

**Evaluating the biocatalytic potential
of a small molecule 2-oxoglutarate
dependent halogenase**

**Ellen Gallimore
Merton College, University of Oxford**

Trinity Term 2018



**A thesis submitted to the University of Oxford for the degree of
Doctor of Philosophy**

Evaluating the biocatalytic potential of a small molecule 2-oxoglutarate dependent halogenase

Ellen Gallimore

Merton College, University of Oxford

Submitted for the degree of Doctor of Philosophy

Trinity Term 2018

The late-stage oxidative functionalisation of C-H bonds can be used to circumvent synthetic route re-design and to generate molecules not easily accessible *via* conventional methods. Such reactions can be problematic due to a lack of site-selectivity and the poor intrinsic reactivity of the C-H bond. Enzymes may provide a solution to both these issues, and enhancements in activity, selectivity, temperature stability and organic solvent tolerance over the wild-type enzyme are achievable by protein engineering.

The stereoselective chlorination of unactivated hydrocarbons is challenging from a synthetic perspective, yet this type of reaction is performed in nature by enzymes of the iron and 2-oxoglutarate (2OG)-dependent halogenase family. By utilising a radical mechanism, these enzymes remove the requirement of existing chlorination biocatalysts for inherent substrate activation (in the form of aromatic or alkene functionalities), and demonstrate the potential for the evolution of new and interesting C-H functionalisation capabilities. Historically, enzymes from the iron and 2OG-dependent halogenase family were unsuitable for biocatalysis due to the necessity of a substrate-bound carrier protein for activity. Identified in 2014, WelO5 was the first enzyme from this family shown to be capable of carrier protein-independent chlorination.

This thesis has evaluated the tractability of developing synthetically useful biocatalysts based on the 2OG-dependent halogenase WelO5, with the aim of adding new reaction capabilities to the synthetic toolbox of the future. To achieve this, WelO5 was produced in high yields and purity for characterisation and crystallisation trials. Three novel WelO5 crystal structures were determined, which aided the rational selection of active site residues for replacement. The natural substrate for WelO5, (+)-12-*epi*-fischerindole U isonitrile, was synthetically produced and used for the development of a screening assay for assessing WelO5 variant activity against substrate-like analogues. A total of 40 WelO5 variants were produced and tested for activity towards the natural substrate and structurally similar analogues. Variant I161A was found to introduce a new hydroxylation activity to the enzyme. Activity was also seen for WelO5 and variants against two structurally unrelated compounds.

Overall, the generated variants of WelO5 have shown promise as biocatalysts, with the main limitation to further progress being the throughput of the available screening methods. Novel activities have been discovered which merit further investigation into small molecule 2OG-halogenases as biocatalysts and this thesis provides the tools with which to do so.

Acknowledgements

First and foremost, I need to thank UCB for giving me the opportunity to carry out a PhD whilst under their employment. I am extremely grateful to the UCB Fellows for supporting the PhD scheme and to Rikki Alexander for providing me with the encouragement and support to apply - I wouldn't be here without your guidance.

I would like to thank Professor Chris Schofield, Dan Brookings and Dave McMillan for their supervision; also, Mike McDonough and Alistair Henry for providing their supervisory guidance in an unofficial capacity. A special mention must also go to Jürgen Brem for all the help and advice he provided when I first joined the Schofield group and for his unwavering attempts at trying to get me to sit up straight.

To the past and present members of the Schofield group, most of whom have helped me in some capacity over the past four years - I have really enjoyed working with such an interesting, fun and diverse group of people and I hope we cross paths again one way or another in the future.

Many of my colleagues at UCB deserve specific mention for their scientific help, including (in no particular order): Prash Mori (protein expression & purification), Neesha Dedi (PhyNexus system), Rob Griffin (expression of labelled proteins for NMR), Sebastian Kelm (homology modelling), Rachel Davis (protein MS), Justin Staniforth & Sarah Taylor (preparative SFC and HPLC), Adam Hold & Victoria Ellis (HRMS), Christine Prosser (protein NMR) and Harry Mackenzie (small-molecule NMR).

I would also like to thank the members of L102 and L242b for tolerating me when my work wasn't going according to plan; and to the wider UCB chemistry department for taking such an interest in my work, as well as the kind enquiries about my write-up progress and the (occasionally bizarre) conversations and free cake.

I am very extremely thankful to Jenny, Laura and Jo for sending the generous fruit basket during my write-up period. The fruit-based puns weren't berry good but it was lovely to know you were thinking of me. Now I'm not writing all hours of the day we have lots to catch up on!

Finally, a special mention must go to Pete. I can't capture here how much you mean to me and how much I have appreciated your help in all things scientific and non-scientific. I am so grateful for your support and patience and, of course, your proof-reading skills. Let's get started on those peanut butter M&Ms!

Contents

Abbreviations	viii
Chapter 1. Introduction	1
1.1. Biocatalysis	1
1.1.1. Enzyme engineering	1
1.1.2. Advantages & challenges	2
1.1.3. Industrial applications	4
1.2. Enzymatic halogenation.....	6
1.2.1. Nucleophilic halogenases	6
1.2.2. Electrophilic halogenases	7
1.2.2.1. Haloperoxidases.....	7
1.2.2.2. Flavin-dependent halogenases.....	9
1.2.3. Radical halogenases.....	9
1.2.4. Biocatalytic applications of halogenase enzymes.....	11
1.3. The Fe ²⁺ & 2OG-dependent oxygenase family.....	14
1.3.1. The consensus mechanism for 2OG-oxygenases	16
1.3.2. Small-molecule Fe ²⁺ & 2OG-dependent oxygenases	19
1.3.3. Fe ²⁺ & 2OG-dependent halogenases.....	22
1.3.3.1. Discovery of WelO5 and AmbO5.....	25
1.4. Aims & objectives	29
Chapter 2. Production of WelO5 & related proteins	31
2.1. Introduction.....	31
2.2. Bioinformatics & biocatalysis.....	32
2.2.1. WelO5 primary sequence analysis.....	33
2.2.1.1. Identification of similar enzymes	33
2.2.1.2. Comparison of WelO5 to enzymes of known function	40
2.2.2. WelO5 tertiary structure prediction.....	42
2.3. WelO5 cloning, expression & purification.....	45
2.3.1. Construction of WelO5 recombinant expression vector	45
2.3.2. WelO5 expression.....	46
2.3.2.1. Expression trials	46
2.3.2.2. Large scale expression of WelO5	48
2.3.3. WelO5 purification.....	49
2.3.3.1. Immobilized Metal Ion Affinity Chromatography (IMAC)	50
2.3.3.2. Cleavage with the TEV protease	52

2.3.3.3. Size-Exclusion Chromatography (SEC)	53
2.3.3.4. Condensed purification protocol	54
2.4. Cloning & expression of WelO5 related proteins	59
2.5. Summary of the production of WelO5 & related proteins.....	67
Chapter 3. Synthesis of WelO5 endogenous substrate and substrate analogues	68
3.1. Introduction.....	68
3.2. Synthesis of endogenous substrate.....	71
3.2.1. Step 1: Indole-carvone coupling.....	76
3.2.2. Step 2: Alkene installation	83
3.2.3. Step 3: Cyclisation.....	85
3.2.4. Step 4: Reductive amination	87
3.2.5. Step 5: Isonitrile formation.....	90
3.3. Synthesis of simplified substrate analogues	94
3.4. Initial activity screening.....	95
3.5. Synthetic summary for WelO5 endogenous substrate and analogues	103
Chapter 4. Characterisation of WelO5	105
4.1. Introduction.....	105
4.2. Initial WelO5 characterisation	106
4.2.1. Molecular weight	106
4.2.2. Oligomerisation	107
4.2.2. Circular dichroism	109
4.2.3. UV-Vis spectroscopy.....	110
4.3. Structural characterisation of WelO5	112
4.3.1. X-ray crystallography	112
4.3.1.1. ICP-MS	112
4.3.1.2. Thermal shift assay (metal-binding).....	113
4.3.1.3. Thermal shift assay (ligand-binding).....	115
4.3.1.4. WelO5 structure determination.....	117
4.3.1.5. Observations from WelO5 crystal structures	123
4.3.1.5.1. Overall structure	123
4.3.1.5.2. WelO5 active site.....	130
4.3.1.5.3. Literature structure comparison.....	135
4.3.2. Preliminary NMR studies	140
4.4. Summary of WelO5 characterisation studies	144
Chapter 5. Enzyme evolution & results	146
5.1. Introduction.....	146

5.2. Mutagenesis strategy	149
5.3. Screening protocol.....	153
5.3.1. Screening protocol requirements	153
5.3.2. Method development & optimisation	154
5.3.2.1. Expression studies.....	155
5.3.2.2. Protein purification	158
5.3.2.3. Assay set-up.....	160
5.3.2.4. Analysis.....	163
5.3.3. Finalising the testing protocol	164
5.4. Design of WelO5 variants	165
5.5. Substitution of WelO5 active site residues	168
5.6. Alternative substrates	180
5.6.1. (+)-12-epi-fischerindole U isonitrile ‘substrate-like’ compounds	180
5.6.2. New substrate choices.....	183
5.7. Summary of WelO5 evolution	186
Chapter 6: Summary & future work.....	188
Chapter 7. Materials & methods.....	194
7.1. Bioinformatics	194
7.1.1. Homology modelling	194
7.2. Biological methods	194
7.2.1. Solutions & media	194
7.2.2. Gel electrophoresis	195
7.2.2.1. Agarose (DNA).....	195
7.2.2.2. SDS-PAGE (protein).....	196
7.2.3. Transformations.....	196
7.2.4. Starter cultures	197
7.2.5. Glycerol stocks	197
7.2.6. Plasmid DNA stocks	197
7.2.7. DNA quantification	197
7.2.8. Ligation Independent Cloning (LIC)	197
7.2.8.1. Synthetic genes	197
7.2.8.2. Expression vectors.....	198
7.2.8.3. Primers.....	198
7.2.8.4. Protocol.....	198
7.2.9. OPPF cloning & expression.....	201
7.2.9.1. Synthetic genes	201
7.2.9.2. Expression vectors.....	204

7.2.9.3. Primers.....	204
7.2.9.4. Protocol.....	205
7.2.10. Site-directed mutagenesis.....	210
7.2.11. Sequencing.....	212
7.2.12. Protein expression.....	212
7.2.12.1. Expression strains.....	212
7.2.12.2. Cell density determination.....	212
7.2.12.3. Expression trials.....	212
7.2.12.4. Large scale protein expression.....	213
7.2.12.5. Batch method (12L scale).....	213
7.2.12.6. Fed-batch method (5L scale).....	213
7.2.12.7. Expression of selenomethionine WelO5 (2.4L scale).....	214
7.2.12.8. Expression of isotopically labelled WelO5.....	215
7.2.12.8.1. ¹⁵ N, ² H-labelled WelO5 (2L scale).....	215
7.2.12.8.2. ¹⁵ N, ¹³ C, ² H-labeled WelO5 (4L scale).....	217
7.2.13. Protein purification.....	218
7.2.13.1. Batch purification.....	218
7.2.13.1.1. Nickel-Affinity Chromatography (IMAC).....	218
7.2.13.1.2. Size-Exclusion Chromatography (SEC).....	219
7.2.13.1.3. Condensed purification protocol.....	219
7.2.14. Protein quantification.....	220
7.3. Chemical synthesis.....	221
7.3.1. Synthesis of (+)-12-epi-fischerindole U isonitrile and analogues.....	222
7.3.2. Simplified substrate analogues.....	233
7.3.3. Re-synthesis of Thermofluor assay hits.....	238
7.4. WelO5 assays & characterisation.....	242
7.4.1. Mass spectrometry.....	242
7.4.1.1. LC-MS.....	242
7.4.1.2. SEC-MALS.....	242
7.4.1.3. ICP-MS.....	242
7.4.2. Circular dichroism (CD).....	243
7.4.3. UV-Vis.....	243
7.4.4. X-Ray crystallography.....	244
7.4.4.1. Crystallisation.....	244
7.4.4.2. Data collection and structure determination.....	245
7.4.4.3. Crystallisation, data collection and refinement statistics.....	247
7.4.5. Thermal shift assays.....	248
7.4.6. NMR.....	249
7.4.7. Activity screening.....	250

7.4.7.1. Initial WelO5 activity screening protocol	250
7.4.8. Finalised WelO5 variant screening	251
7.4.7.3. UPLC Analysis.....	252
7.4.7.3.1. UPLC method gradient: turnover of WelO5 natural substrate	254
7.4.3.7.2. UPLC method gradient: alternative substrates	254
Chapter 8. References	255

Abbreviations

2OG	2-oxoglutarate
2xTY	2x tryptone yeast extract
API	Active pharmaceutical ingredient
BME	β -mercaptoethanol
BSA	Bovine serum albumin
CAST	Combinatorial active-site saturation test
CD	Circular dichroism
CDI	Carbonyldiimidazole
CDMT	2-chloro-4,6-dimethoxy-1,3,5-triazine
COSY	Correlation spectroscopy
C-terminus	Carboxy terminus
CV	Column volume
Da	Dalton
DCM	Dichloromethane
DIPEA	<i>N,N'</i> -diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMSO	Dimethylsulfoxide
DMT-MM	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride
dNTP	Deoxyribonucleotide triphosphate
DPPA	Diphenylphosphoryl azide
DSBH	Double-stranded beta-helix
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
<i>e.e.</i>	Enantiomeric excess
ESI	Electrospray ionisation
ESRF	European Synchrotron Radiation Facility
FPLC	Fast protein liquid chromatography
FT	Flow-through

GFP	Green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His₆	Hexa-histidine tag
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum coherence
I3C	5-amino-2,4,6-triiodoisophthalic acid
ICP-MS	Inductively coupled plasma mass spectrometry
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISM	Iterative scanning mutagenesis
LB	Luria broth
LC-MS	Liquid chromatography-mass spectrometry
LIC	Ligation independent cloning
LiHMDS	Lithium hexamethyldisilazane
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MLCT	Metal-ligand charge transfer
MOPS	(3-(<i>N</i> -morpholino)propanesulfonic acid)
MS	Mass spectrometry
NCBI	National Centre for Biotechnology Information
NMM	<i>N</i> -methyl morpholine
NMR	Nuclear magnetic resonance
NOG	<i>N</i> -oxalyl glycine
<i>N</i>-terminus	Amino terminal
OD	Optical density
OPPF	Oxford Protein Production Facility
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
Ph	Phenyl
PMSF	Phenylmethylsulfonyl fluoride

ppm	Parts per million
RCSB	Research Collaboratory for Structural Bioinformatics
R_f	Retention factor
RT	Retention time
SAM	(<i>S</i>)-adenosyl-L-methionine
SAR	Structure-activity relationship
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SEC-MALS	Size-exclusion chromatography-multi-angle light scattering
SeMet	Selenomethionine (protein)
SFC	Supercritical fluid chromatography
SGC	Structural Genomics Consortium
SOC	Super optimal broth with catabolite repression
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
TOCSY	Total correlation spectroscopy
Tris	Tris-(hydroxymethyl)aminomethane
UPLC	Ultra-performance liquid chromatography
UV	Ultraviolet
wt	Wild-type

Chapter 1. Introduction

1.1. Biocatalysis

The use of enzymes to perform chemical processes in the production and manufacture of fine chemicals and pharmaceuticals has had major successes, but still has enormous untapped potential. Indeed, the field of biocatalysis has rapidly grown in the past decade as chemistry looks for new methodologies which can offer advantages over existing chemical procedures in terms of chemoselectivity, regioselectivity, stereoselectivity, C-H activation capabilities and greener chemistry.¹

1.1.1. Enzyme engineering

Enzymes are very efficient catalysts, often capable of achieving turnovers far superior to that of non-biological chemical catalysts. However, their excellent catalytic activity and selectivity often exists against a singular substrate or substrate type and tends to come at the price of low promiscuity. Chemical methods are often much more generally applicable, especially in research synthesis. By making modifications at the amino acid sequence level, enzymes can be made more amenable to the production of compounds desired by synthetic chemists. Which substitutions are made very much depends upon the property being evolved for and the availability of structural and mechanistic knowledge of the enzyme. In addition to activity against non-native substrates, examples of evolvable enzyme properties include thermostability,² selectivity,^{3,4} non-native substrate activity⁵ and organic solvent tolerance.^{6,7}

Active-site modifications are most commonly used to alter reaction selectivity or to develop non-native substrate activities.⁸ If available, crystal structures or

homology models can be exploited to guide the rational re-design of an active site through the substitution of key residues.^{9,10} However, crystal structures only provide a static picture and the success rate for substitutions at the active site can be difficult to predict, particularly when attempting to design synergistic combinations of active site substitutions.

Directed evolution aims to overcome this limitation by experimentally mimicking the natural evolutionary process. Iterative rounds of diversification and screening/selection are used to bias enzyme evolution towards the enhancement of a given enzyme property.^{11,12}

Directed evolution can be used to broaden substrate scope, enhance stereoselectivity and to optimise enzyme characteristics such as solvent tolerance and thermostability, which are typically a more significant challenge to improve by rational design. The directed evolution of an entire enzyme sequence does not require knowledge of enzyme structure or mechanism. However, due to the large numbers involved, a high-throughput screening or selection method is essential.¹³

1.1.2. Advantages & challenges

The chemical industry has a substantial interest in processes which are both economically and environmentally sustainable and the adoption and implementation of new technologies can be highly influenced by how 'green' they are determined to be. Biocatalysis is widely regarded as a 'green' technology due to the minimal use of hazardous and 'non-green' solvents and reagents, energy efficiency (ambient temperature and pressure) and employing catalytic rather than stoichiometric reagents.^{14,15}

Biocatalytic processes can be superior to chemical methods in terms of specificity and selectivity but significant resource (i.e. directed evolution) is often required to achieve these properties on non-natural substrates. However, the ability to produce large quantities of a fine chemical or pharmaceutical intermediate often justifies this investment, particularly if no practical alternative synthetic route exists for synthesis on scale.¹¹

Enzymes typically demonstrate their greatest catalytic power in aqueous media, unfortunately this is not conducive to performing reactions on organic compounds which typically demonstrate poor solubility in water. The addition of co-solvents to the reaction such as IPA, DMSO or methanol can help with substrate solubility.¹⁶ Alternatively, reactions have been successfully performed in biphasic solvent systems by adding water immiscible solvents such as toluene, heptane or ethyl acetate to the aqueous reaction mixture.^{17,18}

Enzymes generally operate within pH 5-8 and 20-40 °C. Whilst the ability to carry out reactions at neutral pH, ambient temperature and low pressure is advantageous (particularly on scale), it does mean that enzymatic reactions have a narrow window for optimisation.¹⁹ Although exceptions do exist.²⁰⁻²³

Reactions with cofactor-dependent enzymes require supplementation with stoichiometric cofactor, which can be prohibitively expensive on scale (depending on the cofactor). In some cases, cofactors can be regenerated *in situ* - although this is not applicable to every cofactor.^{24,25}

An important area of biocatalysis research involves the identification of enzymes which can perform reactions not possible using existing synthetic

methods. Such reactions can introduce new disconnection approaches and show promise for addition to the synthetic chemistry toolbox of the future.²⁶

1.1.3. Industrial applications

Biocatalytic processes are already employed in industry for the synthesis of fine chemicals, pharmaceuticals and APIs (active pharmaceutical intermediates).²⁷⁻³¹ Perhaps one of the most highly cited examples of the application of biocatalysis to an industrial process is the Codexis developed synthesis of Sitagliptin, the API in Januvia®.¹⁶ Guided by dockings of the substrate into a transaminase homology model and using their proprietary directed evolution platform, Codexis developed an enzymatic process which circumvented a high pressure, rhodium metal-catalysed hydrogenation and achieved superior enantioselectivity (>99.95% *e.e.*) (comparisons with the chemocatalytic process are summarised in Figure 1.1.).^{16,32}

Chemo-catalytic route:

- High pressure hydrogenation
- Rh catalysed (cost & removal)
- Crystallisation required for improvement of chiral purity

Biocatalytic route:

- Hydrogenation eliminated (cost & safety)
- Very high selectivity (> 99.9% e.e.)
- Improved overall yield
- Reduced waste
- Overall reduction in manufacturing cost

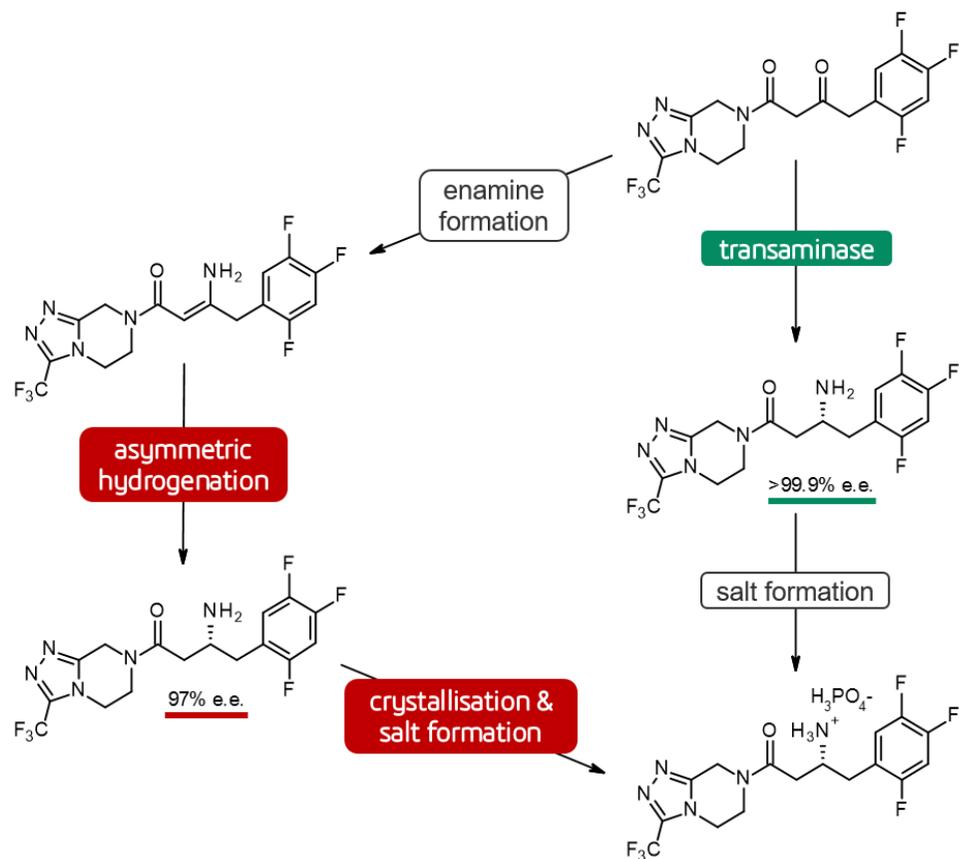


Figure 1.1. A summary of the advantages of the Codexis transaminase-catalysed route to Sitagliptin, over the optimised Merck chemo-catalytic route. The chemo-catalytic route employed a rhodium-catalysed asymmetric hydrogenation step at 250 psi to achieve the desired product in 97% e.e., with a 0.15 mol% catalyst loading. Subsequent treatment with activated carbon was required to remove the residual rhodium, followed by a crystallisation step to achieve a chiral purity of 99.9% e.e. In contrast, the biocatalytic route obtained optically pure material (undesired enantiomer was not detected) at 50% DMSO at 50 °C.

1.2. Enzymatic halogenation

Nature employs several different approaches to install halogens, mostly *via* oxidative mechanisms.³³ Chlorine and bromine are the most prevalent halogens in natural products. Iodination is rare as a direct result of the low natural abundance of this element.³⁴ Due to the high electronegativity of fluorine, reactions proceeding *via* an oxidative mechanism would have to overcome a significant energy barrier and currently no enzymes are known to be able to perform this process.³⁵ Instead, the introduction of enzymatic fluorine has been shown to proceed *via* a nucleophilic mechanism.³⁶ The approaches taken by nature to introduce halogens by nucleophilic, electrophilic and radical mechanisms are discussed.

1.2.1. Nucleophilic halogenases

A single characterised example exists for enzymatic fluorination - the fluorinase from the bacterium *Streptomyces cattleya*.³⁶ This unique enzyme has evolved the ability to use hydrogen bonding to sufficiently desolvate the fluoride and perform nucleophilic displacement under aqueous conditions.³⁵ Fluoride attacks the C-5' of (*S*)-adenosyl-L-methionine (SAM), displacing methionine to generate 5'-fluoro-5'-deoxyadenosine (Figure 1.2.). This enzyme has been shown to accept and use chloride instead of fluoride (albeit at significantly reduced activity), but not bromide.^{36,37}

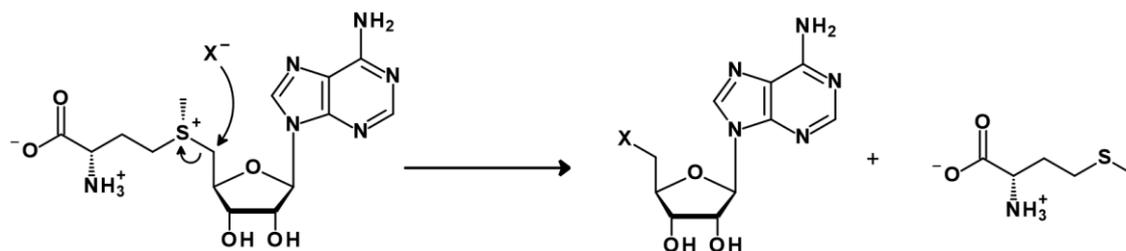


Figure 1.2. Nucleophilic mechanism employed by the fluorinase from *Streptomyces cattleya*. *X* is representative of fluoride and chloride.

The nucleophilic chlorinase from *Salinispora tropica* (SalL) also mediates an analogous S_N2 reaction with SAM.³⁸ This enzyme shares 35% identity with the fluorinase, but they differ in the key residues responsible for co-ordinating their halide ions. For SalL, chloride has been successfully replaced with bromide and iodide, but not fluoride.³⁸

1.2.2. Electrophilic halogenases

Electrophilic halogenases use an effective 'X⁺' equivalent (hypohalous acid; HO-X) for halogenation (Figure 1.3).³⁹ These enzymes are categorised into haloperoxidase and flavin-dependent halogenase subfamilies depending on how the hypohalous acid is generated. The nature of the 'X⁺' equivalent means that electrophilic halogenases are restricted to electron-rich substrates (i.e. possessing aromatic or double-bond functionalities).⁴⁰

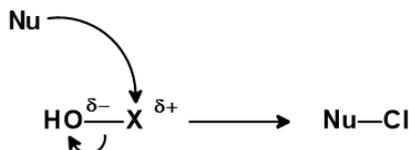


Figure 1.3. Generic outline of the hypohalous acid halogenation mechanism.

1.2.2.1. Haloperoxidases

In haem-dependent haloperoxidases, the porphyrin ligand co-ordinates to a ferric iron (i.e. Fe³⁺) held within the enzyme active site by a cysteine residue (*cf.* cytochrome P450 oxidases).⁴⁰ Hydrogen peroxide binds to the resting ferric (Fe³⁺) state of the haem group, forming a ferryl-oxo Fe(IV)=O species. This species then oxidises the halide ion to generate hypohalous acid (Figure 1.4.).³⁹ Chlorination, bromination and iodination have been observed for haem-dependent haloperoxidases.³⁹

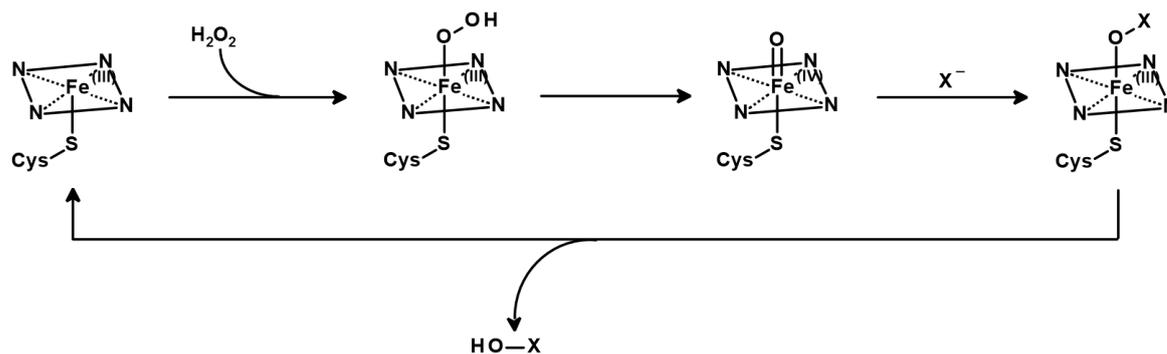


Figure 1.4. Halogenation mechanism employed by haem-dependent haloperoxidases for generation of hypohalous acid.

In the vanadium haloperoxidases, a trigonal bipyramidal vanadium is anchored into the active site by a histidine.⁴⁰ Hydrogen peroxide binds to the vanadium, followed by halide addition and the release of the hypohalous acid (Figure 1.5).^{40,41} In nature, bromination reactions dominate for vanadium haloperoxidases, however, a singular example of a vanadium chloroperoxidase has been found in the fungus *Curvularia inaequalis*.⁴²

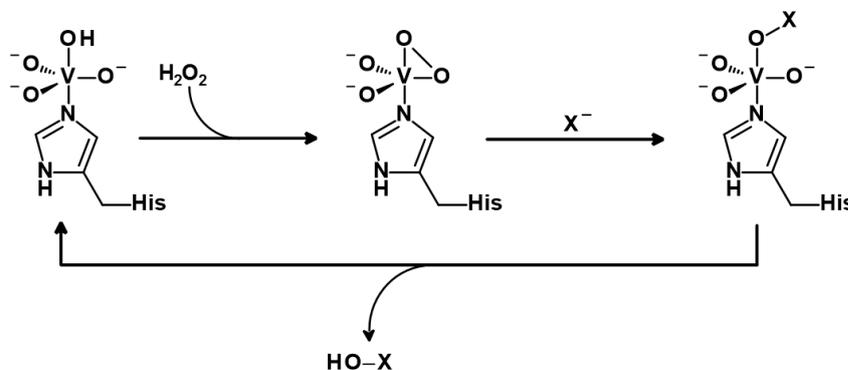


Figure 1.5. Mechanism employed by vanadium haloperoxidases for generation of hypohalous acid.

Whilst haloperoxidase enzymes are responsible for the formation of hypohalous acid, subsequent halogenation of the substrate is not enzyme mediated, resulting in poor regioselectivities.³⁹

1.2.2.2. Flavin-dependent halogenases

From a biocatalytic perspective, the flavin-dependent halogenases have been the most widely investigated for their ability to introduce bromine and chlorine atoms. This is due to their broader substrate scope and their improved regioselectivities relative to the haloperoxidases.^{43,44}

Flavin-dependent halogenases use molecular oxygen as a co-substrate and FADH₂ as a cofactor.⁴⁰ Reaction of the cofactor with molecular oxygen generates a peroxide intermediate which then forms HO-X through reaction with a halogen (Figure 1.6.). The addition of a flavin reductase enzyme is required for such reactions to recycle the flavin cofactor back to FADH₂.⁴⁰

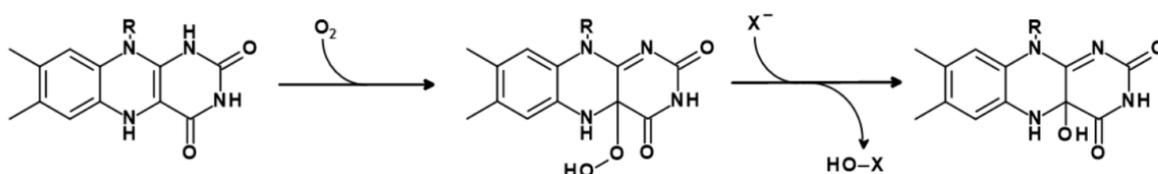


Figure 1.6. Mechanism employed by flavin-dependent halogenases.

Structural studies into the flavin-dependent tryptophan halogenase PrnA proposed that the generated hypohalous acid does not leave the enzyme, and is instead guided through a tunnel towards the substrate.⁴⁵ Similar mechanisms may help to rationalise the improved regioselectivities seen for flavin-dependent halogenases over the haloperoxidases.⁴⁶

1.2.3. Radical halogenases

Also known as Fe²⁺ and 2-oxoglutarate (2OG)-dependent halogenases, these radical halogenases belong to the Fe²⁺ and 2OG-dependent oxygenase superfamily, more detailed discussion on which can be found in Section 1.3.

The 2OG-dependent halogenases are proposed to employ a radical mechanism, using 2OG and molecular oxygen as co-substrates (Figure 1.7).⁴⁷ The radical mechanism allows enzymes from this family to halogenate un-activated, aliphatic carbons. These halogens are also introduced regio- and stereospecifically.⁴⁷ In nature, 2OG-dependent halogenases use chlorine. However, it has been shown that activity with bromine is also achievable.^{48,49}

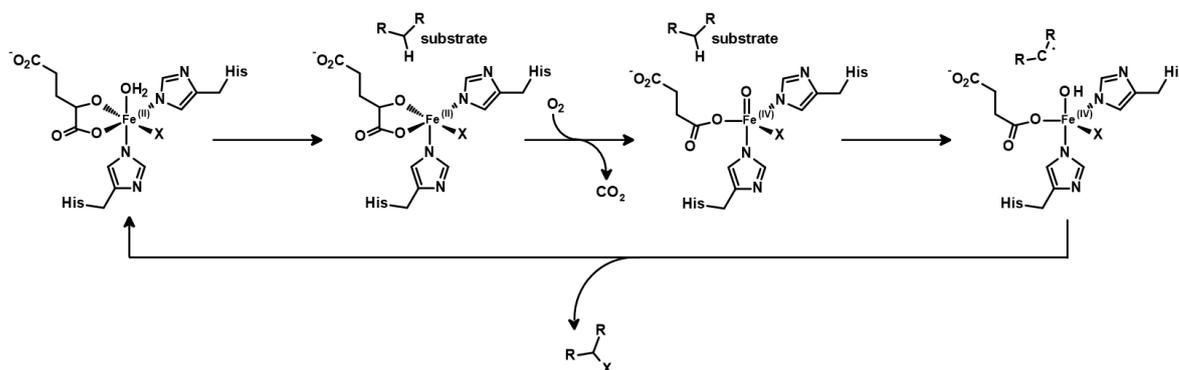


Figure 1.7. Halogenation mechanism employed by Fe^{2+} and 2OG-dependent halogenases.

Historically, 2OG-dependent halogenases have only been able to chlorinate carrier protein-tethered substrates (Figure 1.8.), limiting investigations into these enzymes as biocatalysts.⁴⁷

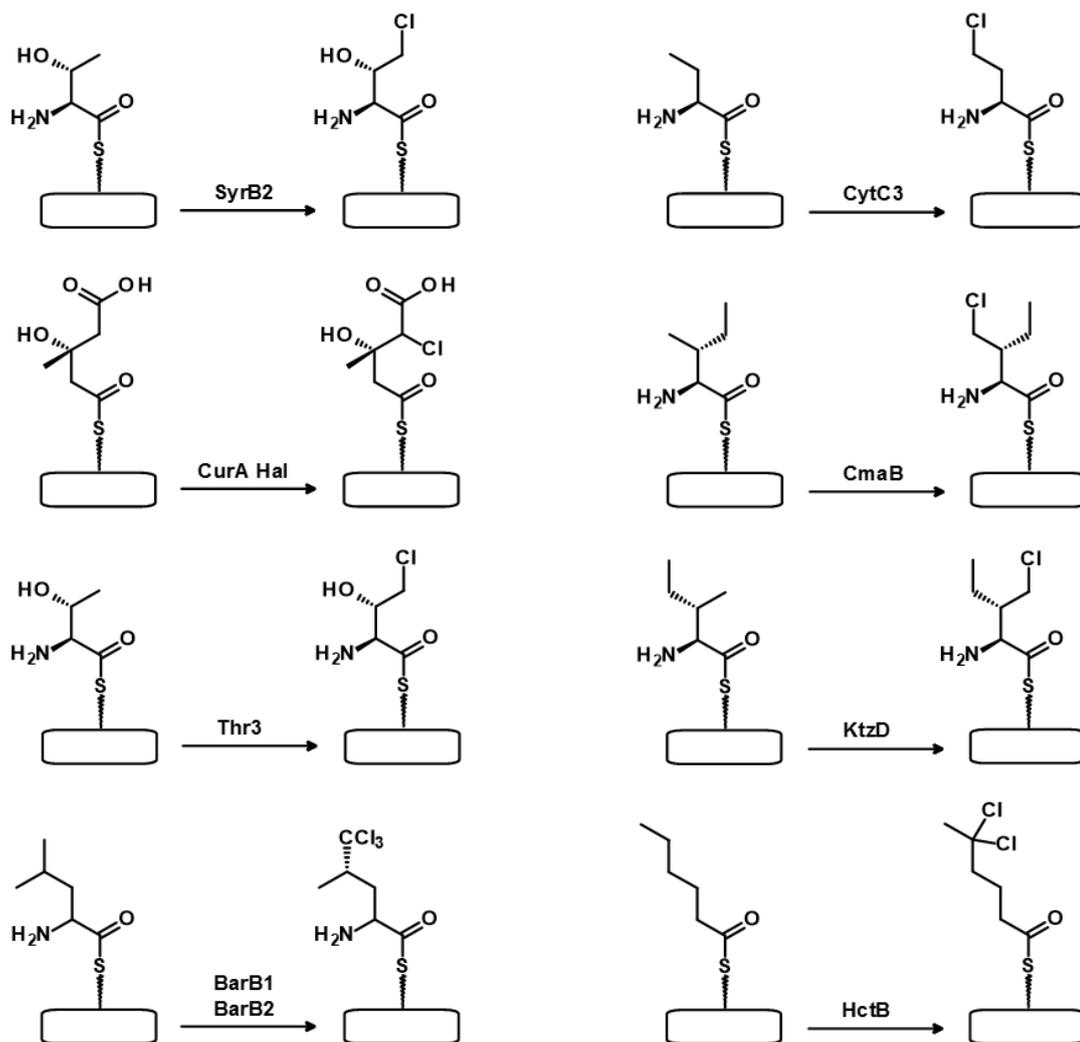


Figure 1.8. Summary of the substrates of known 2OG-halogenases. All substrates are tethered to a carrier-protein (represented by a rectangle) and are small and amino-acid like in structure.

The recent discoveries of WelO5⁵⁰ & AmbO5,⁵¹ 2OG-dependent halogenases which do not require carrier protein pre-functionalisation for activity, show promise for the engineering of 2OG-dependent halogenases as biocatalysts. WelO5 & AmbO5 will be covered in more detail in Section 1.3.3.1.

1.2.4. Biocatalytic applications of halogenase enzymes

It has been shown that halogens can be introduced to small molecules *via* a variety of mechanisms, however, not all are practical from a biocatalytic perspective. By virtue of their substrate scope and regioselectivity, the majority

of enzymes investigated for biocatalytic halogenation to date have been from the flavin-dependent halogenase enzyme family. The mechanism of these enzymes employs electrophilic chlorine, meaning that only electron rich substrates (containing aromatic or double bond functionalities) have been investigated for halogenation activity.⁵²⁻⁵⁶

Discovery of the carrier-protein independent 2OG-halogenases, WelO5 and AmbO5, has opened up the possibility of developing biocatalysts which operate through a different mechanism, expanding the substrate scope of biocatalytic halogenases to include unactivated, aliphatic substrates (highlighted in green, Figure 1.9).

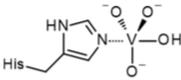
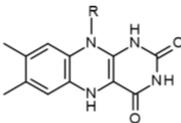
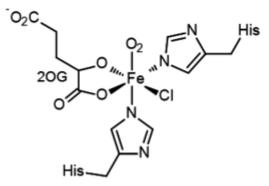
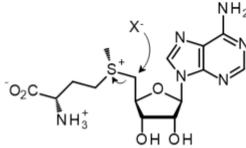
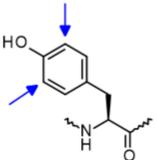
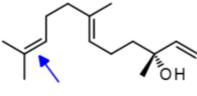
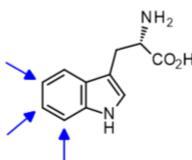
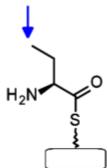
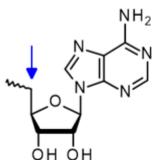
	 Iron & haem-dependent	 Vanadium dependent	 Flavin-dependent	 Iron & 2OG-dependent	 Nucleophilic
Oxidant	H ₂ O ₂	H ₂ O ₂	O ₂	O ₂	N/A
Halogen Species	X⁺	X⁺	X⁺	X[•]	X⁻
Substrate Requirements	aromatic electron rich	activated electron rich	aromatic electron rich	aliphatic un-activated	electrophilic good leaving group (SAM)
Product Examples					

Figure 1.9. A summary of nature's halogenation enzymes. Electrophilic halogenation on electron rich substrates dominates. Nucleophilic halogenases are highly-substrate specific, working only on SAM as a substrate. Iron and 2OG-dependent halogenases (highlighted by the green box) employ a radical mechanism, enabling halogenation of aliphatic and un-activated substrates. Blue arrows indicate the positions where halogens have been introduced. References for product examples: iron & haem-dependent,⁵⁷ vanadium dependent,⁵⁸ flavin-dependent,^{45,59,60} iron & 2OG-dependent⁶¹ and nucleophilic.⁶²

1.3. The Fe²⁺ & 2OG-dependent oxygenase family

The non-haem Fe²⁺ and 2OG-dependent oxygenase family (commonly referred to as 2OG-dependent oxygenases or 2OG-oxygenases) is a ubiquitous family of enzymes that catalyses oxidative reactions on a variety of large and small molecule substrates, including proteins, peptides, amino acids, nucleic acids, lipids, antibiotics and signalling molecules.⁶³ In higher organisms (i.e. humans & animals), known 2OG-oxygenase enzyme activity is limited to hydroxylation reactions (and *N*-demethylation proceeding *via* hydroxylation), but a wider range of reactions are catalysed by 2OG-oxygenases in plants and microorganisms.^{64,65} Interestingly, despite the varied substrates and range of reaction types, members of this superfamily have several common active-site structural features, therefore possessing a well-adapted structural platform for the evolution of new catalytic activities.⁶⁴

All studied 2OG-oxygenases require Fe²⁺ as a cofactor. Two histidine residues and the carboxylate of an aspartate/glutamate bind to the Fe²⁺ and form a highly (but not universally) conserved facial triad of residues.⁶⁶ This motif is supported within a conserved double-stranded beta-helix (DSBH, a.k.a. jelly-roll or cupin) fold.⁶³ The DSBH fold is constructed from two β -sheets, with each sheet consisting of four anti-parallel β -strands denoted by the roman numerals I-VIII. β -strands I, VIII, III and VI form the major sheet and β -strands II, VII, IV and V form the minor sheet (Figure 1.10.).⁶³ Variation in the secondary structural elements extending from DSBH, particularly between strands IV and V (termed the β IV-V insert) in part leads to the different sub-families of 2OG-oxygenases. This insert can vary in size from just a few residues to a whole domain.^{63,67} 2OG is employed as a co-substrate by this enzyme family, in a

mechanism which couples oxidation of the substrate with deoxygenation of 2OG to yield succinate and carbon dioxide.⁶⁴

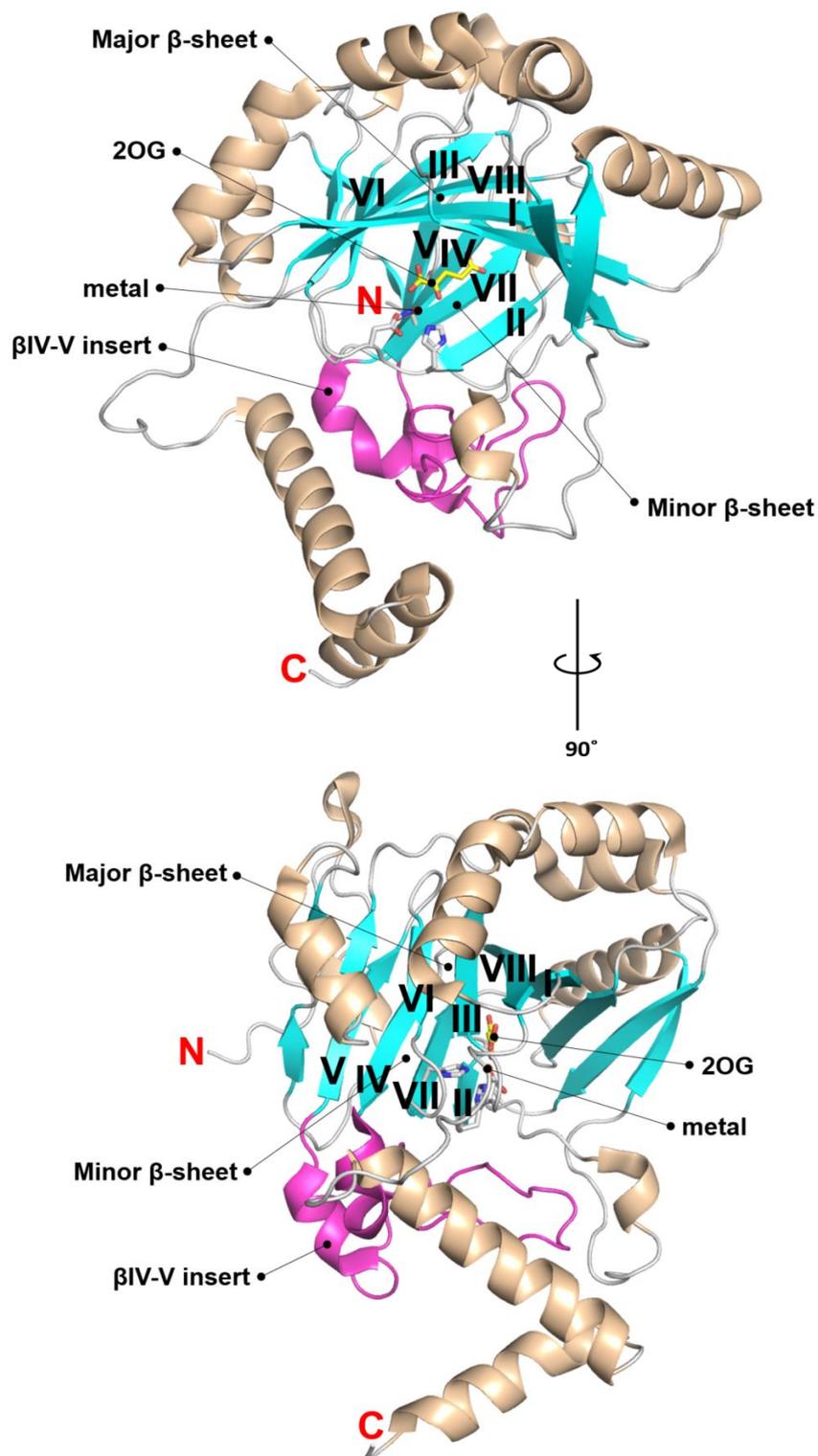
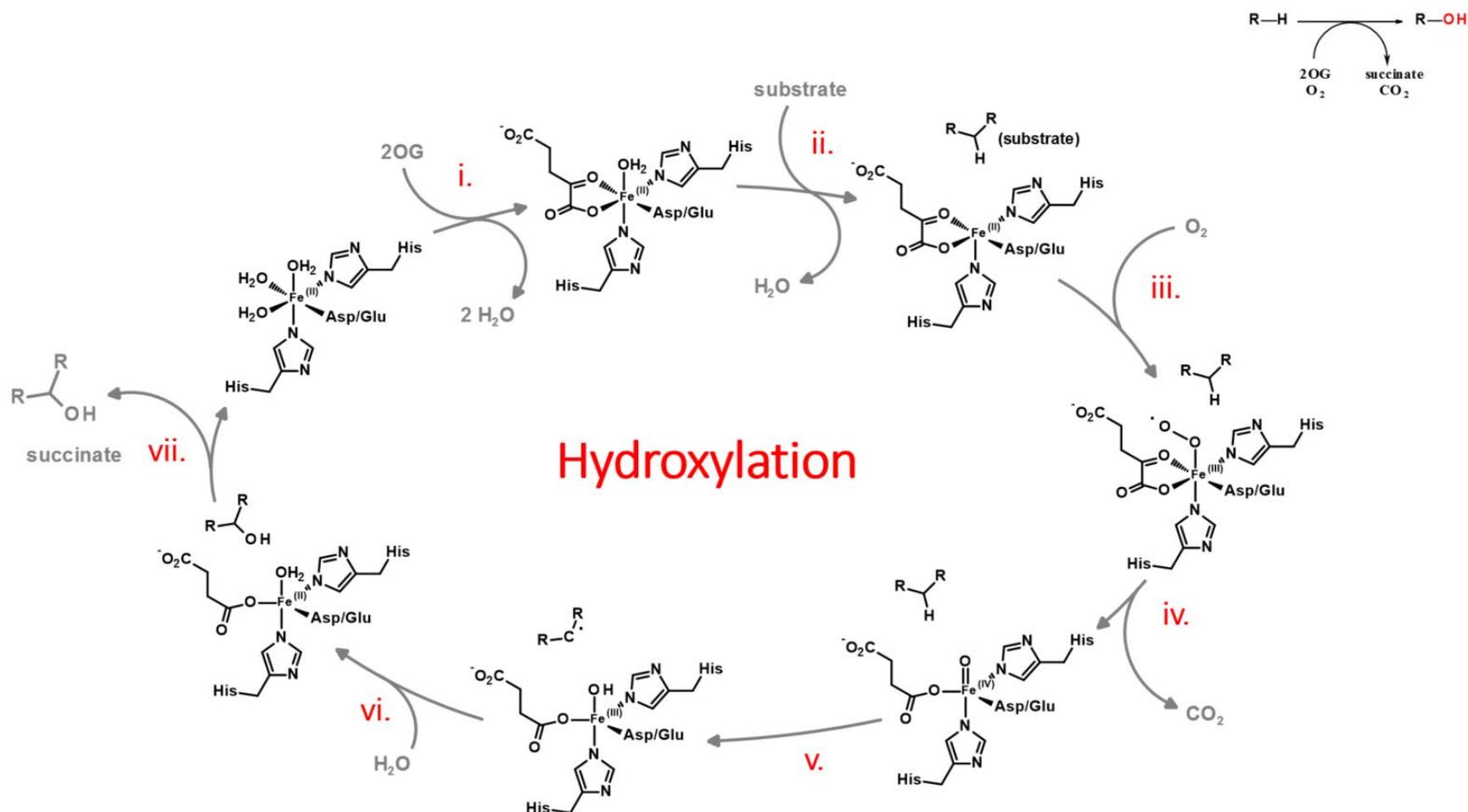


Figure 1.10. Views from a crystal structure of FIH (PDB ID: 4Z2W) with Fe^{2+} metal (red sphere) and 2OG (yellow) bound illustrate the conserved structural features of 2OG-oxygenase enzymes. The N- and C-termini are labelled in red. The core DSBH strands (I-VIII) are highlighted in cyan and the β IV-V insert is coloured pink. The metal sphere is shown coordinated to an aspartate and two histidine residues (sticks), typical of the 2OG-oxygenase active site facial triad. Secondary structural elements common among most 2OG-oxygenases are labelled.

1.3.1. The consensus mechanism for 2OG-oxygenases

The most well studied reaction performed by 2OG-oxygenases is hydroxylation, a mechanism for which is outlined in Scheme 1.1.⁶⁸ A 2OG-dependent oxygenase is considered in its resting state when the Fe^{2+} is held within the active site, octahedrally coordinated by the side chains of two histidine residues and an aspartate or glutamate. The three remaining Fe^{2+} coordination sites are occupied by water molecules. Binding of 2OG displaces two waters, forming a bidentate interaction with the Fe^{2+} , and initiating the catalytic cycle (i).⁶⁸ Positioning of the substrate within the active site proximal to the catalytic triad leads to displacement of the third water, activating the enzyme towards oxygen binding (ii). Binding of molecular oxygen at the metal centre generates a Fe(III)-superoxo intermediate (iii), which attacks the C-2 position of 2OG, leading to the formation of a highly reactive Fe(IV)=O (ferryl) intermediate⁶⁹ with concomitant decarboxylation of the 2OG to succinate and carbon dioxide (iv). This Fe(IV)=O intermediate abstracts a hydrogen from a substrate C-H bond, leading to a substrate radical and an Fe(III)-hydroxyl intermediate (v). In what has been termed a radical rebound mechanism (although direct insertion is also possible), the substrate radical abstracts the resulting hydroxyl (formed in the previous abstraction step) to form a hydroxylated product (vi), which dissociates from the active site (vii).⁶⁸ The catalytic cycle resumes at this point by displacement of the succinate with 2OG, or the resting state is restored by the binding of waters in the vacant coordination sites. If the position of hydroxylation is next to a nitrogen (i.e. N-Me), a hemiaminal intermediate is formed which spontaneously fragments to deliver *N*-demethylation as the net process.

Hydroxylation (and *N*-demethylation) proceeding *via* the consensus mechanism dominates in human 2OG-oxygenases, with reaction substrates ranging from proteins to nucleic acids and lipids. It is this activity which is responsible for the diverse roles of 2OG-oxygenases in humans, with roles in collagen biosynthesis, oxygen sensing, DNA repair and the regulation of gene expression (epigenetics).⁶⁷ Small molecule 2OG-oxygenases display a wider range of reactions (i.e. beyond hydroxylation) with implications for biocatalytic engineering.



Scheme 1.1. The 2OG-oxygenase consensus mechanism. The catalytic cycle starts with the binding of 2OG (i), followed by substrate (ii) within the enzyme active site. Molecular oxygen binding at the Fe(II) centre generates a Fe(III)-superoxo intermediate (iii) which attacks the C-2 position of 2OG, resulting in a Fe(IV)=O intermediate and decarboxylation of the 2OG to succinate (iv). The Fe(IV)=O intermediate abstracts a hydrogen from a substrate C-H bond, to give a substrate radical and a Fe(III)-hydroxyl intermediate (v). Subsequent hydroxyl abstraction by a substrate radical (vi) generates the hydroxylated product, which dissociates from the active site (vii).

1.3.2. Small-molecule Fe²⁺ & 2OG-dependent oxygenases

The range of small molecule reactions performed by plant and bacterial 2OG-oxygenases are of particular interest to the synthetic chemist, with examples of desaturation,⁷⁰ epoxidation,⁷¹ halogenation,⁶¹ rearrangements,⁷² epimerisations,⁷³ and cyclisations⁷⁴ documented in the literature. By varying the non-essential active site residues of enzymes from this family, it should be possible to engineer in promiscuity whilst maintaining catalytic activity. CAS, BBOX and AsqJ are used to illustrate the flexibility in catalysis demonstrated by small molecule 2OG-oxygenases.

Clavaminc acid synthase (CAS)

CAS, from the bacterium *Streptomyces clavuligerus*, has been shown to mediate three separate oxidative reactions in clavaminic acid biosynthesis; hydroxylation, cyclisation and desaturation (Figure 1.11.).⁷⁵ All three reactions are coupled with the decarboxylation of 2OG to succinate and take place at the same active site.⁷⁵

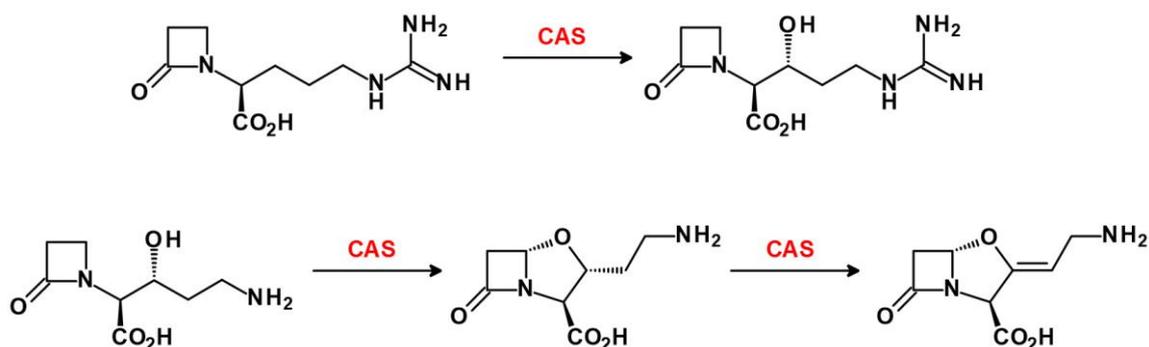


Figure 1.11. The three oxidative reactions performed by CAS (clavaminc acid synthase) all occur at a single active site.

Additionally, crystal structures of CAS demonstrated that 2OG has the capacity to co-ordinate in more than one geometry around the metal centre (Figure 1.12.).⁷⁵ Crystals of CAS exposed to nitric oxide (a surrogate for the binding of

molecular oxygen) showed binding in the axial coordination site and movement of the 2OG C-1 carboxylate into an alternative coordination site vacated by the coordinated water.⁷⁶ This coordination flexibility indicates the possibility of active site rearrangements during catalysis which may aid the ability of CAS to act upon multiple substrates.

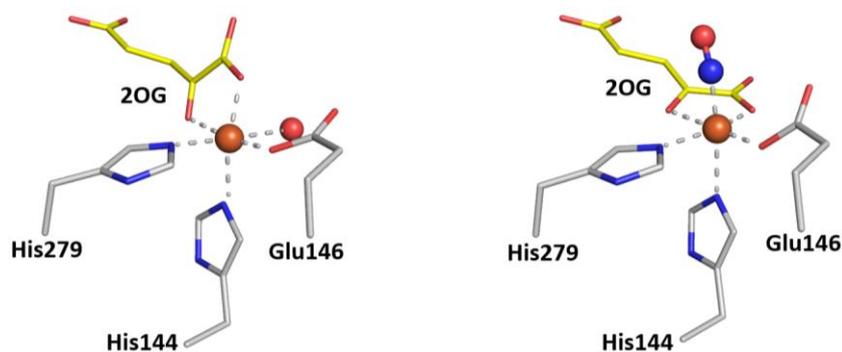


Figure 1.12. Active site views of the crystal structures of CAS with (left; PDB ID: 1GVG) and without the dioxygen surrogate nitric oxide (right; PDB ID: 1DRY, nitric oxide shown as red-blue spheres), reveal different coordination geometries for 2OG (yellow).

Observations for CAS indicate the potential for other 2OG-oxygenases to operate as promiscuous enzymes, showing promise for the development of biocatalysts which exploit the 2OG-oxygenase catalytic motif.

γ -butyrobetaine hydroxylase (BBOX)

Human γ -butyrobetaine hydroxylase (hBBOX) and its more promiscuous homologue PsBBOX (from the bacterium *Pseudomonas* sp. AK1) were found to be able to catalyse the hydroxylation of a variety of different substrate-like compounds (including fluorinated analogues) in addition to their endogenous substrate, γ -butyrobetaine.⁷⁷ The ability of hBBOX to catalyse the oxidative desymmetrisation of achiral *N,N*-dialkyl piperidine-4-carboxylates is of particular interest from a biocatalytic perspective due to the level of structural differentiation between these piperidine substrates and γ -butyrobetaine (Figure 1.13.).⁷⁸

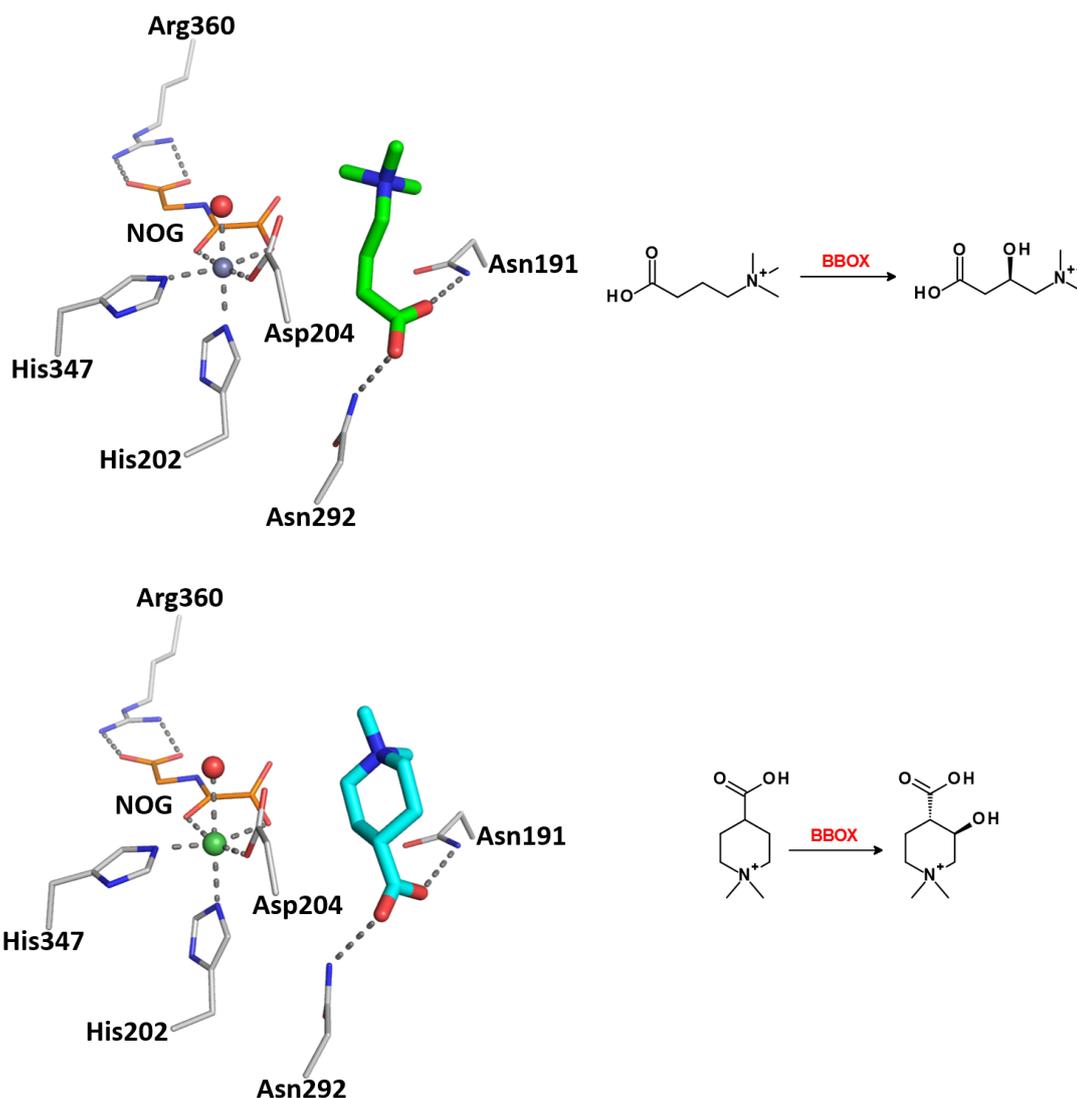


Figure 1.13. Active site views of the crystal structures of hBBOX complexed with the substrates γ -butyrobetaine (top; PDB ID: 3O2G) and *N,N*-dimethylpiperidine (bottom; PDB ID: 4CWD).

AsqJ

AsqJ, from the fungus *Aspergillus nidulans*, catalyses both the desaturation and epoxidation of a benzodiazepinedione within the same active site (Figure 1.14).^{71,79}

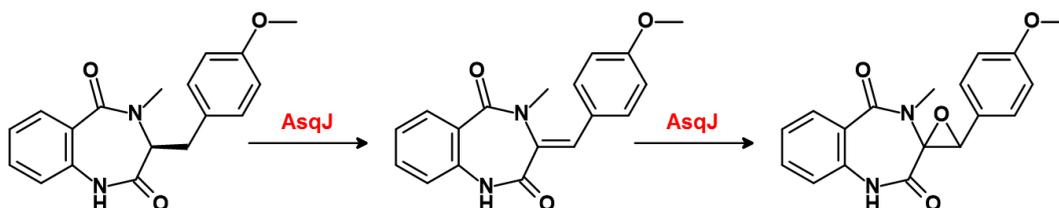


Figure 1.14. Desaturation and epoxidation reactions catalysed in the active site of the 2OG-dependent oxygenase AsqJ.

The presence of a methyl group at the *N*-4 position of the substrate (Figure 1.15.) was found to be key to AsqJ activity.⁸⁰ However, the V72I variant was shown to be able to catalyse desaturation and epoxidation of the des-methyl benzodiazepinedione, in addition to the wild-type activity (effectively doubling the number of reactions being performed at the active site).⁸⁰ This increased promiscuity was rationalised in terms of the isoleucine side chain being able to compensate for the lack of methyl group.

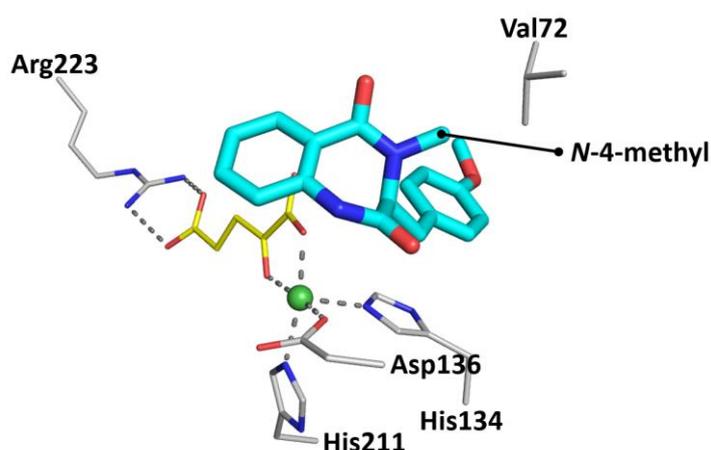


Figure 1.15. *AsqJ* variant V72I compensates for the lack of methyl group, catalysing the desaturation and epoxidation of des-methyl benzodiazepinedione substrates.

1.3.3. Fe²⁺ & 2OG-dependent halogenases

Despite the conserved structural elements within the active sites of 2OG-oxygenases, members of this enzyme family show activity against a remarkable range of substrates. Individually, many of these enzymes have also been shown to catalyse the turnover of non-natural substrates (exemplified in Section 1.3.2). The active site of 2OG-halogenases is an additional example of catalytic flexibility for 2OG-oxygenases. A glycine/alanine residue takes the place of the glutamate/aspartate in the active site of 2OG-halogenases.⁴⁰ Absence of the side chain carboxylate of the iron binding triad creates a vacant Fe²⁺ coordination site which is occupied by a chlorine atom. As for the 2OG-

oxygenases, the Fe^{2+} centre is bound to two active site histidine residues and is chelated by 2OG in the same bidentate manner (Figure 1.16.).⁴⁰

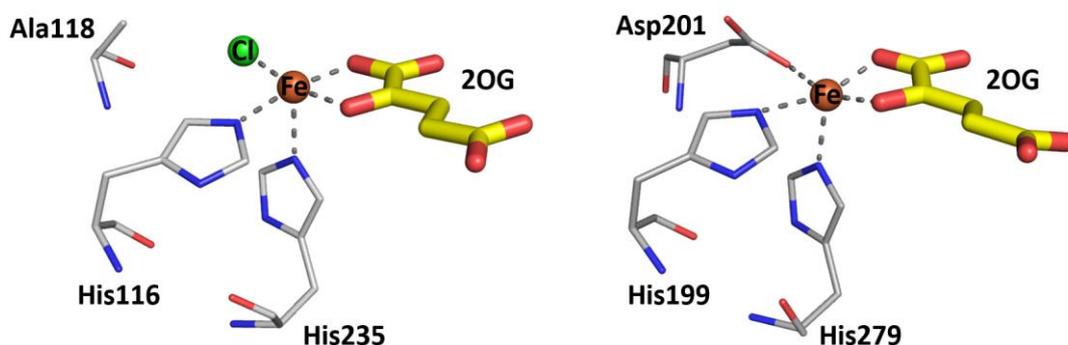
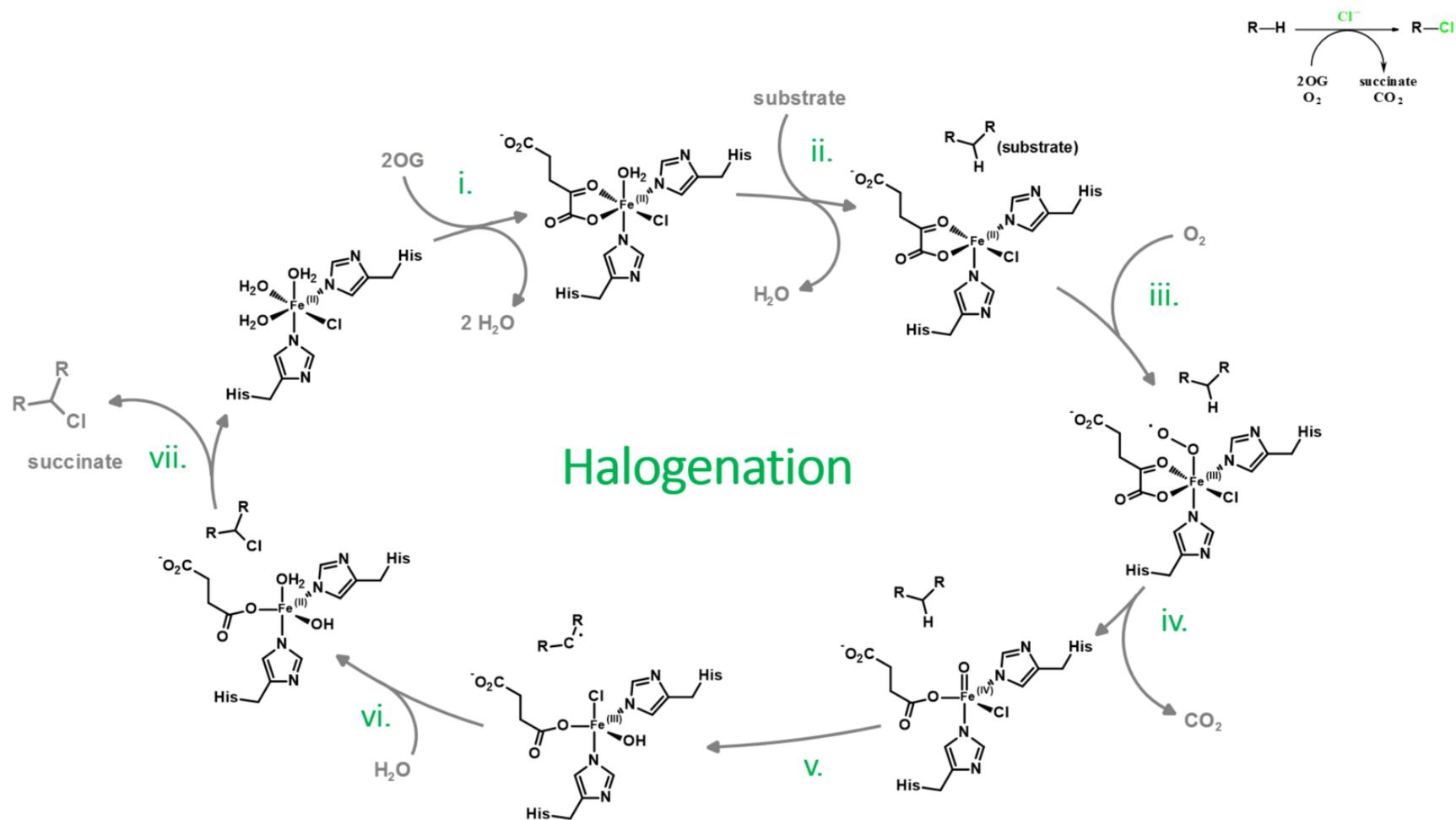


Figure 1.16. Typical active site motifs for the 2OG-halogenases on the left (as exemplified by SyrB2, PDB ID: 2FCT) and the 2OG-hydroxylases on the right (as exemplified by FIH, PDB ID: 1MZF). Coordination of the two histidine residues to the Fe^{2+} (brown sphere) and bidentate binding of 2OG (yellow) is conserved. Substitution of the Asp/Glu in 2OG-oxygenases for an Ala/Gly residue in 2OG-halogenases results in a coordination site suitable for halogen binding.

The mechanism of halogenation is therefore comparable to that of the 2OG-dependent oxygenases, with both mechanisms proceeding to the hydrogen abstraction step in an analogous manner but diverging at the radical rebound step.⁶⁹ In the halogenases, the substrate radical abstracts the Fe^{2+} bound chlorine at this point, and not the hydroxyl (Scheme 1.2.).⁶⁹ This radical mechanism permits the introduction of chlorine at an un-activated carbon centre. As this reaction is enzyme-mediated, chlorine introduction is regio- and stereoselective. The ability to perform such a reaction, also in the presence of other active functionalities, is not currently possible synthetically.



Scheme 1.2. The 2OG-halogenase catalytic cycle. Catalysis is initiated by the binding of 2OG (i), followed by substrate (ii) within the enzyme active site. Molecular oxygen binding at the Fe(III) centre generates a Fe(III)-superoxo intermediate (iii) which attacks the C-2 position of 2OG, resulting in a Fe(IV)=O intermediate and decarboxylation of the 2OG to succinate (iv). The Fe(IV)=O intermediate abstracts a hydrogen from a substrate C-H bond, to give a substrate radical and a Fe(III)-hydroxyl-chloro intermediate (v). Subsequent chloride (and not hydroxyl) abstraction by a substrate radical (vi) generates the chlorinated product, which dissociates from the active site (vii).

Studies on the 2OG-halogenase SyrB2 have provided evidence for the proposal that substrate positioning relative to the Fe(III) intermediate leads to chlorination over hydroxylation.^{81,82} SyrB2 was shown to exhibit a preference for chlorination, rather than it being the exclusive reaction outcome. This is also assumed to be the case for the other 2OG-halogenases.⁸¹ Further indications toward the catalytic flexibility of 2OG-halogenases were illustrated by the ability of SyrB2 to accept and use alternative anions in place of chloride.⁸³ This shows potential for the development of 2OG-halogenase based biocatalysts capable of regio- and stereoselective installation of hydroxyl, bromide, chloride, nitrite and azide functionalities.

However, as previously mentioned in Section 1.3.3., 2OG-halogenases have seen little investigation for biocatalytic applications due to the requirement of a substrate-bound carrier protein for activity. The discovery of the homologous 2OG-halogenases WelO5 and AmbO5, which are capable of the chlorination of isolated small molecules, has fuelled interest in 2OG-halogenase-catalysed halogenation for biocatalytic purposes.

1.3.3.1. Discovery of WelO5 and AmbO5

Through efforts to understand the biosynthesis of compounds from the hapalindole-type natural product family (Figure 1.17.), the group of Xinyu Liu at the University of Pittsburgh identified a 36kb welwitindolinone gene cluster in the cyanobacterium *Hapalosiphon welwitschii* UTEX B1830⁸⁴ and a 42kb ambiguine gene cluster in the cyanobacterium *Fischerella ambigua* UTEX 1903.⁸⁵ Using the identity of compounds isolated from these organisms and bioinformatic analysis of the genes within each cluster, proposals were made for the enzymes involved at different stages of compound biosynthesis.

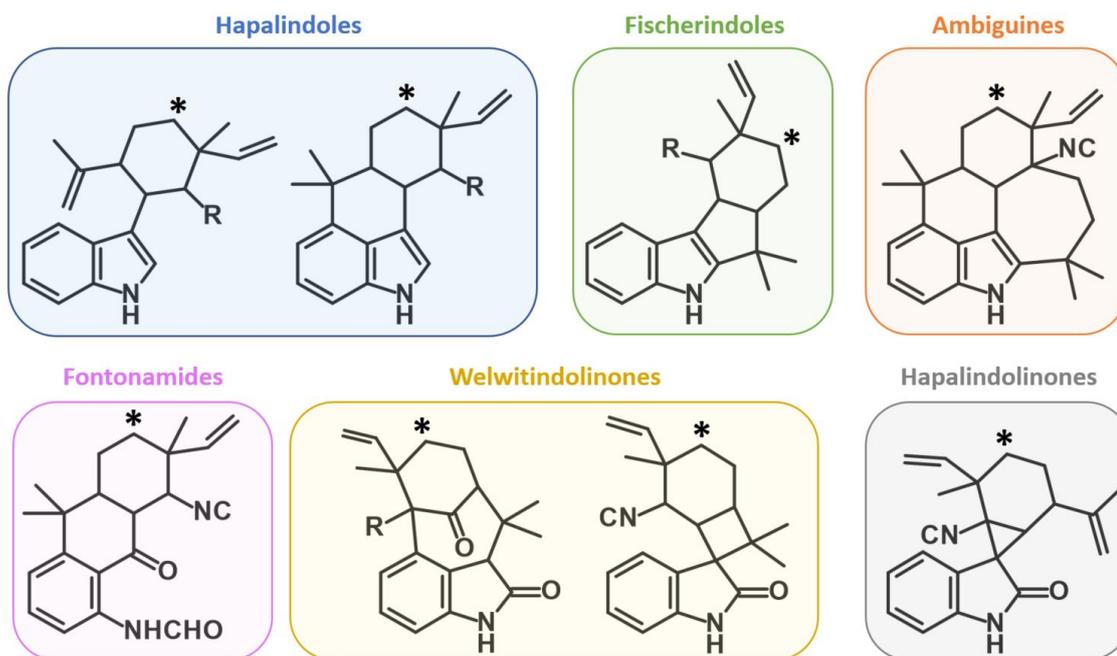


Figure 1.17. Generic structures representing the carbon skeletons of over 80 compounds of the hapalindole-type natural product family, with sites of chlorination indicated by stars. Stereochemistry excluded for clarity. R = isocyanate (-NC) or isothiocyanate (-NCS).

These included two putative 2OG-dependent halogenases, WelO5 (*Hapalosiphon welwitschii* UTEX B1830)⁸⁴ and AmbO5 (*Fischerella ambigua* UTEX 1903).⁸⁵ Recombinant protein production and activity testing confirmed the functional assignment of these enzymes, which were found to share ~ 80% sequence identity.⁵¹

In 2014, WelO5 (the first confirmed member of this family) was shown to regio- and stereoselectively chlorinate the un-activated carbon of 12-*epi*-fischerindole U to generate 12-*epi*-fischerindole G (Figure 1.18).⁵⁰ This activity was also shown to be dependent upon Fe²⁺ and 2OG.

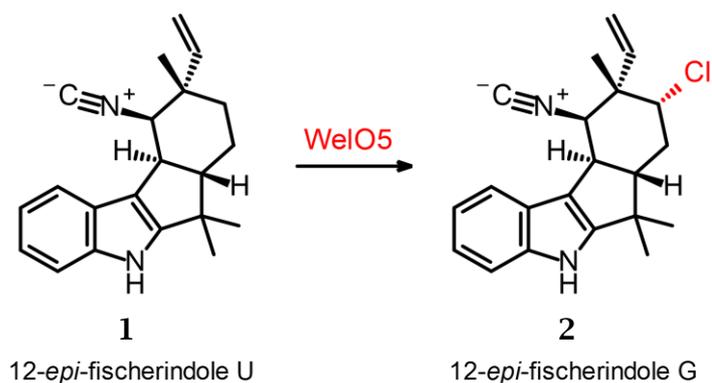


Figure 1.18. WelO5-catalysed formation of 12-*epi*-fischerindole G, **2** by the late-stage chlorination of 12-*epi*-fischerindole U, **1**.

WelO5 was also shown to chlorinate 12-*epi*-hapalindole C, albeit with considerably reduced activity relative to 12-*epi*-fischerindole U (Figure 1.19).⁵⁰

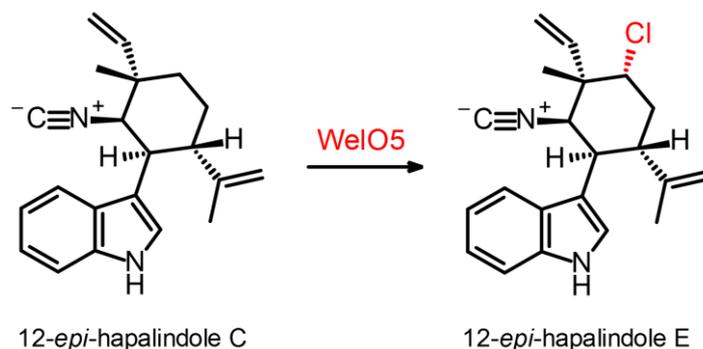


Figure 1.19. WelO5 is also capable chlorinating 12-*epi*-fischerindole C to generate 12-*epi*-fischerindole E.

In 2016, AmbO5 was characterised and confirmed to be a 2OG-dependent halogenase.⁵¹ AmbO5 was found to be capable of chlorinating five different hapalindole-type natural products isolated from *Fischerella ambigua* UTEX 1903 to varying extents (in addition to those chlorinated by WelO5; Figure 1.20.), therefore the natural substrate for AmbO5 could not conclusively be assigned. However, this demonstrated that for compounds of the hapalindole-type family, AmbO5 was more promiscuous than WelO5.⁵¹

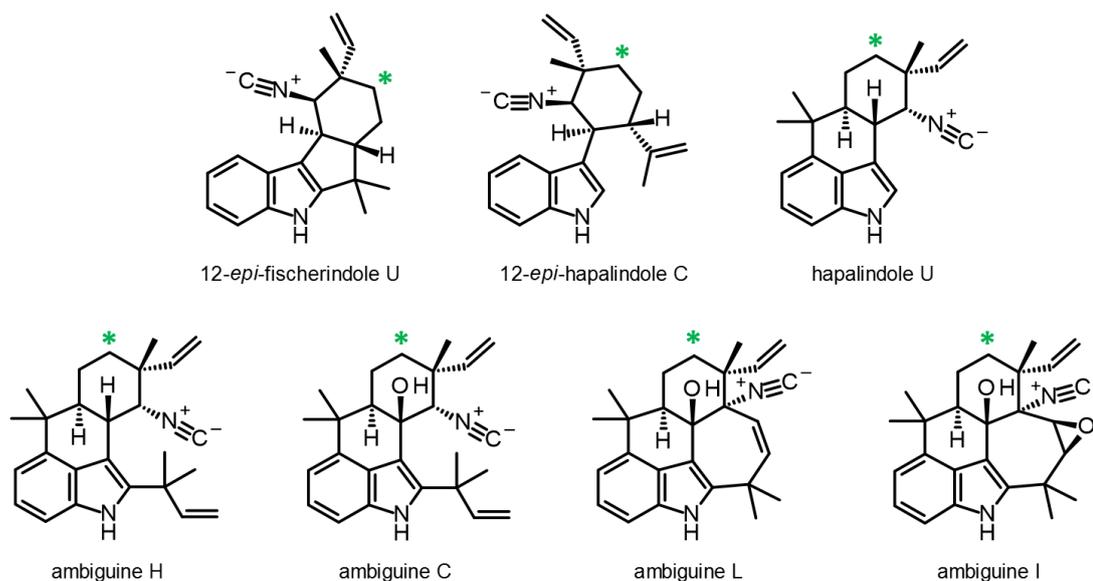


Figure 1.20. Structures of compounds which undergo AmbO5 chlorination, with sites of chlorination indicated by green stars. Ambiguines H, C, L and I were shown to be equally competent substrates for AmbO5, with significantly reduced activity demonstrated for hapalindole U. AmbO5 showed comparable activity to WelO5 for substrates 12-epi-hapalindole C and 12-epi-fischerindole U.

A sequence alignment of AmbO5 with WelO5 (both 290 amino acids in size) showed that a large number of non-identical residues were clustered in the region between residues 213 and 235.⁵¹ By mutating this 23 residue region of WelO5 to match that of AmbO5, the activity of the resulting variant was increased against the ambigaine compounds (albeit with reduced activity relative to AmbO5) but also decreased against 12-epi-fischerindole U and 12-epi-hapalindole C.⁵¹ This chimeric variant possessed an expanded substrate scope relative to WelO5, but a different selectivity profile to both WelO5 and AmbO5 – showing promise for the evolution of these enzymes against alternative substrates.

The discovery and functional assignment of AmbO5 and WelO5 established the existence of a new family of 2OG-dependent halogenases. For all previously characterised Fe^{2+} & 2OG-dependent halogenases, the binding of a carrier protein to the substrate is necessary for activity.^{61,86-92} Therefore, whilst WelO5

and AmbO5 were not the first Fe²⁺ & 2OG-dependent halogenases to be identified, their ability to act upon isolated small molecules renders them highly suitable for biocatalytic evaluation.

1.4. Aims & objectives

Biocatalytic methods have been successfully developed for a number of pharmaceutical and fine chemical industrial processes but an increase in the breadth of available reaction types would help to increase the routine application of biocatalysis in synthetic chemistry.

One enzyme family demonstrating potential to deliver in this regard is that of the iron and 2-oxoglutarate (2OG)-dependent oxygenases, due to the variety of different synthetically interesting reactions achievable using a common active site motif.⁶⁴ The discovery of WelO5 (and later AmbO5) has opened up the possibility of developing biocatalytic 2OG-dependent small-molecule halogenases for the chlorination of unactivated carbons.

This thesis outlines the development of tools used to evaluate the tractability of producing synthetically useful biocatalysts based on WelO5. This work could lead to the development of a series of engineered enzymes capable of accepting a variety of small molecules for late-stage halogenation. Novel enzymes of this type could be used to explore new vectors on existing lead compounds and provide access to chemical matter currently unobtainable by other means. The work described in this thesis focuses predominantly on WelO5 because AmbO5 was not characterised until 2016.⁵¹ However, additional rationale exists for the selection of WelO5 over AmbO5 as a biocatalytic starting point. Firstly, the natural substrate for AmbO5 has not been

conclusively assigned. AmbO5 has been shown to catalyse the chlorination of multiple compounds but none to the levels of conversion seen for WelO5 on 12-epi-fischerindole U (AmbO5 had comparable activity to WelO5 for this substrate). Secondly, the development of enzymes against non-hapalindole-like compounds may not necessarily be more successful for AmbO5 than WelO5 due to the high homology shared by these enzymes and the reasonable similarity of the substrates used to define AmbO5 promiscuity.

To investigate the suitability of WelO5 as a starting point for biocatalytic development, work was initiated in the following areas:

- (1) Use of bioinformatic tools (database searches, sequence comparisons and structural prediction) to analyse WelO5 and search for related enzymes (it is unlikely that WelO5 is an isolated case and by looking for similar enzymes it may be possible to cover a wider range of substrates).
- (2) WelO5 evaluation (enzyme production and characterisation, structural studies, probing of wild-type enzyme promiscuity).
- (3) The production of WelO5 mutants with new activities and modified specificities.

Chapter 2. Production of WelO5 & related proteins

2.1. Introduction

Towards the end of 2014, WelO5 was reported in the literature as a member of the Fe²⁺ and 2OG-dependent halogenase family.⁵⁰ The ability of WelO5 to act upon an untethered substrate molecule had not been seen before for a member of the halogenase subfamily of 2OG-oxygenase enzymes, possibly indicating novel structural and mechanistic features. Although the biochemical function of WelO5 had been assigned, structural information beyond its primary sequence was not available at the time. Further bioinformatic and experimental investigation into the structure-function relationship of WelO5 was therefore warranted. It was hoped that three-dimensional structural information for WelO5 could be used to relate structure to function,^{86,89,93} provide insight into how WelO5's novel activity is achieved, and possibly indicate the feasibility of developing novel small-molecule biocatalysts based on WelO5.

In this Chapter, the primary sequence of WelO5 is compared with known Fe²⁺ and 2OG-dependent halogenase enzyme family members. WelO5 homologues were identified from a literature search^{94,95} with the aim of providing further insight into the relationship between the structure of WelO5 and its novel activity. The ability to predict the three-dimensional structure of WelO5 is evaluated and efforts towards producing protein in suitable quantities for assays, and in sufficiently high purity for crystallisation trials are described.

2.2. Bioinformatics & biocatalysis

To engineer in unique activities and/or alter substrate specificities, existing enzymes can be modified by directed evolution or by rational design.^{3,9,96}

Directed evolution does not necessarily require prior structural knowledge and can rapidly generate enzyme variants with desired characteristics. However, the large numbers involved demand an efficient and robust selection method or high-throughput screening capabilities.^{13,97} Unfortunately, the outcome of WelO5's chlorination reaction does not obviously lend itself to the use of selection processes and the development of a robust, high-throughput screen for directed evolution was assumed to be a significant and likely expensive challenge within the designated time constraints. As the eventual goal was to evolve WelO5 to act upon alternative substrates, the development of a universal, medium-throughput, LC-MS based screening method was envisaged as a suitable alternative. By designing specific amino acid substitutions in the WelO5 active site, it was hoped that an understanding of the biocatalytic potential of WelO5 could be obtained in the absence of high-throughput screening methods. Whilst efforts towards a WelO5 crystal structure were ongoing, bioinformatic tools were used for several aspects of WelO5 structural analysis and for the identification of WelO5 homologues.

Bioinformatic tools were anticipated to help guide the rational re-design of the WelO5 active site in the absence of three-dimensional biophysical information,⁹⁸⁻¹⁰⁰ with the analysis of WelO5 being divided into the following parts:

- *Identification of similar enzymes.* The 'genome mining' of predicted and known protein sequence databases could identify enzymes with reasonable sequence similarity to WelO5. Similar proteins without

functional assignment could be identified and investigated as putative novel Fe²⁺ and 2OG-dependent halogenase enzymes. Proteins possessing structural information and sharing a certain level of sequence identity with WelO5 may provide guidance on the three-dimensional structure of WelO5, with any differences between the sequences providing ideas for amino acid substitution.

- *Comparison of WelO5 primary sequence to enzymes of known and related function.* Structural information (principally crystallographic) on other members of the Fe²⁺ and 2OG-dependent enzyme family could provide insight into the structure of WelO5.^{63,64}
- *WelO5 structural prediction.* It was envisaged that structural information from related Fe²⁺ and 2OG-dependent oxygenases could be used to develop a homology model in the event of a WelO5 crystal structure being unobtainable. Structural prediction algorithms (covered in more detail later), which attempt to simulate protein tertiary structure using primary sequence as an input could also be used.¹⁰¹⁻¹⁰⁴ In the absence of a WelO5 crystal structure, a homology model or predicted three-dimensional structure based on knowledge of crystallised Fe²⁺ and 2OG-dependent oxygenases could help to find enzymes with a similar structure, but relatively low sequence identity/similarity to WelO5. Insight into the tertiary structure of these enzymes may then provide ideas for WelO5 active site substitutions.

2.2.1. WelO5 primary sequence analysis

2.2.1.1. Identification of similar enzymes

WelO5, from cyanobacterium *Hapalosiphon welwitschii*, is a 290 amino acid protein.⁸⁴ Substituting each residue in turn to the other 19 proteogenic amino

acids (i.e. saturation mutagenesis) would require the generation of 5,510 (290 x 19) unique proteins. Furthermore, performing multiple simultaneous amino acid substitutions would expand the total number exponentially to the point of impracticality. Ideally, a crystal structure would help to narrow down the number of residues for selection. However, structural information beyond the primary sequence was not available for WelO5 at the start of this work (i.e. late 2014). Although predictions onto regions of WelO5 substrate binding residues may be made on the basis of comparisons with known 2OG-dependent oxygenase structures, the substrate binding elements of 2OG-dependent oxygenases tend to be variable, especially compared to the core Fe²⁺ and 2OG binding elements.⁶³ By identifying proteins with a certain level of sequence similarity to WelO5, it was envisaged that the number of amino acid substitutions could be narrowed down. Between these similar enzymes, areas of high sequence similarity could be disregarded, whilst areas of low similarity could relate to differences in substrate selectivity and show areas tolerant to variation without abolishing enzymatic function.

To initiate a comparison of WelO5 with other proteins, a PHI-BLAST^{94,95} search was performed on all available non-redundant protein sequences using the WelO5 primary sequence as a search query. PHI-BLAST is a sub-category of the protein-BLAST search program which uses a specific motif or pattern as part of the search query to act as an automatic constraint when filtering results.

Similar enzymes from the Fe²⁺ and 2OG-dependent enzyme family were desired, therefore searches were performed using the conserved iron-binding catalytic motif as an input pattern; the H-X-A/G ...H motif¹⁰⁵ was used for the identification of potential halogenases and the H-X-D/E...H motif⁶⁶ for the identification of potential hydroxylases, respectively. EMBOSS Needle¹⁰⁶ was

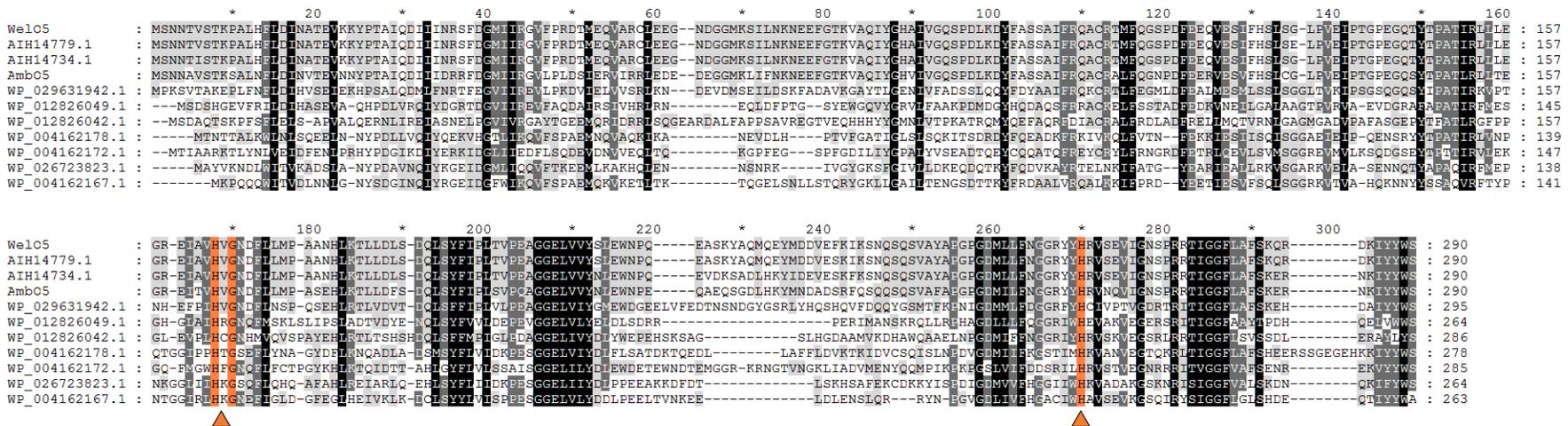
used to generate pairwise sequence alignments of WelO5 with each of the PHI-BLAST results and to subsequently calculate sequence identity (using the equation below) and similarity (scored using the BLOSUM62 substitution matrix).¹⁰⁷

$$\text{Percentage identity} = \frac{\text{number of identical amino acid matches}}{\text{length of alignment}} \times 100$$

Only one enzyme sequence identified from the PHI-BLAST searches had already been reported in the literature, AmbO5.⁸⁵ Like WelO5, AmbO5 is a cyanobacterial halogenase enzyme known to be involved in alkaloid biosynthesis.⁵¹ The 316 amino acid sequence initially published for AmbO5 was found to share ~70% identity and ~85% similarity with WelO5 (EMBOSS Needle; Needleman-Wunsch algorithm). The AmbO5 sequence was later revised at the *N*-terminus to give a 290 amino acid protein.⁵¹ This revised sequence was found to possess ~80% identity to, and ~90% similarity with WelO5. In addition to AmbO5, nine putative halogenases with >25% identity (range 29-99%) and >45% similarity (range 46-99%) to WelO5 were identified from PHI-BLAST searches (Table 2.1.). Using Clustal Omega,¹⁰⁸ the sequences of these putative halogenases were used to generate a multiple sequence alignment with WelO5 (Figure 2.1.) for the identification of areas of high sequence conservation as well as sequence diversity.

	NCBI Protein Identifier	Source	Annotation	% Identity shared with WelO5	% Similarity to WelO5
1	AIH14779.1	<i>Westiella intricata</i> UH HT-29-1	oxidoreductase	99	99
2	AIH14734.1	<i>Hapalosiphon welwitschii</i> UH IC-52-3	oxidoreductase	95	98
3	AHB62755.1	<i>Fischerella ambigua</i> UTEX 1903	AmbO5	79	90
4	WP_029631942.1	<i>Scytonema hofmannii</i> UTEX B 1581	hypothetical protein	45	63
5	WP_012826049.1	<i>Haliangium ochraceum</i>	hypothetical protein	33	52
6	WP_012826042.1	<i>Haliangium ochraceum</i>	hypothetical protein	32	52
7	WP_004162178.1	<i>Microcystis aeruginosa</i>	hypothetical protein	32	50
8	WP_004162172.1	<i>Microcystis aeruginosa</i>	hypothetical protein	32	48
9	WP_026723823.1	<i>Fischerella sp.</i> PCC 9431	hypothetical protein	29	48
10	WP_004162167.1	<i>Microcystis aeruginosa</i>	hypothetical protein	29	46

Table 2.1. WelO5 PHI-BLAST halogenase motif (HXG/A...H) search results. Proteins are labelled per their annotation within the NCBI protein database. Identity to, and similarity shared with WelO5 were calculated by EMBOSS Needle pairwise sequence alignment using the default BLOSUM62 substitution matrix.



We can initially see from the alignment in Figure 2.1. that the sequences of these putative halogenases are similar in length (between 260-290 residues), with two regions showing noticeable variability (residues 60-90 and 215-255). Three highly conserved regions are also present, corresponding to residues 190-215, 255-265 and 280-290 in WelO5. In the absence of three-dimensional structural information, predicting the role of these conserved regions is extremely challenging (they are not part of the core DSBH of the 2OG-dependent oxygenases). However, it is possible that regions of variability may be involved in selectively binding substrate, whilst areas of conservation may be involved in enzyme halogenation activity. What is also notable for these proposed halogenases is that most of them are cyanobacterial in origin - together with the high levels of sequence similarity found, this may suggest a common evolutionary ancestor. The source species for each of these halogenases was mapped to a phylogenetic tree to help visualise the evolutionary relationships between these bacterial halogenases (Figure 2.2.).

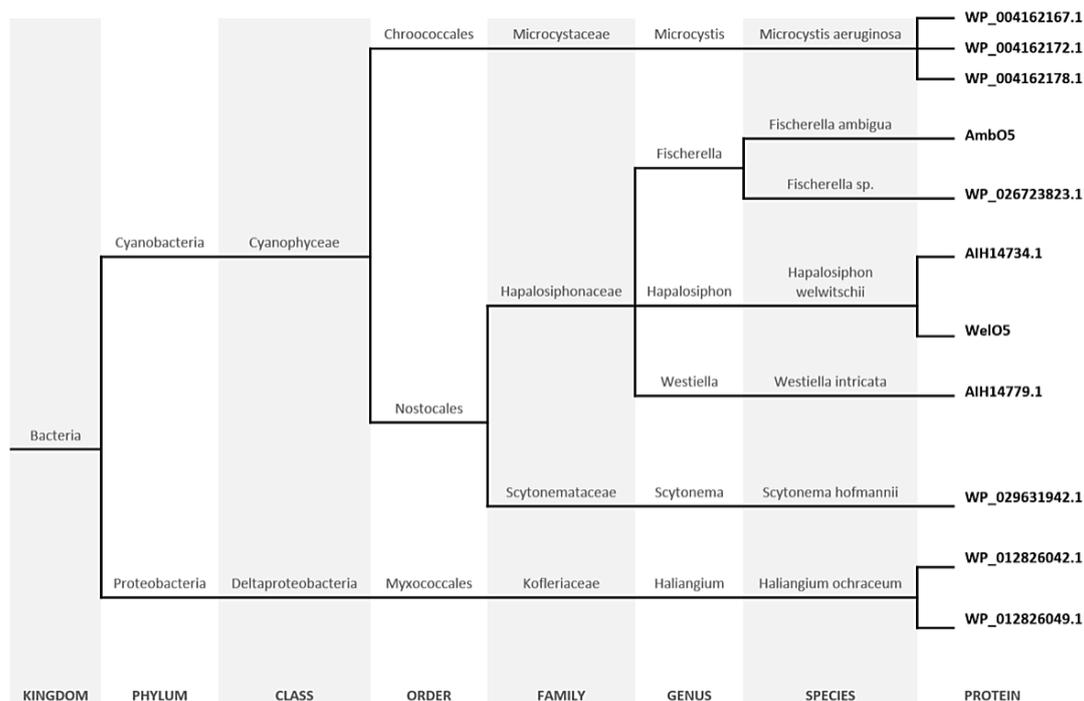


Figure 2.2. Phylogenetic tree constructed from putative halogenases identified by a WelO5 PHI-BLAST search and adapted from a tree generated in NCBI Taxonomy Browser. In the absence of a given name, proteins are identified by their NCBI database ID numbers. Proteins are mapped to their species of origin (as per Figure 2.1.). Changes in taxonomic rank are shown by horizontal movements across the table and are represented by changes in column shading.

In addition to these ten putative halogenases, a small number of homologous hydroxylases were identified by PHI-BLAST (Table 2.2.). To identify enzymes considered likely to demonstrate hydroxylase functionality, only sequences consistent with the hydroxylase motif (HXD/E...H) as well as a '2OG-oxygenase' (or related) annotation were selected from the PHI-BLAST results list. Due to the lower levels of WelO5 similarity (relative to the previous halogenase result list), only the top five results were selected. These share >20% identity (range 23-34%) and >35% similarity (range 38-50%) with WelO5. As was true for the halogenases, the alignment (Figure 2.3.) shows that sequences of these hydroxylase enzymes are similar in length to WelO5 (between 270-290 residues), with the same two regions showing noticeable variability relative to WelO5 (residues 60-90 and 215-255) – however, it is important to note that the hydroxylase sequences do show greater variability overall because of the reduced levels of WelO5 homology.

Protein Identifier	Source	Annotation	% Identity shared with WelO5	% Similarity to WelO5	
1	WP_002634351.1	<i>Myxococcus hansupus</i>	2OG-Fe ²⁺ oxygenase	34	50
2	WP_002613971.1	<i>Stigmatella aurantiaca</i>	2OG-Fe ²⁺ oxygenase	33	50
3	WP_026105571.1	<i>Methylobacterium</i> sp. MB200	proline hydroxylase	23	39
4	WP_003600403.1	<i>Methylobacterium extorquens</i>	prolyl 4-hydroxylase subunit α	23	38
5	WP_012253837.1	<i>Methylobacterium</i> sp.	prolyl 4-hydroxylase subunit α	23	38

Table 2.2. WelO5 PHI-BLAST hydroxylase motif (HXD/E...H) search results. Proteins are labelled per their annotation within the NCBI protein database. Identity to, and similarity shared with WelO5 were calculated by EMBOSS Needle pairwise sequence alignment using the default BLOSUM62 substitution matrix.

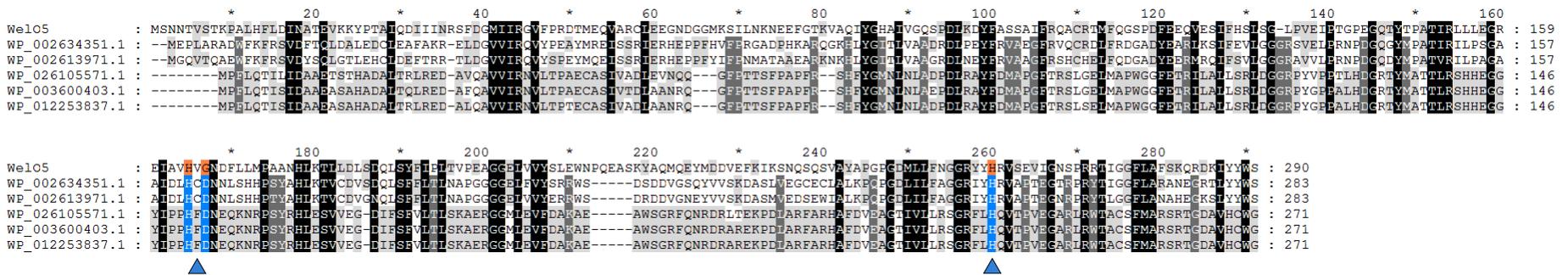


Figure 2.3. Sequence alignment of WelO5 PHI-BLAST hydroxylase motif (HXD/E...H) search results. Putative hydroxylase sequences from Table 2.2. are shown aligned to WelO5 and arranged in order of decreasing WelO5 identity & similarity. Residues are shaded using the amino acid similarity groupings DN, EQ, ST, KR, FYW and LIVM; with black indicating 100% conservation, dark grey indicating 80% conservation and light grey indicating ~40% conservation within each similarity group. The position of the HXD/E...H metal binding motif of the 2OG-dependent hydroxylases is highlighted in blue with the analogous WelO5 HXG...H halogenase motif highlighted in orange. Alignment generated by Clustal Omega and figure created in GeneDoc.

2.2.1.2. Comparison of WelO5 to enzymes of known function

The ability to develop an accurate homology model tends to correlate with the level of sequence identity shared between the template and target, with >30% identity ideally required for a reasonably accurate and credible model.¹⁰⁹⁻¹¹¹ The previously discussed BLAST searches were unable to identify any literature known Fe²⁺ and 2OG-dependent halogenases sharing $\geq 20\%$ identity with WelO5 (AmbO5 had been proposed but not confirmed as a halogenase at this time).

To investigate how different WelO5 was at the sequence level from other members of the family, an alignment was performed with literature-confirmed Fe²⁺ and 2OG-dependent halogenases (AmbO5 was also included for comparative purposes). The similarity to WelO5 across these enzymes was shown to be very low (Figure 2.4.) - of note are the enzymes SyrB2, CurA Hal and CytC3, the only family members with published crystallographic information.^{89,93,105} Due to the low-level homology (Table 2.3.), plausible models using any of these three templates could not be generated.

Description	Source	PDB IDs	% Identity shared with WelO5	% Similarity to WelO5
SyrB2	<i>Pseudomonas syringae</i>	2FCT 2FCU 2FCV	14	25
CurA Hal	<i>Lyngbya majuscula</i>	3NNF 3NNJ 3NNL 3NNM	11	20
CytC3	<i>Streptomyces sp.</i>	3GJA 3GJB	7	11

Table 2.3. Comparison of WelO5 with 2OG-dependent halogenases of known structure, SyrB2, CytC3 and CurA Hal. Values reported for identity and similarity were calculated using EMBOSS Needle pairwise alignment.

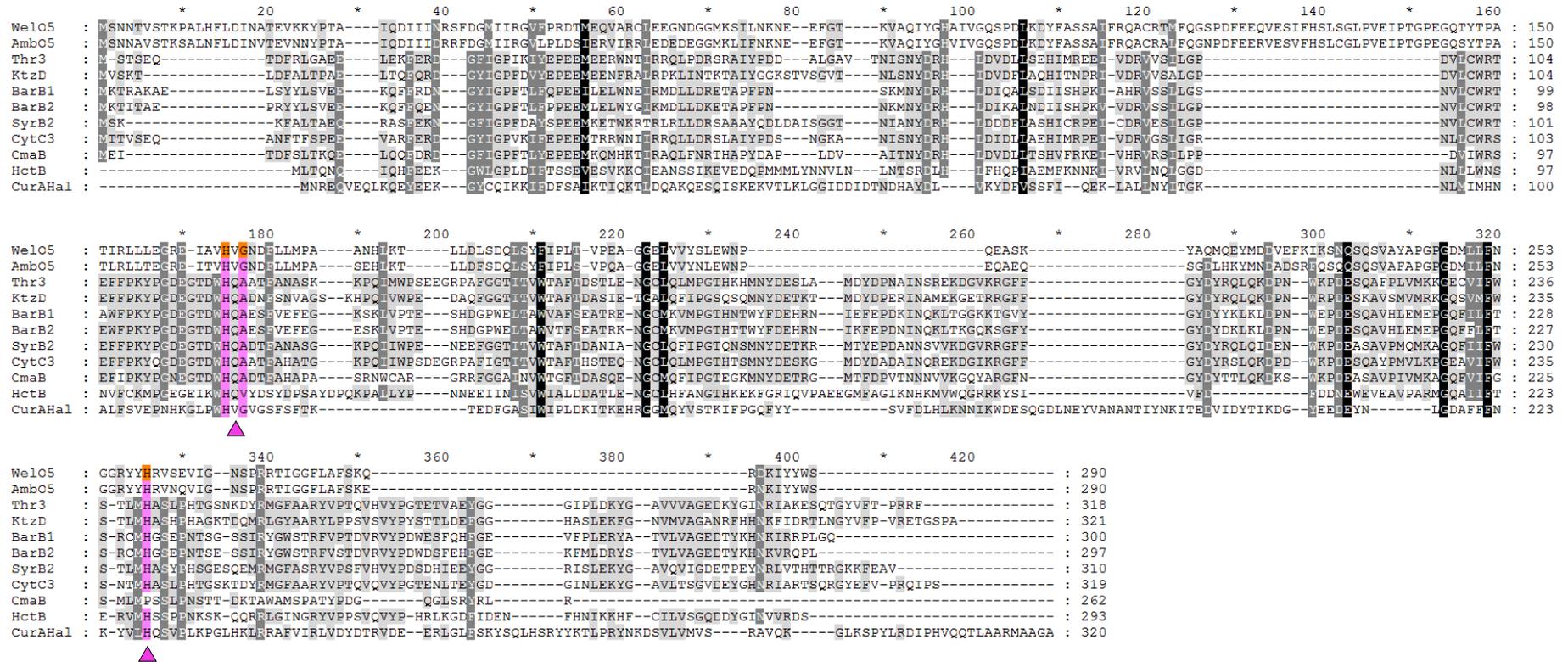


Figure 2.4. Sequence alignment of WelO5 & AmbO5 with other known 2OG-dependent halogenases. Residues are shaded using the amino acid similarity groupings DN, EQ, ST, KR, FYW and LIVM; with black indicating 100% conservation, dark grey indicating 80% conservation and light grey indicating 40% conservation within each similarity group. The position of the HXG/A...H metal binding motif of the 2OG-dependent halogenases, including AmbO5, is highlighted in pink with the analogous WelO5 motif highlighted in orange. Alignment generated by Clustal Omega and figure created in GeneDoc.

It is known that WelO5 possesses the ability to chlorinate an isolated small molecule whereas other superfamily members require a tethered protein/peptide for activity.^{50,61,87,93} AmbO5 shares a high level of identity and similarity with WelO5 (79% and 90%, respectively) and is also reported to possess the ability to chlorinate an isolated small molecule.⁵¹ This unusual activity and the low levels of homology shown by WelO5 and AmbO5 to other small molecule 2OG-dependent enzyme family members is unlikely to be a coincidence. Therefore, it seems entirely plausible that both WelO5 and AmbO5 have evolved independently from the other Fe²⁺ and 2OG-dependent halogenases identified in the literature thus far.

In the absence of three-dimensional structural information for WelO5 or AmbO5, rationalising their ability to act upon isolated small molecule substrates is extremely challenging. Whilst efforts towards the crystallisation of WelO5 were ongoing, a parallel work strand was established to try to accurately predict the structure of WelO5 computationally.

2.2.2. WelO5 tertiary structure prediction

It was initially thought possible that a protein with a reasonable level of sequence similarity to WelO5, preferably also from the Fe²⁺ and 2OG-dependent enzyme superfamily, could be used to generate a homology model. However, as previously discussed, WelO5 exhibited very low homology to all superfamily members with publicly available crystallographic structural information. The absence of structural information for homologous proteins raised significant doubts about the feasibility of creating a credible homology model for WelO5. Consideration was then given to protein structure prediction methods. Such

methods employ algorithms to computationally estimate and model a protein's tertiary structure using the primary sequence as an input.^{112,113}

Proteins sharing a similar arrangement of secondary structural elements and the same topological connections can be identified as sharing a common fold. For crystal structures within the Protein Data Bank (PDB),¹¹⁴ it has been observed that all proteins are comprised of a limited set of folds.¹¹⁵⁻¹¹⁷ As a result, protein fold recognition can be used as a tool to predict structure. Even with low sequence similarities, proteins with similar folds may demonstrate similar functions - meaning that other 2OG-dependent oxygenases could still possess a similar tertiary structure to WelO5, despite sharing very low sequence homology. By computationally aligning sections of amino acid sequence to known structural motifs/templates (instead of a global alignment), models of protein tertiary structure can be constructed from templates with low sequence similarity.¹¹² Consequently, such prediction methods can become useful in the absence of structural information for homologous proteins.

The web-interfaces for several open-source algorithms (Phyre2,¹⁰¹ RaptorX,¹⁰² SWISS-MODEL¹⁰³ & I-TASSER¹⁰⁴) were used to automate the generation of three-dimensional models of WelO5. The production of very similar looking models by different algorithms increased confidence in the predicted structures obtained. The protein SadA was chosen by several algorithms as a template for their WelO5 model, despite each computational method differing in their prediction calculations. SadA (found in rhizosphere organism *Burkholderia ambifaria* AMMD) is a bacterial Fe²⁺ and 2OG-dependent hydroxylase shown to hydroxylate several different *N*-succinyl L-amino acids.^{118,119} Interestingly, despite their shared structural similarities, SadA was found to share only 19%

sequence identity with WelO5 - which explains why SadA was not identified by previous primary sequence-reliant BLAST searches.

The main purpose of a tertiary structural model of WelO5 was to guide the selection of active site residues for variation. It was hoped that protein structure prediction methods could generate a model with sufficient accuracy to achieve this aim in the absence of a crystal structure. Due to their inherent flexibility, the structure of loop regions can prove very difficult to predict computationally.¹²⁰ However, based upon knowledge of other members of the Fe²⁺ and 2OG-dependent oxygenase family, it is extremely likely that the active site of WelO5 is sandwiched within a double-stranded beta-helix (DSBH, a.k.a. jelly-roll or cupin fold).^{63,64} As this structural motif is relatively rigid and well defined it was hoped that structural predictions for the active site would be sufficiently accurate, even if the global structural prediction was not.

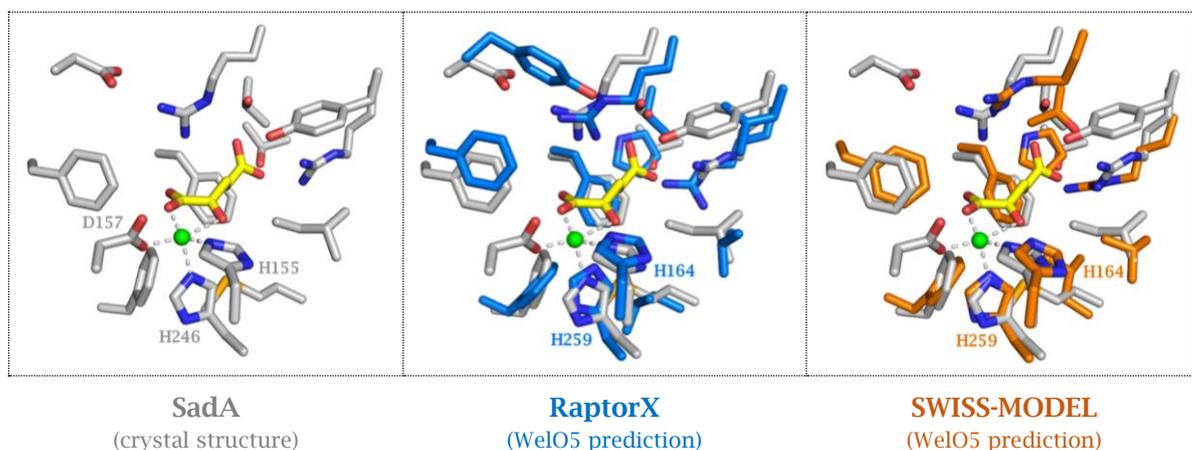


Figure 2.5. Overlays of the active site residues of SadA (PDB: 3W21) shown in grey, with those from the models of predicted WelO5 tertiary structure generated by RaptorX (shown in blue) and SWISS-MODEL (shown in orange) respectively. Fe²⁺ is represented by a green sphere and 2OG is coloured yellow. Figure generated in PyMOL.

All models generated for WelO5 were aligned in PyMOL using the key Fe²⁺ coordinating histidine ligands (H164 & H259) as anchor points and overlaid with the crystal structure of SadA. From this alignment, the active site structural

predictions generated by RaptorX and SWISS-MODEL appeared very similar to one another as well as to SadA (Figure 2.5.). Both models were therefore used to guide the selection of active site amino acid residues for variation.

2.3. WelO5 cloning, expression & purification

Reasonable quantities of high purity WelO5 were required for crystallisation trials and to enable protein characterisation and activity studies. The following section describes the cloning, expression and purification methods developed for accessing WelO5 protein.

2.3.1. Construction of WelO5 recombinant expression vector

The literature protocol for production of WelO5 involved PCR amplification from the genomic DNA of source organism *H. welwitschii*, prior to restriction enzyme digest and ligation into the pQTEV vector.⁵⁰ As this prior work resulted in publication of the WelO5 sequence, the WelO5 gene was accessed *via* commercial gene synthesis (GeneArt). The synthetic WelO5 gene was ordered codon optimised for expression in *E. coli* and sub-cloned from the supplier plasmid into the pNIC28-Bsa4 expression vector by Ligation Independent Cloning (LIC), an efficient PCR based cloning method.¹²¹ LIC was thought to offer relative simplicity, reduced cost and increased speed over the published restriction digest/ligation method - particularly for the cloning of multiple sequences in parallel (a likely prospect for this project considering the anticipated levels of mutagenesis).

The LIC protocol employed in this work was developed by the SGC.¹²² It employs the pNIC28-Bsa4 vector; a kanamycin resistant pET28a-derived *E.coli* expression vector which introduces an *N*-terminal, TEV-cleavable His₆-tag to

the protein to facilitate purification. An *N*-terminal tag was important because the *C*-terminus of other members of the Fe²⁺ and 2OG-dependent oxygenase family tends to be involved in substrate binding.⁶³ After cloning and sequence verification, the desired pNIC28-Bsa4[*WelO5*] construct (Figure 2.6.) was transformed into BL21(DE3) chemically competent cells (NEB) for *WelO5* protein expression trials.

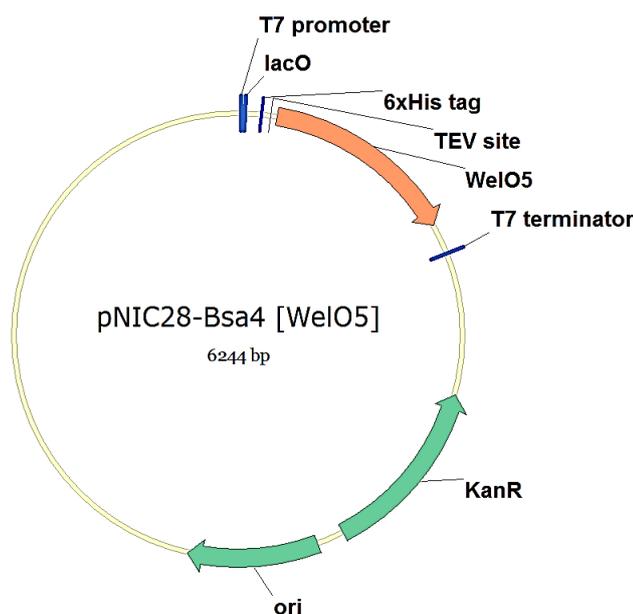


Figure 2.6. Vector map of the recombinant pNIC28-Bsa4[*WelO5*] construct. This vector incorporates an *N*-terminal His₆-tag, followed by a TEV cleavage site prior to the gene of interest (in this case, *WelO5*). Expression lies under control of the T7-lacO promoter. The vector also contains a kanamycin resistance gene. Figure generated in VectorNTI.

2.3.2. *WelO5* expression

2.3.2.1. Expression trials

Expression trials were initiated to determine the best conditions for soluble recombinant *WelO5* production by varying both the temperature at induction and expression time post-induction. Protein expression in pNIC28-Bsa4 lies under the control of the T7 promoter, meaning that gene transcription can be induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to cell cultures. For these trials, *WelO5* plasmid-containing *E. coli* were grown in 100 mL cultures at 37 °C until OD₆₀₀ ~ 0.6-0.8 was reached. Pairs of flasks were then

placed in one of three incubators set at 18 °C, 28 °C or 37 °C, respectively. A culture capable of expressing an unrelated protein from a validated plasmid was also included as a positive control. Allowing 15 min to equilibrate to their respective temperatures, the control cultures and one flask from each WelO5 expression pair were induced with 0.5 mM IPTG (the second flask was deliberately not induced to examine WelO5 baseline expression in the absence of IPTG). At 4-hr and 20-hr post-IPTG induction timepoints, analytical samples were removed from the flasks at all three temperatures (the positive control samples were collected at the 20-hr timepoint only). Each sample was pelleted in a benchtop centrifuge and then frozen at -80 °C.

For analysis, the samples were defrosted and re-suspended in BugBuster® Master Mix protein extraction reagent (Novagen). A portion of this re-suspended sample was then removed to represent the [TOTAL] protein fraction. After centrifuging, a sample of the supernatant was removed to represent the [SOLUBLE] protein fraction. Each of the IPTG induced samples were also subjected to a small-scale purification with Ni²⁺ resin which binds to the *N*-terminal His₆-tag present on the over-expressed protein to give an indication of [PURIFIED] protein levels. All fractions were analysed by SDS-PAGE to compare expression levels and the relative amounts of soluble protein (Figure 2.7.).

The positive control successfully showed overexpression of the expected protein (observed at 73 kDa) and WelO5 expression was found to be inducible at all temperatures tested. All 20-hr post-induction timepoints resulted in greater yields of protein at their respective temperatures; however, higher temperatures tended to correlate with increased proportions of insoluble

protein. Inducing the cells at 28 °C and harvesting at the 20-hr timepoint was deemed an appropriate compromise between expression levels and the proportion of soluble protein. Further optimisation (e.g. varying IPTG concentration, growth media, cell type) was not deemed necessary at this stage. These conditions were subsequently used to scale up to a 12L expression volume.

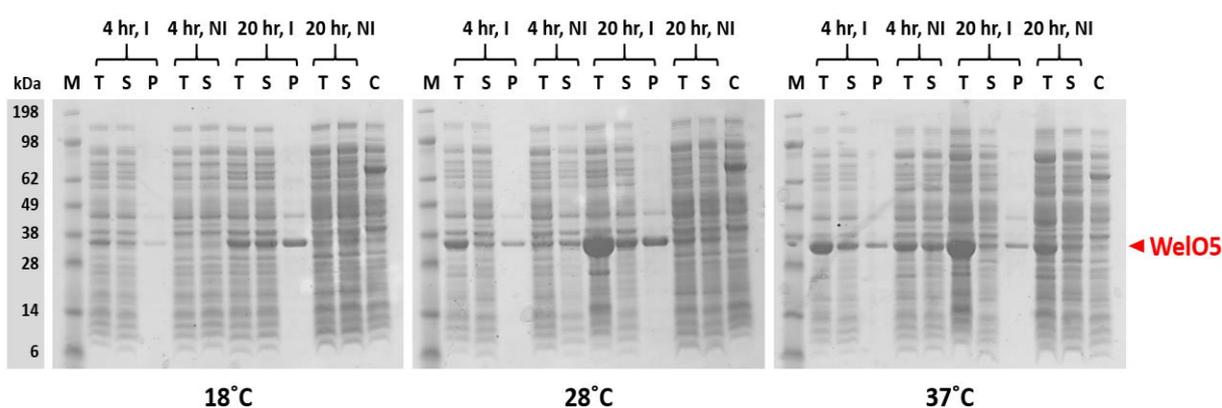


Figure 2.7. SDS-PAGE gels for expression trials performed at three different temperatures. His₆-WelO5: 35 kDa. The protein molecular weight marker is labelled M. [TOTAL] samples are labelled T, [SOLUBLE] fractions are labelled S and [PURIFIED] samples are labelled P. Samples were collected at 4-hr and 20-hr timepoints post-induction (I); with comparator non-induced samples (NI) collected simultaneously. A plasmid proven to express an unrelated protein (73 kDa) was also included in these trials to act as a positive control, labelled C.

2.3.2.2. Large scale expression of WelO5

12L of WelO5 culture was grown using the conditions selected from the previously mentioned expression trials and the cell pellet frozen at -80 °C. After purification and confirmation that the desired protein had been isolated, it was thought that optimisation and refinement of the existing protocol would be beneficial if access to multiple related proteins may be required later down the line. The potential for a fed-batch method was explored because it can produce comparable (or increased) protein yields from reduced culture volumes relative to the standard batch process.¹²³ By supplementing the cells with extra nutrients as they approach stationary phase, the log phase of growth is extended and cell densities are increased (Figure 2.8).

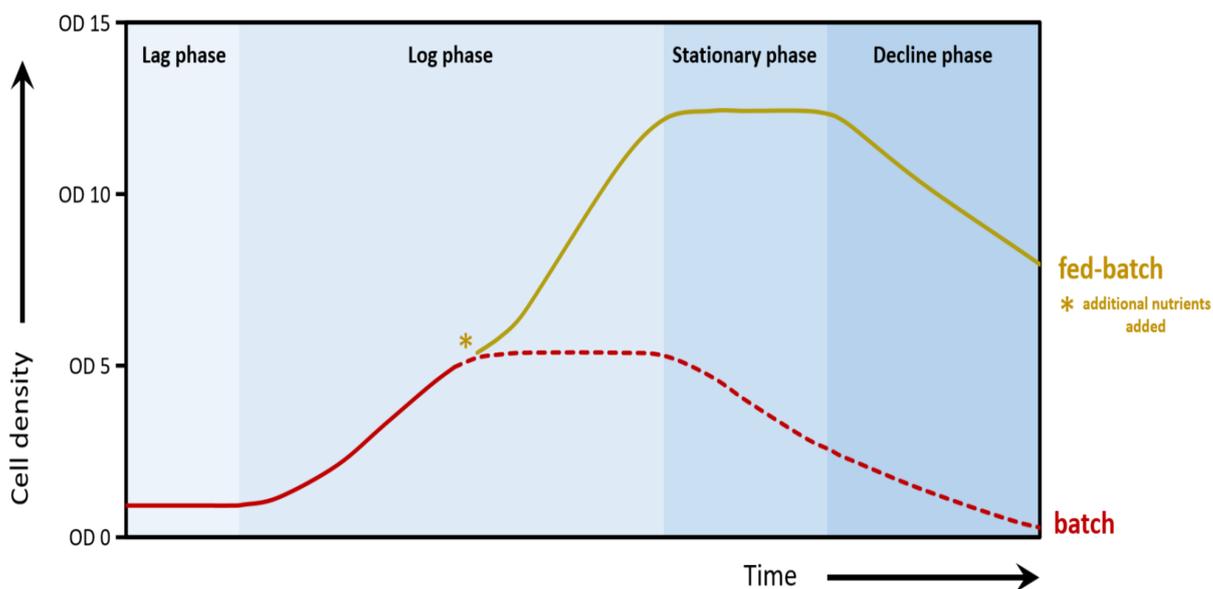


Figure 2.8. Graphical comparison of batch vs. fed-batch expression cultures. The addition of nutrients in the fed-batch protocol prolongs the log phase of growth to increase cell density. Greater yield of protein are therefore obtainable for smaller/reduced culture volumes.

For WelO5, a fed-batch procedure using a 5L culture volume produced higher yields of purified protein than the previous 12L batch expression. This protocol was therefore selected as the primary method for subsequent large-scale protein expressions.

2.3.3. WelO5 purification

The literature method for WelO5 purification was optimised to improve protein yield and purity (suitable for crystallisation).⁵⁰ An adapted immobilized metal ion affinity chromatography (IMAC) protocol for isolation of the His₆-tagged construct did not yield protein of sufficiently high purity (>95%) for crystallisation trials. However, it was found that by running a second Ni²⁺ affinity purification after cleavage of the His₆-tag removed nearly all impurities. A final gel filtration polishing step yielded 45 mg L⁻¹ WelO5 (12L expression) with sufficient purity for crystallisation trials. Using cell pellet isolated from the fed-batch expression method and condensing the initial purification protocol into fewer steps substantially increased the purified yield

to 75 mg L⁻¹ (5L expression) of WelO5. This protein was also of sufficient purity for use in crystallisation trials.

2.3.3.1. Immobilized Metal Ion Affinity Chromatography (IMAC)

IMAC¹²⁴ is commonly used as an initial protein purification technique due to the ease of optimisation, high levels of selectivity and good recoveries.¹²⁵ For purification *via* the IMAC method, the desired protein needs to have been expressed with a polyhistidine tag. Upon application of the cell lysate to the column, the tag forms a highly specific, reversible interaction with the immobilised metal ions of the chromatography matrix. Ni²⁺ ions are most commonly used in IMAC and are also used in this WelO5 protocol.¹²⁵ After application of the cell lysate, the column is washed to remove impurities before the desired tagged-protein is eluted with a buffer containing high concentrations of imidazole. Imidazole competes out the histidine residues by disrupting the reversible interaction between the protein's polyhistidine tag and the column matrix, releasing the protein from the column.¹²⁵

Two different but comparable resins (Ni-NTA agarose, Qiagen & Ni-Sepharose FF, GE Healthcare) were used interchangeably for WelO5 IMAC purification, depending on the equipment available at the time. In each case, the purity of the resulting His₆-WelO5 was found to be unsatisfactory for the set-up of crystallisation trials and therefore assumed to not be related to the choice of resin used. Minimal changes to the literature IMAC protocol were made apart from minor alterations to the composition of the buffers:

- Initial purification using beta-mercaptoethanol (BME)-containing buffers resulted in a noticeable browning of the nickel resin. This raised concerns about the possibility of reduction of the Ni²⁺ which could have

a negative impact upon resin binding capacity. BME was then exchanged for tris(2-carboxyethyl)phosphine (TCEP) which minimised this discolouration effect. TCEP was also preferable to BME as it can be used at lower concentrations, is thiol-free (i.e. odourless) and stable in solution for longer periods of time.

- Removal of the Tween-20 detergent from the buffer was found to have no negative effects and aided buffer preparation due to the elimination of foaming.

A typical IMAC elution profile for WelO5 is shown in Figure 2.9. The fractions isolated from this purification (B5-C7) were combined and concentrated in preparation for a final size-exclusion (SEC) step. However, as the His₆-WelO5 purified by SEC was not clean enough, another strategy needed to be considered.

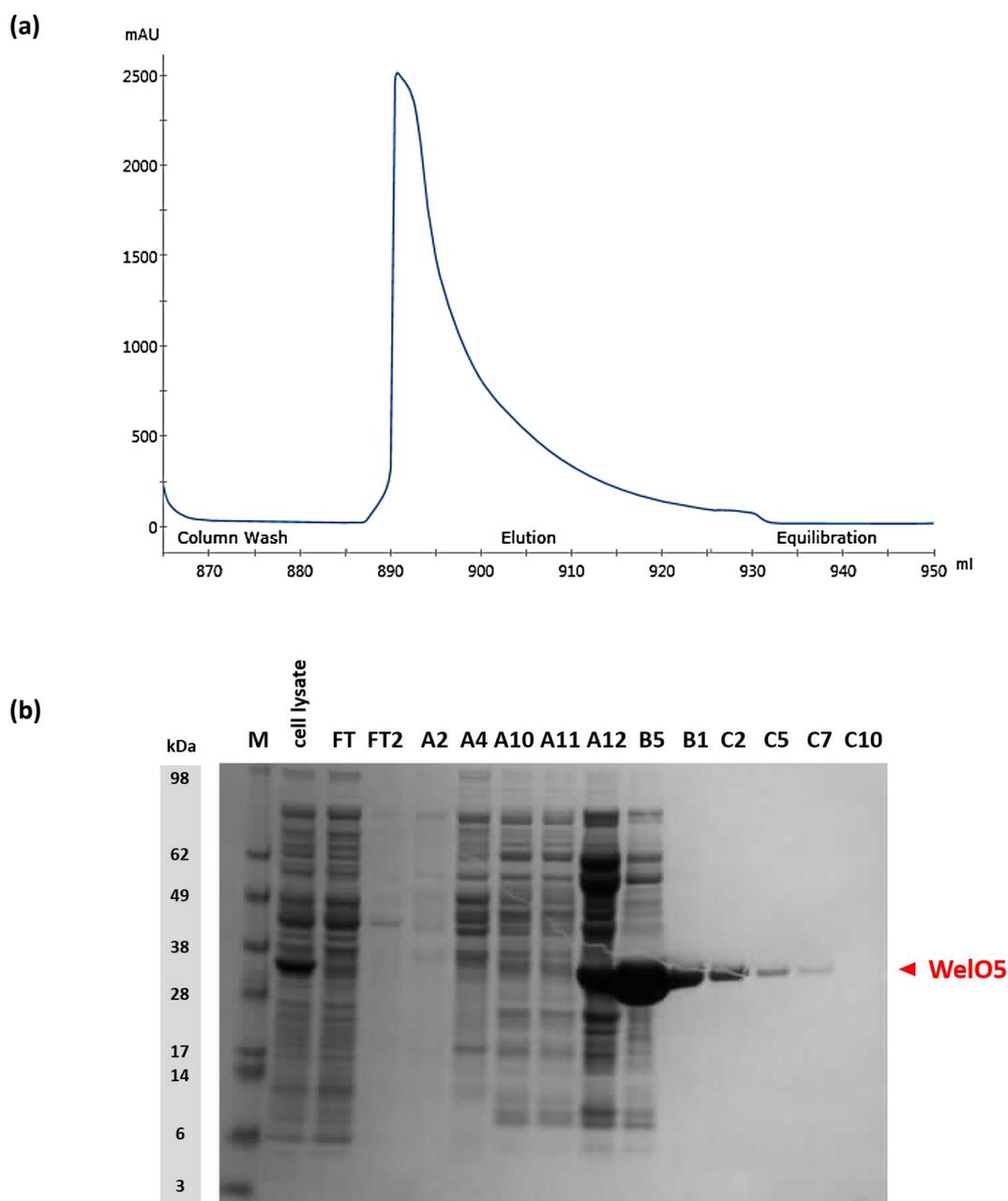


Figure 2.9. WelO5 purification by IMAC: (a) Elution profile of WelO5; (b) SDS-PAGE gel analysis of fractions A2-C10, indicating that the peak shown in the elution profile corresponds to purified His₆-WelO5 at 35 kDa. The protein molecular weight marker is labelled M. Flow-through collected during column loading is labelled FT & FT2.

2.3.3.2. Cleavage with the TEV protease

The addition of a His₆-tag cleavage step after gel filtration, followed by a second round of IMAC was found to purify the protein sufficiently for crystallisation trials. The concentrated fractions from the previous (IMAC) step were buffer exchanged to transfer the protein into the recommended reaction

buffer for TEV protease cleavage¹²⁶ and to remove the imidazole, which at high concentrations could interfere with the cleavage reaction. The reduced volume also facilitated direct loading of the cleavage reaction onto a second IMAC column. As mentioned previously, the second IMAC column was found to remove the impurities remaining from the first IMAC step - presumably because these impurities were retained by the column whilst the cleaved WelO5 was not. Later, it was found that imidazole concentration did not noticeably interfere with the cleavage reaction and the reaction could be loaded directly onto the gel filtration column and still achieve the desired purity. After analysis by SDS-PAGE (Figure 2.10.), the fractions corresponding to the cleaved WelO5 were combined and concentrated.

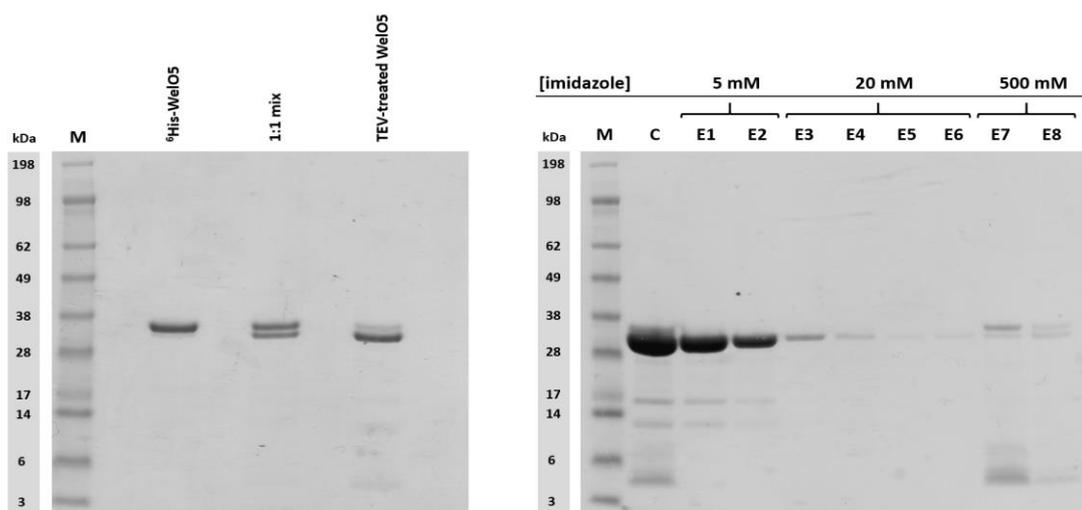


Figure 2.10. SDS-PAGE gel analysis of TEV cleavage reaction and subsequent Ni²⁺ purification. The protein molecular weight marker is labelled M. TEV protease-treated WelO5 is run alongside un-cut WelO5 and a mixture of both samples ('1:1 mix') for comparative purposes due to the minimal changes in molecular weight achieved by tag-cleavage. Crude TEV cleavage reaction (C) is run alongside elution fractions 1-8. Cut WelO5 elutes from the column at the 5mM imidazole loading concentration (fractions E1 + E2). Un-cut WelO5 and other impurities elute at higher imidazole concentrations (fractions E7 + E8).

2.3.3.3. Size-Exclusion Chromatography (SEC)

Size exclusion chromatography (or gel filtration chromatography)¹²⁷ separates proteins by molecular size and shape. It is often used as an orthogonal purification technique to IMAC because multimers and aggregates of the same

protein species can be separated, leading to sample homogeneity and higher purities.¹²⁸ The stationary phase used for SEC is comprised of a matrix of inert polymer beads, through which molecules of differing size and shape will pass at different speeds. Larger molecules which cannot enter the pores in the matrix will elute early on, whilst the ability of smaller molecules to enter the pores leads to slower movement through the matrix and to later elution times.¹²⁸

SEC was initially employed as a polishing step for the IMAC-purified His₆-WelO5. However, as previously mentioned, sufficiently clean protein was not achievable until tag-cleavage and a second IMAC column were introduced. A typical SEC elution profile is shown in Figure 2.11.(a), WelO5 SEC did not show protein multimers and the retention time of the protein eluting from the column (at approx. 235 mL) was consistent with that of WelO5 monomer (35 kDa for His₆-WelO5 or 32.5 kDa for cleaved WelO5).

2.3.3.4. Condensed purification protocol

The complete WelO5 purification protocol involved several concentration, buffer exchange and purification steps. An effort was made to condense the number of steps to increase the yields of purified protein obtained. The TEV protease used was found to have activity in the presence of imidazole, which meant exchange into the reaction buffer was not required - instead, WelO5-containing fractions from the IMAC column were minimally concentrated and the TEV protease added directly. After overnight incubation at 4 °C, the cleavage reaction could be loaded directly onto a size-exclusion column, with gel filtration acting as a combined purification and buffer exchange step. The

resulting protein appeared highly pure (by SDS-PAGE analysis) with separation from any uncleaved His₆-WelO5 and the TEV protease achieved (Figure 2.11.).

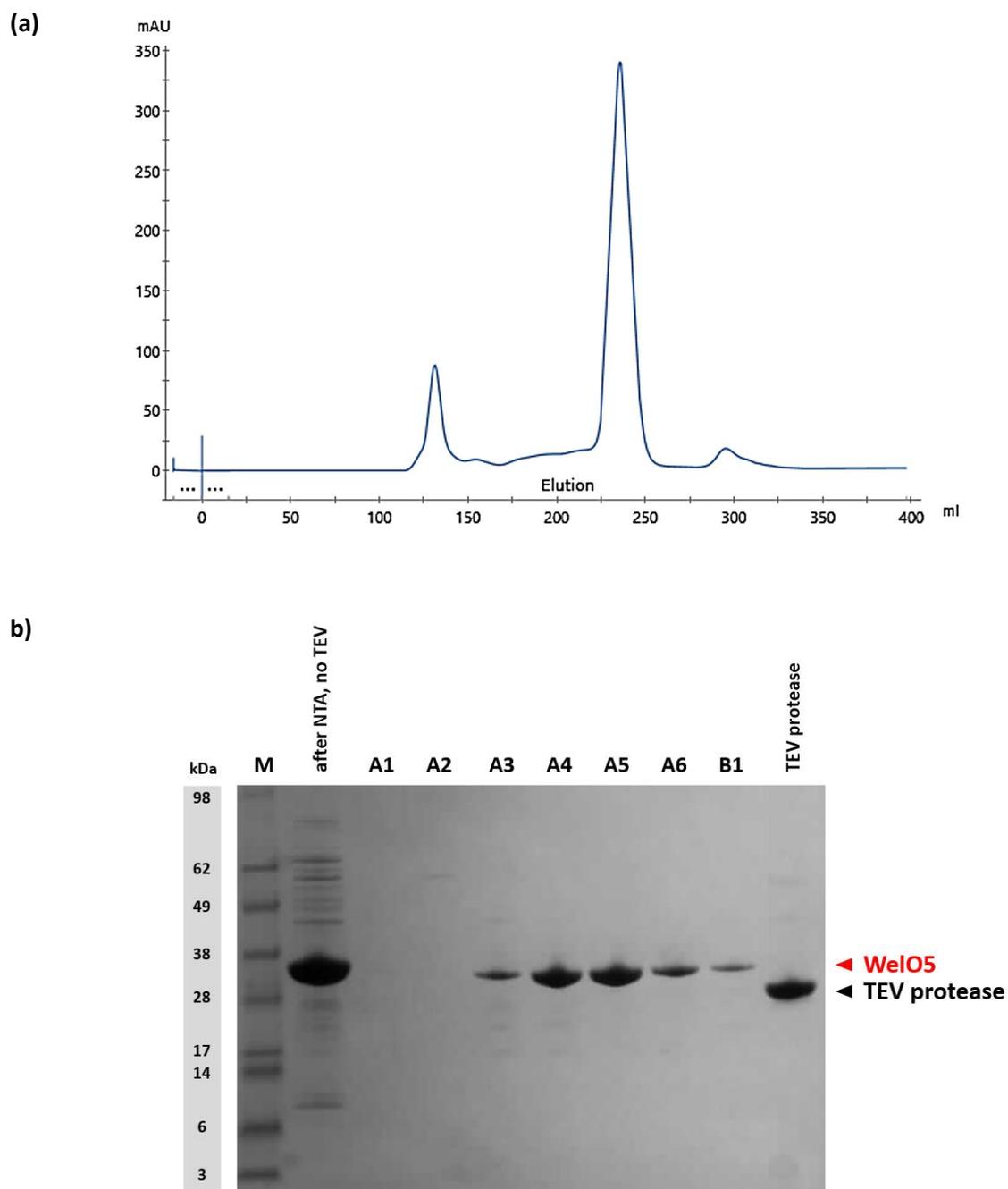


Figure 2.11. WelO5 purification by SEC: (a) Elution profile of WelO5; (b) SDS-PAGE gel analysis of fractions A1-B1, showing high purity WelO5 at 32.5 kDa and clear separation from TEV protease. Single bands indicate separation from any uncleaved His₆-WelO5 (if present) has also been achieved.

Specific comparisons between the different expression and purification protocols used to access WelO5 are detailed in Table 2.4., with Figure 2.12. acting as a graphical representation of the overall protocol optimisation process. The finalised fed-batch expression and optimised purification protocol provided 75 mg L⁻¹ of WelO5 in sufficient purity for crystallisation trials.

	Published Expression & Purification Procedure (A)	Adapted Expression & Purification Procedure (B)	Optimised Expression & Purification Procedure (C)	Fed Batch Expression & Purification (D)
Vector	pQTEV (His ₇ , ampicillin)	pNIC28-Bsa4 (His ₆ , kanamycin)		
Expression Strain	C43 (DE3)	BL21 (DE3)		
Induction	1 mM IPTG, 16 °C	0.5 mM IPTG, 28 °C		0.3 mM IPTG, 17 °C (fed-batch nutrient buffer added)
Affinity Purification	Ni-NTA (Qiagen)	HisTrap™ FF (GE Healthcare)		Ni-NTA (Qiagen)
Affinity Purification Buffers	50 mM Tris 500 mM NaCl 10 mM BME 0.1% Tween-20 pH 7.4	50 mM Tris 500 mM NaCl 10 mM BME 0.1% Tween-20 pH 7.5	50 mM Tris 500 mM NaCl 1 mM TCEP pH 7.5	10 mM PBS 500 mM NaCl 1 mM TCEP pH 7.5
[imidazole] Buffer A	20 mM	5 mM	5 mM	Imidazole free
[imidazole] Buffer B	250 mM	500 mM	500 mM	500 mM
TEV cleavage?	N	Y, manual HisTrap™ to purify		Y, load directly onto GF column
Size Exclusion?	N	Y, S200-300 mL	Y, S75-300 mL	Y, S200-300 mL
Size Exclusion Buffer	N/A	50 mM Tris 500 mM NaCl 10 mM BME pH 7.5	50 mM Tris 500 mM NaCl 1 mM TCEP pH 7.5	20 mM HEPES 150 mM NaCl pH 7.5
No. of purification steps	X 1	X 3	X 3	X 2
No. of buffer exchange/ concentration steps	X 1	X 4	X 3	X 2
Yield (mg L⁻¹)	25	15	45	75

Table 2.4. Specific details of the protein expression and purification protocol optimisation. The published method (A) was adapted (method B), and then optimised (method C) by reducing the number of steps, removing Tween-20 and exchanging BME for TCEP. Method D involves a fed-batch expression, followed by a condensed purification protocol to maximise the yields of WelO5. A comparison of the yields obtained from each method is shown in the final row of the table.

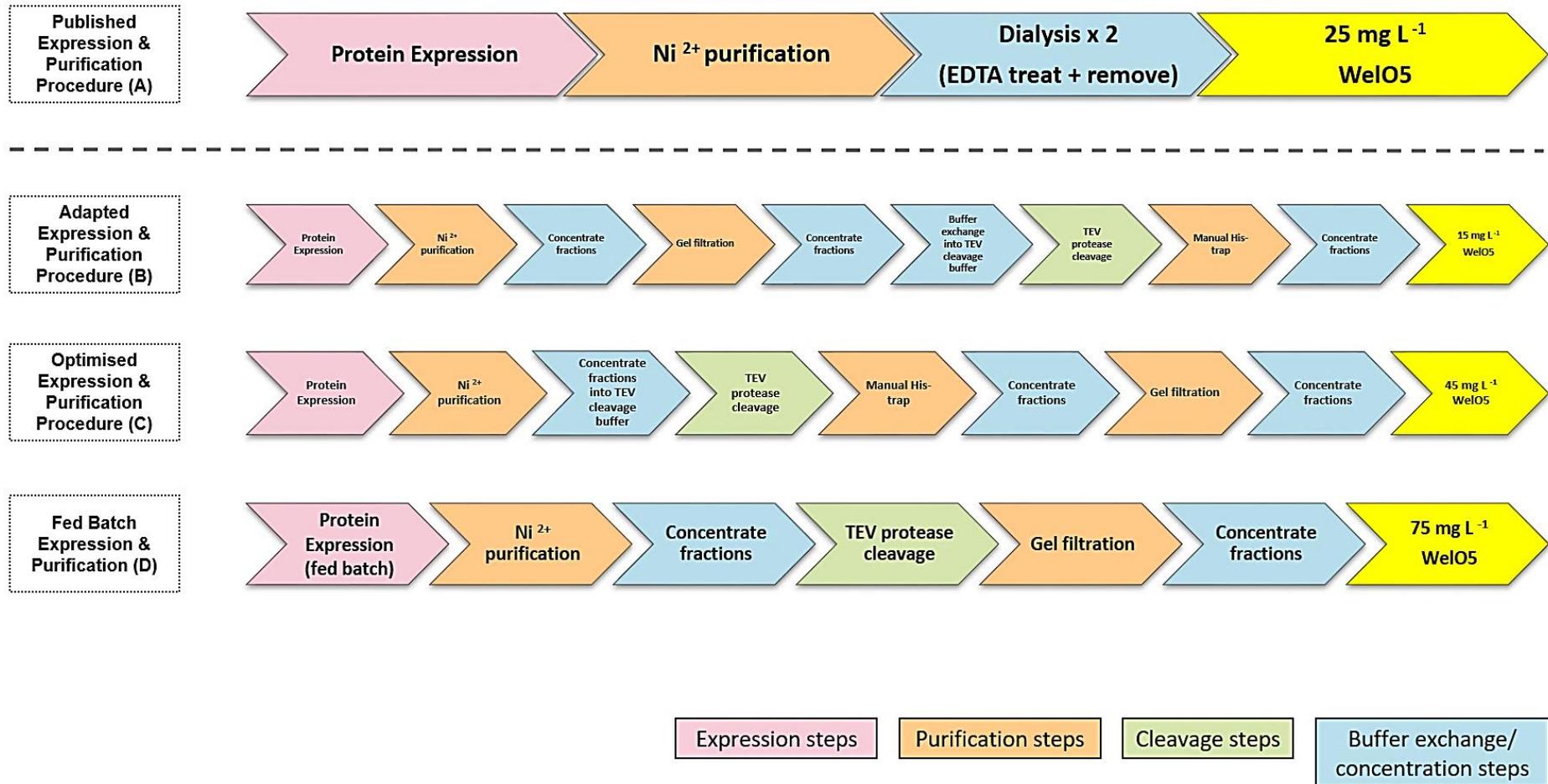


Figure 2.12. Graphical representation of the protein expression and purification optimisation performed for WelO5. Each step within the flow chart illustrates a single step within the protocol, showing a decrease in the overall number of steps from nine in the adapted literature protocol (B) to the six used for the finalised protocol (D).

2.4. Cloning & expression of WelO5 related proteins

To mitigate for potential failure in obtaining well-diffracting WelO5 crystals, the homologous proteins identified from WelO5 primary sequence analysis and previously discussed in Section 2.2.1. were investigated further. If the crystal structure of one of these homologues was faster and easier to obtain, it could act as a template for a WelO5 homology model and help provide insight into the structure of WelO5.

By leveraging technology and expertise in high-throughput protein production at the OPPF (Oxford Protein Production Facility), multiple novel constructs were generated and expressed in a parallel manner¹²⁹ – spreading the risk, maximising research efficiency and theoretically tipping the scales in favour of obtaining at least one novel crystal structure. In addition, access to a selection of pre-prepared constructs, as well as two literature-known enzymes CytC3⁹³ and SyrB2¹⁰⁵ could facilitate further research into novel Fe²⁺ and 2OG-dependent halogenases, including confirmation of function, enzyme characterisation and WelO5 comparator studies - with selection of the most suitable and well-behaved enzymes for further investigation guided by the results generated at the OPPF.

Synthetic genes (codon optimised for *E. coli* expression) were ordered for six of the homologous proteins described in Section 2.2.1.1. (Table 2.5.), as well as for the known Fe²⁺ and 2OG-dependent halogenases CytC3 and SyrB2. One or more *N*-terminal truncations were designed for each WelO5 homologue and two OPPF vectors were selected for the trials: pOPINE (*C*-terminal His₆-fusion) or pOPINF (*N*-terminal His₆-3C-fusion.¹²⁹ As the *C*-terminus of Fe²⁺ and 2OG-dependent oxygenases can be involved in substrate recognition, only full-

length proteins were included in the pOPINE selection as a general test for expression. Investigation of more complex tags was not deemed necessary based on the ease of expression and purification encountered with His₆-WelO5. The previously produced pNIC28-Bsa4[WelO5] construct was included as an expression control alongside the OPPF's GFP control, giving a total of 32 constructs - sufficient to fill one third of a 96-well plate and facilitate plate-to-plate manipulations. A summary of these constructs is included near the end of this Chapter (Table 2.6.), with full protein sequence information included in the supplementary experimental information.

NCBI Protein Identifier	Source	% Identity shared with WelO5	% Similarity to WelO5
AIH14779.1	<i>Westiella intricata</i> UH HT-29-1	99	99
AIH14734.1	<i>Hapalosiphon welwitschii</i> UH IC-52-3	95	98
AHB62755.1	<i>Fischerella ambigua</i> UTEX 1903	79	90
WP_029631942.1	<i>Scytonema hofmannii</i> UTEX B 1581	45	63
WP_012826049.1	<i>Haliangium ochraceum</i>	33	52
WP_012826042.1	<i>Haliangium ochraceum</i>	32	52
WP_004162178.1	<i>Microcystis aeruginosa</i>	32	50
WP_004162172.1	<i>Microcystis aeruginosa</i>	32	48
WP_026723823.1	<i>Fischerella</i> sp. PCC 9431	29	48
WP_004162167.1	<i>Microcystis aeruginosa</i>	29	46

Table 2.5. Table of homologous proteins taken from Section 2.2.1.1.; proteins selected for further work at the OPPF are highlighted in yellow.

The Clontech In-Fusion® method, a ligation-independent cloning method analogous to that used in Section 2.3.1., was used to insert the commercially supplied synthetic genes into the designated OPPF vectors.¹²⁹

- Inserts of the desired genes were generated by PCR, DpnI treated to digest any remaining unwanted template vector and then purified.¹³⁰

Analysis of the products on an agarose gel (Figure 2.13.) confirmed that all reactions produced a single band consistent with the expected insert

sizes (750-1000 bp; see Table 2.6.), with sufficient quantities for the subsequent transformation step.

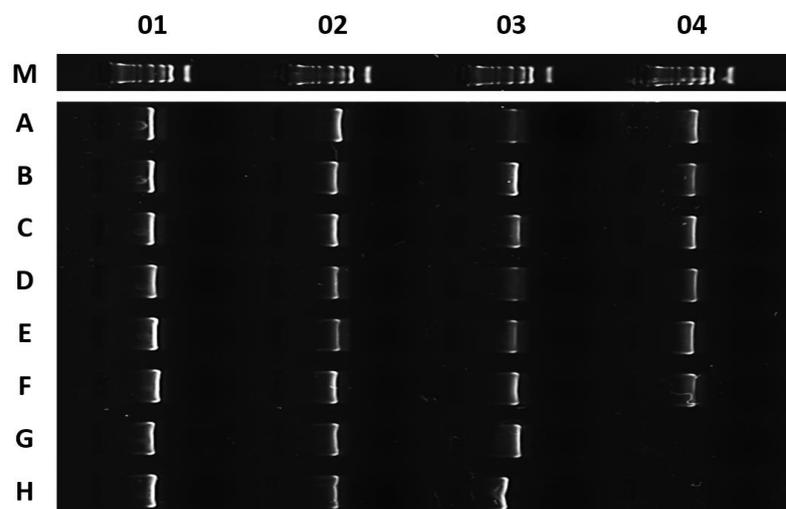


Figure 2.13. Agarose gel analysis of PCR-generated gene inserts. Analysis was performed in a multiple lane, high-throughput format. Gel markers were loaded in lanes M01-M04 and the purified PCR products loaded from the purification plate to their corresponding lanes using a multi-channel pipette. All products found to be present and consistent with expected insert size. Lanes G04 and H04 are empty as they will be later occupied by the WelO5 and GFP expression controls.

- The purified inserts were added to the linearized pOPINX vectors and incubated for 30 min at 42 °C before transformation into OmniMax II cells, spread over agar plates and incubated overnight.
- From these plates, three colonies were picked for each construct, grown overnight and mini-prepped. The constructs were verified by PCR with the relevant pOPINX primers and the reaction products analysed by agarose gel. For each construct, at least one clone was found to produce a band of the desired size (Figure 2.14.). Using these results, one verified plasmid for each construct was selected for use in expression trials.

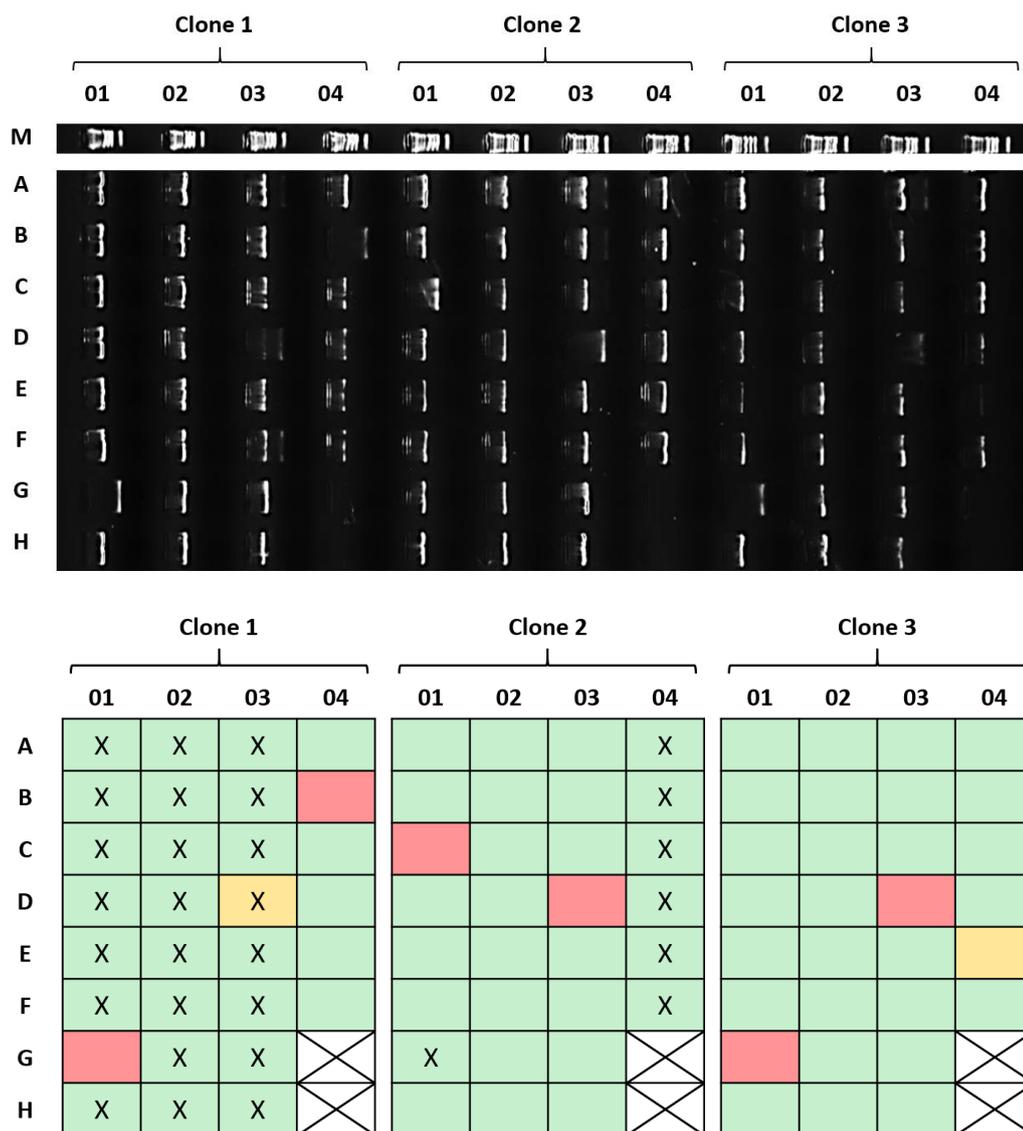


Figure 2.14. Construct verification: agarose gel showing PCR products (A01-F04) from each of the three clones selected for each construct. Lanes G04 and H04 are empty as they will be later occupied by the WelO5 and GFP expression controls. Construct verification results are summarised in the table as follows: green indicates formation of a band of the expected size, red indicates a reaction failure and amber indicates detection of a faint band of the desired size. The plasmids selected for subsequent expression trials are marked with an X.

Expression trials were set up using four different conditions resulting from a combination of two different cell types (Lemo™ 21(DE3), NEB and Rosetta™ 2(DE3) pLacI, Novagen) with two different induction conditions (IPTG or auto-induction media). After overnight incubations, the harvested expression cultures were isolated by small scale Ni²⁺ purification and analysed by SDS-PAGE to compare expression levels. Table 2.6. provides a tabulated overview of the cloning and expression work carried out at the OPPF.

Well	Gene Identifier	Description	No. amino acids	No. amino acids (includes His ₆ -tag)	Number of base pairs	Protein MW (Da, estimated)	OPPF Vector	Cloning			Expression Trial				
								PCR	Construct Verification			Lemo21		Rosetta	
									Clone 1	Clone 2	Clone 3	IPTG	AI	IPTG	AI
A01	CytC3	N-His, full length gene	318	337	1011	37070	pOPINF	X							
B01	SyrB2	N-His, full length gene	309	328	984	36080	pOPINF	X							
C01	AmbO5	N-His, full length gene	315	334	1002	36740	pOPINF	X							
D01	AmbO5	N-His, N-terminal truncation A	290	309	927	33990	pOPINF	X							
E01	AmbO5	N-His, N-terminal truncation B	272	291	873	32010	pOPINF	X							
F01	AmbO5	N-His, N-terminal truncation C	252	271	813	29810	pOPINF	X							
G01	WP1942	N-His, full length gene	294	313	939	34430	pOPINF		X						
H01	WP1942	N-His, N-terminal truncation A	271	290	870	31900	pOPINF	X							
A02	WP1942	N-His, N-terminal truncation B	251	270	810	29700	pOPINF	X							
B02	WP2172	N-His, full length gene	284	303	909	33330	pOPINF	X							
C02	WP2172	N-His, N-terminal truncation A	269	288	864	31680	pOPINF	X							
D02	WP2172	N-His, N-terminal truncation B	250	269	807	29590	pOPINF	X							
E02	WP2172	N-His, N-terminal truncation C	243	262	786	28820	pOPINF	X							
F02	WP3823	N-His, full length gene	263	282	846	31020	pOPINF	X							
G02	WP3823	N-His, N-terminal truncation A	247	266	798	29260	pOPINF	X							
H02	WP3823	N-His, N-terminal truncation B	232	251	753	27610	pOPINF	X							
A03	WP6042	N-His, full length gene	285	304	912	33440	pOPINF	X							
B03	WP6042	N-His, N-terminal truncation A	274	293	879	32230	pOPINF	X							
C03	WP6042	N-His, N-terminal truncation B	254	273	819	30030	pOPINF	X							
D03	WP6042	N-His, N-terminal truncation C	245	264	792	29040	pOPINF	X							
E03	WP6049	N-His, full length gene	263	282	846	31020	pOPINF	X							
F03	WP6049	N-His, N-terminal truncation A	245	264	792	29040	pOPINF	X							
G03	WP6049	N-His, N-terminal truncation B	224	243	729	26730	pOPINF	X							
H03	AmbO5	C-His, full length gene	315	323	969	35530	pOPINE	X							
A04	AmbO5	C-His, N-terminal truncation A	290	298	894	32780	pOPINE		X						
B04	WP1942	C-His, full length gene	294	302	906	33220	pOPINE		X						
C04	WP2172	C-His, full length gene	284	292	876	32120	pOPINE		X						
D04	WP3823	C-His, full length gene	263	271	813	29810	pOPINE		X						
E04	WP6042	C-His, full length gene	285	293	879	32230	pOPINE		X						
F04	WP6049	C-His, full length gene	263	271	813	29810	pOPINE		X						
G04	GFP control	control	-	-	-	27000	N/A	N/A		X					
H04	WetO5 control	control	290	312	936	34320	N/A	N/A		X					

Table 2.6. Summary of the high-throughput molecular biology work carried out at the Oxford Protein Production Facility (OPPF). The selected eight genes and their specific truncated forms are described (number of amino acids, number of base pairs and molecular weight), alongside the OPPF vectors they were cloned into. Bands consistent with the correct construct (from analysis by agarose and SDS-PAGE gel) are colour-coded green, orange colour-coding indicates that isolation was inconclusive (weak bands, poor purity etc.) and red colour-coding indicates failure. The pOPINE vector adds an additional nineteen amino acids to the protein of choice, whilst the pOPINE vector adds eight. The molecular weights were estimated by multiplication of the number of base pairs by 110 (the weighted average amino acid molecular weight). The 30 designed constructs were expressed alongside GFP and WetO5 controls.

Overall, expression levels were generally moderate to poor depending on the construct (Table 2.6.). The internal OPPF GFP control showed good expression for all four conditions used, indicating a successful trial. However, the WelO5 control demonstrated poor expression for both cell types under the IPTG induction conditions. This was despite the high expression levels seen previously for analogous WelO5 IPTG induction conditions (Section 2.3.2.). The known halogenases CytC3⁹³ and SyrB2¹⁰⁵ also showed sub-optimal expression levels under the same OPPF conditions. There are several differences between the conditions used for these expression trials and those used in the literature for WelO5, SyrB2 and CytC3 and one or more of these differences could have contributed to the moderate to poor expressions seen.

The poor expression levels achieved at the OPPF were surprising considering how well WelO5 had been previously expressed. It was anticipated that the WelO5-homologous halogenases (AmbO5, WP1942, WP2172, WP3823, WP6042, WP6049) and their truncated analogues may express better when using the conditions specifically developed for WelO5. An additional expression trial was then initiated, using a limited set of OPPF constructs to see if better expression levels could be achieved with conditions closer to those used previously for WelO5. SyrB2 and CytC3 were also included.

To limit numbers, only the full-length *N*-terminal His₆-tag (pOPINF) constructs were selected for the trial. 'AmbO5 *N*-terminal truncation A' (Table 2.6., ID: D01) was selected over its full-length analogue because it had demonstrated relatively good expression levels at the OPPF and, based on sequence analysis and alignment with WelO5, was already suspected to be the actual sequence for AmbO5 (the published sequence for AmbO5 included an extra 26 *N*-

terminal amino acids). This *N*-terminal truncation was eventually confirmed as the correct AmbO5 sequence by an addendum within the literature and updated within publicly available databases.⁵¹

All full-length *N*-terminal His₆-tag (pOPINF) constructs were sequenced prior to the expression trial to confirm identity and all constructs, apart from WP1942, were confirmed to be correct. As shown in Figure 2.14., only one full-length construct of approximately the correct size for full-length WP1942 (ID: G01) was identified by PCR verification (clone 2). However, sequence analysis revealed the insertion of an extra guanine base after the His₆-tag, moving the desired WP1942 sequence out of frame. This frame shift mutation would have been introduced by a mis-priming event during the gene-insert generation by PCR. As a complete repeat of the LIC protocol would be required for the correct full-length WP1942 construct, this target was temporarily set aside and not carried through into the mini expression trial.

The remaining seven pOPINF constructs (CytC2, SyrB2, AmbO5, WP2172, WP3823, WP6042, WP6049) were expressed alongside pNIC28-Bsa4[WelO5] in a plate-based format using a protocol modified from that used previously (Section 2.3.2.). As before, the cell pellets were harvested and re-suspended in BugBuster® Master Mix protein extraction reagent (Novagen). A portion of this re-suspended sample was removed to represent the [TOTAL] protein fraction. After centrifuging, a sample of the supernatant was removed to represent the [SOLUBLE] protein fraction and the remaining clarified supernatant purified with Ni²⁺ resin to give the [PURIFIED] samples. The [TOTAL], [SOLUBLE] and [PURIFIED] samples for each protein were then analysed by SDS-PAGE to compare expression levels (Figure 2.15.).

WelO5 expressed well under these conditions, as expected. Bands of comparable size can be seen for WelO5 in the soluble fraction (S) and in the total cell lysate (T), showing that the majority of the WelO5 protein produced is in a soluble form. We know that we can isolate good yields of purified WelO5 despite what the purified band indicates - the small band size is most likely due to the quantity of protein exceeding the binding capacity and/or quantity of Ni²⁺ resin used.

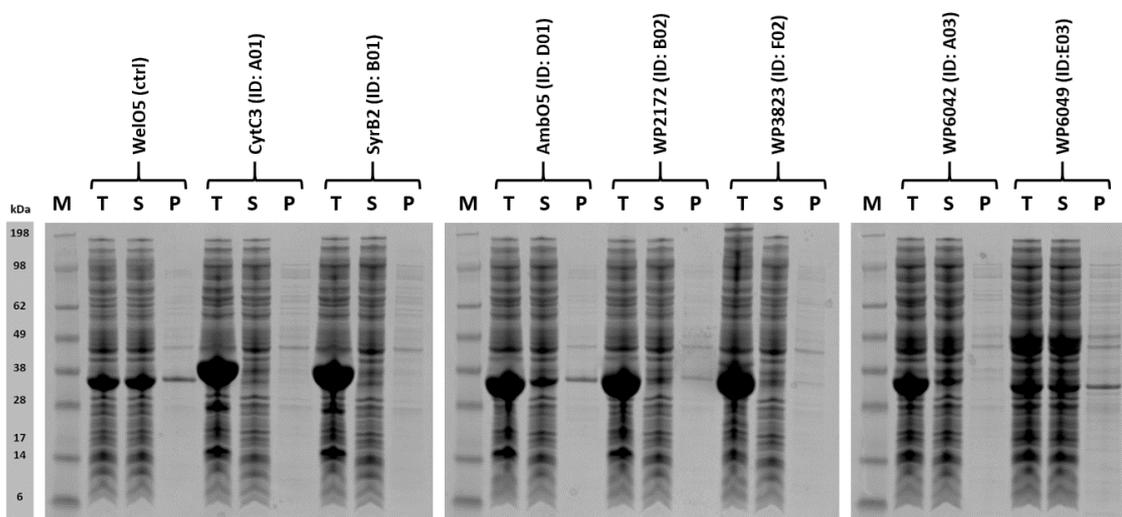


Figure 2.15. SDS-PAGE gel illustrating results of the mini-expression trial of full-length WelO5-homologue constructs performed under conditions previously optimised for WelO5 expression. Proteins were expressed in auto-induction media over 20 hr at 30 °C. The protein molecular weight marker is labelled M. For each sample the total cell extract was run alongside the soluble fraction and the purified protein to examine overexpression levels for each protein. [TOTAL] samples are labelled T, [SOLUBLE] fractions are labelled S and [PURIFIED] samples are labelled P.

In contrast, five of the seven pOPINF constructs (CytC3, SyrB2, WP2172, WP3823 & WP6042) yielded exclusively insoluble proteins. For AmbO5 and WP6049 it was possible to isolate some soluble protein. Clearly the conditions for expression of these OPPF constructs require further optimisation. As only the purified fractions were analysed in the initial OPPF expression trial, the formation of insoluble protein as seen here could explain why we saw very little purified protein on those gels. Further optimisation work was put on hold, pending the outcome of ongoing WelO5 crystallisation trials.

2.5. Summary of the production of WelO5 & related proteins

In this Chapter, the primary sequence of WelO5 was analysed and compared to characterised enzymes of known, related function. WelO5 was found to share low homology with all structurally solved Fe²⁺ and 2OG-dependent oxygenase enzymes. Following a search of the wider literature by PHI-BLAST, proteins displaying greater similarity to WelO5 were identified. However, due to a lack of three-dimensional structural information for any of these homologues, the development of a homology model was not possible. Instead, these proteins were produced in parallel at the OPPF with a view to providing alternative options if well-diffracting WelO5 crystals could not be obtained. Additionally, a successful protocol was developed for production of suitable quantities of high purity WelO5 protein for further studies, including crystallisation trials. Whilst awaiting the growth of well-diffracting WelO5 crystals, open-source structure prediction algorithms were evaluated for their ability to produce a three-dimensional representation of the WelO5 active site. Such models were used as an interim solution to guide enzyme evolution by active-site residue substitution.

Chapter 3. Synthesis of WelO5 endogenous substrate and substrate analogues

3.1. Introduction

In 1994, fifteen complex natural products from the extracts of cyanobacterium *H. welwitschii* (W. & G.S. West UH strain IC-52-3) were isolated and structurally assigned (Figure 3.1.).¹³¹

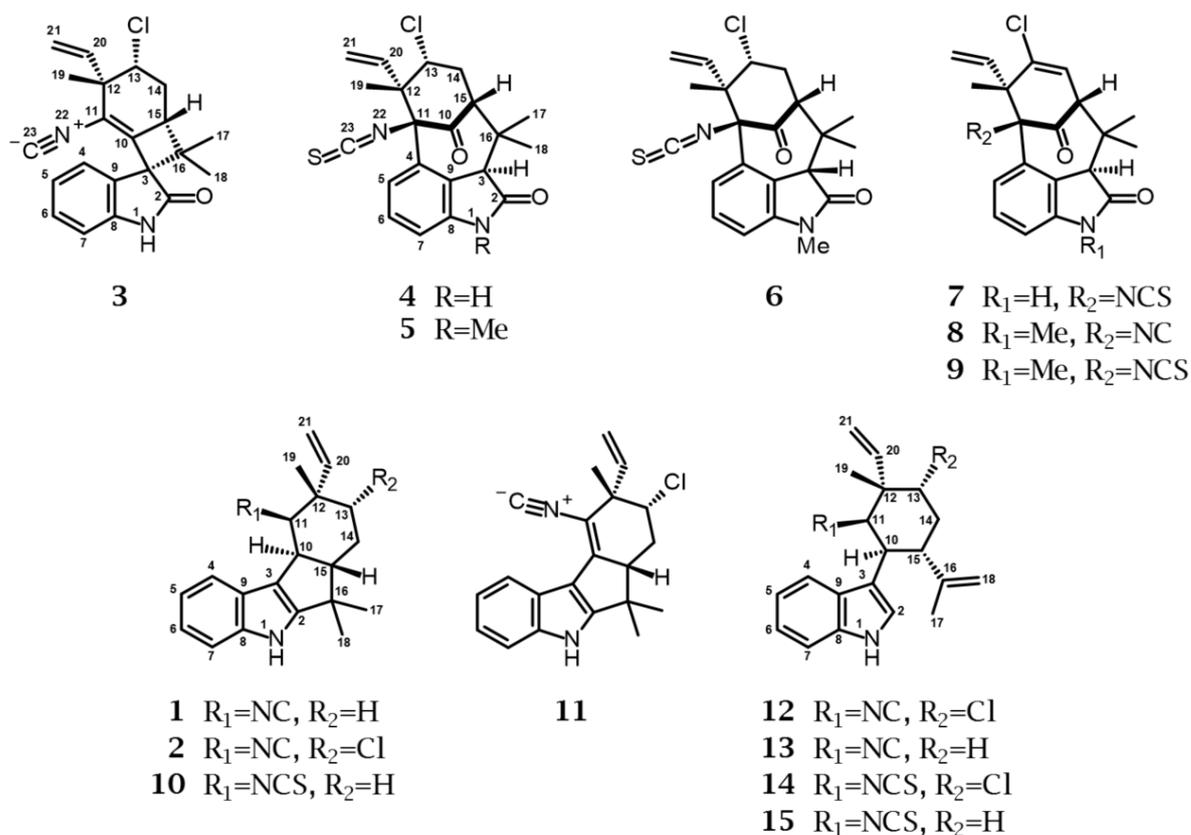


Figure 3.1. Structures of the fifteen compounds isolated from *H. welwitschii* (W. & G.S. West UH strain IC-52-3).¹³¹ Structures are numbered as per the traditional hapalindole labelling convention.^{132,133} The same numbering system is used for all compounds of this series within the literature and within this thesis.

Of these isolated compounds, four were fischerindoles (Figure 3.1., compounds 1, 2, 10 & 11), proposed to be formed from their analogous acyclic hapalindole precursors *via* an enzyme-controlled, acid-catalysed cyclisation of the isopropenyl group onto the 2-position of the indole.¹³¹ The acyclic hapalindole

precursors themselves, 12-*epi*-hapalindole E isonitrile (**12**) and 12-*epi*-hapalindole C isonitrile (**13**), were proposed to be formed from reactive intermediates derived from tryptophan and geraniol phosphate in the presence of either a chloronium or hydrogen ion (Figure 3.2).^{131,133}

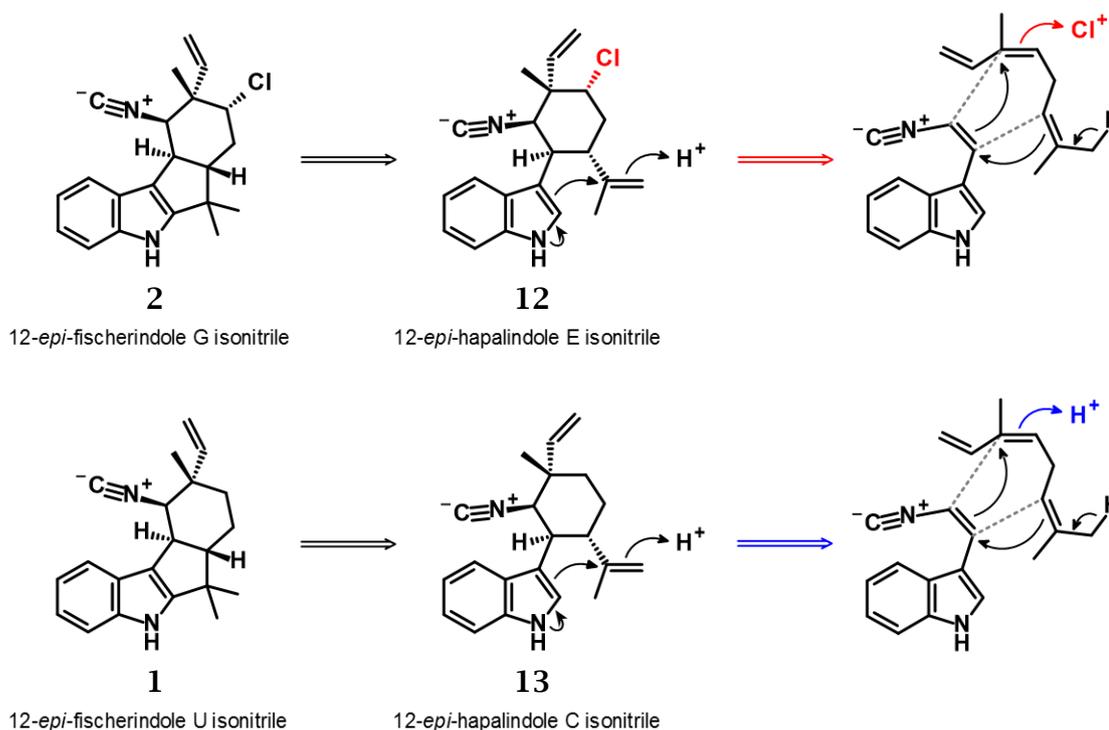


Figure 3.2. Proposed biosynthetic route to *H. welwitschii* fischerindoles, as proposed by Moore *et al.*¹³¹

The hypothesis that chlorine is introduced at the start of the biosynthetic route to these fischerindole species was not disproven until some 20 years later, following the identification and functional assignment of the Fe²⁺ and 2OG-dependent halogenase, WelO5.⁵⁰ Instead of being derived from compound **12**, compound **2** was experimentally shown to form from compound **1** *via* a WelO5-catalysed, late-stage regio- and stereo-specific C-13 chlorination event (Figure 3.3).⁵⁰ The report of this synthetically interesting reaction raised the question as to whether WelO5-type halogenations could be performed upon alternative substrates.

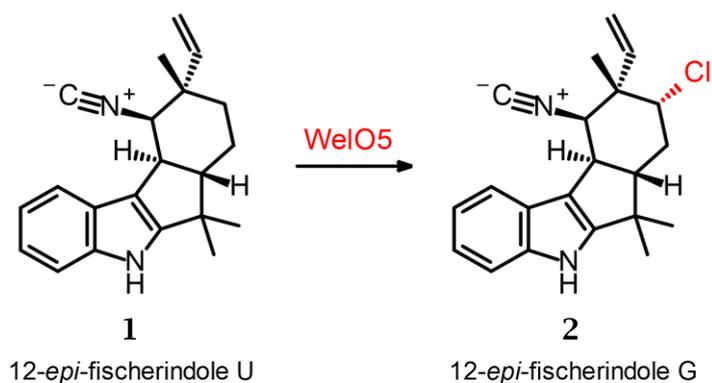


Figure 3.3. WelO5-catalysed formation of 12-*epi*-fischerindole G by regio- and stereo-specific late-stage C-13 chlorination of (+)-12-*epi*-fischerindole U isonitrile.

With a successful protocol in hand for the production of large quantities of high purity, wild-type WelO5 (Chapter 2, Section 2.3.), it was clear that access to the enzyme's natural substrate was key to confirming activity of the purified WelO5. Access to **1** could also enable crystallisation studies (soaking, co-crystallisation) and assay development for evaluation of the activity of wild-type WelO5 and its variants against alternative substrates.

Potential approaches for accessing 12-*epi*-fischerindole U isonitrile, **1** were evaluated:

- **1** and any highly similar precursors/derivatives were unavailable commercially (this is perhaps not unsurprising considering their low natural abundance).¹³¹
- The large-scale fermentation protocol used in the literature to access **1** was determined physically impractical, in part because it was reported to be low yielding (≤ 1 mg of **1** was isolatable from a 20L algal culture volume).⁵⁰ The necessary experimental work for this project (particularly co-crystallisation/crystal soaking attempts) was anticipated to require much greater quantities of **1** than this fermentation process could feasibly provide.

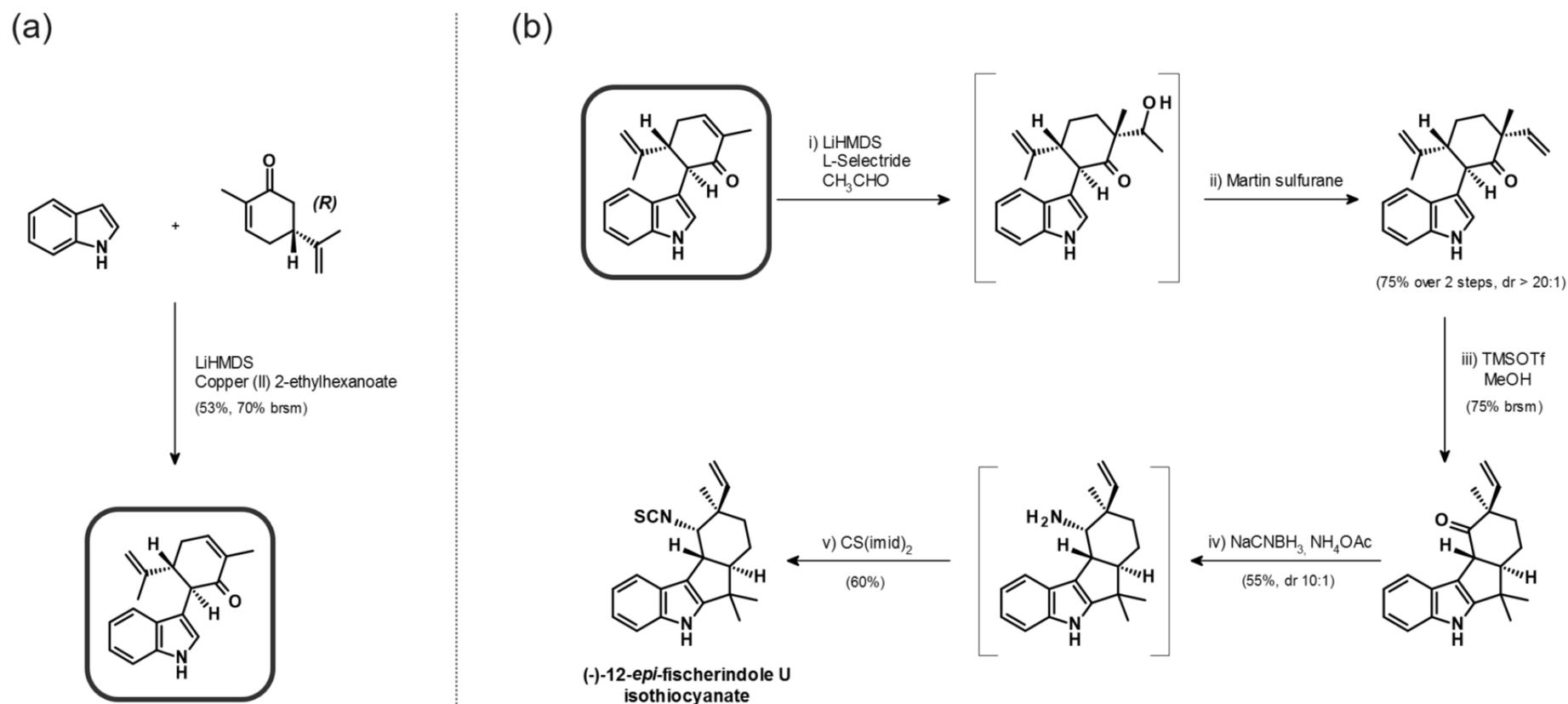
- Given the complex structure of **1**, a polycyclic indole with multiple stereocentres, access *via* organic synthesis was initially ruled out on the grounds of synthetic complexity and the predicted time it would take to evaluate and work through a suitable route. However, a subsequent search of the literature found a relatively recent publication with a concise synthesis to an almost identical compound, (-)-12-*epi*-fischerindole U isothiocyanate.¹³⁴ This route appeared to be amenable to adaptation for the generation of **1**. The synthetic intermediates from this route could also provide additional value if WelO5 was found to possess halogenation activity for any of them.

This Chapter outlines the modifications and revisions made to the published synthetic route to (-)-12-*epi*-fischerindole U isothiocyanate for the generation of milligram quantities of the desired WelO5 endogenous substrate, (+)-12-*epi*-fischerindole U isonitrile, **1**. The catalytic activity of wild-type WelO5 produced by the newly developed protocol (Chapter 2, Section 2.3.) is confirmed using this synthetically produced substrate and results from the incubation of WelO5 with simple substrate analogues and intermediates isolated during substrate synthesis are also described.

3.2. Synthesis of endogenous substrate

The literature route to (-)-12-*epi*-fischerindole U isothiocyanate involves a five step synthesis from cheap & readily available starting materials (Scheme 3.1.).¹³⁴ This route employs a novel copper-catalysed indole-carvone coupling reaction, followed by the addition of acetaldehyde at C-12, subsequent dehydration to install the alkene, and a biomimetic acid-catalysed cyclisation to create the carbocyclic scaffold.¹³⁴ Manipulation of the ketone moiety by

reductive amination, followed by treatment of the resulting amine with thiocarbonyl diimidazole (CDI) yielded the isothiocyanate in an overall yield of 22%.¹³⁴



Scheme 3.1. Baran's total synthesis of (-)-12-*epi*-fischerindole U isothiocyanate (2004).¹³⁴ Reagents and conditions: **(a)** indole (2.0 equiv.), *R*-carvone (1.0 equiv.), LiHMDS (3.0 equiv.), THF, -78 °C, 30 min; then copper(II)2-ethylhexanoate (1.5 equiv.), -78 °C, 12 h, **(b)** (i) LiHMDS (1.5 equiv.), THF, -78 °C, 20 min; then *L*-Selectride (1.05 equiv.), -78 °C, 1 h; then CH₃CHO (6.0 equiv.), -78 °C, 15 min; (ii) Martin Sulfurane (1.1 equiv.), CHCl₃, rt, 10 min; (iii) TMSOTf (3.0 equiv.), MeOH (1.1 equiv.), DCM, 0 °C, 1 h; (iv) NaCNBH₃ (10 equiv.), NH₄OAc (40 equiv.), MeOH/THF, rt, 7 d; (v) CS(imid)₂ (1.1 equiv.), DCM, rt, 4 h.

To access WelO5 substrate **1**, modifications were made to the final step of the published route to yield the analogous isonitrile instead of the isothiocyanate. The specific rotation of the synthetically produced 12-*epi*-fischerindole U isothiocyanate was given as $[\alpha]_D -200$ (DCM, c 0.020),¹³⁴ identifying itself as the opposite enantiomer to that isolated from *H. welwitschii*, $[\alpha]_D +231$ (DCM, c 0.035)(Figure 3.4).¹³¹

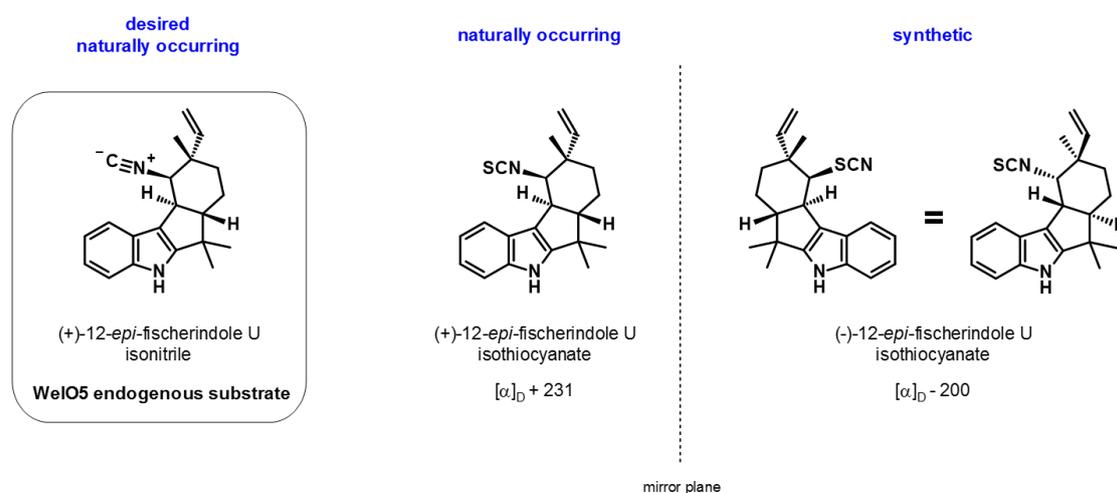
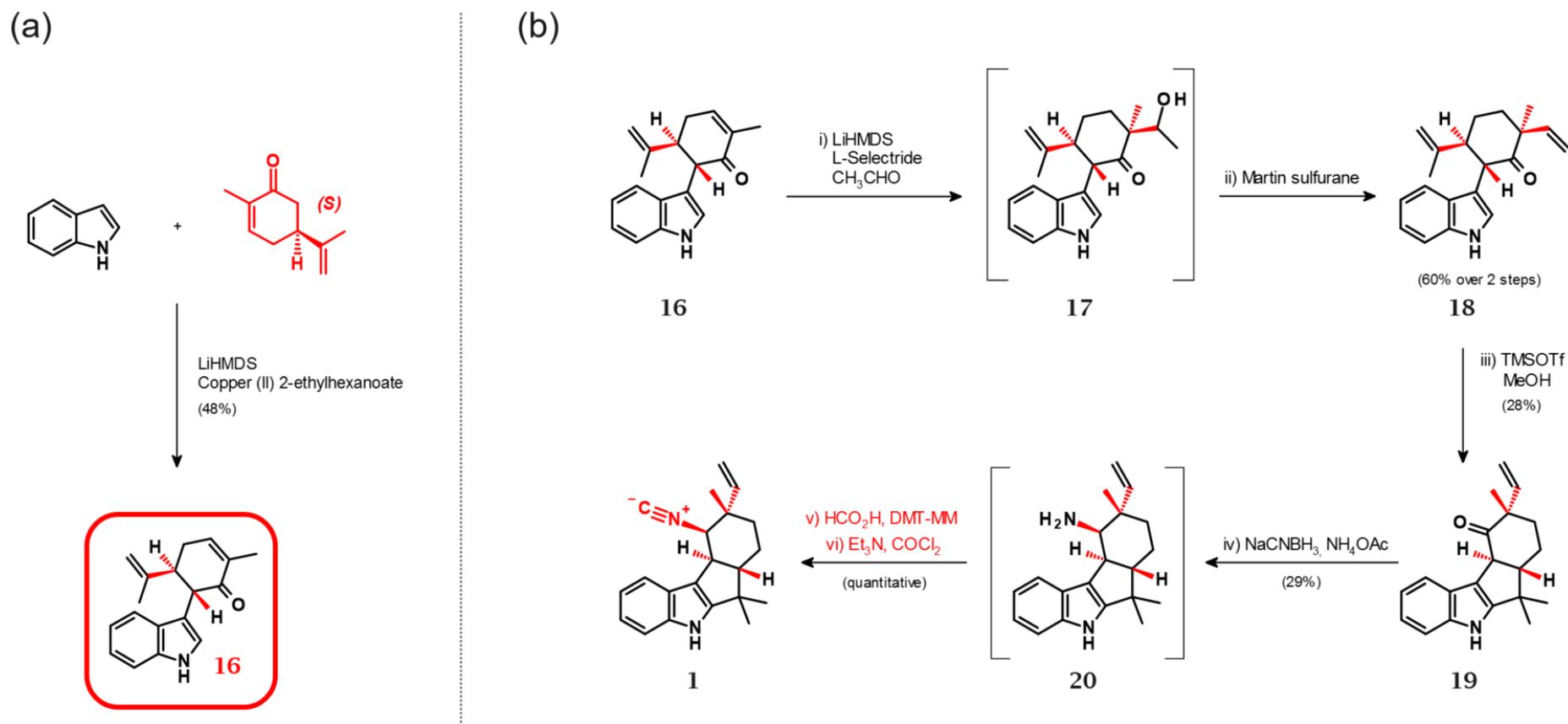


Figure 3.4. Specific rotations for both enantiomers of 12-*epi*-fischerindole U isothiocyanate as reported in the literature,^{131,134} compared to the WelO5 endogenous substrate. The desired synthesis needs to produce the isonitrile analogue of (+)-12-*epi*-fischerindole U isothiocyanate, as drawn and highlighted in the box. For this, it was envisaged that the synthetic route reported in the literature by P. Baran et al.¹³⁴ could be followed by replacing the (*R*)-carvone starting material used in the first step with (*S*)-carvone.

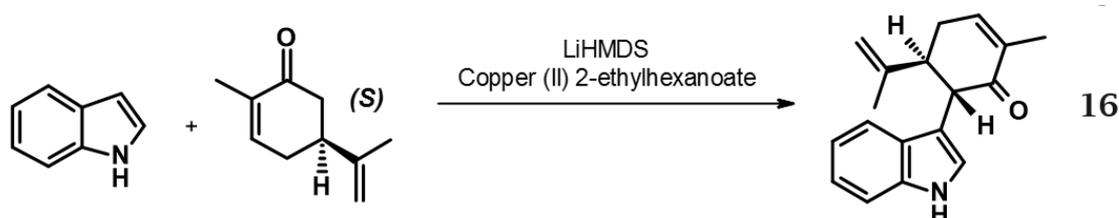
Fortunately, the published route was apparently easily amenable to synthesis of the opposite enantiomer. Exchanging the (*R*)-carvone starting material for (*S*)-carvone should lead to the opposite enantiomer when following the same reaction scheme. The final, modified synthetic route used for the synthesis of **1** is outlined in Scheme 3.2., with deviations from the literature protocol highlighted in red. The isolation of by-products and intermediates enabled a spectrum of structurally-related compounds for testing against WelO5 to investigate structure-activity relationships.



Scheme 3.2. Total synthesis of (+)-12-epi-fischerindole U isonitrile, adapted from that described in Figure 3.3. Reagents and conditions: (a) indole (2.0 equiv.), *S*-carvone (1.0 equiv.), LiHMDS (3.0 equiv.), THF, -78°C , 30 min; then copper(II)-2-ethylhexanoate (1.5 equiv.), -78°C , 12 h, (b) (i) LiHMDS (1.5 equiv.), THF, -78°C , 20 min; then *L*-Selectride (1.05 equiv.), -78°C , 1 h; then CH_3CHO (6.0 equiv.), -78°C , 30 min; (ii) Martin Sulfurane (1.1 equiv.), CHCl_3 , rt, 15 min; (iii) TMSOTf (3.0 equiv.), MeOH (1.1 equiv.), DCM, 0°C , 1 h; (iv) NaCNBH_3 (10 equiv.), NH_4OAc (40 equiv.), MeOH/THF, rt, 5 d; (v) HCO_2H (1.2 equiv.), DMT-MM: formed in situ from CDMT (1.5 equiv.) & NMM (1.5 equiv.), DMAP (0.1 equiv.), DCM, rt, 1 h; (vi) Et_3N (20 equiv.), COCl_2 (3.0 equiv.), DCM, 0°C , 15 min.

3.2.1. Step 1: Indole-carvone coupling

The first step in the synthesis of **1** involves the addition of a solution of LiHMDS to a mixture of carvone and indole starting materials, followed by the addition of a super-stoichiometric quantity of copper(II)-2-ethylhexanoate oxidant (Scheme 3.3).¹³⁴ The literature-proposed mechanism for this reaction proceeds *via* formation of a copper-chelated carvone radical which then undergoes nucleophilic attack by an indole anion to give the desired 6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one, compound **16**.¹³⁵



Scheme 3.3. Copper-catalysed formation of 6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one (**16**) from indole and carvone.

This reaction could be considered the most important in the synthetic scheme because of the formation of a product with two fixed stereocentres in a single step without pre-functionalisation of the starting materials, or the requirement for chiral reagents or catalysts. All the stereocentres in final product **1** (except C-11) are determined by this first step and one can select which 12-*epi*-fischerindole U isonitrile enantiomer is formed by simply exchanging the carvone starting material enantiomers. As the rest of the synthetic route relies upon this first step and there is no comparable alternative reaction, formation of the desired product was very important. Concerns appeared during characterisation of the isolated material that the ¹H NMR spectrum compound **16**, prepared from (*S*)-carvone, showed inconsistencies with its opposite enantiomer (prepared from (*R*)-carvone), the characterisation of which had already been reported in the literature.¹³⁴ Two protons at 6.91 ppm (assigned as

NH-CH=C) and 6.82 ppm (assigned as CH=CMe-C=O) were found to differ to the literature spectra in both their chemical shift and multiplicity; 6.91 (d, $J=2.4$ Hz, 1H) vs. lit. 6.82 (s, 1H) & 6.84-6.75 (m, 1H) vs. lit. 6.71 (d, $J=2.0$ Hz, 1H) (Figure 3.5.).

When chemical shifts from the entire spectrum were compared with the literature values, differences (albeit more subtle ones) were also seen for other protons (Figure 3.5.). These differences vary in magnitude, and their direction of movement (upfield vs. downfield) is inconsistent across the spectrum. However, the ^{13}C NMR spectra of the (*S,S*)- and (*R,R*)-diastereomers appear to be identical (Figure 3.6.). Variations in chemical shift observed in the ^1H NMR spectra of samples prepared from the same solid stock led to the hypothesis that a concentration-dependent effect could be responsible.

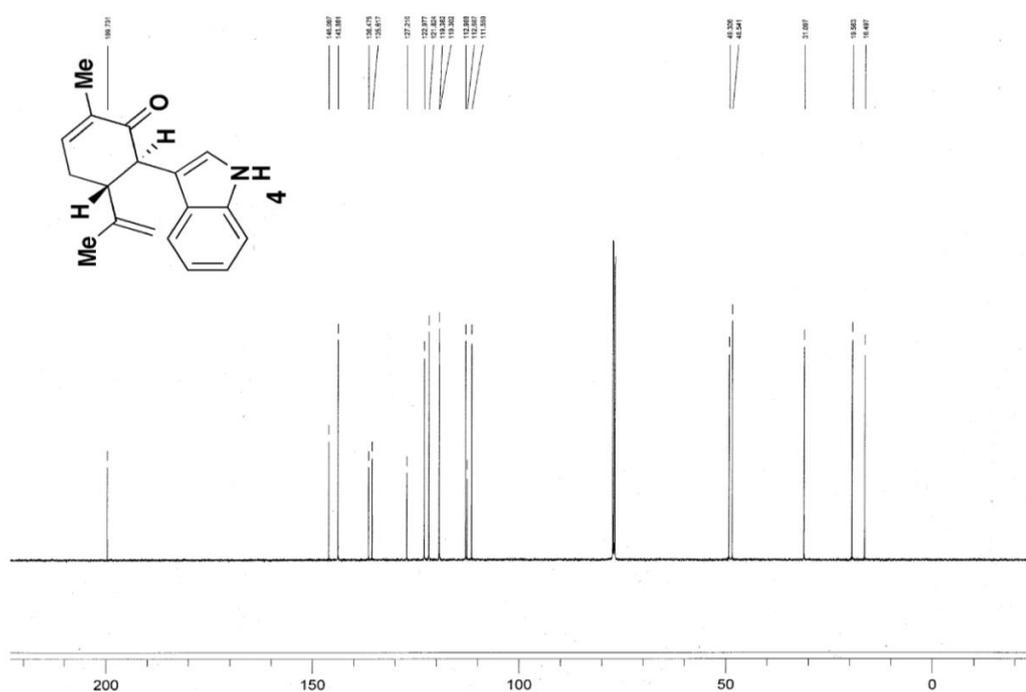
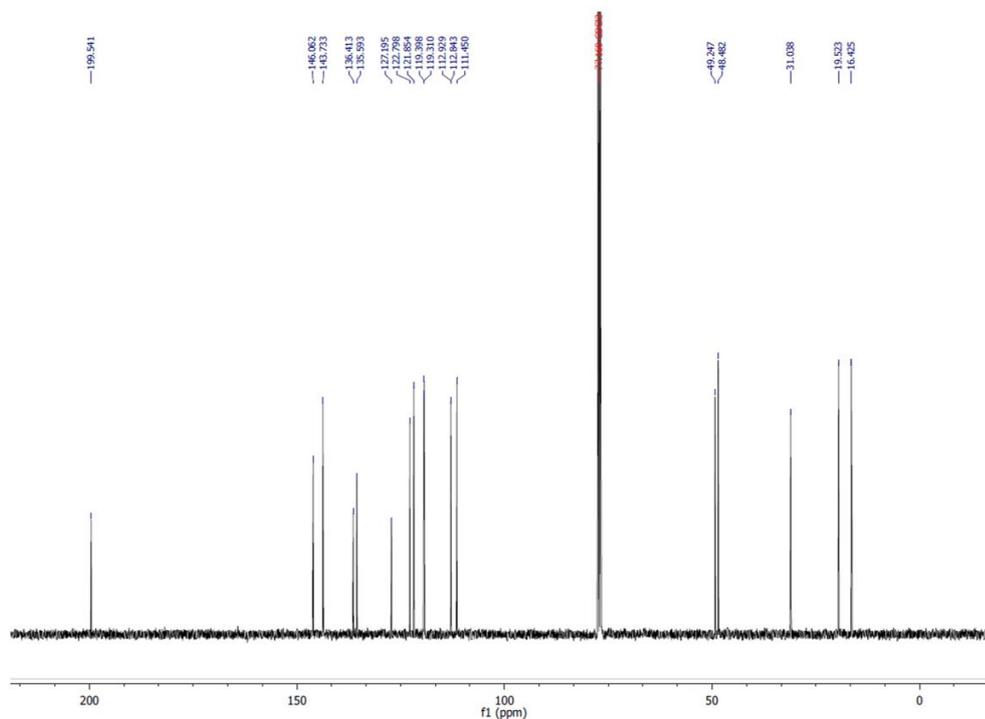
Literature:**This work:**

Figure 3.6. Comparison of ^{13}C NMR spectra for (5R,6R)-6-(1H-indol-3-yl)-5-isopropenyl-2-methylcyclohex-2-en-1-one (literature)¹³⁴ with that of synthetically produced **16** (this work) shows no obvious differences. Both spectra were obtained in the same deuterated solvent, CDCl_3 .

To investigate this proposal, ^1H NMR spectra were obtained for samples of compound **16** over a specified concentration range and the spectra compared (Figure 3.7.). Noticeable changes in chemical shift were seen across the range (1 mg/mL – 100 mg/ mL), with the spectrum prepared at 50 mg/mL appearing near-identical to that of the literature. It was concluded that these changes are likely the result of intermolecular hydrogen bonding effects. By increasing concentration, molecules in the NMR sample develop an increased propensity towards intermolecular hydrogen bonding due to their proximity. Changes in the electron density around the nuclei of atoms directly involved in hydrogen bonding (or indirectly affected as a result of changes in the electronics of the molecule) manifest in changes in chemical shift. A reduction in electron density around the nucleus results in a downfield movement to higher chemical shift. Similarly, an increase in electron density results in greater shielding of the nucleus from the externally applied magnetic field, leading to an upfield movement to lower chemical shift values. The presence of an intermolecular hydrogen bond between the indole NH and the enone carbonyl oxygen in more concentrated samples could explain the observations seen in the ^1H NMR spectra. From the crystal structure of *(5R,6R)*-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one supplied with the synthetic publication,¹³⁴ the proximity of the NH and C=O of two adjacent molecules is consistent with a hydrogen bond in both length and geometry (Figure 3.8.).

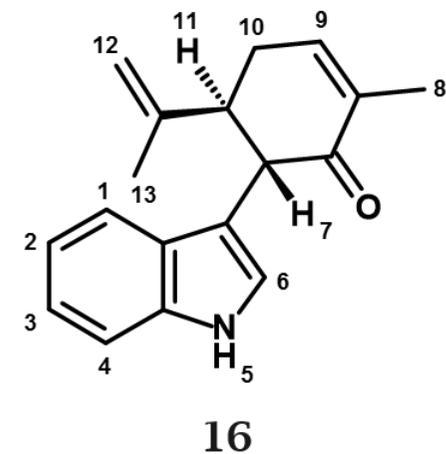
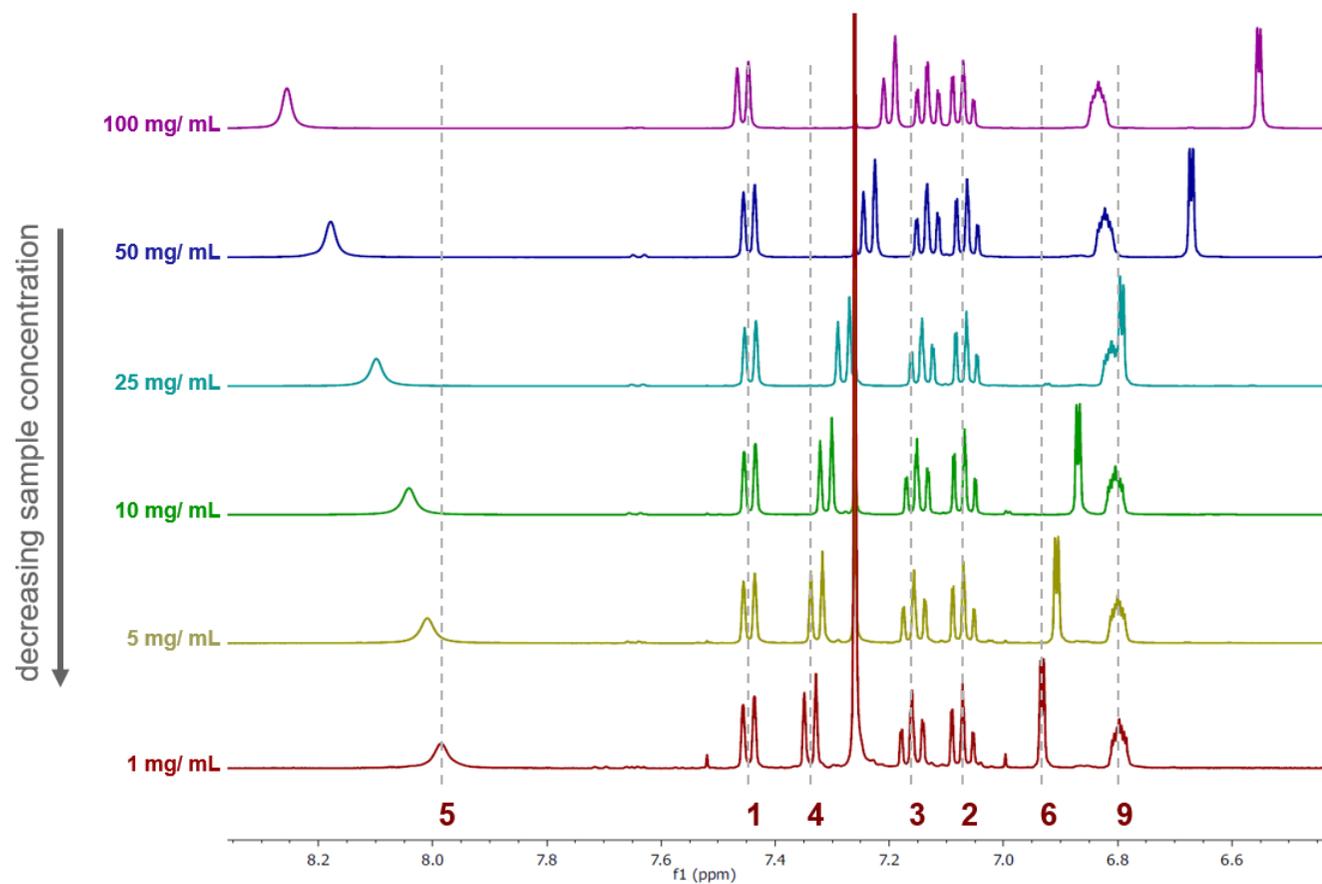


Figure 3.7. Overlay of proton NMR spectra for compound **16** collected over a range of concentrations. Samples were dissolved in deuterated chloroform (δ , 7.26 ppm). The greatest changes in chemical shift can be seen for protons **5** and **6**, and to a lesser extent proton **4**. Increasing the sample concentration from 1 mg/mL to 100 mg/mL results in the movement of NH proton **5** downfield by 0.26 ppm. Across the same concentration differential, aromatic protons **6** and **4** move upfield by 0.38 ppm and 0.14 ppm respectively. More moderate movements are shown by aromatic proton **3** (0.03 ppm upfield) as well as the methyl protons at **8** (spectral region not shown, 0.07 ppm downfield) and vinylic proton **9** (0.03 ppm downfield). The remaining protons of the molecule do not exhibit $\Delta\delta \pm 0.01$ ppm over the concentration gradient.

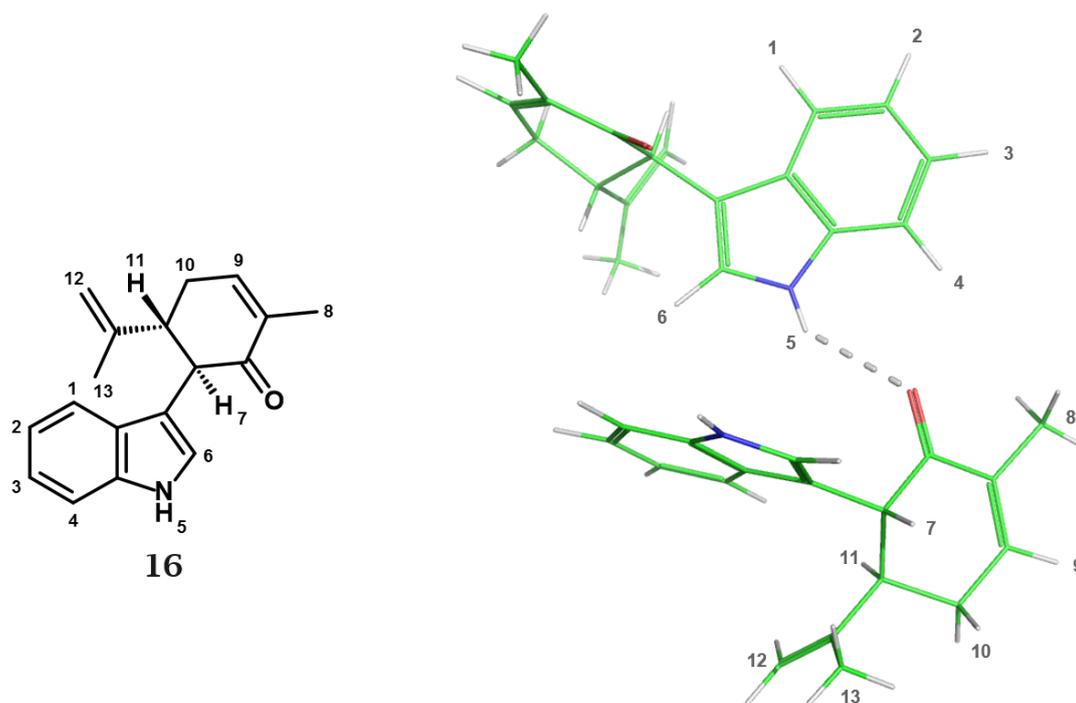


Figure 3.8. Proposed intermolecular hydrogen bond between two molecules of compound **16**. The measured donor-acceptor distance is 3.1Å. These intermolecular hydrogen bonds are responsible for the changes in proton shifts seen at different concentrations in the NMR spectra of this compound, with the most pronounced changes in shift (ppm) seen for the hydrogen bond acceptor (NH). Image rendered in PyMOL using the CIF file supplied with the Baran group synthetic paper.¹³⁴

A second issue arose during characterisation of compound **16**, which was more challenging to rationalise than the ^1H NMR discrepancies: the specific rotation matched the literature value for the opposite (undesired) enantiomer. For the (*R,R*)-enantiomer (literature), $[\alpha]_{\text{D}} = +55$ (DCM, *c* 3.6);¹³⁴ whereas for (*S,S*)-enantiomer **16**, (this work), $[\alpha]_{\text{D}} = +56$ (DCM, *c* 3.9). To confirm that the correct enantiomer had been used in the synthesis of **16**, the specific rotation of the carvone starting material was also measured. A value of $[\alpha]_{\text{D}} = +55$ (neat) was obtained for *S*-carvone, in agreement with the manufacturer's specification.¹³⁶ The reaction was then repeated using *R*-carvone and the specific rotation of the starting material and product both measured, giving *R*-carvone, $[\alpha]_{\text{D}} = -58$ (neat) and 6-(1*H*-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one, $[\alpha]_{\text{D}} = -56$ (DCM, *c* 1.3) respectively (Figure 3.9.). The suspected error in the literature-

reported value was proven to be correct, providing assurance that the desired enantiomer had been isolated.

Both enantiomers of 6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one (the products of step 1) were also analysed by chiral chromatography [SFC; 10-25% MeOH (0.1% NH₄OH); Hichrom CHIRALCEL OJ-3: 4.6 x 150 mm, 3 μm].

Chiral purity calculated by relative UV peak area was determined to be

97.7% e.e. for the (*R,R*)-enantiomer (from *R*-carvone, manufacturers

specification $\geq 97.5\%$)¹³⁷ and 95.7% e.e. for the (*S,S*)-enantiomer, **16** (from *S*-carvone, manufacturers specification $\geq 95.5\%$).¹³⁶

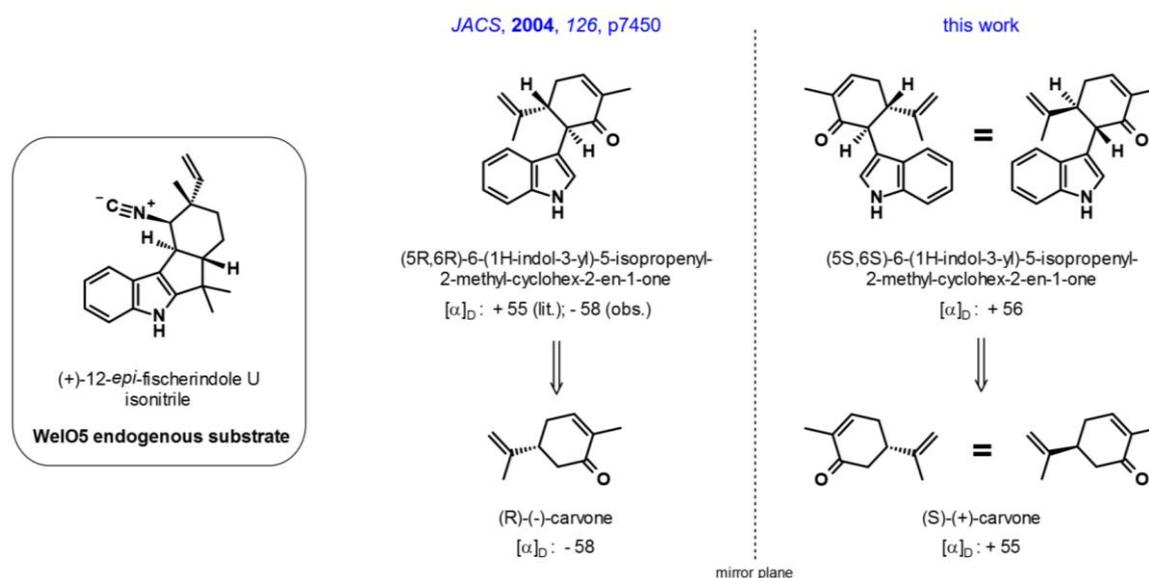
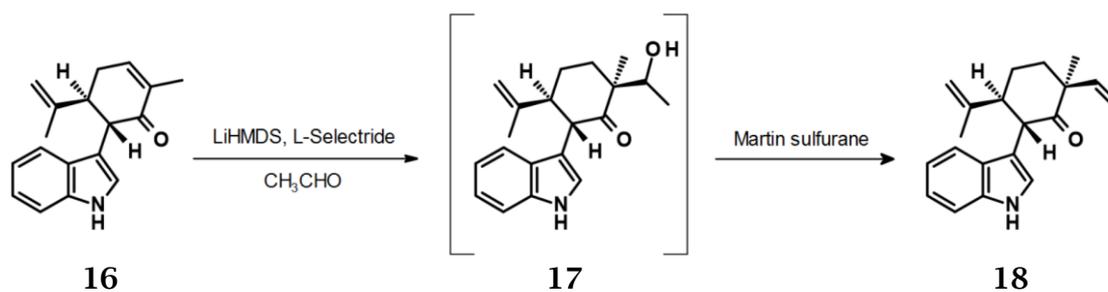


Figure 3.9. Experimentally determined specific rotations are stated alongside the literature quoted values for 6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one enantiomers. The specific rotations of both enantiomers of carvone were also measured and confirmed to match the supplier specification.

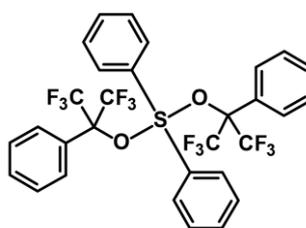
3.2.2. Step 2: Alkene installation

Installation of the fischerindole C-12 alkene proceeds *via* 1,4 conjugate addition of hydride to the enone of compound **16**, with L-selectride (a bulky hydride source) used to achieve diastereoselectivity.



Scheme 3.4. Formation of C-12 secondary alcohol and subsequent dehydration to give the alkene of (2S,5S,6S)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone.

The lithium enolate adds into the carbonyl of acetaldehyde, with dehydration of the resulting secondary alcohol **17** yielding the desired alkene (Scheme 3.4). Martin sulfurane (diphenylbis(1,1,1,3,3,3-hexafluoro-2-phenyl-2-propoxy)sulfurane, Figure 3.10.) was used to perform the alcohol dehydration.¹³⁸



Martin Sulfurane

Figure 3.10. Structure of Martin Sulfurane. This reagent dehydrates a secondary alcohol as part of the C-12 alkene installation process.

Initial dehydration reaction failures were found to be due to the presence of ethanol as an additive (0.5-1%) in the anhydrous reaction solvent (chloroform); this was then replaced with solvent containing unreactive amylene stabilisers. After this correction, the desired product **18** was formed in an overall 60% yield (2 steps), with analytical data consistent with the literature.¹³⁴ The specific rotation for this product was measured as $[\alpha]_D = -126$ (DCM, *c* 0.15), indicating formation of the opposite enantiomer to the literature, for which the specific rotation value is quoted as $[\alpha]_D = +119$ (DCM, *c* 1.3).¹³⁴ This provides additional evidence for isolation of the desired enantiomer of **16** in the preceding indole-

carvone coupling step. Omitting the addition of acetaldehyde entirely led to 1,4-reduction product, compound **21** (Figure 3.11.) during the reaction quench.¹³⁹ This compound was isolated in an inseparable 85:15 ratio of diastereoisomers and retained for WelO5 reactivity screening.

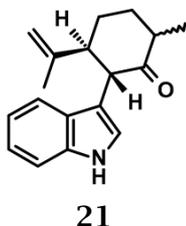
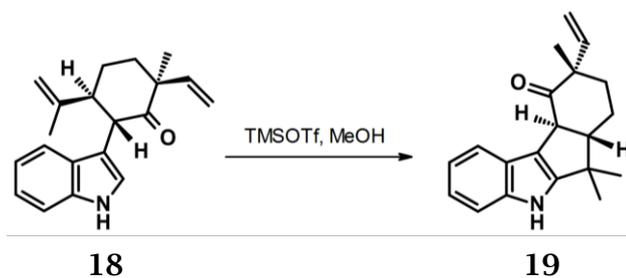


Figure 3.11. Structure of **21**, the by-product produced during the synthesis of **18**. Modification of the reaction conditions resulted exclusively in this conjugate reduction product which was retained for future testing against WelO5.

3.2.3. Step 3: Cyclisation

The final reaction used to construct the fischerindole carbocyclic skeleton was an acid-catalysed cyclisation of the isopropenyl group onto the 2-position of the indole (Scheme 3.5.).¹³⁴



Scheme 3.5. Acid catalysed biomimetic cyclisation of indole onto the isobutylene unit of (2*S*,5*S*,6*S*)-6-(1*H*-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone to give fischerindole (6*aS*,9*S*,10*aS*)-6,6,9-trimethyl-9-vinyl-6*a*,7,8,10*a*-tetrahydro-5*H*-indeno[2,1-*b*]indol-10-one.

In the proposed mechanism, the indole reacts at the 3-position with the tertiary carbocation formed from protonation of the isopropenyl group. Rearrangement of the resulting four-membered ring onto the indole 2-position generates the desired 6,5,5,6 scaffold, **19**.¹⁴⁰ Catalytic triflic acid, formed *in situ* from the addition of methanol to trimethylsilyl trifluoromethanesulfonate (TMSOTf), was used for this step due to reported stability issues with the

fischerindole product under prolonged exposure to acid.¹³⁹ Recycling of the unreacted starting material after purification improved the overall reaction yield. By-products **22**, **23** and **24** were also isolated from the reaction (Figure 3.12.).

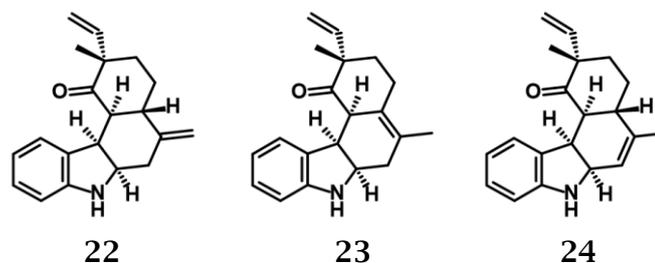


Figure 3.12. Structures of the three by-products formed during acid-catalysed cyclisation of 2*S*,5*S*,6*S*)-6-(1*H*-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone with triflic acid. Samples of each were retained for future testing against WelO5.

The mechanism for formation of these by-products has already been proposed in the literature,¹³⁹ differentiating itself from the previously described mechanism *via* cyclisation onto the far end of the isopropenyl group to give a five-membered ring (formation of the desired product proceeds *via* a four-membered ring) which re-arranges onto the indole 2-position to generate the 6,5,6,6- by-products **22**, **23** and **24** (Figure 3.13.)¹³⁹

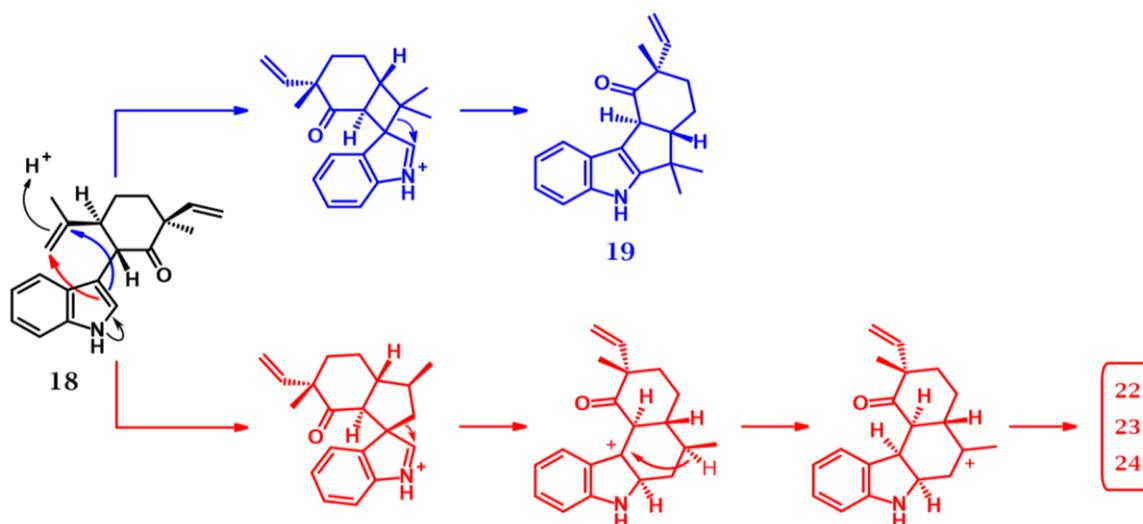
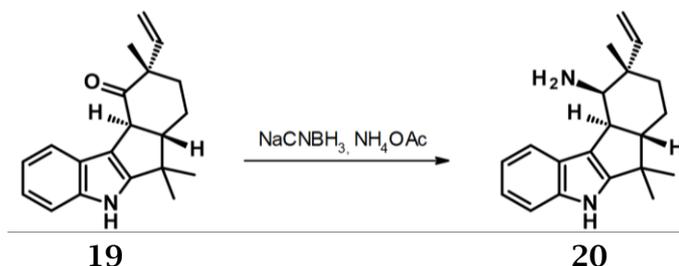


Figure 3.13. Proposed reaction mechanisms for the formation of desired product (6*aS*,9*S*,10*aS*)-6,6,9-trimethyl-9-vinyl-6*a*,7,8,10*a*-tetrahydro-5*H*-indeno[2,1-*b*]indol-10-one (pathway coloured blue), and the formation of minor by-products 2-4 (pathway coloured red).¹³⁹

A small-scale cyclisation using microwave irradiation and acidic Montmorillonite K-10 clay was also attempted and demonstrated similar yields to the triflic acid.¹⁴¹ However, this method was not adopted because it was not easily scalable and the previously mentioned by-products were not detected (the production of multiple structurally-related molecules in a single step has added-value in terms of being able to probe WelO5 structure-activity relationships). Multi-milligram quantities of acyclic material were cyclised using the original triflic acid-recycling method, with samples of all products retained for WelO5 screening. Analytical data for the desired fischerindole reaction product was found to be consistent with the literature, apart from the specific rotation (comparable magnitude, opposite sign).^{134, 139}

3.2.4. Step 4: Reductive amination

The published reaction conditions for the reductive amination step in Scheme 3.6. involve long reaction times (48 hr) and an excess of both reagents (10 equiv. sodium cyanoborohydride, 40 equiv. ammonium acetate) to yield 55% isolated amine (10:1 diastereomeric ratio) and 10% recovered starting material.¹³⁴



*Scheme 3.6. Formation of (6a*S*,9*S*,10*R*,10a*S*)-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1-*b*]indol-10-amine from reductive amination of the ketone to a primary amine.*

However, the superb selectivity achieved by the literature conditions was unrepeatable. A selectivity ratio of 3:2 in favour of the undesired diastereoisomer was estimated from NMR and LC-MS analysis of the crude

isolated reaction mixture. Increasing the reaction time to five days achieved an acceptable balance between starting material consumption and reaction impurity level to give an 85% (crude) yield of amino diastereoisomers, with approximately 10% unreacted starting material (LC-MS). Due to the reduced selectivity observed, preparative chromatography (SFC) was used to separate the constituent isomers. Purification at the amine stage was found to be more favourable than carrying the mixture through and separating the fischerindole diastereoisomers at the end of the synthetic route as it made the reactions easier to follow and gave better overall yields (25-30% isolated yield achieved from the ketone for each amine isomer).

Several alternative conditions were tried for this reductive amination, all without success. The obvious exchange of sodium cyanoborohydride for sodium triacetoxyborohydride failed to yield desired product. Attempts to pre-form the imine with titanium (IV) isopropoxide followed by reduction with sodium borohydride¹⁴² yielded only the corresponding alcohol, suggesting imine formation was either very slow or had not occurred. As a result, an attempt was made to produce the oxime which could then be reduced to the desired amine. Oxime formation was unsuccessful at room temperature but was found to proceed after heating to 60 °C. Attempts to reduce the oxime under mild conditions using zinc and ammonium formate were unsuccessful. Re-analysis of the oxime analytical data suggested that the starting material had undergone a Beckmann rearrangement, explaining the reduction reaction failure.

With a view to forming amine **20** *via* reduction to the alcohol and displacement with azide, ketone **19** was treated with sodium borohydride.¹⁴¹ This reaction

was complete within 10 minutes, suggesting that imine formation may be reaction limiting for the reductive amination. The overall yield for the alcohol mixture was 80%, in a 2:1 ratio of diastereoisomers. Isomer ratios for the ketone reduction and reductive amination reactions were estimated from the ^1H NMR of the crude reaction mixture (Figure 3.14.).

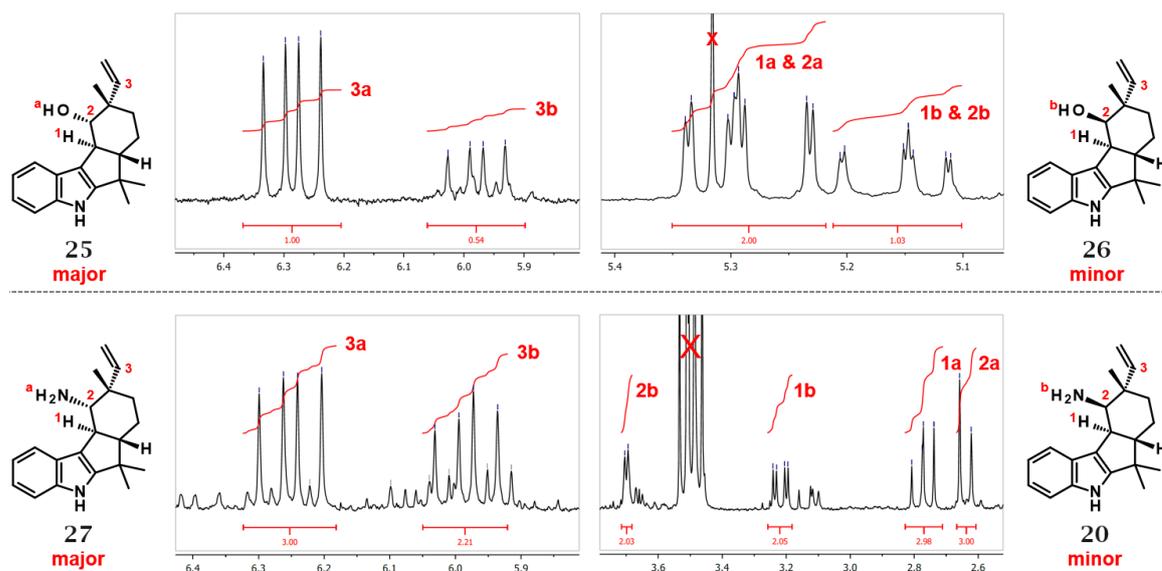


Figure 3.14. Selected regions of ^1H NMR spectra (CDCl_3) of crude product mixtures for both the sodium borohydride ketone reduction (2:1; desired: undesired) and sodium cyanoborohydride reductive amination reaction (3:2; undesired: desired). Key protons used to calculate the approximate isomer ratio are highlighted in red.

The major isomer produced from the sodium borohydride ketone reduction (25) had the same relative stereochemistry as the major isomer from the reductive amination (27). However, whilst the C-11 (*S*)-amine (27) was undesired, the C-11 (*S*)-alcohol (25) was apparently perfectly set up for inversion at C-11 with azide. Initial attempts at a one-step reaction with DPPA were unsuccessful, yielding only starting material. Following the literature protocol, a two-step process *via* the mesylate and treatment with LiN_3 initially appeared successful.¹⁴¹ However, attempts to avoid the use of Na/Hg amalgam¹⁴¹ by reducing the azide with triphenylphosphine (Staudinger reduction)¹⁴³ led only to elimination of the azide. In the interest of time, and

due to the initial failures with this route, the reductive amination remained the preferred method for accessing (+)-12-*epi*-fischerindole U isonitrile, **1**. Both alcohol diastereoisomers were retained for WelO5 activity testing.

3.2.5. Step 5: Isonitrile formation

The final step of the published synthesis produces (-)-12-*epi*-fischerindole U isothiocyanate by reaction of the previously synthesised amine with carbonyldiimidazole (CDI).¹³⁴ At this point, alternative chemistry was required to install an isonitrile in place of the isothiocyanate.

Isonitriles can be prepared from primary amines in two sequential steps, formylation then dehydration.¹⁴⁴ Different combinations of reagents have been used in the literature to make isonitriles this way. In this specific example, formylation was performed *via* amide-coupling of the fischerindole primary amine with formic acid (Scheme 3.7.)^{134,145} (alternative procedures exist where formate esters, acetic formic anhydride or the activated formic ester are used).¹⁴⁶⁻¹⁴⁹



Scheme 3.7. Formation of the (6*aS*,9*S*,10*R*,10*aS*)-6,6,9-trimethyl-9-vinyl-5,6*a*,7,8,10,10*a*-hexahydroindenof[2,1-*b*]indol-10-amine formamide, then dehydration to corresponding isonitrile of WelO5 substrate (+)-12-*epi*-fischerindole U isonitrile, **1**.

The dehydration step was initially performed with the Burgess reagent (methoxycarbonylsulfamoyl)triethylammonium hydroxide),^{150,151} but this method was abandoned in favour of phosgene.¹⁵² Unlike the Burgess reagent,

the use of phosgene does not require flash column chromatography to remove reaction by-products (CO_2 and HCl in the case of phosgene, see Figure 3.15.). Phosgene solution is also substantially cheaper and more atom economical than the Burgess reagent.

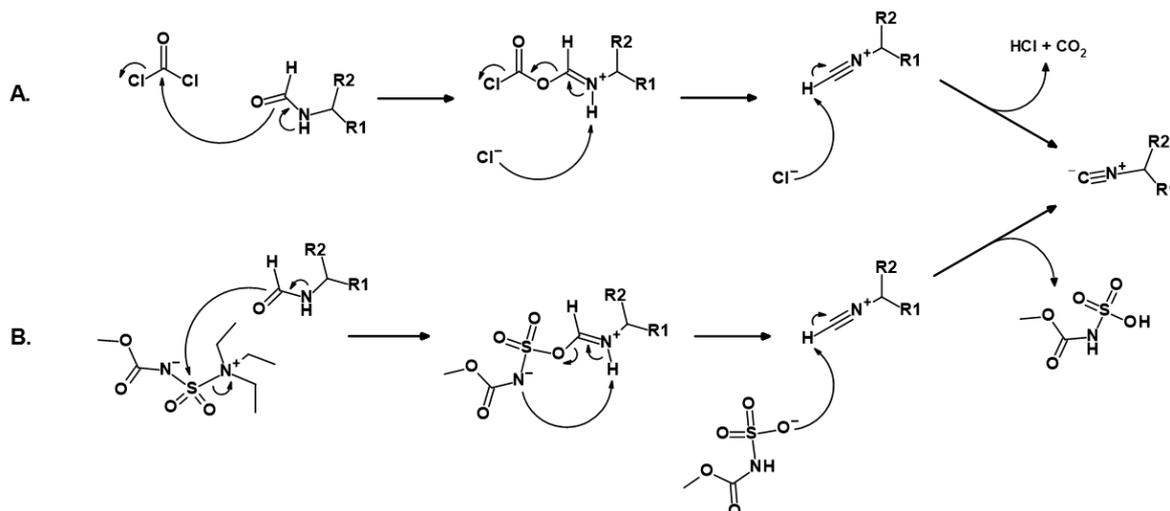
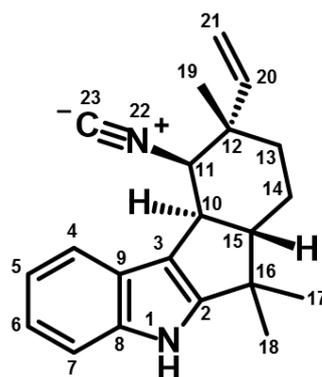


Figure 3.15. Mechanisms and by-product comparison for the two synthetic methods trialled for isonitrile synthesis; phosgene (method A) and Burgess reagent (method B). Use of phosgene results in easily removable HCl and CO_2 and was therefore selected for the final synthetic route.

Both fischerindole isonitrile diastereoisomers were synthesised from their corresponding *C*-11 (*S*)- and *C*-11 (*R*)-amines (**20** and **27** respectively) and retained for WelO5 activity testing. Analytical data for the isolated **1** was consistent with the original 1994 isolation paper, with the NMR data compared in Figure 3.16. and the ^1H NMR spectra aligned with that from the literature in Figure 3.17.¹³¹ The specific rotation for synthetic **1** was measured as $[\alpha]_{\text{D}} +60$ (DCM, *c* 0.30). A reference value does not exist in the literature, however, the analogous 12-*epi*-fischerindole U isothiocyanate (**10**) has a specific rotation of the same sign, $[\alpha]_{\text{D}} +231$ (DCM, *c* 0.035) and 12-*epi*-fischerindole G (**2**) has a specific rotation of the same sign and magnitude, $[\alpha]_{\text{D}} +67$ (DCM, *c* 0.09).¹³¹

12-*epi*-fischerindole U isonitrile (1)

Chemical Shift (lit), ppm	Multiplicity (lit)	Proton Identity	Chemical Shift (obs), ppm	Multiplicity (obs)
8.03	br s	1 (1H)	8.03	br s
7.44	d (J 7.0 Hz)	4 (1H)	7.44	m
7.35	dd (J 7.0, 0.8 Hz)	7 (1H)	7.35	m
7.14	m	5 (1H)	7.08	m
7.14	m	6 (1H)	7.08	m
5.89	dd (J 17.9, 11.0 Hz)	20 (1H)	5.89	dd (J = 17.6, 10.9 Hz)
5.22	d (J 11.0 Hz)	21E (1H)	5.22	d (J = 11.0 Hz)
5.23	d (J 17.9 Hz)	21Z (1H)	5.24	d (J = 17.7 Hz)
4.37	br s	11 (1H)	4.37	s
3.17	dtd (J 11.5, 3.5 Hz)*	10 (1H)	3.17	dtd (J 11.1, 4.1, 3.1 Hz)
2.32	td (J 11.5, 3.3 Hz)	15 (1H)	2.32	td (J = 11.0, 3.9 Hz)
1.91	dd (J 12.5 Hz)	13 _{eq} (1H)	1.92	m
1.71	m	13 _{ax} (1H)	1.71	m
1.71	m	14 _{eq} (1H)	1.66	m
1.62	m	14 _{ax} (1H)	1.60	m
1.40	s	18 (3H)	1.40	s
1.26	s	19 (3H)	1.26	s
1.04	s	17 (3H)	1.04	s

* insufficient coupling data provided by literature paper.¹³¹

Chemical Shift (lit), ppm	Carbon Identity	Chemical Shift (obs), ppm
158.2	C22	158.2
152.9	C2	153.0
143.2	C20	143.2
140.0	C8	140.0
124.3	C9	124.3
121.0	C6	121.1
120.1	C5	120.2
118.4	C4	118.4
115.0	C3	115.1
114.7	C21	114.7
112.1	C7	112.1
62.8	C11	62.9
55.6	C15	55.7
42.5	C10	42.5
41.2	C12	41.2
40.9	C16	40.6
32.3	C13	32.3
28.2	C19	28.2
25.2	C18	25.2
20.92	C14	20.97
20.89	C17	20.92

Figure 3.16. Comparison of ¹H and ¹³C NMR spectral data for synthetic (+)-12-*epi*-fischerindole U isonitrile relative to the literature spectra of the isolated (+)-12-*epi*-fischerindole U isonitrile from *H. wehwitschii*. Spectra are identical (within instrumental variability). Assignments made as per the literature.¹³¹

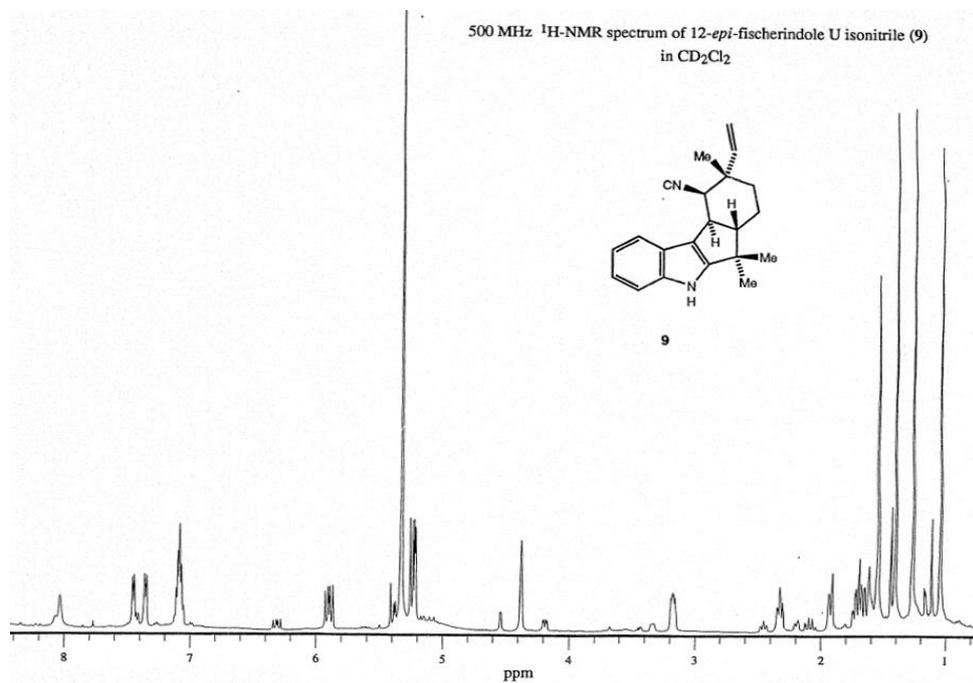
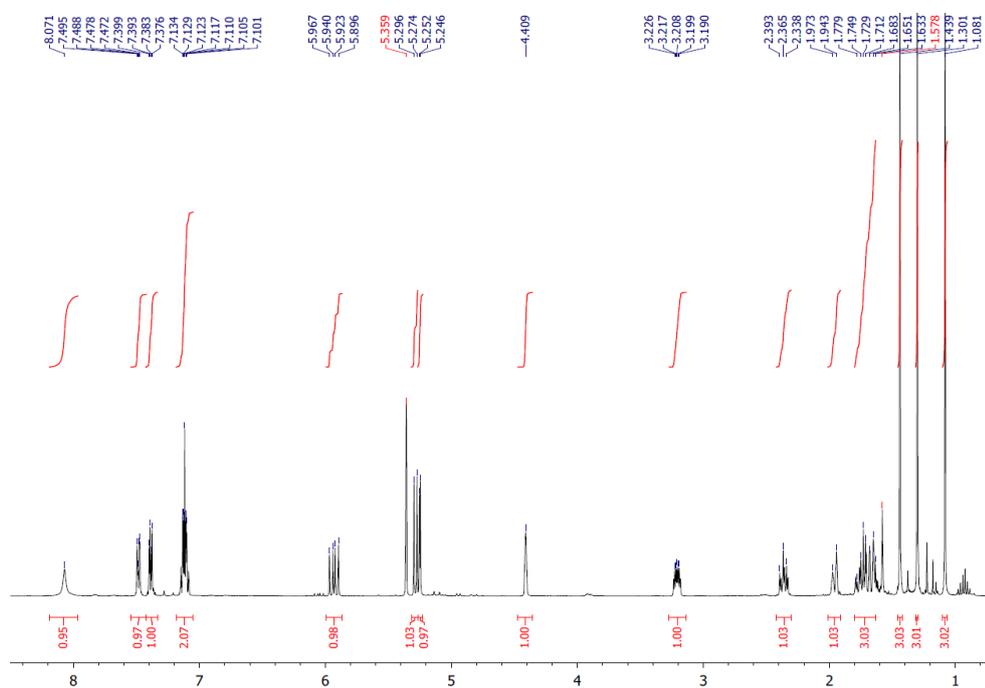
Literature:**This work:**

Figure 3.17. Comparison of $^1\text{H-NMR}$ spectra for the isolated, naturally occurring (+)-12-*epi*-fischerindole U isonitrile (literature) with that of synthetically produced (+)-12-*epi*-fischerindole U isonitrile (this work) shows no obvious differences apart from sample purity. Both spectra were collected in the same solvent, CD_2Cl_2 .

3.3. Synthesis of simplified substrate analogues

A limited set of ‘cut-back’, simplified analogues of the WelO5 natural substrate were synthesised in parallel with (+)-12-*epi*-fischerindole U isonitrile, **1** (Figure 3.18.). These were produced to examine whether halogenase activity could be seen for compounds with reduced functionality as well as for the highly complex (+)-12-*epi*-fischerindole U isonitrile (and intermediates and by-products thereof). Without a WelO5 crystal structure, it was hoped that these additional compounds could help to establish some form of structure-activity relationship and indicate which structural elements or functional groups might be required for WelO5 halogenation activity.

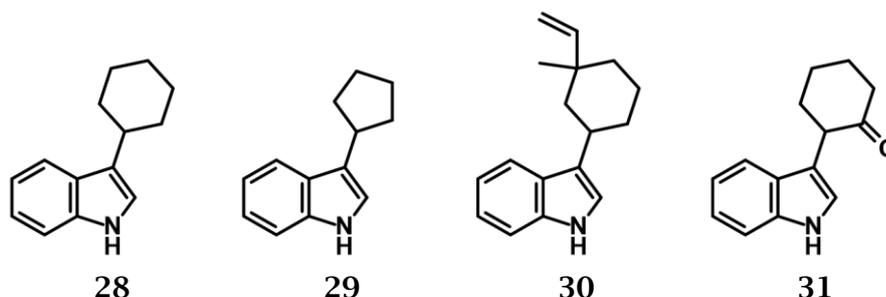


Figure 3.18. Structurally simplified analogues of natural substrate (+)-12-*epi*-fischerindole U isonitrile and its synthetic intermediates.

Compounds **28-30** were all synthesised by a one-step reductive alkylation of indole with the corresponding ketones using trichloroacetic acid and triethylsilane.⁸ This gave exclusively the 3-position alkylated indoles without requiring protection of the indole nitrogen. The ketone used for the synthesis of **30** was not commercially available but was easily synthesised *via* a copper-catalysed 1,4-addition of vinyl Grignard to 3-methylcyclohex-2-en-1-one. Synthesis of **31** was unsuccessful *via* this reductive alkylation method. Instead, indole was coupled with 2-chlorocyclohexanone, using sodium carbonate in TFE, to yield the desired product (in a reaction thought to proceed *via* oxyallyl cation formation).¹⁵³

Plans for the synthesis of further analogues were put on hold due to the lack of activity seen for these compounds in the initial WelO5 activity screen (Section 3.4.).

3.4. Initial activity screening

The mechanism used by Fe²⁺ and 2OG-dependent halogenases for substrate chlorination does not easily lend itself to the development of high-throughput screening or selection methods for activity determination. As a result, a low-throughput, enzyme-compound incubation method has been used in the literature, with WelO5 activity assessed by LC-MS.⁵⁰ Providing sufficient resolution can be obtained by the LC-method, conversion from one compound to another can be calculated using peak area. Product identity can then be elucidated from the mass spectrum. LC-MS based methods are low-throughput but have the advantage that multiple different products can be identified at the same time. By contrast, high-throughput methods are less generalised and are usually designed to look for the formation of a single compound of interest.

The literature protocol for WelO5 activity testing incubates purified, de-oxygenated WelO5 (20 µM) with compound (1 mM), 2OG (2 mM), NaCl (10 mM) and Fe²⁺: Mohr's salt; ammonium iron sulfate (500 µM) in 20 mM HEPES buffer on 100 µL scale, at 30 °C for 1hr. The reactions are set-up in a glove-box using nitrogen pre-purged solutions and the reaction is initiated upon exposure to air (oxygen). After 1hr, the reaction is stopped by extraction with ethyl acetate. The reactions extracts are then combined, dried under a stream of nitrogen gas and re-dissolved in methanol prior to LC-MS analysis.⁵⁰ To assess the activity of WelO5 towards the assembled collection of synthetic intermediates and

simplified substrate analogues, a modified version of this literature screening method was performed. The initial protocol was revised to increase throughput (eliminating glove-box set-up and reducing sample manipulations).

The following alternations were made:

- Purified WelO5 eluted from the gel filtration column in 20 mM HEPES, 150 mM NaCl buffer (pH 7.4); reactions could therefore be performed in the gel filtration buffer. The addition of salt to the reaction mixture was not deemed necessary (excess already present).
- The substrate DMSO stock concentration was not noted in the literature.⁵⁰ Initial experiments incubating WelO5 with natural substrate **1** (1 mM, lit.) in the presence of a 1-5% DMSO final concentration did not achieve complete enzyme turnover. This was thought to be due to a combination of poor solubility (compound precipitation detected) and prior exposure of the reaction to air (oxygen) by not using a glove-box for reaction setup. Compounds were instead used at 100 μ M final concentration, from a dilution of the 1 mM DMSO stock. The inclusion of 10% DMSO in the reaction aided compound solubility and was not found to negatively impact reaction turnover. The use of lower compound concentrations was also valuable for preserving stocks of synthetic substrate **1** for further experiments.
- Preparation of the Fe²⁺ solution was not described in the literature.⁵⁰ A concentrated (50 mM) solution of ammonium iron sulfate was freshly prepared in 20 mM HCl, diluted to 5 mM with Milli-Q[®] water and used at 500 μ M Fe²⁺ reaction concentration. By preparing the Fe²⁺ solution in acid, the speed of oxidation to the catalytically inactive Fe(III) was reduced, relative to preparation in water or buffer (qualitatively

determined from the speed of colour change of the solution due to Fe(III) formation). Addition of the mildly acidic solution did not negatively impact reaction turnover (presumably the small volume added can be compensated for by the buffering capacity of HEPES).

- Incubations were performed at 37 °C overnight. The increased temperature (vs. 30 °C, lit.) was not found to negatively impact reaction turnover and was a more practical choice (easy access to 37 °C incubator space). Reactions were performed overnight to account for the possibility of slower WelO5 turnover on non-natural substrates.
- Reactions were quenched by the addition of an equivalent reaction volume (100 µL) of cold methanol and centrifuged. This step removed sufficient protein and precipitants to enable direct injection of the supernatant for LC-MS analysis without requiring further sample manipulation.

A summary of the reaction components used for this initial compound screen is detailed in Table 3.1. This reaction setup was later modified and optimised to incorporate WelO5 enzyme variants and increase throughput further (see Chapter 5).

Reagent	Stock Solutions	Final Assay Concentration	Volume (µL)
WelO5	100 µM (3.2 mg/mL)	25 µM	25
Small molecule	1 mM (in DMSO)	100 µM	10
(NH ₄) ₂ Fe(SO ₄) ₂	5 mM (in 20 mM HCl)	500 µM	10
2OG	20 mM (in reaction buffer)	2 mM	10
NaCl	150 mM (present in 100 µM enzyme stock solution)	37.5 mM	-
Reaction buffer	20 mM HEPES, pH 7.4	-	45
Total volume:			100

Table 3.1. Summary of the reaction composition used in initial activity screen. Preparation of the individual reaction components is described alongside the final reaction concentrations used.

New components in the reaction mixture were identified by subtracting the chromatograms of the enzyme-free and compound-free controls.

Hydroxylation or chlorination events were confirmed for UV peaks with the correct m/z which were not present in the control reactions.

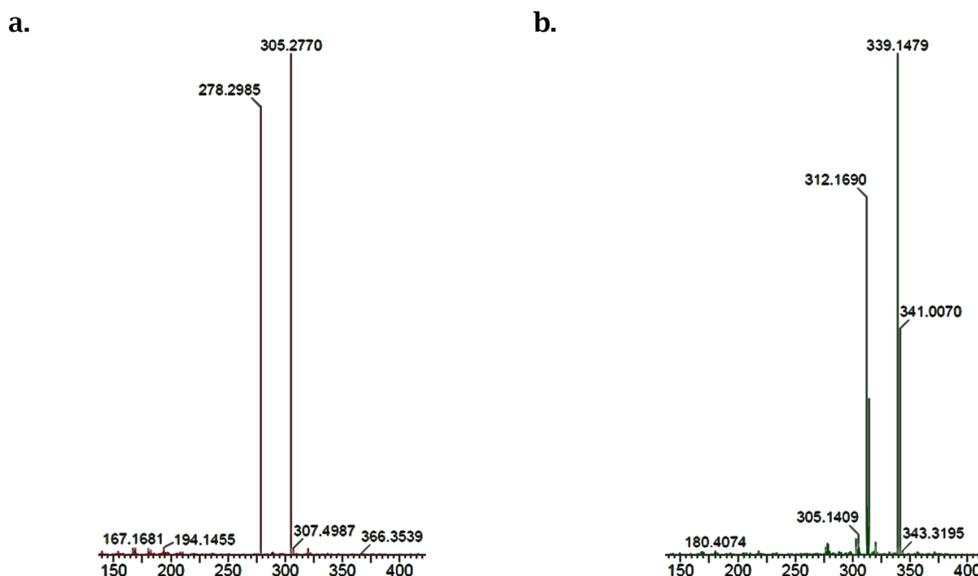
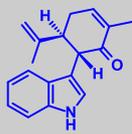
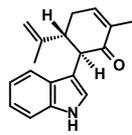
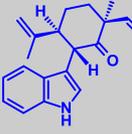
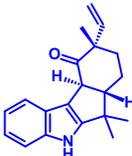
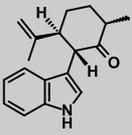
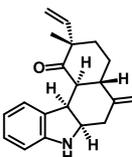
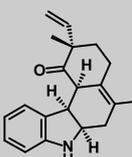
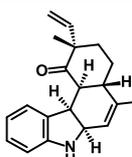


Figure 3.19. Extracted mass spectra for WelO5 natural substrate (+)-12-*epi*-fischerindole U isonitrile, **1** [a.] and chlorinated product (+)-12-*epi*-fischerindole G isonitrile, **2** [b.] (formed after incubation with WelO5 under the previously described reaction conditions). The molecular weight for **1** is 304, appearing as m/z 305 $[M+H]^+$ in mass spectrum a. The molecular weight for **2** is 338, appearing as m/z 339/341 $[M+H]^+$ in mass spectrum b. due to the 3:1 chlorine isotope pattern.

Chlorination events were also confirmed by the presence of the distinctive 3:1 chlorine isotope ionisation pattern (Figure 3.19.), resulting from the relative natural abundance ratio of the ^{35}Cl (75%) and ^{37}Cl (25%) isotopes. The starting material peak and any new peaks in the UV trace were integrated and the percentage conversion calculated from the relative peak areas. The results of this initial WelO5 activity screen are listed in Tables 3.2a. and 3.2b.

(+)-12-*epi*-fischerindole U isonitrile synthetic intermediates & derivatives:

Entry	Name	Structure	Observed Chlorination Y (%) / N	Observed Hydroxylation Y (%) / N
A	(5 <i>S</i> ,6 <i>S</i>)-6-(1 <i>H</i> -indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one, 16		N	N
B	(5 <i>R</i> ,6 <i>R</i>)-6-(1 <i>H</i> -indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one (lit.)		N	N
C	(2 <i>S</i> ,5 <i>S</i> ,6 <i>S</i>)-6-(1 <i>H</i> -indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone, 18		Y (1%)	N
D	(6 <i>aS</i> ,9 <i>S</i> ,10 <i>aS</i>)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5 <i>H</i> -indeno[2,1- <i>b</i>]indol-10-one, 19		Y (75%) *	N
E	(2 <i>S</i> ,3 <i>S</i>)-2-(1 <i>H</i> -indol-3-yl)-3-isopropenyl-6-methyl-cyclohexanone, 21		N	N
F	(2 <i>S</i> ,4 <i>aS</i> ,6 <i>aR</i> ,11 <i>bR</i> ,11 <i>cS</i>)-2-methyl-5-methylene-2-vinyl-3,4,4a,6,6a,7,11b,11c-octahydrobenzo[<i>c</i>]carbazol-1-one, 22		N	N
G	(2 <i>S</i> ,6 <i>aR</i> ,11 <i>bR</i> ,11 <i>cR</i>)-2,5-dimethyl-2-vinyl-4,6,6a,7,11b,11c-hexahydro-3 <i>H</i> -benzo[<i>c</i>]carbazol-1-one, 23		N	N
H	(2 <i>S</i> ,4 <i>aS</i> ,6 <i>aR</i> ,11 <i>bR</i> ,11 <i>cS</i>)-2,5-dimethyl-2-vinyl-4,4a,6a,7,11b,11c-hexahydro-3 <i>H</i> -benzo[<i>c</i>]carbazol-1-one, 24		N	N

I	(6 <i>aS</i> ,9 <i>S</i> ,10 <i>S</i> ,10 <i>aS</i>)-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1- <i>b</i>]indol-10-ol, 25		N	N
J	(6 <i>aS</i> ,9 <i>S</i> ,10 <i>R</i> ,10 <i>aS</i>)-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1- <i>b</i>]indol-10-ol, 26		Y (?) ^{**}	N
K	(6 <i>aS</i> ,9 <i>S</i> ,10 <i>R</i> ,10 <i>aS</i>)-10-isocyano-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1- <i>b</i>]indole (+)-12-<i>epi</i>-fischerindole U isonitrile, 1		Y (100%)	N
L	(6 <i>aS</i> ,9 <i>S</i> ,10 <i>S</i> ,10 <i>aS</i>)-10-isocyano-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1- <i>b</i>]indole, 32		N	N

Table 3.2a. Results from initial WelO5 activity screen. Intermediates used in the synthesis of (+)-12-*epi*-fischerindole U isonitrile, **1** are highlighted in blue. Chlorination and hydroxylation activity are indicated with a Y, with the observed percentage conversion listed in brackets. N indicates no activity seen.

* 75 % conversion to a peak with a chlorine isotope MS pattern but the molecular weight does not correspond to chlorination alone. Complete starting material desaturation seen in the enzyme-free control (loss of two protons). The assay result is consistent with 75 % chlorination of the non-enzymatically desaturated starting material.

** Chlorine isotope MS pattern observed, but product peak overlaps with starting material peak in the UV spectrum, preventing calculation of conversion.

Simplified acyclic analogues:

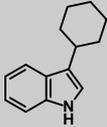
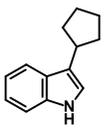
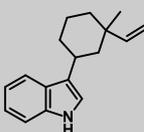
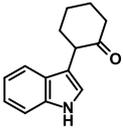
Entry	Name	Structure	Observed Chlorination Y (%) / N	Observed Hydroxylation Y (%) / N
M	3-cyclohexyl-1H-indole, 28		N	N
N	3-cyclopentyl-1H-indole, 29		N	N
O	3-(3-methyl-3-vinyl-cyclohexyl)-1H-indole, 30		N	N
P	2-(1H-indol-3-yl)cyclohexanone, 31		N †	N †

Table 3.2b. Results from initial WelO5 activity screen using simplified acyclic analogues of (+)-12-*epi*-fischerindole U isonitrile. Chlorination and hydroxylation activity are indicated with a Y, with the observed percentage conversion listed in brackets. N indicates no activity seen.

† new peaks seen in UV spectrum but unable to determine identity due to poor ionisation in MS

From these results, we can see that wild-type WelO5 appears to be highly selective for its natural substrate. Apart from the natural substrate control (Table 3.2a., entry K), activity was only detectable within the LC-MS limits of detection (LOD) for **18** (Table 3.2a., entry C), **19** (Table 3.2a., entry D) and **26** (Table 3.2a., entry J) analogues of the natural substrate.

The carbocyclic skeleton of (+)-12-*epi*-fischerindole isonitrile alone is relatively flat (although decorated with three-dimensional functionality). If the saturated ring is not constrained by cyclisation, the acyclic analogues may not position a C-H bond in the same orientation for hydrogen abstraction as the fischerindoles (Figure 3.20.). The addition of some restricted rotation *via*

'substrate-like' peripheral methyl-vinyl and isopropenyl groups (Table 3.2a., entry C), may encourage the molecule to sit in the necessary orientation for some of the time - leading to the small percentage of chlorinated product observed for compound **18**.

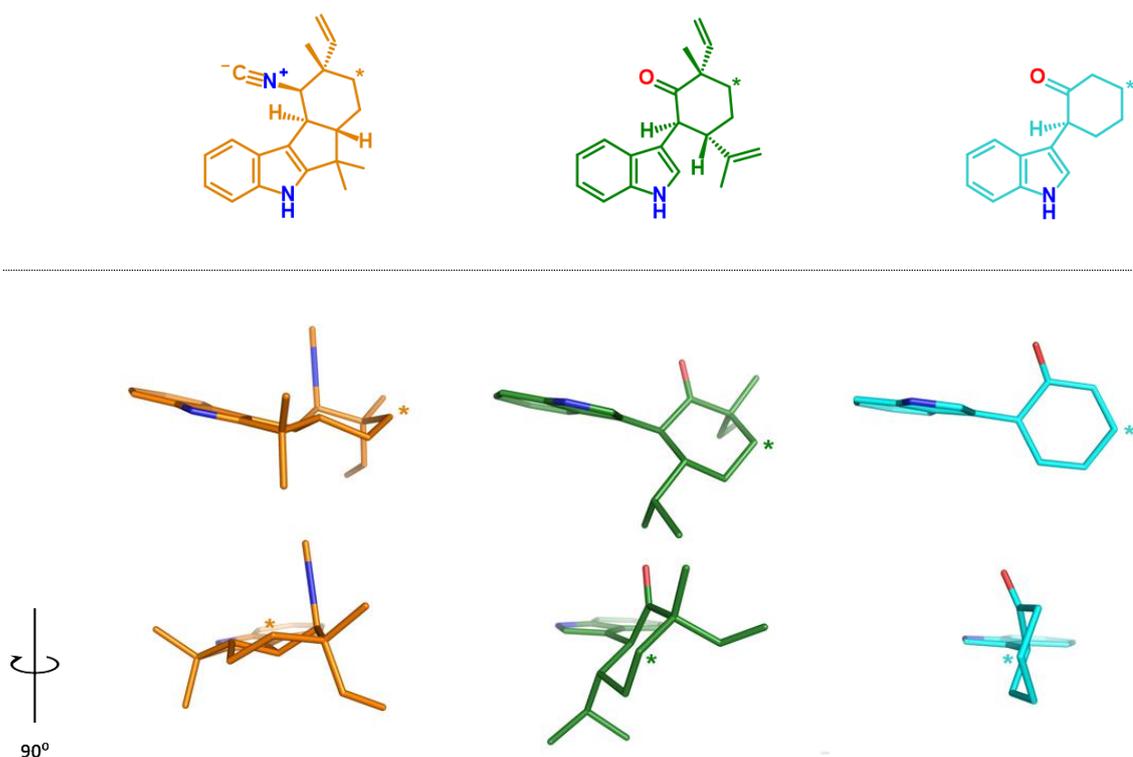


Figure 3.20. Three-dimensional compound structural comparison. Compounds were aligned at their indole rings to illustrate the relative orientation of the cyclohexyl ring. Starred positions show the location (or potential location) of WelO5 chlorination. The carbocyclic skeleton of WelO5 natural substrate (+)-12-epi-fischerindole isonitrile, **1** (orange; Table 3.2a., entry K) is reasonably flat. The fischerindole constrains the cyclohexyl ring; holding it in an optimal orientation for WelO5 H \cdot abstraction and chlorination. Acyclic analogues (**18**, green and **S-31**, blue) are not expected to orient the cyclohexyl ring in a similar way, rationalising the lack of WelO5 activity seen for these compounds. A highly substituted cyclohexyl ring, such as for Table 3.2a., entry C, may provide some restricted rotation, leading to the small percentage of WelO5-chlorinated product observed for this compound. Figure generated in PyMOL.

The isonitrile group itself does not seem to be essential for activity, as chlorination activity is seen for both ketone **19** (Table 3.2a.; entry D) and hydroxy **26** (Table 3.2a., entry J) analogues of the fischerindole substrate. However, the orientation of the group placed at this position does appear to be important. Inverting the hydroxy- and isonitrile- stereo centres (Table 3.2a., compound **25**, entry I & compound **32**, entry L) eliminates activity entirely. One

can therefore assume that an interaction between the isonitrile and the protein is key for activity (possibly involved in orientation of the substrate for successful hydrogen abstraction). This interaction appears to be maintained by one of the lone pairs on compound **19** (Table 3.2a., entry D). The combination of both the ‘flat’ shape of the carbocyclic scaffold and an ability to mimic the interaction made by the isonitrile of the substrate appears to be key for WelO5 activity. Compounds lacking in both these elements do not show any chlorination. We can conclude from this initial activity screen that WelO5 is highly specific for its natural substrate and amino acid substitutions will be necessary to make this enzyme more promiscuous.

3.5. Synthetic summary for WelO5 endogenous substrate and analogues

This Chapter has outlined the synthesis of the natural substrate for WelO5, (+)-12-*epi*-fischerindole U isonitrile, **1** *via* a modified literature route. The intermediates and by-products from this route were retained and tested for WelO5 activity alongside a small number of structurally simpler analogues. Limited WelO5 activity was seen for highly structurally similar compounds, specifically fischerindole scaffolds possessing a group at *C-11* orientated in such a way that it can form an analogous protein interaction to the isonitrile. Activity was not seen for compounds with reduced complexity and functionality.

WelO5 has been shown to be highly specific for **1**, meaning that amino acid substitutions to alter the shape of the active site will be necessary to engineer in activity against other compounds. The selection of amino acids for mutation is complicated by the lack of three-dimensional structural information for WelO5 and the absence of such information for any WelO5-homologous

enzymes. Dockings of **1** into a model of predicted WelO5 tertiary structure (Chapter 2, Section 2.2.2.) were initially used to guide the selection of residues for substitution whilst efforts towards a crystal structure were ongoing. In the meantime, access to milligram quantities of **1** meant that development could start towards a medium-throughput assay for screening larger numbers of compounds and/or different enzyme variants in parallel.

Chapter 4. Characterisation of WelO5

4.1. Introduction

The discovery and functional assignment of WelO5 was first reported in late 2014.⁵⁰ However, this report did not provide detailed characterisation of the WelO5 protein, including its three-dimensional structure. The ability of WelO5 to act upon an untethered substrate molecule had not been seen before for other members of the halogenase enzyme subfamily of 2OG-oxygenases, possibly indicating novel structural and mechanistic features. To investigate whether this proposal was true, the determination of one or more WelO5 crystal structures became necessary.

This Chapter reports studies on the characterisation of the WelO5 enzyme which was purified as described in Chapter 2 (Section 2.3.3.4.). The natural substrate, (+)-12-*epi*-fischerindole U isonitrile, was not immediately or easily accessible and whilst the substrate was being generated synthetically (Chapter 3), several characterisation techniques were used to investigate the biophysical properties of WelO5:

- *mass spectrometry* for confirmation of WelO5 molecular weight;
- *SEC-MALS (Size-Exclusion Chromatography-Multi-Angle Light Scattering)* for assessment of WelO5 aggregation and oligomerisation;
- *circular dichroism (CD)* for estimation of the WelO5 secondary structure;
- *anaerobic UV-Vis spectroscopy* for detection of the characteristic Fe²⁺ and 2OG-dependent oxygenase MLCT (Metal-to-Ligand Charge-Transfer) transitions.

Successful efforts towards the isolation of X-ray diffraction-quality crystals and the determination of three novel WelO5 crystal structures are described.

These structures are analysed for insights into the structure-function relationship of WelO5 and are compared with the crystal structures of 2OG-oxygenase wider family members. Initial solution NMR work is also described, with a 2D spectrum of ^{15}N -labeled WelO5 demonstrating good chemical shift dispersion and minimal overlapping peaks. Preliminary NMR experiments for the detection of WelO5 conformational changes upon substrate binding will also be shown, justifying further work (beyond this research project) towards the assignment of a WelO5 NMR structure.

4.2. Initial WelO5 characterisation

4.2.1. Molecular weight

The calculated (ExPASy ProtParam)¹⁵⁴ molecular weight of His₆-WelO5 is 34991.5 Da; which becomes 32525.8 Da after cleavage of the His₆-tag with the TEV-protease (a serine residue remains on the *N*-terminus).¹²⁶ LC-MS analysis of the final purified protein (isolated after gel filtration) confirmed the predicted WelO5 molecular weight of 32525.8 Da (Figure 4.1.)

through the sample and the presence of protein molecules in solution cause the beam to scatter. The intensity and angular dependence of the scattered light relative to the incoming light beam is then measured at the detector. The intensity of light scattered by a molecule (beyond that seen in the background buffer) is directly proportional to the molar mass and molecular concentration of the sample being measured. SEC-MALS also provides information on the homogeneity of a sample, reflected in the reported polydispersity index.¹⁵⁵

SEC-MALS analysis of WelO5 was performed by Weston Struwe at the Department of Biochemistry, Oxford. From SEC-MALS analysis (Figure 4.2.), the molecular weight of WelO5 was reported as 32.6 ± 0.1 kDa, consistent with monomeric WelO5. Higher molecular weight species were not detected and the polydispersity index of the sample was reported as 1.00 ± 0.04 , indicating a homogenous sample. Monodisperse samples are thought to have greater success in obtaining diffraction-quality crystals,¹⁵⁶ therefore the WelO5 sample analysed here (isolated by purification as per Chapter 2, Section 2.3.3.4.) was used for the set-up of protein crystallisation trial plates.

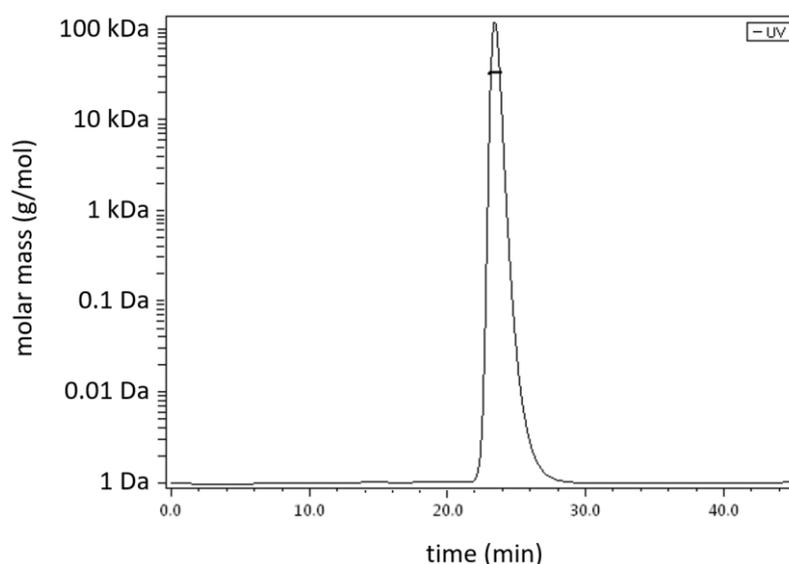


Figure 4.2. UV chromatogram from SEC-MALS analysis of TEV protease-cleaved and gel-filtration purified WelO5.

4.2.2. Circular dichroism

Circular dichroism (CD) describes a property exhibited by a molecule possessing one or more chiral chromophores.¹⁵⁷ Chirality results in differences in the absorption of left-handed and right-handed circularly polarised light, generating a non-zero CD signal (i.e. for chiral molecules, left-handed and right-handed light do not cancel each other out). This spectroscopic technique can be used to estimate protein structure because the highly-ordered arrangements of peptide bonds in protein secondary structural motifs give rise to characteristic signals in the far-UV range (190-260 nm) of CD spectra.¹⁵⁷

- α -helices show negative bands at 208 and 222 nm and a positive band at 193 nm.
- β -sheets show a negative band at 218 nm and a positive band at 195 nm.

Established algorithms can estimate the secondary structural composition of an unknown protein from the far-UV CD spectrum by comparison with databases of proteins of known structure.¹⁵⁷

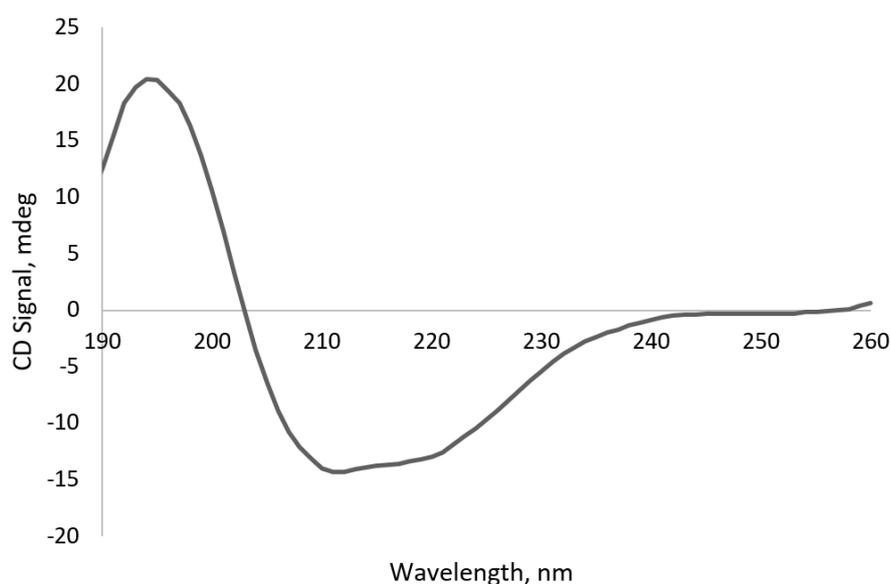


Figure 4.3. Far-UV CD spectrum of WelO5 (0.2 mg/mL) collected at 25 °C in 10 mM sodium phosphate, pH 7.5.

The far-UV CD spectrum of WelO5 was collected at 25 °C and deconvoluted using the DichroWeb web server (Figure 4.3.).¹⁵⁸ The CDSSTR algorithm was found to produce the best fit with the experimental data, estimating the secondary structural composition of WelO5 to be 17-24% α -helical and 28-32% β -sheet.¹⁵⁹ This result was compared to secondary structural analysis of the known 2OG-dependent halogenases SyrB2,¹⁰⁵ CytC3⁹³ and CurA Hal⁸⁹, as well as the 2OG-dependent hydroxylase SadA¹¹⁹ (previously identified as a template for WelO5 tertiary structural prediction - see Chapter 2, Section 2.2.2.). Secondary structural information for these four enzymes (Table 4.1.) was obtained from the DSSP¹⁶⁰ analysis associated with the PDB deposited crystal structures. The estimated α -helical and β -sheet composition for WelO5 is closest to SyrB2 (17-18% α -helical and 27-28% β -sheet), but is also not too different to that of SadA (26% α -helical and 26% β -sheet).

Description	PDB IDs	% Identity shared with WelO5	% Similarity to WelO5	α -helical % (DSSP)	β -sheet % (DSSP)
SyrB2	2FCT 2FCU 2FCV	14	25	17-18	27-28
CurA Hal	3NNF 3NNJ 3NNL 3NNM	11	20	36-38	16-18
CytC3	3GJA 3GJB	7	11	16-17	24-25
SadA	3W20 3W21	19	33	26	26

Table 4.1. Comparison of WelO5 with 2OG-dependent halogenases of known structure, SyrB2, CytC3 and CurA Hal and 2OG-dependent oxygenase SadA. Values reported for identity and similarity were calculated using EMBOSS Needle pairwise alignment. Secondary structural assignment ranges were summarised for each listed crystal structure from the DSSP analysis attached to the structures within the PDB.

4.2.3. UV-Vis spectroscopy

Anerobic UV-Vis spectroscopy was performed using 200 μ M WelO5. Spectra were collected for the apo enzyme (.....), and after the addition of 2OG (- - -),

then Fe^{2+} (— · ·) and chloride (—) respectively. The overlaid spectra are shown in Figure 4.4.

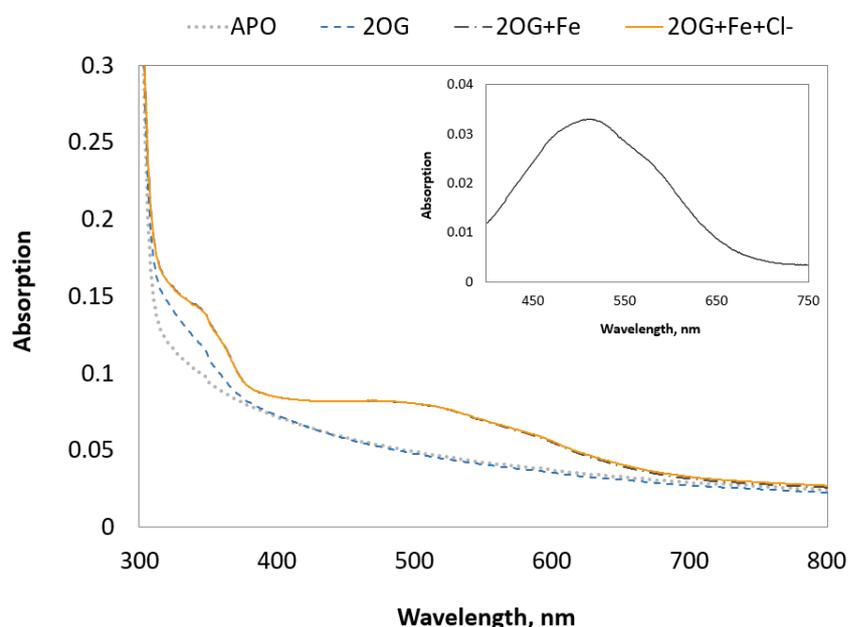


Figure 4.4. UV-Vis spectrum of 200 μM anaerobic apo WelO5 (grey dots), after the addition of 2 μM 2OG (blue dashes), then 1.25 μM Fe^{2+} (grey dash), then 2 μM chloride (solid orange). MLCT band appears at approximately 500 nm after the addition of iron. Further increases are not seen upon the addition of chloride. Inset: difference spectrum showing the MLCT for Fe^{2+} -2OG binding at approximately 515 nm.

Upon the addition of Fe^{2+} to the WelO5-2OG mixture, a metal to ligand charge transfer (MLCT) band was observed at approx. 520 nm, a characteristic transition for the Fe^{2+} and 2OG complex in 2OG-dependent oxygenases (Figure 4.4., — · ·).¹⁶¹ This indicates the presence of metal-ligand orbital overlap, resulting from the chelation of 2OG to the metal centre at the WelO5 active site (a pre-requisite for oxygen binding). This MLCT transition is similar to that reported for other 2OG-oxygenases.^{162,163} Addition of an excess of chloride (2 μM NaCl vs. 1.25 μM iron) did not result in further changes to the absorption spectrum at this enzyme concentration (Figure 4.4., —).

4.3. Structural characterisation of WelO5

Whilst preliminary characterisation was being carried out for WelO5, work towards a WelO5 crystal structure was ongoing (crystal structures for WelO5 were not published in the literature until 2016).¹⁶⁴ It was hoped that crystal structures of this carrier-protein independent 2OG-halogenase would provide some insight into WelO5's novel function and guide the rational selection of active site residues for substitution.

4.3.1. X-ray crystallography

The WelO5 protein isolated in Chapter 2 was confirmed to be homogenous (SEC-MALS), highly pure (SDS-PAGE) and of the correct molecular weight (LC-MS). Trace metal analysis (ICP-MS) and metal and compound-induced binding stability experiments (thermal shift assays) were performed. Finally sitting-drop crystallisation screens were set up using purified WelO5.

4.3.1.1. ICP-MS

Inductively coupled plasma mass spectrometry (ICP-MS) uses plasma to ionise the trace elements present in a sample.¹⁶⁵ ICP-MS is commonly used for trace metal analysis due to its high sensitivity and the ability to measure multiple elements simultaneously.¹⁶⁶ It is particularly useful for protein samples as the presence of trace metals can be quantitatively determined.¹⁶⁶

The WelO5 active site contains two histidine residues which form the enzyme's catalytic centre through co-ordination to Fe^{2+} . In the absence of iron, these two histidines have the capacity to bind other divalent metals. Any bound metal present in the crystal structures would need to be correctly assigned.

Additionally, the binding of a non-iron divalent metal would inactivate the

enzyme - impacting WelO5 activity studies. ICP-MS was therefore used to confirm the absence of bound metals. The apo protein used to set up crystallisation screens was analysed by ICP-MS to detect and quantify the trace metal content and assess whether the WelO5 present in the sample had bound to trace divalent metals during expression or purification (Ni^{2+} was a particular concern due to the nickel-charged resin used for IMAC purification).

ICP-MS sample preparation and analysis was performed by Kathrin Abraham & Phil Holdship at the Department of Earth Sciences, Oxford. Samples were analysed for iron, nickel, manganese, zinc, cobalt and copper content. The sample of purified WelO5 was found to contain less than 0.05% of each of the analysed metals, with the levels of manganese and cobalt below the method detection limit. Therefore, the active site metal ion in any crystal structure produced could be correctly assigned by doping the apo protein solution with a selected divalent metal prior to crystallisation.

4.3.1.2. Thermal shift assay (metal-binding)

Whilst it may inactivate the enzyme, the introduction of a non-iron divalent metal species to WelO5 could aid crystallisation by helping to stabilise the structure of WelO5 (WelO5 catalytic inactivity due to iron replacement with Mn^{2+} , Ni^{2+} , Co^{2+} and Cu^{2+} has already been demonstrated in the literature).⁵⁰ Access to a catalytically inactive (but biologically relevant) form of WelO5 would facilitate the set-up of crystallisation screens under aerobic conditions using a semi-automated process, avoiding the practical challenges associated with setting up anaerobic crystallisation screens in a glove-box.

A ThermoFluor thermal shift assay was used to assess the stability introduced by the binding of different divalent metals to WelO5.¹⁶⁷ In this assay, protein samples are heated to induce unfolding. A hydrophobic dye (SYPRO® Orange)¹⁶⁸ is present which binds to the increasingly exposed hydrophobic regions of the protein as it unfolds, resulting in a measurable increase in fluorescence.¹⁶⁹ Changes in fluorescence intensity are plotted as a function of temperature to calculate the melting temperature of the protein (T_m ; the mid-point between folded and fully unfolded states).¹⁶⁹ Increases or decreases in the T_m value reflect a change in stability, with an increase in T_m indicative of stabilisation of the folded protein form (and an increased propensity to form crystals).¹⁶⁹ This technique can be used to assess the impact of different conditions upon protein stability (e.g. buffer optimisation, ligand stabilisation, enzyme mutations).¹⁷⁰

Divalent metal ions of manganese, zinc, nickel, cobalt and copper (50 μ M), were tested with WelO5 (6 μ M) in the presence and absence of 5 mM NOG (*N*-oxalyl glycine; a catalytically inactive 2OG analogue),¹⁷¹ with the results summarised in Figure 4.5. Iron was included in this experiment but, as oxygen was not excluded during the assay, was assumed to be present in the oxidised (and likely catalytically inactive) Fe^{3+} form. The T_m of apo WelO5 was measured as 54.5 °C. Nickel was found to provide the greatest level of protein stabilisation relative to the other metals tested (+/- NOG; 10.7 °C and 4.5 °C) and was selected as an Fe^{2+} replacement for the set-up of aerobic crystallisation plates.

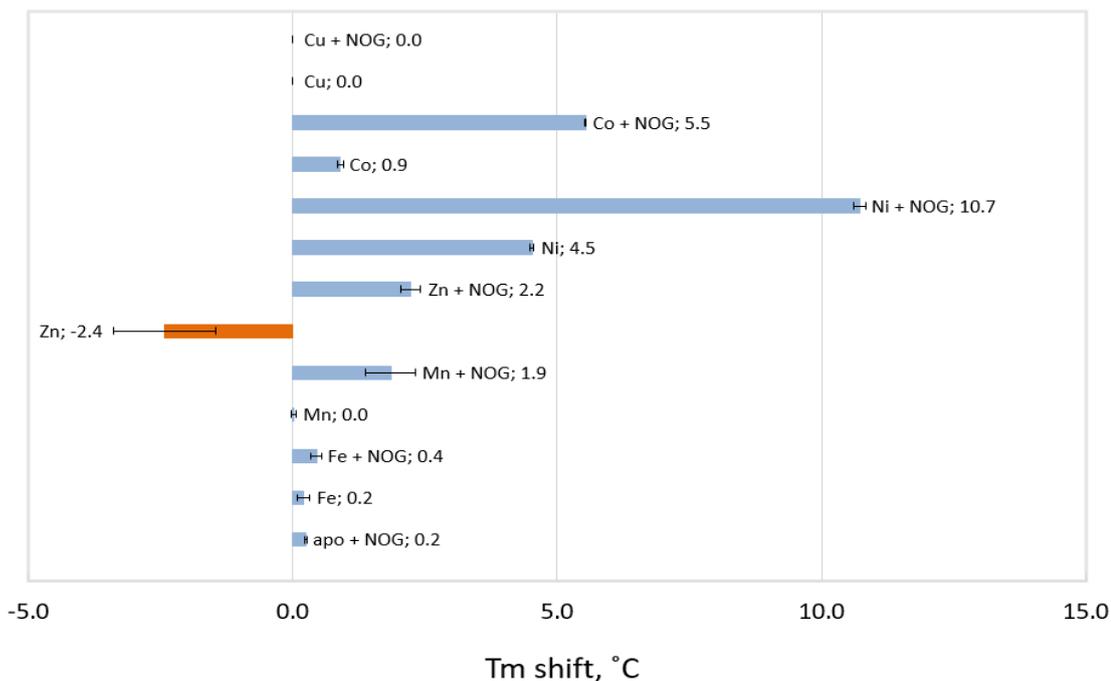


Figure 4.5. Thermal shift results reported for varied metal ThermoFluor assay. Changes in T_m were calculated relative to apo (i.e. metal and ligand free) WelO5, values are reported next to data bars. Error bars reflect the standard deviation of the data. The assay was performed using 6 μM WelO5 in the presence and absence of 5 mM NOG and 50 μM metal.

4.3.1.3. Thermal shift assay (ligand-binding)

WelO5 crystallisation plates set up with Ni^{2+} alone as well as both Ni^{2+} and NOG failed to yield any diffraction-quality protein crystals (salt crystals or microcrystals only). Alternative ligands were sought to introduce stability upon binding to WelO5, encouraging crystallisation. Whilst synthesis of the enzyme's natural substrate was ongoing, a few milligrams of the synthetic ketone precursor, compound **19** (Figure 4.6.) was available.

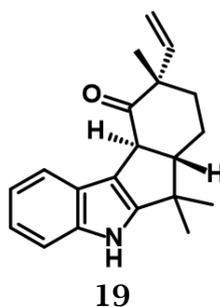


Figure 4.6. Structure of (6a*S*,9*S*,10a*S*)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5*H*-indeno[2,1-*b*]indol-10-one, compound **19**.

Compound **19** is structurally related to the WelO5 substrate (+)-12-*epi*-fischerindole U isonitrile and WelO5 chlorination activity on **19** had been demonstrated in the initial activity screens (Chapter 3, Section 3.4.). Both these factors created assurances about WelO5 active site-binding for this compound. Co-crystallisation plates were set up using WelO5 with compound **19**, Ni²⁺ and 2OG.

A second ThermoFluor assay was used to assess the WelO5 stability-induced binding of an existing panel of 2OG-mimetics (Schofield lab 2OG-oxygenase inhibitor collection). The compounds within this set were known to (or were designed to) bind in the same manner as 2OG, by chelation to the histidine coordinated metal within the catalytic site of a 2OG-oxygenase. As a result, the ThermoFluor screen was performed in the presence of Ni²⁺ (without a divalent metal present, the compounds did not demonstrate a T_m shift).

Approximately one hundred 2OG-mimetic compounds were screened in this ThermoFluor assay using the same conditions as for the divalent metal screen, except for a reduced compound concentration (the 2OG-mimetic panel was available as 1mM stocks in DMSO; used at 100 µM assay concentration). Under these assay conditions, NOG produced a T_m shift of 2.1 °C (*cf.* Section 4.3.1.2.) and 2OG produced a T_m shift of 6.3 °C. From the 100 compounds screened, 15 were found to produce a shift of ≥ 3 °C. The top five were re-screened for confirmation and their identities confirmed by LC-MS of the 1mM DMSO stocks. Three structurally differentiated compounds were selected for co-crystallisation with WelO5 (Table 4.2.), based upon their T_m shift values. Two compounds required re-synthesis and the third, Ciclopirox, was commercially available.

ID	33	34	35
Name	5-[2-(isoquinoline-3-carbonyl)hydrazino]-5-oxopentanoic acid	(2S)-3-[(3-methoxyphenyl)methylsulfonyl]-2-(oxalamino)propanoic acid	6-cyclohexyl-1-hydroxy-4-methyl-pyridin-2-one (Ciclopirox)
T _m shift, °C	+ 5.1	+ 5.5	+ 4.4
Target	HIF Prolyl Hydroxylases ¹⁷²	JMJD2 Histone Demethylases ¹⁷³	Mammalian Prolyl Hydroxylase, HIF PH ¹⁷⁴

Table 4.2. The three compounds selected from the ThermoFluor 2OG-mimetic assay. Screening was performed with 6 μM WelO5, 50 μM Ni²⁺ and 100 μM 2OG-mimetic. Values shown are the average of three separate assay runs (i.e. n=3), one of which included the re-synthesised compounds.

However, before crystallisation trials were initiated with these compounds, the 2OG and compound **19** co-crystallisation plates had yielded well diffracting WelO5 crystals. Whilst co-crystallisation with these 2OG-mimetics was not carried out, it was considered that crystal structures of WelO5 with these compounds bound may still be useful in guiding the future design and selection of putative substrates for WelO5 (and WelO5 variants).

4.3.1.4. WelO5 structure determination

As previously mentioned (Section 4.3.1.3.), co-crystallisation attempts with WelO5 in the presence of Ni²⁺ alone or with Ni²⁺ and NOG failed to yield any protein crystals. This was despite the large number of screening conditions tested (96-3 well plates using SaltRx™, PACT premier™, PEG/Ion™, JCSG-*plus*™, MIDAS™, Wizard™ and Index™ at 4 °C and 20 °C; 4,032 conditions). The synthetic ketone precursor **19** (Figure 4.6.) was used to set up further WelO5 co-crystallisation screens.

Protein samples were prepared for crystallisation by incubating with Ni²⁺, 2OG and 5 mM ketone compound **19** in 20 mM HEPES buffer, passed through a PD-10 desalting column (GE Healthcare) and concentrated back down to the initial protein concentration (Millipore Amicon® Ultra 10 kDa, 4 mL). An initial broad screen of crystallisation conditions (96-3 well plates using SaltRx™, PACT premier™, PEG/Ion™, JCSG-*plus*™, Index™ and Crystal Screen™ at 4 °C and 20 °C; 3,456 conditions) yielded five hits (all from 20 °C plates) in under seven days. Two of these conditions were chosen for optimisation: 0.1 M MES pH 6.0, 0.2 M NaCl, 20% w/v PEG 6000 (PACT premier™) and 0.2 M CaCl₂, 20% w/v PEG 3350 (PEG/Ion™) (Figure 4.7.).

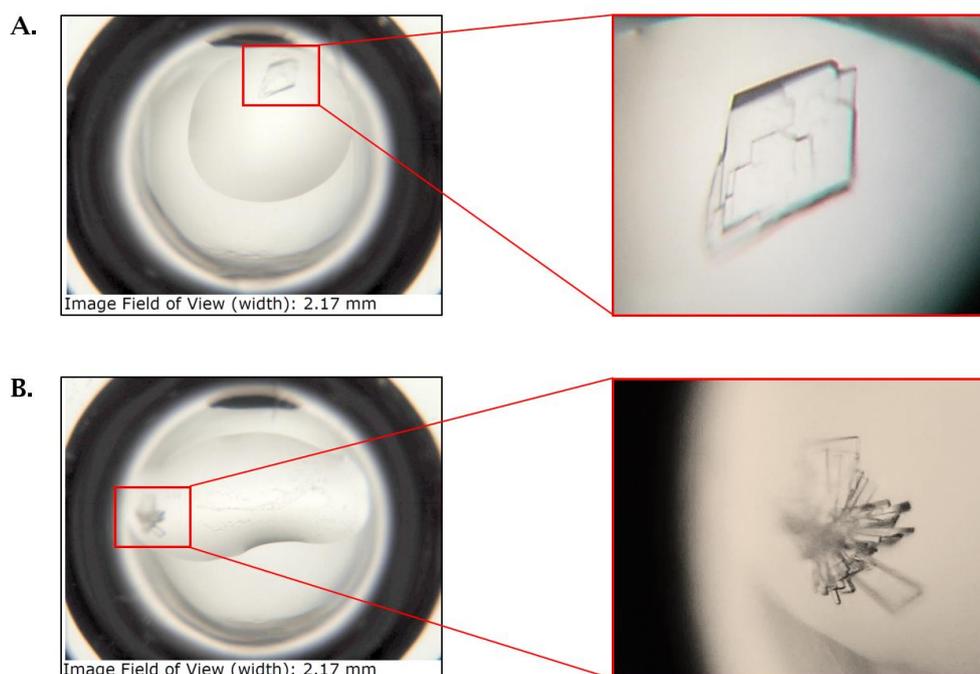


Figure 4.7. Images of the two hits selected from the WelO5 broad screen for further optimisation. Image A: Single crystal grown in a well solution of 0.1 M MES pH 6.0, 0.2 M NaCl, 20% w/v PEG 6000 (PACT premier™) and Image B: Crystals grown in a well solution of 0.2 M CaCl₂, 20% w/v PEG 3350 (PEG/Ion™). The crystals shown in image B were used to obtain the WelO5 structure 5T22.

Optimisation plates developed from these initial hit conditions were then set up at 20 °C. Variable condition matrix plates were created based upon the PACT premier™ broad-screen hit (Figure 4.7., image A.), containing 0-0.35 M NaCl, 10-32% w/v PEG 6000 in 0.1 M MES at pH 5.5, pH 6.0 and pH 6.5 (plates

named OP_1, OP_2 and OP_3, respectively). The pH 5.5 plate (OP_1) yielded six hits and the pH 6.0 plate (OP_2) yielded two hits, all within three days. A single variable condition optimisation matrix plate was also created based upon the PEG/Ion™ broad-screen hit (Figure 4.7., image B.), containing 0-0.35 M CaCl₂ and 10-32% w/v PEG 3350 (named OP_4). This yielded two further hits, also within three days. Data sets were collected (Michael McDonough, Schofield Group) for crystals harvested from these optimisation plates and any crystals which had subsequently grown in the initial broad-screen plates (PACT premier™, PEG/Ion™, JCSG-*plus*™ and Crystal Screen™). During harvesting, some of the crystals were soaked in NaBr or 5-amino-2,4,6-triiodoisophthalic acid (I3C). The presence of heavy atoms has been shown to aid structure determination by introducing strong anomalous signals which act as markers for phase determination.¹⁷⁵ I3C contains three equally spaced iodine atoms which form an easily identifiable, characteristic 'magic triangle'.¹⁷⁶

The data set for PDB ID: 5J4R was obtained from crystals grown within three days from a well solution of 0.1 M MES pH 5.5, 0.1 M NaCl, 22% w/v PEG 6000 (I3C soak). The data set for PDB ID: 5T22 was obtained from the original broad screen PEG/Ion™ hit (Figure 4.7., image B.). Unfortunately, WelO5 was not structurally similar enough to protein structures available from the PDB (including SadA: PDB 3W20 & 3W21) to be able to solve the structure by molecular replacement. An ensemble approach using the top ten most similar protein structures as suggested by Phyre2 also failed (low identity resulted in a poor ensemble structure). In addition, strong anomalous signals were not visible from any of the NaBr or I3C soaked crystals.

Selenomethionine analogues of proteins are routinely used to phase protein structures.^{177,178} The recombinant introduction of selenomethionine is relatively straightforward and results in a site-specific source of anomalous phasing signal, which can be used to determine protein structure. SeMet-WelO5 was expressed on a 2.4L scale (BL21 (DE3) cells, 0.5 mM IPTG at 30 °C, 20 hr) and purified as per Chapter 2 (Section 2.3.3.4) for use in crystallisation trials (Figure 4.8.). 3 mL of Se-Met WelO5 was isolated at 35 mg/mL (expression yield of 45 mg L⁻¹).

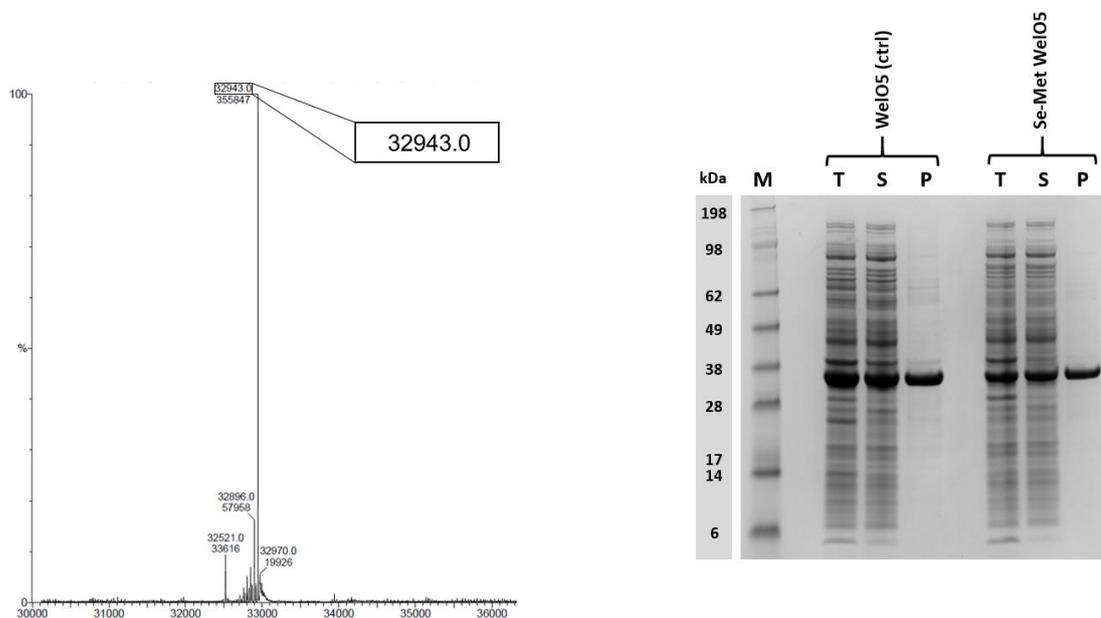


Figure 4.8. Left: Mass spectrum showing the molecular weight of the TEV protease-cleaved and SEC purified Se-Met WelO5 used for the set-up of crystallisation trials. Se-Met WelO5: 32947 Da. Right: SDS-PAGE gel comparing expression levels for WelO5 against those of Se-Met WelO5, both expressed and purified under the same conditions [BL21 (DE3) cells, 30 °C, 0.5 mM IPTG, 20 hr; Ni-NTA & gel filtration]. The protein molecular weight marker is labelled M. Re-suspended samples are labelled T, the soluble fractions are labelled S and the purified samples labelled P.

Screening conditions were selected based on their prior ability to yield more than one WelO5 crystal per plate. Three commercially available broad screens (PACT premierTM, PEG/IonTM and JCSG-*plus*TM) and three optimised matrix condition plates (OP_1, OP_2 and OP_4) were selected and set up in a manner analogous to before (using compound **19**). Multiple SeMet-WelO5 crystals grew across these plates, most of which possessed similar morphology (examples

shown in Figure 4.9.). A SeMet-WelO5 data set was obtained (Ian Clifton & Michael McDonough, Schofield Group) from crystals grown in five days from a well solution of 0.2 M ammonium chloride, 20% w/v PEG 3350 (JCSG-*plus*[™]) (Figure 4.10.).

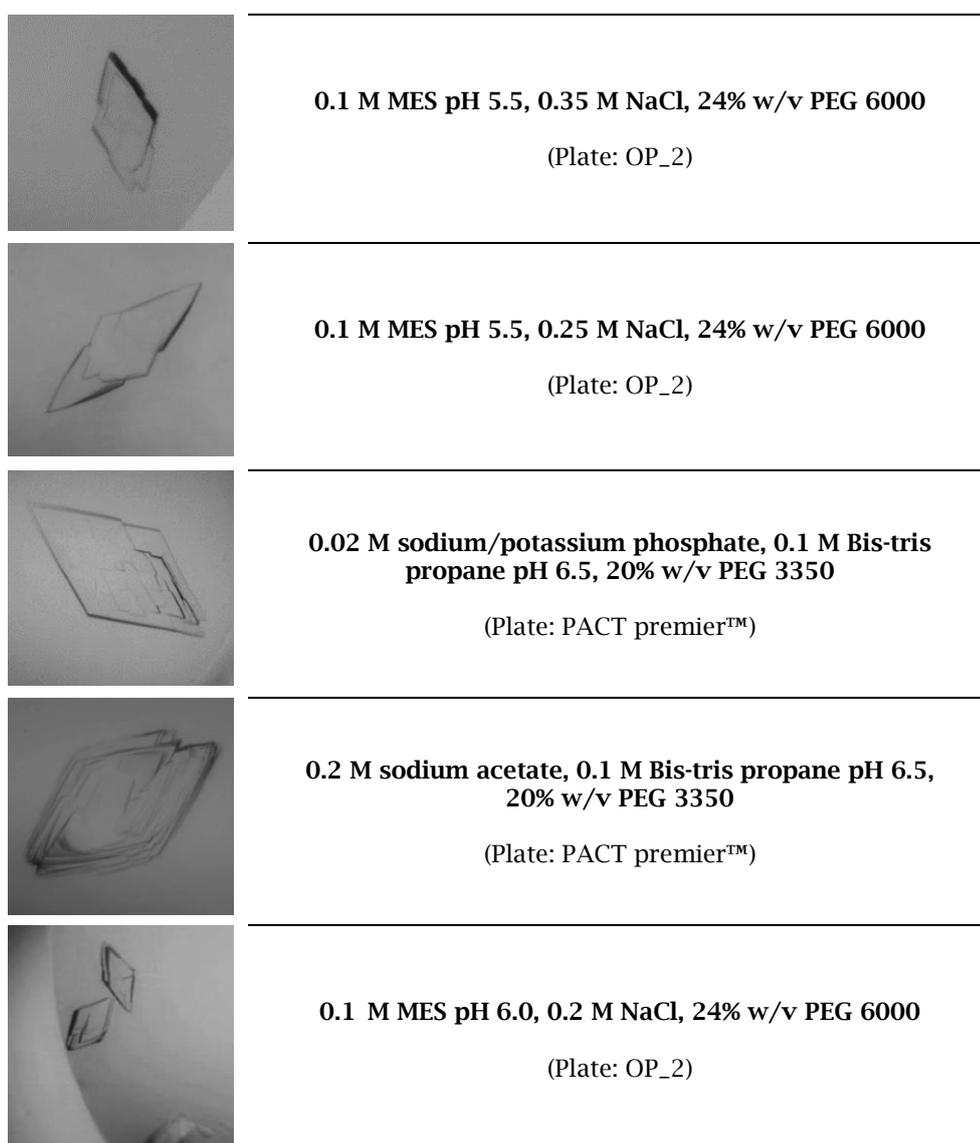


Figure 4.9. Examples of SeMet-WelO5 crystals demonstrating similar morphology. These were obtained from crystallisation trials set up with screening plates selected for their proven ability to yield WelO5 crystals.



Figure 4.10. Crystals of SeMet-WelO5 found in a JCSG-plus™ broad screening plate. The single crystal used to solve the SeMet structure by single-wavelength anomalous diffraction (SAD) was harvested from this well.

A broad crystallisation screen was also set up for WelO5 in the presence of Ni²⁺ and succinate, leading to a crystal structure with both these components present in the active site (PDB ID: 5TRQ; data collection Michael McDonough, Schofield Group). 5TRQ crystals grew within 5 days from a well solution of 0.1 M MES pH 6.0, 0.2 M Ca(OAc)₂, 20% w/v PEG 8000 (Wizard™).

Despite the presence of the ketone compound **19** in the crystallisation conditions used for SeMet-WelO5, 5J4R and 5T22, the compound was not found to be present in any of the resultant structures (a region of electron density consistent with the size and/or shape of this compound was not seen at the active site). By better simulating the conditions used for the WelO5 chlorination assay, it was proposed that crystals may be more likely to grow with **19** bound in the WelO5 active site. 24-well hanging drop co-crystallisation plates were set up with compound **19**, Fe²⁺ and 2OG under anaerobic conditions (i.e. glove box) to prevent catalytic turnover through the exclusion of oxygen. Three different ratios of protein to well solution were set up per well reservoir (2:1, 1:1 and 1:2; drop size 200-300 μL) using a subset of the conditions from the pH 5.5 PACT premier™ optimisation plate (OP_1) and pH

6.0 PACT premier™ optimisation plate (OP_2) developed previously. Crystals appeared in needle-like clusters after 10 days and were harvested after 14 days. The harvested crystals demonstrated very poor diffraction, i.e. 5Å to 10Å. However, before these conditions could be optimised and crystal soaking experiments carried out, four WelO5 crystal structures were released to the PDB (Liu group: 5IQS, 5IQT, 5IQU & 5IQV).¹⁶⁴ Three of these were found to contain the natural substrate (+)-12-*epi*-fischerindole U isonitrile in the enzyme active site. As these crystal structures could guide rational re-design of the WelO5 active site through amino acid substitutions, further crystallisation work with **19** or the natural substrate was deemed unnecessary.

4.3.1.5. Observations from WelO5 crystal structures

4.3.1.5.1. Overall structure

The three crystal structures of WelO5 (summarised alongside the Liu group PDB structures in Table 4.3.) were obtained in different space groups; *I*222 (5J4R), *P*12₁1 (5T22) and *P*1 (5TRQ), none of which had been described previously (the Liu group structures are *P*2₁2₁2₁; *C*121 for the G166D variant).¹⁶⁴ These structures were also of higher resolution than those in the literature (range 2.00-2.51Å); with 5J4R at 1.65Å, 5T22 at 1.75Å and 5TRQ at 1.30Å. All three structures were modelled with Ni²⁺ in the active site, bound to 2OG (5J4R and 5T22) or succinate (5TRQ). Unfortunately, despite using compound **19**, a substrate-like molecule with demonstrated WelO5 activity for co-crystallisation (Figure 4.6.), none of the structures contained this ligand.

PDB Code	5J4R	5T22	5TRQ	5IQS	5IQT	5IQU (G166D variant)	5IQV
Author	Schofield group			Liu group			
Space group	<i>I</i> 222	<i>P</i> 12 ₁	<i>P</i> 1	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>C</i> 121	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Resolution	1.65 Å	1.75 Å	1.30 Å	2.00	2.40	2.51	2.40
Active site	Ni ²⁺ + AKG	Ni ²⁺ + AKG	Ni ²⁺ + SIN	Fe ²⁺ + AKG + Cl	Fe ²⁺ + AKG + Cl + 6CU	Fe ²⁺ + AKG + 6CU	Fe ²⁺ + AKG + Cl + 6CU + NO
Molecules per ASU	1	2	2	3	3	1	3
Chain breaks (no. of missing residues)	I232-Q236 (5)	B: S216-Y224 (9)	none	none	C: K217-A219 (3)	V90-K97 (8) F169-P173 (5)	A: A215-K217 (3) B: S216-A219 (4) C: Y218-E223 (6)
Deposition date	2016-04-01	2016-08-23	2016-10-27	2016-03-11	2016-03-11	2016-03-11	2016-03-11

Table 4.3. Summary of the WelO5 crystal structures deposited in the PDB. PDB ligand codes used: AKG = 2-oxoglutaric acid (2OG); SIN = succinic acid/succinate; 6CU: (+)-12-epi-fischerindole U isonitrile.

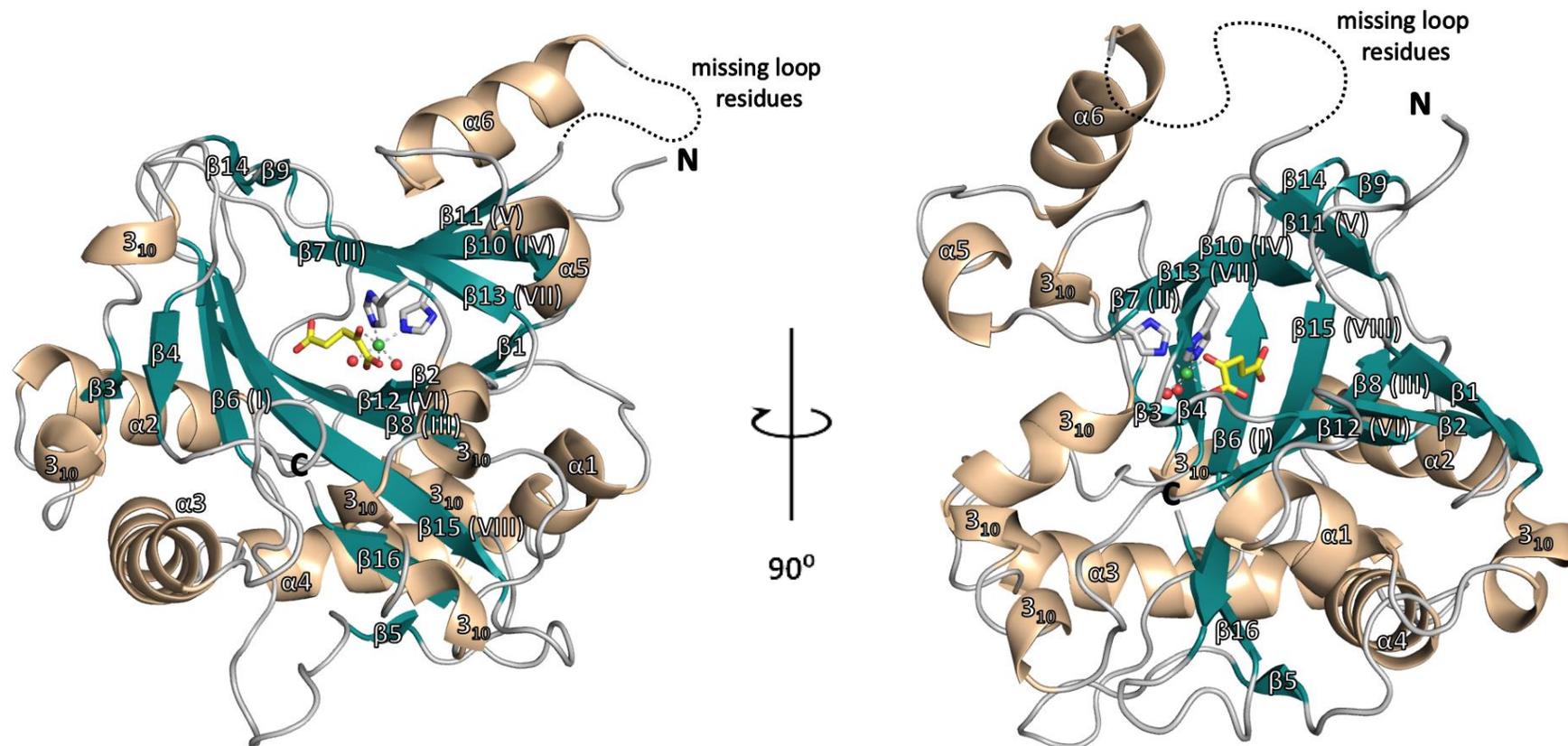


Figure 4.11. An overall view of the structure of WelO5 (5J4R is shown; but the structure is representative of 5T22 and 5TRQ). The side chains of key metal binding residues His164 and His259 are shown as sticks, 2OG shown in yellow, the Ni^{2+} centre as a green sphere and the two co-ordinating waters as red spheres. Secondary structure is numbered sequentially, with the eight β -strands of the DSBH also annotated with numerals I-VIII. The β IV-V insert residues missing in this crystal structure (Ile232-Gln236) are simulated with a dashed line. Figure generated in PyMOL.

The overall structure of WelO5 (Figure 4.11.) was found to be consistent with other members of the 2OG-oxygenase superfamily; with the active site sandwiched between two sheets of a double-stranded beta-helix (DSBH) consisting of eight anti-parallel β -strands (typically denoted by roman numerals I-VIII).⁶³ WelO5 consists of sixteen β -strands, six α -helices and six 3_{10} helices (28-31% α -helical and 24-26% β -sheet by DSSP analysis¹⁶⁰; *cf.* Section 4.2.2.). The characteristic DSBH fold is comprised of a major β -sheet formed from β -strands $\beta 6$ (I), $\beta 15$ (VIII), $\beta 8$ (III) and $\beta 12$ (VI) and a minor β -sheet formed from strands $\beta 7$ (II), $\beta 13$ (VII), $\beta 10$ (IV) and $\beta 11$ (V).

For all three of these WelO5 structures, up to ten of the N-terminal residues could not be modelled due to disorder. Interestingly, the C-terminus and the $\beta 16$ strand immediately preceding it coils back in toward the active site, with the $\beta 16$ strand sitting anti-parallel with $\beta 15$ (VIII) of the major β -sheet of the DSBH. The C-terminal Ser290 side chain is positioned to form a hydrogen bond to the hydroxyl of Thr151 (2.9 Å), as well as interactions with the Ser290 carboxylate, the hydroxyl of Tyr99 (2.6 Å) and backbone nitrogen of Ile89 (2.8 Å). However, from these structures alone and despite its location proximal to the active site, the role the C-terminus might play in substrate binding (if at all) is not immediately obvious.

On a global level, 5J4R, 5T22 and 5TRQ appear consistent in all but three areas of their tertiary structures;

- i) Glu75 – Ala82; the loop joining $\beta 4$ and $\beta 6$ (I);
- ii) Ile89 – Gln92; the loop joining $\beta 4$ and $\alpha 3$;
- iii) Trp210 – Glu238; part of a helical insert between $\beta 10$ (IV) and $\beta 11$ (V).

These three areas all border entry to the active site, with Trp210 - Glu238 showing the most significant conformational change between structures (Figure 4.12.). This region, between β 10(IV) and β 11(V) (henceforth called the β IV-V insert), is a common area of variability amongst the 2OG-oxygenases, varying in size from a small loop to an entire domain depending on the sub-family.⁶³ Differences in this region relative to other 2OG-halogenases could potentially be responsible (at least in part) for the novel carrier-protein independent activity of WelO5. However, this theory is not conclusive due to the challenges presented by drawing structural comparisons from proteins sharing low overall homology with WelO5.

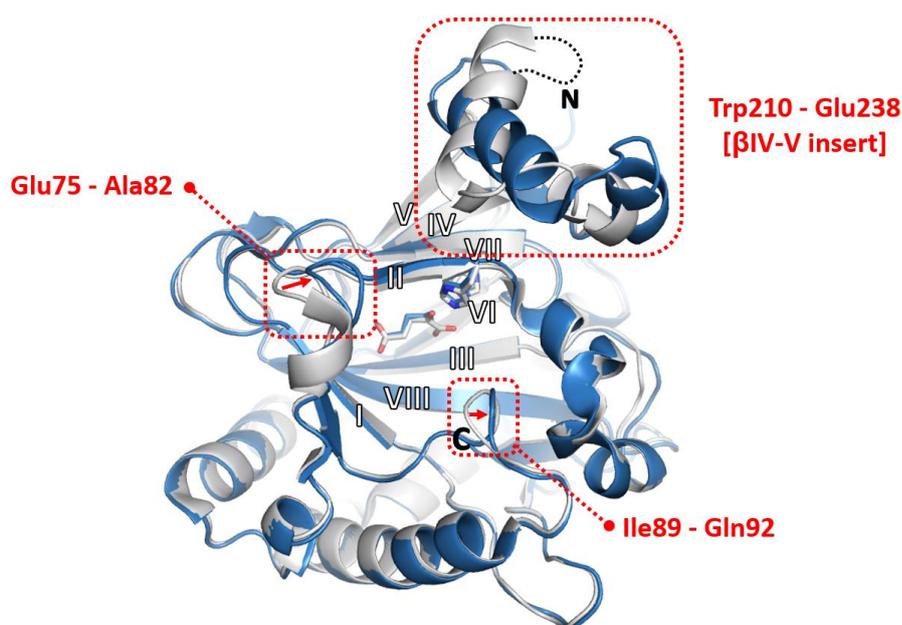


Figure 4.12. Overlays of WelO5 structures 5J4R (white) and 5TRQ (blue) illustrating the conformational changes seen between the three structures of WelO5 (structure 5T22 has been omitted for clarity but both chains of 5T22 exists in conformations somewhere between these two extremes). Regions of conformational change are highlighted in red boxes, with the range of motion indicated by red arrows. Missing residues are represented with a black dashed line. Figure generated in PyMOL.

In the 5J4R and 5T22 (chain B) structures, some of the residues of the β IV-V insert could not be modelled due to disorder, implying a level of flexibility. Flexibility for the β IV-V insert, as well as for the Glu75 - Ala82 region (part of loop connected to β I of the DSBH) was also hinted at in the crystal structures

by higher B -factors (Figure 4.13).¹⁷⁹ For example, in the 5J4R structure, the B -factors in some parts of the Glu75 - Ala82 loop and the β IV-V insert region are as high as 81 and 90, respectively - relative to an average of 30 for the whole of WelO5.

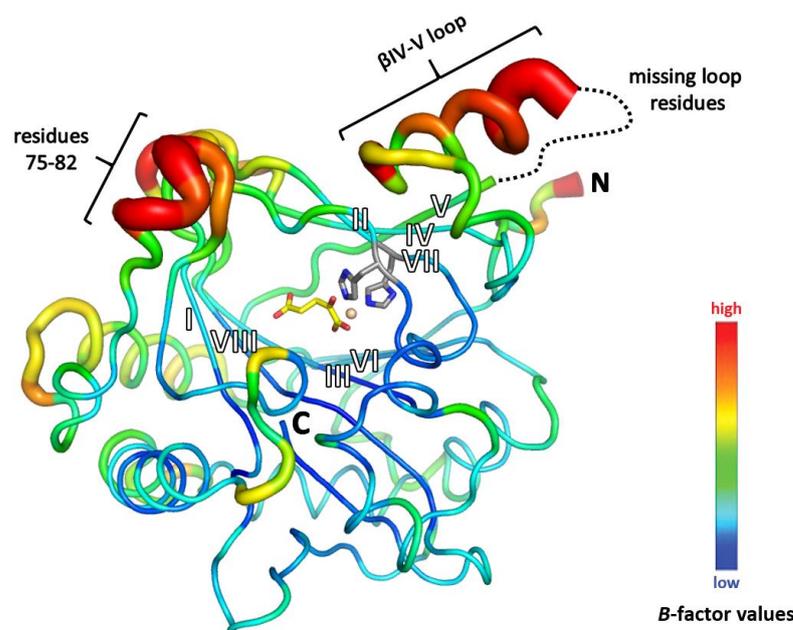


Figure 4.13. View from the WelO5 crystal structure 5JQR, coloured according to the PyMOL B -factor putty script. Areas with higher temperature factors correspond to larger fluctuations and are represented by thicker cylinders and red-orange shading. Higher average B -factors are seen for the β IV-V loop and the loop composed of residues 75-82. β IV-V insert residues 232-236 were highly disordered, with insufficient electron density to model into. These residues are therefore missing from the crystal structure and are represented by a dashed line. Figure generated in PyMOL.

The Liu group structures were found to show similar conformational changes (both in size and directionality) for the regions highlighted in Figures 4.12. and 4.13.; however, two of their structures (5IQT, chain B; 5IQV, chains A & B) also show more significant β IV-V insert movements relative to the 5J4R and 5T22 structures.¹⁶⁴

The two conformational extremes of the WelO5 β IV-V loop are illustrated in Figure 4.14. In the 'closed' form structures (exemplified here by 5IQT, chain B), the β IV-V loop almost completely excludes the WelO5 active site from the bulk

solvent, increasing the number of residues in contact with the substrate. If movement of the β IV-V insert is induced by substrate binding, this could explain why we did not see this ‘closed’ conformation for our substrate-free structures.

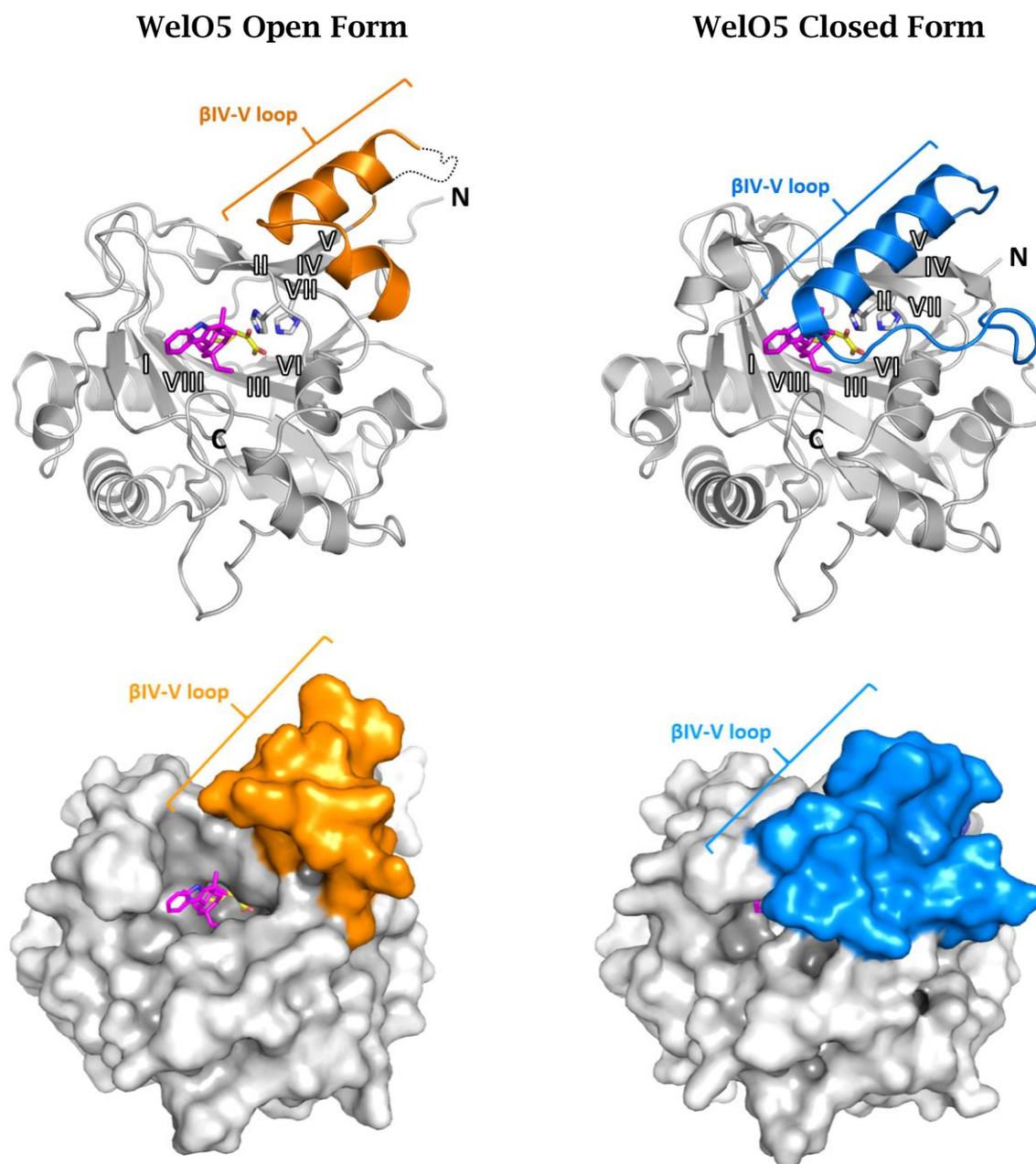


Figure 4.14. Two WelO5 structures, each shown in both cartoon and surface representation, illustrate the most extreme ‘open’ (PDB: 5JQR) and ‘closed’ (PDB: 5IQT, chain B) enzyme conformations seen. The β IV-V loop region is highlighted in orange in the open form and blue in the closed form. WelO5 natural substrate (+)-12-epi-fischerindole U isonitrile is coloured pink, 2OG is coloured yellow and the metal co-ordinating His164 and His259 residues are shown in stick form. For 5JQR, β IV-V loop residues 232-236 were highly disordered with insufficient electron density to model into. These residues are missing from this crystal structure and represented by a dashed line. Figure generated in PyMOL.

However, the existence of exclusively 'open' forms for our structures could (at least, in part) also be a crystallisation artefact, caused by blocking of the active site region by a symmetry mate. This proposal is supported by the fact that not all the substrate-containing protein chains of the Liu structures exist in the 'closed' conformation. WelO5-substrate solution studies (e.g. NMR) would be required at this point to further investigate the relationship between conformation and substrate binding.

4.3.1.5.2. WelO5 active site

As described in Section 4.3.1.2., Ni²⁺ was found (by ThermoFluor assay) to provide significant structural stabilisation to WelO5. Using a non-Fe²⁺ metal also allowed the crystallisation screen to be set up aerobically. For these reasons WelO5 was crystallised with Ni²⁺ and this metal is therefore modelled in place of Fe²⁺ in our crystal structures.

The WelO5 active site exhibits the typical 2OG-oxygenase (halogenase) octahedrally co-ordinated metal ion, chelated in a bidentate manner by 2OG and by two histidine ligands, His164 and His259. His164 is part of strand β 7(II) of the DSBH, and His256 part of strand β 13(VII). In addition to binding of the 2OG C-1 and C-2 carbonyls to the metal centre, the C-1 carbonyl binds to Ser189 (2.6 Å) and the 2OG C-5 carboxylate binds to Arg153 (2.9 Å), Arg270 (2.8 Å & 2.9 Å) and Thr272 (2.5 Å) (Figure 4.15.). Two water molecules occupy the vacant fifth and sixth metal co-ordination sites. The water in the axial co-ordination site, trans to His259 (W1), is predicted to be displaced upon substrate binding and, following the 2OG-oxygenase consensus mechanism, subsequently replaced by an oxygen molecule. The water molecule in the same plane as 2OG and His164 (W2) occupies the halide binding site, permitted by

the existence of Gly166 (aspartate or glutamate residues exist in the analogous position in most non-halogenase 2OG-oxygenases). WelO5 variant G166D (PDB ID: 5IQU) has been shown to reconstruct the 2OG-oxygenase facial triad, converting WelO5 into a hydroxylase.¹⁶⁴

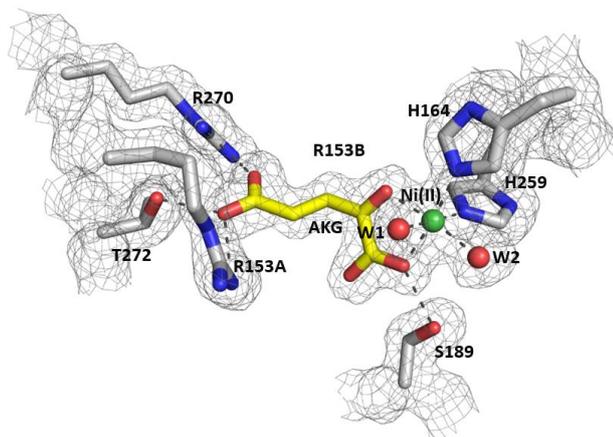


Figure 4.15. Key WelO5 active site residues are shown as sticks with the corresponding $2F_o - F_c$ electron density map contoured to 2σ (grey). His164 and His259 co-ordinate the metal centre. 2OG (AKG), in yellow, binds to Arg270, Thr272, Arg 153 and Ser189. The active site of 5J4R is shown, but is also representative of the active site of 5T22. Figure generated in PyMOL.

In 5TRQ, succinate exists in a similar conformation to 2OG. The C-4 carboxylate (analogous to the 2OG C-5) binds to the same residues as 2OG; Arg153 (2.8 Å), Arg270 (2.7 Å & 2.9 Å) and Thr272 (2.6 Å), however, there is no longer a carbonyl group available on the succinate to hydrogen bond to the Ser189 as for the 2OG C-1 (Figure 4.16.).

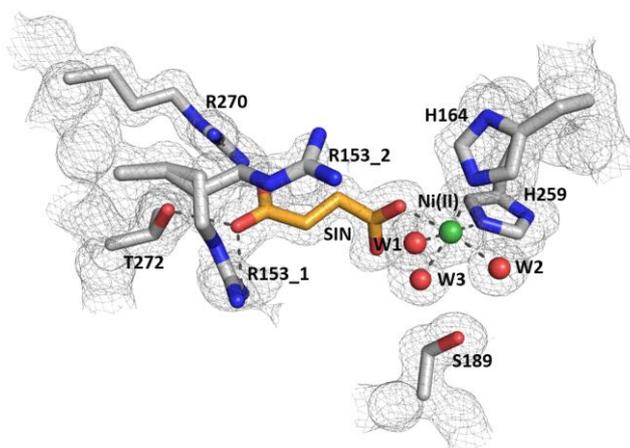


Figure 4.16. Key WelO5 active site residues of 5TRQ are shown as sticks with the corresponding $2F_o - F_c$ electron density map contoured to 2σ (grey). His164 and His259 co-ordinate the metal centre. Succinate (SIN), in orange, binds to Arg270, Thr272 and Arg 153. Two conformations of Arg153 are displayed, named R153_1 and R153_2. Figure generated in PyMOL.

Disorder seen for Arg153 in the active site of crystal structure 5TRQ suggested that this side chain may exist in more than one conformation when succinate is bound (Figure 4.16.), therefore two conformations were modelled for this residue to satisfy the electron density. The results of an overlay with the WelO5-substrate structure 5IQT, imply that the new Arg153 conformation will clash with (+)-12-*epi*-fischerindole U isonitrile (Figure 4.17.), indicating a possible role for Arg153 in release of the product from active site.

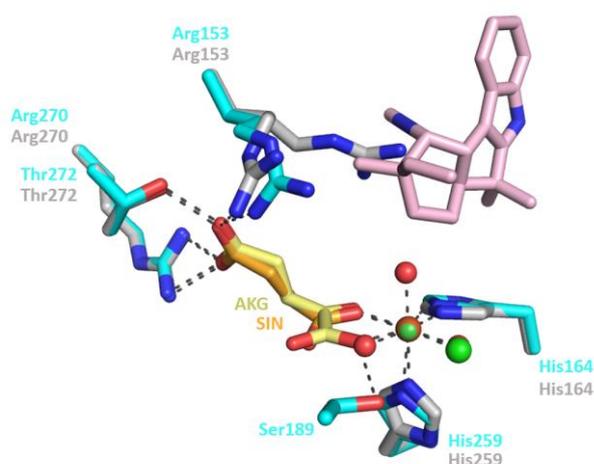


Figure 4.17. 5IQT structure of WelO5 (cyan), containing 2OG (yellow) and (+)-12-*epi*-fischerindole U isonitrile (pink) overlaid with 5T22 (grey), with succinate (orange) bound. Figure generated in PyMOL.

All available WelO5 crystal structures show that the substrate binding site of WelO5 is mainly composed of non-polar amino acid side chains, consistent with the lipophilic nature of the endogenous substrate. From overlays of our crystal structures with the substrate-bound structures in the literature, we can see that apart from the key active site residues discussed previously, residues Phe77, Ala82, Ile84, Val90, Ile161, Ala162 and Phe276 all have side chains within 4 Å of the substrate, forming a hydrophobic pocket. In addition, the Phe77 appears to form an edge-to-face π -interaction with the substrate indole. From the Liu group's 'closed' form WelO5 structures, the side chains of residues Val81, Ala88, Phe169, Met221 and Met225 are also seen to lay within

4 Å of the substrate (Met221 and Met225 are part of the β IV-V insert). The Liu group proposed two key protein-substrate interactions using their substrate-bound ‘closed’ form structures; the first occurring between the substrate isonitrile and the Ala82 backbone nitrogen (3.5 Å), the second between the substrate indole nitrogen and the sulphur of the Met225 side chain (3.7 Å).¹⁶⁴

Assuming the chlorination mechanism of other 2OG-halogenases is operational in WelO5, examination of the active site shows that the Fe²⁺ co-ordinated chlorine atom sits towards the back of the pocket and not within a suitable distance for reaction with the substrate (Figure 4.18.).

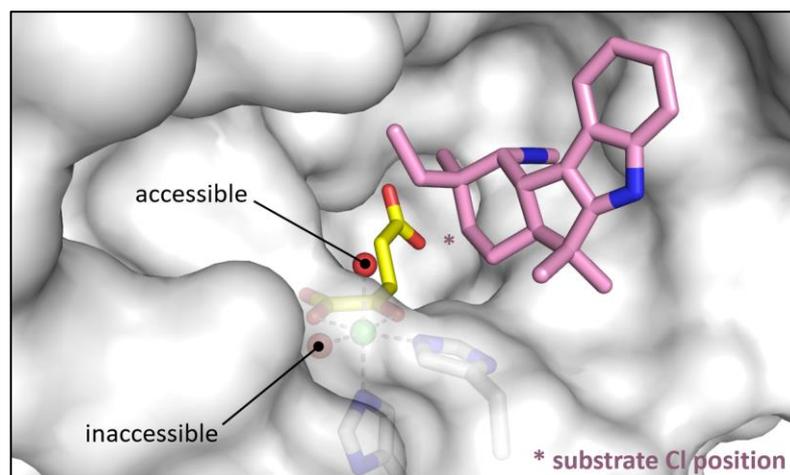


Figure 4.18. Analysis of the active site of substrate-bound WelO5 (PDB ID: 5IQT) reveals only one position accessible for reaction with (+)-12-epi-fischerindole U isonitrile (pink). Chlorination occurs at the substrate in the position highlighted by a star. The WelO5 protein surface is shown in grey, with 2OG (yellow) and the active site histidines, His164 and His259, shown in stick form. The accessible and inaccessible metal co-ordination positions are mapped onto the water molecules present in the crystal structure at these sites, represented by red spheres. Figure generated in PyMOL.

It has been proposed that isomerisation occurs at the iron, moving the chlorine into the axial co-ordination site (opposite His259; occupied by W1 in Figure 4.15.), mediated by hydrogen bonding to Ser189.¹⁶⁴ As the substrate is unable to move into a more suitable position within the active site (without a substantial and improbable conformational rearrangement of the DSBH), and

hydrogen abstraction and chlorination occur at the same carbon, isomerisation likely has to occur between the abstraction step and the chlorination step (i.e. both reaction steps occur from the axial position).

Studies on 2OG-halogenase SyrB2 have suggested that reaction outcome (i.e. hydroxylation or chlorination) is favoured by close proximity of the substrate towards the chloride rather than the hydroxyl.⁸¹ However, in WelO5 it appears that the substrate can position itself proximal to only one of these positions (the axial co-ordination site) and yet some, albeit minimal, hydroxylation activity is still observed. Isomerisation around the metal in WelO5 must have a part to play in determining whether chlorination or hydroxylation is the reaction outcome. Indeed, the S189A variant was found to relax reaction specificity for chlorination, indicating that interactions with the hydroxyl side chain of this residue are key for promoting the reaction intermediate responsible for chlorination.¹⁶⁴ The analogous substitution in CurA Hal (S132A) was found to abolish chlorination activity (however, enzyme activity was also abolished as a whole for this variant).⁸⁹

4.3.1.5.3. Literature structure comparison

The WelO5 structure 5J4R was aligned by the active site histidine residues with crystal structures of known 2OG-halogenases SyrB2, CytC3 and Cur Hal. As predicted from the difficulties in aligning their sequences (Chapter 2, Section 2.2.2.), the structure of WelO5 shared very little overall similarity with these 2OG-halogenases. However, some similarities were seen at the active site level. SadA was also aligned with WelO5 to see if the structural similarities which led to its selection as a template for WelO5 structure prediction extended to the active site (Figure 4.19.).

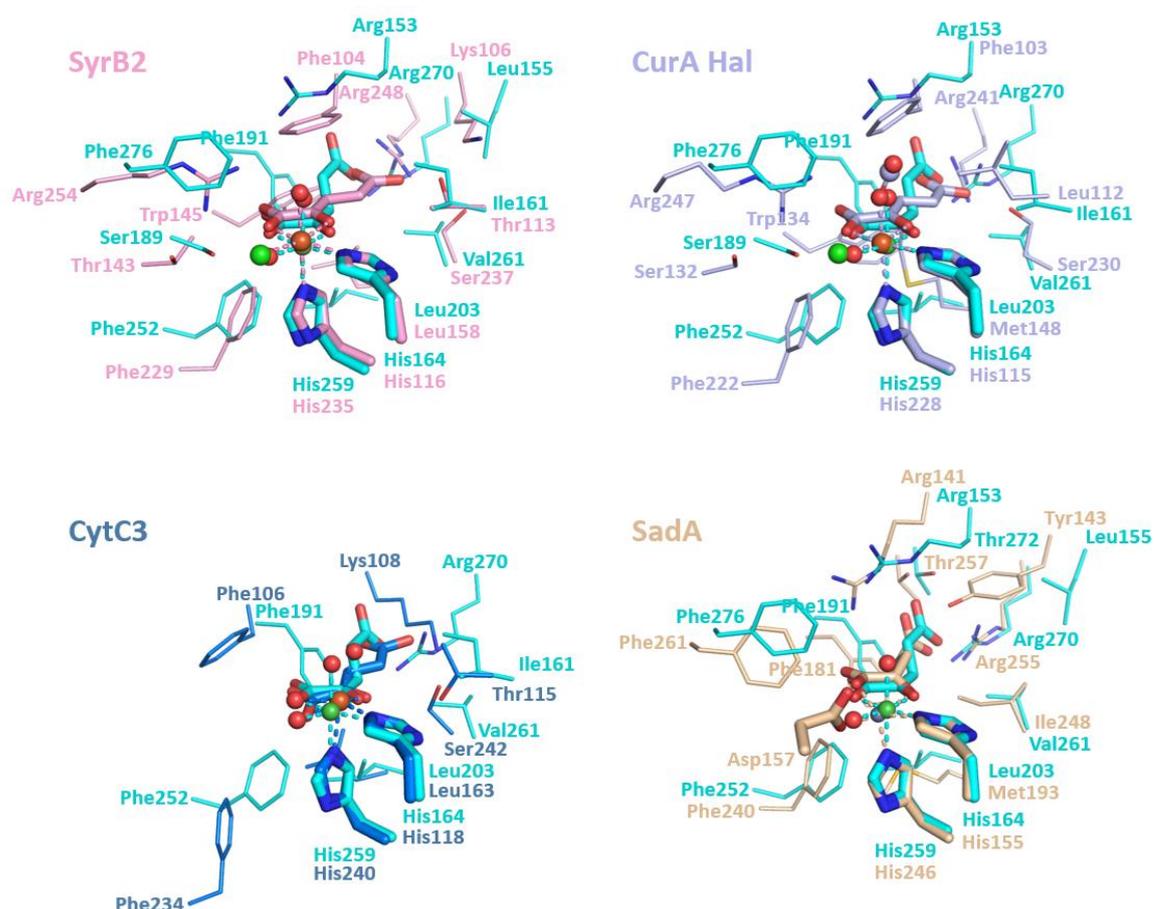


Figure 4.19. Structural alignments of the active site of WelO5 (PDB ID: 5J4R) in cyan with other 2OG-halogenases CurA Hal (lilac), CytC3 (blue), SyrB2 (pink) and 2OG-halogenase SadA (wheat). Structures were aligned by the active side histidine residues. Active site residues which overlaid with one another are highlighted for comparative purposes. Figure generated in PyMOL.

In SadA, Asp157, His155 and His246 form the characteristic 2OG-oxygenase facial triad of metal binding residues.⁶⁶ For the 2OG-halogenases, a vacant space created by the existence of glycine (CurA Hal, WelO5) or alanine (CytC3, SyrB2) in the equivalent position allows the binding of a chlorine atom. Indeed, SadA has been successfully converted to a halogenase by substitution D157G.¹⁸⁰ The salt bridge interaction from the C-5 of 2OG to an arginine residue (corresponding to Arg270 in WelO5) is conserved across all four enzymes; however, the conformation of 2OG differs in SadA and WelO5 compared to the other three enzymes. In these halogenases, a phenyl ring (Phe104 in SyrB2 and Phe103 in CurA Hal) sits above a planar 2OG molecule. Substitution of this phenylalanine side chain for an arginine in WelO5 and SadA (Arg153 and Arg141, respectively), allows rotation about the 2OG C-3 to C-4 bond, with this conformation now arranged in parallel with a different phenylalanine side chain (Phe191 in WelO5 and Phe181 in SadA). A comparison of the two different 2OG conformations, exemplified by SyrB2 and WelO5, is shown in Figure 4.20.

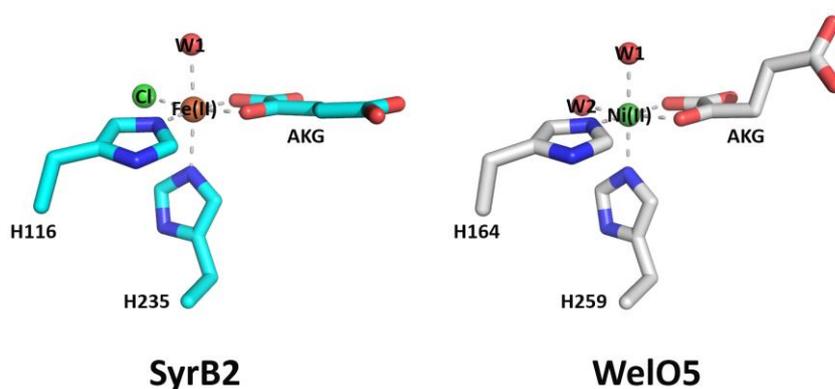


Figure 4.20. A structurally aligned view of the active sites of SyrB2 and WelO5 reveals two different conformations for 2OG (PDB ID: AKG). Figure generated in PyMOL.

Like WelO5, the C-terminus of SadA also folds 'back' towards the active site, forming interactions corresponding to those seen for the WelO5 terminal Ser290 (Figure 4.21.).

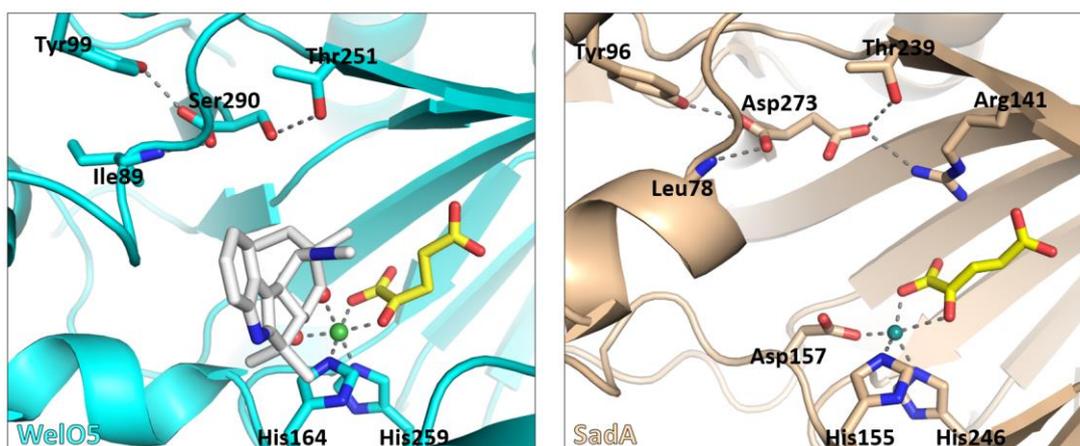


Figure 4.21. Active site view of WelO5 (cyan) and SadA (wheat) illustrating the proximity of the C-terminus to the catalytic centre and the interactions formed by C-terminal residues Ser290 (WelO5) and Asp273 (SadA). The WelO5 substrate is shown in grey, 2OG in yellow, the metal centres represented by green spheres and the metal co-ordinating waters represented by red spheres. Figure generated in PyMOL.

In the case of SadA, the C-terminal residue is an aspartate (Asp273) which is positioned to form interactions with the side chain of Arg141 (3.0 Å) and Thr139 (2.8 Å; *cf.* WelO5 Ser290-Thr151). The C-terminal carboxylate of Asp273 also interacts with the Leu78 backbone nitrogen (2.8 Å; *cf.* WelO5 Ser290-Ile89) and the Tyr96 hydroxyl (2.7 Å; *cf.* Ser290-Tyr99). This is not the case for CytC3, SyrB2 or CurA Hal, all of which display the C-terminus on the outside of the protein, exposed to solvent (Figure 4.22.).

Significantly, all three halogenases have a hydroxyl-containing side chain in the position equivalent to Ser189 in WelO5, consistent with the relationship proposed between S189 and the halogenation activity for WelO5.

None of the available structures of SyrBr, CurA Hal, CytC3 or SadA contain bound substrate, which makes the comparison of substrate-interacting residues a significant challenge.

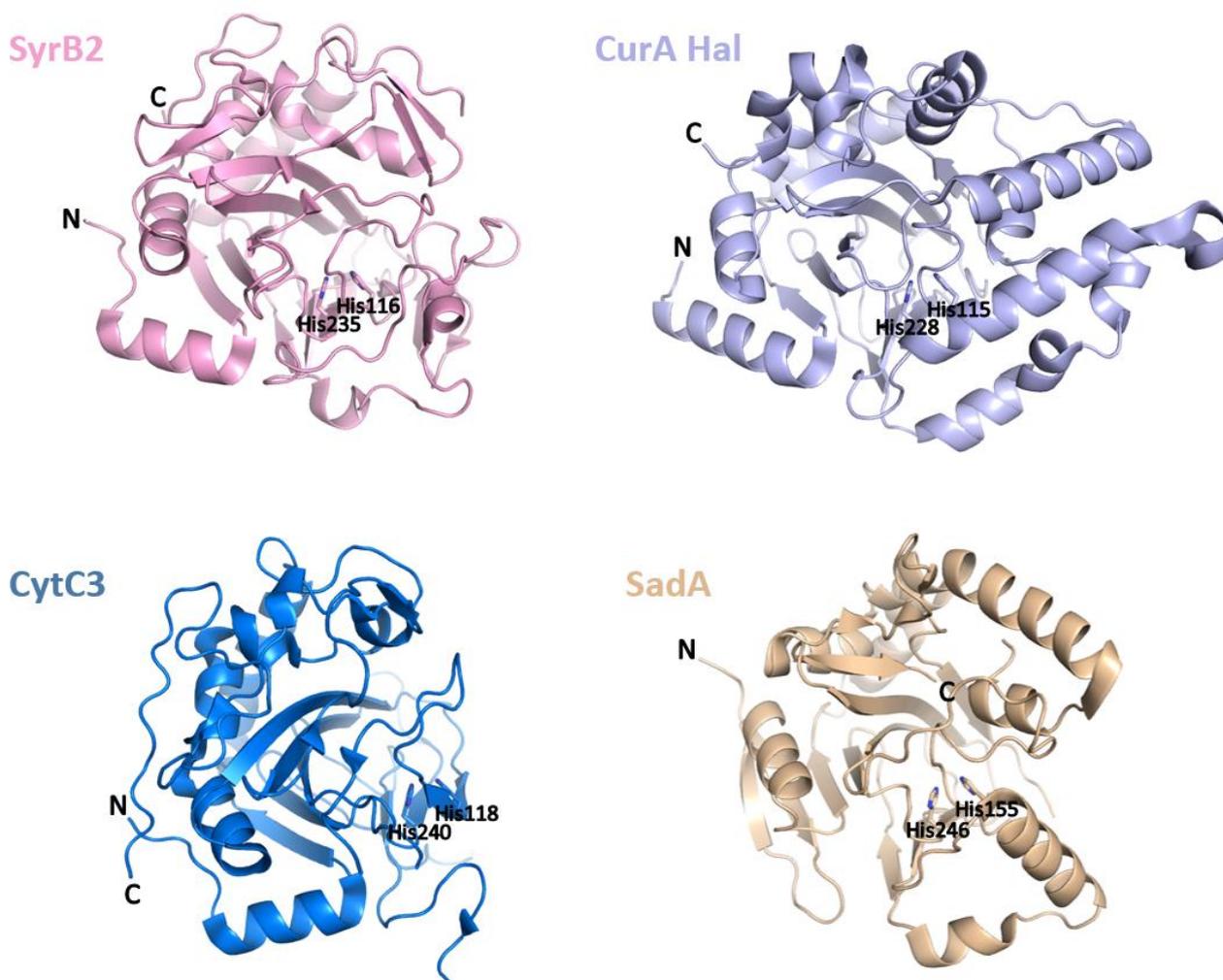


Figure 4.22. Overall structures of enzymes SyrB2, CurA Hal and CytC3 reveal solvent exposed N- and C-termini, in contrast to the structures of enzymes WelO5 and SadA which reveal an internalised C-terminus. Figure generated in PyMOL.

After the ('open' form) WelO5 structure 5J4R was determined, it was used to develop homology models (Sebastian Kelm, UCB) for the putative halogenases identified from the PHI-BLAST search for similar enzymes in Chapter 2 (Section 2.2.1.1.). These structures were aligned and compared with WelO5 to identify active site residues for variation (Figure 4.23.).

Overlay of the ten putative halogenase models showed variability in the two loop regions previously suggested to be involved in enzyme selectivity and substrate binding, corresponding to Glu75-Ala82 and Trp210-Glu238 (the β IV-V insert) in WelO5.

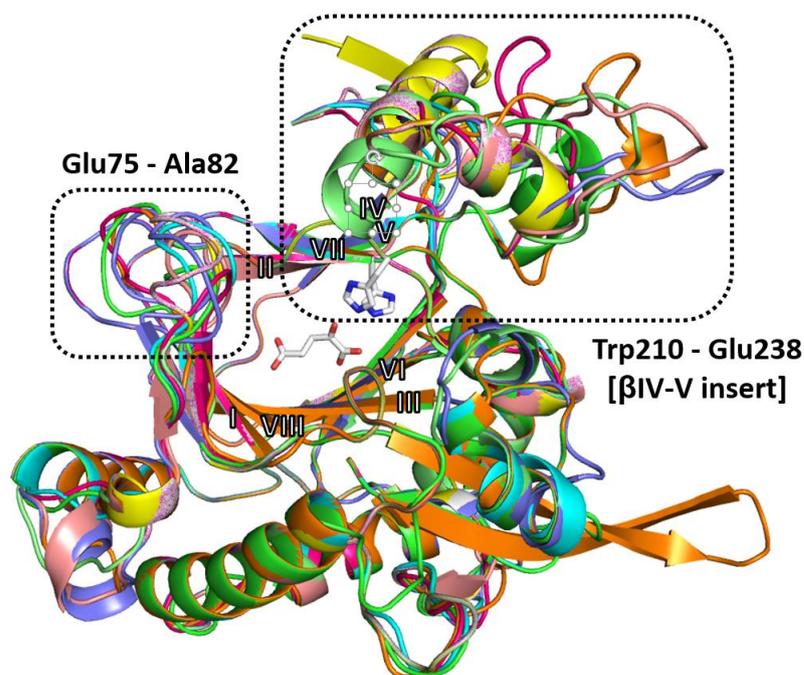


Figure 4.23. Overlays of homology models created for the putative halogenase enzymes identified in Chapter 2 with WelO5 (5J4R) in white. Two key regions of variability are demonstrated between these enzymes, suggesting a relationship between these regions and enzyme selectivity. Figure generated in PyMOL.

Assuming these halogenases have different substrates, it was surprising that most of the active site residues were conserved across this subset of enzymes. Variations were only found to occur at a small number of positions (Figure 4.24.), all of which were proximal to the superimposed WelO5 substrate. Residues surrounding the catalytic metal centre (i.e. His164, His 259 and 2OG) were conserved across all of these putative halogenase enzymes. However, these homology models were developed from the ‘open’ form of WelO5. More information could be gained by developing models from the ‘closed’ form and comparing the residues proximal to the active site from the more variable

β IV-V insert region.

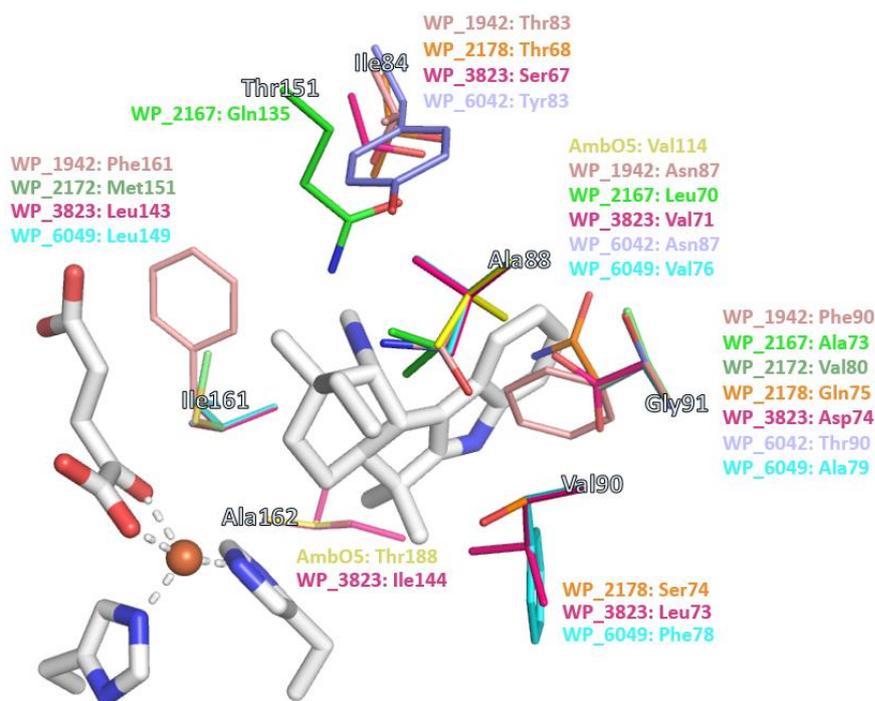


Figure 4.24. Active site view of ‘open’ form WelO5 (5J4R) in white, the structure used for homology modelling. (+)-12-epi-fischerindole U isonitrile is superimposed from WelO5 structure 5IQT. The positions demonstrating amino acid variability across the series and their corresponding enzyme identities are highlighted in colour, with the WelO5 residue identity labelled in white. Figure generated in PyMOL.

Structural overlays of WelO5 with other enzymes can provide a starting point for enzyme evolution *via* active site residue substitution. However, for these homologous (putative) halogenases it is likely that the substrates they act upon are very similar in structure. By introducing new substitutions, different in both identity and location to those in Figure 4.24., more radical changes in substrate scope may be achievable.

4.3.2. Preliminary NMR studies

X-ray crystallography is a powerful tool for the determination of three-dimensional protein structures. However, crystal structures only provide a static picture of a given protein. Flexible or dynamic regions can appear fixed in specific conformations, which may not be representative of the protein

structure in solution. NMR, a solution-based technique, can complement the knowledge extracted from protein crystal structures by examining the behaviour of the molecules in solution.¹⁸¹ Protein NMR experiments can provide insights into dynamics and flexibility, conformation distributions and protein-ligand interactions.^{182,183,184}

Given the substantial movement seen for the WelO5 β IV-V insert (residues 210-238) between crystal structures,¹⁶⁴ solution NMR was an obvious choice for the investigation of this conformational change and the examination of its relationship with substrate binding. However, limitations exist on the size of the protein which can be examined by NMR. For proteins 30-35 kDa and above, spectra start to become significantly more complicated due to greater spectral overlap (the result of an increasing number of peaks, reflecting the increasing number of atoms), peak broadening and decreased signal to noise.¹⁸¹ WelO5 has a molecular weight of 32.5 kDa, approaching the higher end of the range possible for solution NMR experiments. ¹⁵N-WelO5 was expressed and purified to assess whether WelO5 was amenable to NMR spectroscopic investigation. An initial ¹⁵N-WelO5 TROSY¹⁸⁵ experiment (Christine Prosser, UCB) implied some protein folding, but a significant number of peaks were missing considering the size of the protein. Approximately 150 peaks were present, falling far short of the number expected for WelO5 (290 amino acids; ideally expect one peak per amino acid). The addition of Ni²⁺ (already known to bind and stabilise WelO5; thermal shift and crystallisation studies) and 2OG increased the number of peaks observed to over 200 (with the exact number indeterminable due to some peak overlap), presumably by creating order within the active site and DSBH, stabilising the tertiary structure. The 2D TROSY spectrum of ¹⁵N-WelO5 in the presence of Ni²⁺ and 2OG showed (mostly) well-resolved peaks

with good chemical shift dispersion, justifying the expression of treble-labelled WelO5 for structural assignment (N.B. treble-labelled WelO5 was expressed but further experiments were not performed).

The ^{15}N -WelO5 protein was used for titration experiments with (+)-12-*epi*-fischerindole U isonitrile, **1** to evaluate the impact of substrate binding upon the structure of WelO5 (Christine Prosser, UCB). Five samples were prepared with increasing concentration of **1** and a constant enzyme concentration (300 μM), Ni^{2+} (500 μM) and 2OG (400 μM) concentrations. The substrate:enzyme concentration ratios used for this experiment were 0:1 (substrate free), 1:10 (30 μM substrate), 1:2 (150 μM substrate), 1:1 (300 μM substrate) and excess: 1 (500 μM substrate). Nominal changes in the ^{15}N - ^1H TROSY spectrum were seen within the 0-30 μM substrate range but noticeable spectral changes appeared at 150 μM substrate concentration (with minimal changes in chemical shift seen beyond 300 μM substrate). Figure 4.25. provides a side-by-side comparison of the 2D TROSY spectrum of the substrate-free ^{15}N -WelO5 sample (A) with the spectrum collected for the 1:1 substrate-enzyme ratio sample (300 μM substrate) (B).

This initial NMR work has demonstrated that it is possible to isolate pure isotopically-labelled WelO5 in reasonable yield, that the 2D TROSY ^{15}N -WelO5 spectra were reasonably well dispersed and that residue movement was detectable by NMR upon the addition of WelO5 natural substrate (+)-12-*epi*-fischerindole U isonitrile, **1**. These initial studies justify further work to fully assign the structure of WelO5 by NMR.

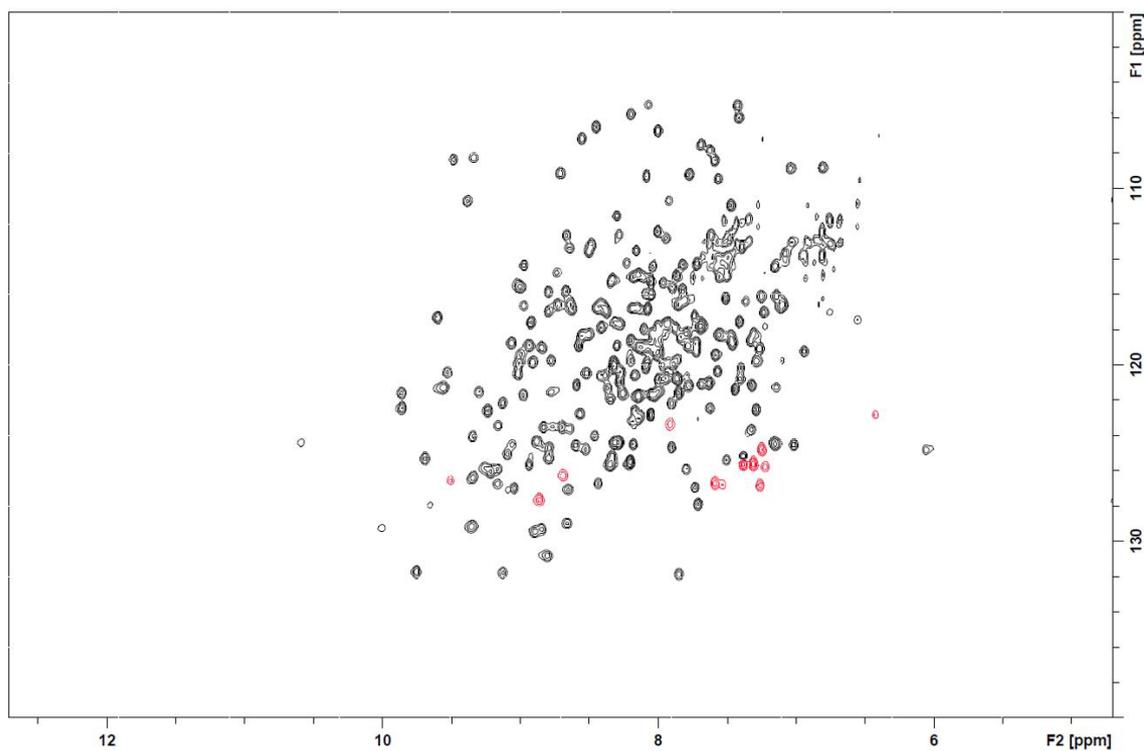
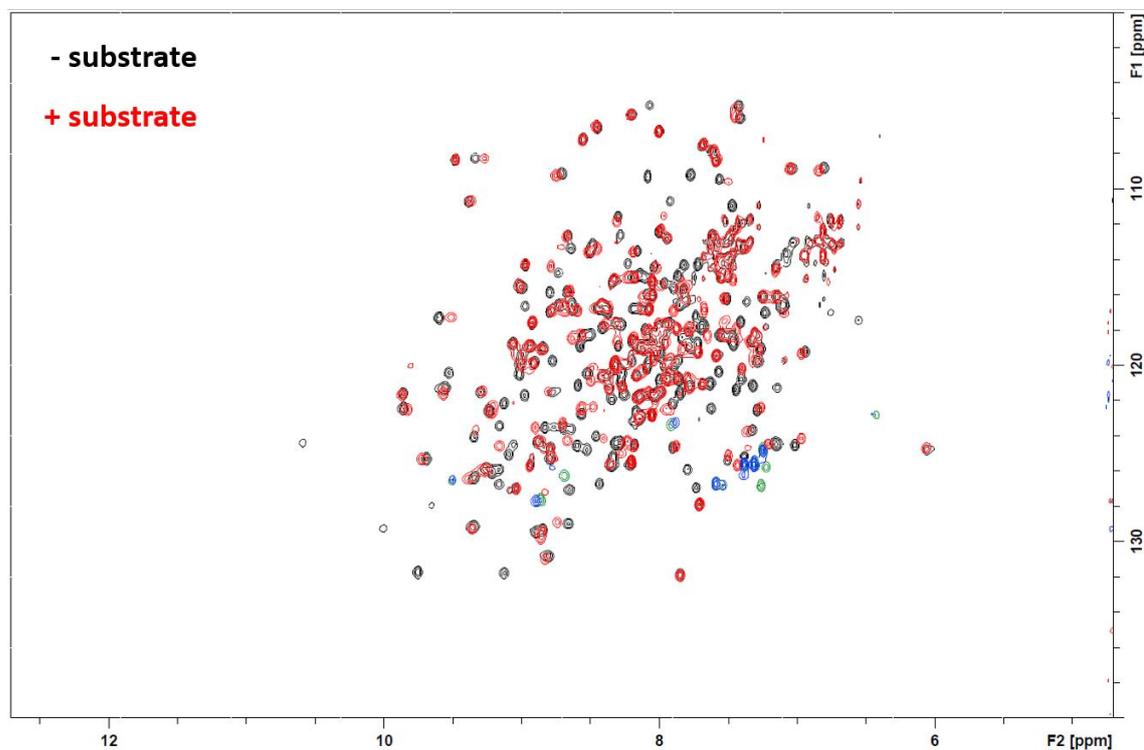
A. ^{15}N - ^1H TROSY: ^{15}N WelO5 + Ni^{2+} + 2OGB. ^{15}N - ^1H TROSY: ^{15}N WelO5 + Ni^{2+} + 2OG with (+)-12*epi*-fischerindole U isonitrile

Figure 4.25. (A) ^{15}N - ^1H TROSY spectra of $300\ \mu\text{M}$ ^{15}N -WelO5 with $500\ \mu\text{M}$ Ni^{2+} and $400\ \mu\text{M}$ 2OG in the absence of substrate, with negative peaks coloured red. (B) Overlay of (A) (with negative peaks in green); overlaid with a spectrum of ^{15}N -WelO5 in the presence of $300\ \mu\text{M}$ (+)-12-*epi*-fischerindole U isonitrile (negative peaks in blue). 600 MHz spectra were collected at 37°C over 90 min.

4.4. Summary of WelO5 characterisation studies

WelO5's ability to chlorinate an isolated small molecule had not been seen before in other members of the 2OG-halogenase family, justifying the characterisation of WelO5 and investigation into the structure-function relationship of this novel enzyme. This Chapter has described preliminary characterisation of WelO5, following on from the initial 2014 WelO5 publication.⁵⁰

Initial solution-based characterisation of WelO5 has been described, specifically molecular weight by mass spectrometry, assessment of aggregation and oligomerisation by SEC-MALS and secondary structural composition by CD. Anaerobic UV-Vis spectroscopy experiments were also able to demonstrate the characteristic 2OG-oxygenase MLCT transition upon binding of Fe²⁺ to 2OG.

Multiple crystallisation conditions have been identified for the growth of diffraction quality crystals, resulting in the determination of three novel WelO5 crystal structures; 5J4R (WelO5 + Ni²⁺ + 2OG, 1.65 Å), 5T22 (WelO5 + Ni²⁺ + 2OG, 1.75 Å) and 5TRQ (WelO5 + Ni²⁺ + succinate, 1.30 Å). Unfortunately, none of these structures contained a ligand in the active site. However, three-dimensional visualisation of the WelO5 active site was still extremely useful in guiding the rational design of active site variants. Four WelO5 crystal structures (PDB ID: 5IQS, 5IQT, 5IQU & 5IQV) were released into the PDB by the Liu group in mid-2016 which confirmed the substantial conformational change of the β IV-V insert region hinted at in our 5J4R and 5T22 structures. The WelO5 natural substrate was present in several of these 2016 crystal structures, confirming its orientation within the active site. The crystal structures described in this Chapter, alongside those published in the

literature, will serve to guide the selection of active site residues for variation, with the aim of introducing new activities (Chapter 5). Initial NMR studies have justified further effort towards solving a NMR structure of WelO5. By studying WelO5 in solution, protein conformation changes and the dynamics involved in enzyme catalysis can be interrogated. The analysis of which might provide further evidence of the β IV-V insert movement seen in the presence and absence of WelO5 substrate.

Chapter 5. Enzyme evolution & results

5.1. Introduction

Changes in enzyme properties can be brought about by directed evolution or by rational design.^{3,9,96} Amino acid substitutions, deletions, or insertions that alter active site shape and introduce or remove substrate-interacting side chains have the capacity to introduce new activities by modifying enzyme selectivity and specificity.^{96,186,187} Three-dimensional structural information (e.g. from crystal structures and homology models) and computational simulations can help to narrow down the overall number of residues selected for substitution and predict which variants should be made.^{98,188} Enzyme activities have been successfully modified by rational design, despite the low level of sequence space explored and the improbability of identifying synergistic mutations.^{186,187,189} Rational design becomes a useful strategy when structural information is available but resources (e.g. time, cost, equipment) are limited and high-throughput screening or selection methods are unavailable.

Relative to specificity and selectivity, properties such as thermal stability or solvent tolerance can be significantly harder, if not impossible, to predict.^{190,2} Such characteristics are at present more likely to be optimised by directed evolution than rational design because static structural and/or functional information is often not necessary or predictive of success for these enzyme properties.⁸ Examples also exist where enzyme specificity and selectivity have been modified by directed evolution.¹⁹¹⁻¹⁹³ Directed evolution accelerates the natural evolutionary process through iterative cycles of mutagenesis and selection to improve a given enzyme property or function. Mutations can be introduced randomly (error-prone PCR), by homologous recombination (e.g.

DNA shuffling) or by site saturation mutagenesis.¹³ Enzyme variants with new or improved properties are identified through screening or selection processes from the pool of generated proteins and selected for the next round of evolution.

For every amino acid selected for substitution, saturation mutagenesis requires the production of 19 variants (there are 20 common proteogenic amino acids), resulting in large libraries. The mutagenesis effort required to generate such libraries can be simplified using degenerate codons. In the genetic code there are more codons ($4^3=64$) than there are encodable amino acids (20, plus three stop codons), resulting in an inbuilt degeneracy - multiple codon combinations encode for the same amino acid (e.g. ATT, ATC and ATA all code for isoleucine). By using degenerate codons, libraries for directed evolution can be created with reduced mutagenesis effort.¹⁹⁴ For example, the degenerate codons NNK (where N = C/G/A/T and K = G/T) and NNS (where S = C/G) both encode for all 20 proteogenic amino acids using just 32 codons.¹⁹⁴ Mixtures of compressed codons can also be designed for defined amino acid libraries (Table 5.1.)¹⁹⁵

Degenerate codon	Number of codons	Encoded amino acids	Library Properties
NNN	64	All (20)	All proteogenic amino acids
NNK	32	All (20)	All proteogenic amino acids
NNS	32	All (20)	All proteogenic amino acids
NTN	16	M, F, L, I, V (5)	Non-polar side chains
NAN	16	Y, H, Q, N, K, D, E (7)	Charged, larger side chains
NDT	12	F, L, I, V, Y, H, N, D, C, R, S, G (12)	Mixture of properties (polarity, size and charge)
RST	4	A, G, S, T (4)	Small side chains

Table 5.1. Some examples of commonly used degenerate codons for directed evolution. Codons are named according to IUPAC terminology: N = C/G/A/T, K = G/T, S = C/G, R = A/G, D = A/G/T^{194,196}

Semi-rational evolutionary approaches also exist, where a specific number of enzyme residues are selected for randomisation based upon structural information.¹⁰ Examples of such approaches include ISM (Iterative Saturation Mutagenesis)^{197, 198} and CASTing.¹⁹⁹ For ISM, structural and/or functional information is used to identify specific amino acid positions suitable for variation. The first position is randomised by saturation mutagenesis and the best assay hit used as a template for randomisation at the second site.¹⁹⁸ The CASTing (Combinatorial Active-site Saturation Test) method was designed specifically for expanding the substrate scope of a given enzyme.^{199,200} Using structural information, sets of amino acids (between 2-10 residues) with side chains directed into the active site are selected for the generation of focused libraries.²⁰¹ When combined with ISM, more efficient optimisation can be achieved wherein the best hit from the first library is used as a template for a second library designed around a different set of residues. This process is then repeated until the desired property/function level has been achieved.²⁰² These methods were designed in order to increase the efficiency of directed evolution by identifying synergistic active site substitutions and minimising the number of enzyme variants required to be screened (hundreds vs. thousands) before one with suitable properties is identified.²⁰³ However, the numbers are still very high relative to those generated from rational design approaches, meaning that a high-throughput screening or selection method is critical for evaluation of these variant libraries.

When embarking upon a new biocatalysis development programme, several factors determine which evolutionary strategy is nominated. These include the enzyme property chosen for optimisation, the anticipated mutant library size, the availability of high-throughput screening/selection methods and access to

three-dimensional structural information. This Chapter describes the strategy selected for assessment of the biocatalytic potential of WelO5, alongside some preliminary mutagenesis results. The development of an optimised protocol for producing multiple enzyme variants in parallel is described. Combining this with an LC-MS based screen yielded a robust screening method with moderate throughput, used for the assessment of a small library of designed WelO5 variants. A strategy involving alanine scanning mutagenesis at the WelO5 active site was selected, leading to the discovery of I161A - an enzyme variant with unexpected activity. In addition, single point active site residue substitutions were designed and tested for activity against compounds structurally related to the WelO5 natural substrate, as well as for two structurally differentiated compounds.

5.2. Mutagenesis strategy

The initial screening performed in Chapter 3 demonstrated a very narrow substrate specificity range for WelO5. Activity was only seen for the substrates most structurally similar to the natural substrate, (+)-12-*epi*-fischerindole U isonitrile. However, amino acid substitutions within active site have the potential to introduce promiscuity (providing that they are tolerated by the enzyme).

Whilst structural and functional information was available for WelO5, the screening process represented a significant limitation for the evaluation of large libraries of WelO5 variants. The 2OG-dependent halogenase reaction mechanism does not lend itself towards a high-throughput assay (due to difficulty in incorporating a chromophore for colourimetric assays and an inability to introduce criteria for colony growth selection; the enzyme's oxygen

requirement and potential need for a glove box are also potential limitations). Significant consideration was given to assay selection and design because throughput would determine the mutagenesis strategy selected. Considering the WelO5 reaction mechanism and outcome (i.e. chlorination), an LC-MS based assay was chosen. In some cases, LC-MS based screens can deliver an advantage over more high-throughput colourimetric methods as they can have general applicability to multiple substrates (i.e. the presence of specific functional groups is not required and substrate pre-functionalisation or product derivatisation is not necessary). Products can be (tentatively) identified from the mass spectrum, with the presence of a chlorine isotope pattern particularly characteristic of a successful enzymatic transformation. Unexpected side-products can also be identified which are likely to be undetected or overlooked when higher-throughput colourimetric or fluorescent assays are used.

Due to the restrictions placed upon library size by LC-MS screening throughput, directed evolution methods were not perceived to be feasible. Instead, a strategy was envisaged involving the introduction of specific, rationally designed amino acid substitutions. Proximal active site residues are likely to be the most influential on enzyme specificity and selectivity.⁸ Therefore, it was anticipated that residue substitutions in the WelO5 active site, guided by crystal structures (see Chapter 4; PDB IDs 5J4R, 5T22, 5TRQ), could produce variants with activity against non-natural substrates. Unfortunately, whilst these crystal structures did model a histidine-chelated metal ion and 2OG (or succinate) in the active site, they did not contain any natural substrate or substrate-like compounds. Using existing knowledge of the stereochemistry of the chlorine introduced at C-13, the 2OG-halogenase

chlorination reaction mechanism and the approximate distance between the substrate and the catalytic Fe^{2+} species,^{63,204} the endogenous substrate was oriented into the WelO5 active site (Figure 5.1.). It was hoped this would be sufficient to design rational amino acid substitutions whilst efforts towards obtaining substrate-bound WelO5 structures were ongoing.

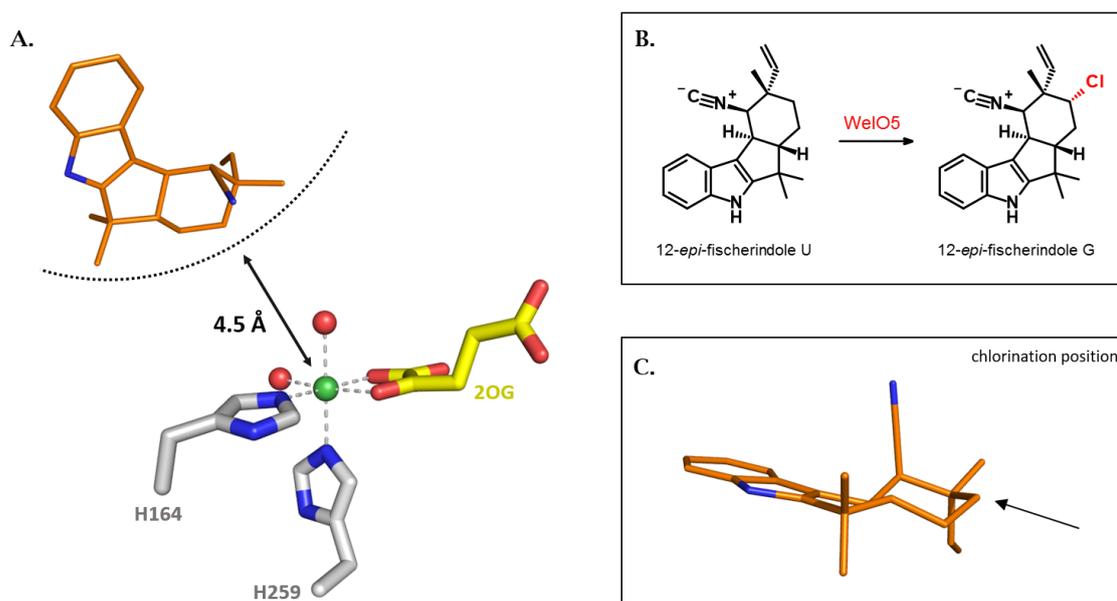


Figure 5.1. Illustration of the guidelines and constraints used for orienting WelO5 natural substrate, (+)-12-epi-fischerindole U isonitrile into the active site of WelO5 crystal structure 5J4R. Image A: C-13 of (+)-12-epi-fischerindole U isonitrile (in orange) was placed 4.5 Å from the metal (green sphere), in an orientation which would allow chlorination to occur from the opposite face of the cyclohexyl ring to the isonitrile (shown in 2D, image B. & in 3D, image C.), to give (+)-12-epi-fischerindole G. Figure generated in PyMOL.

Initial anaerobic co-crystallisation of purified WelO5 with (+)-12-epi-fischerindole U isonitrile yielded poorly diffracting crystals ($> 5\text{Å}$). Before further optimisation was performed, crystal structures of WelO5 with (+)-12-epi-fischerindole U isonitrile bound were published.¹⁶⁴ These structures confirmed the position of the substrate, which was found to sit within the active site in a very similar orientation to that predicted in Figure 5.1. These new substrate-bound crystal structures were used to help select active site residues for rational substitution.

The natural substrate for WelO5 is very different to that of other known 2OG-halogenase enzymes by virtue of its size, lipophilicity and lack of peptide-carrier protein (Figure 5.2).^{89,90,105.}

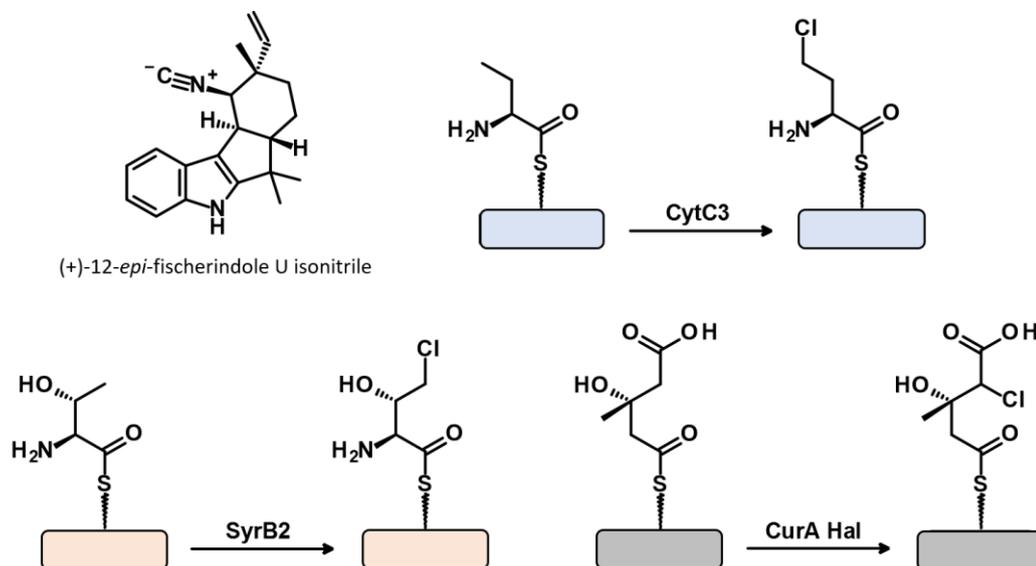


Figure 5.2. Structure of WelO5 natural substrate 12-epi-fischerindole U isonitrile compared with the substrates of other known 2OG-halogenase enzymes, CytC3, SyrB2 and CurA Hal. Rectangles represent carrier proteins.

Crystal structures of WelO5 reveal a correspondingly large, hydrophobic active site, without obvious interactions marking key residues for potential substitution. It was thought that the generation of WelO5 variants by performing saturation mutagenesis at a small number of active site residues could be informative. However, the selection of all active site residues within 5Å of the natural substrate in both 'open' and 'closed' WelO5 structures, with exclusion of those with side chains pointing away from the active site, identified over 20 residues. Given the LC-MS screening method envisaged and that saturation mutagenesis would generate 19 substitutions per selected residue, this number was considered too high. To narrow down the total number of residues for substitution it was decided that an alanine scan (i.e. systematic alanine substitutions) would be performed on the active site residues selected. The effect of each alanine substitution upon WelO5 (+)-12-

epi-fischerindole U isonitrile chlorination activity could be assessed, with noticeable changes (both increases and decreases) in activity marking the residues for further investigation. Residues demonstrating very little change could then be removed from the pool.

5.3. Screening protocol

To assess the impact of alanine substitutions upon WelO5 wild-type activity and against alternative substrates, the use of a broadly applicable LC-MS method was envisaged. Minimal changes could then be made to the LC gradient to accommodate different substrates without significant alterations being required to the overall screening protocol. New UV-active peaks in the LC chromatogram would flag up the production of new compounds, with percentage conversion calculatable from relative peak area. Tentative product identity could then be assigned based on the mass spectrum.

5.3.1. Screening protocol requirements

The following areas were evaluated when designing the screening protocol:

- *The production of variants.* Whilst the total number of variants had already been reduced by selecting only active site residues for alanine substitution, the ability to mutate and express several constructs still needed to be taken into consideration. It was envisaged that learnings from the work performed at the OPPF (Chapter 2) could be applied to perform mutagenesis, expression and purification in parallel, in a plate-based format.
- *Whole cell vs. isolated extract vs. purified enzyme.* It was proposed that the use of whole cells or crude cell extracts would facilitate the screening of higher numbers due to reduced sample manipulation

requirements. However, initial tests showed that cleaner protein samples would simplify the LC-MS analysis (improved MS sensitivity and new, small peaks in the UV trace easier to identify).

- *Enzyme purification.* It was proposed that the expression of all variants with a His₆-tag would enable parallel Ni²⁺-NTA purifications. A tip-based or plate-based purification method was considered superior to a resin-based method to reduce the number of pipetting steps and reduce sample-to-sample variability. With a guaranteed binding capacity per well/tip, low concentrations of isolated protein (relative to wild-type) would be indicative of poor expression and/or protein solubility issues (undesirable properties for variant progression). The isolation of purified enzyme could also be used for further analysis (e.g. mass spectrometry, CD, thermal shift) if required. Initial tests were needed to compare tip-based and plate-based methods and determine whether either method could yield sufficient quantities of enzyme to perform the required screening assay.
- *Analysis.* Consideration of sample processing was necessary to ensure robustness in analysis, reduce system failures (e.g. column/system blockages) and to guarantee visible peaks with detectable ionisation in the UPLC-MS traces. The finalised LC method would need to balance resolution between starting material and the chlorinated product, without excessively long run times.

5.3.2. Method development & optimisation

Assay development and optimisation was performed with wild-type WelO5 prior to testing WelO5 variants.

5.3.2.1. Expression studies

pNIC28-Bsa4[WelO5] was transformed into BL21(DE3) cells (NEB) to evaluate modifications to the conditions previously used for WelO5 expression (Chapter 2, Section 2.3.2.) for use in a plate-based format. Colonies were picked into 1 mL 2xTY media (50 µg/mL kanamycin) in 96 deep-well blocks, covered with a gas-permeable seal and incubated overnight at 37 ° C (220 rpm). Overnight cultures were diluted (20 µL) into 24 deep-well blocks containing 2 mL media per well.

For IPTG induction, 2xTY medium was used. Blocks were covered with a gas-permeable seal and incubated at 37 ° C (220 rpm) until $OD_{600} \sim 0.5$ was reached. The culture blocks were cooled to their respective induction temperatures (20 ° C or 30 ° C) for 15 min and then induced by the addition of IPTG (to a final concentration of 0.1, 0.5 or 1.0 mM). The culture blocks were incubated overnight at their respective induction temperatures and harvested at the 20 hr post-induction timepoint (as for IPTG-induced WelO5 expression; Chapter 2, Section 2.3.2.). Blocks were stored at -80 ° C.

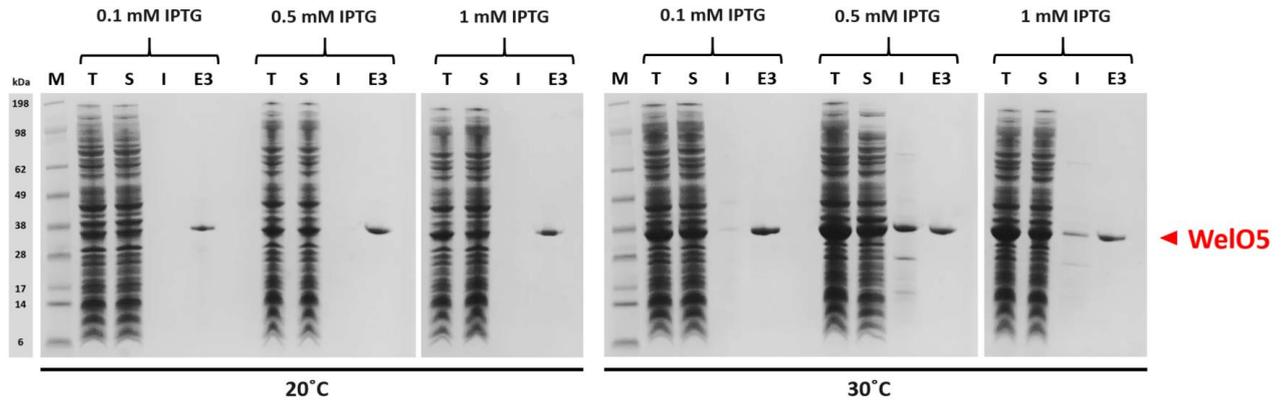
For auto-induction, 2xTY Overnight Express™ Autoinduction System 1 (Novagen®) medium was used. Cells growing in auto-induction media are spontaneously induced when they reach a certain density, meaning that even though cells containing the plasmids of different WelO5 variants may grow at different rates, induction will occur at approximately the same cell density for each culture. Selection with kanamycin is much less effective in auto-induction media so 100 µg/mL kanamycin media was used (as recommended by the manufacturer).²⁰⁵ Auto-induction media may require a longer incubation time to reach stationary phase when using temperatures lower than 37 ° C ²⁰⁵ so the

blocks were first incubated at 37 °C (220 rpm) for 2.5 hr and then cooled to their respective induction temperatures (20 °C or 30 °C). Representative samples were taken after the 2.5 hr at 37 °C and analysed by SDS-PAGE to confirm induction had not yet started. Cultures were harvested at 16, 20 and 24 hr timepoints (OD₆₀₀ measurements were collected from each timepoint to determine if/when stationary phase had been reached). Very little differences in OD₆₀₀ measurements were observed between the 20 and 24 hr timepoints indicating that stationary phase had been reached by 20 hr. After harvesting, the blocks were stored at -80 °C.

All culture blocks were defrosted and the cell pellets purified for comparison of protein yields by SDS-PAGE and Nano-drop. Proteins were isolated in parallel by miniaturised IMAC pipette-tip based purification (Phy-Tips) on a PhyNexus automated purification system, with a three-step elution (3 x 30 µL). Being the most dilute (and therefore showing the greatest cross-sample variability), only the third elution steps (E3) were analysed by SDS-PAGE (Figure 5.3.).

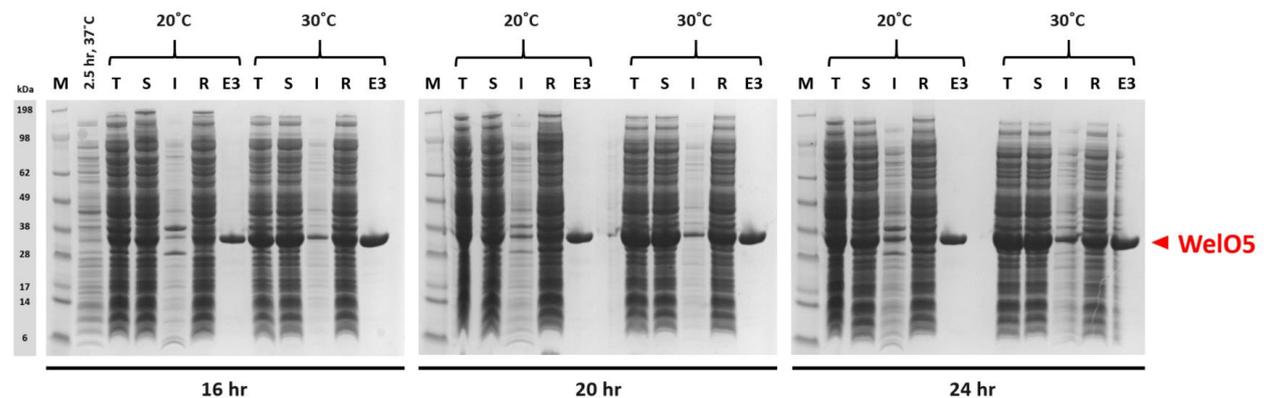
Comparison of E3 for each set of conditions showed that greater protein yields were achievable with auto-induction media. As stationary phase in auto-induction media had been established by 20 hr; expression at 30 °C in auto-induction media over 20 hr were the conditions selected for the production of WelO5 variants.

A) IPTG Induction:



Temperature	[Protein] A280 mg/mL					
	20 ° C			30 ° C		
[IPTG], mM	0.1	0.5	1	0.1	0.5	1
Elution 1	0.8	0.9	0.9	1.1	1.4	1.1
Elution 2	0.5	0.6	0.6	0.8	1.0	0.9
Elution 3	0.3	0.4	0.3	0.5	0.5	0.5

B) Auto-induction:



Harvest timepoint	[Protein] A280 mg/mL					
	16 hr		20 hr		24 hr	
Temperature	20 ° C	30 ° C	20 ° C	30 ° C	20 ° C	30 ° C
Elution 1	2.2	3.6	2.7	4.0	2.9	3.9
Elution 2	2.2	2.8	2.2	2.9	1.9	3.0
Elution 3	1.2	1.6	1.3	1.6	1.3	1.6

Figure 5.3. SDS-PAGE gels and protein concentration measurements (by Nano-drop, A280) for IPTG (A) and auto-induction media (B) plate-based expression trials. His₆-WelO5: 35 kDa. The protein molecular weight marker is labelled M. Total samples are labelled T, soluble fractions are labelled S and insoluble fractions are labelled I. A sample of auto-induction media taken after 2.5 hr at 37 ° C is also included to confirm induction had not already started before the blocks were cooled. Label R describes samples of lysate after Phy-Tip loading, with the presence of excess WelO5 indicating saturation of the binding capacity of the Ni²⁺ resin tips. Wells containing samples from the third elution step are labelled E3.

5.3.2.2. Protein purification

As previously mentioned, miniaturised parallel protein purifications were initially performed using the PhyNexus automated purification system. Using this setup, up to twelve samples can be purified at any one time. Cell pellets produced from 2mL cultures in 24 deep-well blocks were defrosted and lysed by treatment with 500 μ L BugBuster[®] master mix (Novagen). After 30 min, the blocks were spun down and the soluble fractions loaded into a 96-well block for compatibility with the tip-spacing of the PhyNexus automated pipette. 1 mL PhyTips (containing 10 μ L IMAC resin) were equilibrated (5 mM imidazole buffer) and the soluble fractions loaded. The tips were washed (2 x 20 mM imidazole buffer) and then the protein eluted (3 x 30 μ L; 200 mM imidazole buffer).

PhyNexus equilibration buffer: 10 mM NaH₂PO₄, 300 mM NaCl, 5 mM imidazole, pH 7.4

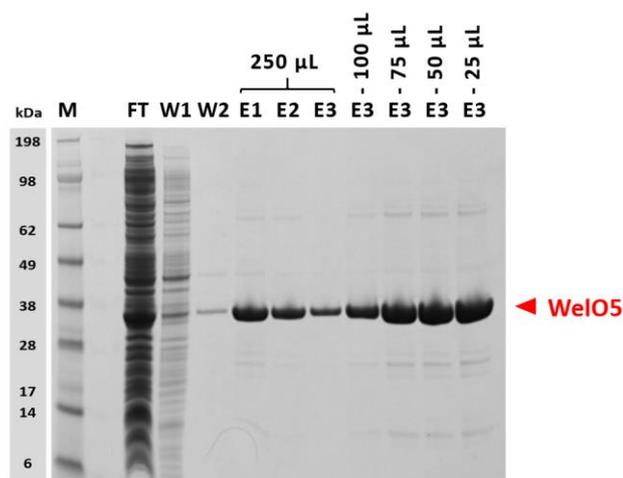
PhyNexus wash buffer: 10 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 7.4

PhyNexus elution buffer: 10 mM NaH₂PO₄, 300 mM NaCl, 200 mM imidazole, pH 7.4

Analysis by SDS-PAGE of His₆-WelO5 purified from cultures grown in auto-induction media (Figure 5.3.) revealed that the amount of protein expressed exceeded the capacity of the tips. However, the amount of purified protein was still sufficient for the required assay. PhyNexus purification was later abandoned in favour of HIS-Select filter plates, which can purify up to 96 His₆-tagged proteins at a time with a binding capacity of > 2mg per well (as determined with a 30 kDa histidine-tagged protein; His₆-WelO5 = 35 kDa).²⁰⁶

Samples were prepared for HIS-Select filter-plate purification in the same way as for PhyNexus purification except that the entire cell extract was loaded directly into the pre-equilibrated plate wells after lysis (no centrifugation required). Samples were purified according to the manufacturer's centrifuge

protocol,²⁰⁶ washing twice with 5 mM imidazole buffer and eluting in 250 mM imidazole buffer. Elution buffer volumes were varied (25–250 μ L) for comparison by SDS-PAGE and Nano-drop to the PhyNexus conditions (Figure 5.4.).



Protein; A280 mg/mL (μ g):

	HIS-Select					PhyNexus
Elution volume, μ L	250	100	75	50	25	30
Elution 1	0.5 (125)	0.2 (20)	0.0 (0)	0.0 (0)	0.0 (0)	4.0 (120)
Elution 2	0.3 (75)	0.9 (90)	1.0 (75)	0.9 (45)	0.9 (23)	2.9 (87)
Elution 3	0.1 (25)	0.5 (50)	0.9 (68)	1.1 (60)	1.2 (30)	1.6 (48)

	HIS-Select					PhyNexus
Combined volume, μ L	750	300	225	150	75	90
Total protein yield, μ g	225	160	143	105	53	255

Figure 5.4. SDS-PAGE gel and protein concentration measurements (by Nano-drop) for HIS-Select filter plate purification of His₆-WelO5 expressed in auto-induction media at 30 °C, 20hr. His₆-WelO5: 35 kDa. The protein molecular weight marker is labelled M. Flow-through from the loading of samples is labelled FT. Both wash steps are labelled W1 and W2. Protein was then eluted in 3 x 250 μ L volumes. Lower volumes were assumed to be too low for the plate and did not elute the greatest amount of protein until the third elution step. The third elution step (E3) for 100, 75, 50 and 25 μ L elution volumes was loaded onto the gel. Protein yields in μ g were calculated from Nano-drop measurements to compare with PhyNexus purification.

The HIS-Select filter plates were unable to elute very low volumes. However, the higher elution volumes were found to produce comparable quantities of protein relative to the PhyNexus, albeit in higher overall volumes (Figure 5.4.;

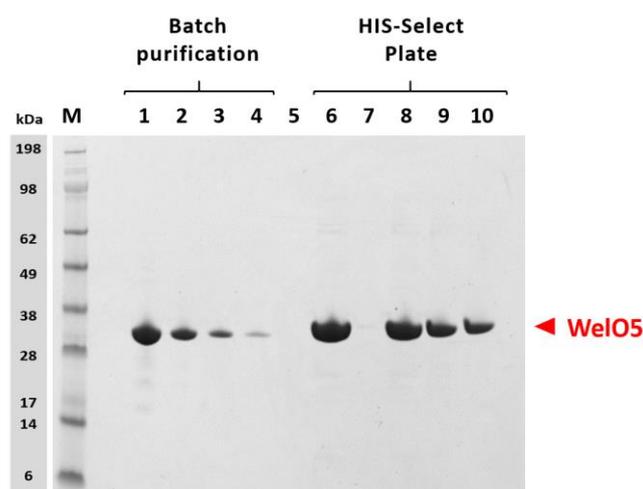
225 µg in 750 µL by HIS-Select vs. 255 µg in 90 µL for PhyNexus). A 250 µL elution volume was found to be optimal in terms of protein recovery (vs. volume) and could be used directly in the screening assay at the eluted concentration (250 µL E1 estimated at 0.5 mg/mL = 14 µM).

5.3.2.3. Assay set-up

Considering the use of protein directly eluted from HIS-Select plates in the screening assay, it needed to be established whether the presence of high concentrations of imidazole (250 mM) would have an inhibitory effect on WelO5 catalytic turnover. To investigate this, reactions were performed with batch-purified WelO5 (Chapter 2, Section 2.3.) under the previously used screening conditions (Chapter 3, Section 3.4.), with increasing concentrations of imidazole in the reaction buffer (0 mM, 50 mM, 100 mM, 200 mM and 300 mM final concentrations). From this, imidazole was confirmed to not interfere with wild-type WelO5 catalytic turnover of natural substrate (+)-12-*epi*-fischerindole U isonitrile (data not shown).

As imidazole was not found to affect WelO5 activity, the assay protocol used for initial screening with batch-purified WelO5 (Chapter 3, Section 3.4.) was modified to use protein eluted directly from the HIS-Select purification plate whilst maintaining the 100 µL assay reaction volume. SDS-PAGE (gel densitometry) was used to calibrate the HIS-Select plate elution volume and concentration with that of batch-purified WelO5 (Chapter 2, Section 2.3.). The initial activity screening performed in Chapter 3, Section 3.4. used purified WelO5 at a 25 µM concentration. Four reactions were prepared with batch-purified WelO5 as previously described (Chapter 3, Section 3.4.) but with varied enzyme concentration; WelO5 was used at 25 µM and at 1/3 (8 µM), 1/9 (4 µM)

and 1/27 (1 μ M) dilutions. Similarly, reactions were prepared with enzyme isolated by HIS-Select purifications performed using 85 μ L and 250 μ L elution volumes. Protein elution was not detected in E1 for the 85 μ L elution so the E2 fraction was assayed instead. Reactions were performed with 85 μ L HIS-Select E2, 250 μ L HIS-Select E1 and with 250 μ L HIS-Select E1 sample dilutions (1/2 and 1/3). Samples prepared with the same dilution factors were also analysed by SDS-PAGE and gel densitometry was used to quantify the amount of protein present. This was compared to the percentage catalytic turnover seen for samples prepared at the same protein concentration (Figure 5.5.).



ID	Description	Estimated concentration	Densitometric Volume	Turnover
1	[assay]	25 μ M	15107816	97%
2	1/3 of 1	8 μ M	7680669	86%
3	1/9 of 1	4 μ M	3535172	35%
4	1/27 of 1	1 μ M	1050710	9%
5	-	-	-	-
6	85 μ L E2	25 μ M	23021294	95%
7	-	-	-	-
8	250 μ L E1	14 μ M	18686448	92%
9	1/2 of 8	7 μ M	11750854	47%
10	1/3 of 8	2 μ M	9131702	26%

Figure 5.5. SDS-PAGE and densitometry, alongside the estimated protein concentrations and catalytic turnover exhibited by screening assays performed with titrated down WelO5 - either from batch purification or from HIS-Select plate elution. Elution from HIS-Select in 250 μ L (E1) (ID 8) yields sufficient protein (concentration and volume) to perform assay repeats in duplicate.

Eluting in an 85 μL (E2) volume from the HIS-Select filter plate (Figure 5.5., Sample ID 6) yielded a similar concentration to that used in the initial screening assay (Sample ID 1). However, this quantity of protein appeared to be in excess because elution in 250 μL (Sample ID 8) to give an approximate concentration of 14 μM still gave approximately 90% catalytic turnover. By eluting in 250 μL instead of the required 85 μL , there was sufficient volume to perform assays in duplicate without significant reductions in reaction turnover.

To examine catalytic turnover by alanine and single-point substitutions relative to wild-type WelO5, the eluted variants and wild-type controls (250 μL E1) were diluted down to achieve 30-50% turnover in the wild-type controls. This allowed the detection of both positive and negative changes in reaction turnover.

Proteins were initially eluted from the HIS-Select plates in the manufacturer recommended phosphate buffer (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0); however, it was found that in this buffer enzyme activity was noticeably reduced relative to the analogous HEPES buffer. Subsequent HIS-Select purifications were then performed using HEPES buffer (20 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 7.4), with the purified variants diluted by 1/3 to enable the assay to be performed in triplicate for each variant (the activity of a 1/2 dilution in phosphate buffer was found to be equivalent to the activity of a 1/3 dilution in HEPES buffer). A summary of the optimised screening assay set-up is described in Table 5.2., with deviations from the original assay highlighted in red.

Composition of screening assay (*cf.* conditions from Chapter 3, Section 3.4.):

Reagent	Stock Solutions	Final Assay Concentration	Volume (µL)
His₆-WelO5	14 µM (estimated)	12 µM (estimated)	85
Small molecule	1 mM (in DMSO)	100 µM	10
(NH ₄) ₂ Fe(SO ₄) ₂	50 mM (in 20 mM HCl)	500 µM	1
2OG	50 mM (in 20 mM HEPES)	2 mM	4
Reaction buffer	20 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 7.4	-	-
Total volume:			100

Table 5.2. Summary of the reaction components used in the WelO5 variant screening assay. Modifications made to the original screening assay are highlighted in red. Preparation of the individual reaction components is described alongside the final reaction concentrations used.

5.3.2.4. Analysis

Both the screening of alternative substrates and assessment of the impact of amino acid substitutions can be performed with this screening assay set-up. However, two different gradients were required for the analysis of samples by UPLC. For both methods, any new peaks ionising with the desired molecular weight could be integrated alongside the starting material to quantify enzyme activity by percentage conversion.

The analysis of samples for percentage chlorination of WelO5 natural substrate by WelO5 variants was performed using a focussed UPLC gradient to achieve resolution between the starting material and chlorinated product whilst minimising run times.

By employing a high aqueous to high organic gradient, a second UPLC method was deliberately more general (i.e. not optimised to a single substrate) and used to analyse a range of different compounds as potential WelO5 or WelO5-variant substrates.

5.3.3. Finalising the testing protocol

This finalised testing protocol can be used for the assessment of WelO5 variants or for the evaluation of new compounds as potential substrates. The protocol is comprised of three components, expression, purification and screening:

Pick BL21 (DE3) colonies into 1 mL volumes of 2TY media (100 µg/mL kanamycin) in 96 deep-well blocks and incubate overnight at 37 °C (220 rpm). Inoculate 20 µL of overnight culture into 2 mL volumes of 2TY Overnight Express™ System 1 auto-induction media (Novagen®) (100 µg/mL kanamycin) in 24 deep-well blocks. Incubate for 2.5 hr at 37 °C (220 rpm), then cool the shaker to 30 °C and incubate a further 20 hr (220 rpm). After this time, harvest the cultures and freeze the culture blocks at -80 °C for a minimum of 30 min.

Defrost the 24 deep-well blocks at room temperature for 30 min. Add 500 µL BugBuster® Master Mix per well and leave at 4 °C for 30 min. Purify the cell-extracts at room temperature using 96-well HIS-Select® filter plates (Sigma-Aldrich) according to the manufacturer's centrifuge protocol (only one wash step required).

Equilibration buffer: 20 mM HEPES, 500 mM NaCl, pH 7.4

Wash buffer: 20 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 7.4

Elution buffer: 20 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.4

Elute the proteins in 250 µL elution buffer.* Variant identity can be confirmed at this stage by mass spectrometry. Transfer 85 µL enzyme to a 96 deep-well block. Add 4 µL 2OG stock (20 mM) and 10 µL (+)-12-epi-fischerindole U isonitrile substrate stock (1 mM in DMSO) per well. Centrifuge for 30 seconds to ensure the wells are well mixed. Add 1 µL freshly prepared Fe²⁺ solution

(50 mM) into each well. Foil-seal the plates and incubate at 37 °C for 90 min. Add 100 µL cold methanol per well and centrifuge for 20 min at 4 °C (4000 g) to quench the reactions by protein precipitation. Transfer the clarified samples to a 96-well plate (150 µL) and analyse by UPLC-MS.

* For the testing of alternative substrates, the eluted protein can be used directly to perform reactions in duplicate. For the testing of variants against the enzyme's natural substrate (where a 30-50 % turnover range is desired for the wild-type controls), the eluted protein can be diluted 1/3 in HEPES elution buffer to perform reactions in triplicate.

5.4. Design of WelO5 variants

As described in the mutagenesis strategy (Section 5.2.), it was decided that the initial variation of the WelO5 active site would involve substitution of selected active site residues with alanine. The substrate pocket has been shown to be large and hydrophobic and generally lacking in any specific charged or hydrogen bonding interactions. Therefore, residues were selected for substitution if their side chains lay within 5Å of the predicted natural substrate binding site region in either (or both) of the two most extreme 'open' (PDB: 5JQR) or 'closed' (5IQT, chain B) WelO5 structures (Figure 5.6., residues highlighted in bold on the 5IQT, chain B structure) - of these residues, three already existed as alanine and so were excluded from further study.

The interaction proposed in the literature to occur between the substrate isonitrile and the Ala82 backbone nitrogen cannot be probed by side chain substitution, however the second proposed interaction between the substrate indole nitrogen and the sulphur of the Met225 side chain is covered in this study.

To assess their importance to WelO5 catalytic activity, metal and 2OG binding residues were also added to the library (Figure 5.6.; 2OG binding residues highlighted in yellow, metal binding histidines highlighted in pink). Two non-alanine substitutions were also added: G166D (G166 aligns with the D157 of SadA) and Y224F (Y224 hydrogen bond was proposed in the literature to have a role in WelO5 β IV-V insert 'lid' closure).¹⁶⁴ A summary of the residues selected for substitution is shown in Figure 5.6. and is shown in three-dimensions in Figure 5.7.

```

      10          20          30          40          50
MSNNTVSTKP ALHFLDINAT EVKKYPTAIQ DIIINRSFDG MIIRGVFPRD

      60          70          80          90         100
TMEQVARCLE EGNDGGMKSI LNKNEEFGTK VAQIYGHAIIV GQSPDLKDYF

      110         120         130         140         150
ASSAIFRQAC RTMFQGS PDF EEQVESIFHS LSGLPVEIPT GPEGQTYTPA

      160         170         180         190         200
TIRLLLEGRE IAVHVCNDFL LMPAANHLKT LLDLSDQLSY FIPLTVPEAG

      210         220         230         240         250
GELVVYSLEW NPQEASKYAQ MQEYMDDVEF KIKSNQSQSV AYAPGPGDML

      260         270         280         290
LFNGGRYYHR VSEVIGNSPR RTIGGFLAFS KQRDKIYYWS

```

Figure 5.6. WelO5 primary sequence, with residues within 5Å of the natural substrate highlighted in bold. 2OG binding residues selected for substitution are coloured yellow, iron binding residues are coloured pink and additional residue G166 is coloured green. Residues comprising the β IV-V insert region are underlined.

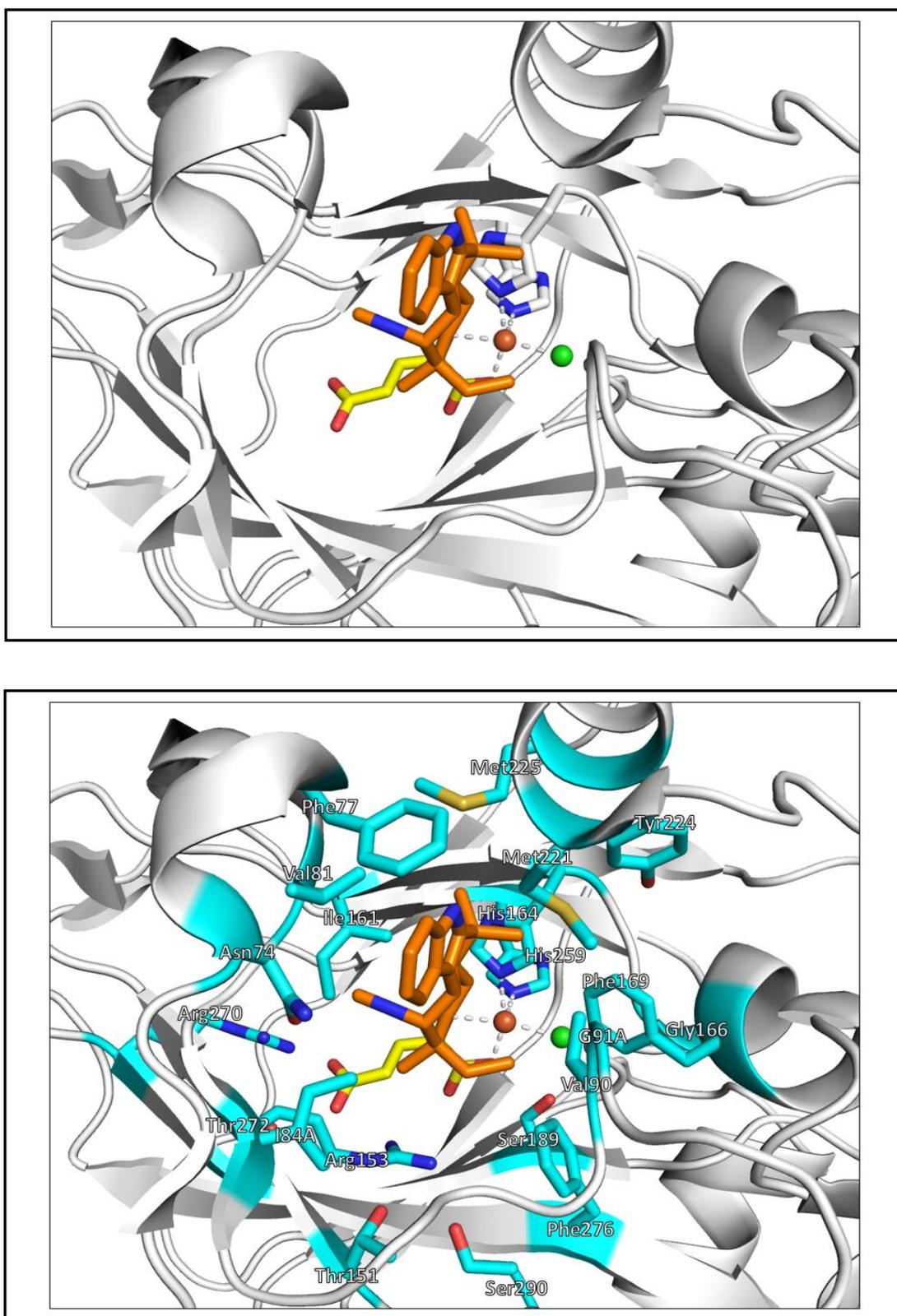


Figure 5.7. Close-up views of the active site of 'closed' form WelO5, with natural substrate (+)-12-epi-fischerindole U isonitrile (orange), 2OG (yellow) and His164 and His259 shown in stick representation. Active-site residues selected for substitution are coloured cyan. Figure generated in PyMOL using the 5IQT (chain B) structure of 'closed' form WelO5.

5.5. Substitution of WelO5 active site residues

WelO5 variants were generated in parallel by site-directed mutagenesis from pNIC28-Bsa4[WelO5] (Chapter 2), using a NEB Q5[®] Site-Directed Mutagenesis Kit (primer sequences in Chapter 7). Variants were expressed and purified according to the finalised protocol (Section 5.3.3.) and expression levels compared by SDS-PAGE analysis of the purified proteins (Figure 5.8.). Levels comparable to wild-type WelO5 were achieved for all variants, except for H259A and T272A. Comparative analysis of the un-purified cell pellets against the wild-type confirmed the lower protein levels seen by SDS-PAGE for H259A and T272A were due to reduced expression, not poor protein solubility.

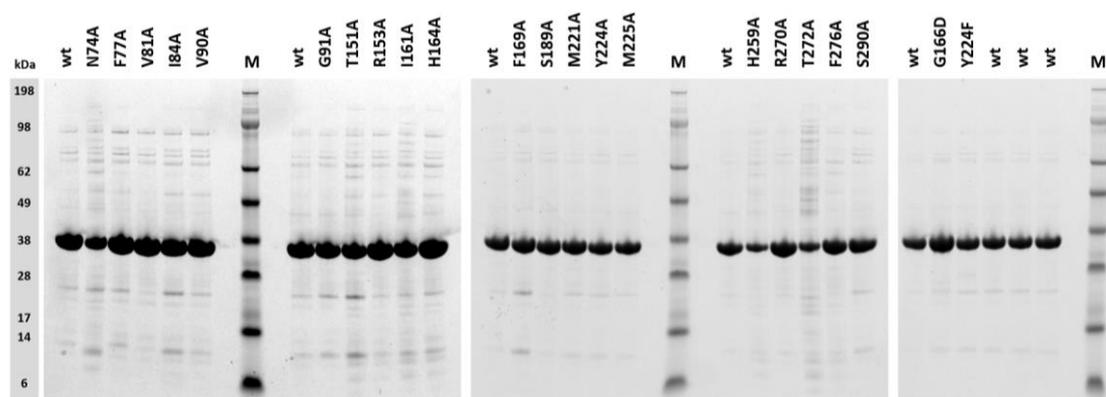


Figure 5.8. SDS-PAGE gels for the comparison of expression levels of alanine scanning mutagenesis variants, as well as G166D and Y224F. Proteins were expressed in parallel in 2TY Overnight Express[™] System 1 auto-induction media at 30 °C over 20 hr. The protein molecular weight marker is labelled M.

All variants were tested for their activity against WelO5's natural substrate to assess the impact of specific active site substitutions upon conversion to chlorinated product (+)-12-*epi*-fischerindole G. The relative percentage peak areas of the starting material and any new products were recorded, with results shown in Figure 5.9. and Table 5.3. Figure 5.9. and Table 5.3. display the results from a single data set, but are representative of the results obtained from performing the finalised protocol (Section 5.3.3.) three times, i.e. n = 3.

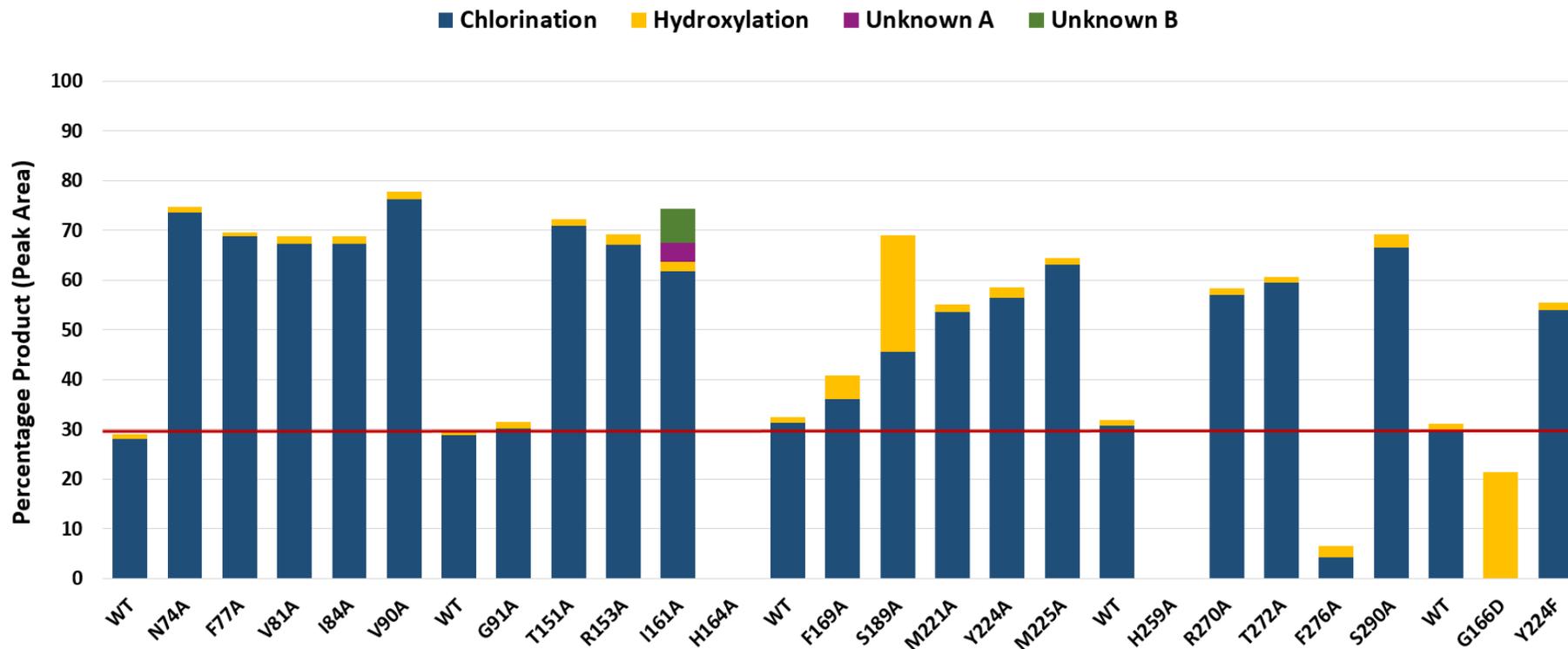


Figure 5.9. Mean percentage conversion of (+)-12-epi-fischerindole U isonitrile to new products by WelO5 variants. Each bar shows the mean average percentage peak area calculated from reactions performed in triplicate for each variant. Chlorination refers to formation of (+)-12-epi-fischerindole G and is represented by a blue bar. Hydroxylation is represented by a yellow bar; and new unknown products A and B are coloured purple and green, respectively. The red line represents the averaged activity levels of the wild-type WelO5 assay controls. Raw data is shown in Table 5.3.

Sample	Chlorination				Hydroxylation				Unknown A				Unknown B			
	R1	R2	R3	Average	R1	R2	R3	Average	R1	R2	R3	Average	R1	R2	R3	Average
WT	28.53	29.81	25.52	27.95	0.91	1.23	1.01	1.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N74A	73.74	72.34	74.36	73.48	1.13	1.10	1.14	1.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
F77A	67.09	68.31	70.72	68.71	0.86	0.94	0.87	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V81A	65.74	67.84	67.99	67.19	1.62	1.63	1.59	1.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I84A	67.72	68.42	65.58	67.24	1.39	1.87	1.29	1.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V90A	78.37	82.60	67.49	76.15	1.62	1.73	1.28	1.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WT	27.20	30.43	28.50	28.71	1.16	1.14	1.08	1.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G91A	28.61	29.84	31.79	30.08	1.37	1.38	1.49	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T151A	69.95	70.68	72.10	70.91	1.28	1.32	1.39	1.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R153A	65.83	67.88	67.27	66.99	2.15	2.20	2.22	2.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I161A	62.38	61.03	61.82	61.74	1.81	1.81	1.96	1.86	3.91	3.71	3.91	3.84	6.94	7.11	6.57	6.87
H164A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WT	30.93	31.47	31.51	31.30	1.18	1.20	1.21	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
F169A	34.88	35.77	37.23	35.96	4.71	4.75	5.07	4.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
S189A	47.54	45.35	43.78	45.56	24.21	22.96	22.85	23.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
M221A	55.51	50.55	54.36	53.47	1.66	1.62	1.65	1.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Y224A	58.51	56.32	54.65	56.49	1.94	2.01	1.90	1.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
M225A	62.11	64.08	63.02	63.07	1.34	1.47	1.43	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WT	30.36	34.15	27.71	30.74	1.05	1.31	1.05	1.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H259A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R270A	57.15	57.35	56.62	57.04	1.21	1.24	1.28	1.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T272A	59.44	59.98	59.07	59.50	1.11	1.10	1.04	1.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
F276A	4.51	4.07	4.15	4.24	2.12	2.38	2.16	2.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
S290A	65.81	66.83	66.91	66.52	2.57	2.80	2.75	2.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WT	33.83	29.60	26.37	29.93	1.42	1.10	0.80	1.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G166D	0.00	0.00	0.00	0.00	22.59	20.22	21.49	21.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Y224F	59.45	53.13	49.36	53.98	1.63	1.42	1.28	1.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.3. Raw data for the percentage conversion of (+)-12-epi-fischerindole U isonitrile to new products by WelO5 variants. Chlorination refers to formation of (+)-12-epi-fischerindole G. Reactions were run for each variant in triplicate (R1-3) and the mean averages calculated for use in Figure 5.9.

Under the conditions of the assay developed here, and in the example data set of Figure 5.9. and Table 5.3., wild-type WelO5 produces, on average, 30% chlorinated (+)-12-*epi*-fischerindole U isonitrile (i.e. (+)-12-*epi*-fischerindole G) and 1% hydroxylated (+)-12-*epi*-fischerindole U isonitrile, respectively. As predicted, substitution of either of the iron co-ordinating histidine residues (H164 & H259) to alanine abolished WelO5 catalytic activity. The WelO5 H259F variant had already been made within the literature, producing the same result.⁵⁰ Variants H164A and H259A demonstrate the critical importance of both of these histidine residues to WelO5 catalytic activity and are useful negative controls for future screening. The results for G166D and S189A were consistent with previously published data.¹⁶⁴ By blocking the chloride-binding co-ordination site, G166D converts WelO5 exclusively to a hydroxylase – abolishing chlorination activity and yielding hydroxylated (+)-12-*epi*-fischerindole U isonitrile as the only product (21%, in this data set). S189 forms a hydrogen bond with C-1 of 2OG and lies near the chloride-binding co-ordination site (Figure 5.10.); variant S189A relaxes functional selectivity between chlorination and hydroxylation, yielding a mixture of both products (46% chlorinated and 23% hydroxylated in this data set) and showing consistency with the literature.¹⁶⁴

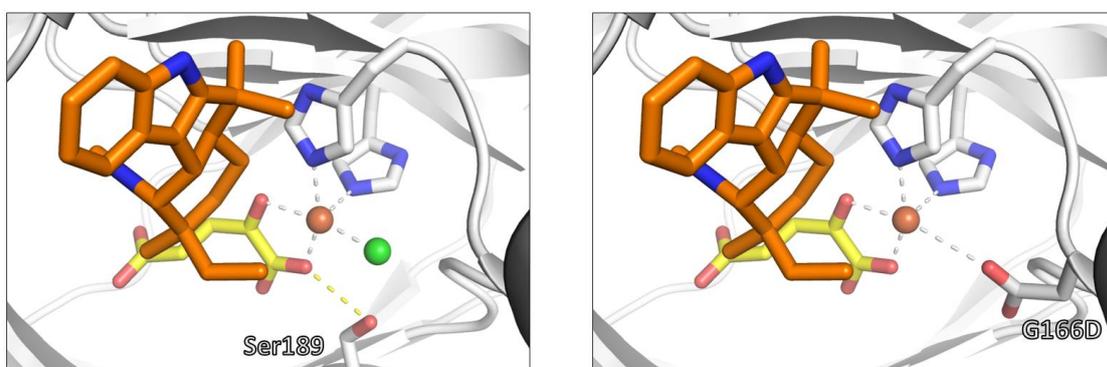


Figure 5.10. WelO5 active site, with 12-*epi*-fischerindole U isonitrile shown in orange, 2OG in yellow and His164 and His259 shown as sticks. S189 sits close to both the chlorine atom and to 2OG. Substitution S189A leads to relaxation of the enzyme's preference for chlorination over hydroxylation. G166D (right) blocks the chloride binding site, abolishing chlorination activity and converting WelO5 into a hydroxylase. Figure generated in PyMOL.

As for H164A and H259A, variants G166D and S189A make suitable assay controls, allowing identification of the hydroxylated fischerindole product (by UPLC retention time) in previously un-investigated variants.

Apart from negative controls H164A and H259A, nearly all alanine substitutions appeared to increase enzyme activity under the tested assay conditions. Interestingly, residues N74, F77, V81, I84 and V90 all exist in a single loop region. This is not the previously discussed prominent β IV-V insert (Chapter 4, Section 4.3.1.5.), but is a region of the WelO5 protein which still demonstrates notable conformational change between crystal structures (comprised of residues 75-82). In the literature, the backbone amide nitrogen of A82 is proposed to interact with the substrate isonitrile.¹⁶⁴ This loop could be responsible for controlling substrate access to the active site of WelO5, with observed increases in WelO5 activity (Figure 5.9., Table 5.3.) the result of additional conformational flexibility introduced by alanine substitution.

Hydrogen bonds from the Y224 side chain hydroxyl to the L165 NH backbone and the D168 side chain were suggested in the literature to have a role in promoting closure of the WelO5 β IV-V insert loop upon substrate binding,¹⁶⁴ however, substitution of Y224 for alanine or phenylalanine did not have a detrimental impact upon enzyme activity. Therefore, assuming these hydrogen bonds exist, they do not appear to have functional importance for WelO5 enzyme activity, at least under the tested assay conditions. Substitution of other WelO5 β IV-V insert residues proximal to the substrate (M221 & M225) for alanine also has no negative impact on enzyme activity.

Apart from Fe²⁺-chelating H164 and H259, the only active site residue where substitution for alanine was found to have a negative impact upon enzyme activity was F276A. Rationalisation of this result is challenging from crystal structures alone; however, this substitution does appear to drastically re-shape the active site pocket surrounding where chloride and oxygen would be expected to bind, and may also interfere with productive 2OG binding (Figure 5.11.).

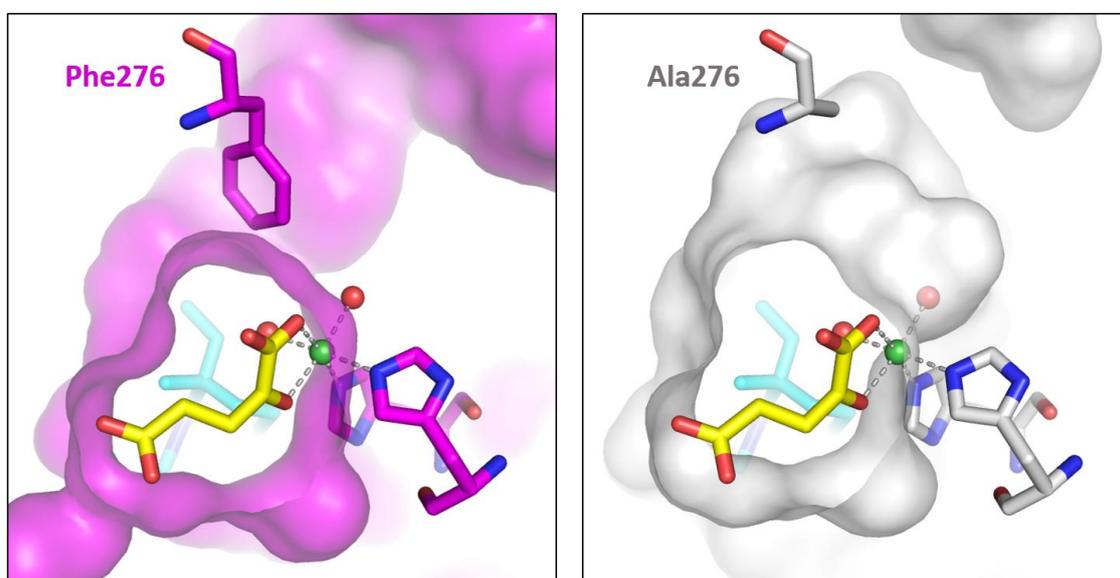


Figure 5.11. Left: WelO5 active site shown in magenta, with residue F276 highlighted. Active site waters are shown as red spheres, with Fe²⁺ represented by a green sphere. 2OG is shown in yellow and WelO5 natural substrate (+)-12-epi-fischerindole U isonitrile is shown in cyan. Right: WelO5 F276A variant with drastically re-shaped active site (due to alanine substitution) shown in grey. Figure generated in PyMOL.

It is also interesting to note that alignments of WelO5 with the active sites of 2OG-halogenases CurA Hal (PDB ID: 3NNF) and SyrB2 (PDB ID: 2FCT) reveal an arginine residue in the position corresponding to F276, with a phenylalanine in the position corresponding to R153 in WelO5 i.e. exchanged residues, which could suggest cooperative binding for these side chains.

Substitution of I161 for alanine resulted in two new products of as yet unknown identity, which were formed exclusively by this alanine variant.

Under the conditions of the assay developed here, and in the example data set used in Figure 5.9. and Table 5.3., WelO5 variant I161A produces, on average, 62% chlorinated product (12-*epi*-fischerindole G), 2% hydroxylated product, 4% unknown product A and 7% unknown product B (Figure 5.12.). From the mass spectrum, product A was found to have a molecular weight of 320 Da and product B a molecular weight of 318 Da. Unfortunately, the identity of products A and B could not be elucidated by UPLC-MS analysis alone. However, a molecular weight of 320 Da is consistent with hydroxylation. A different retention time (0.90 min) to the hydroxylated product already identified in the literature (seen at 0.60 min) is indicative of the formation of a new isomer.¹⁶⁴ In addition, the molecular weight of product B (318 Da) is consistent with ketone formation.

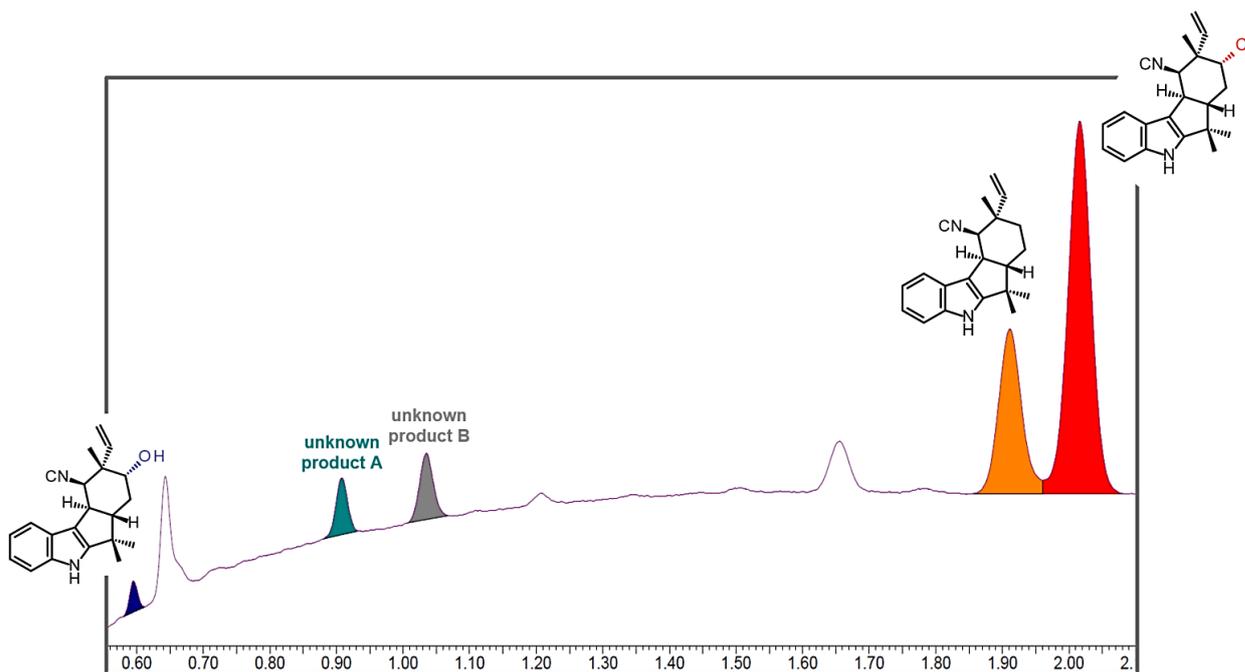


Figure 5.12. UPLC trace of I161A reaction described in Figure 5.9. and Table 5.3., with the relevant peaks integrated and highlighted. WelO5 substrate (and the reaction starting material) is coloured orange (RT 2.22 min, 25%); literature known chlorinated and hydroxylated products are coloured red (RT 2.35 min, 62%) and blue (RT 0.66 min, 2%) respectively.⁵⁰ Unknown product A (RT 1.04 min, 4%) is coloured green and unknown product B (RT 1.19 min, 7%) is coloured grey.

Isolating and identifying unknown products A and B from the I161A reaction mixture was deemed impractical due to the reactions being performed on a

100 μ L, 100 μ M substrate scale and the mixture of multiple products present. By producing the doubly substituted variant I161A_G166D, it was hoped that the chlorination reaction would be eliminated (as seen for G166D) and the percentage peak areas of the unknown products would increase as a result, aiding their eventual isolation when performed on scale (Figure 5.13.).

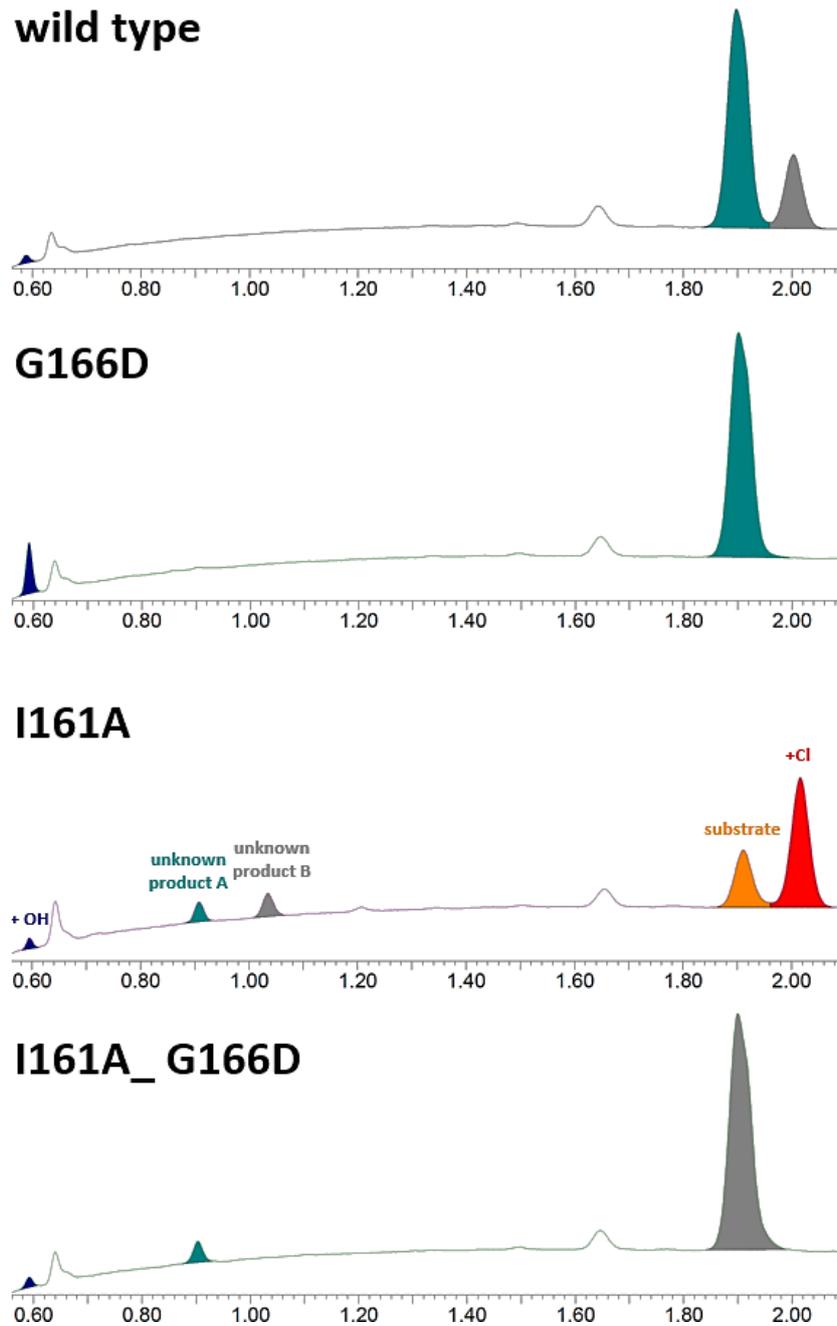


Figure 5.13. Comparison of the UPLC reaction profile of the wild type WelO5 enzyme with that of variants G166D, I161A and doubly substituted variant I161A_G166D.

However, whilst the I161A variant I161A_G166D eliminated chlorination activity as expected, the production of product B was also abolished, indicating that product B may be formed from chlorinated product 12-*epi*-fischerindole G. The lack of a distinctive 3:1 chlorine isotope ionisation pattern for product B, combined with the observed molecular weight of 318 Da, led to the hypothesis that formation of product B proceeds *via* the same hydroxylation as unknown product A but instead involving the loss of chloride (hypothesized mechanisms shown in Figure 5.14.).

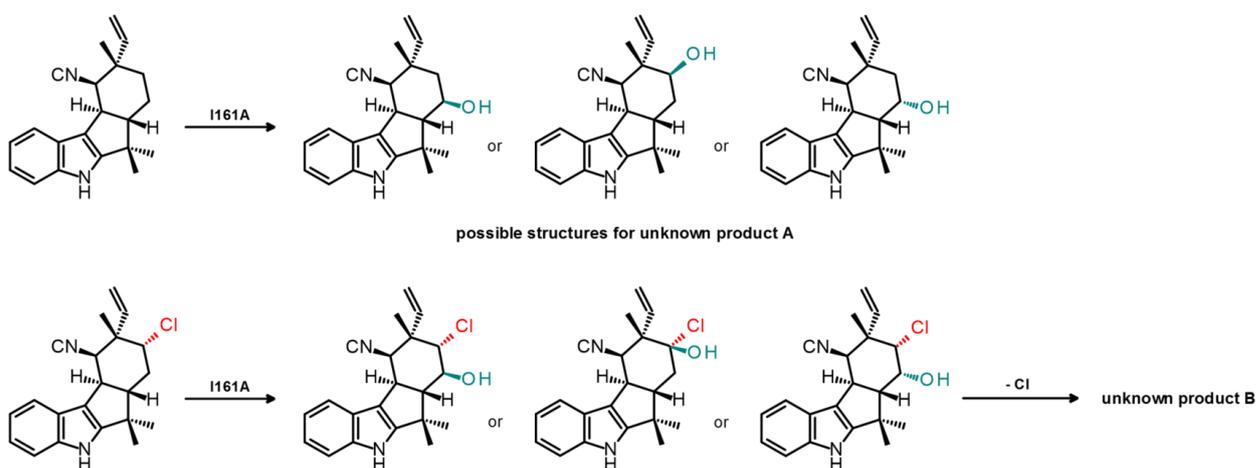


Figure 5.14. Proposed reaction pathways for the formation of unknown products A and B from (+)-12-*epi*-fischerindole U isonitrile by WelO5 variant I161A.

The loss of chloride by one of the following pathways could lead to the structures proposed for product B in Figure 5.15., consistent with the observed molecular weight of 318 Da:

- Pathway 1: Spontaneous chloride elimination, followed by tautomerisation to the ketone.
- Pathway 2: Epoxide formation by displacement of chloride by the vicinal hydroxyl. This pathway is perhaps less likely; although interestingly, 2OG-oxygenase catalysed epoxidation has been described in the literature.^{207,208}

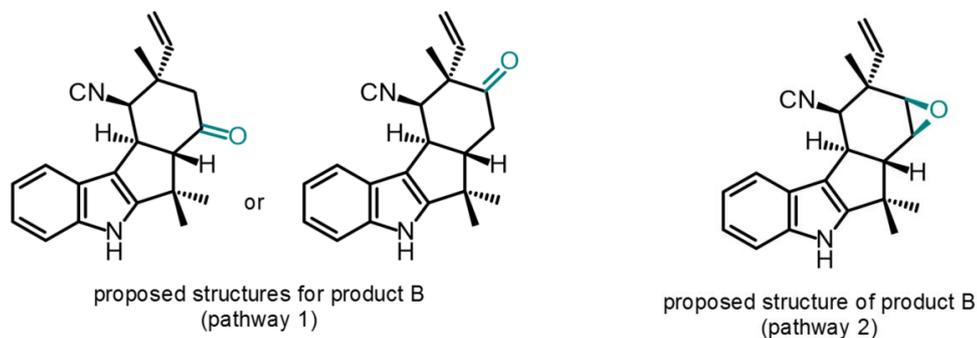


Figure 5.15. Proposed structures for the identity of I161A reaction product B.

Hydroxylation of the opposite face of the cyclohexyl ring to the isonitrile could proceed if the natural substrate (+)-12-*epi*-fischerindole U isonitrile was able to rotate 180 degrees within the active site. Variant I161A would allow this to occur by removing a clash between the isoleucine side chain and the vinyl group of the rotated substrate (Figure 5.16.).

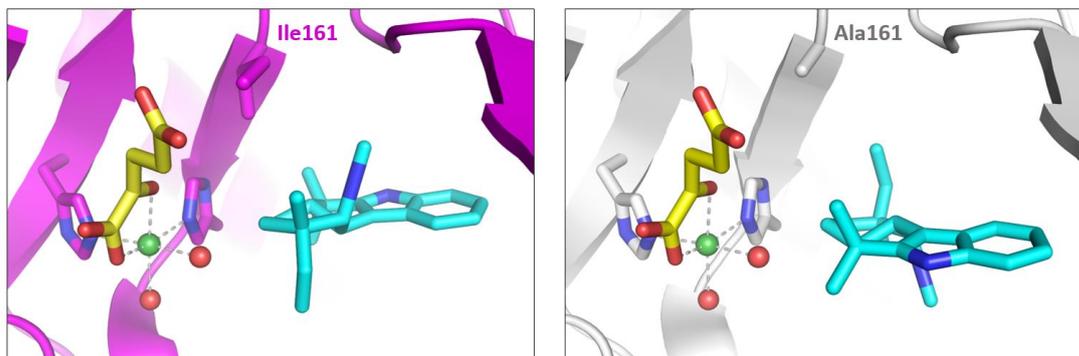


Figure 5.16. On the left is the wild-type WelO5 active site coloured magenta, with residue I161 highlighted. Active site waters are shown as red spheres, with Fe²⁺ represented by a green sphere. 2OG is shown in yellow and WelO5 natural substrate (+)-12-*epi*-fischerindole U isonitrile is shown in cyan. Rotation of the natural substrate by 180 degrees appears to be permitted within the active site of WelO5 variant I161A (shown right, in grey) due to the removal of a steric clash (existing in this substrate orientation only) between the substrate vinyl group and the I161 side chain. Figure generated in PyMOL.

Further work is required at this point to scale up and isolate unknown products A and B to confirm which of these proposed hypotheses is correct.

In an attempt to see if the relative proportions of unknown products A and B could be increased relative to I161A, several different amino acid substitutions

were made at the I161 position. These were designed using structural overlays with the active site of other 2OG-oxygenase crystal structures. Similar band sizes were seen by SDS-PAGE for these purified variants suggesting no issues with expression, however, none were found to be greatly superior to I161A in terms of reactivity under the tested assay conditions (Figure 5.17. & Table 5.4.). I161N was the only variant found to produce detectable quantities of unknown product A (1% vs. 4% for I161A) as well as comparable quantities of unknown product B (8% vs. 5% for I161A). However, the increases seen were not significant enough to warrant scale-up of the I161N reaction.

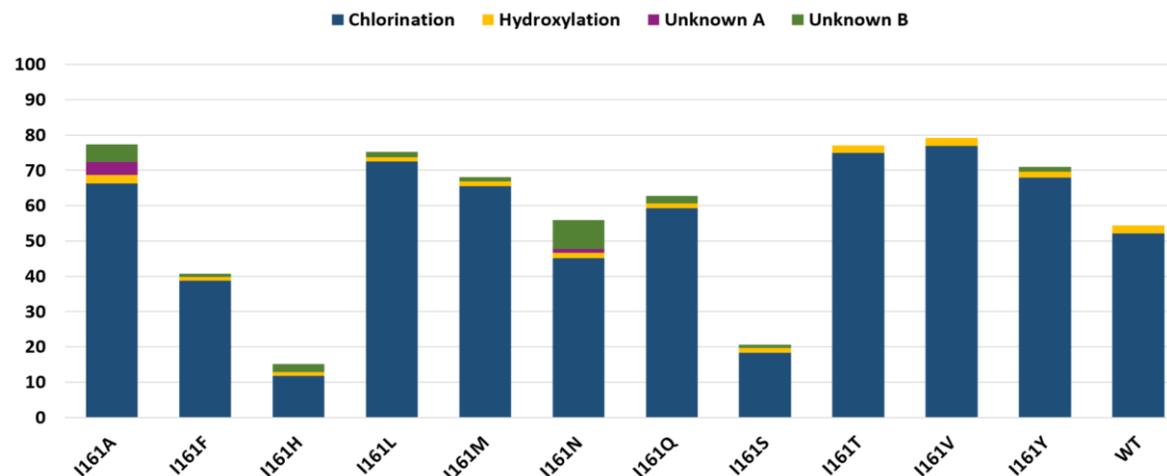


Figure 5.17. Mean percentage conversion of (+)-12-epi-fischerindole U isonitrile to new products by I161X variants. Each bar shows the mean average percentage peak area calculated from reactions performed in triplicate for each variant. Chlorination refers to formation of (+)-12-epi-fischerindole G and is represented by a blue bar. Hydroxylation is represented by a yellow bar; and new unknown products A and B are coloured purple and green respectively. Raw data is shown in Table 5.4.

179

Sample	Chlorination				Hydroxylation				Unknown A				Unknown B			
	R1	R2	R3	Average	R1	R2	R3	Average	R1	R2	R3	Average	R1	R2	R3	Average
I161A	65.57	66.75	66.60	66.31	2.38	2.25	2.40	2.34	4.03	3.69	3.35	3.69	5.32	4.95	4.83	5.03
I161F	42.07	38.80	35.20	38.69	1.16	0.95	1.34	1.15	0.00	0.00	0.00	0.00	1.05	1.00	0.79	0.95
I161H	11.34	12.91	11.14	11.80	0.84	0.96	1.37	1.06	0.00	0.00	0.00	0.00	1.99	2.91	2.06	2.32
I161L	72.60	73.68	71.13	72.47	1.04	1.30	1.21	1.18	0.00	0.00	0.00	0.00	1.76	1.34	1.51	1.54
I161M	64.39	66.90	65.03	65.44	1.35	1.35	1.35	1.35	0.00	0.00	0.00	0.00	1.30	1.30	1.18	1.26
I161N	41.83	47.30	46.46	45.20	1.31	1.37	1.59	1.42	1.16	1.03	1.29	1.16	8.38	8.45	7.51	8.11
I161Q	58.01	61.52	58.29	59.27	1.32	1.31	1.36	1.33	0.00	0.00	0.00	0.00	2.41	2.43	1.82	2.22
I161S	18.21	19.02	18.02	18.42	1.32	1.37	1.48	1.39	0.00	0.00	0.00	0.00	0.84	0.88	0.60	0.77
I161T	75.62	74.38	74.62	74.87	2.24	1.97	2.19	2.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I161V	76.92	76.66	77.09	76.89	1.86	2.09	2.89	2.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I161Y	65.73	66.76	71.12	67.87	1.29	1.79	2.30	1.79	0.00	0.00	0.00	0.00	1.33	1.22	1.16	1.24
WT	52.57	49.63	53.96	52.05	1.68	2.01	3.22	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.4. Raw data for the percentage conversion of (+)-12-epi-fischerindole U isonitrile to new products by I161X variants. Reactions were run for each variant in triplicate (R1-3) and the mean averages calculated.

5.6. Alternative substrates

5.6.1. (+)-12-*epi*-fischerindole U isonitrile ‘substrate-like’ compounds

From the substrate-related compounds in Chapter 3, three were selected for activity testing against the existing panel of WelO5 variants (Figure 5.18.) A small number of additional active site substitutions were designed using WelO5 crystal structures to introduce or increase activity relative to the wild-type.

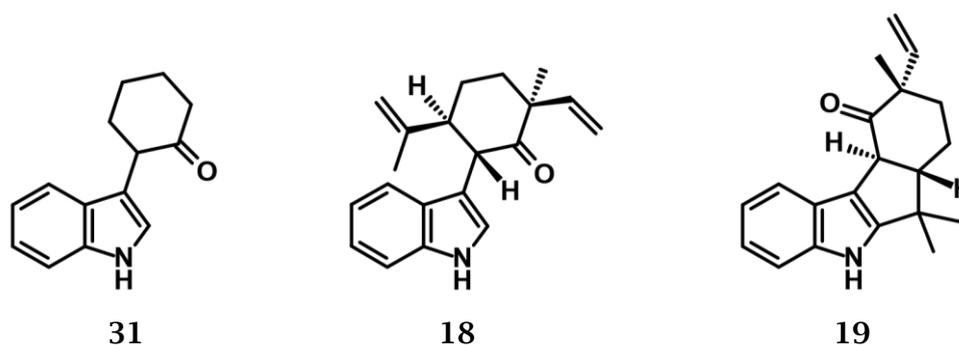


Figure 5.18. Structures of the three compounds selected for activity testing against a panel of WelO5 variants.

2-(1H-indol-3-yl) cyclohexanone (**31**):

Wild-type WelO5 activity was not previously seen for this compound (Chapter 3, Section 3.4.). Variants V81S and I161N were designed to pick up hydrogen bonding interactions, whilst A88V was designed to alter the active site shape to better fit the unsubstituted cyclohexanone ring (Figure 5.19.).

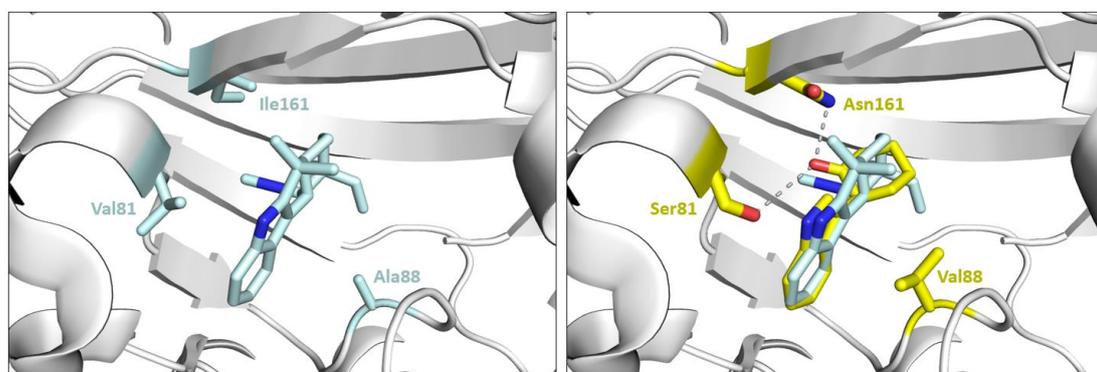


Figure 5.19. Left: wild-type WelO5 with natural substrate (+)-12-*epi*-fischerindole U isonitrile bound and residues selected for substitution shown as sticks coloured pale cyan. Right: 2-(1H-indol-3-yl) cyclohexanone (overlaid with the WelO5 natural substrate) and designed substitutions shown in stick form and coloured yellow. Figure generated in PyMOL.

Under the tested assay conditions, none of the WelO5 variants from the pre-existing panel, or specifically designed variants V81S, A88V & I161N displayed activity towards this compound.

(2S,5S,6S)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone (18):

Previously, wild-type WelO5 had been found to yield approximately 1% chlorinated product with this compound as a substrate (Chapter 3, Section 3.4.). I84N and I161S were designed to pick up hydrogen bonding interactions, whilst V90G and F159V were designed to increase space to better accommodate the isopropenyl unit (Figure 5.20.). Under the tested assay conditions, no improvements were seen over the wild-type chlorination level for any of variants from the pre-existing panel or for specifically designed variants I84N, V90G, I161S & F169V. Hydroxylation activity was also not detected.

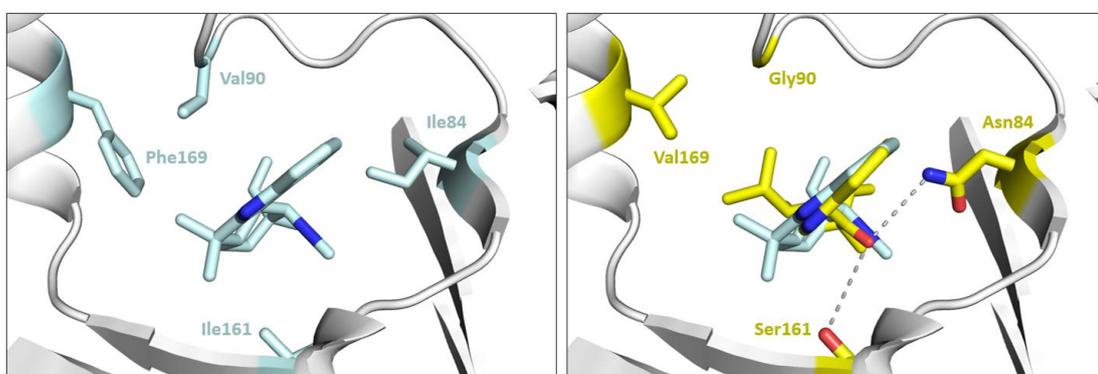


Figure 5.20. Left: wild-type WelO5 with natural substrate (+)-12-epi-fischerindole U isonitrile bound and residues selected for substitution shown as sticks coloured pale cyan. Right: *(2S,5S,6S)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone* (overlaid with the WelO5 natural substrate) and designed substitutions shown in stick form and coloured yellow. Figure generated in PyMOL.

(6aS,9S,10aS)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one (19):

The most ‘natural substrate-like’ of the three selected compounds had been found to undergo non-enzymatic desaturation (loss of two protons) under the assay conditions, with subsequent chlorination of the desaturated compound by wild-type WelO5 (Chapter 3, Section 3.4.). Wild-type WelO5 produced an average of 43% chlorinated-desaturated product and 4% hydroxylated-desaturated product. A88N and I84T were designed to pick up hydrogen bonding interactions with the ketone (Figure 5.21.). Under the tested assay conditions, no improvements were seen over the wild-type chlorination level for any of variants from the pre-existing panel or for specifically designed variants A88N and I84T.

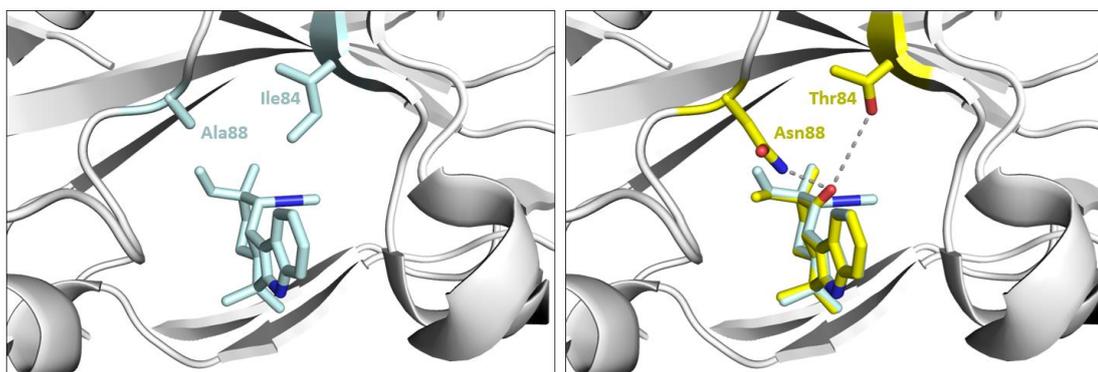


Figure 5.21. Left: wild-type WelO5 with natural substrate (+)-12-epi-fischerindole U isonitrile bound and residues selected for substitution shown as sticks coloured pale cyan. Right: *(6aS,9S,10aS)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one* (overlaid with the WelO5 natural substrate) and designed substitutions shown in stick form and coloured yellow. Figure generated in PyMOL.

Interestingly, under the tested assay conditions S189A was able to significantly increase hydroxylation activity against the desaturated starting material (45% hydroxylation achieved), whereas G166D hydroxylation was comparable to wild-type (Figure 5.22.). S189A also introduced a different hydroxylation event (8% peak area, different retention time - *cf.* I161A).

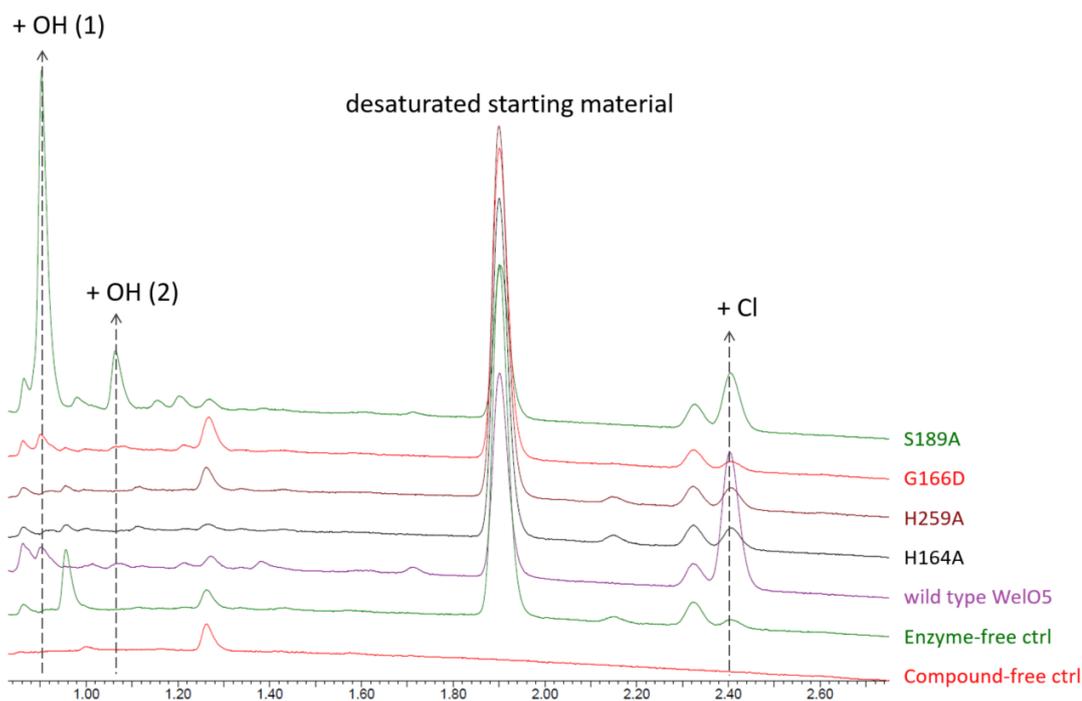


Figure 5.22. Overlay of key UV traces from WelO5 variant activity testing against desaturated-(6a*S*,9*S*,10a*S*)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5*H*-indeno[2,1-*b*]indol-10-one (RT 1.90 min). S189A demonstrates notably higher hydroxylated product (RT 0.90 min) than all other variants, as well as a novel hydroxylation product at RT 1.06 min. G166D and assay controls are also included for comparative purposes. Chlorinated desaturated-product co-elutes with an unknown at RT 2.40 min (peak seen at 2.40 min in the controls but absence of chlorinated product confirmed by MS).

5.6.2. New substrate choices

Two compounds, structurally unrelated to the enzyme natural substrate, were then selected for activity testing using the pre-existing panel of WelO5 variants. A UCB-project compound (generic structure shown) and the natural product eburnamonine were selected upon the basis of containing a reasonably flat aromatic-saturated ring system and a ketone for hydrogen-bonding in a manner analogous to compound **19** (Figure 5.23.).

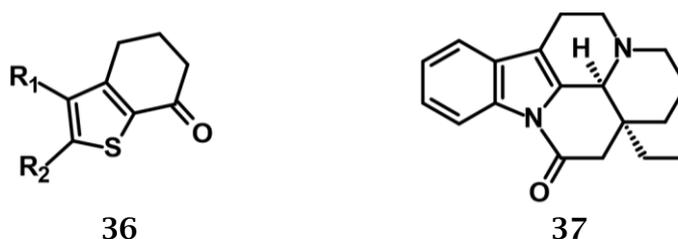


Figure 5.23. UCB 5,6-dihydro-4*H*-benzothiophen-7-one project compound (left), and natural product eburnamonine (right) were selected for activity testing against a panel of WelO5 variants.

Both compounds were selected based on overlays with compound **19** in the active site of WelO5 crystal structures (structure of UCB project compound not shown; eburnamonine overlay shown in Figure 5.24.).

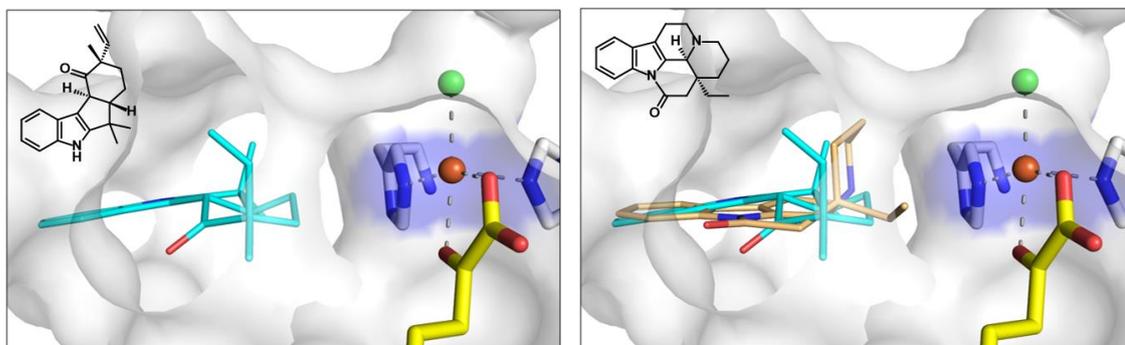


Figure 5.24. Structural overlay of eburnamonine (orange) with compound **19** (cyan) within the active site of WelO5 crystal structure 5J4R. The metal centre is shown as an orange sphere, with the co-ordinating chlorine shown in green. 2OG (yellow) and the two metal-chelating active site histidine residues (H164 and H259) are shown in stick form. Figure generated in PyMOL.

UCB 5,6-dihydro-4H-benzothiophen-7-one project compound (**36**):

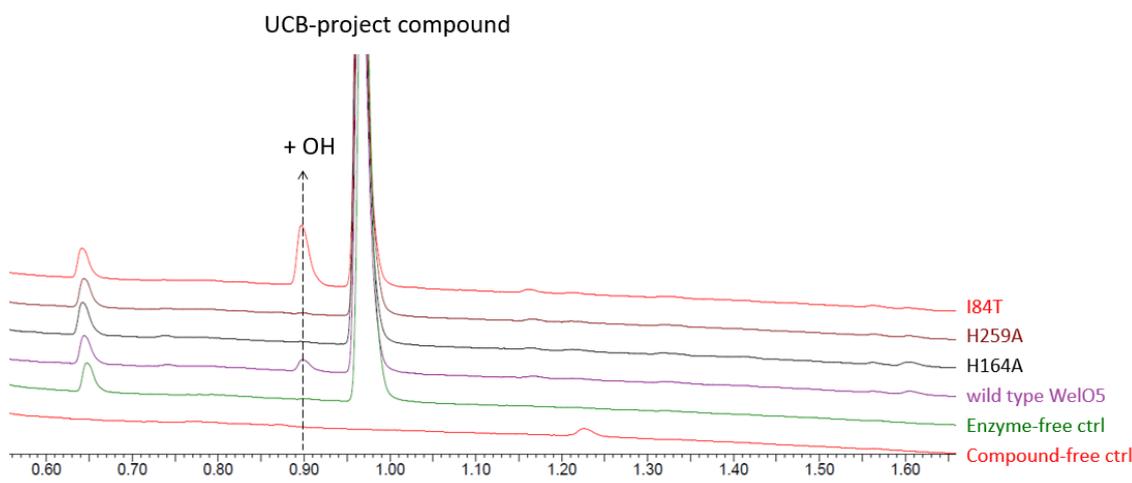


Figure 5.25. Overlay of key UV traces from WelO5 variant activity testing against UCB 5,6-dihydro-4H-benzothiophen-7-one project compound (RT 0.97 min). I84T showed the highest percentage of hydroxylation product (RT 0.90 min) of all variants tested. Assay controls are also included for comparative purposes.

Under the tested assay conditions, wild-type WelO5 hydroxylation activity was seen for this compound - with an average of 3% hydroxylated product seen across the wild-type assay controls. No chlorination was detected. The entire panel of WelO5 variants demonstrated comparable or reduced hydroxylation

activity relative to the wild-type except for I84T (12% hydroxylation, Figure 5.25.) and A88V (6% hydroxylation).

Eburnamonine (37):

Under the tested assay conditions, wild-type WelO5 hydroxylation activity was seen for this compound, with an average of 4% hydroxylated product seen across the wild-type assay controls. Two new unknown products were also identified, with m/z corresponding to single (-2H) and double (-4H) desaturation of the starting material at 2% and 5%, respectively. As these unknowns were not present in any of the controls (enzyme free, H164A or H259A), the formation of these compounds also appears to be enzyme catalysed. The entire panel of WelO5 variants demonstrated comparable or reduced activity relative to the wild-type, with the exception of G166D. G166D showed an average of 12% hydroxylated product, 5% unknown A (SM-4H) and 11% (SM-2H) unknown B (Figure 5.26.)

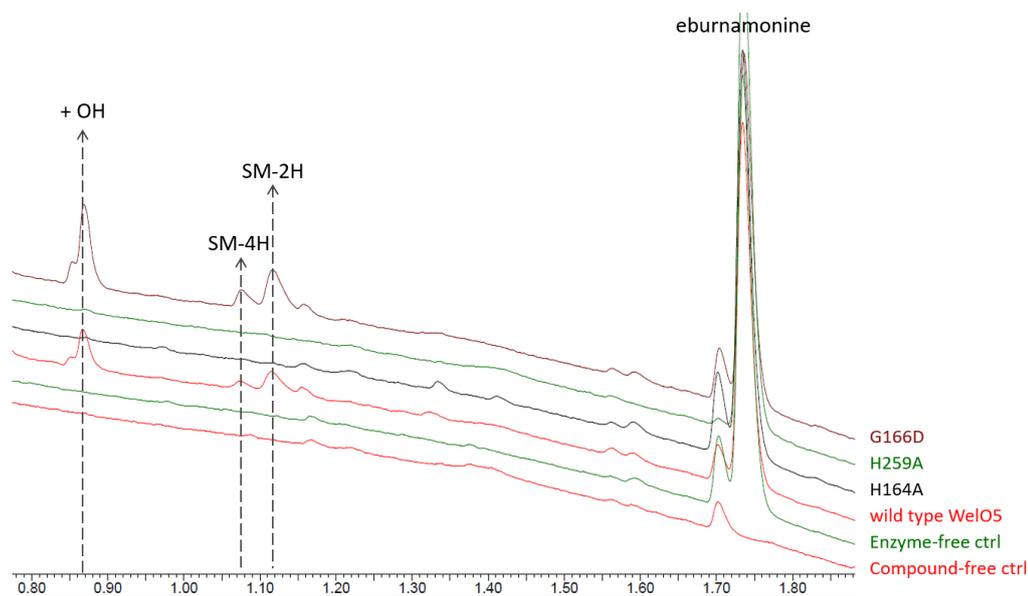


Figure 5.26. Overlay of key UV traces from WelO5 variant activity testing against eburnamonine (RT 1.73 min). G166D showed the highest percentage of hydroxylation product (RT 0.86 min) of all the variants tested. The formation of two enzyme-catalysed desaturation products (RT 1.07 min, 1.12 min) was also found to increase over wild-type WelO5. Assay controls are also included for comparative purposes.

5.7. Summary of WelO5 evolution

In this Chapter, initial efforts towards the evolution of WelO5 variants with new activities are described. Analysis of crystal structures obtained for WelO5 (and those published in the literature) confirmed a large hydrophobic pocket, likely responsible for substrate binding at the active site. An initial alanine scan of selected active site residues was implemented, with significant assay method development carried out to enable the production, purification and screening of multiple WelO5 variants in parallel.

The active site of WelO5 was shown to be highly tolerant of variation - aside from the catalytically important histidine residues, only F276A resulted in a reduction in WelO5 activity, with the vast majority of alanine substitutions actually responsible for increases in WelO5 activity. Identified from the initial active-site alanine scan, I161A was found to introduce a novel enzyme activity. This was assigned as a new hydroxylation event, with the I161A variant able to act upon both the natural substrate, (+)-12-*epi*-fischerindole U isonitrile, and the enzyme-chlorinated product, (+)-12-*epi*-fischerindole G. The increased space created by swapping the bulky isoleucine for the significantly smaller alanine was thought to allow a new orientation of the compound in the active site, leading to a hydroxylation reaction in a different position. Further variants were designed at position 161 based upon crystal structures of other 2OG-oxygenase family members, however, none could significantly improve upon the activity seen for I161A.

A small number of both substrate-like and structurally unrelated compounds were selected for activity testing against the developed panel of WelO5 variants and some additional single-point amino acid substitutions designed

from the WelO5 crystal structures. Activity upon acyclic substrate-analogues **18**, **19** & **30** could not be introduced or improved upon. However, for compound **19**, S189A was found to significantly increase the proportion of hydroxylated product formed (45% vs. wild-type 4%), with a new hydroxylation product also being detected. Wild-type WelO5 hydroxylation activity was also identified for compounds structurally unrelated to the natural substrate, a UCB project compound (**36**) and eburnamonine (**37**). G166D was found to improve eburnamonine hydroxylation (12% vs. wild-type 4%), and I84T and A88V improved hydroxylation of UCB compound **36** (12% and 6% respectively vs. wild-type 3%).

The design of successful WelO5 active site substitutions based upon the crystal structure alone was more significantly challenging than anticipated. None of the designed single point amino acid substitutions were able to introduce or improve wild-type WelO5 activity for the compound which they were designed - serendipity had to be allowed for by screening the new compounds against the entire WelO5-variant panel, not just the specifically designed ones. The WelO5 variants discussed in this Chapter could serve as starting points for future evolution, with significant improvements in enzyme activity theoretically achievable from additive combinations of multiple substitutions. By improving upon the activities described in this Chapter, the enzymatically-produced compounds could be produced as major reaction components and isolated for structural confirmation.

Chapter 6: Summary & future work

The capacity of 2OG-oxygenases acting on small molecules to perform a wide range of different reactions, despite their common active-site structural features,^{64,67} indicates they may represent a well-adapted structural platform for the development of biocatalysts. WelO5, a 2OG-halogenase which performs a regio- and stereo-selective chlorination on a lipophilic small molecule,⁵⁰ was envisaged to be a promising starting point for the evolution of novel C-H functionalisation biocatalysts. The work described in this thesis has begun to investigate the tractability of developing synthetically useful biocatalysts based upon WelO5.

Very little was known about the structure of WelO5 in 2014. Given that the reactions of many 2OG-oxygenases are not amenable to high-throughput screening or selection methods, it was envisaged that structure-guided design of active site substitutions would be the most predictable way of modifying WelO5 selectivity. In Chapter 2, alignments with 2OG-halogenases of known structure (SyrB2, CytC3 & CurA Hal) revealed a distinct lack of homology (< 15% identity and \leq 25% similarity shared with WelO5) (Table 2.3. & Figure 2.4.) and therefore the development of a credible homology model was considered unlikely from these analyses. PHI-BLAST searches were also unable to identify any 2OG-oxygenases closely homologous to WelO5 with available three-dimensional structural information (Table 2.2. & Fig 2.3.) However, evaluation of protein tertiary structure prediction with open-source algorithms (Phyre2, RaptorX, SWISS-MODEL & I-TASSER) led to the identification of 2OG-hydroxylase SadA¹¹⁹ as a template for WelO5 structural prediction. SadA was found to share only ~20% sequence identity with WelO5 (explaining why it had

not been identified previously from the PHI-BLAST), but was later confirmed by structural overlays with WelO5 crystal structures to possess significant structural similarities to WelO5 (Figures 4.19. & 4.21.). A model of WelO5 based on the SadA template was constructed to help visualise the three-dimensional arrangement of amino acid side chains within the WelO5 active site in the absence of crystallographic information.

Whilst unable to identify any proteins with known structures similar to WelO5, a PHI-BLAST search did identify ten WelO5 homologues (minimum 45% similarity and 30% identity to WelO5), proposed to be 2OG-halogenases (Table 2.1. & Figure 2.1.). Only one (AmbO5) had previously been discussed in the literature, opening up the other nine for future investigation. Interestingly, most of these WelO5-homologous putative 2OG-halogenases were identified from cyanobacterial species, suggesting a possible evolutionary origin independent of the 2OG-halogenases already published in the literature. Production of these ten WelO5 homologues was investigated at the OPPF; however, soluble protein expression levels were generally found to be poor. These constructs were retained for future work, which could initially involve either the optimisation of expression conditions or LIC cloning into the vector successfully used for WelO5 expression. Production of these proteins would enable their characterisation and confirmation as novel 2OG-halogenase enzymes and they may also be able to demonstrate complementary selectivities to WelO5. Their high sequence similarity to WelO5 means that the homology models developed in this thesis are likely sufficient for the design of active site variations for these enzymes without the requirement for X-ray crystallography.

The natural substrate for WelO5, (+)-12-*epi*-fischerindole U isonitrile, is a complex natural product which was not immediately or easily accessible. The 20L algal fermentation process used in the literature to access ≤ 1 mg of the desired WelO5 substrate was determined physically impractical. However, modification of the Baran group's synthesis¹³⁴ of (-)-12-*epi*-fischerindole U isothiocyanate yielded milligram quantities of (+)-12-*epi*-fischerindole U isonitrile for screening assay development and activity testing (Scheme 3.2.).

Issues were encountered with the repeatability of the superb selectivities achieved in the literature for the reductive amination of (6*aS*,9*S*,10*aS*)-6,6,9-trimethyl-9-vinyl-6*a*,7,8,10*a*-tetrahydro-5H-indeno[2,1-*b*]indol-10-one and several alternative approaches were tried without achieving noticeable improvements. The material was pushed through this route regardless, and the undesired amine isomer also used to synthesise an alternative, related compound - ((6*aS*,9*S*,10*S*,10*aS*)-10-isocyano-6,6,9-trimethyl-9-vinyl-5,6*a*,7,8,10,10*a*-hexahydroindeno[2,1-*b*]indole; **32**), for activity testing and structure-activity relationship deductions. In the future, a more thorough investigation into alternative conditions for accessing the amine is justified. This would be facilitated by improvements in yield for the preceding acid-catalysed cyclisation reaction. In the literature, degradation was reported for the product under the reaction conditions,¹³⁹ which was reduced by quenching the reaction early and recycling the starting material. However, in practice the multiple recycling steps were not suitable for the generation of large amounts of intermediate. Adapting this reaction to perform in flow may improve the reaction yield and have a significant downstream impact on the limitations of WelO5 natural substrate availability.

A gene encoding for WelO5 was found to express very well in BL21 (DE3) *E. coli* cells from an expression vector constructed *via* LIC (ligation-independent cloning) and was purified to a high purity by affinity and gel filtration chromatography for the set-up of crystallisation trials. Initial attempts to crystallise WelO5 in the presence of Ni²⁺ and NOG were unsuccessful; however, the presence (by co-crystallisation) of an analogue of the natural substrate (**19**) did lead to the growth of a number of crystals.

A structure of WelO5 was solved by single-wavelength anomalous diffraction (SAD) using the anomalous signal from the selenium atoms of recombinantly expressed SeMet-WelO5. This structure was then used to solve three novel crystal structures of WelO5; 5J4R at 1.65Å and 5T22 at 1.75Å (containing Ni²⁺ and 2OG in the active site) and 5TRQ at 1.30Å (containing Ni²⁺ and succinate). Unfortunately, despite co-crystallisation with **19**, 5J4R and 5T22 did not contain any electron density within the active site consistent with the binding of this ligand. However, these crystal structures did reveal a large, open hydrophobic active site (reflective of the large, hydrophobic substrate) and hinted at flexible regions bordering the active site (Glu75 - Ala82; Ile89 - Gln92 and Trp210 - Glu238; part of a helical insert between β10(IV) and β11(V) (see Figures 4.12. & 4.13.). It was already known that the βIV-V insert region varies greatly between members of the 2OG-oxygenase superfamily, and is therefore likely to have a role in substrate specificity. Indeed, re-examination of the sequence alignments generated for the ten WelO5-homologous putative 2OG-halogenases revealed the greatest cross-sequence variability corresponded to these proposed flexible regions (Figure 2.1.). Structure 5T22 (succinate) also revealed conformational changes involving Arg153 in the active site which may be involved in substrate binding (Figures 4.16. & 4.17.). Insights generated

from investigation into the relationship between conformational change and substrate binding by a solution-based technique would justify investment in generating an NMR assignment for WelO5. If successful, this would be the first small molecule 2OG-halogenase NMR 'structure' in the literature.

Co-crystallisation hanging-drop screening plates were set up in the glove-box with WelO5, Fe²⁺, 2OG and **19** in an attempt to obtain a ligand-bound structure by simulating the active enzyme reaction conditions as much as feasibly possible. Poorly diffracting, needle-like clusters were obtained but further optimisation was not performed due to publication of the Liu group WelO5-substrate crystal structures.

Considerable effort was invested into the development of a robust method suitable for the screening of multiple enzyme variants in parallel.

Intermediates isolated during synthesis of the WelO5 natural substrate and a small number of simplified structural analogues were tested for WelO5 halogenation activity. WelO5 was found to be highly specific for substrate-like molecules, which further galvanised investment in active site re-design. An initial alanine-scan of the WelO5 active site revealed that WelO5 is highly tolerant to variation, with most alanine substitutions maintaining or increasing enzyme activity (Table 5.3. & Figure 5.9.). The most interesting variant from the alanine scan (I161A) was found to introduce a new hydroxylation activity (Figures 5.12. to 5.15.). Further work is needed to scale up this reaction for product identification and confirmation of reaction mechanism. Additional active site substitutions will likely be necessary to improve the turnover of this reaction sufficiently to generate the product in appropriate quantities for characterisation and structural conformation. A total of 40 WelO5 variants

were produced and tested for activity on the natural substrate and structurally similar analogues. Two structurally unrelated compounds; a UCB project compound and eburnamonine were also tested. Variants demonstrating modest increases over the wild-type turnover for both substrates were found. However, the introduction of activities not already seen in the wild-type incubations for other compounds was not achievable - possibly because the compounds were not binding within the active site in the first place. As the developed LC-MS based screening method is generally applicable, further activities on multiple different compounds can be explored with additional single point and combination substitutions. Additional changes to the WelO5 reaction outcome (e.g. hydroxylation, bromination,⁴⁹ nitration and azidation⁸³) by replacement of the Cl for a different Fe²⁺ co-ordinating ligand could also be evaluated using this method.

WelO5 has been shown as an excellent starting point for the development of new biocatalysts for C-H functionalisation. The enzyme has a large, lipophilic active site - ideal for the binding of small molecules of interest to industry. Novel crystal structures are available to facilitate active site re-design and WelO5 has been shown to be highly tolerant of (and responsive to) active site substitution. Whilst this thesis has developed a platform to facilitate WelO5 evolution (crystal structures, robust expression and screening protocol, initial NMR work), further work will be required in this area to add such enzymes to the biocatalytic toolbox of the future.

Chapter 7. Materials & methods

7.1. Bioinformatics

7.1.1. Homology modelling

Homology models were created for the ten putative Fe²⁺ and 2OG-dependent halogenases listed in Chapter 2 (Section 2.2.1.1.) by using PROMALS3D²⁰⁹ to generate a multiple sequence alignment. For this, the halogenase primary sequences and SeMet-WelO5 structure file were used as inputs. Homology models were created by Sebastian Kelm (UCB), using the multiple sequence alignment generated by PROMALS3D and the SeMet-WelO5 crystal structure as a template. Models were created using proprietary UCB homology modelling software.

7.2. Biological methods

7.2.1. Solutions & media

Solutions, media and buffers were prepared using Milli-Q® Ultrapure water and autoclaved if required sterile. Heat sensitive solutions were sterilised by filtration (0.22 µm pore size).

Growth media: Media was prepared as below and diluted to 1L with Milli-Q® water prior to autoclaving.

Per 1L:	2xTY	LB
Tryptone	16 g	10 g
Yeast	10 g	5 g
NaCl	5 g	10 g

For solid media (i.e. plates), 15 g agar was added before diluting to the 1L final volume. 5% sucrose selection plates (LIC cloning) were prepared by diluting an autoclaved 2X LB agar media solution using an equal volume of filter-sterilised 10% sucrose solution.

Antibiotics: Kanamycin solutions were prepared in Milli-Q® water and filter-sterilised. Carbenicillin solutions were prepared in 50% ethanol. Antibiotic solutions were stored in aliquots at -20 °C until required.

Antibiotic	Stock Concentration (mg/mL)	Working Concentration (µg/mL)	Dilution factor	µL per 100 mL media
Carbenicillin	100	50	1:2000	50
Carbenicillin	100	100	1:1000	100
Kanamycin	50	50	1:1000	100
Kanamycin	50	100	1:500	200

7.2.2. Gel electrophoresis

7.2.2.1. Agarose (DNA)

Gels were prepared by dissolving 1% agarose (w/v) in TAE buffer (1X) with microwave heating. Invitrogen™ SYBR™ Safe DNA gel stain (10,000X) was added just prior to gel pouring to give a 1X working concentration.

Samples were prepared by adding NEB Gel Loading Dye, Purple (6X), no SDS and run alongside NEB Quick-Load Purple 1kb DNA Ladder in TAE buffer (1X) on a Bio-Rad system at 80V. Gels were run at 120 V for approximately 35 min and visualised using a Bio-Rad ChemiDoc™ XRS+ Imaging System.

7.2.2.2. SDS-PAGE (protein)

Loading buffer (4X): 1 mL NuPAGE™ LDS Sample Buffer (4X) was added to 250 µL NuPAGE™ Sample Reducing Agent (10X).

NuPAGE™ 4-12% Bis-Tris Pre-Cast Protein Gels (1.0 mm) were run in an Invitrogen Mini Gel tank using NuPAGE™ MES SDS Running Buffer (1X). 4X loading buffer was added to each sample, before heating to 70 °C for 5 min. 10 µL (5 µL for 15 well gels) sample was loaded per well and run alongside SeeBlue™ Plus2 Pre-stained Protein Standard. (NuPAGE™ reagents: Invitrogen).

Gels were run for 35 min at 180 V, then stained with InstantBlue™ (Expedeon). Gel images were captured using a GE Healthcare ImageQuant LAS 4000 imaging System.

7.2.3. Transformations

A 50 µL aliquot of chemically competent *E. coli* cells was thawed on ice (approx. 10 min). 1 µL of plasmid DNA (50-100 ng/µL) was added to the cells and the mixture kept on ice for 30 min. The cells were heat shocked at 42 °C (water bath) for 30 seconds, followed by a 5 min recovery period on ice. 200 µL of SOC was added and the transformed cells incubated at 37 °C for 1 hr (220 rpm). Transformants were spread over two agar plates (1/5 and 4/5 of transformation respectively) containing the appropriate antibiotic and allowed to dry before incubating either at 37 °C overnight or at room temperature for 48 hr.

7.2.4. Starter cultures

A single colony was picked from a freshly streaked agar plate (prepared from glycerol stock or fresh transformation) into 2xTY media containing the required antibiotic and incubated overnight at 37 °C (220 rpm).

7.2.5. Glycerol stocks

Glycerol stocks were prepared for long term plasmid storage. 750 µL of non-induced overnight cell culture was added to a sterile cryotube containing 250 µL of sterile glycerol and stored at -80 °C.

7.2.6. Plasmid DNA stocks

5 mL starter cultures were harvested by centrifugation (6000 g, 10 min). Plasmid DNA was isolated from the cell pellet by mini-prep (QIAprep Spin Miniprep Kit, Qiagen) according to the manufacturer's protocol. Plasmids were isolated by elution in 50 µL of the supplied elution buffer and stored at -20 °C.

7.2.7. DNA quantification

DNA samples were quantified using a ThermoFisher Scientific NanoDrop™ ND-1000 UV-Vis spectrophotometer.

7.2.8. Ligation independent cloning (LIC)

7.2.8.1. Synthetic genes

The WelO5 gene was ordered codon optimised for *E. coli* from GeneArt® (Life Technologies™). The supplied glycerol stock was plated out (50 µg/mL kanamycin plates) and a DNA stock prepared as described in Section 8.2.6.

Codon optimised WelO5 DNA sequence:

```

ATGAGCAATAATACCGTTAGCACCAAACCGGCACTGCATTTTCTGGATATTAATGCCACCGAGGTGAAAA
AATACCCGACCGCAATTCAGGATATTATCATCAATCGTAGCTTCGACGGTATGATTATTCGTGGTGTTTT
TCCGCGTGATACAATGGAACAGGTTGCACGTTGTCTGGAAGAGGGTAATGATGGTGGTATGAAAAGCATC
CTGAACAAAAACGAAGAGTTTGGCACCAAAGTTGCCAGATTTATGGTCATGCAATTGTTGGTCAGTCTC
CGGATCTGAAAGATTATTTTGAAGAACAGGTGGAAAGCATTTTTTCATAGCCTGAGCGGTCTGCCGGTTGAAATTCCGACA
TCCGGATTTTGAAGAACAGGTGGAAAGCATTTTTTCATAGCCTGAGCGGTCTGCCGGTTGAAATTCCGACA
GGTCCGGAAGGCCAGACCTATACACCGGCAACCATTTCGTCTGCTGCTGGAAGGTCGTGAAATTGCAGTTC
ATGTTGGTAATGATTTTCTGCTGATGCCTGCAGCAAATCATCTGAAAACCCTGCTGGATCTGAGCGATCA
GCTGAGCTATTTTATCCCGCTGACCGTTCCGGAAGCCGGTGGTGAAGTGGTTGTTTATAGCCTGGAATGG
AATCCGCAAGAGGCAAGCAAATATGCACAGATGCAAGAGTATATGGACGATGTCGAGTTCAAAATCAAAA
GCAATCAGAGCCAGAGCGTTGCCTATGCACCGGGTCCGGGTGATATGCTGCTGTTTAAATGGTGGTTCGTTA
TTATCACCGTGTAGCGAAGTTATTGGTAATAGTCCGCGTCGCACCATTGGTGGTTTTCTGGCATTTAGC
AAACAGCGCGACAAAATCTATTATTGGAGC

```

7.2.8.2. Expression vectors

Genes were cloned into the pNIC28-Bsa4 expression vector (SGC) by LIC.

Vector	Antibiotic resistance	Tag	Tag cleavage	Source	Addgene plasmid #
pNIC28-Bsa4	Kanamycin	N-terminal His ₆ -tag	TEV protease	SGC	26103

7.2.8.3. Primers

Primers were obtained *via* Sigma-Aldrich custom DNA synthesis services and re-suspended in nuclease-free water (Invitrogen™) for a working concentration of 100 µM. The WelO5 gene was amplified from the supplied pMK-T vector using primers designed for LIC:

```

WelO5_LIC_F      TACTTCCAATCCATGAGCAATAATACCGTTAGCACCAAAC
WelO5_LIC_R      TATCCACCTTTACTGTTAGCTCCAATAATAGATTTTGTGCGCGCT

```

7.2.8.4. Protocol

Vector preparation: 5 µg of pNIC28-Bsa4 vector was digested with BsaI in CutSmart® Buffer (NEB) for 3 hr at 50 °C and purified using a GeneJET PCR Purification kit (ThermoFisher Scientific) according to the manufacturer's protocol, eluting in a 50 µL volume of the supplied elution buffer. N.B. BsaI digestion removes the sacB gene from the vector. Presence of the sacB vector

renders cells unable to grow in the presence of sucrose, allowing for negative selection on sucrose-containing agar plates.

WelO5-insert generation: To a 200 μ L PCR microtube (on ice) was added:

	Volume (μL)
Nuclease-free water (Invitrogen™)	13.5
5X Q5® polymerase reaction buffer (NEB)	5.0
50 mM MgSO ₄	0.5
10 mM dNTP mix (NEB)	0.75
10 μ M primer stock*	2.5
10 ng/ μ L WelO5 DNA	2.5
Q5® DNA polymerase (NEB)	0.2
Total	25

*Primer stock (10 μ M): 5 μ L 100 μ M WelO5_LIC_F primer + 5 μ L 100 μ M WelO5_LIC_R primer + 40 μ L nuclease-free water.

The PCR reaction was run in a Techne Genius thermal cycler as outlined below:

	Cycles	Temperature	Time
Pre-heat lid	1	Ramp to 95 °C	∞ (insert samples)
Denaturation	1	95 °C	1 min
Denaturation		94 °C	30 sec
Annealing	30	52 °C	30 sec
Extension		68 °C	1.5 min
Extension	1	68 °C	5 min
Hold	1	4 °C	∞

Generation of cohesive ends (pNIC28-Bsa4 and WelO5-insert): To a 200 μL

PCR microtube (on ice) was added as follows:

For WelO5-insert cohesive ends:

	Volume (μL)
Nuclease-free water (Invitrogen™)	2.15
5X T4® polymerase reaction buffer (NEB)	1.0
100 mM DTT	0.5
BSA (NEB, 10 mg/mL)	0.1
dCTP (25 mM)	1.0
WelO5-insert	5.0
T4 DNA polymerase (NEB)	0.25
Total	10

For linearised pNIC28-Bsa4 cohesive ends:

	Volume (μL)
Nuclease-free water (Invitrogen™)	21.5
5X T4® polymerase reaction buffer (NEB)	10
100 mM DTT	5
BSA (NEB, 10 mg/mL)	1
dGTP (25 mM)	10
linearised pNIC28-Bsa4	50
T4 DNA polymerase (NEB)	2.5
Total	100

PCR reactions were run in a Techne Genius thermal cycler as outlined below:

	Cycles	Temperature	Time
Incubation (generation of LIC overhangs)	1	22 °C	30 min
Polymerase inactivation	1	75 °C	20 min
Hold	1	4 °C	∞

Annealing & transformation:

On ice, 5 μ L T4 DNA polymerase-treated WelO5-insert was added to 2 μ L T4 DNA polymerase-treated vector. The DNA was left to anneal for 15 min at room temperature before returning to ice and then transformed into XL10-Gold cells (Agilent). Transformants were spread over 5% sucrose agar plates (50 μ g/mL kanamycin) and incubated overnight at 37 °C. A single pNIC28-Bsa4[WelO5] XL10-Gold colony was used to prepare a DNA stock as described in Section 8.2.6.

7.2.9. OPPF cloning & expression**7.2.9.1. Synthetic genes**

Synthetic genes were codon optimised for *E. coli* and ordered from GeneArt® (Life Technologies™). For each gene the supplied glycerol stock was plated out (50 μ g/mL kanamycin plates) and a DNA stock prepared as described in Section 8.2.6.

Codon optimised DNA sequence FASTAs:

```
>AmbO5
ATGAAAAACAACACCCGCAACTTTAACATCGGCAATGTGATTCATGGTAAACGCCTGATTATCACCGAAG
AAAAACTGATGAGCAATAATGCCGTTAGCACAAAAGCGCACTGAACTTTCTGGATATTAATGTGACCGA
GGTGAATAACTATCCGACCGCAATTCAGGACATTATTATCGATCGTCGTTTTGACGGTATGATTATTCGT
GGTGTTCGCCGCTGGATAGCATTGAACGTGTTATTCGTCTGCTGGAAGATGAAGATGAGGGTGGTATGA
AACTGATCTTCAACAAAAACGAAGAGTTCGGCACCAAAGTTGCACAGATTTATGGTCATGTTATTGTTGG
TCAGTCTCCGGATCTGAAAGATTATTTTGCAAGCAGCGCAATTTTTTCGTCAGGCATGTCGTGCACTGTTT
CAGGGTAACCCGGATTTTGAAGAACGTGTTGAAAGCGTTTTTTCATAGCCTGTGTGGTCTGCCGGTTGAAA
TTCCGACAGGTCCGGAAGGTCAGAGTTATACACCCGCAACCCTGCGTCTGCTGACCGAAGGTTCGTGAAAT
TACCGTTCATGTTGGTAATGATTTTCTGCTGATGCCTGCAAGCGAACATCTGAAAACCCTGCTGGATTTT
AGCGATCAGCTGAGCTATTTTATCCCGCTGAGCGTTCCGCAGGCAGGCGGTGAACTGGTTGTTTATAATC
TGGAATGGAATCCGGAACAGGCAGAACAGAGCGGTGATCTGCACAAATATATGAATGATGCAGATAGCCG
TTTTTCAGAGCCAGCAGAGTCAGAGCGTTGCATTTGCACCGGGTCCGGGTGATATGATTCTGTTAATGGT
GGTCGCTATTACCATCGTGTTAATCAGGTTATTGGTAATAGTCCGCGTCTGACCATTTGGTGGTTTTCTGG
CATTTAGCAAAGAACGCAACAAAATCTACTATTGGAGC
```

>CytC3

ATGACCACCGTTAGCGAACAGGCCAAATTTTACCTTTAGTCCGGAAGAAGTTGCACGCTTTGAACGTGATG
 GTTATATTGGTCCGGTGAAAATCTTTGAACCGGAAGAAATGACCCGTCGCTGGAACATTATTTCGTGCTCA
 GCTGCTGGATCGTAGCCTGGCCATTTATCCGGATTCAAATGGTAAAGCCAACATCAGCAATTATGATCGC
 CATCTGGATATTGATCTGCTGGCAGAACATATTATGCGTCCGGAATTTGTTGATCGTGTTGGTAGCCTGA
 TTGGTCGTAATCTGCTGTGCTGGCGTAGCGAATTTTTTCCGAAATATCAGGGTGATGAAGGCACCGATTG
 GCATCAGGCAGCAACCTTGCACATGCAACCGGTAAACCGCAGATTATTTGGCCGTCAGATGAAGGTCGT
 CCGGCATTTATTGGCACCATTACCGTTTGGACCGCATTACCCATAGCACCCGAACAGAATGGTTGTCTGC
 AGCTGATGCCGTCACCCATACCAGCATGAATTATGATGAAAGCAAAGGCATGGATTATGACGCCGATGC
 AATTAATCAGCGTGAAAAAGATGGTATCAAACGCGTTTTTTTCGGCTATGATTATCGTAGTCTGCAGAAA
 GATCCGGATTGGAACCGGATGAAAGCCAGGCATATCCGATGGTTCTGAAACCGGGTGAAGCAGTGATTT
 TTTGGAGCAATACCATGCATGCAAGCCTGCCGCATACCGGTAGCAAAACCGATTATCGCATGGGTTTTGC
 AGCACGTTATGTTCCGACCCAGGTTACGGTTTATCCGGGTACAGAAAATCTGACCGAATATGGTGATGGC
 ATCAACCTGGAAAAATATGGTGCAGTTCTGACCAGCGGTGTTGATGAATATGGTCATAATCGTATTGCAC
 GTACCAGCCAGCGTGGTTATGAATTTGTTCCGCGTCAGATTCCGAGC

>SyrB2

ATGAGCAAAAAATTCGCACTGACCGCAGAACAGCGTGCCAGCTTTGAAAAAATGGTTTTATTGGTCCGT
 TCGATGCCATATAGTCCGGAAGAAATGAAAGAAACCTGGAACGTACCCGTCGCTGCTGGATCGTAG
 CGCAGCAGCATATCAGGATCTGGATGCAATTAGCGGTGGCACCAATATTGCAAATTATGATCGTCATCTG
 GACGATGATTTTCTGGCAAGCCATATTTGTGCTCCGGAATTTGTGATCGTGTTGAAAGCATTCTGGGTC
 CGAATGTTCTGTGTTGGCGTACAGAATTTTTTCCGAAATATCCGGGTGATGAAGGCACCGATTGGCATCA
 GGCAGATACCTTTCGAAATGCAAGCGGTAAACCGCAGATTATTTGGCCTGAAAATGAAGAATTTGGTGGC
 ACAATTACCGTTTGGACCGCATTACCGATGCAAATATTGCCAATGGTTGCCTGCAGTTTATTCCGGGTA
 CACAGAATAGCATGAACTACGATGAAACCAAACGCATGACCTATGAACCGGATGCCAATAATAGCGTTGT
 TAAAGATGGTGTTCGTCGCGGTTTTTTTCGGTTATGATTATCGTCAGCTGCAGATCGATGAAAATGGAAA
 CCGATGAAGCAAGCGCAGTTCGATGCAGATGAAAGCAGGTGAGTTTATTATCTTTTGAGACCCCTGA
 TGCATGCAAGCTATCCGCATAGCGGTGAAAGCCAAGAAATGCGTATGGGTTTTGCAAGCCGTTATGTTCC
 GAGCTTTGTTTCATGTTTATCCGGACAGCGATCACATTGAAGAATATGGCGGTGCTATTAGCCTGGAAAAA
 TATGGTGCAGTTCAGGTGATTGGAGATGAAACACCGGAATATAATCGTCTGGTTACCCATACCACCCGTG
 GCAAAAAATTCGAAGCAGTT

>WP1942

ATGCCGAAAAGCGTTACCGCAAAGAACCCTGTTAACTTTCTGGATATTCACGTGAGCGAAATCGAAA
 AACATCCGAGCGCACTGCAGGATATGCTGTTAATCGTACCTTTGAAGGTGTGATTATTCGTGAAGTTCT
 GCCGAAAGATGTTATTGAACTGGTTGTTAGCCGTCGAAAAACGATGAAGTTGATATGTCCGAAATCCTG
 GATTCCAAATTTGCAGATGCAGTTAAAGGTGCATATACCCTGGGTGAAAATATTGTTTTTCCGATAGCA
 GCCTGCAGCAGTATTTGATTATGCAGCAATTTTTCGTCAGAAATGCCGTACCCTGTTTGAAGGTATGCT
 GGATTTTGAAGCACTGATGGAAAGCATGCTGAGCAGCCTGAGCGGTGGTCTGACCGTTAAAATCCGAGC
 GGTAGCCAGGGTCAGAGCTATATCCGGCAACCATTTCGTAAAGTTCCGACCAATCATGAATTTCCGATTC
 ATGTGGGTAACGATTTTCTGAATAGTCCGCAGAGCGAACATCTGCGTACCCTGGTGGATGTTACCGATCA
 GCTGAGCTTTTTATCCCGCTGGTTCTGCCGGAAGCCGGTGGTGAAGTGGTATTTATGGTATGGAATGG
 GATGGTGAAGAACCTGGTTTTTTGAAGATACCAATAGCAATGATGGTTATGGCAGCCGCTGTATCATCAGA
 GCCATCAGGTTTTTATCAGCAGTATGGTAGCATGACCTTTAAACCGAATATTGGCGATATGATGCTGTT
 TGATGGTGGTCGTTTTTATCATTGTATTGTTCCGACCGTTGGTGATCGTACCCTATTACCATTGGTGGT
 TTTCTGGCATTTAGCAAAGAACACGATGCCATCTATTATTGGAGC

>WP2172

ATGACCATTGCAGCACGTA AAAACCCTGTATAATCTGGTGGAAATCGACTTTGAAAATCTGCCTCGTCATT
 ATCCGGATGGCATTAAAGATATTTACGAGCGCAA AATTGATGGCCTGATCATCGAAGATTTTCTGAGCCA
 GGATGAAGTTGATAATGTTGTTGAACAGCTGACCCAGAAAGGTCCGTTTGAAGGTAGCCCGTTTGGTGTAT
 ATCTCGATTTATGGTCCGGCACTGTATGTTAGCGAAGCAGATACCCAAGAATATTTGTGAGCAGGCAACCC
 AGTTTCGTGAATATTGCCGTTACCTGTTTCGTAATGGTCGTGATTTTGA AACCCGCTGCAAGAGGTTCT
 GAGCGTTATGAGCGGTGGTTCGTGAAGTTATGGTTCTGAAAAGTCAGGATGGTAGCGAATATACCCCGACC
 ACCATTCGTGTTCTGGAAAAAGGTCAGTTTATGGGTTGGCATT TTTGGCAATCAGTTTCTGTTTTGTACCC
 CTGGCTATAAACATCTGAAAACCCAGATTGATACCACCGCACATCTGGGTTATTTTCTGGTGTGAGCAG
 CGCAATTAGCGGTGGTGA ACTGATTCTGTATGATCTGGAATGGGATGAAACCGAATGGAATGATACAGAA
 ATGGGTGGTCGTAAACGTAATGGCACCGTTAATGGTAAACTGATCGCAGATGTGATGGAAAATTATCAGC
 AGATGCCGATTA AACCGAAACCGGGTAGCCTGGTTATTTTTGATGATAGCCGATTTCTGCATCGTGTAG
 CACCGTTGAAGGTAATCGTCGTCTGATTACC GTTGGTGGTTTTGTTGCATTTAGCGAAAATCGCGAGAAA
 GTGTATTATTGGAGC

>WP3823

ATGGCCTATGTGAAAAACGATCTGTGGATTACCGTTAAAGCAGATAGCCTGGCAAATTATCCGGATGCAG
 TTAACCAGATTTACAAAGGTGAAATTGATGGCATGCTGATCCAGCAGGTTTTTACCAAAGAAGAAATGCT
 GAAAGCCAAACACCAGCTGGAAAATAACAGCAATCGTAAAATTGTGGGCTATGGCAAAGTTTTGGTATT
 GTGCTGCTGGATAAAGAACAGGATCAGACCAAATATTTCCAGGATGTTAAAGCCTATCGCACCCGA ACTGA
 ATAAAATCTTTGCAACCGGTTATGAAGCACGTATTGAAGCACTGCTGCGTAAAGTTAGCGGTGCACGTAA
 AGTTGAAGTGGCAAGCGAAAATAATCAGACCTATGCACCGGCACAGATTCGCTTTATGGAACCGGAATAAA
 GGTGGCCTGATTATTCATAAAGGTAGCCAGTTTTCTGACCGCATCAGGCATTTGCACATCTGCGTGA AATTG
 CACGTCTGCAAGAACATCTGAGCTATTTTTCTGATCATCGATAAACCGGAAAGCGGTGGTGA ACTGATTAT
 CTATGATCTGCCTCCGGAAGAAGCCAAAAAGATTTTTGATACCCTGAGCAAACACAGCGCCTTTGAAAAA
 TGTGACAAAAAGTATATCAGTCCGGACATTGGTGTATGTTGTTTTTTCATGGTGGCATCATTTGGCATA
 AAGTTGCAGATGCAAAGGCAGCAAAAACCGTATTAGCATTGGTGGTTTTGTTGCGCTGAGCAAAGATAA
 CCAGAAAATCTTCTATTGGAGC

>WP6042

ATGAGTGATGCACAGACCAGCAAACCGTTTTAGCTTTCTGGA ACTGAGCGCACCCGGTTGCACTGCAAGAAC
 GTAATCTGATTCGTGAAATTGCAAGCAATGAACTGCCTGGTGTATTGTTTCGTGGTGCATATACCCGGTGA
 AGAAATGCAGCGTATTGATCGTCTGAGCCAGGGTGAAGCACGTGATGCACTGTTTGCACCCGCCTAGC
 GCAGTTTCGTGAAGGCACCGTTGAACAGCATCATCATTATTATGGTATGAATCTGGTTACCCCGAAAGCAA
 CCCGTCAGATGTATCAAGAATTTGCACAGCGTTTTGATATTGCATGTCGTGCCCTGTTTCGTGATCTGGC
 AGATTTTCGTGAACTGCTGATGCAGACCGTTTCGTAATCTGGGTGCAGGTATGGGTGCAGATGTTCCGGCA
 TTTGCAAGCGGTGAACCGTATACCTTTGCAACCCCTGCGTGGTTTTTCCGCCTGGTCTGGAAGTTCCGCTGC
 ATTTGTGTAATCACATGGTTTCAGTTAGCCCTGCCTATGAGCATCTGCGTACCCTGACCAGCCATAGCCA
 TGATCAGCTGAGCTTTTTATGCCGATTGGTCTGCCGGATGCCGGTGGTGA ACTGATTGTTTTATGATCTG
 TATTGGGAACCGGAACATAGCAAAAGCGCAGGTAGCCTGCATGGTGTGATGCAGCAATGGTTAAAGATCATG
 CATGGCAGGCAGCAGA ACTGAATCCGGGTGATATGATTTTTTTCAATGGCGGTGATCTATACCCGTGT
 TAGCAAAGTTGAAGGTAGCCGCTGCGTCTGACATTGGTGGTTTTTCTGAGCGTTAGCAGCGATCTGGAA
 CGTGCATATCTGTATAGC

>WP6049

ATGAGCGATAGTCATGGTGAAGTGTTCGCATTCTGGATATTCATGCAAGCGAAGTTGCACAGCATCCGG
 ATCTGGTTTCGTGAGATTTATGATGGTTCGTACCGATGGTGTATTATCCGTGAAGTTTTTGCACAGGATGC
 CATTTCGTAGCATTGTTTCATCGTCTGCGTAATGAACAGCTGGATTTTCCGACCGGTAGCTATGAATGGGGT
 CAGGTTTTATGGTTCGTGTTCTGTTTGCAGCAAAAACCGGATATGGATGGTTATCATCAGGATGCACAGAGCT
 TTCGTTCGTGCATGTAAAGAACTGTTTAGCAGCACCCGAGATTTTGAAGATAAAGTGAATGAAAATCTGGG
 TGCCTGGCAGCAGGTACACCGGTTTCGTGTTGCCGAAGTGGATGGTTCGTGCATTTGCACCCGGCAACCAT
 CGTTTTATGGAAAGCGGTTCATGGTCTGGCAATTCATCGTGGTAATCAGTTTATGAGCAA ACTGAGCCTGA
 TTCCGAGCCTGGCAGATACCGTTGATTATGAAAATCAGCTGAGCTATTTTGTGGTTCTGGATGAACCGGA
 AGTTGGTGGTGA ACTGGTGTGTATGAACTGGATCTGAGCGATCGTCTGCCGAACGTATTATGGCAAAT
 AGCAAACCTCAGCTGCTGCCGATGCCGGTGCATCTGCTGTTTTCAGGGTGGTTCGTATTTGGCATG
 AAGTTGCAAAAAGTTGAAGGTGAACGTAGCCGATTACCATTGGTGGTTTTTGCAGCATATACACCCGGATCA
 TCAAGA ACTGGTTTTGGTGGTCA

7.2.9.2. Expression vectors

Genes were cloned into the following expression vectors by the *In-Fusion* Cloning (LIC) method (Clontech).

Vector	Antibiotic resistance	Tag	Tag cleavage	Source	Addgene plasmid #
pOPINE	Ampicillin	C-terminal His ₆ -tag	none	OPPF	26143
pOPINF	Ampicillin	N-terminal His ₆ -tag	3C protease	OPPF	26142

7.2.9.3. Primers

100 µM primers stocks were ordered by the OPPF in 96-well plate format and were used to prepare plates at a working concentration of 10 µM. Primers are named according to the well IDs listed in Table 2.6. (Chapter 2), with ‘F’ annotations denoting forward primers and ‘R’ annotations denoting reverse primers.

A01_F	AAGTTCTGTTTCAGGGCCCCGACCACCGTTAGCGAACAGGCAAATTTTACC
A01_R	ATGGTCTAGAAAGCTTTAGCTCGGAATCTGACGCGGAACAAATTC
B01_F	AAGTTCTGTTTCAGGGCCCCGAGCAAAAAATTCGCACTGACCGCAG
B01_R	ATGGTCTAGAAAGCTTTAAACTGCTTCGAATTTTTTGGCCACGG
C01_F	AAGTTCTGTTTCAGGGCCCCGAAAAACAACACCCGCAACTTTAACATCGGC
C01_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGTAGATTTTTGTTGCGTTCTTTTC
D01_F	AAGTTCTGTTTCAGGGCCCCGATGAGCAATAATGCCGTTAGCACCAAAAGC
D01_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGTAGATTTTTGTTGCGTTCTTTTC
E01_F	AAGTTCTGTTTCAGGGCCCCGGTGACCGAGGTGAATAACTATCCGACCGC
E01_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGTAGATTTTTGTTGCGTTCTTTTC
F01_F	AAGTTCTGTTTCAGGGCCCCGACGGTATGATTATTCGTGGTGTCTGCCGC
F01_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGTAGATTTTTGTTGCGTTCTTTTC
G01_F	AAGTTCTGTTTCAGGGCCCCGCGAAAAGCGTTACCGCAAAGAAC
G01_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATAGATGGCATCGTGTCTTTTC
H01_F	AAGTTCTGTTTCAGGGCCCCGCATCCGAGCGCACTGCAGGATATG
H01_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATAGATGGCATCGTGTCTTTTC
A02_F	AAGTTCTGTTTCAGGGCCCCGGAAGTTCTGCCGAAAGATGTTATTGAACTGGTTG
A02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATAGATGGCATCGTGTCTTTTC
B02_F	AAGTTCTGTTTCAGGGCCCCGACCATTGCAGCACGTAACCCCTGTATAATC
B02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATACACTTTCTCGCGATTTTCG
C02_F	AAGTTCTGTTTCAGGGCCCCGTTTGAATACTGCCTCGTCATTATCCGGATGGC
C02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATACACTTTCTCGCGATTTTCG
D02_F	AAGTTCTGTTTCAGGGCCCCGATTGATGGCCTGATCATCGAAGATTTTCTG
D02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATACACTTTCTCGCGATTTTCG
E02_F	AAGTTCTGTTTCAGGGCCCCGATTTTCTGAGCCAGGATGAAGTTGATAATGTTG
E02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAATACACTTTCTCGCGATTTTCG

F02_F	AAGTTCTGTTTCAGGGCCCCGGCCTATGTGAAAAACGATCTGTGGATTACCG
F02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGAAGATTTTCTGGTTATCTTTGCTCAG
G02_F	AAGTTCTGTTTCAGGGCCCCGGCAAATTATCCGGATGCAGTTAACAGATTTAC
G02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGAAGATTTTCTGGTTATCTTTGCTCAG
H02_F	AAGTTCTGTTTCAGGGCCCCGGATGGCATGCTGATCCAGCAGGTTTTTAC
H02_R	ATGGTCTAGAAAGCTTTAGCTCCAATAGAAGATTTTCTGGTTATCTTTGCTC
A03_F	AAGTTCTGTTTCAGGGCCCCGAGTGATGCACAGACCAGCAAACCG
A03_R	ATGGTCTAGAAAGCTTTAGCTATACAGATATGCACGTTCCAGATCGCTGC
B03_F	AAGTTCTGTTTCAGGGCCCCGCTGGAAGTGCAGCGCACCGGTTG
B03_R	ATGGTCTAGAAAGCTTTAGCTATACAGATATGCACGTTCCAGATCGCTGC
C03_F	AAGTTCTGTTTCAGGGCCCCGAATGAACTGCCTGGTGTATTGTTTCGTGGTGC
C03_R	ATGGTCTAGAAAGCTTTAGCTATACAGATATGCACGTTCCAGATCGCTGC
D03_F	AAGTTCTGTTTCAGGGCCCCGGTGCATATACCGGTGAAGAAATGCAGCG
D03_R	ATGGTCTAGAAAGCTTTAGCTATACAGATATGCACGTTCCAGATCGCTGC
E03_F	AAGTTCTGTTTCAGGGCCCCGAGCGATAGTCATGGTGAAGTGTTCGC
E03_R	ATGGTCTAGAAAGCTTTATGACCACCAAACAGTTCTTGATGATCCGG
F03_F	AAGTTCTGTTTCAGGGCCCCGGCACAGCATCCGGATCTGGTTTCG
F03_R	ATGGTCTAGAAAGCTTTATGACCACCAAACAGTTCTTGATGATCCGG
G03_F	AAGTTCTGTTTCAGGGCCCCGGAAGTTTTTGCACAGGATGCCATTCCG
G03_R	ATGGTCTAGAAAGCTTTATGACCACCAAACAGTTCTTGATGATCC
H03_F	AGGAGATATACCATGAAAAACAACCCCGCAACTTTAACATCGGC
H03_R	GTGATGGTGTGTTTTGCTCCAATAGTAGATTTTGTGCGTTCTTTGC
A04_F	AGGAGATATACCATGATGAGCAATAATGCCGTTAGCACCAAAGC
A04_R	GTGATGGTGTGTTTTGCTCCAATAGTAGATTTTGTGCGTTCTTTGC
B04_F	AGGAGATATACCATGCCGAAAAGCGTTACCGCAAAGAACC
B04_R	GTGATGGTGTGTTTTGCTCCAATAATAGATGGCATCGTGTCTTTG
C04_F	AGGAGATATACCATGACCATTGCAGCACGTAAAACCCTGTATAATC
C04_R	GTGATGGTGTGTTTTGCTCCAATAATACACTTTCTCGCGATTTTTCG
D04_F	AGGAGATATACCATGGCCTATGTGAAAAACGATCTGTGGATTACCG
D04_R	GTGATGGTGTGTTTTGCTCCAATAGAAGATTTTCTGGTTATCTTTGCTCAG
E04_F	AGGAGATATACCATGAGTGATGCACAGACCAGCAAACCG
E04_R	GTGATGGTGTGTTTTGCTATACAGATATGCACGTTCCAGATCGCTGC
F04_F	AGGAGATATACCATGAGCGATAGTCATGGTGAAGTGTTCGC
F04_R	GTGATGGTGTGTTTTGACCACCAAACAGTTCTTGATGATCCGG

7.2.9.4. Protocol

Following OPPF-UK Standard Protocols: Cloning and Expression Screening (version: 16 April 2015).

Insert generation: Into a 96-well PCR microplate (on ice) was added 3 μ L forward primer (10 μ M), 3 μ L reverse primer (10 μ M), 2 μ L template plasmid (20 ng/ μ L) and 42 μ L master mix [Master mix: 875 μ L 2X Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher Scientific) + 595 μ L nuclease-free water (Invitrogen™) (35 reactions)].

The PCR reaction was run in a Veriti PCR machine as per the table below.

	Cycles	Temperature	Time
Pre-heat lid	1	Ramp to 98 °C	∞ (insert samples)
Denaturation	1	98 °C	10 sec
Denaturation	30	98 °C	1 sec
Annealing		60 °C	5 sec
Extension		72 °C	1 min
Extension	1	72 °C	2 min
Hold	1	4 °C	∞

To confirm reaction success, 5 µL PCR reaction was added to 2 µL 5X DNA loading buffer (0.25% w/v bromophenol blue in 30% glycerol) and run on a 1.6% TBE agarose gel for 30 min at 100V, alongside Hyperladder™ 1kb (Bioline). Gels were run on Electro-Fast® Stretch 108 (ABgene) systems to conserve the 8 x 12 well plate format.

PCR products were treated with DpnI (5 units DpnI in CutSmart® buffer per well) and incubated at 37 °C for 1 hr. The treated PCR products were then purified using AMPure XP magnetic beads (Beckman Coulter):

- 90 µL AMPure was added to each reaction with mixing and the plate incubated at room temperature for 5 min.
- The plate was placed onto a magnet (BILATEST™ M96 small volume plate magnet) and the supernatant aspirated and discarded. 200 µL 70% ethanol was added to each well and incubated at room temperature for 30 seconds before the ethanol was aspirated and discarded. This was repeated two more times and then the plate left to air-dry to evaporate any residual ethanol.
- The plate was removed from the magnet and 30 µL elution buffer (10 mM Tris pH 8.0) added to each well, with mixing. The plate was placed back

onto the magnet and 30 μ L supernatant transferred to a new 96-well plate. 5 μ L purified PCR reaction was then run on a gel as before to confirm successful insert purification.

Annealing & transformation:

1 μ L (100 ng) linearised pOPINX vector (pre-prepared stocks, OPPF) and purified insert (200 ng) were mixed and diluted to 10 μ L final volume with Milli-Q® water and transferred to an In-Fusion dry-down plate. After incubation at 42 °C for 30 min, the In-Fusion reactions were cooled on ice and diluted with 40 μ L TE. 3 μ L was immediately transformed into OmniMaxII cells (Invitrogen) and the transformants spread over 1 mL agar plates* (50 μ g/mL carbenicillin) which were incubated overnight at 37 °C.

* 1 mL antibiotic agar per well of a sterile 24-well tissue culture plate (Corning® Costar® #3524).

Preparation of DNA stocks:

Colonies were picked into 1.2 mL volumes of Power Broth™ (Molecular Dimensions) (50 μ g/mL carbenicillin) in 96 deep-well blocks, covered with a gas-permeable seal and incubated overnight at 37 °C (220 rpm). Cultures were harvested by centrifugation (6000 g, 10 min) and plasmid mini-preps performed on a Qiagen BioRobot 8000 (manufacturer's standard protocol).

Construct verification:

Constructs were verified by PCR with the relevant pOPINX primers. Into a 96-well PCR microplate (on ice) was added 1.5 μ L pOPINX reverse primer (10 μ M), 1.5 μ L plasmid and 22 μ L master mix [Master mix: 15 μ L pOPIN general forward primer (100 μ M) + 1250 μ L 2X Phusion Flash High-Fidelity PCR Master

Mix (ThermoFisher Scientific) + 935 μL nuclease-free water (Invitrogen™); 100 reactions].

Thermal cycling was performed in a Veriti PCR machine as per the table below.

	Cycles	Temperature	Time
Pre-heat lid	1	Ramp to 98 °C	∞ (insert samples)
Denaturation	1	98 °C	10 sec
Denaturation	30	98 °C	1 sec
Annealing		60 °C	5 sec
Extension		72 °C	1 min
Extension	1	72 °C	2 min
Hold	1	4 °C	∞

6.5 μL 5X DNA loading buffer (0.25% w/v bromophenol blue in 30% glycerol) was added to each PCR reaction well and 6.5 μL of the mix run on a 1.6% TBE agarose gel (as before). For each construct, at least one clone was found to produce a band of the desired size.

Expression trials:

Expression trials were set up using four different conditions resulting from a combination of two different cell types with two different induction conditions.

3 μL PCR-verified plasmid was transformed into Lemo™ 21(DE3) (NEB) and Rosetta™ 2(DE3) pLacI (Novagen) competent cells. The transformants were spread over 1 mL agar plates (50 $\mu\text{g}/\text{mL}$ carbenicillin + 35 $\mu\text{g}/\text{mL}$ chloramphenicol) and incubated overnight at 37 °C. Colonies were picked into 0.7 mL volumes of Power Broth™ (50 $\mu\text{g}/\text{mL}$ carbenicillin + 35 $\mu\text{g}/\text{mL}$ chloramphenicol) in 96 deep-well blocks, covered with a gas-permeable seal

and incubated overnight at 37 °C (220 rpm). Overnight cultures were diluted (150 µL for Lemo™ 21(DE3) or 250 µL for Rosetta™ 2(DE3) pLacI) into 24 deep-well blocks containing 3 mL media (50 µg/mL carbenicillin + 35 µg/mL chloramphenicol).

For IPTG induction, Power Broth™ medium was used - blocks were covered with a gas-permeable seal and incubated at 37 °C (220 rpm) for 3.5 hr before cooling the cultures at 20 °C (220 rpm) for 20 min. IPTG was added to a final concentration of 1 mM (6 µL 500 mM IPTG stock) and the cultures incubated overnight at 20 °C (220 rpm).

For auto-induction, Overnight Express™ Instant TB medium was used - blocks were covered with a gas-permeable seal and incubated at 37 °C (220 rpm) for 3.5 hr before reducing the shaker temperature to 25 °C and incubating for a further 24 hr.

1 mL culture volumes were transferred into 96 deep-well blocks, harvested by centrifugation (6000 g, 10 min) and frozen at -80 °C for a minimum of 30 min. The cell pellets were defrosted and the deep-well blocks transferred to a Qiagen BioRobot 8000 for protein isolation by small scale Ni²⁺-NTA purification (manufacturer's standard protocol) using magnetic Ni-NTA beads (Qiagen Ni-NTA Superflow 96 BioRobot Kit, Qiagen). Proteins were eluted in a 50 µL volume (elution buffer: 50 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole, 0.05% Tween, pH 8.0). Protein expression levels were analysed by SDS-PAGE on pre-cast 26-well gels to conserve the plate format (2 markers and 3 x 8 samples per gel). 10 µL eluted protein was added to 10 µL SDS-PAGE loading buffer (100 mM Tris pH 6.8, 4% w/v SDS, 0.2% w/v bromophenol blue and 20% glycerol)

and heated at 95 °C for 3 min. 10 µL sample was loaded per well and the gel run at 200V, alongside SigmaMarker Wide Range (MW 6.5–200 kDa) and SigmaMarker Low Range (MW 6.5–66 kDa) (Sigma-Aldrich). Gels were stained with InstantBlue™ (Expedeon).

7.2.10. Site-directed mutagenesis

Site-directed mutagenesis was performed with a Q5® Site-Directed Mutagenesis Kit (NEB), according to the manufacturer's protocol. Mutagenesis primers were designed using NEBaseChanger™ online primer design software (NEB) and obtained *via* Sigma-Aldrich custom oligo service. Primers were re-suspended in nuclease-free water (Invitrogen™) to give a standard concentration of 100 µM.

Mutagenesis primer sequences are listed here (named according to the specific mutation being performed; with 'F' annotations denoting forward primers and 'R' annotations denoting reverse primers):

Alanine scanning mutagenesis primers:

N74A_F	CCTGAACAAAGCTGAAGAGTTTGGCAC
N74A_R	ATGCTTTTCATACCACCATC
F77A_F	AAACGAAGAGGCTGGCACCAAAG
F77A_R	TTGTTTCAGGATGCTTTTC
V81A_F	TGGCACCAAAGCTGCCAGATTT
V81A_R	AACTCTTCGTTTTTGTTCAGGATGC
I84A_F	AGTTGCCAGGCTTATGGTCATGC
I84A_R	TTGGTGCCAAACTCTTCG
V90A_F	TCATGCAATTGCTGGTCAGTCTC
V90A_R	CCATAAATCTGGGCAACTTTG
G91A_F	TGCAATTGTTGCTCAGTCTCCGG
G91A_R	TGACCATAAATCTGGGCAAC
T151A_F	TACACCGGCAGCTATTTCGTCTGC
T151A_R	TAGGTCTGGCCTTCCGGA
R153A_F	GGCAACCATTGCTCTGCTGCTGGAAG
R153A_R	GGTGTATAGGTCTGGCCT
I161A_F	AGGTCGTGAAGCTGCAGTTCATGTTGG
I161A_R	TCCAGCAGCAGACGAATG
H164A_F	AATTGCAGTTGCTGTTGGTAATGATTTTCTGCTG
H164A_R	TCACGACCTTCCAGCAGC

F169A_F	TGGTAATGATGCTCTGCTGATGC
F169A_R	ACATGAACTGCAATTTTCAC
S189A_F	CGATCAGCTGGCTTATTTTATCCCGCTG
S189A_R	CTCAGATCCAGCAGGGTT
M221A_F	ATATGCACAGGCTCAAGAGTATATGGACGATGTC
M221A_R	TTGCTTGCCTCTTGCGGA
Y224A_F	GATGCAAGAGGCTATGGACGATGTC
Y224A_R	TGTGCATATTTGCTTGCC
M225A_F	GCAAGAGTATGCTGACGATGTCGAG
M225A_R	ATCTGTGCATATTTGCTTG
H259A_F	TCGTTATTATGCTCGTGTAGCGAAGTTATTG
H259A_R	CCACCATTAAACAGCAGC
R270A_F	TAATAGTCCGGCTCGCACCATTGG
R270A_R	CCAATAACTTCGCTAACAC
T272A_F	TCCGCGTCGCGCTATTGGTGGTT
T272A_R	CTATTACCAATAACTTCGCTAACACGG
F276A_F	CATTGGTGGTGCTCTGGCATTTAGC
F276A_R	GTGCGACGCGGACTATTA
S290A_F	CTATTATTGGGCTTAACAGTAAAGGTGGATAC
S290A_R	ATTTTGTGCGCGCTGTTTG

Additional single point variant mutagenesis primers:

V81N_F	TGGCACCAAAAACGCCAGATTTATG
V81N_R	AACTCTTCGTTTTTGTTCAG
I84N_F	AGTTGCCCAGAACTATGGTCATGC
I84T_F	AGTTGCCCAGACATATGGTCATGCAATTG
I84X_R	TTGGTGCCAAACTCTTCG
A88N_F	TTATGGTCATAACATTGTTGGTCAGTCTC
A88V_F	TTATGGTCATGTTATTGTTGGTCAGTC
A88X_R	ATCTGGGCAACTTTGGTG
V90G_F	TCATGCAATTGGCGGTCAGTCTCC
V90G_R	CCATAAATCTGGGCAACTTTG
I161F_F	AGGTCGTGAATTTGCAGTTCATG
I161H_F	AGGTCGTGAACATGCAGTTCATGTTGGTAATG
I161L_F	AGGTCGTGAACTTGCAGTTCATG
I161M_F	AGGTCGTGAAATGGCAGTTCATG
I161N_F	AGGTCGTGAAAATGCAGTTCATG
I161Q_F	AGGTCGTGAACAAGCAGTTCATGTTGGTAATG
I161S_F	AGGTCGTGAATCTGCAGTTCATGTTGG
I161T_F	AGGTCGTGAACTGCAGTTCATG
I161V_F	AGGTCGTGAAGTCGCAGTTCATGTTG
I161Y_F	AGGTCGTGAATATGCAGTTCATGTTGG
I161X_R	TCCAGCAGCAGACGAATG
G166D_F	AGTTCATGTTGACAATGATTTTCTGC
G166D_R	GCAATTTACGACCTTCC
F169V_F	TGGTAATGATGTACTGCTGATGC
F169V_R	ACATGAACTGCAATTTTCAC
Y224F_F	GATGCAAGAGTTTATGGACGATG
Y224F_R	TGTGCATATTTGCTTGCC

7.2.11. Sequencing

DNA sequences were confirmed by sequencing (Macrogen) with T7F and T7R universal primers.

T7F	TAATACGACTCACTATAGGG
T7R	GCTAGTTATTGCTCAGCGG

7.2.12. Protein expression

7.2.12.1. Expression strains

BL21 (DE3) chemically competent *E. coli* (NEB) were used for the expression of WelO5 and WelO5 variants.

7.2.12.2. Cell density determination

Optical densities (OD₆₀₀) of cell cultures were measured at 600 nm using 1 mL sample (1.5 mL cuvette) against a cell-free reference sample of growth media.

7.2.12.3. Expression trials

Six flasks containing 100 mL 2xTY media (50 µg/mL kanamycin) were each inoculated with 1 mL overnight starter culture and incubated at 37 °C (220 rpm). When OD₆₀₀ 0.6-0.8, the flasks were split evenly between three incubators set at 18 °C, 28 °C or 37 °C (220 rpm). After 20 min temperature equilibration, one flask from each pair was induced with 0.5 mM IPTG. At both 4 hr and 20 hr time points, the OD₆₀₀ was measured and 35 mL culture removed and the cells harvested by centrifugation (6000 g, 10 min, 4 °C). 1 mL cell culture from each time point was also pelleted for SDS-PAGE analysis. Samples were stored at -80 °C.

7.2.12.4. Large scale protein expression**7.2.12.5. Batch method (12L scale)**

100 mL 2xTY media (50 µg/mL kanamycin) was inoculated with 1 mL starter culture and incubated at 37 °C (220 rpm). After 4 hr, 5 mL was added to each flask of 1L 2xTY media (50 µg/mL kanamycin) and incubated at 37 °C (220 rpm) until OD₆₀₀ 0.6-0.8. The flasks were cooled to 30 °C for 20 min before induction with 0.5 mM IPTG. The induced cultures were then incubated at 30 °C (220 rpm) for 20 hr before being harvested by centrifugation (6000 g, 10 min, 4 °C) and frozen at -80 °C.

7.2.12.6. Fed-batch method (5L scale)

This method is designed to increase expression yields by supplementing the cells with extra nutrients, prolonging the log phase. Nutrients are added in the form of a buffer to achieve a dilution of approximately 1/20 within the growth media.

	Fed-Batch Buffer Stock (pH 7.0-7.4)	Fed-Batch Buffer Final Conc.
MOPS Buffer	1M	50 mM
MgCl ₂	20 mM	1 mM
MgSO ₄	20 mM	1 mM
glycerol	40%	2%

A loopful of colonies from a freshly prepared agar plate was used to inoculate 10 mL 2xTY media (30 µg/mL kanamycin) and incubated at 37 °C (220 rpm) for 6 hr. The entire starter culture was then added to 400 mL 2xTY media (30 µg/mL kanamycin) and incubated at 37 °C (220 rpm) overnight. The OD₆₀₀ of the overnight culture was measured and a sufficient amount transferred into flasks containing 450 mL 2xTY media (30 µg/mL kanamycin) to give a final OD₆₀₀ of at least 0.2. The flasks were incubated at 37 °C (220 rpm) until the

average OD₆₀₀ was around 3.5-4 and then cooled to 17 °C (220 rpm) for 20 min. 300 µM IPTG (150 µL 1M stock) and 25 mL fed-batch buffer stock were added to each flask before incubation at 17 °C (220 rpm) overnight. A final OD₆₀₀ was measured (9-10) and the cells harvested by centrifugation (6000 g, 10 min, 4 °C), before storing at -80 °C.

7.2.12.7. Expression of selenomethionine WelO5 (2.4L scale)

Selenomethionine WelO5 was expressed under the conditions used in Section 7.2.12.5., using a SelenoMet Medium Complete™ kit (Molecular Dimensions, MD12-500) for the preparation of growth media.

- 43.2 g SelenoMet Medium Base™ was diluted up to 2.4 L with Milli-Q® water, divided between three shake flasks and autoclaved.
- 12.24 g SelenoMet Nutrient Mix™ was diluted up to 120 mL with Milli-Q® water and filter-sterilised.

BL21 (DE3) cells (NEB) were used for expression (an auxotrophic strain was not necessary). A single colony was picked from a freshly streaked agar plate into 10 mL 2xTY media (50 µg/mL kanamycin) and incubated for 5 hr at 37 °C (220 rpm). 2.5 mL culture was diluted into each one of three shake flasks containing 80 mL 2xTY media (50 µg/mL kanamycin) and incubated at 37 °C (220 rpm) overnight. The overnight cultures were centrifuged (6000 g, 10 min), the media decanted and the cell pellets re-suspended in 8 mL Milli-Q® water (repeated x 3).

Three shake flasks were prepared as follows:

	Per flask:
SelenoMet Medium Base™	800 mL
SelenoMet Nutrient Mix™	40 mL
Selenomethionine solution (250X)	3.2 mL
re-suspended cell pellet	8 mL
50 µg/mL kanamycin	800 µL

The flasks were incubated at 37 °C (220 rpm) until an OD₆₀₀ of 1 was reached, then cooled to 30 °C for 20 min before induction with 0.5 mM IPTG. The induced cultures were incubated at 30 °C (220 rpm) for 20 hr before being harvested by centrifugation (6000 g, 10 min, 4 °C) and frozen at -80 °C. Purification according to Section 7.2.13.2. yielded 3 mL protein at 35 mg/mL (approx. 1 mM) which was used to prepare crystallisation trial plates.

7.2.12.8. Expression of isotopically labelled WelO5

7.2.12.8.1. ¹⁵N,²H-labelled WelO5 (2L scale)

2L of ¹⁵N,²H-media was prepared as per the table, using (¹⁵NH₄)₂SO₄ and D₂O as sources of ¹⁵N and ²H respectively (pH 7.4; no pH adjustment necessary). A kanamycin stock solution was added to give a 50 µg/mL working concentration and the media filter-sterilised before use.

Labelling Media Recipe	per 1L
$(^{15}\text{NH}_4)_2\text{SO}_4/(\text{NH}_4)_2\text{SO}_4$	1 g
$\text{Na}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$	6.8 g
KH_2PO_4	3 g
NaCl	500 mg
^{13}C -glucose/glucose	4 g
EDTA	50 mg
$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$	16 mg
FeCl_3	5 mg
ZnCl_2	50 μL 10 mg/mL stock
CuCl_2	10 μL 10 mg/mL stock
CoCl_2	10 μL 10 mg/mL stock
H_3BO_3	20 μL 5 mg/mL stock
d-Biotin	1 mg
Thiamine	1 mg
Na_2SO_4	42.6 mg
MgSO_4	120.4 mg
$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	65.7 mg
$\text{D}_2\text{O}/\text{H}_2\text{O}$	dilute up to 1L

N.B. Depending on the required isotopic labelling ($(^{15}\text{NH}_4)_2\text{SO}_4$ (Sigma-Aldrich, 299286), ^{13}C -glucose (Sigma-Aldrich, 389374) and D_2O were substituted for $(\text{NH}_4)_2\text{SO}_4$, glucose and water respectively.

One colony was picked from a pNIC28-Bsa4[WelO5] transformation plate into 50 mL LB media (50 $\mu\text{g}/\text{mL}$ kanamycin) and incubated at 37 °C (220 rpm) overnight. A sufficient volume was seeded into 4 x 500 mL $^{15}\text{N}, ^2\text{H}$ -media to give an OD_{600} of 0.1. The cultures were incubated at 37 °C (220 rpm) until an average OD_{600} of 0.8 was reached and then the incubator temperature dropped to 30 °C. After 30 min the cells were induced with 0.5 mM IPTG. At 20 hr post-induction, the cultures were harvested by centrifugation (6000 g, 10 min) and stored at -80 °C. Purification according to Section 7.2.13.2. yielded 10 mL protein at 13 mg/mL (approx. 400 μM).

7.2.12.8.2. ¹⁵N, ¹³C, ²H-labeled WelO5 (4L scale)

4L of ¹⁵N, ¹³C, ²H-media was prepared as per the table, using (¹⁵NH₄)₂SO₄, ¹³C-glucose and D₂O as sources of ¹⁵N, ¹³C and ²H respectively (pH 7.2; no pH adjustment necessary). 200 mL unlabelled media was prepared in the same manner (adjusted to pH 7.2 with 6M NaOH). 300 µL 50 µg/mL kanamycin stock solution was added (per litre) to give a 15 µg/mL working concentration. Media was filter-sterilised before use.

30% and 70% labelled D₂O media solutions were prepared as follows:

	D₂O media (mL)	Unlabelled media (mL)	Final Volume (mL)
30% D ₂ O media	15	35	50
70% D ₂ O media	175	75	250

Five colonies were picked from a pNIC28-Bsa4[WelO5] transformation plate into 5 x 1 mL LB media (30 µg/mL kanamycin) and incubated at 37 °C (220 rpm) for 8 hr. The cultures were pooled and 1 mL seeded into 5 x 10 mL 30% D₂O media before incubating at 37 °C (220 rpm). After 3 hr, an average OD₆₀₀ of 2 was measured. The cultures were pooled and 1 mL seeded into 5 x 50 mL 70% D₂O media before incubating at 37 °C (220 rpm) overnight. An average OD₆₀₀ of 4 was measured and the cultures pooled. A sufficient volume was seeded into 8 x 500 mL 100% D₂O media to give an OD₆₀₀ of 0.1. The cultures were incubated at 37 °C (220 rpm) until an average OD₆₀₀ of 0.6 was reached. The incubator temperature was dropped to 30 °C and after 1 hr an average OD₆₀₀ of 0.8 was reached. The cells were induced with 0.5 mM IPTG and incubated at 30 °C (220 rpm). At 20 hr post-induction, the cultures possessed an average OD₆₀₀ of 5 and were harvested by centrifugation (6000 g, 10 min).

The cell pellets were re-suspended in PBS, centrifuged again (6000 g, 10 min, 4 ° C) and the media decanted before storing at -80 ° C.

7.2.13. Protein purification

All large scale automated purification was carried out on ÄKTA™ FPLC systems (GE Healthcare) at 4 ° C.

7.2.13.1. Batch purification

7.2.13.1.1. Nickel-Affinity Chromatography (IMAC)

25 g cell pellet (3L culture) was suspended in 100 mL buffer A with 1 mg DNase and 100 mM PMSF and stirred at 4 ° C until a smooth suspension was formed. This was sonicated over ice for total of 8 min (2 x 4 min; 60% amplification; 9.9 sec on, 9.9 sec off). The lysate was clarified by centrifugation (40,000 g, 30 min, 4 ° C), filtered (0.45 µM) and loaded onto a 5 mL HisTrap™ FF column (GE Healthcare) at 1 mL/min. The method was run at 1 mL/min, starting with a column wash of 5 CV 5 mM imidazole (100% buffer A) and 15 CV 25 mM imidazole (96% buffer A) before eluting the protein in 20 CV 500 mM imidazole (100% buffer B).

	Imidazole	Tris	NaCl	TCEP
Buffer A (pH 7.5)	5 mM	50 mM	500 mM	1 mM
Buffer B (pH 7.5)	500 mM	50 mM	500 mM	1 mM

N.B. TCEP added just prior to running the purification.

The desired fractions (SDS-PAGE) were combined and concentrated into 10 mL TEV protease cleavage buffer (25 mM HEPES, 500 mM NaCl, 5% glycerol and 0.5 mM TCEP, pH 7.5) using a Millipore Amicon® Ultra 10 kDa, 15 mL (according to the manufacturer's instructions). 500 µL TEV protease (3 mg/mL) was added and the protein rotated at 4 ° C overnight. The cleaved protein was

manually loaded onto a 5 mL HisTrap™ FF column pre-equilibrated with 5 mM imidazole (buffer A) and washed with 25 mM and 500 mM imidazole buffers, collecting 5 mL fractions. Cleaved WelO5 eluted in the 5 mM imidazole flow through, which was concentrated to 2 mL prior to purification by size exclusion.

7.2.13.1.2. Size-exclusion chromatography (SEC)

A Superdex® 75-300 mL column was equilibrated with 1.1 CV of gel filtration buffer (filtered and degassed prior to use) at 2 mL/min.

	Tris	NaCl	TCEP
Gel Filtration Buffer (pH 7.5)	50 mM	500 mM	1 mM

The protein was loaded using a 2 mL injection loop and eluted isocratically at 2 mL/min. The clean fractions were concentrated (Millipore Amicon® Ultra 10 kDa, 15 mL) to approximately 4 mL at 35 mg/mL, giving a total expression yield of 45 mg/L.

7.2.13.1.3. Condensed purification protocol

Two cell pellets (2L culture) were suspended in 50 mL buffer A with a Pierce™ Protease Inhibitor mini-tablet (EDTA-free). This was stirred at 4 °C until a smooth suspension was formed and then passed twice through a cell disruptor (35 psi, 4.5 °C). The lysate was clarified by centrifugation (40,000 g, 30 min, 4 °C) and tumbled (2 hr, 4 °C) with 60 mL Ni-NTA superflow beads (Qiagen) before loading the column. The method was run at 5 mL/min, starting with a column wash of 5 CV 10 mM imidazole (98% buffer A), before eluting with 5 CV 400 mM imidazole (20% buffer A) and 2 CV of 500 mM imidazole (100% buffer B).

	Imidazole	PBS	NaCl	TCEP
Buffer A (pH 7.5)	-	10 mM	500 mM	1 mM
Buffer B (pH 7.5)	500 mM	10 mM	500 mM	1 mM

The desired fractions (SDS-PAGE) were pooled and concentrated to 150 mL (Millipore Amicon® Plus-70, 10 kDa). 500 µL of TEV protease (3 mg/mL) was added and the reaction incubated for 2hr at room temperature before size exclusion chromatography. A Superdex® 200-320 mL was equilibrated with 1.1 CV of gel filtration buffer (filtered and degassed prior to use).

	HEPES	NaCl
Gel Filtration Buffer (pH 7.4)	20 mM	150 mM

The TEV cleavage reaction mixture was purified in four injections, eluting isocratically at 3 mL/min. The desired fractions were pooled and concentrated (Millipore Amicon® Plus-70, 10 kDa) to approximately 43 mL at 3.5 mg/mL.

7.2.14. Protein quantification

Protein concentration was determined using a ThermoFisher Scientific NanoDrop™ ND-1000 UV-Vis spectrophotometer at 280 nm using the estimated extinction coefficient (ExpASy ProtParam)¹⁵⁴ and calculated molecular weight.

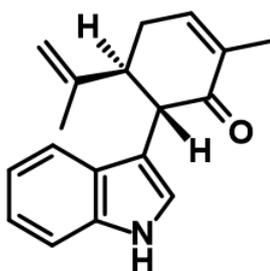
	⁶His-WelO5	Cleaved WelO5
Molecular Weight (Da)	34991.5	32525.8
Molar extinction coefficient (M ⁻¹ cm ⁻¹)	31860	30370

7.3. Chemical synthesis

General procedures: All analytical grade solvents and commercially available reagents were used without further purification. Commercially available anhydrous solvents were used when required. Reactions were performed under nitrogen unless otherwise stated and monitored by TLC on aluminium backed Merck KGaA 60-F₂₅₄ silica gel plates. Visualisation was achieved with UV light, or with potassium permanganate or ceric ammonium molybdate stains for low/non-UV active compounds. Purifications were performed on a Biotage® Isolera™ Prime system using Biotage® SNAP Ultra cartridges. Preparative HPLC and SFC purifications were performed by Justin Staniforth (UCB). All isolated final compounds were freeze-dried from an acetonitrile-water mixture to aid aqueous solubility in subsequent enzyme activity assays. Specific rotations were collected on an Optical Activity polAAR 2001 polarimeter. IR spectra were recorded on a ThermoFisher Scientific Nicolet™ iS5™ FT-IR Spectrometer fitted with an iD5 ATR accessory. ¹H (400 MHz) and ¹³C NMR (100 MHz) spectra were recorded on a Bruker AVIII 400 MHz spectrometer, fitted with a BBFO room temperature probe. Chemical shifts are reported in ppm, using the residual solvent signal as an internal reference. The multiplicity of the signal is indicated as s (singlet), br.s (broad singlet), d (doublet), t (triplet), q (quartet), m (multiplet), dd (doublet of doublets), dt (doublet of triplets) etc. Coupling constants (*J*) are quoted to the nearest 0.1 Hz. The LC-MS purity of UV active compounds was obtained using both acidic (pH3) and basic (pH10) methods on an Agilent 1200RR LC system fitted with a DAD UV detector and a Waters X-Bridge C18 column (2.1 x 20 mm, 2.5 μm), coupled to a 6140 single quadrupole mass spectrometer using dual mode ESI + APCI ionisation. HRMS results were generated by Victoria Ellis and Adam Hold (UCB). HRMS data was collected on a Waters ACQUITY® UPLC system fitted with a DAD UV detector and a Waters

ACQUITY® BEH C18 column (2.1 x 50mm, 1.7 μm), coupled to a Waters Xevo G2 QToF mass spectrometer using ESI ionisation.

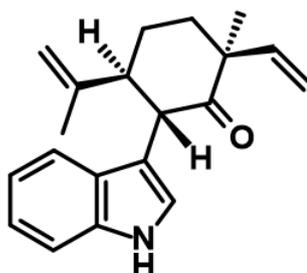
7.3.1. Synthesis of (+)-12-*epi*-fischerindole U isonitrile and analogues



(5*S*,6*S*)-6-(1*H*-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one (16)

Indole (66.5 mmol, 7.87 g) and (*S*)-carvone (33.3 mmol, 5.42 mL) were combined and azeotroped with toluene. The resulting residue was dissolved in THF (200 mL), cooled to -78°C and a solution of LiHMDS (1M in THF, 100 mmol, 100 mL) added dropwise *via* cannula. After 30 min, copper(II) 2-ethylhexanoate (49.7 mmol, 17.4 g) was added directly to the solution (flask open to air) and the reaction left to stir at -78°C overnight with the rubber septum replaced. The reaction was permitted to warm to room temperature, quenched by the addition of 1M HCl and partitioned with EtOAc. The layers were separated and the organic phase washed sequentially with 1M HCl, 1M NaOH, water and brine before being dried and concentrated *in vacuo*. The crude material was purified by flash column chromatography (10-40% EtOAc/hexane) and triturated in the minimal volume of ether to give the *title compound*, 4.98 g as a beige solid (53%). $R_f = 0.27$ (30% EtOAc/hexane); $[\alpha]_D^{25} = +56$ (DCM, c 3.9); IR ν_{max} 3405, 2915, 1678, 1644, 1456, 738 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 8.00 (br.s, 1H), 7.45 (d, $J = 7.8$ Hz, 1H), 7.33 (d, $J = 8.1$ Hz, 1H), 7.16 (t, $J = 7.5$ Hz, 1H), 7.07 (t, $J = 7.5$ Hz, 1H), 6.91 (d, $J = 2.4$ Hz, 1H), 6.84 - 6.75 (m, 1H), 4.67 (s, 1H), 4.66 - 4.63 (m, 1H), 3.92 (d, $J = 10.7$ Hz, 1H), 3.32 - 3.22 (m, 1H), 2.64 - 2.51 (m, 1H), 2.53 -

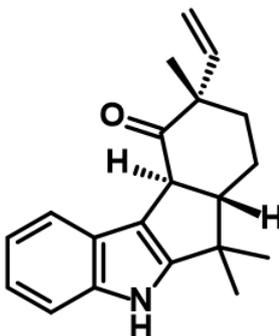
2.40 (m, 1H), 1.87 (s, 3H), 1.63 (s, 3H). ^{13}C NMR (100 MHz, CDCl_3) δ 199.5, 146.1, 143.7, 136.4, 135.6, 127.2, 122.8, 121.9, 119.4, 119.3, 112.9, 112.8, 111.5, 49.3, 48.5, 31.0, 19.5, 16.4; LC-MS (pH3) RT 2.32 min, 100.0%; LC-MS (pH10) RT 2.41 min, 100.0 %; HRMS (ESI) calc. $\text{C}_{18}\text{H}_{20}\text{NO}$ $[\text{M}+\text{H}]^+$ 266.1545, found 266.1523.



**(2S,5S,6S)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone
(18)**

(5S,6S)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one (5.73 mmol, 1.52 g) was azeotroped with toluene. The resulting residue was dissolved in THF (80 mL), cooled to -78°C and a solution of LiHMDS (1M in THF, 6.0 mmol, 6.0 mL) added dropwise. After 20 min, a solution of L-selectride (1M in THF, 2.97 mmol, 2.97 mL) was added and the reaction stirred for an additional hour. Acetaldehyde (34.4 mmol, 1.9 mL) was added and stirred for 30 min at -78°C before quenching by the dropwise addition of a solution of 30% H_2O_2 (20 mL) and 2M NaOH (30 mL) (reaction allowed to warm to room temperature during the quench to aid solubility). After stirring for 5 hr, the reaction was extracted with EtOAc. The combined organic extracts were washed with water and brine before drying and concentrating *in vacuo*. The residue was azeotroped with toluene, concentrated *in vacuo* and dissolved in CHCl_3 (60 mL). Martin sulfurane (6.33 mmol, 4.30 g) was added and after 15 min, the solvent was removed. The crude material was purified by flash column chromatography (0-40% EtOAc/hexane) and then by reverse phase

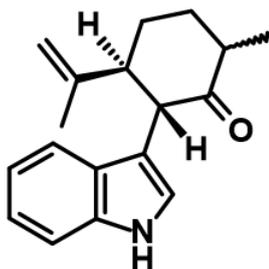
column chromatography (0-100% acetonitrile/water) to give the *title compound*, 1.02 g as a beige solid (61% over two steps). $R_f = 0.30$ (25% EtOAc/hexane); $[\alpha]_D = -128$ (DCM, c 1.3); IR ν_{\max} 3347, 2928, 1700, 1459, 741 cm^{-1} ; $^1\text{H NMR}$ (400 MHz, CDCl_3) δ 8.00 (br.s, 1H), 7.37 - 7.29 (m, 2H), 7.15 (t, $J = 7.6$ Hz, 1H), 7.05 (t, $J = 7.5$ Hz, 1H), 6.94 (d, $J = 2.4$ Hz, 1H), 6.16 (dd, $J = 17.7, 10.7$ Hz, 1H), 5.33 (d, $J = 10.7$ Hz, 1H), 5.18 (d, $J = 17.7$ Hz, 1H), 4.62 (s, 1H), 4.57 - 4.51 (m, 1H), 4.29 (d, $J = 12.5$ Hz, 1H), 2.91 (td, $J = 12.3, 3.4$ Hz, 1H), 2.26 - 2.10 (m, 2H), 1.88 - 1.74 (m, 2H), 1.54 (s, 3H), 1.21 (s, 3H); $^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 210.9, 146.9, 143.1, 136.2, 127.7, 123.3, 121.7, 119.3, 119.0, 116.2, 112.1, 111.9, 111.3, 53.6, 52.4, 48.9, 39.3, 28.8, 25.1, 18.5; LC-MS (pH3) RT 2.76 min, 100.0 %; LC-MS (pH10) RT 2.78 min, 99.1 %; HRMS (ESI) calc. $\text{C}_{20}\text{H}_{24}\text{NO}$ $[\text{M}+\text{H}]^+$ 294.1828, found 294.1858.



(6a*S*,9*S*,10a*S*)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one (19)

(2*S*,5*S*,6*S*)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-2-vinyl-cyclohexanone (4.66 mmol, 1.38 g) was dissolved in DCM (45 mL) and cooled to 0°C. Trimethylsilyl trifluoromethanesulfonate (13.8 mmol, 2.50 mL) was added, followed by MeOH ((5.18 mmol, 210 μL) and the reaction left to stir for 1 hr at 0°C before quenching with a 5% solution of sodium bicarbonate. The reaction was passed through a phase separator and the organic layer reduced *in vacuo*. The aqueous layer was extracted with EtOAc and the combined organic extracts

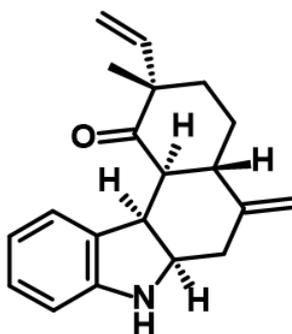
washed with water and brine before being dried, filtered and reduced under vacuum and combined with the previously collected material. The crude material was purified by flash column chromatography (0-30% EtOAc/hexane) to give the *title compound* (225 mg, 17%) as a beige solid and 928 mg (67%) recovered starting material. N.B. The starting material was recycled through the same process to increase the overall product yield. $R_f = 0.34$ (20% EtOAc/hexane); $[\alpha]_D = +56$ (DCM, c 1.3); IR ν_{\max} 3366, 2957, 1708, 1447, 758 cm^{-1} ; $^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.87 (br. s, 1H), 7.74 – 7.67 (m, 1H), 7.34 – 7.23 (m, 1H), 7.14 – 7.07 (m, 2H), 6.07 (dd, $J = 17.6, 10.7$ Hz, 1H), 5.23 (d, $J = 17.6$ Hz, 1H), 5.21 (d, $J = 10.7$ Hz, 1H), 4.04 (d, $J = 12.0$ Hz, 1H), 2.38 (td, $J = 12.3, 2.8$ Hz, 1H), 2.33 – 2.27 (dq, $J = 13.7, 2.3$ Hz, 1H), 1.98 (qd, $J = 12.7, 3.8$ Hz, 1H), 1.83 – 1.73 (m, 1H), 1.70 (td, $J = 13.2, 4.5$ Hz, 1H), 1.37 (s, 3H), 1.25 (s, 3H), 1.13 (s, 3H); $^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 210.8, 151.3, 143.1, 139.8, 124.6, 121.1, 120.3, 115.3, 113.4, 111.5, 64.7, 52.8, 52.5, 41.3, 40.5, 25.3, 24.3, 22.3, 20.6; LC-MS (pH3) RT 2.30 min, 98.0 %; LC-MS (pH10) RT 2.25 min, 98.5 %; HRMS (ESI) calc. $\text{C}_{20}\text{H}_{24}\text{NO}$ $[\text{M}+\text{H}]^+$ 294.1858, found 294.1860.



(2S,3S)-2-(1H-indol-3-yl)-3-isopropenyl-6-methyl-cyclohexanone (21)

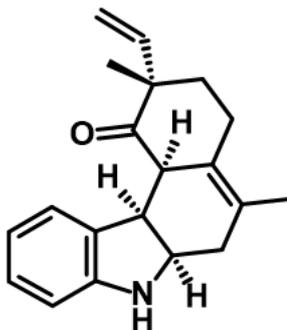
(5S,6S)-6-(1H-indol-3-yl)-5-isopropenyl-2-methyl-cyclohex-2-en-1-one (3.77 mmol, 1.00 g) was azeotroped with toluene. The resulting residue was dissolved in THF (55 mL), cooled to -78°C and a solution of LiHMDS (1M in THF, 5.65 mmol, 5.65 mL) added dropwise. After 20 min, a solution of L-selectride (1M in THF, 3.96 mmol, 3.96 mL) was added and the reaction stirred for an

additional hour. The reaction was quenched by the dropwise addition of a solution of 30% H₂O₂ (13 mL) and 2M NaOH (18 mL) (reaction allowed to warm to room temperature during the quench to aid solubility). After stirring for 5 hr, the reaction was extracted with EtOAc. The combined organic extracts were washed with water and brine before drying and concentrating *in vacuo*. This was purified by flash column chromatography (10-40% EtOAc/hexane) and then further purified by prep. HPLC to give the *title compound*, 642 mg (58%) as a white solid in an inseparable mixture of diastereoisomers (6:1). $R_f = 0.27$ (25% EtOAc/hexane); $[\alpha]_D = -49$ (DCM, c 0.05); IR ν_{\max} 3354, 2922, 1709, 1457, 740 cm^{-1} ; ¹H NMR (400 MHz, CDCl₃) (major) δ 8.00 (br.s, 1H), 7.39 (d, $J = 7.9$ Hz, 1H), 7.32 (d, $J = 8.1$ Hz, 1H), 7.14 (d, $J = 7.5$ Hz, 1H), 7.06 (d, $J = 7.9$ Hz, 1H), 6.96 (d, $J = 2.4$ Hz, 1H), 4.60 (s, 1H), 4.56 - 4.50 (m, 1H), 3.93 (d, $J = 12.4$ Hz, 1H), 2.90 (td, $J = 12.1, 3.8$ Hz, 1H), 2.72 - 2.63 (m, 1H), 2.27 - 2.19 (m, 1H), 2.07 - 1.94 (m, 2H), 1.67 - 1.56 (m, 1H), 1.56 (s, 3H), 1.10 (d, $J = 6.5$ Hz, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 211.0, 146.7, 136.2, 127.6, 123.3, 121.7, 119.3, 119.2, 112.1, 111.7, 111.3, 54.1, 52.5, 45.5, 35.8, 32.4, 18.7, 15.0; LC-MS (pH3) RT 2.48 min, 99.6 %; LC-MS (pH10) RT 2.44 min, 99.2 %; HRMS (ESI) calc. C₁₈H₂₂NO [M+H]⁺ 268.1701, found 268.1682.



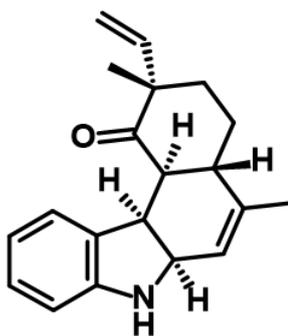
(2*S*,4*aS*,6*aR*,11*bR*,11*cS*)-2-methyl-5-methylene-2-vinyl-3,4,4*a*,6,6*a*,7,11*b*,11*c*-octahydrobenzo[*c*]carbazol-1-one (22)

The *title compound* was isolated as a by-product from the previous acid-catalysed cyclisation reaction, (53 mg, 4%). $R_f = 0.26$ (20% EtOAc/hexane); $[\alpha]_D = -226$ (DCM, c 0.52); IR ν_{\max} 3359, 2925, 1698, 1605, 1479, 750 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.02 (t, $J = 7.6$ Hz, 1H), 6.89 (d, $J = 7.5$ Hz, 1H), 6.69 (t, $J = 7.5$ Hz, 1H), 6.65 (d, $J = 7.7$ Hz, 1H), 5.90 (dd, $J = 17.6, 10.7$ Hz, 1H), 5.16 (d, $J = 10.7$ Hz, 1H), 5.02 (d, $J = 17.6$ Hz, 1H), 4.79 (s, 1H), 4.73 (s, 1H), 4.02 (dd, $J = 7.6, 3.6$ Hz, 1H), 3.92 – 3.72 (m, 1H), 2.72 (dd, $J = 12.8, 3.7$ Hz, 1H), 2.40 (dd, $J = 13.5, 5.4$ Hz, 1H), 2.37 – 2.30 (m, 1H), 2.09 (dt, $J = 13.6, 3.2$ Hz, 1H), 2.02 (dd, $J = 13.5, 8.2$ Hz, 1H), 1.96 – 1.86 (m, 1H), 1.85 – 1.79 (m, 1H), 1.73 (td, $J = 13.5, 3.5$ Hz, 1H), 1.28 (s, 3H). ^{13}C NMR (100 MHz, CDCl_3) δ 211.1, 151.5, 148.0, 142.9, 128.7, 127.5, 126.6, 119.1, 115.7, 110.6, 108.2, 60.52, 51.8, 42.1, 41.9, 39.1, 37.2, 25.2, 25.0; LC-MS (pH3) RT 2.14 min, 100.0 %; LC-MS (pH10) RT 2.17 min, 100.0 %; HRMS (ESI) calc. $\text{C}_{20}\text{H}_{24}\text{NO}$ $[\text{M}+\text{H}]^+$ 294.1858, found 294.1860.



(2*S*,6*aR*,11*b**R*,11*c**R*)-2,5-dimethyl-2-vinyl-4,6,6*a*,7,11*b*,11*c*-hexahydro-3*H*-benzo[*c*]carbazol-1-one (23)**

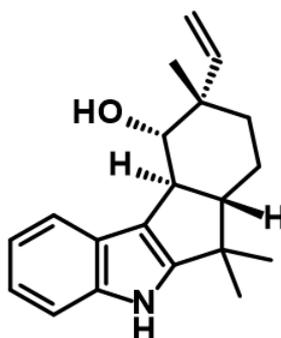
The *title compound* was isolated as a by-product from the previous acid-catalysed cyclisation reaction, (30 mg, 2%). Unfortunately, attempted purification by preparative HPLC at pH 10 resulted in complete degradation of this compound. Minimal analysis of the material performed prior to preparative HPLC. ¹H NMR (300 MHz, CDCl₃) δ 7.08 – 6.95 (m, 2H), 6.70 (td, *J* = 7.5, 0.9 Hz, 1H), 6.65 (d, *J* = 7.7 Hz, 1H), 5.89 (dd, *J* = 17.6, 10.7 Hz, 1H), 5.08 (d, *J* = 10.7 Hz, 1H), 4.97 (d, *J* = 17.6 Hz, 1H), 4.03 – 3.95 (m, 1H), 3.95 – 3.88 (m, 1H), 3.35 (br. s, 1H), 2.57 – 2.35 (m, 2H), 2.28 (m, 1H), 2.17 – 2.05 (m, 2H), 1.67 (br. s, 3H), 1.63 – 1.53 (m, 1H), 1.20 (s, 3H); LC-MS (pH10) RT 2.70 min, 82.3 %.



(2*S*,4*aS*,6*a**R*,11*b**R*,11*c**S*)-2,5-dimethyl-2-vinyl-4,4*a*,6*a*,7,11*b*,11*c*-hexahydro-3*H*-benzo[*c*]carbazol-1-one (24)**

The *title compound* was isolated as a by-product from the previous acid-catalysed cyclisation reaction, (56 mg, 4%). *R_f* = 0.24 (20% EtOAc/hexane);

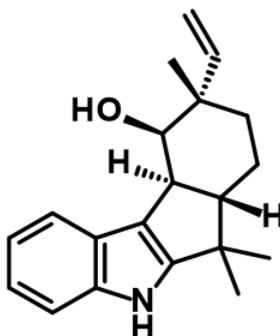
$[\alpha]_D = -178$ (DCM, c 0.65); IR ν_{\max} 3372, 2934, 1697, 1601, 1480, 751 cm^{-1} ;
 ^1H NMR (400 MHz, CDCl_3) δ 7.00 (t, $J = 7.6$ Hz, 1H), 6.91 (d, $J = 7.4$ Hz, 1H), 6.68 (t, $J = 7.4$ Hz, 1H), 6.63 (d, $J = 7.7$ Hz, 1H), 5.94 (dd, $J = 17.6, 10.7$ Hz, 1H), 5.18 (d, $J = 10.7$ Hz, 1H), 5.13 (s, 1H), 5.07 (d, $J = 17.6$ Hz, 1H), 4.23 - 4.15 (m, 1H), 4.02 (d, $J = 7.7$ Hz, 1H), 2.93 (dd, $J = 12.0, 2.6$ Hz, 1H), 2.35 (t, $J = 11.5$ Hz, 1H), 2.19 - 2.09 (m, 1H), 2.05 - 1.98 (m, 1H), 1.73 - 1.65 (m, 1H), 1.62 - 1.55 (m, 1H), 1.61 (d, $J = 1.2$ Hz, 3H), 1.26 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 211.6, 151.4, 142.69, 139.7, 128.6, 127.4, 126.4, 125.3, 119.2, 115.7, 110.9, 59.3, 51.6, 50.5, 40.7, 39.8, 38.3, 25.4, 24.9, 20.9; LC-MS (pH3) RT 2.04 min, 92.8 %; LC-MS (pH10) RT 2.12 min, 96.0 %; HRMS (ESI) calc. $\text{C}_{20}\text{H}_{24}\text{NO}$ $[\text{M}+\text{H}]^+$ 294.1858, found 294.1873.



(6a*S*,9*S*,10*S*,10a*S*)-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1-b]indol-10-ol (25)

(6a*S*,9*S*,10a*S*)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one (0.48 mmol, 140 mg) was dissolved in MeOH (2 mL) and sodium borohydride (0.72 mmol, 28 mg) added. After stirring for 2 hr, the reaction was diluted with saturated aqueous ammonium chloride and extracted with EtOAc. The combined organic extracts were washed with brine before drying and concentrating *in vacuo*. The crude material (as a mixture of isomers) was purified by flash column chromatography (5-15% EtOAc/hexane) to give the *title compound* (80 mg, 57%) as the major isomer. $R_f = 0.43$ (DCM); $[\alpha]_D = -34$

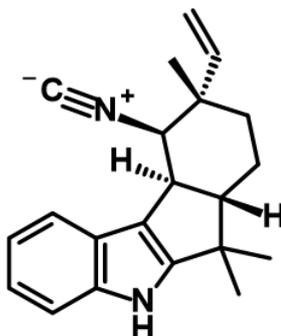
(DCM, c 0.37); IR ν_{\max} 3375, 2954, 1454, 1028, 749 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.85 – 7.74 (m, 2H), 7.33 – 7.27 (m, 1H), 7.11 – 7.04 (m, 2H), 6.27 (dd, $J = 17.6, 11.0$ Hz, 1H), 6.27 (dd, $J = 17.6, 11.0$ Hz, 1H), 5.30 (dd, $J = 11.0, 1.6$ Hz, 1H), 5.30 (dd, $J = 11.0, 1.6$ Hz, 1H), 5.24 (dd, $J = 17.6, 1.6$ Hz, 1H), 3.46 (dd, $J = 10.8, 9.0$ Hz, 1H), 2.81 (t, $J = 10.6$ Hz, 1H), 2.05 – 1.97 (m, 2H), 1.95 (d, $J = 9.0$ Hz, 1H), 1.66 (td, $J = 13.0, 3.7$ Hz, 1H), 1.55 – 1.45 (m, 2H), 1.36 (s, 3H), 1.23 (s, 3H), 1.06 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 152.0, 140.6, 139.7, 124.7, 120.7, 120.6, 120.0, 118.1, 115.9, 111.5, 81.2, 62.1, 47.0, 43.5, 40.8, 38.9, 25.9, 25.5, 21.0, 20.8; LC-MS (pH3) RT 2.20 min, 94.5 %; LC-MS (pH10) RT 2.15 min, 94.5 %; HRMS (ESI) calc. $\text{C}_{20}\text{H}_{26}\text{NO}$ $[\text{M}+\text{H}]^+$ 296.2014, found 296.2015.



(6a*S*,9*S*,10*R*,10a*S*)-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1-b]indol-10-ol (26)

The *title compound* was isolated as the minor isomer from the previous reduction reaction, (27 mg, 19%). $R_f = 0.32$ (DCM); $[\alpha]_D = -17$ (DCM, c 0.23); IR ν_{\max} 3393, 2955, 1455, 1067, 740 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.90 (br. s, 1H), 7.54 – 7.46 (m, 1H), 7.36 – 7.30 (m, 1H), 7.14 – 7.06 (m, 2H), 5.96 (dd, $J = 17.7, 11.0$ Hz, 1H), 5.16 (dd, $J = 17.7, 1.1$ Hz, 1H), 5.12 (dd, $J = 11.0, 1.2$ Hz, 1H), 4.37 (s, 1H), 3.18 (dd, $J = 11.0, 2.2$ Hz, 1H), 2.40 (td, $J = 11.9, 3.3$ Hz, 1H), 1.83 – 1.69 (m, 2H), 1.71 – 1.59 (m, 2H), 1.36 (s, 3H), 1.17 (s, 3H), 1.03 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 153.7, 145.4, 139.7, 124.4, 120.8, 120.1, 118.5,

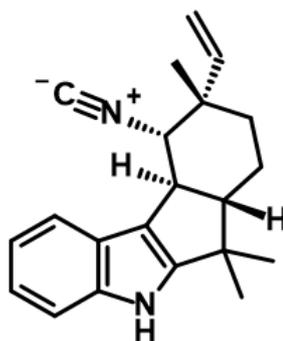
115.1, 112.9, 111.9, 74.8, 54.2, 44.8, 42.0, 40.5, 32.1, 26.8, 25.3, 21.2, 20.7;
 LC-MS (pH3) RT 2.29 min, 97.5 %; LC-MS (pH10) RT 2.25 min, 99.2 %; HRMS (ESI)
 calc. C₂₀H₂₆NO [M+H]⁺ 296.2014, found 296.2017.



(6*aS*,9*S*,10*R*,10*aS*)-10-isocyano-6,6,9-trimethyl-9-vinyl-5,6*a*,7,8,10,10*a*-hexahydroindeno[2,1-*b*]indole [(+)-12-*epi*-fischerindole U isonitrile, (1)]

(6*aS*,9*S*,10*aS*)-6,6,9-trimethyl-9-vinyl-6*a*,7,8,10*a*-tetrahydro-5*H*-indeno[2,1-*b*]indol-10-one (0.43 mmol, 125 mg) was dissolved in 5/1 v/v MeOH/THF (20 mL) and ammonium acetate (17.0 mmol, 1.31 g) and sodium cyanoborohydride (4.27 mmol, 268 mg) added. After five days, the reaction was quenched with saturated aqueous sodium bicarbonate and extracted with diethyl ether. The combined organic extracts were dried, filtered and reduced *in vacuo* and the crude material separated by SFC [10-25% MeOH (0.1% NH₄OH); Phenomenex LUX-Cellulose 3: 4.6 x 150 mm, 3 μm] to yield 16 mg (26%) and 18 mg (29%) of amine *C*-11 isomers. The minor isomer was sequentially treated with formic acid (0.06 mmol, 2.5 μL), CDMT (0.077 mmol, 13.9 mg), DMAP (0.008 mmol, 1 mg), NMM (0.077 mmol, 8.5 μL) and DCM (0.5 mL). After stirring for 30 min, the reaction was diluted with DCM and washed with 1M HCl, saturated sodium bicarbonate solution then brine. The solvent was reduced *in vacuo* and the residue purified by flash column chromatography (20-50% EtOAc/hexane) to give the formamide.

The formamide was dissolved in DCM (4 mL) and cooled to 0 °C. Triethylamine (1.57 mmol, 220 µL) and then phosgene (20% solution in toluene, 0.38 mmol, 200 µL) were added and stirred for 20 min. The reaction was quenched with saturated aqueous sodium bicarbonate and extracted with EtOAc. The combined organic phase was reduced *in vacuo* and purified by flash column chromatography (50-100% DCM/hexane) to give the *title compound* as a white solid (7.5 mg, 12% from the ketone precursor). $R_f = 0.54$ (DCM); $[\alpha]_D = +60$ (DCM, c 0.30); IR ν_{\max} 3317, 2963, 2154, 1456, 713 cm^{-1} ; ^1H NMR (400 MHz, CD_2Cl_2) δ 8.05 (br. s, 1H), 7.51 - 7.44 (m, 1H), 7.42 - 7.36 (m, 1H), 7.16 - 7.07 (m, 2H), 5.93 (dd, $J = 17.6, 11.0$ Hz, 1H), 5.27 (d, $J = 17.7$ Hz, 1H), 5.26 (dd, $J = 11.0, 0.6$ Hz, 1H), 4.40 (s, 1H), 3.26 - 3.15 (m, 1H), 2.36 (td, $J = 11.0, 3.9$ Hz, 1H), 1.99 - 1.91 (m, 1H), 1.80 - 1.72 (m, 1H), 1.72 - 1.63 (m, 2H), 1.44 (s, 3H), 1.30 (s, 3H), 1.08 (s, 3H); ^{13}C NMR (100 MHz, CD_2Cl_2) δ 158.2, 153.0, 143.2, 140.0, 125.2, 124.3, 121.1, 120.2, 118.4, 115.1, 114.7, 112.1, 63.0, 55.7, 53.8, 42.5, 41.2, 40.6, 32.3, 28.2, 25.2, 21.0, 20.9; LC-MS (pH10) RT 2.38 min, 95.2 %; HRMS (ESI) calc. $\text{C}_{21}\text{H}_{25}\text{N}_2$ $[\text{M}+\text{H}]^+$ 305.2018, found 305.2016.

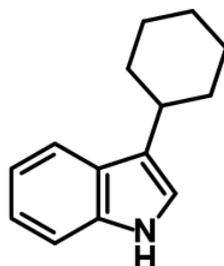


(6a*S*,9*S*,10*S*,10a*S*)-10-isocyano-6,6,9-trimethyl-9-vinyl-5,6a,7,8,10,10a-hexahydroindeno[2,1-b]indole (32)

The *title compound* was synthesised from (6a*S*,9*S*,10a*S*)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one in the same manner and with comparable yield to (+)-12-*epi*-fischerindole U isonitrile; using the major

amine isomer isolated from the reductive amination reaction. $R_f = 0.62$ (DCM); $[\alpha]_D = -162$ (DCM, c 0.20); IR ν_{\max} 3261, 2961, 2164, 1449, 743 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 8.20 – 8.09 (m, 1H), 7.88 (br. s, 1H), 7.35 – 7.28 (m, 1H), 7.18 – 7.08 (m, 2H), 6.25 (dd, $J = 17.5, 11.1$ Hz, 1H), 5.32 (d, $J = 11.1$ Hz, 1H), 5.27 (dd, $J = 17.6, 0.8$ Hz, 1H), 3.48 (d, $J = 11.3$ Hz, 1H), 3.18 (t, $J = 10.8$ Hz, 1H), 2.18 (dt, $J = 13.9, 2.9$ Hz, 1H), 1.98 – 1.87 (m, 1H), 1.66 (qd, $J = 12.7, 3.4$ Hz, 1H), 1.59 – 1.39 (m, 2H), 1.36 (s, 3H), 1.30 (s, 3H), 1.03 (s, 3H). ^{13}C NMR (100 MHz, CDCl_3) δ 157.5, 152.2, 139.7, 139.0, 124.3, 121.5, 121.1, 120.5, 115.8, 115.6, 111.5, 66.3, 62.5, 46.2, 42.0, 40.3, 36.9, 27.7, 25.3, 20.8, 20.6; LC-MS (pH10) RT 2.39 min, 100.0 %; HRMS (ESI) calc. $\text{C}_{21}\text{H}_{25}\text{N}_2$ $[\text{M}+\text{H}]^+$ 305.2018, found 305.2016.

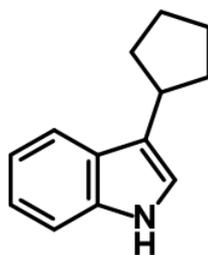
7.3.2. Simplified substrate analogues



3-cyclohexyl-1H-indole (28)

Trichloroacetic acid (13.0 mmol, 2.07 g) and triethylsilane (25.4 mmol, 4.10 mL) were dissolved in toluene (10 mL) and heated to 70°C. A mixture of indole (8.79 mmol, 1.04 g) and cyclohexanone (9.40 mmol, 0.98 mL) in toluene (5 mL) was added dropwise. Heating continued at 70°C for 2 hr before cooling to room temperature. The reaction was quenched with saturated aqueous sodium bicarbonate and extracted with EtOAc. The combined organic extracts were dried and concentrated *in vacuo* prior to flash column chromatography (0-15% EtOAc/hexane). The isolated material was freeze-dried overnight to give the *title compound*, 1.34 g (77%) as an off-white solid. $R_f = 0.27$ (10%

EtOAc/hexane); IR ν_{\max} 3403, 2920, 2846, 1456, 736 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.87 (br. s, 1H), 7.66 (d, $J = 7.9$ Hz, 1H), 7.35 (d, $J = 8.1$ Hz, 1H), 7.17 (t, $J = 7.5$ Hz, 1H), 7.10 (t, $J = 7.5$ Hz, 1H), 6.98 – 6.92 (m, 1H), 2.92 – 2.77 (m, 1H), 2.18 – 2.05 (m, 2H), 1.89 – 1.75 (m, 3H), 1.53 – 1.40 (m, 4H), 1.36 – 1.26 (m, 1H); ^{13}C NMR (100 MHz, CDCl_3) δ 136.5, 126.9, 123.4, 121.9, 119.5, 119.5, 119.1, 111.2, 35.6, 34.2, 27.1, 26.7; LC-MS (pH3) RT 2.78 min, 99.4%; LC-MS (pH10) RT 2.88 min, 98.8%; HRMS (ESI) calc. $\text{C}_{14}\text{H}_{18}\text{N}$ $[\text{M}+\text{H}]^+$ 200.1439, found 200.1440.

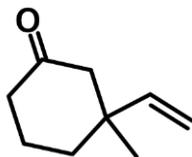


3-cyclopentyl-1H-indole (29)

Trichloroacetic acid (13.0 mmol, 2.07 g) and triethylsilane (25.4 mmol, 4.10 mL) were dissolved in toluene (10 mL) and heated to 70°C. A mixture of indole (8.62 mmol, 1.02 g) and cyclopentanone (9.40 mmol, 0.84 mL) in toluene (5 mL) was added dropwise. Heating continued at 70°C for 2 hr before cooling to room temperature. The reaction was quenched with saturated aqueous sodium bicarbonate and extracted with EtOAc. The combined organic extracts were dried and concentrated *in vacuo* prior to flash column chromatography (0-10% EtOAc/hexane). The isolated material was freeze-dried overnight to give the *title compound*, 618 mg (39%) as pale yellow solid. $R_f = 0.30$ (10%

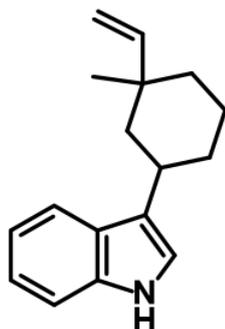
EtOAc/hexane); IR ν_{\max} 3398, 2955, 2862, 1456, 741 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.86 (s, 1H), 7.66 (d, $J = 7.9$ Hz, 1H), 7.35 (d, $J = 8.1$ Hz, 1H), 7.18 (t, $J = 7.4$ Hz, 1H), 7.10 (t, $J = 7.5$ Hz, 1H), 7.00 – 6.95 (m, 2H), 3.34 – 3.21 (m, 1H), 2.22 – 2.10 (m, 2H), 1.87 – 1.64 (m, 6H). ^{13}C NMR (100 MHz, CDCl_3) δ 136.8, 127.5, 122.0, 121.5, 119.8, 119.7, 119.1, 111.2, 37.1, 33.3, 25.4; LC-MS (pH3) RT

2.59 min, 100.0%; LC-MS (pH10) RT 2.67 min, 100.0%; HRMS (ESI) calc. $C_{13}H_{16}N$ $[M+H]^+$ 186.1269, found 186.1283.



3-methyl-3-vinylcyclohexanone

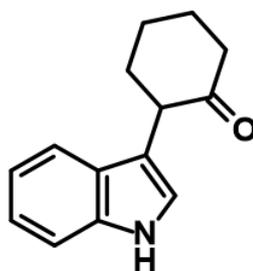
Vinylmagnesium bromide (1M in THF, 27.2 mmol, 27.2 mL) was added to a suspension of cuprous iodide (3.56 mmol, 678 mg) in THF (70 mL) at -78°C and the mixture stirred for 30 min at -78°C . 3-methyl-2-cyclohexen-1-one (17.8 mmol, 2.06 mL) was added dropwise and the reaction stirred for 1 hr before warming to room temperature overnight. Saturated aqueous ammonium chloride was added and the reaction extracted with EtOAc. The combined organic extracts were dried and reduced *in vacuo* before purification by flash column chromatography (0-20% EtOAc/hexane) to give the *title compound*, 1.50 g (61%) as a yellow liquid. $R_f = 0.32$ (10% EtOAc/hexane); IR ν_{max} 2957, 2871, 1707, 1423, 916 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 5.70 (dd, $J = 17.5, 10.9$ Hz, 1H), 5.00 (dd, $J = 10.9, 0.7$ Hz, 1H), 4.96 (dd, $J = 17.5, 0.7$ Hz, 1H), 2.44 (dt, $J = 14.0, 1.5$ Hz, 1H), 2.35 - 2.17 (m, 2H), 2.16 (d, $J = 14.0$ Hz, 1H), 1.90 - 1.78 (m, 2H), 1.76 - 1.68 (m, 1H), 1.66 - 1.57 (m, 1H), 1.06 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 211.4, 145.9, 112.8, 51.8, 41.7, 41.0, 36.7, 27.4, 22.2; HRMS (ESI) calc. $C_9H_{15}O$ $[M+H]^+$ 139.1123, found 139.1120.



3-(3-methyl-3-vinyl-cyclohexyl)-1H-indole (30)

Trichloroacetic acid (12.7 mmol, 2.08 g) and triethylsilane (25.4 mmol, 4.10 mL) were dissolved in toluene (10 mL) and heated to 70°C. A mixture of indole (8.79 mmol, 1.04 g) and 3-methyl-3-vinyl-cyclohexanone (9.26 mmol, 1.28 g) in toluene (5 mL) was added dropwise. Heating continued at 70°C for 2 hr before cooling to room temperature. The reaction was quenched with saturated aqueous sodium bicarbonate and extracted with EtOAc. The combined organic extracts were dried and concentrated *in vacuo* prior to flash column chromatography (0-12% EtOAc/hexane). This was further purified by prep. HPLC to give the *title compound*, 1.24 g (59%) as a yellow oil as an inseparable mixture of diastereoisomers (2:1). $R_f = 0.28$ (10% EtOAc/hexane); IR ν_{\max} 3415, 2921, 2858, 1456, 908, 737 cm^{-1} ; $^1\text{H NMR}$ (400 MHz, CDCl_3) (major) δ 7.88 (br.s, 1H), 7.65 (d, $J = 7.9$ Hz, 1H), 7.35 (d, $J = 8.1$ Hz, 1H), 7.18 (t, $J = 7.5$ Hz, 1H), 7.11 (t, $J = 7.5$ Hz, 1H), 6.95 (s, 1H), 5.97 (dd, $J = 17.6, 11.0$ Hz, 1H), 5.15 (dd, $J = 11.1, 1.4$ Hz, 1H), 5.14 (dd, $J = 17.6, 1.4$ Hz, 1H), 3.03 (m, 1H), 2.19 – 2.04 (m, 1H), 1.90 – 1.61 (m, 3H), 1.59 – 1.23 (m, 3H), 1.02 (s, 3H); $^1\text{H NMR}$ (400 MHz, CDCl_3) (minor) δ 7.88 (br.s, 1H), 7.66 (d, $J = 8.0$ Hz, 1H), 7.35 (d, $J = 8.0$ Hz, 1H), 7.18 (t, $J = 7.5$ Hz, 1H), 7.11 (t, $J = 7.5$ Hz, 1H), 6.95 (s, 1H), 5.89 (dd, $J = 17.6, 10.9$ Hz, 1H), 4.96 (dd, $J = 17.6, 1.4$ Hz, 1H), 4.88 (dd, $J = 10.8, 1.4$ Hz, 1H), 3.12 (m, 1H); 2.19 – 2.04 (m, 1H), 1.90 – 1.61 (m, 3H), 1.59 – 1.23 (m, 3H), 1.21 (s, 3H); $^{13}\text{C NMR}$ (100 MHz, CDCl_3) (mixture) δ 151.4, 146.8, 136.5, 126.9, 126.9,

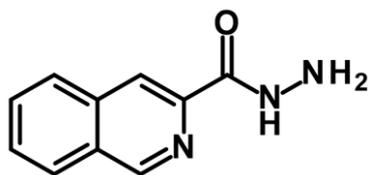
123.1, 122.9, 122.0, 122.0, 119.5, 119.4, 119.4, 119.3, 119.1, 119.1, 112.5, 111.3, 111.3, 109.0, 45.6, 44.4, 37.9, 37.8, 37.2, 36.8, 33.8, 33.7, 32.1, 31.2, 30.6, 23.1, 22.5, 22.3; LC-MS (pH3) RT 3.06 min, 100.0%; LC-MS (pH10) RT 3.14 min, 99.7%; HRMS (ESI) calc. $C_{17}H_{22}N$ $[M+H]^+$ 240.1752, found 240.1738.



2-(1H-indol-3-yl)cyclohexanone (31)

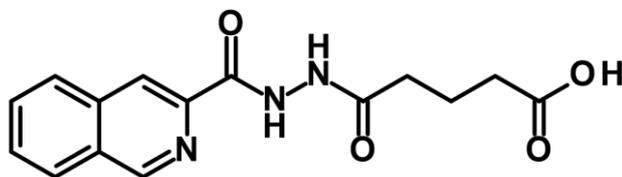
Indole (5.11 mmol, 605 mg) and 2-chlorocyclohexanone (6.09 mmol, 0.71 mL) were dissolved in 2,2,2-trifluoroethanol (10 mL) and sodium carbonate (7.12 mmol, 762 mg) added. The reaction was stirred overnight at room temperature before the addition of further 2-chlorocyclohexanone (3.10 mmol, 0.35 mL). The reaction was stirred for 2 days before filtering through celite and concentrating *in vacuo*. Purification by flash column chromatography (0-50% EtOAc/hexane) gave the *title compound*, 805 mg (74%) as a beige solid. $R_f = 0.37$ (40% EtOAc/hexane); IR ν_{max} 3333, 2927, 2858, 1701, 1460, 740 cm^{-1} ; 1H NMR (400 MHz, $CDCl_3$) δ 8.15 (br.s, 1H), 7.47 (d, $J = 7.9$ Hz, 1H), 7.31 (d, $J = 8.1$ Hz, 1H), 7.17 (t, $J = 7.5$ Hz, 1H), 7.10 (t, $J = 7.5$ Hz, 1H), 7.05 (d, $J = 2.3$ Hz, 1H), 3.93 (dd, $J = 11.3, 5.5$ Hz, 1H), 2.63 - 2.47 (m, 2H), 2.44 - 2.35 (m, 1H), 2.21 - 1.98 (m, 3H), 1.95 - 1.82 (m, 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 211.0, 136.3, 127.1, 122.1, 122.1, 119.5, 119.2, 113.8, 111.4, 48.9, 42.1, 35.1, 28.3, 25.4; LC-MS (pH3) RT 1.80 min, 100.0%; LC-MS (pH10) RT 1.85 min, 99.1%; HRMS (ESI) calc. $C_{14}H_{16}NO$ $[M+H]^+$ 214.1232, found 214.1223.

7.3.3. Re-synthesis of ThermoFluor assay hits



Isoquinoline-3-carbohydrazide

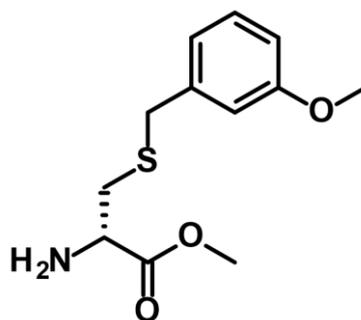
A solution of methyl isoquinoline-3-carboxylate (10.3 mmol, 1.97 g) and hydrazine monohydrate (64-65%, 12.6 mmol, 0.95 mL) in MeOH (20 mL) was heated to reflux for 6 hr. After cooling to room temperature, the solvent was removed *in vacuo* and the residue triturated in ether. Filtration yielded the *title compound*, 1.82 g (94%) as an off-white solid. $R_f = 0.29$ (5% MeOH/DCM); IR ν_{\max} 3289, 1671 cm^{-1} ; $^1\text{H NMR}$ (400 MHz, $\text{d}^6\text{-DMSO}$) δ 9.96 (br. s, 1H), 9.34 (s, 1H), 8.51 (s, 1H), 8.23 (d, $J = 8.2$ Hz, 1H), 8.18 (d, $J = 7.9$ Hz, 1H), 7.87 (t, $J = 7.6$ Hz, 1H), 7.79 (t, $J = 7.6$ Hz, 1H), 4.62 (d, $J = 4.2$ Hz, 2H); $^{13}\text{C NMR}$ (100 MHz, $\text{d}^6\text{-DMSO}$) δ 162.8, 151.6, 143.5, 135.3, 131.3, 129.1, 128.9, 127.8, 127.7, 119.3; LC-MS (pH3) RT 0.85 min, 99.4%; LC-MS (pH10) RT 0.96 min, 98.6%; HRMS (ESI) calc. $\text{C}_{10}\text{H}_{10}\text{N}_3\text{O}$ $[\text{M}+\text{H}]^+$ 188.0824, found 188.0815.



5-[2-(isoquinoline-3-carbonyl)hydrazino]-5-oxo-pentanoic acid (33)

Isoquinoline-3-carbohydrazide (8.55 mmol, 1.60 g) and glutaric anhydride (8.74 mmol, 1.05 g) were stirred in CHCl_3 (150 mL) overnight and then left to stand (without stirring) for 3 hr. The precipitate was collected and recrystallised from EtOH to give the *title compound*, 1.30 g (51 %) as a white solid. $R_f = 0.37$

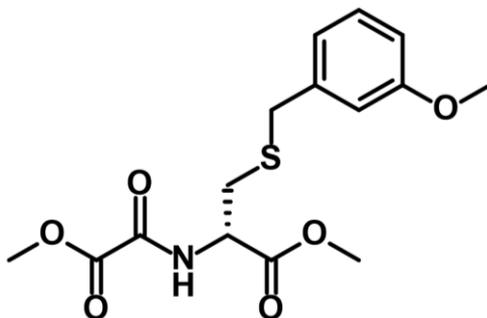
(2% AcOH in 20% MeOH/EtOAc); IR ν_{\max} 3367, 3276, 2972, 1724, 1681, 1644 cm^{-1} ; ^1H NMR (400 MHz, d^6 -DMSO) δ 12.05 (br. s, 1H), 10.49 (s, 1H), 10.07 (br. s, 1H), 9.40 (s, 1H), 8.57 (s, 1H), 8.27 (d, $J = 8.1$ Hz, 1H), 8.21 (d, $J = 7.8$ Hz, 1H), 7.90 (t, $J = 7.8$ Hz, 1H), 7.83 (t, $J = 7.6$ Hz, 1H), 2.32 (t, $J = 7.5$ Hz, 2H), 2.25 (t, $J = 7.3$ Hz, 2H), 1.85 - 1.72 (m, 2H); ^{13}C NMR (100 MHz, d^6 -DMSO) δ 174.2, 170.5, 162.8, 151.7, 142.9, 135.2, 131.4, 129.4, 129.3, 128.0, 127.8, 120.3, 32.8, 32.4, 20.5; LC-MS (pH3) RT 1.02 min, 99.6%; LC-MS (pH10) RT 0.55 min, 99.0%; HRMS (ESI) calc. $\text{C}_{15}\text{H}_{16}\text{N}_3\text{O}_4$ $[\text{M}+\text{H}]^+$ 302.1141, found 302.1109.



Methyl (2S)-2-amino-3-[(3-methoxyphenyl)methylsulfanyl]propanoate

3-methoxybenzyl bromide (15.7 mmol, 2.20 ml) was added dropwise to a solution of *D*-cysteine methyl ester hydrochloride salt (14.2 mmol, 2.49 g) in ammonia solution (7N in MeOH, 350 mmol, 50 mL) at -20°C . The reaction was stirred at 0°C for 3 hr and allowed to warm to room temperature overnight. The solvent was reduced *in vacuo* and partitioned between EtOAc and water, the layers were separated and the aqueous layer extracted with further EtOAc. The combined organic extracts were washed with water and brine before being dried and reduced *in vacuo*. The crude material was then purified by flash column chromatography (0-5% MeOH/EtOAc) to give the *title compound*, (2.45 g, 68%) as a pale yellow oil. $R_f = 0.34$ (2% [5% NH_3 in MeOH]/EtOAc); $[\alpha]_D = -21$ (MeOH, c 0.15); IR ν_{\max} 3374, 2951, 2835, 1734 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.22 (t, $J = 7.8$ Hz, 1H), 6.92 - 6.84 (m, 2H), 6.79 (dd, $J = 8.2, 1.9$ Hz,

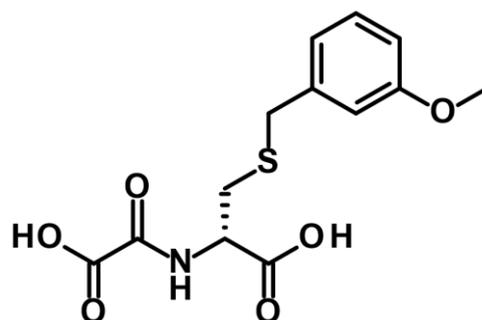
1H), 3.80 (s, 3H), 3.72 (s, 3H), 3.70 (s, 2H), 3.60 (dd, $J = 7.5, 4.7$ Hz, 1H), 2.83 (dd, $J = 13.5, 4.7$ Hz, 1H), 2.67 (dd, $J = 13.5, 7.5$ Hz, 1H), 1.73 (s, 2H); ^{13}C NMR (100 MHz, CDCl_3) δ 174.6, 159.9, 139.6, 129.7, 121.4, 114.5, 112.9, 55.3, 54.2, 52.3, 36.8, 36.7; LC-MS (pH3) RT 0.79 min, 99.2%; LC-MS (pH10) RT 1.55 min, 98.8%; HRMS (ESI) calc. $\text{C}_{12}\text{H}_{18}\text{NO}_3\text{S}$ $[\text{M}+\text{H}]^+$ 256.1007, found 256.0988.



Methyl (2S)-2-[(2-methoxy-2-oxo-acetyl)amino]-3-[(3-methoxyphenyl)methylsulfanyl]propanoate

Methyl (2S)-2-amino-3-[(3-methoxyphenyl)methylsulfanyl]propanoate (9.32 mmol, 2.38 g) was dissolved in DCM (45 mL) and cooled to 0°C. Triethylamine (14.2 mmol, 2.00 mL) was added, followed by the dropwise addition of methyl oxalyl chloride (10.4 mmol, 1 mL) and the reaction permitted to warm to room temperature with stirring over 3 hr. The reaction was quenched saturated aqueous sodium bicarbonate, passed through a phase separator and the organic layer reduced *in vacuo*. The aqueous layer was extracted with EtOAc and the combined organic extracts washed with water and brine before being dried, reduced under vacuum and combined with the previously collected material from the DCM layer. The crude material was purified by flash column chromatography (20-60% EtOAc/hexane) to give the *title compound*, 2.55 g (80 %) as a yellow oil. $R_f = 0.25$ (40% EtOAc/hexane); $[\alpha]_D^{25} = +56$ (MeOH, c 0.10); IR ν_{max} 3350, 1739, 1699 cm^{-1} ; ^1H NMR (400 MHz, CDCl_3) δ 7.69 (d, $J = 7.7$ Hz, 1H), 7.22 (t, $J = 7.8$ Hz, 1H), 6.89 - 6.83 (m, 2H), 6.79 (dd, $J = 8.3, 2.6$ Hz, 1H),

4.84 - 4.74 (m, 1H), 3.92 (s, 3H), 3.81 (s, 3H), 3.77 (s, 3H), 3.69 (s, 2H), 2.95 (dd, $J = 14.0, 5.2$ Hz, 1H), 2.89 (dd, $J = 14.0, 5.2$ Hz, 1H); ^{13}C NMR (100 MHz, CDCl_3) δ 170.1, 160.4, 160.0, 156.0, 138.4, 129.7, 121.4, 114.4, 113.2, 55.3, 53.9, 53.0, 52.2, 36.7, 33.1; LC-MS (pH3) RT 1.78 min, 98.6%; LC-MS (pH10) RT 1.86 min, 96.4%; HRMS (ESI) calc. $\text{C}_{15}\text{H}_{20}\text{NO}_6\text{S}$ $[\text{M}+\text{H}]^+$ 342.1011, found 342.0980.



(2S)-3-[(3-methoxyphenyl)methylsulfanyl]-2-(oxaloamino)propanoic acid (34)

A solution of sodium hydroxide (1M, 10 mmol, 10 mL) was added dropwise to a solution of methyl (2S)-2-[(2-methoxy-2-oxo-acetyl)amino]-3-[(3-methoxyphenyl)methylsulfanyl]propanoate (6.44 mmol, 2.20 g) in MeOH (20 mL) at 0°C. The reaction was permitted to warm to room temperature overnight. The MeOH was removed *in vacuo* and the resulting residue partitioned between EtOAc and water. The organic layer was discarded and the aqueous layer adjusted to pH 2 with 1M HCl. The aqueous layer was extracted with EtOAc and the combined organic phase dried and reduced under vacuum to give the *title compound*, 1.79 g (90 %) as a viscous yellow oil. $R_f = 0.2$ (50% MeOH/EtOAc); $[\alpha]_D = +18$ (MeOH, c 0.11); IR ν_{max} 2939, 1732, 1682 cm^{-1} ; ^1H NMR (400 MHz, $\text{d}^4\text{-MeOD}$) δ 7.20 (t, $J = 8.1$ Hz, 1H), 6.93 - 6.86 (m, 2H), 6.83 - 6.75 (m, 1H), 4.63 (dd, $J = 8.1, 4.7$ Hz, 1H), 3.78 (s, 3H), 3.74 (s, 2H), 3.01 (dd, $J = 14.1, 4.7$ Hz, 1H), 2.87 (dd, $J = 14.1, 8.0$ Hz, 1H); ^{13}C NMR (100 MHz,

d⁴-MeOD) δ 172.8, 162.3, 161.3, 159.9, 140.8, 130.5, 122.4, 115.4, 114.0, 55.6, 53.5, 36.9, 33.2; LC-MS (pH3) RT 0.69 min, 97.1%; LC-MS (pH10) RT 0.15 min, 96.0%; HRMS (ESI) calc. C₁₃H₁₆NO₆S [M+H]⁺ 314.0698, found 314.0681.

7.4. WelO5 assays & characterisation

7.4.1. Mass spectrometry

7.4.1.1. LC-MS

Protein intact masses were determined by LC-MS analysis on a Waters ACQUITY® UPLC I-Class system coupled to a Waters Xevo G2 QTOF MS, fitted with a Waters ACQUITY® BEH C300 column (2.1 x 100mm, 1.7 μ m). Method: 5% – 95 % acetonitrile/water (0.1% TFA), 8 min run time, 0.3 mL/min flow rate. Batch-purified WelO5 samples were analysed at 20 μ M concentration (50 μ L volume). For His-SELECT purified variants, the 250 μ L elution fraction (E1) was used for analysis directly. Data were processed with Waters MassLynx™ MS software and the protein peaks deconvoluted using MaxEnt1.

7.4.1.2. SEC-MALS

SEC-MALS analysis was performed by Weston Struwe at the Department of Biochemistry, Oxford. Samples were analysed on a Wyatt Dawn HELEOS-II 8-angle light scattering detector and Wyatt Optilab® T-rEX™ refractive index monitor linked to a Shimadzu HPLC system comprising LC-20AD pump, SIL-20A autosampler and SPD20A UV-Vis detector.

7.4.1.3. ICP-MS

ICP-MS sample preparation and analysis was performed by Kathrin Abraham & Phil Holdship at the Department of Earth Sciences, Oxford. Samples were

analysed for Mn, Fe, Co, Ni, Cu and Zn content using a Perkin Elmer NexION® 350D ICP-MS system.

7.4.2. Circular dichroism (CD)

The far-UV CD spectrum of WelO5 was obtained from 185-260 nm on an Applied Photophysics Chirascan™ CD spectrometer using a cell of 0.1 cm pathlength at 25 °C. Samples were prepared at 0.2 mg/mL in 10 mM sodium phosphate buffer, pH 7.5 (1 mL volume). The DichroWeb server was used for data processing and estimation of WelO5 secondary structure, with the best data fit provided by the CDSSTR algorithm.

7.4.3. UV-Vis

Samples were prepared in a low (approx. 1 ppm) oxygen glove box. WelO5 enzyme in 20 mM HEPES buffer, pH 7.5 was deoxygenated with 30 cycles of argon and vacuum. 2OG, NaCl and $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ solutions were prepared by transferring the pre-weighed solids into the glove box and diluting up to volume with deoxygenated 20 mM HEPES buffer, pH 7.5 (or water for Fe^{2+}).

Reagent	Stock Solution	Final Concentration	Volume (μL)
WelO5	210 μM	200 μM	476
2OG sodium salt	500 mM	1 mM	1
$(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$	50 mM	250 μM	2.5
NaCl	500 mM	1 mM	1
20 mM HEPES buffer	-	-	20
Total			500.5

Spectra were collected at ambient temperature from 200-800 nm for four different conditions; apo WelO5, WelO5 + 2OG, WelO5 + 2OG + Fe^{2+} and WelO5 + 2OG + Fe^{2+} + Cl⁻ by successive addition of the relevant solutions (specified in

the table). Spectra were obtained using a Varian CARY® UV-Vis 4000 spectrophotometer and exported to Microsoft Excel for analysis.

7.4.4. X-Ray crystallography

7.4.4.1. Crystallisation

WelO5 and selenomethionine-derivatised WelO5 (SeMet-WelO5) were expressed as per Chapter 7 (Sections 7.2.12.5. & 7.2.12.7. respectively) and purified according to the Section 7.2.13.1.3. protocol.

For WelO5·Ni²⁺·2OG (PDB ID: 5J4R) and WelO5·Ni²⁺·2OG (PDB ID: 5T22):

Purified WelO5 protein (35 mg/mL in 20 mM HEPES, 150 mM NaCl buffer, pH 7.5) was supplemented with 5 mM NiCl₂, 5 mM 2OG and 5 mM of substrate analogue (2*S*,4*aS*,6*aR*,11*bR*,11*cS*)-2,5-dimethyl-2-vinyl-4,4*a*,6*a*,7,11*b*,11*c*-hexahydro-3*H*-benzo[*c*]carbazol-1-one; **19** and diluted to a final volume of 1 mL with 20 mM HEPES buffer, pH 7.5 before incubating on ice for 1 hr. To remove excess ligands that remained in solution (not bound to protein), the solution was passed through a PD-10 column (GE Healthcare) pre-equilibrated with 20 mM HEPES buffer, pH 7.5 and concentrated down (Millipore Amicon® Ultra 10 kDa, 4 mL) resulting in approximately the same protein concentration as used initially (i.e. 35 mg/mL). The sample was centrifuged in a microfuge for 10 min (10 K) to remove any potential aggregates prior to the setup of crystallisation trial plates. The sample prepared for SeMet-WelO5 crystallisation trials was set up in a similar manner.

For WelO5·Ni²⁺·succinate (PDB ID: 5TRQ):

The purified WelO5 protein (30 mg/mL in 20 mM HEPES, 150 mM NaCl buffer, pH 7.5) was supplemented with 5 mM NiCl₂ and 5 mM succinate before incubating on ice for 1 hr. The sample was centrifuged in a microfuge for 10 min (10 K) to remove any potential aggregates prior to the setup of crystallisation trial plates. Crystals were grown by the sitting drop vapour diffusion method (drop size 200-300 nL) in 96-3 well INTELLI-PLATE® low-profile protein crystallisation plates (Art Robbins: 102-0001-13) set up at 293K with a Phoenix® RE (Art Robbins/Rigaku) liquid handling system. Three different drop ratios of precipitant to protein (2:1 300 nL, 1:1 200 nL and 1:2 300 nL) were set up per well reservoir. All crystals were harvested with nylon loops after briefly (10-60 seconds) soaking in cryoprotectant (25 % v/v glycerol in precipitant solution) and cryo-cooled by flash-freezing in liquid nitrogen (for 5J4R; 20 mM 5-amino-2,4,6-triiodoisophthalic acid (I3C) was added to the cryoprotectant in the hope that it could be used for phasing based on the electron rich iodine atoms.¹⁷⁶

Broad screening was performed using the following commercially available sparse matrix screens:

Commercial Screens	Supplier
JCSG- <i>plus</i> [™] , PACT premier [™] , MIDAS [™]	Molecular Dimensions
Index [™] , SaltRx [™] , PEGRx [™] , PEG/Ion [™] , Crystal Screen [™]	Hampton Research
Wizard [™]	Rigaku

7.4.4.2. Data collection and structure determination

X-ray diffraction data were collected at the ESRF, European Synchrotron Radiation Facility (SeMet-WelO5; Ian Clifton & Michael McDonough, Schofield

Group) & Diamond Light Source (5J4R, 5T22 & 5TRQ; Michael McDonough, Schofield Group).

The initial SeMet-WelO5 structure was solved by the single-wavelength anomalous diffraction (SAD) method using the anomalous signal from the protein selenium atoms. Data were collected at ESRF BM14 using X-ray energy at the selenium peak (0.976 Å) resulting in a 2.0 Å anomalous data set (Ian Clifton & Michael McDonough, Schofield Group). Selenium sites were identified using SHELXD, initial phases were calculated using these sites and HKL2MAP resulted in an electron density map (Figure of Merit 0.53, CC 56.5 %). Density modification using SHELXE improved the phases (Figure of Merit 0.69, CC 73.6 %). Initial model building using the SeMet electron density maps was performed in COOT with refinement carried out using PHENIX. The high resolution 5J4R structure was then solved by molecular replacement in PHASER using the SeMet-WelO5 structure as a search model. Similarly, 5T22 and 5TRQ were solved by molecular replacement using 5J4R as a search model with refinement performed using PHENIX. All structures were deposited in the RCSB PDB.

7.4.4.3. Crystallisation, data collection and refinement statistics

	SeMet-WelO5 + Ni ²⁺ + 2OG, 2.00 Å	WelO5 + Ni ²⁺ + 2OG, 1.65 Å	WelO5 + Ni ²⁺ + 2OG, 1.75 Å	WelO5 + Ni ²⁺ + succinate, 1.30 Å
PDB Code	N/A	5J4R	5T22	5TRQ
Crystallisation				
Protein conc. (mg/mL)	30 (in 20 mM HEPES, pH 7.5)	27 (in 20 mM HEPES, pH 7.5)	27 (in 20 mM HEPES, pH 7.5)	27 (in 20 mM HEPES, pH 7.5)
Well solution	0.2 M ammonium chloride, 20% PEG 3350	0.1 M MES pH 5.5, 0.1 M NaCl, 22% PEG 6000	0.2 M CaCl ₂ , 20% PEG 3350	0.1 M MES pH 6.0, 0.2 M Ca(OAc) ₂ , 20% PEG 8000
Additives	5 mM nickel chloride, 5 mM 2-oxoglutarate, 5mM (6aS,9S,10aS)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one	5 mM nickel chloride, 5 mM 2-oxoglutarate, 5mM (6aS,9S,10aS)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one	5 mM nickel chloride, 5 mM 2-oxoglutarate, 5 mM (6aS,9S,10aS)-6,6,9-trimethyl-9-vinyl-6a,7,8,10a-tetrahydro-5H-indeno[2,1-b]indol-10-one	5 mM nickel chloride, 5 mM succinic acid
Cryoprotection conditions	25% glycerol	20 mM I3C soak (30 sec), 25% glycerol	25% glycerol	25% glycerol
Data Collection				
X-Ray Source	ESRF BM14	Diamond, I04	Diamond, I04	Diamond, I04
Temperature (K)	100	100	100	100
Wavelength (Å)	0.98	1.54	0.98	0.93
Space group	I222	I222	P12 ₁ 1	P1
Unit cell parameters				
a, b, c, (Å)	51.61, 102.08, 123.88	52.03, 102.77, 123.98	45.95, 80.04, 73.50	46.06, 46.07, 74.44
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 97.76, 90.00	91.77, 90.54, 108.71
Molecules per ASU	1	1	2	2
Resolution range (Å)	51.04-1.78 (1.84-1.78)	47.98-1.56 (1.60-1.56)	72.99-1.75 (1.78-1.75)	38.22-1.30 (1.32-1.30)
R_{merge} (%)	0.119 (1.59)	5.9 (95.4)	18.6 (132.5)	4.5 (95.9)
Mean I/σ(I)	16.2 (1.1)	29.7 (1.5)	6.0 (N/A)	10.2 (1.1)
Completeness (%)	98.12 (86.23)	92.3 (51.3)	100.0 (99.9)	95.1 (92.2)
Redundancy	13.4 (7.5)	20.8 (6.1)	6.1 (5.8)	3.5 (3.5)
Refinement				
Data range (Å)	N/A	47.98-1.65	72.99-1.75	38.22-1.30
Number of unique reflections	N/A	75312	55523	133078
R_{work} (%)	N/A	13.8	22.1	18.8
R_{free} (%)	N/A	16.0	18.9	20.7
RMSD bonds, (Å)	N/A	0.005	0.011	0.014
RMSD angles (°)	N/A	1.030	1.07	1.21
Ramachandran favoured (%)	N/A	99	99	99
Ramachandran allowed (%)	N/A	1	1	1
Ramachandran outliers (%)	N/A	0	0	0

N.B. Highest resolution shell shown in brackets. N/A = not available

7.4.5. Thermal shift assays

ThermoFluor experiments were performed with a Bio-Rad MiniOpticon™ Real-Time PCR System (BioRad) using white 8 x 6 PCR plates. SYPRO® Orange was used to indicate thermally induced protein unfolding by binding to the exposed hydrophobic protein regions, leading to an increase in fluorescence. Protein unfolding was monitored by measurement of fluorescence over a temperature range of 25 °C and 85 °C (temperature increasing at 1 °C/min). The melting curves were exported to GraphPad Prism and a Boltzmann curve fitted between the minimum and maximal intensities, allowing calculation of the midpoint (T_m value). Reactions for the two ThermoFluor assays were prepared as follows:

Metal binding-induced stability assay:

Reagent	Stock Solution	Final Concentration	Volume (µL)
WelO5	30 µM	6 µM	10
Metal	1mM	50 µM	2.5
NOG*	100 mM	5 mM	2.5
Reaction buffer**	-	-	35
Total			50

* NOG was exchanged for Milli-Q® water for NOG-free reactions.

** Reaction buffer: 2µL SYPRO® Orange + 2.5 mL Chelex® 100-treated 200 mM HEPES buffer, pH 7.5. Control reaction: as above, but metal replaced with Milli-Q® water.

Ligand binding-induced stability assay:

Reagent	Stock Solution	Final Concentration	Volume (µL)
WelO5	30 µM	6 µM	10
Metal	1 mM	50 µM	5
Small molecule	1 mM (10% DMSO)	100 µM (1% DMSO)	5
Reaction buffer*	-	-	30
Total			50

* Reaction buffer: 2µL SYPRO® Orange + 2.5 mL Chelex® 100-treated 200 mM HEPES buffer, pH 7.5. Control reaction: as above, but small molecule replaced with 1% DMSO solution.

7.4.6. NMR

See Section 7.2.12.8.1. for expression of ^{15}N -WelO5.

^{15}N - ^1H TROSY spectra of ^{15}N -WelO5 in the presence of variable concentrations of (+)-12-*epi*-fischerindole isonitrile were collected at 37 °C on a 600 MHz Bruker Avance III HD spectrometer fitted with a cryoprobe. ^{15}N -WelO5 was buffer exchanged into 50 mM Tris pH 6.6 (10% D_2O) and concentrated using a 10,000 Da MWCO Amicon Ultra 4 mL centrifugal filter unit (Millipore) to a concentration of 330 μM . 500 μM NiCl_2 , 400 μM 2OG and 15 μL d^6 -DMSO (or d^6 -DMSO substrate stock) were added (see sample composition tables below) and transferred to a 5 mm NMR tube. Acquisition times were 40 ms in F_1 (^{15}N) and 80 ms in F_2 (^1H) with a total experiment time of around 90 min. Solvent suppression was carried out using a WATERGATE pulse sequence and data were processed using TopSpin® 3.0 software (Bruker BioSpin Ltd).

0 μM substrate (0:1)	Stock Solution	Final Concentration	Volume (μL)
WelO5	330 μM	300 μM	450
Metal	10 mM	500 μM	25
2OG	20 mM	400 μM	10
d^6 -DMSO	-	-	15
Total			500

30 μM substrate (1:10)	Stock Solution	Final Concentration	Volume (μL)
WelO5	330 μM	300 μM	450
Metal	10 mM	500 μM	25
2OG	20 mM	400 μM	10
d^6 -DMSO substrate stock solution	10 mM	30 μM	1.5
Total			500

150 μM substrate (1:2)	Stock Solution	Final Concentration	Volume (μL)
WelO5	330 μ M	300 μ M	450
Metal	10 mM	500 μ M	25
2OG	20 mM	400 μ M	10
d ⁶ -DMSO substrate stock solution	10 mM	150 μ M	7.5
Total			500

300 μM substrate (1:2)	Stock Solution	Final Concentration	Volume (μL)
WelO5	330 μ M	300 μ M	450
Metal	10 mM	500 μ M	25
2OG	20 mM	400 μ M	10
d ⁶ -DMSO substrate stock solution	20 mM	300 μ M	7.5
Total			500

500 μM substrate (xs:1)	Stock Solution	Final Concentration	Volume (μL)
WelO5	330 μ M	300 μ M	450
Metal	10 mM	500 μ M	25
2OG	20 mM	400 μ M	10
d ⁶ -DMSO substrate stock solution	20 mM	500 μ M	12.5
Total			500

7.4.7. Activity screening

7.4.7.1. Initial WelO5 activity screening protocol

The initial WelO5 screening reactions in Chapter 3, Section 3.4. were performed in 1.5 mL microtubes on a 100 μ L scale. Batch-purified WelO5 was used for these reactions (Chapter 2, Section 2.3.3.4.). Reactions were prepared as per the table, initiated by Fe²⁺ addition and incubated at 37 °C overnight alongside the relevant compound-free and enzyme-free controls.

Compound-free control: compound solutions replaced with 10 μL DMSO.

Enzyme-free control: enzyme stock replaced with 25 μL reaction buffer.

Reagent	Stock Solutions	Final Assay Concentration	Volume (μL)
WelO5	100 μM (3.2 mg/mL)	25 μM	25
Small molecule	1 mM (in DMSO)	100 μM	10
$(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2^*$	5 mM (in 20 mM HCl)	500 μM	10
2OG	20 mM (in reaction buffer)	2 mM	10
NaCl	150 mM (included in 100 μM enzyme stock solution)	37.5 mM	-
Reaction buffer	20 mM HEPES, pH 7.4	-	45
Total volume:			100

* 50 mM solution freshly prepared in 20 mM HCl and diluted to 5 mM with Milli-Q® water

Reactions were quenched by the addition 100 μL of cold methanol and centrifuged 4 °C (benchtop microfuge, 13,500 rpm, 5 min). This step removed sufficient protein and precipitants to enable direct injection of the supernatant for LC-MS analysis without further sample manipulation.

7.4.8. Finalised WelO5 variant screening

Repeated from Chapter 5, Section 5.3.3. BL21 (DE3) colonies were picked into 1 mL 2TY media (100 $\mu\text{g}/\text{mL}$ kanamycin) in 96 deep-well blocks and incubated overnight at 37 °C (220 rpm). 20 μL of overnight culture was inoculated into 2 mL volumes of 2TY Overnight Express™ System 1 auto-induction media (Novagen®) (100 $\mu\text{g}/\text{mL}$ kanamycin) in 24 deep-well blocks. The blocks were incubated for 2.5 hr at 37 °C (220 rpm), then cooled to 30 °C and incubated a further 20 hr (220 rpm). The cultures were harvested (6000 g, 10 min) and the blocks frozen at -80 °C for a minimum of 30 min. The 24 deep-well blocks were defrosted by warming to room temperature for 30 min, then 500 μL BugBuster® Master Mix added per well. After 30 min at 4 °C, the cell-extracts

were purified at room temperature using 96-well HIS-Select® filter plates (Sigma-Aldrich) according to the manufacturer's centrifuge protocol (only one wash step used).

Equilibration buffer: 20 mM HEPES, 500 mM NaCl, pH 7.4

Wash buffer: 20 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 7.4

Elution buffer: 20 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.4

The proteins were eluted in 250 μ L elution buffer* and 85 μ L was transferred to a 96 deep-well block. 4 μ L 2OG stock (20 mM) and 10 μ L (+)-12-epi-fischerindole U isonitrile substrate stock (1 mM in DMSO) were added per well. The blocks were centrifuged for 30 seconds to ensure the wells were well mixed and the reactions initiated by the addition of 1 μ L freshly prepared Fe²⁺ solution (50 mM) into each well. The blocks were sealed and incubated at 37 °C for 90 min before quenching with 100 μ L cold methanol per well. The blocks were centrifuged at 4 °C (6000 g, 20 min) and the clarified samples transferred to a 96-well plate (150 μ L) for analysis by UPLC-MS.

* For the testing of alternative substrates, this eluted protein was used directly to perform reactions in duplicate. For the testing of variants against the enzyme's natural substrate (where a 30-50 % turnover range is desired for the wild-type controls), the eluted protein was diluted 1/3 in HEPES elution buffer to perform the reactions in triplicate.

7.4.7.3. UPLC Analysis

Two methods were used depending on the reaction components being analysed. Both methods had the same system set-up, column and mobile phase but with differing gradients. The assessment of non-natural substrate compounds was performed with a general reverse-phase gradient (i.e. 5 - 95%

organic). Reactions assessing the activity of WelO5 variants against (+)-12-*epi*-fischerindole U isonitrile were run on a more focussed gradient to achieve resolution between the starting material and chlorinated product.

UPLC-MS system set-up:

Waters ACQUITY® UPLC I-Class	
Column	Waters XBridge® BEH™ C18 2.5µm, (2.1mm x 50mm)
Flow rate	1 mL/min
Injection volume	3 µL
Column temperature	40°C
UV detection wavelength	210 to 400 nm
Solvent A	10 mM ammonium formate + 0.1% ammonia solution
Solvent B	acetonitrile + 5% solvent A + 0.1% ammonia solution

New components in the reaction mixtures were identified by subtracting the UV-chromatograms of the relevant enzyme-free and compound-free controls. The new peaks were then analysed by MS to aid product elucidation. The starting material peak and any new peaks in the UV-chromatogram were integrated and the reaction profile calculated from the relative percentage peak areas.

7.4.7.3.1. UPLC method gradient: turnover of WelO5 natural substrate

Time (min)	A%	B%
0.00	95.0	5.0
0.10	95.0	5.0
2.60	5.0	95.0
2.75	5.0	95.0
2.80	95.0	5.0
3.00	95.0	5.0

7.4.3.7.2. UPLC method gradient: alternative substrates

Time (min)	A%	B%
0.00	95.0	5.0
0.50	55.0	45.0
3.50	40.0	60.0
3.80	5.0	95.0
4.00	95.0	5.0

Chapter 8. References

1. Truppo, M. D., Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8* (5), 476-480.
2. Soh, L. M. J.; Mak, W. S.; Lin, P. P.; Mi, L.; Chen, F. Y.; Damoiseaux, R.; Siegel, J. B.; Liao, J. C., Engineering a thermostable keto acid decarboxylase using directed evolution and computationally directed protein design. *ACS Synth. Biol.* **2017**, *6* (4), 610-618.
3. Wang, J. B.; Li, G.; Reetz, M. T., Enzymatic site-selectivity enabled by structure-guided directed evolution. *Chem. Commun.* **2017**, *53* (28), 3916-3928.
4. Balke, K.; Bäumgen, M.; Bornscheuer, U. T., Controlling Regioselectivity of Baeyer-Villiger Monooxygenases by Mutations of Active Site Residues. *Chembiochem* **2017**, *18* (16), 1627-1638.
5. Romney, D. K.; Murciano-Calles, J.; Wehrmuller, J. E.; Arnold, F. H., Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **2017**, *139* (31), 10769-10776.
6. Wehrmann, M.; Klebensberger, J., Engineering thermal stability and solvent tolerance of the soluble quinoprotein PedE from *Pseudomonas putida* KT2440 with a heterologous whole-cell screening approach. *Microb. Biotechnol.* **2018**, *11* (2), 399-408.
7. Chen, K.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5618-5622.
8. Morley, K. L.; Kazlauskas, R. J., Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **2005**, *23* (5), 231-236.
9. Dourado, D. F. A. R.; Pohle, S.; Carvalho, A. T. P.; Dheeman, D. S.; Caswell, J. M.; Skvortsov, T.; Miskelly, I.; Brown, R. T.; Quinn, D. J.; Allen, C. C. R.; Kulakov, L.; Huang, M.; Moody, T. S., Rational Design of a (S)-Selective-Transaminase for Asymmetric Synthesis of (1S)-1-(1,1'-biphenyl-2-yl)ethanamine. *ACS Catal.* **2016**, *6* (11), 7749-7759.
10. Chica, R. A.; Doucet, N.; Pelletier, J. N., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **2005**, *16* (4), 378-384.
11. Porter, J. L.; Rusli, R. A.; Ollis, D. L., Directed Evolution of Enzymes for Industrial Biocatalysis. *Chembiochem* **2016**, *17* (3), 197-203.
12. Turner, N. J., Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **2009**, *5* (8), 567-573.
13. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015**, *16* (7), 379-394.

14. Sheldon, R. A.; Woodley, J. M., Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2017**, *118* (2), 801-838.
15. Anastas, P. T.; Warner, J. C., Principles of Green Chemistry. In *Green Chemistry: Theory and Practice*, Oxford University Press: New York, 1998; pp 29-54.
16. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, *329* (5989), 305-309.
17. He, J.-Y.; Sun, Z.-H.; Ruan, W.-Q.; Xu, Y., Biocatalytic synthesis of ethyl (S)-4-chloro-3-hydroxy-butanoate in an aqueous-organic solvent biphasic system using *Aureobasidium pullulans* CGMCC 1244. *Process Biochemistry* **2006**, *41* (1), 244-249.
18. Kiljunen, E.; Kanerva, L. T., Chloroperoxidase-catalysed oxidation of alcohols to aldehydes. *J. Mol. Catal. B: Enzym.* **2000**, *0*, 163-172.
19. Kaushik, N.; Biswas, S.; Singh, J., Biocatalysis and Biotransformation Processes – An Insight. *The Scitech Journal* **2014**, *1* (8), 15-22.
20. Raddadi, N.; Cherif, A.; Daffonchio, D.; Neifar, M.; Fava, F., Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl. Microbiol. Biotechnol.* **2015**, *99* (19), 7907-7913.
21. Elleuche, S.; Schroder, C.; Sahm, K.; Antranikian, G., Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotechnol.* **2014**, *29*, 116-123.
22. Ferrandi, E. E.; Previdi, A.; Bassanini, I.; Riva, S.; Peng, X.; Monti, D., Novel thermostable amine transferases from hot spring metagenomes. *Appl. Microbiol. Biotechnol.* **2017**, *101* (12), 4963-4979.
23. Karan, R.; Capes, M. D.; Dassarma, S., Function and biotechnology of extremophilic enzymes in low water activity. *Aquat. Biosyst.* **2012**, *8* (1), 4.
24. V., U.; S., B.; Noronha, S. B., Cofactor regeneration – an important aspect of biocatalysis. *Curr. Sci.* **2014**, *106* (7), 946-957.
25. Zhao, H.; van der Donk, W. A., Regeneration of cofactors for use in biocatalysis. *Curr. Opin. Biotechnol.* **2003**, *14* (6), 583-589.
26. Green, A. P.; Turner, N. J., Biocatalytic retrosynthesis: Redesigning synthetic routes to high-value chemicals. *Perspectives in Science* **2016**, *9*, 42-48.
27. Hayes, S. T.; Assaf, G.; Checksfield, G.; Cheung, C.; Critcher, D.; Harris, L.; Howard, R.; Mathew, S.; Regius, C.; Scotney, G.; Scott, A., Commercial Synthesis of (S,S)-Reboxetine Succinate: A Journey To Find the Cheapest Commercial Chemistry for Manufacture. *Org. Process Res. Dev.* **2011**, *15* (6), 1305-1314.

28. Li, T.; Liang, J.; Ambrogelly, A.; Brennan, T.; Gloor, G.; Huisman, G.; Lalonde, J.; Lekhal, A.; Mijts, B.; Muley, S.; Newman, L.; Tobin, M.; Wong, G.; Zaks, A.; Zhang, X., Efficient, chemoenzymatic process for manufacture of the Boceprevir bicyclic [3.1.0]proline intermediate based on amine oxidase-catalyzed desymmetrization. *J. Am. Chem. Soc.* **2012**, *134* (14), 6467-6472.
29. Shibasaki, T.; Mori, H.; Ozaki, A., Enzymatic production of trans-4-hydroxy-L-proline by regio- and stereospecific hydroxylation of L-proline. *Biosci., Biotechnol., Biochem.* **2014**, *64* (4), 746-750.
30. Panke, S.; Held, M.; Wubbolts, M., Trends and innovations in industrial biocatalysis for the production of fine chemicals. *Curr. Opin. Biotechnol.* **2004**, *15* (4), 272-279.
31. Choi, J. M.; Han, S. S.; Kim, H. S., Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnol. Adv.* **2015**, *33* (7), 1443-1454.
32. Desai, A. A., Sitagliptin manufacture: a compelling tale of green chemistry, process intensification, and industrial asymmetric catalysis. *Angew. Chem. Int. Ed.* **2011**, *50* (9), 1974-1976.
33. Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T., Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem. Rev.* **2006**, *106*, 3364-3378.
34. Murphy, C. D., New frontiers in biological halogenation. *J. Appl. Microbiol.* **2003**, *94* (4), 539-548.
35. Dong, C.; Huang, F.; Deng, H.; Schaffrath, C.; Spencer, J. B.; O'Hagan, D.; Naismith, J. H., Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **2004**, *427* (6974), 561-565.
36. Schaffrath, C.; Deng, H.; O'Hagan, D., Isolation and characterisation of 5'-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Letters* **2003**, *547* (1-3), 111-114.
37. Deng, H.; Cobb, S. L.; McEwan, A. R.; McGlinchey, R. P.; Naismith, J. H.; O'Hagan, D.; Robinson, D. A.; Spencer, J. B., The fluorinase from *Streptomyces cattleya* is also a chlorinase. *Angew. Chem. Int. Ed.* **2006**, *45* (5), 759-762.
38. Eustaquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S., Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **2008**, *4* (1), 69-74.
39. Smith, D. R.; Gruschow, S.; Goss, R. J., Scope and potential of halogenases in biosynthetic applications. *Curr. Opin. Biotechnol.* **2013**, *17* (2), 276-283.
40. Blasiak, L. C.; Drennan, C. L., Structural Perspective on Enzymatic Halogenation. *Acc. Chem. Res.* **2009**, *42* (1), 147-155.

41. Winter, J. M.; Moore, B. S., Exploring the chemistry and biology of vanadium-dependent haloperoxidases. *J. Biol. Chem.* **2009**, *284* (28), 18577-18581.
42. van Schijndel, J. W.; Vollenbroek, E. G.; Wever, R., The chloroperoxidase from the fungus *Curvularia inaequalis*; a novel vanadium enzyme. *Biochim. Biophys. Acta* **1993**, *1161* (2-3), 249-256.
43. van Pee, K. H.; Patallo, E. P., Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbiol. Biotechnol.* **2006**, *70* (6), 631-641.
44. Weichold, V.; Milbredt, D.; van Pee, K. H., Specific Enzymatic Halogenation-From the Discovery of Halogenated Enzymes to Their Applications in vitro and in vivo. *Angew. Chem. Int. Ed.* **2016**, *55* (22), 6374-6389.
45. Dong, C.; Flecks, S.; Unversucht, S.; Haupt, C.; van Pée, K. H.; Naismith, J. H., Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science* **2005**, *309* (5744), 2216-2219.
46. Zhu, X.; De Laurentis, W.; Leang, K.; Herrmann, J.; Ihlefeld, K.; van Pee, K. H.; Naismith, J. H., Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J Mol. Biol.* **2009**, *391* (1), 74-85.
47. Smith, J. L.; Khare, D., Recent Advances in the Structural and Mechanistic Biology of Non-Haem Fe(II), 2-Oxoglutarate and O₂-Dependent Halogenases. In *2-Oxoglutarate-Dependent Oxygenases*, Hausinger, R. P.; Schofield, C. J., Eds. Royal Society of Chemistry: 2015; pp 401-413.
48. Vaillancourt, F. H.; Vosburg, D. A.; Walsh, C. T., Dichlorination and bromination of a threonyl-S-carrier protein by the non-heme Fe(II) halogenase SyrB2. *Chembiochem* **2006**, *7* (5), 748-52.
49. Zhu, Q.; Hillwig, M. L.; Doi, Y.; Liu, X., Aliphatic Halogenase Enables Late-Stage C-H Functionalization: Selective Synthesis of a Brominated Fischerindole Alkaloid with Enhanced Antibacterial Activity. *Chembiochem* **2016**, *17* (6), 466-70.
50. Hillwig, M. L.; Liu, X., A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **2014**, *10* (11), 921-923.
51. Hillwig, M. L.; Qin, H. M.; Ittiamornkul, K.; Liu, X., Discovery of a Promiscuous Non-Heme Iron Halogenase in Ambiguine Alkaloid Biogenesis: Implication for an Evolvable Enzyme Family for Late-Stage Halogenation of Aliphatic Carbons in Small Molecules. *Angew. Chem. Int. Ed. Engl.* **2016**, *55* (19), 5780-5784.
52. Lang, A.; Polnick, S.; Nicke, T.; William, P.; Patallo, E. P.; Naismith, J. H.; van Pee, K. H., Changing the regioselectivity of the tryptophan 7-halogenase PrnA by site-directed mutagenesis. *Angew. Chem. Int. Ed.* **2011**, *50* (13), 2951-2953.

53. Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C., Directed Evolution of RebH for Catalyst-Controlled Halogenation of Indole C-H Bonds. *Chem. Sci.* **2016**, *7* (6), 3720-3729.
54. Shepherd, S. A.; Karthikeyan, C.; Latham, J.; Struck, A.-W.; Thompson, M. L.; Menon, B. R. K.; Styles, M. Q.; Levy, C.; Leys, D.; Micklefield, J., Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci.* **2015**, *6* (6), 3454-3460.
55. Menon, B. R. K.; Brandenburger, E.; Sharif, H. H.; Klemstein, U.; Shepherd, S. A.; Greaney, M. F.; Micklefield, J., RadH: A Versatile Halogenase for Integration into Synthetic Pathways. *Angew. Chem. Int. Ed.* **2017**, *56* (39), 11841-11845.
56. Fraley, A. E.; Garcia-Borras, M.; Tripathi, A.; Khare, D.; Mercado-Marin, E. V.; Tran, H.; Dan, Q.; Webb, G.; Watts, K.; Crews, P.; Sarpong, R.; Williams, R. M.; Smith, J. L.; Houk, K. N.; Sherman, D. H., Function and structure of MalA/MalA', iterative halogenases for late-stage C-H functionalization of indole alkaloids. *J. Am. Chem. Soc.* **2017**, *139* (34), 12060-12068.
57. Morrison, M.; Schonbaum, G. R., Peroxidase-catalyzed halogenation. *Annu. Rev. Biochem.* **1976**, *45*, 861-868.
58. Carter-Franklin, J. N.; Butler, A., Vanadium Bromoperoxidase-Catalyzed Biosynthesis of Halogenated Marine Natural Products. *J. Am. Chem. Soc.* **2004**, *126* (46), 15060-15066.
59. Zehner, S.; Kotzsch, A.; Bister, B.; Sussmuth, R. D.; Mendez, C.; Salas, J. A.; van Pee, K. H., A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* **2005**, *12* (4), 445-452.
60. Milbredt, D.; Patallo, E. P.; van Pée, K. H., A tryptophan 6-halogenase and an amidotransferase are involved in thienodolin biosynthesis. *Chembiochem* **2014**, *15* (7), 1011-1020.
61. Vaillancourt, F. H.; Yin, J.; Walsh, C. T., SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O₂-dependent halogenase. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (29), 10111-10116.
62. Deng, H.; O'Hagan, D., The fluorinase, the chlorinase and the duf-62 enzymes. *Curr. Opin. Chem. Biol.* **2008**, *12* (5), 582-92.
63. Aik, W.; McDonough, M. A.; Thalhammer, A.; Chowdhury, R.; Schofield, C. J., Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **2012**, *22* (6), 691-700.
64. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J., Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* **2006**, *100* (4), 644-669.

65. Kal, S.; Que, L., Dioxygen activation by nonheme iron enzymes with the 2-His-1-carboxylate facial triad that generate high-valent oxoiron oxidants. *J Biol. Inorg. Chem.* **2017**, *22* (2-3), 339-365.
66. Hegg, E. L.; Que JR, L., The 2-His-1-carboxylate facial triad. *Eur. J. Biochem.* **1997**, *250* (3), 625-629.
67. McDonough, M. A.; Loenarz, C.; Chowdhury, R.; Clifton, I. J.; Schofield, C. J., Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr. Opin. Struct. Biol.* **2010**, *20* (6), 659-672.
68. Bollinger Jr, J. M.; Chang, W. C.; Matthews, M. L.; Martinie, R. J.; Boal, A. K.; Krebs, C., Mechanisms of 2-Oxoglutarate-Dependent Oxygenases: The Hydroxylation Paradigm and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, Hausinger, R. P.; Schofield, C. J., Eds. Royal Society of Chemistry: 2015; pp 95-122.
69. Krebs, C.; Fujimori, D. G.; Walsh, C. T.; Bollinger Jr, J. M., Non-Heme Fe(IV)-Oxo Intermediates. *Acc. Chem. Res.* **2007**, *40* (7), 484-492.
70. Chang, W. C.; Guo, Y.; Wang, C. C.; Butch, S. E.; Rosenzweig, A. C.; K, B. A.; Krebs, C.; Bollinger Jr., J. M., Mechanism of the C5 Stereoinversion Reaction in the Biosynthesis of Carbapenem Antibiotics. *Science* **2014**, *343* (6175), 1140-1144.
71. Chang, W. C.; Li, J.; Lee, J. L.; Cronican, A. A.; Guo, Y., Mechanistic Investigation of a Non-Heme Iron Enzyme Catalyzed Epoxidation in (-)-4'-Methoxycyclophenin Biosynthesis. *J. Am. Chem. Soc.* **2016**, *138* (33), 10390-10393.
72. Henry, L.; Leung, I. K.; Claridge, T. D.; Schofield, C. J., γ -butyrobetaine hydroxylase catalyses a Stevens type rearrangement. *Bioorg. Med. Chem. Lett.* **2012**, *22* (15), 4975-4978.
73. Siitonen, V.; Selvaraj, B.; Niiranen, L.; Lindqvist, Y.; Schneider, G.; Metsä-Ketela, M., Divergent non-heme iron enzymes in the nogalamycin biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (19), 5251-5256.
74. Lloyd, M. D.; Lee, H.; Harlos, K.; Zhang, Z.; Baldwin, J. E.; Schofield, C. J.; Charnock, J. M.; Garner, C. D.; Hara, T.; Terwisscha van Scheltinga, A. C.; Valegard, K.; Viklund, J. A. C.; JHajdu, J.; Andersson, I.; Danielsson, A.; Bhikhabhai, R., Studies on the Active Site of Deacetoxcephalosporin C synthase. *J. Mol. Biol.* **1999**, *287*, 943-960.
75. Zhang, Z.; Ren, J.; Stammers, D. K.; Baldwin, J. E.; Harlos, K.; J, S. C., Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Bio.* **2000**, *7* (2), 127-133.
76. Zhang, Z.; Ren, J.; Harlos, K.; McKinnon, C. H.; Clifton, I. J.; Schofield, C. J., Crystal structure of a clavamate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centered rearrangements. *FEBS Lett.* **2002**, *517* (1-3), 7-12.

77. Rydzik, A. M.; Leung, I. K.; Kochan, G. T.; Loik, N. D.; Henry, L.; McDonough, M. A.; Claridge, T. D.; Schofield, C. J., Comparison of the substrate selectivity and biochemical properties of human and bacterial gamma-butyrobetaine hydroxylase. *Org. Biomol. Chem.* **2014**, *12* (33), 6354-6358.
78. Rydzik, A. M.; Leung, I. K.; Kochan, G. T.; McDonough, M. A.; Claridge, T. D.; Schofield, C. J., Oxygenase-catalyzed desymmetrization of N,N-dialkylpiperidine-4-carboxylic acids. *Angew. Chem. Int. Ed.* **2014**, *53* (41), 10925-10927.
79. Brauer, A.; Beck, P.; Hintermann, L.; Groll, M., Structure of the Dioxygenase AsqJ: Mechanistic Insights into a One-Pot Multistep Quinolone Antibiotic Biosynthesis. *Angew. Chem. Int. Ed.* **2016**, *55*, 422-426.
80. Mader, S. L.; Brauer, A.; Groll, M.; Kaila, V. R. I., Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. *Nat. Commun.* **2018**, *9* (1), 1168.
81. Matthews, M. L.; Neumann, C. S.; Miles, L. A.; Grove, T. L.; Booker, S. J.; Krebs, C.; Walsh, C. T.; Bollinger, J. M., Jr., Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (42), 17723-17728.
82. Martinie, R. J.; Livada, J.; Chang, W. C.; Green, M. T.; Krebs, C.; Bollinger, J. M., Jr.; Silakov, A., Experimental Correlation of Substrate Position with Reaction Outcome in the Aliphatic Halogenase, SyrB2. *J. Am. Chem. Soc.* **2015**, *137* (21), 6912-6919.
83. Matthews, M. L.; Chang, W. C.; Layne, A. P.; Miles, L. A.; Krebs, C.; Bollinger, J. M., Jr., Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **2014**, *10* (3), 209-215.
84. Hillwig, M. L.; Fuhrman, H. A.; Ittiarnornkul, K.; Sevco, T. J.; Kwak, D. H.; Liu, X., Identification and characterization of a welwitindolinone alkaloid biosynthetic gene cluster in the stigonematalean Cyanobacterium *Hapalosiphon welwitschii*. *ChemBioChem* **2014**, *15* (5), 665-669.
85. Hillwig, M. L.; Zhu, Q.; Liu, X., Biosynthesis of ambiguine indole alkaloids in cyanobacterium *Fischerella ambigua*. *ACS Chem. Biol.* **2014**, *9* (2), 372-377.
86. Fullone, M. R.; Paiardini, A.; Miele, R.; Marsango, S.; Gross, D. C.; Omura, S.; Ros-Herrera, E.; Bonaccorsi di Patti, M. C.; Lagana, A.; Pascarella, S.; Grgurina, I., Insight into the structure-function relationship of the nonheme iron halogenases involved in the biosynthesis of 4-chlorothreonine --Thr3 from *Streptomyces* sp. OH-5093 and SyrB2 from *Pseudomonas syringae* pv. *syringae* B301DR. *FEBS J.* **2012**, *279* (23), 4269-4282.
87. Galonic, D. P.; Vaillancourt, F. H.; Walsh, C. T., Halogenation of Unactivated Carbon Centers in Natural Product Biosynthesis: Trichlorination of Leucine during Barbamide Biosynthesis. *J. Am. Chem. Soc.* **2006**, *128* (12), 3900-3901.

88. Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T., Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **2005**, *436* (7054), 1191-4.
89. Khare, D.; Wang, B.; Gu, L.; Razelun, J.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L., Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (32), 14099-14104.
90. Ueki, M.; Galonic, D. P.; Vaillancourt, F. H.; Garneau-Tsodikova, S.; Yeh, E.; Vosburg, D. A.; Schroeder, F. C.; Osada, H.; Walsh, C. T., Enzymatic generation of the antimetabolite γ,γ -dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chem. Biol.* **2006**, *13* (11), 1183-1191.
91. Neumann, C. S.; Walsh, C. T., Biosynthesis of (-)-(1*S*,2*R*)-Allocoronamic Acyl Thioester by an Fe(II)-Dependent Halogenase and a Cyclopropane-Forming Flavoprotein. *J. Am. Chem. Soc.* **2008**, *130* (43), 14022-14023.
92. Ramaswamy, A. V.; Sorrels, C. M.; Gerwick, W. H., Cloning and Biochemical Characterization of the Hectochlorin Biosynthetic Gene Cluster from the Marine Cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2007**, *70* (12), 1977-1986.
93. Wong, C.; Fujimori, D. G.; Walsh, C. T.; Drennan, C. L., Structural Analysis of an Open Active Site Conformation of Nonheme Iron Halogenase CytC3. *J. Am. Chem. Soc.* **2009**, *131* (13), 4872-4879.
94. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403-410.
95. Zhang, Z.; Schaffer, A. A.; Miller, W.; Madden, T. L.; Lipman, D. J.; Koonin, E. V.; Altschul, S. F., Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* **1998**, *26* (17), 3986-3990.
96. Park, S.; Morley, K. L.; Horsman, G. P.; Holmquist, M.; Hult, K.; Kazlauskas, R. J., Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations. *Chem. Biol.* **2005**, *12* (1), 45-54.
97. Xiao, H.; Bao, Z.; Zhao, H., High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54* (16), 4011-4020.
98. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.* **2009**, *13* (1), 26-34.
99. Lukk, T.; Sakai, A.; Kalyanaraman, C.; Brown, S. D.; Imker, H. J.; Song, L.; Fedorov, A. A.; Fedorov, E. V.; Toro, R.; Hillerich, B.; Seidel, R.; Patskovsky, Y.; Vetting, M. W.; Nair, S. K.; Babbitt, P. C.; Almo, S. C.; Gerlt, J. A.; Jacobsen, M. P., Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (11), 4122-4127.

100. Suplatov, D. A.; Arzhanik, V. K.; Svedas, V. K., Comparative Bioinformatic Analysis of Active Site Structures in Evolutionarily Remote Homologues of α,β -Hydrolase Superfamily Enzymes. *Acta Naturae* **2011**, *3* (1), 93-98.
101. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10* (6), 845-858.
102. Kallberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J., Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **2012**, *7* (8), 1511-1522.
103. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296-W303.
104. Yang, J.; Zhang, Y., I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43* (W1), W174-W181.
105. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L., Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006**, *440* (7082), 368-371.
106. Li, W.; Cowley, A.; Uludag, M.; Gur, T.; McWilliam, H.; Squizzato, S.; Park, Y. M.; Buso, N.; Lopez, R., The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **2015**, *43* (W1), W580-W584.
107. Pearson, W. R., Selecting the Right Similarity-Scoring Matrix. *Curr. Protoc. Bioinf.* **2013**, *43*, 3.5.1-3.5.9.
108. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
109. Rost, B., Twilight zone of protein sequence alignments. *Protein Eng.* **1999**, *12* (2), 85-94.
110. Khor, B. Y.; Tye, G. J.; Lim, T. S.; Choong, Y. S., General overview on structure prediction of twilight-zone proteins. *Theor. Biol. Med. Modell.* **2015**, *12*, 15.
111. Xiang, Z., Advances in homology protein structure modeling. *Curr. Protein Pept. Sci.* **2006**, *7* (3), 217-227.
112. Dorn, M.; MB, E. S.; Buriol, L. S.; Lamb, L. C., Three-dimensional protein structure prediction: Methods and computational strategies. *Comput. Biol. Chem.* **2014**, *53PB*, 251-276.

113. Moulton, J., A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **2005**, *15* (3), 285-289.
114. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-242.
115. Wolf, Y. I.; Grishin, N. V.; Koonin, E. V., Estimating the Number of Protein Folds and Families from Complete Genome Data. *J. Mol. Biol.* **2000**, *299* (4), 897-905.
116. Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C., SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* **1995**, *247* (4), 536-540.
117. Dawson, N. L.; Lewis, T. E.; Das, S.; Lees, J. G.; Lee, D.; Ashford, P.; Orengo, C. A.; Sillitoe, I., CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* **2017**, *45* (D1), D289-D295.
118. Hibi, M.; Kawashima, T.; Kasahara, T.; Sokolov, P. M.; Smirnov, S. V.; Koder, T.; Sugiyama, M.; Shimizu, S.; Yokozeki, K.; Ogawa, J., A novel Fe(II)/ α -ketoglutarate-dependent dioxygenase from *Burkholderia ambifaria* has β -hydroxylating activity of N-succinyl L-leucine. *Lett. Appl. Microbiol.* **2012**, *55* (6), 414-419.
119. Qin, H. M.; Miyakawa, T.; Jia, M. Z.; Nakamura, A.; Ohtsuka, J.; Xue, Y. L.; Kawashima, T.; Kasahara, T.; Hibi, M.; Ogawa, J.; Tanokura, M., Crystal structure of a novel N-substituted L-amino acid dioxygenase from *Burkholderia ambifaria* AMMD. *PLoS One* **2013**, *8* (5), e63996.
120. Shehu, A.; Kavraki, L. E., Modeling Structures and Motions of Loops in Protein Molecules. *Entropy* **2012**, *14* (2), 252-290.
121. Aslanidis, C.; de Jong, P. J., Ligation-independent cloning of PCR products. *Nucleic Acids Res.* **1990**, *18* (20), 6069-6074.
122. Savitsky, P.; Bray, J.; Cooper, C. D.; Marsden, B. D.; Mahajan, P.; Burgess-Brown, N. A.; Gileadi, O., High-throughput production of human proteins for crystallization: the SGC experience. *J. Struct. Biol.* **2010**, *172* (1), 3-13.
123. Yee, L.; Blanch, H. W., Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Nat. Biotechnol.* **1992**, *10* (12), 1550-1556.
124. Hochuli, E.; Bannwarth, W.; Dobeli, H.; Gentz, R.; Stuber, D., Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Nat. Biotechnol.* **1988**, *6* (11), 1321-1325.
125. Affinity Chromatography, Vol. 2: Tagged Proteins. GE Healthcare: Vol. 18114275.

126. Parks, T. D.; Leuther, K. K.; Howard, E. D.; Johnston, S. A.; Dougherty, W. G., Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* **1994**, *216* (2), 413-417.
127. Lathe, G. H.; Ruthven, C. R. J., The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water. *Biochem. J.* **1956**, *62* (4), 665-674.
128. Size Exclusion Chromatography: Principles and Methods. GE Healthcare: Vol. 18102218.
129. Berrow, N. S.; Alderton, D.; Sainsbury, S.; Nettleship, J.; Assenberg, R.; Rahman, N.; Stuart, D. I.; Owens, R. J., A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **2007**, *35* (6), e45.
130. Lacks, S.; Greenberg, B., A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. *J. Biol. Chem.* **1975**, *250* (11), 4060-4066.
131. Stratmann, K.; Moore, R. E.; Bonjouklian, R.; Deeter, J. B.; Patterson, G. M. L.; Shaffer, S.; Smith, C. D.; Smitka, T. A., Welwitindolinones, Unusual Alkaloids from the Blue-Green Algae *Hapalosiphon welwitschii* and *Westiella intricata*. Relationship to Fischerindoles and Hapalindoles. *J. Am. Chem. Soc.* **1994**, *116* (22), 9935-9942.
132. Moore, R. E.; Cheuk, C.; Patterson, G. M. L., Hapalindoles: New Alkaloids from the Blue-Green Alga *Hapalosiphon fontinalis*. *J. Am. Chem. Soc.* **1984**, *106* (21), 6456-6457.
133. Park, A.; Moore, R. E.; Patterson, G. M. L., Fischerindole L, a new isonitrile from the terrestrial blue-green alga *fischerella muscicola*. *Tet. Lett.* **1992**, *33* (23), 3257-3260.
134. Richter, J. M.; Baran, P. S., Direct Coupling of Indoles with Carbonyl Compounds: Short, Enantioselective, Gram-Scale Synthetic Entry into the Hapalindole and Fischerindole Alkaloid Families. *J. Am. Chem. Soc.* **2004**, *126* (24), 7450-7451.
135. Richter, J. M.; Whitefield, B. W.; Maimone, T. J.; Lin, D. W.; Castroviejo, M. P.; Baran, P. S., Scope and Mechanism of Direct Indole and Pyrrole Couplings Adjacent to Carbonyl Compounds: Total Synthesis of Acremoauxin A and Oxazinin 3. *J. Am. Chem. Soc.* **2007**, *129* (42), 12857-12869.
136. Sigma-Aldrich, Product Specification: S-(+)-Carvone, 96% (product number 435759).
137. Sigma-Aldrich, Product Specification: R-(-)-Carvone, 98% (product number 124931).
138. Roden, B. A., Diphenylbis(1,1,1,3,3,3-hexafluoro-2-phenyl-2-propoxy)sulfurane. In *e-EROS: Encyclopedia of Reagents for Organic Synthesis*, John Wiley & Sons: 2001.

139. Richter, J. M.; Ishihara, Y.; Masuda, T.; Whitefield, B. W.; Llamas, T.; Pohjakallio, A.; Baran, P. S., Enantiospecific Total Synthesis of the Hapalindoles, Fischerindoles, and Welwitindolinones *via* a Redox Economic Approach. *J. Am. Chem. Soc.* **2008**, *130* (52), 17938-17954.
140. Fukuyama, T.; Chen, X., Stereocontrolled Synthesis of (-)-Hapalindole G. *J. Am. Chem. Soc.* **1994**, *116* (7), 3125-3126.
141. Richter, J. M.; Baran, P. S., Enantioselective Total Syntheses of Welwitindolinone A and Fischerindoles I and G. *J. Am. Chem. Soc.* **2005**, *127* (44), 15394-15396.
142. Bhattacharyya, S., Titanium(IV) isopropoxide and sodium borohydride : A reagent of choice for reductive amination. *Tet. Lett.* **1994**, *35* (15), 2401-2404.
143. Meyer, J.; Staudinger, H., Über neue organische phosphorverbindungen III. Phosphinmethylenderivate und phosphinimine. *Helv. Chim. Acta* **1919**, *2*, 635-646.
144. Guchait, S. K.; Priyadarshani, G.; Chaudhary, V.; Seladiya, D. R.; Shah, T. M.; Bhogayta, N. P., One-pot preparation of isocyanides from amines and their multicomponent reactions: crucial role of dehydrating agent and base. *RSC Advances* **2013**, *3*, 10867-10874.
145. De Luca, L.; Giacomelli, G.; Porcheddu, A.; Salaris, M., A New, Simple Procedure for the Synthesis of Formyl Amides. *Synlett* **2004**, *14*, 2570-2572.
146. Krishnamurthy, S., A highly efficient and general N-monomethylation of functionalized primary amines *via* formylation-borane methyl sulfide reduction. *Tett. Lett.* **1982**, *23* (33), 3315-3318.
147. Thomas, J. O., The preparation of N-formyl derivatives of amino acids using N,N-dicyclohexylcarbodiimide. *Tett. Lett.* **1967**, *4*, 335-336.
148. Kajanus, J.; Jacobson, I.; Åstrand, A.; Olsson, R. I.; Gran, U.; Björe, A.; Fjellström, O.; Davidsson, Ö.; Emtenäs, H.; Dahlén, A.; Löfberg, B.; Yuan, Z. Q.; Sundell, J.; Cassel, J.; Gyll, J.; Iliefski, T.; Högberg, Å.; Lindhardt, E.; Malmberg, J., Isoindolinone compounds active as Kv1.5 blockers identified using a multicomponent reaction approach. *Bioorg. Med. Chem. Lett.* **2016**, *26* (8), 2023-2029.
149. Gerack, C. J.; McElwee-White, L., Formylation of Amines. *Molecules* **2014**, *19* (6), 7689-7713.
150. Pascale, T.; Mobashery, S.; Hart, A., (Methoxycarbonylsulfamoyl)triethylammonium Hydroxide In *e-EROS: Encyclopedia of Reagents for Organic Synthesis*, John Wiley & Sons: 2008.
151. Creedon, S. M.; Crowley, H. K.; McCarthy, D. G., Dehydration of formamides using the Burgess Reagent: a new route to isocyanides. *J. Chem. Soc., Perkin Trans. 1* **1998**, (6), 1015-1018.

152. Reisman, S. E.; Ready, J. M.; Hasuoka, A.; Smith, C. J.; Wood, J. L., Total Synthesis of (\pm)-Welwitindolinone A Isonitrile. *J. Am. Chem. Soc.* **2006**, *128* (5), 1448-1449.
153. Tang, Q.; Chen, X.; Tiwari, B.; Chi, Y. R., Addition of Indoles to Oxyallyl Cations for Facile Access to α -Indole Carbonyl Compounds. *Org. Lett.* **2012**, *14* (7), 1922-1925.
154. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvand, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press: 2005.
155. Folta-Stogniew, E., Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods Mol. Biol.* **2006**, *328*, 97-112.
156. Zulauf, M.; D'Arcy, A., Light Scattering of Proteins as a Criterion for Crystallization. *J. Cryst. Growth* **1992**, *122* (1-4), 102-106.
157. Martin, S. R.; Schilstra, M. J., *Circular dichroism and its application to the study of biomolecules*. 2008; Vol. 84, p 263-293.
158. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004**, *32*, W668-673.
159. Johnson, W. C., Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Struct. Funct. Genet.* **1999**, *35* (3), 307-312.
160. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22* (12), 2577-2637.
161. Pavel, E. G.; Zhou, J.; Busby, R. W.; Gunsior, M.; Townsend, C. A.; Solomon, E. I., Circular Dichroism and Magnetic Circular Dichroism Spectroscopic Studies of the Non-Heme Ferrous Active Site in Clavaminase Synthase and Its Interaction with α -Ketoglutarate Cosubstrate. *J. Am. Chem. Soc.* **1998**, *120*, 743-753.
162. Ryle, M. J.; Padmakumar, R.; Hausinger, R. P., Stopped-Flow Kinetic Analysis of *Escherichia coli* Taurine/ α -Ketoglutarate Dioxygenase: Interactions with α -Ketoglutarate, Taurine, and Oxygen. *Biochemistry* **1999**, *38*, 15278-15286.
163. Pratter, S. M.; Ivkovic, J.; Birner-Gruenberger, R.; Breinbauer, R.; Zangger, K.; Straganz, G. D., More than just a halogenase: modification of fatty acyl moieties by a trifunctional metal enzyme. *Chembiochem* **2014**, *15* (4), 567-74.
164. Mitchell, A. J.; Zhu, Q.; Maggiolo, A. O.; Ananth, N. R.; Hillwig, M. L.; Liu, X.; Boal, A. K., Structural basis for halogenation by iron- and 2-oxoglutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **2016**, *12* (8), 636-640.

165. Houk, R. S.; Fassel, V. A.; Flesch, G. D.; Svec, H. J., Inductively Coupled Argon Plasma as an Ion Source for Mass Spectrometric Determination of Trace Elements. *Anal. Chem.* **1980**, *52* (14), 2283-2289.
166. Beauchemin, D., Inductively Coupled Plasma Mass Spectrometry. *Anal. Chem.* **2006**, *78* (12), 4111-4135.
167. Dupeux, F.; Rower, M.; Seroul, G.; Blot, D.; Marquez, J. A., A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67* (11), 915-919.
168. Hawe, A.; Sutter, M.; Jiskoot, W., Extrinsic fluorescent dyes as tools for protein characterization. *Pharm. Res.* **2008**, *25* (7), 1487-1499.
169. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007**, *2* (9), 2212-2221.
170. Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P., Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **2006**, *357* (2), 289-98.
171. Cunliffe, C. J.; Franklin, T. J.; Hales, N. J.; Hill, G. B., Novel Inhibitors of Prolyl 4-Hydroxylase. 3.1 Inhibition by the Substrate Analogue N-Oxaloglycine and Its Derivatives. *J. Med. Chem.* **1992**, *35* (14), 2652-2658.
172. Yeoh, K. K.; Chan, M. C.; Thalhammer, A.; Demetriades, M.; Chowdhury, R.; Tian, Y. M.; Stolze, I.; McNeill, L. A.; Lee, M. K.; Woon, E. C. Y.; Mackeen, M. M.; Kawamura, A.; Ratcliffe, P. J.; Mecinovic, J.; Schofield, C. J., Dual-action inhibitors of HIF prolyl hydroxylases that induce binding of a second iron ion. *Org. Biomol. Chem.* **2013**, *11* (5), 732-745.
173. Rose, N. R.; Woon, E. C.; Kingham, G. L.; King, O. N.; Mecinovic, J.; Clifton, I. J.; Ng, S. S.; Talib-Hardy, J.; Oppermann, U.; McDonough, M. A.; Schofield, C. J., Selective inhibitors of the JMJD2 histone demethylases: combined nondenaturing mass spectrometric screening and crystallographic approaches. *J. Med. Chem.* **2010**, *53* (4), 1810-1818.
174. Ivan, M.; Haberberger, T.; Gervasi, D. C.; Michelson, K. S.; Gunzler, V.; Kondo, K.; Yang, H.; Sorokina, I.; Conaway, R. C.; Conaway, J. W.; Kaelin, W. G., Jr., Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (21), 13459-13464.
175. Dauter, Z.; Dauter, M.; Rajashankar, K. R., Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *D56*, 232-237.
176. Beck, T.; Krasauskas, A.; Gruene, T.; Sheldrick, G. M., A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *D64*, 1179-1182.

177. Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M., Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **1990**, *9* (5), 1665-1672.
178. Walden, H., Selenium incorporation using recombinant techniques. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *D66*, 352-357.
179. Yuan, Z.; Zhao, J.; Wang, Z., Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng., Des. Sel.* **2003**, *16* (2), 109-114.
180. Mitchell, A. J.; Dunham, N. P.; Bergman, J. A.; Wang, B.; Zhu, Q.; Chang, W. C.; Liu, X.; Boal, A. K., Structure-Guided Reprogramming of a Hydroxylase To Halogenate Its Small Molecule Substrate. *Biochemistry* **2017**, *56* (3), 441-444.
181. Yee, Y. A.; Savchenko, A.; Ignachenko, A.; Lukin, J.; Xu, X.; Skarina, T.; Evdokimova, E.; Liu, C. S.; Semesi, A.; Guido, V.; Edwards, A. M.; Arrowsmith, C. H., NMR and X-ray Crystallography, Complementary Tools in Structural Proteomics of Small Proteins. *J. Am. Chem. Soc.* **2005**, *127*, 16512-16517.
182. Markwick, P. R.; Malliavin, T.; Nilges, M., Structural biology by NMR: structure, dynamics and interactions. *PLoS Comput. Biol.* **2008**, *4* (9), e1000168.
183. Kleckner, I. R.; Foster, M. P., An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta, Proteins Proteomics* **2011**, *1814* (8), 942-968.
184. Cala, O.; Guilliere, F.; Krimm, I., NMR-based analysis of protein-ligand interactions. *Anal. Bioanal. Chem.* **2014**, *406* (4), 943-956.
185. Peruvshin, K.; Riek, R.; Wider, G.; Wüthrich, K., Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366-12371.
186. van Den Heuvel, R. H.; Fraaije, M. W.; Ferrer, M.; Mattevi, A.; van Berkel, W. J., Inversion of stereospecificity of vanillyl-alcohol oxidase. *Proc. Natl. Acad. Sci. USA* **2000**, *97* (17), 9455-9460.
187. Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H., Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015**, *1* (2), 89-93.
188. Acebes, S.; Fernandez-Fueyo, E.; Monza, E.; Lucas, M. F.; Almendral, D.; Ruiz-Dueñas, F. J.; Lund, H.; Martinez, A. T.; Guallar, V., Rational Enzyme Engineering Through Biophysical and Biochemical Modeling. *ACS Catal.* **2016**, *6* (3), 1624-1629.

189. Balke, K.; Baumgen, M.; Bornscheuer, U. T., Controlling the Regioselectivity of Baeyer–Villiger Monooxygenases by Mutation of Active-Site Residues. *ChemBioChem* **2017**, *18* (16), 1627-1638.
190. Li, G.; Maria-Solano, M. A.; Romero-Rivera, A.; Osuna, S.; Reetz, M. T., Inducing high activity of a thermophilic enzyme at ambient temperatures by directed evolution. *Chem. Commun.* **2017**, *53* (68), 9454-9457.
191. Varadarajan, N.; Gam, J.; Olsen, M. J.; Georgiou, G.; Iverson, B. L., Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (19), 6855-6860.
192. Sun, Z.; Lonsdale, R.; Wu, L.; Li, G.; Li, A.; Wang, J.; Zhou, J.; Reetz, M. T., Structure-Guided Triple-Code Saturation Mutagenesis: Efficient Tuning of the Stereoselectivity of an Epoxide Hydrolase. *ACS Catal.* **2016**, *6* (3), 1590-1597.
193. Alexeeva, M.; Enright, A.; Dawson, M. J.; Mahmoudian, M.; Turner, N. J., Deracemization of α -Methylbenzylamine Using an Enzyme Obtained by In Vitro Evolution. *Angew. Chem. Int. Ed. Engl.* **2002**, *2002* (41), 17.
194. Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B., Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **2015**, *44* (5), 1172-1239.
195. Halweg-Edwards, A. L.; Pines, G.; Winkler, J. D.; Pines, A.; Gill, R. T., A Web Interface for Codon Compression. *ACS Synth. Biol.* **2016**, *5* (9), 1021-1023.
196. Nomenclature Committee of the International Union of Biochemistry (NC-IUB). Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *Eur. J. Biochem.* **1985**, *150* (1), 1-5.
197. Reetz, M. T.; Prasad, S.; Carballeira, J. D.; Gumulya, Y.; Bocola, M., Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity. *J. Am. Chem. Soc.* **2010**, *132* (26), 9144-9152.
198. Reetz, M. T.; Carballeira, J. D., Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007**, *2* (4), 891-903.
199. Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A., Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chem. Int. Ed. Engl.* **2005**, *44* (27), 4192-4196.
200. Sun, Z.; Lonsdale, R.; Kong, X. D.; Xu, J. H.; Zhou, J.; Reetz, M. T., Reshaping an enzyme binding pocket for enhanced and inverted stereoselectivity: use of smallest amino acid alphabets in directed evolution. *Angew. Chem. Int. Ed. Engl.* **2015**, *54* (42), 12410-12415.
201. Parra, L. P.; Agudo, R.; Reetz, M. T., Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *ChemBiochem* **2013**, *14* (17), 2301-2309.

202. Reetz, M. T.; Carballeira, J. D.; Peyralans, J.; Hobenreich, H.; Maichele, A.; Vogel, A., Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chem. Eur. J.* **2006**, *12* (23), 6031-8.
203. Sun, Z.; Wikmark, Y.; Backvall, J. E.; Reetz, M. T., New Concepts for Increasing the Efficiency in Directed Evolution of Stereoselective Enzymes. *Chem. Eur. J.* **2016**, *22* (15), 5046-5054.
204. Koketsu, K.; Shomura, Y.; Moriwaki, K.; Hayashi, M.; Mitsunashi, S.; Hara, R.; Kino, K.; Higuchi, Y., Refined regio- and stereoselective hydroxylation of L-pipecolic acid by protein engineering of L-proline cis-4-hydroxylase based on the X-ray crystal structure. *ACS Synth. Biol.* **2015**, *4* (4), 383-392.
205. Novagen Overnight Express™ Autoinduction Systems 1 and 2 User Protocol TB383 Rev. I 0209.
206. Sigma-Aldrich HIS-Select Filter Plate Technical Bulletin, Catalog Number H0413.
207. Cardillo, A. B.; Talou, J. R.; Giulietti, A. M., Expression of *Brugmansia candida* Hyoscyamine 6beta-Hydroxylase gene in *Saccharomyces cerevisiae* and its potential use as biocatalyst. *Microb. Cell Fact.* **2008**, *7*, 17.
208. Li, J.; van Belkum, M. J.; Vederas, J. C., Functional characterization of recombinant hyoscyamine 6 β -hydroxylase from *Atropa belladonna*. *Bioorg. Med. Chem.* **2012**, *20* (14), 4356-4363.
209. Pei, J.; Kim, B. H.; Grishin, N. V., PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **2008**, *36* (7), 2295-300.