Discovery and Development of Class I Lyase-like Enzymes for Biotechnological Applications

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (Ph.D.) in the Faculty of Engineering and Physical Sciences

2015

Nicholas James Weise

School of Chemistry
Contents

List of Figures .................................................................................................................. 6
List of Tables ..................................................................................................................... 14
Declaration ....................................................................................................................... 17
Copyright Statement ....................................................................................................... 18
Acknowledgements ......................................................................................................... 19

1. Introduction .................................................................................................................. 20
  1.1 The Class I Lyase-like Enzyme Family ..................................................................... 21
    1.1.1 Histidine Ammonia Lyases ................................................................................ 22
    1.1.2 Phenylalanine Ammonia Lyases ....................................................................... 23
    1.1.3 Tyrosine Ammonia Lyases ................................................................................ 24
    1.1.4 Phenylalanine Aminomutases .......................................................................... 26
    1.1.5 Tyrosine Aminomutases ................................................................................... 27
  1.2 Biomedical Applications of Class I Lyase-like Enzymes ........................................ 28
    1.2.1 Amino Acid Depletion Therapy ....................................................................... 29
    1.2.2 Treatment of Phenylketonuria ......................................................................... 30
  1.3 Class I Lyase-like Enzymes as Biocatalysts .............................................................. 31
    1.3.1 Unnatural Amino Acids ................................................................................... 32
      1.3.1.1 Phenylalanine Derivatives as Chiral Building Blocks ............................... 34
      1.3.1.2 Phenylalanine Derivatives as Chemical Biology Tools ......................... 36
    1.4 Objectives .............................................................................................................. 37

2. Computational Analyses of Class I Lyase-like Enzymes ............................................ 39
  2.1 Background ............................................................................................................... 40
    2.1.1 Electronic Prediction of Gene Function .............................................................. 40
    2.1.2 Annotation of Class I Lyase-like Enzyme Sequences ....................................... 40
  2.2 Results and Discussion ............................................................................................ 42
5.2.4 Bacterial Transformation ................................................................. 151
5.2.5 Small Scale Bacterial Cultures .......................................................... 152
5.2.6 Plasmid DNA Extraction ................................................................. 152
5.2.7 Restriction Enzyme Analysis .............................................................. 153
5.2.8 DNA Sequencing ........................................................................... 153
5.2.9 Biocatalyst Production Cultures ........................................................ 154
5.2.10 Protein Purification ...................................................................... 154
5.2.11 Sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) .................................................................................. 155
5.2.12 Whole Cell Biotransformations ....................................................... 155
5.3 Analytical Methods ........................................................................... 156
5.3.1 Non-chiral Analyses of Biotransformations ....................................... 156
5.3.2 Chiral Analysis of Amino Acid Products ........................................... 158
5.4 Protein and Nucleotide Sequences ..................................................... 166
References .............................................................................................. 168
Appendix ................................................................................................. 176

Final Word Count: 49,522
List of Figures

Figure 1 || The five enzyme activities found in members of the class I lyase-like family including histidine ammonia lyase (HAL), phenylalanine ammonia lyase (PAL), phenylalanine aminomutase (PAM), tyrosine ammonia lyase (TAL) and tyrosine aminomutase (TAM). ..........................................................21

Figure 2 || The proposed catalytic mechanism of MIO / tyrosine-mediated ammonia elimination or amination in class I lyase-like enzymes. The enzyme is shown in black, the substrate in blue and the transferred groups in red. For α-amino acids R₁ = aryl and R₂ = carboxyl. For β-amino acids R₁ = carboxyl and R₂ = aryl. ........................................22

Figure 3 || The primary metabolic pathway in bacteria by which histidine is converted to glutamate. This process is initiated by the HAL HutH with subsequent reactions mediated through additional enzymes in the Hut pathway. Additions to the initial histidine scaffold are shown in red. ............................................................................................................23

Figure 4 || A selection of secondary metabolites biosynthesised from the cinnamate product of PAL-mediated deamination of phenylalanine. The parts of the precursor phenylalanine and derived products cognate to the cinnamate constituents are highlighted in red. ........24

Figure 5 || Three examples of utilisation of the TAL product para-coumarate in bacterial natural products and proteins. The parts of the precursor tyrosine and derived natural products cognate to the cinnamate constituents are highlighted in red. With saccharomicins the R group represents a complex branched polysaccharide. .................................................................25

Figure 6 || The only two natural products known to contain aminomutase-derived β-phenylalanine. The incorporation of (S)-β-phenylalanine by Pantoea agglomerans into the antibiotic andrimid is shown on the left. The synthesis of taxol from (R)-β-phenylalanine in various species of yew tree (Taxus spp.) is shown on the right. ...............................27

Figure 7 || A summary of the tyrosine conversions of characterised enzymes known to have TAM activity. Predominant activities are shown thick bold arrows whereas minor side activities are shown with thinner arrows. .................................................................28

Figure 8 || Possible causes, effects and treatment of the inborn error of metabolism phenylketonuria (PKU). ........................................................................................................................................30

Figure 9 || The reversibility of the natural reactions (thin arrows) of PAL and PAM, whereby the equilibrium can be pushed toward amino acid products starting with cinnamate and excess ammonia (thick arrows). .................................................................................31

Figure 10 || A selection of (S)- and (R)-β-amino acids synthesised by the enantiocomplementary aminomutases (AdmH and TcPAM respectively) via the isomerisation of corresponding (S)-α-regioisomers. .................................................................32

Figure 11 || Examples of unnatural amino acid derivatives of phenylalanine (left) and the general back bone structures of α- and β-peptides (right). .............................................................................34
Figure 12 /// Six examples of pharmaceutical compounds synthesised from phenylalanine derivatives (the phenylalanine component of each is highlighted in green). ………………………35

Figure 13 /// Three examples of site-specific incorporation of unnatural phenylalanine derivatives into various proteins / peptides for chemical biological studies (the phenylalanine component of each is highlighted in green). ………………………………………………………………………………37

Figure 14 /// A cladogram of 31 characterised class I lyase-like enzymes based on a full sequence alignment, of which characteristic sections (1, 2 and 3) are shown. Section 1 is the region surrounding the inner active site loop, with loop residues, as evident from various crystal structures, highlighted in yellow. Section 2 shows the aryl binding pocket specificity residues of each enzyme. Section 3 contains the homologous position proposed to allow differentiation between HAL and the other catalytic activities. The distinguishing glutamate is highlighted in blue for each of the 6 HAL enzymes in the alignment. The enzymes in the alignment are: MdpC4 from Actinomadura madurae, SgTAM from Streptomyces globisporus, CmTAL from Cupriavidus metallidurans, CmdF from Chondromyces croatus, KedY4 from Streptoclooteichus sp., MtTAM from Myxococcus fulvus, MxTAM from Myxococcus sp., EaHAL from Enterobacter aerogenes, StHAL from Salmonella typhimurium, MmHAL from Mus musculus, RnHAL from Rattus norvegicus, SgHAL from Streptomyces griseus, RsTAL from Rhodobacter sphaeroides, AdmH from Pantoea agglomerans, StlA from Photobacterium luminescens, PpHAL from Pseudomonas putida, EncP from Streptomyces maritimus, AvPAL from Anabaena variabilis, NpPAL from Nostoc punctiforme, TcPAM from Taxus canadensis, TwPAM from Taxus wallichiana var. chinensis, PcPAL1 from Petroseolum crispm, AtPAL1-4 from Arabidopsis thaliana, ZmPAL1 from Zea mays, RtPAL from Rhodosporidium toruloides, RgTAL from Rhodotorula graminis, BagA from Streptomyces sp. and Sam8 from Saccharothrix espanaensis. ………………………………………………………………………………43

Figure 15 /// A section of the sequence alignment for AvPAL from Anabaena variabilis and its closest tBLASTn hits. The highly conserved loop region (as evident in the structure of AvPAL) of the first six sequences is enclosed within a dark blue box with amino acid differences highlighted in light blue. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). The hits include the characterised NpPAL from Nostoc punctiforme, TcPAM from Taxus canadensis, TwPAM from Taxus wallichiana var. chinensis, PcPAL1 from Petroseolum crispm, AtPAL1-4 from Arabidopsis thaliana, ZmPAL1 from Zea mays, RtPAL from Rhodosporidium toruloides, RgTAL from Rhodotorula graminis, BagA from Streptomyces sp. and Sam8 from Saccharothrix espanaensis. ……………………………………………………………………………………………………48

Figure 16 /// A flowchart showing the possible mutations in the selectivity residues of class I lyase-like enzymes to allow conversion of PAL=TAL activity and vice versa. The histidine residue responsible for tyrosine substrate preference in highlighted in green. …………………49

Figure 17 /// A section of the sequence alignment for StlA from Photobacterium luminescens and its closest tBLASTn hits. The 4 aa loop lid motif, as inferred from other family members are enclosed within a dark blue box. The specificity residues are shown within the orange
box and sequences displaying potential tyrosine specificity residues at these positions are highlighted (also in orange). The hits include the following putative proteins: PaPAL from Photorhabdus asymbiotica, YePAL from Yersinia enterocolitica, DpPAL from Dictyostelium purpureum, DdPAL from Dictyostelium discoideum, ScTAL from Streptomyces clavuligerus, MaTAL from Methylomicrobium album, SnPAM from Stackebrandtia nassauensis, BiPAL from Brevibacillus laterosporus and ToHAL from Thermosediminibacter oceanii. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Figure 18 || A section of the sequence alignment for RtPAL from Rhodosporidium toruloides and selected tBLASTn hits. The inner active site loop residues for RtPAL, as evident from the structure, are highlighted in yellow. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). The hits include the following putative proteins: PmPAL1 and PmPAL2 from Pencillium marneffei, UmPAL from Ustilago maydis, PgPAL from Puccinia graminis, CcPAL from Coprinopsis cinerea, EnPAL from Emericella nidulans, AoPAL1-3 from Aspergillus oryzae, PcPAL from Pencillium chrysogenum, PaPAL from Podospora aserina, PnPAL from Phaeosphaeria nodorum and AnPAL from Aspergillus niger. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Figure 19 || Similarities between the fungus- and plant-associated cyclochlorotine and astin B cytotoxic agents isolated from Pencillium islandicum and Aster tartaricus respectively. In both structures the (R)-β-phenylalaninyl portion is shown in blue and differences in astin B with respect to cyclochlorotine are highlighted in red.

Figure 20 || Sections of the sequence alignment for TcPAM from Taxus canadensis and its closest tBLASTn hits. The inner active site loop residues of the hit, TwPAM from Taxus wallichiana var. chinensis, as evident from the structure, are highlighted in yellow. The specificity residues are shown within the orange box and sequences displaying characteristic plant mutase residues at these positions are highlighted (also in orange). Salt bridges predicted from a model of TcPAM are shown between an R and D residue (red box) and an R and E residue (blue box). The hits include two further characterised mutases: TmPAM from the hybrid yew tree Taxus x media and TbPAM from Taxus baccata. Also included in the hits are the following putative PALs: PpPAL from Physcomitrella patens, IIIPAL from Isoetes lacustris, DfPAL from Diphasiastrum tristachyum, PmPAL from Pinus massonia, GbPAL from Gingko biloba, PtPAL from Pinus taeda and LkPAL from Larix kaempferi. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Figure 21 || The loop sections of the sequence alignment between TwPAM from Taxus wallichiana var. chinensis and PcPAL1 from Petroselinum crispum used to guide engineering efforts of TwPAM in a recently reported study.

Figure 22 || Sections of the sequence alignment for EncP from Streptomyces maritimus and its closest tBLASTn hits. The inner active site loop residues of the top hit, AdmH from...
Pantoea agglomerans, as evident from the structure, are highlighted in yellow. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). A common DFQD motif for all the putative PAMs is enclosed within a dark blue box with the characteristic glutamate in the corresponding HAL sequences highlighted in light blue. The hits include the following putative genes: VbPAM from Vibrionales bacterium, BrPAM from Burkholderia rhizoxinica, PIPAM from Pseudomonas fluorescens, KpPAM from Klebsiella pneumoniae, BsPAM from Bacillus subtilis, DtPAM from Desulfobacula toluolica., PmHAL from Paenibacillus mucilaginosus, BmHAL from Bacillus megaterium and CsHAL from Clostridium symbiosum. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Figure 23 || Genomic context of the gene from Bacillus subtilis predicted to encode an (S)-PAM. Possible functions associated with surrounding genes are labelled in red, as inferred by conserved domain search tools (top) and placed in the framework of a putative biosynthetic pathway for the synthesis and secretion of pyloricidin antibiotics from proteinogenic amino acids and the primary metabolite UDP-GlcNAc (bottom).

Figure 24 || One proposed retrobiosynthesis of the pyloricidin 5-amino-2,3,4,6-tetrahydroxyhexanoic acid precursor amino acid from the primary metabolite UDP-GlcNAc, as supported by genomic context analysis of the putative secondary metabolite gene cluster in Bacillus subtilis. The proposed steps are reduction (1), deacetylation (2), ring opening (3), dehydrogenation / oxidation (4) and epimerisation (5).

Figure 25 || A section of the sequence alignment for BagA from Streptomyces sp. and its closest tBLASTn hits. The recognisable 4aa motif of the loop lid, as inferred from other family member sequences, is enclosed within a dark blue box. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). The hits include the following putative genes: AcTAL from Amycolatopsis decaplanina, RxPAL from Rubrobacter xylophilus, BIPAL from Brevibacillus laterosporus, ScTAL from Streptomyces clavuligerus, SeTAL2 from Saccharothrix espanaensis., SmTAL from Streptomyces mobaraensis, MpTAL from Microlunatus phosphovorus, SrPAL from Streptomyces rimosus, SnPAM from Stackebrandtia nassauensis and XaTAL from Xanthobacter autotrophicus. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Figure 26 || Chemical structures of chondramide C produced by Chondromyces crocatus and jaspamide H as isolated from the marine sponge Jaspis splendidens. The structures both seem to be built from an (R)-β-tyrosine precursor (shown in blue) and are otherwise very similar (differences in jaspamide are highlighted in red).

Figure 27 || Sections of the sequence alignment for CmdF from Chondromyces crocatus and its closest tBLASTn hits. The inner active site loop residues of the hit, SgTAM from Streptomyces globisporus, as evident from the structure, are highlighted in yellow with likely loop residues from the other aligned sequences contained in the yellow box. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange).
A potential position influencing stereoselectivity in TAMs is shown within the purple box. A common [DQS]NQD motif to distinguish putative TAMs from HALs is enclosed within a dark blue box with the characteristic glutamate in the corresponding HAL sequences highlighted in light blue. The hits include other characterised enzymes: MtTAM from Myxococcus fulvus, MxTAM from Myxococcus sp., CmTAL from Cupriavidus metallidurans, RmTAL from Ralstonia metallidurans, MdpC4 from Actinomadura madurai and KedY4 from Streptothiobacillus sp. Also included in the hits are the following putative enzymes: KaTAM from Kutzneria albida, SxTAM and SxTAM2 from Streptomycetes sp., ShTAM and ShTAM2 from Streptomycetes ghanaensis, MxxTAM from Microbacterium sp., AdTAM from Amycolatopsis decaplanina, SITAM from Streptomycetes albulus, SaTAM from Salinispora arenicola, AcHAL from Aminobacterium colombiense, TaHAL from Thermanaerovibrio acidaminovorans, TiHAL from Thermovigia lienii, TiHAL from Thermosiphon africanus, ApHAL from Aminomonas paucivorans, and GxTAM from Geobacter sp.. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Figure 28 || A PyMOL visualisation of the inner active site loop (yellow) and surrounding area (green) from the crystal structure of SgTAM, PDB ID: 2RJR (top left). Each of the six surrounding panels shows a rotamer of the Q76W mutation made using the PyMOL software and all clashes made by these with the rest of the molecule are shown as red discs. ……..70

Figure 29 || The possible biocatalytic routes from ammonia and arylacrylates to all four stereoisomers of the corresponding phenylalanine derivatives. Methods yielding (S)-α-(A: PAL alone), (R)-β- (B: engineered PAM) and (R)-α- (C: PAL + deracemisation cycle) isomers have been previously reported, leaving only the missing (S)-β-form which cannot be synthesised in this way (D). ……………………………………………………………………………………………………………………………73

Figure 30 || SDS-PAGE analysis of selected fractions from the purification of EncP produced with autoinduction media (right - whole flask cell pellet, left - 1/3 flask cell pellet). The labels over the lanes indicate the marker (M), and fractions from: the column flow-through (F), the wash with buffer A (A), the wash with an AB mixture (A:B) and the elution with buffer B (B). …………………………………………………………………………………………………………………………………………………75

Figure 31 || The target biotransformation with EncP involving the previously reported substrate ((S)-α-phenylalanine 3a) and products ((S)-β-phenylalanine 2a and cinnamate 1a) of the aminomutase / ammonia lyase reactions, but in the amination direction. ……………………..76

Figure 32 || Substrates tested with EncP (α- and β-methylcinnamate 1w and 1x) and found to give no conversion under general reaction conditions. ……………………………………………………………84

Figure 33 || The dependence of β- vs. α-amination by EncP on the core electron binding energy shift (ΔCEBE) due to substrate ring substituents. ……………………………………………………………………………………………………………………………………………………………………………………………………..86

Figure 34 || EncP active site catalytic and substrate positioning residues as inferred by homology modelling based on the previously solved structure of AdmH. (a) shows the interactions and reaction mechanism hypothesised from the visualised model (b). ………..88

Figure 35 || Comparison of the side chains of functionally-equivalent, basic arginine (R) and lysine residues (K) along with the glutamate (E) and the most similarly-shaped amino acids -
glutamine (Q) and methionine (M).

Figure 36 || The effect of temperature on the overall conversion and regioselectivity of amination reactions catalysed by EncP and its rationally-designed variants. .........................90

Figure 37 || The dependence of β- vs. α-amination by EncP rationally designed variants on the core electron binding energy shift (ΔCEBE) due to substrate ring substituents. Trends for the wild-type enzyme are shown as dotted lines for comparison. .........................99

Figure 38 || The hypothesised effects of varying active site residues and para-substituents on substrate positioning and ammonia addition preference in EncP and variants. (a) represents the positioning in the wild-type enzyme, (b) shows the α-selective variants E293Q / E293M and (c) shows the β-selective variant R299K. .........................101

Figure 39 || Use of PAL in the synthesis of perindopril (R = Me) and indolapril (R = Ph) pharmaceuticals from inexpensive 2-chloro- or 2-bromocinnamic acids. .........................108

Figure 40 || Use of AvPAL for gram-scale conversion of 2-chloro- or 4-trifluoromethyl-cinnamic acid to unnatural amino acid derivatives. .............................................109

Figure 41 || The hypothesised fast (MIO-dependent) and slow (MIO-independent) amination pathways in AvPAL leading to (S)- and (R)-α-phenylalanine enantiomers respectively. ...111

Figure 42 || The dependence of β- vs. α-amination by AvPAL on the core electron binding energy shift (ΔCEBE) due to substrate ring substituents. The plot on the left includes all reaction values, whereas the plot on the right includes only reactions were β-amino acid products are observed as more than traces. .............................................115

Figure 43 || The percentage composition of AvPAL-catalysed biotransformations with β-forming substrates and cinnamate over the course of 144h. .............................................117

Figure 44 || The MIO-tyrosine-catalysed enzymatic processes required for PAL (1), (S)-PAM (2) and (R)-PAM (3) activities starting from (S)-α-phenylalanine. All processes involve deamination of the starting material via formation of an amino-MIO adduct and tyrosine-mediated proton abstraction. Release of ammonia from the MIO results in PAL activity. (S)-PAM activity is characterised instead by amino-MIO amination at the β-carbon and tyrosine protonation at the α-carbon on the same face as the preceding deamination and deprotonation steps. Alternatively 180° rotation of the cinnamate intermediate in the active site followed by amino-MIO amination at the β-carbon and tyrosine protonation at the α-carbon on the opposite face gives (R)-PAM activity. .............................................120

Figure 45 || (top) Overlaid active site residues from the crystal structures of TcPAM (PDB: 3NZ4) and AvPAL (PDB: 3CZO). (middle) Protein sequence alignment sections of a selection of distantly related Class I Lyase-like Enzymes. The sequences are: TcPAM from Taxus canadensis, PcPAL1 from Petroselinum crispum, RtPAL (bifunctional PAL / TAL) from Rhodosporidium toruloides, AvPAL from Anabaena variabilis, StlA (PAL) from Photorhabdus luminescens and RcTAL from Rhodobacter capsulatus. Consensus sites, homologous to those in the active site structure are also shown (highlighted in yellow). The percentage

11
sequence identity with respect to the TcPAM sequence is shown in the final column of the alignment table for each entry (Seq. ID). (bottom) A cladogram showing the hypothesised evolutionary history of the aligned enzymes as inferred from sequence similarity. Possible amino acid substitutions in the active site residues are used to label the branches where they are most likely to have occurred.

Figure 46 || The effect of temperature and time of incubation on overall conversion of AvPAL-catalysed amination of 5 mM ortho-fluorocinnamate.

Figure 47 || The effect of substrate loading on overall conversion of AvPAL-catalysed amination of fluorocinnamates.

Figure 48 || The effect of batch addition of 5 mg mL\(^{-1}\) biocatalyst on overall conversion of AvPAL-catalysed amination of fluorocinnamates. Each catalyst addition event is indicated by a black arrow.

Figure 49 || The effect of varying biocatalyst loading on overall conversion of AvPAL-catalysed amination of fluorocinnamates.

Figure 50 || (top) Carbon and (bottom) proton NMR spectra of the isolated 3-fluorophenylalanine product of the AvPAL-catalysed amination reaction (supporting data in chapter 5, section 5.3).

Figure 51 || An example of the FASTA format for a text file containing a sequence identifier encp-pET28a gene-vector construct containing the codon optimised DNA sequence encoding EncP.

Figure 52 || The oligonucleotide primer sets used to introduce the three active site substitutions into the coding sequence for EncP as aligned with the section of plasmid DNA specific to each. The codon for each specific amino acid position to be mutated is highlighted in yellow with the base changes in the primers coloured in red.

Figure 53 || The oligonucleotide primer sets used to introduce the three active site substitutions into the coding sequence for AvPAL as aligned with the section of plasmid DNA specific to each. The codon for each specific amino acid position to be mutated is highlighted in yellow with the base changes in the primers coloured in red.

Figure 54 || HPLC chromatograms showing the separation of authentic standards of amino acid enantiomers using the methods from the table above.

Figure 55 || HPLC chromatograms showing the separation of authentic standards of amino acid regioisomers and enantiomers using the low temperature method for the three sets of compounds with which this was found to work.

Figure 56 || The amino acid sequences of the enzymes EncP from Streptomyces maritimus and AvPAL from Anabaena variabilis in FASTA format.

Figure 57 || The codon optimised nucleotide sequences encoding the enzymes EncP from
Streptomyces maritimus and AvPAL from Anabaena variabilis in FASTA format. ........167
List of Tables

Table 1 || Examples of characterised class I lyase-like enzymes whose molecular functions have been misannotated in the Universal Protein Database (UniProt - entries retrieved May 2014). .................................................................41

Table 2 || The effect of ammonium donor salt on the conversion of cinnamate and $\beta:\alpha$ ratio of products in EncP-catalysed biotransformations. .................................................................76

Table 3 || The effect of ammonium sulphate concentration on the conversion of cinnamate in EncP biotransformations. .................................................................78

Table 4 || The effect of reaction buffer pH on the conversion of EncP biotransformations. ..78

Table 5 || The effect of incubation temperature on the conversion of cinnamate and $\beta:\alpha$ ratio of products in EncP-catalysed biotransformations. .................................................................79

Table 6 || EncP-catalysed amination of cinnamate and various ring-substituted fluorocinnamates. .................................................................80

Table 7 || EncP-catalysed amination of additional ring-substituted halocinnamates. ..........81

Table 8 || EncP-catalysed amination of ring-substituted nitro- and methoxycinnamates. ...82

Table 9 || EncP-catalysed amination of miscellaneous ring-substituted cinnamates...83

Table 10 || Amination of various arylacrylic acids to the corresponding amino acids by the EncP-R299K rationally designed enzyme variant. .................................................................93

Table 11 || Amination of various arylacrylic acids to the corresponding amino acids by the EncP-E293Q rationally designed enzyme variant. .................................................................95

Table 12 || Amination of various arylacrylic acids to the corresponding amino acids by the EncP-E293M rationally designed enzyme variant. .................................................................97

Table 13 || Time course study of the amination of 2-methylcinnamate to the 2-methyl-$\beta$-phenylalanine by the EncP-R299K variant. .................................................................105

Table 14 || Time course study of the amination of 3-methylcinnamate to the 3-methyl-$\beta$-phenylalanine by the EncP-R299K variant. .................................................................105

Table 15 || Time course study of the amination of 4-methylcinnamate to the 4-methyl-$\beta$-phenylalanine by the EncP-R299K variant. .................................................................105

Table 16 || AvPAL-catalysed amination of a panel of ring-substituted cinnamates......112
Table 17 || AvPAL-catalysed amination of ring-substituted cinnamates showing detectable levels of β-amino acid side product. ..............................................................114

Table 18 || AvPAL-catalysed deamination of ring-substituted (S)-α-amino acids and amination of the corresponding cinnamate derivatives. ........................................118

Table 19 || Amination of cinnamate and a selection of 2- and 4-ring-substituted derivatives catalysed by two AvPAL single active site variants (F107C and R317K). ...............125

Table 20 || Amination of cinnamate and a selection of 4-ring-substituted derivatives catalysed by an AvPAL single active site variant (Q311M). ........................................126

Table 21 || Amination of cinnamate and a selection of 4-ring-substituted derivatives catalysed by an AvPAL double active site variant (F017C / Q311M). .........................127

Table 22 || AvPAL-catalysed amination of fluorocinnamates at 5 mM concentration. ....127

Table 23 || The effect of varying pH on overall conversion of AvPAL-catalysed amination of 3-fluorocinnamate. ..............................................................134

Table 24 || Effect of varying substrate concentration on intensified, small scale aminations of 3-fluorocinnamate by AvPAL. ..............................................................136

Table 25 || Reaction parameters giving at least 95% conversion of monosubstituted halocinnamates in AvPAL-catalysed biotransformations. .................................138

Table 26 || PCR reaction set up to a total reaction volume of 50 μL for mutagenesis of EncP and AvPAL encoding genes in a 0.2 mL Eppendorf tube. .....................................150

Table 27 || Standard PCR reaction conditions for all mutagenesis reactions as programmed into the Eppendorf Mastercycler Gradient PCR machine. .................................151

Table 28 || Diagnostic fragment sizes for corresponding double digestions of the avpal-pET-16b and encp-pET-28a constructs. ..............................................................153

Table 29 || DNA sequences of the oligonucleotides required to prime sequencing of DNA subcloned into pET vectors. ..............................................................154

Table 30 || Additional parameters of HPLC analyses of enzyme biotransformations on a non-chiral phase. ..............................................................157

Table 31 || Additional parameters of the HPLC analyses of enzyme biotransformations on a chiral phase. ..............................................................159
Abstract

This thesis reports the investigations of class I lyase-like enzymes with particular focus on aiding discovery of family members, and variants thereof, for biotechnological applications, such as biocatalytic synthesis of value-added compounds.

Chapter 2 details computational investigations of class I lyase-like enzymes based on sequence and structural data. Using an initial set of structurally- and biochemically-characterised class I lyase like enzymes, patterns and relationships were identified and used to annotate publically-available sequences. This allowed the discovery of potential enzyme-coding genes for use in areas of biotechnology, e.g. as biotherapeutics for the treatment of cancer or as biocatalysts for the production of valuable unnatural amino acids. The search also aided elucidation of putative biosynthetic pathways, including one for a narrow spectrum antibiotic, and highlighted possible mechanisms of functional evolution within the family.

In chapter 3 the characterisation and engineering of the bacterial ammonia lyase EncP for the production of (S)-β-amino acids is reported. This enzyme, although previously reported in the literature, had ever been investigated in a biocatalytic context. Creation of a biotransformation method allowed the broad substrate scope and clear enantiopreference of the enzyme to be uncovered. By combining electronic effects of substrates with structural inference, it was possible to create enzyme variants with shifted regioselectivity, including EncP-R299K - a biocatalyst catalysing the (S)-β-selective amination of a range of acrylic acids. This result is complementary to previous work as the (S)-β-products were not previously obtainable using already characterised ammonia lyase biocatalysts.

Chapter 4 is about the use of another biocatalyst, AvPAL, to perform preparative scale synthesis of (S)-α-amino acids. Upon investigation of the substrate scope of this enzyme, imperfect enantio- and regioselectivity were uncovered. Further investigation of the product mixtures revealed that the enzyme had unreported mutase-like side activity, pointing to evolutionary mechanisms of functionalisation, as relating to chapter 2. Unfortunately engineering efforts to augment these activities were relatively unsuccessful. By choosing optimal substrates and reaction conditions, a biotransformation method was developed, allowing industrially relevant space time yields (up to 60 g L⁻¹ d⁻¹) to give crude isolated amino acids in sufficient purity.

Chapter 5 provides further details on exact computational and experimental methods used throughout the investigations.
Declaration

The author of this thesis hereby declares that:

no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgements

First and foremost I would like to thank Professor Nicholas Turner for giving me the opportunity to pursue this programme of study, for his support and guidance but also for allowing me the freedom to choose my own direction and avenues of investigation.

I would like to thank Dr. Fabio Parmeggiani and Mr. Syed Ahmed for all the help with project ideas, experimental work and collaboratory projects. Thanks must also go to Drs. Rachel Heath, Bas Groenendaal and Sarah Lovelock for their preceding work in the area, which influenced the direction and focus of my project.

I must also express my gratitude to many members of the Turner-Flitsch lab. and Manchester Institute of Biotechnology for creating a fun and supportive working environment.

I also acknowledge the generous financial support of the European Union's 7th Framework Programme and members of the KYROBIO industrial-academic consortium for their collaboration and input.
1. Introduction
1.1 The Class I Lyase-like Enzyme Family

Enzymes of the class I lyase-like family are a group of structurally and mechanistically related proteins capable of catalysing the overall elimination of ammonia from the aromatic amino acids histidine, phenylalanine or tyrosine. Enzymes which proceed no further than this initial deamination are said to have ammonia lyase (AL) activity and give free ammonia and the corresponding ary lacrylate (urocanate, cinnamate or coumarate respectively) as products. Some family members are also able to direct the subsequent reamination of the unsaturated intermediate at the neighbouring carbon, thus completing the isomerisation of the α-amino acid substrate to give the β-regioisomer. These enzymes with aminomutase (AM) activity are only known to be specific for either phenylalanine or tyrosine.\(^2\)

---

**Figure 1** The five enzyme activities found in members of the class I lyase-like family including histidine ammonia lyase (HAL), phenylalanine ammonia lyase (PAL), phenylalanine aminomutase (PAM), tyrosine ammonia lyase (TAL) and tyrosine aminomutase (TAM).

---

All solved structures of class I lyase-like enzymes have been shown to be homotetramers of identical polypeptides with almost exclusively alpha-helical propensity.\(^3\) As such, these enzymes are thought to have four active sites, each composed of residues from three neighbouring monomers, complete with inner and outer active site loops which control substrate entry / exit.\(^4,5\) The unique chemistry of this class of biological catalysts is mediated through an active site 5-methylene-3,5-dihydroimidazol-4-one (MIO) moiety.\(^6,7\) This family-specific feature is formed via the posttranslational, autocatalytic cyclisation and double dehydration of a semi-conserved [AT]SG tripeptide.\(^8,9\) The MIO acts in conjunction with an activated catalytic tyrosine, present on the inner active site loop, and together these allow...
abstraction of the α-amine and β-proton (when the loop lid is closed). Release of ammonia from the MIO and loop opening allows release of the ammonia lyase product. Retention of these as intermediates allows the mutase product to be formed without release and rebinding. In both cases the MIO and tyrosine are regenerated as part of the catalytic cycle.\textsuperscript{10–14}

![Figure 2](image_url)

\textbf{Figure 2} || The proposed catalytic mechanism of MIO / tyrosine-mediated ammonia elimination or amination in class I lyase-like enzymes. The enzyme is shown in black, the substrate in blue and the transferred groups in red. For α-amino acids $R_1 = \text{aryl}$ and $R_2 = \text{carboxyl}$. For β-amino acids $R_1 = \text{carboxyl}$ and $R_2 = \text{aryl}$.

\subsection*{1.1.1 Histidine Ammonia Lyases}

The histidine utilization pathway (hut) is an integral part of metabolism in bacteria, allowing scavenging of two nitrogen atoms (in the form of ammonia) from the starting amino acid by mediating its overall conversion to glutamate. The first reaction and initial ammonia-releasing step in this pathway is catalysed by HutH, a histidine ammonia lyase which allows the deamination of histidine to yield urocanate which is then further oxidised.\textsuperscript{15} As histidine utilisation is a primary metabolic pathway in bacteria, all members of this kingdom are predicted to have HutH orthologues. This has also led to the hypothesis that HAL is the ancestral function of all class I lyase-like enzymes, with other enzyme activities evolving subsequently\textsuperscript{16,17} as a means of siphoning other proteinogenic amino acids (namely phenylalanine and tyrosine) into specific secondary metabolic pathways.
Figure 3 | The primary metabolic pathway in bacteria by which histidine is converted to glutamate. This process is initiated by the HAL HutH with subsequent reactions mediated through additional enzymes in the Hut pathway. Additions to the initial histidine scaffold are shown in red.

HAL enzymes are also found in archaea and many subgroups of eukaryote,\textsuperscript{18} implying their ancestral importance in central metabolism. Whilst HALs are retained in many of these organisms, they are not always essential for normal functioning. An example of this is the inborn error of metabolism in humans histidinemia, wherein HsHAL (from \textit{Homo sapiens}) is faulty and does not perform efficient catabolism of histidine.\textsuperscript{19} This genetic disease has been shown to be relatively benign,\textsuperscript{19} indicating the reduced importance of HAL activity in human metabolism. Land plants and members of the fungal subkingdom Dikarya are notable examples of eukaryotic multicellular organisms which lack HAL enzymes, probably due to gene loss.\textsuperscript{18} Oddly these are also the two clades where PAL, TAL and in some cases PAM enzymes are found.

1.1.2 Phenylalanine Ammonia Lyases

The phenylalanine degradation product cinnamate acts as a precursor in the biosynthesis of a variety of natural products. In all cases the formation of this starting material is catalysed by PAL enzymes specific to the secondary metabolic pathway of each in a selection of specialised organisms. In bacteria these metabolites have been shown in some cases to have antibacterial activity, such as the slightly modified cinnamamide in \textit{Streptomyces verticillatus}.\textsuperscript{20} Another example is the production of a stilbene antibiotic by \textit{Photorhabdus luminescens} using cinnamate and the transaminated ketoacid analogue of leucine as starting materials.\textsuperscript{21} In other cases the cinnamate is further modified to a benzoyl-CoA module to prime the synthesis of polyketide-type products. Examples of these include the
soraphens produced by Sorangium cellulosum,\textsuperscript{22} as well as the wailupemycin and enterocin antibiotics from Streptomyces maritimus.\textsuperscript{9,23}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{A selection of secondary metabolites biosynthesised from the cinnamate product of PAL-mediated deamination of phenylalanine. The parts of the precursor phenylalanine and derived natural products cognate to the cinnamate constituents are highlighted in red.}
\end{figure}

In land plants the phenylpropanoid pathway underlies the biosynthesis of numerous phenolic and polyphenolic compounds.\textsuperscript{18,24} An important example of this is the biopolymer lignin.\textsuperscript{18} The pathway normally proceeds via the hydroxylation of cinnamate first at the 4-position to give \textit{para}-coumarate. Some species of monocot plants, however, contain ammonia lyase isoforms capable of deaminating both phenylalanine (to give cinnamate) and tyrosine (to give coumarate)\textsuperscript{25} as routes to the first and second chemicals in phenylpropanoid biosynthesis. Enzymes with similar bifunctionality are also found in the fungal kingdom,\textsuperscript{3,26} although the biosynthetic pathways with which they may be associated are yet to be uncovered.

1.1.3 Tyrosine Ammonia Lyases

As well as the plant and fungal PAL enzymes with additional TAL activity discussed in the previous section, there also exist ammonia lyases with strict tyrosine specificity in species of bacteria. Two examples are reported of these enzymes giving \textit{para}-coumarate as a precursor to antibiotic biosynthesis. The saccharomicin class of bactericidal agents are characterised by a heptadecasaccharide structure linked to a terminal tauryl caffeate. The
Caffeic acid precursor in this instance represents a 3-hydroxylation product of coumarate, which in turn has been shown to be produced from tyrosine by the ammonia lyase Sam8. These compounds have been shown to have activity against multiply antibiotic resistant strains of *Enterococcus* and *Staphylococcus aureus*. In a strain of *Streptomyces*, another example of TAL in actinomycetes was characterised in connection with the production of antibacterial and antifungal bagremycins. Here, following deamination of tyrosine by the enzyme (BagA) the coumarate undergoes decarboxylation and O-linkage to an aminomethylbenzoate in an undetermined order.

**Figure 5** Three examples of utilisation of the TAL product para-coumarate in bacterial natural products and proteins. The parts of the precursor tyrosine and derived natural products cognate to the cinnamate constituents are highlighted in red. With saccharomicins the R group represents a complex branched polysaccharide.

One unusual fate of a tyrosine-derived ammonia lyase product is the incorporation of coumarate as the chromophore of photoactive yellow protein (PYP) in species of purple phototrophic bacteria. Upon identification of the putative PYP-encoding gene in the whole genome sequence of *Rhodobacter capsulatus*, flanking acyl ligase and class I lyase-like sequences were found. It was shown via cloning and characterisation of this second open
reading frame that the enzyme was a TAL. These data revealed that the modified cysteine in
the light absorbing region of PYP was derived from the acrylate product of the RcTAL
deamination and incorporated into the polypeptide by the remaining protein (para-coumarate
ligase or pCL). It is proposed that this chromophore allows the protein to function as a light-
sensing molecule, possibly to regulate phototrophic processes in its host organism.

1.1.4 Phenylalanine Aminomutases

There exist only two examples of secondary metabolites where an incorporated β-
phenylalanine is known to be the product of a PAM reaction. Interestingly these natural
products are found to have very different structures, are present in distantly related
organisms and contain opposite enantiomers of the β-amino acid building block.

Andrimid is a potent, broad-spectrum antibiotic of clinical interest produced by many species
of bacteria. As such, its effectiveness as an antimicrobial agent is predicted to give
production strains a selective advantage. Andrimid acts as a nanomolar inhibitor of
prokaryotic acetyl-CoA carboxylase - an enzyme essential for fatty acid biosynthesis in
bacteria. The molecule has been shown in Pantoea agglomerans to be assembled via a
mixed non-ribosomal peptide / polyketide synthase from various precursor molecules. One
of these, (S)-β-phenylalanine, has been shown to be the product of the aminomutase AdmH
which acts to isomerise primary metabolic phenylalanine as a gateway reaction for andrimid
biosynthesis.
The only two natural products known to contain aminomutase-derived β-phenylalanine. The incorporation of (S)-β-phenylalanine by Pantoea agglomerans into the antibiotic andrimid is shown on the left. The synthesis of taxol from (R)-β-phenylalanine in various species of yew tree (Taxus spp.) is shown on the right.

Figure 6

Taxol is one of the most widely used anti-cancer agents in the world, acting as a potent cytotoxic agent. This activity is conferred via the binding and stabilisation of microtubules in eukaryotic cells, disrupting the dynamic instability of the cytoskeleton, thus halting the normal cell cycle and preventing proliferation. The toxin is present in the leaves, stem and seeds of various species of yew tree (Taxus spp.) from which it can be extracted for medicinal use. Taxol is comprised of a complex caged bacchatin structure linked to an α-hydroxylated, N-benzoylated (R)-β-phenylalanine, which is in turn derived from the mutase reaction of PAM orthologues in each species.

1.1.5 Tyrosine Aminomutases

Enzymes with detectable TAM activity have only ever been characterised in the eubacterial kingdom. As with the other aminomutase enzymes in the family, there exist enzymes which perform the enantiocomplementary rearrangement of (S)-α-tyrosine to give either (S)- or (R)-configured β-products. TAM derived (R)-β-tyrosine is only known to be incorporated into one suite of natural products: the chondramide cytotoxic agents produced by Chondromyces croatus. Further mechanistic studies of the enzyme responsible (CmdF) have, however, revealed that the reaction can proceed with either retention or inversion of stereochemistry, with the detection of (S)-β-tyrosine as a minor product upon incubation with the common starting material. The closest relative of the Chondromyces enzyme, CmTAL from...
Cupriavidus metallidurans, has also been shown to allow \((R)\)-\(\beta\)-readdition of ammonia despite the discovery that the major product of the reaction with this enzyme is para-coumarate. As of yet, this enzyme has not been linked to biosynthesis of a particular natural product in its host organism through either the major arylacrylate or minor amino acid product.

![Figure 7](image-url)

**Figure 7** A summary of the tyrosine conversions of characterised enzymes known to have TAM activity. Predominant activities are shown thick bold arrows whereas minor side activities are shown with thinner arrows.

Other TAM enzymes have been shown to give the \((S)\)-product in association with known secondary metabolic pathways. MfTAM and MxTAM, both from the Myxococcus genus related to Chondromyces, have been demonstrated to be essential to the production of a large \((S)\)-\(\beta\)-tyrosine-containing nonribosomal peptide called myxovalargin. The more distantly related enzymes SgTAM and MdpC4 (from Streptomyces globisporus and Actinomadura madurae) are both involved in biosynthesis of enediyne antitumor antibiotics, providing the precursor for the alkoxy / hydroxyl / chloro tri-substituted amino acid moiety present in this class of natural products. Using a retrobiosynthetic approach a further enzyme was discovered, relating to the production of another enediyne compound kedarcidin in a species of Streptoalloteichus. Interestingly the enzyme KedY4 was found to be \((R)\)-selective, unlike its orthologues, and to accept 2-aza-tyrosine in lieu of the proteinogenic analogue. Both of these observations were consistent with the unusual heterocyclic \(\beta\)-amino acid portion of kedarcidin. This is also the only example of a class I lyase-like enzyme with a natural substrate that is not histidine, phenylalanine or tyrosine.

1.2 Biomedical Applications of Class I Lyase-like Enzymes

Many diseases are characterised by or result in changes in flux through biosynthetic and catabolic pathways, causing an imbalance in metabolites. Examples include inborn errors of
metabolism\textsuperscript{37–39} and uncontrolled cell proliferation in cancerous tissue.\textsuperscript{40–42} One of the most desirable ways to counteract such deviations in metabolic profile would be the introduction of molecules with enzymatic activity capable of removing or producing compounds of medical significance \textit{in vivo}. Enzymes should be well suited to this application as they are highly selective, so as to ensure only specific metabolites are targeted, and are of biological origin so they can be produced renewably and work well under physiological conditions.\textsuperscript{42,43} Members of the class I lyase-like family, particularly ammonia lyases, are well placed to be developed into viable therapeutic enzymes, as they accept primary metabolic aromatic amino acids, whose misregulation is known to be associated with various diseases.\textsuperscript{39,42}

\textbf{1.2.1 Amino Acid Depletion Therapy}

Amino acid depletion therapy is a potential strategy to combat tumour growth in cancer patients. It relies on the fact that certain essential amino acids, not biosynthesised in human cells, are utilised extensively by fast replicating tumour cells compared to normal, resting cells. Some tumour subtypes even display dependency on specific non-essential amino acids, having lost the ability to metabolise them effectively.\textsuperscript{42} A relevant example is the decrease in metastatic phenotype of B16BL5 melanoma upon depletion of the essential amino acid phenylalanine and the non-essential amino acid tyrosine\textsuperscript{41} (although TALs have not yet been used as cancer therapeutics as PALs have). In the 1970s a fungal PAL was shown to prevent growth of abnormal lymphocytes both \textit{in vitro}\textsuperscript{44} and in leukemic mice\textsuperscript{45} by removing free phenylalanine from the available substrate pool. Although promising, the enzyme was found to be highly immunogenic and unstable, being cleared effectively from tumour cells even after repeated injections.\textsuperscript{46} Mast cell sarcomas have also been shown to be sensitive to treatment with a HAL \textit{in vitro}\textsuperscript{40} via depletion of the essential amino acid histidine.

The problems associated with these kinds of therapies may have been due to the limited number of class I lyase-like enzymes available for use and the infancy of biomolecular engineering techniques to improve the therapeutic properties of proteins. Nowadays these do not pose such an issue as demonstrated by ongoing work of engineering reduced immunogenicity, increased stability and substrate specificity of enzymes for cancer therapy\textsuperscript{42} and a recent patent application (2009) regarding the use of more stable prokaryotic PALs as cancer therapeutics.\textsuperscript{47}
1.2.2 Treatment of Phenylketonuria

Phenylketonuria or PKU is an inborn error of metabolism characterised by high levels of phenylalanine in the blood (hyperphenylalaninemia) and decreased levels of tyrosine. Ordinarily in humans, tyrosine is biosynthesised from the essential amino acid phenylalanine via the enzyme PAH (phenylalanine 4-hydroxylase). This process requires the cofactor tetrahydrobiopterin which enables molecular oxygen to be used for the oxidation reaction of PAH. The cofactor is then recycled by dihydrobiopterin reductase (DHBR). A fault in either the hydroxylation or cofactor regeneration steps of this pathway are known to lead to PKU. The heightened blood phenylalanine has been shown to cause mental retardation if not treated from birth. The current most effective treatments are dietary, with intake of phenylalanine being greatly restricted and tyrosine being supplemented in food. Even though the symptoms are thought to be developmental, therapy has been shown to be most effective if continued into adulthood.

![Possible causes, effects and treatment of the inborn error of metabolism phenylketonuria (PKU).](image)

As phenylalanine cannot be biosynthesised by humans and must be ingested, ineffective removal of this amino acid through tyrosine biosynthesis causes levels of these amino acids to be abnormally high and abnormally low respectively. As a phenylalanine-metabolising enzyme, PAL has been extensively investigated as a potential therapeutic to be introduced into patients for the removal of excess phenylalanine. Development of this therapy has advanced from initial use of a fungal PAL, which was effective but cleared quickly by the immune system, to PEGylated forms of the enzyme and use of a more stable, less immunogenic cyanobacterial PAL. It has been reported that clinical trials with this enzyme are currently underway.
1.3 Class I Lyase-like Enzymes as Biocatalysts

Enzymes have been deemed to be desirable as biologically-derived catalysts in the organic synthesis of various pharmaceutical, agro- and fine chemical products.\textsuperscript{49–51} Enzymatic transformations can often be performed in mild aqueous conditions with low temperature and mitigate the need for petrochemicals such as organic solvents. They are also a greener alternative to toxic and expensive metal and organocatalysts and can be produced sustainably via fermentation from renewable feedstocks.\textsuperscript{50} Use of biocatalysts is often driven by their tendency to be highly stereo-, regio- and chemoselective\textsuperscript{51} and the ability of some enzymes to catalyse reactions not possible through synthetic methods. PAL and PAM enzymes of the class I lyase-like family are examples of this, as they allow the selective amination of simple arylacrylates to yield high value unnatural amino acids under an excess of ammonia.\textsuperscript{2,43} This is a process which cannot be achieved through simple chemical methods and the lack of need for a cofactor makes the process industrially-relevant.\textsuperscript{52}

![Diagram](image.png)

**Figure 9** || The reversibility of the natural reactions (thin arrows) of PAL and PAM, whereby the equilibrium can be pushed toward amino acid products starting with cinnamate and excess ammonia (thick arrows).

Phenylalanine aminomutases from *Taxus canadensis* (TcPAM) and *Pantoea agglomerans* (AdmH) have also been surveyed for their ability to isomerise unnatural phenylalanine derivatives to give the β-regioisomers. The *Taxus* enzyme was found to accept a range of starting materials including the unusual (S)-β-styrylalanine to give (R)-configured products.\textsuperscript{53,54} Differences in the catalytic efficiency of the isomerase for various substrates pointed to discrepancies in the accommodation of each in the active site. This data was used
to design an enzyme variant PAmeLA (TcPAM-L104A), which showed a marked decrease in $K_M$ for 3-methylphenylalanine, a compound disfavoured relative to other derivatives by the wild-type enzyme. Similar investigations of AdmH revealed comparable yet different fluctuations in production of (S)-β-phenylalanines via transfer of the amino group. This was explained by the alternative active site architectures of the two aminomutases as had been previously shown in superimposed crystal structures. In this study the electronic and resonance effects of substrate ring-substituents were also used to rationalise, not only substrate binding but also the incidence of α-deamination / β-reamination thus explaining the different conversions from the starting materials to the acrylic acid or β-amino acid products. Issues with the isomerisation reactions with respect to production of β-amino acids include maximum possible conversion of only 50% and requirement of enantiopure starting material. This second point was addressed with the (R)-selective enzyme when it was shown that racemic phenylalanine derivatives could be used as starting material via dynamic kinetic resolution aided by the addition of alanine racemase.

**Figure 10** A selection of (S)- and (R)-β-amino acids synthesised by the enantiocomplementary aminomutases (AdmH and TcPAM respectively) via the isomerisation of corresponding (S)-α-regioisomers.

### 1.3.1 Unnatural Amino Acids

Unnatural amino acids, or non-proteinogenic amino acids, are a diverse range of compounds of ever increasing importance in medicinal chemistry. Despite the name, there are many
examples of these types of compound present in the secondary metabolism of various organisms. Their value is demonstrated by their synthetic utility, particularly for pharmaceutical products. The 22 proteinogenic amino acids are all of a single stereoconfiguration and all have the amino group present on the α-carbon. Unnatural amino acids, however, can be more diverse, ranging from opposite enantiomers, structural and regioisomers or substituted derivatives of a standard amino acid to unusual synthetic structures containing an amino and carboxylic acid group.

Of particular interest in this respect are β-regioisomers of amino acids and their various derivatives. This group of non-proteinogenic amino acids is characterised by the presence of the amino moiety on the β-carbon of the structure. They have received much attention as an emerging class of important targets, particularly in pharmaceutical and natural product chemistry. β-amino acids with larger R groups, and thus chiral centres, are useful precursor molecules. An example of this is the synthesis of single enantiomers of 4-phenylazetidin-2-one, a component of serine and cysteine protease inhibitors, via dehydration of optically pure β-phenylalanine. β-amino acids are also more widely applicable as building blocks for peptidomimetic therapeutics. Polymers of these, known as β-peptides, display great potential as peptide mimics, due to their ability to form protein-like conformations spontaneously in water. In this way, peptidomimetics are able to achieve many of the protein-protein interactions essential for regulation of biological processes.

Synthetic β-peptides have been designed as analogues of bacterial host-defence molecules and have shown significant antimicrobial activity. There are a large number of structures possible from the scaffold of a β-amino acid monomer, including substitution on the main chain α-carbon as well as on side chain residues. This increases the potential for molecular design with this class of compounds, allowing more versatile structures to be built up from simple peptide precursors. Protein analogues containing β-amino acids are also more proteolytically stable in vivo as demonstrated by the creation of β- / α-hybrid peptides.
Figure 11 // Examples of unnatural amino acid derivatives of phenylalanine (left) and the general back bone structures of α- and β-peptides (right).

Phenylalanine and tyrosine (which is itself a 4-hydroxylated form of phenylalanine), in particular, have many non-standard derivatives of chemical and biological relevance. The variety of structures possible from the phenylalanine template is evident from the structure: two possible enantiomers and either the β- or α-position plus 5 ring carbons, which can be substituted with various chemical groups. As such derivatives of phenylalanine are found in natural secondary metabolites, in chiral pools for drug synthesis\textsuperscript{65,74} and have more recently found use as chemical tools for the study of biology.\textsuperscript{61,75}

1.3.1.1 Phenylalanine Derivatives as Chiral Building Blocks

As chiral organic molecules, amino acids are useful compounds in the synthesis of many pharmaceuticals and, as such, are used as chiral precursors from which lead compounds are generated via combinatorial chemistry.\textsuperscript{65} Use of unnatural amino acids broadens the potential properties of a synthesised compound with respect to pharmacokinetics, pharmacodynamics and bioavailability. As analogues of standard amino acids, some synthesised phenylalanine derivatives display similar functional properties to their natural counterparts. Examples of these include substrate mimic peptides containing unnatural amino acids which act as inhibitors for proteolytic enzymes. As specific proteases are utilised by many viruses to mediate infection a large number of widely-used antiviral drugs are designed and manufactured in this way.\textsuperscript{76}

More specific examples include 4-substituted Phe-Pyr-CN compounds, which act by inhibiting dipeptidyl peptidase 4 (dpp4), an enzyme implicated in diabetes and tumour cell proliferation, as well as the HIV infection process. The mechanism of this inhibition is evident from crystal structures of the enzyme with 4-iodophenylalanyl-(S)-2-cyanopyrrolidine.\textsuperscript{77} Binding studies have shown that the presence of 4-iodophenylalanine in this compound
increases the affinity of the enzyme for the drug significantly when compared with the unsubstituted version. Affinity is further improved with the presence of para-substituted amino acids as large as biphenylalanine. Similarly a range of cathepsin A (CatA) inhibitors show better pharmacological properties on substitution of 4-fluoro- for 2-methyl-(S)-β-phenylalanine. CatA is a serine carboxypeptidase shown to be implicated in cardiovascular diseases such as cardiac hypertrophy. As such, development of inhibitory compounds has the potential for new treatments for these pathologies.

**Figure 12** // Six examples of pharmaceutical compounds synthesised from phenylalanine derivatives (the phenylalanine component of each is highlighted in green).

Quinapril is a well-established antihypertensive drug also used to treat congestive heart failure. Upon absorption it is converted to quinaprilat which acts to inhibit angiotensin converting enzyme (ACE), preventing further production of active angiotensin hormone. The drug contains 1,2,3,4-tetrahydroisoquinoline-(S)-carboxylic acid, a 2-substituted, N-cyclised from of (S)-α-phenylalanine. Melphalan is a 4-substituted (S)-α-phenylalanine derivative used as a chemotherapeutic agent. It acts by binding covalently to DNA bases and preventing DNA replication and thus cell proliferation in cancerous tissue. Oddly, despite the fact that the phenylalanine portion of the molecule is not thought to be involved in contacting the DNA, use of medphalan (the enantiocomplementary form of melphalan) is found to be far less effective against cancer cells. As well as medphalan, the antidiabetic drug nateglinide is an example of a therapeutic molecule synthesised from an (R)-α-phenylalanine precursor. It acts by inhibiting ATP-sensitive potassium channels in the presence of insulin, thus stimulating further insulin release in pancreatic beta-cells.
1.3.1.2 Phenylalanine Derivatives as Chemical Biology Tools

Recent advances in synthetic biology mean that it is now possible to incorporate unnatural amino acids into proteins as they are being produced in the cell. This breakthrough has allowed non-proteinogenic compounds to be used as chemical probes to study cell and protein biochemistry.\textsuperscript{75} One example of this is a novel approach to protein spin labelling via incorporation of 4-acetyl-(S)-α-phenylalanine site-specifically. This allows the unnatural substituent to be reacted selectively with a hydroxylamine-functionalised spin label, creating a ketoxime-link between the two. The new labelling system shows advantages in electron paramagnetic resonance (EPR) detection of local conformational protein changes over cysteine-mediated disulphide-linked labels.\textsuperscript{80}

Other examples include studies of protein substrates of tyrosine kinase or tyrosine phosphatase enzymes. Protein phosphorylation is a process central to cell signalling / proliferation and is often misregulated in cancerous cells.\textsuperscript{81} As phosphotyrosine (Ptr) is short-lived in vivo. It is difficult to probe the changes tyrosine phosphorylation elicits in local protein conformation and intermolecular interactions. Several synthetic derivatives have been investigated as potential phosphotyrosine mimics, including 4-phosphonomethyl-, 4-phosphonodifluoromethyl- and 4-phosphonobromomethyl-(S)-α-phenylalanine.\textsuperscript{61,81} As such, the production of semisynthetic proteins containing a non-hydrolysable analogue of Ptr is an instrumental method in the study of cellular processes. Incorporation of 4-benzoyl-(S)-α-phenylalanine into an SH2 (Src Homology 2) protein domain, for example, not only mimics the phosphorylated form of a surface tyrosine but also allows phototrapping of otherwise transient binding partners for the modified interaction interface.\textsuperscript{82}
Unnatural amino acid-containing proteins are also a novel tool for inducing degradation of a particular cellular component. Studies with 2-nitro-(S)-α-phenylalanine revealed that inclusion of this amino acid in a model peptide allowed cleavage of the backbone immediately N-terminal to the site in response to light (300-365nm). This site-specific photochemistry was also shown to allow cleavage of T4 lysozyme into two individual polypeptides. Technologies such as this may allow inactivation of proteins in vivo or more likely cleavage of precursor peptides in to active fragments.83

1.4 Objectives

Reported study of class I lyase-like enzyme has contributed to knowledge of natural product biosynthesis as well as enabling development of novel biotherapeutics and demonstration of new green catalysts for the production of value-added chemicals. As such, the objectives of this thesis are to expand research into these areas.

As only certain family members (namely PAL and PAM enzymes) have been shown to be useful for biomedicine and biocatalysis, there is a need to discover new enzymes, particularly of these classes to expand the scope of research in these areas. For this the

---

**Figure 13** Three examples of site-specific incorporation of unnatural phenylalanine derivatives into various proteins / peptides for chemical biological studies (the phenylalanine component of each is highlighted in green).
ability to analyse data from genomic and metagenomics sequencing efforts and pick out class I lyase-like enzymes with certain functions needs to be assessed. In this way it is envisioned that the natural biodiversity of this enzyme family could be accessed to obtain new templates for future biotechnologically-applicable enzymes.

Endeavours to use and engineer members of the class I lyase-like family as biocatalysts have primarily focussed on a subset of enzymes. As such, a further aim of this thesis will be to investigate different PALs and / or PAMs as template enzymes to be developed for use in biocatalytic processes. This aim will be supplemented by additional efforts to create intensified and / or scaled up processes involving these biocatalysts. It is hoped that these objectives will help to fill the gaps in current biocatalytic methods involving these classes of enzyme whilst keeping the research industrially-relevant and amenable to commercial use.
2. Computational Analyses of Class I Lyase-like Enzymes
2.1 Background

2.1.1 Electronic Prediction of Gene Function

The central dogma of molecular biology dictates that DNA is transcribed into various RNA types, mRNA is translated to protein and proteins fold into an active conformation to enable a biological function. This flow of information alludes to a long standing endeavour of biologists; to be able to predict the function of a protein from its sequence. A solution to this problem is required more and more each year, due to the exponential growth in genomic sequence data. Advances in sequencing technologies mean that genome and metagenome sequences are being elucidated far faster than it has ever been possible to characterise them. As such, crude electronic gene annotations, based on sequence similarities to infer evolutionary relationships, are employed in genomic databases.

The enormous potential of the wealth of sequence data in biological research depends on accurate annotation of putative gene sequences. Unfortunately, misannotations occur frequently due to lack of specialist knowledge of specific enzyme family evolution and poor curation of protein coding sequences. This is exemplified by the construction of RNRdb, a manually curated database of ribonucleotide reductase enzymes, in which 77% of the known sequences were found to be misannotated in some form in GenBank. With enzymes, the prediction of function is particularly difficult as even single amino acid changes in the active site can result in different catalytic activity.

2.1.2 Annotation of Class I Lyase-like Enzyme Sequences

In the class I lyase-like family misannotations are common. Even characterised enzymes are designated automatically with HAL or even just ‘ammonia lyase’ activity based on sequence similarity (examples from UniProt are displayed in Table 1). A representative case is the discovery of the cyanobacterial ammonia lyases AvPAL and NpPAL from Anabaena variabilis and Nostoc punctiforme respectively. These enzymes were found by searching the genome of the two organisms with the protein sequence of the bacterial ammonia lyase EncP from Streptomyces maritimus. The genes encoding these ammonia lyases are reported to be located within the genetic context of putative secondary metabolic pathways (non-ribosomal peptide synthesis for AvPAL and unusual fatty acid biosynthesis for NpPAL). These findings were suspect seeing as the enzyme sequences were annotated as HALs - enzymes central to bacterial primary metabolism of amino acids. This automatic designation of putative HAL activity, particularly in bacteria, is most likely due to the fact that, as primary
metabolic enzymes, a greater number of organisms should have genes coding for these enzymes than for the more specialist PAL, TAL, PAM and TAM enzymes. HALs, however, are less useful for biocatalytic purposes (unlike PALs / PAMs) and treatment of inborn errors of metabolism (e.g. PAL for phenylketonuria). This means that discovery of new enzymes in the family, to address pure and applied research problems, requires the construction of a more comprehensive annotation system. This has been done with other enzyme families using combinations of phylogenetics and knowledge of specific amino acid functions and their conservation within the family.\textsuperscript{85,87}

Table 1 || Examples of characterised class I lyase-like enzymes whose molecular functions have been misannotated in the Universal Protein Database (UniProt - entries retrieved May 2014).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Species</th>
<th>UniProt ID</th>
<th>Annotated Function</th>
<th>Actual Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvPAL</td>
<td>Anabaena variabilis</td>
<td>Q3M5Z3</td>
<td>HAL</td>
<td>PAL\textsuperscript{88}</td>
</tr>
<tr>
<td>TcPAM</td>
<td>Taxus canadensis</td>
<td>Q68G83</td>
<td>PAL</td>
<td>PAM\textsuperscript{55}</td>
</tr>
<tr>
<td>StlA</td>
<td>Photorhabdus luminescens</td>
<td>Q7N4T3</td>
<td>[ammonia-lyase]</td>
<td>PAL\textsuperscript{21}</td>
</tr>
<tr>
<td>RsTAL</td>
<td>Rhodobacter sphaeroides</td>
<td>B9KU73</td>
<td>HAL</td>
<td>TAL\textsuperscript{11}</td>
</tr>
<tr>
<td>RxPAL</td>
<td>Rubrobacter xylophilus</td>
<td>Q1AV79</td>
<td>HAL</td>
<td>PAL\textsuperscript{89}</td>
</tr>
<tr>
<td>AdmH</td>
<td>Pantoea agglomerans</td>
<td>Q84FL5</td>
<td>[ammonia-lyase]</td>
<td>PAM\textsuperscript{28}</td>
</tr>
<tr>
<td>EncP</td>
<td>Streptomyces maritimus</td>
<td>Q9KJH9</td>
<td>[ammonia-lyase]</td>
<td>PAL\textsuperscript{9}</td>
</tr>
<tr>
<td>MdpC4</td>
<td>Actinomadura madurae</td>
<td>B0BLP0</td>
<td>[ammonia-lyase]</td>
<td>TAM\textsuperscript{33}</td>
</tr>
</tbody>
</table>
2.2 Results and Discussion

2.2.1 Characterised Class I Lyase-like Enzymes

In order to be able to ascertain the function of a class I lyase-like enzyme from sequence information alone, previously characterised family members were studied and characteristics of their sequences rationalised in terms of catalytic activity. Phylogenetics was also used to map the likely evolutionary history of the enzyme family, based on sequence identity. Enzyme evolution has been hypothesised to follow a classical functionalisation pattern, whereby acquisition of new activity followed by diversification leads to clades of similar enzymes. These functional subgroups can often be used to distinguish function solely based on sequence similarity\(^8\) and it was predicted that this might be the case with the class I lyase-like family.

Knowledge of ‘function-discriminating residues’ for a particular enzyme family has also been shown to allow more accurate annotation of an enzyme’s catalytic activity when compared to phylogenetics alone, as sometimes a change in only a few key residues can be enough to modify the function of an enzyme dramatically.\(^8\) Examples of such residues have been reported in the class I lyase-like family, e.g. two aryl binding pocket residues (specificity or selectivity residues) which are almost always FL in PALs,\(^89\) CL in (\(R\))-selective PAMs,\(^90\) HL in bacterial TALs (and mutation to FL switches TAL to PAL activity),\(^91\) and SH in TAMs and HALs.\(^33\) HALs are also reported to have a conserved glutamate (E) in the active site which allows them to be distinguished from other members of the family, as these have glutamine (Q) at the homologous position.\(^3,33\) Structural studies of AdmH also show that the residue before this glutamine may help distinguish (\(S\))- from (\(R\))-selective PAM activity (F455 in AdmH vs. N458 in TcPAM).\(^13\) The active site loop regions are also thought to have functional significance, controlling substrate exit/entry and possibly mutase-lyase switching.\(^4\)

The sequences of 31 characterised class I lyase-like enzymes of known function were collected as available from the Universal Protein Resource (UniProt) and aligned for analysis using the ClustalW2 sequence alignment tool. By combining the alignments with structural data, phylogenetic inference and knowledge of function-discriminating residues, systems for the annotation of enzyme sequences were compared for this family. Key points of the analysis are shown in Figure 14.
**Figure 14** | A cladogram of 31 characterised class I lyase-like enzymes based on a full sequence alignment, of which characteristic sections (1, 2 and 3) are shown. Section 1 is the region surrounding the inner active site loop, with loop residues, as evident from various crystal structures, highlighted in yellow. Section 2 shows the aryl binding pocket specificity residues of each enzyme. Section 3 contains the homologous position proposed to allow differentiation between HAL and the other catalytic activities. The distinguishing glutamate is highlighted in blue for each of the 6 HAL enzymes in the alignment. The enzymes in the alignment are: MdpC4 from Actinomadura madurae, SgTAM from Streptomyces globisporus, CtM from Cupriavidus metallidurans, CmT from Chondromyces croatius, KedY from Streptomonas fulvus, MxTAM from Myxococcus sp., EaH from Enterobacter aerogenes, StHAL from Salmonella typhimurium, MmHAL from Mus musculus, RnHAL from Rattus norvegicus, SgHAL from Streptomyces griseus, RstHAL from Rhodobacter sphaeroideae, AmH from Pantoea agglomerans, StIA from Photorhabdus luminescens, PpHAL from Pseudomonas putida, EncP from Streptomyces maritimus, AvPAL from Anabaena variabilis, NpPAL from Nostoc punctiforme, TtPAM from Taxus canadensis, TwPAM from Taxus wallichiana var. chinensis, PpP from Petroselimum crispum, AtPAL1-4 from Arabidopsis thaliana, ZmPAL1 from Zea mays, RpP from Rhodopseudomonas toruloides, RstHAL from Rhodotorula graminis, BgA from Streptomyces sp. and Ssp8 from Saccharothrix espanaensis.

<table>
<thead>
<tr>
<th>Cladogram</th>
<th>Function</th>
<th>Sequence Section 1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>StIa</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>AvPAL</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MsPAM</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>TwPAM</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>TpPAM</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>PpPAM</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>ApPAL1</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>ApPAL2</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>ApPAL3</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>ApPAL4</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>ZmPAL1</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>RpPAM</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>RgPAM</td>
<td>PAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>BagA</td>
<td>TAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MdpC4</td>
<td>TAL</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>SgTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>KedY4</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>CtM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>CmT</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
<tr>
<td>MxTAM</td>
<td>TAM</td>
<td>LKKEWNLGEGEVTGYTQGPPGNNANVYVFFETIKKREHPV</td>
<td>FL</td>
<td>YGQ</td>
</tr>
</tbody>
</table>
2.2.2 Phylogenetics

The most striking feature of Figure 14 is the distribution of enzyme function across the cladogram. Although functional subgroups are present (e.g. all HALs cluster together), there are also examples of monophyletic groups within clades of otherwise functionally similar enzymes. Examples in the cladogram include the presence of CmTAL within the TAM clade and TcPAM / TwPAM surrounded by plant and bacterial PALs. These examples show that it is possible for mutases to evolve lyase activity (CmTAL) and vice versa as with the plant PAMs. It also demonstrates that the enantiocomplementary PAMs from plants (TwPAM/TcPAM) and bacteria (AdmH) do not cluster together and are only distantly related. TAL activity is probably the most extreme example, with TALs being present in all three of the subtrees in the cladogram which split off from the root. The lowest of these subtrees contains only RsTAL and Sam8. The middle subtree is made up mostly of TAMs and HALs but CmTAL is present as a monophyletic group. The top subtree is made up mostly of PALs but there are at least three instances of acquisition of TAL activity (one in BagA - a monofunctional TAL, one leading to RgPAL and RtPAL and one in ZmPAL - all bifunctional PAL/TALs).

These findings demonstrate that the class I lyase-like family has an interesting and complex evolutionary history, with some distantly related enzymes displaying similar activities and other more closely related enzymes being catalytically diverse. It can be inferred from this that many of the enzymes have exhibited convergent evolution to arrive at the same function from different starting points and via different mutations. It is also evident that, while phylogenetic relationships may have some use for annotating these enzymes, inference from sequence similarity alone will likely lead to further misannotations.

2.2.3 Active Site Loop Residues

Due to the presence of structurally characterised enzymes with different functions across the cladogram it is possible to analyse features not evident from the sequence alone, such as inner active site loop composition. As such, the eight structures relevant to the sequence selection were visualised and inspected to allow identification of which residues were present in the inner active site loop. Although loop composition varies, all contain a conserved YG flanked either side by a small amino acid (e.g. IYG, SYGV, AYG). The tyrosine is likely conserved because of its essentiality to the mechanism of the reaction and the glycine (as well as an additional fully conserved glycine 6 residues further down the loop).
probably allow two characteristic turns in the closed conformation of the loop (various structures)\textsuperscript{1,5,13,35,55} due to the high conformational flexibility of this amino acid.

One hypothesis would be that mutases have shorter or more rigid loops, as they require a slower turnover to keep the substrate in the active site longer and allow enough time for re-amination at the β-position. From Figure 14 it can be seen that loop length does not seem to be important for this, as evidenced by comparison of the mutase with the longest loop (SgTAM with 26 residues) and the two lyases with the shortest loops (RtPAL and AvPAL with 22 and 21 residues respectively). These PALs actually have loops which are a similar length to their much closer mutase relative TwPAM (22 residues). This implies that the rigidity or closed conformation interactions of the loops are likely to play a larger role in discriminating mutases and lyases. Until more is understood about this, it is probably not a useful descriptor for sequence annotation.

2.2.4 Aryl Binding Pocket Selectivity Residues

The selectivity residues (Figure 14, sequence section 2) show more promise for use in discriminating function. It can be seen, as reported, that all monofunctional PALs display the FL residues. The main exception is EncP, a physiological PAL that also displays aminomutase activity at low temperature. This indicates that PALs can be identified by these selectivity markers but also that enzymes similar to EncP may also display PAL activity. The PAMs from Taxus spp., although more closely related to plant PALs than bacterial ones, display CL rather than FL. This adds further evidence to the importance of the F residue in PALs, probably aiding the ammonia lyase activity with the substrate specificity being determined merely by the hydrophobicity of the aryl binding pocket. As previously described but on a smaller set of sequences,\textsuperscript{33} the TAMs and HALs all have SH specificity residues. CmTAL also has SH which, due to its relationship to the TAMs, is likely to help confer specificity for tyrosine, with ammonia lyase activity being due to other differences – possibly in the active site loop region. The other enzymes with TAL activity seem to be identifiable by the presence of an H residue at the first position of the specificity residues (RtPAL/RgPAL - HQ, ZmPAL/RsTAL/Sam8 - HL). This adds credence to experimental data on the effects of exchanging H for F in RsTAL.\textsuperscript{91} It does not explain, however, the acceptance of both phenylalanine and tyrosine by ZmPAL vs. the monofunctionality of RsTAL and Sam8, despite them sharing the same residues for substrate specificity (HL). This may be an area where additional use of sequence identity may be useful in distinguishing bifunctional PAL/TALs from other TALs. BagA does not have H at the first position of the specificity
residues despite being a TAL; it displays instead the residues YH. This reinforces the idea that BagA, which clusters with the PALs, evolved to TAL activity convergently with other, more distantly related enzymes via different amino acid changes at a different position. It also illustrates how the presence of a histidine residue at either selectivity position could be used to identify TAL activity in an enzyme sequence.

2.2.5 HAL Differentiation Residues

True to what was reported the HAL-discriminating glutamate can be seen to be present in all HALs in the alignment (Figure 14, section 3), with all other enzymes displaying Q at this position. It can also be seen that the two enzyme exhibiting (S)-selective PAM activity - EncP and AdmH - have an F before this Q, which in all other enzymes (except the two mammalian HALs) is an N. A bulky residue at the homologous position in AdmH (F455) is thought to orient the substrate in such a way as to allow access to both α- and β-positions on the same face and thus discriminate between (S)-PAMs and other enzymes in the family.13

These key residues allow section 3 to be used to identify different subgroups of enzymes which are closely related even if not all of them are functionally related. This four residue motif may be helpful in combination with other descriptors to distinguish possible enzyme functions. The motifs are: [YF]NQD for bacterial PALs, HNQD for plant enzymes and BagA, [AG]NQA for fungal enzymes, DNQD for a subclade of TAMs, [QS]NQD for the other subclade of TAMs, NQED for bacterial HALs, ATED for mammalian HALs, DFQD for (S)-PAMs and [WA]NQD for bacterial TALs (excluding the more distantly related BagA).

2.2.6 Searching Sequence Space

In the post-genomic age more and more genes are being sequenced in the genomes and metagenomes of all manner of organisms at a rate far faster than manual curation allows. This has implications for the study of all enzymes including those of the class I lyase-like family. As important primary and secondary metabolic enzymes in a variety of organisms, there are likely to be numerous uncharacterised class I lyase-like enzyme sequences waiting to be discovered. Using bioinformatic software such as the basic local alignment search tool (BLAST) it is possible to search for new enzymes using a template protein sequence. The majority of the hits in this case will be translated genomic DNA, as identified by a database’s open reading frame prediction, showing reasonable identity, similarity and coverage when compared to the query sequence.
These sequences can then be analysed visually using sequence alignments and the characteristic sequence sections identified in Section 2.2.1. In this way their function can be predicted, if the residues match, or designated as enzymes with potentially new or previously undiscovered functions if this is unclear from the sequence. tBLASTn searches (using a protein queries against an \textit{in silico} translated nucleotide database) were performed using 7 of the 31 characterised sequences from Section 2.2.1. The hits from these were used as a starting point for discovery of new HALs, PALs, PAMs, TALs and TAMs.

\textbf{2.2.6.1 AvPAL}

The cyanobacterial PALs, particularly AvPAL, have attracted much attention for their superior therapeutic properties such as increased stability and lower immunogenicity.\textsuperscript{48} NpPAL and AvPAL are also at an interesting point the evolutionary story of class I lyase-like enzymes, being of bacterial origin but more closely related to eukaryotic ammonia lyases and aminomutases than to other prokaryotic PALs.\textsuperscript{18} AvPAL was used as a query sequence to find closely related enzymes in a tBLASTn search. The results are summarised in Figure 15.
A section of the sequence alignment for AvPAL from Anabaena variabilis and its closest tBLASTn hits. The highly conserved loop region (as evident in the structure of AvPAL) of the first six sequences is enclosed within a dark blue box with amino acid differences highlighted in light blue. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). The hits include the characterised NpPAL from Nostoc punctiforme and the following putative proteins: GxTAL from Gloeocapsa sp., CtTAL from Chroococcidiopsis thermalis, ScTAL from Stanieria cyanosphaera, HaTAL from Herpetosiphon aurantiacus, RxTAL from Rivularia sp., OxPAL from Oscillatoria sp., LxTAL from Leptolyngbya sp., MxPAM from Methylobacterium sp. and PbPAL from Planctomyces brasiliensis. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Although many of the hits were found to be from other cyanobacteria, some of the more distantly related enzymes were from other bacterial species. This may well be due to horizontal gene transfer between different bacterial phyla, given the relatively high sequence identities exhibited. The selectivity residues of some of the more closely related hits displayed YH residues, the same as those shown by the bacterial TAL BagA. The presence of the histidine implies that these enzymes may accept tyrosine as substrates despite being up to 73% identical to the query PAL sequence (e.g. CtTAL). It is also interesting that the less similar OxPAL (70%) displays FL residues and so is likely to be a PAL. This is also the case for PbPAL but this sequence displays an even lower identity (50%). This enzyme is also interesting as it is from the genome of Planctomyces brasiliensis - a halotolerant organism first isolated from a hypersaline pond. This may point to the enzyme being more amenable to use in high salt concentrations; a valuable characteristic for biocatalytic PAL aminations, which require concentrated ammonium salt / pH-adjusted ammonia solutions. As well as PbPAL, other new phenylalanine-metabolising enzymes found in this search could...
be potential templates for the development of new therapeutics for the treatment of phenylketonuria and cancer, as they may have more appropriate starting characteristics with regards to stability, activity and immunogenicity. This is particularly likely as both AvPAL and NpPAL have been reported to have higher catalytic activity than other bacterial PALs. 

The findings here show that there may have been one or more substrate switching events in this area of sequence space and also reveal a new clade of potential TALs, evolutionarily distinct from all others. Another interesting sequence is LxTAL which displays FH residues; exactly halfway between the PAL-like FL and the TAL-conferring YH. As such residues have not been seen before, the function of LxTAL is not known but the presence of a histidine in the selectivity residues has been associated with TAL activity in other, characterised enzymes. This may be an intermediate sequence that has acquired an L=>H mutation from an ancestral PAL sequence and from which an F=>Y mutation would yield a BagA-like TAL motif. This process could also happen in reverse, i.e. from a former YH to a future FL motif.

Another sequence of interest is MxPAM which has selectivity residues GL. As there is no F at the first position this enzyme may not be a PAL, but the absence of H at either position makes it unlikely to be a TAL. In fact it is most similar to the TcPAM / TwPAM residue combination CL, lacking an aromatic F or Y at the first selectivity residue. This could make MxPAM a phylogenetically-distinct bacterial aminomutase, possibly with (R)-selectivity like that of the plant mutases.

The loop region of the top nine hits are very similar and in fact better conserved than the surrounding sequences. This is odd as a non-structure-forming feature on a protein, such as a mobile loop, might be expected to be more tolerant to variation. It is also true that species such as *Chroococcidiopsis thermalis* and members of the genera *Lyngbya* and *Oscillatoria* are known to thrive in hot springs. These findings may be linked in that the enzymes from high temperature dwelling microbes and their close relatives are all thermoactive and / or

**Figure 16** | A flowchart showing the possible mutations in the selectivity residues of class I lyase-like enzymes to allow conversion of PAL=>TAL activity and vice versa. The histidine residue responsible for tyrosine substrate preference in highlighted in green.
thermostable. If true this would go some way to explaining why AvPAL has superior stability compared with more distantly related PALs.\textsuperscript{94}

\subsection{2.2.6.2 StlA}

The StlA sequence is not closely related to any of the other characterised class I lyase-like enzymes as can be seen from the cladogram in Figure 14. It occupies an interesting position; branching off before the divergence of all other PALs from cyanobacteria, plants and fungi but just after BagA, a monophyletic TAL more closely related to the other PALs than any of the characterised tyrosine-specific enzymes. This hints at a large amount of sequence space between StlA and the other branches, with potentially very many uncharacterised bacterial PALs to be discovered. The highest scoring hits from the tBLASTn search for StlA are shown in Figure 17.

<table>
<thead>
<tr>
<th>Name</th>
<th>60 aa Sequence Surrounding Loop and Selectivity Residues</th>
<th>Seq ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>StlA</td>
<td>LEELNSGEMIYGNTGFGGNANLVVPPFEKIAEHQONLTPSAGTDIMSKPCIKASQF</td>
<td>109</td>
</tr>
<tr>
<td>PaPAL</td>
<td>LEERLNOEIMIGNTGFNNANLVVPPFEKISEHQQNLTPSAGTDIMSKCCTIRAQF</td>
<td>109</td>
</tr>
<tr>
<td>YePAL</td>
<td>LEKLDKNEEMIYTGFGGNNANVPPAILEAHQCNMLTPSAGTGLPIFPEEHRAQQF</td>
<td>109</td>
</tr>
<tr>
<td>DpPAL</td>
<td>LDANLKENEGIYGNTGFNNADLPIPPFESLHQSNLDPALSAGTGDCSSYKVRASQF</td>
<td>104</td>
</tr>
<tr>
<td>DdPAL</td>
<td>LENKLNENYAIYGNTGFNGNGLIPDFMKDLHQQNLDPLTCGGDFFDQYVRGIFQF</td>
<td>110</td>
</tr>
<tr>
<td>ScTAL</td>
<td>RDDLUATCRIYVGTITFSRQPVSTEGALTIQLGQNLGPTASESTVRATLM</td>
<td>77</td>
</tr>
<tr>
<td>MaTAL</td>
<td>IDKLSELQEGYVTITGDSCTYPPLDRNRLSHPLTVCBLGELFNAEITAPAILA</td>
<td>111</td>
</tr>
<tr>
<td>SnPAM</td>
<td>RDKLIAITGQYVGTITGDSWYMLSPQAAYLDNLVLVPLTVLTVGEEAFKSVAML</td>
<td>116</td>
</tr>
<tr>
<td>BiPAL</td>
<td>KHEINMQNNGYVGTTGDSVHQIGGEAAMDQSLNPLSCGVMVPDARVAMML</td>
<td>120</td>
</tr>
<tr>
<td>ToHAL</td>
<td>VENMIEREGYVVTGFKCNVVYRREDVKKQLINLISHSVVGVGDYFPEEVRAMML</td>
<td>99</td>
</tr>
</tbody>
</table>

Figure 17 || A section of the sequence alignment for StlA from Photorhabdus luminescens and its closest tBLASTn hits. The 4 aa loop lid motif, as inferred from other family members are enclosed within a dark blue box. The specificity residues are shown within the orange box and sequences displaying potential tyrosine specificity residues at these positions are highlighted (also in orange). The hits include the following putative proteins: PaPAL from Photorhabdus asymbiotica, YePAL from Yersinia enterocolitica, DpPAL from Dictyostelium purpureum, DdPAL from Dictyostelium discoideum, ScTAL from Streptomyces clavuligerus, MaTAL from Methylmicrobium album, SnPAM from Stackebrandtia nassauensis, BiPAL from Brevibacillus laterosporus and ToHAL from Thermosediminibacter oceani. The percentage sequence identity to the query sequence is shown in the last column for each hit.

Only eight non-HAL sequences were discovered in the search, which could imply that the species harbouring the sequence space around StlA have not been extensively sequenced. It also shows how evolutionarily diverse bacterial PALs are, with some such as StlA being
seemingly no more closely related to the others than to HALs. Again there was evidence of substrate switching events as in Section 2.2.6.1, with ScTAL displaying a BagA-like TAL motif, MaTAL with an intermediary FH and BlPAL, a more distant relation at just 34% identity, still displaying FL residues. The six PAL sequences shown here could be useful for biomedical or industrial applications, as mentioned in the introduction. This includes StlA, which, although a characterised enzyme, has never been reported to have been used in applied research. Another interesting sequence was SnPAM, displaying neither F nor H residues and instead LL. Although an uncharacterised combination, this motif does seem similar to CL from *Taxus* PAMs and GL from the potential aminomutase MxPAM and so may well be another bacterial mutase.

The more closely related sequences were all designated as PALs and, oddly, two of the sequences were from eukaryotes, despite the bacterial origin of StlA. These were DpPAL from *Dictyostelium purpureum* and DdPAL from *Dictyostelium discoideum*, both amoebozoan from the protist kingdom. At around 50% identity to StlA, these represent a fairly recent horizontal gene transfer event to eukaryotes that must be distinct from that to plants and fungi. It was shown in 2011 by Brock and co-workers that *Dictyostelium discoideum* exhibits bacterial ‘husbandry’ with its prey. Adding to previous knowledge that *Dictyostelium* spp. take up bacteria as nutrients via phagocytosis, it seems that *D. discoideum* is able to form an endosymbiotic relationship with ingested prokaryotes. In this way the prey can be kept alive within the amoeba when there is plenty of sustenance for digestion later when food is scarce. It is possible that prolonged periods of symbiosis between amoebozoan and bacteria have allowed horizontal gene transfer. This would explain why StlA-like sequences associated with enterobacteria can be found in two species of *Dictyostelium*, despite the organisms lying in completely separate domains of life. These would also be the first reports of PALs in the protist kingdom.

Further investigations of the two sequences from *Dictyostelium* were undertaken to ensure they did, in fact, constitute a second horizontal gene transfer event from eubacteria to eukaryotes. The fact that the sequences were from transcriptomic studies posed the possibility of cross-contamination of these data with the mRNA of any endosymbionts the subject amoebas may have harboured during the study. To address this, the DictyBase online database of *Dictyostelium* genomes was searched, using the nucleotide sequences for DpPAL and DdPAL (those originally translated by the tBLASTn search tool). Due to incomplete assembly of the genomic data for *Dictyostelium purpureum*, no such annotated gene was found matching DpPAL. One scaffold of the shotgun sequencing data was, however, shown to be 98% identical to the cDNA nucleotide data. The 2% discrepancy could be due to post-transcriptional differences between the genomic sequence and the mRNA
sequence it encodes (such as splicing of introns or RNA editing). In the *Dictyostelium discoideum* genome the sequence was found to be the same as the open reading frame DDB_G0273081; a sequence derived from gene prediction and supported by expressed sequence tags. It is present on the Watson strand of chromosome 2 and in *D. discoideum* strains AX3 and AX4 there are two copies of this sequence, probably as the result of a small-scale duplication. As inferred from homology, the DictyBase electronic annotation lists the gene as a putative histidine ammonia lyase, reiterating the lack of knowledge of specificity residues and phylogenetics used to annotate predicted class I lyase-like sequences. These further analyses of DpPAL and DdPAL sequences reveal that these putative PAL genes have indeed been transferred to the genome of *Dictyostelium* species from bacteria and are not false positive hits from bacterial remnants present in these organisms.

### 2.2.6.3 RtPAL

Despite statements in the literature about the presence of PALs in fungi very few of these enzymes have been characterised in any detail. It is known that RtPAL *Rhodosporidium toruloides* is bifunctional with detectable TAL activity but it is not known if this is the case for many other fungal species. As a diverse group of organisms the subkingdom Dikarya are likely to harbour many class I lyase-like enzymes. If bifunctional PAL-TAL has evolved within this sequence space, there may also be classical PALs, TALs or even aminomutases, as have evolved in plants and bacteria via separate functionalisation events. The tBLASTn hits for RtPAL numbered 157 from Dikarya. They displayed an unexpected diversity in the combination of selectivity residues compared to characterised enzymes or hits from other areas of the cladogram. As such, selected hits representative of this diversity were chosen for alignment and further analysis.
A section of the sequence alignment for RtPAL from Rhodosporidium toruloides and selected tBLASTn hits. The inner active site loop residues for RtPAL, as evident from the structure, are highlighted in yellow. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). The hits include the following putative proteins: PmPAL1 and PmPAL2 from Penicillium marneffei, UmPAL from Ustilago maydis, PgPAL from Puccinia graminis, CcPAL from Coprinopsis cinerea, EnPAL from Emericella nidulans, AoPAL1-3 from Aspergillus oryzae, PcPAL from Penicillium chrysogenum, PaPAL from Podospora aserina, PnPAL from Phaeosphaeria nodorum and AnPAL from Aspergillus niger. The percentage sequence identity to the query sequence is shown in the last column for each hit.

None of the hits for RtPAL were above 42% sequence identity and all were from relatively distantly related fungi, perhaps highlighting poor sequence coverage of *Rhodosporidium* and its closest relatives in genomic studies. By relating the structure of RtPAL to the alignment it seems that the enzymes have loops of different lengths which may have implications for variance in catalytic activity. The most striking feature of the set of hits was the diversity of the selectivity residues. Only very few recognisable combinations were present (HQ from PmPAL2 and HL from PgPAL and PaPAL) which could be labelled as potential bifunctional PAL-TALs. Some enzymes have partially familiar residues, such as FI in PcPAL (similar to FL in PALs) and HT in PmPAL1 (similar to HL / HQ in TALs or bifunctional PAL / TALs). Others are completely different with odd combinations of smaller residues, possibly hinting to PAL or maybe PAM activity. Three PALs, all from *Aspergillus oryzae*, are particularly noticeable as they possess a diversity of residues at these positions. AoPAL1 displays LL and so is possibly some kind of PAL or PAM, AoPAL2 has HH presenting the possibility of either histidine allowing accommodation of tyrosine in the active site and AoPAL has EL which is difficult to rationalise in terms of an amino acid substrate. It is possible that these...
residues are so unfamiliar because they do not accommodate traditional class I lyase-like substrates (like KedY4 in the TAM clade).

Figure 19 || Similarities between the fungus- and plant-associated cyclochlorotine and astin B cytotoxic agents isolated from Penicillium islandicum and Aster tartaricus respectively. In both structures the (R)-β-phenylalaninyl portion is shown in blue and differences in astin B with respect to cyclochlorotine are highlighted in red.

The possibility of acquisition of PAM activity among these organisms could be significant, due to the incidence of mutase products as components of fungal secondary metabolites. The hepatotoxic, chlorinated pentapeptide, cyclochlorotine, has been demonstrated to include a single monomer of (R)-β-phenylalanine. Despite its discovery in 1978, the biosynthetic pathway leading to this compound in the producer, Penicillium islandicum, has not yet been uncovered. Interestingly, the toxin bears a remarkable structural similarity to antitumour agents isolated from the flowering plant Aster tartaricus. Astin B, for example, is made up of the same amino acids as cyclochlorotine (including the (R)-β-phenylalanine, serine, α-aminobutyric acid and characteristic dichloroproline residues), but includes a threonine rather than a second serine. These findings may point to the fact that the astin compounds might be produced by a species of fungal plant-symbiont (endophyte) living within A. tartaricus plants. This seems more likely than the plant possessing a similar biosynthetic pathway to a distantly related fungus, especially in light of the fact that Penicillium and related genera are known to form endophytic relationships with land plants. Unfortunately the genome sequence of neither Penicillium islandicum nor Aster tartaricus was available to enable testing of this hypothesis, by searching for an open reading frame with the predicted sequence characteristics of an (R)-selective aminomutase. This evidence does however point to the fact that some of the enzymes discovered in this section may be (R)-PAMs and associated with the production of bioactive compounds. This would constitute
a separate, convergent acquisition of such activity, which at present has only been found in one genus of plant.

### 2.2.5.4 TcPAM

The occurrence of (R)-PAM activity seems to be limited to the *Taxus* genus of plants in characterised enzymes. The cladogram in Figure 14 shows that these mutases seem to have diverged from plant PALs after the horizontal gene transfer from bacteria. As modern day cyanobacterial PALs and plant PALs show FL selectivity residues and the monophyletic PAMs show CL, it is reasonable to suggest that the evolution from PAL to PAM required only a few sequence changes. The tBLASTn hits for TcPAM yielded a few PAMs all from *Taxus* spp. and many potential plant PALs. Only the top 7 PAL hits are shown in Figure 20.

All the *Taxus* enzymes, displaying the characteristic mutase CL residues, had been previously characterised, and, as such, no new aminomutases were uncovered in the search. Interestingly all of the hits with highest sequence identity are from divisions of non-flowering plants. This is significant as most studied plant ammonia lyases are from eudicots and monocots - orders of flowering plants (angiosperms).
Sections of the sequence alignment for TcPAM from Taxus canadensis and its closest tBLASTn hits. The inner active site loop residues of the hit, TwPAM from Taxus wallichiana var. chinensis, as evident from the structure, are highlighted in yellow. The specificity residues are shown within the orange box and sequences displaying characteristic plant mutase residues at these positions are highlighted (also in orange). Salt bridges predicted from a model of TwPAM are shown between an R and D residue (red boxes) and an R and E residue (blue box). The hits include two further characterised mutases: TmPAM from the hybrid yew tree Taxus x media and TbPAM from Taxus baccata. Also included in the hits are the following putative PALs: PpPAL from Physcomitrella patens, IIIPAL from Isoetes lacustris, DtPAL from Diphasiastrum tristachyum, PmPAL from Pinus massonia, GbPAL from Gingko biloba, PtPAL from Pinus taeda and LkPAL from Larix kaempferi. The percentage sequence identity to the query sequence is shown in the last column for each hit.

It has been proposed by those studying PAMs that inner active site loop dynamics play a large role in mutase-lyase switching. In fact molecular dynamics simulations carried out on the structure of TwPAM identified two salt bridges present in the closed conformation of the loop which were thought to aid mutase activity. One interaction, between two loop residues (red boxes in Figure 20), was found to decrease catalytic activity severely when mutated. The other, between a loop arginine and a glutamate in the core structure (blue boxes in Figure 20) was found to increase the ratio of lyase to mutase activity. This fits well with the sequence alignment of TcPAM’s closest relatives as the intra-loop residues are conserved in all hits. In fact these residues seem to be conserved across the characterised plant enzymes as can be seen from Figure 14, sequence section 1. This invariance in nature, combined with the mutagenesis data would imply that this first ionic interaction is essential to the catalytically-active closed conformation of the loop in plants. The arginine of the second interaction is also conserved across the TcPAM hits and other characterised plant enzymes,
but the glutamate in the core structure of PAMs is invariably an aspartate in all plant PAL sequences encountered in this Chapter. As both E and D have acidic side chains it is reasonable to assume that they are functionally equivalent - acting to form a clamp-like salt bridge to stabilise the closed conformation. It seems that an E at this position aids loop closure, allowing mutase activity via re-amination of the cinnamate intermediate. This may be due to the increased side chain length and greater number of rotamers of this amino acid, perhaps making the salt bridge more likely to form or stable for a longer period of time.

Further reported endeavours to alter loop dynamics in TwPAM made use of sequence comparisons with PcPAL1 from *Petroselinum crispum*. It was reasoned that, since the loops of these enzymes were so similar, exchanging all of the homologous loop amino acids in PAM to those in PAL should increase the lyase:mutase activity ratio of the former. Although seven differences exist in the loop regions of the two enzymes studied, only specific combinations of two or three of these substitutions gave the desired change in activity. From the double sequence alignment used in study it is not clear why this might be (Figure 21). Using the tBLASTn hit alignment in this section, however, reveals that many of the differences between TwPAM and PcPAL1 do not exist between the aminomutases and putative ammonia lyases found in the database. An example of this is I79 in TwPAM which, despite being S109 in PcPAL1 and a similar T in many of the plant PALs uncovered, is also I in the putative PpPAL (as well as in many of the other PALs studied in this Chapter from other areas of the class I lyase-like family tree). This can help explain why the I79S variant in TwPAM gave no improvement in lyase activity. Other examples include N95 and R96, the first of which is again different in PcPAL1 (K), but N is found at this position in many of the putative PAL hits. R96 is interesting as it is found to be different between TwPAM and all the uncharacterised PALs (as well as PcPAL1), but it is also not consistent with any of the other PAMs in the alignment. As all other PAMs and PALs have Q at this position, it is unlikely that this difference is of functional significance. In contrast TwPAM loop positions 77, 89, 91 and 97 were found to be conserved in all mutase sequences (A, C, S and L) and different to all predicted PAL sequences and PcPAL1 (T / S, T, H and G). It is a combinations of three of
these substitutions in TwPAM (A77T, C89T and L97G) that gave the desired increases in lyase activity. Interestingly mutagenesis to alter residue S91 is not reported, with a lack of difference in hydrophobicity of this residue and the homologous H121 being cited as he reason. It may be that a substitution here would have a similar effect to the others, given that the choice of the ineffective I76S was based on hydrophobicity analyses but not on phylogenetic / functional inference. These examples demonstrate that several of the amino acid variations between TwPAM and PcPAL1 can be found to be of phylogenetic origin and, as such, do not pertain to structural determinants of mutase / lyase activity. It also shows that such distinctions (between functional and phylogenetic differences) can be made, even within small sequence clades, using multiple alignments with effectively annotated putative genes and that these simple computational analyses can be used to guide, rationalise and reduce more complex and time-consuming laboratory work.

2.2.6.5 EncP

EncP and AdmH are closely related enzymes which display interesting thermal properties; acting as lyases at high temperature and mutases at low temperature. Despite this, AdmH is the only example of a physiological (S)-PAM in nature, with EncP acting as a lyase in the metabolism of its host organism. As bacteria are able to share genes via horizontal gene transfer more easily than other organisms and thus incorporate enzymes into different biosynthetic pathways, it is possible that other (S)-PAMs exist in the sequence space around these enzymes. As such, the sequence for EncP was used in a tBLASTn search and the top 10 hits are shown in Figure 22.
Figure 22 | Sections of the sequence alignment for EncP from Streptomyces maritimus and its closest tBLASTn hits. The inner active site loop residues of the top hit, AdmH from Pantoea agglomerans, as evident from the structure, are highlighted in yellow. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). A common DFQD motif for all the putative PAMs is enclosed within a dark blue box with the characteristic glutamate in the corresponding HAL sequences highlighted in light blue. The hits include the following putative genes: VbPAM from Vibrionales bacterium, BrPAM from Burkholderia rhizoxinica, PIPAM from Pseudomonas fluorescens, KpPAM from Klebsiella pneumoniae, BsPAM from Bacillus subtilis, DtPAM from Desulfovibacula toluolica, PmHAL from Paenibacillus mucilaginosus, BmHAL from Bacillus megatherium and CsHAL from Clostridium symbiosum. The percentage sequence identity to the query sequence is shown in the last column for each hit.

<table>
<thead>
<tr>
<th>Name</th>
<th>60 aa Sequence Surrounding Loop and Selectivity Residues</th>
<th>Seq ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>EncP</td>
<td>DERVIYGNTSMMGFDHLPVSQARQLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>453 100%</td>
</tr>
<tr>
<td>AdmH</td>
<td>DERNIVGYNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>477 63%</td>
</tr>
<tr>
<td>VbPAM</td>
<td>DERNIVGYNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>477 63%</td>
</tr>
<tr>
<td>BrPAM</td>
<td>QCQNYGNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>456 51%</td>
</tr>
<tr>
<td>PIPAM</td>
<td>QCQNYGNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>464 51%</td>
</tr>
<tr>
<td>KpPAM</td>
<td>EGSQYIVNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>465 50%</td>
</tr>
<tr>
<td>BmHAL</td>
<td>NGRQYIVNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>459 50%</td>
</tr>
<tr>
<td>PmHAL</td>
<td>ERSQYIVNTSSMGFDHLPVARISELQNLQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>460 45%</td>
</tr>
<tr>
<td>BmHAL</td>
<td>ERQVYSLTTCGKFSDTVIADGDLLEQNLNQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>440 39%</td>
</tr>
<tr>
<td>CspHAL</td>
<td>EKRVYSLTTCGKFSDTVIADGDLLEQNLNQENLINDATNVGAYLD...TSTDFQGTVSFVFAAR</td>
<td>432 39%</td>
</tr>
</tbody>
</table>

Among the hits were six potential (S)-PAMs, as evident from the presence of the DFQD motif. This contains the non-HAL glutamine and AdmH phenylalanine, proposed to be important for an orientation of the substrate which allows the MIO cofactor access to both α- and β-positions of the substrate on the same face ((S)-selectivity). These hits are likely to be (S)-PAMs under physiological conditions if they share the same temperature dependent switching of activity as EncP and AdmH. This is because they are all present in the genomes of mesophilic bacteria. The alignment section corresponding to the loop of AdmH is quite well conserved among the putative PAMs however it may be that subtle changes here have allowed low temperature lyase activity in some of the hits. A proline residue is conserved at the C-terminal end of the loop in all DFQD containing hits. As the most rotamERICALLY restricted amino acid it may be that this ‘hinge residue’ confers rigidity to the loop and is thus conserved for this function.

Of the six putative enzymes, three were from bacteria already known to produce natural products containing (S)-β-phenylalanine, which supports the prediction that the hits they harbour are (S)-PAMs, rather than HALs as annotated. Marine isolates of Pseudomonas...
fluorescens have been shown to produce moiramides (A & B) which are structurally similar to andrimid. The precursor (S)-β-phenylalanine is likely to be made by PfPAM, which may be present in a gene cluster similar to the Adm operon in Pantoea agglomerans. Vibrio bacterium, a sponge-associated microorganism, has also been shown to be responsible for the presence of andrimid isolated from the sponge tissue. This knowledge would make VbPAM the probable orthologue of AdmH in this organism. Due to the close sequence identity of VbPAM and AdmH it is likely that the entire andrimid gene cluster has been transferred between the organisms recently in evolutionary history.

In 2001 a narrow spectrum antibiotic with potent activity against Helicobacter pylori was isolated from Bacillus sp.. Several versions of these peptide natural products, dubbed pyloricidins, were characterised and all were found to contain a C-terminal (S)-β-phenylalanine, linked to an unusual 5-amino-2,3,4,6-tetrahydroxyhexanoic acid molecule, with some versions also containing one or more branched amino acids toward the N-terminus. It is possible that the β-amino acid used in the synthesis of this antibiotic is made by BsPAM, as it comes from the subtilis species of the genus Bacillus. No investigation seems to be made into how pyloricidins are biosynthesised. To test this theory of BsPAM's involvement in pyloricidin biosynthesis, a genomic context analysis of the gene was performed. The BsPAM-encoding open reading frame was found in the context of an operon-like cluster of protein coding sequences. From the electronic annotations of many of the surrounding genes, in combination with conserved domain searches for each, protein functions consistent with a putative biosynthetic pathway for pyloricidin could be constructed.
Retro-biosynthetic analysis of the unusual 5-amino-2,3,4,6-tetrahydroxyhexanoic acid precursor revealed a possible route, with correct stereochemistry, from the primary metabolite uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) as inferred from the gene annotations in the cluster. This would involve an N-acetylgalactosamine (GalNAc) yielding hexose epimerase combined with nucleotide sugar dehydrogenation, amino acid deacetylation and reduction of the ring-opened sugar aldehyde moiety in an undetermined order. Assuming the production of the (S)-β-phenylalanine via BsPAM from L-phenylalanine (available along with L-branch chained amino acids from primary metabolism), all amino acid precursors to various pyloricidins could potentially be present in *B. subtilis* courtesy of this gene cluster. Moreover, these putative enzymes were also found in the genetic context of non-ribosomal peptide peptidase and aryl-CoA ligase-like enzymes, possibly capable of assembling the amino acid monomers into the final antibiotic. As well as these, all three subunits required to produce ATP-binding cassette (ABC) transmembrane transporter...
complexes were also found in the middle of this cluster, hinting at the possibility of antibiotic-specific efflux mechanisms as a means of host resistance to the antimicrobial agent and delivery to extracellular locations. All of this points to a likely operon in *B. subtilis* responsible for the production and secretion of pyloricidin-like secondary metabolites.

![Diagram of proposed retrobiosynthesis](image)

**Figure 24** | One proposed retrobiosynthesis of the pyloricidin 5-amino-2,3,4,6-tetrahydroxyhexanoic acid precursor amino acid from the primary metabolite UDP-GlcNAc, as supported by genomic context analysis of the putative secondary metabolite gene cluster in Bacillus subtilis. The proposed steps are reduction (1), deacetylation (2), ring opening (3), dehydrogenation/oxidation (4) and epimerisation (5).

Researchers seeking enzymes capable of producing pyloricidin precursors, such as (S)-β-phenylalanine, could easily overlook BsPAM, due to its probable misannotation as an ammonia lyase with the submitted name of "histidine ammonia lyase". These findings reiterate how detrimental poorly-annotated enzyme families can be in biological research - in this case, studies of novel antibiotics. BsPAM is also an interesting example as it contains ASG at the positions homologous to the MIO-forming residues in AdmH. Although ASG is found in most characterised class I lyase-like enzymes, in AdmH, EncP and all DFQD containing hits, bar BsPAM, the residues are TSG.

Aside from the putative aminomutases, all subsequent hits were putative histidase enzymes, containing the bacterial HAL motif NQED. As these are the closest relatives of the (S)-PAMs, it is likely that their activity evolved directly from an ancestral HAL, with possible changes to the selectivity residues being SH (HALs) => SV (BsPAM, DtPAM) => AV / GV
(AdmH, EncP / KpPAM, PfPAM). Of the three histidine-specific enzymes shown in Figure 22, CsHAL from *Clostridium symbiosum* has unfamiliar SD selectivity residues, as opposed to the SH normally found. It is known from the active site characterisation of PpHAL that the S residue forms a hydrogen bond with its neighbouring H, allowing the ε-nitrogen of this residue to coordinate a divalent metal ion. This in turn may aid binding of the imidazole ring of the substrate histidine in the active site. It may well be that the D residue in CsHAL also allows coordination of a metal ion to aid binding or it may interact directly with the substrate. Either way, this sequence represents a potentially new type of HAL enzyme not previously reported and with a possibly different means of substrate binding. Further study of the sequence revealed that the MIO forming residues were not ASG or TSG, but SSG; a combination never before reported. These results seem to indicate that the MIO can be post-translationally formed as an active catalytic group with variation at the first position of the precursor amino acid triad.

### 2.2.6.6 BagA

The enzyme BagA occupies a unique position in the cladogram (Figure 14), being the only bacterial TAL in a large clade otherwise occupied by bacterial PALs and plant PALs / PAMs. As discussed before, BagA is likely to have evolved independently from the other bacterial TALs (RsTAL and Sam8), to which it is only distantly related, and converged on the same activity via a different mechanism (YH vs. HL selectivity residues). BagA was chosen as a good template sequence for enzyme discovery as it was hoped there would be a large amount of unexplored sequence space around this enzyme and a potentially interesting evolutionary history behind the acquisition of TAL activity.
Figure 25 || A section of the sequence alignment for BagA from Streptomyces sp. and its closest tBLASTn hits. The recognisable 4aa motif of the loop lid, as inferred from other family member sequences, is enclosed within a dark blue box. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). The hits include the following putative genes: AcTAL from Amycolatopsis decaplanina, RxPAL from Rubrobacter xylophilus, BiPAL from Brevibacillus laterosporus, ScTAL from Streptomyces clavuligerus, SeTAL2 from Saccharothrix espanaensis, SmTAL from Streptomyces mobaraensis, MpTAL from Microlunatus phosphovorus, SrPAL from Streptomyces rimosus, SnPAM from Stackebrandtia nassauensis and XaTAL from Xanthrobacter autotrophicus. The percentage sequence identity to the query sequence is shown in the last column for each hit.

The tBLASTn search using BagA gave around 500 hits before sequences of putative HAL activity were seen. Periodic samples of these first 500 sequences (every 50th hit collected and aligned) showed only YH and FH selectivity residues, implying the existence of many previously unexplored TALs or even bifunctional TAL-PALs from bacteria. Of the non-HAL hits, over 450 were from Proteobacteria so the large number of hits may be due to intense sequencing efforts in this phylum of eubacteria. Nonetheless it is a huge amount of sequence space information for enzyme discovery and study of natural enzyme evolution.

One interesting hit is a partial sequence from Natrinema versiforme – an extremely halophilic member of the Archaea domain. Unfortunately the loop and specificity residues are in the missing section but the ~50% identity of the remaining sequence to BagA and presence of a HNQD motif are the first evidence for the presence of non-HAL class I lyase-like enzymes in Archaea.

Among the top hits shown in Figure 25 there are examples of FL, FH and YH selectivity residues implying similar substrate switching events to those proposed in Sections 2.2.6.1 and 2.2.6.2. In fact one hit from the latter section, which was present in the hits for StlA, is also present in the top hits for BagA - SnPAM. The presence of the same sequence hit for
both queries implies that BagA and StlA are just two examples of enzymes from a large unexplored section of sequence space containing PALs, TALs and possible enzymes with both functions or mutase activity. One of the top hits, SmTAL, has the unusual combination of GH at the corresponding aryl binding pocket positions. The presence of an H residue suggests tyrosine substrate specificity but the absence of a large Y / F residue (as in BagA and other hits) may be for the same reason as the FL / CL difference in plant enzymes; a switch between mutase and lyase activity. In this case it would be the first example of an independently-evolved TAM, as all others have been shown to cluster together and have SH selectivity residues.

The presence of a potential TAL in the hits from *Saccharothrix espanaensis* is unusual as this species is already known to produce the bacterial TAL, Sam8. As SeTAL2 is a high scoring hit (56% identity) for BagA which is only distantly related to the other bacterial TALs, it is likely that *Saccharothrix espanaensis* acquired genes with the same activity via horizontal transfer of gene clusters for different natural products. This is one hypothesis as to why the bacterium has kept both genes, as their regulation would be tied to their specific biosynthetic pathways. As *Saccharothrix* and *Streptomyces* (which contains only BagA and no Sam8-like sequences) are from are closely related families of Actinobacteria, it is likely that SeTAL2 was originally present when Sam8 was acquired as a pseudoparalogue.

Another interesting hit is RxPAL, which is listed as a putative enzyme in Figure 25. This is because at the time of the tBLASTn search the enzyme's characterisation had not been published and so was given the assignment of a putative PAL, courtesy of its FL residues. The fact that this enzyme has now been characterised as a PAL supports the use of function-discriminating residue identification as a means of annotating putative class I lyase-like enzymes.

### 2.2.6.7 CmdF

CmdF is the only example of a truly (R)-selective TAM with all other known members producing the (S)-enantiomer of β-tyrosine. Two other enzymes from the TAM clade have also been shown to produce (R)-β-amino acids, KedY4 converts an unusual tyrosine derivative and CmTAL shows only minor TAM side activity. But there are examples of other natural products which contain (R)-β-tyrosine. One such example is jaspamide, a macrolide isolated from marine sponges with potent antifungal, insecticidal and anthelmintic activity. Jaspamide is actually remarkably similar to the chondramide end products of the biosynthetic pathway CmdF is involved in. This hints at the possibility that jaspamide could
be produced by a sponge-associated bacterium harbouring an (R)-TAM that has not yet been characterised (as predicted with VbPAM-mediated andrimid production in Section 2.2.6.5). Finding such an enzyme would help identify the microbe responsible and unravel the metabolic processes leading to jaspamide biosynthesis.

![Chemical structures](image)

**Figure 26** Chemical structures of chondramide C produced by Chondromyces crocatus and jaspamide H as isolated from the marine sponge Jaspis splendens. The structures both seem to be built from an (R)-β-tyrosine precursor (shown in blue) and are otherwise very similar (differences in jaspamide are highlighted in red).

Attempts to discover additional TAMs have already been undertaken using genomic database searches. However with the ever increasing numbers of sequenced genomes to be explored the availability of potential gene hits can change year by year. Studies such as these have also shown that the TAM clade of enzymes shows a large amount of sequence and evolutionary diversity given that most of them exhibit the same activity. The top tBLASTn hits for the CmdF protein sequence are shown in Figure 27.

Ten examples of potential uncharacterised TAMs were found in the hits for CmdF with the next closest relatives all being HALs after this. In a similar way to EncP in Section 2.2.6.5, this implies that this clade of enzymes evolved directly from an ancestral HAL, possibly via retention of the common SH selectivity residues and small changes to the active site (e.g. NQED=>DNQD). It was found that *Streptomyces ghanaensis* contained two distinct enzymes, both with 41-43% sequence identity to CmdF, DNQD HAL-discriminating motifs and SH selectivity residues. Due to their similar sequences these are likely to be paralogues rather than pseudoparalogues as with Sam8 and SeTAL2 (Section 2.2.6.6). Interestingly AdTAM is from the same organism as AcTAL from Section 2.2.6.6 (*Amycolatopsis decaplanina*).
Figure 27 | Sections of the sequence alignment for CmdF from Chondromyces crocatus and its closest tBLASTn hits. The inner active site loop residues of the hit, SgTAM from Streptomyces globisporus, as evident from the structure, are highlighted in yellow with likely loop residues from the other aligned sequences contained in the yellow box. The specificity residues are shown within the orange box and sequences displaying an as of yet uncharacterised combination of residues at these positions are highlighted (also in orange). A potential position influencing stereoselectivity in TAMs is shown within the purple box. A common [DQS]NQD motif to distinguish putative TAMs from HALs is enclosed within a dark blue box with the characteristic glutamate in the corresponding HAL sequences highlighted in light blue. The hits include other characterised enzymes: MiTAM from Myxococcus fulvus, MxTAM from Myxococcus sp., CmTAL from Cupriavidus metallidurans, RmTAL from Ralstonia metallidurans, MdpC4 from Actinomadura madurae and KedY4 from Streptoaelliteichus sp.. Also included in the hits are the following putative enzymes: KaTAM from Kutzneria albida, SxTAM and SxTAM2 from Streptomyces sp., ShTAM and ShTAM2 from Streptomycyes ghanaensis, MxxTAM from Microbispora sp., AdTAM from Amycolatopsis decaplanina, SfTAM from Streptomyces albus, SaTAM from Salinispora arenicola, ACHAL from Aminobacterium colombiense, TaHAL from Thermaerovibrio acidaminovorans, TIHAL from Thermovibrio lienii, TIHAL from Thermosiphon africanus, APhAL from Aminomonas paucivorans, and GxTAM from Geobacter sp.. The percentage sequence identity to the query sequence is shown in the last column for each hit.
As a way of predicting enzyme stereochemistry, a position in the alignment corresponding to residues 399 of CmdF was compared across all hits. This residue is shown to alter the stereoselectivity of CmdF in either direction when mutated from the native E and so has stereochemical and possible substrate-directing functions. The only hits found with an E at this position were MxTAM and MfTAM - previously characterised enzymes already known to be (S)-selective. However many hits with an SgTAM / MdpC4-like A at this position were found - also enzymes with demonstrated (S)-TAM activity.

An interesting variation at this position (T) was found in KedY4, which is known to accept an unnatural amino acid (2-aza-tyrosine) as its wild-type substrate and is also (R)-selective like CmdF. It may be that this position is key to allowing unusual substrates to be accepted by TAMs as well as its proposed role in product stereochemistry. An S is found at this position in the putative KaTAM which also has the unfamiliar NH (vs. the traditional TAM SH) selectivity residues, so this enzyme may also have a tyrosine derivative as its substrate and / or may be (R)-selective. The other variation at this position is in GxTAM; a gene with lower sequence identity to CmdF than some of the HALs found in the BLAST search. It has an M at this position which may make it (R)-selective, but this residue is stated in the literature (and backed up by the alignments in this Chapter) to be present here in HAL enzymes. GxTAM does not, however, display the characteristic glutamate of a HAL (it has SNQD) and so may be a surviving evolutionary intermediate between HALs and TAMs. An ancestral GxTAM-like mutase may have diverged giving the variation seen at this position today. It could also be that GxTAM has an unnatural substrate like KedY4. The possible involvement of the potentially (R)-selective TAMs (KaTAM and GxTAM) in the biosynthesis of sponge-associated metabolites was investigated. While no direct evidence of Kutzneria or Geobacter species interactions with sponge hosts are reported, an uncultured bacterium closely related to Geobacter has been observed in tissue samples from the marine sponge Rhopaloëides odorabile.

The occurrence of a mutate-lyase switching event in the TAM subgroup has not been investigated in the literature. Using the sequence alignments between different members of the clade and the structure of SgTAM it is possible to try and rationalise the structural determinants of differences in function between TAM / TAL enzymes. Visualisation of the PDB file 2RJR in PyMOL revealed a possible salt clamp between E57 and K82 at opposite ends of the loop. Other possible interactions include those with part of a neighbouring monomer of the enzyme. It is clear in the structure that hydrogen bonding occurs between the main chain atoms, forming a stable antiparallel beta sheet section between the inner and outer active site loops. The proximity of K84 from the end of the inner loop and D296 from the middle of the outer loop presents the possibility of an additional salt clamp. Although the
alignments do not show conservation of these charged residues, all do show the presence of oppositely charged residues at either end of the loop region. As such, it is possible that replacements at interacting positions in the loops over evolutionary time have served to maintain stabilisation of the closed conformation of the enzyme and thus preserve mutase, as opposed to lyase, activity. CmTAL however does not have any acidic or basic amino acids at the N-terminal end of the loop which may allow greater loop flexibility or be an indication of a lack of such stabilising interactions for the closed conformation of this enzyme.

Another feature of the CmTAL loop is the presence of a W, not found in any of the other TAM loops. As the bulkiest and rarest amino acid with the largest hydrophobic surface, the presence of tryptophan in a dynamic, often solvent exposed part of the enzyme was investigated further. This was achieved via the mutagenesis function in PyMOL, using the closed-loop structure of SgTAM in the absence of a solved structure for CmTAL. As the enzymes display 59% sequence identity it is likely that, excluding the loop, the enzymes have almost identical structures. The homologous residue in SgTAM (Q76) was exchanged with W and the resulting discrepancies visualised in all 6 backbone-dependent rotamers for tryptophan at this position. It can be seen from Figure 28 that the Q76W in silico variant displays extensive clashing between the inner active site loop and remainder of the enzyme in the closed conformation, regardless of which rotamer the residue adopts. This is likely to make the closed conformation far less stable in CmTAL than in SgTAM (assuming structurally similar closed conformations), such that the equilibrium of loop dynamics may be shifted further towards the open conformation. This provides a further possible explanation as to why CmTAL displays enhanced lyase activity and diminished mutase activity, despite its TAM-like protein sequence.
2.3 Conclusions and Further Work

An improved method of annotating class I lyase-like enzymes was designed from sequence analysis of 31 characterised enzymes. This allowed the discovery of many potential PALs, TALs, PAMs and HALs from different organisms. Many of these may be potentially useful for the various applications of class I lyase-like enzymes as discussed in Chapter 1.

By studying the distribution of enzyme function across the enzyme family and looking at sequences of closely related enzymes, patterns in the switching of enzyme function were observed and some rationalised theoretically. This analysis of sequence space has also provided possible insights into the biosynthesis of different natural products. This work could also provide the basis for a number of enzyme characterisation experiments to ascertain the validity of using function-discriminating residues and phylogenetic analyses to assign enzyme function computationally.
3. Development of EncP as a Novel Biocatalyst
3.1 Background

3.1.1 The PAL / PAM Toolbox of Biocatalysts

The direct amination of arylacrylic acids - using free ammonia to yield phenylalanine derivatives - is an attractive route to valuable chiral chemicals, if it can be achieved with high regio- and enantioselectivity. This is because, for this reaction, there exist four individual stereoisomers that could be produced via ammonia addition across the double bond. These are the (S)- and (R)-β-aromatic amino acids, via addition on either face of the carbon adjacent to the aromatic ring, and the (S)- and (R)-α-amino acids, via stereoselective addition to the next carbon along, nearest the carboxyl group. As this is a reaction which cannot be achieved through simple chemical methods, there has been much interest in the direct amination route, catalysed by class I lyase-like family members.

A lot of work has been undertaken using PAL enzymes from various organisms, as a synthetic strategy for (S)-α-phenylalanine and its derivatives (Figure 29A). The first example of this was the use of *Rhodotorula glutinis* yeast cells to perform biotransformations on trans-cinnamate, yielding L-phenylalanine, courtesy of an endogenous ammonia lyase enzyme. Later this approach was used to synthesise many ring-substituted phenylalanine analogues, starting with the corresponding cinnamate derivatives, with varying levels of conversion.

More recently, use of recombinant plant and bacterial PALs has allowed for more optimised syntheses of specific unnatural amino acids, namely using halocinnamates as starting materials.

With the success of PAL-mediated amino acid production over the past few decades, efforts to develop complementary routes to the other isomers of these compounds have been sought in recent years. The first report of β-amino acid production by a class I lyase-like enzyme from similar starting materials (Figure 29B) was in 2009, with the use of an (R)-selective PAM from *Taxus wallichiana var. chinensis*. Various rounds of engineering on this enzyme have enabled its use in the synthesis of (R)-β-phenylalanine derivatives with high optical purity, little α-amino acid side product and improved turnover rates. Lack of a reported wild-type family member with considerable (R)-α-selectivity shifted the focus to other approaches to the remaining α-optical isomer. After unsuccessful attempts to augment the (R)-α-forming side activity of a bacterial ammonia lyase AvPAL, a multienzymatic cascade was developed (Figure 29C) to allow production of the opposite enantiomer of the PAL-catalysed process. This involved initial PAL amination of various arylacrylic acids with subsequent enantiospecific oxidation - with an L-amino acid deaminase (LAAD from *Proteus mirabilis*) - followed by a non-selective chemical reduction. By coupling the initial amino acid
formation to a ‘deracemisation’ cycle, \((R)\)-\(\alpha\)-phenylalanine derivatives could be obtained with good yield and excellent enantiopurity.\(^{107}\)

![Diagram Figure 29](image)

**Figure 29** The possible biocatalytic routes from ammonia and arylacrylates to all four stereoisomers of the corresponding phenylalanine derivatives. Methods yielding \((S)\)-\(\alpha\)- (A: PAL alone), \((R)\)-\(\beta\)- (B: engineered PAM) and \((R)\)-\(\alpha\)- (C: PAL + deracemisation cycle) isomers have been previously reported, leaving only the missing \((S)\)-\(\beta\)-form which cannot be synthesised in this way (D).

Despite the recent advances in this area of research, there still exists no reported method allowing \((S)\)-\(\beta\)-selective amination of acrylic acids, in the same way as has been demonstrated for the other three forms. With a view to completing the PAL / PAM biocatalytic toolbox, use of wild-type or engineered enzymes, possibly in tandem with other bio- and / or chemocatalysts, must be considered. In this way, all four isomers of various aromatic amino acids could be produced from the same starting material, merely by varying the biocatalytic method employed.

### 3.1.2 EncP from *Streptomyces maritimus*

The bacterial ammonia lyase, EncP, was discovered in connection with enterocin and wailupemycin antibiotic metabolism in the thermotolerant bacterium *Streptomyces maritimus*.\(^9\) The oddly low turnover rate for this enzyme under experimental conditions\(^{23}\) was later rationalised by the discovery that the enzyme was in fact thermobifunctional. Studies with a closely related enzyme, AdmH - an \((S)\)-selective PAM from *Pantoea agglomerans*,\(^{29}\) revealed that both enzymes act primarily as aminomutases at ambient temperature, with a slight ‘leaking’ of the cinnamate intermediate. At higher temperatures, however, cinnamate production becomes more and more prevalent, resulting in predominant ammonia lyase activity at temperatures above \(70^\circ\mathrm{C}\).\(^4\) As AdmH is known to allow the desired \((S)\)-\(\beta\)-amination as the second step in its aminomutase catalytic cycle\(^{29}\) and the enzymes possess
high sequence identity (63%) and similar enzymatic properties,\textsuperscript{4} it was hypothesised that either would be a good starting point for the development of the missing biocatalyst in the PAL / PAM toolbox. The thermal properties of these enzymes were also deemed to be of interest in this regard. Whilst PAL enzymes typically show high catalytic constants compared to PAMs,\textsuperscript{5} they are not as of yet reported to allow formation of β-amino acids. Previous attempts at PAM-mediated syntheses have in effect been hindered by the low turnover rates of these enzymes, with engineering efforts specifically undertaken to improve their efficiency in this respect.\textsuperscript{5,9,9} As such, it may be that a thermobifunctional PAL / PAM may allow the best of both worlds; the ability to form β-regioisomers of amino acids with high temperature being used to overcome low turnover rates.

\subsection*{3.2 Results and Discussion}

\subsubsection*{3.2.1 Development of a Biocatalytic Method with EncP}

A pET-28a plasmid vector containing the \textit{encP} gene\textsuperscript{9} codon-optimised for laboratory strain \textit{E. coli} was obtained and used as previously reported\textsuperscript{4} without modification. This was used to transform an expression strain to prepare a potential whole cell biocatalyst using a method shown to work with another bacterial ammonia lyase, AvPAL.\textsuperscript{107} As this method was not the same as others reported for production of active EncP enzyme,\textsuperscript{4,9} a protein purification was attempted to assay for soluble EncP. This was achieved through nickel affinity chromatography courtesy of the hexa-histidine tag contained in the EncP-encoding construct. The bound protein was first washed with a \textasciitilde{}50 mM imidazole solution followed by elution with a higher concentration (1 M). Representative fractions of the purification procedure, along with column flow-through and a protein ladder, were run on SDS-PAGE and the diagnostic band at \textasciitilde{}60 kDa observed.
Figure 30 || SDS-PAGE analysis of selected fractions from the purification of EncP produced with autoinduction media (right - whole flask cell pellet, left - 1/3 flask cell pellet). The labels over the lanes indicate the marker (M), and fractions from: the column flow-through (F), the wash with buffer A (A), the wash with an AB mixture (A:B) and the elution with buffer B (B).

From the protein stain pattern of the SDS-PAGE it was evident that a polypeptide of correct size and with high affinity for the nickel stationary phase was being overproduced by the E. coli cells. An initial preparation from the total protein content of one 800 mL production culture yielded such a large quantity of protein that overproduced bands were apparent in the wash fractions of the purification. A repetition of this procedure with one third of such a culture gave a more classic purification profile.

Upon confirmation of protein production, the activity of EncP in the synthetically-relevant amination direction was tested on a substrate known to be bound by the enzyme (cinnamate 1a - an intermediate in the mutase and product in the lyase reaction). Reports from the literature with a related, enantiocomplementary enzyme with PAM activity suggested that formation of both β- and α- phenylalanine (2a and 3a) would be detectable, if the reaction were to work (Figure 31). Initially the whole cell biocatalyst was tested with a wet weight of 40 mg mL\(^{-1}\) in a 750 μL reaction left overnight (22 hours). Two potential buffers were used as reported for past amination experiments within the family of enzymes (13\% ammonia solution, pH 11\(^{52}\) and 5 M ammonium hydroxide, pH 9.6\(^{107}\). A final substrate concentration of 1 mM cinnamate was used in both cases. A temperature of 70°C was chosen for initial studies, as the enzyme was reported to be active and predominantly an ammonia lyase at this temperature.\(^4\) This high temperature incubation required use of an Eppendorf thermomixer comfort which, due to the small gyration radius, required an agitation of 400 rpm to keep the cells suspended in the reaction mixture. Unfortunately, upon reverse-phase HPLC analysis of the appropriately prepared biotransformation, no traces of β- or α-
phenylalanine were detected - only a single peak corresponding to trans-cinnamate (all in comparison to retention times of authentic standards).

Figure 31 || The target biotransformation with EncP involving the previously reported substrate ((S)-α-phenylalanine 3a) and products ((S)-β-phenylalanine 2a and cinnamate 1a) of the aminomutase / ammonia lyase reactions, but in the amination direction.

In order to find conditions suitable for detection of cinnamate amination activity, a selection of ammonium salts readily available in the lab was tested for suitability as ammonia donors for the reaction. These were ammonium acetate, ammonium chloride, ammonium nitrate and ammonium sulphate. These were tested at 1 M concentration in a manner similar to the initial biotransformations with the EncP-containing whole cells. As each salt produced an acidic solution, pH adjustment was performed with KOH to allow reaction buffers to be in a similar range and all above pH 8 (the value used from lyase / mutase studies of EncP). This gave solutions of pH 8.6 for the acetate salt, 8.5 for the chloride, 8.4 for the nitrate and 8.6 for the sulphate.

Table 2 || The effect of ammonium donor salt on the conversion of cinnamate and β:α ratio of products in EncP-catalysed biotransformations.

<table>
<thead>
<tr>
<th>Salt</th>
<th>Conv.</th>
<th>β- : α- ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ammonium Acetate, (1M, pH 8.6)</td>
<td>23%</td>
<td>62 : 38</td>
</tr>
<tr>
<td>Ammonium Chloride, (1M, pH 8.5)</td>
<td>25%</td>
<td>53 : 47</td>
</tr>
<tr>
<td>Ammonium Nitrate, (1M, pH 8.4)</td>
<td>23%</td>
<td>49 : 51</td>
</tr>
<tr>
<td>Ammonium Sulphate, (1M, pH 8.6)</td>
<td>34%</td>
<td>55 : 45</td>
</tr>
</tbody>
</table>

Reaction conditions: 40 mg mL⁻¹ wet cells, 0.8 mL, 70°C, 22h
After the 22 hour reaction time, HPLC analysis showed production of both regioisomers of phenylalanine from the starting material. Conversions were found to be quite poor but above 20% in all cases with ammonium sulphate giving by far the best value of 34% consumption of cinnamate. Interestingly, although most reactions were found to give relatively equal mixtures of the two phenylalanine isomers, reactions left to proceed in ammonium acetate gave a β:α ratio of 62:38; a definite favouring of the β-product. It is not clear why the anion of the ammonium salt should have such an effect on the regioselectivity of the EncP-catalysed reaction. It may be that altering these changes the specific interactions of the substrate and / or products in solution, possibly via a stabilisation of one product relative to another. It is also odd that a range of ammonium salts can be used as reaction buffer for this enzyme system, whereas the traditional pH adjusted ammonia solutions (used with other family members) yield no discernible conversion. This may be due to the amount of anion present in case of the dissolved salts versus acidified ammonia i.e. a solution of ammonium sulphate basified to higher pH will have a higher relative concentration of sulphate ions than an ammonia solution pH adjusted with sulphuric acid. It could also be that the enzyme was inhibited by the higher concentration and pH of the buffers that work well with PAL enzymes.

As ammonium sulphate had been demonstrated to give superior results at similar concentration and pH to the other candidate buffers, this salt was used in all subsequent optimisation efforts. To test the effect of buffer concentration on conversion, a range of molarities of ammonium sulphate were prepared and used in whole cell biotransformations of cinnamate. Due to the solubility limit of ammonium sulphate in the buffer used, a maximum concentration of 4 M was employed. In this case reactions were also performed in triplicate to test the replicability of such experiments at this scale. The results (summarised in Table 3) showed that, with higher ammonium sulphate concentration relative to that of the substrate, conversion increased - with 4 M reaction buffer giving a reasonable 74% mean conversion from all 22 hour time samples. This is in line with reports of the class I lyase-like enzyme amination reactions being equilibrium driven and thus requiring large excesses of ammonia to reach reasonable conversion. It also shows that the higher concentration of reaction buffer is not detrimental to catalytic activity as previously speculated. Variance between the triplicated reactions was found to be minimal with all giving similar conversion values.
Table 3 || The effect of ammonium sulphate concentration on the conversion of cinnamate in EncP biotransformations.

<table>
<thead>
<tr>
<th>Conc.</th>
<th>0.5 M</th>
<th>1.0 M</th>
<th>1.5 M</th>
<th>2.0 M</th>
<th>3.0 M</th>
<th>4.0 M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv.</td>
<td>rep. 1</td>
<td>15%</td>
<td>23%</td>
<td>35%</td>
<td>47%</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td>rep. 2</td>
<td>15%</td>
<td>23%</td>
<td>38%</td>
<td>46%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>rep. 3</td>
<td>15%</td>
<td>25%</td>
<td>34%</td>
<td>44%</td>
<td>65%</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>15%</td>
<td>24%</td>
<td>35%</td>
<td>46%</td>
<td>64%</td>
</tr>
</tbody>
</table>

Reaction conditions: 40 mg mL⁻¹ wet cells, 0.8 mL (NH₄)₂SO₄ (pH 8.3) 70°C, 22h

Next the effect of pH on the reaction was tested via pH adjustment of 3 M ammonium sulphate to 7.0, 8.0, 9.0 and 10.0 using an appropriate base (KOH). These reactions were performed as before and again in triplicate. The results (Table 4) showed that pH 7 gave a moderate conversion of 45-50% in all cases with all other reactions displaying 58-63% (with, again, very little variation between identical reactions). As such, no distinct improvement or impairment was observed upon changing the pH between 8 and 10. This demonstrates that high pH is not fully detrimental to the activity of the enzyme, possibly due to the inherent nature of the enzyme or to a balancing effect of increased reaction rate and reduced biocatalyst stability with increasing pH. As no advantage is evident in further pH adjustment from 8 to 10, and the initial pH of the dissolved ammonium salt in around 5.6, a pH of around 8 was deemed adequate for all further investigations. To get a final ammonium sulphate concentration of 4 M, it was found that dissolving the corresponding weight of the salt in pH 12 borate buffer consistently gave a buffer of pH 8.3. Due to the ease of creating this solution it was chosen for use in all successive EncP reactions.

Table 4 || The effect of reaction buffer pH on the conversion of EncP biotransformations.

<table>
<thead>
<tr>
<th>pH</th>
<th>7.0</th>
<th>8.0</th>
<th>9.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rep. 1</td>
<td>50%</td>
<td>60%</td>
<td>60%</td>
<td>62%</td>
</tr>
<tr>
<td>rep. 2</td>
<td>45%</td>
<td>61%</td>
<td>58%</td>
<td>60%</td>
</tr>
<tr>
<td>rep. 3</td>
<td>45%</td>
<td>62%</td>
<td>62%</td>
<td>63%</td>
</tr>
<tr>
<td>mean</td>
<td>46%</td>
<td>61%</td>
<td>60%</td>
<td>62%</td>
</tr>
</tbody>
</table>

Reaction conditions: 40 mg mL⁻¹ wet cells, 0.8 mL 4M (NH₄)₂SO₄, 70°C, 22h

Next the temperature profile of the new biocatalyst was probed using the new reaction buffer with optimised pH and ammonium salt concentration. From a previous report it was evident that EncP was active across a range of temperatures but it was unclear how these might affect the newly demonstrated amination reaction. Identical biotransformations were set up and each was incubated at a different sample temperature (25-85°C in increments of 15°C).
From these results, shown in Table 5, it was evident that at low temperature the mutase-like turnover rate of EncP was predominant, showing a low conversion of just 22%. This was found to increase at 40 and 55°C to 52 and 83%, presumably owing to the greater lyase-like dynamics of the enzyme with increased thermal energy. At temperatures of 70°C and above, however, the consumption of the substrate was found to fall. The mild decrease in conversion from 55 to 70°C, along with the sharp drop from 70 to 85°C, could be indicative of loop dynamics increasing to such an extent that correct / prolonged active site closure becomes less and less likely, resulting in fewer productive binding events between the substrate and the enzyme. Alternatively it could be due to decreased stability of the enzyme at higher temperatures causing the biocatalyst to become inactive more quickly, thus providing less time in which the products can be synthesised.

Table 5 || The effect of incubation temperature on the conversion of cinnamate and β:α ratio of products in EncP-catalysed biotransformations.

<table>
<thead>
<tr>
<th>Temp.</th>
<th>Conv.</th>
<th>β : α ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>25°C</td>
<td>22%</td>
<td>52 : 48</td>
</tr>
<tr>
<td>40°C</td>
<td>52%</td>
<td>57 : 43</td>
</tr>
<tr>
<td>55°C</td>
<td>83%</td>
<td>58 : 42</td>
</tr>
<tr>
<td>70°C</td>
<td>74%</td>
<td>58 : 42</td>
</tr>
<tr>
<td>85°C</td>
<td>24%</td>
<td>75 : 25</td>
</tr>
</tbody>
</table>

Reaction conditions: 40 mg mL⁻¹ wet cells, 0.8 mL 4M (NH₄)₂SO₄ (pH 8.3), 22h

In light of the data, the incubation temperature for biotransformations with EncP was changed from 70 to 55°C. This gave an optimised biocatalytic method of 4 M ammonium sulphate, pH 8.3, 1 mM substrate concentration, 40 mg mL⁻¹ wet whole cell biocatalyst, 750 μL reaction volume, 70°C incubation, 400 rpm agitation, 22 hour reaction time.

3.2.2 Substrate Scope of EncP

Using the whole cell biotransformation procedure developed in the preceding section, the substrate scope of the EncP was tested. Previous tests with PALs and TwPAM revealed broad substrate scopes, specifically with regards to arene substitutions on the ring of the cinnamate scaffold.14,103–105 As such, a range of 23 commercially available cinnamate derivatives (1b-x) was used and each incubated at 1 mM substrate concentration using the reaction conditions mentioned above. The enzyme was found to convert 21 of the compounds along with the unsubstituted 1a, with all the accepted substrates containing one
or more ring substituents at all positions. Interestingly the conversion and β:α-amino acid product ratio of the biotransformations were found to vary greatly with the different substrates tested. As predicted, the enzyme was found to be (S)-selective for the amination reaction of almost every single compound tested, regardless of the regioselectivity of the reaction.

Table 6 || EncP-catalysed amination of cinnamate and various ring-substituted fluorocinnamates

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>βee</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>81%</td>
<td>56 : 44</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1b</td>
<td>2-F</td>
<td>92%</td>
<td>25 : 75</td>
<td>75% (S)</td>
<td>83% (S)</td>
</tr>
<tr>
<td>1c</td>
<td>3-F</td>
<td>83%</td>
<td>45 : 55</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>78%</td>
<td>55 : 45</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1e</td>
<td>3,5-F₂</td>
<td>75%</td>
<td>33 : 67</td>
<td>n.d.</td>
<td>72% (S)</td>
</tr>
<tr>
<td>1f</td>
<td>2,3,4,5,6-F₅</td>
<td>95%</td>
<td>2 : 98</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

Comparison of the biotransformation of the unsubstituted acrylic acid 1a with several ring-fluorinated derivatives 1b-f (Table 6) revealed the extent to which the number and position of substituents seemed to influence the reaction. The presence of a single fluorine at the ortho-position (1b) or on all five ring carbons (1f) was found to give higher conversion than for the unsubstituted substrate. Otherwise the conversion of monofluorinated cinnamates followed the pattern ortho- > meta- > para-. With regards to regioselectivity, the presence of any number of electron-withdrawing fluorines on the ring was observed to give higher levels of α-product. The number of fluorines (one for 1c, two for 1e and five for 1f) was also found to correlate with increased α-selectivity, with more fluorines giving less β-product in each case (β:α ratios of 45:55, 33:67 and 2:98 respectively). This is consistent with previous theories that electronic effects of ring substituents of a substrate influences the addition preference of an enzyme with aminomutase activity. Oddly the 4-fluorocinnamate 1d does not demonstrate the same switch in regioselectivity as the other substituted compounds, with α-addition seeming less prevalent from ortho- to meta- to para-compounds. The enzyme was found to exhibit complete enantioselectivity (within the limit of HPLC detection) with three
compounds (1a, c and d) at both β- and α-positions. The lowered αee for difluorocinnamate 1e is possibly cognate to that found with this compound in reactions with another bacterial ammonia lyase, AvPAL from *Anabaena variabilis*. EncP also displayed imperfect enantioselectivity with the 2-fluoro-compound (1b), giving detectable traces of the (R)-form of both β- and α-products. Unfortunately adequate separation of enantiomers of 2e-f and 3f on the chiral stationary phase used for the other compounds was not possible and, as such, enantiomeric excess values for these could not be calculated.

**Table 7** | *EncP-catalysed amination of additional ring-substituted halocinnamates.*

<table>
<thead>
<tr>
<th>1</th>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>βee</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1g</td>
<td>2-Cl</td>
<td>92%</td>
<td>18 : 82</td>
<td>n.d.</td>
<td>81% (S)</td>
</tr>
<tr>
<td>1h</td>
<td>3-Cl</td>
<td>79%</td>
<td>47 : 53</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>58%</td>
<td>57 : 43</td>
<td>&gt;90% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>90%</td>
<td>18 : 82</td>
<td>&gt;90% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>79%</td>
<td>48 : 52</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>45%</td>
<td>61 : 39</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined

Analysis of biotransformations with additional halocinnamates (chlorinated 1g-i and brominated 1j-l, Table 7) revealed similar patterns to those with their fluorinated analogues. The conversion and α-regioselectivity of the reactions followed the same pattern of ortho- > meta- > para-, but with subtle variations. Whilst substituents at the 2- and 3-positions gave similar conversion and α-regioselectivity values to each other across all halogens, the para-substrates showed greater disparities. This subset (1d, i and l) displayed decreased conversion and increased β-selectivity as the substituent became larger and less electronegative. The 4-bromocinnamate 1l showed a significantly increased excess of the β-product, even in comparison to the unsubstituted arylacrylic acid 1a. Again the enzyme was shown to have perfect (S)-enantiopreference for all reactions with substrates (1h-l), bar the slightly lower enantiomeric excess for the α-product of the 2-chloro-compound (1g). As before, the βee of two compounds (2g and 2l) could not be determined with the general analytical methods.
In addition to the halocinnamates, 2-, 3- and 4-substituted nitro- and methoxy-compounds were also found to be accepted by the enzyme (Table 8). The effect of varying the position of a nitro- substitution was found to affect conversion of the substrate in a similar manner to the halogenated compounds. A considerable decrease was observed between the ortho- (1m), meta- (1n) and para- (1o) isomers, with conversions of 93, 67 and 49% respectively. The regioselectivity displayed by the enzyme with these compounds was, however, found to follow a different pattern. Although all three gave high excess of α-amino acid, as coherent with theories for electron-withdrawing groups, the selectivity was found to be higher for the 2- and 4-nitrocinnamates than for 3-nitrocinnamate (5:95 and 18:82 versus 30:70). This could possibly be indicative of resonance effects due to the linkage between the positions incurring greater shifting of regioselectivity. Of the two compounds yielding enough β-amino acid to warrant chiral analysis, ee values for one (1n) could be determined, with only presence of the (S)-form detected (>95% ee). Similar analyses for the α-products revealed imperfect enantiopreference, reminiscent of studies of these substrates with AvPAL, where the enzyme was found to give racemic mixtures for the ortho- and para- isomers, whilst retaining good selectivity with 3-nitrocinnamate.

For EncP the value of 65% (S) for 4-nitrocinnamate is not quite as low as the cognate reaction with AvPAL, possibly due to the low conversion of this substrate (49%). If EncP were to display a build-up of the (R)-α-amino acid with nitro-containing substrates, in a similar manner to AvPAL, it would stand to reason that this would also be conversion dependent, thus giving lower ee values for a higher converted substrate (1m: 92% conv., 6% ee) and vice versa (1o: 49% conv., 64% ee).

Table 8 || EncP-catalysed amination of ring-substituted nitro- and methoxy-cinnamates.

<table>
<thead>
<tr>
<th>1</th>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>β ee</th>
<th>α ee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m</td>
<td>2-NO₂</td>
<td>93%</td>
<td>5 : 95</td>
<td>-</td>
<td>6% (S)</td>
</tr>
<tr>
<td>1n</td>
<td>3-NO₂</td>
<td>67%</td>
<td>30 : 70</td>
<td>&gt;95% (S)</td>
<td>92% (S)</td>
</tr>
<tr>
<td>1o</td>
<td>4-NO₂</td>
<td>49%</td>
<td>18 : 82</td>
<td>n.d.</td>
<td>64% (S)</td>
</tr>
<tr>
<td>1p</td>
<td>2-OCH₃</td>
<td>71%</td>
<td>87 : 13</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1q</td>
<td>3-OCH₃</td>
<td>71%</td>
<td>62 : 38</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1r</td>
<td>4-OCH₃</td>
<td>47%</td>
<td>93 : 7</td>
<td>0%</td>
<td>-</td>
</tr>
</tbody>
</table>

n.d. = not determined, - = not measured
The methoxycinnamates displayed mirrored regioselectivity patterns to their nitro-analogues, giving excesses of the β-product and to a greater extent for ortho- and para-substitutions (87:13 and 93:7 versus 62:38 for 1p and 1q versus 1r). This adheres to observations of similarly electron-donating substituents with other aminomutases14,106 and could also suggest the role of electronic effects at resonant positions (as speculated for the nitrocinnamate substrates). The pattern of conversion was found again to be lowest for the 4-substituted cinnamate derivative (1r: 47%), but was the same for both 2- and 3-methoxycinnamates, contravening the ortho- > meta- precedent set by the nitro-, bromo-, chloro- and fluoro- arine isomers. This may be due to the proximity of the bulky, rotationally-independent nature of the -OCH₃ group to the activated β-carbon of the substrate, causing some sort of steric and / or torsional hindrance to the catalytic residues of the enzyme in the active site. The enzyme was found to give perfect enantiocontrol in the synthesis of the minor products 2p-q for the reactions analysed, and for the excess of product 3q (all >95% (S)). The 4-methoxy-β-product 2r, however, was found to be completely racemic. This could possibly be an artefact of the discriminatory mechanism of PAL and PAM enzymes between 4-hydroxylated and unsubstituted compounds (i.e. tyrosine regioisomers / coumarate and phenylalanine regioisomers / cinnamate). As the two forms of 4-methoxyphenylalanine 2r and 3r (also called O-methyltyrosines) bear structural and electronic similarities to β- and α-tyrosine, it may be that the enzyme disfavours correct binding of these and their acrylic acid analogues, 4-methoxycinnamate and coumarate (or 4-hydroxycinnamate). This may account for the difference in enzymatic activity seen with 1r compared to the other substrates tested thus far. Enantiomers of 2-methoxy-β-phenylalanine 2p were unfortunately not separated with the HPLC method employed for chiral analyses and thus βee values for the reaction with 2-methoxycinnamate remain undetermined.

Table 9 || EncP-catalysed amination of miscellaneous ring-substituted cinnamates.

<table>
<thead>
<tr>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>βee</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Ch₃</td>
<td>80%</td>
<td>41 : 59</td>
<td>85% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>3-Ch₃</td>
<td>71%</td>
<td>77 : 23</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>4-Ch₃</td>
<td>44%</td>
<td>97 : 3</td>
<td>84% (S)</td>
<td>-</td>
</tr>
<tr>
<td>3-CN</td>
<td>28%</td>
<td>21 : 79</td>
<td>n.d.</td>
<td>86% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined, - = not measured
The methylcinnamate substrates 1s-u (Table 9) revealed catalytic differences with EncP that were yet again dissimilar to those seen with other sets of arene substitution isomers. The hierarchy of conversions followed the familiar pattern seen with other subsets of compounds, bar the methoxy-substrates: ortho- (1s: 80%) > meta- (1t: 71%) > para- (1u: 44%). β-selectivity, in line with the electron-donating nature of the methyl moiety, was observed for both 3- and 4-methylcinnamates with a similar pattern as observed with the preceding methoxy-analogues (93:7 versus 97:3 for 4-methoxy- versus 4-methyl- and 62:83 versus 77:23 for 3-methoxy- versus 3-methyl-substituents). Unlike the methoxy-triad, however, there was apparent α-selectivity of the 2-methyl-compound (41:59). Oddly, the β:α ratio with this substrate was found to be lower than some pertaining to substrates with electron-withdrawing groups, including all the para- and meta-halocinnamates (1c-d, h-i and k-l). For biotransformations yielding sufficient α-product (3s-t) enantiomeric excess values were found to be excellent (>95% (S) in both cases). The same was found to be true for 2t, but with 2- and 4-methylcinnamates 1s and 1u, the β-addition reactions were found to proceed with good but imperfect enantioselectivity (under 90% (S) for both). The last compound tested was 3-cyanocinnamate 1v, which, despite giving a low conversion compared with other meta-substrates, gave a β:α product ratio of 21:79; a selectivity consistent with the electron-withdrawing tendency of an -CN group. Although chiral analysis of the β-product of this reaction was not possible, the moderate ee (86% (S)) proved consistent with the slow racemisation of 3v seen in biotransformations with AvPAL.107

![Figure 32](image-url) **Figure 32** | Substrates tested with EncP (α- and β-methylcinnamate 1w and 1x) and found to give no conversion under general reaction conditions.

As well as the 21 ring-substituted cinnamate derivatives, α- and β-methylated arylacrylic acids 1w and 1x were also tested with the EncP whole cell system. The use of these compounds in failed biotransformations with another enantiocomplementary mutase has been reported in the literature and, likewise, neither substrate was found to be accepted by this enzyme.
3.2.3 EncP Regioselectivity Analyses

Of the substrates tested with EncP in the preceding section, none was found to be converted solely to one regioisomer of the corresponding phenylalanine derivative. In general, substrates with electron-donating substituents were found to give more of the β-product relative to the α-, with the reverse being true for compounds with electron-withdrawing groups. This is likely due to the mechanism of nucleophilic attack by the amine moiety of the MIO-NH₂ adduct during the proposed catalytic cycle of the enzyme. Janssen et al. have postulated, with the catalytically similar enzyme TwPAM, that amination is directed initially via MIO-NH₂ adduct formation with the substrate at the more activated double bond carbon of the substrates. This relative activation is influenced by the ability of the double bond-conjugated systems at either end of the substrate molecule to pull electron density away from the central π-system. If the aromatic ring is electron-deficient enough to compete with the carboxyl then a partial positive charge is more likely to be stabilised on the α-carbon of the substrate. With sufficient electron-richness the aromatic ring becomes a poorer sink compared to the carboxyl, causing activation of the β-carbon instead.⁴,¹⁰⁶ Although this theory is consistent with many of the data obtained with EncP, there are a handful of examples which seem to disrupt the pattern. As a way of better understanding the ability of the ring of each substrate to pull electrons from the centre of the molecule, a quantitative (rather than arbitrary) measure was sought.

Position- and substituent-specific core electron binding energy shift (ΔCEBE) values have been demonstrated to be useful chemical descriptors in theoretical analyses of monosubstituted benzenes.¹⁰⁸ These values are calculated computationally as the difference between the core electron binding energy (CEBE: energy required to remove a single core electron) at the ipso-carbon of an aryl ring, compared to that at the equivalent carbon in a substituted ring.¹⁰⁸ A positive shift (i.e. an increase in CEBE) indicates removal of electron density from the ring by a substituent, causing the remaining electrons to be more tightly bound by the six carbons in the ring system. A negative shift, or decrease in CEBE, shows substituent electronic contribution to the aromatic system, resulting in more electron density to be bound by the same number of nuclei. The extent to which the CEBE shifts in either direction is proportional to the electron richness of the ring, as affected by the type and position of specific arene substitutions. As the electron richness of the ring relates directly to its ability to withdraw electron density relative to the carboxyl, ΔCEBE values were used in an attempt to describe the regioselectivity pattern of EncP with various substrates in a more quantitative manner.
The dependence of β- vs. α-amination by EncP on the core electron binding energy shift (ΔCEBE) due to substrate ring substituents.

As a test of the contribution of electronic effects to the regioselectivity of EncP amination, ΔCEBE values corresponding to all monosubstituted substrates were plotted against the percentage β-amino acid composition of the product from the various biotransformations. Overall, the data points displayed a fair degree of scatter best explained by two trend lines: one for the majority of ortho-substituents and another steeper line for the remainder of the data points. The incidence of these two subgroups, both with varying negative correlations, suggests that electronic effects contribute to the regioselectivity of EncP-catalysed amination for all substrates, but that they are less influential for ortho-substituted compounds. This lack of a single strong negative correlation led to speculation that other factors must affect the selectivity of the enzyme. Both lines gave reasonable R² values of >0.9, revealing a couple of seemingly outlying examples. The unsubstituted cinnamate 1a, for example, was found to give similar or even lower amounts of β-product compared to the para-halocinnamates (1d, i and l) despite their more electron-deficient rings. The 4-methyl- (1u) and 3-methoxycinnamates (1q) were also found to lie away from the trend line. These anomalies possibly point to interactions of different substituents in the active site affecting substrate positioning. Likewise, the clustering of substrates with similarly-positioned substituents supported the hypothesis of two binding modes in the active site of the enzyme, with one more prevalent for ortho-substituted cinnamates, favouring α- over β-addition.
3.2.4 Rational Design of EncP

The mixtures of products given for each substrate make the wild-type enzyme an undesirable biocatalyst for the production of (S)-β-amino acids, due to the contaminating α-form in all reactions. As such, it was deemed necessary to undertake engineering of the enzyme to diminish unwanted side activities. The potential interplay of substrate electronics and binding hinted at the possibility of altering the enzyme active site such as to provide better catalyst control over the amination reaction. This already seems to be the case with certain substrates which deviate from the correlation shown in the previous section, where the different way in which the enzyme binds these substrates may change the regioselectivity predicted by electronic effects alone. To explore the potential role of substrate positioning in this regard, structural data for the related aminomutase AdmH were analysed. The structure of this enzyme with both regioisomers of the phenylalanine covalently attached to the MIO group (PDB ID: 3UNV)\textsuperscript{13} was downloaded and used to create a homology model of EncP. The active site regions of the model and template structure were found to be almost identical. As such, the model was analysed and substrate binding interactions predicted via visual inspection. The most prominent feature in this regard was found to be the coordination of each oxygen in the carboxyl group: one by a monodentate salt bridge to R299 and the other via hydrogen bonding to nearby N340. These residues in turn were inferred to interact with E311 via an ionic interaction with the arginine and a further hydrogen bond with the asparagine. In both the model and original structure of AdmH, these three residues appeared to be responsible for correct positioning of the substrate between the MIO and catalytic tyrosine, such as to allow amination at either the β- or α-carbon.
EncP active site catalytic and substrate positioning residues as inferred by homology modelling based on the previously solved structure of AdmH. (a) shows the inferred interactions and reaction mechanism hypothesised from the visualised model (b).

To probe the effect of some of these positions on the substrate binding and possibly regioselectivity of EncP, structure- and sequence-guided design of the enzyme active site was undertaken. An attempt to engineer the carboxyl-binding pocket of a plant PAL (PcPAL1 from Petroselinum crispum) described many difficulties with retaining enzyme activity. This was rationalised to be due to a delicate hydrogen bonding network in this area of the active site, sensitive to modification. The first position targeted for mutagenesis was R299; a residue completely conserved at all homologous positions in other class I lyase-like enzymes (as observed in the sequence analyses of Chapter 2). As such, the charge interaction enabled between this residue and the substrate was deemed to be integral to binding and it was reasoned that changing it would significantly alter positioning of any substrates bound. The obvious choice for mutagenic studies was the introduction of a lysine at this position to
conserve, yet alter the nature of the substrate binding interaction at this position. In the literature there was evidence of mutations at the homologous position in a HAL yielding enzyme variants with residual catalytic activity. Studies of the ammonia lyase from *Pseudomonas putida* revealed that substituting R283 for the functionally-equivalent K still allowed detectable formation of urocanate upon incubation with (S)-α-histidine, albeit at a rate 20 times lower. Removal of the interaction by means of an R283I substitution, however, caused an approximate 1600-fold decrease in activity.\(^\text{12}\) Although enzyme activity was retained in the R=>K substitution, it was the much faster, natural deamination reaction which was tested. It was possible that a similar change in EncP would prove detrimental to the slower, less favourable amination reaction.

Additionally, residue E293 was considered for mutational analysis. This was due to its proximity to and theorised interactions with both of the carboxyl binding residues (R299 and N340). It was hoped that by varying this residue, the positioning of the arginine and asparagine would be shifted in such a way as to reposition the substrate in the active site and possibly influence regioselectivity. From the sequence alignments in Chapter 2 it was evident that this position was only ever glutamate in EncP, AdmH and the putative (S)-PAMs discovered in Section 2.2.6.5. In all the other PALs, TALs, HALs, (R)-PAMs and TAMs studied this position was invariably a glutamine. As well as being the only two amino acids present at this position across the enzyme family, Q also shows the closest shape complementarity to E of all the other proteinogenic amino acids. As the amino acids were so similar and Q was seen to be tolerated at this position in other family members, an E293Q substitution was decided upon.

![Comparison of the side chains of functionally-equivalent, basic arginine (R) and lysine residues (K) along with the glutamate (E) and the most similarly-shaped amino acids - glutamine (Q) and methionine (M).](image)

**Figure 35** || Comparison of the side chains of functionally-equivalent, basic arginine (R) and lysine residues (K) along with the glutamate (E) and the most similarly-shaped amino acids - glutamine (Q) and methionine (M).
Both EncP-R299K and EncP-E293Q were created via mutagenesis of the specific-codon for each amino acid in the *encP* gene. This was performed using two specifically designed oligonucleotides (one containing the desired codon change) to prime a polymerase chain reaction (PCR) of the entire template plasmid. Correct replication and presence of the desired mutation(s) were confirmed by restriction enzyme analysis and nucleotide sequencing respectively. The mutated plasmids were used to create whole cell biocatalyst containing the enzyme variants in the same manner as for the wild-type enzyme, without optimisation of gene expression conditions. Both biocatalysts were tested with cinnamate in the same conditions as for the wild-type biotransformations and found to shift regioselectivity in opposing directions. The R299K variant gave a β:α ratio of 90:10 - an improvement in β-selectivity compared to the more equal distribution with the wild-type (56:44). Conversely, the product ratio obtained with EncP-E293Q was 43:57 - a switch in selectivity from the slight β-excess obtained with the wild-type. For this variant the change from a charged to a polar residue at position 293 was the only difference that resulted in a switching of regioselectivity. To extend this pattern, a less-polar residue with the closest shape similarity possible was sought for further substitution of this residue. Methionine was chosen for this due to its lack of polarity and similar side chain length to glutamate and glutamine. The EncP-E293M variant was created and tested in the same way as the other engineered enzymes and found to give a further shift in regioselectivity (with a ratio of 18:82 β- to α-phenylalanine).

**Figure 36** || The effect of temperature on the overall conversion and regioselectivity of amination reactions catalysed by EncP and its rationally-designed variants.

Although all three variants gave promising shifts in regioselectivity, highlighting the importance of the residues studied in substrate positioning, conversion was found to be lower for reactions with each than for those with the wild-type enzyme. This was
hypothesised to be due to the probable disruption of the hydrogen bonding network described in previously reported work\textsuperscript{109} and the effect of this on catalytic activity. To see if improvements in conversion could be achieved, the temperature optimum of the reaction for each variant was tested as for the wild-type enzyme (Section 3.2.1) and the results compared (Figure 36). It was found that, for the R299K variant reactions, optimal conversion was achieved at lower temperature with activity decreasing for reactions above 40°C and no detectable activity at 85°C. This may have been due to the probable deformation of the carboxyl binding site in EncP. Interestingly the reactions with EncP-E293Q were found to give better conversion at higher temperature, with only a small difference in conversion between the 70 and 85°C biotransformations compared to the sizeable drop for the wild-type enzyme (and to a lesser extent with the E293M variant) between these two temperatures. This improved thermoactivity possibly indicates that the presence of a Q at this position is beneficial to the integrity of the active site, whereas an E is detrimental. This might explain the conservation of glutamine at corresponding positions in all class I lyase-like enzymes, other than the characterised and putative (S)-PAMs (including PcPAL1 - the enzyme where the carboxyl binding hydrogen bonding network was originally studied). As this position is also seen to affect regioselectivity it may be that the presence of an E here in (S)-PAMs promotes addition at the β-position, this providing a physiologically significant enhancement of aminomutase activity selected for in these enzymes despite impairment of their stability / catalytic efficiency.

### 3.2.5 Substrate Scope of EncP Regioselective Variants

With cinnamate 1a it would be expected that an excess of β-amino acid would be produced (assuming equal access of the catalytic residues to both carbons), due to the ability of the carboxyl group to act as an electron sink. This seemed to be the case with the wild-type enzyme when first tested (Section 3.2.3, Figure 33). The rationally designed variants from Section 3.2.4 were demonstrated to give shifted regioselectivity with this substrate, possibly due to changes in active site positioning aiding (in the case of R299K) or overcoming (E293 variants) this tendency. To see if similar overriding of electronic effects via catalyst control could be achieved with compounds of differing electronics, the three variants with shifted regioselectivity were tested with the substrate panel accepted by the wild-type enzymes in Section 3.2.2. The reaction conditions were kept the same as for EncP-WT, as changes in temperature had been shown to alter regioselectivity to some extent with all EncP variants. In the interest of production of (S)-β-phenylalanine derivatives in particular, the original
temperature of 55°C was used as this was demonstrated to give the best β-selectivity with EncP-R299K (Section 3.2.4, Figure 36).
Table 10 || Amination of various arylacrylic acids to the corresponding amino acids by the EncP-R299K rationally designed enzyme variant.

\[
\text{EncP} \quad \text{(wet cells, 40 mg mL}^{-1}\text{)} \quad 0.8 \text{ mL} - 4 \text{ M (NH}_4\text{)}_2\text{SO}_4 \\
(\text{pH 8.3}) 55^\circ \text{C}, 22 \text{ h}
\]

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Conv.</th>
<th>$\beta : \alpha$ ratio</th>
<th>$\beta \text{ee}$</th>
<th>$\alpha \text{ee}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>31%</td>
<td>88:12</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1b</td>
<td>2-F</td>
<td>50%</td>
<td>48:52</td>
<td>65% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1c</td>
<td>3-F</td>
<td>49%</td>
<td>52:48</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>30%</td>
<td>&gt;99:1</td>
<td>&gt;95% (S)</td>
<td>-</td>
</tr>
<tr>
<td>1e</td>
<td>3,5-F$_2$</td>
<td>74%</td>
<td>19:81</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1f</td>
<td>2,3,4,5,6-F$_5$</td>
<td>95%</td>
<td>1:99</td>
<td>-</td>
<td>n.d.</td>
</tr>
<tr>
<td>1g</td>
<td>2-Cl</td>
<td>58%</td>
<td>51:49</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1h</td>
<td>3-Cl</td>
<td>38%</td>
<td>78:22</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>36%</td>
<td>&gt;99:1</td>
<td>&gt;90% (S)</td>
<td>-</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>48%</td>
<td>55:45</td>
<td>&gt;90% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>31%</td>
<td>88:12</td>
<td>&gt;95% (S)</td>
<td>n.d.</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>27%</td>
<td>&gt;99:1</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>1m</td>
<td>2-NO$_2$</td>
<td>72%</td>
<td>2:98</td>
<td>-</td>
<td>68% (S)</td>
</tr>
<tr>
<td>1n</td>
<td>3-NO$_2$</td>
<td>16%</td>
<td>17:83</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1o</td>
<td>4-NO$_2$</td>
<td>27%</td>
<td>7:93</td>
<td>-</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1p</td>
<td>2-OCH$_3$</td>
<td>10%</td>
<td>&gt;99:1</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>1q</td>
<td>3-OCH$_3$</td>
<td>13%</td>
<td>&gt;99:1</td>
<td>&gt;95% (S)</td>
<td>-</td>
</tr>
<tr>
<td>1r</td>
<td>4-OCH$_3$</td>
<td>26%</td>
<td>&gt;99:1</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>1s</td>
<td>2-CH$_3$</td>
<td>36%</td>
<td>&gt;99:1</td>
<td>84% (S)</td>
<td>-</td>
</tr>
<tr>
<td>1t</td>
<td>3-CH$_3$</td>
<td>28%</td>
<td>&gt;99:1</td>
<td>&gt;95% (S)</td>
<td>-</td>
</tr>
<tr>
<td>1u</td>
<td>4-CH$_3$</td>
<td>25%</td>
<td>&gt;99:1</td>
<td>86% (S)</td>
<td>-</td>
</tr>
<tr>
<td>1v</td>
<td>3-CN</td>
<td>n.c.</td>
<td>n.d.</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

n.d. = not determined, - = not measured
3.2.5.1 EncP-R299K

The R299K variant of EncP gave product(s) with all of the substrates accepted before, expect for 3-cyanocinnamate 1v, where neither of the counterpart amino acids (2v or 3v) could be detected. In all instances, apart from one, the conversion observed was lower than for the original enzyme. In fact only 4 of the 22 substrates were found to give a conversion of greater than 50% (3,5-difluoro- 1e: 74%, 2,3,4,5,6-pentafluoro- 1f: 95% - the same as for EncP-WT, 2-chloro- 1g: 58% and 2-nitrocinnamate 1m: 72%). In this respect, the results were similar to the initial test with cinnamate (1a), where lower conversion was explained by the possible disruption of key active site interactions, likely caused by the removal of the arginine side chain from position 299. The pattern of conversion was also similar to that for the wild-type enzyme, with ortho-arene substitutions giving the highest consumption of starting material and their meta- and para- isomers giving lower, more comparable values.

The enzyme was also found to give increased β-selectivity for the majority of substrates, with no detectable trace of the α-regioisomer observed for 9 of the compounds (the para-halocinnamates 1d, i and l and all methyl- / methoxycinnamates 1p-u). Substrates showing little shift in regioselectivity - such as the nitro-compounds 1m-o and multifluorinated cinnamates 1e and f - all possessed classically electron-deficient aryl rings. This shows that even though the engineering efforts with this enzyme to yield increased excess of β-product were successful, substrates with rings subject to stronger electron-withdrawal still undergo primarily α-amination.

The enzyme was also shown to be perfectly (S)-selective in the formation of the majority of analyzable products. Exceptions to this included 3m, the highest converted of the nitrocinnamates, which was the only α-product to show evidence of the (R)-enantiomer. As this was the α-amino acid with the lowest enantiomeric excess in aforementioned experiments (6% (S) with EncP-WT) it is possible that this result is likewise explained by the presence of MIO-independent, non-enantioselective amination activity. The α-ee may be higher in this instance due to the lower conversion of this substrate with the R299K variant (72 versus 93%). For the β-amination reactions only 2-fluoro- 1b and 2- and 4-methylcinnamates 1s and u gave more moderate enantiomeric excess values. For the methyl-compounds the βees were almost identical to reactions with the wild-type, whereas for the fluoro-compound selectivity was somewhat lower - 65 versus 75% (S) - in spite of lower conversion with the more regioselective variant. The 4-methoxy-β-phenylalanine 2r produced with EncP-R299K was also found to be racemic, as before.
Table 11 | Amination of various arylacrylic acids to the corresponding amino acids by the EncP-E293Q rationally designed enzyme variant.

![Diagram](#)

<table>
<thead>
<tr>
<th>1</th>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>βee</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>64%</td>
<td>43 : 57</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1b</td>
<td>2-F</td>
<td>86%</td>
<td>15 : 85</td>
<td>76% (S)</td>
<td>50% (S)</td>
</tr>
<tr>
<td>1c</td>
<td>3-F</td>
<td>77%</td>
<td>32 : 68</td>
<td>&gt;98% (S)</td>
<td>54% (S)</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>68%</td>
<td>51 : 49</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1e</td>
<td>3,5-F₂</td>
<td>69%</td>
<td>26 : 74</td>
<td>n.d.</td>
<td>61% (S)</td>
</tr>
<tr>
<td>1f</td>
<td>2,3,4,5,6-F₅</td>
<td>89%</td>
<td>2 : 98</td>
<td>-</td>
<td>n.d.</td>
</tr>
<tr>
<td>1g</td>
<td>2-Cl</td>
<td>90%</td>
<td>12 : 88</td>
<td>n.d.</td>
<td>47% (S)</td>
</tr>
<tr>
<td>1h</td>
<td>3-Cl</td>
<td>81%</td>
<td>33 : 67</td>
<td>&gt;95% (S)</td>
<td>61% (S)</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>56%</td>
<td>53 : 47</td>
<td>&gt;90% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>89%</td>
<td>13 : 87</td>
<td>&gt;90% (S)</td>
<td>54% (S)</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>84%</td>
<td>30 : 70</td>
<td>&gt;95% (S)</td>
<td>61% (S)</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>44%</td>
<td>56 : 44</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1m</td>
<td>2-NO₂</td>
<td>93%</td>
<td>4 : 96</td>
<td>-</td>
<td>10% (S)</td>
</tr>
<tr>
<td>1n</td>
<td>3-NO₂</td>
<td>73%</td>
<td>10 : 90</td>
<td>&gt;95% (S)</td>
<td>76% (S)</td>
</tr>
<tr>
<td>1o</td>
<td>4-NO₂</td>
<td>31%</td>
<td>1 : 99</td>
<td>-</td>
<td>73% (S)</td>
</tr>
<tr>
<td>1p</td>
<td>2-OCH₃</td>
<td>54%</td>
<td>37 : 63</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1q</td>
<td>3-OCH₃</td>
<td>73%</td>
<td>49 : 51</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1r</td>
<td>4-OCH₃</td>
<td>46%</td>
<td>80 : 20</td>
<td>0%</td>
<td>n.d.</td>
</tr>
<tr>
<td>1s</td>
<td>2-CH₃</td>
<td>64%</td>
<td>35 : 65</td>
<td>73% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1t</td>
<td>3-CH₃</td>
<td>70%</td>
<td>58 : 42</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1u</td>
<td>4-CH₃</td>
<td>44%</td>
<td>80 : 20</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1v</td>
<td>3-CN</td>
<td>37%</td>
<td>4 : 96</td>
<td>-</td>
<td>86% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined, - = not measured
3.2.5.2 EncP-E293Q

Unlike EncP-R299K, EncP-E293Q was found to convert all the substrates accepted by the wild-type enzyme. The conversions were also more comparable to the original enzyme, with some results a little lower than the biotransformations with EncP-WT and one substrate, 3-cyanocinnamate 3v, more highly converted than previously (37 versus 26%). The trend in conversion values of ortho- > meta- > para- for homosubstituted substrate isomers was maintained with the fluoro- (1b-d), chloro- (1g-i), bromo- (1j-l) and nitro-cinnamates (1m-o). In this case however the ortho- and meta-compounds gave more comparable conversions with much less of the para-isomer being converted in each case. This is opposite to the more comparable meta- and para-cinnamate conversions seen with these substituents with the R299K variant, and different again from the more evenly spaced conversion values for these substrates with wild-type EncP. For compounds with electron-donating substituents (the methoxy- and methylcinnamates 1p-u), however, the 3-substitution seemed to aid conversion as compared to those at the 2- or 4-position. This is a step beyond the pattern with the methoxycinnamates in Section 3.2.2 where the ortho- and meta-compounds gave identical conversions, higher than the para-, similar to that of the halo- and nitrocinnamates in this section.

Regioselectivity was subtly different with this variant, with most compounds giving only minor decreases in β:α product ratio. For 2-methoxycinnamate 1p this resulted in a switch from mostly β- (87:13) to predominantly α-product (37:63), as shown before with the initial test of this variant with cinnamate. With regards to enantioselectivity the enzyme was shown to catalyse completely (S)-selective amination at the β-carbon in all cases apart from two ortho-substituted compounds: 2-fluoro- 1b and 2-methylcinnamate 1s with moderate βee values of 76 and 74% (S) respectively. As before the enzyme was completely non-selective in the production of 2r. At the α-position, however, only half of the ee values that could be obtained were >95% (S). Of these compounds 1b, e, g, m-o and v had already been shown to give imperfect αees with EncP (Section 3.2.2) but with better selectivity than with the E293Q variant in many cases - e.g. 2-fluorocinnamate 1b giving just 50 as opposed to 83% (S) earlier. In addition to these the 3-halocinnamates 1c, h and k gave αees of 54, 61 and 61% (S) which contrasts greatly with the perfect enantioselectivity demonstrated for these same substrates with EncP-WT (all >95%). This apparent lowering of enantioselectivity could be due to the different active site architecture enhancing a MIO-independent, non-selective α-amination via access to different conformations of the substrates in the active site.
Table 12 || Amination of various arylacrylic acids to the corresponding amino acids by the EncP-E293M rationally designed enzyme variant.

Table:<br>
<table>
<thead>
<tr>
<th>1</th>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>βee</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>54%</td>
<td>18 : 82</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1b</td>
<td>2-F</td>
<td>82%</td>
<td>5 : 95</td>
<td>-</td>
<td>56% (S)</td>
</tr>
<tr>
<td>1c</td>
<td>3-F</td>
<td>73%</td>
<td>16 : 84</td>
<td>n.d.</td>
<td>87% (S)</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>61%</td>
<td>30 : 70</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1e</td>
<td>3,5-F2</td>
<td>77%</td>
<td>10 : 90</td>
<td>-</td>
<td>48% (S)</td>
</tr>
<tr>
<td>1f</td>
<td>2,3,4,5,6-F5</td>
<td>92%</td>
<td>1 : 99</td>
<td>-</td>
<td>n.d.</td>
</tr>
<tr>
<td>1g</td>
<td>2-Cl</td>
<td>89%</td>
<td>6 : 94</td>
<td>-</td>
<td>60% (S)</td>
</tr>
<tr>
<td>1h</td>
<td>3-Cl</td>
<td>81%</td>
<td>11 : 89</td>
<td>n.d.</td>
<td>74% (S)</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>55%</td>
<td>28 : 72</td>
<td>&gt;80% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>86%</td>
<td>5 : 95</td>
<td>-</td>
<td>72% (S)</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>82%</td>
<td>7 : 93</td>
<td>-</td>
<td>62% (S)</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>45%</td>
<td>25 : 75</td>
<td>n.d.</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1m</td>
<td>2-NO2</td>
<td>93%</td>
<td>3 : 97</td>
<td>-</td>
<td>20% (S)</td>
</tr>
<tr>
<td>1n</td>
<td>3-NO2</td>
<td>71%</td>
<td>2 : 98</td>
<td>-</td>
<td>78% (S)</td>
</tr>
<tr>
<td>1o</td>
<td>4-NO2</td>
<td>35%</td>
<td>3 : 97</td>
<td>-</td>
<td>70% (S)</td>
</tr>
<tr>
<td>1p</td>
<td>2-OCH3</td>
<td>20%</td>
<td>1 : 99</td>
<td>-</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1q</td>
<td>3-OCH3</td>
<td>59%</td>
<td>18 : 82</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1r</td>
<td>4-OCH3</td>
<td>35%</td>
<td>71 : 29</td>
<td>0%</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1s</td>
<td>2-CH3</td>
<td>80%</td>
<td>24 : 76</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1t</td>
<td>3-CH3</td>
<td>58%</td>
<td>26 : 74</td>
<td>&gt;95% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1u</td>
<td>4-CH3</td>
<td>54%</td>
<td>59 : 41</td>
<td>82% (S)</td>
<td>&gt;95% (S)</td>
</tr>
<tr>
<td>1v</td>
<td>3-CN</td>
<td>29%</td>
<td>5 : 95</td>
<td>-</td>
<td>&gt;95% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined, - = not measured
3.2.5.3 EncP-E293M

The final biocatalyst EncP-E293M was found to work with the full panel of 22 substrates, like the parent enzyme and the other α-selective variant. The conversions were comparable to these other two and did not seem to suffer a decrease as with R299K. EncP-E293M also shared the pattern of conversion between 2-, 3- and 4-halo- and nitro- and methoxycinnamate isomers with the E293Q variant. Oddly with the methyl-compounds 1s-u the pattern was not the same as EncP-E293Q or EncP-WT, but more similar to the β-selective variant (with conversions of 80, 58 and 54% as compared to 36, 28 and 25% for ortho-, meta- and para-methylcinnamates 1s, t and u respectively). The shift towards α-selectivity was more apparent across the entire substrate panel than with the glutamine variant and only two compounds 1r and u, both with an electron-donating substituent at the 4-position of the ring, gave the β-amino acid as the major product. All other compounds gave β:α ratios of less than or equal to 30:70. The two predominantly β-forming substrates were also the only cases where βee was shown to be imperfect (82% (S) for 1u and 0% for 1r). In all other reactions, where chiral analyses of the β-product were warranted and achieved, only the (S)-enantiomer of 2 could be identified. With the α-regioisomers, analyses revealed that, like with EncP-E293Q, only 10 substrates were aminated with perfect enantiocontrol. This seems to reinforce the trend of enantioselectivity of α-amination being affected negatively by decreased β:α ratio (R299K > WT > E293Q / E293M). Again this may be due to a reduction / exacerbation of non-selective amination at the α-position by these amino acid variations causing an accumulation of the opposite enantiomer.

3.2.6 Substrate Binding versus Electronic Effects in EncP

It could be seen in a mostly qualitative manner that the electronic effects on regioselectivity, as discussed before, seemed to have been minimised for at least a subset of substrates with each rationally-designed enzyme variant. The link between inherent electronic and catalyst-conferring substrate positioning effects was probed once again by comparisons with ΔCEBEs as with the original enzyme. As a test of the significance of the two trend lines observed in the initial analyses, these were superimposed onto the new plots to allow any differences to be seen more clearly (Figure 37).
Figure 37 || The dependence of β- vs. α-amination by EncP rationally designed variants on the core electron binding energy shift (ΔCEBE) due to substrate ring substituents. Trends for the wild-type enzyme are shown as dotted lines for comparison.
From the new graphs it was evident that the correlation between position and substituent specific ΔCEBEs and prevalence of β-amino acid product had been disrupted with each biocatalyst. With EncP-R299K, although the withdrawing effect of any nitro-groups seemed to outweigh substrate repositioning, all other data points had experienced a significant upward shift from their original positions. All 2-halo-examples, for instance, now gave results more consistent with the steeper trend line, as opposed to the ortho-substituent trend line they had previously occupied. Most other compounds gave 99:1 β:α product ratios irrespective of ΔCEBE values (from ~-0.4 to 0.2 eV). Interestingly the unsubstituted compound and 3-halocinnamates were not acted on with complete regiocontrol (in fact 3-fluoro- seemed to occupy a position very similar to before. Overall this demonstrates that the R299K amino acid exchange mitigates electronic effects to a large extent for the majority of compounds, with a few discrepancies where specific substrate electronics still allow varying levels of α-product formation (10 examples, β:α ratios between 88:12 and 2:98).

For EncP-E293Q the disruption of the original correlation was less considerable. The para-data points were still seen on the steeper trend line and the ortho-points just below the shallower line, but the majority of 3-substituted compounds and the unsubstituted cinnamate now gave values which placed them at an intermediary position between the two correlation lines. One exception to this was 3-cyanocinnamate which gave a prevalence of β-product more consistent with the correlation for ortho-compounds. The most interesting examples with this variant was the regioselectivity shift observed with 2-methoxycinnamate. This compound had been seen to behave like the para- and meta-compounds with EncP-WT but was shown to migrate fully to the ortho-trend line after biotransformation with the glutamine variant. This splitting of data points into ring position groups possibly indicates the presence of more binding modes - one most favoured for each substituent position - which have small effects on the ease of amination at either the β- or α-carbon. Alternatively it could indicate the two originally proposed modes, with the intermediate data points (3-substituted arylacrylic acids) representing more equal incidents of binding between the two as compared to substrate accommodation in the wild-type active site. In any case, the presence of a para-substituent seemed to favour β-relative to α-addition, with the opposite being true for ortho-substituents, even when ΔCEBE values were the same between ring isomers.

The E293M variant gave results where most of the data were clustered around or under the shallower trend line from the wild-type reactions. This, like with the variant of opposite regioselectivity, was evidence of disruption of electronic favouring of regioselective amination as only a slight correlation could be observed with a large degree of scatter. Despite this, it seemed again that there were still position-specific determinants of regioisomeric product prevalence not fully consistent with ΔCEBE values.
The hypothesis is that varying active site residues and para-substituents on substrate positioning and ammonia addition preference in EncP and variants. (a) represents the positioning in the wild-type enzyme, (b) shows the α-selective variants E293Q/E293M and (c) shows the β-selective variant R299K.

The effects of substrate electronics on the amination preference of aminomutase enzymes has been extensively studied\textsuperscript{14,56,106} and the results obtained are further substantiated by the work in this chapter. Theories as to the relative contribution of variations in substrate binding on this are, however, somewhat lacking in the literature. In an attempt to rationalise the deviations seen in this study from patterns based purely on electronic effects, the predicted substrate positioning interactions mentioned previously in the chapter were analysed in conjunction with the regioselectivity data obtained with the all four versions of EncP on substrates with different substitution patterns.
For the wild-type enzyme the prevalence of β-addition was correlated strongly with the electron-richness of the ring of each substrate. This implied that the natural substrate positioning (Figure 38a) gave more or less equal access to both potentially electrophilic carbons, allowing the maximum distribution of β:α ratios courtesy of electronics (97:3 for 4-methyl-\textit{1u} to 5:95 for 2-nitrocinnamate \textit{1m}). However the \textit{para}-halocinnamates \textit{1d}, \textit{i} and \textit{l} with electron-deficient rings (ΔCEBEs of 0.163 0.143 and 0.120 eV respectively) were shown to give comparable or even greater excess of the β-product when compared to the unsubstituted substrate \textit{1a} - which has a less electron-deficient ring in comparison (0 eV). In fact the larger the atomic radius of the halogen at this position, the more β-addition was observed. This implied that larger \textit{para}-substituents, which protrude in such a way as to elongate the arylacrylic acid, cannot be accommodated in the aryl binding pocket as easily in a mode which allows neutral β- / α-amination. Hence they are probably more likely to be placed in a more β-favouring manner, where the α-carbon is pushed away from the catalytic MIO moiety. This may also explain why \textit{para}-substituted cinnamate derivatives were generally less highly converted than their \textit{meta}-isomers (where the substituent juts out at a less laterally restrictive angle) and why \textit{ortho}-substituents seemed to give the highest conversion. In fact the larger the ring-substituent, the greater the disparity between \textit{ortho}- and \textit{para}-conversion values was seen to be, revealing that the larger the \textit{para}-substituent, the less well it fit in the active site and the lower its rate of turnover. The higher α-addition prevalence of \textit{ortho}-substituted cinnamates, when compared to substrates with similar ΔCEBE values but substituents at different positions, could be due to the less intrusive angle of a group at the 2-ring carbon. This would allow a binding mode giving better accommodation of a 2-substituted ring in the aryl binding pocket, thus pushing the α-carbon further over the MIO and the β-carbon out from this position. This may also provide an explanation as to why 2-methylcinnamate (\textit{1s}), with a considerably electron-rich ring (-0.276 eV) paradoxically gave an excess of its α-amino acid product \textit{3s}. Better binding in the active site of \textit{ortho}-substitutions may also shed light on the fact that 2-halo- and nitrocinnamates \textit{1b}, \textit{g}, \textit{j} and \textit{m} gave essentially equally high conversions, despite having very different sizes of substituents. The slightly lower conversion for 2-methylcinnamate \textit{1s} (80 versus <90%) may be due to a combination of electronic deactivation of the α-carbon by the very electron-rich ring (ΔCEBE: -0.249 eV) and amination being favoured at this position by substrate positioning.

For the E293 variants it could be hypothesised that decreasing the interaction between this residue and the positively charged R299 (with E interacting more than Q and both interacting more than M) changed the positioning of this residue. A weaker interaction here was predicted to allow greater movement of this residue (and the substrates it binds) away from
the carboxyl-binding pocket, allowing better access to the α-carbon by the MIO group (Figure 38b). This could also be the case for N240, the other residue thought to interact with both the substrate and position 293 through hydrogen bonding interactions. Repositioning of either of these substrate-binding residues could also go some way to explaining why EncP-E293Q and EncP-E293M suffered greater losses in conversion between 3- and 4-substituted cinnamates making the 2- and 3-isomers more comparable in this respect. Assuming the previous hypothesis that para-substituents are less well accommodated due to the shape of the aryl binding pocket, anything which pushes the substrate further into this region would be most detrimental to productive binding of 4-substituted substrates, as they protrude out along the plane of movement. This may also be the reason why the para-substituted cinnamates still showed the steepest dependence on substrate electronics with the α-selective variants, as the enzyme was less able to position them in such a way as to occlude amination at the β-carbon; they would be more likely to be bound in a more neutral / more wild-type mode.

With the β-selective variant the different chain length and charge density of arginine versus lysine was likely the factor which changed positioning of any accepted substrates in the active site. This could have been through variation of the interaction strength / distance between either position 299 and the substrate, positions 299 and 293 or a combination of the two. In any case, altered ionic interactions in this region would result in a pulling of the ligand closer to the carboxyl binding pocket in such a way as to situate the β-carbon in a more favourable position over the MIO residues relative to the α-, increasing β-amination (Figure 38c). Such circumstances in the active site would also reduce any unfavourable interactions of arene substituents of the substrate with the aryl binding pocket. This may be the reason for the more comparable meta- and para-conversions for homosubstituted substrates (in contrast to the wild-type and E293 variants), as a β-favouring mode of substrate binding would be predicted to allow equally favourable accommodation of both isomers in the active site. It may also be why, in the ΔCEBE plots (Figure 37) the R299K data points shows less scatter than the other graphs - the reaction is less affected by interactions of ring substituents with the aryl binding pocket of the enzyme active site.

3.2.7 Time Course Experiments with Methylcinnamates

Any deviations from excellent enantioselectivity with EncP and its rationally designed variants, with regards to α-amination, could be largely compared with the MIO-independent, non-selective side activity of bacterial ammonia lyases with substrates with electron-
withdrawing substituents. For what concerns amination at the β-position, however, no such studies have ever been undertaken. It seems unlikely that the imperfect enantipreference of the enzymes could be due to a MIO-independent, Michael-type addition of ammonia within the active site of the enzyme. To ascertain if the minor (R)-enantiomer was seen to build up in a manner similar to previous reports, time course experiments were performed with a subgroup of substrates. EncP-R299K was chosen with the substrates 2-, 3- and 4-methylcinnamate 1s-u. These reactions were chosen as they were seen to be completely regioselective in earlier studies but the enantiopurity of the β-products was not found to be perfect in all cases. Reactions were set up as for previous experiments, but with six samples per substrate, so as to allow time points to be taken at 1, 2, 4, 8, 22 and 48 hours. The results are summarised in Tables 13-15.

In all cases the enzyme was found to give no detectable trace of the α-regioisomer, even after 48 hours, demonstrating the excellent catalyst control in this designed enzyme variant. For the 2-methyl-product 2s, chiral analyses revealed that the enantiomeric excess was seen to decrease over time with increasing conversion. This did seem to indicate an accumulation of the minor enantiomer in a way similar to that observed with AvPAL with α-amino acids.\textsuperscript{10,107} For the meta- and para-isomers, however, this did not seem to be the case. Instead the enantiomeric excess for both was found to be relatively stable throughout the entire time course experiment. This possibly indicates that the amination and deamination rates are comparable on either face of the substrate, but that (S)-selective reactions are favoured overall, giving a constant enantiomeric excess without the (R)-form building up over time. Unfortunately not all of the enantiomeric excess values for the product of the EncP-R299K reaction with 4-methylcinnamate 1u could be obtained. The analyses which gave meaningful results do, however, allow a representative view of the enantiomeric excess of the product throughout the course of the reaction.
Table 13 || *Time course study of the amination of 2-methylcinnamate to the 2-methyl-β-phenylalanine by the EncP-R299K variant.*

<table>
<thead>
<tr>
<th>Time / h</th>
<th>Conv.</th>
<th>β : α-ratio</th>
<th>βee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10%</td>
<td>&gt;99 : 1</td>
<td>&gt;96% (S)</td>
</tr>
<tr>
<td>2</td>
<td>19%</td>
<td>&gt;99 : 1</td>
<td>95% (S)</td>
</tr>
<tr>
<td>4</td>
<td>22%</td>
<td>&gt;99 : 1</td>
<td>92% (S)</td>
</tr>
<tr>
<td>8</td>
<td>24%</td>
<td>&gt;99 : 1</td>
<td>90% (S)</td>
</tr>
<tr>
<td>22</td>
<td>26%</td>
<td>&gt;99 : 1</td>
<td>84% (S)</td>
</tr>
<tr>
<td>48</td>
<td>28%</td>
<td>&gt;99 : 1</td>
<td>76% (S)</td>
</tr>
</tbody>
</table>

Table 14 || *Time course study of the amination of 3-methylcinnamate to the 3-methyl-β-phenylalanine by the EncP-R299K variant.*

<table>
<thead>
<tr>
<th>Time / h</th>
<th>Conv.</th>
<th>β : α-ratio</th>
<th>βee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6%</td>
<td>&gt;99 : 1</td>
<td>&gt;90% (S)</td>
</tr>
<tr>
<td>2</td>
<td>12%</td>
<td>&gt;99 : 1</td>
<td>90% (S)</td>
</tr>
<tr>
<td>4</td>
<td>16%</td>
<td>&gt;99 : 1</td>
<td>90% (S)</td>
</tr>
<tr>
<td>8</td>
<td>18%</td>
<td>&gt;99 : 1</td>
<td>90% (S)</td>
</tr>
<tr>
<td>22</td>
<td>20%</td>
<td>&gt;99 : 1</td>
<td>90% (S)</td>
</tr>
<tr>
<td>48</td>
<td>21%</td>
<td>&gt;99 : 1</td>
<td>90% (S)</td>
</tr>
</tbody>
</table>

Table 15 || *Time course study of the amination of 4-methylcinnamate to the 4-methyl-β-phenylalanine by the EncP-R299K variant.*

<table>
<thead>
<tr>
<th>Time / h</th>
<th>Conv.</th>
<th>β : α-ratio</th>
<th>βee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9%</td>
<td>&gt;99 : 1</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>11%</td>
<td>&gt;99 : 1</td>
<td>87% (S)</td>
</tr>
<tr>
<td>4</td>
<td>18%</td>
<td>&gt;99 : 1</td>
<td>n.d.</td>
</tr>
<tr>
<td>8</td>
<td>21%</td>
<td>&gt;99 : 1</td>
<td>n.d.</td>
</tr>
<tr>
<td>22</td>
<td>21%</td>
<td>&gt;99 : 1</td>
<td>86% (S)</td>
</tr>
<tr>
<td>48</td>
<td>21%</td>
<td>&gt;99 : 1</td>
<td>87% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined
3.3 Conclusions

The suitability of the bacterial ammonia lyase EncP as a potentially useful biocatalyst has been investigated in this chapter. A whole cell biocatalytic method was developed and optimised, which allowed the broad substrate scope of the enzyme to be uncovered. As with related aminomutases, the amination reaction was found to give a mixture of regioisomeric products in accordance with electronic effects specific to each substrate. Using core electron binding energy shift (ΔCEBE) values as a chemical descriptor, deviations from and patterns within the correlation between regioselectivity and substrate electronics were found. These prompted sequence- and structure-guided engineering efforts to shift the often undesirable regiopreference of the wild-type enzyme, with a view to creating a selective biocatalyst. Three new enzymes were designed, two with enhanced α- and one with enhanced β-selectivity. These variants also displayed broad substrate acceptance, along with altered regioselectivity and generally good enantioselectivity, with any differences being rationalised via substrate binding and / or electronic effects. The two substitutions E293Q and E293M represent the first report of a substantial shift towards α-selective amination engineered into an enzyme with phenylalanine aminomutase activity. The R299K substitution is the only example of a class I lyase-like enzyme capable of selective (S)-β-amination of various acrylic acids to yield high value unnatural amino acids. This makes it the concluding addition to the PAL / PAM toolbox of biocatalysts, as it allows synthesis of the previously missing optical isomer of phenylalanine derivatives from common starting materials.
4. Exploring the Potential of AvPAL as an Industrial Scale Biocatalyst
4.1 Background

4.1.1 Use of PAL for Preparative Scale Synthesis

Despite the reported prospects of PAL-catalysed reactions in chemical synthesis, there exist only a handful of examples of intensified and larger scale amination reactions mediated through these biocatalysts. A widely cited example of this is the development of an industrial scale method by life science and materials company Royal DSM, to synthesise a pharmaceutical precursor. In this case the starting point of the reaction is either 2-chloro- or 2-bromocinnamic acid, which can be both regio- and enantioselectively aminated to yield the corresponding (S)-α-amino acid. The product can then be cyclised to give (S)-2-indolinecarboxylic acid which, through further chemistry, can be modified and incorporated into the angiotensin converting enzyme (ACE) inhibitors indolapril and perindopril. The catalyst for the amination step is reported to be an undisclosed ammonia lyase gene expressed heterologously in *E. coli* and introduced to the reaction as whole cell slurry (130 g in a 1 L reaction vessel). The reaction is incubated at 30°C with 20 x batch additions of the substrate in order to give a final concentration of 100 mM whilst minimising substrate inhibition. An ammonia source of 13% ammonium hydroxide pH adjusted to 11 with sulphuric acid is used to drive the reaction and increase the solubility of the substrate. Overall the reaction is reported to give 18.1 g 2-chloro-(S)-α-phenylalanine (91% conversion) within an 8.5 hour reaction time, equating to a theoretical space time yield of around 51.1 g L⁻¹ d⁻¹.52

![Chemical Reaction](image)

**Figure 39** | Use of PAL in the synthesis of perindopril (R = Me) and indolapril (R = Ph) pharmaceuticals from inexpensive 2-chloro- or 2-bromocinnamic acids.
A second, more recent example is the use of the bacterial ammonia lyase AvPAL (from *Anabaena variabilis*) to enable gram-scale synthesis of 2-chloro- and 4-trifluoromethyl-(S)-α-phenylalanine from their counterpart cinnamic acid derivatives (Figure 40). This reaction is also mediated through whole cell *E. coli* containing the recombinant PAL protein (7.5 g in a 500 mL reaction) but this time, batch addition is not required due to the lower concentration of substrate used (2 g L\(^{-1}\)). Again a 30°C incubation is used for a comparable reaction period (8 hours) but in this instance the ammonia source is 2.5 M ammonium carbonate (pH 9.1 unadjusted) to allow for ease of purification via thermal decomposition and evaporation of the reaction buffer. The two biotransformations gave 80 and 89% conversion (2-chloro- and 4-trifluoromethyl- respectively). Even with full conversion these reaction parameters would result in the much lower theoretical space time yield of 6 g L\(^{-1}\) d\(^{-1}\).

**Figure 40** \| *Use of AvPAL for gram-scale conversion of 2-chloro- or 4-trifluoromethyl-cinnamic acid to unnatural amino acid derivatives.*

These documented uses of ammonia lyases in larger scale amination reactions demonstrate the potential of these enzymes but also point to certain drawbacks. The DSM method gives industrially-applicable space time yields but relies on batch substrate additions and suffers from difficult reaction work-up.\(^{52}\) Although product isolation is easier for the method reported with AvPAL, due to the volatility of the reaction buffer, incomplete consumption of starting material still necessitates removal of leftover cinnamate and the lower intensity of the reactions requires a large volume for much less product.\(^{104}\) In an effort to build on both studies in collaboration with Johnson Matthey Catalysis and Chiral Technologies, a new method was sought to use AvPAL to give more intensified reactions with improved production rates.
4.1.2 AvPAL from the Cyanobacterium *Anabaena variabilis*

The bacterial ammonia lyase AvPAL from *Anabaena variabilis* was first reported - along with a homologue from the related cyanobacterium *Nostoc punctiforme* - in reference to unpublished data in 2006. The full discovery, biochemical and structural characterisation of both were reported in 2008, revealing structural and probable mechanistic relationships to plant PALs and bacterial HALs. Substrate interactions with the active site of AvPAL were also predicted in the study via the modelling of (S)-α-phenylalanine based on the structure of a bacterial TAL-(S)-α-tyrosine complex already reported. This revealed a putative bidentate salt-bridge between R317 and the carboxyl of the modelled substrate as well as a possible ‘edge on’ interaction between the aryl rings of the substrate and residue F107. Later crystallographic studies of an AvPAL double variant protein (C503S / C565S) gave a structure of the enzyme with the active site loop in its closed conformation. As of yet there remains no co-crystallised structure of a PAL enzyme with a substrate in its active site.

Although deamination of (S)-α-phenylalanine by AvPAL had been published along with the structure later studies also revealed that the enzyme could be used for amination of cinnamic acid derivatives, as was well known for related enzymes. The enzyme has been described as a promising candidate for development as a biocatalyst, due to its superior stability and substrate scope compared with other catalysts of the same class. Oddly, in the amination direction, it was found that 4-nitro- and 4-trifluoromethylcinnamic acids, along with 2- and 4-pyridylacrylic acids, gave very low enantiomeric excess values for the α-amino acid products formed. In fact, time course experiments with the 4-nitro- compound revealed that the ee dropped from >99 % (S) to racemic over the course of the reaction. Mutations of the active site residues revealed that formation of the unnatural enantiomer was more or less limited to substrates with electron deficient aromatic rings and proceeded, at least in part, in a MIO-independent fashion with absolute requirement for the catalytic active site loop tyrosine.
Figure 41 || The hypothesised fast (MIO-dependent) and slow (MIO-independent) amination pathways in AvPAL leading to (S)- and (R)-α-phenylalanine enantiomers respectively.

The pre-existing characterisation and enzymatic properties of AvPAL make it a good starting point for development of an industrial process. The enzyme has already been shown to be stable, easy to produce in high quantities and applicable across a broad substrate range with the potential for preparative scale biotransformations.\textsuperscript{104} The availability of structural data also makes it possible to rationalise substrate interactions, to perform structural comparisons with related enzyme scaffolds and may also give scope for rational design of the enzyme to improve catalytic properties.

4.2 Results and Discussion

4.2.1 AvPAL Candidate Substrates

The plasmid construct used in previous studies of AvPAL was obtained and introduced into a competent \textit{E. coli} expression strain. Whole cell biocatalyst was prepared as described before\textsuperscript{10,104,107} and used to probe the full substrate scope of the enzyme in the synthetically-useful amination direction. AvPAL had already been tested with a subset of arylacrylic acids and shown to exhibit imperfect enantiocontrol in some cases despite perfect regioselectivity.\textsuperscript{10,104,107} In order to select the best possible substrates for scale up and intensification efforts, the biotransformations performed in the preceding Chapter with EncP were repeated with AvPAL. For this, a panel of 21 substrates was used with 20 mg mL\textsuperscript{-1} wet whole cells in each 750 μL reaction. As the enzyme had been shown to display PAL activity...
without external heating\textsuperscript{88} (unlike EncP),\textsuperscript{4} a reaction temperature of 30\textdegree{}C was employed instead of the established 55\textdegree{}C incubation. All other parameters for the reactions and analyses were retained from the EncP biotransformation procedure (4M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, pH 8.3, 550rpm agitation, 22 h, analysis by HPLC on chiral and non-chiral stationary phases).

Table 16 || AvPAL-catalysed amination of a panel of ring-substituted cinnamates.

<table>
<thead>
<tr>
<th>1</th>
<th>R</th>
<th>Conv.</th>
<th>(\text{ae}e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>64%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1b</td>
<td>2-F</td>
<td>90%</td>
<td>96% (S)</td>
</tr>
<tr>
<td>1c</td>
<td>3-F</td>
<td>78%</td>
<td>96% (S)</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>72%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1e</td>
<td>3,5-F\textsubscript{2}</td>
<td>86%</td>
<td>78% (S)</td>
</tr>
<tr>
<td>1f</td>
<td>2,3,4,5,6-F\textsubscript{5}</td>
<td>&gt;99%</td>
<td>n.d.</td>
</tr>
<tr>
<td>1g</td>
<td>2-Cl</td>
<td>85%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1h</td>
<td>3-Cl</td>
<td>68%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>54%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>66%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>55%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>42%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1m</td>
<td>2-NO\textsubscript{2}</td>
<td>88%</td>
<td>52% (S)</td>
</tr>
<tr>
<td>1n</td>
<td>3-NO\textsubscript{2}</td>
<td>89%</td>
<td>83% (S)</td>
</tr>
<tr>
<td>1o</td>
<td>4-NO\textsubscript{2}</td>
<td>83%</td>
<td>49% (S)</td>
</tr>
<tr>
<td>1p</td>
<td>2-CH\textsubscript{3}</td>
<td>64%</td>
<td>90% (S)</td>
</tr>
<tr>
<td>1q</td>
<td>3-CH\textsubscript{3}</td>
<td>58%</td>
<td>94% (S)</td>
</tr>
<tr>
<td>1r</td>
<td>4-CH\textsubscript{3}</td>
<td>59%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1s</td>
<td>2-OCH\textsubscript{3}</td>
<td>58%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1t</td>
<td>3-OCH\textsubscript{3}</td>
<td>25%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1u</td>
<td>4-OCH\textsubscript{3}</td>
<td>2%</td>
<td>&gt;99% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined

The enzyme was found to give variable conversion with all of the substrates tested, from 2\% with the 4-methoxy- compound 1u to >99\% with 2,3,4,5,6-pentafluorocinnamate 1f. Of these, 10 compounds had not been reported as substrates for this enzyme in the literature (2,3,4,5,6-pentafluorocinnamate 1f, bromocinnamates 1j-1, methyl- and methoxycinnamates
The lowest conversion obtained fits with existing observations that AvPAL does not accept tyrosine well as a deamination substrate and this finding may be linked to the chemical similarities between coumarate / L-tyrosine and 4-methoxycinnamate / 4-methoxy-(S)-α-phenylalanine. As with EncP the pattern of conversions for substrate isomers with the same ring substituent tended to follow the trend: ortho - > meta - > para- (e.g. a chloro-substituent giving 85%, 68% and 54% for 1g, h and i). The nitro- and methyl- substituted compounds, however, deviated slightly from this pattern, with all homo-substituted isomers showing conversions within the same range (88%, 89% and 83% for nitrocinnamates 1m, n and o / 64%, 58% and 59% for methylcinnamates 1p, q and r).

The enantiomeric excess values for the (S)-α-amino acid products were found to be excellent (>99%) for only 12 of the substrates tested, with significant amounts of the (R)-enantiomer detected for all others. The four compounds giving products with the lowest enantiopurities in this work (difluorocinnamate 1e and nitrocinnamates 1m-o) had been shown to do so in aforementioned studies with the same enzyme. Additionally to these, ortho- and meta-, fluoro- and methyl- substituted compounds (1b-c and p-q) gave a small amount of the (R)-α-amino acid product. For the methyl- compound this is the first time that substrates with an electron-rich aromatic ring have been shown to allow formation of the unnatural enantiomer. It is possible that this is due to different positioning of these substrates in the active site, although this does not seem to be the case with most other 2- and 3-substituted arylacrylic acids tested. The findings for the fluorinated substrates also differ from literature results, where the enantioselectivity of the equivalent reactions have been shown to be perfect. This may be explained by the differences in reaction conditions - 5 mM substrate concentration, 8 hour reaction time and 5 M pH 9.6 NH₄OH reaction buffer versus 1 mM, 22 hours and 4 M pH 8.3 (NH₄)₂SO₄. From this comparison it can be inferred that the longer reaction time / lower substrate concentration (and possibly different reaction buffer) would result in a larger amount of the (R)-α-amino acid building up relative to the (S)-enantiomer. This could account for the lower ee values, assuming the change in ee over time is consistent with the aforementioned results for all (R)-forming compounds.

### 4.2.2 Side Products of AvPAL Biotransformations

Upon non-chiral HPLC analysis of the AvPAL biotransformations with the substrate panel above, it was found that 8 of the compounds gave 2 product peaks. Strikingly, although one peak corresponded to the expected α-amino acid product, the additional peak was recognisable from biotransformation analyses with EncP as the β-amino acid (confirmed by
retention time comparisons with authentic standards). $\beta$:$\alpha$ product ratios of 2:98 were observed in reactions with the 2- and 3-bromocinnamates (1j and k) and only traces of the side product were detectable with 3-methyl- / methoxy- compounds (1q and t). For the para-halogenated and methylated cinnamic acids (1d, i, l and r) more $\beta$-formation was observed. The occurrence of $\beta$-addition with AvPAL seemed to be specific to these compounds with no evident relationship to substrate conversion. In fact, some of the highest converted of the $\beta$-forming substrates, 4-fluorocinnamate 1d, seemed to give a lower ratio than 4-methylcinnamate 1r (with the least selective ratio of 18:82), but a higher ratio than the lowest converted substrate 1t (3-methoxy-).

**Table 17 || AvPAL-catalysed amination of ring-substituted cinnamates showing detectable levels of $\beta$-amino acid side product.**

<table>
<thead>
<tr>
<th>1</th>
<th>R</th>
<th>Conv.</th>
<th>$\beta$ : $\alpha$ ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1d</td>
<td>4-F</td>
<td>72%</td>
<td>4 : 96</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>54%</td>
<td>9 : 91</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>66%</td>
<td>2 : 98</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>55%</td>
<td>2 : 98</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>42%</td>
<td>9 : 91</td>
</tr>
<tr>
<td>1q</td>
<td>3-CH$_3$</td>
<td>58%</td>
<td>1 : 99</td>
</tr>
<tr>
<td>1r</td>
<td>4-CH$_3$</td>
<td>59%</td>
<td>18 : 82</td>
</tr>
<tr>
<td>1t</td>
<td>3-OCH$_3$</td>
<td>25%</td>
<td>1 : 99</td>
</tr>
</tbody>
</table>

Evidence of wavering regioselectivity in amination reactions with AvPAL is surprising as this enzyme is an ammonia lyase and has never been shown to have discernable aminomutase activity.$^{88}$ Ammonia addition at both positions has only been observed in the enzyme family before with EncP and TwPAM; enzymes known to catalyse aromatic amino acid isomerisation reactions in the absence of ammonia under certain conditions. These wild-type mutases have also both been shown to give roughly equal mixtures of both phenylalanine regioisomers in amination reactions with cinnamate (in the literature$^{14}$ and the previous Chapter), whereas PALs (AvPAL and others) are reported to give only $\alpha$-phenylalanine.$^{26,105}$ In kinetic studies, however, it has been reported that a eukaryotic PAL from *Rhodotorula graminis* was able to deaminate (S)-$\beta$-phenylalanine, albeit at a rate 800 times lower than with the natural lyase reaction.$^4$ This highlights the fact that the reverse may be possible,
even with the cinnamate as starting material. Interestingly the subset of compounds where β-amino acids were detected as side products with AvP AL was also found to give greater β:α ratios than the unsubstituted natural acrylic acid (1a) with EncP in the previous Chapter. However EncP also gave similar regioselectivity changes with 2- and 4-methoxycinnamate (1s and u) not observed here. The latter substrate presumably gives no detectable quantities of β-product in this instance due it being poorly accepted by AvP AL, as compared to its higher conversion with EncP. Oddly the 2-methoxy- compound 1s shows no detectable side product despite a moderate conversion of 58%.

The regioselectivity data obtained from biotransformations with AvP AL may point to possible interplay between substrate electronic effects and substrate positioning in the enzyme active site, as inferred with EncP. In order to ascertain the possible contributions of these, substituent- and position-specific core electron binding energy shift (ΔCEBE) values for each substrate were plotted against the percentage contribution of the β-amino acid to product resulting from each biotransformation with AvP AL. The results showed no correlation when all data points were taken into account, but a strong negative correlation when substrates affording no or only traces of β-amino acid side product were excluded. This implies that the substrate electronic effects on enzyme regioselectivity are overridden for the majority of compounds converted by the regiocontrol of the catalyst, but do contribute, at least in part, for a small section of the substrate panel (Figure 42 left vs. right).
4.2.3 Time Course AvPAL Amination Experiments

To test the relative rates of regioisomer production for the reaction, the four substrates giving the highest $\beta$:$\alpha$ product ratios were retested, along with the unsubstituted substrate $1a$ as a control experiment. For these reactions the catalyst loading was increased to 50 mg mL$^{-1}$ to enhance the rate of product formation and 7 identical 750 $\mu$L reactions set up per substrate. To take time points, a single reaction was removed, worked up and the conversion / $\beta$:$\alpha$ ratio determined by HPLC. Time points were taken at 45 minutes, 2 hours 15 minutes, 5 hours 15 minutes, 22 hours, 48 hours, 72 hours and 144 hours. The results for all are shown in Figure 43.

The biocatalyst was found to give maximal conversion of all five substrates (80-90%) within the first day of the reaction. Although conversion remained more or less constant after this, the levels of each amino acid were altered significantly throughout the time of the experiment. Biotransformations stopped after the first few time points revealed that most of the starting material had been converted to the $\alpha$-regioisomer in a short space of time, with the four $\text{para}$-substituted substrates also showing small quantities of $\beta$-amino acid. With these compounds the level of $\alpha$-product was observed to be lower and lower with each later time point (5 hours 15 minutes and after) with levels of $\beta$-product gradually increasing with reaction time. With three of the compounds the percentage composition of the $\beta$-amino acid actually overtook that of the $\alpha$-amino acid - after 144 hours for reactions with 4-chloro- $1i$ and after 48 hours for 4-fluoro- $1d$ and 4-methylcinnamate $1r$. These compounds gave high $\beta$:$\alpha$ product ratios after 144 hours of 62:38, 54:46 and 68:32 respectively with the 4-bromocinnamate giving a slightly lower 48:52. Even with the unsubstituted cinnamate, which had previously yielded only the single expected regioisomer, $\beta$-product could be detected, first in the 48-hour, and then increased amounts in the 72- and 144-hour biotransformations (with a final $\beta$:$\alpha$ ratio of 13:87 for the last time point). This production of first one regioisomer followed by the other is very different to time course experimental results TwPAM, where the $\beta$:$\alpha$ ratio has been shown to remain more or less constant, even with increasing conversion over time.$^5$ It is possibly indicative of two distinct types of substrate positioning in the active site of AvPAL: one major binding mode for $\alpha$-addition and another for $\beta$-addition which is more disfavoured by the active site.
Figure 43 // The percentage composition of AvPAL-catalysed biotransformations with β-forming substrates and cinnamate over the course of 144h.
The production of predominantly β-amino acid by the end of the time course experiment may be due to this regioisomer being in some way favoured thermodynamically, possibly by higher solubility in the ammonium sulphate buffer or by some manner of product stabilising interaction with one or more components of the reaction mixture. It is also possible that the rate of deamination of the β-product is so slow compared to its amination (and compared to the cognate rates for the α-regioisomer) that it becomes trapped and builds up over time as more and more β-promoting binding modes are sampled.

To test the enantioselectivity of the AvPAL amination reaction at both positions, the 5 biotransformations were run again for 4 days and the results analysed by HPLC on both chiral and non-chiral stationary phases. Conversions and product ratios were found to be comparable to data from earlier experiments. Interestingly after the longer reaction time there were detectable amounts of (R)-α-amino acid with all five substrates, giving ee values of 80-92% (S). This, combined with some lower ee values with methyl- and fluoro-substrates in the initial amination experiments, implies that the formation of the ‘unnatural’ enantiomer may be possible with all substrates, if the reaction proceeds for long enough. The (R)-selective amination at the α-position seemed to be faster with substrates with an electron-deficient ring, but did not occur exclusively with them. This may explain why these findings have only been reported with such compounds under the shorter reaction times (and with higher substrate concentrations) in preceding work.10,107

Table 18 || AvPAL-catalysed deamination of ring-substituted (S)-α-amino acids and amination of the corresponding cinnamate derivatives.

<table>
<thead>
<tr>
<th>3</th>
<th>R</th>
<th>Conv.</th>
<th>Conv.</th>
<th>β : α ratio</th>
<th>βee</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>H</td>
<td>&gt;99%</td>
<td>73%</td>
<td>25 : 75</td>
<td>14% (R)</td>
<td>92% (S)</td>
</tr>
<tr>
<td>3d</td>
<td>4-F</td>
<td>&gt;99%</td>
<td>83%</td>
<td>66 : 34</td>
<td>16% (R)</td>
<td>91% (S)</td>
</tr>
<tr>
<td>3i</td>
<td>4-Cl</td>
<td>&gt;99%</td>
<td>84%</td>
<td>69 : 31</td>
<td>8% (R)</td>
<td>85% (S)</td>
</tr>
<tr>
<td>3l</td>
<td>4-Br</td>
<td>&gt;99%</td>
<td>81%</td>
<td>67 : 33</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>3r</td>
<td>4-CH₃</td>
<td>&gt;99%</td>
<td>83%</td>
<td>72 : 28</td>
<td>44% (R)</td>
<td>80% (S)</td>
</tr>
</tbody>
</table>

n.d. = not determined
Enantiomeric excess values for four of the five β-products (2a, d, i and r) could be obtained with the HPLC method used. These were found to be poor but, interestingly, each showed an excess of the (R)-configured amino acid with values ranging from almost racemic (2i: 8% (R)) to moderately enriched (2r: 44% (R)). This suggests that the ammonia lyase active site may impart selectivity at the β- (as well as the α-) positions in a mode similar to the (R)-selective aminomutase TwPAM. To confirm this catalytic similarity, five further biotransformations were set up using the same reaction conditions as before but with 1 mM of the (S)-α-amino acids (3a, d, i, l and r) instead of the counterpart acrylic acids, and 0.1 M borate buffer (pH 8.3) to discourage reamination. Unsurprisingly AvPAL was found to deaminate all 5 amino acids fully after a 22 hour incubation. The ability of an ammonia lyase to allow both deamination of (S)-α-phenylalanine and derivatives, and β-amination with slight (R)-selectivity under similar conditions, points to the evolutionary origins of (R)-selective PAM activity.

4.2.4 Similarities between AvPAL and (R)-selective PAMs

For an enzyme to exhibit phenylalanine aminomutase activity it must be able to mediate three processes. The first of these is the deamination of phenylalanine to give cinnamate and ammonia. Secondly the enzyme must be able to retain these species as intermediates in the active site before the third step - ammonia addition to the β-carbon of the substrate. Aminomutases are presumed to be able to do this courtesy of the shape of their active sites, which position the substrate and intermediate such as to allow amination / deamination by catalytic residues at both positions. These enzymes have also been reported to have inner active site loops with properties that slow the turnover of the enzyme such as to allow time for full isomerization, whilst minimising loss of the cinnamate and ammonia intermediates. A similar result can be achieved with the Taxus mutase at lower temperatures by mutating the active site loop to make it less constrained or more PAL-like. After the first demonstration of the third trait (addition of ammonia at the β-position) in a PAL, it is clear that ammonia lyases may already possess two of the characteristics of PAMs. However the ability of PALs to catalyse mutase-like amination can only be observed under high ammonia concentration from the cinnamic acid, as the active site loops of characterised PALs do not seem to allow retention of amino acid deamination products for subsequent reamination. So while the activity exists in wild-type PALs, it is unlikely to happen under natural conditions and thus has probably not been
tailored by natural selection. This could explain why the selectivity of the amination reaction for the PAL studied here is much poorer than true aminomutases at the β-position.

Figure 44 | The MIO-tyrosine-catalysed enzymatic processes required for PAL (1), (S)-PAM (2) and (R)-PAM (3) activities starting from (S)-α-phenylalanine. All processes involve deamination of the starting material via formation of an amino-MIO adduct and tyrosine-mediated proton abstraction. Release of ammonia from the MIO results in PAL activity. (S)-PAM activity is characterised instead by amino-MIO amination at the β-carbon and tyrosine protonation at the α-carbon on the same face as the preceding deamination and deprotonation steps. Alternatively 180° rotation of the cinnamate intermediate in the active site followed by amino-MIO amination at the β-carbon and tyrosine protonation at the α-carbon on the opposite face gives (R)-PAM activity.

It is stated in the literature that the active sites of PAL and PAM are nearly identical, in reference to comparisons between two (R)-selective aminomutases from Taxus spp.
(TcPAM and TwPAM) and another plant enzyme (PcPAL).\textsuperscript{5} It is also attested that the \((S)\)-selective bacterial PAM AdmH has a very different active site to its enantiocomplementary family members.\textsuperscript{13} The mode of \((R)\)-\(\beta\)-phenylalanine formation in eukaryotic PAMs has been postulated to involve rotation of the cinnamate in the active site such as to present the opposite faces of the \(\text{C}\alpha\text{-C}\beta\) double bond to the protonating tyrosine and MIO-\(\text{NH}_2\) adduct respectively.\textsuperscript{55,106,110} In fact computational studies of this process have further theorised that this rotation is most likely to happen only around the \(\text{C}1\text{-C}\alpha\) bond in such a way as to only allow addition on the \(re\) face at the \(\beta\)-position but no longer at the \(\alpha\)-position in this flipped state (even though deamination at the \(\alpha\)-position is tolerated in the unflipped state).\textsuperscript{74} As AvPAL has been shown to be more closely related to the eukaryotic enzymes than the microbial mutase, it is likely that the mechanism of selectivity of ammonia addition with this enzyme closely resembles that of its more immediate relatives. To assess the possible contributors to differences in stereo- and regioselectivity between PAL and PAM, the crystal structures of a cinnamate-TcPAM complex and the AvPAL C503S/C565S double variant were visualised and overlaid. Manual inspection of active site residues was limited to those within 6 Ångströms of the TcPAM-bound cinnamate ligand, with TcPAM residue C107 also added (despite lying further out) to overlay with the homologous residue F107 already included in the AvPAL residue selection. To assess the significance of any variations, sections of the sequence alignment of characterised class I lyase-like enzymes from Chapter 2 were used to compare the same inner shell active site residues across a range of more distantly related PAL enzymes. The TAL from \textit{Rhodobacter capsulatus} was also included in the sequence analysis as this was the enzyme previously used as a template to create a model of AvPAL with a substrate docked in the active site.\textsuperscript{88}
**Figure 45**  
(top) Overlaid active site residues from the crystal structures of TcPAM (PDB: 3NZ4) and AvPAL (PDB: 3CZO).  
(middle) Protein sequence alignment sections of a selection of distantly related Class I Lyase-like Enzymes. The sequences are: TcPAM from Taxus canadensis, PcPAL1 from Petroselinum crispum, RtPAL (bifunctional PAL / TAL) from Rhodospirillum toruloïdes, AvPAL from Anabaena variabilis, StlA (PAL) from Photorhabdus luminescens and RcTAL from Rhodobacter capsulatus. Consensus sites, homologous to those in the active site structure are also shown (highlighted in yellow). The percentage sequence identity with respect to the TcPAM sequence is shown in the final column of the alignment table for each entry (Seq. ID).  
(bottom) A cladogram showing the hypothesised evolutionary history of the aligned enzymes as inferred from sequence similarity. Possible amino acid substitutions in the active site residues are used to label the branches where they are most likely to have occurred.
Of all the residues in this first shell (around the cinnamate) of TcPAM and AvPAL, only two were found to be different in the overlapped active sites of the enzymes, with the other 16 amino acids found to be indistinguishable in both identity and rotameric position. The deviations were A88 and C107 in TcPAM versus G86 and F107 in AvPAL. The former amino acid position could be seen from the structural analyses to be located on the inner active site loop, 8 residues C-terminal to the catalytic tyrosine. The sequence alignment revealed that this position was an alanine in both plant enzymes (PcPAL1 and TcPAM), a glycine in the remaining PAL enzymes (RtPAL, AvPAL and StIA) and a proline in RcTAL. This pattern of the residue being identical in the enzyme with the highest identity yet different catalytic activity to TcPAM and being different between plant and fungal / bacterial PALs (and different again in the most distantly related enzyme - RcTAL), means that the difference is more likely to be of phylogenetic, rather than function-discriminating significance. The other difference (C107 / F107) was more promising as it fell within the selectivity residues discussed in Chapter 2 - positions shown to display variance and correlate with enzyme catalytic activity. This position was found to be different not only between the PAM and PALs (cysteine versus phenylalanine) but also between these and the enzymes known to have TAL activity (histidine in RtPAL and RcTAL). In the model based on an RcTAL-tyrosine co-crystal produced by Moffitt et al. AvPAL was hypothesised to form an ‘edge on’ interaction between the aryl ring of the substrate and residue F107.88 Visualisation of molecular and Van der Waals surfaces these two features in the TcPAM / AvPAL overlay gave similar molecular contacts. It is therefore possible that this edge on interaction, not present between C107 and cinnamate in the TcPAM structure, may constrain any rotation of the substrate in the active site and thus disfavour formation of the β-amino acid (in particular the (R)-form). A similar steric effect may occur between the aromatic ring of the histidine residue in enzymes with TAL activity deterring any mutase-like binding modes.

4.2.5 Mutational Analysis of AvPAL Regioselectivity

The occurrence of enantioselective β-amino acid formation by AvPAL presented an opportunity to engineer a biocatalyst for the formation of (R)-β-aryllalanines from cinnamate derivatives. This has been previously carried out by Janssen et al. with TwPAM as a starting template. While this enzyme already had excellent enantioselectivity, it formed a regioisomeric mixture of (R)-β- and (S)-α-phenylalanine derivatives with most substrates tested and suffered very low turnover rates for all. These hindrances were overcome by directed evolution of the active site and extensive rational design of the inner active site loop. Saturation libraries across three pairs of active site residues (approximately 1200 variants)
yielded a handful of hits with more desirable β-selectivity - the best of which being TwPAM-Q319M.\textsuperscript{106} After initial computational modelling of the complex dynamics of loop opening in TwPAM, mutations were introduced to remove any molecular constraints on this process, of which one, TwPAM-R92S, was effective.\textsuperscript{5} In a subsequent study additional dynamics simulations, combined with loop hydrophobicity studies and comparison to a related ammonia lyase, were used to create more variants with increased amination rates.

Due to the complexity of studying protein dynamics with a view to enzyme redesign, it is desirable to have a starting point where inner active site loop properties are already suited to amination reactions. As AvPAL already had good turnover rates, courtesy of its lyase-like catalytic cycle at ambient temperature, work was initiated on the enhancement of the potentially useful side activities of the enzyme discovered in this Chapter. To begin with, two amino acid substitutions were selected for introduction into the enzyme. The first was F107C, chosen to make the active site of AvPAL more similar to that of TcPAM and TwPAM (with the only difference being the active site loop glycine / alanine variation previously dismissed by sequence analyses). It was hoped that this change would improve the enantio- and initial regioselectivity of the enzyme to make it more like the wild-type TwPAM. The second variant to be made was AvPAL-R317K, as this is the homologous residue to that mutated in Chapter 3 in EncP (R299K). This variant was shown to have increased β-selectivity and so it was predicted that the same mutation in AvPAL might have a comparable effect.
Table 19 || Amination of cinnamate and a selection of 2- and 4-ring-substituted derivatives catalysed by two AvPAL single active site variants (F107C and R317K).

<table>
<thead>
<tr>
<th></th>
<th>AvPAL-F107C</th>
<th>AvPAL-R317K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>&lt;1 : 99</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1j</td>
<td>&lt;1 : 99</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1g</td>
<td>&lt;1 : 99</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1b</td>
<td>&lt;1 : 99</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1p</td>
<td>&lt;1 : 99</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1l</td>
<td>2 : 98</td>
<td>2 : 98</td>
</tr>
<tr>
<td>1i</td>
<td>2 : 98</td>
<td>2 : 98</td>
</tr>
<tr>
<td>1d</td>
<td>2 : 98</td>
<td>2 : 98</td>
</tr>
<tr>
<td>1r</td>
<td>2 : 98</td>
<td>2 : 98</td>
</tr>
</tbody>
</table>

The mutants were created and variant proteins produced for use in whole-cell biocatalytic conversions of various arylacrylic acids as before. In this case the classically β-forming substrates (1l, i, d and r) were tested against their ortho-substituted arene isomers (1j, g, b and p) which had shown no β-production with the wild-type enzyme. Cinnamate (1a) was also included as a control substrate. Reaction conditions were kept the same as before with 50 mg mL⁻¹ to ensure conversion was seen after unoptimised protein production conditions. AvPAL-F107C showed conversion with all substrates tested, but the regioselectivity was not found to be affected as predicted. Within the 22-hour reaction time no β-amino acids were seen for biotransformations with cinnamate, the 2-substituted derivatives or 4-fluorocinnamate. The R317K variant showed low conversions with most substrates, however, no products were detectable in reactions with the 2-methyl compound 1p. Nonetheless β-amination was observed with all four para-substituted substrates as well as with ortho-fluorocinnamate (β:α ratio = 15:85); a substrate with which only α-amination had been observed with the wild-type enzyme. The regioselectivity for the 4-methylcinnamate reaction was particularly interesting as only β-amino acid could be detected as product. However the low conversion of 4% meant that the β:α ratio could not be assigned to greater than 95:5. Unfortunately a combination of poor conversions and / or low regioselectivity meant that determination of ee values for the β-products formed in this suite of reactions was not possible. The R317K variant was found to be temperamental to handle as a whole cell biocatalyst as freeze-thawing the cells or storing them at 4°C for subsequent
biotransformations resulted in a complete loss of enzyme activity. However a re-growth of
the protein production culture led to similar activity as the first cell batch but with the same
storage difficulties. This could be akin to the drop in temperature optimum and substrate
 conversions seen when comparing the EncP-R299K and EncP-WT whole cell biocatalysts in
the previous Chapter.

In the interests of ascertaining possible synergistic effects of site-directed mutagenesis the
AvPAL double variant F107C / R317K was also made and tested with the same panel of
compounds as the single variants. The enzyme was found to have no activity with any of the
compounds under the same reaction conditions. The presence of the protein in the whole
cells was confirmed by observation of an overproduced band of correct molecular weight in
an SDS-PAGE analysis following cell lysis.

**Table 20 || Amination of cinnamate and a selection of 4-ring-substituted derivatives
catalysed by an AvPAL single active site variant (Q311M).**

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Conv.</th>
<th>β : α ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>60%</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>70%</td>
<td>6 : 94</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>75%</td>
<td>9 : 91</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>80%</td>
<td>10 : 90</td>
</tr>
<tr>
<td>1r</td>
<td>4-CH₃</td>
<td>33%</td>
<td>50 : 50</td>
</tr>
</tbody>
</table>

Inspired by the directed evolution efforts detailed with TwPAM, an additional AvPAL variant
was created. AvPAL-Q311M was designed in the image of the TwPAM-Q319M variant,
which had been shown to give much improved β-regioselectivity with cinnamate and a small
group of derivatives.⁴⁶ This was attempted due to the high similarity of the active sites of
AvPAL and TcPAM when compared to that of AdmH (the basis for the R299K substitution in
EncP). It was predicted that this active site alteration might have a similar effect on the
regioselectivity of AvPAL to the R317K mutation, but without the stability / activity issues
inferred by transferring mutations between these more distantly related family members.
Initial tests with AvPAL-Q311M were performed as for the time course experiments and with
the same substrates for 22 hours. Oddly regioselectivity was scarcely found to be better than
initial substrate screening studies with the wild-type enzyme. None of the compounds gave any improvement in regioselectivity over the best \( \beta:\alpha \) ratios observed in other biotransformations in this Chapter. Again the combination of conversion and \( \beta:\alpha \) ratio for each substrate contravened determination of \( \beta_{ee} \) values.

The Q311M substitution was, however found to allow multiple freeze-thaw cycles of the whole cell biocatalyst whilst retaining activity. In the same vein as with the previous double variant, both Q311M and F107C were introduced into the final PAL protein via plasmid mutagenesis. This created an enzyme with an identical active site to TwPAM-Q319M (minus the discounted A88/G86 variation). Testing of this final variant with the 5 substrates from before gave moderate conversions of 4 of them, but again no improvements in \( \beta \)-selectivity (the best substrate being 4-fluorocinnamate 1d, with 52\% conversion and a 17:83 \( \beta:\alpha \) product ratio). Oddly neither \( \beta \)- nor \( \alpha \)-amino acid product could be detected after the 22 hour reaction time with the 4-methylated compound 1r.

**Table 21** || Amination of cinnamate and a selection of 4-ring-substituted derivatives catalysed by an AvPAL double active site variant (F017C/Q311M).

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Conv.</th>
<th>( \beta : \alpha ) ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>H</td>
<td>43%</td>
<td>&lt;1 : 99</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>52%</td>
<td>17 : 83</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>64%</td>
<td>4 : 96</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>76%</td>
<td>8 : 92</td>
</tr>
<tr>
<td>1r</td>
<td>4-CH(_3)</td>
<td>&lt;1%</td>
<td>- : -</td>
</tr>
</tbody>
</table>

The discrepancies in behaviour of AvPAL, EncP, TwPAM and their many variants compound existing evidence that the transition from PAL to PAM activity is poorly understood. Even enzymes with nearly identical active sites can have vastly different activities and endeavours to make them more similar or transfer homologous amino acid substitutions do not seem to result in similar activities. So whilst it seems clearer that some PALs may possess enantioselective mutase-like side activities as starting points from which new PAM function could evolve, the mechanisms by which this may occur still require more study. The molecular determinants of these related functions must lie elsewhere in the enzyme.
structures, possibly within outer active site shells to influence binding pockets with altered dynamics such as to promote / disfavour appropriate substrate binding modes. With more work into this area and / or more in-depth engineering efforts, AvPAL could well be developed into a biocatalyst for the asymmetric synthesis of β-amino acids. But based on current work it is more feasible to focus on its application in α-amino acid production.

4.2.6 Intensification of AvPAL Reactions

4.2.6.1 Optimising Reaction Conditions

From the collection of substrates tested with AvPAL, candidates forming products of industrial interest could be selected for intensification. A target biocatalytic manufacture was developed in partnership with collaborators at Johnson Matthey Catalysts and Chiral Technologies, who were interested in incorporating PAL into their portfolio of enzyme catalysts. As such, an outline of desirable bioprocess parameters was jointly agreed upon to ensure the process would be efficient, viable and easy to insert into the infrastructure available within an industrial setting. These were:

- Catalyst usable in lyophilised form
- Robust biocatalyst formulation
- Substrate concentration 50-100 mM
- High conversion
- Theoretical space time yield of around 20 g l⁻¹ d⁻¹
- Catalyst:substrate w / w ratio of 1:1 or less
- Volatile reaction buffer
- Excellent enantiomeric excess of final product
- Perfect regioselectivity of amination

Use of catalyst in lyophilised form was decided upon due to ease and consistency of measuring quantities to be added to biotransformations. Use of dry whole cells was recommended, at least for initial studies, due to the simplicity of catalyst removal and possible reuse. This formulation was also hoped to allow storage of the biocatalyst after its production. Target substrate concentrations of 50-100 mM were initially set to allow productive synthesis in combination with high conversions, relative to the scale of the reaction. These parameters would be predicted to allow an industrially-relevant theoretical space time yield of around 20 g L⁻¹ d⁻¹. An aim was also set to use less than a 1:1 ratio of substrate weight to catalyst weight, although this target was more flexible as the additional
weight of *E. coli* whole cells (rather than just the isolated protein catalyst) compared to small acrylic acid substrates has been shown to give high ratios in previous instances. In the interests of pure products the regio- and enantioselectivity of the reactions (as well as the conversions) were specified to be high enough such as to mitigate the need for lengthy and / or complex separation methods. To this end a volatile reaction buffer was deemed appropriate, so as to allow removal upon heating, leaving only product (and possible side products and starting material). The industrial collaborators were particularly interested in developing a route to fluoro phenylalanines due to their pre-existing work with structurally-related fluoroamphetamine fine chemicals.

In light of the factors set by the industrial supervisors it was decided first to try lyophilisation of the whole cell biocatalyst used in the earlier sections of this Chapter. The reaction buffer was replaced with a 13% w / v ammonia solution (as used in a documented industrial process) pH adjusted by bubbling CO₂ to create ammonium carbonate, similar to that reported in an earlier scale up attempt with AvPAL. The advantage of pH adjusting ammonia to create a volatile reaction buffer rather than dissolving the solid carbonate salt is that a higher starting concentration of ammonia can be used (13%) and the pH can be more easily adjusted to test the effect of this parameter on the biotransformation. Of the fluorinated cinnamates shown to work with AvPAL, the difluoro-compound 1e was discarded due to the lower αee seen with the product of this reaction. The three monofluorinated isomers 1b-d were all selected as the enzyme showed good to excellent enantioselectivity with these compounds. Even though the para-fluorocinnamate 1d was shown to form quantities of β-amino acid product before, it was hoped that under intensified reaction conditions and in a different reaction buffer, this side activity could be minimised.

### Table 22 || AvPAL-catalysed amination of fluorocinnamates at 5 mM concentration.

<table>
<thead>
<tr>
<th></th>
<th>Conv.</th>
<th>αee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>98%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1c</td>
<td>94%</td>
<td>&gt;99% (S)</td>
</tr>
<tr>
<td>1d</td>
<td>86%</td>
<td>&gt;99% (S)</td>
</tr>
</tbody>
</table>
To prepare the dry catalyst, *E. coli* BL21(DE3) whole cells harbouring wild-type AvPAL were flash frozen, lyophilised and ground into a fine powder. As an initial test 750 μL reactions were set up using 5 mg mL⁻¹ lyophilised whole cell powder and 5 mM of substrate in 13% ammonium hydroxide, pH adjusted to 10.2 with CO₂. The reactions were incubated at 30°C for 22 hours with 250 rpm agitation and subsequently analysed as before. All three substrates were found to give good to excellent conversion, following the same trend of ortho- > meta- > para- as before, and resulted in enantiopure (S)-α-amino acid. Although a small amount of β-side product was detectable after the biotransformation with 1d, it accounted for around 0.6% of the total product and 0.5% of the overall composition of the reaction. This was a much lower ratio than seen with 1 mM substrate concentration and ammonium sulphate (pH 8.3), despite higher conversion. This may be due to effects of the different pH or the presence / absence of sulphate, borate and carbonate ions in the reaction mixture which may act to favour or disfavour dynamically the amino acid regioisomers relative to each other.

![Figure 46](image.png)

**Figure 46** // The effect of temperature and time of incubation on overall conversion of AvPAL-catalysed amination of 5 mM ortho-fluorocinnamate.

To test the effect of temperature on conversion, the reaction with the highest-converted substrate 1b was repeated at 25, 37, 55, and 70°C. This time, 5 separate reactions were set up at each temperature to allow time points to be taken at 50 min, 2, 4, 8 and 22 hours. All incubations gave lower conversion than the original 30°C biotransformation after 22 hours, with a temperature of 37°C resulting in 91% conversion, 55°C in 76%, 25°C in 59% and 70°C in just 40%. From a plot of the time course experiments for each temperature it could be seen that the higher temperatures give faster initial rates but conversions plateau more
suddenly resulting in the 70°C reactions all reaching the same conversion, irrespective of incubation time, and the 55°C samples showing no increase in conversion between the 8 and 22 hour reaction times. This points to a balancing of increased rate of conversion with decreased biocatalyst longevity at higher temperatures. From these data it can be seen that improvement in conversion is unlikely to result from increasing or decreasing the incubation temperature of AvPAL biotransformations from the preliminary 30°C set point.

In the interest of intensifying the reactions before attempting anything at larger scale, substrate concentration analyses were performed. First the maximum solubility of each of the three substrates in the concentrated ammonia buffer was tested via stepwise addition of each to a set volume of liquid until precipitate remained upon mixing at 30°C. This resulted in stable solutions of ~200 mM 2-fluoro-, ~150 mM 3-fluoro- and ~100 mM 4-fluorocinnamate in the pH 10.2 13% ammonium hydroxide to be used in subsequent biotransformations. Substrate loading was tested by diluting each of these stock solutions with neat reaction buffer 1:40, 1:20, 1:10 and 1:5. These were then set off as 22 hour biotransformations at 750 μL scale, as before.

![Figure 47](image)

**Figure 47** // The effect of substrate loading on overall conversion of AvPAL-catalysed amination of fluorocinnamates.

Analysis of the biotransformations revealed that, whilst high conversions could be achieved at low concentrations, increased substrate loading impacted negatively on overall conversion. The relationship between substrate concentration and conversion for each isomer of fluorocinnamate was not found to be linear i.e. a doubling of substrate
concentration did not result in a halving of conversion (except in one single case of 4-fluorocinnamate increasing from 10 to 20 mM). This means that the same amount of catalyst under the same reaction conditions did not convert the same amount of substrate in the same amount of time. In fact the 1:40 / 1:20 / 1:10 / 1:5 dilutions gave consumption of 4.9 / 8.3 / 10.3 / 12.2 mM of 1b, 3.5 / 7.1 / 11.5 / 15.3 mM of 1c and 2.2 / 3.7 / 5.4 / 5.4 mM of 1d respectively. This implies that the overall conversion is not only limited by the amount of catalyst but possibly also by equilibrium or substrate / product inhibition. It was also evident that, despite 1b being the highest converted compound at 5 mM substrate concentration, 1c came to outperform both other substrates at concentrations above 7.5 mM.

As a means of increasing overall conversion and substrate concentration, the reaction time and amount of catalyst were varied using the undiluted stock solutions of each substrate in the reaction buffer. Initial reactions were performed for 22, 48 and 78 hours as before but with 100 mM 1d, 150 mM 1c and 200 mM 1b. An additional 5 mg mL⁻¹ catalyst was then added and incubated for 0, 22, and 48 hours. The batch addition was then repeated with incubation times of 26 and 50 hours, before another addition with an incubation of 55 hours. Finally for the 3-fluorocinnamate biotransformation a fifth and final batch addition was performed with incubation for 22, 48 and 60 hours.

![Figure 48](image-url)  
*Figure 48* // The effect of batch addition of 5 mg mL⁻¹ biocatalyst on overall conversion of AvPAL-catalysed amination of fluorocinnamates. Each catalyst addition event is indicated by a black arrow.
As expected using 5 mg mL\(^{-1}\) catalyst with 100+ mM substrate resulted in very poor conversions. In spite of the increase in conversion over the time points for the first catalyst batch, overall transformation of starting material remained poor (below 12% for all substrates). As before, 1c gave the highest conversion at each time point (around twice the conversion of the others), despite this substrate being more concentrated for this reaction than that with 1d. With the second batch of catalyst before incubation the conversions were calculated to be within 1% of the last time point for batch 1. After 22 and 48 hour incubations increases were seen to 31% for 1d and 12-13% for 1b and d. The third batch of catalyst started to give reasonable conversions for 3-fluorocinnamate (57 and 62% after 26 and 50 hour incubations) with increased but still low values for the other two substrates (19-23%).

By addition 4, a 55 hour incubation gave 88% conversion for the meta- compound but only 24% for the ortho- and 35% for the para-. The final addition, giving a total catalyst loading of 25 mg mL\(^{-1}\), gave 91-93% conversion for the 3-fluorocinnamate which could not be increased further, with little difference between the 22, 48 and 60 hour incubation periods.

Whilst batch addition of catalyst gave very good conversion of 3-fluorocinnamate 1c, the entire experiment took over two weeks to carry out. In the interests of condensing the reaction time, the catalyst loading was increased from 5 to 20 and 40 mg mL\(^{-1}\) and experiments repeated for various time points with all other conditions remaining constant. Even with the increase of catalyst loading to near and above the final batch addition in the preceding experiment, conversions of above 90% for the most promising substrate were not observed. There were, however improvements in conversion for the 2- and 4-fluorocinnamates although these remained modest (still <50%).

![Figure 49: The effect of varying biocatalyst loading on overall conversion of AvPAL-catalysed amination of fluorocinnamates.](image)
To assess the extent to which the reaction was equilibrium-limited two additional biotransformations were performed, this time using 3- and 4-fluoro-(S)-α-phenylalanine as starting materials (150 and 100 mM respectively). All parameters were kept the same as before and 20 mg mL\(^{-1}\) dry cell biocatalyst was used. The samples were incubated for 22, 30 and 48 hours and showed conversions of 8-10% for the 3-fluoro- (3c) and 12-13% for the 4-fluoro-compound (3d). The catalyst was then removed via centrifugation and the supernatant added to 20 mg mL\(^{-1}\) fresh catalyst with incubation for a further 22 hours. This process was then repeated once more to give a total of two catalyst replacements. At the end of the third biotransformation (using the same reaction mix) conversion was 9% for 3-fluorophenylalanine and 13% for 4-fluorophenylalanine. These findings highlight that even after 3 batches of biocatalyst the reaction goes no further than the equilibrium position, which for the synthetic reaction would be the equivalent of 90-92% conversion for 1c (as seen previously with the 5 x batch addition) and 87-88% for 1d in the same reaction buffer.

Having shown that conversions of 80-90% should be possible in theory, at least with the meta- and para-fluorinated compounds, the pH of the reaction buffer was varied by adjustment with different amounts of CO\(_2\). It was found that pH adjustment could be performed as far as 9.5, with further addition of CO\(_2\) causing white precipitate to form in the buffer, presumably due to the lower solubility limit of ammonium carbonate in water as compared to ammonia gas. Using increments of 0.5 from pH 9.5 to 12.0, six biotransformations were performed using 150 mM 3-fluorocinnamate and 20 mg mL\(^{-1}\) lyophilised whole cell biocatalyst, with all other parameters kept the same as before.

**Table 23** // The effect of varying pH on overall conversion of AvPAL-catalysed amination of 3-fluorocinnamate.

<table>
<thead>
<tr>
<th>pH</th>
<th>9.5</th>
<th>10.0</th>
<th>10.5</th>
<th>11.0</th>
<th>11.5</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv. (8 h)</td>
<td>76%</td>
<td>36%</td>
<td>6%</td>
<td>5%</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>Conv. (22 h)</td>
<td>93%</td>
<td>68%</td>
<td>14%</td>
<td>7%</td>
<td>5%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Samples taken after 8 and 22 hours revealed that conversion was better in lower pH reaction buffer with an ‘equilibrium point’ conversion of 93% observed after 22 hours at pH 9.5.
Interestingly the 8 hour sample taken from the lowest pH reaction buffer showed higher conversion than the 22 hour sample taken from the pH 10.0 reaction buffer. Higher pH buffers (10.5-12.0) all gave poor results, which may be due to the stability of the biocatalyst as indicated by the decreasing difference in conversion between 8 and 22 hour samples with increasing pH. An additional biotransformation at pH 9.5 with the 4-fluorocinnamate was found to give a similarly high result, comparable to the ‘equilibrium point’ conversion inferred before for this substrate.

4.2.6.2 Altering Reaction Buffer

On discussion of intensification efforts with collaborators at Johnson Matthey, it was deemed that the production of a volatile ammonia buffer using CO\(_2\) to reach the correct pH was not a scalable process. In particular the amount of gas required to reach the optimal pH range (<10) was considered to be too much for an industrial setting. As such, alternative means were sought. Ammonium carbonate as a salt to be dissolved in water was not considered feasible for investigations in this study, due to the low solubility of the solid when compared with the current reaction buffer. Another option was ammonium carbamate - a salt that is inexpensive, very soluble in water and also readily decomposes into ammonia and carbon dioxide upon heating. As an initial test, 4 M ammonium carbamate solution was prepared and found to form a solution of pH 9.9. This buffer was utilised in lieu of the pH adjusted 13% ammonia for biotransformations with 3-fluorocinnamate at 150 mM substrate concentration. Assessment of the conversion after 24 hours was found to result in a conversion of 98%. This higher conversion implied that a change in the ammonium donor salt altered the position of the equilibrium. To test the extent to which the equilibrium position had been shifted, the biotransformation was repeated and more performed with increments of 50 mM concentration of 3-fluorocinnamate, up to 350 mM. Above around 200 mM the substrate was found not to be completely soluble in the reaction buffer. However, analysis of the soluble and insoluble fractions of the 22 hour biotransformations revealed that the majority of the substrate had been converted to product and most of this was in the soluble fraction. This suggested that the higher solubility of the product than the starting material at pH 9.9 may act to drive the reaction, even when the cinnamate derivative is not completely soluble to begin with.
Table 24 // Effect of varying substrate concentration on intensified, small scale aminations of 3-fluorocinnamate by AvPAL

<table>
<thead>
<tr>
<th>[Substrate]</th>
<th>Conv.</th>
<th>αee</th>
<th>Theoretical Space-time Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 mM</td>
<td>&gt;99%</td>
<td>&gt;99% (S)</td>
<td>~9 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>100 mM</td>
<td>99%</td>
<td>&gt;99% (S)</td>
<td>~18 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>150 mM</td>
<td>98%</td>
<td>99% (S)</td>
<td>~26 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>200 mM</td>
<td>98%</td>
<td>99% (S)</td>
<td>~34 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>250 mM</td>
<td>97%</td>
<td>98% (S)</td>
<td>~43 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>300 mM</td>
<td>97%</td>
<td>98% (S)</td>
<td>~51 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>350 mM</td>
<td>81%</td>
<td>98% (S)</td>
<td>~50 g L⁻¹ d⁻¹</td>
</tr>
</tbody>
</table>

The productivity of the intensified, small scale reactions was found to be relatively good, with substrate concentrations of above 100 mM giving hypothetical production rates of more than the desired 20 g L⁻¹ d⁻¹. To prove the scalability of this system the highest substrate concentration tested was used to design a gram-scale synthesis of 3-fluoro-(S)-α-phenylalanine 3c from the acrylic acid precursor 1c. The method required only the simple addition of 0.34 g dried whole cells, 1 g of the starting material and 5.3 g of ammonium carbamate salt to 17 mL of dH₂O. This was performed in a round-bottom 50 mL flask with magnetic stirring at 200 rpm and temperature controlled at 30°C. After a 24 hour reaction time a small sample was taken and prepared for HPLC analyses. Remarkably the reaction showed a considerable increase in conversion compared to the small scale 22 hour reaction (93% versus 81% conversion) and a comparable enantiomeric excess of the final product. 93% conversion of the 1 g of starting material was calculated to equate to 1.025 g of product within 24 hours which would give a theoretical yield of 60.3 g if scaled to a total volume of 1 L.
Figure 50 // $^{13}$C (top) and $^1$H (bottom) NMR spectra of the isolated 3-fluorophenylalanine product of the AvPAL-catalysed amination reaction (supporting data in chapter 5, section 5.3).
As a test of the generality of this biocatalytic method, developed with the most promising substrate 1c, biotransformations were performed with all monohalogenated arylacrylic acids (1b-d and g-l). In the interests of mitigating awkward purification steps for the removal of the starting material, reaction conditions giving at least 95% conversion were sought. It was envisaged that this would allow simple separation of the catalyst through centrifugation followed by removal of reaction buffer through heating to yield a crude product with only a small percentage of starting material. A summary of the conditions giving the desired outcomes for each substrate is presented in table 25. It was found that, for the most soluble (fluoro-) cinnamates, concentrations of 100 mM could be used with lower concentrations of 50 and 25 mM being more suitable for the less soluble chloro- and bromocinnamates respectively. All reactions were found to give high conversions after just 24 hours (80-99%), however in the interests of minimising starting material as an impurity, suboptimal conversions (for 1d, h, j and l) were improved via an extra day of reaction time. This was done for all four para-halogenated substrates, but also for meta-chlorocinnamate which gave 93% conversion after the first 24 hour sample. For the 3-fluoro- compound a quantitative isolated yield was obtained using the isolation method above and NMRs run (Figure 50).

Table 25 // Reaction parameters giving at least 95% conversion of monosubstituted halocinnamates in AvPAL-catalysed biotransformations.

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Time</th>
<th>[Substrate]</th>
<th>Conv.</th>
<th>αee</th>
<th>Theoretical Space-time Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>2-F</td>
<td>1 d</td>
<td>100 mM</td>
<td>&gt;99%</td>
<td>&gt;99%  (S)</td>
<td>18.3 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1c</td>
<td>3-F</td>
<td>1 d</td>
<td>100 mM</td>
<td>99%</td>
<td>&gt;99%  (S)</td>
<td>18.3 g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1d</td>
<td>4-F</td>
<td>2 d</td>
<td>100 mM</td>
<td>96%</td>
<td>99%  (S)</td>
<td>9.2  g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1g</td>
<td>2-Cl</td>
<td>1 d</td>
<td>50 mM</td>
<td>&gt;99%</td>
<td>&gt;99%  (S)</td>
<td>10.0  g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1h</td>
<td>3-Cl</td>
<td>2 d</td>
<td>50 mM</td>
<td>96%</td>
<td>98%  (S)</td>
<td>5.0  g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1i</td>
<td>4-Cl</td>
<td>2 d</td>
<td>50 mM</td>
<td>95%</td>
<td>98%  (S)</td>
<td>5.0  g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1j</td>
<td>2-Br</td>
<td>1 d</td>
<td>25 mM</td>
<td>98%</td>
<td>99%  (S)</td>
<td>6.3  g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1k</td>
<td>3-Br</td>
<td>1 d</td>
<td>25 mM</td>
<td>96%</td>
<td>99%  (S)</td>
<td>6.3  g L⁻¹ d⁻¹</td>
</tr>
<tr>
<td>1l</td>
<td>4-Br</td>
<td>2 d</td>
<td>25 mM</td>
<td>95%</td>
<td>98%  (S)</td>
<td>3.1  g L⁻¹ d⁻¹</td>
</tr>
</tbody>
</table>
4.3 Conclusions

In this Chapter the cyanobacterial ammonia lyase AvPAL has been used to catalyse intensified and industrially-relevant aminations of arylacrylic acids to yield α-amino acids with high conversion and optical purity. Initial studies revealed that the enzyme accepted a broad selection of 20 ring-substituted cinnamate derivatives. Conversion and enantioselectivity of the reaction was found to be good in most cases with a few lower values. Interestingly the enzyme was also shown to allow formation of small quantities of corresponding β-amino acids for a subset of compounds - an observation never before reported with a strict ammonia lyase. Differences were found in the relative rates of β- and α-amination for these compounds and comparisons with related (R)-selective aminomutase enzymes revealed striking structural similarities. These findings imply that mutase-like side reactions enabled by PAL enzymes could be reminiscent of ancestral enzymes from which (R)-PAMs evolved via enhancement of this minor activity under selection pressures. Attempts to use structure, sequence and literature-guided mutagenesis with a view to creating an (R)-β-selective catalyst gave little or no improvement. With input from industry the major (S)-α-selective amination activity of the enzyme was exploited to give intensified production of various halophenylalanine derivatives in high purity and with excellent enantiomeric excess values. Scale up of the most promising substrate, 3-fluorocinnamate, resulted in gram scale synthesis of the counterpart amino acid, with high conversion to the optically-pure (S)-isomer and an significant theoretical space time yield of >60 g L⁻¹ d⁻¹.
5. Materials and Methods
5.1 Computational Methods

5.1.1 Web Resources


Clustal W2: Multiple Sequence Alignment - http://www.ebi.ac.uk/Tools/msa/clustalw2/


StarORF: open reading frame identification application - www.star.mit.edu/orf/

DictyBase: Resource for Dictyostelid Genomics - www.dictybase.org


UniProt: Universal Protein resource database - www.uniprot.org

ENA: European Nucleotide Archive - http://www.ebi.ac.uk/ena

PDB: RCSB Protein Databank - www.rcsb.org

5.1.2 Molecular Modelling and Visualisation Software

PyMOL version 0.99 from DeLano Scientific

YASARA (Yet Another Scientific Artificial Reality Application) Structure version 14.7.17

5.1.3 Biological Data

All information regarding gene and protein names, species of origin, annotated functions, amino acid sequences and database cross references were obtained from the Universal Protein resource database (UniProt) as provided by the European Bioinformatics Institute (EBI), Protein Information Resource (PIR) and Swiss Institute of Bioinformatics (SIB). All protein sequence data was downloaded as provided and constructed as text files in FASTA format for further analyses.
>AdmH
MSIVNESGSQPVRDSLETSLQIERTSHISGSDLSEDLAIIAMADHQPVTLLHEDEVNRYT
RSRSILESVMSTERVITGNYNTMGSFVNYIVPIIRASAELLQNNMLINAYAVNGKTYEDDD
VRATMLARIVSLSRGGNAISIVNFRLIFEIINOQGIVCPFIERGSLGISSDLGLALALILV
CTGQMRKARYQEGÖSNGGAMALERAGISPMEDSFREGLALINGTSAMYSLGLYLLYDEVVRLF
DTTLTVTSLSIGLHGRPPFEDADVHRMKPHGQLEVAATTITWETLDASLGVNEHEVEXKL
LAEEMGLYKASNRQ1EDAYSIRCFTQILGFVADTLKNIKQVTLNENSSDNPLIDQTT
EEVFHHGFHGQYVSMAMCHLNTALVMMNLDNRRIDRFDKSNNSNLFGFLCAENAGLR
LGLGSGSFSMTASTAESAERASCMPMSIQSLSTTGDFQIDVSFGLVAARRVREQLK3NLKVF
SFELLCAEOAVDIRGTAGLSKRTRALYDKTRTLVFYLEEDKTISDYESTIAQVTILKNSD

Figure 51 || An example of the FASTA format for a text file containing a sequence identifier (in this case the protein name ‘AdmH’) line-separated from the sequence data.

All protein structures were available through the Protein Data Bank as provided by the Research Collaboratory for Structural Bioinformatics (RCSB PDB). Structural data for each individual biomolecule was downloaded in PDB text file for use in subsequent visualisation and modelling experiments.

5.1.4 Genome Mining

Discovery of uncharacterised potential enzyme sequences was undertaken using an appropriate query (i.e. a known enzyme of interest) to perform a sequence similarity search within the UniProt public database. In each case the basic local alignment search tool (BLAST) was used to find regions of sequence similarity between an amino acid sequence and the in silico translated DNA sequences of all available genomes and metagenomes (tBLASTn). All searches were within the UniProt Knowledgebase (UniProtKB), gapped and unfiltered with a statistical significance value threshold of E=10. The choice of protein substitution matrix was set to automatic and thus assigned computationally based on the length of sequence.

5.1.5 Sequence Alignments and Phylogenies

All sequence alignments were performed through use of the W2 command line interface for the Clustal multiple sequence alignment computer programme, as available online. All alignments made use of the Gonnet protein weight matrix with the ‘gap open’ penalty score set to 10 in all cases. The initial pairwise alignment type was set to slow with a gap
extension score of 0.1. With subsequent multiple alignments the gap extension score was set to 0.2 in addition to a gap distance penalisation value of 5, without end gap penalisation or iteration. Sequences were clustered via the neighbour-joining method.

Sequence phylogenies were constructed from the automatically-generated phylogenetic trees from the Clustal W2 programme. These were inspected visually for congruence with protein sequence identity matrices and represented as cladograms from which evolutionary relationships were drawn.

5.1.6 Genomic Context Analysis

The sequence context of any genes of interest within their host genomes was determined by cross-referencing UniProt entries with appropriate genomic databases. The *in silico* sequence annotation data of the surrounding DNA was then inspected to find predicted genes within the region. Putative gene functions were assigned for open reading frames (ORFs) based on the annotations already present in the corresponding database. These protein coding regions were also downloaded and their possible functions confirmed and / or expanded upon through use of a conserved domain search, as made available through the National Center for Biotechnology Information (NCBI). All searches were performed specifically against the conserved domain database (CDD v3.14 - 47363 PSSMs) with a composition based statistics adjustment and an expected value threshold of E = 0.01.

5.1.7 Structural Analyses

Structures were visualised for inspection using the YASARA Structure version 14.7.17 molecular modelling software. In each case the PDB file to be studied was opened and all features hidden initially. Amino acid, cofactor and ligand of interest were visualised in ‘stick’ format and appropriate residues hidden / shown to create images for Figures. Correct interface contacts were predicted - when not in the original PDB files - using the ‘oligomerise’ function to visualise crystal contact data. Structures were overlaid for comparison using the ‘superpose’ option with additional manual adjustments.

Clashes between residues modelled in YASARA in all possible side chain-dependent rotamers were represented visually using the PyMOL molecular visualisation data of modified PDB files.
5.1.8 Homology Modelling

Homology models were created using the experiment option in the YASARA Structure version 14.7.17 using the 'slow' modelling method. For this the amino acid sequence of the query enzyme was input in FASTA format along with the unmodified PDB file for the appropriate structure. The number of PSI-BLAST iterations to build a position-specific scoring matrix from related sequences was set to 3 with an E-value cut off of 0.5. Sequence profiles were obtained from YASARA's PSSM (Profiles and Sequence- and Structurally-related Proteins). The maximum number of top scoring templates was set to 5, with a maximum of 5 alignments per template and a maximum oligomerisation state for templates of 4. The number of sampling experiments to optimise loop structures was set to 50 with a terminal extension value of 10. Side-chain rotamers were selected using electrostatic and knowledge-based packing interactions as well as solvation effects with implicit solvent. Hydrogen bonding networks were optimised to include pH-dependence and ligands bound via hydrogen prediction and visualisation. A high-resolution energy minimization with a shell of explicit solvent molecules was run using YASARA version-specific knowledge-based force field parameters.

5.2 Experimental Methods

5.2.1 Materials

DNA Nucleotides:
- avpal-pET16b gene-vector construct containing the codon optimised DNA sequence encoding AvPAL - ampicillin selection marker (as used in previous studies)\textsuperscript{10,107}
- encp-pET28a gene-vector construct containing the codon optimised DNA sequence encoding EncP - kanamycin selection marker (as used in a previous study)\textsuperscript{4}
- Various oligonucleotide primers from Eurofins MWG Operon
- Deoxyribonucleotide triphosphate solutions (dNTPs, N = guanine, cytosine, thymine or adenine) from NEB

\textit{E. coli} expression and cloning strains:
- BL21(DE3) from New England Biolabs (NEB) - for gene expression
- DH5\textalpha{} from NEB - for non-methylated DNA production
- XL1 Blue from Agilent - for methylated DNA production

**Enzyme Substrates:** Cinnamic acid, racemic and (S)-configured α- and β-phenylalanine and all derivatives thereof and other miscellaneous chemicals were used as purchased from Sigma Aldrich, Peptech Corp. or Alfa Aesar.

**Cloned and modified enzymes** from New England Biolabs (NEB):
- NdeI - restriction enzyme originally from *Neisseria denitrificans*
- XhoI - restriction enzyme originally from *Xanthomonas holcicola*
- BamHI - restriction enzyme originally from *Bacillus amyloliquefaciens H*
- DpnI - restriction enzyme originally from *Streptomyces pneumoniae*
- Phusion® - high fidelity DNA polymerase and processivity domain fused protein
- T4 Ligase - double-stranded DNA / RNA ligase originally from T4 bacteriophage
- T4 PNK - polynucleotide kinase originally from T4 bacteriophage

**Enzyme Reaction Buffers and Supplements** from New England Biolabs (NEB):
- NEBuffer 4 - general reaction buffer
- T4 Ligase Buffer - supplemented with ATP
- Phusion® HF Buffer - all-purpose high-fidelity buffer
- Phusion® GC Buffer - buffer for replication of GC-rich DNA
- DMSO - dimethyl sulphoxide

QIAprep Spin Miniprep Kit from Qiagen

SYBR® safe DNA stain from Life Technologies

Purple DNA loading dye from NEB

Parafilm M® from Bemis, NA

Thomson 0.45 μm PVDF filter vial

Absolute Ethanol from Sigma Aldrich

HPLC Grade Methanol from Sigma Aldrich

Laemmli buffer from NEB

Pre-stained protein marker from NEB

InstantBlue Coomassie stain from Expedeon

0.45 μm and 0.20 μm PTFE syringe filters from Sigma Aldrich
Hitrap Chelating column (5 mL) from GE Healthcare Bio-Science.

Antibiotics: kanamycin sulphate and ampicillin from Sigma Aldrich

Disposable Containers: Eppendorf 0.2, 1.5 and 2 mL tubes, Falcon 15 mL and 50 mL tubes, sterile petri dishes (15 x 150 mm).

Microbiological Media: all in dH₂O with autoclave sterilisation at 121°C for 20 minutes

SOC outgrowth media from Sigma Aldrich

- Tryptone (20 g L⁻¹)
- Yeast extract (5 g L⁻¹)
- Glucose (3.6 g L⁻¹)
- MgSO₄ (4.8 g L⁻¹)
- NaCl (0.5 g L⁻¹)
- KCl (0.2 g L⁻¹)

Lysogeny Broth (LB medium)

- Tryptone (10 g L⁻¹)
- Yeast extract (5 g L⁻¹)
- NaCl (10 g L⁻¹)

Lysogeny Broth (LB) agar

- Tryptone (10 g L⁻¹)
- Yeast extract (5 g L⁻¹)
- NaCl (10 g L⁻¹)
- Agar (15 g L⁻¹)

Auto-induction Medium (AIM) - LB Broth Base including trace elements from Formedium™

- Tryptone (10 g L⁻¹)
- Yeast extract (5 g L⁻¹)
- (NH₄)₂SO₄ (3.3 g L⁻¹)
- KH₂PO₄ (6.8 g L⁻¹)
- Na₂HPO₄ (7.1 g L⁻¹)
- Glucose (0.5 g L⁻¹)
- α-Lactose (2.0 g L⁻¹)
- MgSO₄ (0.15 g L⁻¹)
- Trace Elements (0.03 g L⁻¹)
General Buffers and Solutions: in dH₂O unless otherwise specified.

0.1M Borate Buffer

- Boric Acid (6.18 g L⁻¹), pH adjustment with KOH (8, 8.3, 10, 12)

1M Ammonium Acetate, pH 8.6

- Ammonium Acetate (77.1 g L⁻¹ in 0.1M borate buffer, pH 10)

1M Ammonium Chloride, pH 8.4

- Ammonium Chloride (53.5 g L⁻¹ in 0.1M borate buffer, pH 10)

1M Ammonium Nitrate, pH 8.5

- Ammonium Nitrate (80.0 g L⁻¹ in 0.1M borate buffer, pH 10)

1M Ammonium Sulphate, pH 8.6

- Ammonium Sulphate (132.0 g L⁻¹ in 0.1M borate buffer, pH 10)

Ammonium Sulphate

- Ammonium Sulphate (0.5M: 66 g L⁻¹, 1.5M: 198 g L⁻¹, 2M: 264 g L⁻¹, 3M: 396 g L⁻¹), pH adjustment with KOH (7, 8, 9, 10)

4M Ammonium Sulphate, pH 8.3

- Ammonium Sulphate (529 g L⁻¹ in 0.1M borate buffer, pH 12)

5M Ammonium Hydroxide, pH 9.6

13% Ammonium Hydroxide

- 37:63 mixture of 35% ammonium hydroxide and dH₂O, pH adjusted with dry ice (9.5, 10, 10.5, 11, 12)

Tris-Acetate-EDTA (TAE) Running Buffer

- Tris base (4.7 g L⁻¹)
- Glacial acetic acid (1.14 % w / w)
- EDTA (2.92 g L⁻¹)

Purification Buffer A

- NaCl (17.5 g L⁻¹)
- Imidazole (1.36 g L\(^{-1}\))
- KH\(_2\)PO\(_4\) (0.72 g L\(^{-1}\))
- K\(_2\)HPO\(_4\) (16.5 g L\(^{-1}\))

Purification Buffer B

- NaCl (17.5 g L\(^{-1}\))
- Imidazole (68 g L\(^{-1}\))
- KH\(_2\)PO\(_4\) (0.72 g L\(^{-1}\))
- K\(_2\)HPO\(_4\) (16.5 g L\(^{-1}\))

Sodium Dodecyl Sulphate (SDS) Gel Running Buffer

- Tris base (3 g L\(^{-1}\))
- Glycine (1.44 g L\(^{-1}\))
- SDS (1 g L\(^{-1}\))

5.2.2 Equipment

General biological laboratory equipment: Nanodrop spectrophotometer, microbiological static and shaking incubators, electrophoresis gel tanks (protein and DNA) and power pack, Eppendorf benchtop centrifuge, Eppendorf mastercycler gradient, Beckmann Coulter floor-standing centrifuge and JA-10.10 / 25.50 fix angle rotors, Beckmann Coulter 500 mL and 50 mL centrifuge tubes, peristaltic pump, 2L baffled Erlenmeyer flasks, standard laboratory lyophiliser, Soniprep 150 sonicator (MSE UK Ltd.), Safe Imager™ 2.0 Blue Light Transilluminator from Life Technologies, 50mL round-bottom flask, clamp and clamp stand, stirrer hot plate with DrySyn Hot heating block, magentic stirrer bar.

5.2.3 Site Directed Mutagenesis

Mutations were introduced into methylated plasmid constructs with the desired enzyme coding gene used as a template. Introduction of missense changes in the DNA was implemented courtesy of short forward and reverse non-overlapping primers (shown in the Figures below) mixed together in distilled water to create a primer mix specific to each mutagenesis reaction. Each oligonucleotide pair mix (10 pmol µL\(^{-1}\) stock) was used prime a whole plasmid polymerase chain reaction to yield a blunt-ended PCR product containing the desired nucleotide substitution (set up as shown in the tables below).
R299K

...TAGCCTGGCAATTGAAGATGCATATAGCATTCGTTGTACACCGCAGAT...  
5’-AGCATTACGTGTACACCG-3’
3’-CCGTAAACTTCTACGTATA-5’

E293Q

...GGGCACCGTTGCAAAAGCAGGTAGCCTTGCAATTGAAGATGCATATAGCATTCG...  
5’-GGCAATTACCGATGCATATAGC-3’
3’-CAACGTTTTCTCGCATCCGGA-5’

E293M

...GGGCACCGTTGCAAAAGCAGGTAGCCTTGCAATTGAAGATGCATATAGCATTCG...  
5’-GGCAATTACCGATGCATATAGC-3’
3’-CAACGTTTTCTCGCATCCGGA-5’

**Figure 52** || The oligonucleotide primer sets used to introduce the three active site substitutions into the coding sequence for EncP as aligned with the section of plasmid DNA specific to each. The codon for each specific amino acid position to be mutated is highlighted in yellow with the base changes in the primers coloured in red.
F107C

...GAACAGGCAAGCGAACTGCAGACCAATCTGGTTTGGCTTGCTGAAACCAGGGTCAGTAAAT...

5′-CTGGTTTGGCTGAAACCAGGGT-3′

3′-CCGGTCTGACGTCTGGTTA-5′

R317K

ATTATCGTGATCATGAACTGATCCAGGATCGTTATAGCCTGCGTGTCTGCCCGCAGTATCTGGGTCCG

5′-TTATAGCCTGAACTGATCTGCC-3′

3′-GTACTTGACTAGGTCTAGC-5′

Q311M

GGTAAACATGATTATCGTGATCATGAACTGATCAGGATCGTTATAGCCTGCGTGTCTGCCCGCAGTATCTGGGTCCG

5′-CTGATCCTGATCTGCCCGCAGTATCTGGGTCCG-3′

3′-GTACTAATAGCACTAGTACTT-5′

**Figure 53** || The oligonucleotide primer sets used to introduce the three active site substitutions into the coding sequence for AvPAL as aligned with the section of plasmid DNA specific to each. The codon for each specific amino acid position to be mutated is highlighted in yellow with the base changes in the primers coloured in red.

**Table 26** || PCR reactions set up to a total reaction volume of 50 μL for mutagenesis of EncP and AvPAL encoding genes in a 0.2 mL Eppendorf tube.
Table 27 || Standard PCR reaction conditions for all mutagenesis reactions as programmed into the Eppendorf Mastercycler Gradient PCR machine.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Number of Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>98°C</td>
<td>30s</td>
<td>1</td>
</tr>
<tr>
<td>98°C</td>
<td>15s</td>
<td></td>
</tr>
<tr>
<td>58°C</td>
<td>30s</td>
<td>30</td>
</tr>
<tr>
<td>98°C</td>
<td>150s</td>
<td></td>
</tr>
<tr>
<td>72°C</td>
<td>300s</td>
<td>1</td>
</tr>
<tr>
<td>4°C</td>
<td>∞</td>
<td></td>
</tr>
</tbody>
</table>

Any fragments resulting from reactions 1-4 were assayed by taking a 5 μL sample, mixing 5:1 with NEB purple gel loading dye for separation on a 1% agarose gel supplemented with SYBR® safe DNA stain in TAE running buffer. After electrophoresis at 90V for 40 minutes the gel was imaged on a blue light transilluminator. In each case the reaction yielding in the cleanest gel image (with the major product evident as a single fragment at 7-7.5 kbp) was subjected to a two-step phosphorylation-ligation to yield circular plasmid. For this an extra 5 μL sample was transferred to a new 0.2 mL Eppendorf tube along with 1 μL T4 PNK, and 1:10 dilutions of NEBuffer 4 and T4 ligase buffer in a total reaction volume of 19 μL. The mix was placed in a static incubator at 37°C for 1h before addition of 1 μL T4 ligase and a further 2h incubation at ambient temperature. The ligated plasmid was transformed directly into chemically competent E. coli lab strains. The remaining reaction mixes were stored at 4°C until needed. From the resulting colonies, 2 or 3 were used to grow small scale bacterial cultures from which the plasmid DNA was extracted and verified by restriction enzyme analysis and DNA sequencing. In cases where only wild-type sequences were found the original reaction mix was subjected to digestion with DpnI (1 μL) at the same time as the phosphorylation and the process repeated as before.

5.2.4 Bacterial Transformation

Plasmid constructs were introduced into E. coli BL21(DE3), DH5α or XL1 Blue chemically competent cells as required for heterologous gene expression, unmethylated or methylated plasmid production respectively. The heat shock transformation protocols from the respective manufacturers were modified to allow introduction of DNA to any of the strains. In each case an aliquot of competent cell was allowed to thaw on ice for 30 minutes. A 20 μL sample was transferred to a previously autoclaved 1.5 mL Eppendorf tube using sterile
pipette tips along with 1 μL of the plasmid to be transformed (30-100ng μL⁻¹ DNA concentration). The mixture was left on ice for 30 minutes before a 30 second heat shock step at 42°C. The mixture was then transferred immediately back to incubation on ice for 2 minutes before the addition of 250 μL SOC outgrowth media (super optimal broth supplemented with 20 mM glucose from NEB). The outgrowth culture was placed in a microbiological incubator (37°C, 250 rpm agitation) before spreading on agar plates (20mL LB agar in 150 x 15 mm petri dishes) supplemented with an appropriate antibiotic (50 mg L⁻¹ ampicillin or 25 mg L⁻¹ kanamycin). The plates were incubated for 16h at 37°C in a static incubator before being store at 4°C until needed (up to 4 weeks).

5.2.5 Small Scale Bacterial Cultures

A single colony from a pre-prepared bacterial plate was picked using a sterile loop to inoculate 5 mL LB medium supplemented with the appropriate antibiotic (50 mg L⁻¹ ampicillin or 25 mg L⁻¹ kanamycin) in a 50 mL Falcon tube. The bacteria were grown in a microbiological incubator for 16h at 37°C with 250 rpm agitation. The resulting cell suspension was stored at 4°C for up to one week until needed.

5.2.6 Plasmid DNA Extraction

Circular plasmid DNA was isolated from bacterial cell cultures using the QIAprep Spin Miniprep Kit with a modified protocol. 2 mL of a small scale bacterial culture was transferred to a 2 mL Eppendorf tube and centrifuged at 13000 rpm for 10 minutes in an Eppendorf benchtop centrifuge. The cell pellet was resuspended in 300 μL pre-chilled Qiagen miniprep buffer 1 after addition of the RNase provided. 300 μL buffer 2 was then added and the mixture incubated on ice for 3 minutes. 300 μL pre-chilled buffer N3 was then added and the newly formed precipitate pelleted by centrifugation. The plasmid DNA was then bound to a Qiagen miniprep spin column by further centrifugation followed by two rounds of column washing with a 3:7 Qiagen buffer PB and ethanol mixture via 1 minute centrifugation with the flow-through being discarded after each step. After 1 minute spin drying of the column the DNA was eluted into a 1.5 mL Eppendorf tube with 50 μL distilled water (preheated to ~70°C) via a final 1 minute centrifugation. DNA concentration was ascertained using a Nanodrop spectrophotometer with measurement at 260nm against a distilled water blank measurement. With low yielding extractions the final concentration could be increased via up
to three re-elution steps from the spin column with the original eluent. Plasmid DNA was stored indefinitely at -20°C until required.

5.2.7 Restriction Enzyme Analysis

Double digestion of pET vector containing the gene of interest was performed using the restriction enzyme NdeI for 5' digestion either XhoI for pET-28a or BamHI for pET-16b 3’ digestion. Plasmid DNA (final concentration <10 ng μL⁻¹) was incubated with 10U of each enzyme and NEBuffer 4 (diluted 1 in 10) for 2-3 hours at 37°C. A 10 μL sample was taken and mixed 5:1 with NEB purple gel loading dye and the fragments separated by electrophoresis on a 1% agarose gel supplemented with SYBR® safe DNA stain in TAE running buffer. After electrophoresis at 90V for 40 minutes the gel was imaged on a blue light transilluminator and the diagnostic fragments observed.

Table 28 || Diagnostic fragment sizes for corresponding double digestions of the avpal-pET-16b and encp-pET-28a constructs.

<table>
<thead>
<tr>
<th>Plasmid Feature</th>
<th>Fragment Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>avpal gene</td>
<td>1710 bp</td>
</tr>
<tr>
<td>pET-16b</td>
<td>~5700 bp</td>
</tr>
<tr>
<td>encP gene</td>
<td>1572 bp</td>
</tr>
<tr>
<td>pET-28a</td>
<td>~5400 bp</td>
</tr>
</tbody>
</table>

5.2.8 DNA Sequencing

All DNA sequences were ascertained through the Sanger sequencing services of GATC Biotech Ltd. For this 25 μL of purified plasmid in distilled water was prepared with dilution as necessary (DNA concentration 30-100 ng L⁻¹). The coding region between the multiple cloning sites of the plasmid vectors used was targeted for sequencing using the ‘T7’ and ‘pET-RP’ oligonucleotide primers (sequences shown below). The sequence results were received as text files in FASTA format and inspected visually or via sequence alignment.
following *in silico* translation with the StarORF open reading frame identification application (as provided by the Massachusetts Institute of Technology’s Software Tools for Academics and Researchers).

Table 29 DNA sequences of the oligonucleotides required to prime sequencing of DNA subcloned into pET vectors.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7</td>
<td>5′-TAATACGACTCACTATAGGG-3′</td>
</tr>
<tr>
<td>pET-RP</td>
<td>5′-CTAGTTATTCGTCAGCGG-3′</td>
</tr>
</tbody>
</table>

5.2.9 Biocatalyst Production Cultures

Expression of genes encoding EncP, AvPAL and variants thereof was conducted according to previously reported methods. An *E. coli* BL21(DE3) starter culture harbouring the appropriate plasmid construct was added to a 2L baffled Erlenmeyer flask along with 800 mL autoinduction media with trace elements supplemented with the required antibiotic (concentration of 50 mg L⁻¹ ampicillin or 25 mg L⁻¹ kanamycin) under sterile conditions. The culture was grown in a microbiological incubator for 4 days at 18°C with 250 rpm agitation. The resulting cell suspension was divided between three Beckmann Coulter 500 mL centrifuge tubes and pelleted using a JA-10 fixed angle rotor in a floor-standing centrifuge (4000rpm, 12 min). The supernatant was then discarded and the cell mass transferred to 50 mL Falcon tubes for storage at -20°C until required.

Alternatively the Falcon tubes containing the transferred cells were flash frozen in liquid nitrogen or a dry ice / acetone mix and freeze dried using a standard laboratory lyophiliser for 16 - 24h. The dry cell mass was then ground into a fine powder and stored at -20°C until required.

5.2.10 Protein Purification

Frozen *E. coli* BL21(DE3) wet cell mass harbouring the appropriate biocatalyst was thawed at room temperature for 30 minutes before resuspension in 0.1 M borate buffer pH 8. Cells
were lysed with addition of 1 mg mL\(^{-1}\) of lysozyme powder (from hen egg white) and placed in a microbiological incubator for 45 minutes at 37°C with 250 rpm agitation. The resulting suspension was disrupted by sonication using a Soniprep 150 sonicator (10 cycles at 12 micron amplitude for 20s with 20s intervals) (MSE UK Ltd.). The cell debris was then removed by dividing between Beckmann Coulter 50 mL centrifuge tubes and pelleting in a JA-25.50 fixed angle rotor in a floor-standing centrifuge (20000rpm, 20 min). The cell-free supernatant was then stored on ice and subsequently filtered through 0.45 μm and then 0.20 μm syringe filters. This was then loaded on a GE Healthcare Bio-Science Hitrap Chelating column (5 mL) charged with 100 mM NiSO\(_4\) and pre-equilibrated with 10 column volumes of purification buffer A, using a variable-speed peristaltic pump. The column was then washed with 10 column volumes of a 20:1 mix of buffer A and B before elution with 5 column volumes of buffer B.

5.2.11 Sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE)

SDS-PAGE separation of denatured protein fragments was performed using pre-prepared 15-lane 10% polyacrylamide gels and running in SDS gel running buffer. Cell lysate, column elution and flow-through samples of 15 μL were transferred to clean 1.5 mL Eppendorf tubes and mixed with a further 15 μL of Laemmli buffer before heating (5 minutes, 100°C) in an Eppendorf Thermomixer® Comfort. The samples were then centrifuged briefly (15 seconds, 13000 rpm) in an Eppendorf benchtop centrifuge and then ran down 10% SDS-PAGE gels at 250V for 35 minutes alongside a pre-stained protein marker. Polypeptides were stained within the gel by submerging in InstantBlue Coomassie stain for 30 minutes and then destained by submerging in dH\(_2\)O for a further 30 minutes.

5.2.12 Whole Cell Biotransformations

For small scale reactions, frozen \textit{E. coli} BL21(DE3) wet cell mass harbouring the appropriate biocatalyst was thawed at room temperature for 30 minutes before transfer of the desired cell weight to a 1.5 mL Eppendorf tube. For dried cells the powdered biocatalyst was used directly as stored. This was then resuspended in 750 μL of the desired general buffer supplemented with the relevant substrate at the concentration required. Where appropriate, negative control experiments were set up to lack the whole cell catalyst, the substrate or (in the case of the amination reactions) the ammonia source. The tube lid was then sealed with Parafilm and the reaction incubated at the appropriate temperature using an Eppendorf
Thermomixer® Comfort. After the reaction time the biotransformation was mixed 1:1 with methanol and the catalyst and salts removed via centrifugation (3 min, 13000 rpm) in a benchtop centrifuge. The supernatant was prepared for analysis of the soluble fraction of the reaction through the use of a Thomson 0.45 μm PVDF filter vial. Analysis of the insoluble fraction was enabled by resuspending the pellet in pH 12 borate buffer (0.1M) and repeating the centrifugation and filtration steps as before.

Gram-scale biotransformations were performed in a 50mL round-bottom flask positioned by a clamp and stand over a stirrer hot plate with DrySyn Hot heating block and magnetic stirrer bar. Results were analysed by taking 750 μL samples and preparing as before.

5.3 Analytical Methods

Reverse phase HPLC analyses were performed on an Agilent 1200 Series system equipped with a G1379A degasser, G1312A binary pump, a G1329 autosampler unit, a G1316A temperature controlled column compartment and a G1315B diode array detector. Where appropriate, an external column cooling jacket was employed instead of the temperature controlled column compartment. HPLC grade water and methanol mobile phases were used as provided by Sigma Aldrich. The 0.35% ammonium hydroxide eluent (pH10) was prepared via a 100-fold dilution of 35% ammonium hydroxide from Sigma Aldrich (with pH adjustment with concentrated HCl) which was then vacuum filtered for use.

NMR analyses of the preparative scale amination of 3-fluorocinnamate (chapter 4, section 4.2.6.2) were performed by standard methods and instruments with D2O solvent and water suppression methods employed for proton NMR. The supporting data for the spectra in figure 50 are: $^1$H NMR (400 MHz, D2O) 7.20-7.26 (1H, m), 6.89-6.96 (3H, m), 3.38 (1H, m), 2.84-2.89 (1H, dd, J = 16 Hz, 12 Hz), 2.71-2.76 (1H, dd, J= 16 Hz, 12 Hz) $^{13}$C NMR (400 MHz, D2O) 182.17, 162.56 ($^3$JCF = 242 Hz), 140.85 ($^3$JCF = 8 Hz), 130.06 ($^3$JCF = 8 Hz), 125.19 ($^4$JCF = 2 Hz), 115.91 ($^2$JCF = 21 Hz), 113.25 ($^2$JCF = 21 Hz), 57.32, 40.49.

5.3.1 Non-chiral Analyses of Biotransformations

Conversion and product ratio analyses were ascertained via reverse-phase high performance liquid chromatography (HPLC) on a ZORBAX Extend-C18 column (50 mm × 4.6 mm × 3.5 μm Agilent) to allow baseline separation of the cinnamate, β- and α-amino acid derivatives (2, 3, and 1 respectively). In each case the mobile phase consisted of mix of
NH₄OH buffer (0.35% w/v, pH 10.0) and varying amounts of methanol to keep all run times below 10 minutes (see table below). The flow rate was set to 1 mL min⁻¹ and the temperature controlled at 40°C. Peaks were detected at a wavelength of 210 nm and assigned via comparison with commercially available standards. Retention times (shown in the table below) all followed the pattern: β-amino acid < α-amino acid < arylacrylic acid.

Conversions and product ratios were calculated from peak area integrations with use of appropriate response factors where needed. For each ring substituted derivative a specific response factor (RF) was calculated from a standard mix of the cinnamate and β- and α-amino acids to give a solution containing 1 mM of each. A serial 1 in 2 dilution was then performed to yield 5 standards which were then analysed by HPLC as described above. The peak areas for β- and α-amino acids were found invariably to be within 1% of each other and so a response factor of 1 was assumed. The peak areas of the α-amino acid and corresponding arylacrylic acid were plotted and the trend line gradient taken as RF.

Table 30 || Additional parameters of HPLC analyses of enzyme biotransformations on a non-chiral phase.

<table>
<thead>
<tr>
<th>R</th>
<th>% Methanol</th>
<th>RF</th>
<th>Retention time / min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>β-Amino Acid (2)</td>
</tr>
<tr>
<td>a</td>
<td>H</td>
<td>10</td>
<td>2.3</td>
</tr>
<tr>
<td>b</td>
<td>2-F</td>
<td>10</td>
<td>2.7</td>
</tr>
<tr>
<td>c</td>
<td>3-F</td>
<td>10</td>
<td>2.7</td>
</tr>
<tr>
<td>d</td>
<td>4-F</td>
<td>10</td>
<td>2.8</td>
</tr>
<tr>
<td>e</td>
<td>3,5-F₂</td>
<td>20</td>
<td>2.9</td>
</tr>
<tr>
<td>f</td>
<td>2,3,4,5,6-F₅</td>
<td>30</td>
<td>1.0</td>
</tr>
<tr>
<td>g</td>
<td>2-Cl</td>
<td>20</td>
<td>2.2</td>
</tr>
<tr>
<td>h</td>
<td>3-Cl</td>
<td>30</td>
<td>2.1</td>
</tr>
<tr>
<td>i</td>
<td>4-Cl</td>
<td>30</td>
<td>2.0</td>
</tr>
<tr>
<td>j</td>
<td>2-Br</td>
<td>30</td>
<td>1.4</td>
</tr>
<tr>
<td>k</td>
<td>3-Br</td>
<td>30</td>
<td>1.5</td>
</tr>
<tr>
<td>l</td>
<td>4-Br</td>
<td>30</td>
<td>1.4</td>
</tr>
<tr>
<td>m</td>
<td>2-NO₂</td>
<td>10</td>
<td>1.5</td>
</tr>
<tr>
<td>n</td>
<td>3-NO₂</td>
<td>10</td>
<td>1.4</td>
</tr>
<tr>
<td>o</td>
<td>4-NO₂</td>
<td>10</td>
<td>2.1</td>
</tr>
</tbody>
</table>
5.3.2 Chiral Analysis of Amino Acid Products

The β- and α-amino acids were separated using the non-chiral method with an injection volume of 50 μL and 150 μL fractions collected manually in HPLC vials with inserts. These were then used to ascertain separate enantiomeric excess values via reverse-phase high performance liquid chromatography (HPLC) on a CROWNPAK CR(+) column, (150 mm x 4 mm x 5 µm, Daicel) to allow baseline separation of the amino acid enantiomers (\((R)- / (S)-2\) and \((R)- / (S)-3\)). In each case the mobile phase consisted of mix of aq. HClO₄ (1.14% w / v, pH 2.0) and varying amounts of methanol (see table below). All α- and some β-amino acid enantiomers were separated according to the column manufacturer’s protocol with a flow rate of 1 mL min⁻¹ at 25°C. The remaining β- amino acid enantiomers were separated using a method developed specifically for this study using a flow rate of 0.5 mL min⁻¹ at -8°C. Peaks were detected at a wavelength of 210 nm and were assigned via comparison with commercially available standards (retention times shown in the table below). Enantiomeric excess values were calculated by dividing the difference between the \((R)-\) and \((S)\)-peak areas by the sum of the \((R)-\) and \((S)\)-peak areas.

<table>
<thead>
<tr>
<th></th>
<th>2-(\text{OCH}_3)</th>
<th>3-(\text{OCH}_3)</th>
<th>4-(\text{OCH}_3)</th>
<th>2-(\text{CH}_3)</th>
<th>3-(\text{CH}_3)</th>
<th>4-(\text{CH}_3)</th>
<th>3-(\text{CN})</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>20</td>
<td>2</td>
<td>1.7</td>
<td>2.3</td>
<td>4.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q</td>
<td>20</td>
<td>1.9</td>
<td>1.5</td>
<td>1.9</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>10</td>
<td>2.1</td>
<td>1.7</td>
<td>2.2</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>20</td>
<td>2</td>
<td>2.1</td>
<td>2.8</td>
<td>6.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>20</td>
<td>2</td>
<td>2.5</td>
<td>3.2</td>
<td>7.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>20</td>
<td>2</td>
<td>2.4</td>
<td>3.2</td>
<td>7.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>10</td>
<td>1.2</td>
<td>1.8</td>
<td>2.5</td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 31 || Additional parameters of the HPLC analyses of enzyme biotransformations on a chiral phase.

<table>
<thead>
<tr>
<th>R</th>
<th>% Methanol</th>
<th>Temp. /°C</th>
<th>Retention time / min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(R)-enantiomer</td>
</tr>
<tr>
<td>2a</td>
<td>H</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3a</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>2-F</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3b</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2c</td>
<td>3-F</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>3c</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2d</td>
<td>4-F</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3d</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2e</td>
<td>3,5-F₂</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3e</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2f</td>
<td>2,3,4,5,6-F₅</td>
<td>–</td>
<td>–8</td>
</tr>
<tr>
<td>3f</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2g</td>
<td>2-Cl</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td>3g</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2h</td>
<td>3-Cl</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3h</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2i</td>
<td>4-Cl</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3i</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2j</td>
<td>2-Br</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3j</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2k</td>
<td>3-Br</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td>3k</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2l</td>
<td>4-Br</td>
<td>14</td>
<td>–8</td>
</tr>
<tr>
<td>3l</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2m</td>
<td>2-NO₂</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>3m</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2n</td>
<td>3-NO₂</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>3n</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2o</td>
<td>4-NO₂</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3o</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2p</td>
<td>2-OCH₃</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3p</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2q</td>
<td>3-OCH₃</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3q</td>
<td>4</td>
<td>25</td>
<td>13.9</td>
</tr>
<tr>
<td>2r</td>
<td>14</td>
<td>-8</td>
<td>162.1</td>
</tr>
<tr>
<td>3r</td>
<td>4</td>
<td>25</td>
<td>14.1</td>
</tr>
<tr>
<td>2s</td>
<td>4</td>
<td>25</td>
<td>15.0</td>
</tr>
<tr>
<td>3s</td>
<td>4</td>
<td>25</td>
<td>12.1</td>
</tr>
<tr>
<td>2t</td>
<td>4</td>
<td>25</td>
<td>17.6</td>
</tr>
<tr>
<td>3t</td>
<td>4</td>
<td>25</td>
<td>15.4</td>
</tr>
<tr>
<td>2u</td>
<td>14</td>
<td>25</td>
<td>168.4</td>
</tr>
<tr>
<td>3u</td>
<td>4</td>
<td>-8</td>
<td>14.9</td>
</tr>
<tr>
<td>2v</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3v</td>
<td>4</td>
<td>25</td>
<td>5.0</td>
</tr>
</tbody>
</table>

![Graphs of (S)-2a and (R)-2a](image1.png)

![Graphs of (R)-3a and (S)-3a](image2.png)

![Graphs of (S)-2b and (R)-2b](image3.png)

![Graphs of (R)-3b and (S)-3b](image4.png)

![Graphs of (R)-2c and (S)-2c](image5.png)
(R)-3e
(S)-3e

(R)-2d
(S)-2d

(R)-3d
(S)-3d

(R)-3g
(S)-3g

(R)-2h
(S)-2h

(R)-3h
(S)-3h
Figure 54 || HPLC chromatograms showing the separation of authentic standards of amino acid enantiomers using the methods from the table above.

Figure 55 || HPLC chromatograms showing the separation of authentic standards of amino acid regioisomers and enantiomers using the low temperature method for the three sets of compounds with which this was found to work.
5.4 Protein and Nucleotide Sequences

>EncP
MITVIELDNSNLDDQLEAARQRARTFVNLSPAVTSLAFSRDLVKVQDERWVGYVNTSM
GGFVDHLVPYSQARQLQENLNAVATVNGAYLDDTTARTMLSRIVSLARENSITIPANL
DNLVALVNLAGIPVCFEKSGLTSGDGLPLAAALVCAAGONFKLYNGIQNPGRQALSEG
VEMPESYKOOAIIGGSQMVLGQAAARVQVRVLVQSVALSVEGLAMKIPFDFR
VHGKFRHGRQRQVASLREWGLPAMLAVHELDEQSTLAGEMTSTVAKAGSLATEDYSIRC
TFQILGPPVDDLVLRIGATLQDELNSNDHFIQLPEAEEVFHGHIFHGGQYVAMADHLNMA
LATVINLNRRVDRDLDKSNMLFAPICREDPGLRLGLGGQMMTAATAERTRTLTFPM
SVQSLISTADFOEDIVSGFVVARRRERDVLINAYYVFAELLCACOOAVDIRGADKL55FR
FLYERIKIVEFVFDFDRDETIDPVEKLAADLIAEGFVDAVAH

>AvPAL
MKTSQGQSDKSTQSFSTGNSSVYIIGNQKLTDNVARVARNGTLYSLTNNDILOGI
QRASCYIHNAVSEGIYCVTSQFQGMNVAISREQASELQINLWFLKTIAGCNKLPLAD
VRAAMLRRHSHMGRGASGRKLKIERE1FLNAGVTPYVFKEGSSGDSLVPBYITS1GS
LGLDFSFKVDNGKEMDAPTALRQLNL9FLTLLFEKEGLAMMNSTSVNTGIAANCYVTIQ
ILTAIACGGVQAIQALNGINQSFHIHNSKPKHFGQWLAAQMISSLANSLQVRDELDG
KNHYRDHELICDRYSRLRCPLQYGLFVVDIGSQTAKQIEIRINSVTNIPLDVQDGASYHG
GNTLQCGVHMENHDLRYIIGELAKHDLVDQIALLFSESNGFPSILLNERKVKVNGLKG
LQICGNSINPLTFYNSIADREFPHTAEQFHNIGQNGYTSALTARRSVDIFQNYVAIL
MFGQVQAVELATYKETQMYDARACLSFATERLYSAKVHVQKFTP5DRPYWNDNEQGLED
H!ARISADIAAAGSVTVQAVQDIIFCELH

**Figure 56** | The amino acid sequences of the enzymes EncP from *Streptomyces maritimus* and AvPAL from *Anabaena variabilis* in FASTA format.
Figure 57 || The codon optimised nucleotide sequences encoding the enzymes EncP from Streptomyces maritimus and AvPAL from Anabaena variabilis in FASTA format.
References


Title: Synthesis of D- and L-Phenylalanine Derivatives by Phenylalanine Ammonia Lyases: A Multienzymatic Cascade Process

Authors: Dr. Fabio Parmeggiani, Dr. Sarah L. Lovelock, Nicholas J. Weise, Syed T. Ahmed and Prof. Nicholas J. Turner


Impact Factor: 11.34

Volume: 54

Pages: 4608-4611

Submission Date: 2nd November 2014

Publication Date: 1st April 2015
Title: Chemoenzymatic Synthesis of Optically Pure L- and D-Biarylalanines through Biocatalytic Asymmetric Amination and Palladium-Catalyzed Arylation

Authors: Syed T. Ahmed, Dr. Fabio Parmeggiani, Nicholas J. Weise, Prof. Sabine L. Flitsch and Prof. Nicholas J. Turner

Journal: ACS Catalysis

Impact Factor: 9.31

Volume: 5

Pages: 4410-4413

Submission Date: 1st June 2015

Publication Date: 10th August 2015
Title: The Bacterial Ammonia Lyase EncP: A Tunable Biocatalyst for the Synthesis of Unnatural Amino Acids

Authors: Nicholas J. Weise, Dr. Fabio Parmeggiani, Syed T. Ahmed and Prof. Nicholas J. Turner

Journal: Journal of the American Chemical Society

Impact Factor: 12.11

Volume: 137

Pages: 12977-12983

Submission Date: 14th July 2015

Publication Date: 21st September 2015