THE UNIVERSITY
of EDINBURGH

Developing Dynamic Combinatorial Chemistry as a Platform for Drug Discovery

Thesis presented for the degree of

Doctor of Philosophy in Chemistry

by

Alexander G. Ekström

School of Chemistry

University of Edinburgh

Declaration

I herewith declare that all of the scientific work and experiments reported in this thesis are my own work unless otherwise noted. None of this work has been submitted in any other application for a higher degree.

Alexander G. Ekström

Table of Contents

Declaration 1

Table of Contents 2

Abbreviations 7

Preface 9

Acknowledgements 10

Lay Abstract 11

Abstract 12

Chapter 1: Dynamic Combinatorial Chemistry **14**

1.1 Introduction 14

1.1.1 Dynamic combinatorial chemistry 14

1.1.2 Protein-templated DCLs 15

1.1.3 The DCC experiment 16

1.1.4 *N*-Acylhydrazone (NAH) Exchange 19

1.1.5 *N*-Acylhydrazones as drugs 28

1.2 Challenges in DCC 29

1.3 A brief history of antibiotics 30

Chapter 2: Monitoring a DCL by ^{19}F NMR **35**

Abstract 35

2.1 Introduction 35

2.1.1 Fatty acid biosynthesis 35

2.1.2 FabH (β-ketoacyl-ACP synthase III) 37

2.1.3 Inhibiting FAS II 40

2.1.4 Inhibitors of FabH 41

2.1.5 ^{19}F NMR for fragment screening 43

2.1.6 Aims 46

2.2 Results and Discussion 47

2.2.1 Expression and purification of ecFabH 47

2.2.2 Nucleophilic catalysis of NAH exchange 49

2.2.3 Library selection 55

2.2.4 ^{19}F NMR pulse sequence 57
2.2.5 Blank DCL construction 58
2.2.6 Protein-templated DCL 58
2.2.7 DTNB assay and results 60
2.3 Conclusion 62
2.4 Experimental 64
2.4.1 Microbiology 64
2.4.1.1 Transformation of *ecFabH*/pUC 64
2.4.1.2 PCR amplification of *ecFabH* 64
2.4.1.3 Ligation of *ecFabH* into pGEM-T easy 65
2.4.1.4 Transformation and analytical restriction digest 65
2.4.1.5 Sequencing of *ecFabH*/pGEM 66
2.4.1.6 Restriction digest of *ecFabH* and pET-28a 66
2.4.1.7 Ligation of *ecFabH* into pET-28a 67
2.4.1.8 Transformation and analytical restriction digest of *ecFabH*/pET-28a 67
2.4.2 Expression and purification of ecFabH 68
2.4.3 Nucleophilic catalysis of NAH exchange 69
2.4.3.1 NAH formation from hydrazide 4 and aldehyde A 69
2.4.3.2 NAH formation from 3-membered forward library 69
2.4.3.3 NAH formation from 3-membered reverse library 70
2.4.4 DCL conditions 70
2.4.5 Length of relaxation time in ^{19}F NMR pulse sequence 70
2.4.6 FabH enzymatic DTNB assay 71
2.4.7 Synthetic chemistry 71
Chapter 3: Mechanistic Studies on a Promising DCC Core Scaffold 75
Abstract 75
3.1 Introduction 75
3.1.1 Covalent drugs 75
3.1.2 1,2-Dithiole-3-ones 78

3.1.3 4,5-Dichloro-1,2-dithiole-3-one 79
3.1.4 Aims 83
3.2 Results and discussion 84
3.2.1 Model reaction 84
3.2.2 Expression and purification of saFabH 87
3.2.3 HR45 modification 90
3.2.4 Digestion of saFabH/HR45 adduct 92
3.2.5 Structural studies 98
3.3 Conclusion 100
3.4 Experimental 101
3.4.1 Mass spectrometry 101
3.4.1.1 Peptide analysis 101
3.4.1.2 LCMS of denatured protein 101
3.4.2 NMR of model reaction of HR45 with *N*-acetylcysteine 102
3.4.3 Modification of *sa*FabH with HR45 103
3.4.4 Modification of saFabH with *N*-ethylmaleimide (NEM) 103
3.4.5 Removal of saFabH/HR45 modification with dithiothreitol (DTT) 104
3.4.6 Trypsin digest and thiol bead purification 104
3.4.7 Pepsin digest 105
3.4.8 Cloning of *saFabH* into pET-28a 106
3.4.8.1 Transformation of *saFabH*/pUC-19 106
3.4.8.2 Restriction digest of *saFabH* and pET-28a 107
3.4.8.3 Ligation of *saFabH* into pET-28a 107
3.4.8.4 Transformation and analytical restriction digest of *saFabH*/pET-28a 108
3.4.8.5 PCR Amplification of *saFabH* 108
3.4.9 General method for expression and purification of saFabH 109
3.4.10 Expression and purification of TEV protease 111

Chapter 4: Monitoring a Pyridoxal 5'-Phosphate DCL by Mass Spectrometry **113**

Abstract 113

4.1 Introduction 113

4.1.1 Tuberculosis 113

4.1.2 Biotin biosynthesis 114

4.1.3 BioA 116

4.1.4 PLP 118

4.1.5 Inhibitors of mtBioA targeting PLP 121

4.1.6 Mass spectrometry for protein analysis 124

4.1.7 ESI-MS in DCC 127

4.1.8 Aims 130

4.2 Results and Discussion 131

4.2.1 Protein expression and purification 131

4.2.2 Library selection 135

4.2.3 Library validation 137

4.2.4 DCC experiment 140

4.2.5 Control experiments 143

4.2.6 Results of protein-templated DCC experiments 148

4.2.6.1 mtBioA 149

4.2.6.2 K283A control 153

4.2.6.3 ecBioA 154

4.2.6.4 ecBioA reversibility control 156

4.2.7 *In vivo* assays 157

4.2.8 Enzyme assays 163

4.3 Conclusion 164

4.4 Experimental 166

4.4.1 Expression and purification of ecBioA 166

4.4.2 Expression and purification of mtBioA 167

4.4.3 Cloning, expression and purification of K283A mtBioA 168

4.4.4 Bradford Assay 168
4.4.5 Differential Scanning Fluorimetry (DSF) screen 169
4.4.6 DCC experiments 169
4.4.6.1 DCL conditions 169
4.4.6.2 DCL reversibility 170
4.4.7 Structural studies 170
4.4.8 Disc diffusion assays 171
4.4.9 *In vivo* assays 171
4.4.9.1 *E. coli* 171
4.4.9.2 *M. smegmatis* 172
4.4.9.3 *A. baumannii* 172
4.4.10 Synthetic Chemistry 173
5 Summary and Outlook **184**
6 Appendix 1 Protein amino acid sequences **186**
7 References **188**

Abbreviations

4-APA	4-amino-L-phenylalanine
ACE	acetylcholinesterase
ACM	amiclenomycin
ADT	anethole dithiolethione
AON	8-amino-7-oxononanoic acid
CA	carbonic anhydrase
CAD	collision-activated dissociation
CCDC	Cambridge Crystallographic Data Centre
COSY	homonuclear correlation spectroscopy
CV	column volumr
DAN	7,8-diaminononanoic acid
DCC	dynamic combinatorial chemistry
DCL	dynamic combinatorial library
DCMS	dynamic combinatorial mass spectrometry
DCX	dynamic combinatorial X-ray crystallography
DFT	density functional theory
Dha	dehydroalanine
DSF	differential scanning fluorimetry
DTB	dethiobiotin
DTNB	5,5-dithio-bis-(2-nitrobenzoic acid)
DTT	dithiothreitol
ESI-FTMS	electrospray ionisation Fourier-transform mass spectrometry
ESI-MS	electrospray ionisation mass spectrometry
FDA	US Food and Drug Administration
GST	glutathione-*S*-transferase
HPLC	high performance liquid chromatography
HTS	high throughput screen
IPTG	isopropyl β-D-1-thiogalactopyranoside
LCMS	liquid chromatography mass spectrometry

MIC	minimum inhibitory concentration
Mtb	*Mycobacterium tuberculosis*
MWCO	molecular weight cut-off
NAH	*N*-acylhydrazone
NCI	National Cancer Institute (US)
NEM	*N*-ethylmaleimide
NMR	nuclear magnetic resonance
PAIN	pan assay interference compound
PDB	Protein Databank
PLP	pyridoxal 5'-phosphate
PMP	pyridoxamine phosphate
SAM	*S*-adenosyl methionine
SAR	structure-activity relationship
SBD	structure-based design
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
STD	saturation transfer difference
TCI	targeted covalent inhibitors
TLM	thiolactomycin
VCL	virtual combinatorial library
VT	variable temperature

Preface

At the time of submission, parts of this thesis have either been published or manuscripts are in the process of being submitted to peer reviewed journals:

1. Ekström, A. G.; Kelly, V.; Marles-Wright, J.; Cockroft, S. L. and Campopiano, D. J., *Org. Biomol. Chem.,* **2017**, 15 (30), 6310-6313
2. Ekström, A. G.; Wang, J. T.; Campopiano, D. J. *et al., 'Non-invasive analysis of a protein-templated DCL by ^{19}F NMR*' - Manuscript ready for submission
3. Ekström, A. G.; Wang, J. T.; Serpico, A.; Marles-Wright, J. and Campopiano, D. J., *'Antimicrobial compounds from a pyridoxal 5'-phosphate DCL targeting BioA'* - Manuscript ready for submission

Acknowledgements

Firstly, I would like to thank Professor Dominic Campopiano for the opportunity to work on a very interesting and challenging project. I would also like to thank the members of the Campopiano group, past and present, for teaching me the art of carefully selecting the prize bottles from the proverbial wine cellar of knowledge, advice and enthusiasm into which our esteemed leader welcomes us with open arms.

I would like to thank my M.Sc. project students from over the years, especially Theresa Jue Wang for many insightful discussions. I would also like to thank the ever-helpful members of our NMR and MS facilities, without whom this PhD would have taken far longer. This work would not have been possible without my funding body SULSA.

Finally, I would like to thank my parents for their unwavering support and encouragement, and of course M^{lle} Pauline Coste for always providing a ray of sunshine through the clouds.

Lay Abstract

In the war against bacterial infections we are living in a time of growing unease, as we realise that our waning arsenal may not hold for much longer against the resurgence of illnesses previously consigned to history. Bacteria have succeeded in evolving strains resistant to almost every drug that we have thrown at them, and therein lies the challenge ahead. Scientists across many fields must work together in order to stay one step ahead of the constantly evolving bacteria. We must develop new diagnostics to detect bacterial infections, new antibiotic drugs to treat these infections and we must educate the global community on how to use them effectively so that we can reverse the bleak predictions for global public health.

The work described in this thesis explores a nascent method of drug discovery known as dynamic combinatorial chemistry (DCC). In DCC, many small 'drug-like' molecules are allowed to react with one another like Lego© bricks. The internal engine of an enzyme, the *active site*, can then act as a template to promote a specific combination of these molecules. Because this templated combination blocks the active site, it prevents the normal jobs of enzyme from taking place, deactivating the enzyme. We have chosen a selection of enzymes that are critical for the survival of various pathogenic bacterial strains, and attempted to deactivate these enzymes using molecular combinations from DCC experiments.

Because molecules are far too small to be observed with the naked eye, we rely on sophisticated techniques to 'see' how the molecules are behaving, specifically in our case to monitor which molecular combinations are being preferentially formed in a DCC experiment. In this thesis, we describe how we have developed these DCC-monitoring techniques to allow us to use larger numbers of compounds, giving the enzyme a greater selection and increasing the chance it of templating a combination that fits snugly within the active site. Through applying our work to our chosen bacterial targets, we identified new antibiotic compounds that either deactivated the target enzyme or stopped the bacteria from growing entirely.

Abstract

Dynamic combinatorial chemistry (DCC) is a powerful tool to identify new ligands for biological targets. In the technique, library synthesis and hit identification are neatly combined into a single step. A labile functionality between fragments allows the biological target to self-select binders from a dynamic combinatorial library (DCL) of interconverting building blocks. The scope of suitable reversible reactions that proceed under thermodynamic control in physiological conditions has been gradually expanded over the last decades, however DCC has thus far failed to gain traction as a technique appropriate for drug discovery in the pharmaceutical industry. The constraints placed on library size by validated analytical techniques, and the effort-intensive reality of this academically elegant concept have not allowed DCC to develop into a broad-platform technique to compete with the high-throughput screening campaigns favoured by medicinal chemists. This thesis seeks to develop DCL analysis techniques, in an effort to increase the library size and accelerate the analysis of DCC experiments.

Using a ^{19}F-labelled core scaffold, we constructed a DCL that could be monitored non-invasively by ^{19}F NMR. Building on NMR techniques developed by fragment screening and non-biological DCC campaigns, the method was developed to circumvent the undesired equilibrium-perturbing side effects arising from sample-consuming analytical methods. The *N*-acylhydrazone (NAH) DCL equilibrated rapidly at pH 6.2 using 4-amino-L-phenylalanine (4-APA) as a novel, physiologically benign, nucleophilic catalyst. The DCL was designed to target β-ketoacyl-ACP synthase III (FabH), an essential bacterial enzyme and antibiotic target. From the 5-membered DCL, a single combination was identified as a privileged structure by our ^{19}F NMR method. The result correlated well with an *in vitro* assay, validating ^{19}F NMR as a tool for DCL screening.

During the ^{19}F NMR study we identified an established antimicrobial compound, 4,5-dichloro-1,2-dithiole-3-one (HR45), to have potential as a core scaffold from which to

develop future DCLs targeting FabH. Despite the potentially tractable chemistry of HR45 for DCC, lack of knowledge around the inhibitory mechanism of the compound prevented us from proceeding. Thus, we used mass spectrometry, NMR and molecular modelling to show that HR45 acts by forming a covalent adduct with *S. aureus* FabH. The 5-chloro substituent directs attack from the nucleophilic thiol side chain of the essential active site cysteine-112 residue via a Michael-type addition elimination mechanism. Although interesting, this mechanism disfavoured the use of HR45 as a core scaffold for NAH exchange in a DCC campaign.

Electrospray ionisation mass spectrometry (ESI-MS) is a powerful technique that allows for larger DCLs by eliminating the size-limitations imposed by the need for spectral or chromatographic resolution of DCL members. We developed a 4-APA-catalysed NAH library targeting the pyridoxal 5'-phosphate (PLP) dependent enzyme 7,8-diaminopelargonic acid synthase (BioA), an essential enzyme in the biotin biosynthesis pathway. We exploited the aldehyde moiety of PLP to form an NAH DCL with a panel of hydrazides, and used the BioA isozymes from *M. tuberculosis* (*Mtb*) and *E. coli* to template the library. A combination of buffer exchange and denaturing ESI-MS allowed us to conduct a DCC experiment with a 29-member DCL. Hits from the DCC experiment correlated well with differential scanning fluorimetry (DSF) results. Of these hits, 5 compounds were selected for further study. *In vivo* activity was displayed by 2 compounds against *E. coli* and the *ESKAPE* pathogen *A. baumannii*. The identification of compounds with antibacterial activity from a DCL further validates ESI-MS as a platform technology for drug discovery.

Chapter 1: Dynamic Combinatorial Chemistry

1.1 Introduction

The field of combinatorial chemistry has received widespread attention since its advent in the late 1980s[1,2]. In cleverly applying synthetic organic chemistry, combinatorial chemistry presents a rapid and cost-effective route to large libraries through the rational selection of irreversible chemical reactions that proceed with high yield and specificity. The technique has been particularly instrumental in the field of drug discovery, where split-and-pool synthesis has been used to generate diverse arrays of drug-like molecules for high throughput synthesis (HTS) and lead refinement.

A significant hurdle in split-and-pool combinatorial synthesis lies in the requirement for each compound to be isolated and characterised before being screened for activity against a target, a process not entirely trivial considering the complexity of the molecular arrays. Dynamic combinatorial chemistry (DCC) attempts to solve this problem by combining library synthesis and activity screening into a single step. A dynamic combinatorial library (DCL) is a reversible combinatorial library under thermodynamic control, whose composition is dynamic in the sense that it is able to reorganise through the stabilizing influence of external stimuli[3].

1.1.1 Dynamic combinatorial chemistry

The DCC concept was first introduced by the groups of Sanders and Lehn in the 1990s, who sought to combine the recent developments in combinatorial chemistry with the error-checking and proof-reading functions carried out in nature by reversible reactions[4–6]. To further develop and demonstrate the ingenuity of the concept, DCC has since been successfully applied by many academic groups to a variety of different fields including dynamic polymers[7], self-assembly systems[8–11] and sensor design[12]. The application of DCC with the most potential impact is in drug discovery, where DCLs have been assembled and templated by biological targets of interest to identify biologically active small molecules[13,14]. Despite the academic validation of protein-

templated DCC, the technique has not been adopted by the pharmaceutical industry and has thus far failed to deliver any drug candidates to the clinic. A significant weakness of DCC lies in the lack of a robust and widely reproducible method to analyse the outcome of an experiment. Our work has focused on exploring and developing the analysis of protein-templated DCLs to aid the transition of DCC from an academic exercise to a results-focused drug discovery platform.

A DCL is composed of functionalised building blocks that can react reversibly with one another to form a dynamic library of interconverting species whose distribution strives towards the thermodynamic minimum of the system (see fig. 1.1). Because the library members are constantly interconverting, under the correct conditions the introduction of a selection pressure can perturb the equilibrium. For example, if a template (such as a receptor or an enzyme) is added, one or more library members may bind to the template through non-covalent interactions, stabilising those library members. In which case, the equilibria between library members will reorganise to find the new thermodynamic minimum, resulting in an increased concentration (amplification) of the energetically stabilised member at the expense of other library members.

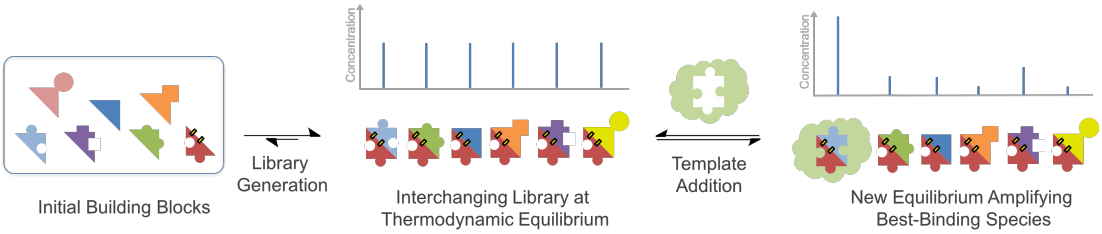


Figure 1.1 - Scheme describing the DCC concept.

1.1.2 Protein-templated DCLs

The application of DCC to drug discovery was pioneered by Huc and Lehn, and since then many examples of protein-templated DCLs have been published[4]. The approach has proved successful for identifying ligands for biomolecules and mapping the active

site of proteins for which structural information is not available[13,14]. In this application of DCC, the 3D architecture of the protein active site acts as a template on the interconverting DCL, stabilising any members that bind to it. This results in amplification of the best binding species at the expense of other library members. What results is a target-driven, self-screening process that combines the attributes of combinatorial chemistry with HTS in a single step[13]. However, it must be noted that the tightest binding member is not necessarily always also the most biologically active[15,16]. The largest obstacle to protein-templated DCC is the requirement to conduct the experiment under conditions that are not detrimental to the protein. The spectrum of reversible covalent reactions that operate on a reasonable timescale under physiological conditions is narrow, and much work has been done to expand this toolbox. A further challenge lies in conducting a rapid and reproducible analysis of protein-templated DCLs without further perturbing the finely balanced equilibrium obtained in the experiment. This final challenge forms the central theme of this thesis.

1.1.3 The DCC experiment

A DCC experiment is typically designed around three dominating factors: (i) the choice of library members or 'building blocks', (ii) the choice of reversible reaction, and (iii) the method by which the experiment will be analysed.

Firstly, the building blocks must be carefully selected. Either by building on the knowledge of known substrates and inhibitors, by *in silico* screening of fragment libraries, or by rational library design using 3D X-ray structures, the library members must be chosen to cover a wide range of chemical and molecular space within the receptor or binding site, considering the important residues and binding interactions present. For these interactions to be optimised, geometric organisation can be achieved by using a core scaffold. This may be a known substrate of the target, whose binding properties can be exploited and improved, or simply a central organisational structure with carefully determined vectors and topicity[17]. Most importantly, the

building blocks must all be sufficiently soluble and functionalised with a chemical handle to allow them to take part in the selected reversible reaction. This functionality should not interfere with the recognition event, and they should be orthogonal in that there are no competing reactions with other functionalities on library members, or with residues on the target. Ideally the building blocks should be isoenergetic, so that library equilibration is not diverted by kinetic traps. This condition is practically challenging, given the desire to impart chemical diversity on the DCL, however a near-isoenergetic DCL can be assembled by careful selection of building blocks[3].

The choice of reversible chemical reaction that mediates the interconversion of DCL members is critical. The ideal reaction should proceed under mild or physiological conditions, and do so on a reasonable timescale to avoid degradation of the target. It should not be central to the recognition events between the DCL and the target, and the equilibration rate of the reaction should be slower than the on/off rate of the target, as this would prevent recognition events from taking place. In many literature examples, it has been necessary to freeze the exchange process after equilibrium has been reached, either chemically (for example by reducing an imine DCL to its amine form with $NaBH_3CN$), or kinetically, to effectively halt the exchange process (for example by a change in pH or temperature), aiding analysis of the DCL composition[14]. A list of reversible covalent reactions that have been demonstrated in protein-templated DCC is shown in fig. 1.2. Of these, *N*-acylhydrazone (NAH) exchange represents one of the most useful and well-characterised reactions for DCC. The work in this thesis concentrates primarily on this reaction, which will be discussed in detail in the following section.

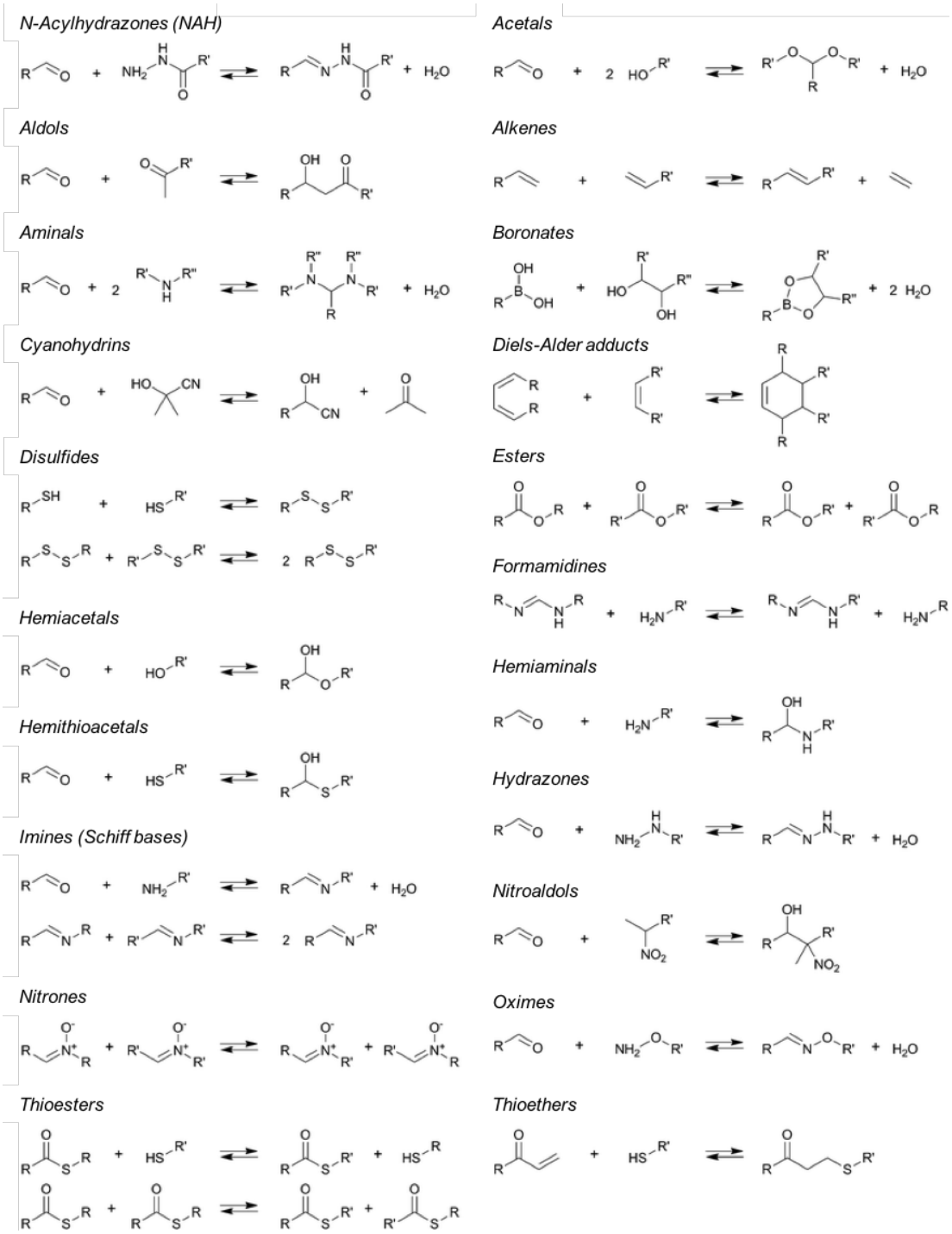


Figure 1.2 - Reversible covalent reactions used in DCC to find bioactive molecules[13].

Finally, the method employed to analyse the DCL is central to the DCC experiment. Depending on the size of the DCL, many different methods to identify or isolate best binding or amplified hits have been used. DCLs have been analysed by HPLC[4,18–28],

size exclusion chromatography (SEC)[29], nuclear magnetic resonance (NMR) spectroscopy[30,31], X-ray crystallography[32], fluorescence assay[33–35], dynamic deconvolution[22,36,37] and mass spectrometry (MS)[38–42]. While each technique has its benefits, many suffer from drawbacks that occur with increasing library size. Techniques involving spectral or chromatographic identification often require library members to be individually synthesised for characterisation, and other techniques are simply not amenable to developing a results-driven drug discovery platform. These cumbersome and time consuming DCL analysis methods detract from the elegance and efficiency of a DCC experiment, and may carry some of the blame for the failure of DCC to be accepted as a drug discovery technique by the pharmaceutical industry. It is the central theme of this thesis to explore and improve DCC analyses.

1.1.4 *N*-Acylhydrazone (NAH) Exchange

Prior to the development of NAH for DCC, other C=N forming reactions were used. In the early days of DCC, imine virtual combinatorial libraries (VCLs) were pioneered by the Lehn group to isolate inhibitors of a carbonic anhydrase II (CA), a well-characterised zinc dependent enzyme[4]. The 15-member library was synthesised with a 3-fold excess of $NaBH_3CN$, reducing the imines to aldehydes as soon as they were formed. The blank library equilibrated at physiological pH within 24 h, however the same library synthesised in the presence of CA required 14 days at 25 °C before a stable product distribution had been reached. The reduction step was necessary because the kinetic barrier to imine hydrolysis is low, and the products proved too transient to analyse by HPLC, representing is a serious drawback to the reaction.

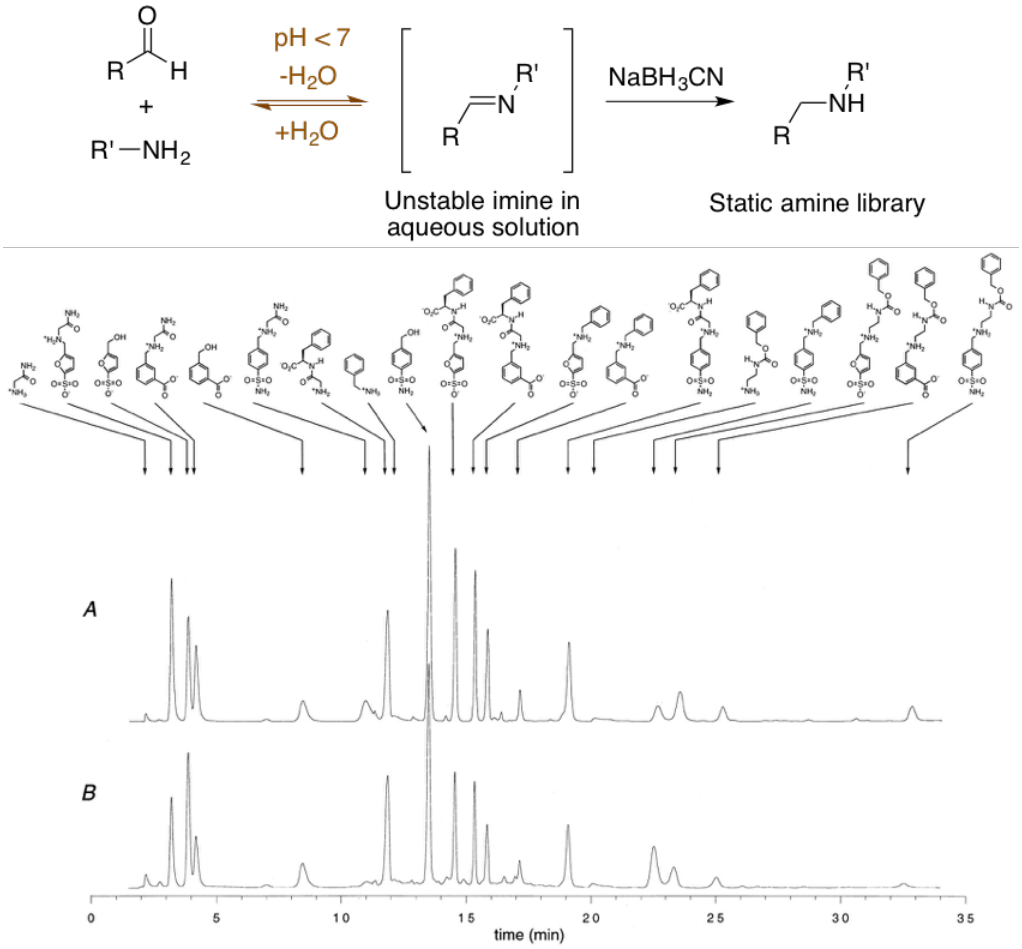
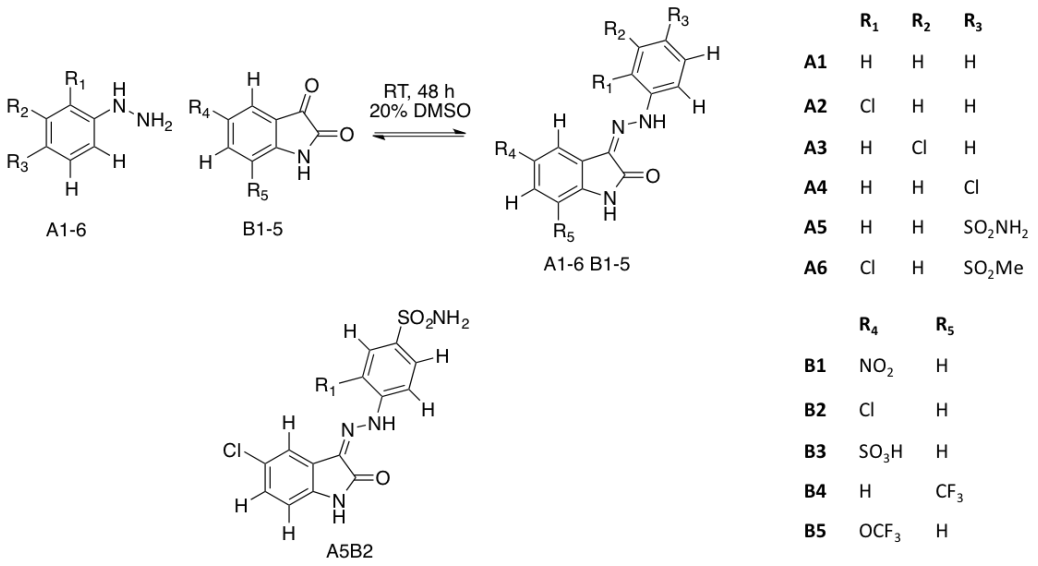


Figure 1.3 - The imine exchange reaction, followed by reduction with $NaBH_3CN$, as used by Lehn in his original protein-templated DCL targeting CA. Chromatogram **A** and **B** show the distribution of 15 reduction products and 7 building blocks in the absence and presence of CA, respectively. Image modified from Huc and Lehn[4].

Congreve *et al.* developed a technique called dynamic combinatorial X-ray crystallography (DCX) to analyse a hydrazone DCL targeting cyclin-dependent kinase 2 (CDK2), an enzyme implicated in a number of cancers[32]. Their 30-member hydrazone library was equilibrated over 48 h in the presence of CDK2 crystals. Electron density maps confirmed that a single hydrazone **A5B2** had bound to the crystal, which was found to have an IC_{50} of 30 μM. Congreve *et al.* recognised that a drawback of previously published DCC experiments was that amplification could only be detected by comparing identical libraries in the presence and absence of the

target protein. Instead they sought to use the dynamic nature of the library to discriminate against non-binding species, shrinking the effective number of library members. This was the first example of a direct screening method for analysing a DCL, circumventing the need to laboriously interrogate or characterise the library members prior to analysis.



	R_1	R_2	R_3
A1	H	H	H
A2	Cl	H	H
A3	H	Cl	H
A4	H	H	Cl
A5	H	H	SO_2NH_2
A6	Cl	H	SO_2Me

	R_4	R_5
B1	NO_2	H
B2	Cl	H
B3	SO_3H	H
B4	H	CF_3
B5	OCF_3	H

Figure 1.4 - The hydrazone DCX assembled by Congreve *et al.* targeting CDK2.

Taking inspiration from an earlier example of a self-assembling NAH kinase C inhibitor[43], the introduction of NAH exchange to DCC (Fig. 1.4), first by Cousins *et al.* in an abiotic DCL[44], and then by Bunyapaiboonsri *et al.* in a protein-templated DCL targeting acetylcholinesterase (ACE), represented a big step forward in DCC. The NAH products show higher hydrolytic stability than imines in aqueous solution, with the equilibrium constant lying much further to the right[45]. However, the reaction kinetics of NAH exchange are extremely slow at physiological pH and require a pH < 4 to equilibrate on a reasonable timescale, making the reaction unsuitable for most proteins[36]. To combat this, Bunyapaiboonsri *et al.* assembled a DCL of 66 possible products, which was equilibrated for 1 week at pH 4. Once the DCL had equilibrated, the pH was increased to 7.2 and interacted with ACE. The advantage of this

experiment was that many compounds could be synthesised in a single step, however this removes the adaptability of the library and is more comparable with static combinatorial libraries.

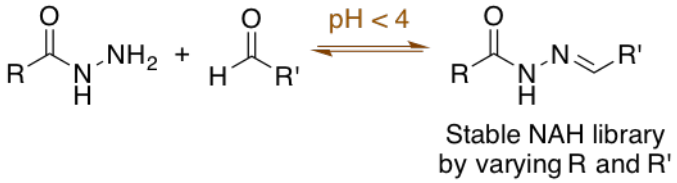


Figure 1.5 - The NAH exchange reaction used Bunyapaiboonsri *et al.* to assemble a pre-equilibrated library targeting ACE.

The elegance of the experiment by Bunyapaiboonsri *et al*. was in the analysis of the pre-equilibrated DCL. In what they called *dynamic combinatorial deconvolution*, they assayed the entire DCL against the target, sequentially removing single building blocks from the pre-equilibrated DCL to quantify the effect that each building block contributed to overall inhibition. They were therefore able to map the potential of each building block, asserting a structure-activity relationship (SAR) and identifying an ACE inhibitor with an IC_{50} of 2.3 nM[36].

The Lehn group went on to demonstrate dynamic combinatorial deconvolution by assembling a 440-member NAH DCL targeting HPr kinase/phosphatase (HPrK/P) from *B. subtilis*, resulting in an inhibitor with an IC_{50} value of 17 μM. This was followed by a 474-member NAH DCL targeting the plant lectin concanavalin A, resulting in an inhibitor with an IC_{50} value of 22 μM. It would have been incredibly challenging to deconvolute such a large library using spectral or chromatographic techniques due to the likelihood of signal overlap. Dynamic combinatorial deconvolution also has the benefit that it selects privileged library members on the grounds of inhibitory activity rather than binding strength, an asset that differentiates the technique from most other analysis methods[21,22,36,37].

The next significant evolution in NAH exchange for DCC came when Dawson and colleagues applied earlier work by Cordes and Jencks on the use of aniline to catalyse semicarbazone formation (Fig. 1.6)[46,47]. Dirksen *et al.* investigated the nucleophilic catalysis of aniline on the formation and hydrolysis of NAH peptides. A common problem among previous NAH examples was that the formation of hydrazones was relatively rapid (ca. 24 h), but the hydrolysis reaction was slow (days to weeks). Dirksen *et al.* showed that aniline could impart a 70-fold rate enhancement with no effect on the rate constant, expediting NAH equilibration by allowing both the hydrolysis and formation reactions to complete within the space of hours at near-physiological pH (5.7). They proposed that aniline-mediated transimination proceeded via a transient Schiff base intermediate, whose aqueous instability allowed relatively high catalytic loading. This discovery opened the doors of NAH DCC to many previously unsuitable protein candidates.

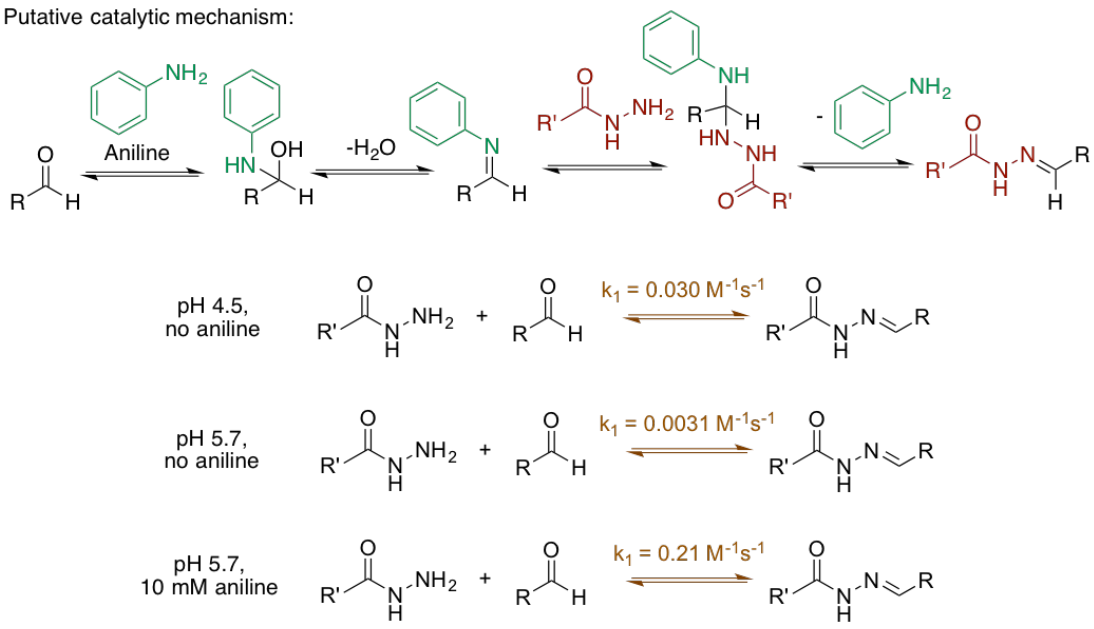


Figure 1.6 - The use of aniline by Dirksen *et al.* to catalyse NAH exchange. In this example, the aldehyde and hydrazide are both at a concentration of 1 mM, and the rate of hydrazone formation is enhanced 70-fold at pH 5.7 by the addition of 10 mM aniline.

Aniline-mediated NAH catalysis was leveraged by Greaney and Campopiano, who assembled a 10-membered DCL to target the enzyme glutathione-*S*-transferase (GST)[20]. They demonstrated that aniline could indeed be used for DCC, allowing their fully reversible NAH library to equilibrate within 6 h in the presence of 10 mM aniline at pH 6.2, compared to 5 days in the absence of catalyst. The library was assembled from 10 commercially available hydrazides and an aldehyde derivative of the know GST substrate, chloro-2,4-dinitrobenzene (Fig. 1.7). To kinetically freeze their equilibrated DCL for HPLC analysis, they increased the pH to 8 with the addition of NaOH. This effectively stops NAH exchange, even in the presence on aniline. The protein was then removed from the sample by ultrafiltration and analysed by HPLC. Elegantly, the DCL was separately templated by two different GST isoforms, each resulting in different amplified species. NAH **3g** was amplified in the presence of the helminth worm *Schistosoma Japonicum* isoform (sjGST), and NAH **3c** was amplified by the human isoform (hGST). In an important control experiment, a catalytically-inactive sjGST mutant also amplified NAH **3g**, demonstrating that the amplification was due to a perturbation of the thermodynamic equilibrium distribution rather than a kinetic effect of the enzyme. Owing to poor solubility, Bhat *et al*. were unable to assay the amplified hits, so they conjugated aldehyde **1** to glutathione (GS), increasing both solubility and recognition elements for the enzyme. The second-generation GS-conjugated library afforded the same hits, with IC_{50} values of 22 μM and 57 μM against sjGST and hGST, respectively (see fig. 1.7)[20]. This demonstrates both the sensitivity of a DCL, and its use not only in identifying binders, but also mapping active site interactions. GST has a highly specific G-site and non-specific and notoriously hard to target H-site. The GS-conjugated DCL allowed Bhat *et al.* to interrogate the H-site through being simultaneously being anchored to the G-site. The same group went on to develop a bivalent NAH DCL templated by four different GST isoforms, identifying a number of inhibitors in the nM range[18].

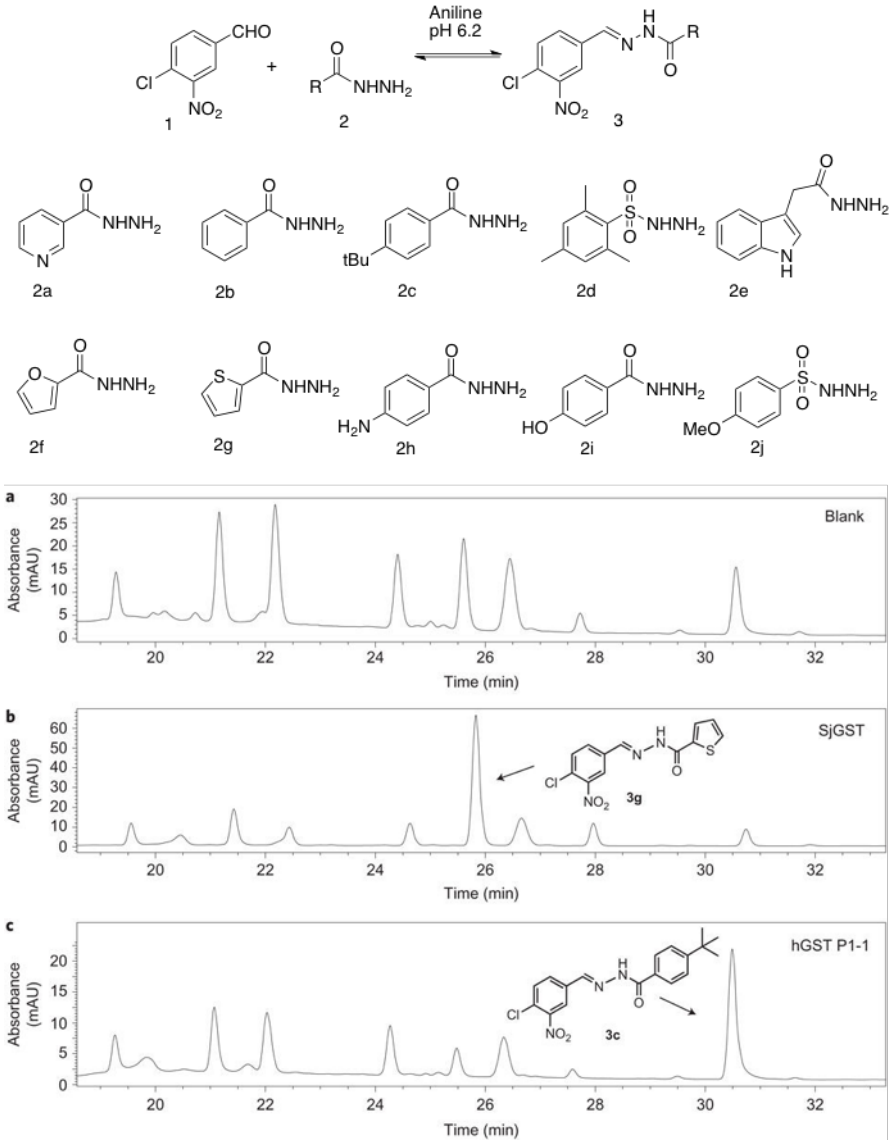


Figure 1.7 - The aniline-catalysed NAH DCL by Bhat *et al.* targeting GST, showing isoform selectivity from a 10-membered library. Adapted from Bhat *et al.*[20].

Mondal *et al.* combined *de novo* structure-based design (SBD) and DCC to assemble a DCL targeting the aspartic protease endothiapepsin[30]. By analysing X-ray crystal structures of the protein in different binding modes and co-crystallised with 11 fragments, they identified moieties capable of binding in different conformations of the 2 discreet binding pockets. Using molecular modelling, they proposed a series of 5 NAH-based inhibitors, which were retrosynthesised into 5 aldehydes and 5 hydrazides. Each aldehyde was mixed with the 5 hydrazides, forming 5 separate 5-

membered libraries (fig. 1.8). The 5 DCLs were pre-equilibrated over 24 h in the absence of aniline before the enzyme was added and each system was separately analysed by ^{1}H saturation transfer difference (STD) NMR to identify binders. Since the optimum operating pH of endothiapepsin is 4.6, it allowed Mondal *et al.* to equilibrate the library in the absence of aniline[30]. This represents a rare example of a protein that is amenable to catalyst-free NAH DCC. From the 5 sub-libraries, they identified 8 binders which were validated as inhibitors by enzymatic assay with IC_{50}s in the low micromolar range. The binding mode of two of the identified NAH inhibitors was confirmed by co-crystal structures with the protein. Whereas ^{1}H-STD NMR would not be appropriate for analysing larger, more diverse DCLs due to issues with spectral overlap, cost and the quantity of protein required, Mondal *et al.* demonstrated an elegant use of the technique in analysing a small, rationally designed DCL.

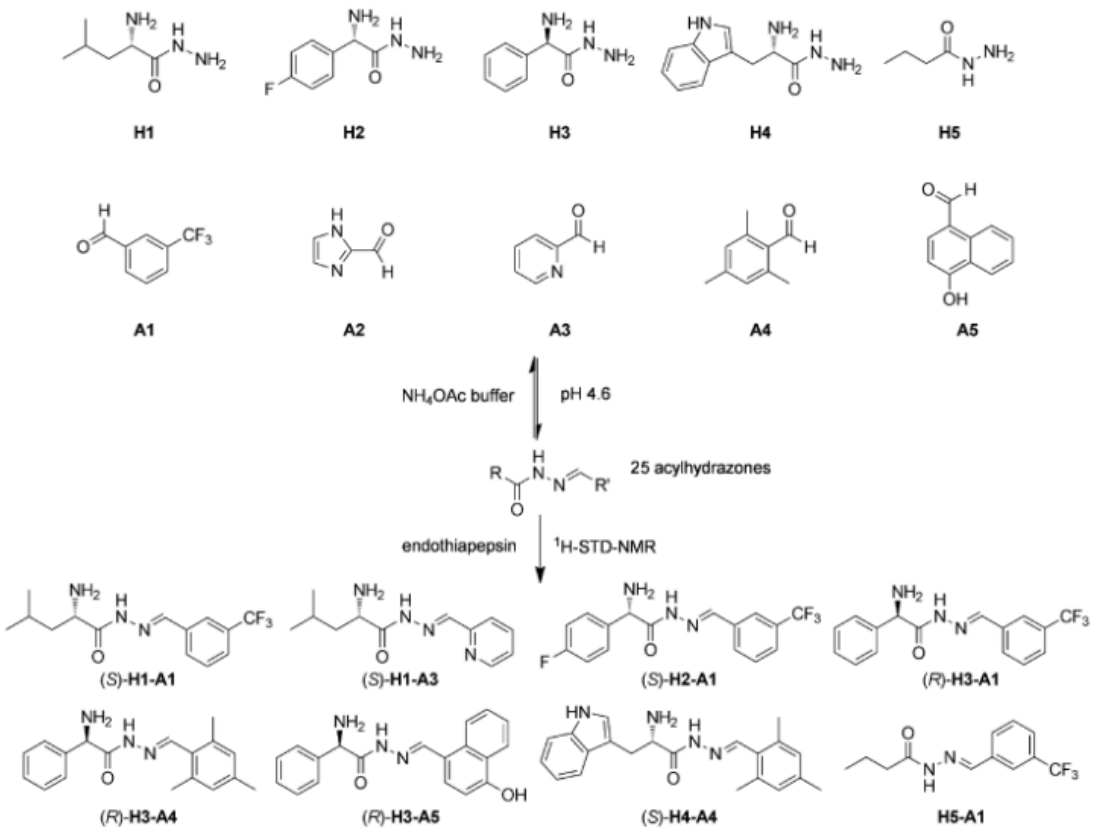


Figure 1.8 - The 25-member DCL from Mondal *et al.* composed of 5 pre-equilibrate 5-member DCLs targeting endothiapepsin, with the 8 binding NAH hits identified by ^{1}H STD-NMR[30].

While many of DCC analysis techniques of NAH DCLs have been discussed in this section (also tabulated in table 1.1), we have omitted mass spectrometry related techniques as these are reviewed in depth at the beginning of Chapter 4.

Target	DCL size	Deconvolution technique	IC_{50} or K_I of best binder (μM)	Ref.
Acetylcholinesterase	66	Dynamic deconvolution	0.0023	[36]
Carbonic anhydrase II	10	ESI-MS	0.0106	[42]
Concanavalin A	474	Dynamic deconvolution	22	[22]
Glutathione *S*-transferase	10	HPLC	22	[20]
Glutathione *S*-transferase	24	HPLC	0.05	[18]
HPr kinase/phosphatase	440	Dynamic deconvolution	17	[37]
Aspartic protease endothiapepsin	25	^{1}H STD NMR	12.8	[30]
Aspartic protease endothiapepsin	9	Fluorescence assay	85	[33]
GABA transporter 1 (GAT1) [a]	36	LC-ESI-MS/MS	0.66	[41]
Cyclin-dependent kinase 2 (CDK2) [a]	30	X-ray crystallography	0.03	[32]

Table 1.1 - DCC studies searching for biologically active compounds using DCLs based on NAH-exchange. [a] The GAT1 and CDK2 DCLs were built on hydrazone, not NAH exchange.

1.1.5 *N*-Acylhydrazones as drugs

NAHs are currently enjoying a renaissance in drug discovery and have recently been dubbed *privileged structures* owing to the wide scope of activity that NAH lead compounds have shown[48]. There are literature examples of NAHs with anticonvulsant, antidepressant, analgesic, antimalarial, antimicrobial, antimycobacterial, antitumoural and antiviral activities[49]. Recently, NAHs have been employed against more fashionable targets including histone deacetylase 6/8 (HDAC6/8), influenza virus PA endonuclease and *M. tuberculosis* targets[48,50,51]. The

presence of numerous literature examples confirms the biological compatibility of the NAH moiety and validates NAH DCC as a promising technique for finding biologically active lead compounds.

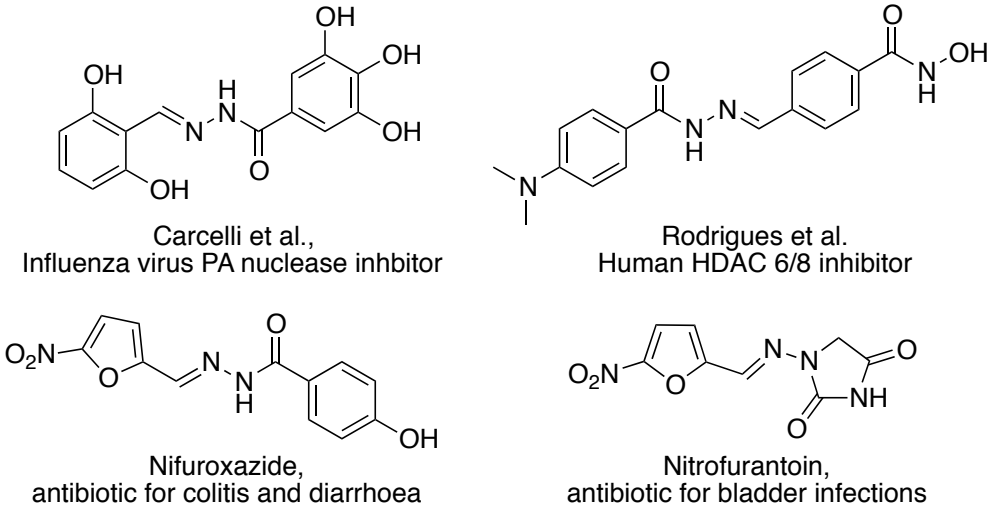


Figure 1.9 - Examples of biologically active NAHs with different targets[48,50].

1.2 Challenges in DCC

Since the advent of fragment-based drug discovery (FBDD) 30 years ago, several industrial and academic groups have adopted the approach and it has resulted in over 30 candidates reaching the clinic, with Vemurafenib (Plexxikon) and Venetoclax (AbbVie, Genentech) winning FDA approval[52]. The idea of dynamic combinatorial chemistry (DCC) in the context of drug discovery was born at approximately the same time, and the two have evolved alongside one another. The striking difference between the two is that despite numerous academic examples of success in demonstrating the DCC approach to find inhibitors, the pharmaceutical industry has by and large declined to adopt the approach and a DCC-derived compound is yet to reach the clinic. DCC took a brief foray into the industrial sector with Therascope AG, who developed imine dynamic combinatorial libraries (DCLs) against the influenza target neuraminidase[27]. The company was subsequently bought by Alantos (Amgen), but the DCC project was abandoned.

For DCC to have an impact on drug discovery, it must first be established why it has thus far failed to do so. At the advent of DCC, one of the cornerstones of the principle was that the approach combined synthesis and screening into a single step, supposedly simplifying the process of FBDD followed by fragment linking[53]. Instead, the field has developed into a screening approach that measures ligand enrichment from a dynamic library, whereby the library is screened in the absence and presence of a target protein, and the distribution of ligands in each is compared[54]. The complexity of a DCC experiment increases with library size, owing to chromatographic or spectral overlap in common deconvolution techniques (HPLC, NMR)[54]. Additionally, the requirement to synthesise each possible library combination to confirm its assignment undermines the argument that DCC makes drug discovery more efficient by combining synthesis and screening.

Only when direct screening methods are developed that are not inhibited by library size or the requirement to synthesise each member individually, will DCC become a viable option for industrial drug discovery. The work in this thesis seeks to develop existing DCL analysis and devise new methods to address existing issues, with the hope of taking DCC a step further towards acceptance as a results-driven drug discovery technique. By applying DCC to different antibiotic targets, we hope to demonstrate the versatility of the technique.

1.3 A brief history of antibiotics

Almost all antibiotics that are used in the clinic today are derivatives of compounds that were first discovered in what is now referred to as the *golden era* of antibiotic discovery. This period of antimicrobial natural product discovery spanning the 1940s-1960s was ignited by the famously serendipitous discovery of penicillin by Alexander Fleming in 1928[55], which was later developed for use in the clinic by Florey and Chain. The three scientists were awarded the Nobel Prize for Medicine in 1945 in recognition of their pioneering work on the discovery and applications of the β-lactam penicillin, the natural and synthetic derivatives of which are credited with saving millions of

lives. A few years later, the sulfadrug prontosil was discovered by Domagk and co-workers in a small screen of dyes for antibacterial activity[56]. In the early 1940s, Selman Waksman developed the first platform to screen soil bacteria for antimicrobial activity, leading to the discovery of streptomycin[57]. This screening approach was broadly adopted by academia and industry, and led to the discovery of several new classes of antibiotic.

The quinolone class of antibiotics was developed in the early 1960s after discovery of the antimicrobial activity of nalidixic acid, a compound found in the distillate of an attempted synthesis of the antimalarial drug chloroquine[58]. By the late 1960s, the discovery output from the Waksman platform had dried up. In response, the community began filling the pipeline with synthetic analogues of previously discovered natural products. This was to spell the end of the *golden era* of antibiotic discovery. Since the 1960s no new class of broad-spectrum antibiotics has been discovered[56], and the most recent clinically useful class of antibiotic discovered was the lipopeptide daptomycin in 1986[56].

Towards the end of the twentieth century, the community turned its interest towards synthetic small-molecules to expand search for new antibiotics, hoping to exploit the rapid advances in synthetic medicinal chemistry. However, despite initial promise and huge investment, the small-molecule approach in the search for antibiotics has proved astoundingly barren. High-throughput screening (HTS) failed to produce a single compound with a reasonable spectrum of *in vivo* activity until 1997, when bedaquiline was discovered to have specific antibacterial activity in a HTS campaign against mycobacteria[59]. Many lead compounds with *in vitro* activity have been generated against both Gram-positive and Gram-negative bacterial targets, however none of these have been translated into clinically useful molecules. This failure has been attributed to the fact that bacterial cells do not conform to the rules that medicinal chemists have successfully established to improve efficacy against eukaryotic targets[60].

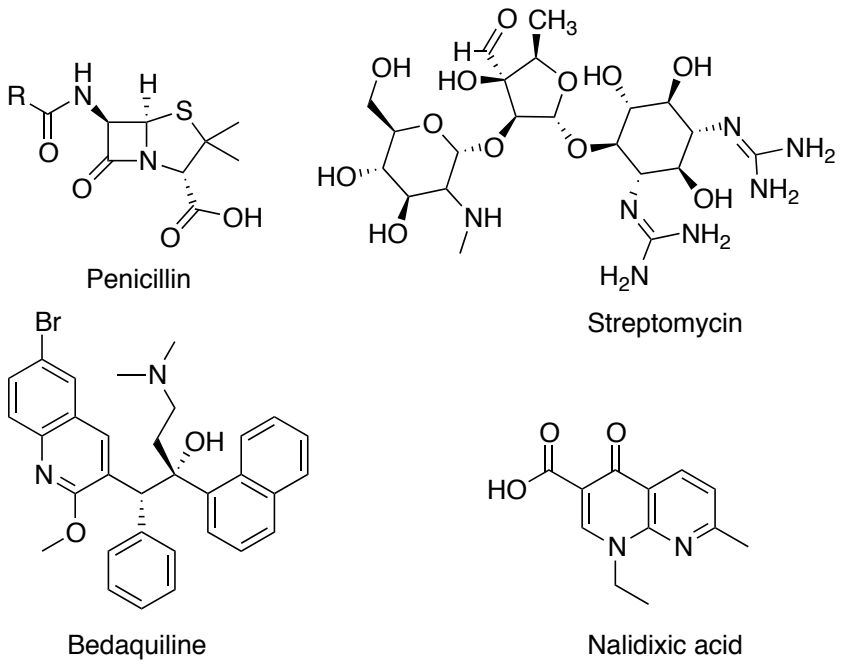


Figure 1.10 - The structures of mentioned antibiotics.

Bacterial cells have evolved ingenious, multi-faceted methods of keeping unwanted compounds out or breaking them down. Gram-negative bacteria are particularly well protected, with an outer membrane that prevents the passage of amphipathic molecules (a feature of most small-molecule drugs to increase their bio-availability). Bacteria have also evolved multidrug-resistant (MDR) efflux pumps that remove a wide variety of structurally unrelated compounds from the cell, with high specificity for hydrophobic cations[61]. Finally, to complete the defence system the inner membrane poses a barrier to hydrophilic compounds. Almost all established antibiotic compounds do not obey Lipinski's rules, therefore a generic medicinal chemistry approach is not suitable for antibiotic development.

Modern medicine has striven to maintain high standards in clinical trials and drug safety studies. While this is by no means a bad thing, it has added to the demise of antibiotic discovery. New antibiotics are produced to fight resistant pathogens, which currently only infect a small minority of patients. Ethical concerns also prevent placebo studies on patients with MDR-infections, making clinical trials of new

antibiotics difficult to define and therefore expensive. This only worsens the already poor return on investment in antibiotic research; a therapy course is typically 1-2 weeks long, and owing to the inevitable problem of antibiotic resistance the drug will have a finite useful lifetime. The global community is beginning to respond with initiatives such as the approval of the *GAIN (Generate Antibiotic Incentives Now) Act* in the US in 2012, and the WHO *Global Action Plan on Antimicrobial Resistance, which* aim to generate economic incentives or at least lower the financial barriers to antibiotic research.

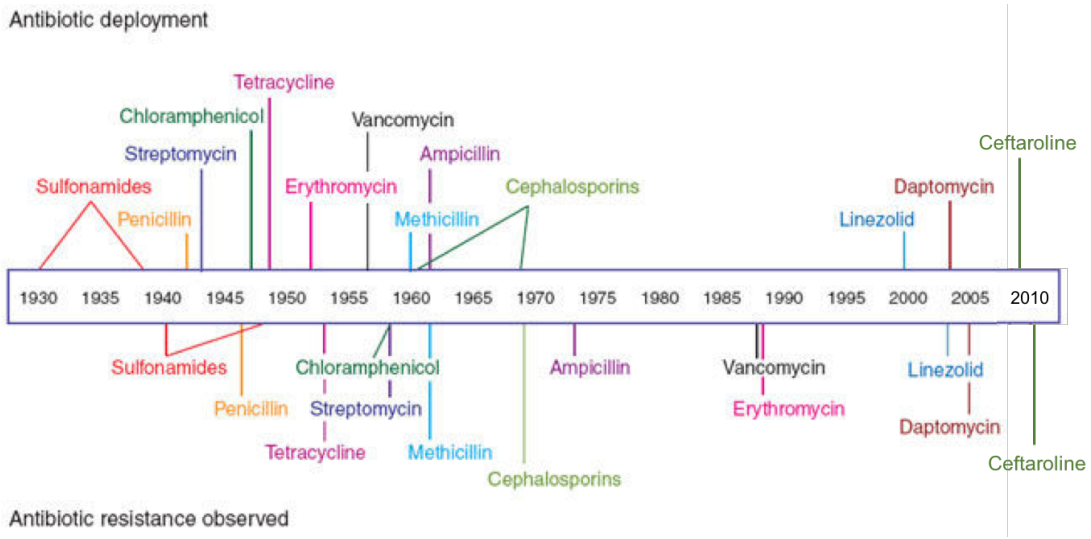


Figure 1.11 - A timeline showing the rapid emergence of resistance to antibiotics. The top line indicates when the antibiotic was first discovered, and the bottom line indicates when resistance was first observed[62]. Image reproduced with permission from Springer Nature, license number 4317580582151.

In his Nobel Prize acceptance speech, Fleming had the foresight to warn that *'…the ignorant man may easily under-dose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant'*. Ever since the first antibiotics were brought into the public domain, resistance has emerged in every case (see fig. 1.11) and today we are faced with the rapid spread of MDR or extensively drug-resistant (XDR) species. These include both Gram-positive and Gram-negative bacteria that

commonly cause nosocomial or community infections. These have been dubbed *ESKAPE* pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species)[63]. Until recently, the glycopeptide vancomycin was thought of as the antibiotic of last resort. The first case of a commonly acquired *S. aureus* infection with reduced susceptibility to vancomycin was reported in 1996[64], followed by a strain of fully vancomycin-resistant *S. aureus* in 1999[65]. There are now reports of resistance emerging in nosocomial *A. baumannii* infections to all known antibiotics[66].

Bacteria have managed to exploit every possible method of undermining the activity of antibiotics. By destruction, modification, efflux or subverting specific cellular functions bacteria rapidly develop resistance to new antimicrobial compounds[56]. Current interest is focused on persisters, these are dormant, non-dividing phenotypic variations of bacterial cells that escape the action of antibiotics that target active cell processes. Persisters can reactivate and begin dividing as antibiotic levels fall, causing a relapse of infection and accelerates the development of antibiotic-resistant mutants[56]. We are faced, therefore, with an urgent need to develop not only new antibiotics, but drugs that target novel pathways in both actively dividing and dormant cells.

Chapter 2: Monitoring a DCL by ^{19}F NMR

Abstract

Dynamic combinatorial chemistry (DCC) is a powerful tool to identify new ligands for biological targets, however the technique has failed to gain traction in industrial drug discovery. Only when rapid, reproducible and cost-effective methods to analyse DCC experiments are developed, will DCC become a genuine alternative to traditional drug discovery platforms. In this chapter, we discuss the construction of a dynamic combinatorial library (DCL) targeting β-ketoacyl-ACP synthase III (FabH), an essential bacterial enzyme involved in fatty acid biosynthesis. The DCL was developed specifically to contain fluorine-labelled products. This allowed the DCC reaction to be monitored non-invasively by ^{19}F NMR without the need for destructive analytical methods, which we believe perturb the finely balanced equilibrium attained during a DCC experiment. From the 5-membered DCL, a single combination was identified as a privileged structure by ^{19}F NMR analysis, and it was confirmed to be the most efficacious inhibitor of FabH by an *in vitro* enzymatic assay.

2.1 Introduction

2.1.1 Fatty acid biosynthesis

Fatty acid synthesis (FAS) is the essential biological process by which fatty acids are derived from short-chain acyl-CoAs. Higher eukaryotes possess a single gene that encodes a large protein termed fatty acid synthase I (FAS I), which contains all of the functionalities required to produce a fatty acid[67]. In bacteria, FAS is carried out in an analogous process by fatty acid synthase II (FAS II), a dissociated system where individual proteins, each responsible for a single step in the pathway, are coded for by separate genes. Whilst there is homologous architecture between functionalities in FAS I and FAS II, there are also significant differences that make the systems distinct. FAS II is not only responsible for producing long-chain fatty acids, as when the growing fatty acid chain is carried between FAS II proteins by the acyl carrier protein (ACP), the intermediates may be diverted into other biosynthetic pathways[67,68]. These attributes make the essential bacterial process of FAS II an

attractive target for the development of new antimicrobials[69]. FAS II in *E. coli* has been regarded as a model system, and although the biosynthetic pathway is analogous amongst all bacteria, there are slight variations in substrate specificity regarding carbon-chain length and branching. Unless stated otherwise, the current chapter addresses enzymes of the *E. coli* FAS II system.

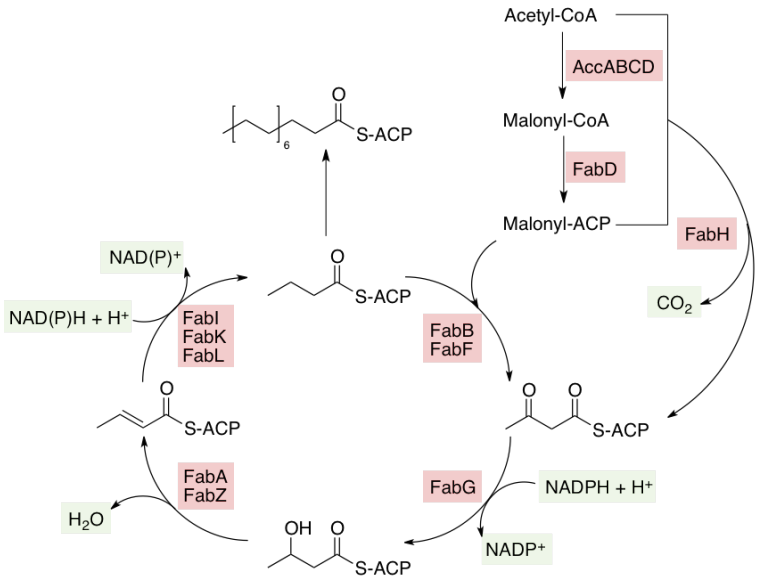


Figure 2.1 – The FAS II cycle. The enzymes are highlighted in red, in order of increasing chain length specificity when more than one is listed. With each turn of the cycle, the chain length is increased by two carbon units in the form of malonyl-ACP, catalysed by FabB/F.

ACP is a small acidic protein (ca. 10 kDa) that is highly abundant in cells and has a vital role in transporting pathway intermediates between domains of the FAS II complex[70]. All FAS II intermediates are attached via a thioester linkage to the terminal sulfhydryl of the 4'-phosphopantetheine prosthetic arm of ACP[67,71]. Malonyl-CoA transacylase (FabD) carries out the initial transthioesterification to transfer the malonate moiety from CoA to ACP. Malonyl-ACP is subsequently used by the condensing enzymes (FabH/B/F) to extend the growing fatty acid chain by two carbon units with each turn of the cycle.

FabH is the first condensing enzyme in the pathway[72]. It is an essential protein that facilitates an enzyme-catalysed Claisen-like condensation between acetyl-CoA and malonyl-ACP[73,74]. The reaction follows a two-step mechanism whereby the acetyl unit first undergoes transacylation and is sequestered on the active site cysteine-112 residue with the release of CoA. Malonyl-ACP is then decarboxylated and stabilised by active site histidine and asparagine residues before being condensed with the cysteine-sequestered acyl chain. Each of the three residues in this catalytic triad has been shown to be essential for activity[75].

Subsequent condensing enzymes (FabB, FabF) are specific for longer-chain acyl groups and have a low affinity for acetyl-CoA as a substrate, making FabH essential for initiation of fatty acid biosynthesis. FabB and FabF differ significantly from FabH in both active site chemistry and architecture. Both use a Cys-His-His catalytic triad, differing from the Cys-His-Asn triad evolved in FabH, and both have larger binding pockets to sequester the growing acyl chain[76].

2.1.2 FabH (β-ketoacyl-ACP synthase III)

Along with being essential for the initiation of FAS II by condensing acetyl-CoA and malonyl-ACP to produce the C_4 β-ketoacyl-ACP, FabH plays an important role in mediating the fatty acid composition of the organism. Although the active site residues and peptide sequence remain highly conserved across both Gram-positive and Gram-negative bacteria[77], small structural differences in the acyl-binding pocket determine the substrate specificity of FabH[78,79,80]. Gram-positive bacteria tend to favour methyl-branched fatty acids, in which case FabH uses branched short-chain CoA primers. Conversely, Gram-negative bacteria typically favour straight-chain fatty acids and employ acetyl-CoA. *E. coli* FabH (ecFabH) shows strong selectivity for acetyl-CoA, and its active site architecture prevents cysteine-112 bound thioesters longer than 4 carbons from being accommodated[67].

Structural data of ecFabH was the first to become available[75,81,82] and subsequently data has been published with bound inhibitors[83,84], detailing the interaction of the substrate with the acyl binding-channel. Native ecFabH behaves as a homodimer with a monomeric mass of 33,515 Da.

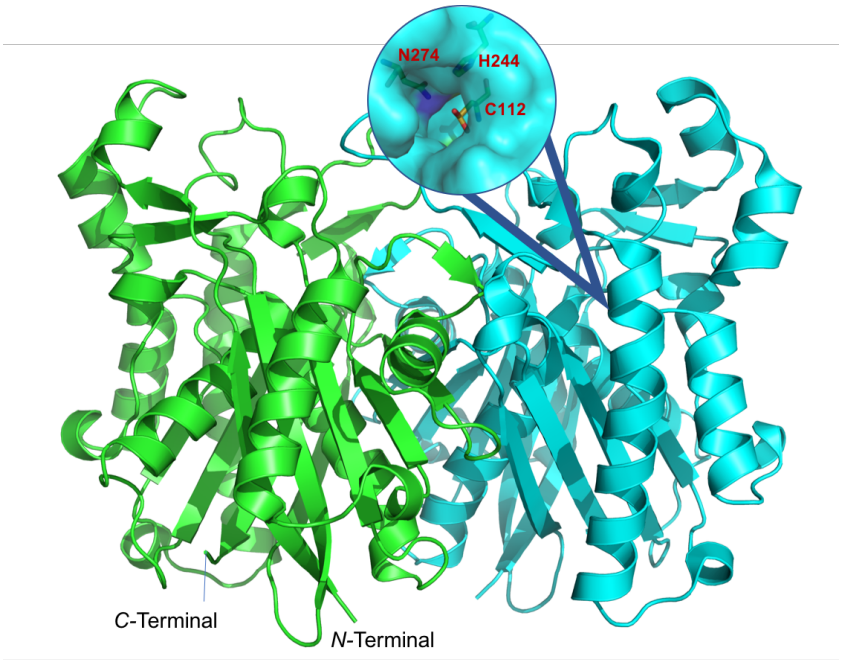


Figure 2.2 - Crystal structure of ecFabH (PDB: 3IL9) with an enlargement of the active site tunnel and the C112, H244, N247 catalytic triad. Note C112 is oxidised to cysteine sulfinic acid due to crystallisation under non-reducing conditions.

The active site of FabH is located at the end of a tunnel reached by a prosthetic 4'-phosphopantheteine (4'-PP) arm, a feature shared by both ACP and CoA. Co-crystal structures of FabH with acetyl-CoA show the substrate binding in a 'fish-hook' configuration with the adenine group partaking in π-stacking interactions on the surface of the protein, while the 4'-PP arm is anchored into the active site binding channel by hydrogen bonds augmented by Van der Waals interactions along its entire length[67,75]. Although there is no crystal data detailing the binding mode of the ACP substrate, the binding mode of the 4'-PP group is likely to closely resemble the binding revealed by the acetyl-CoA bound FabH structure[67].

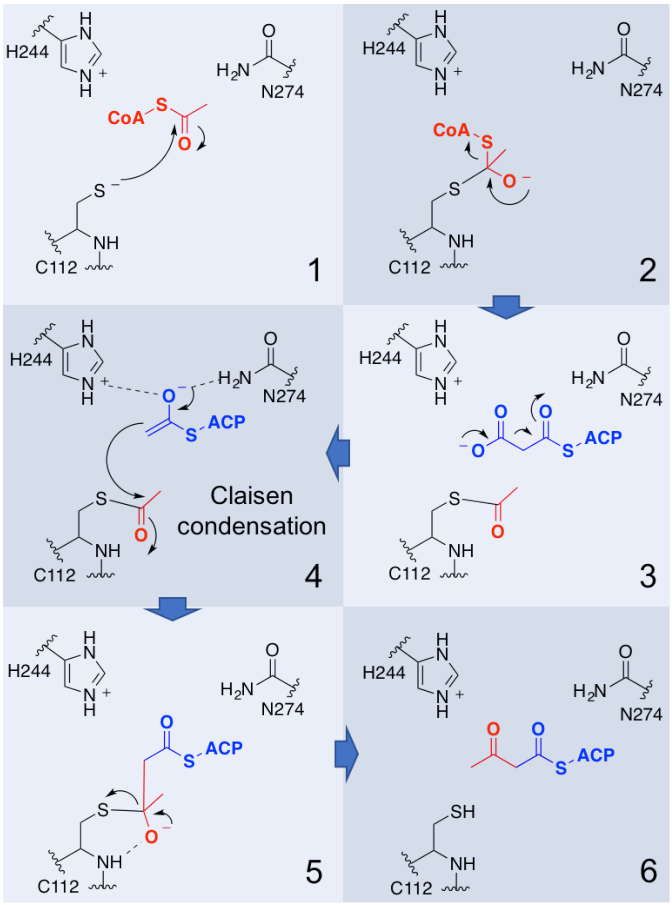


Figure 2.3 - Schematic representation of the catalytic mechanism of the two-step condensation performed by FabH. **1** - Acetyl-CoA is sequestered in the active site and attacked by nucleophilic C112. **2** - The tetrahedral intermediate is stabilised by C112. **3** - Decarboxylation of malonyl-ACP is facilitated by N274. **4** - The reactive enol is sequestered in the oxyanion hole formed between H244 and N274 in a position to attack the acetylated C112 in an enzyme-catalysed Claisen condensation. **5** - C112 is detached from the growing carbon chain. **6** - β-Ketoacyl-CoA leaves the active site.

At the base of the CoA binding channel sits the catalytic triad of Cys-His-Asn. Substitution studies have shown that each residue is fundamental for turnover and each has its own role in the two-step catalytic process. The C112 residue is responsible for sequestering acetyl-CoA by transacylation. The nucleophilicity of the C112 thiol is controlled by the macrodipole effect, a phenomenon caused by the cumulative dipole formed from the alignment of amide bonds in the coiling of an α-

helix. C112 is located at the N-terminus of a long α-helix resulting in a pK_a decrease from 8.8 to 7.2, sufficient to facilitate transacylation of the thioester[67]. When this residue is exchanged for serine (C112S), transacylation does not take place and the reaction is stopped before the first step[75]. With a similar dependence of the first step on C112, mutagenesis of H244 and N274 results in no decarboxylation of malonyl-ACP or subsequent condensation with the C112-sequestered acetyl group[75]. The orientations of H244 and N274 are fixed by a network of hydrogen bonds, resulting in the proton donating moieties of both residues pointing in the same direction forming an oxyanion hole that stabilises the developing negative charge on the thioester carbonyl to promote decarboxylation and enolate formation. The nucleophilic enol then displaces the acyl-enzyme intermediate thioester to form the C_4 product β-ketoacyl-ACP[75].

2.1.3 Inhibiting FAS II

FAS II is an attractive target for antimicrobial therapeutics as the pathway enzymes can be selectively targeted with minimal effect in humans[85,86,87]. There are many known FAS II inhibitors, with a large number targeting the enoyl-ACP reductase enzymes, acetyl-CoA carboxylase, and longer-chain condensing enzymes[88].

As FabH is the only FAS II condensing enzyme able to accept CoA-bound substrates, it is therefore essential for the initial condensation reaction of the pathway. Recently, the strategy of targeting FAS II in Gram-positive bacteria for antimicrobial therapeutics was questioned after findings that FAS II was not essential in *S. agalactiae* when the bacteria were supplied with exogenous fatty acids, as would be the case in a mammalian host[89,90]. Whether this conclusion can be extended to all Gram-positive bacteria is a subject of debate with special interest in *S. aureus*[91,92]**.** There is no evidence suggesting that FAS II is non-essential in Gram-negative bacteria, as they lack the ability to transfer exogenous fatty acid chains from CoA to ACP[89]. Regardless of the result of the dispute over Gram-positive bacteria, there remains

little doubt that targeting FAS II in Gram-negative bacteria is a scientifically lucrative oportunity[89].

2.1.4 Inhibitors of FabH

A limited number of natural products and small molecules have been reported to inhibit FabH, although most with poor efficacy. Thiolactomycin (TLM) is a natural product that mimics malonyl-ACP binding in the condensing enzymes of FAS II. TLM has been shown to be specific for FAS II and shows highest efficacy against FabB, followed by FabF and FabH with IC_{50} values of 26 µM, 60 µM and 158 µM, respectively[69].

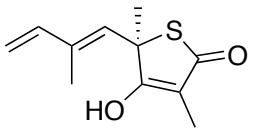


Figure 2.4 - Thiolactomycin (TLM).

The low toxicity of TLM along with the opportunity for accessible derivatisation has led to a number of thiolactomycin analogues that show improved activity against FabH[93].

Platensimycin is a potent natural product inhibitor of FabF isolated from *Streptomyces platensis* with an IC_{50} of 48 nM against *S. aureus* and 160 nM against *E. coli* FabF, however very low activity towards FabH. Structural data detailing the binding of platensimycin to FabF suggested that the acyl binding pocket of FabH would not be able to physically accommodate the inhibitor. Further screening of the same source resulted in the discovery of platencin, a structurally similar compound that targets both FabF and FabH with IC_{50} values of 4.6 µM and 9.2 µM respectively. A multiple enzyme assay was devised using FabD, FabF and FabH to determine the target of the compound by the products produced. It was determined that platencin

targets both FabF and FabH in a synergistic fashion, as IC_{50} values were lower in whole-cell assays compared to individual enzyme assays[85,86].

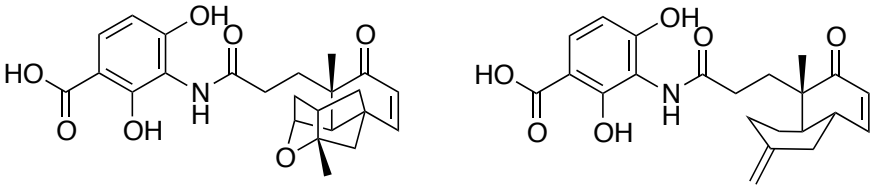


Figure 2.5 - The structures of platensimycin (left) and platencin (right).

Quorex Pharmaceuticals developed a library of small molecule inhibitors targeting FabH in Gram-positive bacteria aided by structural bioinformatics[84]. This resulted in a series of promising leads with IC_{50} values in the low nM range against *E. faecalis* and *S. pyogenes* and in the low μM range against *S. aureus* and *H. influenzae*[84]. Their co-crystal structure of FabH with the inhibitor (see fig. 2.6) showed that the A-ring was located deep in the FabH active site and the carboxyl and hydroxyl groups played important roles coordinating to histidine residues in the active site. The rest of the molecule was optimised to maximise binding interactions with residues in the acyl binding-channel, and the C-ring was vital for π-stacking interactions on the surface of the protein. Structural data was not released into the public domain.

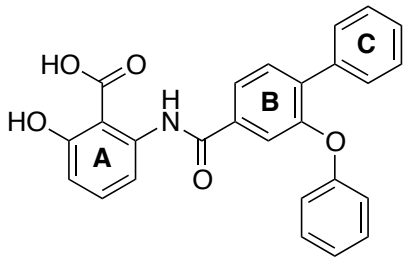


Figure 2.6 - Compound 33 from Nie *et al.*[84], IC_{50} 4 nM (*E. faecalis*), 4 nM (*S. pyogenes*), 3.8 μM (*S. aureus*), 2.4 μM (*H. influenzae*).

Zhu and co-workers proposed a number of small molecule FabH inhibitors following *in silico* screening of the crystal structure of ecFabH against a small molecule library (fig. 2.7)[77]. These were composed of two substituted aromatic rings joined by an aliphatic linker, Schiff base or hydrazone moiety[94–100]. A combination of *in vitro* and whole-cell assays showed promising inhibition data, suggesting that this class of small molecule could be further exploited. Although no structural data was obtained, modelling suggested that the compounds bound in the 4'-PP tunnel.

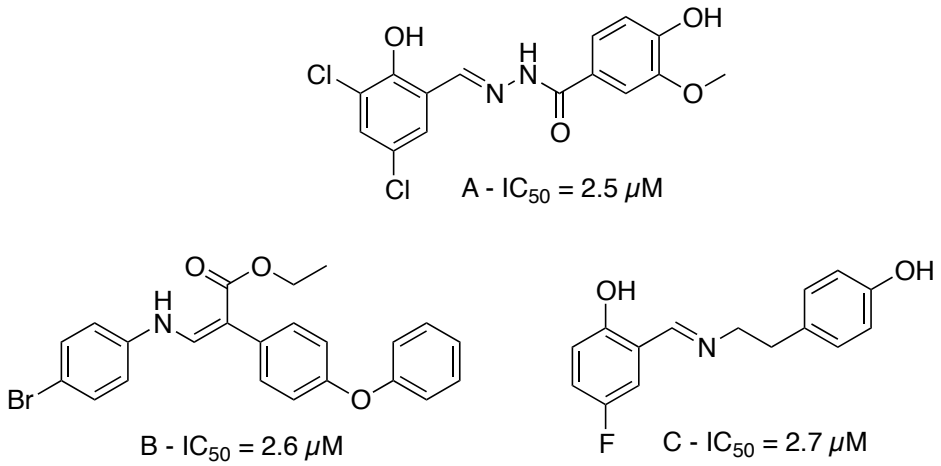


Figure 2.7 - Inhibitors proposed by Wang *et al.* (A)[94], Li *et al.* (B)[95] and Shi *et al.* (C)[96] with IC_{50} values against *E. coli*.

The vanillic *N*-acylhydrazone inhibitors proposed by Wang *et al.* stirred our interest because we believed the molecules were an ideal starting point for building a dynamic combinatorial library targeting ecFabH[94].

2.1.5 ^{19}F NMR for fragment screening

Fluorine is a highly abundant halogen and the most electronegative of the known elements. In the last decades, fluorine has become a central part of drug discovery, with approximately 25% of all marketed drugs containing at least one fluorine atom[101]. The inclusion of fluorine can help medicinal chemists to modulate the pharmacokinetic, pharmacodynamic and physical binding properties of a

compound[101]. Several examples of blockbuster fluorine-containing drugs exist, spanning many different classes of target (see fig. 2.8).

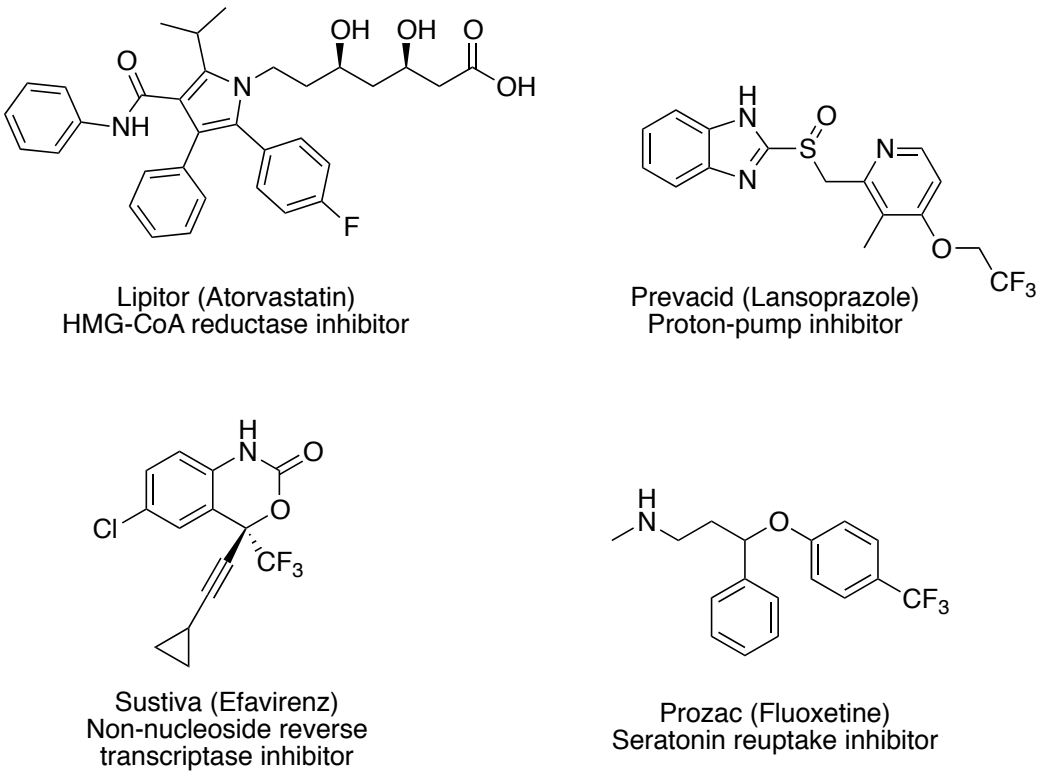


Figure 2.8 - Examples of blockbuster drugs containing a fluorine atom.

Fluorine has a spin quantum number of ½ and a high gyromagnetic ratio, making it one of the most useful nuclei in NMR spectroscopy (^{19}F NMR). This feature of fluorine, combined with its increasing prevalence in drug discovery research has resulted in the pharmaceutical industry adopting ^{19}F NMR as a robust and reliable technique for library screening and performing biochemical- and binding assays[101]. ^{19}F NMR screening offers some unique advantages. The high sensitivity and 100% natural abundance of the spin ½ NMR active isotope means that the technique can be used with relatively low ligand concentration in NMR terms. Factors that render ^{1}H NMR screening problematic such as protonated solvents or buffers do not interfere with the ^{19}F spectrum and avoid the use of expensive deuterated reagents while not compromising the stability of the protein by allowing the use of optimal buffer

solutions. The combination of rational experimental design and a broad chemical shift range minimises the possibility of spectral overlap, while the absence of fluorine in most naturally occurring macromolecules allows the use of the full spectral range. The transverse relaxation rate (R_2) of the ^{19}F nucleus is a highly sensitive observable parameter of binding events due to both the large chemical shift anisotropy of the nucleus, and the exchange between the free and unbound state of the ligand and their respective different chemical shifts[101,102].

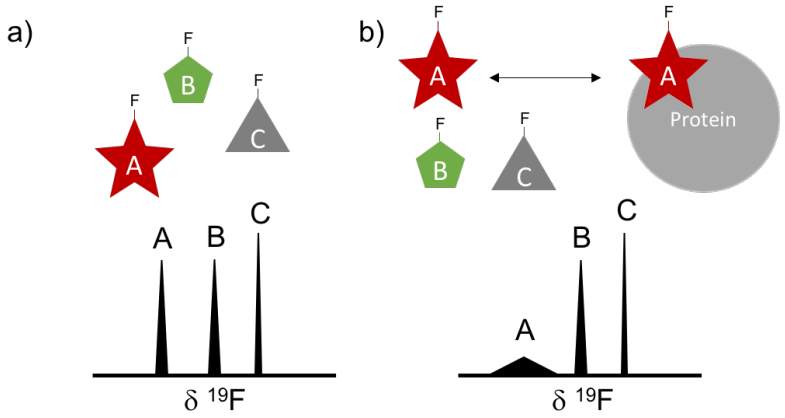


Figure 2.9 - Illustration of a ^{19}F NMR screening experiment for binders. a) Two fluorinated molecules to be screened against a protein target (A and B), and a fluorinated internal reference (C) that is known not to bind. b) The fluorinated internal reference molecule (C) does not bind to the protein and the breadth of the signal remains unchanged. Due to a combination of the chemical shift anisotropy and exchange between unbound states, the signal for a binding molecule (A) is broad, while the signal width for a non-binding molecule (B) remains unchanged.

Thanks to several academic and industrial groups, ^{19}F NMR has developed into a popular technique for screening fluorine-based fragment libraries, with the ability to screen more than 20 fragments in one experiment[101]. Due to the biological scarcity of ^{19}F, a background-free spectrum can be obtained with a low-field spectrometer in a matter of minutes, rather than the hours required on a high-field spectrometer for a saturation-transfer difference (STD) NMR experiment.

Once binders have been identified from an initial screen of fluorinated fragments, non-fluorinated fragments may be screened in competition assays against an identified, fluorinated binder. This assay is known as fluorine chemical shift and anisotropy exchange for screening (FAXS), where the known fluorinated binder and a non-binding fluorinated control are incubated with the protein, resulting in a broad signal for the binding molecule (see fig. 2.9b). When the rest of the fragment library is introduced, a stronger binder may outcompete the fluorinated binder, resulting in a narrowing of its ^{19}F signal because there is no longer an exchange process occurring between bound and unbound states[103]. The anisotropy of the fluorine atom does not contribute to line broadening in the unbound state as the molecule is rotating freely in solution.

As with all spectroscopic and chromatographic screening methods, the caveat of ^{19}F screening is the requirement for individual well-resolved signals. This can be achieved though intelligent library design aided by increasingly accurate chemical shift prediction software. More complex NMR experiments have been used to resolve complex libraries containing members with overlapping signals, both by 2D homonuclear correlation-ordered experiments (^{1}H-^{19}F COSY) and pseudo-2D diffusion-ordered (^{19}F DOSY) experiments, where the signals are separated on second axis by diffusion coefficient[101,104].

2.1.6 Aims

Having identified ^{19}F NMR as a technique that offered promising potential in protein-templated DCC, we sought to demonstrate the power of the ^{19}F NMR analysis by constructing a ^{19}F-labelled DCL against the antibiotic target protein FabH. We proposed that this would allow us to monitor the DCL equilibration process in pseudo real time, and identify privileged binding combinations from the DCL without the need for destructive analytical techniques.

2.2 Results and Discussion

2.2.1 Expression and purification of ecFabH (UniProtKB: P0A6R0)

The *FabH* gene from *E. coli* was gifted to us by the Burkhart group (UCSD, California) in an unspecified vector with chloramphenicol resistance. PCR primers were designed to introduce 5' Nde1 and 3' Xho1 restriction sites, and the gene was amplified from the construct by PCR and cloned into the linearised pGEM-T Easy vector with 3' terminal thymidine (T) overhangs. The pGEM vector allows blue/white screening of recombinant colonies as the multiple cloning site is located within the α-peptide coding region of β-galactosidase. If no recombinant gene is present, β-galactosidase is expressed and hydrolyses a chromogenic agar plate additive called X-gal, resulting in a blue colour. Colonies containing the recombinant gene are not able to hydrolyse X-gal and remain white.

The Nde1 and Xho1 restriction sites were used to clone *FabH* into the pET-28a vector containing an N-terminal hexahistidine tag. The protein was expressed in BL21 (DE3) *E. coli* cells, which were grown in LB under conditions that had previously been optimised for ecFabH, and protein overexpression was induced through the T7 promoter by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After the cells were harvested, a two-stage purification was conducted using nickel affinity chromatography followed by size exclusion chromatography (SEC). SDS-PAGE analysis showed that the protein had been overexpressed and purified without issue (see fig. 2.10).

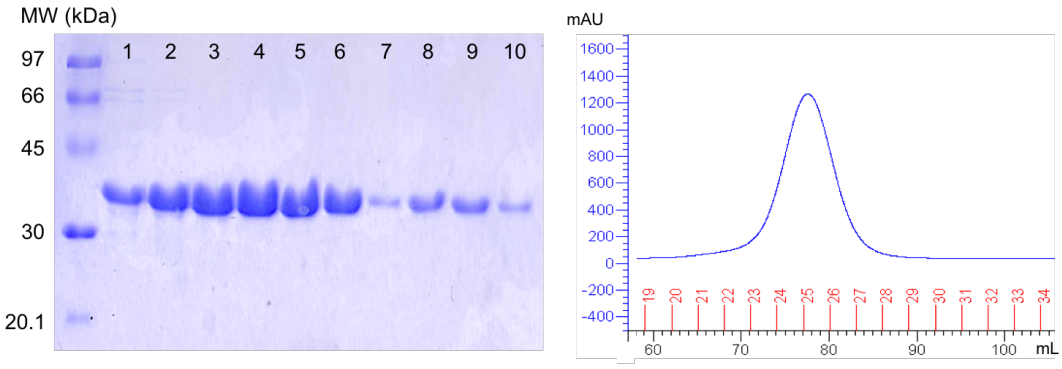


Figure 2.10 - SDS-PAGE of fractions containing ecFabH from nickel affinity chromatography (lanes 1-6) and SEC (lanes 7-10, corresponding to fractions 23-26 in chromatogram). SEC chromatogram shows ecFabH eluting as a single peak at ca. 78 mL.

The protein was characterised by ESI LC-MS under denaturing conditions, giving a mass of 35551.31 ± 0.18 Da which corresponds to the sequence-predicted mass of 35547.25 Da (ExPASy).

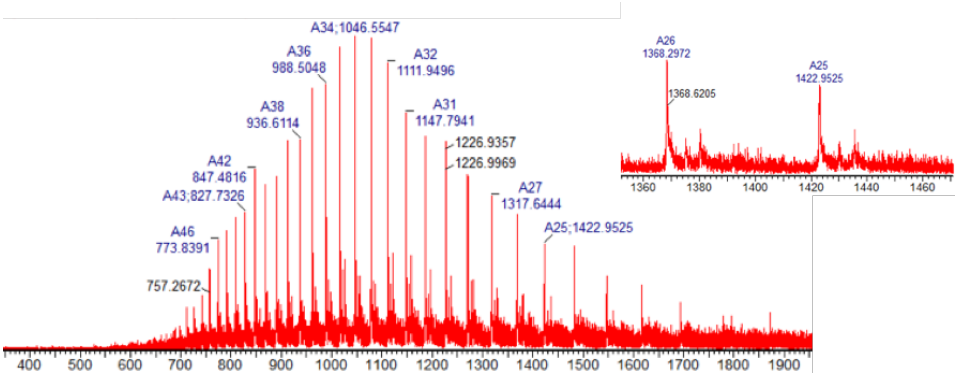


Figure 2.11 - Mass spectrum of ecFabH, with a magnification of +25 and +26 charge states.

Prior to *in vitro* assay, it was necessary to transfer the protein into reducing agent (DTT) free buffer. After 6 hours in DTT-free buffer, a new mass spectrum was obtained with a deconvoluted mass shift of +48 Da, indicating quantitative oxidation

of the catalytic C112 residue to cysteic acid (see fig. 2.12), for which there is literature precedent[105]. This suggests that the activity of ecFabH rapidly decreases outside of a reducing environment, requiring assays to be performed immediately after buffer exchange.

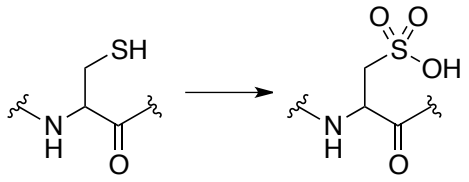


Figure 2.12 - The oxidation of cysteine to cysteic acid under non-reducing conditions

2.2.2 Nucleophilic catalysis of NAH exchange

As discussed in Chapter 1, the use of aniline as a nucleophilic catalyst for *N*-acylhydrazone (NAH) exchange was introduced by Dawson[46], based on work from Jencks on the use of aniline promote semicarbazone formation[47]. Aniline has since been successfully applied to DCC experiments[18,20].

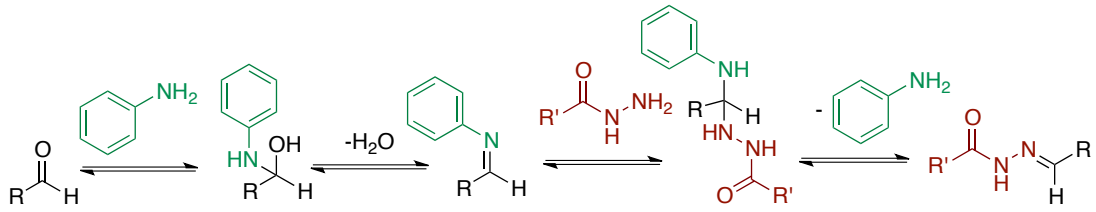


Figure 2.13 - Generally accepted mechanism of aniline-catalysed NAH formation.[106]

Serendipitously, we discovered that aniline was binding to ecFabH in a ^{1}H saturation transfer difference (STD) NMR experiment. This led us to seek out the work of Blanden *et al*. who had noticed that the use of aniline to promote biological oxime-formation ligations was detrimental to the structure of their target, α,β-tubulin[107]. They characterised 4-amino-L-phenylalanine (4-APA) as a suitable alternative catalyst

for promoting oxime exchange at physiological pH. Many other catalysts promoting oxime and hydrazone exchange have since been proposed for different applications[106,108].

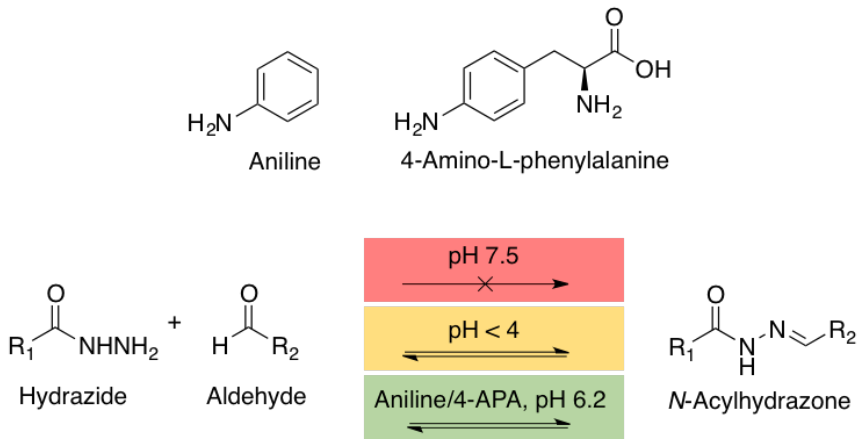


Figure 2.14 - The process of NAH exchange occurs on a reasonable timescale below pH 4, but it can also occur under physiological pH with the aid of a nucleophilic catalyst.

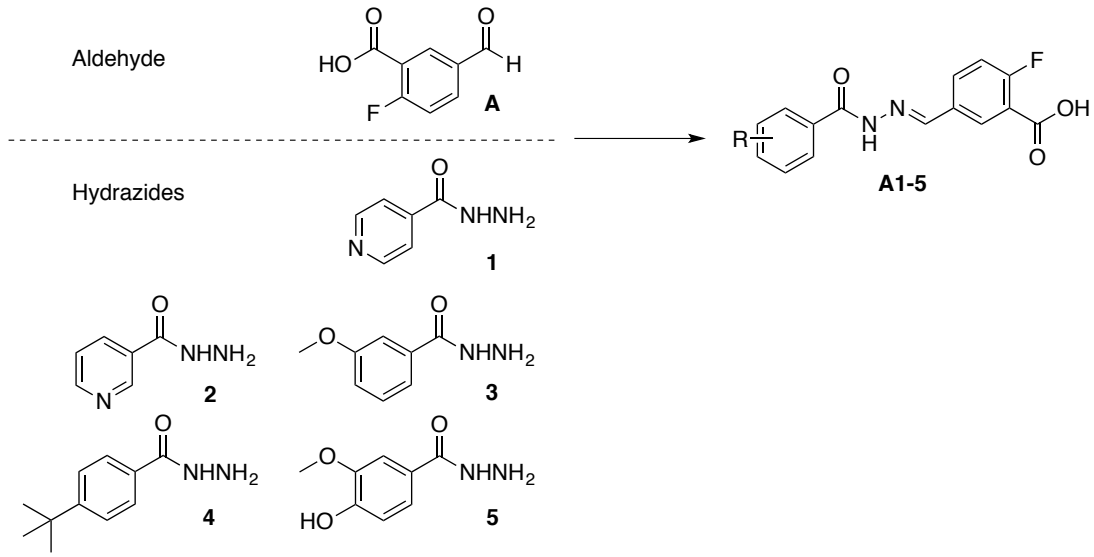


Figure 2.15 - The aldehyde and hydrazides used in this study.

We investigated the suitability of 4-APA to replace aniline as a catalyst for our DCL. Aside from its catalytic properties, 4-APA has many important benefits over aniline including high water solubility, increased biocompatibility and lower safety requirements. Whereas aniline is a known toxin, carcinogen, mutagen and corrosive compound with harmful environmental effects, 4-APA is only reported to cause irritation. We characterised 4-APA catalysis of NAH formation by conducting a simple reaction between 2-fluoro-5-formylbenzoic acid **A** and vanillic acid hydrazide **5** in the presence of different concentrations of aniline and 4-APA (see fig. 2.16). HPLC chromatograms were collected at time points over 6 h, and the integral under the peak corresponding to NAH **A5** was calculated and plotted as a function of time for each reaction (see fig. 2.17).

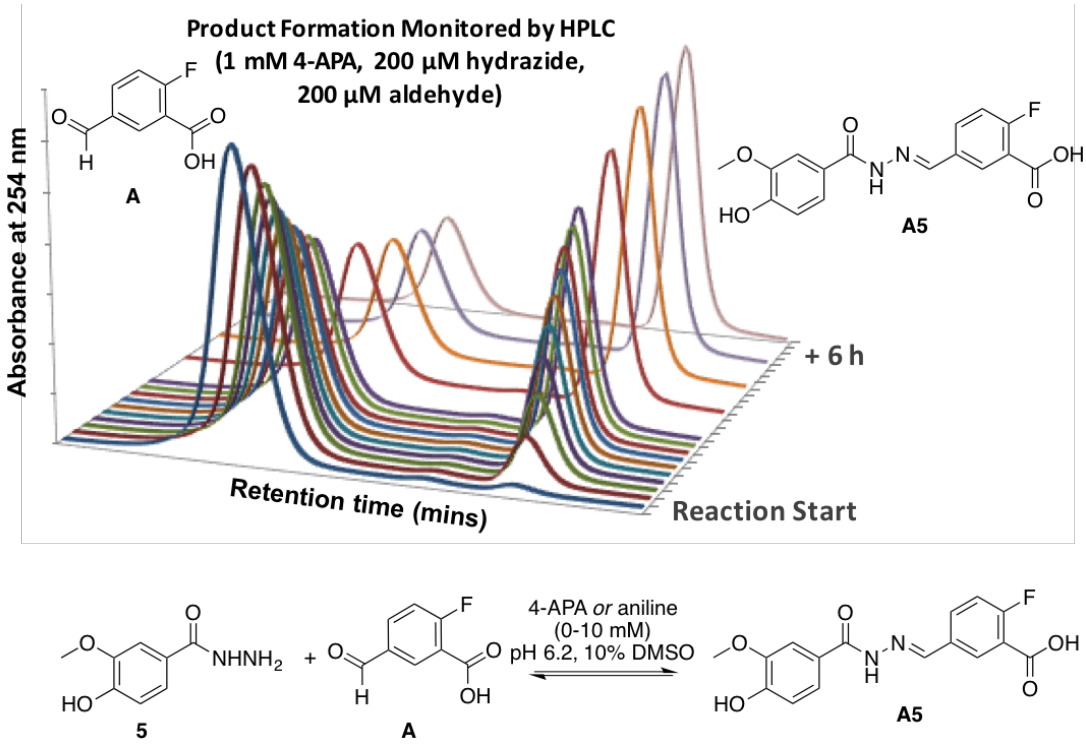


Figure 2.16 - A 3D-stacked representation of HPLC chromatograms over a period of 6 h monitoring the reaction shown above between aldehyde **A** and hydrazide **5** in the presence of 1 mM 4-APA. Retention times: aldehyde **A** 5.74 mins, NAH **A5** 6.28 mins, hydrazide **5** did not bind to the C18 column.

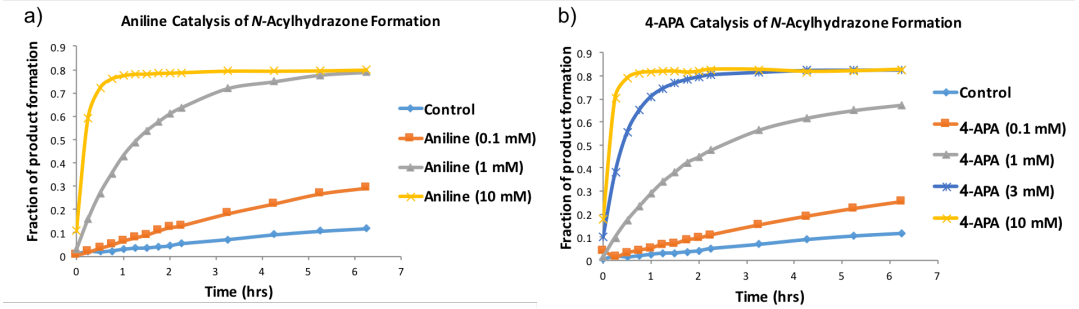


Figure 2.17 - Comparison plots derived from the chromatograms shown in fig. 2.16, repeated with differing catalytic loading. Each point on the chart represents the amount of NAH **A5** formed at a given time. Plots compare aniline and 4-APA catalysis of the forward reaction between hydrazide **5** and aldehyde **A**. Graph a) shows product formation over time for 0 mM, 0.1 mM, 1 mM and 10 mM aniline. Graph b) shows product formation over time for 0 mM, 0.1 mM, 1 mM, 3 mM and 10 mM aniline.

Graphs a) and b) of fig. 2.17 show that 4-APA exerts a comparable albeit marginally weaker rate enhancement when compared to aniline. Concentrations of 3 mM and 10 mM 4-APA accelerated the forward reaction on a satisfactory timescale, both reaching equilibrium within 2 h. The reaction does not drive all the way to completion, as illustrated by fig. 2.17, but reaches a plateau at approximately 80% product concentration.

In a DCC experiment, it is essential that the reverse reaction is on a comparable timescale to the forward reaction, otherwise a thermodynamic equilibrium will struggle to establish upon template addition. It was therefore essential that we also characterise the catalytic power of 4-APA in the reverse reaction. This was done by comparing the rates of the forward and reverse reactions in a model 3-membered library composed of aldehyde **A**, and hydrazides **3**, **4** and **5**. To monitor the reverse reaction, the library was composed of the same fundamental building blocks, however in the form of hydrazides **3** and **4**, and NAH **A5** (see fig. 2.18), such that formation of NAHs **A3** and **A4** must result from the catabolism of **A5**. Reactions were

run in parallel in the presence of either 3 mM or 10 mM 4-APA and compared to a control containing no catalyst to highlight the importance of its use. A HPLC chromatogram of each reaction was collected every 60 minutes over 9 h, and the integral under respective peaks was calculated and plotted as a function of time.

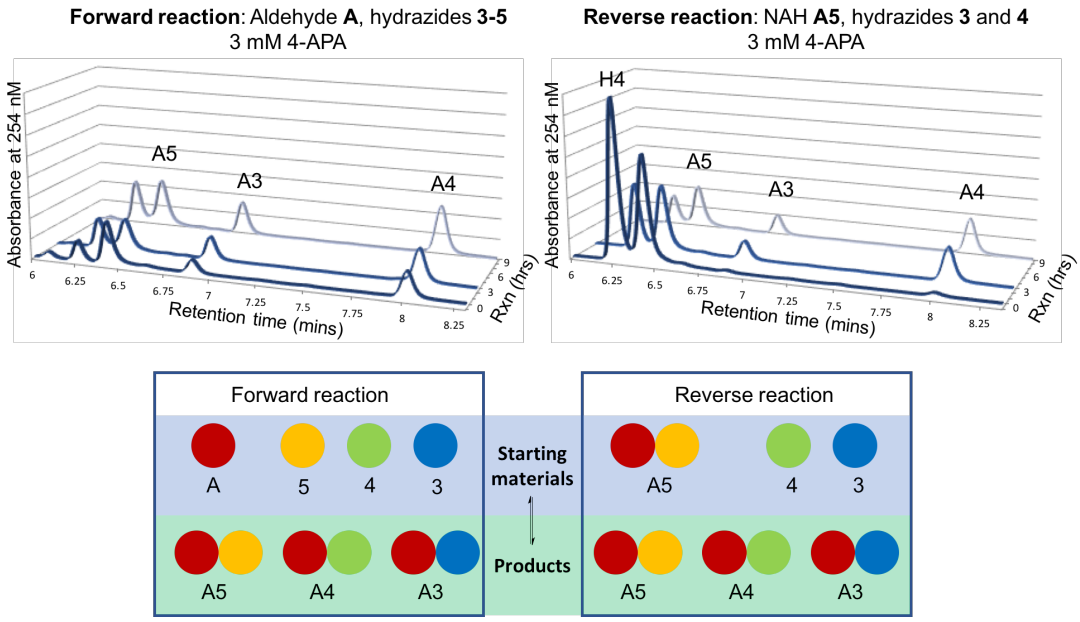


Figure 2.18 - 3D-stacked representations of HPLC chromatograms over a period of 9 h monitoring the catalysis of both the forward and reverse NAH exchange reactions shown above between aldehyde **A** and hydrazides **3**, **4** and **5** in the presence of 1 mM 4-APA. Retention times: aldehyde **A** 5.74 mins, NAH **A5** 6.28 mins, NAH **A3** 6.88 mins, NAH **A4** 7.95 mins, hydrazide **4** 6.23 mins, hydrazides **3** and **5** eluted on the solvent front.

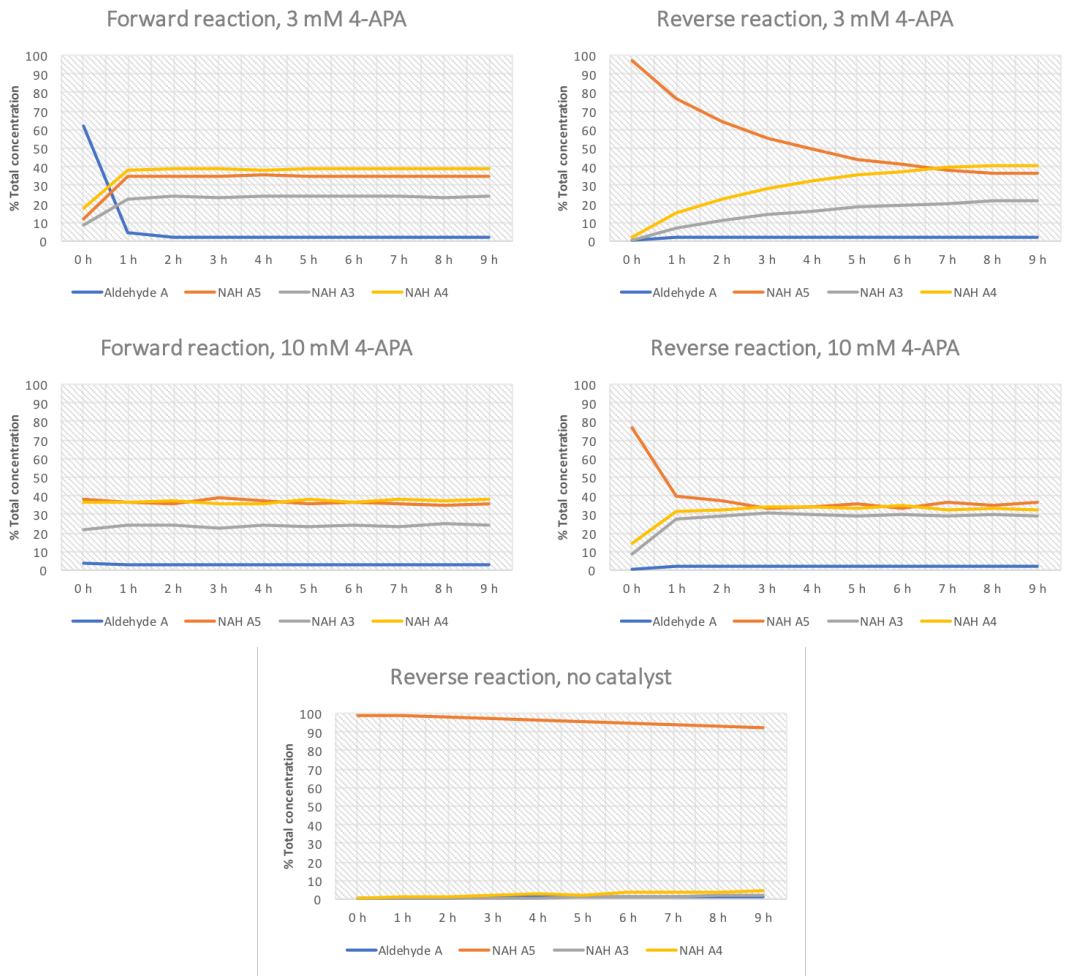


Figure 2.19 - Forward reaction composed of aldehyde **A** (200 μM) with hydrazides **3**, **4** and **5** (each 200 μM). Reverse reaction composed of NAH **A5** (200 μM) with hydrazides **3** and **4** (each 200 μM). Charts show the relative concentration of each component over time. Concentrations determined by HPLC at hourly time points.

When using 3 mM 4-APA as the catalyst, the forward reaction was equilibrated within 2 h. The reverse reaction required 8 h to attain the same equilibrium. It is interesting to note that not all NAHs appear to be of equal concentration. This may be due to both small differences in reactivity and misrepresentation due to unequal absorbance at 254 nm. With 10 mM 4-APA the forward reaction had reached equilibrium in the time that it took to inject the sample onto the HPLC column, and the reverse reaction had stabilised after approximately 2 h. While the relative

concentrations of the products in the 3 mM reactions remained relatively stable once they had reached equilibrium, there was continual fluctuation in the 10 mM experiments.

As already discussed, one caveat of a DCC experiment is that the rate of re-equilibration should not be faster than the on/off rate of the ligand binding to the protein otherwise the library would be in flux whilst members were bound, preventing the desired templating effect. For this reason, we decided that 3 mM 4-APA provided sufficient catalytic enhancement for our NAH DCL. Gratifyingly, 4-APA was not observed to bind to the protein in subsequent ^{1}H STD NMR experiments, allowing us to confirm the suitability of 4-APA as an equally potent, yet biologically benign, alternative to aniline in protein-templated NAH DCL catalysis.

2.2.3 Library selection

The structures of both platencin and the FabH inhibitors proposed by Nie *et al.* suggest that a carboxylic acid group proximal to the base of the binding pocket forms essential H-bonds with active-site residues[84]. Wang *et al.* conducted preliminary structure-activity relationship studies on their initial vanillic acylhydrazone hits, showing that hydroxyl groups on the aldehyde-derived ring were improved the potency of the hits[94]. In addition to *in vitro* enzymatic assays, their compounds underwent cell-based assays that showed that their title compound E9 had greater antimicrobial activity than kanamycin in both Gram-positive and Gram-negative bacteria, suggesting that *N*-acylhydrazones can penetrate bacterial cells (see fig. 2.20).

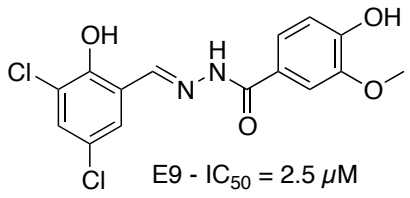


Figure 2.20 - Compound E9 from Wang *et al.*[94] with inhibition data against *E. coli*.

A DCL based on literature FabH inhibitors was designed consisting of two substituted aromatic rings connected by an *N*-acylhydrazone linker. We began with the commercially available, fluorinated, carboxyl-functionalised aromatic aldehyde 2-fluoro-5-formylbenzoic acid **A**. This core scaffold contained the important carboxyl group[84], along with a fluorine label that was used to monitor the formation of each *N*-acylhydrazone combination (see fig. 2.21).

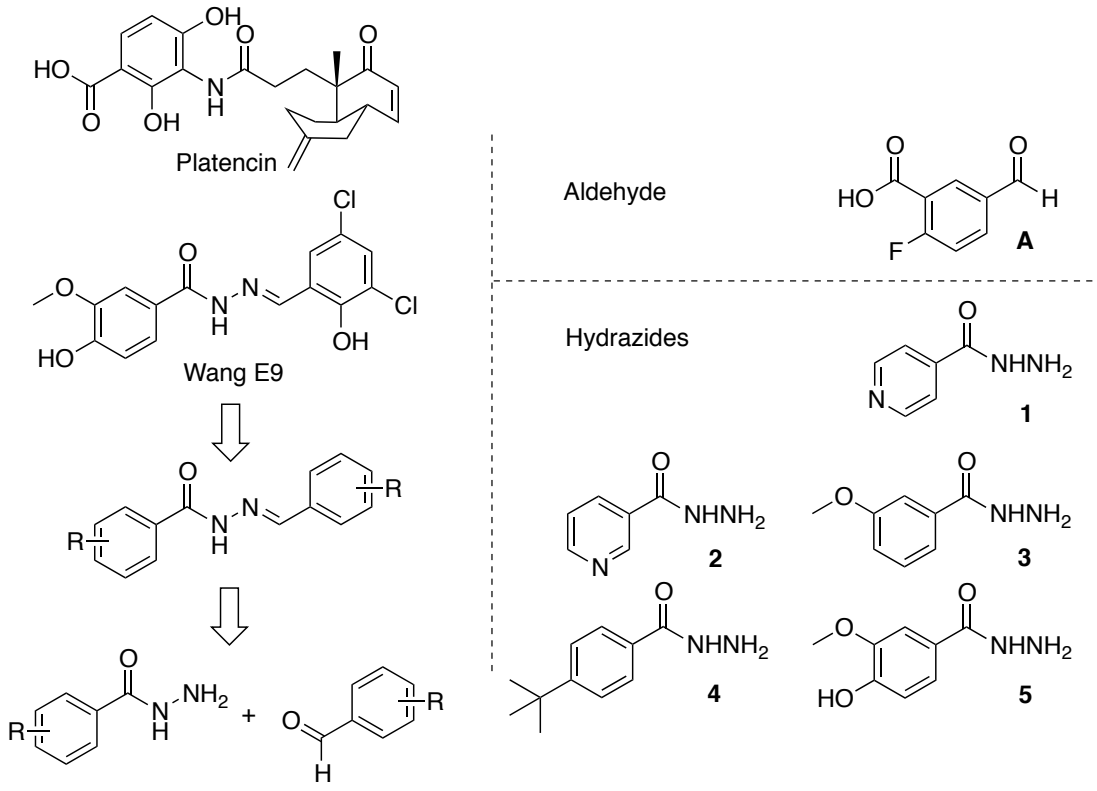


Figure 2.21 – Left: rationale of DCL, Right: assembly of ^{19}F-labelled DCL.

NMR-scale reactions were set up to determine the ^{19}F chemical shift of each of the five-membered library **1-5** in combination with fluorinated aldehyde **A**. This requirement to manually determine chemical shifts is a drawback to all NMR screening campaigns, however this may soon be superfluous with the increasing sophistication of chemical shift prediction software. Interestingly, the sensitivity of the ^{19}F nucleus betrayed small changes in the molecular structure of some *N*-

acylhydrazones up to 13 bonds away[109]. The five hydrazides **1-5** formed an NAH library of **A1-5** with individually resolved ^{19}F NMR signals (see fig. 2.22).

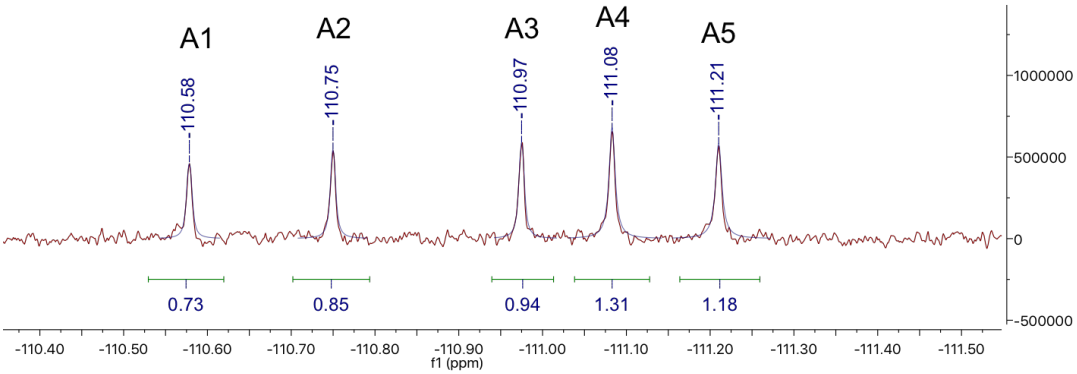


Figure 2.22 - ^{19}F NMR spectrum of un-templated five-membered (**A1-A5**) DCL at equilibrium showing well-resolved signals for each *N*-acylhydrazone library member.

2.2.4 ^{19}F NMR pulse sequence

The relaxation rate of a nucleus differs between molecules, and this can result in changes to the intensity of the ^{19}F signal. It was therefore important to conduct a series of experiments with increasing lengths of relaxation time in the pulse sequence to determine if there was any observable difference in the ^{19}F signals between library members. Experiments were conducted with a pre-equilibrated 3-member DCL assembled from hydrazides **3**, **4** and **5** with aldehyde **A** using 512 scans with relaxation times of 1, 2, 3 *or* 4 seconds. The ratio of signal integrals was compared between different experiments but it was apparent that the relaxation time had no effect on the outcome. This is perhaps not surprising as the immediate chemical environment surrounding the ^{19}F nucleus was identical in each compound. This remains, however, a necessary experiment whose importance grows as the DCL increases in diversity. A relaxation time of 1 s was used in subsequent ^{19}F NMR pulse sequences.

2.2.5 Blank DCL construction

The blank DCL was assembled with each hydrazide and aldehyde at a concentration of 200 μM. This corresponds to a stoichiometry of 1:5 of aldehyde to total hydrazide. 80% of the hydrazide would therefore be free in solution, but was invisible due to the absence of ^{19}F. This stoichiometry allows the DCL equilibrium distribution to reorder upon introduction of a template, with the possibility of exclusively amplifying a single combination at the total cost of the other library members.

The DCL was assembled under predetermined optimal physiological conditions[20], in 50 mM sodium phosphate buffer at pH 6.2, 50 mM NaCl with 3 mM 4-APA and a final concentration of 10% DMSO to promote the solubility of the NAH and 200 μM 5-fluorouracil (5-FU) as a non-competing internal reference. The library was transferred to an NMR tube and ^{19}F NMR spectra were recorded every 2 h over a 12 h period. The results agreed with the 4-APA catalyst characterisation, with the forward equilibrium being reached after approx. 2-3 h.

2.2.6 Protein-templated DCL

The protein-templated DCL was assembled in tandem with a control blank DCL, again at a final concentration of 200 μM of each hydrazide and aldehyde. The library was prepared as previously, at 1.5x concentration. To two 400 μL aliquots was added either 200 μL ecFabH (600 μM, 150 mM NaCl, 20 mM Tris pH 7.6), or 200 μL of buffer (150 mM NaCl, 20 mM Tris pH 7.6) resulting in a 1.0x cocktail. ^{19}F NMR spectra were recorded every 2 h over a 12 h period for both samples.

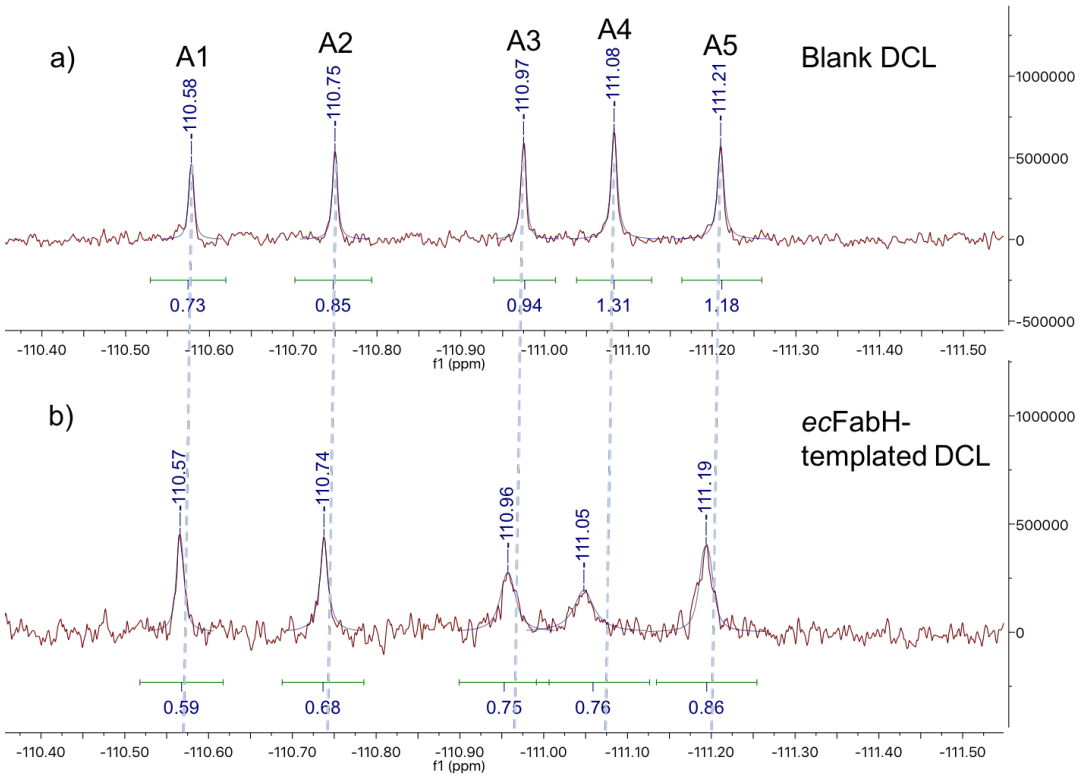


Figure 2.23 - a) shows the blank DCL after 8 h in the presence of 3 mM 4-APA. **b**) shows the ecFabH-templated DCL after 8 h in the presence of 3 mM 4-APA. Integrals and chemical shift values are relative to the 5-FU control (-169.19 ppm).

Fig. 2.23a shows the ^{19}F spectrum of the control DCL in the absence of protein template, 8 hours after mixing. It is interesting to note that not all products are of equal concentration, suggesting the electronic substituents of the ring have a pronounced effect on the thermodynamic stability of individual *N*-acylhydrazones.

Fig. 2.23b shows the ^{19}F spectrum of the library templated by the target, ecFabH. As discussed in Section 2.1.5 (see fig. 2.9), the event of a fluorinated ligand binding to a protein results in line broadening for two distinct reasons. Firstly, the exchange between bound and unbound states of the fluorinated ligand will cause the observed unbound signal to broaden and begin to shift either upfield or downfield, depending on the resonance of the bound state nucleus. Secondly, the anisotropy (uneven distribution of electrons) around the ^{19}F nucleus results in a difference in chemical

shift depending on the orientation of the molecule in the magnetic field. When the ligand is free in solution it is freely rotating, resulting in an averaging of the chemical shift signals and a relatively tight signal. In the bound state, the ligand is constrained and we therefore observe a much broader signal due to a wider distribution of chemical shifts.

In fig. 2.23b we see that the signals corresponding to **A1** and **A2** remain unchanged in appearance, despite a slight loss in integral relative to the 5-FU control. The remaining signals report the occurrence of a binding event. Most significantly the ^{19}F signal corresponding to **A4** has downfield shifted from -111.08 ppm to -111.05 ppm with an integral decrease of 42%. Both factors indicate that the ligand is present largely in its bound state resonance, however this corresponding bound state signal was not detected. This may be as it was too broad to be distinguished from the baseline. Subsequent 2-dimensional ^{19}F homonuclear correlation (COSY) NMR experiments also failed to identify the bound state resonance. Broadening is also seen in the signals corresponding to **A3** (-110.97 ppm) and **A5** (-111.21 ppm), indicating that they are also binding, but to a lesser extent.

Our DCC experiment indicates that compounds **A3**, **A4** and **A5** interact with ecFabH, and our analysis of the ^{19}F NMR spectrum suggests that the strongest interaction is with compound **A4**. Our interpretation of this data suggests that **A4** would be the most potent inhibitor of ecFabH, followed by **A3** and **A5**.

2.2.7 DTNB assay and results

To validate the results obtained from the DCC experiment, an *in vitro* coupled assay was used to characterise the inhibitory activity of compounds **A1-5**. The coupled assay utilises the free thiol produced when malonyl-ACP and acetyl-CoA are condensed by FabH. The evolved free CoA-SH displaces the disulfide of Ellman's reagent (5,5-dithio-bis-(2-nitrobenzoic acid), DTNB), creating a dual-negatively charged species that absorbs at 412 nm (ε = 14,150 L/mol/cm). The progress of the

reaction was monitored by an increase in the yellow colour of the solution in 96-well format (see fig. 2.24).

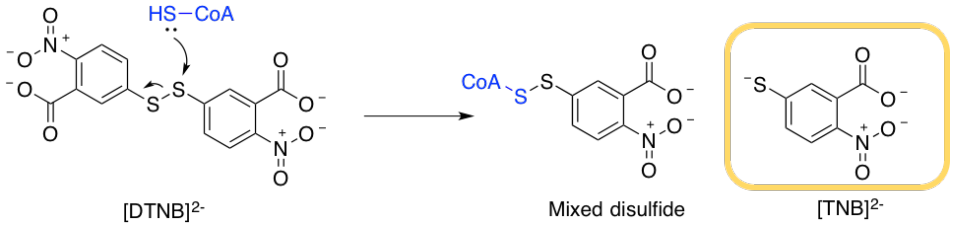


Figure 2.24 - The DTNB free thiol assay, monitoring the increase in absorbance at 412 nm resulting from the formation of the $[TNB]^{2-}$ dianion.

Fig. 2.25 shows the assay data for the 5 *N*-acylhydrazone DCL members. Pleasingly, the data corroborates our interpretation of the ^{19}F DCL experiment, with NAH **A4** showing the highest inhibitory activity, followed by NAH **A3** and NAH **A5**. Despite this agreement, the inhibitory activity of these compounds is very weak, with the assay requiring compound concentrations of 3 mM before any inhibitory activity was detected.

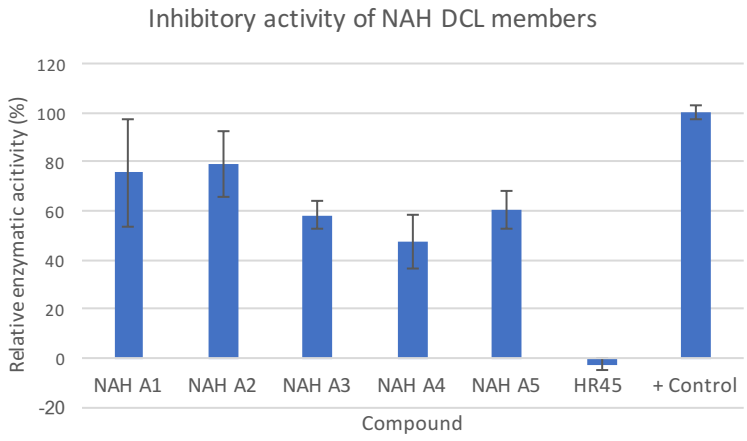


Figure 2.25 - Initial rates of ecFabH after 20 mins of incubation with NAH **A1**-**A5** (3 mM final) or HR45 (200 μM final), relative to a DMSO positive control. Error bars refer to standard error (n=3).

2.3 Conclusion

The adaptation of ^{19}F fragment screening to the field of DCC will allow a broader range of targets and library members to be explored, where limiting factors such as chromatographic and mass resolution have been problematic. Several issues with our approach remain to be tackled before this technique can become widely applicable. For example, the inability to account for the entire population of ^{19}F-containing molecules due to bound-unbound state coalescence represents both a downfall and an opportunity to develop the technique for quantitative DCL analysis. The problem arises because by modern analytical chemistry standards, NMR is incredibly insensitive. It is entirely plausible that by a combination of improved methodology and variable temperature (VT) experiments, it may be soon possible to locate the bound state signals for ligands at DCL concentrations. This will be essential, not only for comparing the ratio of bound-to-unbound, but also to qualify ^{19}F NMR as a true protein-templated DCC monitoring technique and quantify the concentration of species in the perturbed equilibrium relative to the blank control experiment.

The ideal ^{19}F NMR-monitored DCL would consist of a protein with a known ligand that binds tightly to the protein that can be used as a scaffold, containing a handle that can be exploited for a reversible DCC reaction. The library would consist of a selection of fluorinated fragments that do not bind to the protein by themselves, may or may not augment the binding of the scaffold. A quantitative NMR DCL study could then be conducted by monitoring the signal broadening of the DCL products. Any loss in integral could be compared to the ^{19}F signal from the free fragment to determine whether it is due to bound-unbound exchange, or to catabolism of that library member in favour of another combination.

In this chapter, we described the construction of a DCL targeting ecFabH that was designed to be monitored by ^{19}F NMR. The non-destructive nature of the analysis allowed us to monitor the compositions of both the blank and templated DCL equilibria without the level of uncertainty that is introduced when the experiment is

deconvoluted by destructive techniques such as HPLC and LC-MS. The ^{19}F NMR analysis of our semi-rationally constructed DCL identified a binding hierarchy among the 5 products, which we corroborated by *in vitro* enzymatic assay. To our knowledge, this is one of the first examples of using ^{19}F NMR to non-invasively deconvolute a protein-templated DCL.

2.4 Experimental

2.4.1 Microbiology

2.4.1.1 Transformation of *ecFabH*/pUC-9 derived UCSD plasmid

The *ecFabH* gene was kindly donated in a chloramphenicol resistant vector by the Burkhart group (UCSD, California). The *ecFabH*/unknown plasmid (2 µL) was added to an aliquot (25 µL) of DH5α competent cells and set on ice for 30 minutes. The cells were heat shocked at 42 °C for 30 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (34 µg/mL chloramphenicol) and incubated overnight at 37 °C. The biggest colonies were picked into separate universal tubes containing LB broth (5 mL, 34 µg/mL chloramphenicol) and cultured overnight at 37 °C with agitation. DNA was isolated using the Qiagen Spin MiniPrep kit.

2.4.1.2 PCR amplification of *ecFabH*

The *ecFabH* gene from the *ecFabH*/unknown construct was amplified by PCR, using primers to introduce 5' NdeI and 3' XhoI restriction sites.

ecFabH NdeI Fwd: 5'-CATATGTATACGAAGATTATTGGTACTGGCAGCTATCTG-3'

ecFabH XhoI Rev: 5'-CTCGAGTTAGAAACGAACCAGCGCGGA-3'

The PCR reaction contained:

5.0 µL	PFu DNA Polymerase buffer (2x)
2.5 µL	*ecFabH*/unknown construct DNA
1.5 µL	*ecFabH NdeI Fwd* primer (10 µM)
1.5 µL	*ecFabH XhoI Rev* primer (10 µM)
1.0 µL	dNTP mix (10 µM)
1.0 µL	PFu DNA Polymerase (1 U/µL)
37.5 µL	ddH_2O

The PCR reaction was conducting using the following thermal cycle: 98 °C for 2 minutes, (98 °C for 30 seconds, 60 °C for 30 seconds, 72 °C for 4 minutes) x 30, 72 °C for 10 minutes. A' overhangs were introduced to the amplified DNA strands by adding a *Pure Taq Ready to Go* bead (GE Healthcare) to the finished PCR reaction and the

mixture was incubated at 72 °C for 25 minutes. The PCR product was purified on a 1% agarose gel (TAE buffer, 0.01% gel red) and the DNA was isolated using the Qiagen QIAquick gel extraction kit.

2.4.1.3 Ligation of *ecFabH* into pGEM-T easy

The amplified *ecFabH* DNA with T' overhangs was ligated into the pGEM-T easy vector using the following protocol:

1.0 µL pGEM-T easy vector

1.0 µL T4 DNA ligase

3.0 µL *ecFabH* PCR product DNA

5.0 µL Rapid ligation buffer (2x)

The reaction proceeded at room temperature for 30 minutes.

2.4.1.4 Transformation and analytical restriction digest

The product from the ligation reaction (4 µL) was used to transform an aliquot (50 µL) of *E. coli* C2987 high-efficiency competent cells and set on ice for 25 minutes. The cells were heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on S-gal LB agar (100 µg/mL ampicillin) and incubated overnight at 37 °C. White colonies were selected over black colonies, picked into separate universal tubes containing LB broth (5 mL, 100 µg/mL ampicillin) and cultured overnight at 37 °C with agitation. DNA was isolated using the Qiagen Spin MiniPrep kit. The presence of a DNA insert was determined by analytical restriction digest at the EcoRI restriction sites in the pGEM vector using the following protocol:

1.0 µL EcoRI

1.0 µL EcoRI buffer (10x)

8.0 µL *ecFabH*/pGEM construct DNA (1/10 dilution of ligation MiniPrep)

The reaction proceeded at 37 °C for 1 hour, and was analysed on a 1% agarose gel (TAE buffer, 0.01% gel red).

2.4.1.5 Sequencing of *ecFabH*/pGEM

Plasmids giving the correct bands (approx. 1000 bp for gene and approx. 3000 bp for the empty pGEM vector) were confirmed using the Sanger dideoxy chain-terminating sequencing method. Two simultaneous PCR reactions were performed with the forward pGEM sequencing primer and reverse pGEM sequencing primer respectively:

pGEM T7 Fwd: TAATACGACTCACTATAGGG

pGEM T7 Rev: ATTTAGGTGACACTATAGAA

Sanger sequencing reactions used the following protocol:

1.0 µL Fwd or Rev pGEM T7 primer (10 µM)

1.0 µL Big Dye

2.0 µL Terminator 3.1 buffer (2x)

3.0 µL *ecFabH*/pGEM construct DNA

3.0 µL ddH_2O

The reaction was conducted using the following thermal cycle: 94 °C for 30 s, (94 °C for 30 s, 50 °C for 20 s, 60 °C for 20 s) x 24. The sample was submitted for sequencing by an ABI Prism 377 DNA sequencer.

2.4.1.6 Restriction digest of *ecFabH* and pET-28a

The *ecFabH* gene was then excised from the pGEM-T easy vector using the NdeI and XhoI restriction sites previously introduced and inserted into a pET-28a vector with an N-terminal His-tag. The restriction digests of the *ecFabH*/pGEM and empty pET-28a constructs were done according to the following protocol:

1.5 µL NdeI

1.5 µL XhoI

6.0 µL CutSmart buffer (10x)

50.0 µL *ecFabH*/pGEM construct *or* pET-28a

The reactions were incubated at 37 °C for 2 hours and loaded directly onto a 1% agarose gel (TAE buffer, 0.01% gel red). A band from each experiment corresponding to the *ecFabH* gene and the empty pET-28a vector (at approx. 1000 bp and approx.

5400 bp, respectively) were excised and purified using the Qiagen QIAquick gel extraction kit.

2.4.1.7 Ligation of *ecFabH* into pET-28a

The *ecFabH* gene was then ligated into the prepared pET-28a vector according to the following protocol:

1.0 µL	T4 DNA ligase
1.0 µL	Digested pET-28a
3.0 µL	Digested *ecFabH*
5.0 µL	Rapid ligation buffer (2x)

The reaction proceeded at room temperature for 30 minutes.

2.4.1.8 Transformation and analytical restriction digest of *ecFabH*/pET-28a

The product from the ligation reaction (4 µL) was used to transform an aliquot (50 µL) of C2987 high-competency cells and set on ice for 25 minutes. The cells were then heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (30 µg/mL kanamycin) and incubated overnight at 37 °C. Transformants were picked into LB broth (5 mL, 30 µg/mL kanamycin) and cultured overnight at 37 °C with agitation. DNA was isolated using the Qiagen Spin MiniPrep kit, and the success of the ligation was determined by analytical restriction digest according to the following protocol:

1.0 µL	NdeI
1.0 µL	XhoI
1.0 µL	CutSmart buffer (10x)
7.0 µL	*ecFabH*/pET-28a

The reaction was incubated at 37 °C for 2 hours and loaded directly onto a 1% agarose gel (TAE buffer, 0.01% gel red) for analysis.

2.4.2 Expression and purification of ecFabH

The *ecFabH*/pET-28a construct (4 µL) was transformed into an aliquot (50 µL) of BL21(DE3) cells and set on ice for 25 minutes. The cells were heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (30 µg/mL kanamycin) and incubated overnight at 37 °C. A single transformant was used to inoculate two seed cultures of sterile LB broth (2 x 250 mL, 30 µg/mL kanamycin) and agitated overnight at 37 °C. One of the overnight seed cultures was used to sub-culture sterile LB broth (5 x 500 mL, 30 µg/mL kanamycin) to an OD_{600} of 0.1. The cultures were agitated at 37 °C until the OD_{600} reached 0.6, at which point expression was induced by addition of IPTG (final conc. 0.1 mM). Cells were harvested after a further 3 hours at 30 °C and subsequently stored at -20 °C.

N-Terminal histidine-tagged ecFabH was purified at 4 °C by Ni-affinity chromatography followed by size exclusion chromatography. The BL21 (DE3) cell pellet expressing ecFabH was resuspended in lysis buffer (30 mL, 20 mM Tris-HCl pH 7.6, 300 mM NaCl, 5 mM imidazole) and lysed for 15 minutes with rounds of 30 second of sonication followed by 30 seconds of rest. The cell lysate was clarified by centrifugation (*18,000 g*, 30 minutes, 4 °C) and the cell-free extract was injected onto a HisTrap 5 mL (GE Healthcare) Ni^{2+}-affinity chromatography column pre-equilibrated in lysis buffer. The column was washed with lysis buffer (5 CV) before the histidine-tagged protein was eluted using a gradient (0-100%) of lysis buffer to elution buffer (20 mM Tris-HCl pH 7.6, 300 mM NaCl, 400 mM imidazole) over 20 CV.

Each elution fraction was analysed by SDS-PAGE, and the fractions containing His-tagged ecFabH were pooled, and concentrated to a volume less than 5 mL. ecFabH was further purified by size exclusion chromatography (HiLoad Superdex 200 16/60, GE Healthcare) with an isocratic elution of mobile phase buffer (20 mM Tris-HCl pH 7.6, 100 mM NaCl, 10% glycerol) at 1 mL/min over 120 minutes. ecFabH eluted at

approximately 70 minutes and the most concentrated fractions were pooled and flash frozen and stored at -80 °C.

2.4.3 Nucleophilic catalysis of NAH exchange

2.4.3.1 NAH formation from hydrazide 4 and aldehyde A

The following reaction was prepared to characterise the catalysis of aniline and 4-APA at different catalytic loading concentrations:

2-Fluoro-5-formylbenzoic acid **A** (200 μM), 4-tert-butylbenzohydrazide **4** (200 μM), aniline (0 mM, 0.1 mM, 1 mM, 10 mM) *or* 4-APA (0 mM, 0.1 mM, 1 mM, 3 mM, 10 mM), sodium phosphate buffer (50 mM, pH 6.2), 10% DMSO.

A HPLC chromatogram was collected every 15 minutes for the first 2 h, followed by every 60 minutes up to 6 h.

Column: Luna 5μ C18(2) 50 mm x 4.6 mm, flow rate 1 mL/min, 30 °C

Solvents: A = Water (0.1% formic acid), B = Acetonitrile (0.1% formic acid)

Method: 5% B for 1 minute, 5-95% B over 7 minutes, 95% B for 1 minute, re-equilibrate at 5% B for 5 minutes.

2.4.3.2 NAH formation from 3-membered forward library

The following reaction was prepared to characterise the catalysis of the forward reaction with catalyst 4-APA at both 3 mM and 10 mM:

2-Fluoro-5-formylbenzoic acid **A** (200 μM), 4-tert-butylbenzhydrazide **4** (200 μM), 3-methoxybenzhydrazide **3** (200 μM), 4-hydroxy-3-methoxybenzhydrazide **5** (200 μM), 4-APA (3 mM *or* 10 mM), sodium phosphate buffer (50 mM, pH 6.2), 10% DMSO.

A HPLC chromatogram was collected every 60 minutes up to 9 h. The integral under each peak was calculated using LabSolutions (Shimadzu) and plotted as a percentage of total concentration for each time point.

Column: Luna 5μ C18(2) 50 mm x 4.6 mm, flow rate 1 mL/min, 30 °C

Solvents: A = Water (0.1% formic acid), B = Acetonitrile (0.1% formic acid)

Method: 5% B for 1 minute, 5-95% B over 7 minutes, 95% B for 1 minute, re-equilibrate at 5% B for 5 minutes.

2.4.3.3 NAH formation from 3-membered reverse library

The following reaction was prepared to characterise the catalysis of the forward reaction 4-APA at 3 mM and 10 mM:

NAH **A5** (200 µM), 4-tert-butylbenzhydrazide **4** (200 µM), 3-methoxybenzhydrazide **3** (200 µM), 4-APA (3 mM *or* 10 mM), sodium phosphate buffer (50 mM, pH 6.2), 10% DMSO.

A HPLC chromatogram was collected every 60 minutes up to 9 h. The integral under each peak was calculated using LabSolutions (Shimadzu) and plotted as a percentage of total concentration for each time point.

Column: Luna 5µ C18(2) 50 mm x 4.6 mm, flow rate 1 mL/min, 30 °C

Solvents: A = Water (0.1% formic acid), B = Acetonitrile (0.1% formic acid)

Method: 5% B for 1 minute, 5-95% B over 7 minutes, 95% B for 1 minute, re-equilibrate at 5% B for 5 minutes.

2.4.4 DCL conditions

The library was assembled with each of the hydrazides **1-5** and aldehyde **A** at a final concentration of 200 µM, buffered at pH 6.2 using 50 mM sodium phosphate with 50% D_2O, 10% DMSO, 50 mM NaCl and 3 mM 4-APA. The library cocktail also contained 200 µM 5-fluorouracil as a non-competing internal reference and either 200 µM ecFabH or an equivalent volume of enzyme purification buffer. Each library was transferred to an NMR tube and a ^{19}F spectrum was acquired every two hours over a twelve-hour period.

2.4.5 Length of relaxation time in ^{19}F NMR pulse sequence

A three-member library was assembled from NAHs **3**, **4** and **5** (200 µM each) in sodium phosphate buffer (50 mM, D_2O, pH 6.2), 5-fluorouracil (internal standard, 200 µM) and a total of 10% DMSO. A series of ^{19}F NMR experiments were conducted with 512 scans and relaxation times of 1,2,3 *or* 4 seconds. The integral of the signals corresponding each NAH relative to the internal standard were compared at different

relaxation times to determine the required relaxation time for the DCL ^{19}F NMR experiment.

2.4.6 FabH enzymatic DTNB assay

An assay cocktail was prepared from DTNB (1 mL, 4 mM in 50 mM NH_4OAc), HEPES buffer (1 mL, 1 M, pH 8.0) and ddH_2O (8 mL). The assay was run in 96-well plate format with the following protocol (the following bracketed figures refer to final concentrations):

100 µL DTNB assay cocktail

25 µL ecFabH (0.63 µM)

25 µL Acetyl-CoA (375 µM)

25 µL Malonyl-CoA (375 µM)

5 µL ddH2O

20 µL NAH **A1-A5** (3 mM) *or* HR45 (200 µM) in DMSO *or* DMSO control

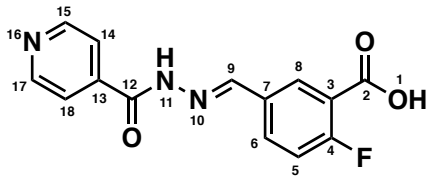
Following a 20 min incubation at 37 °C, the reaction was initiated by the addition of malonyl-CoA and followed by monitoring absorbance change at 412 nm over 20 mins at 37 °C. The DMSO concentration was maintained at 10% to aid the solubility of the NAH compounds.

2.4.7 Synthetic Chemistry

General procedure for the synthesis of *N*-acylhydrazones A1-5:

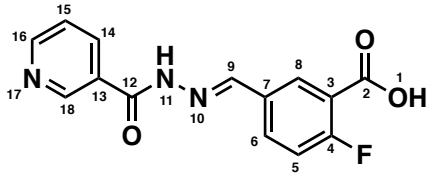
2-Fluoro-5-formylbenzoic acid **A** (50 mg, 0.30 mmol, 1.0 eq.) and the hydrazide **1-5** (0.33 mmol, 1.1 eq.) were dissolved in ethanol (2 mL). A few drops of glacial acetic acid were added and the mixture was stirred overnight at room temperature. The solid formed was collected by vacuum filtration and washed in diethyl ether (5 mL) and water (5 mL) to yield the target **NAH** as a white solid.

A1 - *2-fluoro-5-[(1E/Z)-[(pyridin-4-yl-formamido)imino]methyl]benzoic acid*



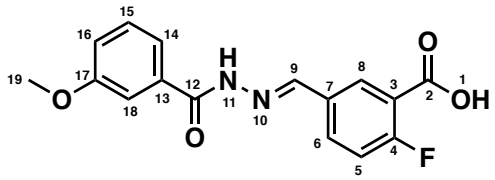
(54 mg, 0.188 mmol, 63%): **m.p.** 263-264°C; **IR** (nujol, cm^{-1}) 3497, 3402, 3196, 3073, 3028, 1657, 1612, 1564; **^{1}H NMR** (500 MHz, DMSO) δ 12.17 (1H, s, OH-1), 8.83–8.78 (2H, m, CH-15,17), 8.50 (1H, s, CH-9), 8.27 (1H, m, CH-5), 8.02–7.99 (1H, m, CH-8), 7.86–7.81 (2H, m, CH-14,18), 7.44 (1H, dd, *J* = 10.6, 8.6 Hz, CH-6); **^{13}C NMR** (126 MHz, DMSO) δ 165.1, 162.3, 150.8, 147.5, 140.8, 133.8, 131.1, 130.9, 112.0, 118.4, 118.2; **^{19}F NMR** (470 MHz, DMSO) δ -110.58; **HRMS** m/z (ESI^+) calcd. $C_{14}H_{10}FN_3O_3$ $[M+Na]^+$ requires 310.0598, found 310.0599.

A2 - *2-fluoro-5-[(1E/Z)-[(pyridin-3-ylformamido)imino]methyl]benzoic acid*



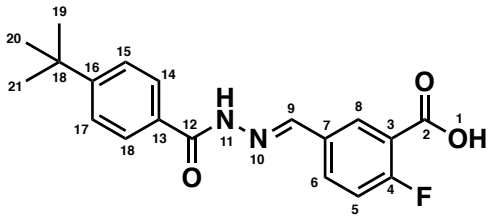
(72 mg, 0.251 mmol, 84%): **m.p.** 265-266°C; **IR** (nujol, cm^{-1}) 3196, 3065, 2446, 2426, 1647, 1618, 1601, 1553; **^{1}H NMR** (500 MHz, DMSO) δ 12.12 (1H, s, OH-1), 9.10–9.05 (1H, m, CH-18), 8.78 (1H, d, *J* = 3.7 Hz, CH-16), 8.48 (1H, s, CH-9), 8.28–8.26 (2H, m, CH-5,14), 7.59 (1H, dd, *J* = 7.8, 4.9 Hz, CH-15), 7.44 (1H, dd, *J* = 10.5, 8.6 Hz, CH-6); **^{13}C NMR** (126 MHz, DMSO) δ 165.6, 165.1, 152.9, 149.1, 146.9, 136.0, 133.7, 131.2, 130.8, 129.6, 124.1, 120.5, 118.4, 118.4, 118.2, 31.2; **^{19}F NMR** (470 MHz, DMSO) δ -110.75; **HRMS** m/z (ESI^+) calcd. $C_{14}H_{10}FN_3O_3$ $[M+H]^+$ requires 288.0779, found 288.0786.

A3 - *2-fluoro-5-[(1E/Z)-{[(3-methoxyphenyl)formamido]imino}methyl]benzoic acid*



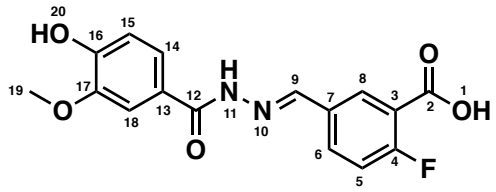
(82 mg, 0.259 mmol, 87%): **m.p.** 223-224°C; **IR** (nujol, cm^{-1}) 3401, 3225, 2995, 1701, 1659, 1618, 1589, 1545; **^{1}H NMR** (500 MHz, DMSO) δ 11.91 (1H, s, OH-1), 8.49 (1H, s, CH-9), 8.27–8.21 (1H, m, CH-5), 7.98 (1H, s, CH-8), 7.51–7.38 (4H, m, CH-14,15,16,18), 7.21–7.15 (1H, m, 6), 3.85 (3H, s, CH_3-19); **^{13}C NMR** (126 MHz, DMSO) δ 165.1, 163.47, 161.3, 159.7, 146.3, 135.2, 133.6, 131.4, 130.7, 130.2, 120.5, 120.3, 118.3, 118.1, 118.1, 113.4, 55.9, 31.2; **^{19}F NMR** (470 MHz, DMSO) δ -110.97; **HRMS** m/z (ESI^+) calcd. $C_{16}H_{13}FN_2O_4$ $[M+Na]^+$ requires 339.0752, found 339.0755.

A4 - 5-[(1E/Z)-{[(4-*tert*-butylphenyl)formamido]imino}methyl]-2-fluorobenzoic acid



(33 mg, 0.096 mmol, 32%): **m.p.** 230-232°C; **IR** (nujol, cm^{-1}) 3441, 3289, 2953, 2619, 2596, 1701, 1612, 1557, 1503; **^{1}H NMR** (500 MHz, DMSO) δ 11.88 (1H, s, OH-1), 8.48 (1H, s, CH-9), 8.24 (1H, d, *J* = 6.9 Hz, CH-5), 7.99–7.96 (1H, m, CH-8), 7.86 (2H, d, *J* = 8.1 Hz, CH-14,18), 7.56 (2H, d, *J* = 8.3 Hz, CH-15,17), 7.42 (1H, m, CH-6), 1.33 (9H, s, CH_3-19,20,21); **^{13}C NMR** (126 MHz, DMSO) δ 172.4, 165.1, 164.0, 146.0, 144.2, 130.6, 128.0, 125.8, 120.4, 118.1, 35.2, 31.4, 31.2, 21.5; **^{19}F NMR** (470 MHz, DMSO) δ -111.08; **HRMS** m/z (ESI^+) calcd. $C_{19}H_{19}FN_2O_3$ $[M+H]^+$ requires 343.1453, found 343.1455.

A5 - *2-fluoro-5-[(1E/Z)-{[(4-hydroxy-3-methoxyphenyl)formamido]imino}methyl] benzoic acid*



(72 mg, 0.226 mmol, 73%): **m.p.** 260-263°C; **IR** (nujol, cm^{-1}) 3292, 3092, 1695, 1651, 1618, 1589, 1557, 1520; **^{1}H NMR** (500 MHz, DMSO) δ 11.71 (1H, s, OH-1), 9.71 (1H, s, OH-20), 8.47 (1H, s, CH-5), 8.23 (1H, dd, *J* = 7.3, 2.3 Hz, CH-8), 7.95 (1H, d, *J* = 6.7 Hz, CH-6), 7.48–7.44 (1H, m, CH-18), 7.44–7.38 (1H, m, CH-14), 6.89 (1H, d, *J* = 8.2 Hz, CH-15), 3.86 (3H, s, CH_3-19); **^{13}C NMR** (126 MHz, DMSO) δ 165.1, 163.2, 161.1, 150.7, 147.8, 133.4, 133.3, 131.6, 130.5, 124.5, 121.9, 120.4, 120.3, 118.3, 118.1, 115.5, 112.2, 56.3, 19.0; **^{19}F NMR** (470 MHz, DMSO) δ -111.21; **HRMS** m/z (ESI^+) calcd. $C_{16}H_{13}FN_2O_5$ $[M+Na]^+$ requires 355.0701, found 355.0705.

Chapter 3: Mechanistic Studies on a Promising DCC Core Scaffold

Abstract

There is a great need for new chemical agents to combat the spread of antimicrobial resistance. Fundamental to the success of developing potent lead molecules is understanding the inhibition mechanisms of active compounds, such that their potential may be exploited in a rational manner. In this chapter, we describe an investigation into the inhibitory mechanism of 4,5-dichloro-1,2-dithiole-3-one (HR45), a halogenated 1,2-dithiole-3-one commonly used as a positive control for β-ketoacyl-ACP synthase III (FabH) inhibition. We believed that the structure of HR45 showed potential as a core scaffold for a DCL targeting FabH, but first it was necessary to understand the inhibitory mechanism of the molecule. We used mass spectrometry, NMR and molecular modelling to show the established antimicrobial inhibitor HR45 acts by forming a covalent adduct with the target protein. The 5-chloro substituent directs attack from the nucleophilic thiol side chain of the essential active site cysteine-112 residue via a Michael-type addition elimination mechanism. This work was done with the assistance of Dr Van Kelly in mass spectrometry, Dr Scott Cockroft in density functional theory (DFT) calculations, and Dr Jon Marles-Wright in structural modelling.

3.1 Introduction

3.1.1 Covalent Drugs

Drugs that form covalent bonds with their targets make up a class of molecules known as covalent inhibitors. Covalent inhibitors are traditionally composed of a recognition element and a reactive, often electrophilic warhead that forms a covalent bond with nucleophilic residues on the target protein. The class can be divided into two subsets, depending on whether the covalent inhibition is reversible on a biological timescale. Reversible covalent inhibitors typically bind to their target and subsequently dissociate from it at a rate faster than that at which the enzyme turns over[110]. These inhibitors contain electrophilic warheads at the softer end of the spectrum, like aldehydes, activated ketones or boronic acid derivatives. Irreversible

covalent inhibitors typically include a hard electrophile warhead like an epoxide, α-haloketone, or most commonly Michael-type acceptors based on the acrylamide moiety[110]. These irreversible covalent inhibitors do not dissociate from the target. These two classes may be further divided, depending on whether the drug is administered in its active form, or the administered drug is metabolised *in vivo* to unveil the reactive warhead.

Covalent inhibitors have attracted some scepticism from the scientific community since they elicit their efficacy through covalent modification, and justifiably the pharmaceutical industry has avoided developing covalent inhibitors due to largely unpredictable and potentially devastating off-target effects and idiosyncratic toxicity. This can occur through insufficiently specific compounds or the formation of chemically active metabolites, leading either to direct tissue damage or eliciting an immune response through haptenisation of proteins[111].

Despite the drawbacks, the top-10 list of highest grossing drugs in the United States contains three blockbuster drugs that have a covalent mode of action against their targets[111]. However, a common feature among the 39 covalent inhibitors that have received approval from the US FDA, is that their precise mode of action was discovered years after their therapeutic value was demonstrated. For this reason, the overarching statistics concerning covalent inhibitors must be interpreted with caution, as their mode of action is serendipitous rather than by design.

Aspirin is the most widely used medication in the world, first marketed over 100 years ago[112]. It was only in the late 1970s that it was discovered that the primary mode of action was the selective acetylation of a serine residue proximal to the active site of cyclooxygenase 1 and 2 (COX-1 and COX-2)[113]. Another example is the β-lactam containing class of antibiotics, which were found, years after their discovery by Fleming[55], to acetylate bacterial D-ala-D-ala-transpeptidase (also known as the

penicillin binding protein), preventing an essential cross-link from forming in the peptidoglycan layer of the bacterial cell wall.

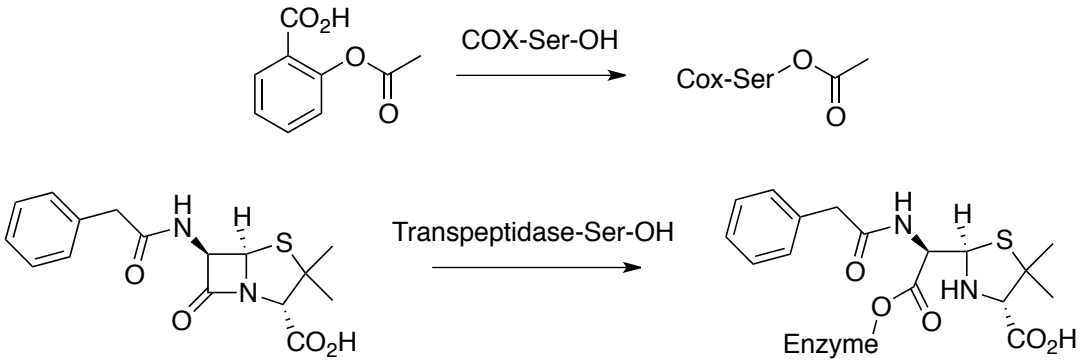


Figure 3.1 - Two examples of popular drugs which were found to have covalent modes of action.

Since the late 1990s, developments in bioinformatics have led to the rational design of covalent inhibitors gaining ground in drug discovery, with a new class of drug known as targeted covalent inhibitors (TCIs)[110]. The idea behind TCIs is to develop a highly selective molecule that can bind to a target and place a weakly reactive warhead in sufficient proximity to a non-conserved nucleophilic residue on the protein, which then spontaneously forms a covalent adduct. This approach has been heralded as driving the resurgence of covalent drugs, with several TCIs either in late-stage clinical trials or approved for use.

Advances in analytical techniques have allowed the scientific community to identify the structures of reactive drug-derived metabolites, and retro-analysis of these species has armed medicinal chemists with lists of functional groups (toxicophores) that should be avoided to minimise attrition due to off-target effects[114].

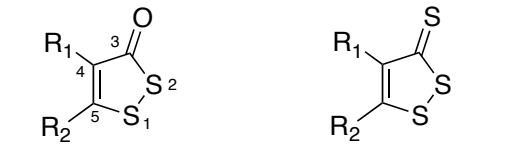
There are several benefits that successful covalent inhibitors could offer, including high potency due to complete blockade of the target[115], and as a result low dosing.

Unlike in non-covalent inhibition, the duration of action would be dependent on the turnover rate of the protein, rather than the pharmacokinetics of the drug[110].

The promising future of covalent drugs is still dawning, and their success depends not only on the continued marriage of bioinformatics and structure-based drug design, but also development of the analytical methods by which we can characterise the covalent modification of targets.

3.1.2 1,2-Dithiole-3-ones

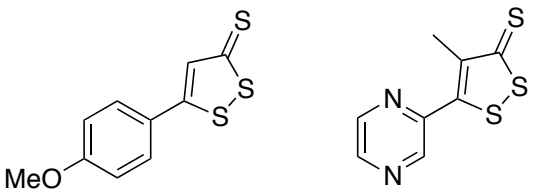
1,2-Dithiole-3-ones are a synthetic class of compounds centred around a 5-membered asymmetric ring containing a disulfide adjacent to a carbonyl. Analogous to the better studied 1,2-dithiole-3-thiones, a class of compound that occurs naturally in cruciferous vegetables such as cabbage and Brussels sprouts[116], dithiole-3-thiones also have clinical precedent.



1,2-Dithiole-3-one 1,2-Dithiole-3-thione

Figure 3.2 - Generic structures of 1,2-dithiole-3-ones and 1,2-dithiole-3-thiones.

Anethole dithiolethione (ADT) has been used for decades as a choleretic and a sialogogue, and is marketed in France as Sulfarlem for the treatment of dyspepsia[116]. ADT, along with other members of the class has been shown to increase cellular glutathione levels to protect against oxidative stress[117].



Anethole dithiolone (ADT) Oltipraz

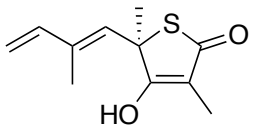
Figure 3.3 - The structures of ADT and Oltipraz, examples of 1,2-dithiole-3-thione drugs.

Oltipraz is another dithiole-3-thiolone that has an interesting biological profile. Originally identified and developed as a drug to treat schistosomiasis (snail fever), a disease caused by a parasitic tapeworm endemic to tropical developing countries, Oltipraz has since been shown to protect against carcinogenesis in several organs, including breast, colon, pancreas, lung, stomach, skin, and the bladder[117,118]. Oltipraz has also been linked to preventing aflatoxin-induced liver cancer, a carcinoma induced by aflatoxin B1, a toxin released by some strains of *Aspergillus*[118]. Despite phase I and II clinical trials against various cancer targets, campaigns were abandoned due to gastrointestinal- and neuro-toxicity. However, at the time of writing phase III trials of Oltipraz as a fatty acid synthesis inhibitor against non-alcoholic fatty acid liver disease are currently underway in South Korea.

3.1.3 4,5-Dichloro-1,2-dithiole-3-one

Of the reported natural product inhibitors of FabH, the thiolactone antibiotic thiolactomycin (TLM, fig. 3.4) has selectivity for bacterial FAS II condensing enzymes by mimicking malonyl-ACP binding[119,120]. In an effort to develop FabH inhibitors by exploiting the known binding mode of TLM, He *et al*. conducted an *in silico* screen of the National Cancer Institute (NCI) database for structurally similar molecules[121–123]. Their initial modelling highlighted three important structural components: the 5-membered thiolactone ring was important for mimicking malonyl-ACP, the carbonyl of the thiolactone formed hydrogen bonds with the histidine-244 and asparagine-274

of ecFabH, and the alkene functionality at the distal end of the isoprenoid tail incurred a hydrophobic stabilisation in the binding channel[119,123]. These structural features were integrated into their NCI database search resulting in 72 hits, of which 39 were assayed against ecFabH. At screening concentrations between 75-100 μM, 9 of the compounds demonstrated inhibition of over 50% of the enzyme activity. Of these 9 compounds, 4 were 1,2-dithiole-3-ones.



Thiolactomycin (TLM)

Figure 3.4 - The structure of the thiolactone antibiotic thiolactomycin upon which He *et al.* based their screen of the NCI database[123].

A similarity search on these four 1,2-dithiole-3-ones yielded a further nine compounds. In total, six of these compounds were tested. The most potent hit was 4,5-dichloro-1,2-dithiol-3-one (also referred to in the literature as HR45 and DDCP, fig. 3.5). Subsequent structure-activity relationship studies showed that the chlorine in the 5-position (C5) was found to be essential for irreversible inhibition of FabH isoforms from both saFabH and ecFabH with HR45 effecting reported IC_{50} values of 156 nM and 2.0 μM, respectively[123]. Analogues in which the C5 chlorine was replaced the isosteric and isoelectronic trifluoromethyl group showed a marked reduction in activity, leading the authors to believe that the C5 chlorine played a larger role than modulating the electron density of the ring[123]. Additionally, when the enzyme was inhibited using the C5 trifluoromethyl analogues it was possible to regain significant activity after multiple washing steps. This was not the case when the enzyme was inhibited by analogues with a C5 chlorine substituent. The authors postulated that the presence of a chlorine at C5 ensured a slowly releasing complex[123].

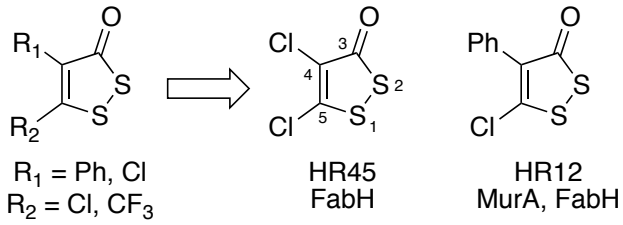


Figure 3.5 - The structure of some of the 1,2-dithiole-3-ones investigated by He *et al.*[123], specifically HR45 (4,5-dichloro-1,2-dithiole-3-one).

A less efficacious hit from the same study was HR12 (also known as RWJ-3981, fig. 3.5), an analogue of HR45 with a phenyl group replacing the chlorine in the 4-position (C4), with reported IC_{50} values of 980 nM and 5.7 µM for saFabH and ecFabH respectively[121,123]. Interestingly, in a separate study HR12 was identified as a hit against *E. coli* uridine diphosphate-*N*-acetyl glucosamine enolpyruvyl transferase (ecMurA), a cysteine-dependent enzyme involved in the first committed step of peptidoglycan cell wall biosynthesis that shares no structural homology with FabH apart from an active-site cysteine residue[124]. The study in question also determined that the C5 chlorine was essential for potent activity.

Both the studies on ecMurA, and ec- and saFabH were unable to provide empirical data to demonstrate the mode of binding of these potent halogenated 1,2-dithiole-3-ones. However, both suggested that the essentiality of the C5 chlorine at the distal end of an α,β-unsaturated ketone functionality could point towards a covalent modification model via a Michael-type mechanism[123,124]. The authors of the ecMurA study proposed that HR12 formed a covalent adduct with cysteine-115, but were unable to provide evidence from mass spectrometry (MS) analysis to back up the fact. A key point to note from the ecMurA study is that although significant activity could not be regained after washing the inhibited enzyme, the presence of dithiothreitol (DTT) prevented inhibition, either before or after a pre-incubation period of the enzyme with HR12[124].

The FabH study put their active compounds through a cell based assay, with HR45 having the lowest MIC against *E. coli* but also showing activity against *S. aureus,* with an order of magnitude greater activity than TLM against both organisms, but a lower activity than vancomycin and oxacillin[123]. As noted by both studies, it is unlikely that the antibacterial activity derived from these compounds is specific against either MurA or FabH. However, the cell-based assays show that the compounds are able to enter the cell and exert significant bactericidal activity[123,124]. It is also worth considering that the hydrolysis pathway of dithiole-3-ones is a complex one, and it is likely therefore that reactive metabolites of these molecules react in several different ways with enzyme targets[125].

The commercially available 4,5-dichloro-1,2-dithiol-3-one has since been frequently reported as a positive control for FabH inhibition[94,98,126–129], and also has FDA approval as a slimicide additive in the paper industry for food packaging (FDA docket number 99F-1423).

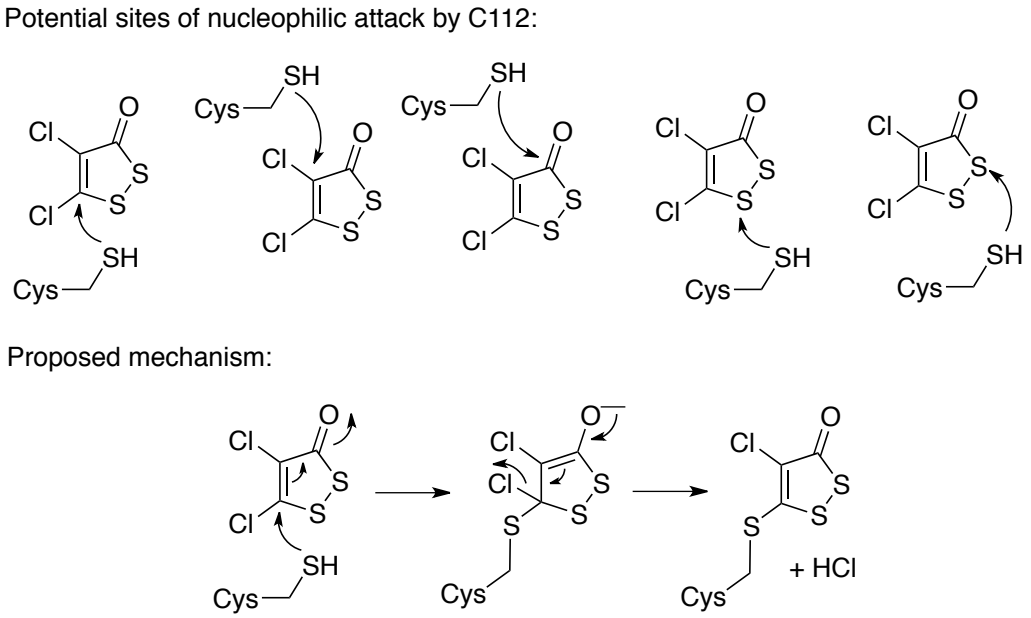


Figure 3.6 - Potential sites of nucleophilic attack by C112 and a proposed mechanism of Michael-type addition-elimination at the C5 position of the dithiole-3-one with the loss of HCl.

Upon initial examination, the Michael-type acceptor nature of the cyclic α,β-unsaturated ketone of HR45 suggested to us that it may act as a pan assay interference compound (PAIN)[130] with non-specific activity towards nucleophilic amino acid residues. Whilst PAINs are often removed from traditional small molecule screens, it has been suggested that this should not be the case when screening for antimicrobial activity[56]. This is because the small molecule screens that have been assembled by academia and the pharmaceutical industry are largely based on Lipinski's rules[60], which we now know should not govern antibiotic research due to the different stipulations on molecules entering eukaryotic and prokaryotic cells[56].

3.1.4 Aims

In order to evaluate the suitability of HR45 as a core scaffold for a DCL, it was essential to elucidate the mode of action of the compound on FabH. The aim of the investigation was to discern the previously elusive mode of inhibition of HR45 on FabH, and in doing so attempt to capture the adduct by MS and explain why previous groups had failed to do so.

3.2 Results and discussion

3.2.1 Model reaction

At the start of our attempt to capture the FabH/HR45 adduct, the first step was to develop a simplistic model reaction that would be simple to monitor and characterise. Working on the hypothesis that HR45 did indeed react with the active site cysteine of FabH, a stoichiometric amount of *N*-acetyl cysteine was incubated with HR45 in phosphate buffer at pH 7.5 and the reaction was followed by ^{1}H NMR. This model reaction had gone to completion by the time the reaction had been transferred to an NMR tube and obtain a ^{1}H spectrum. This was seen by a shift in the two resonances corresponding to the two protons of the pro-chiral CH_2 group of *N*-acetylcysteine, and a shift in the resonance of the single proton on the α-carbon of the protected amino acid. Subsequent ^{13}C and 2D HMBC NMR experiments were done to determine the regioselectivity of the reaction. Cross-peaks between C-5 and CH_2-6 (3.53, 163.7) and CH_2-6' (3.72, 163.7), and absence of cross-peaks between C-4 and CH_2-6 and CH_2-6' confirm that the addition-elimination reaction has taken place at the C5 position of HR45 (see fig. 3.7).

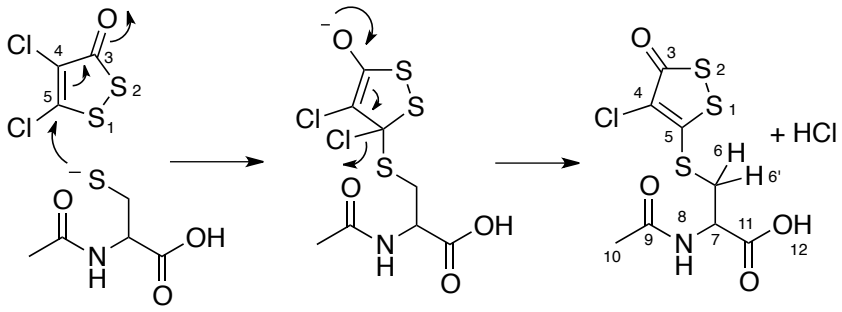


Figure 3.7 - Proposed mechanism and numbering of model reaction of HR45 with *N*-acetylcysteine.

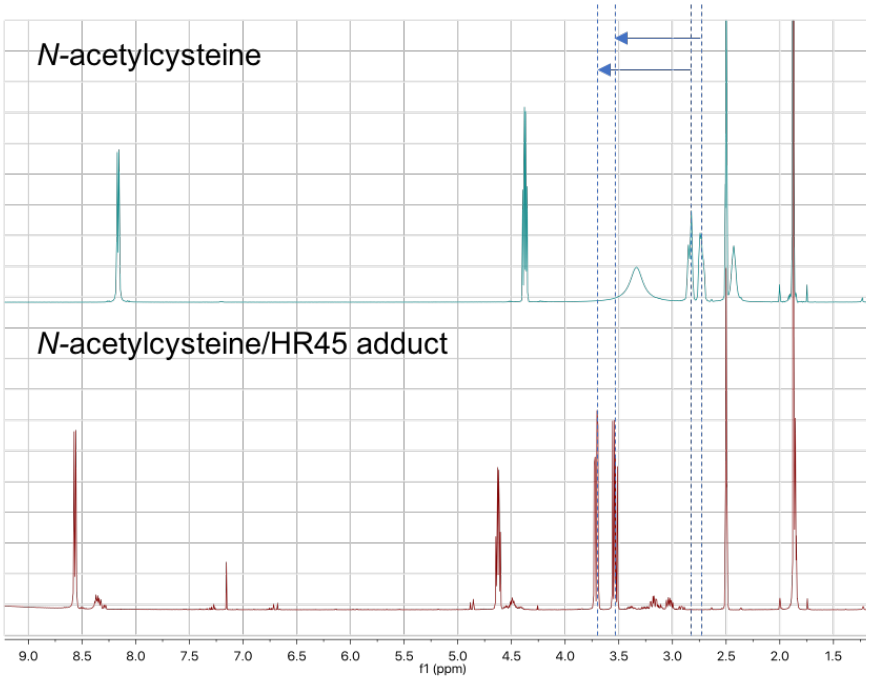


Figure 3.8 - Stacked ^{1}H NMR spectra of *N*-acetylcysteine (above) and the HR45/*N*-acetylcysteine adduct. The dotted lines and arrows indicate the downfield shift of the prochiral CH_2 protons (CH_2-6 and CH_2-6') in the HR45/*N*-acetylcysteine adduct.

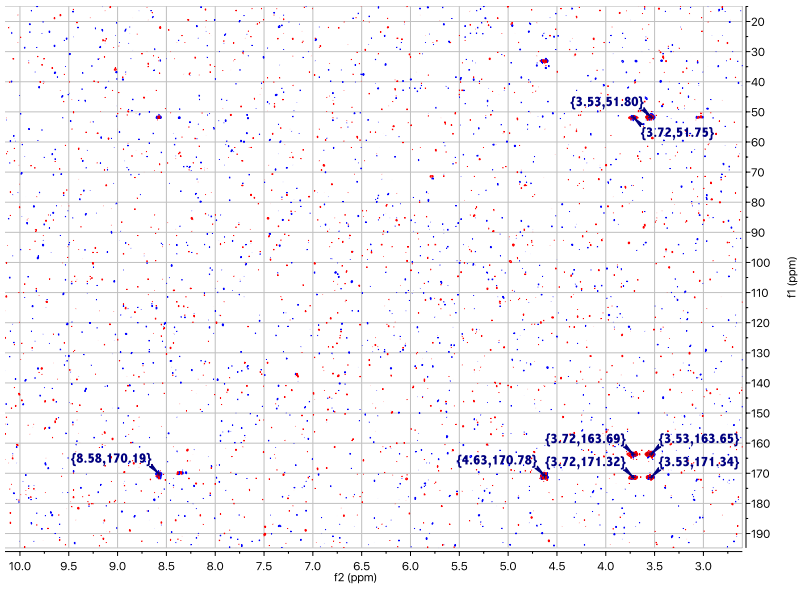


Figure 3.9 - $^{1}H^{13}C$ HMBC NMR spectrum showing cross-peaks between C-5 and CH_2-6 (3.53, 163.7) and CH_2-6' (3.72, 163.7), and absence of cross-peaks between C-4 and

CH_2-6 and CH_2-6' confirm that the addition-elimination reaction has taken place at the C-5 position.

To rationalise this data, Dr Scott Cockroft at the University of Edinburgh conducted DFT calculations to determine the plausibility of a Michael-type attach at C5 of HR45. '*The structure of HR45 was minimised using B3LYP/6-311G*. The LUMO and LUMO+1 orbitals were degenerate in energy. The LUMO+1 is shown below and is consistent with a substrate that would undergo Michael addition at C5. According to frontier molecular orbital theory, the shape of this orbital should approximate the localisation of the additional pair of electrons following nucleophilic attack of the S atom of a cysteine residue on HR45. The largest orbital coefficient (lobe) associated with a single atom is located over C5, consistent with the site that would favour formation of a new σ-bond (i.e. between C5 of HR45 and the S atom of a cysteine residue). Furthermore, this orbital (and therefore, the pair of electrons accepted during the addition reaction) is further delocalised over C4 and C3 (with double-bond character) and onto the carbonyl oxygen, consistent with the enolate intermediate formed during the proposed addition reaction*' (see fig. 3.10).

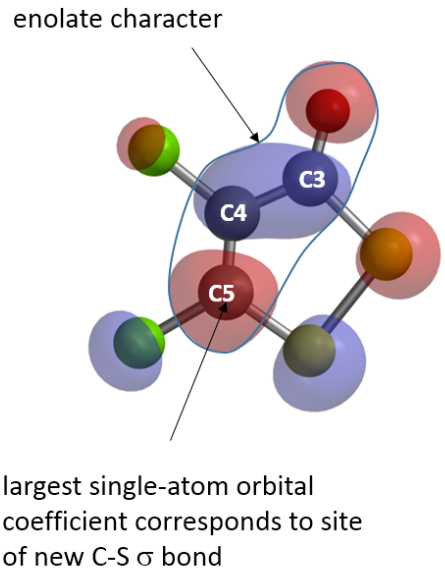


Figure 3.10 - Localisation of LUMO+1 of HR45 calculated using B3LYP/6-311G* indicating Michael acceptor character. Courtesy of Dr Scott Cockroft.

Encouraged by the corroboration of our model reaction with the DFT calculations, we proceeded with the hypothesis that HR45 asserted its inhibition on FabH by covalently modifying the active site cysteine-112. We elected to work with FabH from *S. aureus* (saFabH), an isoform of the enzyme containing only a single cysteine residue that formed an essential part of the catalytic triad. Recombinant saFabH had been easily expressed and characterised in *E. coli* by several different groups, suggesting that the enzyme was easy to handle. He *et al.* had also reported that HR45 had the greatest *in vitro* activity on saFabH out of the isoforms of the enzyme that they tested[123].

3.2.2 Expression and purification of saFabH

The *saFabH* gene amino acid sequence was obtained from Uniprot (P9WNG3) and purchased as a codon-optimised gene from Genscript with the introduction of 5'-Nco1 and 3'-BamH1 restriction sites. The synthetic gene was excised from the pUC-57 vector in which it was delivered and cloned into pET-28a with an N-terminal hexahistidine tag. The protein was expressed in BL21 (DE3) *E. coli* cells, which were grown in LB under conditions that had previously been optimised for ecFabH, and protein overexpression was induced through the T7 promoter by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After the cells were harvested, a two-stage purification was conducted using nickel affinity chromatography followed by size exclusion chromatography (SEC). SDS-PAGE showed that the protein had been overexpressed and purified without issue (see fig. 3.11).

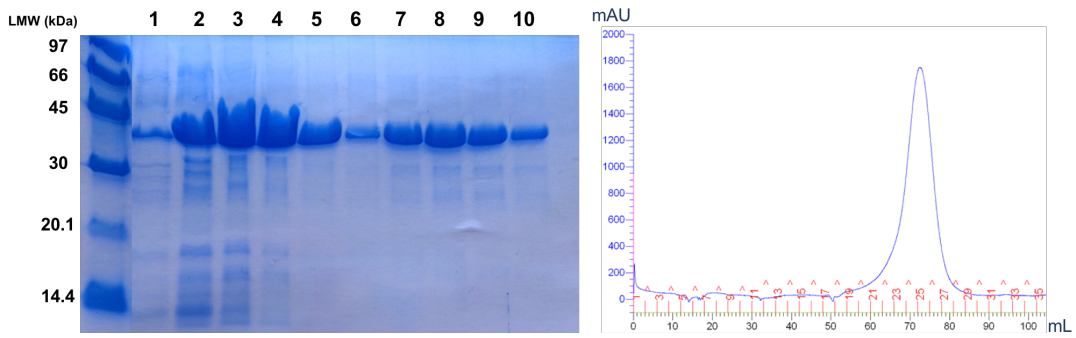


Figure 3.11 - SDS-PAGE of fractions containing saFabH from nickel affinity chromatography (lanes 1-5) and SEC (lanes 6-10), and the SEC chromatogram showing saFabH eluting at 70 mL as a single peak from the Superdex S200 SEC column.

The protein was characterised by LC-MS under denaturing conditions, giving a mass of 35911.00 ± 0.69 Da which corresponds to the predicted mass of 35911.58 Da. On closer inspection of the data, it was apparent that a second major form of the protein was present with a mass difference of +178 Da. This corresponded to gluconoylation of the N-terminal histidine tag, a common post-translational modification of recombinant proteins expressed in *E. coli*[131]. There were also further peaks in the mass spectrum that were not identified (see fig. 3.12). The presence of multiple species in the mass spectrum not only weakens the intensity of the signal as each charge state is divided in two, but it also adds complicating factors when trying to analyse the spectrum for a covalent adduct.

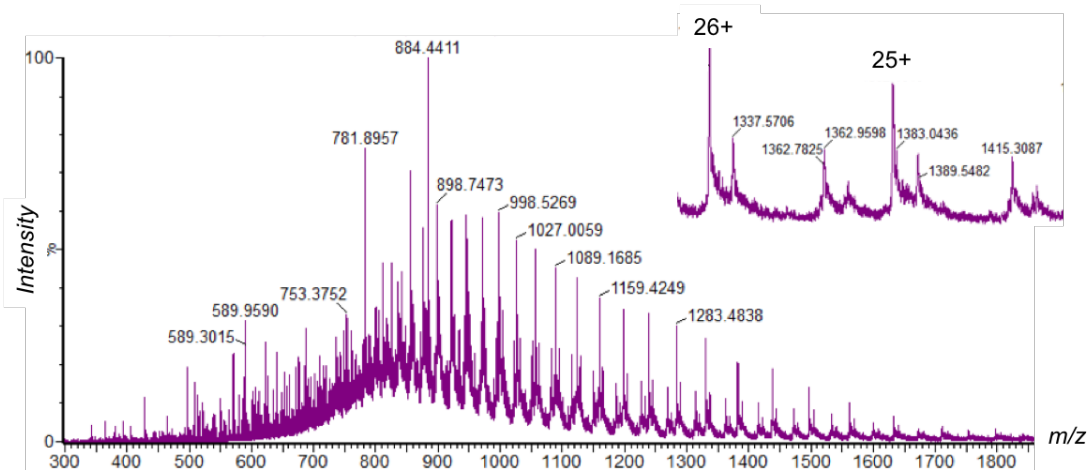


Figure 3.12 - Mass spectrum of saFabH with magnification of 26+ and 25+ charge states showing a second peak corresponding to gluconoylation of the affinity tag, and two further unidentified species in each charge state.

We therefore decided that before we began trying to capture the proposed covalent adduct we must begin with an uncomplicated native protein mass spectrum. To do this the *FabH* gene was amplified from the pET-28a vector by PCR using primers designed to insert 5'-Nco1 and 3'-Xho1 restriction sites, and cloned into the pET-HISTEV vector. The pET-HISTEV vector is a second-generation plasmid based on pET-28a that was gifted by Prof Jim Naismith at the University of St. Andrews, which has a tobacco etch virus nuclear-inclusion-a endopeptidase (TEV protease) recognition site introduced between the gene and the N-terminal hexahistidine tag. This allows the protein to be expressed and purified as previously, however after the nickel affinity chromatography purification step the protein can be incubated with TEV protease to cleave the hexahistidine tag prior to SEC.

The yield from the expression and purification of saFabH from the *FabH*/pET-HISTEV vector was significantly lower than that of the previous purification. Presumably this was due to incomplete cleavage of the affinity tag by TEV protease, in our hands a relatively difficult protein to handle, leaving a portion of saFabH bound to the nickel affinity column with the histidine tag still attached. However, MS characterisation of

the purified protein provided well-resolved data with a single peak for each charge state and a deconvoluted mass of 34063.73 ± 0.51 Da, matching the predicted mass of 34064.63 Da.

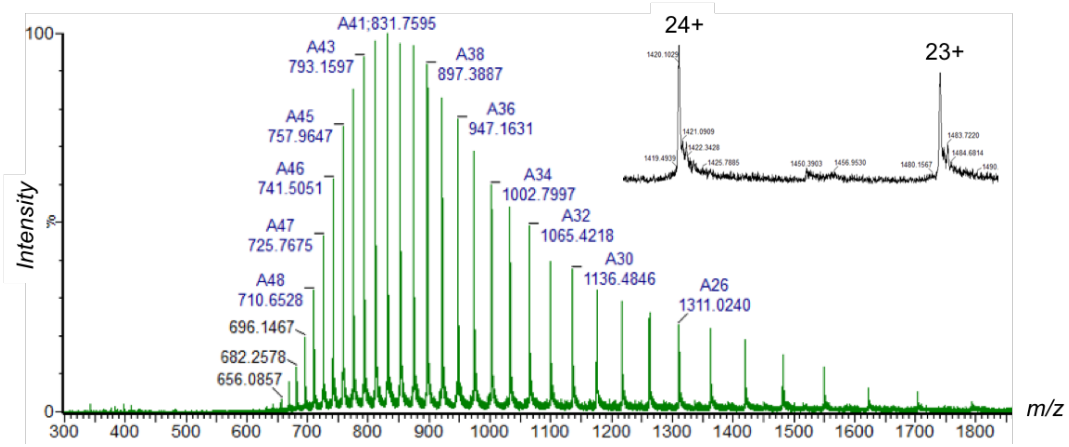


Figure 3.13 – Mass spectrum of tagless saFabH with magnification of the 24+ and 23+ charge states, showing only a single form of the protein.

3.2.3 HR45 modification

With a clean mass spectrum of the native protein in hand, we began to investigate the existence of a covalent reaction between saFabH and HR45. We began by incubating a 12 μM sample of saFabH that had been exchanged into fresh buffer containing no reducing agent with HR45 in approximately 20-fold excess for 30 minutes at room temperature. This resulted in a quantitative shift of the deconvoluted mass of +150 Da, a mass shift corresponding to the addition of HR45 with the loss of a HCl. This data gratifyingly supported our working hypothesis of a Michael-type addition-elimination reaction involving attack of the nucleophilic cysteine residue at C5 of HR45, followed by elimination of the *ipso*-chlorine.

In support of the observation by Baum *et al.* that inhibition of ecMurA by HR12 was prevented by the presence of DTT, we showed that incubating the HR45-modified

protein sample with 2 mM DTT resulted in a quantitative mass shift of -150 Da to the native protein mass, indicating complete removal of the modification.

To determine that the target was indeed cysteine-112 and not another nucleophilic residue on the protein, labelling experiments were performed with *N*-ethylmaleimide (NEM). NEM is a reagent commonly used in MS to selectively alkylate cysteine residues. On incubation of saFabH with 10 mM NEM for 30 minutes a mass shift of +125 Da was observed, corresponding to the alkylation of a single cysteine residue. Incubation of the NEM-alkylated sample with HR45 for 3 h resulted in no further change in mass, indirectly indicating that HR45 was unable to form an adduct with saFabH because NEM had irreversibly alkylated the crucial cysteine-112 residue.

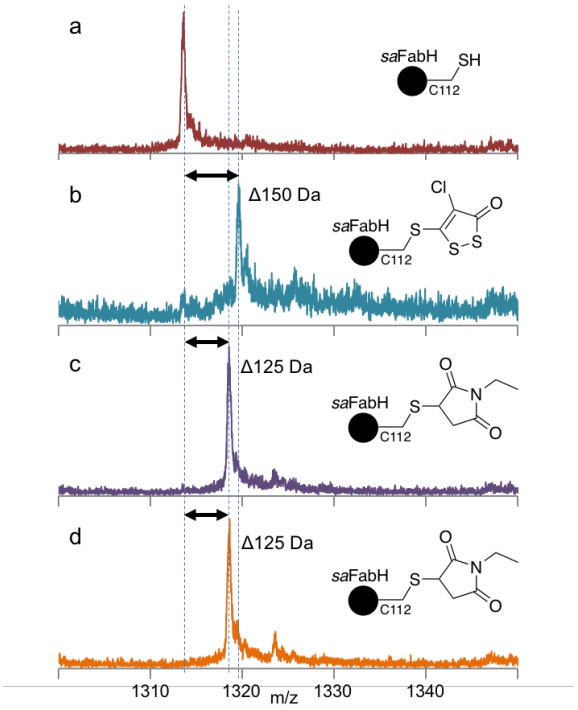


Figure 3.14 - The LC-MS spectra (26+ charge state) of saFabH a) unmodified control, b) protein incubated with HR45 for 30 mins at RT resulting in complete modification (+150 Da), c) protein incubated with N-ethylmaleimide (NEM) resulting in complete alkylation of the single cysteine residue (+125 Da), and d) NEM-alkylated protein incubated with HR45 for 3 hours at RT resulting in no mass change.

3.2.4 Digestion of saFabH/HR45 adduct

To obtain direct evidence that HR45 was modifying C112, we decided to enzymatically digest the modified protein and conduct MS/MS on the modification-containing fragment, in the hope of obtaining consecutive ions that would prove the site of modification beyond doubt.

Trypsin is a serine protease commonly used in proteomics for protein digestion and cleaves proteins at the C-terminal side of lysine and arginine residues, and has an optimal operating temperature of 37 °C at pH 8. A tryptic digest of both the HR45-modified and native saFabH was carried out overnight under optimal conditions (ammonium bicarbonate, pH 8.0, 37 °C). The mixture was cleaned up using a c18 ZipTip cartridge and infused into a 12T SolariX FT-ICR spectrometer. As well as the absence of the ion corresponding to the C112-containing tryptic peptide (103-VASMDQLAACSGFMYSMITAK-123, predicted monoisotopic mass 2224.0036 Da) in the control digest, we also failed to detect a peptide corresponding to the expected mass of the C112-HR45 adduct. This observation was consistent with the MurA study by Baum *et al.* that failed to detect MurA/HR12 adducts[124].

Because of the complexity of the obtained mass spectrum, caused by the presence of multiple charge states of each tryptic fragment, we were unable to conclusively identify any new signals in the HR45-modified sample that were absent in the control digest. We solved this problem by conducting a large-scale trypsin digest of unmodified-saFabH, followed by incubation with activated thiol sepharose. This was to allow us to capture the cysteine-containing peptide (103-VASMDQLAACSGFMYSMITAK-123) on the activated thiol sepharose solid phase and wash away all other non-cysteine containing peptides, before eluting the desired peptide by adding the reducing agent DTT. The peptide was desalted, lyophilised and resuspended in either ammonium bicarbonate (pH 8.0) or ammonium acetate (pH 7.0) buffer, in the presence or absence of HR45. All four of the samples were incubated at 37 °C overnight to simulate the conditions used in a tryptic digest. This

would allow us to clearly monitor any change in the mass spectrum, and determine if the slightly alkaline pH of the trypsin digest influenced the HR45 adduct.

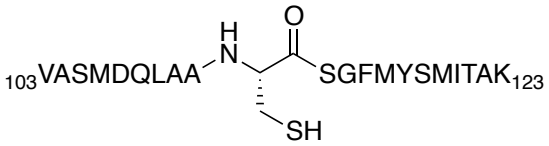


Figure 3.15 - Purified cysteine-112 containing peptide from large scale trypsin digest and activated thiol sepharose purification.

In ammonium acetate (pH 7.0), both the unmodified and HR45-modified forms were observed at 1113.0042 Da $[M+2H]^{2+}$ and 1187.9439 Da $[M+2H]^{2+}$, respectively. The latter corresponds to the addition of 150 Da, consistent with observations on the intact protein (fig. 3.16). In ammonium bicarbonate (pH 8.0) the unmodified peptide was found to exist predominantly as a $[M+4H]^{4+}$ disulfide dimer, as disulfide formation is favoured at alkaline pH approaching the pK_a of cysteine (8.3). Perplexingly, incubation of the HR45-modified peptide in ammonium bicarbonate resulted in a mass shift of -184 Da from the expected HR45 modified peptide mass, or -34 Da from the unmodified peptide. Using CID tandem MS, a fragmentation spectrum of this new peptide was recorded with good coverage. Simultaneous ions indicated the 34 Da loss was from cysteine, which we interpreted as the conversion of C112 to dehydroalanine (Dha). The observed Dha peptide mass of 1095.9991 Da $[M+2H]^{2+}$ was within 15 ppm of the expected mass. It was interesting to note that a small amount of the Dha peptide was also present in the unmodified sample at pH 8.0, possibly derived from the disulfide dimer (see fig. 3.16).

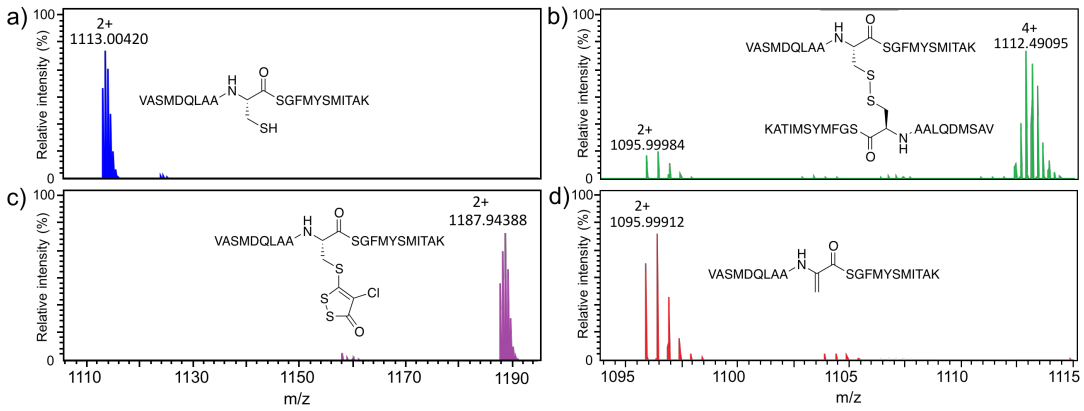


Figure 3.16 - Purified tryptic peptide (103-VASMDQLAACSGFMYSMITAK-123) in a) ammonium acetate (100 mM, pH 7.0) for 16 h at 37 °C and b) ammonium bicarbonate (100 mM, pH 8.0) for 16 h at 37 °C. The purified peptide modified with HR45 in c) ammonium acetate (100 mM, pH 7.0) for 16 h at 37 °C, and d) in ammonium bicarbonate (100 mM, pH 8.0) for 16 h at 37 °C. All masses were observed within 15 ppm of expected.

The conversion of cysteine to Dha under typical trypsin digest conditions may help explain previous failures to observe the covalent modification of saFabH and ecMurA by HR45 and HR12. The alkaline hydrolysis pathway of 1,2-dithiol-3-ones is a highly complex one, with several known products[125]. Although we do not yet have sufficient data to speculate on the mechanism of the C112-HR45 to Dha conversion, it is likely that one of the hydrolysis products facilitates the transformation, possibly by a similar mechanism to other Dha formation chemistry[132]. However, when we conducted analogous pH-dependence experiments on the intact protein level with the saFabH/HR45 adduct, no Dha-containing protein signal was detected.

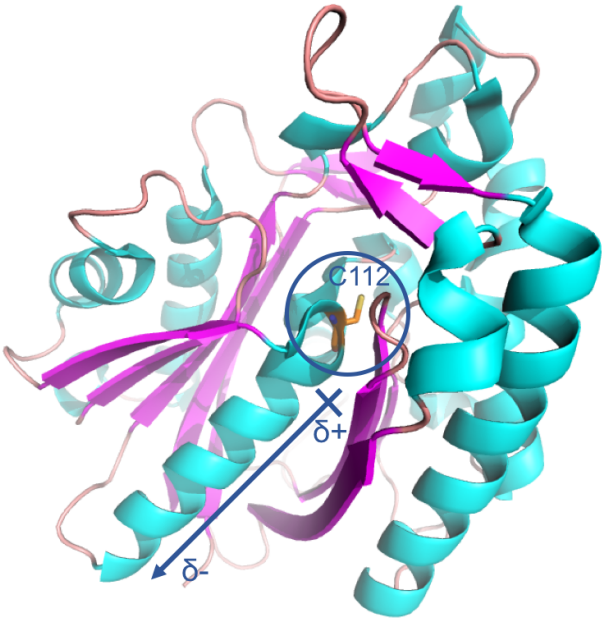


Figure 3.17 - The α-helix macrodipole effect illustrated on the structure of saFabH (PDB: 3IL7).

In the folded protein, C112 is located at the N-terminus of a long α-helix which lowers the pK_a of C112 from 8.8 to 7.2 in ecFabH[67,133]. The same phenomenon is observed in the structure of saFabH (see fig. 3.17). This is known as the α-helix dipole effect, and is essential for sufficiently increasing the nucleophilicity of the active site cysteine residue. Whether the increase in pK_a of C112 due to the absence of this effect in the tryptic peptide helps to facilitate the cysteine to Dha conversion requires further investigation.

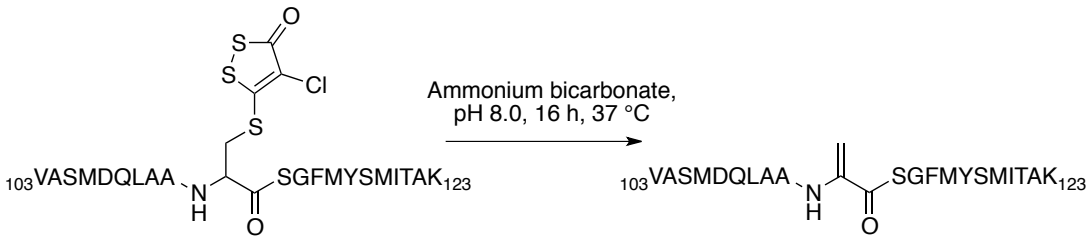


Figure 3.18 - The elimination of the HR45 adduct to form Dha under typical trypsin digest conditions.

The instability of the adduct under standard trypsin digest conditions led us to seek an alternative method of digesting the saFabH/HR45 adduct, rather than basing our investigation on the purified tryptic peptide. Pepsin is an aspartic protease that cleaves proteins at the C-terminal of tyrosine, phenylalanine, tryptophan and leucine residues with occasional non-specific cleavage after other residues. We elected to digest with pepsin as it has an optimum operating pH of 2.5 at 4 °C, which we hoped would leave the HR45-modification intact.

Unmodified saFabH was digested with pepsin (0.04 N HCl, pH 2.5, 4 °C, 60 mins). After desalting and infusion into a 12T SolariX FT-ICR mass spectrometer, the crucial cysteine containing tryptic peptides AACSGF and AACSGFM^{Ox} (M^{OX} denotes oxidation of the M residue) were identified. An identical digest was carried out with HR45-modified saFabH. Upon MS analysis of the digested HR45-modified sample, it was apparent that the two unmodified peptides were absent and had been replaced by two new peptides in the EIC. The masses of the two new peptides matched AAC*SGF and AAC*SGFM^{Ox}, where C* denotes a cysteine modification of +150 Da, corresponding to addition of HR45 with the loss of HCl. The two most abundant isotopes of chlorine are ^{35}Cl and ^{37}Cl, in an approximate ratio of 3:1. This gives the presence of chlorine a very distinctive appearance in a mass spectrum, with an M+2 peak of a third of the intensity. Gratifyingly, the HR45-modified peptide signals displayed a diagnostic isotope pattern consistent with the ^{35}Cl and ^{37}Cl isotopes in a 3:1 ratio, a characteristic not present in the unmodified peptides. To confirm the cysteine residue as the site of modification, CID tandem MS was used to fragment AAC*SGFM^{Ox}. A Mascot search (Matric Science) identified the generated ions and confirmed that the modification was indeed on the cysteine residue (see fig. 3.19 and fig. 3.21).

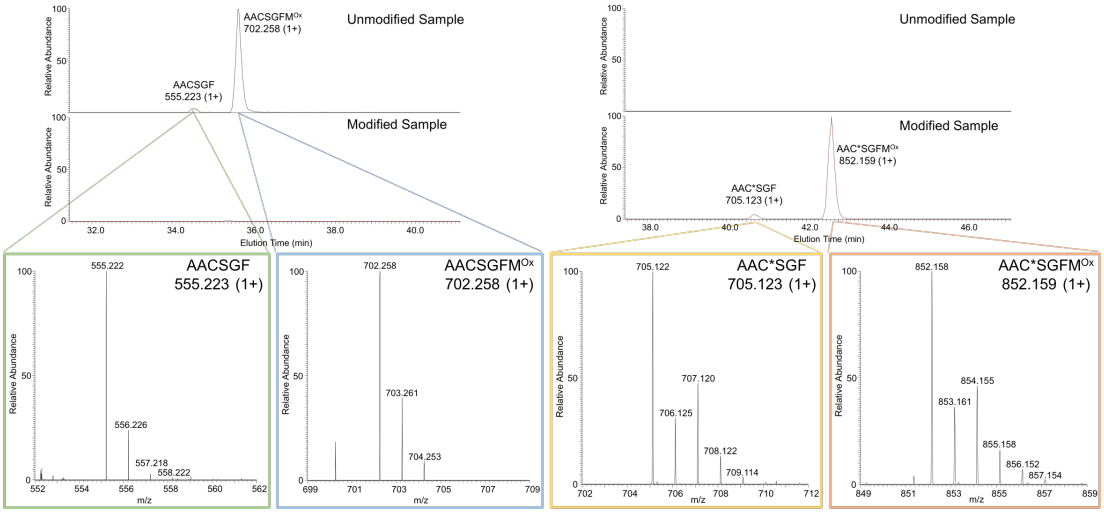


Figure 3.19 - The EICs of cysteine-112 containing peptides after pepsin digest. *Left* unmodified saFabH with expanded MS spectra of two separate unmodified peptides. *Right* HR45-modified saFabH with expanded MS spectra of two separate modified peptides, C* corresponding to an addition of +150 Da at C112. Note the appearance of the diagnostic chlorine isotope pattern in both modified ions, corroborating the proposed addition/elimination mechanism (see fig. 3.21). Note also that no Dha-containing fragments were identified, suggesting the previously discussed phenomenon is indeed pH-dependent.

3.2.5 Structural studies

Despite observing quantitative modification of saFabH with HR45, we were unable to obtain crystals of the HR45/saFabH adduct using both literature conditions[105] and commercial screens (JCSG+, PEG-ION, MIDAS). With the help of Dr Jon Marles-Wright at the University of Newcastle we docked the recently deposited small molecule crystal structure of HR45[134] into the active site of saFabH[105] (PDB: 3IL7) guided by our proposed mechanism.

'The X-ray structure of HR45 was obtained from the CCDC (CCDC Number: 1517056)[134]. The coordinates for a Cysteine-HR45 C5-SG adduct were produced using JLigand[135]; and the adduct was then manually docked into the crystal structure of saFabH to replace Cys112 (PDBID: 3IL7)[105] and adjusted to minimise steric clashes and to maximize hydrogen bonding potential using coot[136]. The resulting model was then energy minimized using refmac5[137]. Model images were produced using PyMol (Schrödinger LLC).'

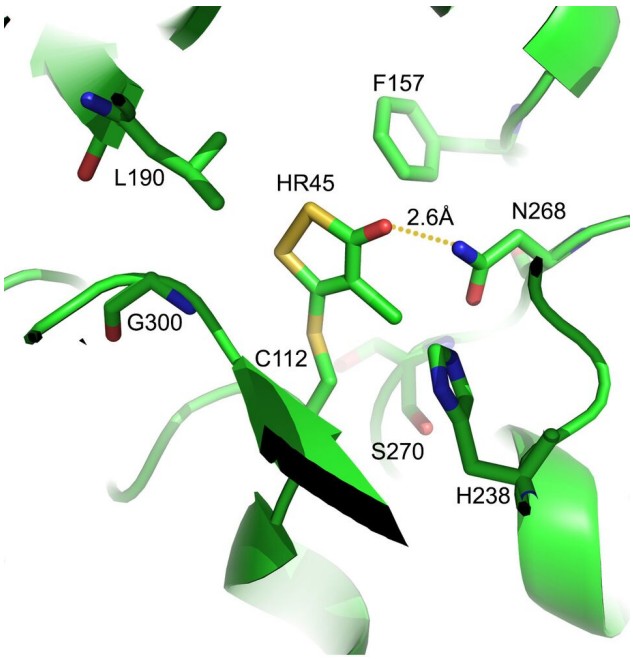


Figure 3.20 - Model of HR45 small molecule crystal structure (CCDC: 1517056) docked into saFabH protein crystal structure (PDB:3IL7).

Although there is ample space for HR45 in the active site, catalytically essential cysteine-112 residue has little freedom of movement, limiting the likely positions of the ligand. Interestingly, our model suggested there was hydrogen bonding between the C3 carbonyl of HR45 and the side chain of catalytically essential N268[67]. This agrees with the work done by He *et al.* who identified *in silico* that this was an essential interaction between thiolactomycin and FabH[123].

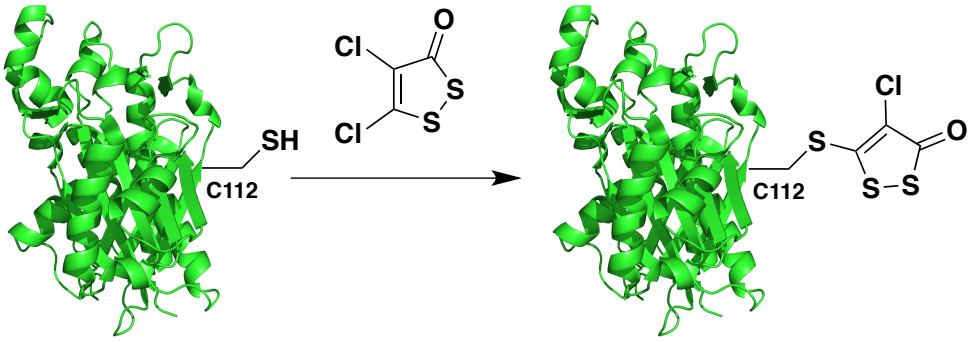


Figure 3.21 - Summary of the mechanism of action of HR45 on FabH.

3.3 Conclusion

In this study, we determined that 1,2-dithiol-3-one HR45 covalently modifies saFabH through a Michael-type addition elimination at the catalytic residue C112. FabH from *S. aureus* was chosen as the target enzyme due to the presence of only one cysteine residue in its amino acid sequence. FabH from *E. coli* has 5 cysteine residues with varying degrees of solvent exposure. We briefly attempted MS studies with the *E. coli* isoform, however the MS spectra were complicated and unreproducible, suggesting some degree of non-specific adduct formation. Despite this, we suggest that the major mode of inhibitory action of HR45 on FabH from *E. coli* and other bacteria remains covalent modification of the catalytically essential active site cysteine. We also suggest that this is likely to be the mode of action of HR12 on ecMurA in the study by Baum *et al.*, owing to the similarity of a catalytically essential active site cysteine residue[124].

We have also shown that the likely reason that this covalent modification has previously escaped detection is that under normal tryptic digest conditions, the cysteine-112 residue of HR45-modified saFabH is converted to Dha, either through the complex alkaline hydrolysis pathway of HR45, a change in the electronics of the sulfur atom at peptide level, or a combination of the two. The precise mechanism of how this elimination reaction occurs was not investigated, however this would make for interesting future study and may lead to new methods of generating unnatural amino acid containing proteins *in situ*.

Describing the precise mode of action of molecules isolated from screening campaigns that failed to make the clinic may have impact on the design and discovery of new antimicrobial leads[56]. Our detailed mechanistic study of HR45 provides insight for the design of new antimicrobials, and this approach will no doubt be useful for other systems.

3.4 Experimental

3.4.1 Mass Spectrometry

3.4.1.1 Peptide analysis

Purified tryptic peptide samples were analysed by direct infusion electrospray ionisation using a TriVersa nanomate (Advion) coupled to a Bruker 12T SolariX with a 10cm infinity cell. Spectra were acquired in positive polarity over 100-2000m/z with a 2megaword data size. 20 spectra were summed, each with a 50 ms accumulation time. Pepsin digests were analysed by C18 LC-MS/MS over a 50 min gradient from 5% to 45% acetonitrile/0.1% formic acid. MS spectra were acquired on an LTQ Orbitrap in positive polarity over a 333-1800 m/z range using a 371.10124 lockmass. MS scans were acquired at 60k resolution followed by 6 data depended ion trap MSMS spectra, or at 30k resolution followed by 4 7500 resolution Orbitrap MSMS spectra. In additional analyses, candidate modified peptides were fragmented in the Orbitrap, as above, but using a data independent inclusion list. MSMS spectra were converted to mgf format and peptides assigned using the Mascot search engine (www.matrixscience.com).

3.4.1.2 LCMS of denatured protein

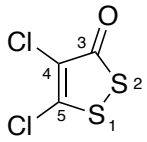
2 μL of each protein sample (10 μM) was injected onto a 1 x 250 mM analytical ProSwift RP-4H (Thermo) column and eluted at a flow rate of 250 μL/min with a gradient between buffer A (0.1% v/v formic acid) and buffer B (acetonitrile, 0.1% v/v formic acid): Initial 0-3 min at 5% B to waste, 3-8 min 5%-95% B, 8-8.5 min 95% B, 8.5-9 min 95%-5% B, 9-10 min 5% B. Electrospray ionization was coupled to a Synapt G2 Q-TOF (Waters) with a 120 °C source temperature, 2 kV capillary voltage, and sampling cone voltage of 35. MS spectra were acquired over a 150–2,000 *m/z* mass range in positive polarity and in sensitivity mode. A lockmass correction was applied once at the beginning of each analysis against leucine enkephalin peptide.

3.4.2 NMR of model reaction of HR45 with *N*-acetylcysteine

Nuclear magnetic resonance (NMR) spectra were recorded at 298 K on Bruker PRO500 or AVA500 spectrometers running at 500 MHz (^{1}H spectra) or 126 MHz (^{12}C spectra). Chemical shift values (δ) are reported in parts per million (ppm) relative to tetramethylsilane (δ_{TMS} = 0) and are referenced to the residual solvent peak. ^{1}H NMR data are reported in the format: chemical shift, relative intensity, multiplicity (s = singlet, d = doublet, t = triplet, m = multiplet, br = broad), coupling constant (*J* value, Hz), and assignment. ^{13}C NMR data are reported in the format: chemical shift and assignment. ^{1}H/^{13}C HMBC NMR data are reported qualitatively.

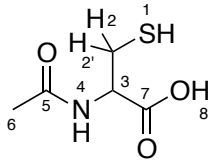
To a solution of *N*-acetylcysteine (15 mM) in ammonium bicarbonate (100 mM, D_2O) was added 4,5-dichloro-1,2-dithiole-3-one (15 mM) and DMSO was added to a final concentration of 10% (v/v). The reaction was transferred to an NMR tube, and ^{1}H spectra were recorded at 6 h intervals overnight. ^{13}C and ^{1}H^{13}C HMBC spectra were recorded after 12 h. A ^{13}C spectrum of 4,5-dichloro-1,2-dithiole-3-one, and ^{1}H and ^{13}C spectra of *N*-acetylcysteine were recorded as references.

4,5-Dichloro-1,2-dithiole-3-one (HR45)



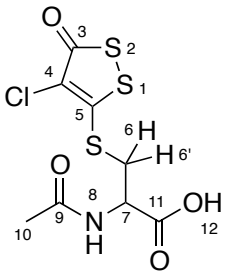
^{13}C NMR (126 MHz, DMSO-d_6) δ 121.48 (C-4), 153.7 (C-5), 184.6 (C-3)

***N*-Acetylcysteine**



^{1}H NMR (500 MHz, DMSO-d_6) 1.88 (3H, s, CH_3-6), 2.36-2.48 (1H, m, SH-1), 2.68-2.78 (1H, m, CH_2-2), 2.78-2.88 (1H, m, CH_2-2'), 4.37 (1H, ddd, ^{3}J = 4.7, 7.7, 7.7 Hz, CH-3), 8.16 (1H, d, ^{3}J = 7.7 Hz, NH-4), 12.80 (1H, s (br), OH-8)

HR45/*N*-acetylcysteine adduct



^{1}H NMR (500 MHz, DMSO-d_6) 1.87 (3H, s, CH_3-10), 3.55 (1H, dd, ^{2}J = 13.5 Hz, ^{3}J = 8.5 Hz, CH_2-6), 3.70 (1H, dd, ^{2}J = 13.5 Hz, ^{3}J = 4.5 Hz, CH_2-6'), 4.59-4.65 (1H, m, CH-7), 8.54 (1H, d, ^{3}J = 8.1 Hz, NH-8), 9.65 (1H, s, OH-12); **^{13}C NMR** (126 MHz, DMSO-d_6) δ 22.3 (CH_3-10), 32.7 (CH_2-6), 51.3 (CH-7), 115.7 (C-4), 163.2 (C-5), 169.7 (C-9), 170.9 (C-11), 184.3 (C-3)

^{1}H^{13}C HMBC Cross-peaks between C-5 and CH_2-6 (3.53, 163.7) and CH_2-6' (3.72, 163.7), and absence of cross-peaks between C-4 and CH_2-6 and CH_2-6' confirm that the addition-elimination reaction has taken place at the C-5 position.

3.4.3 Modification of *sa*FabH with HR45

A stock solution of HR45 (25 mM) in DMSO was always prepared immediately before use. In a typical modification experiment, HR45 (25 mM, DMSO) was added to an aliquot of saFabH (12 μM) such that the final concentration of HR45 was 250 μM with a final concentration of 1% DMSO (v/v). The solution was mixed and left for 0.5 h at 37 °C.

3.4.4 Modification of saFabH with *N*-ethylmaleimide (NEM)

In a NEM-labelling experiment, NEM (10 mM, 100 mM PBS, pH 7.4) was added to an aliquot of saFabH (12 μM, 100 mM PBS, pH 7.4) such that the final concentration of NEM was 1 mM. The solution was mixed and left for 2 h at room temperature. The protein was desalted into fresh buffer to remove NEM before analysis.

3.4.5 Removal of saFabH/HR45 modification with dithiothreitol (DTT)

To an aliquot of saFabH/HR45 (12 μM) was added DTT to a final concentration of 2 mM. The solution was mixed and left for 15 mins at room temperature. The protein was desalted into fresh buffer to remove NEM before analysis.

3.4.6 Trypsin digest and thiol bead purification

saFabH (2.5 mL, 0.4 mg/mL) in non-reducing buffer (20 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol) was incubated with trypsin (40 μL, 0.5 mg/mL, Promega proteomics grade) overnight at 37 °C with shaking at 300 rpm. The reaction was then incubated with 4B-activated thiol resin (100 mg dry weight, pre-equilibrated in non-reducing buffer) and slowly rotated for 1 h. The slurry was then added to a 1 mL spin-cartridge fitted with a frit. The slurry was washed with non-reducing buffer (3 x 400 μL) before the bound cysteine-containing peptide was eluted with elution buffer (200 μL, 20 mM DTT, 20 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol). The eluate was then desalted into MeCN:H_2O (2:1, 0.1% formic acid, 200 μL) using a C18 SPE cartridge and used immediately or lyophilised and stored at -20 °C.

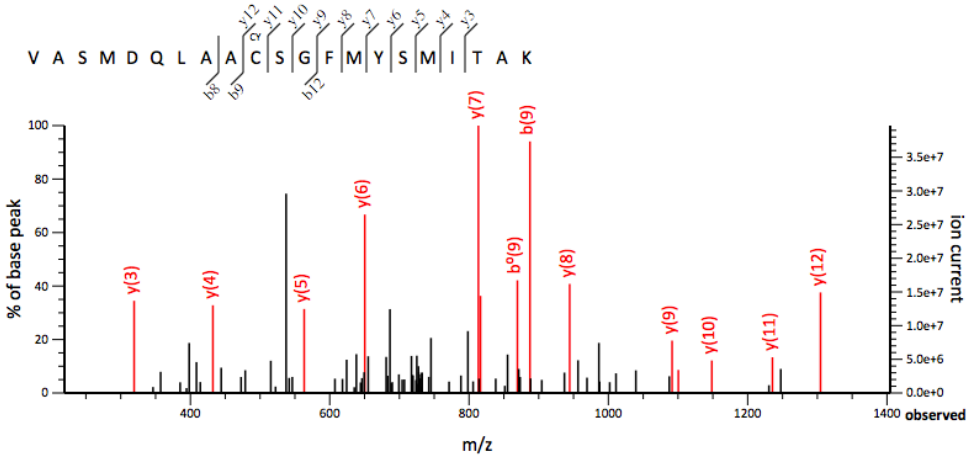
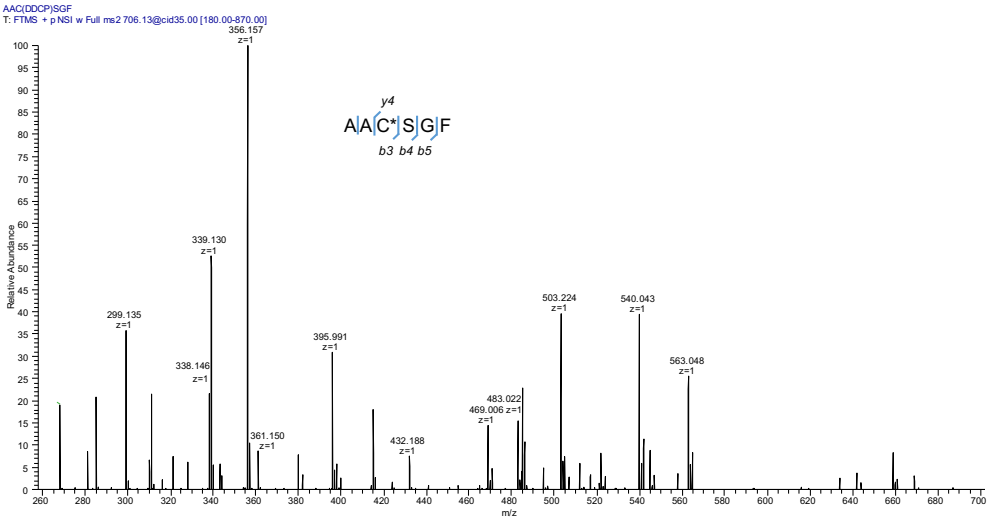


Figure 3.21 - MS^2 fragmentation of purified tryptic peptide/HR45 adduct after Dha forming conditions.

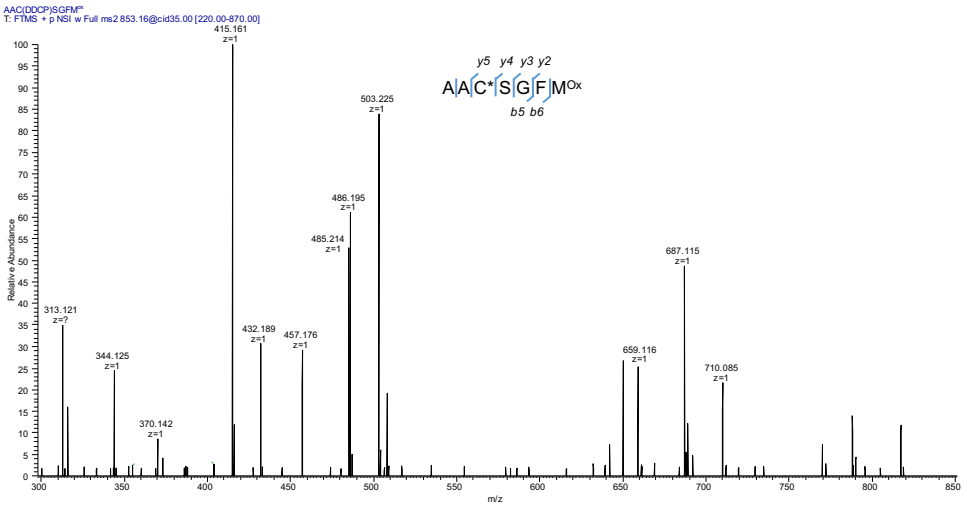
3.4.7 Pepsin digest

To an aliquot of saFabH (2.5 mL, 0.4 mg/mL) in non-reducing buffer (20 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol) was added HCl (100 μL, final conc. 0.04 N). Lyophilised pepsin (Promega) was resuspended in ddH_2O (1 mg/mL, pH 5.5), and added (25 μL, 1:40 w/w) to the acidified protein aliquot and incubated overnight at 37 °C with shaking at 300 rpm. The reaction was stopped by heating to 95 °C for 10 minutes. The peptic digest was then desalted into MeCN:H_2O (2:1, 0.1% formic acid, 200 μL) using a C18 SPE cartridge and used immediately or lyophilised and stored at -20 °C.



299.1353455	*B4-nHR45*
338.1459045	*B5-nHR45-water*
339.1297302	*B5-nHR45-ammonia*
356.1565552	*B5-nHR45*
361.1502686	*P-nHR45-H2O-Ala-Ala*
395.9905701	*B3*
432.1878052	*P-nHR45-water-Ala*
483.0220337	*B4*
503.2238464	*P-nHR45-water*
540.0432739	*B5*
563.0484619	*Y4*

Figure 3.22 - MS^2 of peptic peptide 706.13, AAC*SGF and assignment**.**



313.1210327	*Y2*
344.1246033	*B5-nHR45-ammonia-Ala*
370.1424561	*Y3*
415.1611023	*B5-nHR45-ammonia*
432.1890259	*B6-nHR45-ammonia-Ala*
457.1759338	*Y4*
485.214325	*B6-nHR45-water*
486.195282	*B6-nHR45-ammonia*
503.2246094	*B6-nHR45*
659.1160889	*B5-CO*
687.1149292	*B6*
710.0846558	*Y5*

Figure 3.23 - MS^2 of peptic peptide 853.16, $AAC*SGFM^{Ox}$ and assignment.

3.4.8 Cloning of *saFabH* into pET-28a

3.4.8.1 Transformation of *saFabH*/pUC-19

saFabH was purchased as a codon-optimised gene (Genscript) in the pUC-19 with 5'-Nco1 and 3'-BamH1 restriction sites. *saFabH*/pUC-19 (2 µL) was added to an aliquot (25 µL) of DH5α competent cells and set on ice for 30 minutes. The cells were heat shocked at 42 °C for 30 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (100 µg/mL ampicillin) and incubated overnight at 37

°C. The biggest colonies were picked into separate universal tubes containing LB broth (5 mL, 100 µg/mL ampicillin) and cultured overnight at 37 °C with agitation. DNA was isolated using the Qiagen Spin MiniPrep kit.

3.4.8.2 Restriction digest of *saFabH* and pET-28a

The *saFabH* gene was excised from the pUC-19 vector using the Nco1 and BamH1 restriction sites previously introduced and inserted into a pET-28a vector with an N-terminal His-tag. The restriction digests of the *saFabH*/pUC-19 and empty pET-28a constructs were done according to the following protocol:

1.5 µL	Nco1
1.5 µL	BamH1
6.0 µL	CutSmart buffer (10x)
50.0 µL	*saFabH*/pUC-19 construct or pET-28a

The reactions were incubated at 37 °C for 2 hours and loaded directly onto a 1% agarose gel (TAE buffer, 0.01% gel red). A band from each experiment corresponding to the saFabH gene and the empty pET-28a vector (at approx. 1000 bp and approx. 5400 bp, respectively) were excised and purified using the Qiagen QIAquick gel extraction kit.

3.4.8.3 Ligation of *saFabH* into pET-28a

The *saFabH* gene was ligated into the prepared pET-28a vector according to the following protocol:

1.0 µL	T4 DNA ligase
1.0 µL	Digested pET-28a
3.0 µL	Digested *saFabH*
5.0 µL	Rapid ligation buffer (2x)

The reaction proceeded at room temperature for 30 minutes.

3.4.8.4 Transformation and analytical restriction digest of *saFabH*/pET-28a

The product from the ligation reaction (4 µL) was used to transform an aliquot (50 µL) of C2987 high-competency cells and set on ice for 25 minutes. The cells were then heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (30 µg/mL kanamycin) and incubated overnight at 37 °C. Transformants were picked into LB broth (5 mL, 30 µg/mL kanamycin) and cultured overnight at 37 °C with agitation. DNA was isolated using the Qiagen Spin MiniPrep kit, and the success of the ligation was determined by analytical restriction digest according to the following protocol:

1.0 µL Nco1

1.0 µL BamH1

1.0 µL CutSmart buffer (10x)

7.0 µL *FabH*/pET-28a

The reaction was incubated at 37 °C for 2 hours and loaded directly onto a 1% agarose gel (TAE buffer, 0.01% gel red) for analysis.

3.4.8.5 PCR Amplification of *saFabH*

The *saFabH* gene from the *saFabH*/pET-28a construct was amplified by PCR, using primers to introduce 5'-Nco1 and 3'-XhoI restriction sites.

saFabH Nco1 Fwd: 5'-CCGGCCATGGATGAATGTGGGTATTAAGGGTTT-3'

saFabH Xho1 Rev: 5'-GGCCCTCGAGTTATTTGCCCCAC-3'

The PCR reaction contained:

5.0 µL PFu DNA Polymerase buffer (2x)

2.5 µL *saFabH*/pET-28a construct DNA

1.5 µL *saFabH Nco1 Fwd* primer (10 µM)

1.5 µL *saFabH Xho1 Rev* primer (10 µM)

1.0 µL dNTP mix (10 µM)

1.0 µL PFu DNA Polymerase (1 U/µL)

37.5 µL ddH_2O

The PCR reaction was conducting using the following thermal cycle: 98 °C for 2 minutes, (98 °C for 30 seconds, 60 °C for 30 seconds, 72 °C for 4 minutes) x 30, 72 °C for 10 minutes. The PCR product was purified on a 1% agarose gel (TAE buffer, 0.01% gel red) and the DNA was isolated using the Qiagen QIAquick gel extraction kit. The DNA was digested and ligated into pET-HISTEV as previously described, using Nco1 and Xho1 restriction enzymes. The product from the ligation reaction was transformed into C2987 cells and selected colonies were analysed by analytical restriction digest as previously described.

3.4.9 General method for expression and purification of saFabH

Note: Prior to establishing the importance of removing the affinity tag from saFabH, the protein was expressed in the pET-28a vector followed by a two-step purification by affinity chromatography and SEC. All methods were analogous to the following protocol, emitting the incubation step with TEV-protease.

The *saFabH*/pET-HISTEV construct (4 µL) was transformed into an aliquot (50 µL) of BL21(DE3) (New England BioLabs) cells and set on ice for 25 minutes. The cells were then heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (30 µg/mL kanamycin) and incubated overnight at 37 °C. A single transformant was used to inoculate two seed cultures of sterile LB broth (2 x 250 mL, 30 µg/mL kanamycin) and agitated overnight at 37 °C. One of the overnight seed cultures was used to sub-culture sterile LB broth (5 x 500 mL, 30 µg/mL kanamycin) to an OD_{600} of 0.1. The cultures were agitated at 37 °C until the OD_{600} reached 0.6, at which point expression was induced by addition of IPTG (final conc. 0.1 mM). Cells were harvested by centrifugation (30 minutes at 5000 rpm) after a further 3 hours at 30 °C and subsequently stored at -20 °C.

N-Terminal histidine-tagged saFabH was purified at 4 °C by Ni-affinity chromatography followed by size exclusion chromatography. The BL21 (DE3) cell pellet expressing FabH was resuspended in lysis buffer (30 mL, 20 mM Tris-HCl pH 7.6, 300 mM NaCl, 5 mM imidazole) and lysed for 15 minutes with rounds of 30 second of sonication followed by 30 seconds of rest. The cell lysate was clarified by centrifugation (*18,000 g*, 30 minutes, 4 °C) and the cell-free extract was injected onto a HisTrap 5 mL (GE Healthcare) Ni^{2+}-affinity chromatography column pre-equilibrated in lysis buffer. The column was washed with lysis buffer (5 CV) before the histidine-tagged protein was eluted using a gradient (0-100%) of lysis buffer to elution buffer (20 mM Tris-HCl pH 7.6, 300 mM NaCl, 400 mM imidazole) over 20 CV.

Each elution fraction was analysed by SDS-PAGE, and the fractions containing His-tagged saFabH were pooled, and the protein concentration was determined by a Bradford assay (Thermo Fisher). In order to remove the *N*-terminal histidine tag, a 5 mL aliquot of histidine-tagged saFabH was combined with an aliquot of TEV protease (1:5 w/w) and dialysed against size exclusion mobile phase buffer (500 mL, 20 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol) at 4 °C for 16 h using 8 kDa MWCO dialysis tubing. The reaction mixture was then injected onto a HisTrap 5 mL (GE) Ni^{2+}-affinity chromatography column pre-equilibrated in lysis buffer (30 mL, 20 mM Tris-HCl pH 7.6, 300 mM NaCl, 5 mM imidazole). The flow-through was collected and concentrated to a volume of 2 mL using a 30 kDa MWCO spin-filter.

The concentrated non-histidine-tagged saFabH sample was further purified by size exclusion chromatography (HiLoad Superdex 200 16/60, GE Healthcare) with an isocratic elution of mobile phase buffer (500 mL, 20 mM Tris-HCl pH 7.6, 100 mM NaCl, 10% glycerol) at 1 mL/min over 120 minutes. The saFabH protein eluted at approximately 70 minutes and the most concentrated fractions were pooled and flash cooled in 500 µL aliquots in liquid nitrogen before storage at -80 °C.

3.4.10 Expression and purification of TEV protease

The TEV/GST fusion construct (4 µL) was transformed into an aliquot (50 µL) of BL21 (DE3) Rosetta Gami (Millipore) cells and set on ice for 25 minutes. The cells were then heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (30 µg/mL kanamycin, 25 µg/mL chloramphenicol) and incubated overnight at 37 °C. A single transformant was used to inoculate two seed cultures of sterile LB broth (2 x 250 mL, 30 µg/mL kanamycin, 25 µg/mL chloramphenicol) and agitated overnight at 37 °C. One of the overnight seed cultures was used to sub-culture sterile LB broth (5 x 500 mL, 30 µg/mL kanamycin, 25 µg/mL chloramphenicol) to an OD_{600} of 0.1. The cultures were agitated at 37 °C until the OD_{600} reached 0.6, at which point the temperature was reduced to 20°C and expression was induced by addition of IPTG (final conc. 0.4 mM). Cells were harvested after a further 16 h at 20 °C and subsequently stored at -20 °C.

N-Terminal histidine-tagged TEV protease was purified at 4 °C by Ni-affinity chromatography followed by size exclusion chromatography. The BL21 (DE3) Rosetta Gami cell pellet expressing FabH was resuspended in lysis buffer (PBS, 300 mM NaCl, 10 mM imidazole, 1 mM PMSF, 1 mM benzamidine) and lysed for 15 minutes with rounds of 30 second of sonication followed by 30 seconds of rest. Excess cell matter was removed by centrifugation (*18,000 g*, 30 minutes, 4 °C) and the supernatant was injected onto a HisTrap 5 mL (GE) Ni^{2+}-affinity chromatography column pre-equilibrated in lysis buffer. The column was washed with lysis buffer (5 CV) before the histidine-tagged protein was eluted using a gradient (0-100%) of lysis buffer to elution buffer (PBS, 300 mM NaCl, 500 mM imidazole, 1 mM PMSF, 1 mM benzamidine) over 20 CV.

Each elution fraction was analysed by SDS-PAGE, and the fractions containing His-tagged TEV protease were pooled and dialysed against self-cleavage buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM PMSF) at 4 °C for 2 h using 8 kDa MWCO dialysis

tubing in the absence of benzamidine to allow self-cleavage of the GST solubility-tag domain. The reaction was concentrated to 2 mL using a 30 kDa MWCO spin filter and further purified by size exclusion chromatography (HiLoad Superdex 200 16/60) with an isocratic elution of mobile phase buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% glycerol) at 1 $mL.min^{-1}$ over 120 minutes. TEV protease eluted at approx. 80 minutes and the most concentrated fractions were pooled, and dialysed against storage buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 50% glycerol). TEV protease was flash cooled in 500 µL aliquots in liquid nitrogen and stored at -80 °C.

Chapter 4: Monitoring a Pyridoxal 5'-Phosphate DCL by Mass Spectrometry

Abstract

Dynamic combinatorial chemistry is a powerful tool to identify new ligands for biological targets, however the technique has failed to gain traction in industrial drug discovery due to difficulties in deconvoluting large DCLs. In this chapter, we seek to demonstrate the power of ESI-MS in allowing for greater library sizes in DCC, circumventing the size-limiting implications resulting from the need for spectral or chromatographic resolution. Here, we discuss the construction of a DCL targeting 7,8-diaminononanoic acid synthase (BioA) from *M. tuberculosis* (*Mtb*) and compare the results to those obtained from the same library templated by the *E. coli* isoform. BioA is a pyridoxal 5'-phosphate (PLP) dependent enzyme that catalyses the second step in the biotin biosynthetic pathway, and is essential in both latent and actively dividing forms of *Mtb*. The DCL was developed around the well-studied *N*-acylhydrazone (NAH) exchange reaction to explore the interaction of hydrazide inhibitors with the aldehyde moiety of the PLP cofactor. From the 28-membered DCL, 4 privileged PLP-NAHs were selected by both isoforms. Although no *in vivo* activity was detected against the non-pathogenic *Mtb* substitute *M. smegmatis*, the results from the DCL experiment were corroborated by *in vivo* activity against *E. coli* and *A. baumannii*.

4.1 Introduction

4.1.1 Tuberculosis

Tuberculosis (TB) is the disease caused by the bacterium *Mycobacterium tuberculosis* (*Mtb*), and is one of the top-10 killers worldwide[138]. The World Health Organisation (WHO) estimates that 1 in 3 people is infected with TB, although 90% of these infections are latent and not immediately contagious. In 2015, there were 8 million newly infected patients and almost 2 million deaths, with 95% of these cases occurring in developing countries[138]. TB has a particularly high mortality rate in immune-compromised patients, whose alveolar macrophages are less effective at preventing the bacterial infection from taking hold when *Mtb* bacilli are inhaled. TB accounts for more than 35% AIDS-associated mortalities[139,140]. The current first line

treatment for TB is a 6-month combination therapy of isoniazid, rifampicin, pyrazinamide and ethambutol (HRZE)[141]. A combination of the protracted nature of this therapy with unpleasant side effects and inadequate public health information often results in low patient compliance. This, combined with a limited and dated arsenal of anti-TB drugs has led to rapid emergence of both multidrug resistant TB (MDR-TB), and extensively drug-resistant TB (XDR-TB)[142]. The incidence of drug resistant forms is on a rapid and worrying incline, making the need for new drugs that exploit different targets in *Mtb* increasingly urgent.

Latent infections provide another hurdle in successfully treating TB. Latent pathogens are often non-replicating, making them difficult to treat with conventional antibiotics that target mechanisms involved in replication. A rational approach in designing new anti-TB drugs is to target metabolic processes that are essential in both latent and actively dividing *Mtb* infections[141]. Biotin biosynthesis is an example of a pathway that has been shown to be essential in both latent and active *Mtb*, and thus has been identified as a promising target for new anti-TB drugs[143].

4.1.2 Biotin biosynthesis

Biotin (vitamin B7) is an essential cofactor used by several carboxylase enzymes. It is a bicyclic small molecule composed of fused ureido and tetrahydrothiophene rings with a valeric acid tail attached to the latter (see fig. 4.1). Many plants, fungi and micro-organisms are biotin autotrophs, synthesising the cofactor *de novo*. In contrast, mammals are unable to synthesise biotin and must obtain it from exogenous sources, such as from their gut microbiota or dietary intake[144]. The absence of an analogous biotin biosynthesis pathway in mammals makes the pathway a potential target for developing new antibiotics[143].

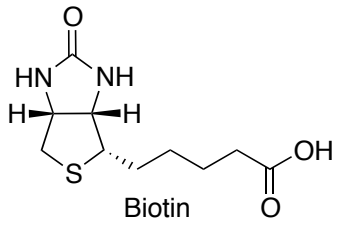


Figure 4.1 - Biotin.

Mtb has attracted attention in the search for antibiotics targeting biotin biosynthesis, as it has been shown that the bacterium is unable to import exogenous biotin. Other bacteria have evolved methods to recruit biotin from outside of the cell, the most well documented of which is the biotin transporter protein (BioY), however homologues in the *Mtb* genome remain elusive[143,145]. In *Mtb*, biotin is essential for the enzymes pyruvate carboxylase and acyl-CoA carboxylase, fundamental proteins involved in the Krebs cycle and fatty acid biosynthesis, respectively[143]. Knockout studies of the biotin biosynthesis enzymes of *Mtb* have shown the pathway is essential for survival both *in vitro* and *in vivo*[146,147]. The final four steps in biotin biosynthesis are conserved across all biotin autotrophs, which are catalysed by 8-amino-7-oxononanoic acid synthase (BioF), 7,8-diaminononanoic acid synthase (BioA), dethiobiotin synthase (BioD) and biotin synthase (BioB), respectively. These four enzymes sequentially transform thioester-linked pimeloyl-ACP into biotin (see fig. 4.2).

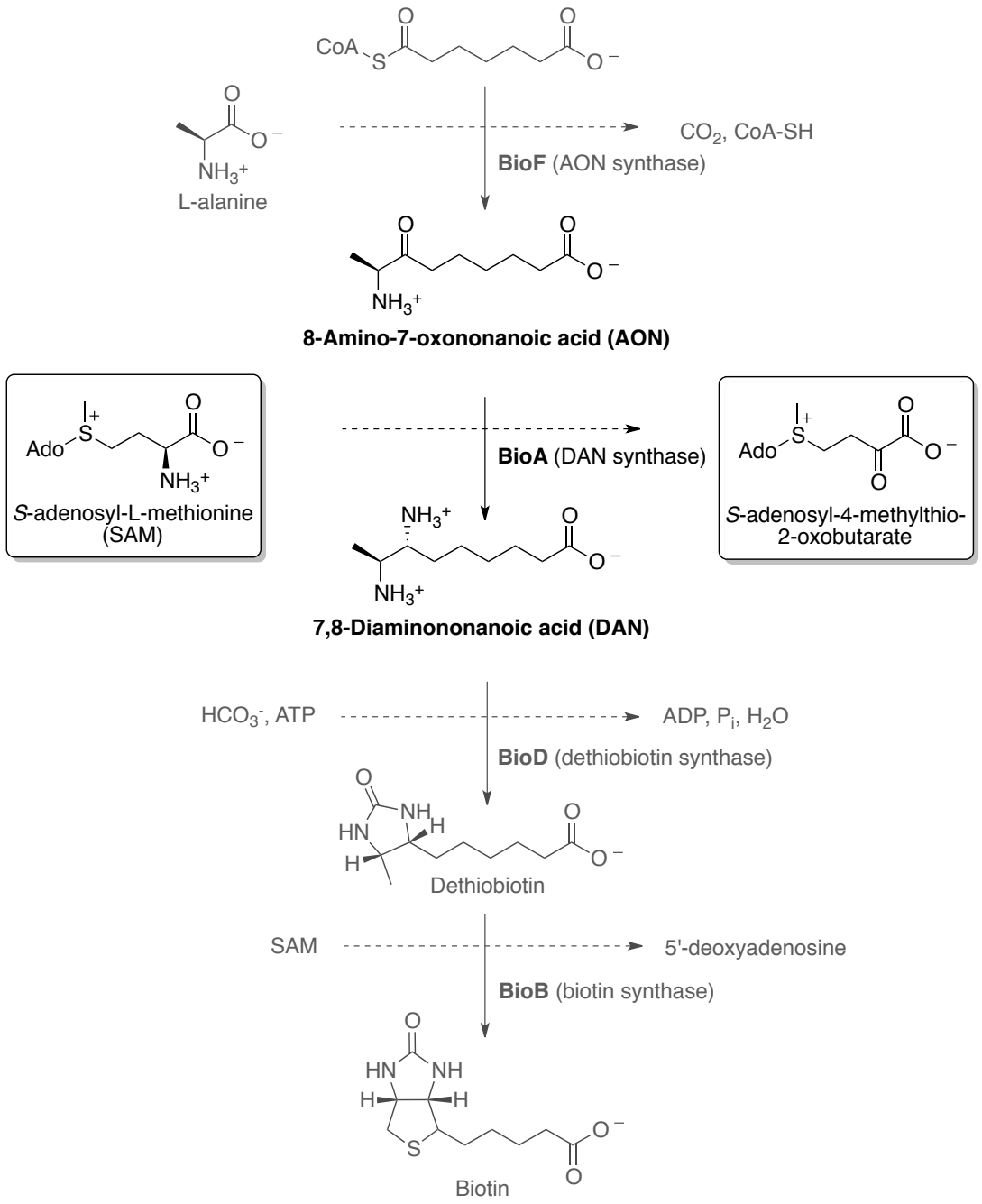


Figure 4.2 - Final four conserved steps in the biotin biosynthesis pathway.

4.1.3 BioA

In its native state, *Mtb* BioA (mtBioA) is found as a 93 kDa homodimer. It catalyses the second step of the final sequence of biotin biosynthesis, transforming 8-amino-7-oxononanoic acid (AON) into 7,8-diaminononanoic acid (DAN) in a transamination

reaction. Mouse models using tetracycline-regulated BioA expression have shown that the enzyme is essential not only for establishing an infection, but also maintaining bacterial persistence[147]. Park *et al.* also generated a deletion mutant Δ*bioA*, showing it was not viable in biotin-free media, even when supplemented with AON. Perplexingly, they showed that growth could be rescued on addition of dethiobiotin (DTB) or biotin, suggesting that *Mtb* was indeed able to recruit exogenous biotin, although it required a concentration at least 25x that found in human serum to have any effect[147]. Unusually, biotin starvation of Δ*bioA* resulted in cell death, whereas *Mtb* auxotrophs would normally enter a state of growth arrest[148]. Park *et al.* determined through variable expression experiments that the wild type expression of BioA required a 99% reduction before *Mtb* lost viability. Collectively, their findings suggest that BioA could be a promising TB drug target, yet with challenging efficacy requirements[51,147].

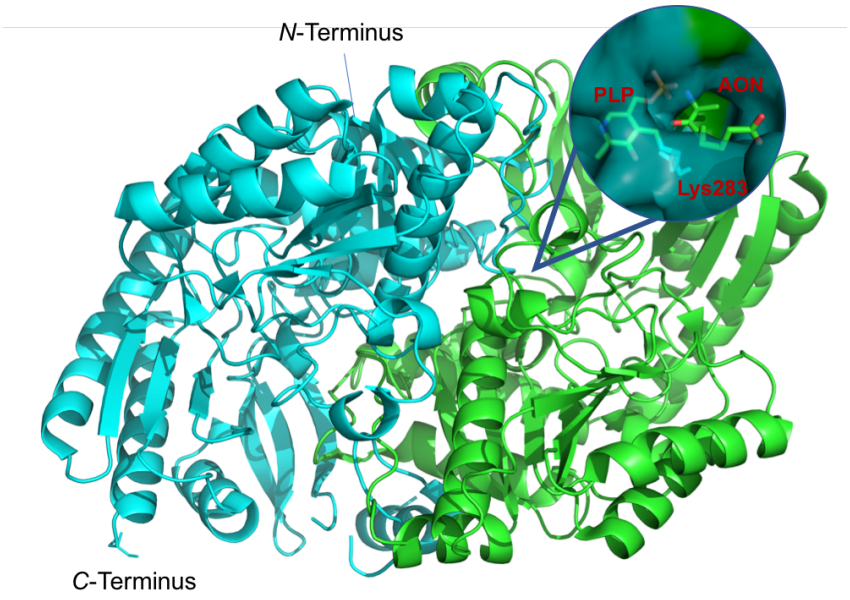
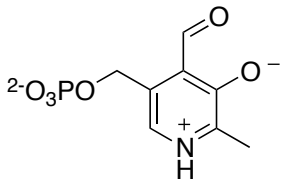


Figure 4.3 - The crystal structure of the *Mtb* BioA (mtBioA) homodimer (PDB: 4CXQ)[51]. The protein was crystallised in the internal aldimine form, the magnification shows the entrance to the active site on the dimer interface with PLP attached to Lys283 via a Schiff base. AON can also be observed, bound at the entrance to the active site.

BioA is a class I, pyridoxal 5'-phosphate (PLP) dependent transaminase that transforms AON into DAN, using *S*-adenosyl-L-methionine (SAM) as an amino donor. In what is known as a *Ping-Pong Bi-Bi mechanism*, SAM initially displaces the internal aldimine between PLP and lysine residue 283 (Lys283), donating an amino group to the cofactor forming pyridoxamine phosphate (PMP), which is sequestered in the active site. PMP then forms an imine with the second substrate, AON, which retains the amino group when it is displaced by Lys283, forming DAN and returning BioA to its original PLP-bound holo form (see fig. 4.5)[149,150]. It is interesting that BioA has evolved high specify for SAM as an amino donor, not only because SAM is most commonly used as a methyl donor, but also most transaminases can utilise a variety of endogenous amino acids as amino donors[51]. Another interesting artefact of this amino donor selection is a large structural difference between the two substrates, AON and SAM, raising the question of how the BioA active site manages to retain high specificity for both substrates. By analysing 3D structures, Dey *et al.* suggested that BioA relies upon active site structural changes to accommodate different states of the catalytic cycle[51,148]. This characteristic has reportedly made *in silico* drug design very complex[151], however fragment-based approaches for finding new inhibitors have been explored[51,152].

4.1.4 PLP

Pyridoxal 5'-phosphate (PLP, vitamin B6) is an enzyme cofactor involved in a large array of biological reactions, including all known transamination reactions as well as many amino acid racemisation reactions, decarboxylations and deaminations[153].



Pyridoxal 5'-phosphate (PLP)

Figure 4.4 - PLP.

This highly versatile ‘electron sink’ covalently binds the substrate and activates it to the enzyme-mediated chemistry, acting as an electrophilic catalyst. The mechanism of the PLP-dependent transamination reaction carried out by mtBioA is shown in fig. 4.5. In the resting state of the catalytic cycle, PLP is covalently bound as an internal aldimine to BioA via a Schiff base with the ε-amino group of active site Lys283.

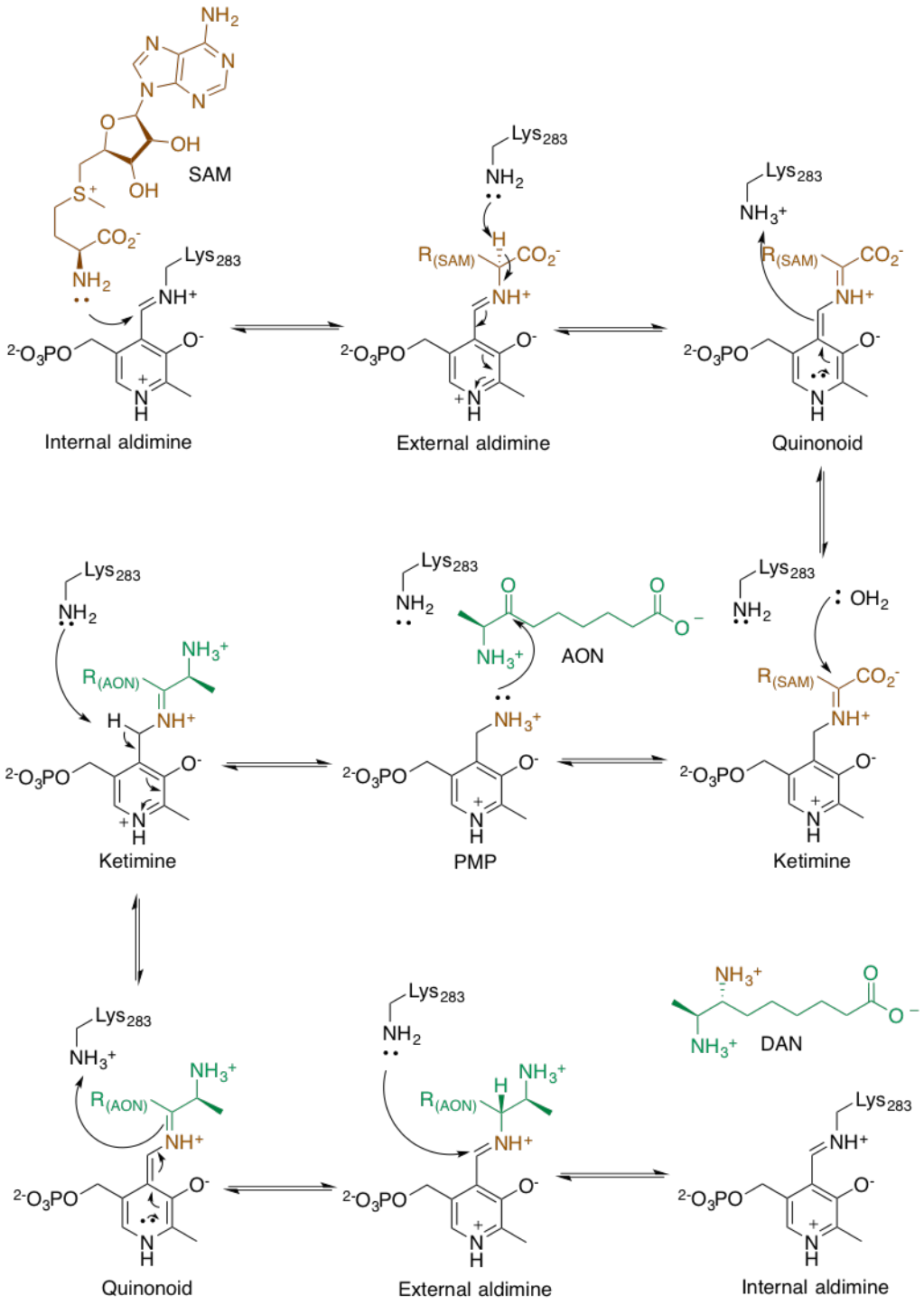


Figure 4.5 - The role of PLP in the catalytic mechanism of mtBioA.

4.1.5 Inhibitors of mtBioA targeting PLP

Since BioA has only gained significant interest as an anti-tuberculosis target in the last 15 years, there are only a few known natural products that inhibit the enzyme. One such example is amiclenomycin (ACM), which acts by forming a covalent adduct with PLP, driven by aromatisation of the cyclohexadiene ring[154]. The compound showed specific activity against *Mtb* among a selection of 40 bacterial strains analysed[155]. Shi *et al.* characterised a reduced version of ACM lacking the primary amine with better whole-cell potency, and developed a mechanism-based BioA inhibitor with better chemical stability based around an isostere of the ACM core scaffold (see Fig. 4.6). Their compound showed comparable activity to ACM, and they proved through whole-cell controlled expression experiments that BioA was indeed the target. The enantiomer of their compound showed no activity, supporting the idea that the BioA active site is highly specific[155].

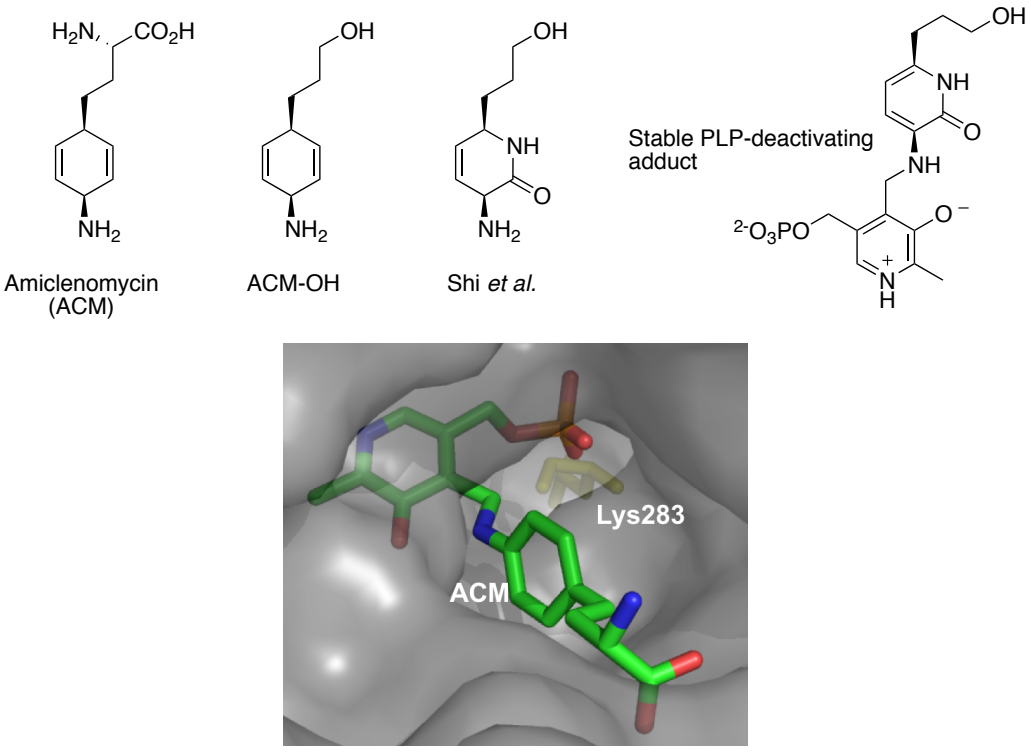


Figure 4.6 - Natural product BioA inhibitor ACM and derivatives, all of which inhibit the enzyme by forming a covalent adduct with the cofactor PLP, as shown in the crystal structure (PDB: 1MLZ)[154].

The knowledge that PLP reacts with hydrazines and hydrazides, deactivating transaminase enzymes has been held for a long time[156]. Recently, Dai *et al.* carried out a differential scanning fluorimetry (DSF) screening campaign on a library of ca. 1000 small molecules to identify compounds that bound to BioA. From this screen, they identified 2-(aminomethyl)-benzothiazole **C1** and confirmed their findings by ^{1}H-STD NMR and X-ray crystallography. They developed a second screen of commercially available compounds structurally similar to their initial hit, identifying other binding compounds (see Fig. 4.7). Whilst many of the hydrazines and hydrazides that they explored appeared to form covalent adducts with PLP, none of the hydrazides had any inhibitory activity. They hypothesised that the hydrazones formed had similar stability to the Schiff base formed between the cofactor and Lys283, so they were readily reversible in the presence of substrates. Despite this, they were able to structurally characterise the covalent PLP-adducts formed by 4 library members (amine **C1**, PDB: 4MQ0; hydrazine **C2**, PDB: 4MQP; hydrazide **C3**, PDB: 4MQQ; hydrazide **C4**, PDB: 4MQR). The binding mode of the PLP motif remains unchanged in all of the structures, and the hydrazine and hydrazide motifs all retain the same binding conformation through the nitrogen-nitrogen double bond and carbonyl functionality (in the case of 4MQQ and 4MQR), however the distal rings of each hit all explore different space within the same binding pocket. Although no potent inhibitors were obtained by the study, it provided vital information on the conformations that inhibitors could take within the active site, serving as a basis for future rational structure-based inhibitor design studies[51,141,152,157].

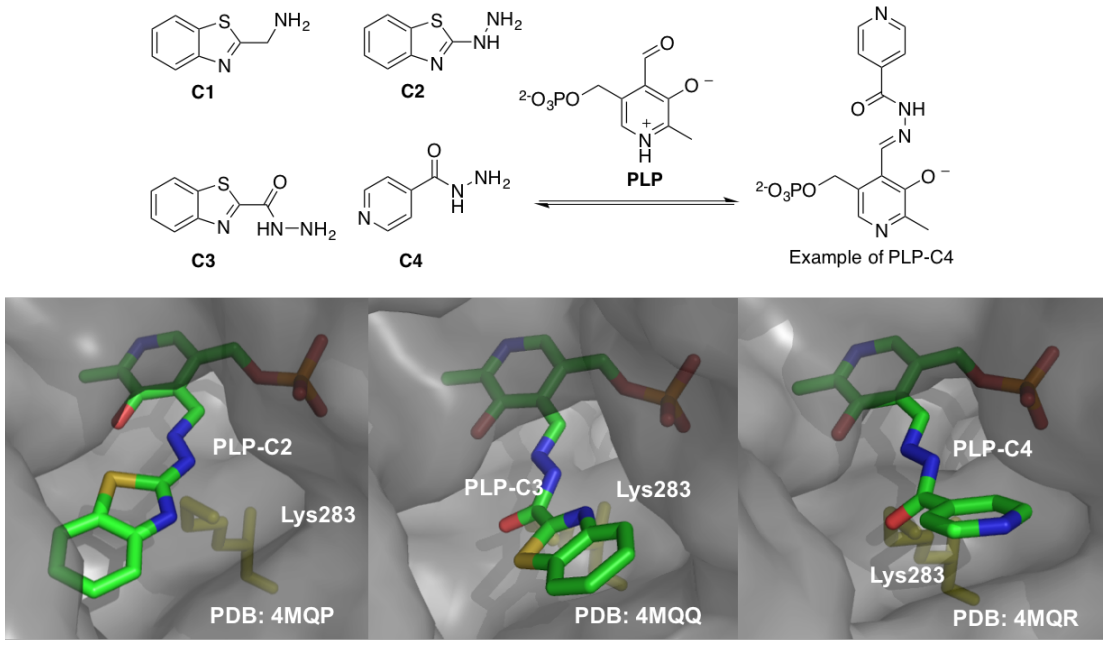


Figure 4.7 - DSF hits from Dai *et al.* with examples of covalent PLP adduct formed with **C4**. The PLP adducts of **C2, C3** and **C4** are shown bound to mtBioA[51].

Using a different approach in a whole-cell phenotypic screening study, Zlitni *et al.* screened their in-house library of ca. 30,000 commercially available small molecules for growth inhibitory activity against *E. coli* MG1655 in a nutrient deficient medium. Metabolic suppression profiles of 74 hits were recorded and hydrazide MAC13772 was identified as an inhibitor of analogous *E. coli* BioA (ecBioA)[158]. A library of 24 analogues of MAC13772 was developed to characterise the structure-activity relationship (SAR) of the molecule. Zlitni *et al.* found that the hydrazide moiety was essential for inhibition, but the ring could be manipulated. In some cases, replacing the ring with an aliphatic chain retained activity although they concluded that the presence of the benzyl ring carried high importance. They were also able to change the substituents on the ring, or replace it with a heterocycle and observe inhibition. Despite the SAR studies the original molecule was the most potent, and through UV-vis spectroscopic analysis they determined that it acted by forming an adduct with PLP (see fig. 4.8). In contrast with the hydrazides proposed by Dai *et al.*, hydrazide MAC13772 demonstrated inhibitory activity with an IC_{50} of 250 nM[158]. This suggests

that hydrazides may not only be suitable for mapping active site interactions, but could also be potent inhibitors of BioA. However, subtle differences between the *Mtb* and *E. coli* isoforms could have large effects on the inhibitory activity of the hydrazides.

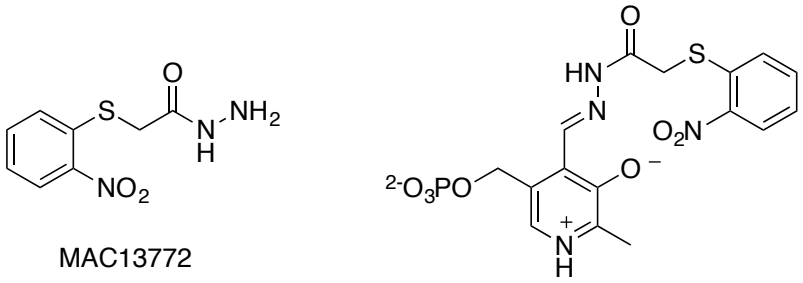


Figure 4.8 - ecBioA inhibitor MAC13772, discovered in a whole-cell phenotypic screen by Zlitni *et al.*, and the structure of the MAC13772-PLP adduct.

There is therefore a strong precedent in the literature that aryl hydrazides could provide a basis for the development of BioA inhibitors. By using diverse hydrazides, it may be possible to further map favourable interactions that lead to strong binding and develop mtBioA selective inhibitors that deactivate the enzyme by forming a covalent adduct with PLP in the active site. The chemistry involved in this interaction makes in amenable to study using DCC, where the enzyme is used as a template to amplify the best binding combination, or combinations, from a pool of dynamically interchanging building blocks.

4.1.6 Mass spectrometry for protein analysis

Electrospray ionisation mass spectrometry (ESI-MS) is a relatively gentle ionisation process that is used to visualise intact proteins. It is favourable over other ionisation techniques that are prone to causing the protein to fragment. In ESI-MS, a high voltage is applied to the protein solution, forming highly charged droplets. As the solvent in each droplet evaporates, the charge density increases, resulting in

coulombic expulsion of yet smaller droplets. The process continues until a fully desolvated gas phase protein remains, which is guided into a mass analyser (see fig. 4.9). The spectrometer then determines the mass to charge ratio (*m/z*) of the protein by various methods, depending on the nature of the machine[54,159].

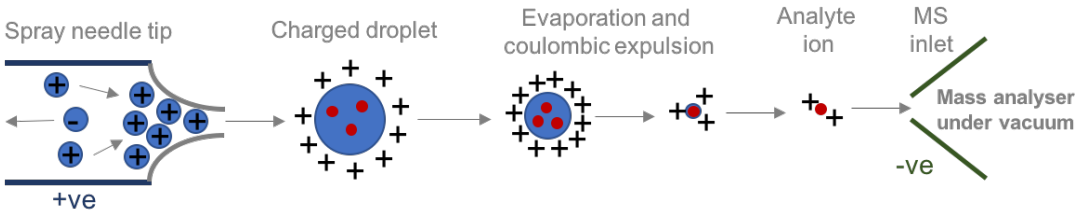


Figure 4.9 - Sample desolvation by ESI-MS in positive ion mode.

To determine the mass of a protein, including any stable covalent adducts and post-translational modifications, the protein is typically dissolved in a 1:1 methanol-water with 0.1% formic or acetic acid. This acidification disrupts the hydrogen bonding maintaining the tertiary structure of the macromolecule, causing the protein to unfold and expose many basic groups that can be protonated, resulting in a wide charge state distribution (see fig. 4.10).

It is also possible to visualise a protein in its folded state under non-denaturing conditions. By incubating the target with a compound, or library of compounds, and then buffer exchanging the complex mixture into a volatile buffer (ammonium acetate or bicarbonate, pH dependent) the protein can be desolvated and analysed by ESI-MS. Once the conditions are optimised, this also allows for the visualisation of ligands that are non-covalently bound to the protein, providing they have a K_d lower than 1 mM[160]. Because the protein remains folded, only a few groups capable of carrying a charge are exposed, resulting in a much narrower charge state distribution than under denaturing conditions. Naturally, this makes the ionisation and desolvation process more difficult, resulting in lower signal strength and mass

resolution. Different ESI sources that operate at lower flow rates have been designed to combat this issue, creating smaller droplets that lead to better signal intensity.

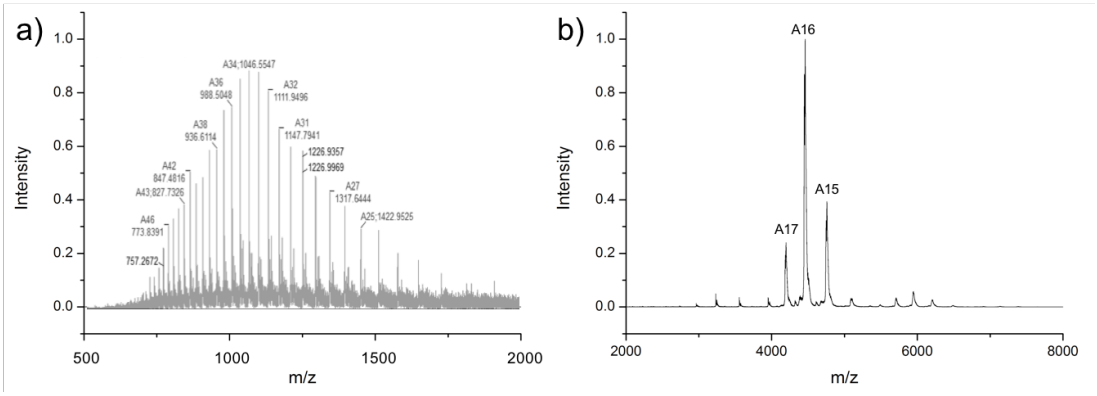


Figure 4.10 - ESI-MS spectra of *E. coli* FabH, **a)** under denaturing conditions, showing charge state distribution of +21 to +47, and **b)** under non-denaturing conditions, showing the +15, +16 and +17 charge states of the protein in its native dimeric state.

Non-denaturing 'native' ESI-MS has been extensively employed to study proteins in complex with substrates, natural products and small molecules[54,160–162]. The huge benefit in using non-denaturing ESI-MS in a screening campaign, is that binding ligands can be immediately identified by the *m/z* increase that they impart on the protein mass. In the case of isobaric ligands, tandem MS experiments can be used to correctly assign binders. A further benefit of this technique is that by adding a compound with a known K_d as an internal calibrant, it is possible to infer the binding constant for ligands that form non-covalent complexes with the target[54,160].

ESI-MS is an elegant, direct-screening technique that can be applied to DCC, negating the need for individual synthesis of library members and leveraging the original promise of DCC to combine library synthesis and screening into a single step.

4.1.7 ESI-MS in DCC

There have been numerous successful examples of using MS to deconvolute a DCL. The first example was from Poulsen, who developed an NAH DCL targeting carbonic anhydrase II (CA)[42]. The library was based around an aromatic sulfonamide that was a known recognition element for CA inhibition. The library was composed of 2 aldehydes, one with the sulfonamide and one without, and 5 hydrazides. After allowing the DCL to equilibrate for 40 h in the absence of any nucleophilic catalyst, the library was infused under non-denaturing conditions and analysed using ESI Fourier-transform mass spectrometry (ESI-FTMS) in negative ion mode. The non-denatured control mass spectrum of CA showed a tight charge state envelope of 9-, 10- and 11- charge states. The CA-templated DCL non-denaturing mass spectrum showed the same charge states, each being composed of a group of signals corresponding to bound *N*-acylhydrazones.

Poulsen assigned the CA-bound *N*-acylhydrazones by tandem mass spectrometry, tuning the collision energy to cause the protein-ligand complexes to dissociate, resulting in a spectrum with both signals corresponding to apo-CA, and to the NAHs that had been ejected. Of the 10 possible *N*-acylhydrazones, only the 5 containing the sulfonamide recognition element were detected, and a CA binding assay confirmed that these had an increased K_I of up to 10x that of the sulfonamide by itself.

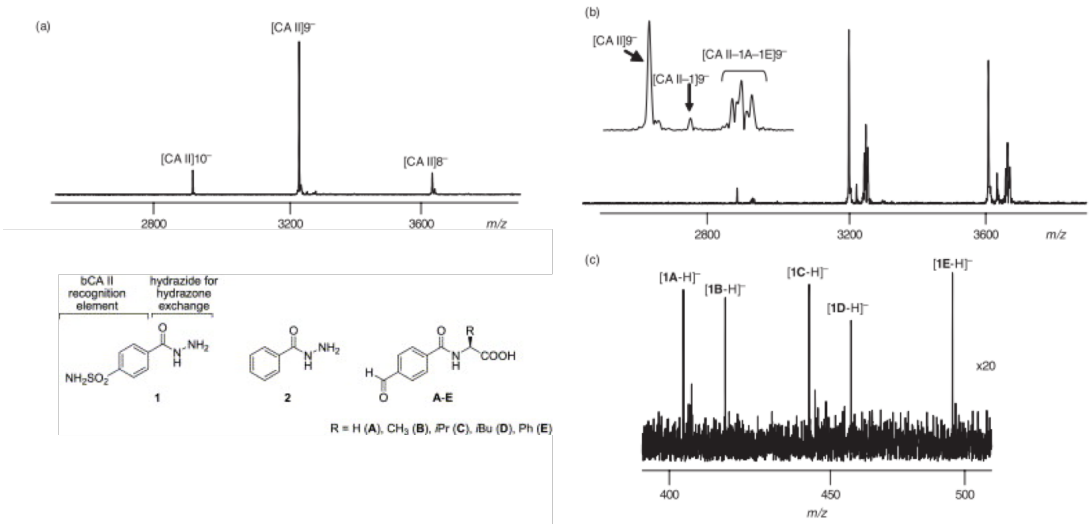


Figure 4.11 - The ESI-FTMS analysis by Poulsen of a DCL targeting CA. **a)** the non-denaturing mass spectrum of CA, **b)** the non-denaturing mass spectrum of the CA-templated DCL, **c)** the *N*-acylhydrazones obtained after collision activated dissociation (CAD).

Further examples came from the Schofield group who used non-denaturing MS to analyse a thiol-disulfide exchange DCL targeting the zinc-dependent enzyme metallo-β-lactamase, extending their DCL from a thiol-containing support ligand that bound to the metal[39]. They then characterised boronate ester exchange as a reaction suitable for protein-templated DCC by NMR, and developed this chemistry into a DCL targeting prolyl hydroxylase domain isoform 2 (PHD2), a 2-oxoglutarate (2OG) oxygenase of therapeutic interest because it regulates the human hypoxic response[40,163]. They successfully used non-denaturing MS, which was coined dynamic combinatorial mass spectrometry (DCMS), to identify potent inhibitors from their library (see fig. 4.12). A caveat that they conceded was that DCMS analysis in the gas phase does not necessarily reflect what happens in solution phase, so results should be validated by solution studies[40,164].

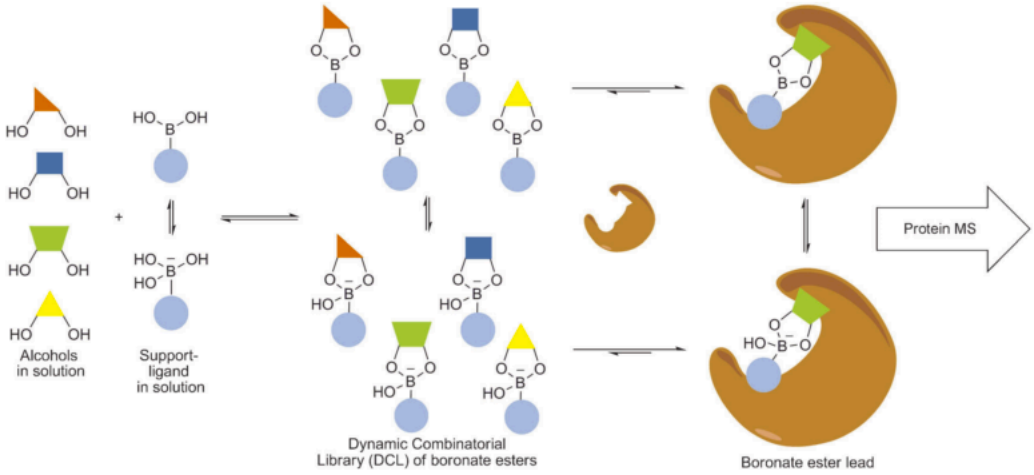


Figure 4.12 - Boronate ester DCL analysed by DCMS[40].

Schofield *et al.* used DCMS to deconvolute an impressive 38-membered thiol-disulfide DCL targeting another 2OG oxygenase from *E. coli*, AlkB[38]. Their DCMS results agreed with solution-based DSF and IC_{50} values, and they obtained crystallographic data of one of their compounds in complex with the target protein. This further validated the technique as a tool for results-driven drug discovery.

An alternative example of using MS to analyse a DCL was proposed by Guo and co-workers who designed a DCL experiment using size exclusion chromatography (SEC) and MS to isolate ligand-target adducts (SEC-MS)[29]. Using the well-studied hen egg-white lysozyme (HEWL), they first confirmed the technique using a known feedback inhibitor *N*-acetyl-D-glucosamine (NAG), with a weak K_d of 20-60 mM[29]. The molecule was incubated with HEWL, and the complex was passed down a SEC column. The eluent protein was denatured with acetonitrile to release any bound molecules, followed by MS analysis. The presence of NAG was confirmed in the analysis, so they developed an imine DCL of 4 amines including glucosamine, and 10 aldehydes. They found that it was necessary to chemically freeze the library with $NaBH_3CN$, as the imines were exchanging and being lost during SEC. The static amine library delivered 3 binding hits in the SEC-MS analysis, validating this technique as a valid approach to isolate ligand-target adducts, providing they are stable and remain bound under SEC

conditions. They further refined the technique by using small buffer-exchange spin columns in place of a SEC column, and applied their approach to more complex systems[165,166].

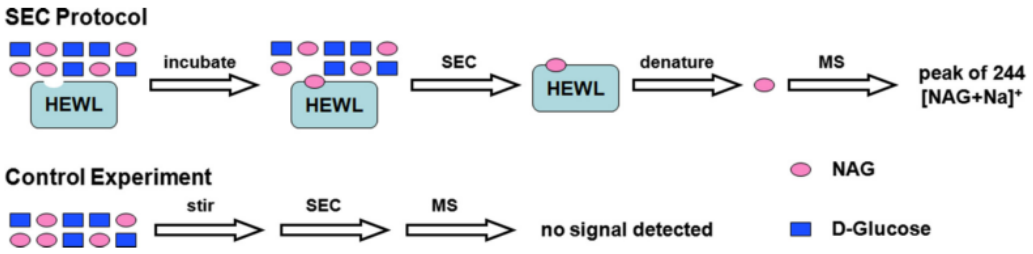


Figure 4.13 - The SEC-MS protocol proposed by Guo and co-workers for identifying ligand-target adducts[29].

The various MS techniques that have been devised to analyse DCC experiments over the last decade suggest a promising outlook for the field. The direct-measurements obtained by MS analyses circumvent the limitations on library size and the requirement to individually synthesise library members that are often imposed by spectral and chromatographic analysis techniques. Further validation of DCL MS analysis will allow DCC to gain lost momentum and take another step towards acceptance as a results-oriented drug discovery technique.

4.1.8 Aims

The aim of this investigation was to demonstrate the power of ESI-MS in analysing larger complex DCLs. Based on literature precedent of hydrazides inhibiting BioA by forming NAHs with the PLP cofactor, we set out to construct a DCL based on NAH exchange between a library of hydrazides and the aldehyde moiety of PLP. Our goal was to discover compounds active against the recently validated anti-mycobacterial target BioA.

4.2 Results and Discussion

4.2.1 Protein expression and purification

The *BioA* gene from *Mtb* was gifted to us by the Aldrich group (University of Minnesota) in a pUC-19 vector with an N-terminal hexahistidine tag. The protein was expressed in BL21(DE3) Rosetta 2 *E. coli* cells using literature conditions with expression under the control of the lac promoter[51]. After the cells were harvested, a two-stage purification was conducted using nickel affinity chromatography followed by SEC. SDS-PAGE confirmed that mtBioA had been overexpressed and purified without issue (see fig. 4.14)

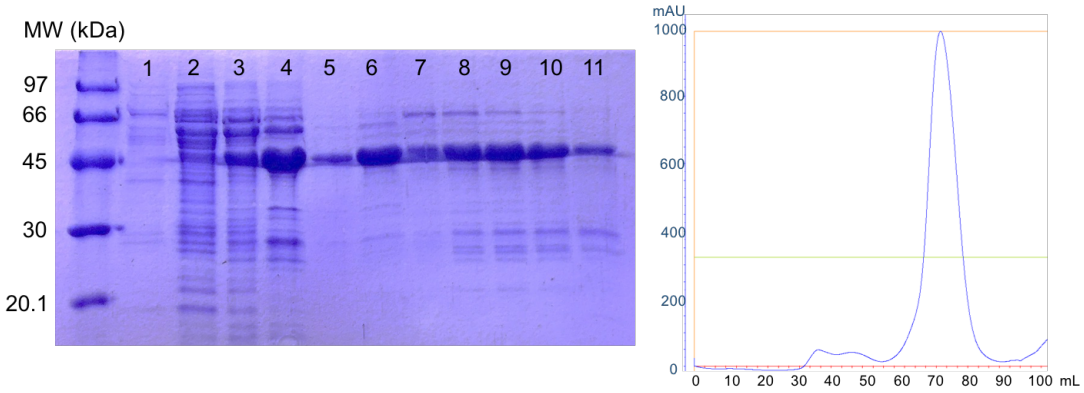


Figure 4.14 - SDS-PAGE of fractions containing mtBioA from nickel affinity chromatography (lanes 1-5) and SEC (lanes 6-11). The SEC chromatogram shows the protein eluting as a single peak at approx. 70 mL.

The protein was characterised by LC-MS under denaturing conditions, giving a mass of 48351.64 ± 0.50 Da, which corresponds to the sequence-predicted mass of 48351.49 Da (ExPASy).

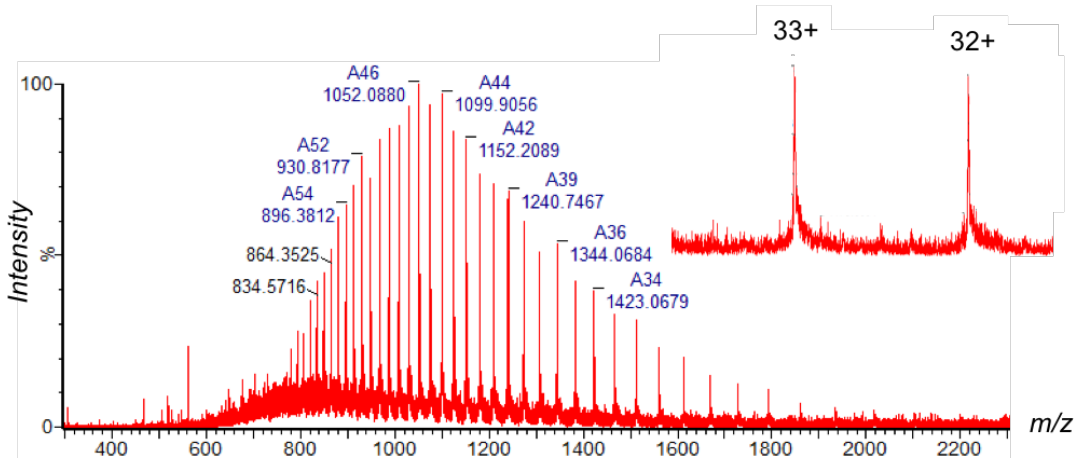


Figure 4.15 - Mass spectrum of mtBioA, with a magnification of the 33+ and 32+ charge states.

A catalytically inactive mutant was constructed from the mtBioA gene, replacing the essential lysine at position 283 with an alanine residue. Multiple attempts at different strategies to introduce a mutation by PCR failed, so a gene fragment was purchased (Genscript) that coded for the desired mutation. The fragment was flanked by Not1 and Age1 restrictions sites intrinsic to the mtBioA gene, but absent from the pUC-19 plasmid. This allowed for a simple restriction digest of the wild type (WT) *mtBioA/pUC-19* construct with the restriction enzymes Not1 and Age1, affording a DNA-strand into which the purchased fragment was ligated.

K283A mtBioA was expressed in BL21(DE3) Gold *E. coli* cells following a poor expression test in the BL21(DE3) Rosetta strain. Aside from the expression strain, the mutant protein was expressed and purified by the same method as the WT enzyme. SDS-PAGE confirmed that K283A mtBioA had been overexpressed and purified without issue (see fig. 4.16)

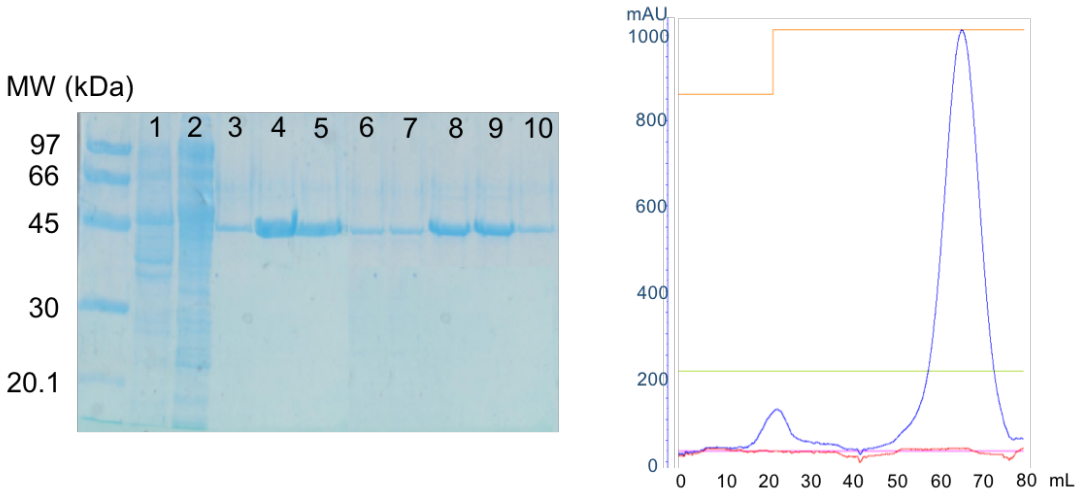


Figure 4.16 - SDS-PAGE of fractions containing ecBioA from nickel affinity chromatography (lanes 3-6) and SEC (lanes 7-10).

The protein was characterised by LC-MS under denaturing conditions, giving a mass of 48295.36 ± 0.43 Da, which corresponds to the sequence-predicted mass of 48294.39 Da (ExPASy).

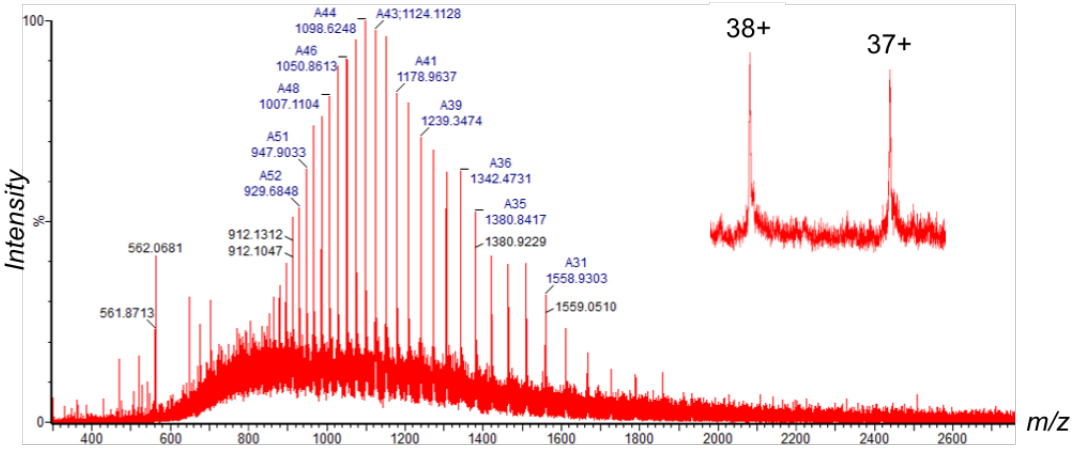


Figure 4.17 - Mass spectrum of K283A mtBioA, with a magnification of the 38+ and 37+ charge states.

The *BioA* gene from *E. coli* was provided by Annabel Serpico (Campopiano group, Edinburgh) in a pET-28a vector with an N-terminal hexahistidine tag. The protein was expressed in BL21(DE3) *E. coli* cells using conditions that had been optimised for ecBioA expression. Protein overexpression was induced through the T7 promoter by the addition of 0.5 mM IPTG followed by agitation at 30 °C for 5 h. After the cells were harvested, a two-stage purification was conducted using nickel affinity chromatography followed by size exclusion chromatography (SEC). SDS-PAGE confirmed that the protein had been overexpressed and purified without issue (see fig. 4.18).

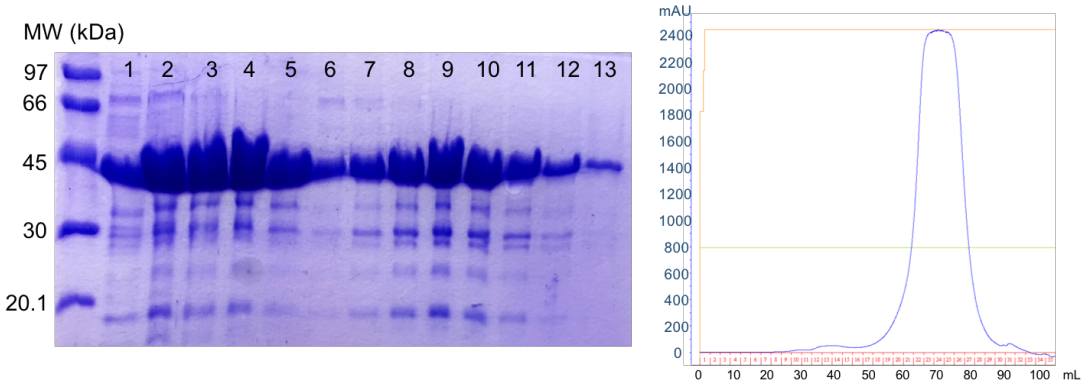


Figure 4.18 - SDS-PAGE of fractions containing ecBioA from nickel affinity chromatography (lanes 1-6) and SEC (lanes 7-13). The SEC chromatogram shows the protein eluting as a single peak at approx. 70 mL.

The protein was characterised by LC-MS under denaturing conditions, giving a mass of 49367.93 ± 0.32 Da, which corresponds to the sequence-predicted mass of 49367.72 Da (ExPASy).

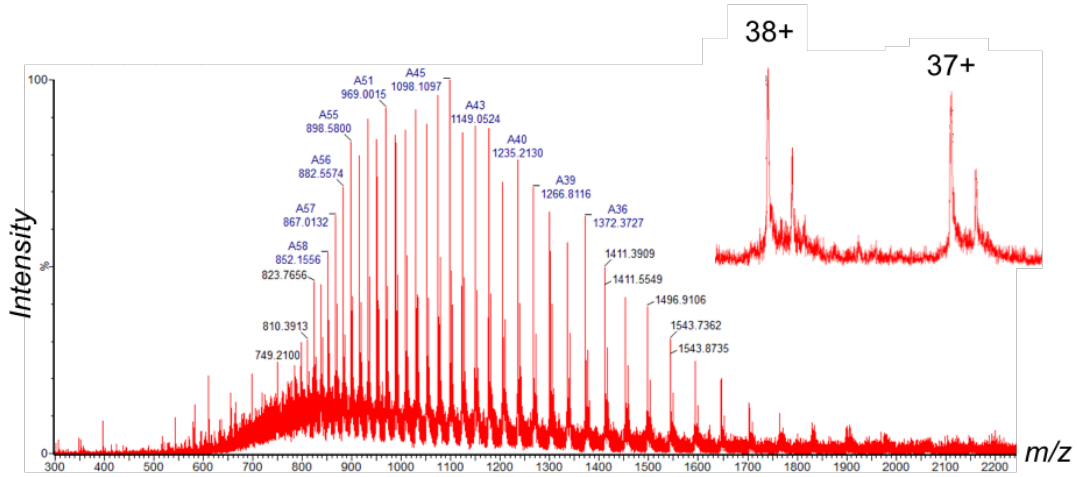


Figure 4.19 - Mass spectrum of ecBioA, with a magnification of the 37+ and 38+ charge states. Note the presence of a second species in each charge state of +176 Da corresponding to *N*-gluconoylation, a common post-translational modification of recombinant hexahistidine-tagged proteins expressed in *E. coli*[131].

4.2.2 Library selection

The elegance of MS as a DCL deconvolution technique is that spectral overlap does not limit the size of the library, providing care is taken to exclude isobaric compounds. As such, a library of hydrazides was assembled from a mix of synthetic and commercially available compounds, along with compounds obtained from the NCI Developmental Therapeutic Program's open compound library (http://dtp.cancer.gov). The hydrazides all had a MW < 250, contained a single hydrazide moiety and no other reactive groups that would interfere with the DCC experiment.

The 5 amino acid-derived hydrazides were synthesised during a Masters research project by Jue Theresa Wang, aiming at increasing the aqueous solubility of subsequent *N*-acylhydrazones. These compounds were included in the library targeting BioA, as the aminotransferase class of enzymes often utilise amino acids as substrates or amino-donors, raising the possibility of advantageous amino acid binding interactions in the enzyme active site. For example, BioA from *Bacillus subtilis*

utilises lysine rather than SAM as amino donor in the transamination reaction[167]. The 28-membered hydrazide library H1-H28 is shown in fig. 4.20.

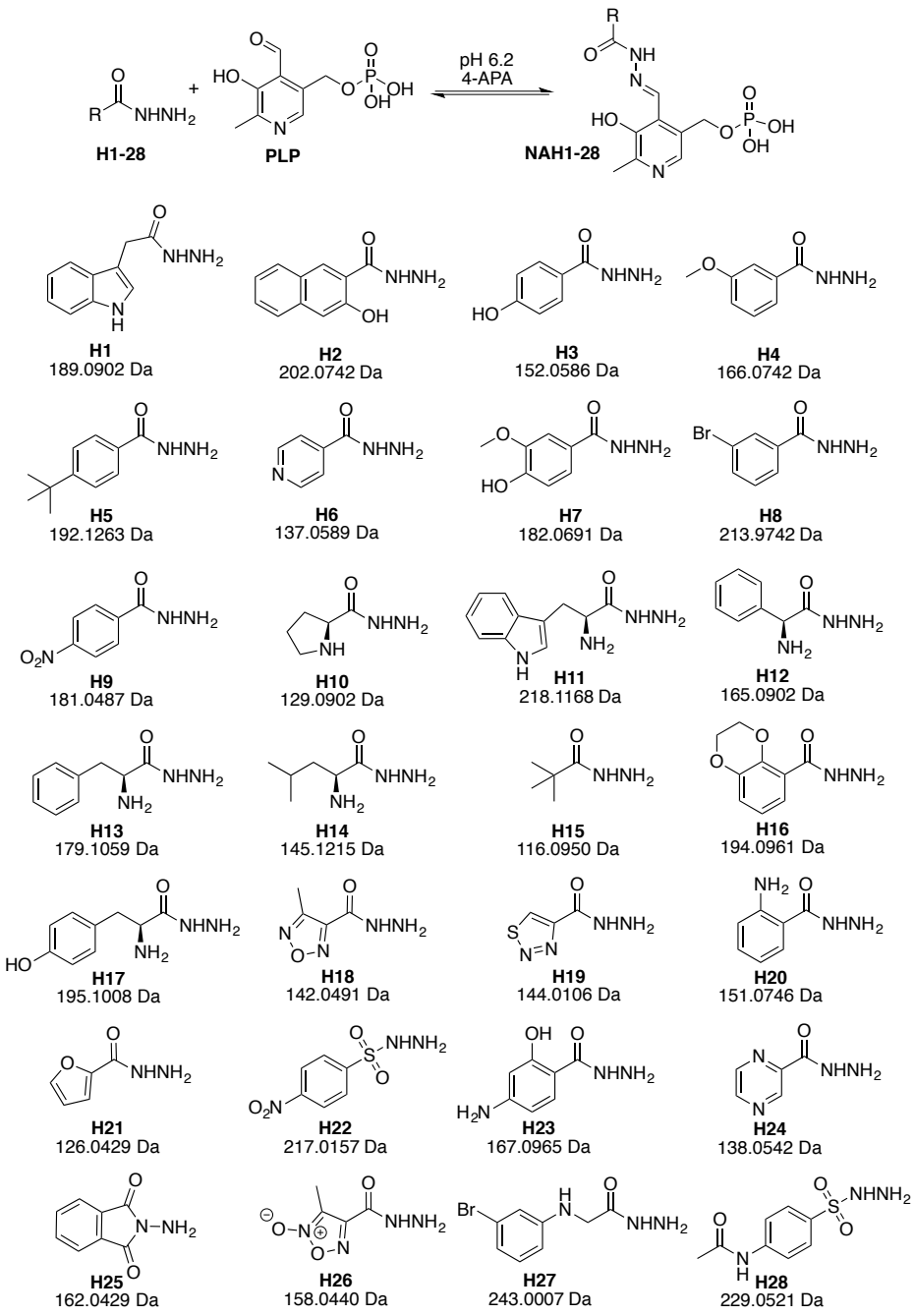


Figure 4.20 - The 28-member hydrazide library with exact mass of each compound. The corresponding PLP-NAHs have an addition of 229.0140 Da.

4.2.3 Library validation

Prior to the DCC experiment, differential scanning fluorimetry (DSF) was used to ensure that all of the assembled library members bound to the enzyme. The technique uses a fluorescent reporter to monitor the state of unfolding of an enzyme and hence determine its melting point. Although traditional inhibitors are expected to stabilise the enzyme and effect a positive shift in melting temperature, compounds that impart destabilising shifts can also be of value. These can promote enzyme degradation and subsequent depletion of the target if the bacterium is in a dormant non-replicating state[51]. Aldrich, Finzel *et al.* have used DSF extensively to investigate binders of mtBioA, showing that the formation of a NAH by a hydrazide displacing PLP from the catalytic lysine results in a destabilising thermal shift[51,152,168]. They report a melting temperature of 86 °C for mtBioA, agreeing with our experimental value of 87 °C (see fig. 4.21). Through altered purification procedures, they also attributed a melting temperature of 68 °C to the *apo* form of the protein, confirming that binding of the PLP-cofactor to the enzyme has a large stabilising influence. Hydrazide hits that resulted in a melting temperature between 68 - 86 °C therefore displaced the PLP from the catalytic lysine, forming a PLP-NAH that remained bound in the active site.

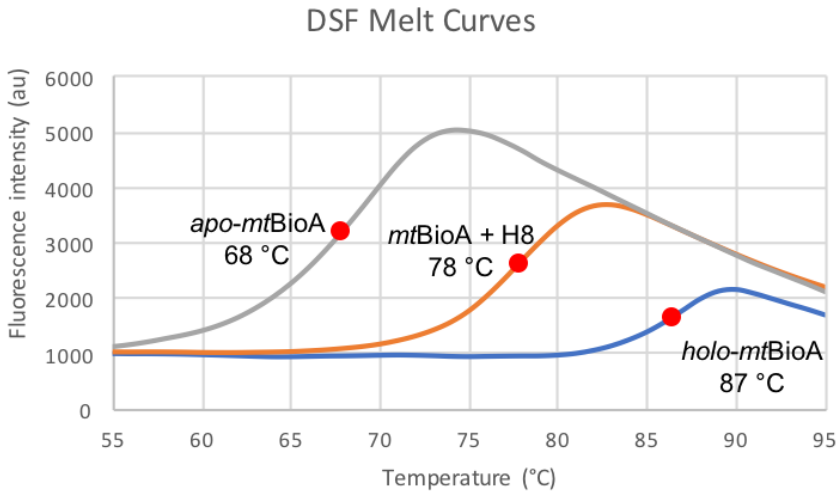


Figure 4.21 - DSF melt curves showing the melting temperatures of *apo*-mtBioA, *holo*-mtBioA, and the destabilising shift of -9 °C caused by binding of hydrazide H8.

We conducted DSF assays with all 28 library members against both mtBioA (*holo* T_m 87 °C, *apo* T_m 68 °C) and ecBioA (*holo* T_m 79 °C, *apo* T_m 65 °C), along with 4-APA to ensure the DCL catalyst did not compete for PLP. Gratifyingly, a destabilising shift was observed for all library members, whilst no shift was observed for 4-APA (see table 4.1). This confirmed that each member of the assembled 28-membered library interacted with the PLP-cofactor. Interestingly, there were distinct variations in the relative melting point shift exerted on each isozyme by many library members. It has been shown through X-ray crystallography that when a hydrazide is introduced the enzyme it forms a PLP-NAH which remains sequestered in the active site, likely due to the strong interactions between PLP and the active site residues that stabilise the external aldimine stage of the BioA catalytic cycle[51]. The polarity of the binding pocket differs between ecBioA and mtBioA, hence a given PLP-NAH may be more destabilising to one isozyme than the other (see fig. 4.22). This pronounced destabilisation is likely due to translocation of PLP and the positioning of the binding pocket at the dimer interface.

Compound	ΔT_m mtBioA	ΔT_m ecBioA
Holo control	0	0
Apo control	-19	-14
H1	-6	-2
H2	-10	-4
H3	-7	-6
H4	-9	-4
H5	-10	-8
H6	-7	-3
H7	-8	-5
H8	-9	-4
H9	-5	+1
H10	-11	-10
H11	-12	-9
H12	-12	-10
H13	-11	-9
H14	-18	-15
H15	-4	-5
H16	-4	-3
H17	-17	-5
H18	-6	-7
H19	-7	-5
H20	-10	-8
H21	-7	-7
H22	-20	-18
H23	-9	-9
H24	-5	-4
H25	-6	-3
H26	0	-1
H27	-3	+1
H28	-18	-8
4-APA	0	0

Table 4.1 - Showing the shift in the melting temperatures of ecBioA and mtBioA caused by members of the 28-member hydrazide library.

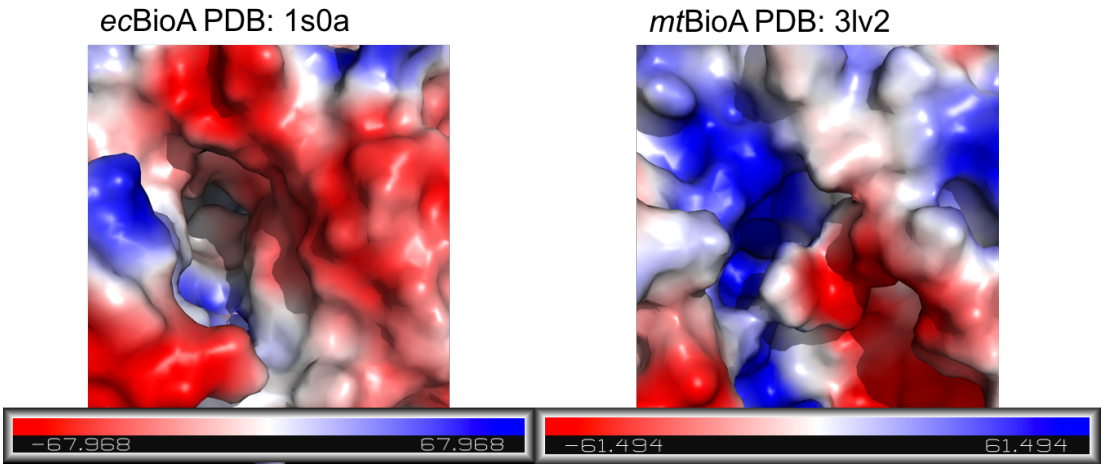


Figure 4.22 – Electrostatic potential maps generated in PyMOL from PDB structures of ecBioA and mtBioA. The substrate binding pocket of ecBioA is relatively neutral in character, compared the binding pocket of mtBioA, which is flanked by positively charged amino acids (units of calculated electrostatic potential K_bT/e_c).

4.2.4 DCC experiment

Initial attempts at monitoring the DCC reaction followed the DCMS protocol, where the target protein is interfaced with the DCL and the mixture is infused under non-denaturing conditions into the spectrometer, before the collision-induced dissociation (CID) energy is tuned to cause bound ligands to dissociate (see section 4.1.7). If done correctly, this results in signals corresponding to the *apo*-protein, and signals in the low mass range corresponding to the previously bound small molecules. Upon attempting this approach with both mtBioA and ecBioA we were unable to achieve sufficient charge state resolution, despite tuning the spectrometer and employing multiple desalting steps to ensure that the protein was in purely volatile buffer with minimal sodium adducts (see fig. 4.23). The signal intensity and resolution were further compromised when CID was applied.

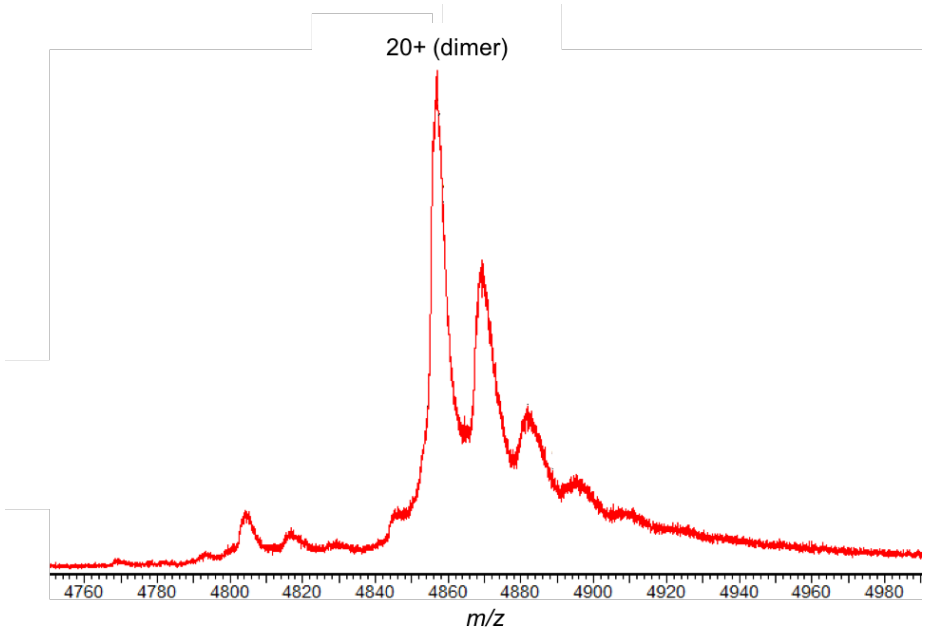


Figure 4.23 - The dimeric 20+ charge state of the non-denaturing mass spectrum of mtBioA. Note breadth of each peak and the series of unidentified adducts, each bringing an addition of ca. 250 Da.

Therefore, we modified the SEC-MS technique proposed by Fang *et al.*, utilising a buffer exchange step to remove unbound DCL members after the library had been incubated with the protein, whilst the bound DCL members remained associated with the protein in a volatile buffer. Infusing the protein ligand complex under denaturing conditions allowed us to directly search for masses corresponding to the NAH DCL products. To begin, BioA was supplemented with 4 molar equivalents of fresh PLP and left for 30 mins at room temperature to ensure that the enzyme was fully loaded in its *holo*-form. A buffer exchange step to remove the excess unbound PLP was carried out using a Bio-Spin P-6 (Bio-Rad) column, which facilitates a buffer exchange of over 99.9%. The BioA-templated 28-member DCL was then assembled with the enzyme and each hydrazide at 30 μM, such that the maximum combined concentration of NAH products was 30 μM. The reaction was buffered with sodium phosphate at pH 6.2 in the presence of 3 mM 4-APA. The 100-fold excess of 4-APA had previously shown to be sufficient to equilibrate both the forward and reverse

reactions of NAH exchange within 2 h. The reaction was incubated overnight at 25 °C with gentle shaking, followed by a P-6 buffer exchange into 50 mM ammonium bicarbonate. This served to both remove any of the hydrazide members that remained unbound, and exchange the protein ligand complex into a volatile buffer system suitable for ESI-MS infusion. Prior to infusion, the sample was diluted to a final protein concentration of 10 μM using a 60% methanol in 0.1% formic acid solution. This caused the protein to denature, releasing any bound NAH DCL products, while also kinetically freezing the reaction by hindering hydrolysis at low pH. All MS data were calibrated using a single-point calibration with the calculated mass of **NAH1** (419.11148 Da).

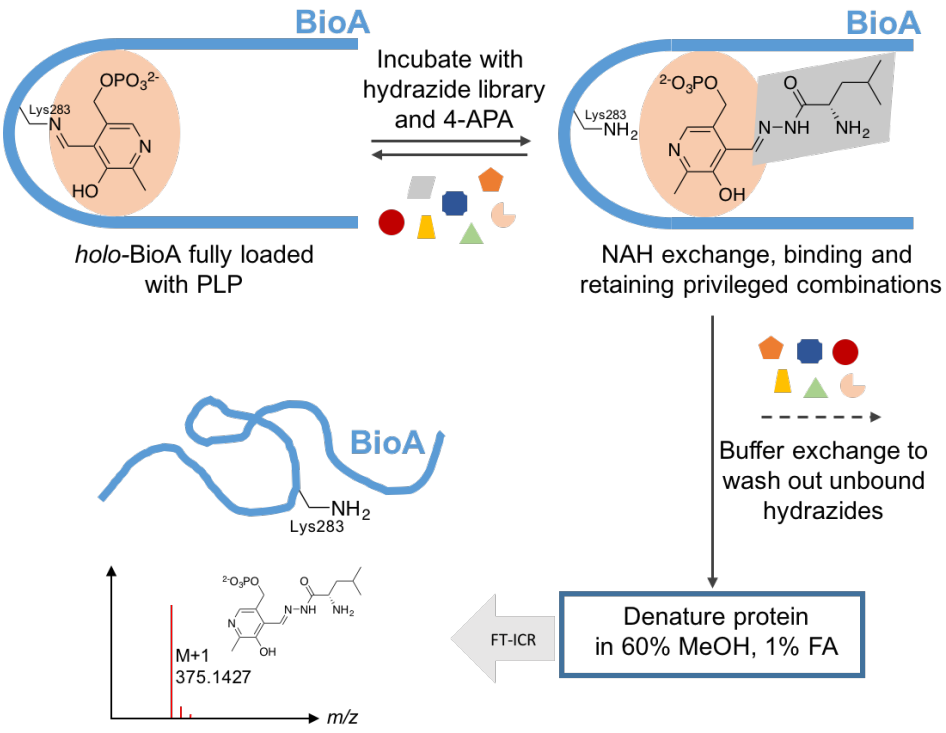


Figure 4.24 - The DCC experiment, using buffer exchange followed by MS to determine the best binders.

4.2.5 Control experiments

The first important control for this method was to ensure that unbound DCL members were entirely removed during the buffer exchange step (see fig. 4.25). To investigate this, PLP was chemically reduced onto the catalytic Lys283 residue using $NaBH_4$, preventing the PLP from taking part in the DCC reaction. To achieve this, an aliquot of BioA was supplemented with a 4-fold molar excess of fresh PLP and left for 30 mins at room temperature. To this aliquot was added 2 mM $NaBH_4$, which was accompanied by an immediate colour change from yellow to colourless, indicating that the reduction had taken place. This was confirmed by LCMS, showing a clear mass increase of +229 Da (see fig. 4.26). A buffer exchange step removed excess unbound PLP and $NaBH_4$, and the method control protein sample was subject to the same protocol as the DCC reaction. Gratifyingly, not a single hydrazide or NAH DCL product was identified in the method control samples from both ecBioA and mtBioA (see fig. 4.27).

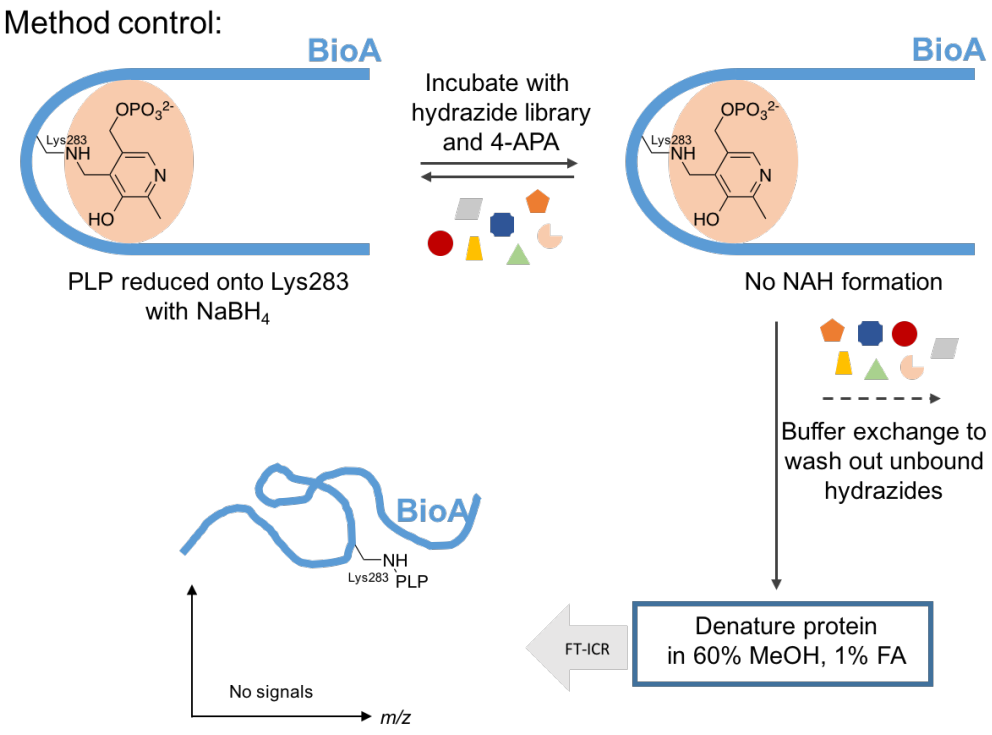


Figure 4.25 - The method control experiment demonstrated that unbound species are entirely removed by the buffer exchange step.

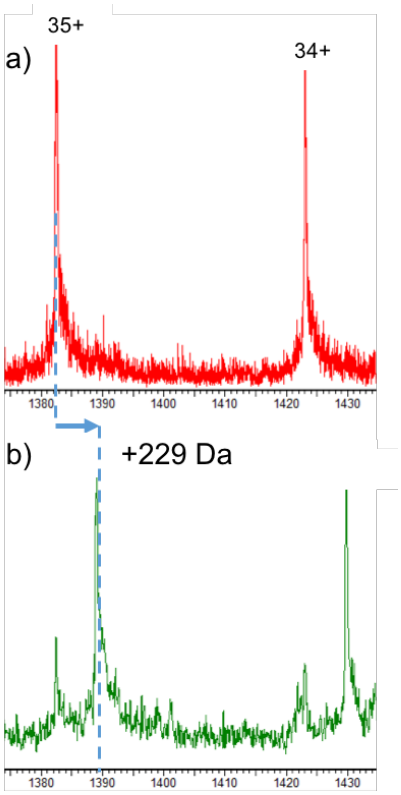


Figure 4.26 - The ESI-MS spectra under denaturing conditions of a) the 36+ and 37+ charge states of mtBioA, and b) the same charge states of the method control sample showing a mass change of +229 Da, corresponding to reduction of the PLP-internal aldimine.

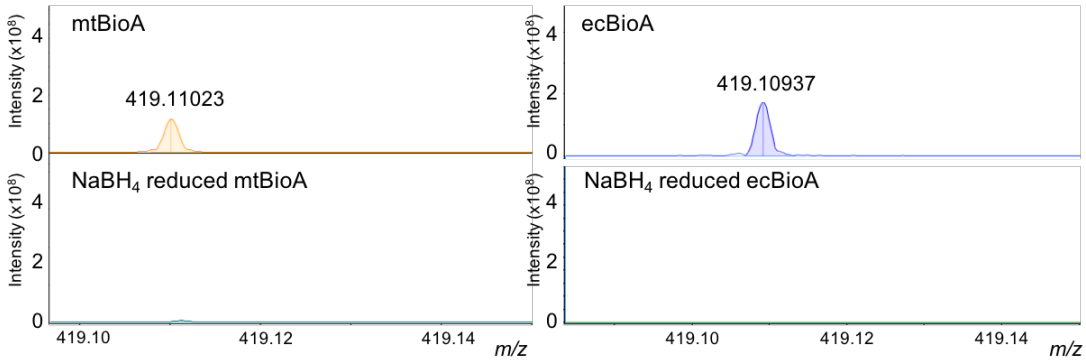


Figure 4.27 - Zoomed mass spectra showing NAH1 present in the analysis of the protein-templated DCL (above) and absent in the analysis of the $NaBH_4$ reduced protein-templated DCL (below).

A second essential control was to ensure that all of the NAH DCL products were forming under the conditions used for the DCL experiment (see fig. 4.28). To demonstrate this, an un-templated 28-member DCL was assembled with each hydrazide and PLP at 30 μM, such that the maximum combined concentration of NAH products was 30 μM to mirror the conditions of the BioA-templated experiment. The reaction was buffered with sodium phosphate at pH 6.2 in the presence of 3 mM 4-APA, and allowed to equilibrate at 25 °C overnight with gentle shaking. Because the NAH DCL products would not bind to a P-6 column for buffer exchange, a different approach was required. The reaction was successfully exchanged into a volatile buffer system using a C18 zip-tip (Millipore). This bound the NAH DCL products but not the unbound hydrazides, which were eluted. The NAH DCL products were then eluted with a 60% methanol in 0.1% formic acid solution at a concentration corresponding to 15 μM total PLP NAH product (derived from the 30 μM PLP initial concentration).

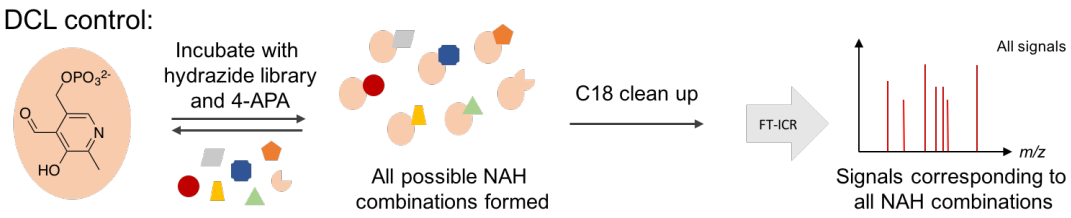


Figure 4.28 - The DCL control experiment demonstrated that all 28 NAH products are identified by the MS method from an un-templated DCL.

Gratifyingly, all PLP NAH products were identified in the DCL control. Observed masses were only assigned if they were within 2 ppm of the predicted mass and a signal to noise ratio greater than 4:1. Due to the low analyte concentration (approx. 0.5 μM per NAH), the signals did not dominate the mass spectrum (see fig. 4.29), and therefore the data were extracted and plotted in a bar chart as a function of relative intensity assigned to each NAH (see fig. 4.30)

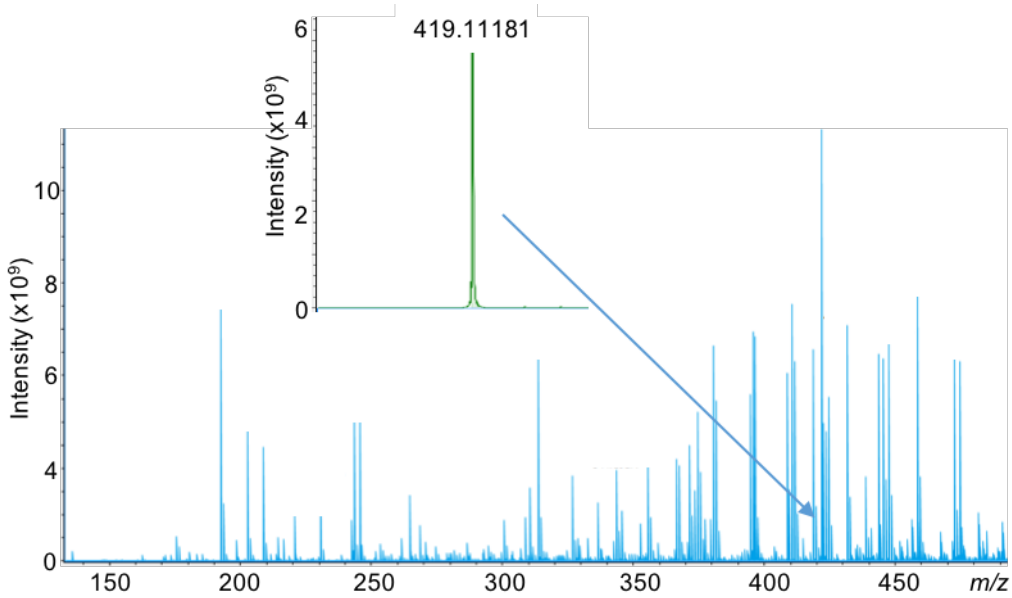


Figure 4.29 - The ESI-MS spectrum of the blank DCL, showing the signal corresponding to NAH1.

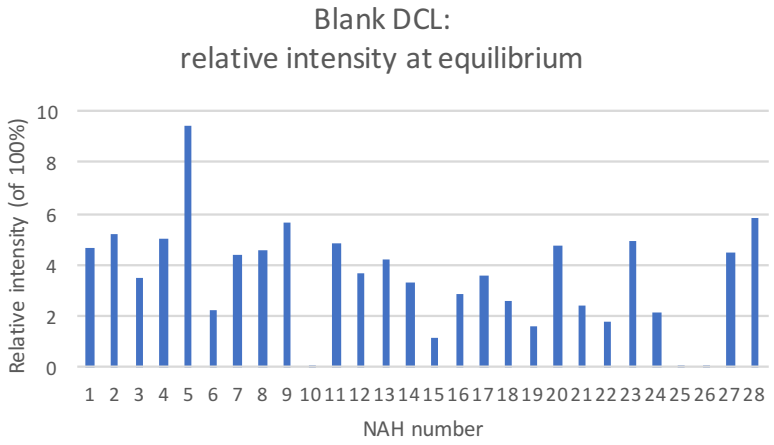


Figure 4.30 - The relative signal intensities of un-templated DCL products at equilibrium. A mass corresponding to each member was detected, although those of H10, H25, and H26 are significantly weaker than others.

To correct for the fact that each compound has a different ionisation efficiency under the ESI-MS condition due to the presence of different chemical motifs, a second blank

DCL was constructed in the presence of excess PLP. This meant that each hydrazide at 30 μM would react with the excess PLP to form an equimolar amount of NAH. This provided a signal intensity for each NAH at equal maximum concentration (see fig. 4.31) which could be used as a reference to replot fig. 4.30 as a function of population distribution at blank equilibrium (see fig. 4.32). Although the concentration of NAHs 1-28 is not equal at equilibrium, they are all within a reasonably narrow range. There do not appear to be any structural features that would explain the preference towards **NAH2**, **11** and **23**.

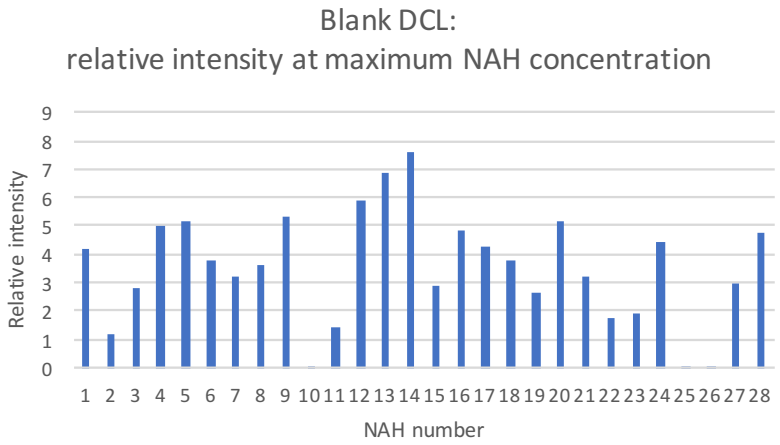


Figure 4.31 - The relative intensity of each NAH formed in an excess of PLP, providing an intensity reference at equal NAH concentration.

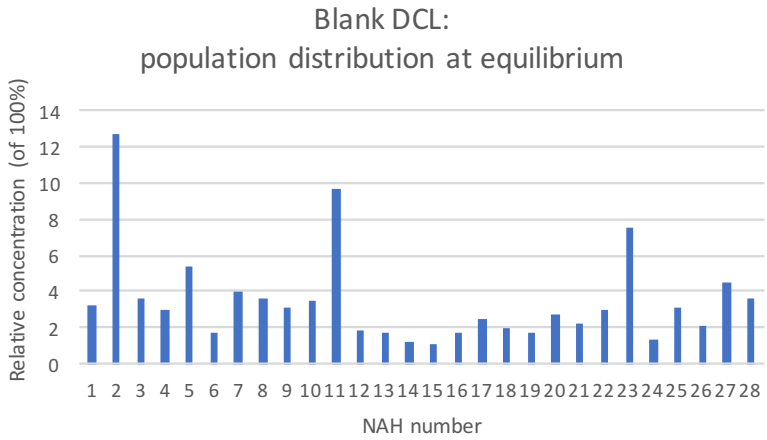


Figure 4.32 - The blank DCL from fig. 4.30, replotted as a function of population distribution.

4.2.6 Results of protein-templated DCC experiments

Relative signal intensities of DCL products from the protein-templated DCC experiments were converted to *% total product population*, correcting for ionisation efficiency based on the relative intensities in the excess PLP blank DCL control. This further allowed for each NAH to be represented in a chart showing *percentage amplification* relative to the blank DCL control. Observed masses were only assigned if they were within 2 ppm of the predicted mass and a signal to noise ratio greater than 4:1. Mass list and corresponding intensities of all DCC experiments are tabulated in Table 4.2, and discussed separately below.

NAH number	Calc. Mass	Equimolar NAH (rel. intensity)	Blank DCL (rel. intensity)	mtBioA (rel. intensity)	K283A mtBioA (rel. intensity)	ecBioA (rel. intensity)	ecBioA reverse (rel. intensity)
1	419.11148	4.15	4.65	8.69	31.69	20.96	12.10
2	432.09548	1.21	5.21	24.38	3.00	14.07	24.41
3	382.07988	2.80	3.51	0.00	5.64	0.44	0.00
4	396.09548	4.97	5.05	3.87	0.00	1.55	1.59
5	422.14758	5.17	9.47	0.00	5.05	0.34	1.01
6	367.08018	3.79	2.26	0.00	0.00	1.21	0.78
7	412.09038	3.25	4.42	10.29	8.88	0.00	0.00
8	443.99548	3.62	4.53	0.00	4.01	1.90	2.07
9	411.06998	5.34	5.67	0.00	3.31	13.93	13.94
10	359.11148	0.02	0.02	0.00	0.00	0.00	0.00
11	448.13808	1.46	4.82	16.61	5.31	0.98	1.73
12	395.11148	5.88	3.70	3.49	1.40	0.60	0.70
13	409.12718	6.86	4.16	2.96	0.96	0.24	0.41
14	375.14278	7.62	3.28	2.11	1.18	0.22	0.37
15	346.11628	2.88	1.11	0.00	0.00	0.27	0.28
16	424.09038	4.82	2.85	10.44	4.17	3.97	2.54
17	425.12208	4.24	3.60	0.00	1.10	0.11	0.35
18	372.07038	3.77	2.55	0.00	0.00	0.38	0.44
19	374.03188	2.62	1.56	0.00	0.00	0.16	0.00
20	381.09588	5.17	4.76	0.00	0.00	0.00	0.00
21	356.06418	3.22	2.42	0.00	2.94	0.23	0.34
22	447.03698	1.75	1.80	0.00	0.00	1.08	3.51
23	397.09078	1.94	4.96	0.00	0.00	0.00	0.00
24	368.07548	4.42	2.10	0.00	1.45	0.26	0.20
25	392.06418	0.07	0.07	0.00	0.00	0.00	0.00
26	388.06528	0.04	0.03	0.00	0.00	0.00	0.00
27	473.02198	2.95	4.48	17.16	15.32	11.31	12.77
28	459.07338	4.76	5.86	0.00	4.60	25.78	20.48

Table 4.2 - The relative intensities of compounds identified in each analysis. The sum of all assigned intensities in each experiment is 100. See below for discussion of each.

4.2.6.1 mtBioA

In the mtBioA-templated DCL, 10 out of a possible 28 compounds were 'amplified' by the enzyme. Of these 10 compounds, there are no obvious structural features that universally qualify their selection. mtBioA appears to favour both larger hydrazides, and those with a primary amine functionality. Hits also appear to have some correlation with results from the DSF screen. NAHs that were amplified all caused a negative ΔT_m of mtBioA between 5-12 °C (see table. 4.1). While the T_m shift indicates that the PLP is likely to have been displaced from the active site lysine, it should not be assumed that the greater the negative ΔT_m, the better the binder. In fact, it is probably the case that there is a window of ΔT_m where the hydrazide forms an NAH with PLP, but remains well stabilised in the active site. This agrees with the observed trend that the concentration of selected DCL members is inversely proportionate to the ΔT_m.

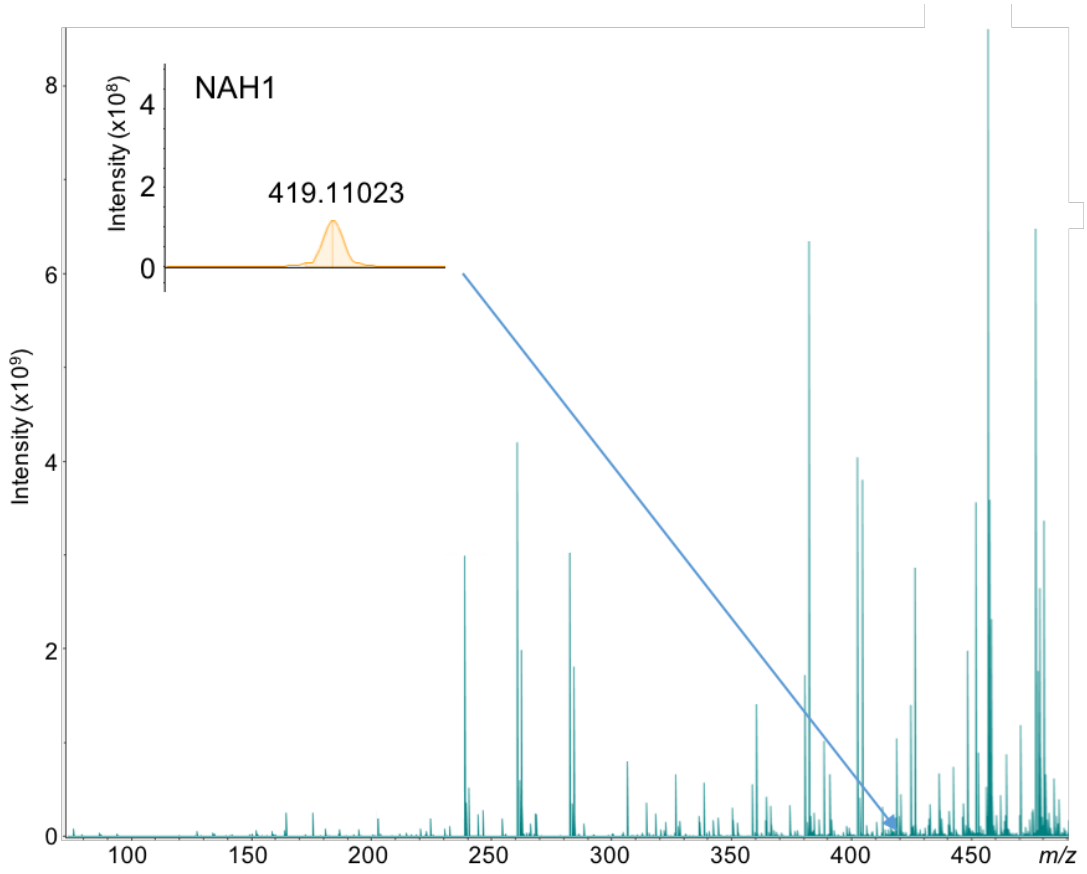


Figure 4.33 - Mass spectrum analysis of the mtBioA templated DCL. Because of the low analyte concentration, signals corresponding to DCL members have an intensity almost two orders of magnitude lower than contaminants in the infusion.

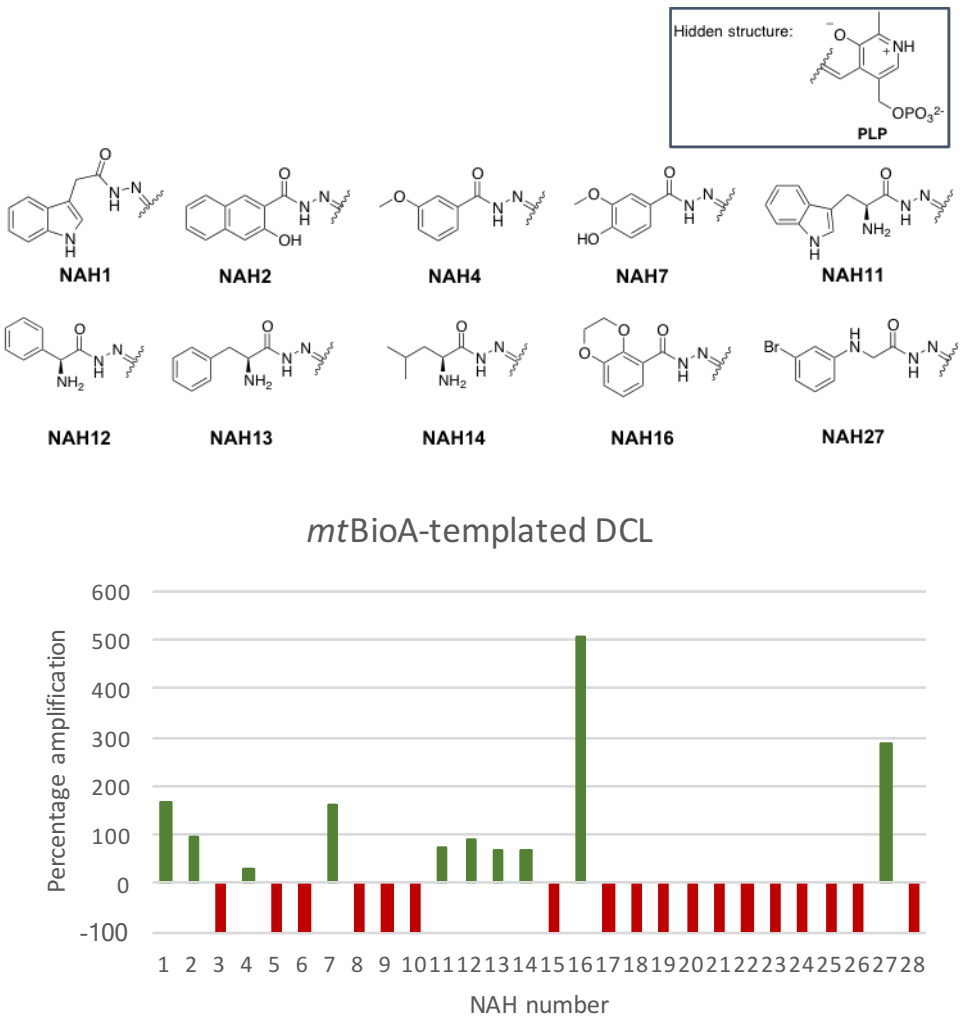


Figure 4.34 - The results of the mtBioA-templated DCC experiment.

To determine whether the preferential selection of hydrazides **11-14** resulted from favourable binding interactions with the amino acid derived primary amine functionality, the compounds were crystallised with mtBioA following literature crystallisation conditions[51]. Despite obtaining crystals from each co-crystallisation well, X-ray diffraction data was only obtained for **H11** in complex with mtBioA. The structure confirms that **H11** forms an NAH by displacing the PLP internal aldimine. The PLP ring is bound in the same position as in structures published by Aldrich and co-workers (see fig. 35), and the NAH functionality remains fixed between 4MQQ, 4MQR and our structure. This is likely due to the hydrogen bonding interactions

between the double-bonded nitrogen and the protonated ε-amino group of Lys283, and between the oxygen of Thr318 and the hydrogen of the NAH (see fig. 4.36). No other polar contacts were identified to justify the importance of the NAH carbonyl or the primary amine group. The indole ring appeared to stack between residues Phe402 and Tyr25, however the indole ring sits approximately 4.5 Å from each residue, therefore outside the π-π stacking distance range of 3.3-3.8 Å. There may also be edge-to-face stacking with Tyr157. The indole ring explores further down the binding channel than either of the compounds in 4MQQ and 4MQR (see fig. 4.35), identifying an area with potential for exploitation in future studies.

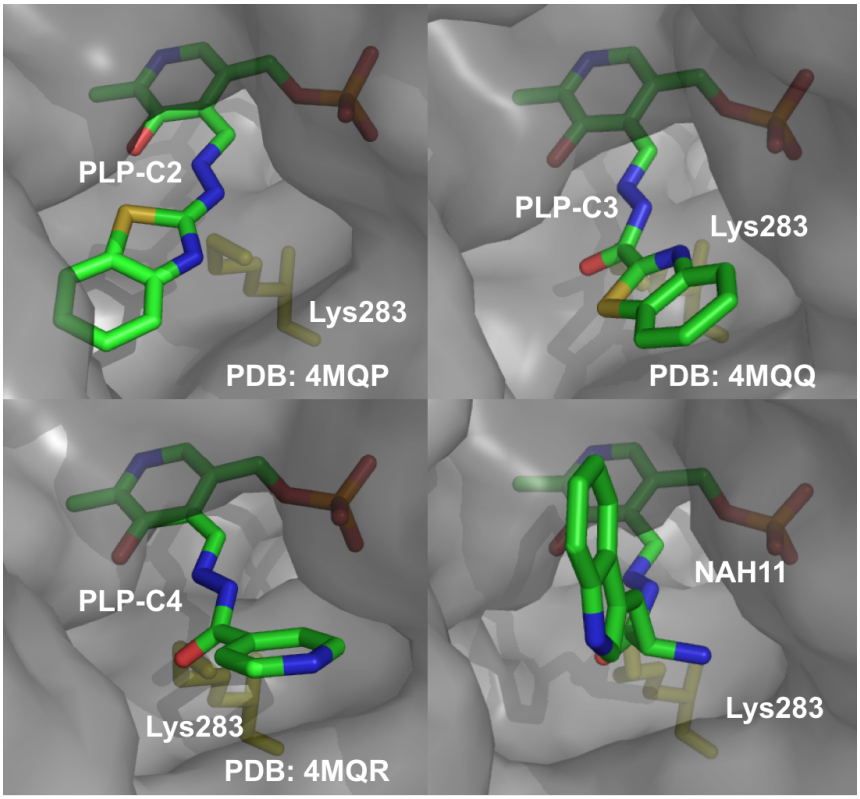


Figure 4.35 - The X-ray crystal structures from Aldrich and co-workers in comparison with the structure of NAH11 bound to mtBioA.

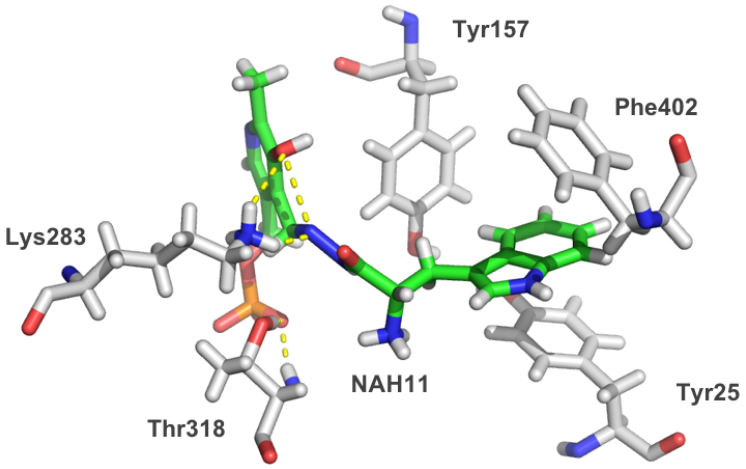


Figure 4.36 - The X-ray crystal structure of NAH11 bound to mtBioA, showing the polar interactions with residues Lys283 and Thr318.

4.2.6.2 K283A control

To confirm that the hits from the mtBioA-templated DCL were due to a perturbation of the dynamic equilibrium rather than a kinetic result of enzymatic activity, we templated the same DCL with catalytically inactive mutant K283A. The K283A mtBioA templated DCL appears to broadly amplify the same compounds as the WT-templated library. There are some differences, for example the relative amplifications of **H1** and **H16** have been swapped, and the concentrations of **H11-H14** appear to have remained static rather than be amplified. These differences may be due to the removal of Lys283, which as shown in fig. 4.36 stabilises the bound NAH though hydrogen bonding with the PLP hydroxyl and the NAH functionality itself. This experiment shows that the hits are due to thermodynamic selection, rather than a kinetic result of enzymatic activity.

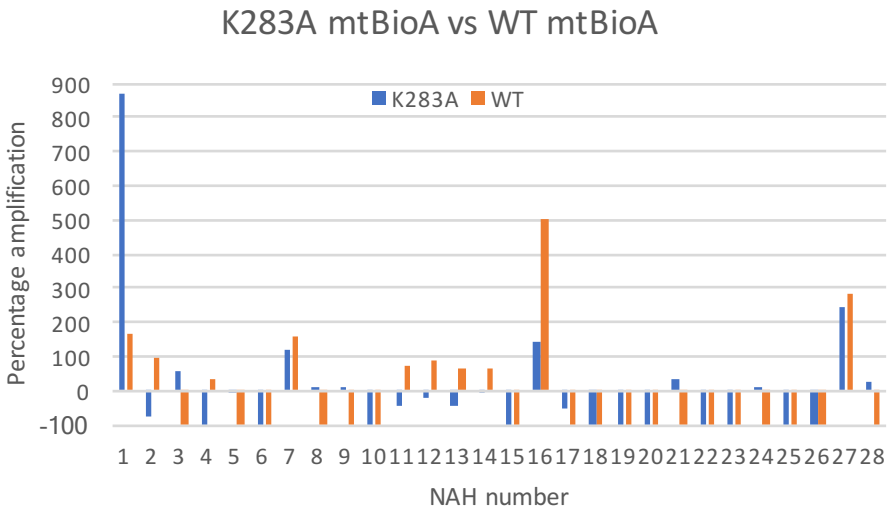


Figure 4.37 - Comparing the difference in result between DCLs templated by WT mtBioA and K283A mtBioA.

4.2.6.3 ecBioA

In the ecBioA-templated DCL, 6 out of a possible 28 compounds were amplified by the enzyme. In general, the ecBioA hits appear to have less steric bulk than those amplified by mtBioA, although there are no obvious structural features that qualify their selection. There also appears to be no correlation between destabilising ΔT_m and selection from the templated DCL.

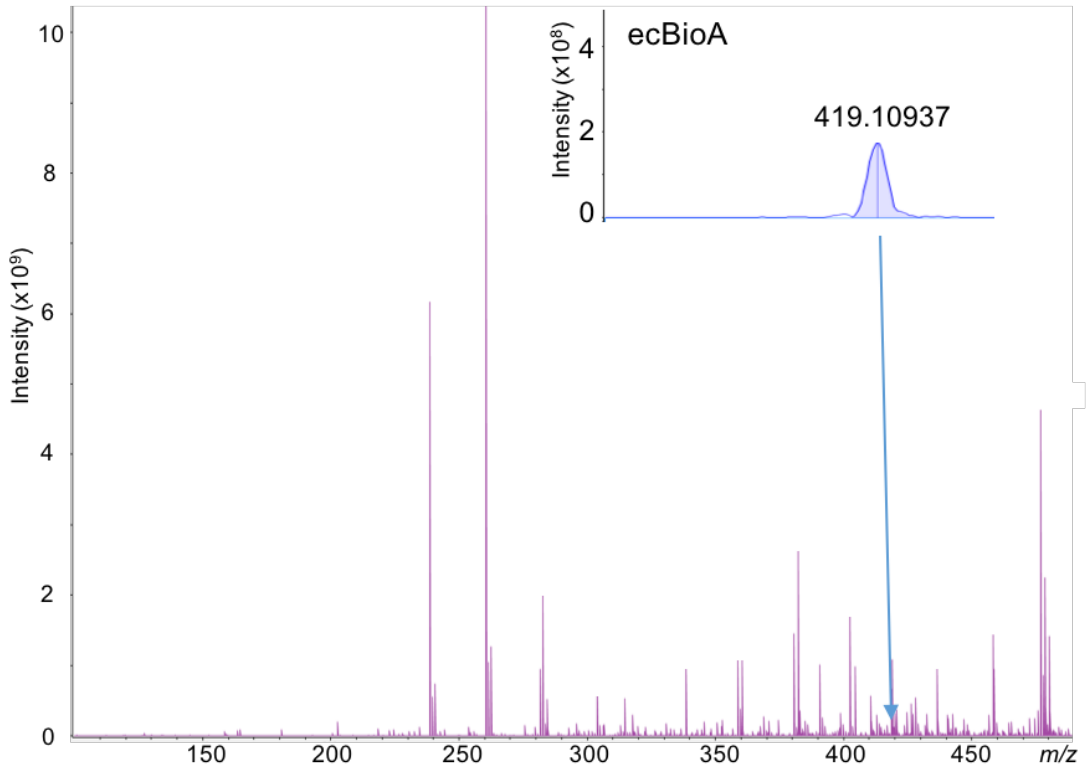


Figure 4.38 - Mass spectrum analysis of the mtBioA templated DCL. Because of the low analyte concentration, signals corresponding to DCL members have an intensity almost two orders of magnitude lower than contaminants in the infusion.

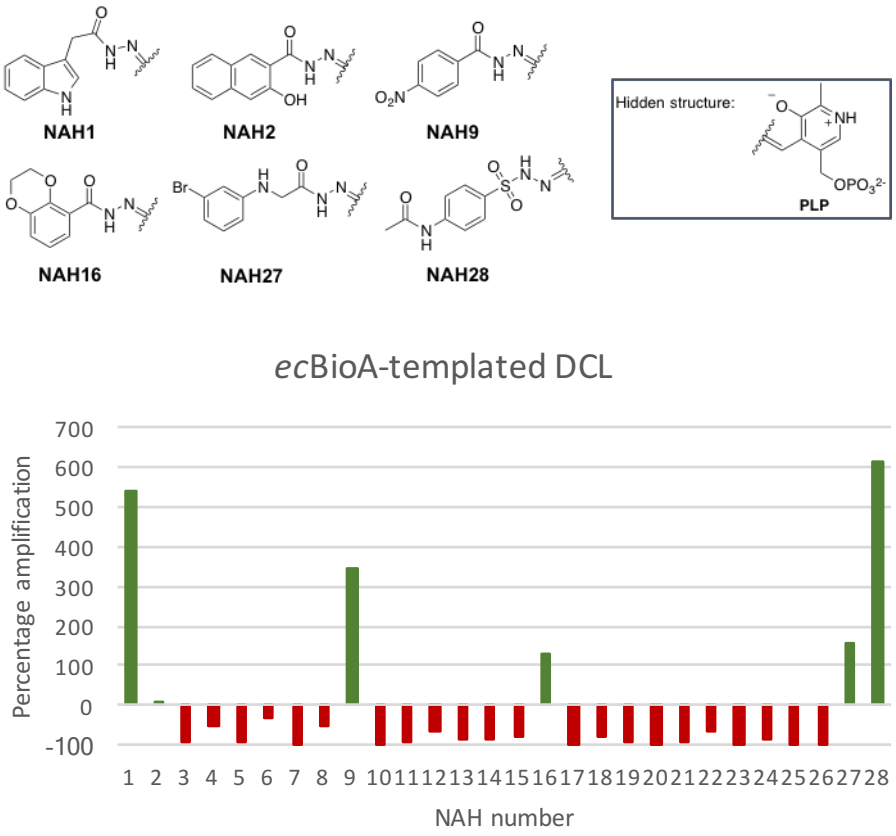


Figure 4.39 - The results of the ecBioA-templated DCC experiment plotted alongside the amplitude of the negative thermal shift caused by each NAH.

There is a striking difference between the DCC results templated by mtBioA and ecBioA. This is not unexpected and is also seen in the DSF screen, possibly due to significant differences between the electrostatic potential of the binding pockets in mtBioA and ecBioA as discussed previously, although there is no significant variation in the architecture of the binding pocket (see fig. 4.22).

4.2.6.4 ecBioA reversibility control

As a final control to demonstrate the reversibility of NAH exchange in equilibrating DCLs, two ecBioA-templated DCLs were set up in parallel. Both were assembled with fully PLP-loaded ecBioA in the same manner as the prior experiments. The first library was set up with 28 hydrazides **H1-H28**. The second was assembled with the single

hydrazide **H1**. Both were allowed to equilibrate overnight with PLP in the presence of ecBioA, and a 50 μL aliquot of each was worked up the following morning, as in previous experiments. Only the single hydrazide **NAH1** was observed in the 1-member library. The 1-member library was augmented with the remaining hydrazides **H2-H28** and allowed to re-equilibrate in the presence of ecBioA overnight. Fig. 4.40 shows that the ecBioA-templated DCLs show the same product distribution, regardless whether the DCL was composed of PLP and 28 hydrazides, or **NAH1** and 27 hydrazides. This experiment validated the reversibility and error-checking ability of PLP-NAH exchange.

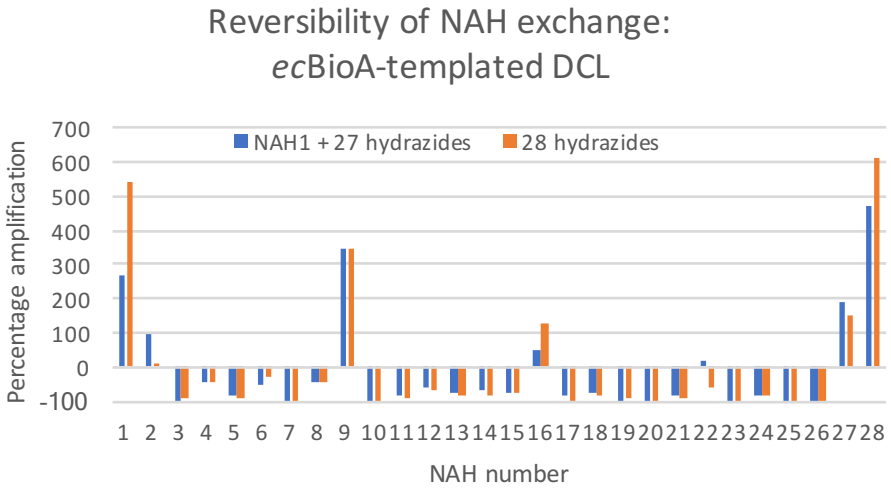


Figure 4.40 - Validating NAH formation reversibility by comparing ecBioA-templated DCLs constructed from **H1-H28** or **NAH1** and **H2-28**.

4.2.7 *In vivo* assays

We sought to implement the hits from our DCL experiments by conducting *in vivo* assays. A series of compounds were characterised for *in vivo* activity in collaboration with the Brown group (McMaster, Hamilton). Compounds **H11**, **H12**, **H13**, **H14**, and **H17** were selected as a closely related set of amino acid-derived compounds that had either been selected by one of the isozymes in the DCC experiment, or shown promising DSF thermal shifts. The compounds were initially tested for *in vivo* activity

by disc diffusion assay against the opportunistic pathogen *E. coli*, and *ESKAPE* pathogens *S. aureus*, *A. baumannii*, and *P. aeruginosa*.

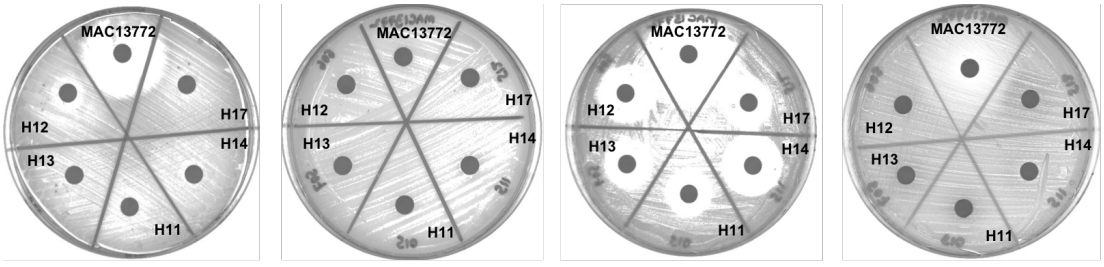


Figure 4.41 - *In vivo* disc diffusion assays of compounds **H11**, **H12**, **H13**, **H14** and **H17** against opportunistic pathogens. MAC 13772 was used as a positive control[158].

The disc diffusion plates in fig. 4.41 show that the tested compounds have no activity against *S. aureus* or *P. aeruginosa*. A weak halo of growth inhibition is seen around all compound-impregnated discs, most markedly for compound **H12**. However, the activity is low compared to the positive control MAC13772. The entire compound set showed comparable activity to the positive control against *A. baumannii*, prompting us to run minimum inhibitory concentration (MIC) assays against *E. coli*, *A. baumannii* and *M. smegmatis*. *M*. *smegmatis* was used in place of *M. tuberculosis*, a common laboratory mycobacterial substitute due to its relative ease of handling and genetic similarity[169].

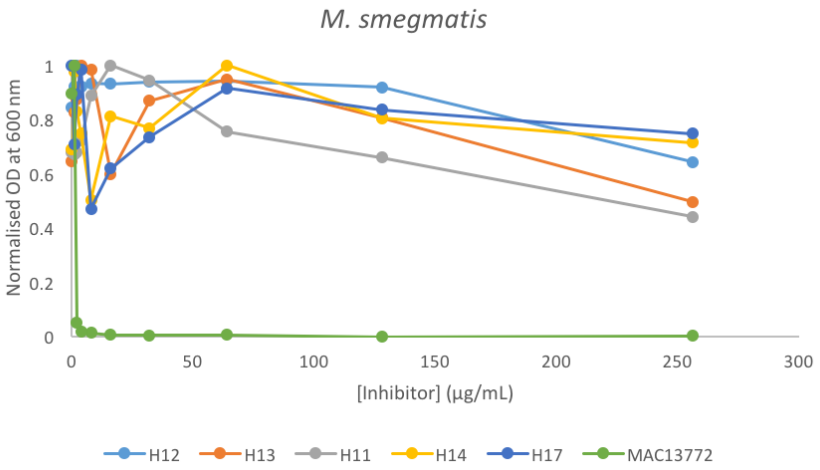


Figure 4.42 - Chart showing the MIC assay data for the compound set against *M. smegmatis*.

Disappointingly, no real activity was observed against *M. smegmatis* by any of the tested compounds, compared to the MIC of MAC13772 (1 µg/mL). Mycobacteria are notoriously difficult to target due to their complex and poorly understood membrane systems. If the compounds succeed in entering the cell, their amino acid derived structure may allow them to be removed by a number of amino acid transporters. The fact that isoniazid and MAC13772 both show excellent activity against *Mtb* suggests that the acylhydrazone moiety does not prevent the compounds from entering the cell.

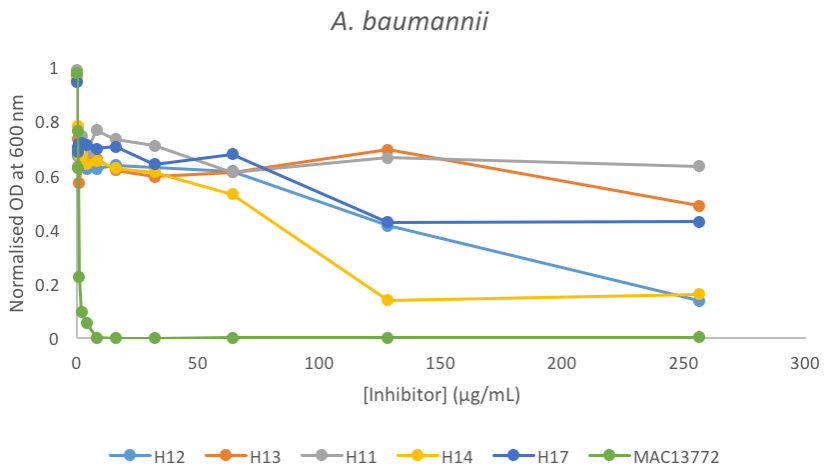


Figure 4.43 - Chart showing the MIC assay data for the compound set against *A. baumannii.*

Although no MIC values could be obtained for *A. baumannii*, compounds **H12**, **H14**, and **H17** caused 50% inhibition of growth at concentrations higher than 125 µg/mL. Compounds **H12** and **H14** showed the highest activity, but were still weak compared to MAC13772 (MIC 8 µg/mL). No structural data exists for *A. baumannii* BioA, so a homology model was constructed using the *Phyre*[2] server in intensive mode[170]. The alignment of the *A. baumannii* BioA sequence with structures of mtBioA (PDB: 3BV0, 3LV2) and ecBioA (PDB: 1S0A) yielded a *highly reliable* model with 99% modelled at a confidence of >90%. While there is slight variation in the structure and calculated electrostatic potential of the entrance to the binding pocket between *A. baumannii* BioA, ecBioA and mtBioA, the active site is structurally almost identical between the three isoforms. This suggests that compounds that bind to ecBioA and mtBioA may bind to the *A. baumannii* protein in a similar manner.

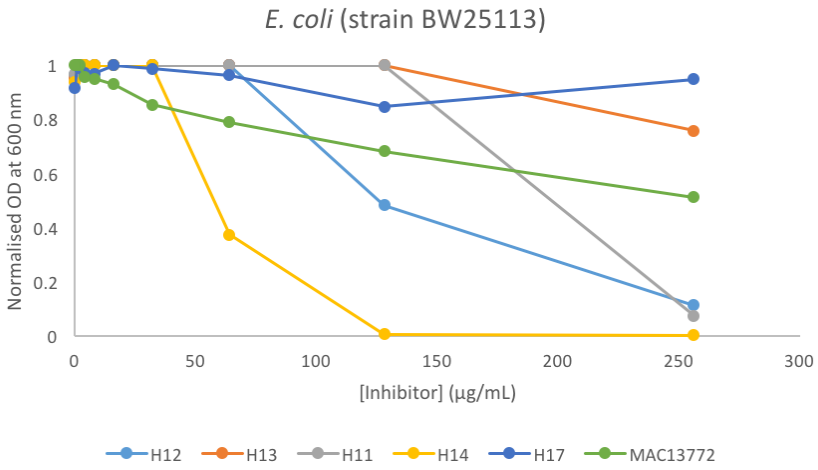


Figure 4.44 - Chart showing the MIC assay data for the compound set against *E. coli*.

An MIC of 125 µg/mL can be ascribed to **H14** against *E. coli*. Although no other compounds had total inhibitory activity, **H12** caused 50% inhibition of growth at concentrations higher than 125 µg/mL. These are the two compounds that showed strongest activity against *A. baumannii* and were selected in the DCC experiments against mtBioA, confirming our prediction that they were likely to be able to enter and interact with critical binding site residues common across BioA isoforms.

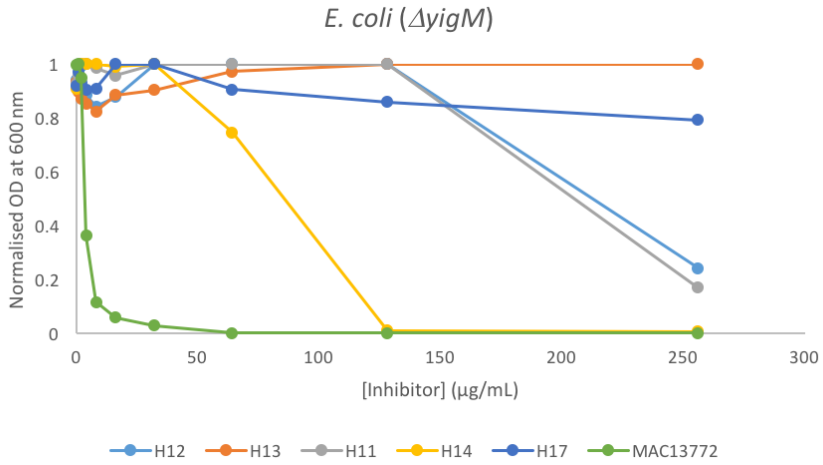


Figure 4.45 - Chart showing the MIC assay data for the compound set against *E. coli* biotin transporter knockout Δ*yigM*.

Neither *E. coli*, *M. tuberculosis* or *A. baumannii* contain the gene coding for the biotin transport protein BioY, and were therefore thought to be dependent on *de novo* biotin biosynthesis[143,145,171]. Recently another biotin transporter protein, YigM, was identified in *E. coli*[172]. Although homologues exist in other bacteria, none were identified in *M. tuberculosis* or *A. baumannii*. Our compound set was therefore also assayed against the biotin auxotrophic *E. coli* knockout mutant Δ*yigM*, against which MAC13772 is known to have a strong efficacy (see fig. 4.45). Because the bacterium can no longer import exogenous biotin, the efficacy of MAC13772 is greatly increased[158]. The lack of biotin transporter in the Δ*yigM* strain appears to have little effect on the potency of our compound set, with **H14** still displaying an MIC of 125 µg/mL. This result suggests that the *in vivo* activity of **H14** may not be entirely as a result of inhibiting BioA, and the compound may interact with other PLP-dependent enzymes or intervene in other cellular processes by hitherto undiscussed mechanisms.

4.2.8 Enzyme assays

At the time of submission, coupled enzymatic assays following a literature protocol[51] had not been attempted due to difficulties in obtaining active BioB. Further efforts will be made in collaboration with the Brown group (McMaster, Hamilton).

4.3 Conclusion

The aim of this project was to demonstrate the power of ESI-MS in facilitating the analysis of larger libraries in DCC. The DCL-size-limiting implications resulting from the need for spectral or chromatographic resolution have restricted the use of DCC as a results-driven drug discovery technique. The use of ESI-MS allows for much larger library sizes. Here we have used the aminotransferase BioA, a validated antibacterial target, to demonstrate the use of ESI-MS for analysing a DCC experiment with protein targets that are not amenable to infusion under non-denaturing conditions.

We demonstrated that each of our 28-member DCL interacted with both mtBioA and ecBioA through DSF experiments. Because of the proposed mechanism of interaction of stripping the PLP from the active site lysine and forming an NAH, which remained sequestered in the active site, each DCL member effected a destabilising thermal shift on the protein melting temperature. While destabilising thermal shifts are still of interest in drug discovery, their correlation to binding strength is not as clear cut as the more traditional approach of identifying fragments that exert stabilising thermal shifts. For this reason, we should consider the relative values of the destabilising shifts with caution, and use the results mainly as a confirmation that a PLP-NAH is being formed.

We showed that every PLP-NAH product was detectable under DCC reaction conditions. On the assumption that the concentration of the PLP-NAHs was approximately equal at equilibrium, this allowed us to convert an intensity value in the mass spectrum to a relative concentration value, making the results easier to compare. Interestingly the DCC experiments with ecBioA and mtBioA selected 6 and 10 PLP-NAH combinations, respectively, out of a possible 28. Of these privileged combinations, only 4 were selected by both isoforms. This difference in selectivity was attributed to subtle differences in the architecture and calculated electrostatic potential of residues at the entrance to the binding channel. It was further inferred

that the commonly selected combinations were therefore likely to interact with other homologous BioA isoforms.

Compounds **H11**, **H12**, **H13**, **H14**, and **H17** were submitted for *in vivo* characterisation, and although there was no detectable activity against *M. smegmatis*, two of the compounds showed *in vivo* activity against *E. coli*. Interestingly, inhibitory activity was also discovered against *A. baumannii*, an ESKAPE pathogen with strains reported to be resistant to all known antibiotics.

This work further validates the use of ESI-MS as a method for protein-templated DCC analysis in identifying lead compounds. Although there are limitations to the technique, libraries far greater in size to the one that we have demonstrated may achieved.

4.4 Experimental

4.4.1 Expression and purification of ecBioA (UniProtKB: P12995)

An aliquot (50 µL) of BL21(DE3) cells was transformed with the *ecBioA/p*ET-28a plasmid and set on ice for 25 minutes. The cells were then heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (30 µg/mL kanamycin) and incubated overnight at 37 °C. A single transformant was used to inoculate two seed cultures of sterile LB broth (2 x 250 mL, 30 µg/mL kanamycin) and agitated overnight at 37 °C. One of the overnight seed cultures was used to sub-culture sterile LB broth (4 x 500 mL, 30 µg/mL kanamycin) to an OD_{600} of 0.1. The cultures were agitated at 37 °C until the OD_{600} reached 0.6, at which point the temperature was lowered to 25 °C and expression was induced by addition of IPTG (final conc. 0.5 mM). Cells were harvested after a further 5 hours at 30 °C and subsequently stored at -20 °C.

N-Terminal histidine-tagged ecBioA was purified at 4 °C by Ni-affinity chromatography followed by size exclusion chromatography. The BL21 (DE3) cell pellet expressing ecBioA was resuspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM $MgCl_2$, 1 mg/mL lysozyme, 0.1 mM TCEP) lysed for 15 minutes with rounds of 30 second of sonication followed by 30 seconds of rest. Excess cell matter was removed by centrifugation (*18,000 g*, 30 minutes, 4 °C) and the supernatant was injected onto a HisTrap 1 mL (GE) Ni^{2+}-affinity chromatography column pre-equilibrated in wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole). The column was washed with wash buffer (5 CV) before both the histidine-tagged protein was eluted using a gradient (0-100%) of wash buffer to elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole) over 20 CV.

Each elution fraction was analysed by 15% SDS-PAGE, and the fractions containing His-tagged ecBioA were pooled, concentrated to 2 mL using a 30 kDa MWCO spin filter and further purified by size exclusion chromatography (HiLoad Superdex S200

16/60) with an isocratic elution using mobile phase buffer (25mM HEPES pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 10% glycerol) at 1 mL/min over 120 minutes. ecBioA eluted at approx. 70 mL and the most concentrated fractions were pooled, cryogenically frozen in 1 mL aliquots and stored at -80 °C.

4.4.2 Expression and purification of mtBioA (UniProtKB: P9WQ81)

An aliquot (50 µL) of BL21 Rosetta 2 (DE3) cells was transformed with the *mtBioA/pUC-19* plasmid and set on ice for 25 minutes. The cells were then heat shocked at 42 °C for 40 seconds and set back on ice for a further 2 minutes. SOC media (100 µL) was added and the mixture was agitated at 37 °C for 1 hour. The mixture was spread on LB agar (100 µg/mL ampicillin) and incubated for 72 hours at 37 °C. A single transformant was used to inoculate two seed cultures of sterile LB broth (2 x 250 mL, 100 µg/mL ampicillin) and agitated overnight at 37 °C. One of the overnight seed cultures was used to sub-culture sterile TB broth (1 L, 100 µg/mL ampicillin) to an OD_{600} of 0.1. The cultures were agitated at 37 °C for 16 hours after which the cells were harvested and subsequently stored at -20 °C.

Histidine-tagged mtBioA was purified at 4 °C by Ni-affinity chromatography followed by size exclusion chromatography. The BL21 (DE3) cell pellet expressing mtBioA was resuspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM $MgCl_2$, 1 mg/mL lysozyme, 0.1 mM TCEP) lysed for 15 minutes with rounds of 30 second of sonication followed by 30 seconds of rest. Excess cell matter was removed by centrifugation (*18,000 g*, 30 minutes, 4 °C) and the supernatant was injected onto a HisTrap 1 mL (GE) Ni^{2+}-affinity chromatography column pre-equilibrated in wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole). The column was washed with wash buffer (5 CV) before both the histidine-tagged protein was eluted using a gradient (0-100%) of wash buffer to elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole) over 20 CV.

Each elution fraction was analysed by 15% SDS-PAGE, and the fractions containing His-tagged mtBioA were pooled, concentrated to 2 mL using a 30 kDa MWCO spin filter and further purified by size exclusion chromatography (HiLoad Superdex S200 16/60) with an isocratic elution using mobile phase buffer (25mM HEPES pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 10% glycerol) at 1 mL/min over 120 minutes. mtBioA eluted at approx. 70 mL and the most concentrated fractions were pooled, cryogenically frozen in 1 mL aliquots and stored at -80 °C.

4.4.3 Cloning, expression and purification of K283A mtBioA

Note: preparation of the K283A mutant gene was done by Annabel Serpico.

A gene fragment was purchased (Genscript) containing the K283A mutation flanked with Not1 and Age1 restriction sites intrinsic to the *mtBioA* gene, but absent from the pUC-19 plasmid. A restriction digest of WT mtBioA was carried out with restriction enzymes Not1 and Age1, allowing the backbone to be isolated and subsequently ligated with the synthetic fragment.

The K283A *mtBioA/pUC-19* plasmid was transformed into an aliquot (50 µL) of BL21 Gold (DE3) cells and expressed and purified following the same protocol as WT mtBioA (section 4.4.2).

4.4.4 Bradford Assay

The Bradford assay reagent was purchased from Sigma Aldrich and used according to the supplied protocol in 96-well format. Briefly, a binary serial dilution was prepared in the linear range of 1.4 – 0.1 mg/mL (1.4, 1.0, 0.7, 0.5, 0.35, 0.25 mg/mL). The protein sample concentration was estimated using the Beer-Lambert equation with the calculated extinction coefficient of the protein (http://web.expasy.org/protparam) and the absorbance of the sample at 280 nm. The sample was diluted accordingly to be within the linear range of the Bradford assay. Wells were prepared containing 5 µL of sample or protein standard, and 250 µL of Bradford reagent. All wells were prepared in triplicate. The reaction was left for

15 mins before the absorbance of each well was read at 595 nm, and a standard curve was plotted using the protein standards. The concentration of the protein sample was determined using the standard curve.

4.4.5 Differential Scanning Fluorimetry (DSF) screen

A DSF screen was conducted against ecBioA and mtBioA using the in-house hydrazide library. The screen was carried out in 50 μL 96-well PCR plate format with a final concentration of 200 nM protein (42.5 μL ecBioA or mtBioA, freshly dialysed against 50 μM PLP and desalted into 25 mM HEPES pH 7.5, 50 mM NaCl), 500 μM compound (2.5 μL 10 mM DMSO stock H1-H28) and 5x SYPRO Orange (5 μL 50x SYPRO Orange prepared in 25 mM HEPES pH 7.5, 50 mM NaCl) at a final concentration of 5% DMSO. The plate was incubated at room temperature for 30 mins and centrifuged (2 mins, 2000 x g, 21 °C) to remove bubbles. The screen was conducted using a Bio-Rad iQ5 Thermocycler scanning a temperature range from 20 – 95 °C with increments of 1 °C every 30 seconds. Data were collected using the Bio-Rad iQ5 software, and replotted as the derivative of the change in fluorescence intensity over time to give the melting temperature of the protein with each compound.

4.4.6 DCC experiments

4.4.6.1 DCL conditions

The following procedure was used for both ecBioA and mtBioA. The protein was concentrated to 200 μM using a 30 kDa MWCO spin filter. The protein sample was then dialysed against fresh PLP (600 μM) for 30 minutes to fully load the enzyme, desalted (25 mM HEPES pH 7.5, 50 mM NaCl) and divided into aliquots (2 x 200 μL). As a control, one aliquot was incubated with $NaBH_4$ (2 mM) for 30 mins to reduce the PLP internal aldimine, and desalted (25 mM HEPES pH 7.5, 50 mM NaCl). The concentration of each protein sample was determined by Bradford assay, and diluted to 120 μM with desalting buffer. The samples were used to prepare protein-templated and control (with reduced internal aldimine) DCLs in parallel, which contained 30 μM BioA, 30 μM each of hydrazides H1-H28, 3 mM 4-APA, and 100 mM

pH 6.2 sodium phosphate buffer containing 1.5% DMSO. The DCLs were incubated at room temperature with gentle shaking overnight and then desalted (50 mM NH_4HCO_3). Each sample was diluted two-fold with a 60% MeOH, 1% formic acid solution and were analysed by direct infusion electrospray ionisation using a Bruker 12T SolariX with a 10cm infinity cell. Spectra were acquired in positive polarity over 100-500 m/z with a 2megaword data size. 400 spectra were summed, each with a 50 ms accumulation time. A 1-point calibration was carried out on each dataset using the calculated mass (ExPASy) of NAH1 (419.11148 Da). Assigned peaks were within 2 ppm of the predicted mass values.

4.4.6.2 DCL reversibility

An aliquot of ecBioA was concentrated to 200 μM using a 30 kDa MWCO spin filter. The protein sample was then dialysed against fresh PLP (600 μM) for 30 minutes to fully load the enzyme and desalted (25 mM HEPES pH 7.5, 50 mM NaCl). The concentration of each protein sample was determined by Bradford assay, and diluted to 120 μM with desalting buffer. A 200 μL DCL was prepared containing 30 μM BioA, 30 μM hydrazide **H1**, 3 mM 4-APA, and 100 mM pH 6.2 sodium phosphate buffer containing 1.5% DMSO. The DCL was incubated at room temperature with gentle shaking overnight and then a 50 μL aliquot was removed and desalted (50 mM NH_4HCO_3). To the remaining 150 μL sample was added 22.5 μM each of hydrazide **H2-28** to maintain an equal DCL concentration. This DCL was incubated at room temperature with gentle shaking overnight and then a 50 μL aliquot was removed and desalted (50 mM NH_4HCO_3). Both samples were analysed as previously (see section 4.4.6.1).

4.4.7 Structural Studies

mtBioA was co-crystallised with compounds **H11-14** using previously published conditions, in a 24-well plate format using the hanging-drop method[51]. mtBioA was dialysed against fresh PLP (50 μM), desalted (25 mM HEPES pH 7.5, 50 mM NaCl) and concentrated to 10 mg/mL using a 30 kDa MWCO filter (Sartorius). The well solutions

were prepared freshly and filtered through a 0.2 μm cellulose membrane. Each well solution consisted of 100 mM HEPES pH 7.5, 100 mM $MgCl_2$, 5 mM of respective small molecule, and rows 1-6 contained a range of 9-14% PEG 8000, respectively. Crystals were frozen in corresponding well solution supplemented with 15% PEG 400. The co-crystal structure of **H11** in mtBioA was solved and optimised by Dr Jon Marles-Wright (Newcastle University).

4.4.8 Disc diffusion assays

Compounds **H11**, **H12**, **H13**, **H14**, and **H17** were tested for growth inhibitory activity by the Brown group (McMaster, Hamilton) against *E. coli, P. aeruginosa, A. baumannii and S. aureus* alongside MAC13772 (positive control), ddH_2O (negative control), and DMSO (negative control). Bacterial cultures were grown in LB to an OD_{600} value of approx. 0.1 and streaked on LB + 0.1 u/mL avidin plates. Paper discs impregnated with 10 μL of 25.6 mg/mL of sample (256 μg of sample) were placed on the streaked plate. The plates were incubated at 37 °C for 18 hours and the growth inhibition around the discs was monitored visually.

4.4.9 *In vivo* assays

Compounds **H11**, **H12**, **H13**, **H14**, and **H17** were kindly assayed against *E. coli, M. smegmatis and A. baumannii* by the Brown group (McMaster, Hamilton) using the following protocols:

4.4.9.1 *E. coli*

The 5 compounds provided were tested for *in vivo* activity against BW25113 (orange) and *ΔyigM* (blue) strains of *Escherichia coli* (*E. coli*). *ΔyigM E. coli* is the biotin transporter deletion *E. coli* mutant, making it significantly more sensitive to biotin biosynthesis inhibitors. Minimum inhibitory concentration (MIC) values were determined in M9 minimal media. E. coli strains were diluted 1 in 1000-fold from a mid-log subculture of overnight cultures grown at 37 °C. The diluted bacteria were added to flat-bottomed 96 well plates containing two-fold serial dilution of the tested

compounds (256 - 1µg/mL) and incubated at 37 °C for 16 hours. Optical density at 600 nm was measured after incubation and MIC was determined as the lowest concentration that inhibits visible growth.

4.4.9.2 *M. smegmatis*

The 5 compounds provided were tested for *in vivo* activity against *M. smegmatis*. MIC values were determined in LB media in the presence of 0.1 u/mL avidin to sequester any free biotin in the media. *M. smegmatis* was diluted 1 in 10000-fold from a mid-log subculture of overnight cultures grown at 37 °C. The diluted bacteria were added to flat-bottomed 96 well plates containing two-fold serial dilution of the tested compounds (256 - 1µg/mL) and incubated at 37 °C for 40 hours. Optical density at 600 nm was measured after incubation and MIC was determined as the lowest concentration that inhibits visible growth.

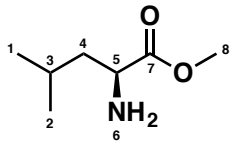
4.4.9.3 *A. baumannii*

The 5 compounds provided were tested for *in vivo* activity against *A. baumannii.* MIC values were determined in LB media in the presence of 0.1 u/mL avidin to sequester any free biotin in the media. *A. baumannii* was diluted 1 in 10000-fold from a mid-log subculture of overnight cultures grown at 37 °C. The diluted bacteria were added to flat-bottomed 96 well plates containing a two-fold serial dilution of the tested compounds (256 - 0.25 µg/mL) and incubated at 37 °C for 18 hours. Optical density at 600 nm was measured after incubation and MIC was determined as the lowest concentration that inhibits visible growth.

4.4.10 Synthetic Chemistry

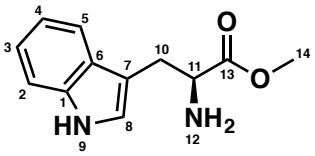
Note: Compounds AE01-AE18 were initially synthesised by J. T. Wang during her 5th year M.Sc. project to increase the aqueous solubility of DCLs.

Compound AE01: methyl (2*S*)-2-amino-4-methylpentanoate



Thionyl chloride (5.0 eq., 4.53 g, 38.1 mmol) was added dropwise to a solution of L-leucine (1.0 eq., 1.00 g, 7.6 mmol) in methanol (20 ml) at 0°C. The solution was stirred at room temperature for 21 hours. The solvent was removed *in vacuo* and the residue was resuspended in acetone (30 ml) and stirred for 15 minutes. The sample was further sonicated for 10 minutes at an amplitude of 20 microns. The solid formed was collected by vacuum filtration and washed with acetone (3 x 10 ml) to yield the target compound **AE01** (1.09 g, 7.5 mmol, 99%) as a colourless solid: **MP** 142-143°C; **IR** (cm^{-1}) 2957, 2843, 1738, 1595, 1506, 1227; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.86 (2H, b, NH_2-6), 4.04 (1H, s, CH-5), 3.84 (3H, s, CH_3-8), 2.00 (2H, m, CH-4), 1.87 (1H, m, CH-3), 1.02 (6H, d, *J* = 0.9 Hz, CH_3-1,2); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 170.05, 53.07, 51.79, 39.55, 24.43, 22.30, 21.96; **HRMS** m/z (ESI^+) calcd. $C_7H_{15}NO_2$ $[M+Na]^+$ requires 168.0995, found 168.1000.

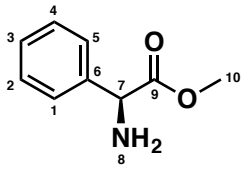
Compound AE02: methyl (2*S*)-2-amino-3-(1H-indol-3-yl)propanoate



Thionyl chloride (5.0 eq., 2.91 g, 24.5 mmol) was added dropwise to a solution of L-tryptophan (1.0 eq., 1.00 g, 4.9 mmol) in methanol (20 ml) at 0°C. The solution was stirred at room temperature for 19 hours. The solvent was removed *in vacuo* and the residue was resuspended in acetone (30 ml) and stirred for 15 minutes. The precipitate was collected by vacuum filtration and washed with acetone (3 x 10 ml) to yield the target compound **AE02** (0.80 g, 3.7 mmol, 75%) as a white solid: **MP** 204-

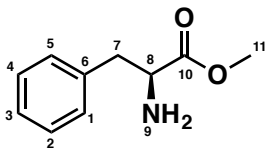
206°C; **IR** (cm^{-1}) 3260, 2999, 2866, 1748, 1578, 1501, 730; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 11.06 (1H, s, NH-9), 8.41 (2H, s, NH_2-12), 7.50 (1H, d, *J* = 7.9, 0.7 Hz, CH-5) 7.38 (1H, dt, *J* = 8.0, 0.9 Hz, CH-2), 7.23 (1H, d, *J* = 2.5 Hz, CH-8), 7.11 (1H, ddd, *J* = 8.1, 7.0, 1.2 Hz, CH-3), 7.03 (1H, ddd, *J* = 8.0, 7.0, 1.1 Hz, CH-4), 4.27 (1H, t, *J* = 6.3 Hz, CH-11), 3.68 (3H, s, CH_3-14), 3.37 – 3.22 (2H, m, CH-10); **^{13}C NMR** (126 MHz, D_2O) δ 170.8, 136.3, 126.4, 125.4, 122.3, 119.6, 118.1, 112.0, 106.1, 53.6, 53.4, 25.8; **HRMS** m/z (ESI^+) calcd. $C_{12}H_{14}N_2O_2$ $[M+H]^+$ requires 219.1128, found 219.1156.

Compound AE03: methyl (2*S*)-2-amino-2-phenylacetate



Thionyl chloride (5.0 eq., 3.9 g, 33.1 mmol) was added dropwise to a solution of D-phenylglycine (1.0 eq., 1.0 g, 6.6 mmol) in methanol (20 ml) at 0°C. The solution was stirred at room temperature for 20 hours. The solvent was removed *in vacuo* and the residue was resuspended in acetone (30 ml) and stirred for 15 minutes. The precipitate was collected by vacuum filtration and washed with acetone (3 x 10 ml) to yield the target compound **AE03** (1.24 g, 6.2 mmol, 94%) as a colourless hydrochloride salt: **MP** 200-201°C; **IR** (cm^{-1}) 2837, 2623, 1737, 1568, 1501; **^{1}H NMR** (500 MHz, DMSO) δ 8.96 (2H, s, NH_2-8), 7.52 – 7.46 (5H, m, CH-1,2,3,4,5), 5.30 (1H, s, CH-7), 3.73 (3H, s, CH_3-10); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 169.7, 131.08, 130.4, 129.7, 128.0, 56.4, 53.91; **HRMS** m/z (ESI^+) calcd. $C_9H_{11}NO_2$ $[M+Na]^+$ requires 188.0682, found 188.0687.

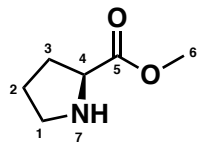
Compound AE04: methyl (2*S*)-2-amino-3-phenylpropanoate



Thionyl chloride (5.0 eq., 3.6 g, 30.5 mmol) was added to a solution of L-phenylalanine (1.0 eq., 1.0 g, 6.1 mmol) in methanol (20 ml) at 0 °C and stirred at room temperature

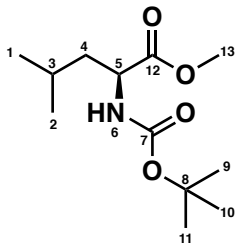
for 24 hours. The solvent was removed *in vacuo* to give a white residue, which was resuspended in acetone (30 ml) and stirred for 15 minutes. The precipitate was collected by vacuum filtration and washed with acetone (3 x 10 ml) to yield the target compound **AE04** (786 mg, 4.4 mmol, 72%) as a colourless solid: **MP** 143-145 °C; **IR** (cm^{-1}) 2833, 1746, 1584, 1495; **^{1}H NMR** (500 MHz, DMSO) δ 8.64 (2H, s, NH_2-9), 7.36 – 7.31 (2H, m, CH-2,4), 7.31 – 7.26 (1H, m, CH-3), 7.25 – 7.22 (2H, m, CH-1,5), 4.26 (1H, dd, *J* = 6.0, 7.3, CH-8), 3.18 (1H, dd, *J* = 14.0, 5.9, CH-7), 3.09 (1H, dd, *J* = 14.0, 7.4, CH-7); **^{13}C NMR** (126 MHz, DMSO) δ 169.31, 134.8, 129.3, 128.6, 127.2, 53.2, 52.5, 35.8; **HRMS** m/z (ESI^+) calcd. $C_{10}H_{13}O_2N_1Na$ $[M+Na]^+$ requires 202.08385, found 202.08650.

Compound AE05: methyl (2*S*)-pyrrolidine-2-carboxylate



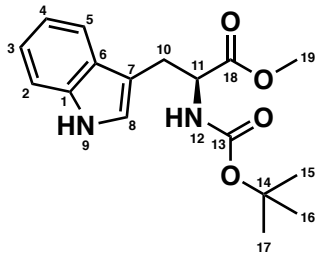
Thionyl chloride (5.0 eq., 5.17 g, 43.4 mmol) was added dropwise to a solution of L-proline (1.0 eq., 1.00 g, 8.7 mmol) in methanol (20 ml) at 0°C. The solution was stirred at room temperature for 25 hours. The solvent was removed *in vacuo* and the residue was taken up in water and the mixture was extracted with ethyl acetate (3 x 10 ml). The combined organic layers were dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the target compound **AE05** (1.10 g, 8.5 mmol, 98%) as a light brown oil: **IR** 2870,1742, 1445; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 4.52 (1H, m, CH-4), 3.86 (3H, s, CH_3-6), 3.65 – 3.50 (2H, m, CH-1), 2.48 – 2.41 (1H, m, CH-3), 2.22 – 2.17 (1H, m, CH-3), 2.14 – 2.04 (2H, m, CH-2); **^{13}C NMR** (126 MHz, D_2O) δ 169.3, 59.3, 53.6, 46.2, 28.8, 23.7; **HRMS** m/z (ESI^+) calcd. $C_6H_{11}NO_2$ $[M+H]^+$ requires 130.0863, found 130.0871.

Compound AE06: methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-4-methylpentanoate



Sodium bicarbonate (2.0 eq., 1.15 g, 13.7 mmol) and di-tert-butyl dicarbonate (1.1 eq., 1.64 g, 7.5 mmol) were added to a solution of methyl (2*S*)-2-amino-4-methylpentanoate **AE01** (1.0 eq., 0.99 g, 6.8 mmol) in a 1:1 mixture of tetrahydrofuran and water (20 ml). The mixture was stirred at room temperature for 24 hours. The volatile solvent was removed *in vacuo* and then taken up in sodium bicarbonate and ethyl acetate. The organic layer was separated from the aqueous layer. The aqueous was extracted with ethyl acetate (2 x 10 ml). The combined organic layers were dried over magnesium sulfate, filtered and concentrated *in vacuo* to yield the target compound **AE06** (1.51 g, 6.2 mmol, 90%) as a colourless oil: **IR** (cm^{-1}) 2957, 2872, 1738, 1713, 1508; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 4.90 (1H, d, *J* = 8.6 Hz, NH-6), 4.34 (1H, q, *J* = 8.5, 7.9 Hz, CH-5), 3.75 (3H, s, CH_3-13), 1.71 (1H dq, *J* = 15.0, 8.5, 7.6 Hz, CH-3), 1.67 – 1.56 (2H, m, CH_2-4), 1.55 (6H, d, *J* = 0.9 Hz, CH_3-1,2), 1.47 (9H, s, CH_3-9,10,11); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 155.4, 146.8, 85.2, 52.1, 41.9, 28.3, 27.4, 24.8, 22.8, 21.9; **HRMS** m/z (ESI^+) calcd. $C_{12}H_{23}NO_4$ $[M+Na]^+$ requires 268.1519, found 268.1518.

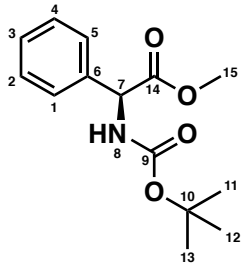
Compound AE07: methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-3-(1H-indol-3-yl)propanoate



Sodium bicarbonate (2.0 eq., 0.54 g, 6.4 mmol) and di-tert-butyl dicarbonate (1.1 eq., 0.77 g, 3.5 mmol) were added to a solution of methyl (2*S*)-2-amino-3-(1H-indol-3-

yl)propanoate **AE02** (1.0 eq., 0.70 g, 3.2 mmol) in a 1:1 mixture of tetrahydrofuran and water (20 ml). The mixture was stirred at room temperature for 4 hours. The volatile solvent was removed *in vacuo* and then taken up in sodium bicarbonate and ethyl acetate. The organic layer was separated from the aqueous. The aqueous was extracted two more times with ethyl acetate. The combined organic layers were dried over magnesium sulfate, filtered and concentrated *in vacuo* to yield the target compound **AE07** (0.96 g, 3.01 mmol, 94%) as a colourless oil: **IR** (cm^{-1}) 3383, 3314, 1736, 1690, 1522; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.09 (1H, s, NH-12), 7.58 (1H, dt, *J* = 7.9, 1.0 Hz, CH-5), 7.38 (1H, dt, *J* = 8.1, 0.9 Hz, CH-2), 7.22 (1H, ddd, *J* = 8.2, 7.0, 1.2 Hz, CH-4), 7.14 (1H, ddd, *J* = 8.0, 7.0, 1.1 Hz, CH-3), 7.04 (1H, s, CH-8), 4.67 (1H, m, CH-11), 3.70 (3H, s, CH_3-19), 3.31 (2H, m, CH_2-10), 1.55 (9H, s, CH_3-15,16,17); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 172.9, 155.4, 136.3, 127.9, 122.8, 122.4, 119.8, 118.9, 111.3, 110.5, 80.0, 73.4, 54.3, 52.4, 28.5, 28.1; **HRMS** m/z (ESI^+) calcd. $C_{17}H_{22}N_2O_4$ $[M+Na]^+$ requires 341.1472, found 341.1480.

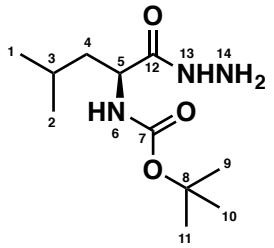
Compound AE08: methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-2-phenylacetate



Sodium bicarbonate (2.0 eq., 1.16 g, 13.8 mmol) and di-tert-butyl dicarbonate (1.1 eq., 1.65 g, 7.5 mmol) were added to a solution of methyl (2*S*)-2-amino-2-phenylacetate **AE03** (1.0 eq., 1.14 g, 6.9 mmol) in a 1:1 mixture of tetrahydrofuran and water (20 ml). The mixture was stirred at room temperature for 21 hours. The volatile solvent was removed *in vacuo* and then taken up in sodium bicarbonate and ethyl acetate. The organic layer was separated from the aqueous layer. The aqueous was extracted with ethyl acetate (2 x 10 ml). The combined organic layers were dried over magnesium sulfate, filtered and concentrated *in vacuo* to yield the target compound **AE08** (1.65 g, 6.2 mmol, 90%) as a colourless solid: **MP** 99-101°C; **IR** (cm^{-}

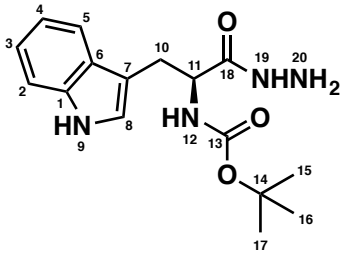
[1]) 3401, 2982, 1728, 1697, 1501; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.41 – 7.30 (5H, m, CH-1,2,3,4,5), 5.35 (1H, d, *J* = 7.2 Hz, CH-7), 3.74 (3H, s, CH_3-15), 1.55 (9H, s, CH_3-11,12,13); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 137.0, 128.9, 128.4, 127.1, 80.2, 57.6, 52.7, 28.3; **HRMS** m/z (ESI^+) calcd. $C_{14}H_{19}NO_4$ $[M+Na]^+$ requires 288.1206, found 288.1209.

Compound AE09: *t*-butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-3-methylbutyl]carbamate



Hydrazine solution (24-26% in water, 5.0 eq., 6.14 g, 30.7 mmol) was added to a solution of methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-4-methylpentanoate **AE06** (1.0 eq., 1.50 g, 6.1 mmol) in methanol (15 ml) and stirred at reflux for 4 hours. The solution was concentrated *in vacuo* to yield the target compound **AE09** (1.30 g, 5.3 mmol, 87%) as a light yellow oil: **IR** (cm^{-1}) 3319, 2959, 1697, 1655, 1519, 1366; **^{1}H NMR** (500 MHz, DMSO) δ 9.01 (1H, s, NH-6), 6.76 (d, *J* = 8.5 Hz, 1H), 3.93 (2H, td, *J* = 9.1, 5.5 Hz, CH_2-4), 1.55 (1H, dq, *J* = 13.5, 6.8 Hz, CH-3), 1.37 (9H, s, CH_3-9,10,11), 0.85 (6H, dd, *J* = 12.5, 6.6 Hz, CH_3-1,2); **^{13}C NMR** (126 MHz, DMSO) δ 172.2, 155.6, 78.3, 51.8, 41.6, 28.7, 24.7, 23.3, 22.2; **HRMS** m/z (ESI^+) calcd. $C_{11}H_{23}N_3O_3$ $[M+Na]^+$ requires 268.1632, found 268.1632.

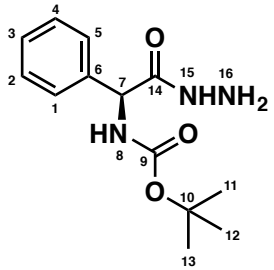
Compound AE10: *t*-butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-2-(1H-indol-3-yl)ethyl]carbamate



Hydrazine solution (24-26% in water, 5.0 eq., 3.00 g, 15.0 mmol) was added to a solution of methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-3-(1H-indol-3-

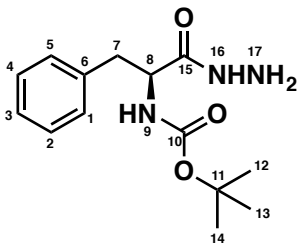
yl)propanoate **AE07** (1.0 eq., 0.96 g, 3.0 mmol) in methanol (10 ml) and stirred at reflux for 15 hours. Another 5 equivalents of hydrazide solution were added and the mixture was stirred at reflux for another 3.5 hours. The solution was concentrated *in vacuo* to yield the target compound **AE10** (0.78 g, 2.5 mmol, 81%) as an off-white sticky solid: **IR** (cm^{-1}) 3343, 3210, 2984, 1688, 1672, 1655, 1522, 1504; **^{1}H NMR** (500 MHz, DMSO) δ 10.78 (1H, s, NH-9), 9.12 (1H, s, NH-12), 7.59 (1H, d, *J* = 7.9 Hz, CH-5), 7.32 (1H, d, *J* = 8.1 Hz, CH-2), 7.13 (1H, s, NH-19), 7.06 (1H ddd, *J* = 8.1, 6.9, 1.2 Hz, CH-4), 6.98 (1H, ddd, *J* = 8.0, 6.9, 1.0 Hz, CH-3), 6.73 (1H, d, *J* = 8.4 Hz, CH-8), 4.16 (2H, m, CH_2-10), 3.01 (1H, m, NH-19), 2.89 (1H, m, CH-11), 1.32 (9H, s, CH_3-15,16,17); **^{13}C NMR** (126 MHz, D_2O) δ 168.1, 156.1, 136.2, 126.4, 125.5, 122.3, 119.7, 118.0, 112.1, 105.9, 72.1, 62.5, 52.5, 26.8; **HRMS** m/z (ESI^+) calcd. $C_{16}H_{22}N_4O_3$ $[M+Na]^+$ requires 341.1584, found 341.1562.

Compound AE11: *t*-butyl *N*-[(*S*)-(hydrazinecarbonyl)(phenyl)methyl]carbamate



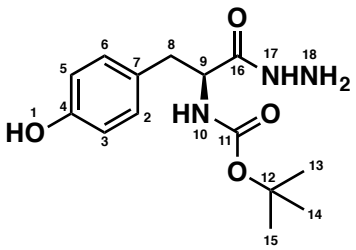
Hydrazine solution (24-26% in water, 5.0 eq., 6.22 g, 31.1 mmol) was added to a solution of methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-2-phenylacetate **AE08** (1.0 eq., 1.6 g, 6.2 mmol) in methanol (15 ml) and stirred at reflux for 11 hours. Another 5 equivalents of hydrazide solutions were added and the mixture was stirred at reflux for another 3 hours. The solution was concentrated *in vacuo* to yield the target compound **AE11** (1.23 g, 4.7 mmol, 75%) as a light yellow solid: **MP** 104-106°C; **IR** (cm^{-1}) 3333, 2974, 1686, 1653, 1626, 1518, 1499; **^{1}H NMR** (500 MHz, DMSO) δ 7.41 – 7.31 (5H, m, CH-1,2,3,4,5), 5.80 (1H, d, *J* = 6.9 Hz, CH-7), 5.24 (1H, s, NH-8), 1.43 (9H, s, CH_3-11,12,13); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 133.4, 129.6, 129.0, 127.7, 55.4; **HRMS** m/z (ESI^+) calcd. $C_{13}H_{19}N_3O_3$ $[M+Na]^+$ requires 288.1319, found 288.1332.

Compound AE12: *t*-butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-2-phenylethyl]carbamate



Hydrazine solution (24-26% in water, 5.0 eq., 3.58 g, 17.9 mmol) was added to a solution of methyl (2*S*)-2-{[(*tert*-butoxy)carbonyl]amino}-2-phenylacetate **AE04** (1.0 eq., 1.00 g, 3.6 mmol) in methanol (10 ml) and stirred at reflux for 5 hours. The solution was then concentrated *in vacuo* to yield the target compound **AE12** (0.78 g, 2.8 mmol, 78%) as an off-white solid: **MP** 132-133°C; **IR** (cm^{-1}) 3325, 1690, 1659, 1647, 1502; **^{1}H NMR** (500 MHz, DMSO) δ 9.10 (1H, s, NH-9), 7.30 – 7.15 (5H, m, CH-1,2,3,4,5), 6.87 (1H, d, *J* = 8.7 Hz, NH-16), 4.23 (1H, m, CH-7), 4.12 (1H, td, *J* = 9.4, 4.7 Hz, CH-7), 2.88 (1H, dd, *J* = 13.7, 4.8 Hz, CH-8), 2.74 (1H dd, *J* = 13.7, 10.0 Hz, NH-17), 1.30 (9H, s, CH_3-12,13,14); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 171.4, 155.6, 138.6, 129.6, 128.5, 126.6, 78.4, 54.9, 38.3, 28.6; **HRMS** m/z (ESI^+) calcd. $C_{14}H_{21}N_3O_3$ $[M+Na]^+$ requires 280.1656, found 280.1668.

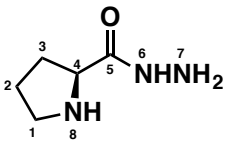
Compound AE13: *t*-butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-2-(4-hydroxyphenyl)ethyl]carbamate



Hydrazine solution (24-26% in water, 5.0 eq., 1.68 g, 8.4 mmol) was added to a solution of *tert*-butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-2-(4-hydroxyphenyl)ethyl]carbamate (1.0 eq., 0.50 g, 1.7 mmol) in a 1:1 mixture of tetrahydrofuran and methanol (10 ml) and stirred at room temperature overnight. The solution was then concentrated *in vacuo* to yield the target compound **AE13** (0.65 g, 2.2 mmol, 65%) as a colourless solid: **MP** 207-209°C; **IR** (cm^{-1}) 3329, 3304,

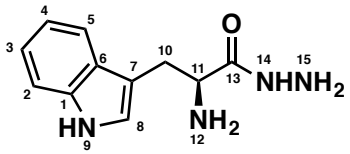
1682, 1663, 1611, 1512; **^{1}H NMR** (500 MHz, DMSO) δ 9.13 (1H, s, OH-1), 9.04 (1H, s, NH-10), 7.06 – 6.99 (2H m, CH-2,6), 6.77 (1H, d, *J* = 8.6 Hz, NH-17), 6.67 – 6.61 (2H, m, CH-3,5), 4.19 (2H, m, CH_2-8), 4.11 – 3.98 (1H, m, CH-9), 2.75 (1H, dd, *J* = 13.8, 5.0 Hz, NH-18), 2.63 (1H, dd, *J* = 13.8, 9.6 Hz, NH-18), 1.31 (9H, s, CH_3-13,14,15); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 171.6, 156.2, 155.6, 130.5, 128.6, 115.3, 78.3, 55.2, 37.6, 28.7; **HRMS** m/z (ESI^+) calcd. $C_{14}H_{21}N_3O_4$ $[M+Na]^+$ requires 318.1424, found 318.1435.

Compound AE14 (H10): (2*S*)-pyrrolidine-2-carbohydrazide



Hydrazine solution (24-26% in water, 5 eq., 5.83 g, 5.5 ml, 43.5 mmol) was added to a solution of methyl (2*S*)-pyrrolidine-2-carboxylate **AE05** (1.0 eq., 1.37 g, 8.7 mmol) in methanol (15 ml) and stirred at reflux for 3.5 hours. The solution was then concentrated *in vacuo* to give the target compound **AE14** (0.67 g, 5.2 mmol, 60%) as a purple oil: **IR** (cm^{-1}) 3361, 2473, 2241, 2070; **^{1}H NMR** (500 MHz, CD_3OD) δ 4.46 – 4.37 (1H, m, CH-4), 3.52 – 3.37 (2H, m, CH_2-1), 2.50 (1H, m, CH-3), 2.22 – 2.01 (3H, m, CH_2-2 CH-3); **^{13}C NMR** (126 MHz, CD_3OD) δ 167.6, * 46.1, 29.4, 23.6; **HRMS** m/z (ESI^+) calcd. $C_5H_{11}N_3O$ $[M+H]^+$ requires 130.0975, found 130.0975. * One carbon signal obscured by CD_3OD.

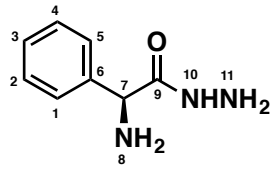
Compound AE15 (H11): (2*S*)-2-amino-3-(1H-indol-3-yl)propanehydrazide



t-Butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-2-(1H-indol-3-yl)ethyl]carbamate **AE10** (1.0 eq., 100 mg, 0.31 mmol) was stirred in 1 M HCl (2 ml) at room temperature overnight. The solvent was removed *in vacuo* to give the target compound **AE15** (102 mg, 0.31 mmol, 99%) as an off-white solid: **MP** decomposes at 240°C; **IR** (cm^{-1}) 3117, 3021, 2805, 1757, 1711, 1399; **^{1}H NMR** (500 MHz, CD_3OD) δ 7.73 – 7.71 (1H, m, CH-8), 7.51

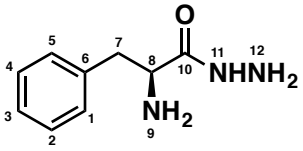
– 7.37 (3H, m, CH-3,4 NH-9), 7.31 – 7.29 (2H b, NH_2-12), 7.30 (d, *J* = 4.1 Hz, 2H), 7.18 – 7.09 (2H, m, CH-2,5), 4.32 (1H, dd, *J* = 7.8, 6.8 Hz, CH-11), 3.51 – 3.37 (2H, m, CH_2-10); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 136.9, 126.8, 124.5, 121.6, 119.1, 117.7, 111.3, 105.9, 52.4, 27.3; **HRMS** m/z (ESI^+) calcd. $C_{11}H_{14}N_4O$ $[M+H]^+$ requires 219.1240, found 219.1239.

Compound AE16 (H12): (2*S*)-2-amino-2-phenylacetohydrazide



t-Butyl *N*-[(*S*)-(hydrazinecarbonyl)(phenyl)methyl]carbamate **AE11** (1.0 eq., 100 mg, 0.38 mmol) was stirred in 1 M HCl (2 ml) at room temperature overnight. The solvent was removed *in vacuo* to give the target compound **AE16** (94 mg, 0.34 mmol, 91%) as a light yellow solid: **MP** decomposes at 260°C; **IR** (cm^{-1}) 3119, 3021, 2805, 1751, 1670, 1397; **^{1}H NMR** (500 MHz, CD_3OD) δ 7.60 – 7.52 (2H, m, CH-1,5), 7.52 – 7.43 (3H, m, CH-2,3,4), 4.84 (1H, s, CH-7); **^{13}C NMR** (126 MHz, CD_3OD) δ 133.4, 129.6, 129.0, 127.7, 55.4; **HRMS** m/z (ESI^+) calcd. $C_8H_{11}N_3O$ $[M+Na]^+$ requires 188.0794, found 188.0812.

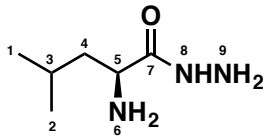
Compound AE17 (H13): (2*S*)-2-amino-3-phenylpropanehydrazide



t-Butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-2-phenylethyl]carbamate **AE12** (1.0 eq., 100 mg, 0.36 mmol) was stirred in 1 M HCl (2 ml) at room temperature overnight. The solvent was removed *in vacuo* to give the target compound **AE17** (103 mg, 0.36 mmol, 99%) as an off-white solid: **MP** decomposes at 260°C; **IR** (cm^{-1}) 3119, 3026, 2805, 1748, 1392; **^{1}H NMR** (500 MHz, CD_3OD) δ 7.42 – 7.32 (3H, m, CH-2,3,4), 7.32 – 7.26 (2H, m, CH-1,5), 3.95 (1H, t, *J* = 7.3 Hz, CH-8), 3.17 (1H, dd, *J* = 13.8, 7.3 Hz, CH-7) 3.06 (1H, dd, *J* = 13.7, 7.3 Hz, CH-7); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 129.1, 128.6,

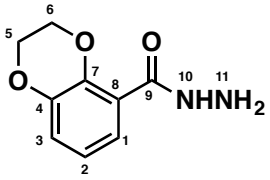
127.3, 53.6, 37.7; **HRMS** m/z (ESI^+) calcd. $C_9H_{13}N_3O$ $[M+H]^+$ requires 180.1131, found 180.1104.

Compound AE18 (H14): (2*S*)-2-amino-4-methylpentanehydrazide



t-Butyl *N*-[(1*S*)-1-(hydrazinecarbonyl)-3-methylbutyl]carbamate **AE09** (1.0 eq., 100 mg, 0.41 mmol) was stirred in 1 M HCl (2 ml) at room temperature overnight. The solvent was removed *in vacuo* to give the target compound **AE18** (103 mg, 0.40 mmol, 99%) as a light yellow oil: **IR** (cm^{-1}) 3115, 3011, 2959, 2805, 1753, 1692, 1670, 1391; **^{1}H NMR** (500 MHz, CD_3OD) δ 3.84 (1H, t, *J* = 7.1 Hz, CH-5), 1.86 – 1.74 (1H, m, CH-3), 1.74 – 1.63 (2H, m, CH_2-4), 1.05 – 0.98 (6H, m, CH_3-1,2); **^{13}C NMR** (126 MHz, CD_3OD) δ 168.5, 50.6, 40.2, 24.1, 21.4, 21.1; **HRMS** m/z (ESI^+) calcd. $C_6H_{15}N_3O$ $[M+Na]^+$ requires 168.1107, found 168.1097.

Compound AE19 (H16): 2,3-dihydrobenzo-b[1,4]-dioxine-5-carbohydrazide



Hydrazine solution (24-26% in water, 5.0 eq., 2.40 g, 12.0 mmol) was added to a solution of ethyl-2,3-dihydrobenzo-1,4-dioxine-5-carboxylate (1.0 eq, 0.50 g, 2.4 mmol) in ethanol (10 mL) and heated to reflux overnight. The solution was then concentrated *in vacuo* to yield the target compound **AE19** (0.41 g, 2.2 mmol, 90%) as an off-white solid: **MP** 70-72 °C; **IR** (cm^{-1}) 2982, 1721, 1468, 1252; **^{1}H NMR** (500 MHz, DMSO) δ 7.21 (1H, dd, *J* = 8.1, 1.6 Hz, CH-3), 7.05 (1H, dd, *J* = 8.1, 1.6 Hz, CH-5), 6.87 (1H, dd, *J* = 8.1 Hz, CH-4), 4.33-4.21 (4H, m, CH_2-5, -6); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 165.0, 143.9, 143.2, 123.2, 122.5, 120.6, 120.2, 64.1, 63.6; **HRMS** m/z (ESI^+) calcd. $C_9H_{10}N_2O_3$ $[M+H]^+$ requires 195.0764, found 195.0754.

5 Summary and Outlook

Dynamic combinatorial chemistry is a powerful tool to identify new ligands for biological targets. The scope of suitable reversible reactions that proceed under thermodynamic control in physiological conditions has been gradually expanded over the last decades, however DCC has thus far failed to gain traction as a technique appropriate for drug discovery in the pharmaceutical industry. The constraints placed on library size by validated analytical techniques, and the effort-intensive reality of this academically elegant concept have not allowed DCC to develop into a broad-platform technique to compete with the high-throughput screening campaigns favoured by medicinal chemists.

While the scope of reversible reactions that are suitable for protein-templated DCC has been steadily growing over the last two decades, the methods used to analyse these elegant experiments have not developed in a way that has made them attractive for industrial drug discovery. In this thesis, we have attempted to address this shortcoming by exploring DCL analysis methods that have both a broad utility and the potential to analyse larger more complex libraries of compounds.

In the first chapter of this thesis, we demonstrated the power of ^{19}F NMR for screening DCLs. This non-invasive analysis method was used in a proof-of-concept study to identify inhibitors of the antimicrobial target FabH from *E. coli*. In recent years this analysis method has been adopted and developed for the screening of fragment libraries by the pharmaceutical industry. While DCL size and complexity remain limited by spectral overlap, it is our hope that the application of ^{19}F NMR to protein-templated DCC will evolve in line with its use in fragment-based drug discovery.

In the second chapter, we investigated the use of a known antimicrobial compound, HR45, as a core scaffold for a DCL targeting FabH. In doing so, we identified the inhibitory mode of action of HR45 on FabH from *S. aureus*. We observed by ESI-MS

that the compound formed a covalent adduct with a catalytically essential cysteine residue (Cys112). We determined that the failure of previous investigations to capture this adduct was due to the HR45-induced degradation of Cys112 to dehydroalanine under tryptic digest conditions. The deduced mode of action rendered HR45 unsuitable for our DCC approach, but our detailed mechanistic study will provide insight in the design of new antimicrobials.

In the third chapter, we demonstrated the use of ESI-MS to analyse complex DCLs. Using a 28-member DCL we identified hits against the recently validated antimicrobial target enzyme BioA from *M. tuberculosis* and *E. coli*. X-ray crystallography allowed us to demonstrate the binding mode of one of our hits in the *M. tuberculosis* isoform, information that will be useful in the development of future BioA inhibitors. A collaboration with Dr Eric Brown (McMaster) determined that two of our hits demonstrated *in vivo* activity against a selection of pathogenic strains.

Since its advent, the DCC concept has developed in many different directions. The primal evolutionary concepts at the core of the technique provide intrigue for exploring different chemical systems, and elegant experiments can give us the most basic insight into the mysteries of cellular processes. However, only when sophisticated screening methods are developed in line with the requirements of pharmaceutical industry, will the latent power of DCC be harnessed for drug discovery.

6 Appendix 1 Protein amino acid sequences

ecFabH/pET-28a
predicted MW (ExPASy): 35678.45 Da
MGSSHHHHHHSSGLVPRGSHMYTKIIGTGSYLPEQVRTNADLEKMVDTSDEWIVTRTGIRER
HIAAPNETVSTMGFEAATRAIEMAGIEKDQIGLIVVATTSATHAFPSAACQIQSMLGIKGCPAF
DVAAACAGFTYALSVADQYVKSGAVKYALVVGSDVLARTCDPTDRGTIIIFGDGAGAAVLAASE
EPGIISTHLHADGSYGELLTLPNADRVNPENSIHLTMAGNEVFKVAVTELAHIVDETLAANNLD
RSQLDWLVPHQANLRIISATAKKLGMSMDNVVVTLDRHGNTSAASVPCALDEAVRDGRIKPG
QLVLLEAFGGGFTWGSALVRF

saFabH/pET-28a
predicted MW (ExPASy): 36042.77 Da
MGSSHHHHHHSSGLVPRGSHMNVGIKGFGAYAPEKIIDNAYFEQFLDTSDEWISKMTGIKER
HWADDDQDTSDLAYEASVKAIADAGIQPEDIDMIIVATATGDMPFPTVANMLQERLGTGKV
ASMDQLAACSGFMYSMITAKQYVQSGDYHNILVVGADKLSKITDLTDRSTAVLFGDGAGAVII
GEVSEGRGIISYEMGSDGTGGKHLYLDKDTGKLKMNGREVFKFAVRIMGDASTRVVEKANLTS
DDIDLFIPHQANIRIMESARERLGISKDKMSVSVNKYGNTSAASIPLSIDQELKNGKLKDDDTIVL
VGFGGGLTWGAMTIKWGK

saFabH/pET-HISTEV (cleavable tag underlined)
predicted MW (ExPASy): 38346.38 Da (35378.21 Da cleaved)
MSYYHHHHHHDYDIPTTENLYFQGAMGNVGIKGFGAYAPEKIIDNAYFEQFLDTSDEWISKM
TGIKERHWADDDQDTSDLAYEASVKAIADAGIQPEDIDMIIVATATGDMPFPTVANMLQERL
GTGKVASMDQLAACSGFMYSMITAKQYVQSGDYHNILVVGADKLSKITDLTDRSTAVLFGDG
AGAVIIGEVSEGRGIISYEMGSDGTGGKHLYLDKDTGKLKMNGREVFKFAVRIMGDASTRVVE
KANLTSDDIDLFIPHQANIRIMESARERLGISKDKMSVSVNKYGNTSAASIPLSIDQELKNGKLKD
DDTIVLVGFGGGLTWGAMTIKLQITRAPPPPPLRSGC

mtBioA/pUC-19
predicted MW (ExPASy): 48482.68 Da
MGSSHHHHHHSSGLVPRGSHMAAATGGLTPEQIIAVDGAHLWHPYSSIGREAVSPVVAVAA
HGAWLTLIRDGQPIEVLDAMSSWWTAIHGHGHPALDQALTTQLRVMNHVMFGGLTHEPAA
RLAKLLVDITPAGLDTVFFSDSGSVSVEVAAKMALQYWRGRGLPGKRRLMTWRGGYHGDTFL
AMSICDPHGGMHSLWTDVLAAQVFAPQVPRDYDPAYSAAFEAQLAQHAGELAAVVVEPVV
QGAGGMRFHDPRYLHDLRDICRRYEVLLIFDEIATGFGRTGALFAADHAGVSPDIMCVGKALT
GGYLSLAATLCTADVAHTISAGAAGALMHGPTFMANPLACAVSVASVELLLGQDWRTRITELA
AGLTAGLDTARALPAVTDVRVCGAIGVIECDRPVDLAVATPAALDRGVWLRPFRNLVYAMPP
YICTPAEITQITSAMVEVARLVGSLP

mtBioA/pUC-19 K283A mutant
predicted MW (ExPASy): 48425.58 Da (purchased fragment underlined)
MGSSHHHHHHSSGLVPRGSHMAAATGGLTPEQIIAVDGAHLWHPYSSIGREAVSPVVAVAA
HGAWLTLIRDGQPIEVLDAMSSWWTAIHGHGHPALDQALTTQLRVMNHVMFGGLTHEPAA
RLAKLLVDITPAGLDTVFFSDSGSVSVEVAAKMALQYWRGRGLPGKRRLMTWRGGYHGDTFL
AMSICDPHGGMHSLWTDVLAAQVFAPQVPRDYDPAYSAAFEAQLAQHAGELAAVVVEPVV

QGAGGMRFHDPRYLHDLRDICRRYEVLLIFDEIATGFGRTGALFAADHAGVSPDIMCVG**A**ALT
GGYLSLAATLCTADVAHTISAGAAGALMHGPTFMANPLACAVSVASVELLLGQDWRTRITELA
AGLTAGLDTARALPAVTDVRVCGAIGVIECDRPVDLAVATPAALDRGVWLRPFRNLVYAMPP
YICTPAEITQITSAMVEVARLVGSLP

ecBioA/pET-28a
predicted MW (ExPASy): 49498.92 Da
MGSSHHHHHHSSGLVPRGSHMTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDG
RRLVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNSM
HSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAGGMR
MYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKALTGGTMTLSATL
TTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQQVADIEVQLREQLAP
ARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRPFGKLIYLMPPYIILPQQLQR
LTAAVNRAVQDETFFCQ

7 References

1 R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo, *Nature*, 1991, **354**, 84–86.

2 Á. Furka, F. Sebestyén, M. Asgedom and G. Dibó, *Int. J. Pept. Protein Res.*, 1991, **37**, 487–493.

3 S. R. Beeren and J. K. M. Sanders, in *Dynamic Combinatorial Chemistry*, eds. J. N. H. Reek and S. Otto, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2010, vol. 8, pp. 1–22.

4 I. Huc and J. M. Lehn, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 2106–2110.

5 S. J. Rowan, D. G. Hamilton, P. A. Brady and J. K. M. Sanders, *J. Am. Chem. Soc.*, 1997, **119**, 2578–2579.

6 P. A. Brady and J. K. M. Sanders, *Chem. Soc. Rev.*, 1997, **26**, 327.

7 S. Fujii and J. M. Lehn, *Angew. Chemie - Int. Ed.*, 2009, **48**, 7635–7638.

8 J. W. Sadownik and D. Philp, *Angew. Chemie - Int. Ed.*, 2008, **47**, 9965–9970.

9 J. W. Sadownik and D. Philp, *Org. Biomol. Chem.*, 2015, **13**, 10392–10401.

10 R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das and R. V Ulijn, *Nat Nano*, 2009, **4**, 19–24.

11 R. F. Ludlow and S. Otto, *Components*, 2008, **3**, 12218–12219.

12 S. Hagihara, H. Tanaka and S. Matile, *J. Am. Chem. Soc.*, 2008, **130**, 5656–5657.

13 A. Herrmann, *Chem. Soc. Rev.*, 2014, **43**, 1899–933.

14 M. Mondal and A. K. H. Hirsch, *Chem. Soc. Rev.*, 2014, **44**, 2455–2488.

15 R. F. Ludlow and S. Otto, *Chem. Soc. Rev.*, 2008, **37**, 101–108.

16 K. Severin, *Chem. - A Eur. J.*, 2004, **10**, 2565–2580.

17 O. Ramström and J.-M. Lehn, *Nat. Rev. Drug Discov.*, 2002, **1**, 26–36.

18 A. J. Clipson, V. T. Bhat, I. McNae, A. M. Caniard, D. J. Campopiano and M. F. Greaney, *Chem. - A Eur. J.*, 2012, **18**, 10562–10570.

19 B. Shi, R. Stevenson, D. J. Campopiano and M. F. Greaney, *J. Am. Chem. Soc.*, 2006, **128**, 8459–8467.

20 V. T. Bhat, A. M. Caniard, T. Luksch, R. Brenk, D. J. Campopiano and M. F. Greaney, *Nat. Chem.*, 2010, **2**, 490–497.

21 O. Ramström and J. M. Lehn, *Chembiochem*, 2000, **1**, 41–48.

22 O. Ramström, S. Lohmann, T. Bunyapaiboonsri and J. M. Lehn, *Chem. - A Eur. J.*, 2004, **10**, 1711–1715.

23 S. Zameo, B. Vauzeilles and J. M. Beau, *Angew. Chemie - Int. Ed.*, 2005, **44**, 965–969.

24 A. Bugaut, J. J. Toulmé and B. Rayner, *Angew. Chemie - Int. Ed.*, 2004, **43**, 3144–3147.

25 A. Bugaut, K. Bathany, J. M. Schmitter and B. Rayner, *Tetrahedron Lett.*, 2005, **46**, 687–690.

26 B. Shi and M. F. Greaney, *Chem. Commun. (Camb).*, 2005, 886–888.

27 M. Hochgürtel, H. Kroth, D. Piecha, M. W. Hofmann, C. Nicolau, S. Krause, O. Schaaf, G. Sonnenmoser and A. V Eliseev, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 3382–3387.

28 M. Hochgürtel, R. Biesinger, H. Kroth, D. Piecha, M. W. Hofmann, S. Krause, O. Schaaf, C. Nicolau and A. V. Eliseev, *J. Med. Chem.*, 2003, **46**, 356–358.

29 Z. Fang, W. He, X. Li, Z. Li, B. Chen, P. Ouyang and K. Guo, *Bioorganic Med. Chem. Lett.*, 2013, **23**, 5174–5177.

30 M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe and A. K. H. Hirsch, *Angew. Chemie - Int. Ed.*, 2014, **53**, 3259–3263.

31 R. Caraballo, H. Dong, J. P. Ribeiro, J. Jiménez-Barbero and O. Ramström, *Angew. Chemie - Int. Ed.*, 2010, **49**, 589–593.

32 M. S. Congreve, D. J. Davis, L. Devine, C. Granata, M. O'Reilly, P. G. Wyatt and H. Jhoti, *Angew. Chemie Int. Ed.*, 2003, **42**, 4479–4482.

33 M. Mondal, D. E. Groothuis and A. K. H. Hirsch, *Med. Chem. Commun.*, 2015, **6**, 1267–1271.

34 M. F. Schmidt, A. Isidro-Llobet, M. Lisurek, A. El-Dahshan, J. Tan, R. Hilgenfeld and J. Rademann, *Angew. Chemie Int. Ed.*, 2008, **47**, 3275–3278.

35 M. F. Schmidt and J. Rademann, *Trends Biotechnol.*, 2009, **27**, 512–521.

36 T. Bunyapaiboonsri, O. Ramström, S. Lohmann, J. M. Lehn, L. Peng and M. Goeldner, *Chembiochem*, 2001, **2**, 438–444.

37 T. Bunyapaiboonsri, H. Ramström, O. Ramström, J. Haiech and J.-M. Lehn, *J. Med. Chem.*, 2003, **46**, 5803–5811.

38 E. C. Y. Woon, M. Demetriades, E. A. L. Bagg, W. Aik, S. M. Krylova, J. H. Y. Ma, M. Chan, L. J. Walport, D. W. Wegman, K. N. Dack, M. A. McDonough, S. N. Krylov and C. J. Schofield, *J. Med. Chem.*, 2012, **55**, 2173–2184.

39 B. M. R. Liénard, N. Selevsek, N. J. Oldham and C. J. Schofield, *ChemMedChem*, 2007, **2**, 175–179.

40 M. Demetriades, I. K. H. Leung, R. Chowdhury, M. C. Chan, M. A. McDonough, K. K. Yeoh, Y. M. Tian, T. D. W. Claridge, P. J. Ratcliffe, E. C. Y. Woon and C. J. Schofield, *Angew. Chemie - Int. Ed.*, 2012, **51**, 6672–6675.

41 M. Sindelar and K. T. Wanner, *ChemMedChem*, 2012, **7**, 1678–1690.

42 S.-A. Poulsen, *J. Am. Soc. Mass Spectrom.*, 2006, **17**, 1074–1080.

43 S. A. Rotenberg, T. Calogeropoulou, J. S. Jaworski, I. B. Weinstein and D. Rideout, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 2490–4.

44 G. R. L. Cousins, S.-A. Poulsen and J. K. M. Sanders, *Chem. Commun.*, 1999, 1575–1576.

45 J. Kalia and R. T. Raines, *Angew. Chemie - Int. Ed.*, 2008, **47**, 7523–7526.

46 A. Dirksen, S. Dirksen, T. M. Hackeng and P. E. Dawson, *J. Am. Chem. Soc.*, 2006, **128**, 15602–15603.

47 E. H. Cordes and W. P. Jencks, *J. Am. Chem. Soc.*, 1962, **84**, 826–831.

48 D. A. Rodrigues, G. À. Ferreira-Silva, A. C. S. Ferreira, R. A. Fernandes, J. K. Kwee, C. M. R. Sant'Anna, M. Ionta and C. A. M. Fraga, *J. Med. Chem.*, 2016, **59**, 655–670.

49 S. Rollas and Ş. G. Küçükgüzel, *Molecules*, 2007, **12**, 1910–1939.

50 M. Carcelli, D. Rogolino, A. Gatti, L. De Luca, M. Sechi, G. Kumar, S. W. White, A. Stevaert and L. Naesens, *Sci. Rep.*, 2016, **6**, 31500.

51 R. Dai, D. J. Wilson, T. W. Geders, C. C. Aldrich and B. C. Finzel, *ChemBioChem*, 2014, **15**, 575–586.

52 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discov.*, 2016, **15**, 605–619.

53 A. Ganesan, *Angew. Chemie Int. Ed.*, 1998, **37**, 2828–2831.

54 S.-A. Poulsen and H. Vu, in *Dynamic Combinatorial Chemistry*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010, pp. 201–228.

55 A. Fleming, *Br J Exp Pathol*, 1929, 10, 226–236.

56 K. Lewis, *Nat. Rev. Drug Discov.*, 2013, **12**, 371–387.

57 A. Schatz, E. Bugle and S. A. Waksman, *Exp. Biol. Med.*, 1944, **55**, 66–69.

58 J. S. Wolfson and D. C. Hooper, *Clin. Microbiol. Rev.*, 1989, **2**, 378–424.

59 K. Andries, *Science (80-.).*, 2005, **307**, 223–227.

60 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 1997, **23**, 3–25.

61 X. Li and H. Nikaido, *Efflux-mediated drug resistance in bacteria*, 2004, vol. 69.

62 A. E. Clatworthy, E. Pierson and D. T. Hung, *Nat. Chem. Biol.*, 2007, **3**, 541–548.

63 H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg and J. Bartlett, *Clin. Infect. Dis.*, 2009, **48**, 1–12.

64 K. Hiramatsu, H. Hanaki, T. Ino, K. Yabuta, T. Oguri and F. C. Tenover, *J Antimicrob Chemother*, 1997, **40**, 135–136.

65 T. L. Smith, M. L. Pearson, K. R. Wilcox, C. Cruz, M. V. Lancaster, B. Robinson-Dunn, F. C. Tenover, M. J. Zervos, J. D. Band, E. White and W. R. Jarvis, *N. Engl. J. Med.*, 1999, **340**, 493–501.

66 P. G. Higgins, C. Dammhayn, M. Hackel and H. Seifert, *J. Antimicrob. Chemother.*, 2009, **65**, 233–238.

67 S. W. White, J. Zheng, Y.-M. Zhang and Rock, *Annu. Rev. Biochem.*, 2005, **74**, 791–831.

68 J. Beld, D. J. Lee and M. D. Burkart, *Mol. BioSyst.*, 2015, **11**, 38–59.

69 J. W. Campbell and J. E. Cronan Jr., *Annu. Rev. Microbiol.*, 2001, **55**, 305–32.

70 C. Nguyen, R. W. Haushalter, D. J. Lee, P. R. L. Markwick, J. Bruegger, G.

Caldara-Festin, K. Finzel, D. R. Jackson, F. Ishikawa, B. O'Dowd, J. A. McCammon, S. J. Opella, S.-C. Tsai and M. D. Burkart, *Nature*, 2013, **505**, 427–431.

71 D. J. Prescott and P. R. Vagelos, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1972, **36**, 269–311.

72 S. Jackowski, C. M. Murphys, J. E. Cronan and C. Rock, *J. Biol. Chem.*, 1989, **264**, 7624–7629.

73 C. Y. Lai and J. E. Cronan, *J. Biol. Chem.*, 2003, **278**, 51494–51503.

74 S. S. Khandekar, R. A. Daines and J. T. Lonsdale, *Curr. Protein Pept. Sci.*, 2003, **4**, 21–29.

75 C. Davies, R. J. Heath, S. W. White and C. O. Rock, *Structure*, 2000, **8**, 185–195.

76 S. S. Khandekar, D. R. Gentry, G. S. Van Aller, P. Warren, H. Xiang, C. Silverman, M. L. Doyle, P. A. Chambers, A. K. Konstantinidis, M. Brandt, R. A. Daines and J. T. Lonsdale, *J. Biol. Chem.*, 2001, **276**, 30024–30030.

77 J. Y. Lee, K. W. Jeong, J. U. Lee, D. Il Kang and Y. Kim, *Bioorganic Med. Chem.*, 2009, **17**, 1506–1513.

78 K.-H. H. K. Choi, R. J. Heath and C. O. Rock, *J. Bacteriol.*, 2000, **182**, 365–370.

79 X. Qiu, A. E. Choudhry, C. a Janson, M. Grooms, R. a Daines, J. T. Lonsdale and S. S. Khandekar, *Protein Sci.*, 2005, **14**, 2087–2094.

80 J. H. Pereira, E. B. Goh, J. D. Keasling, H. R. Beller and P. D. Adams, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2012, **68**, 1320–1328.

81 X. Qiu, C. a. Janson, A. K. Konstantinidis, S. Nwagwu, C. Silverman, W. W. Smith, S. Khandekar, J. Lonsdale and S. S. Abdel-Meguid, *J. Biol. Chem.*, 1999, **274**, 36465–36471.

82 X. Qiu, C. a Janson, W. W. Smith, M. Head, J. Lonsdale and A. K. Konstantinidis, *J. Mol. Biol.*, 2001, **307**, 341–356.

83 R. A. Daines, I. Pendrak, K. Sham, G. S. Van Aller, A. K. Konstantinidis, J. T. Lonsdale, C. A. Janson, X. Qiu, M. Brandt, S. S. Khandekar, C. Silverman and M. S. Head, *J. Med. Chem.*, 2003, **46**, 5–8.

84 Z. Nie, C. Perretta, J. Lu, Y. Su, S. Margosiak, K. S. Gajiwala, J. Cortez, V. Nikulin, K. M. Yager, K. Appelt and S. Chu, *J. Med. Chem.*, 2005, **48**, 1596–1609.

85 J. Wang, S. M. Soisson, K. Young, W. Shoop, S. Kodali, A. Galgoci, R. Painter, G. Parthasarathy, Y. S. Tang, R. Cummings, S. Ha, K. Dorso, M. Motyl, H. Jayasuriya, J. Ondeyka, K. Herath, C. Zhang, L. Hernandez, J. Allocco, A. Basilio, J. R. Tormo, O. Genilloud, F. Vicente, F. Pelaez, L. Colwell, S. H. Lee, B. Michael, T. Felcetto, C. Gill, L. L. Silver, J. D. Hermes, K. Bartizal, J. Barrett, D. Schmatz, J. W. Becker, D. Cully and S. B. Singh, *Nature*, 2006, **441**, 358–361.

86 J. Wang, S. Kodali, S. H. Lee, A. Galgoci, R. Painter, K. Dorso, F. Racine, M. Motyl, L. Hernandez, E. Tinney, S. L. Colletti, K. Herath, R. Cummings, O. Salazar, I. González, A. Basilio, F. Vicente, O. Genilloud, F. Pelaez, H. Jayasuriya, K. Young, D. F. Cully and S. B. Singh, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7612–7616.

87 M. M. Alhamadsheh, F. Musayev, A. a. Komissarov, S. Sachdeva, H. T. Wright, N. Scarsdale, G. Florova and K. a. Reynolds, *Chem. Biol.*, 2007, **14**, 513–524.

88 R. J. Heath, S. W. White and C. O. Rock, *Appl. Microbiol. Biotechnol.*, 2002, **58**, 695–703.

89 J. B. Parsons and C. O. Rock, *Curr. Opin. Microbiol.*, 2011, **14**, 544–549.

90 S. Brinster, G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss and C. Poyart, *Nature*, 2009, **458**, 83–86.

91 W. Balemans, N. Lounis, R. Gilissen, J. Guillemont, K. Simmen, K. Andries and A. Koul, *Nature*, 2010, **463**, E3; discussion E4.

92 S. Brinster, G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss and C. Poyart, *Nature*, 2010, **463**, E4–E4.

93 X. He, A. M. Reeve, U. R. Desai, G. E. Kellogg and K. A. Reynolds, *Antimicrob. Agents Chemother.*, 2004, **48**, 3093–3102.

94 X. L. Wang, Y. Bin Zhang, J. F. Tang, Y. S. Yang, R. Q. Chen, F. Zhang and H. L. Zhu, *Eur. J. Med. Chem.*, 2012, **57**, 373–382.

95 H. Q. Li, Y. Luo and H. L. Zhu, *Bioorganic Med. Chem.*, 2011, **19**, 4454–4459.

96 L. Shi, R.-Q. Fang, Z.-W. Zhu, Y. Yang, K. Cheng, W.-Q. Zhong and H.-L. Zhu, *Eur. J. Med. Chem.*, 2010, **45**, 4358–4364.

97 P.-C. Lv, K.-R. Wang, Y. Yang, W.-J. Mao, J. Chen, J. Xiong and H.-L. Zhu, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 6750–6754.

98 Y. Li, C. P. Zhao, H. P. Ma, M. Y. Zhao, Y. R. Xue, X. M. Wang and H. L. Zhu, *Bioorganic Med. Chem.*, 2013, **21**, 3120–3126.

99 F. Zhang, Q. Wen, S.-F. Wang, B. Shahla Karim, Y.-S. Yang, J.-J. Liu, W.-M. Zhang and H.-L. Zhu, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 90–5.

100 X. Song, Y. Yang, J. Zhao and Y. Chen, *Chem. Pharm. Bull.*, 2014, **62**, 1110–1118.

101 A. Vulpetti and C. Dalvit, *ChemMedChem*, 2013, **8**, 2057–2069.

102 C. Dalvit, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2007, **51**, 243–271.

103 C. Dalvit, P. E. Fagerness, D. T. A. Hadden, R. W. Sarver and B. J. Stockman, *J. Am. Chem. Soc.*, 2003, **125**, 7696–7703.

104 C. Dalvit and A. Vulpetti, *Magn. Reson. Chem.*, 2012, **50**, 592–597.

105 K. S. Gajiwala, S. Margosiak, J. Lu, J. Cortez, Y. Su, Z. Nie and K. Appelt, *FEBS Lett.*, 2009, **583**, 2939–2946.

106 D. K. Kölmel and E. T. Kool, *Chem. Rev.*, 2017, **117**, 10358–10376.

107 A. R. Blanden, K. Mukherjee, O. Dilek, M. Loew and S. Bane, *Bioconjug. Chem.*, 2011, **22**, 1954–1961.

108 P. Crisalli and E. T. Kool, *J Org Chem*, 2013, **78**, 1184–1189.

109 J. T. Gerig, *Methods Enzymol.*, 1989, **177**, 3–23.

110 T. A. Baillie, *Angew. Chemie - Int. Ed.*, 2016, **55**, 13408–13421.

111 J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat. Rev. Drug Discov.*, 2011, **10**, 307–317.

112 J. R. Vane and R. M. Botting, *Thromb. Res.*, 2003, **110**, 255–258.

113 G. Roth and P. W. Majerus, *J. Clin. Invest.*, 1975, **56**, 624–632.

114 T. A. Baillie, *Chem. Res. Toxicol.*, 2006, **19**, 889–893.

115 A. J. T. Smith, X. Zhang, A. G. Leach and K. N. Houk, *J. Med. Chem.*, 2009, **52**, 225–233.

116 B. Drukarch, E. Schepens, J. C. Stoof and C. H. Langeveld, *Eur. J. Pharmacol.*, 1997, **329**, 259–262.

117 G. Fischer, *Recent Progress in 1,2-Dithiole-3-thione Chemistry*, Elsevier, 2013, vol. 109.

118 Y. Zhang and G. B. Gordon, *Mol. Cancer Ther.*, 2004, **3**, 885–93.

119 A. C. Price, K. H. Choi, R. J. Heath, Z. Li, S. W. White and C. O. Rock, *J. Biol. Chem.*, 2001, **276**, 6551–6559.

120 I. Nishida, A. Kawaguchi and M. Yamada, *J. Biol. Chem.*, 1986, **1611**, 1605–1611.

121 X. He and K. A. Reynolds, *Antimicrob. Agents Chemother.*, 2002, **46**, 1310–1318.

122 X. He, J. P. Mueller and K. A. Reynolds, *Anal. Biochem.*, 2000, **282**, 107–114.

123 X. He, A. M. Reeve, U. R. Desai, E. Glen, K. A. Reynolds and G. E. Kellogg, *Antimicrob. Agents Chemother.*, 2004, **48**, 3093.

124 E. Z. Baum, D. A. Montenegro, L. Licata, I. Turchi, G. C. Webb, B. D. Foleno and K. Bush, *Antimicrob. Agents Chemother.*, 2001, **45**, 3182–3188.

125 F. Boberg, *Justus Liebigs Ann. Chem.*, 1965, **178**, 132–148.

126 Y. S. Yang, F. Zhang, C. Gao, Y. Bin Zhang, X. L. Wang, J. F. Tang, J. Sun, H. Bin Gong and H. L. Zhu, *Bioorganic Med. Chem. Lett.*, 2012, **22**, 4619–4624.

127 Y. Li, Y. Luo, Y. Hu, D. Di Zhu, S. Zhang, Z. J. Liu, H. Bin Gong and H. L. Zhu, *Bioorganic Med. Chem.*, 2012, **20**, 4316–4322.

128 Z. L. Li, Q. S. Li, H. J. Zhang, Y. Hu, D. Di Zhu and H. L. Zhu, *Bioorganic Med. Chem.*, 2011, **19**, 4413–4420.

129 X. Song, Y. Yang, J. Zhao and Y. Chen, *Chem. Pharm. Bull.*, 2014, **62**, 1110–1118.

130 J. B. Baell and G. A. Holloway, *J. Med. Chem.*, 2010, **53**, 2719–2740.

131 K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V Kamath and J. G. Stroh, *Anal. Biochem.*, 1999, **267**, 169–184.

132 J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-

González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield and B. G. Davis, *Chem. Sci.*, 2011, **2**, 1666.

133 T. Kortemme and T. E. Creighton, *J. Mol. Biol.*, 1995, **253**, 799–812.

134 K. Boukebbous, E. A. Laifa, A. De Mallmann and M. Taoufik, *IUCrData*, 2016, **1**, x161820.

135 A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2012, **68**, 431–440.

136 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 486–501.

137 G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 355–367.

138 World Health Organisation, *Global Tuberculosis Report 2016*, 2016.

139 H. C. Mwandumba, D. G. Russell, M. H. Nyirenda, J. Anderson, S. A. White, M. E. Molyneux and S. B. Squire, *J. Immunol.*, 2004, **172**, 4592–4598.

140 United Nations, *HIV Factsheet — 2015 UNAIDS*, 2016.

141 F. Liu, S. Dawadi, K. M. Maize, R. Dai, S. W. Park, D. Schnappinger, B. C. Finzel and C. C. Aldrich, *J. Med. Chem.*, 2017, **60**, 5507–5520.

142 D. G. Russell, C. E. Barry and J. L. Flynn, *Science (80-.).*, 2010, **328**, 852–856.

143 W. Salaemae, A. Azhar, G. W. Booker and S. W. Polyak, *Protein Cell*, 2011, **2**, 691–695.

144 H. M. Said, *J. Nutr.*, 2008, **139**, 158–162.

145 D. A. Rodionov, P. Hebbeln, A. Eudes, J. ter Beek, I. A. Rodionova, G. B. Erkens, D. J. Slotboom, M. S. Gelfand, A. L. Osterman, A. D. Hanson and T. Eitinger, *J. Bacteriol.*, 2009, **191**, 42–51.

146 C. M. Sassetti and E. J. Rubin, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12989–94.

147 S. Woong Park, M. Klotzsche, D. J. Wilson, H. I. Boshoff, H. Eoh, U. Manjunatha, A. Blumenthal, K. Rhee, C. E. Barry, C. C. Aldrich, S. Ehrt and D. Schnappinger, *PLoS Pathog.*, 2011, **7**, e1002264.

148 S. Dey, J. M. Lane, R. E. Lee, E. J. Rubin and J. C. Sacchettini, *Biochemistry*, 2010, **49**, 6746–6760.

149 A. C. Eliot, J. Sandmark, G. Schneider and J. F. Kirsch, *Biochemistry*, 2002, **41**, 12582–12589.

150 S. Mann and O. Ploux, *FEBS J.*, 2006, **273**, 4778–4789.

151 J. Billones, M. C. Carrillo, V. Organo, J. B. Sy, N. A. Clavio, S. J. Macalino, I. Emnacen, A. Lee, P. Ko and G. Concepcion, *Drug Des. Devel. Ther.*, 2017, **11**, 563–574.

152 R. Dai, T. W. Geders, F. Liu, S. W. Park, D. Schnappinger, C. C. Aldrich and B. C. Finzel, *J. Med. Chem.*, 2015, **58**, 5208–5217.

153 A. C. Eliot and J. F. Kirsch, *Annu. Rev. Biochem.*, 2004, **73**, 383–415.

154 J. Sandmark, S. Mann, A. Marquet and G. Schneider, *J. Biol. Chem.*, 2002, **277**, 43352–43358.

155 C. Shi, T. W. Geders, S. W. Park, D. J. Wilson, H. I. Boshoff, O. Abayomi, C. E. Barry, D. Schnappinger, B. C. Finzel and C. C. Aldrich, *J. Am. Chem. Soc.*, 2011, **133**, 18194–18201.

156 P. Holtz and D. Palm, *Pharmacol. Rev.*, 1964, **16**, 113–178.

157 S. W. Park, D. E. Casalena, D. J. Wilson, R. Dai, P. P. Nag, F. Liu, J. P. Boyce, J. A. Bittker, S. L. Schreiber, B. C. Finzel, D. Schnappinger and C. C. Aldrich, *Chem. Biol.*, 2015, **22**, 76–86.

158 S. Zlitni, L. F. Ferruccio and E. D. Brown, *Nat. Chem. Biol.*, 2013, **9**, 796–804.

159 J. Fenn, M. Mann, C. Meng, S. Wong and C. Whitehouse, *Science (80-.).*, 1989, **246**, 64–71.

160 S. A. Hofstadler and K. A. Sannes-Lowery, *Nat. Rev. Drug Discov.*, 2006, **5**, 585–595.

161 A. J. R. Heck, *Nat. Methods*, 2008, **5**, 927–933.

162 C. Schmidt and C. V. Robinson, *FEBS J.*, 2014, **281**, 1950–1964.

163 I. K. H. Leung, T. Brown Jr, C. J. Schofield and T. D. W. Claridge, *Medchemcomm*, 2011, **2**, 390.

164 J. M. Daniel, S. D. Friess, S. Rajagopalan, S. Wendt and R. Zenobi, *Int. J. Mass*

Spectrom., 2002, **216**, 1–27.

165 W. He, Z. Fang, Z. Yang, D. Ji, K. Chen and K. Guo, *RSC Adv.*, 2015, **5**, 23224–23228.

166 Z. Yang, Z. Fang, W. He, Z. Wang, H. Gan, Q. Tian and K. Guo, *Bioorg. Med. Chem. Lett.*, 2016, **26**, 1671–4.

167 S. W. Van Arsdell, J. B. Perkins, R. R. Yocum, L. Luan, C. L. Howitt, N. P. Chatterjee and J. G. Pero, *Biotechnol. Bioeng.*, 2005, **91**, 75–83.

168 T. W. Geders, K. Gustafson and B. C. Finzel, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 2012, **68**, 596–600.

169 J. M. Reyrat and D. Kahn, *Trends Microbiol.*, 2001, **9**, 472–4.

170 L. A. Kelly, S. Mezulis, C. Yates, M. Wass and M. Sternberg, *Nat. Protoc.*, 2015, **10**, 845–858.

171 D. A. Rodionov, A. A. Mironov and M. S. Gelfand, *Genome Res.*, 2002, **12**, 1507–1516.

172 F. Finkenwirth, F. Kirsch and T. Eitinger, *Bioengineered*, 2014, **5**, 129–132.