1 Rv1155 and Rv2991 do not exhibit PNPOx activity

To determine if either Rv1155 or Rv2991 are capable of PNPOx activity, as is Rv2607 (Chapter 3), each enzyme was incubated with PNP, PMP, and PLP in the presence of F_{420} or $F_{420}H_2$ (Figure 4.11 A-C) and the reaction was followed by UV/Vis spectroscopy. The substrate, PNP, the product, PLP, and coenzyme, F_{420} , have absorbance maxima at 325 nm, 388 nm, and 420 nm respectively at pH 7.5. Oxidation of PNP and its depletion results in a decrease in absorbance at 325 nm, an increase in absorbance at 388nm with the formation of the product PLP, and a decrease in the absorbance at 420 nm indicating the reduction of

F₄₂₀. No spectral changes were observed in the presence of untagged Rv1155 or Rv2991 that had been thawed from frozen storage or freshly purified. Clarified lysate from bacteria overexpressing Rv1155-GST or His_{6x}-Rv2991 was also tested for activity to determine if Rv1155 requires another protein or small molecule endogenous in *E. coli* for PNPOx activity. The clarified lysate did not result in spectral changes consistent with PNPOx enzymatic activity. The results indicate that Rv1155 and Rv2991 likely do not function as a PNPOxs. Unlike Ddn, Rv1155 and Rv2991 cannot metabolise prodrug PA-824 (Figure 4.11 D). However, because these are negative results, it is difficult to conclusively rule out other factors for inactivity, such as the requirement of a small molecule or partner protein from *Mtb*.

4.6.2 Rv1155 and Rv2991 do not recognize aflatoxin as a substrate

Mycobacterium smegmatis homologues of enzymes in the DDN and PNPOx families have been linked to F₄₂₀ metabolism with transposon mutagenesis studies designed to identify enzymes in *M. smegmatis* responsible for the degradation of aflatoxin, a food crop mycotoxin^{175,167}. Of the five *M. smegmatis* isolates that did not degrade aflatoxin, four insertions disrupted the *fbjC* gene, which encodes the enzyme that catalyses the last step of the biosynthesis of FO, the deazaflavin precursor to F₄₂₀.

One isolate disrupted the gene that encodes FGD. The enzymes responsible for this activity were identified by fractionating protein extracts, testing for aflatoxin degradation, and identifying active enzymes with proteomic tools. Five enzymes belonging to the DDN family and four enzymes in the PNPOx family were expressed recombinantly, purified, and shown to catalyse the F₄₂₀-dependent reduction of aflatoxins and analogues containing the coumarin core¹⁷⁵.

Rv1155 and Rv2991 were tested for activity with coumarin analogues (Figure 4.11 E-H) by monitoring spectroscopic changes, as described above. No

spectral changes were observed in the presence of any of the coumarin compounds, nor did any of these compounds bind by STD NMR (Figure 4.11 E-H).

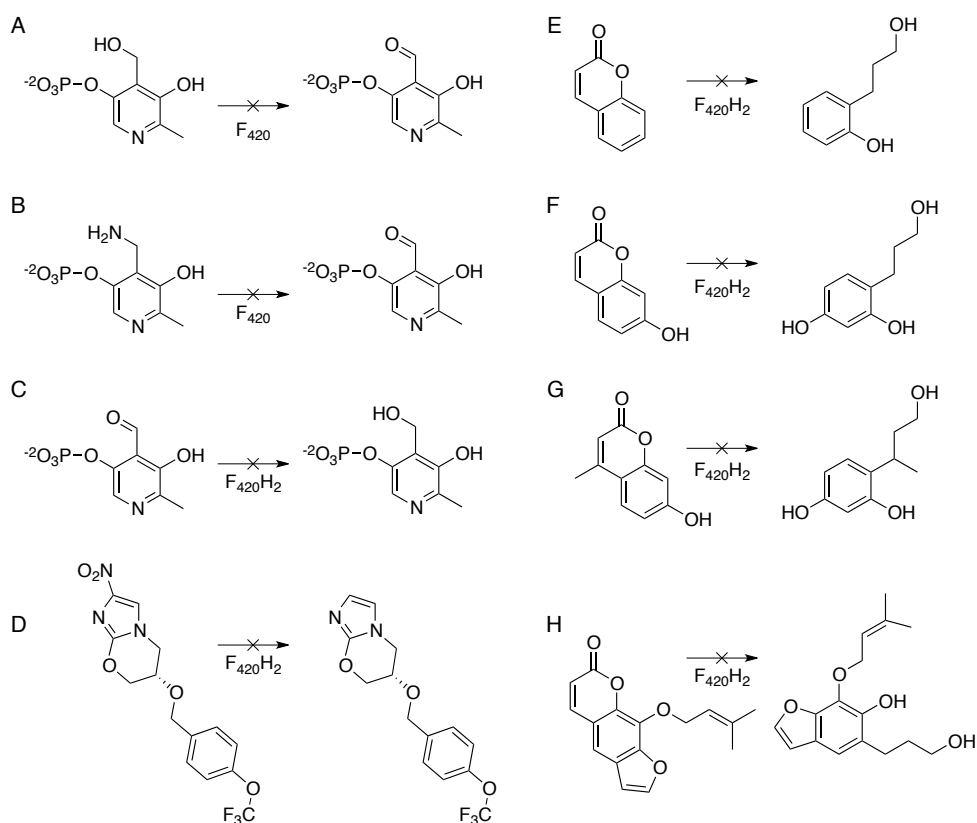


Figure 4.11. Test enzymatic reactions conducted with Rv1155 and Rv2991. The following substrates were screened for activity with Rv1155 and Rv2991 separately in the presence of either F_{420} or $F_{420}H_2$: (A) pyridoxine 5'-phosphate, (B) pyridoxamine 5'-phosphate, (C) pyridoxal 5'-phosphate, (D) PA-824, (E) coumarin, (F) 7-hydroxycoumarin, (G) 7-hydroxy-4-methylcoumarin, and (H) imperatorin.

These results are maybe not surprising as aflatoxin and coumarin analogues are unlikely natural substrates for these mycobacterial enzymes, as neither *Mtb* and its close relative, *M. smegmatis*, would normally encounter these toxins in their natural environment.

4.7 Conclusions

Rv1155 and Rv2991 were previously annotated in the *Mtb* genome as probable FMN-binding PNPOx-like proteins with unknown function. The biophysical and structural results described here suggest that they are not FMN-dependent PNPOxs, as is Rv2607, but are instead members of a novel F₄₂₀-dependent class of proteins that share a global fold similarity. The trial enzymatic reactions conducted with Rv1155 and Rv2991 rule out any obvious function these enzymes may possess. The catalytic activities of Rv1155 and in particular, Rv2991, are further explored in Chapter 5.

Chapter 5

Design, development, and screening of a targeted flavoprotein fragment library

5.1 Background

Functional genome annotation is usually achieved by employing computational methods to infer evolutionary relationships between protein sequences¹⁷⁶. While this information can guide initial hypotheses of protein function, protein characterisation via biochemical methods remains essential to definitive gene annotation. This is especially important in light of the widespread misannotation of genes¹⁷⁷ and the abundance of gene products with homologues of unknown function, also known as “conserved hypothetical” proteins (CHPs)¹⁷⁸. CHPs are proteins encoded in the genomes of organisms from multiple phylogenetic lineages, none of which have been functionally characterised¹⁷⁹. Any newly

sequenced bacterial genome contains about 30-40% genes encoding CHPs¹⁸⁰. As discussed in Chapter 1, Section 1.3, the *Mtb* genome includes over a thousand CHPs¹⁹. Partial or inaccurate gene function information poses serious challenges to the study of this human pathogen and can impede the discovery of novel drug targets.

Few gene products in *Mtb*, and even fewer CHPs, have been assigned function based on experimental data. One major reason for this is the paucity of experimental techniques to elucidate the function of CHPs. Knock-out or knock-down mutation studies¹⁸¹, transcriptional profiling¹⁸², and more recently, metabolomics technologies¹⁸³ are some existing experimental approaches to studying protein function. While these techniques can provide evidence of CHP function, they tend to be technically challenging and low-throughput.

5.2 Objective and rationale

The objective of this thesis work is to explore the applicability of fragment-based methods to protein function prediction. Fragment-based approaches, as described in Chapter 1, Section 1.2.2.2, involve the use of sensitive biophysical techniques to detect relatively weak non-covalent binding interactions between a target protein and low molecular weight compounds (<300 Da) known as 'fragments'. As described previously, these fragments are typically elaborated with the guidance of protein structural information into inhibitors with greater complexity and affinity for the target¹⁰¹. The advantage of biophysical screening over traditional approaches to function prediction is that it does not require time-consuming genetic experiments. Further, *a priori* protein structure information is an advantage, but is not necessary to conduct the screen. This work provides the first example, to our knowledge, of the application of fragment-approaches to investigate the natural substrates of CHPs. This will be referred to herein as the 'fragment-to-substrate' approach.

5.3. Methodological approach

Because fragments tend to interact with a target macromolecule weakly, fragment screening must be conducted using sensitive biophysical techniques. Proteins of unknown function have been previously studied with biophysical screening methods, DSF¹⁸⁴, SPR¹⁸⁵, and MS¹⁸⁶. There are several examples of *in silico* structure-based function prediction approaches, which are predicated on screening potential substrates by docking them into a proposed active site^{187,188,189,190}. As described in Chapter 1, ligand-binding screens can provide clues to protein function because they probe the molecular recognition properties of the target.

5.3.1 Fragments as chemical tools

Fragments may be useful chemical tools to investigate substrate specificity. Although fragments tend to be weak binders by virtue of their limited size and chemical complexity, they are known to have high ligand efficiency ($\Delta G/\text{number of non-hydrogen atoms}$)¹⁹¹. Fragments of a natural substrate can recapitulate the binding properties of the parent substrate^{192,102,193}. Early studies showed that fragments derived from substrates of beta-hydroxy-betamethylglutaryl coenzyme A reductase (and analogues) have additive free energies¹⁹². The total free energy of these fragments approximates the binding energy of the natural substrate, after correction for the linking energy. Other work demonstrated that the fragmented substrates of thymidylate synthase can bind the active site in a similar fashion to analogous moieties of the original substrate¹⁰². It has been reported that fragments of NADPH binding to *E. coli* ketopantoate reductase¹⁰³ can identify the 'hot spots' of the binding site, which contribute most to the binding energy of the cofactor. These studies retrospectively analyse the binding of a known substrate by determining the free energy contribution and binding mode of the substrate's constituent fragments. This thesis work aims to

determine if the prospective case is also true, i.e., whether fragments binding to a target enzyme can predict substrates for that enzyme.

Central to the success of this approach is a fragment library that is reflective of the chemical space occupied by natural substrates, which is markedly different from that of drug molecules. For example, natural products tend to have a greater prevalence of features such as chirality¹⁹⁴. Some commercial providers have developed fragment libraries for drug design inspired by natural ligands, including BioFragments developed by Asinex (<http://www.asinex.com/BioFragments.html>) and Fragments of Life™ from Emerald Biosciences¹⁹⁵ (<http://www.emeraldbiostructures.com/services/fragmentscreening/>). There has also been recent interest in developing metabolite-like fragment libraries in academic laboratories, with particular emphasis on introducing fragments with more sp³ centres^{196,197}. Fragment libraries targeted to specific classes of proteins have also been created¹⁹⁸, perhaps most notably for kinases¹⁹⁹. These targeted and metabolite-like fragment libraries have been designed to broaden the chemical space sampled by fragments used to design inhibitors.

The goal of this work is to develop a biologically-relevant, targeted fragment library that can probe the substrate specificity of enzymes in *Mtb*. This fragment library was assembled based on natural substrates of proteins functionally related to the target enzyme. The Protein Databank (PDB) was a source of natural ligand information. Using a computational technique called 'virtual fragment linking' (VFL)²⁰⁰, statistical models were created based on binding fragments, which were used to predict likely substrates from a database of known metabolites in the cell. Figure 5.1 summarises this 'fragment-to-substrate' workflow.

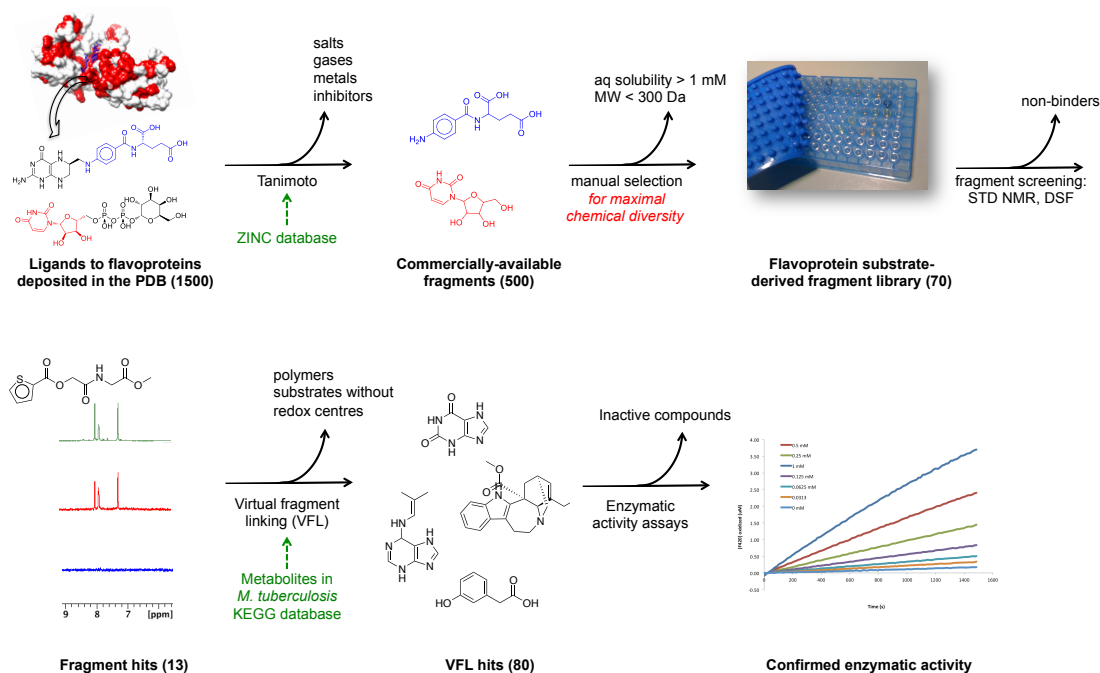


Figure 5.1. A schematic describing the ‘fragment-to-substrate’ workflow. The protein targets selected for this study are known flavoproteins; therefore, the ligands of flavin (FMN, FAD, and F₄₂₀)-dependent proteins were selected from the PDB (1500 ligands). Molecules such as salts, gases, metals, and inhibitor compounds were excluded. The remaining ligands were compared to low molecular weight (<300 Da) compounds with good aqueous solubility (>1 mM) in the ZINC database²⁰¹ (<http://zinc.docking.org/>) using Tversky²⁰⁰ similarity scores. Commercially-available fragments within the desired size and solubility ranges were manually selected to construct a fragment library that is representative of diverse chemical scaffolds. The library was screened against the target proteins using STD NMR. Fragments that bind to the target were analysed by VFL and compared to all known metabolites in the *Mtb* cell from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database²⁰² (approximately 4000 molecules). From the top ranking VFL hits, 80 molecules were manually selected as potential substrates and a subset were tested for enzymatic activity.

5.4 F₄₂₀-dependent enzymes Rv1155 and Rv2991 as model CHPs

As described in Chapter 2, Rv2991 and Rv1155 are annotated as PNPOx-like proteins with unknown function. The work presented in this thesis demonstrates that these two enzymes do not exhibit PNPOx activity (Chapter 4, Figure 4.11), nor do they have affinity for predicted cofactor FMN (Chapter 4, Figure 4.4). Rv2991 and Rv1155 both bind strongly to the mycobacterial redox coenzyme F₄₂₀. These two mycobacterial proteins will be used as model systems because

they are members of the same structural class of enzymes and each have no known function. Perhaps the most well-studied F_{420} -dependent system in *Mtb* is an enzyme called *Ddn*, which is responsible for activating a pro-drug PA-824 currently in clinical trials for TB^{25,147} (see Chapter 4 for a description of the mechanism of action). As novel F_{420} -dependent enzymes in *Mtb*, Rv2991 and Rv1155 are interesting targets.

A structural superimposition of Rv2991 and Rv1155 (Figure 5.2) reveals that they share similar global fold topology with an RMSD = 1.45 Å (number of residues: Rv2991, 153; Rv1155, 141). Although these enzymes share global fold similarity, they have relatively low sequence identity (33%). A region with high sequence identity between the two proteins is the polyglutamate binding site, which is consistent with the observation that these proteins both bind F_{420} (Figure 5.2, conserved residues are shown in surface representative, dark blue). The dimer interface is another region of high sequence similarity between these two proteins. However, the predicted active sites of Rv2991 and Rv1155 are each comprised of non-conserved residues (Figure 5.2, white residues and inset images). These observations suggest that Rv1155 and Rv2991 could bind different classes of ligands and may be functionally divergent enzymes within an evolutionarily-related class. The structural similarities and differences of Rv1155 and Rv2991 make them interesting model systems for exploring the applicability of fragments to probe the substrate specificity of related enzymes.

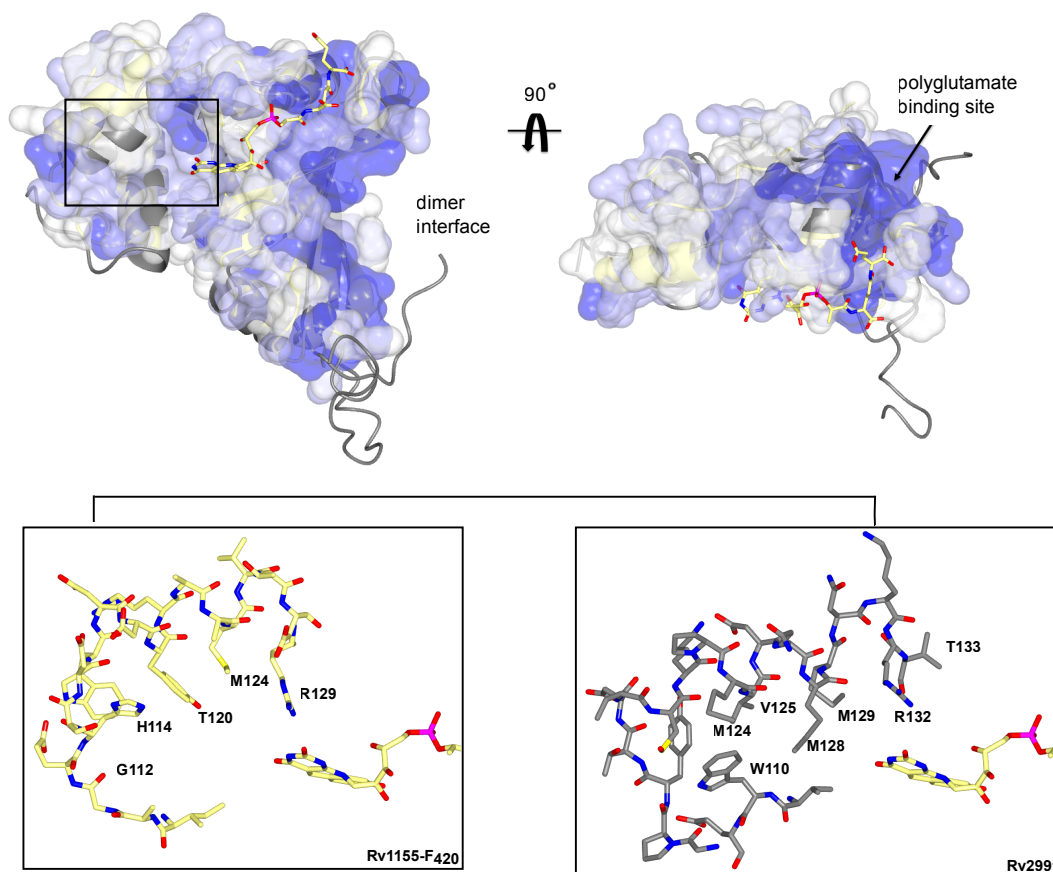


Figure 5.2. Structural comparison of F_{420} -dependent Rv1155 and Rv2991. Rv1155 (yellow) with bound F_{420} (C atoms) overlaid with Rv2991 (grey) (PDB ID: 1RFE). Residues with surface representation are colored according to sequence similarity between Rv2991 and Rv1155. Dark blue residues are conserved, light blue residues have a conserved residue type, and white residues are not conserved. The small domain adjacent to the F_{420} ring system (the predicted substrate binding site) is comprised of residues that are not conserved. The Rv1155- F_{420} and Rv2991 active site are shown in the inset images. F_{420} is present in the Rv1155 structure only. The image was generated using CCP4mg¹²⁵.

5.5 Library design and development

A fragment library to target F_{420} -dependent Rv1155 and Rv2991 was designed based on known natural ligands bound to flavoproteins deposited in the PDB. The ligands of FMN-, FAD-, or F_{420} -dependent proteins in the PDB were selected, which led to a set of 810 compounds. This set was further filtered

based on molecular weight (>5 heavy atoms), solubility (>500 μM in water), and by excluding compounds containing inorganic atoms. Protein precipitants and crystal cryoprotectants were removed with a custom Pipeline Pilot²⁰³ structure filter and drug molecules were manually excluded. The filtering process retained 350 molecules, which comprise a diverse dataset of natural ligands to known flavoproteins. This work was done in collaboration with Dr. Andreas Bender and Sean Hudson (Department of Chemistry, University of Cambridge).

This set of query molecules was employed for similarity searching in the ChemBridge, ChemDev, Ryan Scientific, Enamine, Life Chemical, ChemDiv, and Specs libraries obtained from the ZINC²⁰¹ database and filtered based on molecular weight and solubility (<300 Da and >1 mM). ECFP4 fingerprints were employed as descriptors, which belong to the class of circular fingerprints that were shown to capture structure-bioactivity relationships rather well in previous studies²⁰⁴. The similarity coefficient chosen was the Tversky coefficient²⁰⁵ as implemented in PipelinePilot (due to the assumption that fragments represent substructures of the ligands from the PDB set). The Tversky coefficient is calculated as shown in the equation below.

$$S(Tversky)_{A,B} = \frac{c}{\alpha * a + \beta * b + c}$$

Here, 'A' refers to a library compound while 'B' refers to a natural flavoprotein ligand from the query dataset derived from the PDB. The variable a describes the number of molecular features only in A (and not in B), while b describes molecular features only in B (and not in A). The variable c is the number of features in common between A and B²⁰⁰. The weighting factors were set to $\alpha = 1$ and $\beta = 0$ in order to bias the selection of fragments towards those that are as much a complete substructure of the query compounds as possible, while omitting those parts of the experimental ligands that are not captured by the fragment from the calculation, given that a fragment is meant to represent only a

part of the high-affinity binder. A complete list of all fragments in the target library can be found in Appendix B.

5.6 Fragment screen of Rv1155 and Rv2991

Rv2991 and Rv1155 were each screened against the flavoprotein fragment library in the presence and absence of F_{420} first by DSF and then by STD NMR. These techniques are described in Chapter 1. A fragment-only control screen revealed that 13% (9/70) of the fragments in the library are not assayable by DSF using SYPRO® Orange dye system due to the intrinsic fluorescence of these compounds.

For Rv2991- F_{420} , 46% (28/61) of fragments in the library increased the T_m by least twice the standard deviation of the DMSO control (0.4). Please refer to Appendix C, Table C1 for full results. Among the fragments that did not increase the T_m of the protein, 15% (9/61) decreased the T_m of the protein by more than 1 °C, 38% (23/61) of fragments had no effect on the T_m , slightly increased it, or slightly lowered it ($\pm <1$ °C), and 1 fragment produced a complex, multi-transition melt curve. Screening the same fragment library against Rv1155- F_{420} revealed that only 15% (9/61) of fragments increased the T_m by least twice the standard deviation of the DMSO control (0.6). The T_m of Rv1155 was significantly lowered (>1 °C) in the presence of 38% (23/61) of fragments. There was little change in the T_m of Rv1155 with 46% (28/61) of fragments ($\pm <1$ °C), and 1 fragment produced an uninterpretable melt curve. Because most fragments either increased or did not significantly decrease the T_m of Rv1155 or Rv2991, all 70 fragments were analysed for binding by STD NMR.

A screen of the flavoprotein library against Rv2991 by STD NMR revealed fragment hits in the presence and absence of F_{420} with a hit rate of 17% (12/70) (Figure 5.3). Complete NMR spectra for all fragment hits against Rv2991 can be found in Appendix C. Binding was observed only in the presence of F_{420} for 25% of the binding fragments (Figure 5.3 C). For fragments that showed binding by STD in both presence and the absence of F_{420} , enhanced binding

signal was observed when F_{420} was present in the reaction mixture (Figure 5.2 A). Thermal unfolding studies with Rv2991 indicate that the Rv2991- F_{420} complex is significantly stabilised relative to the apoenzyme (Chapter 4, Figure 4.4), which may increase the amount of functional protein available to engage in a binding event. For some fragments, the π -stacking interaction between the deazaisalloxazine ring system of F_{420} and the fragment may be required for the fragments to bind. For all fragments screened, no STD signal was observed in the presence of F_{420} and fragment only.

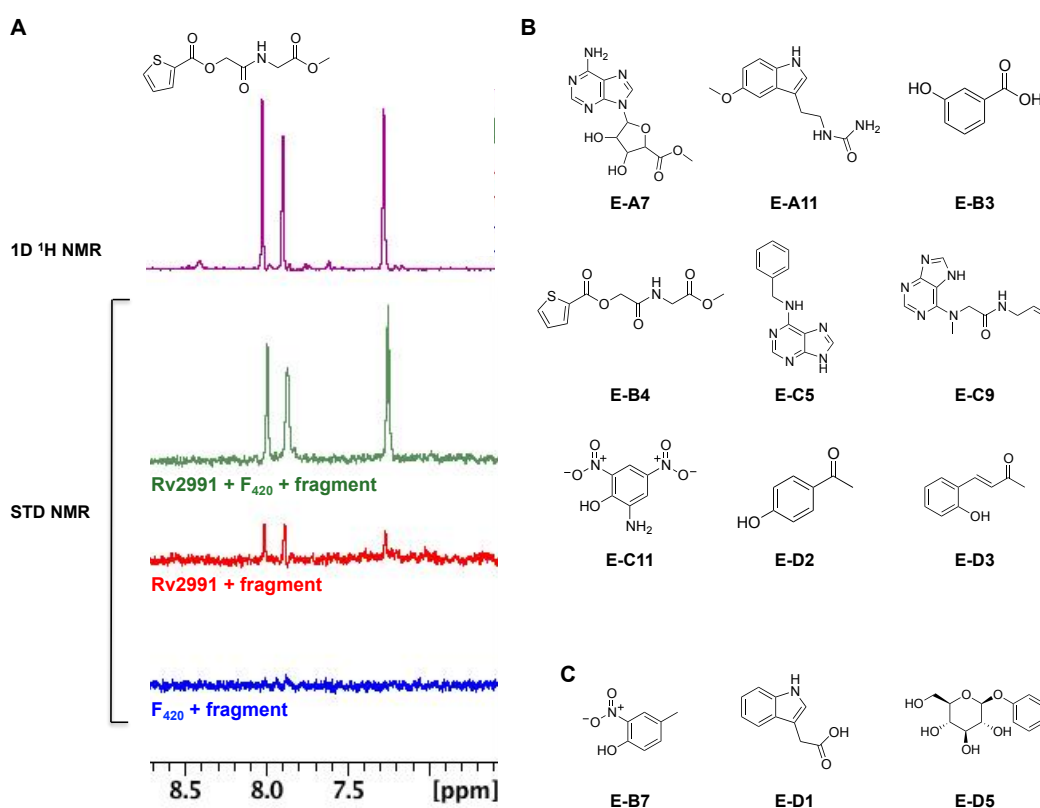


Figure 5.3. STD NMR fragment hits for Rv2991. (A) Example STD NMR spectra for Rv2991 with fragment E-B4. Where it is present, the reaction components exist in the following concentrations: Rv2991 (25 μM), fragment (1 mM), F_{420} (50 μM). One dimensional ^1H NMR spectrum of aromatic proton resonances (purple); STD NMR spectrum of reaction mixture containing Rv2991, F_{420} , and fragment E-B4 (green); STD NMR spectrum of reaction mixture containing Rv2991 and fragment E-B4 (red); STD NMR spectrum of reaction mixture containing F_{420} and fragment E-B4 only (blue). Fragment E-B4 binds in both the presence and absence of F_{420} , although the signal is stronger in the presence of F_{420} . (B) Fragments binding to Rv2991 in both the presence and absence of F_{420} . (C) Fragments binding to Rv2991 in the presence but not in the absence of F_{420} . The flavoprotein fragment library contains 70 fragments, which were all screened.

There were significantly fewer clear STD NMR fragment hits for both apo- and holoenzyme Rv1155 (Figure 5.4) than there were for Rv2991. Only 2 fragments from the targeted flavoprotein library were binding to Rv1155 by NMR (hit rate: 3%) and none of these showed binding to the protein in the absence of F_{420} . The structural basis for this relatively lower hit rate is unclear. It may be that the chemical space occupied by the flavoprotein fragment library insufficiently samples potential Rv1155 binders. To expand the repertoire of fragments binding to Rv1155, a commercially-available fragment library containing 1249 fragments (Maybridge Fragment Libraries) was screened against Rv1155 by DSF (see Appendix C, Table C.2 for full results). Of the fragments in the library, 74 increased the T_m beyond twice the standard deviation of the fragment-negative control. These 74 fragments were further screened by STD NMR and two additional ligand-observed NMR experiments, CPMG and WaterLOGSY (Appendix C, Figure C.3). This screening process yielded 9 fragment hits (Figure 5.4 C) that share some structural similarities. The Maybridge library fragment hits against Rv1155 are specific to Rv1155; they do not bind Rv2991 as assessed by using DSF. Unlike Rv2991, no fragments have been discovered for Rv1155 that bind without F_{420} . No binding signal is observed for these fragments in the absence of F_{420} . Future structural studies with Rv1155 and Rv2991 will elucidate the basis of the differential selectivity observed between these two proteins.

A comparison of DSF and NMR results for Rv1155 and Rv2991 reveals that there is limited overlap in hits between the two techniques. For example, some fragments that had little effect on the T_m of the protein showed clear binding by NMR. Conversely, some fragments that seemed to increase the T_m were not hits by STD. This result is consistent with observations that orthogonal techniques for fragment screening do not always agree, particularly if DSF is used^{206,79}. Each technique can identify unique hits²⁰⁶. Because ligand-observed NMR spectroscopy is a more established screening technique for fragments⁵², STD NMR fragments hits were prioritised.

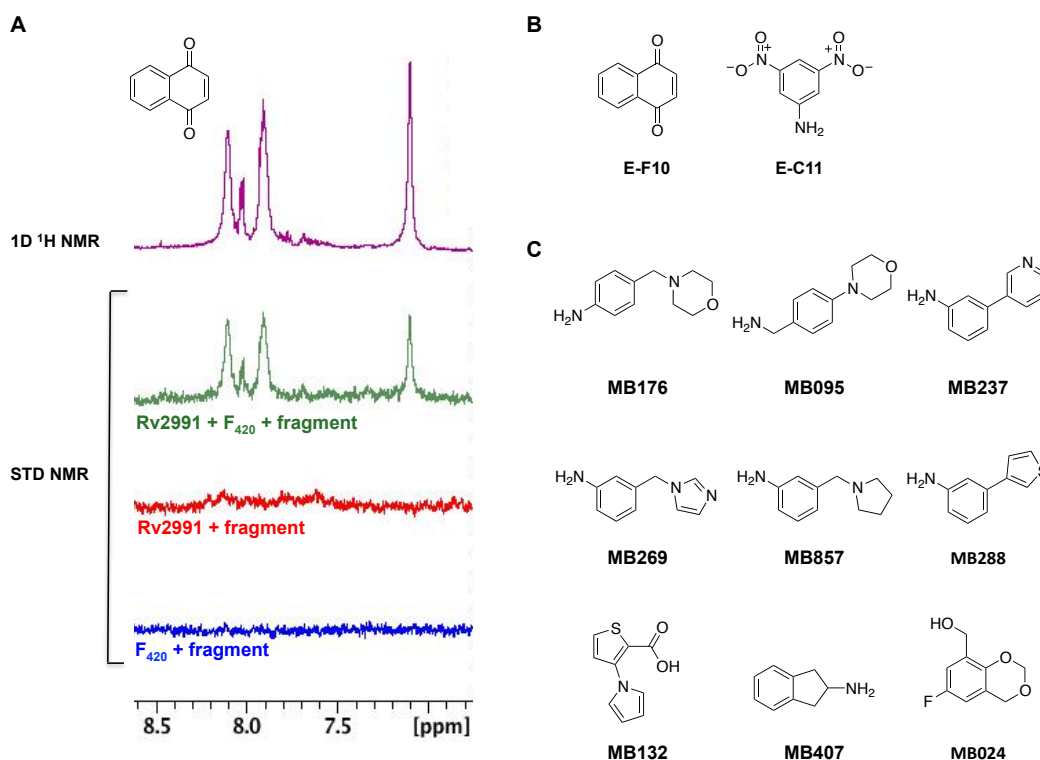


Figure 5.4. STD NMR fragment hits for Rv1155. (A) Example STD NMR spectra for Rv1155 with fragment E-F10. Where present, reaction components exist in the following concentrations: Rv1155 (25 μM), fragment (1 mM), F₄₂₀ (50 μM). One dimensional ^1H NMR spectrum of aromatic proton resonances (purple); STD NMR spectrum of reaction mixture containing Rv1155, F₄₂₀, and fragment E-F10 (green); STD NMR spectrum of reaction mixture containing Rv1155 and fragment E-F10 (red); STD NMR spectrum of reaction mixture containing F₄₂₀ and E-F10 only (blue). Fragment E-F10 does not bind in the absence of F₄₂₀. (B) Fragment hits against Rv1155-F₄₂₀ from flavoprotein fragment library (total number of fragments screened: 70). Fragment hits bind in the presence, but not the absence of F₄₂₀. (C) Fragment hits against Rv1155-F₄₂₀ from the Maybridge fragment library. Total number of fragments screened by STD NMR: 74 (fragments selected for screening by DSF). Fragment hits bind in the presence, but not the absence of F₄₂₀.

5.7 Virtual fragment linking results and preliminary enzyme assays

VFL is a computational approach that analyses fragment screening data for a given target to predict molecules with greater affinity for that target²⁰⁰. It has been described as an *in silico* method to ‘bridge the gap’ between low micromolar or millimolar fragments and nanomolar ligands. The technique is predicated on generating statistical models based on the chemical features, or ‘molecular

fingerprints', of binding fragments. These models are then used to screen large libraries containing compounds that are more chemically complex than fragments. 'VFL hits' are predicted to bind the target with high affinity because they have a combination of molecular fingerprints highly represented in the set of binding fragments. The VFL approach has been used in the pharmaceutical industry to discover potent HTS hits based on fragment binding data. The schematic shown in Figure 5.5 illustrates the VFL approach using fragment and HTS data for the target nitric oxide synthase (iNOS).

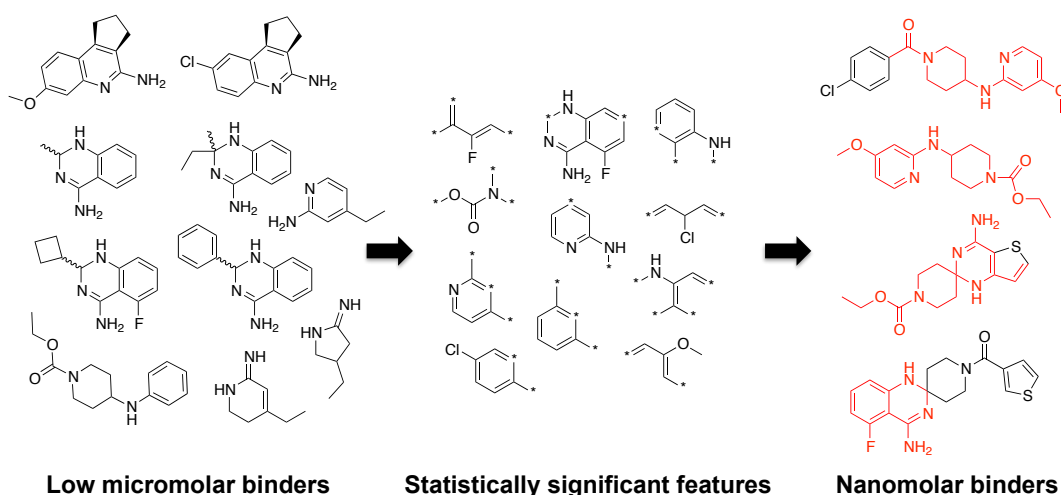


Figure 5.5. Schematic demonstrating the VFL approach. Fragments binding to iNOS (low micromolar binders) are analysed to determine statistically significant features in the set (middle). This information is used to create models that can predict nanomolar binders from an HTS library. The substructures highlighted in red are represented in the fragment set. Figure adapted from Crisman *et al. J. Med. Chem.* 2008²⁰⁰.

VFL was used to create statistical models of the 12 metabolite-like fragments binding to Rv2991. In order to identify potential substrates for Rv2991, the models were compared to all small molecules found in *Mtb* cells. A curated library of *Mtb* metabolites does not currently exist for academic use. Therefore, a dataset of metabolites was compiled using substrates in the KEGG database²⁰² (<http://www.genome.jp/kegg/>), which included 3856 molecules. For the 12 fragments binding to Rv2991 versus the 58 non-binding fragments, a Bayesian

model based on Molprint 2D fingerprint was generated²⁰⁷. The features differentiating the active and inactive datasets (74) were used to generate a second Bayesian model, which was used to screen the *Mtb* metabolome dataset. This work was done in collaboration with Dr. Andreas Bender (Department of Chemistry, University of Cambridge). Of the 120 top VFL hits, 80 were selected because they were small molecules involved in intermediary metabolism. The remaining 80 compounds were queried using KEGG Ligand SIMCOMP²⁰⁸ (<http://www.genome.jp/tools/simcomp/>) to identify all known enzymatic reactions associated with each substrate. Six compounds out of the 80 were purchased as preliminary set of test substrates based on their predicted redox properties (see Chapter 6, Table 6.3).

A high-throughput spectrophotometric assay was developed to assess Rv2991 enzymatic activity. Because each test substrate has a unique spectrophotometric signature, the oxidation state of the cofactor was monitored. It reported that oxidised F_{420} can be detected sensitively with fluorescence spectroscopy²⁰⁹. In a pilot experiment, the test substrates were incubated with Rv2991 and the reduced form of the cofactor, $F_{420}H_2$. Reduced cofactor was generated enzymatically as reported previously¹⁶⁹ using F_{420} -dependent glucose 6-phosphate dehydrogenase 1 (FGD-1), which oxidises glucose 6-phosphate (G6P) and thereby produces $F_{420}H_2$ (Figure 5.6 A; see Chapter 6, Figure 6.3 for the enzymatic reduction of F_{420}). FGD-1 is removed from the reaction mixture prior to incubation with Rv2991. Of the 6 predicted substrates tested, one demonstrated Rv2991-dependent re-oxidation of $F_{420}H_2$ over time: 5-methoxytryptamine (5-MT), an intermediate in tryptophan metabolism (Figure 5.6 B). Substrate, cofactor, and enzyme, concentration dependence of this activity was observed (Figure 5.6 C-F). Preliminary structure activity relationship (SAR) experiments indicate that only tryptamine (TP) also exhibits activity (Figure 5.6 B & G). Similar rates of $F_{420}H_2$ oxidation are observed when either 5-MT or TP are used as substrates.

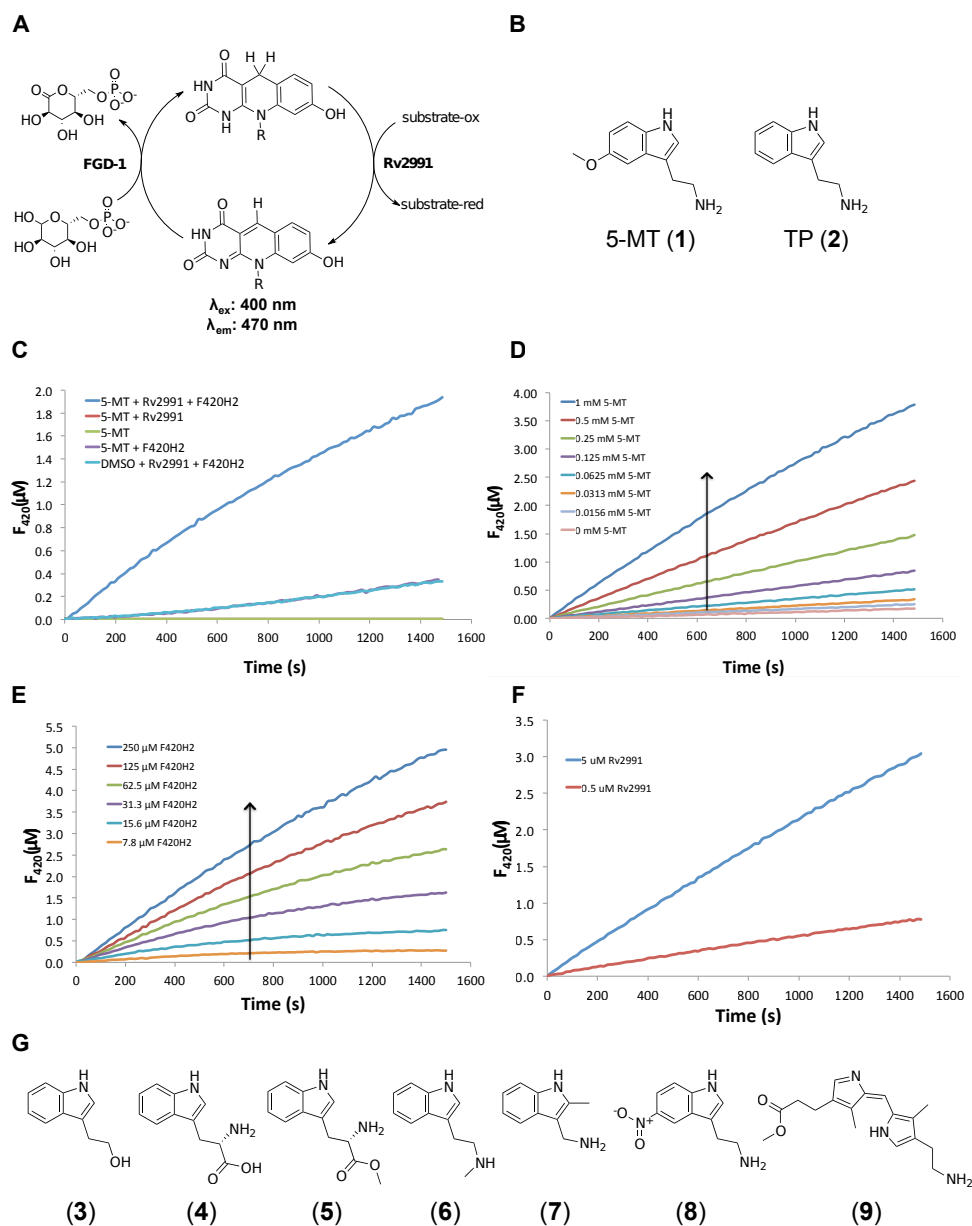


Figure 5.6. Rv2991- $F_{420}H_2$ enzyme assay for predicted substrates. (A) Schematic representing the fluorescence-based assay to test the enzymatic activity of Rv2991 with test substrates predicted by VFL. (B) Active compounds 5-MT and TP, which show similar enzymatic profiles. (C) Time-dependent oxidation of $F_{420}H_2$ in the presence of Rv2991 and 5-MT (blue trace). No activity is observed for control reactions without Rv2991 (green and purple traces), $F_{420}H_2$ (red and green traces), or 5-MT (cyan trace). Reaction components, where present, are at the following concentrations: Rv2991 (5 μM), $F_{420}H_2$ (50 μM), and 5-MT (1 mM). (D) 5-MT concentration dependence on the rate of F_{420} formation. (E) $F_{420}H_2$ concentration dependence on the rate of F_{420} formation. (F) Rv2991 concentration dependence on the rate of F_{420} formation. (G) Inactive tryptamine analogues.

The enzyme, cofactor, and substrate dependence on the rate of $F_{420}H_2$ oxidation are promising indicators of enzymatic activity. The reaction appears to be dependent on Rv2991- F_{420} , as no activity was observed when Rv1155 was used rather than Rv2991. Further, no change was observed when NADPH (λ_{ex} : 340 nm λ_{em} : 450 nm)²¹⁰ was used as an electron carrier, instead of $F_{420}H_2$. However, downstream experiments to validate and characterise this activity have been unsuccessful. The product of the reaction (with either 5-MT or TP as substrate) cannot be identified by LC-MS and 1D 1H NMR spectroscopy. Further, these methods show little decrease in the substrate signal over time. A major complicating factor is that the rate of the observed activity is very slow. Although the activity appears to be enzymatic based on the relevant controls, the k_{cat} and K_M of this reaction are 0.0033 s^{-1} and 1.5 mM , respectively, for 5-MT and 0.0020 s^{-1} and 1.0 mM , respectively, for TP. This exceedingly slow conversion rate requires high concentrations of enzyme and cofactor, which impose a number of technical challenges on LC-MS and NMR-based assays. Preliminary SAR studies (Figure 5.6 G, compounds **2-9**) have failed to identify tryptamine analogues that exhibit a higher rate of enzymatic activity. These results provide interesting clues about the enzymatic properties of Rv2991, but the rate of the reaction is too slow to be considered biologically-relevant. Six of the 80 predicted VFL substrates have been tested for enzymatic activity. Additional VFL-predicted substrates must be screened to arrive at substrates that are turned over at faster rates.

5.8 Conclusions and future work

The work presented in this chapter demonstrates some foundational, proof-of-concept studies to develop a 'fragment-to-substrate' method. A targeted fragment library for flavoproteins was designed and assembled. Biophysical techniques were employed to detect fragments binding to two structurally related F_{420} -binding proteins, Rv2991 and Rv1155, with unknown function. Fragments bound

to each protein, but fewer fragment hits were observed for Rv1155, as compared to Rv2991. A high hit rate for Rv2991 was observed (17%) against the flavoprotein fragment library. Bayesian models were created based on the hits against Rv2991, which were then used to screen all known metabolites in the *Mtb* cell. Among the VFL-predicted substrates tested, two demonstrated enzymatic activity via a fluorescence-based assay that monitors the redox state of the cofactor. This enzymatic activity, however, falls short of rates that are expected from natural substrates. However, this procedure will be used in future studies to screen additional predicted substrates for activity.

Chapter 6

Experimental

6.1 Protein overexpression and purification

6.1.1 Purification of Rv2607

The clone of Rv2607 was obtained as an amino-terminal 6x-histidine (His_{6x}) tag fusion construct with a thrombin cleavage site in a modified pET28b vector from the laboratory of Dr. Thomas C. Terwilliger, Los Alamos National Laboratories, New Mexico, USA¹¹⁷. The construct also included a kanamycin resistance gene for selection. Rv2607 was transformed in *E. coli* BL21-Gold(DE3) cells (Stratagene). An overnight culture of a single colony grown at 37 °C with agitation was diluted 1:100 into 500 mL Luria Broth (LB) supplemented with 40 µg/mL kanamycin. The culture was incubated at 37 °C with shaking. Protein

expression was induced when the culture reached an OD_{600nm} of 0.6 with the addition of isopropyl thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The culture was further incubated at 20 °C for 20 hrs with agitation. Cells were harvested by centrifugation (14000 xg, 15 min). The bright yellow cell pellet was resuspended in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl. The cells underwent one freeze-thaw cycle (-80 °C) and were further lysed by sonication (Misonix S-4000, amplitude 60%, 1 s on, 1 s off for 30 s, repeated 3 times). The cell lysate was centrifuged (48,000 xg, 30 min) and the clarified yellow supernatant was applied to a Ni-NTA affinity column (5 mL) pre-equilibrated with buffer A (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 20 mM imidazole). The column was washed with equilibration buffer (150 mL) until the flow through no longer contained any protein, as determined by a Bradford test¹³⁶. The protein was eluted with elution buffer (20 mM Tris-HCl, 500 mM NaCl, 300 mM imidazole, pH 8.0). The eluted protein was buffer exchanged into phosphate buffered saline (PBS), pH 7.6, containing 10% glycerol and 1 mM β -mercaptoethanol (β -ME) using 10 kDa molecular weight cut off centrifugal filters (Millipore). SDS-PAGE analysis showed that the protein was purified to >95% homogeneity and the yield of the purified protein was determined to be 140 mg per 1 L of cells with the Bradford assay.

6.1.2 Purification of Rv1155 and Rv2991

The expression clone of Rv1155 was obtained as an aminoterminal His_{6x}-MBP fusion construct with TEV cleavage site from Dr. Yves Bourne (Architecture et Fonction des Macromolécules Biologiques, CNRS, Marseille, France)¹¹⁸ and the clone of Rv2991 was obtained as an aminoterminal His_{6x} fusion construct with TEV cleavage site from the laboratory of Prof. Pedro Alzari (Institut Pasteur, Paris, France). Rv1155 and Rv2991 were each transformed in *E. coli* BL21(DE3) cells. An overnight culture of a single colony from each transformation was diluted 1:100 into LB (2 L) supplemented with 100 μ g/mL ampicillin. The cultures

were incubated with shaking at 37 °C. Protein expression was induced when the cultures reached an OD_{600nm} of 0.4-0.6 with the addition of IPTG to a final concentration of 1 mM. The cultures were further incubated at 25 °C for 16 hrs with agitation. The cells were then harvested by centrifugation and were resuspended in equilibration buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM imidazole) and subjected to sonication on ice. The lysate was cleared by centrifugation (50 min, 10000 xg) and the supernatant was loaded onto HiTrap HP nickel affinity column 5 mL (GE Healthcare) pre-equilibrated with equilibration buffer. The column was washed with equilibration buffer (20x column volume). The target proteins were eluted with elution buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 300 mM imidazole). His_{6x}-MBP-Rv1155 and His_{6x}-Rv2991 protein were buffer exchanged into a buffer optimized for the TEV protease cleavage reaction (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT) using 10 kDa centrifugation filters (Millipore). The tags were cleaved from each fusion by incubating TEV protease with Rv1155 and Rv2991 at 1:100 molar ratio each at 25 °C for 2 hrs. The TEV and any uncleaved protein were removed by loading the reaction mixture onto a second HiTrap HP nickel affinity column with a 50 mL superloop (GE Healthcare). The eluate was retained and any uncleaved Rv1155 or Rv2991 bound on the resin electrostatically was removed with high salt buffer (50 mM Tris-HCl, pH 7.5, 600 mM NaCl, 10 mM imidazole). The eluted protein was pooled and purified by SEC with a 16/60 Superdex-75 column (GE Healthcare) pre-equilibrated with 50 mM potassium phosphate buffer, pH 7.5, 100 mM NaCl for Rv1155 and 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 20 mM dithiothreitol (DTT) for Rv2991. Fractions eluted from column were analysed by SDS-PAGE. Fractions containing pure Rv1155 or Rv2991 (corresponding to a molecular weight of ~16 kDa and ~18 kDa, respectively) were pooled and concentrated with 10 kDa cutoff centrifugation filters (Millipore). The protein concentration was determined by Abs_{280} with the calculated extinction coefficient based on the untagged sequence, $18450 \text{ cm}^{-1} \text{ M}^{-1}$ for Rv1155 and $27960 \text{ cm}^{-1} \text{ M}^{-1}$

for Rv2991. The yield was determined to be 10 mg per L of cells for Rv1155 and 14 mg per 1 L of cells for Rv2991.

6.2 Biophysical characterisation of Rv2607, Rv1155, and Rv2991 by MS

6.2.1 Confirmation of molecular weight

The calculated and observed MW of Rv2607, Rv1155, and Rv2991 can be found in the Table 6.1. The calculated molecular weight was based the translated DNA sequencing results. All samples were prepared for MS by desalting with C18 Zip Tip® (Millipore) pipette tips according to the manufacturer's specifications and eluted in a solution of acetonitrile and 0.05% trifluoroacetic acid (TFA). The samples were analysed by electrospray ionisation positive ion mode with a Q Star Elite Mass Spectrometer (Applied Biosystems/MDS Sciex).

6.2.2 MS analysis of oligomerisation state and cofactor stoichiometry

A Synapt HDMS system optimised for the transmission of non-covalent complexes was used to record nESI-MS spectra²¹¹. All samples were prepared in 100 mM ammonium acetate. Rv2607 (20 μ M) was analysed with co-purified FMN (no exogenous FMN was added). Rv2991 (12 μ M) incubated with F₄₂₀ prior to analysis (1:5 molar ratio of enzyme:coenzyme). Rv1155 (10 μ M) was prepared as described in Section 6.6. Each protein sample (2.5 μ L) was electrosprayed from gold-coated glass capillaries²¹². To preserve the non-covalent interactions between the protein and cofactor, the MS parameters used for the Synapt were: capillary voltage, 1.7 kV; sample cone, 80-100 V; trap and transfer collision energy, 15 V and 12 V respectively; backing pressure 3.8 mbar; trap and IMS pressure $5e^{-2}$ and $5e^{-1}$ mbar, respectively; time-of-flight analyser pressure $1.16e^{-6}$ mbar. For the Rv2607 collision-induced dissociation (CID) experiment, the 'trap and transfer collision energies' were 150 V and 110 V respectively to

affect dissociation. Data were processed with MassLynx 4.0 software (Waters/Micromass).

Protein	Calculated MW (Da)	Observed MW (Da)	Comments
Rv2607	27493.8	27438.0	The experimental value is consistent with the molecular weight of the His _{6x} -Rv2607 fusion protein (27493.8 Da), minus the aminoterminal methionine (-131 Da), and with an added molecule of β -Me (+76 Da).
Rv1155	16226.3	16227.0	The calculated molecular weight of Rv1155 is based on the WT sequence of Rv1155 (untagged) without the aminoterminal methionine (-131 Da) and with an added glycine (+74), which is retained from TEV cleavage.
Rv2991	18032.4	18034.0	The calculated molecular weight of Rv2991 is based on the WT sequence of Rv1155 (untagged) without the aminoterminal methionine (-131 Da) and the carboxyterminal proline and with an added glycine (+74), which is retained from TEV cleavage.

Table 6.1. Electrospray ionisation MS analysis Rv2607, Rv1155, and Rv2991 proteins used in the biophysical and enzymatic assays. The experimentally-observed MW of each protein is consistent with the calculated MW, which is derived from the expected protein sequence based on DNA sequencing data.

6.2.3 Analysis of co-purified cofactor from Rv2607

Extracted cofactor from Rv2607 and authentic FMN were analyzed with a Q Star Elite Mass Spectrometer (Applied Biosystems/MDS Sciex) by electrospray ionisation positive ion mode with a gold-plated nanospray tip. To extract the bound cofactor from Rv2607, a concentrated solution of purified Rv2607 (270 μ M) in 25 mM potassium phosphate buffer, pH 7.8, was heat-denatured for 5 min at 100 °C. To the denatured protein was added 300 μ L 25 mM potassium

phosphate buffer, pH 7.8, and the mixture was filtered using a 10 K molecular weight cutoff Vivaspin 500 centrifugation filter (15 min at 17,900 xg). The flow through was desalted with C18 Zip Tip® pipette tips (Millipore) previously washed with acetonitrile and equilibrated with 0.1% TFA in ultrapure water. The analyte was eluted in a solution of acetonitrile and 0.05% TFA. The FMN standard sample was prepared by acidifying it in 0.1% TFA to a final concentration of 500 µM. All spectra were calibrated internally using a solution of cesium iodide (100 mg/mL).

6.3 Enzymatic characterisation of Rv2607

6.3.1 Synthesis of PNP

PNP was synthesised based on a published procedure²¹³ with some modifications. Under argon atmosphere, freshly prepared 0.5 M NaOH (3.2 mL) was added to PLP (200 mg, 0.8 mmol). The solution was deoxygenated by bubbling with argon for 5 min at room temperature (25 °C), after which sodium borohydride (12 mg, 0.32 mmol) was added portionwise with stirring. The reaction was stirred until complete as judged by TLC (SiO₂, 80:20 methanol:ethyl acetate v:v). The reaction was quenched with acetone (120 µL, 1.6 mmol) and allowed to stir at 25 °C for 30 min, after which NaOH (147 mg, 3.7 mmol) was added. After stirring for another 15 min at 25 °C, the reaction mixture was loaded directly onto a column containing 5 mL Amberlite IRA-743 (5 mL in a 1.0 x 11.0 cm column) to remove borate salts. The resin had been conditioned by washing with absolute ethanol (25 mL), water (3 x 25 mL), 1 M HCl (25 mL), water (25 mL), 1 M NaOH (25 mL), and finally equilibrated with water (50 mL). The flow through was collected and the resin was washed with enough water (5-7 mL) so that PNP could no longer be detected spectrophotometrically ($\lambda_{\text{max}} = 325 \text{ nm}$ at pH 7.5)²¹⁴. The flow-through and wash were pooled and this solution was further purified with weakly acidic cation exchange resin Amberlite IRC-76 (10 mL in a

1.30 x 12.5 cm column). The resin had been conditioned with 1 M NaOH (50 mL), water (50 mL), 1 M HCl (50 mL), and finally equilibrated with water (100 mL). The flow through was collected and pooled with two elution steps: 0.5 M NaOH (5 mL) and water (10 mL). The resin was washed with water (10 mL) until PNP could not be detected spectrophotometrically. The eluted product was pooled, lyophilised, and resuspended in 100 mM potassium phosphate buffer, pH 7.8. ^1H NMR (D_2O , 400 MHz): δ 7.77 (1H, s), 4.88 (2H, d, $J = 6.9$), 4.77 (2H, s), 2.41 (3H, s).

6.3.2 LC-MS assay for Rv2607-dependent PLP formation

Proposed Rv2607 substrates, PNP and PMP, and the product, PLP, co-migrated by reverse-phase high-pressure liquid chromatography (HPLC). In order to resolve the product and substrates, an adduct was formed between PLP and DNPH. DNPH reactivity with PLP, PNP, and PMP was tested. To 1 mM solutions of PLP (100 μL), PNP (100 μL), and PMP (100 μL) in 25 mM potassium phosphate buffer (pH 7.8), a saturated solution of DNPH in 3 M HCl (10 μL) was added to a final concentration of 0.7 mM. As a negative control, 3 M HCl (10 μL) was substituted for DNPH. The PLP, PNP and PMP negative control reactions were carried out by substituting the substrate with 100 μL of 25 mM potassium phosphate buffer, pH 7.8. All reactions were carried out at 25 $^\circ\text{C}$ for 3 hrs.

Reaction mixtures (150 μL each) containing 1 mM PNP or 1 mM PMP, 10 μM FMN, and 10 μM Rv2607 in 25 mM potassium phosphate buffer were incubated at 25 $^\circ\text{C}$ for 3 hrs. Enzyme, cofactor, and substrate negative control reactions were treated in a similar manner. After 3 hrs incubation at 25 $^\circ\text{C}$, the reaction was quenched with addition of 10 μl of DNPH in 3 M HCl to a final concentration of 0.7 mM. The reaction mixture was filtered with 10 kDa MW cutoff centrifugation filters and 10 μl of the reaction was analyzed by LC-MS.

Analysis of the enzymatic reactions and controls were performed on Agilent 1100 LC-MS instrument using Luna (3 μm) C18 column (50 x 2 mm,

Phenomenex). Solution A contained water and 0.1% formic acid. Solution B contained acetonitrile and 0.1% formic acid. The column was equilibrated with 95% solution A for 1 min. The following gradient was used: 5% to 95% solution B from 1 to 11 min; 95% solution B from 11 to 15 min; 95% to 5% solution B from 15 to 17 min; 5% solution B from 17 to 21.5 min.

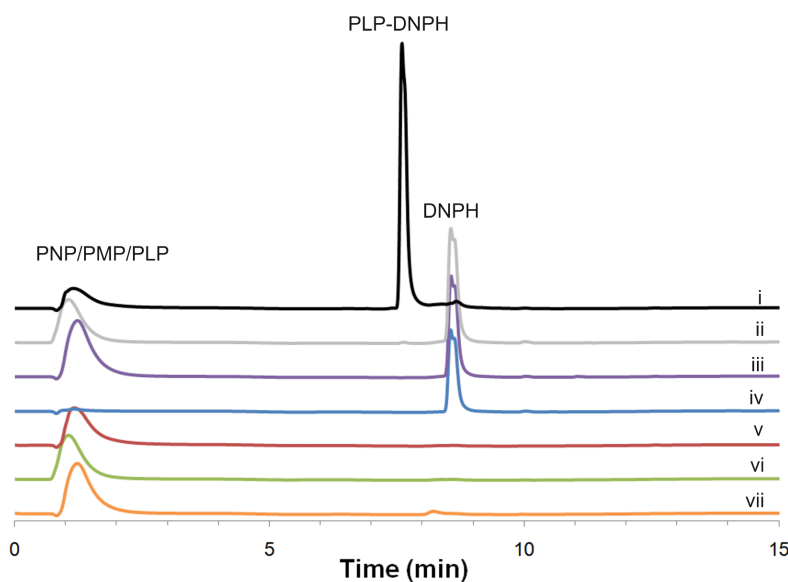


Figure 6.1. Reverse-phase HPLC analysis of control reactions between PLP, PNP, or PMP, and DNPH. HPLC chromatograms (270 nm) of reaction mixtures containing PLP and DNPH (i), PMP and DNPH (ii), PNP and DNPH (iii), DNPH (iv), PLP (v), PMP (vi), and PNP (vii). Reaction components, were at the following final concentrations: 1 mM PLP, PMP, PNP and 0.7 mM DNPH. All reactions were carried out in 25 mM potassium phosphate buffer, pH 7.8 at 25 °C for 3 hrs. The identity of each peak was assigned by electrospray ionisation in the positive ion mode (PLP-DNP $m/z = 428$).

6.3.3 NMR time-course assay of Rv2607-dependent PLP formation

The enzymatic reaction was monitored by ^1H NMR spectroscopy using a Bruker Avance 700 MHz spectrometer with a TXI cryoprobe. The reaction mixture contained 1 mM PNP, 10% deuterium oxide, and 20 μM 2,2,3,3-D₄-3-(trimethylsilyl)propionic acid (TSP-d₄) as an internal standard in 25 mM potassium phosphate buffer, pH 7.6. The reaction was initiated with the addition

of enzyme to a final concentration of 28 μM or an equivalent volume of buffer for the control. The enzymatic reaction and enzyme negative control (200 μL) were added to 3 mm capillary tubes (Bruker), which were maintained at constant temperature (25 $^{\circ}\text{C}$). The spectrometer conditions were optimised with the enzyme negative control sample. The sample was locked, shimmed, and the pulses were calibrated for water suppression using excitation sculpting with gradients²¹⁵. At selected time points, ^1H NMR spectra were collected for 18 hrs. Each experiment consisted of 4 dummy scans followed by 8 collected scans, resulting in 63 s of acquisition time per data point. Chemical shifts are reported in ppm relative to TSP ($\delta = 0$ ppm). The spectra were processed and analysed using TopSpinTM 1.8 (Bruker, MA, USA). Substrate conversion was determined by comparing integrals of the C2- ^1H signals associated with PLP (7.80 ppm) to that of PNP (7.75 ppm). To confirm the identity of the product formed, the reaction mixture was enriched with authentic PLP after 18 hrs.

6.3.4 Steady state kinetic characterisation of Rv2607

The rate of product formation was monitored spectrophotometrically with a Varian Cary 400 UV/Vis spectrophotometer 1 cm pathlength quartz cuvettes (Varian). The product, PLP, has a λ_{max} at 388 nm and $4900 \text{ cm}^{-1}\text{M}^{-1}$ is the extinction coefficient for that wavelength²¹⁴. The rate of PLP production was measured for various concentrations of PNP: 50 μM , 100 μM , 175 μM , 250 μM , 500 μM , 750 μM , 1000 μM , 1500 μM . No exogenous FMN was added. All solutions were made in 100 mM potassium phosphate buffer, pH 7.8. The K_{M} and k_{cat} for Rv2607 were determined by fitting the initial rate (less than 10% conversion) of product formation as a function of substrate concentration to the Michaelis-Menten equation by non-linear regression using GraphPad Prism 5 (GraphPad Software, Inc., La Jolla, California, USA).

6.4 Test enzymatic reactions with Rv1155 and Rv2991

All activity assays were conducted using Varian Cary 300 Bio UV/Vis Spectrophotometer in 1 cm pathlength quartz cuvettes (Varian). Where present, the reagents were at the following concentrations: 1 μM enzyme (Rv1155 or Rv2991), 100 μM substrate (PNP, PMP, coumarin, 7-hydroxycoumarin, 7-hydroxy-4-methylcoumarin, imperatorin, PA-824), 20 μM cofactor (F_{420} or FMN). Reaction mixtures (150 μl total) contained: 1) enzyme, substrate, and cofactor, 2) enzyme and substrate, 3) substrate and cofactor, 4) enzyme and cofactor. All reactions were carried out in 25 mM potassium phosphate buffer, pH 7.5. The reactions were initiated with the addition of enzyme and a wavelength scan (200-500 nm) was collected every 2 min for 1 hr.

6.5 Purification of F_{420}

6.5.1 Isolation and purification of F_{420} from *M. smegmatis*

The protocol for F_{420} purification was developed by modifying existing procedures^{172,173}. *M. smegmatis* cells were transformed with genes *fbiA*, *fbiB*, and *fbiC*, which were obtained in a pYUBDuet vector from Prof. Edward N. Baker (University of Auckland, New Zealand). The genes *fbiA*, *fbiB*, and *fbiC* encode enzymes involved in the biosynthesis of F_{420} ²¹⁶. It was confirmed that *M. smegmatis* cells transformed with the vector containing *fbiABC* have about five times higher F_{420} production levels than do the wild type (WT) cells (Figure 6.2). The vector was electroporated into the *M. smegmatis* strain mc²155 as previously described¹⁷³.

M. smegmatis cells overexpressing the F_{420} biosynthetic genes were grown in kilogram quantities in fermenters in LB broth. When the culture reached an $\text{OD}_{600\text{nm}}$ of 0.6, IPTG was added to a final concentration of 0.25 mM and further incubated for 2 days. For one round of purification, typically around 800 g

of *M. smegmatis* transformed with *fbiABC* was thawed at 4 °C and resuspended in a minimal volume of 25 mM potassium phosphate buffer, pH 7.5, being careful to avoid large clumps of cells (~1.6 L). The resuspension was autoclaved for 20 min, cooled at 4 °C, and centrifuged at 40,000 xg for 50 min. The bright yellow supernatant was filtered with a 0.45 µm pore membrane filter (Millipore). The cell pellet was again resuspended in a minimal volume of 25 mM potassium phosphate buffer, pH 7.5 (~1 L) and the mixture was again autoclaved, centrifuged, and filtered as before.

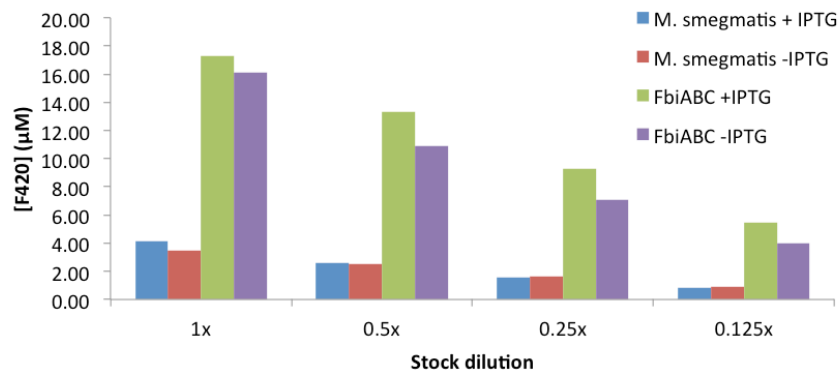


Figure 6.2. F_{420} concentration found in *M. smegmatis* cells transformed with FbiABC overexpression vector compared to WT. Transformants with either *fbiABC* or a control vector were cultured (5 mL) at 37 °C with shaking in the presence of kanamycin (50 µg/mL) until an OD_{600nm} of 0.6 was reached. An aliquot of each culture was removed and treated with IPTG (0.25 mM final concentration) and further incubated at 37 °C with shaking for 4 days. The cells were harvested by centrifugation and the wet cell pellets were weighed to normalise the fluorescence signal to the mass of the cell pellet. The cells were resuspended in 1 mL 100 mM potassium phosphate buffer pH 7.5, autoclaved for 20 min, centrifuged, and the supernatant was analysed by fluorescence spectroscopy (λ_{ex} : 400 nm, λ_{em} : 470 nm). The concentration of F_{420} present was quantified with a calibration curve of a known standard of F_{420} . Each sample was measured at 4 dilutions. There is approximately 5.5x greater F_{420} production in the +*fbiABC* strain in the presence of IPTG.

The supernatant from both autoclave cycles was pooled and loaded onto a quaternary aminoethyl ion exchange column equilibrated with 25 mM potassium phosphate buffer pH 7.5 buffer at flow rate of 0.5 mL/min at 4 °C. The column was then washed at a constant flow rate of 1 mL/min with 25 mM potassium phosphate buffer, pH 7.5, each containing increasing concentrations of NaCl. The eluate was monitored spectrophotometrically at $\lambda = 420$ nm (the

λ_{\max} of F_{420} at pH 7.5) and was collected in 500 mL fractions. Each eluted fraction was analysed by LC-MS. The column was first washed with 100 mM NaCl in potassium phosphate buffer (2 L). FO, a precursor to F_{420} , was eluted with 250 mM NaCl in potassium phosphate buffer (2 L). The column was further washed with 2.0 L 400 mM NaCl in potassium phosphate buffer. F_{420} eluted from the column with 2.0 L 600 mM NaCl potassium phosphate buffer. The presence of F_{420} species with 5 glutamates, F_{420-5} (m/z: 1160), and with 6 glutamates, F_{420-6} (m/z: 1290) was detected by LC-MS.

The fractions containing F_{420} were removed from the high salt buffer using solid phase extraction (SPE) C18 cartridges according to the manufacturer's protocol (Waters SEP-Pak Long). Conductivity measurements indicated that the SPE cartridges removed 97% of salts in the sample. The desalted F_{420} fractions were lyophilised, dissolved in a minimal volume of water.

The F_{420} was further purified by reverse-phase HPLC on a Varian Prep Star instrument using a Phenomenex Luna C18 (250 x 21.20 mm, 10 μ m) column. Solution A was water and solution B was acetonitrile. The column was equilibrated with 95% solution A and 5% solution B for 4 minutes. The following linear gradient was used: 95% to 80% solution A and 5% to 20% solution B from 2 to 32 min, then 80% to 5% solution A and 20% to 95% solution B from 32 to 35 min, and 5% solution A and 95% solution B for 5 min. The fractions containing F_{420} were pooled, lyophilised, and dissolved in a minimal volume of water. The concentration of F_{420} was determined with the extinction coefficient²¹⁷ at 420 nm, which is 25.9 mM⁻¹ cm⁻¹. The product purity was assessed by LC-MS. Yield was determined to be 104 mg per 800 g *M. smegmatis* + *fbiABC* cells.

6.5.2. Enzymatic reduction of F_{420}

Reduced F_{420} ($F_{420}H_2$) was prepared enzymatically using recombinant F_{420} -dependent glucose 6-phosphate dehydrogenase 1 (FGD-1) based on published protocols^{169,147}. F_{420} was typically reduced at 50-250 μ M concentrations with

FGD-1 and excess and glucose 6-phosphate (G6P) in 50 mM potassium phosphate buffer, pH 7.5 without exposure to light (total reaction volume 100-500 μ L). Before use, FGD-1 was removed from the reaction mixture by filtering with a 10 kDa cutoff centrifugation filter (Millipore).

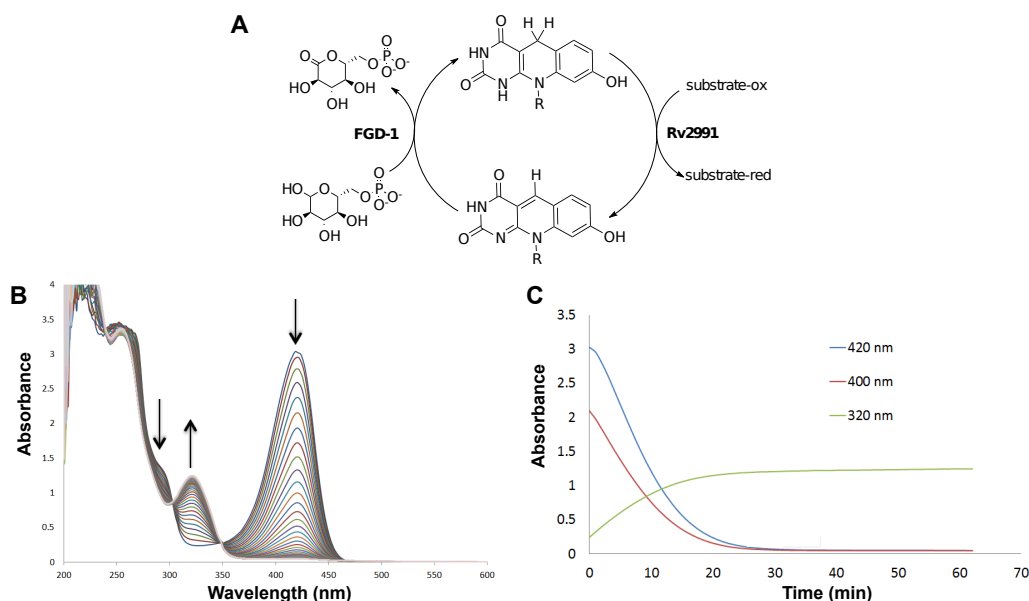


Figure 6.3. Enzymatic reduction of F_{420} by FGD-1. (A) Schematic representing FGD-dependent reduction of F_{420} in the presence of G6P. (B) UV/Vis scans (200-500 nm) of F_{420} reduction catalysed by FGD-1. The reaction mixture contained 100 μ M F_{420} , 6.25 mM G6P, and FGD-1 and was monitored for 1 hr with a scan taken every min. (C) Time-dependent change in select wavelength of F_{420} in the presence of FGD-1 and G6P. A decrease in the absorbance at 420 nm and 400 nm and concurrent increase in the absorbance at 325 nm is observed as $F_{420}H_2$ is formed.

6.6 Protein crystallography

Rv1155 and Rv2991 were prepared for crystallographic screens by overexpressing, isolating, and cleaving the His_{6x}-MBP-Rv1155 and His_{6x}-Rv2991 as described in section 6.1.2. For crystallographic studies with Rv1155, after the second nickel affinity column to remove TEV and any tagged protein, the protein was incubated with F_{420} (1:1 ratio) at 4 °C overnight. The protein was then purified by SEC with Superdex 75 10/300 GL (GE Healthcare), pre-equilibrated with 10 mM Tris-HCl, pH 7.5, 150 mM NaCl. Prior to injecting the protein, the

equilibration buffer containing 25 μM F_{420} (15 mL) was loaded onto the column. The eluate was monitored at two wavelengths ($\lambda_{F_{420}} = 420$ nm, $\lambda_{F_{420}, \text{protein}} = 280$ nm) and collected in 1 mL fractions (refer to Figure 6.4 for a chromatogram of Rv1155 co-purified with F_{420} by SEC).

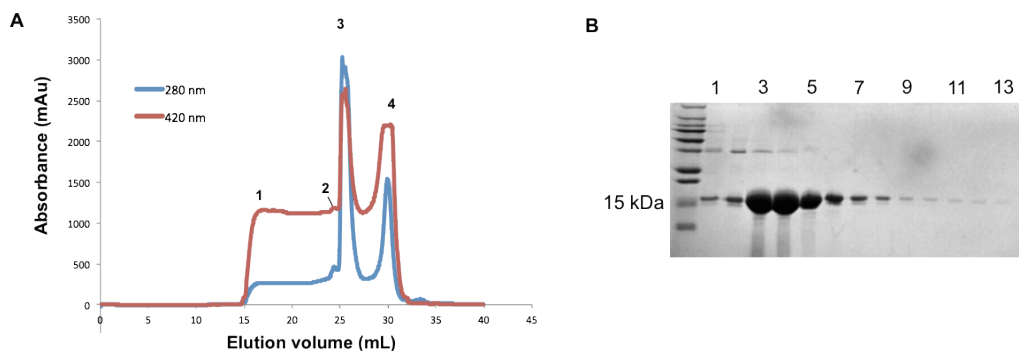


Figure 6.4. Chromatogram and SDS-PAGE analysis of Rv1155- F_{420} co-purification by SEC. (A) Chromatogram of Rv1155 co-eluting with F_{420} . Wavelengths monitored are 420 nm (red) and 280 nm (blue). (B) SDS-PAGE analysis of purified fractions. F_{420} begins to elute at 15 mL (peak 1). A small contaminant protein elutes from the column (peak 2), which corresponds to fractions 1 and 2 in panel B. Peak 3 is the co-migrating complex of Rv1155 and F_{420} (fractions 3–7 in panel B), as the absorbance at 280 nm is enhanced by the presence of F_{420} . Excess F_{420} is eluted from the column at ~30 mL. The SDS-PAGE (9–13) shows the absence of protein in these fractions.

The peak corresponding to the co-elution of Rv1155 and F_{420} (peak 3) was pooled, concentrated to 500 μM , and used to screen commercially-available crystallisation conditions (384 in total) at a ratio of 1:1 protein to screening condition by sitting drop vapor diffusion (Crystal Phoenix, Art Robbins Instruments LLC, CA, USA). A number of conditions produced crystal hits, which are summarised in Table 6.2.

As discussed in Chapter 4, Section 4.4.2, crystals obtained for the Rv1155- F_{420} complex had only one molecule of F_{420} bound, even though there are two coenzyme binding sites in the Rv1155 homodimer and nESI-MS analysis indicates that it predominantly binds two molecules of F_{420} . The purification procedure described above produces Rv1155- F_{420} at about 1:1 molar ratio of enzyme to coenzyme. Future studies will be directed towards optimising the

conditions shown in Table 6.2 and screening for new crystallisation conditions for Rv1155-F₄₂₀, where F₄₂₀ is added in molar excess of Rv1155.

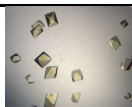
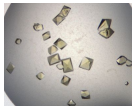
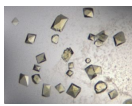
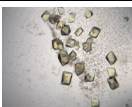
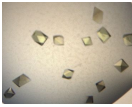
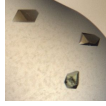
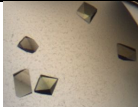
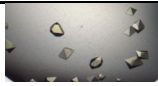
Condition	Example Crystal image
0.1 M MES pH 6.5, 25% PEG 3000	
0.1 M MES pH 6.5, 25% PEG 4000	
0.1 M MES pH 6.5, 25% PEG 6000	
0.1 M MES pH 6.5, 25% PEG 8000	
0.2M NaF, 20% PEG 3350	
0.2M KF, 20% PEG 3350	
0.2M sodium formate, 20% PEG 3350	
0.2M Potassium formate, 20% PEG 3350	

Table 6.2. Crystallographic conditions that produced crystals for Rv1155-F₄₂₀. All conditions were identified in a 96 well commercial crystallographic screen, PEGs Suite (Qiagen). Images were taken by Rock Imager 500 (Formulatrix, Inc, Waltham, MA, USA).

Co-crystallisation studies with Rv2991 and F₄₂₀ have been unsuccessful in producing diffraction-quality crystals of the enzyme-coenzyme complex. However, apoenzyme crystals of Rv2991 could be reproduced based on known

crystallisation conditions (Alzari, P. *et al.*, unpublished results), which demonstrates that the protein is properly folded and is suitable for crystallisation trials. Figure 6.5 shows example Rv2991 apoenzyme crystals, which formed in the presence of 100 mM TRIS-HCl, pH 8.5, 200 mM sodium acetate, 20% PEG-4000, and 10 mM MgCl₂. Optimisation of the protein to mother liquor ratio dramatically affects the morphology of the crystals. A ratio of 1:1 or 1:2 protein to crystallisation condition yields thin needles that appear after 1 day (Figure 6.5 A). Larger needles were observed with 2:1 ratio of protein to mother liquor. Long three-dimensional crystals were obtained with a ratio of 1:3 protein to crystallisation condition in the drop (Figure 6.5 B), which appear after about 3 weeks.

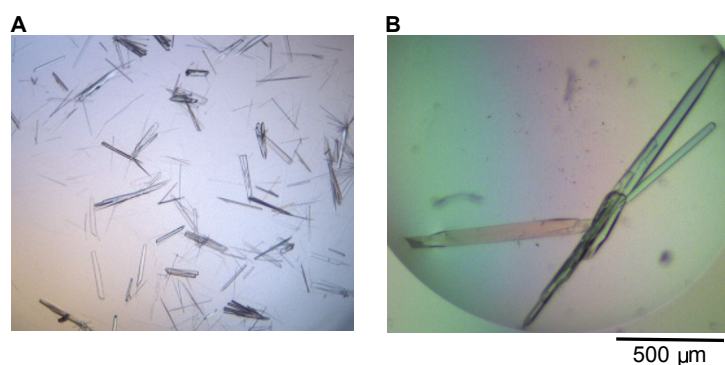


Figure 6.5. Optimisation of apoenzyme Rv2991 crystals. (A) Rv2991 apoenzyme crystals obtained from hanging drops at a ratio of 1:1 protein to mother liquor in the drop. (B) Rv2991 apoenzyme crystals obtained from hanging drops at 1:3 protein to mother liquor in the drop. Crystallisation condition: 100 mM TRIS-HCl, pH 8.5, 200 mM sodium acetate, 20% PEG-4000, and 10 mM MgCl₂.

6.7 Isothermal titration calorimetry

Untagged Rv1155 and Rv2991 were prepared as described in 6.1.2 and underwent dialysis overnight at 4 °C with stirring in 600 mL 100 mM potassium phosphate buffer, pH 7.4, 20 mM NaCl with 10 KDa MW cutoff Slide-A-Lyzer dialysis cassette (ThermoScientific, MA, USA). For Rv2991, 5 mM DTT was added to the dialysis buffer. F₄₂₀ was dialysed into the same buffer. In general using 500 Da MW cutoff Float-A-Lyzer dialysis cassette (Spectrum Laboratories,

Inc, CA, USA). The dialysate was retained to rinse the instrument before use. For the Rv1155-F₄₂₀ titration, 10 μ M Rv1155 and 275 μ M were used and for the Rv2991-F₄₂₀ titration, 51.3 μ M Rv2991 and 1.15 mM F₄₂₀.

All titrations were performed on a MicroCal iTC₂₀₀ calorimeter (GE Healthcare). In general, experimental parameters were set as follows: 17 1.5 μ L injections (0.5 μ L for first injection) with 3 s duration (1 s for the first injection), 5-7 μ cal/s reference power, 4 s filter period, and 100-180 s spacing. Data was analysed and figures were generated using Origin (OriginLab Corp, MA, USA).

6.8 Optimisation of DSF experiments

Optimal conditions for DSF experiments with Rv1155 and Rv2991 were determined by varying buffer type (potassium phosphate buffer and TRIS-HCl), pH (7.0-8.5), ionic strength (0-350 mM NaCl), protein concentration (1 μ M-10- μ M), and SYPRO® Orange dye concentration (1x-90x). Because high baseline fluorescence signal was observed in the Rv2991 melt curve, a number of additives were also tested, such as 1-15% DMSO, 1-15% dioxane, and 1-30 mM DTT. The conditions that produced the optimal melt curve for Rv1155 were 100 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2.5x SYPRO® Orange dye concentration, and 5 μ M protein. An interpretable melt curve was obtained for Rv2991 under the following conditions: 20 mM potassium phosphate buffer, pH 7.5, 20 mM NaCl, 1x SYPRO® Orange dye, 5 μ M protein, and 20 mM DTT. All samples (100 μ L total volume) were tested in triplicate and incubated over a temperature range of 37-90 °C with a ramp rate of 0.01 °C/s, collecting a scan every 1 s. The data was analysed using Protein Thermal Shift v1.0 (Applied Biosciences, CA, USA). The T_m was determined by calculating the negative derivative of the melt curve and identifying the local minima⁷⁷.

6.9 Fragment library design, development, and screening

6.9.1 Fragment library design

The targeted fragment library was designed based on ligands of known flavoprotein (FMN-, FAD-, and F₄₂₀-binding). Ligands from each of the targeted classes were obtained from the PDB (<http://www.rcsb.org/pdb>). Ions, gases, buffer additives, metal complexes were excluded from the flavoprotein ligands. The filtering process retained 1500 substrates. Commercially-available fragment-like molecules were created by compiling the ChemBridge, ChemDiv, Enamine, Life Chemicals, Ryan Scientific and Specs supplier databases from the ZINC repository (<http://zinc.docking.org/>, Bioinformatics and Chemical Informatics Research Center, University of California San Francisco, CA, USA) and filtering by predicted solubility in water >1 mM and MW <300 Da using Pipeline Pilot v6.1 (Accelrys, CA, USA). The commercial fragments that represented substructures of the ligands were identified by calculating Tversky coefficients in Pipeline Pilot using the extended connectivity molecular fingerprint (ECFP4) generated for each molecule. Representative fragments (70) were ordered to create the focused flavoprotein library.

6.9.2 Fragment library construction

Approximately 10 mg quantities of each fragment were weighed using an analytical balance and deposited in removable glass vials that are stored in a 96 well plate holder with an air-tight seal. Enough DMSO-d₆ was added to the fragment to yield a solution of 100 mM final concentration. The solution was pipetted up and down repeatedly to mix. The plate was covered and stored at 25 °C overnight. The following morning, the plates were stored at -80 °C. As a quality control measure, a 1D ¹H NMR spectra of each fragment in buffer was obtained.

6.9.3 NMR screen of fragment library

Reaction mixtures (200 μ L) contained 1 mM fragment with and without 25 μ M Rv2991 or 25 μ M Rv1155 were prepared in 50 mM Tris-HCl buffer at pH 7.5, 50 mM NaCl, 10% (v/v) D₂O, 20 μ M deuterated TSP-d₄ and 2.5% (v/v) DMSO-d₆. All fragments were screened in the presence and absence of 50 μ M F₄₂₀. The samples were then pipetted into 3-mm NMR capillaries (Bruker) and placed in 5 mm NMR tubes (Wilmad-LabGlass, NJ, USA). STD and WaterLOGSY 1D ¹H NMR spectra were acquired at 278 K on a Bruker DRX 700 MHz NMR spectrometer (Bruker, MA, USA) equipped with a 5 mm triple resonance inverse (TXI) cryoprobe with z-gradients. STD experiments employed a 40 ms selective Gaussian 180° pulse at a frequency alternating between 'on-resonance' (0.9 ppm) and 'off-resonance' (40 ppm) after every scan. WaterLOGSY experiments employed a 40 ms selective Gaussian 180° shaped pulse at the water signal frequency and a NOE mixing time of 1 s. Water signal suppression was achieved using a W5 Watergate gradient spin-echo pulse sequence. The spectra were processed using TopSpin 3.0 software (Bruker UK, Coventry, UK) and WaterLOGSY spectra were scaled relative to the TSP-d₄ peak intensity at 0 ppm.

6.9.4. DSF screen of fragment library

Samples (100 μ L total volume) were prepared in optical 96-well reaction plates (Applied Biosystems, CA, USA), each containing 5 mM fragment, 5 μ M Rv2991, 2.5x SYPRO® Orange (Invitrogen, OR, USA), in the presence or absence of 50 μ M F₄₂₀. Fluorescence data were collected on 7500 Real Time PCR with instrument software v2.0.6 (Applied Biosciences, CA, USA) using the ROX filter (λ_{ex} : 587 nm, λ_{em} : 607 nm). The data was analysed using Protein Thermal Shift v1.0 (Applied Biosciences, CA, USA). The T_m was determined by calculating the negative derivative of the melt curve and identifying the local minima⁷⁷.

6.10 Rv2991 substrate prediction

6.10.1. Virtual fragment linking

In order to generate a dataset of putative metabolites, the *Mtb* H37Rv metabolome was derived from the KEGG database²⁰², comprising a total of 3,856 endogenous metabolites with valid structural entries. For the 12 binding fragments, a Bayesian model based on Molprint 2D fingerprint was generated²⁰⁷ using circular fingerprints with a radius of 2. The 74 features possessing highest information gain between the binding fragments and non-binding fragments were used to generate a Bayesian model, which was used to screen the *Mtb* metabolome dataset. The 120 endogenous metabolites scoring highest in the model were subject to human inspection. This approach is analogous to previous 'Virtual Fragment Linking' efforts, where it could be shown that information about fragment binding can be joined to retrieve small molecules with high bioactivity from a molecular database²⁰⁰.

6.10.2 Test enzymatic assays with Rv2991

Test enzymatic reactions (100 μ L total volume) contained Rv2991 (5 μ M), F₄₂₀H₂ (50 μ M), 1 mM test substrate dissolved in DMSO (1% DMSO). Enzyme, coenzyme, and substrate negative control reactions were prepared by adding an equivalent volume of buffer or DMSO in the place of the reagent. The preliminary set of test substrates are shown in Table 6.3. Samples were prepared in 100 mM potassium phosphate buffer, pH 7.4, 50 mM NaCl. F₄₂₀ was reduced enzymatically as described in Section 6.5.2 using FGD-1, which was removed by 10 kDa MW cutoff filtration before it was added to the reaction mixture. All reaction mixtures were monitored by fluorescence spectroscopy (λ_{ex} = 400 nm, λ_{em} = 470 nm) to detect the formation of oxidised F₄₂₀.

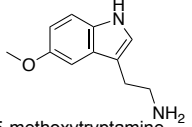
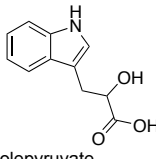
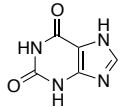
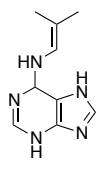
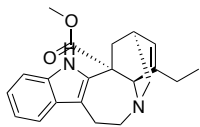
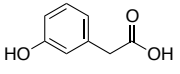
Compound ID	Structure and name	Metabolic pathway	Example reduction-oxidation reaction
ES-1	 5-methoxytryptamine	Tryptophan metabolism Melatonin degradation	$\text{Tryptamine} + \text{Oxygen} + \text{NADPH} \rightleftharpoons$ $\text{N-Hydroxyl-tryptamine} + \text{H}_2\text{O} + \text{NADP}^+$
ES-2	 Indolepyruvate	Tryptophan metabolism	$\text{Indolepyruvate} + \text{NADH} + \text{H}^+ \rightleftharpoons$ $\text{Indolelactate} + \text{NAD}^+$
ES-3	 Xanthine	Purine metabolism	$\text{Xanthine} + \text{NADH} + \text{H}^+ \rightleftharpoons$ $\text{Hypoxanthine} + \text{NAD}^+ + \text{H}_2\text{O}$
ES-4	 N6-(delta2-Isopentenyl)-adenosine	Biosynthesis of secondary metabolites Zeatin biosynthesis	$\text{N6-(delta2-Isopentenyl)-adenosine 5'-monophosphate} + \text{Reduced flavoprotein} + \text{Oxygen}$ $\rightleftharpoons \text{trans-Zeatin riboside monophosphate} + \text{Oxidized flavoprotein} + \text{H}_2\text{O}$
ES-5	 Catharanthine	Indole alkaloid biosynthesis Biosynthesis of secondary metabolites	$\text{Geissoschizine} + \text{NADP}^+ \rightleftharpoons$ $4,21\text{-Dehydrogeissoschizine} + \text{NADPH}$
ES-6	 3-Hydroxyphenylacetate	Tyrosine metabolism	$3\text{-Hydroxyphenylacetate} + \text{Oxygen} + \text{NADPH} + \text{H}^+$ $\rightleftharpoons \text{Homogentisate} + \text{NADP}^+ + \text{H}_2\text{O}$

Table 6.3. Select VFL hits that were tested for enzymatic activity with Rv2991 and F₄₂₀H₂. These potential substrates were selected because they participated in known reduction-oxidation reactions, as documented by KEGG database.

One substrate, 5-methoxytryptamine (5-MT), demonstrated time-dependent re-oxidation of F₄₂₀ in the presence of enzyme, substrate, and coenzyme, but not in the absence of any reagent components (named compound ES-1 in Table 6.3). To confirm that this change is enzymatic, a range of concentrations of substrate (1000-15.6 μM), enzyme (5-0.5 μM), and coenzyme

(250-7.8 μM) were tested. Analogues of 5-MT were tested for activity as described above. K_M and k_{cat} for the 5-MT and TP reactions were determined by fitting the initial as a function of substrate concentration to the Michaelis-Menten equation by non-linear regression using GraphPad Prism 5 (GraphPad Software, Inc., La Jolla, California, USA).

Appendix A

Crystallographic refinement statistics

Data Collection	DS1	DS2	DS3
X-ray source	Grenoble	Grenoble	Grenoble
Resolution limit (Å)	30-2.0 (2.07-2.00)*	30-2.6 (2.69-2.60)	30-1.9 (1.97-1.90)
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions (Å)	55.3, 66.5, 77.0	54.5, 66.2, 76.8	54.5, 66.2, 76.8
Completeness (%)	94.7 (99.9)	99.9 (100.0)	99.8 (99.8)
<I/σ(I)>	26.97 (6.13)	40.08 (15.07)	25.00 (5.98)
Wilson B-factor (Å ²)	18.55	48.37	25.64
Refinement statistics			
Resolution limits (Å)	30-2.00(2.07-2.00)	30-2.6 (2.69-2.60)	30-1.9 (1.97-1.90)
Number of reflections in working set	17238 (1780)	8209 (793)	20694 (2033)
Number of reflections in test set	1469 (162)	746 (65)	1772 (166)
Rwork (%)	25.0 (26.0)	24.1 (32.0)	21.8 (27.9)
Rfree (%)	29.0 (31.0)	29.9 (43.0)	24.8 (28.6)
Number of atoms:			
Protein; solvent; ligand/ion	2122; 89; 39	2103; 0; 0	2156; 148; 72
Geometry r.m.s deviations:			
Bond lengths (Å)	0.006	0.007	0.006
Bond angle (°)	1.62	1.40	1.34

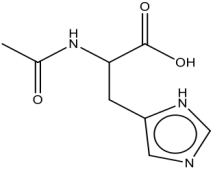
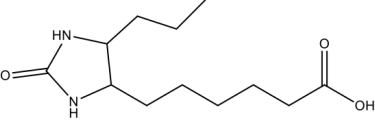
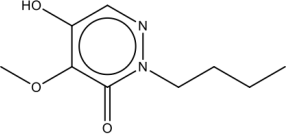
Table A.1. Crystallographic refinement statistics for Rv1155-F₄₂₀ datasets 1-3. Highest resolution shell is shown in parenthesis. Rmerge, multiplicity of measurements, and mosaicity are not shown, as unmerged reflections and diffraction images are no longer available. All datasets contain F₄₂₀.

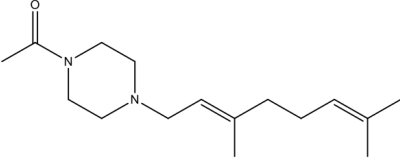
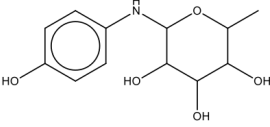
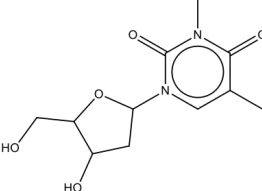
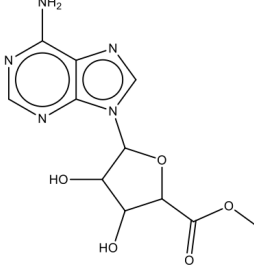
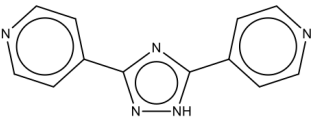
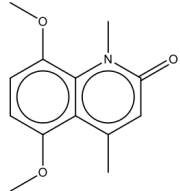
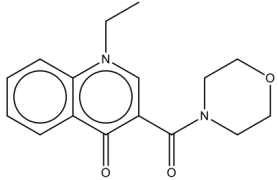
Data Collection	DS4	DS5
X-ray source	Grenoble	Grenoble
Resolution limit (Å)	40-2.4 (2.49-2.40)	35-2.8 (2.90-2.80)
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions (Å)	55.3, 66.5, 76.9	54.0, 66.8, 76.6
Completeness (%)	99.5 (100.0)	100.0 (100.0)
<I/σ(I)>	38.20 (14.44)	31.42 (6.80)
Wilson B-factor (Å ²)	36.95	66.34
Refinement statistics		
Resolution limits (Å)	40-2.4 (2.49-2.40)	35-2.8 (2.90-2.80)
Number of reflections in working set	10567 (1027)	6498 (623)
Number of reflections in test set	923 (90)	612 (66)
Rwork (%)	21.8 (28.5)	20.9 (32.2)
Rfree (%)	28.4 (42.7)	27.4 (40.1)
Number of atoms:		
Protein; solvent; ligand/ion	2144; 45; 47	2148; 12; 54
Geometry r.m.s deviations:		
Bond lengths (Å)	0.007	0.007
Bond angle (°)	1.30	1.35

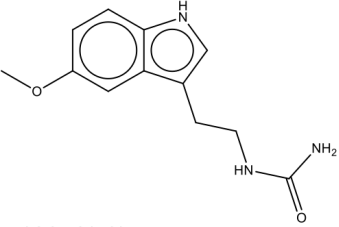
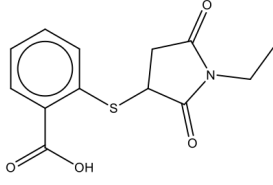
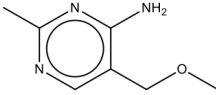
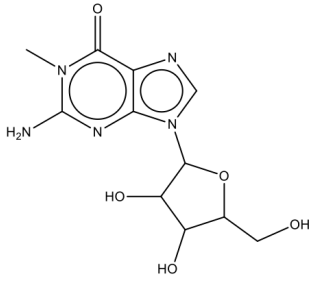
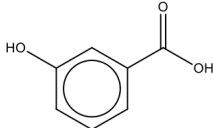
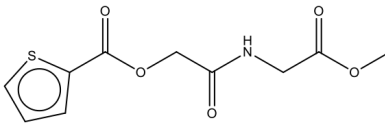
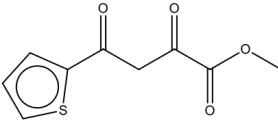
Table A.2. Crystallographic refinement statistics for Rv1155-F₄₂₀ datasets 4-5. Highest resolution shell is shown in parenthesis. Rmerge, multiplicity of measurements, and mosaicity are not shown, as unmerged reflections and diffraction images are no longer available. All datasets contain F₄₂₀.

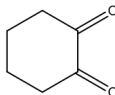
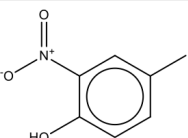
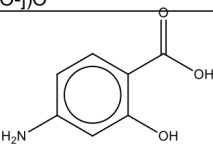
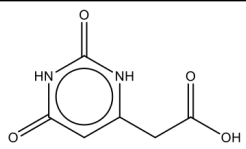
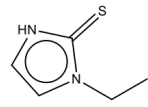
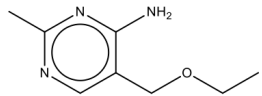
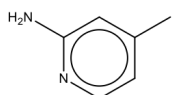
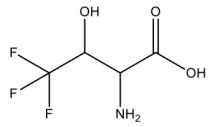
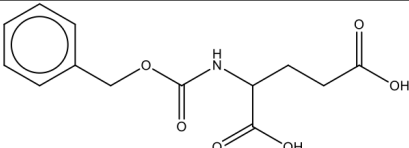
Appendix B

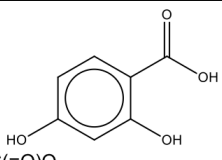
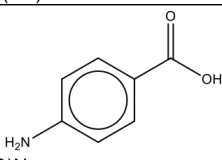
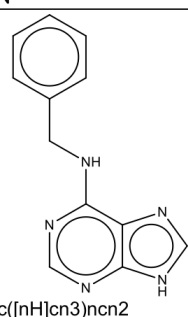
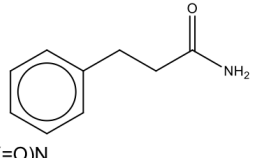
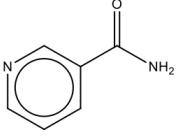
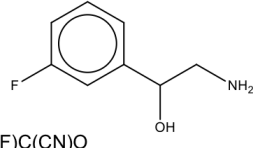
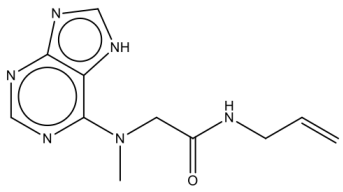
Targeted enzyme fragment library

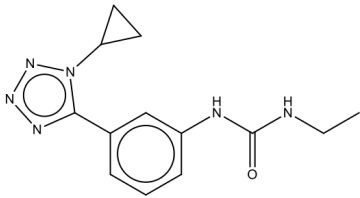
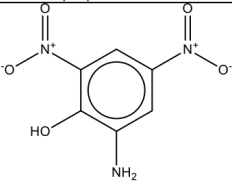
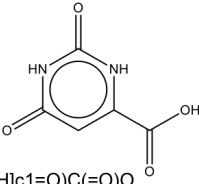
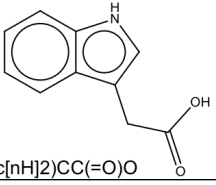
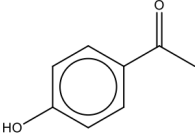
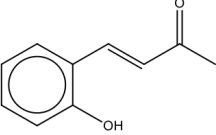
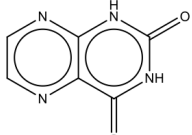
ID	Plate Location	Fragment structure and SMILES string
E-A1	A1	 <chem>CC(=O)NC(Cc1cnc[nH]1)C(=O)O</chem>
E-A2	A2	 <chem>CCCC1C(NC(=O)N1)CCCCC(=O)O</chem>
E-A3	A3	 <chem>CCCCn1c(=O)c(c(cn1)O)OC</chem>

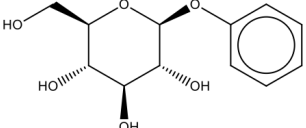
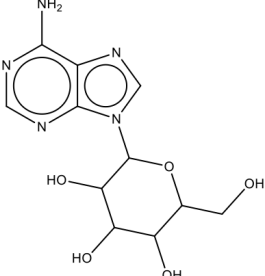
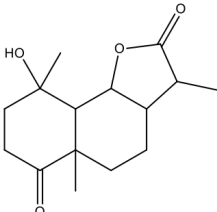
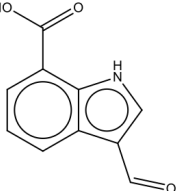
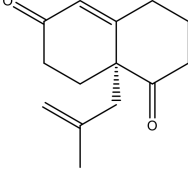
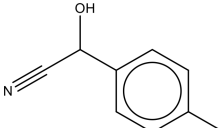
E-A4	A4	 <chem>CC(=CCC/C=C/CN1CCN(CC1)C(=O)C)/C</chem>
E-A5	A5	 <chem>CC1C(C(C(O1)Nc2ccc(O)c2)O)O</chem>
E-A6	A6	 <chem>Cc1cn(c(=O)n(c1=O)C)C2CC(C(O2)CO)O</chem>
E-A7	A7	 <chem>COC(=O)C1C(C(C(O1)n2cnc3c2ncnc3N)O)O</chem>
E-A8	A8	 <chem>c1cnccc1c2[nH]nc(n2)c3ccncc3</chem>
E-A9	A9	 <chem>Cc1cc(=O)n(c2c1c(O)c(O)c2)OC</chem>
E-A10	A10	 <chem>CCn1cc(c(=O)c2c1cccc2)C(=O)N3CCOCC3</chem>

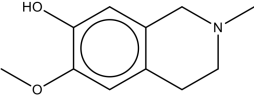
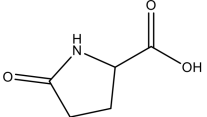
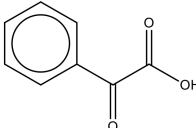
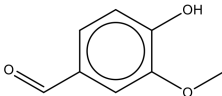
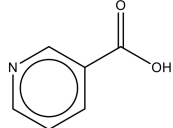
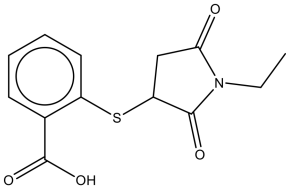
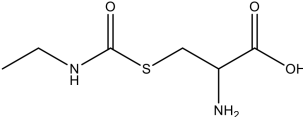
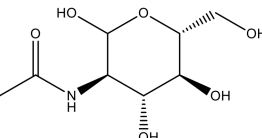
E-A11	A11	
<chem>COC1ccc2c(c1)c(c[nH]2)CCNC(=O)N</chem>		
E-A12	A12	
<chem>CCN1C(=O)CC(C1=O)Sc2ccccc2C(=O)O</chem>		
E-B1	B1	
<chem>Cc1ncc(c(n1)N)COC</chem>		
E-B2	B2	
<chem>Cn1c(=O)c2c(nc1N)n(cn2)C3C(C(C(O3)CO)O)O</chem>		
E-B3	B3	
<chem>c1cc(cc(c1)O)C(=O)O</chem>		
E-B4	B4	
<chem>COC(=O)CNC(=O)COC(=O)c1cccs1</chem>		
E-B5	B5	
<chem>COC(=O)C(=O)CC(=O)c1cccs1</chem>		

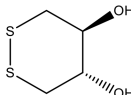
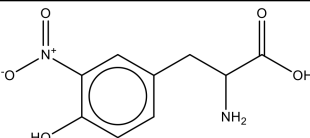
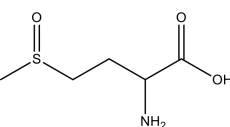
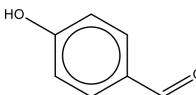
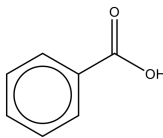
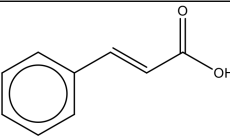
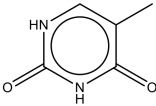
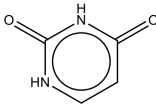
E-B6	B6		<chem>C1CCC(=O)C(=O)C1</chem>
E-B7	B7		<chem>Cc1ccc(c(c1)[N+](=O)[O-])O</chem>
E-B8	B8		<chem>c1cc(c(cc1N)O)C(=O)O</chem>
E-B9	B9		<chem>c1c([nH]c(=O)[nH]c1=O)CC(=O)O</chem>
E-B10	B10		<chem>CCn1cc[nH]c1=S</chem>
E-B11	B11		<chem>CCOCc1cnc(nc1N)C</chem>
E-B12	B12		<chem>Cc1ccnc(c1)N</chem>
E-C1	C1		<chem>C(C(C(F)(F)O)(C(=O)O)N</chem>
E-C2	C2		<chem>c1ccc(cc1)COC(=O)NC(CCC(=O)O)C(=O)O</chem>

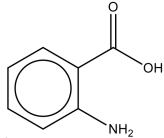
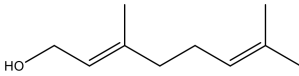
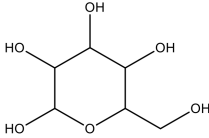
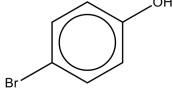
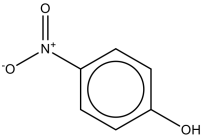
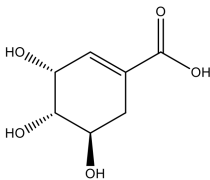
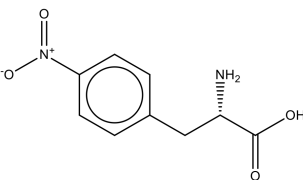
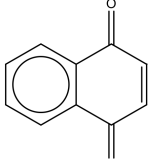
E-C3	C3	
<hr/>		
E-C4	C4	
<hr/>		
E-C5	C5	
<hr/>		
E-C6	C6	
<hr/>		
E-C7	C7	
<hr/>		
E-C8	C8	
<hr/>		
E-C9	C9	
<hr/>		

E-C10	C10	
<hr/>		
<chem>CCNC(=O)Nc1cccc(c1)c2nnnn2C3CC3</chem>		
E-C11	C11	
<hr/>		
<chem>c1c(cc(c1N)O)[N+](=O)[O-][N+](=O)[O-]</chem>		
E-C12	C12	
<hr/>		
<chem>c1c([nH]c(=O)[nH]c1=O)C(=O)O</chem>		
E-D1	D1	
<hr/>		
<chem>c1ccc2c(c1)c([nH]2)CC(=O)O</chem>		
E-D2	D2	
<hr/>		
<chem>CC(=O)c1ccc(O)cc1</chem>		
E-D3	D3	
<hr/>		
<chem>CC(=O)/C=C/c1cccc(O)c1</chem>		
E-D4	D4	
<hr/>		
<chem>c1cnc2c(n1)c(=O)[nH]c(=O)[nH]2</chem>		

E-D5	D5		<chem>c1ccc(cc1)O[C@H]2[C@@H](O)[C@H](O)[C@@H](O)[C@H](O2)CO</chem>
E-D6	D6		<chem>c1nc(c2c(n1)n(cn2)C3C(C(C(C(O3)CO)O)O)O)N</chem>
E-D7	D7		<chem>CC1C2CCC3(C(C2OC1=O)C(CCC3=O)(C)O)C</chem>
E-D8	D8		<chem>c1cc2c(c[nH]c2c(c1)C(=O)O)C=O</chem>
E-D9	D9		<chem>CC(=C)C[C@]12CCC(=O)C=C1CCCC2=O</chem>
E-D10	D10		<chem>Cc1ccc(cc1)C(C#N)O</chem>

E-D11	D11	
<hr/>		
<chem>CN1CCc2cc(c(cc2C1)O)OC</chem>		
E-D12	D12	
<hr/>		
<chem>C1CC(=O)NC1C(=O)O</chem>		
E-E1	E1	
<hr/>		
<chem>c1ccc(cc1)C(=O)C(=O)O</chem>		
E-E2	E2	
<hr/>		
<chem>COC1cc(ccc1O)C=O</chem>		
E-E3	E3	
<hr/>		
<chem>c1cc(cnc1)C(=O)O</chem>		
E-E4	E4	
<hr/>		
<chem>CCN1C(=O)CC(C1=O)Sc2ccccc2C(=O)O</chem>		
E-E5	E5	
<hr/>		
<chem>CCNC(=O)SCC(C(=O)O)N</chem>		
E-E6	E6	
<hr/>		
<chem>CC(=O)N[C@@H]1[C@H]([C@@H]([C@@H]([C@H]1O)CO)O)O</chem>		

E-E7	E7		<chem>C1[C@H]([C@@H](CSS1)O)O</chem>
E-E8	E8		<chem>c1cc(c(cc1CC(C(=O)O)N)[N+](=O)[O-])O</chem>
E-E9	E9		<chem>CS(=O)CCC(C(=O)O)N</chem>
E-E10	E10		<chem>c1cc(ccc1C=O)O</chem>
E-E11	E11		<chem>c1ccc(cc1)C(=O)O</chem>
E-E12	E12		<chem>c1ccc(cc1)/C=C/C(=O)O</chem>
E-F1	F1		<chem>Cc1c[nH]c(=O)[nH]c1=O</chem>
E-F2	F2		<chem>c1c[nH]c(=O)[nH]c1=O</chem>

E-F3	F3	
<hr/>		
		<chem>c1ccc(c(c1)C(=O)O)N</chem>
E-F4	F4	
<hr/>		
		<chem>CC(=CCC/C=C/CO)C</chem>
E-F5	F5	
<hr/>		
		<chem>C(C1C(C(C(C(O1)O)O)O)O)O</chem>
E-F6	F6	
<hr/>		
		<chem>c1cc(ccc1O)Br</chem>
E-F7	F7	
<hr/>		
		<chem>c1cc(ccc1[N+](=O)[O-])O</chem>
E-F8	F8	
<hr/>		
		<chem>C1[C@H]([C@@H]([C@H](C=C1C(=O)O)O)O)O</chem>
E-F9	F9	
<hr/>		
		<chem>c1cc(ccc1C[C@H](C(=O)O)N)[N+](=O)[O-]</chem>
E-F10	F10	
<hr/>		
		<chem>O=C1C=CC(=O)C=C2C=CC(=O)C=C12</chem>

Appendix C

Fragment screening data

Fragment	ΔT_m Rv2991		ΔT_m Rv1155		Fragment	ΔT_m Rv2991		ΔT_m Rv1155	
	- F ₄₂₀	+ F ₄₂₀	- F ₄₂₀	+ F ₄₂₀		- F ₄₂₀	+ F ₄₂₀	- F ₄₂₀	+ F ₄₂₀
E-A1	-1.2	-0.7	0.2	2.0	E-C12	-0.9	-0.1	-0.1	-0.2
E-A2	-2.7	0.1	0.5	0.3	E-D1	-2.4	3.4	1.9	-1.2
E-A3	-1.8	2.2	0.8	-1.5	E-D2	-0.6	0.7	-0.1	0.5
E-A4	-1.5	-0.4	1.1	-1.5	E-D3	N/A	N/A	N/A	N/A
E-A5	N/A	N/A	N/A	N/A	E-D4	0.0	0.7	0.8	2.5
E-A6	-0.9	-0.1	0.2	-0.7	E-D5	0.6	0.7	0.5	0.8
E-A7	-0.6	-0.4	0.8	1.3	E-D6	N/A	N/A	N/A	N/A
E-A8	N/A	N/A	N/A	N/A	E-D7	0.3	0.1	0.5	-0.7
E-A9	N/A	N/A	N/A	N/A	E-D8	-1.8	0.4	0.2	-1.2
E-A10	-0.3	-0.4	-0.4	-2.2	E-D9	-1.8	-0.7	1.1	-0.7
E-A11	-1.2	-1.3	-1.3	-1.9	E-D10	-0.9	-0.4	1.1	-1.0
E-A12	-2.7	0.7	1.4	-1.7	E-D11	N/A	N/A	N/A	N/A
E-B1	0.0	-0.1	-0.4	-1.5	E-D12	-2.7	0.7	1.4	-1.0
E-B2	-0.6	-0.4	1.1	-0.2	E-E1	-1.2	0.4	1.1	0.0
E-B3	-1.8	1.6	1.9	-0.7	E-E2	-2.1	1.3	1.4	-1.7
E-B4	-0.3	0.4	-0.1	-1.0	E-E3	-2.1	9.6	1.4	0.8
E-B5	-4.5	-29.6	0.8	4.0	E-E4	-0.9	1.3	1.4	-0.2
E-B6	1.7	-0.4	1.4	0.0	E-E5	-0.6	1.0	1.4	0.3
E-B7	-0.9	1.0	1.1	-0.7	E-E6	0.3	0.4	1.7	0.5
E-B8	0.3	0.7	-0.1	2.3	E-E7	0.3	0.4	1.4	-0.2
E-B9	-0.9	0.4	0.2	-1.5	E-E8	-0.9	1.0	1.4	1.3
E-B10	-0.6	-0.4	-0.1	-1.0	E-E9	-0.3	0.7	1.1	0.0
E-B11	-0.3	-0.4	-0.7	-1.5	E-E10	-6.3	-0.7	0.8	-1.7
E-B12	1.1	-1.9	-2.2	-1.5	E-E11	-22.0	-26.1	0.8	0.0
E-C1	-0.6	-0.1	0.8	-0.7	E-E12	N/A	-29.1	0.2	-1.7
E-C2	-0.6	1.3	-0.7	-0.5	E-F1	-0.3	-0.1	-0.1	0.5
E-C3	-2.1	1.3	1.4	-0.7	E-F2	0.3	0.1	0.2	0.3
E-C4	-1.8	1.3	2.5	-1.5	E-F3	-2.1	-28.8	2.8	-0.5
E-C5	N/A	N/A	N/A	N/A	E-F4	N/A	N/A	N/A	N/A
E-C6	0.0	0.1	-0.1	-3.7	E-F5	0.6	0.4	0.8	-0.2
E-C7	0.3	0.1	0.5	-0.2	E-F6	-3.6	-1.9	-4.0	-3.4
E-C8	N/A	N/A	N/A	N/A	E-F7	-1.8	1.0	-0.4	-1.2
E-C9	-17.2	-23.2	-12.6	1.5	E-F8	-0.6	0.4	0.8	0.0
E-C10	-0.6	0.1	0.8	-0.2	E-F9	-0.6	0.1	0.8	0.3
E-C11	N/A	N/A	N/A	N/A	E-F10	-2.2	-21.7	0.7	-1.0

Table C.1. Melting temperatures of Rv2991 and Rv1155 in the presence of fragments from the targeted flavoenzyme library.

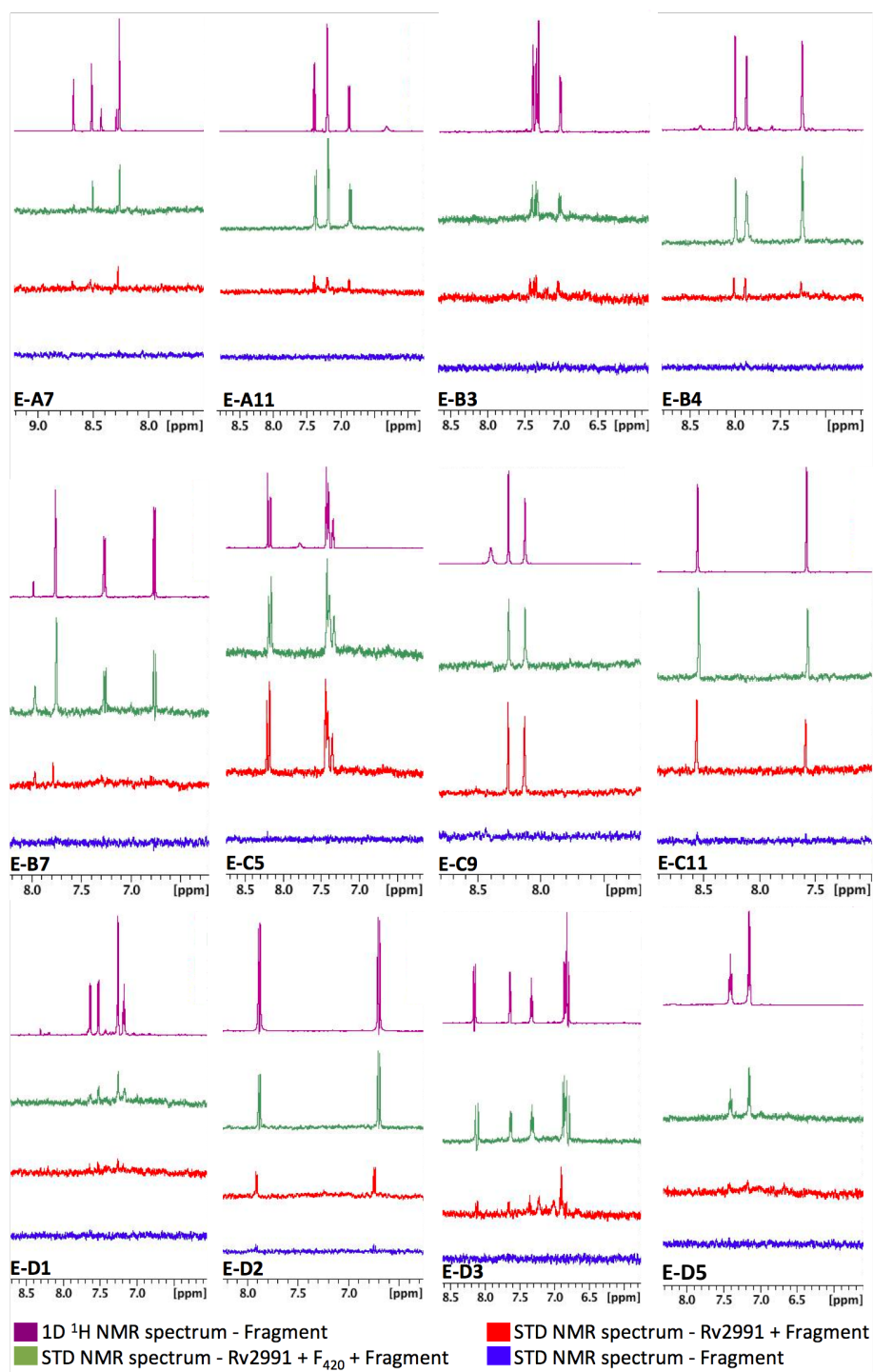


Figure C.1. STD NMR spectra for fragment hits against Rv2991 from the targeted flavoenzyme fragment library.

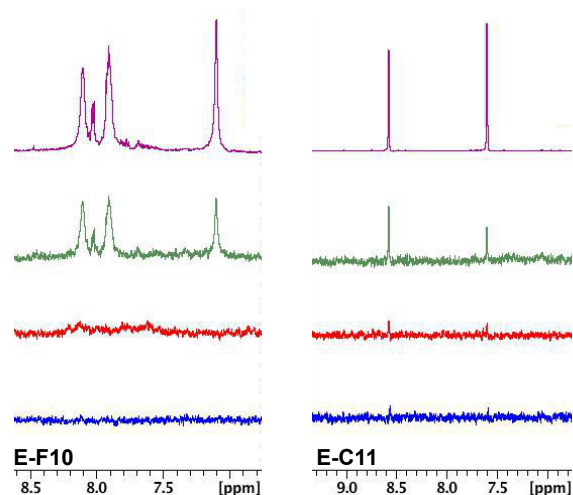


Figure C.2. STD NMR spectra for fragment hits against Rv1155 from the targeted flavoenzyme fragment library. The spectra are colour-coded as in Figure C2.

ID	ΔT_m	ID	ΔT_m
MB022	0.5	AT0544	1.3
MB095	0.3	AT1772	1.5
MB031	1.9	AT0541	0.8
MB024	0.7	CA206	0.6
MB027	0.8	CA208	0.6
MB077	0.5	CA010	0.8
MB143	0.3	MB590	0.8
MB137	0.3	MB600	1.2
MB176	0.3	MB759	0.7
MB131	0.8	MB775	0.7
MB179	0.3	MB770	0.9
MB132	1.2	MB880	0.5
MB199	1.7	MB857	1.0
MB248	1.1	MB875	1.2
MB202	0.2	MB950	0.3
MB210	0.8	MB911	2.0
MB211	0.2	MB920	0.7
MB235	0.4	MB937	0.7
MB237	0.6	MB941	1.6
MB330	0.8	MB992	0.8
MB288	2.1	MB1078	0.3
MB281	0.5	MB1135	1.1
MB320	0.5	MB1120	0.3
MB269	0.5	MB1081	0.3
MB299	0.8	MB1121	0.3
MB366	0.2	MB1145	1.1
MB343	0.7	MB1146	0.5
MB375	1.3	MB1159	0.7
MB383	3.1	MB1201	1.4
MB407	1.3	MB1154	0.7
MB392	0.6	MB1164	2.3
MB401	0.7	MB1238	0.9
MB362	0.4	MB1270	0.9
MB379	0.4	MB1298	0.9
MB432	0.8	MB1276	0.7
CA045	1.9	MB1335	0.5
CA212	0.8	MB1313	0.5

Table C.2. Fragments from the Maybridge library that increase the T_m of Rv1155 by at least twice the standard deviation of the DMSO-only control on each screening plate.

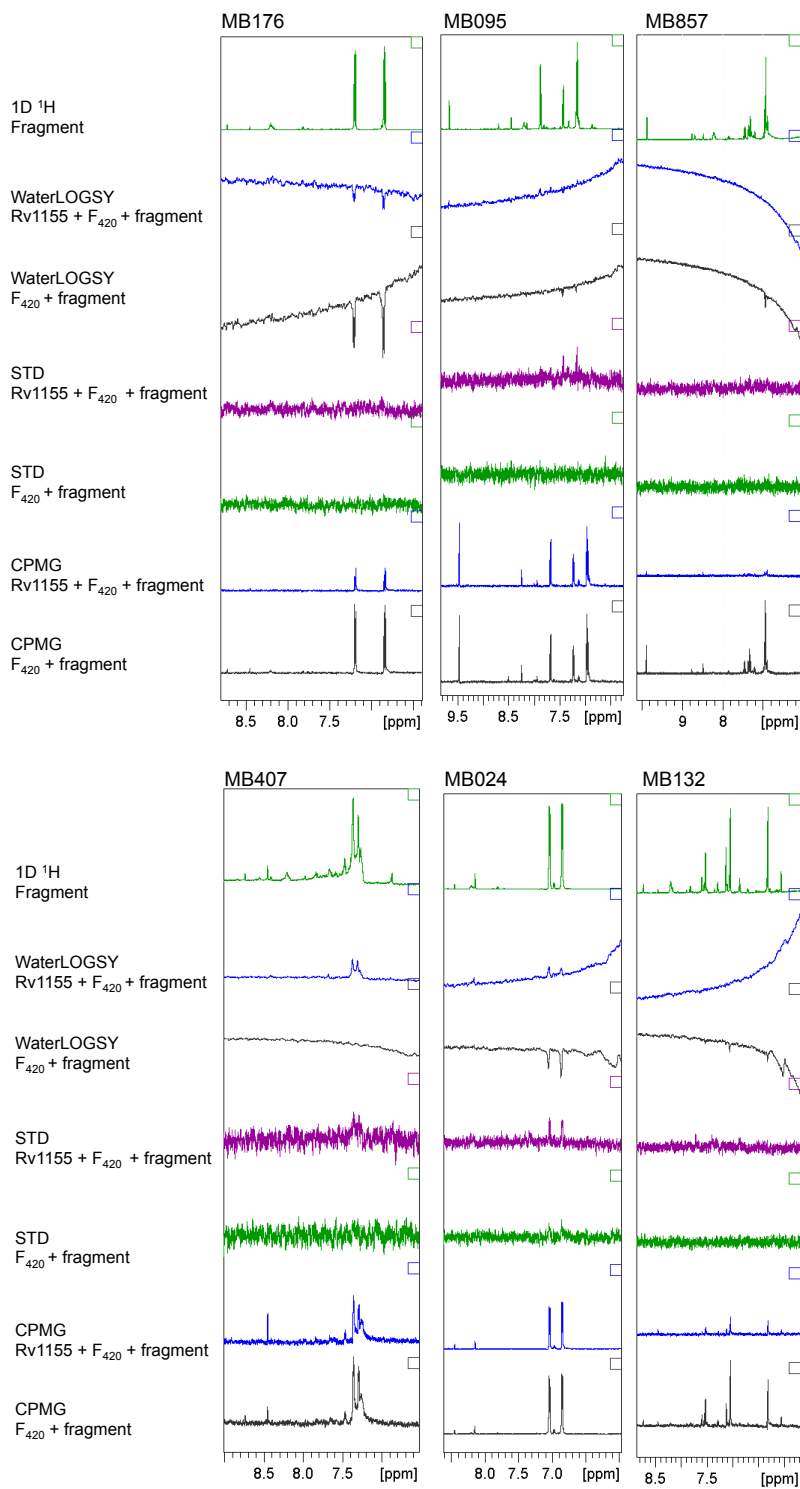


Figure C.3. STD NMR spectra for fragment hits against Rv1155 from the Maybridge library.

References

1. Hershkovitz, I. *et al.* Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS ONE* **3**, e3426 (2008).
2. *WHO Report Global Tuberculosis Control. 2012*
3. Hirsh, A. E., Tsolaki, A. G., DeRiemer, K., Feldman, M. W. & Small, P. M. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4871–4876 (2004).
4. World Health Organization. Stop TB Partnership. The Global Plan to Stop TB 2011-2015. 1–101 (2011).
5. Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. The challenge of new drug discovery for tuberculosis. *Nature* **469**, 483–490 (2013).
6. Mitchison, D. & Davies, G. The chemotherapy of tuberculosis: past, present and future. *Int J. Tuberc. Lung Dis.* **16**, 724–732 (2012).
7. L'homme, R. F. *et al.* Clinical experience with the combined use of lopinavir/ritonavir and rifampicin. *AIDS* **23**, 863–865 (2009).
8. Kapur, A. & Harries, A. D. The double burden of diabetes and tuberculosis - Public health implications. *Diabetes Res. Clin. Pract* 1–10 (2013).
9. Raviglione, M. *et al.* Scaling up interventions to achieve global tuberculosis control: progress and new developments. *Lancet* **379**,

- 1902–1913 (2012).
10. Centers for Disease Control and Prevention National Center for HIV/AIDS Viral Hepatitis STD and Tuberculosis Prevention. *TB Elimination. Multidrug-Resistant Tuberculosis (MDR TB)*. 1–2 (2012).
 11. Centers for Disease Control and Prevention National Center for HIV/AIDS Viral Hepatitis STD and Tuberculosis Prevention. *TB Elimination. Extensively Drug-Resistant Tuberculosis (XDR TB)*. 1–3 (2012).
 12. Udwadia, Z. F., Amale, R. A., Ajbani, K. K. & Rodrigues, C. Totally drug-resistant tuberculosis in India. *Clin Infect Dis* **54**, 579–581 (2012).
 13. Hoffner, S. Unexpected high levels of multidrug-resistant tuberculosis present new challenges for tuberculosis control. *Lancet* **380**, 1367–1369 (2012).
 14. Duncan, K. & Barry, C. E., III. Prospects for new antitubercular drugs. *Curr. Opin. Microbiol.* **7**, 460–465 (2004).
 15. Russell, D. G., Barry, C. E. 3. & Flynn, J. L. Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**, 852–856 (2010).
 16. Pelaez, F. The historical delivery of antibiotics from microbial natural products--can history repeat? *Biochem. Pharmacol.* **71**, 981–990 (2006).
 17. 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.* **6**, 29–40 (2006).
 18. Cole, S. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
 19. Camus, J.-C., Pryor, M. J., Médigue, C. & Cole, S. T. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiol.* **148**, 2967–2973 (2002).
 20. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12712–12717 (2001).
 21. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiol.* **48**, 77–84 (2003).
 22. Sala, C. & Hartkoorn, R. C. Tuberculosis drugs: new candidates and how to find more. *Future Microbiol.* **6**, 617–633 (2011).
 23. Makarov, V. *et al.* Synthesis and antileprosy activity of some dialkylthiocarbamates. *J. Antimicrob. Chemother.* **57**, 1134–1138 (2006).
 24. Makarov, V. *et al.* Benzothiazinones kill *Mycobacterium tuberculosis* by blocking arabinan synthesis. *Science* **324**, 801–804 (2009).
 25. Manjunatha, U. H. *et al.* Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 431–436 (2006).
 26. Sacksteder, K. A., Protopopova, M., Barry, C. E. 3., Andries, K. & Nacy, C. A. Discovery and development of SQ109: a new antitubercular drug

- with a novel mechanism of action. *Future Microbiol* **7**, 823–837 (2012).
27. Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**, 223–227 (2005).
 28. Cohen, J. Infectious disease. Approval of novel TB drug celebrated--with restraint. *Science* **339**, 130 (2013).
 29. Caviedes, L., Delgado, J. & Gilman, R. H. Tetrazolium microplate assay as a rapid and inexpensive colorimetric method for determination of antibiotic susceptibility of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **40**, 1873–1874 (2002).
 30. Gengenbacher, M., Rao, S. P. S., Pethe, K. & Dick, T. Nutrient-starved, non-replicating *Mycobacterium tuberculosis* requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability. *Microbiol.* **156**, 81–87 (2010).
 31. Rao, S. P. S., Alonso, S., Rand, L., Dick, T. & Pethe, K. The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11945–11950 (2008).
 32. Pethe, K. *et al.* A chemical genetic screen in *Mycobacterium tuberculosis* identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy. *Nat. Commun.* **1**, 57 (2010).
 33. Collins, L. A., Torrero, M. N. & Franzblau, S. G. Green fluorescent protein reporter microplate assay for high-throughput screening of compounds against *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **42**, 344–347 (1998).
 34. Terstappen, G. C., Schlupen, C., Raggiaschi, R. & Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov.* **6**, 891–903 (2007).
 35. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
 36. 3rd, C. E. B. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat. Rev. Micro.* **7**, 845–855 (2009).
 37. Sacchettini, J. C., Rubin, E. J. & Freundlich, J. S. Drugs versus bugs: in pursuit of the persistent predator *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **6**, 41–52 (2008).
 38. Serhan, C. N. Signalling the fat controller. *Nature* **384**, 23–24 (1996).
 39. Goodwill, K. E., Tennant, M. G. & Stevens, R. C. High-throughput x-ray crystallography for structure-based drug design. *Drug Discov. Today* **6**, S113–118 (2001).
 40. Verlinde, C. L. & Hol, W. G. Structure-based drug design: progress, results and challenges. *Structure* **2**, 577–587 (1994).
 41. Chim, N. *et al.* The TB Structural Genomics Consortium: a decade of progress. *Tuberculosis* **91**, 155–172 (2011).
 42. He, X., Alian, A. & Ortiz de Montellano, P. R. Inhibition of the *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase InhA by

- aryl amides. *Bioorg. Med. Chem.* **15**, 6649–6658 (2007).
43. Magnet, S. *et al.* Leads for antitubercular compounds from kinase inhibitor library screens. *Tuberculosis* **90**, 354–360 (2010).
 44. Hegymegi-Barakonyi, B. *et al.* Signalling inhibitors against *Mycobacterium tuberculosis*--early days of a new therapeutic concept in tuberculosis. *Curr. Med. Chem.* **15**, 2760–2770 (2008).
 45. Szekely, R. *et al.* A novel drug discovery concept for tuberculosis: inhibition of bacterial and host cell signalling. *Immunol. Lett.* **116**, 225–231 (2008).
 46. Li, X. *et al.* Synthesis and SAR studies of 1,4-benzoxazine MenB inhibitors: novel antibacterial agents against *Mycobacterium tuberculosis*. *Bioorg. Med. Chem. Lett.* **20**, 6306–6309 (2010).
 47. White, E. L. *et al.* A novel inhibitor of *Mycobacterium tuberculosis* pantothenate synthetase. *J. Biomol. Screen.* **12**, 100–105 (2007).
 48. Chessari, G. & Woodhead, A. J. From fragment to clinical candidate--a historical perspective. *Drug Discov. Today* **14**, 668–675 (2009).
 49. Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **51**, 4990–5003 (2012).
 50. FDA approves vemurafenib for treatment of metastatic melanoma. *Oncology* **25**, 906 (2011).
 51. Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* **12**, 5–7 (2012).
 52. Ciulli, A. & Abell, C. Fragment-based approaches to enzyme inhibition. *Curr. Opin. Biotechnol.* **18**, 489–496 (2007).
 53. Hung, A. *et al.* Application of fragment growing and fragment linking to the discovery of inhibitors of *Mycobacterium tuberculosis* pantothenate synthetase. *Angew. Chem. Int. Ed. Engl.* **48**, 8452–8456 (2009).
 54. Hudson, S. A. *et al.* Application of fragment screening and merging to the discovery of inhibitors of the *Mycobacterium tuberculosis* cytochrome P450 CYP121. *Angew. Chem. Int. Ed. Engl.* **51**, 9311–9316 (2012).
 55. Hajduk, P. J. *et al.* Design of adenosine kinase inhibitors from the NMR-based screening of fragments. *J. Med. Chem.* **43**, 4781–4786 (2000).
 56. Gill, A., Cleasby, A. & Jhoti, H. The discovery of novel protein kinase inhibitors by using fragment-based high-throughput x-ray crystallography. *ChemBioChem* **6**, 506–512 (2005).
 57. Liu, G. *et al.* Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. *J. Med. Chem.* **46**, 4232–4235 (2003).
 58. Chen, L., Cressina, E., Leeper, F. J., Smith, A. G. & Abell, C. A fragment-based approach to identifying ligands for riboswitches. *A.C.S. Chem. Biol.* **5**, 355–358 (2010).
 59. Petros, A. M. *et al.* Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *J. Med. Chem.* **49**, 656–663 (2006).

60. Scott, D. E. *et al.* Using a fragment-based approach to target protein-protein interactions. *ChemBioChem*. **14**, 332-342 (2013).
61. Mochalkin, I. *et al.* Discovery of antibacterial biotin carboxylase inhibitors by virtual screening and fragment-based approaches. *A.C.S. Chem. Biol.* **4**, 473–483 (2009).
62. Mpamhanga, C. P. *et al.* One scaffold, three binding modes: novel and selective pteridine reductase 1 inhibitors derived from fragment hits discovered by virtual screening. *J. Med. Chem.* **52**, 4454–4465 (2009).
63. Pieters, J. Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. *Cell Host Microbe* **3**, 399–407 (2008).
64. Rawls, K. A. *et al.* Fragment-based discovery of selective inhibitors of the Mycobacterium tuberculosis protein tyrosine phosphatase PtpA. *Bioorganic & Medicinal Chemistry Letters* **19**, 6851–6854 (2009).
65. Soellner, M. B., Rawls, K. A., Grundner, C., Alber, T. & Ellman, J. A. Fragment-based substrate activity screening method for the identification of potent inhibitors of the Mycobacterium tuberculosis phosphatase PtpB. *J. Am. Chem. Soc.* **129**, 9613–9615 (2007).
66. Khazak, V., Astsaturov, I., Serebriiskii, I. G. & Golemis, E. A. Selective Raf inhibition in cancer therapy. *Expert Opin. Ther. Targets* **11**, 1587–1609 (2007).
67. Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3041–3046 (2008).
68. Sledz, P., Abell, C. & Ciulli, A. Ligand-observed NMR in fragment-based approaches. "NMR of biomolecules: towards mechanistic systems biology", First Ed. Edited by Bertini, I. McGreevy, K.S., Parigi, G. Wiley-VCH Verlag GmbH & Co. KGaA. 265-281 (2012).
69. Kranz, J. K. & Schalk-Hihi, C. Protein thermal shifts to identify low molecular weight fragments. *Methods Enzymol.* **493**, 277–298 (2011).
70. Hartshorn, M. J. *et al.* Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **48**, 403–413 (2005).
71. Blundell, T. L., Jhoti, H. & Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* **1**, 45–54 (2002).
72. Navratilova, I. & Hopkins, A. L. Fragment Screening by Surface Plasmon Resonance. *A.C.S. Med. Chem. Lett.* **1**, 44–48 (2010).
73. Hofstadler, S. A. & Sannes-Lowery, K. A. Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat. Rev. Drug Discov.* **5**, 585–595 (2006).
74. Poklar, N., Lah, J., Salobir, M., Macek, P. & Vesnaver, G. pH and temperature-induced molten globule-like denatured states of equinatoxin II: a study by UV-melting, DSC, far- and near-UV CD spectroscopy, and ANS fluorescence. *Biochemistry* **36**, 14345–14352 (1997).
75. Pantoliano, M. W. *et al.* High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **6**, 429–440 (2001).

76. Lo, M.-C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **332**, 153–159 (2004).
77. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).
78. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534 (1996).
79. *Practical Fragments*. <http://practicalfragments.blogspot.com/>
80. Lepre, C. A., Moore, J. M. & Peng, J. W. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.* **104**, 3641–3676 (2004).
81. Angulo, J. & Nieto, P. M. STD-NMR: application to transient interactions between biomolecules—a quantitative approach. *Eur. Biophys. J.* **40**, 1357–1369 (2011).
82. 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* **21**, 349–359 (2001).
83. Carr, Y. H. & Purcell, M. E. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **94**, 630–638 (1954).
84. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
85. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **38**, 1784–1788 (1999).
86. Dalvit, C. *et al.* Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **18**, 65–68 (2000).
87. Hajduk, P. J., Olejniczak, E. T. & Fesik, S. W. One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J. Am. Chem. Soc.* **119**, 12257–12261 (1997).
88. Hung, A. *et al.* Application of fragment growing and fragment linking to the discovery of inhibitors of Mycobacterium tuberculosis pantothenate synthetase. *Angew. Chem. Int. Ed.* **48**, 8452–8456 (2009).
89. Sledz, P. *et al.* Optimization of the interligand Overhauser effect for fragment linking: Application to inhibitor discovery against Mycobacterium tuberculosis pantothenate synthetase. *J. Am. Chem. Soc.* **132**, 4544–4545 (2010).
90. Abrahams, G. L. *et al.* Pathway-selective sensitization of Mycobacterium tuberculosis for target-based whole-cell screening. *Chem. Biol.* **19**, 844–854 (2012).
91. Li, X.-Z. & Nikaido, H. Efflux-mediated drug resistance in bacteria: an update. *Drugs* **69**, 1555–1623 (2009).

92. Pichota, A. *et al.* Peptide deformylase inhibitors of *Mycobacterium tuberculosis*: synthesis, structural investigations, and biological results. *Bioorg. Med. Chem. Lett.* **18**, 6568–6572 (2008).
93. Boshoff, H. I. M. *et al.* Biosynthesis and recycling of nicotinamide cofactors in *mycobacterium tuberculosis*. An essential role for NAD in nonreplicating bacilli. *J. Biol. Chem.* **283**, 19329–19341 (2008).
94. Gerdes, S. Y. *et al.* From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J. Bacteriol.* **184**, 4555–4572 (2002).
95. Mikusova, K. *et al.* Decaprenylphosphoryl arabinofuranose, the donor of the D-arabinofuranosyl residues of mycobacterial arabinan, is formed via a two-step epimerization of decaprenylphosphoryl ribose. *J. Bacteriol.* **187**, 8020–8025 (2005).
96. Batt, S. M. *et al.* Structural basis of inhibition of *Mycobacterium tuberculosis* DprE1 by benzothiazinone inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11354–11359 (2012).
97. Neres, J. *et al.* Structural basis for benzothiazinone-mediated killing of *Mycobacterium tuberculosis*. *Sci. Transl. Med.* **4**, 150ra121 (2012).
98. Lew, J. M., Kapopoulou, A., Jones, L. M. & Cole, S. T. TubercuList--10 years after. *Tuberculosis* **91**, 1–7 (2011).
99. Gellman, S. H. Introduction: Molecular Recognition. *Chem. Rev.* **97**, 1231–1232 (1997).
100. Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **78**, 269–281 (2000).
101. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187–192 (2009).
102. Stout, T. J., Sage, C. R. & Stroud, R. M. The additivity of substrate fragments in enzyme–ligand binding. *Structure* **6**, 839–848 (1998).
103. 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.* **49**, 4992–5000 (2006).
104. Erlanson, D. A. *et al.* Discovery of a new phosphotyrosine mimetic for PTP1B using breakaway tethering. *J. Am. Chem. Soc.* **125**, 5602–5603 (2003).
105. Bateman, A. The Pfam protein families database. *Nucleic Acids Res.* **32**, 138D–141 (2004).
106. Dick, T., Manjunatha, U., Kappes, B. & Gengenbacher, M. Vitamin B6 biosynthesis is essential for survival and virulence of *Mycobacterium tuberculosis*. *Mol. Microbiol.* **78**, 980–988 (2010).
107. Di Salvo, M. L., Contestabile, R. & Safo, M. K. Vitamin B(6) salvage enzymes: mechanism, structure and regulation. *Biochim. Biophys. Acta.* **1814**, 1597–1608 (2011).
108. Di Salvo, M. L., Safo, M. K., Musayev, F. N., Bossa, F. & Schirch, V. Structure and mechanism of *Escherichia coli* pyridoxine 5'-phosphate oxidase. *Biochim. Biophys. Acta.* **1647**, 76–82 (2003).

109. Eliot, A. C. & Kirsch, J. F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **73**, 383–415 (2004).
110. Fitzpatrick, T. B. *et al.* Two independent routes of de novo vitamin B6 biosynthesis: not that different after all. *Biochem. J.* **407**, 1 (2007).
111. Webb, M. E., Marquet, A., Mendel, R. R., Rebeille, F. & Smith, A. G. Elucidating biosynthetic pathways for vitamins and cofactors. *Nat. Prod. Rep.* **24**, 988–1008 (2007).
112. Ehrenshaft, M., Bilski, P., Li, M. Y., Chignell, C. F. & Daub, M. E. A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9374–9378 (1999).
113. Cole, S. T. *et al.* Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011 (2001).
114. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
115. Ohno, S. *Evolution by Gene Duplication*. Springer-Verlag, (1970).
116. Moreno-Hagelsieb, G., Trevino, V., Perez-Rueda, E., Smith, T. F. & Collado-Vides, J. Transcription unit conservation in the three domains of life: a perspective from *Escherichia coli*. *Trends Genet.* **17**, 175–177 (2001).
117. Pédelacq, J.-D. *et al.* Crystal structure of a putative pyridoxine 5'-phosphate oxidase (Rv2607) from *Mycobacterium tuberculosis*. *Proteins* **62**, 563–569 (2005).
118. Cnaan, S. *et al.* Crystal structure of the conserved hypothetical protein Rv1155 from *Mycobacterium tuberculosis*. *FEBS Lett.* **579**, 215–221 (2005).
119. Biswal, B. K., Au, K., Cherney, M. M., Garen, C. & James, M. N. G. The molecular structure of Rv2074, a probable pyridoxine 5'-phosphate oxidase from *Mycobacterium tuberculosis*, at 1.6 Å resolution. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **62**, 735–742 (2006).
120. Safo, M. K., Musayev, F. N. & Schirch, V. Structure of *Escherichia coli* pyridoxine 5'-phosphate oxidase in a tetragonal crystal form: insights into the mechanistic pathway of the enzyme. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 599–604 (2005).
121. Musayev, F.N., Di Salvo, M.L., Ko, T.P., Schirch, V. & Safo, M.K. Structure and properties of recombinant human pyridoxine 5'-phosphate oxidase. *Protein Sci.* **12**, 1455-1463 (2003).
122. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
123. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–9 (2010).
124. Taly, J.-F. *et al.* Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat Protoc.* **6**, 1669–1682 (2011).

125. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 386–394 (2011).
126. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–8 (2006).
127. Biswal, B. K., Cherney, M. M., Wang, M., Garen, C. & James, M. N. G. Structures of Mycobacterium tuberculosis pyridoxine 5'-phosphate oxidase and its complexes with flavin mononucleotide and pyridoxal 5'-phosphate. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 1492–1499 (2005).
128. *Nomenclature for vitamins B-6 and related compounds*. International Union of Pure and Applied Chemistry and International Union of Biochemistry. <http://www.chem.qmul.ac.uk/iupac/misc/B6.html>
129. Strohmeier, M. *et al.* Structure of a bacterial pyridoxal 5'-phosphate synthase complex. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19284–19289 (2006).
130. Zein, F. *et al.* Structural insights into the mechanism of the PLP synthase holoenzyme from *Thermotoga maritima*. *Biochemistry* **45**, 14609–14620 (2006).
131. Gengenbacher, M. *et al.* Vitamin B6 biosynthesis by the malaria parasite *Plasmodium falciparum*: biochemical and structural insights. *J. Biol. Chem.* **281**, 3633–3641 (2006).
132. Wrenger, C., Eschbach, M.-L., Muller, I. B., Warnecke, D. & Walter, R. D. Analysis of the vitamin B6 biosynthesis pathway in the human malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **280**, 5242–5248 (2005).
133. Boshoff, H. I. M. *et al.* The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action. *J. Biol. Chem.* **279**, 40174–40184 (2004).
134. Schnappinger, D. *et al.* Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *J. Exp. Med.* **198**, 693–704 (2003).
135. Tanaka, T., Tateno, Y. & Gojobori, T. Evolution of vitamin B6 (pyridoxine) metabolism by gain and loss of genes. *Mol. Biol. Evol.* **22**, 243–250 (2005).
136. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
137. Churchich, J. E. Brain pyridoxine-5-phosphate oxidase. A dimeric enzyme containing one FMN site. *Eur. J. Biochem.* **138**, 327–332 (1984).
138. Choi, S. Y., Churchich, J. E., Zaiden, E. & Kwok, F. Brain pyridoxine-5-phosphate oxidase. Modulation of its catalytic activity by reaction with pyridoxal 5-phosphate and analogs. *J. Biol. Chem.* **262**, 12013–12017 (1987).
139. Kazarinoff, M. N. & McCormick, D. B. Rabbit liver pyridoxamine (pyridoxine) 5'-phosphate oxidase. Purification and properties. *J. Biol.*

- Chem.* **250**, 3436–3442 (1975).
140. Benesch, J. L. P. Collisional activation of protein complexes: picking up the pieces. *J. Am. Soc. Mass Spectrom.* **20**, 341–348 (2009).
 141. Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682 (1999).
 142. Wada, H. & Snell, E. E. The enzymatic oxidation of pyridoxine and pyridoxamine phosphates. *J. Biol. Chem.* **236**, 2089–2095 (1961).
 143. Di Salvo, M., Yang, E., Zhao, G., Winkler, M. E. & Schirch, V. Expression, purification, and characterization of recombinant *Escherichia coli* pyridoxine 5'-phosphate oxidase. *Protein Expr. Purif.* **13**, 349–356 (1998).
 144. Choi, J. D., Bowers-Komro, M., Davis, M. D., Edmondson, D. E. & McCormick, D. B. Kinetic properties of pyridoxamine (pyridoxine)-5'-phosphate oxidase from rabbit liver. *J. Biol. Chem.* **258**, 840–845 (1983).
 145. Vilcheze, C. *et al.* Transfer of a point mutation in *Mycobacterium tuberculosis inhA* resolves the target of isoniazid. *Nat. Med.* **12**, 1027–1029 (2006).
 146. Belanger, A. E. *et al.* The embAB genes of *Mycobacterium avium* encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11919–11924 (1996).
 147. Singh, R. *et al.* PA-824 kills nonreplicating *Mycobacterium tuberculosis* by intracellular NO release. *Science* **322**, 1392–1395 (2008).
 148. Darwin, K. H., Ehrt, S., Gutierrez-Ramos, J.-C., Weich, N. & Nathan, C. F. The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. *Science* **302**, 1963–1966 (2003).
 149. MacMicking, J., Xie, Q. W. & Nathan, C. Nitric oxide and macrophage function. *Annu. Rev. Immunol.* **15**, 323–350 (1997).
 150. Nathan, C. Microbiology. An antibiotic mimics immunity. *Science* **322**, 1337–1338 (2008).
 151. COUSINS, F. B. The prosthetic group of a chromoprotein from mycobacteria. *Biochim. Biophys. Acta* **40**, 532–534 (1960).
 152. Sutton, W. B. Properties of a new TPN-like electron transport component from *Mycobacterium phlei*. *Biochem. Biophys. Res. Commun.* **15**, 414–419 (1964).
 153. Cheeseman, P., Toms-Wood, A. & Wolfe, R. S. Isolation and properties of a fluorescent compound, factor 420, from *Methanobacterium* strain M.o.H. *J. Bacteriol.* **112**, 527–531 (1972).
 154. Walsh, C. Naturally occurring 5-deazaflavin coenzymes: biological redox roles. *Acc. Chem. Res.* **19**, 216–221 (2012).
 155. Bair, T., Isabelle, D. & Daniels, L. Structures of coenzyme F(420) in *Mycobacterium* species. *Arch. Microbiol.* **176**, 37–43 (2001).
 156. Massey, V. The chemical and biological versatility of riboflavin. *Biochem. Soc. Trans.* **28**, 283–296 (2000).
 157. Boshoff, H. I. M. & Barry, C. E. 3. Tuberculosis - metabolism and

- respiration in the absence of growth. *Nat. Rev. Micro.* **3**, 70–80 (2005).
158. Hasan, M. R., Rahman, M., Jaques, S., Purwantini, E. & Daniels, L. Glucose 6-phosphate accumulation in mycobacteria: implications for a novel F420-dependent anti-oxidant defense system. *J. Biol. Chem.* **285**, 19135–19144 (2010).
159. Guerra-Lopez, D., Daniels, L. & Rawat, M. Mycobacterium smegmatis mc2 155 fbiC and MSMEG_2392 are involved in triphenylmethane dye decolorization and coenzyme F420 biosynthesis. *Microbiol.* **153**, 2724–2732 (2007).
160. Purwantini, E. & Mukhopadhyay, B. Conversion of NO₂ to NO by reduced coenzyme F420 protects mycobacteria from nitrosative damage. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6333–6338 (2009).
161. Anderson, R. F. Energetics of the one-electron reduction steps of riboflavin, FMN and FAD to their fully reduced forms. *Biochim. Biophys. Acta* **722**, 158–162 (1983).
162. Hagemeyer, C. H. *et al.* Coenzyme F420-dependent methylenetetrahydromethanopterin dehydrogenase (Mtd) from Methanopyrus kandleri: a methanogenic enzyme with an unusual quarternary structure. *J. Mol. Biol.* **332**, 1047–1057 (2003).
163. Shima, S. *et al.* Structure of coenzyme F(420) dependent methylenetetrahydromethanopterin reductase from two methanogenic archaea. *J. Mol. Biol.* **300**, 935–950 (2000).
164. Nakano, T. *et al.* Identification and cloning of the gene involved in the final step of chlortetracycline biosynthesis in Streptomyces aureofaciens. *Biosci. Biotechnol. Biochem.* **68**, 1345–1352 (2004).
165. Li, W., Chou, S., Khullar, A. & Gerratana, B. Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an SJG-136 monomeric analog. *Applied Environ. Microbiol.* **75**, 2958–2963 (2009).
166. Heiss, G. *et al.* npd gene functions of Rhodococcus (opacus) erythropolis HL PM-1 in the initial steps of 2,4,6-trinitrophenol degradation. *Microbiol.* **148**, 799–806 (2002).
167. Graham, D. E. A new role for coenzyme F420 in aflatoxin reduction by soil mycobacteria. *Mol. Microbiol.* **78**, 533–536 (2010).
168. Purwantini, E. & Daniels, L. Molecular analysis of the gene encoding F420-dependent glucose-6-phosphate dehydrogenase from Mycobacterium smegmatis. *J. Bacteriol.* **180**, 2212–2219 (1998).
169. Bashiri, G., Squire, C. J., Moreland, N. J. & Baker, E. N. Crystal structures of F420-dependent glucose-6-phosphate dehydrogenase FGD1 involved in the activation of the anti-tuberculosis drug candidate PA-824 reveal the basis of coenzyme and substrate binding. *J. Biol. Chem.* **283**, 17531–17541 (2008).
170. Purwantini, E. & Daniels, L. Purification of a novel coenzyme F420-dependent glucose-6-phosphate dehydrogenase from Mycobacterium smegmatis. *J. Bacteriol.* **178**, 2861–2866 (1996).
171. Selengut, J. D. & Haft, D. H. Unexpected abundance of coenzyme

- F(420)-dependent enzymes in *Mycobacterium tuberculosis* and other actinobacteria. *J. Bacteriol.* **192**, 5788–5798 (2010).
172. Isabelle, D., Simpson, D. R. & Daniels, L. Large-scale production of coenzyme F420-5,6 by using *Mycobacterium smegmatis*. *Applied and Environ. Microbiol.* **68**, 5750–5755 (2002).
173. Bashiri, G., Rehan, A. M., Greenwood, D. R., Dickson, J. M. J. & Baker, E. N. Metabolic engineering of cofactor F420 production in *Mycobacterium smegmatis*. *PLoS ONE* **5**, e15803 (2010).
174. Cellitti, S. E. *et al.* Structure of Ddn, the deazaflavin-dependent nitroreductase from *Mycobacterium tuberculosis* involved in bioreductive activation of PA-824. *Structure* **20**, 101–112 (2012).
175. Taylor, M. C. *et al.* Identification and characterization of two families of F420 H₂-dependent reductases from *Mycobacteria* that catalyse aflatoxin degradation. *Mol. Microbiol.* **78**, 561–575 (2010).
176. Lee, D., Redfern, O. & Orengo, C. Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* **8**, 995–1005 (2007).
177. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* **5**, e1000605 (2009).
178. Furnham, N., Garavelli, J. S., Apweiler, R. & Thornton, J. M. Missing in action: enzyme functional annotations in biological databases. *Nat. Chem. Biol.* **5**, 521–525 (2009).
179. Galperin, M. Y. & Koonin, E. V. ‘Conserved hypothetical’ proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* **32**, 5452–5463 (2004).
180. Bork, P. Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res.* **10**, 398–400 (2000).
181. Griffiths, A.J.F. *et al.* An Introduction to Genetic Analysis. Seventh Ed. W.H. Freeman (2000). Available from <http://ncbi.nlm.nih.gov/books/NBK21766>.
182. Schulze, A. & Downward, J. Navigating gene expression using microarrays—a technology review. *Nat. Cell Biol.* **3**, E190–5 (2001).
183. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).
184. Carver, T. E. *et al.* Decrypting the biochemical function of an essential gene from *Streptococcus pneumoniae* using ThermoFluor technology. *J. Biol. Chem.* **280**, 11704–11712 (2005).
185. Williams, C. Biotechnology match making: screening orphan ligands and receptors. *Curr. Opin. Biotechnol.* **11**, 42–46 (2000).
186. Yang, M., Brazier, M., Edwards, R. & Davis, B. G. High-throughput mass-spectrometry monitoring for multisubstrate enzymes: determining the kinetic parameters and catalytic activities of glycosyltransferases. *ChemBioChem* **6**, 346–357 (2005).
187. Macchiarulo, A., Nobeli, I. & Thornton, J. M. Ligand selectivity and competition between enzymes in silico. *Nat. Biotechnol.* **22**, 1039–1045

- (2004).
188. Zarembinski, T. I. *et al.* Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15189–15193 (1998).
 189. Hermann, J. C. *et al.* Structure-based activity prediction for an enzyme of unknown function. *Nature* **448**, 775–779 (2007).
 190. Sadowski, M. I. & Jones, D. T. The sequence-structure relationship and protein function prediction. *Curr. Opin. Structural Biol.* **19**, 357–362 (2009).
 191. Hopkins, A. L., Groom, C. R. & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9**, 430–431 (2004).
 192. Nakamura, C. E. & Abeles, R. H. Mode of interaction of beta-hydroxy-beta-methylglutaryl coenzyme A reductase with strong binding inhibitors: compactin and related compounds. *Biochemistry* **24**, 1364–1376 (1985).
 193. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4046–4050 (1981).
 194. Larsson, J., Gottfries, J., Muresan, S. & Backlund, A. ChemGPS-NP: tuned for navigation in biologically relevant chemical space. *J. Nat. Prod.* **70**, 789–794 (2007).
 195. Davies, D. R. *et al.* Discovery of leukotriene A4 hydrolase inhibitors using metabolomics biased fragment crystallography. *J. Med. Chem.* **52**, 4694–4715 (2009).
 196. Hung, A. W. *et al.* Route to three-dimensional fragments using diversity-oriented synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6799–6804 (2011).
 197. Over, B. *et al.* Natural-product-derived fragments for fragment-based ligand discovery. *Nat. Chem.* **5**, 21–28 (2013).
 198. Orry, A. J. W., Abagyan, R. A. & Cavasotto, C. N. Structure-based development of target-specific compound libraries. *Drug Discov. Today* **11**, 261–266 (2006).
 199. Akritopoulou-Zanze, I. & Hajduk, P. J. Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors. *Drug Discov. Today* **14**, 291–297 (2009).
 200. Crisman, T. J. *et al.* ‘Virtual fragment linking’: an approach to identify potent binders from low affinity fragment hits. *J. Med. Chem.* **51**, 2481–2491 (2008).
 201. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model* (2012). doi:10.1021/ci3001277
 202. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
 203. *Pipeline Pilot Student Edition v6.1.*
 204. Hert, J. *et al.* Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* **2**, 3256–3266 (2004).
 205. Crisman, T. J., Sisay, M. T. & Bajorath, J. Ligand-target interaction-

- based weighting of substructures for virtual screening. *J. Chem. Inf. Model.* **48**, 1955–1964 (2008).
206. Basse, N. *et al.* Toward the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant. *Chem. Biol.* **17**, 46–56 (2010).
207. Bender, A., Mussa, H. Y., Glen, R. C. & Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J. Chem. Inf. Comput. Sci.* **44**, 170–178 (2004).
208. Hattori, M., Tanaka, N., Kanehisa, M. & Goto, S. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* **38**, W652–6 (2010).
209. Eirich, L. D., Vogels, G. D. & Wolfe, R. S. Distribution of coenzyme F420 and properties of its hydrolytic fragments. *J. Bacteriol.* **140**, 20–27 (1979).
210. PistonKnobel. Real-time Analysis of Glucose Metabolism by Microscopy. *Trends Endocrinol. Metab.* **10**, 413–417 (1999).
211. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 (2002).
212. Nettleton, E. J. *et al.* Protein subunit interactions and structural integrity of amyloidogenic transthyretins: evidence from electrospray mass spectrometry. *J. Mol. Biol.* **281**, 553–564 (1998).
213. Argoudelis, C. Preparation of Crystalline Pyridoxine 5'-Phosphate and Some of Its Properties. *J. Agr. Food Chem.* **34**, 995–998 (1986).
214. Kwok, F. & Churchich, J. E. Interaction between pyridoxal kinase and pyridoxine-5-P oxidase, two enzymes involved in the metabolism of vitamin B6. *J. Biol. Chem.* **255**, 882–887 (1980).
215. Hwang, T. & Shaka, A. Water suppression that works - excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson. Ser. A.* **112**, 275–279 (1995).
216. Choi, K.-P., Kendrick, N. & Daniels, L. Demonstration that fbiC is required by *Mycobacterium bovis* BCG for coenzyme F(420) and FO biosynthesis. *J. Bacteriol.* **184**, 2420–2428 (2002).
217. Ashby, K. D., Casey, T. A., Rasmussen, M. A. & Petrich, J. W. Steady-state and time-resolved spectroscopy of F420 extracted from methanogen cells and its utility as a marker for fecal contamination. *J. Agr. Food Chem.* **49**, 1123–1127 (2001).