Biocatalytic Imine Reduction: Isolation, Application, Structure and Mechanism of Imine Reductases

Elizabeth Rachel Wells

PhD

University of York

Chemistry

May 2017

Abstract

Chiral amines are industrially useful chemicals found in the pharmaceutical, chemical and agrochemical industries. While many excellent methods have been developed for their synthesis using abiotic catalysis, these are often reliant on precious metals the global supply of which is becoming increasingly scarce. Additionally, those catalysts often require harsh, non-environmentally friendly reaction conditions such as high temperatures and pressures.

Biocatalytic imine reduction, which involves the asymmetric reduction of imines to form chiral amines using imine reductases (IREDs), offers an efficient and sustainable synthesis of chiral amines which overcomes many of the limitations encountered in abiotic catalysis.

Two NADPH-dependent IREDs were expressed and purified; *SkR*-IRED from *Streptomyces kanamyceticus* and *SS*-IRED from *Streptomyces* sp. GF3546, which catalyse the reduction of the model imine 2-methylpyrroline to (*R*)- and (*S*)- amine products respectively.

The *SkR*-IRED monomer consists of an *N*-terminal Rossman fold motif and a *C*-terminal helical domain. *SkR*-IRED exists as a homodimer of two monomers which are linked by an unusual reciprocal domain sharing arrangement. The structure of *SkR*-IRED was used as a basis to study mechanism using mutagenesis experiments which suggested that residue Asp187 may be important for catalysis.

The structure of *SS*-IRED was solved to a resolution of 3.2 Å in its *apo* form and revealed that the enzyme shares the structural features of *SkR*-IRED. Mutagenesis experiments suggested that residue Tyr169 is necessary for catalysis.

Last, a novel IRED from the moss *Physcomitrella patens* (*Pp*-IRED) was expressed and purified. *Pp*-IRED was active towards the amine *N*-methyl-1-quinolin-6-ylmethanamine in the oxidative direction, the first IRED active towards this substrate. The structure of *Pp*-IRED was solved in both its *apo* form and in complex with NADPH (2.5 Å resolution). The structure was distinct from other IREDs as it did not display the reciprocal domain sharing arrangement seen in *SkR*-IRED and *SS*-IRED.

Table of Contents

Abstract i

Table of Contents ii

Author's Declaration vii

List of Figures viii

List of Tables xxii

Acknowledgements xxv

1. Introduction 1

1.1 Biocatalysis 1

1.1.1 The history of biocatalysis 1

1.1.2 The advantages of biocatalysis 2

1.2 Chiral amines 3

1.3 Chemical synthesis of chiral amines 4

1.3.1 Reduction of enamides 5

1.3.2 Reductive amination 6

1.3.3 Reduction of imines 8

1.3.4 Limitations of chemical chiral amine synthesis 8

1.4 Biocatalytic routes towards chiral amines 9

1.4.1 Lipases 10

1.4.2 Transaminases 12

1.4.3 Amine oxidases 14

1.4.4 Amine dehydrogenases 17

1.4.5 Imine reductases (IREDs) 19

1.5 Project aims 25

1.6 Addendum to the Introduction 26

2. Materials and Methods 30

2.1 Materials 30
2.1.1 Bacterial growth media 30
2.1.2 Buffers for protein purification and enzyme assays 30
2.1.3 Gel electrophoresis buffers 32
2.1.4 Substrates and products 35
2.2 Molecular cloning 38
2.2.1 Polymerase chain reaction 40
2.2.2 Ligation-independent cloning (LIC) 43
2.3 Protein production 46
2.3.1 Transformation 46
2.3.2 Small scale expression tests 46
2.3.3 Large scale protein production 47
2.4 Protein purification 49
2.4.1 Cell lysis 49
2.4.2 Nickel affinity chromatography 49
2.4.3 Size exclusion chromatography (SEC) 50
2.4.4 Protein concentration 52
2.4.5 Protein storage 52
2.5 6His tag cleavage 52
2.6 Analytical techniques 53
2.6.1 Ultraviolet-visible (UV) spectrophotometry 53
2.6.2 HPLC 54
2.6.3 GC 56
2.7 Enzyme kinetics 57
2.7.1 Michaelis-Menten kinetics 57
2.7.2 Substrate inhibition kinetics 59
2.7.3 NADPH concentration assays 60

2.7.4 Data analysis 62
2.8 Biotransformations 62
2.8.1 Biotransformations by SkR-IRED 62
2.8.4 Biotransformations using mutants of SS-IRED and SR-IRED 64
2.8.5 Biotransformations using mutants of Pp-IRED 65
2.9 Site-directed mutagenesis (SDM) 66
2.9.1 PCR 68
2.9.2 DpnI Digest 69
2.9.3 Transformation 69
2.9.4 DNA isolation 69
2.9.5 DNA Sequencing 69
2.10 Sample preparation for electrospray mass spectrometry 69
3. Isolation, application, structure and mechanism of SkR-IRED, an (R)-selective NADPH-dependent imine reductase from Streptomyces kanamyceticus 70
3.1 Introduction 70
3.2 Preliminary work 71
3.2.1 Crystallization and data collection 71
3.2.2 Structure solution 72
3.3 Aims 74
3.4 Results and discussion 74
3.4.1 Large scale expression and purification 74
3.4.2 Kinetic analysis and pH behaviour of SkR-IRED 77
3.4.3 Biotransformations using SkR-IRED 79
3.4.5 Mechanism proposal for SkR-IRED 81
3.4.6 Investigations towards the proposed mechanism of SkR-IRED 85
3.4.7 Investigation of physiological role of SkR-IRED 88
3.5 Conclusion 95

4. Isolation, application, structure and mechanism of SS-IRED, an (S)-selective NADPH-dependent imine reductase from Streptomyces sp. GF3546 96

4.1 Introduction 96

4.2 Aims 96

4.3 Results 97

4.3.1 Amplification of target genes 97

4.3.2 Subcloning into pET-YSBLIC3C expression vector 97

4.3.3 Large scale expression and purification 98

4.3.4 Cleavage of 6His tag 101

4.4.5 Structure determination 103

4.4.6 SS-IRED+, a modified protein for the crystallization of SS-IRED 112

4.4.7 Kinetic analysis and pH behaviour of SS-IRED 119

4.4.8 Proposing a mechanism of SS-IRED 120

4.4.9 Mutational studies towards the mechanism of SS-IRED 124

4.4.10 Conclusion 130

5. Investigating the basis for enantioselectivity of NADPH-dependent IREDs 132

5.1 Introduction 132

5.2 Aims 132

5.2 Comparison of the active site of an (R)- and (S)-selective IRED 133

5.3 Analysis of bacterial protein-sequence space of imine reductases for the basis of enantioselectivity 140

5.4 Discussion 152

5.5 Conclusion 154

6. Characterization of an PpIRED: an Imine Reductase from Physcomitrella patens 156

6.1 Introduction 156

6.2 Aims 156

6.3 Results 157
6.3.1 Expression and purification of PpIRED 157
6.4 Crystal structure of PpIRED in apo form 159
6.5 Crystal structure of PpIRED in complex with $NADP^+$ 165
6.6 Kinetic analysis of PpIRED activity towards amines, ketones and alcohols 168
6.7 GC analysis of biotransformations of quinoline derivatives by PpIRED 175
6.8 Conclusion 179
7. Conclusion 180
Appendix 183
Publications relating to thesis 183
Primer sequences 183
Glossary of Abbreviations and Symbols 186
References 188

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

List of Figures

Figure 1.1: The principles of green chemistry. The use of enzymes is particularly relevant for certain green chemistry criteria. Their use reduces the need for auxiliary substances (i.e. organic solvents). They are also energy efficient, operating optimally at atmospheric temperature and pressure.[4] 2

Figure 1.2: An array of pharmaceutical products containing optically active amine functional groups including diabetes, Alzheimer's and malaria treatments. The image illustrates the high demand for chiral amines in the pharmaceutical industry.[8]..... 4

Figure 1.3: Synthetic routes towards primary chiral amines from N-acetylenamides. Synthesis of the enamide occurs using either a nitrile or ketone starting material. The enamide is then asymmetrically reduced with hydride using a Rhodium catalyst.[8].... 6

Figure 1.4: The first example of an enantioselective reductive amination: a synthesis of an active ingredient in commercial herbicides, metolachlor. The reaction involves the conjugation of methoxyacetone with 2-methyl-6-ethylaniline to form an intermediary imine in situ, before reduction using H_2 in the presence of an Ir-complexed chiral catalyst, Xyliphos.[15] 7

Figure 1.5: The synthesis of (S)-metolachlor using an asymmetric reduction step. Chirality is introduced by use of an asymmetric catalyst; the Ir-xyliphos (an iridium-complexed iron-containing chiral phosphine complex). 8

Figure 1.6: Enzymatic routes for the production of chiral amines. Current methods include kinetic resolution of racemic amines by hydrolases, dynamic kinetic resolution (DKR) and deracemization of racemic amines by hydrolases and oxidases, asymmetric amination of ketones by transaminases, decarboxylation of carboxylic acids by decarboxylases and the asymmetric reduction of imines by imine reductases (IREDs).[20] 10

Figure 1.7: A general illustration of the lipase-catalysed reversible acetylation of an amide substrate to form the corresponding enantiomerically pure amine alongside the amide with opposite chirality.[20] 11

Figure 1.8: Dynamic kinetic resolution of a racemic amine by the selective acylation of the (R)-amine by a lipase and the racemisation of the unreacted (S)-amine using an achiral Pd/C catalyst.[26] 11

Figure 1.9: The two-step separation of racemic phenylethylamine to both the (S)- and (R)-phenylethylamine in their enantiomerically pure forms. The (R)-amine is initially

converted to the corresponding amide by a lipase from P. plantarii. The (R)-amide product is subsequently separated by distillation and converted back to the (R)-amine by basic hydrolysis. The optically pure unreacted (S)-amine remains in the reaction mixture after removal of the (R)-amide by distillation.[27] 12
Figure 1.10: The general scheme of a reaction catalysed by a transaminase. An amino group is transferred from the pyridoxamine cofactor which acts as an amino donor to a ketone or aldehyde substrate, which facilitates the formation of a chiral amine when performed asymmetrically.[20] 13
Figure 1.11: A biocatalytic route towards the synthesis of Januvia®, an antidiabetic drug used to lower the blood sugar level in patients, using a mutant rationally evolved from transaminase ATA-117. 14
Figure 1.12: A general reaction catalysed by a flavin-dependent amine oxidase. An imine is reversible oxidised to an amine with concurrent reduction of oxygen to hydrogen peroxide.[20] 15
Figure 1.13: The deracemization of primary, secondary or tertiary amines by MAO-N variants with concurrent achiral reduction of the imine product back to the racemized amine starting material. The amine enantiomer which is not converted by MAO-N accumulates in the reaction mixture with increasing levels of optical purity.[39] 16
Figure 1.14: A biocatalytic step in the synthesis of Telaprevir, a Hepatitis C viral protease inhibitor, mediated by an engineered variant of MAO-N from Aspergillus niger.[41] 17
Figure 1.15: The reactions catalysed by amino acid dehydrogenases (AADHs) (top) and amine dehydrogenases (AmDHs) (bottom). AADHs catalyse the NADH-dependent interconversion of keto-acids and amino acids while AmDHs catalyse the interconversion of ketones and amines. 18
Figure 1.16: A cascade reaction employing an ADH and AmDH to convert racemic alcohols to chiral amines using NAD(H) in a single reaction vessel. 19
Figure 1.17: Imine reduction as a route for the synthesis of amines. When done asymmetrically, such as by using an enzyme, the reduction of imines (which are readily obtainable from ketones) can be used to generate optically pure amines. 20
Figure 1.18: Reduction of imine moiety of dihydrofolate by DHFR to form tetrahydrofolate; an intermediary reaction in the biosynthesis of DNA.[44, 45] 21

Figure 1.19: The reduction of the flavin moiety of the enzyme cofactor F_{420} to produce the reduced form, $F_{420}H_2$.[46-48] 21
Figure 1.20: The role of F420-dependent IREDs TomJ and SibT in the biosynthesis of PBDs Tomaymycin and Sibiromycin in Streptomyces achromogenes and Streptosporanium sibiricum respectively. TomJ and SibT are each responsible for the imine reduction step in these biosynthetic pathways. 22
Figure 1.21: Reduction of imine moiety of dihydrofolate by DHFR to form tetrahydrofolate; an intermediary reaction in the biosynthesis of DNA.[50] 23
Figure 1.22: Reduction of imine moiety of beta-carboline imines by cell free extract of strains of Saccharomyces to form the corresponding (R)-amines.[51] 23
Figure 1.23: Reduction of imine moiety of beta-carboline imines by cell free extract of E. foetida to form the corresponding (R)-amines.[52] 24
Figure 1.24: Reduction of imine moiety of 2-methyl-1-pyrroline (R)-selectively and (S)-selectively by enantiocomplementary strains of Streptomyces. Reduction of 1a using produces amine (R)-1b with Streptomyces sp. GF3587 and (S)-1b using Streptomyces sp. GF3546 [53-56] 25
Figure 1.25: A schematic view of the aims of the first part of this project, which will involve the study of NADPH-dependent IREDs. 26
Figure 2.1: The pET-YSBLIC3C expression vector, based on the widely used pET28a vector. pET28a carries an N-terminal 6His tag coding sequence, a kanamycin antibiotic resistance marker, a multiple cloning site, the T7 promoter sequence, the lac operator coding sequence and the lacI gene which codes for the lactose repressor protein, Additionally to the pET28a features, pET-YSBLIC3C contains the coding sequence for a human rhinovirus 3C (HRV3C) protease cleavage site downstream of the 6His tag, allowing for the cleavage of the N-terminal 6His tag from expressed proteins. The pET-YSBLIC3C expression vector has also been modified to include a ligation independent cloning (LIC) site with a BseRI cleavage site.[66] 39
Figure 2.2: The repeating cycle of denaturing the target DNA template, annealing of complementary oligonucleotide primers and extension of new DNA using DNA polymerase in a PCR reaction for the amplification of target DNA sequences 41

Figure 2.3: A schematic representation of a nickel affinity chromatography column. Target proteins tagged with the 6His tag have a high affinity for the Ni^{2+} resin, therefore binding to the column medium. The remaining proteins are eluted through the column without binding, allowing them to be removed from the mixture. The 6His-tagged target protein is then displaced from the column by imidazole, which has a higher affinity for Ni^{2+}, and is eluted from the column as pure protein. 50
Figure 2.4: A schematic representation of a SEC column. The resin contains small pores through which smaller proteins are able to flow, causing them to take a longer route through the column. Larger proteins do not fit through these pores and therefore bypass them, resulting in a faster route through the column. Larger proteins are therefore eluted more quickly than smaller ones. 51
Figure 2.5: A schematic diagram of the experimental setup of an HPLC. The sample is injected into the column, through which a gradient of buffers is pumped. The speed at which analytes travel through the column is determined by the solubility of the compound in the solvent and the degree of interaction of the analyte to the stationary phase. Analytes eluted from the column are detected, producing a signal that is visualised using a processing display unit. 55
Figure 2.6: A schematic diagram of the experimental setup of a GC. The sample is injected and transported through the column via a carrier gas. The speed at which analytes are eluted from the column is determined by their boiling point and the degree of affinity to the stationary phase. Analytes eluted from the column are detected at the detector, producing a signal that are then visualised using a processing display unit. 56
Figure 2.7: A representation of a typical Michaelis-Menten plot for enzyme kinetics. Reaction velocity is plotted against substrate concentration. A maximum velocity is reached as a certain substrate concentration where the enzyme becomes saturated. This velocity is termed V_{max}. The substrate concentration at which an enzyme velocity of half the V_{max} is achieved is termed the K_M, and is a measure of enzyme affinity for the substrate. 59
Figure 2.8: A representation of a substrate inhibition model reaction (dotted blue line) superposed onto a typical Michaelis-Menten model (red line). Reaction velocity is plotted against substrate concentration. Unlike in a Michaelis-Menten system, maximum velocity (V_{max}) is not reached at higher substrate concentrations as the

substrate begins to exhibit an inhibitory effect on the system. V_{max} and K_M are determined by fitting a curve using a model which includes an inhibition constant, K_I, to account for substrate inhibition.....60
Figure 2.9: The oxidation of the enzyme cofactor NADPH to form $NADP^+$. This oxidation can be detected spectophotometrically, as NADPH absorbs at 340 nm while $NADP^+$ does not.....61
Figure 2.10: A schematic representation of the steps involved in site-directed mutagenesis (SDM). The template DNA is denatured at high temperatures, allowing the mutagenic primers to anneal. The plasmid is then amplified from the mutagenic primers. The template DNA is removed by digestion by DnpI, and E. coli is transformed with the newly synthesised mutant DNA for the repair of nicks and cloning.....67
Figure 3.1: The NADPH-mediated reduction of imine 1a to form enantiomerically pure (R)-1b by SkR-IRED.....71
Figure 3.2: The crystal structure of a single monomeric subunit of SkR-IRED at a resolution of 2.6 Å. The monomer consists of an N-terminal Rossman fold motif (residues 1 – 181) and a C-terminal helical domain (residues 210 – 306) connected by a long inter-domain helix (Leu182 – Gly209). [60] 72
Figure 3.3: The crystal structure of apo SkR-IRED at 2.6 Å resolution. The monomer consists of an N-terminal Rossman fold motif (1 – 181) and a C-terminal helical domain (210 – 306) connected by a long inter-domain helix (Leu182 – Gly209). Dimer formation is facilitated by the reciprocal domain sharing of the two monomeric subunits A (green) and B (orange).[60] 73
Figure 3.4: SkR-IRED in complex with NADPH cofactor at 2.7 Å resolution. The positioning of NADPH indicates that the active site is formed in a cleft between the N-terminal domain of one subunit and the C-terminal domain of the opposite subunit. NADPH is illustrated by ball and stick representation.[60] 74
Figure 3.5: SDS-PAGE gel analysis of purification of SkR-IRED by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: unbound flow-through collected during loading of nickel column, lanes 3 – 14: fractions containing purified protein eluted during increasing imidazole gradient. Purified SkR-IRED is observed to be approximately 32 kDa by SDS-PAGE. 75

Figure 3.6: Chromatogram of the purification of SkR-IRED by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 76
Figure 3.7: SDS-PAGE gel analysis of purification of SkR-IRED by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified protein eluted between 55 and 65 mL buffer. Purified SkR-IRED is observed to be approximately 32 kDa by SDS-PAGE.....77
Figure 3.8: Michaelis-Menten plot for the reduction of 1a by SkR-IRED. 78
Figure 3.9: The range of imines tested as substrates for SkR-IRED.....79
Figure 3.10: A summary of substrates accepted by SkR-IRED in biotransformation experiments. High ee values are observed for the conversion of 1a and 8a. For the biotransformation of 1a, an NADPH recycling system was used where *G6P is glucose-6-phosphate, **6PGL is 6-phosphoglucanolactone, and ***G6PDH is glucose-6-phosphate dehydrogenase. 80
Figure 3.11: The superimposition of the structures of SkR-IRED (red) and HIBDH (blue). The overlay demonstrates very high structural similarity between the two enzymes, where there is almost complete conservation between the two N-terminal Rossman fold subdomains (right). There is also significant conservation between the two C-terminal helical domains. However, in HIBDH a sharp β-turn occurs at residue Ser204, thereby turning C-terminal domain back towards the N-terminal domain. Such a feature is not observed in SkR-IRED, causing the N-terminal domain to protrude from the C-terminal domain.[67] 82
Figure 3.12: The mechanism for the reduction of 2-methyl-3-oxopropoanoate by β-hydroxyisobutyrate dehydrogenase from Thermus thermophilus (HIBDH). Residue Lys165, which is proximal to both the NADPH cofactor and β-hydroxyisobutyrate substrate behaves as a proton donor to facilitate the formation of the corresponding alcohol.[67].....83
Figure 3.13: The superimposition of the active sites of 2CVZ (purple) and SkR-IRED (red) shows that residue Asp187 occurs in the equivalent position in SkR-IRED to the catalytically active Lys165 residue of 2CVZ. 84
Figure 3.14: The mechanism proposed for the reduction of imines (such as 1a) by SkR-IRED. A proton is donated to 1a by catalytic residue Asp187, resulting in the formation

of an activated iminium ion. Hydride is then delivered to the iminium ion by NADPH to form the corresponding (R)-amine.....85
Figure 3.15: Chromatogram of the purification of SkR-IRED Asp187Ala by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 86
Figure 3.16: SDS-PAGE gel analysis of purification of SkR-IRED Asp187Ala by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: crude cell lysate, lane 3: sample of protein pooled following nickel-affinity chromatograpy, lanes 3-15: fractions containing purified protein eluted between 49 and 66 mL buffer. Purified SkR-IRED Asp187Ala is observed to be approximately 32 kDa by SDS-PAGE. 87
Figure 3.17: The active site of SkR-IRED. The dashed line illustrates the bond distance of 7.9 Å between the C4 atom of NADPH (from which hydride is delivered to the substrate) and the potentially catalytic residue Asp187, further than the 4.4 Å Lys165-NADPH C4 distance observed in the homologue HIBDH, where direct protonation from the protic Lys165 to the substrate occurs.....88
Figure 3.18: A region of unassigned electron density in the active site of SkR-IRED in the vicinity of the putative catalytic residue Asp187.....89
Figure 3.19: The chemical structure of dihydrofolate, a known substrate for the imine reductase DHFR. 89
Figure 3.20: The chemical structure of methotrexate, a known inhibitor for the imine reductase DHFR. 90
Figure 3.21: Mass spectrum of purified SkR-IRED using electrospray ionisation. Two intense peaks are observed in a feasible mass region for potential substrates which may explain the unassigned density in the enzyme active site. The accurate masses of the peaks suggest a molecular formula of $C_{27}H_{33}N_5O_5$ and $C_{22}H_{21}N_3O_5$ for m/z = 507.24806 and 408.15574 respectively. 91
Figure 3.22: The condensation and reduction of an L-amino acid and α-keto acid by an opine dehydrogenase (OpDH) and NADPH to form a product containing a newly formed amine moiety. 92
Figure 3.23: The chemical structure for reductasporine; a tryptophan dimer (TD) with antifungal properties.[70] 94

Figure 3.24: An imine reduction step performed by IRED RedE in the biosynthetic pathway of reductasporine, a tryptophan dimer (TD) with antifungal properties.[70] 94
Figure 4.1: Agarose gel visualising the amplification of the gene coding for SS-IRED by PCR. Lane 1: NEB 1 kB ladder, lane 2: PCR product of gene coding for SS-IRED 97
Figure 4.2: Agarose gel visualising restriction digest of gene coding for SS-IRED subcloned into the pET-YSBLIC3C expression vector. Lane 1: NEB 1 kB ladder. Lane 2: pET-YSBLIC3C plasmid containing S-IRED gene, digested with NcoI and NdeI. Two bands are observed in lane 2; the pET-YSBLIC3C vector backbone (~5.5 kB) and the SS-IRED gene (~900 bp) 98
Figure 4.3: SDS-PAGE gel analysis of purification of SS-IRED by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: unbound flow-through collected during loading of nickel column, lanes 3 – 14: fractions containing purified protein eluted during increasing imidazole gradient. Purified SS-IRED is observed to be approximately 32 kDa by SDS-PAGE. 99
Figure 4.4: Chromatogram of the purification of SS-IRED by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 100
Figure 4.5: SDS-PAGE gel analysis of purification of SS-IRED by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified protein. Purified SS-IRED is observed to be approximately 32 kDa by SDS-PAGE. 101
Figure 4.6: Chromatogram of the purification of SS-IRED, after cleavage of its 6His tag, by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 102
Figure 4.7: SDS-PAGE gel analysis of purification of SS-IRED by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified protein. Purified SS-IRED is observed to be approximately 30 kDa by SDS-PAGE. 103
Figure 4.8: Crystals of SS-IRED cleaved of N-terminal 6His tag obtained with protein at 120 mg ml^{-1} in 2.1 M DL-malic acid pH 7.0 (left) and 0.1 M HEPES pH 7.5, 1.4 M sodium citrate tribasic dehydrate (right) 104
Figure 4.9: Protein sequence alignment of SS-IRED (top) and SkR-IRED (bottom). The GXGXXG NADPH-binding motif which is conserved between the two sequences is

highlighted in blue. The 'catalytic' Asp187 residue of SkR-IRED aligns with Tyr169 in SS-IRED, outlined in red. 105
Figure 4.10: Graphical representation of the 2D-structural description of SS-IRED. β-helices are represented by blue arrows and α-helices are represented by red barrels. 108
Figure 4.11: Graphical representation of the 2D-structural description of SS-IRED. β-helices are represented by blue arrows and α-helices are represented by red barrels.[71-74] 109
Figure 4.12: The crystal structure of SS-IRED in its apo form at 3.2 Å resolution. The monomer consists of an N-terminal Rossman fold motif and a C-terminal helical domain connected by a long interdomain helix (Ala164-Ala192). Dimer formation is facilitated by the reciprocal domain sharing of the two monomeric subunits A (green) and B (blue). 110
Figure 4.13: Structural overlay of the putative active sites of SkR-IRED and SS-IRED. The image shows that the possible catalytic residue Asp187 of SkR-IRED is replaced by Tyr169 in SS-IRED. 111
Figure 4.14: A sequence alignment of SkR-IRED and SS-IRED. The comparison shows that SkR-IRED has a significantly longer N-terminus (highlighted in blue). The N-terminus of SS-IRED, by contrast, consisted of only 5 amino acids (highlighted in grey). 112
Figure 4.15: The sequence of SS-IRED+ (middle), a modified version of the SS-IRED enzyme, in alignment with the original SS-IRED sequence (top) and the sequence of SkR-IRED (bottom). The N-terminus of SS-IRED has been truncated to the consensus sequence of the two wild-type enzymes SS-IRED and SkR-IRED, and the N-terminus of SkR-IRED has been inserted in its place (highlighted in blue). 113
Figure 4.16: Agarose gel visualising the amplification of the gene coding for SS-IRED+ by PCR. Lane 1: NEB 1 kB ladder, lane 2: PCR product of gene coding for SS-IRED+. 114
Figure 4.17: Agarose gel visualising restriction digest of gene coding for SS-IRED+ subcloned into the pET-YSBLIC3C expression vector. Lane 1: NEB 1 kB ladder. Lane 2: pET-YSBLIC3C plasmid containing SS-IRED+ gene, digested with NcoI and NdeI. Two bands are observed in lane 2; the pET-YSBLIC3C vector backbone (~5.5 kB) and the SS-IRED+ gene (~1000 bp) 115

Figure 4.18: SDS-PAGE analysis of small scale expression test for the production of SS-IRED+. Lane 1: BioRad low range molecular weight marker. Lane 2: Insoluble fraction for uninduced control cells. Lanes 3 – 5: Insoluble fractions for expression at 16, 30 and 37 °C respectively. Lane 6: Soluble fraction for uninduced control cells. Lanes 7 – 9: Soluble fractions for expression at 16, 30 and 37 °C respectively. The target protein is present in both soluble and insoluble fractions for all induced samples. 116
Figure 4.19: SDS-PAGE gel analysis of purification of SS-IRED+ by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: crude cell lysate, lane 3: unbound flow-through fraction collected during loading of nickel column, lanes 4 – 14: fractions containing purified protein eluted during increasing imidazole gradient. Purified SS-IRED+ is observed to be approximately 32 kDa by SDS-PAGE..... 117
Figure 4.20: Chromatogram of the purification of SS-IRED+ by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 118
Figure 4.21: SDS-PAGE gel analysis of purification of SS-IRED+ by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified. Purified SS-IRED is observed to be approximately 32 kDa by SDS-PAGE. 119
Figure 4.22: Michaelis-Menten kinetic plot of SS-IRED towards substrate 1a. 120
Figure 4.23: Sequence alignment of genes coding for SkR-IRED (top) and SS-IRED (bottom). The two genes are 36 % sequence similar. The putative catalytic Asp187 residue of SkR-IRED is replaced by residue Tyr169 in SS-IRED (highlighted blue). 121
Figure 4.24: A structural overlay between the active site region of SkR-IRED (in complex with NADPH) and the equivalent position in SS-IRED in the active site region of SkR-IRED. The figure demonstrates that Asp187 (SkR-IRED) and Tyr169 (SS-IRED) occur in the same position. 122
Figure 4.25: The mechanism proposed for the reduction of imines (such as 1a) by SS-IRED. A proton is donated to 1a by catalytic residue Tyr169, resulting in the formation of an activated iminium ion. Hydride is then delivered to the iminium ion by NADPH to form the corresponding (S)-amine. 122

Figure 4.26: The mechanism of the imine reductase PTR1 in the first step in the NADPH-assisted reduction of an oxidised pterin to the active tetrahydro-form..... 123
Figure 4.27: The putative active site region of SS-IRED. Residue Tyr169, which may have a role as a proton donor in the reduction of imines, is neighboured by residues His243 and Ser94 which may assist Tyr169 in catalysis. 124
4.4.9 Mutational studies towards the mechanism of SS-IRED..... 124
Figure 4.28: SDS-PAGE gel analysis of purification of SS-IRED Tyr169Phe by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: crude cell lysate, lane 3: Unbound flow-through fractions on column binding, lanes 4 - 15: fractions containing purified protein..... 125
Figure 4.29: SDS-PAGE gel corresponding to the purification of SS-IRED Ser94Ala by nickel affinity chromatography. Lane 16: Bio-Rad low range molecular weight marker, lane 17: crude cell lysate, lane 18: Unbound flow-through fractions on column binding, lanes 19-28: fractions containing purified protein,..... 126
Figure 4.30: SDS-PAGE gel corresponding to the purification of SS-IRED His243Ala by nickel affinity chromatography. Lane 29: Bio-Rad low range molecular weight marker, lane 30: crude cell lysate, lane 31: Unbound flow-through fraction on column binding, lanes 32-43: fractions containing purified protein..... 126
Figure 4.31: Chromatogram of the purification of SS-IRED Tyr169Phe by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 127
Figure 4.32: Chromatogram of the purification of SS-IRED Ser94Ala by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 127
Figure 4.33: Chromatogram of the purification of SS-IRED His243Ala by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace. 128
Figure 4.34: SDS-PAGE gel analysis of purification of mutants of SS-IRED by size exclusion chromatography. Top: SDS-PAGE gel corresponding to SS-IRED Tyr169Phe. Lane 1: Bio-Rad low range molecular weight marker, lanes 2-10: fractions containing purified protein, Bottom left: SDS-PAGE gel corresponding to SS-IRED Ser94Ala. Lane 11: Bio-Rad low range molecular weight marker, lanes 11 - 16: fractions containing purified protein, Bottom right: SDS-PAGE gel corresponding to SS-IRED His243Ala.

Lane 17: Bio-Rad low range molecular weight marker, lanes 18 - 24: fractions containing purified protein, ... 129
Figure 5.1: The active site position in the crystal structure of AoIRED in complex with (R)-1-methyl-1,2,3,4-tetrahydroisoquinoline ((R)-4b). The figure demonstrates the nature of imine substrate binding in AoIRED. The amine product (R)-4b is bound in a cleft between the C-terminal domain of subunit A (green) and the N-terminal domain of subunit B (purple). 133
Figure 5.2: Active site overlay of SR-IRED and SS-IRED at the putative amine binding site, based on the location of imine binding in AoIRED. The figure details the positions at which residues are not conserved between enzymes. 134
Figure 5.3. SDS-PAGE analysis of expression of mutants A1 - B5. Samples 1 and 7 correspond to a low range molecular weight marker supplied by NEB. Samples 2 – 6 and 8 - 12 correspond to the soluble and insoluble fractions for expression of SR-IRED mutants A1 – 5 in E. coli BL21(DE3) respectively. Soluble expression is seen for mutants A2, A3 and A5. No clear soluble expression is seen for mutant A4, and no expression at all is seen for mutant A1. 135
Figure 5.4. SDS-PAGE analysis of expression of mutants B1-B5. Samples 1 and 7 correspond to a low range molecular weight marker supplied by NEB. Samples 2 – 6 and 8 - 12 correspond to the soluble and insoluble fractions for expression of SS-IRED mutants B1 – 5 in E. coli BL21(DE3) respectively. Soluble expression is seen for mutants B2, B3 and B5. Only insoluble expression is seen for mutant B4, and no expression at all is seen for mutant B1. 136
Figure 5.5: Chiral GC analysis of 1a (5 mM) in EtOAc. The figure demonstrates that imine 1a has a retention time of 16.37 min. 137
Figure 5.6: Chiral GC analysis of (R)-1b (5 mM) in EtOAc. The figure demonstrates that the optically pure amine (R)-1b has a retention time of 20.26 min. 137
Figure 5.7: Chiral GC analysis of racemic 1b (5 mM) in EtOAc. The figure demonstrates that the optically pure amine (S)-1b has a retention time of 18.72 min, as the peak at 20.26 minutes corresponds to the (R)-1b enantiomer (Figure 5.5). 138
Figure 5.8: Substrates selected for IRED biotransformations in study selected by Wetzl and coworkers. 143
Figure 5.9: List of sequences which display consistent (R)- or (S)-enantiopurity of products formed by tested imine substrates, where strongly selective (>80 % ee for

all products) IREDs are listed in black and more weakly selective IREDs (<80 % ee for one or more products) are listed in grey italics. Residues are numbered according to SkR-IRED sequence positions. 147
Figure 5.10: An alignment of sequences of strongly (S)-selective (blue) and (R)-selective (red) IREDs with amino acid residues colour coded according to level of conservation between equivalent positions in similarly selective IREDs according to the key in Table 5.5. The most highly conserved residues within each group were used to form a consensus sequence for (S)- and (R)-selective IREDs..... 150
Figure 6.1: Chromatogram recorded during size exclusion chromatography purification of PpIRED. PpIRED is eluted in a broad peak between approximately 45 – 59 mL elution volume. The peak exhibits some shoulders which may have arisen due to irregular protein concentration caused by dilution during loading. 157
Figure 6.2: SDS-PAGE analysis of fractions collected during nickel-affinity chromatography purification of PpIRED. 158
Figure 6.3: SDS-PAGE analysis of fractions collected during size exclusion chromatography purification of PpIRED. 158
Figure 6.4: A single cystal of PpIRED in its apo form in the INDEX screen in 0.2 M NaCl, 0.1 M BIS-TRIS pH 6.5, 25 % w/v PEG 3350. 159
Figure 6.5: Protein sequence alignment of PpIRED and 3DOJ (glyoxylate reductase from Arabidopsis thaliana). 160
Figure 6.6: Crystal structure of a single monomeric subunit of PpIRED, consisting of an N-terminal Rossman fold motif and a C-terminal helical domain..... 163
Figure 6.9: A dimer of PpIRED formed by the association of two monomeric subunits A and B. The dimer does not exhibit any domain sharing effects which are observed in other IREDs. 164
Figure 6.10: Crystals of Pp-IRED in complex with $NADP^+$ formed using Pp-IRED incubated with 10 mM $NADP^+$ in 0.2 M L-proline in the Hampton INDEX screen in a 96-well sitting drop format. 166
Figure 6.11: The crystal structure of PpIRED in complex with $NADP^+$. The cofactor binds in a cleft formed between the N- and C-terminal domains of a single subunit. 168

Figure 6.12: Amine substrates which are oxidised to the corresponding imine by PpIRED with the cofactor $NADP^+$. These were identified as substrates for PpIRED during $NADP^+$ depletion assays carried out by GlaxoSmithKline. 168
Figure 6.13: Non-linear regression analysis of activity levels of PpIRED towards increasing concentrations of 11a fitted using the substrate inhibition model. The presence of substrate inhibition is visible from the marked decrease in reaction velocity at the highest substrate concentration (60 mM). This conclusion is confirmed in a quantifiable way using an extra sum of squares F test ($F = 25.07$, $P = 0.0001$). 170
Figure 6.14: Non-linear regression analysis of reaction velocity of PpIRED towards increasing concentrations of substrate 13a using the Michaelis-Menten model..... 171
Figure 6.15: Non-linear regression analysis of reaction velocity of PpIRED towards increasing concentrations of substrate 14a using the Michaelis-Menten model..... 171
Figure 6.16: Non-linear regression analysis of reaction velocity of PpIRED towards increasing concentrations of substrate 15a using the Michaelis-Menten model..... 172
Figure 6.17: Non-linear regression analysis of reaction velocity of PpIRED towards increasing concentrations of substrate 16a using the Michaelis-Menten model..... 172
Figure 6.18: Chromatogram recorded from GC analysis of sample containing commercially prepared 11a (5 mM) in EtOAc 176
Figure 6.19: Chromatogram recorded from GC analysis of sample containing commercially prepared pentadecane (5 mM) in EtOAc. Pentadecane acts as an internal standard in the analysis. 176
Figure 6.20: Chromatogram recorded from GC analysis of samples drawn from a reaction of 11a with 0.2 mg mL^{-1} PpIRED and $NADP^+$ at 0 min (top) and 24 h (bottom). There is no evidence of the depletion of the peak corresponding to 11a after 24 h reaction time. 177
Figure 6.21: Chromatogram recorded from GC analysis of samples drawn from a reaction of 11a with 5 mg mL^{-1} PpIRED and $NADP^+$ at 0 min (top) and 24 h (bottom). There is, again, no evidence of the depletion of the peak corresponding to 11a after 24 h reaction time despite an increased enzyme concentration. 178

List of Tables

Table 2.1: Components required to make LB medium....30
Table 2.2: Components required to make TB medium....30
Table 2.3: Components required to make buffer A....31
Table 2.4: Components required to make buffer B....31
Table 2.5: Components required to make 1 L buffer C....32
Table 2.7: Components required to make TAE buffer for agarose gel electrophoresis 32
Table 2.8: Components required to make 4 x loading buffer for SDS-PAGE....33
Table 2.9: Components required to make running buffer for SDS-PAGE....33
Table 2.10: Components required to make sufficient resolving gel for SDS-PAGE for 2 SDS-PAGE gels 33
Table 2.11: Components required to make sufficient gel for SDS-PAGE for 2 SDS-PAGE gels....34
Table 2.12: Components required for preparation of glycin-NaOH buffer (0.1 M, pH 9.5)....34
Table 2.13: Components required for preparation of phosphate buffer (0.1 M, pH 7.5) 34
Table 2.14: Imine, aldehyde and alcohol substrates employed in IRED reactions....35
Table 2.15: Products for IRED reactions. Products have not been included where the product compound is already included in the list of substrates 37
Table 2.16: The reaction components used in standard PCR for the amplification of target DNA sequences 42
Table 2.17: Cycling parameters used as standard PCR protocol for the amplification of target DNA sequences in a thermal cycler. 43
Table 2.18: Reaction components used in enzyme kinetic assays....61
Table 2.19: HPLC conditions used for the analysis of imines 63
Table 2.20: HPLC conditions used for the analysis of amines 64
Table 2.21: Chromatography conditions employed in the chiral GC analysis of biotransformations of 1a by mutants of SS-IRED and SR-IRED....65
Table 2.22: Chromatography conditions employed in the chiral GC analysis of biotransformations of 11a – 13a by Pp-IRED 66
Table 2.23: PCR reaction components employed for site directed mutagenesis....68

Table 2.24: PCR cycling parameters for site directed mutagenesis 68
Table 3.1: Kinetic parameters determined during non-linear regression analysis of experimental data of the reduction of 1a by SkR-IRED at varying substrate concentrations using the substrate inhibition model 78
Table 3.2: Summary of data obtained from HPLC analysis of biotransformations of SkR-IRED towards imines 1a – 8a. 79
Table 3.3: Combinations of α-keto acids and L-amino acids used in opine dehydrogenase activity tests with SkR-IRED 93
Table 4.1: Data collection and refinement statistics for SS-IRED 107
Table 4.2: Kinetic parameters of SS-IRED towards substrate 1a. 120
Table 5.1: The results of GC analysis of biotransformations of imine 1a with mutants SR-IRED A1 – A5 and SS-IRED B1 – B5. The analysis demonstrates the enantioselectivity of each mutant by demonstrating the stereochemistry of the amine products formed in each reaction. Mutant reactions were performed in duplicate, and have been referred to as repeat 1 and repeat 2. 139
Table 5.2: Names and source organisms of enzymes included in analysis of sequence basis for enantioselective in IREDs. 142
Table 5.3: Enantioselectivity of selected IREDs towards six varying imine substrates. Enantioselectivity values for formation of (S)-amines are highlighted in blue and listed with positive values, enantioselectivity values for formation of (R)-amines are highlighted in red and listed with negative values 144
n.d. = not determined. This is stated where conversion is low (>25 %) or data are not available (where reaction data was obtained from sources external to Wetzl publication). 145
* Reaction data obtained from Man and coworkers[78] 145
**Reaction data obtained from results outlined in chapter 3 145
***Reaction data obtained from Hussain and coworkers[59] 145
Table 5.4: List of IREDs which consistently display >80 % enantioselectivity for the formation of only (R)- or only (S)-amine for all substrates tested. Enzymes with <80 % enantioselectivity for any substrate, or displayed mixed enantioselectivity (formation of different enantiomers depending on substrate) were omitted. 145

Table 5.5: Groups of amino acids treated as similar in sequence similarity analysis of enantioselective IREDs. Amino acid side chains are grouped according to hydrophobicity, size, charge and aromaticity. 148
Table 5.6: Key describing degree of conservation correlating to colour used to highlight residues in a sequence alignment 149
Table 5.7: An outline of amino acid residues conserved in active region of (R)- and (S)-selective IREDs. 151
Table 6.1: Data Collection and Refinement Statistics for apo-PpIRED 162
Table 6.2: Data collection and refinement statistics for PpIRED in complex with $NADP^+$ 167
Table 6.3: Values for kinetic parameters calculated from activity of PpIRED towards a range of amine, ketone and aldehyde-containing substrates (11a – 16a). 173
Table 8.1: Primer sequences used for all sub-cloning experiments 183
Table 8.2: Primer sequences used for all site-directed mutagenesis experiments. 183

Acknowledgements

I have been aspiring to achieve a PhD since early in my undergraduate career, and completing the writing of a PhD thesis is something of which I am immensely proud. It has been an enormous undertaking which has been intellectually, physically and psychologically testing, but an extremely worthwhile endeavour which has unlocked many doors for me as I go forward in life. Completing my PhD has brought important opportunities that have hugely benefited my life.

Since I began writing my PhD in 2012 I have experienced some life changing events, specifically the birth of my son in 2014 and the death of my partner in 2016. Without the extraordinary levels of support I received from my supervisor, family and colleagues, it would not have been possible to complete this work throughout this time, and I wish to take this opportunity to thank them.

First I thank my PhD supervisor Professor Gideon Grogan, not only for providing me with the opportunity to carry out this PhD, but also for inspiring me to carry out research in this area through his undergraduate lectures and for being such an approachable, compassionate and professional supervisor who has a great deal of concern for the success of his students. I truly believe that without Gideon's extraordinary supervision, it would not have been possible for me to finish my PhD under the challenging circumstances I sometimes faced.

I also sincerely thank my colleagues in the Grogan group who during my PhD have been not just co-workers but also great friends who have made work so enjoyable and provided so much moral support, particularly Muhiadin Omar, Henry Man, Amina Frese, Hamid-Reza Danesh Azari and Claudia Spandolf. These people have not only been colleagues, but amazing friends with whom I hope to keep in touch with over future years. I would especially like to thank Muhiadin for his assistance with my lab work when I had safety restrictions during my pregnancy, and providing a listening ear during times when life was tough.

Importantly I would also like to thank my family, particularly my late partner David Smith, my mother Rachel Wells and my son Rory's Godmother Jenefried Gay for their kindness and moral support during my PhD as well as their assistance with looking after Rory when I needed to work late or work additional days. Both David and my

mother have made personal sacrifices that have made the completion of my PhD possible, for which I will always be immensely grateful.

Finally I would like to thank Andrew Taylor for lifting my spirits immeasurably during my write up period, giving me the ultimate motivation to finish my PhD, which has helped far more than he knows.

1. Introduction

1.1 Biocatalysis

1.1.1 The history of biocatalysis

Biocatalysis is the use of biological entities such as enzymes, either in their isolated form or as part of their host organism, to catalyse chemical reactions. Enzymes, which usually take the form of proteins, are biological catalysts that underpin a great majority of vital biological reaction processes within organisms, from the digestion of food sources to biosynthesis of tissues.

The use of biological systems in the conversion of one chemical species to another is a process is known as a biotransformation, and is a process that has been biotechnologically exploited outside its natural physiological context for thousands of years. For instance, *Saccharomyces cerevisiae*, known as baker's yeast, has been used in the fermentation of sugars to form carbon dioxide and ethanol in the production of bread products and alcoholic drinks.[1] Enzymatic reactions also play an important role in the production of other food products such as the use of enzyme cocktail, rennet, in the production of cheese and the use of *Lactobacillus* in the controlled fermentation processes to produce yoghurt.[1]

In more recent years, biotransformations have been exploited for a broader range of applications, including manufacture of bulk chemicals and pharmaceuticals. One of the first examples of this was the discovery of penicillin, the first antibiotic used as a pharmaceutical, by Dr Alexander Fleming in 1928.[2] Microbes for its bulk manufacture were subsequently developed within the pharmaceutical industry.[1]

This early interest in biotechnology for the production of pharmaceuticals drove a wave of research into the use of microorganisms in the production of chemicals, underpinning our knowledge of the large scale production of relevant microorganisms using fermentation equipment, and the biochemical behaviour of enzymes. In particular, the development of technologies such as PCR and heterologous expression of genes using convenient host organisms and improved DNA sequencing technologies have allowed for extremely rapid progress in Biocatalysis research in recent decades.[1]

1.1.2 The advantages of biocatalysis

The field of biocatalysis in organic synthesis has expanded rapidly in recent decades due to advancements in microbiology that have made the production of enzymes easier. There has been motivation to develop this field due to the distinct advantages offered by enzymes towards many of the challenges faced by the synthetic chemist.

Biocatalysis is becoming increasingly realised as a solution to a number of issues facing the synthetic chemist. Firstly, they adhere strongly to the principles of green chemistry. Due to their inherently low toxicities, preference for mild reaction conditions (such as atmospheric temperature and pressure), and their ability to function optimally in water-based media without the requirement for organic solvents, they comply well with the principles of green chemistry (Figure 1.1). This is highly advantageous when they are being applied as catalysts on a large scale in an industrial setting.[3]

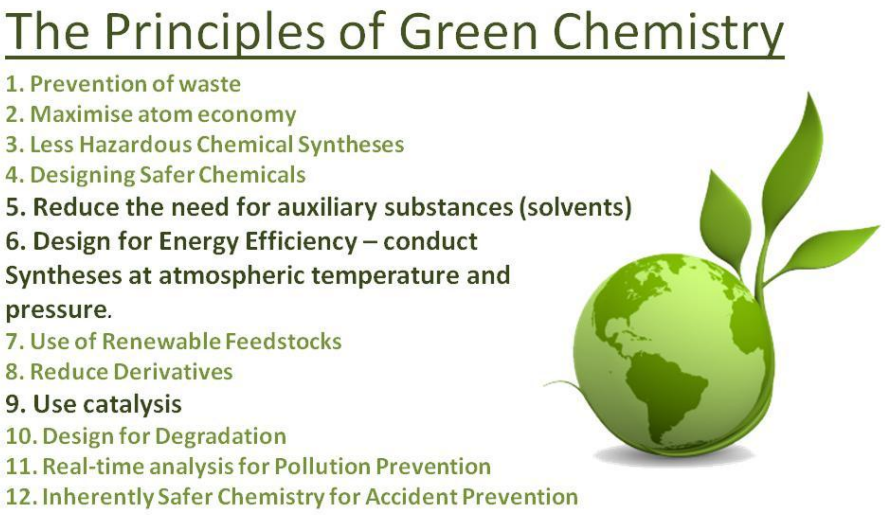


Figure 1.1: The principles of green chemistry. The use of enzymes is particularly relevant for certain green chemistry criteria. Their use reduces the need for auxiliary substances (i.e. organic solvents). They are also energy efficient, operating optimally at atmospheric temperature and pressure.[4]

Moreover, enzymes are inherently highly selective. They have often evolved in nature to transform a substrate in a very specific manner, with tight binding pockets that allow chemical changes to be carried out in a certain region of a molecule and chemical environments within the protein conferring very specific reaction environments into their active sites, making them regio-, chemo- and enantioselective. While impressive advancements have been made in regioselective and asymmetric chemocatalysis, it is often more difficult to introduce the levels of selectivity into chemical catalysis that are present in their biological counterparts.

With the increasing demand for pharmaceuticals and other fine chemicals to be dispensed in their enantiomerically pure forms, this high selectivity is an increasingly attractive feature in enzymes as catalysts.

1.2 Chiral amines

Optically pure chiral amines are highly valuable compounds due to their multitude of applications. They are applied in the pharmaceutical, chemical and agrochemical industries due to their uses as chiral auxiliaries, resolving agents and catalysts for asymmetric synthesis and account for many alkaloid natural products.[3, 5-8] Chiral amine centres are particularly common in pharmaceutical molecules. This is unsurprising due to the biological potency of this chemical class arising from their structural influence delivered by their propensity to form hydrogen bonds to neighbouring chemical entities such as proteins. A number of important examples of pharmaceuticals containing a chiral amine centre are illustrated in Figure 1.2.[8]

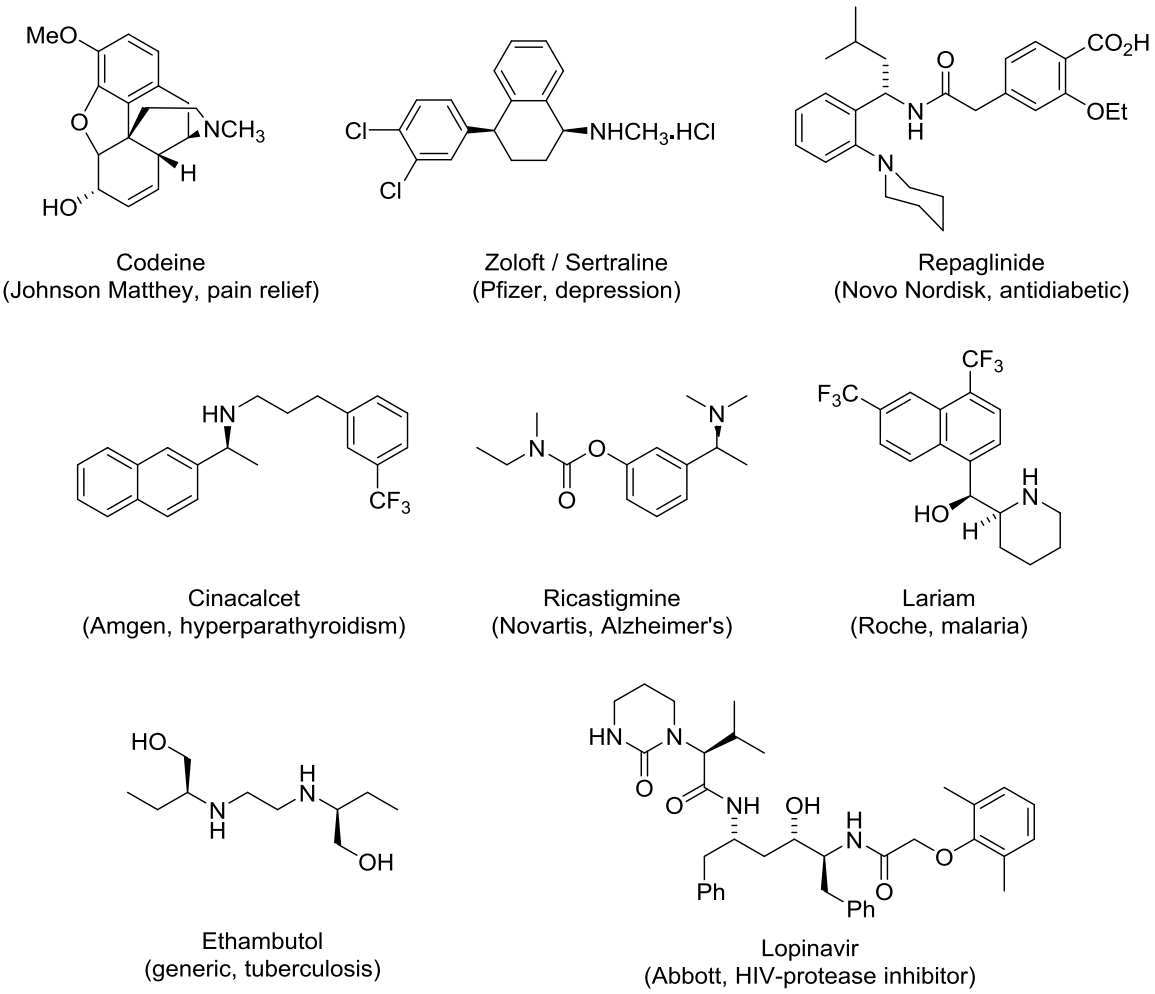


Figure 1.2: An array of pharmaceutical products containing optically active amine functional groups including diabetes, Alzheimer's and malaria treatments. The image illustrates the high demand for chiral amines in the pharmaceutical industry.[8]

1.3 Chemical synthesis of chiral amines

The synthesis of chiral amines in a regio-, stereo- and chemoselective manner is highly challenging. However, a number of chemical methods have been developed for their synthesis.[8]

Synthetic efforts have been focussed primarily on α-chiral primary amines, as most amine-containing pharmaceuticals and natural products are α-chiral secondary or tertiary amines which can be easily derived from α-chiral primary amines by simple alkylation or amidation. Therefore, α-chiral primary amines are the most convenient chiral building blocks used by chemists for the synthesis of pharmaceuticals and alkaloid natural products.

Traditionally, the desired enantiomer was isolated using techniques such as classical revolution and preparative HPLC.[9] However, such techniques are not green due to the necessity for additional reaction steps and wastage of the unwanted product enantiomer.[8]

The most efficient method to produce chiral amines is by asymmetric synthesis, where the desired enantiomer can be produced with high optical purity. A number of chemical routes of this nature have been developed; most commonly the reduction of enamides, reductive amination and the reduction of imines. Each of these methods share the asymmetric addition of hydride as the stereochemistry-determining step.[8]

1.3.1 Reduction of enamides

The pathways for the synthesis of primary amines from enamides are pictured in Figure 1.3. The most common enamide substrate employed in the synthesis of chiral amines is an *N*-acetylenamide, such as that which is pictured in the centre right of Figure 1.3.

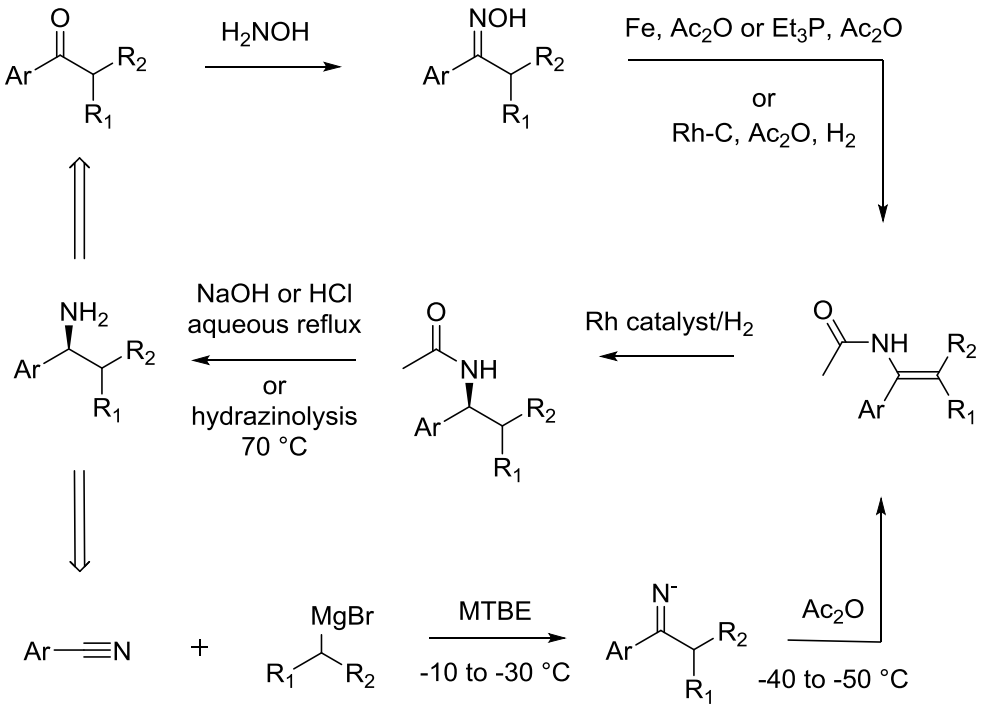


Figure 1.3: Synthetic routes towards primary chiral amines from *N*-acetylenamides. Synthesis of the enamide occurs using either a nitrile or ketone starting material. The enamide is then asymmetrically reduced with hydride using a Rhodium catalyst.[8]

While the asymmetric hydrogenation step generally provides 100 % conversion of enamide to the corresponding secondary amine product, the less reactive nature of the enamide substrate dictates that high hydrogen pressures (5 – 25 bar) are required for this reaction step.[10, 11] Fortunately this use of higher pressures does not adversely affect *ee*, but temperature must be controlled to ensure that sufficiently high product *ee*s are obtained.[11, 12] While this reaction step can be performed in a number of solvents, the most favourable *ee*s of the secondary amine product are obtained when CH_2Cl_2 or MeOH is employed as a solvent.[12, 13]

1.3.2 Reductive amination

Reductive amination is the one-pot conversion of a ketone and an amine starting material to the conjugated amine product,[14] as opposed to the reduction of preformed imines, which will be discussed in section 1.3.3.

An important example of an asymmetric reductive amination was reported in 1999 by Blaser and co-workers as a route towards the synthesis of an active ingredient in commercial herbicides, (*S*)-metolachlor. The synthesis involved a transition metal catalyst; an Ir-xyliphos complex under high pressures (80 bar) at 50 °C. Xyliphos is an

iron-containing chiral diphosphine complex used for the substrate-specific enantioselective catalysis of this reaction.[15]

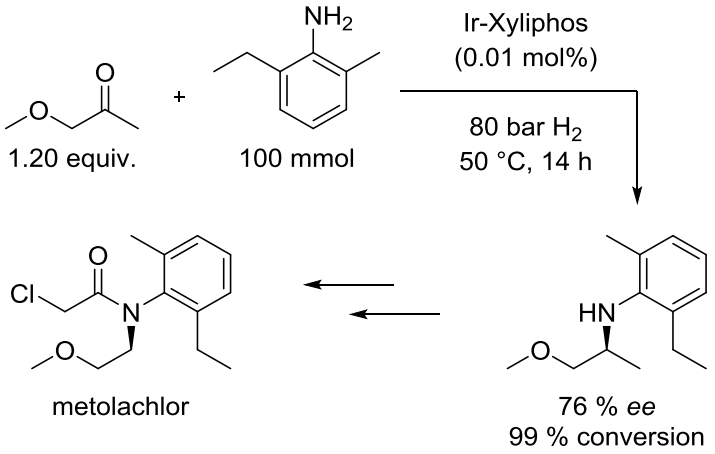


Figure 1.4: The first example of an enantioselective reductive amination: a synthesis of an active ingredient in commercial herbicides, metolachlor. The reaction involves the conjugation of methoxyacetone with 2-methyl-6-ethylaniline to form an intermediary imine *in situ*, before reduction using H_2 in the presence of an Ir-complexed chiral catalyst, Xyliphos.[15]

Ultimately, this reaction pathway was disfavoured for industrial application in light of a synthesis using a direct imine reduction which required decreased catalyst loadings (Section 1.3.3).[16]

As well as using transition metal-based asymmetric catalysts including the above example, reductive amination has also been performed asymmetrically using asymmetric organocatalysts such as chiral phosphoric acids and chiral auxiliaries such as chiral ammonia.[17, 18] Despite the need for removal of a chiral auxiliary using an additional deprotection step, this method is still feasible on an industrial scale provided the auxiliary is inexpensive.[8]

Despite its convenience, asymmetric reductive amination is less thoroughly explored than direct imine reduction due to some limitations in its implantation. Firstly, ketone starting materials involved in this reaction type are incompatible with transition metal hydride catalysts used in asymmetric hydrogenation reactions due to the formation of alcohol by-products. Additionally, transition metal catalysts are inhibited by complexation with amines, which act as both starting materials and

products in reductive amination reactions.[8, 16] However, recent developments in enzymatic asymmetric reductive amination brings a great deal of promise to the use of this pathway in an industrial context.[8]

1.3.3 Reduction of imines

The direct, asymmetric reduction of imines is a direct route towards the formation of chiral amines. The most common strategy employed for this synthetic route is homogeneous catalysis using a transition-metal based catalyst, however, the use of asymmetric organocatalysts is making an increasingly important contribution towards the field of asymmetric imine reduction.[8]

One of the most successful examples of an asymmetric imine reduction in an industrial context is the asymmetric reduction step in the synthesis of (*S*)-metolachlor, which is used in favour of alternative techniques using reductive amination due to the decreased catalytic load required as mentioned in Section 1.3.2 (Figure 1.5).[19]

Metolachlor is used as a herbicide in volumes in excess of 20,000 tonnes per year. The use of enantiomerically enriched (*S*)-metolachlor in place of a racemic mixture greatly reduces the necessary environmental load of the herbicide, greatly increasing the need for its efficient asymmetric synthesis.[19]

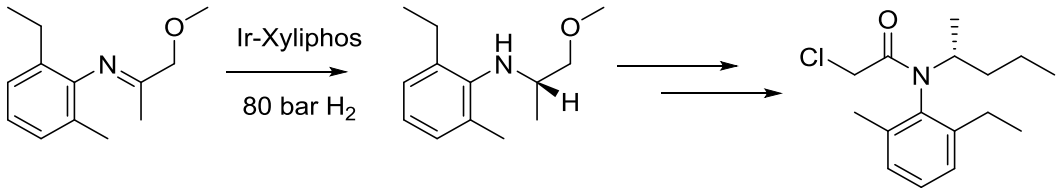


Figure 1.5: The synthesis of (*S*)-metolachlor using an asymmetric reduction step. Chirality is introduced by use of an asymmetric catalyst; the Ir-xyliphos (an iridium-complexed iron-containing chiral phosphine complex).

1.3.4 Limitations of chemical chiral amine synthesis

A number of approaches to the synthesis of chiral amines using chemical methods have been discussed in this section. While many remarkable examples where impressive yields and rates of conversion to products of interest have been obtained,

there are a number of limitations in the implementation of these reactions on an industrial scale.

Usually, asymmetric synthesis of chiral amines requires the use of transition metal catalysts or organocatalysts whose syntheses are complex in their own right. In the case of transition metal catalysts, many such metals are becoming increasingly expensive as global supply dwindles, and there is therefore a pressure to avoid their usage. Moreover, reduction reactions commonly necessitate high pressures and temperatures which further increase the environmental impact of these chemical processes. Deprotection steps are also sometimes necessary, particularly with the use of chiral auxiliaries, which further reduce the efficiency of such reactions by increasing the number of necessary steps and reducing atom economy.[8, 15-19]

Due to these factors, there is a strong market for the use of biocatalytic routes towards the synthesis of chiral amines, as this offers an opportunity to develop sustainable routes towards the synthesis of chiral amines which circumvent some of the disadvantages of chemical routes listed above.

1.4 Biocatalytic routes towards chiral amines

Because enzymes offer many industrial advantages (Section 1.1.2) as well as being highly selective, they offer a desirable tool towards the synthesis of chiral amines. There have therefore been substantial efforts to establish biocatalytic methods for the synthesis of chiral amines, and a number of classes of enzymes have been employed for this purpose (Figure 1.6).

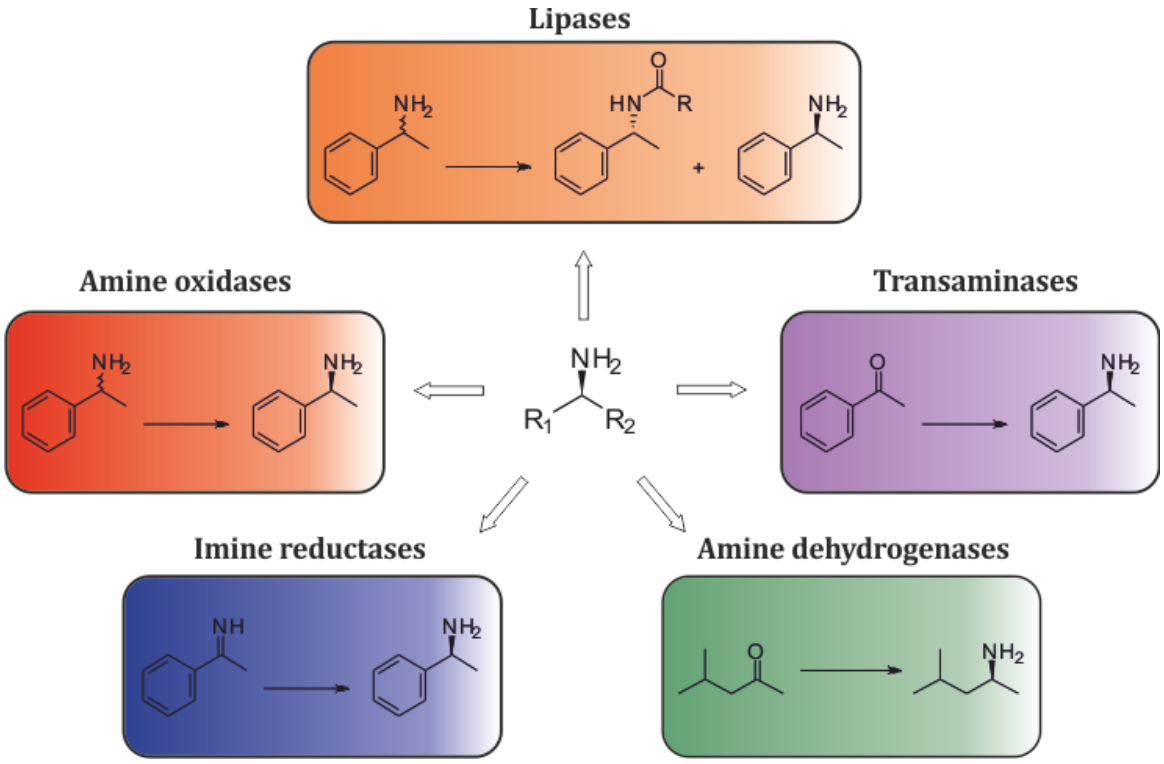


Figure 1.6: Enzymatic routes for the production of chiral amines. Current methods include kinetic resolution of racemic amines by hydrolases, dynamic kinetic resolution (DKR) and deracemization of racemic amines by hydrolases and oxidases, asymmetric amination of ketones by transaminases, decarboxylation of carboxylic acids by decarboxylases and the asymmetric reduction of imines by imine reductases (IREDs).[20]

1.4.1 Lipases

Lipases are enzymes of the hydrolase family which can catalyse the reversible hydrolysis and acylation of amides and esters (Figure 1.7).[20-22] As a class of enzymes, they have been thoroughly explored in terms of their applications and present many advantages for use in industrial biotechnology. They can be produced produced in high yield in bacteria and fungi, possess extremely high levels of chemo-, regio- and stereoselectivity and also maintain activity in organic solvents.[23-25]

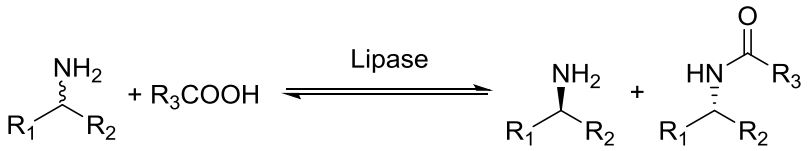


Figure 1.7: A general illustration of the lipase-catalysed reversible acetylation of an amide substrate to form the corresponding enantiomerically pure amine alongside the amide with opposite chirality.[20]

Lipases may be exploited for the synthesis of chiral amines using a kinetic resolution approach. The kinetic resolution approach limits the potential yield to only 50 % as only half of the starting material will be hydrolysed or acetylated. However, such limitations can be overcome using two different approaches.

Firstly, a dynamic kinetic resolution technique may be employed where unconverted starting material is racemised *in situ* to facilitate a theoretical 100 % yield. An example of this was reported by Reetz and coworkers in 1996, where racemic phenylethylamine was converted to the enantiomerically pure *N*-acetylated product using a strongly selective lipase from *C. antarctica* with palladium on charcoal (Pd/C) as a racemization agent (Figure 1.8).[26]

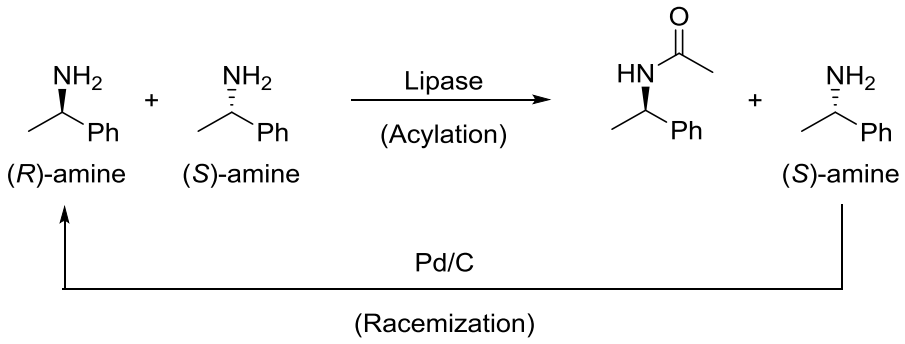


Figure 1.8: Dynamic kinetic resolution of a racemic amine by the selective acylation of the (*R*)-amine by a lipase and the racemisation of the unreacted (*S*)-amine using an achiral Pd/C catalyst.[26]

Secondly, an approach may be used where both enantiomerically pure products can be formed and separated to maximise the utility of the racemic starting material. A reaction using this method was reported by Balkenhohl and coworkers in 1997 (Figure 1.9). In the first reaction step a lipase in *P. plantarii* converts the (*R*)-amine to the corresponding amide, while enantiomerically pure, unreacted (*S*)-amine remains

in the reaction mixture. The amide is then separated from the reaction mixture by distillation or extraction and converted to the corresponding (*R*)-amine by basic hydrolysis.[27]

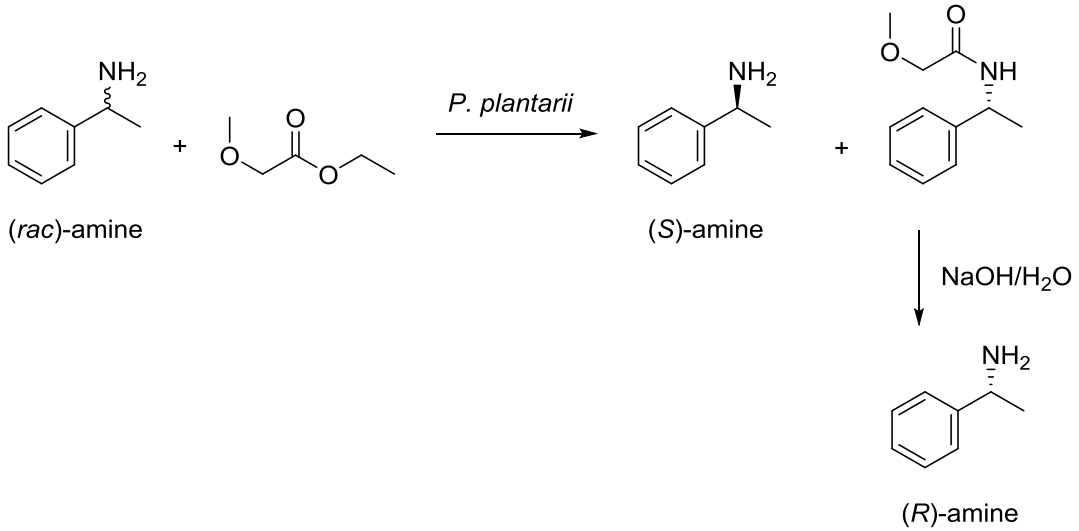


Figure 1.9: The two-step separation of racemic phenylethylamine to both the (*S*)- and (*R*)-phenylethylamine in their enantiomerically pure forms. The (*R*)-amine is initially converted to the corresponding amide by a lipase from *P. plantarii.* The (*R*)-amide product is subsequently separated by distillation and converted back to the (*R*)-amine by basic hydrolysis. The optically pure unreacted (*S*)-amine remains in the reaction mixture after removal of the (*R*)-amide by distillation.[27]

1.4.2 Transaminases

Transaminases are enzymes that catalyse the transfer of an amino group from an amino donor to a ketone or aldehyde. Transaminases are dependent on a pyridoxamine cofactor, which is converted to pyridoxal phosphate following the amination of the carbonyl-containing substrate (Figure 1.10). Transaminases can be broadly categorised into two classes; α-transaminases, which require a carboxylic acid moiety in the α-position relative to the aldehyde of ketone group and ω-transaminases, which can potentially accept any aldehyde or ketone. The latter are considered more useful from a biotechnological standpoint due to their increased versatility due to a broader substrate range as well as their ability to also catalyse the kinetic resolution of amides in the deamination direction.[20, 28]

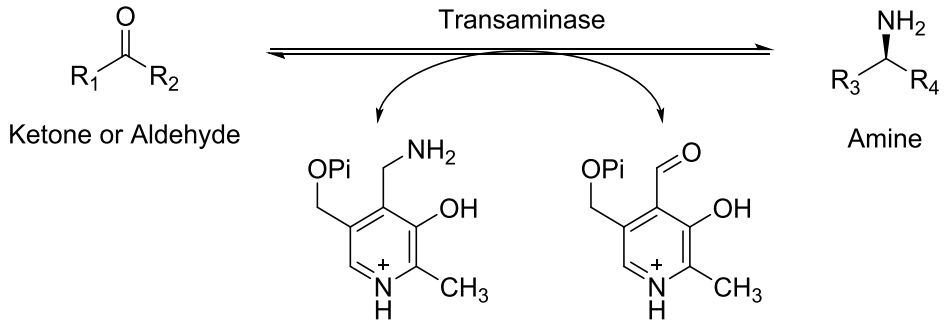


Figure 1.10: The general scheme of a reaction catalysed by a transaminase. An amino group is transferred from the pyridoxamine cofactor which acts as an amino donor to a ketone or aldehyde substrate, which facilitates the formation of a chiral amine when performed asymmetrically.[20]

When transaminases are enantioselective, they offer a route towards the synthesis of chiral amines. Historically, a number of limitations have inhibited their application as industrial biocatalysts; firstly, the presence of product inhibition due to the reversible nature of the enzymes and secondly, the limited substrate scope of wild-type enzymes.[20]

Successful efforts have been made to overcome limitations with product inhibition. An example an effective approach reported by Shin and co-workers where the inhibitory ketone product was extracted from the aqueous phase in a membrane reactor during kinetic resolution.[29]

An instance of a successful application in a commercial setting is where a transaminase was engineered for the synthesis of Januvia®, an anti-diabetic drug which lowers the blood sugar levels of patients (Figure 1.11).[30, 31] The wild-type enzyme was an (*R*)-selective ω-transaminase, ATA-117, which initially displayed activity towards small substrates which exhibited similarity towards the target molecule. A rational mutagenesis approach was employed during a collaboration between Merck and Codexis to target the precise substrate of interest, followed by additional rounds of mutation to improve properties of the enzyme specifically for industrial application.[20, 32]

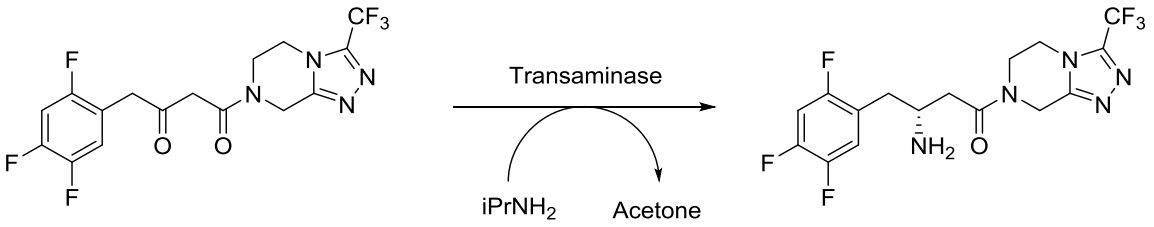


Figure 1.11: A biocatalytic route towards the synthesis of Januvia®, an antidiabetic drug used to lower the blood sugar level in patients, using a mutant rationally evolved from transaminase ATA-117.

This is a very good example to illustrate the use of enzymes in the commercial synthesis, as the biocatalytic route towards Januvia® using a transaminase increased the overall yield by 10 % while eliminating the need to use precious metal catalysts compared to the original chemocatalytic route using a Rh-based catalyst.[33]

1.4.3 Amine oxidases

Amine oxidases are oxidoreductases, which are enzymes that catalyse redox reactions. This class of enzymes is largely made up of monoamine oxygenases (MAOs) which catalyse the oxidation of amines to imines with concurrent reduction of oxygen to hydrogen peroxide (Figure 1.12). Amine oxidases can be grouped into two categories; Type I and Type II. Type I amine oxidases are dependent on both copper and pyrroloquinolinequinone (PQQ) cofactors, and are involved in biological pathways such as the metabolism of amino acids and biosynthesis of alkaloid natural products in bacteria, fungi, plants and animals. Type II amine oxidases are flavin-dependent enzymes found in mammals, fungi and bacteria which in humans are membrane bound-proteins which catalyse the inactivation of neurotransmitters such as seretonin and dopamine.[20]

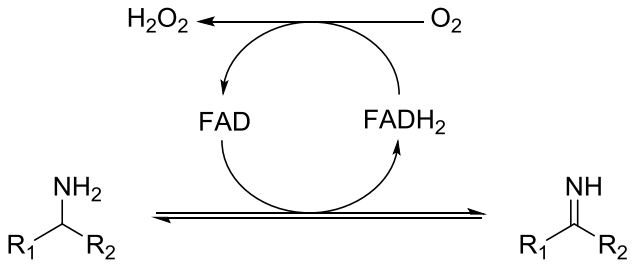


Figure 1.12: A general reaction catalysed by a flavin-dependent amine oxidase. An imine is reversible oxidised to an amine with concurrent reduction of oxygen to hydrogen peroxide.[20]

The most commonly applied amine oxidase in a biotechnological setting is the MAO from *Aspergillus niger* (MAO-N). The organism contains both species of amine oxide (Type I and Type II). Type I amine oxygenases form covalently bound imine products, therefore their use for biocatalytic applications is limited. Type II amine oxygenases generate free imines which may be used as substrates in subsequent biocataytic reaction steps. Therefore, although *A. niger* contains both species of amine oxidase, only those of the Type II category are applied.

MAO-N may be applied to obtain optically pure amines.[34] A library of MAO-N variants has been developed by directed evolution experiments to enable the synthesis of a diverse range of chiral amines, as well as variants with increased levels of activity and selectivity.[35-38] Such mutants have been applied in a novel deracemization process for the production of chiral amines where amines are enantioselectively oxidised by MAO-N variants in combination with an achiral chemical reduction of the prochiral imine (Figure 1.13).[39]

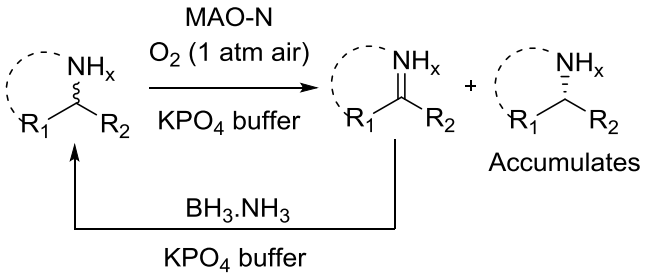


Figure 1.13: The deracemization of primary, secondary or tertiary amines by MAO-N variants with concurrent achiral reduction of the imine product back to the racemized amine starting material. The amine enantiomer which is not converted by MAO-N accumulates in the reaction mixture with increasing levels of optical purity.[39]

The extensive evolution efforts involving MAO-N have presented the enzyme as a promising catalyst for many industrially relevant reactions. An example of an application of the use of the deracemization chemistry of MAO-N in the context of pharmaceutical synthesis is the use a variant termed MAO-N D5 in the desymmetrisation of 3,4-substituded *meso*-pyrrolidines *via* the formation of the corresponding pyrroline imine.[40] This reactivity was exploited in the synthesis of pharmaceuticals such as Telaprevir, a Hepatitis C viral protease inhibitor (Figure 1.14).[41]

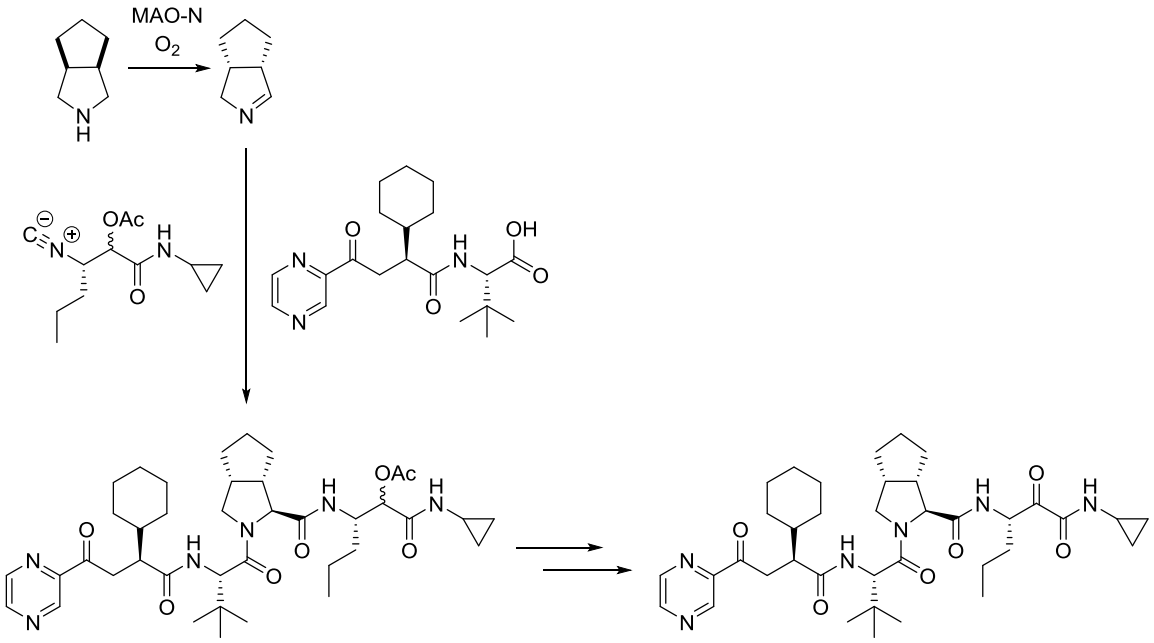


Figure 1.14: A biocatalytic step in the synthesis of Telaprevir, a Hepatitis C viral protease inhibitor, mediated by an engineered variant of MAO-N from *Aspergillus niger*.[41]

1.4.4 Amine dehydrogenases

Amine dehydrogenases (AmDHs) are NAD(P)H-dependent enzymes which catalyse the enantioselective reductive amination of ketones to produce the corresponding chiral amine via an imine intermediate. The majority of AmDHs have been engineered from wild-type amino acid dehydrogenases (AADHs), which catalyse the reductive amination of keto-acids using ammonia to form amino acids *via* an imino-acid intermediate (Figure 1.15). However, there also exist a small number of wild-type AmDHs including an AmDH from *Streptomyces virginiae* with limited selectivity.[42] AmDHs offer a direct route from ketones to chiral amines using a single enzyme.

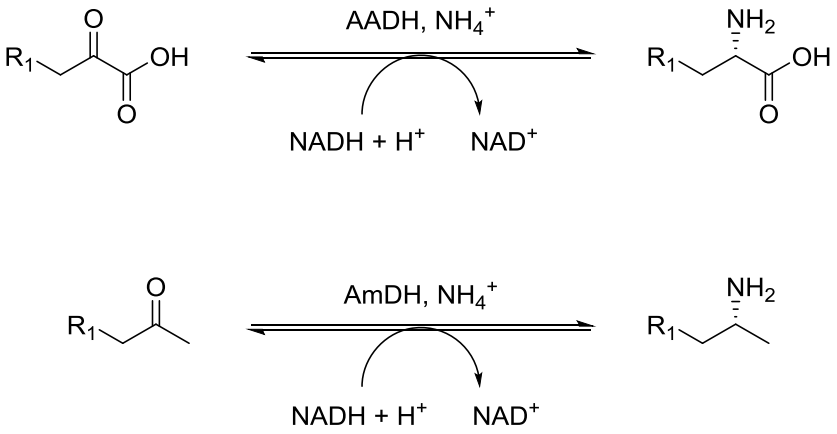


Figure 1.15: The reactions catalysed by amino acid dehydrogenases (AADHs) (top) and amine dehydrogenases (AmDHs) (bottom). AADHs catalyse the NADH-dependent interconversion of keto-acids and amino acids while AmDHs catalyse the interconversion of ketones and amines.

AmDHs were engineered to target substrates which are precursors to active pharmaceutical ingredients or speciality chemicals that contain chiral amines, and have primarily been evolved from leucine dehydrogenases, phenylalanine dehydrogenases and chimeric amino acid dehydrogenases possessing both of these activities.

In 2015 Mutti and co-workers reported the use of AmDHs in the synthesis of chiral amines from primary and secondary alcohols as part of a cascade reaction. The reaction involves the concurrent use of an alcohol dehydrogenase (ADH) and an AmDH to afford the amination of a broad range of aliphatic and aromatic alcohols to the corresponding chiral amines with high levels of conversion and optical purity (Figure 1.16).[43]

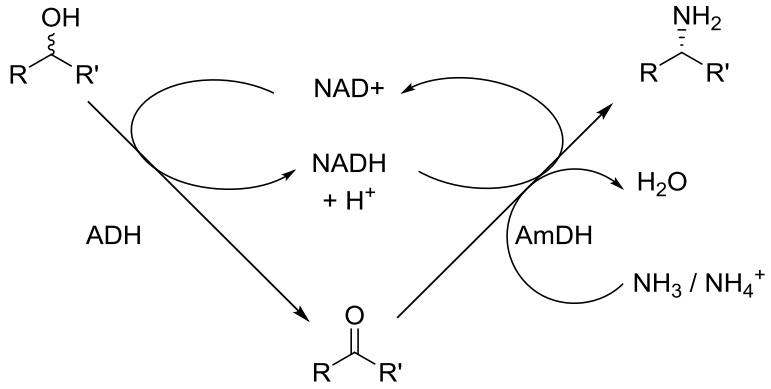


Figure 1.16: A cascade reaction employing an ADH and AmDH to convert racemic alcohols to chiral amines using NAD(H) in a single reaction vessel.

1.4.5 Imine reductases (IREDs)

This section focuses on the use of biocatalytic imine reduction to produce chiral amines. This involves the conversion of a prochiral imine to a chiral amine using an IRED. The ability to convert a prochiral starting material to an optically pure amine with a theoretical yield of 100 % makes this a potentially very efficient route towards the synthesis of chiral amines, and it has therefore attracted a large amount of interest in the Industrial Biotechnology sector.

Biocatalytic imine reduction offers a particularly attractive route for the synthesis of chiral amines. Imines are prochiral, so when their reduction is carried out asymmetrically, such as by using a selective IRED, this offers a means of directly introducing chirality into the achiral imine starting material, facilitating a theoretical yield of 100 % (Figure 1.17).

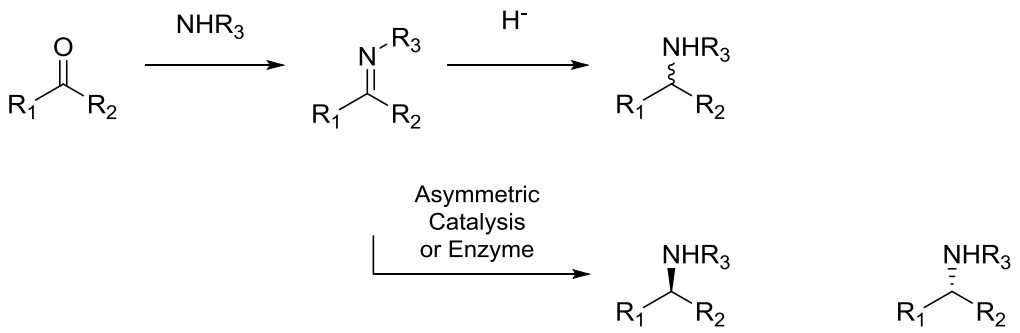


Figure 1.17: Imine reduction as a route for the synthesis of amines. When done asymmetrically, such as by using an enzyme, the reduction of imines (which are readily obtainable from ketones) can be used to generate optically pure amines.

Despite the desirability of IREDs as catalysts, this is a technology which had until recently not been well explored. Because many imines are labile under aqueous conditions, few organisms have evolved to metabolise them, and therefore few naturally occurring IREDs are known. As a result, some doubt initially existed as to whether IREDs exist as an enzyme class. However, increasing numbers of examples of IREDs arose in the literature in recent years, and knowledge about this enzyme class continues to rapidly expand.

A well-studied example of a naturally occurring imine reductase is dihydrofolate reductase (DHFR) from *E. coli*, the enzyme responsible for the reduction of dihydrofolate, a folic acid derivative, to form tetrahydrofolate, a key intermediate in the biosynthesis of DNA (Figure 1.18).[44] While this enzyme has not been exploited for biotechnological purposes, aside from in the formation of its natural product (*S*)-tetrahydrofolate,[45] this is a very widely known example which proved the principle that genuine IRED activity does indeed exist and occur naturally.

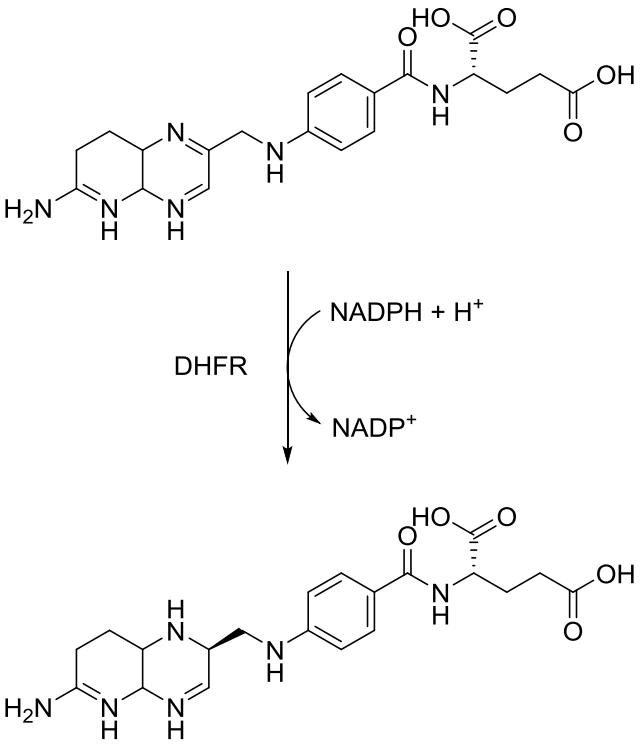


Figure 1.18: Reduction of imine moiety of dihydrofolate by DHFR to form tetrahydrofolate; an intermediary reaction in the biosynthesis of DNA.[44, 45]

A further example of naturally occurring IRED activity arises in the biosynthetic pathways of pyrrolobenzodiazepines (PDBs), which are natural products that exhibit antibiotic and anti-tumour properties. Imine reducing activity in species which biosynthesise PDBs is dependent on an unusual deazaflavin cofactor F-420, which acts as a redox cofactor in biocatalytic reactions (Figure 1.19).[46-48]

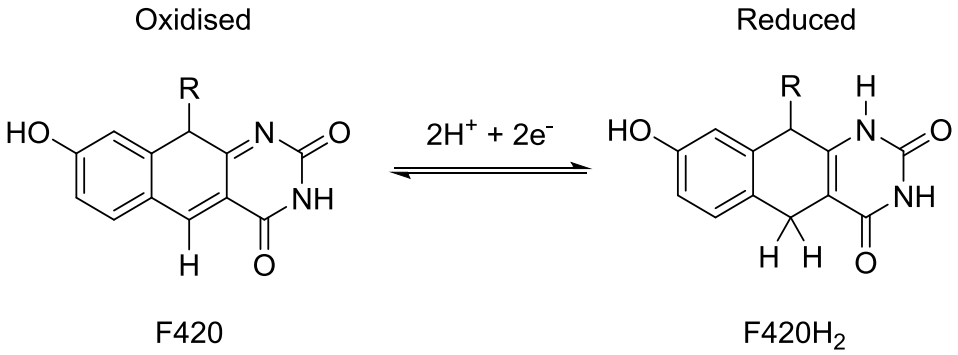


Figure 1.19: The reduction of the flavin moiety of the enzyme cofactor F_{420} to produce the reduced form, $F_{420}H_2$.[46-48]

Two examples of F420-dependent IREDs are TomJ, an F420-dependent IRED in the biosynthesis of the PBD Tomaymycin from *Streptomyces achromogenes* and SibT, an F420-dependent IRED in the biosynthesis of the PBD Sibyromycin from *Streptosporangium sibiricum*, whose biosynthetic roles are outlined in Figure 1.20.

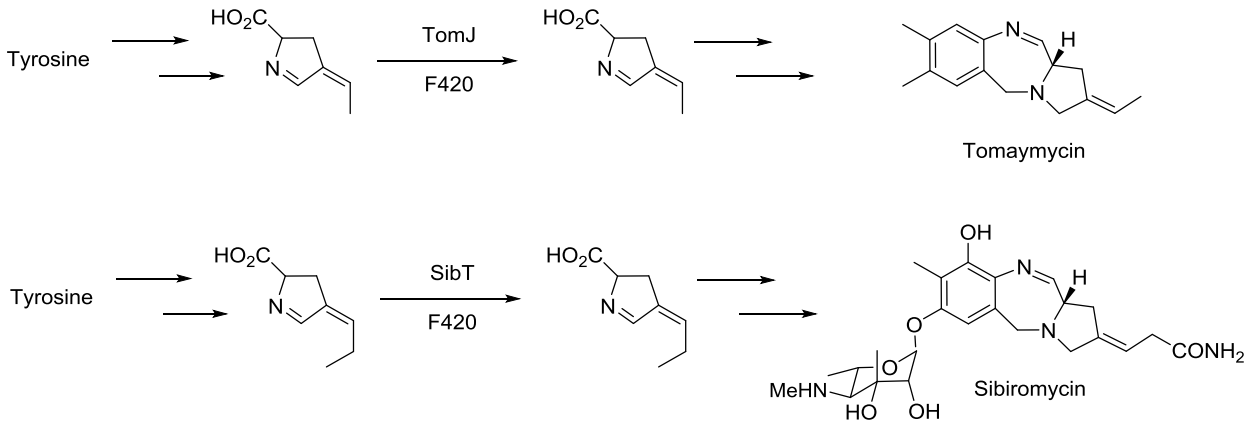


Figure 1.20: The role of F420-dependent IREDs TomJ and SibT in the biosynthesis of PBDs Tomaymycin and Sibiromycin in *Streptomyces achromogenes* and *Streptosporanium sibiricum* respectively. TomJ and SibT are each responsible for the imine reduction step in these biosynthetic pathways.

More recently, IREDs have been sought specifically for their biotechnological applications. *Candida parapsilosis* ATCC 7330, which is known for its ability to reduce α- and β-keto esters[49], was found to also possess IRED activity towards aryl imines, forming the corresponding (*R*)-amine with high enantioselectivity (Figure 1.21).

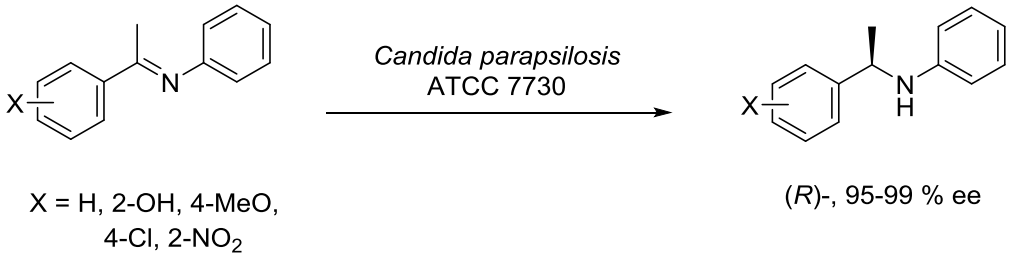


Figure 1.21: Reduction of imine moiety of dihydrofolate by DHFR to form tetrahydrofolate; an intermediary reaction in the biosynthesis of DNA.[50]

Later, multiple strains of *Saccharomyces* were screened for activity towards β-carboline imines (Figure 1.22). Reduction was successfully carried out with moderate to high enantioselectivity in multiple strains of *Saccharomyces,* with *S. bayanus* proving the most effective whole cell biocatalyst of all the strains tested.[51]

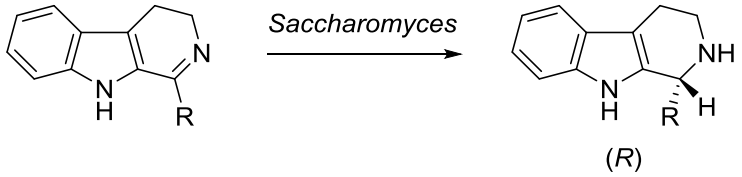


Figure 1.22: Reduction of imine moiety of beta-carboline imines by cell free extract of strains of *Saccharomyces* to form the corresponding (*R*)-amines.[51]

The enantioselective reduction of β-carboline imines was also subsequently carried out successfully using cell free extract from *E. foetida* (earthworms) in the presence of NADPH (Figure 1.23).[52]

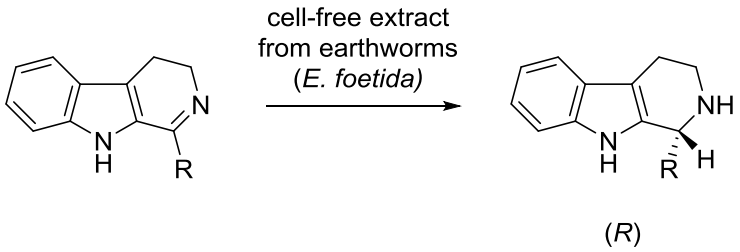


Figure 1.23: Reduction of imine moiety of beta-carboline imines by cell free extract of *E. foetida* to form the corresponding (*R*)-amines.[52]

Most interestingly from a biotechnological standpoint, a broad-ranging screen of activity of organisms towards the imine 2-methyl-pyrroline (**1a**) by Mitsukura and co-workers led to the discovery of two enantiocomplementary strains of *Streptomyces* with IRED activity towards **1a**; *Streptomyces* sp. GF3587 and *Streptomyces* sp. GF3546, which reduced **1a** to enantioselectively form the corresponding amine (*R*)-**1b** and (*S*)-**1b** with 99.2 % ee and 92.3 % ee respectively (Figure 1.24).[53, 54] The sequences of these IREDs, which will be referred to as *SR*-IRED and *SS*-IRED respectively, were determined by the procurement of *N*-terminal amino acid sequence data from purified enzyme. They were found to exclusively exhibit NADPH cofactor dependence.[53, 55]

The gene encoding *SR*-IRED was used to identify an additional homologue, an IRED from *Streptomyces kanamyceticus* (*SkR*-IRED), which was also found to (*R*)-selectively reduce **1** to form (*R*)-**1b** with 99.6 % ee.[55] *SR*-IRED and *SS*-IRED have been recombinantly expressed in *E. coli* and purified to homogeneity by Mitsukura and co-workers.[53, 55] The expression, purification and characterization of *SkR*-IRED had thus far not been reported.

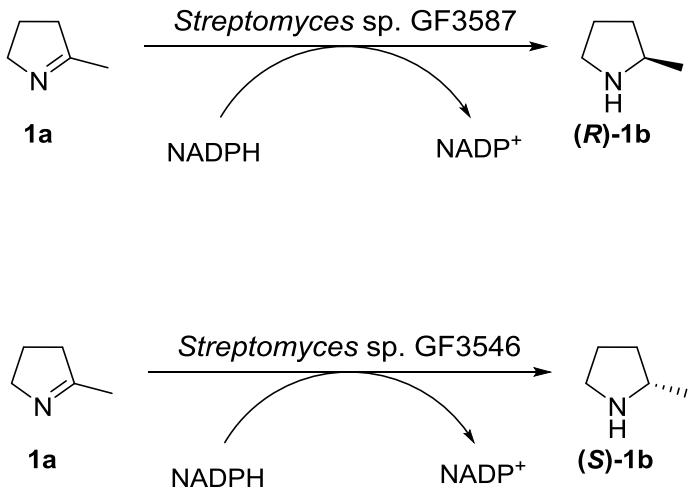


Figure 1.24: Reduction of imine moiety of 2-methyl-1-pyrroline (*R*)-selectively and (*S*)-selectively by enantiocomplementary strains of Streptomyces. Reduction of **1a** using produces amine (*R*)-**1b** with *Streptomyces* sp. GF3587 and (*S*)-**1b** using *Streptomyces* sp. GF3546 [53-56]

1.5 Project aims

The aim of this project will be to increase knowledge about NADPH-dependent IREDs as a class of enzymes. In order to do this, the structure and mechanism of a number of IREDs of this family will be studied.

The practical aims of this project are the cloning, recombinant expression and purification of these enzymes with a view for obtaining crystal structures to inform mechanistic studies as well as engineering for improved catalytic properties and increased substrate range, as outlined in Figure 1.25. Analysis of the structure and function of a range of IREDs will also facilitate investigation into what directs the selectivity of this class of enzymes.

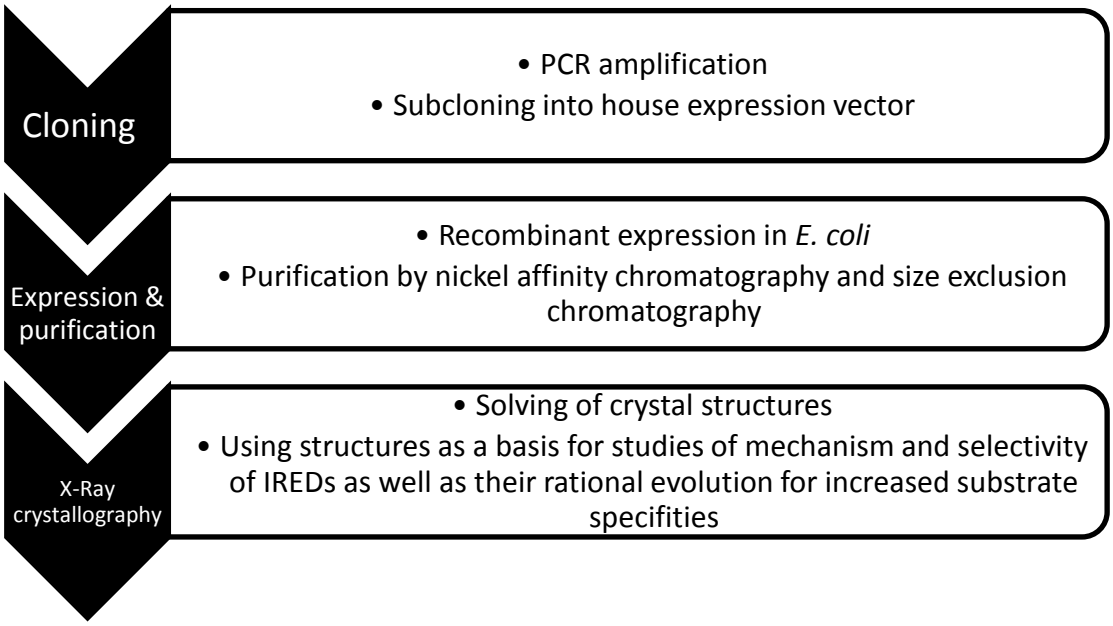


Figure 1.25: A schematic view of the aims of the first part of this project, which will involve the study of NADPH-dependent IREDs.

As the cloning, expression and purification of *SkR*-IRED has not yet been reported, this IRED was investigated first. The crystal structure of this enzyme was solved in preparation for this project, and will now be used to inform mechanistic studies of this enzyme and, therefore, (*R*)-selective IREDs more broadly.

The applications and substrate range of *SS*-IRED has been investigated.[55, 57] However, an *S*-selective NADPH-dependent IRED had not yet been characterised structurally and mechanistically. This will therefore form the next part of the focus of this project in the context of NADPH-dependent IREDs.

Studies of the structures and applications of *SkR*-IRED and *SS*-IRED, which exhibit opposite enantioselectivity, will facilitate investigation into the basis of selectivity in NADPH-dependent IREDs. Additionally, gaining an insight into the mechanism of NADPH-dependent IREDs may ultimately facilitate a sufficient understanding of this enzyme class that this can be used to inform rational mutagenesis experiments to improve their catalytic properties.

1.6 Addendum to the Introduction

The initial discovery of NADPH-dependent IREDs *SR*-IRED and *SS*-IRED by Mitsukura and co-workers attracted the further investigation of these enzymes by a large

number of groups worldwide over subsequent years as this finding newly unlocked the real potential for the use of IREDs in an industrial context. This is because, unlike in the examples outlined in Section 1.4.5, the enzymes responsible for these imine reductions have been successfully identified which therefore facilitates further study, while also operating on a simple, biotechnologically promising substrate.[58]

As a result, this subsection of biocatalysis is a field which grew rapidly throughout the timespan of this PhD, and continual publication of new literature heavily influenced the direction of the project. It is therefore appropriate to include in this introduction literature which has been published since the start of this PhD project began.

The 5-membered cyclic imine **1a** was used as a model substrate in the initial screening of organisms for imine reducing activity due to its stability in aqueous environments as well as its commercial availability. However, further investigations towards the applications of *SR*-IRED and *SS*-IRED carried out by the Turner group at the Manchester Institute of Biotechnology indicated that the substrate specifity of these enzymes is much broader, and includes both substituted and unsubstituted 6-membered and 7-membered rings as well as quinoline derivatives and, in the case of *SS*-IRED, β-carboline derivatives. These findings demonstrated that IREDs of this class offer a broader substrate scope than initially realised, this further increasing their potential for industrial application.[57, 59]

The structure of *SkR*-IRED, which shares high levels of homology with *SS*-IRED and, particularly, *SR*-IRED was solved in both its *apo* form and in complex with NADPH. This structural facilitated the beginnings of investigation towards an understanding of how this class of enzymes works.[60] Subsequently, the structure of *SS*-IRED in complex with NADPH was also determined by Huber and co-workers. The findings demonstrate that the structure of a monomeric subunit of this class of IREDs consists of an *N*-terminal helical domain and a *C*-terminal Rossman-fold motif connected by a long inter-domain helix. The active form of these enzymes consists of a closely formed dimeric complex of two monomers connected by a reciprocal domain sharing arrangement. Dimer formation results in the creation of a hydrophobic cleft which forms at the interface of the *N*-terminal domain of one subunit and the *C*-terminal

domain of the opposite subunit, which contains the active site region of the protein.[61] It is observed that the inter-domain cleft closes upon binding of the NADPH cofactor.[58]

The identification of the genes coding for *SR*-IRED, *SS*-IRED and *SkR*-IRED has facilitated the identification of a substantial numbers of related genes available for sub-cloning, expression, purification and screening for IRED activity. This effort was championed by the Hauer group who developed a database of putative IRED sequences grouped into sub-families according to sequence homology.[62]

A substantial number of these newly discovered genes have since been characterised. Examples include the isolation and characterization of IREDs from *S. roseum*, *S. turgidiscabies* and *Paenibacillus elgii* by Hauer and co-workers. The former two IREDs exhibited (*R*)-selectivity towards **1a** while the latter exhibited (*S*)-selectivity. Höhne and co-workers demonstrated that IREDs could also operate on acyclic imines when they isolated and characterized IREDs from *P. elgii* B69, *S. ipomoeae* 91-03 and *Pseudomonas putida* KT2440. Assays were performed in the oxidative direction and showed that these IREDs could oxidize the acyclic amine, *N*-methyl-2-pentanamine.[63] More recently, Xu and co-workers successfully exploited an IRED from *Paenibacillus lactis* to reduce indoline-based imines and iminium ions to the corresponding optically pure (*S*)-amines.[64]

Grogan and co-workers subsequently cloned and expressed IREDs from *Bacillus cereus* (*BcS*-IRED) and *Nocardiopsis halophilia* (*Nh*-IRED). While *Bc*-SIRED was found to be predominantly (*S*)-selective towards the usual panel of imines, the investigation of *Nh*IRED, in particular, demonstrated that the (*R*)- and (*S*)-selective nature of IREDs can be highly substrate dependent. Selectivity was found to change from (*S*)- to (*R*)- with more sterically demanding substrates.

Furthermore, Iding and co-workers based at Roche characterized 20 IRED enzymes in terms of their substrate scope and selectivity. These findings confirmed that while many IREDs are uniformly selective, there are number of IREDs for which selectivity is substrate dependent. This provided an opportunity to examine the function of many sequences from the available database of theoretical IREDs and seek evidence as to what determines the enantioselectivity of an IRED.[65]

The first two chapters of this thesis are based on the findings of Mitsukura and co-workers, building on their initial characterization of *SS*-IRED, *SR*-IRED and *SkR*-IRED. The literature presented in this addendum had an impact on the final two results chapters of this PhD thesis, as the research reported in these sections builds in the knowledge acquired from the characterization of additional IRED homologues reported in later work.

2. Materials and Methods

2.1 Materials

2.1.1 Bacterial growth media

Lysogeny broth (LB) growth medium was used as standard for the growth of *E. coli* cultures. LB components (Table 2.1) were solubilised in Milli-Q water.

Table 2.1: Components required to make LB medium

Component	Concentration (g L^{-1})
Sodium chloride	10
Yeast extract	5
Tryptone	10

Where necessary, LB medium was supplemented with 30 μg mL^{-1} kanamycin (LB-Kan).

Where increased cell yield was required, the richer medium terrific broth (TB) was used. TB components (Table 2.2) were solubilised in Milli-Q water.

Table 2.2: Components required to make TB medium

Component	Concentration
Yeast extract	24 g L^{-1}
Tryptone	12 g L^{-1}
Glycerol	5 g L^{-1}
NaH_2PO_4	17 mM
Na_2HPO_4	72 mM

2.1.2 Buffers for protein purification and enzyme assays

Buffer A (Table 2.3) was used as a buffer for protein storage and size exclusion chromatography.

Table 2.3: Components required to make buffer A

Component	Final concentration
Tris	50 mM
NaCl	300 w/v
Adjusted to pH 7.5 with conc. HCl	
Diluted to total volume in Milli-Q water	

Buffer B (Table 2.4) was used as a buffer for the resuspension of expression cells, and in nickel affinity chromatography.

Table 2.4: Components required to make buffer B

Component	Final concentration
Tris	50 mM
NaCl	300 w/v
Imidazole	30 mM
Adjusted to pH 7.5 with conc. HCl	
Diluted to total volume in Milli-Q water	

Buffer C (Table 2.5) was used as an elution buffer in nickel affinity chromatography.

Table 2.5: Components required to make 1 L buffer C

Component	Final concentration
Tris	50 mM
NaCl	300 w/v
Imidazole	500 mM
Adjusted to pH 7.5 with conc. HCl	
Diluted to 1 L total volume in Milli-Q water	

2.1.3 Gel electrophoresis buffers

Agarose gel electrophoresis

Table 2.7: Components required to make TAE buffer for agarose gel electrophoresis

Component	Concentration
Tris	40 mM
Acetic acid	20 mM
EDTA	1 mM

To prepare a single agarose gel, 0.6 g agarose was dissolved in 60 mL TAE buffer by heating in a microwave.

Sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE) buffers

Table 2.8: Components required to make 4 x loading buffer for SDS-PAGE

Component	Concentration
Tris	120 mM
Glycerol	20 % w/v
SDS	4 % w/v
β-mercaptoethanol	20 % w/v
Bromophenol blue	0.04 % w/v
Adjusted to pH 8.0 with conc. HCl	

Table 2.9: Components required to make running buffer for SDS-PAGE

Component	Concentration
Tris	25 mM
Glycine	200 mM
SDS	0.1% w/v

Table 2.10: Components required to make sufficient resolving gel for SDS-PAGE for 2 SDS-PAGE gels

Component	Concentration	Volume
Deionised water	-	3.2 mL
Acrylamide	40 % w/v	4.5 mL
Resolving buffer (1.5 M Tris.HCl containing 0.1 % w/v SDS, pH 8.8)	-	2.5 mL
APS	10 % w/v	20 μL
TEMED	-	8 μL

Table 2.11: Components required to make sufficient gel for SDS-PAGE for 2 SDS-PAGE gels

Component	Concentration	Volume
Deionised water	-	3.2 mL
Acrylamide	40 % w/v	0.5 mL
Stacking buffer (0.5 M Tris.HCl containing 0.4 % w/v SDS, pH 6.8)	-	1.3 mL
APS	10 % w/v	20 µL
TEMED	-	8 µL

Table 2.12: Components required for preparation of glycin-NaOH buffer (0.1 M, pH 9.5)

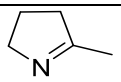
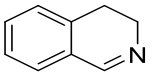
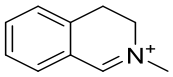
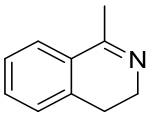
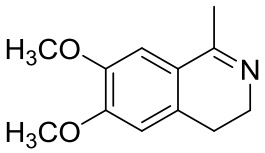
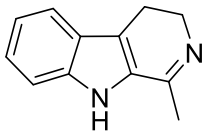
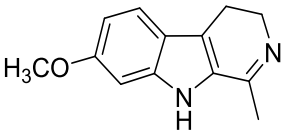
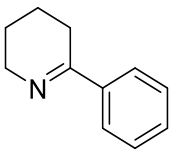
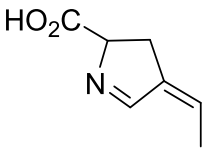
Component	Concentration
Glycine	100 mM
Adjusted to pH 9.5 with 0.2 M NaOH	
Dissolved to 1 L total volumn in milliQ water	

Table 2.13: Components required for preparation of phosphate buffer (0.1 M, pH 7.5)

Component	Mass
Na_2HPO_4	13.40 g L^{-1}
NaH_2PO_4	2.63 g $^{-1}$
Dissolved to 1 L total volume in milliQ water	

2.1.4 Substrates and products

Table 2.14: Imine, aldehyde and alcohol substrates employed in IRED reactions

Abbreviation	Molecular Formula	Structure	MW (g mol^{-1})
1a	C_5H_9N		83.07
2a	C_9H_9N		131.07
3a	$C_{10}H_{12}N$		146.10
4a	$C_{10}H_{11}N$		145.09
5a	$C_{12}H_{15}NO_2$		205.11
6a	$C_{12}H_{12}N_2$		184.10
7a	$C_{13}H_{14}N_2O$		214.11
8a	$C_{11}H_{13}N$		159.10
9a	$C_7N_9NO_2$		139.15

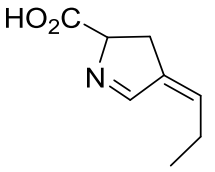
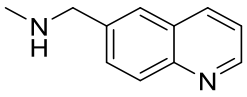
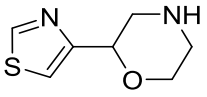
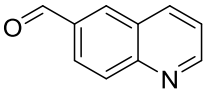
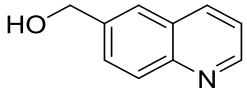
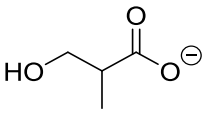
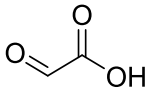
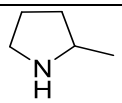
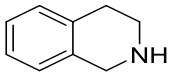
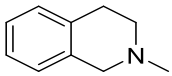
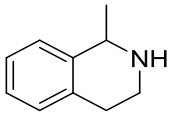
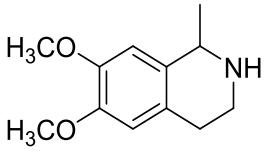
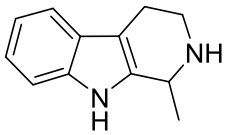
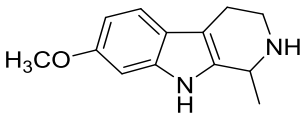
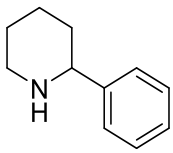
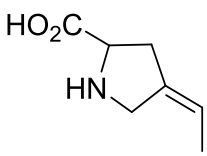
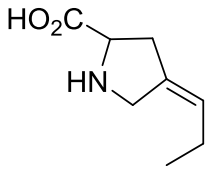
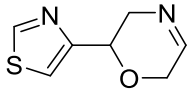
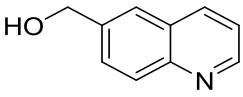
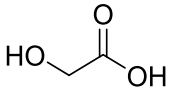
Abbreviation	Molecular Formula	Structure	MW (g mol^{-1})
10a	$C_8H_{11}NO_2$		153.18
11a	$C_{11}H_{12}N_2$		172.10
12a	$C_7H_{10}N_2OS$		170.23
13a	$C_{10}H_7NO$		157.17
14a	$C_{10}H_9NO$		159.07
15a	$C_4H_7O_3$		103.10
16a	$C_2H_2O_2$		74.04

Table 2.15: Products for IRED reactions. Products have not been included where the product compound is already included in the list of substrates

Abbreviation	Molecular Formula	Structure	MW (g mol^{-1})
1b	C_5H_9N		85.09
2b	C_9H_9N		133.09
3b	$C_{10}H_{12}N$		148.12
4b	$C_{10}H_{13}N$		147.11
5b	$C_{12}H_{17}NO_2$		207.13
6b	$C_{12}H_{14}N_2$		186.12
7b	$C_{13}H_{16}N_2O$		216.13
8b	$C_{11}H_{15}N$		161.12
9b	$C_7H_{11}NO_2$		141.17

Abbreviation	Molecular Formula	Structure	MW (g mol^{-1})
10b	$C_8H_{13}NO_2$		155.20
12b	$C_7H_8N_2OS$		168.04
15b	$C_{10}H_9NO$		159.19
16b	$C_2H_4O_3$		76.05

2.2 Molecular cloning

Target genes were subcloned into pET-YSBL-LIC3C (Figure 2.1).[66] The vector confers kanamycin resistance onto *E*. *coli* cells, and a cleavable *N*-terminal 6His tag onto gene products. The vector is adapted from the pET system, which is designed to facilitate the production of large copy numbers of the target protein within the host cell.

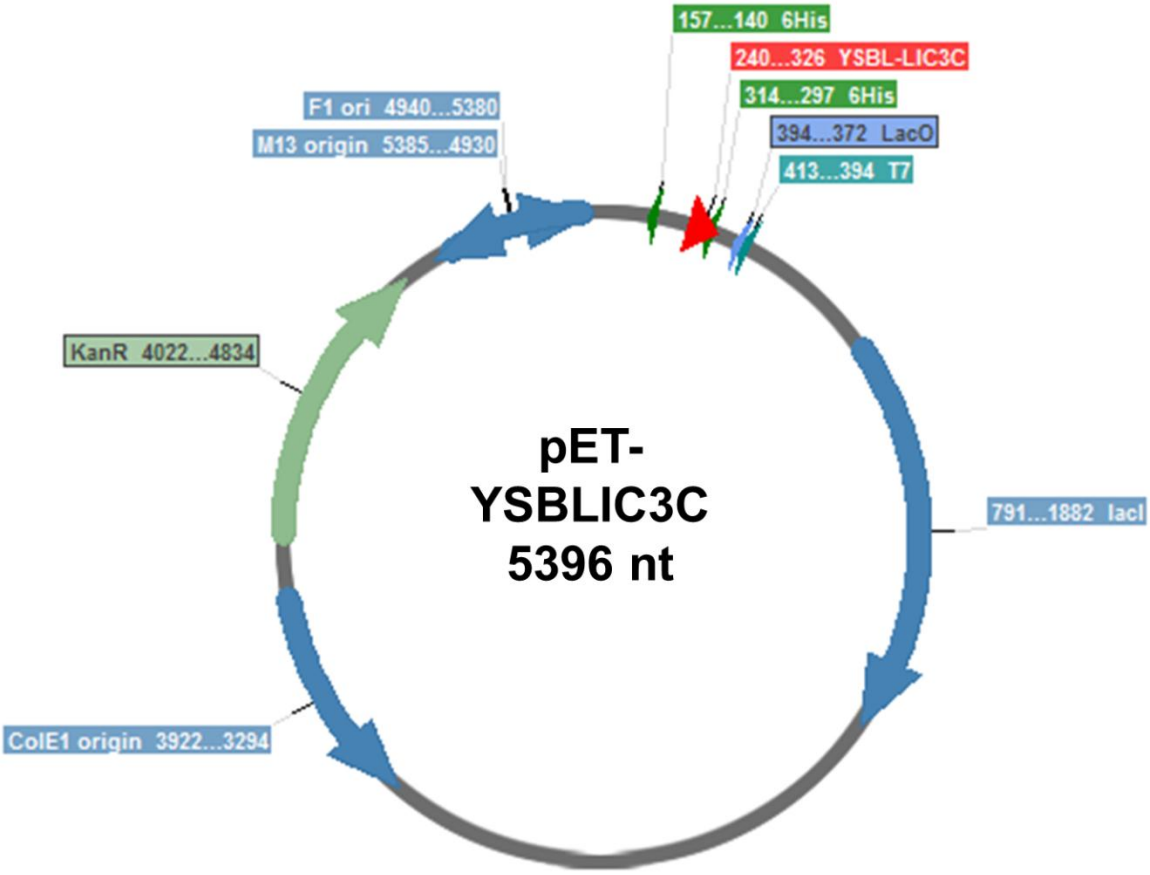


Figure 2.1: The pET-YSBLIC3C expression vector, based on the widely used pET28a vector. pET28a carries an N-terminal 6His tag coding sequence, a kanamycin antibiotic resistance marker, a multiple cloning site, the T7 promoter sequence, the lac operator coding sequence and the lacI gene which codes for the lactose repressor protein, Additionally to the pET28a features, pET-YSBLIC3C contains the coding sequence for a human rhinovirus 3C (HRV3C) protease cleavage site downstream of the 6His tag, allowing for the cleavage of the N-terminal 6His tag from expressed proteins. The pET-YSBLIC3C expression vector has also been modified to include a ligation independent cloning (LIC) site with a BseRI cleavage site.[66]

2.2.1 Polymerase chain reaction

The polymerase chain reaction (PCR) is a temperature-controlled, molecular biological technique for generating large copy numbers of the desired template gene using a three-step process; denaturation, annealing and extension.

Initially the two strands of parent DNA are reversibly separated at approximately 95 °C in the denaturation step. Custom oligonucleotides designed to anneal to both termini of the target DNA sequence then bind to the sense and antisense strands at approximately 55 °C in the annealing step. The annealing temperature is dictated by the melting temperature (T_m) of the primers. In the extension step, a new strand of DNA is synthesised by a DNA polymerase by adding dNTPs complementary to the template from the 3' end of each oligonucleotide using deoxynucleotide triphosphate monomeric units (dNTPs). The extension step is usually carried out at 72 °C, depending on the DNA polymerase used. This process is repeated over 25-35 cycles until many copies of the desired gene have been achieved. PCR is carried out using a thermal cycler, which facilitates the rapid temperature changes required by the reaction. A schematic diagram of the denaturing, annealing and extending step cycle is shown in Figure 2.2.

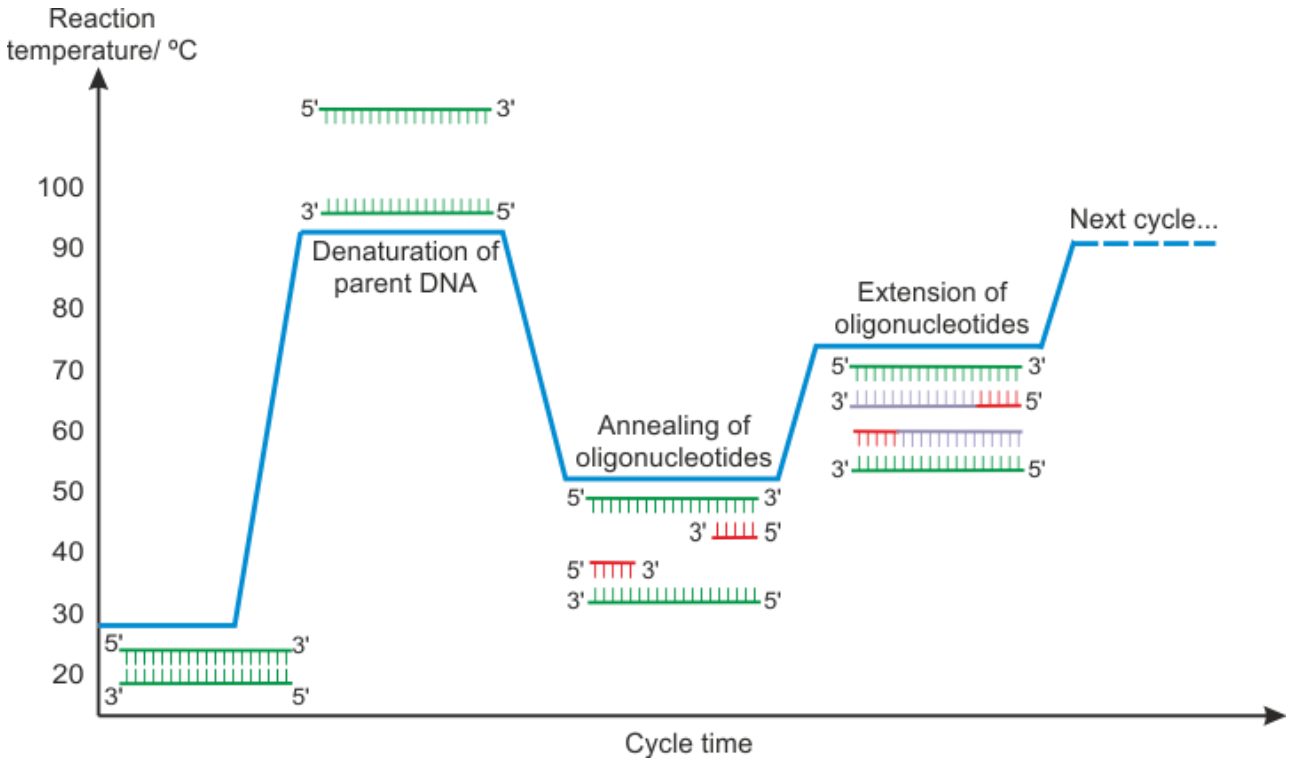


Figure 2.2: The repeating cycle of denaturing the target DNA template, annealing of complementary oligonucleotide primers and extension of new DNA using DNA polymerase in a PCR reaction for the amplification of target DNA sequences

PCR components

To a 500 µL PCR tube, the components listed in Table 2.16 were added.

Table 2.16: The reaction components used in standard PCR for the amplification of target DNA sequences

Component	Stock concentration	Volume
Plasmid DNA template	100 ng μL^{-1}	1 µL
Forward primer	20 pmol μL^{-1}	1 µL
Reverse primer	20 pmol μL^{-1}	1 µL
dNTPs	2 mM (each nucleotide)	5 µL
$MgSO_4$	25 mM	3 µL
Merck KOD Hot Start DNA polymerase buffer	10 x	5 µL
Merck KOD Hot Start DNA polymerase	1.0 U μL^{-1}	1 µL
Milli-Q water to 50 µL		

PCR cycling parameters

The following parameters were employed as standard for PCR using a thermal cycler as outlined in Table 2.17.

Table 2.17: Cycling parameters used as standard PCR protocol for the amplification of target DNA sequences in a thermal cycler.

Step	Temperature/ °C	Time
Initial denature	94	180 s
Denature	94	30 s
Annealing	55	30 s
Extension	72	60 s
Final extension	72	180 s

2.2.2 Ligation-independent cloning (LIC)

Ligation-independent cloning is a method for molecular cloning without the use of restriction endonucleases or DNA ligases. This technique was used as standard for all instances of sub-cloning of target genes into the pET-YSBLIC3C expression vector.

This technique exploits the 3'-5' exonuclease activity of T4 DNA polymerase in order to generate lengthy (10-15 bp) overhangs in both the target gene and expression vector which facilitate the spontaneous annealing of gene and vector.

Linearisation of expression vector

To a 1.5 mL Eppendorf, pET-YSBLIC3C vector DNA (12.5 µg), *Bse*RI (12.5 µL), NEB 10 x Buffer 2 (25 µL) were added. The volume was diluted to 250 µL with the appropriate volume of Milli-Q water. The reaction was placed in an incubator (37 °C, 110 min). The digested vector DNA was loaded onto an agarose gel (1 % agarose TAE, 100 V, 90 min) and the band corresponding to linearized DNA excised and purified using a Sigma gel extraction it according to the manufacturer's instructions. DNA concentration was then determined using an Eppendorf Biophotometer.

T4 DNA polymerase treatment of expression vector

To a 1.5 mL Eppendorf, *Bse*RI-linearised vector (1.0 pmol), Merck T4 DNA polymerase 10 x buffer (10 µL), dTTP (25 mM, 10 µL) and DTT (100 mM, 5 µL). The reaction volume was diluted to 250 µL with the appropriate volume Milli-Q water. The reaction mixture was divided into multiple 50 µL aliquots in 500 µL PCR cups and incubated in a PCR thermocycler at 22 °C for 30 min followed by 75 °C for 20 min. The reaction was purified using a Qiagen PCR clean up kit according to manufacturer's instructions, and eluted in autoclaved Milli-Q water. DNA concentration was measured using an Eppendorf Biophotometer with Milli-Q water as a blank.

T4 DNA polymerase treatment of target gene

To a 500 µL PCR tube, the PCR amplified target gene (0.2 pmol), T4 10 x DNA polymerase buffer (2 µL), dATP (25 mM, 2 µL), DTT (100 mM, 1 µL) and LIC qualified T4 DNA polymerase (2.5 U μL^{-1}, 0.5 µL). The reaction mixture was diluted to 20 µL with the appropriate volume Milli-Q water and incubated in a PCR thermocycler at 22 °C for 30 min followed by 75 °C for 20 min.

2.1.2.4 Annealing

To a 500 µL Eppendorf tube, 3 µL T4 DNA polymerase-treated gene insert and 2 µL T4 DNA polymerase-treated linearised pET-YSBLIC3C was added. The reaction was incubated at 22 °C for 10 min before EDTA (25 mM, 1 µL) was added. The reaction was then incubated for a further 10 min.

2.1.2.5 Transformation

To a 1.5 mL Eppendorf, chemically competent *E. coli* XL10 Gold cells (50 µL) and SDM DNA product (5 µL) containing the desired mutation was added. The mixture was incubated on ice for 20 min followed by heat shock (42 °C, 45 s) before returning to ice for 2 min. To the heat shocked cells, SOC medium was added (150 µL) and the resulting mixture incubated in an orbital shaker (37 °C, 60 min, 180 rpm). The cells were subsequently plated out onto LB+Kan plates and incubated (37 °C, 16 h).

DNA isolation

From each plate, two colonies were selected and cultured in LB-Kan medium (5 mL) and incubated in an orbital shaker (37 °C, 16 h, 180 rpm). Cells were harvested by centrifugation (13,000 g, 1 min), from which plasmid DNA was isolated using a Sigma miniprep kit according to manufacturer's instructions. DNA was eluted in 50 µL Milli-Q water. DNA concentration was measured using an Eppendorf biophotometer.

Diagnostic restriction digest

A restriction digest was carried out to determine the presence of an appropriately sized insert in the purified plasmid DNA. To a 1.5 mL Eppendorf, plasmid DNA (7 µL, 80 - 100 ng μL^{-1}), *Nde*I (1 µL), *Nco*I (1 µL) and 10 x NEB CutSmart buffer (1 µL) was added. The digestion reaction was incubated at 37 °C for 3 h. The sample was mixed with 6 x DNA loading dye (2 µL) and loaded onto an agarose gel (1 % agarose TAE, 100 V, 25 min). DNA bands were visualised using a Blue Light transilluminator and assessed for the presence of an appropriately sized DNA insert.

Sequencing

Clones that appeared positive according to diagnostic restriction digest were diluted to 20 µL at 60 ng μL^{-1} and submitted to sequencing at GATC Biotech to verify the presence of the correct insert.

2.3 Protein production

2.3.1 Transformation

To a 1.5 mL Eppendorf, chemically competent *E. coli* BL21 (DE3) cells (25 µL) and pET-YSBLIC3C vector containing the target gene (100 ng) was added. The mixture was incubated on ice (20 min) and heat shocked (42 °C, 45 s) before returning to ice (2 min). To the heat shocked cells, SOC medium was added (150 µL) and the resulting mixture incubated in an orbital shaker (37 °C, 60 min, 180 rpm). The cells were subsequently plated out onto LB+Kan plates and incubated (37 °C, 16 h).

2.3.2 Small scale expression tests

For novel genes for which expression conditions had not previously been established, small scale expression tests were carried out.

Inoculation of liquid cultures

Cell material from a single colony was used to inoculate an LB-Kan preculture (10 mL). The precultures were incubated in an orbital shaker (37 °C, 16 h, 180 rpm). Four aliqots of LB-Kan medium (10 mL) in 50 mL falcon tubes were inoculated with preculture (100 µL) and incubated in an orbital shaker (37 °C, 180 rpm) until an OD_{600} of 0.8 was reached.

Induction of protein expression

At an OD_{600} of 0.8, three of the four cultures were induced with 200 µg mL^{-1} Isopropyl β-D-1-thiogalactopyranoside (IPTG) and one culture incubated in separate orbital shakers at 16 °C, 30 °C and 37 °C (180 rpm, 20 h).

Harvesting of cells

The induced cell cultures were harvested by centrifugation (15 min, 5000 g). The fourth culture was harvested at an OD_{600} of 0.8 by centrifugation without induction by IPTG (15 min, 5000 g).

Cell lysis

Cell pellets were resuspended in buffer A (1 mL) and disrupted by sonication (3 x 30 s) with incubation on ice during and between pulses. Sonicated cells were subjected to centrifugation (30 min, 13,000 g, 4 °C).

Preparation of samples for SDS-PAGE analysis

Soluble fractions were decanted into separate 2 mL Eppendorf tubes, and the insoluble pellets suspended in buffer A (1 mL). For both soluble and insoluble samples, samples were taken (10 µL) and mixed with 2 x SDS-PAGE (10 µL) loading buffer. All samples were then heat denatured (94 °, 5 min).

SDS-PAGE analysis

To prepare a 12 % acrylamide SDS-PAGE gel, the resolving gel components (Table 2.9) were mixed and loaded into a Mini-PROTEAN® Tetra Cell casting stand loaded with two 1.0 mm thickness BioRad Mini-PROTEAN® spacer plates and two short glass plates. The resolving gel was topped with butanol (100 µL). After the resolving gel had set (approx. 30 min), the butanol layer was removed with compressed air. The stacking gel components (Table 2.10) were mixed and loaded atop the resolving gel. A BioRad Mini-PROTEAN® comb (1.0 mM thickness, 10-15 well) was inserted into the stacking gel. After the stacking gel had set (approx. 45 min), the comb was removed.

To run a 12 % acrylamide SDS-PAGE gel, the gels were transferred within their glass plates to a BioRad electrophoresis cell. Samples were loaded into sample wells and the gel was run (200 V, 50 min).

Samples were run alongside a BioRad low range molecular weight marker.

2.3.3 Large scale protein production

Inoculation of liquid cultures

Cell material from a single colony on an agar plate was used to inoculate 2 LB-Kan precultures (10 mL). Precultures were incubated in an orbital shaker (37 °C, 16 hours, 180 rpm) and used to inoculate two LB-Kan cultures (1 L) in baffled 2 L Erlenmyer flasks. Flasks were incubated in an orbital shaker (37 °C, 180 rpm) until an OD_{600} of 0.8 was reached.

Induction of protein expression

Cells were then induced with ITPG (200 µg mL^{-}) and incubated in an orbital shaker (16 °C, 20 h, 180 rpm) at an OD_{600} of 0.8.

Harvesting of cells

Cells were harvested by centrifugation in Beckman 1 L centrifuge bottles (10 min, 5000 g, 4 °C).

2.4 Protein purification

Target proteins were routinely purified from cell lysate in two steps; nickel affinity chromatography and size exclusion chromatography.

2.4.1 Cell lysis

Cell pellets were resuspended in buffer B (40 mL) and disrupted by sonication (13,000 amplitude microns, 6 x 30 s) with incubation on ice. Cell free extract was obtained in supernatant after centrifugation (16,000 g, 45 min, 4 °C).

2.4.2 Nickel affinity chromatography

Nickel affinity chromatography (figure 2.3) exploits the 6His tag conferred onto proteins by the pET-YSBLIC3C vector. The resin of the nickel affinity column is charged with Ni^{2+} ions, which chelate the 6His tag at the terminus of the protein while untagged host proteins from *E. coli* are not bound and elute through the column. Proteins are eluted in an increasing gradient of imidazole, for which the Ni^{2+} ions have a very high affinity. Imidazole displaces the bound 6His tagged proteins, allowing them to be eluted from the column.

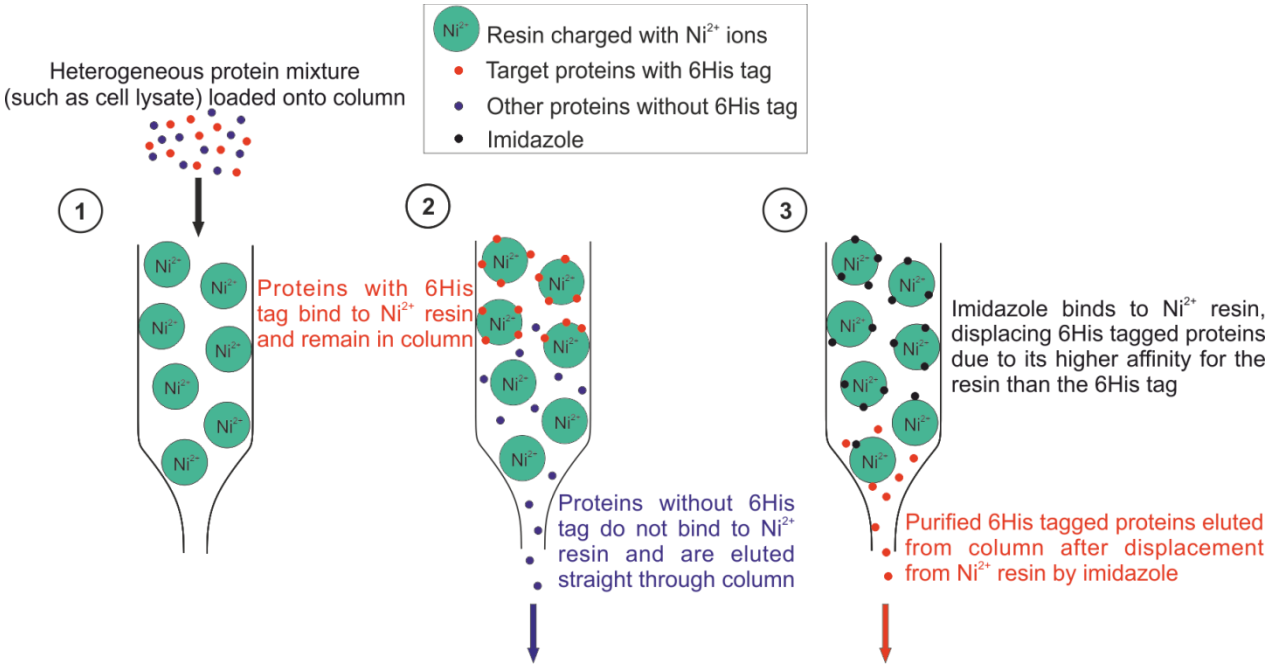


Figure 2.3: A schematic representation of a nickel affinity chromatography column. Target proteins tagged with the 6His tag have a high affinity for the Ni^{2+} resin, therefore binding to the column medium. The remaining proteins are eluted through the column without binding, allowing them to be removed from the mixture. The 6His-tagged target protein is then displaced from the column by imidazole, which has a higher affinity for Ni^{2+}, and is eluted from the column as pure protein.

Lysate was loaded onto a HisTrap HP column (5 mL) charged with nickel sulfate using a bench-top peristaltic pump (5 mL min^{-1}). Bound proteins were then eluted with an increasing imidazole gradient on an Äkta purifier (30 – 500 mM imidazole, 100 mL, 2 mL min^{-1}, 2 mL fractions). Fractions were analysed by SDS-PAGE, and those containing the target protein were pooled.

2.4.3 Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) (figure 2.4) is a technique used to separate proteins according to their molecular weight. It is particularly useful for obtaining proteins exclusively in their correct oligomeric state, as different oligomerisations will elute at distinct elution volumes. SEC columns contain a fine resin through which heavier proteins elute faster, as they are unable to pass through the narrower channels within the beads. Smaller proteins are able to pass through narrower pores in the resin of the column, and therefore traverse a longer route before eluting from the column, thus eluting more slowly.

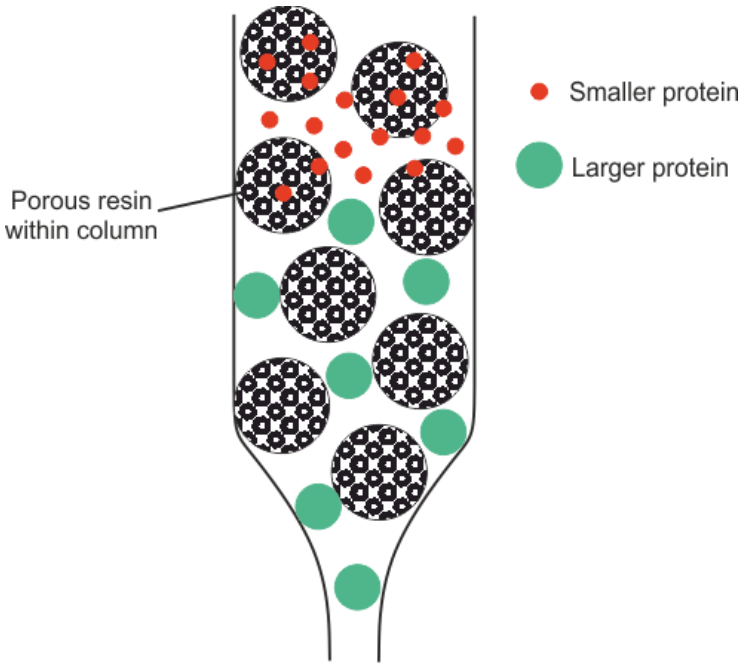


Figure 2.4: A schematic representation of a SEC column. The resin contains small pores through which smaller proteins are able to flow, causing them to take a longer route through the column. Larger proteins do not fit through these pores and therefore bypass them, resulting in a faster route through the column. Larger proteins are therefore eluted more quickly than smaller ones.

After purification by nickel affinity chromatography, the target protein was concentrated in a 30 kDa Amicon Ultra-15 centrifugal filter (5,000 g, 4 °C) to a volume of 2 mL. The protein was filtered using membrane syringe filter (0.45 μM) to remove particulate matter. The protein was loaded onto a SuperDex HiLoad 16/60 S75 size exclusion column via a 2 mL injection loop and eluted in buffer A (120 mL, 0.5 mL min^{-1}). Fractions were analysed by SDS-PAGE. Those containing pure protein in its correct oligomeric state were pooled for subsequent use.

2.4.4 Protein concentration

Protein concentration was determined by measuring absorbance at 280 nm using an Eppendorf Biophotometer with buffer A as a blank. The concentration could be calculated using the Beer-Lambert law,

$$A = \varepsilon . c . l$$

A = Measured absorbance

ε = Exctinction coefficient

l = Path length

The absorption coefficient was ascertained using the ExPASy ProtParam tool, and the pathlength of the measurement cuvette was 1 cm.

2.4.5 Protein storage

For long term storage, protein was aliquoted flash frozen in liquid nitrogen and stored (-20 °C) until use.

2.5 6His tag cleavage

The pET-YSBLIC3C expression vector codes for a cleavage site (Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro-Ala) recognised by the human rhinovirus 3C protease (HRV 3C) downstream from the *N*-terminal 6His tag. It was therefore possible to cleave the 6His tag from proteins expressed in the pET-YSBLIC3C expression vector using HRV 3C. Proteins were diluted to a concentration of 1 mg mL^{-1}, and 1 mg HRV 3C protease per 100 mg protein substrate was added. The reaction was incubated (4 °C, 16 h). The cleaved protein was purified using nickel-affinity chromatography (section 2.3.2), where the cleaved protein was eluted before the imidazole gradient due to its lack of binding capacity due to the absence of the 6His tag.

2.6 Analytical techniques

2.6.1 Ultraviolet-visible (UV) spectrophotometry

Ultraviolet-visible (UV-Vis) spectrophotometry is routinely used in analytical chemistry to characterise molecules that absorb light within the ultraviolet or visible regions. Many transition metal and unsaturated organic compounds are able to absorb UV or visible light as they contain π- or non-bonding electrons that can absorb energy from photons with wavelengths in the UV or visible regions that excite them into higher energy anti-bonding molecular orbitals.

The UV-Vis spectrophotometer instrument operates by generating a beam of light at a given wavelength. Both the intensity of the incident beam before it reaches the sample (I_0) and the intensity of the beam transmitted from the sample (I) are measured. The ratio I/I_0 is known as the transmittance and can be used to calculate the absorbance, a quantitative measure of the amount of radiation absorbed by a sample:

$$A = -log_{10}\left(\frac{I}{I_0}\right)$$

The Beer-Lambert states that the absorbance of a solution is directly proportional to the concentration of an absorbing species and the path length, allowing the absorbance to be directly converted into the concentration of the species of interest.

$$A = \varepsilon . c . l$$

A = Measured absorbance

I_0 = Incident intensity of light at a given wavelength (λ)

I = Transmitted intensity of light at a given wavelength (λ)

l = Path length of the measurement vessel

c = Concentration of sample

ε = Extinction coefficient

UV-Vis spectrophotometry may also be used to follow the progression of a reaction using time course measurements, where absorbance and thus reactant or product concentration is measured at regular intervals. This technique can be used to determine reaction rates, and is therefore useful for determining kinetic parameters in enzyme reactions.

2.6.2 HPLC

HPLC is an analytical technique for the separation of chemicals in a mixture. It is a form of liquid column chromatography achieved by pumping solvents through the column at high pressures, thus greatly increasing the speed of analysis.

An HPLC system consists of a sampler, pumps and a detector (figure 2.5). The sampler introduces the sample mixture into the mobile phase (solvent). The sample is carried through the column via the mobile phase. The pump dictates the flow rate and composition of the mobile phase as it passes through the column. The detector detects peaks corresponding to molecules from the sample as they are eluted from the column, generating a signal that is proportional in size to the corresponding molecule.

HPLC can be carried out using a normal phase or reversed phase setup. In the normal phase, the stationary phase (the column) is polar and the mobile phase (solvent) is non-polar. Non-polar molecules elute more slowly through the column due to increased Van der Waals interactions with the non-polar mobile phase. Reversed phase HPLC refers to a non-polar stationary phase and a polar mobile phase. Polar molecules in the liquid sample passing through the column display an increased affinity towards the polar mobile phase, leading to longer retention times relative to non-polar molecules. This principle allows for the separation of molecules according to their polarity.

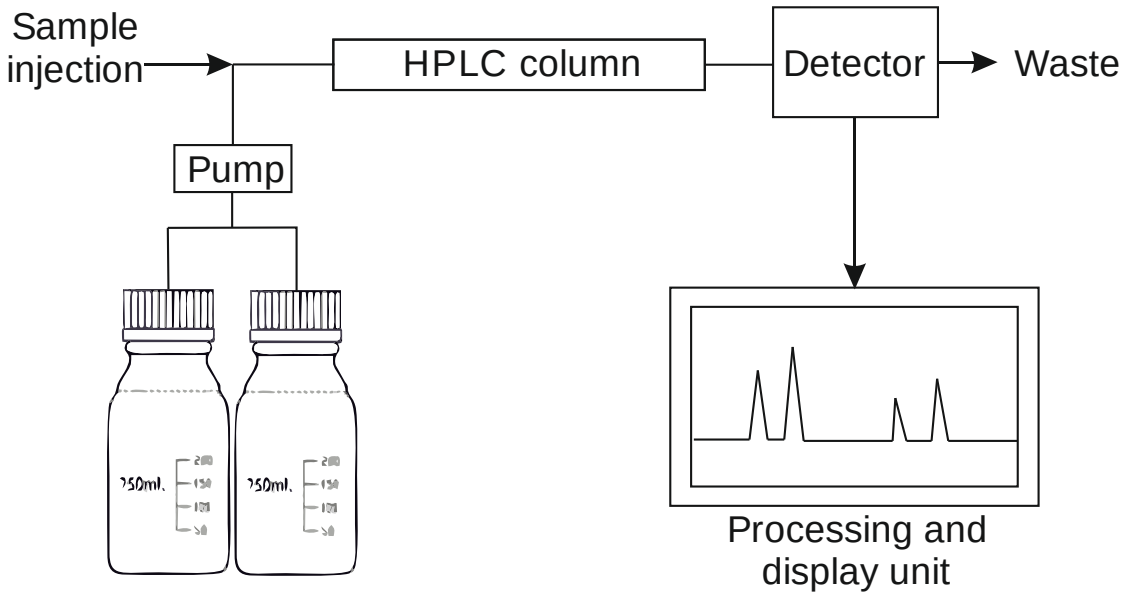


Figure 2.5: A schematic diagram of the experimental setup of an HPLC. The sample is injected into the column, through which a gradient of buffers is pumped. The speed at which analytes travel through the column is determined by the solubility of the compound in the solvent and the degree of interaction of the analyte to the stationary phase. Analytes eluted from the column are detected, producing a signal that is visualised using a processing display unit.

2.6.3 GC

Gas chromatography (GC) is an analytical technique for the separation of chemicals in a mixture. The equipment consists of a sample injector, a column where separation occurs, a detector and a processing and display unit (figure 2.6). Small sample volumes are injected at the injector and analytes are carried through the column by a gaseous mobile phase. The stationary phase is situation within the column and is normally a high boiling point liquid absorbed onto a solid. The speed at which the analytes pass through the column is determined by how strongly they are attracted to the stationary phase. Compounds more strongly attracted to the stationary phase pass through the column more slowly as they more readily condense or dissolve in the liquid of the stationary phase.

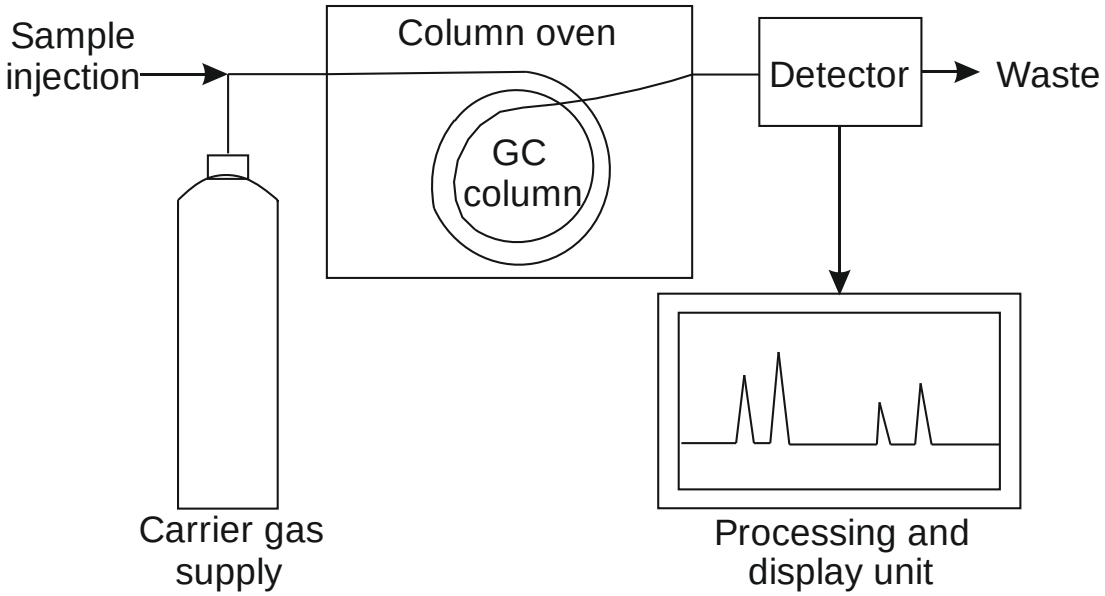


Figure 2.6: A schematic diagram of the experimental setup of a GC. The sample is injected and transported through the column via a carrier gas. The speed at which analytes are eluted from the column is determined by their boiling point and the degree of affinity to the stationary phase. Analytes eluted from the column are detected at the detector, producing a signal that are then visualised using a processing display unit.

2.7 Enzyme kinetics

All assays were carried out with purified enzyme unless otherwise specified.

2.7.1 Michaelis-Menten kinetics

The Michaelis-Menten model is widely used for describing ideal enzyme kinetic behaviour. An enzymatic reaction in equilibrium may be described by the equation:

$$E + S \rightleftharpoons ES \rightarrow E + P$$

Where substrate, *S*, and enzyme, *E*, bind reversibly to form the enzyme-substrate complex, *ES*, and the reaction of the *ES* complex to form the product, *P*, is assumed to be irreversible under a large excess of concentration of *S* relative to *P*.

For such a system, the Michaelis-Menten equations are derived:

$$v = \frac{V_{max}[S]}{K_{\mathrm{M}} + [S]}$$

$$V_{max} = k_{cat}[E_0]$$

Where,

V_{max} = The maximum reaction velocity of the system

K_M = The Michaelis-Menten constant

[S] = Substrate concentration

k_{cat} = Catalytic turnover number of enzyme

$[E]_0$ = Initial enzyme concentration

The V_{max} describes the maximum reaction velocity for the system at saturating substrate concentrations.

The k_{cat} may be calculated directly from the V_{max} and the enzyme concentration. This is the turnover number of the enzyme, and describes the number of molecules turned over by one molecule of enzyme per second.

The K_M is the substrate concentration at which half the V_{max} is achieved. This serves as a measure of the affinity of the substrate to an enzyme, where a lower K_M indicates tighter binding of substrate and enzyme, and a higher K_M indicates weaker binding, as more substrate is required to achieve the maximum reaction rate.

A general measure of the efficiency of an enzyme can be measured using the specifity constant, k_{cat}/K_M. This value may be used to compare the efficiency of enzyme-substrate systems.

The kinetic parameters V_{max}, k_{cat} and K_M can be determined using an experimental setup where reaction velocity is measured with constant concentration of enzyme at increasing substrate concentrations. The Michaelis-Menten model predicts that the enzyme will adopt the kinetic behaviour depicted graphically in Figure 2.7 where reaction rate increases steadily until the enzyme reaches substrate saturation.

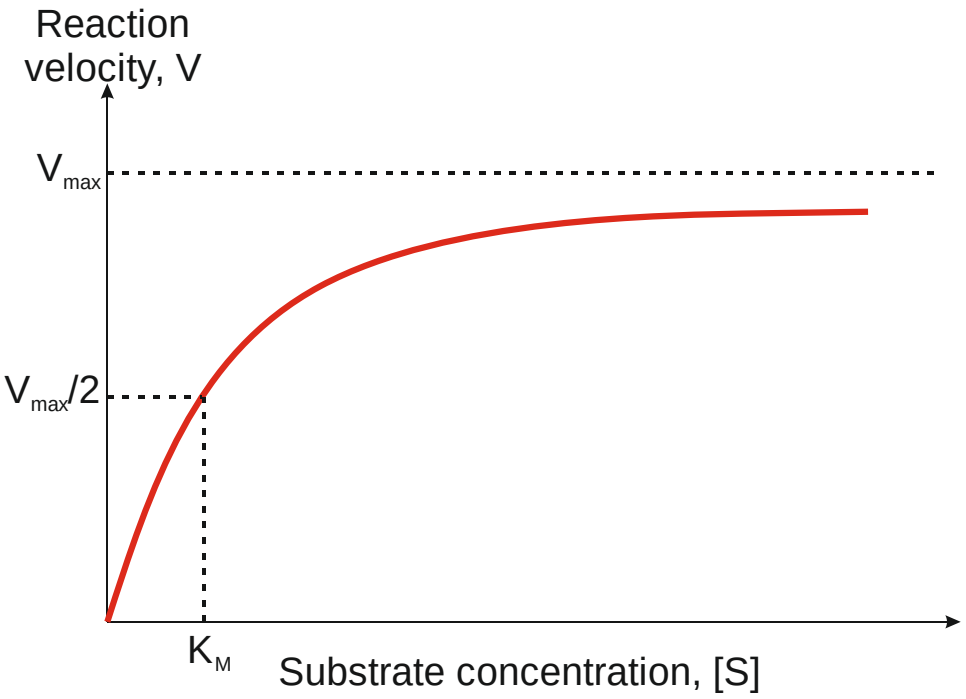


Figure 2.7: A representation of a typical Michaelis-Menten plot for enzyme kinetics. Reaction velocity is plotted against substrate concentration. A maximum velocity is reached as a certain substrate concentration where the enzyme becomes saturated. This velocity is termed V_{max}. The substrate concentration at which an enzyme velocity of half the V_{max} is achieved is termed the K_M, and is a measure of enzyme affinity for the substrate.

2.7.2 Substrate inhibition kinetics

In some cases, as substrate concentration increases, reaction velocity does not converge to a maximum value as is the case in the Michaelis Menten model, but instead begins to decrease (figure 2.8). In such cases an inactive enzyme complex, ESS, is said to be formed at high concentrations of substrate, where two substrate molecules bind to the enzyme in the following manner.

$$E + S \rightleftharpoons ES \rightarrow E + P$$
$$\uparrow\downarrow +S$$
$$ESS$$

From this, an equation describing the substrate inhibition model is derived,

$$v = \frac{V_{max}\,[S]}{K_M + [S]\left(\frac{1 + [S]}{K_i}\right)}$$

The substrate inhibition equation shares the parameters V_{max} and K_M with the Michaelis Menten equation with the additional parameter, K_i, which represents the inhibition constant, corresponding to rate constant of formation of the inactive ESS complex. K_i acts as a measure of the potency of substrate inhibition, with higher values indicating a stronger inhibition effect.

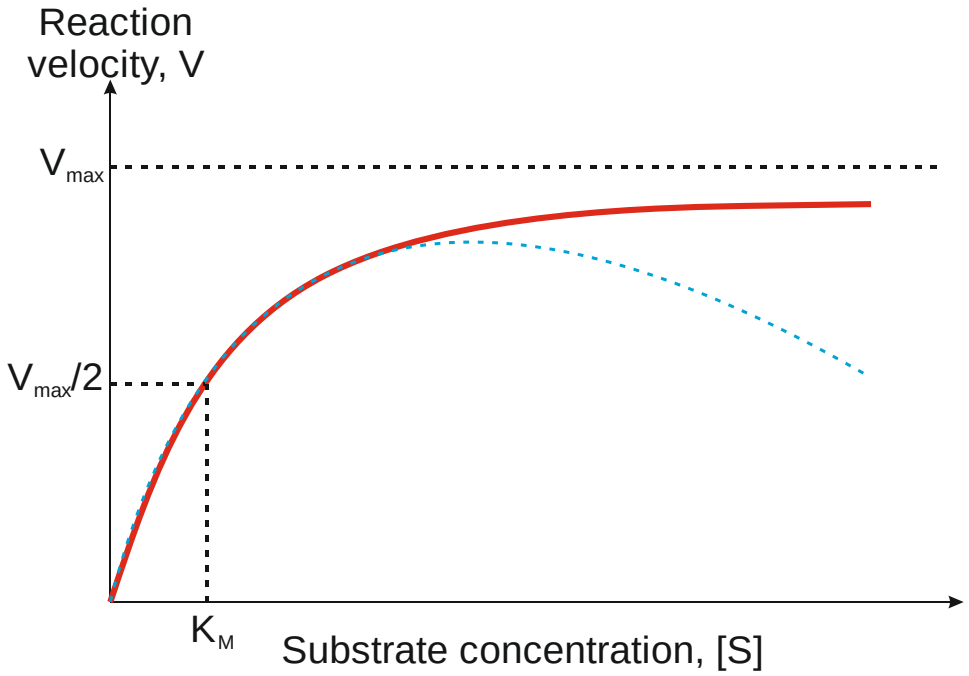


Figure 2.8: A representation of a substrate inhibition model reaction (dotted blue line) superposed onto a typical Michaelis-Menten model (red line). Reaction velocity is plotted against substrate concentration. Unlike in a Michaelis-Menten system, maximum velocity (V_{max}) is not reached at higher substrate concentrations as the substrate begins to exhibit an inhibitory effect on the system. V_{max} and K_M are determined by fitting a curve using a model which includes an inhibition constant, K_I, to account for substrate inhibition.

2.7.3 NADPH concentration assays

The cofactor NADPH is reversibly oxidised to $NADP^+$ in enzymatically catalysed reduction reactions. Both NADPH and $NADP^+$ absorb strongly in the UV region due to the presence of the adenine moiety, however their absorbance behaviours differ considerably. NADPH has a λ_{max} of 340 nm, while $NADP^+$ does not absorb at this wavelength. Therefore, the concentration of NADPH may be monitored by measuring the depletion in absorbance at 340 nm. This allows for the indirect monitoring of the

progression of the reaction, as the concentration of NADPH is directly proportional to substrate depletion. Such a technique may be used to measure both the depletion of NADPH in reductive reactions and the depletion of $NADP^+$ in oxidative reactions by measuring the decrease or increase in absorbance at 340 nm respectively.

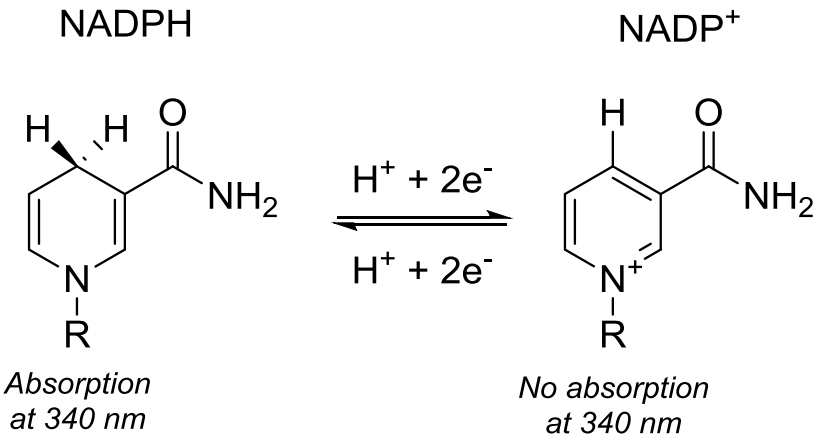


Figure 2.9: The oxidation of the enzyme cofactor NADPH to form $NADP^+$. This oxidation can be detected spectophotometrically, as NADPH absorbs at 340 nm while $NADP^+$ does not.

All NADPH concentration assays were carried out in a GCB-DBUV spectrophotometer (Cintra10) in a quartz cuvette (1 mL). Reactions (1 mL) were carried out with the components outlined in Table 2.16.

Table 2.18: Reaction components used in enzyme kinetic assays

Component	Volume	Final concentration
*NADPH/$NADP^+$	20 µL	0.2 mM
Substrate (solubilised in ethanol)	10 µL	Increasing from 0.5 mM until saturation achieved
Enzyme	100 µL	0 – 10 mg mL^{-1}
Tris.HCl pH 7.5	870 µL	50 mM

All reaction components excluding enzyme were mixed in a quartz cuvette (1 mL). Enzyme was then added to initiate the reaction, which was monitored at 370 nm over a time course of 3 – 15 min depending on reaction rate. Assuming the absorption coefficient of NADPH to be 2300 M^{-1} cm^{-1}, the raw absorbances were converted to the concentration of NADPH at a given time point. Using the assumption that the change in concentration of NADPH is inversely proportional to change in substrate concentration for reductive reactions, or directly proportional for oxidative reactions, reaction velocities could be determined.

2.7.4 Data analysis

With the GraphPad Prism software suite, non-linear regression analysis was used to determine best fit values for the kinetic parameters K_M, k_{cat} and K_i for datasets of reaction rates vs. substrate concentration for given enzyme-substrate systems using either the Michaelis-Menten or substrate inhibition model.

An F-test was performed to compare which model (Michaelis-Menten or substrate inhibition) most closely fit the data, where:

H_0 = Michaelis Menten model

H_a = Substrate inhibition model

Critical value = 0.05.

2.8 Biotransformations

2.8.1 Biotransformations by *SkR*-IRED

Biotransformations using *SkR*-IRED were carried out to ascertain the selectivity and substrate scope of the enzyme. Reactions containing imines **1a** – **8a** (5 mM), NADPH (0.2 mM) and purified enzyme (0.8 mg mL^{-1}) were incubated (20 h, 25 °C, 250 rpm).

GITC derivatization procedure

For analysis of biotransformations of **1a**, functionalisation with the chiral derivatizing agent 2,3,4,6-Tetra-*O*-acetyl-β-D-glucopyranosyl isothiocyanate (GITC) was carried out, generating diastereomers that could be separated by an achiral reverse phase HPLC column.

An aqueous sample (100 µL) of the **1a** biotransformation was mixed with triethylamine (100 µL, 0.8% (w/v)) in acetonitrile and 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl Isothiocyanate (GITC) (200 µL, 0.8% (w/v)). The mixture was incubated (37°C, 30 min). Insoluble fractions were removed by centrifugation and the supernatant analysed by HPLC.

Extraction procedure

For samples of biotransformations of **2a** – **8a**, extraction with *t*-butyl methyl ether (TBME) was carried out in preparation for analysis by HPLC. Reactions (1 mL) were extracted with an equal volume of TBME by vortexing. Samples were centrifuged (4,000 g, 10 min) and the TBME fraction decanted into a clean vial for HPLC analysis.

Chromatography conditions

For analysis of biotransformations of *SkR*-IRED towards **1a**, reverse phase HPLC for samples of GITC-derivatised **1a** and **1b** was performed on an Agilent system equipped with a Luna® C18(2) column (3 µm particle size) was used. The typical injection volume was 10 µL and chromatograms of absorbance at 254 nm were recorded.

For analysis of biotransformations with *SkR*-IRED towards imines **2a – 8a**, normal phase chiral HPLC was performed on an Agilent system equipped with either a CHIRALPAK®IC Analytical (5 µm particle size) or CHIRALCEL®OD-H (5 µm particle size). The typical injection volume was 10 µL and chromatograms of absorbance at 265 nm were recorded.

Table 2.19: HPLC conditions used for the analysis of imines

Compound	Column	Solvent	Solvent ratio/ v/v	Flow/ mL min^{-1}	Retention time/ min
2a	CHIRALPAK®IC[	Hexane:*i*PrOH:DEA	90:10:0.1	1	14.03
3a	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	90:10:0.1	1	5.21
4a	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	97:03:0.1	1	17.12
5a	CHIRALCEL®OD-H	Hexane:*i*PrOH:DEA	90:10:0.1	1	8.75
6a	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	90:10:0.1	1	8.60
7a	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	80:20:0.1	1	7.22
8a	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	95:05:0.1	1	9.63

Table 2.20: HPLC conditions used for the analysis of amines

Compound	Column	Solvent	Solvent ratio/ v/v	Flow/ mL min^{-1}	Retention time (*S*)/ min	Retention time (*R*)/ min
1b-GITC	Eclipse XDB-C18	KPB pH 2.5:MeOH	55:45	1	16.067	16.899
2	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	90:10:0.1	1	10.777 (prochiral)	
3	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	90:10:0.1	1	6.571 (prochiral)	
4	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	97:03:0.1	1	9.829	10.699
5	CHIRALCEL®OD-H	Hexane:*i*PrOH:DEA	90:10:0.1	1	12.606	16.781
6	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	90:10:0.1	1	10.056	11.154
7	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	80:20:0.1	1	11.867	14.097
8	CHIRALPAK®IC	Hexane:*i*PrOH:DEA	95:05:0.1	1	5.976	5.372

2.8.4 Biotransformations using mutants of *SS*-IRED and *SR*-IRED

Biotransformations of **1a** using mutants of *SS*-IRED and *SR*-IRED were carried out in order to assess the effect of mutations on enantioselectivity using chiral GC.

Extraction and acetylation procedure

Reactions containing **1a** (5 mM), NADPH (20 mM) and purified enzyme (0.2 – 0.5 mg mL^{-1}) were diluted to a volume of 500 µL with phosphate buffer (100 mM, pH 7.5) and incubated (25 °C, 250 rpm). Extract was transferred to a GC vial containing a 200 µL inlet. Imines and amines in the reaction mixture were functionalised with an acetyl group to improve resolution of separation in GC analysis. Samples were acetylated by addition of acetic anhydride (15 µL) trimethylamine (30 µL).

Chromatography conditions

For analysis of biotransformations of **1a** by mutants of *SS*-IRED and *SR*-IRED, samples containing acetylated **1a** and the corresponding product **1b** were analysed using an Agilent 6890 GC equipped with a chiral SGE Analytical Science BP20. A split ratio of 5.0:1 and a flow rate of 9.7 mL min^{-1} was employed. Column conditions and analyte retention times are outlined in Table 2.20.

Table 2.21: Chromatography conditions employed in the chiral GC analysis of biotransformations of **1a** by mutants of *SS*-IRED and *SR*-IRED

Conditions	Retention time (min)		
	Imine (**1a**)	(*R*)-Amine ((*R*)-**1b**)	(*S*)-Amine ((*S*)-**1b**)
150 °C 22 min, 20 °C/min to 200 °C, 5 min at 200 °C	16.3	20.1	18.7

2.8.5 Biotransformations using mutants of *Pp*-IRED

Biotransformations were carried out with substrates **11a** – **13a** (5 mM), $NADP^+$ (for the oxidation of **11a** and **13a**) and NADPH (for the reduction of **12a**) and purified enzyme (0.5 mg mL^{-1}) diluted to a volume of 500 µL with phosphate buffer (100 mM, pH 7.5). Reactions were incubated (20 h, 25 °C, 250 rpm).

Extraction procedure

Reactions were extracted with TBME (250 µL) containing pentadecane (0.15 % v/v) as an internal standard. Extract was transferred to GC vials containing 200 µL inlets.

Chromatography conditions

Biotransformations of substrates **11a** – **13a** by *Pp*-IRED were analysed using an Agilent 6890 GC equipped with an HP-5 column. A split rate of 5.0:1 and a flow rate of 26.7 mL min^{-1} were used. The analyte retention times and column conditions applied are outlined in Table 2.21.

Table 2.22: Chromatography conditions employed in the chiral GC analysis of biotransformations of **11a** – **13a** by Pp-IRED

Conditions	Retention time (min)		
	11a	**12a**	**13a**
160 °C, 10 min	5.6	8.4	9.3

2.9 Site-directed mutagenesis (SDM)

Site-directed mutagenesis is a molecular biological tool for introducing a specific and targeted mutation into a DNA sequence allowing the mutation of a chosen single amino acid residue to a chosen target residue. SDM is commonly used for investigating the importance of certain residues in enzymatic activity or to alter the properties of a protein in a targeted way.

To carry out SDM, primers complementary to the target DNA sequence containing the desired mutation were designed. The primers anneal to the target DNA and copies containing the desired mutation are synthesised using PCR. The template DNA is digested by the restriction endonuclease *Dpn*I, which exclusively digests methylated DNA. The genetic material is then only the unmethylated DNA synthesised during the PCR, which can then be used to transform a host organism for cloning. Selected colonies are cultured and the transformed is DNA isolated using a miniprep procedure. Isolated DNA is sequence to screen for the presence of the desired mutation (Figure 2.10).

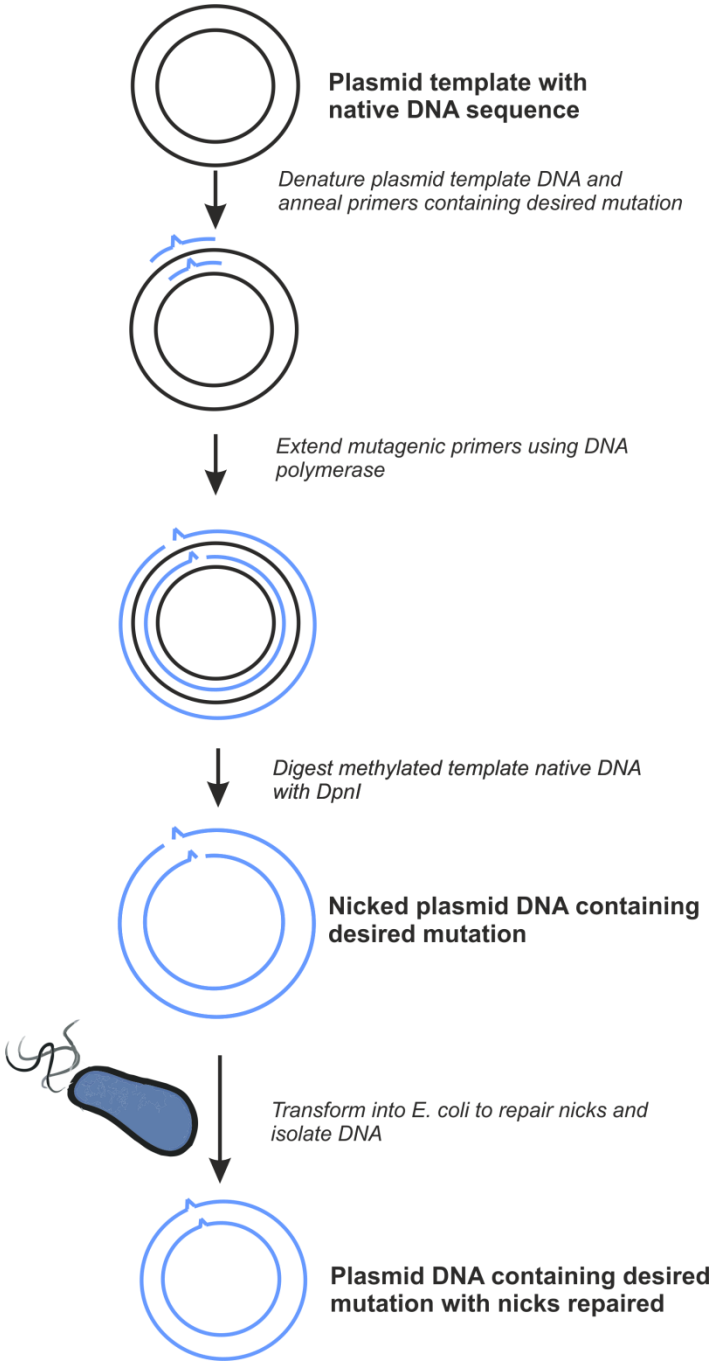


Figure 2.10: A schematic representation of the steps involved in site-directed mutagenesis (SDM). The template DNA is denatured at high temperatures, allowing the mutagenic primers to anneal. The plasmid is then amplified from the mutagenic primers. The template DNA is removed by digestion by *Dnp*I, and *E. coli* is transformed with the newly synthesised mutant DNA for the repair of nicks and cloning.

2.9.1 PCR

PCR reaction components

Table 2.23: PCR reaction components employed for site directed mutagenesis

Component	Stock concentration	Volume/ µL
Plasmid DNA template	100 ng μL^{-1}	1
Forward primer	20 pmol μL^{-1}	1
Reverse primer	20 pmol μL^{-1}	1
dNTPs	2 mM (each nucleotide)	5
Agilent Pfu Turbo DNA polymerase buffer	10 x	5
Agilent Pfu Turbo DNA polymerase	1.0 U μL^{-1}	1
Milli-Q water to 50 µL		

PCR cycling parameters

Table 2.24: PCR cycling parameters for site directed mutagenesis

Step	Temperature/ °C	Time
Initial denature	95	300 s
Denature	95	60 s
Annealing	55	60 s
Extension	72	600 s
Final extension	72	300 s

2.9.2 *Dpn*I Digest

To the PCR reaction, *Dpn*I (2 units μL^{-1}, 1.5 μL) is added to digest template DNA. The digest reaction is incubated (37 °C, 4.5 h).

2.9.3 Transformation

To a 1.5 mL Eppendorf, chemically competent *E. coli* XL10 Gold cells (50 μL) and SDM DNA product (5 μL) containing the desired mutation was added. The mixture was incubated on ice for 20 min followed by heat shock (42 °C, 45 s) before returning to ice for 2 min. To the heat shocked cells, SOC medium was added (150 μL) and the resulting mixture incubated in an orbital shaker (37 °C, 60 min, 180 rpm). The cells were subsequently plated out onto LB+Kan plates and incubated (37 °C, 16 h).

2.9.4 DNA isolation

From each plate, two colonies were selected and cultured in LB-Kan medium (5 mL) and incubated in an orbital shaker (37 °C, 16 h, 180 rpm). Cells were harvested by centrifugation (13,000 rpm, 1 min), from which plasmid DNA was isolated using a Sigma miniprep kit according to manufacturer's instructions. DNA was eluted in 50 μL Milli-Q water. DNA concentration was measured using an Eppendorf biophotometer.

2.9.5 DNA Sequencing

Purified plasmids were diluted to 20 μL at 60 ng μL^{-1} and submitted to sequencing at GATC Biotech to verify the presence of the desired mutation.

2.10 Sample preparation for electrospray mass spectrometry

SkR-IRED was highly concentrated to 100 mg mL^{-1} and resuspended in ddH_2O to remove Tris buffer which might interfere with mass spectrometry analysis. The sample was ionized by electrospray ionisation (ESI) and analysed using FT-MS by the Technology Facility at the University of York.

3. Isolation, application, structure and mechanism of *SkR*-IRED, an (*R*)-selective NADPH-dependent imine reductase from *Streptomyces kanamyceticus*

3.1 Introduction

The asymmetric reduction of the prochiral C=N bond in imines by an enantioselective enzyme offers a convenient route for the synthesis of chiral amines, which offer a broad range of applications in the pharmaceutical, agrochemical and chemical industries. Prior to 2010, there existed few reports of imine-reducing enzymes (IREDs) suited for industrial biotechnological applications.

In an effort to identify IREDs with applicability to biocatalysis, a broad ranging screen of organisms for activity towards imine **1a** was carried out by Mitsukura and co-workers, which led to the discovery of two organisms that catalysed the reduction of the imine to the corresponding amine **1b**. The two strains produced enantiocomplementary amines; *Streptomyces sp.* GF3587 and *Streptomyces sp.* GF3546, which catalysed the formation of amines (*R*)-**1b** and (*S*)-**1b** with 99.2 % and 92.3 % ee respectively.[54] The enzymes responsible for these reactions, which will be referred to as *SR*-IRED and *SS*-IRED respectively, were isolated and their sequence was identified using *N*-terminal sequence acquisition. The enzymes were characterized and reported to be NADPH-dependent.[53, 55] Knowledge of the sequence of *SR*-IRED and *SS*-IRED presented an opportunity to identify homologues which could be screened for additional imine reductase activities. Knowledge of the gene sequences of *SR*-IRED and *SS*-IRED facilitated the discovery of an additional (*R*)-selective IRED in *Streptomyces kanamyceticus,* which has been termed *SkR*-IRED, which was also found to catalyse the reduction of **1a** to form enantiomerically pure (*R*)-**1b** (Figure 3.1).[56]

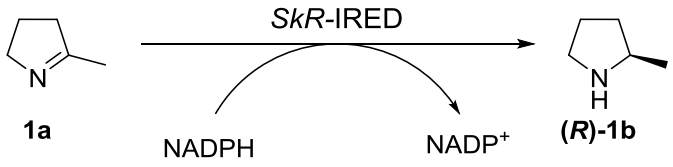


Figure 3.1: The NADPH-mediated reduction of imine **1a** to form enantiomerically pure (*R*)-**1b** by *SkR*-IRED.

Prior to the beginning of this project, no characterization of *SkR*-IRED had been reported, presenting the opportunity to study a novel IRED, which forms the basis of this chapter. This section describes the characterization of the structure, substrate scope, kinetic behaviour and mechanistic studies of *SkR*-IRED.

This enzyme was the first enzyme that became the primary focus of this project not only because it was novel, but because it was readily crystallisable. Attempts to crystallise *SR*-IRED, which had been characterized by Mitsukura and coworkers, were not successful and therefore *SkR*-IRED was a more ideal target for mechanistic studies.

3.2 Preliminary work

Preceding this PhD project, preliminary work was carried out on *SkR*-IRED by M. Rodriguez-Mata.

3.2.1 Crystallization and data collection

Purified *SkR*-IRED was subjected to broad ranging crystallization screens. An initial hit was achieved in a 300 nL drop containing 1:1 v/v ratio of protein and precipant using a protein concentration of 20 mg mL^{-1} and tacsimate (pH 7.0, 35 % v/v) in a 96-well sitting drop format in the commercially available INDEX screen (Hampton research). The crystallization scale was increased using 2 µL hanging drops containing a 1:1 v/v ratio of protein and precipitate. A selection of crystals was soaked in mother liquor containing NADPH (10 mM) in order to obtain a cofactor complex. Crystallographic data was collected at Diamond light source using the I03 and I04-1 beamlines for *apo* and NADPH-complexed crystals respectively. Crystals of *SkR*-IRED diffracted to a resolution of 2.6 Å in its *apo* form and 2.7 Å in complex with NADPH.

3.2.2 Structure solution

The structure of *SkR*-IRED was solved by molecular replacement using tartronic semialdehyde reductase from *Salmonella typhymurium* as a model (PDB accession code 1YB4), with which *SkR*-IRED shares 21.1 % identity and 54.6 % similarity. The unit cell contained four monomeric units arranged as two pairs of dimers.

SkR-IRED monomer

A monomeric subunit of *SkR*-IRED (Figure 3.2) consists of an *N*-terminal Rossman fold domain characteristic of NADPH-binding proteins (residues 1-181) and a *C*-terminal helical domain (residues 210 – 306) connected by a long inter-domain helix (residues 182 – 209).[60]

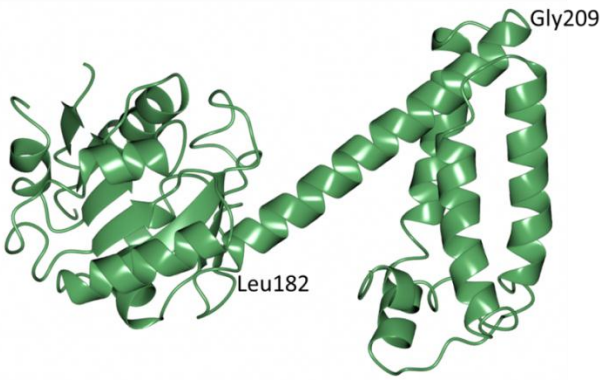


Figure 3.2: The crystal structure of a single monomeric subunit of *SkR*-IRED at a resolution of 2.6 Å. The monomer consists of an *N*-terminal Rossman fold motif (residues 1 – 181) and a C-terminal helical domain (residues 210 – 306) connected by a long inter-domain helix (Leu182 – Gly209). [60]

Dimer formation

SkR-IRED exists as a dimer of two monomeric subunits (A and B) (Figure 3.3) participating in reciprocal domain sharing. Analysis using PISA demonstrated that the two subunits share a contact area of 3761 Å^2. The surface of the inter-domain helix is largely hydrophobic, which assists in its binding into a hydrophobic channel of the opposite subunit in dimer formation. In the dimer, a cleft is formed between the *N*-terminal subunit of monomer A and the *C*-terminal subunit of monomer B, with the interlocking *C*-terminal domains forming the central region of the dimeric complex and the *N*-terminal domains existing at the terminal regions.[60]

Figure 3.3: The crystal structure of apo *SkR*-IRED at 2.6 Å resolution. The monomer consists of an *N*-terminal Rossman fold motif (1 – 181) and a C-terminal helical domain (210 – 306) connected by a long inter-domain helix (Leu182 – Gly209). Dimer formation is facilitated by the reciprocal domain sharing of the two monomeric subunits A (green) and B (orange).[60]

NADPH binding

NADPH binds in a cleft between the *N*-terminal domain of one subunit and the *C*-terminal domain of the opposite subunit (Figure 3.4) at the consensus GXGXXG NADPH-binding motif, which in *SkR*-IRED is GLGMLG 26-31.

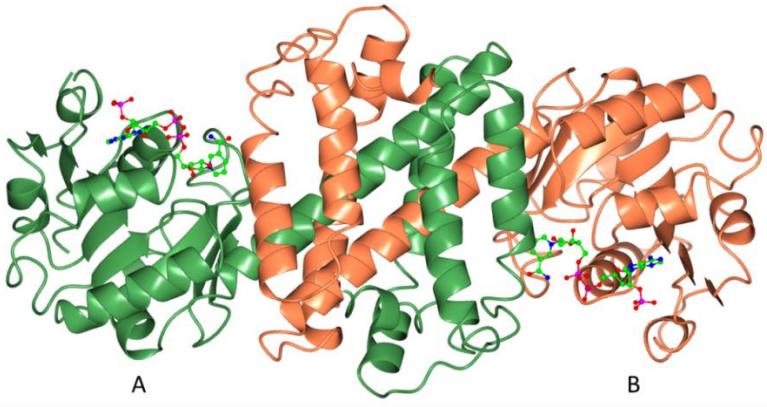


Figure 3.4: *SkR*-IRED in complex with NADPH cofactor at 2.7 Å resolution. The positioning of NADPH indicates that the active site is formed in a cleft between the N-terminal domain of one subunit and the C-terminal domain of the opposite subunit. NADPH is illustrated by ball and stick representation.[60]

3.3 Aims

The crystal structure of *SkR*-IRED will be used as a basis to study the mechanism of the enzyme and to investigate its physiological role. Additionally, the substrate scope and catalytic properties of *SkR*-IRED will be characterized. Ultimately, this structural, mechanism and activity data may be used as a basis for the rational engineering of the enzyme towards improved catalytic properties.

3.4 Results and discussion

3.4.1 Large scale expression and purification

SkR-IRED was expressed in the pET-YSBLIC3C vector in *E. coli* BL21 (DE3) cells in 2 L LB-Kan medium at the previously established optimal expression conditions (16 °C, 180 rpm, 20 h).[60] Protein was purified from lysed, resuspended cells using nickel affinity chromatography, and protein samples were analysed by SDS-PAGE to assess protein purity (Figure 3.5).

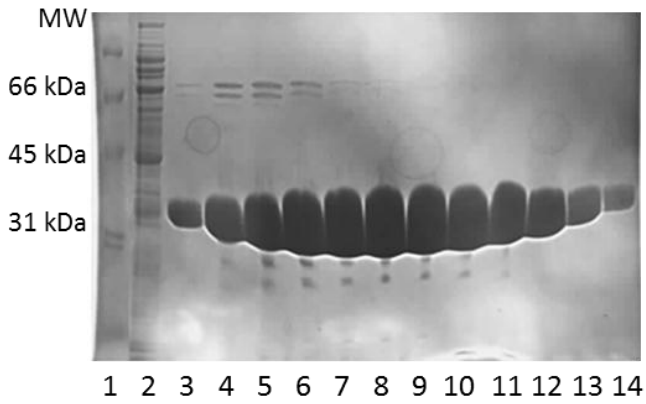


Figure 3.5: SDS-PAGE gel analysis of purification of *SkR*-IRED by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: unbound flow-through collected during loading of nickel column, lanes 3 – 14: fractions containing purified protein eluted during increasing imidazole gradient. Purified *SkR*-IRED is observed to be approximately 32 kDa by SDS-PAGE.

SkR-IRED was further purified by size exclusion chromatography (figure 3.6) using a Superdex® HiLoad 75 gel filtration column to remove remaining host proteins and *SkR*-IRED in different oligomeric states. Fractions which exhibited absorbance at 280 nm, indicating presence of protein, were analysed by SDS-PAGE (Figure 3.7).

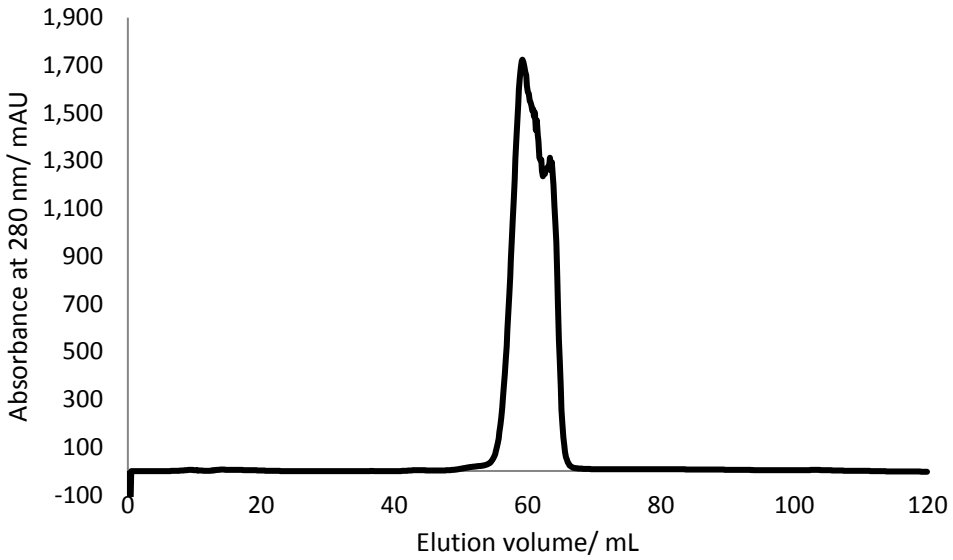


Figure 3.6: Chromatogram of the purification of SkR-IRED by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

SkR-IRED was eluted in a large, shouldered peak with 55 – 65 mL buffer, indicating that the protein existed in only one oligomeric state. The noise visible at the highest points in the peak may be due to the absorbance value exceeding the detector limit, and the shoulder at a higher elution volume may be due to the presence of impurities in the sample. Using a calibration curve provided by GE Healthcare, the elution volume was used to estimate a protein molecular weight of approximately 60 kDa. The molecular weight of a monomer of 6His-tagged *SkR*-IRED is approximately 33 kDa, indicating that *SkR*-IRED was most likely to be present as a dimer in solution. Protein-containing fractions were analysed by SDS-PAGE (Figure 3.7). The protein was of high homogeneity, with very minor quantities of *E. coli* host proteins present relative to the quantity of *SkR*-IRED.

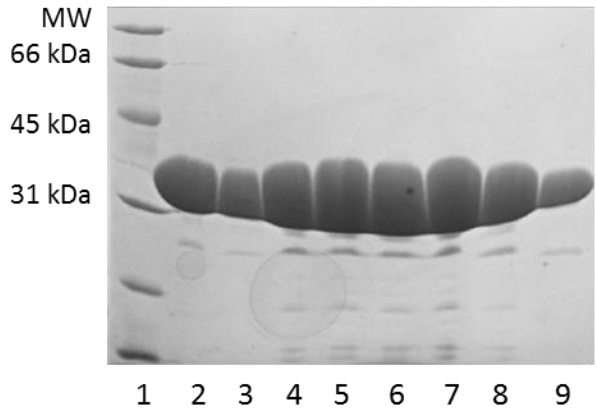


Figure 3.7: SDS-PAGE gel analysis of purification of *SkR*-IRED by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified protein eluted between 55 and 65 mL buffer. Purified *SkR*-IRED is observed to be approximately 32 kDa by SDS-PAGE.

3.4.2 Kinetic analysis and pH behaviour of *SkR*-IRED

A kinetic assay was carried out with increasing concentrations of substrate **1a** to determine the k_{cat} and K_M of the enzyme in the imine reduction reaction as described in section 2.5.3.

Visual inspection of the data indicated that the system exhibits significant levels of substrate inhibition with reaction velocities beginning to decrease at substrate concentrations greater than approximately 7.5 mM.

In an extra sum of squares F test, the substrate inhibition model was convincingly favoured over the Michaelis-Menten model (F = 19.82, P = 0.0005). Curve fitting was therefore carried out using non-linear regression analysis with the substrate inhibition model (Figure 3.8).

When non-linear regression was carried out using the substrate inhibition model without constraints on parameter values, an unreasonably high value of K_M was calculated (47.7 mM). Therefore, K_M was artificially constrained to be equal to the literature value of 8.21 mM in the calculation of best-fit values for k_{cat} and K_i.[60] The best fit values for kinetic parameters calculated during non-linear regression are outlined in Table 3.1.

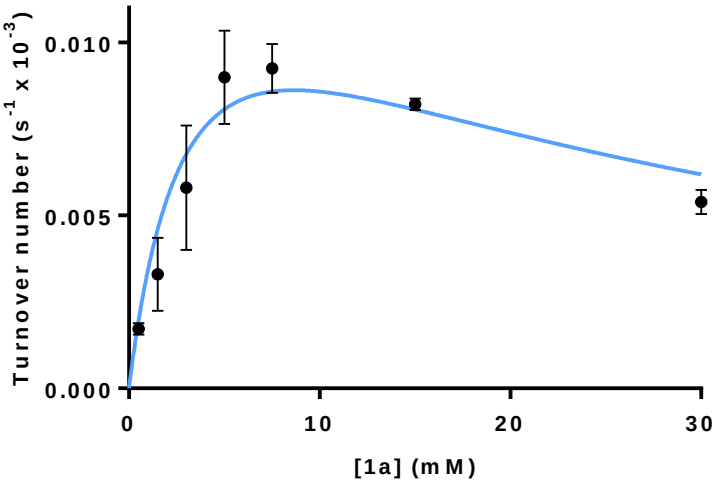


Figure 3.8: Michaelis-Menten plot for the reduction of **1a** by *SkR*-IRED.

Table 3.1: Kinetic parameters determined during non-linear regression analysis of experimental data of the reduction of **1a** by *SkR*-IRED at varying substrate concentrations using the substrate inhibition model

	Value	**Standard error**
k_{cat}	0.0264 s^{-1}	±0.0018 s^{-1}
K_M	8.21 mM	N/A*
K_i	9.05 mM	1.74 mM

*No standard error may be calculated for K_M as this value has been artificially constrained, not calculated

The relatively small standard errors for the values of k_{cat} and K_i indictate that the calculated values are likely to be reliable and that 8.21 mM was a reasonable estimate of K_M. The k_{cat} value of 0.0264 s^{-1} is low, indicating poor catalytic efficiency of *SkR*-IRED towards **1a**. The high K_M value also suggests that *SkR*-IRED exhibits low affinity towards substrate **1a**. These findings suggest that **1a** is not likely to be a natural substrate for *SkR*-IRED.

3.4.3 Biotransformations using *SkR*-IRED

A range of imines (Figure 3.9) were investigated as substrates for *SkR*-IRED in the presence of NADPH. The reactions were analysed using HPLC. For substrate **1a**, the reaction was carried out on a larger scale with the use of a glucose-6-phosphate dehydrogenase-based NADPH recycling system.

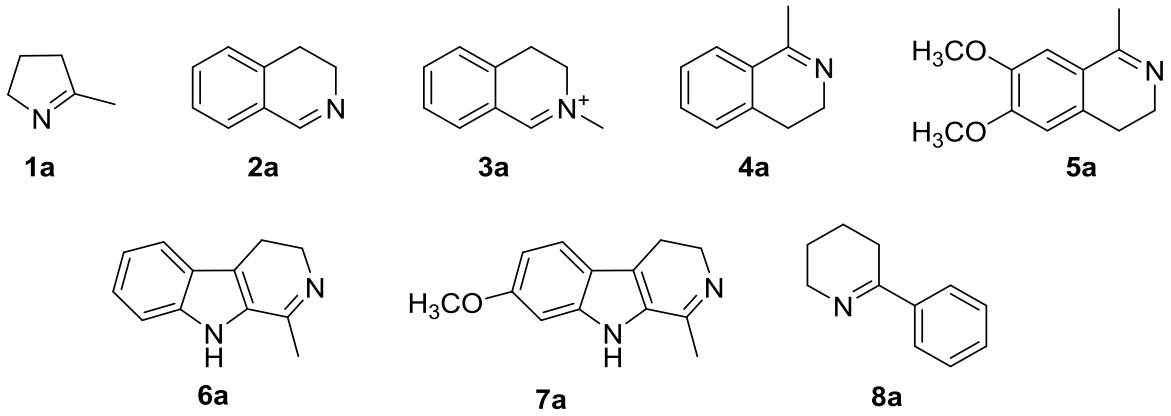


Figure 3.9: The range of imines tested as substrates for *SkR*-IRED

The outcome of the biotransformation experiments is outlined in Table 3.1. Reactions for which conversion is observed are outlined in Figure 3.10.

Table 3.2: Summary of data obtained from HPLC analysis of biotransformations of *SkR*-IRED towards imines **1a** – **8a**.

Substrate	Conversion %	*ee* %	Amine Configuration
1a	22.6	100	*R*
3a	15.7	25	*R*
4a	5.0	21	*R*
8a	Not measurable*	100	*S*

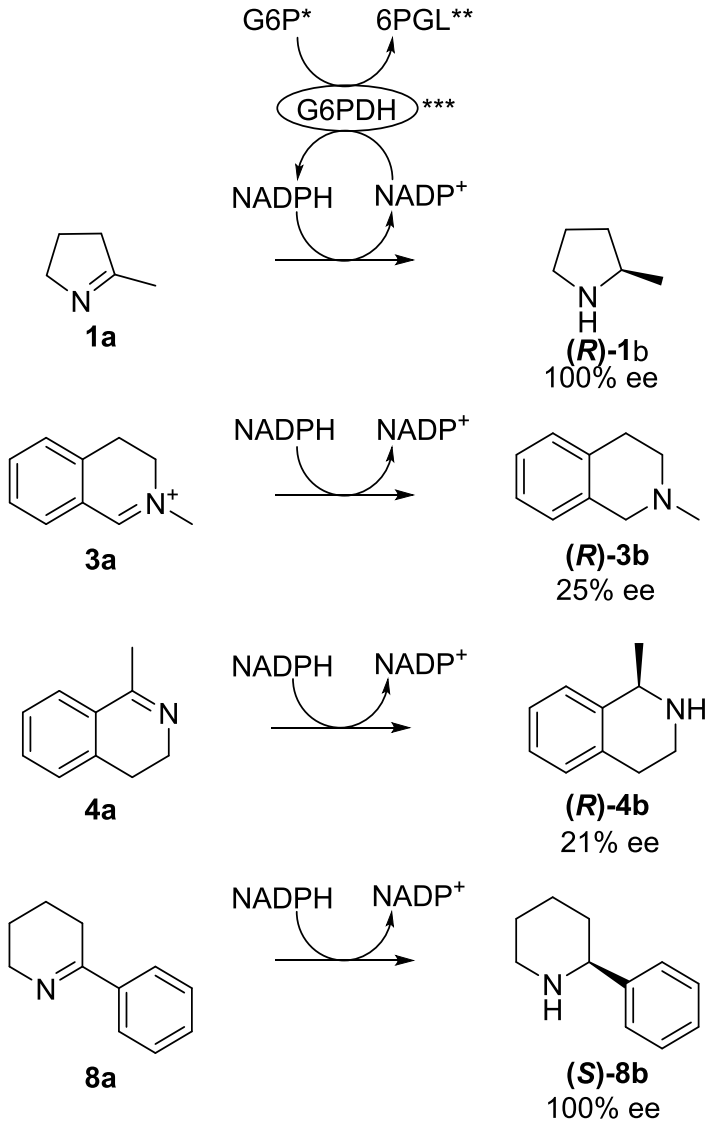


Figure 3.10: A summary of substrates accepted by *SkR*-IRED in biotransformation experiments. High ee values are observed for the conversion of 1a and 8a. For the biotransformation of 1a, an NADPH recycling system was used where *G6P is glucose-6-phosphate, **6PGL is 6-phosphoglucanolactone, and ***G6PDH is glucose-6-phosphate dehydrogenase.

IRED activity of *SkR*-IRED was found to be confined mostly to substrate **1a** and **8a,** suggesting a possible preference of *SkR*-IRED for substituted 5- and 6-membered cyclic imines. *SkR*-IRED displayed either modest or limited activity towards fused ring systems **2a** – **5a**, and no activity towards β-carboline imines **6a** and **7a**.

For substrates **1a** and **8a**, *SkR*-IRED displayed a strong preference for the formation of the corresponding (*R*)- and (*S*)-amine respectively. This apparent reversal of stereoselectivity for substrate **8a** is an artefact of Cahn-Ingold-Prelog nomenclature rules as opposed to any fundamental changes in enantioselectivity.

3.4.5 Mechanism proposal for *SkR*-IRED

Structurally related enzymes

Structurally similar enzymes were sought in order to provide insight into the previously uncharacterised mechanism of *SkR*-IRED. A DALI structural alignment showed that *SkR*-IRED shares the most similarity with members of the gamma-hydroxybutyrate dehydrogenase (GHBDH) and β-hydroxyisobutyrate dehydrogenase (HIBDH) families. HIBDH from *Thermus thermophilus* HB8 was used as an illustrative example.[67]

2CVZ and *SkR*-IRED were structurally aligned (Figure 3.11). Their superimposition highlights the structural similarity of the two enzymes, where the structure of the *N*-terminal helical domain is almost totally conserved. While the *C*-terminal helical domain also shares a high degree of structural similarity, their orientations are flipped with respect to one another.

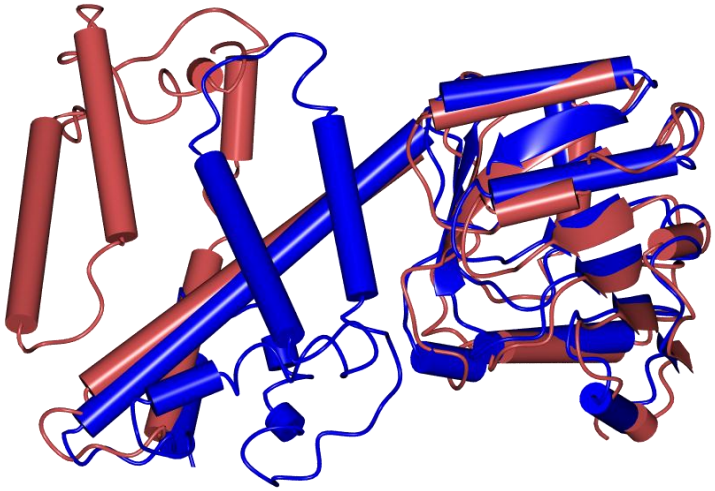


Figure 3.11: The superimposition of the structures of *SkR*-IRED (red) and HIBDH (blue). The overlay demonstrates very high structural similarity between the two enzymes, where there is almost complete conservation between the two N-terminal Rossman fold subdomains (right). There is also significant conservation between the two C-terminal helical domains. However, in HIBDH a sharp β-turn occurs at residue Ser204, thereby turning C-terminal domain back towards the N-terminal domain. Such a feature is not observed in *SkR*-IRED, causing the N-terminal domain to protrude from the C-terminal domain.[67]

The differences in the orientation of the *C*-terminal helical domain of *SkR*-IRED and 2CVZ explain why the unusual reciprocal domain sharing phenomenon is possible in *SkR*-IRED, but not in structurally related enzymes such as HIBDH.

Mechanistic information from 2CVZ

In other NADPH-dependent oxidoreductases such as HIBDH, a protic amino acid residue in the vicinity of the nicotinamide ring of the NADPH cofactor and the region occupied by the substrate is required in order to provide a proton to the hydroxyl group of the substrate in the reductive direction. In HIBDH, this role is fulfilled by Lys165, which donates a proton to β-hydroxyisobutyrate to form the corresponding terminal alcohol as shown in Figure 3.12.[67]

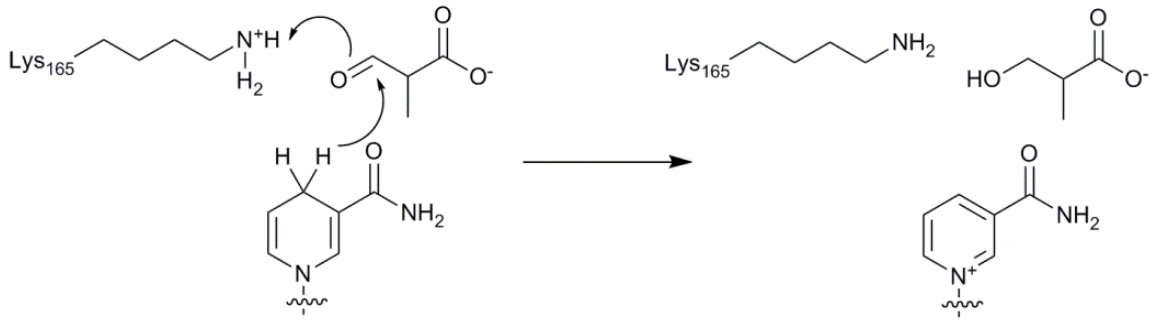


Figure 3.12: The mechanism for the reduction of 2-methyl-3-oxopropoanoate by β-hydroxyisobutyrate dehydrogenase from *Thermus thermophilus* (HIBDH). Residue Lys165, which is proximal to both the NADPH cofactor and β-hydroxyisobutyrate substrate behaves as a proton donor to facilitate the formation of the corresponding alcohol.[67]

The superimposition of the active sites of HIBDH and *SkR*-IRED showed that the catalytically active Lys165 residue of 2CVZ is substituted by residue Asp178 in the equivalent location within the active site of *SkR*-IRED (Figure 3.13).

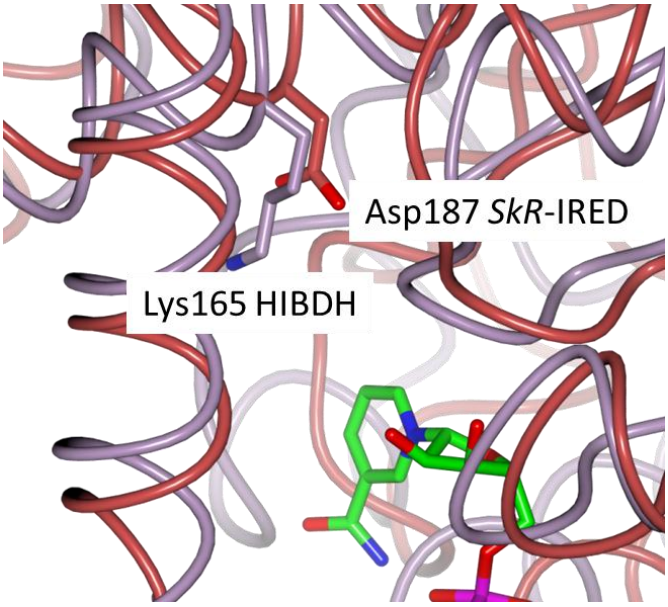


Figure 3.13: The superimposition of the active sites of 2CVZ (purple) and *SkR*-IRED (red) shows that residue Asp187 occurs in the equivalent position in *SkR*-IRED to the catalytically active Lys165 residue of 2CVZ.

Based on the similarities of the overall structures and active site topology of *SkR*-IRED and 2CVZ, it was proposed that the mechanism of *SkR*-IRED could proceed in a similar manner, where residue Asp187 of *SkR*-IRED could be catalytically active in the reduction of imines such as **1a**, as opposed to residue Lys165 of HIBDH in the reduction of β-hydroxyisobutyrate (Figure 3.14). The proposed mechanism is outlined in Figure 3.12.

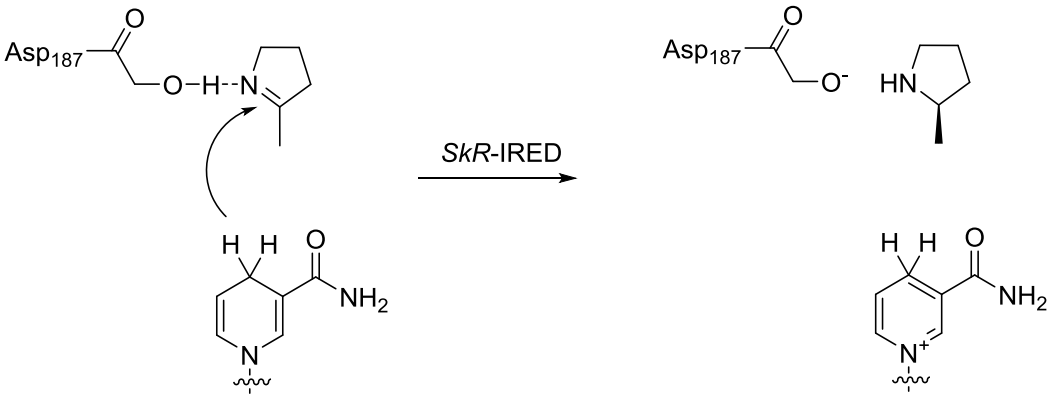


Figure 3.14: The mechanism proposed for the reduction of imines (such as **1a**) by *SkR*-IRED. A proton is donated to **1a** by catalytic residue Asp187, resulting in the formation of an activated iminium ion. Hydride is then delivered to the iminium ion by NADPH to form the corresponding (R)-amine.

3.4.6 Investigations towards the proposed mechanism of *SkR*-IRED

The *SkR*-IRED mutant Asp187Ala was generated for use in imine reduction assays to assess whether the Asp187 residue is necessary for catalysis. The Ala sidechain is aprotic, and incapable of carrying out the proposed protonation for the reduction of imines as shown in the mechanism in figure 3.14. Therefore, if the resultant mutant does not possess imine reduction activity, this would support the suggestion that residue Asp187 is required for catalysis.

Site-directed mutagenesis

The wild-type gene was used as a template for the generation of mutant Asp187Ala by site-directed mutagenesis (SDM). The SDM reaction was digested with *Dpn*I to remove template DNA and used to transform *E. coli* cells. DNA was isolated from a resultant colony using a miniprep kit, and submitted for sequencing at GATC Biotech. The gene in the isolated DNA was found to have 100 % sequence identity to the target Asp187Ala mutated gene.

Expression and purification of SkR-IRED mutant

SkR-IRED was expressed in the pET-YSBLIC3C vector in *E. coli* BL21 (DE3) cells in 2 L LB-Kan medium at the optimal expression temperature for the wild-type protein of 16 °C. Protein was initially purified from lysed, resuspended cells by nickel affinity chromatography.

SkR-IRED Asp187Ala was subsequently purified by size exclusion chromatography (Figure 3.15) using a Superdex® HiLoad 75 gel filtration column. Protein-containing fractions were analysed by SDS-PAGE (Figure 3.16).

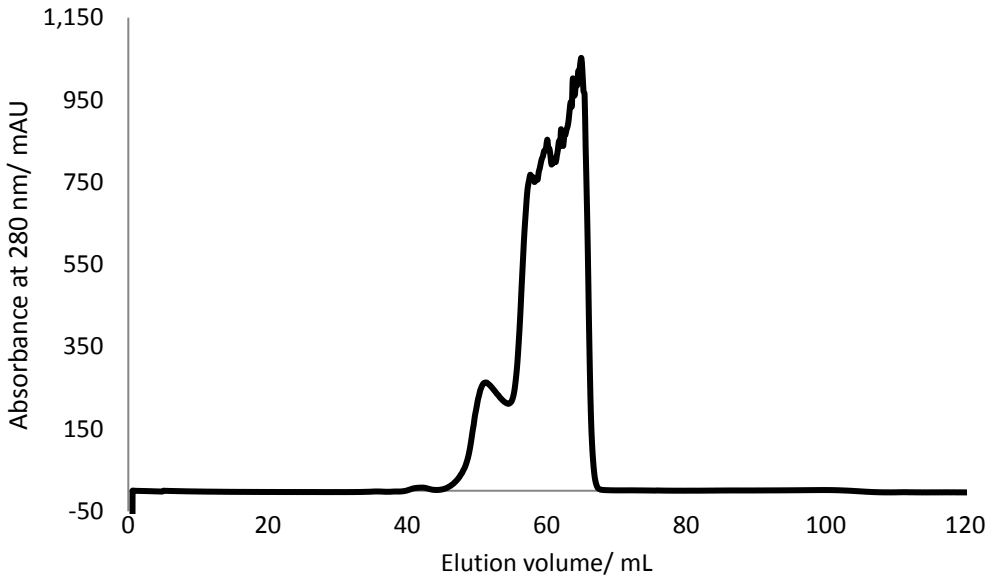


Figure 3.15: Chromatogram of the purification of *SkR*-IRED Asp187Ala by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

The chromatogram from SEC purification contained a cluster of peaks with an elution volume of 55 – 67 mL. There is a shoulder main peak at 50 – 55 mL elution volume which may be due to impurities. Using a calibration curve provided by GE Healthcare, the elution volume could be used to estimate a protein molecular weight of approximately 60 kDa that the dimeric structure of the wild-type protein had been preserved.

Protein-containing fractions from both the main peak and shouldering peak were analysed by SDS-PAGE (figure 3.14). All fractions appeared to contain pure *SkR*-IRED. Proteins eluted in the shoulder peak may have corresponded to heavier oligomers of *SkR*-IRED Asp187Ala and were therefore discarded. Only minor quantities of *E. coli* host proteins were visible relative to the quantity of *SkR*-IRED.

.

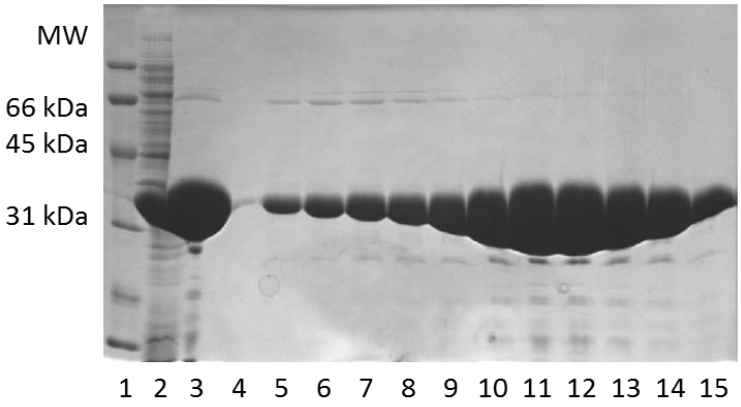


Figure 3.16: SDS-PAGE gel analysis of purification of *SkR*-IRED Asp187Ala by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: crude cell lysate, lane 3: sample of protein pooled following nickel-affinity chromatograpy, lanes 3-15: fractions containing purified protein eluted between 49 and 66 mL buffer. Purified *SkR*-IRED Asp187Ala is observed to be approximately 32 kDa by SDS-PAGE.

***Kinetic assays using SkR-IRED* mutant**

A spectrophotometric-based NADPH-depletion assay was used to assess the kinetic parameters of *SkR*-IRED Asp187Ala towards the standard substrate **1a** under the same conditions used for the wild-type protein. No NADPH-depletion was observed for **1a** concentrations between 1 and 30 mM. The results suggested that residue Asp187 is necessary for catalysis.

Asp187 as a catalytic residue

The generation of mutant Asp187Ala resulted in an enzyme that was devoid of activity according to a spectrophotometric-based NADPH-depletion assay, which supports that this residue may be necessary for catalysis. However, the bond distance between the Asp187 residue and the C4 atom of the nicotinamide ring of the NADPH cofactor is 7.86 Å (Figure 3.17). In HIBDH, the distance between the catalytic Lys165 residue and the C4 of NADPH is 4.4 Å, which facilitates Lys165 to act as a direct proton donor[67]. This suggests that Asp187 may protonate **1a** directly, unless a large domain shift occurs upon binding of the imine substrate. It may be the case that mutation of Asp187 renders the mutant inactive due to a change in the physical properties of the enzyme. Alternatively, it is possible that Asp187 may protonate **1a** via an activated

water molecule. Further investigation would be required in order to ascertain the precise role of Asp187 in catalysis.

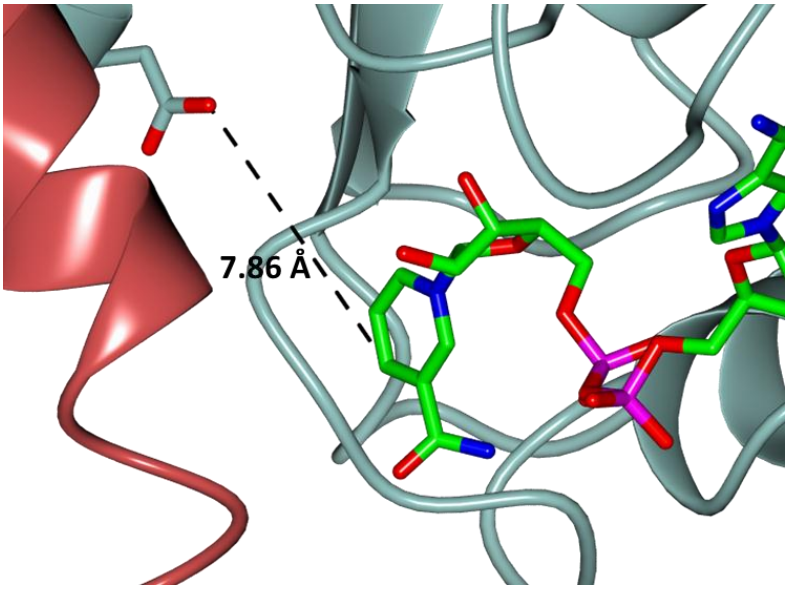


Figure 3.17: The active site of *SkR*-IRED. The dashed line illustrates the bond distance of 7.9 Å between the C4 atom of NADPH (from which hydride is delivered to the substrate) and the potentially catalytic residue Asp187, further than the 4.4 Å Lys165-NADPH C4 distance observed in the homologue HIBDH, where direct protonation from the protic Lys165 to the substrate occurs.

3.4.7 Investigation of physiological role of *SkR*-IRED

The limited catalytic efficiency of *SkR*-IRED towards **1a** described in the kinetic analysis of this reaction in section 3.4.2 indicates that this is not likely to be a natural substrate of *SkR*-IRED. The physiological function of *SkR*-IRED is therefore currently unknown. If a natural substrate is determined, this will provide information on the type of substrate that is most readily accepted by *SkR*-IRED, and provide insight towards substrates which may be efficiently converted by the enzyme.

In the structural solution for *SkR*-IRED, an unassigned region of electron density is observed in the active site (Figure 3.18). It was proposed that this may provide insight into the physiological role of *SkR*-IRED, as the density may belong to a natural substrate.

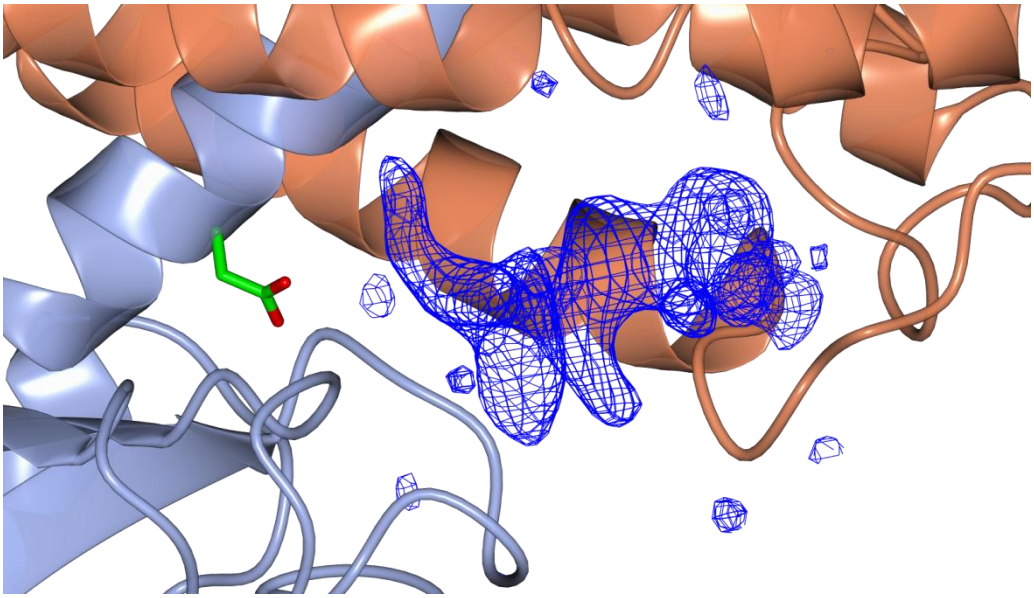


Figure 3.18: A region of unassigned electron density in the active site of *SkR*-IRED in the vicinity of the putative catalytic residue Asp187.

Screening for activity towards dihydrofolate and methotrexate

It was suggested that the unassigned electron density in the active site of *SkR*-IRED could be attributed to a folate derivative such as dihydrofolate (Figure 3.19), as this is a known substrate for the naturally occurring imine reductase dihydrofolate reductase (DHFR).

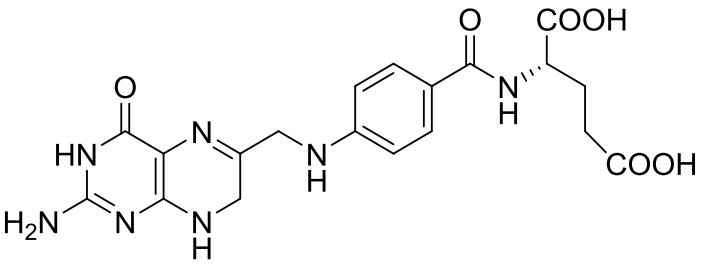


Figure 3.19: The chemical structure of dihydrofolate, a known substrate for the imine reductase DHFR.

A UV-Vis spectrophotometric-based NADPH depletion assay was used to assess whether dihydrofolate behaved as a substrate for *SkR*-IRED. Additionally, it was assessed whether the presence of DHFR inhibited the activity of *SkR*-IRED towards a known substrate **1a**, as this would also indicate active site binding. However, no NADPH depletion was observed for *SkR*-IRED in the presence of NADPH and dihydrofolate.

The dihydrofolate mimic methotrexate (Figure 3.20) was also investigated for both substrate and inhibitor behaviour towards *SkR*-IRED, as this is a known inhibitor of the imine reductase DHFR. However, no substrate activity was observed, and the presence of methotrexate had no effect on the activity of *SkR*-IRED towards **1a**.

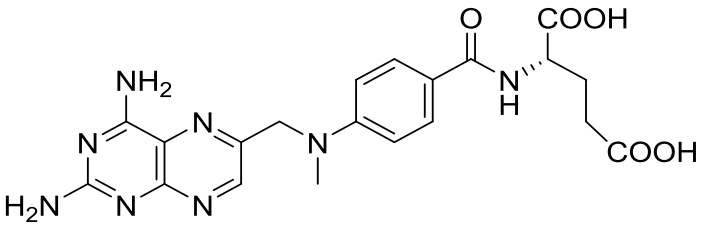


Figure 3.20: The chemical structure of methotrexate, a known inhibitor for the imine reductase DHFR.

Mass spectrometry investigations

In an effort to identify the unassigned electron density in the active site of *SkR*-IRED, mass spectrometry analysis was carried out on purified protein (Figure 3.21).

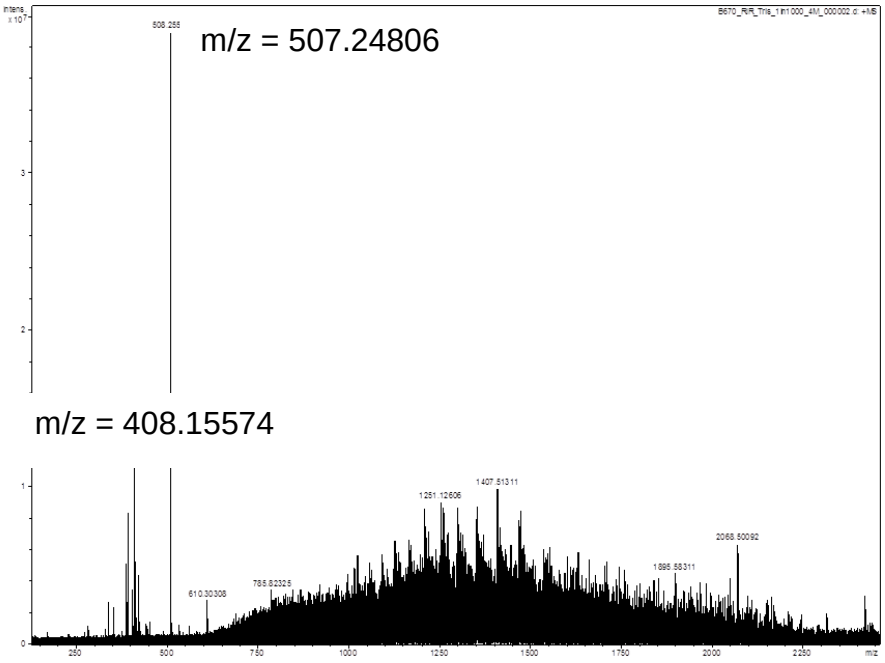


Figure 3.21: Mass spectrum of purified *SkR*-IRED using electrospray ionisation. Two intense peaks are observed in a feasible mass region for potential substrates which may explain the unassigned density in the enzyme active site. The accurate masses of the peaks suggest a molecular formula of $C_{27}H_{33}N_5O_5$ and $C_{22}H_{21}N_3O_5$ for m/z = 507.24806 and 408.15574 respectively.

The two most intense peaks in the lower mass region of the mass spectrum correspond to molecular formulae of $C_{27}H_{33}N_5O_5$ and $C_{22}H_{21}N_3O_5$ according to their accurate masses. Both ions were resistant to further fragmentation with increasing amounts of collision energy, which would not be expected with an imine, as fragmentation should occur readily across the carbon-nitrogen double bond. No matches were found for the *E. coli* metabolome database (ECMB) for the masses of formulae corresponding to the two intense peaks.

This prevented any chemical structures being proposed from the molecular formulae.

Screening for reductive amination activity

Opine dehydrogenases (OpDHs) are NAD(P)H-dependent enzymes that catalyse the reductive condensation of L-amino acids with α-keto acids. The primary function of OpDHs is to oxidise glycolytically generated NAD(P)H to form $NADP^+$, which sustains the glycolytic cycle which generates ATP under anaerobic conditions during short, intense periods of muscular activity in invertebrates.[68] In such reactions, an imine is

formed in-situ by the condesnation of L-amino acid and α-keto acid, which is subsequently reduced by OpDH and NADPH (Figure 3.22).

It was proposed that the density observed in the active site could represent a conjugated L-amino acid and α-keto acid, which would indicate that the natural function of *SkR*-IRED was that of an OpDH.

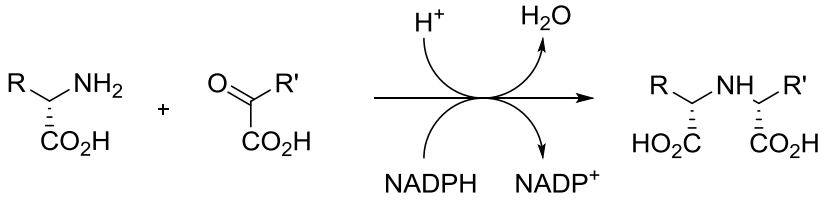


Figure 3.22: The condensation and reduction of an L-amino acid and α-keto acid by an opine dehydrogenase (OpDH) and NADPH to form a product containing a newly formed amine moiety.

In order to investigate the presence of OpDH activity in *SkR*-IRED, activities towards several combinations of L-amino acids and α-keto acids outlined in Table 3.3 were measured. Amino acids were selected to include a broad range of side chain properties, including; positively charged (arginine and histidine), hydrophobic (leucine), aromatic (phenylalanine), negatively charged (glutamic acid), and compact (alanine) acids.

One mL reactions containing *SkR*-IRED (0.44 mg mL^{-1}), NADPH (0.2 mM), L-Amino acid (10 mM) and α-keto acid (10 mM) in Tris.HCl buffer (pH 7.5, 50 mM) were measured by detecting NADPH depletion in UV-vis spectroscopy.

Table 3.3: Combinations of α-keto acids and L-amino acids used in opine dehydrogenase activity tests with *SkR*-IRED

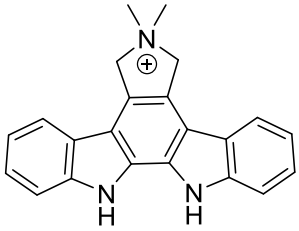
		α-Keto acid		
		Succinic acid	Pyruvic acid	2-Oxoglutaric acid
L-Amino acid	Arginine	No activity	No activity	No activity
	Isoleucine	No activity	No activity	No activity
	Phenylalanine	No activity	No activity	No activity
	Glutamic acid	No activity	No activity	No activity
	Histidine	No activity	No activity	No activity
	Alanine	No activity	No activity	No activity

No NADPH depletion was detected in the assay for any combinations of α-keto acid and L-amino acid, indicating that *SkR*-IRED does not display opine dehydrogenase activity for the substrates tested. Therefore, this is unlikely to be the physiological role of *SkR*-IRED.

Possible role of SkR-IRED in typtophan dimer formation

Tryptophan dimers (TDs) are formed by the oxidation and subsequent dimerization of tryptophan, and are precursors to a diverse collection of bacterial natural products. TDs are frequently associated with biological activity, such as antibacterial and antifungal properties, which suggests that these may be a useful scaffold for bioactive natural products.[69, 70]

Brady and co-workers reported the characterization of the pathway towards the TD reductasporine (Figure 3.23), which acts as a biosynthetic precursor for a natural product possessing antifungal bioactivity. This gene cluster was identified during metagenome mining of environmental DNA (eDNA) libraries.[70]



reductasporine

Figure 3.23: The chemical structure for reductasporine; a tryptophan dimer (TD) with antifungal properties.[70]

The pathway involves an imine reduction step, performed by an IRED which is termed RedE (Figure 3.24).[70]

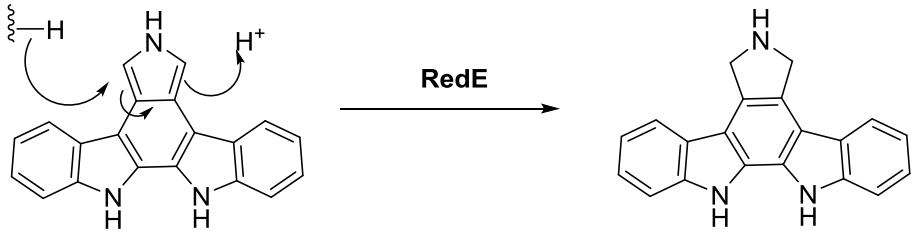


Figure 3.24: An imine reduction step performed by IRED RedE in the biosynthetic pathway of reductasporine, a tryptophan dimer (TD) with antifungal properties.[70]

The closest relative of RedE was found to be *SkR*-IRED. The role of RedE in the reduction of the pyrrole ring in the synthesis of reductasporine is consistent with the ability of *SkR*-IRED to reduce the similar pyrroline **1a**. It is possible that the natural function of *SkR*-IRED is in TD formation in a biosynthetic pathway similar to that which is catalysed by RedE.

.

3.5 Conclusion

During this study, *SkR*-IRED has been recombinantly expressed and purified. Purified protein has been used in assays to assess the substrate scope and kinetic parameters for the enzyme. Biotransformations with a number of imines demonstrated that *SkR*-IRED is able to convert substituted 5-membered pyrrolines and 6-membered piperideines to the corresponding (*R*)-amines (or (*S*)-amines in instances where nomenclature is reversed according to Cahn-Ingold-Prelog rules).

For the conversion of **1a** to the corresponding amine **1b**, the K_M was 8.21 mM and the k_{cat} 0.0264 $s^{-1,}$ indicating a relatively low catalytic efficiency that suggests that **1a** is not a natural substrate for the enzyme. *SkR*-IRED converted imines **1a** and **8a** to the corresponding (*R*)- and (*S*)-amine with 100 % ee.[60]

Structural observations indicated that Asp187 may provide a role in catalysis, which was reinforced by the fact that the mutant Asp187Ala lost activity towards **1a**. However, bond distances between the Asp187 hydroxyl and the C4 carbon of NADPH, from which hydride is delivered, were too large to suggest direct protonation from Asp187Ala. Further investigation would be required to determine the exact role of Asp187 in catalysis.[60]

Despite investigations into the physiological role of *SkR*-IRED by screening for activity towards or inhibitory effects from dihydrofolate and methotrexate, screening for opine dehydrogenase activity and investigations of active site density using mass spectrometry, no native substrates were identified for *SkR*-IRED from experimental work. However, research work was published subsequently by Brady and co-workers revealed a close homologue to *SkR*-IRED which plays a role in the formation of reductasporine, a tryptophan dimer (TD) with antifungal properties, which might suggest that *SkR*-IRED plays a similar biosynthetic role.[70]

4. Isolation, application, structure and mechanism of *SS*-IRED, an (*S*)-selective NADPH-dependent imine reductase from *Streptomyces* sp. GF3546

4.1 Introduction

The characterization of an (*R*)-selective imine reductase (IRED) has been described in section 3. At the start of this project, there were no known 3D-structures of an (*S*)-selective IRED. This will outline a a collaborative effort with colleagues at the Manchester Institute of Biotechnology to thoroughly characterise an (*S*)-selective imine reductase. As described in section 3, a broad ranging screen of organisms for activity towards imine **1a** lead to the discovery of two organisms that catalysed the reduction of this substrate to the corresponding amine **1b**; *Streptomyces sp.* GF3587 and *Streptomyces sp.* GF3546, which catalysed the formation of (*R*)-**1a** and (*S*)-**1b** with 99.2 % and 92.3 % ee respectively.[54] The two enzymes associated with the reactions were termed *SR*-IRED and *SS*-IRED. As at the start of this project, *SS*-IRED was the only imine reductase with confirmed (*S*)-selective activity towards substrate **1a**, this enzyme will be used as model for the study for (*S*)-selective imine reductases.

4.2 Aims

The primary aim of this project will be to characterise the structure of *SS*-IRED using X-ray crystallography, which may be used to inform the mechanism of (*S*)-selective imine reductases, rationalise the differences in selectivity to (*R*)-selective imine reductases and serve as a basis for rational engineering towards increased substrate scope and improved catalytic properties. This will be carried out as complementary work to collaborators at the Manchester Institute of Biotechnology, who will focus mainly on the chemical applications of the enzyme.

The gene will be subcloned into the house expression vector pET-YSBLIC3C, recombinantly expressed in *E. coli* and purified. Crystallization conditions will then be established, and any crystals formed will be used in X-ray diffraction experiments to determine the enzyme structure.

Enzyme kinetic behaviour will be compared with *SkR*-IRED, the (*R*)-selective IRED described in Chapter 3. The two enzyme structures will be compared to rationalise

these differences as well as differences in enantioselectivity, which will contribute towards an overall understanding of NADPH-dependent IREDs as an enzyme family.

4.3 Results

4.3.1 Amplification of target genes

Primers with LIC extensions were used for the amplification of the gene coding for *SS*-IRED. A synthetic gene with coding for *SS*-IRED with codon optimisation for expression in *E. coli* was used as a template for PCR amplification. The PCR reaction product was visualised by agarose gel electrophoresis (Figure 4.1).

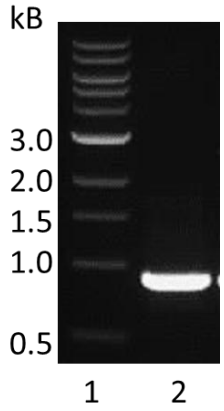


Figure 4.1: Agarose gel visualising the amplification of the gene coding for *SS*-IRED by PCR. Lane 1: NEB 1 kB ladder, lane 2: PCR product of gene coding for *SS*-IRED

A bright band at the expected length of approximately 900 bp is observed in the analysis of the amplification of *SS*-IRED. The band was excised and purified.

4.3.2 Subcloning into pET-YSBLIC3C expression vector

The purified *SS*-IRED gene was treated with T4 DNA polymerase and annealed to linearized T4 DNA polymerase-treated pET-YSBLIC3C vector. *E. coli* was transformed with the annealed plasmid and DNA isolated. A diagnostic restriction digest was carried out on the purified plasmid to assess for the presence of an appropriately sized insert (Figure 4.2).

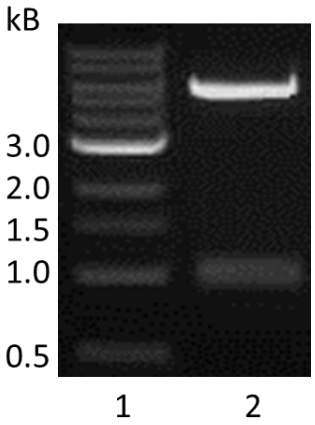


Figure 4.2: Agarose gel visualising restriction digest of gene coding for *SS*-IRED subcloned into the pET-YSBLIC3C expression vector. Lane 1: NEB 1 kB ladder. Lane 2: pET-YSBLIC3C plasmid containing *S*-IRED gene, digested with *Nco*I and *Nde*I. Two bands are observed in lane 2; the pET-YSBLIC3C vector backbone (~5.5 kB) and the *SS*-IRED gene (~900 bp)

The restriction digest showed an appropriately sized insert of approximately 900 bp arising from the gene coding for *SS*-IRED. The agarose gel indicated that the cloning experiment had been successful, and the corresponding plasmid was submitted for sequencing analysis.

DNA sequencing

The DNA sample corresponding to lane 2 in figure 4.1 was submitted for sequencing at GATC Biotech. DNA sequencing confirmed that the gene present in the sample had 100 % sequence identity with the target sequence.

4.3.3 Large scale expression and purification

SS-IRED was expressed in the pET-YSBLIC3C vector in *E. coli* BL21 (DE3) cells in 2 L LB-Kan medium at the previously established optimal expression temperature of 16 °C. Protein was initially purified from lysed, resuspended cells by nickel affinity chromatography, and protein samples were analysed by SDS-PAGE to assess protein purity (Figure 4.3).

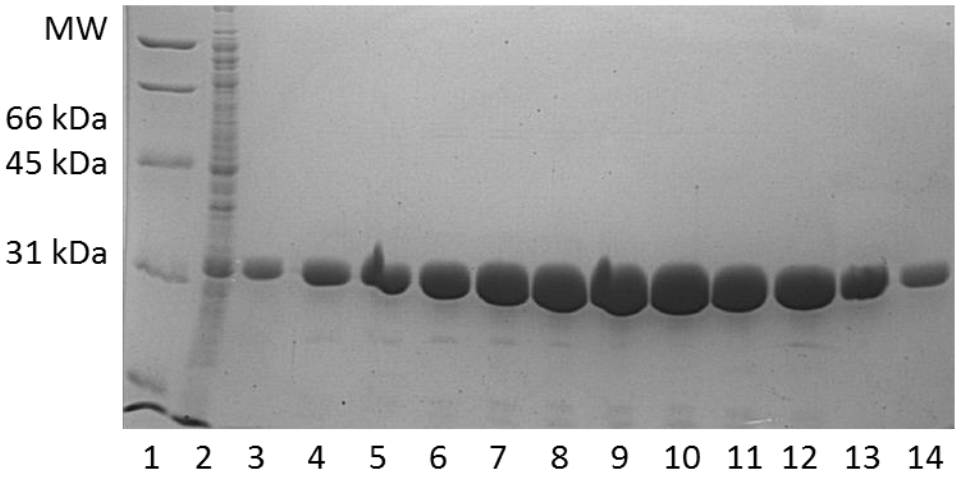


Figure 4.3: SDS-PAGE gel analysis of purification of *SS*-IRED by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: unbound flow-through collected during loading of nickel column, lanes 3 – 14: fractions containing purified protein eluted during increasing imidazole gradient. Purified *SS*-IRED is observed to be approximately 32 kDa by SDS-PAGE.

SS-IRED was subsequently purified by size exclusion chromatography (Figure 4.4) using a Superdex® HiLoad 75 gel filtration column. Fractions which showed absorbance at 280 nm were analysed by SDS-PAGE (Figure 4.5).

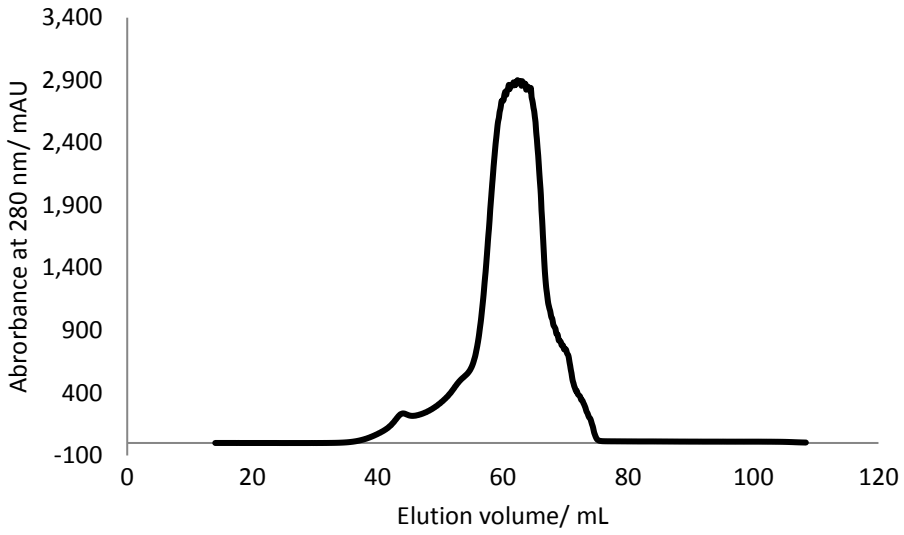


Figure 4.4: Chromatogram of the purification of *SS*-IRED by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

SS-IRED was eluted in a large peak between 40 – 67 mL buffer. The peak was broader due to shoulders at both sides, possibly due to high protein volumes loaded onto the column. The elution volume of the protein corresponded to a molecular weight of approximately 60 kDa in a calibration curve provided by GE healthcare, indicating that *SS*-IRED was present as a dimer in solution, similarly to *SkR*-IRED (Chapter 3). Protein-containing fractions were analysed by SDS-PAGE (Figure 4.5).

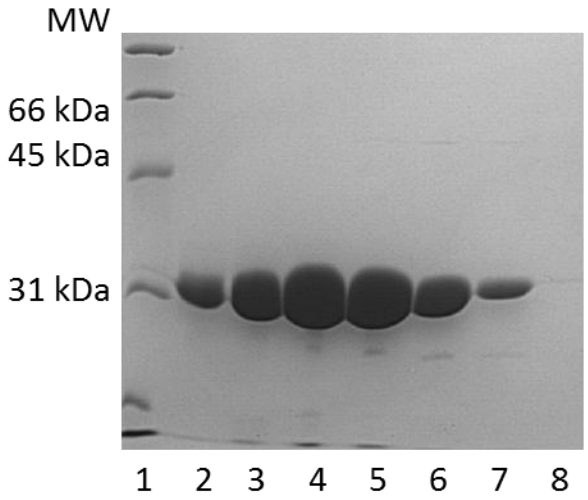


Figure 4.5: SDS-PAGE gel analysis of purification of *SS*-IRED by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified protein. Purified *SS*-IRED is observed to be approximately 32 kDa by SDS-PAGE.

SDS-PAGE analysis indicated that purified *SS*-IRED was highly homogeneous and free of contaminant host proteins from *E. coli*.

4.3.4 Cleavage of 6His tag

For use in crystallization trials, *SS*-IRED with its *N*-terminal 6His tag removed was prepared for use in addition to the *N*-terminally tagged construct, as the presence or absence of this tag can have a strong influence on the crystallization behaviour of a protein. After purification by nickel-affinity chromatography, an aliquot of *SS*-IRED was removed. The imidazole from the protein buffers was removed by dialysis overnight. *SS*-IRED was then incubated with HRV3C protease. Uncleaved *SS*-IRED and HRV3C protease, which are both 6His-tagged, were separated from *SS*-IRED using a second round of nickel-affinity chromatography. Cleaved *SS*-IRED was found in the flowthrough fractions, as its lack of 6His tag prevented binding to the column.

Cleaved *SS*-IRED was then purified by size exclusion chromatography using a Superdex® HiLoad 75 gel filtration column (Figure 4.6). Protein-containing fractions were analysed by SDS-PAGE (Figure 4.7).

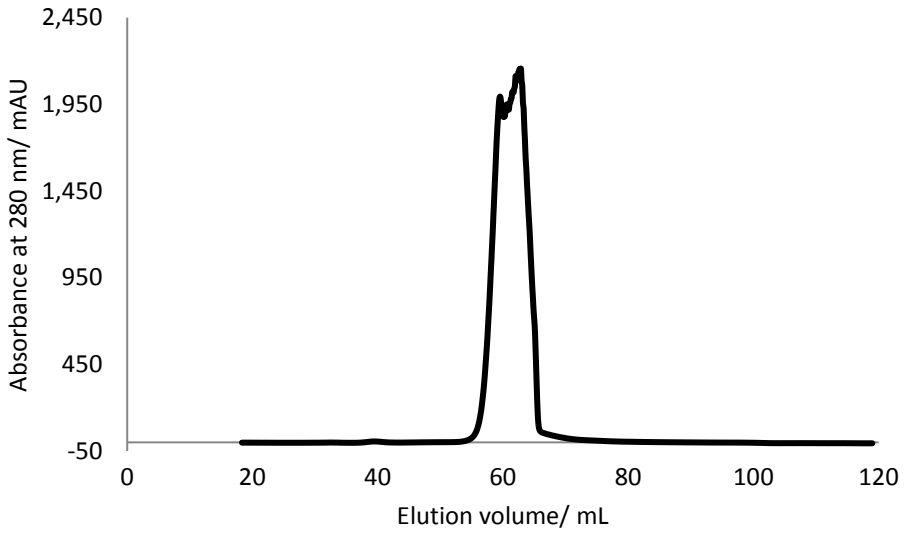


Figure 4.6: Chromatogram of the purification of *SS*-IRED, after cleavage of its 6His tag, by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

SS-IRED was eluted in a narrow peak with 57 – 63 mL buffer. Protein-containing fractions were analysed by SDS-PAGE (Figure 4.7). Distortion is visible at the tip of the peak due to the absorbance value exceeding the capacity of the UV detector. The protein appeared to be homogeneous, and had lost the expected molecular weight of approximately 2 kDa relative to uncleaved *SS*-IRED after cleavage of the 6His tag.

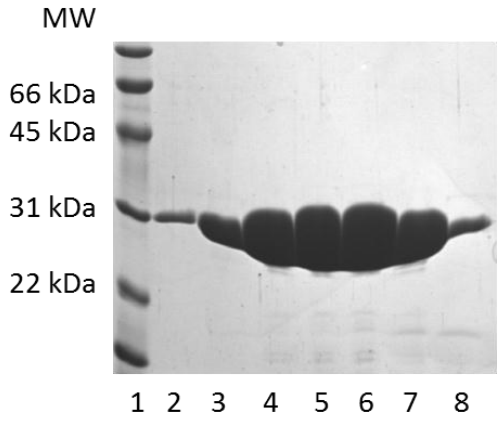


Figure 4.7: SDS-PAGE gel analysis of purification of *SS*-IRED by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified protein. Purified *SS*-IRED is observed to be approximately 30 kDa by SDS-PAGE.

4.4.5 Structure determination

The structure of *SS*-IRED was investigated using X-ray crystallography in order to rationalise differences in substrate specificity and selectivity compared to (*R*)-selective IDEDs such as *SkR*-IRED (Chapter 3) and (*S*)-selective IREDs. The ultimate aim is to provide a sufficient understanding of *SS*-IRED to serve as a basis for the rational engineering of the enzyme

Crystallization of SS-IRED

SS-IRED both with and without its *N*-terminal 6His tag was subjected to a broad range of commercially available crystallization screens in a 96-well sitting drop format; PACT, INDEX, CSS 1 and 2, Hampton 1 and 2, the PEG ion screen and the MPD screen. Screens were dispensed with 300 nL drops consisting of a 1:1 v/v ratio of purified protein and precipitate. Crystals for *SS*-IRED were obtained with 6His-cleaved protein at a concentration of 120 mg mL^{-1} in DL-malic acid (2.1 M, pH 7.0) and sodium citrate tribasic dihydrate (1.4 M) (Figure 4.8). As crystals could not be reproduced at increased scale, data collection was carried out on crystals obtained directly from the initial crystallization screen.

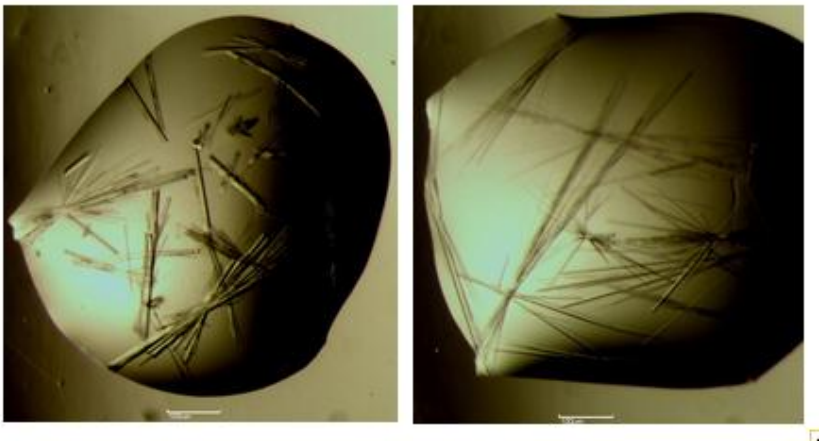


Figure 4.8: Crystals of *SS*-IRED cleaved of *N*-terminal 6His tag obtained with protein at 120 mg ml^{-1} in 2.1 M DL-malic acid pH 7.0 (left) and 0.1 M HEPES pH 7.5, 1.4 M sodium citrate tribasic dehydrate (right).

Data collection

Crystals could not be reproduced in a 24-well hanging drop format. Therefore, the original crystals in the 96-well sitting drop crystallization screen pictured in Figure 4.8 were used for X-ray diffraction experiments at Diamond synchrotron using the I-04 beamline. The crystals obtained in 2.1 M DL-malic acid pH 7.0 diffracted to 3.2 Å resolution.

Structural solution

The structure was solved using a molecular replacement strategy based on the structure of *SkR*-IRED.

The alignment of the protein sequences of *SS*-IRED and *SkR*-IRED (Figure 4.9) shows that the two IREDs share 68.9 % sequence similarity and 38.2 % sequence identity. The GXGXXG motif is conserved between the two enzymes as GLGPMG in *SS*-IRED and GLGLMG in *SkR*-IRED. In place of the 'catalytic' *SkR*-IRED-Asp187 residue in the alignment of both protein sequences, *SS*-IRED contains residue Tyr169.

```
SS-IRED       ----------------MSKQSVTVIGLGPMGQAMVNTFLDNGHEVTVWNRTASKAEALVA 44
SkR-IRED      MPDNPSTKGRMMRNQQAEHTPVTVIGLGLMGQALAGAFLGAGHPTTVWNRTAAKAEPLVA 60
                             .:  ******* ****:..:**. ** .*******:*** ***

SS-IRED       RGAVLAPTVEDALSANELIVLSLTDYDAVYAILEPVTGS-LSGKVIANLSSDTPDKAREA 103
SkR-IRED      RGAKSAGSVAEAVAASPLVVVCVSDYDAVHALLDPLDGTALQGRTLVNLTSGTSAQARER 120
              ***  * :* :*::*. *:*:.::*****:*:*:*: *: *.*:.:.**:*.*  :***

SS-IRED       AKWAAKHGAKHLTGGVQVPPPLIGKPESSTYYSGPKDVFDAHEDTLKVL-TNADYRGEDA 162
SkR-IRED      AAWADGRGADYLDGAILAGPAAIGTADAVVLLSGPRSAFDPHASALGGLGAGTTYLGADH 180
              * **  :**.:* *.: . *  **. :: .  ***:..** * .:*  * :.: * * *

SS-IRED       GLAAMYYQAQMTIFWTTMLSYYQTLALGQANGVSAKELLPYATMMTSMMPHFLELYAQHV 222
SkR-IRED      GLASLYDAAGLVMMWSILNGFLQGAALLGTAGVDATTFAPFITQGIGTVADWLPGYARQI 240
              ***::*  * :.::*: : .: *  **  : **.*. : *: *   . : .:*  **:::

SS-IRED       DSADYPGDVDRLAMGAASVDHVLHTHQDAGVSTVLPAAVAEIFKAGMEKGFAENSFSSLI 282
SkR-IRED      DDGAYPADDAAIDTHLATMEHLIHESEFLGVNAELPRFIKALADRAVADGHGGSGYPALI 300
              *.. **.*   :    *:::*::*  :  **.: **  :  : . .: .*.. ..: :**

SS-IRED       EVLKKPAV-    290
SkR-IRED      EQFRTHSGK    309
              * ::. :
```

Figure 4.9: Protein sequence alignment of *SS*-IRED (top) and *SkR*-IRED (bottom). The GXGXXG NADPH-binding motif which is conserved between the two sequences is highlighted in blue. The 'catalytic' Asp187 residue of *SkR*-IRED aligns with Tyr169 in *SS*-IRED, outlined in red.

The unit cell contained four monomers (A – D) arranged as two pairs of homodimers (A-B and C-D). A monomer of *SS*-IRED, represented by subunit A in Figure 4.10, shares many of its structural features with *SkR*-IRED; an *N*-terminal Rossman fold motif charactistic of NADPH binding proteins (Met1 - Gly163) and a *C*-terminal helical domain (Gly194 – Val290), connected by a long inter-domain helix (Leu164 - Asn193). The data collection and refinement statistics are outlined in Table 4.1.

Table 4.1: Data collection and refinement statistics for *SS*-IRED

Date	19-04-13
	5888
Beamline	i03
Wavelength (Å)	0.97626
Resolution (Å)	73.80-3.19 (3.27-3.19)
Space Group	$P2_12_12_1$
Unit cell (Å)	a = 85.9, b = 87.7; c = 136.7
	$\alpha = \beta = \gamma = 90°$
No. of molecules in the asymmetric unit	4
Unique reflections	17782 (1292)
Completeness (%)	99.9 (99.9)
R_{merge} (%)	0.13 (0.89)
$R_{p.i.m.}$	0.08 (0.56)
Multiplicity	6.5 (6.4)
$\langle I/\sigma(I) \rangle$	12.9 (2.4)
$CC_{1/2}$	1.00 (0.72)
Overall B factor from Wilson plot (Å^2)	58
R_{cryst}/ R_{free} (%)	25.4/32.7
r.m.s.d 1-2 bonds (Å)	0.012
r.m.s.d 1-3 angles (^{o})	1.56
Avge main chain B (Å^2)	61
Avge side chain B (Å^2)	62

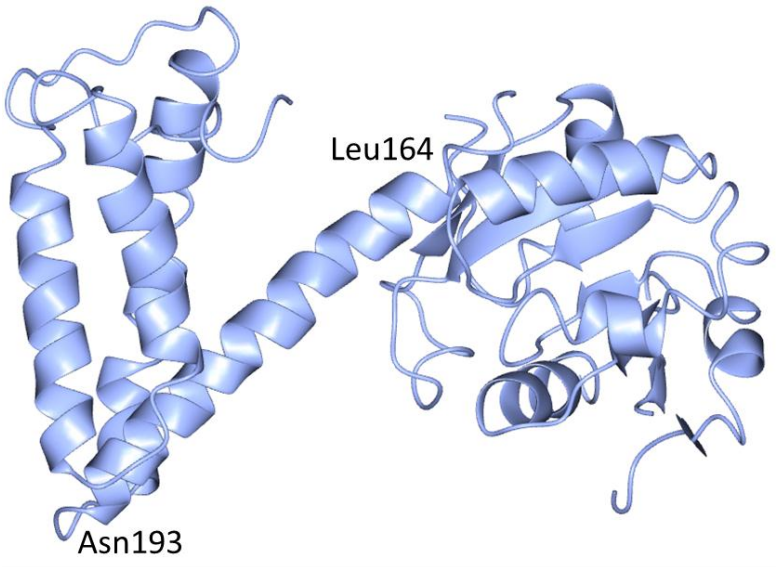


Figure 4.10: Graphical representation of the 2D-structural description of SS-IRED. β-helices are represented by blue arrows and α-helices are represented by red barrels.

The *SS*-IRED monomer consists of twelve alpha helices: α1 (residues Pro13-Asn20), α2 (Val53-Ser58), α3 (Tyr70-Leu77), α4 (Pro97-Lys109), α5 (Pro123-Leu125), α6 (Val141-Ala144), α7 (Glu146-Leu149), α8 (Ala164-Ala192), α9 (Ala197-Ser224), α10 (Leu234-Gln248) α11 (Val256-Glu170) and α12 (Ile282-Leu285) and seven beta sheets: β1 (Val6-Ile9), β2 (Val29-Val31), β3 (Leu62-Val64), β4 (Val87-Asn90), β5 (Lys113-Val119), β6 (Ser132-Ser136) and β7 (Asn154-Asp156). This information is also outlined graphically in Figure 4.11, which emphasises that the *N*-terminal Rossman fold domain includes a mixture of α-helices and β-sheets, while the *C*-terminal domain consists solely of α-helices.

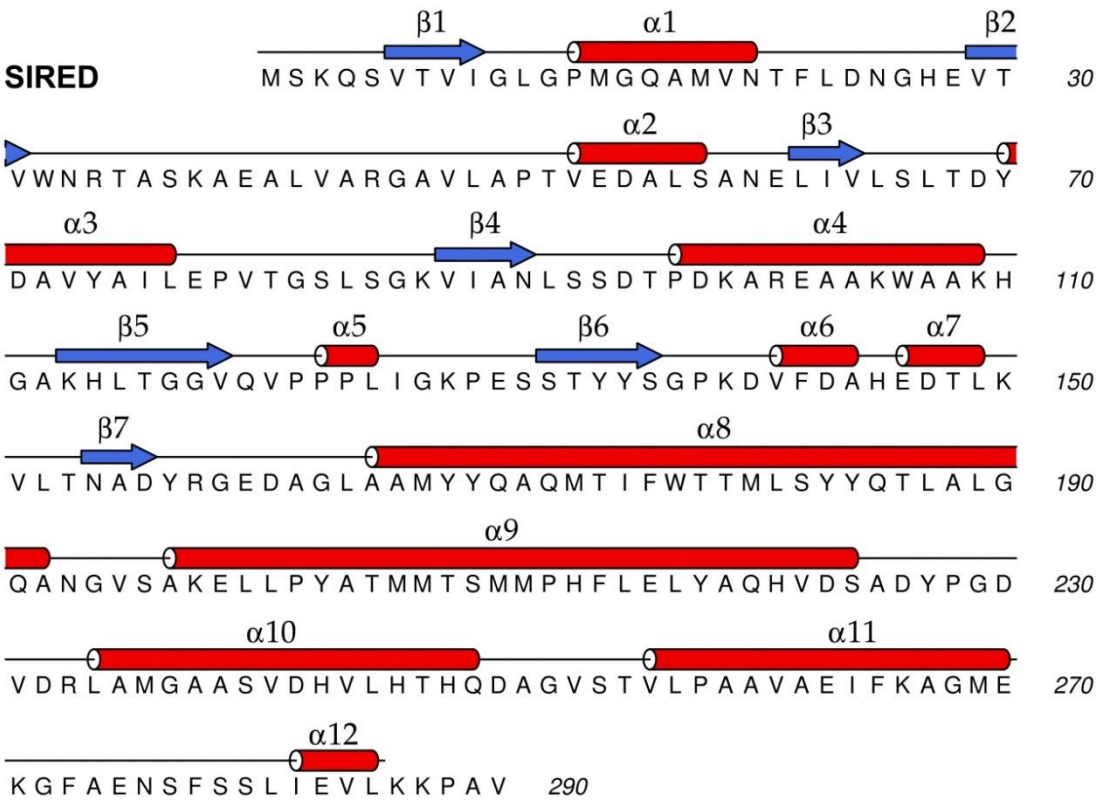


Figure 4.11: Graphical representation of the 2D-structural description of SS-IRED. β-helices are represented by blue arrows and α-helices are represented by red barrels.[71-74]

The active form of the enzyme is a dimer which is represented in Figure 4.12 by subunits A and B. As is the case for *SkR*-IRED, the *SS*-IRED dimer exhibits a domain swapping arrangement where the inter-domain helix of each subunit interlocks with the helical domain of the opposite subunit. Analysis by the EBI PISA tool showed that the surface area of the interface between chains A and B is 3867 Å^2, indicating substantial contact between the two subunits of comparable magnitude to the 3761 Å^2 contact area between monomers of an *SkR*-IRED dimer.[75, 76]

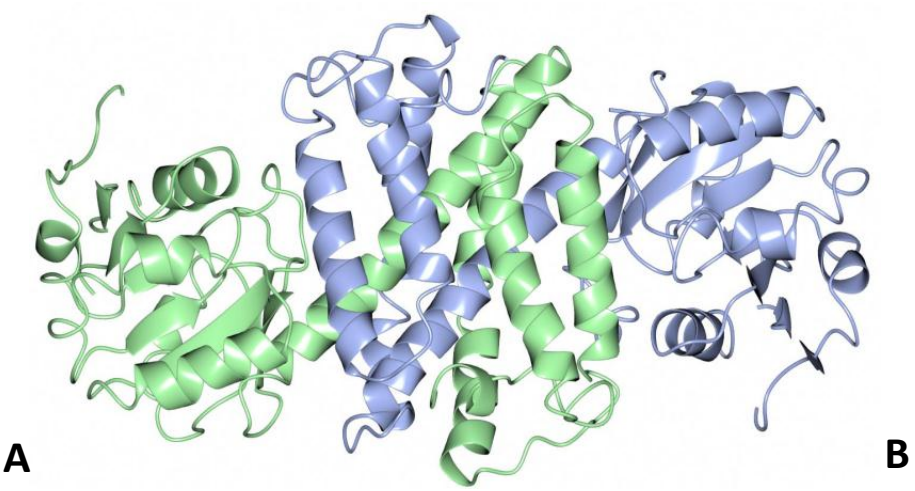


Figure 4.12: The crystal structure of *SS*-IRED in its *apo* form at 3.2 Å resolution. The monomer consists of an *N*-terminal Rossman fold motif and a *C*-terminal helical domain connected by a long interdomain helix (Ala164-Ala192). Dimer formation is facilitated by the reciprocal domain sharing of the two monomeric subunits A (green) and B (blue).

A structural overlay of *SS*-IRED and *SkR*-IRED is pictures in Figure 4.13. The overlay highlights that *SS*-IRED-Tyr169, which aligns with *SkR*-IRED-Asp187 in the sequence alignment in Figure 4.9, also aligns structurally. In chapter 3, it was concluded that Asp187 may have some importance in enzyme mechanism. It is a possibility that Tyr169 may be of catalytic importance in *SS*-IRED in place of Asp187 in *SkR*-IRED. This suggestion will be investigated in more detail in sections 4.4.8 and 4.4.9.

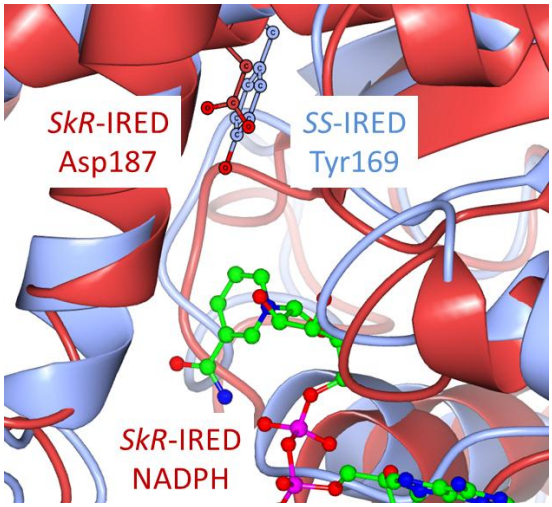


Figure 4.13: Structural overlay of the putative active sites of *SkR*-IRED and *SS*-IRED. The image shows that the possible catalytic residue Asp187 of *SkR*-IRED is replaced by Tyr169 in *SS*-IRED.

A structural similarity search on the structure of *SS*-IRED using the DALI search yielded that, aside from structures of other IREDs that have recently been solved, *SS*-IRED shares the most structural similarity with the protein with PDB accession number 316D, a putative oxidoreductase from *Pseudomonas putida* KT2440. Interestingly, 316D shares the unusual domain-swapping arrangements exhibited in IREDs, but as yet its function is unknown.

Knowledge of the *apo* structure of *SS*-IRED has confirmed that its 3D structure is very similar to that of *SkR*-IRED. It would now be useful to obtain a complex of the enzyme with NADPH and its imine substrates, as this would assist in indicating mechanistically important amino acids, as well as the configuration of the imine in the active site to help rationalise the selectivity of the enzyme.

However, attempts to reproduce crystals of *SS*-IRED with the wild-type protein in the original crystallization conditions were unsuccessful. Further techniques, including macro-seeding of lower quality crystals of *SS*-IRED, manipulation of protein and precipitant volume ratios, total drop volumes, crystal growth temperatures and the use of larger scale hanging drops also failed to produce more *SS*-IRED crystals. Therefore, no complex of *SS*-IRED with either NADPH or imine substrates was obtained.

In an effort to remedy this, modified constructs of *SS*-IRED were therefore designed in an effort to obtain complexed crystals.

4.4.6 *SS*-IRED+, a modified protein for the crystallization of *SS*-IRED

Gene design

SkR-IRED (Chapter 3) could be crystallized readily. However, the crystallization of *SS*-IRED was notably more challenging. Therefore, features of *SkR*-IRED which differed heavily from *SS*-IRED which may contribute towards the relative ease of crystallization for this enzyme were sought.

An alignment of the genes coding for *SkR*-IRED and *SS*-IRED showed that *SkR*-IRED has a much longer *N*-terminus upstream to the consensus sequence (21 amino acids, highlighted in blue in Figure 4.14) than *SS*-IRED (5 amino acids, highlighted in grey in Figure 4.14).

```
SkR-IRED   MPDNPSTKGRMMRNQQAEHTPVTVIGLGLMGQALAGAFLGAGHPTTVWNRTAAKAEPLVA 60
SS-IRED    -----------MSKQ-----SVTVIGLGPMGQAMVNTFLDNGHEVTVWNRTASKAEALVA 44
                      * :*      .******* ****:..:**. ** .*******:***.***

SkR-IRED   RGAKSAGSVAEAVAASPLVVVCVSDYDAVHALLDPLDGTALQGRTLVNLTSGTSAQARER 120
SS-IRED    RGAVLAPTVEDALSANELIVLSLTDYDAVYAILEPVTG-SLSGKVIANLSSDTPDKAREA 103
           ***  * :* :*::*. *:*:.::*****:*:*:*: * :*.*:.:.**:*.*. :***

SkR-IRED   AAWADGRGADYLDGAILAGPAAIGTADAVVLLSGPRSAFDPHASALGGLGAGTTYLGADH 180
SS-IRED    AKWAAKHGAKHLTGGVQVPPPLIGKPESSTYYSGPKDVFDAHEDTLKVLTN-ADYRGEDA 162
           * **  :**.:* *.: . *. **.::: .  ***:..**.* .:*  *   : * * *

SkR-IRED   GLASLYDAAGLVMMWSILNGFLQGAALLGTAGVDATTFAPFITQGIGTVADWLPGYARQI 240
SS-IRED    GLAAMYYQAQMTIFWTTMLSYYQTLALGQANGVSAKELLPYATMMTSMMPHFLELYAQHV 222
           ***::*  * :.::*: : .: *  **  : **.*. : *: *   . :..:*  **:::

SkR-IRED   DDGAYPADDAAIDTHLATMEHLIHESEFLGVNAELPRFIKALADRAVADGHGGSGYPALI 300
SS-IRED    DSADYPGDVDRLAMGAASVDHVLHTHQDAGVSTVLPAAVAEIFKAGMEKGFAENSFSSLI 282
           *.. **.*   :    *:::*::*  :  **.: **  :  : . .: .*.. ..:.:**

SkR-IRED   EQFRTHSGK 309
SS-IRED    EVLKKPAV- 290
           * ::. :
```

Figure 4.14: A sequence alignment of *SkR*-IRED and *SS*-IRED. The comparison shows that *SkR*-IRED has a significantly longer *N*-terminus (highlighted in blue). The *N*-terminus of *SS*-IRED, by contrast, consisted of only 5 amino acids (highlighted in grey).

In the crystal packing of *SkR*-IRED, this lengthy *N*-terminus forms protein-protein crystal- contacts with its neighbouring crystallographic subunits, possibly facilitating crystallization. A modified construct of *SS*-IRED with the *N*-terminus of *SkR*-IRED was

therefore designed in an effort to produce a more readily crystallisable analogue of *SS*-IRED.

The sequence of *SS*-IRED was truncated to the consensus sequence of the two proteins by removal of its original *N*-terminus (highlighted grey in Figure 4.14). The *N*-terminus of *SkR*-IRED (highlighted blue in Figure 4.14) was then added to the *N*-terminus of *SS*-IRED. The resultant construct was termed *SS*-IRED+ (Figure 4.15).

```
SS-IRED     -----------MSKQ-----SVTVIGLGPMGQAMVNTFLDNGHEVTVWNRTASKAEALVA 44
SS-IRED+    MPDNPSTKGRMMRNQQAEHTPVTVIGLGPMGQAMVNTFLDNGHEVTVWNRTASKAEALVA 60
SkR-IRED    MPDNPSTKGRMMRNQQAEHTPVTVIGLGLMGQALAGAFLGAGHPTTVWNRTAAKAEPLVA 60
                       * :*     .******* ****:..:**. ** .*******:***.***

SS-IRED     RGAVLAPTVEDALSANELIVLSLTDYDAVYAILEPVTG-SLSGKVIANLSSDTPDKAREA 103
SS-IRED+    RGAVLAPTVEDALSANELIVLSLTDYDAVYAILEPVTG-SLSGKVIANLSSDTPDKAREA 119
SkR-IRED    RGAKSAGSVAEAVAASPLVVVCVSDYDAVHALLDPLDGTALQGRTLVNLTSGTSAQARER 120
            ***  * :* :*::*. *:*:.::*****:*:*:*: * :*.*:.:.**:*.*. :***

SS-IRED     AKWAAKHGAKHLTGGVQVPPPLIGKPESSTYYSGPKDVFDAHEDTLKVLTN-ADYRGEDA 162
SS-IRED+    AKWAAKHGAKHLTGGVQVPPPLIGKPESSTYYSGPKDVFDAHEDTLKVLTN-ADYRGEDA 178
SkR-IRED    AAWADGRGADYLDGAILAGPAAIGTADAVVLLSGPRSAFDPHASALGGLGAGTTYLGADH 180
            * **  :**.:* *.: . *. **..:: .  ***:..**.* .:*  *   : * * *

SS-IRED     GLAAMYYQAQMTIFWTTMLSYYQTLALGQANGVSAKELLPYATMMTSMMPHFLELYAQHV 222
SS-IRED+    GLAAMYYQAQMTIFWTTMLSYYQTLALGQANGVSAKELLPYATMMTSMMPHFLELYAQHV 238
SkR-IRED    GLASLYDAAGLVMMWSILNGFLQGAALLGTAGVDATTFAPFITQGIGTVADWLPGYARQI 240
            ***::*  * :.::*: : .: *  **  : **.*. : *: *   . :..:*  **:::

SS-IRED     DSADYPGDVDRLAMGAASVDHVLHTHQDAGVSTVLPAAVAEIFKAGMEKGFAENSFSSLI 282
SS-IRED+    DSADYPGDVDRLAMGAASVDHVLHTHQDAGVSTVLPAAVAEIFKAGMEKGFAENSFSSLI 298
SkR-IRED    DDGAYPADDAAIDTHLATMEHLIHESEFLGVNAELPRFIKALADRAVADGHGGSGYPALI 300
            *.. **.*   :    *:::*::*  :  **.: **  :  : . .: .*.. ..:.:**

SS-IRED     EVLKKPAV- 290
SS-IRED+    EVLKKPAV- 306
SkR-IRED    EQFRTHSGK 309
            * ::. :
```

Figure 4.15: The sequence of *SS*-IRED+ (middle), a modified version of the *SS*-IRED enzyme, in alignment with the original *SS*-IRED sequence (top) and the sequence of *SkR*-IRED (bottom). The *N*-terminus of *SS*-IRED has been truncated to the consensus sequence of the two wild-type enzymes *SS*-IRED and *SkR*-IRED, and the *N*-terminus of *SkR*-IRED has been inserted in its place (highlighted in blue).

The gene coding for *SS*-IRED+ was codon-optimised for expression in *E. coli* and commercially synthesised.

Amplification of target gene

Primers with LIC extensions were designed and commercially synthesised for the amplification of the gene coding for *SS*-IRED+. The synthetic gene coding for *SS*-IRED

with codon optimisation for expression in *E. coli* was used as a template for PCR amplification. The PCR reaction product was visualised by agarose gel electrophoresis (Figure 4.16).

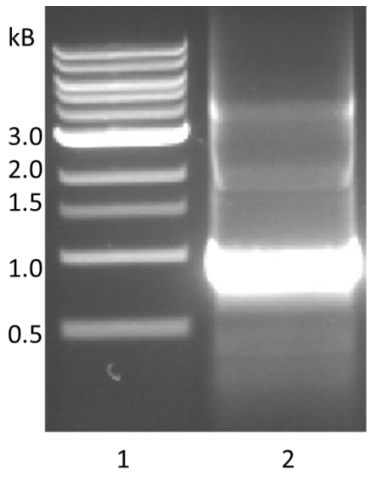


Figure 4.16: Agarose gel visualising the amplification of the gene coding for *SS*-IRED+ by PCR. Lane 1: NEB 1 kB ladder, lane 2: PCR product of gene coding for *SS*-IRED+.

A bright band at the expected length of approximately 900 bp is visible in the analysis of PCR amplification of the *SS*-IRED+ gene, which was then excised and purified.

Subcloning into pET-YSBLIC3C expression vector

The purified *SS*-IRED+ gene was treated with T4 DNA polymerase and annealed with linearised T4 DNA polymerase-treated pET-YSBLIC3C vector. *E. coli* was transformed with the annealed plasmid. DNA was isolated and a diagnostic restriction digest carried out on the purified plasmid to assess for the presence of the *SS*-IRED+ gene (Figure 4.17).

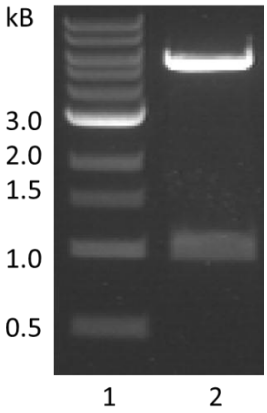


Figure 4.17: Agarose gel visualising restriction digest of gene coding for *SS*-IRED+ subcloned into the pET-YSBLIC3C expression vector. Lane 1: NEB 1 kB ladder. Lane 2: pET-YSBLIC3C plasmid containing *SS*-IRED+ gene, digested with *Nco*I and *Nde*I. Two bands are observed in lane 2; the pET-YSBLIC3C vector backbone (~5.5 kB) and the *SS*-IRED+ gene (~1000 bp)

The restriction digest showed an appropriately sized insert of approximately 1000 bp arising from the gene coding for *SS*-IRED+. The agarose gel indicated that the cloning experiment had been successful, and the corresponding plasmid was submitted for sequencing analysis.

DNA sequencing

The DNA sample corresponding to lane 2 in Figure 4.14 was submitted for sequencing at GATC Biotech. DNA sequencing confirmed that the gene present in the sample had 100 % sequence identity with the target *SS*-IRED+ sequence.

Small-scale expression testing

As *SS*-IRED+ was a novel construct, expression was tested in 10 mL scale in order to determine the optimum growth temperature for the production of soluble recombinant protein. Samples were incubated (37 °C, 180 rpm) until an OD_{600} of 0.8 was reached. Cultures were induced with IPTG and incubated (180 rpm, 20 h) at multiple growth temperatures; 16, 30 and 37 °C. Samples from both soluble and insoluble fractions for all expression tests and controls were analysed by SDS-PAGE (Figure 4.18).

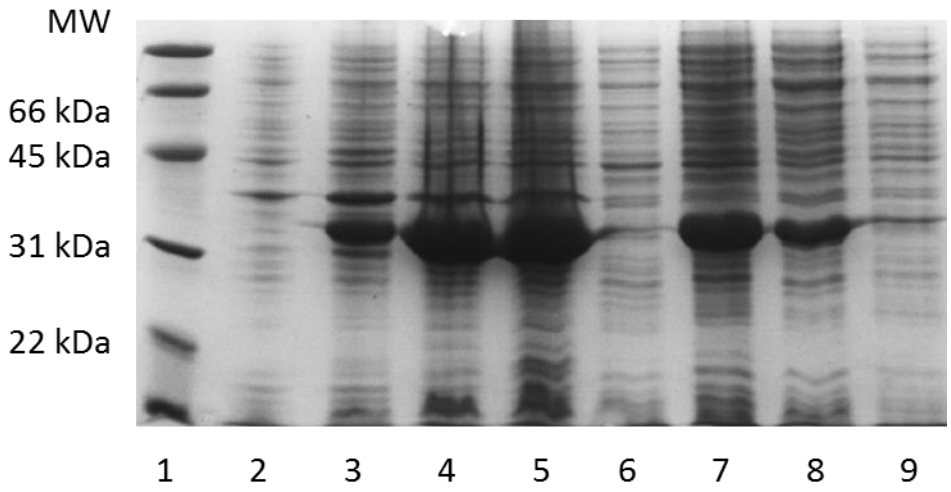


Figure 4.18: SDS-PAGE analysis of small scale expression test for the production of *SS*-IRED+. Lane 1: BioRad low range molecular weight marker. Lane 2: Insoluble fraction for uninduced control cells. Lanes 3 – 5: Insoluble fractions for expression at 16, 30 and 37 °C respectively. Lane 6: Soluble fraction for uninduced control cells. Lanes 7 – 9: Soluble fractions for expression at 16, 30 and 37 °C respectively. The target protein is present in both soluble and insoluble fractions for all induced samples.

The expression tests demonstrated that soluble expression was present at all growth temperatures. The highest levels of soluble expression were achieved at a growth temperature of 16 °C (lane 7). This condition was therefore employed for large scale expression of *SS*-IRED+.

Large scale expression and purification

SS-IRED+ was expressed in the pET-YSBLIC3C vector in *E. coli* BL21 (DE3) cells in 2 L LB-Kan medium at the previously established optimal expression temperature of 16 °C. Protein was initially purified from lysed, resuspended cells by nickel affinity chromatography, and protein samples were analysed by SDS-PAGE to assess protein purity (Figure 4.19).

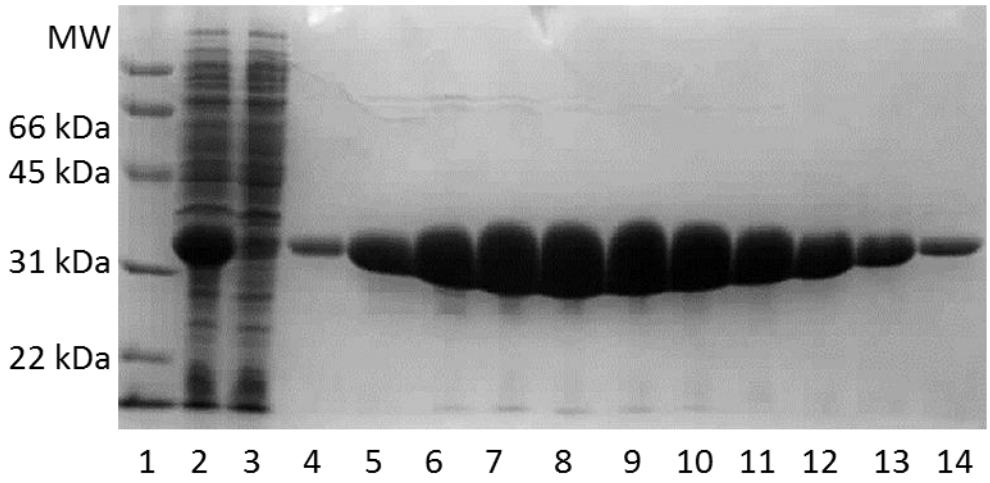


Figure 4.19: SDS-PAGE gel analysis of purification of *SS*-IRED+ by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: crude cell lysate, lane 3: unbound flow-through fraction collected during loading of nickel column, lanes 4 – 14: fractions containing purified protein eluted during increasing imidazole gradient. Purified *SS*-IRED+ is observed to be approximately 32 kDa by SDS-PAGE.

SS-IRED+ was subsequently purified by size exclusion chromatography (Figure 4.20) using a Superdex® HiLoad 75 gel filtration column. Protein-containing fractions were analysed by SDS-PAGE (Figure 4.21).

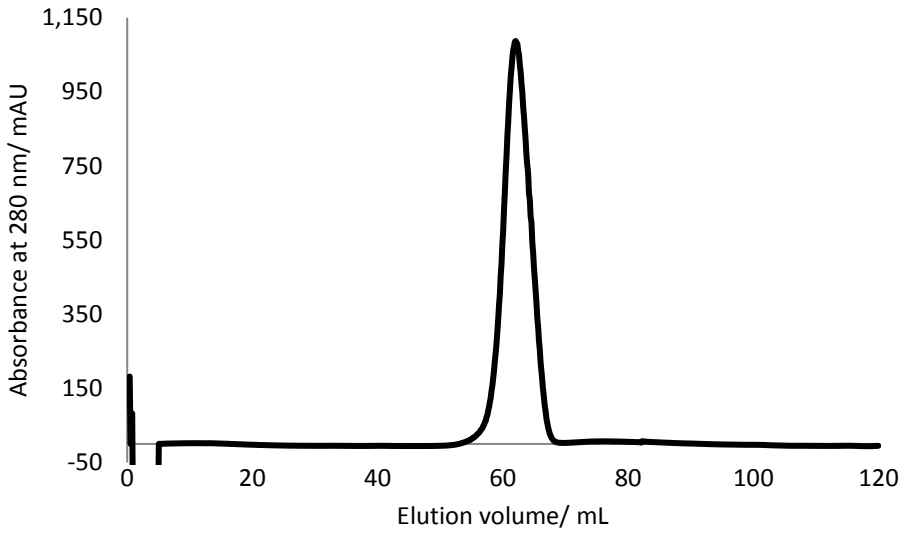


Figure 4.20: Chromatogram of the purification of *SS*-IRED+ by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

SS-IRED was eluted in a single sharp peak at 58 - 66 mL buffer. Using a calibration curve provided by GE Healthcare, the elution volume could be used to estimate a protein molecular weight of approximately 60 kDa, indicating that the dimeric structure of *SS*-IRED had not been disrupted by the modification of the *N*-terminus. Protein-containing fractions were analysed by SDS-PAGE (figure 4.18). The protein appeared to be of high purity.

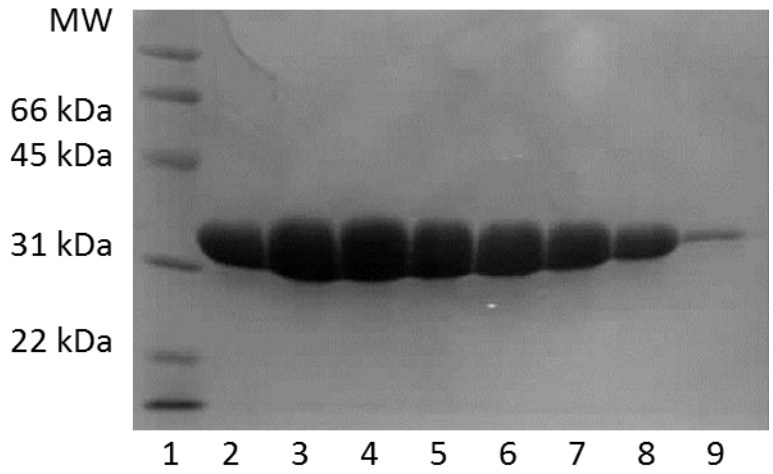


Figure 4.21: SDS-PAGE gel analysis of purification of *SS*-IRED+ by size exclusion chromatography. Lane 1: Bio-Rad low range molecular weight marker, lanes 2 – 9: fractions containing purified. Purified *SS*-IRED is observed to be approximately 32 kDa by SDS-PAGE.

Purified protein corresponding to fractions 2-9 was pooled and used for crystallization trials.

Protein crystallization

Purified protein both with and without the *N*-terminal 6His tag was subjected to a broad range of crystallization screens; PACT, INDEX, CSS 1 and 2, Hampton 1 and 2, the PEG ion screen and the MPD screen in a 96-well sitting drop format. A range of protein concentrations (20 – 120 mg ml^{-1}) were employed. No crystals were obtained in any of the attempted conditions.

Alternative methods for the crystallization of *SS*-IRED were therefore attempted.

4.4.7 Kinetic analysis and pH behaviour of *SS*-IRED

The kinetic analysis of *SS*-IRED towards the standard substrate **1a** was carried out using an spectrophotometric NADPH-depletion assay as described in section 2.5.3 to determine k_{cat} and K_m. The outcomes of the experiment are outlined in Figures 4.22 and Table 4.2.

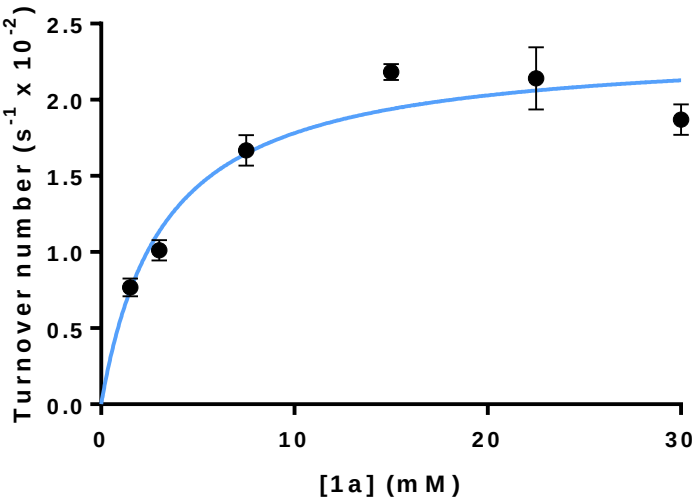


Figure 4.22: Michaelis-Menten kinetic plot of *SS*-IRED towards substrate **1a**.

Table 4.2: Kinetic parameters of *SS*-IRED towards substrate **1a**.

	Value	**Standard Error**
$\mathbf{k_{cat}}$ (s^{-1})	0.0236	±0.0011
$\mathbf{K_m}$ (mM)	3.24	±0.60

The kinetic parameters calculated by curve fitting using non-linear regression analysis returned reasonable standard errors relative to their values, suggesting that the data can be treated with confidence. The relatively low k_{cat} and high K_M suggest that the catalytic efficiency and affinity of *SS*-IRED towards **1a** is poor, and it is therefore unlikely to be a natural substrate for the enzyme.

4.4.8 Proposing a mechanism of *SS*-IRED

The mechanism of an (*R*)-selective IRED has been studied in section 3.4.6. However, the mechanism of an (*S*)-selective NADPH-dependent IRED has so far not been reported. Therefore, *SS*-IRED was used as a model for mechanistic studies of (*S*)-selective NADPH-dependent imine reductases.

A sequence alignment of the genes coding for *SkR*-IRED and *SS*-IRED showed that in place of the putative catalytic residue Asp187 of *SkR*-IRED, *SS*-IRED possesses a tyrosine residue, Tyr169 (Figure 4.21).

```
SkR-IRED  MPDNPSTKGRMMRNQQAEHTPVTVIGLGLMGQALAGAFLGAGHPTTVWNRTAAKAEPLVA 60
SS-IRED   -----------MSKQ-----SVTVIGLGPMGQAMVNTFLDNGHEVTVWNRTASKAEALVA 44
                     * :*     .******* ****:..:**. ** .*******:***.***

SkR-IRED  RGAKSAGSVAEAVAASPLVVVCVSDYDAVHALLDPLDGTALQGRTLVNLTSGTSAQARER 120
SS-IRED   RGAVLAPTVEDALSANELIVLSLTDYDAVYAILEPVTG-SLSGKVIANLSSDTPDKAREA 103
          ***  * :* :*::*. *:*:.::*****:*:*:*: * :*.*:.:.**:*.*. :***

SkR-IRED  AAWADGRGADYLDGAILAGPAAIGTADAVVLLSGPRSAFDPHASALGGLGAGTTYLGADH 180
SS-IRED   AKWAAKHGAKHLTGGVQVPPPLIGKPESSTYYSGPKDVFDAHEDTLKVLTN-ADYRGEDA 162
          * **  :**.:* *.: . *. **.::: .  ***:..**.* .:*  *   : * * *

SkR-IRED  GLASLYDAAGLVMMWSILNGFLQGAALLGTAGVDATTFAPFITQGIGTVADWLPGYARQI 240
SS-IRED   GLAAMYYQAQMTIFWTTMLSYYQTLALGQANGVSAKELLPYATMMTSMMPHFLELYAQHV 222
          ***::*  * :.::*: : .: *  **  : **.*. : *: *   . :..:*  **:::

SkR-IRED  DDGAYPADDAAIDTHLATMEHLIHESEFLGVNAELPRFIKALADRAVADGHGGSGYPALI 300
SS-IRED   DSADYPGDVDRLAMGAASVDHVLHTHQDAGVSTVLPAAVAEIFKAGMEKGFAENSFSSLI 282
          *.. **.*   :    *:::*::*  :  **.: **  :  : . .: .*.. ..:.:**

SkR-IRED  EQFRTHSGK 309
SS-IRED   EVLKKPAV- 290
          * ::. :
```

Figure 4.23: Sequence alignment of genes coding for *SkR*-IRED (top) and *SS*-IRED (bottom). The two genes are 36 % sequence similar. The putative catalytic Asp187 residue of *SkR*-IRED is replaced by residue Tyr169 in *SS*-IRED (highlighted blue).

An alignment of the 3D structures of *SkR*-IRED (section 3.1) and *SS*-IRED (section 4.4.5) could be used to investigate the relative positions of Asp178 (*SkR*-IRED) and Tyr169 (*SS*-IRED). The structural overlay was examined in the active site region of *SkR*-IRED (Figure 4.24).

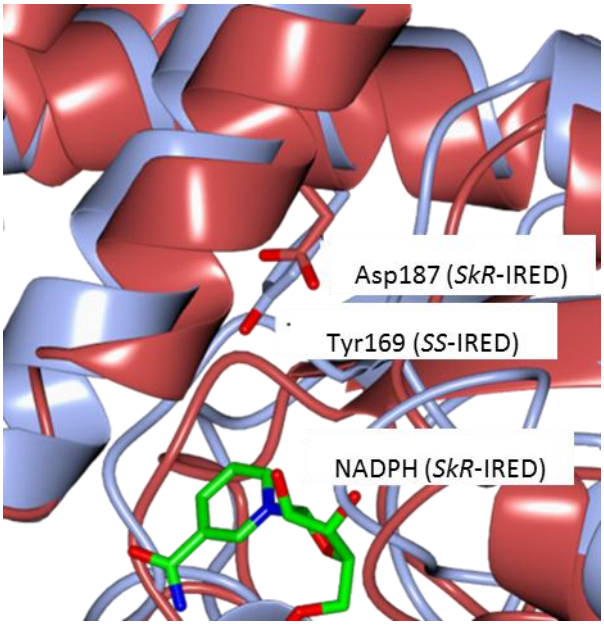


Figure 4.24: A structural overlay between the active site region of *SkR*-IRED (in complex with NADPH) and the equivalent position in *SS*-IRED in the active site region of *SkR*-IRED. The figure demonstrates that Asp187 (*SkR*-IRED) and Tyr169 (*SS*-IRED) occur in the same position.

The alignment demonstrates that residues Asp187 (*SkR*-IRED) and Tyr169 (*SS*-IRED) align structurally. It is possible, therefore, that the two residues perform a similar mechanistic role, where Tyr169 behaves as a proton donor in the reduction of imines, a mechanism for which is outlined in Figure 4.25.

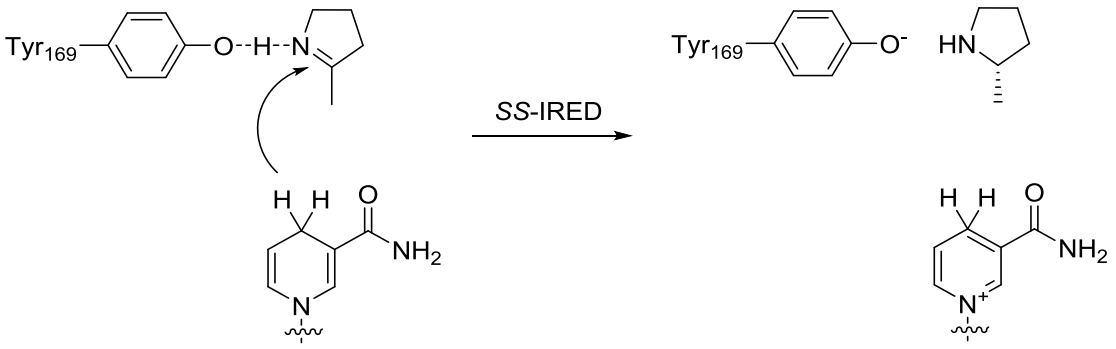


Figure 4.25: The mechanism proposed for the reduction of imines (such as **1a**) by *SS*-IRED. A proton is donated to 1a by catalytic residue Tyr169, resulting in the formation of an activated iminium ion. Hydride is then delivered to the iminium ion by NADPH to form the corresponding (*S*)-amine.

The behaviour of tyrosine as a proton donor in the reduction of imines has precedent in the naturally occurring imine reductase pteridine reductase (PTR1) (Figure 4.26).

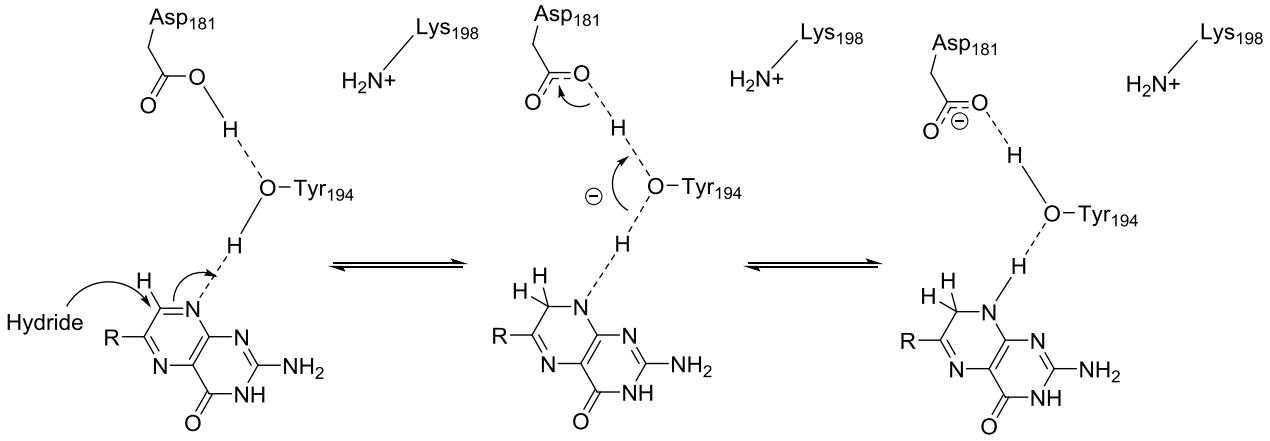


Figure 4.26: The mechanism of the imine reductase PTR1 in the first step in the NADPH-assisted reduction of an oxidised pterin to the active tetrahydro-form.
Here, residue Tyr194 hydrogen bonds to the nucleophilic centre of the pterin substrate and provides a proton for the reduction of its imine moiety. A neighbouring residue Lys198 was suggested to lower the pK_a of Tyr194 due to its positively charged side chain which would stabilise the intermediary tyrosinate ion, thus favouring its formation. It was proposed that Asp181 may also help stabilise the charged tyrosinate intermediate while providing a role in proton relay.

In light of the PTR1 mechanism, the active site of *SS*-IRED was therefore more closely in order to investigate whether, like PTR1, the active site of *SS*-IRED contains residues that could assist Tyr169 in the transfer of a proton to its imine substrate.

Active site of SS-IRED

The neighbouring residues of Tyr169 were examined in the putative active site region of *SS*-IRED. It was noted that residues Ser94 and His243 are found in the vicinity of Tyr169, which could behave as analogues to PTR1-Asp181 and PTR1-Lys198 respectively (Figure 4.27).

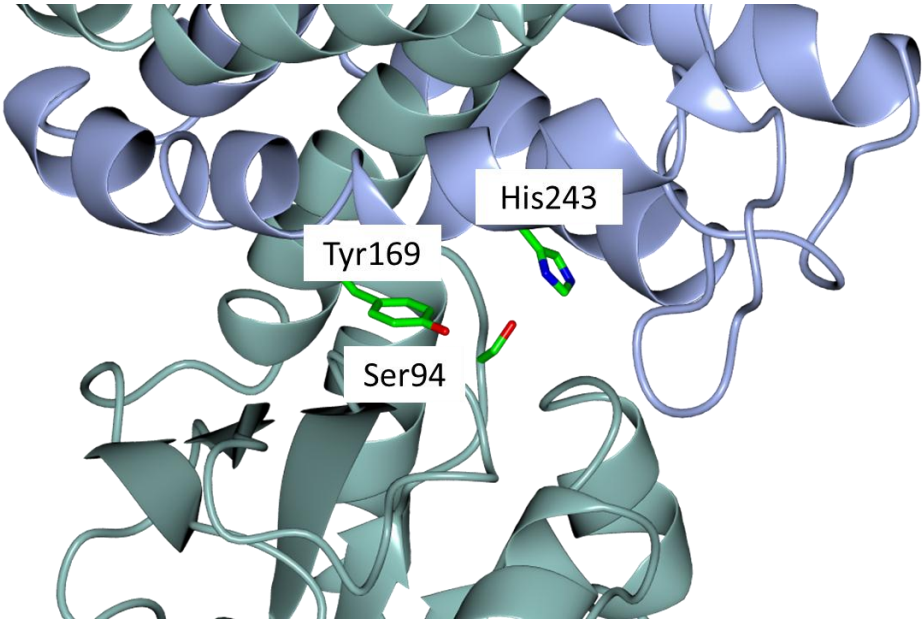


Figure 4.27: The putative active site region of *SS*-IRED. Residue Tyr169, which may have a role as a proton donor in the reduction of imines, is neighboured by residues His243 and Ser94 which may assist Tyr169 in catalysis.

Serine and histidine residues have a precedent of playing important roles in catalytic triads.[77] *SS*-IRED-Tyr169 may behave as a proton donor in the reduction of imines. This residue may be assisted by Ser94 and His243 similarly to residues Asp181 and Lys198 in PTR1. Positively charged His243 may lower the pK_a of Tyr169 by stabilising the intermediary tyrosinate ion by electrostatic interactions. Ser94 may play a role in proton relay as well as in the stabilisation of Tyr169.

4.4.9 Mutational studies towards the mechanism of *SS*-IRED

In order to assess whether residue Tyr169 could be a catalytically active residue, the wild-type gene was used as a template for the generation of the mutant Tyr169Phe. The Phe sidechain is sterically similar to that of Tyr, but lacks the hydroxyl functional group and is therefore aprotic and incapable of carrying out the proposed protonation step in the reduction of imines as shown in the mechanism in Figure 4.13. Therefore, if the resultant mutant does not possess imine reduction activity, this would support the suggestion that residue Tyr169 may behave as a proton donor. Mutants Ser94Ala and His243Ala were also generated in order to assess whether these may play a role in assisting residue Tyr169 in catalysis as described in section 4.4.6.

Site-directed mutagenesis

The wild-type gene was used as a template for the generation of mutants Tyr169Phe, Ser94Ala and His243Ala by site-directed mutagenesis (SDM). The SDM reaction was digested with *Dpn*I and used to transform *E. coli* cells. DNA was isolated from resultant colonies for each mutant using a miniprep kit, and submitted for sequencing at GATC Biotech. The genes in the isolated DNA were found to have 100 % sequence identity to the target *SS*-IRED Tyr169Phe, Ser94Ala and His243Ala mutated genes.

Expression and purification of SS-IRED mutants

SS-IRED Tyr169Phe, Ser94Ala and His243Ala were expressed in the pET-YSBLIC3C vector in *E. coli* BL21 (DE3) cells in 2 L LB-Kan medium at the optimal expression temperature for the wild-type protein of 16 °C. Protein was initially purified from lysed, resuspended cells by nickel affinity chromatography. Protein-containing fractions were analysed by SDS-PAGE (Figures 4.28 – 4.30)

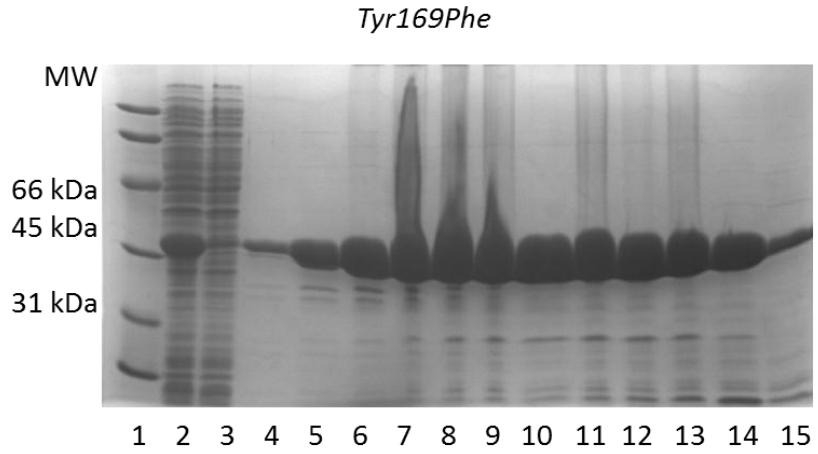


Figure 4.28: SDS-PAGE gel analysis of purification of *SS*-IRED Tyr169Phe by nickel affinity chromatography. Lane 1: Bio-Rad low range molecular weight marker, lane 2: crude cell lysate, lane 3: Unbound flow-through fractions on column binding, lanes 4 - 15: fractions containing purified protein.

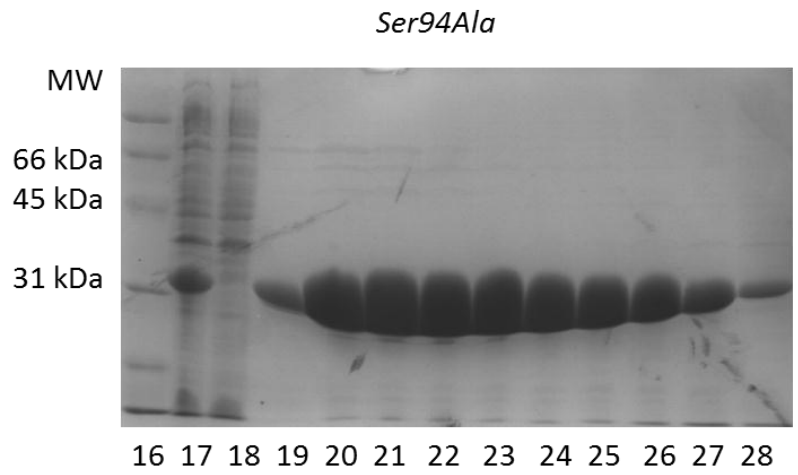


Figure 4.29: SDS-PAGE gel corresponding to the purification of *SS*-IRED Ser94Ala by nickel affinity chromatography. Lane 16: Bio-Rad low range molecular weight marker, lane 17: crude cell lysate, lane 18: Unbound flow-through fractions on column binding, lanes 19-28: fractions containing purified protein,

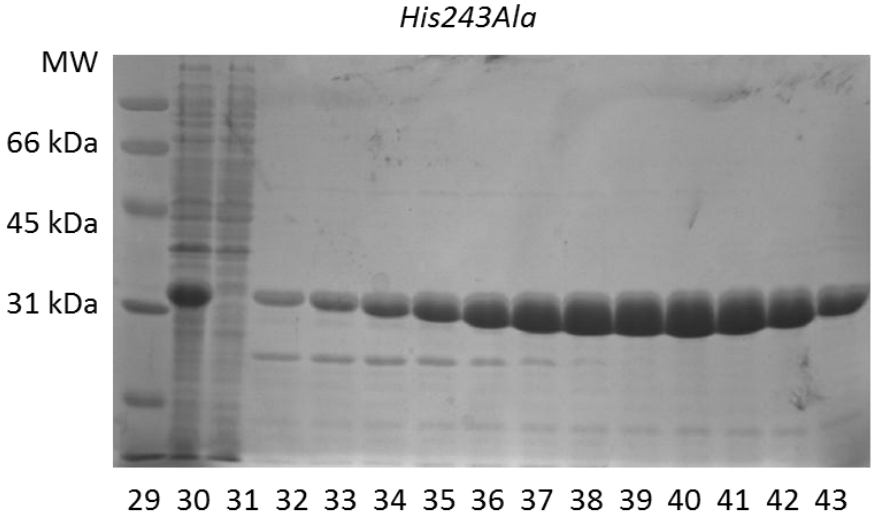


Figure 4.30: SDS-PAGE gel corresponding to the purification of *SS*-IRED His243Ala by nickel affinity chromatography. Lane 29: Bio-Rad low range molecular weight marker, lane 30: crude cell lysate, lane 31: Unbound flow-through fraction on column binding, lanes 32-43: fractions containing purified protein.

SS-IRED Tyr169Phe, Ser94Ala and His243Ala were subsequently purified by size exclusion chromatography (Figures 4.31 - 33) using a Superdex® HiLoad 75 gel filtration column. Protein-containing fractions were analysed by SDS-PAGE (Figure 4.34).

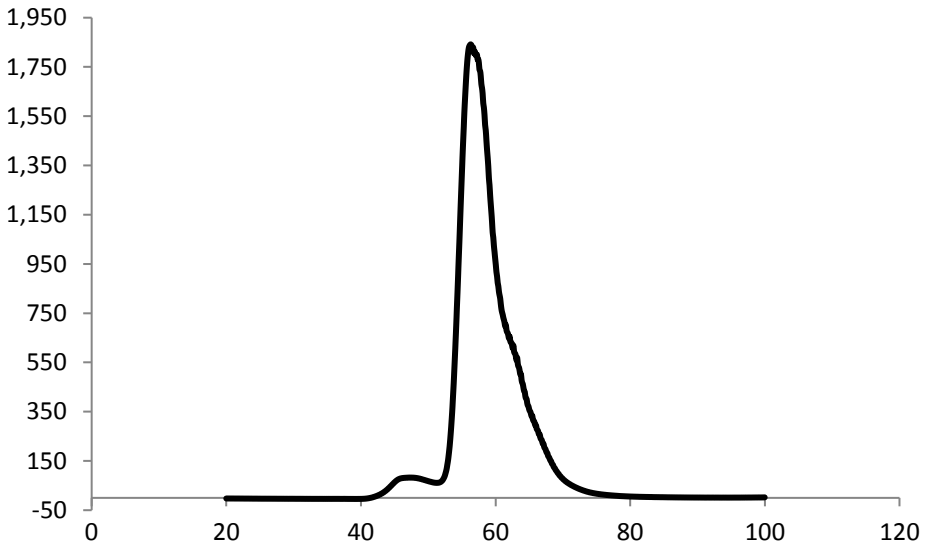


Figure 4.31: Chromatogram of the purification of *SS*-IRED Tyr169Phe by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

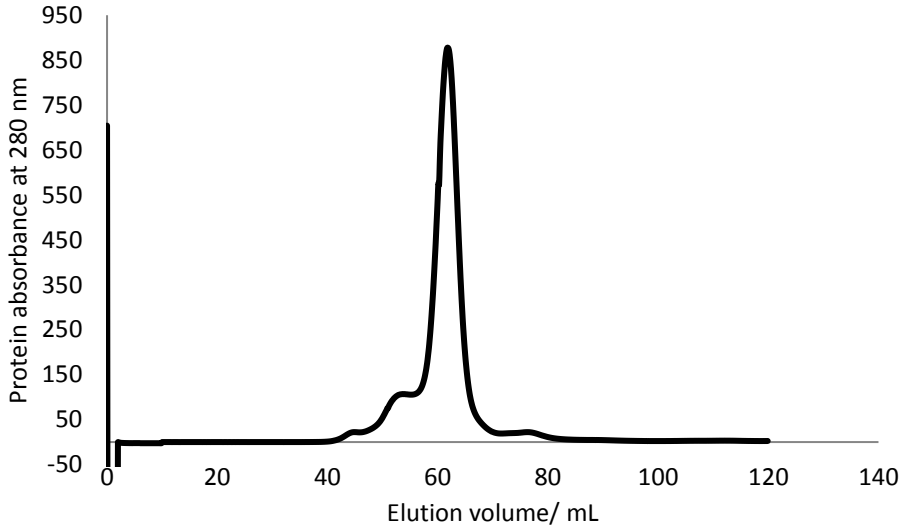


Figure 4.32: Chromatogram of the purification of *SS*-IRED Ser94Ala by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

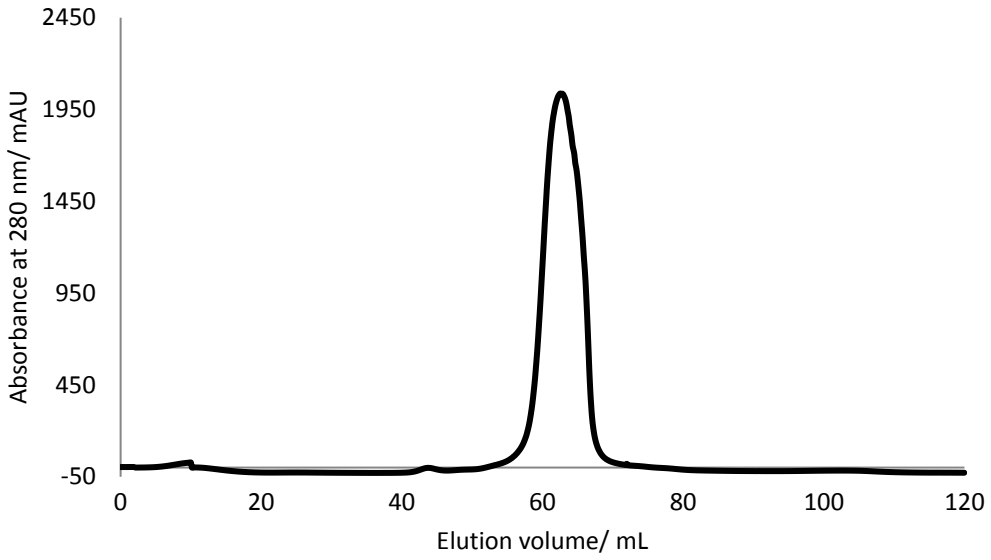


Figure 4.33: Chromatogram of the purification of *SS*-IRED His243Ala by size exclusion chromatography. The presence of protein is detected using absorbance measurements at 280 nm, which is indicated by the blue trace.

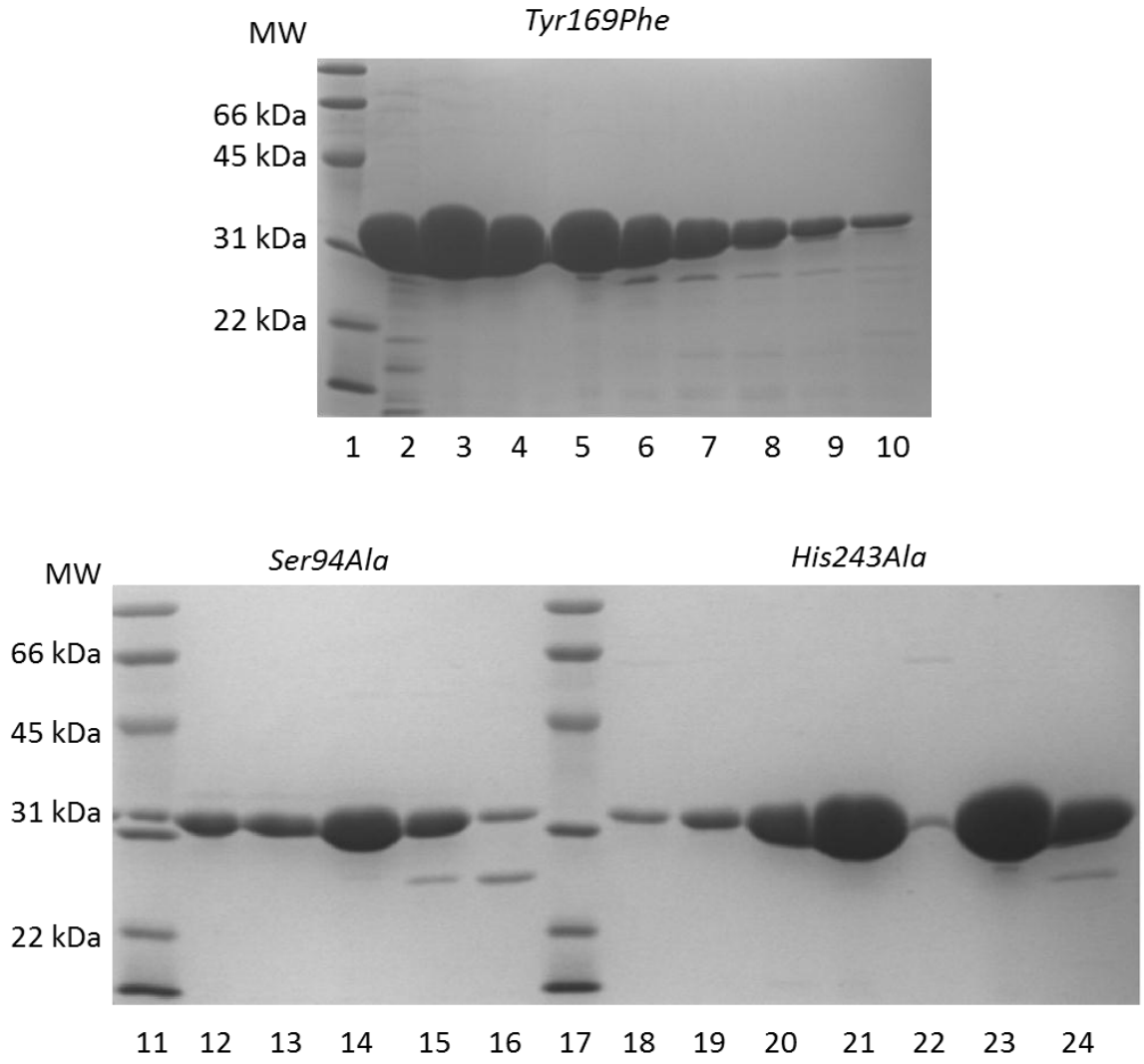


Figure 4.34: SDS-PAGE gel analysis of purification of mutants of *SS*-IRED by size exclusion chromatography. **Top:** SDS-PAGE gel corresponding to *SS*-IRED Tyr169Phe. Lane 1: Bio-Rad low range molecular weight marker, lanes 2-10: fractions containing purified protein, **Bottom left:** SDS-PAGE gel corresponding to *SS*-IRED Ser94Ala. Lane 11: Bio-Rad low range molecular weight marker, lanes 11 - 16: fractions containing purified protein, **Bottom right:** SDS-PAGE gel corresponding to *SS*-IRED His243Ala. Lane 17: Bio-Rad low range molecular weight marker, lanes 18 - 24: fractions containing purified protein,

All *SS*-IRED mutants were eluted in a single peak with maxima at 57, 62 and 63 mL for Tyr169Phe, Ser94Ala and His243Ala respectively. Using a calibration curve provided by GE Healthcare, the elution volume could be used to estimate a protein molecular

weight of approximately 60 kDa for mutants Ser94Ala and His243Ala, indicating that the dimeric structure of *SS*-IRED had not been preserved during point mutations. However, the earlier elution volume for *SS*-IRED Tyr169Phe suggested a lower molecular weight, which could correspond to a protein monomer. This would suggest that dimerization had been disrupted by the mutation. This is investigated further in section 4.4.8.

For subsequent use, Protein corresponding to fractions 2-10, 12-15 and 18-23 was pooled. Fractions containing considerable levels of non-target proteins were discarded.

***Kinetic assays using SkR-IRED* mutant**

A spectrophotometric-based NADPH depletion assay was used to assess the kinetic parameters of *SS*-IRED Tyr169Phe, Ser94Ala and His243Ala towards the standard substrate **1a** under the same conditions used for the wild-type protein. No NADPH-depletion was observed for **1a** concentrations between 1 and 30 mM. The results suggested that these residues may have importance for catalysis. However, further investigation will be required to confirm a precise role that they play for imine reductase activity, as the elimination of activity may have arisen for other reasons.

4.4.10 Conclusion

An (*S*)-selective NADPH-dependent imine reductase from *Streptomyces* sp. GF3546 has been sub-cloned, recombinantly expressed in *E. coli*, purified to homogeneity and crystallized. A 3.2 Å resolution structure was obtained of *SS*-IRED in the *apo* form.

A probable active site region of *SS*-IRED was identified using a structural alignment with the previously characterised (*R*)-selective imine reductase *SkR*-IRED (chapter 3). A tyrosine residue, Tyr169, is observed in the active site in place of the putative catalytic residue of *SkR*-IRED. It was suggested that this may act as a proton donor in imine reduction in a similar manner to the naturally occurring imine reductase PTR1, which reduces oxidised pteridines. The PTR1 mechanism involves assistance from neighbouring aspartate and lysine residues. Tyr169 of *SS*-IRED is neighboured by residues Ser94 and His243 in the active site cavity, which could assist the tyrosine residue in catalysis in a similar manner to Asp and Lys in PTR1.

In order to assess whether this proposed mechanism is feasible, *SS*-IRED mutants Tyr169Phe, Ser94Ala and His243Ala were generated. Samples of mutants were loaded onto an analytical gel filtration column to assess their homogeneity and molecular weight, to test whether the dimeric structure had been preserved. The mutants were screened for activity towards imine **1a**. A spectrophotometric-based NADPH depletion assay showed that none of the mutants displayed activity towards the imine substrate.

This supports the hypothesis that Tyr169 may have a role as a proton donor with assistance from Ser94 and His243. However, more data is required in order to confirm whether this is the case, as elimination of activity could occur for any number of reasons, such as disruptions in the 3D structure of the protein which may inhibit binding of the substrate.

To investigate the mechanism and selectivity of *SS*-IRED further, it is important to obtain a structure of the enzyme in complex with NADPH and an imine substrate, as this will provide conclusive evidence towards the residues that are necessary for catalysis, as well as helping to rationalise the selectivity of the enzyme by observing the orientation adopted by the substrate in the active site. As *SS*-IRED is difficult to crystallize, it has so far not been possible to obtain such a complex.

Therefore, modifications and homologues of *SS*-IRED have been explored in an effort to obtain an improved structure of an (*S*)-selective imine reductase. So far, crystallization has not been achieved in this project using this method. This goal is therefore ongoing within the scope of this project.

5. Investigating the basis for enantioselectivity of NADPH-dependent IREDs

5.1 Introduction

Many IREDs that have been characterized display a high degree of enantioselectivity. Examples include *SkR*-IRED from *Streptomyces kanamyceticus* and *SR*-IRED from *Streptomyces* sp. GF3587, which both catalyse the formation of (*R*)-**1b** with greater than 98 % enantiomeric excess,[53, 54, 59, 60] and *SS*-IRED from *Streptomyces* sp. GF3546 and *BcS*-IRED from *Bacillus cereus* which catalyse the formation of (*S*)-**2b** also with at least 98 % optical purity.[54, 57, 78] However, despite high levels of enantioselectivity commonly displayed by IREDs, the reasons for this selectivity have not yet been determined.

Gaining an understanding of their enantioselectivity would provide an insight into the nature of substrate binding in IREDs, which has thus far been difficult to characterize, in part due to the typically large active sites and high degrees of dynamicity belonging to these enzymes, as well as generally low affinities of IREDs for imine substrates which have thus far been identified, which makes it challenging to obtain complexes of IREDs with their substrates.[59, 60, 78] Additionally, knowledge of the residues responsible for directing enantioselectivity could provide a basis for customising IREDs by rational mutagenesis in terms of which amine enantiomer they produce and improving the optical purity of products formed.

5.2 Aims

This chapter will describe investigations into the differences between (*R*)- and (*S*)-selective IREDs using two approaches. Firstly, the substrate binding pocket of (*R*)- and (*S*)-selective IREDs will be examined to determine structural differences in the substrate binding region between these two groups of IREDs that may direct enantioselectivity. Secondly, an analysis by Wetzl and coworkers (Wetzl) on the bacterial protein-sequence space of imine reductases will be reviewed and examined for sequence differences between (*R*)- and (*S*)-selective IREDs.[65]

5.2 Comparison of the active site of an (*R*)- and (*S*)-selective IRED

Identifying residues that direct enantioselectivity in IREDs

The structure of an IRED from *Amycolatopsis orientalis* (*Ao*IRED) was recently reported in complex with an amine product, (*R*)-1-methyl-1,2,3,4-tetrahydroisoquinoline (**4b**). This structure revealed the location of the previously unknown binding site of substrates and products to the enzyme active site.

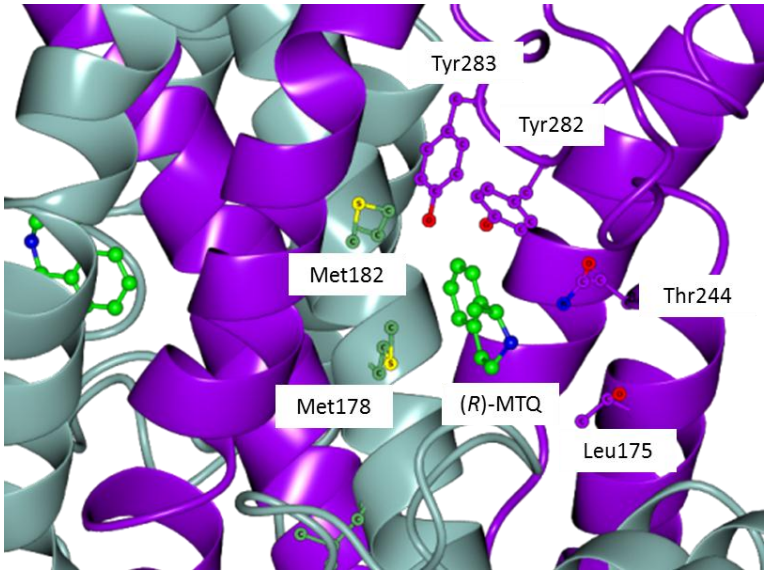


Figure 5.1: The active site position in the crystal structure of *Ao*IRED in complex with (*R*)-1-methyl-1,2,3,4-tetrahydroisoquinoline ((*R*)-**4b**). The figure demonstrates the nature of imine substrate binding in *Ao*IRED. The amine product (*R*)-**4b** is bound in a cleft between the *C*-terminal domain of subunit A (green) and the *N*-terminal domain of subunit B (purple).

The enantioselectivity of *Ao*IRED is substrate-dependent. While it displays predominantly (*S*)-selectivity, there are a number of substrates, including **4a**, for which it shows selectivity for the formation of the (*R*)-enantiomer. This enzyme therefore does not provide a clear subject for the study of enantioselectivity in IREDs.

Instead, two more uniformly selective IREDs will be used as models in this study; *SS*-IRED and *SR*-IRED from *Streptomyces sp.* GF3546 and 3587 respectively. *SR*-IRED has been selected for this study because of its superior activity towards imine substrates in comparison with *SkR*-IRED.[59] An overlay of the putative substrate binding sites of

SS-IRED and *SR*-IRED is pictured in figure 5.2. The location of the active site has been proposed on the assumption that it lays in the equivalent region of the protein to that of *Ao*IRED. Inspection of the binding sites of each enzyme reveals five positions at which residues differ between the two IREDs; L176, M179, L183, H242 and Y283 in *SR*-IRED which structurally overlay with M173, F176, M180, G237 and F278 in *SS*-IRED.

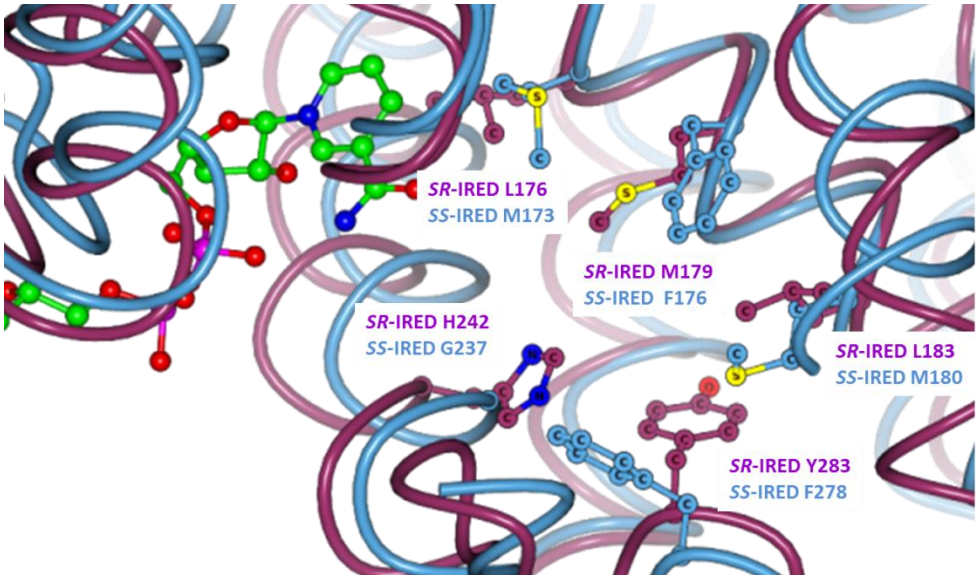


Figure 5.2: Active site overlay of *SR*-IRED and *SS*-IRED at the putative amine binding site, based on the location of imine binding in *Ao*IRED. The figure details the positions at which residues are not conserved between enzymes.

Preparation of reciprocal mutants of SR-IRED and SS-IRED

Mutagenesis experiments were performed to assess whether these five residues direct enantioselectivity in either enzyme. For each enzyme at each of the five positions, reciprocal mutants were generated (*SR*-IRED L176M, M179F, L183M, H242G and Y283F and *SS*-IRED M173L, F176M, M180L, G237H and F278Y) by site-directed mutagenesis. Each mutant will be used in biotransformations of model imine **1a**, and the resultant enantioselectivity measured by chiral GC analysis. This will demonstrate whether enantioselectivity has been changed as a result of mutation to the residue found in the IRED with opposite enantioselectivity.

Heterologous expression tests were carried out in *E. coli* BL21(DE3) and presence of soluble expression was assessed using SDS-PAGE analysis (Figure 5.3).

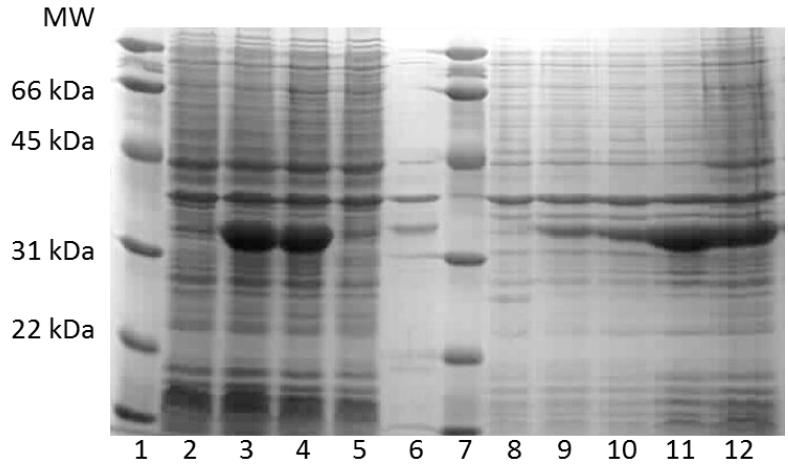


Figure 5.3. SDS-PAGE analysis of expression of mutants A1 - B5. Samples 1 and 7 correspond to a low range molecular weight marker supplied by NEB. Samples 2 – 6 and 8 - 12 correspond to the soluble and insoluble fractions for expression of *SR*-IRED mutants A1 – 5 in *E. coli* BL21(DE3) respectively. Soluble expression is seen for mutants A2, A3 and A5. No clear soluble expression is seen for mutant A4, and no expression at all is seen for mutant A1.

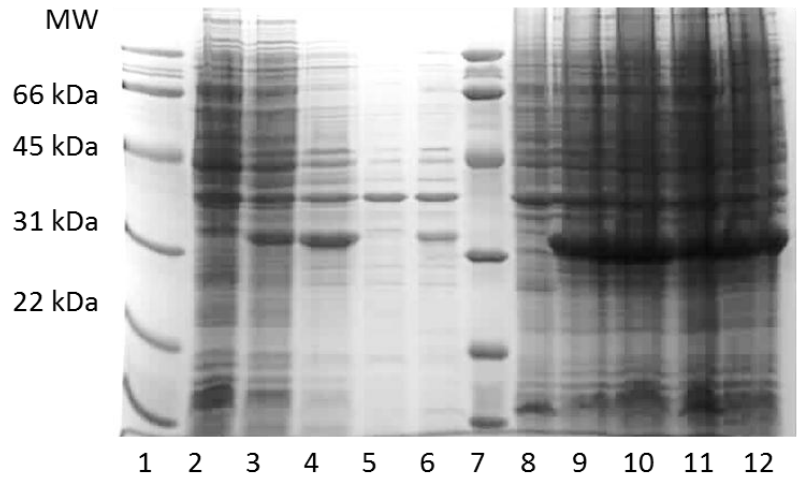


Figure 5.4. SDS-PAGE analysis of expression of mutants B1-B5. Samples 1 and 7 correspond to a low range molecular weight marker supplied by NEB. Samples 2 – 6 and 8 - 12 correspond to the soluble and insoluble fractions for expression of *SS*-IRED mutants B1 – 5 in *E. coli* BL21(DE3) respectively. Soluble expression is seen for mutants B2, B3 and B5. Only insoluble expression is seen for mutant B4, and no expression at all is seen for mutant B1.

Expression tests revealed that seven out of ten mutants expressed soluble protein, with particularly high levels of soluble expression for mutants *SS*-IRED B2 and B3 and *SR*-IRED A2 and A3. Mutants *SR*-IRED A1 and B1 did not express at all, and mutant A4 expressed only insolubly.

Mutants which expressed solubly during expression tests (*SR*-IRED A2, A3 and A5, and *SS*-IRED B2, B3 and B5) were expressed on 1 L scale and purified using Ni-affinity chromatography.

GC analysis of biotransformations with mutants of SR-IRED and SS-IRED

Purified mutants were used in biotransformations of **1a** and reaction products were analysed by chiral-GC after 24 h reaction time. The retention times of imine **1a** and the corresponding products (*R*)-**1b** and (*S*)-**1b** were identified by analysing the GC traces recorded from commercially prepared **1a**, **1b** and (*R*)-**1b** (Figures 5.5, 5.6 and 5.7). Biotransformations of **1a** performed using mutants of *SR*-IRED and *SS*-IRED were analysed under the same conditions as the substrate and product standards using chiral GC. The results of these biotransformations are outlined in Table 5.1.

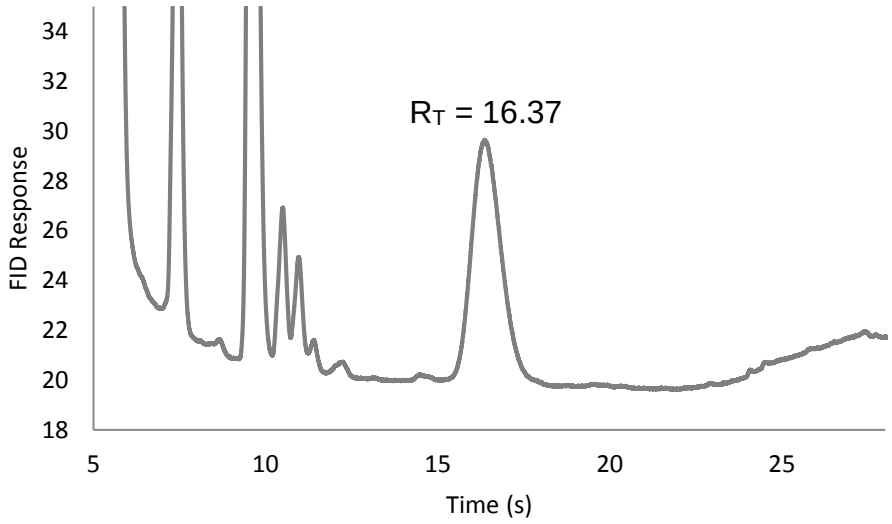


Figure 5.5: Chiral GC analysis of **1a** (5 mM) in EtOAc. The figure demonstrates that imine **1a** has a retention time of 16.37 min.

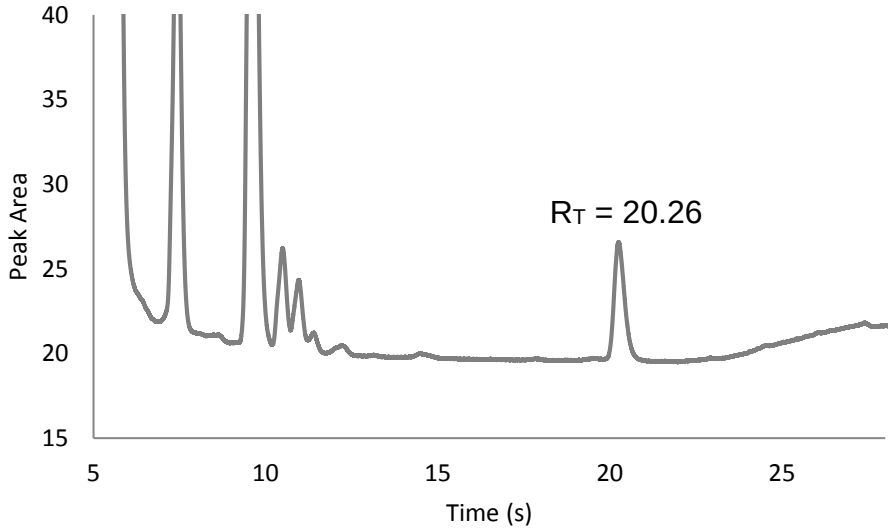


Figure 5.6: Chiral GC analysis of (*R*)-**1b** (5 mM) in EtOAc. The figure demonstrates that the optically pure amine (*R*)-**1b** has a retention time of 20.26 min.

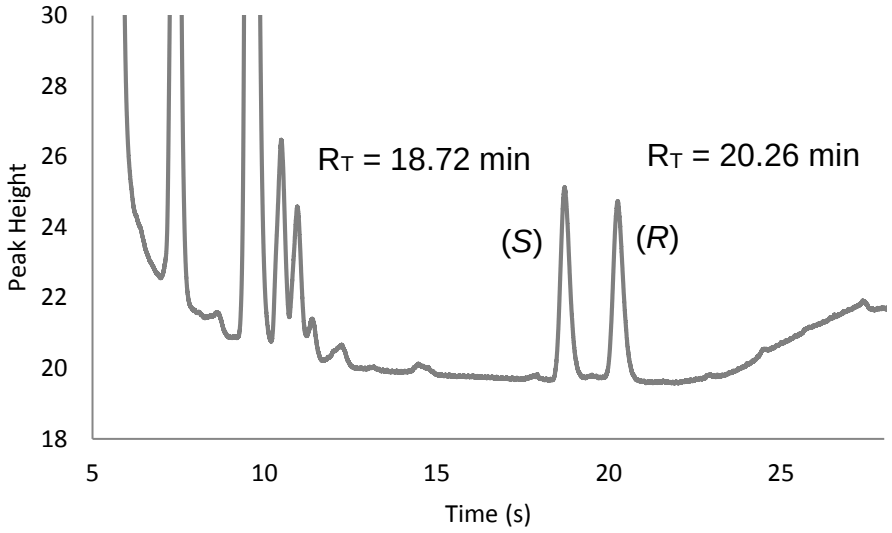


Figure 5.7: Chiral GC analysis of racemic **1b** (5 mM) in EtOAc. The figure demonstrates that the optically pure amine (*S*)-**1b** has a retention time of 18.72 min, as the peak at 20.26 minutes corresponds to the (*R*)-**1b** enantiomer (Figure 5.5).

Table 5.1: The results of GC analysis of biotransformations of imine **1a** with mutants *SR*-IRED A1 – A5 and *SS*-IRED B1 – B5. The analysis demonstrates the enantioselectivity of each mutant by demonstrating the stereochemistry of the amine products formed in each reaction. Mutant reactions were performed in duplicate, and have been referred to as repeat 1 and repeat 2.

Enzyme		**Peak area of imine 1a**	**Peak area of amine (*R*)-1b**	**Peak area of amine (*S*)-1b**	**ee 1b/ %**
SR-IRED A2	Repeat 1:	101	735	7	98 (*R*)
	Repeat 2:	160	522	0	>99 (*R*)
SR-IRED A3	Repeat 1:	177	716	2	>99 (*R*)
	Repeat 2:	188	551	0	>99 (*R*)
SR-IRED A5	Repeat 1:	173	534	2	>99 (*R*)
	Repeat 2:	176	495	2	>99 (*R*)
SS-IRED B2	Repeat 1:	679	0	50	>99 (*S*)
	Repeat 2:	816	0	56	>99 (*S*)
SS-IRED B3	Repeat 1:	884	0	62	>99 (*S*)
	Repeat 2:	745	0	22	>99 (*S*)
SS-IRED B5	Repeat 1:	554	0	398	>99 (*S*)
	Repeat 2:	353	12	353	94
SR-IRED wild-type		246	741	0	>99 (*R*)
SS-IRED wild-type		174	29	732.7	92 (*S*)

Chiral GC data for the conversion of **1a** shows that no mutations resulted in reduced enantioselectivity. This may indicate that none of these residues are responsible for directing enantioselectivity. However, this could indicate that more than a single mutation is necessary in order to alter the enantiospecificity of an IRED, particularly if broader structural differences are implicated.

5.3 Analysis of bacterial protein-sequence space of imine reductases for the basis of enantioselectivity

The availability of IRED protein sequences whose activity has been confirmed in a laboratory setting have been used as a reference in BLAST searches that have facilitated the discovery of a large number of sequences of putative IREDs in addition to those already characterised in the literature. Most notably, such investigations were used by Hauer and coworkers to construct an invaluable database of over 350 protein sequences, termed the Imine Reductase Engineering Database (IBD), which putatively possess imine reducing activity.[62]

However, although many potential imine reductase genes had been identified, the number of imine reductases that had been experimentally isolated and characterized remained low. In an effort to increase the number of characterized imine reductases, Wetzl and coworkers selected a total of an additional 23 putative IRED genes to express, purify and characterize in terms of their substrate scope and enantioselectivity.[65]

Because protein sequences that were selected were highly diverse and sourced from multiple different organisms, and involved practical characterizations of each enzyme, this study gave rise to an unprecedented opportunity to allow investigators to make links between sequence properties and the catalytic properties of an IRED.

In the scope of this chapter, the focus will be on the relation of (*R*)- and (*S*)-selectivity to sequence motifs and individual residues within the enzyme active site. In the study by Wetzl and coworkers, it was hypothesised that residues occurring in the C-terminal domain would be more important in influencing the catalytic properties of an IRED than the N-terminal Rossman fold domain, whose purpose is NADPH-binding and is more highly conserved between IREDs.

In this study, a profile Hidden Markov Model (HMM) was built using the sequence-alignment of a group of six IREDs previously characterized in the literature. The HMM was used to find protein sequences with residues conserved at the most statistically important positions, which led to the identification of 702 protein sequences. Sequences were allocated to clusters based on their homology, with the majority of these protein sequences being organised into 85 clusters, only 15 of which contained 3 or more IREDs. Genes were selected across multiple clusters and organisms of origin to maximise sequence diversity.

All proteins were heterologously expressed and purified and biotransformations were carried out in a range of substrates, displayed in Figure 5.8. Biotransformations were analysed by chiral HPLC in order to determine level of conversion and enantiomeric purity of enzymatically-formed products.

For some enzymes enantioselectivity was substrate-dependent, where they displayed strong (*R*)-selectivity toward some substrates and strong (*S*)-selectivity for other substrates. In order to carry out a clearer analysis on the sequence basis of enantioselectivity for the purposes of this study, only enzymes which were consistently (*R*)- or (*S*)-selective across the entire range of tested substrates have included. All enzymes that fit these criteria are listed in Table 5.2. The table includes both reaction data from the study by Wetzl and coworkers, as well as data from *SkR*-IRED from the results section of chapter 3 in this thesis, *BcS*-IRED as described by Man and coworkers and *SR*-IRED as described by Hussain and co-workers.[59, 65, 78]

Table 5.2: Names and source organisms of enzymes included in analysis of sequence basis for enantioselective in IREDs.

IRED Name	Organism of Origin	IRED Name	Organism of Origin
IRED_1	*Streptomyces sp. Mg1*	IRED_11	*Frankia sp. QA3*
IRED_2	*Streptomyces fulvissimus*	IRED_12	*Verrucosispora maris*
IRED_3	*Bacillales*	IRED_13	*Streptomyces sp. CNBO91*
IRED_4	*Kribella flavida DSM 17836*	IRED_14	*Cellulosimicrobium cellulans*
IRED_5	*Streptomyces aurantiacus*	IRED_15	*Paenibacillus ehimensis*
IRED_6	*Mycobacterium smegmatis*	IRED_16	*Streptomyces sp. CNH287*
IRED_7	*Saccharopolyspora*	*SS*-IRED	*Streptomyces sp. GF3546*
IRED_8	*Actinomadura rifamycini*	*SR*-IRED	*Streptomyces sp. GF3587*
IRED_9	*Mycobacterium vaccae*	*BcS*-IRED	*Bacillus cereus*
IRED_10	*Mycobacterium sp. JLS*	*SkR*-IRED	*Streptomyces kanamyceticus*

A range of cyclic imine substrates were selected for this screen, including 5-, 6- and 7-membered rings and number of quinoline derivatives. Chosen substrates are outlined in Figure 5.8.

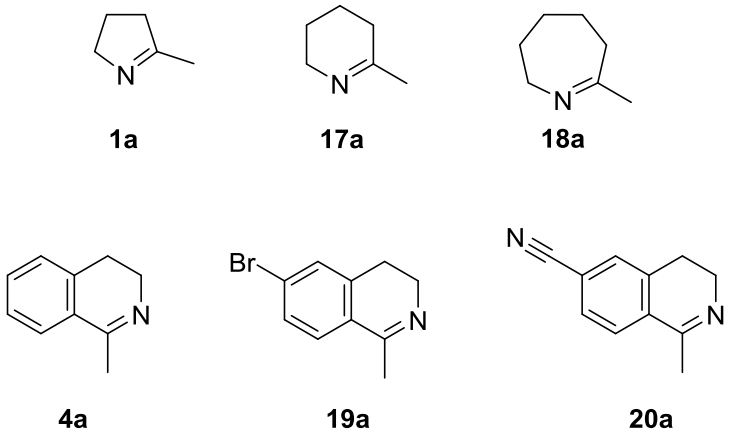


Figure 5.8: Substrates selected for IRED biotransformations in study selected by Wetzl and coworkers.

Enantiopurity of products produced from IREDs from substrates **1a**, **4a** and **9-12a** are recorded in Table 5.3. Some enzymes displayed substrate-dependent enantioselectivity, where the stereochemistry of the enantiomer formed is reversed for given substrates relative to the usual selectivity of the enzyme. For clarity reasons, only uniformly selective IREDs were included in this analysis.

Table 5.3: Enantioselectivity of selected IREDs towards six varying imine substrates. Enantioselectivity values for formation of (*S*)-amines are highlighted in blue and listed with positive values, enantioselectivity values for formation of (*R*)-amines are highlighted in red and listed with negative values

Substrate:	**1a**	**17a**	**18a**	**4a**	**19a**	**20a**
IRED_1	99	98	99	99	96	99
IRED_2	n.d.	99	n.d.	n.d.	11	69
IRED_3	n.d.	n.d.	n.d.	86	37	93
IRED_4	95	85	97	99	98	99
IRED_5	99	99	96	99	97	99
IRED_6	97	98	95	99	98	94
IRED_7	44	97	98	68	90	79
IRED_8	99	99	99	99	18	17
IRED_9	98	99	97	97	97	92
IRED_10	98	99	93	99	98	99
SS-IRED	99	99	96	92	91	98
BcS-IRED*	99	99	n.d.	n.d.	81	n.d.
IRED_11	-95	-90	-99	-99	-98	-92
IRED_12	-98	-69	-88	-88	-21	-6
IRED_13	-97	-81	-73	-85	-33	-56
IRED_14	-80	-82	-25	-89	-98	-90
IRED_15	n.d.	-99	-99	n.d.	-97	-97
IRED_16	-93	-92	-99	-96	-76	-72

Substrate:	**1a**	**17a**	**18a**	**4a**	**19a**	**20a**
SkR-IRED**	-99	n.d.	n.d.	n.d.	n.d.	n.d.
SR-IRED***	-99	-99	n.d.	-99	n.d.	n.d.

n.d. = not determined. This is stated where conversion is low (>25 %) or data are not available (where reaction data was obtained from sources external to Wetzl publication).

* Reaction data obtained from Man and coworkers[78]

**Reaction data obtained from results outlined in chapter 3

***Reaction data obtained from Hussain and coworkers[59]

Some enzymes displayed substrate-dependent enantioselectivity, where the preference for the formation of the opposite enantiomers can be formed dependent on substrate. Based on these results, only enzymes which displayed consistent enantioselectivity were included any subsequence analysis of sequence basis for enantioselectivity (Table 5.4).

Table 5.4: List of IREDs which consistently display >80 % enantioselectivity for the formation of only (*R*)- or only (*S*)-amine for all substrates tested. Enzymes with <80 % enantioselectivity for any substrate, or displayed mixed enantioselectivity (formation of different enantiomers depending on substrate) were omitted.

Strongly (*R*)-selective IREDs*	Strongly (*S*)-selective IREDs*
IRED_11	IRED_1
IRED_15	IRED_4
SkR-IRED	IRED_5
SR-IRED	IRED_6
	IRED_9
	IRED_10
	SS-IRED
	SkR-IRED

Sequences of all IREDs were listed and separated according to (*R*)- or (*S*)-selectivity, where sequences highlighted in blue are (*S*)-selective and sequences highlighted in red are (*R*)-selective (figure 5.3). Sequences of IREDs that displayed consistently at least 80 % ee of products formed from all substrates are listed in black font. Sequences of IREDs that displayed consistently (*R*)- or (*S*)-selectivity but less than 80 % ee for one or more products are listed in grey italics.

	27	34	81	87	108	116	134	143	184	195	222	228	231,232	247	265
>1	LGPMGQAM		LSLTDYR		NLSSDTPDR		GGVMVPAPMV		QLMYQAQLAVFL		LLSSAD-AIG		-----A-ILRAG		GDLSTVTMMGATADHIVET
>2	*LGPMGRAM*		*ISLTDYA*		*NLTSATPEE*		*GGVNSPPSGI*		*ALLYQIGVGMFW*		*ADDTAN-SLT*		*-----Q-FFAFY*		*GDVDRLAMGMASVEHVLRT*
>3	*LGPMGKAI*		*LSLTDYR*		*NLSSDTPEK*		*GGVLASPPGI*		*MLYYQLQIDVFW*		*VSDILS-TMP*		*-----K-LLEFY*		*GDVEKLAMGLASVEHVVQT*
>4	LGPMGQAM		LSLTDYQ		NLSSDTPDR		GGVMIPAPMV		QLMYQAQLDVFL		LIGMLR-TVP		-----A-MLEAG		GDLSTITMMGATADHIVGA
>7	LGPMGQAM		LSLIDYD		NLSSDTPEK		GGVLSPPPGI		ALMYQLNMVVFW		VSENFA-GMG		-----Q-FIDFY		GDVDRLSMGVASMEHVVHT
>8(GF3546)	LGPMGQAM		LSLTDYD		NLSSDTPDK		GGVQVPPPLI		AMYYQAQMTIFW		ATMMTS-MMP		-----H-FLELY		GDVDRLAMGAASVDHVLHT
>10	LGPMGQAL		VNVVDHD		GLSSDTPDR		GAIMTPIDTI		AAFDMALLDLFW		AHGIVG-ILS		-----P-IFTEV		DASASVSSVASSVRHLIAA
>13	*LGPMGYAL*		*VCVADYA*		*NLCSGTPQE*		*GAIMVPVEVI*		*VLHNTALLGLMW*		*AVDWFLPSVT*		*-----GEILRAE*		*GDGGTLAMCLTAIEHIVRT*
>15	*LGPMGKAM*		*ISQTDYK*		*NLSSGSPDE*		*GGVMVPPPGI*		*NLYYQAQLYLFW*		*ATETVT-SLGVDGPMG*		*-FLRIL*		*GGENSMLMMAVGADHMVEA*
>17	LGPMGRAL		INVVDHD		GLSSDTPDR		GAIMTPTDVV		AAYDMALLDVFW		AVGIAA-ILP		-----P-VFSEV		DANATVSSAAASLSHLVAT
>18	QGPMGQAL		VNVVDQA		GLASDTPDT		GAIMTPTDTI		AAFDMALLDLFW		AQGIVG-ILP		-----P-IVDEL		DSRAQVASVAASVRHLIAA
>Bcer	LGPMGQAM		LSLTDYN		NLSSDTPEK		GGVQVPPSGI		MLYYQIQMDIFW		ASAMMS-SLP		-----K-FVEFY		GDVDRLAMGLASVEHVVHT
>9	LGAMGTAL		VCLLDDA		NLTTGSPAQ		GGIMAVPPMV		ALHDVALLSAMT		LVEWLR-AMA		-----T-SASAT		GVMSNLAMQVAGNATLLRT
>11	*LGAMGSAL*		*ICVLDYR*		*NLTSGTPGD*		*GAIMATPSMI*		*SLYDVALLGLMW*		*AQAWFEYVIS*		*-----P-EVPNL*		*DNDSTLGMQTVAIEHLVEA*
>12	*LGAMGSAL*		*TCLTTYE*		*TLNSGSPAG*		*GAIKNVPPAV*		*ALYEMAVGGTLL*		*SEQWLQ-MIA*		*-----S-VLPVL*		*EPLSSVNVFVAGAAHDAEL*
>19	*LGAMGRAL*		*AVLLDHA*		*NLVTTTPEE*		*GGIMAVPGMI*		*PLYDLGLLAGMY*		*AAPWIA-AMT*		*-----A-ELAGY*		*GGPGQQSLEFSDLSDMVRA*
>21	LGMMGSAL		VCVLDYE		NLTNGKPEQ		GGIMAVPQMI		ALYDLALLTTAY		VIPWLQ-AMI		-----A-SLPSQ		TDVSSLNINKVGFVNLIEA
>23	*TGAMGSAL*		*ACLLDDD*		*NLTTGTPGQ*		*GGIMAVPPMI*		*ALHDVALLSAMY*		*LVSWLT-AMA*		*-----P-AALES*		*DVVSNLAMQVAGIPTFLRT*
>Skan	LGLMGQAL		VCVSDYD		NLTSGTSAQ		GAILAGPAAI		SLYDAAGLVMMW		ITQGIG-TVA		-----D-WLPGY		ADDAAIDTHLATMEHLIHE
>SPGF3587	LGLMGQAL		VCLSTYD		NLTSGSSDQ		GAIMITPPGI		ALYDVSLLGLMW		AHMWLE-AIK		-----M-FTADY		ANDATLETHLAALKHLVHE

Figure 5.9: List of sequences which display consistent (*R*)- or (*S*)-enantiopurity of products formed by tested imine substrates, where strongly selective (>80 % ee for all products) IREDs are listed in black and more weakly selective IREDs (<80 % ee for one or more products) are listed in grey italics. Residues are numbered according to *SkR*-IRED sequence positions.

An analysis was carried out on the differences between sequences of (*R*)- and (*S*)-selective IREDs. For the purposes of this analysis, only IREDs which were deemed to be strongly selective (>80 % ee for all products) were included. For each group of IREDs ((*R*)- and (*S*)-selective), the extent to which each residue is conserved was assessed manually. Amino acid residues treated as similar are listed in Table 5.5. Groupings were made based on hydrophobicity, size, charge and aromaticity of amino acid side chains, as these are believed to affect substrate binding. All other residues (H, C, G, and P) were deemed not to have unique properties which meant they are not sufficiently similar to other residues to be placed in any group, and were treated individually.

Table 5.5: Groups of amino acids treated as similar in sequence similarity analysis of enantioselective IREDs. Amino acid side chains are grouped according to hydrophobicity, size, charge and aromaticity.

Amino acid type	Amico acids in group
Amino acids with aromatic side chains	F, Y, W
Amino acids with hydrophobic, aliphatic side chains	V, I, L, A
Amino acids with large polarised, uncharged side chains	N, Q, M
Amino acids with small, polarised, uncharged side chains	S, T
Amino acids with positively charged side chains	K, R
Amino acids with negatively charged side chains	D, E

In each sequence alignment, level of conservation for each position was calculated and highlighted according to the colour code in Table 5.6. Similar amino acid residues were regarded as conserved, and the percentage conservation was calculated as the number of amino acids identical or similar to the most abundant amino acid.

Using this analysis, a consensus sequence for both (*R*)- and (*S*)-selective IREDs was constructed using the most abundant amino acid for each position between sequences. Where more than one amino acids were equally highly abundant, the least

functionalised amino acid was used in the consensus sequence (for example, Ala would be used in the consensus sequence instead of Val, if the two were equally the most abundant).

Table 5.6: Key describing degree of conservation correlating to colour used to highlight residues in a sequence alignment

Degree of conservation	Colour code
0 – 19 %	
20 – 39 %	
40 – 59 %	
60 – 79 %	
80 – 99 %	
100 %	

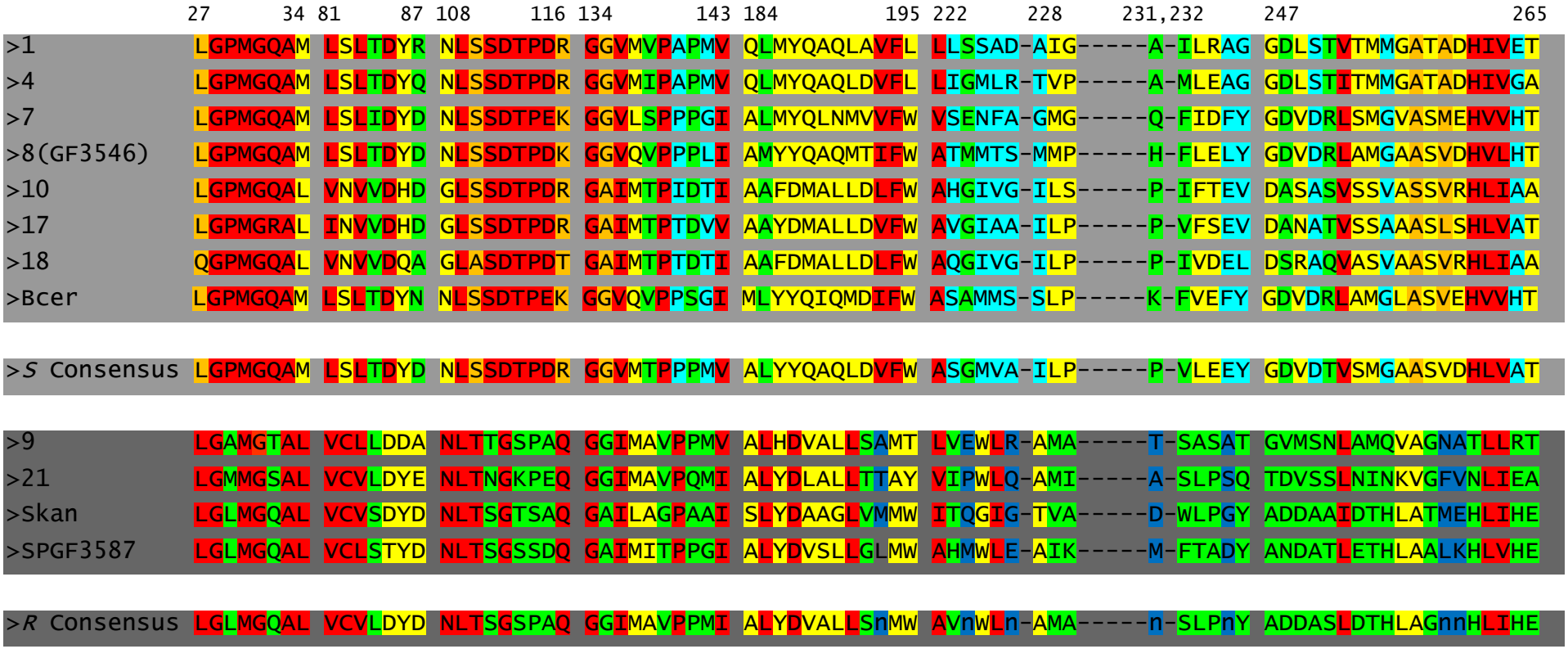


Figure 5.10: An alignment of sequences of strongly (*S*)-selective (blue) and (*R*)-selective (red) IREDs with amino acid residues colour coded according to level of conservation between equivalent positions in similarly selective IREDs according to the key in Table 5.5. The most highly conserved residues within each group were used to form a consensus sequence for (*S*)- and (*R*)-selective IREDs.

(*R*)- and (*S*)-selective consensus sequences were inspected for differences. Attention was drawn to residues that were highly conserved within one group, but were different or non-conserved in the opposite group. The most notable differences occur at amino acids Pro29, Asp112, Lys116 and Phe194 which are strongly conserved in (*S*)-selective IREDs but not present in (*R*)-selective IREDs, and residues Cys82, Gly112 and Gln116 which are strongly conserved in (*R*)-selective IREDs but not (*S*)-selective IREDs. Some of these strongly conserved residues are reciprocal (positions 112 and 116), where each group of IREDs had strong conservation of different residues at these positions. Other residues were strongly conserved in one group of IREDs but had no particular consensus amino acid at the equivalent position in the other group. The results of these findings are outlined in Table 5.7.

Table 5.7: An outline of amino acid residues conserved in active region of (*R*)- and (*S*)-selective IREDs.

Amino Acid Position	Amino acid present in (*R*)-seletive IREDs (amino acid, percentage conservation)	Amino acid present in (*S*)-selective IREDs (amino acid, level of conservation)
29	No strong conservation	Pro, 100 %
82	Cys, 100 %	No strong conservation
112	Gly, 100 %	Asp, 100 %
116	Arg, 88 %*	Gln, 100 %
194	Phe, 100 %	No strong conservation

5.4 Discussion

As there are known sequences for more weakly selective IREDs with (*R*)- and (*S*)-selectivity (<80 % product ee formed for one or more substrates, sequences in grey italics in figure 5.4), it is possible to provisionally inspect whether the absence of these conserved residues make affect enantioselectivity.

At position 29, there are no strongly or weakly (*S*)-selective sequences that do not have the conserved Pro residue. Equally, Cys82 is conserved among all (*R*)-selective IREDs. This could perhaps be because these residues are essential for (*S*)- and (*R*)-selectivity, respectively, or these could simply be artefacts of evolution within (*R*)- and (*S*)-selective IREDs.

At other positions, however, residues that were highly conserved in strongly selective IREDs were not always conserved in weakly selective IREDs. At position 112, strongly (*S*)-selective IREDs possess a highly conserved Asp residue, whereas strongly (*R*)-selective IREDs have a strongly conserved Gly residue. Out of the four more weakly (*S*)-selective IREDs, three do not have the conserved Asp112 residue (IRED2, IRED13 and IRED15), but instead either a Gly or Ala residue. Conversely, one of four weakly (*R*)-selective IREDs possessed a Thr residue in place of the Gly present in strongly (*R*)-selective IREDs.

At position 116, all strongly (*S*)-selective IREDs have a highly conserved Arg (or similar Lys) residue. In the four weakly (*S*)-selective IREDs, three instead have a Glu residue (IRED2, IRED7 and IRED8).

At position 194, strongly (*S*)-selective IREDs have a highly conserved Phe residue. In one of the four weakly (*S*)-selective IREDs (IRED13), this is replaced by a Met residue. A Met residue at this position is found in 75 % of the strongly (*R*)-selective IREDs, suggesting that mutation of this residue to Met could result in weakened (*S*)-selectivity.

These observations suggest that Gly112, Arg (or Lys)116 and Phe194 may play a role in directing (*R*)-selectivity, as there are examples of IREDs without these residues whose enantioselectivity is compromised. Residues Asp112 and Gln116 may assist in

directing (*S*)-selectivity, as some IREDs where these residues were not conserved also displayed weakened enantiopreference.

However, while IREDs without these residues sometimes displayed weakened enantioselectivity, this may not indicate that these residues are directly responsible for this effect. Evidence would be required from mutagenesis experiments that, for more strongly selective IREDs, removal of these residues results in weakened enantioselectivity. The opportunity to conduct these mutagenesis experiments is prohibited by time limitations in this project, but would be a valuable future experiment.

Finally, it is important to make note that, aside from IRED7, all (*R*)- and (*S*)-selective IREDs studied in this section exhibited consistently high enantioselectivity towards substrate **1a**. Only the larger substrates **10a, 11a** and **12a** exposed the weakened enantioselectivity of the less selective IREDs. This could suggest that it is the steric bulk surrounding the imine moiety interacting with the enzyme active sites in such a way that enantioselectivity can be reversed, given certain sequence differences.

5.5 Conclusion

IREDs are commonly highly enantioselective enzymes, yet the reason for this is not currently known. This chapter outlines investigations into the basis for the enantiospecificity of IREDs.

Firstly, comparison of the amine binding sites of (*R*)-selective *SR*-IRED and (*S*)-selective *SS*-IRED revealed five residues which were not conserved between the two enzymes. In order to assess whether these residue differences were responsible for the opposing enantioselectivities of these IREDs, reciprocal mutants were made for both enzymes at each site (*SR*-IRED L176M, M179F, L183M, H242G and Y283F and *SS*-IRED M173L, F176M, M180L, G237H and F278Y). Soluble expression was achieved for seven out of ten of these mutants (all except *SR*-IRED L176M and *SS*-IRED M173L and M180L). Soluble mutants were expressed and purified and used in biotransformations of imine **1a**. The enantiomeric configuration of amine products were assessed using chiral GC. The findings demonstrated that these point mutations did not alter enantioselectivity relative to the wild-type enzymes. This suggests that these residues do not individually direct enantioselectivity. This could mean that other residues are instead responsible, or that more broad structural changes are necessary to alter enantioselectivity.

Secondly, a publication reporting the sequences of 19 (*R*)- or (*S*)-selective IREDs was analysed for further sequence-based differences between IREDs with opposite enantioselectivities. A consensus sequence of IREDs which formed amines with consistently >80 % ee across the full range of substrates tested in this study was constructed. The consensus sequence of (*R*)- and (*S*)-selective IREDs was compared and a number of sequence differences were observed. Cys82, Gly112, Arg116 and Phe194 was strongly conserved in (*R*)-selective IREDs but not present in selective IREDs. Conversely, Pro29, Asp112 and Gln116 were strongly conserved in (*S*)-selective IREDs but not in (*R*)-selective IREDs. By observing sequences of IREDs which were still consistently (*R*)- or (*S*)-selective, but whose enantiopreferences were not as strong as those used in the construction of the consensus sequences, it was possible to see that the absence of Gly112, Arg116 and Phe194 in (*R*)-selective IREDs and Asp112 and Gln116 in (S)-selective IREDs could result in reduced enantiopreference,

suggesting that these residues may contribute towards directing enantioselectivity in IREDs.

6. Characterization of an *Pp*IRED: an Imine Reductase from *Physcomitrella patens*

6.1 Introduction

This chapter outlines the results from an industrial collaboration with GlaxoSmithKline (GSK).

A broad screen of novel IREDs towards a range of substrates was carried out by collaborators at GSK. The screen involved activity tests towards both imine and amine substrates in the reductive and oxidative directions respectively. During this screening effort, an IRED from *Physcomitrella patens* (*Pp*IRED) displayed markedly different substrate selectivity to other IREDs. It was active towards substrates towards which no other IREDs had been reported to work, either during this screening effort or previously in literature.

The distinctly different substrate scope of *Pp*IRED means that it is complementary to existing IREDs which catalyse the transformation of other substrates. It would therefore be of interest to characterize this enzyme further.

This chapter outlines the characterization of *Pp*IRED in terms of its substrate scope and 3D structure as determined using X-ray Crystallography. This provides the first opportunity to characterize an IRED deriving from a plant organism as well as an IRED which so far operates exclusively on a unique set of substrates.

6.2 Aims

Following interesting preliminary results suggesting a unique substrate scope for *Pp*IRED, a fuller characterization of this IRED will be carried out. The crystal structure will be determined, including a cofactor complex, in order to gain a mechanistic understanding of the enzyme. Furthermore, the reactivity profile of *Pp*IRED will be carried out and detailed kinetic analysis will be performed for each reaction in order to ascertain the potential applications of the enzyme. GC analysis will be used to confirm the activity of *Pp*IRED in amine oxidation reactions to confirm its identity as an IRED.

6.3 Results

6.3.1 Expression and purification of *Pp*IRED

*Pp*IRED was expressed at 2 L scale in LB medium in shake flasks and purified by nickel-affinity chromatography and size exclusion. Further purification was carried out by size exclusion chromatography (Figure 6.1). After each purification step, fractions were analysed by SDS-PAGE (Figures 6.2 and 6.3).

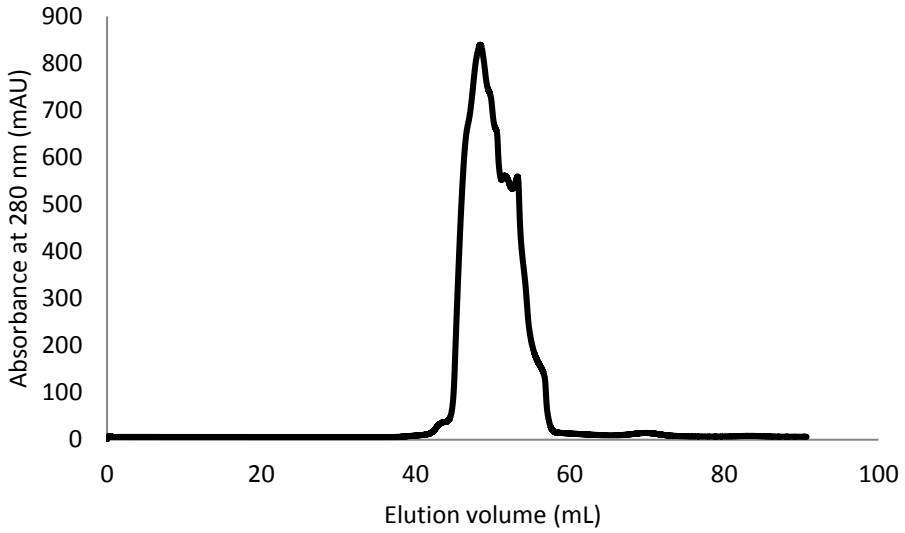


Figure 6.1: Chromatogram recorded during size exclusion chromatography purification of *Pp*IRED. *Pp*IRED is eluted in a broad peak between approximately 45 – 59 mL elution volume. The peak exhibits some shoulders which may have arisen due to irregular protein concentration caused by dilution during loading.

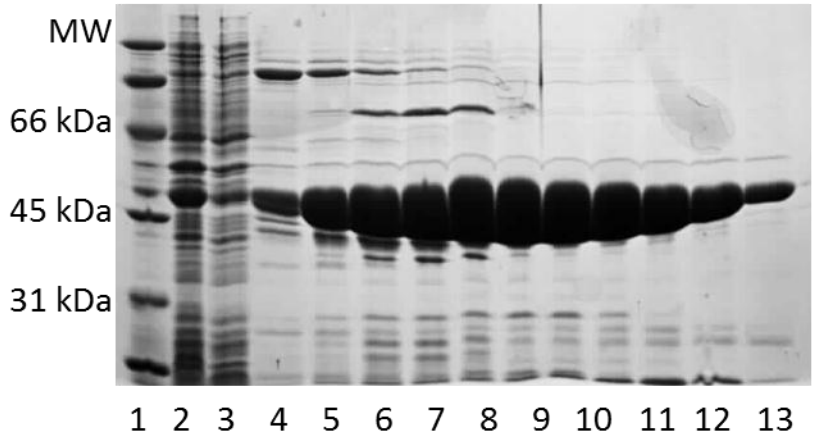


Figure 6.2: SDS-PAGE analysis of fractions collected during nickel-affinity chromatography purification of *Pp*IRED.

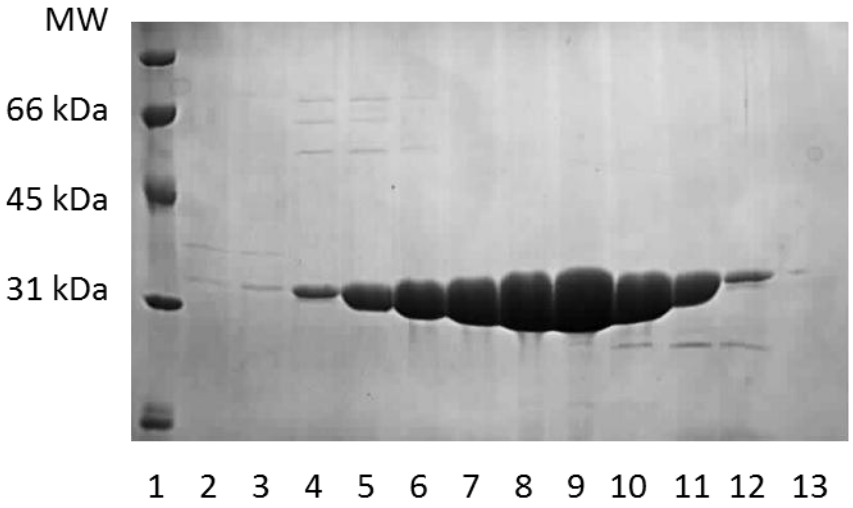


Figure 6.3: SDS-PAGE analysis of fractions collected during size exclusion chromatography purification of *Pp*IRED.

Pure protein was pooled and concentrated to 50 mg ml^{-1} prior to being flash frozen and stored at -20 °C for subsequent use.

6.4 Crystal structure of *Pp*IRED in *apo* form

Crystallization

Purified *Pp*IRED was subjected to the commercial crystallization screens INDEX (Hampton Research) and CSS (Molecular Dimensions) in a 96 well sitting drop format. The protein was readily crystallized, with a large number of crystals forming in a range of conditions in the INDEX and CSS screens within 24 hours. The highest quality crystals were obtained in the INDEX screen condition 0.2 M NaCl, 0.1 M BIS-TRIS pH 6.5, 25 % w/v PEG 3350 (Figure 6.4).

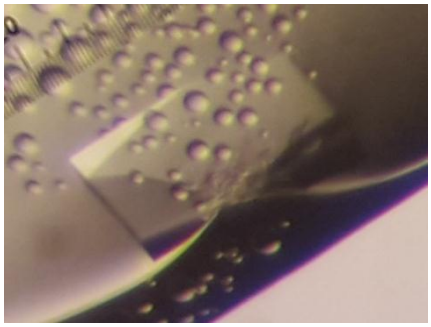


Figure 6.4: A single cystal of *Pp*IRED in its *apo* form in the INDEX screen in 0.2 M NaCl, 0.1 M BIS-TRIS pH 6.5, 25 % w/v PEG 3350.

The condition was repeated in 48- and 96-well scale without success. Therefore, crystals from the initial 96-well crystallization screen were used for diffraction experiments.

Structural solution

The structure of *Pp*IRED was solved using a molecular replacement strategy using the glyoxylate reductase 1 (GLYR1) enzyme from *Arabidopsis thaliana*, which was the protein for which *Pp*IRED shared the highest level of protein sequence similarity of all proteins whose structures had been deposited in the PDB.

GLYR1 from *Arabidopsis thaliana*, which will hereon be referred to by its PDB accession code, 3DOJ, shares 68.0 % protein sequence identity and 88.7 % protein sequence similarity with *Pp*IRED, suggesting a very closely related enzyme.

GLYR1 type enzymes such as 3DOJ are one of two possible isoforms of GLYR in plants (GYLR1 or GLYR1).[79] GLYR enzymes are found in a plant's cytoplasm and catalyse the NADH- or NADPH-dependent reduction of glyoxylate, a photorespiration metabolite, to glycolate. High levels of glyoxylate in plants is an indication of stress and glyoxylate can cause damage in the plant by reacting with DNA, cell membrane lipids and plant proteins as well as by increasing transcription of stress-related genes.[80] This highlights the importance of 3DOJ and other GLYR1s as detoxifying enzymes in their removal of harmful excesses of glyoxylate.

```
PpIRED      MGSDGMDIGFLGLGIMGIAMARNLLKKGFNVTVWNRSPGKCEELAKEGASIGSSPADVVK
3DOJ        ----MMEVGFLGLGIMGKAMSMNLLKNGFKVTVWNRTLSKCDELVEHGASVCESPAEVIK
                 *::********** **: ****:**:******: .**:**.:.***: .***:*:*

PpIRED      KCSITIAMLADPSVAISVALGPGGVVEGITPGKGYVDMSTVDSATSSQIAKAIEAKGGDF
3DOJ        KCKYTIAMLSDPCAALSVVFDKGGVLEQICEGKGYIDMSTVDAETSLKINEAITGKGGRF
            **. *****:**..*:**.:. ***:* *  ****:******: ** :* :** .*** *

PpIRED      LEAPVSGSKQPAEAGTLIILAAGSESLFAKVKPAFDAMGKKSFFLNDVGSGAKMKLVVNM
3DOJ        VEGPVSGSKKPAEDGQLIILAAGDKALFEESIPAFDVLGKRSFYLGQVGNGAKMKLIVNM
            :*.******:*** * *******.::** :  ****.:**:**:*.:**.******:***

PpIRED      IMGSMMTAFSEGLALADKAGLSQQTLLEVLELGAIGNPMFKLKGPSMISGKCPTAFPLKH
3DOJ        IMGSMMNAFSEGLVLADKSGLSSDTLLDILDLGAMTNPMFKGKGPSMNKSSYPPAFPLKH
            ******.******.****:***.:***::*:***: ***** ***** ... * ******

PpIRED      QQKDMRLALALGDEVGQSMPVAAAANEIYKKARSLGLGEQDFSSVHQALTQEAANASDSA
3DOJ        QQKDMRLALALGDENAVSMPVAAAANEAFKKARSLGLGDLDFSAVIEAVKFSRE------
            ************** . ********** :*********: ***:* :*:. .

PpIRED      KAAA
3DOJ        ----
```

Figure 6.5: Protein sequence alignment of *Pp*IRED and 3DOJ (glyoxylate reductase from *Arabidopsis thaliana*).
Both enzymes possess the GxGxxG binding motif, which presents as GLGIMG in both *Pp*IRED and 3DOJ.

The unit cell of *Pp*IRED determined using a molecular replacement strategy with 3DOJ contained two monomers (A and B) arranged as a single homodimer. A monomer of *Pp*IRED is represented by subunit A in Figure 6.6.

The monomer of *Pp*IRED possesses several structural features which are consistent with other IREDs including *SS*-IRED and *SkR*-IRED;[60, 61, 78] *Pp*IRED consists of an *N*-terminal Rossman fold domain (Met1 – Val165) and a *C*-terminal helical domain (Gly200 – Ala304) connected by a long interdomain helix (Gly169 – Ala199). This is

described visually by a secondary structure sequence annotation in Figure 6.7. The data collection and refinement statistics are outlined in Table 6.1.

Table 6.1: Data Collection and Refinement Statistics for *apo-Pp*IRED

Date	14-02-16
	9262
Beamline	I24
Wavelength (Å)	0.96856
Resolution (Å)	50.06-2.53 (2.64-2.53)
Space Group	*I*222
Unit cell (Å)	a = 82.5, b = 87.2; c = 182.3
	$\alpha = \beta = \gamma = 90°$
No. of molecules in the asymmetric unit	2
Unique reflections	22294 (2665)
Completeness (%)	99.9 (100.0)
R_{merge} (%)	0.16 (0.73)
$R_{p.i.m.}$	0.09 (0.42)
Multiplicity	7.3 (7.3)
$\langle I/\sigma(I) \rangle$	10.0 (3.2)
$CC_{1/2}$	0.99 (0.85)
Overall *B* factor from Wilson plot (Å^2)	22
R_{cryst}/ R_{free} (%)	19.5/2396
r.m.s.d 1-2 bonds (Å)	0.015
r.m.s.d 1-3 angles (^{o})	1.91
Avge main chain B (Å^2)	37
Avge side chain B (Å^2)	38
Avge water B (Å^2)	23

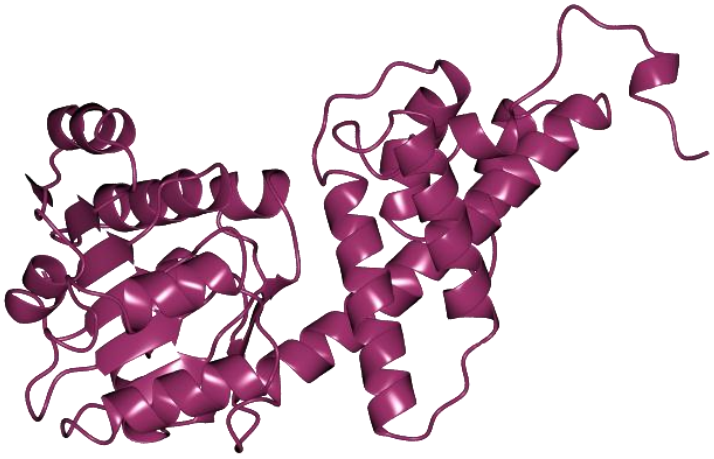


Figure 6.6: Crystal structure of a single monomeric subunit of *Pp*IRED, consisting of an *N*-terminal Rossman fold motif and a *C*-terminal helical domain

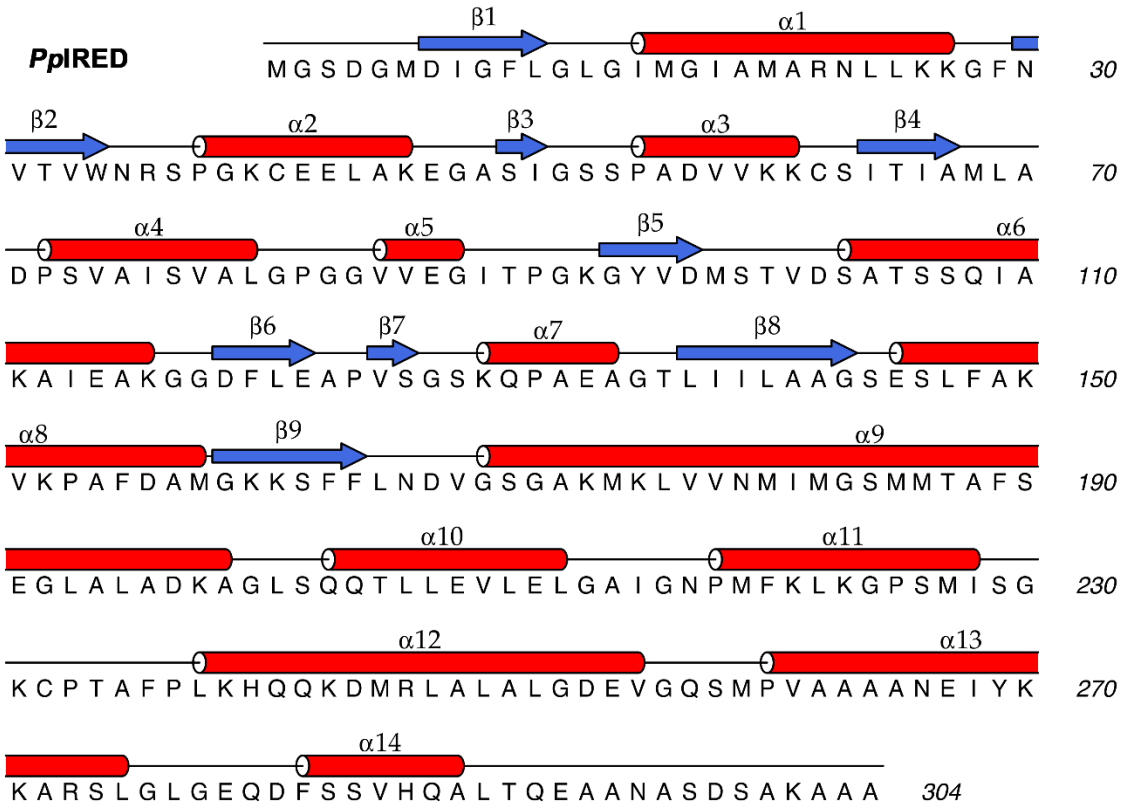


Figure 6.7: Secondary structure sequence annotation for PpIRED. The figure depicts that the structure consists of an N-terminal Rossman fold motif, C-terminal helical domain connected by a long interdomain helix.[71-74]

However, *Pp*IRED differs structurally to other IREDs in the orientation of its *N*- and *C*-terminal domains. While in previously studied IREDs such as *SkR*-IRED and *SS*-IRED, the *C*-terminal helical domain protrudes outwards from the inter-domain helix, in *Pp*IRED this *C*-terminal domain reverts backwards towards the *N*-terminal domain of the same monomeric subunit.

This has a strong impact in the nature of dimer formation. In IREDs such as *SkR*-IRED and *SS*-IRED, the protrusion of the *C*-terminal domain away from the *N*-terminal domain and inter-domain helix causes the formation of a large cleft through which the opposite subunit binds during dimer formation. This phenomenon is responsible for the characteristic reciprocal domain sharing observed in many IREDs.[60, 61, 78] However, in *Pp*IRED the closer proximity of the *N*- and *C*-terminal domains of the same subunit means this phenomenon is not observed, and dimer formation occurs through the association of two monomers without interlocking of domains (Figure 6.8). The nature of dimer formation exhibited by *Pp*IRED is common in enzymes such as 3DOJ and HIBDH (Chapter 2) which catalyse the reversible reduction of carbonyls.

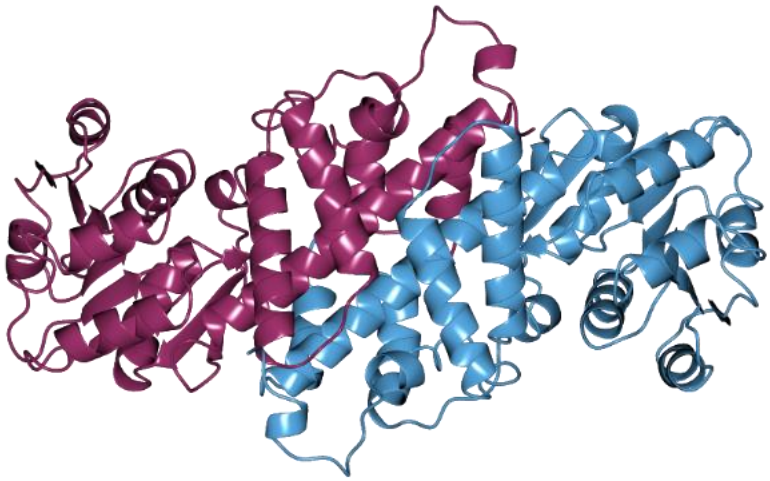


Figure 6.9: A dimer of *Pp*IRED formed by the association of two monomeric subunits A and B. The dimer does not exhibit any domain sharing effects which are observed in other IREDs.

An analysis of the structure using the EBI PISA tool revealed that the contact surface area of the interface between subunits A and B was 2681 Å^2.[75, 76] This is significantly lower than the ~3800 Å^2 interface areas exhibited by other IREDs such as *SkR*-IRED and *SS*-IRED and reflects the less intimate nature of domain sharing in *Pp*IRED relative to other IREDs.

The structure of subunit A of the *Pp*IRED monomer was submitted to the DALI server to search for the most structurally similar protein. The protein with the highest level of structural homology was 3PDU; the NADP$^+$ bound form of GLYR1 from *Arabidopsis thaliana*, which had a rmsd of 0.7 Å relative to *Pp*IRED. This result is unsurprising given the very high level of sequence homology between the two enzymes, and also strongly suggests that the two enzymes share similar physiological roles.

6.5 Crystal structure of *Pp*IRED in complex with NADP$^+$

Crystallization

In order to gain an insight into the active site mechanism of *Pp*IRED, a structure of the enzyme in complex with its putative cofactor, NADPH, was sought. Pure, concentrated *Pp*-IRED protein was incubated with either NADPH or NADP$^+$ with final cofactor concentrations ranging from 0.5 – 10 mM and subjected to a range of commercially available 96-well crystallization screens in the sitting drop format.

Crystals were obtained in numerous crystallization conditions with all cofactor concentrations but diffraction experiments at Diamond synchrotron demonstrated that a complex was only formed with crystals grown in 0.2 M L-proline with *Pp*-IRED incubated with 10 mM NADP$^+$ in the Hampton INDEX screen (Figure 6.10). All other crystallization conditions, including those with NADPH and lower concentrations of NADP$^+$ resulted in *apo* complexes.

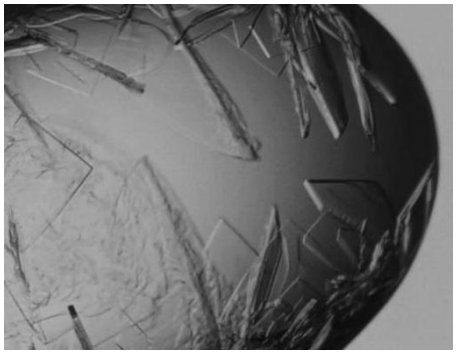


Figure 6.10: Crystals of *Pp*-IRED in complex with $NADP^+$ formed using *Pp*-IRED incubated with 10 mM $NADP^+$ in 0.2 M L-proline in the Hampton INDEX screen in a 96-well sitting drop format.

Crystal structure

The crystal structure of *Pp*-IRED in complex with $NADP^+$ was solved with a molecular replacement strategy using the *apo* structure of *Pp*-IRED. The unit cell contained 20 subunits (A – J) arranged as 10 dimers, each containing the $NADP^+$ cofactor. The structure is represented by subunits A and B in Figure 6.11. The data collection and refinement statistics are outlined in Table 6.2.

Table 6.2: Data collection and refinement statistics for *Pp*IRED in complex with $NADP^+$

Date	08-05-16
	8786
Beamline	i02
Wavelength (Å)	0.97949
Resolution (Å)	65.49-2.50 (2.56-2.50)
Space Group	$P2_1$
Unit cell (Å)	a = 106.2, b = 143.3; c = 215.5
	$\alpha = \gamma = 90°$; $\beta = 104°$;
No. of molecules in the asymmetric unit	20
Unique reflections	213205 (15837)
Completeness (%)	98.6 (99.0)
R_{merge} (%)	0.16 (0.81)
$R_{p.i.m.}$	0.14 (0.70)
Multiplicity	2.0 (2.2)
$<I/\sigma(I)>$	6.7 (1.5)
$CC_{1/2}$	0.980 (0.55)
Overall B factor from Wilson plot ($Å^2$)	42
R_{cryst}/ R_{free} (%)	24.6/27.4
r.m.s.d 1-2 bonds (Å)	0.012
r.m.s.d 1-3 angles (°)	1.88
Avge B ($Å^2$)	56
Avge water B ($Å^2$)	39
Avge ligand B ($Å^2$)	56

The structure shows that, similarly to other IREDs, NADPH binds in a cleft formed between the *N*- and *C*-terminal domains. However, unlike in other IREDs where $NADP^+$ binds between domains of opposite subunits, in *Pp*IRED the cofactor binds between domains of the same subunit. This is as a result of the lack of reciprocal domain sharing in *Pp*IRED which is observed in other IREDs.

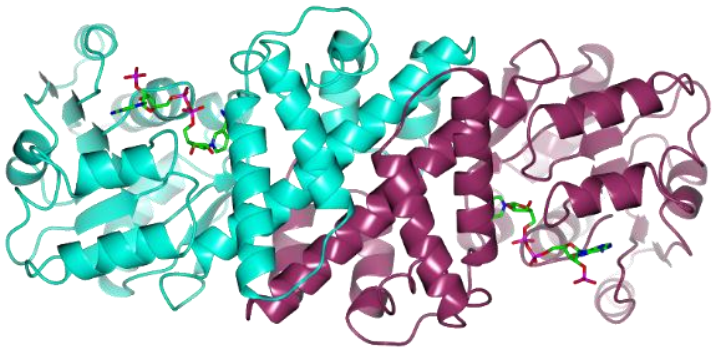
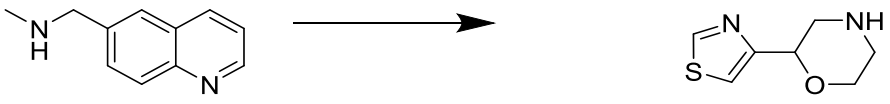


Figure 6.11: The crystal structure of *Pp*IRED in complex with $NADP^+$. The cofactor binds in a cleft formed between the *N*- and *C*-terminal domains of a single subunit.

6.6 Kinetic analysis of *Pp*IRED activity towards amines, ketones and alcohols

During a screen of NADPH concentration assays to test for oxidation activity towards a range of imine and amine substrates.

*Pp*IRED was shown to be uniquely active towards amines **11a** and **12a** during an activity screen carried out by GlaxoSmithKline (Figure 6.12).



11a **12a**

Figure 6.12: Amine substrates which are oxidised to the corresponding imine by *Pp*IRED with the cofactor $NADP^+$. These were identified as substrates for *Pp*IRED during $NADP^+$ depletion assays carried out by GlaxoSmithKline.

To fully characterize the reactivity profile of *Pp*IRED, the enzyme was tested on a range of substrates using a UV-vis based NADPH concentration assay to test for the presence of activity and to determine values for kinetic parameters.

Initially, the activity of *Pp*IRED towards the amines **11a** and **12a** (Figure 6.12). As *Pp*IRED shares high levels of sequence identity with 3DOJ, a glyoxylate reductase which catalyses the reversible reduction of C=O bonds, *Pp*IRED was also tested towards several carbonyl-and hydroxyl-containing substrates; **13a** and **14a** (the aldehyde and alcohol analogues of **11a**), **15a** (hydroxyisobutyrate) and **16a** (glyoxylate – the substrate for the most closely related enzyme to *Pp*IRED, 3DOJ).

Kinetic parameters were determined by measuring the activity towards each substrate at increasing substrate concentrations. Parameters were calculated by fitting the data using non-linear regression to the Michaelis-Menten model or, where substrate inhibition is present, the substrate inhibition model.

The outcome of kinetic analyses towards substrates **11a** – **16a** is pictured below in Figures 6.13 – 6.17. The values for kinetic parameters K_M, k_{cat}/K_M and, where substrate inhibition is present, K_I are listed in Table 6.3.

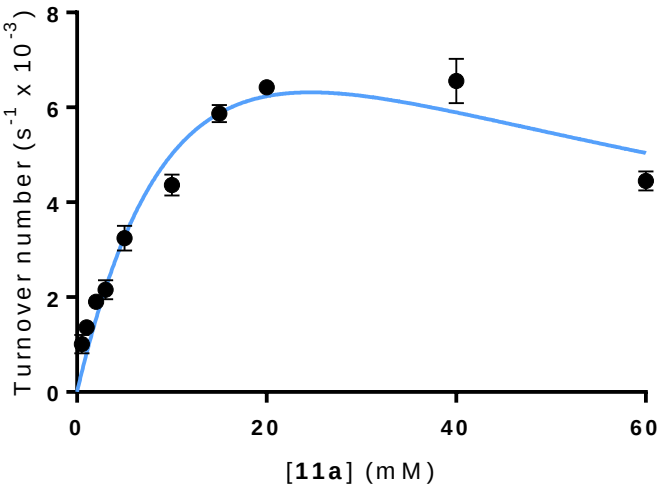


Figure 6.13: Non-linear regression analysis of activity levels of *Pp*IRED towards increasing concentrations of **11a** fitted using the substrate inhibition model. The presence of substrate inhibition is visible from the marked decrease in reaction velocity at the highest substrate concentration (60 mM). This conclusion is confirmed in a quantifiable way using an extra sum of squares F test (F = 25.07, P = 0.0001).

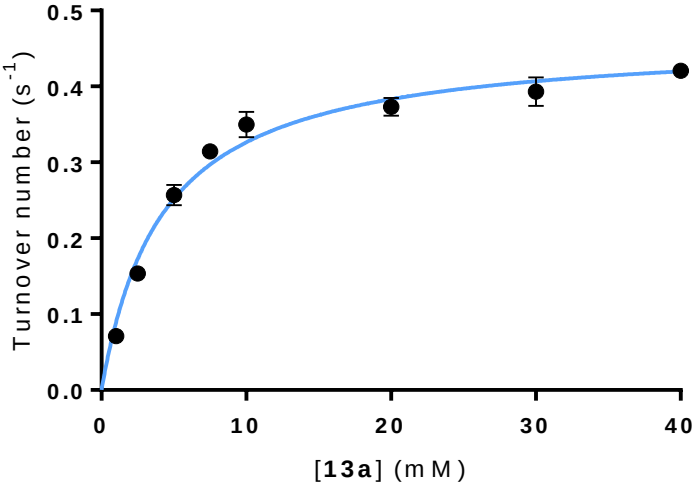


Figure 6.14: Non-linear regression analysis of reaction velocity of *Pp*IRED towards increasing concentrations of substrate **13a** using the Michaelis-Menten model.

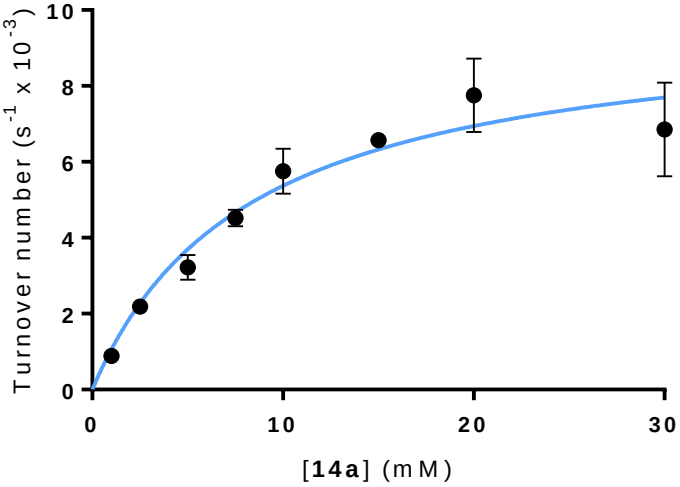


Figure 6.15: Non-linear regression analysis of reaction velocity of *Pp*IRED towards increasing concentrations of substrate **14a** using the Michaelis-Menten model.

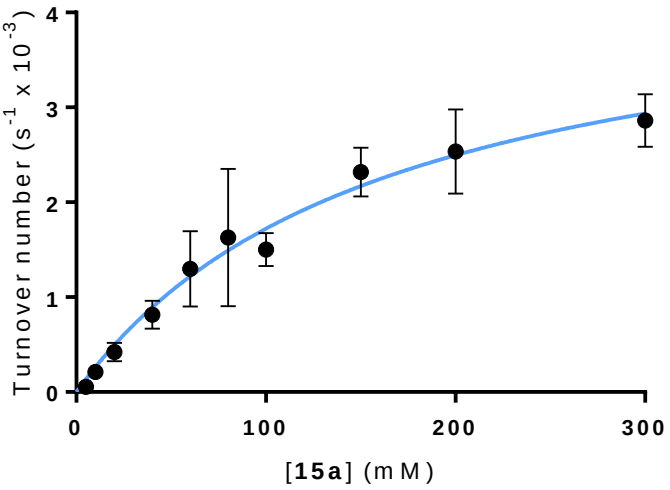


Figure 6.16: Non-linear regression analysis of reaction velocity of *Pp*IRED towards increasing concentrations of substrate **15a** using the Michaelis-Menten model.

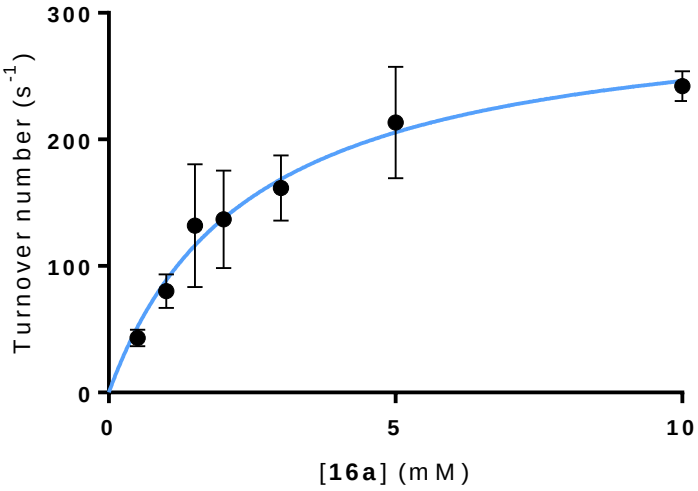
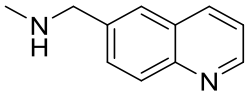
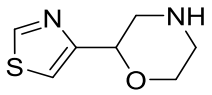
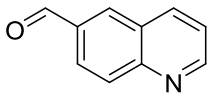
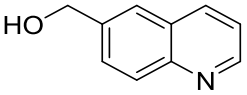
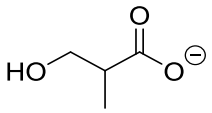
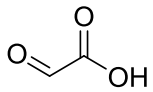


Figure 6.17: Non-linear regression analysis of reaction velocity of *Pp*IRED towards increasing concentrations of substrate **16a** using the Michaelis-Menten model.

Table 6.3: Values for kinetic parameters calculated from activity of *Pp*IRED towards a range of amine, ketone and aldehyde-containing substrates (**11a** – **16a**).

Substrate	Chemical Structure	k_{cat} (s^{-1})	K_M (mM)	K_I (mM)	k_{cat}/K_M (s^{-1} mM^{-1})
11a		$1.577 \times 10^{-2} \pm 0.484 \times 10^{-2}$	18.44 ± 8.04	32.92 ± 16.97	8.55×10^{-4}
12a		n.d.*			
13a		0.464 ± 0.012	4.22 ± 0.41	-†	0.110
14a		$8.30 \times 10^{-3} \pm 1.94 \times 10^{-3}$	9.83 ± 0.90	-†	4.23×10^{-3}
15a		$4.53 \times 10^{-3} \pm 0.63 \times 10^{-3}$	163.1 ± 43.5	-†	2.78×10^{-5}
16a		306.8 ± 32.3	2.46 ± 0.63	-†	124.72

* No activity was recorded for substrate **12a** due to limited substrate availability

† K_I not calculated unless substrate inhibition model used

Activity was observable for *Pp*IRED in the oxidation of amine **11a** with a modest k_{cat}. The high K_M of 18.44 mM suggests that, while **11a** is a substrate for *Pp*IRED, the affinity of the enzyme towards **11a** is weak. The K_I of 32.92 mM also indictates strong levels of substrate inhibition, so the maximum reaction velocity reported by the k_{cat} is not realised under real reaction conditions.

No activity was detectable towards amine **12a**. However, as very limited quantities of substrate were available, it was neither possible to test at higher substrate concentrations (> 10 mM) nor to retest at higher enzyme concentrations. It is possible that activity may be observed if these experiments are later performed.

*Pp*IRED is strongly active towards **13a** with a k_{cat} of 0.464 s^{-1}. The reductive, rather than oxidative nature of this reaction may contribute towards this increased activity, as reductions generally proceed faster than oxidations in IREDs.

*Pp*IRED also displays modest levels of activity towards alcohol **14a**. Because this is also an oxidation reaction, kinetic parameters for this reaction may be directly compared to the oxidation of **11a**. Interestingly, *Pp*IRED displays a higher catalytic turnover number for **11a** than **14a** which initially suggests a preference for amines over hydroxyls. However, the presence of substrate inhibition in reactions with **11a** makes this comparison difficult, as the theoretical maximum turnover number is never achieved due to decreasing reaction velocities at higher substrate concentrations. Additionally, the K_M of *Pp*IRED is approximately twofold higher for **11a** than **14a**, indicating a tighter binding preference for the alcohol- rather than the amine-containing substrate.

Finally, *Pp*IRED displays activity towards **16a** which is many magnitudes greater than for any other substrate tested. This result can be expected as the closest homologue to *Pp*IRED in both sequence and structure is the glyoxylate reductase 3DOJ, for which **16a** is a natural substrate.

In summary, *Pp*IRED does display modest activity in the oxidation of amine **11a** which supports the identification of *Pp*IRED as an IRED. However, the equivalent imine is not available for testing. *Pp*IRED displays the greatest activity in the reduction of carbonyls, particularly **16a**, which may be a natural substrate for *Pp*IRED.

6.7 GC analysis of biotransformations of quinoline derivatives by *Pp*IRED

To quality control findings from Section 6.6 that *Pp*IRED displays oxidative activity towards amine **11a**, a GC assay was carried out. This was to measure the strength of the evidence that *Pp*IRED acts as an IRED. While imine **11b** is not commercially available and therefore directly testing imine reduction is not possible, the assumption was made that if *Pp*IRED successfully oxidises an amine with $NADP^+$, then the reverse reaction (imine reduction using NADPH) would also be possible. As **11a** was the only amine substrate for which activity was confirmed using an NADPH-formation assay, and the corresponding imine is not commercially available, GC analysis was used to verify these findings.

Samples of reactions containing *Pp*IRED, **11a** (5 mM) and $NADP^+$ were collected periodically and analysed by GC.

GC chromatograms of standards containing the substrate **11a** and the internal standard pentadecane were recorded to determine their retention times (8.03 min and 4.33 min respectively) (Figure 6.18 and 6.19). This allowed the substrate and internal standard peaks to be identified in subsequent chromatograms of reactions using *Pp*IRED. No standard could be recorded for the imine product, **11b**, as this compound was not commercially available.

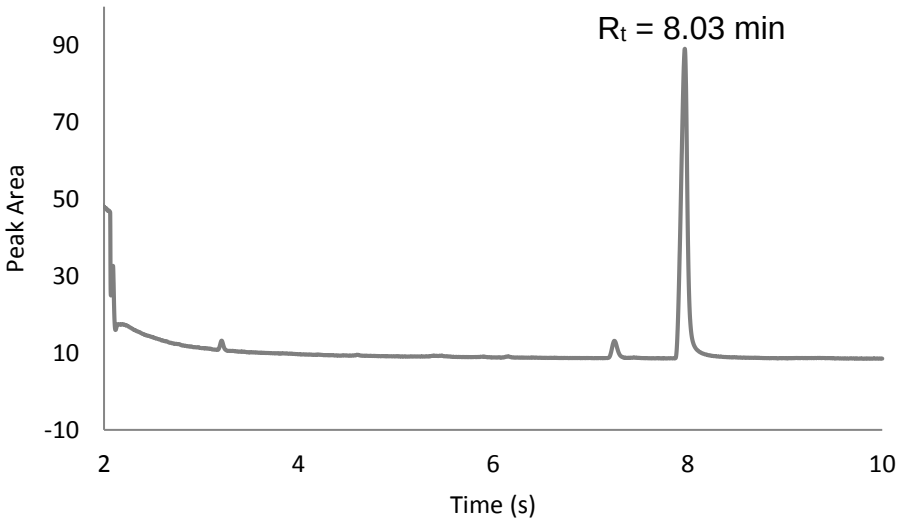


Figure 6.18: Chromatogram recorded from GC analysis of sample containing commercially prepared **11a** (5 mM) in EtOAc

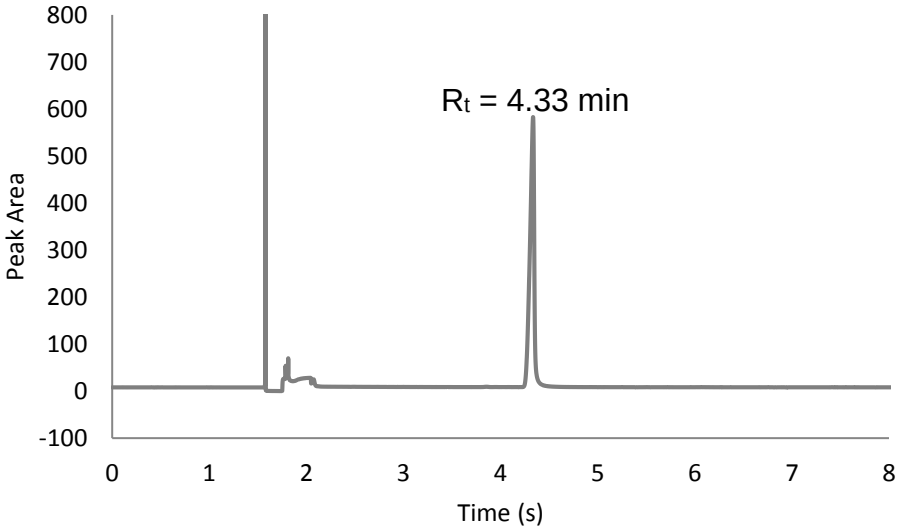


Figure 6.19: Chromatogram recorded from GC analysis of sample containing commercially prepared pentadecane (5 mM) in EtOAc. Pentadecane acts as an internal standard in the analysis.

Samples drawn from the reaction at time points (0 min, 20 min, 40 min, 1 h, 2h, 4h and 24h) were analysed by GC. Analysis demonstrated that, after 24 h of reaction time, there was no change in relative peak area for substrate **11a** (Figure 6.20). The reaction was therefore repeated at increased enzyme concentration (5 mg mL^{-1}).

However, again there was no reduction in peak area for substrate **11a** and no formation of any apparent product peaks (Figure 6.21).

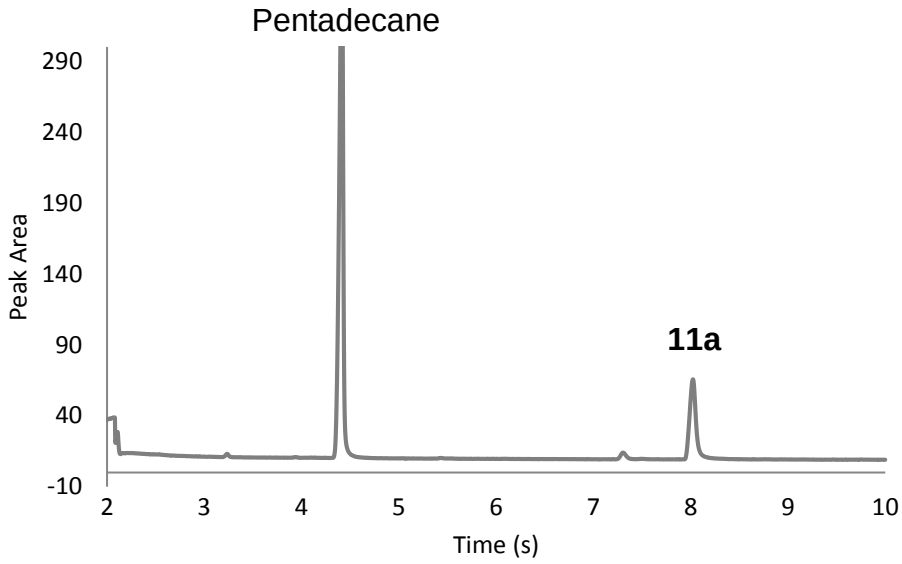


Figure 6.20: Chromatogram recorded from GC analysis of samples drawn from a reaction of **11a** with 0.2 mg mL^{-1} *Pp*IRED and NADP^{+} at 0 min (top) and 24 h (bottom). There is no evidence of the depletion of the peak corresponding to **11a** after 24 h reaction time.

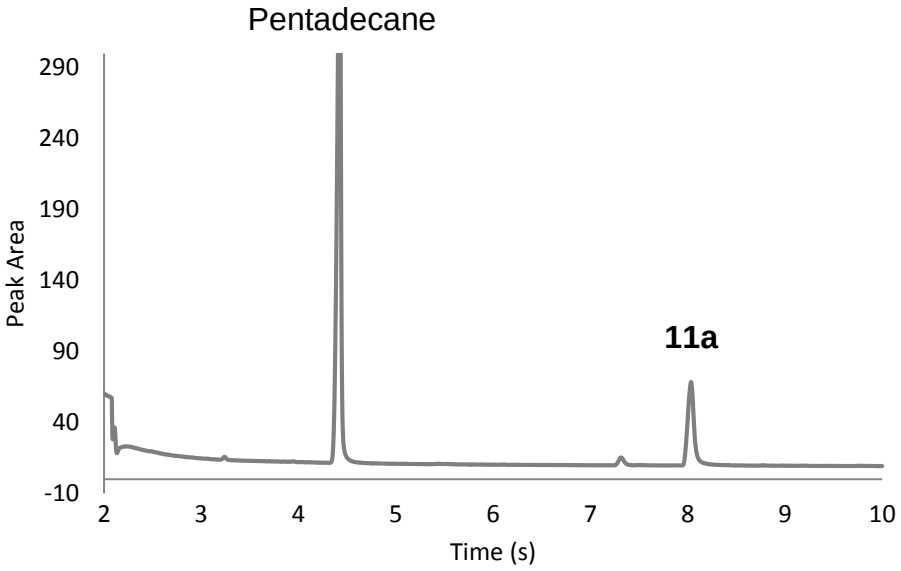


Figure 6.21: Chromatogram recorded from GC analysis of samples drawn from a reaction of **11a** with 5 mg mL^{-1} *Pp*IRED and $NADP^+$ at 0 min (top) and 24 h (bottom). There is, again, no evidence of the depletion of the peak corresponding to **11a** after 24 h reaction time despite an increased enzyme concentration.

The lack of change in chromatograms between 0 min and 24 h reaction time for oxidation of **11a** with *Pp*IRED and $NADP^+$ is inconsistent with the findings of the NADPH-formation kinetic assay carried out in Section 6.6, the results of which suggested that *Pp*IRED does indeed oxidise **11a** to the corresponding imine **11b**. There are multiple possibilities as to why this is not seen during GC analysis. Firstly, a UV-Vis based NADPH formation assay is far more sensitive than a GC assay, and it is possible to detect only very minute levels of substrate conversion. It is possible that the levels of conversion are insufficient to be detected using GC. Additionally, because there is no commercially available standard for the product of the reaction, imine **11b**, it is impossible to assess where the product peak should appear. There is a small possibility that the substrate and product peaks co-elute very closely and are therefore indistinguishable from each other.

6.8 Conclusion

A novel IRED from *Physcomitrella patens*, a plant often used as a model organism in studies of plants. was identified during a UV-Vis based NADPH-concentration screen carried out by GSK towards a broad range of amine and imine substrates. *Pp*IRED was uniquely active towards two amine substrates; **11a** and **12a**. Its unique reactivity profile made it interesting as a target for further study.

The structure of *Pp*IRED was solved both in its *apo* form and in complex with its cofactor, $NADP^+$. *Pp*IRED shares most similarity, both in terms of its sequence and structure, with a Glyoxylate reductase from *Arabidopsis thaliana* (PDB accession number 3DOJ); also a plant plant-based organism. Like other IREDs, *Pp*IRED possesses an *N*-terminal Rossman fold motif and a *C*-terminal domain. However, unlike other IREDs, *Pp*IRED does not form a dimer with a reciprocal domain sharing arrangement. The cofactor $NADP^+$ binds in a cleft formed between the *N*- and *C*-terms domains of the same subunit unlike in other IREDs, where this cleft is formed between domains of opposite subunits.

The reactivity profile of *Pp*IRED was characterised further using UV-Vis based NADPH-concentration assays. Activity was tested not only towards amines **11a** and **12a** but also carbonyl- and hydroxyl-containing compounds **13a** – **16a**. Analysis showed that *Pp*IRED indeed catalyses the oxidation of amine **11a**, but displays preferential activity towards the equivalent alcohol **14a**. *Pp*IRED displayed particularly strong activity towards the reduction of aldehydes the quinoline derivative **13a** and glyoxylate, **16a**. The levels of activity of *Pp*IRED towards **16a** were several orders of magnitude higher than for any other substrate, suggesting that this may be a natural substrate for the enzyme. This result is expected due to the high levels of sequence identity and structural similarity of *Pp*IRED with 3DOJ.

GC analysis was used to verify the fact that *Pp*IRED can indeed oxidise amines, as the ability to convert both amines (and therefore imines) and aldehydes/alcohols was unprecedented in any other IRED. However, GC analysis remains inconclusive as findings did not suggest **11a** was converted, potentially due to very low levels of substrate conversion.

7. Conclusion

Imine reductases (IREDs) are enzymes of the oxidoreductase group that catalyse the asymmetric reduction of imines to form chiral amines. They offer a convenient, sustainable method for the synthesis of chiral amines, compounds which are widely applied in the chemical, agrochemical and pharmaceutical industries.

During the course of this project, NADPH-dependent IREDs *SkR*-IRED from *Streptomyces kanamyceticus* and and *SS*-IRED from *Streptomyces* sp. GF3546 were expressed and purified. These IREDs catalyse the reduction of the model imine substrate 2-methylpyrroline to form the corresponding (*R*)- and (*S*)-amine respectively.

The structure of *SkR*-IRED was used as a basis to study the mechanism of NADPH-dependent IREDs. It was proposed that residue Asp187 may play a role in the protonation step of imine reduction during catalysis. It was found that mutants in which Asp187 had been reduced to Ala or Asn were devoid of activity, which suggests that Asp187 is important for catalysis. However, due to the large interatomic distance between the C4 carbon of NADPH and the Asp187 residue (approximately 8 Ångstrom), the exact of involvement of this residue in catalysis remains unclear, particularly in the absence of a structure in which *SkR*-IRED is complexed with an imine. The substrate scope of *SkR*-IRED was also investigated, and it was found to be active towards 6-membered rings as well as quinoline-based imines.[60]

The structure of *SS*-IRED was solved in its *apo* form to a resolution of 3.2 Å, and permitted a comparison of the structure of an (*R*)- and (*S*)-selective IRED. *SS*-IRED was proposed to be dependent on a catalytic Tyr169 residue, and indeed the mutant Tyr169Phe was found to be devoid of activity. The direct involvement of Tyr169 was considered to be more feasible, as there is precedent for tyrosine as a proton donor in in other reductase mechanisms.[81] Furthermore the interatomic distance between the protic Tyr169 hydrogen and the C4 carbon of NADPH is considerably shorter (approximately 5 Ångstrom), in *SS*-IRED than in *SkR*-IRED

The prior study of both an (*R*)- and (*S*)-selective IREDs prompted an investigation into factors that may influence the selectivity of IREDs. Reciprocal mutants of an (*R*)- and (*S*)-selective IRED were generated in what is believed to be the substrate binding site

by sequence comparison with other IREDs.[82] None of these point mutations influenced selectivity, suggesting that the factors which might direct selectivity may involve more complex factors such as protein dynamics, or be affected by residues elsewhere in the protein.

Last, a novel IRED, *Pp*IRED from the moss *Physcomitrella patens*, which displayed activity towards a unique panel of substrates, was studied. While it appears that *Pp*IRED is active towards amines, producing certain imines in the oxidative direction, it was found to display greater activity towards carbonyl and hydroxyl substrates. The substrate towards which *Pp*IRED was most active was glyoxylate, and the homology of *Pp*IRED to known glyoxylate reductases suggests that *Pp*IRED is a glyoxylate reductase with promiscuous imine-reducing capabilities. The structure of *Pp*IRED was solved in both its *apo* form and in complex with NADPH, and demonstrated that it is structurally somewhat different to other IREDs. Most noticeably, it does not share the reciprocal domain sharing arrangement upon dimer formation that is usually characteristic of IREDs.

This work has contributed towards the understanding of IREDs as an enzyme class, a field which was in its infancy at the beginning of this project. The work reported in this thesis provided some of the first structural and mechanistic insights of IREDs.

In conclusion, IREDs have shown great promise as a class of enzymes for industrial application in the production of chiral imines. Wild-type IREDs are often active towards a structurally diverse selection of cyclic imine substrates, including substituted pyrrolines, substituted piperideines, isoquinoline derivatives and beta-carboline imines as well as a small number of acyclic imines with high enantioselectivity.[59-61, 78]

Furthermore, a subsection of this class of enzymes known as RedAms has evolved to catalyse the direct reductive amination of carbonyl-containing substrates. This reaction involves the *in situ* formation of an imine from the conjugation of an imine and ketone substrate, and removes the need for the separate formation of an imine intermediate.[61,83] This reaction is even more direct than that which is offered by traditional IREDs, and offers the potential for a highly efficient synthesis of chiral amines on an industrial scale.[61, 83]

Appendix

Publications relating to thesis

Throughout the duration of this PhD, three publications which relate to the project have been published.

1. Rodriguez-Mata, M., Frank, A., Wells, E., Leipold, F., Turner, N. J., Hart, S., Turkenburg, J. P., and Grogan, G. (2013) Structure and Activity of NADPH-Dependent Reductase Q1EQE0 from Streptomyces kanamyceticus, which Catalyses the R-Selective Reduction of an Imine Substrate, *ChemBioChem, 14*, 1372-1379.
2. Hussain, S., Leipold, F., Man, H., Wells, E., France, S. P., Mulholland, K. R., Grogan, G., and Turner, N. J. (2015) An (R)-Imine Reductase Biocatalyst for the Asymmetric Reduction of Cyclic Imines, *ChemCatChem, 7*, 579-583.
3. Man, H., Wells, E., Hussain, S., Leipold, F., Hart, S., Turkenburg, J. P., Turner, N. J., and Grogan, G. (2015) Structure, Activity and Stereoselectivity of NADPH-Dependent Oxidoreductases Catalysing the S-Selective Reduction of the Imine Substrate 2-Methylpyrroline, *ChemBioChem*, *16*, 1052-1059.

Primer sequences

Table 8.1: Primer sequences used for all sub-cloning experiments

Enzyme	Primer Purpose	Primer Sequence
SS-IRED	Sub-cloning into pET-YSBL-LIC3C vector, forward primer	CCAGGGACCAGCAATGAGCAAACAGTCA GTTACGGTGATTGGTCTG
SS-IRED	Sub-cloning into pET-YSBL-LIC3C vector, reverse primer	GAGGAGAAGGCGCGTTACACTGCCGGTT TTTTCAGGACTTCAATCAGAG

Table 8.2: Primer sequences used for all site-directed mutagenesis experiments

Enzyme	Mutation and Primer Direction	Primer Sequence
SkR-IRED	D178A, forward primer	CAAGCCTGTATGCGGCCGCAGGTCTGG
SkR-IRED	D178A, reverse primer	CCAGACCTGCGGCCGCATACAGGCTTG
SkR-IRED	D178N, forward primer	GGCAAGCCTGTATAATGCCGCAGGTCTGGTC
SkR-IRED	D178N, reverse primer	GACCTGCGGCATTATACAGGCTTGCC
SS-IRED	Y169F, forward primer	CTGGCCGCAATGTATTTTCAGGCGCAAATGACC
SS-IRED	Y169F, reverse primer	GGTCATTTGCGCCTGAAAATACATTGCGGCCAG

SS-IRED	M179F, forward primer	CGTGTCACTGCTGGGTCTGTTTTGGGGCAC GCTGAACTCG
SS-IRED	M179F, reverse primer	CGAGTTCAGCGTGCCCCAAAACAGACCCAG CAGTCACACG
SS-IRED	L176M, forward primer	CTACGACGTGTCACTGATGGGTCTGATGTGGG
SS-IRED	L176M, reverse primer	CCCACATCAGACCCATCAGTGACACGTCGTAC
SS-IRED	Y283F, forward primer	CACGCTAAAAACAGCTTTGCGGCAGTCCTGAAAG
SS-IRED	Y283F, reverse primer	CTTTCAGGACTGCCGCAAAGCTGTTTTTAGCGTG
SS-IRED	H242G, forward primer	CGCTACGCTGGAAACCGGCCTGGCGGCCCTGAAAC
SS-IRED	H242G, reverse primer	GTTTTCAGGGCCGCCAGGCCGGTTTCCAGCGTAGG
SR-IRED	L183M forward primer	CTGATGTGGGGCACGATGAACTCGTTTCTGC
SR-IRED	L183M reverse primer	GCAGAAACGAGTTCATCGTGCCCCACATCAG
SS-IRED	F176M forward primer	CCAGGCGCAAATGACCATTATGTGGACCACGATC TGAGC
SS-IRED	F176M reverse primer	GCTCAGCATCGTGGTCCACATAATGGTCATTTGCC CTGG
SS-IRED	M173L forward primer	GTATTACCAGGCGCAACTGACCATTTTCTGGACC
SS-IRED	M173L reverse primer	GGTCCAGAAAATGGTCAGTTGCGCCTGGTAATAC
SS-IRED	F278Y forward primer	GCTTTGCTGAAAATTCGTATTCCTCTCTGATTGA A GTCC
SS-IRED	F278Y reverse primer	GGACTTCAATCAGAGAGGAATACGAATTTTCAGC A AAGC
SS-IRED	G237H forward primer	GTGGACCGTCTGGCGATGCATGCAGCTTCAGTCG AT CACG
SS-IRED	G237H reverse primer	CGTGATCGACTGAAGCTGCATGCATCGCCAGACG GT CCAC
SS-IRED	M180L reverse primer	CATTTTCTGGACCACGCTGCTGAGCTATTACC
SS-IRED	M180L forward primer	GCTAATAGCTCAGCAGCGTGGTCCAGAAAATG

Glossary of Abbreviations and Symbols

Abbreviation	Denotation
ADH	Alcohol dehydrogenase
AmDH	Amine dehydrogenase
bp	Base pairs
DHFR	Dihydrofolate reductase
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
EDTA	Ethylenediaminetetraacetic acid
g	grams
HIBDH	Hydroxyisobutyrate dehydrogenase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRED	Imine reductase
L	Litre
LB	Lysogeny broth
LB-Amp	LB supplemented with 100 μg mL^{-1} ampicillin
LB-Kan	LB supplemented with 30 μg mL^{-1} kanamycin
M	Molar
m/z	Mass charge ratio
MAO	Monoamine oxidase
mg	Milligram
mL	Millilitre
mM	Millimolar
mol	Mole

$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PTR1	Pteridine reductase
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate protein agarose gel electrophoresis
SkR-IRED	(*R*)-selective imine reductase from *Streptomyces kanamyceticus*
SOC	Super optimal broth with catabolite repression
SR-IRED	(*R*)-selective imine reductase from *Streptomyces* sp GF3587
SS-IRED	(*S*)-selective imine reductase from *Streptomyces* sp GF3546
TEMED	Tetramethylethylenediamine
U	Units (of enzyme activity)
v/v	Volume per volume
w/v	Weight per volume
µL	Microliter
DCM	Dichloromethane

References

[1] Grogan, G. (2009) *Practical Biotransformations: A Beginner's Guide*, Wiley-Blackwell.

[2] Bigwood, E. J., and Demerre, L. J. (1945) The Discovery of Antibiotics in General and Penicillin in Particular, *J. Am. Med. Assoc.*, *128*, 461-462.

[3] Hohn, M., and Bornscheuer, U. T. (2009) Biocatalytic Routes to Optically Active Amines, *ChemCatChem*, *1*, 42-51.

[4] Warner, J. C. (2012) Green Chemistry: Theory and practice, *Abstr. Pap. Am. Chem. Soc.*, *244*.

[5] Nieuwenhuijzen, J. W., Grimbergen, R. F. P., Koopman, C., Kellogg, R. M., Vries, T. R., Pouwer, K., van Echten, E., Kaptein, B., Hulshof, L. A., and Broxterman, Q. B. (2002) The role of nucleation inhibition in optical resolutions with families of resolving agents, *Angew. Chem. Int. Ed.*, *41*, 4281-4286.

[6] Henderson, K. W., Kerr, W. J., and Moir, J. H. (2000) Enantioselective deprotonation reactions using a novel homochiral magnesium amide base, *Chem. Commun.*, 479-480.

[7] Adamo, M. F. A., Aggarwal, V. K., and Sage, M. A. (2000) Epoxidation of alkenes by amine catalyst precursors: Implication of aminium ion and radical cation intermediates, *J. Am. Chem. Soc.*, *122*, 8317-8318.

[8] Nugent, T. C. (2010) *Chiral Amine Synthesis: Methods, Developments and Applications*, Wiley-VCH.

[9] Blacker, A. J., Stirling, M. J., and Page, M. I. (2007) Catalytic racemisation of chiral amines and application in dynamic kinetic resolution, *Org. Process Res. Dev.*, *11*, 642-648.

[10] Minnaard, A. J., Feringa, B. L., Lefort, L., and De Vries, J. G. (2007) Asymmetric hydrogenation using monodentate phosphoramidite ligands, *Acc. Chem. Res.*, *40*, 1267-1277.

[11] van den Berg, M., Minnaard, A. J., Haak, R. M., Leeman, M., Schudde, E. P., Meetsma, A., Feringa, B. L., de Vries, A. H. M., Maljaars, C. E. P., Willans, C. E., Hyett, D., Boogers, J. A. F., Henderickx, H. J. W., and de Vries, J. G. (2003) Monodentate phosphoramidites: A breakthrough in rhodium-catalysed asymmetric hydrogenation of olefins, *Adv. Synth. Catal.*, *345*, 308-323.

[12] Jia, X., Guo, R. W., Li, X. S., Yao, X. S., and Chan, A. S. C. (2002) Highly enantioselective hydrogenation of enamides catalyzed by rhodium-monodentate phosphoramidite complex, *Tetrahedron Lett.*, *43*, 5541-5544.

[13] Hu, A. G., Fu, Y., Xie, J. H., Zhou, H., Wang, L. X., and Zhou, Q. L. (2002) Monodentate chiral spiro phosphoramidites: Efficient ligands for rhodium-catalyzed enantioselective hydrogenation of enamides, *Angew. Chem. Int. Ed.*, *41*, 2348-2350.

[14] Emerson, W. S. (1947) The Preparation of Amines by Reductive Alkylation, *Org. React.*, *4*, 174-255.

[15] Blaser, H. U., Buser, H. P., Jalett, H. P., Pugin, B., and Spindler, F. (1999) Iridium ferrocenyl diphosphine catalyzed enantioselective reductive alkylation of a hindered aniline, *Synlett*, 867-868.

[16] Blaser, H. U., Buser, H. P., Coers, K., Hanreich, R., Jalett, H. P., Jelsch, E., Pugin, B., Schneider, H. D., Spindler, F., and Wegmann, A. (1999) The chiral switch of metolachlor: The development of a large-scale enantioselective catalytic process, *Chimia*, *53*, 275-280.

[17] Hoffmann, S., Seayad, A. M., and List, B. (2005) A powerful bronsted acid catalyst for the organocatalytic asymmetric transfer hydrogenation of imines, *Angew. Chem. Int. Ed.*, *44*, 7424-7427.

[18] Nugent, T. C., and Seemayer, R. (2006) An efficient enantiopure synthesis of a pivotal precursor to substance P antagonists, *Org. Process Res. Dev.*, *10*, 142-148.

[19] Blaser, H. U., and Spindler, F. (1997) Enantioselective catalysis for agrochemicals: The case history of the DUAL MAGNUM(R) herbicide, *Chimia 51*, 297-299.

[20] Ghislieri, D., and Turner, N. J. (2014) Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines, *Top. Catal.*, *57*, 284-300.

[21] Uenishi, J., Hiraoka, T., Hata, S., Nishiwaki, K., Yonemitsu, O., Nakamura, K., and Tsukube, H. (1998) Chiral Pyridines: Optical resolution of 1-(2-pyridyl)- and 1-[6-(2,2 '-bipyridyl)]ethanols by lipase-catalyzed enantioselective acetylation, *J. Org. Chem.*, *63*, 2481-2487.

[22] Gonzalez-Sabin, J., Gotor, V., and Rebolledo, F. (2002) CAL-B-catalyzed resolution of some pharmacologically interesting beta-substituted isopropylamines, *Tetrahedron: Asymmetry*, *13*, 1315-1320.

[23] Jaeger, K. E., and Eggert, T. (2002) Lipases for biotechnology, *Curr. Opin. Biotechnol.*, *13*, 390-397.

[24] Zaks, A., and Klibanov, A. M. (1984) Enzymatic Catalysis in Organic Media at 100-Degrees-C, *Science*, *224*, 1249-1251.

[25] Kumar, A., Dhar, K., Kanwar, S. S., and Arora, P. K. (2016) Lipase catalysis in organic solvents: advantages and applications, *Biol. Proced. Online*, *18*.

[26] Reetz, M. T., and Schimossek, K. (1996) Lipase-catalyzed dynamic kinetic resolution of chiral amines: Use of palladium as the racemization catalyst, *Chimia, 50*, 668-669.

[27] Balkenhohl, F., Ditrich, K., Hauer, B., and Ladner, W. (1997) Optically active amines via lipase-catalyzed methoxyacetylation, *J. Prakt. Chem./Chem. Ztg.*, *339*, 381-384.

[28] Ricca, E., Brucher, B., and Schrittwieser, J. H. (2011) Multi-Enzymatic Cascade Reactions: Overview and Perspectives, *Adv. Synth. Catal.*, *353*, 2239-2262.

[29] Shin, J. S., Kim, B. G., Liese, A., and Wandrey, C. (2001) Kinetic resolution of chiral amines with omega-transaminase using an enzyme-membrane reactor, *Biotechnol. Bioeng.*, *73*, 179-187.

[30] Thornberry, N. A., and Weber, A. E. (2007) Discovery of JANUVIA (TM) (Sitagliptin), a selective dipeptidyl peptidase IV inhibitor for the treatment of type2 diabetes, *Curr. Top. Med. Chem.*, *7*, 557-568.

[31] Weber, A. E., and Thornberry, N. (2007) Case History: JANUVIA (TM) (Sitagliptin), a Selective Dipeptidyl Peptidase IV Inhibitor for the Treatment of Type 2 Diabetes, *Annu. Rep. Med. Chem.*, *42*, 95-109.

[32] 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., and Hughes, G. J. (2010) Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture, *Science*, *329*, 305-309.

[33] Desai, A. A. (2011) Sitagliptin Manufacture: A Compelling Tale of Green Chemistry, Process Intensification, and Industrial Asymmetric Catalysis, *Angew. Chem. Int. Ed.*, *50*, 1974-1976.

[34] Li, M., Binda, C., Mattevi, A., and Edmondson, D. E. (2006) Functional role of the "aromatic cage" in human monoamine oxidase B: Structures and catalytic properties of Tyr435 mutant proteins, *Biochemistry*, *45*, 4775-4784.

[35] Schilling, B., and Lerch, K. (1995) Cloning, Sequencing and Heterologous Expression of the Monoamine-Oxidase Gene from Aspergillus-Niger, *Mol. Genet. Genomics*, *247*, 430-438.

[36] Schilling, B., and Lerch, K. (1995) Amine Oxidases from Aspergillus-Niger - Identification of a Novel Flavin-Dependent Enzyme, *Biochim. Biophys. Acta, Gen. Subj.*, *1243*, 529-537.

[37] Sablin, S. O., Yankovskaya, V., Bernard, S., Cronin, C. N., and Singer, T. P. (1998) Isolation and characterization of an evolutionary precursor of human monoamine oxidases A and B, *Eur. J. Biochem., 253*, 270-279.

[38] Alexeeva, M., Enright, A., Dawson, M. J., Mahmoudian, M., and Turner, N. J. (2002) Deracemization of alpha-methylbenzylamine using an enzyme obtained by in vitro evolution, *Angew. Chem. Int. Ed.*, *41*, 3177-+.

[39] Turner, N. J. (2011) Enantioselective Oxidation of C-O and C-N Bonds Using Oxidases, *Chem. Rev.*, *111*, 4073-4087.

[40] Kohler, V., Bailey, K. R., Znabet, A., Raftery, J., Helliwell, M., and Turner, N. J. (2010) Enantioselective Biocatalytic Oxidative Desymmetrization of Substituted Pyrrolidines, *Angew. Chem. Int. Ed.*, *49*, 2182-2184.

[41] Znabet, A., Polak, M. M., Janssen, E., de Kanter, F. J. J., Turner, N. J., Orru, R. V. A., and Ruijter, E. (2010) A highly efficient synthesis of telaprevir by strategic use of biocatalysis and multicomponent reactions, *Chem. Commun.*, *46*, 7918-7920.

[42] Itoh, N., Yachi, C., and Kudome, T. (2000) Determining a novel NAD(+)-dependent amine dehydrogenase with a broad substrate range from Streptomyces virginiae IFO 12827: purification and characterization, *J. Mol. Catal. B: Enzym.*, *10*, 281-290.

[43] Mutti, F. G., Knaus, T., Scrutton, N. S., Breuer, M., and Turner, N. J. (2015) Conversion of alcohols to enantiopure amines through dual-enzyme hydrogen-borrowing cascades, *Science*, *349*, 1525-1529.

[44] Smith, D. R., and Calvo, J. M. (1980) Nucleotide-Sequence of the Escherichia-Coli Gene Coding for Dihydrofolate-Reductase, *Nucleic Acids Res.*, *8*, 2255-2274.

[45] Eguchi T., O. T., Kuge Y., Mochida K., Uwajima T. (1990) Process for producing L(-)tetrahydrofolic acid.

[46] Li, W., Chou, S. C., Khullar, A., and Gerratana, B. (2009) Cloning and Characterization of the Biosynthetic Gene Cluster for Tomaymycin, an SJG-136 Monomeric Analog, *Appl. Environ. Microbiol., 75*, 2958-2963.

[47] Li, W., Khullar, A., Chou, S., Sacramo, A., and Gerratana, B. (2009) Biosynthesis of Sibiromycin, a Potent Antitumor Antibiotic, *Appl. Environ. Microbiol., 75*, 2869-2878.

[48] Taylor, M., Scott, C., and Grogan, G. (2013) F-420-dependent enzymes - potential for applications in biotechnology, *Trends Biotechnol.*, *31*, 63-64.

[49] Chadha, A., and Baskar, B. (2002) Biocatalytic deracemisation of alpha-hydroxy esters: high yield preparation of (S)-ethyl 2-hydroxy-4-phenylbutanoate from the racemate, *Tetrahedron: Asymmetry*, *13*, 1461-1464.

[50] Vaijayanthi, T., and Chadha, A. (2008) Asymmetric reduction of aryl imines using Candida parapsilosis ATCC 7330, *Tetrahedron: Asymmetry*, *19*, 93-96.

[51] Espinoza-Moraga, M., Petta, T., Vasquez-Vasquez, M., Laurie, V. F., Moraes, L. A. B., and Santos, L. S. (2010) Bioreduction of beta-carboline imines to amines employing Saccharomyces bayanus, *Tetrahedron: Asymmetry*, *21*, 1988-1992.

[52] Mirabal-Gallardo, Y., Soriano, M. D. C., and Santos, L. S. (2013) Stereoselective bioreduction of beta-carboline imines through cell-free extracts from earthworms (Eisenia foetida), *Tetrahedron: Asymmetry*, *24*, 440-443.

[53] Mitsukura, K., Suzuki, M., Shinoda, S., Kuramoto, T., Yoshida, T., and Nagasawa, T. (2011) Purification and Characterization of a Novel (R)-Imine Reductase from Streptomyces sp GF3587, *Biosci., Biotechnol., Biochem., 75*, 1778-1782.

[54] Mitsukura, K., Suzuki, M., Tada, K., Yoshida, T., and Nagasawa, T. (2010) Asymmetric synthesis of chiral cyclic amine from cyclic imine by bacterial whole-cell catalyst of enantioselective imine reductase, *Org. Biomol. Chem.*, *8*, 4533-4535.

[55] Mitsukura, K., Kuramoto, T., Yoshida, T., Kimoto, N., Yamamoto, H., and Nagasawa, T. (2013) A NADPH-dependent (S)-imine reductase (SIR) from Streptomyces sp GF3546 for asymmetric synthesis of optically active amines: purification, characterization, gene cloning, and expression, *Appl. Microbiol. Biotechnol.*, *97*, 8079-8086.

[56] Nagasawa T., Y. T., Ishida K., Yamamoto H., Kimoto N. (2011) Process for production of optically active amine derivative.

[57] Leipold F., H. S., Ghislieri D., Turner N. J. (2013) Asymmetric Reduction of Cyclic Imines Catalyzed by a Whole-Cell Biocatalyst Containing an (S)-Imine Reductase, *ChemCatChem*., *12*, 3505-3508.

[58] Grogan, G., and Turner, N. J. (2016) InspIRED by Nature: NADPH-Dependent Imine Reductases (IREDs) as Catalysts for the Preparation of Chiral Amines, *Chem. - Eur. J.*, *22*, 1900-1907.

[59] Hussain, S., Leipold, F., Man, H., Wells, E., France, S. P., Mulholland, K. R., Grogan, G., and Turner, N. J. (2015) An (R)-Imine Reductase Biocatalyst for the Asymmetric Reduction of Cyclic Imines, *ChemCatChem, 7*, 579-583.

[60] Rodriguez-Mata, M., Frank, A., Wells, E., Leipold, F., Turner, N. J., Hart, S., Turkenburg, J. P., and Grogan, G. (2013) Structure and Activity of NADPH-Dependent Reductase Q1EQE0 from Streptomyces kanamyceticus, which Catalyses the R-Selective Reduction of an Imine Substrate, *ChemBioChem, 14*, 1372-1379.

[61] Huber, T., Schneider, L., Prag, A., Gerhardt, S., Einsle, O., and Muller, M. (2014) Direct Reductive Amination of Ketones: Structure and Activity of S-Selective Imine Reductases from Streptomyces, *ChemCatChem, 6*, 2248-2252.

[62] Scheller, P. N., Fademrecht, S., Hofelzer, S., Pleiss, J., Leipold, F., Turner, N. J., Nestl, B. M., and Hauer, B. (2014) Enzyme Toolbox: Novel Enantiocomplementary Imine Reductases, *ChemBioChem*, *15*, 2201-2204.

[63] Gand, M., Muller, H., Wardenga, R., and Hohne, M. (2014) Characterization of three novel enzymes with imine reductase activity, *J. Mol. Catal. B: Enzym.*, *110*, 126-132.

[64] Li, H., Luan, Z. J., Zheng, G. W., and Xu, J. H. (2015) Efficient Synthesis of Chiral Indolines using an Imine Reductase from Paenibacillus lactis, *Adv. Synth. Catal.*, *357*, 1692-1696.

[65] Wetzl, D., Berrera, M., Sandon, N., Fishlock, D., Ebeling, M., Muller, M., Hanlon, S., Wirz, B., and Iding, H. (2015) Expanding the Imine Reductase Toolbox by Exploring the Bacterial Protein-Sequence Space, *ChemBioChem*, *16*, 1749-1756.

[66] Bonsor, D., Butz, S. F., Solomons, J., Grant, S., Fairlamb, I. J. S., Fogg, M. J., and Grogan, G. (2006) Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes, *Org. Biomol. Chem.*, *4*, 1252-1260.

[67] Lokanath, N. K., Ohshima, N., Takio, K., Shiromizu, I., Kuroishi, C., Okazaki, N., Kuramitsu, S., Yokoyama, S., Miyano, M., and Kunishima, N. (2005) Crystal structure of novel NADP-dependent 3-hydroxyisobutyrate dehydrogenase from Thermus thermophilus HB8, *J Mol Biol 352*, 905-917.

[68] Smitst, S. H. J., Mueller, A., Schmitt, L., and Grieshaber, M. K. (2008) A structural basis for substrate selectivity and stereoselectivity in octopine dehydrogenase from Pecten maximus, *J. Mol. Biol.*, *381*, 200-211.

[69] Chang, F. Y., Ternei, M. A., Calle, P. Y., and Brady, S. F. (2013) Discovery and Synthetic Refactoring of Tryptophan Dimer Gene Clusters from the Environment, *J. Am. Chem. Soc.*, *135*, 17906-17912.

[70] Chang, F. Y., Ternei, M. A., Calle, P. Y., and Brady, S. F. (2015) Targeted Metagenomics: Finding Rare Tryptophan Dimer Natural Products in the Environment, *J. Am. Chem. Soc.*, *137*, 6044-6052.

[71] Bond, C. S., and Schuttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments, *Acta Crystallogr.*, *65*, 510-512.

[72] Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res.*, *42*, W320-W324.

[73] Kabsch, W., and Sander, C. (1983) Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features, *Biopolymers*, *22*, 2577-2637.

[74] Joosten, R. P., Beek, T. A. H. T., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R., Sander, C., and Vriend, G. (2011) A series of PDB related databases for everyday needs, *Nucleic Acids Res.*, *39*, D411-D419.

[75] Du, X. Z., Wang, Y. C., Ding, Y. H., and Guo, R. (2007) Protein-directed assembly of binary monolayers at the interface and surface patterns of protein on the monolayers, *Langmuir*, *23*, 8142-8149.

[76] Krissinel, E. (2010) Crystal Contacts as Nature's Docking Solutions, *J. Comput. Chem.*, *31*, 133-143.

[77] Harris, T. K., and Turner, G. J. (2002) Structural basis of perturbed pK(a) values of catalytic groups in enzyme active sites, *IUMBM Life*, *53*, 85-98.

[78] Man, H., Wells, E., Hussain, S., Leipold, F., Hart, S., Turkenburg, J. P., Turner, N. J., and Grogan, G. (2015) Structure, Activity and Stereoselectivity of NADPH-Dependent Oxidoreductases Catalysing the S-Selective Reduction of the Imine Substrate 2-Methylpyrroline, *ChemBioChem*, *16*, 1052-1059.

[79] Ching, S. L. K., Gidda, S. K., Rochon, A., van Cauwenberghe, O. R., Shelp, B. J., and Mullen, R. T. (2012) Glyoxylate Reductase Isoform 1 is Localized in the Cytosol and Not Peroxisomes in Plant Cells, *J. Integr. Plant Biol.*, *54*, 152-168.

[80] Allan, W. L., Clark, S. M., Hoover, G. J., and Shelp, B. J. (2009) Role of plant glyoxylate reductases during stress: a hypothesis, *Biochem. J.*, *423*, 15-22.

[81] Gourley, D. G., Schuttelkopf, A. W., Leonard, G. A., Luba, J., Hardy, L. W., Beverley, S. M., and Hunter, W. N. (2001) Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites, *Nat. Struct. Biol.*, *8*, 521-525.

[82] Aleku, G. A., Man, H., France, S. P., Leipold, F., Hussain, S., Toca-Gonzalez, L., Marchington, R., Hart, S., Turkenburg, J. P., Grogan, G., and Turner, N. J. (2016) Stereoselectivity and Structural Characterization of an Irvine Reductase (IRED) from Amycolatopsis orientalis, *ACS Catal.*, *6*, 3880-3889.

[83] Scheller, P. N., Lenz, M., Hammer, S. C., Hauer, B., and Nestl, B. M. (2015) Imine Reductase-Catalyzed Intermolecular Reductive Amination of Aldehydes and Ketones, *ChemCatChem*, *7*, 3239-3242.