



**MONASH** University

**Application of orthogonal  
fragment-based design strategies to  
*EcDsbA***

Rebecca Lee Whitehouse

*Bachelor of Pharmaceutical Science Advanced with Honours (Medicinal Chemistry)*

A thesis submitted for the degree of Doctor of Philosophy at  
Monash University in 2021

Department of Medicinal Chemistry  
Faculty of Pharmacy and Pharmaceutical Sciences  
Monash Institute of Pharmaceutical Sciences  
**Monash University**

## Copyright notice

© Rebecca Lee Whitehouse (2021).

*I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.*

## **Declaration**

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Rebecca Whitehouse

Date: 07/04/2021

# Table of Contents

<b>Publications/Presentations during enrolment</b> .....	vii
<b>Acknowledgements</b> .....	ix
<b>Abbreviations</b> .....	xi
<b>Abstract</b> .....	xiv
<b>Chapter 1: Introduction</b> .....	1
1.1 Fragment-based Drug Design (FBDD).....	2
1.1.1 Advantages of fragment-based drug design .....	3
1.1.2 Strategies of fragment-based drug design.....	6
1.2 Covalent inhibitors .....	29
1.2.1 Advantages of covalent inhibition .....	30
1.2.2 Challenges of covalent inhibition .....	32
1.2.3 Inhibition and reactive functional groups .....	33
1.3 Antibacterial resistance and bacterial virulence .....	39
1.4 <i>Escherichia coli</i> DsbA ( <i>EcDsbA</i> ) .....	43
1.4.1 DsbA as an antivirulence target.....	43
1.4.2 Mechanism and catalytic cycle of <i>EcDsbA</i> .....	44
1.4.3 Structure and function of <i>EcDsbA</i> .....	46
1.4.4 Dynamics of <i>EcDsbA</i> .....	54
1.4.5 Small molecule inhibitors of <i>EcDsbA</i> .....	56
1.5 Project hypotheses and aims.....	59
<b>Chapter 2: Characterisation of a mono-cysteine peptide substrate</b> .....	61
2.1 Structural characterisation of peptide binding .....	67
2.1.1 Effects of non-covalent peptides against <i>EcDsbA</i> .....	67
2.1.2 Effects of a covalent peptide against <i>EcDsbA</i> .....	70
2.2 Characterisation of reaction kinetics for peptide oxidation .....	76
2.3 Conclusions and future directions.....	79
<b>Chapter 3: Investigation of covalent thiol fragments</b> .....	81



3.1 Structural analysis of small molecule inhibitors .....	84
3.2 Design of targeted thiol fragment series .....	88
3.3 Parallel synthesis of thiol fragments.....	93
3.3.1 Synthesis of diaryl ether precursors .....	93
3.3.2 Microscale synthesis of thiol functionalised analogues .....	94
3.3.3 Covalent labelling of <i>EcDsbA</i> by unpurified reaction products .....	98
3.4 Synthesis of thiol fragments as pure compounds.....	99
3.5 Conclusions .....	104
<b>Chapter 4: Design and screening of an electrophilic warhead library .....</b>	<b>106</b>
4.1 Electrophilic warhead library design.....	110
4.2 Pure warhead screen.....	112
4.3 Synthesis and characterisation of non-covalent fragment .....	115
4.4 Parallel synthesis of covalent fragment library .....	118
4.5 Crude parallel synthesis library screen .....	123
4.6 Purified compound validation .....	127
4.6.1 Selection of compounds for pure compound screening .....	127
4.6.2 Synthesis and compound solubility .....	128
4.6.3 Pure compound screen .....	130
4.7 Conclusions and future directions .....	136
<b>Chapter 5: Protein hot spot identification by organic solvent screening .....</b>	<b>137</b>
5.1 Computational binding site predictions .....	141
5.1.1 Analysis of FTMap clusters .....	141
5.1.2 Analysis of DoGSiteScorer simulation .....	148
5.1.3 Computational prediction conclusions .....	152
5.2 Solvent library selection.....	153
5.3 X-ray crystallography solvent screen .....	155
5.3.1 Sites occupied by more than one solvent .....	157
5.3.2 Sites occupied by multiple copies of the same solvent .....	160
5.4 <sup>1</sup> H- <sup>15</sup> N HSQC NMR solvent screen .....	162

5.4.1 Hot spots identified by chemical shift perturbation mapping .....	164
5.4.2 Hot spot identification by grid point analysis .....	166
5.4.3 $K_D$ determination of binding hot spots .....	168
5.5 Comparison of hot spot identification techniques .....	170
5.6 Conclusions and future directions .....	175
<b>Chapter 6: Design and screening of a MicroFrag library .....</b>	<b>177</b>
6.1 MicroFrag library design .....	179
6.1.1 Privileged MicroFrag .....	180
6.1.2 MicroFrag library selection .....	183
6.2 X-ray crystallography MicroFrag screen .....	188
6.2.1 Assay optimisation .....	188
6.2.2 MicroFrag screen by X-ray crystallography .....	191
6.3 $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR MicroFrag screen .....	199
6.3.1 Assay optimisation .....	199
6.3.2 MicroFrag screen by $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR .....	202
6.4 Assay comparisons .....	206
6.4.1 Comparison of the $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR and X-ray crystallography MicroFrag screens .....	206
6.4.2 Comparison of MicroFrag and solvent screens .....	210
6.5 Conclusions and future directions .....	215
<b>Chapter 7: Conclusions and future directions .....</b>	<b>218</b>
<b>Chapter 8: Experimental methods .....</b>	<b>226</b>
8.1 Chemistry .....	227
8.1.1 Instrumentation .....	227
8.1.2 Microscale synthesis .....	229
8.1.3 General procedures .....	230
8.1.4 Compound characterisation .....	235
8.2 Small molecule solubility .....	252
8.3 Protein production and purification .....	252

8.4 NMR parameters and sample preparation .....	254
8.4.1 Peptide binding characterisation .....	254
8.4.2 Backbone and aliphatic side chain assignments .....	255
8.4.3 Covalent library screen.....	258
8.4.4 Organic solvent screen.....	258
8.4.5 MicroFrag screen .....	258
8.4.6 Chemical shift perturbations.....	259
8.5 Protein mass spectrometry .....	259
8.6 Crystallisation and X-ray diffraction experiments .....	259
8.6.1 Refinement Statistics – Table 1 Organic solvent screen.....	261
8.7 Computational calculations .....	267
8.7.1 Hot spot calculations .....	267
8.7.2 Covalent molecular docking experiments .....	267
8.8 MicroFrag library design .....	268
8.9 In-house custom scripts.....	270
8.9.1 $^1\text{H}$ - $^{15}\text{N}$ HSQC peak list RMSD alignment.....	270
8.9.2 Grid point clusters for HSQC binding site analysis .....	271
8.10 Multisequence alignment .....	271
<b>References</b> .....	272
<b>Appendices</b> .....	295
Appendix 1: Non-covalent peptide solubility .....	296
Appendix 2: $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra and computational analysis of organic solvents.....	296
Appendix 3: Binding site definitions .....	307
Appendix 4: SMILES list of MicroFrag library.....	325
Appendix 5: $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra of MicroFrag library .....	327

## **Publications/Presentations during enrolment**

Work conducted during this PhD has resulted in a number of publications and presentations to date and these are listed below:

### **Papers:**

Doak BC, Whitehouse RL, Rimmer K, Williams M, Heras B, Caria S, Ilyichova O, Vazirani M, Mohanty B, Harper JB, Scanlon MJ, Simpson JS. Fluoromethylketone-fragment conjugates designed as covalent modifiers of *EcDsbA* are atypical substrates. Manuscript in preparation. 2021

Bentley MR, Ilyichova OV, Wang G, Williams ML, Sharma G, Alwan WS, Whitehouse RL, Mohanty B, Scammells PJ, Heras B, Martin JL, Totsika M, Capuano B, Doak BC, Scanlon MJ. Rapid Elaboration of Fragments into Leads by X-ray Crystallographic Screening of Parallel Chemical Libraries (REFiLX). *J Med Chem*. 2020;63(13):6863-75.

### **Oral presentations:**

#### **3<sup>rd</sup> Fragment-Based Drug Design Down Under (2019)**

“Identification of novel binding sites by high-concentration NMR and X-ray crystallography”

### **Poster presentations:**

#### **15<sup>th</sup> Annual Drug Discovery Chemistry conference (2020)**

“Identification of novel binding sites by high-concentration NMR and X-ray crystallography”

#### **15<sup>th</sup> Annual Drug Discovery Chemistry conference (2020)**

“Identification of selective covalent fragment inhibitors from parallel reaction mixtures”

#### **45<sup>th</sup> Lorne Conference on Protein Structure and Function (2020)**

“Identification of novel binding sites by high-concentration NMR and X-ray crystallography”

**44<sup>th</sup> Lorne Conference on Protein Structure and Function (2019)**

“Design of a “MiniFrag” library for the identification of protein hot spots by X-ray crystallography”

**43<sup>rd</sup> Annual RACI Synthesis Symposium (2018)**

“Optimisation of covalent *EcDsbA* inhibitors by fragment-based drug design”

**42<sup>nd</sup> Annual RACI Synthesis Symposium (2017)**

“Optimisation of *EcDsbA* inhibitors by fragment-based drug design”

**RACI Centenary Chemistry Congress (AIMECS Medicinal chemistry for diverse targets theme) (2017)**

“Optimisation of *EcDsbA* inhibitors by fragment-based drug design”

## Acknowledgements

I would like to whole heartedly thank my supervisors Prof. Martin Scanlon, Dr Bradley Doak and Prof. Peter Scammells. I appreciate you welcoming me into your research group and giving me the opportunity and support to explore all facets of drug design that have ever interested me. Without your guidance, advice and unwavering belief I would never have been able to achieve any of the work in this thesis or be the scientist I am today and for that I will be forever grateful.

I would like to thank all members of the *EcDsbA* team for all of the support, suggestions and comradery through all of the ups and downs of my PhD including Assoc. Prof. Ben Capuano, Dr Biswaranjan Mohanty, Dr Olga Ilyichova, Dr Wesam Alwan, Dr Matthew Bentley, Dr Gaurav Sharma, Dr Martin Williams, Dr Anitha Kopinathan and Yildiz Tasdan. I would like to thank my panel review members Dr Elizabeth Yuriev and Dr Christopher MacRaid for your advice and input regarding my research throughout the course of my PhD. I would like to thank every single one of the Scanlon group members and members of the chemistry lab for their friendship and support. Any time one of you made me laugh or smile it made the science that little bit easier and much more enjoyable.

An immense thank you to Brad for the countless times I have bothered you with questions about everything from chemistry to chemoinformatics to life; you have taught me so much and with such patience. I would also especially like to thank Biswa and Olga for their crash courses in NMR and X-ray crystallography; thank you for the escape from organic chemistry it was honestly career changing.

I would like to thank all of the friends I have made along the way. Shane Dawson, Ana Grasso, Rory Freeman, Kyle Wright, Grace Curtis, Jack Phelps, Ashley Taylor and Stefan Nebl – thank you for laughing with me, for make-believing with me, for celebrating with me, for crying with me and for holding me up all of these years. Life would not be the same without you.

I would like to thank the love of my life Stephanie Trezise. There is not enough room in this thesis to express how grateful for you I am. Thank you for all the support, the unconditional love, for not letting me give up, and for believing in me when I didn't believe in myself. Thank you for everything. There is no way I would have made it this far without you.

I would like to thank my mother Wendy Wright, my sister Dr Sarah Whitehouse, and my grandfather Ken Lee for all of their love and for every sacrifice they have ever made that has led me here. Thank you for encouraging me, pushing me and supporting me. It is all of your values and hard work that have helped me get here.

Thank you all so much. Making you all proud is all I could have hoped for.

## Abbreviations

2D	2-dimensional
3D	3-dimensional
ADMET	Absorption, distribution, metabolism, excretion and toxicity
ATP	Adenosine triphosphate
Boc	<i>tert</i> -Butyloxycarbonyl
<i>Bps</i>	<i>Burkholderia pseudomallei</i>
BQCA	Benzyl quinolone carboxylic acid
CDK2	Cyclin-dependent kinase 2
ClogP	Calculated octanol-water partition co-efficient
CPMG	Carr-Purcell-Meiboom-Gill
CSP	Chemical shift perturbation
CYP51	Lanosterol 14 $\alpha$ -demethylase
DMF	<i>N,N</i> -Dimethylformamide
DMSO	Dimethyl sulfoxide
DoGSS	DoGSiteScorer
Dsb	Disulfide bond
DSF	Differential scanning fluorometry
DSS	Sodium trimethylsilylpropanesulfonate
DTNB	5,5'-Dithio-bis-(2-nitrobenzoic acid)
<i>Ec</i>	<i>Escherichia coli</i>
EDC	<i>N</i> -(3-Dimethylaminopropyl)- <i>N'</i> -ethylcarbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
EtOAc	Ethyl acetate
EtOH	Ethanol
FBDD	Fragment-based drug design
FDA	United States Food and Drug Administration
FMK	Fluoromethylketone
FP	Fluorescence polarisation
GDB	Generated DataBase
GPCR	G-protein coupled receptor
GSH	Glutathione



HAC	Heavy atom count
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HCl	Hydrochloric acid
HPLC	High performance liquid chromatography
HSP90	Heat shock protein 90
HSQC	Heteronuclear single quantum coherence
HTS	High-throughput screening
IC50	Concentration required to produce 50 % inhibition
IPA	Isopropanol
ITC	Isothermal calorimetry
$K_D$	Equilibrium dissociation constant
$K_i$	Equilibrium inhibition constant
$k_{inact}$	Inhibition rate constant
$K_m$	Michaelis constant
$k_{off}$	Dissociation rate constant
$k_{on}$	Association rate constant
LC	Liquid chromatography
LE	Ligand efficiency
LEM	Ligand efficiency metric
$m/z$	Mass to charge ratio
M <sub>1</sub> mAChR	M <sub>1</sub> muscarinic acetylcholine receptor
MCSS	Multiple copy simultaneous search
MeCN	Acetonitrile
MeOH	Methanol
MS	Mass spectrometry
MSCS	Multiple solvent crystal structures
$Nm$	<i>Neisseria meningitidis</i>
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effects
nPMI	Normalised principle moment of inertia
PAINS	Pan-assay interfering compounds
PDB	Protein Data Bank

PDHK	Pyruvate dehydrogenase kinase
PPI	Protein-protein interface
RD	Relaxation-dispersion
RMSD	Root mean square deviation
Ro3	Rule of three
SAR	Structure-activity relationship
SBDD	Structure-based drug design
SMCM	Synthetic and molecular complexity
SPR	Surface plasmon resonance
STD	Saturation transfer difference
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TNB	2-Nitrothiobenzoate
TPSA	Topological polar surface area
UV/Vis	Ultraviolet visible
<i>V. cholerae</i>	<i>Vibrio cholerae</i>
WAC	Weak-affinity chromatography

## Abstract

Fragment-based drug design (FBDD) has become a mainstream alternative to more traditional drug discovery approaches. Fragment screening uses very small molecules, typically containing fewer than 20 heavy atoms. Due to their small size fragments tend to bind to their target proteins with low affinities. As a result, highly sensitive bioassays such as nuclear magnetic resonance (NMR), X-ray crystallography, and mass spectrometry are required for the characterisation of these binding events. Development of the fragment hits with intrinsically low affinities (low millimolar) into high affinity (nanomolar) lead-like compounds typically requires multiple iterative rounds of medicinal chemistry.

A major bottleneck in the rounds of medicinal chemistry is the purification and characterisation of compounds, which is typically conducted prior to testing for every compound that is synthesised. Workflows have recently been developed, which synthesise compounds in parallel microscale reactions and screen them as unpurified products to identify compounds that bind without extensive purification or characterisation. Binding assays to support the testing of unpurified products have been developed for surface plasmon resonance, mass spectrometry and X-ray crystallography. These strategies exploit the protein's ability to isolate the most slowly dissociating component from parallel reaction mixtures. As dissociation is often the key driver of binding affinity, the compounds isolated in this way are usually those that bind with the highest affinity. These approaches have accelerated the design-synthesis-testing cycle and have resulted in the efficient identification in improved fragment analogues. Another emerging strategy for the rapid optimisation of fragment efficacy is the introduction of electrophilic covalent warheads to low affinity fragments. Covalent attachment of the fragments can enhance the inhibition of challenging targets that lack a well-defined small molecule binding pocket. Structure-based drug design (SBDD) is used heavily in FBDD to expand fragments within binding sites. Protein interactions and binding hot spots to target by SBDD can be identified by screening high concentrations of low molecular weight probes by X-ray crystallography and NMR. Orthogonal biophysical techniques can reveal complimentary information about different aspects of binding and using these

techniques and design strategies in concert can expedite the development of fragments into lead-like compounds.

The work in this thesis applies these orthogonal approaches of early fragment elaboration to the development of inhibitors of the antivirulence target *Escherichia coli* DsbA (*EcDsbA*). The active site of *EcDsbA* comprises a shallow hydrophobic substrate binding site, which makes it an extremely challenging target for small molecule inhibition. In Chapter 2, the interaction of *EcDsbA* with a mono-cysteine peptide substrate was characterised and the insight obtained through this covalent adduct was used to inform the design of thiol functionalised fragments. Chapter 3 describes efforts to synthesise a series of thiols for testing as *EcDsbA* inhibitors. It was found that these compounds were encumbered by issues in stability and purification and so efforts were undertaken to identify a more tractable covalent warhead. Chapter 4 describes the design of a ‘target agnostic’ library of electrophilic warheads. This library was designed to be suitable for testing against multiple other targets in addition to *EcDsbA*. The library was designed to contain a range of electrophilic warheads, linked via spacers of different lengths, to a reactive functionality which was amenable to coupling the covalent warheads with known non-covalent fragment binders. A library was constructed of a known *EcDsbA* ligand coupled to each of the covalent warhead/linkers. This library was synthesised in parallel and the unpurified products of these reactions were screened by  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) NMR in an effort to identify a selective and stable covalent inhibitor of *EcDsbA*. In a complementary approach, a suite of methods was employed to identify hot spots on the surface of *EcDsbA*. Chapter 5 describes the screening of solvents against *EcDsbA* using NMR and X-ray crystallography. Chapter 6 describes the design of a library of diverse and very low molecular weight probes containing 5 – 8 heavy atoms. These compounds, termed MicroFrag, were screened against *EcDsbA* at high concentrations using X-ray crystallography and NMR and were able to identify conserved residue interactions in multiple binding hot spots. Together, these target agnostic electrophilic warhead and MicroFrag libraries have provided data that can be used to expedite the development of fragment inhibitors against *EcDsbA* and these strategies should be equally applicable to other challenging targets.

# Chapter 1:

# Introduction

## 1.1 Fragment-based Drug Design (FBDD)

High-throughput screening (HTS) is one of the most common methods for generating a starting lead compound for medicinal chemistry campaigns. Although HTS libraries are very large (up to millions of compounds), there are many instances where the screens conducted identify little to no hits and the identified hits often have suboptimal physicochemical properties. Furthermore, purchasing, maintaining and screening these libraries is often beyond the scope of academic research, as it requires dedicated infrastructure and is very costly and time consuming before compound optimisation has even begun. Since its inception in 1996 by Fesik and colleagues <sup>(1)</sup>, fragment-based drug design (FBDD) has rapidly gained popularity and become a mainstream alternative to HTS. The shortcomings of HTS campaigns are addressed in FBDD by screening smaller libraries (up to thousands) of small and simple compounds (“fragments”), which increases the chances of finding a “hit” <sup>(2)</sup>. The diminished size of these fragment hits results in an intrinsically lower affinity than their HTS counterparts and requires a larger design effort to develop lead-like compounds. However, this can afford more control over compound properties during optimisation and can improve the physicochemical property profiles of the resulting lead-like compounds <sup>(3, 4)</sup>. Fragments have been defined in many ways, however, the core guidelines are widely accepted as the Astex “rule of three” (Ro3) <sup>(5, 6)</sup>. These guidelines suggest that fragments have a molecular weight < 300 Da, ≤ 3 hydrogen bond donors, ≤ 3 hydrogen bond acceptors, ≤ 3 rotatable bonds, a ClogP ≤ 3, and a total polar surface area of ≤ 60 Å<sup>2</sup>. While these rules suggest that compounds could be ~ 22 heavy atoms <sup>(7)</sup>, screening libraries are often < 17 heavy atoms and many have begun to shift toward even smaller compounds to conform with the hit rates observed in historical campaigns <sup>(6, 8, 9)</sup>.

Fragment-based strategies have waged successful campaigns against challenging targets some consider “undruggable” such as protein-protein interactions (PPIs), transcription factors, Ras proteins and others <sup>(10-24)</sup>. Furthermore, this approach has already resulted in four fragment-based compounds approved for therapeutic use (erdafitinib <sup>(25)</sup>, pexidartinib <sup>(26)</sup>, vemurafenib <sup>(27)</sup>, venetoclax <sup>(12)</sup>) and over 40 compounds have progressed into clinical trials <sup>(28)</sup>.

### 1.1.1 Advantages of fragment-based drug design

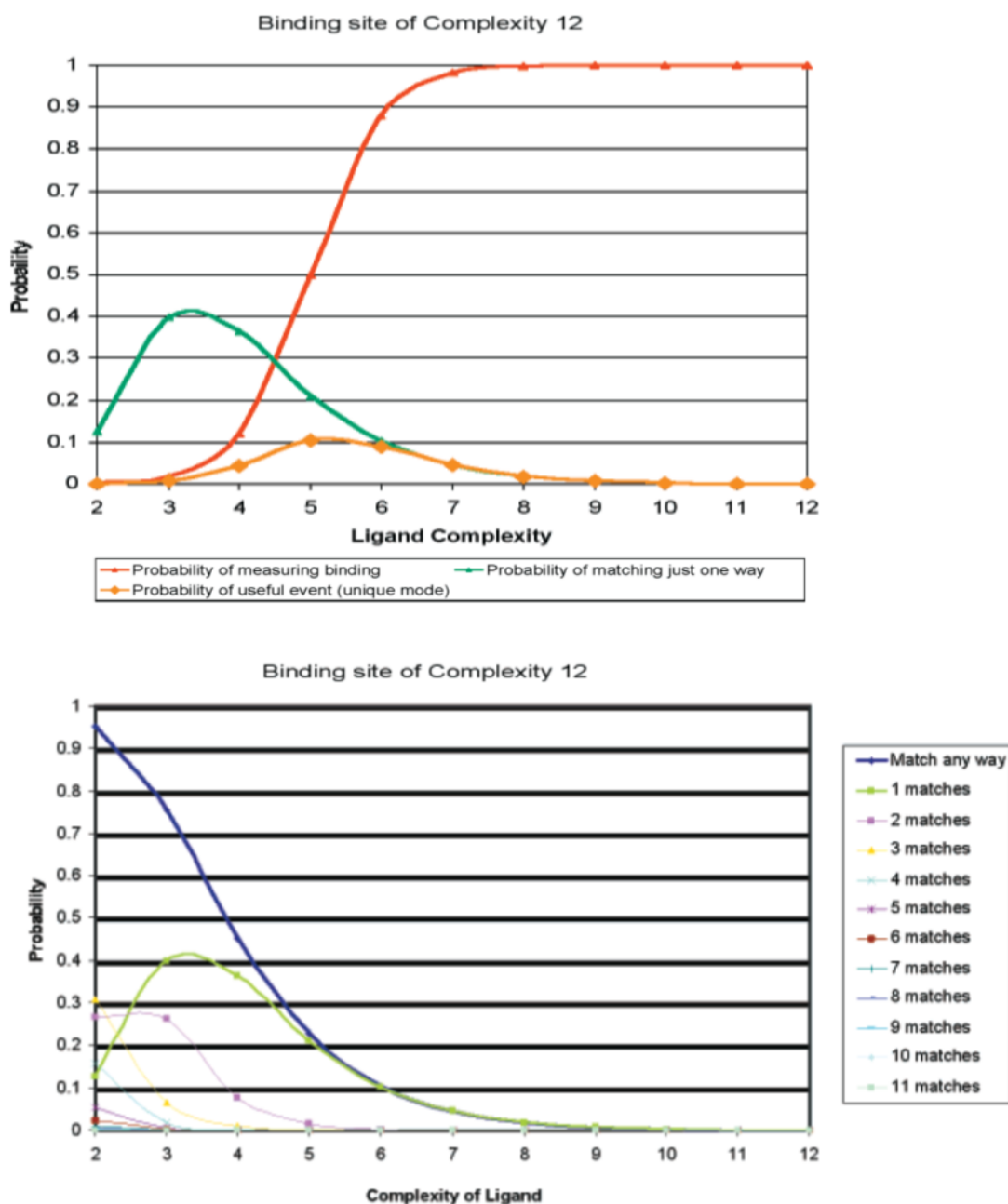
The main ideology of fragment-based design revolves around the ability of smaller compounds to cover chemical space more effectively, which increases the chances of finding a binding site on the protein where there is complementarity between the fragment and the surface of the protein. Chemical space refers to the exhaustive combinations of chemical molecules given a specified set of parameters <sup>(29)</sup>. The size of chemical space has been estimated as  $> 10^{60}$  molecules (heavy atom count  $\leq 30$ ) <sup>(30)</sup>, however imparting stricter parameters to conform with the Lipinski “rule of 5” this chemical space is closer to the order of  $10^{33}$  (heavy atom count  $\approx 36$ ) <sup>(31)</sup>. Similar computational enumerations of atoms to generate all possible structures have been conducted by Reymond and colleagues for the Generated DataBase (GDB) of chemical space for molecules more akin to fragments <sup>(32-36)</sup>. These studies implement very conservative restrictions for inclusion and still claims between 26.4 million (11 heavy atoms <sup>(33)</sup>) and 166 billion (17 heavy atoms <sup>(32)</sup>) molecules for fragment-like chemical space. Analysis of the GDB-13 <sup>(34)</sup> suggests that with the addition of each heavy atom the number of possible compounds increase  $\sim 8$ -fold, however, this is a gross underestimation as the calculation only includes the elements C, N, O, S and Cl <sup>(8, 34, 37)</sup>. Moreover, analysis of the GDB libraries revealed that some known molecules were not predicted in the calculations. Nonetheless, based on these estimations, this equates screening a library of 1000 fragments with 14 heavy atoms to screening  $10^{18}$  HTS compounds of 32 heavy atoms <sup>(37)</sup>. As less compounds are required for these libraries they are more cost effective to both maintain and test, thereby making fragment-based drug design more accessible to academic groups and small pharmaceutical companies.

The higher hit rates for fragment screens is not just due to the statistical coverage of chemical space, but is also largely contributed to the simplicity of the compounds tested. Hann *et al.* have described a model which details the probability of finding useful compounds based on the complexity of the target and ligand, and the sensitivity of the assay being used <sup>(2)</sup>. Using +’s and –’s as minimalistic representations of surface properties of ligands and receptors (topologies, pharmacophores, physicochemical properties), they explored how the number of features (complexity) affects molecular recognition. A range of ligands were compared to a target with a complexity of 12

(12 + or – features) and the number of direct matches between the two were enumerated (Figure 1.1). A “useful event” was used to describe the detectable binding of a singular and unique binding mode. Moreover, when selecting an appropriate compound complexity, one must consider the sensitivity of assays being used to detect the binding event.

As the complexity of a ligand increased, so too did the probability of a “mismatch”, or unfavourable interaction. Consequentially, this resulted in the compound not being identified as binding (as exemplified with the low hit rates of HTS). Conversely, as the ligands complexity decreased it became less likely to be detected, and more likely to display multiple binding modes (as exemplified by solvents, see section 1.1.2.4.6 Hot spot identification). The efficiency of fragment strategies arises from balancing these considerations of complexity and sensitivity thresholds of bioassays to achieve an improved library hit rate. Indeed, the rise of FBDD has been made possible by advances in biophysical screening methods that have enabled weak binding interactions to be detected with confidence. Furthermore, as the compounds identified are of lower complexity and are more developable, a greater control over physicochemical and pharmacokinetic properties can be obtained during lead optimisation.





**Figure 1.1:** Hann molecular complexity model for protein-ligand complementarity. Complexity is defined as a binding site or ligand features including different pharmacophores, topologies, physicochemical properties and matches refers to the number of potential binding modes. Reprinted (adapted) with permission from Hann MM, Leach AR, Harper G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *Journal of Chemical Information and Computer Sciences*. 2001;41(3):856-64 Copyright 2001 American Chemical Society. <sup>(2)</sup>

Attrition of drugs in clinical trials is often linked to poor absorption, distribution, metabolism, excretion and toxicity (ADMET) profiles, which in turn is principally controlled by their physicochemical properties. Lipophilicity and molecular weight have been widely implicated in detrimental solubility, bioavailability, toxicity and promiscuity of failed drug candidates <sup>(38-44)</sup>. The size and lipophilicity of compounds in many HTS libraries result in their initial hits having less than optimal physicochemical property profiles <sup>(45)</sup>. This is exacerbated by the fact that medicinal chemists often increase ligand affinity through the introduction of hydrophobic groups, which in turn increases both size and lipophilicity <sup>(4)</sup>. Screening smaller, more polar ligands can lead to the identification of more efficient binders, where ideally each heavy atom in the molecule makes a positive contribution to affinity <sup>(46)</sup>. These starting points begin with more desirable physicochemical properties which can then be closely controlled during optimisation, to lead compounds with improved drug-likeness <sup>(3, 4)</sup>. Analysis of marketed drugs and drug candidates from large pharmaceutical companies has recognised that adopting fragment-like strategies could aid in clinical attrition of new drug candidates <sup>(3, 40, 47, 48)</sup>.

While there are many advantages that arise from the screening of small fragments, their size is also the source of some of the limitations of FBDD. The low affinity inherent in the fragment hits requires highly sensitive techniques to detect binding and a substantial medicinal chemistry effort is required to obtain a high affinity lead-like compound. However, there have been great advancements in assay sensitivity, acquisition and automation <sup>(49-52)</sup>, and fragment elaboration is being expedited through the implementation of new and innovative strategies.

### **1.1.2 Strategies of fragment-based drug design**

Successful fragment campaigns are a result of the culmination of intelligent library design, screening the library for hits – which usually involves implementation of biophysical techniques to measure binding, hit evaluation, and fragment optimisation strategies. There is no one set way to approach these components and many unique styles based on expertise and personal preferences can dictate which pathway a campaign takes.

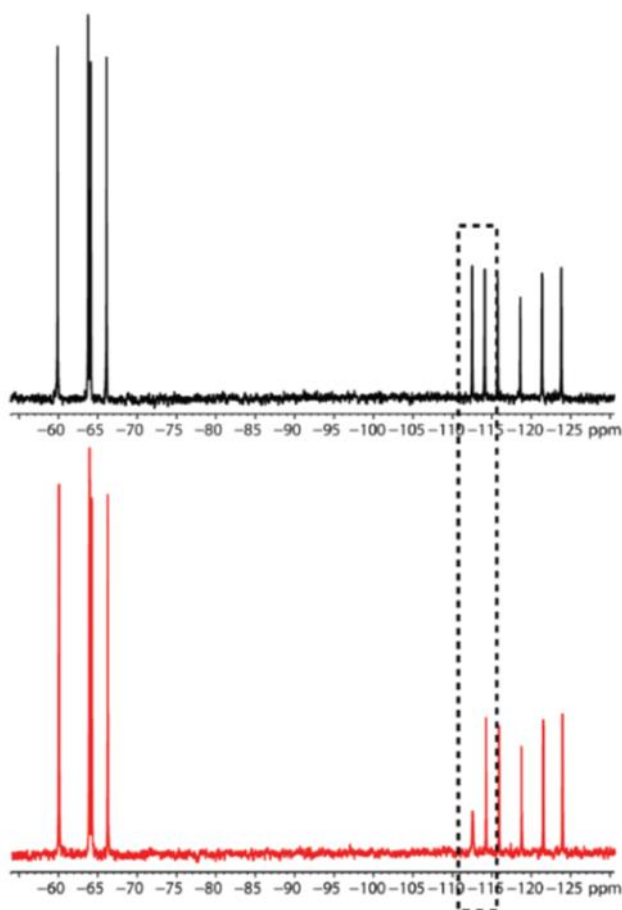
### 1.1.2.1 Library design

Fragment libraries are curated to contain a diverse collection of compounds which follow the general guidelines and principles discussed above. These libraries can be manipulated to enhance certain properties or contain functionalised compounds for specific assays or labelling. Generally, libraries stick fairly close to the Ro3<sup>(5)</sup> when selecting compounds, however they are not viewed as strict rules. Many groups tend to include ~1000 – 2000 compounds<sup>(9)</sup> and are beginning to skew their libraries toward smaller and polar fragments. This is due to historical data showing that these compounds are over-represented amongst the observed hits and they are more likely to have sufficient solubility to be suitable for the screening assays<sup>(6, 8, 9)</sup>. Diversity and developability in a library are key for the efficient detection and evolution of an initial fragment. Broad coverage of the desired physicochemical property space, pharmacophores and chemical fingerprints ensure the highest chance of finding fragments with protein complementarity and ensure the transferability of the ligand set across different classes of targets<sup>(53)</sup>. As a result of the higher hit rates observed in FBDD, it is often the case that the same fragment can be observed binding to multiple targets. Although this promiscuity can be perceived as undesirable, specificity can be introduced during fragment evolution. However, the high hit rates also mean that many fragments are identified in each screen. The inherent ligand diversity in fragment libraries often allows for the identification of new chemical classes binding to the protein target and provides novelty in the development of lead compounds. Libraries are often designed to ensure the commercial availability of analogues of library members to allow for “structure-activity relationship (SAR) by catalogue” as the first stage of hit progression<sup>(9)</sup>. This allows expansion vectors of fragment hits to be assessed through purchased compounds prior to expending substantial chemistry efforts. Furthermore, historical data can be used to identify and remove from undevelopable scaffolds such as 2-aminothiazoles (PrATs) as well as pan-assay interfering compounds (PAINS) from screening libraries. Unwanted properties and functional groups can also be avoided to prevent significant resources being devoted to hits that represent futile starting points<sup>(54-58)</sup>.

Natural product-like fragment libraries have been designed with the aim of enhancing the coverage of 3-dimensional (3D) space<sup>(59-61)</sup>. This is due in part to the observation

that the small and simple molecules that are typically selected in fragment libraries also tend to be relatively flat. It has been suggested that more complex natural product-like compounds can improve solubility and selectivity, and are more useful for challenging targets <sup>(62-64)</sup>, however with the caveat of lower hit rates. These starting points are designed to emulate drugs and drug candidates, which suggest that stereochemistry and aromaticity play a role in clinical success <sup>(65, 66)</sup>. However, the hypothesis that more 3D fragments are better starting points for optimisation has been challenged by observations which suggest that the 3D characteristics of the developed lead compounds are independent from the planarity of the initial hit fragment <sup>(8, 67)</sup>. Nonetheless, these libraries present geometrically different expansion vectors and explore chemical space often missed by *sp*<sup>2</sup> rich fragment libraries.

Libraries which consist of fluorinated fragments have been gaining popularity for their use in <sup>19</sup>F nuclear magnetic resonance (NMR) screening <sup>(68-70)</sup>. Fluorine has many advantages in the field of NMR, namely that it is a 100 % naturally abundant isotope, it is absent from natural biomolecules and buffers that are typically used in screening, it is highly sensitive, has a chemical shift range over hundreds of ppm and narrow linewidth (Figure 1.2) <sup>(68, 69)</sup>. This translates into fast screening, because the sensitivity allows low compound concentrations and due to the wide chemical shift range, <sup>19</sup>F spectra of mixtures can be designed with no overlapping resonances, so they are easily deconvoluted. Together this means that numbers of fluorinated compounds can be combined into larger screening cocktails <sup>(68-70)</sup>. Furthermore, these properties of fluorine atoms can also be used to establish sensitive and robust assays for determination of ligand affinities, saving resources in comparison to protein-detect approaches that require isotope labelling of protein samples <sup>(68, 71)</sup>. While <sup>19</sup>F compounds show great promise, the requirement that the molecules must contain at least one fluorine atom limits the coverage of chemical space <sup>(68, 69)</sup> and care must be taken when designing the library to ensure sufficient diversity within the compound set.



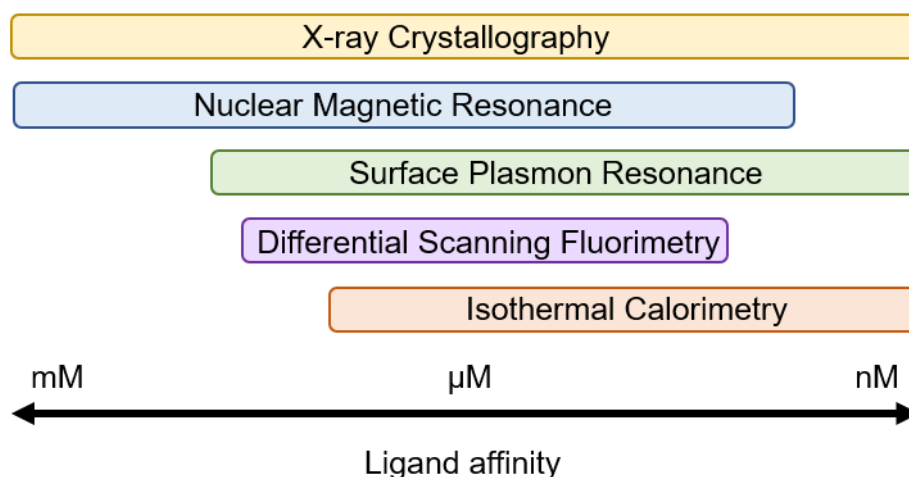
**Figure 1.2:** Typical  $^{19}\text{F}$  NMR fragment screen spectra. Two identical sets of spectra are acquired, one containing 10  $\mu\text{M}$  protein (in red) and the other containing no protein (control plate, in black). Compound hits can be determined by changes in peak intensity or chemical shift in the presence of protein. Reprinted (adapted) with permission from Jordan JB, Poppe L, Xia X, Cheng AC, Sun Y, Michelsen K, et al. Fragment Based Drug Discovery: Practical Implementation Based on  $^{19}\text{F}$  NMR Spectroscopy. *Journal of Medicinal Chemistry*. 2012;55(2):678-87. Copyright 2012 American Chemical Society <sup>(68)</sup>

The resurgence of covalently labelling drugs (discussed in more detail in section 1.2) has also seen the growing use of covalent libraries in fragment-based drug design. They have been employed in many different strategies, including but not limited to; screening of electrophilic ligands <sup>(72, 73)</sup>, tethering <sup>(74)</sup>, photoaffinity labelling <sup>(75)</sup>, whole cell screening <sup>(76-78)</sup> and *in silico* screening <sup>(79)</sup>. The design of these covalent libraries, builds upon the same factors used in the design of traditional FBDD libraries, but must give consideration to the reactivity of any covalent groups that are employed <sup>(80)</sup>. Reactivity of the functional groups must be tempered to limit non-specific labelling, however, factors such as electronic effects of the fragment, and the nucleophilicity of

the target residue require that any library must contain a diverse range of warhead reactivities. Furthermore, an appropriate type of warhead (irreversible, reversible, photoactivated, residue specific) must be selected for the targets that are likely to be screened and the assays that will be employed. It should be noted that reactivity can also be tuned in development by modifying the affinity of any non-covalent interactions and properties of the fragment. This can provide a unique strategy for developing highly efficacious and specific fragment binders.

#### **1.1.2.2 Biophysical techniques**

The use of biophysical techniques in fragment-based drug design has been extensively reviewed and will not be discussed in detail <sup>(22, 28, 81, 82)</sup>. Highly sensitive and robust assays are required for testing due to the low binding affinities of fragments for their protein targets. In many fragment screening campaigns, biochemical assays are complemented by the use of biophysical techniques. The most popular biophysical techniques <sup>(67)</sup> are nuclear magnetic resonance (NMR) <sup>(83, 84)</sup>, surface plasmon resonance (SPR) <sup>(85)</sup> and X-ray crystallography <sup>(86)</sup>. However, many research groups also identify and characterise fragments through isothermal calorimetry (ITC) <sup>(87)</sup>, mass spectrometry (MS) <sup>(88, 89)</sup>, differential scanning fluorometry (DSF or thermal shift) <sup>(90)</sup> and virtual screening <sup>(91)</sup>. As technology advances, the introduction of cryogenic electron microscopy <sup>(92)</sup>, weak-affinity chromatography (WAC) <sup>(93)</sup> and microscale thermophoresis <sup>(94)</sup> into screening cascades has also been reported. Orthogonal biophysical techniques can identify different hits through their ability to detect different ranges of affinity (owing to occupancy and sensitivity, Figure 1.3), tolerance of fragment solubility, and assay artefacts which generate false-positives and/or false-negatives <sup>(6)</sup>. While not all orthogonal techniques are suitable for each target, they are often used in tandem for cross-validation <sup>(95, 96)</sup> and can also provide characterisation of different aspects of binding. Structural information, binding affinity, thermodynamics, kinetics and stoichiometric characterisation can be obtained through the application of complimentary techniques. This characterisation is vital in evaluation and prioritisation of fragment hits and analogues in medicinal chemistry campaigns.



**Figure 1.3:** Typical affinity ranges for biophysical techniques used in fragment-based drug discovery.

### 1.1.2.3 Hit evaluation

Hit compounds, either from initial screens or series optimisation, require extensive evaluation to allow for their prioritisation. Compounds are conventionally ranked using factors such as structure-activity relationships, synthetic viability, selectivity, affinity, and the availability of structural data in complex with their target. However, factors such as physicochemical properties and relative size also need to be considered as advancing compounds based solely on affinity tends to result in the prioritisation of the largest compounds, rather than the most efficient. Ligand efficiency metrics (LEMs) have been embraced in fragment-based drug design in order to monitor and normalise affinity relative to the size and properties of the fragment.

<b>Equation 1.1</b>	$\Delta G = -RT\ln(K_D) \approx -RT\ln(IC_{50})$	
<b>Equation 1.2</b>	$LE = \Delta G/HAC$	(7)
<b>Equation 1.3</b>	$LLE = pK_D \text{ (or } pIC_{50}) - ClogP \text{ (or } logD)$	(40)
<b>Equation 1.4</b>	$LLE_{AT} = 0.11 + 1.34(LLE/HAC)$	(97)
<b>Equation 1.5</b>	$LELP = ClogP/LE$	(98)
<b>Equation 1.6</b>	$GE = \Delta(\Delta G)/\Delta HAC$	(99)
<b>Equation 1.7</b>	$SILE = pK_D/HAC^{0.3}$	(100)
<b>Equation 1.8</b>	$FQ = LE/(0.0715 + (7.5328/(HAC))) + (25.7079/(HAC^2)) - (361.4722/(HAC^3))$	(101)

$\Delta G$  = Gibbs free energy,  $R$  = ideal gas constant ( $1.987 \times 10^{-3}$  kcal  $K^{-1}$  mol $^{-1}$ ),  $T$  = temperature in Kelvin (K),  $K_D$  = dissociation constant,  $IC_{50}$  = half-maximal inhibitory concentration,  $LE$  = ligand efficiency,  $HAC$  = heavy atom count,  $LLE$  = lipophilic ligand efficiency,  $LLE_{AT}$  = lipophilic ligand efficiency adjusted for heavy atom count or Astex  $LLE$ ,  $LELP$  = lipophilicity-corrected ligand efficiency,  $GE$  = group efficiency,  $SILE$  = size-independent ligand efficiency,  $FQ$  = fit quality.

The concept of comparing compounds based on the contribution of each atom or functional group to the free energy of binding was first proposed by Andrews *et al.* <sup>(102)</sup>, and demonstrated by Kuntz *et al.* <sup>(46)</sup>. However, it was ultimately Hopkins *et al.* <sup>(7)</sup> who defined “ligand efficiency” ( $LE$ ) (Equation 1.2) and suggested the use of the simple term in compound development. An estimation using a 10 nM ligand of 38 heavy atoms (~ 500 Da) was used to obtain a target value and suggested that ligands should aim for a  $LE$  of ~ 0.29 kcal mol $^{-1}$  HAC $^{-1}$ . Although 0.3 kcal mol $^{-1}$  HAC $^{-1}$  is often used as a benchmark for success, it should be noted that the optimal  $LE$  value will be target dependent <sup>(103)</sup>. Small molecule inhibitors of protein-protein interactions will often have lower ligand efficiency due to the size of the binding interfaces, which often requires larger, more lipophilic compounds to obtain a reasonable affinity. Conversely, targets such as aminergic G protein-coupled receptors (GPCRs) have well defined binding sites and are able to achieve high affinity with small molecules. GPCR ligands with high  $LE$  values are often achievable as the endogenous ligands are similarly small, tight binders. Furthermore, ligand efficiency has been shown to be size dependent. Average  $LE$  values are systematically higher for compounds with a low heavy atom count ( $HAC$ ) <sup>(101)</sup>, a consideration especially pertinent to FBDD. The metric also does not discriminate between the addition of carbon, heteroatoms or halogens, nor does it

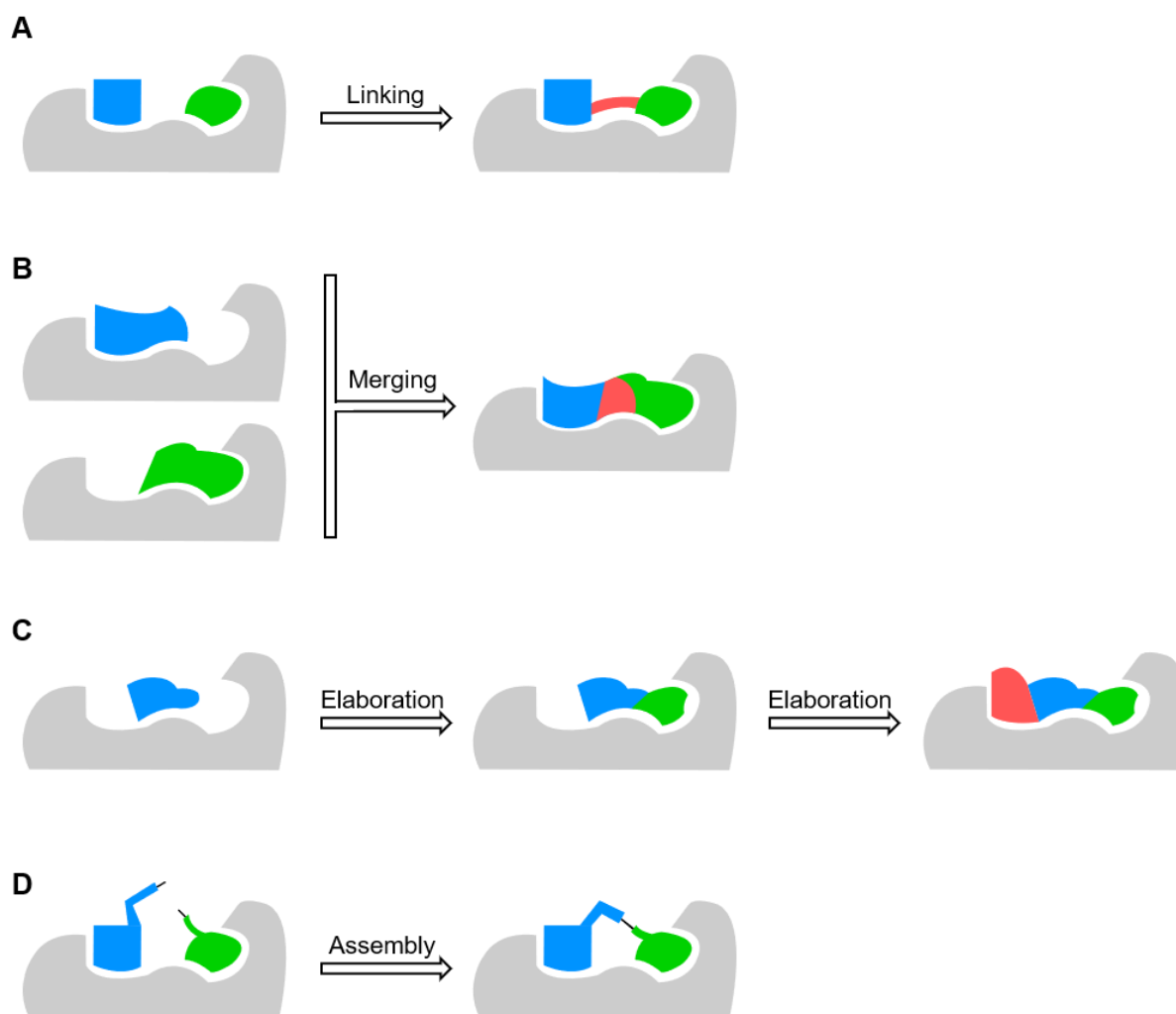


account for the beneficial or detrimental changes in ADMET profiles that this can cause. In response, LEMs have been developed to compensate for lipophilicity (LLE)<sup>(40)</sup>, polar surface area (SEI)<sup>(104)</sup>, molecular size (FQ, SILE)<sup>(100, 101)</sup>, thermodynamics (SIHE)<sup>(105)</sup> and combinations of these properties (LLE<sub>AT</sub>, LELP)<sup>(97, 98)</sup>. Group efficiency<sup>(99)</sup> is a metric very similar to ligand efficiency, however, it is specifically used to compare incremental changes to the ligand and aid in the interpretation of structure-activity relationships.

When selecting which metric to use for compound comparison the chemotype, druggability and binding site of the intended target, affinity range and size of the ligands, and the required ADMET properties should be considered. Striving to maintain constant efficiency during optimisation is a useful tactic and, where possible, the indicators should be used in concert to provide more confidence in the advancement of compounds.

#### **1.1.2.4 Fragment development and optimisation**

Beginning from fragment hits can require many iterative rounds of medicinal chemistry before obtaining a high affinity lead-like compound. There is no right or wrong way to arrive at this point, however, there are established approaches that are often used for fragment development. The general approaches to optimisation are often described as fragment linking, merging, elaboration and self-assembly. However, more than one strategy may be available for a given target and in practice they often overlap during campaigns (Figure 1.4)<sup>(106, 107)</sup>. Accompanying strategies such as hot spot identification, screening of crude reaction mixtures and covalently functionalising compounds can be employed in parallel to these approaches and aid in the efficient advancement of fragments<sup>(107, 108)</sup>.



**Figure 1.4:** General fragment-based drug design strategies. A) Fragment linking B) Fragment merging C) Fragment elaboration D) Fragment self-assembly

#### 1.1.2.4.1 Fragment Linking

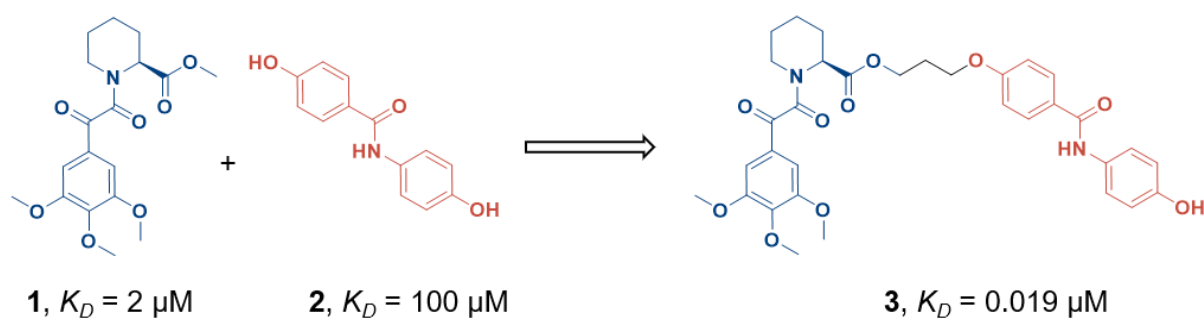
Fragment linking was the strategy employed in the first reported fragment-based ligand optimisation <sup>(1)</sup>, and involves the identification and joining of two fragments binding to the target in close proximity to each other. While linking can be very efficient in boosting affinity, there are challenges inherent with this approach. Identifying two adjacent fragments is not always possible, is reliant on structural information and can require multiple screening campaigns. Furthermore, once discovered, these fragments may not contain expansion vectors that allow the two fragments to be linked <sup>(109)</sup>.

The binding energetics of such linked molecules was first described by Jencks. In this analysis, the binding free energy of a compound is a summation of the free energies of its fragments or functional groups <sup>(102, 110)</sup>. The summation of these intrinsic free energies of binding must compensate for the benefits or costs incurred by linking the individual compound components. Thus, Jencks <sup>(110)</sup> describes the energetic contributions of a two component molecule as:

$$\text{Equation 1.9} \quad \Delta G_{AB}^0 = \Delta G_A^i + \Delta G_B^i + \Delta G^s \quad (110)$$

Where  $\Delta G_{AB}^0$  is the overall binding energy,  $\Delta G_A^i$  and  $\Delta G_B^i$  are the binding energy of the individual components and  $\Delta G^s$  is the connection energy. This energy of connection can be detrimental due to the loss of translational or rotational entropy, or geometric and conformational strain <sup>(102, 110, 111)</sup>. Conversely, if optimised, the linker attaching the fragments can increase the affinity of the ligand. Although rare, compound linkers can obtain “superadditivity” (also referred to as positive cooperation) and further increase affinity through the formation of their own protein interactions <sup>(112, 113)</sup>.

In the seminal ‘SAR by NMR’ FBDD paper, Fesik and colleagues <sup>(1)</sup> describe the use of  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single-quantum correlation (HSQC) NMR to identify two low affinity fragments binding in adjacent sites of FK506 binding protein. The initial fragment was a pipercolinic acid derivative (**1**,  $K_D = 2 \mu\text{M}$ ) and subsequent screening in the presence of saturating concentrations of **1** identified a benzanilide derivative binding in a neighbouring site. The binding orientations of these compounds were modelled using intermolecular  $^{15}\text{N}$ - $^{13}\text{C}$  filtered nuclear Overhauser effects (NOE) data. These models were then used to identify appropriate positions on the fragments for linking. Using simple alkyl chain linkers, fragment **1** and **2** (an analogue of the original benzanilide fragment, which bound with  $K_D = 100 \mu\text{M}$ ) were coupled. This resulted in a high affinity inhibitor ( $K_D = 19 \text{ nM}$ ) which was developed with minimal rounds of medicinal chemistry (Figure 1.5).



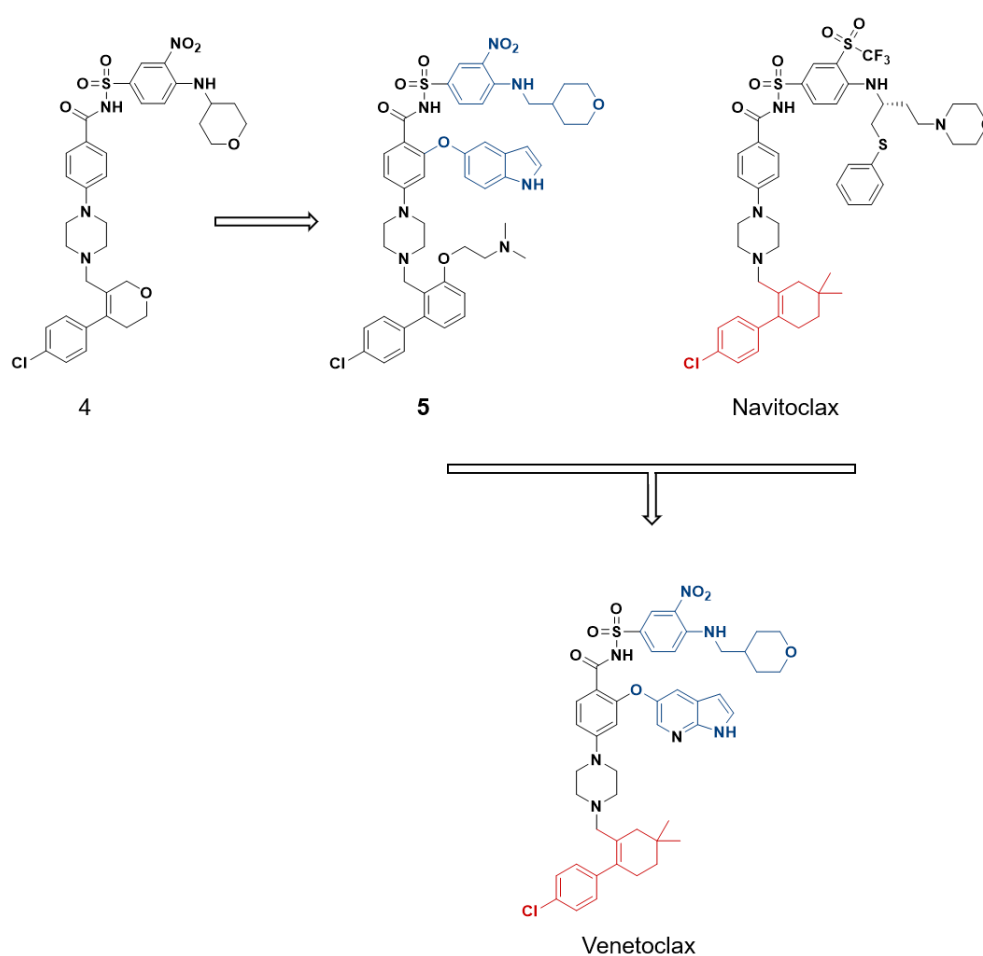
**Figure 1.5:** Fragment linking strategy against FK506 binding protein implemented by NMR in the seminal “SAR by NMR” report. <sup>(1)</sup>

#### 1.1.2.4.2 Fragment merging

Fragment merging combines elements of one or more fragments which bind to the target in overlapping poses. The challenges faced in merging compounds are very similar to fragment linking. This largely revolves around the merging process introducing geometric strain into the merged compound, which compromises the initial binding interactions of the individual components <sup>(102, 110, 111)</sup>. As a result, the expected increases in potency of the merged compound may not be achieved. Merging also heavily relies on the use of structural data to identify the binding orientation of appropriate fragments to guide their subsequent fusion. In the absence of structural characterisation, the SAR from separate or literature compounds can be combined to gain insight into potential merging opportunities <sup>(114)</sup>.

Venetoclax, one of the four fragment-derived United States Food and Drug Administration (FDA) approved therapies, was derived from a fragment merging campaign (Figure 1.6) <sup>(12)</sup>. A preclinical anticancer agent navitoclax has a high affinity for both BCL-2 and BCL-X<sub>L</sub> <sup>(115, 116)</sup>. A side effect of navitoclax is the decrease in the number of circulating platelets within patients’ blood caused by the inhibition of BCL-X<sub>L</sub> <sup>(117, 118)</sup>. To reduce this effect and maintain its efficacy, this drug was optimised to be selective to BCL-2 <sup>(12)</sup>. A crystal structure of BCL-2 complexed with an acylsulfonamide analogue (**4**) resulted in a similar binding pose as adopted by navitoclax. However, the P4 region which is occupied by the thiophenyl group of

navitoclax was now occupied by a tryptophan sidechain from a second protein molecule. This amino acid sidechain was incorporated into the design of a series of compounds and produced an indole containing derivative (**5**). Compound **5** gained an electrostatic interaction through this appendage which gave the desired selectivity for BCL-2 over BCL-XL. Optimisation of the indole resulted in an azaindole derivative which was then merged with navitoclax to become the BCL-2 selective anticancer therapy venetoclax.

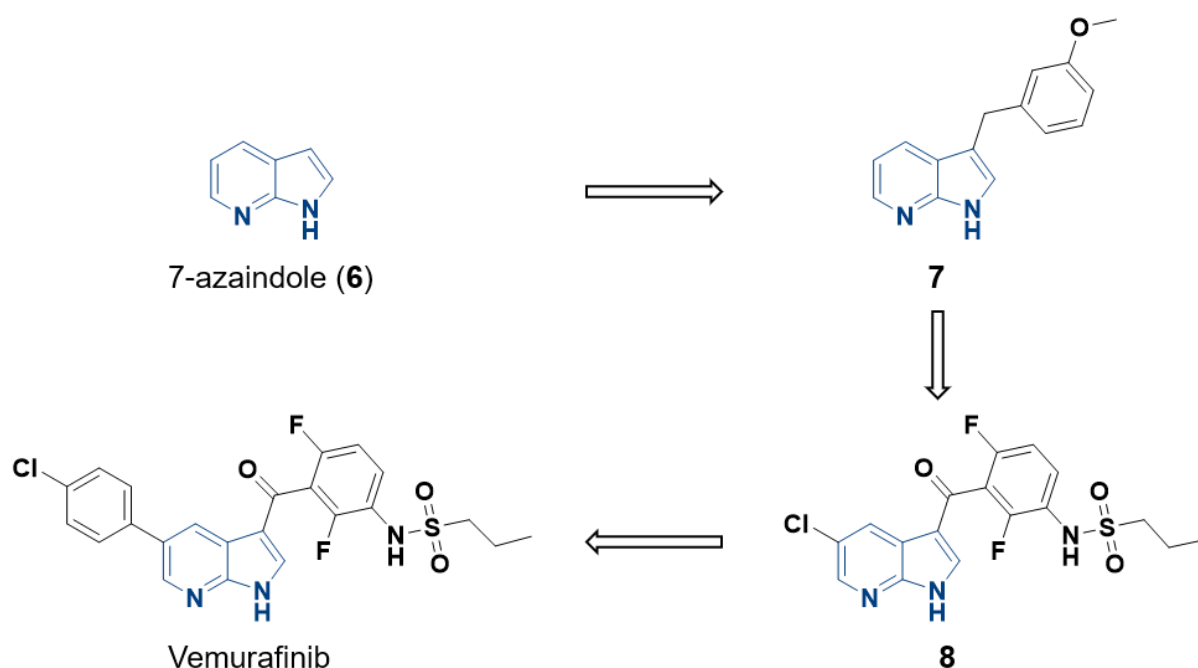


**Figure 1.6:** Fragment merging strategy which lead to the FDA approved BCL-2 inhibitor Venetoclax.

#### 1.1.2.4.3 Fragment elaboration

Fragment elaboration is by far the most common fragment optimisation strategy <sup>(109)</sup>. Evolution of the compound through this method requires the sequential addition of binding motifs and functional groups. The general workflow of this approach involves iterative rounds of SAR (by biophysical or biochemical assays), structural characterisation and analogue synthesis.

Vemurafinib, the first fragment derived drug on the market, was developed by the growth of a 7-azaindole fragment into a highly potent and selective B-Raf<sup>V600E</sup> inhibitor (Figure 1.7) <sup>(119)</sup>. B-Raf<sup>V600E</sup> is one of the most commonly oncogenic mutations observed for the B-Raf protein kinase and is a key driver of melanoma. Selectivity to this mutation over the wild type kinase is key for the drugs toxicity profile. 20,000 fragments were screened against a panel of kinases and hits from the initial biochemical screen were validated and characterised by co-crystallisation. The 7-azaindole scaffold (**6**, IC<sub>50</sub> > 200 µM) was selected for optimisation in spite of the observation of multiple binding modes. Monosubstituted compounds resolved this issue and through crystal structures an important hydrophobic Raf-specific pocket was identified as a potential source of selectivity. Extension into this pocket was achieved through the formation of hydrogen bonds which anchored the 7-azaindole into the hinge region and a sulfonamide linker which directed an alkyl chain into the Raf-specific pocket. These analogues allowed the development of a compound with 10-fold selectivity for B-raf<sup>V600E</sup> (**8**). Optimisation of the ADMET profile of this compound, resulted in the development and approval of vemurafenib as a treatment of therapy resistant melanomas <sup>(27)</sup>.

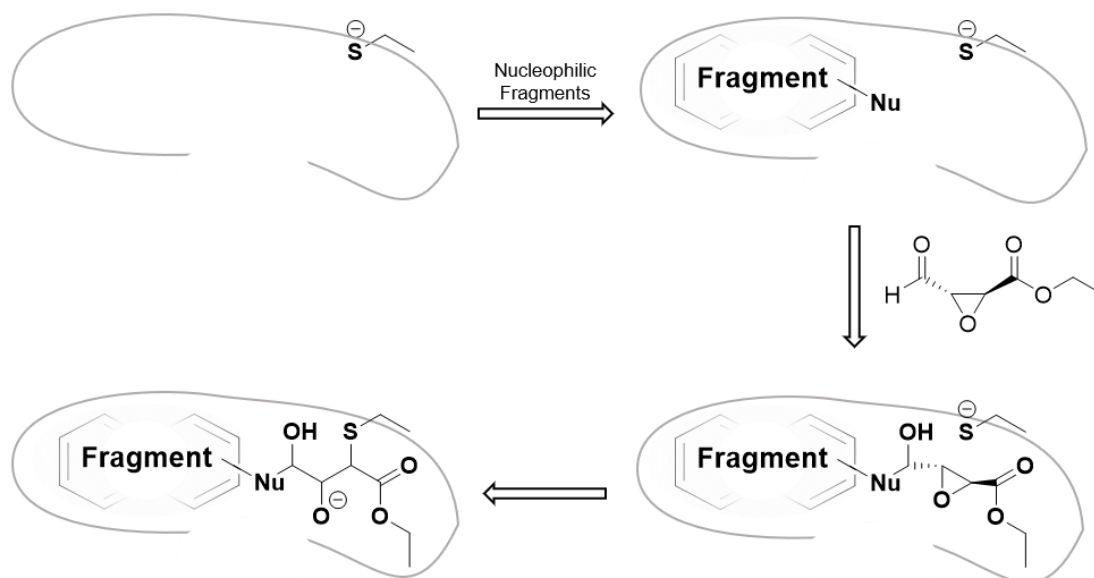


**Figure 1.7:** Fragment elaboration campaign against B-Raf<sup>V600E</sup> which led to the first fragment-derived therapeutic Vemurafinib.

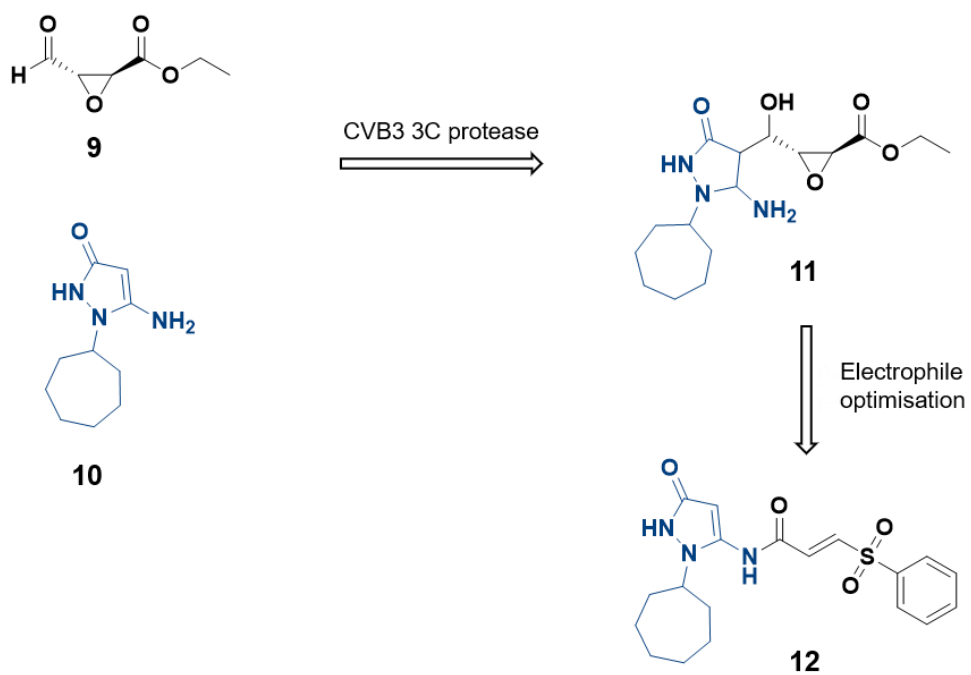
#### 1.1.2.4.4 Fragment self-assembly

Similar to fragment linking, fragment self-assembly requires two reactive fragments to bind in adjacent protein subpockets. Fragments bind to a target simultaneously, positioning their reactive handles in close proximity and utilise the proteins surface to catalyse their ligation. This can reduce the effort required to identify and optimise linkers that maintain the initial fragment binding poses. Furthermore, using the protein surface to template the reaction allows synthesis and evaluation of activity to be done in one step, resulting in an efficient optimisation of fragments. Target guided synthesis can employ either reversible or irreversible linking between both the individual fragment components or between the fragment and protein <sup>(120, 121)</sup>. Unfortunately, the range of reactions suitable for this approach is limited by the required assay conditions, which must maintain the protein in a folded state, and this is problematic especially with respect to C-C bond formation and cross coupling reactions <sup>(120, 121)</sup>.

A



B



**Figure 1.8:** Identification of inhibitors by fragment self-assembly. A) General concept of ligation assay  
B) Fragment progression against Coxsackie virus B3 3C protease.



Rademann and colleagues used the Coxsackie virus B3 3C protease to catalyse the assembly of broad spectrum enteroviral protease inhibitors (Figure 1.8) <sup>(122)</sup>. Initially a bis-electrophile fragment (**9**) was designed through the combination of a cysteine labelling warhead (epoxide) and a reactive handle for ligation with nucleophilic amines (aldehyde) and was validated using a fluorescence-resonance energy transfer-based assay. This assay was then used to identify a nucleophilic amine which enhanced the reduction of enzymatic activity when ligated to electrophile **9**. Cycloheptyl amine **10** when screened alone reduced the activity of the protease < 10 %, however the epoxide coupled product **11** resulted in complete inhibition. The reaction was quantified by high performance liquid chromatography – mass spectrometry (HPLC/MS) with and without the presence of the protein which confirmed the templating effect of the enzyme. Once amine **10** was validated, a series of electrophilic warheads were explored, resulting in a vinyl sulfone inhibitor (**12**). The vinyl sulfone analogue showed submicromolar IC<sub>50</sub> values against multiple genetically related viral proteins and was inactive when tested against human proteases.

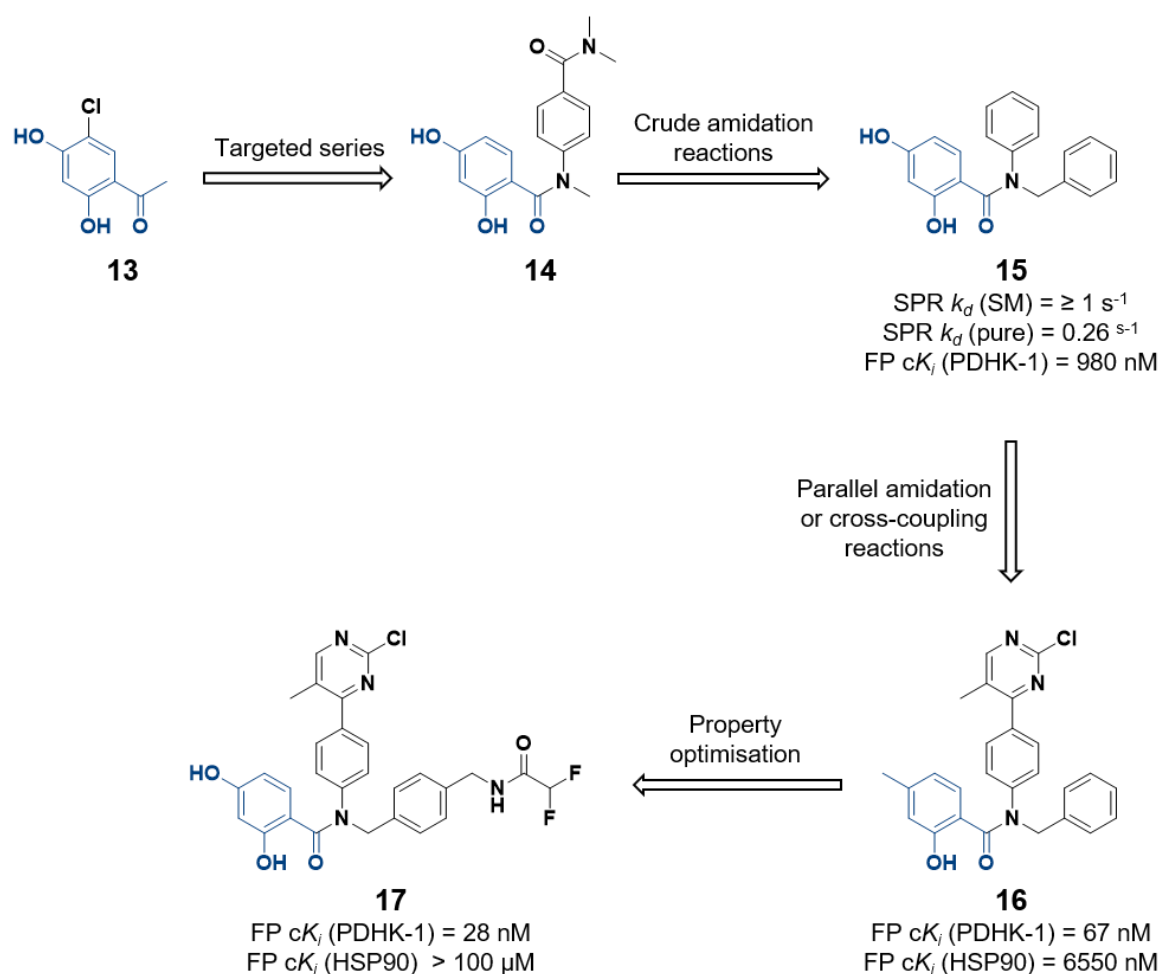
#### **1.1.2.4.5 Crude reaction mixture screening**

A major bottleneck in the development of fragments is the synthesis and purification of compounds for testing. Techniques which employ microscale parallel synthesis speed up the process of compound synthesis and a number of approaches have been reported that enable subsequent screening to be performed with minimal purification, thereby removing the need for compound purification and characterisation prior to testing. Together, these techniques have begun to navigate these bottlenecks. In these assays the target protein is used to resolve the component in a crude reaction mixture which binds with the highest affinity. This is achieved as the highest affinity compounds tend to have slower dissociation rates and longer residence times at the binding site. Therefore, this approach rapidly identifies compounds of interest without the need for purification. Currently there have been reports of screening crude reaction mixtures by SPR <sup>(123-125)</sup>, affinity-selection mass spectrometry <sup>(126, 127)</sup> and X-ray crystallography <sup>(128, 129)</sup>. These assays are limited to high-yielding chemistries and robust protein systems to ensure sufficient reaction conversion to achieve observable ligand occupancies and protein target viability. With the improvement in the sensitivity

of many biophysical assays this approach will likely be extended to other techniques and more widely implemented in medicinal chemistry campaigns.

A library of amide resorcinol derivatives was developed into inhibitors of pyruvate dehydrogenase kinase (PDHK) by using an SPR technique termed “off-rate screening” (Figure 1.9) <sup>(124)</sup>. PDHK has four mammalian isoforms (PDHK-1, -2, -3 and -4), which have been implicated in glucose regulation. The role of PDHK isoforms have been studied in multiple disease states including diabetes and cancer, and has sparked interest in the development of pan-PDHK inhibitors as potential therapeutics. Heat shock protein 90 (HSP90) is a member of the same kinase super family as PDHK and these enzymes show very similar ATP binding sites. Immense care was taken to ensure that the structure-based designs exploited the subtle differences in binding sites to optimise fragment selectivity for PDHK over HSP90.

An NMR fragment screen was conducted against PDHK-2 which identified 78 initial hits which produced 43 compounds structures bound to PDHK-3. A resorcinol derivative **13** was identified in these screens, however, the binding pose for the fragment was virtually identical for PDHK-3 and HSP90. Further docking studies using these structures suggested that replacing the methylketone with an amide provided expansion vectors which could be used to design in selectivity. A small series of *N*-substituted compounds, such as disubstituted amide **14**, showed activity against all four isoforms with some selectivity over HSP90. These results suggested that an off-rate screen based on a series of amidation reactions could be used to advance this fragment series. Acyl chlorides were reacted with a library of secondary amine reagents to explore direct attachment to the amide or expansion through an *N*-benzamide vector. This off-rate screen identified the directly attached *N*-benzyl compound **15** which gave a slow off-rate and high ligand efficiency. This compound was therefore selected for further parallel chemistry.



**Figure 1.9:** Application of off-rate screening against pyruvate dehydrogenase kinase (PDHK).

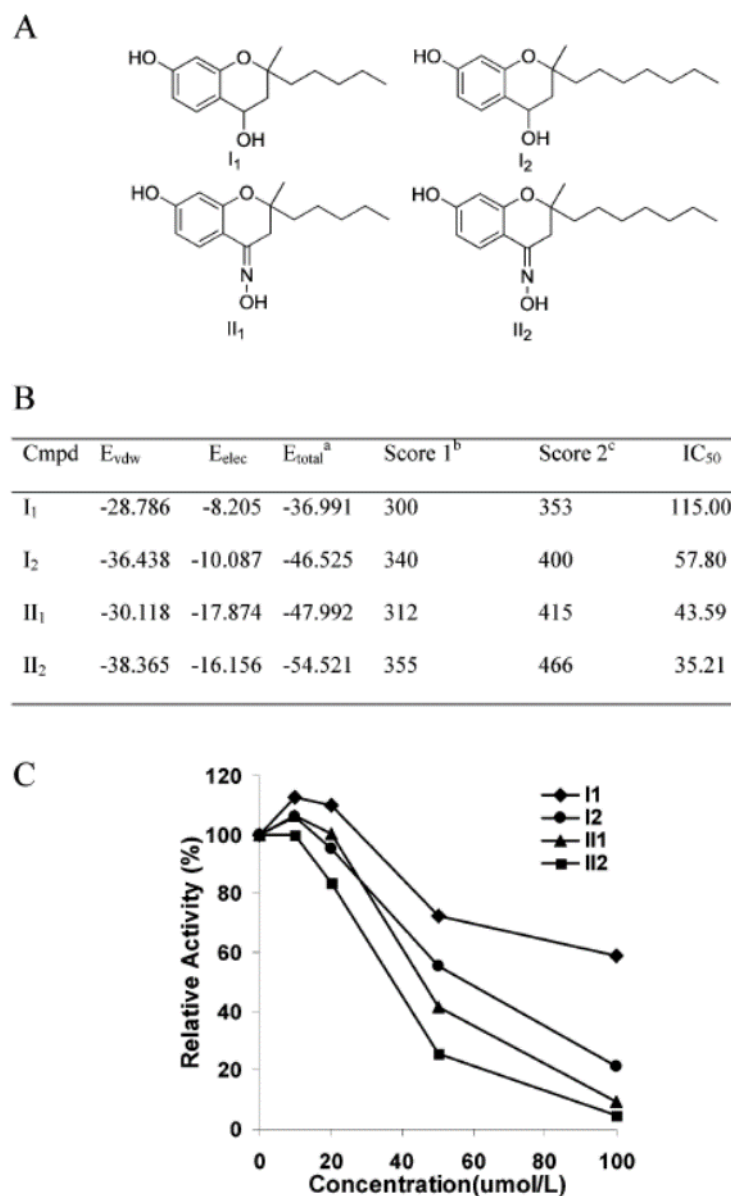
Carboxylic acid and boronate ester analogues of **15** were synthesised and used as the starting materials for a series of amidation and cross coupling reactions respectively. One cross coupling product, 2-chloro,5-methylpyrimidine analogue **16**, resulted in 100-fold selectivity for PDHK-1 over HSP90 by fluorescence polarisation (FP). Pyrimidine **16** was optimised for potency and physicochemical properties using traditional medicinal chemistry strategies and resulted in the identification of compound **17**. This final difluoromethyl compound gave an FP  $K_i$  value of 28 nM against PDHK-1 and > 100  $\mu\text{M}$  against HSP90. While *in vivo* studies ultimately attributed the antiproliferative activity to off-target effects, compound **17** was optimised

through off-rate screening of crude mixtures and was able to obtain high potency against all PDHK isoforms and retained high selectivity over HSP90.

#### **1.1.2.4.6 Hot spot identification**

FBDD strategies rely on the efficient protein interactions made by fragments. Binding at protein “hot spots” allows these fragments to make the most energetically favourable interactions. Initial fragment screens uncover the main ligand binding spot, however, in order to improve affinity, it is necessary to identify favourable interactions at neighbouring sites. Hot spot identification extends the core idea of fragment-based discovery which uses small compounds to efficiently cover chemical space and increase the possibility of target complementarity. The Hann complexity model limits the size of fragments based on the definition of useful binding as a single site event <sup>(2)</sup>. In hot spot identification this limit no longer applies. As the aim is to discover multiple interaction sites for ligand expansion, it is favourable to detect more than one binding event and therefore smaller compounds can be used.

Organic solvents (such as acetonitrile, methanol, etc) are considered minimal pharmacophores as they are often a single functional group and can represent a single pharmacophore <sup>(102)</sup>. The implementation of solvents as binding site probes originated as multiple copy simultaneous search (MCSS) computational modelling <sup>(130)</sup>. This approach identified the lowest energy poses of multiple solvents and used clusters of solvent poses to prioritise protein hot spots and interactions. Shortly after, experimental methods emerged for screening of solvents using X-ray crystallography <sup>(131)</sup> and NMR <sup>(132)</sup>. Low molecular weight probes bind in affinity ranges even lower than fragments and require high sensitivity in the biophysical technique being used. To compensate for the low affinity and protein occupancy very high concentrations of solvent are required (up to 100 % solvent in crystallography <sup>(133)</sup>).



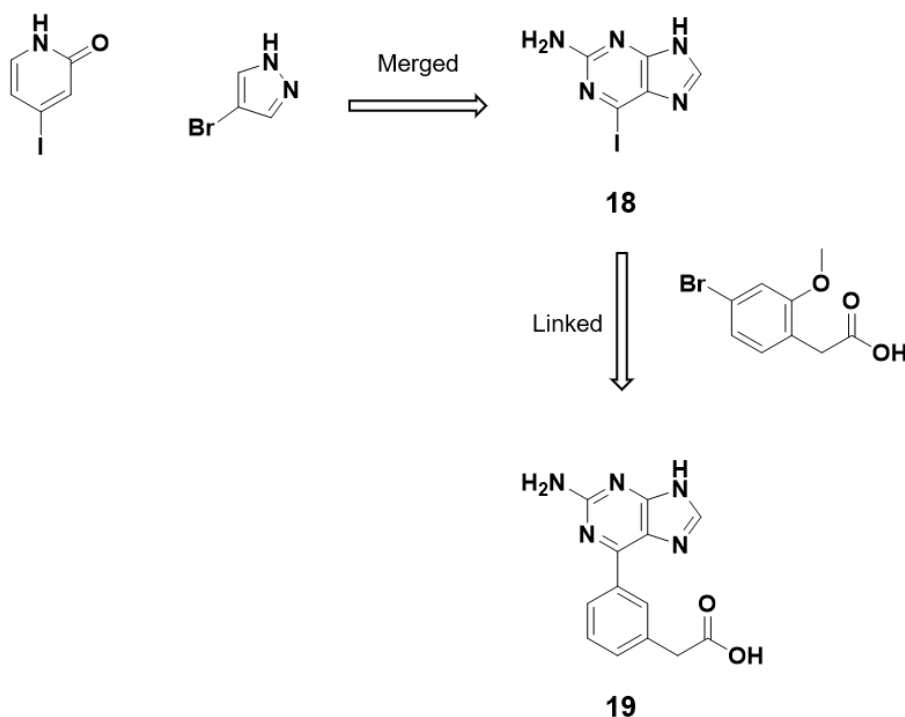
**Figure 1.10:** Development of lanosterol 14 $\alpha$ -demethylase (CYP51) antifungal agents using multiple copy simultaneous searching (MCSS). A) Chemical structure of the designed lead compounds. B) Calculated interaction energies (kcal/mol) for the complexes of the lead compounds with the active site of CYP51 of *Candida albicans*, Ludi scores, and experimentally determined binding affinity (IC<sub>50</sub>,  $\mu$ mol/L) of the lead compound for the reconstituted CYP51 from *Candida albicans*. *a*:  $E_{total} = E_{vdw} + E_{elec}$ . C) Inhibition of the activity of CYP51 from *Candida albicans* by lead compounds. Reprinted (adapted) with permission from Ji H, Zhang W, Zhang M, Kudo M, Aoyama Y, Yoshida Y, et al. Structure-Based de Novo Design, Synthesis, and Biological Evaluation of Non-Azole Inhibitors Specific for Lanosterol 14 $\alpha$ -Demethylase of Fungi. *Journal of Medicinal Chemistry*. 2003;46(4):474-85. Copyright 2003 American Chemical Society <sup>(134)</sup>

A MCSS solvent model was used in the *de novo* design of specific lanosterol 14 $\alpha$ -demethylase (CYP51) antifungal agents <sup>(134)</sup>. Azole containing CYP51 antifungal treatments show toxicity due to their ability to coordinate the heme group of the CYP51 mammalian counterpart. Therefore, it is desirable to identify compounds which contain a novel scaffold and to avoid this interaction. An MCSS using benzene, propane, cyclohexane, phenol, methanol, ether and water identified 4 binding subsites outside of the heme adjacent site. Fragments were designed by connecting the most favourable functional groups from each subsite and subsequent testing of the derivatives showed micromolar activity against CYP51.

More recently there has been a move toward identifying hot spots with compounds which are between solvents and fragments in size and complexity <sup>(135-138)</sup>. Currently, these techniques focus on using X-ray crystallography to identify the ligand binding motifs. Concentrations and conditions do not need to be as extreme as those used in organic solvent screening due to the increased affinities of these compounds. Nonetheless, libraries often contain polar compounds to aid in solubility and often incorporate halogen atoms to exploit anomalous scattering for unambiguous binding pose identification.

A set of "FragLites" were screened by X-ray crystallography to map the protein hot spots of cyclin-dependent kinase 2 (CDK2) <sup>(138)</sup>. The FragLite library consists of 31 halogenated fragments with either a hydrogen bond donor and acceptor or two hydrogen bond acceptor motifs with pharmacophore connectivities between one and five bonds. They were screened at a concentration of 50 mM by soaking each compound individually into crystals of CDK2. Electron density was observed for nine of these 31 compounds. The two known binding sites of CDK2 (the orthosteric and allosteric sites) as well as an additional four previously unidentified sites were found to be occupied by FragLites. The proximity of some of these unique binding sites and interaction motifs led to the design of compounds which merged aspects of the hits (Figure 1.11). The interactions made by the original FragLite hits were conserved by the new compound **18** and afforded a  $K_D$  of 350  $\mu$ M by ITC. Linking this fragment to another of the original crystallographic hit which displayed an extra interaction in the

proteins hinge region led to the FragLite derived fragment **19** with a  $K_D$  of 50  $\mu\text{M}$  for CDK2.



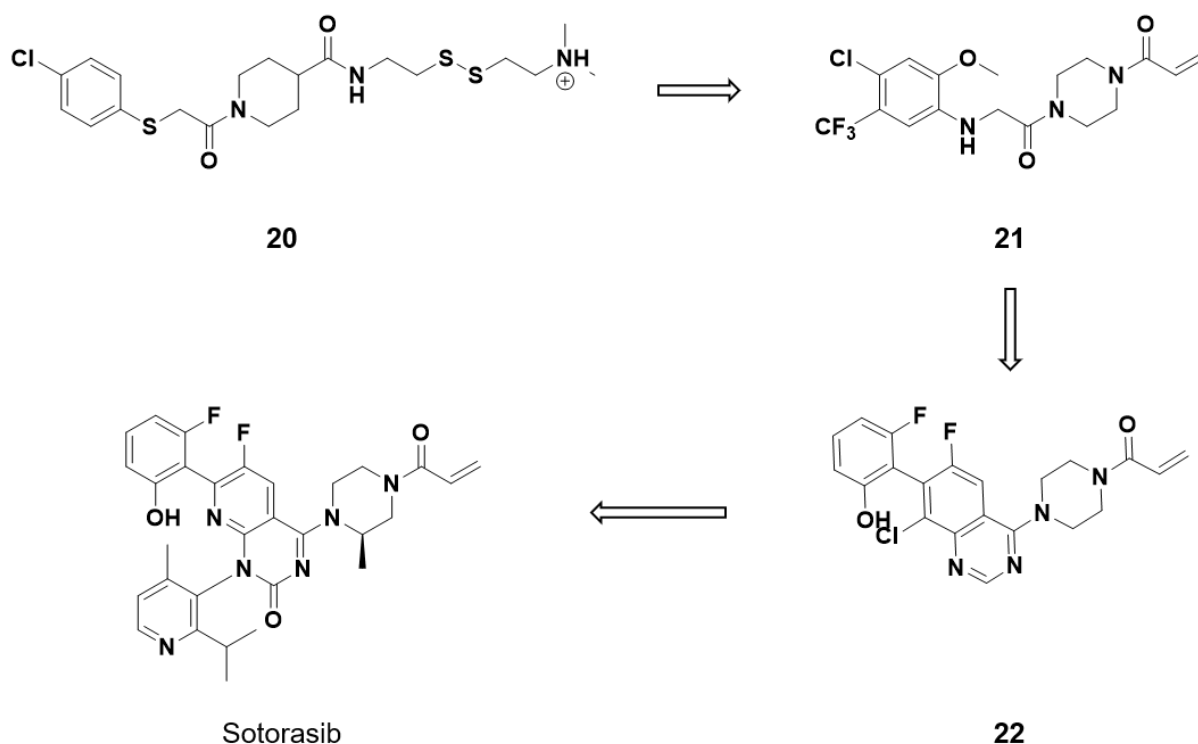
**Figure 1.11:** Optimisation of cyclin-dependent kinase 2 (CDK2) inhibitors by merging high concentration X-ray crystallography hits (FragLites).

#### 1.1.2.4.7 Covalent fragments

Covalent adaptation of fragments has been gaining great momentum in recent years with a large number of publications exemplifying this developmental strategy<sup>(67)</sup>. The development of covalent fragments can be approached from either optimising the affinity of a covalent fragment hit or attaching a covalent warhead onto an advanced non-covalent fragment. These compounds contain electrophilic warheads with the ability to either reversibly or irreversibly label nucleophilic residues. The low affinity of fragments can be supplemented by the enthalpic gain of covalently labelling the protein. It is often applied to inhibitors of difficult or “undruggable” targets to address

flat, featureless protein-protein binding interfaces <sup>(139, 140)</sup>. Furthermore, covalent anchoring of compounds can assist with multiple or unreliable binding modes during fragment optimisation <sup>(141)</sup>. Toxicity caused by non-specific binding is one of the main concerns in the field of covalent inhibition <sup>(142-144)</sup>. In order to reduce ligand promiscuity frequent fragment hitters and warheads with an intrinsically high reactivity must be carefully monitored or avoided if possible <sup>(141)</sup>.

GTPase K-Ras is susceptible to oncogenic mutations where a G12C mutation is highly common and usually results in poor response to oncotherapies <sup>(145)</sup>. Replacement of glycine with a nucleophilic cysteine provides an interesting opportunity to use covalent inhibition to infer compound selectivity. Furthermore, Cys12 of K-Ras is located in close proximity to regions of the protein involved in effector interactions <sup>(146)</sup>. Disulfide tethering <sup>(74)</sup> was used to identify an initial fragment hit (**20**) which bound specifically to K-Ras<sup>G12C</sup> in an allosteric ligand-induced cryptic pocket <sup>(146)</sup>.



**Figure 1.12:** Development of Sotorasib a covalent K-Ras<sup>G12C</sup> inhibitor derived from fragments identified by tethering.



Structural guided development into this pocket and replacement of the disulfide with an electrophilic warhead yielded acrylamide **21**. The compound was modified to improve its affinity and ADMET profile by scaffold hopping, expanding into a hydrophobic pocket and removal of the amide bond linkage to become analogue **22** which achieved potency akin to a drug candidate <sup>(147, 148)</sup>. Acrylamide **22** was further optimised for potency and became Sotorasib, a first-in-class K-Ras<sup>G12C</sup> inhibitor, which is currently in phase 2 clinical trials <sup>(149)</sup>.

## 1.2 Covalent inhibitors

Covalent enzyme inhibition has been shown as a successful therapeutic strategy and is utilised by many commonly used drugs. Highly popular drugs such as aspirin and penicillin have covalent modes of action, however, this mechanism was not intended and only discovered many years later <sup>(150, 151)</sup>. Nonetheless, there has historically been much skepticism to this type of inhibition due to concerns over potential off-target effects and promiscuity <sup>(143, 144)</sup>. These concerns are being addressed through rational drug design, with an emphasis on taming and tuning ligand reactivity and thorough establishment of mechanisms of action. While some remain sceptical, advancements in the field have seen covalent inhibitors undergo a great resurgence with both publications and clinical candidates booming <sup>(67, 143, 144, 152)</sup>. Although already used extensively in activity based protein profiling <sup>(153)</sup>, covalent functionalisation is readily being implemented into fragment-based design <sup>(141)</sup>, DNA-encoded libraries <sup>(154)</sup>, *in silico* screening <sup>(155)</sup> and other drug design strategies.

Covalent inhibition by compounds can arise from chemoreactive and photoactivatable functional groups or as prodrugs that are activated by enzymatic modification for covalent binding. The reactions may be either reversible or irreversible and generally target nucleophilic amino acids. While cysteine and serine residues make up over half of the targeted residues in FDA approved drugs, there has been a growing interest in labelling threonine and lysine residues <sup>(152)</sup>. It is possible to label other amino acids and protein cofactors, however, they are rarely the focus of drug development <sup>(152, 156)</sup>.

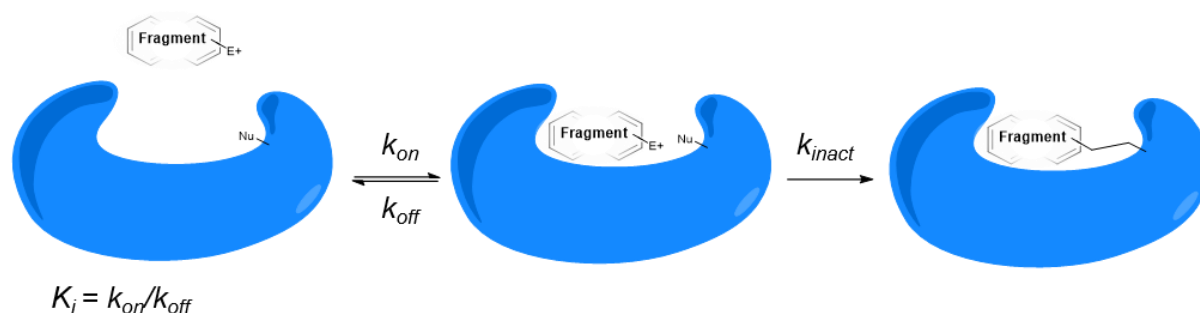
Reversibility of Inhibition	Class of Inhibitor	Scheme	Major Selectivity Determinant
Reversible	Covalent Reversible	$E + I \rightleftharpoons E \cdot I \rightleftharpoons E-I$	Thermodynamic equilibrium
	Slow Substrate	$E + I \rightleftharpoons E \cdot I \rightarrow E-I \rightarrow E + P$	Enzymatic catalysis
Irreversible	Residue-Specific Reagent	$E + I \rightarrow E-I$	Chemical reactivity
	Affinity Label: Classical	$E + I \rightleftharpoons E \cdot I \rightarrow E-I$	Effective molarity & chemical reactivity
	Affinity label: Quiescent	$E + I \rightleftharpoons E \cdot I \xrightarrow[\text{"off pathway"}]{\text{catalyzed}} E-I$	Effective molarity & "off pathway" cat.
	Affinity label: Photoaffinity	$E + I \rightleftharpoons E \cdot I \xrightarrow{h\nu} E-I$	Effective molarity & light activation
	Mechanism-Based	$E + I \rightleftharpoons E \cdot I \xrightarrow[\text{"on pathway"}]{\text{catalyzed}} E \cdot I^* \rightarrow E-I$ $\downarrow$ $E + P$	Enzymatic catalysis

**Figure 1.13:** General reactions of covalent inhibitors (I) with their target proteins (E). Reprinted (adapted) with permission from Tuley A, Fast W. The Taxonomy of Covalent Inhibitors. *Biochemistry*. 2018;57(24):3326-37. Copyright 2018 American Chemical Society. <sup>(157)</sup>

### 1.2.1 Advantages of covalent inhibition

The renewed interest in covalently modifying compounds has arisen due to their potential to address drug resistance and to address some of the common factors that lead to compound failure in clinical trials (such as lack of potency and poor ADMET profiles). Covalent inhibition of a target protein can lead to the separation of pharmacokinetic and pharmacodynamic profiles. When a drug encounters its target protein, an equilibrium is established between the bound and unbound states, which is described by an equilibrium constant ( $K_D$  or  $K_i$ ). This equilibrium is affected by the affinity of the drug to the target, the concentration of any endogenous ligands that compete with the drug for binding and the remaining concentrations of the drug at the target site. Covalent drugs undergo a subsequent reaction with the enzyme where they transition to form an essentially irreversible covalent complex. The formation of this covalent complex is described by the rate constant  $k_{inact}$  (Figure 1.14). Due to this final step, covalent drugs operate under nonequilibrium kinetics. Furthermore, this step increases the residence time of the compounds and duration of inhibition is no longer determined by compound clearance, but by target resynthesis. This often translates to higher potency and less systemic drug exposure <sup>(158, 159)</sup>. In turn, this can result in fewer side-effects, less frequent dosing, and ultimately higher patient

compliance <sup>(159, 160)</sup>. Compounds with less desirable physicochemical properties may be less problematic for clinical development due to the diminished role that pharmacokinetics play in covalent inhibition <sup>(161)</sup>.



**Figure 1.14:** Binding kinetics of irreversible covalent inhibition.

In addition, covalent drugs have the ability to address problematic targets, namely, those susceptible to resistance by mutation and ‘undruggable’ proteins with shallow binding sites that do not easily accommodate conventional small molecule drugs. Many enzymes require a nucleophilic residue in their catalytic motif, therefore, this residue is essential and cannot be mutated without loss of function, which eliminates one potential route to acquired resistance. Furthermore, resistance to therapies has also been linked to the short residence times that are associated with high  $k_{off}$  values that are sometimes observed for non-covalent inhibitors <sup>(144, 162)</sup>. In contrast, covalent inhibition results in long residence times. Together, this suggests that covalent mechanisms of action are a valid strategy to combat therapeutic resistance. Like fragments, covalent compounds have found applications in targeting difficult binding sites. Forming an irreversible bond with the protein has been identified as a way to overcome the low affinity interactions inherent with flat and solvent exposed binding sites that are typically found at protein-protein interaction sites <sup>(163, 164)</sup>. Furthermore, as described in the case study against K-Ras<sup>G12C</sup>, exploitation of unique residues (mutations or distinctions between homologues) and their intrinsic reactivity can provide an orthogonal way to obtain target selectivity <sup>(149)</sup>.

### 1.2.2 Challenges of covalent inhibition

Some of the advantages of covalent inhibition can also present challenges for the implementation of covalent inhibition. While prolonged residence time has benefits with respect to ADMET profiles, it requires great understanding of the implications of continuous suppression of the target. If the ramifications of this are poorly understood or some function is still required, total blockage may result in unacceptable toxicity profiles <sup>(161, 165, 166)</sup>. For example, glutamate – related neuronal cell injuries have been implicated in many neurodegenerative disorders such as dementia <sup>(167)</sup>. These injuries can be exacerbated by the overexcitation of *N*-methyl-D-aspartate-sensitive glutamate receptors which results in an excessive influx of calcium through associated channels. However, glutamate is crucial for the regulation of the nervous system. Complete or even excessive inactivation of this receptor is not appropriate due to the receptors vital role in the nervous system. Furthermore, the intrinsic reactivity of the covalent warhead needs to be closely examined. Compounds with highly reactive functional groups or active metabolites can cause adverse drug reactions due to off-target effects <sup>(142)</sup>.

Compounds which label non-catalytic residues as a means to achieve selectivity can be susceptible to resistance. For example, osimertinib is a covalent inhibitor that is used as a treatment for therapy resistant lung cancer. It acts by targeting a non-catalytic cysteine on the epidermal growth factor receptor (EGFR). However, since this cysteine residue is not essential for EGFR function, resistance has been found to emerge through a cysteine point mutation (C797S) <sup>(168)</sup>. This mutation removes the covalent mechanism of the drug without affecting protein function and thereby results in resistance to the treatment. However, the development of drug resistance through mutation of target residues is not an issue unique to covalent inhibitors, and despite this limitation, osimertinib is a clinically approved drug.

Lastly, development of covalent compounds cannot be ranked and evaluated using the same metrics as their non-covalent counterparts. The nonequilibrium binding and time dependent effects make the use of  $K_D$  and  $IC_{50}$  values inappropriate or difficult to implement and interpret <sup>(144, 169)</sup>. The use of  $k_{inact}/K_i$  examines the kinetics and efficiency of the formation of the covalent complex <sup>(169)</sup>.  $k_{inact}$  describes the rate of

covalent bond formation and the  $K_i$  value describes the concentration required to achieve half maximal occupancy, the metric can be used to describe the relationship between drug concentration and rate of reaction. Another commonly used metric for comparisons are the classical Michaelis-Menten kinetics ( $K_m$ )<sup>(170, 171)</sup>.  $K_m$  demonstrates the relationship between the substrate concentration and half maximal velocity of bond formation using the following relationships:

$$\text{Equation 1.10} \quad K_m = (k_{off} + k_{inact}) / k_{on} \quad (170, 171)$$

$$\text{Equation 1.11} \quad V_0 = (V_{max}[S]) / (K_m + [S]) \quad (170, 171)$$

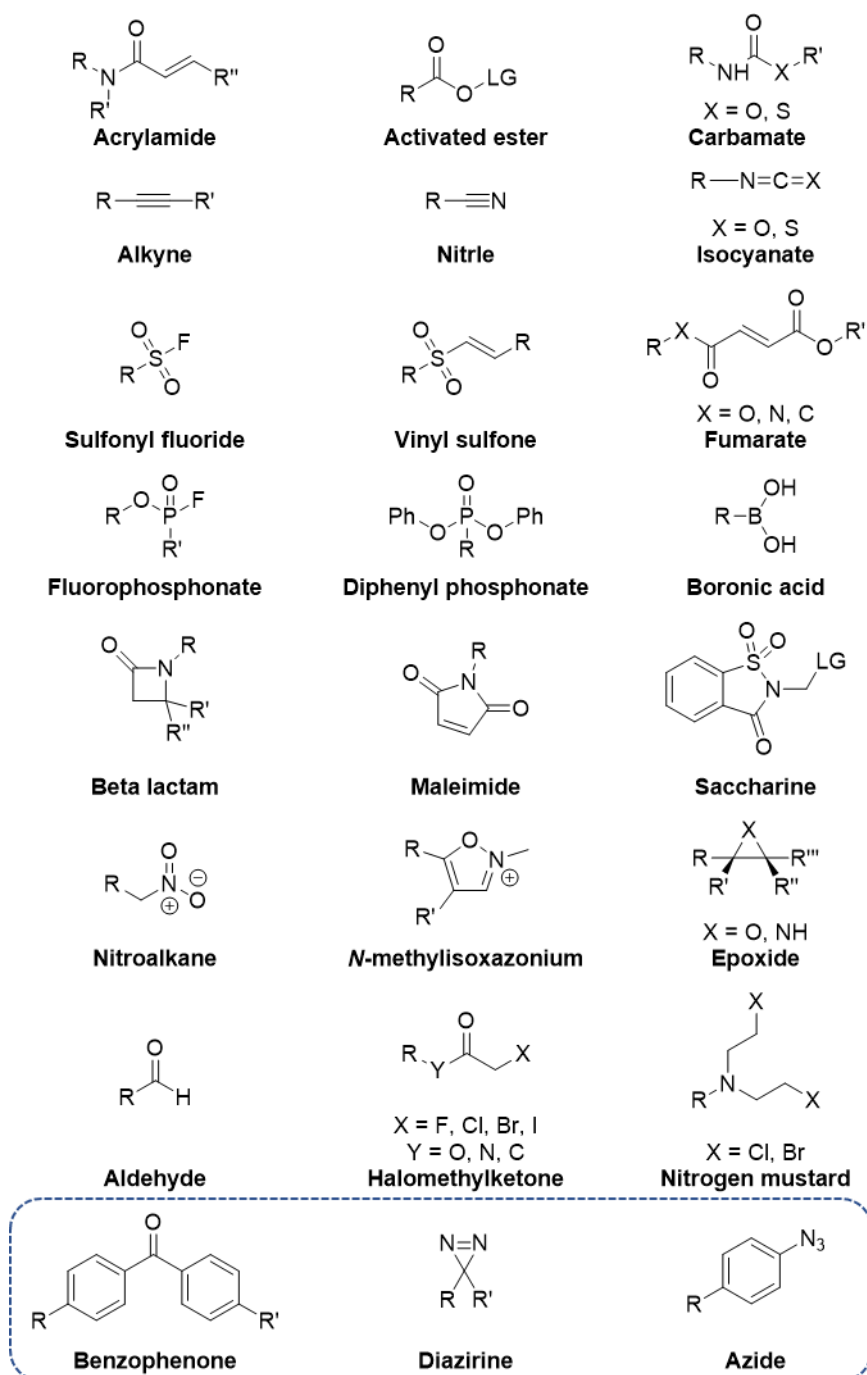
Where  $K_m$  is the Michaelis constant,  $V_0$  is the initial velocity and  $V_{max}$  the maximum velocity of inaction and  $[S]$  is the substrate concentration. These values can be more time consuming to obtain, however, they provide greater insights to structure-activity relationship, selectivity, pharmacokinetics and target occupancy<sup>(169)</sup>.

### 1.2.3 Inhibition and reactive functional groups

There are many different electrophiles which have the ability to label nucleophilic residues of proteins, as well as cofactors and DNA. Of the 88 FDA approved drugs which work by a covalent mechanism, there are over 35 different types of warheads<sup>(152)</sup>. In spite of this,  $\beta$ -lactams (generally antibiotic drugs), acrylamides and aziridiniums (prodrugs) made up 39 % of the warheads. Furthermore, cysteine and serine residues are the target for 62 % of compounds, representing a clear bias for warhead and target<sup>(152)</sup>. Despite the commonality in target residue and warhead, reactivity can be difficult to predict. Reactivity can be dictated by the intrinsic properties of the functional group or by the affinity-driven placement of warheads in close proximity to nucleophiles.

Selectivity and potency can be challenging to tune however, it is generally approached by optimising a non-covalent motif for high affinity and functionalisation with a relatively unreactive warhead. This is the general concept and strategy behind targeted covalent inhibitors drug discovery which has championed the resurgence of covalent inhibition

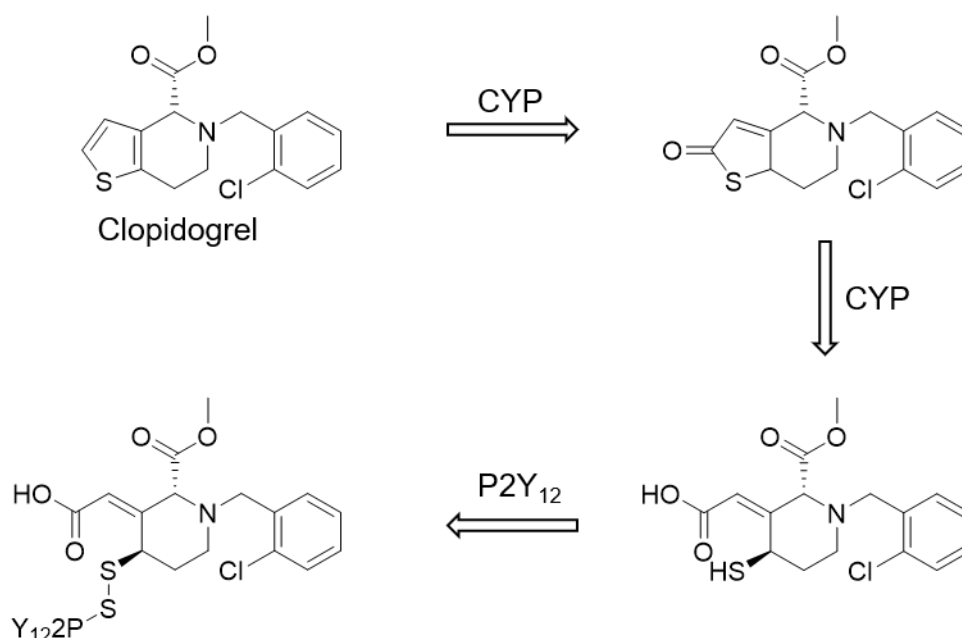
in drugs. The electrophilic warheads operate by one of three types of inhibition; mechanism-based (enzyme catalysed), irreversible or reversible.



**Figure 1.15:** Non-exhaustive examples of covalent warheads. Warheads which require photoactivation are denoted by the blue dashed box, leaving groups denoted by LG. <sup>(152, 156, 172-174)</sup>

### 1.2.3.1 Mechanism-based inhibition

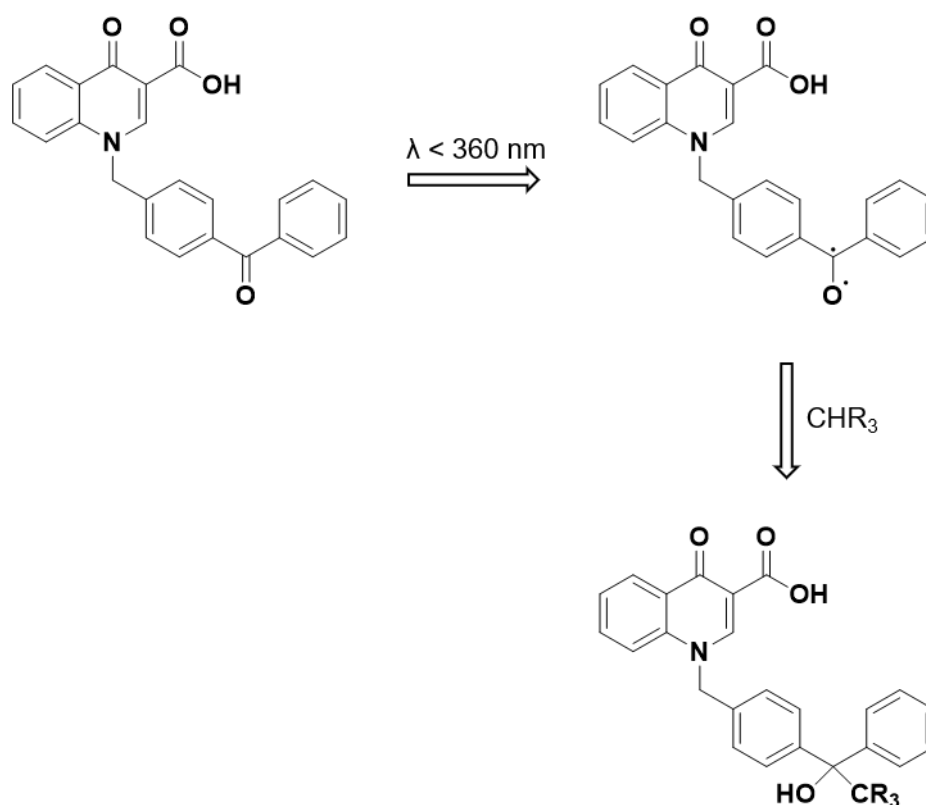
Mechanism-based inhibitors are considered difficult to design and, consequently, they have often been identified phenotypically, with their mode of action discovered retrospectively. These inhibitors bind at the active site of an enzyme and are processed as a substrate. This enzyme may be the intended target or an enzyme used for compound metabolism. This process yields a functional group which covalently labels the desired protein. For example, the P2Y<sub>12</sub> purinoceptor plays an important role in platelet aggregation and has been implicated in strokes. Clopidogrel is oxidised to a thiophenone by CYP450 and subsequent hydrolysis to the open ring results in a free thiol metabolite (<sup>175</sup>). This active metabolite can act as a covalent inhibitor and form a disulfide bond to a cysteine of the P2Y<sub>12</sub> purinoceptor (Figure 1.16) (<sup>176</sup>). This type of enzyme catalysed mechanism generally uses highly reactive functional groups. This can lead to a lack of selectivity, where the activated functional group labels the most accessible residues, and this in turn can result in unwanted toxicity (<sup>144, 175</sup>). For these reasons modern strategies tend to avoid designing compounds in this way.



**Figure 1.16:** Mechanism-based inhibition of P2Y<sub>12</sub> by Clopidogrel.

### 1.2.3.2 Photoactivated inhibition

Photoreactive groups can be useful in targeting residues which are normally unreactive. An inert functional group is activated through exposure to a particular wavelength (functional group dependent) to form a highly reactive motif. Due to the high reactivity obtained it is possible to target unreactive, and even non-nucleophilic, residues <sup>(172, 177)</sup>. This allows for more flexibility in positioning of this motif as it does not require proximity to a specific nucleophilic residue. However, the requirement for UV dependent activation means that these compounds are not generally suitable as therapeutics or tool compounds for *in vivo* experiments <sup>(172, 178)</sup>. However, they can be useful compounds to probe mechanisms because they do not require the covalent warhead to be positioned close to a nucleophilic residue on the protein.



**Figure 1.17:** Photoactivated inhibition of the M<sub>1</sub> muscarinic acetylcholine receptor.



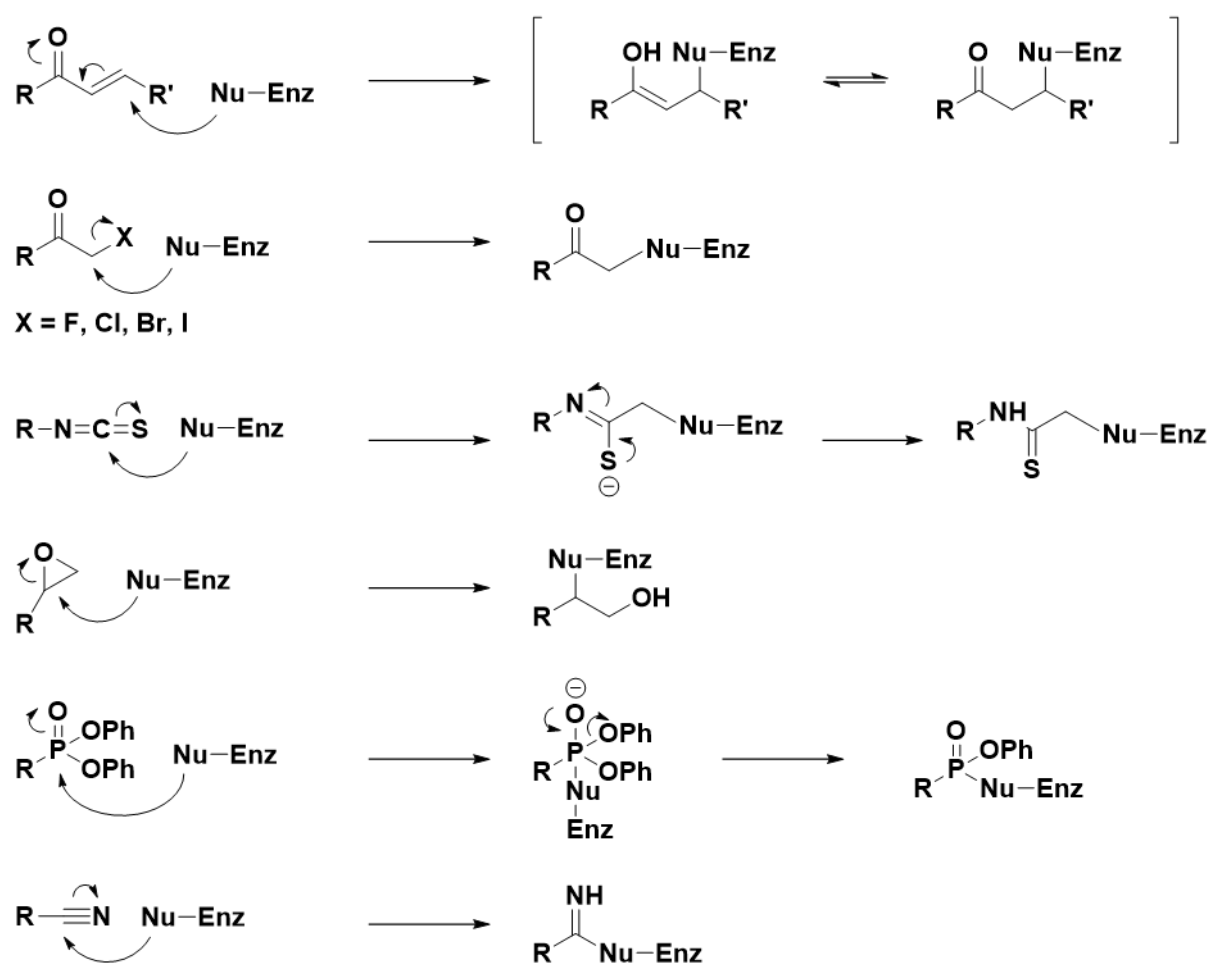
A group at Monash University developed a photoactivated covalent ligand to bind to and probe the function of an allosteric site on the M<sub>1</sub> muscarinic acetylcholine receptor (M<sub>1</sub> mAChR) (Figure 1.17) <sup>(179)</sup>. This family of receptors is implicated in neurological disorders such as schizophrenia and Alzheimer's disease and characterisation of the structure and mechanism of these GPCRs could advance the design of cognitive therapeutics <sup>(180)</sup>. A known positive allosteric modulator of M<sub>1</sub> mAChR (benzyl quinolone carboxylic acid, BQCA <sup>(181)</sup>) was modified to contain a photoactivatable benzophenone component <sup>(182)</sup>. This analogue was shown to have a similar binding, cooperativity and specificity profile of BQCA against the mAChR family. The ability to mimic the binding of BQCA was thought to be an important characteristic in characterising the effects of BQCA on the structure and function of the GPCR. In contrast, a previous covalent ligand designed by the same group contained an isothiocyanate warhead <sup>(182)</sup>. While this compound was also an irreversible allosteric binder of M<sub>1</sub> mAChR, it was likely that significant conformational change in the mode of binding was necessary to facilitate the covalent labelling of nearby nucleophilic residues. Due to the importance of conformational change and protein dynamics in GPCR signalling it was reasoned that the benzophenone was a better tool compound for probing the effects of BQCA.

### 1.2.3.3 Chemoreactive inhibition

Chemoreactive functional groups are generally electrophilic in nature and are used to target nucleophilic amino acids <sup>(157)</sup>. The implementation of this covalent inhibition generally uses weakly reactive functional groups to allow for compound stability and target selectivity. While these warheads need to be able to readily undergo reactions with their target protein they must remain intact in water prior to reaching their destination <sup>(172)</sup>. Moreover, these reactions must be selective enough to their target in order to avoid adverse side effects through promiscuity.

Chemoreactive inhibition can occur via various reactions, including but not limited to; Michael addition, alkylation, acylation, sulfonylation, phosphorylation, oxidation and Pinner-type reactions <sup>(156, 172, 174)</sup>. This broad range of reactions allow an expansive number of warheads to be used. Furthermore, based on reaction type and intrinsic

reactivity of the warhead, these reactions can either be reversible or irreversible. The reactivity of the functional groups is often tempered by the compound motif responsible for molecular recognition <sup>(157)</sup>. Not only can this part of the compound alter the electrostatic properties of the warheads, higher affinity compounds can increase the rate of reaction through an increase in local concentrations. For this reason, weakly reactive warheads can be employed in covalent inhibitors.



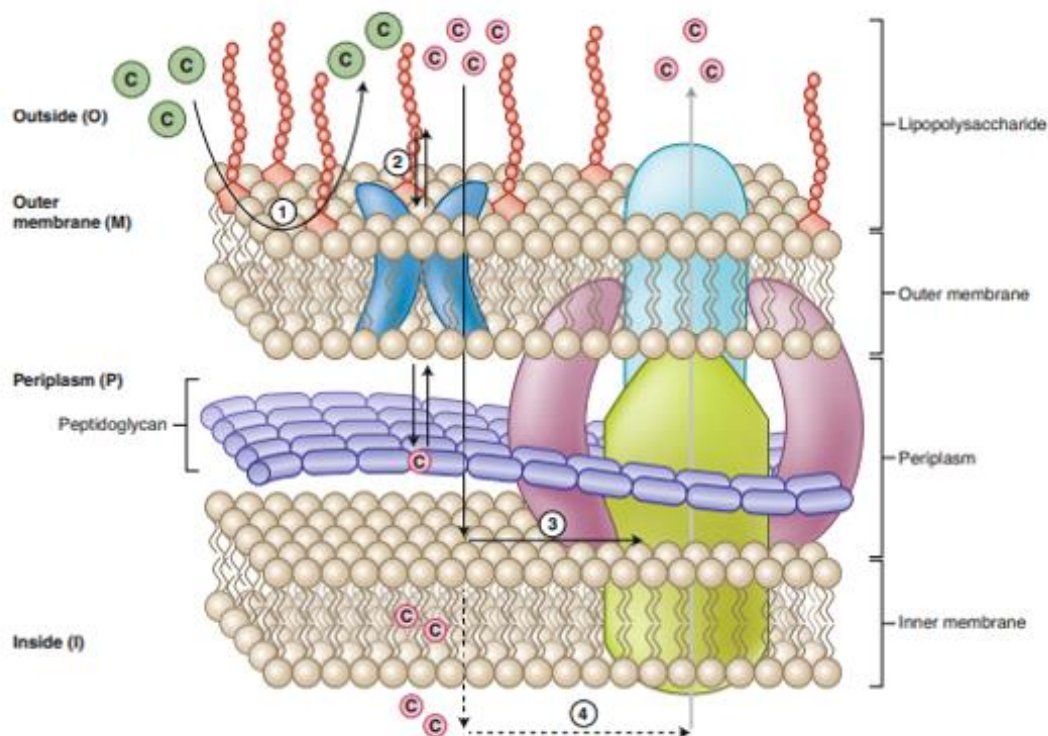
**Figure 1.18:** Non-exhaustive examples of chemoreactive inhibition reaction mechanisms.

### 1.3 Antibacterial resistance and bacterial virulence

One of the most significant discoveries of modern medicine has been the use of antibacterial agents to treat infections, conversely, the emergence of widespread antibacterial resistant strains of bacteria is considered to be one of the largest health challenges to be faced in the 21<sup>st</sup> century. Despite the growing need, introduction of new treatments have dwindled as large pharmaceutical companies have abandoned this area of research in lieu of more lucrative projects <sup>(183)</sup>. No new classes of antibiotics against Gram-negative bacteria have been discovered for more than 50 years <sup>(184)</sup> and only two new classes of antibiotics in general have been discovered in the last 20 years <sup>(183)</sup>. Although the pipeline of antibacterial development is undergoing a revitalisation with 30 new chemical entities currently in clinical trials <sup>(185, 186)</sup>, the problem is mostly addressed through repurposing of, or improving upon, old antibiotics, known scaffolds and existing drug classes <sup>(187)</sup>. This lack of investment in new compounds that address novel targets and mechanisms of action is why the trajectory for antibacterial resistance still stands to see annual deaths increase from the current 700,000 <sup>(188)</sup> to 10 million by the year 2050 <sup>(189)</sup> at an estimated cost of \$100 trillion USD <sup>(189)</sup>.

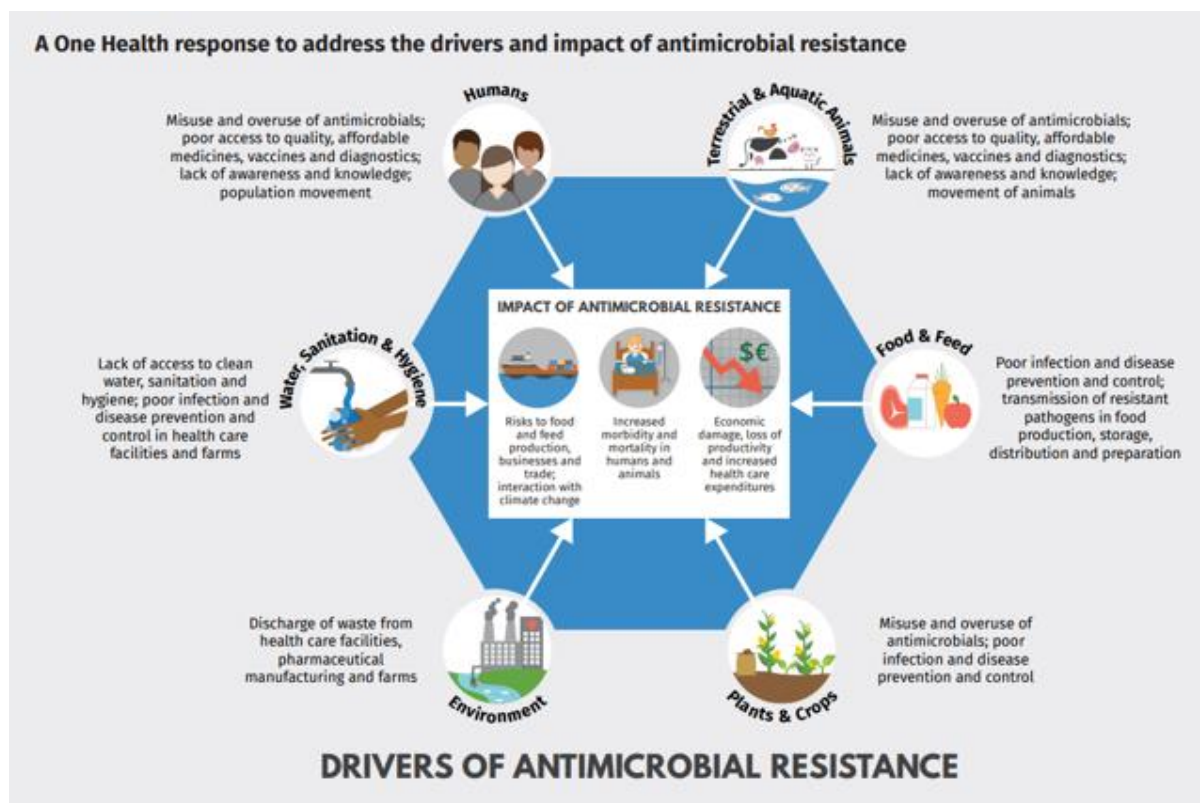
Existing antibiotics work by either adopting a bacteriostatic (inhibiting the growth of bacteria) or bactericidal (reduction in bacterial viability) approach. These drugs target processes which play a role in essential bacterial pathways such as peptidoglycan synthesis ( $\beta$ -lactams), DNA replication and transcription (quinolones), protein synthesis (tetracyclines), and folate synthesis (sulfonamides) <sup>(190)</sup>. Resistance to these drug classes is most concerning for Gram-negative bacteria which display high morbidity and mortality rates. *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Klebsiella* (species plural), *Enterobacter* (species plural) have demonstrated resistance to multiple antibiotic classes and are amongst the highest priority species for research and development <sup>(191)</sup>. The development of drugs for the treatment of Gram-negative infections is particularly challenging due to their unique cellular structure whereby two cellular membranes are present. The outer membrane is rich in lipopolysaccharides, which presents an effective barrier against large and hydrophobic compounds, and limits drug permeability. The inner membrane presents an effective barrier to the small and polar compounds, which are able to cross the

outer membrane. In addition, efflux pumps are able to remove antibiotics from both periplasm and cytosol (Figure 1.19) <sup>(192)</sup>.



**Figure 1.19:** The Gram-negative cell envelope and pathways of drug fluxes across it. The envelope includes a network of lipopolysaccharides that limits penetration of large and hydrophobic compounds (green C; 1), porin channels that permit passage of small, hydrophilic molecules (red C; 2), and trans-envelope efflux pumps that capture their substrates from the periplasm (3) or cytosol (4) and pump them out of the cell. Reprinted by permission from Springer Nature: Nature Chemical Biology, Defining new chemical space for drug penetration into Gram-negative bacteria. Zhao, S., Adamiak, J.W., Bonifay, V. *et al.* Copyright 2020 <sup>(192)</sup>

Antibacterial resistance arises due to a variety of reasons, including in response to poor infection and disease control, inappropriate waste disposal of antibiotic compounds by pharmaceutical manufacturers and health care facilities, and most importantly, the abuse and misuse of current antibiotics in humans, animals and farming of livestock and crops (Figure 1.20) <sup>(185, 188, 189)</sup>. The overexposure to antibiotics imposes severe evolutionary pressure on the bacterial cells and results in the appearance of antimicrobial resistant strains.



**Figure 1.20:** Main causes, drivers and impacts of antimicrobial resistance. Reprinted from World Health Organisation No time to wait: securing the future from drug-resistant infections. World Health Organisation, Geneva, 2019 <sup>(188)</sup>

Although resistance can occur naturally, the major resistance issues stem from acquired or transferred means <sup>(193)</sup>. Bacteria can acquire resistance through a variety of different mechanisms. One is to reduce the amount of active antibiotic that is present in the cell. This can involve changes to the influx and efflux of drugs as well as their efficacy or metabolism within the cell <sup>(193)</sup>. As mentioned above Gram-negative bacteria are able to reduce the permeability of drugs through the cell membrane through lipopolysaccharides, however, this can be further exacerbated by changes in the type, amount or function of the water-filled diffusion channels that are present in the outer membrane and are known as porins. These changes are able to reduce the amount of a drug that enters that bacterial cell. In addition, upregulation or increased activity of efflux pumps can reduce the effective drug concentration by increasing the rate at which bacteria expel the antibiotics from within the cell. Other mechanisms that reduce the efficacy of antibiotics have also been observed. For example, bacteria have

been shown to overproduce proteins that are the targets of an antibiotic, which requires more of the antibiotic to be present to achieve the same effect, or overproduce non-essential proteins that bind to the antibiotic and thereby negate their effects on the desired metabolic pathway. Bacteria are also able to produce modifying enzymes which can chemically modify antibiotics, thereby decreasing the affinity and efficacy of the drug for its target. Furthermore, conformational changes or mutations at non-essential residues in the binding site of the target can further decrease the affinity of the drug and prevent functional inhibition. These adaptations and advantages can be passed on using horizontal gene transfer between different organisms to rapidly spread resistance.

Addressing the problem of bacterial resistance to current antibiotics will require the introduction of new treatments, preferably with novel mechanisms of actions, and ideally ones which do not exert profound selective pressures. Virulence, or pathogenicity, is the ability of an organism to cause disease and infect the host, however, virulence is not an essential function for bacterial survival. This is potentially an attractive trait, as it means that targeting virulence can inhibit disease without exerting a strong selection pressure for the development of resistance. Many factors are involved in virulence and so this strategy also offers an abundance of new pharmacological targets. Antivirulence compounds should allow preservation of native microbiota thereby leading to a reduction of colonisation by the pathogenic bacterium (194, 195). It is hypothesised that this combination of factors would allow the host to clear the infection (194-198). Virulence factors include proteins or protein complexes which mediate various functions including adhesion, biofilm formation and colonisation, toxins, their transcription factors and secretory pathways (type II and III secretion systems), quorum sensing, two-component response systems and motility (194-199). Disruption of any of these virulence pathways has the potential to neutralise the pathogen and either make it more susceptible to antibiotic treatment or clearance by the host (195, 197, 200).

Although the FDA has approved antibody treatments targeting pathogen specific exotoxin virulence factors, there are presently no small molecule antivirulence inhibitors approved for therapeutic use (199). Of the 30 new chemical antibacterial

entities which are progressing through the clinical pipeline, none work via an antivirulence mechanism <sup>(185, 186)</sup> and the 33 antivirulence projects in preclinical trials are the currently most advanced treatments of this kind <sup>(200)</sup>. Therefore, targeting virulence factors or regulators presents an exciting new opportunity for novel treatments against multidrug resistant bacteria.

## **1.4 *Escherichia coli* DsbA (EcDsbA)**

### **1.4.1 DsbA as an antivirulence target**

In Gram-negative bacteria a family of oxidoreductase enzymes are responsible for the introduction and isomerisation of disulfide bonds in peptides and proteins. These proteins form the Dsb enzyme family. Once exported into the periplasm, protein substrates require these Dsb enzymes to catalyse formation of intramolecular disulfide bonds that are necessary for their correct folding, stability and function. In *Escherichia coli* (*E. coli*) there are potentially over 300 proteins that contain disulfide bonds, many of which play a role in virulence <sup>(201, 202)</sup>. These proteins include FlgI, which is a component of the flagellar motor and is necessary for motility <sup>(203)</sup>, PapD the P fimbriae molecular chaperone <sup>(204)</sup>, the heat-labile enterotoxin (LT) <sup>(205)</sup>, and outer membrane secretin EscC <sup>(206)</sup>. Many of these virulence proteins require DsbA in order to fold correctly <sup>(207)</sup>, thereby making DsbA a key regulator of bacterial virulence.

Bacteria in which the *dsbA* gene is either deleted or inactivated by mutation have been shown to give pleiotropic phenotypes in multiple strains of Gram-negative bacteria. This is consistent with the involvement of DsbA in the folding of multiple proteins. It is almost always observed that bacteria lacking a functional DsbA are avirulent in both *in vitro* and *in vivo* infection models <sup>(208-217)</sup>. One such study investigated the role of DsbA in the virulence of *Burkholderia pseudomallei* (*Bps*), the causative pathogen of the melioidosis <sup>(217)</sup>. A *dsbA* deletion strain of *Bps* was shown to be defective in both secretion and motility. In animal infection models, mice which were subjected to the wild type strain of *Bps* showed 100 % mortality rate at 42 days post injection, whereas the mice which were challenged with the  $\Delta dsbA$  *Bps* did not result in any deaths. The

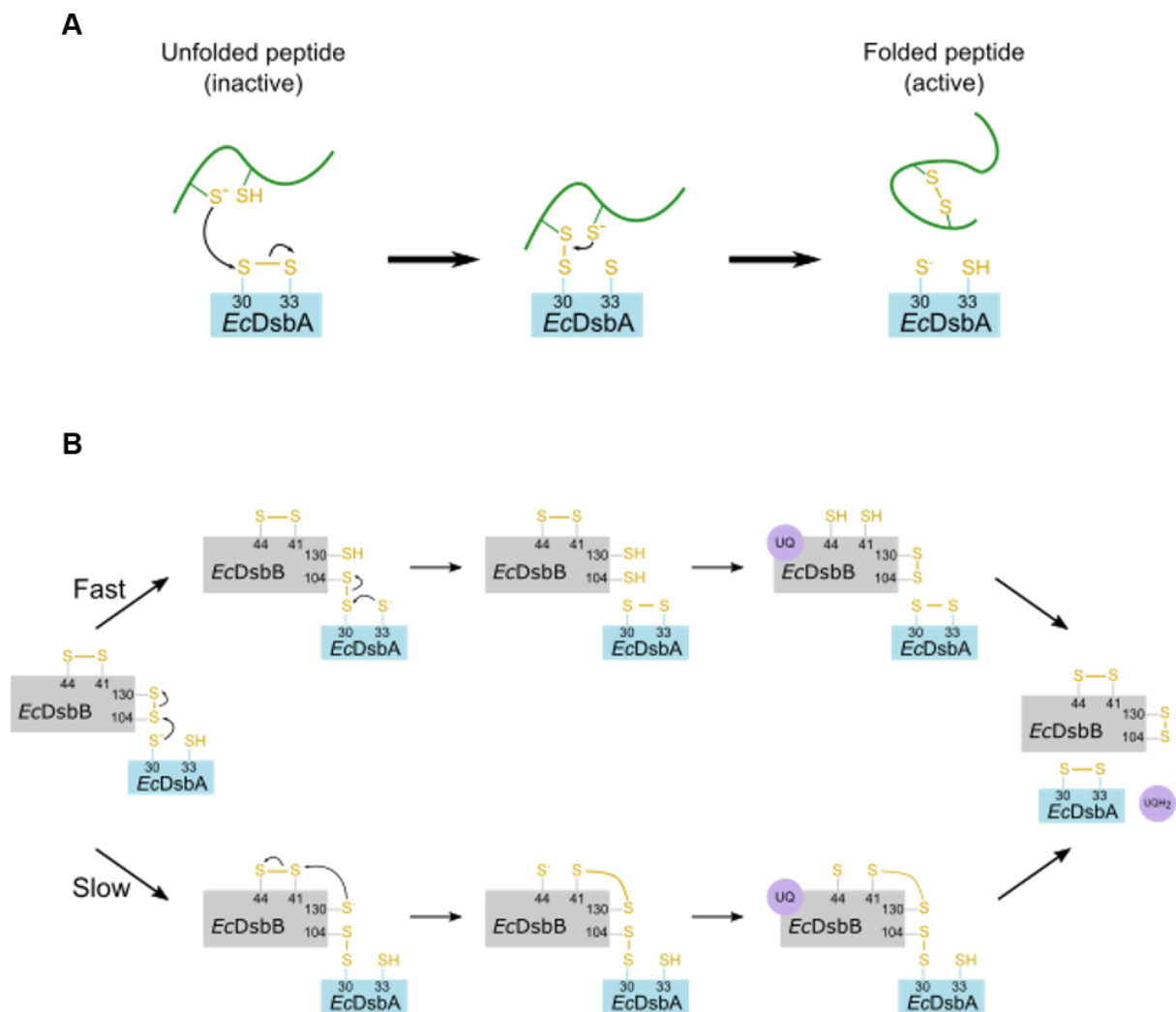
surviving animals had failed to clear the infection, (with bacterial colonies observed in the liver, lungs and spleen) proving that this mutant did not affect bacterial viability. Therefore, the study infers that targeting DsbA would be able to suppress pathogenicity but would not result in the rapid development of resistance against DsbA based treatments.

DsbA from *E. coli* K-12 (*EcDsbA*) is considered the archetypal enzyme and is the most characterised DsbA to date. Due to its prevalence, place on the World Health Organisation priority list and its extensive characterisation, the *E. coli* strain of DsbA is the main focus of this body of work.

#### **1.4.2 Mechanism and catalytic cycle of *EcDsbA***

*EcDsbA* is able to introduce disulfide bonds into unfolded substrates through a series of disulfide exchanges between the protein, *EcDsbA*, and *EcDsbB*, with the oxidising potential of the system being derived from ubiquinone (Figure 1.21) <sup>(218)</sup>. The substrate is drawn into a hydrophobic groove on the surface of *EcDsbA* where its thiol attacks the solvent exposed Cys30 of the *EcDsbA* active site in a nucleophilic substitution reaction ( $S_N2$ ) <sup>(219, 220)</sup>. The formation of this bond requires a planar arrangement of the sulfur atoms involved and results in an intermediate where all residues share the charge <sup>(219, 220)</sup>. This intermediate contains an intermolecular disulfide bond which is subsequently attacked by the second cysteine of the substrate, releasing the oxidised substrate protein and reduced *EcDsbA*.





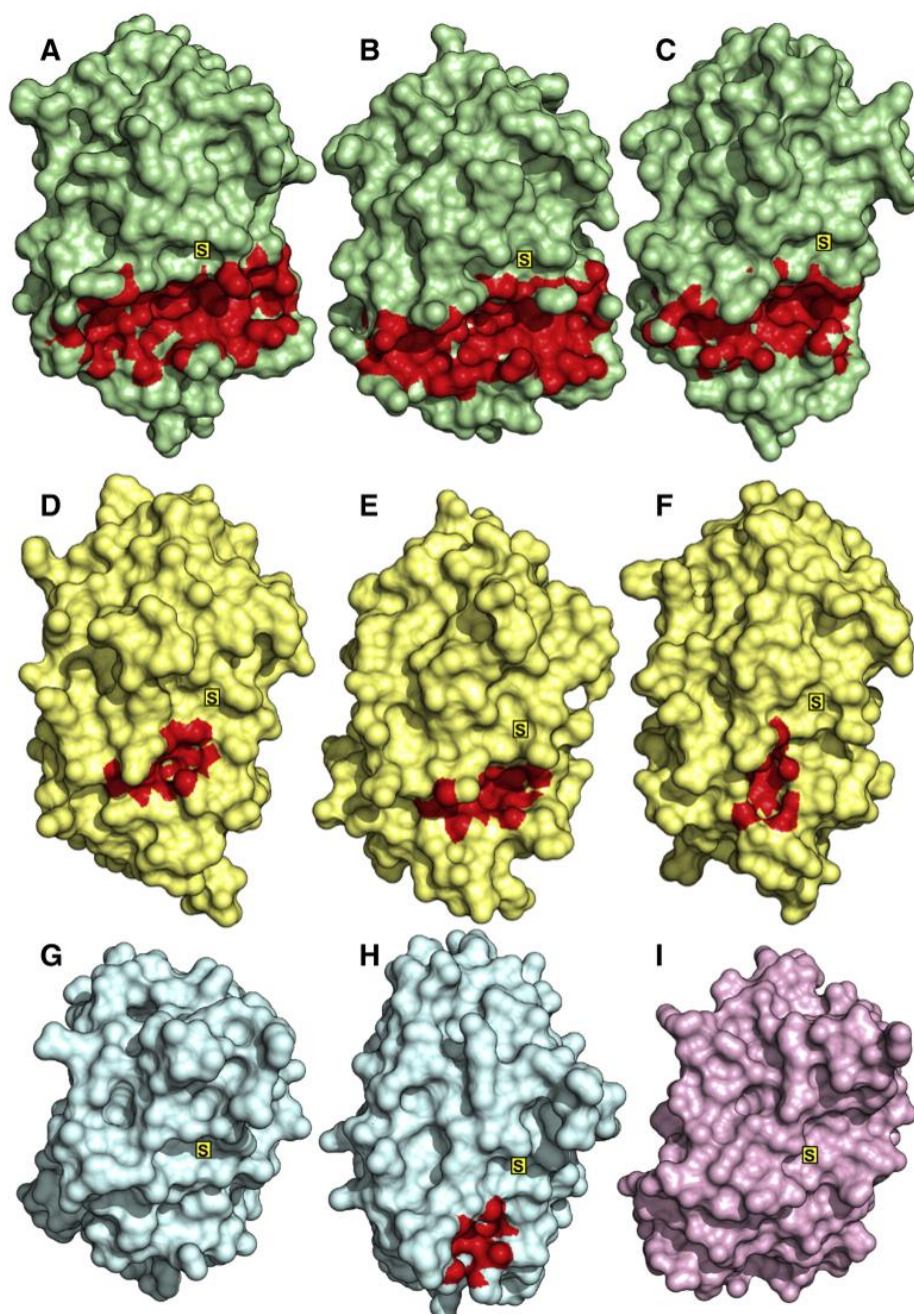
**Figure 1.21:** Catalytic cycle of *EcDsbA* and redox partner *EcDsbB*. A) Introduction of disulfide bond into peptide substrates by *EcDsbA* B) Oxidation of *EcDsbA* by *EcDsbB* and ubiquinone

*EcDsbB* has two disulfide bonds which participate in the redox cycle, Cys104-Cys130 and Cys41-Cys44<sup>(221)</sup>. Electrons flow from *EcDsbA* through these bonds sequentially, through the DsbB co-factor ubiquinone and on through the rest of the respiratory electron transport chain<sup>(218)</sup>. The unidirectional flow of electrons in *EcDsbB* is thought to be promoted by conformational changes in the protein, including the relocation of Cys130 to the vicinity of Cys41-Cys44, and the forced spatial separation of Cys104 and Cys130 by *EcDsbA*'s Met64<sup>(222)</sup>. This then promotes the attack of the intermolecular bond by Cys33 of DsbA (fast pathway)<sup>(223)</sup> or the attack of Cys41 by

Cys130 of DsbB (slow pathway) <sup>(221, 224)</sup> and continues the remainder of the redox cycle to finally result in the regeneration of oxidised *EcDsbA* and *EcDsbB*.

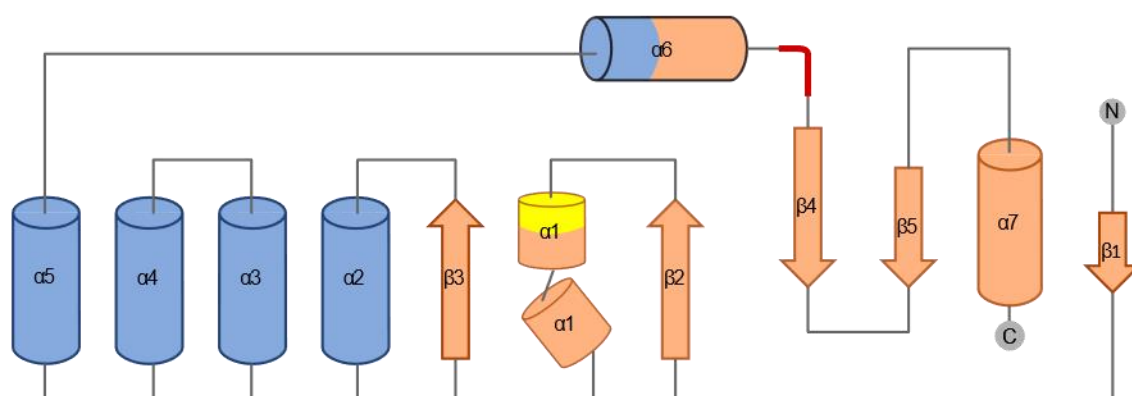
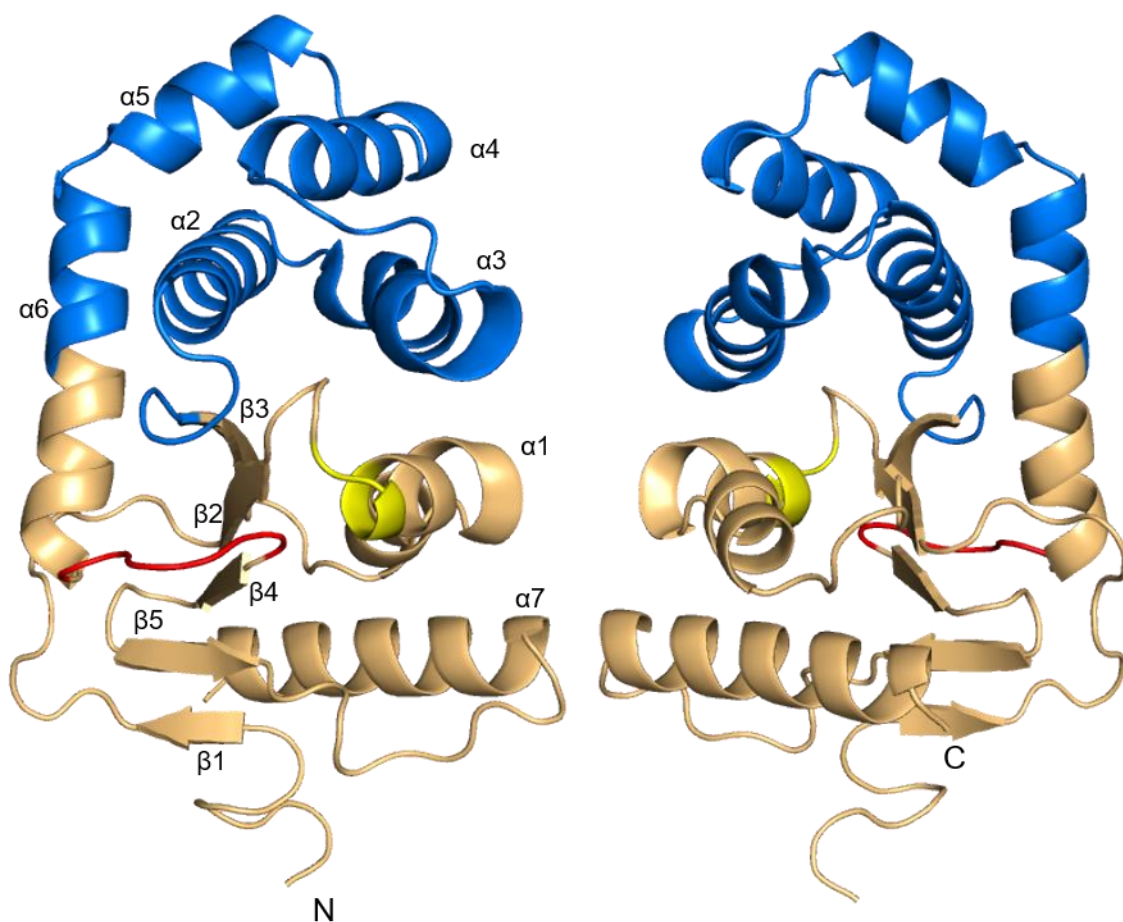
### 1.4.3 Structure and function of *EcDsbA*

The DsbA family of enzymes from different bacteria often do not share a high sequence identity, however, the main structural architecture remains the same <sup>(225, 226)</sup>. These proteins comprise of a thioredoxin fold and an inserted  $\alpha$ -helical domain with highly conserved sequence of residues at the catalytic site CPHC which is located in spatial proximity to a *cis*-Pro loop motif. Although most DsbA enzymes maintain these two features, the topology and interactions between  $\beta$ -sheets in the structure have led to the characterisation of the enzymes into two classes (I and II) <sup>(227)</sup>. These classes are further separated into two subclasses (a and b) based on the surface loops surrounding the catalytic cysteines. These subclasses also result in distinct  $\alpha$ -helical features (Figure 1.22) and surface topology which has been implicated in substrate recognition <sup>(225, 227-231)</sup>. This suggests that it would be difficult to develop a broad spectrum DsbA inhibitor, however, it may be possible to develop a subclass specific inhibitor <sup>(227, 232)</sup>.



**Figure 1.22:** Structural characteristics of DsbA subclasses. A large groove on the catalytic surface is common to the DsbA-Ia enzymes (green): A) *EcDsbA* B) *SeDsbA* C) *VcDsbA*. In contrast, DsbA-Ib members (yellow) typically have a small surface pocket on the catalytic surface located between loops L2 and L3: D) *BpsDsbA* E) *XfDsbA* F) *PaDsbA*. DsbA-II enzymes (IIa =blue, IIb = purple) do not generally have a defined pocket near the catalytic site: G) *MtbDsbA* and I) *WpDsbA1*, with the exception of H) *BsDsbA*. In each case, the position of the active site cysteine is indicated by “S” and the region of the surface groove or pocket is shown in red. Reprinted by permission from Elsevier: *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics*, Four structural subclasses of the antivirulence drug target disulfide oxidoreductase DsbA provide a platform for design of subclass-specific inhibitors. McMahon, R. M, Premkumar, L., Martin, J. L. Copyright 2014 <sup>(227)</sup>

*EcDsbA* contains five  $\beta$  sheets ( $\beta 1 - 5$ ) and seven  $\alpha$  helices ( $\alpha 1 - 7$ ). The thioredoxin fold encompasses  $\beta$  sheets  $\beta 2 - \beta 5$  and  $\alpha$ -helices  $\alpha 1$ ,  $\alpha 7$  and partially includes  $\alpha 6$ , whereas the inserted  $\alpha$ -helical domain spans from  $\alpha 2 - \alpha 6$  (Figure 1.23) <sup>(230)</sup>. The active site, <sup>30</sup>CPHC<sup>33</sup>, is located at the *N*-terminal end of the  $\alpha 1$  helix, and although distant in sequence, the *cis*-Pro loop lies adjacent to the active site on the loop connecting  $\alpha 6$  and  $\beta 4$ . *EcDsbA* is part of the Ia subclass <sup>(227)</sup>, and therefore has a large and solvent exposed hydrophobic groove on the catalytic face of the protein (referred to as the “front” of the protein for the remainder of the thesis). This hydrophobic groove is thought to be the main substrate binding site and has been shown to be the interface where *EcDsbA* interacts with *EcDsbB* (Figure 1.24) <sup>(222)</sup>. Furthermore, recent studies have identified an internal cryptic pocket whose entrance is guarded by two structural waters adjacent to the *cis*-Pro loop <sup>(233)</sup>. Currently the mechanistic or biological relevance of this internal pocket is unknown, however two phenylalanine residues (Phe26 and Phe93), which make up a large portion of the internal surface, have been shown to be highly conserved throughout the DsbA family <sup>(225)</sup> and may play a potential role in function.

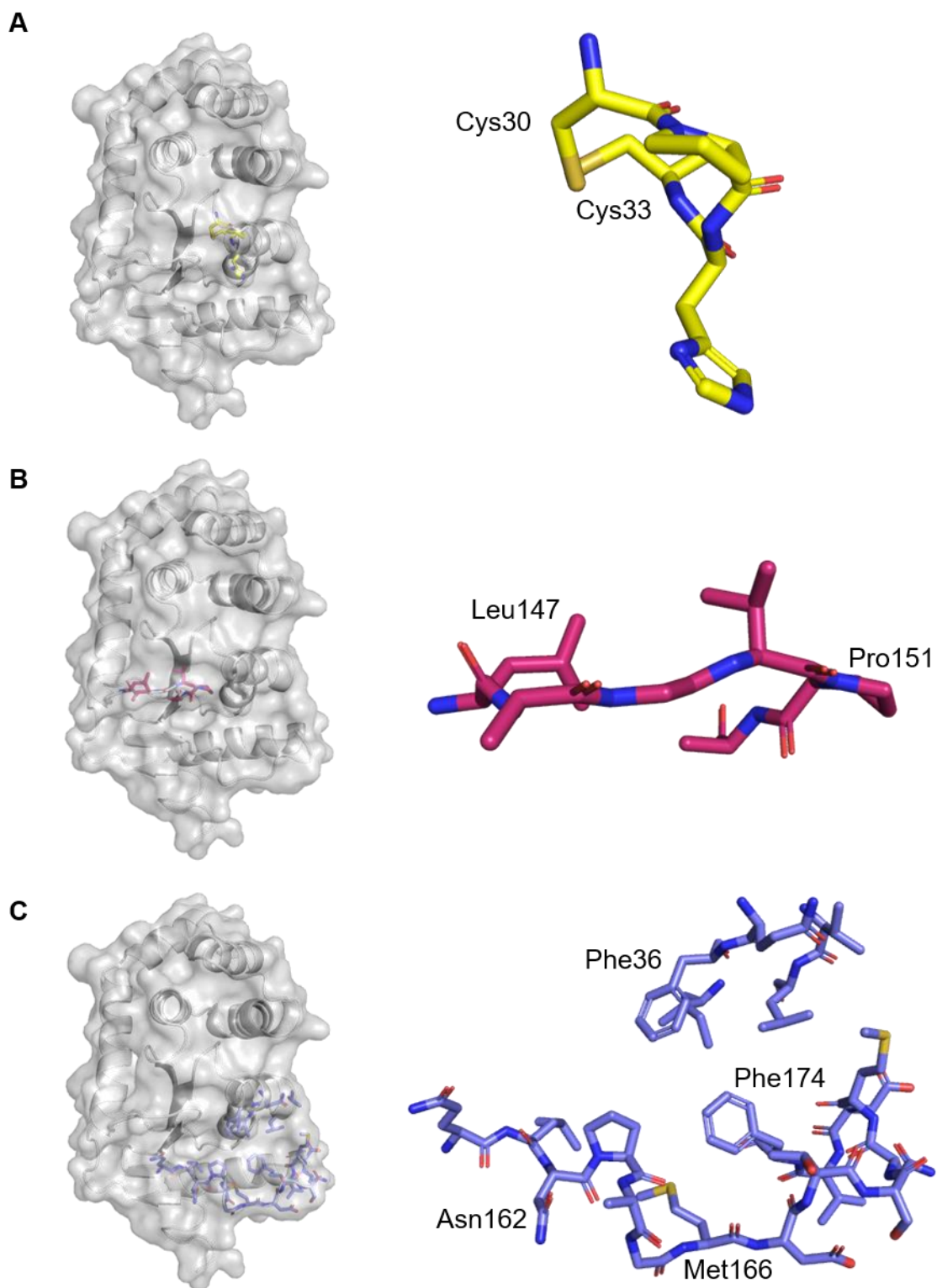


**Figure 1.23:** Structural domains of *EcDsbA* (PDB ID: 1FVK). Thioredoxin fold shown in orange, inserted  $\alpha$  helical domain shown as blue, active site CPHC shown in yellow and *cis*-Pro loop shown in red.

The active site cysteine of *EcDsbA* is highly reactive and has an oxidative potential of -122 mV and an uncommonly low cysteine pKa (Cys30  $\approx$  3.5, average cysteine  $\approx$  9) <sup>(234, 235)</sup>. The electrostatic interactions of the dipeptide sequence between the cysteines in the catalytic CXXC motif have been shown to modulate the cysteine pKa and contribute to the greater thermodynamic stability of the reduced state <sup>(234-236)</sup>. The stability of the reduced redox state of *EcDsbA* over oxidised further promotes the rapid disulfide exchange with unfolded substrates <sup>(237)</sup>.

The roles of the residues within the *cis*-Pro loop for the catalytic cycle was investigated using single point mutagenesis. Variability in these residues reduces enzymatic activity and protein stability, and sees the accumulation of intermolecular complexes between *EcDsbA* and either peptide substrates or redox partner DsbB <sup>(238, 239)</sup>. It was suggested that the *cis*-Pro conformation is important in maintaining the correct backbone orientation of the preceding loop. Crystal structures have been obtained of *EcDsbA* trapped in intermolecular complexes with DsbA and either substrates or *EcDsbB* <sup>(222, 229, 240-242)</sup>. These structures show that the periplasmic loop of *EcDsbB* is drawn into the putative substrate binding site and is anchored there by the intermolecular disulfide bond <sup>(222, 240, 241)</sup>. The section of *EcDsbB* which binds within this site forms an antiparallel  $\beta$  sheet with Arg148 and Val150 of *EcDsbA*. These interactions with the main chain of the *cis*-Pro loop are also observed for binding of covalently bound peptide substrates <sup>(229, 242)</sup>. The conservation of these hydrogen bonds implies that the substrate-DsbA interactions at this loop are vital to the oxidative cycle <sup>(239)</sup>.





**Figure 1.24:** Main structural features of EcDsbA (grey, PDB ID: 1FVK). A) CPHC catalytic site (yellow sticks) B) *cis*-Pro loop (magenta sticks) C) substrate binding groove (purple sticks).

The groove is approximately 6 Å tall, 7 Å wide, 5 Å deep and comprised of residues from the  $\alpha 1$  helix,  $\beta 5$  sheet,  $\alpha 7$  helix and the loop connecting  $\beta 5$  and  $\alpha 7$  (Figure 1.24) <sup>(243)</sup>. Binding at this site is driven by hydrophobic interactions and, as a consequence, generally results in weak binding <sup>(222, 242, 244, 245)</sup>. The bottom of the groove consists of residues from a flexible, unstructured loop and, together with the weak interactions with peptide substrates, is thought to facilitate the broad substrate specificity of *EcDsbA*. On the other hand, these interactions within the hydrophobic groove have been proven to be paramount for substrate binding. The reactions of *EcDsbA* with covalent peptides which make low affinity non-covalent interactions and *EcDsbA* with reduced glutathione (GSH) have been compared <sup>(245, 246)</sup>. The reactions with GSH displayed significantly slower kinetics than those with the peptides. This suggested that the formation of the mixed disulfide bond is proximity driven and promoted by non-covalent interactions within the hydrophobic groove. Therefore, despite the low affinity (often in the millimolar range <sup>(242, 246)</sup>) of these interactions, they still provide sufficient complex stability to enhance the oxidation of unstructured substrates.

Comparisons of DsbA sequences from different bacteria have identified a number of conserved residues <sup>(225, 243, 247)</sup> (Figure 1.25) and their importance to *EcDsbA* are slowly becoming evident. Valine residues Val22 and Val155 maintain the fold and positioning of the *cis*-Pro loop <sup>(225)</sup>, however surrounding residues <sup>156</sup>NGKY<sup>159</sup> are also highly conserved. These residues are flexible in solution (discussed below) <sup>(233)</sup> and potentially play a role in the release of substrates. Many charged and invariant residues are grouped in an acidic patch on the back of the protein, however, they do not directly affect the reactivity of DsbA. While they were shown to stabilise the oxidised protein, no definitive function has been identified for the residues <sup>(243, 247)</sup>. Furthermore, conserved phenylalanine residues Phe26 and Phe93 have been shown to form part of a recently discovered internal cryptic pocket <sup>(233)</sup>. While more extensive characterisation of these residues is required, the high conservation between DsbA variants suggests they hold an important role in protein function.



UniProt ID	Identity (%)	Strain	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
P0AEG4	100	<i>E. coli</i>	P	Q	V	L	E	F	F	S	F	F	C	P	H	C	Y	Q	F	E	E	V	L
P32557	38.3	<i>V. cholerae</i>	P	V	V	N	E	F	F	S	F	Y	C	P	H	C	N	T	F	E	P	I	I
P52235	99.5	<i>S. Flexneri</i>	P	Q	V	L	E	F	F	S	F	F	C	P	H	C	Y	Q	F	E	E	V	L
P52234	68.8	<i>D. dadantii</i>	P	Q	V	L	E	F	F	S	F	Y	C	P	H	C	Y	Q	F	A	Q	V	Y
P31810	44.8	<i>H. influenzae</i>	K	E	V	I	E	F	F	S	F	Y	C	P	H	C	Y	A	F	E	M	E	Y
P0A2H9	86.1	<i>S. typhimurium</i>	P	Q	V	L	E	F	F	S	F	Y	C	P	H	C	Y	Q	F	E	E	V	L
Q44504	28.4	<i>Az. Vinelandii</i>	I	E	V	V	E	L	F	W	Y	G	C	P	H	C	Y	Q	F	E	P	S	I
P50024	23.4	<i>L. pneumophila</i>	P	L	I	T	E	F	F	S	Y	G	C	P	W	C	Y	K	I	D	A	P	L
Q63Y08	24.6	<i>B. pseudomallei</i>	V	E	V	I	E	F	F	W	Y	G	C	P	H	C	Y	E	F	E	P	T	I
Q02DMO	25.1	<i>P. aeruginosa</i>	I	E	V	V	E	L	F	W	Y	G	C	P	H	C	Y	A	F	E	P	T	I
Q9EYL5	14.9	<i>S. aureus</i>	P	L	V	V	V	Y	G	D	Y	K	C	P	Y	C	K	E	L	D	E	K	V
80% consensus			.	.	V	I	E	a	F	.	a	h	C	P	H	C	Y	p	F	-	.	.	I

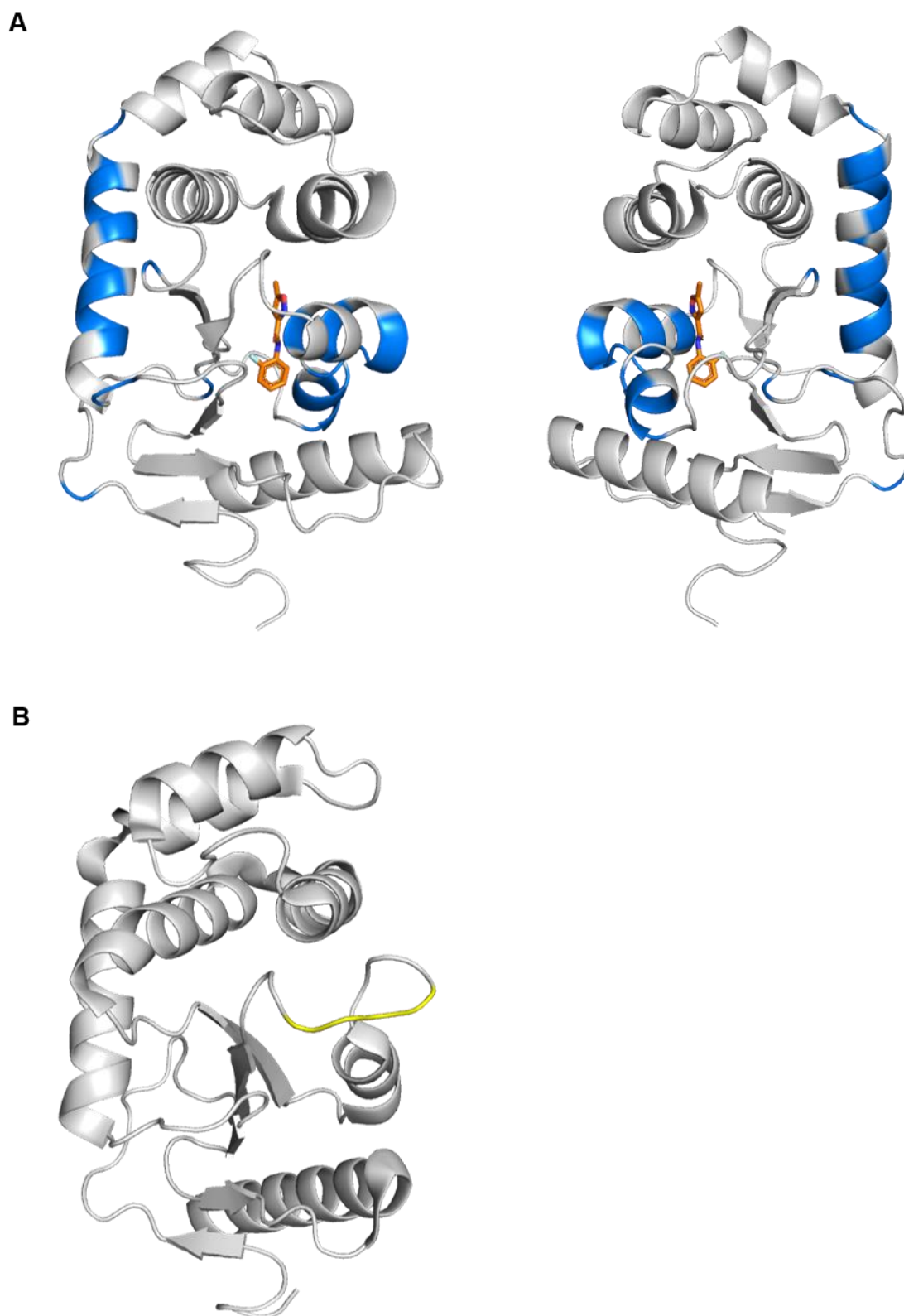
	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109			
<i>E. coli</i>	V	P	L	F	E	G	V	Q	K	T	Q	-	T	I	R	S	A	S	D	I	R		
<i>V. cholerae</i>	P	V	M	F	N	R	I	H	T	L	R	K	P	P	K	D	E	Q	E	L	R		
<i>S. Flexneri</i>	V	P	L	F	E	G	V	Q	K	T	Q	-	T	I	R	S	A	S	D	I	R		
<i>D. dadantii</i>	P	L	M	F	D	A	V	Q	K	T	Q	-	T	V	K	Q	P	Q	D	I	R		
<i>H. influenzae</i>	S	P	L	F	E	A	A	Q	K	D	-	-	A	L	K	S	M	D	D	I	R		
<i>S. typhimurium</i>	V	P	L	F	E	A	V	Q	K	T	Q	-	T	V	Q	S	A	A	D	I	R		
<i>Az. Vinelandii</i>	T	A	I	F	D	A	I	H	K	D	G	K	K	L	A	T	P	E	E	M	A		
<i>L. pneumophila</i>	P	I	L	F	K	A	I	Q	E	D	K	N	P	L	A	T	K	Q	S	M	V		
<i>B. pseudomallei</i>	P	A	V	F	N	A	I	H	K	E	K	N	Y	L	L	T	P	Q	A	Q	A		
<i>P. aeruginosa</i>	N	A	V	F	E	A	I	H	K	E	H	K	K	L	A	T	P	E	E	M	A		
<i>S. aureus</i>	K	Q	L	F	A	A	Q	Q	D	E	N	K	E	W	L	T	K	E	L	L	D		
80% consensus			s	.	I	F	p	u	I	p	c	p	p	.	.	I	t	o	.	p	p	h	t

	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	
<i>E. coli</i>	K	A	A	A	D	V	Q	L	R	G	V	P	A	M	F	V	N	G	K	Y	Q	
<i>V. cholerae</i>	K	Q	F	Q	D	S	G	L	T	G	V	P	A	V	V	V	N	N	R	Y	L	
<i>S. Flexneri</i>	K	A	A	A	D	V	Q	L	R	G	V	P	A	M	F	V	N	G	K	Y	Q	
<i>D. dadantii</i>	K	A	A	A	D	L	Q	L	R	G	V	P	A	V	F	V	N	G	K	Y	M	
<i>H. influenzae</i>	N	A	A	E	Q	F	K	V	R	G	V	P	D	F	Y	V	N	G	K	F	R	
<i>S. typhimurium</i>	K	A	A	A	D	L	Q	L	Q	G	V	P	A	M	F	V	N	G	K	Y	Q	
<i>Az. Vinelandii</i>	K	L	A	I	A	Y	Q	I	S	G	V	P	V	M	V	V	N	G	K	Y	R	
<i>L. pneumophila</i>	S	L	M	A	H	Y	Q	I	N	A	V	P	A	F	V	V	N	N	K	Y	K	
<i>B. pseudomallei</i>	E	L	L	K	N	Y	N	I	D	G	V	P	T	I	V	V	Q	G	K	Y	K	
<i>P. aeruginosa</i>	K	L	A	M	A	Y	Q	V	T	G	V	P	T	M	V	V	N	G	K	Y	R	
<i>S. aureus</i>	K	I	A	K	D	N	H	I	K	T	T	P	T	A	F	I	N	G	E	K	V	
80% consensus			c	h	h	t	s	h	p	I	p	G	V	P	s	h	h	V	N	G	K	Y

**Figure 1.25:** Sequence conservation of DsbA homologues. Truncated sequence alignments of *EcDsbA* and 10 DsbA homologues showing residue and characteristic conservation as the consensus of 80 % of the populations. Residue number is based on the sequence of *EcDsbA*. Each homologue is labelled with the UniProt ID used for analysis and the sequence identity compared to *EcDsbA*. Catalytic cysteines highlighted in yellow. Structural features are denoted as the single letter amino acid codes, “o” for alcohols, “l” for aliphatic, “a” for aromatic, “c” for charged, “h” for hydrophobic, “-” for anion, “+” for cation, “p” for polar, “s” for small, “u” for tiny, and “t” for turnlike as allocated by MView <sup>(248)</sup>.

#### 1.4.4 Dynamics of *EcDsbA*

Previous work has examined the dynamics of *EcDsbA* and close homologues<sup>(231, 233, 249, 250)</sup>. NMR relaxation dispersion experiments have demonstrated that the DsbA proteins from *B. pseudomallei*<sup>(250)</sup>, *V. cholerae*<sup>(249)</sup> and *E. coli*<sup>(233)</sup> undergo concerted dynamics motions that are dependent on the redox state of the protein. Movement is evident immediately prior to and within the active site helix for all of these variants of DsbA. These dynamic residues have also been identified as the binding sites for small molecules that interact with either *BpsDsbA* (adjacent to the active site) or *EcDsbA* (internal cavity behind the active site). The pockets into which the small molecules bind in these two proteins are not evident in the structure of either apo-protein – and are therefore referred to as “cryptic” pockets. These cryptic pockets can either represent a low population state that is present in the protein in the absence of the ligand, or they can form in the presence of the ligand. In either case a conformational change is required for the cryptic pocket to form, and the residues that form the cryptic pockets corresponded well with residues which had displayed dynamics in the NMR data. The cryptic pocket in *EcDsbA* is an internal pocket that is shielded from solvent, which means that the residues which line this pocket must undergo significant conformational change to allow access for small molecule binding. Additional evidence for the presence of this internal cryptic pocket in *EcDsbA* came from a study aimed at the development of *EcDsbB* inhibitors. This had led to the discovery of a quinone derivative which covalently bound cysteines in *EcDsbB* that was subsequently shown to also bind to *EcDsbA*<sup>(251)</sup>. This compound did not react with the solvent exposed Cys30 as expected, however, nanoESI experiments showed that the compound had labelled Cys33. As Cys33 is buried and inaccessible in the crystal structures of *EcDsbA*, these data suggest that the active site is undergoing conformational changes to expose and facilitate alkylation of this residue.



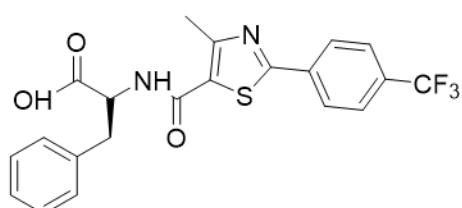
**Figure 1.26:** Dynamics of DsbA. A) Overlay of *EcDsbA* residues which showed dynamic characteristics by  $^1\text{H}^{\text{N}}$  CPMG-RD (blue) and fragment bound within the internal cryptic pocket (orange sticks) B) Cartoon structure of *NmDsbA3* (PDB ID: 2ZNM) with a partially unwound active site helix (yellow).

Both the labelling study and the NMR dynamics experiments are consistent with an unwinding of the active site helix of *EcDsbA*. This would allow binding of small molecules at the cryptic pocket and covalent labelling of Cys33<sup>(233)</sup>. Evidence for this minor population of *EcDsbA* was characterised by <sup>1</sup>H<sup>N</sup> Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG-RD) and hydrogen-deuterium (H/D) exchange experiments by NMR. The relaxation-dispersion NMR data for the residues in the active site helix were consistent with it adopting a disordered secondary structure in the minor state. Furthermore, active site residues which appear to be protected from solvent in the crystal structure of *EcDsbA* underwent H/D exchange when lyophilised protein was exposed to deuterated buffer. These experiments suggest that a minor population of *EcDsbA* exists where the active site helix is partially unwound. This is further supported by a crystal structure obtained for a DsbA homologue from *Neisseria meningitidis* (*NmDsbA3*)<sup>(231)</sup>. A crystal structure of *NmDsbA3* captured a protein conformation where the *N*-terminal end of its active site helix was partially unfolded (Figure 1.26). Taken together, these studies suggest that active site dynamics play an, as yet, undefined role in binding and reactivity of *EcDsbA* with substrates and small molecules.

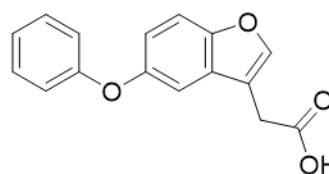
#### 1.4.5 Small molecule inhibitors of *EcDsbA*

Small molecule inhibitors of *EcDsbA* have the potential to be useful as both potential therapeutics and tool compounds to characterise the detailed mechanism of action. As the substrate binding site is rather shallow, fragment-based strategies have been implemented to aid in their discovery. A fragment screen of 1132 compounds against *EcDsbA* was conducted by saturation transfer difference NMR (STD-NMR)<sup>(252)</sup>. This screen identified and validated 37 fragment hits of 8 distinct classes and 11 singletons. Many of these compound classes have undergone further development with varying success (Figure 1.27)<sup>(128, 233, 252-256)</sup>. All crystal structures obtained for these series identified fragments bound within the hydrophobic groove. Unfortunately, some classes, such as thiophenes<sup>(254)</sup> and benzofurans<sup>(256)</sup>, did not maintain binding poses throughout optimisation, making SAR difficult to interpret. Nonetheless, a common feature observed for these fragments is the hydrophobic contacts which drive their affinity. The most important interaction made by all of these compounds is a  $\pi$ - $\pi$

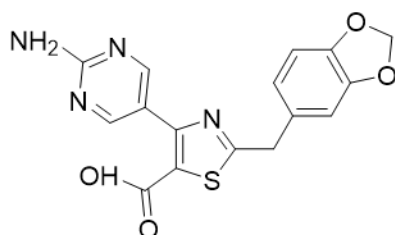
interaction with His32 of the active site. This interaction acts as an anchor point for binding and each series displayed a significant reduction in affinity when it was lost. Conversely, obtaining hydrogen bonds to the carbonyl oxygen of Pro151 and the amide sidechain of Gln164 generally failed to provide any significant improvement in affinity <sup>(233)</sup>. This illustrates that although interactions can be gained, some interactions are more favourable than others. The ability to determine which interactions will boost affinity would greatly improve the ability to design higher affinity compounds.



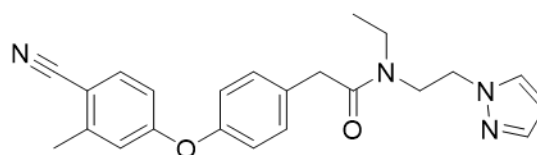
SPR  $K_D = 196 \pm 35 \mu\text{M}$



NMR  $K_D = 326 \pm 25 \mu\text{M}$



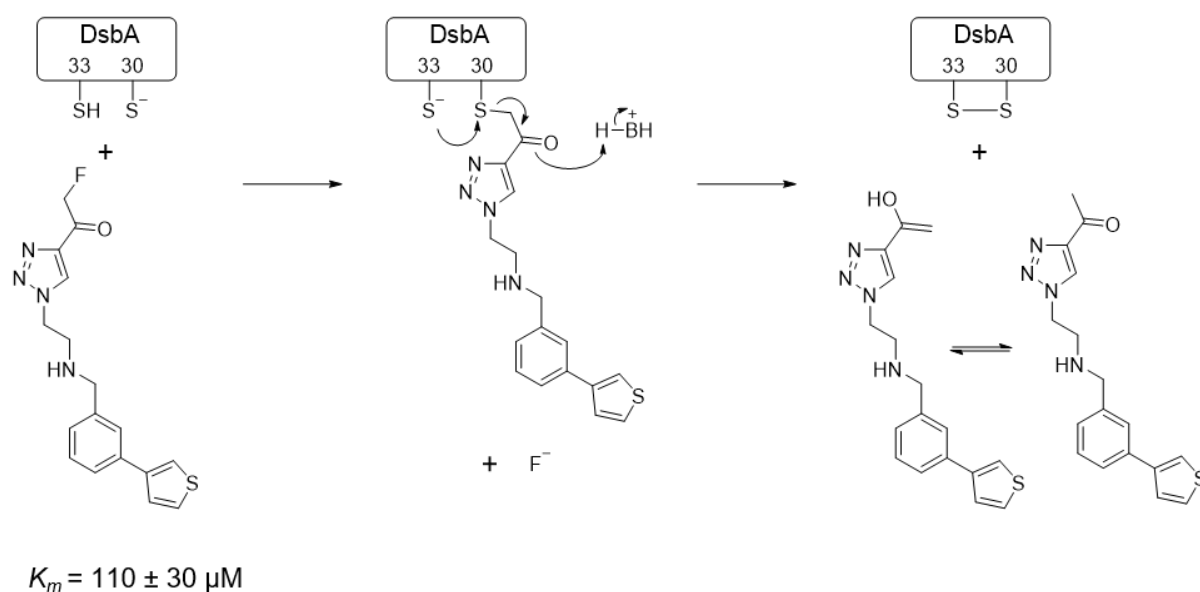
NMR  $K_D = 448 \pm 48 \mu\text{M}$



NMR  $K_D = 63 \pm 9 \mu\text{M}$

**Figure 1.27:** Most advanced small molecule inhibitors of *EcDsbA*.

In an attempt to navigate inconsistent binding poses and unproductive non-covalent interactions a series of covalent fragments were designed and tested <sup>(254, 255)</sup>. A weak fluoromethylketone (FMK) warhead was attached to a thiophene scaffold through click chemistry. These analogues improved the parent compound from an estimated  $K_D$  of  $\sim 6.6 \pm 3 \text{ mM}$  by  $^1\text{H}$ - $^{15}\text{N}$  HSQC to an elaborated aryl FMK with a  $K_m$  of  $110 \pm 30 \mu\text{M}$ . Furthermore, the importance of the thiophene ligand in placement of the FMK warhead adjacent to the active site of *EcDsbA* was demonstrated as significant activity was lost when the FMK was attached to a simple phenyl ring.



**Figure 1.28:** Proposed mechanism for labelling of *EcDsbA* at Cys30 and subsequent turnover of aryl fluoromethylketone fragment by neighbouring residues. (254, 255)

While this validated that attaching a covalent warhead to a weakly binding fragment could greatly improve compound activity, the resultant protein complex was found to be unstable. After the FMK alkylation of the active site Cys30 it was proposed that a nearby weakly acidic residue (possibly His32) facilitated the nucleophilic attack of Cys33 on the protein-fragment conjugate, releasing the desfluoro thiophene (Figure 1.28). There was also evidence of a slow build-up of a protein species which was alkylated at both Cys30 and Cys33, providing further evidence that dynamics in residues around the active site contributes to substrate binding.

Recently a fragment screen using another fragment library of 1118 compounds was conducted against *EcDsbA* in the same manner as the previous screen (233). This screen identified 26 validated fragment hits, two of which contained an isoxazole scaffold. These isoxazoles caused a unique chemical shift perturbation profile by <sup>1</sup>H-<sup>15</sup>N HSQC. Rather than perturbing residues within the hydrophobic groove, peak broadening and perturbations were observed for residues adjacent to the active site cysteines. An X-ray crystal structure discovered that these fragments bound inside the protein, behind the active site, within the cryptic pocket (Figure 1.26). Even more interestingly, binding at this previously unidentified pocket inhibited the ability of

*EcDsbA* to introduce disulfide bonds into substrates. Since the isoxazole fragments do not block the hydrophobic groove or prevent direct access to the active site Cys30, this suggests that the cryptic pocket – and potentially the dynamic processes that allow it to form – plays a functional role in DsbA activity.

Attempts to develop the fragments that bind within the hydrophobic groove and the cryptic pocket have not yet generated the compounds with submicromolar affinities that are expected to be required to achieve functional inhibition of DsbA in bacteria. The efforts to develop DsbA inhibitors have employed structure-based design, as well as parallel synthesis and screening of unpurified reaction mixtures<sup>(128, 233, 253)</sup>. This has enabled the development of a suite of robust biophysical assays to characterise binding to *EcDsbA*. This provided a platform for this research project, which aimed to explore some of the approaches described above, how they might be employed in the development of *EcDsbA* inhibitors, and how they might be implemented more generally to advance programs of FBDD against difficult protein targets.

## **1.5 Project hypotheses and aims**

This thesis explores the hypothesis that the implementation of orthogonal fragment-based design techniques will expedite fragment development against difficult targets. In order to test this theory, multiple techniques were applied to the optimisation of inhibitors of *EcDsbA*.

Previous characterisation of peptide and small molecule binding has implied that there are some protein interactions which are privileged and some that are unproductive for increasing substrate affinity. In my research I explored a range of different methods in an effort to characterise “hot spots” on *EcDsbA* that may be targeted to improve compound binding affinity. Previous studies have also demonstrated that fragment activity can be vastly improved through the addition of a covalently binding functional group. The previous work resulted in the reaction of an FMK covalent warhead with *EcDsbA*. This generated a protein-ligand complex that was ultimately turned over, however, this suggests that identification of a warhead which forms a stable covalent

species may lead to a potent inhibitor of *EcDsbA*. Lastly, a novel cryptic binding pocket was discovered during a fragment screen. Binding at this site requires dynamic protein movement and this pocket is not observable in a static unbound crystal structure. This suggests that there are still aspects of ligand binding which are not understood and that new binding sites could be identified by probing *EcDsbA* with diverse libraries and structural studies.

This body of work aims to:

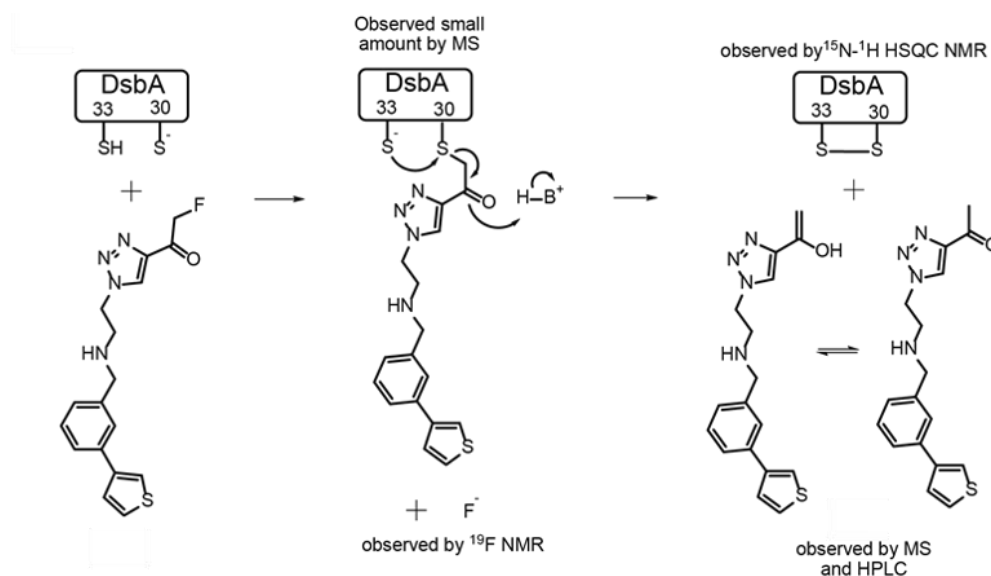
- Interrogate the interaction of *EcDsbA* with a peptide substrate and apply this knowledge to the design of covalent small molecule inhibitors.
- Develop methods to screen crude products by  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR for the identification of selective and stable covalent inhibitors.
- Develop methods to identify protein binding hot spots, their conserved interactions and privileged pharmacophores for use in future fragment development.



# Chapter 2:

## Characterisation of a mono-cysteine peptide substrate

Initial attempts within the group to inhibit *EcDsbA* with covalent substrates included the optimisation of a fluoromethylketone (FMK) inhibitor which primarily binds within the hydrophobic groove <sup>(254, 255)</sup>. This FMK compound successfully labelled the catalytic thiolate Cys30 of the reduced protein and recognition of the fragment binding motif was vital. This was shown by replacement of the fragment extension with a benzyl substituent which significantly reduced the FMK reactivity with reduced *EcDsbA*. However, the covalent complex was unexpectedly turned over by *EcDsbA* to release the desfluoro derivative of the ligand and oxidised *EcDsbA* (proposed mechanism is given in Figure 2.1). Additional testing of maleimide, which formed a similar intermolecular  $\beta$ -keto thioether, also yielded oxidised protein. Furthermore, initial  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR testing of other common thiol labelling motifs, such as acrylamides, demonstrated little to no reactivity with *EcDsbA*. The mechanism by which *EcDsbA* oxidises native peptide substrates was further investigated to understand the reactivity of Cys30 thiol in reduced *EcDsbA*.

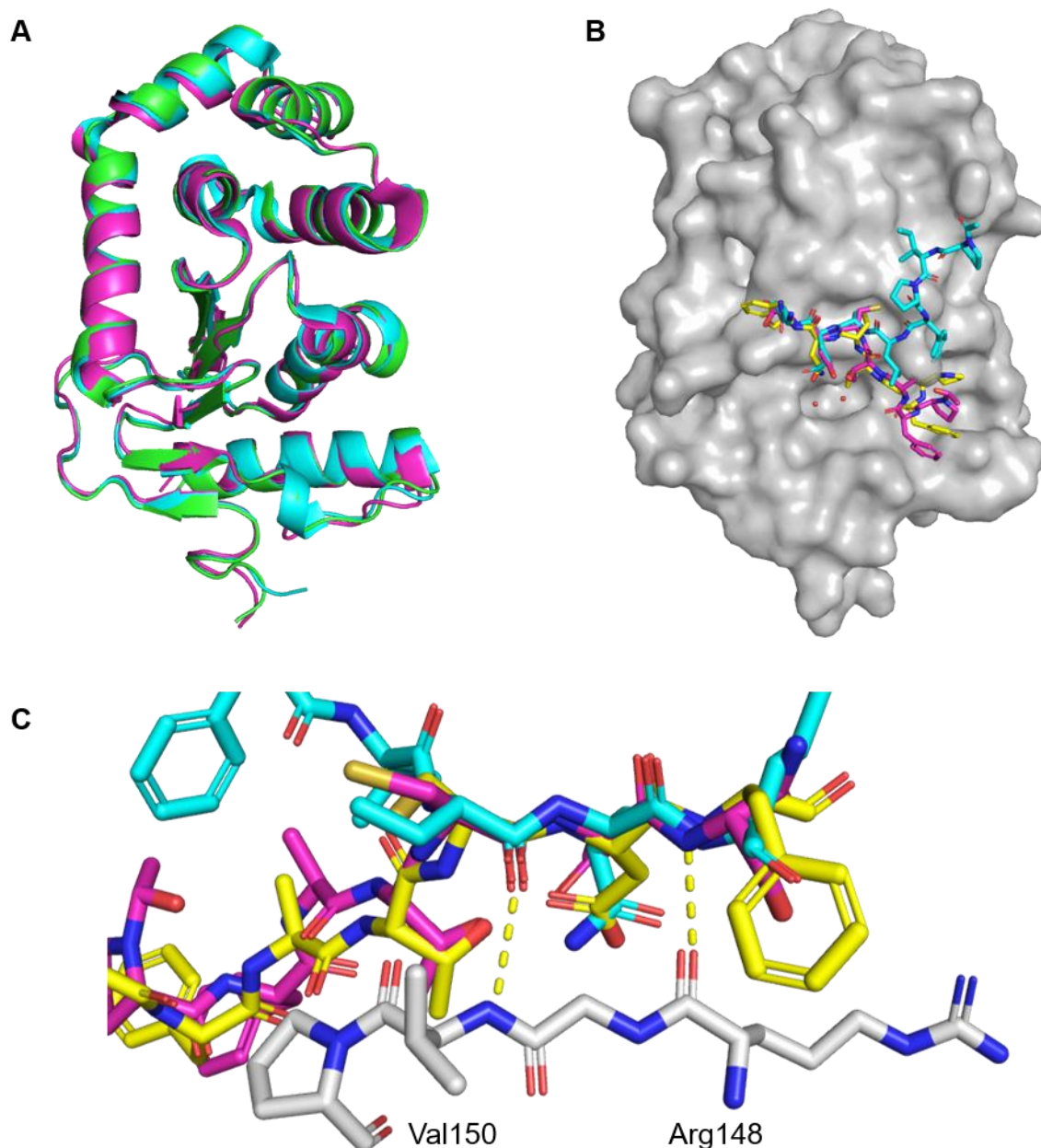


**Figure 2.1:** Proposed mechanism for labelling of *EcDsbA* at Cys30 and subsequent turnover of aryl fluoromethylketone fragment by neighbouring residues. <sup>(254, 255)</sup>

*EcDsbA* has a high redox potential of -122 mV <sup>(257)</sup> and the active site Cys30 has an unusually low pK<sub>a</sub> value of ~ 3.5 <sup>(258)</sup>. This results in the predominance of an active site thiolate for reduced *EcDsbA* at physiological pH. Furthermore, the reduced form of the protein is stabilised due to the electrostatic interaction between the negatively charged cysteine and neighbouring His32 <sup>(236)</sup>. This stability results in the preference of the reduced state over the oxidised and, therefore, drives the reaction of the enzyme and substrates forward <sup>(234, 236)</sup>. Historically, the oxidation of substrates occurs too rapidly for the disulfide intermediate between *EcDsbA* and the unfolded peptides to be isolated. In order to interrogate different rates, stabilities and structures of inter- and intramolecular reactions and complexes it was necessary to employ the use of cysteine mutants of either substrate peptides or *EcDsbA* <sup>(222, 229, 240-242, 244-246)</sup>. DsbA's preferential reaction with peptides over glutathione is attributed to non-covalent interactions formed in addition to the disulfide bond <sup>(246)</sup>. These interactions are important for substrate recognition but are not typically very strong. Affinities of these non-covalent interactions were too weak to measure <sup>(242, 246)</sup>, however, these complexes do not need to be high affinity to catalyse peptide substrate oxidation. Furthermore, weaker interactions which are only mildly stabilising can facilitate the dissociation of folded peptide substrates and continuation of the catalytic cycle.

A crystal structure was obtained for the interaction of *EcDsbA* and *EcDsbB* through mutations of the cysteines involved in the resolution pathways by nucleophilic attack (*EcDsbA* – Cys33Ala, *EcDsbB* – Cys130Ser, Figure 2.2) <sup>(222, 240, 241)</sup>. This structure highlighted a seven-residue sequence from *EcDsbB* which occupied the main substrate binding site of *EcDsbA*, <sup>100</sup>PFATCDF<sup>106</sup>. This structure showed a mixed disulfide bond between Cys30 and Cys104 of DsbA and DsbB respectively (Figure 2.2). A tripeptide sequence of *EcDsbB* (Cys104 – Phe106) forms an antiparallel  $\beta$  sheet with residues in the loop preceding the *cis*-Pro residue (Arg148 – Val150). The hydrogen bonds to the backbone amides of these residues represented the only polar interactions between the two proteins. However, the sidechains of DsbB form many hydrophobic interactions to residues within the hydrophobic groove of DsbA (such as Leu40 and Phe174).

As this segment of *EcDsbB*'s periplasmic loop is drawn into the main substrate binding site of *EcDsbA* attempts have been made to optimise the DsbB sequence to into peptide inhibitors <sup>(242)</sup>. Mono-cysteine peptide analogues were characterised through ITC and thermal shift assays which led to the identification of Ac-PWATCDS-NH<sub>2</sub> (**23**) which had a  $K_D$  of  $5.7 \pm 0.4 \mu\text{M}$  against *EcDsbA*. A structure of a close analogue of this peptide was obtained in which it adopted a highly similar pose as the *EcDsbB* loop (Figure 2.2b). There were no peaks consistent with reduced protein observed by thermal shift and data obtained by ITC supported a reversible 1:1 binding model. Therefore, it was proposed that a second peptide molecule did not attack the mixed disulfide bond to release oxidised dimerised peptide and reduced DsbA. Instead it was suggested that the mixed disulfide was resolved by the nucleophilic attack of *EcDsbA* Cys33 on the mixed disulfide bond to release the original states of reduced peptide and oxidised protein. Although this mechanism does not reflect the enzymatic pathway.



**Figure 2.2:** Crystal structures of *EcDsbA* in complex with peptide substrates. A) Overlay of secondary structures for *EcDsbA* in its unbound oxidised form (green cartoon, PDB ID: 1FVK<sup>(236)</sup>), in complex with *EcDsbB* (magenta cartoon, PDB ID: 2ZUP<sup>(241)</sup>) and in complex with a SigA derivative (cyan cartoon, PDB ID: 3DKS<sup>(229)</sup>). B) Overlay of binding poses of covalently labelled *EcDsbA* (grey surface, PDB ID: 1FVK) for a SigA derivative (cyan sticks, PDB ID: 3DKS), *EcDsbB* periplasmic loop (<sup>100</sup>Pro – Phe<sup>106</sup>, magenta sticks, PDB ID: 2ZUP) and *EcDsbB* derivative PWATCDF (yellow sticks, PDB ID: 4TKY<sup>(242)</sup>) with structural waters (red spheres). C) Conserved polar interactions (yellow dashes) between peptide substrates (coloured sticks) and *cis*-Pro loop of *EcDsbA* (grey sticks).

A nine-residue SigA transporter peptide (a substrate of *EcDsbA*) with the cystine replaced with a homobromoalanine has previously been synthesised and reacted with *EcDsbA* <sup>(229)</sup>. This allowed a crystal structure to be obtained as a model of the catalytic intermediate, by preventing substrate turnover through the replacement of the disulfide bond with an  $\alpha$ -thioether bond. The superposition of this model intermediate with previously recorded structures of the unbound reduced and oxidised protein indicated that no significant global conformational changes occur upon substrate interaction (Figure 2.2a). This peptide mimics DsbB's interactions with Arg148 – Val150 by adopting a similar antiparallel  $\beta$  sheet formation with the *cis*-Pro loop. The only additional interactions made by the peptide are hydrophobic interactions with DsbA sidechains. Interestingly, this crystal structure does not show the SigA substrate occupying the hydrophobic groove (Figure 2.2b). The sidechain of His32 orientates itself downward toward the groove and allows the peptide to interact with residues along the thioredoxin –  $\alpha$  helical domain interface. However, the interaction of *EcDsbA* and a non-covalent homoserine SigA analogue was characterised by <sup>1</sup>H-<sup>15</sup>N HSQC NMR. As the only difference between the peptide substrates is the formation of the thioether linkage with Cys30 it was predicted that the binding orientation and interactions would be consistent with the crystal structure. Mapping of the chemical shift perturbations caused by the presence of the homoserine peptide suggested that the peptide may be binding along the domain interface and also within the hydrophobic groove. Furthermore, it should be noted that the hydrophobic groove is partially occluded in the crystal structure. Together these observations suggest that it is possible for the peptide to present multiple binding poses and could potentially relocate once the intermolecular bond is formed with *EcDsbA*.

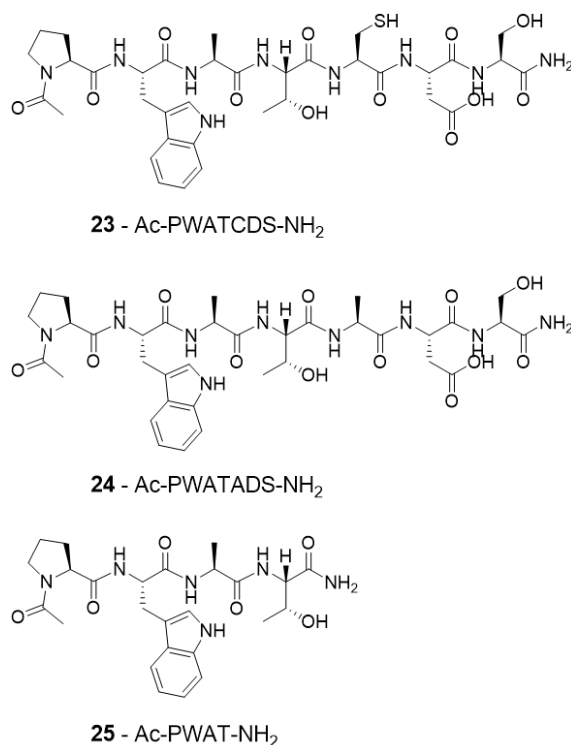
The ambiguity of these studies highlights how much is still to be understood about the unusual and unique chemistry of *EcDsbA*. Additionally, it cannot be overlooked that the structural and mechanistic scrutiny of the system is conducted with either mutated protein or mutated substrates. Non-covalent peptides bound in multiple and unexpected poses raising questions about the importance of the type of interactions made and the residues which make them. In this study the native <sup>30</sup>CPHC<sup>33</sup> *EcDsbA* will be studied in the presence of a mono-cysteine heptapeptide by solution NMR in

order to obtain structural and kinetic insights with the native protein and to evaluate thiols as potential warheads for covalent inhibitor design.

## 2.1 Structural characterisation of peptide binding

### 2.1.1 Effects of non-covalent peptides against *EcDsbA*

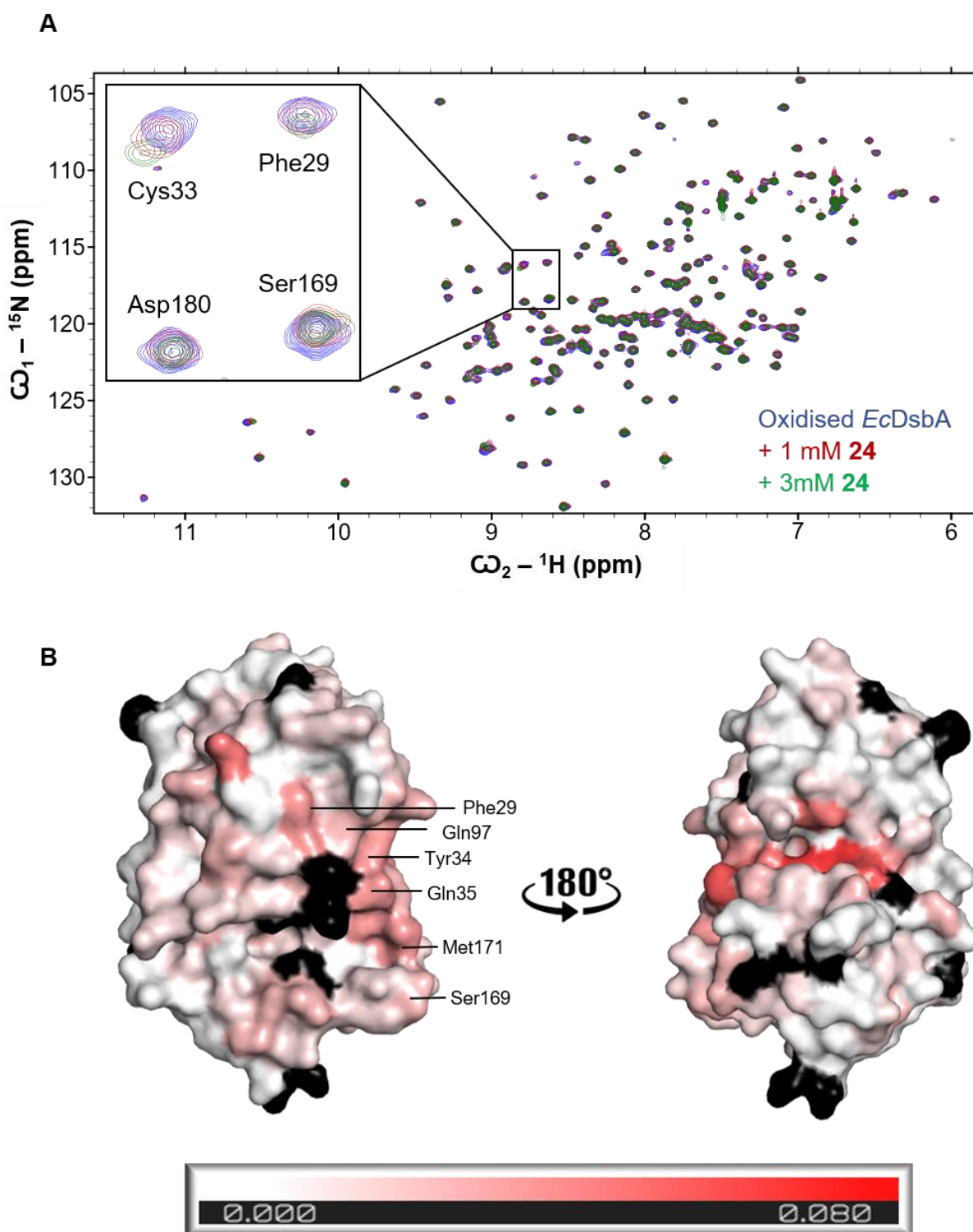
The interactions made by the Ac-<sup>1</sup>PWATCDS<sup>7</sup>-NH<sub>2</sub> peptide (**23**) are dominated by the formation of the intermolecular disulfide bond formed between C5 and Cys30<sup>(242)</sup>. The residues which flank the cysteine are either exposed to the main substrate groove or form an antiparallel  $\beta$  sheet with Arg148 and Val150 from *EcDsbA*. In order to interrogate the overall value of the non-covalent interactions and determine whether binding in the hydrophobic groove or along the *cis*-Pro loop is more favourable, an alanine replacement peptide Ac-PWATADS-NH<sub>2</sub> (**24**) and a truncated Ac-PWAT-NH<sub>2</sub> peptide (**25**) were tested for binding to U-<sup>15</sup>N labelled *EcDsbA* by <sup>1</sup>H-<sup>15</sup>N HSQC (Figure 2.3).



**Figure 2.3:** *EcDsbB* peptide analogues used to characterise substrate oxidation by *EcDsbA*.

The chemical shift perturbations (CSPs) of *EcDsbA* in the presence of the non-covalent peptides were mapped onto the structure of the protein. The truncated peptide **25** showed movement in Arg148 consistent with the antiparallel  $\beta$  sheet formed in previous studies, however no other residues in the vicinity of the *cis-Pro* loop gave rise to CSPs. Perturbations of Gln35, Ser169, and Met171 more strongly supported binding within the hydrophobic groove, as predicted by the crystallographic binding pose. Similarly, binding of C5A peptide **24** was suggested to occur primarily within the main substrate binding site (Figure 2.4). Residues along the  $\alpha$ 1 helix which contains the catalytic motif and forms the top of the groove are affected by the presence of the peptide, however, as with **25**, no residues preceding or within the *cis-Pro* loop show significant shifts. Phe29, which is implicated in binding of the SigA peptide, <sup>(229)</sup> along with adjacent residues of Tyr34 and Gln97 showed movement in the HSQC spectrum. These perturbations in both the hydrophobic groove and domain interface may be explained by the transient and non-selective binding of peptides to DsbA. Reports using different peptide sequences have provided evidence to suggest that peptides may bind in multiple poses along the surface of *EcDsbA*, however this has not been definitively proved <sup>(229, 244)</sup>.



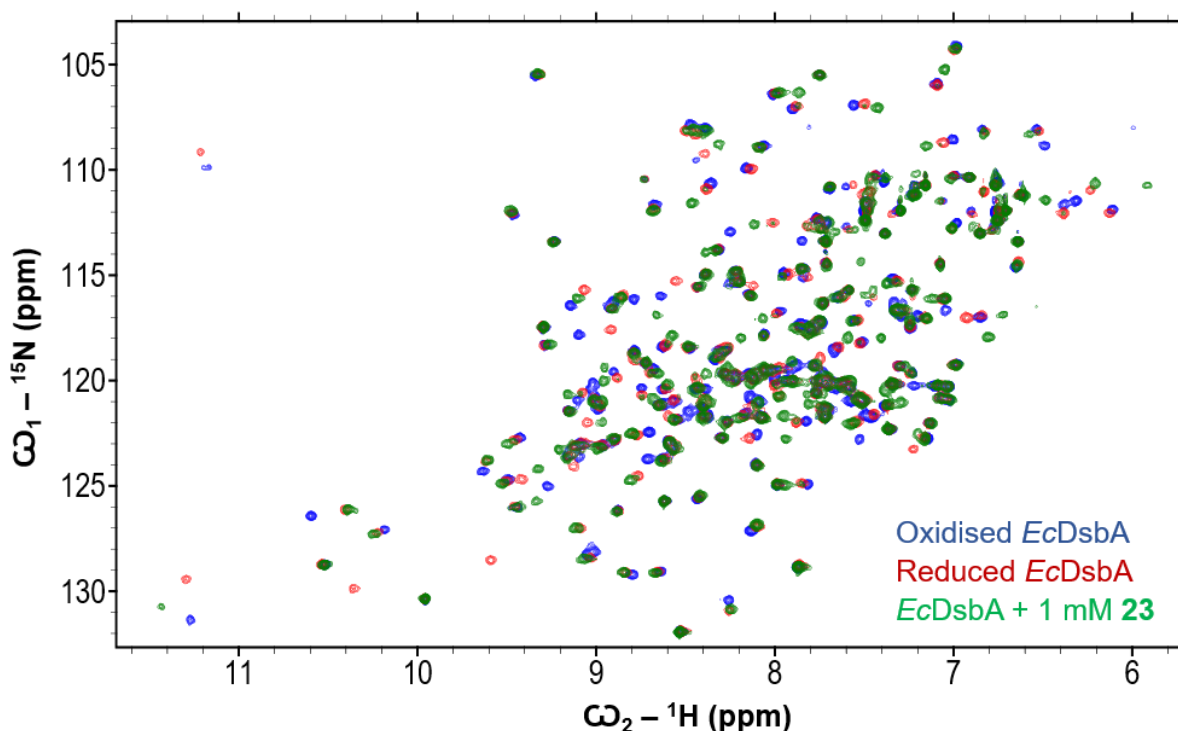


**Figure 2.4:** Characterisation of non-covalent interactions of **24** by  ${}^1\text{H}$ - ${}^{15}\text{N}$  HSQC. A)  ${}^1\text{H}$ - ${}^{15}\text{N}$  spectra of oxidised  ${}^{15}\text{N}$  *EcDsbA* (blue), incubated in the presence of 1 mM **24** (red) and 3 mM **24** (green). B)  ${}^1\text{H}$ - ${}^{15}\text{N}$  HSQC chemical shift perturbations caused by 3 mM **24** onto the crystal structure of *EcDsbA* (PDB ID: 1FVK) and coloured from white to red by the extent of CSP: minimum = 0.0 ppm (white) and maximum = 0.08 ppm (red), and unassigned (black).

The peptides were assessed for aqueous solubility via semi-quantitative 1D  $^1\text{H}$  NMR serial dilutions in  $\text{D}_2\text{O}$  buffer <sup>(259)</sup>. The peptide chemical shifts remained consistent and relative concentrations were calculated by the comparison of the peak integration to an internal standard (sodium trimethylsilylpropanesulfonate, DSS). This was used to determine the maximum assayable concentration with respect to the maximal solubility and no evidence of aggregation.  $K_D$  values could not be assigned to **24** and **25** as the perturbations which occur at the highest peptide solubility are too small for reliable calculations. However, based on the magnitude of historical fragment CSPs the affinity was estimated to be in the low millimolar range ( $K_D > 2 \text{ mM}$ ). This observation is also consistent with other studies being unable to quantify binding affinities for the non-covalent peptide analogues <sup>(242)</sup>.

### 2.1.2 Effects of a covalent peptide against *EcDsbA*

In the presence of Ac-PWATCDS-NH<sub>2</sub> (**23**) the  $^1\text{H}$ - $^{15}\text{N}$  HSQC resonances of oxidised wildtype *EcDsbA* backbone amides changed dramatically, suggesting the formation of an intermolecular disulfide species not previously recorded (Figure 2.5). No cross-peaks were consistent with either oxidised or reduced Cys30 or Cys33, implying that the new species was a complex corresponding to the disulfide intermediate of substrate peptide oxidation with DsbA. The magnitude of the changes observed in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of *EcDsbA* did not allow for nearest neighbour approximations and required the backbone resonances of this complex to be reassigned. However, the acquisition of these experiments provided an opportunity to further interrogate the structural implications of substrate binding.

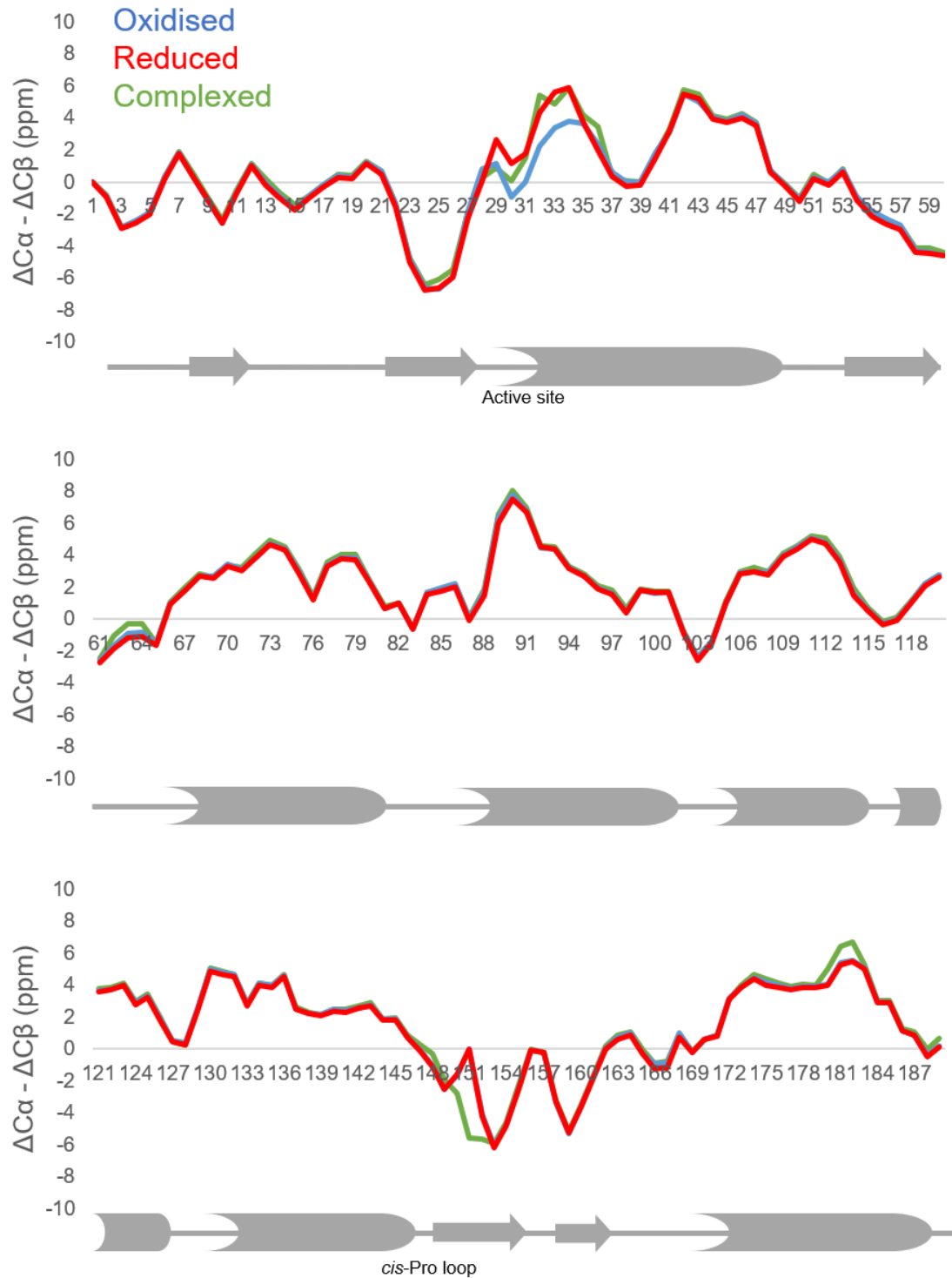


**Figure 2.5:** Comparison of redox states by  $^1\text{H}$ - $^{15}\text{N}$  HSQC. HSQC overlays of unbound oxidised (blue) and reduced (red) *EcDsbA* (100  $\mu\text{M}$ ) and in the presence of 1mM **23** (green).

$\text{U-}^{15}\text{N}$ - $^{13}\text{C}$  labelled *EcDsbA* was complexed with covalent peptide **23** for the acquisition of double and triple resonance experiments used to assign amide backbone,  $\text{C}\alpha$  and  $\text{C}\beta$  resonances (experiments conducted by Dr Biswaranjan Mohanty, see chapter 8 for experimental detail). Backbone amide resonance assignments were determined for 99 % of non-proline residues including His32, Ser106 and Phe129 which are not observed in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC of oxidised *EcDsbA*. Analysis of the H, N,  $\text{C}\alpha$  and  $\text{C}\beta$  chemical shifts showed large changes in the two active site cysteines where Cys30 completely broadened out in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC and the  $\text{C}\beta$  shift for Cys33 was now consistent with the residue being in a reduced-like state. Additionally, His32, which is not observed in oxidised *EcDsbA*  $^1\text{H}$ - $^{15}\text{N}$  HSQC, can now be seen and is in a position which is inconsistent with the peak from the reduced spectrum. Random coil analysis on the  $\text{C}\alpha$  and  $\text{C}\beta$  environments was conducted against all three states; oxidised, reduced and the intermediate complex with the peptide, to determine any changes in secondary structure without the use of nuclear Overhauser effects (NOE)

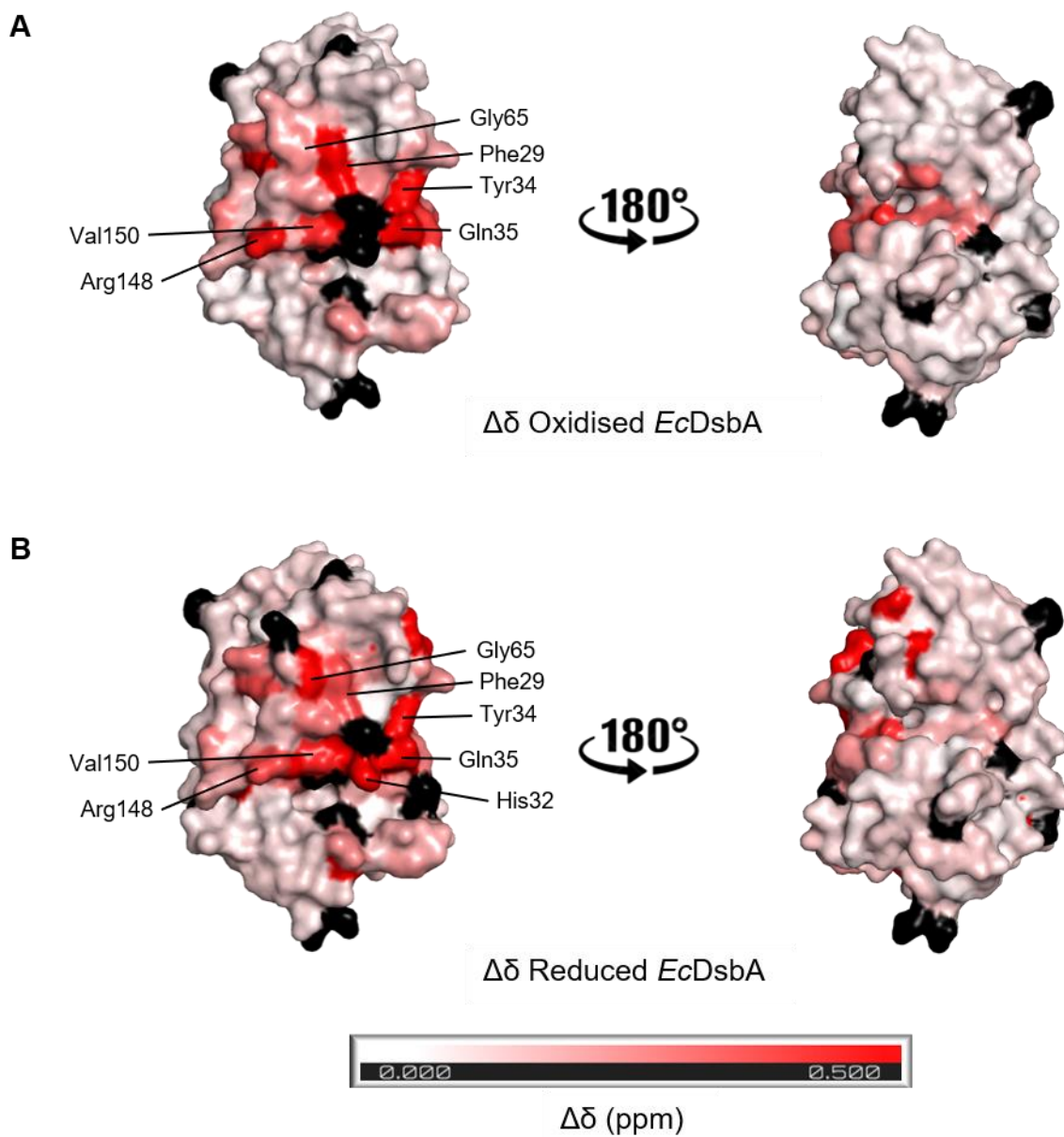
(Figure 2.6) <sup>(260, 261)</sup>. Random coil analysis involves the comparison of known chemical shift ranges for residues in unstructured peptides and the observed chemical shifts for the protein of interest. Secondary structure motifs can be identified by the local densities of positive or negative deviations from the random coil shift values, with a string of positive values indicating the presence of an  $\alpha$  helix and a string of negative values indicating a  $\beta$  sheet. While the global folds between the states were mostly unchanged the regions of biggest discrepancy were at the active site and the *cis*-Pro loop. The structural order of Cys30 and Pro31 in the complex appears to be between those of oxidised and reduced while the adjacent residues along the  $\alpha$ 1 helix (His32 – Gln35) appear almost identical to the state of the reduced protein. The *cis*-Pro loop from Val150 – Ala152 which are the same for oxidised and reduced now shows less flexibility and takes on the properties of a well ordered  $\beta$  sheet. This may be due to the known hydrogen bonding interactions with the backbones of substrates in this area.

## Random coil shift analysis of *EcDsbA* by redox state



**Figure 2.6:** Secondary structure determination by random coil shift analysis. Overlays of  $\Delta C\alpha - \Delta C\beta$  (ppm) for oxidised (blue), reduced (red) and complexed (green) *EcDsbA* against residue number. Estimated positions of  $\alpha$  helices (grey cylinder),  $\beta$  sheets (grey arrow) and unstructured loops (grey line) shown under graph.

Mapping of the difference in the weighted average of  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts from the mixed state against the oxidised and reduced proteins showed affected areas similar to those described above (Figure 2.7). Despite the  $\text{C}\alpha$  and  $\text{C}\beta$  environments for the redox states showing minimal changes, the perturbations between the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra are significant. The CSP profile is similar to that observed for the non-covalent peptides **24** and **25**, where the main effects are along the  $\alpha 1$  helix and Phe29 and Tyr34 amongst the most affected. Residues Phe63 – Gly66 form a type IV  $\beta$  turn to connect the thioredoxin and inserted  $\alpha$  helical domains and this turn is believed to act as a hinge in domain motion <sup>(243)</sup>. These residues show large perturbations in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra for the disulfide bond complex and are visible only in the formation of the intermolecular bond and not detected in the binding of non-covalent peptides **24** and **25**. Although there are some perturbations observed within the hydrophobic groove they are dwarfed by those along the domain interface. This suggests that binding of this peptide is more consistent with the pose for the SigA derived peptide <sup>(229)</sup> than the crystal structure obtained with the DsbB peptide analogue <sup>(242)</sup>. Furthermore, the areas of greatest change coincide with the protein areas where dynamics were observed by relaxation-dispersion NMR studies <sup>(233)</sup>.



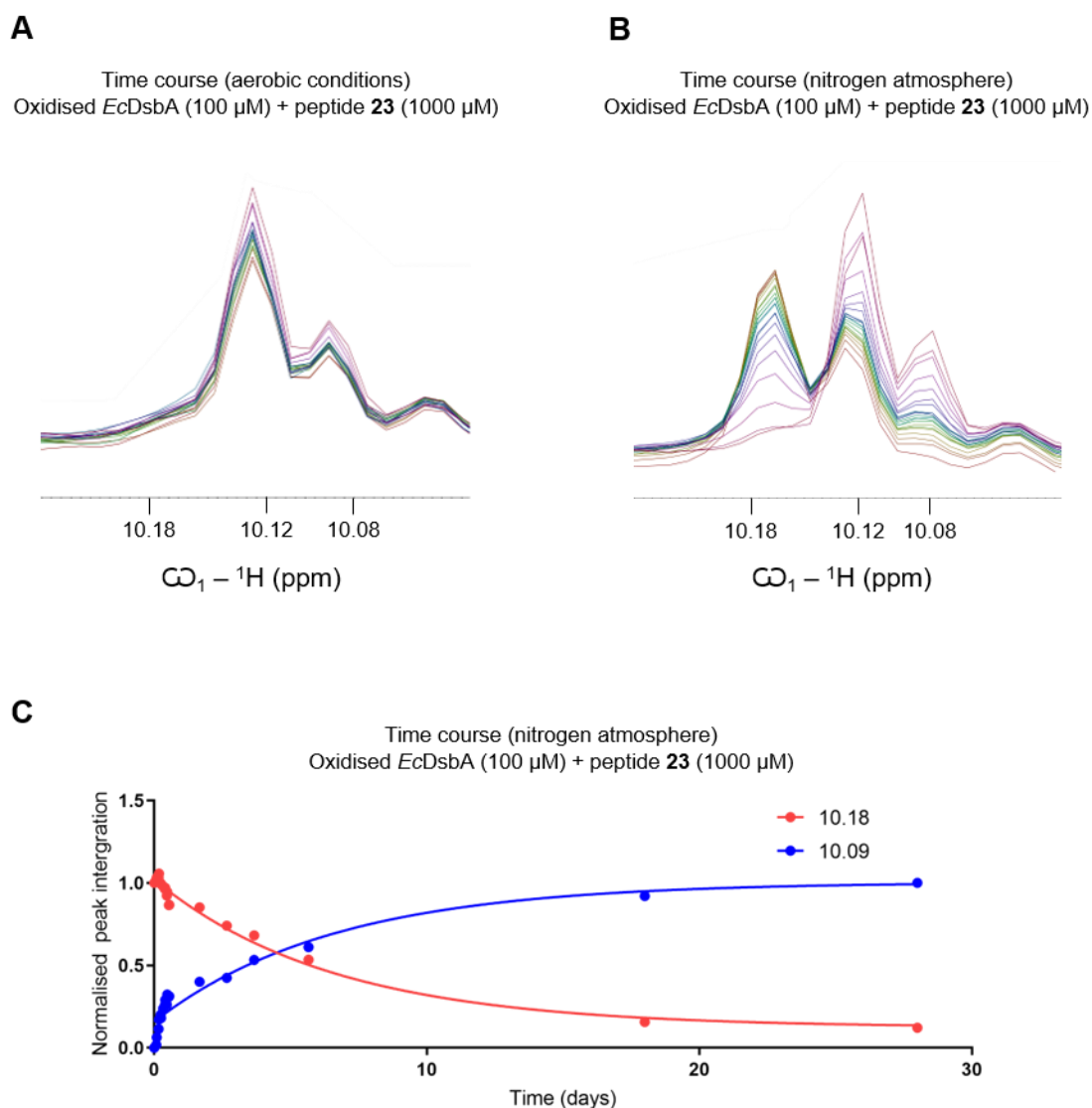
**Figure 2.7:**  $^1\text{H}$ - $^{15}\text{N}$  HSQC chemical shift perturbations caused by **23** labelling. CSP against (A) oxidised and (B) reduced *EcDsbA* mapped onto the crystal structure of *EcDsbA* (PDB ID: 1FVK) and coloured from white to red by the extent of CSP: minimum = 0.0 ppm (white) and maximum = 0.5 ppm (red), and unassigned (black).

## 2.2 Characterisation of reaction kinetics for peptide oxidation

As the peptide contained a cysteine residue it was expected that oxidised *EcDsbA* would facilitate the dimerisation of peptide **23**. However, it was also found that the peptide was susceptible to oxidation by the aqueous buffer conditions in the absence of *EcDsbA*. Therefore, the rate of oxidation in both the presence and absence of *EcDsbA* was monitored and compared. To investigate the rate at which *EcDsbA* catalysed the formation of a disulfide bond between two molecules of mono-cysteine peptide **23** the populations of reduced and oxidised dimerised peptide were followed by 1D  $^1\text{H}$  NMR. Proton signals corresponding to the Trp2 residue of the peptide were isolated from all others and did not overlap when in the monomer or dimer state. Relative peak height and integration of the monomer/dimer signals were used to monitor reaction completion (Figure 2.8). The conversion of **23** to its corresponding dimer by aerial oxidation in aqueous conditions occurred so rapidly (< 10 mins) that reduced peptide was never observed. In order to slow the peptide oxidation, the sample was placed under an atmosphere of nitrogen (in addition to the degassed buffer) which caused the extension of the reduced peptide's half-life to 3.5 days.

The addition of **23** to oxidised *EcDsbA* saw a mixed disulfide intermediate form rapidly (< 30 mins) under both an atmosphere of air (aerobic conditions) or nitrogen. This implies that oxidised DsbA can rapidly react with reduced mono-cysteine peptide **23** to form the complex prior to background peptide dimerisation. No proton signals consistent with the reduced form of peptide **23** were ever observed in the presence of oxidised *EcDsbA* in aerobic conditions and therefore a peptide half-life could not be calculated (Figure 2.8). Nevertheless, the peptide and oxidised *EcDsbA* formed the protein-peptide complex which was stable for 21 – 28 days. Purging the sample's air with nitrogen gas reduced aerial oxidation and prolonged the life of the intermediary complex, which now resolved in 33 – 56 days, with a peptide half-life of 4.3 days. However, the 95% confidence intervals of the half-lives in the presence and absence of *EcDsbA* overlaps. Therefore, this does not allow us to determine if this reaction was being catalysed by the protein.





**Figure 2.8:** Peptide oxidation time course by  $^1\text{H}$  NMR. Overlay of proton signals corresponding to reduced (10.18 ppm) and oxidised (10.08 ppm) peptide **23** in the presence of oxidised *EcDsbA* and under A) an atmosphere of air or B) an atmosphere of nitrogen. C) Half-life of reduced peptide **23** in the presence of oxidised *EcDsbA* and under an atmosphere of nitrogen calculated using the normalised peak integration of  $^1\text{H}$  NMR signals.

Ethylenediaminetetraacetic acid (EDTA) was introduced to further reduce background oxidation by the means of metal coordination. This resulted in an extension of peptide half-life by up to 20 hours, and increased the stability of the disulfide complex by up to seven days. Reducing the stoichiometric ratio of peptide to protein to 1:1 (as opposed to 10:1) still resulted in the mixed disulfide intermediate. As no peaks correlating to the oxidised protein were observed it was assumed that the peptide fully labelled the

protein and no free ligand was left in solution. It was thought that the removal of free peptide would slow the degradation of the complex, however, its residency was reduced by five days.

Surprisingly the resolution of the complex always resulted in the observation of oxidised peptide and oxidised *EcDsbA* with no trace of reduced peptide or reduced *EcDsbA*. Mechanistically it would be expected that the cysteine of peptide **23** would undergo nucleophilic attack by its corresponding cysteine in a free molecule of **23** thereby releasing reduced *EcDsbA* and oxidised peptide. Reduced *EcDsbA* is stable in solution for ~ 10 days before aerial oxidation converts it into oxidised *EcDsbA*. As there was never any indication of the presence of reduced *EcDsbA* in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra, the mechanism of DsbA introducing the peptide's disulfide bond and regenerating oxidised protein through aerial oxidation is not supported. In lieu of this mechanism, it would be expected that the interaction with *EcDsbA* is reversible. This would insinuate that the accumulation of oxidised peptide is caused by background oxidation only and would explain why reduced protein is not observed. Furthermore, it would provide a reason as to why the half-life of the peptide did not decrease in the presence of *EcDsbA*. Conversely, the formation and stability of the mixed disulfide complex under an atmosphere of air does not support this theory. If a reversible reaction was occurring between the reduced peptide and oxidised protein the intermediate state should not be formed under air as the half-life of the reduced peptide in these conditions is < 30 mins.

Although this system was designed to investigate how oxidised DsbA interacts with reduced peptide, it is also possible that any reduced DsbA released into solution was reacting with oxidised peptide. As the peptide is derived from *EcDsbB* it would not be surprising if the oxidised peptide was able to bind to the reduced protein and react rapidly. The addition of reduced peptide to reduced *EcDsbA* generated a  $^1\text{H}$ - $^{15}\text{N}$  HSQC profile identical to the chemical shift profile for the complex obtained by reacting reduced peptide with oxidised *EcDsbA* under the same conditions. The complex was also formed by the reduced protein at the same rate as it was formed when beginning with oxidised protein (< 30 mins). This shows that the bond is not formed exclusively between oxidised *EcDsbA* and reduced peptide as < 30 mins is not long enough for

the reduced protein to have undergone background oxidation. However, the complex formed by mixing reduced peptide and reduced *EcDsbA* was only visible for 10 – 17 days and still resolved to oxidised peptide and oxidised protein. This was 2 – 3 times shorter than the complex formed with the reduced peptide and oxidised *EcDsbA* which was stable for 33 – 56 days. The ability of reduced *EcDsbA* to rapidly react with the peptide may explain why the reduced protein was never observed by HSQC and suggests that the trace amounts of reduced protein may play a role in turnover of the complex.

## 2.3 Conclusions and future directions

Previous studies have attempted to characterise the interactions between *EcDsbA* and its substrates, however, due to the transient nature of their interactions have had to rely on mutated proteins or stabilised non-disulfide covalent (thioether) intermediates. This work details the first time the native enzyme has been captured in an intermolecular disulfide bond with an unmodified peptide in physiologically relevant conditions.

Through the use of HSQC perturbations caused by binding of non-covalent peptides it was confirmed that binding occurs at the hydrophobic groove in a solution state. Although formation of an antiparallel  $\beta$  sheet between substrates and the protein are conserved in crystal structures the residues involved were not affected by the presence of these non-covalent version of the peptide. In concordance with previous studies it was found that the non-covalent interactions were of very low affinity ( $K_D > 3$  mM) and were not able to be quantified. Fragment binders which are known to bind in the hydrophobic groove generally have  $K_D$  values an order of magnitude lower than estimated for these peptides and introduction of a covalent labelling motif is likely to yield an efficacious inhibitor.

It is clear that aerial oxidation plays a major role in the oxidation of the mono-cystine peptide as well as resolution of the covalent intermediate. This was shown by the prolonged presence of the intermolecular disulfide bond when the protein-peptide complex was placed in an inert atmosphere and/or in the presence of EDTA. However,

the mechanism of the resolution of the covalently linked peptide-protein complex remains ambiguous. It is possible that the system is undergoing a 1:1 reversible binding event in a similar fashion observed by Duprez *et al* <sup>(242)</sup>. The mechanism proposed in that study requires Cys33 of *EcDsbA* to attack Cys30 to regenerate the original states of oxidised protein and reduced peptide. Alternatively, the complex could be resolved by the attack of the free thiol of a secondary peptide on the intermolecular disulfide bond to release the dimerised peptide and reduced *EcDsbA*. Each of these potential mechanisms have discrepancies with the observations made by NMR. Firstly, the disulfide bond was formed between the reduced peptide and both oxidised and reduced *EcDsbA*. This implies that if the system begun with the reduced protein a reversible binding event would regenerate reduced peptide and reduced protein. The resolution of the complex did not result in either 1D <sup>1</sup>H or <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra with profiles consistent with the presence of reduced peptide or reduced protein. Furthermore, as the reduced protein was never observed by <sup>1</sup>H-<sup>15</sup>N HSQC it is unlikely that the intermolecular bond was resolved by the mechanism expected by the traditional peptide/*EcDsbA*/*EcDsbB* oxidation cycle. This further highlights the difficulty in determining the exact mechanisms, stabilities and reaction rates for the covalent reactions of substrate or small molecules with *EcDsbA*.

Regardless of the difficulty in determining the exact mechanisms these studies have provided insights that will be useful for future medicinal chemistry campaigns. Notably a peptide functionalised with a free cysteine was able to trap *EcDsbA* in a stable intermediate, that is consistent with inhibition of DsbA activity <sup>(242)</sup>. Furthermore, the complex was formed by targeting both redox states and did not undergo drastic conformational changes upon labelling. High-resolution crystal structures have already been solved and together these observations suggest that a small molecule or fragment could be developed into a covalent inhibitor of *EcDsbA* using a thiol warhead and traditional structure-based drug design techniques.

# Chapter 3:

## Investigation of covalent thiol fragments

A major challenge faced in fragment-based drug design is the low binding affinity inherent in most initial fragment hits. It is often the case that many rounds of medicinal chemistry, generally heavily reliant on a knowledge of the structure of the fragment bound to its protein target, is required for optimisation of fragments into higher affinity compounds <sup>(67)</sup>. Recently there has been a resurgence within the pharmaceutical industry to implement strategies to develop compounds which covalently modify their protein targets <sup>(144, 152)</sup>. These are particularly applicable in the context of fragment-based drug design, as the addition of a covalent “warhead” allows for a rapid increase in potency. There are now several examples where it was possible to generate highly potent, and selective fragments either by attaching a fragment that binds to a target to an initially identified reactive motif or attaching an electrophilic functional group to a known fragment binder <sup>(80, 141)</sup>.

The reactivity of a functional group can be modulated by the properties of the chemical scaffold to which it is attached, greatly affecting the ability of a covalent modifier to label its target. Non-covalent protein interactions involving the chemical scaffold attached to the reactive functional group are thought to increase the rate and selectivity of the reaction by constraining the covalent warhead in close proximity to the reactive residues of the protein target. *EcDsbA* has a uniquely reactive catalytic site cysteine, which contains an active site cysteine residue (Cys30) that has a very low pKa of at ~3.5. This makes it ~1000 times more reactive than normal towards thiol-reactive functional groups <sup>(262)</sup>. This increased reactivity due to the low pKa, results in the thiolate being the predominant form of Cys30 in reduced *EcDsbA* at physiological pH, and provides a great opportunity to design highly selective covalent inhibitors using a weakly thiol reactive functional group as a covalent warhead.

Compounds that covalently bind to *EcDsbA* have been identified through the derivatisation of peptides derived from the sequence of the redox partner protein *EcDsbB* <sup>(242, 263)</sup>, its co-factor ubiquinone <sup>(251)</sup> and an endogenous peptide substrate of *EcDsbA*, SigA <sup>(229)</sup>. Furthermore, the characterisation of the mono-cysteine heptapeptide (chapter 2) has indicated that a stable ligand-protein conjugate can be formed by forming a disulfide bond between *EcDsbA* and a mono-cysteine peptide. This suggests that a similarly stable covalent complex could be replicated if a fragment

that bound adjacent to the active site of *EcDsbA* were to be appropriately functionalised with a free thiol. Furthermore, it was expected that suitable fragments with thiol warheads would act as inhibitors of *EcDsbA*. To test this hypothesis, a range of thiol-functionalised fragments was synthesised and tested.

Purification of synthesised compounds prior to testing represents an extensive bottleneck in medicinal chemistry. The development of biophysical techniques to screen libraries of unpurified single compound reaction mixtures that contain the desired compound would largely mitigate this issue and aid in the prompt identification of high affinity binders. Currently there are examples of this strategy applied using SPR <sup>(123-125)</sup>, affinity-selection mass spectrometry <sup>(127)</sup> and X-ray crystallography <sup>(128, 129)</sup>. SPR can be used to identify high affinity binders in unpurified reaction mixtures via analysis of the dissociation kinetics observed in sensorgrams. However, for this reason it is limited to analysis of compounds that bind non-covalently. X-ray crystallography is useful in characterising binding across a wide range of affinities and can give atomic level detail for binding events. In principle, X-ray crystallography can be used for analysis of both covalent and non-covalent binding. However, some targets are not amenable to crystallisation and the process of data acquisition, analysis and refinement can be time consuming. Mass spectrometry is able to detect the exact mass of protein adducts, unequivocally identifying the bound mixture component and its reaction stoichiometry with the target. It is ideal as only small amounts of reagents (both protein and compound) are required for screening, however in order to determine the site of labelling more involved experiments are necessary (such as trypsin digests). Using NMR to screen a library of reaction mixtures would provide an orthogonal method, with reasonable throughput, for characterisation of compound interactions and covalent complex formation. These experiments would provide residue-specific, structural information about the binding event. The information acquired by mass spectrometry and NMR complement each other well and together they provide a promising avenue for screening libraries of crude products for potential covalently modifying compounds.

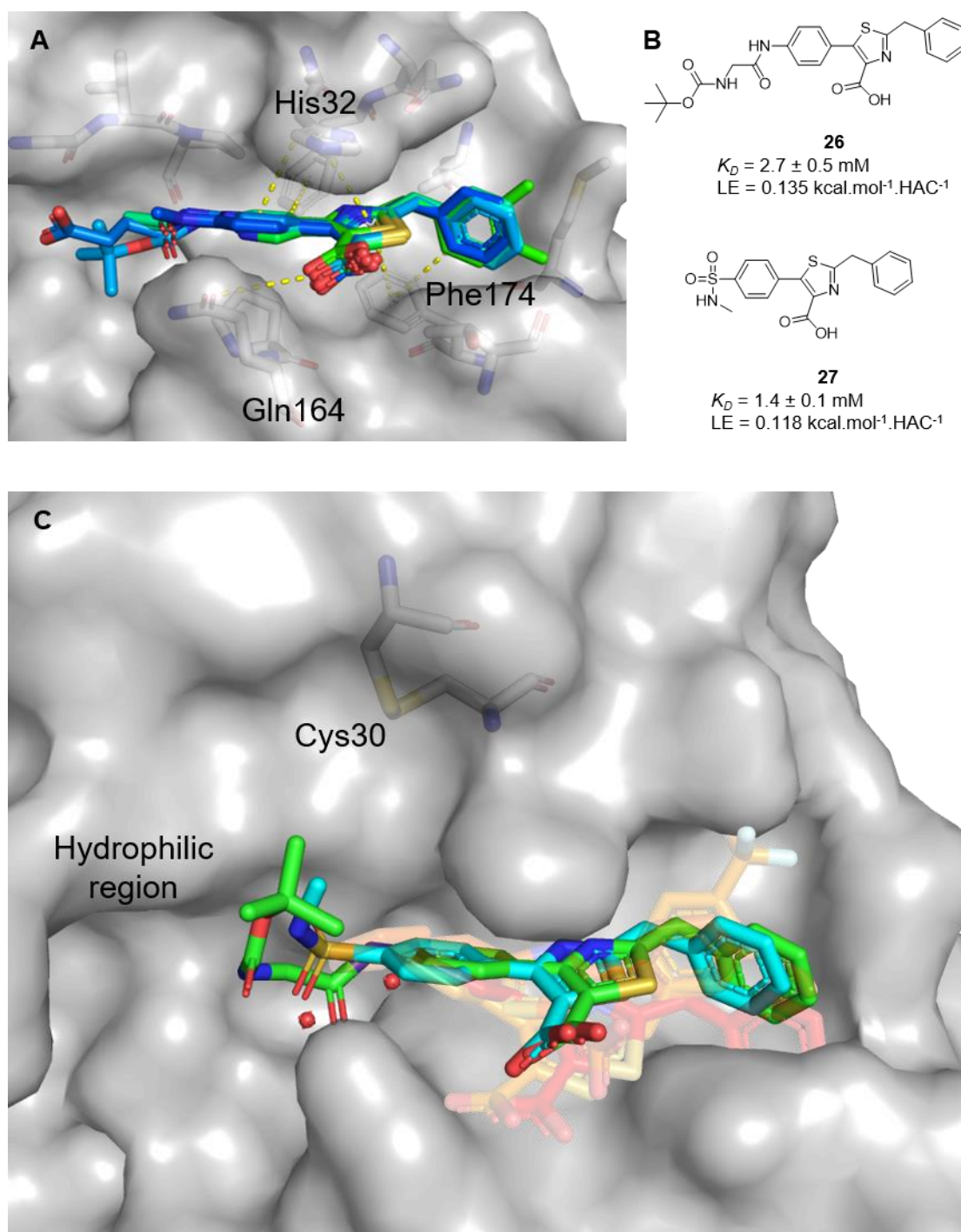
In this chapter we aimed to develop new NMR techniques to screen libraries of reaction mixtures and apply these in conjunction with mass spectrometry to covalently modifying thiol functionalised small molecules as potential *EcDsbA* inhibitors.

### 3.1 Structural analysis of small molecule inhibitors

Previous optimisation of small molecules as *EcDsbA* inhibitors within the group has yielded crystal structures of multiple scaffolds and their elaborated analogues <sup>(128, 233, 252, 253)</sup>. An analysis of in-house crystal structures was undertaken to identify a suitable fragment for expansion with a reactive thiol. Two series of compounds, benzyl thiazoles <sup>(233)</sup> and diaryl ethers <sup>(253)</sup>, are known to bind with mid  $\mu\text{M}$  to low mM affinity along the hydrophobic groove, adjacent to the active site of *EcDsbA* (Figure 3.1, 3.2). Compounds that provide synthetic handles or vectors which could be expanded toward the catalytic cysteines with a thiol warhead were required for development of thiol inhibitors.

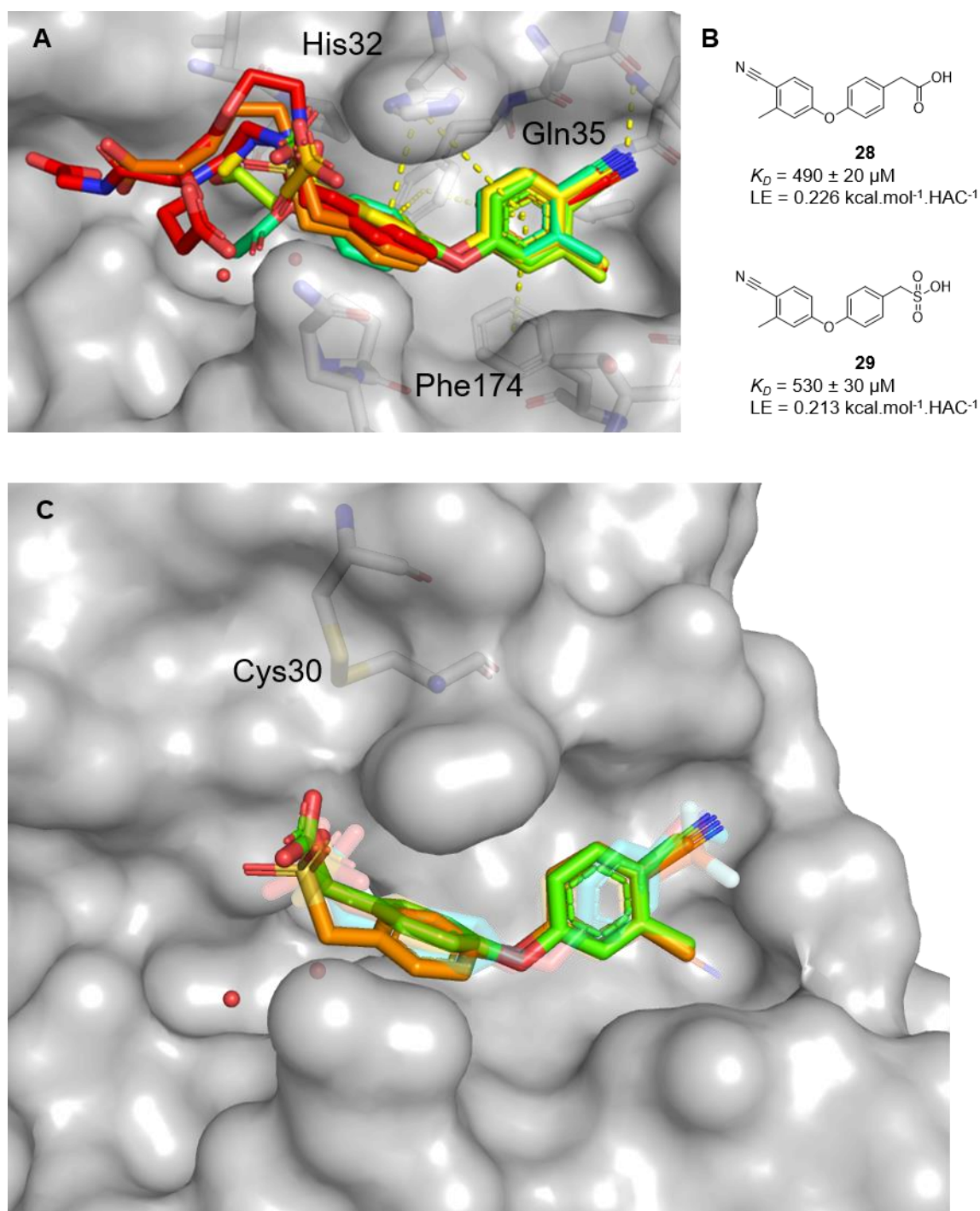
The series built upon the benzyl thiazole core gave 38 high resolution crystal structures where ligands show unambiguous binding poses. The core consistently adopts planar poses along the hydrophobic groove with most optimisation focussing on extending out of the groove toward the hydrophilic region. The consistent interactions made in this series are face-face or face-edge  $\pi$ - $\pi$  stacking interactions with aromatic residues His32, Phe36, Phe174. In addition, the most developed compounds in the series form a hydrogen bond to the sidechain of Gln164 (Figure 3.1). As  $\pi$  stacking interactions with His32 were found to be a key element of the pharmacophore for this binding pocket, most of the efforts to improve the binding affinity of this series involved the design of analogues that were intended to make additional interactions by adding polar substituents to the core that expanded toward the hydrophilic region. As a result of this strategy, there were only five compounds that contained expansion vectors that projected towards the active site of *EcDsbA*, and of these five, only amide **26** and sulfonamide **27** maintained the conserved core binding pose (Figure 3.1). These two benzyl thiazole compounds which provided a potential expansion vector showed low affinity  $K_D$ 's of  $> 1 \text{ mM}$  and  $\text{LE} < 0.15$ .





**Figure 3.1:** Binding modes of benzyl thiazole fragment series. A) Conserved interactions of the benzyl thiazole core with interactions shown as yellow dashes, fragments shown as coloured sticks, protein residues of the hydrophobic groove shown as grey sticks. B) Binding affinity characterisation of fragment analogues C) Overlay of benzyl thiazole analogues with expansion vectors towards the active site. Catalytic cysteines shown as grey sticks, structural waters as red spheres, fragments as coloured sticks, fragments which did not maintain the optimised binding mode shown as transparent.

The second series containing the diaryl ether core, had produced 19 crystal structures where either full or partial ligand density was observed, which were available for analysis (further crystal structures have since been published <sup>(128)</sup>). The diphenyl ether analogues had a conserved binding mode for the two main aryl rings, consistently making  $\pi$  stacking interactions with His32, Phe36, and Phe174. Compounds with a *para*-nitrile group also maintained a hydrogen bond to the sidechain amide of Gln35 (Figure 3.2). Compounds with the 2-methyl benzonitrile ring were found to have generally higher affinities, presumably due to their ability to make the best hydrophobic contacts within the right-hand side of the hydrophobic groove. The opposing ring showed that *para* substituted acids bound in almost identical poses, made a hydrogen bond to His32 and provided a direct vector toward the active site. Additionally, amidation of this carboxylic acid generated many alkyl amides that expanded from this vector and filled part of the channel between His32 and Val150 leading to the active site Cys30 and Cys33, suggesting that optimisation of these compounds would be able to reach the active site (Figure 3.2).



**Figure 3.2:** Binding modes of diaryl ether fragment series. A) Conserved interactions of the diaryl ether core with interactions shown as yellow dashes, fragments shown as coloured sticks, protein residues of the hydrophobic groove shown as grey sticks, structural waters shown as red spheres. B) Binding affinity characterisation of fragment analogues. C) Overlay of diaryl ether analogues with expansion vectors towards the active site. Catalytic cysteines shown as grey sticks, structural waters as red spheres, fragments as coloured sticks, fragments which did not maintain the optimised binding mode shown as transparent.

The carboxylic and sulfonic acid functionalised ethers **28** and **29** had a similar vector as seen in benzyl thiazoles **26** and **27**, however they had higher affinities, with  $K_D$  values of  $490 \pm 20 \mu\text{M}$  and  $530 \pm 30 \mu\text{M}$  respectively, and both had LE values  $> 0.2$  (Figure 3.2). This suggested that these compounds were more developable than the benzyl thiazoles and therefore, the covalent compound library was designed around diphenyl ethers **28** and **29**.

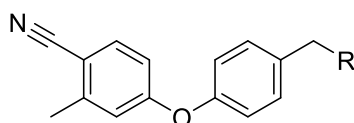
### 3.2 Design of targeted thiol fragment series

The side chains of Cys30 and His32 from the catalytic motif of *EcDsbA* show conformational flexibility in different crystal structures, which include the protein in different redox states as well as bound complexes with peptide substrates (Figure 3.3a). Cys33Ala mutations of *EcDsbA* have also been used to capture complexes with intermolecular disulfide bonds in order to prevent substrate turnover<sup>(222, 229)</sup>. The structure of unbound wildtype *EcDsbA* (PDB ID: 1FVK<sup>(243)</sup>) is most often used for non-covalent docking calculations and is one of the higher resolution structures available. However, it was reasoned that a structure with a covalently labelled Cys30 would provide more relevant protein conformation for designing and evaluating potential covalent fragments. A structure of the complex of *EcDsbA* bound with redox partner *EcDsbB* (PDB ID: 2HI7<sup>(222)</sup>) was considered for use in covalent docking calculations, however due to the low resolution of this structure (3.7 Å) and the uncommon rotamer observed for His32, this was determined to be unsuitable. Therefore, the structure of *EcDsbA* in complex with a short peptide derived from the loop of *EcDsbB* which binds within the hydrophobic groove (PDB ID: 4TKY<sup>(242)</sup>) was selected and prepared for docking studies (see chapter 8 for full experimental details).

The diaryl ether core was enumerated with alkyl thiols either directly attached to the core, or attached via an amide or sulfonamide and covalently docked to *EcDsbA* (Figure 3.3b, table 3.1). A short (1 carbon), medium (5 carbon) or long (10 carbon) alkyl chain was explored to approximate the number of heavy atoms required to retain the binding pose of the parent compounds **28** and **29**. None of the long alkyl chains maintained the correct poses with the diaryl ethers flipping ring orientations, losing the

hydrogen bonding interactions with His32 and the chain itself did not sit close to the protein surface. The 5 carbon linkers either maintained their binding poses or resulted in minimal movement of the core. Although the interactions with His32 were lost, it was often due to the wrong orientation of the linker and not due to the extreme relocation of the amide/sulfonamide as seen with the longer linkers. The shortest linkers caused the compounds to shift leftwards away from the hydrophobic groove and toward the active site. This lost the nitrile interaction with Gln35 and significantly weakened their hydrogen bonds to His32 (now ~ 3.6 – 3.8 Å away), however they did gain a hydrogen bond to the backbone of Val150. These docking results suggest that amines with 2 – 5 heavy atoms between the nitrogen of the linker and the terminal thiol would provide the most ideal geometry without perturbing the fragment binding pose within the groove. The desire to constrain the compound warhead within close proximity of the active site led to the decision to restrict the flexibility of the linker by using 5- and 6-membered rings.

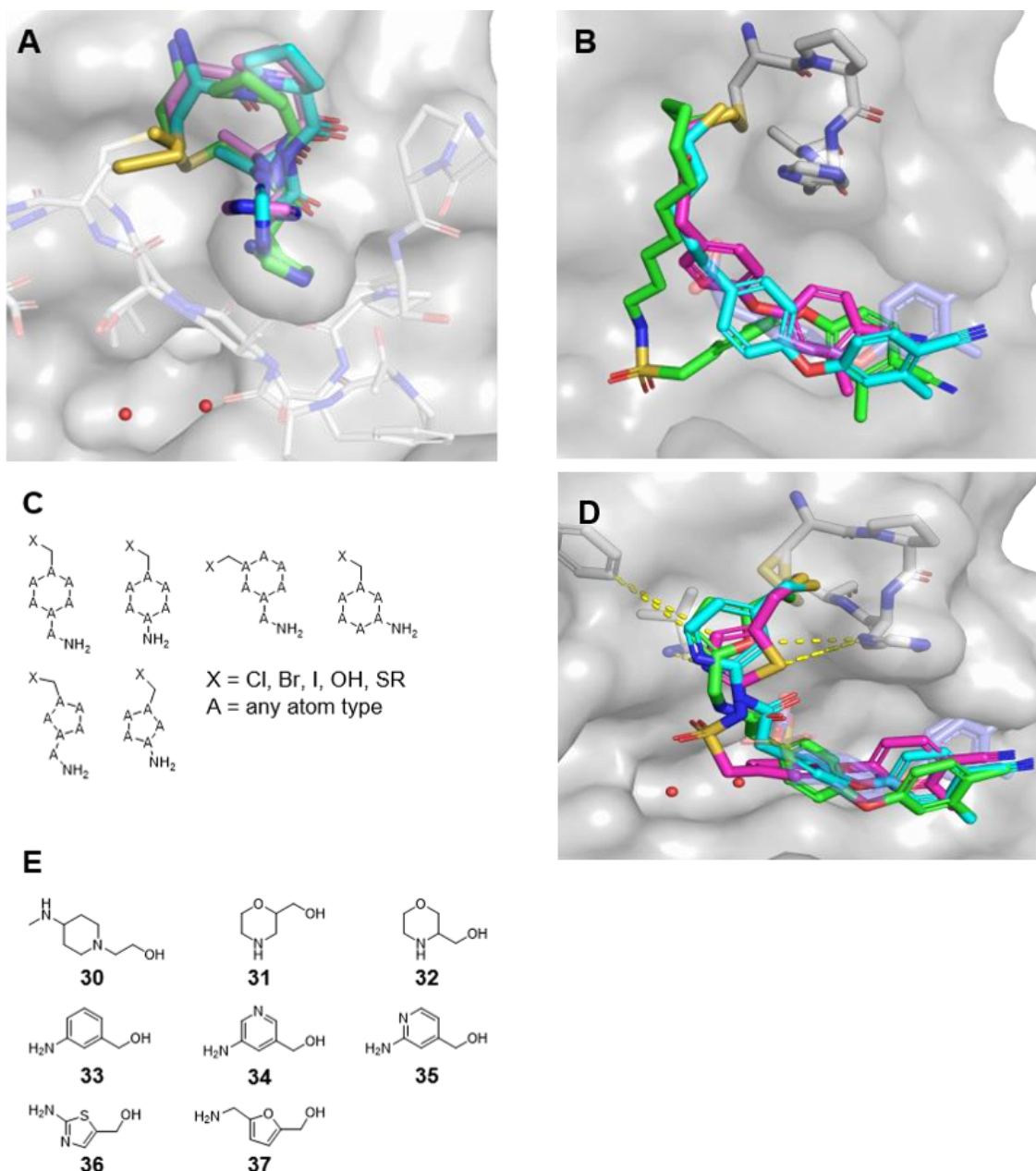
**Table 3.1:** Observations from covalent docking with alkyl thiols.



R =	Observation
	Shifted core to the left, replaced interaction between amide and His32 with Val150
	Shifted core to the right, lost interaction between amide and His32 (wrong orientation)
	Core flipped vertically and shifted to the right, lost interaction between amide and His32 (too far away)
	Shifted core to the left, flipped nitrile ring, replaced interaction between sulfonamide and His32 with Val150
	Core did not shift but flipped vertically, lost sulfonamide interactions to His32 (too far away)
	Core flipped vertically, sulfonamide lost interactions with His32 (too far away)
	Core binding adjacent to His32 instead of inside the hydrophobic groove
	Maintained binding pose and orientations
	Core shifted left (almost out of the hydrophobic groove) and flipped vertically

A targeted library of amines was designed by filtering commercially available compounds for the desired size and functional groups, covalently docked and selected from the best poses. Initially the commercially available building blocks from eMolecule were filtered to identify amines which had a terminal thiol or could undergo minimal functional group manipulations to form a terminal thiol (thioesters, alcohols, and alkyl halides, figure 3.3c). These compounds were then converted to the terminal thiols and grouped by the resulting product. They were then filtered to exclude compounds larger than 15 heavy atoms (this allowed for protection groups to be included), did not include a 5- or 6-membered ring and were manually inspected. A series of 92 reagents were then reacted *in silico* and converted to the desired fragment linked thiol analogues of ethers **28** and **29**. These analogues were then covalently docked to *EcDsbA* using CovDock <sup>(264)</sup> (Figure 3.3d).

The compounds with the lowest energy conformations were able to form additional hydrogen bonds to His32, Val150 or Pro151. The compounds which gained one or more of these interactions and maintained the binding pose of the parent diary ethers were considered for the small thiol library. The functional group manipulations for the compounds was to be conducted by parallel synthesis in 96 well plates and so they were purchased as the alcohol functionalised amines (**30** – **37**, Figure 3.3e).



**Figure 3.3:** Covalent molecular docking of thiol containing compounds. A) Overlay of available *EcDsbA* crystal structures. Overlay of the catalytic motif shown as sticks for oxidised *EcDsbA* (PDB ID: 1FVK, green), in complex with *EcDsbB* (PDB ID: 2HI7, *EcDsbA* as cyan sticks, *EcDsbB* as light grey sticks) and in complex with peptide substrate (PDB ID: 4TKY, *EcDsbA* as magenta sticks, peptide as light grey sticks) with structural waters shown as red spheres. B) Overlay of molecular docking results for alkyl chain linkers with the active site shown as grey sticks, bound poses shown as coloured sticks, and crystal structure of diaryl ether **28** shown as transparent purple sticks. C) Searching parameters for amine reagents for the targeted library. A denotes any atom and X is either a Cl, Br, I, OH or a thioether. D) Covalent docking results for targeted library showing new interactions (yellow dashes) between *EcDsbA* (grey sticks and surface) and predicted analogue poses (coloured sticks). E) Amino-alcohol reagents selected for the synthesis of the targeted thiol library.

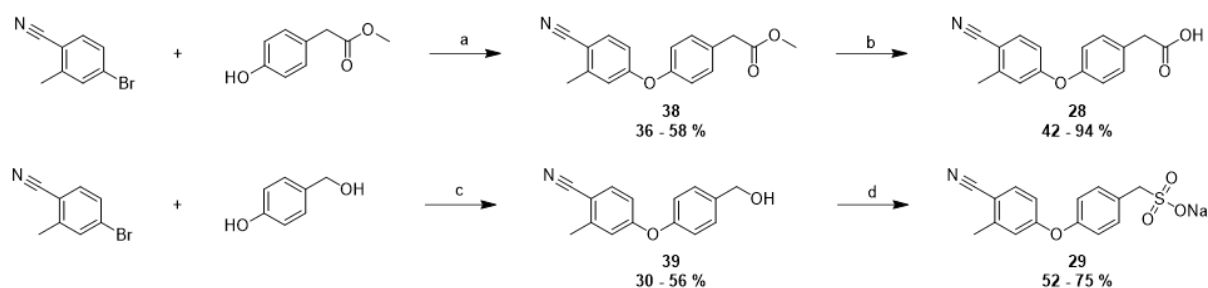


### 3.3 Parallel synthesis of thiol fragments

#### 3.3.1 Synthesis of diaryl ether precursors

Formation of the diaryl ether compounds **38** and **39** was achieved via a copper-mediated Ullmann coupling between 4-bromo-2-methylbenzonitrile and a *para* substituted phenol in moderate yields (30 – 58 %, Scheme 3.1). The methyl ester **38** was hydrolysed to the desired carboxylic acid **28** by lithium hydroxide as confirmed by the loss of the methyl signal in  $^1\text{H}$  NMR at  $\delta$  3.72 ppm and  $m/z$  of 266.8  $[\text{M-H}]^-$ , consistent with current literature <sup>(128)</sup>.

To obtain sulfonic acid **29**, initial Ullmann coupling with 4-bromo-2-methylbenzonitrile and 4-hydroxymethyl phenol gave alcohol **39**. Attempts to then convert alcohol **39** into a suitable leaving group started with bromination using *N*-bromosuccinimide and triphenylphosphine. The reaction was monitored by  $^1\text{H}$  NMR and it was observed that the benzylic protons shift upfield from  $\delta$  4.72 ppm to  $\delta$  2.75 ppm, consistent with bromination. The bromide **40** however, proved to be unstable during purification and the crude bromide product was therefore reacted with sodium sulfite to yield the desired sulfonic acid product **29** in low yields (10 – 23 %). As the sulfonic acid **29** was a key precursor to undergo a further four synthetic steps to obtain the thiol compounds an improved synthesis was investigated. The observed instability of the bromide was initially addressed via formation of the tosylate or mesylate from alcohol **39** with the corresponding sulfonyl chloride in DCM at 0 °C. Formation of the tosylate was observed in LCMS with peaks at 221.9  $m/z$   $[\text{M}]^+$  and 170.8  $m/z$   $[\text{M}]^-$  corresponding to the diphenyl ether benzyl cation and tosylate anion respectively, correlated to a new LC UV/Vis peak. However, upon addition of sodium sulfite no product was formed. Mesylation was conducted under the same conditions, the crude material **41** was treated with sodium sulfite and gave the desired sulfonic acid **29** in good yields (52 – 75 %) over two steps.



**Scheme 3.1:** Synthesis of parallel thiol library and precursors. Reagents and conditions: a) *N,N*-dimethylglycine hydrochloride,  $\text{Cs}_2\text{CO}_3$ ,  $\text{Cu(I)I}$ , 1,4-dioxane,  $120^\circ\text{C}$ , 18 h b)  $\text{LiOH}_{(\text{aq})}$ , THF,  $\text{H}_2\text{O}$ , RT, 18 h c) *N,N*-dimethylglycine hydrochloride,  $\text{Cs}_2\text{CO}_3$ ,  $\text{Cu(I)I}$ , 1,4-dioxane,  $120^\circ\text{C}$ , 18 h d)  $\text{Et}_3\text{N}$ , methanesulfonyl chloride, DCM,  $0^\circ\text{C}$ , 4 h. Sodium sulfite, EtOH,  $\text{H}_2\text{O}$ ,  $60^\circ\text{C}$ , 20 h

### 3.3.2 Microscale synthesis of thiol functionalised analogues

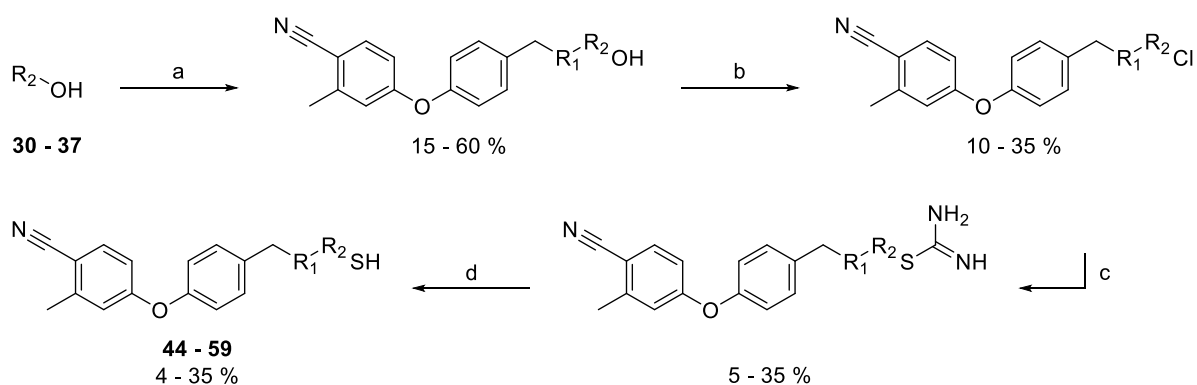
With both the key precursor carboxylic acid **28** and sulfonic acid **29** obtained in moderate yields the synthetic pathway to convert amino-alcohols to the corresponding fragment coupled thiols was investigated as a parallel microscale library synthesis. Initially a four steps synthesis, starting with acylation/sulfonylation of the amine via the acid chloride/sulfonyl chloride, halogenation of the alcohol, *S*-alkylation to the thiourea, and hydrolysis to the free thiol was proposed (Scheme 3.2) <sup>(265)</sup>. This synthesis was selected as the conditions and reagents used in these transformations were generally compatible with the use of microscale in thinly glass-coated 96-well plates, allowed for consistent conditions for both carboxylic acid **28** and sulfonic acid **29**, and used volatile reagents, hoping to reduce the amount of by-products remaining in the parallel microscale reaction mixture for screening.

Acid precursors **28** and **29** were added to triethylamine in thionyl chloride at room temperature to provide the corresponding acid chloride **42** and sulfonyl chloride **43** in quantitative yields (as determined by LCMS). This was followed by microscale acylation/sulfonylation with triethylamine in MeCN and after 18 hours the reaction was evaporated under vacuum. Product purities were estimated using LCMS and all reactions showed an *m/z* corresponding to the desired product with purities between 15 – 60 % measured by UV/Vis peaks at 215/254 nm.

The crude product mixtures were then redissolved in chloroform and treated with thionyl chloride to afford the chlorides in purities between 10 – 35 %, with  $m/z$  profiles showing a 3:1 isotopic abundance consistent with the presence of a single chlorine. Alkylation of thiourea with the crude chlorides in DMF resulted in LCMS spectra with many UV/Vis peaks. Although the thiourea analogues did not always provide distinct peaks the expected  $m/z$  for the desired products were observed. The reactions for which clear UV absorption data were observed saw purities of 4 – 35 %, suggesting almost quantitative conversion of the alkyl chlorides.

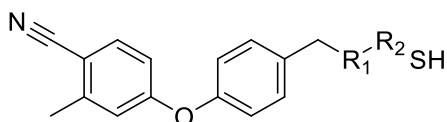
The isothiurea fragments were treated with sodium hydroxide for 4 hours after which the reactions were quenched with hydrochloric acid. Conversion to the free thiol was difficult to determine unambiguously as many by-products and previous intermediates overlapped or obscured peaks in the LCMS. Nonetheless, 50 % of reaction mixtures were determined to have the presence of the thiol compound by observation of the desired  $m/z$  corresponding to a peak in the UV/Vis trace. The carboxylic acid core and its sulfonic acid counterpart performed equally well, however the 6-membered aromatic amines **47 – 49** and **55 – 57** gave the best results having successfully reacted with both cores and giving the best purities over four steps (10 – 35 %).

After concentration of the reactions,  $d_6$ -DMSO stocks of the crude products were made to a nominal 100 mM concentration based on complete conversion over all four steps of the synthesis.

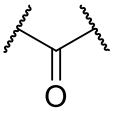
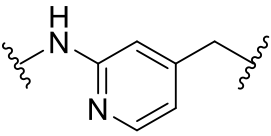
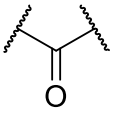
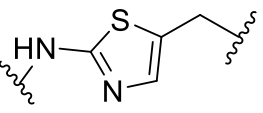
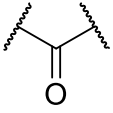
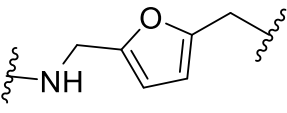
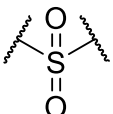
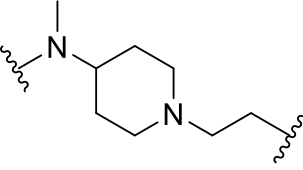
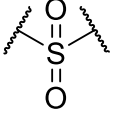
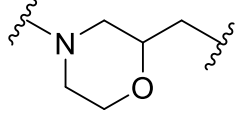
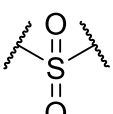
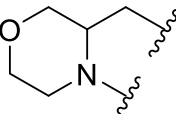
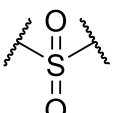
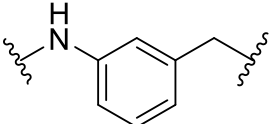
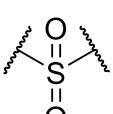
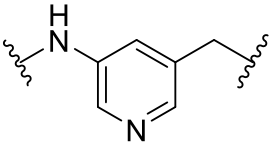
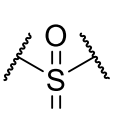
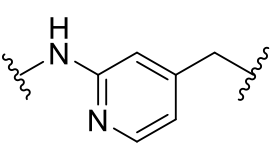
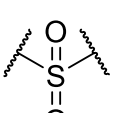
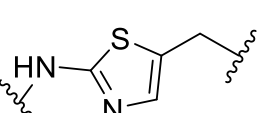
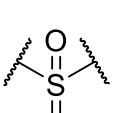
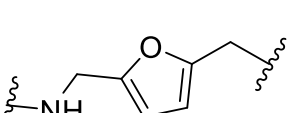


**Scheme 3.2:** Synthetic route for parallel synthesis of the targeted thiol library. a) **28** or **29**, thionyl chloride, RT, 2 h. Amine, Et<sub>3</sub>N, MeCN, 25 °C, 18 h b) thionyl chloride, CHCl<sub>3</sub>, RT, 19 h c) thiourea, DMF, RT, 12 h d) NaOH<sub>(aq)</sub>, RT, 4 h.

**Table 3.2:** Purity of parallel reaction mixtures for compounds **44** – **59**. Purity over the four steps was calculated by UV absorbance at 254 nm by LCMS. The purity was reported as the UV absorbance of the desired product as a percentage of all UV active peaks at 254 nm.



Compound no.	R <sub>1</sub>	R <sub>2</sub>	Reaction purity (%)
<b>44</b>			8
<b>45</b>			4
<b>46</b>			24
<b>47</b>			35
<b>48</b>			31

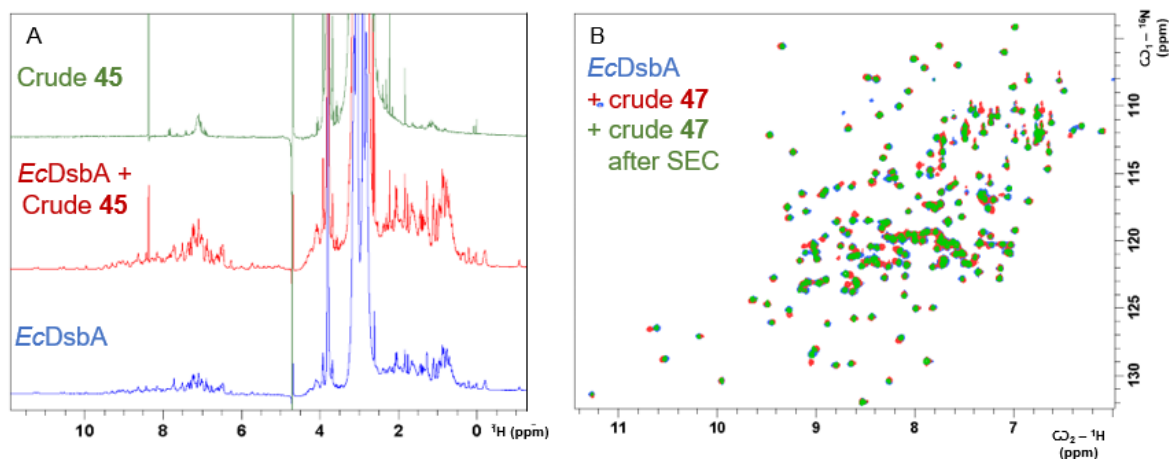
Compound no.	R <sub>1</sub>	R <sub>2</sub>	Reaction purity (%)
49			19
50			15
51			17
52			7
53			0
54			4
55			10
56			10
57			28
58			28
59			0

### 3.3.3 Covalent labelling of *EcDsbA* by unpurified reaction products

To assess the ability of each of the thiol-modified fragments to react with *EcDsbA*, each synthetic product was added at a notional concentration of 1 mM, without further purification from the reaction mixture, to 100  $\mu$ M *EcDsbA*. After 5 hours of incubation, the extent of cysteine labelling was assessed by LCMS. No crude products gave MS peaks expected for the covalent attachment of the corresponding thiol-fragment product.

In order to ensure that the crude products did not cause protein precipitation, unfolding or degradation prior to LCMS analysis, a comparison was conducted of  $^1\text{H}$  NMR spectra of the purified protein alone, the protein incubated with the reaction mixture of **45**, and the reaction mixture in the absence of protein (Figure 3.4a). The spectrum of the reaction mixture suggested the presence of multiple compounds in the aromatic region, consistent with the low conversion to the thiol observed by LCMS. Overlay of the three samples showed that neither the protein nor the crude products showed any clear change when incubated together.

To assess whether the conditions of the LCMS assay was causing the fragmentation of covalently attached adducts,  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments were recorded for *EcDsbA* after incubation for 5 hours with the mixture containing the target thiol which gave the highest purity, **47**. Following acquisition of the HSQC of the protein in the presence of the reaction mixture at a final  $d_6$ -DMSO concentration of 2 %, the sample was subjected to size exclusion chromatography in order to remove any free starting materials, reagents or by-products. Comparison of the spectra prior to and after the size exclusion purification was expected to allow distinction between non-covalent and covalent protein interactions. Although minor CSP were observed in the spectrum of *EcDsbA* after the initial incubation of the reaction mixture, they were not observed following size exclusion chromatography (Figure 3.4b). This indicated that the CSP were not caused by covalent labelling but by non-covalent interactions of reaction mixture components with *EcDsbA*.



**Figure 3.4:** Characterisation of crude products by NMR. A)  $^1\text{H}$  NMR spectra of 1 mM crude product **45** (green), 1 mM crude product in the presence of 100  $\mu\text{M}$  oxidised  $\text{U-}^{15}\text{N}$ -EcDsbA (red) and 100  $\mu\text{M}$  oxidised  $\text{U-}^{15}\text{N}$ -EcDsbA (blue). B)  $^1\text{H-}^{15}\text{N}$  spectra of oxidised 100  $\mu\text{M}$   $\text{U-}^{15}\text{N}$  EcDsbA (blue), incubated in the presence of 1 mM crude product **47** (red) and incubated sample after a size exclusion column (SEC, green).

It was hypothesised that these results may have been caused by false positives in thiol identification by LCMS during parallel synthesis, or dimerisation of the thiol products. We therefore undertook a more standard research scale synthesis and purification of a selection of the fragment-thiol analogues to address this concern. We anticipated that these authentic samples would allow us to validate the suitability of conditions used in the LCMS assay for analysis of the unpurified reaction mixtures. A similar approach has previously been used to validate conditions for screening unpurified reaction products using SPR and X-ray crystallography <sup>(125, 128)</sup>.

### 3.4 Synthesis of thiol fragments as pure compounds

Three amino alcohols were chosen for batch scale synthesis. Amino alcohols **32**, **34** and **36** were selected to sample the different reactivities, linker geometries and ring systems in the targeted library. As non-covalent interactions modulate and contribute to the binding of covalent inhibitors replacement of the diaryl ether fragment was also replaced with an unsubstituted benzene ring as control compounds. Initial attempts to sulfonylate each amine with sulfonyl chloride **43** failed to show any product, hence

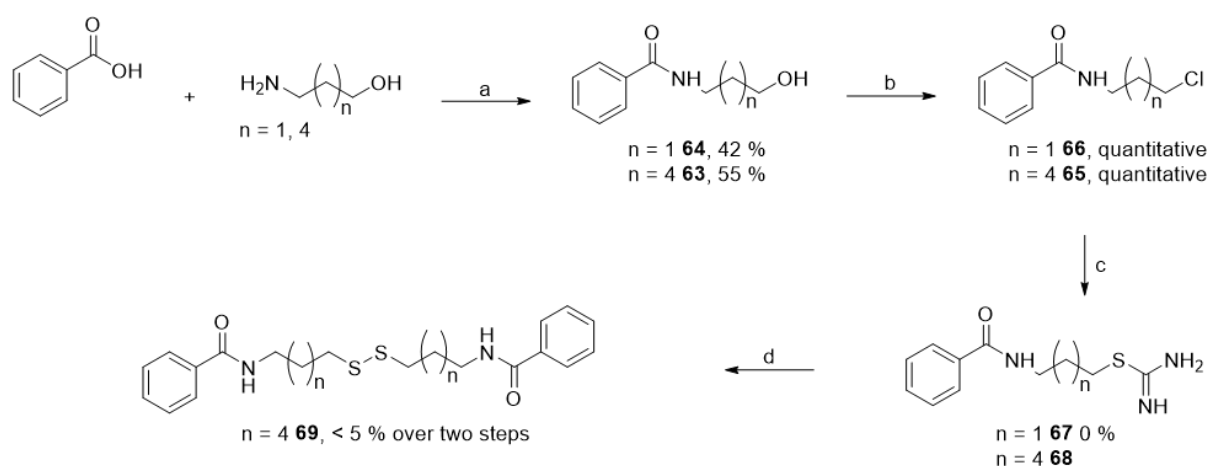
EDC coupling conditions were attempted with both the sulfonic acid **29** and carboxylic acid **28**. EDC coupling with carboxylic acid **28** showed minimal product conversion with amino thiazole **36** and pyridinyl amine **34**, with the major component by LCMS consistent with the EDC activated carboxylic acid intermediate. Although there was some evidence of the desired product in the crude product the pure product was unable to be isolated. EDC coupling of carboxylic acid **28** with morpholine **32** did not show any evidence of product formation.

As only a small amount (500 mg) of the amino alcohol library reagents was purchased and some reagents were sterically hindered, alkyl amino alcohols, 3-aminopropanol and 6-aminohexanol, were used in the further optimisation of reaction conditions as they are also desirable analogues for investigation of covalent binding to *EcDsbA*. The EDC coupling of sulfonic acid **29** and 6-aminohexanol was monitored by TLC and LCMS, however no indication of product formation was observed and only starting material sulfonic acid **29** was recovered. 3-Aminopropanol however, did give the desired sulfonamide **60** in 7 % yield. 6-Aminohexanol and 3-aminopropanol were successfully coupled with carboxylic acid **28** using EDC in moderate yields (**61** and **62**, 49 and 66 % yields, respectively). These amide couplings were replicated using commercially available benzoic acid in place of carboxylic acid **28** with comparable results (**63** and **64**, 55 and 42 % yields, respectively). As sulfonic acid **29** showed little or no conversion to the sulfonamides and carboxylic acid **28** showed moderate yields of the amides, benzoic acid was used to optimise the chemistry and conditions to save resources during synthesis optimisation (Scheme 3.3).

Hence, phenyl amido alcohols **63** and **64** underwent chlorination using thionyl chloride and the transformation was confirmed by the loss of hydroxy protons at  $\delta$  6.71 and 4.31 ppm in their  $^1\text{H}$  NMR spectra, respectively, and LCMS showing  $m/z$  profiles consistent with chlorinated compounds **65** and **66**. Nucleophilic substitution of chloropropyl **66** to the isothiurea **67** failed to yield any desired product as assessed by  $^1\text{H}$  NMR and LCMS with the reaction showing unknown by-products and no starting material was able to be recovered. The treatment of chlorohexyl **60** with thiourea was successful with product identified by LCMS. However subsequent column chromatography purification attempts and characterisation by 2D TLC indicated that



the desired S-alkyl product **68** was unstable on silica. Further attempts to precipitate the desired product, recrystallise or triturate the crude material failed to give high purity isothioureia **68** product. Hence, for these reasons the crude product of **68** was used in the next step without purification.

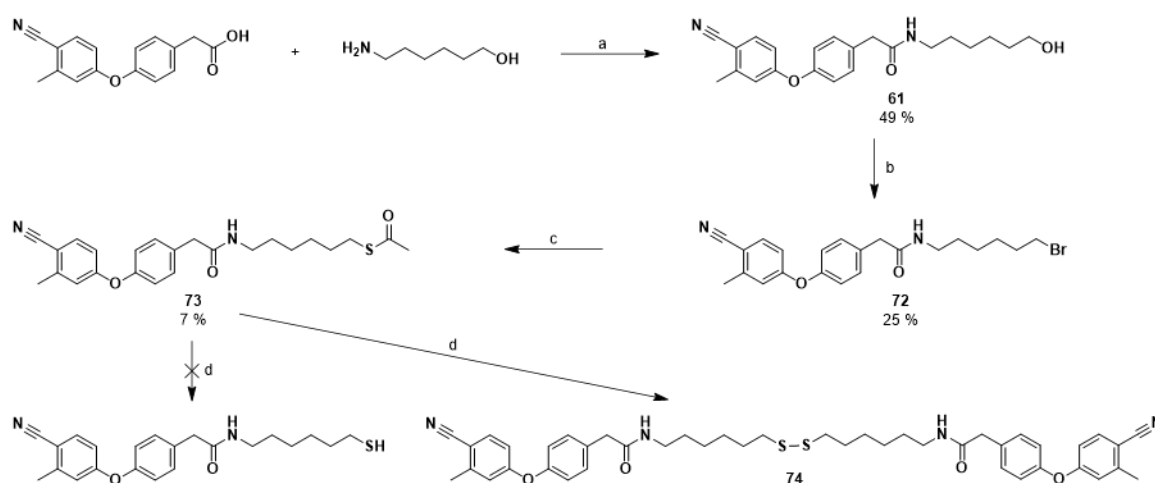


**Scheme 3.3:** Synthesis of benzoic acid analogues. Reagents and conditions: a) EDC.HCl, HOBT, DMF, N<sub>2</sub>, RT, 16 – 18 h b) thionyl chloride, chloroform, 0 °C – RT, 1 h, 80 °C, 16 h c) thiourea, DMF, N<sub>2</sub>, 120 °C, 3 h d) 1 M NaOH<sub>(aq)</sub>, RT, 1 h

Deprotection of **68** was achieved using NaOH (aq. 1 M) and loss of urea was supported by the observed 237.1 *m/z* corresponding to [M+H]<sup>+</sup> of the desired thiol by mass spectrometry. However, a minor *m/z* corresponding the 473.1 [2M+H]<sup>+</sup> ion was also seen and suggested the presence of the disulfide dimer by-product **69**. Unfortunately, the <sup>13</sup>C NMR spectrum showed that the carbon α to the sulfur was at a chemical shift consistent with the disulfide rather than the reduced thiol (generally > δ 35 ppm oxidised, < δ 32 ppm reduced) <sup>(266)</sup>. Although high conversion was observed in the reaction, the disulfide product **69** was only obtained in low yield after purification (< 5 %), however two major by-products were also obtained. These by-products had *m/z* peaks consistent with further oxidation to the sulfone and disulfone or sulfoxide analogues (489.0 [M+H]<sup>+</sup>, and 505.1 [M+H]<sup>+</sup>, respectively). It was clear from the compound instability, and oxidation problems observed while conducting step-wise synthesis of the fragment-thiol and benzamide-thiol analogues that this pathway was

not suitable for microscale parallel synthesis of thiol libraries and a new avenue had to be explored.

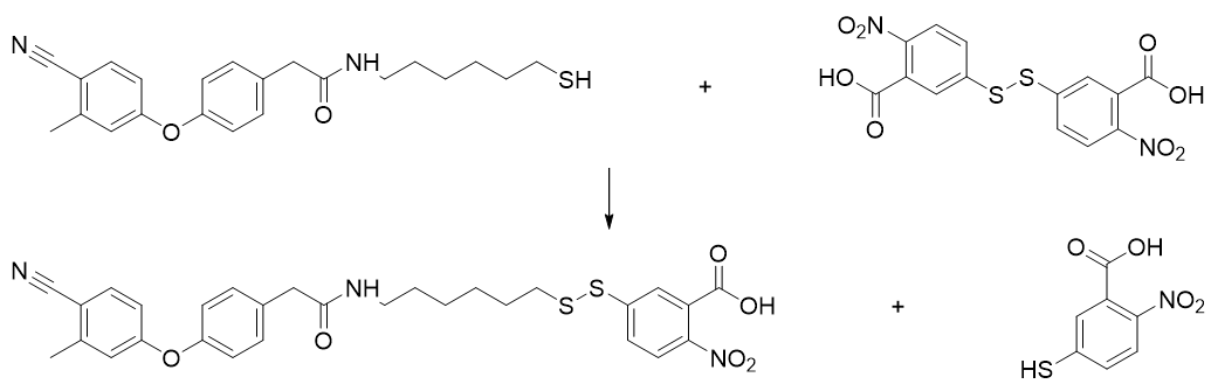
Promising initial results from acylation/sulfonylation on microscale prompted us to focus on optimisation of the S-alkylation and deprotection steps. Utilising thioacetic acid as the source of sulfur nucleophile may alleviate stability issues as well as allow for milder S-deprotection conditions in an attempt to avoid disulfide formation and further oxidations (Scheme 3.4) <sup>(267)</sup>. Alcohol **63** was therefore halogenated via an Appel reaction in the presence of triphenylphosphine and carbon tetrabromide to afford bromide **70** in a moderate yield (48 %). The thioester **71** was then formed by nucleophilic substitution with thioacetic acid and was confirmed by the appearance of acetyl protons at  $\delta$  2.32 ppm in the  $^1\text{H}$  NMR spectra. Subsequent cleavage with NaOH in water and acetonitrile at room temperature for 20 mins was then still found to give the disulfide **69** as the major product, similar to results observed during the thiourea synthesis. However, no evidence of further oxidation to sulfone and possibly sulfoxide species was observed with the milder hydrolysis conditions.



**Scheme 3.4:** Synthesis of hexanethiol diaryl ether analogue. Reagents and conditions: a) EDC.HCl, HOBT, N<sub>2</sub>, DMF, RT, 16 h b) CBr<sub>4</sub>, Ph<sub>3</sub>P, DCM, 0 – RT, 2.5 h c) thioacetic acid, Et<sub>3</sub>N, EtOH, RT, 16 h d) NaOH<sub>(aq)</sub>, MeCN, RT, 30 min

This improved synthetic strategy was then applied to fragment carboxylic acid **28**, where a yield of 7 % was obtained for the Appel reaction halogenation to bromide **72**. S-alkylation of thioacetic acid with bromide **72** to give thioester **73** proceeded with high conversion as judged by HPLC, however only 13 % of the desired product could be isolated. The stability of all intermediates on silica was examined by 2D TLC, which indicated that both the bromide and thioester were degrading under column chromatography conditions and this could explain the low isolated product yields.

A small-scale test hydrolysis of the thioester to the terminal thiol confirmed that the conditions used with benzamide analogue also resulted in the formation of a disulfide bond giving diaryl ether dimer **74**. It was evident that the free thiol would not be formed as the sole product, however an Ellman assay was used to interrogate the rate of the removal of the acetyl group and subsequent dimerisation. Ellman assays can be used to detect free thiols in solution by a change in colour and UV absorption caused by disulfide exchange of Ellman's reagent (5,5'-dithio-bis-(2-nitrobenzoic acid), DTNB) with a free thiol (Scheme 3.5). This releases 2-nitrothiobenzoate (TNB) into the solution which becomes a strong yellow colour and can be tracked by absorbance at 412 nm. It was hoped that we could use Ellman's reagent as a trap for any desired free thiolate product and follow the reaction by LCMS to identify the reaction time of acetyl hydrolysis which would provide the highest yield of thiol. DTNB was reduced using sodium borohydride to obtain a pure sample of the assay product, TNB. The retention times and UV absorbance profiles for both compounds were obtained to be used as standards in the assay. Furthermore, a standard curve was generated using DTNB and a serial dilution of cysteine which was then measured at 412 nm using the same LCMS method to be used to track the reaction.



**Scheme 3.5:** General reaction scheme for free thiol detection by Ellman assay

Aliquots of the acetyl cleavage of **73** were quenched every 5 minutes for 30 minutes using DTNB. Surprisingly none of the samples displayed the change in colour normally observed by the reduction of DTNB in the presence of free thiols. Analysis of the LCMS spectra indicated that the formation of the disulfide homodimer **74** occurred without the detection of the free thiol, indicating that it was unlikely that the free thiol would be able to be isolated. The lack of free thiol stability to even mild aqueous conditions lead us to re-evaluate the thiol warhead as a suitable warhead for covalent inhibitor design with *EcDsbA*.

### 3.5 Conclusions

Fragments containing free thiols as covalent warheads were investigated as potential small molecule inhibitors of *EcDsbA*. Molecular docking indicated that optimal linkers could potentially gain additional hydrogen bonds to His32, Val150 and Pro151. It also suggested linkers containing 5- or 6-membered rings would provide a suitable rigidity, linker length and geometry to place the thiol warhead in close proximity to the active site. A small thiol fragment library was designed and synthesised in parallel microscale reactions. The resultant crude reaction products were screened against *EcDsbA* by LCMS. Unfortunately, no clear evidence of covalent product formation was observed. Using 1D  $^1\text{H}$  and  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments it was clear that this was not due to fragmentation of the covalent adduct during the LCMS assay, nor was it due to the reaction mixture precipitating or unfolding the protein. It was hypothesised that the

desired thiol products were dimerising prior to addition to the protein and purified products could be used to confirm dimerisation and to optimise the MS assay.

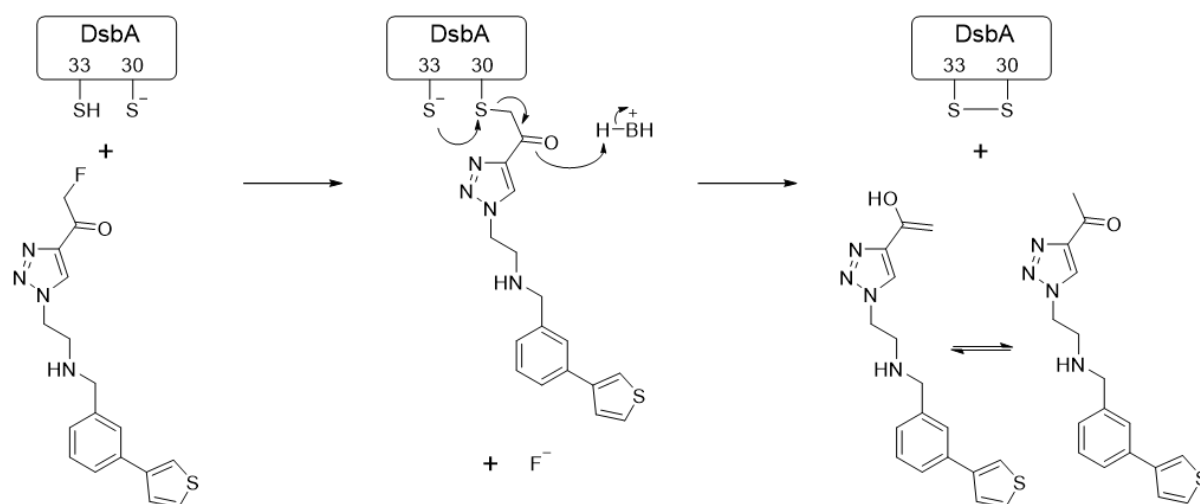
Batch scale synthesis using the diaryl ether cores **28** and **29** often resulted in low yields or no conversion. Two reaction pathways were explored to transform the terminal alcohols to the corresponding thiols. The intermediate halogens, thioureas and thioesters, were found to be unstable in solution, subject to degradation during purification and ultimately resulted in low yields. Furthermore, each attempt at hydrolysing the compounds to the free thiol caused the formation of disulfide homodimers. Careful time course analysis of the final deprotection step revealed that the desired thiols were oxidising equally as fast as deprotection was occurring, suggesting that the desired free thiol compounds would not be able to be isolated.

Although the optimised synthetic scheme could be applied in microscale, the lack of any detectable free thiol would be problematic for assays as well as ongoing covalent inhibitor design. Due to the problems encountered in both the synthesis and testing of fragments containing thiol warheads, it was determined that it would not be productive to follow up this series of thiol compounds with further rounds of medicinal chemistry. Instead we decided to find an approach which could identify a more tractable covalent warhead and chemistry more amenable to parallel or high-throughput chemistry.

# Chapter 4:

## Design and screening of an electrophilic warhead library

We have previously described the development of a reversible covalent inhibitor of *EcDsbA*. This was achieved by elaborating an *EcDsbA*-binding phenylthiophene fragment with an aryl fluoromethylketone (FMK) warhead using copper (I) catalysed azide-alkyne cycloaddition (CuAAC or click) chemistry <sup>(254, 255)</sup>. Characterisation of the reaction between the FMK and *EcDsbA* indicated that initially the warhead reacted with Cys30 to generate a thioether product as predicted. However, subsequent to this initial reaction the adjacent His32 facilitated the formation of the Cys33 thiolate, which resulted in nucleophilic attack on the thioether adduct to release the des-fluoro compound (Figure 4.1).

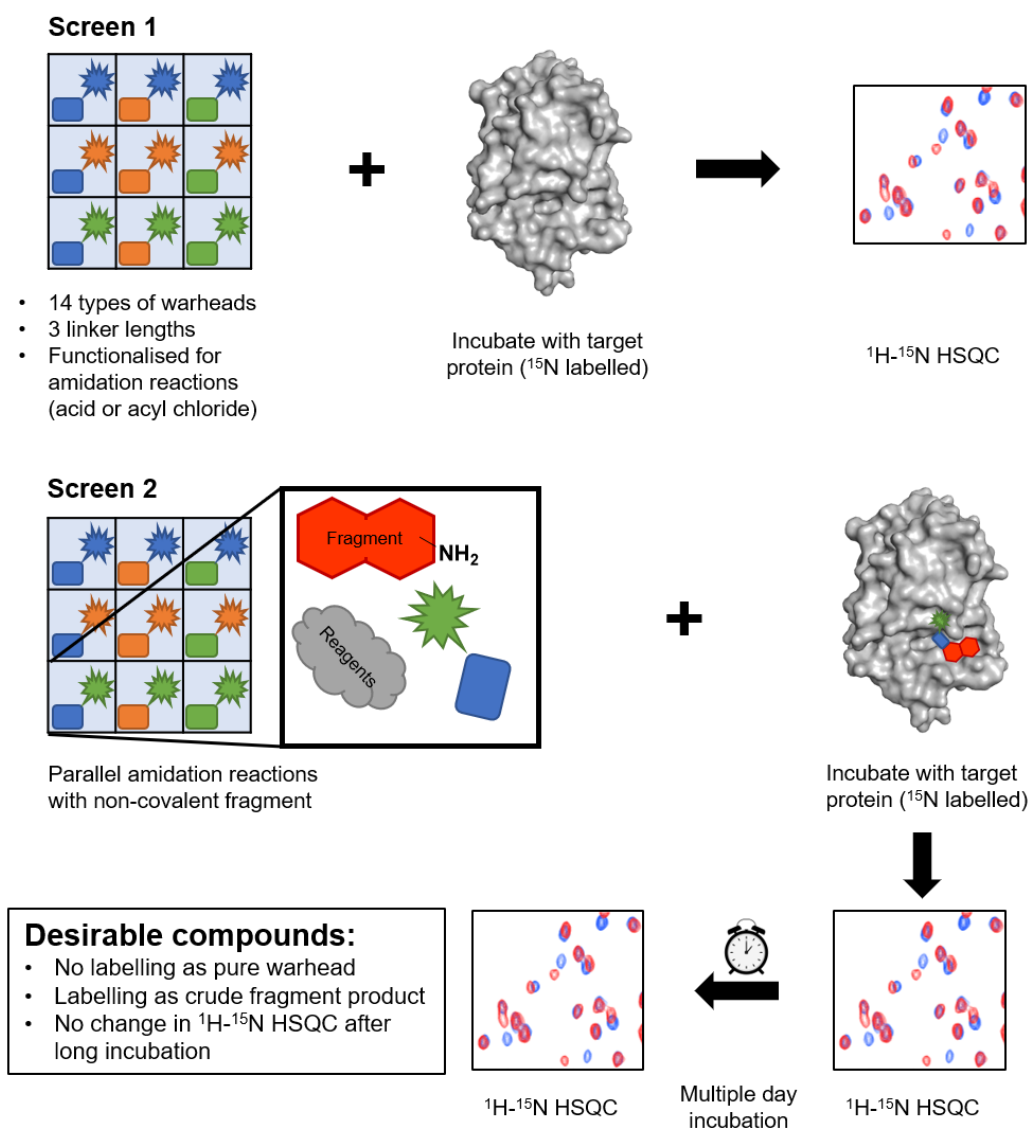


**Figure 4.1:** Proposed mechanism for labelling of *EcDsbA* at Cys30 and subsequent turnover of aryl fluoromethylketone fragment by neighbouring residues <sup>(254, 255)</sup>.

Further investigations revealed that *EcDsbA* demonstrated unusual and unpredictable reactivity. *EcDsbA* was also able to turn over maleimide, which is generally considered to be an irreversible thiol-reactive reagent which forms an  $\alpha$ -thiocarbonyl intermediate, and did not show any covalent labelling by an acrylamide functionalised fragment <sup>(254)</sup>. This highlights the importance of selecting the most appropriate warhead for a specific target as opposed to relying on the most commonly used motif for the desired residue.

As described in the previous chapters, several studies have demonstrated that the screening of crude products expedites the process of fragment optimisation (123, 124, 126-129). Therefore, our goal was to develop a similar approach to enable us to identify covalent inhibitors of *EcDsbA*. It was envisaged that this could be achieved by attaching a suitable electrophile to an *EcDsbA*-binding fragment via an appropriate linker. Based on the unexpected chemistries displayed by *EcDsbA*, a modular library amenable to parallel chemistry and consisting of a wide range of electrophilic warheads was designed. This library was to be tested by biophysical techniques against the protein alone and subsequently retested using a crude product containing the electrophilic warheads coupled to a known fragment binder (Figure 4.2). We sought to identify those warheads that formed covalent adducts with *EcDsbA* only when attached to the fragment.





**Figure 4.2:** General workflow for screening the electrophilic warhead library and the crude amidation products by  $^1\text{H}$ - $^{15}\text{N}$  HSQC. Screen 1) pure warhead screen, screen 2) crude fragment-coupled product screen.

Comparison of the reactivity with the warhead alone and when attached to the fragment would allow for the identification of motifs that are only weakly thiol reactive and may be enhanced by attachment to the fragment. We predicted that this would allow the identification of compounds which would be unlikely to react non-specifically with other cysteine-containing proteins and aid in the development of selective inhibitors for *EcDsbA*. To ensure that the compounds were not being turned over as

seen in previous studies, the stability of any covalent adducts was assessed by comparing the samples after short (hours) and longer (days) incubation times. Reactions with *EcDsbA* were monitored by recording  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the reduced protein following addition of a compound carrying the warhead. It was anticipated that reaction with *EcDsbA* would induce CSPs for resonances of residues close to the active site, and that following the initial reaction the spectra would remain unchanged. The appearance of chemical shifts that were characteristic of the oxidised form of *EcDsbA* would suggest that the compound was most likely being turned over, whereas further changes to the HSQC spectra over time would potentially indicate the formation of multi-adducts due to non-specific reaction with residues other than Cys30. While compounds having low selectivity, or compounds that are substrates may be potentially useful as probes of the scope of *EcDsbA*'s reactivity they were not desired as specific inhibitors.

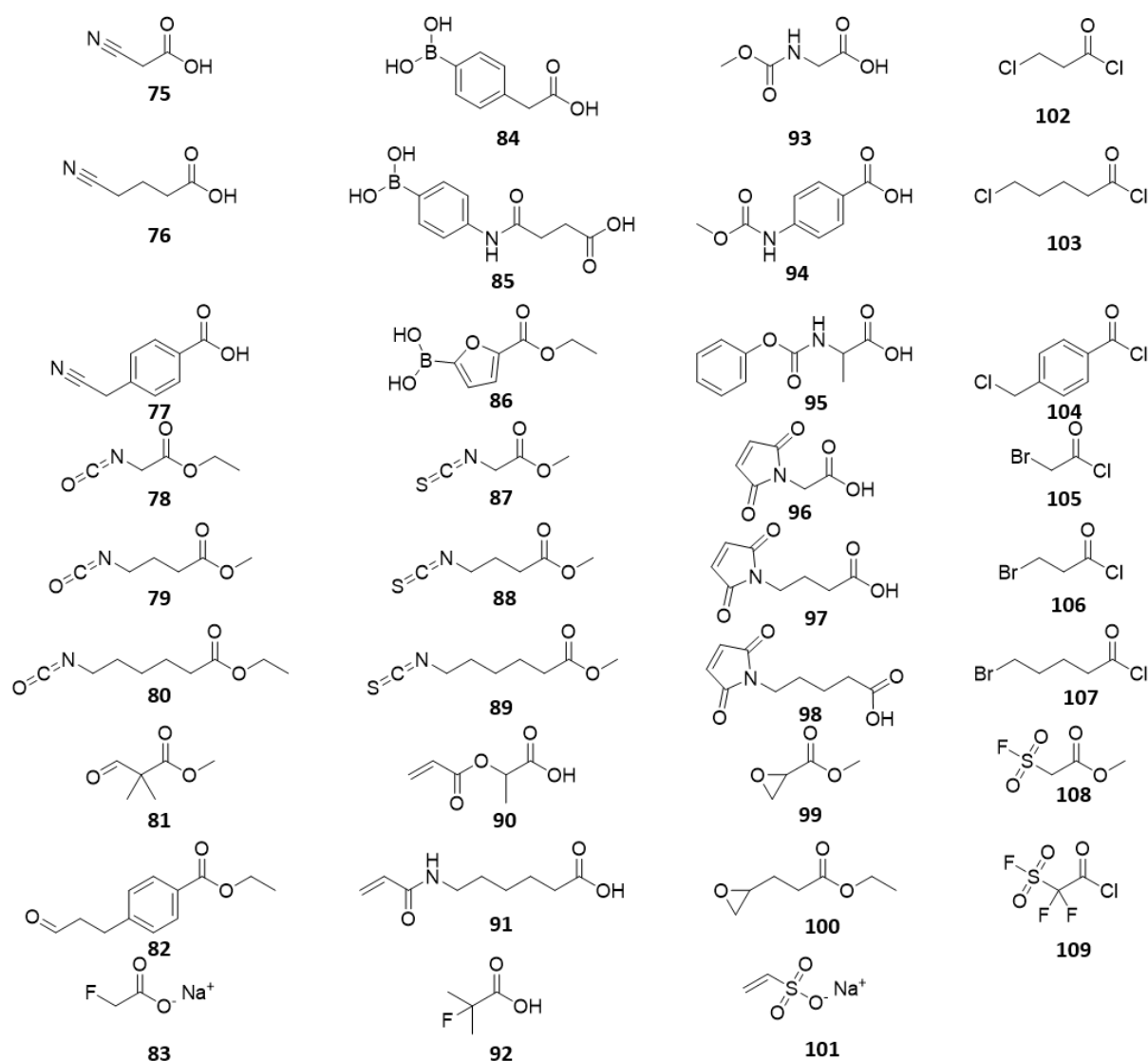
We aimed to design a covalent warhead library that was both suitable for the subsequent analysis of other protein targets and amenable to elaboration by attaching the warhead to a fragment using parallel microscale synthesis. The library was to be tested against *EcDsbA* to identify suitable warheads that could be coupled to *EcDsbA*-binding fragments to generate a stable and selective inhibitor for further medicinal chemistry optimisation.

#### **4.1 Electrophilic warhead library design**

This library was designed to be target agnostic and able to form adducts with any nucleophilic residue, including serine, threonine, lysine, tyrosine and cysteine. It was important for the versatility of this approach that warheads were not excluded due to their selectivity to a certain residue over others and therefore it was desirable to cover many different reactivities in the library while still being amenable to a consistent chemistry for linking to a fragment <sup>(152, 156, 173, 174)</sup>. Furthermore, we did not wish to discriminate between screening reversible or irreversible warheads and therefore many different reaction mechanisms were covered in the small library.

Reviewing the literature of covalent drugs and probes was undertaken to compile lists of covalent reactive warheads. Ultimately, aldehydes, nitriles,  $\alpha$ - $\beta$  unsaturated ketones, alkyl halides, isothiocyanates, isocyanates, boronic acids, vinyl sulfones, sulfonyl fluorides, epoxides, carbamates, maleimides, phosphonates,  $\beta$ -lactams and nitrogen mustards were selected as the reactive groups. Amide coupling and click chemistry were considered for coupling between the fragment and warhead as their conditions are relatively mild and should not compromise the integrity of the warhead during coupling. Therefore, commercially available reagents were filtered for these warheads with the condition that they were also functionalised with either a terminal alkyne, an acid, acyl chloride or sulfonyl chloride and methyl or ethyl ester.

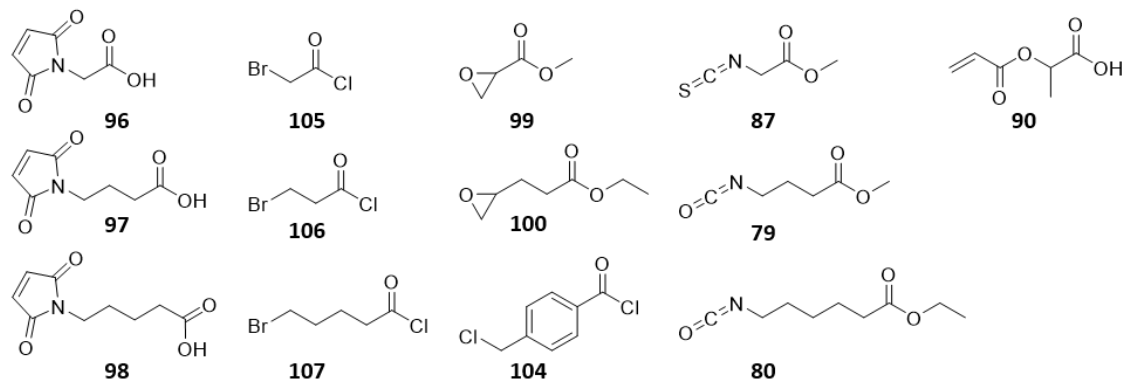
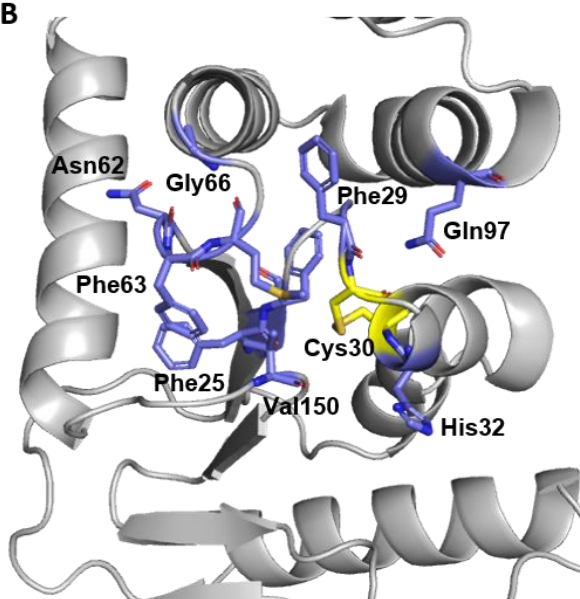
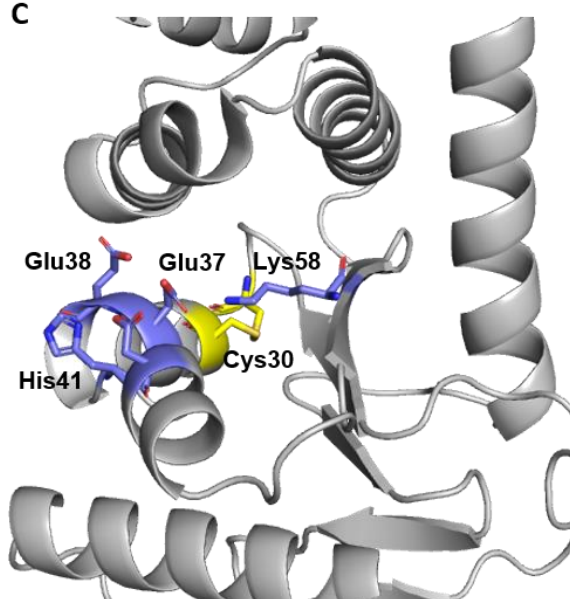
Due to a lack of commercial reagent availability, phosphonates,  $\beta$ -lactams and nitrogen mustards could not be included in the library, which unfortunately excluded the entire phosphorylation reaction class. Additionally, alkyne functionalised analogues were scarce therefore a library based on click chemistry would be difficult to compile commercially, requiring synthesising a bespoke in-house library to sufficiently cover the desired warhead diversity. Upon parallel library synthesis and attachment to a non-covalent fragment binder, it was expected that different length and chemical composition of the linkers between the fragment and the warheads can modulate reactivity, vary rigidity and warhead placement with respect to the reactive residue. It was intended to include in the library the same warhead with linkers of different length and composition. Linkers were generally classified as short (1 – 3 heavy atoms), of medium length (3 – 5 heavy atoms) or long (5+ heavy atoms) and each warhead generally had one analogue selected of each classification. Where highly similar linkers were available, the most chemically simple linker was selected. These considerations resulted in the final selection of an electrophilic warhead library for parallel chemistry that contained 35 compounds with 14 different warheads (**75 – 109**, Figure 4.3).



**Figure 4.3:** Final selection of compounds for the electrophilic covalent warhead library (75 – 109).

## 4.2 Pure warhead screen

The ability of each covalent warhead to react with reduced *EcDsbA* was tested by recording  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments of *EcDsbA* (100  $\mu\text{M}$ ) following addition of the warhead (1000  $\mu\text{M}$ ). Of the 35 library members 13 compounds were shown to affect the intensity of the active site cysteines of *EcDsbA*, which was considered to be indicative of forming a covalent adduct with the active site (Figure 4.4a).

**A****B****C**

**Figure 4.4:** Pure covalent warhead screening results A) Warheads which caused broadening of Cys30 or Cys33 of reduced *EcDsbA* by  $^1\text{H}$ - $^{15}\text{N}$  HSQC B) Residues which perturb or broaden in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra upon covalent modification (purple sticks) located adjacent to the active site cysteines (yellow sticks). C) Residues which perturb or broaden in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra upon covalent modification (purple sticks) located on the opposite side of the protein from the active site cysteines (yellow sticks).

The compounds that were observed to cause changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of *EcDsbA* were all of the maleimides (**96 – 98**), alkyl bromides (**105 – 107**) and epoxides (**99, 100**). It was unsurprising that these motifs appeared to form adducts with the protein as they are recognised as highly reactive groups and are often used as non-specific cysteine labelling agents <sup>(268, 269)</sup>. Alkyl chlorides are less reactive than

their bromide counterparts and chloride **104** was the only compound in the alkyl chloride warhead subset which showed any binding. It is possible that the linker of benzoyl chloride **104** contributes the reactivity of the warhead. The aromatic ring in the linker could activate the warhead through its electron withdrawing effects, however, it is also possible that this group is forming a  $\pi$ - $\pi$  interaction with the sidechain of neighbouring His32. This non-covalent interaction may enhance the extent of labelling by alkyl chloride **104** through proximity of the warhead. In the Michael acceptor subset, the warhead with the acrylate **90** was shown to bind however, the warhead with the acrylamide **91** did not. It is unclear why this would occur based solely on the intrinsic reactivity of the warheads, however this may be affected by thiolate of Cys30 which has been shown to display unpredictable chemical properties. Alternatively, the shorter length between the acrylate warhead of **90** and the carboxylic acid may have allowed for a hydrogen bond with His32 and, as suggested above, enhance labelling through non-covalent interactions. For the isocyanate (**78 – 80**) and isothiocyanate (**87 – 89**) warheads there are matched pairs with the same short (methyl), medium (propyl) and long (pentyl) linkers. Surprisingly, the shortest isothiocyanate **87** and the two longer isocyanates **79** and **80** caused large changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of EcDsbA, however the same linker attached to the opposite warhead (**78**, **88** and **89**) showed no effect at all.

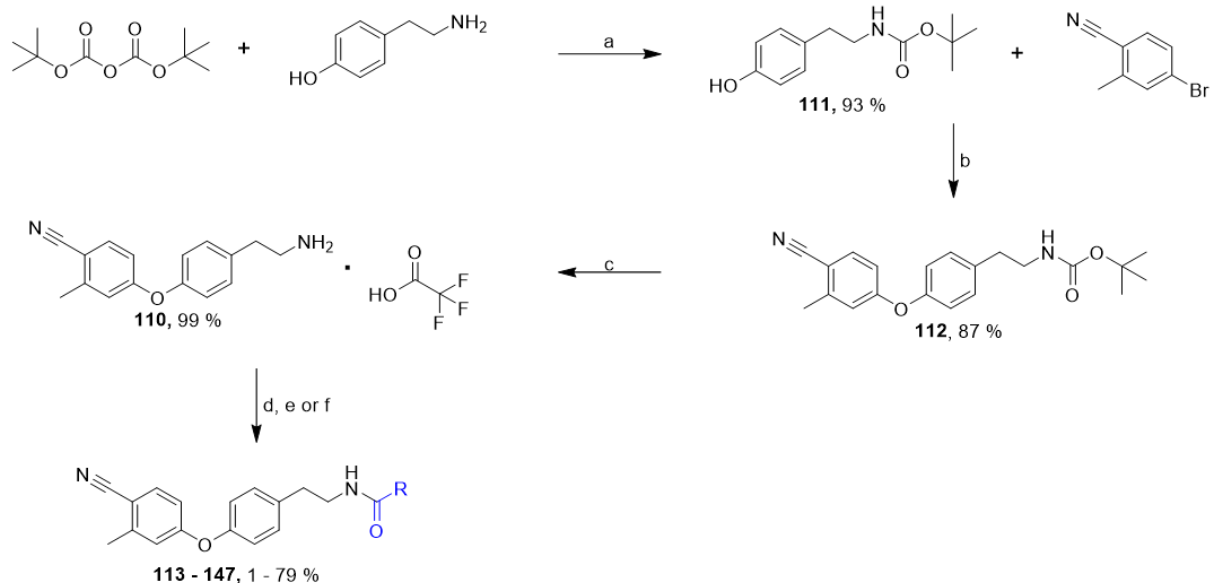
The warheads that did not show signs of a covalent adduct in the screen were often either weakly reactive or are known for forming reversible bonds. Some warheads that were expected to label such as the short and medium length alkyl chlorides (**102** and **103**), aldehydes (**81** and **82**) and vinyl sulfonic acid **101** may have been due to steric hinderance of Cys30 or the unusual nature of EcDsbA's intrinsic reactivity.

Where binding was observed, that pattern of changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of *EcDsbA* were consistent amongst the warheads:

1. All compounds caused perturbation or broadening of residues of the active site (Cys30, His32, Cys33, Figure 4.4b).
2. Large shifts or broadening of peaks also occurred for residues on the domain interface above the active site in most spectra which showed evidence of labelling (Phe25, Phe26, Phe29, Asn62, Phe63, Met64, Gly66, Gln97, Val150, Figure 4.4b). This supports the notion that the cysteine residues are being covalently modified. Many of these residues are not nucleophilic and are likely broadening due to the changes at the adjacent active site.
3. Solvent exposed and nucleophilic residues Glu37, Glu38, His41, Asp44 and Lys58 were broadened or shifted in each spectrum. This group of residues is located at the protein's acidic patch between the inserted  $\alpha$  helical and the thioredoxin domains on the opposite side of the protein to the active site (Figure 4.4c). These resonances for these residues are often affected by binding events at other sites (discussed below), however as these residues can be covalently modified this data may suggest off-target labelling of *EcDsbA*.
4. With the exception of warheads which showed a global loss of signals, residues that are found in the main binding site hydrophobic groove were unaffected in both intensity and chemical shift.

### 4.3 Synthesis and characterisation of non-covalent fragment

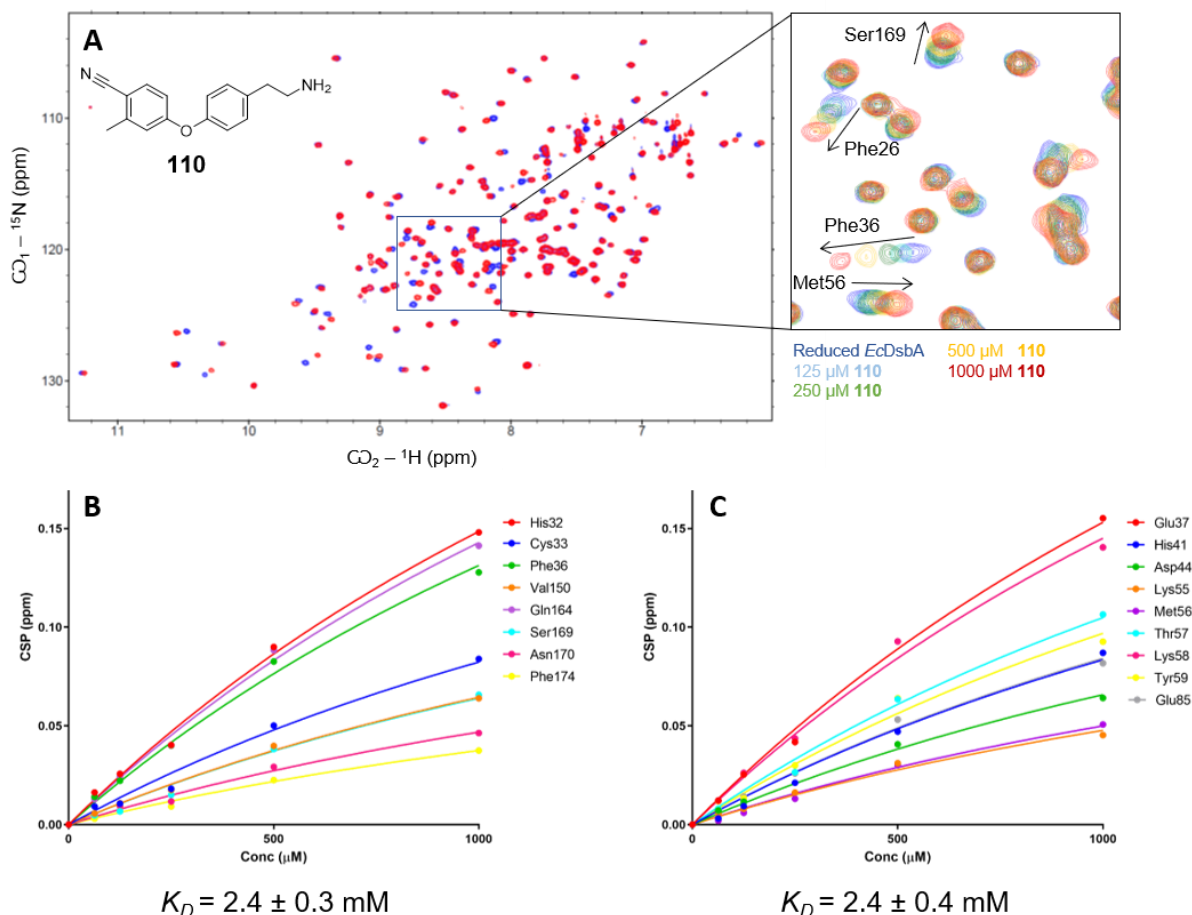
For parallel library synthesis with the warhead library an amine handle was required on a non-covalent fragment binder. Hence synthesis of the ethylamine diaryl ether analogue **110** was designed to be reactive with the functionalised warhead library via acylation. Amine **110** was synthesised as shown in Scheme 4.1. Tyramine was Boc protected to give phenol **111** which was then Ullmann coupled with 4-bromo-2-methylbenzonitrile to give Boc-protected diphenyl ether **112**. Trifluoroacetic acid deprotection then gave the desired ammonium trifluoroacetate salt **110** in overall high yield.



**Scheme 4.1:** Parallel synthesis of covalent fragment library. Reagents and conditions: a) THF, RT, 2 h b) *N,N*-dimethylglycine.HCl, Cu(I)I, Cs<sub>2</sub>CO<sub>3</sub>, 1,4-dioxane, 120 °C, 21 h c) TFA, DCM, RT, 1.5 h d) R-COOH, Et<sub>3</sub>N, HATU or EDC.HCl, DMF, RT, 18 h e) R-COCl, Et<sub>3</sub>N, DMF, RT, 18 h f) R-COOR<sub>2</sub>, Et<sub>3</sub>N, HATU or EDC.HCl, KOH<sub>(aq)</sub>, RT, 18 h

The binding of parent fragment amine **110** to reduced *Ec*DsbA was characterised by recording <sup>1</sup>H-<sup>15</sup>N HSQC titrations (Figure 4.5). The pattern of CSPs observed was to be used for comparison with future crude screening results to discriminate between non-covalent binding and covalent labelling. It was anticipated that similar CSP would be observed in the event of non-covalent binding of either unreacted **110** starting material in the reaction mixture, or non-covalent binding of the desired product. The CSP profile indicated that binding was occurring in the hydrophobic groove, with the largest localised shifts occurring for the resonances of residues His32, Gln164, Phe36 and to a lesser extent Cys33, Val150, Ser169, Asn170 and Phe174. These residues span the entire length of the hydrophobic groove and the region adjacent to the active site. It should be noted that although perturbations were seen for Cys33 of the active site, the peak intensity showed no change and the shifts were not large enough to indicate that the redox state of the thiolate was affected.





**Figure 4.5:** Binding characterisation of amine diaryl ether **110** A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC titration of diaryl ether **110** with reduced U- $^{15}\text{N}$  EcDsbA. Expansion shows contours for ligand concentrations of 0 (blue), 125 (cyan), 250 (green), 500 (yellow), and 1000 (red)  $\mu\text{M}$ . B) Binding isotherm of **110** at the hydrophobic groove plotting concentration ( $\mu\text{M}$ ) vs CSP (ppm) and fit globally using a single site binding model with ligand depletion. C) Binding isotherm of inferred effects of **110** plotting concentration ( $\mu\text{M}$ ) vs CSP (ppm) and fit globally using a single site binding model with ligand depletion.

Another large cluster of CSP were observed for residues on the opposite side of the protein to the active site, namely at residues Glu37, His41, Thr57, Lys58, and Tyr59. These residues are a part of the acidic patch (Glu37, His41) and a ridge which connects the two protein domains (Thr57, Lys58 and Tyr59). They are also commonly perturbed in  $^1\text{H}$ - $^{15}\text{N}$  HSQC titrations of fragments which have crystal structures confirming their binding to the hydrophobic groove (128, 233, 253, 254). Furthermore, analysis of the CSP for these residues indicate that these are unlikely to represent another binding event, since the CSP for these two clusters can be fitted to the same

1:1 binding model and the  $K_D$  value obtained from these calculations are essentially identical (Figure 4.5). It is more likely that these are inferred shifts due to changes in conformation or dynamics of these residues caused by binding at the hydrophobic groove.

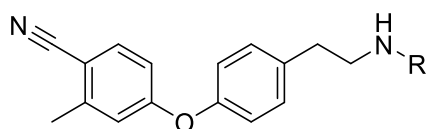
#### 4.4 Parallel synthesis of covalent fragment library

The covalent library was functionalised as acyl chlorides, carboxylic acids, sulfonic acids and esters and as such required three conditions to couple the amine fragment **110** to all warheads. Due to the small size of the warhead library, a separate condition optimisation phase using a few select library members was unnecessary. All reaction conditions were attempted with all relevant warheads to quickly find the highest reaction conversion and increase the chance of finding a condition which maintained the integrity of the reactive group. DMF and acetonitrile were both considered as the solvents as they were compatible with all reactions as they aprotic and therefore would not interfere with amide coupling conditions. However, the parent amine compound **110** was not soluble in acetonitrile and so DMF was selected for the synthesis. Furthermore, DMF is water miscible, which enabled *in situ* ester deprotection in aqueous base. Amide coupling reactions using HATU or EDC as acid activating agents were selected due to their mild nature and previous success in parallel synthesis within our research group for similar compounds <sup>(128, 253)</sup>. Acyl chlorides, which do not require further activation were directly added to fragment amine **110** with triethylamine base. Aqueous bases and concentrations were screened in small scale against epoxides **99** and **100**, boronic acid **86** and isothiocyanate **87** and crude products were analysed by LCMS and <sup>1</sup>H NMR to confirm the integrity of the warhead. Aqueous potassium hydroxide was selected for ester deprotection as these reactions suggested that it was able to hydrolyse the ester without degrading the reactive warheads.

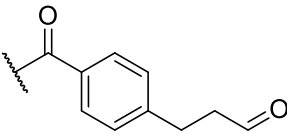
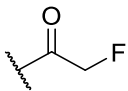
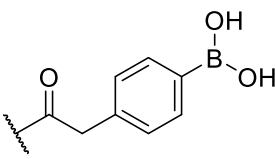
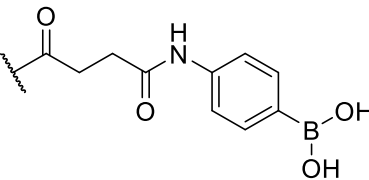
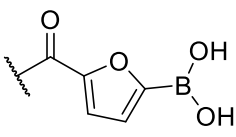
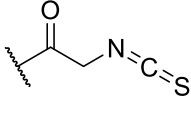
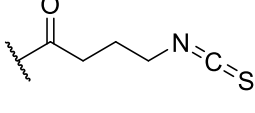
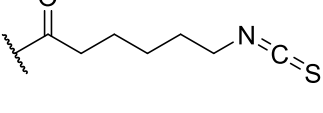
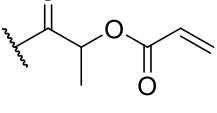
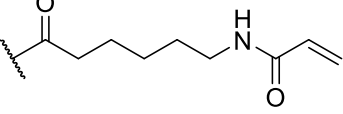
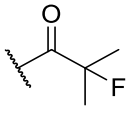
The reaction plates were set up and left at room temperature without agitation for 18 hours. The solvent and volatile reagents were then removed under vacuum and made to 100 mM *d*<sub>6</sub>-DMSO stocks. The stocks were diluted to 5 mM using a 1:1 acetonitrile:water solution, filtered and product purity was estimated from UV

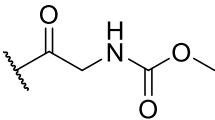
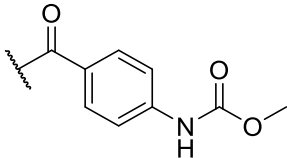
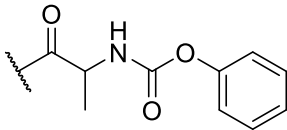
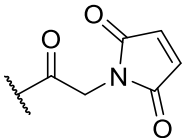
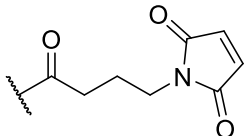
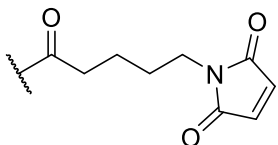
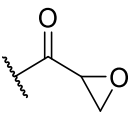
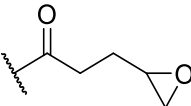
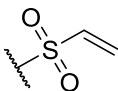
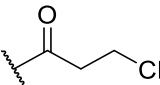
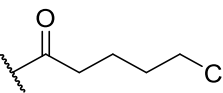
absorbance and  $m/z$  of crude samples by LCMS (Table 4.1). Desired products were observed in one or more reaction conditions for 78 % of the warheads, with 15 of the 35 warheads showing  $\geq 10$  % product purity and an  $m/z$  consistent with the desired product. Successful reactions using acid or acyl chloride functionalised warheads were obtained in moderate purities (average of 44 and 32 % respectively). Although the ester functionalised warheads in the library had low purities ( $\leq 10$  %), the correct  $m/z$  for the desired product was observed for all warheads except **146**.

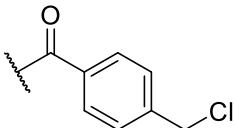
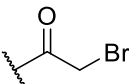
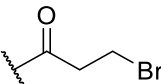
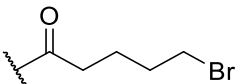
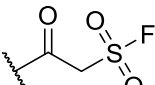
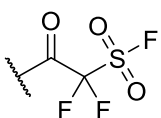
**Table 4.1:** Characterisation of crude parallel mixtures **113 – 147**



Compound no.	R group	Condition <sup>a</sup>	Product purity (%) <sup>b</sup>	Binding (Y/N) <sup>c</sup>
<b>113</b>		EDC	$m/z$ only <sup>d</sup>	N
<b>114</b>		EDC	50	N
<b>115</b>		HATU	37	N
<b>116</b>		HATU	$m/z$ only <sup>d</sup>	N
<b>117</b>		EDC	6	N
<b>118</b>		EDC	7	N
<b>119</b>		EDC	35	N

Compound no.	R group	Condition <sup>a</sup>	Product purity (%) <sup>b</sup>	Binding (Y/N) <sup>c</sup>
120		EDC	1	Y
121		EDC	17	N
122		HATU	48	N
123		EDC	12	N
124		EDC	<i>m/z</i> only <sup>d</sup>	N
125		HATU	2	Y
126		EDC	10	Y
127		EDC	<i>m/z</i> only <sup>d</sup>	Y
128		EDC	<i>m/z</i> only <sup>d</sup>	N
129		HATU	69	Y
130		EDC	0	N

Compound no.	R group	Condition <sup>a</sup>	Product purity (%) <sup>b</sup>	Binding (Y/N) <sup>c</sup>
131		HATU	45	N
132		HATU	4	Y
133		EDC	0	N
134		HATU	38	Y
135		HATU	79	Y
136		HATU	53	Y
137		EDC	28	N
138		EDC	<i>m/z only</i> <sup>d</sup>	N
139		EDC	0	N
140		N/A	40	N
141		N/A	0	N

Compound no.	R group	Condition <sup>a</sup>	Product purity (%) <sup>b</sup>	Binding (Y/N) <sup>c</sup>
142		N/A	0	Y
143		N/A	0	Y
144		N/A	24	Y
145		N/A	0	Y
146		EDC	0	N
147		N/A	<i>m/z</i> only <sup>d</sup>	Y

<sup>a</sup> The coupling condition that led to the highest product purity that was used for screening against *EcDsbA*. <sup>b</sup> LCMS product purity of the UV/Vis peak with the correct corresponding *m/z* of the desired product. <sup>c</sup> Binding results determined by <sup>1</sup>H-<sup>15</sup>N HSQC <sup>d</sup> “*m/z* only” denotes the corresponding desired product *m/z* was obtained but no or overlapping UV/Vis peaks were present at this retention time.

It was hypothesised that the amide coupling reagents, activated intermediates and potential by-products of the amidation reactions may have bound to *EcDsbA* or obscured the covalent modification by the crude products. In order to address this EDC, HATU, acetic acid and benzylamine were used to generate “faux” reaction mixtures using the same conditions as the parallel synthesis of the crude warhead library.

The following combinations were used:

1. Amine parent **110** with coupling agent (EDC or HATU) to determine if the coupling reagents bound to *EcDsbA* or interfered with the binding of amine **110**.
2. Acetic acid and coupling agent (EDC or HATU) to determine if the activated acid intermediates would covalently modify nucleophilic residues of *EcDsbA*.

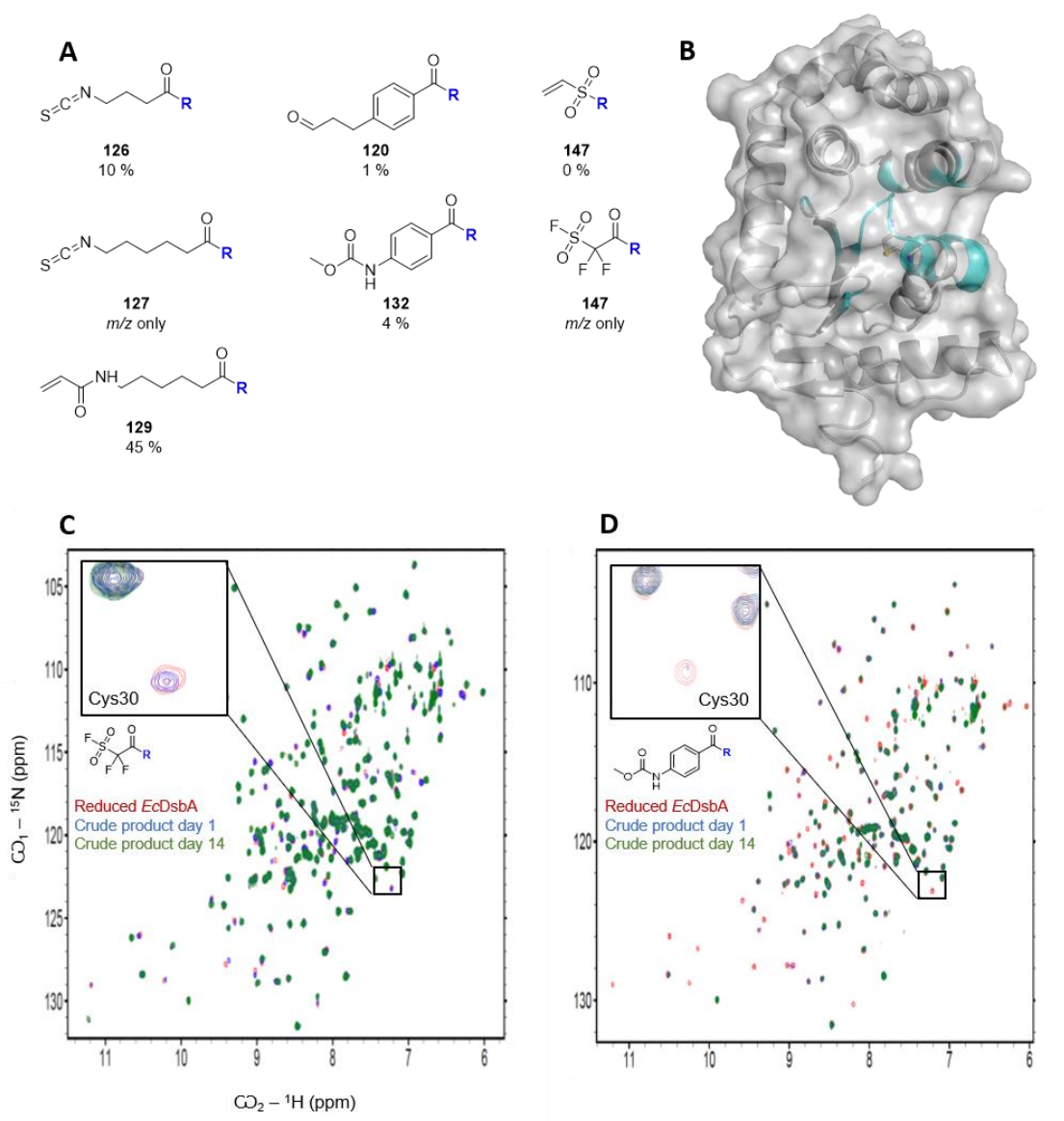
3. Acetic acid, benzylamine and coupling agent (EDC or HATU) to determine if any by-products generated by the consumption of the coupling reagents would interfere with binding.

#### 4.5 Crude parallel synthesis library screen

$^1\text{H}$ - $^{15}\text{N}$  HSQC of U- $^{15}\text{N}$  EcDsbA (100  $\mu\text{M}$ ) were recorded after ~1.5 – 17 hours of incubations with each library crude product at 1 mM (with the assumption of 100 % reaction conversion). Initially the faux reactions were examined for any effects caused by the reaction reagents and by-products. It was found that EDC did not bind to the protein and the CSP in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were consistent with the non-covalent binding profile of the pure amine **110**. HATU was shown to bind weakly to the protein, producing some minor CSP, however, HATU did not cause any changes to peak intensity. Neither the HATU or EDC activated acids or potential by-products formed in the model amidations caused any further effects than those observed in the previous control mixtures. In light of this, analysis of the HSQC data for the crude products was compared to the controls which consisted of the parent amine **110** and the corresponding amide coupling reagent used in the reaction to account for any CSP arising due to the presence of HATU. This revealed that eight crude library products (maleimides **134** – **136**, alkyl bromides **143** – **145**, benzyl chloride **142** and isothiocyanate **125**) of the 13 warheads that showed covalent reactivity in the pure warhead HSQC screen also showed labelling in the synthesised crude fragment-warhead library. The crude library products derived from epoxides **137** and **138**, isocyanates **117** and **118** and acrylate **128** no longer formed covalent adducts with EcDsbA. This may be caused by degradation of the warhead during synthesis and misleading data from the crude product LCMS data which shows the desired products mass in all of these cases or reduced binding in the presence of the fragment.

In addition to those warheads that showed reactivity in the pure warhead screen, seven crude library products showed evidence of labelling the active site of EcDsbA where no reaction was observed in the pure warhead screen (Figure 4.6). These crude library products reduced the peak intensity of Cys30 or Cys33 by > 20 %. The majority of new warheads identified were compounds which were expected to hit without the

presence of the fragment. Acrylamide **129** and aldehyde **120** showed only minor decreases in the peak intensity of Cys33, however did not show any additional CSPs or reductions in peak intensity. Isothiocyanates **126** and **127** showed reduced intensity of both Cys30 and Cys33, as well as CSPs consistent with strong binding to the hydrophobic groove.

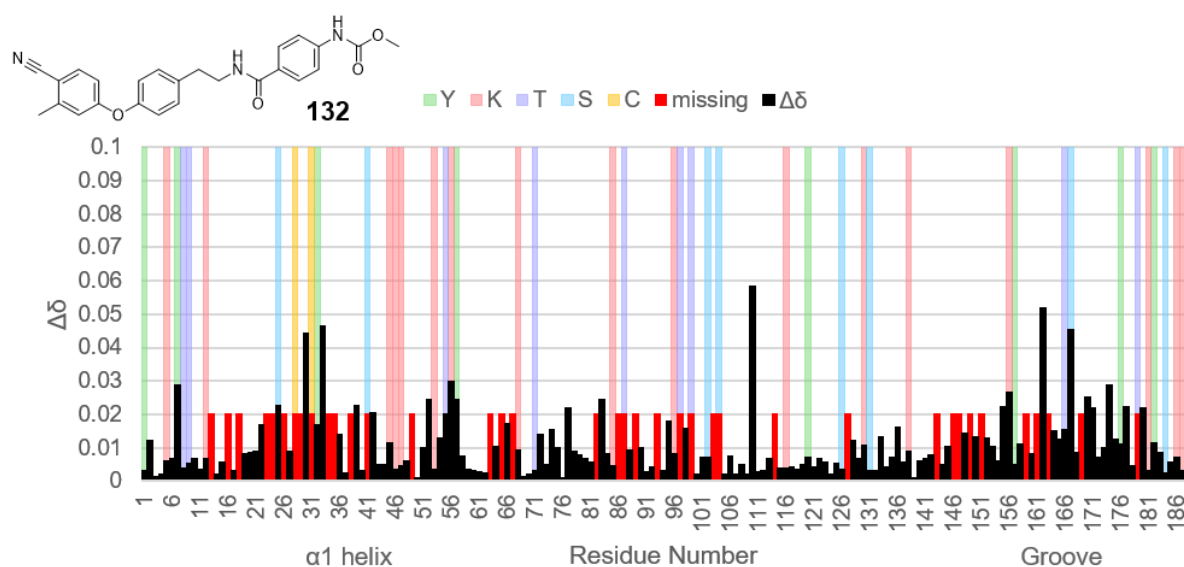


**Figure 4.6:** Crude reaction mixture screening results. A) Warheads which only labelled when conjugated to fragment binder, reaction purity as determined by LCMS. B) Residues whose peaks broadened in the spectra of sulfonylfluoride **147** and carbamate **132** (cyan cartoon). C)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of crude product of sulfonylfluoride **147** with reduced  $^{15}\text{N}$  EcDsbA. Expansion shows contours for faux reaction with no ligand (red), initial timepoint of crude product (blue) and 14-day timepoint (green). D)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of crude product of carbamate **132** with reduced  $^{15}\text{N}$  EcDsbA. Expansion shows contours for faux reaction with no ligand (red), initial timepoint of crude product (blue) and 14-day timepoint (green).



The two warheads which showed the most striking changes in their HSQC spectra after coupling to the diaryl ether precursor were the difluoromethyl sulfonyl fluoride **147** and phenyl-linked methylcarbamate **132** (Figure 4.6c, d). Both crude products caused peak broadening in not only Cys30 and Cys33, but also multiple residues adjacent to the active site in a similar profile to that observed in the covalent modification by pure warheads (Figure 4.6b). The magnitude of the CSPs observed for residues in the hydrophobic groove was increased, however many of the peaks were no longer in fast exchange and as a result could not be assigned in the spectrum. To confirm that the crude covalent warheads formed stable covalent complexes and were specific to Cys30, the crude products were screened again after 14 days of storage at 4 °C.

Five of the seven warheads that were found only to label *EcDsbA* in the crude product screen and not the pure warhead library screen (aldehyde **120**, isothiocyanates **126** and **127**, vinyl sulfone **139** and sulfonyl fluoride **147**), showed additional peak broadening and CSPs at the 14-day timepoint. This was interpreted as potential evidence of a secondary reaction with the protein. Further reduction in the intensity of the HSQC peaks for active site residues was often accompanied with instances of broadening of residues in other regions of *EcDsbA*. These included residues located close to or within the acidic patch of *EcDsbA*, as well as in two regions found far away from the expected binding site at the active site. Residues Asn62, Asp71 and Glu139 are situated along the domain interface on the  $\alpha 2$  and  $\alpha 6$  helices and loop connecting the  $\beta 3$  sheet and  $\alpha 2$  helix. The remaining residue, Asn156, is found on the loop connecting sheets  $\beta 4$  and  $\beta 5$ , adjacent to the C terminus of the protein. These two areas were where many residues showed dynamic characteristics in previous  $^1\text{H}^{\text{N}}$  CPMG-RD NMR experiments (Figure 1.26) <sup>(233)</sup> and these dynamics may account for covalent labelling of the residues over time.



**Figure 4.7:** Broadening and CSP histogram of **132**. Chemical shift perturbations of reduced *EcDsbA* caused by the presence of **132** as a function of residue. CSP (black), missing or unassignable residues (dark red), with nucleophilic residues Y (green), K (pink), T (purple), S (blue) and C (yellow) highlighted. With the active site helix ( $\alpha 1$  helix) and the hydrophobic groove labelled beneath the residue numbers.

Acrylamide **129** and carbamate **132** gave  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra which were consistent with stable single site covalent modification. Carbamate **132** was considered the most interesting compound to be identified in this screen as large changes in the original spectrum were observed (in comparison to acrylamide **129**) and the CSPs were unchanged after 14 days. The residues that had broadened beyond detection or had moved so far that they were no longer able to be assigned in the sample treated with crude product **132** were almost exclusively located in or around the active site (Figure 4.7). This indicated that the covalent complex was not being turned over as seen with previous fluoromethylketone (FMK) compounds as it formed a stable adduct that was unchanged over 14 days, and its reactivity was greatly enhanced by coupling to the diaryl ether fragment. Together these observations suggested that the carbamate was likely to be labelling *EcDsbA* in a site-specific manner. The promising results of this crude library screen were to be validated with resynthesised and purified compounds to confirm and fully characterise the compounds and behaviour.

## 4.6 Purified compound validation

We set out to validate the covalent modification of *EcDsbA* by carbamate **132** through batch synthesis and retesting the pure compound. We also wanted to explore the effects of yields on the kinetics of covalent modification by synthesising isothiocyanate **125**, the effects of reaction purity on binding by synthesising alkyl bromides **143** – **145**, and the effects of linker composition and fragment recognition by synthesising methyl carbamate **131** and a phenethyl coupled warhead **148** respectively.

### 4.6.1 Selection of compounds for pure compound screening

#### 4.6.1.1 Isothiocyanate warhead

The short isothiocyanate **125** was chosen to test the effects of yield in the parallel synthesis. When the isothiocyanate **87** was initially tested in the pure warhead screen, it was found to label *EcDsbA* quickly and lost amide signals in the <sup>1</sup>H-<sup>15</sup>N HSQC suggesting it had caused extensive labelling or precipitation of the reduced protein. The corresponding crude product **125** initially showed evidence of non-covalent interactions, however the minor reduction of the cross-peak intensity for the active site cysteines suggested only low levels of covalent labelling. However, changes observed in the HSQC spectrum of *EcDsbA* after 14 days incubation with isothiocyanate **125** were more consistent with those produced by isothiocyanate **87** in the original pure warhead screen. Evaluation of the crude products by LCMS indicated a product purity of 2 %, suggesting that low conversion could have affected reaction kinetics in this case.

#### 4.6.1.2 Alkyl bromide warhead

The alkyl bromides were chosen to examine the effects of interference due to the presence of by-products in the reaction mixtures. The crude mixtures **143** – **145** generated with each of the alkyl bromides produced much larger changes in the HSQC spectrum of *EcDsbA* in comparison to the pure uncoupled warheads **106** – **108**. The HSQC profile of the reaction mixtures saw large global shifts and loss of amide resonances which was interpreted as protein precipitation or unfolding. Two

of the expected products in the crude reaction with the alkyl bromides could not be verified from the LCMS data as no peaks with the correct  $m/z$  for **143** and **145** were identified and there were no new UV peaks observed. Therefore, changes observed in the HSQC spectra in the presence of the crude products could have resulted from the protein interacting with a reaction by-product. Alternatively, the warheads may have been present but were too reactive or unstable for detection in the reaction assessment by LCMS and the loss of HSQC signal was caused by the increased molecular recognition of the diaryl ether fragment.

#### 4.6.1.3 Carbamate warhead

Carbamate **132** which bound as a crude fragment but not in the pure warhead screen as carbamate **94** was selected for synthesis. The methyl carbamate warhead containing a methylene linker **131**, which did not show any evidence of binding in the pure warhead screen or as a crude product was also selected. This second methyl carbamate-containing compound **131** was resynthesised to ensure that this compound had not given rise to a false negative in the crude product screen. It also provided a means to assess the importance of the linker in the reactivity of the fragment-coupled compound. One additional methyl carbamate analogue was also synthesised. In this analogue the 2-methyl-4-phenoxybenzonitrile section of the diaryl ether fragment was replaced with a non-substituted benzene ring to generate carbamate **148**. This compound was synthesised to determine the contribution of the non-covalent interactions made by the diaryl ether fragment to the reactivity and recognition of the compound.

#### 4.6.2 Synthesis and compound solubility

The analogues were all synthesised using similar conditions to those which gave the best purity in the microscale reactions by LCMS. Nucleophilic substitution of acyl chlorides was conducted in DCM with triethylamine to afford the alkyl halides **143** – **145** in moderate yields of 34 – 68 %. Isothiocyanate ester **125** was deprotected by lithium hydroxide *in situ* and this compound, as well as carbamates **131**, **132** and

**148**, were coupled using standard HATU coupling conditions and obtained in moderate to good yields of 56 – 86 %.

Before further testing, the aqueous solubility of these compounds was determined by semi-quantitative NMR analysis. Solubility was measured by recording 1D <sup>1</sup>H-NMR spectra for each compound at a concentration of 1000, 500, 250, 125 and 62.5 μM, respectively, in a buffer containing 2 % *d*<sub>6</sub>-DMSO, as used in the crude mixture screen. Compounds which showed the expected doubling in intensity without any change in the chemical shift of the resonances were considered to be soluble and not aggregating <sup>(259)</sup>. The compound with the highest solubility calculated from this analysis was the methylene carbamate **131**, which was determined to be soluble at a concentration of 250 μM. Increasing the amount of organic solvent to 5 % aided solubility of some of the compounds, however only minor increases in solubility were observed. To compensate for the poor solubility, the concentration of *EcDsbA* used in the assays was decreased to 50 μM in an effort to maintain an excess of the covalent fragment in the mixture.

The solubility of the most promising compound, carbamate **132** was the lowest of all the compounds tested. The apparent solubility was below 20 μM in 5 % *d*<sub>6</sub>-DMSO, however, a solvent screen with this compound revealed that the highest aqueous solubility was observed in the presence of 5 % *d*<sub>7</sub>-DMF giving ~30 μM. Due to the low solubility of this compound it was not possible to generate a mixture where the covalent fragment was in excess without dropping the *EcDsbA* concentration below what was required to generate sufficiently sensitive HSQC data to allow validation by NMR. Therefore, alternative analytical strategies that would tolerate lower compound concentrations would need to be explored for this compound. Nonetheless, all other diaryl ether analogues were validated by <sup>1</sup>H-<sup>15</sup>N HSQC.

**Table 4.2:** Aqueous solubility of pure compounds estimated from semi-quantitative 1D <sup>1</sup>H NMR

Compound	Estimated solubility (μM)	Solvent
<b>125</b>	< 125	5 % <i>d</i> <sub>6</sub> -DMSO
<b>143</b>	250	5 % <i>d</i> <sub>6</sub> -DMSO
<b>144</b>	< 125	5 % <i>d</i> <sub>6</sub> -DMSO
<b>145</b>	< 125	5 % <i>d</i> <sub>6</sub> -DMSO
<b>131</b>	250	5 % <i>d</i> <sub>6</sub> -DMSO
<b>148</b>	125	5 % <i>d</i> <sub>6</sub> -DMSO
<b>132</b>	< < 125	5 % <i>d</i> <sub>7</sub> -DMF

#### 4.6.3 Pure compound screen

The pure isothiocyanate **125**, alkyl bromides **143** – **145**, and carbamates **131** and **148** were incubated with *EcDsbA* for 4 – 7 hours prior to acquisition of the <sup>1</sup>H-<sup>15</sup>N HSQC. These samples were stored at 4 °C for 14 days, retested by HSQC and compared to the initial spectra of the pure compounds.

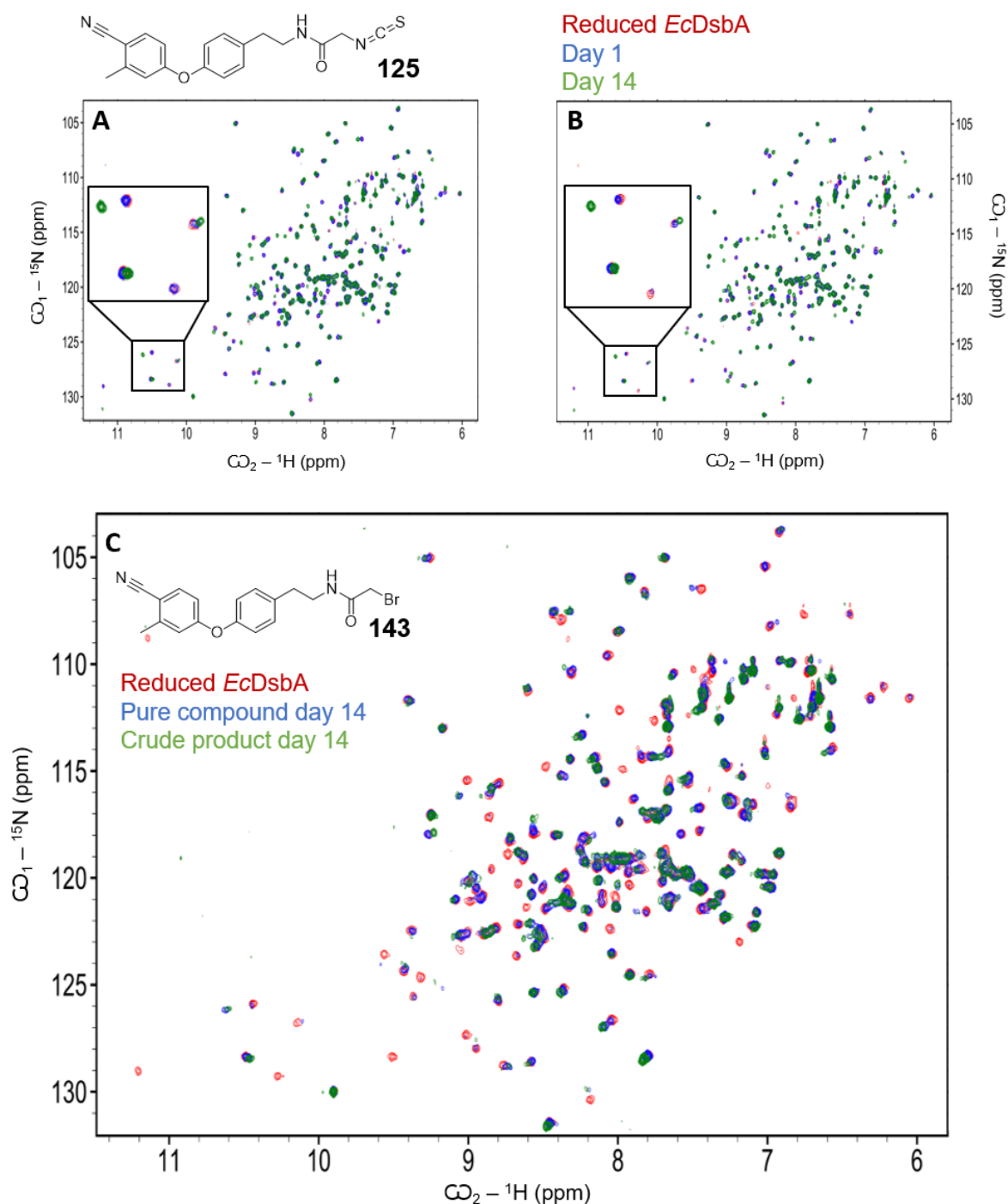
##### 4.6.3.1 Isothiocyanate and alkyl bromide warheads

Analysis of the HSQC data for the purified diaryl ether isothiocyanate **125** revealed a slow rate of reaction for this compound with *EcDsbA*. Only in the HSQC data at the 14-day timepoint was any significant CSP observed in the spectra, suggesting that the protein adduct was formed slowly, consistent with the original crude screen (Figure 4.8 a, b). This indicates that the reaction kinetics observed for this compound were independent of the low product purity seen in the crude mixture and are more likely controlled by the modulation of the warhead placement and reactivity caused by the diaryl ether.

Comparison of the changes observed in the HSQC spectra of *EcDsbA* purified diaryl ether-linked alkyl bromides **143** – **145** with their corresponding crude products revealed similar patterns of CSP were observed in each case. However, in the spectra

for the 14-day timepoint the quality of the spectra for the purified compounds was improved. More of the peaks in the HSQC spectrum could still be observed in the data (Figure 4.8c). This suggests that by-products and/or reagents in the crude reaction mixture were contributing to the degradation in the quality of the HSQC data.

The validation of the purified diaryl ether linked isothiocyanate **125** and bromide compounds **143** – **145** as binding to *EcDsbA* highlights the potential for false negatives in the data analysis. For compounds with low reactivity and/or solubility such as the isothiocyanate, binding may not be evident unless long incubation times are employed. For compounds which are highly reactive, such as the alkyl bromides, it is possible that protein modification is possible, even if there is no evidence in the LCMS data for the formation of the desired product in the crude reaction mixture. This supports the use of data acquisition at two timepoints to monitor the reactivity of warheads and adduct stability. Furthermore, this suggests that the LCMS analysis should not be used to exclude compounds from testing or validation, rather it provides a means to evaluate the suitability of the conditions used for the synthesis of the coupled fragments. Where different synthetic strategies are used to generate the desired products (e.g. EDC vs HATU) the LCMS can indicate the reactions where products are observed with the highest purities. Even where the desired products are not observed in the LCMS, the data can be used to prioritise reactions for analysis by considering the reaction conditions which showed the highest success amongst the other analogues.



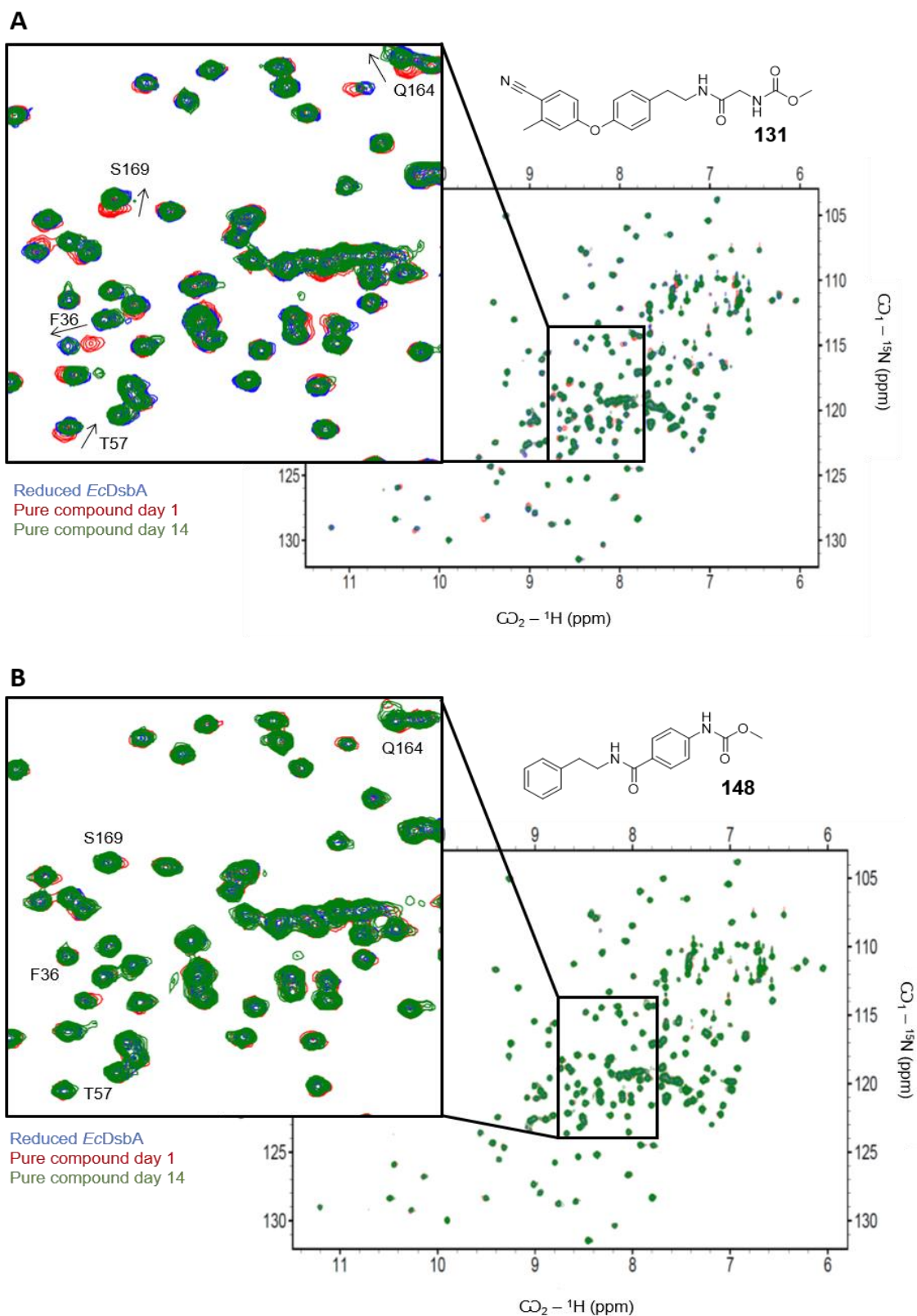
**Figure 4.8:** Validation of pure covalent compounds A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of crude product of isothiocyanate **125** with reduced  $^{15}\text{N}$  *EcDsbA*. Expansion shows contours for faux reaction with no ligand (red), initial timepoint of crude product (blue) and 14-day timepoint (green). B)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of pure isothiocyanate **125** with reduced  $^{15}\text{N}$  *EcDsbA*. Expansion shows contours for no ligand (red), initial timepoint of crude product (blue) and 14-day timepoint (green). C)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of crude product and pure alkyl bromide **143** with reduced  $^{15}\text{N}$  *EcDsbA* after 14 days. Contours for no ligand (red), 14-day timepoint of pure **143** (blue) and 14-day timepoint of **143** crude product (green).



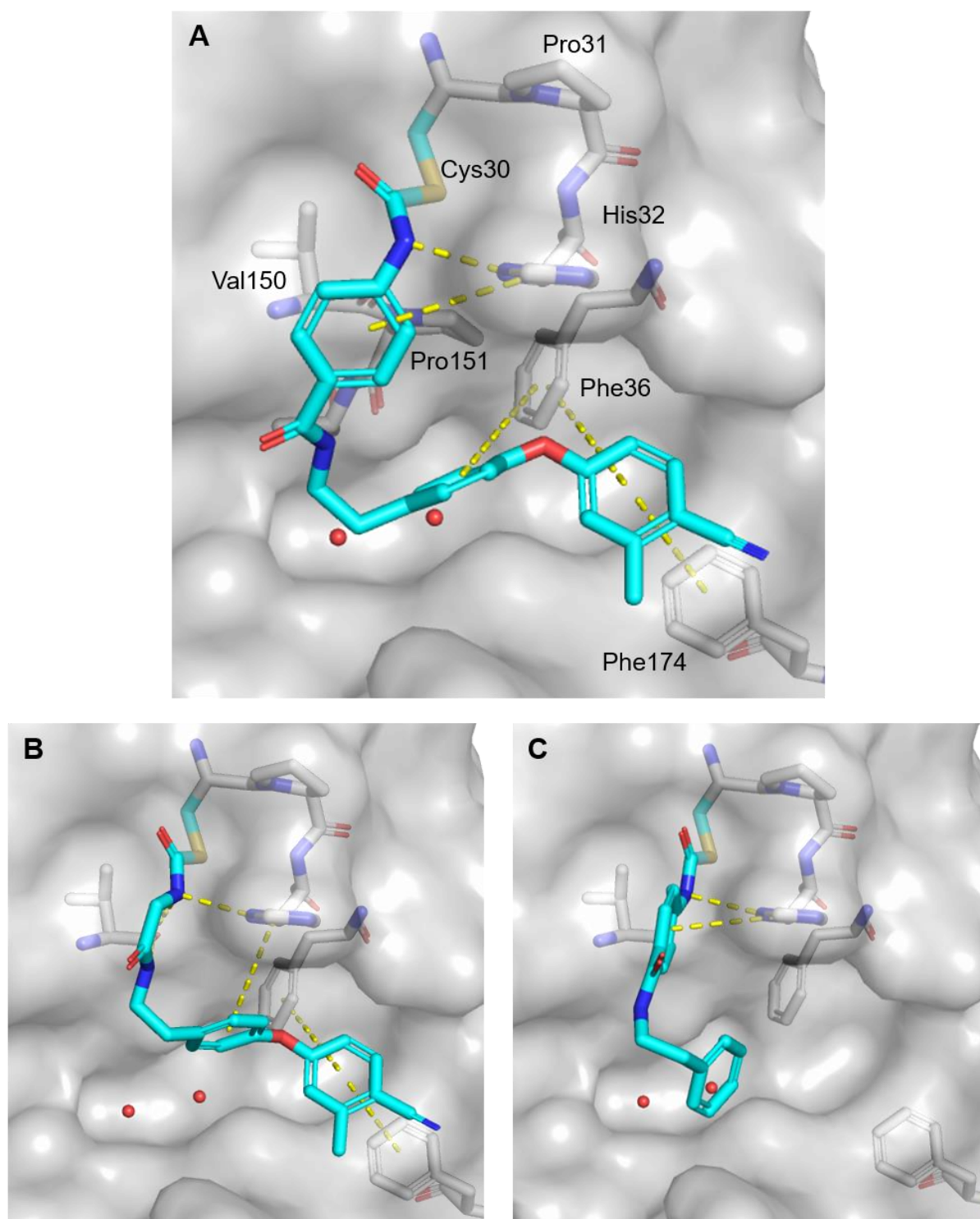
#### 4.6.3.2 Carbamate warheads

Neither methylene carbamate **131** nor the phenethyl carbamate **148** showed evidence of covalent modification of any of the nucleophilic residues in *EcDsbA* (Figure 4.9). The methylene carbamate **131** displayed small perturbations consistent with diaryl ether binding in the hydrophobic groove, however there was no evidence of covalent modification of either active site cysteine. This may be due to the decreased intrinsic reactivity of the methylene linked **131** in comparison to the phenyl linked **132** which may be enhanced by the electron withdrawing effects of the benzene ring in the linker. No perturbations were caused by the phenethyl carbamate **148** in the hydrophobic groove, and this compound showed no sign of non-covalent or covalent interactions to any other region of the protein.

As the phenyl-linked carbamate **132** could not be tested by  $^1\text{H}$ - $^{15}\text{N}$  HSQC a covalent docking study was conducted with **131**, **132** and **148** for further potential insight into the difference in reactivity between analogues (Figure 4.10). These studies identified hydrogen bonds to the amide of Val150 and the sidechain of His32 with the S-thiocarbamate groups of each of the carbamate analogues. They also suggested that the aromatic ring of carbamates **132** and **148** make a  $\pi$ - $\pi$  stacking interaction with the sidechain of His32, an interaction not achievable by the methylene linker of carbamate **131**. However, it supported that phenyl-linked carbamate **132** and the methylene carbamate **131** would have retained the binding pose of the diaryl ether fragment in the hydrophobic groove, whereas phenethyl **148** would have lost all of these non-covalent interactions. It is possible that the loss of non-covalent interactions in the linker of methylene carbamate **131** and the non-covalent interactions within the hydrophobic groove for phenethyl carbamate **148** causes these compounds to no longer react with Cys30 or Cys33, consistent with the lack of reactivity observed in the assays.



**Figure 4.9:** A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of pure methylene carbamate **131** with reduced  $^{15}\text{N}$  *EcDsbA*. Expansion shows contours for no ligand (red), initial timepoint (blue) and 14-day timepoint (green). B)  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of pure phenethyl carbamate **148** with reduced  $^{15}\text{N}$  *EcDsbA*. Expansion shows contours for no ligand (blue), initial timepoint (red) and 14-day timepoint (green).



**Figure 4.10:** Covalent docking of carbamate functionalised compounds against reduced *EcDsbA*. *EcDsbA* (PDB ID: 4TKY) shown as grey surface, interactive residues shown as grey sticks, structural waters shown as red spheres, interactions shown as yellow dashes, docked pose of inhibitors shown as cyan sticks. A) phenyl carbamate hit **132**, B) methylene carbamate **131** C) phenethyl carbamate **148**.

## 4.7 Conclusions and future directions

In this chapter, a library of 35 electrophilic covalent warheads was designed for coupling to a known fragment binder via parallel microscale amidations. A  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR assay which tested crude product fragment linked warheads was implemented in an attempt to identify mildly reactive covalent warheads which labelled reduced *EcDsbA* in a stable and selective manner. This was achieved by analysing the enhanced labelling of reactive functional groups upon the conjugation to a known fragment binder, and the progression of the reaction with *EcDsbA* over an extended period of time. The pure carbamate warhead **94** did not bind in the original pure warhead screen, however when coupled to the diaryl ether fragment the crude carbamate product **132** showed large chemical shift perturbations, broadening of the active site cysteines and remained stable for 14 days. Six analogues and one control compound were resynthesised and tested as purified compounds by HSQC. Upon retesting the compounds, it was found that the crude product results were able to be confirmed with resynthesised and purified samples. However, the potentially selective and stable carbamate **132** showed poor aqueous solubility and was unable to be retested by  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR.

Although this study encountered a number of challenges in validation of the initial data, the implementation of this crude reaction screening of electrophilic fragments was able to quickly identify a promising potentially selective and stable warhead for further optimisation. The design of this warhead library was target agnostic and has the ability to be applied to multiple targets with nucleophilic residues adjacent to the fragment binding site for the discovery of covalent inhibitors of other protein targets.

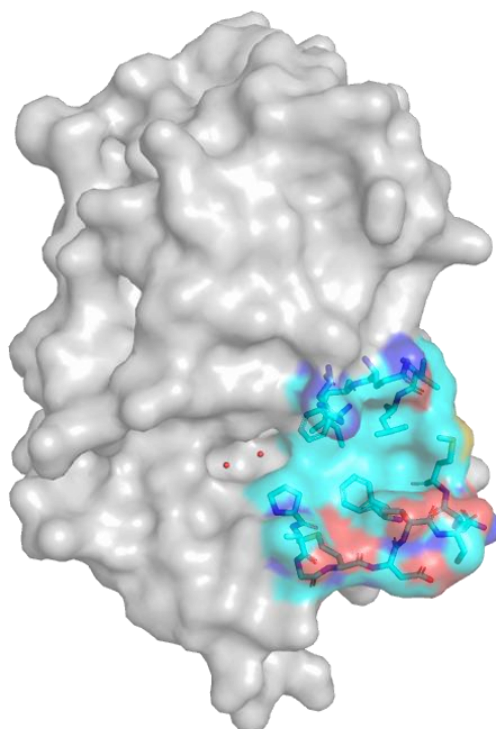
# **Chapter 5:**

## **Protein hot spot identification by organic solvent screening**

In chapter 2, 3 and 4 compounds that can bind covalently to *EcDsbA* were investigated. The unique reactivity of *EcDsbA* active site Cys30 thiol provides an opportunity as well as a challenge for the development of covalent inhibitors. These efforts identified the electrophilic carbamate warhead as a promising candidate that could be used in further covalent inhibitor development. While development of covalent inhibitors can be an attractive strategy for increased potency not all targets are amenable to covalent inhibitor development due to the requirement of a suitable, often nucleophilic, residues in or adjacent to a binding site. In addition, development of covalent inhibitors also generally requires optimisation of the non-covalent interactions to enhance kinetics and selectivity for the desired target. Hence, in parallel to investigating covalent inhibitors we investigated other strategies to provide information for improved non-covalent inhibitor design. Chapters 5 and 6 present efforts to employ existing and novel techniques for identification and characterisation of interaction hot spots on *EcDsbA*.

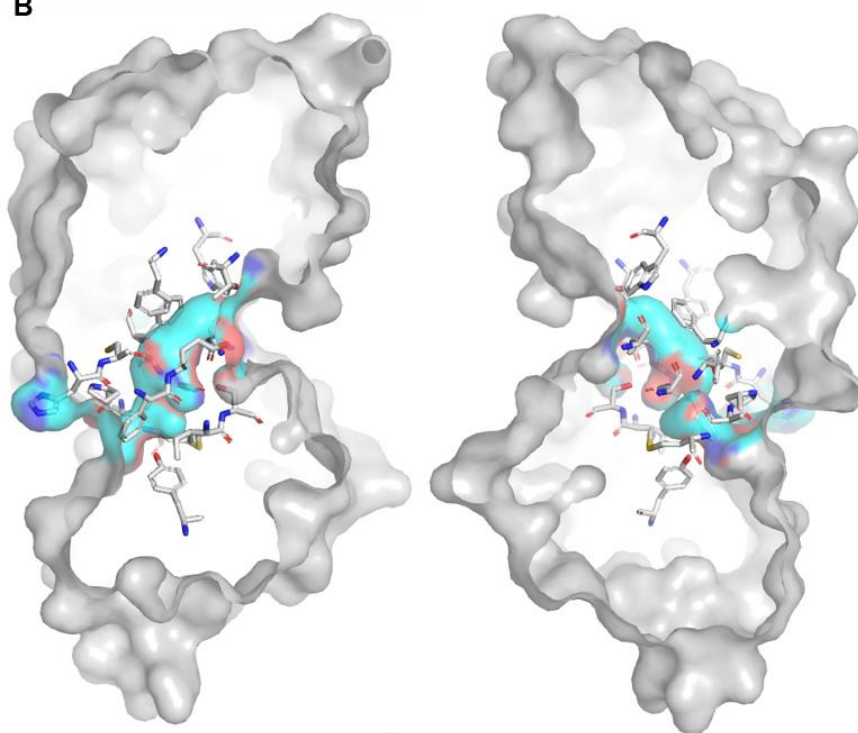
Drug discovery efforts for targets that make protein-protein interactions (PPIs) are often difficult due to the fact that the binding site surface area is often large and commonly devoid of well-defined cavities and proximal hot spots required for high affinity small molecule inhibitor binding. DsbA proteins show low specificity for endogenous peptide substrates <sup>(201, 202)</sup> and non-cysteine peptides have been shown to bind with poor affinity (explored in Chapter 2). NMR studies of *EcDsbA* have also shown significant protein dynamics and flexibility in and around residues that form the main hydrophobic groove binding site <sup>(233)</sup> (Figure 5.1). These observations are further supported by the identification of a previously unidentified internal cryptic pocket <sup>(233)</sup> (Figure 5.2). Furthermore, previous attempts to develop covalent inhibitors of *EcDsbA* resulted in the labelling of the buried active site cysteine residue, Cys33, highlighting the presence of dynamics <sup>(251, 254, 255)</sup>. Finally, crystal structures obtained within the group have demonstrated significant loop and side chain flexibility during medicinal chemistry efforts to develop inhibitors <sup>(233, 253)</sup>. These dynamic conformational changes and motions of the protein makes compound optimisation a challenge due to entropic penalties during compound binding as well as difficulty employing structure-based design to predict the binding site conformation for novel ligands.

A



Hydrophobic groove

B



Internal cryptic pocket

**Figure 5.1:** Known small molecule binding sites of *EcDsbA*. A) Hydrophobic groove and B) internal cryptic pocket (shown with surface highlighted as cyan carbons, red oxygens, blue nitrogens, yellow sulfurs) positioned on *EcDsbA* (PDB ID: 1FVK, shown as grey sticks, grey surface).

Structure-based design using computational docking has previously been used in the design of higher affinity analogues of various small molecule scaffolds. Fragments which bind at the main hydrophobic groove of *EcDsbA* have been well characterised and the structures obtained have been used to design analogues which can form a multitude of new interactions. However, these compounds, when synthesised and tested, failed to improve affinity significantly. For example, compounds which formed hydrogen bonds to Val150 and Pro151 in the *cis*-Pro loop of *EcDsbA*, which were predicted by docking and confirmed by X-ray crystallography only achieved small improvement in affinity (< 5-fold) <sup>(233, 253)</sup>. This suggests that computational docking studies using flexible ligands and a semi-rigid protein conformation (some side chain flexibility is allowed) for *EcDsbA* does not adequately model which interactions can lead to increased binding energy. We therefore sought to supplement docking with additional experimental information on interaction hot spots to prioritise key interactions that may be targeted for improved affinity inhibitor design.

Organic solvents have been used as chemical probes to identify binding hot spots and favourable residue interactions due to their small size, which enables them to provide a good coverage of very small compound chemical space <sup>(130-132)</sup>. As discussed in the introduction the probability of finding a complementary interaction of a compound and target are significantly increased when compound size and complexity are decreased <sup>(2)</sup>. Organic solvents commonly contain single functional groups that are representative of the functional groups in lead-like compounds. They therefore have the potential to identify the position and desired interaction orientations of hot spots on the target that can be used to gain key interactions in the design of high affinity binders.

This idea was first described as a computational method known as multiple copy simultaneous search (MCSS) <sup>(130)</sup> and was shortly followed by experimental techniques developed for X-ray crystallography <sup>(131)</sup> and NMR spectroscopy <sup>(132)</sup>. Clustering of organic solvents in localised areas are hypothesised to be indicative of protein hot spots and key interactions can be determined by the analysis of crystal structures or HSQC amide perturbations.

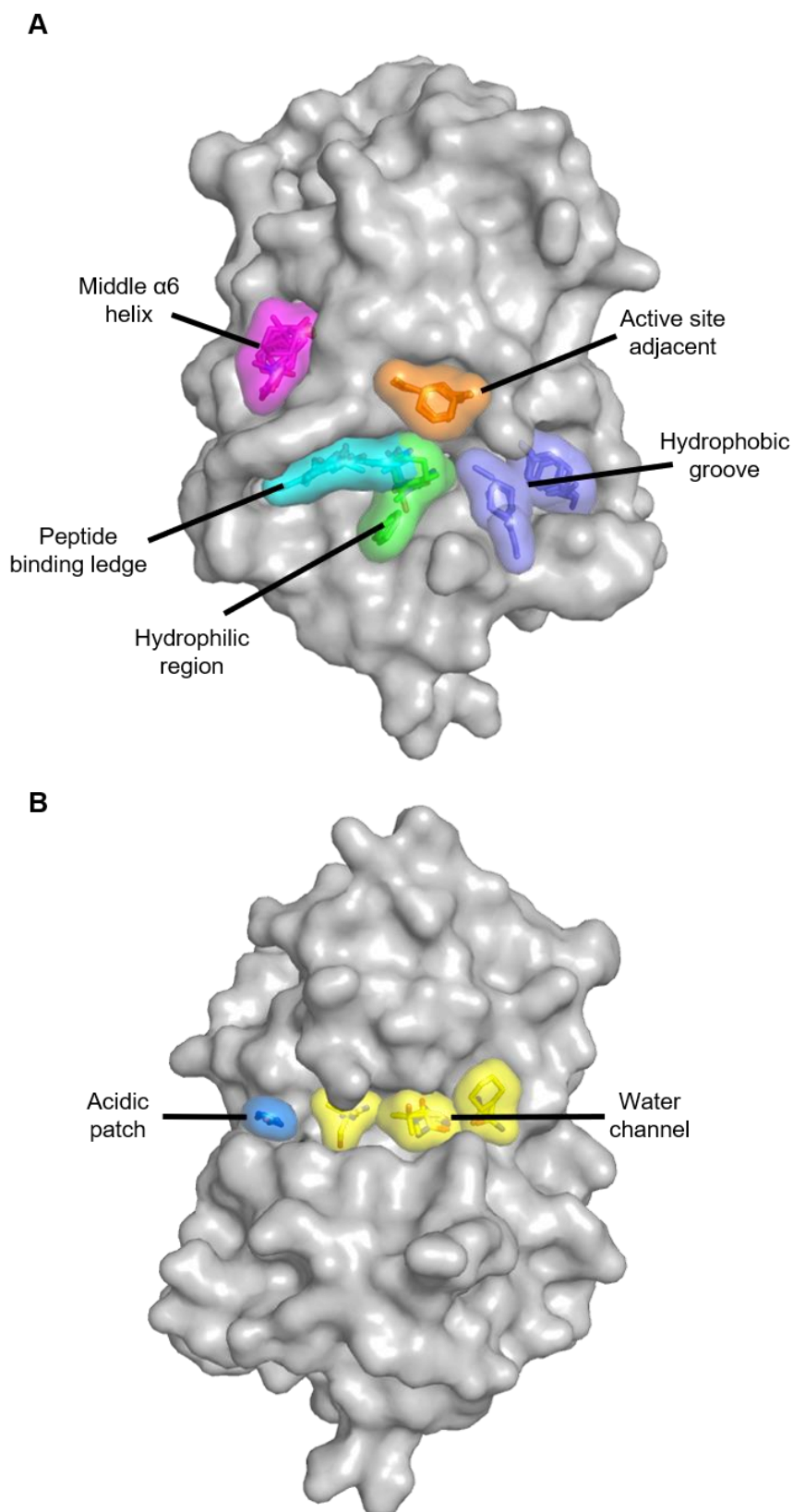


## 5.1 Computational binding site predictions

More recently computational methods such as FTMap <sup>(270, 271)</sup> and DoGSiteScorer (DoGSS) <sup>(272)</sup> have expanded on algorithms similar to MCSS. FTMap emulates the solvent screening technique and provides poses of each probe allowing the key interactions to be identified. DoGSiteScorer provides cluster surfaces that define the shape of the binding pockets without predicting exact interactions. Each of the methods discussed above was conducted on *EcDsbA* (chain A monomer from the 1FVK PDB structure) to predict binding site hot spots <sup>(243)</sup>.

### 5.1.1 Analysis of FTMap clusters

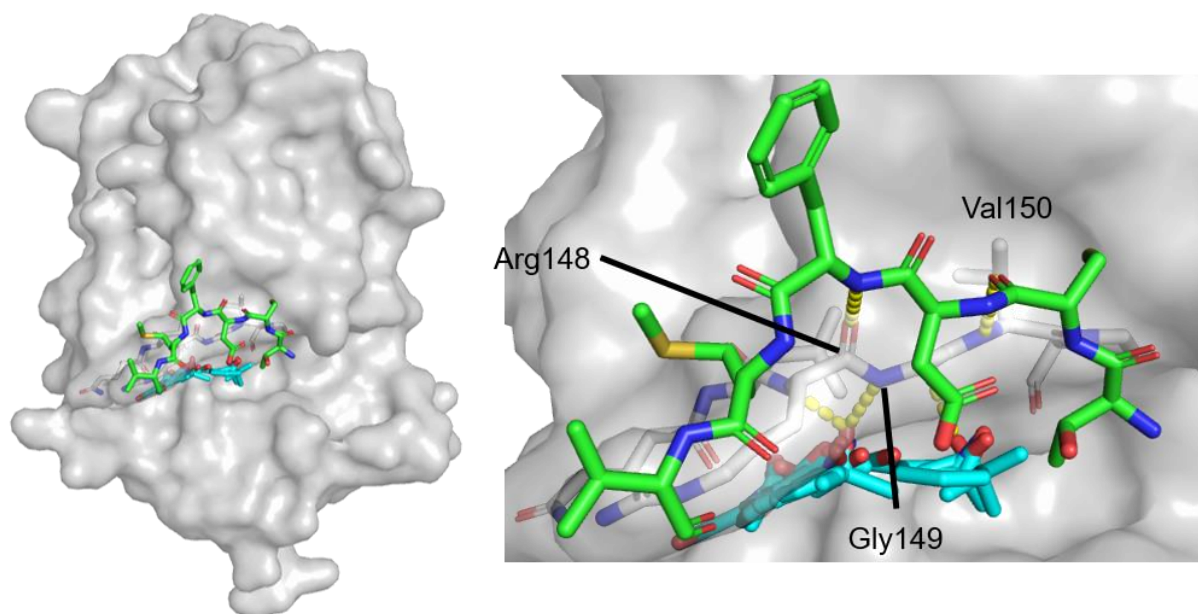
The FTMap algorithm uses 16 small molecule probes; ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide and *N,N*-dimethylformamide (DMF) to predict hot spots. The presence of multiple highly populated and adjacent hot spots has been linked to the ability to develop high affinity drug-like small molecules <sup>(270, 271)</sup>. 12 cluster sites were identified on the protein surface of *EcDsbA*, characterised as seven binding hot spots (Figure 5.2).



**Figure 5.2:** Results from FTMap calculations. Predicted solvent clusters (coloured sticks, surface) bound to *EcDsbA* (PDB ID: 1FVK, shown as grey surface) on A) the front catalytic face and B) the back of the protein.

### 5.1.1.2 Clusters around the substrate peptide binding site

The highest ranked FTMap cluster shows 14 probes bound below a ledge formed from loop residues Val145 – Pro151. These residues make up the loop that forms an antiparallel  $\beta$  sheet interaction with the *EcDsbB* loop (PDB ID: 2HI7<sup>(222)</sup>) and substrate peptides, however, the peptide motifs do not occupy the space under the loop where the cluster is found, but rather bind across the top of the ledge (Figure 5.3). The solvent exposed surface of this ledge is highly polar and has few hydrophobic interactions. Non-polar side chains Leu147, Val150 and Pro151 are pointed inwards while the polar backbone and flexible side chains of Gln146, Arg148 and Gln160 point towards the surface. The hydrogen bonds suggested by the clusters are concentrated at the backbone amides of Arg148 and Gly149, but are also observed to the backbones of Val150, Pro151 and the sidechain of Gln160 which create the right and bottom of the FTMap cluster. To our knowledge no evidence of peptides or small molecules binding to the space under the loop residues has been observed, however analogues previously synthesised within the group in the benzyl thiazole series have made hydrogen bonding interactions to the backbone carbonyl of Pro151 similar to some probes in this cluster.

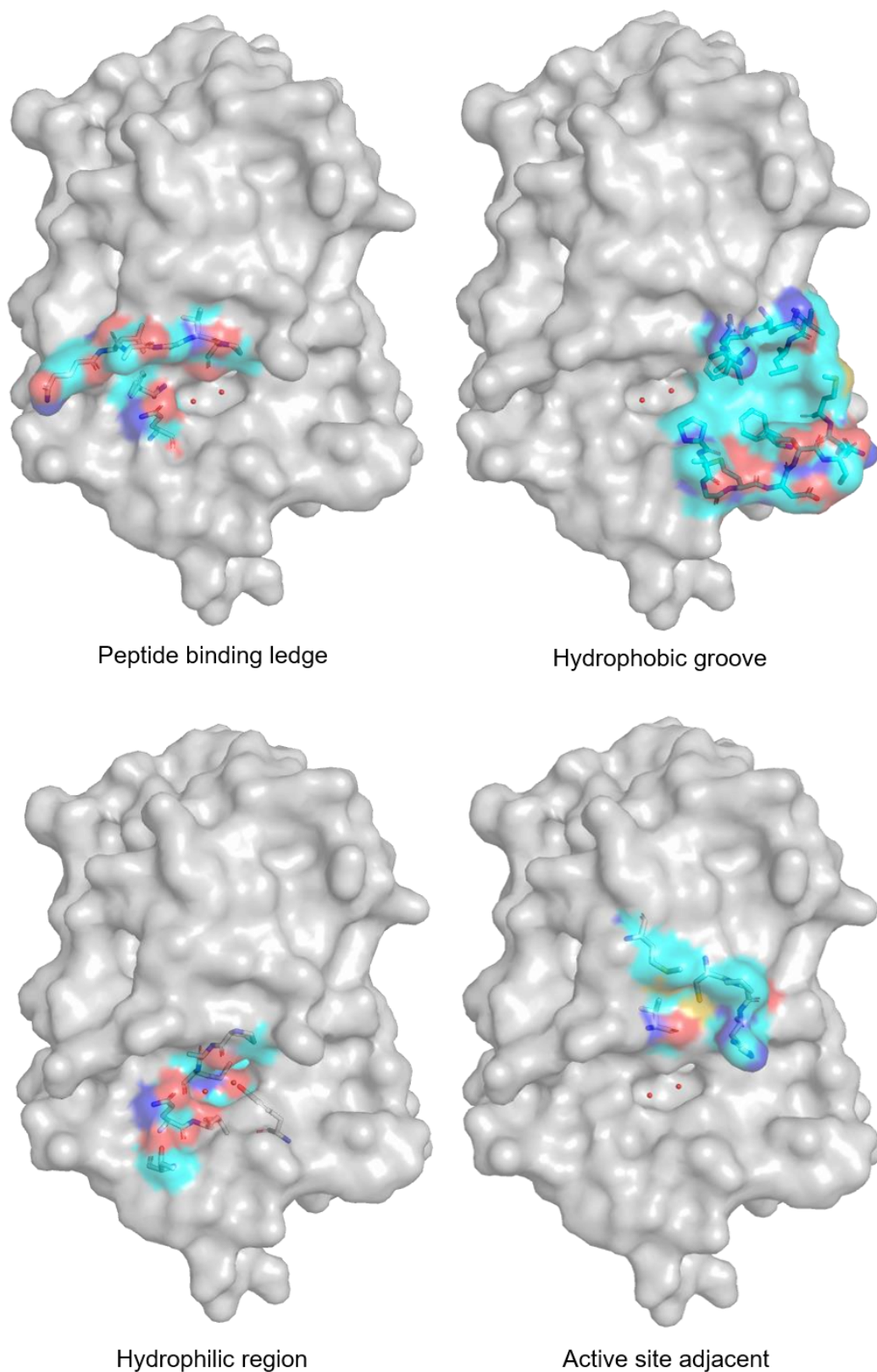


**Figure 5.3:** Overlay of the *EcDsbB* peptide (<sup>103</sup>TCDFMV<sup>108</sup>, green sticks) with the FTMap peptide binding ledge cluster (blue sticks) on the surface of *EcDsbA* (PDB ID: 2HI7, grey surface). Interactions are shown as yellow dashes and residues of *EcDsbA* labelled.

Four clusters of probes were also found in the hydrophobic groove where substrate peptides and *EcDsbB* loops form hydrophobic interactions. As the name suggests, the most common interactions found were hydrophobic interactions to Phe36, Leu40, Pro163, Met171, Phe174, and each were formed by at least 2 solvents. The only polar interactions that were formed were hydrogen bonds to His32 and the backbones of Gln164 and Met166 all found on the left-hand side of the groove proximal to the hydrophilic region of the substrate binding site. These clusters are ranked as 6<sup>th</sup>, 7<sup>th</sup>, 11<sup>th</sup> and 12<sup>th</sup> of the 12 clusters that were identified and have only a few probes per cluster. This is an unexpected result given that all fragments identified to bind to *EcDsbA*, with the exception of the isoxazole series, bind at this hydrophobic groove site.

A cluster is also observed between the hydrophobic groove and the peptide binding ledge, partially overlapping with the peptide binding ledge site. These residues make up what is referred to as the hydrophilic region of the substrate binding site (Figure 5.4). All previously solved high resolution crystal structures have identified two structural waters that make hydrogen bonds to Leu161, Tyr178 and Met153, casting doubt on more than half of the predicted probe binding poses which make similar interactions as the structural waters. FTMap removes all ligands and waters prior to probe sampling and minimisation. Solvents that are predicted to bind without displacing these structural waters make hydrogen bond donor interactions to the backbone of Gln160 and Leu161 or hydrogen bond acceptor interactions with the side chain of Thr10.

A small cluster of compounds (Figure 5.4) also bind at the active site Cys30 and adjacent residues Met64, His32 and Val150. The hydrogen bonds suggested are formed to His32 or Val150, supported by hydrophobic interactions with the sidechains of Met64 and Cys30. This small pocket is directly above the hydrophilic portion of the substrate binding site and together with the top of the peptide binding ledge, hydrophilic and hydrophobic portions of the groove form the main peptide substrate and *EcDsbB* binding site.



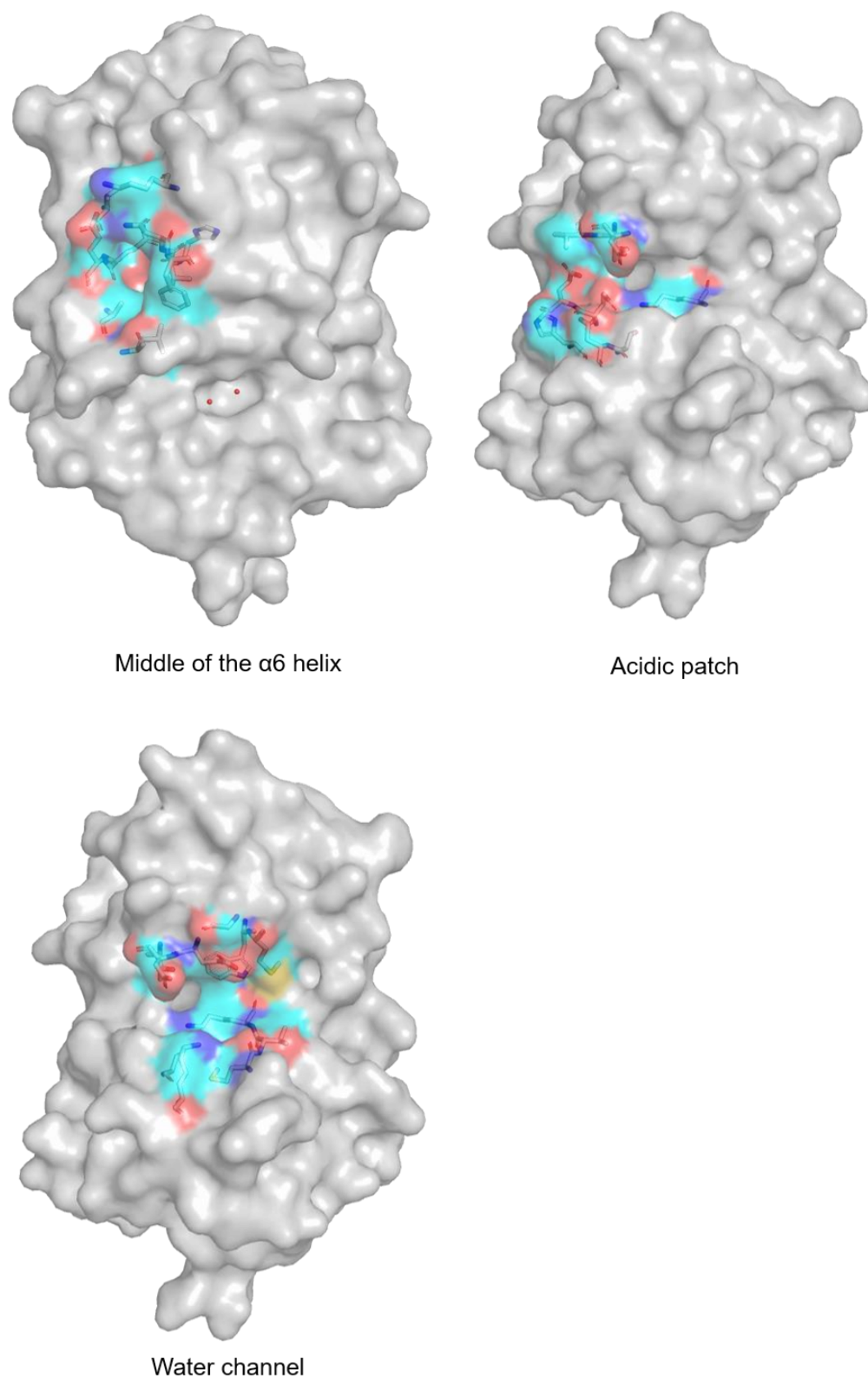
**Figure 5.4:** Binding hot spots identified by FTMap around the substrate binding site. Binding hot spot residues shown as cyan sticks and surface, structural waters shown as red spheres, mapped onto *EcDsbA* shown as grey surface (PDB ID: 1FVK).

### 5.1.1.3 Other FTMap clusters

The second highest ranked FTMap cluster is found on the front left of the protein along the  $\alpha 6$  helix which links the thioredoxin domain and the inserted  $\alpha$  helical domain approximately 9 Å away from the peptide binding ledge in a shallow pocket (Figure 5.5). This cluster makes conserved hydrophobic interactions with a shallow pocket formed by the side chains of Val61, Phe63, and Ala142 as well as two potential hydrogen bonds with a ring of polar atoms surrounding the shallow pocket (Asn62, Phe63, Glu139, Leu147 and Ala142). Similar to the area below the peptide binding ledge, to our knowledge, no small molecules have been found to bind here.

A known structural feature of *EcDsbA* is the acidic patch (Figure 5.5) found on the back of the protein, this patch is formed by the 3 carboxylate sidechains of Asp44, Glu37 and Glu38. Methylamine, urea and ethanol probes were found here making ionic and hydrogen bond donor interactions to Glu37 and Glu38 sidechain as well as the backbone of Glu37. Although this feature is known, only metals have been found to bind here previously.

A water channel runs across the back of the protein between the thioredoxin domain and  $\alpha$ -helical domain, extending the acidic patch around the back of *EcDsbA* (Figure 5.5). Three FTMap clusters are found in this channel spaced across the back of the protein. Several hydrogen bonds are formed to the top (Trp76, Met80, Glu85, and Thr89), as well as the bottom of the channel (Glu37, Lys55, Met56, Lys58 and Tyr59).



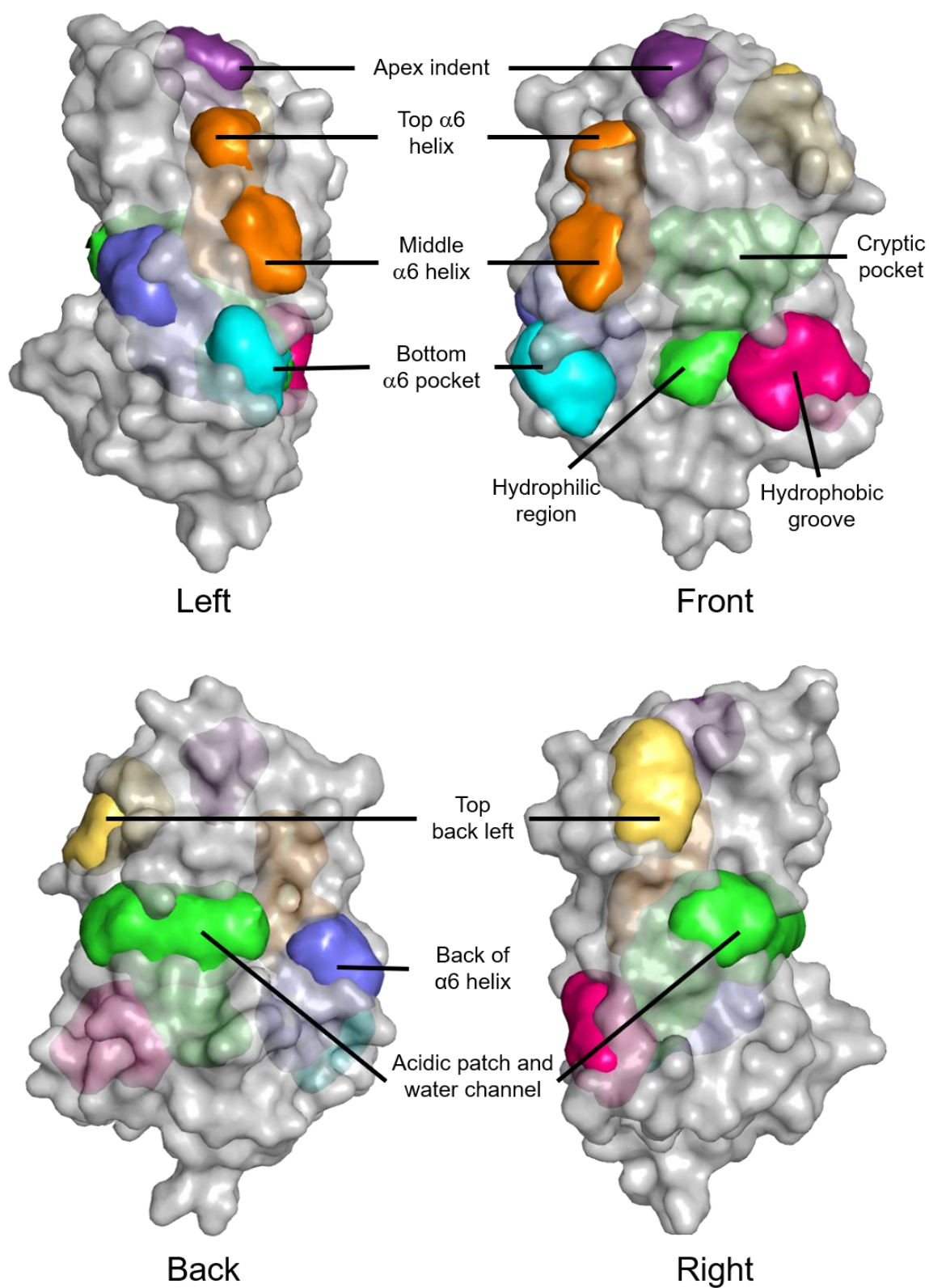
**Figure 5.5:** Binding hot spots identified by FTMap distant from the substrate binding site. Binding hot spot residues shown as cyan sticks and surface, structural waters shown as red spheres, mapped onto *EcDsbA* shown as grey surface (PDB ID: 1FVK).

Although a total of 12 clusters were identified by FTMap, these calculations notably failed to identify the internal cryptic pocket. Furthermore, the well-studied major binding site of *EcDsbA* (the hydrophobic groove) was ranked among as the least important clusters. Finally, interactions suggested by the probes in the hydrophilic region of the substrate binding pocket have been previously investigated. Analogues designed within the group which have formed these hydrogen bonds (confirmed to make these by X-ray crystallography) have yielded only small gains in affinity (2 – 5-fold).

### **5.1.2 Analysis of DoGSiteScorer simulation**

DoGSiteScorer (DoGSS) predicts binding pockets using a series of geometry, property and machine learning steps. Initially a grid is produced around the target, followed by a difference of Gaussians calculation to identify grid points where a sphere like probe may bind. These grid points are then clustered and merged to give potential binding sites and subsites as volumes, where they are then passed to a machine learning algorithm trained on pocket properties to identify druggable pockets. Potential binding sites are scored from 0 – 1 as undruggable to druggable. DoGSS does not identify exact interactions and a single pocket may be made up of multiple interaction hot spots. Despite this DoGSS was conducted to see if more traditional binding site identification algorithms could identify both the main hydrophobic groove binding site as well as the cryptic pocket, something FTMap was unable to do.



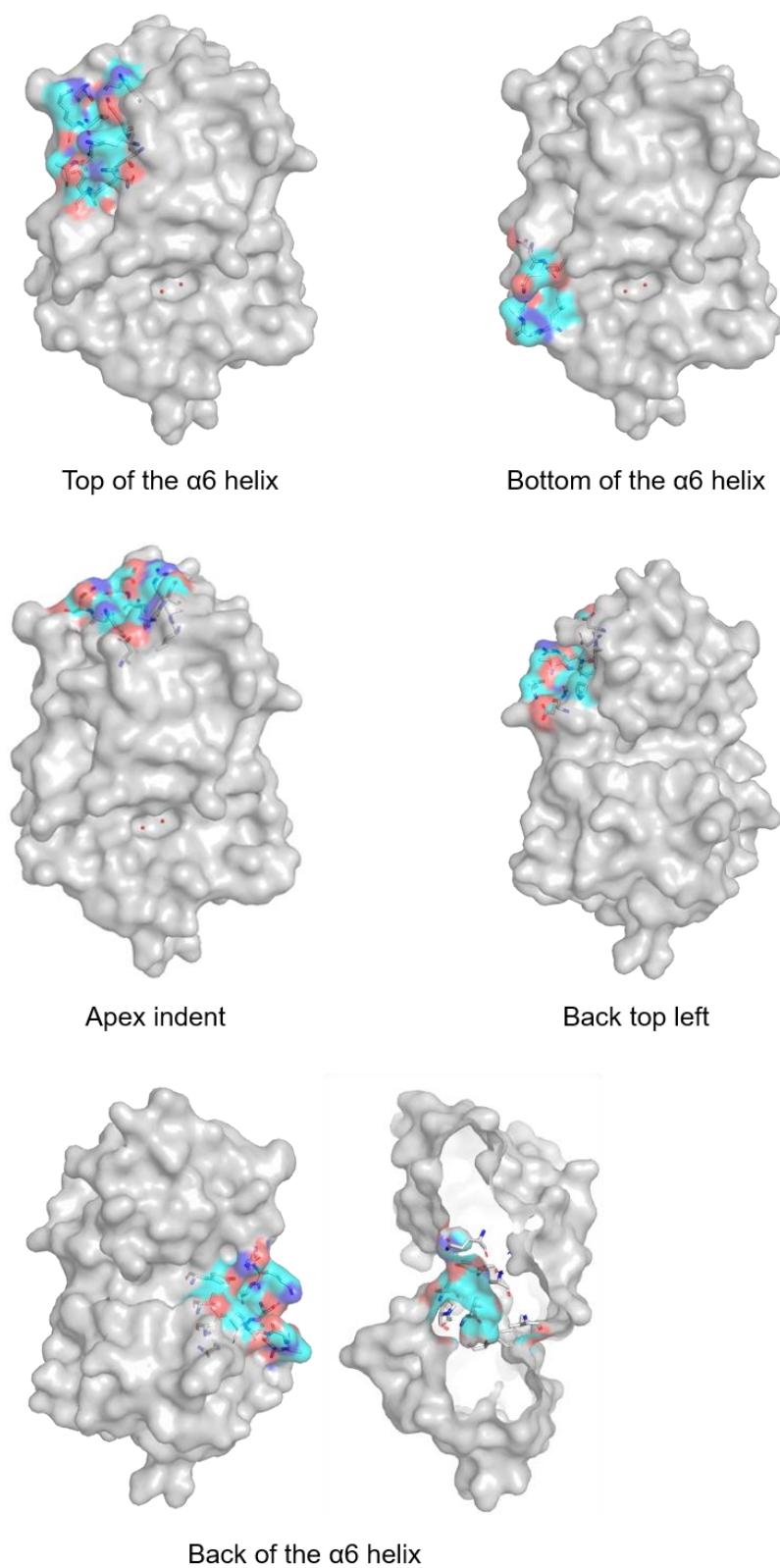


**Figure 5.6:** Results from DoGSiteScorer calculations. Predicted binding hot spots (coloured surfaces) mapped onto *EcDsbA* (PDB ID: 1FVK, shown as grey surface).

The highest ranked binding site from DoGSS was a binding site spanning the front and back of the protein made up of subpockets in the hydrophilic portion of the substrate binding site at the front of the protein, into the cryptic pocket and expanding out the back of the protein into an area spanning both the acidic patch and partially filling the water channel on the back of the protein (Figure 5.6). Similarly, the hydrophobic groove and the shallow hydrophobic pocket previously identified by FTMap on the front of the protein next to  $\alpha 6$  helix linking the thioredoxin and  $\alpha$  helical domains were identified. The hot spot along the  $\alpha 6$  helix was classified as one pocket, however, compared to the FTMap  $\alpha 6$  helix hot spot, it was extended upward and was divided into two subpockets (top and middle of the helix). These DoGSS pockets are all overlapping with clusters identified by FTMap, with the region at the peptide binding ledge the only FTMap cluster not covered by a pocket in DoGSS.

In addition to these pockets, four additional pockets were identified by DoGSS (Figure 5.7):

1. at the bottom of the  $\alpha 6$  helix to the far left the peptide binding ledge. This site is made from Leu12, Glu13, Lys14 and capped by Gln146 to create a small pocket.
2. a narrow hydrophilic indent approximately 5 Å deep in the apex of the protein (Figure 5.7). The indent is surrounded Asp71, Ser106, Arg109, Asp123, Trp126, and Asn127, while the bottom of the indent is formed by hydrophobic residues Trp126, Ile108, Ala105 and the polar phenol of Tyr122.
3. at the top left of the back of the protein above the acidic patch. This pocket is made up of charged residues Glu94, Arg103, Asp107 and Asp110 and the polar sidechains of Thr99 and Thr101.
4. a hollow at the back of the  $\alpha 6$  helix as an extension of the water channel. Similar to the pocket at the apex, the opening is surrounded by polar sidechains and backbone amides of Gly18, Pro20, Tyr59, Gln137, Lys140, and Asp144, while the inside is formed by hydrophobic residues Val16, Ala19, Pro20, Leu23, Val145, Phe154 and Gly157.



**Figure 5.7:** New binding hot spots identified using DoGSiteScorer (DoGSS). Binding hot spot residues shown as cyan sticks and surface, structural waters shown as red spheres, mapped onto *EcDsbA* shown as grey surface (PDB ID: 1FVK).

DoGSiteScorer predicted all but one FTMap cluster/hot spot but also predicted the cryptic pocket as well as three new pockets on *EcDsbA*. While sites are broken down into subpockets and ranked for their druggability, the grid system fails to identify key interactions that may be used for structure-based design and significant interpretation is required for this information to be used directly in analogue design.

### 5.1.3 Computational prediction conclusions

13 combined protein hot spots were predicted using both FTMap and DoGSiteScorer. Of these sites it is encouraging to see that the hydrophobic groove was predicted by both. However, the cryptic pocket was only predicted by DoGSS and while both computational methods predicted binding sites at the back of the protein in the acidic patch and water channel, multiple fragment screens against *EcDsbA* have failed to identify small molecules that bind there. This may be due to *EcDsbA*'s flexibility which could expand and contract this area as the two domains move relative to one another. Key interactions identified by FTMap in the hydrophilic region have already been investigated in two different series of fragments, with only small gains in affinity. Compounds that bind to the cryptic pocket also gain little information from these calculations as the limited flexibility of the protein in these calculations does not allow for possible opening to the cryptic pocket to be explored.



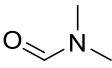



While these algorithms only require a crystal structure of the protein and do provide theoretical hot spots and subpockets to elaborate towards, each algorithm has limitations. We therefore conducted organic solvent screens by both X-ray crystallography and  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR to validate and possibly expand on these set of hot spots.



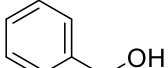
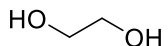
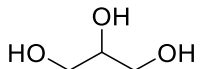
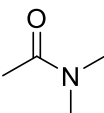
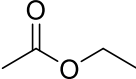
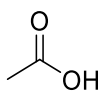
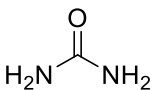
## 5.2 Solvent library selection

Experimentally identifying interaction hot spots using solvent libraries is an attractive technique. However, the high concentrations required to provide enough occupancy of the target for the identification of the interaction (often solvent concentrations of 2 – 5 % in solution NMR and > 50 % in X-ray crystallography) can make it difficult to execute these screens due to the potential for target unfolding, precipitation and non-specific binding.

A list of 15 solvents was selected to cover hydrophilic, hydrophobic, hydrogen bond donors and acceptors as well as aliphatic and aromatic motifs while also being miscible with water for NMR screening. These selected solvents varied in calculated octanol-water partitioning co-efficient (ClogP) from -1.67 to 1.18 (mean -0.01), and topological polar surface area (TPSA) from 9.23 to 69.11 (mean 28.17). These compounds are very simple, synthetic and molecular complexity (SMCM)  $\leq 14$ , and 13/15 compounds contain only 1 or 2 pharmacophores.

**Table 5.1:** Characteristics of organic solvent library

Solvent	Structure	MW (Da)	ClogP	TPSA (Å <sup>2</sup> )	H <sub>Don</sub>	H <sub>Acc</sub>	SMCM
Acetonitrile		41.05	0.53	23.79	0	1	4.44
Tetrahydrofuran (THF)		72.11	0.80	9.23	0	1	7.80
<i>N,N</i> -dimethylformamide (DMF)		73.10	-0.30	20.31	0	1	6.39
Dimethyl sulfoxide (DMSO)		78.14	-0.01	17.07	0	1	1.42
Methanol (MeOH)		32.04	-0.39	20.23	1	1	3.05
Ethanol (EtOH)		46.07	0.00	20.23	1	1	5.05

Isopropanol (IPA)		60.10	0.39	20.23	1	1	9.05
Acetone		58.08	0.60	17.07	0	1	4.67
Benzyl alcohol		108.14	1.18	20.23	1	1	14.05
Ethylene glycol		62.07	-1.03	40.46	2	2	7.094
Glycerol		92.10	-1.67	60.69	3	3	13.14
<i>N,N</i> -dimethylacetamide		87.12	0.10	20.31	1	2	6.39
Ethyl acetate (EtOAc)		88.11	0.57	26.30	0	2	4.47
Acetic acid		60.05	0.09	37.30	1	1	6.72
Urea		60.06	-0.98	-0.98	2	1	4.68

The selected solvents have between 2 and 6 heavy atoms (with the exception of benzyl alcohol, included as a water miscible hydrophobic ring system) which limits possible topologies and covers almost all pharmacophore types (HBD, HBA, negatively charged and hydrophobic aromatic and aliphatic). Although the selection is not exhaustive the majority of hot spots are expected to be able to be characterised using this set of solvents.

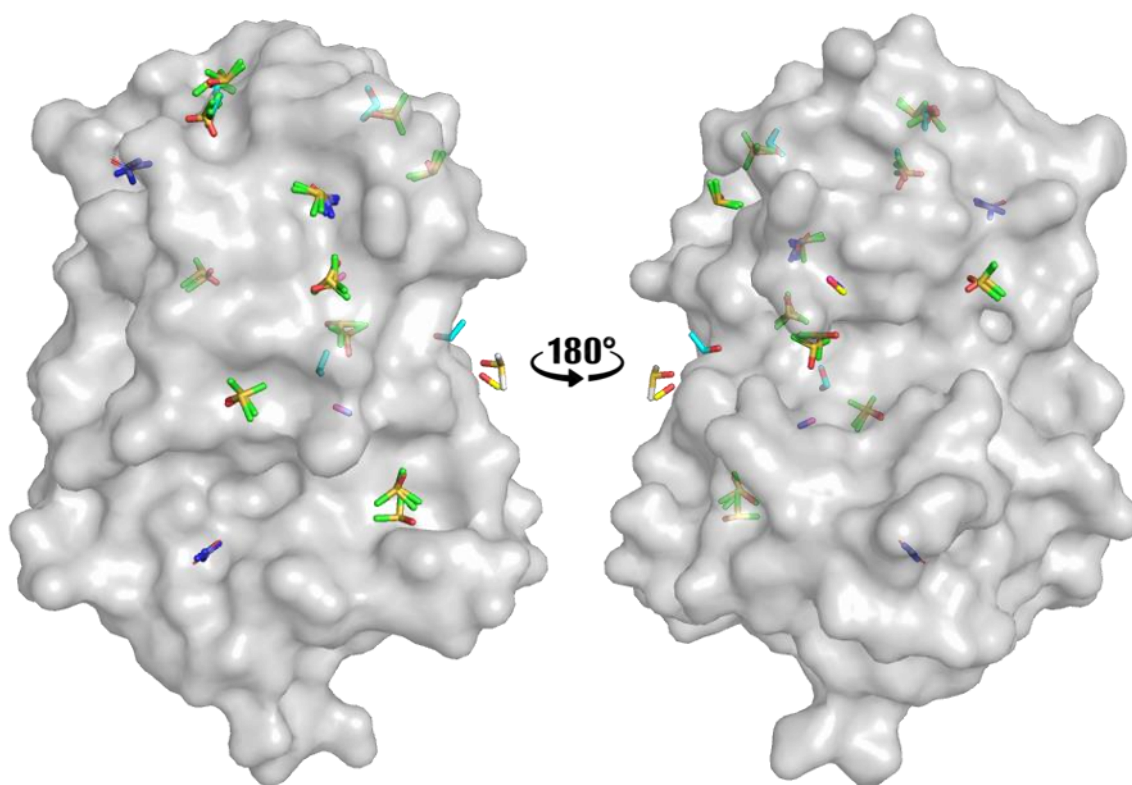
### 5.3 X-ray crystallography solvent screen

Both X-ray crystallography and  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR solvent screens were conducted using the library, discussed here first is the X-ray crystallography solvent screen. In order to obtain enough solvent occupancy, the protein needs to be treated with high percentages (v/v) of solvent. Many crystal systems are not amenable to such harsh conditions with pre-grown crystals often dissolving, cracking or no longer giving high resolution diffraction patterns upon soaking. Although *EcDsbA* has a very robust system for X-ray crystallography, many of the conditions attempted for this organic solvent screen had similar issues. Additionally, the volatility of these solvents caused issues in maintaining the desired final organic solvent concentration since evaporation within the crystal containing drop occurred quickly. Consequently, this may have resulted in some crystals providing false negatives. Preformed crystals of *EcDsbA* were soaked at 50 and 80 % (v/v) concentrations (between 5 – 20 M) of each solvent.

A resolution cut-off of 2.5 Å was implemented for the refinement of structures and these were ultimately solved to high resolutions between 1.8 – 2.5 Å (mean 2 Å, see chapter 8 for experimental details). Some studies have suggested that the high percentage of organic solvents used can cause conformational changes, based on the comparison of the root mean square deviation (RMSD) to the unbound structure <sup>(273)</sup>. However, the effect of the solvents on the main chain of *EcDsbA* was comparable to conformational changes caused by fragment binding. The small size of some of the solvents also often made it difficult to distinguish between a solvent and water molecule, also potentially resulting in false positives for some solvents. Ligands were generally fit with  $\beta$  factors between 50 – 80 Å<sup>2</sup>, however, the smallest solvents, methanol and acetonitrile, were fit with  $\beta$  factors between 35 – 50 Å<sup>2</sup>. When refined with either solvent or water occupying these small densities, similar  $\beta$  factor values were obtained and this did not aid in the discrimination between the two.

Nonetheless, of the 15 compounds soaked, 6 solvents gave structures which diffracted to < 2.5 Å and had density for the solvent (Figure 5.8). For clarity of analysis, ligands which simultaneously made interactions to both molecules of *EcDsbA* in the asymmetric unit or were found at the crystal packing interface were disregarded, as these pockets are not expected to be found in the solution form of *EcDsbA*. As glycerol

exclusively had ligands at crystal packing interfaces this meant that this data set was not included in the analysis. Hot spots which were occupied by more than one solvent were prioritised and overlaying all of the available solvent crystal structures identified seven hot spots in which molecules of two or more solvents were binding (Figure 5.9). Another six clusters were found with two or more overlapping molecules of the same solvent, either from overlaying the chains in the asymmetric unit or where two structures of the same solvent were available (Figure 5.10).



**Figure 5.8:** Overlay of all organic solvent structures solved by X-ray crystallography. Solvents shown as coloured sticks (DMSO = green, urea = purple, ethanol = blue, acetonitrile = pink, methanol = yellow). Molecules which were at crystal contacts or did not overlap with another molecule were removed for clarity.



### 5.3.1 Sites occupied by more than one solvent

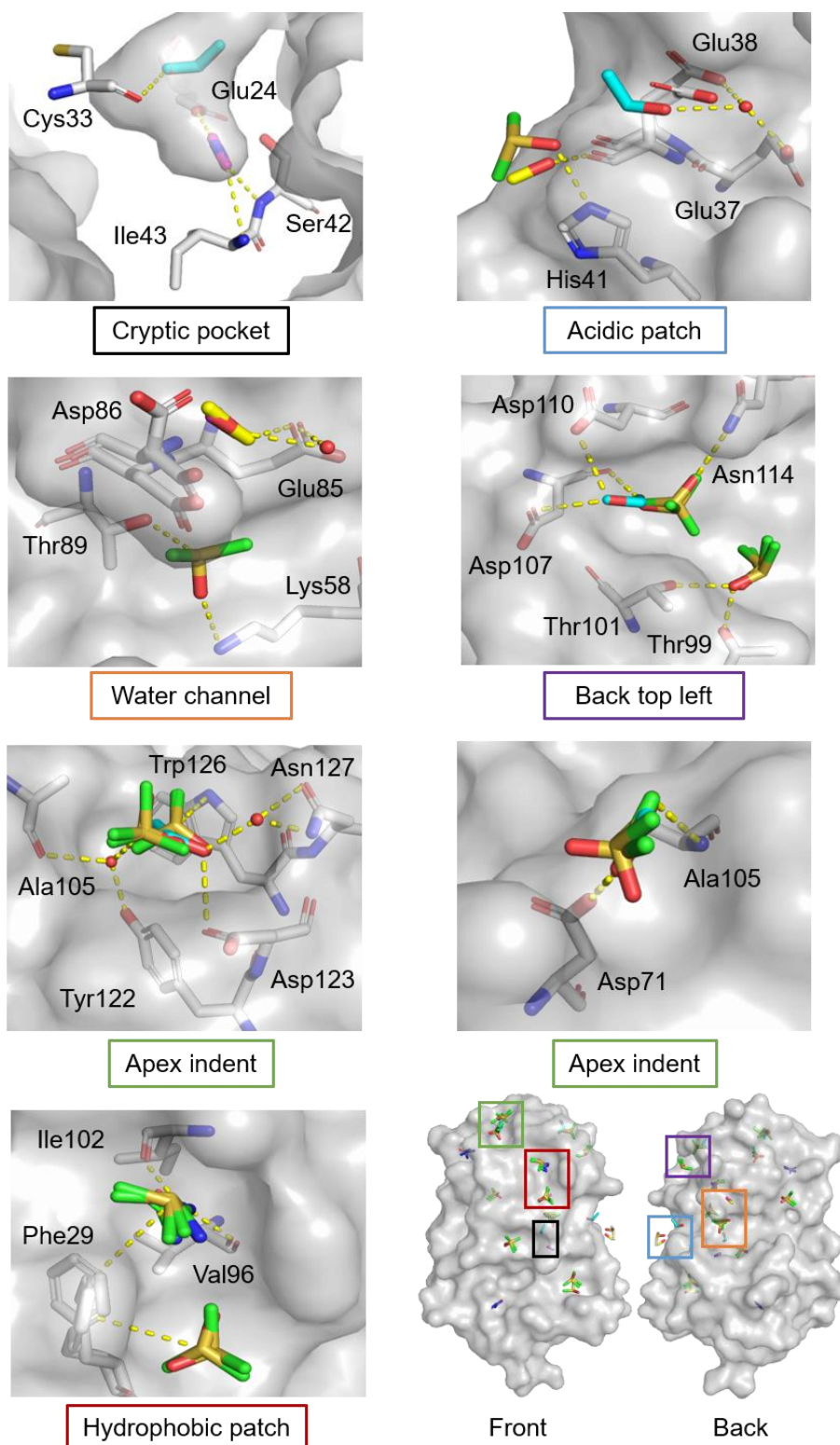
The sites that are occupied by two or more solvents, the cryptic pocket, acidic patch, rear water channel, apex indent and the back top left pocket had previously been predicted by FTMap and DoGSS. An additional site was also observed with solvents bound at a hydrophobic patch on the front top right of *EcDsbA* approximately 10 Å above the active site Cys30 (Figure 5.9).

Both ethanol and acetonitrile were found to bind within the cryptic pocket, making hydrogen bonds with the sidechain of Glu24, and with the backbones of Cys33, Ile42, and Ser43. Crystal structures of fragments bound to the cryptic pocket expand the size of the pocket and also make hydrogen bonds to the acid side chain of Glu24. The solvent screen however, did not identify interactions to Phe25 and Glu37 which are observed in structures with some fragments which bind at this site. It was also noted that although urea was not observed in the cryptic pocket, Glu37 was found in an alternative conformation in these structures resulting in the cavity slightly expanding. It is possible that urea is bound with low solvent occupancy which may have resulted in density too weak to fit a model. There were no interactions in the cryptic pocket that were observed by more than one solvent, despite multiple solvents being able to fit within the pocket and make the interaction. For example, the hydrogen bond observed from ethanol to Cys33 backbone could presumably be formed by methanol, however no density for methanol was found here.

At the acidic patch methanol was found to form a hydrogen bond donor interaction with the amide carbonyl of Glu38. Ethanol was also found to bind at the acidic patch but forms a hydrogen bond to a conserved water that in turn is bound to the side chains of Glu38 and Glu37. Finally, DMSO was found making hydrogen bond acceptor interactions with an altered conformation His41 side chain. The rear water channel has two areas in which solvents bind. Firstly, a DMSO molecule was found making hydrogen bond acceptor interactions with Thr89 and Lys58 at the bottom of the channel. Acetonitrile and methanol were found in overlapping poses, however, were ambiguous in orientation and either made a hydrogen bond donor or a water mediated hydrogen bond to Glu85 side chain from above the channel. These may be the only two solvents binding in this spot due to their small size. Similarly, two solvent clusters

were found at the apex indent, binding both ethanol and DMSO on either side of Ala105. Multiple water mediated hydrogen bonds are observed between the protein and DMSO and ethanol. One of these, the hydrogen bond to the sidechain of Asn127, is the only conserved interaction in this hot spot. Additional water mediated hydrogen bonds are made to Asp71, Ala105, Tyr122 and ethanol also participates directly in a hydrogen bond donor interaction to the sidechain of Asp123.

The hot spot at the back of the protein at the top left was found to bind four DMSO and one ethanol molecule, all making different interactions to Asp107, Asp110, Asn114, Thr99 and Thr101. The DMSO molecules which interact with Asn114 have high  $\beta$  factor values (69 and 89 Å<sup>2</sup>) which may indicate these interactions are less stable. Within the new hydrophobic patch, which was identified at a position ~10 Å above the active site on the front top right of the protein, multiple DMSO molecules are found making hydrophobic interactions with the aromatic ring of Phe29. Urea is also found in an overlapping pose with DMSO at this hot spot, however, it forms two hydrogen bond donor interactions to the carbonyls of Val96 and Ile102.

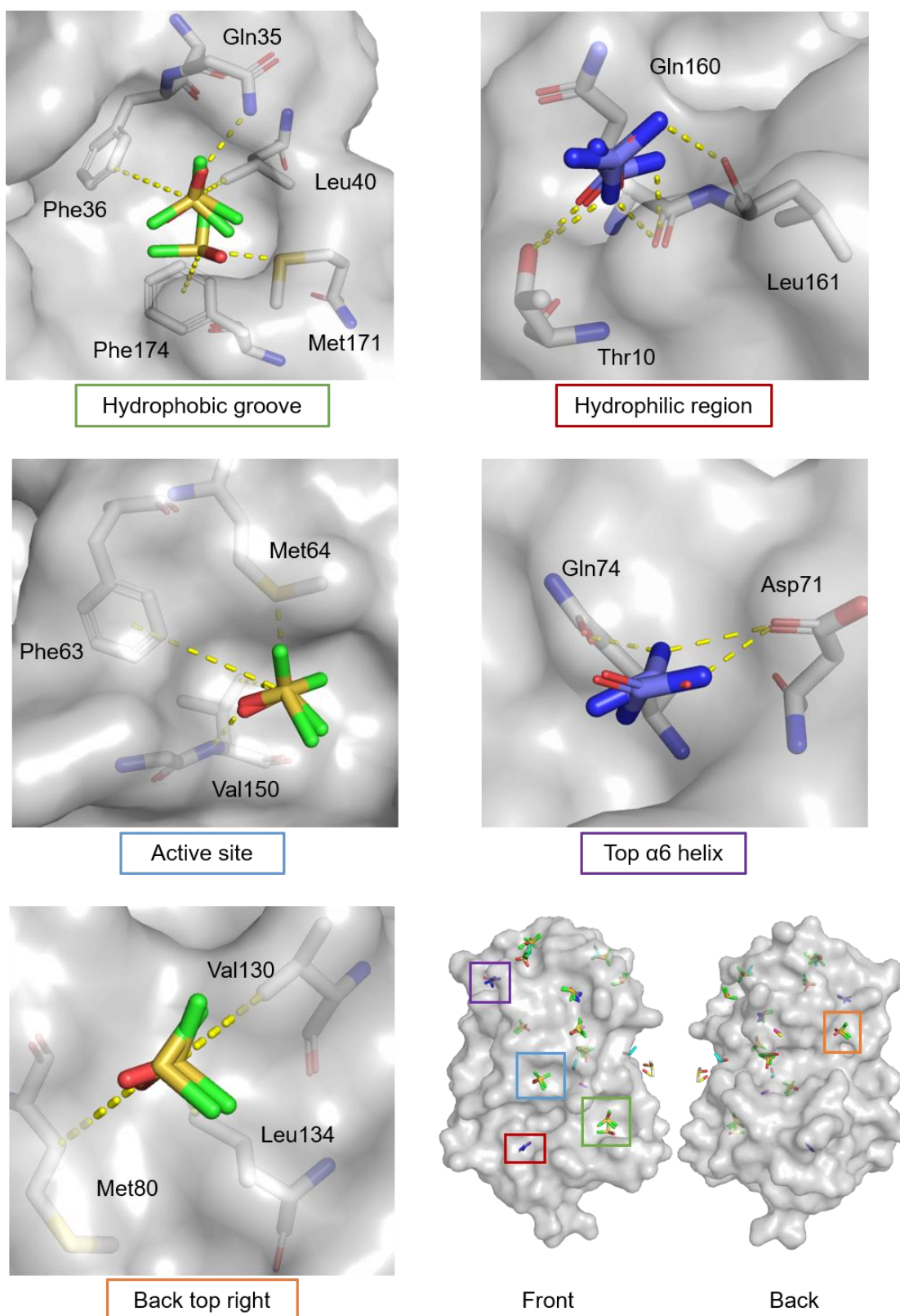


**Figure 5.9:** Binding sites occupied by multiple solvents. Solvents shown as coloured sticks (DMSO = green, urea = purple, ethanol = blue, acetonitrile = pink, methanol = yellow), protein residues shown as grey sticks, water shown as red spheres and interactions shown as yellow dashes. Coloured boxes around the binding site shown on the full view of the protein correspond to the border of the hot spot label.

### 5.3.2 Sites occupied by multiple copies of the same solvent

Hot spots where two copies or more of only a single solvent were observed occur in the hydrophobic groove, hydrophilic region, active site and  $\alpha 6$  helix top pocket as well as a new hot spot on the back top right of the protein that was not predicted by FTMap or DoGSS (Figure 5.10). Three molecules of DMSO were observed in the hydrophobic groove. These made hydrophobic interactions with Met171, Phe36 and Phe174, and one DMSO molecule made a hydrogen bond acceptor interaction with the side chain of Gln35. A similar hydrogen bond has been observed previously with some fragments that bind in the hydrophobic groove. Two molecules of urea bound to the hydrophilic region and showed hydrogen bonds to the side chain of Thr10 as well as the backbone of Leu161 and Gln160. In the  $\alpha 6$  helix top pocket urea made multiple hydrogen bond donor interactions to Asp71 and Glu74. Within the active site pocket two molecules of DMSO made hydrophobic interactions with Met64 and Phe63, as well as a conserved hydrogen bond acceptor interaction with the amide of Val150. Two DMSO molecules were also bound at the novel hot spot on the back top right however, no polar interactions were made to the protein.

X-ray crystallography solvent screening was able to identify the hydrophobic groove and cryptic pocket where fragments have been found to bind. Furthermore, it identified substrate peptide hot spots such as the hydrophilic region and the active site and predicted hot spots on the back of the protein in the acidic patch, rear water channel and at the indent at the protein's apex. However, very few polar interactions were found that were conserved between multiple different solvent molecules, making selection and prioritisation of potential interactions difficult. Of the cases where the same polar interaction is observed in more than one crystal structure, Thr10 and Gln160 in the hydrophilic region and Val150 of the active site are the only interactions in close proximity to fragment binding sites. Based on X-ray crystallography these would be the highest priority hot spots to target for elaboration of the current inhibitor series.

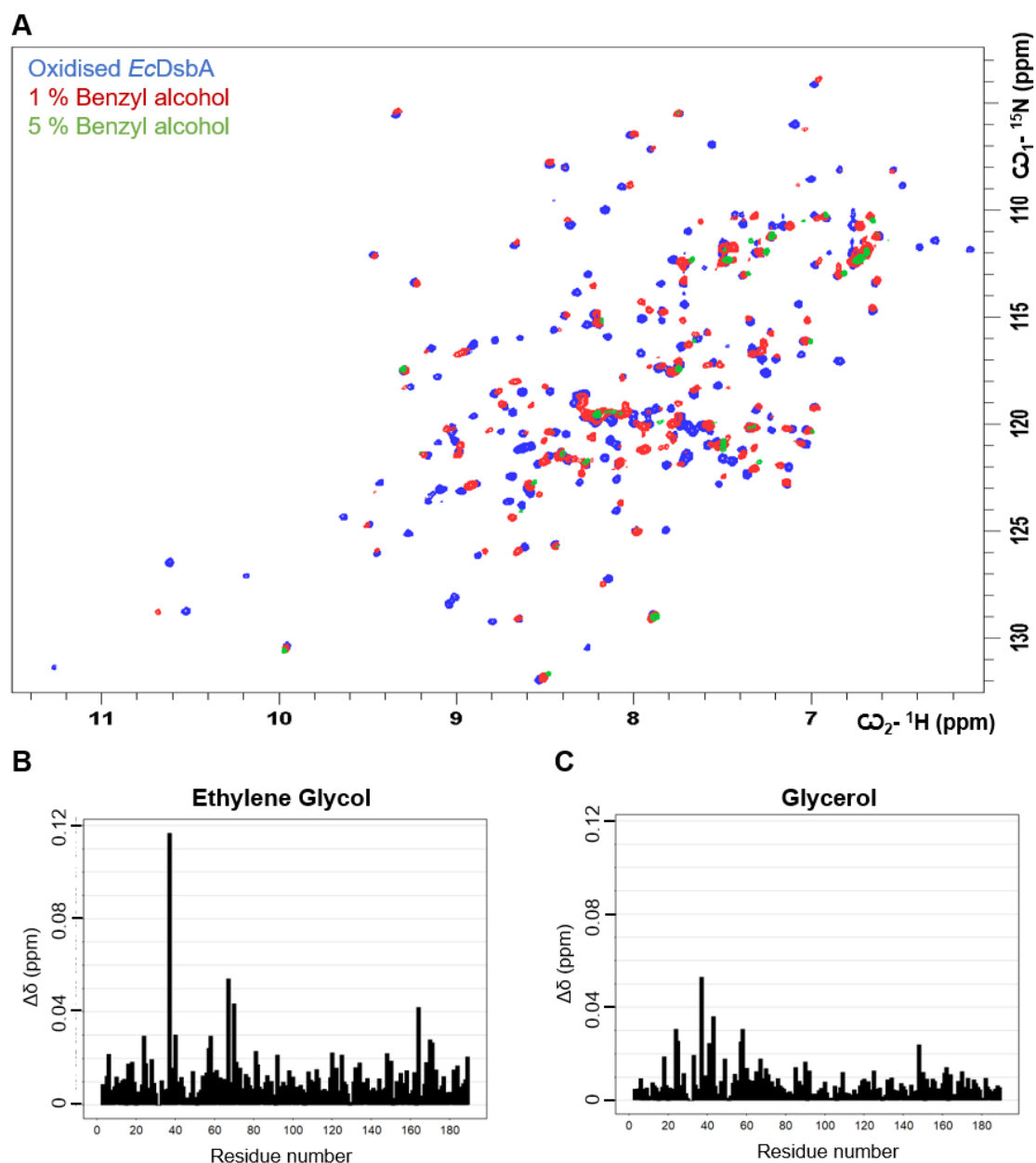


**Figure 5.10:** Binding sites occupied by single solvents. Solvents shown as coloured sticks (DMSO = green, urea = purple, ethanol = blue, acetonitrile = pink, methanol = yellow), protein residues shown as grey sticks and interactions shown as yellow dashes. Coloured boxes around the binding site shown on the full view of the protein correspond to the border of the hot spot label.

#### 5.4 $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR solvent screen

To compliment the X-ray crystallography solvent screen a  $^1\text{H}$ - $^{15}\text{N}$  HSQC screen was also performed with the same series of 15 solvents. HSQCs were acquired with U- $^{15}\text{N}$  *EcDsbA* (100  $\mu\text{M}$ ) in the presence of 1 – 5 % (v/v) solvent (~1 – 12 M). Comparisons of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of *EcDsbA* in the presence and absence of an organic solvent were used to measure chemical shift perturbations (CSPs) for individual residues. To investigate potential binding site locations the CSPs were mapped onto the structure of *EcDsbA*.

Glycerol and ethylene glycol were classified as non-binding as they showed only one and four residues with CSPs over 0.04 ppm, respectively (Figure 5.11). Benzyl alcohol and acetic acid caused loss of intensity for the majority of amide resonances in spectra for all concentrations indicating large changes in dynamics or protein precipitation from solution. These solvents were therefore excluded from the analysis conducted below.



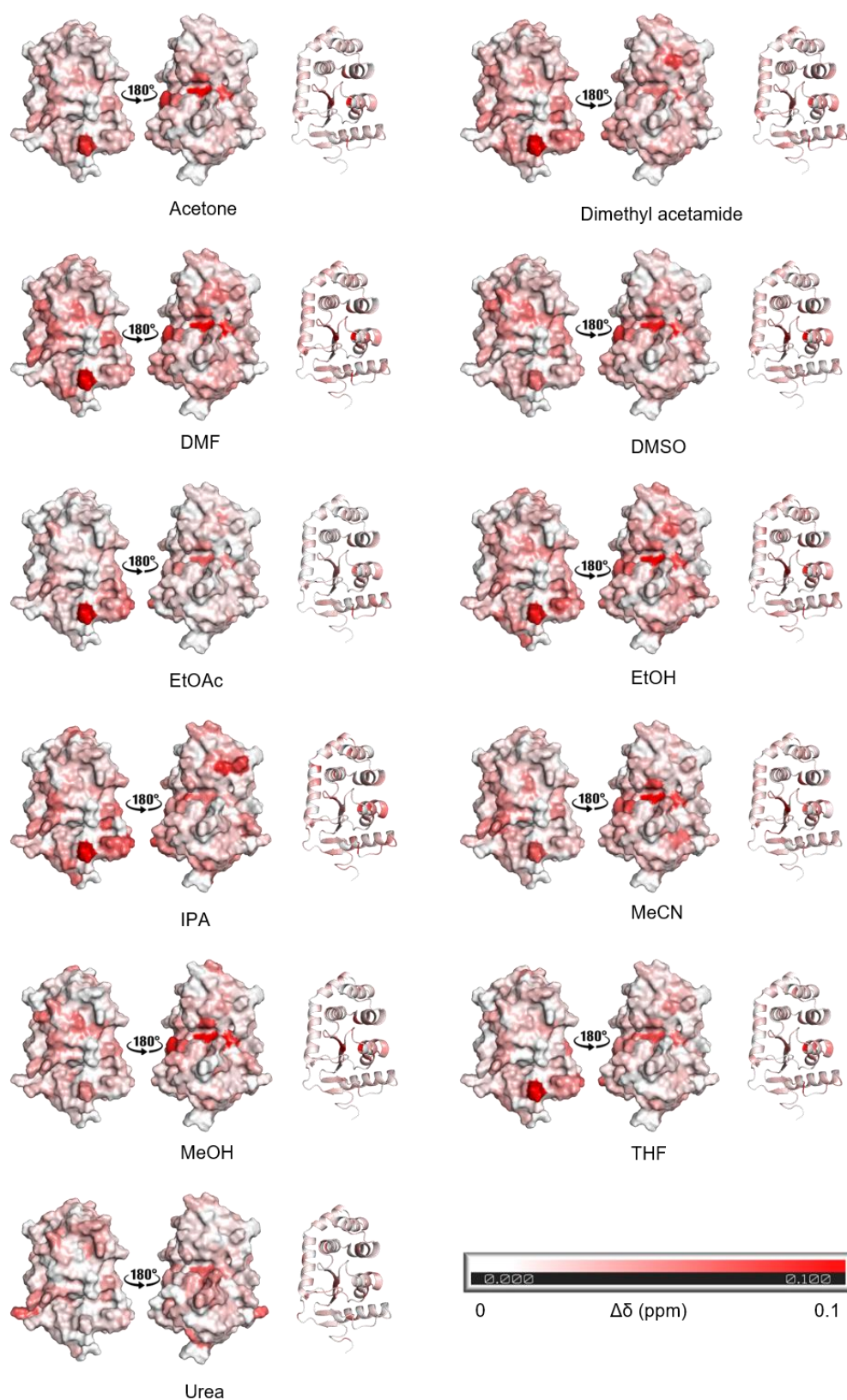
**Figure 5.11:**  $^1\text{H}$ - $^{15}\text{N}$  HSQC binding profiles of benzyl alcohol, ethylene glycol and glycerol  
A)  $^1\text{H}$ - $^{15}\text{N}$  spectra of oxidised  $^{15}\text{N}$  *EcDsbA* (blue), incubated in the presence of 1 % (v/v, red) and 5 % (v/v, green) benzyl alcohol. B) Chemical shift perturbations of *EcDsbA* treated with ethylene glycol as a function of residue. C) Chemical shift perturbations of *EcDsbA* treated with glycerol as a function of residue.

#### 5.4.1 Hot spots identified by chemical shift perturbation mapping

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra showed a large number of CSPs for each of the 11 solvents analysed. Ten residues (Cys33, Val39, Ile42 – Asp44, Thr57 – Tyr59, Val61, Gln164) gave perturbations over 0.03 ppm in every spectrum. The majority of these residues are found at the top of the hydrophobic groove, the acidic patch, the water channel and partially line the cryptic pocket. The perturbations of these residues are very large (up to 0.15 ppm) and when mapped onto the structure of *EcDsbA* (Figure 5.12), they visually dominate the appearance of their respective areas. In addition to these sites, residues which give CSP > 0.03 ppm in at least half of the solvents, when mapped onto the structure of *EcDsbA*, suggest hot spots at the hydrophobic patch and the site midway down the  $\alpha 6$  helix. However, using these criteria, there are one or two residues found in other binding sites such as the back top left and back top right. When inspecting the individual CSP maps these residues appear to be surrounded by residues with weaker CSP.

This highlighted a problem with using a simple threshold to interpret the CSP data. If, in these cases, these weak shifts are considered indicators of binding at this location of the protein, shifts of the same magnitude should also be considered binding events without the presence of one or two strong perturbations. Weak CSP like these are found to cover large portions of the protein surface (eg DMSO, Figure 5.12) and do not allow the determination or prioritisation of discrete hot spots. Furthermore, a number of residues are buried within the protein (such as the back of the  $\alpha 6$  helix), or are NMR silent (proline or unassigned residues) and can obscure the identification of hot spots. Hence a quantitative method for identification of the hot spots based on the CSPs observed in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR was investigated.





**Figure 5.12:**  $^1\text{H}$ - $^{15}\text{N}$  HSQC analysis of organic solvent binding. CSPs of solvents mapped onto the crystal structure of *EcDsbA* (PDB ID: 1FVK) and coloured from white to red by the extent of CSP: minimum = 0.0 ppm (white) and maximum = 0.1 ppm (red).

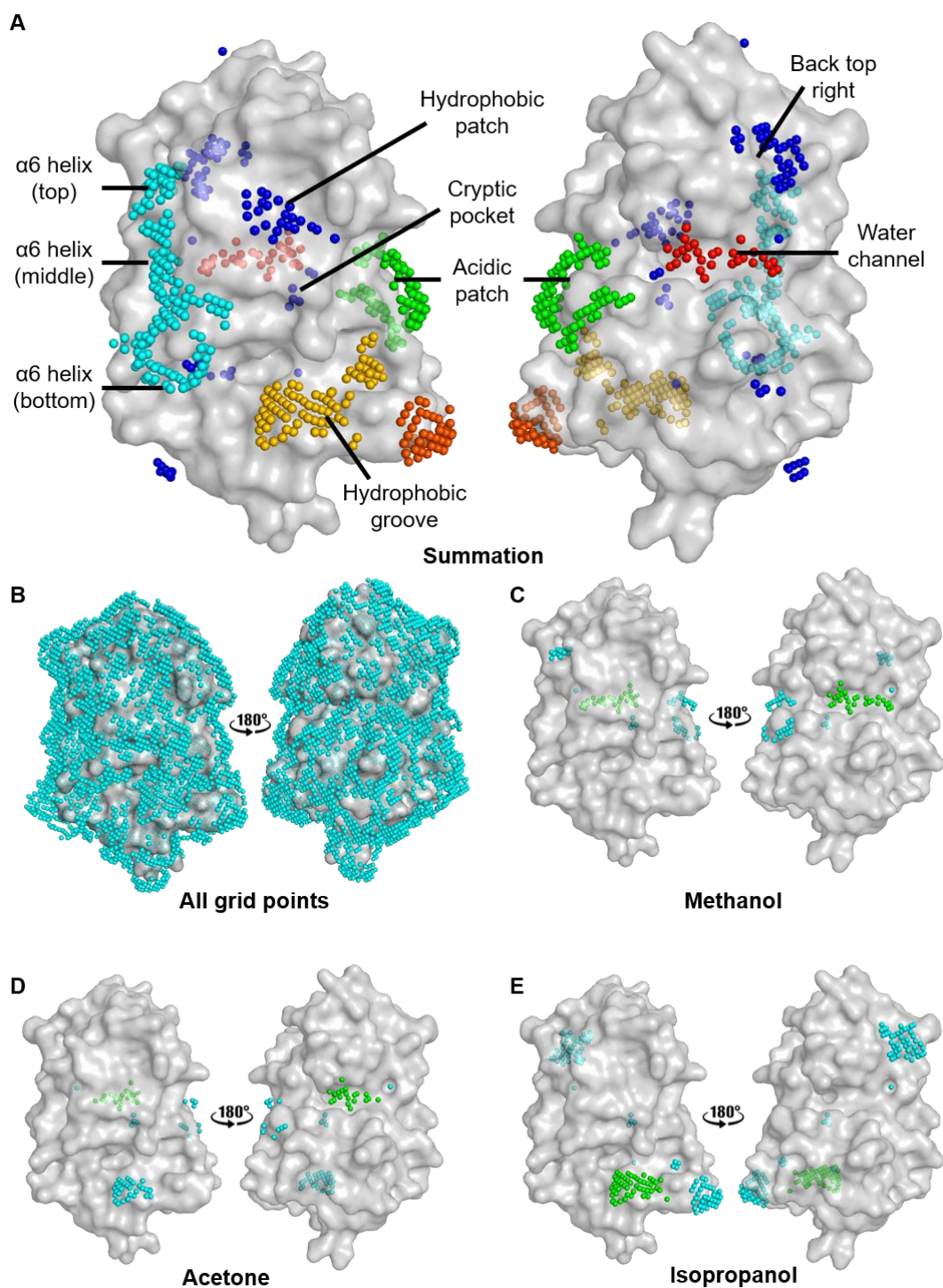
### 5.4.2 Hot spot identification by grid point analysis

In order to capture the potential hot spots and produce volumes in the surface features of the protein that were more consistent with computationally (FTMap and DoGSS) and X-ray crystallography identified hot spots, a grid point system around the protein was used. Multiple methods of scoring the grid points based on nearby CSPs were then investigated. Briefly, a value could be assigned to each grid point using three factors:

1. nearby amide N-H CSPs ( $< 5 \text{ \AA}$ ) were treated either as the magnitude of CSP or simply CSP above a set noise level ( $\geq 0.04 \text{ ppm}$ )
2. scaling the CSP values by their distance to the grid point using binary, linear or exponential scaling factors
3. CSPs from multiple residues to the same grid point were combined by average, geometric mean or summation

The grid points can be optionally smoothed with adjacent grid points and filtered based on their values. An agglomerative clustering method based on the number of adjacent points was then used to identify clusters where a set number of points, average value or maximum value above threshold was observed. Grid points and grid point clusters can then be inspected for each solvent separately or summed together across multiple solvents to provide a global map of potential hot spots. Here we discuss the method which gave the clearest identification of potential hot spots by  $^1\text{H}$ - $^{15}\text{N}$  HSQC solvent screening.

A grid of points  $1 \times 1 \times 1 \text{ \AA}$  spaced covering the protein surface was generated at a maximum of  $5 \text{ \AA}$  away from each residue's amide N-H hydrogen (Figure 5.13). Using an in-house script each solvent's grid points were assigned a value based on the magnitude of a the CSP observed for nearby residues and the distance between the grid point and that residue (see chapter 8 for experimental detail). If a group of adjacent grid points with high values was observed it was taken to indicate that a binding event had occurred in the area. Importance of binding sites can be estimated based on the number and increased value of grid points within a cluster. A summation and clustering of the data for all solvent grid points was also conducted to allow for the identification of consistent binding sites across multiple solvents.



**Figure 5.13:** NMR predicted binding sites of organic solvents. A) Predicted binding sites by grid point clusters (shown as coloured spheres) calculated by the summation of individual solvent clusters. B) All possible grid points (shown as cyan spheres) and their coverage of the surface of *EcDsbA* (PDB ID: 1FVK, shown as grey surface). C) Predicted binding sites by grid point clusters (shown as coloured spheres) for methanol. D) Predicted binding sites by grid point clusters (shown as coloured spheres) for acetone. E) Predicted binding sites by grid point clusters (shown as coloured spheres) for isopropanol.

From analysis of the grid point clusters of the individual solvents it was observed that the two major fragment binding sites, the hydrophobic groove and cryptic pocket, were consistently populated by grid points across the majority of solvents (10/13). Where ethanol and DMSO were not found in the hydrophobic groove and urea did not generate any grid clusters for any hot spots due to the weak CSPs observed. The water channel was also found to be a common binding site with methanol, ethanol, acetonitrile, DMF and DMSO all showing grid clusters here, consistent with inspection of the mapped CSP data. Small solvents such as acetonitrile, methanol, and acetone also gave clusters in the acidic patch, with grid points centred around His41 being the strongest. Isopropanol suggested the presence of a previously unidentified hot spot on the back of the protein at the top right above the water channel consisting of Met80, Ala81, Val130 and Leu134 (Figure 5.13).

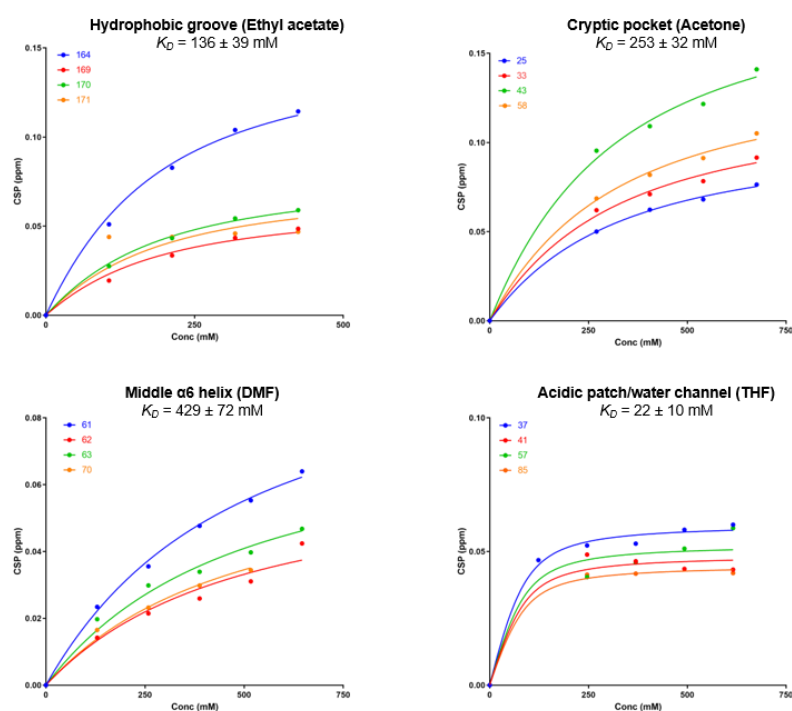
The summation of the grid point values from all solvents also predicted interactions to the two experimentally observed fragment binding sites at the hydrophobic groove and cryptic pocket. In addition, hot spots were identified at the acidic patch and water channel, which was often the strongest cluster for the individual solvents. The  $\alpha 6$  helix grid point cluster was found to be the largest cluster overall, with grid points extending the full length of the helix. This volume encompassed the hot spot at the top of the helix to the hot spot at the bottom, which had previously been identified using DoGSS and the X-ray crystallography solvent screen. Although the site on the back top right of the protein was only evident in the presence of isopropanol, its appearance upon summation of the data suggests that there may be minor interactions at this site for other solvents as well. The hydrophobic patch which was not defined as a hot spot on an individual solvent level, was validated as a hot spot in a similar manner.

#### **5.4.3 $K_D$ determination of binding hot spots**

In addition to the single point analysis conducted based on CSPs at 5 % v/v a 6-point  $^1\text{H}$ - $^{15}\text{N}$ -HSQC titration series from 0 - 5% organic content was also acquired. Using these titrations,  $K_D$  values were estimated for binding at different hot spots around the protein. Residues which gave perturbations less than 0.04 ppm at the highest

concentration or non-linear perturbations were excluded from the calculations and the data were fit to a single site with ligand depletion binding model (Figure 5.14). Some binding sites were unable to have  $K_D$  values fitted (or gave inaccurate fits) due to few residues meeting the minimum CSP criteria or not showing curvature in the binding model.

The CSPs obtained for the hydrophobic groove, cryptic pocket, acidic patch, water channel, middle of the  $\alpha 6$  helix and hydrophobic patch were large enough to enable affinity data to be generated. This observation is also consistent with the rankings obtained by the grid point analysis based on number of grid points and their values. The acidic patch and water channel gave the lowest  $K_D$  values estimated between 10 – 100 mM, followed by the cryptic pocket (50 – 150 mM), and all other binding sites had estimated affinities ranging between 100 – 800 mM. The lower  $K_D$  values could be caused by synergistic chemical shifts between the cryptic pocket and water channel as residues are located in close proximity to both hot spots. Additionally, residues between the two sites overlap and are used in calculations of affinities for both.



**Figure 5.14:** Binding isotherm obtained for each binding site using 0 – 5 % (v/v) solvent  $^1\text{H}$ - $^{15}\text{N}$  HSQC titrations.

## 5.5 Comparison of hot spot identification techniques

Computational calculations of hot spots can be fast and are often provided as free to use services where little to no resources are required to obtain results, and where numerous potential binding sites can be identified. For DoGSS these potential binding sites were often very broad and the calculations provide no details of specific interactions which may be targeted in developing small molecules. While FTMap does provide details of specific interactions formed by the probes, it failed to identify a number of sites that were observed experimentally through X-ray crystallography and NMR solvent screening. Although multiple conformations of a target proteins can be used for FTMap calculations, these conformations must either be generated computationally or through extensive experimental structural studies and even then, they may miss certain populations of structures generated by protein dynamics. This is evident when applying FTMap to *EcDsbA*, which did not identify the cryptic pocket. In contrast, DoGSS identified more hot spots, however the large surfaces and volumes of the suggested sites decreased the confidence in these predicted sites. Furthermore, interactions with structural waters can often be integral for catalytic activity or affinity and their removal prior to the calculations also cause these important interactions to be overlooked.

X-ray crystallography can identify distinct interactions between solvent molecules and the protein, however, the size of the solvents is often comparable to that of a water molecule, making it difficult to distinguish between them in the data. Refinement of the solvent and water often resulted in equally good fit of density and similar  $\beta$  factors meaning that it was not possible to unambiguously assign which molecule was bound. Crystal packing and contacts between multiple copies of the protein in the asymmetric unit may also occlude or enhance binding to some sites, which adds additional complexity to the data. To compensate for low affinities, very high concentrations of solvent are used. Unfortunately, while these conditions can help increase occupancy, which enables the solvent to be observed in the electron density, they are often detrimental to the viability of the crystals. This results in crystals which diffract to lower resolution or do not diffract at all. High solvent concentrations may also induce large conformational changes which can disrupt the integrity of the crystals. Moreover, crystallography does not give an indication of the relative importance of the

interactions in terms of their contribution to binding affinity, and as very few interactions with *EcDsbA* were conserved between the different solvent molecules, this makes ranking of these hot spots for investigation in fragment elaboration strategies problematic.

$^1\text{H}$ - $^{15}\text{N}$  HSQC NMR solvent screening avoided some of these problems by identifying hot spots in samples where the protein was present in aqueous solutions. This potentially allows for sampling of structures that are formed dynamically in solution, which may be difficult to capture with other methods. However, this presents a different problem as solvent binding may perturb the protein dynamics. In this case some of the CSP observed can be due to conformational changes for residues that can be distant from the solvent binding site leading to uncertainty in identifying the true binding site. Large perturbations in lone residues can also obscure the recognition of binding sites which may have multiple residues with consistent weak shifts. Artefacts such as these can lead to false negatives in sites where ligand occupancy may be low. Therefore, to attempt to address some of these issues, a systematic grid point analysis was designed.

While the grid points allow for the ranking of binding sites and clustering of perturbations more objectively than manual inspection, it can also introduce other biases. The grid does not cover the entire exposed surface area of the protein, as the grid is generated based on solvent exposed amide N-H hydrogens only. Unfortunately, some of the areas containing few grid points may correlate to potential binding sites. These sites could be overlooked due to the lack of available grid points in close proximity to the observed CSPs, especially when the CSPs are of a low magnitude. Conversely, pockets of the protein that are surrounded by many amide N-Hs (such as internal or deep pockets) are potentially overrepresented as many more surrounding residues can contribute to the grid point value. Furthermore, specific interactions still need to be inferred from the extent of CSP of residues encompassed in the identified binding sites.

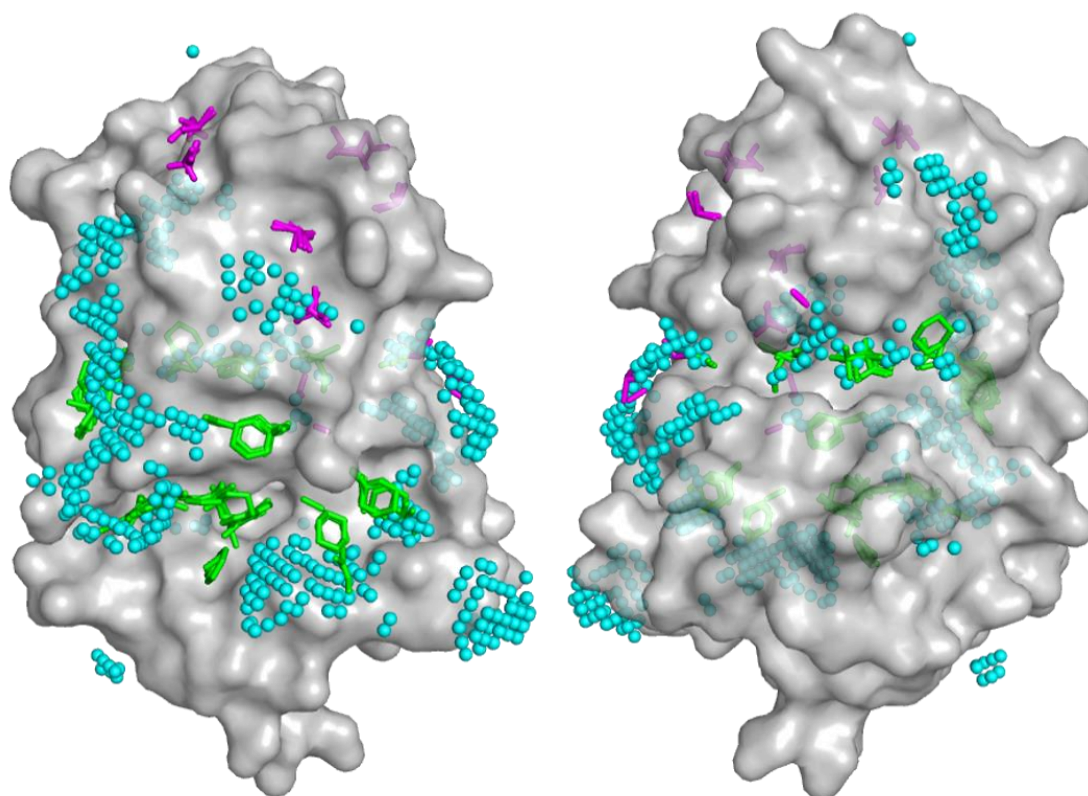
The limitations observed for, and artefacts caused by, the individual techniques described highlight the necessity for validation across multiple methods. Although



FTMap uses six solvents included in the NMR and crystallography screens, data sets for only three solvents, urea, acetonitrile, and ethanol, were obtained across FTMap, NMR and X-ray crystallography. All three of these methods placed ethanol at the acidic patch, however each method suggested different binding poses. NMR clustering indicates that the binding event would be below His41, whereas X-ray and FTMap provide poses above His41. While FTMap and crystallography both suggest the formation of hydrogen bonds to the sidechain of Glu38, the FTMap ethanol forms this bond directly and whereas a water mediated bond is observed by crystallography. This crystallographic water facilitates two hydrogen bonds (to Glu37 and Glu38), however is removed for the FTMap calculations. Removing this water only allows ethanol to form one of these interactions by FTMap and provides a good example of one of this method's drawbacks.

Similarly, all three methods place acetonitrile at the water channel, however X-ray crystallography and FTMap show it binding above and within the channel respectively, and thereby do not maintain consistent interactions. In contrast, NMR clustering covers all of the three poses provided by FTMap and X-ray crystallography. In the cryptic pocket, NMR and X-ray crystallography both suggest that ethanol and acetonitrile are able to bind, with the results for ethanol resulting in a direct overlay. NMR and FTMap suggest that acetonitrile will bind within the hydrophobic groove, however at opposite ends of the hot spot, approximately 9 Å away. Likewise, urea binds within the hydrophilic region at Thr10 by X-ray crystallography, however in the FTMap predictions it binds in the place of the structural waters approximately 4 Å away.





Technique	FTMap	NMR	X-ray crystallography	Total binding sites
FTMap	-	4 (57 %)	2 (29 %)	7
NMR	4 (44 %)	-	4 (44 %)	9
X-ray crystallography	2 (33 %)	4 (67 %)	-	6

**Figure 5.15:** Comparison of binding hot spots identified by different techniques. Overlay of FTMap predicted binding poses (green sticks), NMR predicted binding site clusters (cyan spheres) and multi-solvent clusters solved by X-ray crystallography (magenta sticks) onto *EcDsbA* (PDB ID: 1FVK, grey surface).

The three techniques showed little concordance when viewed on an individual solvent/probe level. Consequently, the hot spots and highly ranked interactions were compared. Overall a total of 15 potential binding hot spots on *EcDsbA* were identified across four methods (FTMap, DoGSS, crystallography and NMR). Of these 15, 11 binding sites were validated in 2 or more of the methods, 13 if the single solvent sites observed by X-ray crystallography are included (Table 5.2). Interestingly, the two sites

where fragments have been observed to bind, the hydrophobic groove and the cryptic pocket were only validated in three approaches each (NMR, FTMap and DoGSS, and NMR, X-ray crystallography, and DoGSS, respectively). The acidic patch and the water channel were the only hot spots to be identified in all methods, suggesting that these would be the most common or strongest binding interaction sites. This is supported by these sites having the lowest estimated  $K_D$  values by NMR titrations ( $22 \pm 10$  mM). Within the acidic patch, interactions with the sidechains of Glu37, Glu38 and His41 are conserved across the techniques. The water channel containing Met56, Lys58, Trp76, and Glu85 was the most commonly observed hot spot across all methods showing interactions with these residues or CSP in the NMR data, for 12 of the solvents. The  $\alpha 6$  helix middle hot spot favoured non-polar compounds interacting with the sidechain of Phe63 and Val61.  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR also show consistent chemical shifts around these residues. The hydrophobic patch above the active site is largely non-polar with a prominent area formed by the aromatic sidechain of Phe29. Although this hot spot was identified by both NMR and crystallography, Phe29 is the only residue that is highlighted in binding across the two methods. The binding hot spots at the back of the  $\alpha 6$  helix and the back top right were deemed the least important binding sites as they are not known interaction sites (unlike the active site and peptide binding ledge) and were only identified by one method.

**Table 5.2:** Binding site cross validation

Binding site	FTMap	DoGSS	NMR	X-ray crystallography
<b>Hydrophobic groove</b>	Y	Y	Y	Single solvent only
<b>Cryptic pocket</b>	N	Y	Y	Y
<b>Peptide binding ledge</b>	Y	N	N	N
<b>Hydrophilic region</b>	Y	Y	N	Single solvent only

<b>Active site</b>	Y	N	N	Single solvent only
<b>Acidic patch</b>	Y	Y	Y	Y
<b>Water channel</b>	Y	Y	Y	Y
<b>α6 helix (top)</b>	N	Y	Y	Single solvent only
<b>α6 helix (middle)</b>	Y	Y	Y	N
<b>α6 helix (bottom)</b>	N	Y	Y	N
<b>α6 helix (back)</b>	N	Y	N	N
<b>Back top left</b>	N	Y	N	Y
<b>Apex indent</b>	N	Y	N	Y
<b>Hydrophobic patch</b>	N	N	Y	Y
<b>Back top right</b>	N	N	Y	Single solvent only

## 5.6 Conclusions and future directions

X-ray crystallography and  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR solvents screens as well as computational hot spot identification techniques were conducted against *EcDsbA* to discover conserved binding hot spots and interactions that may help future fragment optimisation efforts. A total of 15 potential hot spot/interaction sites were suggested, and 11 of these sites were validated across two or more methods. Fragments and analogues binding to the hydrophobic groove did not have any adjacent and validated interactions to exploit in the evolution of these compounds. FTMap did not reveal the

cryptic pocket which may be due to the limited ability of this approach to sample flexibility. Furthermore, X-ray crystallography identified solvents binding within this site, however not all of the interactions made by fragments were observed with the solvents.

To our knowledge no fragments or peptides have been observed to bind to the most prominent hot spot identified (the acidic patch and water channel). Furthermore, due to the extremely low affinity of solvents for the hot spot high solvent concentrations were required and this made it difficult to prepare samples and obtain unambiguous high-quality data for some solvents. Despite these draw backs, screening simple, minimal functional groups for hot spot analysis provided some promising data with *EcDsbA*. Development of these methods to provide higher affinity interactions, clearer data, and improve the information generated for implementation in fragment elaboration strategies was explored further in chapter 6.

# **Chapter 6:**

## **Design and screening of a MicroFrag library**

Solvent screening allows for the identification of energetically favourable interaction hot spots on a protein surface that can be targeted when elaborating fragments. This can aid in the design of higher affinity binders. As solvents bind with very low affinity, the technique requires exposing proteins to very high concentrations of organic solvents (~0.5 – 20 M) in order to achieve sufficient occupancy so that binding can be detected. Not all proteins that will tolerate these conditions, meaning this approach can be limited in its implementation. In addition, the high organic solvent concentrations that are used can alter the targets conformation or dynamics<sup>(273)</sup> and may not provide a comprehensive coverage of binding events that can occur in physiologically relevant aqueous environments. In the solvent screen with *EcDsbA* (Chapter 5) it was found that the most frequent hot spots, bound solvents with  $K_D$  values that were in the range of 100s of mM. This suggests that at the high concentrations that were used for screening, the probes should have had sufficiently high occupancy to be detected, but nonetheless it was observed that the electron density observed in the X-ray crystallography data was ambiguous. The ambiguity arose from both the orientation of the solvent or at times identifying if the density was due to the solvent or a water molecule.

A more recent approach that has been reported in FBDD, involves X-ray crystallographic screening of small libraries of very small compounds with a heavy atom count (HAC) between 5 – 13. These small compounds are screened at high ligand concentration (0.05 – 1 M)<sup>(136-138)</sup>. The advantage of using probes of this size is that the sampling of chemical space is still higher than in conventional fragment libraries while molecules of this size typically bind with higher affinity than organic solvents, which make the binding events easier to detect in biophysical assays. These libraries have been designed to include compounds containing heteroatoms, which aid in maintaining aqueous solubility and provide the ability to form polar interactions with the target. In addition, aromatic and halogenated compounds were incorporated to ensure a wide scope of drug-like interactions are able to be emulated and lipophilic hot spots can be identified. Higher screening hit rates (26 – 60 %)<sup>(136-138)</sup> were observed when screening these very small compounds in comparison with standard fragment screening. Moreover, the screens generally resulted in the identification of more binding pockets (3 – 20)<sup>(137)</sup>. It is likely that the data from this screening

approach, as opposed to solvent screening, can be more easily utilised in medicinal chemistry as a result of these considerations.

While X-ray crystallography provides atomic resolution detail of protein-ligand interactions and compound orientation, not all proteins are amenable to crystallisation. NMR spectroscopy has also been utilised as a biophysical technique for screening organic solvents <sup>(132)</sup> and provides a potential alternative method for screening low molecular weight fragments. NMR provides the ability to screen against the protein in solution, where dynamics allow the protein to sample a range of conformations, which may also affect compound binding. Screening and validating these binding events using multiple techniques offers potential advantages in identifying consistent protein hot spots and interactions that may be more successfully employed in structure-based design.

The chapter describes the design of a library of very small compounds, termed “MicroFrag”, and its screening against *EcDsbA*. The screen was conducted using both X-ray crystallography and protein-detected NMR spectroscopy. The results obtained from these two approaches are described and compared.

## 6.1 MicroFrag library design

The screening techniques described above achieve a high coverage of chemical space with a relatively small number of compounds by utilising chemically diverse compounds of ultra-low molecular weight. To identify protein hot spots by high concentration X-ray crystallography and NMR spectroscopy we designed a “MicroFrag” screening library. The MicroFrag library was designed to be enriched in chemotypes that are either commonly found in oral drugs or have been observed by crystallographic screening. In addition, the MicroFrag were selected to efficiently cover chemical space, contain a variety of pharmacophore elements and be soluble under aqueous conditions. Elements of the design and selection strategy are most similar to reported protocols for REFIL reagent selection <sup>(125, 128)</sup> and the Vertex SHAPES fragment library <sup>(274)</sup>. The REFIL reagent selection protocol uses a flexible, iterative method that considers multiple components such as compound

physicochemical properties, compound similarity, and the addition of new chemotypes that are not yet represented in the set to select optimised small libraries of compounds. This iterative selection protocol is ideally suited to small libraries (< 500 members) and allows for selections that require balancing of multiple factors and fine tuning of library properties. The Vertex SHAPES fragment library was designed to include common scaffolds and motifs already found as substructures of approved drugs. The SHAPES fragment library also forced the inclusion of heteroatoms in compounds to participate in polar interactions and to aid solubility. The final selection of compounds for the MicroFrag library used an iterative selection protocol enriching for drug-like motifs while maintaining properties similar to the Astex “MiniFrag” library <sup>(137)</sup>. The Astex MiniFrag library was designed for X-ray crystallography screening and contains 81 compounds of 5 – 7 heavy atoms (one 8 heavy atom MiniFrag), with a minimum of one heteroatom. The MiniFrag library contained polar compounds with an average ClogP of -0.4 <sup>(137)</sup>.

To be selected in the library MicroFrag ligands had to meet the criteria of containing 5 – 8 heavy atoms, with a minimum of one heteroatom, one 5- or 6-membered ring and a ClogP between -2 and 2. As in the SHAPES and MiniFrag libraries, the insistence upon a minimum of one heteroatom and ClogP between -2 and 2 gave the compounds the ability to participate in hydrogen bonds and a greater chance of aqueous solubility at the concentrations required for screening. Rings were made essential to decrease flexibility, reduce the number of possible compound orientations per binding site and aid crystallographic refinement. Rings smaller than 5 members were excluded from the selection due to their increased ring strain and lower prevalence as privileged scaffolds in drug design compared to 5- or 6-membered rings (see section 6.1.1). The MicroFrag library was biased toward compounds which were found in the Protein Data Bank (PDB) <sup>(275-277)</sup> or substructures of orally available, FDA approved therapeutics from the DrugBank database <sup>(278, 279)</sup>.

### 6.1.1 Privileged MicroFrag

It was envisaged that MicroFrag binding motifs would be identified by X-ray crystallography and be used in the optimisation of fragments into lead-like compounds.



Hence, the Protein Data Bank (PDB) <sup>(275-277)</sup> and DrugBank databases <sup>(278, 279)</sup> were mined to find privileged scaffolds/motifs common to crystallised and therapeutic compounds to inform library selection.

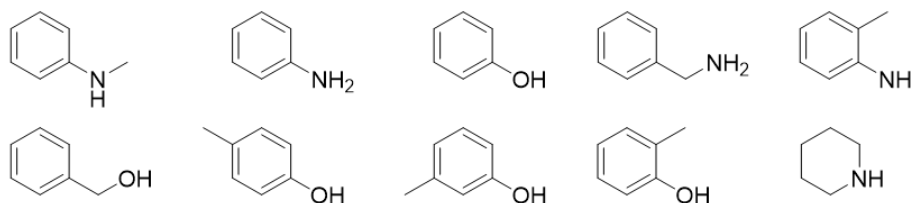
Filtering all PDB ligands for HAC 5 – 8, ClogP  $\leq 2$ ,  $\geq 1$  ring, and  $\geq 1$  heteroatom gave a list of 199 compounds. Removal of ligands by an in house manually curated set of PAINS <sup>(54, 55)</sup>, reactive functional groups (alkyl bromides, Michael acceptors etc.) and  $\geq 3$  alert structures (urea, terminal alkynes etc.) gave 147 molecules. All molecules except two had a ClogP between -2 and 2, however the profile was biased toward the larger more lipophilic compounds with the average ClogP for this set as 0.34 and an average HAC of 7.3. 93 % of the compounds contained either a 5- or 6-membered ring with the 10 additional compounds contained rings with less than five or less than six atoms or complex bridged ring systems. 39 % of the ring systems were saturated while 61 % were aromatic.

The DrugBank database contains 30,118 compounds with 2,603 of them FDA approved, orally available therapeutics. These drugs were filtered for compounds with a ClogP between -2 and 6, and fragmented into 5,250 substructures of 5 – 8 heavy atoms. 2,293 substructures contained at least one heteroatom, a ring and did not contain a PAINS motif, reactive functional groups or  $\geq 3$  alert structures. As the MicroFrag library was designed to be commercially available these substructures were filtered by their availability through the MolPort Building Blocks reagent list to give 1,033 potential compounds.

The property profiles of the potential MicroFragments from the PDB and DrugBank were very similar (Figure 6.1) with 93 and 87 % containing a 5- or 6- membered ring and an average of 7.3 and 7.4 heavy atoms in the datasets, respectively. However, the DrugBank MicroFragments showed a bias toward slightly more lipophilic compounds (PDB mean ClogP = 0.34; DrugBank mean ClogP = 0.72). 2D 2-point pharmacophores were used to compare the coverage of bonding characteristics (hydrogen bond donors and acceptors, positive and negative charges, 5- or 6-membered aromatic rings) and their number of connecting bonds in 2D space. Although the DrugBank had seven times the number of compounds as the PDB dataset, they only provided an extra six unique

2D 2-point pharmacophores, and had three fewer unique topologies. 78 molecules were considered privileged as they were found in both the PDB and DrugBank MicroFrag datasets. Of these, nine of the top ten most common ligands found in oral drugs were benzene rings substituted with a single oxygen or nitrogen containing substituent, among the simplest of possible ring containing MicroFragments. Surprisingly, Astex's MiniFrag library has only one compound containing a benzene ring (phenol) and while many rings with aniline type nitrogens, phenol type oxygens, aminomethyl and hydroxymethyl substituents are present, they are almost always attached to heteroaromatic rings. Interestingly, in the dataset of 78 privileged MicroFragments there were only two halogen containing compounds (one fluorine, one chlorine) and only one compound contained a carboxylic acid. As the datasets are restricted to 5 – 8 heavy atoms, carboxylic acids are only possibly on 5-membered rings and this severely restricts the number of possible compounds that contain a negative charge and their representation in the datasets.

Property	PDB MicroFragments	DrugBank MicroFragments
# of compounds	147	1,033
Average HAC	7.3	7.4
Average ClogP	0.34	0.72
# of unique 2D 2-point pharmacophores	45	51
# of unique topologies	20	17
% with a 5- or 6-membered ring	93 %	87 %



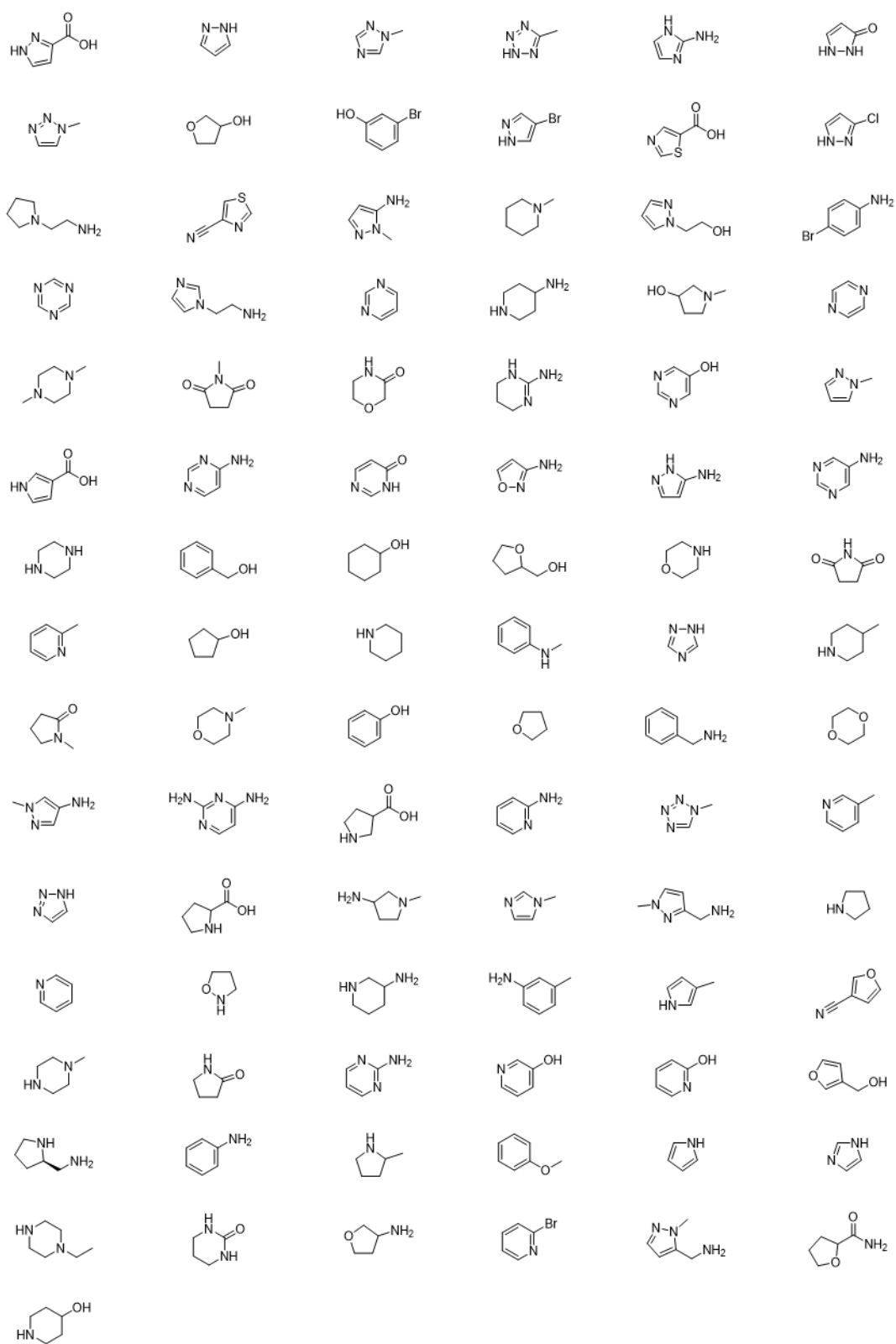
**Figure 6.1:** Privileged MicroFrag scaffolds. Property profiles for privileged MicroFrag scaffolds in the Protein Data Bank (PDB) and DrugBank database. Top 10 compounds of DrugBank database found in crystal structures from the PDB (ordered from highest occurrence, left to right).

### 6.1.2 MicroFrag library selection

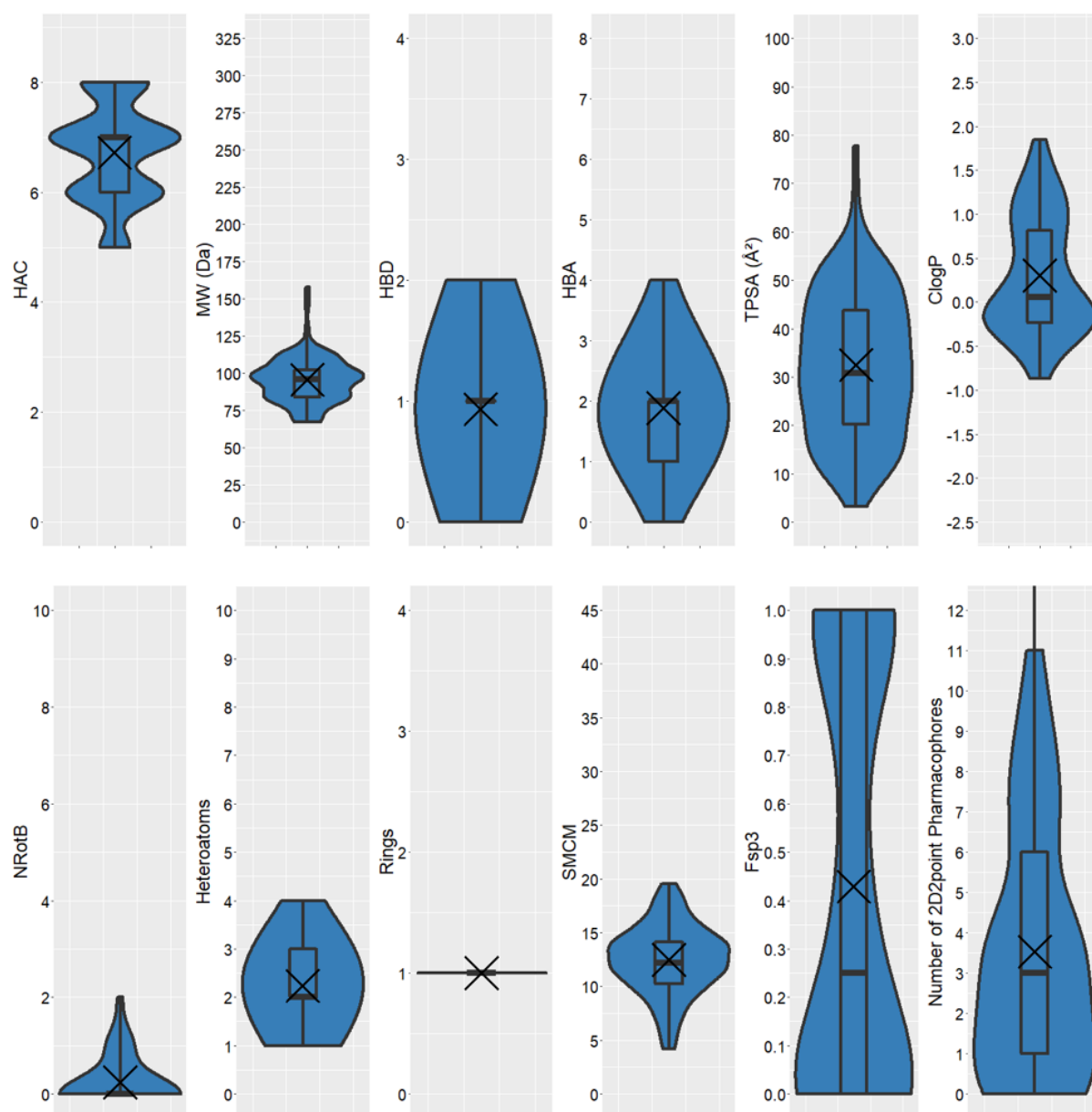
The MicroFrag screening library was designed to consist of compounds which were commercially available, highly soluble, cover as many unique topologies and 2D 2-point pharmacophores as possible with few close neighbours while also, if possible being present in oral drugs and known X-ray crystal structures. The MolPort building blocks reagent list comprising of ~350,000 compounds were refined to a list of 2560 ligands which fit all definitions of a MicroFrag (1 ring,  $\geq 1$  heteroatom, 5 – 8 heavy atoms, ClogP -2 to 2) and did not contain a reactive functional group <sup>(280)</sup>, PAINS <sup>(54, 55)</sup> or  $\geq 3$  structural alerts <sup>(281)</sup>. In addition, their physicochemical properties, 2D 2-point pharmacophores, synthetic molecular complexity (SMCM) <sup>(282)</sup> and chemical fingerprints (ECFP4 <sup>(283)</sup>, FCFP4 <sup>(283)</sup> and MACCS <sup>(284)</sup>) were calculated for use in library design and diversity analysis. An iterative selection protocol was used where a single compound was added to the library at each step. For each step, inclusion in the library was based on a selection score derived from three main criteria:

- 1) physicochemical property and PDB/drug relevance; calculated based on normalised HAC, number of heteroatoms, ClogP, TPSA, number of aromatic rings, number of halogens, the number of unique 2D-2-point pharmacophores in the compounds, whether they were present in the PDB or the fragmented DrugBank oral drug dataset.
- 2) novel diversity; measured as coverage of novel 2D 2-point pharmacophores or topologies compared to the currently selected library.
- 3) similarity/sphere exclusion; measured as maximum Tanimoto similarity based on ECFP4 and 2D 2-point pharmacophore fingerprints of the potential library member to the currently selected library.

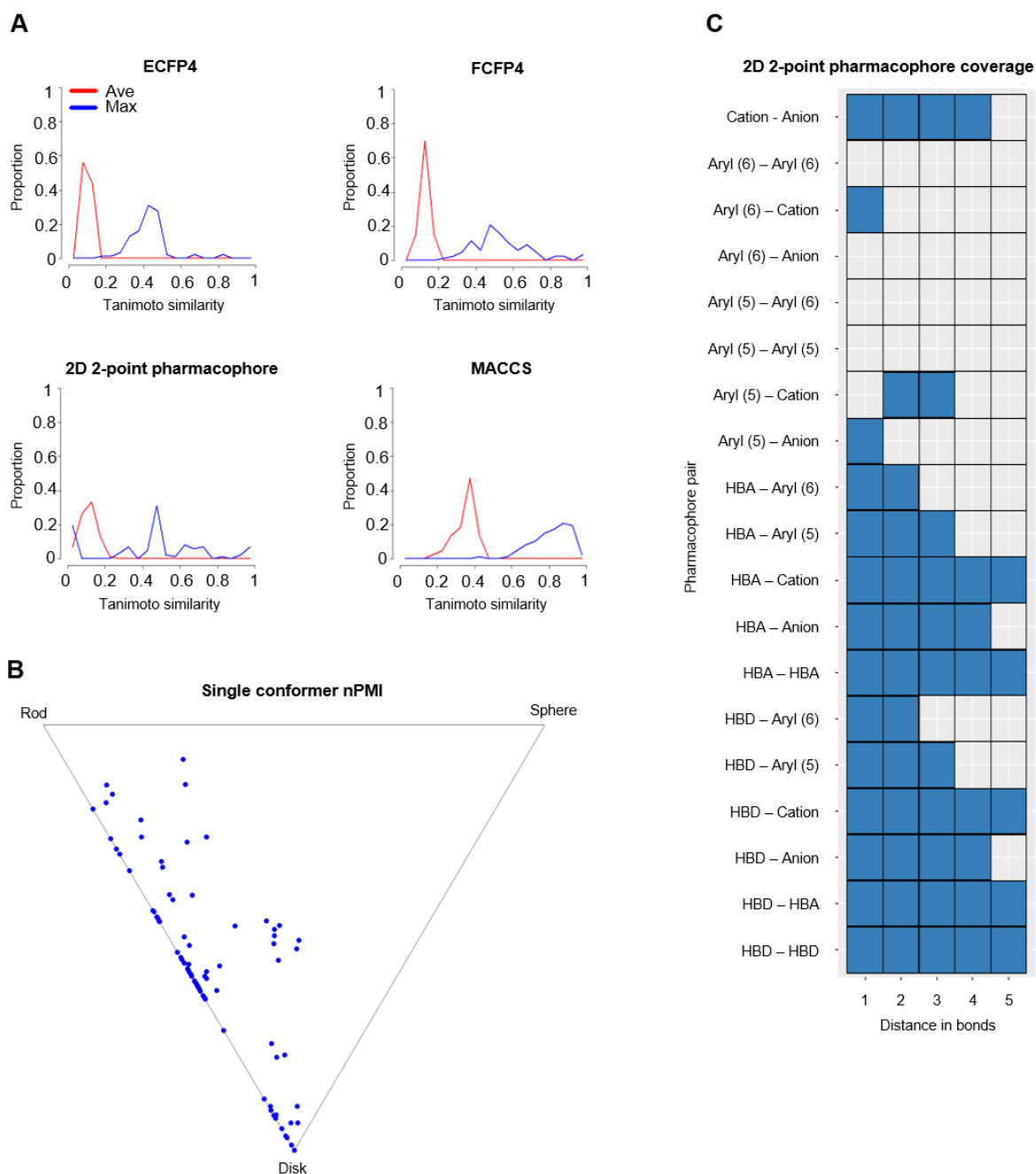
These factors were weighted and optimised based on inspection of trial library selections in order to provide an even balance of the factors (Figure 6.3, Figure 6.4, see chapter 8 for experimental detail). After purchase a 91-member MicroFrag library was obtained (Figure 6.2).



**Figure 6.2:** Selected MicroFrag library



**Figure 6.3:** Physicochemical properties of the MicroFrag library visualised as violin plots with overlaid boxplots. Abbreviations used: heavy atom count (HAC), molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), topological polar surface area (TPSA), number of rotatable bonds (NRotB), synthetic and molecular complexity (SMCM), fraction of  $sp^3$  hybridised carbons (Fsp3), ClogP (calculated octanol-water partitioning co-efficient).



**Figure 6.4:** Diversity of the MicroFrag library. A) Tanimoto similarities of the library based on the calculated fingerprint profiles. B) Normalised principle moments of inertia (nPMI) plot describing compound shape and 3D character. C) Coverage of 2D 2-point pharmacophores by all compounds in the library as a function of number of bonds separating the pharmacophore pairs. Pharmacophore pairs are described using the following abbreviations: 6-membered aromatic ring (Aryl (6)), 5-membered aromatic ring (Aryl (5)), hydrogen bond acceptor (HBA), hydrogen bond donor (HBD).

Comparing the library diversity coverage to the diversity coverage of all commercially available MicroFragments, the final library selection covered 100 % of possible 2D 2-point pharmacophores and 14 % of Murcko scaffolds <sup>(285)</sup> of this size in the MolPort reagent list. These compounds also covered 70 % of commercially available topologies, with some of the missing topologies only found in commercial reagents that contained bridged rings or involved substitution of hydrogens with fluorine or deuterium atoms. In comparison to the theoretical and commercial availability of drug-like compounds ( $\leq 36$  HAC,  $\leq 500$  Da) and fragments ( $\leq 17$  HAC), the MicroFrag library obtains a huge gain in relative chemical space coverage (Table 6.1). Although many of the ligands contain aromatic rings and are inherently planar, the library did not lie along the rod-disk-like axis in the normalised principle moment of inertia (nPMI) <sup>(286)</sup> plot and had an average fraction of Csp<sup>3</sup> of 0.43.

**Table 6.1:** Chemical space coverage of drug-like compounds, fragments and MicroFragments

		<b>“Rule of 5” compounds<sup>a</sup></b>	<b>Fragments<sup>b</sup></b>	<b>MicroFragments</b>
<b>Compounds</b>	Theoretical	$10^{33}$ <sup>(31)</sup>	$166 \times 10^9$ <sup>(32)</sup>	44,000 <sup>(34)</sup>
	Commercially available <sup>c</sup>	6,890,000 ( $0.2 - 2 \times 10^6$ )	440,000 ( $1 - 5 \times 10^3$ )	14,000 (91)
<b>Topologies</b>	Theoretical	Unknown	5,420,000	21
	Commercially available <sup>c</sup>	Unknown	50,000	17 (12)
<b>2D 2-point pharmacophores</b>	Theoretical	95	95	55
	Commercially available <sup>c</sup>	95	95	51 (51)

<b>Murcko scaffolds</b>	Commercially available <sup>c</sup>	Unknown	26,733	214 <b>(31)</b>
-------------------------	-------------------------------------	---------	--------	--------------------

<sup>a</sup> drug-like compounds ( $\leq 36$  HAC,  $\leq 500$  Da), <sup>b</sup> fragments ( $\leq 17$  HAC), <sup>c</sup> available from the MolPort catalogue. Shown in bold within brackets are the typical or included size or coverage of libraries that are used for screening in this chemical space.

## 6.2 X-ray crystallography MicroFrag screen

### 6.2.1 Assay optimisation

X-ray crystallography soaking experiments with the MicroFrag screens required optimisation to find suitable conditions that were tolerated by the crystals. A trial set of 20 MicroFrag screens was used to investigate optimal soaking conditions for *EcDsbA*. Due to the significantly higher concentrations required for the MicroFrag screen, conditions were optimised to maintain crystal stability and quality of diffraction while having enough occupancy to detect binding (estimated as  $10 \times K_D$  for *EcDsbA* X-ray crystallography).

Protein crystals can be sensitive to changes in pH, although previous experience with *EcDsbA* in the group indicated that the crystal form used for soaking of *EcDsbA* tolerates a wide pH range from pH 5 to 9. pH values outside these ranges can cause dissolution, damage or affect the diffraction quality of the crystals. Initially the *EcDsbA* cryoprotectant (24 % PEG 8000, 22 % glycerol, 100 mM sodium cacodylate, pH 6.1) was tested for its ability to resist pH changes caused by the presence of different concentrations of MicroFrag screens using a universal dye indicator assay <sup>(287)</sup>. The results indicated that 500 mM of the trial MicroFrag screens did not cause changes more than  $\pm 1$  pH unit with 250 mM and 100 mM showing less than  $\pm 0.5$  pH units. Therefore, the buffering ability of the cryoprotectant was deemed satisfactory for the assay.

*EcDsbA* crystals of the apo protein as well as co-crystals with the hydrophobic groove binder carboxylic acid diaryl ether **28** or the highest affinity binder of the diaryl ether series were used in the optimisation of the MicroFrag soaking experiments to investigate possible differences when additional higher affinity ligands are bound.



Initial conditions were based on the Astex Therapeutics MiniFrag X-ray crystallography conditions <sup>(137)</sup> and drops were prepared with MicroFrag at a concentration of 1 M in *EcDsbA* cryoprotectant (24 % PEG 8000, 22 % glycerol, 100 mM sodium cacodylate, pH 6.1). Soaking time of the apo crystals in the MicroFrag solution was varied to monitor the effects on crystal stability, quality of diffraction. Crystals were soaked for 30 seconds, 2.5 minutes, 7.5 minutes, 30 minutes and 60 minutes. Almost all apo-crystals were stable and did not dissolve or crack even after 1 hour of soaking with the trial set of MicroFrag. In the case of the preformed *EcDsbA*-diaryl ether co-crystals, soaking with 1 M MicroFrag for 1 hour revealed that two MicroFrag appeared to affect the crystal stability. However, upon acquisition of diffraction data electron density was observed for only 3/20 (15 %) of the tested MicroFrag in the apo crystals. In the case of the co-crystals, either no diffraction was observed or there was no density for the MicroFrag. The Astex Therapeutics MiniFrag crystallography screens suggested that a higher hit rate would be obtained <sup>(137)</sup> and so it was hypothesised that this may have been caused by a low MicroFrag solubility in the relatively non-polar cryoprotectant.

The MicroFrag solubility in the *EcDsbA* cryoprotectant was assessed using semi-quantitative 1D <sup>1</sup>H NMR. The MicroFrag samples were prepared as per the trial soaks, diluted in *d*<sub>6</sub>-DMSO and analysed by semi-quantitative 1D <sup>1</sup>H NMR with 1 mM DSS internal standard. These spectra showed that only three of the MicroFrag tested were visible in the cryoprotectant and happened to be the three MicroFrag for which electron density was observed. These results confirmed that the solubility of the MicroFrag in the cryoprotectant solution was an issue in the previous crystallisation attempts. Furthermore, issues with sample preparation were identified upon visual inspection of the 96-well plates where it appeared that some MicroFrag may have evaporated. The initial plates were prepared by making 1 M MicroFrag stocks in methanol, transferring the required volume to the 96-well plate and removing the solvent under vacuum. The samples were prepared again and were left at atmospheric pressure to evaporate to minimise possible MicroFrag evaporation which, upon visual inspection, showed that some sample remained in all wells. Subsequently, the MicroFrag were no longer concentrated under vacuum. To optimise buffer solubility the MicroFrag were dissolved in either the *EcDsbA* cryoprotectant (with and without

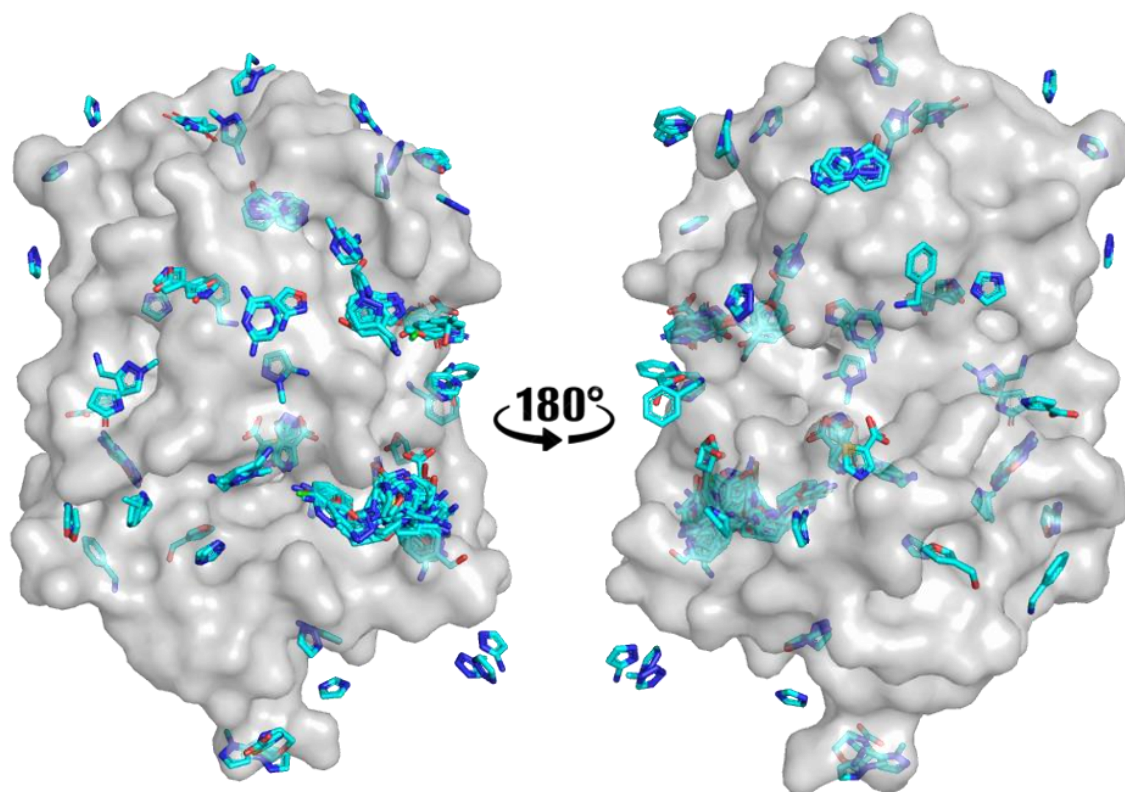
10 % methanol), or the crystallisation mother liquor (13 % PEG 8000, 7.5 % glycerol, 1 mM CuCl<sub>2</sub>, 100 mM sodium cacodylate pH 6.1). Samples of this soaking solution were taken and diluted into *d*<sub>6</sub>-DMSO and analysed as above. The MicroFrag in the cryoprotectant with methanol or in the mother liquor had > 2-fold improved solubility than the cryoprotectant alone. As a result, the MicroFrag samples were prepared using 10 % methanol in the crystallisation mother liquor.

The soaking time of MicroFrag with apo *EcDsbA* crystals was revisited, however, with the solubility issue now resolved the crystals were generally not viable for longer than 15 – 30 minutes. The effect of MicroFrag concentration in the soaking experiments was evaluated at 200 mM, 1 M and 2 M, using the new sample preparation process in crystallisation mother liquor with 10 % methanol. Slightly lower occupancies were observed for the samples containing 200 mM MicroFrag, however, there was no discernible difference for the density obtained at either 1 or 2 M and the final screening concentration remained at 1 M. The optimised soaking conditions also afforded a higher hit rate of 14/20 (70 %) compared to experiments conducted with the cryoprotectant alone (15 %), which is more in line with expected MicroFrag results based on Astex's data.

The full 91-member MicroFrag library was soaked into apo oxidised *EcDsbA* crystals for 5 – 15 minutes at a 1 M MicroFrag concentration using the optimised protocol. Despite optimisation efforts not all crystals were stable in the presence of all MicroFrag at 1 M in the soaking buffer (13 % PEG 8000, 7.5 % glycerol, 1 mM CuCl<sub>2</sub>, 100 mM sodium cacodylate pH 6.1 and 10 % methanol). The MicroFrag which were observed to dissolve or damage the crystals were commonly found to be at a pH of < 3 or > 10 and were adjusted to a pH between 5 – 9. If the soaking buffer with 1 M MicroFrag was already in this pH range, or still damaged the crystal after adjustment, the concentration of the MicroFrag was halved (500 mM). These adjustments allowed for collection of datasets for every ligand in the MicroFrag library.

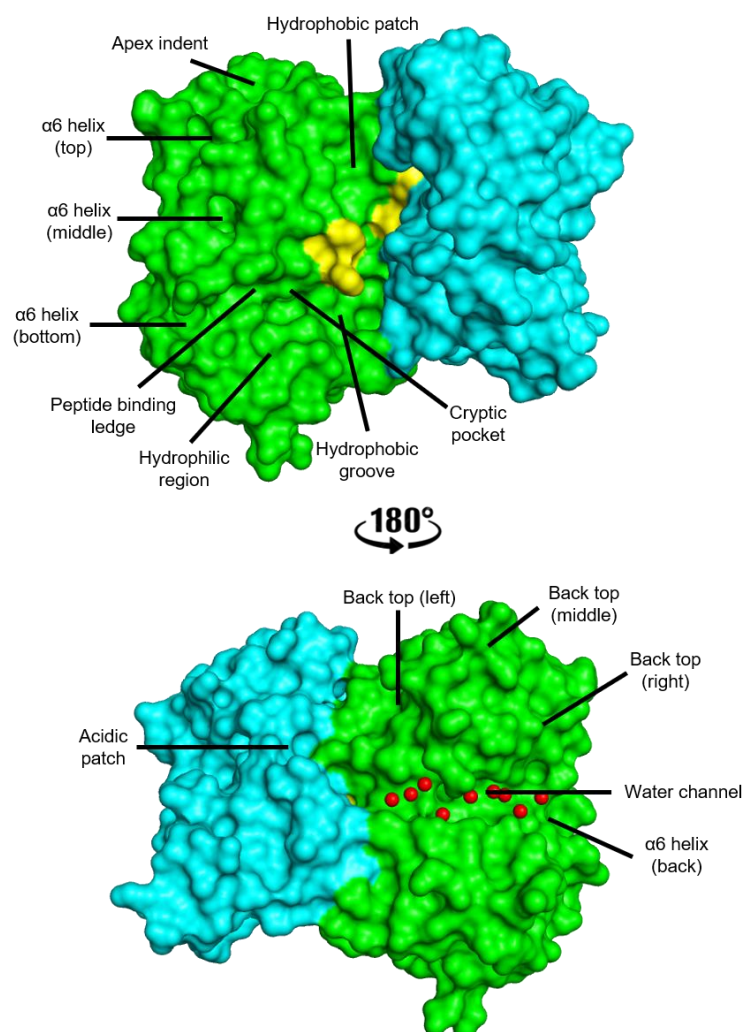
### 6.2.2 MicroFrag screen by X-ray crystallography

The X-ray crystallography screen of the MicroFrag library yielded structures of 47 compounds and 180 poses with an overall hit rate of 52 % (Figure 6.5). All structures had resolutions between 1.47 – 1.88 Å and resulted in an average ligand  $\beta$  factor of  $< 50 \text{ Å}^2$ . Determining the presence and orientation of the MicroFrag was significantly simpler compared to the solvent screen. The increased compound size and varied shapes gave a clear distinction between MicroFrag and water molecules and resulted in more confidence in the MicroFrag poses.

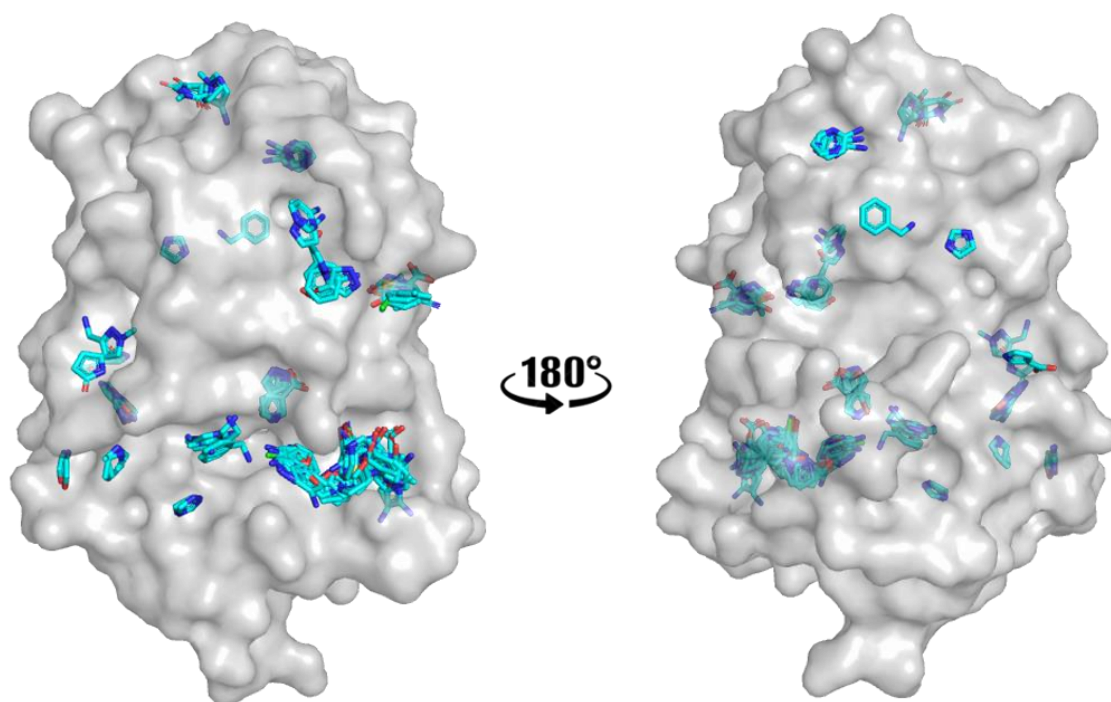


**Figure 6.5:** Overlay of all MicroFrag structures solved by X-ray crystallography. MicroFrag shown as coloured sticks, oxidised *EcDsbA* shown as transparent grey surface.

In this form of *EcDsbA* crystals two monomers are observed in the asymmetric unit. This, along with crystal packing, results in different solvent exposed areas for each monomer and, although the interface blocks part of the protein surface, all previously described binding sites remained accessible by ligands in at least one of the monomers (Figure 6.6). These contacts also cause compounds to potentially bind in pockets which do not occur naturally in solution, which is considered to be one of the limitations of X-ray crystallography. To avoid biasing hot spot identification due to crystal packing artefacts any ligands which did not make any interactions to the protein or made interactions (hydrogen bonds, halogen, ionic, hydrophobic contacts or  $\pi$ - $\pi$  stacking) with more than one *EcDsbA* molecule were removed, resulting in 112 poses for analysis (Figure 6.7). These compounds made a total of 295 interactions (hydrogen bonds, halogen, ionic, hydrophobic contacts or  $\pi$ - $\pi$  stacking) with the protein. 223 of these interactions were made directly between the ligand and protein and the remaining 72 were water mediated hydrogen bonds.



**Figure 6.6:** Crystallographic asymmetric unit of EcDsbA. Surface contacts of the two monomers (coloured as cyan and green surfaces) of *EcDsbA* with active site ( $^{30}\text{CPHC}^{33}$ ) highlighted in yellow and water channel shown as red spheres.



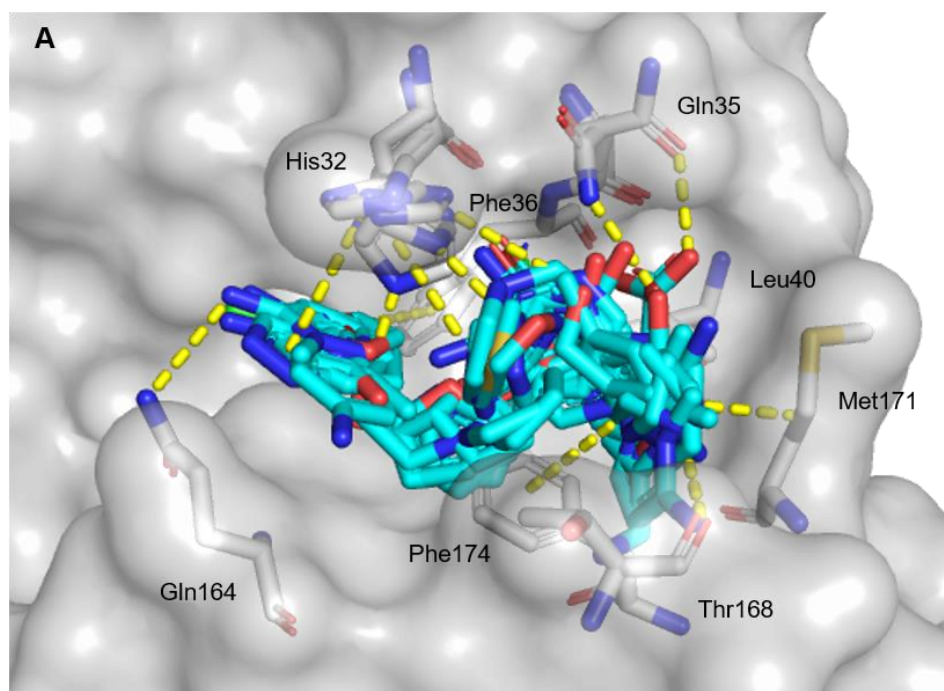
**Figure 6.7:** Overlay of analysed MicroFrag structures solved by X-ray crystallography. Binding poses are shown for MicroFrag (cyan sticks) which did not interact with both of the *EcDsbA* monomers and are a member of a binding site occupied by  $\geq 2$  MicroFrag.

All remaining poses were overlaid and hot spots were identified based on having  $\geq 2$  MicroFrag bound to the site. 11 hot spots were identified using this technique and two hot spots, an extension of the hydrophobic patch and at the middle top back of the protein had not been previously described. Single MicroFrag were found to bind at the active site and the back top left of the protein and no densities were found in the acidic patch, water channel, or the top of the  $\alpha 6$  helix. The most occupied hot spot was the hydrophobic groove where 26 different MicroFrag were found to bind (29 % hit rate), with many showing multiple poses and interactions (Figure 6.8a). Interactions made by the MicroFrag were similar to those made by optimised fragment binders. Hydrophobic  $\pi$ - $\pi$  stacking interactions with Phe36, Phe174 and His32 were the most common, followed by hydrogen bonds to the sidechains of Gln35, Gln164 and the main chain of Thr168, all of which have been previously observed with fragment analogues. Two MicroFrag with two poses each were found bound in the cryptic

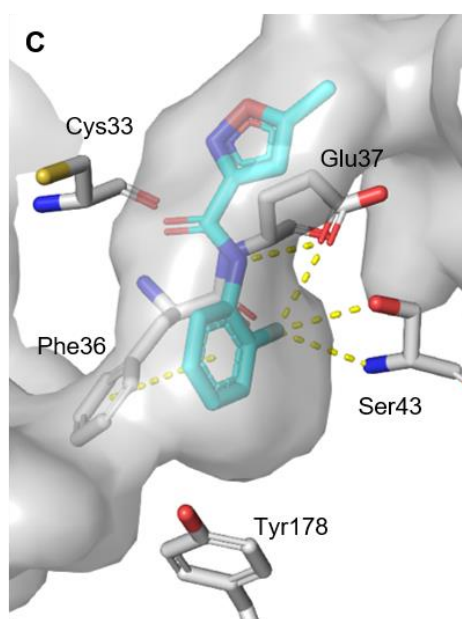
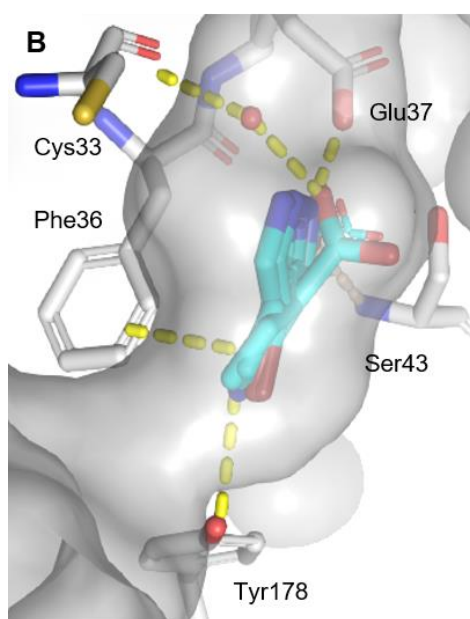
pocket and these overlapped with the lower ring of the known cryptic pocket binders. However, no MicroFragments were found where the upper isoxazole ring of cryptic pocket binders is located (Figure 6.8b, c). During medicinal chemistry development of the cryptic pocket fragments replacement of the isoxazole ring was poorly tolerated <sup>(233)</sup> and possibly reflects the change in conformation of *EcDsbA* required to bind at the upper region of this hot spot. Furthermore, although the MicroFragments made hydrogen bond donor interactions to Glu37, which the fragment analogues also make, the MicroFragments make different interactions with conserved  $\pi$ - $\pi$  stacking to Phe36, a hydrogen bond donor interaction or halogen bonding to Tyr178 side chain, a conserved water mediated interaction with the carbonyl of active site Cys33 and a singular hydrogen bond acceptor interaction with backbone NH of Ser43. Some of these interactions could be incorporated into the design of future cryptic pocket binding analogues.

MicroFragments binding to the peptide binding ledge make hydrogen bond donor interactions with Arg148, Val150 and Pro151 backbone carbonyls as well as Gln160 side chain (Figure 6.9a). In the hydrophilic region two MicroFragments containing a 1-3 diaza-5-membered heteroaromatic motif make two hydrogen bonds to Thr10 and Leu161 (Figure 6.9b). This binding interaction is also found with urea in the solvent screen where a similar pharmacophore and interactions are observed. Two ligands were found at the middle of the  $\alpha$ 6 helix, however, no conserved interactions were made (Figure 6.9c). Binding to the hot spot at the bottom of  $\alpha$ 6 helix shows two distinct interaction locations (Figure 6.9d). The first forms a hydrogen bond donor interaction to Gln146 backbone, where the second makes hydrogen bond interactions with the backbones of both Lys14 and Asp144. The back of the  $\alpha$ 6 helix bound gave 15 MicroFrag poses (Figure 6.9e). Many conserved water-mediated hydrogen bonds were made in this hot spot to Pro20, Lys140, Asp144, Phe154, Asn156 and Gly157 as well as a direct hydrogen bond to the main chain of Ala141 and  $\pi$ - $\pi$  stacking interactions were formed with Phe25 and Phe154.





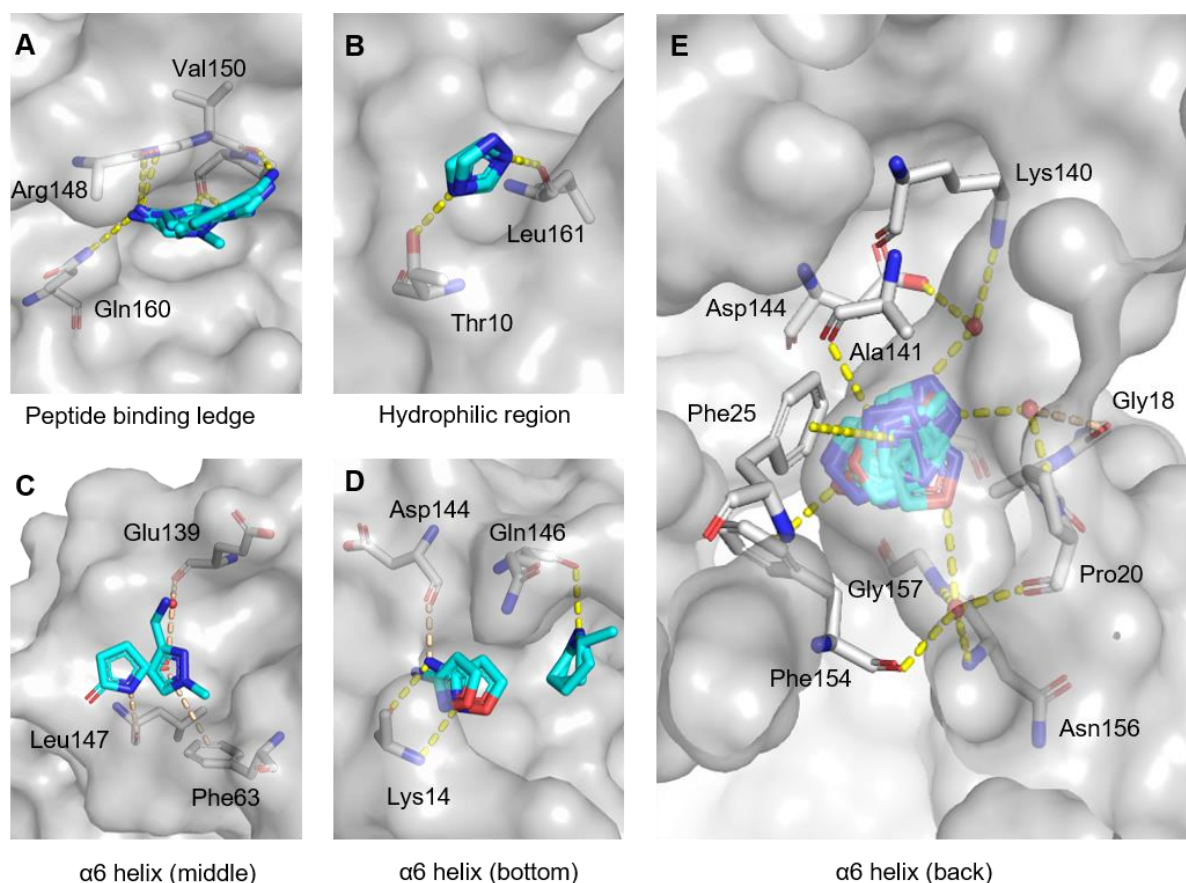
Hydrophobic groove



Cryptic pocket

**Figure 6.8:** Overlay of MicroFrag crystal structures and their protein interactions at the known fragment binding sites. A) MicroFrag interactions with residues of the hydrophobic groove. MicroFrag shown as cyan sticks, conserved interactions as yellow dashes and *EcDsbA* shown as grey sticks and grey surface. Interactions with residues of the cryptic pocket with B) MicroFrag and C) a known cryptic pocket binder. MicroFrag and fragment binder shown as cyan sticks, conserved interactions as yellow dashes, interactions made by single compounds as tan dashes, water molecules as red spheres and *EcDsbA* shown as grey sticks and grey surface.

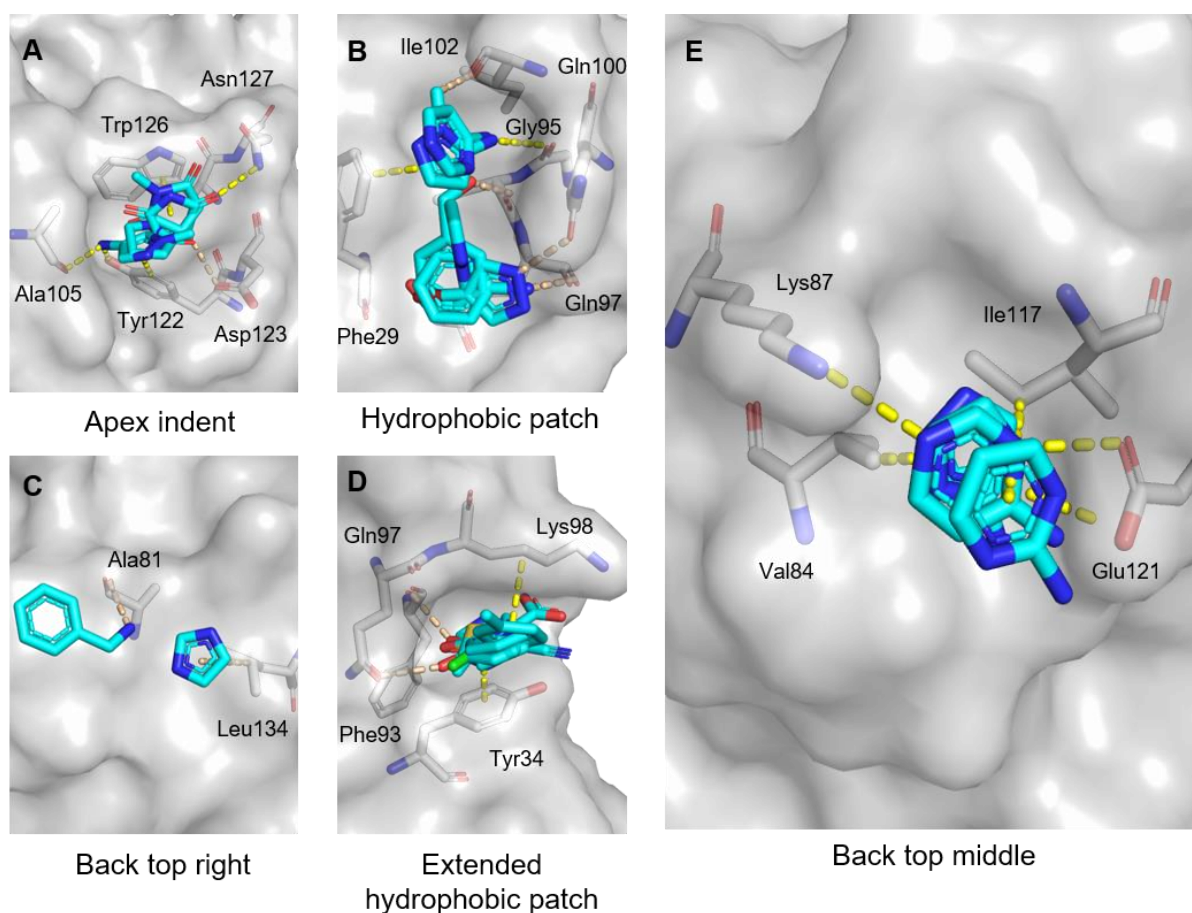




**Figure 6.9:** Overlay of MicroFrag crystal structures and their protein interactions at binding hot spots A) peptide binding ledge, B) hydrophilic region, C)  $\alpha 6$  helix middle, D)  $\alpha 6$  helix bottom and E)  $\alpha 6$  helix back which bind  $\geq 2$  ligands. MicroFrag shown as cyan sticks, conserved interactions as yellow dashes, interactions made by single compounds as tan dashes, water molecules as red spheres and *EcDsbA* shown as grey sticks and grey surface.

Binding to the protein apex is facilitated by two  $\pi$ - $\pi$  interactions with Tyr122 and Trp126, a hydrogen bond to Ala105 (either direct or water mediated) and a hydrogen bond donor interaction with Asn127 (Figure 6.10a). An additional one-off hydrogen bond is made to the sidechain of Asp123. Large clusters of ligands are found at the hydrophobic patch and at an extension of this site to toward the acidic patch. The main section of the hydrophobic patch features conserved  $\pi$ - $\pi$  stacking interactions with Phe29 and a hydrogen bond donor interaction with the backbone of Gly95 (Figure 6.10b). The extension of this site conserves hydrophobic interactions with Tyr34 and the alkyl chain of Lys98 (Figure 6.10d). Taken together however, these hot spots show a conserved hydrogen bond to the sidechain Gln97. The other newly identified hot spot on the back of the protein at the top middle above the water channel

has conserved hydrogen bonds to the sidechains of Lys87 and Glu121 and hydrophobic contacts with Val84 and Ile117 (Figure 6.10e). While two MicroFrag bind to the back top right of the protein, there were no conserved interactions at this site (Figure 6.10c).



**Figure 6.10:** Overlay of MicroFrag crystal structures and their protein interactions at binding hot spots A) apex indent, B) hydrophobic patch, C) back top right, D) extended hydrophobic patch and E) back top middle which bind  $\geq 2$  ligands. MicroFrag shown as cyan sticks, conserved interactions as yellow dashes, interactions made by single compounds as tan dashes, water molecules as red spheres and *EcDsbA* shown as grey sticks and grey surface.

It is clear from these studies with MicroFrag that *EcDsbA* presents many possible hot spots across the protein surface and with residues within small pockets. Among the 11 identified ligand clusters, the peptide binding ledge, hydrophilic region, the hydrophobic patch and its extension are the closest to the fragment series currently

under development. While the conserved hydrogen bond donor interaction to the peptide binding ledge and Thr10-Leu161 di-aza motifs in the hydrophilic region can potentially be targeted from the hydrophobic groove and cryptic pocket, the hydrophobic patch hot spots are located 10 – 14 Å away from the hydrophobic groove binders and would appear to be inaccessible to the cryptic pocket binders. While polar interactions are preferred for boosting affinity and creating selective binders, MicroFrag crystallography clusters also highlight the importance of hydrophobic  $\pi$ - $\pi$  interactions as they are found in all hot spots except the peptide binding ledge, hydrophilic region, bottom of the  $\alpha$ 6 helix and the back top right of the protein. These hydrophobic interactions were not seen consistently with solvent screening, most likely due to the lack of hydrophobic motifs in the solvent set. If analogues properties can be suitably maintained it would be remiss to exclude these conserved interactions from future compound designs.

X-ray crystallography screening of the MicroFrag library identified hot spots and interactions at a similar rate to those in similar screening campaigns <sup>(136-138)</sup>. As some targets are not amenable to crystallography, we wanted to test the capacity to extend this approach to screening by  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR, using the same approach described in chapter 5 for the identification of hot spots by solvent screening.

## **6.3 $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR MicroFrag screen**

### **6.3.1 Assay optimisation**

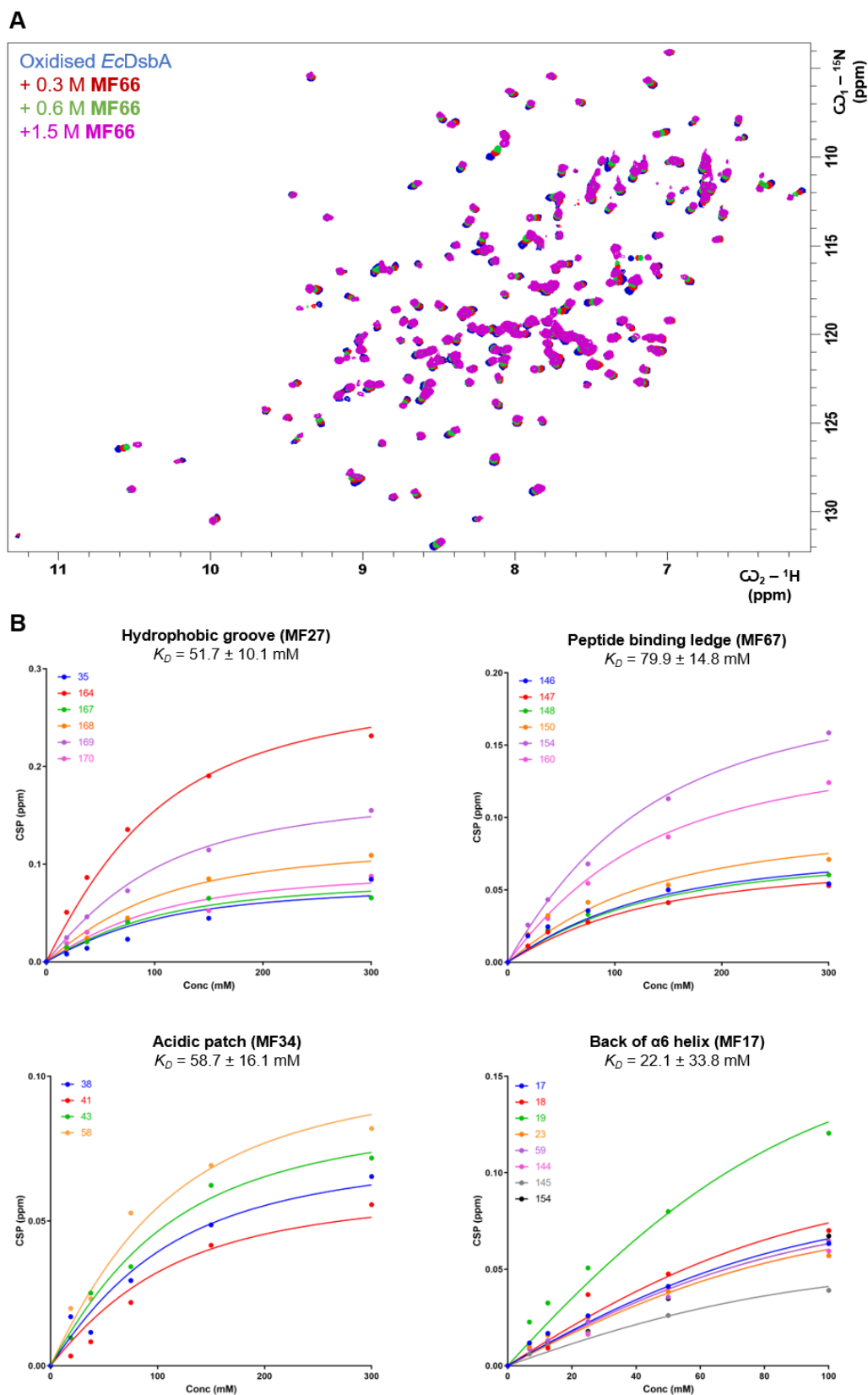
$^1\text{H}$ - $^{15}\text{N}$  HSQC NMR experiments with *EcDsbA* were optimised with a test set of MicroFrag under a variety of conditions to investigate protein stability, the ability to detect binding, and the quality of the spectra at the high ligand concentrations required for MicroFrag screening.

Denaturation and/or precipitation of *EcDsbA* was at the forefront of concerns for this assay. Initially buffering conditions were examined in order to limit the denaturation of protein due to large changes in pH upon addition of ionisable MicroFrag, or

precipitation caused by the pH approaching the isoelectric point of *EcDsbA* (theoretical  $pI = 5.4$ ). As this library is designed to be target agnostic, not all targets will use the same buffer conditions and therefore the compound stocks were prepared in pure water. These stocks were manually pH adjusted as close as possible to pH 7; generally, stocks were within  $\pm 0.5$  pH units, however, due to the lack of buffering capacity of water this was occasionally difficult and was not possible for all MicroFrag.

Previous NMR experiments recorded on *EcDsbA* have typically used a buffer system comprising either 50 mM HEPES and 25 mM NaCl or 50 mM phosphate and 25 mM NaCl at a pH of 6.8. These buffers were tested at normal and double concentrations for their ability to maintain the desired pH after the addition of the MicroFrag stocks. The desired pH is 6.8 and HEPES has the greatest buffering capacity over the pH range of 6.8 – 8.2, while phosphate buffer has a wider buffering capacity over the pH range of 5.7 – 8.0. This was reflected in the observation that phosphate buffer was superior to HEPES in controlling pH upon addition of the ligands. Doubling the concentration of the phosphate buffer (100 mM phosphate, 50 mM NaCl, pH 6.8) improved the buffering capacity and this buffer was used in the final screening of the MicroFrag library. The resonances of some residues in *EcDsbA* are very sensitive to pH and although the buffer system generally maintained the samples close to the desired pH, samples were manually readjusted to a pH of  $6.8 \pm 0.2$  prior to acquisition where necessary.

High concentrations of MicroFrag were expected to be required to achieve high enough occupancy at the binding sites to detect CSP in HSQC spectra. However, high concentrations can have detrimental effects on protein stability and spectrum quality. A range of compound concentrations from 1 mM to 2.5 M was used to probe the effects of MicroFrag ligand concentration on the assay and determine at what concentration significant chemical shifts can be observed as well as where spectral quality may degrade. The high concentrations of MicroFrag caused some 1D  $^1H$  NMR experiments to fail due to receiver overflow, even when receiver gain was at the lowest possible setting.  $^1H$ - $^{15}N$  HSQC experiments did not have the same issue due to the low natural abundance of  $^{15}N$  in the MicroFrag compounds (0.4 %) and high-quality HSQC spectra were obtained even at the highest MicroFrag concentrations tested.

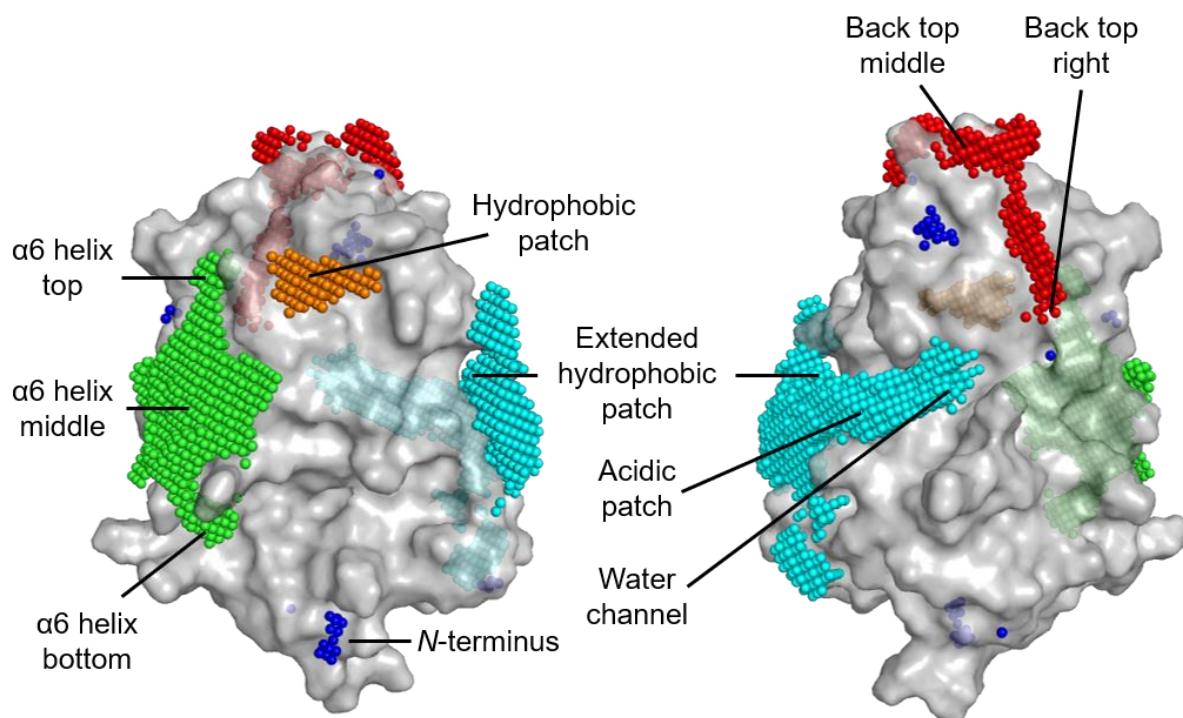


**Figure 6.11:** MicroFrag titrations. A)  $^1\text{H}$ - $^{15}\text{N}$  spectra of oxidised  $^{15}\text{N}$  *EcDsbA* (100  $\mu\text{M}$ , blue), in the presence of 0.3 M (red), 0.6 M (green) and 1.5 M (magenta) **MF66**. B) Binding isotherms obtained for the hydrophobic groove, peptide binding ledge, acidic patch and the back of the  $\alpha 6$  helix by  $^1\text{H}$ - $^{15}\text{N}$  HSQC MicroFrag titrations.

The maximum concentration of MicroFrag that could be tested was generally found to be limited by the aqueous solubility of the MicroFrag, rather than protein stability or spectral quality. For example, high quality spectra were observed for *EcDsbA* upon addition of 2 M succinimide, however, other trial MicroFrag showed maximum solubilities between 100 – 500 mM in the phosphate buffer, limiting the highest concentration at which they could be screened. To estimate the binding affinity of the MicroFrag, a subset of the compounds were titrated against *EcDsbA* (5-point, 2-fold dilution series from 100 mM or 300 mM) and  $K_D$  values in the range of 10 – 100 mM were observed for multiple binding sites (Figure 6.11). These affinities suggested that ligand concentrations between 100 mM – 500 mM were sufficient to achieve occupancies of > 50 %, which was judged to be sufficient to induce CSPs that could be measured reliably. A final concentration of 300 mM was selected for screening of the MicroFrag library as the best condition to ensure protein stability, ligand solubility and hot spot occupancy.

### 6.3.2 MicroFrag screen by $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR

A  $^1\text{H}$ - $^{15}\text{N}$  HSQC was acquired for oxidised *EcDsbA* in the presence of each MicroFrag at a concentration of 300 mM. Two MicroFrag screening samples had significantly reduced peak intensity in the HSQC and a further 21 compounds caused large global chemical shift perturbations. These 23 MicroFrag were excluded from the subsequent analysis. The analysis of binding site hot spots was conducted using the grid point generation, summation and clustering scripts as described in Chapter 5 (Figure 6.12).



**Figure 6.12:** NMR predicted binding sites of MicroFrag. Binding sites were predicted from grid point clusters (coloured spheres) calculated by the summation of individual MicroFrag grid point clusters and mapped onto the surface of oxidised *EcDsbA* (PDB ID: 1FVK, grey surface). The individual grid point clusters were calculated from  $^1\text{H}$ - $^{15}\text{N}$  HSQC chemical shift perturbations upon addition of each MicroFrag to *EcDsbA*.

In total, the  $^1\text{H}$ - $^{15}\text{N}$  HSQC clusters for the individual MicroFrag identified 15 potential hot spots. Seven of these hot spots were identified upon the summation of the grid point clusters of all the individual MicroFrag. One new hot spot that had not been observed in the solvent screening was identified at the *N*-terminus of *EcDsbA* (Figure 6.12) and this was found in the analysis of both the individual and summation clusters. There were four hot spots identified in the combined solvent screening methods (Chapter 5) that were not identified by the NMR MicroFrag screen. These were the cryptic pocket, active site, back top left, and the apex indent. Furthermore, two hot spots from the solvent screen, hydrophilic region and the back of  $\alpha 6$  helix, showed very few cases ( $< 10\%$ ) of MicroFrag binding. This may be rationalised in part because the MicroFrag screen resulted in larger CSPs than the organic solvent screen, which meant that the definitions of clusters were stricter and now required more adjacent grid points above a minimum value threshold to be classified as a hot

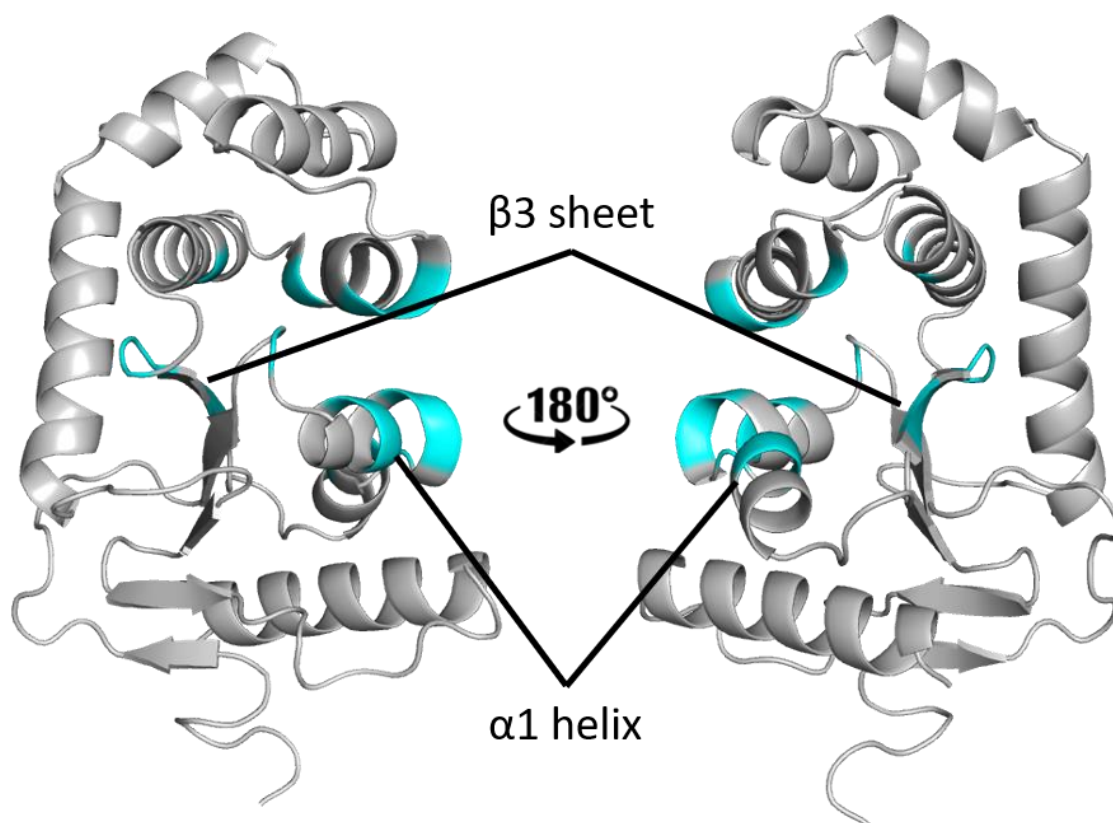


spot. It is likely that the cryptic pocket and hot spot at the back of the  $\alpha 6$  helix were not identified in the clusters due to the small number of adjacent grid points within the site, thereby disqualifying them from the formation of grid point clusters.

The hydrophobic groove is the main substrate and fragment binding pocket, and it was surprising that only 21 % of the MicroFrag gave rise to grid point clusters at this site. The centre of the hydrophobic groove was not often identified by the grid point clusters. However, Ser169 – Val173 and Leu40 of the groove were commonly perturbed and gave rise to clusters around these residues, however on the opposite side of the residues to the hydrophobic groove. If these clusters were considered to be indicative of binding to the hydrophobic groove, this resulted in 97 % of MicroFrag showing evidence of binding at this site, which is more in line with expectations.

The most frequently identified hot spots were the middle of  $\alpha 6$  helix, the water channel, the acidic patch, the extension and the main section of the hydrophobic patch, which were identified in 64 – 99 % of spectra. 18 residues within these sites had a high occurrence (> 20 % of MicroFrag spectra) of perturbations above 0.04 ppm and were generally within close proximity of each other. Many of these residues form parts of multiple different hot spots and as a result they form a continuous stretch of highly perturbed residues from the middle of the  $\alpha 6$  helix, to the hydrophobic patch and around the side of the protein to the acidic patch and water channel. Furthermore, 14/18 of these residues are located on or are directly adjacent to the active site  $\alpha 1$  helix or the  $\beta 3$  sheet which runs through the protein from the water channel to the middle of the  $\alpha 6$  helix (Figure 6.13). It is therefore possible that their high hit rate may be due to altered protein dynamics from compound binding at one position of the helix or sheet and causing CSPs along the secondary structure motif.





**Figure 6.13:** Residues from the top 5 predicted binding sites with chemical shift perturbations frequently  $> 0.04$  ppm. Important residues (cyan) are mapped onto the surface of *EcDsbA* (PDB ID: 1FVK, grey cartoon).

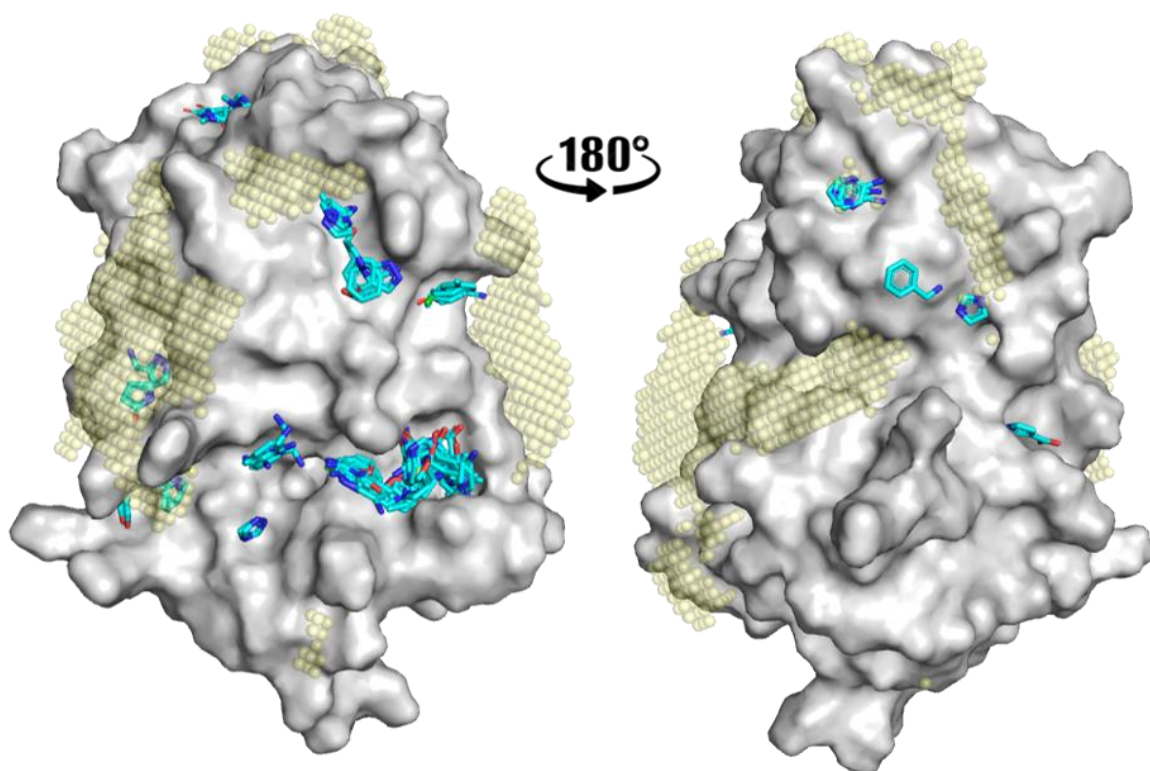
The  $^1\text{H}$ - $^{15}\text{N}$  HSQC MicroFrag screen was able to identify multiple binding hot spots also identified by the solvent screens and the MicroFrag crystallography screen. Furthermore, the NMR MicroFrag screen identified a binding hot spot not identified by any of the other screens. While the results require a more involved interpretation of binding events, the results suggest that this approach is able to provide valuable hot spot information for proteins which are not amenable to X-ray crystallography.

## 6.4 Assay comparisons

### 6.4.1 Comparison of the $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR and X-ray crystallography MicroFrag screens

This MicroFrag library was designed for use in both X-ray crystallography and  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR in order to extend the current applications of similar assays to proteins which are unable to be crystallised or crystal systems are not amenable to high ligand concentrations. Binding was found for 47 and 67 ligands in X-ray and NMR screens, respectively, together identifying 12 binding sites overall. X-ray and NMR MicroFrag screens had six hot spots represented in both screens (the middle and bottom of the  $\alpha 6$  helix, the hydrophobic patch and its extension, and the back top middle and right of the protein). Among the most populated hot spots, binding was not observed to the peptide binding ledge or hydrophilic region in the NMR screen. Similarly, no MicroFrag were found in the water channel by X-ray crystallography. Neither NMR nor X-ray crystallography screening identified MicroFrag binding to the active site (Figure 6.14).

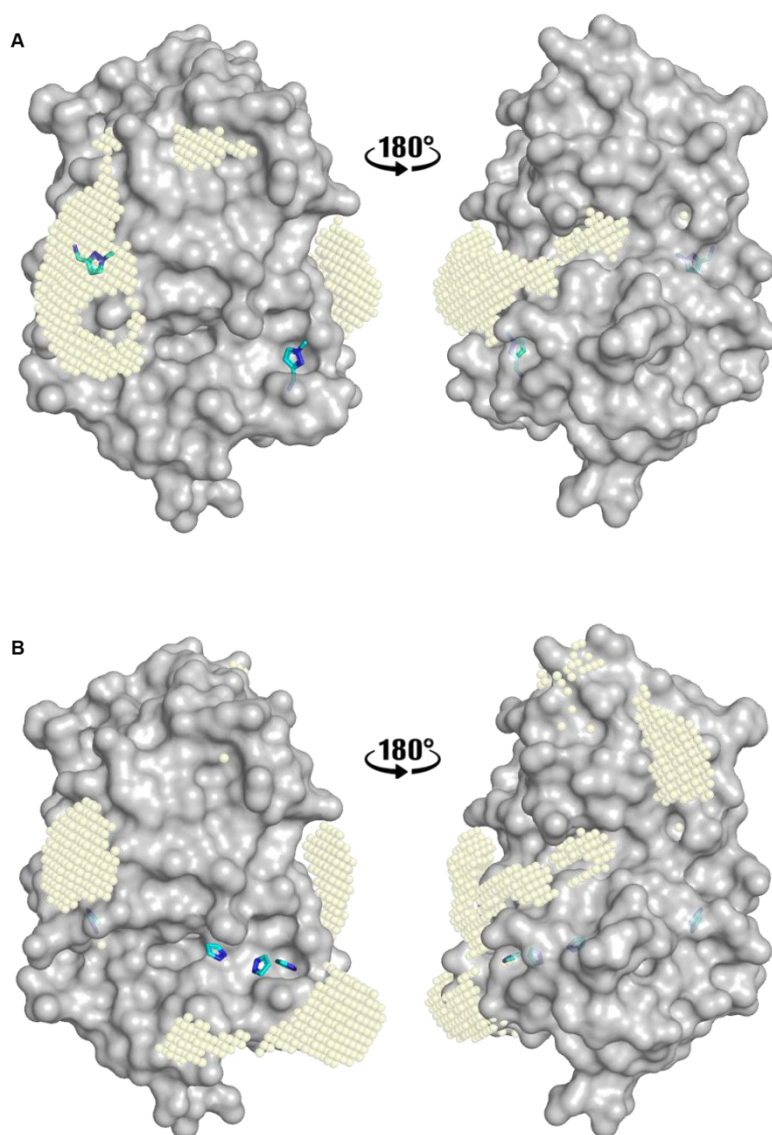
Each MicroFrag crystal structure showing binding was overlayed with their corresponding HSQC NMR grid point clusters and binding profiles obtained in each assay were compared. X-ray poses were found in the same binding hot spot as an NMR cluster 28 % of the time, with 14 % of poses directly overlaying with grid points. Conversely, only 5 % of NMR grid point cluster binding sites were also occupied by a ligand solved by crystallography with 3 % of clusters having a direct overlap of the X-ray crystallography pose with the NMR grid points. On an individual MicroFrag level, there was poor correlation of NMR and X-ray crystallography data. However, when a similar comparison was done with the summation of NMR grid point clusters and the overlayed set of crystal structures, 58 % of binding hot spots are identified in both approaches. Of these 50 % have direct or partial overlap of NMR grid points and X-ray crystallography binding poses. Direct overlays were observed at the middle and the bottom of the  $\alpha 6$  helix, at the top of the hydrophobic patch, the extended hydrophobic patch, the back top middle and back top right of the protein.



**Figure 6.14:** Comparison of global MicroFrag X-ray crystallography poses and corresponding NMR predicted binding sites. NMR grid point clusters (transparent yellow spheres) are overlaid onto binding poses of all crystallography MicroFrag clusters to show hot spot agreement between the two techniques.

MicroFragS which bound within the hydrophobic groove by X-ray crystallography often did not overlay with NMR grid points, however, these MicroFragS did show NMR clusters in the acidic patch and water channel (examples provided in Figure 6.15). This is possibly due to changes in the conformation or dynamics of *EcDsbA*, where binding in the hydrophobic groove results in inferred chemical shifts being observed at the acidic patch and water channel on the opposite face of the protein. A similar phenomenon has been observed previously for fragments binding to the hydrophobic groove (Chapter 4). Furthermore, X-ray crystallography did not find any MicroFragS in the acidic patch or the water channel. The asymmetric unit does not obscure access to these areas in X-ray crystallography (Figure 6.6) and the reason for lack of MicroFragS found bound to these hot spots which have been found by multiple other techniques is not clear. Follow-up studies where longer or higher concentration soaks or co-crystallography could be performed may validate these results with MicroFragS that show CSP for residues at these sites in the HSQC NMR spectra. This analysis

was not performed due to their significant distance from the hydrophobic groove fragment binding site and are unlikely to provide useful information on conserved interactions that may be targeted by novel analogues.



**Figure 6.15:** Comparison of individual MicroFrag X-ray crystallography poses and corresponding NMR predicted binding sites. NMR grid point clusters (transparent yellow spheres) are overlaid onto binding poses of A) **MF37** and B) **MF6** (cyan sticks) to show hot spot agreement between the two techniques.

Residues that showed interactions to MicroFags by X-ray crystallography also had CSP > 0.04 ppm in > 10 HSQC spectra 47 % of the time. These overlapping residues can be considered validated hot spot interaction points and those in close proximity to the known analogue binding sites may be used to prioritise expansion vectors and target interactions (Table 6.2).

**Table 6.2:** Correlation of X-ray crystallography interactions and NMR chemical shift perturbations (CSP) by binding site.

<b>Binding site</b>	<b>X-ray interactions with matched NMR CSP</b>	<b>NMR CSP with matched X-ray interactions</b>	<b>Conserved residues</b>
<b>Hydrophobic groove</b>	4/8	4/6	Gln35, Leu40, Gln164, Thr168
<b>Cryptic pocket</b>	2/5	2/6	Cys33, Ser43
<b>Peptide binding ledge</b>	1/5	1/1	Arg148
<b>Hydrophilic region</b>	1/2	1/1	Leu161
<b>Hydrophobic patch</b>	2/6	2/2	Phe29, Gln97
<b>Extended hydrophobic patch</b>	2/4	2/4	Tyr34, Lys98

#### 6.4.2 Comparison of MicroFrag and solvent screens

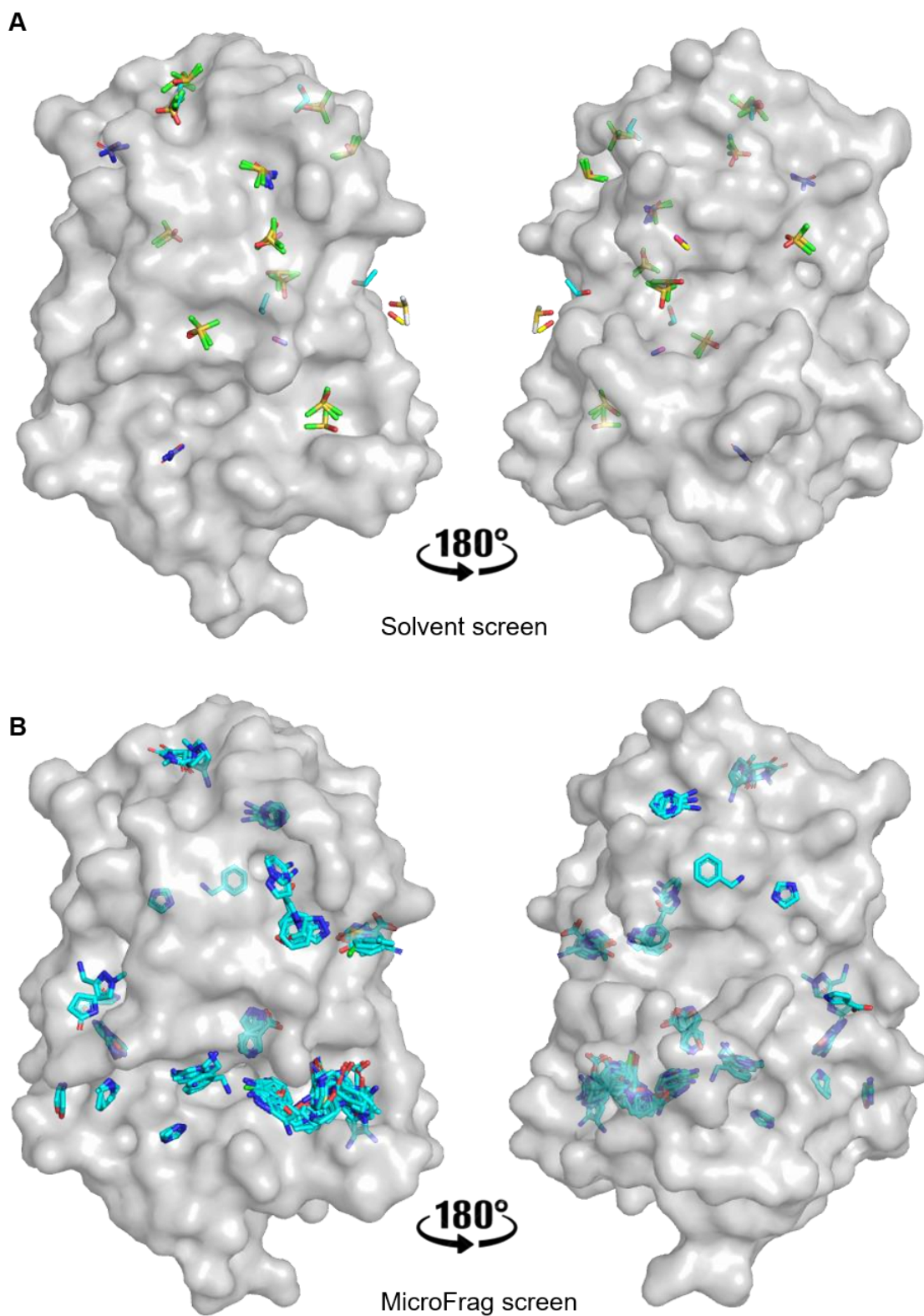
Both MicroFrag and solvent screens are based on the use of simple, low molecular weight probes to identify protein binding hot spots. After conducting both  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR and X-ray crystallography screens it is clear that the high concentrations of MicroFrag were more well tolerated by *EcDsbA* than the extremely high concentrations of solvents required to identify hot spots. X-ray crystallography analysis of MicroFrag also gave data that was easier to analyse and refine. Determining hot spots by the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR solvent and MicroFrag screens was more difficult due to the extensive interpretation of possible binding events by CSPs and required the use of additional grid point clustering analysis. The NMR screens, however, were useful in validating key residues and identified more hot spots than X-ray crystallography in both solvent and MicroFrag screens.

Solvent screening by X-ray crystallography was technically challenging due to solvent volatility and uncertainty in the orientation of the solvent or sometimes its identity, where either the solvent molecule or water could be fit to the electron density in refined structures. Using MicroFrag, which generally have more atoms than the solvents and reduced flexibility, lower ligand concentrations could be used and the lower volatility of the MicroFrag diminished these issues. The higher affinity of MicroFrag compared to solvents also possibly resulted in higher hot spot occupancy which gave less ambiguous electron density in crystallography and allowed the use of lower ligand concentrations that were better tolerated than the solvents, although with the caveat of requiring a larger library to effectively cover their respective chemical spaces. The use of lower ligand concentrations resulted in more samples with interpretable data obtained by crystallography and NMR for MicroFrag.

Benzyl alcohol, which is a member of both the MicroFrag and solvent screening libraries, gave different results depending on the conditions used. The conditions used in the solvent NMR screen saw concentrations of 3 % (289 mM) start to cause loss of intensity in  $^1\text{H}$ - $^{15}\text{N}$  HSQC and by 5 % (481 mM) protein signals were no longer detectable by  $^1\text{H}$ - $^{15}\text{N}$  HSQC. Although the MicroFrag screen was run at a concentration comparable to the sample with 3 % solvent v/v, using a stronger buffer system and adding ligand which had already been pH adjusted to ~7 allowed for

interpretable data to be obtained, which resulted in the identification of three binding sites by grid point clustering. In the X-ray crystallography analysis, the high volume of solvent (50 – 80 %, 4.8 – 7.7 M) used in the solvent soaking experiments did not result in any compound density, however in MicroFrag soaking conditions at 1 M gave a crystal structure which contained five ligands and one pose which was part of a multi-ligand cluster. This may be due to multiple factors including the high volatility of benzyl alcohol (causing total evaporation from the solvent screen soaking buffer), low solubility in the cryoprotectant compared to the MicroFrag soaking buffer which notably contains 10 % methanol or the stability and diffraction of the crystal under the concentrations used.

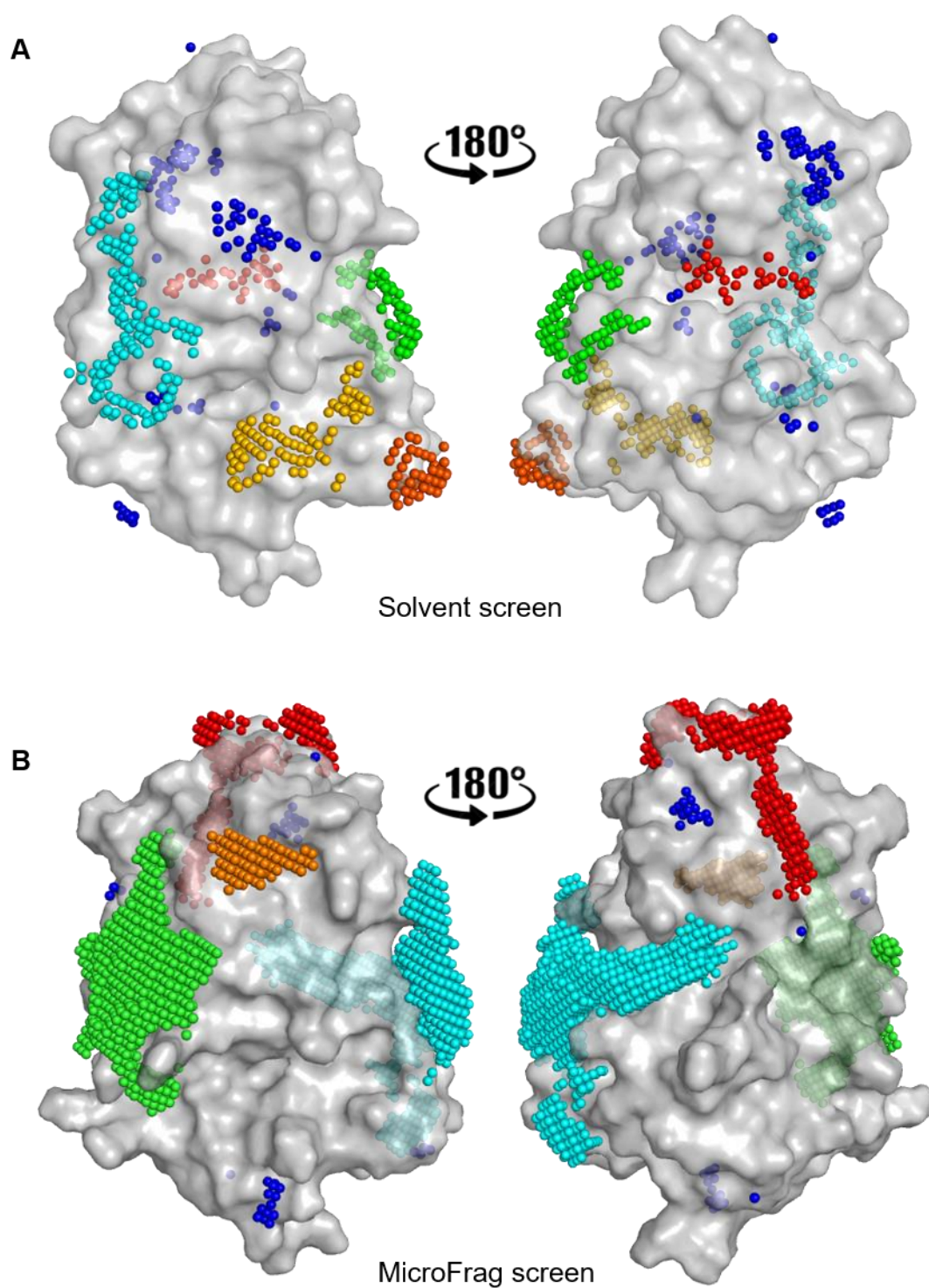




**Figure 6.16:** Comparison of global organic solvent screen and MicroFrag X-ray crystallography clusters. Overlay of all binding site clusters obtained from A) the organic solvent screen (Figure 5.21) and B) the MicroFrag screen (Figure 6.12a).



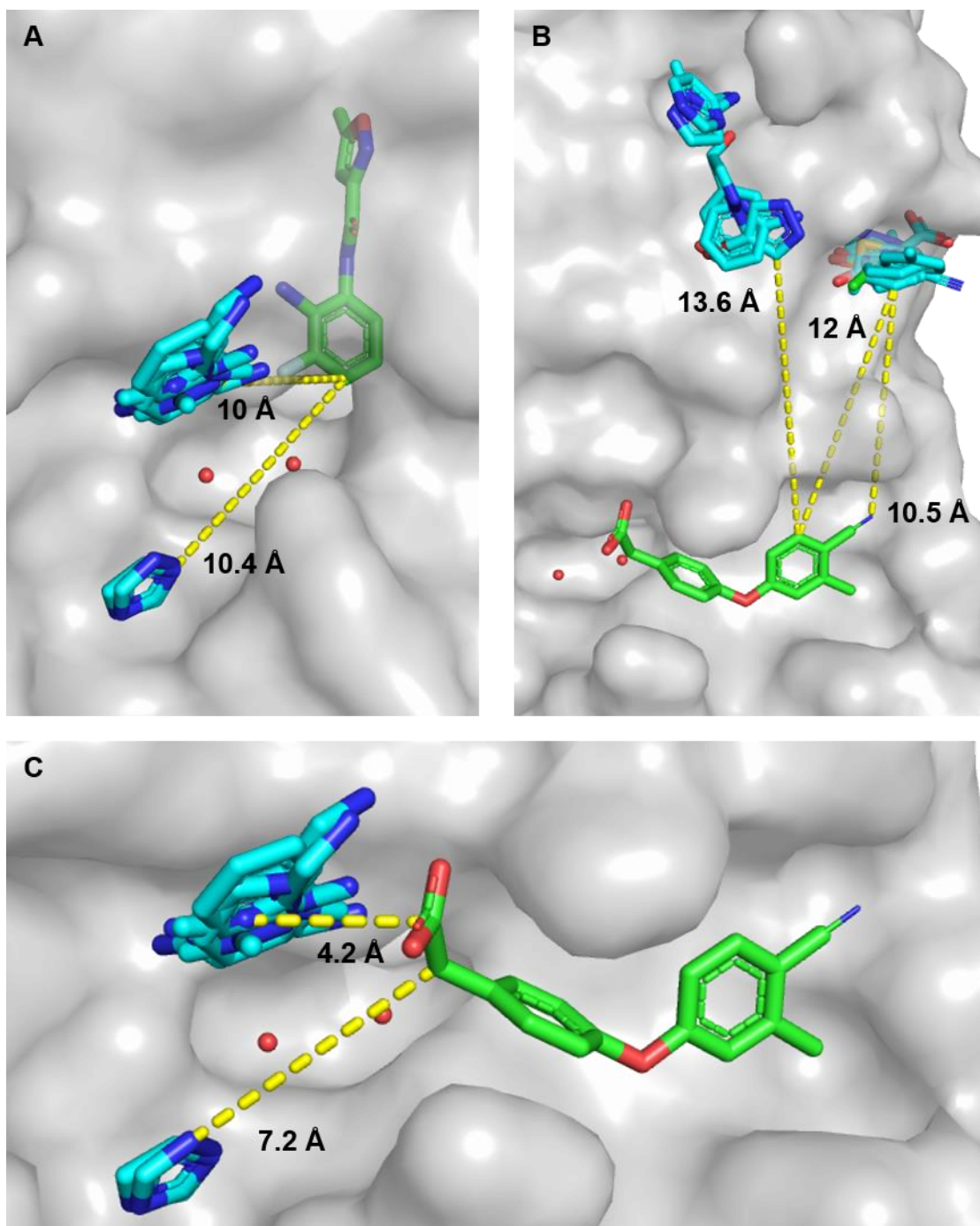
Hit rates for the two screening libraries varied. Solvent crystal structures had an average of 5.2 ligands per structure, whereas an average of 2.3 ligands was observed per MicroFrag structure. In contrast, for the NMR analysis an average of 2.7 HSQC clusters were observed per solvent, which is lower than the average of 7.4 HSQC clusters per MicroFrag. The number of refined solvent molecules in each crystal structure may be inflated by incorrectly assigned water molecules, however it is also possible that additional binding sites may result from their small size and ability to occupy regions of the proteins inaccessible to even slightly larger compounds. These solvent screen X-ray structures identified six multi-solvent hot spots, having three (the cryptic pocket, apex indent and hydrophobic patch) also identified with the MicroFrag X-ray screen. Solvent X-ray crystallography, however, also identified hot spots at the acidic patch and water channel, sites which were not identified in the MicroFrag crystallography screen. Conversely, eight hot spots identified in the MicroFrag crystallography screen, including the hydrophobic groove, were not found in the solvent X-ray screen. Many of these sites that were not identified by the solvent screen featured interactions with hydrophobic residues which are sparingly covered by the organic solvent library. The NMR data for the MicroFrag screen identified all clusters from the NMR solvent screen except the cryptic pocket. This may be due to the higher affinity MicroFrag resulting in the need for more stringent requirements imposed on calculation of grid point clusters. These requirements are implemented to restrict large and poorly defined hot spots into discrete binding sites. However, this may bias the results by unintentionally removing hot spots which are represented by only a few grid points, such as the cryptic pocket and the back of the  $\alpha 6$  helix.



**Figure 6.17:** Comparison of global organic solvent screen and MicroFrag  $^1\text{H}$ - $^{15}\text{N}$  HSQC grid point clusters. Binding site clusters obtained by the summation of the individual NMR clusters obtained from A) the organic solvent screen (Figure 5.18b) and B) the MicroFrag screen (Figure 6.6).

## 6.5 Conclusions and future directions

A screening library of 5 – 8 heavy atom compounds (MicroFrag) was designed for testing by  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR and X-ray crystallography assays. The library was chosen by iterative rounds of diversity selections skewed for the inclusion of compounds which were found within structures from the PDB <sup>(275, 276)</sup> and fragments of orally approved drugs from the DrugBank database <sup>(278, 279)</sup>. Of the 47 X-ray crystallography hits, 15 were members of both databases and a further 25 are substructures of the DrugBank database. Once the screening conditions for the library had been optimised it was logistically easier to implement these screens compared to solvent screening and the data generally produced more reliable results. Having more confidence in hot spots and interactions and identification of higher affinity ligands should also aid in the incorporation of this data into structure-based drug design strategies for new analogues. The NMR assay was successful in identifying binding hot spots through the use of the summation of ligand grid point clusters and specific interactions were highlighted by inspecting chemical shift perturbations. This assay could be used as a stand-alone screening technique for proteins not amenable to X-ray crystallography, however the data is most compelling in conjunction with the results obtained by X-ray crystallography. Of the 18 binding hot spots described within this body of work, 16 were identified by the MicroFrag screens (all except the active site and back top left). Seven sites were cross-validated by the  $^1\text{H}$ - $^{15}\text{N}$  HSQC clusters and crystallography poses with two additional X-ray crystallography observed hot spots being supported by CSPs but not grid point analysis.



**Figure 6.18:** Overlay of known fragment binders and nearby protein hot spots. A) Cryptic pocket binder (green sticks) overlayed with the peptide binding ledge and hydrophilic region clusters (cyan sticks) with distances labelled and shown as yellow dashes, structural waters as red spheres. Hydrophobic groove binder **28** (green sticks) overlayed with the B) main and extended hydrophobic patch and C) peptide binding ledge and hydrophilic region clusters (cyan sticks) with distances labelled and shown as yellow dashes, structural waters as red spheres.

Four protein hot spots in two groups are located near the known hydrophobic groove binding diaryl ether series and cryptic pocket isoxazole series (Figure 6.18). The peptide binding ledge and hydrophilic region are 5 – 10 Å away from these ligand series and hydrogen bonds identified in the MicroFrag analysis suggest that targeting Arg148 as well as Thr10-Leu161 may be useful in developing inhibitors with improved affinity. Although the residues in the peptide binding ledge are already implicated in substrate binding, neither the MicroFrag orientations at this hot spot, nor the interactions and di-azo motif at the hydrophilic region have been targeted with fragment binders. The hydrophobic patch is also within 10 – 14 Å away from the diaryl ether compounds in the hydrophobic groove and expansion into these sites would likely require longer linkers and be more difficult. Nonetheless, Phe29 and Tyr34 appear to make conserved interactions to MicroFrag, are highly conserved across DsbA homologues (Figure 1.25) <sup>(243)</sup> and are also implicated in peptide binding (Chapter 2).

Hot spot identification requires a balance of compound affinity, library size, physicochemical properties and the assay technologies available. While fragment screening can identify multiple starting points for medicinal chemistry efforts, not all binding hot spots are identified. Similarly, solvent screening identified a number of hot spots, however failed to find  $\pi$ - $\pi$  interactions within hot spots. In addition, the binding poses obtained from solvent screening were sometimes ambiguous and the technique was limited by the requirement of protein stability and solubility at the high concentrations required. MicroFrag balanced these requirements with the property and complexity profiles of compounds in the library lying between those for the compounds used for solvent and fragment screening. Screening moderately sized compound libraries by X-ray crystallography is becoming quicker and more accessible with the advancements in automation and data collection. In this case, the library of MicroFrag was able to be screened in ~16 hours of synchrotron time. Furthermore, we showed that screening MicroFrag by  $^1\text{H}$ - $^{15}\text{N}$  HSQC could provide hot spot information for targets which are not amenable to crystallography. These MicroFrag screens provided valuable information on binding hot spots and conserved interactions that may lead to higher affinity inhibitors of *EcDsbA* and would be equally applicable to screening of other challenging drug targets.

# **Chapter 7:**

## **Conclusions and future directions**

This body of work aimed to explore orthogonal strategies towards the optimisation of compounds that bound to *EcDsbA*. This research was informed by previous work that had enabled the identification of peptides and small molecules that bound to *EcDsbA* both covalently and non-covalently <sup>(128, 229, 233, 242, 251-256)</sup>. The initial strategy was to examine the interaction of *EcDsbA* with a peptide containing a single cysteine residue that was based on the sequence of a loop in *EcDsbB* that has been shown to interact with *EcDsbA* in the crystal structure of their complex. The heptapeptide **23**, Ac-PWATCDS-NH<sub>2</sub>, derived from the sequence of *EcDsbB* was able to form a stable mixed disulfide complex with native <sup>30</sup>CPHC<sup>33</sup> *EcDsbA*. Most previous studies of mixed disulfide complexes with *EcDsbA* have used an active site mutant, <sup>30</sup>CPH(S/A)<sup>33</sup>. I found that the mixed disulfide formed with native *EcDsbA* was surprisingly stable and this complex was characterised using NMR. The catalytic intermediate was stable for multiple weeks and could be formed from either the oxidised or the reduced state of *EcDsbA*. Analysis of the chemical shifts for the C $\alpha$ , C $\beta$  and NH resonances of the peptide-protein complex in comparison to the unbound oxidised and reduced *EcDsbA* showed that the protein did not undergo a large conformational change upon formation of this intermolecular bond. Furthermore, comparison of **23** with a related peptide, **24** Ac-PWATADS-NH<sub>2</sub>, that was unable to form a covalent complex, revealed that the chemical shift perturbations that were observed for residues surrounding the active site were much larger for **23**. This suggested that the non-covalent interactions of the peptide with residues in the hydrophobic groove of *EcDsbA*, which are thought to contribute to substrate recognition, were greatly enhanced by the addition of the covalent bond. The stability of the mixed-disulfide peptide complex suggested that it may be possible to functionalise small molecules with thiols as warheads to generate covalent inhibitors of oxidised *EcDsbA* or alternatively to use small molecules with suitable electrophilic warheads to target reduced *EcDsbA*.

In order to test the utility of small molecules with thiol warheads as inhibitors, a previously reported molecule based on a diaryl ether fragment core was selected <sup>(128, 253)</sup>. The aim was to generate a library of small molecule inhibitors which expanded on diaryl ether fragment with the addition of the thiol by using a microscale parallel synthesis approach. It was envisaged that these could be tested as crude

products, since it was expected that the most potent thiols would form stable adducts with *EcDsbA*. The parallel synthesis protocol involved a multistep conversion of terminal alcohols to free thiols. It was observed that this approach resulted in products that were present in low purity and were potentially formed as symmetrical disulfide-linked dimers. Resynthesis on traditional batch scales (> 50 mg) revealed that both the intermediates and products in this reaction were unstable and confirmed the rapid oxidation to disulfide dimers, which indicated that the approach of testing thiols as covalent *EcDsbA* warheads was not a viable strategy. Therefore, an alternative approach was developed. A target agnostic library of electrophilic covalent warheads, which were able to label a range of nucleophilic residues, was designed. The warheads were designed to include a reactive handle which facilitated coupling of the warheads to target-selective fragments. Again, it was anticipated that this elaboration could be performed using microscale parallel reactions. The covalent library contained warheads that varied both mechanism of reaction with the protein and warhead reactivity. These warheads were attached to linkers of different lengths, and all reagents were functionalised to allow coupling to a fragment via amidation reactions.

The covalent library was tested for its ability to form stable adducts with *EcDsbA*. The covalent binding assay tested whether reduced *EcDsbA* was able to bind to either the warhead alone, or the warhead conjugated to diaryl ether amine **110**. The conjugated warheads were synthesised in parallel using microscale chemistry and the products were tested without further purification. The warheads alone and the diaryl ether conjugated products (at two timepoints) were tested against *EcDsbA* by <sup>1</sup>H-<sup>15</sup>N HSQC. This identified those warheads that were only reactive in the presence of the fragment binding motif and also tested the stability of the intermolecular complex formed. Sulfonyl fluoride **147** and carbamate **132** produced substantial labelling as a crude product, however, of the two, carbamate **132** was the only compound to be unchanged at the second timepoint. This represented the desired outcome – that the warhead was only sufficiently reactive to form a covalent adduct when positioned close to the active site of *EcDsbA* by the fragment, and that it formed a single stable adduct. A series of compounds from the covalent library were resynthesised for testing as pure compounds, however the carbamate of interest **132**, was not soluble enough in



aqueous conditions to be validated by the NMR experiment. Neither control compound **131** or **148**, which replaced either the linker or fragment component of carbamate **132** respectively, showed the ability to covalently react with the reduced protein. Assuming that the initial result with carbamate **132** was not a false positive, this demonstrated the importance of the fragment motif for binding, and the presence and length of the linker for modulation of warhead reactivity. Although it was not possible to validate covalent inhibitors of *EcDsbA*, the results of this study provided some interesting data for further study. Firstly, chemical modification of carbamate **132** to improve its solubility without impacting its binding to *EcDsbA* will be required to validate the original findings of the crude reaction screening. Alternatively, the sulfonyl fluoride **147** may be a suitable candidate for developing a covalent *EcDsbA* inhibitor if the reactivity of this warhead can be tuned to generate a single stable adduct. Further support for the use of sulfonyl fluorides as a covalent warhead arises from preliminary results of second covalent warhead screen that has recently been undertaken against *EcDsbA* where sulfonyl fluoride compounds were again identified as hits. Therefore, it is possible that sulfonyl fluorides are suitable starting points for developing covalent inhibitors of *EcDsbA*. However, it is likely that further development of the non-covalent portion of any such inhibitors will be required in order to maximise their selectivity for *EcDsbA*.

A range of different approaches was tested in order to generate data to aid in the structure-based design of compounds that bound with higher affinity to *EcDsbA*. Protein “hot spots” or pockets on *EcDsbA* that were able to form interactions with specific functional groups were identified using high concentrations of low molecular weight probes in NMR and X-ray crystallography assays. Firstly, organic solvents were employed in computational studies similar to multiple copy simultaneous search (MCSS) to assess their ability to bind to *EcDsbA*. In addition, experimental approaches to monitor solvent binding were employed including <sup>1</sup>H-<sup>15</sup>N HSQC NMR and X-ray crystallography to generate multiple solvent crystal structures (MSCS). While the organic solvents provided insight into the potential binding sites of oxidised *EcDsbA* the intrinsically harsh conditions of the experimental screens were not always tolerated by the protein and crystal systems, were difficult to implement and produced some ambiguous results. As a compromise between traditional fragment screening and

solvent screening, a MicroFrag library consisting of compounds with between 5 – 8 heavy atoms was designed. This represents a chemical space which is between organic solvents and fragments in terms of size and complexity. The MicroFrag were tested for their ability to bind *EcDsbA* using NMR spectroscopy and X-ray crystallography. The MicroFrag were found to have higher affinity than solvents for the protein hot spots. This allowed them to be tested at lower concentrations than the solvents and resulted in more reliable electron density maps and chemical shift perturbation profiles by crystallography and  $^1\text{H}$ - $^{15}\text{N}$  HSQC, respectively. This improved data quality and allowed the binding sites and privileged protein interactions to be identified and prioritised with more confidence. In-house scripts were developed to convert HSQC chemical shift perturbations into grid point clusters indicative of binding events, and together with clusters of ligand poses observed by X-ray crystallography and computational simulations, the MicroFrag data suggested a number of different hot spots on the surface of *EcDsbA* that are capable of binding to small molecules. Encouragingly, this analysis identified both of the previously known fragment binding sites for *EcDsbA*, namely the hydrophobic groove and the internal cryptic pocket, which provided some confidence that the analysis was providing useful information. Four additional MicroFrag binding sites were identified within close proximity to current fragment series. These data obtained for the MicroFrag binding at these sites can potentially be used to expand compounds in the current fragment series by targeting these sites. This could either be achieved by the direct incorporation of MicroFrag hits into current compounds using an approach analogous to fragment linking or by designing analogues which encompass pharmacophore elements observed in MicroFrag co-structures, using a fragment growing approach.

The first of these additional sites was observed at the peptide binding ledge of *EcDsbA*, which is  $\sim 4 - 10$  Å away from current groove and cryptic pocket fragments. The ledge has previously been identified as an important region for substrate recognition and binding. Crystal structures show that peptide substrates and *EcDsbB* form an antiparallel  $\beta$  sheet interaction with Arg148 and Val150, which are within this hot spot <sup>(222, 229, 242)</sup>. Furthermore, these residues have been implicated in the substrate specificity of *EcDsbA* <sup>(238, 239)</sup>. The MicroFrag screens identified a hydrogen bond donor interaction to the backbone carbonyl of Arg148 as the most important

interaction of this hot spot. A large chemical shift perturbation was also observed for Arg148 in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the *EcDsbA* complex with covalent peptide **23**. The MicroFrag data suggested other interactions that may be important at this site, including those involving the mainchain nitrogen of Gly149, the carbonyl oxygen of Pro151 and the sidechain of Gln160. Previous efforts to elaborate the current series of *EcDsbA* fragments were able to generate compounds that formed a hydrogen bond with Pro151, but these did not improve the binding affinity. Efforts to target Arg148 and Gly149 did not produce compounds that were able to make interactions with these residues. This may indicate that the geometry of the expansion vector used in the previous studies was not ideal, or the compounds did not contain the most suitable functional groups. Analysis of the MicroFrag data reveals that the hydrogen bonds observed in the crystal structures of *EcDsbA* with the MicroFragments originate from aryl amines which form multiple polar contacts at once and suggests that fragment expansions should aim to place a 5- or 6-membered aromatic ring with a hydrogen bond donor at this position; preferably a primary amine as it appears in multiple structures and has the ability to donate two hydrogen bonding interactions.

The second hot spot identified in the MicroFrag data is present between the  $\beta 1$  and  $\beta 5$  sheets at a hydrophilic region of the substrate binding site and is located  $\sim 7 - 11$  Å away from the two known fragment binding sites. This hot spot has not yet been explored in elaborating any of the current fragment series, however residues in this hot spot contribute to the binding site of peptide **23**, and this hot spot is also identified in FTMap calculations and the organic solvent screen. The MicroFragments which occupy this binding site make hydrogen bonds to the sidechains of Thr10, and the main chain oxygens of Gln160 and Leu161, and they all contain the same 2D 2-point pharmacophore, two hydrogen bond donors separated by one heavy atom. Although only urea was observed at this site in the organic solvent screen, the two MicroFrag ligands that were seen at this site were 5-membered heteroaromatic rings, 1,2,4-triazole and imidazole. This suggests that this type of ring could be beneficial in fragment design.

The remaining two hot spots, are  $\sim 10 - 13$  Å away from the hydrophobic groove binding fragments. The hot spots comprise hydrophobic patches along the interface

of the thioredoxin and  $\alpha$ -helical domains of *EcDsbA*, which previous studies have suggested are influential to ligand binding <sup>(229, 230)</sup>. The residues that are found at this site include Phe29 and Tyr34, which are present on either side of the active site. These residues were implicated in the binding of the mono-cysteine peptide, as well as the covalent small molecule inhibitors and the MicroFragments where Phe29 was also highlighted by the organic solvent screen. Other residues in these hot spots include Gln97 and Lys98. Additional data implicating these residues in intermolecular interactions come from chemical shift perturbations or line broadening of their amide resonances in <sup>1</sup>H-<sup>15</sup>N HSQC experiments in the presence of covalent compounds and MicroFragments. A crystal structure of *EcDsbA* in complex with a substrate peptide derived from SigA revealed a direct hydrophobic contact between *EcDsbA*'s Phe29 and Ile2 of the peptide by X-ray crystallography and the complementary NMR binding experiment resulted in perturbations at Phe29 and Tyr34 <sup>(229)</sup>. Moreover, a comparison of *EcDsbA* and 10 DsbA homologues with sequence identities between 15 – 99 % revealed that the residues at these hot spots, or their properties, are highly conserved. These sequences give rise to an 80 % consensus that the residues corresponding to Phe29 are hydrophobic, Tyr34 are conserved as Tyr, that Gln97 are polar residues (either a Gln or His) and that Lys98 is a charged residue (either cation or anion). Taken together, these data suggest that these residues may be significant to ligand binding and targeting interactions at these sites could provide a productive strategy for the development of the current fragment series.

Although this body of work was conducted in the context of developing *EcDsbA* inhibitors, the libraries and techniques described herein are equally applicable for screening against a variety of protein targets. Preliminary data from screening the covalent library against other targets suggest that it may be suitable for this purpose. Additionally, the MicroFrag library has been and is currently being tested by X-ray crystallography against a range of other targets. Consistent with the previous literature, hit rates of ~ 10 – 50 % have been identified in these screens. This work explored multiple orthogonal synthetic and biophysical techniques which, in conjunction with fragment-based drug design, have the potential to aid in the development of small molecule that bind with higher affinity. It is likely that these

strategies will be particularly useful in medicinal chemistry campaigns against difficult targets, where obvious strategies for fragment elaboration are lacking.

# **Chapter 8:**

# **Experimental methods**

## 8.1 Chemistry

### 8.1.1 Instrumentation

All chemicals and reagents were purchased from commercial suppliers and used without further purification unless otherwise stated. Solvents were HPLC grade unless otherwise stated and used without further preparation. Anhydrous solvents were obtained from MBraun SPS-800 Solvent purification system according to manufacturer instructions. TLCs were run on silica gel 60 F254 pre-coated plates (0.25 mm, Merck) and visualised with UV light. Flash column chromatography was run using Grace HP+ C18 silica gel 60 (40 – 63  $\mu\text{m}$ ; 230 – 400 mesh).

$^1\text{H}$  and  $^{13}\text{C}$  Nuclear Magnetic Resonance (NMR) spectra were obtained at 400 MHz and 100 MHz respectively, on a Bruker Avance III Nanobay 400 MHz spectrometer coupled to a BACS 60 automatic sample changer. All spectra were processed using MestReNova versions 6.0 – 14.1. All chemical shifts were reported as  $\delta$  values in parts per million (ppm) and were referenced to the respective residual proteo-solvent peaks: 7.26 ppm ( $^1\text{H}$  NMR) and 77.16 ppm ( $^{13}\text{C}$  NMR) for deuterated chloroform ( $\text{CDCl}_3$ ), 3.31 ppm ( $^1\text{H}$  NMR) and 49.00 ppm ( $^{13}\text{C}$  NMR) for methanol ( $\text{CD}_3\text{OD}$ ), 8.03 ppm ( $^1\text{H}$  NMR) and 163.15 ppm ( $^{13}\text{C}$  NMR) for *N,N*-dimethylformamide ( $\text{DCON}(\text{CD}_3)_2$ ), 4.79 ppm ( $^1\text{H}$  NMR) for deuterium oxide ( $\text{D}_2\text{O}$ ) and 2.50 ppm ( $^1\text{H}$  NMR) and 39.52 ppm ( $^{13}\text{C}$  NMR) for  $d_6$ -DMSO and 0.030 ppm ( $^1\text{H}$  NMR) for DSS in  $\text{D}_2\text{O}$ . Multiplicities are described as a (apparent), br (broad), s (singlet), d (doublet), t (triplet), q (quartet) p (pentet) and m (multiplet).  $^{13}\text{C}$  NMR assignment of carbon environments based on APT phasing is as follows: C = quaternary carbon, CH = methine carbon,  $\text{CH}_2$  = methylene carbon and  $\text{CH}_3$  = methyl carbon.

Liquid chromatography mass spectrometry (LCMS) was recorded on three systems. System A an Agilent 6120 Series Single Quad LC/MS coupled to an Agilent 1260 Series HPLC. System B an Agilent 6100 Series Single Quad LC/MS coupled to an Agilent 1200 Series HPLC. System C a Shimadzu single Quad LCMS 2020 coupled with a Shimadzu NexeraX2 HPLC. Both system A and B have buffer A: 99.9 %  $\text{H}_2\text{O}$  with 0.1 % formic acid and buffer B: 99.9 % MeCN with 0.1 % formic acid. A flow rate of 0.5 mL/min was used with an Agilent Poroshell 120 EC-C18 50 x 3.0 mm 2.7  $\mu\text{m}$

column and a Phenomenex Luna 3  $\mu\text{m}$  C8(2) 15 x 4.6 mm column for system A and B, respectively. The hydrophobic methods are as follows: System A, 95 % buffer A and 5 % buffer B from 0 – 1 minute, gradient of 5 – 100 % of buffer B from 1 – 2.5 min, 100% buffer B from 2.5 – 3.8 min, 95 % buffer A and 5 % buffer from 3.8 – 5 min. System B 95 % buffer A and 5 % buffer B from 0 – 4 min, gradient of 5 – 100 % buffer B from 4 – 7 min, 95 % buffer A and 5% buffer B from 7 – 12 min. System C is a Shimadzu LCMS with DGU-20A Degasser, NexeraX2 LC-30AD binary pump, NexeraX2 SIL-3AC autosampler, CTO-20A oven, CBM-20A Communication BUS manager, SPD-M30A photodiode array detector, and 2020 MS system fitted with a Merck C18 Chromolith FastGradient RP-18 endcapped 50 mm  $\times$  2.0 mm monolithic column (2  $\mu\text{m}$ ). Buffer A: 0.1% acetic acid in water; buffer B: 0.1% acetic acid in acetonitrile. Gradient: 0 – 0.25 mins isocratic 95% Buffer A and 5% Buffer B, 0.25 – 3.5 mins gradient to 100% Buffer B, 3.5 – 4 mins isocratic 100% Buffer B, 4.0 – 4.01 mins gradient to 95% Buffer A and 5% Buffer B, 4.01 – 5.0 mins isocratic 95% Buffer A and 5% Buffer B at 35 °C and 0.6 mL/min. Data was analysed with LabSolutions v.5.80 software. For all systems mass spectra were acquired in the positive and negative ion mode with a scan range of 100 – 1000  $m/z$  with UV detection at 214 and 254 nm.

Analytical HPLC were performed on an Agilent 1260 HPLC system. A flow rate of 1 mL/min was used with a Zorbax Eclipse Plus C18 Rapid Resolution 4.6 x 100 mm 3.5  $\mu\text{m}$  column. Buffer A: H<sub>2</sub>O with 0.1 % TFA and buffer B: MeCN with 0.1 % TFA. A gradient of 5 – 100 % solvent B in solvent A from 0 – 10 min was used. Agilent LC/MSD Chemstation Rev.B.04/03 coupled with Mass Hunter Easy Access Software was used to run and process the samples.

High resolution mass spectrometry (HRMS) were acquired on an Agilent 6224 TOF LC/MS Mass Spectrometer coupled to an Agilent 1290 Infinity or a H<sub>2</sub>O<sub>s</sub> LCT TOF LC/MS Mass Spectrometer coupled to a 2795 Alliance Separations module. For the Agilent 6224 TOF LC/MS data were mass corrected via a dual-spray electrospray ionisation (ESI) source with the following parameters: Electrospray Ionisation, Drying gas-flow: 11 L/ min; Nebuliser: 45 psi; Drying gas temperature: 325°C; Capillary Voltage (V<sub>cap</sub>): 4000 V; Fragmentor: 160 V; Skimmer: 65 V; OCT RFV: 750 V; Scan



range acquired: 100 – 1500  $m/z$ . Internal Reference ions: Positive Ion Mode =  $m/z$  = 121.050873 & 922.009798. Acquisition and analysis were performed using Agilent Mass Hunter Data Acquisition software vB.05.00 Build 5.0.5042.2 and Mass Hunter Qualitative Analysis vB.05.00 Build 5.0.519.13, respectively. For the H<sub>2</sub>O<sub>s</sub> LCT TOF LC/MS data were acquired and mass corrected via a dual-spray Leucine Enkephaline reference sample with the following parameters: Electrospray Ionisation Desolvation gas flow: 550 L/H; Desolvation temperature: 250 °C; Source temperature: 110 °C; Capillary Voltage: 2400 V; Sample cone voltage: 60 V; Scan range acquired: 100 – 1500  $m/z$  Scan time: 1 s. Internal Reference ions: Positive Ion Mode =  $m/z$  = 556.2771. Acquisition and analysis were performed using Masslynx software v4.1.

### 8.1.2 Microscale synthesis

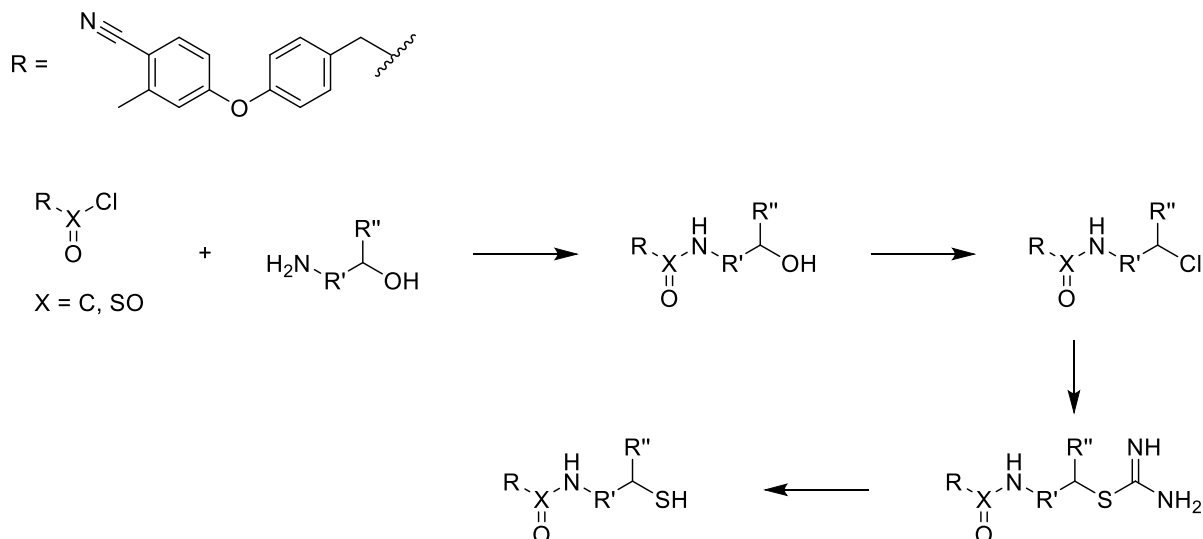
Parallel chemistry was conducted in sealed Grenier 96 well U bottom PP plates. Solvent evaporation from the plates were carried out on a Genevac EZ-2PLUS according to manufacturer's guidelines.

5  $\mu$ L of the crude reaction mixture *d*<sub>6</sub>-DMSO stock (100 mM) was diluted with 95  $\mu$ L of 1:1 MeCN/MilliQ H<sub>2</sub>O to prepare a 5 mM stock. The stock was then centrifuged through a Pall AcroPrep Advance 96 well, 0.2  $\mu$ m pore filter plate. Reaction purity was then assessed by LCMS using a 2  $\mu$ L injection of the filtrate.

LCMS for microscale reactions were performed using system C as described above.

### 8.1.3 General procedures

#### 8.1.3.1 Thiol warhead library general synthesis



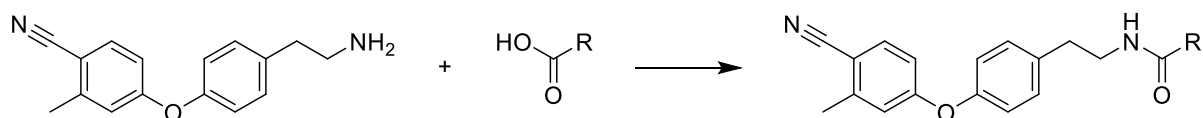
Solid amine (~1 mg, ~8  $\mu\text{mol}$ ), solid acid chloride (~2 mg, ~7  $\mu\text{mol}$ ) and Et<sub>3</sub>N (0.5 M, 1.4  $\mu\text{L}$ , 10  $\mu\text{mol}$ ) were added to wells containing MeCN (18.6  $\mu\text{L}$ ). The plate was sealed and left at room temperature for 18 h without agitation. 1  $\mu\text{L}$  of sample was removed for analysis by LCMS and the remaining solvent was removed *in vacuo*.

The crude solids were redissolved in chloroform (18.6  $\mu\text{L}$ ), and thionyl chloride (0.5 M, 1.4  $\mu\text{L}$ , 10  $\mu\text{mol}$ ) was added to each well. The plate was sealed and left at room temperature for 19 h without agitation. 1  $\mu\text{L}$  of sample was removed for analysis by LCMS and the remaining solvent was removed *in vacuo*.

The crude solids were redissolved in dry DMF (10  $\mu\text{L}$ ) and thiourea (1 M in dry DMF, 10  $\mu\text{L}$ , 10  $\mu\text{mol}$ ) were added to each well. The plate was sealed and left at room temperature for 12 h without agitation. 1  $\mu\text{L}$  of sample was removed for analysis by LCMS and the remaining solvent was removed *in vacuo*.

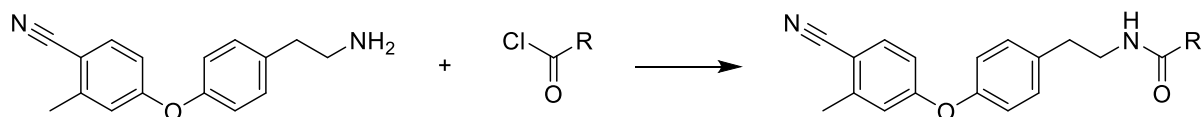
The crude solids were redissolved in NaOH (aq., 1 M, 20  $\mu\text{L}$ , 20  $\mu\text{mol}$ ), the plate was sealed and left at room temperature for 4 h without agitation. After this time, the reactions were quenched with HCl (aq., 1 M, 20  $\mu\text{L}$ , 20  $\mu\text{mol}$ ), the plate was then sealed and left at room temperature for 30 mins without agitation. The solvent was removed *in vacuo* and mixtures were redissolved in *d*<sub>6</sub>-DMSO (20  $\mu\text{L}$ ).

### 8.1.3.2 Acid warhead library general synthesis



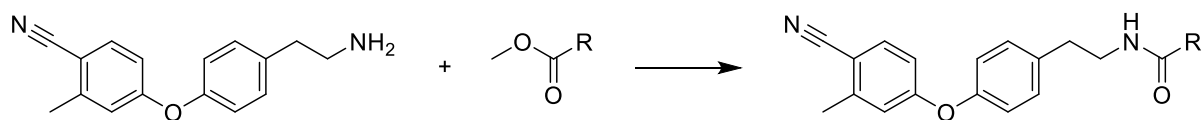
Acid (1 M in dry DMF, 5  $\mu$ L, 1 eq., 5  $\mu$ mol), Et<sub>3</sub>N (1  $\mu$ L, 1.5 eq, 7.5  $\mu$ mol) and coupling reagent (HATU or EDC.HCl 0.5 M in dry DMF, 20  $\mu$ L, 2 eq., 10  $\mu$ mol) was added to each well. The plate was sealed and left at room temperature for 30 minutes without agitation. After this time, amine (1 M in dry DMF, 5 or 10  $\mu$ L, 1 or 2 eq., 5 or 10  $\mu$ mol) and dry DMF was added to each well to a final reaction volume of 50  $\mu$ L, the plate was sealed and left at room temperature for a further 18 h without agitation. The solvent was removed *in vacuo* and mixtures were redissolved in *d*<sub>6</sub>-DMSO (50  $\mu$ L, final conc. 0.1 M).

### 8.1.3.3 Acid chloride warhead library general synthesis



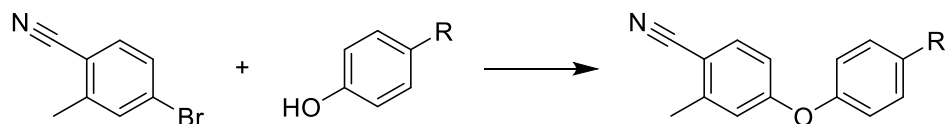
Dry DMF (34 or 39  $\mu$ L), followed by Et<sub>3</sub>N (1  $\mu$ L, 1.5 eq, 7.5  $\mu$ mol) and amine (1 M, 5 or 10  $\mu$ L, 1 or 2 eq., 5 or 10  $\mu$ mol) was added to each of the reaction wells. The acid chloride (1 M in dry DMF, 5  $\mu$ L, 1 eq., 5  $\mu$ mol) was then added to a final reaction volume of 50  $\mu$ L and the plate was sealed and left at room temperature for a further 18 h without agitation. The solvent was removed *in vacuo* and redissolved in *d*<sub>6</sub>-DMSO (50  $\mu$ L, final conc. 0.1 M).

#### 8.1.3.4 Ester warhead library general synthesis



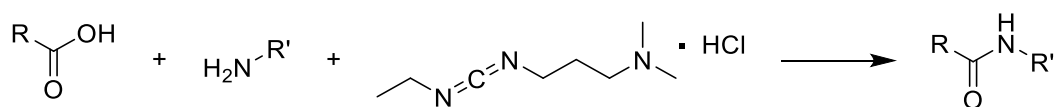
Ester (1 M in dry DMF, 5  $\mu$ L, 1 eq., 5  $\mu$ mol), Et<sub>3</sub>N (1  $\mu$ L, 1.5 eq, 7.5  $\mu$ mol), coupling reagent (HATU or EDC.HCl 0.5 M in dry DMF, 20  $\mu$ L, 2 eq., 10  $\mu$ mol) and KOH (2.5 M in H<sub>2</sub>O, 2  $\mu$ L, 1 eq., 5  $\mu$ mol) were added to each well. The plate was sealed and left at room temperature for 30 minutes without agitation. After this time, amine (1 M in dry DMF, 5 or 10  $\mu$ L, 1 or 2 eq., 5 or 10  $\mu$ mol) and dry DMF was added to each well to a final reaction volume of 50  $\mu$ L and the plate was sealed and left at room temperature for a further 18 h without agitation. The solvent was removed *in vacuo* and redissolved in *d*<sub>6</sub>-DMSO (50  $\mu$ L, final conc. 0.1 M).

#### 8.1.3.5 General procedure A – Ullmann coupling



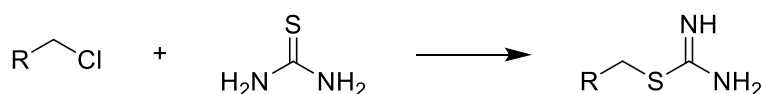
*N,N*-dimethylglycine (0.1 – 0.3 eq.), 4-bromo-2-methyl-benzonitrile (1 eq.), copper(I) iodide (0.1 – 0.3 eq.), caesium carbonate (1.5 – 1.8 eq.), and phenol (1.5 – 1.8 eq.) were dissolved in 1,4-dioxane (0.1 M) and refluxed with stirring for 16-24 h at 120 °C. The reaction was diluted with EtOAc and washed with water (3  $\times$  10 – 50 mL) and brine (3  $\times$  10 – 50 mL), and the organic layer dried over anhydrous magnesium sulfate. The crude reaction mixture was purified using flash column chromatography.

### 8.1.3.6 General procedure B – Amide coupling (EDC.HCl)



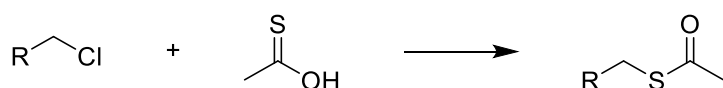
EDC.HCl (1.3 – 1.6 eq.), HOBt (1.5 – 1.6 eq.) and acid (1 eq.) were sealed in a round bottom flask and purged with  $N_2$ . Dry DMF (0.1 M) was added and the reaction mixture was stirred at room temperature for 1 h. After this time amine (1 – 2 eq.) was added to the solution and it was stirred for an additional 16 – 18 h. The reaction was diluted with EtOAc, washed with water (3 × 10 – 50 mL) and brine (3 × 10 – 50 mL) and the organic layer dried over anhydrous magnesium sulfate. The crude reaction mixture was purified using flash column chromatography.

### 8.1.3.7 General procedure C – S-alkylation (thiourea)



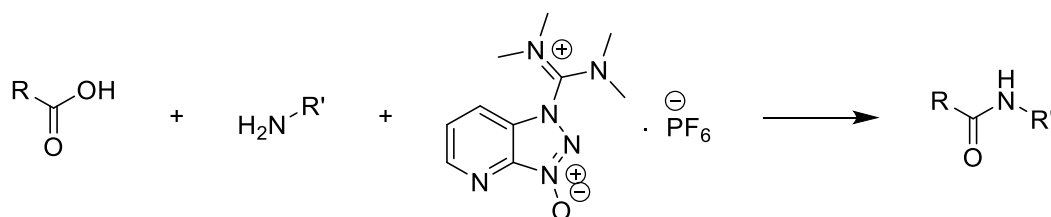
Chloride (1 eq.) and thiourea (3 eq.) were purged with  $N_2$  and dissolved in dry DMF (0.1M). The reaction mixture was refluxed at  $120^\circ C$  for 3 h before concentrating *in vacuo*. The products were used without further purification.

### 8.1.3.8 General procedure D – S-alkylation (thioacetic acid)



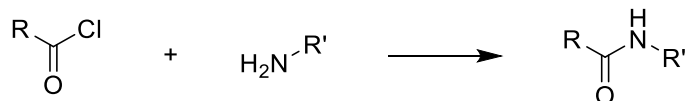
Thioacetic acid (1.8 – 2 eq.) was added to a solution of bromide (1 eq.) and  $Et_3N$  (2 – 2.1 eq.) in EtOH (0.1 M) and stirred at room temperature for 16 – 24 h. The reaction mixture was diluted with THF and EtOAc, washed with water (3 × 10 – 50 mL) and brine (3 × 10 – 50 mL), and the organic layer dried over anhydrous magnesium sulfate. The crude reaction mixture was purified using flash column chromatography.

### 8.1.3.9 General procedure E – Amide coupling (HATU)



Acid (1 eq.), HATU (1.5 – 2 eq.) and Et<sub>3</sub>N (1.4 – 2 eq.) were dissolved in dry DMF (0.1 M) and stirred at room temperature for 30 mins. After this time amine (1.5 – 2 eq.) was added to the solution and the reaction was stirred for an additional 16 – 24 h. The reaction was diluted with EtOAc, washed with water (3 × 10 – 50 mL) and brine (3 × 10 – 50 mL) and the organic layer dried over anhydrous magnesium sulfate. The crude reaction mixture was purified using flash column chromatography.

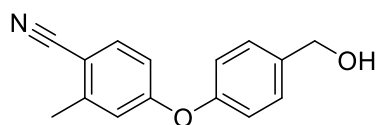
### 8.1.3.10 General procedure F – Amide formation (acyl chloride)



Amine (1 – 2 eq.) and acyl chloride (1 – 2 eq.) were added to a solution of Et<sub>3</sub>N (2 eq.) and DCM (0.1 M) and stirred at room temperature for 45 min – 16 h (reaction monitored by TLC and LCMS for completion). The reaction was washed with water and HCl (aq., 1 M) and the organic layer dried over anhydrous magnesium sulfate. The crude reaction mixture was purified using flash column chromatography.

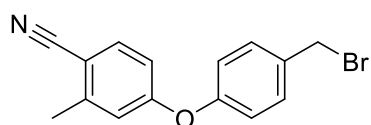
### 8.1.4 Compound characterisation

#### 4-(4-(hydroxymethyl)phenoxy)-2-methylbenzonitrile (39)



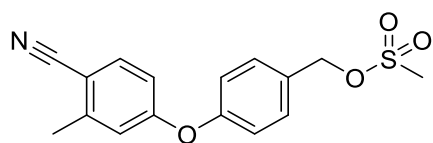
Prepared according to procedure A with 4-bromo-2-methylbenzonitrile (100 mg, 1.00 eq., 0.51 mmol), 4-hydroxybenzyl alcohol (112 mg, 1.80 eq., 0.90 mmol), copper iodide (29 mg, 0.30 eq., 0.15 mmol), *N,N*-dimethylglycine hydrochloride (30 mg, 0.40 eq., 0.22 mmol), and caesium carbonate (300 mg, 1.80 eq., 0.92 mmol). The crude material was purified by flash chromatography using a 1:3 ratio of EtOAc in petroleum spirits to give the product as a yellow residue (65 mg, 53 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.53 (d, *J* = 8.5 Hz, 1H), 7.41 (d, *J* = 8.6 Hz, 2H), 7.05 (dt, *J* = 8.5, 4.7, 2.0 Hz, 2H), 6.85 (d, *J* = 2.4 Hz, 1H), 6.82 (dd, *J* = 8.5, 2.4 Hz, 1H), 4.72 (d, *J* = 5.5 Hz, 2H), 2.49 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 161.54, 154.56, 144.62, 137.71, 134.49, 129.05, 120.60, 118.95, 118.30, 115.48, 106.62, 64.35, 20.77. **LCMS** (*m/z*): 238.6 [M-H]<sup>-</sup>, *t*<sub>R</sub> = 3.57 min, hydrophobic method. **HPLC** *t*<sub>R</sub> = 5.812 min, 99 %, hydrophobic method. The data was consistent with literature <sup>(253)</sup>.

#### 4-(4-(bromomethyl)phenoxy)-2-methylbenzonitrile (40)



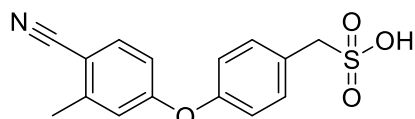
4-(4-(Hydroxymethyl)phenoxy)-2-methylbenzonitrile (60 mg, 1.0 eq., 0.25 mmol), *N*-bromosuccinimide (93 mg, 2.1 eq, 0.52 mmol) and triphenylphosphine (129 mg, 2.00 eq, 0.49 mmol) were added to dry DCM (5 mL) and refluxed for 1 h with stirring. The reaction mixture was diluted with diethyl ether where a brown precipitate was observed. The solution was filtered, evaporated and the crude product was reacted on without further purification. **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.72 – 7.64 (m, 3H), 7.58 (dd, *J* = 7.4, 1.5 Hz, 1H), 7.51 – 7.47 (m, 3H), 3.47 (s, 3H), 2.75 (s, 2H). **LCMS** (*m/z*): did not ionise, *t*<sub>R</sub> = 3.79 min, hydrophobic method. **HPLC** *t*<sub>R</sub> = 7.37 min, 87 %, hydrophobic method.

#### 4-(4-cyano-3-methylphenoxy)benzyl methanesulfonate (41)



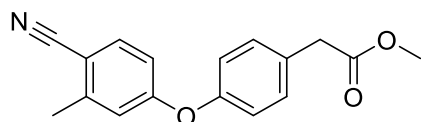
4-(4-(Hydroxymethyl)phenoxy)-2-methylbenzonitrile (391 mg, 1.00 eq., 1.64 mmol) and Et<sub>3</sub>N (218 mg, 1.30 eq, 2.15 mmol) were added to DCM (1 mL) in a round bottom flask and cooled to 0 °C. Methanesulfonyl chloride (375 mg, 2.00 eq, 3.27 mmol) was added and the mixture was stirred for 4 hours. The reaction mixture was slowly warmed to room temperature and diluted with NaHCO<sub>3</sub> (aq., 1 M) and the organic layer was concentrated to give the product as a white solid (510 mg, 98 %). The material obtained was 75 % purity and was used without further purification. **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.54 (d, *J* = 8.5 Hz, 1H), 7.42 (dt, *J* = 8.6, 4.9 Hz, 2H), 7.03 (dt, *J* = 8.6, 4.9 Hz, 2H), 6.88 (d, *J* = 2.5 Hz, 1H), 6.83 (dd, *J* = 8.5, 2.2 Hz, 1H), 4.60 (s, 2H), 2.50 (s, 3H). **LCMS** (*m/z*): 315.9 [M-H]<sup>-</sup>, *t<sub>R</sub>* = 3.51 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 7.18 min, 75 %, hydrophobic method.

#### (4-(4-Cyano-3-methylphenoxy)phenyl)methanesulfonic acid (29)

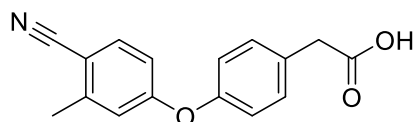


4-(4-Cyano-3-methylphenoxy)benzyl methanesulfonate (540 mg, 1.00 eq., 1.70 mmol) and sodium sulfite (2720 mg, 13.0 eq, 21.6 mmol) were added to ethanol (2 mL) and water (2 mL) and stirred for 20 h at 60 °C. The reaction was then evaporated to dryness and the crude material purified by reverse phase C18 chromatography using a gradient of 10 – 80 % methanol in water to give the product as a white solid (335 mg, 65 %). **<sup>1</sup>H NMR** (401 MHz, D<sub>2</sub>O) δ 7.68 (d, *J* = 8.5 Hz, 1H), 7.48 (d, *J* = 8.3 Hz, 2H), 7.15 (d, *J* = 8.3, 2H), 7.01 – 6.97 (m, 2H), 4.21 (s, 2H), 2.45 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, *d*<sub>6</sub>-DMSO) δ 161.18, 152.88, 144.31, 134.71, 132.54, 132.04, 119.36, 118.54, 117.79, 115.28, 105.46, 56.72, 20.00. **LCMS** (*m/z*): 301.9 [M-H]<sup>-</sup>, *t<sub>R</sub>* = 3.27 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 4.64 min, 98 %, hydrophobic method. The data was consistent with literature <sup>(253)</sup>.



**Methyl 2-(4-(4-cyano-3-methylphenoxy)phenyl)acetate<sup>(253)</sup> (38)**

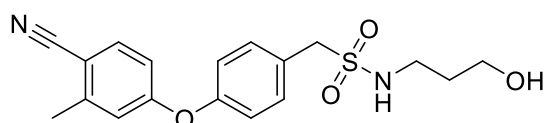
Prepared according to procedure A with 4-bromo-2-methyl-benzonitrile (1081 mg, 1.00 eq., 5.51 mmol), *N,N*-dimethylglycine hydrochloride (150 mg, 0.20 eq., 1.08 mmol), caesium carbonate (3320 mg, 1.80 eq., 10.1 mmol), copper iodide (106 mg, 0.10 eq., 0.551 mmol), and methyl 2-(4-hydroxyphenyl)acetate (1380 mg, 1.50 eq., 8.32 mmol). The crude material was purified by flash chromatography using a gradient of 0 – 5 % acetone in toluene to give the product as a yellow oil (900 mg, 58 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.52 (d, *J* = 8.5 Hz, 1H), 7.31 (dt, *J* = 8.7, 4.9 Hz, 2H), 7.01 (dt, *J* = 8.6, 4.9 Hz, 2H), 6.86 (d, *J* = 2.5 Hz, 1H), 6.82 (dd, *J* = 8.5, 2.4 Hz, 1H), 3.72 (s, 3H), 3.64 (s, 2H), 2.49 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 172.01, 161.42, 154.22, 144.6, 134.46, 131.17, 130.76, 120.53, 119.02, 118.29, 115.52, 106.64, 52.28, 40.49, 20.74. **LCMS** (*m/z*): 280.1 [M-H]<sup>-</sup>, *t*<sub>R</sub> = 3.75 min, hydrophobic method. **HPLC** *t*<sub>R</sub> = 6.74 min, 99 %, hydrophobic method. The data were consistent with the literature <sup>(253)</sup>.

**2-(4-(4-Cyano-3-methylphenoxy)phenyl)acetic acid <sup>(128)</sup> (28)**

Methyl 2-(4-(4-cyano-3-methylphenoxy)phenyl)acetate (900 mg, 1.00 eq., 3.20 mmol) and lithium hydroxide (394 mg, 5.10 eq, 16.5 mmol) were added to THF (15 mL) and water (15 mL) and stirred at room temperature for 18 hours. THF was evaporated from the reaction mixture and the aqueous layer was acidified using HCl (aq., 1 M) to pH 1. The solution was extracted with ethyl acetate (3 × 20 mL), washed with brine (2 × 20 mL) and dried over anhydrous sodium sulfate, filtered and evaporated to give the product as a yellow solid (737 mg, 86 %). **<sup>1</sup>H NMR** (401 MHz, *d*<sub>6</sub>-DMSO) 12.35 (s, 1H), 7.72 (d, *J* = 8.6 Hz, 1H), 7.34 (d, *J* = 8.5 Hz, 2H), 7.06 (d, *J* = 8.5 Hz, 2H), 7.01 (d, *J* = 2.2 Hz, 1H), 6.86 (dd, *J* = 8.6, 2.4 Hz, 1H), 3.60 (s, 2H), 2.42 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, *d*<sub>6</sub>-DMSO) δ 172.64, 160.96, 153.27, 144.34, 134.71, 131.85, 131.36,

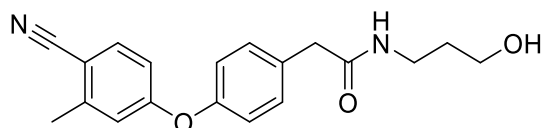
119.96, 118.77, 117.95, 115.35, 105.75, 51.75, 20.01. **LCMS** ( $m/z$ ): 266.8  $[M-H]^-$ ,  $t_R$  = 3.31 min, hydrophobic method. **HPLC**  $t_R$  = 5.922 min, 98 %, hydrophobic method. The data were consistent with the literature <sup>(128)</sup>.

**1-(4-(4-Cyano-3-methylphenoxy)phenyl)-N-(3-hydroxypropyl)methanesulfonamide (60)**



Prepared according to general procedure F with 3-aminopropan-1-ol (491 mg, 10.0 eq., 6.54 mmol) and [4-(4-cyano-3-methyl-phenoxy)phenyl]methanesulfonyl chloride (203 mg, 1.00 eq., 0.631 mmol). The crude material was concentrated to give the product as a yellow solid (16 mg, 7 %). **<sup>1</sup>H NMR** (401 MHz,  $d_6$ -DMSO)  $\delta$  11.07 (s, 1H), 7.78 (d,  $J$  = 8.6 Hz, 1H), 7.67 (d,  $J$  = 8.7 Hz, 2H), 7.19 (d,  $J$  = 8.6 Hz, 2H), 7.09 (d,  $J$  = 2.5 Hz, 1H), 6.94 (dd,  $J$  = 8.4, 2.3 Hz, 1H), 4.71 (s, 1H), 4.28 (s, 2H), 3.46 (dd,  $J$  = 11.9, 5.8 Hz, 2H), 2.82 (dd,  $J$  = 13.3, 7.0 Hz, 2H), 2.45 (s, 3H), 1.71 (dt,  $J$  = 13.0, 6.2 Hz, 2H). **<sup>13</sup>C NMR** (101 MHz,  $d_6$ -DMSO)  $\delta$  160.26, 155.56, 144.49, 134.80, 133.32, 126.92, 120.04, 119.38, 117.84, 115.91, 106.24, 57.77, 41.32, 36.50, 30.02, 19.99. **LCMS** ( $m/z$ ) 358.9  $[M-H]^-$ ,  $t_R$  = 3.01 min, hydrophobic method. **HPLC**  $t_R$  = 4.93 min, 95 %, hydrophobic method. **HRMS** did not ionise.

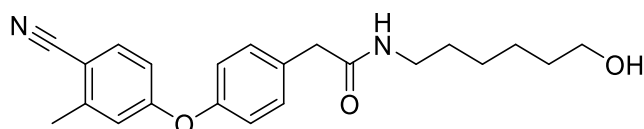
**2-(4-(4-Cyano-3-methylphenoxy)phenyl)-N-(3-hydroxypropyl)acetamide (62)**



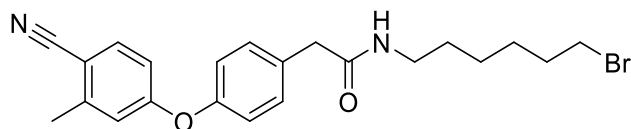
Prepared using general procedure E using 2-(4-(4-cyano-3-methylphenoxy)phenyl)acetic acid (242 mg, 1.00 eq., 0.905 mmol), 3-amino-propanol (99 mg, 1.5 eq., 1.3 mmol), HATU (592 mg, 1.70 eq., 1.56 mmol), and Et<sub>3</sub>N (145 mg, 1.60 eq., 1.44 mmol). The crude material was purified by reverse phase C18 flash chromatography using gradient of 5 – 100 % MeOH in water to give the product as a

white solid (194 mg, 66 %). **<sup>1</sup>H NMR** (401 MHz, *d*<sub>6</sub>-DMSO) δ 8.02 (t, *J* = 5.3 Hz, 1H), 7.74 (d, *J* = 8.6 Hz, 1H), 7.33 (d, *J* = 8.6 Hz, 2H), 7.06 (d, *J* = 8.6 Hz, 2H), 7.01 (d, *J* = 2.4 Hz, 1H), 6.87 (dd, *J* = 8.6, 2.2 Hz, 1H), 4.41 (t, *J* = 5.2 Hz, 1H), 3.41 – 3.38 (m, 4H), 3.10 (dt, *J* = 12.7, 6.9 Hz, 2H), 2.44 (s, 3H), 1.55 (ap, *J* = 6.7 Hz, 2H). **<sup>13</sup>C NMR** (101 MHz, *d*<sub>6</sub>-DMSO) δ 169.90, 161.03, 152.96, 144.32, 134.71, 133.35, 130.83, 119.96, 118.64, 117.93, 115.30, 105.58, 58.37, 41.59, 35.87, 32.37, 19.98. **LCMS** (*m/z*): 325.0 [M+H]<sup>+</sup>, *t*<sub>R</sub> = 3.33 min, hydrophobic method. **HPLC** *t*<sub>R</sub> = 5.38 min, 99 %, hydrophobic method. **HRMS** (*m/z*): requires C<sub>19</sub>H<sub>20</sub>N<sub>2</sub>O<sub>3</sub> 325.1547 [M+H]<sup>+</sup>, found 325.1555.

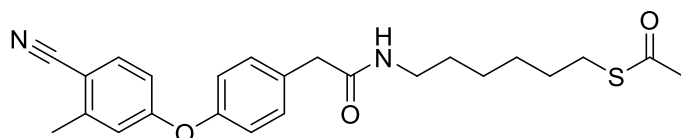
### 2-(4-(4-Cyano-3-methylphenoxy)phenyl)-*N*-(6-hydroxyhexyl)acetamide (61)



Prepared using general procedure B using 2-(4-(4-cyano-3-methylphenoxy)phenyl)acetic acid (47 mg, 1.0 eq., 0.17 mmol), 6-amino-hexanol (47 mg, 2.3 eq., 0.40 mmol), EDC.HCl (54 mg, 1.6 eq., 0.28 mmol), and HOBT (36 mg, 1.5 eq., 0.27 mmol). The crude material was purified by flash chromatography using a gradient of MeOH (0 – 10%) in DCM to give the product as a white solid (31 mg, 49 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.51 (d, *J* = 8.5 Hz, 1H), 7.29 (d, *J* = 8.6 Hz, 2H), 7.00 (d, *J* = 8.6 Hz, 2H), 6.85 (d, *J* = 2.4 Hz, 1H), 6.80 (dd, *J* = 8.6, 2.4 Hz, 1H), 5.81 (s, 1H), 3.59 (t, *J* = 6.5 Hz, 2H), 3.54 (s, 2H), 3.23 (dt, *J* = 13.0, 6.9 Hz, 2H), 2.47 (s, 3H), 1.58 – 1.43 (m, 4H), 1.41 – 1.21 (m, 4H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 170.93, 161.26, 154.26, 144.59, 134.44, 131.82, 131.17, 120.69, 119.06, 118.22, 115.51, 106.64, 62.60, 43.00, 39.67, 32.53, 29.52, 26.50, 25.32, 20.70. **LCMS** (*m/z*): 367.0 [M+H]<sup>+</sup>, *t*<sub>R</sub> = 3.44 min, hydrophobic method. **HPLC** *t*<sub>R</sub> = 5.85 min, 99 %, hydrophobic method. **HRMS** (*m/z*): requires C<sub>22</sub>H<sub>26</sub>N<sub>2</sub>O<sub>3</sub> 367.2016 [M+H]<sup>+</sup>, found 367.2025.

***N*-(6-Bromohexyl)-2-(4-(4-cyano-3-methylphenoxy)phenyl)acetamide (72)**

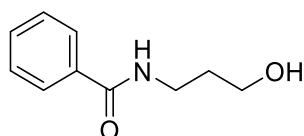
2-(4-(4-Cyano-3-methylphenoxy)phenyl)-*N*-(6-hydroxyhexyl)acetamide (85 mg, 1.0 eq., 0.23 mmol) and carbon tetrabromide (99 mg, 1.3 eq., 0.30 mmol) were added to dry DCM (2.5 mL) and cooled to 0 °C. Triphenylphosphine (72 mg, 1.2 eq., 0.28 mmol) was added and the reaction was slowly warmed to room temperature and stirred for 2.5 h. The reaction was diluted with toluene to precipitate a brown by-product precipitate. The solution was filtered and the filtrate was evaporated and was purified by reverse phase C18 flash chromatography using a gradient of methanol (0 – 10%) in water to give the product as a brown solid (24 mg, 25 %). **<sup>1</sup>H NMR** (401 MHz, *d*<sub>6</sub>-DMSO) δ 8.02 (t, *J* = 5.5 Hz, 1H), 7.74 (d, *J* = 8.6 Hz, 1H), 7.33 (d, *J* = 8.6 Hz, 2H), 7.05 (d, *J* = 8.6 Hz, 2H), 7.01 (d, *J* = 2.3 Hz, 1H), 6.87 (dd, *J* = 8.6, 2.1 Hz, 1H), 3.50 (t, *J* = 6.7 Hz, 2H), 3.41 (s, 2H), 3.04 (dt, *J* = 12.6, 6.8 Hz, 2H), 2.43 (s, 3H), 1.81 – 1.72 (m, 2H), 1.45 – 1.21 (m, 6H). **<sup>13</sup>C NMR** (101 MHz, *d*<sub>6</sub>-DMSO) δ 169.74, 161.02, 152.96, 144.30, 134.70, 133.39, 130.80, 119.94, 118.64, 117.92, 115.28, 105.58, 41.62, 38.47, 35.05, 32.16, 28.86, 27.20, 25.44, 19.98. **LCMS** (*m/z*): 430.0 [M+H]<sup>+</sup>, *t*<sub>R</sub> = 3.55 min, hydrophobic method. **HPLC** *t*<sub>R</sub> = 6.95 min, hydrophobic method. **HRMS** (*m/z*): requires C<sub>22</sub>H<sub>25</sub>BrN<sub>2</sub>O<sub>2</sub> 429.1172 [M+H]<sup>+</sup>, found 429.1181.

***S*-(6-(2-(4-(4-Cyano-3-methylphenoxy)phenyl)acetamido)hexyl)ethanethioate (73)**

Formed using general procedure D using *N*-(6-bromohexyl)-2-(4-(4-cyano-3-methylphenoxy)phenyl)acetamide (1204 mg, 1.00 eq., 2.80 mmol), Et<sub>3</sub>N (782 μL, 2.00 eq., 5.61 mmol) and thioacetic acid (426 mg, 2.00 eq. 5.61 mmol) in ethanol (28 mL). The crude material was purified by flash chromatography using a gradient of

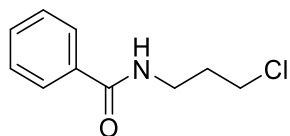
0 – 5 % acetone in 1:1 ethyl acetate:petroleum spirits to give the product as a yellow solid (83 mg, 7 %). **<sup>1</sup>H NMR** (401 MHz, MeOD)  $\delta$  7.62 (d,  $J$  = 8.6 Hz, 1H), 7.38 (d,  $J$  = 8.5 Hz, 2H), 7.05 (d,  $J$  = 8.1 Hz, 2H), 6.96 (d,  $J$  = 2.6 Hz, 1H), 6.88 (dd,  $J$  = 8.6, 2.4 Hz, 1H), 3.53 (s, 2H), 3.44 (t,  $J$  = 6.7 Hz, 2H), 3.26 – 3.16 (m, 2H), 2.48 (s, 3H), 2.30 (s, 3H), 1.85 (dt,  $J$  = 14.5, 6.8 Hz, 2H), 1.62 – 1.50 (m, 2H), 1.51 – 1.42 (m, 2H), 1.41 – 1.32 (m, 2H). **<sup>13</sup>C NMR** (151 MHz, CDCl<sub>3</sub>)  $\delta$  170.82, 160.95, 153.40, 143.97, 133.89, 131.88, 130.63, 130.33, 120.06, 118.45, 117.74, 114.90, 105.81, 42.25, 42.11, 40.01, 39.26, 37.71, 30.19, 29.08, 25.90, 20.14. **LCMS** ( $m/z$ ): 425.0 [M+H]<sup>+</sup>,  $t_R$  = 3.49 min, hydrophobic method. **HPLC**  $t_R$  = 6.83 min, 72 %, hydrophobic method. **HRMS** did not ionise.

#### ***N*-(3-hydroxypropyl)benzamide** <sup>(288)</sup> (**64**)



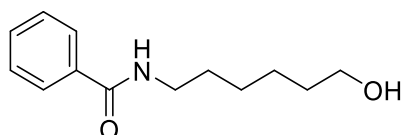
Prepared using general procedure E with benzoic acid (606 mg, 1.00 eq., 4.96 mmol), HATU (2227 mg, 1.20 eq., 5.86 mmol), Et<sub>3</sub>N (980 mg, 2.00 eq., 9.69 mmol) and 3-amino-propan-1-ol (748 mg, 2.00 eq., 9.96 mmol) The crude material was purified by flash chromatography using a gradient of 0-10 % MeOH in 1:1 ethyl acetate : petroleum spirits and to give the product as a yellow oil (371 mg, 42 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>)  $\delta$  7.77 (d,  $J$  = 6.9 Hz, 2H), 7.50 (t,  $J$  = 7.3 Hz, 1H), 7.43 (t,  $J$  = 7.4 Hz, 2H), 6.71 (s, 1H), 3.73 (t,  $J$  = 5.7 Hz, 2H), 3.64 (aq,  $J$  = 6.1 Hz, 2H), 1.80 (ap,  $J$  = 6.0 Hz, 2H). **LCMS** ( $m/z$ ): 178.0 [M-H]<sup>-</sup>,  $t_R$  = 2.78 min, hydrophobic method. **HPLC**  $t_R$  = 3.10 min, 95 %, hydrophobic method. The data were consistent with the literature <sup>(288)</sup>.

### ***N*-(3-chloropropyl)benzamide (66)**



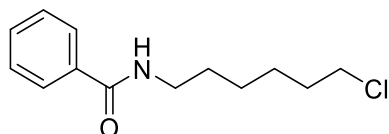
*N*-(3-hydroxypropyl)benzamide (69 mg, 1.0 eq., 0.38 mmol) was added to chloroform (2 mL), cooled to 0 °C and thionyl chloride (0.40 mL, 14 eq., 5.5 mmol) was added dropwise. The reaction was then heated to reflux for 16 h. After this time, the reaction was evaporated under a flow of nitrogen and the product used without further purification. **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.77 (d, *J* = 6.9 Hz, 2H), 7.51 (t, *J* = 7.3 Hz, 1H), 7.44 (t, *J* = 7.3 Hz, 2H), 3.70 – 3.60 (m, 4H), 2.13 (ap, *J* = 6.5 Hz, 2H). **LCMS** (*m/z*): 198.0 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 3.20 min, hydrophobic method.

### ***N*-(3-hydroxyhexyl)benzamide <sup>(289)</sup> (63)**



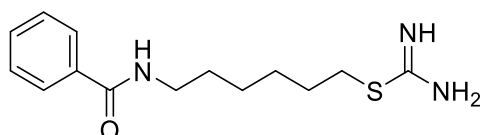
Prepared using general procedure B with benzoic acid (588 mg, 1.00 eq. 4.81 mmol), 6-amino-hexanol (549 mg, 1.00 eq., 4.68 mmol), EDC.HCl (1220 mg, 1.20 eq., 5.85 mmol), and HOBt (989 mg, 1.50 eq., 7.32 mmol). The crude material was purified by flash chromatography using a ratio of 10:90 methanol:DCM to give the product as a white solid (573 mg, 55 %). **<sup>1</sup>H NMR** (401 MHz, *d*<sub>6</sub>-DMSO) δ 8.42 (t, *J* = 5.2 Hz, 1H), 7.83 (d, *J* = 7.2 Hz, 2H), 7.51 (t, *J* = 7.2 Hz, 1H), 7.44 (t, *J* = 7.5 Hz, 2H), 4.33 (t, *J* = 5.1 Hz, 1H), 3.38 (aq, *J* = 6.3 Hz, 2H), 3.33 (s, 1H), 3.24 (aq, *J* = 6.8 Hz, 2H), 1.56 – 1.46 (m, 2H), 1.46 – 1.36 (m, 2H), 1.36 – 1.26 (m, 4H). **LCMS** (*m/z*): 222.1 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 3.16 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 4.23 min, 94 %, hydrophobic method. The data were consistent with the literature <sup>(289)</sup>.

### ***N*-(3-chlorohexyl)benzamide (65)**



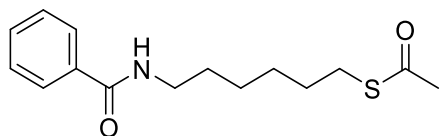
*N*-(3-hydroxyhexyl)benzamide (160 mg, 1.00 eq., 0.723 mmol) was added to chloroform (2 mL), cooled to 0 °C and thionyl chloride (0.50 mL, 9.5 eq., 6.89 mmol) was added dropwise. The reaction was then heated to reflux for 16 h. After this time, the reaction was evaporated under a flow of nitrogen and the product used without further purification. **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.76 (d, *J* = 6.9 Hz, 2H), 7.50 (t, *J* = 7.3 Hz, 1H), 7.43 (t, *J* = 7.3 Hz, 2H), 6.13 (s, 1H), 3.54 (t, *J* = 6.7 Hz, 2H), 3.47 (td, *J* = 7.2, 5.8 Hz, 2H), 1.85 – 1.73 (m, 2H), 1.70 – 1.58 (m, 2H), 1.55 – 1.38 (m, 4H). **LCMS** (*m/z*): 240.1 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 3.46 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 5.81 min, 95 %, hydrophobic method.

### **6-benzamidohexyl carbamimidothioate (68)**



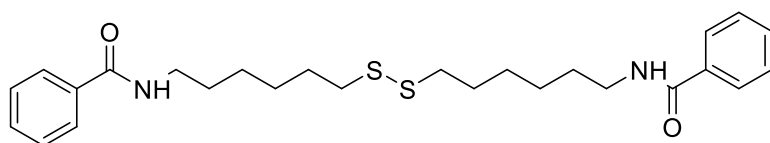
Prepared using general procedure C with *N*-(3-chlorohexyl)benzamide (28 mg, 1.0 eq., 0.11 mmol) and thiourea (42 mg, 4.8 eq., 0.55 mmol). The crude material was used without further purification. **<sup>1</sup>H NMR** (401 MHz, *d*<sub>6</sub>-DMSO) δ 7.86 – 7.78 (m, 2H), 7.55 – 7.40 (m, 3H), 3.27 – 3.20 (m, 2H), 3.18 – 3.06 (m, 2H), 1.63 – 1.48 (m, 4H), 1.45 – 1.29 (m, 4H). **<sup>13</sup>C NMR** (101 MHz, *d*<sub>6</sub>-DMSO) δ 169.97, 166.10, 134.67, 131.00, 128.23, 127.12, 37.76, 30.07, 28.92, 28.30, 27.54, 25.85. **LCMS** (*m/z*): 280.0 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 3.02 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 4.75 min, 44 %, hydrophobic method. **HRMS** (*m/z*): C<sub>14</sub>H<sub>21</sub>N<sub>3</sub>OS requires 280.1478 [M+H]<sup>+</sup>, found 280.1483.

### S-(6-benzamidohexyl) ethanethioate (71)



Prepared using general procedure D with *N*-(3-bromohexyl)benzamide (58 mg, 1.0 eq., 0.20 mmol), Et<sub>3</sub>N (61  $\mu$ L, 2.1 eq., 0.43 mmol) and thioacetic acid (28 mg, 1.8 eq., 0.37 mmol). The crude material was purified by reverse phase C18 chromatography using a gradient of methanol (0 – 10%) in water to give the product as a white solid (37 mg, 64 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>)  $\delta$  7.76 (d,  $J$  = 6.9 Hz, 2H), 7.49 (t,  $J$  = 7.3 Hz, 1H), 7.42 (t,  $J$  = 7.3 Hz, 2H), 6.18 (s, 1H), 3.45 (aq,  $J$  = 6.4 Hz, 2H), 2.87 (t,  $J$  = 7.2 Hz, 2H), 2.32 (s, 3H), 1.70 – 1.52 (m, 4H), 1.47 – 1.37 (m, 4H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>)  $\delta$  196.18, 167.54, 134.8, 131.39, 128.56, 126.98, 40.02, 30.70, 29.53, 29.45, 28.99, 28.38, 26.46 **LCMS** ( $m/z$ ): 280.0 [M+H]<sup>+</sup>,  $t_R$  = 3.50 min, hydrophobic method. **HPLC**  $t_R$  = 5.21 min, 98 %, hydrophobic method. **HRMS** ( $m/z$ ): requires C<sub>15</sub>H<sub>21</sub>NO<sub>2</sub>S 280.1366 [M+H]<sup>+</sup>, found 280.1372.

### *N,N*-(Disulfanediyldis(hexane-6,1-diyl))dibenzamide (69)

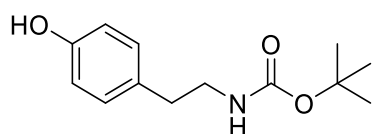


S-(6-Benzamidohexyl)ethanethioate (12 mg, 1.0 eq., 0.043 mmol) was dissolved in NaOH (aq., 1 M, 1 mL) and MeCN (1 mL) and stirred at room temperature for 16 h. After this time, the reaction was quenched with HCl (aq., 1 M, 1 mL) and diluted with DCM (5 mL). The organic layer was dried over anhydrous magnesium sulfate, concentrated and the crude material was purified by reverse phase C18 flash chromatography using a gradient of methanol (0 – 10%) in water to give the product as a white solid (4.3 mg, 21 %). **<sup>1</sup>H NMR** (401 MHz, CD<sub>3</sub>OD)  $\delta$  7.80 (ad,  $J$  = 7.2, 4H), 7.54 – 7.49 (m, 2H), 7.65 – 7.42 (m, 4H), 3.37 (t,  $J$  = 7.2 Hz, 4H), 2.69 (t,  $J$  = 7.0, 4H), 1.73 – 1.59 (m, 8H), 1.50 – 1.36 (m, 8H). **<sup>13</sup>C NMR** (101 MHz, CD<sub>3</sub>OD)  $\delta$  170.25,



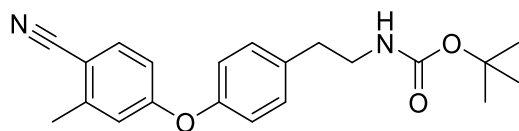
135.90, 132.53, 129.53, 128.22, 40.96, 39.67, 30.39, 30.13, 29.14, 27.67. **LCMS** ( $m/z$ ): 473.1  $[M+H]^+$ ,  $t_R$  = 3.7.26 min, hydrophobic method. **HPLC**  $t_R$  = 6.30 min, 98 %, hydrophobic method. **HRMS** ( $m/z$ ): requires  $C_{26}H_{36}N_2O_2S_2$  473.2291  $[M+H]^+$ , found 473.2303.

***tert*-Butyl (4-hydroxyphenethyl)carbamate <sup>(290)</sup> (111)**



Di-*tert*-butyl dicarbonate (8.925 g, 1.00 eq., 40.89 mmol) and tyramine (5.752 g, 1.02 eq. 41.93 mmol) were dissolved in dry THF (82 mL) and stirred at room temperature for 2 hours. After this time, the solvent was removed and the crude material was purified by flash chromatography using a gradient of acetone (0 – 10 %) in DCM to give the product as a white solid (9.07 g, 94 %). **<sup>1</sup>H NMR** (401 MHz,  $CDCl_3$ )  $\delta$  7.00 (d,  $J$  = 8.5 Hz, 2H), 6.78 (d,  $J$  = 8.5 Hz, 2H), 5.65 (s, 1H), 4.64 (s, 1H), 3.32 (t,  $J$  = 7.0 Hz, 2H), 2.70 (t,  $J$  = 7.1 Hz, 2H), 1.44 (s, 9H). **LCMS** ( $m/z$ ):  $t_R$  = 236.0  $[M-H]^-$ , 2.73 min, hydrophobic method. **HPLC**  $t_R$  = 4.60 min, 98 %, hydrophobic method. The data were consistent with the literature <sup>(290)</sup>.

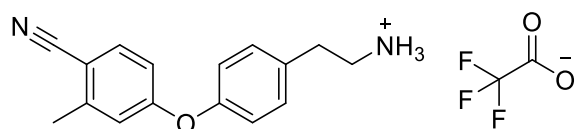
***tert*-Butyl (4-(4-cyano-3-methylphenoxy)phenethyl)carbamate (112)**



Prepared according to general procedure A with *N,N*-dimethylglycine (784 mg, 0.300 eq., 5.62 mmol), 4-bromo-2-methyl-benzonitrile (4475 mg, 1.20 eq., 22.83 mmol), copper iodide (708 mg, 0.200 eq., 3.71 mmol), caesium carbonate (9.29 g, 1.50 eq., 28.5 mmol), and *tert*-butyl(4-hydroxyphenethyl)carbamate (4.50 g, 1.00 eq., 19.0 mmol). The crude material was purified by flash chromatography using a gradient of 0 – 30 % EtOAc in petroleum spirits to give the product as a white solid (5.844 g, 87 %). **<sup>1</sup>H NMR** (401 MHz,  $CDCl_3$ )  $\delta$  7.50 (d,  $J$  = 8.5 Hz, 1H), 7.21

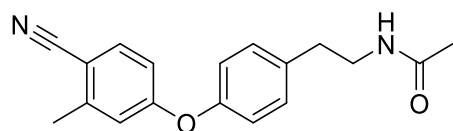
(d,  $J = 8.5$  Hz, 2H), 6.97 (d,  $J = 8.5$  Hz, 2H), 6.83 (d,  $J = 2.4$  Hz, 1H), 6.78 (dd,  $J = 8.5, 2.2$  Hz, 1H), 4.59 (s, 1H), 3.37 (t,  $J = 7.1$  Hz, 2H), 2.79 (t,  $J = 7.1$  Hz, 2H), 2.47 (s, 3H), 1.42 (s, 9H).  **$^{13}\text{C}$  NMR** (101 MHz,  $\text{CDCl}_3$ )  $\delta$  161.57, 155.95, 153.45, 144.47, 135.89, 134.37, 130.53, 120.53, 118.78, 118.27, 115.26, 106.34, 79.44, 42.06, 35.77, 28.47, 20.68. **LCMS** ( $m/z$ ): 351.1  $[\text{M}-\text{H}]^-$ ,  $t_R = 3.26$  min, hydrophobic method. **HPLC**  $t_R = 6.38$  min, 98 %, hydrophobic method. **HRMS** did not ionise.

#### 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (110)



*tert*-Butyl (4-(4-cyano-3-methylphenoxy)phenethyl)carbamate (2.01 g, 1.00 eq. 5.71 mmol) and dissolved 1:4 trifluoroacetic acid:DCM (63 mL) and stirred at room temperature for 1.5 h. The solvent was then evaporated to give the product as a light brown solid (2.08 g, 99 %).  **$^1\text{H}$  NMR** (401 MHz,  $\text{CDCl}_3$ )  $\delta$  7.51 (d,  $J = 8.6$  Hz, 1H), 7.23 (d,  $J = 8.4$  Hz, 2H), 7.00 (d,  $J = 8.4$  Hz, 2H), 6.85 (d,  $J = 2.4$  Hz, 1H), 6.78 (dd,  $J = 8.5, 2.5$  Hz, 1H), 3.31 (s, 3H), 3.03 (t,  $J = 7.3$  Hz, 2H), 2.47 (s, 2H).  **$^{13}\text{C}$  NMR** (101 MHz,  $\text{CDCl}_3$ )  $\delta$  161.29, 154.67, 144.79, 134.57, 131.96, 130.58, 121.03, 119.19, 118.17, 115.53, 106.55, 41.59, 32.91, 20.66. **LCMS** ( $m/z$ ): 254.1  $[\text{M}+\text{H}]^+$ ,  $t_R = 3.07$  min, hydrophobic method. **HPLC**  $t_R = 4.46$ , 98 %, hydrophobic method. **HRMS** ( $m/z$ ):  $\text{C}_{16}\text{H}_{17}\text{N}_2\text{O}$  requires 253.1335  $[\text{M}+\text{H}]^+$ ; found 253.1324.

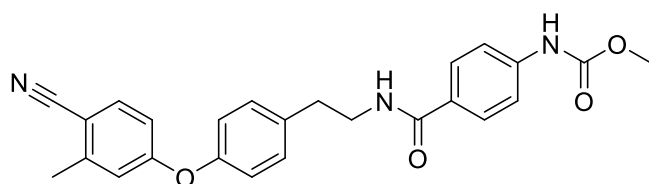
#### *N*-(4-(4-Cyano-3-methylphenoxy)phenethyl)acetamide (149)



Prepared according to general procedure F with 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (94 mg, 1.0 eq., 0.26 mmol), acetyl chloride (55  $\mu\text{L}$ , 3.0 eq., 0.77 mmol) and  $\text{Et}_3\text{N}$  (72  $\mu\text{L}$ , 2.0 eq., 0.51 mmol). The crude material

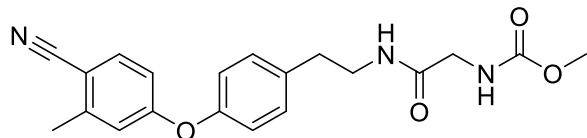
was purified by flash chromatography using a gradient of 0 – 10 % MeOH in EtOAc to give the product as a white solid (37 mg, 48 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.52 (d, *J* = 8.5 Hz, 1H), 7.22 (d, *J* = 8.4 Hz, 2H), 6.99 (d, *J* = 8.5 Hz, 2H), 6.85 (d, *J* = 2.4 Hz, 1H), 6.79 (dd, *J* = 8.5, 2.3 Hz, 1H), 5.86 (s, 1H), 3.52 (q, *J* = 6.7 Hz, 2H), 2.84 (t, *J* = 7.2 Hz, 2H), 2.49 (s, 3H), 2.00 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 170.53, 161.54, 153.69, 144.58, 135.60, 134.45, 130.50, 120.65, 118.91, 118.30, 115.34, 106.51, 40.37, 34.13, 23.27, 20.77. **LCMS** (*m/z*): 295.1 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 3.51, hydrophobic method. **HPLC** *t<sub>R</sub>* = 5.33 min, 98 %, hydrophobic method. **HRMS** (*m/z*): C<sub>18</sub>H<sub>18</sub>N<sub>2</sub>O<sub>2</sub> requires 295.1441 [M+H]<sup>+</sup>; found 295.1448.

**Methyl (4-((4-(4-cyano-3-methylphenoxy)phenethyl)carbamoyl)phenyl)carbamate (132)**



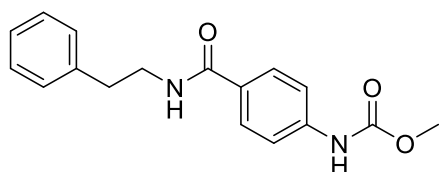
Prepared according to general procedure E with 4-(methoxycarbonylamino)benzoic acid (25 mg, 1.0 eq., 0.13 mmol), Et<sub>3</sub>N (36 μL, 2.0 eq., 0.29 mmol), 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (93 mg, 2.0 eq., 0.25 mmol) and HATU (81 mg, 1.7 eq., 0.21 mmol). The organic layer was concentrated and triturated using DCM to give the product as a white solid (47 mg, 86 %). **<sup>1</sup>H NMR** (401 MHz, MeOD) δ 7.71 (d, *J* = 8.8 Hz, 2H), 7.57 (d, *J* = 8.6 Hz, 1H), 7.51 (d, *J* = 8.7 Hz, 2H), 7.32 (d, *J* = 8.4 Hz, 2H), 7.00 (d, *J* = 8.5 Hz, 2H), 6.89 (d, *J* = 2.5 Hz, 1H), 6.81 (dd, *J* = 8.6, 2.5 Hz, 1H), 3.74 (s, 3H), 3.60 (t, *J* = 7.3 Hz, 2H), 2.93 (t, *J* = 7.3 Hz, 2H), 2.44 (s, 3H). **LCMS** (*m/z*): 430.2 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 3.94 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 5.67 min, 98 %, hydrophobic method. **HRMS** (*m/z*): requires C<sub>25</sub>H<sub>23</sub>N<sub>3</sub>O<sub>4</sub> 430.1761 [M+H]<sup>+</sup>, found 430.1763.

**Methyl (2-((4-(4-cyano-3-methylphenoxy)phenethyl)amino)-2-oxoethyl)carbamate (131)**



Prepared according to general procedure E with 2-(methoxycarbonylamino)acetic acid (27 mg, 1.0 eq., 0.20 mmol), 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (104 mg, 1.40 eq., 0.283 mmol), HATU (152 mg, 2.00 eq., 0.399 mmol), and Et<sub>3</sub>N (112  $\mu$ L, 4.0 eq., 0.80 mmol). The crude material was purified by flash chromatography using a gradient of 0 – 10 % MeOH in chloroform to give the product as a white solid (41 mg, 56 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>)  $\delta$  7.50 (d,  $J$  = 8.5 Hz, 1H), 7.20 (d,  $J$  = 8.5 Hz, 2H), 6.96 (d,  $J$  = 8.5 Hz, 2H), 6.83 (d,  $J$  = 2.4 Hz, 1H), 6.78 (dd,  $J$  = 8.5, 2.2 Hz, 1H), 6.40 (s, 1H), 5.56 (s, 1H), 3.80 (d,  $J$  = 4.4 Hz, 2H), 3.66 (s, 3H), 3.52 (aq,  $J$  = 6.8 Hz, 2H), 2.82 (t,  $J$  = 7.2 Hz, 2H), 2.47 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>)  $\delta$  169.24, 161.45, 157.33, 153.55, 144.46, 135.39, 134.34, 130.43, 120.53, 118.76, 118.22, 115.19, 106.31, 52.59, 44.68, 40.71, 34.99, 20.63. **LCMS** ( $m/z$ ): 368.2 [M+H]<sup>+</sup>,  $t_R$  = 3.60 min, hydrophobic method. **HPLC**  $t_R$  = 5.24 min, 95 %, hydrophobic method. **HRMS** ( $m/z$ ): requires C<sub>20</sub>H<sub>21</sub>N<sub>3</sub>O<sub>4</sub> 368.1605 [M+H]<sup>+</sup>, found 368.1616.

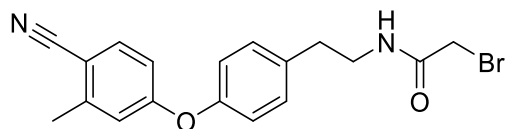
**Methyl (4-(phenethylcarbamoyl)phenyl)carbamate (148)**



Prepared according to general procedure E with 4-(methoxycarbonylamino)benzoic acid (41 mg, 1.0 eq., 0.21 mmol), phenethylamine (53  $\mu$ L, 2.0 eq., 0.42 mmol) HATU (122 mg, 1.50 eq., 0.321 mmol), and Et<sub>3</sub>N (58  $\mu$ L, 2.0 eq., 0.42 mmol). The crude material was purified by flash chromatography using a ratio of 1:9 acetone in DCM to

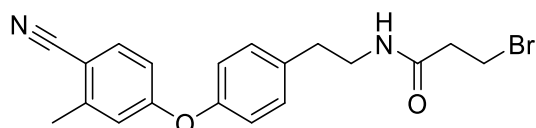
give the product as a yellow solid (47 mg, 75 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.63 (d, *J* = 8.8 Hz, 2H), 7.43 (d, *J* = 8.7 Hz, 2H), 7.35 – 7.29 (m, 2H), 7.26 – 7.20 (m, 3H), 7.11 (s, 1H), 3.77 (s, 3H), 3.70 (aq, *J* = 6.8 Hz, 2H), 2.92 (t, *J* = 7.0 Hz, 2H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 167.23, 153.92, 141.23, 139.00, 137.83, 128.94, 128.84, 128.15, 126.73, 118.10, 52.62, 41.35, 35.81. **LCMS** (*m/z*): 299.1 [M+H]<sup>+</sup>, *t<sub>R</sub>* = 2.67 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 4.78 min, 98 %, hydrophobic method. **HRMS** (*m/z*): C<sub>17</sub>H<sub>18</sub>N<sub>2</sub>O<sub>3</sub> requires 299.139 [M+H]<sup>+</sup>, found 299.1398.

### 2-Bromo-*N*-(4-(4-cyano-3-methylphenoxy)phenethyl)acetamide (143)



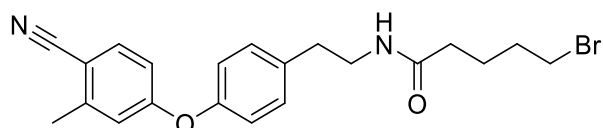
Prepared according to general procedure F with 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (104 mg, 1.28 eq., 0.285 mmol), bromoacetyl chloride (35 mg, 1.0 eq., 0.22 mmol) and Et<sub>3</sub>N (62 μL, 2.0 eq., 0.45 mmol). The crude material was purified by flash chromatography using 1:1 EtOAc in petroleum spirits to give the product as a yellow solid (29 mg, 34 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>) δ 7.53 (d, *J* = 8.5 Hz, 1H), 7.24 (d, *J* = 8.5 Hz, 2H), 7.01 (d, *J* = 8.5 Hz, 2H), 6.84 (d, *J* = 2.4 Hz, 1H), 6.80 (dd, *J* = 8.4, 2.3 Hz, 1H), 6.52 (s, 1H), 3.87 (s, 2H), 3.56 (aq, *J* = 7.0 Hz, 2H), 2.87 (t, *J* = 7.1 Hz, 2H), 2.49 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 165.42, 161.57, 153.86, 144.60, 135.20, 134.48, 130.60, 120.79, 118.90, 118.32, 115.38, 106.57, 41.45, 34.93, 29.44, 20.78. **LCMS** (*m/z*): 373.1 [M+H]<sup>+</sup>, 3.71 min, hydrophobic method. **HPLC** *t<sub>R</sub>* = 5.59 min, 98 %, hydrophobic method. **HRMS** did not ionise.

### 3-Bromo-*N*-(4-(4-cyano-3-methylphenoxy)phenethyl)propenamide (144)



Prepared according to general procedure F with 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (129 mg, 1.50 eq., 0.351 mmol), 3-bromopropenoyl chloride (40 mg, 1.0 eq., 0.23 mmol) and Et<sub>3</sub>N (65  $\mu$ L, 2.0 eq., 0.47 mmol). The crude material was purified by flash chromatography using 1:1 EtOAc in petroleum spirits to give the product as a yellow solid (61 mg, 68 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>)  $\delta$  7.53 (d,  $J$  = 8.5 Hz, 1H), 7.24 (d,  $J$  = 8.6 Hz, 2H), 6.99 (d,  $J$  = 8.5 Hz, 2H), 6.85 (d,  $J$  = 2.4 Hz, 1H), 6.80 (dd,  $J$  = 8.5, 2.1 Hz, 1H), 5.61 (s, 1H), 3.64 (t,  $J$  = 6.5 Hz, 2H), 3.57 (aq,  $J$  = 7.0 Hz, 2H), 2.86 (t,  $J$  = 7.0 Hz, 2H), 2.72 (t,  $J$  = 6.5 Hz, 2H), 2.49 (s, 3H). **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>)  $\delta$  169.74, 161.54, 153.74, 144.59, 135.51, 134.47, 130.59, 120.68, 118.92, 118.32, 115.37, 106.53, 40.98, 39.94, 35.17, 27.66, 20.78. **LCMS** ( $m/z$ ): 387.1 [M+H]<sup>+</sup>, 3.71 min, hydrophobic method. **HPLC**  $t_R$  = 5.705 min, 95 %, hydrophobic method. **HRMS** did not ionise.

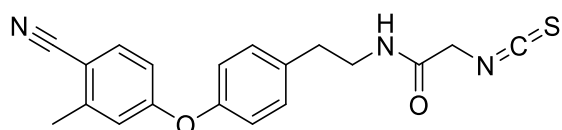
### 5-Bromo-*N*-(4-(4-cyano-3-methylphenoxy)phenethyl)pentanamide (145)



Prepared according to general procedure F with 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (70 mg, 1.1 eq., 0.19 mmol), 5-bromopentanoyl chloride (35 mg, 1.0 eq., 0.18 mmol) and Et<sub>3</sub>N (36 mg, 2.0 eq., 0.35 mmol). The crude material was purified by flash chromatography using 1:1 EtOAc in petroleum spirits to give the product as a yellow solid (43 mg, 60 %). **<sup>1</sup>H NMR** (401 MHz, CDCl<sub>3</sub>)  $\delta$  7.52 (d,  $J$  = 8.5 Hz, 1H), 7.22 (d,  $J$  = 8.5 Hz, 2H), 6.99 (d,  $J$  = 8.5 Hz, 2H), 6.85 (d,  $J$  = 2.4 Hz, 1H), 6.80 (dd,  $J$  = 8.5, 2.2 Hz, 1H), 5.52 (s, 1H), 3.53 (aq,  $J$  = 6.9 Hz, 2H), 3.40 (t,  $J$  = 6.5 Hz, 2H), 2.84 (t,  $J$  = 7.1 Hz, 2H), 2.49 (s, 3H), 2.19 (t,  $J$  = 7.3 Hz, 2H), 1.99 – 1.83 (m, 2H), 1.84 – 1.49 (m, 2H). **<sup>13</sup>C NMR** (151 MHz, CDCl<sub>3</sub>)  $\delta$  172.53, 161.54, 153.71, 144.58, 135.63, 134.45, 130.52,

120.63, 118.94, 118.30, 115.36, 106.52, 40.80, 35.63, 35.22, 33.32, 32.15, 24.33, 20.76. **LCMS** ( $m/z$ ): 416.1  $[M+H]^+$ , 3.83 min, hydrophobic method. **HPLC**  $t_R$  = 5.95 min, 98 %, hydrophobic method. **HRMS** ( $m/z$ ):  $C_{21}H_{23}N_2O_2Br$  requires 415.1016  $[M+H]^+$ ; found 415.1014.

***N*-(4-(4-Cyano-3-methylphenoxy)phenethyl)-2-isothiocyanatoacetamide (125)**



Methyl 2-isothiocyanatoacetate (32 mg, 1.5 eq., 0.24 mmol) and HATU (63 mg, 1.0 eq., 0.17 mmol) were added to DMF (1.5 mL) and stirred until dissolved.  $Et_3N$  (44  $\mu$ L, 2.0 eq., 0.32 mmol) and KOH (aq., 2.5 M, 0.098 mL, 1.5 eq., 0.24 mmol) were added, followed by 4-[4-(2-aminoethyl)phenoxy]-2-methyl-benzonitrile trifluoroacetic acid salt (59 mg, 1.0 eq., 0.16 mmol) and the reaction was stirred at room temperature for 20 h. The reaction was then diluted with EtOAc (10 mL), washed with water (3  $\times$  10 mL) and dried over anhydrous magnesium sulfate. The crude material was purified by flash chromatography using a gradient of 0 – 100 % EtOAc in petroleum spirits to give the product as a yellow oil (31 mg, 56 %).  **$^1H$  NMR** (401 MHz,  $d_6$ -DMSO)  $\delta$  7.74 (d,  $J$  = 8.6 Hz, 1H), 7.32 (d,  $J$  = 8.5 Hz, 2H), 7.06 (d,  $J$  = 8.5 Hz, 2H), 7.01 (d,  $J$  = 1.6 Hz, 1H), 6.86 (dd,  $J$  = 8.6, 2.5 Hz, 1H), 4.23 (d,  $J$  = 5.6 Hz, 2H), 3.63 (s, 3H), 2.83 (t,  $J$  = 7.4 Hz, 2H), 2.43 (s, 2H).  **$^{13}C$  NMR** (101 MHz,  $d_6$ -DMSO)  $\delta$  170.55, 161.14, 152.79, 144.34, 136.16, 134.71, 130.59, 120.44, 120.21, 118.60, 117.97, 115.21, 105.54, 51.65, 44.94, 33.41, 20.02. **LCMS** ( $m/z$ ): 350.1  $[M-H]^-$ ,  $t_R$  = 2.96 min, hydrophobic method. **HPLC**  $t_R$  = 5.67 min, 98 %, hydrophobic method. **HRMS** ( $m/z$ ):  $C_{19}H_{17}N_3O_2S$  requires 725.1975  $[2M + Na]^+$ , found 725.1958.

## 8.2 Small molecule solubility

Purified compounds had their solubility and aggregation evaluated via a semi-quantitative 1D  $^1\text{H}$  NMR titration in aqueous buffer. Compounds were accurately weighed and dissolved in  $d_6$ -DMSO at 100 mM.  $\text{D}_2\text{O}$  buffer was prepared with 50 mM sodium phosphate, 25 mM NaCl at pH 7.4 or 50 mM HEPES, 50 mM NaCl at pH 6.8 with 100  $\mu\text{M}$  4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) internal standard. Serial dilution of the small molecule  $d_6$ -DMSO stocks was used to prepare a 4-point two fold dilution series with final concentrations of 125, 250, 500 and 1000  $\mu\text{M}$  with 2 %  $d_6$ -DMSO in buffer and a total volume of 600  $\mu\text{L}$ . Serial dilution of the peptide  $d_6$ -DMSO stocks was used to prepare a 5-point two fold dilution series with final concentrations of 0.5, 1, 2, 4 and 8 mM with 2 %  $d_6$ -DMSO in HEPES buffer and a total volume of 600  $\mu\text{L}$ . NMR spectra were recorded on a Bruker 400 MHz spectrometer with TD = 32 K, D1 = 5 s, 64 scans and water suppression at 4.7 ppm. NMR spectra were processed in MNOVA (version 6.0 – 14.2) and referenced by the DSS peak at 0.030 ppm. Compounds failed solubility at a given concentration if they did not show a doubling in intensity concordant with doubling concentration or showed chemical shift changes  $> 0.004$  ppm (1.6 Hz) from the lowest recorded concentration. Ligand concentrations were calculated relative to the DSS integral in the 1D  $^1\text{H}$  NMR spectra and used to calculate the stock concentration which was used in  $K_D$  determination experiments.

## 8.3 Protein production and purification

Unlabelled and  $^{15}\text{N}$ -labelled *EcDsbA* were expressed as described previously using autoinduction <sup>(252, 291)</sup>. All media components were sterilised prior to use by autoclave or 0.22  $\mu\text{m}$  filtration as appropriate. Pre-expression cultures were prepared by inoculating 10 mL Luria-Bertani broth supplemented with 50  $\mu\text{g}/\text{mL}$  kanamycin from frozen glycerol stocks of *E. coli* BL21(DE3) carrying the plasmid B0013-(5644bb) coding for *EcDsbA*, as described previously <sup>(229)</sup> and incubating for 16 h at 37 °C with agitation at 220 rpm. ZYM-5052 or N-5052 media supplemented with 50  $\mu\text{g}/\text{mL}$  kanamycin in baffled conical flasks were inoculated with 1 % v/v of pre-culture and incubated for 24-30 h at 37 °C with agitation at 170 rpm, after which the cells were harvested by centrifugation at 3200 $\times g$  at 4 °C for 20 min. The supernatant was



discarded and the pellets were stored at -20 °C for subsequent protein extraction and purification.

Each frozen bacterial pellet was thawed, resuspended in an equal volume of osmotic shock buffer (20 mM Tris (pH 9.0), 10 mM EDTA, 50 % w/v sucrose) and maintained at 4 °C with gentle stirring for 60 min, after which the cell suspension was rapidly diluted to 10 times its volume with deionised H<sub>2</sub>O at 4 °C. The suspension was maintained at 4 °C and stirred for a further 90 min before the cell pellet and lysate were separated by centrifugation at 50000×g at 4 °C for 30 min. The lysate was carefully decanted from the pellet and stored at -20 °C. The cell pellet was resuspended in 10 times its volume of lysis buffer (20 mM Tris (pH 8.0), 25 mM NaCl, 4 mg/mL colistin sulfate) and maintained at room temperature with gentle stirring for 18-24 h. This suspension was then centrifuged at 50000×g at 4 °C for 30 min and the lysate was decanted and added to the thawed osmotic shock lysate. The pellet was discarded and the target protein was purified from the combined lysate. (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was added to the combined lysate with gentle stirring to a concentration of 0.8 M and the solution was syringe filtered (0.22 µm, Millipore) for loading onto a HiLoad 1610 Phenyl Sepharose HP hydrophobic interaction chromatography column (GE Healthcare) using 20 mM Tris (pH 8.0), 50 mM NaCl, 1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. The bound proteins were eluted on a gradient from 1 – 0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Fractions were analysed by SDS-PAGE and those containing target protein were pooled and concentrated to 10 mL using an Amicon centrifugal diafiltration unit (10000 MWCO, Millipore) and buffer exchanged to 50 mM HEPES (pH 6.8), 50 mM NaCl using a HiPrep 2610 desalting column (GE Healthcare). Following buffer exchange, the fractions containing target protein were pooled and loaded onto a MonoQ HR 10/10 anion-exchange column (GE Healthcare). In these buffer conditions, pure target protein was collected in the flowthrough fraction and impurities were eluted using a gradient from 50 mM – 1 M NaCl. Protein purity was confirmed by SDS-PAGE.

Following purification by anion-exchange chromatography, each *EcDsbA* solution was concentrated to approximately 8 – 9 mL by diafiltration (Amicon 3000 MWCO, Millipore) and treated with 0.1 volumes of freshly prepared 15 mM copper-phenanthroline solution. After 1 h reaction time at 4 °C, the copper-phenanthroline was

removed by buffer exchange (50 mM HEPES pH 6.8, 50 mM NaCl) using a HiPrep 2610 desalting column.

If required the purified protein was reduced with a 10 – 20-fold excess of DTT. After 1 hour of incubation the DTT was removed by adding a maximum of 2.5 mL of protein to a PD-10 desalting column (GE Healthcare) and eluted over approximately 5 mL of the desired NMR buffer. This process was repeated twice and the purified protein was concentrated.

Protein samples were adjusted to their required final concentrations by concentration using Amicon centrifugal diafiltration units (3000 MWCO, Millipore) or dilution using storage/NMR buffer (50 mM HEPES (pH 6.8), 50 mM NaCl or 50 mM phosphate, 25 mM NaCl, pH 6.8). For NMR, D<sub>2</sub>O was added to a final concentration of 10 % (v/v). Final protein concentrations were quantified using a NanoDrop lite UV spectrophotometer (ThermoFisher) based on a calculated extinction coefficient for oxidised EcDsbA of  $\epsilon = 28545 \text{ M}^{-1}\text{cm}^{-1}$ .

## 8.4 NMR parameters and sample preparation

### 8.4.1 Peptide binding characterisation

A 100  $\mu\text{M}$  [U-<sup>15</sup>N]-labelled oxidised EcDsbA sample with 1 mM unlabelled peptide **23** in 50 mM HEPES, 50 mM NaCl, 2 mM EDTA, pH 6.8, 10 % D<sub>2</sub>O, 2 % *d*<sub>6</sub>-DMSO, was prepared to a sample volume of 160  $\mu\text{L}$  and acquired in a 3 mm tube. 2D <sup>1</sup>H-<sup>15</sup>N HSQC spectra of [U-<sup>15</sup>N]-labelled oxidised EcDsbA were acquired with and without peptide with number of scans (NS) = 8, dummy scans (DS) = 8, relaxation delay (D1) = 1 sec, total complex points of 2048 (<sup>1</sup>H<sup>N</sup>) and 128 (<sup>15</sup>N), spectral width of 30 ppm (<sup>15</sup>N) x 16.6 (<sup>1</sup>H<sup>N</sup>) with a transmitter frequency offset of 117.5 ppm (<sup>15</sup>N) and 4.7 ppm (<sup>1</sup>H<sup>N</sup>). Experiments were acquired on a 600 MHz spectrometer equipped with CryoProbe at 298 K.

#### 8.4.2 Backbone and aliphatic side chain assignments

Experiments were run on a Bruker 700 MHz NMR spectrometer with a CryoProbe at 298 K. A 300  $\mu\text{M}$  [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labelled oxidised *EcDsbA* sample with 3 mM peptide **23** in 50 mM HEPES, 50 mM NaCl, 2 mM EDTA, pH 6.8, 10 %  $\text{D}_2\text{O}$ , 2 %  $d_6$ -DMSO, was prepared to a sample volume of 330  $\mu\text{L}$  and acquired in a Shigemi tube. The details of NMR data acquisition are reported in the Table 8.1 below. All data were processed by Topspin and analysed by CARA.

Experiment	Linear/Non-uniform sampling (NUS)	Number of scans (NS)	Dummy scans (DS)	Relaxation delay (D1)/sec	Total complex points	Spectral width/ ppm	Offset/ ppm
2D $^{15}\text{N}$ HSQC	Linear	4	8	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 128 ( $^{15}\text{N}$ )	14.3 ( $^1\text{H}$ ) x 30 ( $^{15}\text{N}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ )
2D $^{13}\text{C}^{\text{ali}}$ HSQC	Linear	8	16	1	2048 ( $^1\text{H}^{\text{ali}}$ ) x 256 ( $^{13}\text{C}^{\text{ali}}$ )	14.3 ( $^1\text{H}^{\text{ali}}$ ) x 65 ( $^{13}\text{C}^{\text{ali}}$ )	4.7 ( $^1\text{H}$ ) x 38.5 ( $^{13}\text{C}$ )
3D HNCA	NUS (25 %)	24	16	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 128 ( $^{13}\text{C}^{\alpha}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 32 ( $^{13}\text{C}^{\alpha}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 54.5 ( $^{13}\text{C}^{\alpha}$ )
3D HNCACB	NUS (20 %)	48	32	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 128 ( $^{13}\text{C}^{\alpha\beta}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 61 ( $^{13}\text{C}^{\alpha\beta}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 42 ( $^{13}\text{C}^{\alpha\beta}$ )
3D CBCACONH	NUS (25 %)	24	16	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 96 ( $^{13}\text{C}^{\alpha\beta}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 61 ( $^{13}\text{C}^{\alpha\beta}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 42 ( $^{13}\text{C}^{\alpha\beta}$ )
3D HBHACONH	NUS (20 %)	32	16	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 128 ( $^1\text{H}^{\alpha\beta}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 10 ( $^1\text{H}^{\alpha\beta}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 4.7 ( $^1\text{H}^{\alpha\beta}$ )
3D (H)CC(CO)NH	NUS (20 %)	48	16	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 128 ( $^{13}\text{C}^{\text{ali}}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 65 ( $^{13}\text{C}^{\text{ali}}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 38.5 ( $^{13}\text{C}^{\text{ali}}$ )

3D H(CCCO)NH	NUS (20 %)	48	16	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 128 ( $^1\text{H}^{\text{ali}}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 11 ( $^1\text{H}^{\text{ali}}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 4.7 ( $^1\text{H}^{\text{ali}}$ )
3D HNCACO	NUS (20 %)	64	32	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 96 ( $^{13}\text{C}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 12 ( $^{13}\text{C}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 174.5 ( $^{13}\text{C}$ )
3D HNCO	NUS (20 %)	24	128	1	2048 ( $^1\text{H}^{\text{N}}$ ) x 48 ( $^{15}\text{N}$ ) x 128 ( $^{13}\text{C}$ )	14.3 ( $^1\text{H}^{\text{N}}$ ) x 30 ( $^{15}\text{N}$ ) x 20 ( $^{13}\text{C}$ )	4.7 ( $^1\text{H}^{\text{N}}$ ) x 118 ( $^{15}\text{N}$ ) x 173 ( $^{13}\text{C}$ )

**Table 8.1:** NMR parameters used for backbone and aliphatic side chain assignments

#### 8.4.3 Covalent library screen

A freshly reduced sample of 100  $\mu\text{M}$  U- $^{15}\text{N}$ -labelled reduced *EcDsbA* sample with 1 mM covalent library reaction mixture (based on 100 % reaction conversion) in 50 mM phosphate, 25 mM NaCl, pH 6.8, 10 %  $\text{D}_2\text{O}$ , 2 %  $d_6$ -DMSO, was prepared to a sample volume of 160  $\mu\text{L}$  and acquired in a 3 mm tube.  $^1\text{H}$ - $^{15}\text{N}$  HSQC Acquisition was uniform sampling over 20 mins (NS = 8, DS = 8, D1 = 1), total complex points of 2048 ( $^1\text{H}$ ) and 128 ( $^{15}\text{N}$ ), spectral width of 31 ppm ( $^{15}\text{N}$ ) x 16.6 ( $^1\text{H}$ ) with a transmitter frequency offset of 117.5 ppm ( $^{15}\text{N}$ ) and 4.7 ppm ( $^1\text{H}$ ).

#### 8.4.4 Organic solvent screen

NMR samples were prepared with 100  $\mu\text{M}$  U- $^{15}\text{N}$ -labelled oxidised *EcDsbA* sample and 0, 1, 2, 3, 4 or 5 % (v/v) organic solvent in 50 mM phosphate, 25 mM NaCl, pH 6.8, 10 %  $\text{D}_2\text{O}$ , pH adjusted to 6.8 if required with a total volume of 160  $\mu\text{L}$  and acquired in a 3 mm tube. Organic solvents used were as follows: acetonitrile (MeCN, ACN), tetrahydrofuran (THF), *N,N*-dimethylformamide (DMF), dimethyl sulfoxide (DMSO), methanol (MeOH), ethanol (EtOH), isopropanol (IPA), acetone, benzyl alcohol, ethylene glycol, glycerol, *N,N*-dimethylacetamide, ethyl acetate (EtOAc), acetic acid and urea. A single-point solvent sample for 5% (v/v) acetic acid and 1 M urea were made to the same conditions as above.  $^1\text{H}$ - $^{15}\text{N}$  HSQC acquisition was uniform sampling over 20 mins (NS = 10, DS = 8, D1 = 1), total complex points of 2048 ( $^1\text{H}$ ) and 128 ( $^{15}\text{N}$ ), spectral width of 31 ppm ( $^{15}\text{N}$ ) x 16.6 ( $^1\text{H}$ ) with a transmitter frequency offset of 117.5 ppm ( $^{15}\text{N}$ ) and 4.7 ppm ( $^1\text{H}$ ).

#### 8.4.5 MicroFrag screen

An 80  $\mu\text{M}$  U- $^{15}\text{N}$ -labelled oxidised *EcDsbA* sample with 300 mM MicroFrag in 100 mM phosphate, 50 mM NaCl, pH 6.8, 10 %  $\text{D}_2\text{O}$ , 2 %  $d_6$ -DMSO, was prepared to a sample volume of 160  $\mu\text{L}$  by addition of a 600 mM aqueous stock of MicroFrag pH  $7 \pm 0.5$  to a 200  $\mu\text{M}$  U- $^{15}\text{N}$ -labelled oxidised *EcDsbA* sample in 200 mM phosphate, 100 mM NaCl, pH 6.8, 20 %  $\text{D}_2\text{O}$ , 4 %  $d_6$ -DMSO. Samples were manually adjusted to pH 6.8 if required and acquired in a 3 mm tube. Acquisition was uniform sampling over 20 mins

(NS = 8, DS = 8, D1 = 1), total complex points of 2048 ( $^1\text{H}$ ) and 128 ( $^{15}\text{N}$ ), spectral width of 31 ppm ( $^{15}\text{N}$ ) x 16.6 ( $^1\text{H}$ ) with a transmitter frequency offset of 117.5 ppm ( $^{15}\text{N}$ ) and 4.7 ppm ( $^1\text{H}$ ).

#### 8.4.6 Chemical shift perturbations

Chemical shift perturbations were calculated using the following equation:

$$CSP = \sqrt{\Delta H^2 + (0.2 \times \Delta N)^2}$$

Where  $\Delta H$  and  $\Delta N$  are the observed changes in chemical shift of the backbone amide proton and nitrogens between the unbound and bound  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra.

#### 8.5 Protein mass spectrometry

Protein liquid chromatography mass spectrometry was conducted on a Shimadzu LC-2020 LCMS system with a Zorbax 300SB-C3, 4.6 x 150 mm, 5  $\mu\text{m}$  column. Buffer A: 99.9 %  $\text{H}_2\text{O}$ , 0.1 % acetic acid and buffer B 99.9 % MeCN, 0.1 % acetic acid. The method was as follows: gradient of 20-60 % buffer B in buffer A from 0 – 11 minutes, 60 % buffer B in buffer A from 11 – 12 minutes, 20 % buffer B in buffer A from 12 – 15 minutes all at a flow rate of 0.8 mL/min. All mass spectra were acquired by DUIS in the positive and negative ion mode with a scan range of 100 – 2000  $m/z$  with UV detection at 214, 254 and 280 nm. Data was analysed using LabSolutions v.5.80 software and mMass v 5.5.0 <sup>(292)</sup>.

#### 8.6 Crystallisation and X-ray diffraction experiments

Micro Frag-*EcDsbA* complexes were prepared by crystal soaking. *EcDsbA* was crystallised using hanging drop vapour diffusion. 1  $\mu\text{L}$  of 30 mg/mL protein was mixed with an equal volume of crystallisation buffer (11 – 13 % PEG8000, 5 – 6.5 % glycerol, 1 mM  $\text{CuCl}_2$ , 100 mM sodium cacodylate pH 6.1) and equilibrated against 0.5 mL of reservoir buffer at 20  $^\circ\text{C}$ . Crystals were transferred into 2  $\mu\text{L}$  drops of 13 % PEG8000,

6.5 % glycerol, 1 mM CuCl<sub>2</sub>, 100 mM sodium cacodylate pH 6.1, 9 % MeOH and the Micro Frag of interest at a concentration of 1 M and incubated for 5 minutes. Crystals were mounted on loops and flash-cooled in liquid nitrogen. If the addition of compound caused cracking or dissolution of the crystal the final concentration was adjusted to 0.5 M and the soaking experiment was rerun.

Datasets were collected at the Australian Synchrotron, part of ANSTO, on MX1 and MX2 beamlines made use of the Australian Cancer Research Foundation (ACRF) detector<sup>(293)</sup>. MX1 beamline was equipped with an ADSC Quantum 210r detector and MX2 with an EIGER X 16M pixel detector (Dectris Ltd). Data were processed using the automated data pro-cessing pipeline implemented at the beamline, where data were indexed, integrated and scaled with xdsme<sup>(294)</sup> and AIMLESS<sup>(295)</sup>. The resulting data collection statistics for each dataset were reviewed and if necessary reprocessed with XDS<sup>(296)</sup> or IMOSFLM<sup>(297)</sup> and AIMLESS<sup>(295, 298)</sup> and were scaled using SCALA<sup>(299)</sup>. For all the datasets we chose a resolution cut-off based on the following criteria in the highest resolution range met: CC1/2 was at least 0.6,  $\langle I/\sigma(I) \rangle$  was greater than 1.0 and completeness was greater than 90 %. In cases where data collection statistics did not fulfil our quality requirement or crystals diffracted to lower than 2.5 Å resolution, the crystallographic experiment was repeated.

All datasets were phased by molecular replacement (MR) with Phaser<sup>(300)</sup> using chain A of apo-EcDsbA structure (PDB:1FVK) as a search model. MR and automated refinement steps were carried out using the automated MR pipeline implemented in Auto-Rickshaw<sup>(301, 302)</sup>. Briefly, MR was performed using the MOLREP<sup>(303)</sup> automated refinement protocol. This involved initial refinement in CNS<sup>(304)</sup> followed by refinement in REFMAC5<sup>(305)</sup>. In the CNS rigid body refinement, B-factor (INDIVIDUAL refinement) and positional refinement was undertaken. The resulting model was used in REFMAC5, to perform a maximum likelihood refinement including B-factor refinement.

The final structure was obtained after several rounds of manual refinement using Coot<sup>(306)</sup> and refinement in phenix.refine<sup>(307, 308)</sup>. Ligand restraints were generated in phenix.eLBOW<sup>(309)</sup> and refinement statistics were prepared with “Generate “Table 1” for journal” utility within PHENIX version 1.17.1<sup>(308)</sup>.



### 8.6.1 Refinement Statistics – Table 1 Organic solvent screen.

Crystal structures denoted as % (v/v for liquids, w/v for solids), solvent, and where multiple structures were obtained a structure number.

	80 % Urea (1)	80 % Urea (2)	50 % Glycerol
Wavelength			
Resolution range	37.39 - 1.8 (1.864 - 1.8)	37.54 - 2.5 (2.589 - 2.5)	34.85 - 2.2 (2.279 - 2.2)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	116.009 64.329 75.0501 90 125.315 90	116.341 64.324 75.3464 90 125.318 90	117.147 63.57 75.1367 90 125.492 90
Total reflections	83737 (8268)	31741 (3130)	45611 (4543)
Unique reflections	41881 (4139)	15873 (1565)	22829 (2275)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.98 (100.00)	99.89 (100.00)	99.30 (98.87)
Mean I/sigma(I)	19.56 (2.20)	15.41 (2.57)	8.18 (2.56)
Wilson B-factor	39.46	50.34	47.35
R-merge	0.01526 (0.3046)	0.03076 (0.3038)	0.04471 (0.2603)
R-meas	0.02157 (0.4307)	0.0435 (0.4296)	0.06324 (0.3681)
R-pim	0.01526 (0.3046)	0.03076 (0.3038)	0.04471 (0.2603)
CC1/2	1 (0.883)	0.999 (0.805)	0.995 (0.875)
CC*	1 (0.969)	1 (0.944)	0.999 (0.966)
Reflections used in refinement	41878 (4139)	15862 (1565)	22825 (2275)
Reflections used for R-free	2160 (209)	807 (60)	1152 (124)
R-work	0.2096 (0.3038)	0.2168 (0.3110)	0.2104 (0.3017)

	80 % Urea (1)	80 % Urea (2)	50 % Glycerol
R-free	0.2361 (0.3405)	0.2409 (0.4005)	0.2367 (0.3516)
CC(work)	0.952 (0.576)	0.953 (0.676)	0.948 (0.666)
CC(free)	0.930 (0.639)	0.947 (0.365)	0.914 (0.471)
Number of non-hydrogen atoms	3036	2827	2882
macromolecules	2863	2782	2862
ligands	29	22	20
solvent	144	23	376
Protein residues	374	369	0.026
RMS(bonds)	0.012	0.01	1.45
RMS(angles)	1.3	1.18	96.77
Ramachandran favored (%)	97.57	96.44	2.7
Ramachandran allowed (%)	2.16	3.29	0.54
Ramachandran outliers (%)	0.27	0.27	5.15
Rotamer outliers (%)	2.04	4.96	6.8
Clashscore	4.29	8.13	54.14
Average B-factor	48.1	59.61	53.95
macromolecules	47.88	59.64	80.82
ligands	57.67	68.09	1
solvent	50.52	47.75	

	50 % DMSO (1)	50 % DMSO (2)	50 % MeCN
Wavelength			
Resolution range	37.18 - 1.95 (2.02 - 1.95)	47.86 - 1.95 (2.02 - 1.95)	34.24 - 1.9 (1.968 - 1.9)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	117.246 63.095 74.5471 90 125.253 90	117.214 63.205 74.8476 90 125.254 90	115.653 63.341 74.0794 90 126.067 90
Total reflections	65016 (6416)	65422 (6554)	68495 (6777)
Unique reflections	32513 (3205)	32728 (3277)	34256 (3391)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.92 (99.84)	99.97 (99.97)	99.98 (100.00)
Mean I/sigma(I)	9.51 (1.61)	15.48 (2.20)	15.40 (2.21)
Wilson B-factor	39.71	42.45	35.39
R-merge	0.03108 (0.3731)	0.02087 (0.3418)	0.02154 (0.339)
R-meas	0.04396 (0.5277)	0.02952 (0.4833)	0.03046 (0.4794)
R-pim	0.03108 (0.3731)	0.02087 (0.3418)	0.02154 (0.339)
CC1/2	0.998 (0.817)	0.999 (0.805)	0.999 (0.78)
CC*	1 (0.948)	1 (0.944)	1 (0.936)
Reflections used in refinement	32493 (3205)	32724 (3277)	34254 (3391)
Reflections used for R-free	1728 (164)	1739 (167)	1620 (185)
R-work	0.2110 (0.3022)	0.1991 (0.2852)	0.1954 (0.2871)
R-free	0.2302 (0.3250)	0.2241 (0.3404)	0.2236 (0.2729)
CC(work)	0.935 (0.594)	0.950 (0.695)	0.956 (0.690)

	50 % DMSO (1)	50 % DMSO (2)	50 % MeCN
CC(free)	0.936 (0.502)	0.947 (0.600)	0.946 (0.641)
Number of non-hydrogen atoms	3032	3058	3117
macromolecules	2876	2899	2887
ligands	67	81	18
solvent	89	78	212
Protein residues	376	376	377
RMS(bonds)	0.01	0.014	0.011
RMS(angles)	0.98	1.27	1.03
Ramachandran favored (%)	98.11	98.65	98.92
Ramachandran allowed (%)	1.89	1.35	1.08
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	2.39	2.68	2.01
Clashscore	11.35	6.17	4.25
Average B-factor	43.57	48.96	38.69
macromolecules	43.07	48.06	38.27
ligands	59.75	78.28	46.09
solvent	47.84	52.1	43.82

	50 % MeOH	50 % EtOH (1)	50 % EtOH (2)
Wavelength			
Resolution range	34.36 - 1.9 (1.968 - 1.9)	34.49 - 1.8 (1.864 - 1.8)	34.81 - 1.85 (1.916 - 1.85)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	115.974 63.614 74.275 90 126.03 90	116.356 63.261 74.6006 90 125.834 90	116.552 63.923 75.1037 90 125.588 90
Total reflections	69153 (6822)	79727 (7828)	76783 (7518)
Unique reflections	34588 (3413)	39946 (3918)	38422 (3760)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.97 (99.97)	97.84 (96.65)	99.97 (99.97)
Mean I/sigma(I)	18.86 (3.46)	19.93 (2.14)	17.66 (1.69)
Wilson B-factor	35.11	41.94	50.1
R-merge	0.01957 (0.2181)	0.01248 (0.2943)	0.01572 (0.3931)
R-meas	0.02768 (0.3084)	0.01765 (0.4162)	0.02223 (0.5559)
R-pim	0.01957 (0.2181)	0.01248 (0.2943)	0.01572 (0.3931)
CC1/2	0.999 (0.906)	0.999 (0.883)	0.998 (0.862)
CC*	1 (0.975)	1 (0.968)	0.999 (0.962)
Reflections used in refinement	34582 (3412)	39944 (3918)	38419 (3760)
Reflections used for R-free	1636 (180)	1954 (179)	1797 (166)
R-work	0.1933 (0.2629)	0.2130 (0.3182)	0.2158 (0.3396)
R-free	0.2193 (0.3012)	0.2272 (0.3359)	0.2219 (0.3654)
CC(work)	0.955 (0.705)	0.954 (0.548)	0.941 (0.399)

	50 % MeOH	50 % EtOH (1)	50 % EtOH (2)
CC(free)	0.955 (0.639)	0.943 (0.557)	0.942 (0.280)
Number of non-hydrogen atoms	3121	3002	2937
macromolecules	2874	2862	2853
ligands	18	19	7
solvent	229	121	77
Protein residues	378	377	376
RMS(bonds)	0.011	0.01	0.012
RMS(angles)	1.07	1.16	1.31
Ramachandran favored (%)	98.92	98.39	98.11
Ramachandran allowed (%)	1.08	1.61	1.89
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	2.03	3.06	3.08
Clashscore	2.67	5.73	5.06
Average B-factor	38.24	48.99	57.68
macromolecules	37.72	48.93	57.73
ligands	49.23	57.58	63.3
solvent	43.87	49.11	55.45

## 8.7 Computational calculations

### 8.7.1 Hot spot calculations

FTMap solvent mapping <sup>(270, 271)</sup>, D3Pockets <sup>(310)</sup> and DoGSiteScorer <sup>(272, 311)</sup> were conducted using default settings, using protein chain A for PDB ID: 1FVK with all solvent removed. All calculations conducted on 28/10/2019.

### 8.7.2 Covalent molecular docking experiments

Covalent complexes were docked using Schrödinger Maestro Version 11.0 and 12.5. EcDsbA was prepared using default settings in the protein preparation wizard with a single protein chain (chain B) of the PDB ID: 4TKY structure containing the two structural waters (HOH223, HOH226). The covalently attached peptide was removed and Cys30 was manually ionised to the thiolate and the structure underwent constrained minimisation using default settings. Ligands were prepared using default LigPrep settings and a maximum of 32 states were generated per ligand. The default CovDock settings were applied using Cys30 as the reactive residue, box size of  $\leq 30$  Å, centroid around residues 30, 32, 35, 150 and 164, using a 2.5 kcal/mol energy cutoff for further refinement to give a maximum output of 3 poses per ligand.

Reaction SMARTS used as followed:

Thiols

RECEPTOR_SMARTS_PATTERN	2,[C]-[S;H1,-1]
LIGAND_SMARTS_PATTERN	1,[S;X2;H1]
CUSTOM_CHEMISTRY	("<1>","charge",0,(1)))
CUSTOM_CHEMISTRY	("<2>","charge",0,(1)))
CUSTOM_CHEMISTRY	("<1> <2>","bond",1,(1,2)))

### Isothiocyanates

RECEPTOR_SMARTS_PATTERN	2,[#6]-[#16]
LIGAND_SMARTS_PATTERN	2,[N]=[C]=[S]
CUSTOM_CHEMISTRY	("<2>=[N]",("bond",1,(1,2)))
CUSTOM_CHEMISTRY	("<1> <2>",("bond",1,(1,2)))

### Alkyl

### halides

RECEPTOR_SMARTS_PATTERN	2,[C,c]-[S]
LIGAND_SMARTS_PATTERN	1,[C][F,Cl,Br,I]
CUSTOM_CHEMISTRY	("<1>",("charge",0,(1)))
CUSTOM_CHEMISTRY	("<1> <2>",("bond",1,(1,2)))
CUSTOM_CHEMISTRY	("<2>([F,Cl,Br,I]",("delete",2))

### Carbamates

RECEPTOR_SMARTS_PATTERN	2,[C,c]-[S]
LIGAND_SMARTS_PATTERN	3,[#6]-[#8]-[#6X3H0](=[#8])-[#7X3H1]
CUSTOM_CHEMISTRY	("<1>",("charge",0,(1)))
CUSTOM_CHEMISTRY	("<1> <2>",("bond",1,(1,2)))
CUSTOM_CHEMISTRY	("<2>-[#8]-[#6]",("delete",(2,3)))

## 8.8 MicroFrag library design

All analysis and design were conducted using KNIME Analytic Platform version 4.0.2 with nodes developed by KNIME Analytic Platform, RDKit, CDK toolkits, R, Vernalis and Indigo (EPAM Systems). All theoretical topologies were manually created for compounds containing 5-8 heavy atoms and one 5- or 6-membered ring. Topologies were defined as the theoretical connected scaffold graphs of nodes and connecting edges without bond order or atom type.



The PDB analysis was conducted using the dataset of all ligands accessed in November 2018, where the following filters were applied: 5 – 8 heavy atoms, SlogP < 2, at least 1 heteroatom, at least 1 ring, and a custom list of  $\geq 3$  undesirable alert,  $\geq 1$  reactive functional groups or  $\geq 1$  PAINS. A score of 1 or 0 was given if the fragments were found in the PDB or not.

The DrugBank analysis was conducted using the dataset accessed in November 2018, where the following filters were applied: FDA approved status, oral route of administration, and SlogP between -2 and 6. The remaining 1054 drugs were fragmented into 5250 5 and 8 heavy atom fragments, radicals were converted to hydrogens and aromaticity was resolved using the RDKit normalisation node. Fragments with  $\geq 3$  undesirable alert,  $\geq 1$  reactive functional groups or  $\geq 1$  PAINS were removed as well as any compound that did not contain at least 1 heteroatom. A normalised frequency of fragment occurrence within the oral drug bank set was then calculated with  $sd = 1$ ,  $av = 0$ .

The Micro Frag library was designed from reagents that were commercially available within the MolPort building blocks reagent list accessed in November 2018. The reagent list was filtered for compounds which contained between 5 – 8 heavy atoms, SlogP < 2, one 5- or 6-membered ring, and removed isotopically enriched compounds. All commercially available compounds fitting the criteria were analysed by calculation of physicochemical properties (RDKit Descriptors node), number of 2D 2point pharmacophores (from a list of 95 in house SMARTS substructural 2-point pharmacophore motifs). A scoring function (Micro Frag score) for desired physicochemical properties, presence of Cl, Br or I halogens, complexity and appearance in the PDB dataset or frequency in the DrugBank databases was calculated based on the following formula:

$$(\text{abs}((2.5 - \text{\$NumHeteroAtoms\$})) * -2) + (9 - \text{\$NumHeavyAtoms\$}) * 2 + (\text{abs}(-0.65 - \text{\$SlogP\$})) * -6 + (\text{\$number of 2D 2-point Pharmacophores\$} * -0.5) + (\text{\$whether it was found in the PDB or DrugBank\$} * 1) + (\text{normalised frequency of occurrence in drugs} * 2) + (\text{\$contains a Cl, Br or I\$} * 1) + (\text{\$NumAromaticRings\$} * 2) + (\text{abs}(45 - \text{\$TPSA\$})) * -0.1 + (\text{\$number of alert FG matches\$} * -2)$$

Diversity selection was conducted using an iterative selection process where a balance of normalised Micro Frag score, novel topology coverage, novel 2D 2-point pharmacophore coverage, and sphere exclusion calculated with Tanimoto similarity of ECFP4 and 2D 2-point pharmacophore fingerprints were used to build the library. Diversity selection was calculated based on the following formula:

$$(-1 * \text{number of novel topology or 2D 2-point pharmacophore}) * (\text{weighting of novel topology/pharmacophore}) + (-1 * \text{MicroFrag score}) * (\text{weighting of MicroFrag score}) + (\text{Tanimoto similarity to library (ECFP4)}) * (\text{weighting of Tanimoto similarity}) + (\text{Tanimoto similarity to library (2D 2-point pharmacophores)}) * (\text{weighting of Tanimoto similarity})$$

Weighting of variables:

- Novel topology/pharmacophore = 0.5
- MicroFrag score = 1.2
- Tanimoto similarity to library = 1.2

Physicochemical properties, PMI <sup>(286)</sup>, 2D 2-point pharmacophores, 2D 3-point pharmacophores, SMCM <sup>(282)</sup>, Murcko scaffolds <sup>(285)</sup>, topologies, and Tanimoto similarities of ECFP4, FCFP4, and MACCS fingerprints <sup>(283, 312)</sup> were used to analyse and compare the resultant commercially available, PDB, DrugBank and MicroFrag compound lists.

## 8.9 In-house custom scripts

### 8.9.1 <sup>1</sup>H-<sup>15</sup>N HSQC peak list RMSD alignment

Peak lists were aligned based on Levenberg-Marquardt fitting routine to minimise peak residuals.

### 8.9.2 Grid point clusters for HSQC binding site analysis

Briefly, a set of grid points around the protein were calculated using a 1x1x1 Å grid. Each point was then assigned a value calculated based on the distance of the atom to the grid point and the magnitude of chemical shift perturbation which were summed across all CSPs near that grid point. The grid point values were then smoothed with the 6 adjacent grid points and a threshold was manually set based on inspection of the dataset to filter the grid points. Clusters were then calculated with an in-house agglomerative clustering algorithm with rules dictating thresholds for minimal grid point values and minimal number of adjacent grid points to give the final grid point clusters.

Grid point value =  $((5 / \text{distance in } \text{\AA} \text{ from grid point}) * (\text{weighted CSP} / 0.04))$

### 8.10 Multisequence alignment

Sequences from the following DsbA homologues were used: *E. coli* (UniProt ID: P0AEG4), *V. cholerae* (UniProt ID: P32557), *S. flexneri* (UniProt ID: P52235), *D. dadantii* (UniProt ID: P52234), *H. influenzae* (UniProt ID: P31810), *S. typhimurium* (UniProt ID: P0A2H9), *Az. Vinelandii* (UniProt ID: Q44504), *L. pneumophila* (UniProt ID: P50024), *B. pseudomallei* (UniProt ID: Q63Y08), *P. aeruginosa* (UniProt ID: Q02DMO) and *S. aureus* (UniProt ID: Q9EYL5). The multisequence alignment was generated using Clustal Omega <sup>(313)</sup> and sequence conservation was determined using MView <sup>(248)</sup> using default settings.

# References

1. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science*. 1996;274(5292):1531-4.
2. Hann MM, Leach AR, Harper G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J Chem Inf Comput Sci*. 2001;41(3):856-64.
3. Leeson PD, St-Gallay SA. The influence of the 'organizational factor' on compound quality in drug discovery. *Nat Rev Drug Discov*. 2011;10(10):749-65.
4. Hann MM. Molecular obesity, potency and other addictions in drug discovery. *MedChemComm*. 2011;2(5):349-55.
5. Congreve M, Carr R, Murray C, Jhoti H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today*. 2003;8(19):876-7.
6. Jhoti H, Williams G, Rees DC, Murray CW. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat Rev Drug Discov*. 2013;12(8):644-.
7. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today*. 2004;9(10):430-1.
8. Hall RJ, Mortenson PN, Murray CW. Efficient exploration of chemical space by fragment-based screening. *Prog Biophys Mol Biol*. 2014;116(2):82-91.
9. Keserü GM, Erlanson DA, Ferenczy GG, Hann MM, Murray CW, Pickett SD. Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J Med Chem*. 2016;59(18):8189-206.
10. Davies TG, Wixted WE, Coyle JE, Griffiths-Jones C, Hearn K, McMenamin R, et al. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J Med Chem*. 2016;59(8):3991-4006.
11. Friberg A, Vigil D, Zhao B, Daniels RN, Burke JP, Garcia-Barrantes PM, et al. Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J Med Chem*. 2013;56(1):15-30.
12. Souers AJ, Leverson JD, Boghaert ER, Ackler SL, Catron ND, Chen J, et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat Med*. 2013;19(2):202-8.
13. Chessari G, Buck IM, Day JEH, Day PJ, Iqbal A, Johnson CN, et al. Fragment-Based Drug Discovery Targeting Inhibitor of Apoptosis Proteins: Discovery of a Non-Alanine Lead Series with Dual Activity Against cIAP1 and XIAP. *J Med Chem*. 2015;58(16):6574-88.
14. Demont EH, Chung C-w, Furze RC, Grandi P, Michon A-M, Wellaway C, et al. Fragment-Based Discovery of Low-Micromolar ATAD2 Bromodomain Inhibitors. *J Med Chem*. 2015;58(14):5649-73.

15. Seetoh W-G, Abell C. Disrupting the Constitutive, Homodimeric Protein–Protein Interface in CK2 $\beta$  Using a Biophysical Fragment-Based Approach. *J Am Chem Soc.* 2016;138(43):14303-11.
16. Kessler D, Gmachl M, Mantoulidis A, Martin LJ, Zoephel A, Mayer M, et al. Drugging an undruggable pocket on KRAS. *Proc Natl Acad Sci.* 2019;116(32):15823-9.
17. Maurer T, Garrenton LS, Oh A, Pitts K, Anderson DJ, Skelton NJ, et al. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc Natl Acad Sci.* 2012;109(14):5299-304.
18. Villemagne B, Machelart A, Tran NC, Flipo M, Moune M, Leroux F, et al. Fragment-Based Optimized EthR Inhibitors with in Vivo Ethionamide Boosting Activity. *ACS Infect Dis.* 2020;6(3):366-78.
19. Kirsch P, Jakob V, Oberhausen K, Stein SC, Cucarro I, Schulz TF, et al. Fragment-Based Discovery of a Qualified Hit Targeting the Latency-Associated Nuclear Antigen of the Oncogenic Kaposi's Sarcoma-Associated Herpesvirus/Human Herpesvirus 8. *J Med Chem.* 2019;62(8):3924-39.
20. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature.* 2007;450(7172):1001-9.
21. Price Amanda J, Howard S, Cons Benjamin D. Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem.* 2017;61(5):475-84.
22. Li Q. Application of Fragment-Based Drug Discovery to Versatile Targets. *Front Mol Biosci.* 2020;7(180).
23. Messick TE, Smith GR, Soldan SS, McDonnell ME, Deakyne JS, Malecka KA, et al. Structure-based design of small-molecule inhibitors of EBNA1 DNA binding blocks Epstein-Barr virus latent infection and tumor growth. *Sci Transl Med.* 2019;11(482):eaau5612.
24. Valenti D, Hristeva S, Tzalis D, Ottmann C. Clinical candidates modulating protein-protein interactions: The fragment-based experience. *Eur J Med Chem.* 2019;167:76-95.
25. Perera TPS, Jovcheva E, Mevellec L, Vialard J, De Lange D, Verhulst T, et al. Discovery and Pharmacological Characterization of JNJ-42756493 (Erdafitinib), a Functionally Selective Small-Molecule FGFR Family Inhibitor. *Mol Cancer Ther.* 2017;16(6):1010-20.
26. Tap WD, Wainberg ZA, Anthony SP, Ibrahim PN, Zhang C, Healey JH, et al. Structure-Guided Blockade of CSF1R Kinase in Tenosynovial Giant-Cell Tumor. *N Engl J Med.* 2015;373(5):428-37.
27. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov.* 2012;11(11):873-86.
28. Osborne J, Panova S, Rapti M, Urushima T, Jhoti H. Fragments: where are we now? *Biochem Soc Trans.* 2020;48(1):271-80.

29. Pearlman RS, Smith KM. Novel software tools for chemical diversity. *Perspect Drug Discov Des.* 1998;9(0):339-53.
30. Bohacek RS, McMartin C, Guida WC. The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev.* 1996;16(1):3-50.
31. Polishchuk PG, Madzhidov TI, Varnek A. Estimation of the size of drug-like chemical space based on GDB-17 data. *J Comput Aided Mol Des.* 2013;27(8):675-9.
32. Ruddigkeit L, van Deursen R, Blum LC, Reymond J-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J Chem Inf Model.* 2012;52(11):2864-75.
33. Fink T, Bruggesser H, Reymond J-L. Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons. *Angew Chem Int Ed.* 2005;44(10):1504-8.
34. Blum LC, Reymond J-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J Am Chem Soc.* 2009;131(25):8732-3.
35. Visini R, Awale M, Reymond J-L. Fragment Database FDB-17. *J Chem Inf Model.* 2017;57(4):700-9.
36. Reymond J-L, van Deursen R, Blum LC, Ruddigkeit L. Chemical space as a source for new drugs. *MedChemComm.* 2010;1(1):30-8.
37. Roughley SD, Hubbard RE. How Well Can Fragments Explore Accessed Chemical Space? A Case Study from Heat Shock Protein 90. *J Med Chem.* 2011;54(12):3989-4005.
38. Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD. A Comparison of Physiochemical Property Profiles of Development and Marketed Oral Drugs. *J Med Chem.* 2003;46(7):1250-6.
39. Hughes JD, Blagg J, Price DA, Bailey S, DeCrescenzo GA, Devraj RV, et al. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg Med Chem Lett.* 2008;18(17):4872-5.
40. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov.* 2007;6(11):881-90.
41. Hill AP, Young RJ. Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity. *Drug Discov Today.* 2010;15(15):648-55.
42. Obach RS, Lombardo F, Waters NJ. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. *Drug Metab Dispos.* 2008;36(7):1385-405.
43. Gleeson MP, Hersey A, Montanari D, Overington J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov.* 2011;10(3):197-208.

44. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev.* 1997;23(1):3-25.
45. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov.* 2011;10(3):188-95.
46. Kuntz ID, Chen K, Sharp KA, Kollman PA. The maximal affinity of ligands. *Proc Natl Acad Sci.* 1999;96(18):9997-10002.
47. Hann MM, Keserü GM. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nat Rev Drug Discov.* 2012;11(5):355-65.
48. Teague SJ, Davis AM, Leeson PD, Oprea T. The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed.* 1999;38(24):3743-8.
49. Delaglio F, Walker GS, Farley KA, Sharma R, Hoch JC, Arbogast LW, et al. Non-Uniform Sampling for All: More NMR Spectral Quality, Less Measurement Time. *Am Pharm Rev.* 2017;20(4):339681.
50. Amero C, Schanda P, Durá MA, Ayala I, Marion D, Franzetti B, et al. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *J Am Chem Soc.* 2009;131(10):3448-9.
51. Pearce NM, Krojer T, Bradley AR, Collins P, Nowak RP, Talon R, et al. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat Commun.* 2017;8(1):15123.
52. Owen RL, Juanhuix J, Fuchs M. Current advances in synchrotron radiation instrumentation for MX experiments. *Arch Biochem Biophys.* 2016;602:21-31.
53. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov.* 2016;15(9):605-19.
54. Baell JB, Holloway GA. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J Med Chem.* 2010;53(7):2719-40.
55. Baell JB, Nissink JWM. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. *ACS Chem Biol.* 2018;13(1):36-44.
56. Devine SM, Mulcair MD, Debono CO, Leung EWW, Nissink JWM, Lim SS, et al. Promiscuous 2-Aminothiazoles (PrATs): A Frequent Hitting Scaffold. *J Med Chem.* 2015;58(3):1205-14.
57. Davis BJ, Erlanson DA. Learning from our mistakes: The 'unknown knowns' in fragment screening. *Bioorg Med Chem Lett.* 2013;23(10):2844-52.



58. McGovern SL, Caselli E, Grigorieff N, Shoichet BK. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J Med Chem.* 2002;45(8):1712-22.
59. Over B, Wetzel S, Grütter C, Nakai Y, Renner S, Rauh D, et al. Natural-product-derived fragments for fragment-based ligand discovery. *Nat Chem.* 2013;5(1):21-8.
60. Hung AW, Ramek A, Wang Y, Kaya T, Wilson JA, Clemons PA, et al. Route to three-dimensional fragments using diversity-oriented synthesis. *Proc Natl Acad Sci.* 2011;108(17):6799-804.
61. Morley AD, Pugliese A, Birchall K, Bower J, Brennan P, Brown N, et al. Fragment-based hit identification: thinking in 3D. *Drug Discov Today.* 2013;18(23):1221-7.
62. Kidd SL, Osberger TJ, Mateu N, Sore HF, Spring DR. Recent Applications of Diversity-Oriented Synthesis Toward Novel, 3-Dimensional Fragment Collections. *Front Chem.* 2018;6(460).
63. Prosser KE, Stokes RW, Cohen SM. Evaluation of 3-Dimensionality in Approved and Experimental Drug Space. *ACS Med Chem Lett.* 2020;11(6):1292-8.
64. Yang Y, Engkvist O, Llinàs A, Chen H. Beyond Size, Ionization State, and Lipophilicity: Influence of Molecular Topology on Absorption, Distribution, Metabolism, Excretion, and Toxicity for Druglike Compounds. *J Med Chem.* 2012;55(8):3667-77.
65. Ritchie TJ, Macdonald SJF. The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design? *Drug Discov Today.* 2009;14(21):1011-20.
66. Lovering F, Bikker J, Humblet C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J Med Chem.* 2009;52(21):6752-6.
67. Jahnke W, Erlanson DA, de Esch IJP, Johnson CN, Mortenson PN, Ochi Y, et al. Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J Med Chem.* 2020;63(24):15494-507.
68. Jordan JB, Poppe L, Xia X, Cheng AC, Sun Y, Michelsen K, et al. Fragment Based Drug Discovery: Practical Implementation Based on <sup>19</sup>F NMR Spectroscopy. *J Med Chem.* 2012;55(2):678-87.
69. Vulpetti A, Hommel U, Landrum G, Lewis R, Dalvit C. Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine. *J Am Chem Soc.* 2009;131(36):12949-59.
70. Vulpetti A, Dalvit C. Design and generation of highly diverse fluorinated fragment libraries and their efficient screening with improved (<sup>19</sup>F) NMR methodology. *ChemMedChem.* 2013;8(12):2057-69.
71. Dalvit C. Ligand- and substrate-based <sup>19</sup>F NMR screening: Principles and applications to drug discovery. *Prog Nucl Magn Reson Spectrosc.* 2007;51(4):243-71.

72. Shin Y, Jeong JW, Wurz RP, Achanta P, Arvedson T, Bartberger MD, et al. Discovery of N-(1-Acryloylazetidino-3-yl)-2-(1H-indol-1-yl)acetamides as Covalent Inhibitors of KRASG12C. *ACS Med Chem Lett.* 2019;10(9):1302-8.
73. Resnick E, Bradley A, Gan J, Douangamath A, Krojer T, Sethi R, et al. Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J Am Chem Soc.* 2019;141(22):8951-68.
74. Erlanson DA, Wells JA, Braisted AC. Tethering: Fragment-Based Drug Discovery. *Annu Rev Biophys Biomol Struct.* 2004;33(1):199-223.
75. Grant EK, Fallon DJ, Eberl HC, Fantom KGM, Zappacosta F, Messenger C, et al. A Photoaffinity Displacement Assay and Probes to Study the Cyclin-Dependent Kinase Family. *Angew Chem Int Ed.* 2019;58(48):17322-7.
76. Backus KM, Correia BE, Lum KM, Forli S, Horning BD, González-Páez GE, et al. Proteome-wide covalent ligand discovery in native biological systems. *Nature.* 2016;534(7608):570-4.
77. Parker CG, Galmozzi A, Wang Y, Correia BE, Sasaki K, Joslyn CM, et al. Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell.* 2017;168(3):527-41.e29.
78. Douangamath A, Fearon D, Gehrtz P, Krojer T, Lukacik P, Owen CD, et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat Commun.* 2020;11(1):5047.
79. Chowdhury SR, Kennedy S, Zhu K, Mishra R, Chuong P, Nguyen A-u, et al. Discovery of covalent enzyme inhibitors using virtual docking of covalent fragments. *Bioorg Med Chem Lett.* 2019;29(1):36-9.
80. Keeley A, Petri L, Ábrányi-Balogh P, Keserű GM. Covalent fragment libraries in drug discovery. *Drug Discov Today.* 2020;25(6):983-96.
81. Robson-Tull J. Biophysical screening in fragment-based drug design: a brief overview. *Bioscience Horizons: The International Journal of Student Research.* 2019;11.
82. Coyle J, Walser R. Applied Biophysical Methods in Fragment-Based Drug Discovery. *SLAS Discov.* 2020;25(5):471-90.
83. Ma R, Wang P, Wu J, Ruan K. Process of Fragment-Based Lead Discovery—A Perspective from NMR. *Molecules.* 2016;21(7):854.
84. Gossert AD, Jahnke W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog Nucl Magn Reson Spectrosc.* 2016;97:82-125.
85. Shepherd CA, Hopkins AL, Navratilova I. Fragment screening by SPR and advanced application to GPCRs. *Prog Biophys Mol Biol.* 2014;116(2):113-23.

86. Maveyraud L, Mourey L. Protein X-ray Crystallography and Drug Discovery. *Molecules*. 2020;25(5):1030.
87. Geschwindner S, Ulander J, Johansson P. Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip? *J Med Chem*. 2015;58(16):6321-35.
88. Pedro L, Quinn RJ. Native Mass Spectrometry in Fragment-Based Drug Discovery. *Molecules*. 2016;21(8):984.
89. Chan Daniel S-H, Whitehouse Andrew J, Coyne Anthony G, Abell C. Mass spectrometry for fragment screening. *Essays Biochem*. 2017;61(5):465-73.
90. McMahon RM, Scanlon MJ, Martin JL. Interrogating Fragments Using a Protein Thermal Shift Assay. *Aust J Chem*. 2013;66(12):1502-6.
91. de Souza Neto LR, Moreira-Filho JT, Neves BJ, Maidana RLBR, Guimarães ACR, Furnham N, et al. In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Front Chem*. 2020;8(93).
92. Renaud J-P, Chari A, Ciferri C, Liu W-t, Rémy H-W, Stark H, et al. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov*. 2018;17(7):471-92.
93. Ohlson S, Duong-Thi M-D. Fragment screening for drug leads by weak affinity chromatography (WAC-MS). *Methods*. 2018;146:26-38.
94. Jerabek-Willemsen M, André T, Wanner R, Roth HM, Duhr S, Baaske P, et al. MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct*. 2014;1077:101-13.
95. Scanlon MJ, Norton RS. Fragment-based Drug Discovery, an Accessible Approach to New Therapeutics. *Australian Biochemist*. 2013;44:9-12.
96. Practical Fragments blog [Available from: <http://practicalfragments.blogspot.com/2014/01/poll-results-affiliation-fragment.html>].
97. Mortenson PN, Murray CW. Assessing the lipophilicity of fragments and early hits. *J Comput Aided Mol Des*. 2011;25(7):663-7.
98. Keserü GM, Makara GM. The influence of lead discovery strategies on the properties of drug candidates. *Nat Rev Drug Discov*. 2009;8(3):203-12.
99. Verdonk ML, Rees DC. Group efficiency: a guideline for hits-to-leads chemistry. *ChemMedChem*. 2008;3(8):1179-80.
100. Nissink JWM. Simple Size-Independent Measure of Ligand Efficiency. *J Chem Inf Model*. 2009;49(6):1617-22.
101. Reynolds CH, Tounge BA, Bembenek SD. Ligand Binding Efficiency: Trends, Physical Basis, and Implications. *J Med Chem*. 2008;51(8):2432-8.
102. Andrews PR, Craik DJ, Martin JL. Functional group contributions to drug-receptor interactions. *J Med Chem*. 1984;27(12):1648-57.

103. Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov.* 2014;13(2):105-21.
104. Abad-Zapatero C. Ligand efficiency indices for effective drug discovery. *Expert Opinion on Drug Discovery.* 2007;2(4):469-88.
105. Ferenczy GG, Keseru GM. Enthalpic efficiency of ligand binding. *J Chem Inf Model.* 2010;50(9):1536-41.
106. Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. *Nat Rev Drug Discov.* 2004;3(8):660-72.
107. Congreve M, Chessari G, Tisi D, Woodhead AJ. Recent Developments in Fragment-Based Drug Discovery. *J Med Chem.* 2008;51(13):3661-80.
108. Joseph-McCarthy D, Campbell AJ, Kern G, Moustakas D. Fragment-Based Lead Discovery and Design. *J Chem Inf Model.* 2014;54(3):693-704.
109. Bancet A, Raingeval C, Lomberget T, Le Borgne M, Guichou J-F, Krimm I. Fragment Linking Strategies for Structure-Based Drug Design. *J Med Chem.* 2020;63(20):11420-35.
110. Jencks WP. On the attribution and additivity of binding energies. *Proc Natl Acad Sci.* 1981;78(7):4046-50.
111. Murray CW, Verdonk ML. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J Comput Aided Mol Des.* 2002;16(10):741-53.
112. Ichihara O, Barker J, Law RJ, Whittaker M. Compound Design by Fragment-Linking. *Mol Inform.* 2011;30(4):298-306.
113. Yu HS, Modugula K, Ichihara O, Kramschuster K, Keng S, Abel R, et al. General Theory of Fragment Linking in Molecular Design: Why Fragment Linking Rarely Succeeds and How to Improve Outcomes. *J Chem Theory Comput.* 2021;17(1):450-62.
114. Hubbard R. Fragment approaches in structure-based drug discovery. *Journal of synchrotron radiation.* 2008;15:227-30.
115. Park C-M, Bruncko M, Adickes J, Bauch J, Ding H, Kunzer A, et al. Discovery of an Orally Bioavailable Small Molecule Inhibitor of Prosurvival B-Cell Lymphoma 2 Proteins. *J Med Chem.* 2008;51(21):6902-15.
116. Tse C, Shoemaker AR, Adickes J, Anderson MG, Chen J, Jin S, et al. ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res.* 2008;68(9):3421-8.
117. Wilson WH, O'Connor OA, Czuczman MS, LaCasce AS, Gerecitano JF, Leonard JP, et al. Navitoclax, a targeted high-affinity inhibitor of BCL-2, in lymphoid malignancies: a phase 1 dose-escalation study of safety, pharmacokinetics, pharmacodynamics, and antitumour activity. *Lancet Oncol.* 2010;11(12):1149-59.
118. Roberts AW, Seymour JF, Brown JR, Wierda WG, Kipps TJ, Khaw SL, et al. Substantial Susceptibility of Chronic Lymphocytic Leukemia to BCL2 Inhibition: Results of a

Phase I Study of Navitoclax in Patients With Relapsed or Refractory Disease. *J Clin Oncol*. 2012;30(5):488-96.

119. Tsai J, Lee JT, Wang W, Zhang J, Cho H, Mamo S, et al. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A*. 2008;105(8):3041-6.

120. Jaegle M, Wong EL, Tauber C, Nawrotsky E, Arkona C, Rademann J. Protein-Templated Fragment Ligations-From Molecular Recognition to Drug Discovery. *Angewandte Chemie (International ed in English)*. 2017;56(26):7358-78.

121. Jaegle M, Nawrotsky E, Wong EL, Arkona C, Rademann J. Protein-Templated Fragment Ligation Methods: Emerging Technologies in Fragment-Based Drug Discovery. *Angew Chem Int Ed* 2016:293-326.

122. Becker D, Kaczmarek Z, Arkona C, Schulz R, Tauber C, Wolber G, et al. Irreversible inhibitors of the 3C protease of Coxsackie virus through templated assembly of protein-binding fragments. *Nat Commun*. 2016;7(1):12761.

123. Murray JB, Roughley SD, Matassova N, Brough PA. Off-Rate Screening (ORS) By Surface Plasmon Resonance. An Efficient Method to Kinetically Sample Hit to Lead Chemical Space from Unpurified Reaction Products. *J Med Chem*. 2014;57(7):2845-50.

124. Brough PA, Baker L, Bedford S, Brown K, Chavda S, Chell V, et al. Application of Off-Rate Screening in the Identification of Novel Pan-Isoform Inhibitors of Pyruvate Dehydrogenase Kinase. *J Med Chem*. 2017;60(6):2271-86.

125. Adams L, Wilkinson-White LE, Gunzburg MJ, Headey SJ, Scanlon MJ, Capuano B, et al. Rapid Elaboration of Fragments into Leads Applied to Bromodomain-3 Extra Terminal Domain. *ChemRxiv*. 2020;Preprint.

126. Singh P, Madhaiyan K, Duong-Thi MD, Dymock BW, Ohlson S. Analysis of Protein Target Interactions of Synthetic Mixtures by Affinity-LC/MS. *SLAS Discov*. 2017;22(4):440-6.

127. Gesmundo NJ, Sauvagnat B, Curran PJ, Richards MP, Andrews CL, Dandliker PJ, et al. Nanoscale synthesis and affinity ranking. *Nature*. 2018;557(7704):228-32.

128. Bentley MR, Ilyichova OV, Wang G, Williams ML, Sharma G, Alwan WS, et al. Rapid Elaboration of Fragments into Leads by X-ray Crystallographic Screening of Parallel Chemical Libraries (REFiLX). *J Med Chem*. 2020;63(13):6863-75.

129. Baker LM, Aimon A, Murray JB, Surgenor AE, Matassova N, Roughley SD, et al. Rapid optimisation of fragments and hits to lead compounds from screening of crude reaction mixtures. *Commun Chem*. 2020;3(1):122.

130. Miranker A, Karplus M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins: Structure, Function, and Bioinformatics*. 1991;11(1):29-34.

131. Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, et al. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. *The Journal of Physical Chemistry*. 1996;100(7):2605-11.
132. Liepinsh E, Otting G. Organic solvents identify specific ligand binding sites on protein surfaces. *Nat Biotechnol*. 1997;15(3):264-8.
133. English AC, Done SH, Caves LSD, Groom CR, Hubbard RE. Locating interaction sites on proteins: The crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins: Struct Funct Bioinform*. 1999;37(4):628-40.
134. Ji H, Zhang W, Zhang M, Kudo M, Aoyama Y, Yoshida Y, et al. Structure-Based de Novo Design, Synthesis, and Biological Evaluation of Non-Azole Inhibitors Specific for Lanosterol 14 $\alpha$ -Demethylase of Fungi. *J Med Chem*. 2003;46(4):474-85.
135. Bauman JD, Harrison JJEK, Arnold E. Rapid experimental SAD phasing and hot-spot identification with halogenated fragments. *IUCrJ*. 2016;3(1):51-60.
136. Draxler SW, Bauer M, Eickmeier C, Nadal S, Nar H, Rangel Rojas D, et al. Hybrid Screening Approach for Very Small Fragments: X-ray and Computational Screening on FKBP51. *J Med Chem*. 2020;63(11):5856-64.
137. O'Reilly M, Cleasby A, Davies TG, Hall RJ, Ludlow RF, Murray CW, et al. Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov Today*. 2019.
138. Wood D, Lopez-Fernandez JD, Knight L, Al-Khawaldeh I, Gai C, Lin S, et al. FragLites - minimal, halogenated fragments displaying pharmacophore doublets. An efficient approach to druggability assessment and hit generation. *J Med Chem*. 2019.
139. Erlanson DA, Arndt JW, Cancilla MT, Cao K, Elling RA, English N, et al. Discovery of a potent and highly selective PDK1 inhibitor via fragment-based drug discovery. *Bioorg Med Chem Lett*. 2011;21(10):3078-83.
140. Hacker SM, Backus KM, Lazear MR, Forli S, Correia BE, Cravatt BF. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat Chem*. 2017;9(12):1181-90.
141. Lu W, Kostic M, Zhang T, Che J, Patricelli MP, Jones LH, et al. Fragment-based covalent ligand discovery. *RSC Chem Biol*. 2021.
142. Uetrecht J. Immune-Mediated Adverse Drug Reactions. *Chem Res Toxicol*. 2009;22(1):24-34.
143. Bauer RA. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov Today*. 2015;20(9):1061-73.
144. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. *Nat Rev Drug Discov*. 2011;10:307.

145. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, et al. KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. *PLoS Med.* 2005;2(1):e17.
146. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature.* 2013;503(7477):548-51.
147. Patricelli MP, Janes MR, Li LS, Hansen R, Peters U, Kessler LV, et al. Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discov.* 2016;6(3):316-29.
148. Janes MR, Zhang J, Li LS, Hansen R, Peters U, Guo X, et al. Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell.* 2018;172(3):578-89.e17.
149. Canon J, Rex K, Saiki AY, Mohr C, Cooke K, Bagal D, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature.* 2019;575(7781):217-23.
150. Vane JR, Botting RM. The mechanism of action of aspirin. *Thromb Res.* 2003;110(5-6):255-8.
151. Yocum RR, Rasmussen JR, Strominger JL. The mechanism of action of penicillin. Penicillin acylates the active site of *Bacillus stearothermophilus* D-alanine carboxypeptidase. *J Biol Chem.* 1980;255(9):3977-86.
152. Vita ED. 10 years into the resurgence of covalent drugs. *Future Med Chem.* 2021;13(2):193-210.
153. Deng H, Lei Q, Wu Y, He Y, Li W. Activity-based protein profiling: Recent advances in medicinal chemistry. *Eur J Med Chem.* 2020;191:112151.
154. Zambaldo C, Daguer JP, Saabach J, Barluenga S, Winssinger N. Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *MedChemComm.* 2016;7(7):1340-51.
155. Adeniyi AA, Muthusamy R, Soliman MES. New drug design with covalent modifiers. *Expert Opinion on Drug Discovery.* 2016;11(1):79-90.
156. Gehring M, Laufer SA. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J Med Chem.* 2019;62(12):5673-724.
157. Tuley A, Fast W. The Taxonomy of Covalent Inhibitors. *Biochemistry.* 2018;57(24):3326-37.
158. Copeland RA, Pompliano DL, Meek TD. Drug–target residence time and its implications for lead optimization. *Nat Rev Drug Discov.* 2006;5(9):730-9.
159. Swinney DC. Biochemical mechanisms of drug action: what does it take for success? *Nat Rev Drug Discov.* 2004;3(9):801-8.
160. Claxton AJ, Cramer J, Pierce C. A systematic review of the associations between dose regimens and medication compliance. *Clin Ther.* 2001;23(8):1296-310.

161. Johnson DW, E; Cravatt, B. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem.* 2010;2(6):949-64.
162. Shuman CF, Markgren P-O, Hämäläinen M, Danielson UH. Elucidation of HIV-1 protease resistance by characterization of interaction kinetics between inhibitors and enzyme variants. *Antiviral Res.* 2003;58(3):235-42.
163. Gauthier JY, Chauret N, Cromlish W, Desmarais S, Duong LT, Falgout JP, et al. The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorg Med Chem Lett.* 2008;18(3):923-8.
164. Cleasby A, Yon J, Day PJ, Richardson C, Tickle IJ, Williams PA, et al. Structure of the BTB Domain of Keap1 and Its Interaction with the Triterpenoid Antagonist CDDO. *PLoS One.* 2014;9(6):e98896.
165. Copeland RA. The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety. *Expert Opin Drug Discov.* 2010;5(4):305-10.
166. Ohlson S. Designing transient binding drugs: a new concept for drug discovery. *Drug Discov Today.* 2008;13(9-10):433-9.
167. Lipton SA. Paradigm shift in neuroprotection by NMDA receptor blockade: Memantine and beyond. *Nat Rev Drug Discov.* 2006;5(2):160-70.
168. Leonetti A, Sharma S, Minari R, Perego P, Giovannetti E, Tiseo M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br J Cancer.* 2019;121(9):725-37.
169. Strelow JM. A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *SLAS DISCOVERY: Advancing Life Sciences R&D.* 2016;22(1):3-20.
170. Johnson KA, Goody RS. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry.* 2011;50(39):8264-9.
171. Michaelis L, Menten ML. Die kinetik der invertinwirkung. *Biochem z.* 1913;49(333-369):352.
172. Jörg M, Scammells PJ. Guidelines for the Synthesis of Small-Molecule Irreversible Probes Targeting G Protein-Coupled Receptors. *ChemMedChem.* 2016;11(14):1488-98.
173. Ray S, Murkin AS. New Electrophiles and Strategies for Mechanism-Based and Targeted Covalent Inhibitor Design. *Biochemistry.* 2019;58(52):5234-44.
174. Powers JC, Asgian JL, Ekici ÖD, James KE. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem Rev.* 2002;102(12):4639-750.
175. Potashman MH, Duggan ME. Covalent Modifiers: An Orthogonal Approach to Drug Design. *J Med Chem.* 2009;52(5):1231-46.
176. Savi P, Zachary J-L, Delesque-Touchard N, Labouret C, Hervé C, Uzabiaga M-F, et al. The active metabolite of Clopidogrel disrupts P2Y12 receptor oligomers and partitions them out of lipid rafts. *Proc Natl Acad Sci.* 2006;103(29):11069-74.



177. AE Takemori, Portoghese P. Affinity Labels for Opioid Receptors. *Annu Rev Pharmacol Toxicol.* 1985;25(1):193-223.
178. Fedan JS, Hogaboom GK, O'Donnell JP. Photoaffinity labels as pharmacological tools. *Biochem Pharmacol.* 1984;33(8):1167-80.
179. Davie BJ, Sexton PM, Capuano B, Christopoulos A, Scammells PJ. Development of a Photoactivatable Allosteric Ligand for the M1 Muscarinic Acetylcholine Receptor. *ACS Chem Neurosci.* 2014;5(10):902-7.
180. Langmead CJ, Watson J, Reavill C. Muscarinic acetylcholine receptors as CNS drug targets. *Pharmacol Ther.* 2008;117(2):232-43.
181. Ma L, Seager MA, Wittmann M, Jacobson M, Bickel D, Burno M, et al. Selective activation of the M1 muscarinic acetylcholine receptor achieved by allosteric potentiation. *Proc Natl Acad Sci U S A.* 2009;106(37):15950-5.
182. Davie BJ, Valant C, White JM, Sexton PM, Capuano B, Christopoulos A, et al. Synthesis and pharmacological evaluation of analogues of benzyl quinolone carboxylic acid (BQCA) designed to bind irreversibly to an allosteric site of the M<sub>1</sub> muscarinic acetylcholine receptor. *J Med Chem.* 2014;57(12):5405-18.
183. Luepke KH, Suda KJ, Boucher H, Russo RL, Bonney MW, Hunt TD, et al. Past, Present, and Future of Antibacterial Economics: Increasing Bacterial Resistance, Limited Antibiotic Pipeline, and Societal Implications. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy.* 2017;37(1):71-84.
184. Spellberg B. The future of antibiotics. *Critical Care.* 2014;18(3):228.
185. Theuretzbacher U, Gottwalt S, Beyer P, Butler M, Czaplewski L, Lienhardt C, et al. Analysis of the clinical antibacterial and antituberculosis pipeline. *Lancet Infect Dis.* 2019;19(2):e40-e50.
186. World Health Organisation. 2019 Antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline. 2019.
187. Brown DG, Wobst HJ. A Decade of FDA-Approved Drugs (2010–2019): Trends and Future Directions. *J Med Chem.* 2021.
188. World Health Organisation. No time to wait: securing the future from drug-resistant infections. 2019.
189. Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. 2014.
190. Chopra I, Hesse L, O'Neill AJ. Exploiting current understanding of antibiotic action for discovery of new drugs. *J Appl Microbiol.* 2002;92(s1):4S-15S.
191. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis.* 2018;18(3):318-27.

192. Zhao S, Adamiak JW, Bonifay V, Mehla J, Zgurskaya HI, Tan DS. Defining new chemical space for drug penetration into Gram-negative bacteria. *Nat Chem Biol.* 2020;16(12):1293-302.
193. Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. *Microbiology spectrum.* 2016;4(2):10.1128/microbiolspec.VMBF-0016-2015.
194. Heras B, Scanlon MJ, Martin JL. Targeting virulence not viability in the search for future antibacterials. *Br J Clin Pharmacol.* 2015;79(2):208-15.
195. Rasko DA, Sperandio V. Anti-virulence strategies to combat bacteria-mediated disease. *Nat Rev Drug Discov.* 2010;9(2):117-28.
196. Allen RC, Popat R, Diggle SP, Brown SP. Targeting virulence: can we make evolution-proof drugs? *Nat Rev Microbiol.* 2014;12(4):300-8.
197. Cegelski L, Marshall GR, Eldridge GR, Hultgren SJ. The biology and future prospects of antivirulence therapies. *Nat Rev Microbiol.* 2008;6(1):17-27.
198. Zambelloni R, Marquez R, Roe AJ. Development of Antivirulence Compounds: A Biochemical Review. *Chem Biol Drug Des.* 2015;85(1):43-55.
199. Dickey SW, Cheung GYC, Otto M. Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance. *Nat Rev Drug Discov.* 2017;16(7):457-71.
200. Theuretzbacher U, Outterson K, Engel A, Karlén A. The global preclinical antibacterial pipeline. *Nat Rev Microbiol.* 2020;18(5):275-85.
201. Dutton RJ, Boyd D, Berkmen M, Beckwith J. Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. *Proc Natl Acad Sci.* 2008;105(33):11933-8.
202. Hiniker A, Bardwell JCA. *In Vivo* Substrate Specificity of Periplasmic Disulfide Oxidoreductases \*. *J Biol Chem.* 2004;279(13):12967-73.
203. Macnab RM. Proton-driven bacterial flagellar motor. *Methods Enzymol.* 1986;125:563-81.
204. Jacob-Dubuisson F, Pinkner J, Xu Z, Striker R, Padmanhaban A, Hultgren SJ. PapD chaperone function in pilus biogenesis depends on oxidant and chaperone-like activities of DsbA. *Proc Natl Acad Sci U S A.* 1994;91(24):11552-6.
205. Wülfing C, Rappuoli R. Efficient production of heat-labile enterotoxin mutant proteins by overexpression of dsbA in a degP-deficient *Escherichia coli* strain. *Arch Microbiol.* 1997;167(5):280-3.
206. Miki T, Okada N, Kim Y, Abe A, Danbara H. DsbA directs efficient expression of outer membrane secretin EscC of the enteropathogenic *Escherichia coli* type III secretion apparatus. *Microb Pathog.* 2008;44(2):151-8.
207. Bardwell JC, McGovern K, Beckwith J. Identification of a protein required for disulfide bond formation in vivo. *Cell.* 1991;67(3):581-9.

208. Coulthurst SJ, Lilley KS, Hedley PE, Liu H, Toth IK, Salmond GPC. DsbA plays a critical and multifaceted role in the production of secreted virulence factors by the phytopathogen *Erwinia carotovora* subsp. *atroseptica*. *J Biol Chem*. 2008;283(35):23739-53.
209. Bringer M-A, Rolhion N, Glasser A-L, Darfeuille-Michaud A. The Oxidoreductase DsbA Plays a Key Role in the Ability of the Crohn's Disease-Associated Adherent-Invasive *Escherichia coli* Strain LF82 To Resist Macrophage Killing. *J Bacteriol*. 2007;189(13):4860-71.
210. Burall LS, Harro JM, Li X, Lockett CV, Himpel SD, Hebel JR, et al. *Proteus mirabilis* Genes That Contribute to Pathogenesis of Urinary Tract Infection: Identification of 25 Signature-Tagged Mutants Attenuated at Least 100-Fold. *Infect Immun*. 2004;72(5):2922-38.
211. Lee SH, Butler SM, Camilli A. Selection for in vivo regulators of bacterial virulence. *Proc Natl Acad Sci U S A*. 2001;98(12):6889-94.
212. Lin D, Rao CV, Slauch JM. The *Salmonella* SPI1 type three secretion system responds to periplasmic disulfide bond status via the flagellar apparatus and the RcsCDB system. *J Bacteriol*. 2008;190(1):87-97.
213. Yu J. Inactivation of DsbA, but not DsbC and DsbD, affects the intracellular survival and virulence of *Shigella flexneri*. *Infect Immun*. 1998;66(8):3909-17.
214. Totsika M, Heras B, Werpel DJ, Schembri MA. Characterization of two homologous disulfide bond systems involved in virulence factor biogenesis in uropathogenic *Escherichia coli* CFT073. *J Bacteriol*. 2009;191(12):3901-8.
215. Rosadini CV, Wong SM, Akerley BJ. The periplasmic disulfide oxidoreductase DsbA contributes to *Haemophilus influenzae* pathogenesis. *Infect Immun*. 2008;76(4):1498-508.
216. Miki T, Okada N, Danbara H. Two periplasmic disulfide oxidoreductases, DsbA and SrgA, target outer membrane protein SpiA, a component of the *Salmonella* pathogenicity island 2 type III secretion system. *J Biol Chem*. 2004;279(33):34631-42.
217. Ireland PM, McMahon RM, Marshall LE, Halili M, Furlong E, Tay S, et al. Disarming *Burkholderia pseudomallei*: structural and functional characterization of a disulfide oxidoreductase (DsbA) required for virulence in vivo. *Antioxid Redox Signal*. 2014;20(4):606-17.
218. Bader M, Muse W, Ballou DP, Gassner C, Bardwell JCA. Oxidative Protein Folding Is Driven by the Electron Transport System. *Cell*. 1999;98(2):217-27.
219. Vlamis-Gardikas A. The multiple functions of the thiol-based electron flow pathways of *Escherichia coli*: Eternal concepts revisited. *Biochim Biophys Acta*. 2008;1780(11):1170-200.
220. Fernandes PA, Ramos MJ. Theoretical insights into the mechanism for thiol/disulfide exchange. *Chemistry (Easton)*. 2004;10(1):257-66.
221. Kadokura H, Beckwith J. Four cysteines of the membrane protein DsbB act in concert to oxidize its substrate DsbA. *EMBO J*. 2002;21(10):2354-63.

222. Inaba K, Murakami S, Suzuki M, Nakagawa A, Yamashita E, Okada K, et al. Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. *Cell*. 2006;127(4):789-801.
223. Kishigami S, Kanaya E, Kikuchi M, Ito K. DsbA-DsbB Interaction through Their Active Site Cysteines: EVIDENCE FROM AN ODD CYSTEINE MUTANT OF DsbA (&#x2217;). *J Biol Chem*. 1995;270(29):17072-4.
224. Inaba K, Takahashi Y-h, Ito K. Reactivities of Quinone-free DsbB from *Escherichia coli*\*. *J Biol Chem*. 2005;280(38):33035-44.
225. Shouldice SR, Heras B, Walden PM, Totsika M, Schembri MA, Martin JL. Structure and function of DsbA, a key bacterial oxidative folding catalyst. *Antioxid Redox Signal*. 2011;14(9):1729-60.
226. Heras B, Shouldice SR, Totsika M, Scanlon MJ, Schembri MA, Martin JL. DSB proteins and bacterial pathogenicity. *Nat Rev Microbiol*. 2009;7(3):215-25.
227. McMahon RM, Premkumar L, Martin JL. Four structural subclasses of the antivirulence drug target disulfide oxidoreductase DsbA provide a platform for design of subclass-specific inhibitors. *Biochim Biophys Acta*. 2014;1844(8):1391-401.
228. Haebel PW, Goldstone D, Katzen F, Beckwith J, Metcalf P. The disulfide bond isomerase DsbC is activated by an immunoglobulin-fold thiol oxidoreductase: crystal structure of the DsbC-DsbD $\alpha$  complex. *The EMBO journal*. 2002;21(18):4774-84.
229. Paxman JJ, Borg NA, Horne J, Thompson PE, Chin Y, Sharma P, et al. The structure of the bacterial oxidoreductase enzyme DsbA in complex with a peptide reveals a basis for substrate specificity in the catalytic cycle of DsbA enzymes. *J Biol Chem*. 2009;284(26):17835-45.
230. Martin JL, Bardwell JC, Kuriyan J. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature*. 1993;365(6445):464-8.
231. Vivian JP, Scoullar J, Robertson AL, Bottomley SP, Horne J, Chin Y, et al. Structural and Biochemical Characterization of the Oxidoreductase NmDsbA3 from *Neisseria meningitidis*\*. *J Biol Chem*. 2008;283(47):32452-61.
232. Kurth F, Rimmer K, Premkumar L, Mohanty B, Duprez W, Halili MA, et al. Comparative Sequence, Structure and Redox Analyses of *Klebsiella pneumoniae* DsbA Show That Anti-Virulence Target DsbA Enzymes Fall into Distinct Classes. *PLoS One*. 2013;8(11):e80210.
233. Alwan WS. Discovery & Design of Novel DsbA Inhibitors Using Fragment-Based Design Strategies: Monash University; 2019.
234. Huber-Wunderlich M, Glockshuber R. A single dipeptide sequence modulates the redox properties of a whole enzyme family. *Fold Des*. 1998;3(3):161-71.
235. Grauschopf U, Winther JR, Korber P, Zander T, Dallinger P, Bardwell JC. Why is DsbA such an oxidizing disulfide catalyst? *Cell*. 1995;83(6):947-55.

236. Guddat LW, Bardwell JC, Glockshuber R, Huber-Wunderlich M, Zander T, Martin JL. Structural analysis of three His32 mutants of DsbA: support for an electrostatic role of His32 in DsbA stability. *Protein Sci.* 1997;6(9):1893-900.
237. Zapun A, Bardwell JCA, Creighton TE. The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. *Biochemistry.* 1993;32(19):5083-92.
238. Kadokura H, Tian H, Zander T, Bardwell JCA, Beckwith J. Snapshots of DsbA in Action: Detection of Proteins in the Process of Oxidative Folding. *Science.* 2004;303(5657):534-7.
239. Charbonnier JB, Belin P, Moutiez M, Stura EA, Quéméneur E. On the role of the cis-proline residue in the active site of DsbA. *Protein science : a publication of the Protein Society.* 1999;8(1):96-105.
240. Inaba K, Ito K. Structure and mechanisms of the DsbB–DsbA disulfide bond generation machine. *Biochim Biophys Acta.* 2008;1783(4):520-9.
241. Inaba K, Murakami S, Nakagawa A, Iida H, Kinjo M, Ito K, et al. Dynamic nature of disulphide bond formation catalysts revealed by crystal structures of DsbB. *The EMBO journal.* 2009;28(6):779-91.
242. Duprez W, Premkumar L, Halili MA, Lindahl F, Reid RC, Fairlie DP, et al. Peptide Inhibitors of the Escherichia coli DsbA Oxidative Machinery Essential for Bacterial Virulence. *J Med Chem.* 2015;58(2):577-87.
243. Guddat LW, Martin JL, Bardwell JCA, Zander T. The uncharged surface features surrounding the active site of Escherichia coli DsbA are conserved and are implicated in peptide binding. *Protein Sci.* 1997;6(6):1148-56.
244. Couprie J, Vinci F, Dugave C, Quéméneur E, Moutiez M. Investigation of the DsbA Mechanism through the Synthesis and Analysis of an Irreversible Enzyme–Ligand Complex. *Biochemistry.* 2000;39(22):6732-42.
245. Frech C, Wunderlich M, Glockshuber R, Schmid FX. Preferential binding of an unfolded protein to DsbA. *The EMBO journal.* 1996;15(2):392-8.
246. Darby NJ, Creighton TE. Catalytic Mechanism of DsbA and Its Comparison with That of Protein Disulfide Isomerase. *Biochemistry.* 1995;34(11):3576-87.
247. Jacobi A, Huber-Wunderlich M, Hennecke J, Glockshuber R. Elimination of all charged residues in the vicinity of the active-site helix of the disulfide oxidoreductase DsbA. Influence of electrostatic interactions on stability and redox properties. *J Biol Chem.* 1997;272(35):21692-9.
248. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics.* 1998;14(4):380-1.

249. Horne J, d'Auvergne EJ, Coles M, Velkov T, Chin Y, Charman WN, et al. Probing the Flexibility of the DsbA Oxidoreductase from *Vibrio cholerae*—a  $^{15}\text{N}$  -  $^1\text{H}$  Heteronuclear NMR Relaxation Analysis of Oxidized and Reduced Forms of DsbA. *J Mol Biol.* 2007;371(3):703-16.
250. Nebel S, Alwan WS, Williams ML, Sharma G, Taylor A, Doak BC, et al. NMR fragment screening reveals a novel small molecule binding site near the catalytic surface of the disulfide–dithiol oxidoreductase enzyme DsbA from *Burkholderia pseudomallei*. *J Biomol NMR.* 2020;74(10):595-611.
251. Halili MA, Bachu P, Lindahl F, Bechara C, Mohanty B, Reid RC, et al. Small Molecule Inhibitors of Disulfide Bond Formation by the Bacterial DsbA–DsbB Dual Enzyme System. *ACS Chem Biol.* 2015;10(4):957-64.
252. Adams LA, Sharma P, Mohanty B, Ilyichova OV, Mulcair MD, Williams ML, et al. Application of Fragment-Based Screening to the Design of Inhibitors of *Escherichia coli* DsbA. *Angew Chem Int Ed.* 2015;54(7):2179-84.
253. Bentley MR. Development of the diaryl ether scaffold as inhibitors of *EcDsbA* using a fragment-based approach [Thesis]: Monash University; 2019.
254. Doak BC. Fragment Based Discovery of DsbA Inhibitors: Monash University; 2012.
255. Doak BC, Whitehouse RL, Rimmer K, Williams M, Heras B, Caria S, et al. Fluoromethylketone-fragment conjugates designed as covalent modifiers of *EcDsbA* are atypical substrates. Manuscript in preparation. 2021.
256. Duncan LF, Wang G, Ilyichova OV, Scanlon MJ, Heras B, Abbott BM. The Fragment-Based Development of a Benzofuran Hit as a New Class of *Escherichia coli* DsbA Inhibitors. *Molecules.* 2019;24(20):3756.
257. Wunderlich M, Glockshuber R. Redox properties of protein disulfide isomerase (DsbA) from *Escherichia coli*. *Protein science : a publication of the Protein Society.* 1993;2(5):717-26.
258. Warwicker J, Gane PJ. Calculation of Cys 30  $\Delta\text{pK}_a$ 's and oxidising power for DsbA mutants. *FEBS Lett.* 1996;385(1-2):105-8.
259. LaPlante SR, Carson R, Gillard J, Aubry N, Coulombe R, Bordeleau S, et al. Compound Aggregation in Drug Discovery: Implementing a Practical NMR Assay for Medicinal Chemists. *J Med Chem.* 2013;56(12):5142-50.
260. Wishart DS, Sykes BD. The  $^{13}\text{C}$  Chemical-Shift Index: A simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J Biomol NMR.* 1994;4(2):171-80.
261. Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry.* 1992;31(6):1647-51.

262. Nelson JW, Creighton TE. Reactivity and Ionization of the Active Site Cysteine Residues of DsbA, a Protein Required for Disulfide Bond Formation in vivo. *Biochemistry*. 1994;33(19):5974-83.
263. Duprez W, Bachu P, Stoermer MJ, Tay S, McMahon RM, Fairlie DP, et al. Virtual Screening of Peptide and Peptidomimetic Fragments Targeted to Inhibit Bacterial Dithiol Oxidase DsbA. *PLoS One*. 2015;10(7):e0133805.
264. Zhu K, Borrelli KW, Greenwood JR, Day T, Abel R, Farid RS, et al. Docking Covalent Inhibitors: A Parameter Free Approach To Pose Prediction and Scoring. *J Chem Inf Model*. 2014;54(7):1932-40.
265. Sun M, Hong CY, Pan CY. A unique aliphatic tertiary amine chromophore: fluorescence, polymer structure, and application in cell imaging. *J Am Chem Soc*. 2012;134(51):20581-4.
266. Sharma D, Rajarathnam K. <sup>13</sup>C NMR chemical shifts can predict disulfide bond formation. *J Biomol NMR*. 2000;18(2):165-71.
267. Suzuki T, Kouketsu A, Matsuura A, Kohara A, Ninomiya S-i, Kohda K, et al. Thiol-based SAHA analogues as potent histone deacetylase inhibitors. *Bioorg Med Chem Lett*. 2004;14(12):3313-7.
268. Kim Y, Ho SO, Gassman NR, Korlann Y, Landorf EV, Collart FR, et al. Efficient site-specific labeling of proteins via cysteines. *Bioconj Chem*. 2008;19(3):786-91.
269. Ravasco JMM, Faustino H, Trindade A, Gois PMP. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chemistry – A European Journal*. 2019;25(1):43-59.
270. Brenke R, Kozakov D, Chuang G-Y, Beglov D, Hall D, Landon MR, et al. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics (Oxford, England)*. 2009;25(5):621-7.
271. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc*. 2015;10(5):733-55.
272. Volkamer A, Griewel A, Grombacher T, Rarey M. Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. *J Chem Inf Model*. 2010;50(11):2041-52.
273. Dechene M, Wink G, Smith M, Swartz P, Mattos C. Multiple solvent crystal structures of ribonuclease A: an assessment of the method. *Proteins*. 2009;76(4):861-81.
274. Fejzo J, Lepre CA, Peng JW, Bemis GW, Ajay, Murcko MA, et al. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem Biol*. 1999;6(10):755-69.
275. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res*. 2017;45(D1):D271-D81.

276. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-42.
277. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* 2020;49(D1):D437-D51.
278. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074-d82.
279. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006;34(Database issue):D668-72.
280. Benigni R, Bossa C. Mechanisms of Chemical Carcinogenicity and Mutagenicity: A Review with Implications for Predictive Toxicology. *Chem Rev.* 2011;111(4):2507-36.
281. Sushko I, Salmina E, Potemkin VA, Poda G, Tetko IV. ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions. *J Chem Inf Model.* 2012;52(8):2310-6.
282. Allu TK, Oprea TI. Rapid Evaluation of Synthetic and Molecular Complexity for in Silico Chemistry. *J Chem Inf Model.* 2005;45(5):1237-43.
283. Rogers D, Hahn M. Extended-Connectivity Fingerprints. *J Chem Inf Model.* 2010;50(5):742-54.
284. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL Keys for Use in Drug Discovery. *J Chem Inf Comput Sci.* 2002;42(6):1273-80.
285. Bemis GW, Murcko MA. The Properties of Known Drugs. 1. Molecular Frameworks. *J Med Chem.* 1996;39(15):2887-93.
286. Sauer WHB, Schwarz MK. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J Chem Inf Comput Sci.* 2003;43(3):987-1003.
287. Newman J, Sayle RA, Fazio VJ. A universal indicator dye pH assay for crystallization solutions and other high-throughput applications. *Acta Crystallographica Section D, Biological Crystallography.* 2012(D68):1003-9.
288. Movassaghi M, Schmidt MA. N-Heterocyclic Carbene-Catalyzed Amidation of Unactivated Esters with Amino Alcohols. *Org Lett.* 2005;7(12):2453-6.
289. Kim E, Rye PT, Essigmann JM, Croy RG. A bifunctional platinum(II) antitumor agent that forms DNA adducts with affinity for the estrogen receptor. *J Inorg Biochem.* 2009;103(2):256-61.



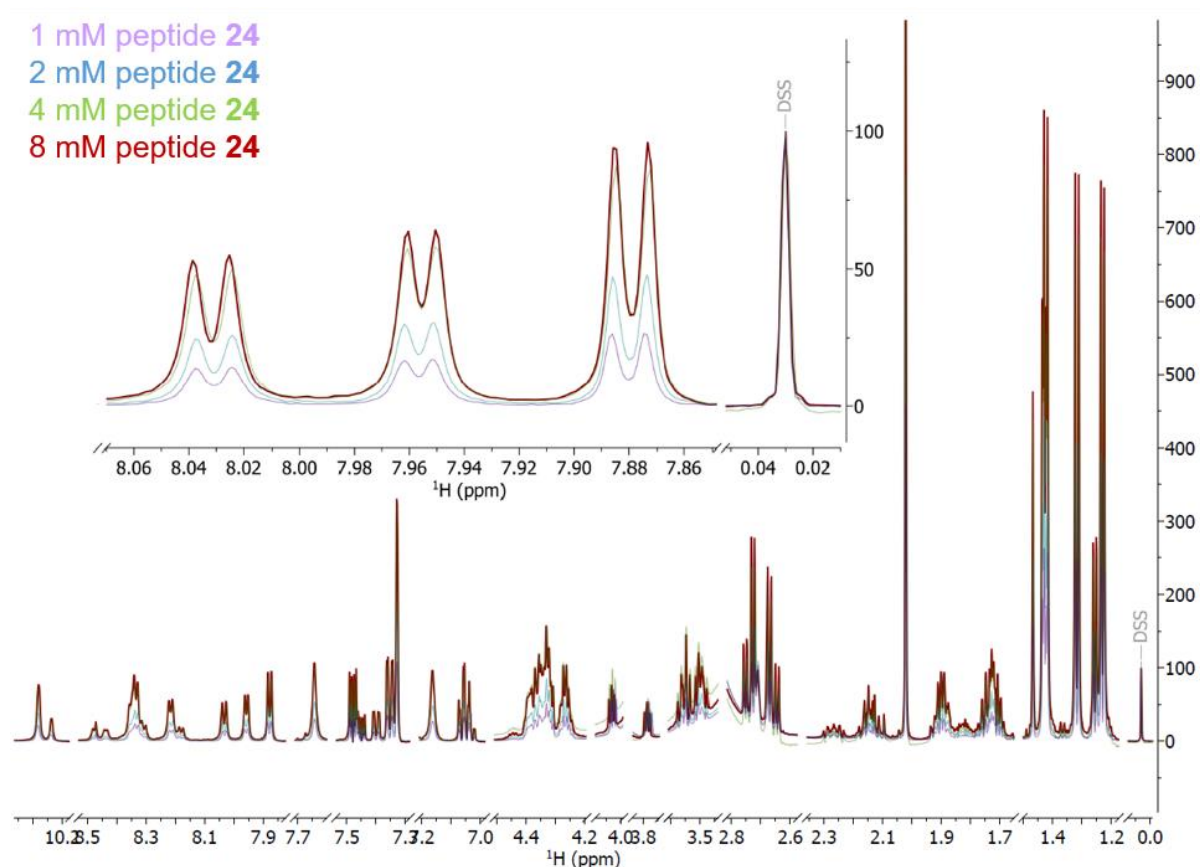
290. Hart ME, Suchland KL, Miyakawa M, Bunzow JR, Grandy DK, Scanlan TS. Trace amine-associated receptor agonists: synthesis and evaluation of thyronamines and related analogues. *J Med Chem*. 2006;49(3):1101-12.
291. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005;41(1):207-34.
292. Strohm M, Kavan D, Novák P, Volný M, Havlíček V. mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal Chem*. 2010;82(11):4648-51.
293. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, et al. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synchrotron Radiat*. 2002;9(Pt 6):401-6.
294. Legrand P. XDSME: XDS Made Easier GitHub repository. 2017.
295. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*. 2013;69(Pt 7):1204-14.
296. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):125-32.
297. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):271-81.
298. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994;50(Pt 5):760-3.
299. Evans P. Scaling and assesment of data quality. *Acta Crystallogr Sect D Biol Crystallogr*. 2006;62(1):72-82.
300. A.J. McCoy RWG-K, P.D. Adams, M.D. Winn, L.C. Storoni, and R.J. Read. Phaser crystallographic software. *J Appl Crystallogr*. 2007(40):659-74.
301. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr*. 2005;61(Pt 4):449-57.
302. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr*. 2009;65(Pt 10):1089-97.
303. Vagin A, Teplyakov A. MOLREP: an Automated Program for Molecular Replacement. *J Appl Crystallogr*. 1997;30(6):1022-5.
304. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*. 1998;54(Pt 5):905-21.
305. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53(Pt 3):240-55.

306. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 12 Pt 1):2126-32.
307. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):213-21.
308. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*. 2019;75(Pt 10):861-77.
309. Moriarty NW, Grosse-Kunstleve RW, Adams PD. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr Sect D Struct Biol*. 2009;65(10):1074-80.
310. Chen Z, Zhang X, Peng C, Wang J, Xu Z, Chen K, et al. D3Pockets: A Method and Web Server for Systematic Analysis of Protein Pocket Dynamics. *J Chem Inf Model*. 2019;59(8):3353-8.
311. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining Global and Local Measures for Structure-Based Druggability Predictions. *J Chem Inf Model*. 2012;52(2):360-72.
312. Morgan HL. The Generation of a Unique Machine Description for Chemical Structures- A Technique Developed at Chemical Abstracts Service. *J Chem Doc*. 1965;5(2):107-13.
313. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.

# Appendices

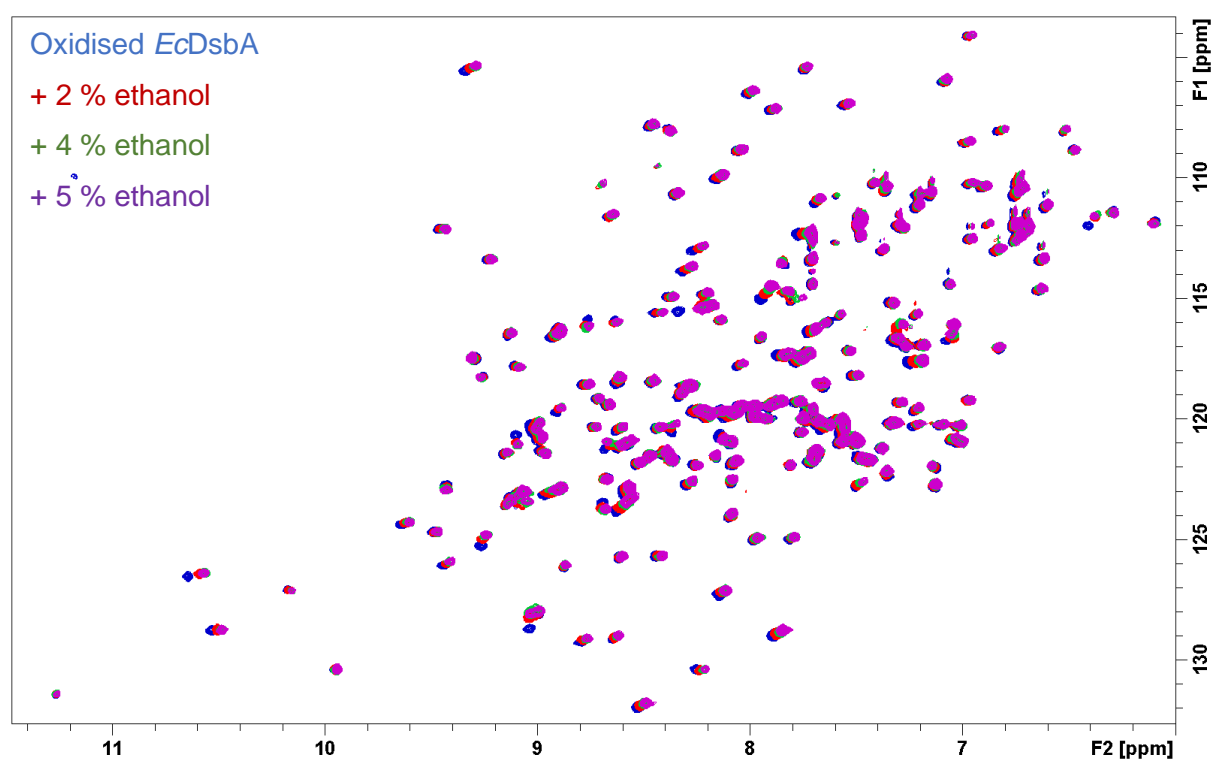
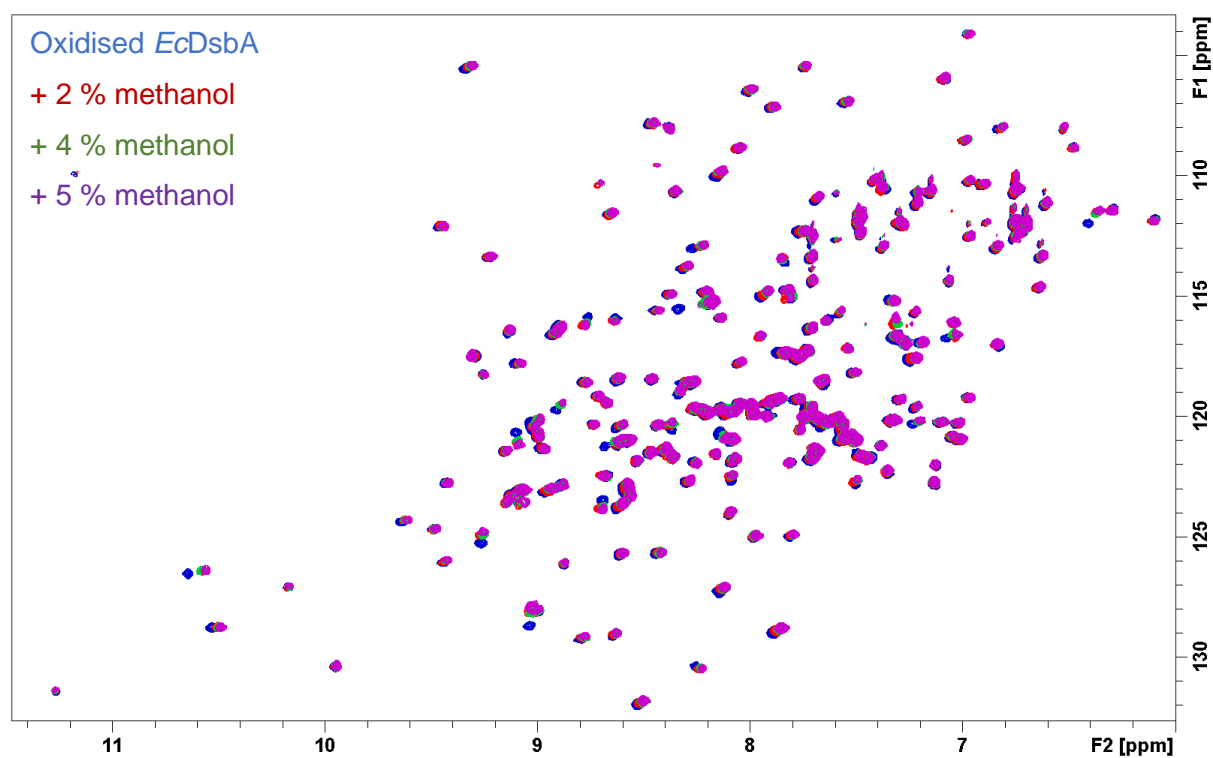
## Appendix 1: Non-covalent peptide solubility

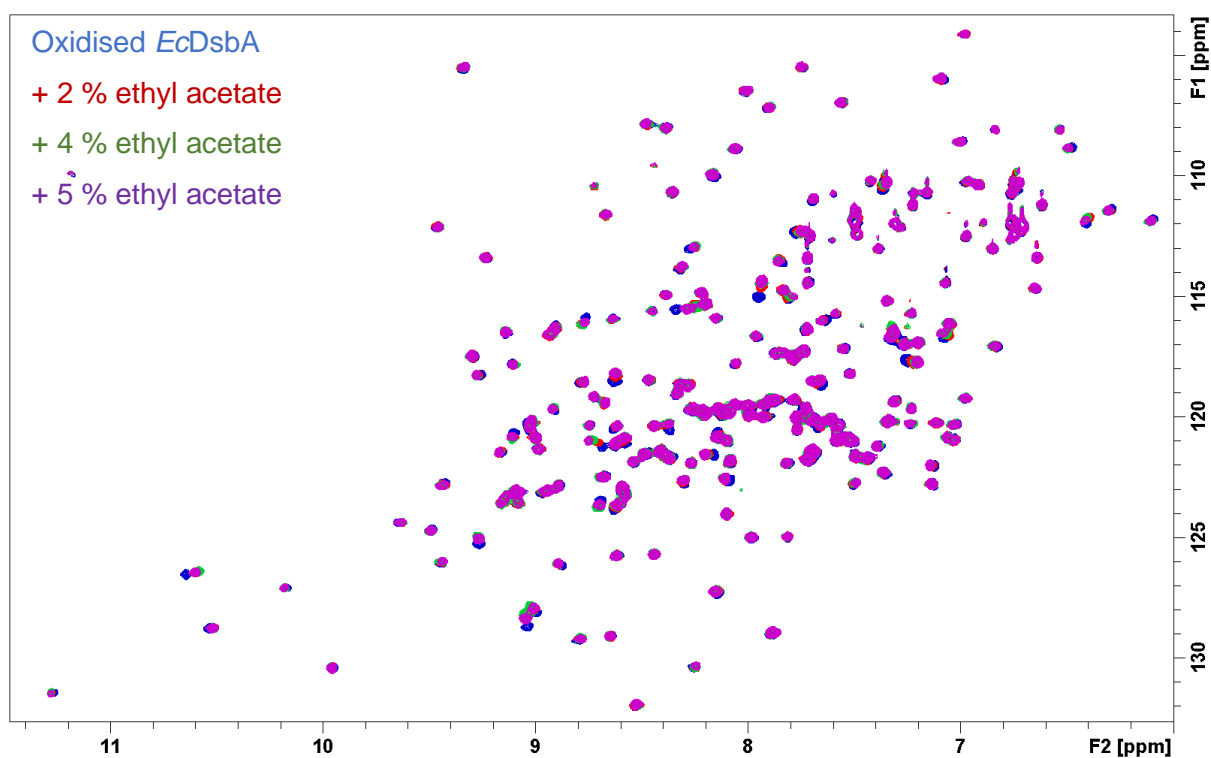
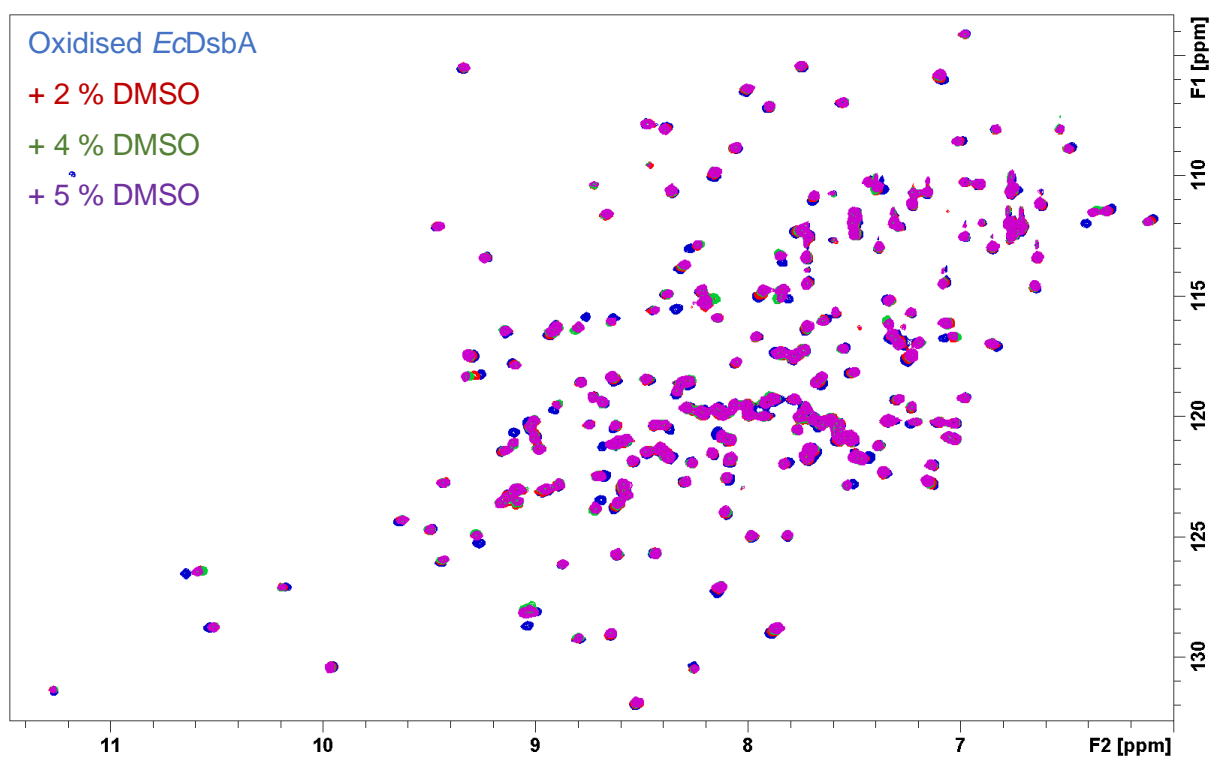
Overlay of semi-quantitative 1D  $^1\text{H}$  NMR serial dilutions in  $\text{D}_2\text{O}$  buffer. Zoom inset demonstrates the relative intensity of the internal standard (sodium trimethylsilylpropanesulfonate, DSS) and select aromatic resonances for samples containing 1 mM (purple), 2 mM (blue), 4 mM (green) and 8 mM (red) peptide **24**. The highest estimated concentration of peptide in the sample was 3 mM.

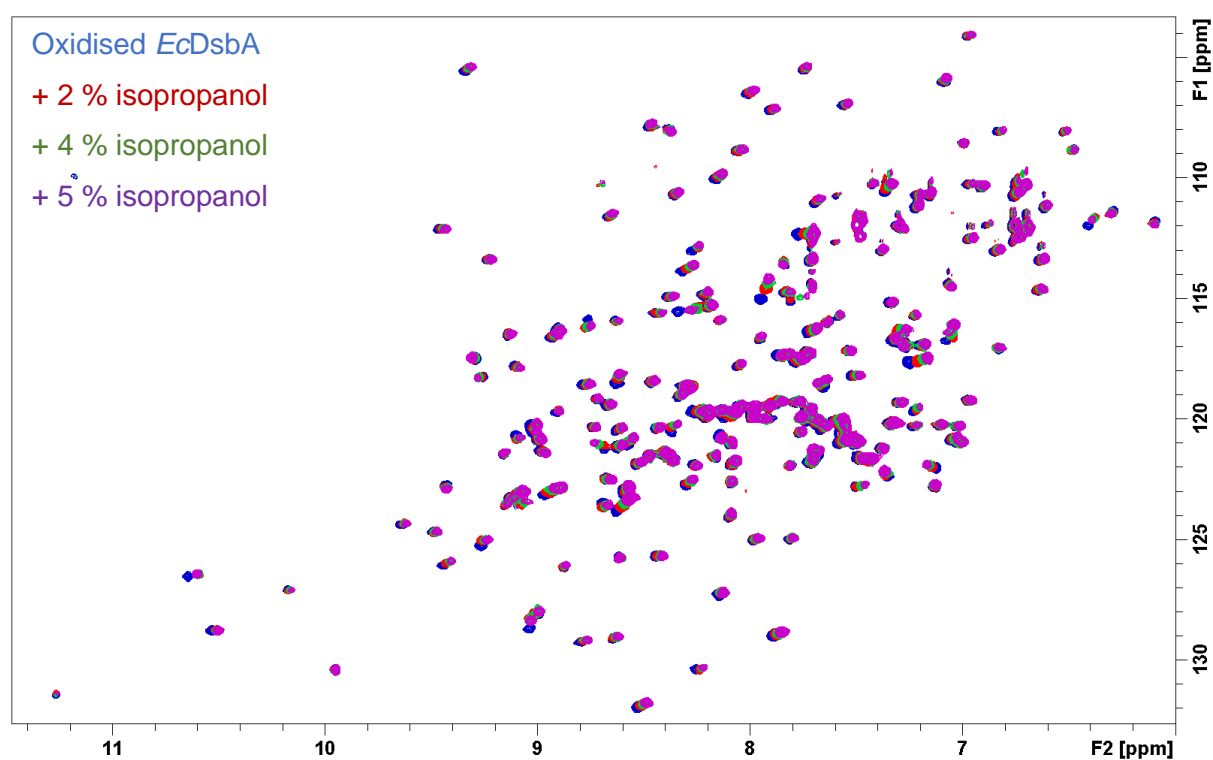
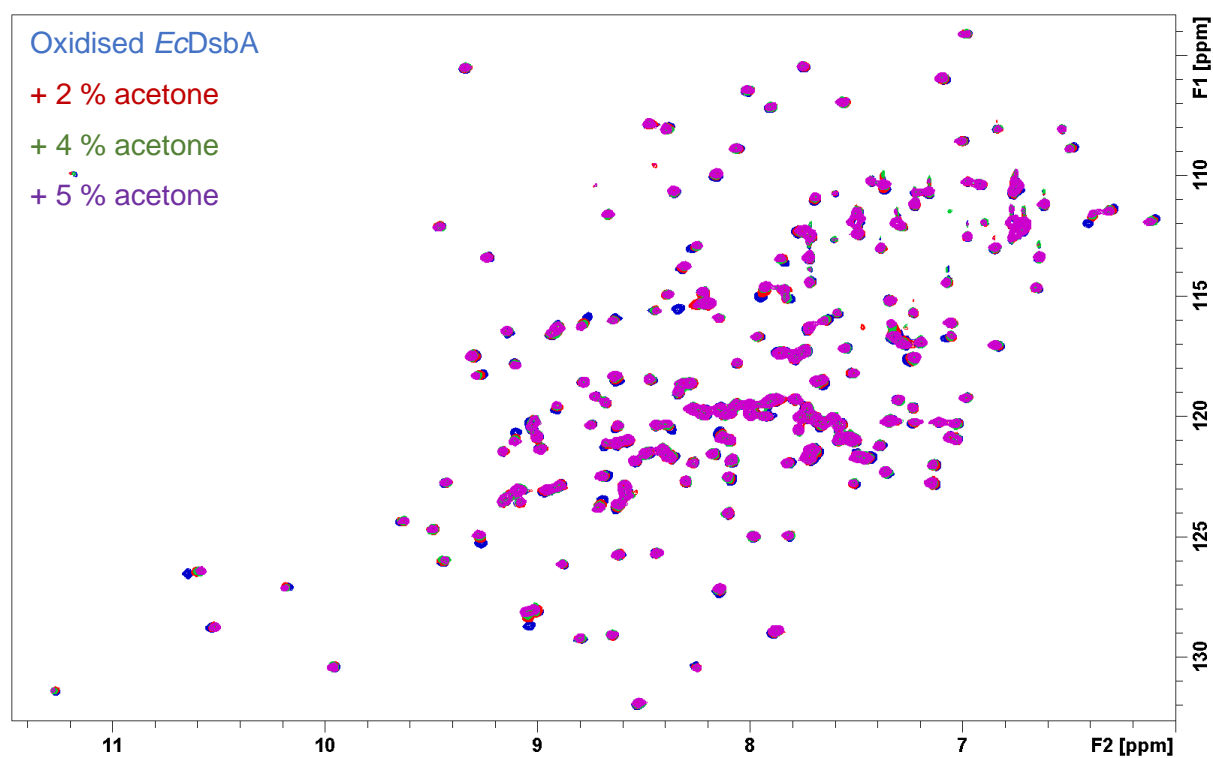


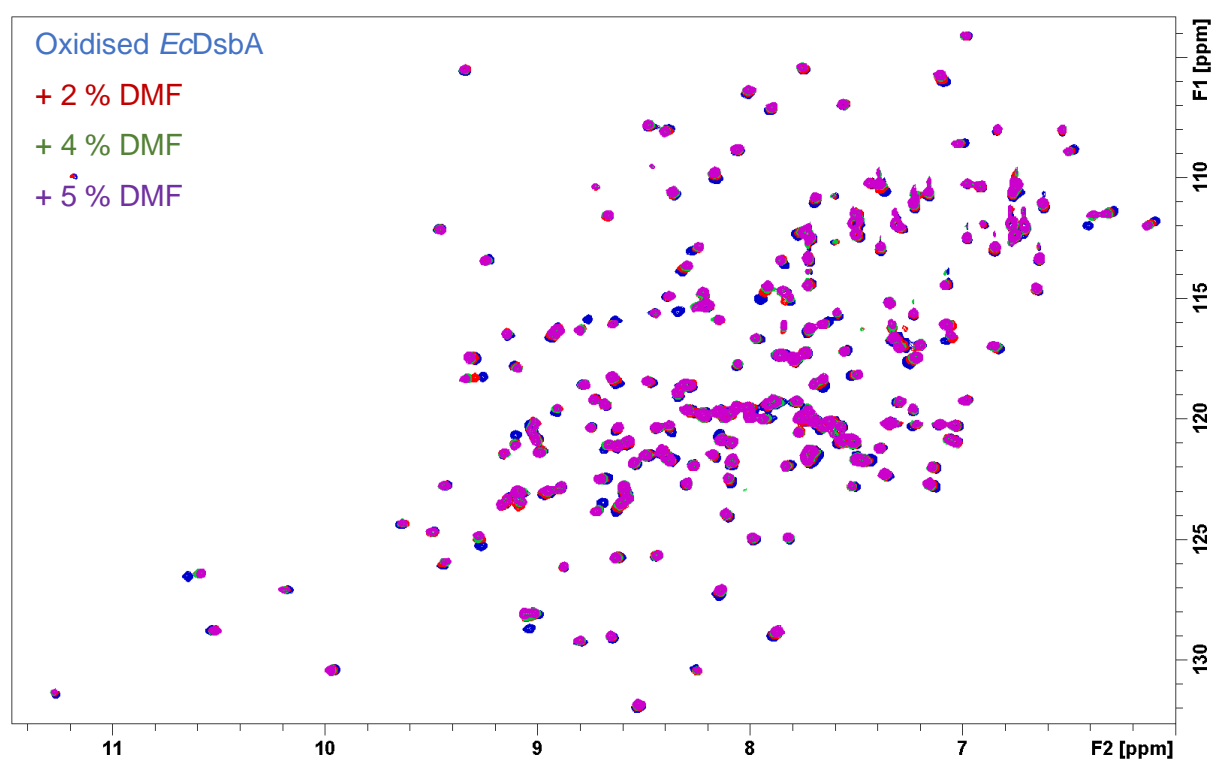
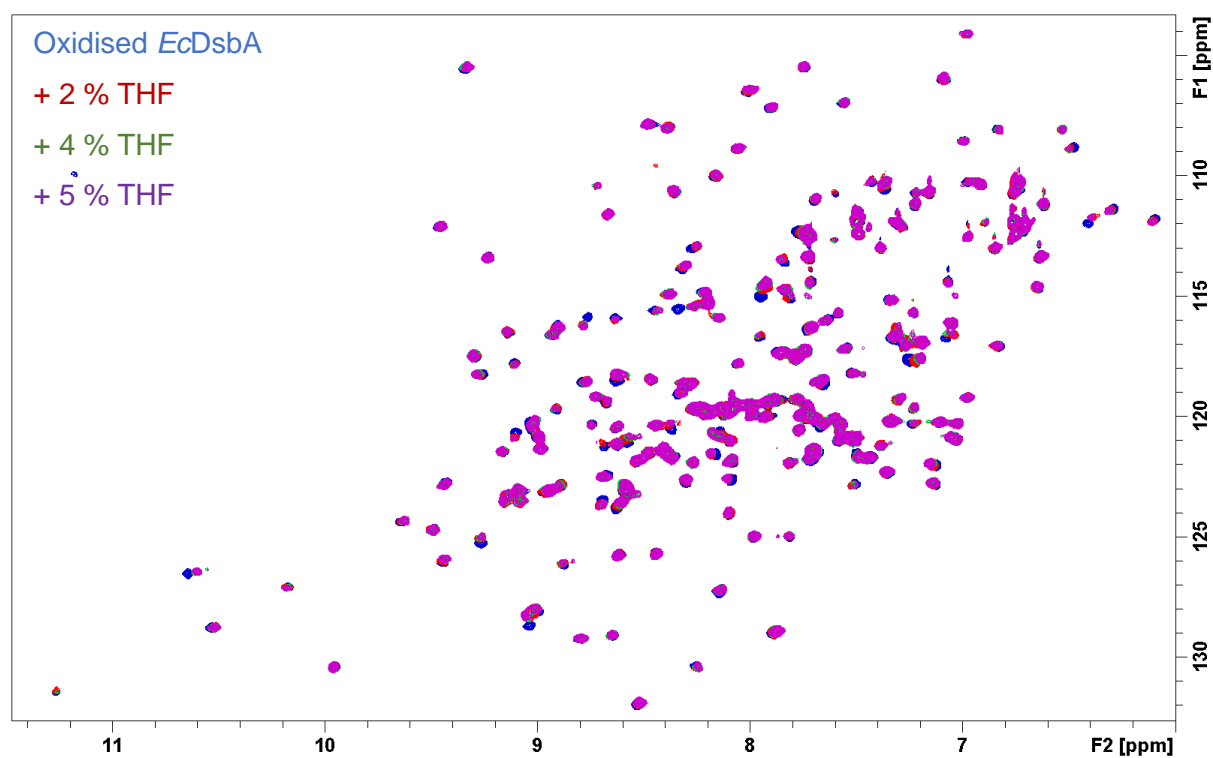
## Appendix 2: $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra and computational analysis of organic solvents

$^1\text{H}$ - $^{15}\text{N}$  HSQC spectra overlay of oxidised *EcDsbA* (100  $\mu\text{M}$ ) in the absence (blue) and presence of 2 % (red), 4 % (green), and 5 % (purple) (v/v) organic solvent. Acetic acid (5 % v/v) and urea (1 M) were conducted as a single point HSQC. F1 =  $^{15}\text{N}$  nucleus, F2 =  $^1\text{H}$  nucleus.

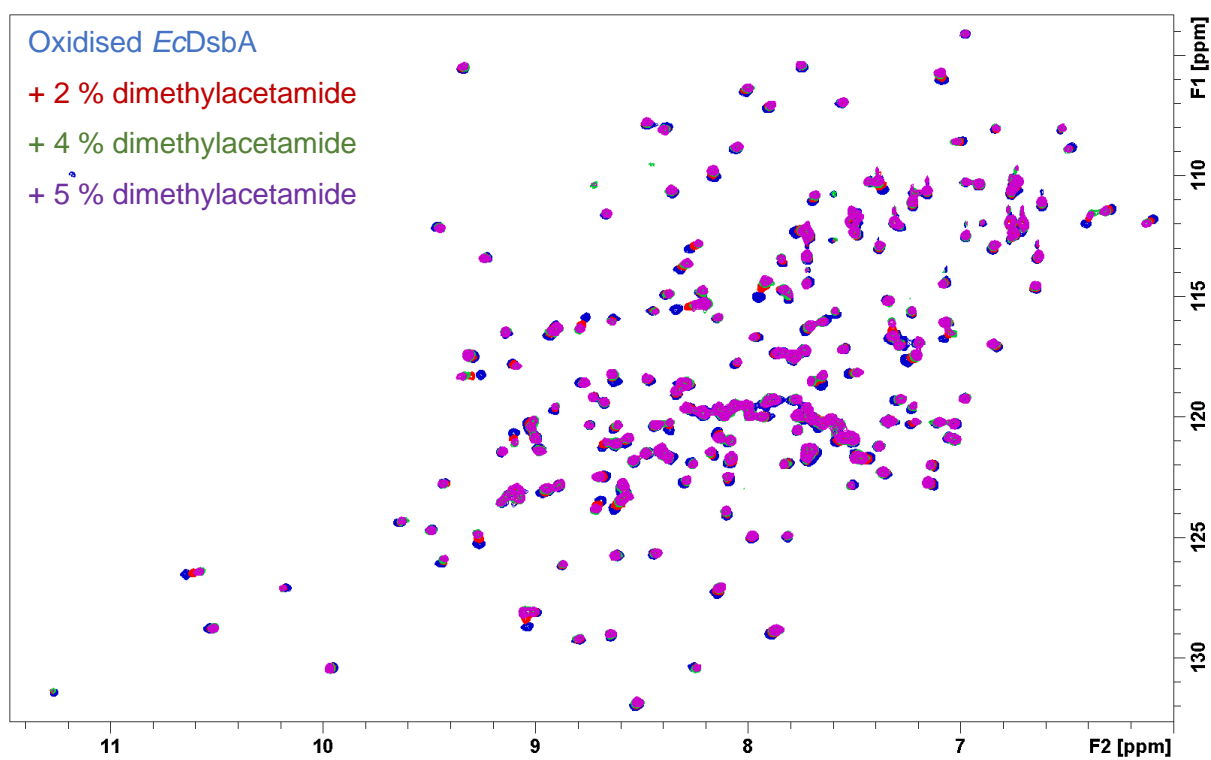
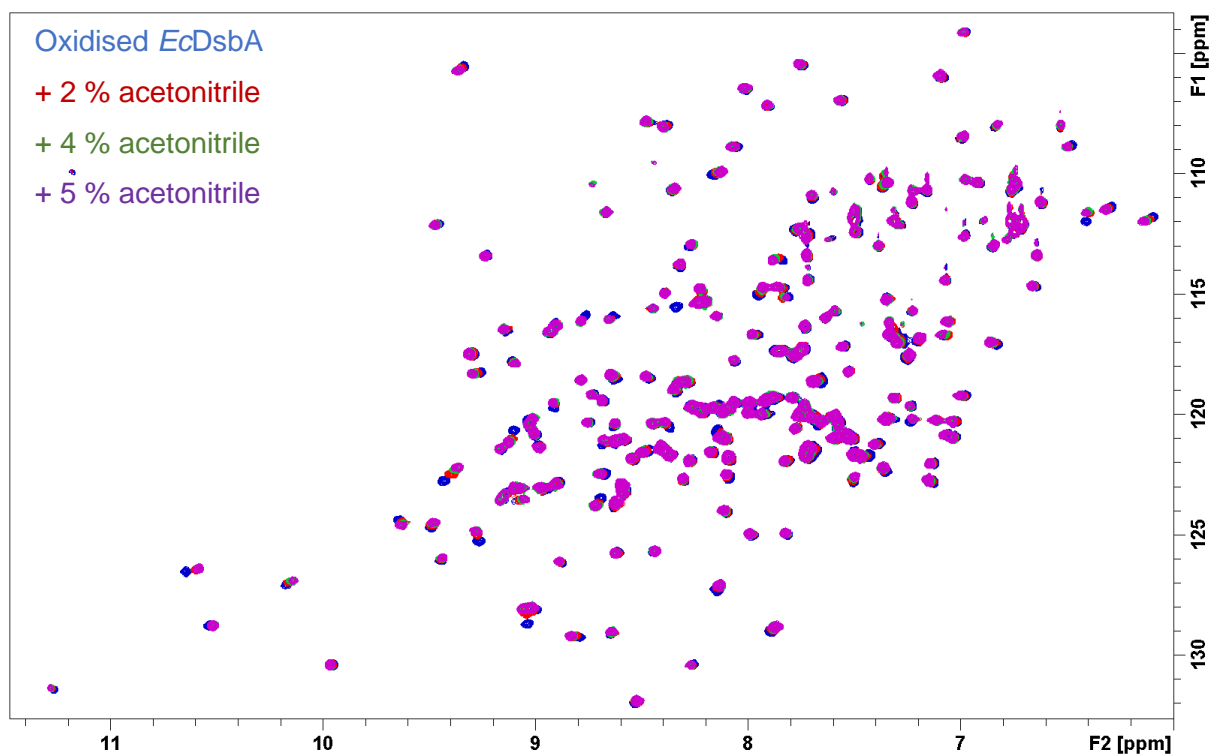


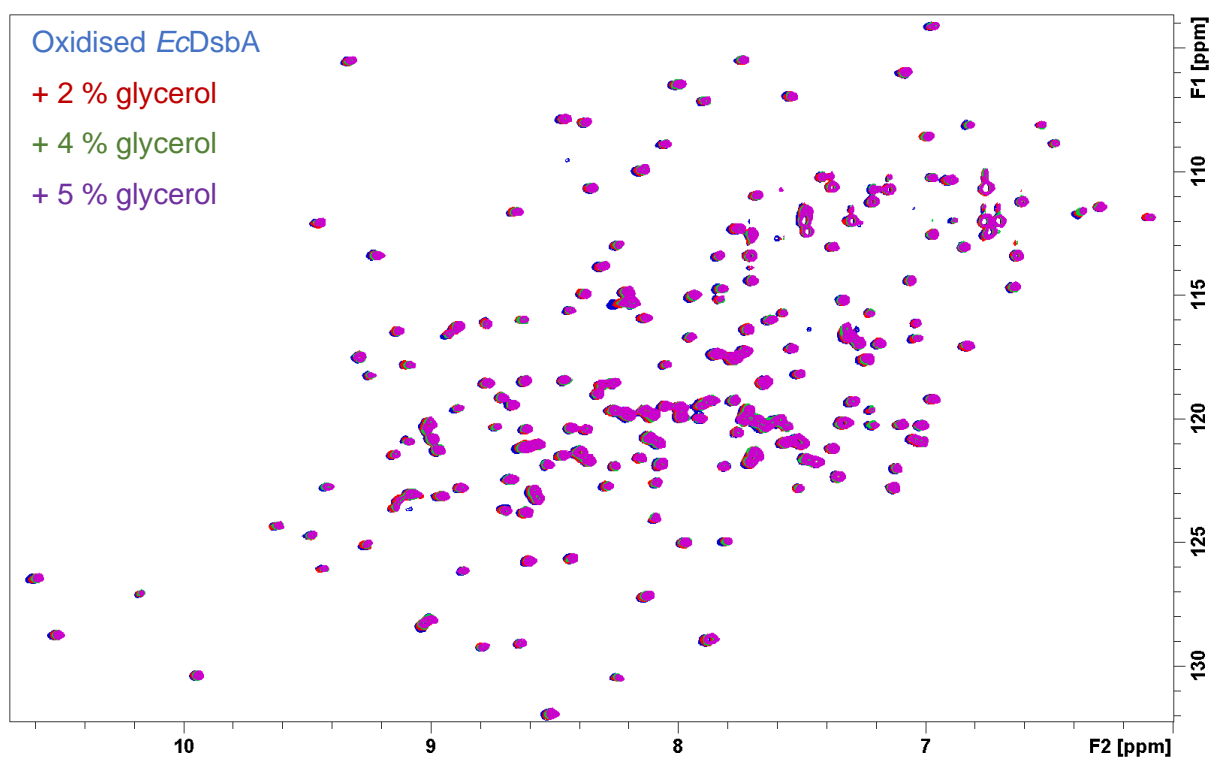
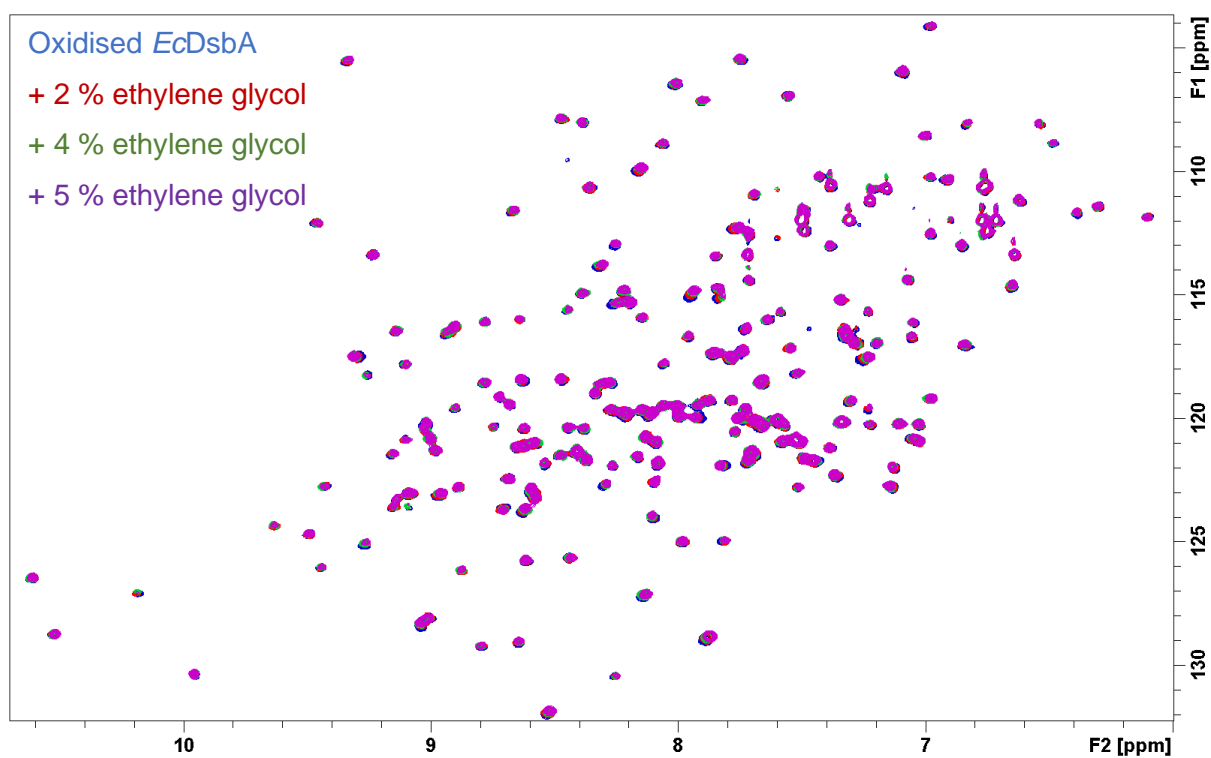


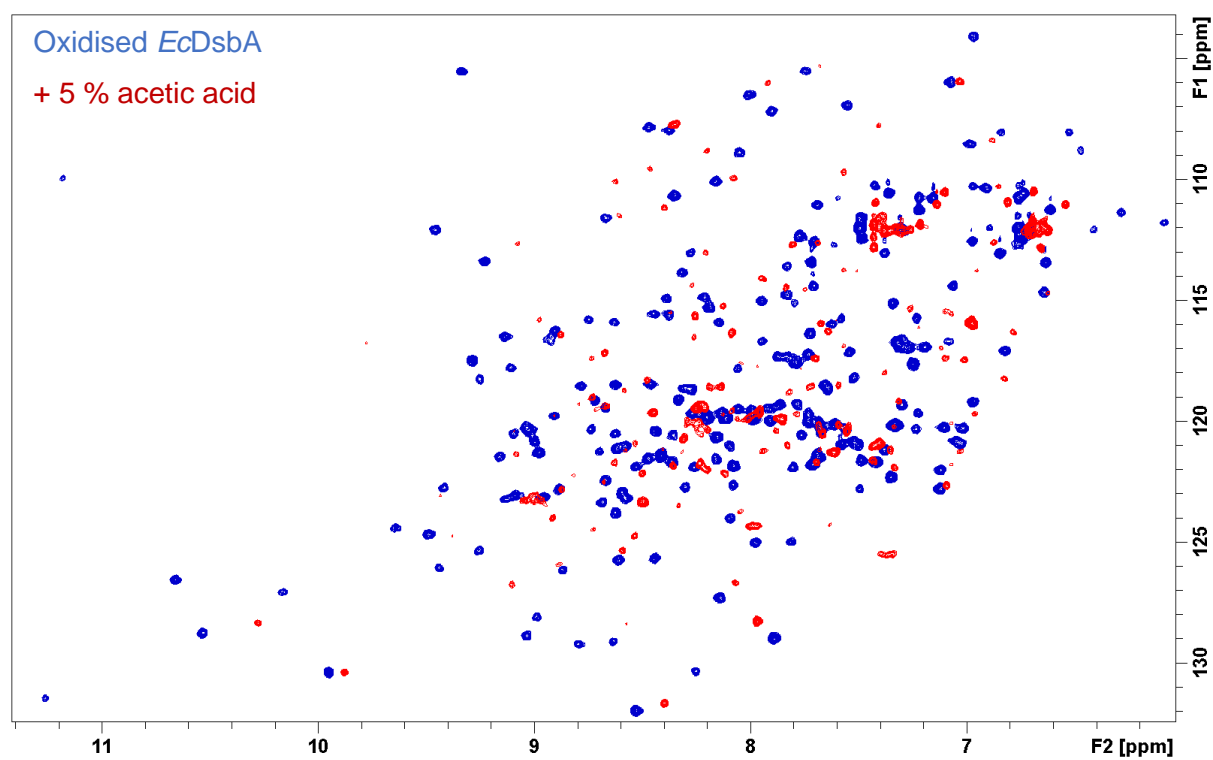
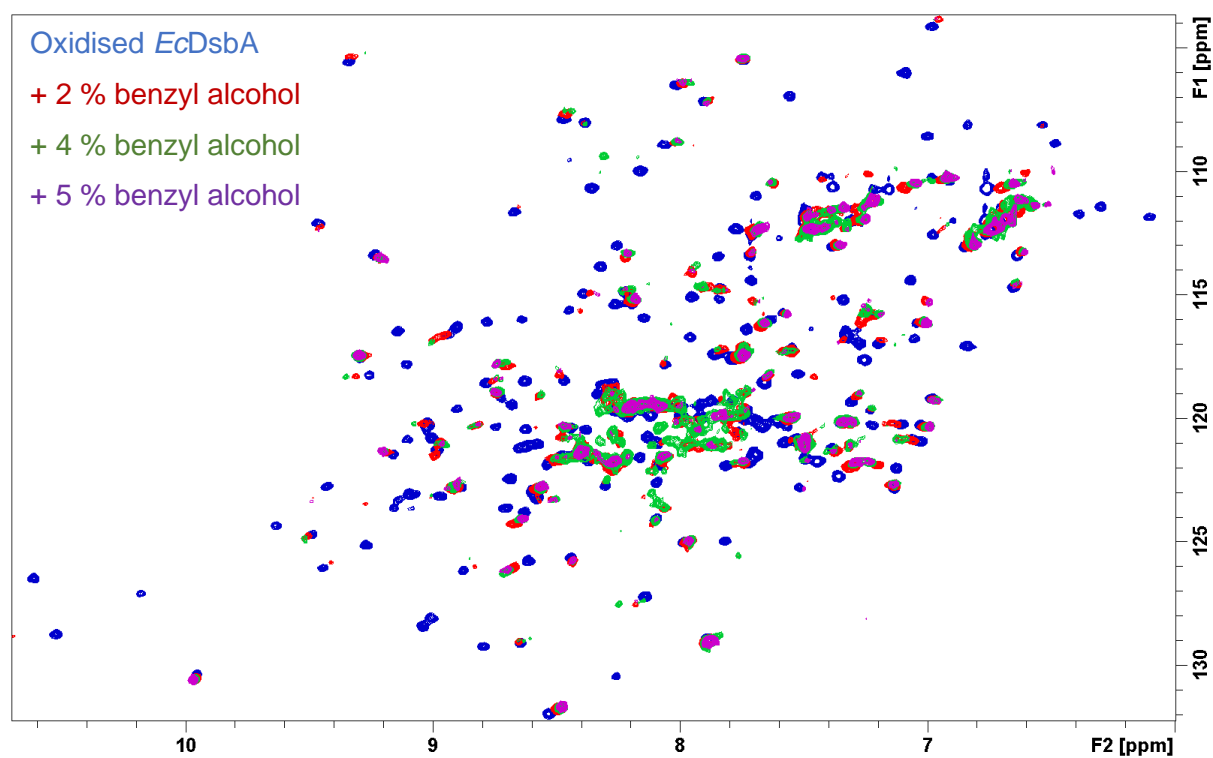


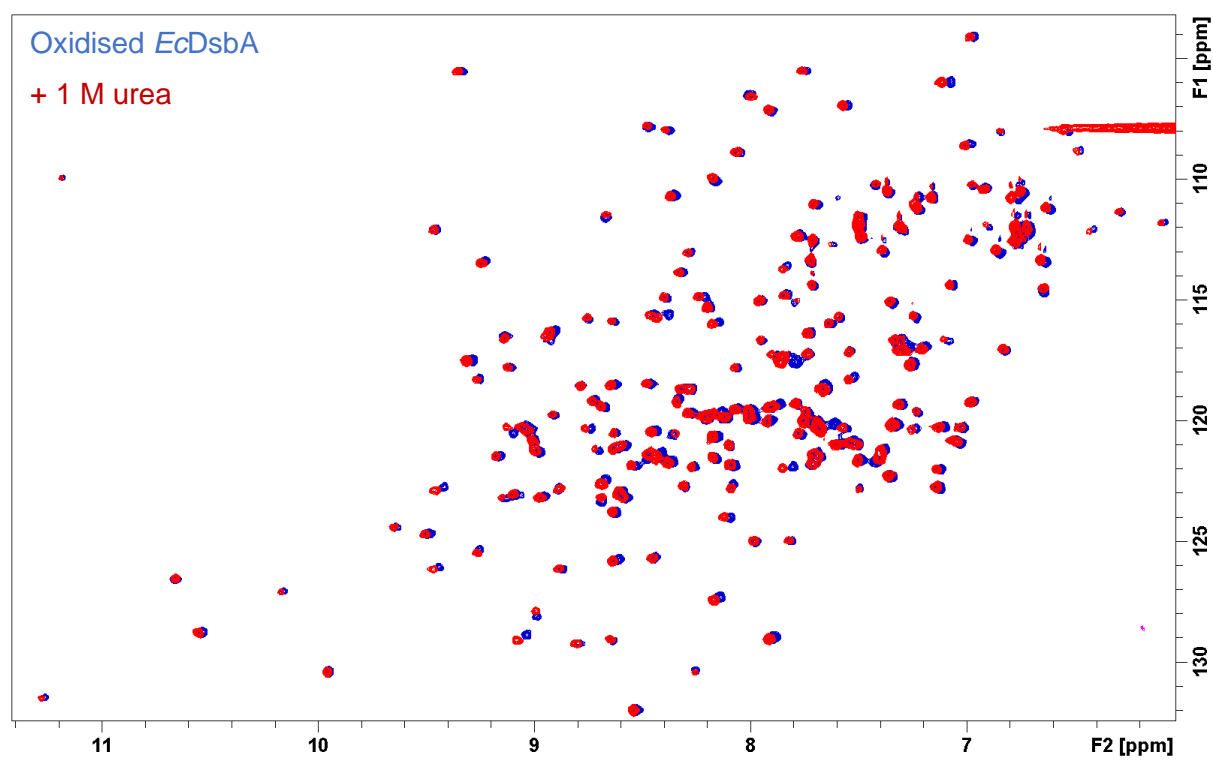




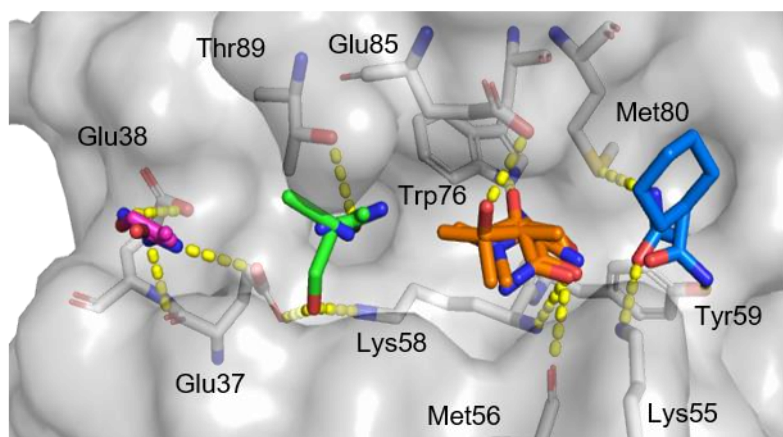
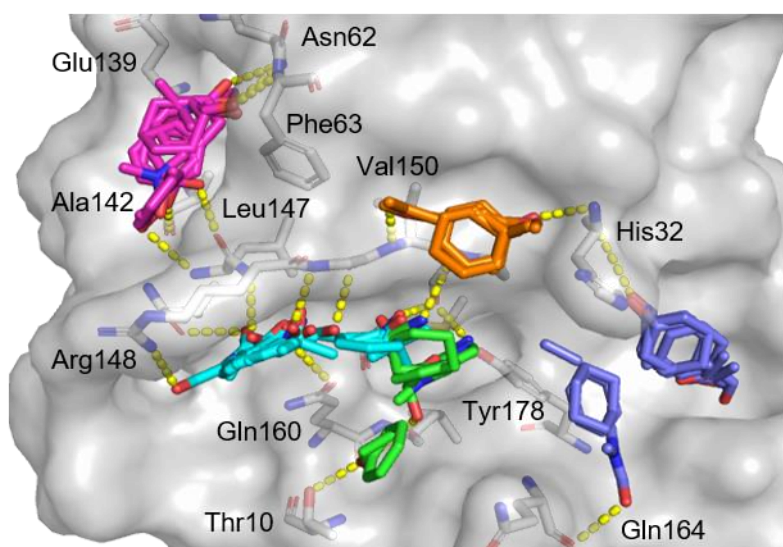
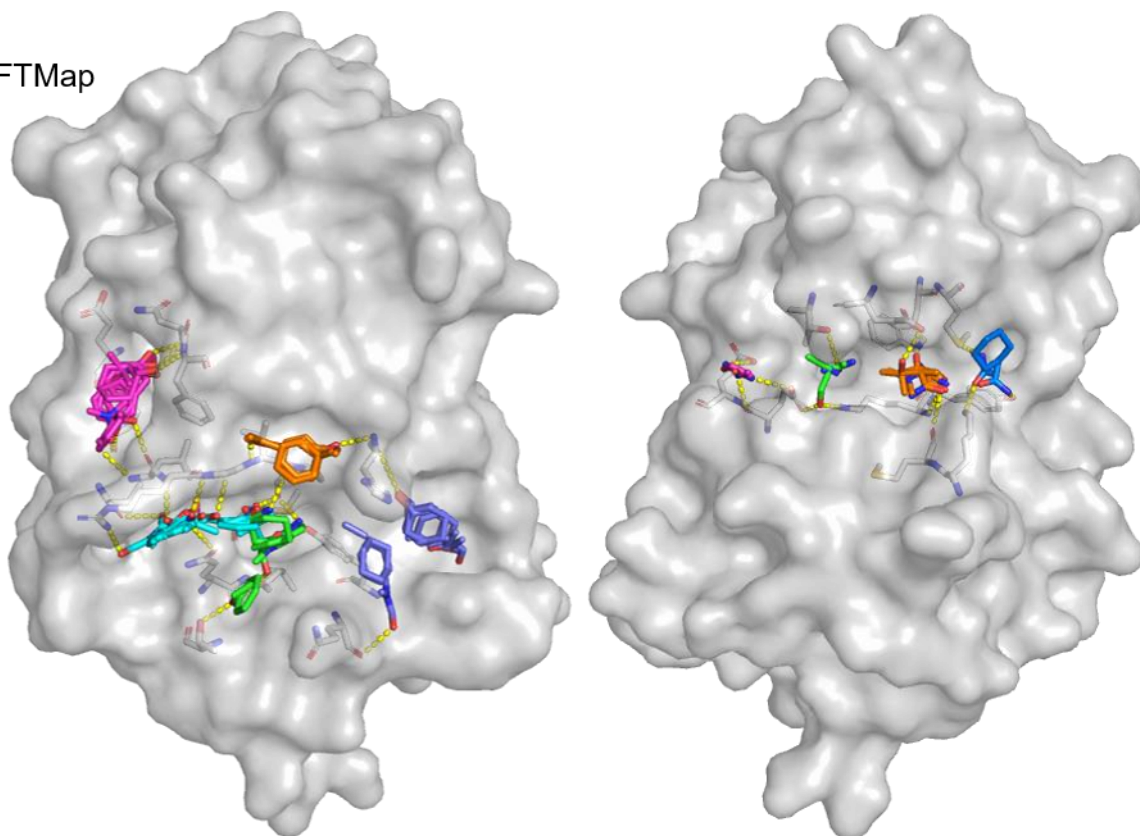




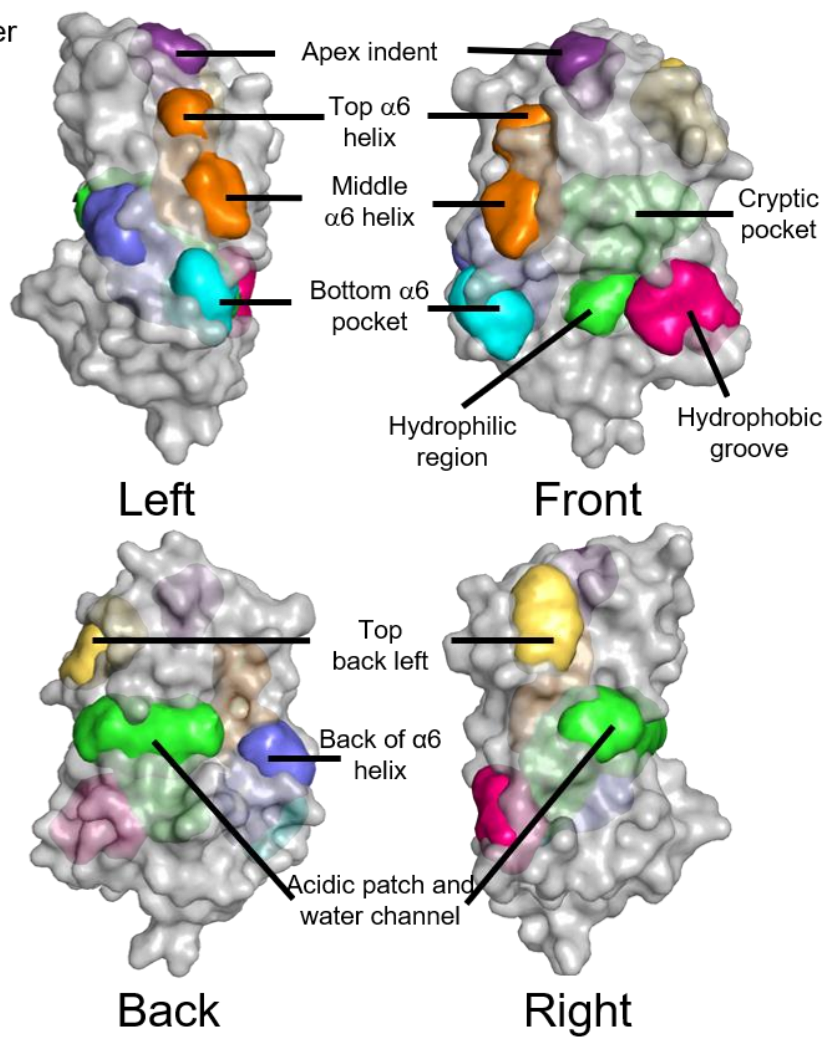




FTMap



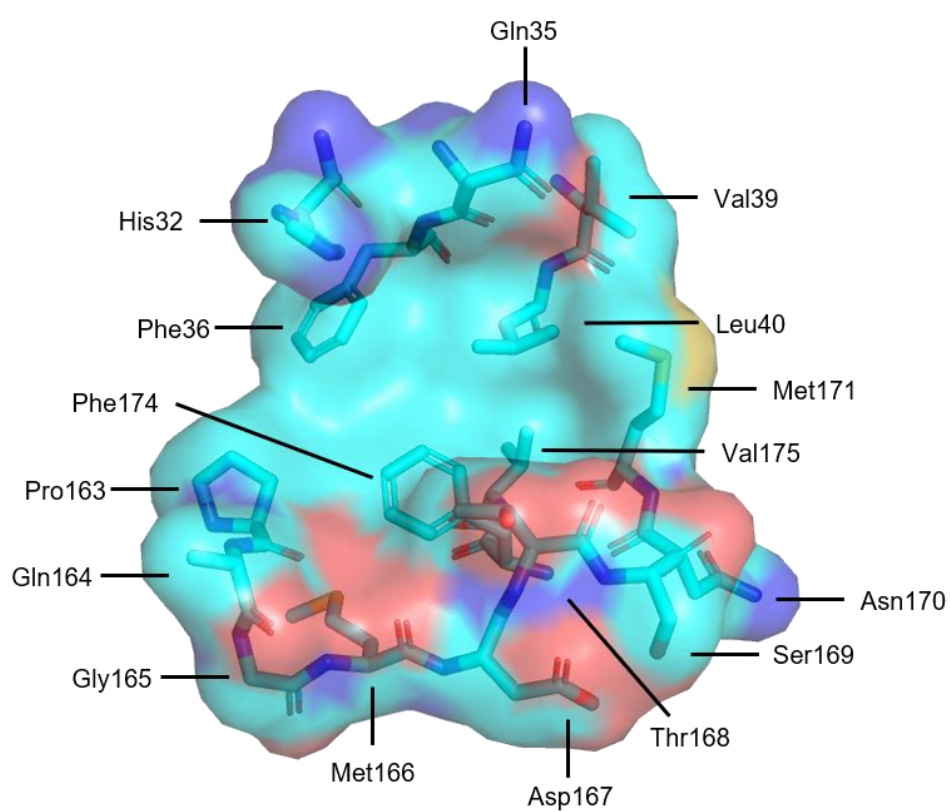
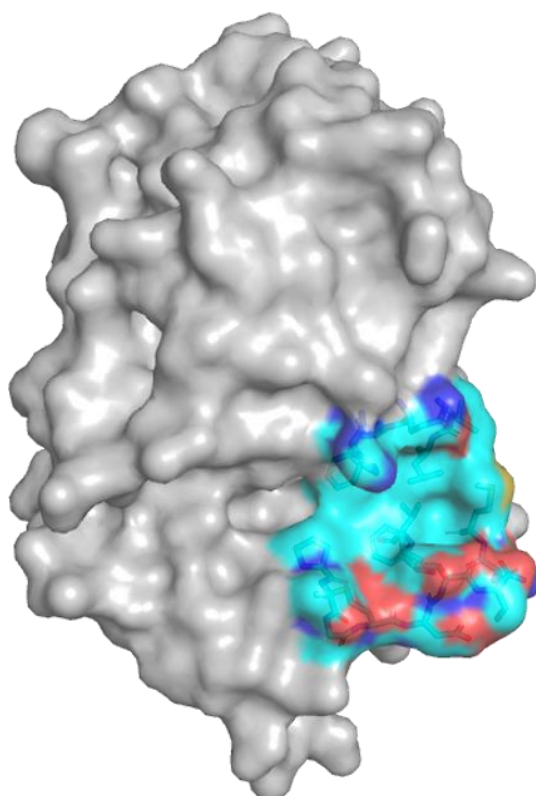
DoGSiteScorer



Hot spot	Surface area ( $\text{\AA}^2$ )	Volume ( $\text{\AA}^3$ )	Druggability score
Hydrophobic groove	233	177	0.35
Cryptic pocket, hydrophilic region, acidic patch and water channel	836	660	0.79
Apex	149	104	0.25
Top and middle $\alpha 6$ helix	441	245	0.59
Bottom $\alpha 6$ helix	507	188	0.49
Back $\alpha 6$ helix	464	252	0.61
Top back left	255	141	0.21

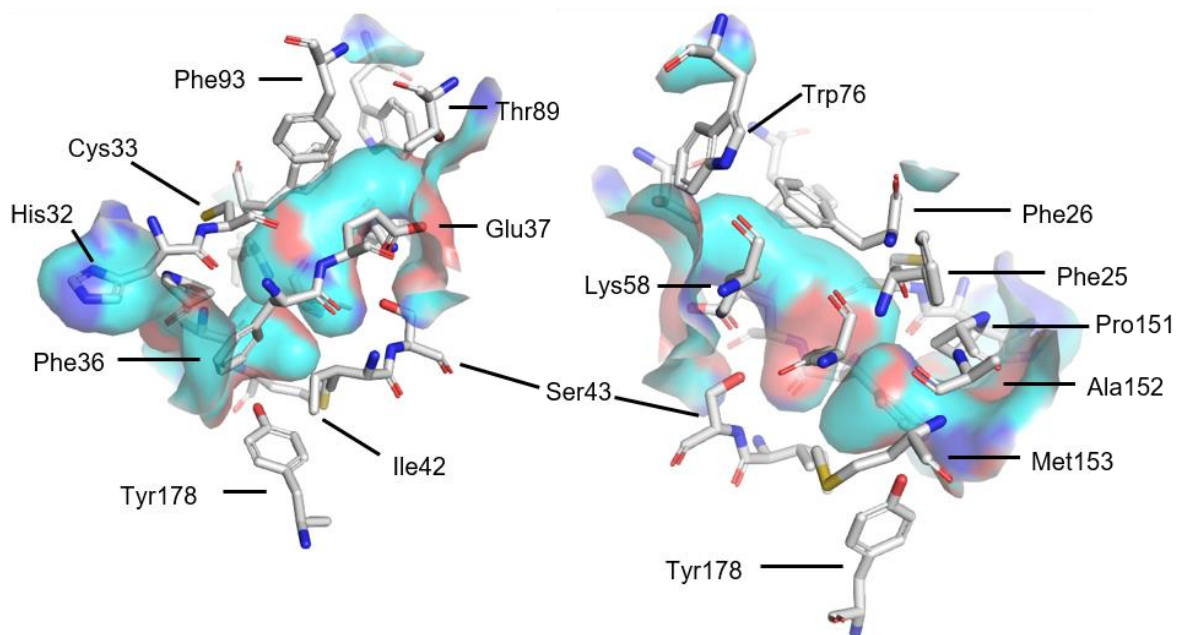
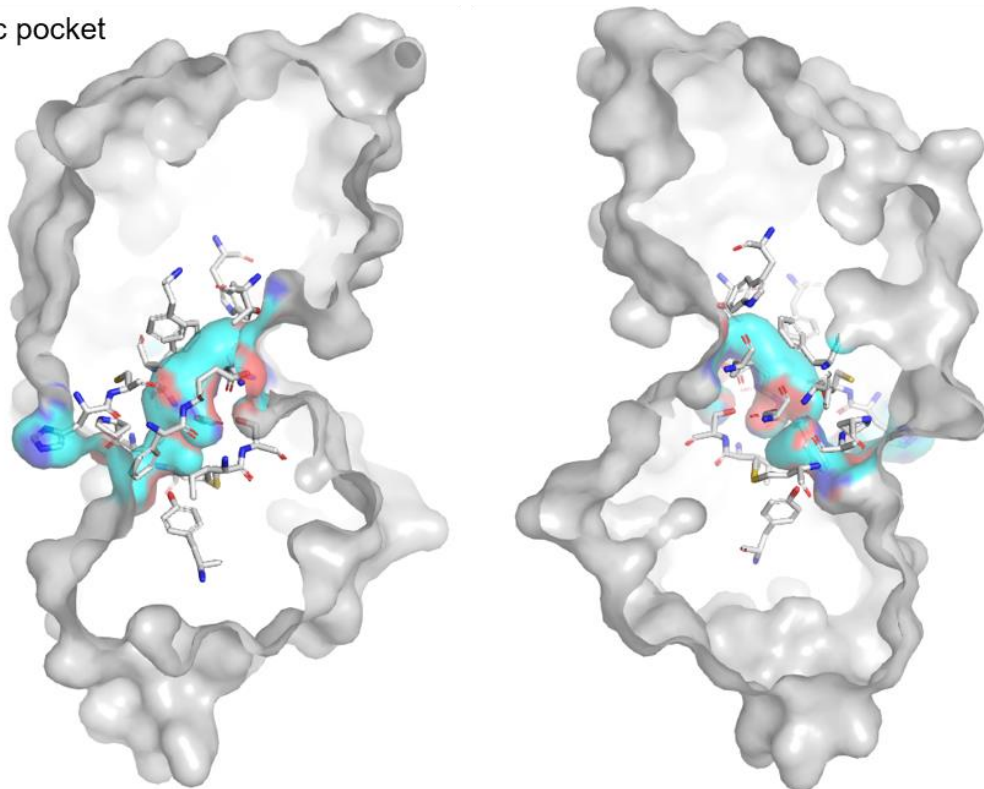
### Appendix 3: Binding site definitions

Hydrophobic groove



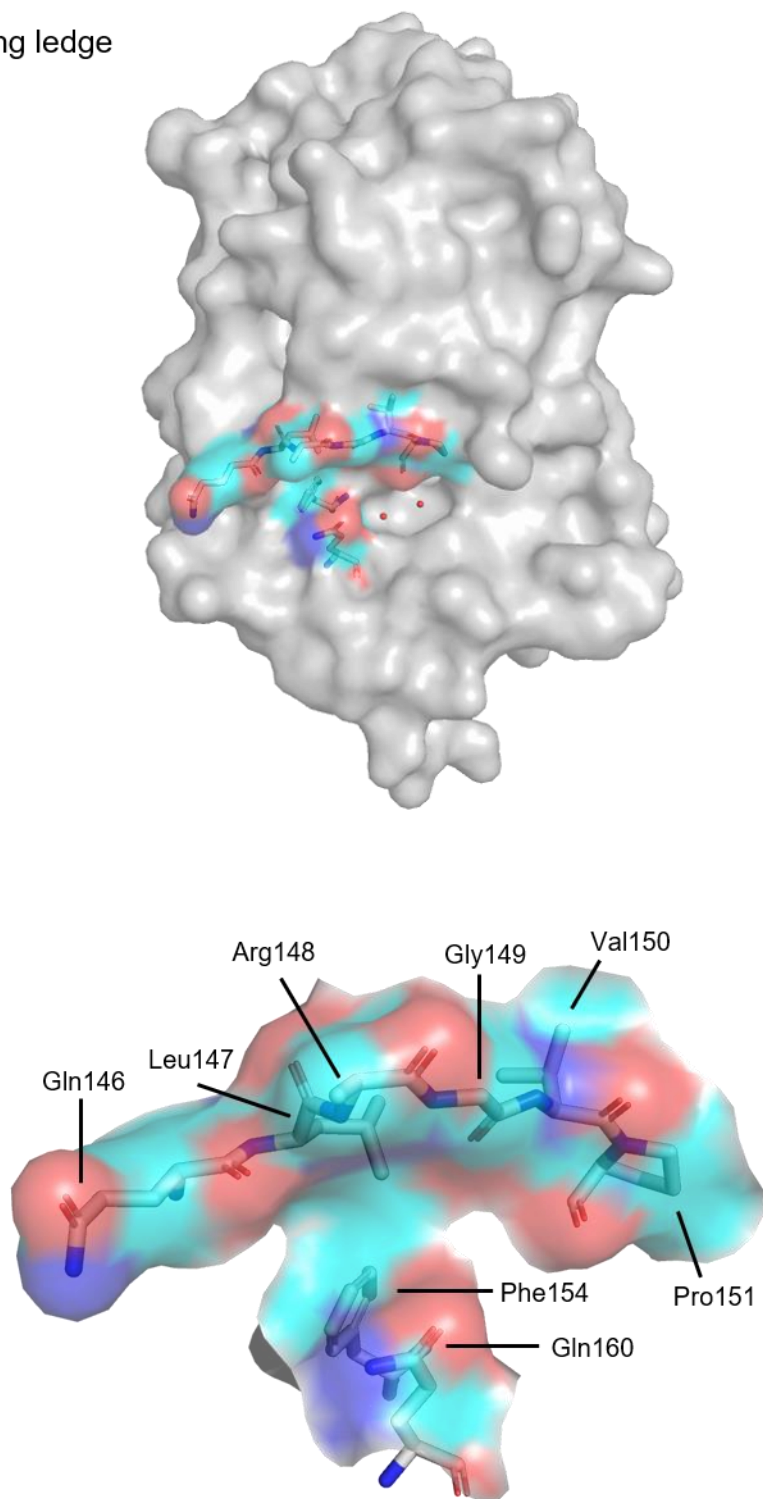


Cryptic pocket

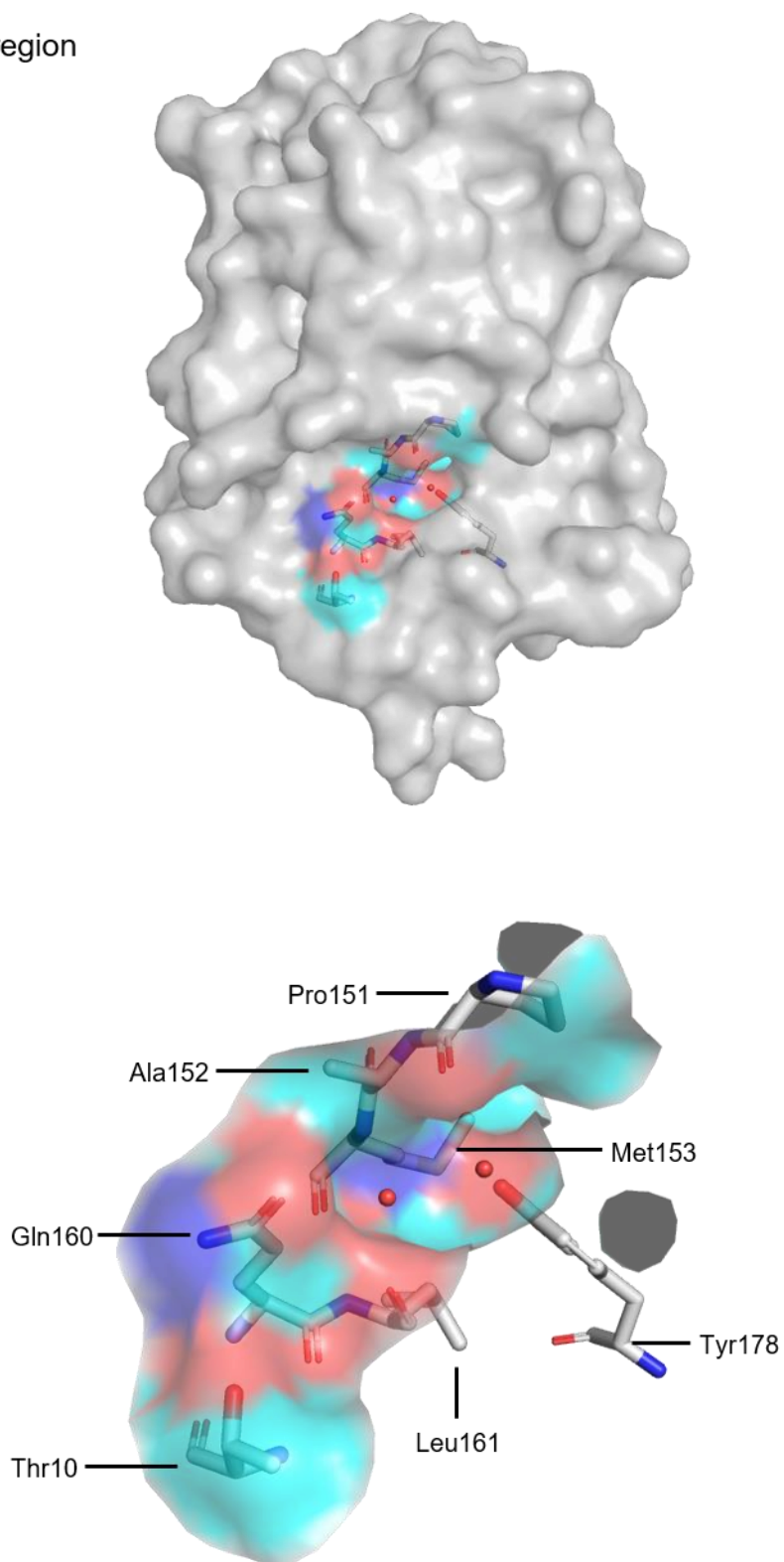




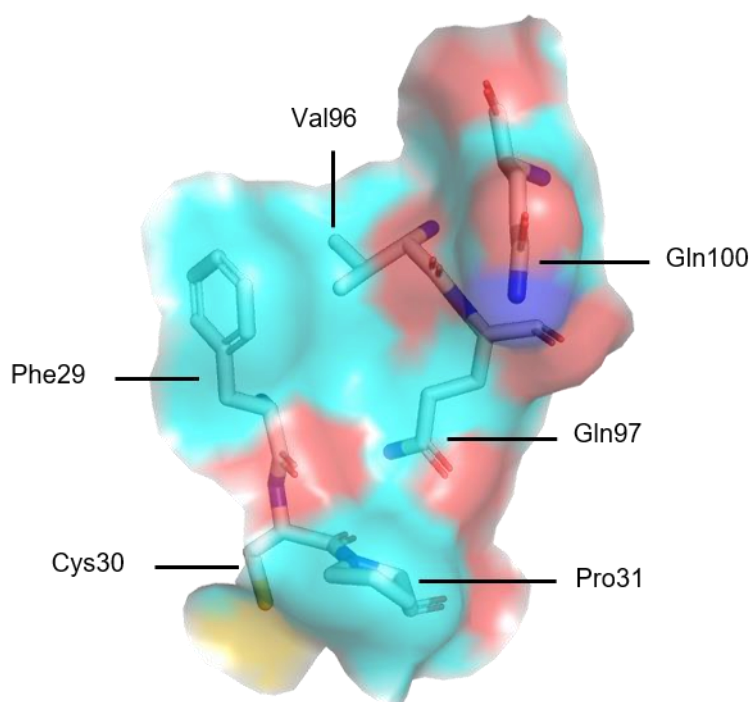
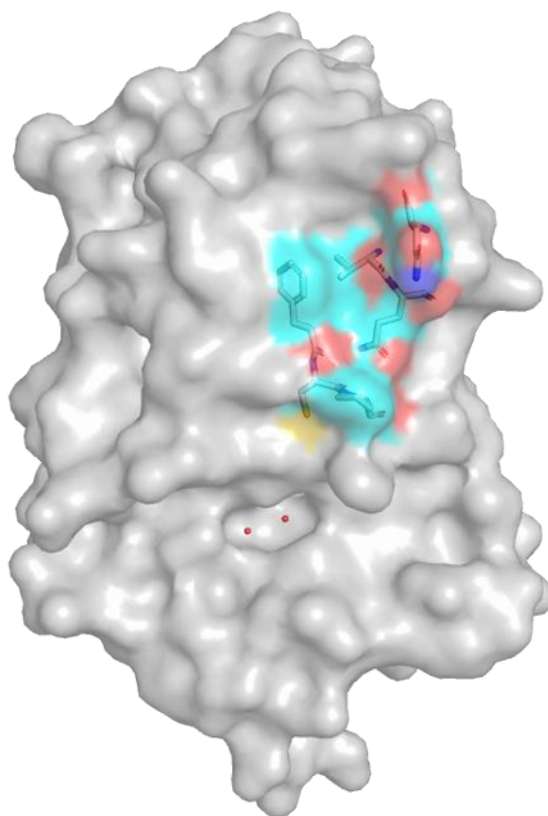
Peptide binding ledge



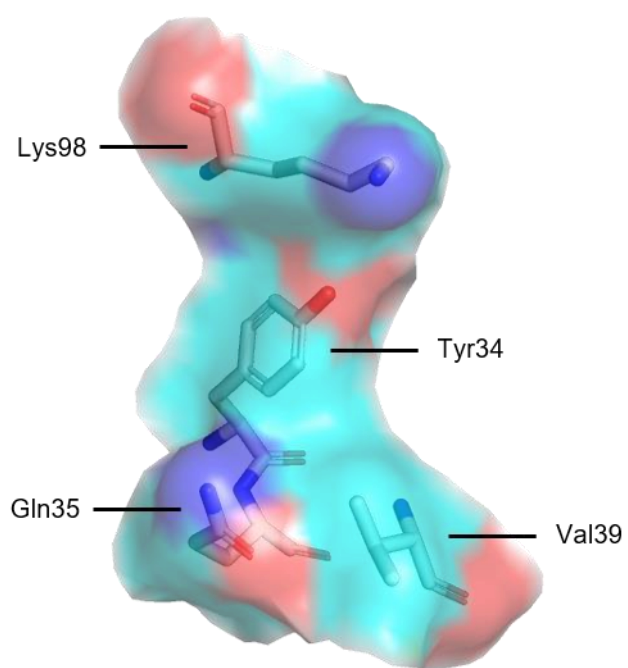
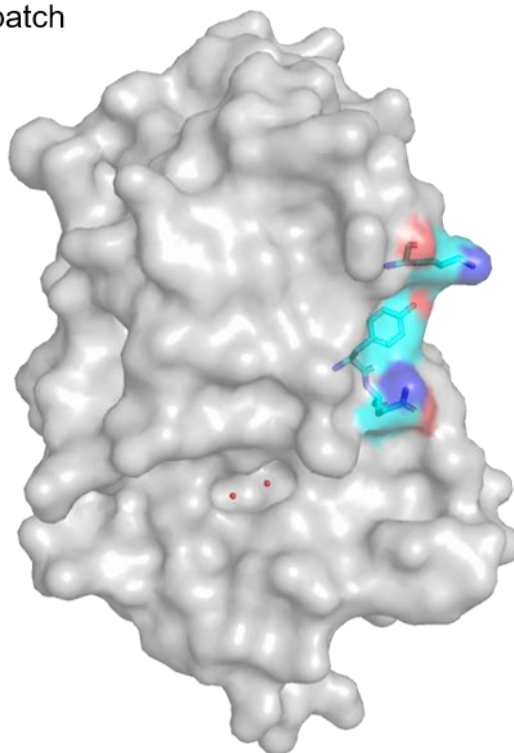
Hydrophilic region



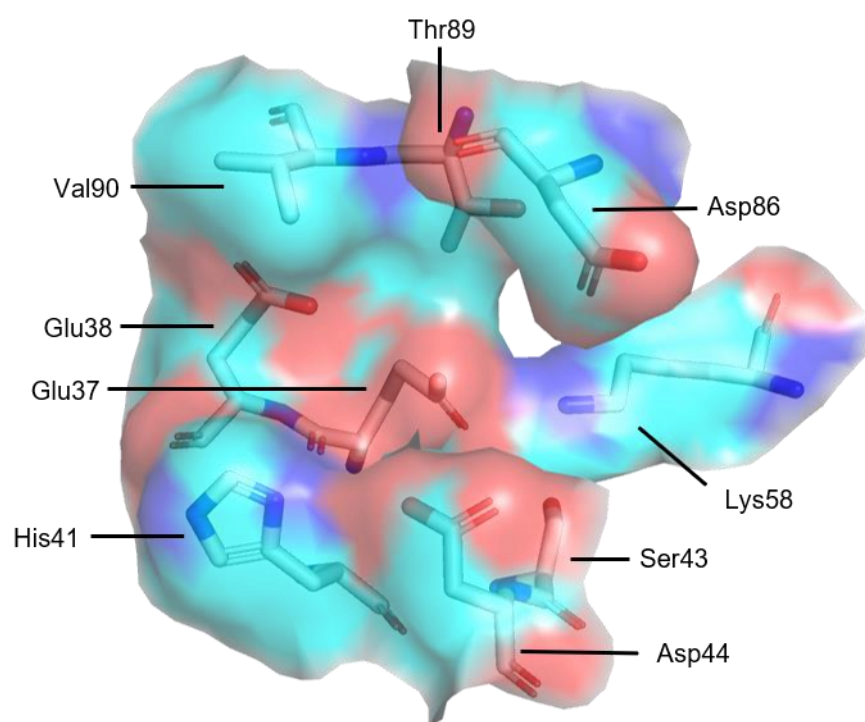
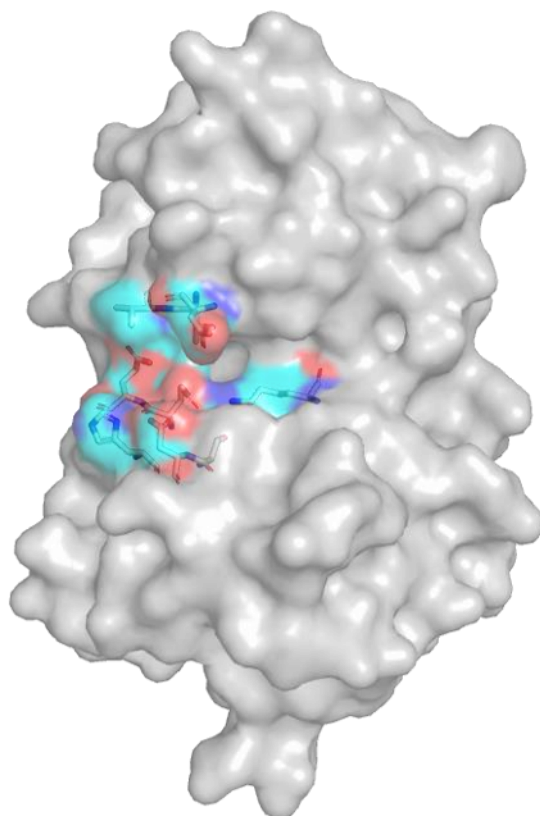
Hydrophobic patch



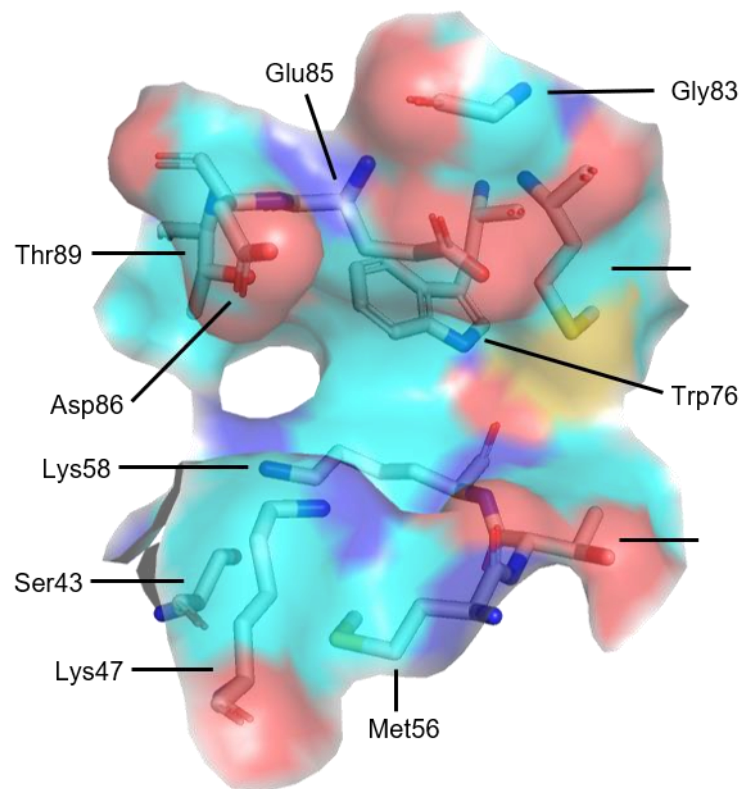
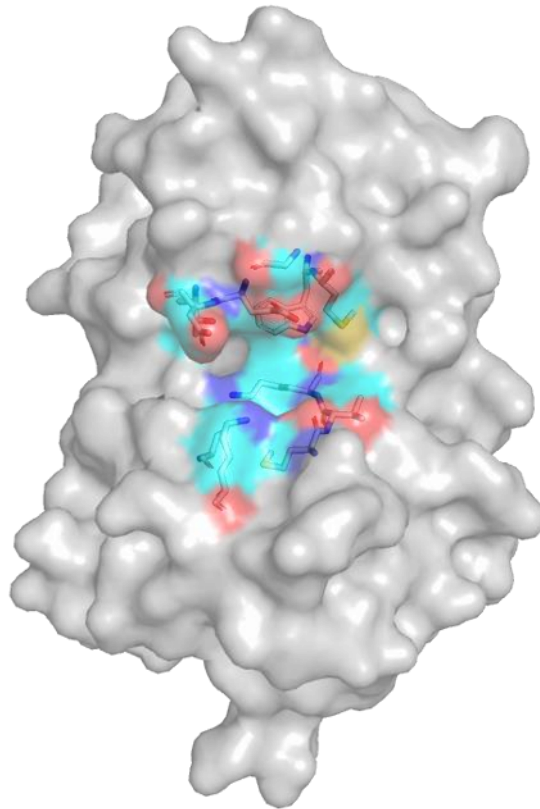
Extended hydrophobic patch



Acidic patch

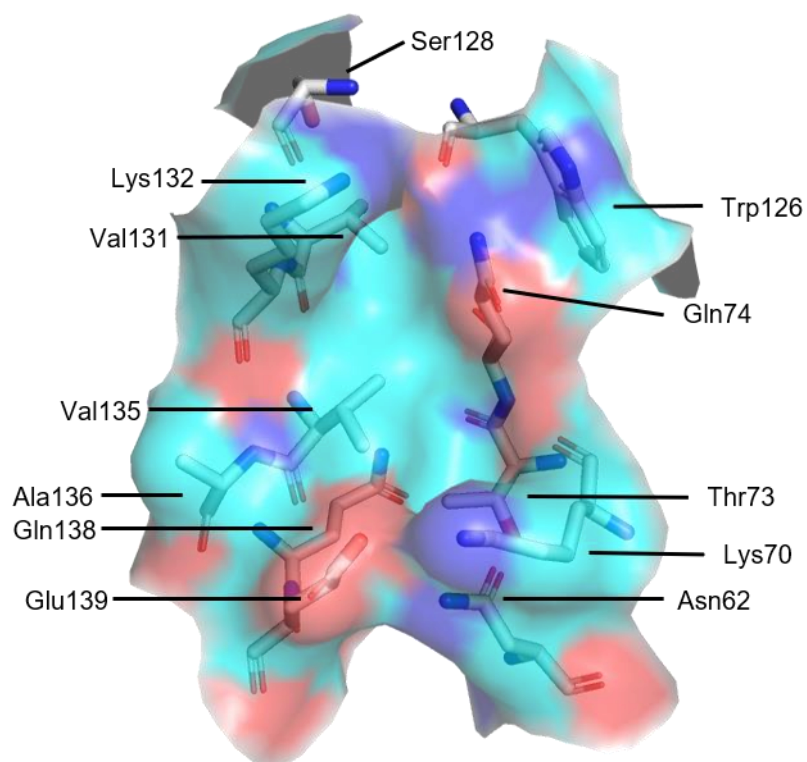
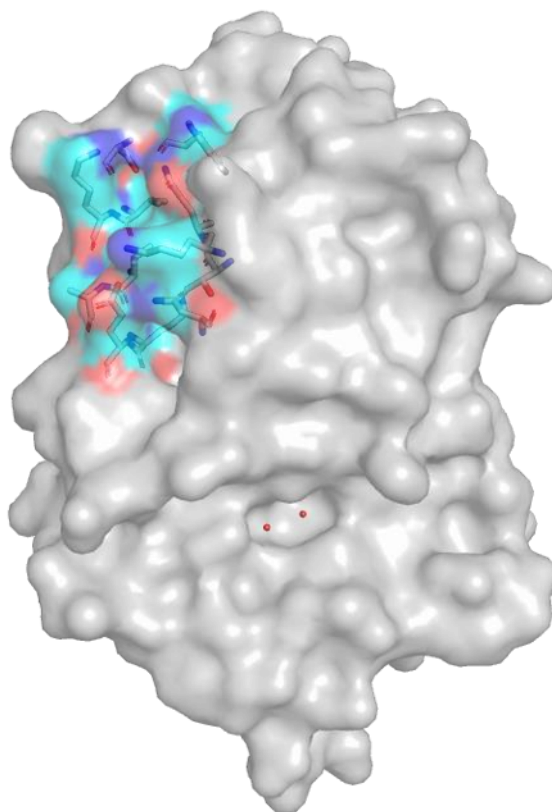


Water channel

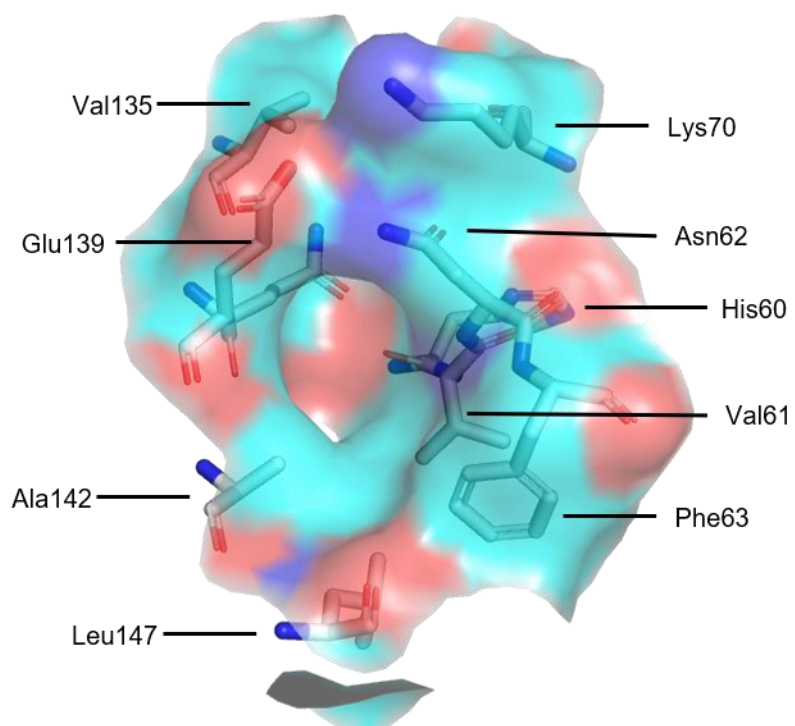
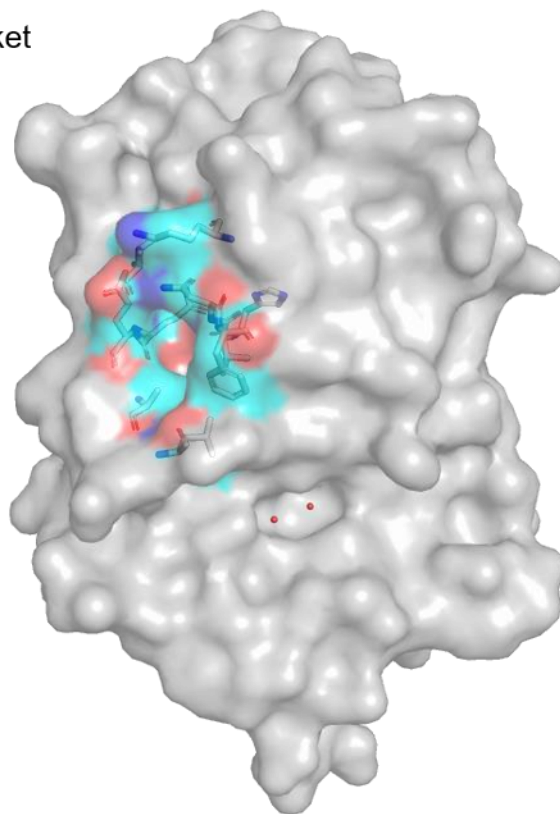




Top  $\alpha 6$  helix pocket

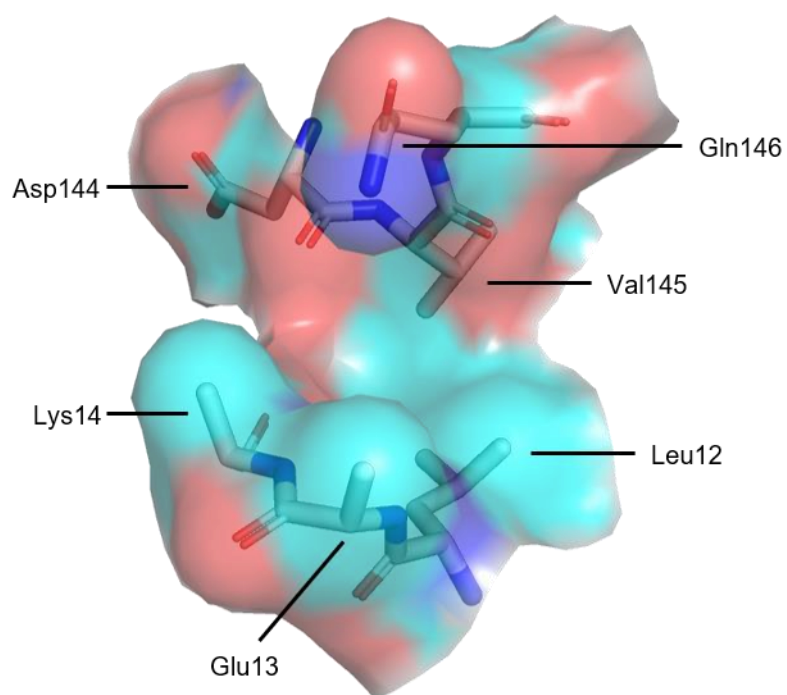
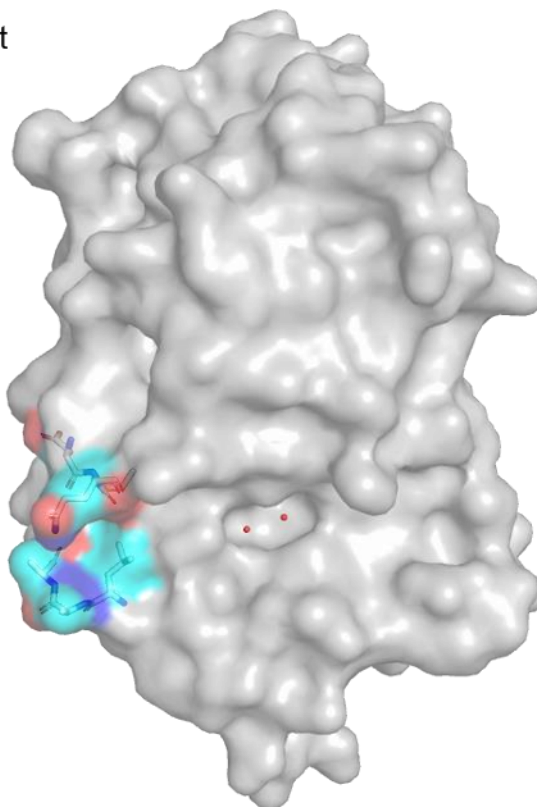


Middle  $\alpha 6$  helix pocket

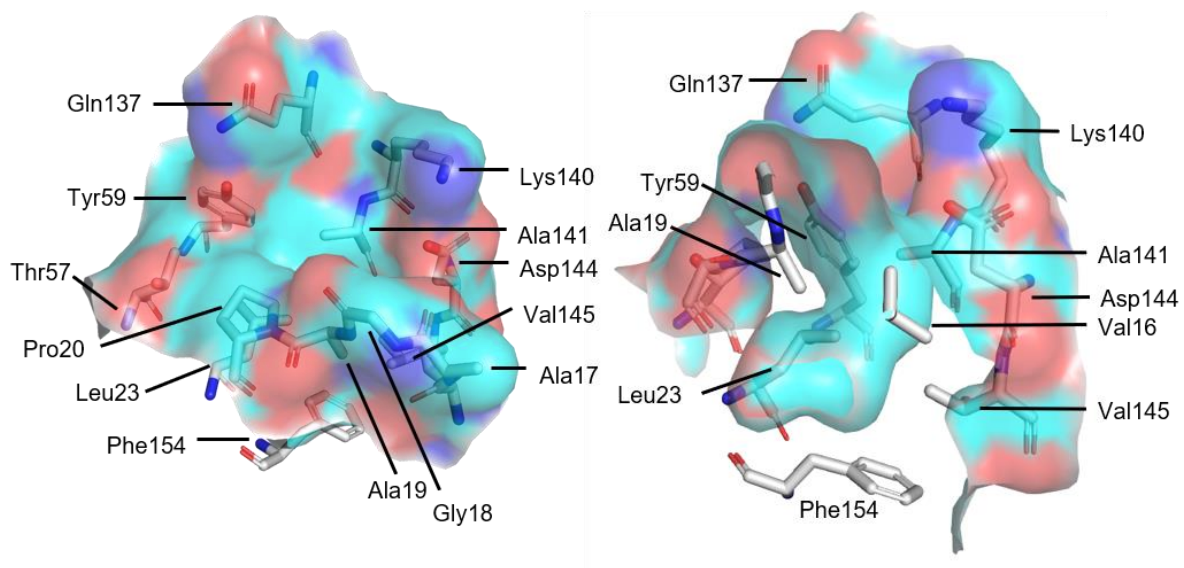
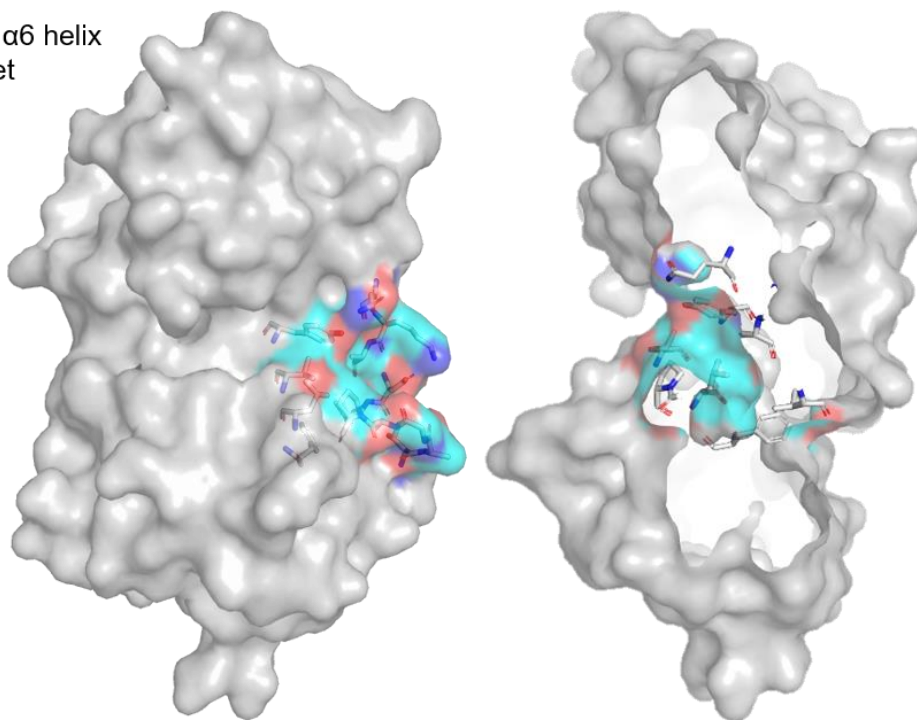




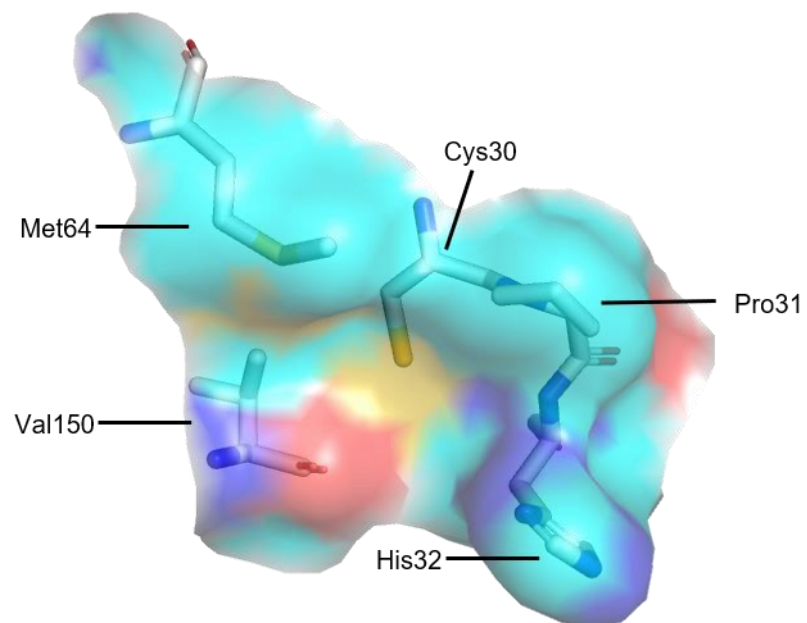
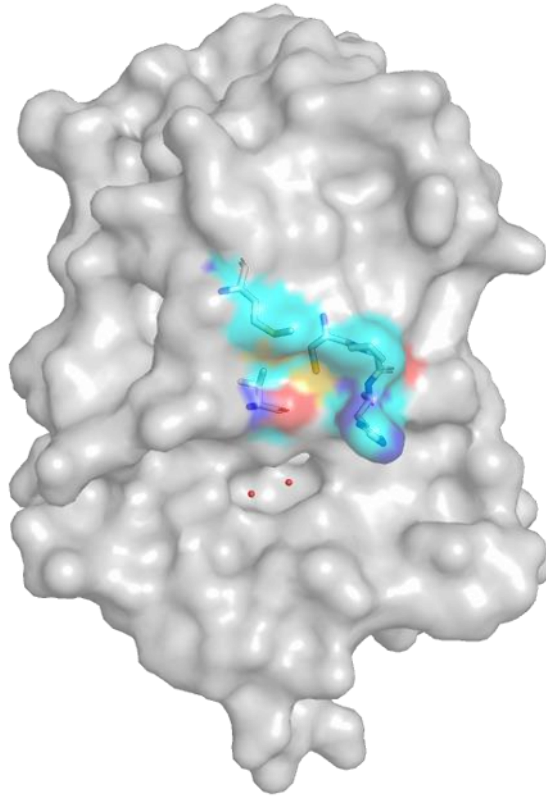
Bottom  $\alpha 6$  helix pocket



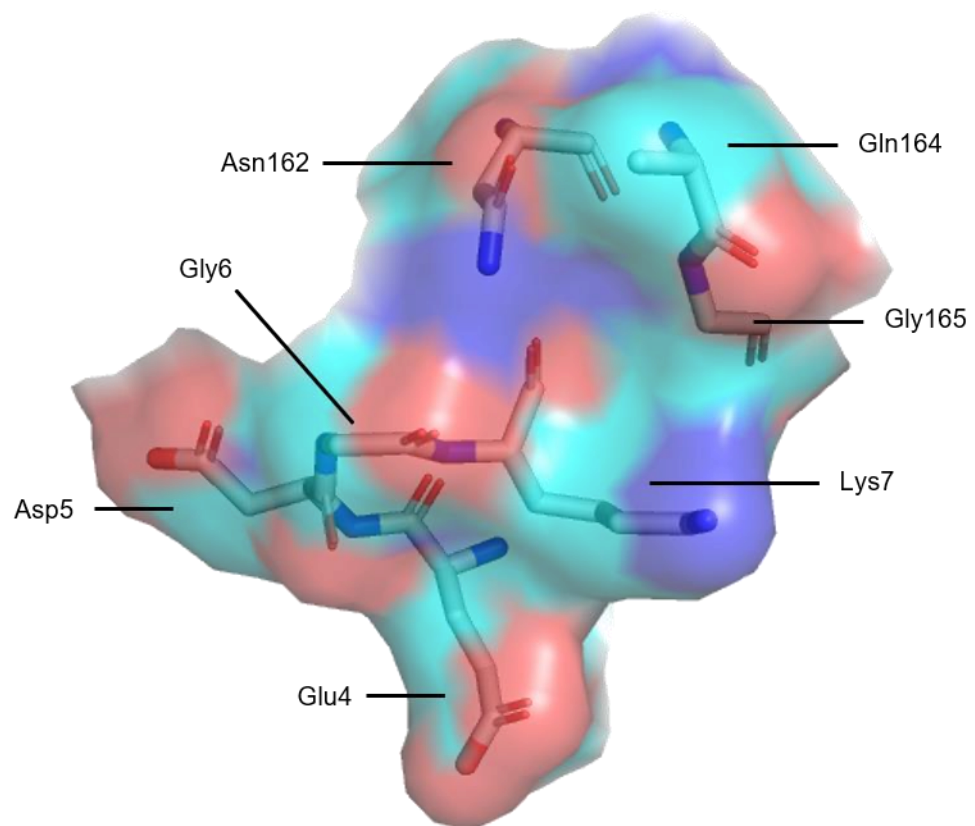
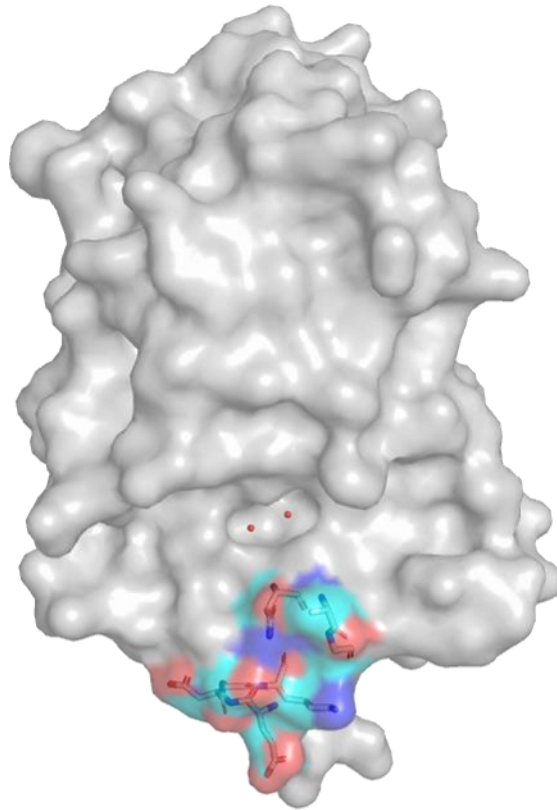
Back  $\alpha 6$  helix  
pocket



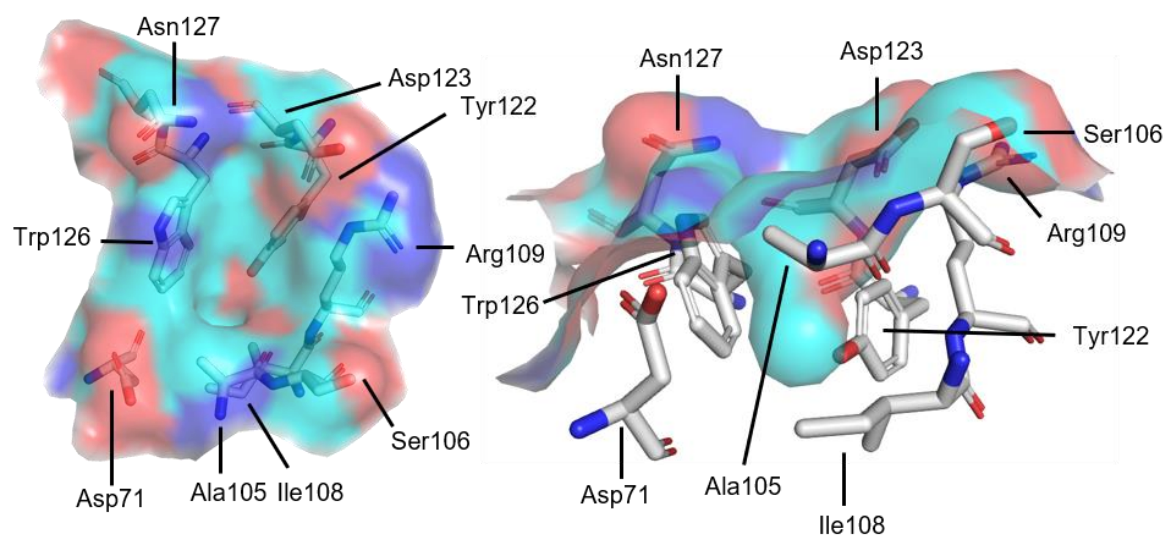
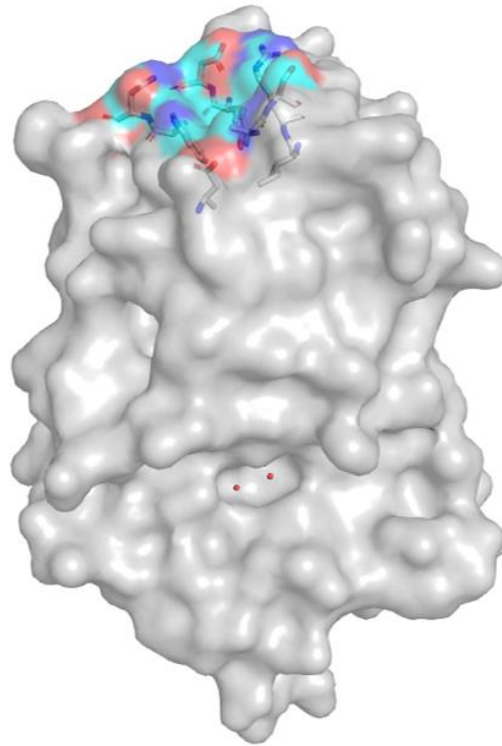
Active site adjacent



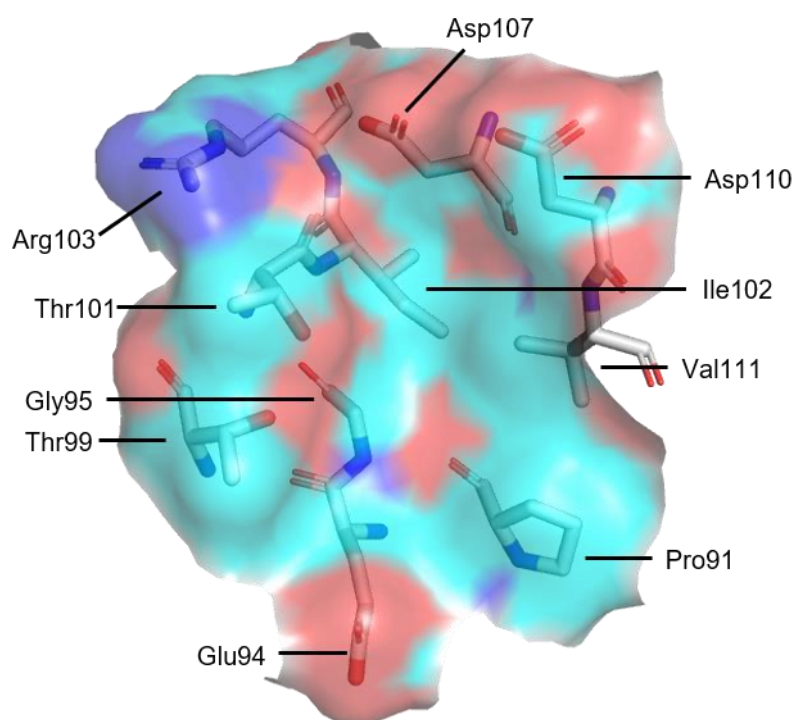
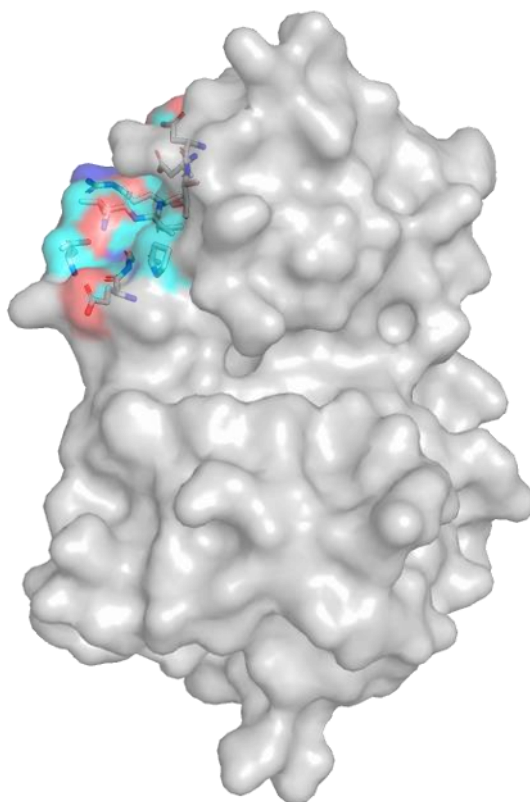
*N*-terminus front



Apex indent

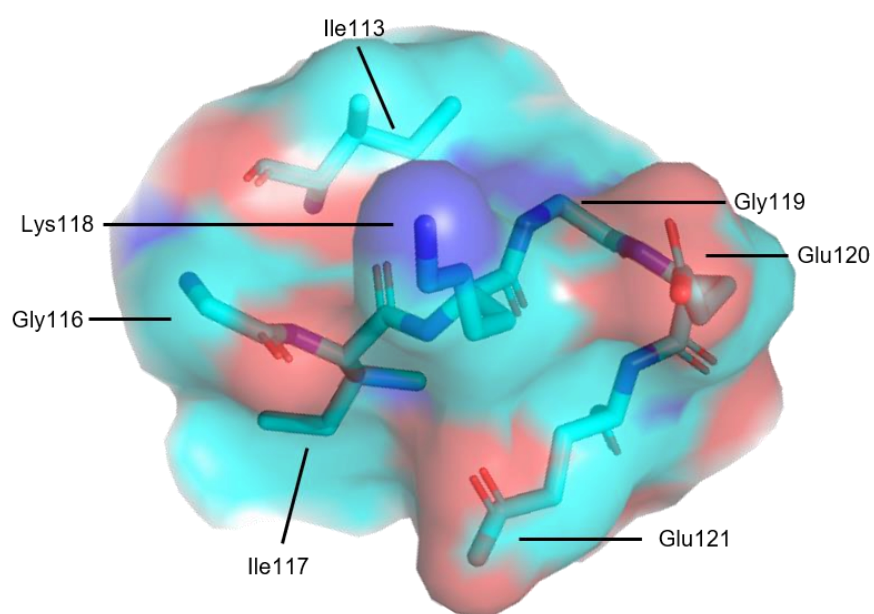
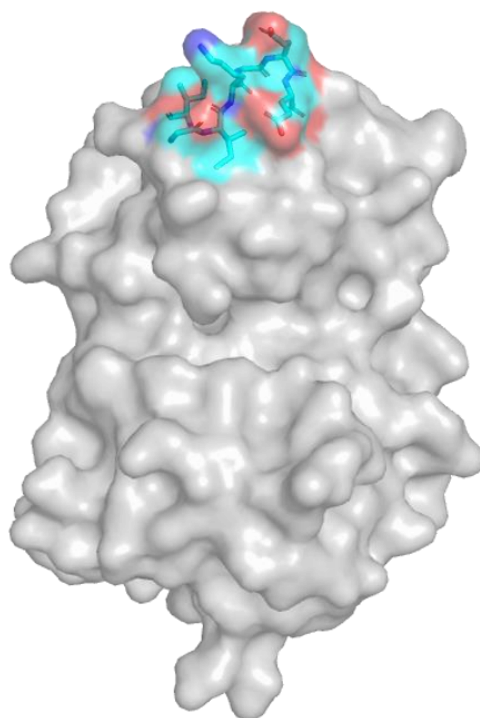


Back top left

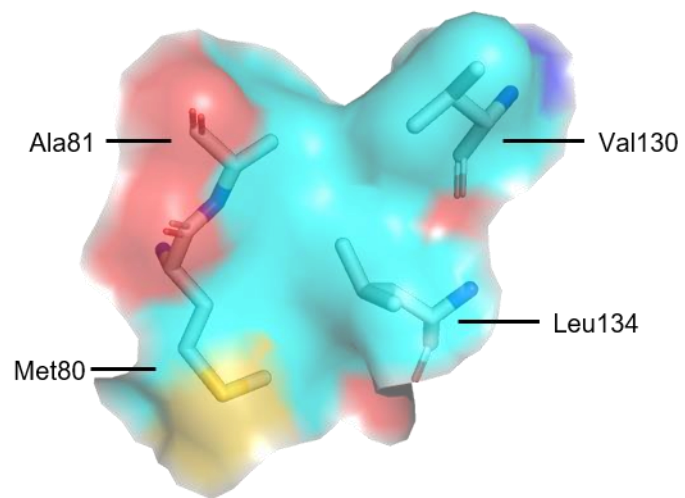
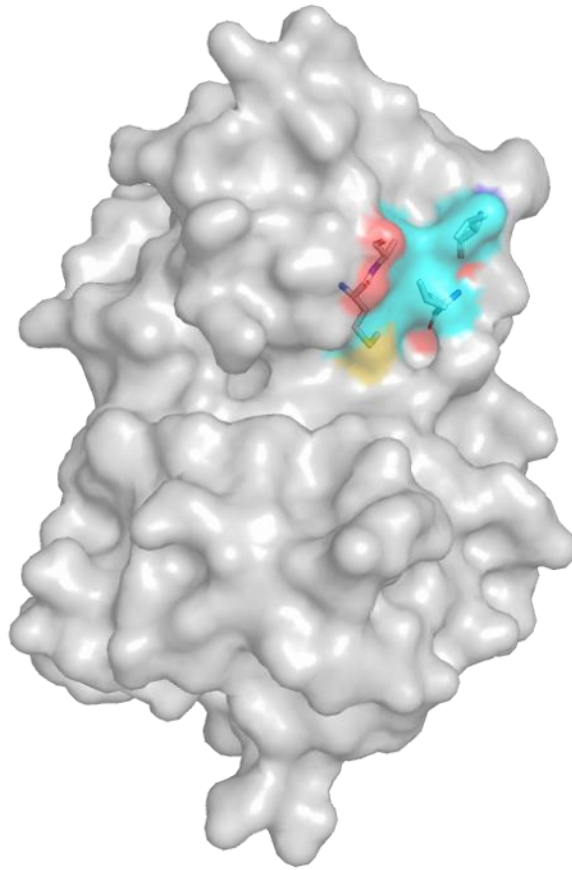




Back top middle



Back top right





#### Appendix 4: SMILES list of MicroFrag library

Compound number	SMILES
MF1	<chem>OC(=O)C1=NNC=C1</chem>
MF2	<chem>CN1C=CN=N1</chem>
MF3	<chem>NCCN1CCCC1</chem>
MF4	<chem>C1=NC=NC=N1</chem>
MF5	<chem>CN1CCN(C)CC1</chem>
MF6	<chem>N1C=CC=N1</chem>
MF7	<chem>OC1CCOC1</chem>
MF8	<chem>N#CC1=CSC=N1</chem>
MF9	<chem>Cl.NCCN1C=CN=C1</chem>
MF10	<chem>CN1C(=O)CCC1=O</chem>
MF11	<chem>CN1C=NC=N1</chem>
MF12	<chem>OC1=CC(Br)=CC=C1</chem>
MF13	<chem>CN1N=CC=C1N</chem>
MF14	<chem>C1=CN=CN=C1</chem>
MF15	<chem>O=C1COCCN1</chem>
MF16	<chem>CC1=NNN=N1</chem>
MF17	<chem>BrC1=CN=C1</chem>
MF18	<chem>CN1CCCCC1</chem>
MF19	<chem>NC1CCNCC1</chem>
MF20	<chem>Cl.NC1=NCCCN1</chem>
MF21	<chem>NC1=NC=CN1</chem>
MF22	<chem>OC(=O)C1=CN=CS1</chem>
MF23	<chem>OCCN1C=CC=N1</chem>
MF24	<chem>CN1CCC(O)C1</chem>
MF25	<chem>OC1=CN=CN=C1</chem>
MF26	<chem>O=C1NNC=C1</chem>
MF27	<chem>ClC1=NNC=C1</chem>
MF28	<chem>NC1=CC=C(Br)C=C1</chem>
MF29	<chem>C1=CN=CC=N1</chem>
MF30	<chem>CN1C=CC=N1</chem>
MF31	<chem>CC1=CC=CN=C1</chem>
MF32	<chem>C1CCNC1</chem>
MF33	<chem>N#CC1=COC=C1</chem>
MF34	<chem>CN1N=CC=C1CN</chem>

Compound number	SMILES
MF35	<chem>NC(=O)C1CCCO1</chem>
MF36	<chem>CN1C=NN=N1</chem>
MF37	<chem>CN1C=CC(CN)=N1</chem>
MF38	<chem>CC1=CNC=C1</chem>
MF39	<chem>OCC1=COC=C1</chem>
MF40	<chem>Cl.CC1CCCN1</chem>
MF41	<chem>Cl.OC(=O)C1CCNC1</chem>
MF42	<chem>Cl.Cl.CN1CCC(N)C1</chem>
MF43	<chem>Cl.Cl.NC1CCCNC1</chem>
MF44	<chem>OC1=CC=CN=C1</chem>
MF45	<chem>Cl.Cl.NC[C@H]1CCCN1</chem>
MF46	<chem>NC1=CC=NC(N)=N1</chem>
MF47	<chem>OC(=O)C1CCCN1</chem>
MF48	<chem>O=C1CCCN1</chem>
MF49	<chem>Cl.C1CNOC1</chem>
MF50	<chem>O=C1NCCCN1</chem>
MF51	<chem>OC(=O)C1=CNC=C1</chem>
MF52	<chem>NC1=CC=NC=N1</chem>
MF53	<chem>O=C1NC=NC=C1</chem>
MF54	<chem>NC1=NOC=C1</chem>
MF55	<chem>NC1=CC=NN1</chem>
MF56	<chem>NC1=CN=CN=C1</chem>
MF57	<chem>OC1CCNCC1</chem>
MF58	<chem>N1C=CN=C1</chem>
MF59	<chem>COC1=CC=CC=C1</chem>
MF60	<chem>BrC1=NC=CC=C1</chem>
MF61	<chem>C1CNCCN1</chem>
MF62	<chem>OCC1=CC=CC=C1</chem>
MF63	<chem>OC1CCCCC1</chem>
MF64	<chem>OCC1CCCO1</chem>
MF65	<chem>C1COCCN1</chem>
MF66	<chem>O=C1CCC(=O)N1</chem>
MF67	<chem>N1C=CN=N1</chem>
MF68	<chem>NC1=NC=CC=N1</chem>
MF69	<chem>NC1=NC=CC=C1</chem>

Compound number	SMILES
MF70	<chem>OC1=NC=CC=C1</chem>
MF71	<chem>CC1=NC=CC=C1</chem>
MF72	<chem>OC1CCCC1</chem>
MF73	<chem>C1CCNCC1</chem>
MF74	<chem>CNC1=CC=CC=C1</chem>
MF75	<chem>N1C=NC=N1</chem>
MF76	<chem>CC1CCNCC1</chem>
MF77	<chem>CN1CCNCC1</chem>
MF78	<chem>CCN1CCNCC1</chem>
MF79	<chem>CN1C=CN=C1</chem>
MF80	<chem>NC1=CC=CC=C1</chem>
MF81	<chem>CN1CCCC1=O</chem>
MF82	<chem>CN1CCOCC1</chem>
MF83	<chem>OC1=CC=CC=C1</chem>
MF84	<chem>C1CCOC1</chem>
MF85	<chem>NCC1=CC=CC=C1</chem>
MF86	<chem>C1COCCO1</chem>
MF87	<chem>C1=CC=NC=C1</chem>
MF88	<chem>NC1CCOC1</chem>
MF89	<chem>CC1=CC=CC(N)=C1</chem>
MF90	<chem>N1C=CC=C1</chem>
MF91	<chem>CN1C=C(N)C=N1</chem>

## Appendix 5: $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra of MicroFrag library

$^1\text{H}$ - $^{15}\text{N}$  HSQC spectra overlay of oxidised *EcDsbA* (100  $\mu\text{M}$ ) in the absence (blue) and presence (red) of a MicroFrag (300 mM). F1 =  $^{15}\text{N}$  nucleus, F2 =  $^1\text{H}$  nucleus

