The Use of Covalent Modification and Genetic Code Expansion to Create Organocatalytic Artificial Enzymes

A Thesis Submitted to Cardiff University for the Degree of Doctor of Philosophy by

Thomas L. Williams

February 2021

Supervisors: Dr Louis Y.P. Luk and Dr Yu-Hsuan Tsai

Acknowledgements

Firstly, I would like to thank Louis and Yu-Hsuan for taking me on as a part-time PhD student and all the help they have provided me throughout my studies. I would also like to thank Dr Alex Nödling and Sanjay Patel for training, help and advice in organic chemistry and molecular biology respectively. I would like to thank all my friends and family for their continued support. I am lucky to have worked in a great research group with brilliant colleagues that always provide entertainment in and out of the lab. I cannot mention everyone but a big thanks goes to Alex, Nicoló, Dai and Sanjay for all the great beer, food, pub times and nights out. I also want to thank Simon, Patrick and Heather for keeping me company when we are in first thing in the morning (early risers) and Luke, Emily and Vicky for the coffee breaks outside (not so early risers). Thank you also goes to Zappa, Lander, Adri and Helen for all the helpful discussions and advice.

I would like to thank all the technicians in the school of chemistry for all the hard work they put in keeping the school running, not just for this thesis but also for the contribution they make to all the research in the chemistry department. Special thanks go to Jamie for letting me jump the queue in stores, James for giving me a substitute access card when I forget mine and Robin for filling my NMR magnets when I have so many experiments on the go (or forgotten).

Finally, I would like to thank Dr Rob Jenkins for hiring me as an analytical services technician, giving me the opportunity to pursue a PhD and always supporting and mentoring me through my work and studies.

Abstract

Cyclic secondary amines can be used to perform enamine and iminium ion catalysis to modify carbonyl compounds through a broad range of chemical reactions. To utilise these compounds in a chemical biology context requires the catalysis to work in an aqueous and biocompatible environment. To accomplish this, the catalysts can be placed into proteins generating new hybrid organocatalytic artificial enzymes. This thesis concerns two strategies to generate these artificial enzymes and the subsequent characterisation of their activity.

Firstly, penicillin binding proteins are covalently modified with carbapenems that carry secondary amine pyrrolidines in their structure. Both the carbapenems alone and the protein-carbapenem hybrids were able to afford a Michael-nitro addition reaction under biocompatible reaction conditions. The second strategy concerns the direct incorporation of secondary amine unnatural amino acids into a protein backbone using genetic code expansion technology. Three unnatural amino acids are incorporated at four different positions in the multidrug binding protein LmrR. The resulting proteins are tested for their activity to perform selective reduction of α,β-unsaturated aldehydes using benzyl-nicotinamide as a hydride source. The catalytic amino acid generating the highest conversion is then incorporated into the natural oxidoreductase dihydrofolate reductase from *E. coli* at three different positions. The EcDHFR variants are then tested for their activity in the hydride transfer reaction, using NADPH as the hydride source. Two highly active enzymes LmrR Phe93DPK and EcDHFR Ala7DPK are identified and further investigated for their enzyme kinetic character and substrate scope. EcDHFR Ala7DPK is then coupled to a natural enzyme system allowing for the co-factor to be recycled *in situ*. Mechanistic studies and kinetic isotope effects are performed that strongly suggest an iminium ion mechanism in the hydride transfer reaction.

Genetic code expansion relies heavily on orthogonal tRNA synthetases. In this work, pyrrolysine tRNA synthetase form *M. bakeri* is used as the pivotal *in vivo* enzyme to incorporate the cyclic secondary amine unnatural amino acids. Although the catalytic C-terminus has been extensively explored, the tRNA binding N-terminus has been less so. Mutations in the N-terminus were performed to determine if they can increase unnatural amino acid incorporation efficiency, however the opposite was found, in contrast to its homolog from *M. mazei*. Five MbPylRS variants, eight unnatural amino acids and two different culture temperatures are explored.

Contents

Table of Contents

Acknowledgements ***II***

Abstract ***III***

Contents ***IV***

List of Figures ***VIII***

List of Tables ***XV***

Abbreviations ***XVI***

1. Introduction ***1***

1.1 Enzyme Catalysis **2**

1.2 Engineering Enzymes **6**
1.2.1 Rational Redesign of Enzymes 6
1.2.2 Directed Evolution 8
1.2.3 Repurposing Enzymes and Expanding their Reaction Profile 11

1.3 *De Novo* Artificial Enzyme Design **14**
1.3.1 Artificial Metalloenzymes 14
1.3.2 Organocatalytic Artificial Enzymes 17
1.3.3 Organocatalytic Artificial Enzymes - Covalent Modification 19
1.3.4 Organocatalytic Artificial Enzymes - Computational Design 21
1.3.5 Organocatalytic Artificial Enzymes - Supramolecular Ligand-Host Interaction 29
1.3.6 Organocatalytic Artificial Enzymes - Genetic Code Expansion 32

1.4 Cyclic Secondary Amine Artificial Enzymes **37**
1.4.1 Organocatalysis with Cyclic Secondary Amines 37
1.4.2 4-Oxalocrotonate Tautomerase 39
1.4.3 Streptavidin and Biotin-Secondary Amine Moieties 42

1.5 Aims **45**

1.6 References **47**

2. Covalently Modifying Penicillin Binding Proteins with Carbapenems to Create Organocatalytic Artificial Enzymes ***52***

2.1 Preface **52**

2.2 Introduction **53**

2.3 Results and Discussion **56**

2.4 Conclusion **64**

2.5 References **66**

3. Genetic Incorporation of Secondary Amine Unnatural Amino Acids to Create Artificial Enzymes ***67***

3.1 Preface **67**

3.2 Introduction **68**

3.3 Results and Discussion **70**

3.4 Conclusion **94**

3.5 References **96**

4. Investigating the Effects of N-terminal Mutations in Methanosarcina bakeri Pyrrolysyl-tRNA Synthetase ***97***

4.1 Preface **97**

4.2 Introduction **98**

4.3 Results and Discussion **104**

4.4 Conclusion **110**

4.5 References **111**

5. General Conclusions, Discussion and Future Work ***113***

5.1 General Conclusion and discussion **114**

5.2 Future work **116**

6. Materials and Method ***117***

6.1 Molecular biology **118**
6.1.1 LB Media 118
6.1.2 LB Agar 118
6.1.3 Chemically Competent *E. coli* Cells 118
6.1.4 Bacterial Transformation 118
6.1.5 Plasmid construction 119
6.1.6 Polymerase chain reaction (PCR) 119
6.1.7 Agarose Gel Electrophoresis 120
6.1.8 Sodium Dodecyl Sulphate Polyacryalmide Gel Electrophoresis (SDS-PAGE) 120
6.1.9 Plasmid extraction and purification 120
6.1.10 Agarose Gel Electrophoresis PCR Product Purification 121

6.2 General Buffer Solutions and Reagents **121**
6.2.1 Tris-Acetate EDTA Buffer for Agarose Gel Electrophoresis (TEA) 121
6.2.2 SDS-PAGE Resolving Buffer 121
6.2.3 SDS-PAGE Stacking Buffer 121
6.2.4 SDS-PAGE Running Buffer 122
6.2.5 Coomassie Brilliant Blue SDS-PAGE Stain 122
6.2.6 SDS-PAGE Loading Buffer (2x) and Sample Preparation 122
6.2.7 Ammonium persulfate (APS) 123
6.2.8 GQ Buffer 123
6.2.9 Antibiotic Solutions 123
6.2.10 Isopropyl β-D-1-thiogalactopyranoside (IPTG) 123

6.3 Cloning, Expression and Purification of Recombinant Proteins **123**
6.3.1 *Mycobacterium tuberculosis* β-lactamase (BlaC) 123
6.3.2 *Mycobacterium tuberculosis* β-lactamase mutant (BlaC Glu166Ala) 125
6.3.3 *Mycobacterium tuberculosis* D,D-Transpeptidase (Rv3330) 125
6.3.4 Superfolder Green Fluorescent Protein (sfGFP Asp150TAG) 126
6.3.5 *Lactococcus lactis* Multidrug Resistant Protein (LmrR) 127
6.3.6 *Lactococcus lactis* Multidrug resistant regulator Protein Mutants (LmrR TAG) 128
6.3.7 *Escherichia coli* Dihydrofolate Reductase (EcDHFR) 128
6.3.8 *Escherichia coli* Dihydrofolate Reductase Mutants (EcDHFR TAG) 129
6.3.9 *Thermoanaerobacter brockii* Alcohol Dehdrogenase (TbADH) 130
6.3.10 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase and tRNA (MbPylRS) 132

6.3.11 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase C-terminal mutants (ThzKRS, AcKRS, PrKRS and PCCRS) 133
6.3.12 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase N-terminal mutants. 133
6.3.13 Primer Table 134
6.3.14 Determination of protein concentration 135

6.4 Procedure for Small Molecule and Enzyme Catalysed Reactions and for the Determination of Enantioselectivity 135
6.4.1 Michael Nitro Addition Reaction (Carbapenems) 135
6.4.2 Michael Nitro Addition Reaction (BlaC-Carbapenem Artificial Enzymes) 136
6.4.3 Determination of Enantioselectivity for the Nitro Michael Addition reaction (Carbapenems and BlaC-Carbapenem Artificial Enzymes) 137
6.4.4 Hydride Transfer Screening Reactions (LmrR and mutants) 138
6.4.5 LmrR Phe93DPK Substrate Scope 139
6.4.6 Hydride Transfer Screening Reactions (EcDHFR and mutants) 140
6.4.7 EcDHFR Ala7DPK Substrate Scope 141

6.5 Enzyme Kinetic Analysis 142
6.5.1 Determination of Vmax, k_{cat} and K_M 142
6.5.3 Kinetic Analysis of LmrR Phe93DPK 143
6.5.3 Kinetic Analysis of EcDHFR Ala7DPK 145

6.6 Determination of Iminium Ion Intermediates 146

6.7 Isotope Incorporation Study 146

6.8 Kinetic Isotope Effect Experiments 147

6.9 Cofactor Recycling Reaction 147

6.10 Whole Cell Lysate Catalysis 148

6.11 Unnatural Amino Acid Incorporation Efficiency Fluorescence Assays 149

6.12 Synthesis of Unnatural amino acids and hydride donors 150
6.12.1 Synthesis of Secondary Amine Unnatural Amino Acids 150
6.12.2 Synthesis of Benzyl-nicotinamide (BNAH) 152
6.12.3 Enzymatic Synthesis of [4*R*-^{2}H]-Nicotinamide Adenine Dinucleotide Phosphate (NADPD) 152

6.13 NMR Spectroscopy 153

6.14 Mass Spectrometry 153
6.14.1 Protein Liquid Chromatography-Mass Spectrometry 153
6.14.2 High Resolution Mass Spectrometry for Characterisation (HR-MS) 154
6.14.3 Gas Chromatography Mass Spectrometry (GC-MS) 154
6.14.4 Liquid Chromatography-Mass Spectrometry (LC-MS) 154

6.15 Gas Chromatography Flame Ionisation Detection (GC-FID) 155

6.16 Analytical Size Exclusion Chromatography 155

6.17 Chiral High-Performance Liquid Chromatography (Chiral HPLC) 155

7. Appendix 157

7.1 Nucleotide and Amino Acid Sequences 158
7.1.1 *Mycobacterium tuberculosis* β-lactamase (BlaC) 158
7.1.2 *Mycobacterium tuberculosis* β-lactamase (BlaC Glu166Ala) 158
7.1.3 *Mycobacterium tuberculosis* D, D – Transpeptidase (Rv3330) 158
7.1.4 Superfolder Green Fluorescent Protein (sfGFP Asp150TAG) 159
7.1.5 *Lactococcus lactis* multidrug resistant regulator protein (LmrR) 159

7.1.6 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Val15TAG) 160
7.1.7 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Asp19TAG) 160
7.1.8 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Met89TAG) 160
7.1.9 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Phe93TAG) 161
7.1.10 *Escherichia coli* Dihydrofolate Reductase (EcDHFR) 161
7.1.11 *Escherichia coli* Dihydrofolate Reductase (EcDHFR Ala7TAG) 161
7.1.12 *Escherichia coli* Dihydrofolate Reductase (EcDHFR Phe31TAG) 162
7.1.13 *Escherichia coli* Dihydrofolate Reductase (EcDHFR Ser49TAG) 162
7.1.14 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase 162
7.1.15 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (ThzKRS) 163
7.1.16 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (AcKRS) 163
7.1.17 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (PrKRS) 164
7.1.18 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (PCCRS) 165
7.1.19 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg) 165
7.1.20 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg ThzKRS) 166
7.1.21 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg AcKRS) 166
7.1.22 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg PrKRS) 167
7.1.23 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg PCCRS) 167
7.1.24 *Thermoanaerobacter brockii* Alcohol Dehydrogenase (TbADH) 168

7.2 Protein Mass Spectra 169

7.3 Analytical Size Exclusion Chromatograms (LmrR) 194

7.4 Chiral HPLC Chromatograms 197

7.5 NMR Spectra 200

7.6 Mass spectra 206

List of Figures

FIGURE 1. **A)** INDUCED FIT MODEL DESCRIBES THE PREORGANISED ACTIVE SITE CONFORMATION OF AN ENZYME FOR A SPECIFIC SUBSTRATE. THE ENZYME BINDS THE SUBSTRATE TO FORM AN ENZYME:SUBSTRATE COMPLEX (E:S) TO CATALYSE THE REACTION FOLLOWED BY PRODUCT RELEASE. **B)** ENZYME CATALYSIS ENERGY DIAGRAM. THE ACTIVATION ENERGY (E_A) OF THE ENZYME CATALYSED REACTION IS LOWER THAN THAT OF THE UNCATALYSED REACTION. BLUE LINE IS THE UNCATALYSED REACTION TRAJECTORY. GREEN LINE IS THE ENZYME CATALYSED REACTION TRAJECTORY. 3

FIGURE 2. EXAMPLES OF ENZYMES APPLIED IN INDUSTRIAL SETTINGS AS BIOCATALYSTS.[8, 9] 4

FIGURE 3. A) STEREOSELECTIVE REDUCTION OF ACETOPHENONE TO THE ALCOHOL CATALYSED BY WILD-TYPE CARBONYL REDUCTASE (CR) USING NADH AS THE HYDRIDE SOURCE. **B)** EFFECT OF MUTATIONS ON SUBSTRATE SCOPE AND STEREOCHEMISTRY ON 24 KETONE SUBSTRATES. **C)** REDUCTION OF ONE SUBSTRATE BY TWO VARIANTS OF THE ENZYME TO YIELD EITHER THE *R* OR *S* ENANTIOMER. **D)** CRYSTAL STRUCTURE OF THE CARBONYL REDUCTASE FROM *C. PARAPSILOSIS* WITH NAD^+ BOUND IN THE ACTIVE SITE (**PDB: 3WLE**) USED TO RATIONALLY RE-ENGINEER THE BINDING SITE. 7

FIGURE 4. A) CHEMICAL SYNTHESIS OF SITAGLIPTIN USING A CHIRAL RHODIUM CATALYST TO REDUCE THE ENAMINE USING HIGH PRESSURE H_2. **B)** BIOCATALYTIC ROUTE USING A TRANSAMINASE AND PLP WITH THE 2-PROPYLAIME AS THE AMINE DONOR. 9

FIGURE 5. DIRECTED EVOLUTION OF A TRANSAMINASE FOR THE BIOCATALYTIC SYNTHESIS OF SITAGLIPTIN. 10

FIGURE 6. THE REPURPOSING OF HAEM BEARING ENZYMES FOR NITRILE AND CARBENE CHEMISTRY. FOLLOWING SELECTION, THE ENZYMES ARE IMPROVED THROUGH DIRECTED EVOLUTION TO GENERATE NEW ENZYME ACTIVITIES. 12

FIGURE 7. A) ALKYLATION AT SP^3 CARBON-HYDROGEN BONDS USING A REPURPOSED AND EVOLVED CYTOCHROME P450 ENZYME. **B)** SUBSTRATE SCOPE OF THE ENZYME USING AMINE BEARING SUBSTRATES AND DIFFERENT ALKYL SUBSTITUENTS. **C)** CHEMOENZYMATIC SYNTHESIS OF (+)-LYNGBIC ACID UTILISING THE REPURPOSED ENZYME TO FORM THE CHIRAL CENTRE. **D)** CHEMOENZYMATIC SYNTHESIS OF (-)-CUSPAREINE UTILISING THE REPURPOSED ENZYME TO FORM THE CHIRAL CENTRE....13

FIGURE 8. A) THE STRUCTURE OF D-BIOTIN. **B)** A RUTHENIUM (RU) CATALYST COMPLEX HOSTED BY TETRAMERIC STREPTAVIDIN (**PDB:1STP**) FOR THE SELECTIVE HYDRATION OF KETONES. **C)** AN IRIDIUM (IR) CATALYST COMPLEX HOSTED BY DIMERIC STREPTAVIDIN (**PDB: 6S4Q**) FOR THE REDUCTION OF CYCLIC IMINES. **D)** A RHODIUM (RH) CATALYST COMPLEX HOSTED BY MONOMERIC STREPTAVIDIN (**PDB:4JNJ**) FOR THE SELECTIVE SYNTHESIS OF Δ-LACTAMS. 15

FIGURE 9. A MAMMALIAN GENE SWITCH DEVISED BY USING A PURPOSEFULLY DESIGNED ARTIFICIAL METALLOENZYME. **A)** TWO BIOTIN CONJUGATES ONE FOR ABIOTIC CATALYSIS AND THE SECOND FOR CELL PENETRATION AND REPORTING. **B)** TETRAMERIC STREPTAVIDIN WITH FOUR LIGAND BINDING POCKETS ALLOWS DUAL OCCUPATION. **C)** SUBSTRATE TARGET WHICH ACTS AS A GENE REPRESSOR. **D**) ASSEMBLY OF THE ARM FOLLOWED BY DELIVERY TO THE CELL AND ACTIVATION OF BIOLUMINESCENT LUCIFERIN. 16

FIGURE 10. THE USE OF DIMERIC LMRR FOR THE DESIGN OF ARMS. **A)** SUPRAMOLECULAR ASSEMBLY TO GENERATE ARTIFICIAL ENZYMES CAPABLE OF CATALYSING FRIEDEL-CRAFT AND CYCLOPROPANATION REACTIONS. **B)** COVALENT MODIFICATION TO AFFORD DIELS-ALDER REACTIONS. **C)** METAL BINDING UNNATURAL AMINO ACID INCORPORATION THROUGH GENETIC CODE EXPANSION TO PERFORM FRIEDEL-CRAFTS AND HYDROXYLATION REACTIONS. 17

FIGURE 11. EXAMPLES OF SMALL MOLECULE ORGANOCATALYSTS. **A)** AMINE BASED ORGANOCATALYSTS THAT CAN ACTIVATE CARBONYL COMPOUNDS THROUGH THE FORMATION OF IMINIUM ION INTERMEDIATES. **B)** (THIO)UREA BASED ORGANOCATALYSTS THAT CAN ACTIVATE CARBONYLS THROUGH NON-COVALENT HYDROGEN BONDING....18

FIGURE 12. FOUR METHODS OF DESIGNING AND GENERATING ORGANOCATALYTIC ARTIFICIAL ENZYMES. 19

FIGURE 13. THE USE OF CYSTEINE AS A REACTIVE THIOL NUCLEOPHILE TO ATTACH A CATALYTICALLY ACTIVE MOLECULE SITE-SPECIFICALLY TO A PROTEIN BACKBONE. 19

FIGURE 14. THREE EXAMPLES OF COVALENT MODIFICATION OF PROTEINS WITH CO-FACTORS TO IMPART NEW REACTIONS TO THE PROTEIN BACKBONES. **A)** COVALENT MODIFICATION OF PAPAIN WITH FLAVIN FOR THE OXIDATION OF BNAH. **B)** COVALENT MODIFICATION OF PAPAIN WITH THIAMINE TO PERFORM CARBON – CARBON BOND FORMING REACTIONS. **C)** COVALENT MODIFICATION OF ADIPOCYTE BINDING PROTEIN WITH PYRIDOXAMINE FOR THE TRANSAMINATION OF KETO ACIDS TO AMINO ACIDS. 20
FIGURE 15. ENZYMES AND ENZYME SECTIONS CAN BE COMPUTATIONALLY DESIGNED USING RELEVANT SOFTWARE. THESE DESIGNS CAN THEN BE PRODUCED THROUGH RECOMBINANT EXPRESSION AND TESTED EXPERIMENTALLY. 21
FIGURE 16. THE KEMP ELIMINATION REACTION. ABSTRACTION OF THE PROTON BY A BASE, RESULTS IN RING OPENING. THE TRANSITION STATE RESULTS IN NEGATIVE CHARGE BUILD UP AT THE PHENOLIC OXYGEN. 22
FIGURE 17. A) THE DIELS-ALDER CYCLOADDITION REACTION BETWEEN THE TARGET SUBSTRATES OF THE ENZYME DESIGN. **B)** DESIGN CONSIDERATIONS FOR INCLUDING HYDROGEN BOND ACCEPTOR AND DONOR TO NARROW THE ENERGY GAP BETWEEN THE LUMO AND HOMO OF THE SUBSTRATES BY CHARGE STABILISATION AND DESIRED GEOMETRIC POSITIONING TO REDUCE ENTROPIC CONTRIBUTIONS. **C)** TYROSINE 121 WAS SHOWN TO BE THE HYDROGEN BOND DONOR AND GLUTAMINE 195 THE HYDROGEN BOND ACCEPTOR. **D)** CRYSTAL STRUCTURE OF THE ARTIFICIAL DIELS-ALDERASE RESULTING FROM THE COMPUTATIONAL DESIGN BASED ON A B-PROPELLER PHOSPHATASE SCAFFOLD (**PDB: 3I1C**). 24
FIGURE 18. **A)** THE MORITA-BAYLISS-HILLMAN REACTION THAT PROCEEDS THROUGH TWO INTERMEDIATES. **B)** CRYSTAL STRUCTURE OF BH25 WITH THE CATALYTIC DYAD CYS39 AND LYS286 HIGHLIGHTED IN BLUE **PDB: 3UW6**). **C)** CRYSTAL STRUCTURE OF BH32 WITH THE CATALYTIC DYAD HIS23 AND GLU46 HIGHLIGHTED IN BLUE (**PDB: 3U26**). 26
FIGURE 19. A) CRYSTAL STRUCTURE OF THE COMPUTATIONALLY DESIGNED RETRO-ALDOLASE RA 95.0 (GREEN) (**PDB: 4A29**) AND RA 95.5 (YELLOW) (**PDB: 5AN7**) **B)** POSITIONS OF THE CATALYTIC LYS210 (RA 95.0) AND LYS83 (RA 95.5) BOUND TO KETONE INHIBITORS. **C)** PROPOSED CATALYTIC CYCLE FOR THE RETRO-ALDOL REACTION OF METHODOL. 27
FIGURE 20. REACTION SCOPE OF THE RA95 FAMILY OF COMPUTATIONALLY DESIGNED RETRO-ALDOLASES. **A)** THE ORIGINAL RETRO ALDOL REACTION USED FOR INITIAL SCREENING, THE CLEAVAGE OF THE STARTING PRODUCT (METHODOL) RESULTS IN A CHANGE IN FLUORESCENCE**. B)** HENRY ADDITION OF NITROMETHANE TO AROMATIC ALDEHYDES. **C)** CONJUGATE ADDITION OF ACETONE TO NITROSTYRENE**.** ENAMINE CATALYSED CONJUGATE ADDITION. **D)** CONJUGATE ADDITION OF CARBON NUCLEOPHILES. **E)** CONJUGATE OF NITROMETHANE TO A,B-UNSATURATED ALDEHYDES. **F)** KNOEVENAGEL CONDENSATIONS OF CARBON NUCLEOPHILES WITH A,B-UNSATURATED AROMATIC ALDEHYDES. 28
FIGURE 21. CARTOON REPRESENTATION OF THE COMPONENTS FOR NON-COVALENT TETHERING THAT CONSISTS OF A PROTEIN HOST, A TIGHT BINDING LIGAND AND A CATALYTIC MOIETY...29
FIGURE 22. DECARBOXYLATIVE ALKYLATION REACTION PERFORMED BY THE NDI Π-ANION BASED STREPTAVIDIN CATALYST. 30
FIGURE 23. **A)** FIVE Π-ANION NDI BASED CATALYSTS CONJUGATED TO BIOTIN. **B)** CRYSTAL STRUCTURE OF TWO BIOTIN MONOMERS WITH BIOTIN ANCHORED INSIDE. RESIDUES IN BLUE WERE CRUCIAL FOR INSIGHTS INTO THE HYBRID CATALYSTS ACTIVITY (**PDB: 1MK5**). 31
FIGURE 24. MICHAEL-ALDOL DOMINO REACTION PERFORMED BY THE NDI Π-ANION BASED STREPTAVIDIN CATALYST. 31
FIGURE 25. ADDITION OF THE UNNATURAL AMINO ACID TO THE CULTURE MEDIUM ALLOWS INCORPORATION INTO THE PROTEIN BACKBONE BY AN ORTHOGONAL TRNA SYNTHETASE/TRNA 33
FIGURE 26. A) CRYSTAL STRUCTURE OF LMRR (**PDB: 3F8F**). **B)** CLOSE UP OF THE DIMER INTERFACE. THE RESIDUES IN ORANGE WERE THOSE TARGETED FOR MUTATION. **C)** REDUCTION OF *P*-AZIDOPHENYLALANINE TO *P*-AMINOPHENYLALANINE WITH TCEP. **D)** HYDRAZONE AND OXIME LIGATION PERFORMED BY THE UNNATURAL AMINO ACID *P*-AMINOPHENYLALANINE AT POSITION V15. 34
FIGURE 27. CRYSTAL STRUCTURE OF LMRR WITH THE UNNATURAL AMINO ACID *P*-AMINOPHENYLALANINE AT POSITION 15 (ORANGE) (**PDB: 6I8N**). RED RESIDUES THAT SURROUND THE ACTIVE SITE WERE TARGETED FOR DIRECTED EVOLUTION. 35

FIGURE 28. **A)** CRYSTAL STRUCTURE OF OA1.3 DERIVED FROM THE PROTEIN BACKBONE BH32. RESIDUES IN RED ARE THOSE RESULTING FROM DIRECTED EVOLUTION (**PDB: 6Q7R**). **B)**. FLUORESCEIN ESTER HYDROLYSIS CATALYSED BY METHYL-HISTIDINE. **C)** UNNATURAL AMINO ACID METHYL-HISTIDINE BOUND TO THE INHIBITOR BROMOACETOPHENONE. **D)** STABLE INTERMEDIATE FORMED BY HIS23 AND REACTIVE INTERMEDIATE FORMED BY M-HIS23. 36
FIGURE 29. THREE COMMON CYCLIC SECONDARY AMINE ORGANOCATALYSTS. **A)** L-PROLINE. **B)** FIRST GENERATION IMIDAZOLIDINONE (MACMILLAN) CATALYST. **C)** JORGENSEN-HAYASHI CATALYST. 37
FIGURE 30. A) IMINIUM ACTIVATION BY CYCLIC SECONDARY AMINES. **B)** DIAGRAM REPRESENTING THE REDUCTION OF THE ENERGY OF THE LUMO DUE TO THE IMINIUM SPECIES WITH RESPECT TO THE PARENT CARBONYL ION. **C)** ENAMINE ACTIVATION BY CYCLIC SECONDARY AMINES. **D)** DIAGRAM REPRESENTING THE INCREASE IN ENERGY OF THE HOMO OF THE ENAMINE SPECIES WITH RESPECT TO THE PARENT CARBONYL. 38
FIGURE 31. A) CRYSTAL STRUCTURE OF 4-OT WITH ONE MONOMER HIGHLIGHTED IN DARK BLUE. **B**) THE ENZYME'S NATIVE REACTION, A BASE CATALYSED TAUTOMERISATION. 40
FIGURE 32. REACTIONS CATALYSED BY 4-OT TAUTOMERASE. **A)** INTERMOLECULAR ALDOL REACTION. **B)** INTRAMOLECULAR ALDOL REACTION. **C)** ENAMINE CATALYSED CONJUGATE ADDITION. **D)** NITRO-MICHAEL ADDITION. 41
FIGURE 33. CHEMOENZYMATIC SYNTHESIS OF PREGABALIN AND THREE ANALOGUES. 42
FIGURE 34. A) THE STRUCTURE OF D-BIOTIN AND EIGHT SECONDARY AMINE ORGANOCATALYSTS THAT WERE TESTED. **B)** TETRAMERIC STRUCTURE OF STREPTAVIDIN (**PDB: 6GH7**) WITH BIOTIN PYRROLIDINE CATALYST (*R*) BOUND IN THE HYDROPHOBIC POCKET **C)** THE NITRO-MICHAEL ADDITION REACTION BETWEEN NITROMETHANE AND A RANGE OF A,B-UNSATURATED ALDEHYDES. 43
FIGURE 35. A) THE GENERAL MECHANISM OF THE PENICILLIN BINDING PROTEIN D-ALA-D-ALANINE TRANSPEPTIDASE THAT FACILITATES THE CROSS LINKING OF PEPTIDOGLYCAN IN BACTERIAL CELL WALL BIOSYNTHESIS (PG = PEPTIDOGLYCAN). **B)** THE GENERAL MECHANISM OF B-LACTAM HYDROLYSIS BY CLASS A B-LACTAMASES. BOTH ENZYMES REQUIRED A CATALYTIC SERINE (SHOWN IN BLUE). 54
FIGURE 36. A) GENERAL STRUCTURE OF THE PENICILLINS. **B)** GENERAL STRUCTURE OF THE CEPHALOSPORINS. **C)** GENERAL STRUCTURE OF THE CARBAPENEMS. **D)** THE STRUCTURES OF THE THREE CARBAPENEMS USED IN THIS STUDY. 55
FIGURE 37. A) CRYSTAL STRUCTURE OF D,D-TRANSPEPTIDASE RV3330 FROM *MYCOBACTERIUM TUBERCULOSIS* (**PDB: 4PPR**) WITH MEROPENEM COVALENTLY BOUND IN THE ACTIVE SITE. **B)** CLOSE UP OF THE MEROPENEM ACYL-ENZYME INTERMEDIATE BOUND TO THE NUCLEOPHILIC SERINE (RED) OF RV3330. **C)** CLOSE UP OF THE MEROPENEM ACYL-ENZYME INTERMEDIATE BOUND TO THE NUCLEOPHILIC SERINE (RED) OF BLAC. **D)** CRYSTAL STRUCTURE OF B-LACTAMASE BLAC FROM *MYCOBACTERIUM TUBERCULOSIS* (**PDB: 3DWZ**) WITH MEROPENEM COVALENTLY BOUND IN THE ACTIVE SITE. **E)** CHEMICAL STRUCTURE REPRESENTATION OF AS AN ESTER ADDUCT OF A NUCLEOPHILIC SERINE BOUND TO MEROPENEM. 56
FIGURE 38. THE NITRO MICHAEL ADDITION BETWEEN NITROMETHANE (**1**) AND CINNAMALDEHYDE (**2**) TO YIELD THE 1, 4 ADDITION PRODUCT (**3**) OR THE 1, 2 ADDITION SIDE PRODUCT (**4**). 57
FIGURE 39. SDS-PAGE ANALYSIS OF RECOMBINANT RV3330. CALCULATED MASS = 35879 DA. **A)** EXPRESSION ANALYSIS OF RV3330, 1 – BEFORE INDUCTION, 2 – AFTER INDUCTION WITH IPTG. **B**) NI-NTA PURIFICATION ANALYSIS OF RV3330, 1 – FLOW THROUGH, 2 – WASH, 3 – ELUTION, 4 – INCLUSION BODIES. PROTEIN LADDERS IN KDA. TWO BANDS ARE PRESENT IN THE ELUTION FRACTION. THE MONOMER OF THE PROTEIN RV3330 AT APPROX. 36 KDA AND THE DISULPHIDE DIMER AT APPROX. 72 KDA. 60
FIGURE 40. DECONVOLUTED ESI-MS OF PURIFIED RV3330 ENZYMES. **A)** ADDUCT FORMATION BETWEEN THE ACTIVE SITE SERINE AND THE CARBAPENEMS. **B)** RV3330 REFERENCE MASS = 35748 DA (35926 DA = RV3330 + FORMYL-MET). **C)** RV3330 + MEROPENEM = 36132 DA (36310 DA = RV3330 + MEROPENEM + FORMYL-MET). **D)** RV3330 + DORIPENEM = 36167 DA (36345 DA = RV3330 + DORIPENEM + FORMYL-MET). 61
FIGURE 41. SDS-PAGE ANALYSIS OF RECOMBINANT BLAC(GLU166ALA). CALCULATED MASS = 30608 DA. **A)** EXPRESSION ANALYSIS OF BLAC(E166A), 1 – BEFORE INDUCTION, 2 – AFTER INDUCTION

WITH IPTG. **B)** NI-NTA PURIFICATION ANALYSIS OF BLAC(E166A), 1 – FLOW THROUGH, 2 – WASH 1, 3 – WASH 2, 4 – ELUTION, 5 – INCLUSION BODIES. PROTEIN LADDER IN KDA. 62
FIGURE 42. DECONVOLUTED ESI-MS OF PURIFIED BLAC ENZYMES. **A)** RETRO ALDOL PRODUCT OF THE CARBAPENEMS FOLLOWING REACTION BETWEEN THE NUCLEOPHILIC SERINE AND THE CARBAPENEM. **B)** BLAC REFERENCE MASS = 30476 DA (30654 DA = BLAC + FORMYL-MET). **C)** BLAC + MEROPENEM = 30859 DA (31038 DA = BLAC + MEROPENEM + FORMYL-MET, 30816 DA = RETRO ALDOL PRODUCT OF MEROPENEM BOUND TO BLAC). **D)** BLAC + DORIPENEM = 30897 DA (31075 DA = BLAC + DORIPENEM + FORMYL-MET, 30853 DA = RETRO ALDOL PRODUCT OF DORIPENEM BOUND TO BLAC).....62
FIGURE 43. LC-MS ANALYSIS OF BLAC FOLLOWING 48 HOURS AFTER LABELLING. A) THE MEROPENEM-BLAC ARTIFICAL ENZYME. B) THE DORIPENEM-BLAC ARTIFICIAL ENZYME. THE MASS INDICATED AT 30476 IS THE CALCULATED MASS OF BLAC WITH NO CARBAPENEM ATTACHED. 64
FIGURE 44. THREE SECONDARY AMINE UNNATURAL AMINO ACIDS. **A)** D-PROLYL-L-LYSINE (DPK). **B)** L-PROLYL-L-LYSINE. **C)** THIAZOLIDINE LYSINE (THZK).....68
FIGURE 45. CRYSTAL STRUCTURE OF LMRR (**PDB:3F8F**). THE FOUR RESIDUES TARGETED FOR MUTATION ARE IN CYAN. 69
FIGURE 46. A) THE CRYSTAL STRUCTURE OF ECDHFR WITH NADPH BOUND IN THE ACTIVE SITE (**PDB: 1RA1**); RESIDUES TARGETED FOR MUTATION ARE IN YELLOW. **B)** CLOSE UP OF ECDHFR'S ACTIVE SITE.....70
FIGURE 47. SYNTHESIS OF THE UNNATURAL AMINO ACIDS DPK (*R*) (R = CH_2), LPK (*S*) (R = CH_2) AND THZK (*S*) (R = S). AN ACTIVATED ESTER IS FORMED BETWEEN THE CARBOXYLIC ACID OF THE PROLYL GROUP AND PENTAFLUOROPHENOL USING DIISOPROPYLCARBODIIMIDE AS THE COUPLING REAGENT. *N*A-BOC-L-LYSINE IS THEN COUPLED THROUGH A PEPTIDE BOND BY NUCLEOPHILIC SUBSTITUTION BETWEEN THE FREE AMINE AND THE ESTER OF THE ACTIVATED PROLYL-GROUP. THE TWO BOC PROTECTED DIPEPTIDE IS THEN DEPROTECTED USING TRIFLUORACETIC ACID. 71
FIGURE 48. SDS-PAGE ANALYSIS FOR THE INCORPORATION OF DPK, LPK AND THZK INTO THE REPORTER PROTEIN SFGFP (ASP150TAG). MASS OF SFGFP APPROX. 28 KDA. PROTEIN LADDER IN KDA.....72
FIGURE 49. SDS-PAGE ANALYSIS OF THE PURIFIED LMRR MUTANTS CONTAINING THE INDICATED UNNATURAL AMINO ACID AT THE SPECIFIED POSITION. AVERAGE MASS OF THE CALCULATED PROTEIN MONOMER = 14757 DA. PROTEIN LADDER IN KDA.....73
FIGURE 50. DECONVOLUTED ESI MASS SPECTRA OF WILD-TYPE LMRR AND EACH MUTANT, DETERMINED FROM THE PROTEIN MASS SPECTROMETRY CHARGED STATE. 74
FIGURE 51. A) NATURAL HYDRIDE DONOR NADH AND SYNTHETIC HYDRIDE DONOR BNAH. **C)** SELECTIVE REDUCTION OF CINNAMALDEHYDE BY BNAH.....75
FIGURE 52. SYNTHESIS OF THE HYDRIDE DONOR BENZYL-NICOTINAMIDE (BNAH). 75
FIGURE 53. A) SELECTIVE REDUCTION OF CINNAMALDEHYDE BY BNAH TO HYDRO-CINNAMALDEHYDE. **B)** PEAK AREA OF THE HYDRIDE TRANSFER PRODUCT HYDRO-CINNAMALDEHYDE. FOR EACH OF THE LMRR BASED ARTIFICIAL ENZYMES, THE WILD-TYPE LMRR, DPK AND THE SUBSTRATES ALONE UNDER THE SAME REACTION CONDITIONS. REACTION CONVERSION DETERMINED BY COMPARISON TO A STANDARD CALIBRATION CURVE FOR THE PRODUCT HYDRO-CINNAMALDEHYDE. EACH CHART ENTRY IS AN AVERAGE OF THREE REPEATS. ERROR BARS ARE A REPRESENTATION OF THE STANDARD DEVIATION. 76
FIGURE 54. SDS-PAGE ANALYSIS OF THE PURIFIED ECDHFR MUTANTS CONTAINING THE INDICATED UNNATURAL AMINO ACID AT THE SPECIFIED POSITION. AVERAGE MASS OF THE PROTEIN IS 19378 DA. PROTEIN LADDER IN KDA. 77
FIGURE 55. DECONVOLUTED ESI MASS SPECTRA OF WILD-TYPE ECDHFR AND EACH MUTANT, DETERMINED FROM THE PROTEIN MASS SPECTROMETRY CHARGED STATE. 78
FIGURE 56. A) SELECTIVE REDUCTION OF CINNAMALDEHYDE BY NADPH TO HYDRO-CINNAMALDEHYDE. **B)** PEAK AREA OF THE HYDRIDE TRANSFER PRODUCT HYDRO-CINNAMALDEHYDE. FOR EACH OF THE ECDHFR BASED ARTIFICIAL ENZYMES, THE WILD-TYPE ECDHFR, DPK AND THE SUBSTRATES ALONE UNDER THE SAME REACTION CONDITIONS. REACTION CONVERSION WAS DETERMINED BY COMPARISON TO A STANDARD CALIBRATION

CURVE FOR THE PRODUCT HYDRO-CINNAMALDEHYDE. EACH CHART ENTRY IS AN AVERAGE OF THREE REPEATS. ERROR BARS ARE A REPRESENTATION OF THE STANDARD DEVIATION. 79
FIGURE 57. TIME COURSE CHART MEASURING THE REACTION OF LMRR-PHE93-DPK AS A CATALYST CONVERTING CINNAMALDEHYDE TO HYDRO-CINNAMALDEHYDE. 50 MM OF LMRR WAS INCUBATED WITH 1 MM OF CINNAMALDEHYDE AND 5 MM OF BNAH. THE REACTIONS WERE QUENCHED AT EACH TIME POINT WITH THE ADDITION OF DCM AND THE CONVERSION OF SUBSTRATE TO PRODUCT DETERMINED USING GC-FID. EACH TIME POINT WAS PERFORMED THREE TIMES AND THE MEAN REPORTED. ERROR BARS ARE A REPRESENTATION OF THE STANDARD DEVIATION. 80
FIGURE 58. KINETIC CHARACTERISATION OF LMRR PHE93DPK AND ECDHFR ALA7DPK. ASSAYS WERE PERFORMED USING 50 MM OF LMRR AND 20 MM OF ECDHFR. THE ASSAYS WERE PERFORMED USING SATURATING CONCENTRATIONS OF CINNAMALDEHYDE (1 MM) AND VARYING CONCENTRATIONS OF BNAH FOR LMRR AND NADPH FOR ECDHFR. THE RATE CONSTANTS FOR LMRR WERE FITTED TO A MICHAELIS-MENTEN CURVE AND VMAX, K_{CAT}, K_M AND K_{CAT}/K_M DETERMINED USING THE SOFTWARE PRIZM. THE RATE CONSTANTS FOR ECDHFR WERE FITTED TO AN ALLOSTERIC SIGMOIDAL CURVE RELEVANT TO THE HILL EQUATION USING THE SOFTWARE PRIZM. 81
FIGURE 59. SUBSTRATE SCOPE OF THE ARTIFICIAL ENZYMES LMRR PHE93DPK AND ECDHFR ALA7DPK. CONVERSION DETERMINED BY ^{1}H NMR BY COMPARISON OF THE INTEGRALS ARISING FROM THE ALDEHYDE PROTONS OF THE SUBSTRATES AND PRODUCTS AND NORMALISED PER MOL% OF ENZYME USED (LMRR = 10 MOL% AND ECDHFR = 1 MOL%) IN THE REACTION WITH RESPECT TO THE CARBONYL SUBSTRATE. REACTIONS WERE PERFORMED IN 50 MM PBS BUFFER (PH 7.0) FOR 48 HOURS AT 25 °C WITH CONSTANT AGITATION. EACH ENTRY IS AN AVERAGE OF THREE REPEATS. 82
FIGURE 60. A) COFACTOR RECYCLING SCHEME EMPLOYED TO REGENERATE NADPH *IN SITU*, DURING THE IMINIUM ION CATALYSED HYDRIDE TRANSFER REACTION CATALYSED ECDHFR ALA7DPK. **B)** CONVERSION OF CINNAMALDEHYDE TO HYDRO-CINNAMALDEHYDE WITH AND WITHOUT THE COFACTOR RECYCLING SCHEME. 83
FIGURE 61. GC-MS CHROMATOGRAMS PRODUCED FROM THE EXTRACTION OF THE ORGANIC COMPOUNDS FROM THE WHOLE CELL LYSATE CATALYSIS USING ECDHFR ALA7DPK. THE DESIRED PRODUCT HYDRO-CINNAMALDEHYDE HAS A RETENTION TIME OF 7.21 MINUTES. BL21 (DE3) CONTROL CELLS EXTRACTED IDENTIFIED PRODUCT COMPOUNDS ARE CINNAMYL ALCOHOL AT RETENTION TIME 8.03 MINUTES AND HYDRO-CINNAMYL ALCOHOL AT 8.80 MINUTES. THE MAJOR EXTRACTABLE COMPOUND IS CINNAMALDEHYDE AT RETENTION TIME 8.37 MINUTES. THE MAJOR EXTRACTABLE PRODUCT IN THE CELLS EXPRESSING ECDHFR ALA7DPK WAS THE SUBSTRATE CINNAMALDEHYDE. 85
FIGURE 62. LC-MS ANALYSIS OF THE CHYMOTRYPSIN DIGEST OF LMRR PHE93DPK. EXTRACTION OF THE PREDICTED PEPTIDE MASS THAT CONTAINS THE UNNATURAL AMINO FROM THE PEPTIDE DIGESTS RESULTED IN EXACT MASS IDENTIFICATION OF THE PREDICTED PEPTIDE MASS WITH DPK MODIFIED WITH CINNAMALDEHYDE. 86
FIGURE 63. LC-MS ANALYSIS OF THE CHYMOTRYPSIN DIGEST OF ECDHFR ALA7DPK. EXTRACTION OF THE PREDICTED PEPTIDE MASS THAT CONTAINS THE UNNATURAL AMINO FROM THE PEPTIDE DIGESTS RESULTED IN EXACT MASS IDENTIFICATION OF THE PREDICTED PEPTIDE MASS WITH DPK MODIFIED WITH CINNAMALDEHYDE. 87
FIGURE 64. NATURAL MECHANISM OF ECDHFR. THE SATURATED IMINE OF THE PTERIN RING OF DIHYDROFOLATE IS REDUCED TO FORM TETRAHYDROFOLATE USING THE PRO-*R* HYDRIDE OF NADPH. 88
FIGURE 65. A) HR-MS ANALYSIS OF THE ENZYMATICALLY SYNTHESISED NADPD. CALCULATE M/Z FOR NADPD = 747.1052 **B)** THE ENZYMATIC REACTION USED TO INSTALL A DEUTERIUM ATOM AT THE PRO-*R* HYDRIDE OF NADPH. 89
FIGURE 66. GC-MS ANALYSIS OF THE PRODUCTS OF THE REACTION BETWEEN CINNAMALDEHYDE AND NADPH AND NADPD. INCORPORATION OF A DEUTERIUM ATOM RESULTS IN A MASS INCREASE OF 1 DA. FRAGMENT ANALYSIS OF THE RADICAL FORMED UPON ELECTRON IMPACT IONISATION SUGGESTS THAT THE DEUTERIUM IS TRANSFERRED TO THE GAMMA CARBON OF THE SUBSTRATE. 90
FIGURE 67. SELECTIVE TRANSFER OF THE PRO-*R* HYDRIDE OF NADPH TO THE GAMMA-CARBON OF CINNAMALDEHYDE BY ECDHFR ALA7DPK. 90

FIGURE 68. PROPOSAL OF THE MECHANISM OF SELECTIVE REDUCTION OF A,B-UNSATURATED CARBONYL COMPOUNDS BY ECDHFR ALA7DPK USING NADPH AS THE COFACTOR**. A)** FORMATION OF THE IMINIUM ION BETWEEN THE CARBONYL COMPOUND AND THE SECONDARY AMINE UNNATURAL AMINO ACID. **B)** TRANSFER OF THE PRO-*R* HYDRIDE OF NADPH TO THE GAMMA-CARBON OF THE A,B-UNSATURATED ALDEHYDE AND FORMATION ON AN ENAMINE. **C)** PROTONATION OF THE ENAMINE AND HYDROLYSIS OF THE RESULTING IMINIUM ION PRODUCING THE SATURATED PRODUCT AND OXIDISED $NADP^+$. 91
FIGURE 69. RATE CURVES DISPLAYING THE OXIDATION OF NADPH (BLUE) AND NADPD (RED) IN THE REACTION CATALYSED BY ECDHFR ALA7DPK. UV ABSORPTION AT 340 NM WAS USED TO MONITOR PRODUCT TURNOVER. TABLE DISPLAYING THE DETERMINATION OF THE KINETIC ISOTOPE EFFECT (KIE). 93
FIGURE 70. A SELECTION OF UNNATURAL AMINO ACIDS THAT HAVE BEEN SUCCESSFULLY INCORPORATED INTO A PROTEIN USING GENETIC CODE EXPANSION. **A)** BOC LYSINE OFTEN USED FOR INITIAL STUDIES. **B)** CYCLOPROPENE LYSINE AND **C)** THIAZOLIDINE LYSINE USED FOR SITE-SELECTIVE BIOCONJUGATION **D)** D-PROLYL LYSINE AN ANALOGUE OF PYRROLYSINE AND CATALYTICALLY COMPETENT. **E)** ACETYL LYSINE**, F)** PROPIONYL LYSINE AND **G)** THIOACETYL LYSINE USED FOR MIMICKING POST-TRANSLATIONAL MODIFICATIONS. **H)** PHOTOCAGED CYSTEINE USED IN PROTEIN ACTIVATION AND CONTROL. 98
FIGURE 71. A) CHARGING OF TRNA WITH AN AMINO ACID BY TRNA SYNTHETASE. **B)** RIBOSOMAL TRANSLATION OF MRNA INTO PROTEIN FROM THE N-TERMINUS TO C-TERMINUS. **C)** THE FULLY FORMED POLYPEPTIDE RELEASED FROM THE RIBOSOME FOLLOWING TERMINATION. 100
FIGURE 72. THE GENERAL MECHANISM OF TYPE II AMINO-ACYL TRNA SYNTHETASES. **A)** NUCLEOPHILIC ATTACK OF THE DEPROTONATED CARBOXYL GROUP TO THE A-PHOSPHATE FORMS A PENTAVALENT INTERMEDIATE, THIS COLLAPSES TO FORM AN ADENYLATE ESTER AND DIPHOSPHATE. **B)** THE AMINO ACID IS THEN CHARGED ONTO THE TRNA BY NUCLEOPHILIC SUBSTITUTION OF THE ACTIVATED ESTER. 101
FIGURE 73. A) CHARGING OF AN ORTHOGONAL TRNA WITH A NON-CANONICAL UNNATURAL AMINO ACID BY AN ORTHOGONAL TRNA SYNTHETASE. B**)** RIBOSOMAL TRANSLATION OF THE UNNATURAL AMINO ACID INTO THE PROTEIN BACKBONE USING A NON-CODING CODON. **C)** THE FULLY TRANSLATED POLYPEPTIDE WITH THE UNNATURAL AMINO ACID SITE-SPECIFICALLY INCORPORATED INTO TO THE PROTEIN. 102
FIGURE 74. A) CRYSTAL STRUCTURE OF MMPYLRS (**PDB: 6ABM**), C-TERMINAL DOMAIN IN RED AND N-TERMINAL DOMAIN IN BLUE. **B)** SEQUENCE ALIGNMENT OF MMPYLRS AND MBPYLRS N-TERMINUS. **C)** THE STRUCTURE OF THE 22^{ND} PROTEINOGENIC AMINO ACID PYRROLYSINE. ..104
FIGURE 75. A) POST-TRANSLATIONAL CHROMOPHORE MATURATION OF GREEN FLUORESCENT PROTEIN PROPOSED MECHANISM.[37-39] **B)** TRANSLATION OF THE NATIVE PROTEIN PRODUCES GREEN FLUORESCENCE, A TAG MUTATION RESULTS IN PREMATURE TERMINATION RESULTING IN NO FLUORESCENCE, ADDITION OF AN UNNATURAL AMINO ACID AND THE CORRESPONDING MACHINERY RESULTS IN FLUORESCENCE CORRESPONDING TO THE INCORPORATION EFFICIENCY. **C)** CRYSTAL STRUCTURE OF SFGFP WITH THE ASN150 HIGHLIGHTED IN RED (**PDB: 4LQT**) 105
FIGURE 76. OUTLINE OF GFP REPORTER ASSAY TO DETERMINE INCORPORATION EFFICIENCY OF THE TRNA SYNTHETASES. CO-TRANSFORMATION OF THE SYNTHETASE/TRNA PLASMID WITH THE REPORTER PLASMID GFP. CULTURE TO AN OD_{600} OF 0.6. INDUCE GFP EXPRESSION WITH IPTG (0.5 MM) AND ADD THE UNNATURAL AMINO ACID (5 MM). SAMPLE THE CULTURE. NORMALISE THE OD_{600} TO 1.0 AND MEASURE THE RELATIVE FLUORESCENCE WITH A PLATE READER. 106
FIGURE 77. COMPARISON OF THE INCORPORATION EFFICIENCY OF THE FIVE TRNA SYNTHETASES WITH RESPECT TO THEIR UNNATURAL AMINO ACIDS. REPORTED AS NORMALISED FLUORESCENCE RELATIVE TO THE MOST INTENSE SIGNAL MEASURED FROM THE GFP PLATE READER ASSAY. MEAN OF THREE BIOLOGICAL REPLICATES ± STANDARD DEVIATION (S.D.). RAW DATA TABLE AND STATISTICAL P VALUES FOR EACH COMPARISON DETERMINED FROM TWO TAILED T TEST ASSUMING UNEQUAL VARIANCE. SIGNIFICANCE LEVEL >0.05. CULTURES PERFORMED AT 20 °C FOR 18 HOURS FOLLOWING REPORTER GENE EXPRESSION RESULTING FROM THE ADDITION OF 0.5 MM OF IPTG AND 5 MM OF THE RELEVANT UNNATURAL AMINO ACID. 107

FIGURE 78. COMPARISON OF THE INCORPORATION EFFICIENCY OF MBPYLRS (LEFT) WITH TWO NEW SUBSTRATES CYPK AND DPK AND ACKRS (RIGHT) WITH THE SUBSTRATES PK AND TACK. REPORTED AS NORMALISED FLUORESCENCE RELATIVE TO THE MOST INTENSE SIGNAL MEASURED FROM THE GFP PLATE READER ASSAY. MEAN OF THREE BIOLOGICAL REPLICATES ± STANDARD DEVIATION (S.D.). RAW DATA TABLE AND STATISTICAL P VALUES FOR EACH COMPARISON DETERMINED FROM TWO TAILED T TEST ASSUMING UNEQUAL VARIANCE. SIGNIFICANCE LEVEL >0.05. CULTURES PERFORMED AT 20 °C FOR 18 HOURS FOLLOWING REPORTER GENE EXPRESSION RESULTING FROM THE ADDITION OF 0.5 MM OF IPTG AND 5 MM OF THE RELEVANT UNNATURAL AMINO ACID. 108
FIGURE 79. COMPARISON OF THE INCORPORATION EFFICIENCY OF MBPYLRS WITH BOCK AND ACKRS WITH RESPECT TO EXPRESSION TEMPERATURES 20 °C AND 37 °C. REPORTED AS NORMALISED FLUORESCENCE RELATIVE TO THE MOST INTENSE SIGNAL MEASURED FROM THE GFP PLATE READER ASSAY. MEAN OF THREE BIOLOGICAL REPEATS. MEAN OF THREE BIOLOGICAL REPLICATES ± STANDARD DEVIATION (S.D.). RAW DATA TABLE AND STATISTICAL P VALUES FOR EACH COMPARISON DETERMINED FROM TWO TAILED T TEST ASSUMING UNEQUAL VARIANCE. SIGNIFICANCE LEVEL >0.05. CULTURES PERFORMED AT 20 °C OR 37 °C FOR 18 HOURS FOLLOWING REPORTER GENE EXPRESSION RESULTING FROM THE ADDITION OF 0.5 MM OF IPTG AND 5 MM OF THE RELEVANT UNNATURAL AMINO ACID. 109
FIGURE 80. SDS-AGE ANALYSIS OF LMRR PHE93DPK FOLLOWING SIZE EXCLUSION CHROMATOGRAPHY. MASS OF LMRR PHE93DPK MONOMER APPROX. 14.5 KDA. PROTEIN LADDER IN KDA. 128
FIGURE 81. SDS-AGE ANALYSIS OF LMRR PHE93DPK FOLLOWING SIZE EXCLUSION CHROMATOGRAPHY. MASS OF LMRR PHE93DPK MONOMER APPROX. 19.5 KDA. PROTEIN LADDER IN KDA. 130
FIGURE 82. SDS-PAGE ANALYSIS FOR THE EXPRESSION AND PURIFICATION OF TBADH. BI = BEFORE INDUCTION OF GENE EXPRESSION BY IPTG. AI = AFTER INDUCTION OF GENE EXPRESSION BY IPTG. SN = SUPERNATANT FOLLOWING LYSIS BY SONICATION. IB = INCLUSION BODIES FOLLOWING LYSIS BY SONICATION. SF = SOLUBLE FRACTION FOLLOWING THERMAL PRECIPITATION. IF = INSOLUBLE FRACTION FOLLOWING THERMAL PRECIPITATION. IX1 = ION EXCHANGE ELUTION 1. IX2 = ION EXCHANGE ELUTION 2. IX3 = ION EXCHANGE ELUTION 3. MASS OF TBADH APPROX. 37 KDA. PROTEIN LADDER IN KDA. 132
FIGURE 83. GC-FID HYDRO-CINNAMALDEHYDE STANDARD CALIBRATION CURVE FOR THE DETERMINATION OF PRODUCT CONCENTRATION. 142

List of Tables

TABLE 1. PERCENTAGE CONVERSION CINNAMALDEHYDE TO NITRO-CINNAMALDEHYDE (**3**) WITH DIFFERENT BUFFERS AT PH 7.0. PERCENTAGE CONVERSION DETERMINED BY ^{1}H NMR AFTER 24 H AT 25 °C. EACH ENTRY IS AN AVERAGE OF THREE REPEATS.....57
TABLE 2. PERCENTAGE CONVERSION OF SUBSTRATE TO PRODUCT AND SIDE PRODUCT IN 10 MM PBS AT INDICATED PH. PERCENTAGE CONVERSION DETERMINED BY ^{1}H NMR AFTER 24 H AT 25 °C. EACH ENTRY IS AN AVERAGE OF THREE REPEATS. 58
TABLE 3. PERCENTAGE CONVERSION OF SUBSTRATE TO PRODUCT IN THE PRESENCE OF CO-SOLVENTS AT 1:1 RATIO. PERCENTAGE CONVERSION DETERMINED BY ^{1}H NMR AFTER 24 H AT 25 °C. EACH ENTRY IS AN AVERAGE OF THREE REPEATS. 58
TABLE 4. ENANTIOMERIC RATIO (*R*:*S*) OF PRODUCT **3**. RATIO CALCULATED FROM THE RATIO OF THE PEAK AREA FROM CHIRAL HPLC. ASSIGNMENT PERFORMED BY COMPARISON OF A PURE CHIRAL STANDARD FROM THE LITERATURE.[10] 59
TABLE 5. PERCENTAGE CONVERSION AND ENANTIOMERIC RATIOS OF **3** IN THE PRESENCE OF BLAC-CARBAPENEM COMPLEXES. CONVERSION DETERMINED USING ^{1}H NMR AFTER 18 H AT 25 °C. ENANTIOSELECTIVITY CALCULATED FROM THE RATIO OF THE PEAK AREA DETERMINED BY CHIRAL HPLC. ASSIGNMENT PERFORMED BY COMPARISON OF A PURE CHIRAL STANDARD FROM THE LITERATURE. 63
TABLE 6. SUBSTRATE TO PRODUCT CONVERSION FOR THE SELECTIVE REDUCTION OF CINNAMALDEHYDE TO HYDRO-CINNAMALDEHYDE USING NADPH AS THE HYDRIDE SOURCE IN THE PRESENCE OF GLUCOSE-6-PHOSPHATE AND GLUCOSE-6-PHOSPHATE DEHYDROGENASE. TOTAL TURNOVER NUMBERS CALCULATED BY DIVIDING MOLES OF PRODUCT BY MOLES OF COFACTOR ADDED TO THE REACTION. REACTIONS WERE PERFORMED IN 50 MM PBS BUFFER WITH 5% METHANOL AND 10 MOL% CATALYST LOADING. SUBSTRATE CONVERSION AND TTN ARE AND AVERAGE OF THREE REPEATS. 84
TABLE 7. POLYMERASE CHAIN REACTION CONDITIONS. 119
TABLE 8. SDS-PAGE RECIPE BUFFERS. 120
TABLE 9. ANTIBIOTIC CONCENTRATIONS. 123
TABLE 10. C-TERMINAL MUTATIONS ALIGNED FOR EACH RESIDUE COMPARED TO THE WILD-TYPE MBPYLRS. 133
TABLE 11. PRIMER TABLE. 134
TABLE 12. TRIPLICATE CONVERSION OF SUBSTRATE TO THE PRODUCT HYDRO-CINNAMALDEHYDE BY LMRR AND VARIANTS DETERMINED BY COMPARISON OF THE PEAK AREA AGAINST A STANDARD CALIBRATION CURVE BY GC-MS. 138
TABLE 13. CONVERSION OF PRODUCTS DETERMINED BY ^{1}H NMR SPECTROSCOPY USING LMRR PHE93DPK AS THE CATALYST. 139
TABLE 14. TRIPLICATE SUBSTRATE TO PRODUCT CONVERSION OF HYDRO-CINNAMALDEHYDE BY ECDHFR AND VARIANTS DETERMINED BY COMPARISON OF THE PEAK AREA AGAINST A STANDARD CALIBRATION CURVE BY GC-MS. 140
TABLE 15. CONVERSION OF PRODUCTS DETERMINED BY ^{1}H NMR SPECTROSCOPY USING ECDHFR ALA7DPK AS THE CATALYST. 141
TABLE 16. TIME COURSE FOR THE REACTION CATALYSED BY LMRR-PHE93-DPK. CONVERSION OF THE SUBSTRATE TO PRODUCT DETERMINED USING GC-FID. 144
TABLE 17. RATE DETERMINATION OF THE REACTION CATALYSED BY LMRR PHE93DPK DEPENDENT ON CONCENTRATION OF BNAH. RATES ARE AN AVERAGE OF THREE REPEATS. 144
TABLE 18. RATE DETERMINATION OF THE REACTION CATALYSED BY ECDHFR ALA7DPK DEPENDANT ON CONCENTRATION OF NADPH. RATES ARE AN AVERAGE OF THREE REPEATS. 145
TABLE 19. PROTEIN LC-MS CHROMATOGRAPHY PARAMETERS. 154
TABLE 20. LC-MS CHROMATOGRAPHY PARAMETERS. 155

Abbreviations

4-OT	4-Oxalocrotonate tautomerase
3D	Three dimensional
AcK	Acetyl lysine
AcKRS	Acetyl lysine tRNA synthetase
Ala	Alanine
ADH	Alcohol dehydrogenase
Arg	Arginine
ArM	Artificial metallo-enzyme
Asp	Aspartic acid
Asn	Asparagine
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BlaC	β-lactamase
Boc	Di-tertbutyl di-carbonate
BocK	*N*-ε-Boc lysine
BNAH	Benzyl-nicotinamide
CyPK	Cycloproene lysine
Cys	Cysteine
DNA	Deoxyribonucleic acid
DCM	Dichloromethane
DPK	D-prolyl-L-lysine
D.R.	Diastereomeric ratio
Ea	Activation energy
e.e.	Enantio-excess
EI	Electron impact ionisation
ESI-MS	Electrospray ionisation-mass spectrometry
EcDHFR	Escherichia coli dihydrofolate reductase
EtOH	Ethanol
GC	Gas chromatography
GC-MS	Gas chromatography-mass spectrometry
Gln	Glutamine
Glu	Glutamate
Gly	Glycine

HEK293T	Human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HOMO	Highest occupied molecular orbital
HPLC	High performance liquid chromatography
IPA	2-propanol
Ile	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k_{cat}	Catalytic turnover
K_d	Dissociation constant
KIE	Kinetic isotope effect
K_M	Michaelis-Menten constant
KPi	Potassium phosphate
L. lactis	Lactococcus lactis
LC-MS	Liquid chromatography-mass spectrometry
Leu	Leucine
LmrR	Lactococcal multidrug resistant regulator
LPK	L-prolyl-L-lysine
LUMO	Lowest unoccupied molecular orbital
Lys	Lysine
MeOH	Methanol
Met	Methionine
M-His	Methyl-histidine
MbPylRS	Methanosarcina bakerii pyrrolysyl-tRNA synthetase
MjTyrRS	Methanococcus jannaschii tyrosyl-tRNA synthetase
MmPylRS	Methanosarcina mazei pyrrolysyl-tRNA synthetase
mRNA	Messenger ribonucleic acid
NaPi	Sodium phosphate
NAD^+	Nicotinamide adenine dinucleotide (oxidised form)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidised form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
NADPD	^{4}C pro-*R* deuterated nicotinamide adenine dinucleotide phosphate
NDI	Naphthalenediimide

Ni-NTA	Nickel-nitriloacetic acid
NMR	Nuclear magnetic resonance spectroscopy
NOX	NAD oxidase
NT	N-terminus
OAE	Organocatalytic artificial enzyme
OD	Optical density
PBS	Phosphate buffered saline
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
PCC	Photocaged cysteine
PCCRS	Photocage cysteine tRNA-synthetase
PDB	Protein databank
Phe	Phenylalanine
Pk_a	Acid dissociation constant
PLP	Pyridoxal phospate
PrK	Proionyl lysine
PrKRS	Propionyl lysine tRNA syntetase
Pro	Proline
PyltRNA	Pyrrolysyl-tRNA
PylRS	Pyrrolysyl-tRNA synthetase
RF1	Release factor 1
Rt	Room temperature
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide electrophoresis
Ser	Serine
sfGFP	Super folder green fluorescent protein
SOMO	Singly occupied molecular orbital
TA	Transaminase
TAcK	Thioacetyl lysine
TbADH	Thermoanaerobacter brockii alcohol dehydrogenase
TFA	Trifluoracetic acid
TIM	Triosephosphate isomerase
Thr	Threonine
ThzK	Thiazolidine lysine
ThzKRS	Thiazolidine lysine tRNA synthetase
Tris	Tris(hydroxymethyl)amino methane
Trp	Tryptophan

TTN	Total turnover number
Tyr	Tyrosine
UV	Ultraviolet
Val	Valine
Vmax	Maximum velocity

1. Introduction

1.1 Enzyme Catalysis

To perform chemical reactions fast enough to sustain life, nature employs protein catalysts known as enzymes. As biological components of all living cells, enzymes catalyse their reactions at moderate temperatures, physiological pH and under largely aqueous environments. Even under these mild conditions, enzymes are able to increase the rate of a given reaction faster than any man-made catalyst. For example carbonic anhydrase, which regulates CO_2 levels in the blood,[1] and β-lactamases, which confer antibiotic resistance by hydrolysing β-lactam antibiotics,[2] can enhance the rate of the catalysed reaction to the uncatalysed counterpart by an order of 10^6. In addition to their remarkable activity, most enzymes are exceptionally specific for their given reaction and highly selective with respect to both regio- and stereochemistry.

How these proteins are able to afford such efficient and selective reactions is not fully understood and is a topic of hot debate and fundamental research. Nevertheless, evidence has come to the fore that suggests amino acid side chains can form highly ordered binding pockets known as active sites that promote favourable electrostatic properties for a given reaction.[3] Furthermore, the size and geometry of this active site produces a suitable fit and space for specific substrates known as the induced fit model (**Figure 1a**). Protein dynamics (the movement of the enzymes themselves) are crucial to substrate recruitment (forming an enzyme:substrate complex) and product release,[4] which are thought to favour the production of the transition state and lower the activation energy (E_a) between substrate and product (**Figure 1b**). Consequently, the reaction rate increases whilst there is exquisite control of the chemistry performed. Based on bioinformatics analysis, enzyme specificity and efficiency are thought to be the product of millions of years of natural evolution.

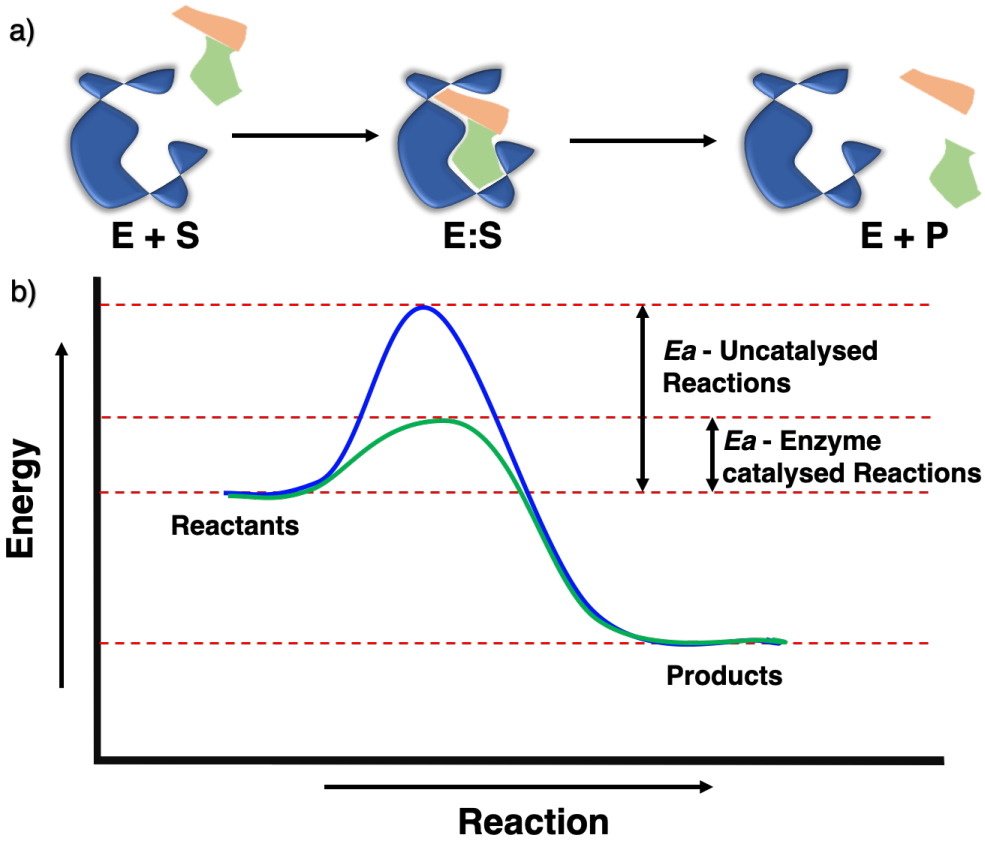


Figure 1. **a)** Induced fit model describes the preorganised active site conformation of an enzyme for a specific substrate. The enzyme binds the substrate to form an enzyme:substrate complex (E:S) to catalyse the reaction followed by product release. **b)** Enzyme catalysis energy diagram. The activation energy (E_a) of the enzyme catalysed reaction is lower than that of the uncatalysed reaction. Blue line is the uncatalysed reaction trajectory. Green line is the enzyme catalysed reaction trajectory.

When enzymes are exploited for synthetic uses and industrial applications they are often referred to as biocatalysts.[5] They can be classed as sustainable catalysts for the following reasons.[6] Firstly, through recombinant protein technology, enzymes can be generated through environmentally benign conditions. Secondly, as mentioned above, they are able to work under mild and safe reaction conditions that require low volumes of toxic or hazardous solvents. Thirdly, they are non-toxic and biodegradable and lastly, due to their intrinsic efficiency and specificity they can reduce the number of side products, boosting yields of desired compounds relative to chemical catalysis. In addition, their modification through genetic manipulation, such as directed evolution, has allowed the identification of modified enzymes with more desired properties.[7] Enzymes have now found prevalent use in industry for a wide range of applications from food and beverage manufacturing to the synthesis of pharmaceuticals and plastics (**Figure 2**).

Figure 2. Examples of enzymes applied in industrial settings as biocatalysts.[8, 9]

A vast number of enzymes are now applied as biocatalysts in industry. Particularly noticeable are those enzymes used in organic synthesis, a prominent example is the enzymatic synthesis of acrylamide. Acrylamide is used in a range of polymers as well as applications in biotechnology and water treatment. The enzyme nitrile hydratase is able to hydrolyse acrylonitrile into acrylamide, supplanting the chemical hydrolysis via sulphuric acid. It has been reported that 30,000 tons per year are produced through the biocatalytic method.[10] Lipases have also found wide applicability in the synthesis of pharmaceutical intermediates.[9] The anti-anxiety drug, pregabalin[11] and the anti-inflammatory compounds ibuprofen,[12] flurbiprofen[13] and ketoprofen,[14] like many biologically active chemicals require enantioselective synthesis. This is where the selectivity of enzymes can be applied. For instance, lipases have been used in the stereoselective synthesis of these compounds under mild and convenient reaction conditions.

Clearly the use of enzymes as catalysts holds vast potential. However, there is a major drawback. Although enzymes are highly efficient at their specific reactions, this hinders their broad application. Indeed, an enzyme that may perform a reductive hydrogenation across a carbon double bond for a specific compound with high fidelity, may have no activity on a similar yet structurally distinct compound. In comparison to heterogeneous catalysts, catalysts such as nickel are able to perform this reaction on many substrates, so long as the double bond is present, however it is often difficult to induce stereo and regio-selectivity.

To this end, many chemists have endeavoured to engineer enzymes to increase their substrate scope, to alter their reactivity (e.g., from hydrolysis to ligation) or to generate entirely new activities by creating artificial enzymes. Hydride transfer enzymes, that activate their substrates via one specific mechanism using a universal hydride source for a vast range of substrates, would be one example. Another would be the conjugate addition of alkyl-nitro groups, rare amongst known enzyme reactions. As previously mentioned, the exact circumstances allowing enzymes to perform with such efficiency and selectivity is not fully understood. During the design, engineering and tuning the reactivity of a natural or artificial enzyme, insights may be gleaned of how these proteins have evolved to perform such transformations. The attempt to generate new artificial enzymes is continually an active area of research. As understanding of the underlying characteristics of Nature's catalysts continues to be explored and the technology employed, from directed evolution and bioconjugation to genetic code expansion and computational design, become more sophisticated, the goal of generating tailor-made biocatalysts is increasingly within reach.

1.2 Engineering Enzymes

For more than three decades, the strive to engineer natural enzymes and design artificial enzymes has been a topic of heated research. Here, I will summarise the current advances in enzyme engineering and artificial enzyme design. Firstly, discussion will concern engineering enzymes to expand their substrate scope to enable a broader profile of substrates to be accepted, or enhancement of characteristics to enable their use in desired applications such as increased thermal stability. Secondly, the repurposing of enzymes to broaden their reaction profile, and finally the *de novo* design of artificial enzymes by introducing new catalytic prowess into a protein backbone often inspired by small molecule chemical catalysis. Relevant examples will be presented that have endeavoured to expand the biocatalysis toolbox.

1.2.1 Rational Redesign of Enzymes

Enzyme and protein engineering generally require detailed information (e.g., structure, activity, kinetic information) of the system to be modified.[15] As described in the induced fit model, the active site of the enzyme in question is optimised for one substrate, to perform a specific chemical transformation. Hence, structural information often provided by protein crystallography, but also derived from NMR and computational modelling, has been pivotal in analysis and engineering. The crystallisation of a protein or enzyme in its apo (free of substrates, intermediates, inhibitors or products) or holo (bound to substrates, intermediates, inhibitors or products) form can aid in the identification of key amino acid residues in the active site that may play a role in the enzyme's catalysis. Detection of such residues can lead to insights into how substrates are bound and how they may proceed through any relevant intermediates producing the product. In turn, target residues can be identified for the modification of the active site to permit a wider range of substrates, broadening the utility of the enzyme as a general catalyst. In addition, kinetic data that describes the rate of an enzyme reaction can aid in understanding of the fundamental chemistry the enzyme performs. Residue substitution is often performed through site-directed mutagenesis, where codons that encode a particular amino acid are exchanged for those that encode a different amino acid. Many enzymes have benefitted from this structure/activity analysis including lipases,[16] transaminases,[17] oxidoreductases[18] and hydrogenases.[19]

A notable example is that of a carbonyl reductase isolated from *Candida parapsilosis*.[20] The enzyme naturally reduces ketones to alcohols using nicotinamide adenine dinucleotide (NADH) as a hydride source (**Figure 3a**). The enzyme has a narrow substrate scope accepting only a handful of small aryl ketones and β-keto esters.[21] However, it was reasoned that the expansion of the substrate scope of this enzyme could allow for a wider range of enantioselective transformation of ketones into their respective chiral alcohols, useful intermediates in biologically active compounds. Upon crystallisation the protein's active site was shown to have three bulky aromatic residues, including Trp116, Phe285 and Trp286.[22] The crystal structure displayed NAD^+ in proximity to these residues (**Figure 3d**). In addition, Trp116 was shown to have an interaction with the aryl ring of a substrate bound in the active site. Mutation of these three residues to alanine, both individually and in combination, increased the catalytic efficiency towards a number of larger substrates. Sixteen substrates tested showed improved catalytic efficiency assessed by measuring their k_{cat} constants compared to those measured with the wild-type enzyme. In addition, it was shown that the Trp116Ala mutant did not affect the stereoselectivity of the enzyme. However, the Phe285Ala and Trp286Ala mutants did have an effect (**Figure 3b**). For eight of the substrates tested the enantioselectivity was reversed and the enantiomeric excess exceeded 80%. For one substrate, the mutant produced a product alcohol with 100% inverted enantioselectivity compared to the wild-type enzyme (**Figure 3c**).

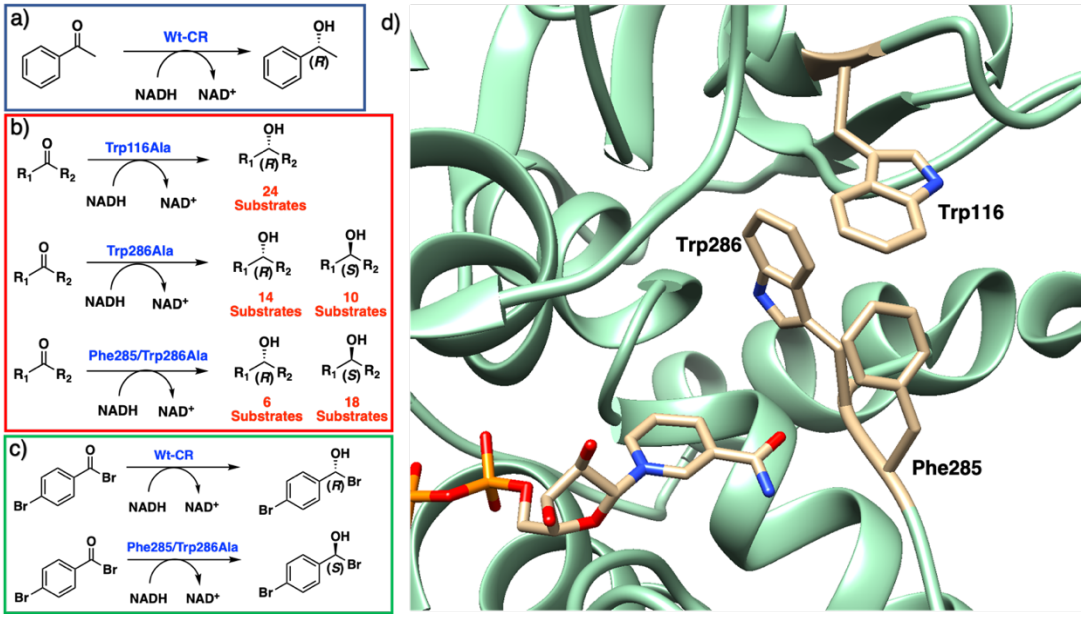


Figure 3. a) Stereoselective reduction of acetophenone to the alcohol catalysed by wild-type carbonyl reductase (CR) using NADH as the hydride source. **b)** Effect of mutations on substrate scope and stereochemistry on 24 ketone substrates. **c)** Reduction of one substrate

by two variants of the enzyme to yield either the *R* or *S* enantiomer. **d)** Crystal structure of the carbonyl reductase from *C. parapsilosis* with NAD^+ bound in the active site (**PDB: 3WLE**) used to rationally re-engineer the binding site.

1.2.2 Directed Evolution

The rational modification of an active site based on its 3D structure is a powerful tool for enzyme engineering. However, it is not always clear exactly how an enzyme can perform its reaction or there is limited data available to assume such modifications will be successful. Hence, there is a need to execute random singular mutation and trial-and-error assays. The advent of directed evolution is often seen as a monumental advancement in protein engineering. Directed evolution aims to replicate natural evolution in the laboratory.[23] Codons of a gene of interest are artificially randomised and following selective screening, variants that produce a desired phenotype can be isolated. Through modern genetic editing techniques such as error-prone PCR and site-saturation mutagenesis, multiple positions throughout the nucleotide chain may be randomised. Suitable screening assays allow for high throughput selection of many mutants simultaneously. Through iterative rounds of variation followed by assay screening, a proteins activity can be tuned, until a reasonable characteristic or change in characteristic has been realised.[24]

Directed evolution has now been applied to a vast number of enzymes to enhance stability or to tune poor efficiency or selectivity for given substrates. A highly successful example is the directed evolution of a ω-transaminase in the synthesis of the anti-diabetic drug sitagliptin phosphate.[25] The key step in the synthesis is to transform pro-sitagliptin to sitagliptin by turning a ketone group into a chiral amine. Under chemical catalysis an enamine is first formed, followed by high pressure (250 psi) reduction to the amine using H_2 and a chiral rhodium catalyst (**Figure 4a**). The catalyst must then be removed from the reaction mixture and recrystallization is performed to enrich the desired enantiomer. To supplant the toxic catalyst and dangerous reaction conditions, a biocatalytic route to the product was designed (**Figure 4b**).

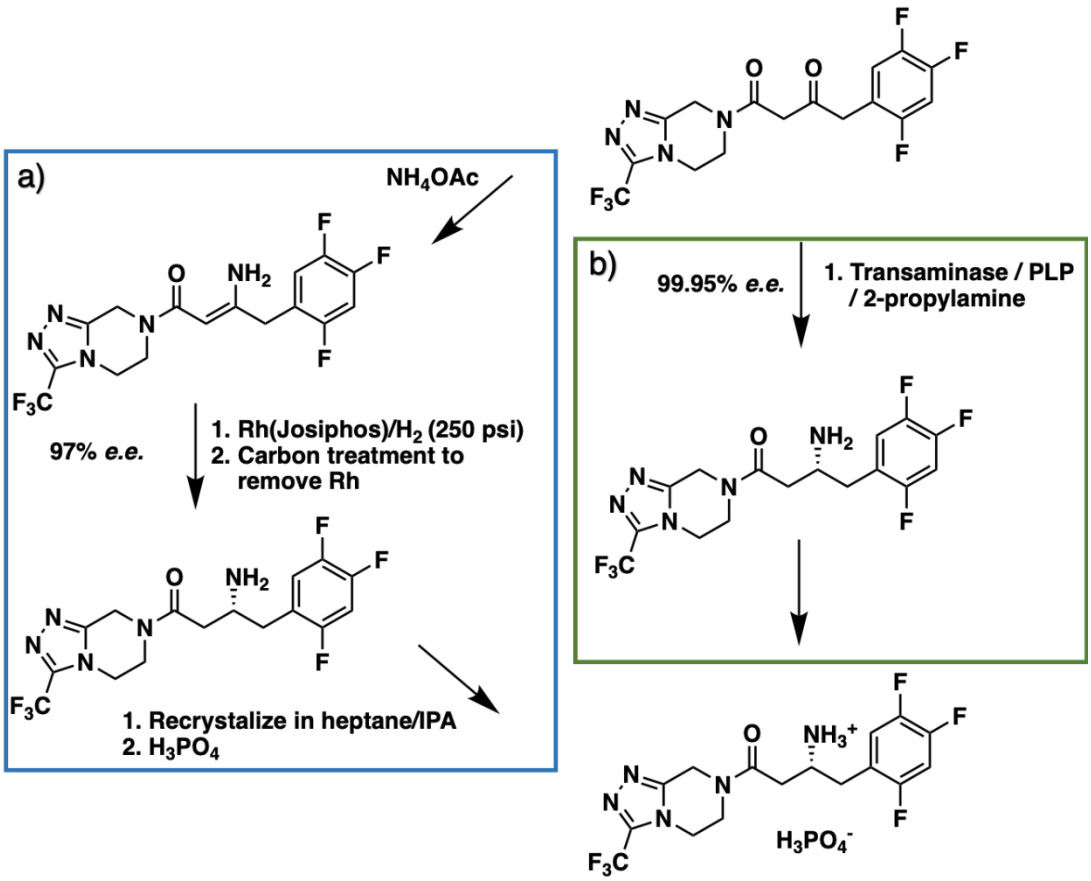


Figure 4. a) Chemical synthesis of sitagliptin using a chiral rhodium catalyst to reduce the enamine using high pressure H_2. **b)** Biocatalytic route using a transaminase and PLP with the 2-propylaime as the amine donor.

A transaminase was engineered that can perform the corresponding reductive amination to generate nearly enantiopure product (99.95% *e.e.*), using the co-factor pyridoxal phosphate and 2-propylamine as the amine donor. A 96-well plate culture using robotic handling was used for mutant expression followed by lysis, addition of the substrates and HPLC analysis to determine the conversion. To find an enzyme capable of performing the reaction, a number of commercially available transaminases were first screened. None, however, was suitable. Next, a transaminase reported to act on small cyclic ketones was tested (**Figure 5a**). A homology model was generated and docking studies suggested that the binding of the substrate in the active site would be unlikely due to steric hindrance. A truncated substrate of the pro-sitagliptin ketone was used to model inside the binding pocket and saturation mutagenesis was performed on residues lining this pocket. A Ser-to-Pro mutation showed limited activity towards the transamination of the truncated substrate (**Figure 5b**). Residues that were shown to be in proximity to the trifluorophenyl group, were targeted next. Single mutation variants as well as

combinatorial libraries were generated and four mutations in the active site resulted in measurable activity on the full substrate.

The most efficient mutant was used as the template for the next round of evolution. Twelve mutations (ten in the active site) resulted in a 75-fold increase in activity. A further nine rounds of directed evolution produced mutations predominantly at the dimer interface of the enzyme. It was reasoned that these mutations strengthen the dimer formation and overall structure of the protein under the industrial conditions that were needed for the drug synthesis. After screening a total of 36,480 enzyme variants it was concluded that the transamination of pro-sitagliptin to sitagliptin by the evolved transaminase, resulted in 27,000-fold improvement compared with the parent enzyme (**Figure 5c**). Directed evolution to tune the protein to the required reaction conditions allowed the enzyme to work efficiently at 45 °C (to solubilise the substrate and drive the reaction forward by evaporating the acetone byproduct) in the presence of 50% dimethyl sulfoxide.

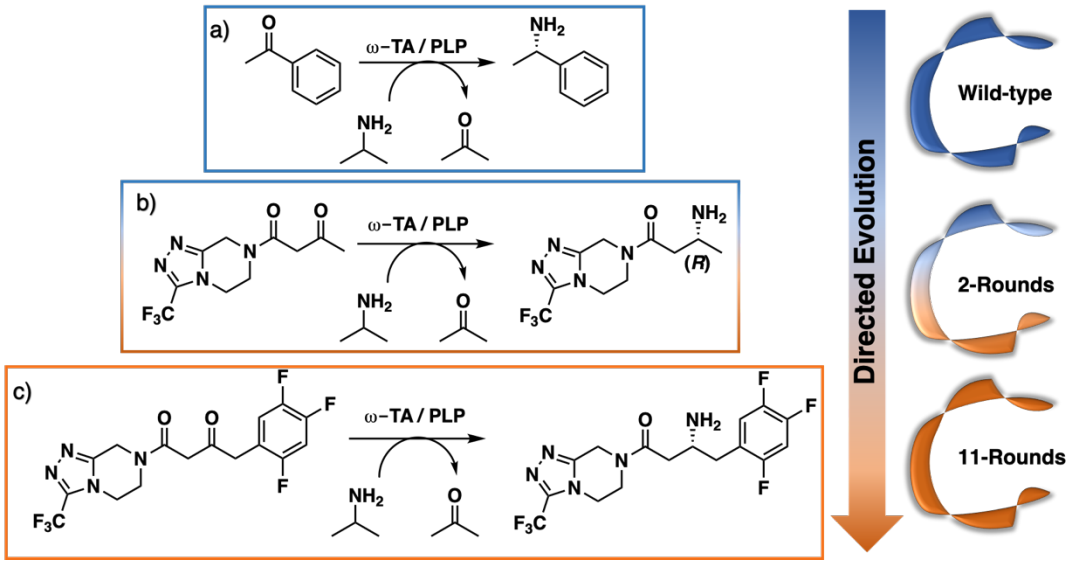


Figure 5. Directed evolution of a transaminase for the biocatalytic synthesis of Sitagliptin.

The above example clearly demonstrates the power of directed evolution in modifying properties of enzymes to adapt to conditions outside their native environment. Tuning of the enzyme from negligible activity for a truncated substrate, to a highly efficient system that can supplant the current chemical method is an impressive feat. However, it is of fundamental interest to explore reactions that are orthogonal to nature. Expanding the repertoire of enzyme catalysis can open up opportunities that will lay the foundations for utilising designer biocatalysts in industrial settings.

Directed evolution is a powerful technique to re-engineer enzymes but requires the use of high through-put screening methods to rapidly identify mutants with desired properties. This often requires spectroscopic or visible light changes in the reactions performed. Although mass spectrometry or chromatographic analysis may also by employed, targeting multiple mutants simultaneously can be cumbersome and difficult to utilise. Variant library size and suitable screening methods must be considered prior to performing directed evolution experiments.

1.2.3 Repurposing Enzymes and Expanding their Reaction Profile

The repurposing of an enzyme to perform a new reaction is an active avenue of research. Biocatalysts that can be used to mimic man-made artificial reactions that have no precedent in nature can be utilised in challenging transformations.[26] Haem-binding enzymes, largely from the cytochrome P450 family, but also other haem binding enzymes such as myoglobin, have been applied for abiotic reactions. P450 enzymes are ubiquitous in nature and allow the activation of oxygen to perform oxidation reactions upon usually unreactive organic scaffolds (**Figure 6**).[27] It was realised that iron at the centre of the haem can also be used for formation of nitrenes and carbenes transformed from diazo precursors. In the case of nitrene activation direct amination[28, 29] and aziridation[30] can be performed and carbene species allow for C-H insertion,[31] strained carbocycle synthesis[32] and direct alkyl fluorination.[33] Tuning of these enzymes for desired substrates and reactivity can be accomplished by directed evolution resulting in high efficiency and accomplished regio- and stereoselectivity.[24, 34]

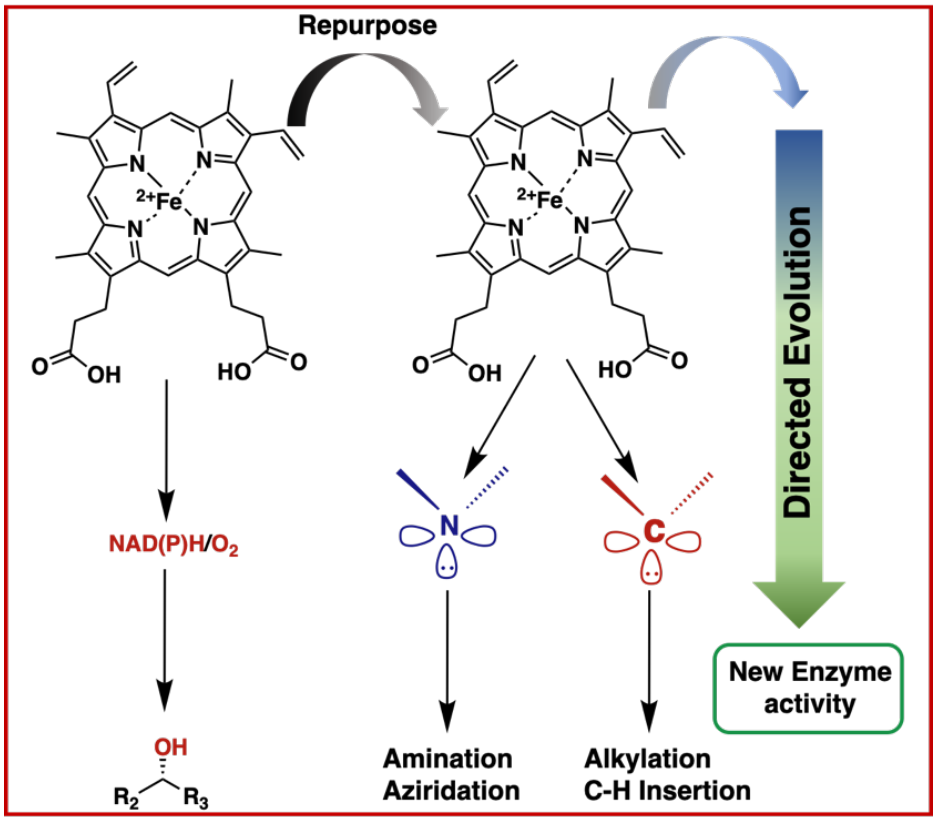


Figure 6. The repurposing of haem bearing enzymes for nitrile and carbene chemistry. Following selection, the enzymes are improved through directed evolution to generate new enzyme activities.

Activation of C-H bonds is considered a challenging area of chemistry. This abundant bond in organic molecules is unreactive and alkylation at sp^3 carbons in nature is also limited. In conventional catalysis these reactions are often accomplished using rhodium, iridium and other transition metals. Iron has also been applied, but reaction temperatures are as high as 80 °C, require stochiometric amounts of the catalyst or are restricted in their versatility. It was envisaged that an iron-containing enzyme could be used to supplant chemical methods.[35] 78 haem bearing proteins were tested for the modification of a methoxybenzyl methyl ester at a sp^3 carbon using ethyl diazoacetate. A variant of a P450 enzyme from *Bacillus megaterium* was reported to be active. The variant harboured a Cys-to-Ser mutation on the axial haem (turning it into a cytochrome P411) plus 18 other mutations throughout the enzyme. This variant was then evolved and truncated to remove the FAD binding domain to yield a significantly improved biocatalyst. A 140-fold improvement over the original template was produced and the enzyme was able to exhibit excellent stereoselectivity. Twelve benzyl substrates were effectively modified at the usually unreactive carbon, generating new chiral centres. The enzyme was further utilised for linear alkyl substrates containing allylic and propargyl groups (**Figure 7a**). Next, substrates

harbouring amines demonstrated the versatility of this enzyme as did modifying the alkylating group (**Figure 7b**).

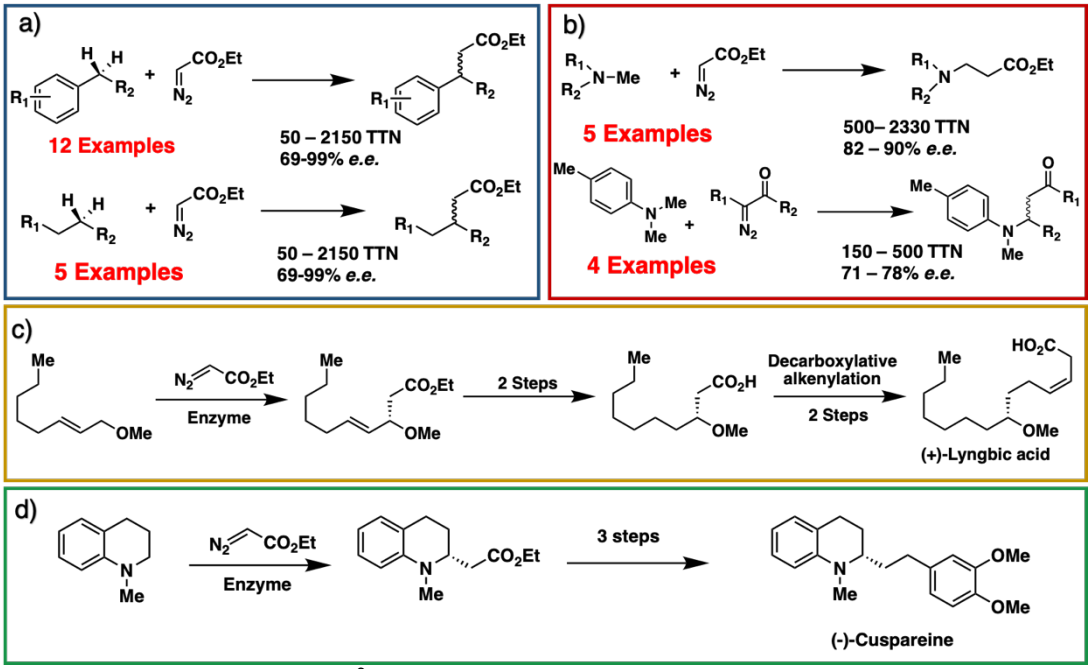


Figure 7. a) Alkylation at sp^3 carbon-hydrogen bonds using a repurposed and evolved cytochrome P450 enzyme. **b)** Substrate scope of the enzyme using amine bearing substrates and different alkyl substituents. **c)** Chemoenzymatic synthesis of (+)-Lyngbic acid utilising the repurposed enzyme to form the chiral centre. **d)** Chemoenzymatic synthesis of (-)-Cuspareine utilising the repurposed enzyme to form the chiral centre.

The enzyme was applied in the chemoenzymatic synthesis of (+)-Lyngbic acid (**Figure 7c**) and (-)-Cuspareine (**Figure 7d**). The realisation that enzymes can be repurposed for desired reactions allows new-to-nature chemically inspired catalysis. The reaction centre of the enzyme is a biologically natural molecule that already holds the mechanistic potential to apply new chemistry and the protein backbone is already optimised for the binding of this cofactor in a beneficial arrangement. The biocatalyst was recruited from a large library of variants already known to have some activity on similar reactions, but the introduction of a wholly new reactive system in a protein requires more extensive designing. One way to achieve this is to merge small molecule catalysts with a suitable protein backbone to form new protein hybrid systems. Here, the potential of chemical catalysts can be merged with enzyme architecture to afford designer biocatalysts that can in theory be targeted for any desired reaction.

1.3 *De Novo* Artificial Enzyme Design

To introduce a new reactive profile into an enzyme, a good starting point is to review small-molecule catalysis or synthetic chemical reactions. It has been proven that orthogonal activation methods and well-characterised organic and inorganic reaction mechanisms can be incorporated into a protein host, forming a hybrid system. Indeed, incorporation of small molecule activation can also improve the biocompatibility of the chemical catalysis, which has underlying importance in the development of chemical and synthetic biology research. In this context, reactions and catalysts that work poorly under physiological conditions (pH 7.4, mostly aqueous solvent and atmospheric temperature and pressure) can be readily applied in biological environments. Compounds of biological origin (e.g. carbohydrates, proteins, nucleic acids) could be used as substrates. In addition, the merger of natural enzyme catalysis into chemoenzymatic cascades requires a mutual environment that will allow for all components in the system to provide optimal efficiency.

1.3.1 Artificial Metalloenzymes

The creation of artificial metalloenzymes (ArMs) has been achieved by exploiting the non-covalent binding affinity of the protein streptavidin with its ligand biotin (**Figure 8a**). The dissociation constant (K_d) of biotin with streptavidin is of an order of magnitude of 10^{-14} mol/L.[36] Biotin can be modified at the carboxylic acid to form an artificial cofactor, which can be recruited into the protein's binding pocket. The flexibility of this approach can be exploited to enable biocompatible catalysis using metal complexes for a vast range of applications. Examples include hydrogenation by ruthenium (Ru),[37, 38] cyclic imine reduction by iridium (Ir)[39] and Suzuki couplings by palladium (Pd).[40] The versatility of this approach is highlighted by the ability to modify the complex or metal whilst maintaining the recruitment mechanism and position in the protein using biotin.[41] Genetic modification can then be used to alter the streptavidin backbone. Crystallography,[37] computational modelling,[42] and directed evolution[43, 44] have all been used to optimise the system. Wild-type streptavidin is tetrameric in nature with four homologous subunits (**Figure 8b**). A dimeric version was engineered allowing insight into key residues in adjacent monomers to the active site (**Figure 8c**).[45] Recently, a monomeric version previously engineered, has been used to host a ruthenium metal complex for the enantioselective synthesis of δ-lactams (**Figure 8d**).[46]

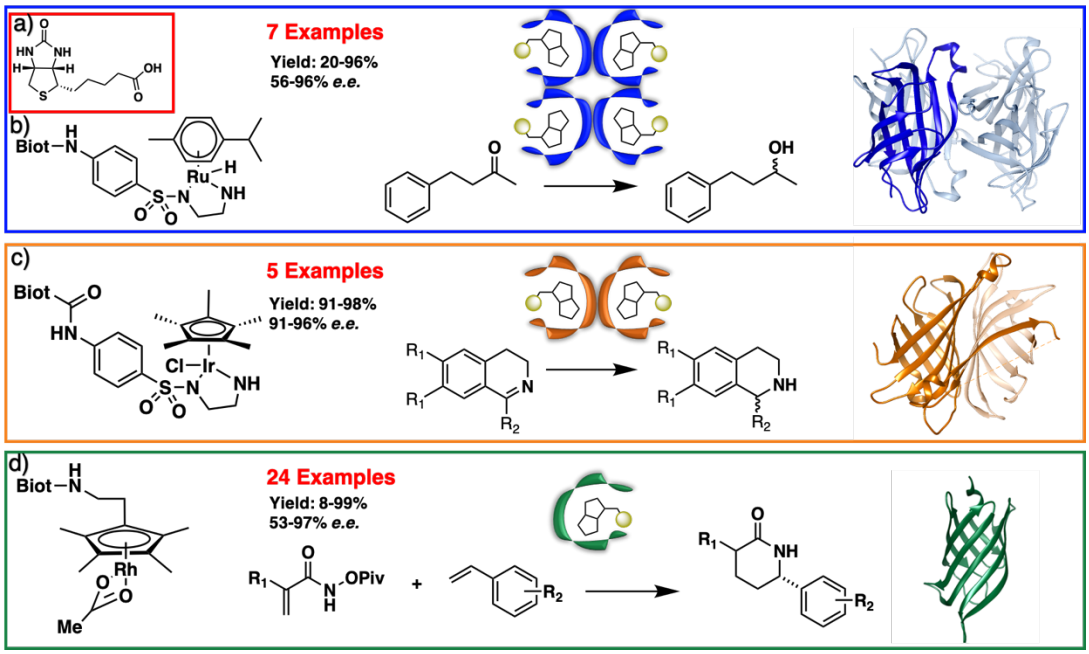


Figure 8. a) The structure of D-Biotin. **b)** A ruthenium (Ru) catalyst complex hosted by tetrameric streptavidin (**PDB:1STP**) for the selective hydration of ketones. **c)** An iridium (Ir) catalyst complex hosted by dimeric streptavidin (**PDB: 6S4Q**) for the reduction of cyclic imines. **d)** A rhodium (Rh) catalyst complex hosted by monomeric streptavidin (**PDB:4JNJ**) for the selective synthesis of δ-lactams.

Use of ArMs based on the streptavidin and biotin technology has extended beyond simple designer biocatalysts. The orthogonality that ArMs can provide has been applied to synthetic biology research. Recently, a ruthenium catalyst hosted by streptavidin was delivered to mammalian cells (HEK293T) by exploiting the tetrameric character of the host protein.[47] The four binding sites were occupied with either the catalyst or a cell penetrating disulfide polymer, conjugated to a fluorescent probe (**Figure 9a** and **9b**). Upon delivery, the designer enzyme was able to perform a bio-orthogonal allylic carbamate cleavage which had been optimised *in vitro* (**Figure 9c**). The substrate **(Figure 9c**), which was supplemented in the media, represses a thyroid responsive hormone gene switch. When present the gene switch is off and the final response which was bioluminescence produced through a luciferase would not express. Upon substrate cleavage, the gene switch is initiated (**Figure 9d**) resulting in downstream expression of the enzyme and hence bioluminescence could be observed. The novelty of this technique was demonstrated by the abiotic reaction that was installed. The catalyst alone could not function under the conditions of the cytosol and there is no known enzyme capable of performing the required transformation, highlighting the importance of combining the two systems into an artificial enzyme.

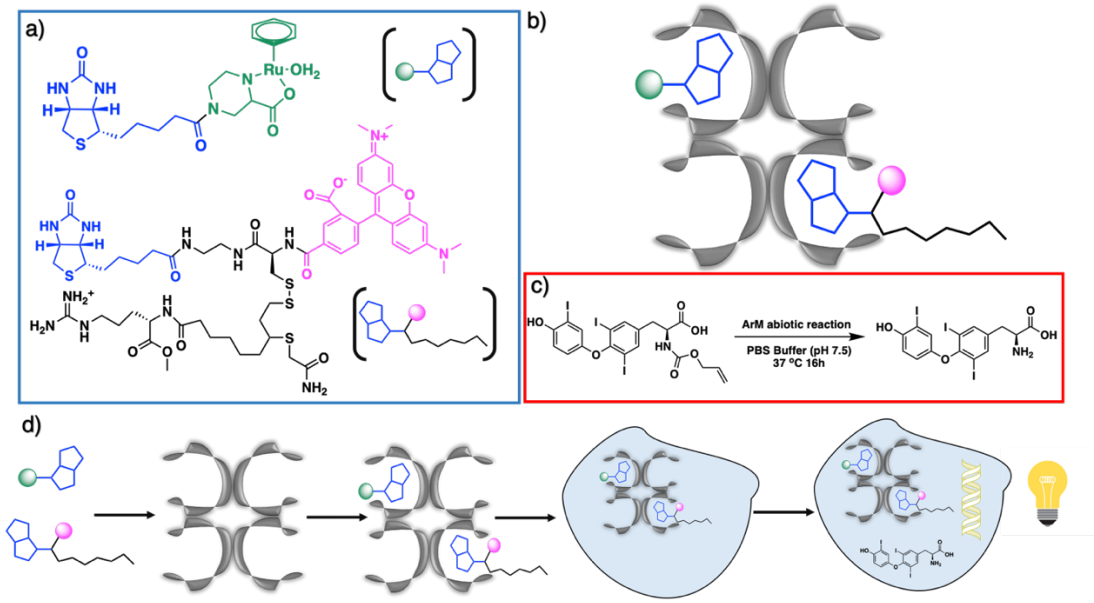


Figure 9. A mammalian gene switch devised by using a purposefully designed artificial metalloenzyme. **a)** Two biotin conjugates one for abiotic catalysis and the second for cell penetration and reporting. **b)** Tetrameric streptavidin with four ligand binding pockets allows dual occupation. **c)** Substrate target which acts as a gene repressor. **d**) Assembly of the ArM followed by delivery to the cell and activation of bioluminescent luciferin.

The streptavidin system has also been applied to the kinetic resolution of prochiral amines. The ArM was coupled with a mono amine oxidase allowing a semisynthetic cascade,[48] again demonstrating the importance of biocompatibility achieved by incorporating the metal into a protein host. Aside from streptavidin and biotin, many other designs have been exploited to engineer artificial metalloenzymes. Adipocyte binding protein was used to bind copper and afford amide and ester hydrolysis[49]. Serum albumins have also been tested, combined with iron and manganese metal complexes to perform oxidation reactions[50]. In addition, periplasmic binding proteins combined with siderophores have been used for the reversible anchoring of a catalytic ligand.[51]

A notable example is the use of the lactococcal multidrug resistant regulator (LmrR) protein isolated from *Lactococcus lactis*. The protein consists of two homologous monomers that non-covalently assemble to form a hydrophobic pore. Small organic molecules can be anchored inside and used for catalysis.[52] Copper complexes have been used for Friedel-Craft alkylations with up to 93% *e.e.*[53], and haem was utilised for cyclopropanations with a 51% *e.e.* (**Figure 10a**).[54] LmrR had also been modified covalently for the introduction of metal cofactors. This can be achieved by introducing cysteine residues within the protein, thereby allowing a bio-conjugation reaction with a copper chelating ligand. Such a Cu^{2+} catalytic system was able to mediate Diels-

Alder cycloaddition reactions (**Figure 10b**).[55] More recently, this ligand was directly inserted into the backbone via genetic code expansion technology, consequently copper was chelated and Friedel-Craft alkylation and hydroxylation reactions could be performed (**Figure 10c**).[56, 57]

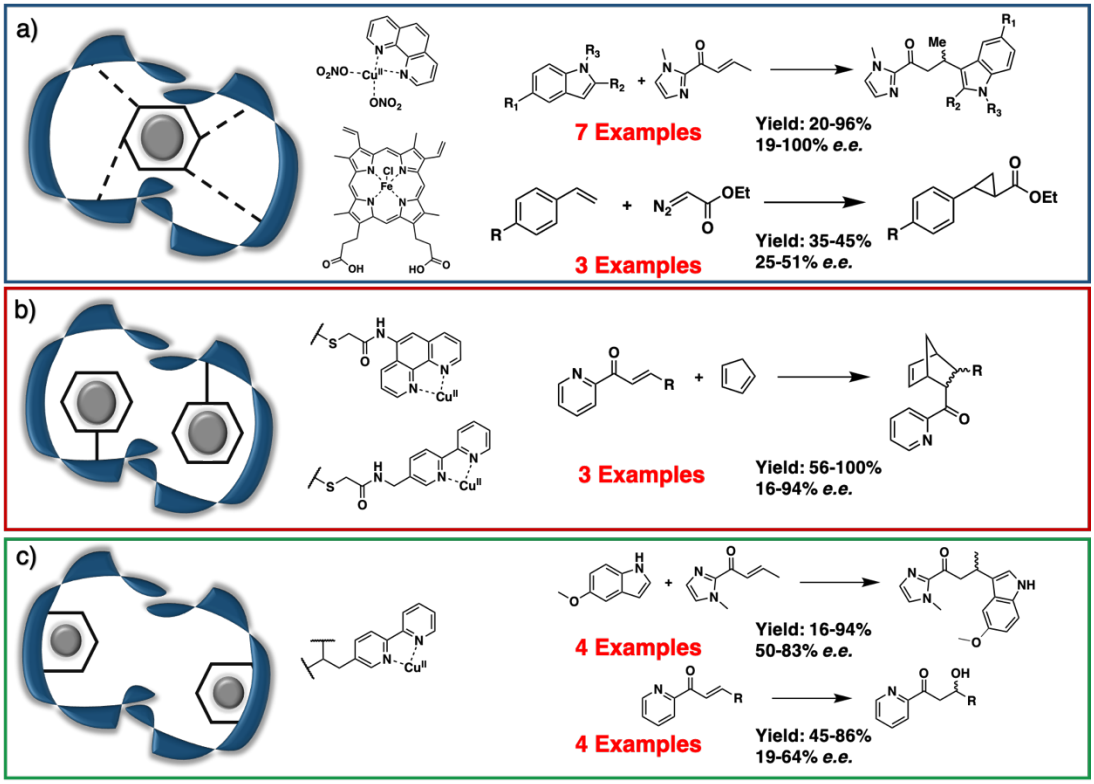


Figure 10. The use of dimeric LmrR for the design of ArMs. **a)** Supramolecular assembly to generate artificial enzymes capable of catalysing Friedel-Craft and cyclopropanation reactions. **b)** Covalent modification to afford Diels-Alder reactions. **c)** Metal binding unnatural amino acid incorporation through genetic code expansion to perform Friedel-Crafts and hydroxylation reactions.

1.3.2 Organocatalytic Artificial Enzymes

Metal free catalysis, such as organocatalysis, has long been an area of interest in the organic chemistry community.[58] However, integration into artificial enzyme design has been far more limited when compared to the use of metals. Organocatalysts can be broadly classified through two activation methods:

- Covalent organocatalysis including enamine, iminium ion (**Figure 11a**) and singly occupied molecular orbital (SOMO) catalysis through the formation of covalent iminium ions or enamine species to allow the functionalisation of carbonyl substrates.[59]

- Non-covalent catalysis such as hydrogen bonding catalysis and pi-anion catalysis such as through (thio)urea-based compounds (**Figure 11b**)[60] or NDI based frameworks respectively.[61]

Organocatalysts benefit from lower toxicity, simple handling, and tolerance to atmospheric pressure and temperature when compared to transition metals.[62] Disadvantages include their generally poor activity in water, high catalyst loading, and extremes of temperature when high selectivity is required. Just as metal catalysis has benefitted from its introduction to protein scaffolds, organocatalysis may also be improved by a similar approach. As in the design of ArMs, the activation methods prevalent to organocatalysts can be transplanted into proteins generating new to nature enzymatic reactions.

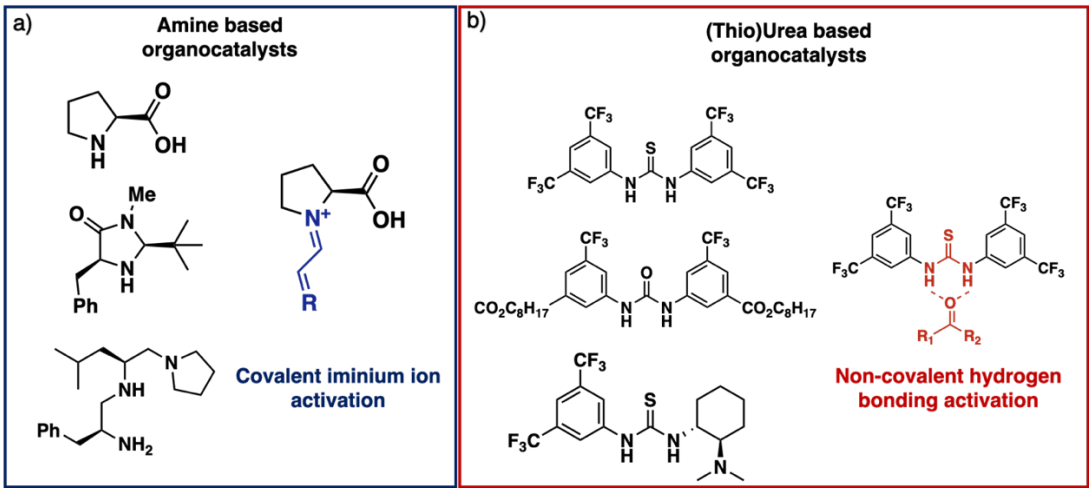


Figure 11. Examples of small molecule organocatalysts. **a)** Amine based organocatalysts that can activate carbonyl compounds through the formation of iminium ion intermediates. **b)** (Thio)urea based organocatalysts that can activate carbonyls through non-covalent hydrogen bonding.

To design and engineer an organocatalytic artificial enzyme (OAE), the approaches used in metalloenzyme design may also be applied, including covalent modification of the protein backbone, computational design using natural residues, introduction of unnatural amino acids and supramolecular non-covalent anchoring **(Figure 12)**. Choice of design and the adoption of a specific catalytic system can be crucial elements in the production of an active hybrid biocatalyst. Herein, I will discuss the different design strategies applied to generate an organocatalytic artificial enzyme with relevant examples to demonstrate the current success in the field.

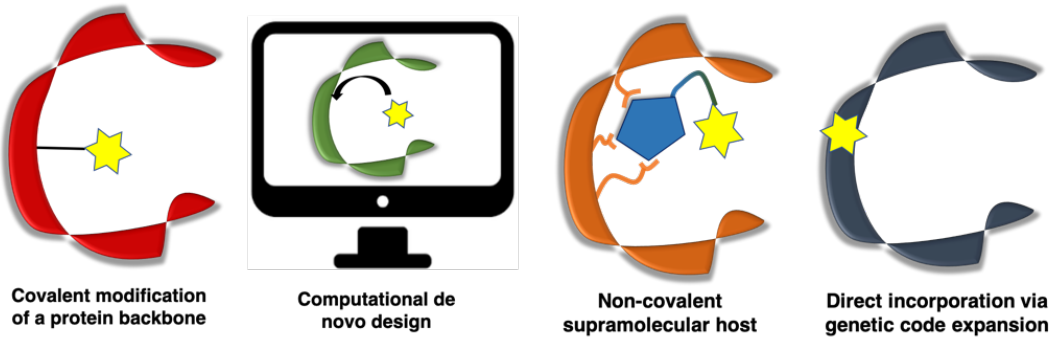


Figure 12. Four methods of designing and generating organocatalytic artificial enzymes.

1.3.3 Organocatalytic Artificial Enzymes - Covalent Modification

Many enzymes recruit cofactors to facilitate their reactions, enabling chemical transformations that could be challenging if only the side chains of the regular 20 proteinogenic amino acids are allowed. Covalent modification of a protein backbone can create a new reactive centre. This reaction centre will be surrounded by the native protein chain. Examples such as covalently linking natural cofactors such as flavins, thiamine or pyridoxamine have proven to be successful. Cysteine, rare in protein chains relative to other amino acids, are an excellent choice for site-specific modification. The thiol side chain can act as a powerful nucleophile, that can react with a large number of reactive groups to form a covalent bond at just one position in a protein chain (**Figure 13**).

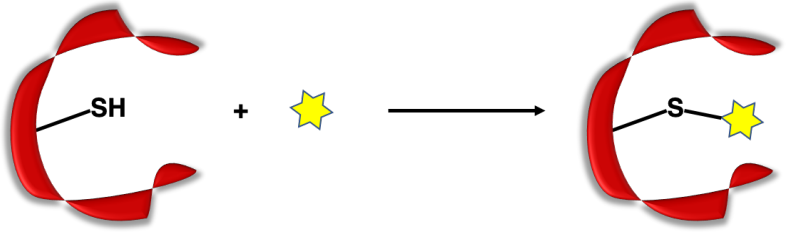


Figure 13. The use of cysteine as a reactive thiol nucleophile to attach a catalytically active molecule site-specifically to a protein backbone.

The natural cofactors flavin, thiamine and pyridoxal phosphate (PLP) are used by natural enzymes to enable reactions that are not possible with the protein backbone alone. These reactive cofactors do not need an enzyme to perform reactions; instead, they can be used as small molecule organic based catalysts. Enzymes are able to utilise the cofactors in a highly specific way to increase the reaction rate and to specifically modify a particular group through either stereo or regio-selectively. Flavins are involved in redox reactions, thiamine in condensation reactions for

carbon-carbon bond forming and PLP for use in a myriad of reactions that require amine, imine and enamine activation. Enzymes that utilise these cofactors have complimentary binding sites that allow for their recruitment in the same way as substrates. The inherent reactions that can be performed can be utilised by other proteins for non-native reactions.

Flavin was attached to the protein papain through its catalytic cysteine. The flavin was modified with a bromo group allowing for nucleophilic substitution to take place. This allowed for the enzyme (naturally a protease) to perform oxidation reactions with BNAH using oxygen as the oxidation source. The OAE was able to enhance the reaction 50-fold when compared with flavin alone (**Figure 14a**).[63] Another example using the same backbone was modification with thiamine. This gave the enzyme the new activity to perform carbon – carbon bond forming reactions.[64] Again, the catalytic cysteine was used as the position of modification (**Figure 14b**). However, the enzyme's activity was low with the cyclisation of 6-oxo heptanal, requiring 6 days, and a large portion of product was identified to be the dimer. Adipocyte binding protein was modified with pyridoxamine to impart it with a catalytic activity. This artificial enzyme was able to perform transamination reactions of keto acids to afford their amino acid equivalents (**Figure 14c**). Good selectivity was also achieved, with some reaching as high as 94% enantiomeric excess.[49]

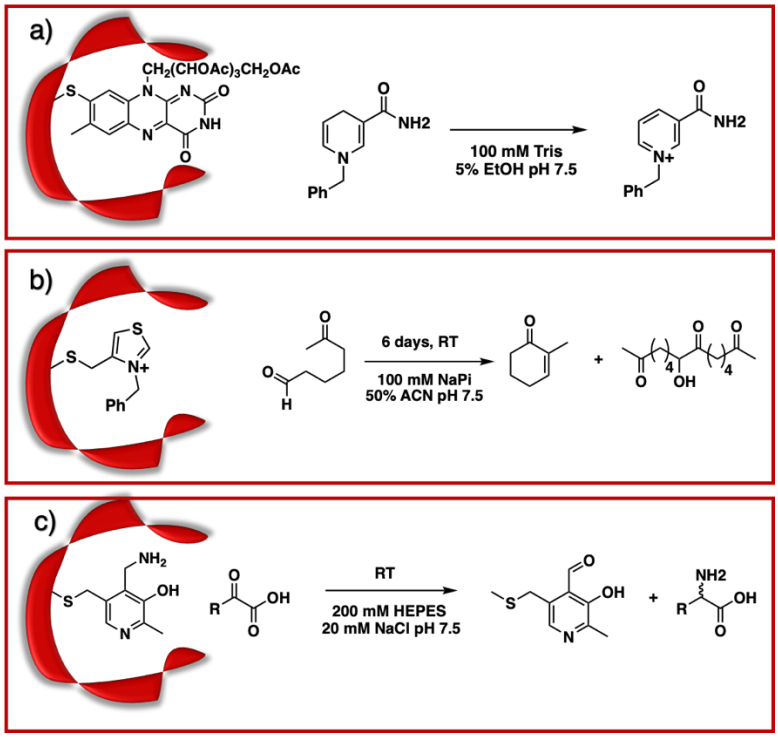


Figure 14. Three examples of covalent modification of proteins with co-factors to impart new reactions to the protein backbones. **a)** Covalent modification of papain with flavin for the oxidation of BNAH. **b)** Covalent modification of papain with thiamine to perform carbon –

carbon bond forming reactions. **c)** Covalent modification of adipocyte binding protein with pyridoxamine for the transamination of keto acids to amino acids.

The use of covalent modification is a fairly simple approach to modify protein function. However, issues arise in the reactivity of the cysteine residue. There is a need to ensure that there is only one cysteine available for modification, and that it is exposed enough to efficiently perform the required ligation. Following this conjugation, removal of the catalysts is then required prior to reactions being performed. In these early systems it was clearly identified that the already catalytic prowess of the reactive cofactors may be a route to creating artificial enzymes that have not evolved for specific substrates. In addition, the advance of bioconjugation chemistry may allow for more strategic labelling strategies in the future, allowing this approach to be further utilised.[65]

1.3.4 Organocatalytic Artificial Enzymes - Computational Design

Computational chemistry has often been a powerful complementary technique to experimental investigations. Protein engineering has benefited widely from computer power. An example is the rendering of homology models produced through related protein families whose structure has already been solved. In addition, small molecule docking and molecular dynamics can be applied to determine the protein geometry and residues that may be crucial to activity. Indeed, software such as PyMOL and Chimera have become pivotal to viewing the 3D structures of proteins derived from NMR and crystallography, allowing the rational modification of active sites.

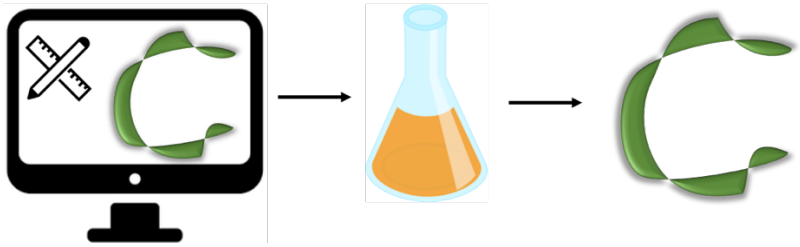


Figure 15. Enzymes and enzyme sections can be computationally designed using relevant software. These designs can then be produced through recombinant expression and tested experimentally.

Computational software, including Rosetta and ORBIT,[66] has accelerated the development of *de novo* enzyme design (**Figure 15**).[67] In the first stage, a "theozyme," where a theoretical arrangement of side chain residues and bioavailable

molecules (water and ions) that can stabilise the rate-limiting transition state(s) of a chosen reaction,[68] is designed. This assembly of a theozyme is subsequently transformed into an experimentally tangible protein structure through evaluations based on calculated parameters (e.g. geometry and energy) by screening of available protein structures available in repositories. Eventually, the best options are recombinantly produced for characterisation. Laboratory evolution can be used to enhance both catalytic activity and stability. This method has led to the formation of highly competent artificial enzymes, some of which have no precedent in nature.

An artificial enzyme for the kemp elimination was produced through computational design.[69] The Kemp elimination requires the deprotonation of carbon and proceeds through a single transition state, which results in the build-up of negative charge (**Figure 16**).[70, 71] To initiate the reaction, abstraction of the proton must be performed by a general base. Two basic residues were considered in the design. A histidine coupled with an aspartate or glutamate as a catalytic dyad can act as a strong base or the direct use of an aspartate or glutamate can be implemented. Additionally, residues that can stabilize the building negative charge were added, along with aromatic residues that can facilitate π-stacking of the substrate and transition state. RosettaMatch was used to fit the active site to protein scaffolds and RosettaDesign to optimise rotamers and geometries, required for this substrate and transition state. In total, 59 designs in 17 different scaffolds were tested experimentally, of those 39 utilised Asp or Glu as a base and the remaining 20 used His-Asp/Glu dyad.

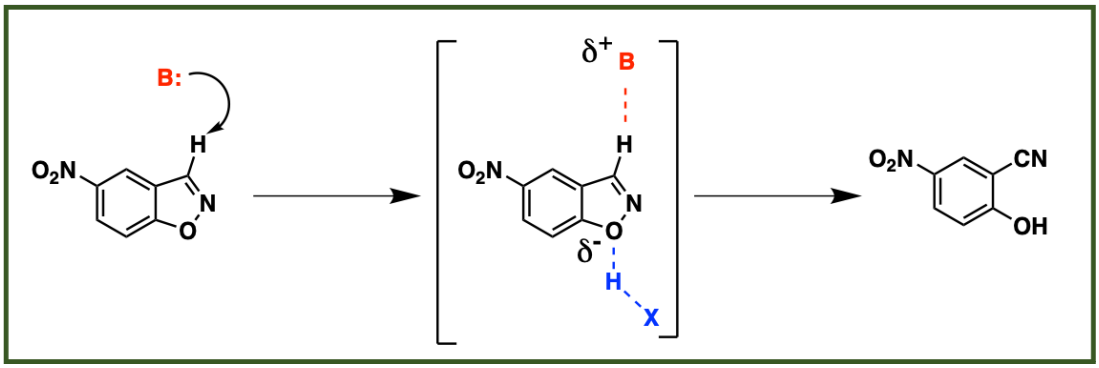


Figure 16. The Kemp elimination reaction. Abstraction of the proton by a base, results in ring opening. The transition state results in negative charge build up at the phenolic oxygen.

The most successful design was an enzyme that utilised a glutamate as the Bronstead base to deprotonate the carbon. This design was inserted into a protein TIM barrel backbone. A tyrosine residue aided in stabilising the transition state through π-stacking interactions. A second successful design was determined to have

a His-Asp dyad. Mutation of the catalytic basic residues abolished the activity of the enzymes. The k_{cat}/K_M of the two enzymes were 163 and 78 $M^{-1} \cdot s^{-1}$ respectively. The activity is impressive considering there is no known enzyme that can perform a Kemp elimination. A third design which utilised a glutamate as the general base was subjected to directed evolution. The reported k_{cat}/K_M for the original template was 12 $M^{-1} \cdot s^{-1}$. A k_{cat}/K_M of 2590 $M^{-1} \cdot s^{-1}$ was achieved following seven rounds of evolution. Residues that were targets in the optimisation included those in the original design and those adopted from the protein scaffold. Interestingly, all of the residues in the original design were conserved in the evolved enzyme suggesting that directed evolution could be used to fine tune the residues and geometry of computationally designed active sites.[72, 73]

The Rosetta software suite has also been used to design an artificial Diels-Alderase.[74]. The substrates 4-carboxybenzyl *trans*-1,3-butadiene-1-carbamate (the diene) and *N,N*-dimethylacrylamide (the dienophile) were used as the model substrates. Understanding of the transition state was pivotal in the enzyme design (**Figure 17a**). Two aspects of the reaction trajectory were considered. Firstly, it was reasoned that the reaction proceeds due to the interaction between the highest occupied molecular orbital (HOMO) of the diene, with lowest unoccupied molecular orbital (LUMO) of the dienophile. Thus, narrowing the energy gap between these two orbitals should, in theory increase the reaction rate. To accomplish this, the designers elected to have a hydrogen bond acceptor interact with the N-H of the diene (stabilising build-up of the positive charge in the transition state) and a hydrogen bond donor to interact with the amide oxygen of the dienophile (stabilising the negative charge build up in the transition state). Secondly, the orientation of the two substrates in the correct geometry was thought to be pivotal to rate increase through entropic reduction (**Figure 17b**).

For the hydrogen bond acceptor, the carbonyl group of a glutamine or asparagine was chosen and for the donor, the hydroxyl group of a serine, threonine or tyrosine. RosettaMatch was used to optimise binding of the substrates in a favourable geometry in 207 protein scaffolds taking into consideration steric clashes and residues available for the aforementioned hydrogen bonding system. 10^{19} theoretical active sites were produced which were then matched to stable protein structures, narrowing the possibilities to 10^6. RosettaDesign was used to maximise transition state binding, catalytic geometry and favourable electrostatic interactions.

Optimisation resulted in 84 possible designs, of which two, displayed activity for the Diels-Alder cycloaddition. The first design was based on a phosphatase with a β-propeller structure that contained 13 mutations compared to the wild-type enzyme (**Figure 17d**). The second was a ketosteroid isomerase that held 14 mutations compared with the wild-type protein.

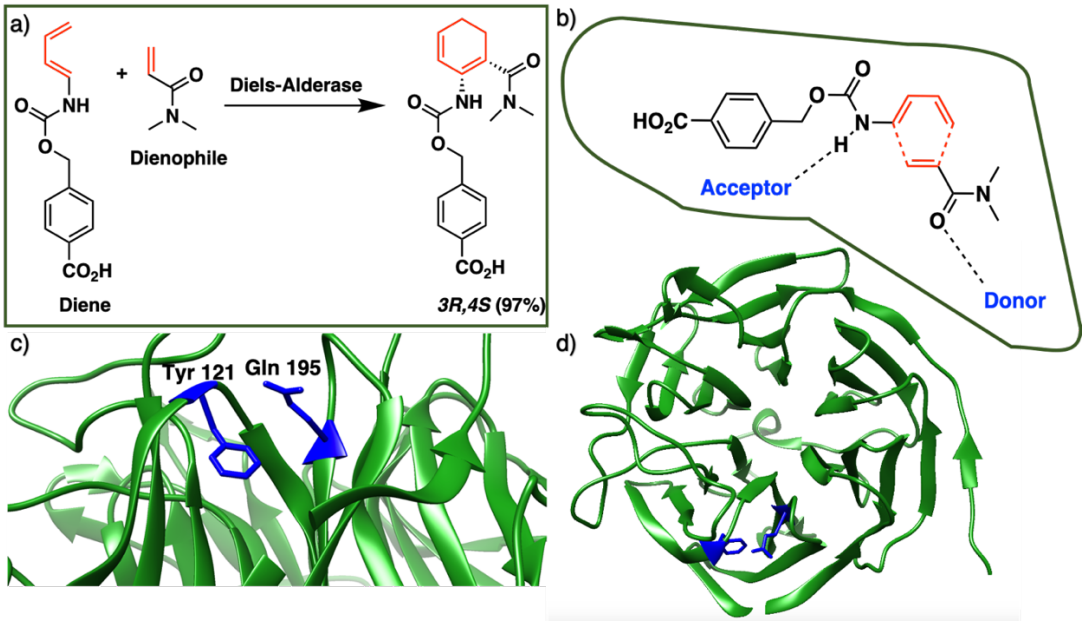


Figure 17. a) The Diels-Alder cycloaddition reaction between the target substrates of the enzyme design. **b)** Design considerations for including hydrogen bond acceptor and donor to narrow the energy gap between the LUMO and HOMO of the substrates by charge stabilisation and desired geometric positioning to reduce entropic contributions. **c)** Tyrosine 121 was shown to be the hydrogen bond donor and glutamine 195 the hydrogen bond acceptor. **d)** Crystal structure of the artificial Diels-Alderase resulting from the computational design based on a β-propeller phosphatase scaffold (**PDB: 3I1C**).

To increase the activity of the artificial enzymes, further mutations that surround the active site were performed. Six mutations in the phosphatase resulted in a 100-fold increase in activity and four mutations in the isomerase demonstrated a 20-fold increase in turnover. Gln195 was considered to be the hydrogen bond acceptor in the phosphatase scaffold and when mutated to a glutamate, it resulted in a 450-fold decrease in enzyme activity. Correspondingly, the residue reported to act as the hydrogen bond donor, which in this case was Tyr121, was mutated to a phenylalanine leading to a 27-fold reduction in activity (**Figure 17c**). Mutational studies support the designer's hypothesis, that narrowing the orbital energy between the LUMO and HOMO by hydrogen bond interactions would increase the rate of the reaction. The optimised design exhibited a k_{cat}/K_M of 8.1 x 10^{-3} $M^{-1} \cdot s^{-1}$ and a selectivity that was 97% in favour of the 3*R*,4*S* product. The substrate scope of the enzyme was poor, however

in this instance, this was in keeping with the intention of the designers, for substrate specificity. In addition to the design of enzyme active sites, the protein backbone was modified through computational means improving the K_M of both substrates.[75] This was largely down to the modification of loop regions of the enzyme that were computationally predicted using Foldit. The resulting computational design was shown to have high similarity with the elucidated crystal structure. Selectivity, specificity, and rate enhancement based on fundamental knowledge of the targeted reaction mechanism clearly demonstrates the power that computational *de novo* enzyme design holds.

A similar design strategy was used for the engineering of an artificial enzyme that can perform the Morita-Bayliss-Hillman reaction.[76] Inspiration for this enzyme came from peptide and small molecule catalysts. The reaction allows carbon-carbon bond formation through formation of two intermediates. Intermediate 1 is formed between a nucleophile and an enone, followed by formation of Intermediate 2 through attack of an aldehyde on Intermediate 1. A proton shuffle results in release of the product and recycling of the catalyst (**Figure 18a**). The theozyme was based on five criteria: (1). A cysteine or histidine residue to act as a nucleophile; (2). Two hydrogen donors to stabilise the first intermediate; (3). Hydrogen bond donor for intermediate 2 stabilisation; (4). A hydrophobic binding pocket for cyclohexanone; and (5). A hydrophobic binding pocket with an aromatic interaction for benzaldehyde (**Figure 18**).

The theozyme included at least 10 residues which would include the initial nucleophile required for the nucleophilic attack on the enone. Either a cysteine activated by a lysine or a histidine activated by either an aspartate or glutamate were considered. 48 designs were determined to be experimentally viable and were expressed and investigated, with two active “hits” identified. An enzyme designated BH25 contained a nucleophilic cysteine based on an alanine racemase (**Figure 18b**), and the second dubbed BH32, exhibited a histidine nucleophile supported by a glutamate, built into the structure of a haloacid dehalogenase (**Figure 18c**). Mutation of the nucleophiles to alanine abolished their activity. Each enzyme was improved by three further mutations. Nucleophilic attack on the enone was believed to be the rate limiting step. To identify improved variants from the BH25/32 parent designs, a fluorescent labelling reaction based on conjugation to Tamra dye was developed. Variants derived from BH25 and BH32 provided yields of 9% and 24% over 28 hours respectively.

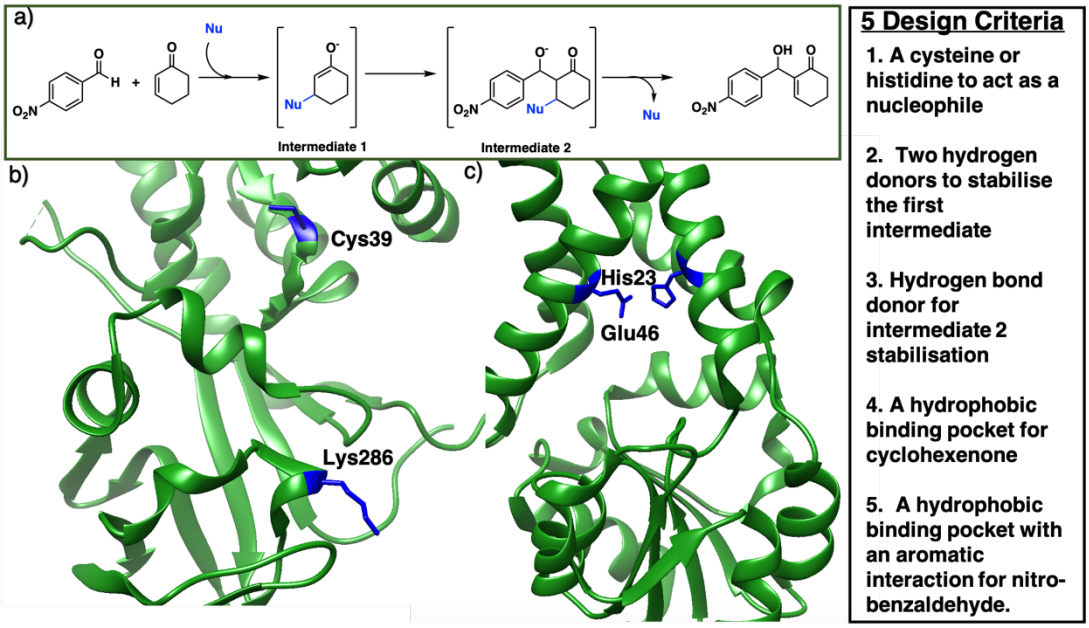


Figure 18. **a)** The Morita-Bayliss-Hillman reaction that proceeds through two intermediates. **b)** Crystal structure of BH25 with the catalytic dyad Cys39 and Lys286 highlighted in blue **PDB: 3UW6**). **c)** Crystal structure of BH32 with the catalytic dyad His23 and Glu46 highlighted in blue (**PDB: 3U26**).

Retro-aldolases are a class of *de novo* designed enzymes capable of catalysing retro-aldol reactions *via* formation of an iminium intermediate.[77] Retro-aldolases have been created from a theozyme that is able to mediate cleavage of the fluorogenic compound methodol (**Figure 19c**). The reaction was selected to allow for facile screening as the retro-aldol product naphthaldehyde is fluorescent. The most effective theozyme in terms of rate enhancement in the recombinantly produced protein contains a catalytically active lysine residue within a hydrophobic binding pocket and a strategically positioned water molecule, that was suggested to mediate formation of the Schiff base intermediate. Interestingly, this designed network was found to be catalytically more active than those made based on naturally found proton shuffle networks. Computational tools, such as RosettaMatch, were recruited to dock the theozyme into a protein scaffold, creating a suitable host for the artificial active site.

Indole-3-glycerol phosphate synthase, a TIM-barrel protein fold, was identified for hosting the theozyme (**Figure 19a**).[78] Further adjustment of the residues surrounding the transition state was made using RosettaDesign.[79] Among these active models,

the variant RA95.0 with the catalytically active lysine at position 210 (**Figure 19b**) with an apparent pK_a of 8.1 was identified as the most promising candidate. Experimentally, RA95.0 was able to mediate cleavage of methodol with catalytic efficiency (k_{cat}/K_M) of 0.19 $M^{-1} \cdot s^{-1}$ and selectivity for *S* over *R* (2:3:1).[80] To enhance the activity, regions at and around the active site were subjected to iterative cassette mutagenesis, a form of saturation mutagenesis where pre-synthesised short oligonucleotides are inserted through modern cloning techniques. The resulting variant showed 73-fold increase in catalytic efficiency when compared to RA 95.0 with 3:1 *R*-*S* selectivity. Upon structural analysis by crystallography, a second reaction centre was revealed. Mutation of Tyr83Lys enabled formation of an iminium intermediate in addition to the original Lys210. It was suggested that this mutation shifted the pK_a of the original Lys210 to 7.6, which resulted in the improved performance.[81]

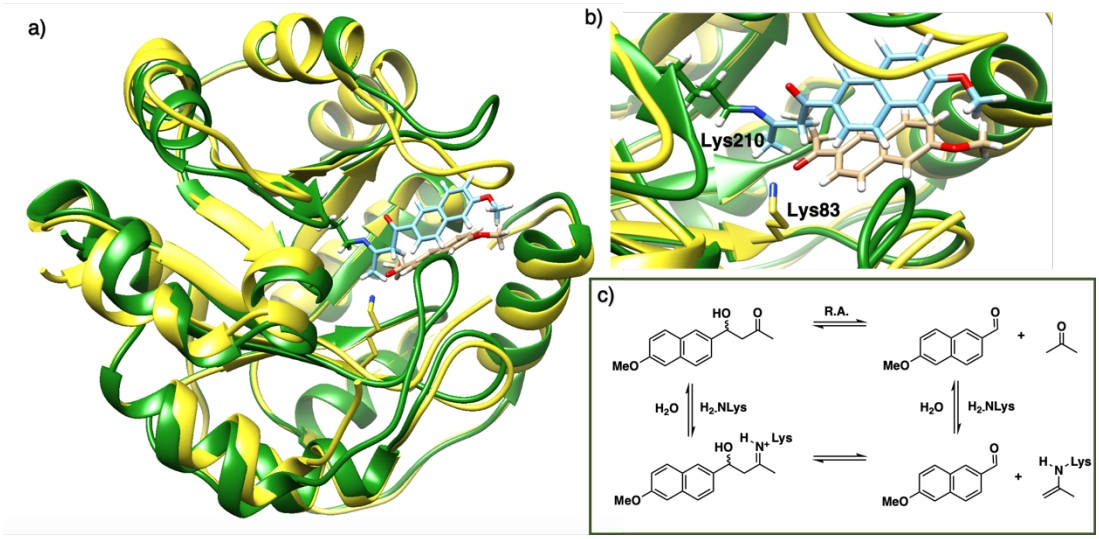


Figure 19. a) Crystal structure of the computationally designed retro-aldolase RA 95.0 (green) (**PDB: 4A29**) and RA 95.5 (yellow) (**PDB: 5AN7**) **b)** Positions of the catalytic Lys210 (RA 95.0) and Lys83 (RA 95.5) bound to ketone inhibitors. **c)** Proposed catalytic cycle for the retro-aldol reaction of methodol.

Additional laboratory evolution of the entire gene through error-prone PCR and DNA shuffling created the variant RA 95.5 (**Figure 19a**). The variant had an additional six mutations and demonstrated Significantly improved activity (>20-fold, k_{cat}/K_M 320 $M^{-1} \cdot s^{-1}$, and selectivity 5:1 *R* over *S*). Crystallographic studies illustrated that Lys83 transformed into the only reaction centre for the methodol cleavage (**Figure 19b**), indicating that there is a switch in location of the residue responsible for catalysis. Restructuring of the active site was likely unpredictable during the initial design, highlighting that randomness is a key element during the evolution of an efficient

enzyme. Lastly, a final three rounds of laboratory evolution yielded the variant RA95.5-8, which contains substitutions at both the active site and distal positions, and its catalytic efficiency (k_{cat}/K_M) was measured to be 850 $M^{-1} \cdot s^{-1}$.

Following successful engineering of the artificial retro-aldolase, the enzyme was employed in multiple reactions. Considering the key step in the methodol cleavage was formation of a Schiff base it was realised that the designer biocatalyst can be applied to iminium and enamine catalysis. Reactions catalysed included conjugate additions[82-84] and Knoevenagel[85] and Henry condensations (**Figure 20**).[86] From the design stages, to optimisation by directed evolution and finally for application to multiple reactions, the design strategy demonstrated by the successful engineering of these *de novo* retro aldolases has paved the way for using computational design to generate ad hoc biocatalysts with wide reaction profiles and expanded substrate scopes.

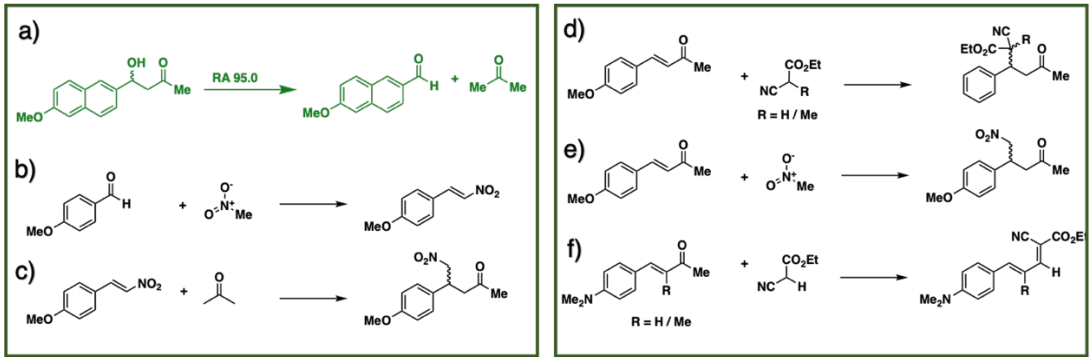


Figure 20. Reaction scope of the RA95 family of computationally designed retro-aldolases. **a)** The original retro aldol reaction used for initial screening, the cleavage of the starting product (methodol) results in a change in fluorescence**. b)** Henry addition of nitromethane to aromatic aldehydes. **c)** conjugate addition of acetone to nitrostyrene**.** Enamine catalysed conjugate addition. **d)** Conjugate addition of carbon nucleophiles. **e)** Conjugate of nitromethane to α,β-unsaturated aldehydes. **f)** Knoevenagel condensations of carbon nucleophiles with α,β-unsaturated aromatic aldehydes.

Computational chemistry has become a powerful tool for the design and production of metal free artificial enzymes.[87] The above examples have shown the variety of mechanisms that can be exploited, from cycloadditions, to acid base catalysis, and reactions facilitated by enamine and iminium ions and demonstrates the versatility of computational *de novo* design. As computing power increases, it is likely that this design strategy will aid in both creating artificial enzymes with diverse reaction profiles and improving our knowledge of protein-based catalysis. Machine learning and artificial intelligence is a new and exciting technology.[88, 89] Machine learning has recently been applied to aid in the directed evolution of proteins by aiding in the choice

of target residues that have a higher probability of improving the desired transformation.[90] Algorithms that can analyse large sequence spaces of proteins can reduce the time taken for a scientist to do the same in the lab. It is reasonable to assume that machine learning will be applied to *de novo* enzyme design in the very near future.

1.3.5 Organocatalytic Artificial Enzymes - Supramolecular Ligand-Host Interaction

Streptavidin and biotin have been extensively applied to artificial metalloenzyme design. The streptavidin system has also applied to metal free catalysis. Using the same methodology to modify the biotin ligand, organocatalytic moieties can be installed into the streptavidin backbone. Versatility of this technique is highlighted by the ability to screen multiple catalysts quickly through simple chemical modification of the ligand (**Figure 21**).

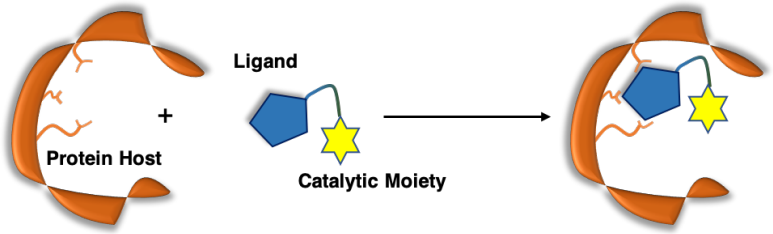


Figure 21. Cartoon representation of the components for non-covalent tethering that consists of a protein host, a tight binding ligand and a catalytic moiety.

Anion–π-catalysis has become a contemporary topic in organocatalysis.[61, 91] During catalysis, anion intermediates that form during the reaction can be stabilised by π-acidic molecules such as naphthalenediimides (NDI), which possess a positive quadrupole moment. This consequently facilitates organic transformations such as conjugate additions. However, all-natural aromatic amino acids are π-basic and interact with cations. Hence, the authors introduced the π-acidic NDI moieties into streptavidin via biotin conjugation. Subsequently, decarboxylative alkylation reactions between thioester malonates and nitrostyrenes could be achieved (**Figure 22**). Both chemical (modification of the catalysts) and genetic (modification of the protein host) techniques were employed to achieve a competent designer enzyme.[92]

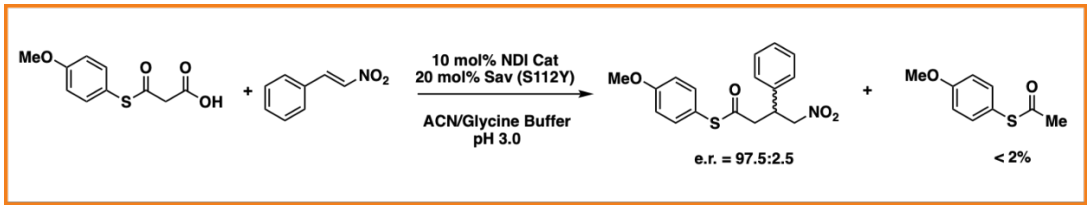


Figure 22. Decarboxylative alkylation reaction performed by the NDI π-anion based streptavidin catalyst.

The design was based on stabilisation of the enolate intermediate by the NDI motif, and tertiary amines included in the catalytic moieties that were able to localise the intermediate in proximity to the NDI. These criteria resulted in five catalysts with different length linkers between the NDI and amine that were screened for activity with tetrameric streptavidin (**Figure 23a**). The most active ligand was then combined with 20 different streptavidin variants. A Ser112Tyr mutation resulted in the highest conversion, showing 90% of substrate transformed into product with an *e.r.* of up to 97.5:2.5. It is interesting to note that the reaction proceeded best at pH 3.0, far away from general enzyme optimal pH of 7.4. Further investigation using site-directed mutagenesis and docking studies suggest that the medium length of the linker between the amine and the NDI avoids steric clashes of the catalysts with the protein. Large electron withdrawing groups on the NDI structure were found to weaken the binding of biotin to its host, whilst a flexible linker such as dimethylene hinders the conversion and selectivity.

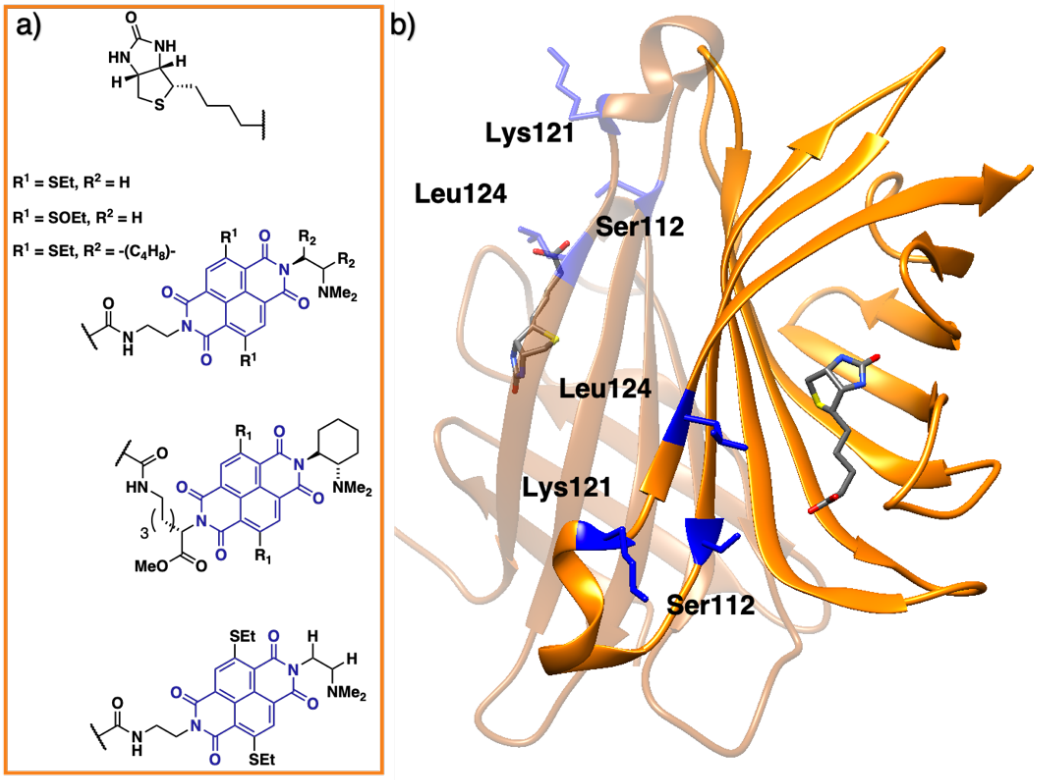


Figure 23. **a)** Five π-anion NDI based catalysts conjugated to biotin. **b)** Crystal structure of two biotin monomers with biotin anchored inside. Residues in blue were crucial for insights into the hybrid catalysts activity (**PDB: 1MK5**).

The organocatalytic artificial enzyme was then applied to domino Michael-aldol reactions (**Figure 24**). Where, yields of approx. 50%, enantioselectivity of 0-80% and diastereoselectivity of >20:1 were obtained.[93]

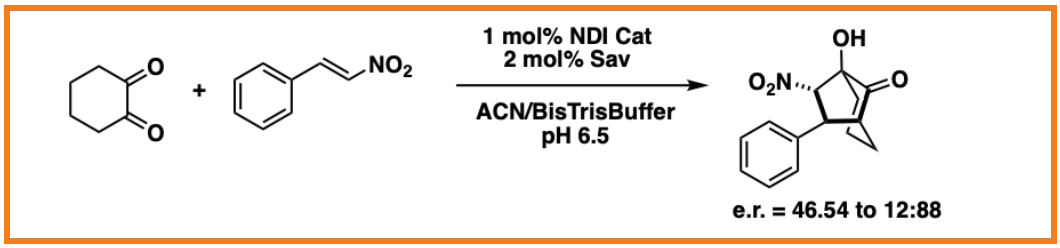


Figure 24. Michael-aldol domino reaction performed by the NDI π-anion based streptavidin catalyst.

The Bayliss-Hillman C-C bond formation reaction has also benefitted from the streptavidin biotin system.[94] Biotin was conjugated to catalytically active moiety 4-(4-aminopipidino) pyridine, known to facilitate this reaction in solution, and anchored inside tetrameric streptavidin. For the model substrates, *p*-nitrobenzaldehyde and cyclopentenone were chosen as model substrates. Preliminary activity showed no observable product beyond that seen in a catalyst free control, and so the hybrid

catalyst was crystallised, and the structure solved. Two residues, Gln114 and Arg121, were seen to be sterically hindering the catalyst. Mutation to alanine along with Ser112, which lay under the catalyst, resulted in no improvement. The designers then decided to conserve the serine residue to force the catalyst forward to the mouth of the binding pocket and also to introduce bulkier amino acids such as isoleucine, phenylalanine and methionine at this position. Activity above the background reaction was achieved by the Ser112Ile mutant with a conversion of 35%. However, the reaction lacked any selectivity. This study did not result in a selective or highly efficient catalyst, but it does illustrate the importance that rational genetic modification does have use in improving organocatalytic artificial enzymes, even when no activity was originally observed.

Secondary amine organocatalysis has been applied to create artificial enzymes using the supramolecular approach to facilitate iminium ion and enamine catalysis.[95, 96] This will be discussed below. Currently only the streptavidin biotin system has been utilised in designing supramolecular based organocatalytic artificial enzymes. Other strong protein-ligand non-covalent interactions may be exploited in the future.

1.3.6 Organocatalytic Artificial Enzymes - Genetic Code Expansion

Genetic code expansion is the introduction of amino acids beyond the 20 proteinogenic into a protein backbone by ribosomal translation.[97] Incorporation of unnatural amino acids offers the opportunity to expand the chemistry that is currently possible by the 20 standard amino acid sidechains. Genetic code expansion enables site-specific incorporation of unnatural amino acids, which can be used to mediate bioorthogonal chemical reactions. To achieve this goal, an orthogonal aminoacyl-tRNA synthetase/tRNA pair is needed.[98] Specifically, the orthogonal tRNA decodes a blank codon, commonly the amber stop codon (TAG) as it is often the least used codon in most organisms. Expression of recombinant proteins that contain unnatural amino acids in *E. coli* can be achieved using pyrrolysyl-tRNA synthetase/tRNA and tyrosyl-tRNA synthetase/tRNA pairs from archaea. The pyrrolysyl-tRNA synthetase/tRNA pair is particularly convenient as it naturally decodes the amber codon, this least used stop codon in *E. coli*.[99]

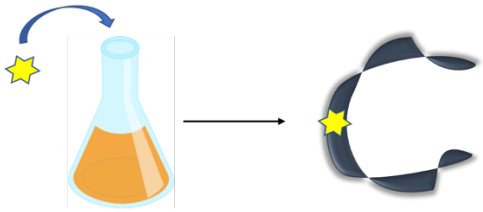


Figure 25. Addition of the unnatural amino acid to the culture medium allows incorporation into the protein backbone by an orthogonal tRNA synthetase/tRNA

Incorporation of the unnatural amino acid requires a stop codon mutation at a desired position in the gene of interest. Double transformation of *E. coli* with plasmids containing the gene of interest and the synthetase are conducted. By including the unnatural amino acid in the medium, the orthogonal synthetase specifically charges the orthogonal tRNA with the unnatural amino acid, which will allow for production of full-length protein with the unnatural amino acid incorporated into the protein backbone (**Figure 25**). To date, over 200 unnatural amino acids can be genetically incorporated into a protein of interest using this technique.[100] Thus there exists a vast opportunity to exploit these unnatural amino acids for metal free transformations.[101] Once incorporated, the unnatural amino acid becomes the centre of the active site. Contemporary synthetic biology techniques can then be applied, such as directed evolution, to tune the environment around the amino acid for a given reaction. Unlike other design methodologies, genetic code expansion allows the position of the active catalyst to be varied throughout the protein backbone, allowing larger creativity with orientation and protein environment.

Lactococcus multidrug binding protein (LmrR), which was previously used in the design of artificial metalloenzymes, was the first scaffold to be used to host a metal free artificial biocatalytic reaction. Four residues that had been explored in the aforementioned design (Val15, Asp19, Met89 and Phe93, **Figure 26a** and **b**). were targeted for substitution by the unnatural amino acid *p*-aminophenylalanine. Incorporation was attempted under the action of an evolved tyrosyl-tRNA synthetase from *Methanococcus jannaschii*.[102] However the incorporation was deemed unsuccessful on this protein, despite success elsewhere. A derivative with an azido group in the para position was then chosen and subsequently reduced to the catalytically active aniline *in vitro*, using Tris(2-carboxyethyl)phosphene (TCEP **Figure 26c**). The designer enzyme was then tested for hydrazone and oxime ligation. It was found that unnatural amino acid replacement at the Val15 position yielded the most promising result (**Figure 26d**).

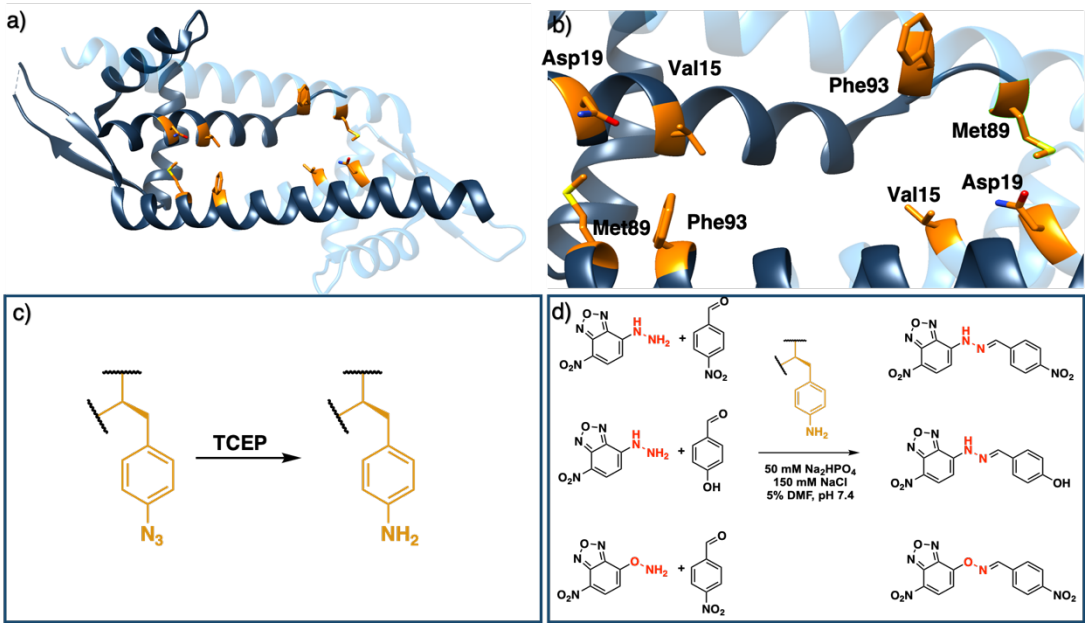


Figure 26. a) Crystal Structure of LmrR (**PDB: 3F8F**). **b)** Close up of the dimer interface. The residues in orange were those targeted for mutation. **c)** Reduction of *p*-azidophenylalanine to *p*-aminophenylalanine with TCEP. **d)** Hydrazone and Oxime ligation performed by the unnatural amino acid *p*-aminophenylalanine at position V15.

Hydrazone formation between 4-hydrazino-7-nitro-2,1,3-benzoxadiazole (HNB) and *p*-nitro benzaldehyde was used as the model reaction. This reaction is known to proceed without catalysts, although slowly, but the rate can be enhanced by aniline. In addition, the wild-type LmrR was able to catalyse the reaction, likely due to hydrophobic pore that can recruit organic molecules, aided by tryptophan residues lining this pocket. The introduction of the unnatural amino acid at Val15 increased the reaction yield from 46% (wild-type LmrR) to 72%. The transient iminium ion that forms as the reaction proceeds was trapped using $NaCNBH_4$, and LC-MS confirmed formation of this ion at that position. Oxime formation was then investigated using the hydroxy equivalent of HNB. Again, the introduction of the unnatural amino acid increased the rate of reaction.

Laboratory evolution was used to screen the library variants in 96 well plates by measuring the loss of the UV absorbance from the substrate. The resulting variant, which carries additional mutations including Ala11Leu, Asp19Met, Ala92Arg and Phe93His, showed a 74-fold increase in catalytic efficiency (**Figure 27**).[103] Based on the knowledge of these positions from previous structures, Leu11 and Met19 are thought to help position the aniline in a more “reaction-ready” position. Furthermore, Arg92 was reasoned to stabilise the build-up of negative charge that appears during the condensation of the aniline with the carbonyl group. Lastly, His93 was proposed

to serve as a proton shuttle assisting in the formation of iminium ion intermediates and promoting the transamination processes.

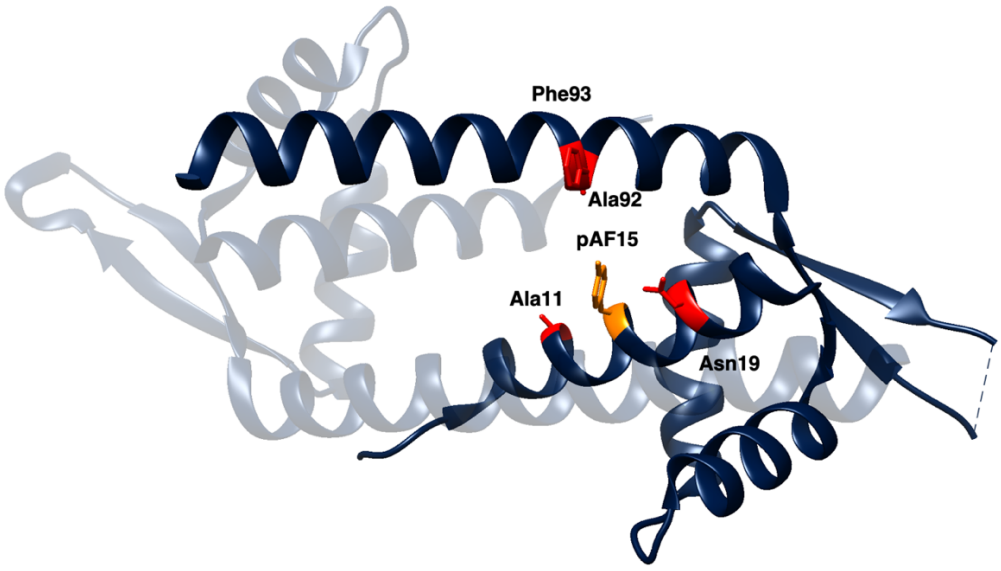


Figure 27. Crystal structure of LmrR with the unnatural amino acid *p*-aminophenylalanine at position 15 (orange) (**PDB: 6I8N**). Red residues that surround the active site were targeted for directed evolution.

Recently, the *p*-aminophenylalanine/LmrR system has been further modified for a novel dual substrate activation strategy. Through combination with a supramolecularly bound Lewis acidic Cu(II) complex, the resulting artificial enzyme was able to mediate a Michael reaction that involves both formation of a Cu-enolate and an organocatalytic iminium intermediate. Yields of this novel reaction mode were up to 90%, with d.r. and *e.e.* up to 9:1 and >99% respectively. This work highlights that importance of developing different approaches to artificial enzyme design (e.g. genetic code expansion and supramolecular assembly), as proteins can be used to host multiple catalytic centres for coupled reaction cascades.[104]

A second design based on the protein backbone BH32, an enzyme originally created by Rosetta to perform the Morita–Baylis–Hillman reaction, has been further re-engineered into a potent hydrolase through the combined use of genetic code expansion and laboratory evolution (**Figure 28a**). Substitution of the catalytic His23 with methyl-histidine was achieved by using an evolved variant of the pyrrolysyl-tRNA synthetase and its cognate tRNA (**Figure 28c**). The resulting enzyme was dubbed OA1.1 and was able to perform ester hydrolysis for a range of compounds that fluoresce upon reaction. Screening for variants with improved activity was performed using 96 well plates on a plate reader where formation of the fluorescein product

could be monitored (**Figure 28b**). Six mutations resulted in a 15-fold increase in enzyme activity (known as OA1.3). Mutations resulting from the evolution were Lys10Pro, Ala19His, Ser22Met, Glu46Asp, Pro63Gly and Asp125Gly. Based on the data derived from crystallography and kinetic investigations, the authors concluded that the aromatic ester formed between the substrate and Me-His was significantly more prone to hydrolysis. In contrast, the neutral acyl enzyme intermediate formed from the natural amino acid histidine hydrolyses slowly under the same condition (**Figure 29d**).[103]

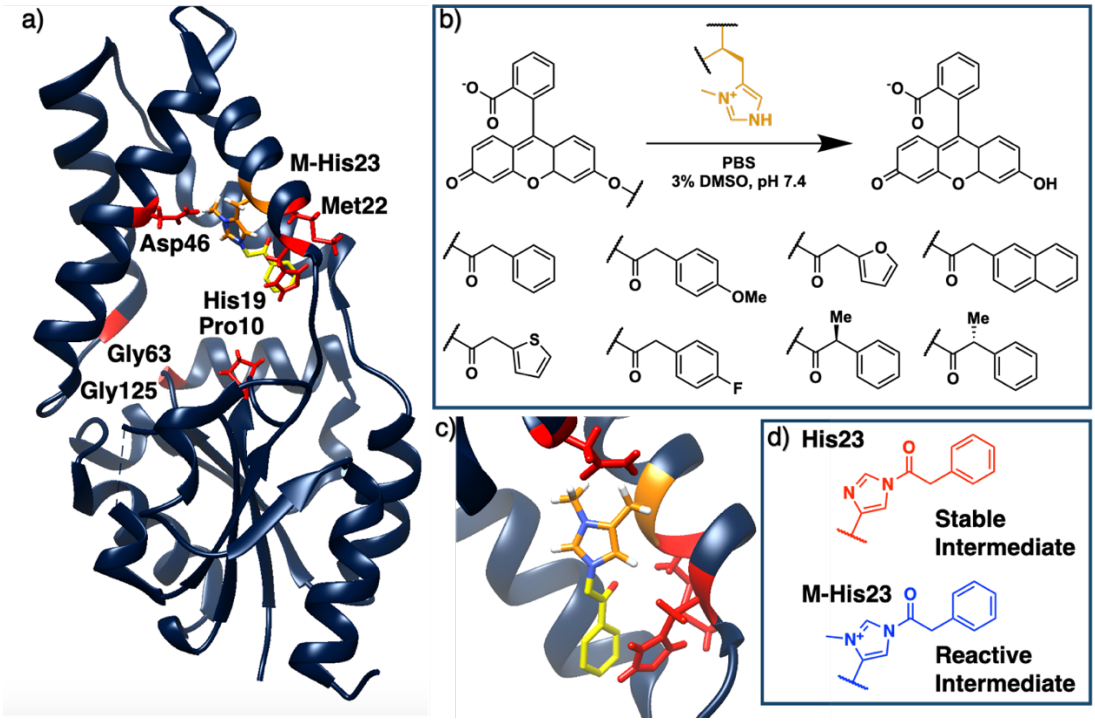


Figure 28. a) Crystal structure of OA1.3 derived from the protein backbone BH32. Residues in red are those resulting from directed evolution (**PDB: 6Q7R**). **b)**. Fluorescein ester hydrolysis catalysed by methyl-histidine. **c)** Unnatural amino acid methyl-histidine bound to the inhibitor bromoacetophenone. **d)** Stable intermediate formed by His23 and reactive intermediate formed by M-His23.

Genetic code expansion for the design and engineering of artificial enzymes has already demonstrated powerful results. Advantages of this method include the ability to transplant a novel catalytic moiety from one back bone to another and to explore many positions in said backbone. In addition, this method allows for genetic modification of the backbone in a simple and efficient way without having to modify any vital interactions to keep the catalyst in place. Evolution of the tRNA synthetases utilised in genetic code expansion technology is now commonplace and robust strategies are continually increasing the catalogue of non-canonical amino acids available for incorporation.

1.4 Cyclic Secondary Amine Artificial Enzymes

1.4.1 Organocatalysis with Cyclic Secondary Amines

Secondary amine catalysis is a branch of organocatalysis. Here, small molecular, sub stoichiometric and metal-free compounds are used to catalyse chemical transformations. Secondary cyclic amines have become popular catalysts in many enantioselective transformations that involve carbonyls as their substrates. Proline (**Figure 29a**) was one of the first such compounds shown to catalyse chemical transformation (the Hajos-Parrish intramolecular aldol reaction). Following this early report, more versatile catalysts were designed and tested. Imidazolidinone catalysts (**Figure 29b**) and the Jorgensen-Hayashi (**Figure 29c**) catalysts are some of the most prevalent. These catalysts have been applied in stereoselective reactions, such as conjugate additions,[105, 106] aldol reactions, α-fluorinations,[107] transfer hydrogenation[108] and the Diels-Alder reaction.[109]

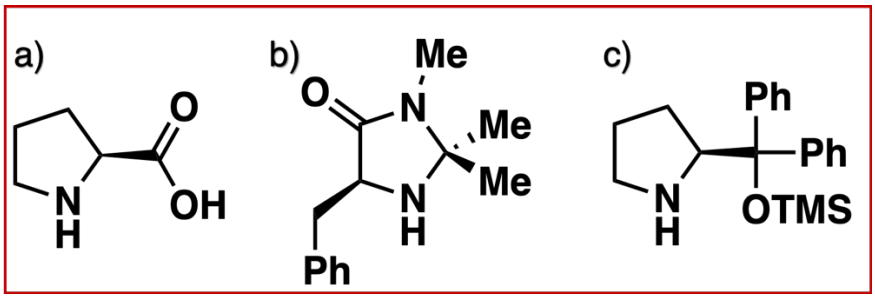


Figure 29. Three common cyclic secondary amine organocatalysts. **a)** L-proline. **b)** First generation imidazolidinone (MacMillan) catalyst. **c)** Jorgensen-Hayashi catalyst.

The mechanism involves activation of the carbonyl and formation of an intermediate, either an iminium ion or an enamine (**Figure 30**). In the case of iminium ion catalysis, this can facilitate nucleophilic addition to the β-carbon of α,β-unsaturated carbonyls (**Figure 30a**). The formation of an enamine can allow electrophilic addition to the α-carbon of a carbonyl (**Figure 30c**). Formation of an iminium ion lowers the energy of the lowest unoccupied molecular orbital (LUMO) relative to the parent carbonyl, thus bringing its energy closer to that of the highest occupied molecular orbital (HOMO) of the attacking nucleophile (**Figure 30b**). The higher electrophilicity of the iminium ion species when compared to that of the carbonyl group facilitates the energy shift. In contrast, enamine catalysis results in the raising of the energy of the HOMO, due to the lower electronegativity of the enamine species relative to the carbonate ion. The

energy difference of the enamine HOMO and the accepting electrophiles LUMO is now closer, allowing the reaction to proceed (**Figure 30d**).

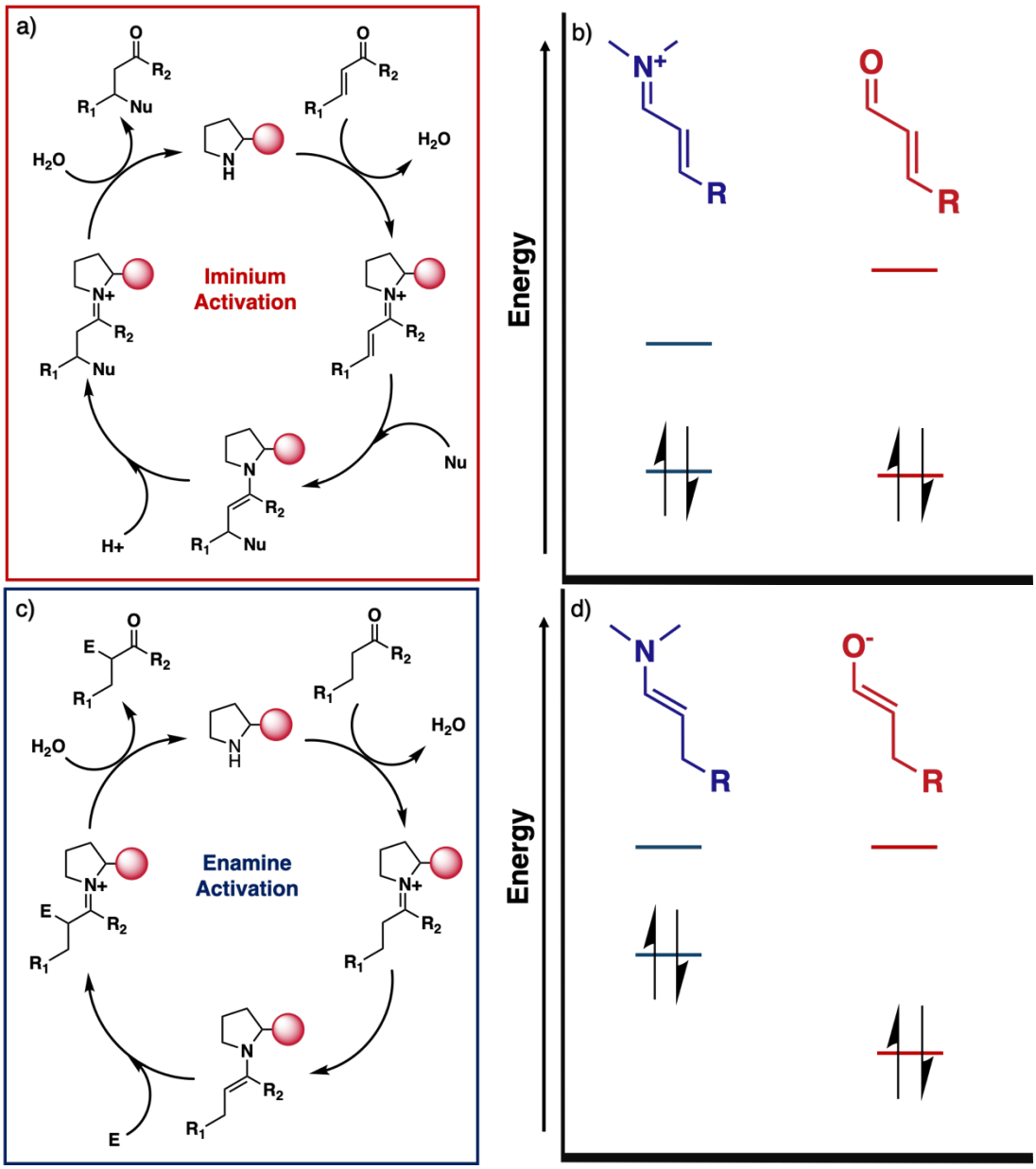


Figure 30. a) Iminium activation by cyclic secondary amines. **b)** Diagram representing the reduction of the energy of the LUMO due to the iminium species with respect to the parent carbonyl ion. **c)** Enamine activation by cyclic secondary amines. **d)** Diagram representing the increase in energy of the HOMO of the enamine species with respect to the parent carbonyl.

Due to the reduced toxicity when compared to metal catalysts it is tempting to utilise secondary amines in biologically compatible reactions. However, the poor activity of the catalysts under aqueous conditions hinders application in biological environments. The myriad of reactions allowed could potentially permit a bottom-up design approach, whereby a desired reaction or set substrates can be the basis of the system. The lack of biocompatibility has been investigated previously and

systems such as supramolecular scaffolds,[110] hydrogels,[111] and emulsions have been attempted,[112] however the use of proteins to overcome such issues is yet to be fully exploited.

The integration of versatile organocatalytic mechanisms of secondary amines and protein scaffolds have previously been explored. Two examples are prominent in the literature. The first is the repurposing of a natural enzyme that bears an N-terminal proline and was subsequently evolved through directed evolution. The second takes advantage of the streptavidin-biotin interaction through the supramolecular host method.

1.4.2 4-Oxalocrotonate Tautomerase

When located at the N-terminus of a protein, proline offers a secondary amine that can be used for iminium ion and enamine based organocatalysis. One such example is 4-oxalocrotonate tautomerase (4-OT) from *M. putida*, which is composed of six homologous monomers carrying a catalytic N-terminal proline (**Figure 31a)**. Naturally, this residue acts as a general base, catalysing the tautomerisation of a dienol into an unsaturated ketone (**Figure 31b**).

As mentioned above, proline alone can catalyse reactions through enamine and iminium ion catalysis. The N-terminal proline residue of 4-OT can form similar iminium ion and enamine intermediates with various carbonyl substrates. Because of its significant substrate promiscuity, 4-OT has been used as an organocatalyst for chemical transformations. It has been demonstrated that 4-OT is able to catalyse enamine-based aldol reactions and conjugate additions (**Figure 32a, b and c**). Additionally, 4-OT has been exploited for iminium ion catalysis, including the conjugate addition of nitromethane (**Figure 33d**). Reduction of the intermediate iminium ion by sodium cyanoborohydride and subsequent mass spectrometry analysis provide evidence that supports the formation of the iminium intermediate.

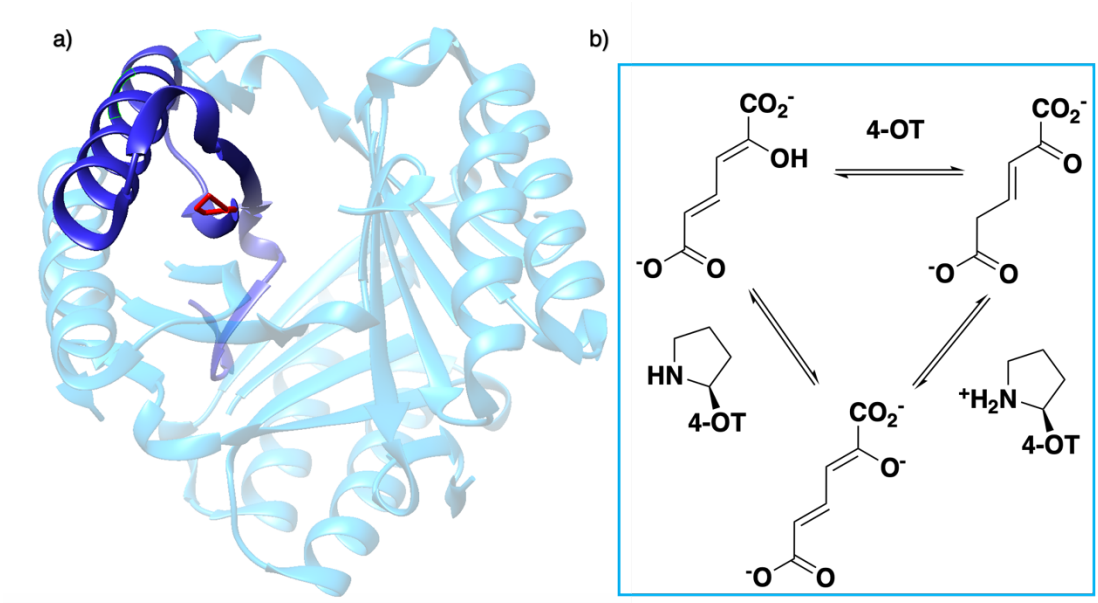


Figure 31. a) Crystal structure of 4-OT with one monomer highlighted in dark blue. **b**) The enzyme's native reaction, a base catalysed tautomerisation.

Mutagenesis via a combined computational and experimental approach has led to the identification of enhanced variants. Three residues in proximity were found to be crucial for catalysis, including Phe50, Met45 and Ala33. Mutability landscapes were used to determine 'residue hotspots.' The experiment consisted of singly mutating all amino acids with the exception of the catalytic Pro1. Protein solubility of single point mutations was first assessed, followed by an activity screen of the tautomerisation reaction and subsequently the Michael addition. A Phe50Ala mutation resulted in an increase of catalytic efficiency by a factor of 600 for cross-coupling aldol reactions. In contrast, when both Phe50 and Met45 were replaced with valine and tyrosine respectively, the resulting variant was more effective at self-condensation reactions. The Phe50Val/Met45Tyr double mutant resulted predominantly in the *R* product, whereas a third mutant Ala33Asp selectively yielded the *S* enantiomer in the conjugate additions of acetaldehyde to β-nitrostyrenes. Crystal structures of the two mutants have been obtained, but the N-terminal region was not resolved, possibly due to its inherent flexibility. Hence, the actual assembly in the active site remains unclear.

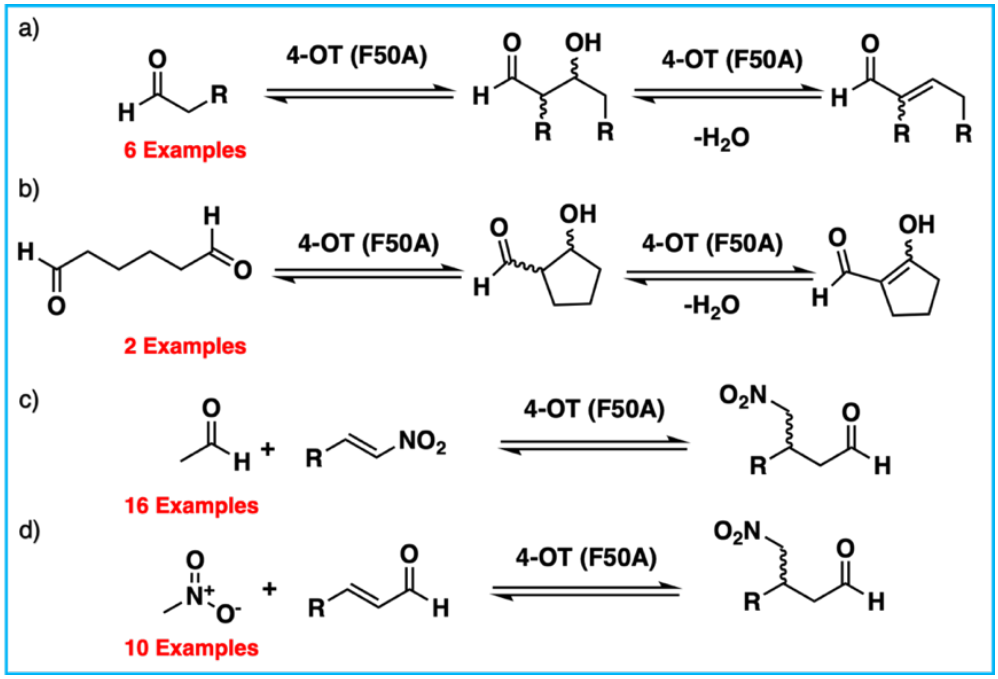


Figure 32. Reactions catalysed by 4-OT tautomerase. **a)** Intermolecular aldol reaction. **b)** Intramolecular aldol reaction. **c)** Enamine catalysed conjugate addition. **d)** Nitro-Michael addition.

4-OT and its variants have been used for a range of applications including enzymatic and chemoenzymatic cascades. The anti-anxiety drug pregabalin and three of its analogues were synthesised by coupling the 4-OT reaction with catalysis by aldehyde dehydrogenase (**Figure 33**). Acetaldehyde was added stereoselectively to α,β-unsaturated nitro substrates under the action of a 4-OT variant, followed by oxidation by ALDH to yield the corresponding carboxylic acid. To recycle NADH, a cofactor recycling system operated by NAD oxidase (NOX) was included. Lastly, the nitro group was reduced to the amine using sodium borohydride in the presence of nickel chloride. These applications present evidence that protein-based organocatalysis can be used in combined synthesis, which may not be readily achievable using traditional organocatalytic systems.

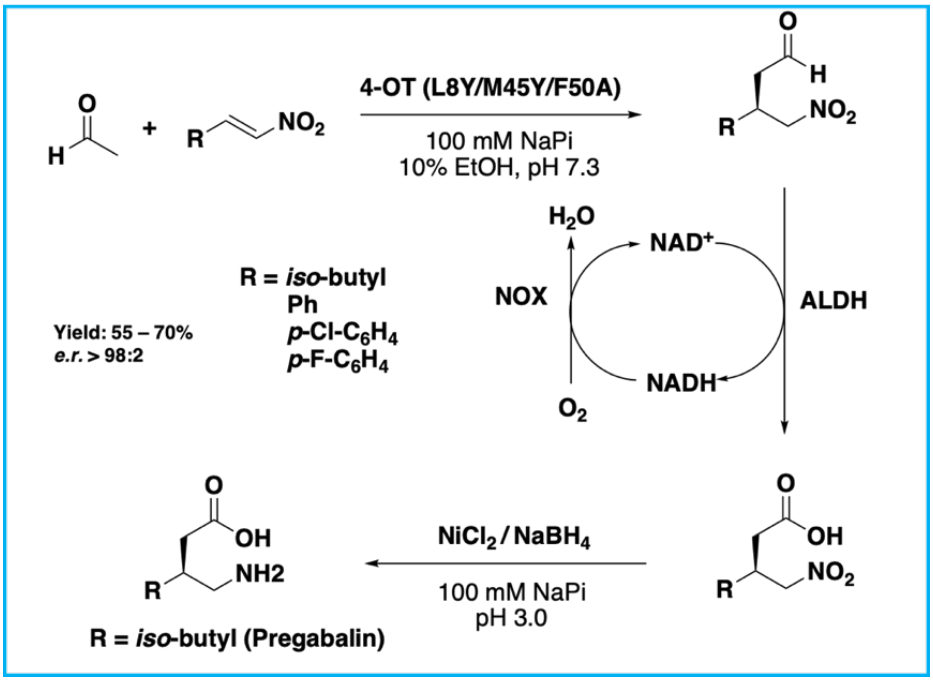


Figure 33. Chemoenzymatic synthesis of Pregabalin and three analogues.

Utilising only natural residues, with no chemical modification needed, the N-terminal proline approach is arguably the simplest in establishing a biocompatible organocatalytic system. As a range of reactions have already been established, 4-OT is an attractive system for performing organic reactions in biological contexts. However, the sequence space of 4-OT is relatively small when compared to many proteins. Each monomer consists of only 63 amino acids, reducing the applicability of extensive engineering. In addition, the catalytic residue of Pro1 cannot be moved, as it will only function as an organocatalyst when the amine of the proline is free, and this is only possible when at the N-terminus.

1.4.3 Streptavidin and Biotin-Secondary Amine Moieties

Recently, a secondary amine artificial enzyme was designed and applied using the streptavidin–biotin technology. Seven biotinylated secondary amines were prepared via either copper-catalysed azide–alkyne cycloaddition or amide bond coupling reactions (**Figure 35a**). These catalysts can be broadly segregated into three types: imidazolidinones, prolines and pyrrolidines. Their ability to catalyse the Michael addition of nitromethane to aromatic α,β-unsaturated aldehydes was investigated (**Figure 35c**).

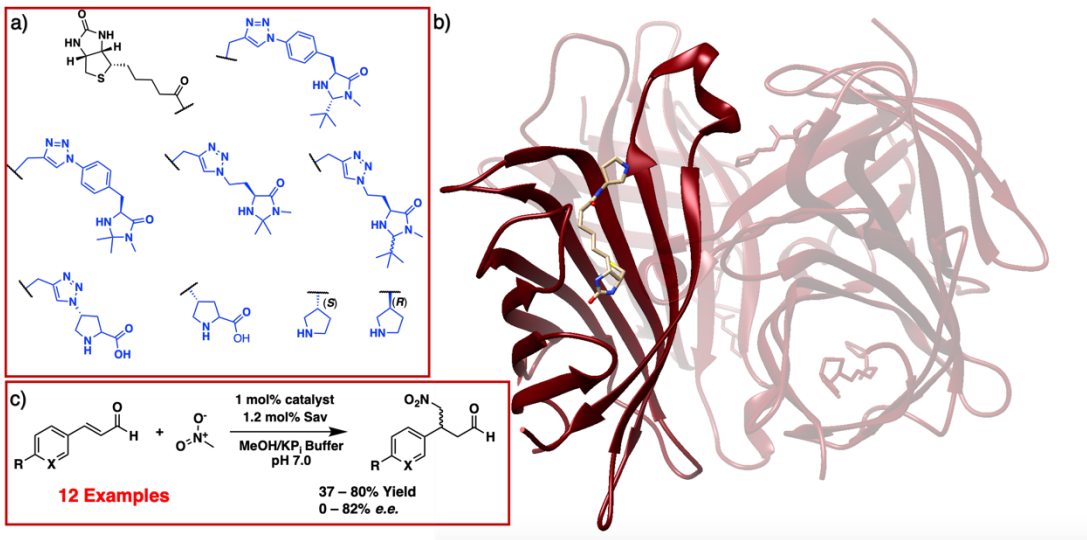


Figure 34. a) The structure of D-biotin and eight secondary amine organocatalysts that were tested. **b)** Tetrameric structure of streptavidin (**PDB: 6GH7**) with biotin pyrrolidine catalyst (*R*) bound in the hydrophobic pocket **c)** The nitro-Michael addition reaction between nitromethane and a range of α,β-unsaturated aldehydes.

It was shown that the two biotin catalysts *R* and *S* pyrrolidines were able to facilitate the reaction alone, however they were not able to induce enantioselectivity. When introduced to the tetrameric streptavidin, they were found to be able to mediate the model reaction with high reactivity and stereoselectivity. Moderate to good yields (37–80%) were obtained using only 1 mol% of protein catalyst when the reaction was performed in 1:1 MeOH/PBS. Despite the fact the two catalysts differed by only one chiral centre, their stereoselectivity was opposite, with the former favouring for the *S* enantiomer and latter for the *R* enantiomer.

Again, it has been shown that introduction of a secondary amine into a protein backbone is able to catalyse a known (and new to nature) reaction under aqueous conditions. However, to increase the yield of the reaction 50% organic co-solvent had to be added. Later, the same artificial enzyme was used to perform an enamine catalysed aldol reaction between nitro-benzaldehyde and acetone, which gave an excellent conversion of 93%. However, a poor *e.e.* of only 34% was obtained.[95]

Both designs have shown that it is possible to utilise secondary cyclic amine organocatalyst mechanisms and activations in protein scaffolds. It is also clear that the reactions can proceed under aqueous conditions and also that the chiral protein scaffold is able to induce enantioselectivity. These are impressive feats, however both systems suffer from the same flaws: both 4-OT and the streptavidin system are limited in their design by the placement of the catalyst and their choice or protein host.

Directed evolution can be used to modify the chiral environment however the overall fold of the proteins, i.e. secondary, tertiary and quaternary structure, are not able to be fully modified. Derivatives of streptavidin have been produced, but the overall protein structure and much of the hydrophobic pore are essential for binding of biotin in the backbone. Similarly, 4-OT lacks design flexibility. Here the backbone is reliant on always having a proline at the N-terminus. It was discovered through mutagenesis landscapes of 4-OT that many variants were insoluble let alone active. In addition, 4-OT has no other binding sites that may allow for the recruitment of co-factors or secondary reaction reagents.

Clearly there is a need to expand the design of integrating secondary amine catalysis with proteins to create hybrid artificial enzymes. The advantages have already been demonstrated by the applications of 4-OT in enzymatic and chemoenzymatic syntheses. As described above protein chemists have explored differing ways to introduce new reactivities into proteins and so it is prudent to pursue other approaches to expand the field of artificial enzymes.

1.5 Aims

Previously, the combination of the distinct catalytic systems of enzyme catalysis and secondary amine organocatalysis into one hybrid artificial system have focussed on two designs. The first was the re-purposing of 4-OT-tautomerase and the second the exploitation of non-covalent tethering of a catalytic moiety into streptavidin through the supramolecular approach. Unlike previous reports the aim here is to exploit two different design methodologies that offer improved flexibility to create new artificial enzymes, that utilise the reactivity of secondary amines and the biocompatibility of proteins. The aims of this thesis can be characterised into three areas:

1. The covalent modification of a protein backbone with organocatalytic moieties. By exploiting the natural metabolism of penicillin binding proteins with carbapenems to afford a nitro-Michael addition of nitromethane to α,β-unsaturated aldehydes under biocompatible conditions. Carbapenems are broad substrates for a vast number of penicillin binding proteins and β-lactamases, allowing a large selection of protein backbones to be exploited using the natural reaction of the host. Three carbapenems that bear secondary amines in their structure will be tested with two protein backbones for the aforementioned reaction. The use of the carbapenems as catalysts alone will be explored under aqueous conditions. Conversion of substrate to product and enantioselectivity will be assessed. In addition, the labelling of the catalytic moieties to the protein backbones will be characterised.

2. Investigation of the use of genetic code expansion to incorporate cyclic secondary amines directly into a protein backbone, thereby catalysing a biomimetic hydride transfer reaction. Three secondary amine unnatural amino acids will be placed into various positions in the multidrug resistant binding protein LmrR. A synthetic hydride donor will be used to selectively reduce α,β-unsaturated carbonyl compounds. To demonstrate the flexibility of this method of enzyme design the amino acids will be transplanted into a different protein backbone. In this case EcDHFR a natural oxidoreductase will be used, and the reaction performed using NADPH as the hydride donor. Kinetic and mechanistic characterisation will be performed, and the substrate scope explored. In addition, biocompatibility of the organocatalytic artificial enzyme

will be demonstrated by coupling of the catalyst to a natural enzyme for *in situ* recycling of the enzyme cofactor.

3. One of the design techniques relies heavily on genetic code expansion. To improve protein yield of recombinant proteins with unnatural amino acids, a brief study will be pursued where the N-terminus of *Methanosarcina bakeri* pyrrolysyl-tRNA synthetase will be mutated, to determine its effects on unnatural amino acid incorporation using *E. coli* as the host cells. Multiple C-terminal mutants and amino acids will be explored in addition to differing culture temperatures.

1.6 References

1. Tu, C. K.; Silverman, D. N.; Forsman, C., et al., Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry* **1989,** *28* (19), 7913-7918.
2. Page, M. G., Extended-spectrum beta-lactamases: structure and kinetic mechanism. *Clin Microbiol Infect* **2008,** *14 Suppl 1*, 63-74.
3. Warshel, A.; Sharma, P. K.; Kato, M., et al., Electrostatic basis for enzyme catalysis. *Chem Rev* **2006,** *106* (8), 3210-3235.
4. Luk, L. Y.; Loveridge, E. J.; Allemann, R. K., Protein motions and dynamic effects in enzyme catalysis. *Phys. Chem. Chem. Phys.* **2015,** *17* (46), 30817-27.
5. Meyer, H.-P.; Eichhorn, E.; Hanlon, S., et al., The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC). *Catal. Sci. Technol.* **2013,** *3* (1), 29-40.
6. Truppo, M. D., Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med Chem Lett* **2017,** *8* (5), 476-480.
7. Turner, N. J., Directed evolution of enzymes for applied biocatalysis. *Trends Biotechnol.* **2003,** *21* (11), 474-478.
8. Singh, R.; Kumar, M.; Mittal, A., et al., Microbial enzymes: industrial progress in 21st century. *3 Biotech* **2016,** *6* (2), 174.
9. Trono, D., Recombinant Enzymes in the Food and Pharmaceutical Industries. In *Advances in Enzyme Technology*, Singh, R. S.; Singhania, R. R.; Pandey, A.; Larroche, C., Eds. Elsevier: 2019; pp 349-387.
10. Zaks, A., Industrial biocatalysis. *Curr. Opin. Chem. Biol.* **2001,** *5* (2), 130-136.
11. Li, X. J.; Zheng, R. C.; Ma, H. Y., et al., Engineering of Thermomyces lanuginosus lipase Lip: creation of novel biocatalyst for efficient biosynthesis of chiral intermediate of Pregabalin. *Appl. Microbiol. Biotechnol.* **2014,** *98* (6), 2473-2483.
12. Benaiges, M. D.; Alarcon, M.; Fucinos, P., et al., Recombinant Candida rugosa lipase 2 from Pichia pastoris: immobilization and use as biocatalyst in a stereoselective reaction. *Biotechnol Prog* **2010,** *26* (5), 1252-1258.
13. Siodmiak, T.; Mangelings, D.; Vander Heyden, Y., et al., High enantioselective Novozym 435-catalyzed esterification of (R,S)-flurbiprofen monitored with a chiral stationary phase. *Appl. Biochem. Biotechnol.* **2015,** *175* (5), 2769-2785.
14. Long, Z.-D.; Xu, J.-H.; Zhao, L.-L., et al., Overexpression of Serratia marcescens lipase in Escherichia coli for efficient bioresolution of racemic ketoprofen. *J. Mol. Catal. B: Enzym.* **2007,** *47* (3-4), 105-110.
15. Brannigan, J. A.; Wilkinson, A. J., Protein engineering 20 years on. *Nat Rev Mol Cell Biol* **2002,** *3* (12), 964-970.
16. Bassegoda, A.; Cesarini, S.; Diaz, P., Lipase improvement: goals and strategies. *Comput Struct Biotechnol J* **2012,** *2*, e201209005.
17. Ahmed, S. T.; Parmeggiani, F.; Weise, N. J., et al., Engineered Ammonia Lyases for the Production of Challenging Electron-Rich l-Phenylalanines. *ACS Catal.* **2018,** *8* (4), 3129-3132.
18. Gillam, E. M., Engineering cytochrome p450 enzymes. *Chem. Res. Toxicol.* **2008,** *21* (1), 220-231.
19. Esmieu, C.; Raleiras, P.; Berggren, G., From protein engineering to artificial enzymes - biological and biomimetic approaches towards sustainable hydrogen production. *Sustain Energy Fuels* **2018,** *2* (4), 724-750.
20. Wang, S.; Nie, Y.; Xu, Y., et al., Unconserved substrate-binding sites direct the stereoselectivity of medium-chain alcohol dehydrogenase. *Chem. Commun.* **2014,** *50* (58), 7770-7772.
21. Nie, Y.; Xu, Y.; Yang, M., et al., A novel NADH-dependent carbonyl reductase with unusual stereoselectivity for (R)-specific reduction from an (S)-1-phenyl-1,2-ethanediol-producing micro-organism: purification and characterization. *Lett. Appl. Microbiol.* **2007,** *44* (5), 555-562.
22. Nie, Y.; Wang, S.; Xu, Y., et al., Enzyme Engineering Based on X-ray Structures and Kinetic Profiling of Substrate Libraries: Alcohol Dehydrogenases for Stereospecific Synthesis of a Broad Range of Chiral Alcohols. *ACS Catal.* **2018,** *8* (6), 5145-5152.

23. Jackel, C.; Kast, P.; Hilvert, D., Protein design by directed evolution. *Annu Rev Biophys* **2008,** *37* (1), 153-173.
24. Arnold, F. H., Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* **2018,** *57* (16), 4143-4148.
25. Savile, C. K.; Janey, J. M.; Mundorff, E. C., et al., Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010,** *329* (5989), 305-309.
26. Chen, K.; Arnold, F. H., Engineering new catalytic activities in enzymes. *Nat. Catal.* **2020,** *3* (3), 203-213.
27. Chen, K.; Zhang, S. Q.; Brandenberg, O. F., et al., Alternate Heme Ligation Steers Activity and Selectivity in Engineered Cytochrome P450-Catalyzed Carbene-Transfer Reactions. *J. Am. Chem. Soc.* **2018,** *140* (48), 16402-16407.
28. Yang, Y.; Cho, I.; Qi, X., et al., An enzymatic platform for the asymmetric amination of primary, secondary and tertiary C(sp(3))-H bonds. *Nat Chem* **2019,** *11* (11), 987-993.
29. Jia, Z. J.; Gao, S.; Arnold, F. H., Enzymatic Primary Amination of Benzylic and Allylic C(sp(3))-H Bonds. *J. Am. Chem. Soc.* **2020,** *142* (23), 10279-10283.
30. Farwell, C. C.; Zhang, R. K.; McIntosh, J. A., et al., Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent Sci* **2015,** *1* (2), 89-93.
31. Zhang, R. K.; Huang, X.; Arnold, F. H., Selective CH bond functionalization with engineered heme proteins: new tools to generate complexity. *Curr. Opin. Chem. Biol.* **2019,** *49*, 67-75.
32. Chen, K.; Huang, X.; Kan, S. B. J., et al., Enzymatic construction of highly strained carbocycles. *Science* **2018,** *360* (6384), 71-75.
33. Zhang, J.; Huang, X.; Zhang, R. K., et al., Enantiodivergent alpha-Amino C-H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **2019,** *141* (25), 9798-9802.
34. Brandenberg, O. F.; Chen, K.; Arnold, F. H., Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. *J. Am. Chem. Soc.* **2019,** *141* (22), 8989-8995.
35. Zhang, R. K.; Chen, K.; Huang, X., et al., Enzymatic assembly of carbon-carbon bonds via iron-catalysed sp(3) C-H functionalization. *Nature* **2019,** *565* (7737), 67-72.
36. Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J., et al., Structural origins of high-affinity biotin binding to streptavidin. *Science* **1989,** *243* (4887), 85-88.
37. Creus, M.; Pordea, A.; Rossel, T., et al., X-ray structure and designed evolution of an artificial transfer hydrogenase. *Angew. Chem. Int. Ed.* **2008,** *47* (8), 1400-1404.
38. Letondor, C.; Pordea, A.; Humbert, N., et al., Artificial transfer hydrogenases based on the biotin-(strept)avidin technology: fine tuning the selectivity by saturation mutagenesis of the host protein. *J. Am. Chem. Soc.* **2006,** *128* (25), 8320-8328.
39. Durrenberger, M.; Heinisch, T.; Wilson, Y. M., et al., Artificial transfer hydrogenases for the enantioselective reduction of cyclic imines. *Angew. Chem. Int. Ed.* **2011,** *50* (13), 3026-3029.
40. Chatterjee, A.; Mallin, H.; Klehr, J., et al., An enantioselective artificial Suzukiase based on the biotin-streptavidin technology. *Chem Sci* **2016,** *7* (1), 673-677.
41. Schwizer, F.; Okamoto, Y.; Heinisch, T., et al., Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. *Chem Rev* **2018,** *118* (1), 142-231.
42. Robles, V. M.; Durrenberger, M.; Heinisch, T., et al., Structural, kinetic, and docking studies of artificial imine reductases based on biotin-streptavidin technology: an induced lock-and-key hypothesis. *J. Am. Chem. Soc.* **2014,** *136* (44), 15676-15683.
43. Liang, A. D.; Serrano-Plana, J.; Peterson, R. L., et al., Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology: Enzymatic Cascades and Directed Evolution. *Acc. Chem. Res.* **2019,** *52* (3), 585-595.
44. Heinisch, T.; Schwizer, F.; Garabedian, B., et al., E. coli surface display of streptavidin for directed evolution of an allylic deallylase. *Chem Sci* **2018,** *9* (24), 5383-5388.
45. Wu, S.; Zhou, Y.; Rebelein, J. G., et al., Breaking Symmetry: Engineering Single-Chain Dimeric Streptavidin as Host for Artificial Metalloenzymes. *J. Am. Chem. Soc.* **2019,** *141* (40), 15869-15878.
46. Hassan, I. S.; Ta, A. N.; Danneman, M. W., et al., Asymmetric delta-Lactam Synthesis with a Monomeric Streptavidin Artificial Metalloenzyme. *J. Am. Chem. Soc.* **2019,** *141* (12), 4815-4819.

47. Okamoto, Y.; Kojima, R.; Schwizer, F., et al., A cell-penetrating artificial metalloenzyme regulates a gene switch in a designer mammalian cell. *Nature Communications* **2018,** *9* (1), 1943.
48. Kohler, V.; Wilson, Y. M.; Durrenberger, M., et al., Synthetic cascades are enabled by combining biocatalysts with artificial metalloenzymes. *Nat. Chem.* **2013,** *5* (2), 93-9.
49. Davies, R. R.; Distefano, M. D., A Semisynthetic Metalloenzyme Based on a Protein Cavity That Catalyzes the Enantioselective Hydrolysis of Ester and Amide Substrates. *J. Am. Chem. Soc.* **1997,** *119* (48), 11643-11652.
50. Mahammed, A.; Gross, Z., Albumin-conjugated corrole metal complexes: extremely simple yet very efficient biomimetic oxidation systems. *J. Am. Chem. Soc.* **2005,** *127* (9), 2883-2887.
51. Raines, D. J.; Clarke, J. E.; Blagova, E. V., et al., Redox-switchable siderophore anchor enables reversible artificial metalloenzyme assembly. *Nat. Catal.* **2018,** *1* (9), 680-688.
52. Roelfes, G., LmrR: A Privileged Scaffold for Artificial Metalloenzymes. *Acc. Chem. Res.* **2019,** *52* (3), 545-556.
53. Bos, J.; Browne, W. R.; Driessen, A. J., et al., Supramolecular Assembly of Artificial Metalloenzymes Based on the Dimeric Protein LmrR as Promiscuous Scaffold. *J. Am. Chem. Soc.* **2015,** *137* (31), 9796-9799.
54. Villarino, L.; Splan, K. E.; Reddem, E., et al., An Artificial Heme Enzyme for Cyclopropanation Reactions. *Angew. Chem. Int. Ed.* **2018,** *57* (26), 7785-7789.
55. Bos, J.; Fusetti, F.; Driessen, A. J., et al., Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew. Chem. Int. Ed.* **2012,** *51* (30), 7472-5.
56. Drienovská, I.; Rioz-Martínez, A.; Draksharapu, A., et al., Novel artificial metalloenzymes by in vivo incorporation of metal-binding unnatural amino acids. *Chem. Sci.* **2015,** *6* (1), 770-776.
57. Drienovska, I.; Alonso-Cotchico, L.; Vidossich, P., et al., Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid. *Chem. Sci.* **2017,** *8* (10), 7228-7235.
58. MacMillan, D. W., The advent and development of organocatalysis. *Nature* **2008,** *455* (7211), 304-308.
59. Erkkila, A.; Majander, I.; Pihko, P. M., Iminium catalysis. *Chem. Rev.* **2007,** *107* (12), 5416-70.
60. Sun, Y.-L.; Wei, Y.; Shi, M., Applications of Chiral Thiourea-Amine/Phosphine Organocatalysts in Catalytic Asymmetric Reactions. *ChemCatChem* **2017,** *9* (5), 718-727.
61. Zhao, Y.; Domoto, Y.; Orentas, E., et al., Catalysis with anion-pi interactions. *Angew. Chem. Int. Ed.* **2013,** *52* (38), 9940-9943.
62. Tavakolian, M.; Vahdati-Khajeh, S.; Asgari, S., Recent Advances in Solvent-Free Asymmetric Catalysis. *ChemCatChem* **2019,** *11* (13), 2943-2977.
63. Levine, H. L.; Nakagawa, Y.; Kaiser, E. T., Flavopapain: synthesis and properties of semi-synthetic enzymes. *Biochem. Biophys. Res. Commun.* **1977,** *76* (1), 64-70.
64. Suckling, C. J.; Zhu, L. M., Carbon-Carbon Bond Formation Mediated by Papain Chemically Modified by Thiazolium Salts. *Bioorg. Med. Chem. Lett.* **1993,** *3* (4), 531-534.
65. Lee, K. J.; Kang, D.; Park, H.-S., Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol Cells* **2019,** *42* (5), 386-396.
66. Hilvert, D., Design of protein catalysts. *Annu. Rev. Biochem.* **2013,** *82* (1), 447-70.
67. Zanghellini, A.; Jiang, L.; Wollacott, A. M., et al., New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **2006,** *15* (12), 2785-2794.
68. Tantillo, D. J.; Chen, J.; Houk, K. N., Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **1998,** *2* (6), 743-750.
69. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M., et al., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-195.
70. Na, J.; Houk, K. N.; Hilvert, D., Transition State of the Base-Promoted Ring-Opening of Isoxazoles. Theoretical Prediction of Catalytic Functionalities and Design of Haptens for Antibody Production. *J. Am. Chem. Soc.* **1996,** *118* (27), 6462-6471.
71. Hu, Y.; Houk, K. N.; Kikuchi, K., et al., Nonspecific medium effects versus specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *J. Am. Chem. Soc.* **2004,** *126* (26), 8197-8205.

72. Khersonsky, O.; Kiss, G.; Rothlisberger, D., et al., Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci U S A* **2012,** *109* (26), 10358-10363.
73. Blomberg, R.; Kries, H.; Pinkas, D. M., et al., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013,** *503* (7476), 418-421.
74. Siegel, J. B.; Zanghellini, A.; Lovick, H. M., et al., Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010,** *329* (5989), 309-313.
75. Eiben, C. B.; Siegel, J. B.; Bale, J. B., et al., Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat. Biotechnol.* **2012,** *30* (2), 190-192.
76. Bjelic, S.; Nivon, L. G.; Celebi-Olcum, N., et al., Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-Hillman reaction. *ACS Chem Biol* **2013,** *8* (4), 749-757.
77. Jiang, L.; Althoff, E. A.; Clemente, F. R., et al., De novo computational design of retro-aldol enzymes. *Science* **2008,** *319* (5868), 1387-91.
78. Hennig, M.; Darimont, B. D.; Jansonius, J. N., et al., The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from Sulfolobus solfataricus with substrate analogue, substrate, and product. *J. Mol. Biol.* **2002,** *319* (3), 757-766.
79. Kuhlman, B.; Dantas, G.; Ireton, G. C., et al., Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** *302* (5649), 1364-1368.
80. Althoff, E. A.; Wang, L.; Jiang, L., et al., Robust design and optimization of retroaldol enzymes. *Protein Sci.* **2012,** *21* (5), 717-726.
81. Giger, L.; Caner, S.; Obexer, R., et al., Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **2013,** *9* (8), 494-8.
82. Garrabou, X.; Macdonald, D. S.; Wicky, B. I. M., et al., Stereodivergent Evolution of Artificial Enzymes for the Michael Reaction. *Angew. Chem. Int. Ed.* **2018,** *57* (19), 5288-5291.
83. Garrabou, X.; Beck, T.; Hilvert, D., A Promiscuous De Novo Retro-Aldolase Catalyzes Asymmetric Michael Additions via Schiff Base Intermediates. *Angew. Chem. Int. Ed.* **2015,** *54* (19), 5609-5612.
84. Garrabou, X.; Verez, R.; Hilvert, D., Enantiocomplementary Synthesis of gamma-Nitroketones Using Designed and Evolved Carboligases. *J. Am. Chem. Soc.* **2017,** *139* (1), 103-106.
85. Garrabou, X.; Wicky, B. I.; Hilvert, D., Fast Knoevenagel Condensations Catalyzed by an Artificial Schiff-Base-Forming Enzyme. *J. Am. Chem. Soc.* **2016,** *138* (22), 6972-4.
86. Garrabou, X.; Macdonald, D. S.; Hilvert, D., Chemoselective Henry Condensations Catalyzed by Artificial Carboligases. *Chemistry* **2017,** *23* (25), 6001-6003.
87. Privett, H. K.; Kiss, G.; Lee, T. M., et al., Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (10), 3790.
88. Russ, W. P.; Figliuzzi, M.; Stocker, C., et al., An evolution-based model for designing chorismate mutase enzymes. *Science* **2020,** *369* (6502), 440.
89. Senior, A. W.; Evans, R.; Jumper, J., et al., Improved protein structure prediction using potentials from deep learning. *Nature* **2020,** *577* (7792), 706-710.
90. Yang, K. K.; Wu, Z.; Arnold, F. H., Machine-learning-guided directed evolution for protein engineering. *Nature Methods* **2019,** *16* (8), 687-694.
91. Zhao, Y.; Cotelle, Y.; Avestro, A. J., et al., Asymmetric Anion-pi Catalysis: Enamine Addition to Nitroolefins on pi-Acidic Surfaces. *J. Am. Chem. Soc.* **2015,** *137* (36), 11582-11585.
92. Cotelle, Y.; Lebrun, V.; Sakai, N., et al., Anion-pi Enzymes. *ACS Cent. Sci.* **2016,** *2* (6), 388-93.
93. Liu, L.; Cotelle, Y.; Klehr, J., et al., Anion-pi catalysis: bicyclic products with four contiguous stereogenic centers from otherwise elusive diastereospecific domino reactions on pi-acidic surfaces. *Chem. Sci.* **2017,** *8* (5), 3770-3774.
94. Lechner, H.; Emann, V. R.; Breuning, M., et al., An Artificial Cofactor catalyzing the Baylis-Hillman Reaction using Designed Streptavidin as Protein Host. *bioRxiv* **2020**, 2020.03.05.978098.
95. Santi, N.; Morrill, L. C.; Luk, L. Y. P., Streptavidin-Hosted Organocatalytic Aldol Addition. *Molecules* **2020,** *25* (10), 2457.

96. Nodling, A. R.; Swiderek, K.; Castillo, R., et al., Reactivity and Selectivity of Iminium Organocatalysis Improved by a Protein Host. *Angew. Chem. Int. Ed.* **2018,** *57* (38), 12478-12482.
97. Chin, J. W., Expanding and reprogramming the genetic code. *Nature* **2017,** *550* (7674), 53-60.
98. Nodling, A. R.; Spear, L. A.; Williams, T. L., et al., Using genetically incorporated unnatural amino acids to control protein functions in mammalian cells. *Essays Biochem.* **2019,** *63* (2), 237-266.
99. Wan, W.; Tharp, J. M.; Liu, W. R., Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **2014,** *1844* (6), 1059-1070.
100. Dumas, A.; Lercher, L.; Spicer, C. D., et al., Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **2015,** *6* (1), 50-69.
101. Drienovská, I.; Roelfes, G., Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nat. Catal.* **2020,** *3* (3), 193-202.
102. Drienovska, I.; Mayer, C.; Dulson, C., et al., A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **2018,** *10* (9), 946-952.
103. Burke, A. J.; Lovelock, S. L.; Frese, A., et al., Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **2019,** *570* (7760), 219-223.
104. Zhou, Z.; Roelfes, G., Synergistic catalysis in an artificial enzyme by simultaneous action of two abiological catalytic sites. *Nat. Catal.* **2020,** *3* (3), 289-294.
105. Notz, W.; Tanaka, F.; Barbas, C. F., 3rd, Enamine-based organocatalysis with proline and diamines: the development of direct catalytic asymmetric Aldol, Mannich, Michael, and Diels-alder reactions. *Acc. Chem. Res.* **2004,** *37* (8), 580-591.
106. Szanto, G.; Bombicz, P.; Grun, A., et al., Highly enantioselective organocatalytic conjugate addition of nitromethane to benzylidene acetones. *Chirality* **2008,** *20* (10), 1120-1126.
107. Kwiatkowski, P.; Beeson, T. D.; Conrad, J. C., et al., Enantioselective organocatalytic alpha-fluorination of cyclic ketones. *J. Am. Chem. Soc.* **2011,** *133* (6), 1738-1741.
108. Ouellet, S. G.; Tuttle, J. B.; MacMillan, D. W., Enantioselective organocatalytic hydride reduction. *J. Am. Chem. Soc.* **2005,** *127* (1), 32-33.
109. Ahrendt, K. A.; Borths, C. J.; MacMillan, D. W. C., New Strategies for Organic Catalysis: The First Highly Enantioselective Organocatalytic Diels−Alder Reaction. *J. Am. Chem. Soc.* **2000,** *122* (17), 4243-4244.
110. Brauer, T. M.; Zhang, Q.; Tiefenbacher, K., Iminium Catalysis inside a Self-Assembled Supramolecular Capsule: Scope and Mechanistic Studies. *J. Am. Chem. Soc.* **2017,** *139* (48), 17500-17507.
111. Lou, J.; Liu, F.; Lindsay, C. D., et al., Dynamic Hyaluronan Hydrogels with Temporally Modulated High Injectability and Stability Using a Biocompatible Catalyst. *Adv. Mater.* **2018,** *30* (22), e1705215.
112. Freund, M.; Schenker, S.; Tsogoeva, S. B., Enantioselective nitro-Michael reactions catalyzed by short peptides on water. *Org Biomol Chem* **2009,** *7* (20), 4279-4284.

2. Covalently Modifying Penicillin Binding Proteins with Carbapenems to Create Organocatalytic Artificial Enzymes

2.1 Preface

The ubiquitous mechanism of penicillin binding proteins to form covalent adducts with β-lactam antibiotics at the active site, offers a unique opportunity to create an artificial enzyme. Exploration of the natural metabolism allows for the anchoring of potentially catalytic moieties into a protein backbone. To test this design strategy two protein back bones, a D,D-transpeptidase and a β-lactamase, from *Mycobacterium tuberculosis* were modified with two carbapenems bearing a potentially catalytic pyrrolidine ring. The modification of the proteins, the activity of the catalysts alone and the enantioselectivity of the reactions were each explored. One designer enzyme was able to perform iminium ion catalysed Michael-nitro addition reactions.

This chapter contributed to the publication: Williams, T. L.; Nodling, A. R.; Tsai, Y. H., et al., Carbapenems as water soluble organocatalysts. *Wellcome Open Res.* **2018,** *3*, 107.

2.2 Introduction

Penicillin binding proteins (PBPs) are proteins that can bind to the β-lactam structure consistent in penicillin.[1] Common proteins in this class of enzymes are transpeptidases that cross-link peptidoglycan in the cell wall of bacteria and β-lactamases that can hydrolyse a β-lactam ring in penicillin-based antibiotics.[2] Nucleophilic attack from a deprotonated serine at the carbonyl carbon in the peptide bond or β-lactam results in a tetrahedral intermediate. Collapse of this intermediate and protonation of the adjacent nitrogen results in the leaving of D-alanine, in the case of D-ala-D-alanine transpeptidase, or the ring opening of the β-lactam, in the case of β-lactamases resulting in the formation of an acyl-enzyme intermediate. With reference to the transpeptidase, the N-terminus of another peptidoglycan molecule can attack the intermediate which results in the formation of a new peptide bond connecting the two molecules together (**Figure 35a**).[3] When the substrate is instead a β-lactam antibiotic, the acyl-enzyme intermediate forms, but no incoming nucleophile breaks the intermediate and the transpeptidase will be inhibited. β-lactamases are able to hydrolyse this intermediate, opening the β-lactam ring and inactivating the drug (**Figure 35b**).[4]

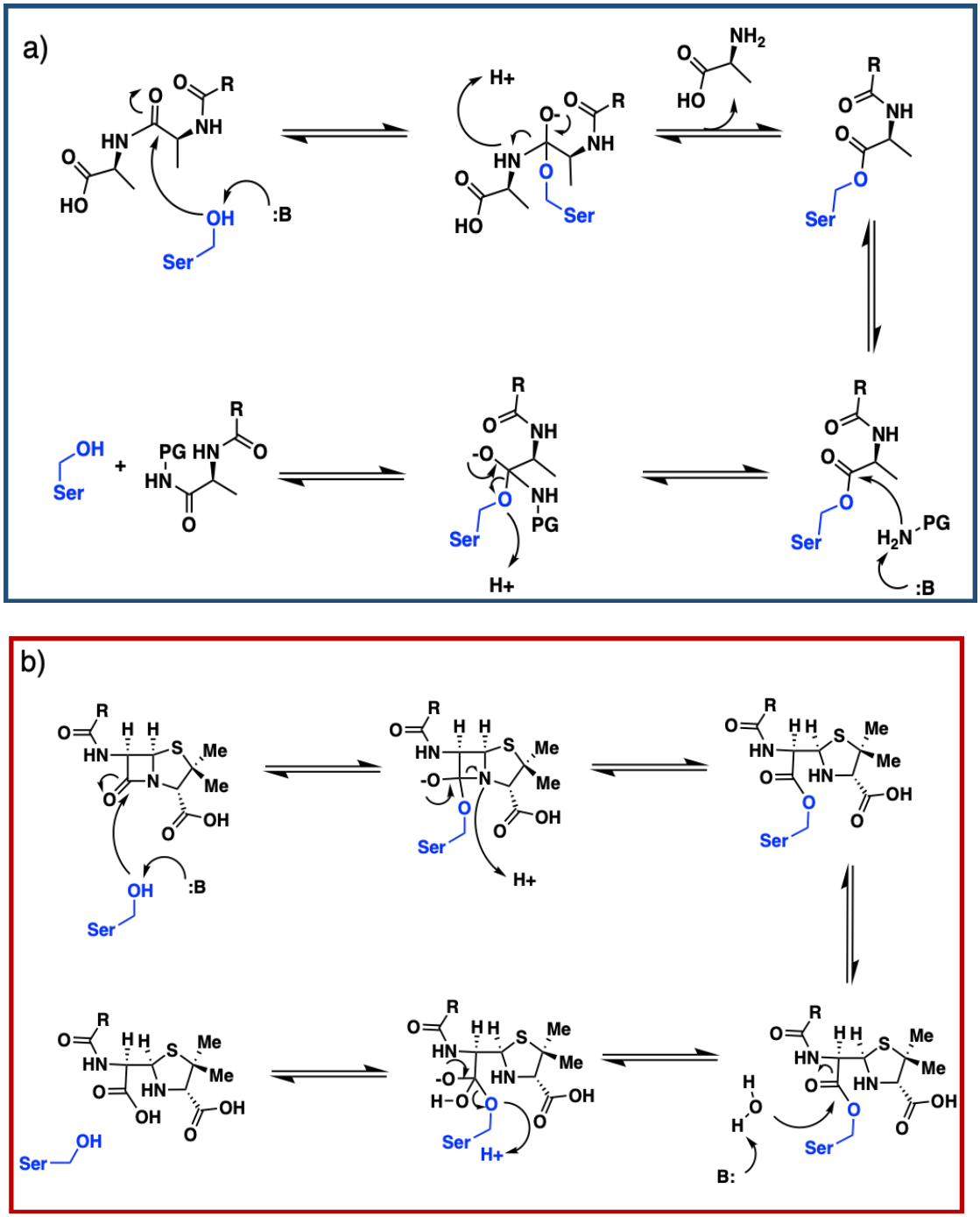


Figure 35. a) The general mechanism of the penicillin binding protein D-ala-D-alanine transpeptidase that facilitates the cross linking of peptidoglycan in bacterial cell wall biosynthesis (PG = peptidoglycan). **b)** The general mechanism of β-lactam hydrolysis by class A β-lactamases. Both enzymes required a catalytic serine (shown in blue).

Due to the rise of β-lactam resistance, chemists aimed to create new structures that cannot be hydrolysed by these enzymes and so can once again overcome the spread of pathogenic microorganisms. Following penicillin based antibiotics (**Figure 36a),** cephalosporins (**Figure 36b**) were first applied and the most recent class were

carbapenems (**Figure 36c**).[5] Carbapenems can form the same covalent intermediate in the active site of β-lactamases and PBPs.[6] Three of these carbapenems meropenem, doripenem and ertapenem have a secondary amine as a pyrrolidine ring in their structure (**Figure 36d**). This offers an opportunity to utilise the three carbapenems for iminium ion or enamine catalysis. Anchoring of these potentially catalytic moieties in the active site of a protein by exploiting their natural metabolism could allow formation of a hybrid secondary amine – protein catalyst. Should the carbapenems be active, the universal mechanism of this class of enzymes allow ease of passing these reactive moieties to many PBPs and β-lactamases allowing the exploration of various protein backbones for different reactions or to accommodate new substrates.

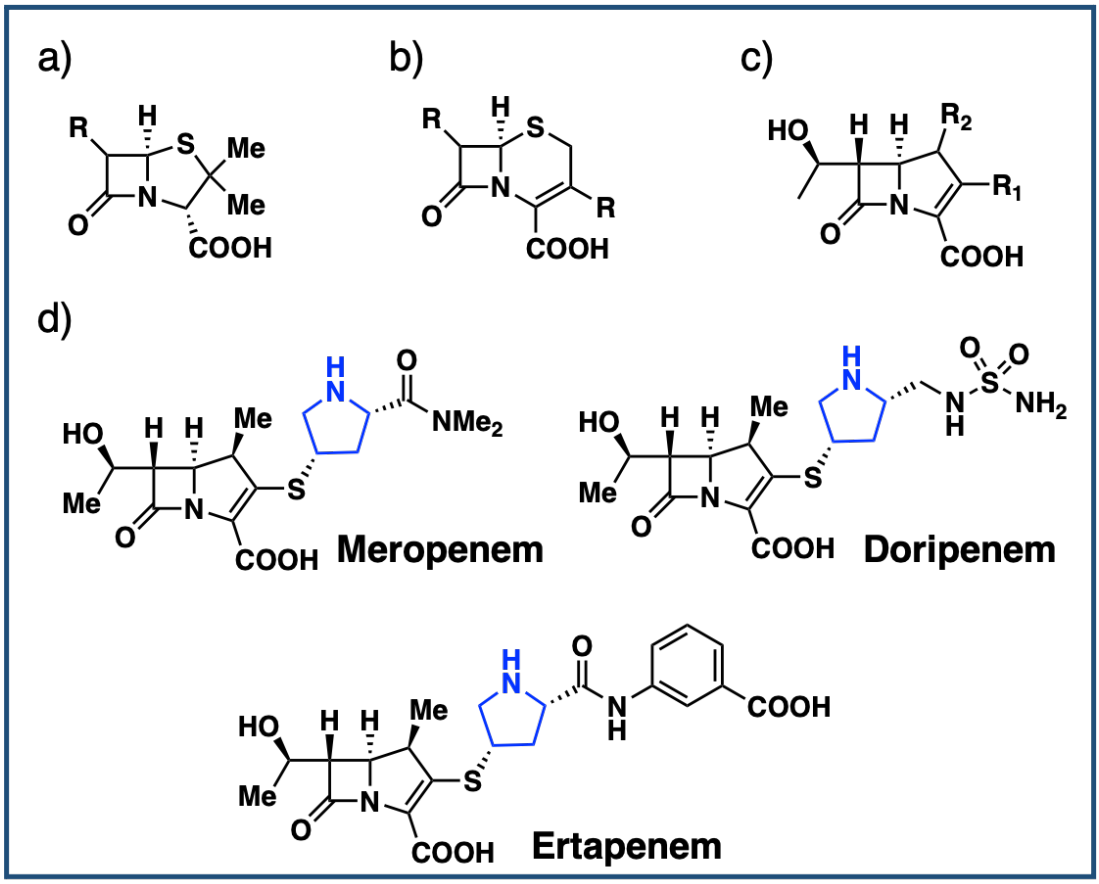


Figure 36. a) General structure of the penicillins. **b)** General structure of the cephalosporins. **c)** General structure of the carbapenems. **d)** The structures of the three carbapenems used in this study.

A suitable protein backbone that can perform the native reaction is required to form a stable intermediate. *Mycobacterium tuberculosis* (*Mt*) is a multi-drug resistant bacterium that is resistant to many antibiotics.[7] The origin of the drug resistance was discovered to be a highly promiscuous class-A β-lactamase dubbed BlaC.[8] BlaC

showed substantial activity towards penicillin, cephalosporin and carbapenem based antibiotics. The enzyme's promiscuity and relatively large substrate scope, along with the crystallography data (**Figure 37d)**, show a large and dynamic active site that could be used to accommodate an organocatalytic reaction, should the carbapenems be active in the chosen reaction. The second protein chosen was Rv3330. This is one of the D,D-transpeptidases that was isolated from *Mt* and again there is crystallographic data with the carbapenems available.[9] This structural insight can aid the design of the artificial enzyme.

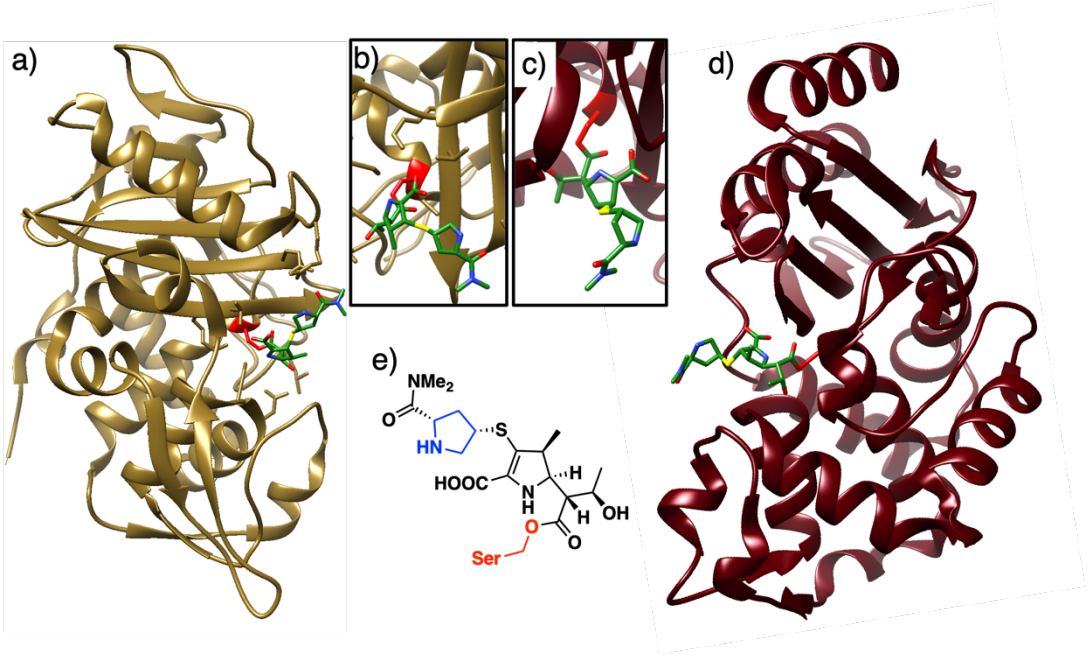


Figure 37. a) Crystal structure of D,D-transpeptidase Rv3330 from *Mycobacterium tuberculosis* (**PDB: 4PPR**) with meropenem covalently bound in the active site. **b)** Close up of the meropenem acyl-enzyme intermediate bound to the nucleophilic serine (red) of Rv3330. **c)** Close up of the meropenem acyl-enzyme intermediate bound to the nucleophilic serine (red) of BlaC. **d)** Crystal structure of β-lactamase BlaC from *Mycobacterium tuberculosis* (**PDB: 3DWZ**) with meropenem covalently bound in the active site. **e)** Chemical structure representation of as an ester adduct of a nucleophilic serine bound to meropenem.

2.3 Results and Discussion

The first step to realising an active artificial enzyme was to determine if the carbapenems were active as catalysts for a secondary amine organocatalytic reaction. To test this hypothesis, the Michael-nitro addition of nitromethane to cinnamaldehyde was used as a model reaction (**Figure 38**).

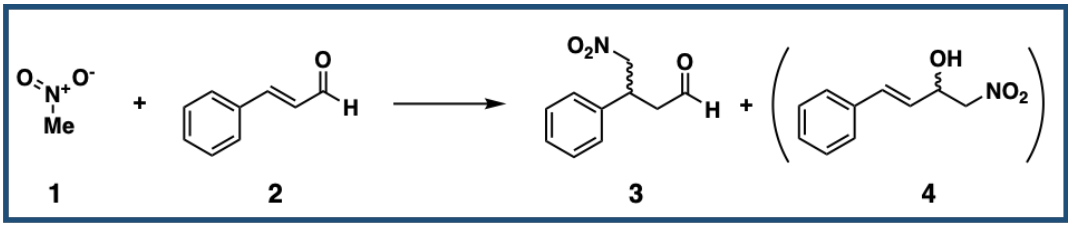


Figure 38. The nitro Michael addition between nitromethane (**1**) and cinnamaldehyde (**2**) to yield the 1, 4 addition product (**3**) or the 1, 2 addition side product (**4**).

The activity of the carbapenems was evaluated under different aqueous conditions including those that mimic biological environments. Proteins require a buffered solution to be active and as such it was important to establish the activity of the carbapenems in the above reaction under these conditions. 20 mol% of each catalyst was used, relative to cinnamaldehyde and the reactions were performed for 24 hours in a thermoshaker at pH 7.0 and 25 °C. The reactions were extracted with $CDCl_3$ and the conversion determined by ^1H NMR (**Table1**).

Table 1. Percentage conversion cinnamaldehyde to nitro-cinnamaldehyde (**3**) with different buffers at pH 7.0. Percentage conversion determined by ^1H NMR after 24 h at 25 °C. Each entry is an average of three repeats.

Catalyst	H_2O	HEPES	KP_i	NaP_i	PBS
Meropenem	0	33	38	33	35
Doripenem	0	38	29	39	31
Ertapenem	1	8	9	8	13

All four buffers that were tested produced similar conversions (**Table 1**) relative to each catalyst. Ertapenem showed poor activity, however doripenem and meropenem showed moderate conversion. This suggests that the carbapenems can be utilised as catalysts. Interestingly, the catalyst requires a buffered environment to facilitate this reaction. When water was used alone no reaction occurred and so it seems to be that buffered conditions are crucial for the reaction to proceed in aqueous environments. The addition of NaCl to the buffer (PBS) resulted in little change to the conversion. PBS is a very common buffer, often used in enzyme reactions and to mimic biological environments and so was chosen for further experiments.

With the knowledge of the importance of buffer in the reaction, it seemed prudent to explore the effect of pH. Knowing that protonation of the secondary amine would result in almost no reactivity and the reaction can be base catalysed at too high a pH or result in the production of a side product (**4**) the 1,2 addition, pH 7.0, 7.5 and 8.0

were chosen to perform the reaction in PBS under the same conditions as the original study.

Table 2. Percentage conversion of substrate to product and side product in 10 mM PBS at indicated pH. Percentage conversion determined by ^{1}H NMR after 24 h at 25 °C. Each entry is an average of three repeats.

Catalyst	pH 7.0		pH 7.5		pH 8.0	
	3	**4**	**3**	**4**	**3**	**4**
Meropenem	35	1	25	75	23	25
Doripenem	31	1	34	26	29	28
Ertapenem	13	0	8	4	5	3

The pH screen revealed that the competing reaction of the addition of nitromethane and 1,2 addition reaction was more prominent above pH 7.0 (**Table 2**). This finding suggests that for the carbapenems to catalyse the reaction through an iminium ion facilitated reaction, pH 7.0 is optimum as it generates solely the wanted product. Ertapenem again, did not perform very well producing very low substrate to product conversion.

To increase the reaction efficiency, it was hypothesised that the substrates could benefit from improved solubility. Organic co-solvent in a 1:1 ratio was added to the reaction. Four solvents were explored. Two miscible solvents, methanol and acetonitrile, and two non-miscible solvents, benzene and chloroform. The reactions were performed at pH 7.0 using the previous conditions.

Table 3. Percentage conversion of substrate to product in the presence of co-solvents at 1:1 ratio. Percentage conversion determined by ^{1}H NMR after 24 h at 25 °C. Each entry is an average of three repeats.

Catalyst	Miscible co-solvent		Non-miscible co-solvent	
	Methanol	**Acetonitrile**	**Benzene**	**Chloroform**
Meropenem	71	19	0	0
Doripenem	74	27	1	2
Ertapenem	22	5	0	6

The addition of the non-miscible co-solvents, benzene and chloroform, abolished the formation of the product completely (**Table 3**). The addition of the miscible solvent acetonitrile however did not improve the substrate conversion. On the other hand, the addition of methanol clearly improved the product yield. This is likely due to the increased solubility of the organic compounds involved. Methanol is also a protic solvent, this may aid in the protonation and deprotonation throughout the catalytic cycle, whereas acetonitrile does not have this characteristic.

The carbapenems used in this study are not unlike traditional organocatalysts. Surrounding the cyclic secondary amine are periphery groups that are thought to induce enantioselectivity. With this in mind, and the potential of the product to be chiral it would be prudent to determine if they can facilitate enantioselective reactions. Only meropenem and doripenem were tested as ertapenem did not produce a significant conversion of substrate to product. To determine the enantioselectivity of the catalysts, the reactions were scaled up 10 times and the product purified following extraction. The product was reduced to the alcohol using sodium borohydride and the resulting alcohol subjected to chiral HPLC analysis.

Table 4. Enantiomeric ratio (*R*:*S*) of product **3**. Ratio calculated from the ratio of the peak area from chiral HPLC. Assignment performed by comparison of a pure chiral standard from the literature.[10]

Catalyst	PBS only (%*e.e.*)	1:1 PBS:MeOH (%*e.e.*)
Meropenem	43:57 (14)	41:59 (12)
Doripenem	43:57 (14)	47:53 (6)

It is clear from the chiral analysis (**Table 4**) that neither meropenem nor doripenem produce a substantial chiral product under these conditions. Meropenem in the presence of methanol and PBS gave the highest *e.r.* at 41:59 in favour of the *S* isomer. Despite poor enantioselectivity it was confirmed that both meropenem and doripenem can catalyse secondary amine organocatalytic reactions, likely via an iminium ion, under aqueous conditions. This is indeed interesting as the carbapenems could now be anchored to a penicillin binding protein or a β-lactamase in the hope of either improving the conversion or stereo selectivity. The transpeptidase Rv3330 and β-lactamase BlaC from *mycobacterium tuberculosis* were chosen to anchor the carbapenems and create an artificial enzyme.

The gene encoding for the transpeptidase Rv3330 was a kind gift from the Alber lab[9] and was recombinantly expressed in *E. coli* and the resulting protein was purified by Ni-NTA affinity chromatography. The carbapenems meropenem and doripenem were added to the lysis buffer. Successful expression and purification were determined by SDS-PAGE (**Figure 40**).

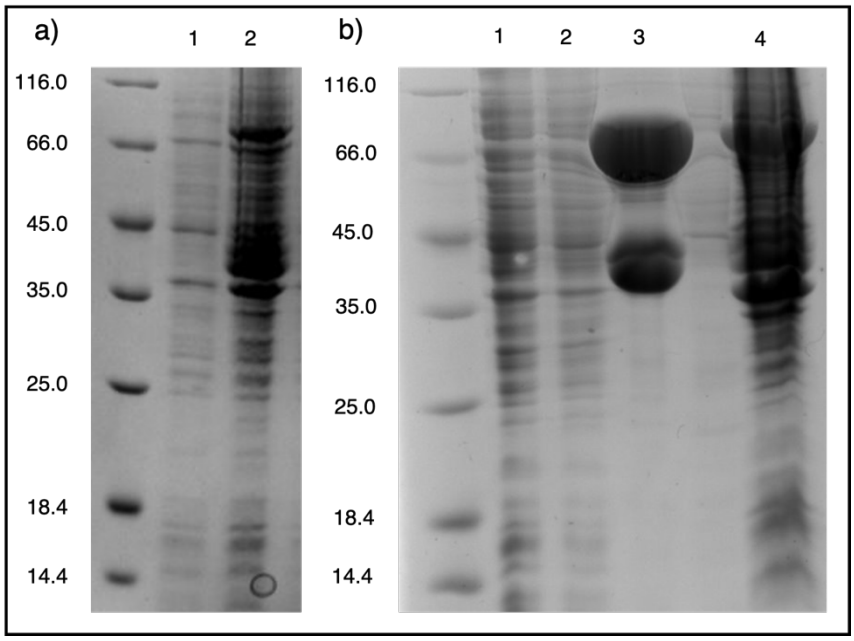


Figure 39. SDS-PAGE analysis of recombinant Rv3330. Calculated mass = 35879 Da. **a)** Expression analysis of Rv3330, 1 – before induction, 2 – after induction with IPTG. **b**) Ni-NTA purification analysis of Rv3330, 1 – flow through, 2 – wash, 3 – elution, 4 – inclusion bodies. Protein ladders in kDa. Two bands are present in the elution fraction. The monomer of the protein Rv3330 at approx. 36 kDa and the disulphide dimer at approx. 72 kDa.

Following successful expression and purification (**Figure 39**), Rv3330 proved difficult to work with for application of an artificial enzyme. Upon purification in different buffers, it was determined that the enzyme was not fully active. In addition, without any disufide reducing agent the protein partially dimerised at high concentrations in solution. The LC-MS shows that the enzyme does not fully react with the carbapenems (**Figure 40**). This results in an incomplete system. As such the enzyme Rv3330 was not pursued further.

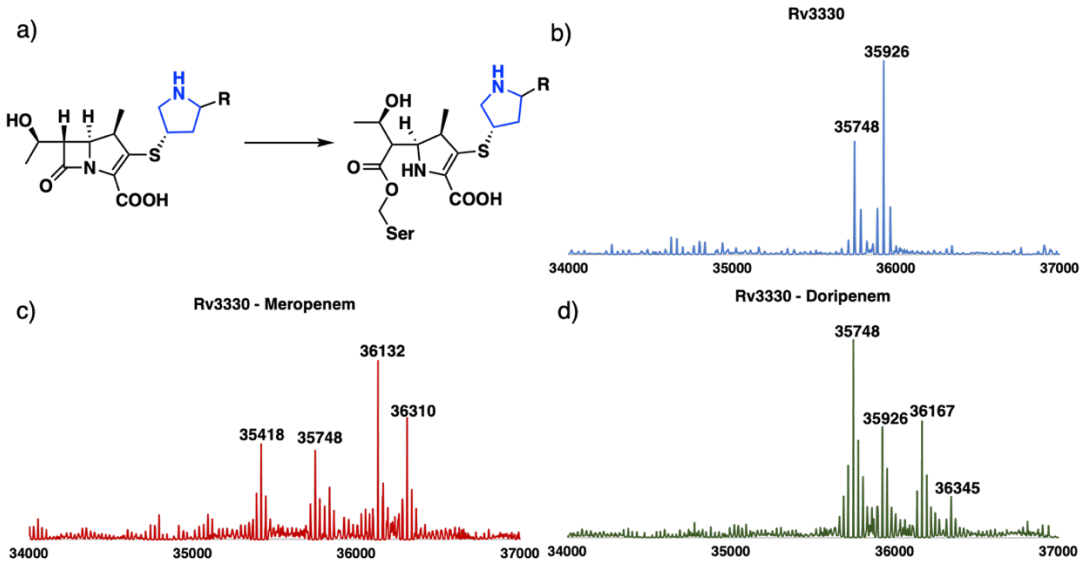


Figure 40. Deconvoluted ESI-MS of purified Rv3330 enzymes. **a)** Adduct formation between the active site serine and the carbapenems. **b)** Rv3330 reference mass = 35748 Da (35926 Da = Rv3330 + Formyl-Met). **c)** Rv3330 + meropenem = 36132 Da (36310 Da = Rv3330 + meropenem + Formyl-Met). **d)** Rv3330 + doripenem = 36167 Da (36345 Da = Rv3330 + doripenem + Formyl-Met).

BlaC was next tested as a suitable backbone. The gene expressing the wild-type protein was recombinantly expressed in *E. coli* and purified by Ni-NTA chromatography with the carbapenems. However, the enzyme was able to hydrolyse the ester bond that anchors the carbapenems to the protein by activation of a water molecule. To circumvent this, a point mutation was introduced into the enzyme using PCR. Glutamate 166 was mutated to an alanine and this greatly reduced the carbapenems hydrolysis, as reported in the literature.[11] Expression and purification of the mutant enzyme was performed using the same procedure as the wild-type enzyme and analysed by SDS-PAGE (**Figure 41**).

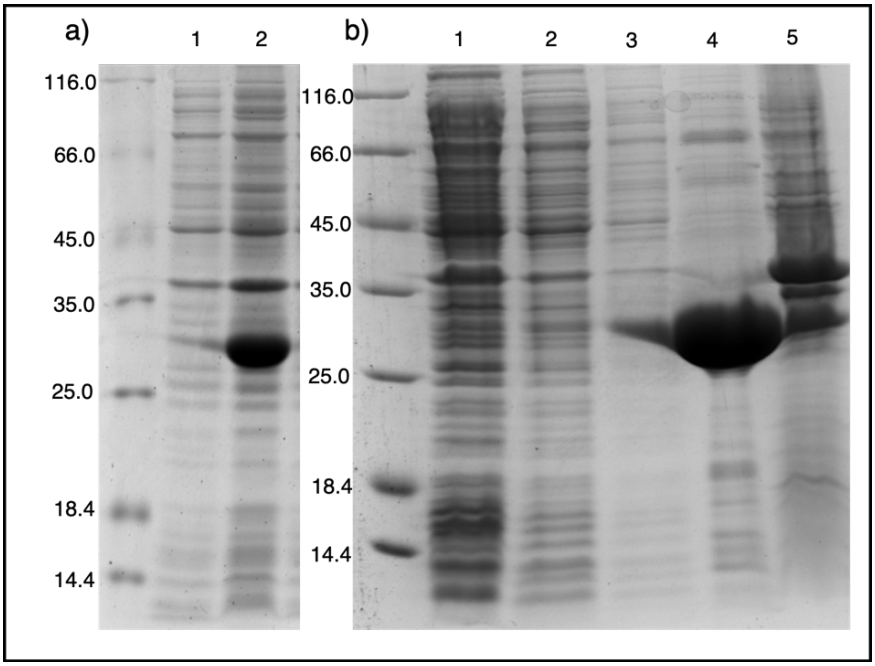


Figure 41. SDS-PAGE analysis of recombinant BlaC(Glu166Ala). Calculated mass = 30608 Da. **a)** Expression analysis of BlaC(E166A), 1 – before induction, 2 – after induction with IPTG. **b)** Ni-NTA purification analysis of BlaC(E166A), 1 – flow through, 2 – wash 1, 3 – wash 2, 4 – elution, 5 – inclusion bodies. Protein ladder in kDa.

Following the purification in the presence of the carbapenems, size exclusion chromatography was then employed to simultaneously buffer exchange the protein into the reaction buffer and remove any residual carbapenem. The BlaC-carbapenem artificial enzymes were then subjected to LC-MS analysis to determine the extent of the labelling (**Figure 42**).

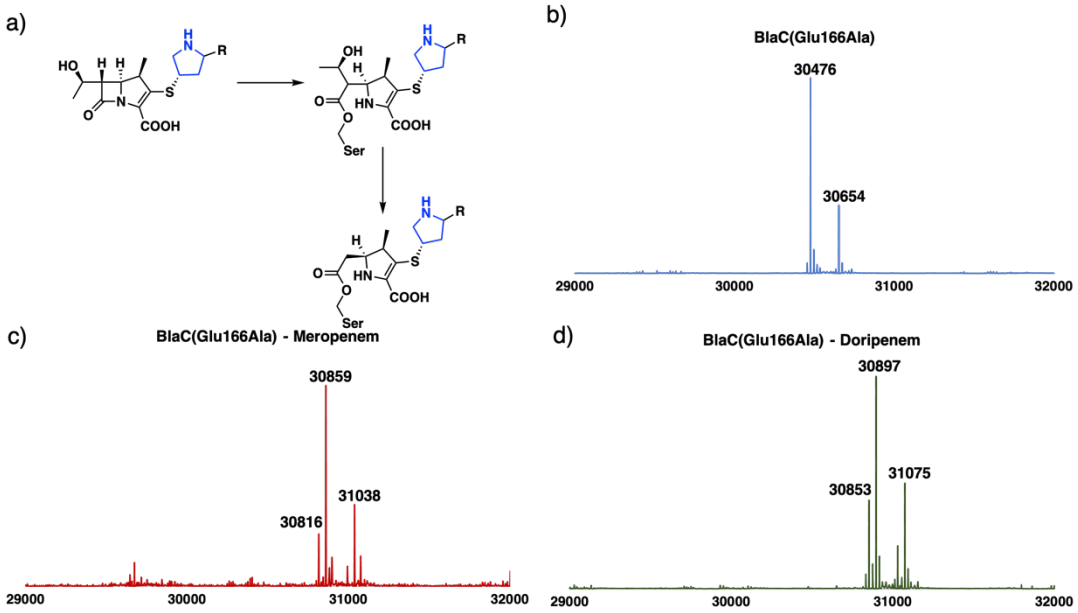


Figure 42. Deconvoluted ESI-MS of purified BlaC enzymes. **a)** Retro aldol product of the carbapenems following reaction between the nucleophilic serine and the carbapenem. **b)** BlaC

reference mass = 30476 Da (30654 Da = BlaC + Formyl-Met). **c)** BlaC + meropenem = 30859 Da (31038 Da = BlaC + meropenem + Formyl-Met, 30816 Da = retro aldol product of meropenem bound to BlaC). **d)** BlaC + doripenem = 30897 Da (31075 Da = BlaC + doripenem + Formyl-Met, 30853 Da = retro aldol product of doripenem bound to BlaC).

The mass spectrometry analysis of the BlaC-carbapenem complexes also revealed an interesting trait of BlaC (**Figure 42**). Following the reaction of the substrate with the enzyme, the carbapenems can undergo a retro-aldol reaction to produce acetaldehyde (**Figure 42a**). Even if that may be the case, it seems that the enzyme is fully labelled with the carbapenem allowing for them to be screened for activity. Upon labelling the enzyme BlaC with meropenem and doripenem the nitro addition reaction was performed. The original reaction conditions used for the carbapenems alone were not tolerated by the enzymes. As such the conditions were modified by reducing the concentration of enzyme and using only 10% methanol. The substrate to product conversion was determined by ^{1}H NMR and upon scale up, the product was purified and the enantioselectivity of the hybrid enzyme determined by chiral HPLC.

Table 5. Percentage conversion and enantiomeric ratios of **3** in the presence of BlaC-carbapenem complexes. Conversion determined using ^{1}H NMR after 18 h at 25 °C. Enantioselectivity calculated from the ratio of the peak area determined by chiral HPLC. Assignment performed by comparison of a pure chiral standard from the literature.

Catalyst	Conversion (%)	*R:S (%e.e.)*
Meropenem	22	43:57 (14)
BlaC – Meropenem	20	55:45 (10)
Doripenem	23	46:54 (8)
BlaC - Doripenem	27	58:42 (16)

The modified reaction conditions resulted in lower conversions as expected due to lower catalyst loading. The artificial enzymes produced approximately the same conversion as the carbapenems alone (**Table 5**). Enantioselectivity, as determined by chiral HPLC was also poor. Little to no improvement was exhibited by the BlaC-carbapenem complexes, but the ratio was inverted, suggesting that the protein does have some influence on the stereoselectivity of the reaction, but it seems that the enzyme and the catalyst are opposed in their selectivity. Even with the glutamate to alanine mutation, after 48 hours background hydrolysis of the carbapenems resulted in the degradation of the artificial enzymes as can be seen from LC-MS analysis **(Figure 43).**

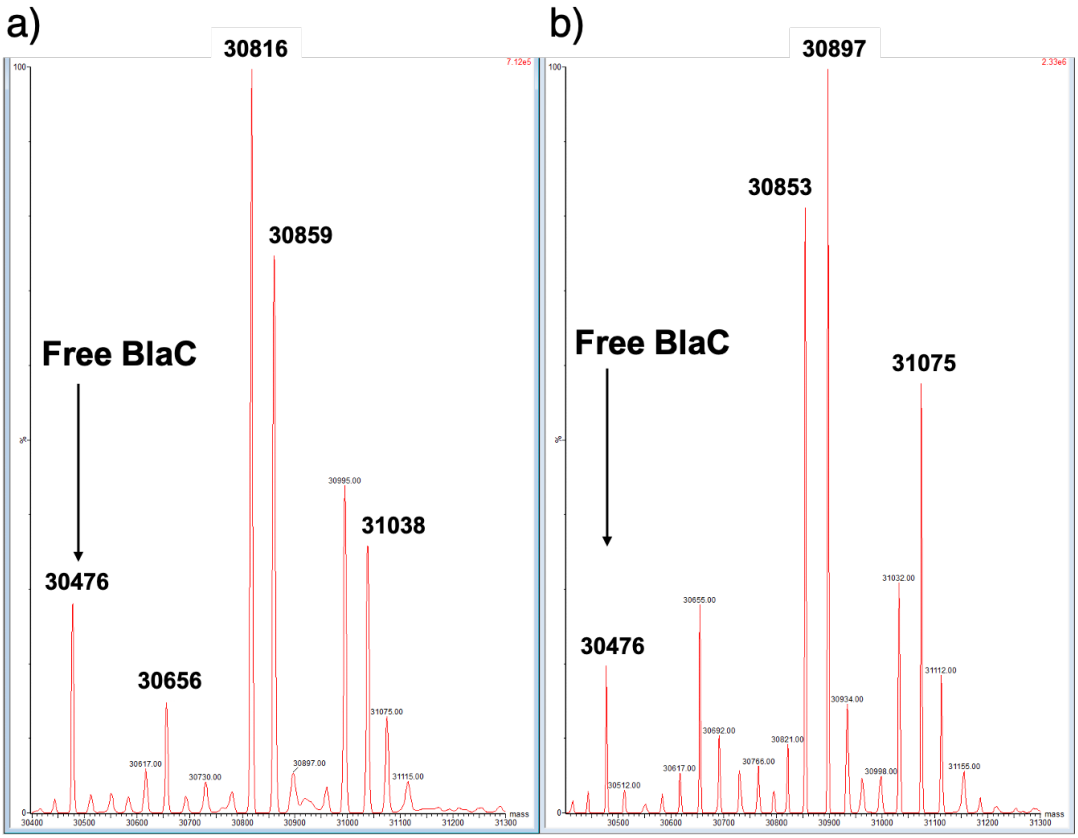


Figure 43. LC-MS analysis of BlaC following 48 hours after labelling with the carbapenems. **a)** The meropenem-BlaC artifical enzyme. **b)** The doripenem-Blac artificial enzyme. The mass indicated at 30476 is the calculated mass of BlaC with no carbapenem attached.

2.4 Conclusion

It has been shown that the carbapenems meropenem and doripenem can be used to facilitate the iminium ion catalysed Michael addition of nitromethane to cinnamaldehyde under aqueous conditions. In addition, by placing the catalysts inside a protein backbone the reaction can still proceed. However, when BlaC is used without the carbapenems no product is formed suggesting that the carbapenems in the active site of the enzyme are essential for conversion.

The data suggest that the protein backbone has little effect on the reaction performed. Firstly, the conversion of substrate to product with and without the enzyme is approximately the same, the conversion is poor due to the instability of the enzyme at higher concentrations. The enzyme requires a significant concentration to produce a small amount of product. Secondly the enzymes are not enantioselective, however the small enantiomeric ratio that is observed is reversed when placed in the enzyme suggesting that the backbone does induce some selectivity. Other residues lining the

active site pocket may need to be modified to reduce the natural hydrolysis or a different PDP may be employed.

This approach can lead to an artificial enzyme however there is much optimisation required before an efficient system is produced. The screening of further backbones could aid in identifying a more efficient system. Mutations around the active site may also be beneficial to both activity and selectivity. The transpeptidase was determined to be a poor choice for this system and so other transpeptidases could be explored.

2.5 References

1. Waxman, D. J.; Strominger, J. L., Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. *Annu. Rev. Biochem.* **1983,** *52* (1), 825-869.
2. Zapun, A.; Contreras-Martel, C.; Vernet, T., Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol. Rev.* **2008,** *32* (2), 361-385.
3. Sauvage, E.; Kerff, F.; Terrak, M., et al., The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **2008,** *32* (2), 234-258.
4. Macheboeuf, P.; Contreras-Martel, C.; Job, V., et al., Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiol. Rev.* **2006,** *30* (5), 673-691.
5. Papp-Wallace, K. M.; Endimiani, A.; Taracila, M. A., et al., Carbapenems: past, present, and future. *Antimicrob. Agents Chemother.* **2011,** *55* (11), 4943-4960.
6. Hugonnet, J. E.; Tremblay, L. W.; Boshoff, H. I., et al., Meropenem-clavulanate is effective against extensively drug-resistant Mycobacterium tuberculosis. *Science* **2009,** *323* (5918), 1215-1218.
7. Wang, F.; Cassidy, C.; Sacchettini, J. C., Crystal structure and activity studies of the Mycobacterium tuberculosis beta-lactamase reveal its critical role in resistance to beta-lactam antibiotics. *Antimicrob. Agents Chemother.* **2006,** *50* (8), 2762-2771.
8. Tremblay, L. W.; Fan, F.; Blanchard, J. S., Biochemical and structural characterization of Mycobacterium tuberculosis beta-lactamase with the carbapenems ertapenem and doripenem. *Biochemistry* **2010,** *49* (17), 3766-3773.
9. Prigozhin DM, K. I., Huizar JP, Mavrici D, Waldo GS, Hung L-W, Sacchetinni JC, Terwilliger TC, Alber T., Subfamily-Specific Adapatations in the Structures of Two Penicillin-Binding Proteins from *Mycobacterium tuberculosis*. *PLoS ONE* **2014,** *9* (12).
10. Mager, I.; Zeitler, K., Efficient, enantioselective iminium catalysis with an immobilized, recyclable diarylprolinol silyl ether catalyst. *Org. Lett.* **2010,** *12* (7), 1480-1483.
11. Tremblay, L. W.; Xu, H.; Blanchard, J. S., Structures of the Michaelis complex (1.2 A) and the covalent acyl intermediate (2.0 A) of cefamandole bound in the active sites of the Mycobacterium tuberculosis beta-lactamase K73A and E166A mutants. *Biochemistry* **2010,** *49* (45), 9685-9687.

3. Genetic Incorporation of Secondary Amine Unnatural Amino Acids to Create Artificial Enzymes

3.1 Preface

To generate an artificial enzyme with an iminium ion-based mechanism, genetic code expansion was used as the design strategy. This method allowed the incorporation of three secondary amine unnatural amino acids into four positions of the multidrug resistant binding protein LmrR. The twelve artificial enzymes were then tested in an iminium ion catalysed selective reduction of cinnamaldehyde to afford the product hydro-cinnamaldehyde using the synthetic hydride donor benzyl nicotinamide. When D-prolyl-L-lysine was incorporated at position Phe93, the resulting artificial protein was able to perform the aforementioned reaction generating a good conversion. D-prolyl-L-lysine was then incorporated into EcDHFR at three separate positions and when incorporated at position Ala7, an excellent conversion was accomplished for the same reaction but here NADPH was used as the hydride source. Kinetic characterisation of the most promising enzymes was performed and application to the selective reduction of a further seven α,β-unsaturated carbonyls was explored. The EcHDFR variant was then applied to whole cell catalysis and a coupling reaction where the cofactor was able to be recycled *in situ*. Further mechanistic characterisation was performed to determine the artificial enzymes catalytic activation with isotope incorporation studies, intermediate analysis and kinetic isotope effects supporting an iminium ion mechanism. This chapter highlights the flexibility of utilising genetic code expansion in generating artificial enzymes and introduces new secondary amine organocatalytic based designer enzymes.

3.2 Introduction

Expanding the genetic code with catalytically competent unnatural amino acids to afford metal free catalysts, has only recently been realised. The introduction of *p*-aminophenylalanine into LmrR[1, 2] and methyl-histidine into BH32,[3] has provided new avenues of designer enzymes. In the first instance, hydrazone and oxime ligations were accomplished, however the amino acid has to be chemically modified from its azido precursor to be able to perform its reactions. Methyl-histidine has exhibited the ability to cleave ester substrates via hydrolysis, although it could be argued that this is an expansion of the enzyme's already known activity rather than a wholly new *de novo* mechanism.

Secondary amines, on the other hand, have a wide and well-known reaction profile.[4, 5] Iminium ion, enamine and singly occupied molecular orbital (SOMO) catalysis can all be accomplished. With the introduction of this catalytic residue into a protein backbone, new artificial enzymes may be realised that can in theory, perform any reactions that have been preceded by secondary amine small molecule catalysis. In this chapter three unnatural amino acids that bear secondary amine prolyl-rings were used as the catalytic residues. The unnatural amino acids are D-prolyl-L-lysine (DPK), L-prolyl-L-lysine (LPK) and thiazolidine lysine (ThzK) (**Figure 44**). The selective reduction of α,β-unsaturated carbonyl compounds at the double bond through an iminium ion catalysed mechanism was performed to prove the concept of this design.

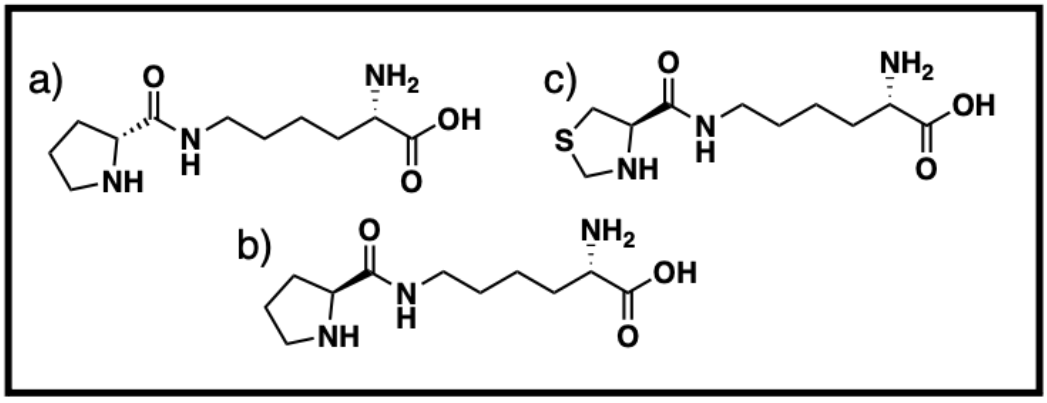


Figure 44. Three secondary amine unnatural amino acids. **a)** D-prolyl-L-lysine (DPK). **b)** L-prolyl-L-Lysine. **c)** Thiazolidine lysine (ThzK).

The benefit of using genetic code expansion over other strategies, such as the N-terminal proline and the supramolecular approach, is that there is no limitation to the

protein backbone that may be exploited and the position of the catalytic reactive centre can be placed anywhere in said backbone.[6] This allows far greater flexibility when assessing which protein will be used to host the catalysis. In this design, two protein architectures that employ strikingly different characteristics, were utilised to perform the aforementioned hydride transfer. In addition, a different hydride source was used for each enzyme.

The first protein to undergo re-design was the homo dimeric lactococcal multidrug resistant regulator, LmrR from *Lactococcus lactis*.[7] The hydrophobic binding pocket can accommodate small hydrophobic organic molecules[8] and so the synthetic hydride donor benzyl-nicotinamide (BNAH) was used. The second protein was the natural oxidoreductase dihydrofolate reductase from *Escherichia coli,* EcDHFR. This enzyme has a native binding site for the natural cofactor nicotinamide adenine dinucleotide phosphate (NADPH) and reduces the cofactor dihydrofolate to tetrahydrofolate.[9]

Four positions in LmrR were chosen for introduction of the unnatural amino acids (**Figure 45**). These positions were chosen based on previous studies in the literature that showed promise in enzyme design and additionally, each lines the hydrophobic binding pocket upon formation of the dimer in solution. The four positions were Val15, Asp19, Met89 and Phe93.[2]

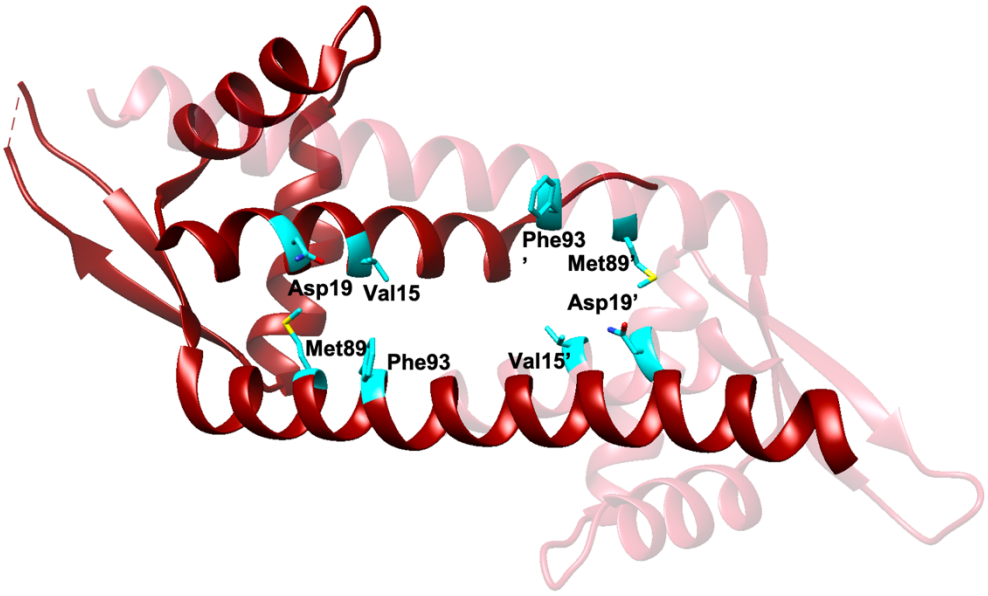


Figure 45. Crystal structure of LmrR (**PDB:3F8F**). The four residues targeted for mutation are in cyan.

The second protein host, EcDHFR exhibits protein motions and dynamics during catalysis.[10]. To exploit the transferability of genetic code expansion, three positions in EcDHFR (**Figure 46**) were chosen to incorporate the unnatural amino acid DPK for use in the hydride transfer reaction, this time using NADPH as the hydride source. The three positions chosen were Ala7 which is located at the end of a dynamic loop that moves upon binding to NADPH, Phe93 that is located at the active site where the natural substrate dihydrofolate occupies and finally Ser49 which is also on the periphery of the active site.

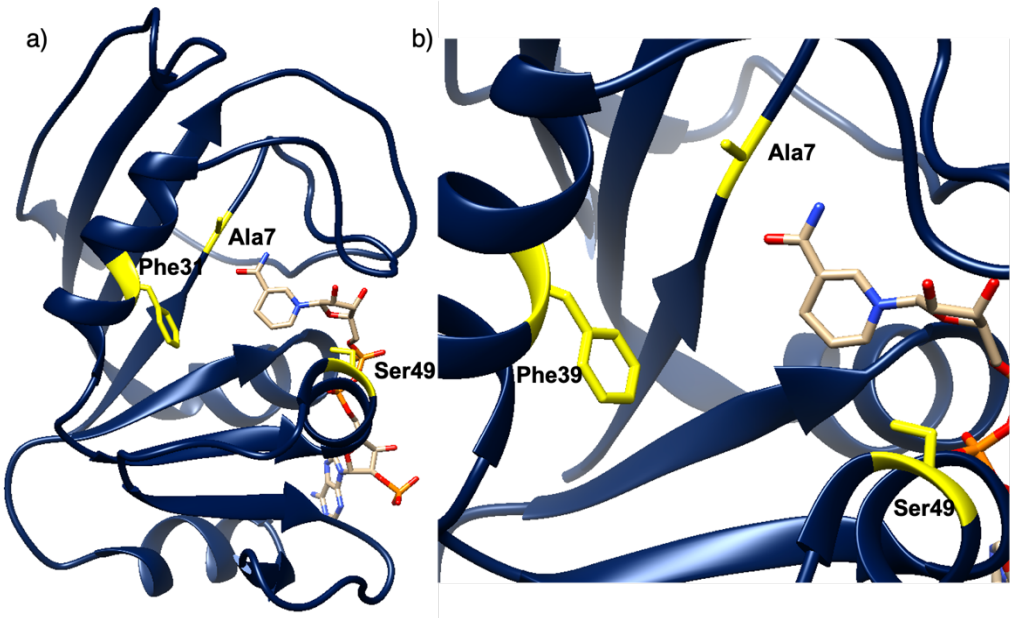


Figure 46. a) The crystal structure of EcDHFR with NADPH bound in the active site (**PDB: 1RA1**); residues targeted for mutation are in yellow. **b)** Close up of EcDHFR's active site.

The enzymes were investigated for their ability to perform iminium ion catalysed hydride transfer reactions and characterised in comparison, to demonstrate the power of exploiting different protein backbones for the same catalytic mechanism. In addition, as the reactions were performed under biocompatible conditions the EcDHFR variant was tested for it's ability to catalyse the organocatalytic reaction in concert with a natural enzyme.

3.3 Results and Discussion

To engineer an artificial enzyme by genetic code expansion, firstly the catalytic amino acids had to be synthesised. Each was synthesised using the same protocol. Firstly, formation of an activated ester of the prolyl carboxylic acids with pentafluoro phenol. This intermediate was then coupled to the epsilon-amine of protected lysine through

nucleophilic substitution. The amino acids were deprotected yielding the final products (**Figure 46**).

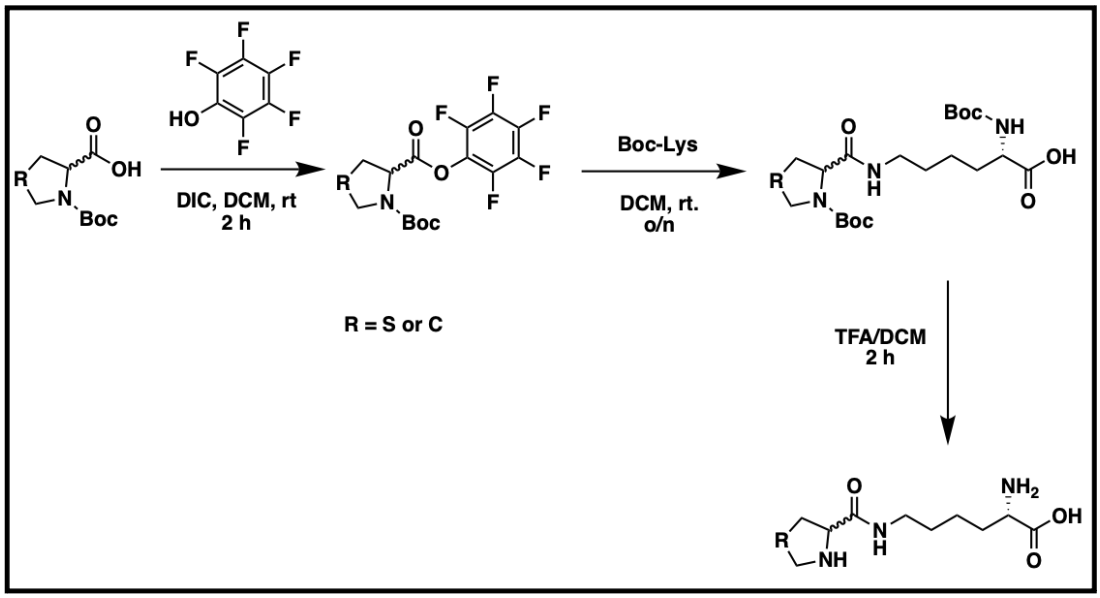


Figure 47. Synthesis of the unnatural amino acids DPK (*R*) (R = CH_2), LPK (*S*) (R = CH_2) and ThzK (*S*) (R = S). An activated ester is formed between the carboxylic acid of the prolyl group and pentafluorophenol using diisopropylcarbodiimide as the coupling reagent. *N*α-Boc-L-Lysine is then coupled through a peptide bond by nucleophilic substitution between the free amine and the ester of the activated prolyl-group. The di-Boc protected dipeptide is then deprotected using trifluoracetic acid.

DPK and LPK had previously been tested as analogues of pyrrolysine for incorporation using pyrrolysyl-tRNA synthetase.[11] It was shown that DPK could be incorporated by the wild-type synthetase, but that LPK could not. ThzK was incorporated using a mutant synthetase of pyrrolysyl-tRNA synthetase that was produced via directed evolution.[12] To determine if the three amino acids could be incorporated as in the literature the reporter protein sfGFP was used (**see section 4.3**). The gene was cloned into a pET-28a vector and so was placed under the control of the lac operon.

A point mutation in this gene was then performed via PCR to generate an Asp150TAG mutation. Incorporation of an unnatural amino acid via amber codon suppression will result in a full-length protein as pyrrolysyl-tRNA naturally recognises the TAG codon. A second plasmid, that was a kind gift from Jason Chin,[6] that harboured the wild-type *Methanosarcina bakeri* pyrrolysyl-tRNA synthetase (MbPylRS) and its cognate tRNA, both under constitutive promoters, was used for incorporation of the amino acid DPK. The wild-type synthetase plasmid was then used as the template to generate the mutant synthetase via PCR, to incorporate ThzK, This mutant was used to test the

incorporation of LPK. The sfGFP plasmid, along with the relevant synthetase plasmid was co-transformed into *E. coli* cells and grown to an OD_{600} of 0.6, then 1 mM of the relevant amino acid was added to the media along with IPTG to induce the gene expression of the reporter protein. The cells were incubated overnight, and the product protein was analysed by SDS-PAGE (**Figure 48**).

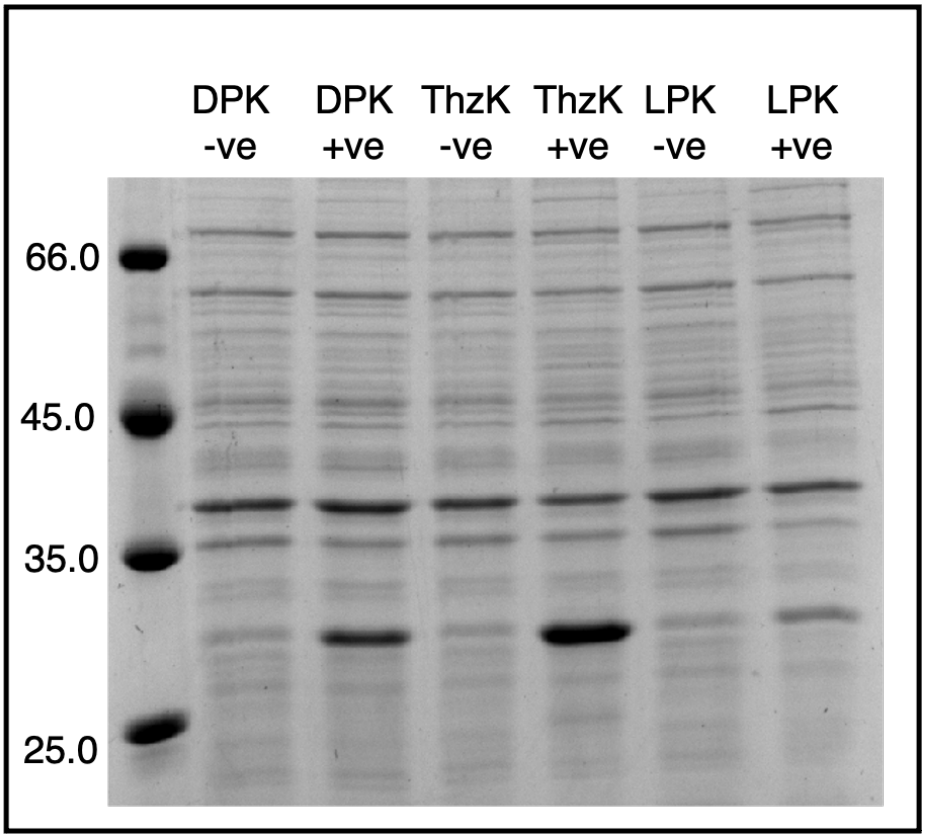


Figure 48. SDS-PAGE analysis for the incorporation of DPK, LPK and ThzK into the reporter protein sfGFP (Asp150TAG). Mass of SfGFP approx. 28 kDa. Protein ladder in kDa.

Each amino acid was successfully incorporated into the reporter protein sfGFP (**Figure 48**). As mentioned above the multi-drug resistant binding protein LmrR form *Lactococcus lactis* has previously been exploited for this method of enzyme design. LmrR forms a homodimer that can recruit hydrophobic compounds.[13] Four previously reported positions were targeted for mutation. Each TAG mutation was introduced by PCR independently and the mutants co-expressed with the relevant synthetase in *E. coli* cells. Each unnatural amino acid was added to a final concentration of 1 mM and the proteins were purified by Ni-NTA affinity chromatography (**Figure 49**). Each LmrR variant, upon purification, was then dialysed overnight into the reaction buffer (50 mM NaPi, 150 mM NaCl, pH 7.0).

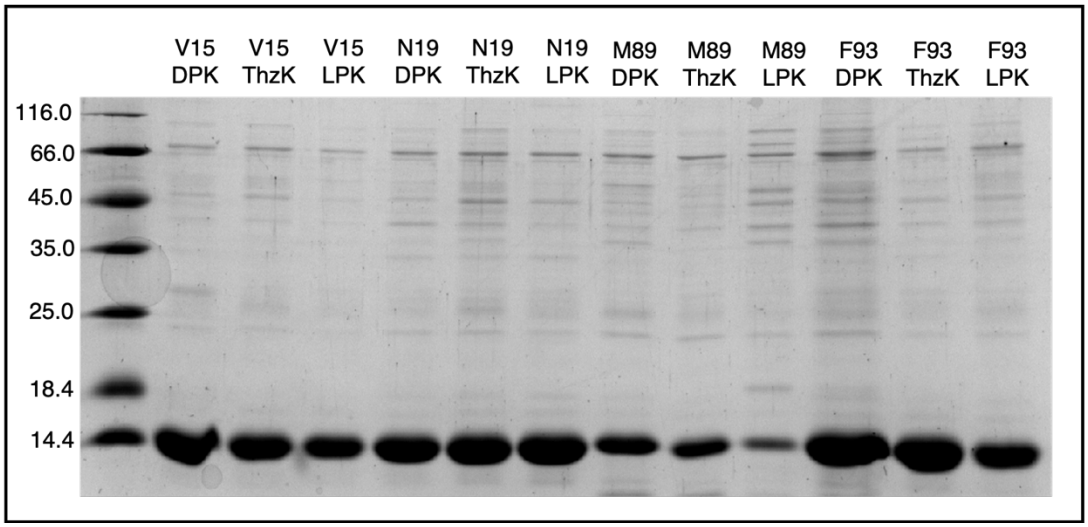
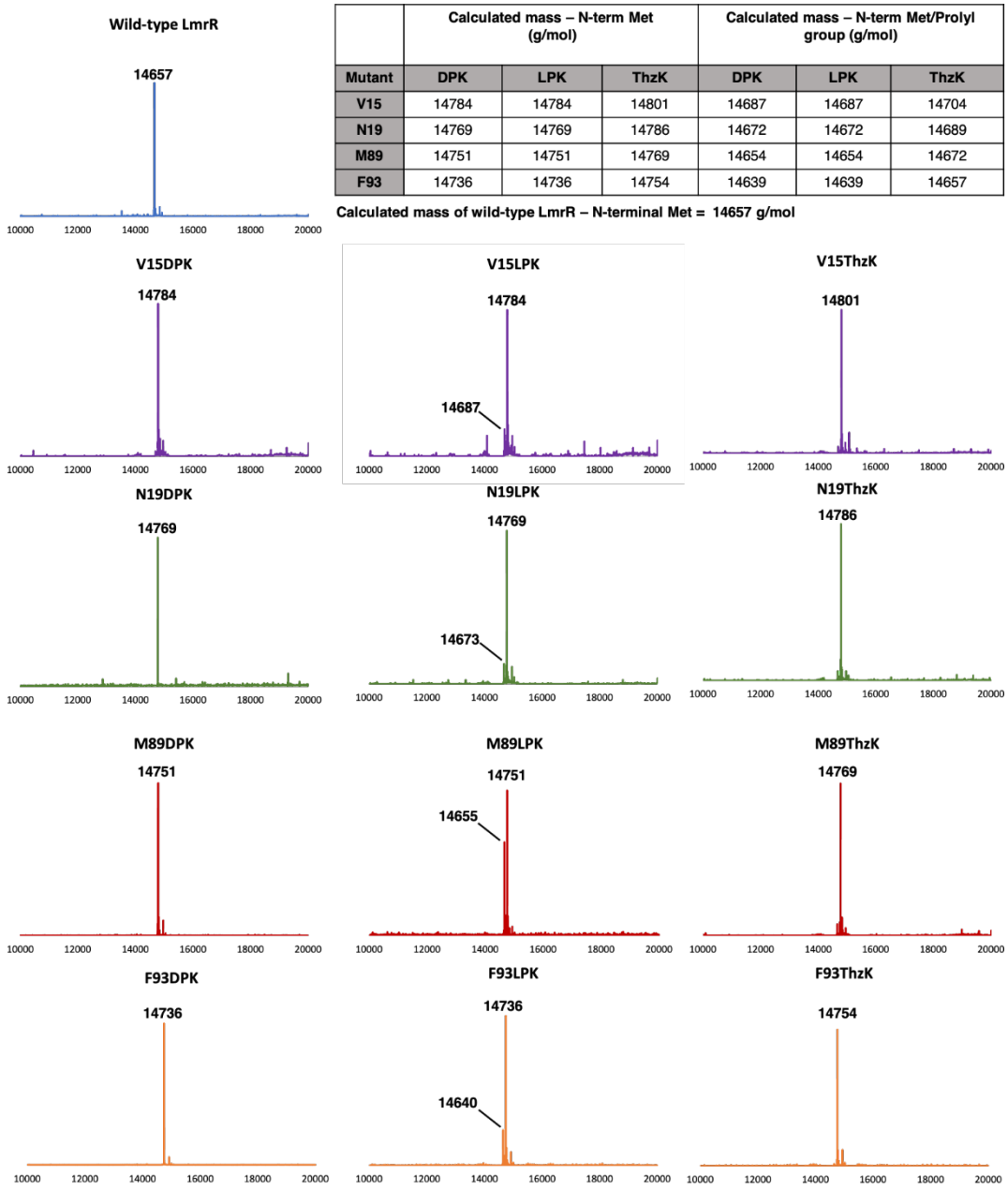


Figure 49. SDS-PAGE analysis of the purified LmrR mutants containing the indicated unnatural amino acid at the specified position. Average mass of the calculated protein monomer = 14757 Da. Protein ladder in kDa.

SDS-PAGE analysis of the twelve LmrR variants demonstrated that each was successfully expressed and purified. To ensure that each LmrR mutant had successfully incorporated the amino acid at the correct position and that the amino acid was intact, protein mass spectrometry was used to measure the accurate mass (**Figure 50**).

Mass spectrometry analysis of the artificial enzymes suggests that all twelve enzymes were successfully expressed with the unnatural amino acids in the positions of interest based on mass difference. However, it was interesting to note that those enzymes that had the amino acid LPK incorporated, did show various levels of truncation (loss of the prolyl group determined by a mass difference of 97 g mol^{-1}). This was likely due to an endogenous enzyme that is able to cleave the peptide bond between a L-proline and L-lysine residue. Even so, the enzymes that contained LPK were tested for their activity.



	Calculated mass – N-term Met (g/mol)			Calculated mass – N-term Met/Prolyl group (g/mol)		
Mutant	**DPK**	**LPK**	**ThzK**	**DPK**	**LPK**	**ThzK**
V15	14784	14784	14801	14687	14687	14704
N19	14769	14769	14786	14672	14672	14689
M89	14751	14751	14769	14654	14654	14672
F93	14736	14736	14754	14639	14639	14657

Calculated mass of wild-type LmrR – N-terminal Met = 14657 g/mol

Figure 50. Deconvoluted ESI mass spectra of wild-type LmrR and each mutant, determined from the protein mass spectrometry charged state.

The reaction chosen to test the artificial enzymes was a biomimetic hydride transfer (**Figure 51b**). Selective hydrogenation of α,β-unsaturated aldehydes has precedent in nature. However, engineering of enzymes to affect efficient reduction, with a universal hydride source, on non-natural substrates is not a trivial task. As such it would be beneficial to explore this reaction via iminium ion catalysis. Here a synthetic hydride donor benzyl-nicotinamide (BNAH) (**Figure 51a**) of the natural hydride source, nicotinamide adenine dinucleotide (phosphate) (NAD(P)H) was used (**Figure 51a**).

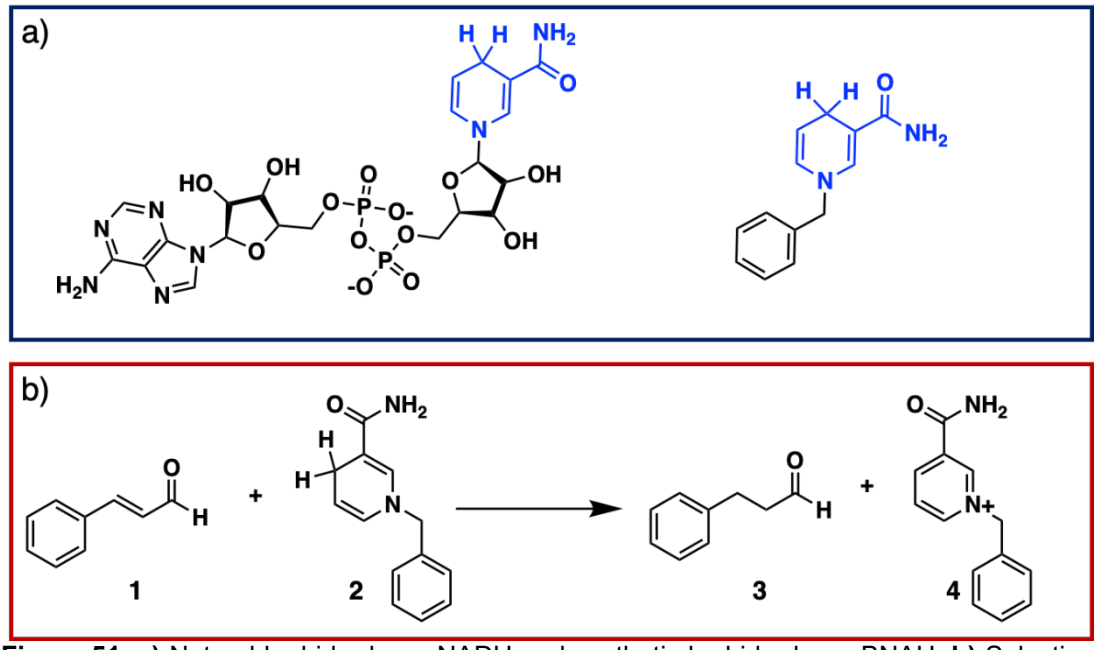


Figure 51. a) Natural hydride donor NADH and synthetic hydride donor BNAH. **b)** Selective reduction of cinnamaldehyde by BNAH.

The hydride donor BNAH was synthesised using a straightforward procedure from the literature (**Figure 52**).[14] It was hypothesised that the hydrophobic nature of the hydride donor would allow localisation into the hydrophobic pore at the dimer interface of LmrR when in an aqueous environment. To determine if each mutant formed the dimer in solution analytical size exclusion chromatography was performed. Each mutant had a retention time that matches that of the wild-type protein (**See appendix section 7.3**).

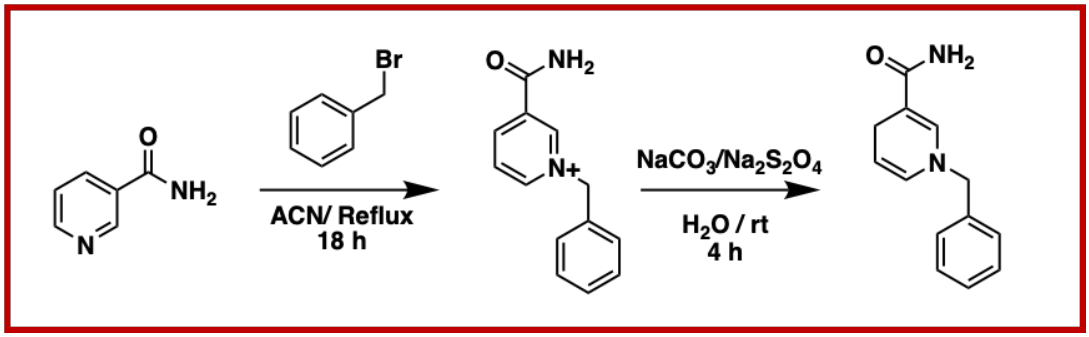


Figure 52. Synthesis of the hydride donor benzyl-nicotinamide (BNAH).

Each enzyme was concentrated following dialysis to a concentration of 2 mg/mL and 5 µL of a stock solution of cinnamaldehyde in methanol was added to the enzyme followed by 5 µL of a stock solution of BNAH also in methanol. This resulted in a catalyst loading of 10 mol% in comparison to cinnamaldehyde and an excess of the hydride donor (2 eq to cinnamaldehyde). The reactions were placed in a

thermoshaker for 18 hours at 25 °C. The substrates and products from each reaction were extracted with 200 µL of DCM and analysed by gas chromatography mass spectrometry (GC-MS) (**Figure 53**).

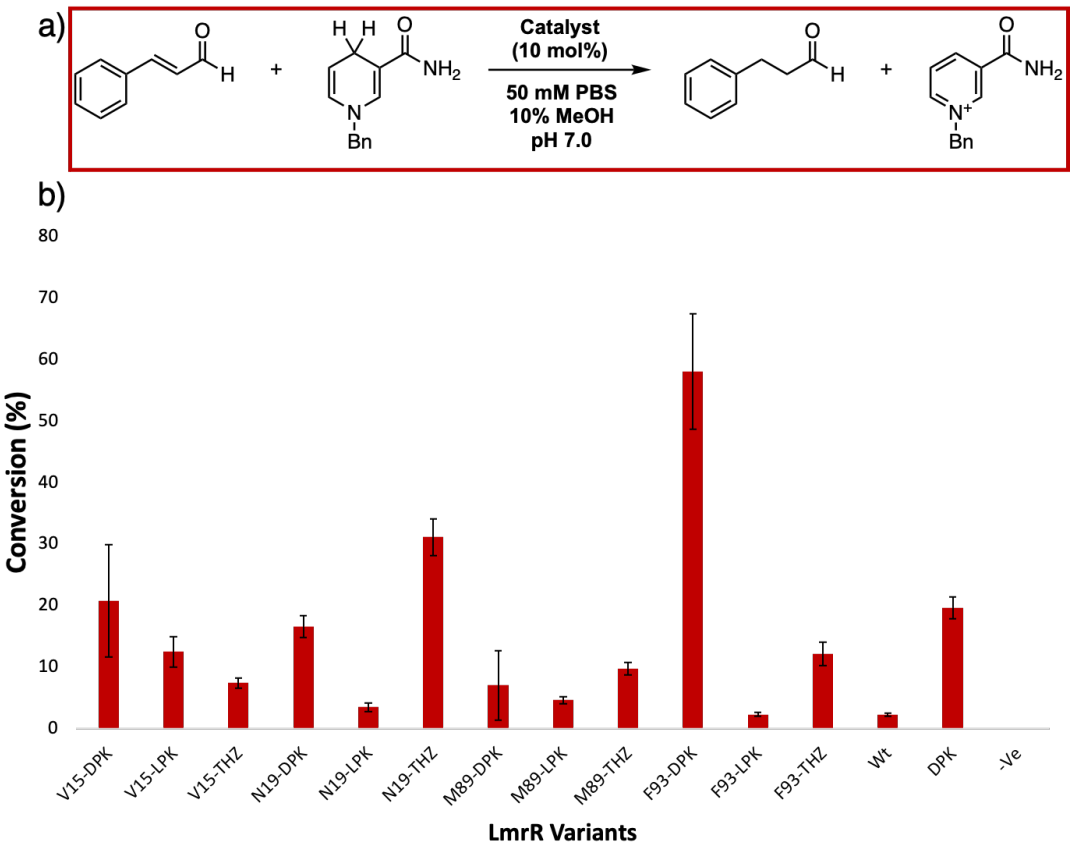


Figure 53. a) Selective reduction of cinnamaldehyde by BNAH to hydro-cinnamaldehyde. **b)** Peak area of the hydride transfer product hydro-cinnamaldehyde. for each of the LmrR based artificial enzymes, the wild-type LmrR, DPK and the substrates alone under the same reaction conditions. Reaction conversion determined by comparison to a standard calibration curve for the product hydro-cinnamaldehyde. Each chart entry is an average of three repeats. Error bars are a representation of the standard deviation.

It can be seen clearly from the conversion determined by GC-MS (**Figure 53b**) that the most successful enzyme was the LmrR Phe93DPK mutant with consistently the highest conversion measured against a standard curve (58%). It is also interesting to note that the negative control and the wild-type exhibited little to no activity under the same conditions. DPK alone produced only 14% conversion. The Phe93DPK artificial enzyme was taken forward to be further investigated.

Following successful screens for the amino acid DPK in LmrR, EcDHFR was next utilised. Here, the three TAG mutations were again introduced by PCR and the enzyme was co-expressed with the wild-type MbPylRS tRNA synthetase and its

cognate tRNA. DPK was added to a final concentration of 1 mM and the proteins again purified by Ni-NTA chromatography (**Figure 54**). Each mutant, upon purification was then dialysed overnight into the reaction buffer (50 mM NaPi, 150 mM NaCl, pH 7.0).

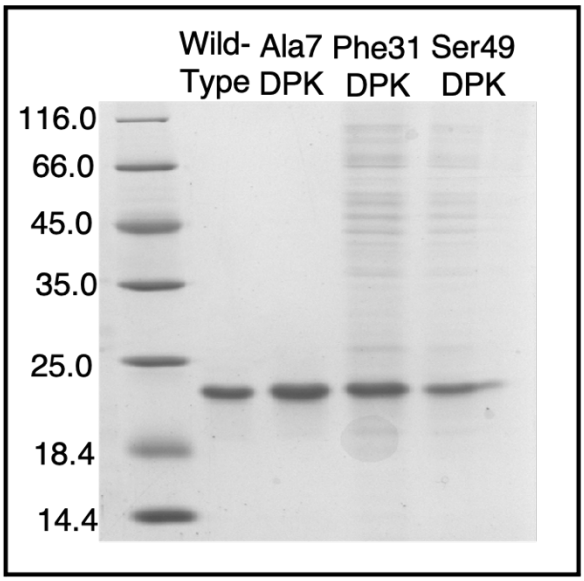
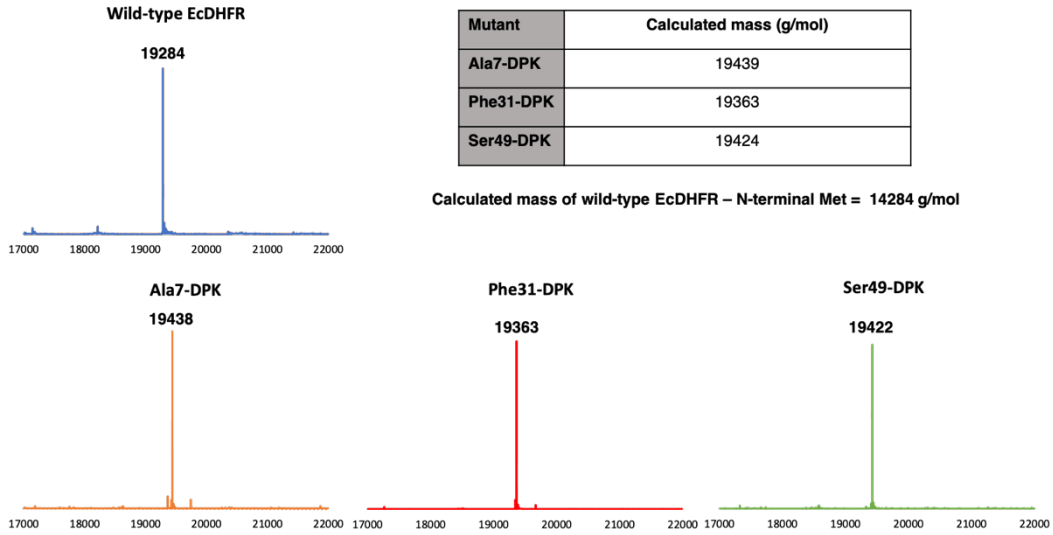


Figure 54. SDS-PAGE analysis of the purified EcDHFR mutants containing the indicated unnatural amino acid at the specified position. Average mass of the protein is 19378 Da. Protein ladder in kDa.

SDS-PAGE analysis of the three modified protetins and the wild-type protein demonstrated that each was successfully expressed and purified. Protein mass spectrometry was again used to measure the accurate mass of each protein (**Figure 55**). Analysis by mass spectrometry revealed that each mutant had the amino acid successfully incorporated. As with the DPK variants of LmrR, no truncation of the amino acid was observed. Each enzyme was then concentrated to a final concentration of 2 mg/mL. This time, the hydride transfer reaction was performed using DHFR's natural cofactor, NADPH.



Mutant	Calculated mass (g/mol)
Ala7-DPK	19439
Phe31-DPK	19363
Ser49-DPK	19424

Calculated mass of wild-type EcDHFR – N-terminal Met = 14284 g/mol

Figure 55. Deconvoluted ESI mass spectra of wild-type EcDHFR and each mutant, determined from the protein mass spectrometry charged state.

To screen the enzyme variants, 5 µL of a stock solution of cinnamaldehyde in methanol was added to the enzyme followed by 5 µL of a stock solution of NADPH in the reaction buffer. This resulted in a catalyst loading of 10 mol% with respect to cinnamaldehyde and an excess of the hydride donor (2 eq to cinnamaldehyde). The reactions were placed in a thermoshaker for 18 hours at 25 °C. The substrates and products from each reaction were extracted with 200 µL of DCM and analysed by gas chromatography mass spectrometry (GC-MS) (**Figure 56b**).

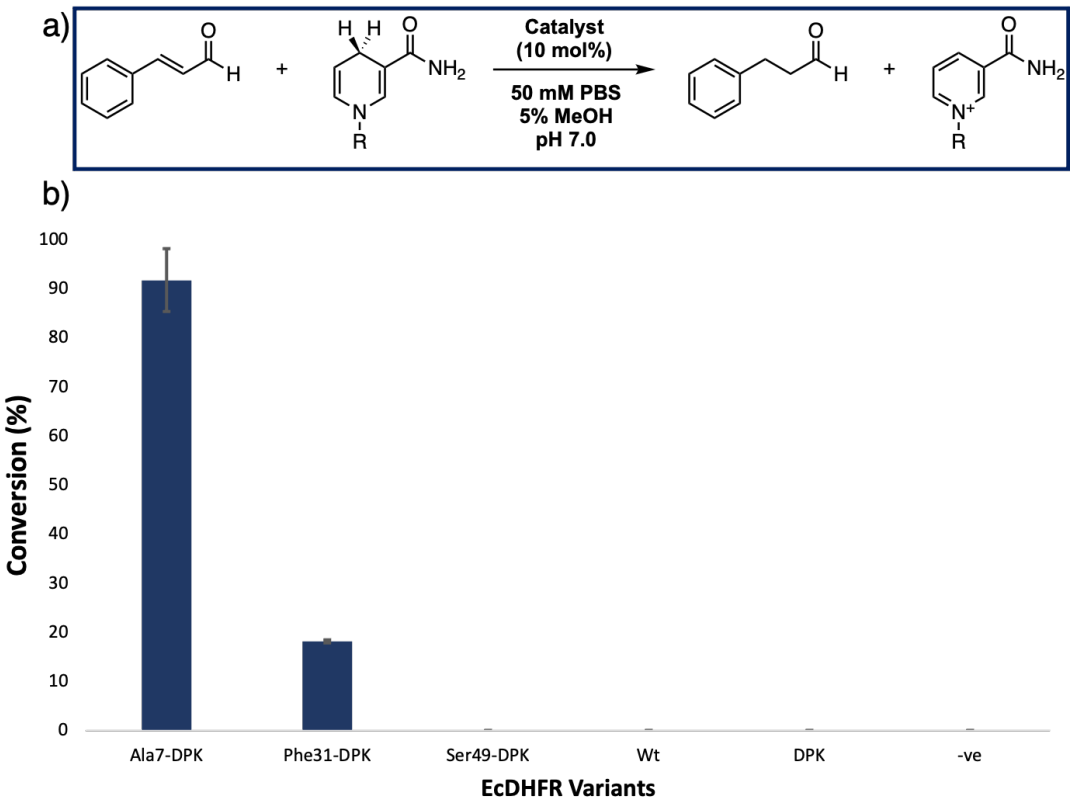


Figure 56. a) Selective reduction of cinnamaldehyde by NADPH to hydro-cinnamaldehyde. **b)** Peak area of the hydride transfer product hydro-cinnamaldehyde. for each of the EcDHFR based artificial enzymes, the wild-type EcDHFR, DPK and the substrates alone under the same reaction conditions. Reaction conversion was determined by comparison to a standard calibration curve for the product hydro-cinnamaldehyde. Each chart entry is an average of three repeats. Error bars are a representation of the standard deviation.

It was clear from the conversion determined by GC-MS, that the Ala7 mutation had the greatest catalytic potential with an average conversion of (90%). This is the residue located at the end of the dynamic loop. Phe31 also showed some activity. Upon inspection of the Ser49 mutant to sample work-up, substantial protein precipitation had appeared, so this could be the reason for no product detected. Neither the wild-type, negative control or DPK alone showed any measurable product.

To approximate the catalytic efficiency of the two most successful enzymes, LmrR Phe93DPK and EcDHFR Ala7DPK were further purified by size exclusion chromatography (**see section 6.3**). To assess the preliminary rate of the artificial enzymes, attempts were made to measure the kinetics of the reaction. An end-point assay was developed where conversion of the substrate to the product over time was used to determine a reaction rate which could be fitted to the Michaelis-Menten equation. Product conversion was determined by quantitative gas chromatography flame ionisation detection (GC). Saturating conditions of cinnamaldehyde were

determined by independently doubling the substrate concentration. LmrR was incubated with 1 mM of cinnamaldehyde and 5 mM of BNAH and time points taken to estimate the initial rate (linear phase) of the reaction (**Figure 57**). Two hours was determined to be the top off the linear phase under these conditions.

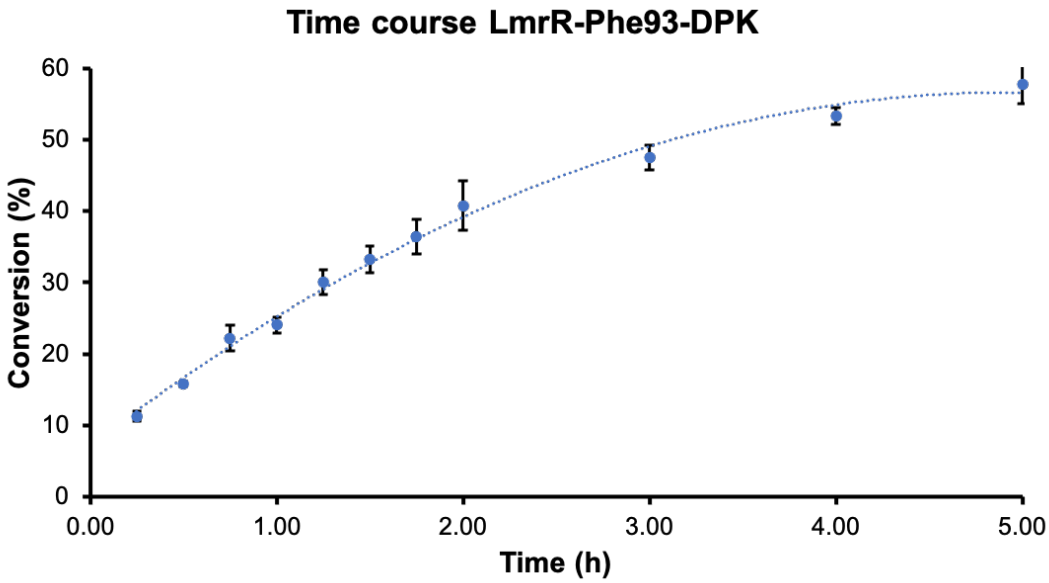


Figure 57. Time course chart measuring the reaction of LmrR-Phe93-DPK as a catalyst converting cinnamaldehyde to hydro-cinnamaldehyde. 50 μM of LmrR was incubated with 1 mM of cinnamaldehyde and 5 mM of BNAH. The reactions were quenched at each time point with the addition of DCM and the conversion of substrate to product determined using GC-FID. Each time point was performed three times and the mean reported. Error bars are a representation of the standard deviation.

As the end-point assay required a fair amount of product to be generated to be detected, the top of the linear phase (2 hours) was chosen to measure the substrate dependant initial reaction rate. Next the hydride donors for each respective enzyme were modified to determine the apparent k_{cat} and K_M. (**Figure 58**).

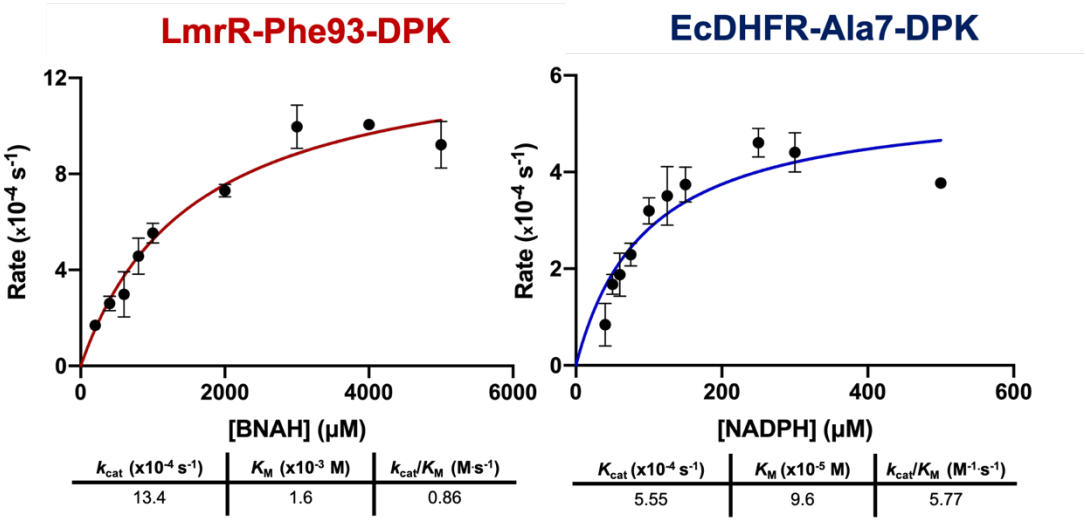


Figure 58. Kinetic characterisation of LmrR Phe93DPK and EcDHFR Ala7DPK. Assays were performed using 50 μM of LmrR and 20 μM of EcDHFR. The assays were performed using saturating concentrations of cinnamaldehyde (1 mM) and varying concentrations of BNAH for LmrR and NADPH for EcDHFR. The rate constants for LmrR were fitted to a Michaelis-Menten curve and Vmax, k_{cat}, K_M and k_{cat}/K_M determined using the software Prizm. The rate constants for EcDHFR were fitted to an allosteric sigmoidal curve relevant to the Hill equation using the software Prizm.

The k_{cat} of LmrR Phe93DPK was determined to be 13.4 x10^{-4} s^{-1} and for EcDHFR Ala7DPK, 5.55 x10^{-4} s^{-1}. This finding suggests that LmrR Phe93DPK has a higher reaction rate than the EcDHFR mutant, and this could likely be due to the hydrophobic property of the protein. Cinnamaldehyde may locate inside the hydrophobic binding pocket faster than it enters the active site of EcDHFR. The apparent $K_{M[BNAH]}$ for LmrR Phe93DPK was determined to be 1.6 x10^{-3} M. The catalytic efficiency (k_{cat}/K_M) of LmrR Phe93DPK was determined to be 0.86 $M^{-1}{\cdot}s^{-1}$. The apparent $K_{M[NADPH]}$ for EcDHFR Ala7DPK was determined to be 9.6 x10^{-5} M. The catalytic efficiency (k_{cat}/K_M) of EcDHFR Ala7DPK was determined to be 5.77 $M^{-1}{\cdot}s^{-1}$ **(Figure 58)**.

The end-point assay used to determine the kinetic characteristics of the artificial enzymes allows only an approximation partly due to the limitation of product detection. Catalytic efficiency of the enzymes in future can be more accurately determined following development of a continuous assay but it is hoped that this preliminary data can aid as a starting point.

Following kinetic analysis, a further 7 α,β-unsaturated carbonyl compounds were tested in the same reaction. NMR was employed to measure the conversion of the parent carbonyl compounds to their saturated products. The conversions were determined as per mol% of enzyme used (**Figure 59**).

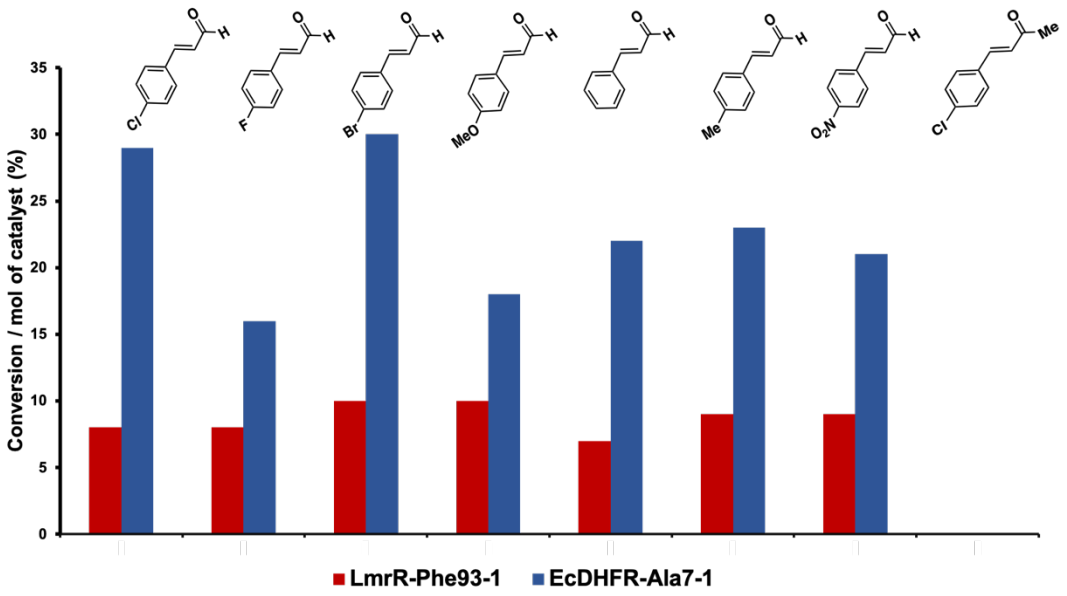


Figure 59. Substrate scope of the artificial enzymes LmrR Phe93DPK and EcDHFR Ala7DPK. Conversion determined by ^{1}H NMR by comparison of the integrals arising from the aldehyde protons of the substrates and products and normalised per mol% of enzyme used (LmrR = 10 mol% and EcDHFR = 1 mol%) in the reaction with respect to the carbonyl substrate. Reactions were performed in 50 mM PBS buffer (pH 7.0) for 48 hours at 25 °C with constant agitation. Each entry is an average of three repeats.

The substrate scope of the artificial enzymes clearly shows a fair turnover of product in 48 hours. The ketone substrate however did not show any reactivity, and this is likely due to low solubility in the reaction medium and lower electrophilicity of the carbonyl carbon. In comparison to LmrR, EcDHFR showed a high conversion to product per mol% of enzyme used. Analysis also reveals the promiscuity of the two enzymes. Incorporation of the unnatural amino acids immediately allowed for a biocatalytic reaction that can accommodate a number of substrates without lengthy rational mutation or directed evolution assays.

EcDHFR naturally recruits NADPH to perform the hydride transfer reaction. This property of the organocatalytic protein can be exploited by adding a cofactor recycling scheme. The ability to perform these reactions in a biocompatible environment gives access to the use of natural enzymes in conjunction with the artificial enzyme. Subsequently The enzyme glucose-6-phosphate dehydrogenase was coupled with the EcDHFR Ala7DPK to drive the reaction forward to generate the product hydro-cinnamaldehyde (**Figure 60a**). To accomplish this the artificial enzyme was incubated with cinnamaldehyde, glucose-6-phosphate and glucose-6-phosphate dehydrogenase, in the presence of NADPH and the product of the reaction quantitatively determined (**Figure 60b**).

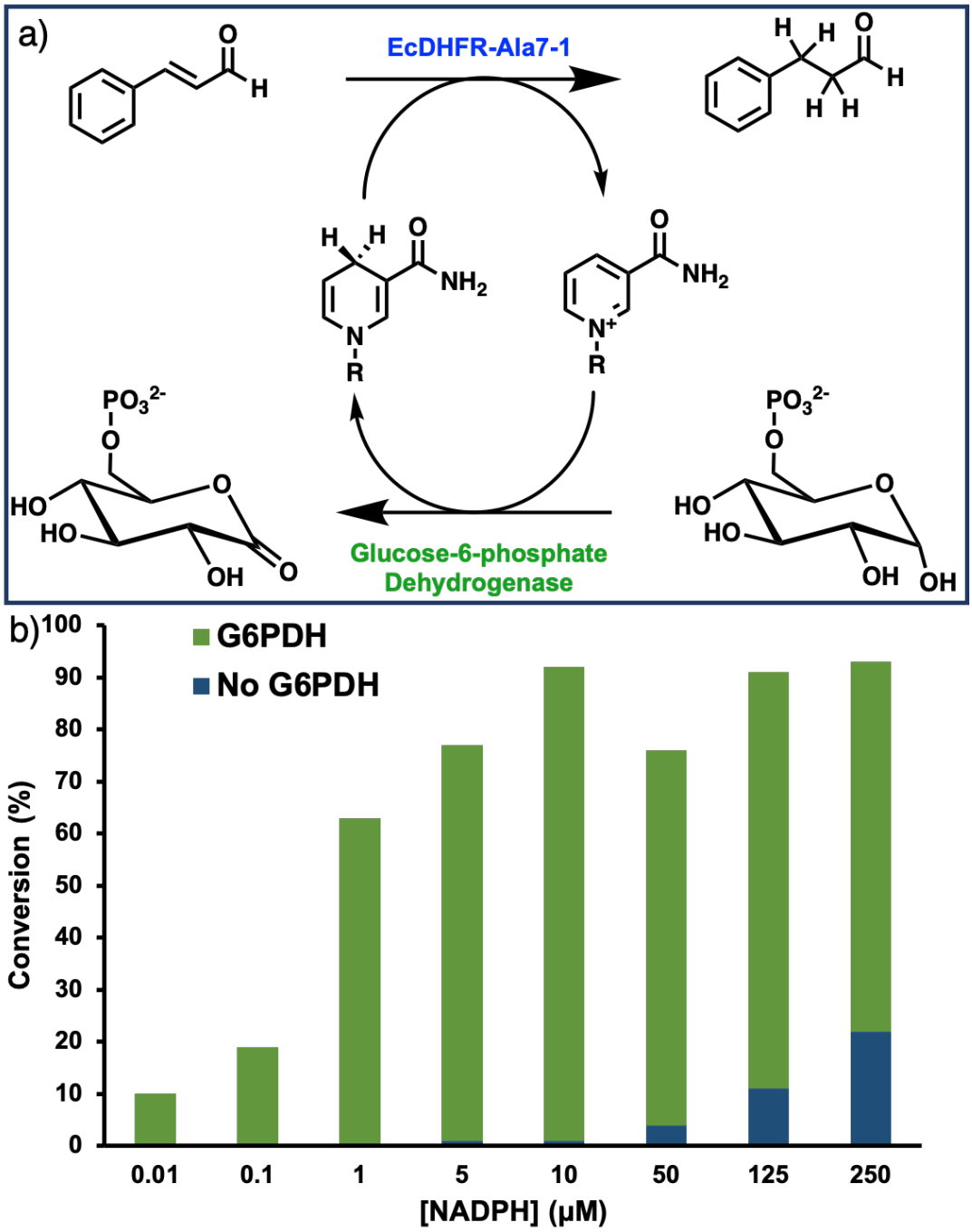


Figure 60. a) Cofactor recycling scheme employed to regenerate NADPH *in situ*, during the iminium ion catalysed hydride transfer reaction catalysed EcDHFR Ala7DPK. **b)** conversion of cinnamaldehyde to hydro-cinnamaldehyde with and without the cofactor recycling scheme.

EcDHFR Ala7DPK was incubated with 1 mM of cinnamaldehyde with 10 mol% catalyst loading and 50 nM of glucose-6-phosphate dehydrogenase and 2 M of glucose-6-phosphate. The concentration of NADPH was then varied to determine the total turnover number (**Table 6**). The reaction was performed in PBS buffer with 5% methanol allowing both reactions to proceed seamlessly. The reactions were performed for 18 hours and the products extracted and analysed by GC.

Table 6. Substrate to product conversion for the selective reduction of cinnamaldehyde to hydro-cinnamaldehyde using NADPH as the hydride source in the presence of glucose-6-phosphate and glucose-6-phosphate dehydrogenase. Total turnover numbers calculated by dividing moles of product by moles of cofactor added to the reaction. Reactions were performed in 50 mM PBS buffer with 5% methanol and 10 mol% catalyst loading. Substrate conversion and TTN are and average of three repeats.

NAPDH Concentration (μM)	Conversion (%)	Total turnover number (TTN)
0.01	10	10460
0.1	19	1880
1.0	63	632
5.0	76	153
10	91	91
125	72	14
250	80	6

With the addition of the cofactor recycling components high conversions and total turnover numbers could be accomplished. The trait of EcDHFR to bind a natural cofactor allows the integration of the natural enzymes to improve the efficiency of organocatalysed reactions.

As NADPH is present in all living systems, it was prudent to determine if EcDHFR Ala7DPK was able to catalyse the artificial reaction in cell lysate. Addition of cinnamaldehyde to lysed cells that had confirmed expression of the artificial enzyme and a control sample of BL21 (DE3) cultured cells was performed. The products were extracted and analysed by GC-MS (**Figure 61**).

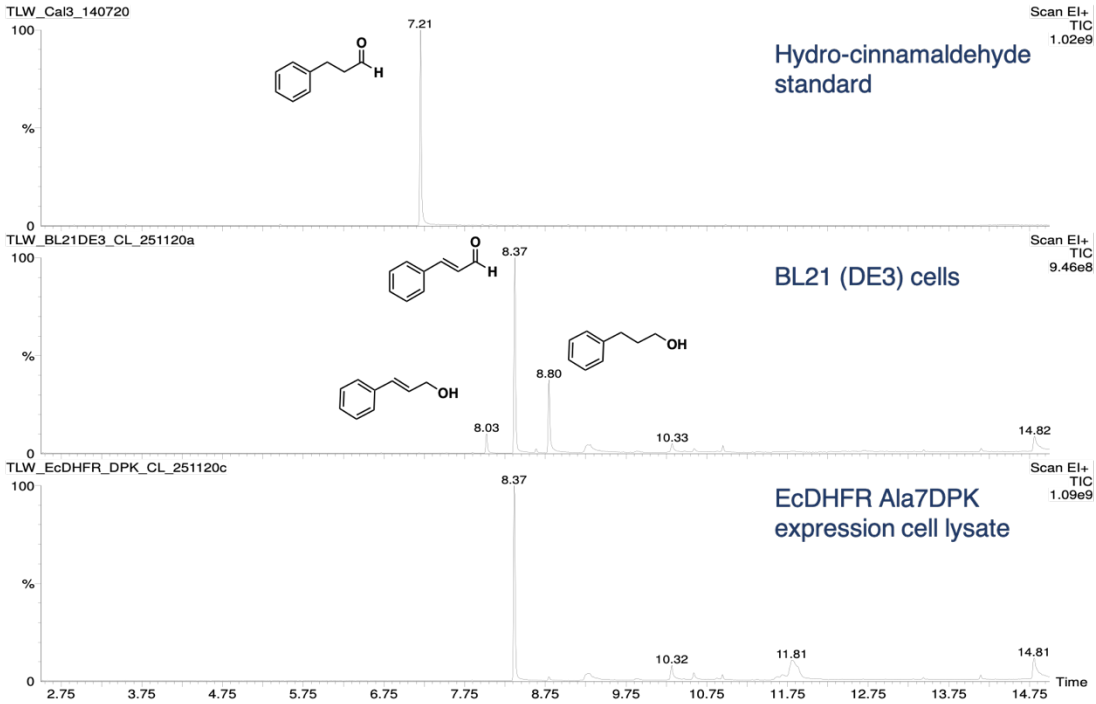


Figure 61. GC-MS chromatograms produced from the extraction of the organic compounds from the whole cell lysate catalysis using EcDHFR Ala7DPK. The desired product hydro-cinnamaldehyde has a retention time of 7.21 minutes. BL21 (DE3) control cells extracted identified product compounds are cinnamyl alcohol at retention time 8.03 minutes and hydro-cinnamyl alcohol at 8.80 minutes. The major extractable compound is cinnamaldehyde at retention time 8.37 minutes. The major extractable product in the cells expressing EcDHFR Ala7DPK was the substrate cinnamaldehyde.

Upon analysis by GC-MS of the extractable products, no trace of the intended product could be detected in either the cells expressing EcDHFR Ala7DPK or the control cells. As the enzyme requires an aldehyde substrate, the reduction of the aldehyde to the alcohol will render the substrate inactive to the designer enzyme. It is likely that the low concentration of protein combined with relatively low rate compared to native enzymes in the cell that can act upon cinnamaldehyde, will prove difficult to perform this reaction using whole cell lysate under current conditions.

The design of these enzymes was intended for iminium ion formation followed by nucleophilic hydride transfer. Strong evidence to support this mechanism is pivotal for proof of this *de novo* enzyme design. If the reaction follows an iminium catalysed cycle, then it is clear that the application may very well be far reaching. Just as the small molecule organocatalysts were applied to a far broader reaction profile after their initial study, so to may any successful design that arises from this work. The two artificial enzymes were incubated overnight with 10 equivalents of cinnamaldehyde at 25 °C. The following day 20 equivalents of sodium cyanoborohydride was added and the protein incubated for 2 hours. The proteins were then digested with

chymotrypsin and analysed by high resolution liquid chromatography mass spectrometry (LC-MS). Digestion yielded both expected peptides and an elemental composition within 1 ppm was determined between the calculated and measured mass (**Figure 62 and 63**).

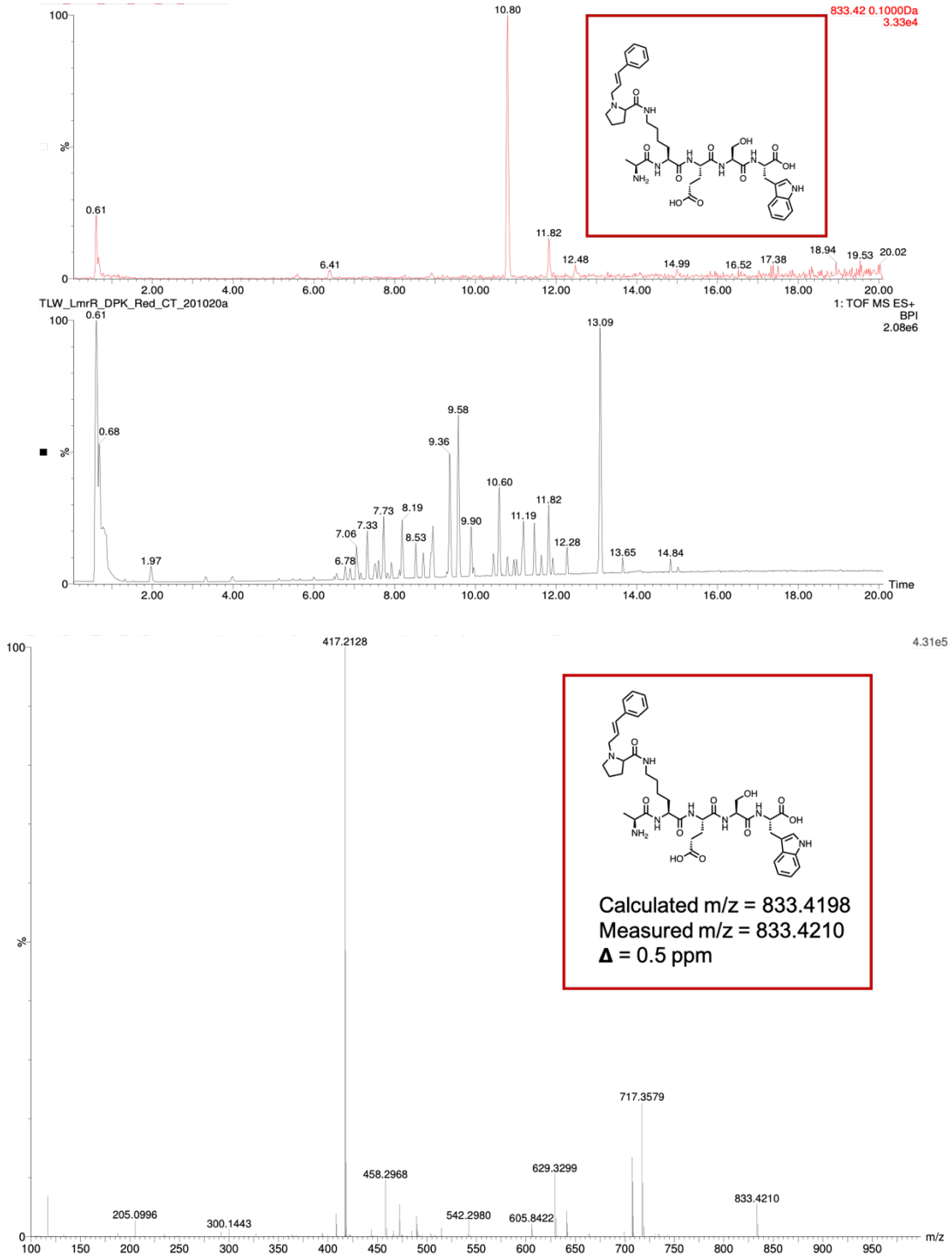


Figure 62. LC-MS analysis of the chymotrypsin digest of LmrR Phe93DPK. Extraction of the predicted peptide mass that contains the unnatural amino from the peptide digests resulted in

exact mass identification of the predicted peptide mass with DPK modified with cinnamaldehyde.

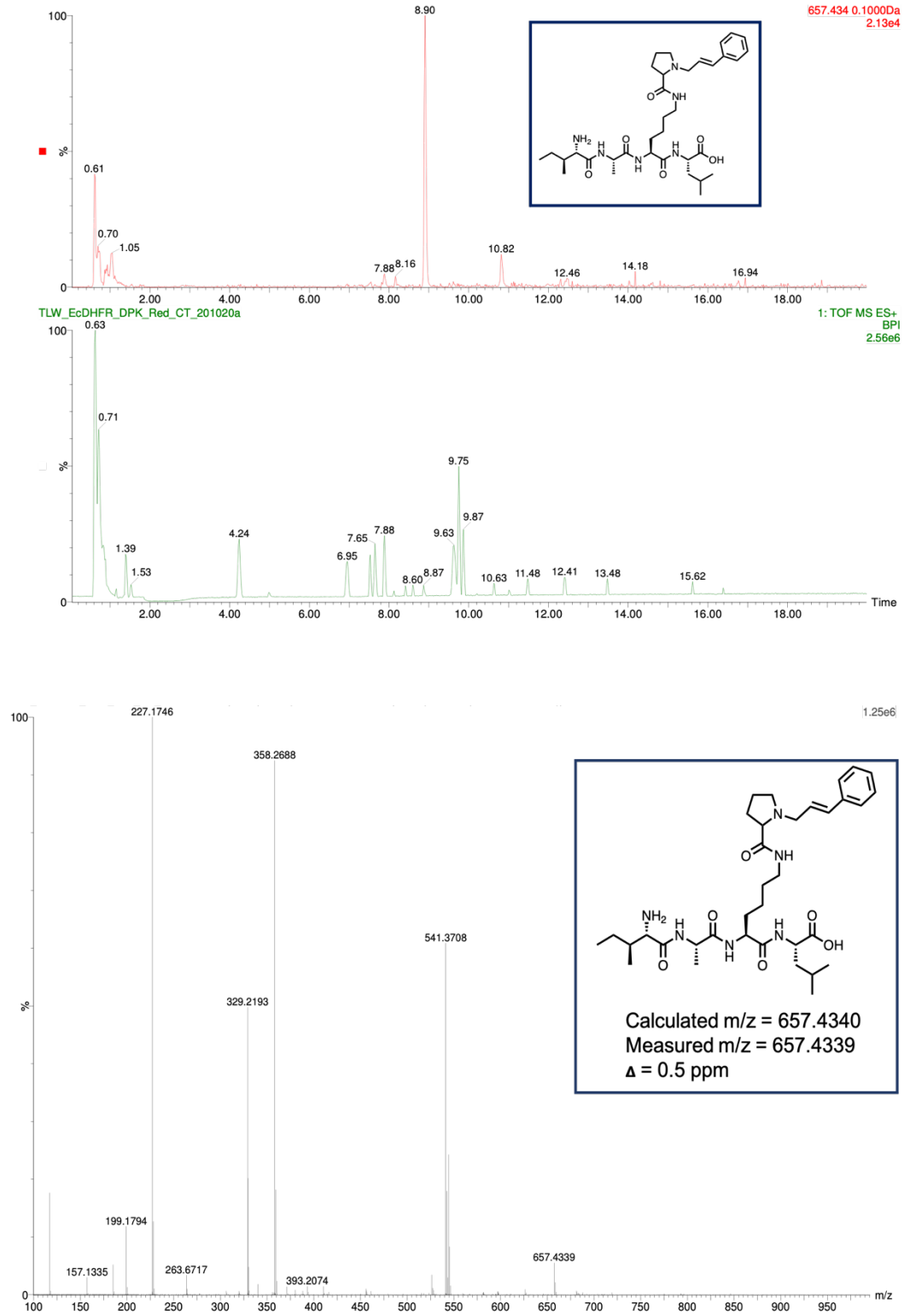


Figure 63. LC-MS analysis of the chymotrypsin digest of EcDHFR Ala7DPK. Extraction of the predicted peptide mass that contains the unnatural amino from the peptide digests resulted in

exact mass identification of the predicted peptide mass with DPK modified with cinnamaldehyde.

Following confirmation of iminium ion formation, the hydride transfer was next probed using an isotopomer of NADPH. The mechanism of wild-type EcDHFR selectively transfers the Pro-*R* hydride of C4 of the nicotinamide ring of NADPH to the saturated imine in the pterin ring of dihydrofolate (**Figure 64**). The oxidised form of NADPH, $NAPD^+$, can be reduced selectively to place a deuterium atom at the pro-*R* position. This will allow the determination of hydride transfer specificity. In addition, if the position of the deuteron can be revealed in the products structure, this may provide strong evidence in the hydride transfer predominantly following an iminium ion hydride transfer mechanism.

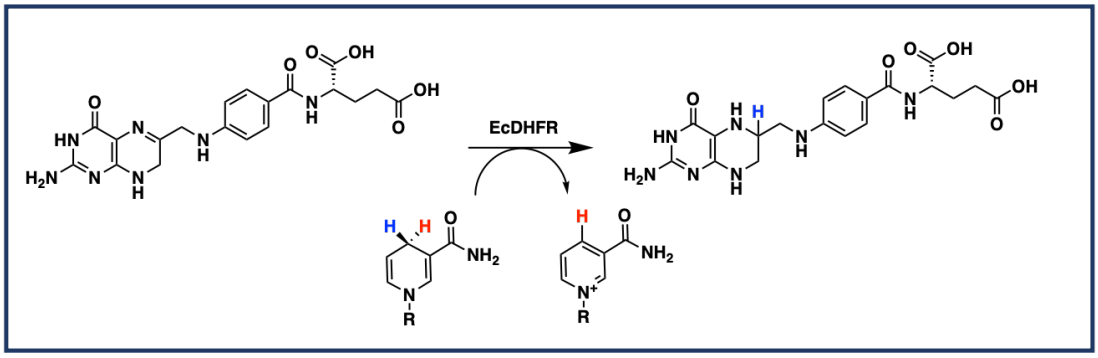


Figure 64. Natural mechanism of EcDHFR. The saturated imine of the pterin ring of dihydrofolate is reduced to form tetrahydrofolate using the Pro-*R* hydride of NADPH.

To selectively reduce the oxidised $NADP^+$ to generate the molecular probe, an enzymatic reaction was employed. An alcohol dehydrogenase (TbADH) from *Thermoanaerobacter brokii* was recombinantly expressed in *E. coli* and purified according to a literature procedure. Using deuterated iso-propyl alcohol and TbADH as catalyst, $NADP^+$ was selectively reduced to generate the deuterated NADPD.[15] Confirmation of enrichment by deuterium was determined by high resolution mass spectrometry (**Figure 65**).

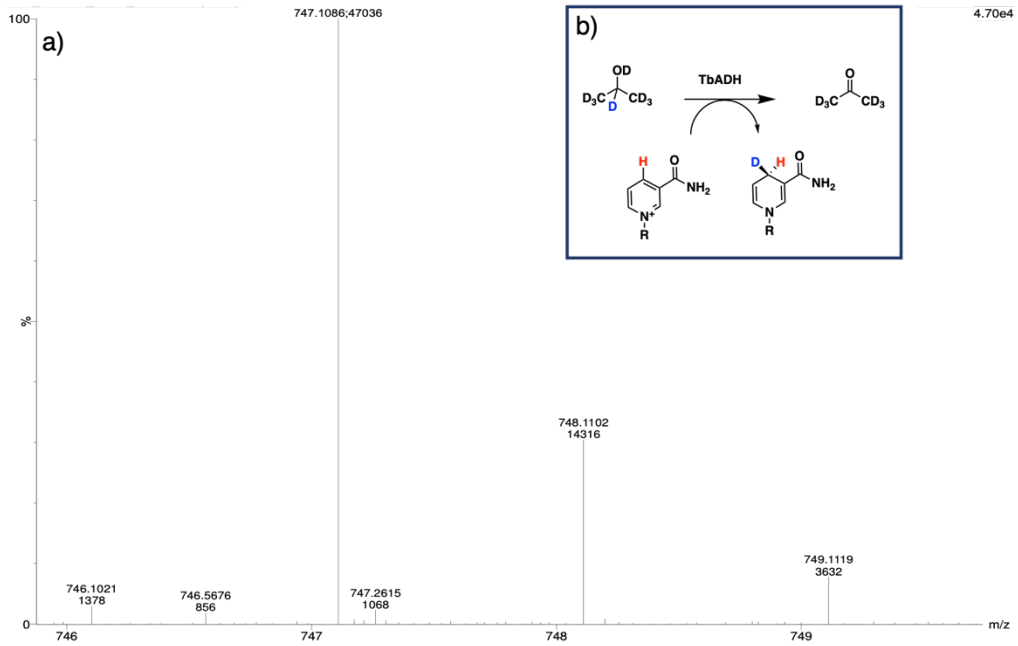


Figure 65. a) HR-MS analysis of the enzymatically synthesised NADPD. Calculate m/z for NADPD = 747.1052 **b)** The enzymatic reaction used to install a deuterium atom at the Pro-*R* hydride of NADPH.

Next the isotopomer NADPD was used as the hydride source in the reduction of the C=C bond of cinnamaldehyde using EcDHFR Ala7DPK as the catalyst. Simultaneously natural abundance NADPH was used in a second reaction. GC-MS analysis was employed to assess the incorporation of deuterium into the product hydro-cinnamaldehyde (**Figure 66**).

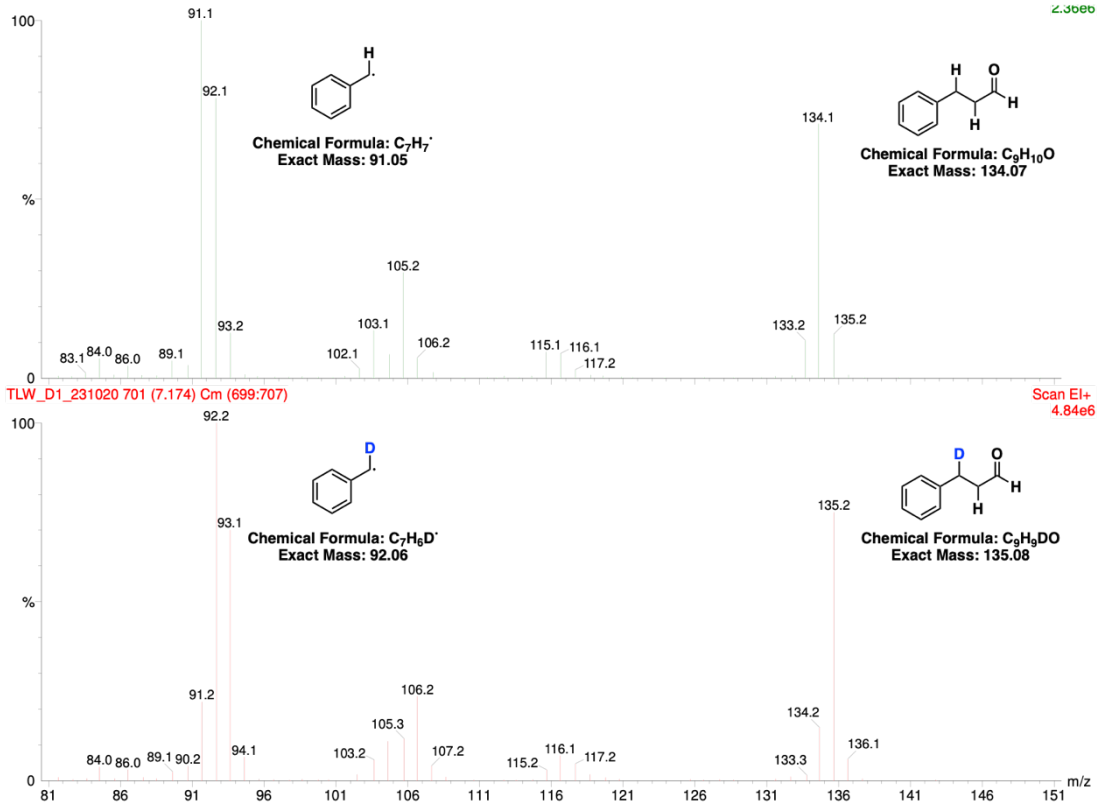


Figure 66. GC-MS analysis of the products of the reaction between cinnamaldehyde and NADPH and NADPD. Incorporation of a deuterium atom results in a mass increase of 1 Da. Fragment analysis of the radical formed upon electron impact ionisation suggests that the deuterium is transferred to the gamma carbon of the substrate.

The GC-MS analysis of the product revealed that that the hydride source originated from the pro-*R* hydride of NADPH in the same way as it does in EcDHFR's natural mechanism. In addition, fragment analysis produced from the source ionisation reveals that the position that the hydride is transferred to is that of the gamma carbon of cinnamaldehyde, as proposed for an iminium ion catalysed mechanism utilising an amine (**Figure 67**).

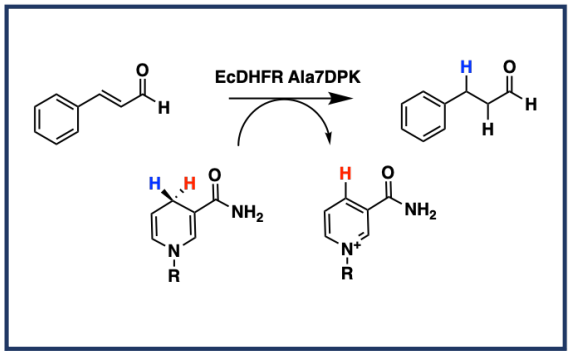


Figure 67. Selective transfer of the pro-*R* hydride of NADPH to the gamma-carbon of cinnamaldehyde by EcDHFR Ala7DPK.

With evidence for both the iminium ion formation as well as the source and trajectory of the hydride ion revealed by mass spectrometry analysis, a detailed mechanism of the artificial enzyme EcDHFR Ala7DPK may be proposed for selective reduction of α,β-unsaturated aldehydes by NADPH (**Figure 68**).

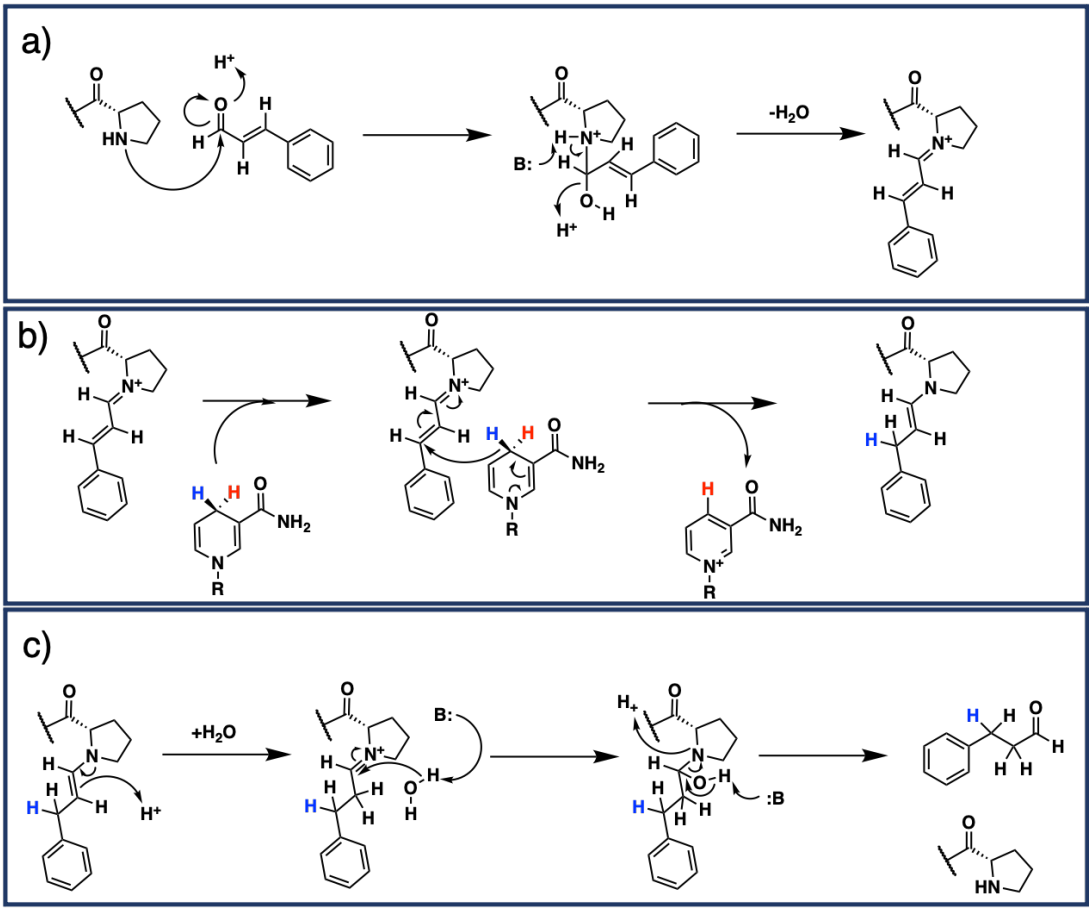
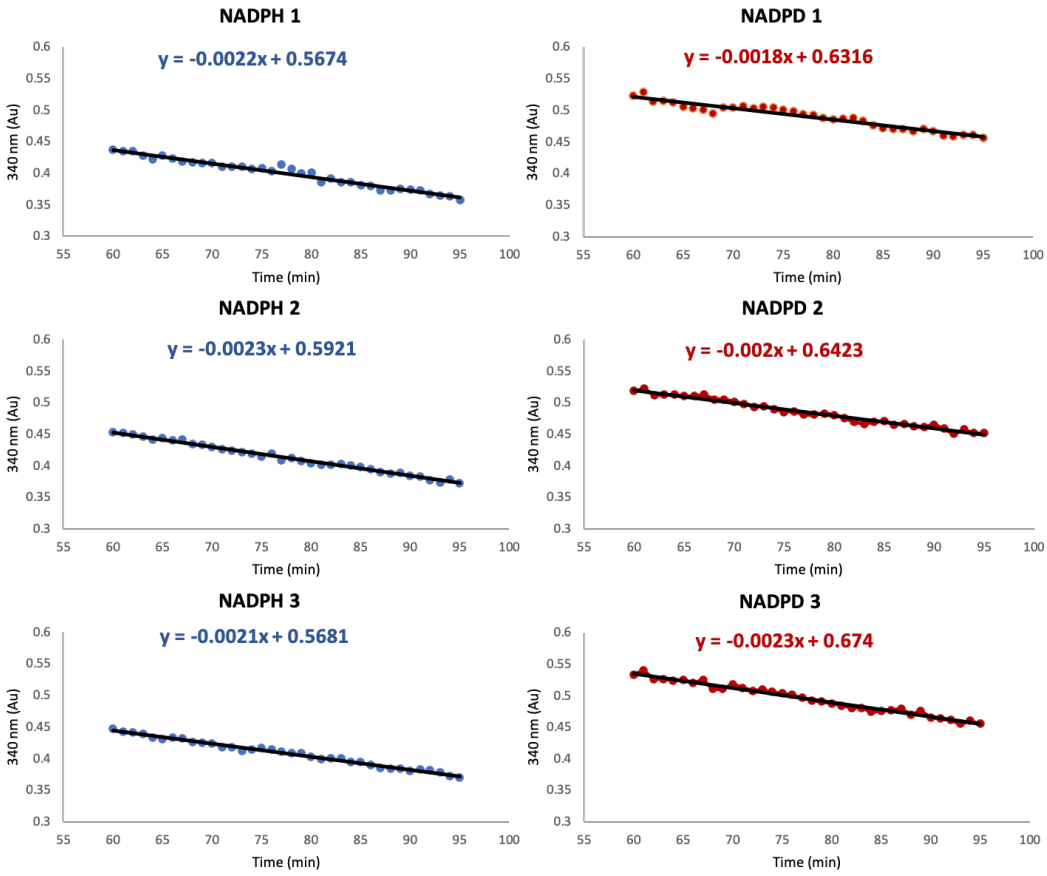


Figure 68. Proposal of the mechanism of selective reduction of α,β-unsaturated carbonyl compounds by EcDHFR Ala7DPK using NADPH as the cofactor**. a)** Formation of the iminium ion between the carbonyl compound and the secondary amine unnatural amino acid. **b)** Transfer of the Pro-*R* hydride of NADPH to the gamma-carbon of the α,β-unsaturated aldehyde and formation on an enamine. **c)** Protonation of the enamine and hydrolysis of the resulting iminium ion producing the saturated product and oxidised $NADP^+$.

The proposed mechanism as outlined above (**Figure 68**) is as follows. First the amine of the prolyl-ring of DPK performs a nucleophilic attack on the carbonyl carbon of the aldehyde and following the elimination of water results in an iminium ion (**Figure 68a**). Next the Pro-*R* hydride of NADPH is transferred in a nucleophilic addition to the gamma-carbon of the iminium ion forming an enamine (**Figure 68b**). The enamine is then protonated at the beta carbon resulting in a second iminium ion. This iminium ion is then hydrolysed from DPK by water resulting in the saturated carbonyl product (**Figure 68b**).

Now that a reasonable mechanism can be proposed further investigations into the artificial enzyme's properties can be probed. The formation of the iminium ion and the transfer of the hydride can be investigated using kinetic isotope effects (KIEs). The use of KIEs for probing the rate determining step in the reaction can aid in future designs of this type of enzyme. For example, should iminium ion formation be rate limiting then increase in catalytic competency may possibly be accomplished through introduction of secondary amines with increased nucleophilicity. If the hydride transfer is rate limiting, then movement of the incorporated amino acid may be beneficial bringing the reactive intermediate into close proximity to the hydride source. Either way, assessing the rate limiting step in this enzymes reaction is of fundamental interest in how this reaction can be achieved.

To measure if the transfer of the hydride has a kinetic isotope effect, a comparative assay was performed under the same reaction conditions using either NADPH or NADPD as the hydride source. The reaction was monitored by measuring the decrease in UV absorbance at 340 nm (**Figure 69**).



	Exp 1.	Exp. 2	Exp. 3	Average	Std. Deviation
NADPH (Slope)	0.0022	0.0023	0.0021	0.0022	0.0001
NADPD (Slope)	0.0018	0.002	0.0023	0.0020	0.0003
KIE	1.22	1.15	0.92	1.10	0.16

Figure 69. Rate curves displaying the oxidation of NADPH (blue) and NADPD (red) in the reaction catalysed by EcDHFR Ala7DPK. UV absorption at 340 nm was used to monitor product turnover. Table displaying the determination of the kinetic isotope effect (KIE).

By determination of the slope of the absorbance curve an assessment can be made as to the rate of the reaction. In both instances, using NADPH or NADPD the slope remains largely the same. Division of the slope arising from NADPH by that of NADPD produces a kinetic isotope effect of 1.10 (± 0.16). This suggests that hydride transfer is not the rate limiting step in the reaction. This observation is also observed with EcDHFR's natural mechanism under physiological conditions. If hydride transfer is not rate limiting, then another step in the enzymes mechanism must account for the rate limitation. This maybe be formation of the iminium ion, hydrolysis of the product or protonation of the enamine. Equally possible is the dynamic movement that

EcDHFR uses to release the oxidised cofactor. This is often cited as the rate limiting in natural catalysis and so it is plausible that the same phenomenon is observed here.

3.4 Conclusion

Genetic code expansion has proven to be a powerful tool to introduce new reactivates into proteins. Three secondary amine unnatural amino acids were incorporated into four different positions in the protein backbone LmrR and screened for their ability to selectively reduce α,β-unsaturated carbonyl compounds, using the synthetic hydride donor BNAH. One variant, LmrR Phe93DPK was able to facilitate the reaction competently. The amino acid DPK was then placed in the backbone EcDHFR at three positions and the hydride transfer reaction performed using NADPH as the hydride donor with the position Ala7 showing the most promise. Mass spectrometry analysis demonstrated the ability of the unnatural amino acid to form an iminium ion intermediate, as was planned in the *de novo* design. Kinetic analysis was able to demonstrate the rate and catalytic efficiency of the designer enzymes, and application to a further seven substrates demonstrated the enzymes promiscuity.

This chapter highlights the advantages of using genetic code expansion in enzyme design. The ability to maintain the same catalytic moiety, but to move around the protein architecture allows a novel approach in rational enzyme engineering. Instead of having to maintain the same environment for the intended catalysis, and randomly mutating around this catalytic spot to tune the environment, one simple mutation can result in a dramatic difference. With relevance to LmrR, the movement of DPK from one position to another, results in far greater activity. Even more impressive, to increase the catalytic activity for the desired reaction, the catalyst was simply transferred to a new protein that has a lower K_M for a hydride donor. This resulted in almost 10 times the catalytic efficiency.

The use of natural NADPH as a cofactor lays the foundations of using secondary amine organocatalysis in chemical biology applications. Thus far there is no precedent of using secondary amines to catalyse reactions with structurally complex biologically derived molecules. This is a step forward to bringing organocatalytic applications closer to biological applications.

Mechanistic studies of the artificial enzyme EcDHFR provides strong evidence of an iminium catalysed reaction utilising the unnatural amino acid. This finding may allow

broad application of this type of artificial enzyme in applications beyond hydride transfer catalysis. Many organocatalytic reactions may be able to be catalysed in this way and so this brings the field of secondary amine organocatalysis closer to the ability to be used in chemical biological applications.

In comparison to previous reports on enzymes designed using this strategy, there is larger possibilities available with iminium ion catalysed reactions compared to oxime and hydrazone ligations and ester hydrolysis. In reference to artificial enzymes that are already able to perform secondary amine catalysis (4-OT tautomerase and the streptavidin-biotin system) this system offers far greater flexibility. In essence this chapter has demonstrated the broad applicability of merging genetic code expansion and secondary amine catalysis in the design of artificial enzymes.

3.5 References

1. Mayer, C.; Dulson, C.; Reddem, E., et al., Directed Evolution of a Designer Enzyme Featuring an Unnatural Catalytic Amino Acid. *Angew. Chem. Int. Ed.* **2019,** *58* (7), 2083-2087.
2. Drienovska, I.; Mayer, C.; Dulson, C., et al., A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **2018,** *10* (9), 946-952.
3. Burke, A. J.; Lovelock, S. L.; Frese, A., et al., Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **2019,** *570* (7760), 219-223.
4. Bertelsen, S.; Jorgensen, K. A., Organocatalysis--after the gold rush. *Chem. Soc. Rev.* **2009,** *38* (8), 2178-2189.
5. MacMillan, D. W., The advent and development of organocatalysis. *Nature* **2008,** *455* (7211), 304-308.
6. Chin, J. W., Expanding and reprogramming the genetic code. *Nature* **2017,** *550* (7674), 53-60.
7. Agustiandari, H.; Lubelski, J.; van den Berg van Saparoea, H. B., et al., LmrR is a transcriptional repressor of expression of the multidrug ABC transporter LmrCD in Lactococcus lactis. *J. Bacteriol.* **2008,** *190* (2), 759-63.
8. Madoori, P. K.; Agustiandari, H.; Driessen, A. J., et al., Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition. *EMBO J.* **2009,** *28* (2), 156-66.
9. Luk, L. Y.; Javier Ruiz-Pernia, J.; Dawson, W. M., et al., Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (41), 16344-9.
10. Luk, L. Y.; Loveridge, E. J.; Allemann, R. K., Protein motions and dynamic effects in enzyme catalysis. *Phys. Chem. Chem. Phys.* **2015,** *17* (46), 30817-27.
11. Polycarpo, C. R.; Herring, S.; Berube, A., et al., Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **2006,** *580* (28-29), 6695-700.
12. Nguyen, D. P.; Elliott, T.; Holt, M., et al., Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* **2011,** *133* (30), 11418-21.
13. Roelfes, G., LmrR: A Privileged Scaffold for Artificial Metalloenzymes. *Acc. Chem. Res.* **2019,** *52* (3), 545-556.
14. Ismail, M.; Schroeder, L.; Frese, M., et al., Straightforward Regeneration of Reduced Flavin Adenine Dinucleotide Required for Enzymatic Tryptophan Halogenation. *ACS Catal.* **2019,** *9* (2), 1389-1395.
15. Loveridge, E. J.; Allemann, R. K., Effect of pH on hydride transfer by Escherichia coli dihydrofolate reductase. *ChemBioChem* **2011,** *12* (8), 1258-62.

4. Investigating the Effects of N-terminal Mutations in *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase

4.1 Preface

The introduction of unnatural amino acids through genetic code expansion has become a pivotal technology in chemical biology. From bio-conjugation to generating novel biocatalysts and everything in between, the applications are wide-reaching and substantial. Pyrrolysyl-tRNA synthetases from Archaea have become the principal tools in the development of this technology. The two homologs from *Methanosarcina mazei* and *Methanosarcina bakeri* are sequentially and structurally similar. These two enzymes are often interchanged with each other and mutations at the catalytic C-terminus have been shown to be successfully translated from one homolog to the other. Beneficial mutations in the N-terminus on the other hand, have not been investigated with regards to movement from one protein to the other. Here, beneficial mutations found in *M. mazei* Pyrrolysyl-tRNA synthetase at the N-terminus are probed and investigated in its homolog originating from *M. bakeri*.

This chapter contributed to the publication: Williams, T. L.; Iskandar, D. J.; Nödling, A. R., et al., Transferability of N-terminal mutations of pyrrolysyl-tRNA synthetase in one species to that in another species on unnatural amino acid incorporation efficiency. *Amino Acids* **2020**.

4.2 Introduction

Genetic code expansion has become a pivotal technology in chemical biology. The technique allows for the site-specific incorporation of unnatural amino acids (also known as non-canonical amino acids) into a protein backbone using the cells translational machinery.[1, 2] The incorporation of unnatural amino acids, beyond the 20 proteinogenic amino acids, through genetic code expansion has opened up a vast avenue of applications ranging from investigating protein-protein interactions and probing post translational modifications, to site-selectively modifying proteins with small molecules and controlling the function of proteins in live cells and animals.[3]

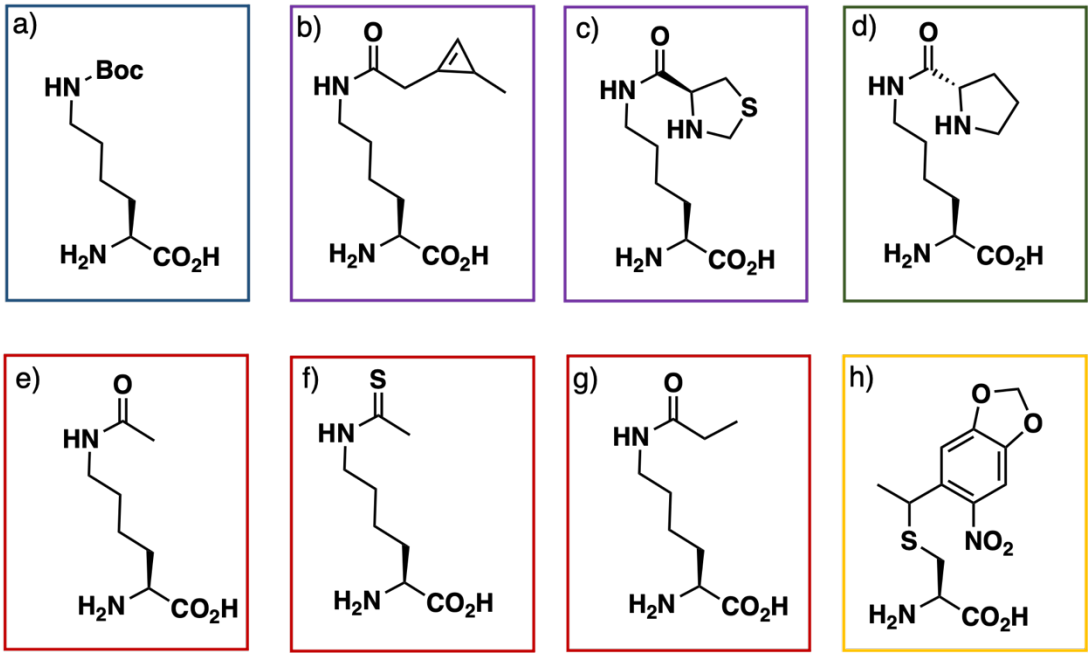


Figure 70. A selection of unnatural amino acids that have been successfully incorporated into a protein using genetic code expansion. **a)** Boc lysine often used for initial studies. **b)** Cyclopropene lysine and **c)** thiazolidine lysine used for site-selective bioconjugation **d)** D-prolyl lysine an analogue of pyrrolysine and catalytically competent. **e)** acetyl lysine**, f)** propionyl lysine and **g)** thioacetyl lysine used for mimicking post-translational modifications. **h)** Photocaged cysteine used in protein activation and control.

Bioconjugation of small molecules to proteins has been extensively exploited using genetic code expansion.[4] The incorporation of unnatural amino acids allows for site-specific reactions to occur at just one residue. Non-biological functional groups can be installed in the protein through the unnatural amino acid that undergo bio-orthogonal reactions, thereby allowing a single modification on a protein backbone.[5] Copper catalysed click chemistry through azide-alkyne coupling is a dominant field where either the azide or the alkyne group have been demonstrated to be incorporated.[6] Recently, catalyst free reactions have been developed allowing for

protein modification in live cells. These include the incorporation of amino acids that bear strained alkenes, such as a cyclopropene group (**Figure 70b**), and alkynes to perform copper free click reactions through inverse electron demand Diels-Alder reactions and strain-promoted alkyne-azide cycloadditions.[7, 8] Further strategies employ the use of amino acids that are furnished with protected functional groups that can be activated and functionalised at a specific time. Thiazolidine (**Figure 70d**) is one such functional group that can be applied in this circumstance.[9, 10]

Post-translational modifications are a critical part of the proteome. Upon translation, a protein can be decorated with extra functional groups at specific residues, that allow for the protein to perform varying activities during the cell cycle. The study of post translational modifications can be difficult using recombinant protein technology. Generally, in order to reach the yields required for adequate study, proteins are produced in *E. coli* and yeast through heterologous over expression. The protein of interest is often not a native protein to the host cell and so natural post translational modifications are difficult to install through such methods. Genetic code expansion may be used to incorporate an amino acid that already holds the modification of interest allowing *in vitro* study of such functionalities.[11] Acetylation (**Figure 70e**)[12] and propionylation (**Figure 70g**)[13] of lysine are prominent examples. These two amino acids may be incorporated site-specifically at a position of interest often previously determined from proteomic studies of the protein's native environment. More recently, analogues of post translational modifications have also been successfully incorporated such as thioacetal lysine (**Figure 70f**).[14]

The ability to control proteins and enzymes through activation of an active residue is a powerful tool. Again, genetic code expansion can play a key role in this avenue of research. Installation of unnatural amino acids with protecting groups that can be activated by various external stimuli can allow for the spatial and temporal control of proteins. Lysine, cysteine and tyrosine are often suitable choices. Deprotection can occur through small molecules,[15-17] enzymes[18] and even light activation.[19-23] Amino acids fundamental to the protein's mechanism can be introduced with a photo labile protecting group intact such as photocaged cysteine (**Figure 70h**). The amino acid is then "de-caged" by light, returning the protein to its natural function.

Proteins are synthesised in the cell through protein transcription followed by translation. Transcription first occurs, where mRNA is synthesised by RNA polymerase producing the respective mRNA sequence. Every three nucleotides

known collectively as codons, encode for one amino acid. The mRNA then moves to the ribosome where translation ensues. Here, tRNA that has been charged with amino acids through a reaction catalysed by amino-acyl tRNA synthetases (**Figure 71a**), match up with the codon in the mRNA. Each tRNA chain has an anticodon allowing for the specific recognition with its codon, allowing for the correct sequence of amino acids to be produced according to the original DNA template. As the tRNA enters the ribosome a peptide bond is formed between the current amino acid and the growing polypeptide chain (**Figure 71b**). Translation is performed from the N-terminus to the C-terminus. Completion of translation of the protein is denoted by a stop codon designated in the genetic code as either TAG, TAA or TGA. The termination of the protein is then induced by the ribosomal protein release factor 1 releasing the fully formed protein from the ribosome (**Figure 71c**).[24]

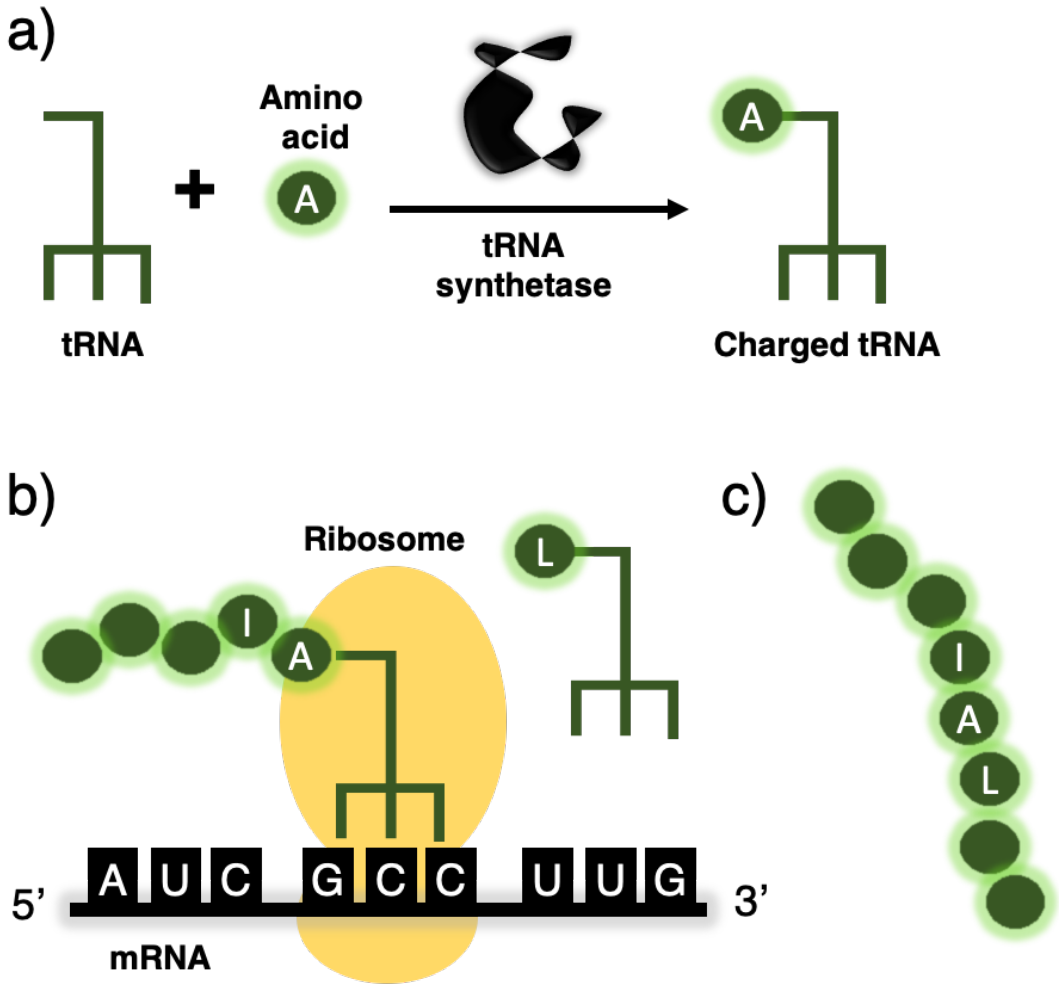


Figure 71. a) Charging of tRNA with an amino acid by tRNA synthetase. **b)** Ribosomal translation of mRNA into protein from the N-terminus to C-terminus. **c)** The fully formed polypeptide released from the ribosome following termination.

Amino-acyl tRNA synthetases are fundamental enzymes in protein synthesis. These are the catalysts that attach amino acids to their cognate tRNA allowing recognition in the ribosome. The mechanism of these enzymes involves the tRNA, the amino acid and adenosine triphosphate (ATP). Firstly, an adenyl-acyl intermediate is formed between adenosine monophosphate and the amino acid releasing diphosphate (**Figure 72a**). Then the 3' end of the tRNA can attack to form an ester bond and release adenosine monophosphate. This results in the tRNA covalently linked to the carboxyl group of the amino acid through an ester bond (**Figure 72b**).

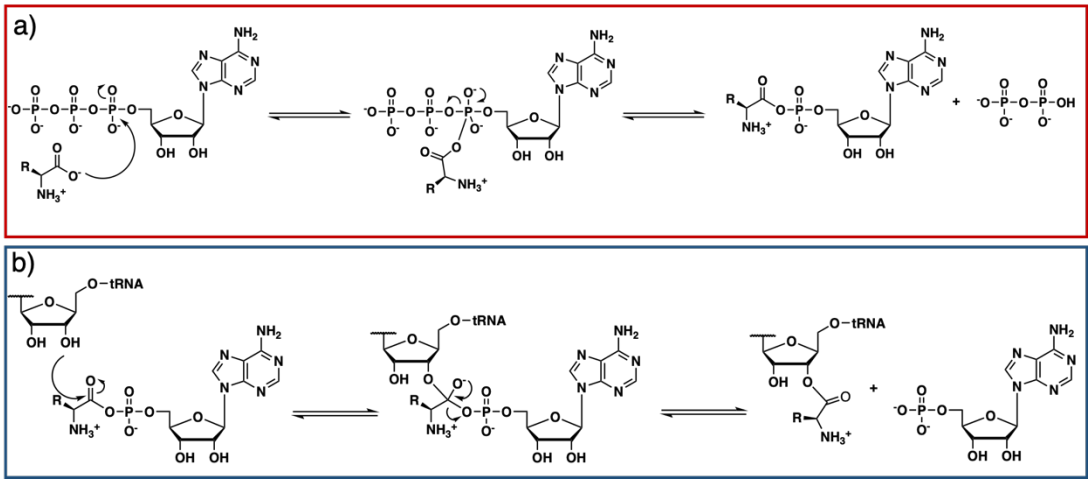


Figure 72. The general mechanism of type II amino-acyl tRNA synthetases. **a)** Nucleophilic attack of the deprotonated carboxyl group to the α-phosphate forms a pentavalent intermediate, this collapses to form an adenylate ester and diphosphate. **b)** The amino acid is then charged onto the tRNA by nucleophilic substitution of the activated ester.

To successfully incorporate an unnatural amino acid through genetic code expansion two things must be considered. Firstly, a tRNA synthetase must be able to recognise an unnatural amino acid but not recognise any endogenous amino acids in the cell (**Figure 73a**). Secondly the synthetase must recognise a cognate tRNA that has an anti-codon that matches a non-coding codon (**Figure 73b**). In other words, both the tRNA synthetase and its cognate tRNA must be orthogonal to the native protein synthesis machinery of the cell.[25] If these conditions are met then an unnatural amino acid may be incorporated into a protein in a live cell (**Figure 73c**). There should be no mis-incorporation of an endogenous amino acid at the position of interest and the non-canonical amino acid should not be placed in any position other position in the remining protein chain.

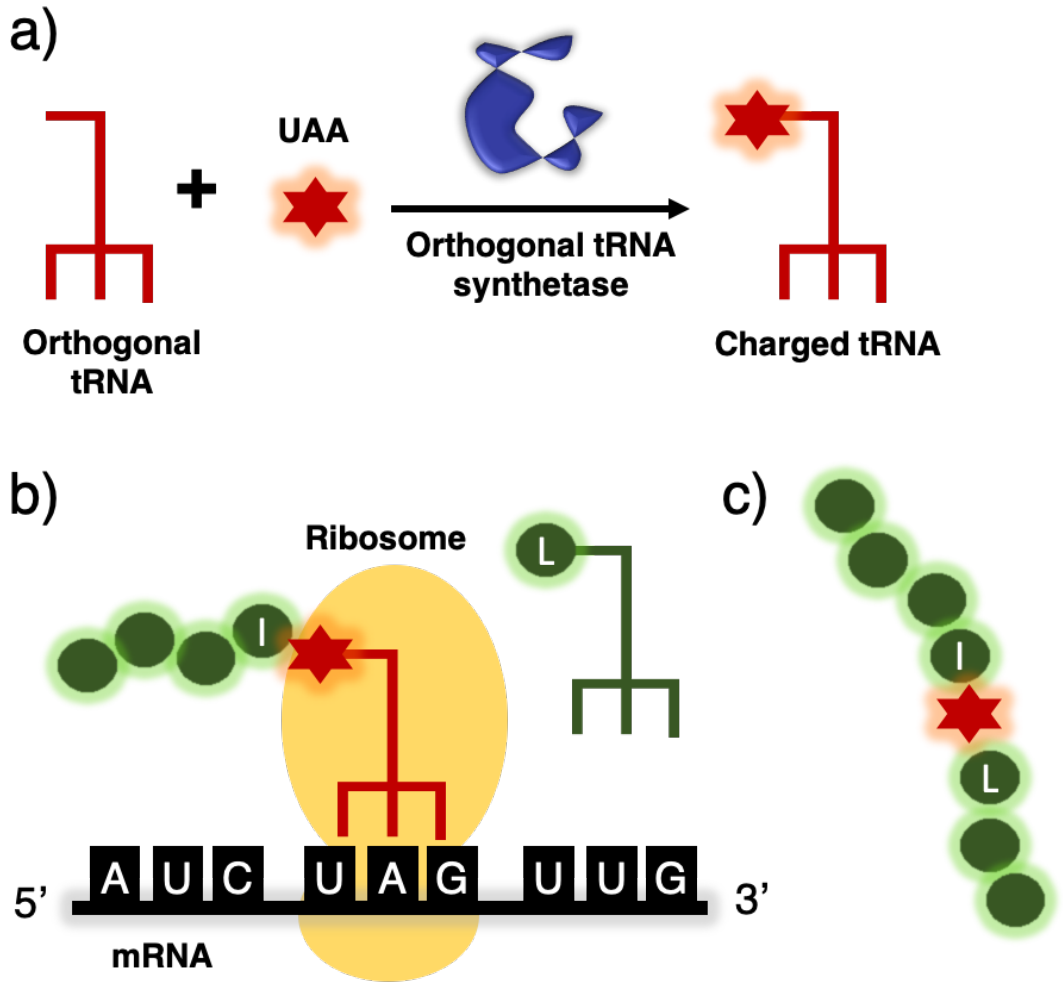


Figure 73. a) Charging of an orthogonal tRNA with a non-canonical unnatural amino acid by an orthogonal tRNA synthetase. b**)** Ribosomal translation of the unnatural amino acid into the protein backbone using a non-coding codon. **c)** The fully translated polypeptide with the unnatural amino acid site-specifically incorporated into to the protein.

Pyrrolysyl-tRNA synthetase (PylRS) is a remarkable enzyme of its class. It was discovered to encode the 22nd amino acid pyrrolysine (**Figure 74c**).[26] It has been discovered in methanogenic archaea and some bacteria such as *Sulfitobacter hafniense.* The enzyme is a type II tRNA synthetase.[27] The synthetase shows high promiscuity with regards to substrate scope and has been extensively engineered towards the incorporation of a vast number of structurally distinct amino acids.[28] In addition, the enzyme's cognate tRNA (PyltRNA) naturally decodes at the UAG codon in mRNA.[29] The UAG codon is analogous to the amber stop codon TAG. This practice known as amber codon suppression simplifies incorporation of an unnatural amino acid that can be recognised by PylRS as it simply requires a TAG mutation at the point of interest in the gene.

Incorporation of unnatural amino acids by PylRS has been extensively practiced. The C-terminal segment of the enzyme is the catalytic site, where the enzyme binds ATP and the amino acid and catalyses the reaction to charge the tRNA. The wild-type enzyme naturally recognises pyrrolysine along with many structurally distinct amino acids. These include ε-*N*-Boc lysine and cyclopropene lysine. To expand the substrate scope, the enzyme has been extensively subjected to directed evolution. An assay first developed for MjTyr tRNA synthetase for a similar application can be used to select for the incorporation of the unnatural amino acids following genetic variation. This has led to a vast number of unnatural amino acids being incorporated with highly varied structures.[30]

However, as with all genetic code expansion technologies, the challenge remains in the incorporation efficiency of the amino acid. Firstly, when utilising a TAG codon for incorporation there is always a competing mechanism to terminate the translation by release factor 1. This can be circumvented by the removal of all the RF1 factors in *E. coli* that recognise TAG and the mutation of all TAG codons to the remaining stop codons TAA or TGA. When this was applied, it was shown that although the incorporation efficiency of multiple amino acids can be improved in the same chain, the incorporation efficiency itself of the amino acid, at a singular position, is not improved.[31, 32] Secondly the enzyme promiscuity with regards to the amino acid is to be considered. Often, the enzymes require modification and lab-based evolution to achieve a substantial level of incorporation.[1]

The two most employed pyrrolysyl-tRNA synthetases are from the methanogenic archaea *Methanosarcina bakerri* (*Mb*) and *Methanosarcina mazei* (*Mm*). These two enzymes have high sequence homology (74%). The enzymes have two sections (**Figure 74b**). The C-terminal catalytic site[33] and the N-terminal segment which is thought to be instrumental in the recognition and binding of the PyltRNA.[34] Transplantation of the C-terminal mutations to increase substrate scope between the two homologs has shown to be highly successful. Often these mutations are interchangeable and have the desired effect of altering the substrate scope. However, transplantation of the N-terminal mutations has yet to be reported. In this instance it could be possible to increase the incorporation efficiency of a particular amino acid. It has been shown that in the *Mm* homolog mutations at the N-terminus are able to improve the incorporation efficiency. When residue Arg19 was mutated to His, His29 was mutated to Arg and Thr122 was mutated to Ser in MmPylRS it was shown that incorporation efficiency of Boc-lysine utilising the wild-type enzyme, was four times

greater than the enzyme with the native sequence.[34] In addition, when these mutations were performed in a variant evolved to incorporate acetyl lysine, an increase in incorporation efficiency was also observed. When comparing the primary sequence of both MmPylRS and MbPylRS Arg19 and His29 are both conserved. Thr122 is present as a native serine in MbPylRS (**Figure 74a**).

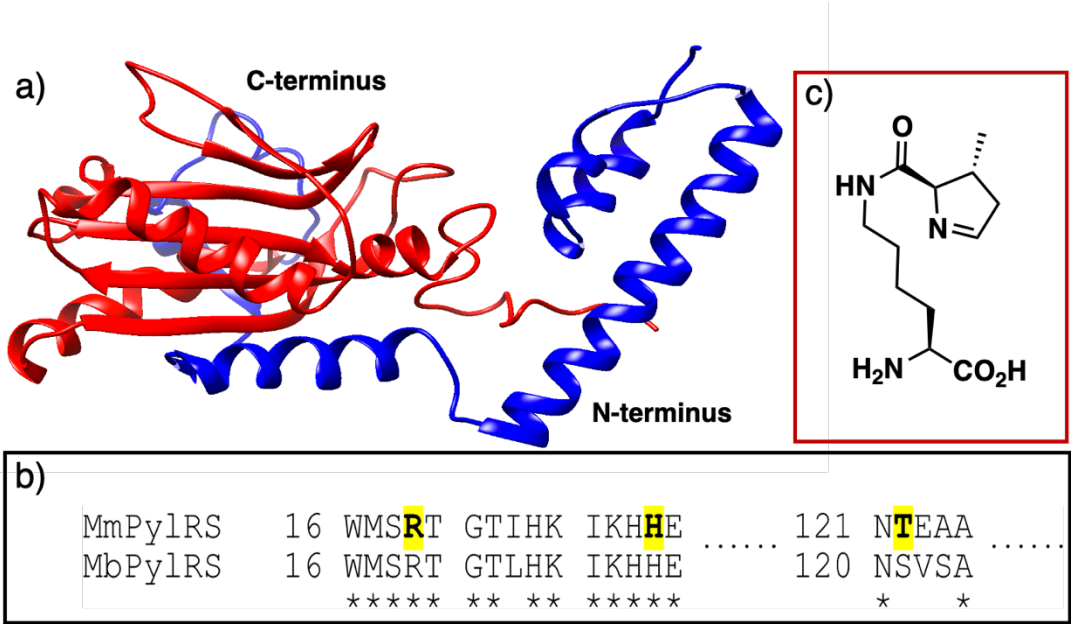


Figure 74. a) Crystal structure of MmPylRS (**PDB: 6ABM**), C-terminal domain in red and N-terminal domain in blue. **b)** Sequence alignment of MmPylRS and MbPylRS N-terminus. **c)** The structure of the 22nd proteinogenic amino acid pyrrolysine.

Here, the beneficial mutations found at the N-terminus of MmPylRS were evaluated in its homolog MbPylRS. In addition, C-terminal mutants were also investigated along with multiple amino acids and two culture temperatures.

4.3 Results and Discussion

To determine the incorporation efficiency of the unnatural amino acids, the reporter protein super folder green fluorescent protein (sfGFP) was used. Originally isolated from the jelly fish *Aequorea victoria*.[35] Upon translation this protein folds and performs a post-translational modification which produces a green fluorescent colour due to the maturation of a chromophore between three residues in the protein Ser65, Tyr66 and Gly67 (**Figure 75a**).[36] For the maturation of the chromophore to occur the full-length protein needs to be translated (**Figure 75b**). When a position in the nucleotide sequence is mutated to a TAG codon this will result in early termination of translation producing a truncated protein and hence the chromophore cannot mature (**Figure**

75b). When in the presence of a tRNA synthetase and a cognate tRNA that can recognise the TAG codon, along with an amino acid that is a substrate of said synthetase and tRNA, the cell machinery can produce a full-length protein which upon folding will mature a chromophore producing a fluorescent green colour (**Figure 75b**). Position Asn150 in GFP which does not take part in the chromophore maturation, was mutated to a TAG codon using PCR. Incorporation of an amino acid at this position would allow for the full expression of the protein resulting in the maturation of the chromophore (**Figure 75c**). The intensity of the fluorescence is dependent on the amount of full-length protein and as such correlates with the incorporation efficiency of the tRNA synthetase and its respective amino acid.

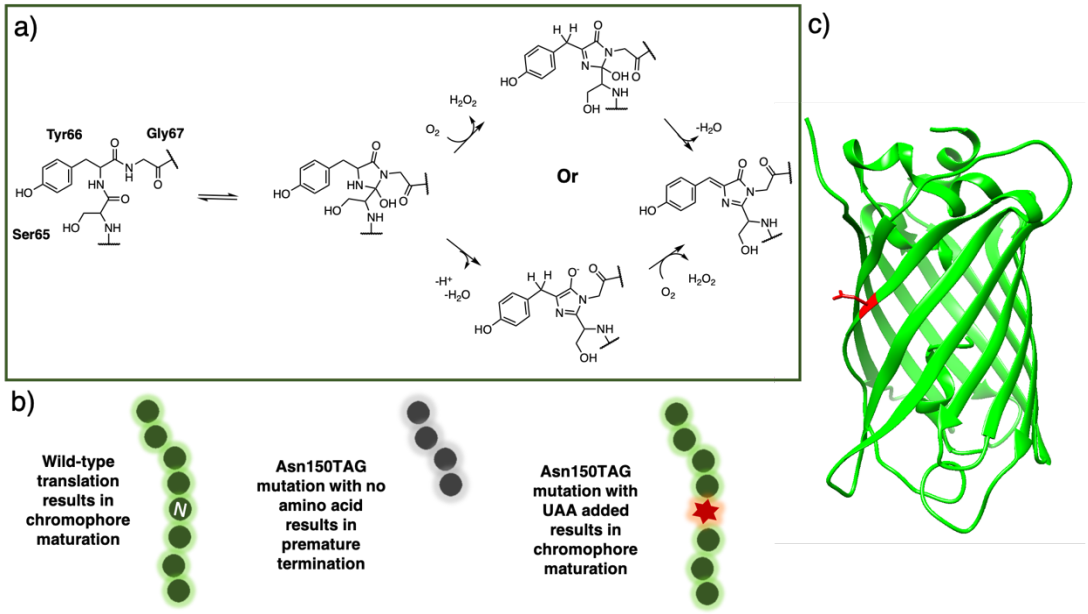


Figure 75. a) Post-translational chromophore maturation of green fluorescent protein proposed mechanism.[37-39] **b)** Translation of the native protein produces green fluorescence, a TAG mutation results in premature termination resulting in no fluorescence, addition of an unnatural amino acid and the corresponding machinery results in fluorescence corresponding to the incorporation efficiency. **c)** Crystal structure of sfGFP with the Asn150 highlighted in red (**PDB: 4LQT**)

To determine the incorporation efficiency of the synthetases BL21 (DE3) cells were co-transformed with the sfGFP mutant reporter plasmid which is under an inducible promoter and a plasmid encoding the PylRS synthetase and PyltRNA both under constitutive promoters. Cells containing the plasmids were then grown in fresh LB media until OD_{600} and induced with IPTG. Simultaneously the unnatural amino acid was added at a concentration of 5 mM and the cells cultured overnight (**Figure 76**). The following day fluorescence measurements were taken in triplicate on a fluorescent plate reader and the data collated and compared.

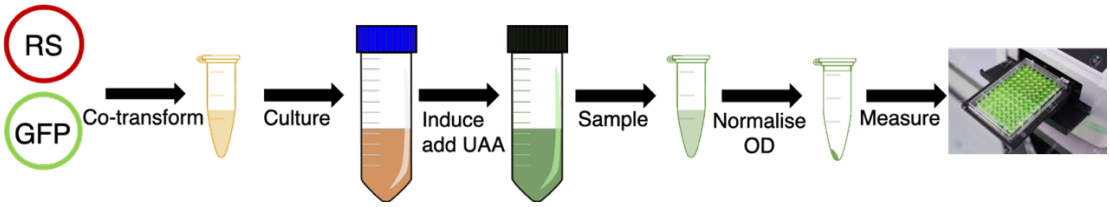
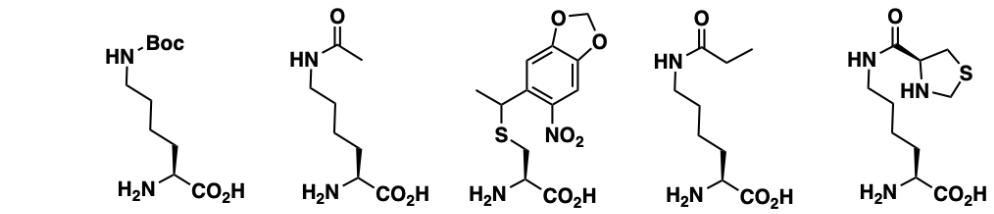
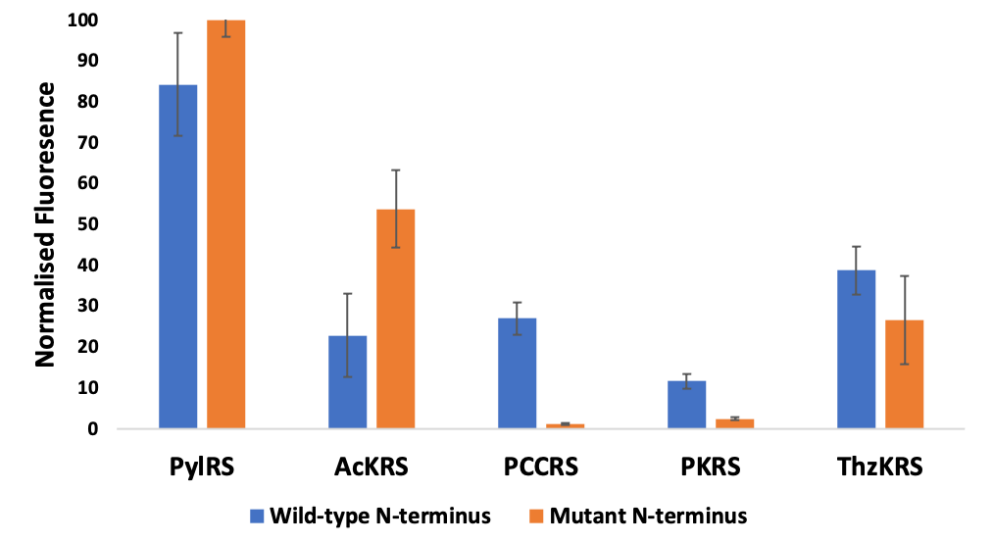


Figure 76. Outline of GFP reporter assay to determine incorporation efficiency of the tRNA synthetases. Co-transformation of the synthetase/tRNA plasmid with the reporter plasmid GFP. Culture to an OD_{600} of 0.6. Induce GFP expression with IPTG (0.5 mM) and add the unnatural amino acid (5 mM). Sample the culture. Normalise the OD_{600} to 1.0 and measure the relative fluorescence with a plate reader.

The His19Arg and Arg29His mutations of MbPylRS and its C-terminal variants were introduced by PCR. In addition to the wild-type enzyme, four reported variants were tested. The wild-type MbPylRS was tested using the substrate Boc-lysine. The variants employed and their respective amino acids (in brackets) were AcKRS (Acetyl lysine), PrKRS (Propionyl lysine), PCCRS (Photocaged cysteine) and ThzKRS (Thiazolidine lysine) (**Figure 77**).





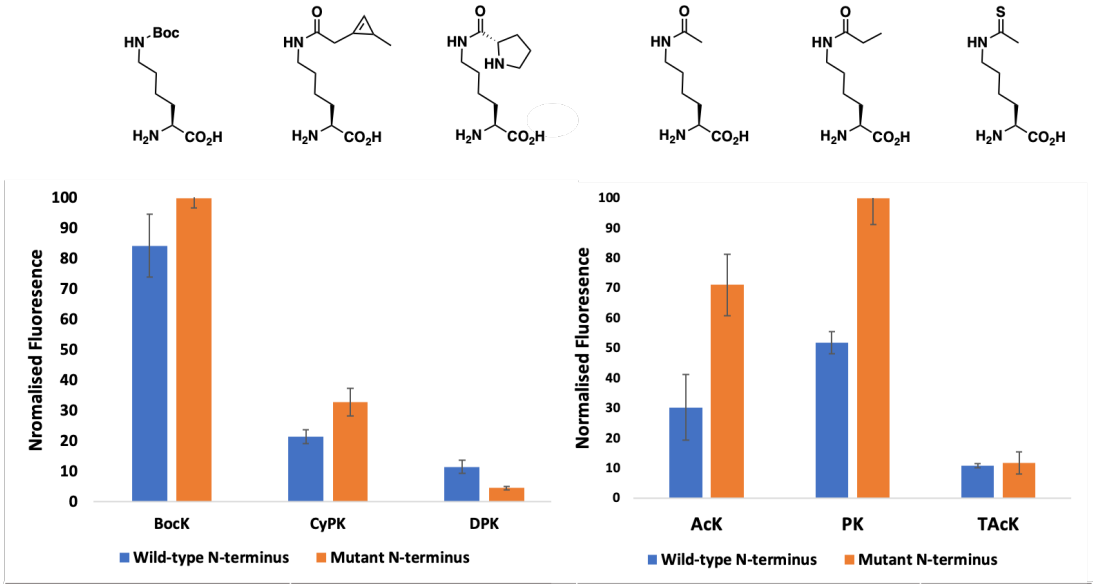
PylRS Variant	NT Wildtype		NT Mutant		P
	Mean	S.D.	Mean	S.D.	
Wild-type	26383	3936	31282	1244	0.09
AcKRS	7184	3160	16849	2955	0.01
PCCRS	8481	1235	411	57	0.00
PKRS	3676	583	810	125	0.01
ThzKRS	12153	1818	8341	3383	0.09

Figure 77. Comparison of the incorporation efficiency of the five tRNA synthetases with respect to their unnatural amino acids. Reported as normalised fluorescence relative to the most intense signal measured from the GFP plate reader assay. Mean of three biological replicates ± standard deviation (S.D.). Raw data table and statistical P values for each comparison determined from two tailed T test assuming unequal variance. Significance level >0.05. Cultures performed at 20 °C for 18 hours following reporter gene expression resulting from the addition of 0.5 mM of IPTG and 5 mM of the relevant unnatural amino acid.

The fluorescence assay suggests that the beneficial mutations found at the N-terminus of MmPylRS were able to increase the incorporation efficiency of AcKRS. The wild-type enzyme does not reach the thresh hold value to be significantly different. The three variants PCCRS, PKRS and ThzKRS demonstrated poorer incorporation efficiency. This is interesting in regard to the role played by both the C-terminus and N-terminus of the enzyme.

The positive effect of the N-terminal mutations in PylRS and AcKRS were investigated further. Both enzymes are able to use multiple amino acids as substrates for

incorporation efficiency. In addition to Boc-lysine (BocK), wild-type MbPylRS is able to incorporate cyclopropene lysine (CyPK) and D-prolyl lysine (DPK). AcKRS is able to incorporate propionyl lysine (PrK) and thioacetyl lysine (TAcK). The same reporter assay was used to determine the substrate effect of the N-terminal mutations (**Figure 78**).



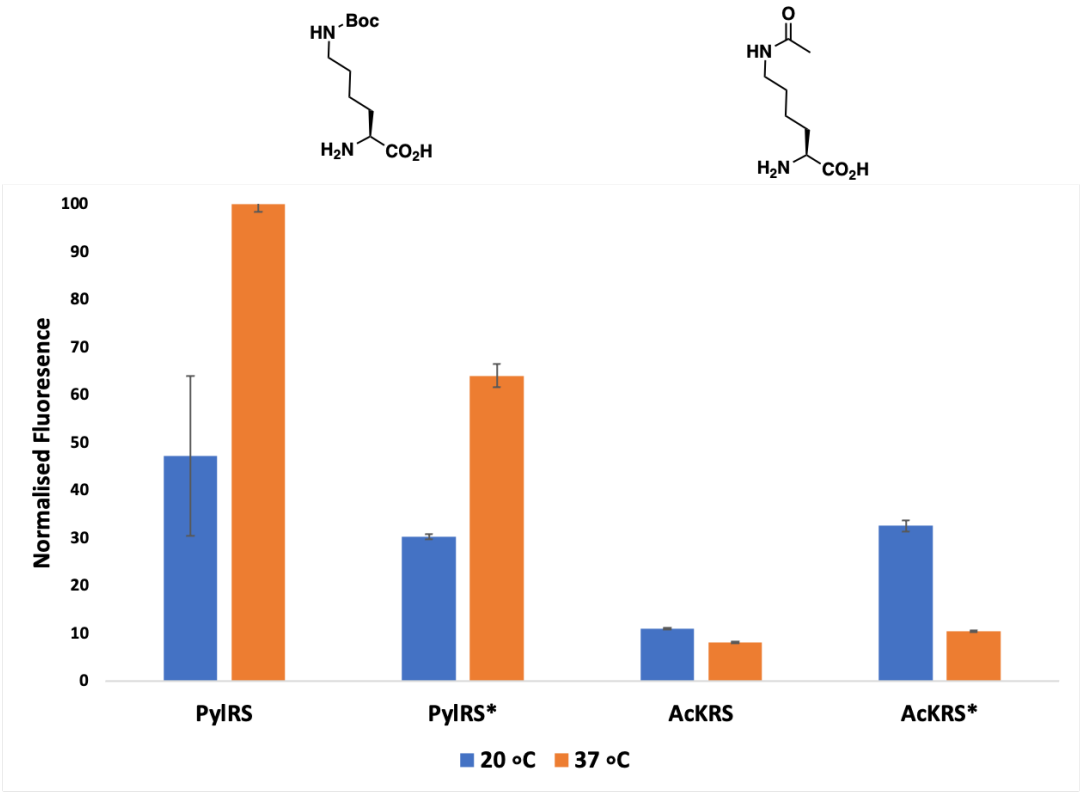
PylRS	NT Wildtype		NT Mutant		P
UAA	Mean	S.D.	Mean	S.D.	
BocK	26383	3936	31282	1244	0.09
CypK	6695	704	10241	1436	0.03
DPK	3593	686	1398	160	0.02
AcKRS	**NT Wildtype**		**NT Mutant**		**P**
UAA	Mean	S.D.	Mean	S.D.	
AcK	7184	3160	16849	2955	0.01
PK	12283	822	23677	2050	0.00
TAcK	2568	150	2809	865	0.37

Figure 78. Comparison of the incorporation efficiency of MbPylRS (left) with two new substrates CypK and DPK and AcKRS (right) with the substrates PK and TAcK. Reported as normalised fluorescence relative to the most intense signal measured from the GFP plate reader assay. Mean of three biological replicates ± standard deviation (S.D.). Raw data table and statistical P values for each comparison determined from two tailed T test assuming unequal variance. Significance level >0.05. Cultures performed at 20 °C for 18 hours following reporter gene expression resulting from the addition of 0.5 mM of IPTG and 5 mM of the relevant unnatural amino acid.

Interestingly the effect of the mutations carried over in to the CyPK substrate but not in the DPK substrate. Similarly, the mutations in AcKRS improved the incorporation of PK but not TAcK. This suggests that in addition to the C-terminal mutations is not the only factor underlying the effect of the mutations in the N-terminus but also the unnatural amino acids acting as the substrates in the reaction.

The temperature at which proteins express under recombinant conditions in *E. coli* can have a substantial effect on the yield. The incorporation of unnatural amino acids through GCE requires an exogenous enzyme reaction to occur in the cell. This is

delivered by the foreign synthetase, tRNA and unnatural amino acid. The incorporation efficiency dependant on the enzyme was assessed with regards to the expression temperature (**Figure 79**).



PylRS Variant	NT Wildtype		NT Mutant		P
	Mean	S.D.	Mean	S.D.	
Wild-type 20 °C	8207	2900	5260	95	0.11
Wild-type 37 °C	17369	260	11136	415	0.00
AcKRS 20 °C	1918	19	5661	202	0.00
AcKRS 37 °C	1419	18	1830	38	0.00

Figure 79. Comparison of the incorporation efficiency of MbPylRS with BocK and AcKRS with respect to expression temperatures 20 °C and 37 °C. Reported as normalised fluorescence relative to the most intense signal measured from the GFP plate reader assay. Mean of three biological repeats. Mean of three biological replicates ± standard deviation (S.D.). Raw data table and statistical P values for each comparison determined from two tailed T test assuming unequal variance. Significance level >0.05. Cultures performed at 20 °C or 37 °C for 18 hours following reporter gene expression resulting from the addition of 0.5 mM of IPTG and 5 mM of the relevant unnatural amino acid.

The wild-type synthetase MbPylRS showed an increased incorporation efficiency when cultured at 37 °C in both the native and mutated N-terminus. AcKRS on the other hand showed an improved incorporation efficiency when cultured at 20 °C for both N-terminal variants. This suggests that temperature has no effect on the incorporation efficiency with regards to the N-terminal mutations.

4.4 Conclusion

Mutations Arg19His/His29Arg in the N-terminus of MmPylRS has proven to significantly improve the incorporation efficiency of Boc-lysine and acetyl lysine. However, the data presented here strongly suggests that these mutations are not beneficial in its homolog MbPylRS. This is in contrast to transplanting mutations in the C-terminus of the enzymes. Here there is clearly an identical structure/activity relationship.

Introduction of the N-terminal mutations showed a very small increase in incorporation efficiency for MbPylRS and MmPylRS. On the other hand, the modification in ThzRS reduced the incorporation efficiency and in PKRS and PCCRS it abolished the incorporation completely. In addition, it appears that the amino acid used as the substrate does not show a clear pattern. This is puzzling as the amino acids have similar structures as they are all based on lysine. In addition, the culture temperature does not have an effect on incorporation efficiency, relative to the N-terminal mutations.

The structural difference between the amino acids and subsequent difference in C-terminal mutations do not allow rationalization of this effect. It is likely that the two closely related homologs, while often used interchangeably in genetic code expansion, are different in their underlying molecular mechanisms for functional regulation and catalytic activity. Indeed, there are 35 extra amino acid residues in the N-terminal domain of MmPylRS, and this difference accounts for the difference in the length of the two homologs (MmPylRS: 454 amino acids; MbPylRS: 419 amino acids). To improve unnatural amino acid incorporation efficiency, it would be necessary to perform detailed investigation of each homolog independently and caution should be taken when transferring mutations between MmPylRS and MbPylRS.

4.5 References

1. Chin, J. W., Expanding and reprogramming the genetic code. *Nature* **2017,** *550* (7674), 53-60.
2. Young, D. D.; Schultz, P. G., Playing with the Molecules of Life. *ACS Chem Biol* **2018,** *13* (4), 854-870.
3. Chin, J. W., Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **2014,** *83* (1), 379-408.
4. Lee, K. J.; Kang, D.; Park, H.-S., Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol Cells* **2019,** *42* (5), 386-396.
5. Lim, S. I.; Kwon, I., Bioconjugation of therapeutic proteins and enzymes using the expanded set of genetically encoded amino acids. *Crit. Rev. Biotechnol.* **2016,** *36* (5), 803-815.
6. Pickens, C. J.; Johnson, S. N.; Pressnall, M. M., et al., Practical Considerations, Challenges, and Limitations of Bioconjugation via Azide-Alkyne Cycloaddition. *Bioconjug Chem* **2018,** *29* (3), 686-701.
7. Tsai, Y. H.; Essig, S.; James, J. R., et al., Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat Chem* **2015,** *7* (7), 554-561.
8. Peng, T.; Hang, H. C., Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* **2016,** *138* (43), 14423-14433.
9. Nguyen, D. P.; Elliott, T.; Holt, M., et al., Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* **2011,** *133* (30), 11418-21.
10. Zheng, X.; Li, Z.; Gao, W., et al., Condensation of 2-((Alkylthio)(aryl)methylene)malononitrile with 1,2-Aminothiol as a Novel Bioorthogonal Reaction for Site-Specific Protein Modification and Peptide Cyclization. *J. Am. Chem. Soc.* **2020,** *142* (11), 5097-5103.
11. Xu, X.; Cao, X.; Yang, J., et al., Proteome-Wide Identification of Lysine Propionylation in the Conidial and Mycelial Stages of Trichophyton rubrum. *Front Microbiol* **2019,** *10* (2613), 2613.
12. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **2008,** *4* (4), 232-234.
13. Gattner, M. J.; Vrabel, M.; Carell, T., Synthesis of epsilon-N-propionyl-, epsilon-N-butyryl-, and epsilon-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun.* **2013,** *49* (4), 379-381.
14. Venkat, S.; Nannapaneni, D. T.; Gregory, C., et al., Genetically encoding thioacetyl-lysine as a non-deacetylatable analog of lysine acetylation in Escherichia coli. *FEBS Open Bio* **2017,** *7* (11), 1805-1814.
15. Li, J.; Yu, J.; Zhao, J., et al., Palladium-triggered deprotection chemistry for protein activation in living cells. *Nat Chem* **2014,** *6* (4), 352-361.
16. Fan, X.; Ge, Y.; Lin, F., et al., Optimized Tetrazine Derivatives for Rapid Bioorthogonal Decaging in Living Cells. *Angew. Chem. Int. Ed.* **2016,** *55* (45), 14046-14050.
17. Luo, J.; Liu, Q.; Morihiro, K., et al., Small-molecule control of protein function through Staudinger reduction. *Nat. Chem.* **2016,** *8* (11), 1027-1034.
18. Reille-Seroussi, M.; Mayer, S. V.; Dorner, W., et al., Expanding the genetic code with a lysine derivative bearing an enzymatically removable phenylacetyl group. *Chem. Commun.* **2019,** *55* (33), 4793-4796.
19. Ren, W.; Ji, A.; Ai, H. W., Light activation of protein splicing with a photocaged fast intein. *J. Am. Chem. Soc.* **2015,** *137* (6), 2155-2158.
20. Nguyen, D. P.; Mahesh, M.; Elsasser, S. J., et al., Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **2014,** *136* (6), 2240-2243.
21. Uprety, R.; Luo, J.; Liu, J., et al., Genetic encoding of caged cysteine and caged homocysteine in bacterial and mammalian cells. *ChemBioChem* **2014,** *15* (12), 1793-1799.
22. Gautier, A.; Deiters, A.; Chin, J. W., Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells. *Journal of the American Chemical Society* **2011,** *133* (7), 2124-2127.

23. Hemphill, J.; Chou, C.; Chin, J. W., et al., Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. *J. Am. Chem. Soc.* **2013,** *135* (36), 13433-13439.
24. Kuru, E.; Maattala, R. M.; Noguera, K., et al., Release Factor Inhibiting Antimicrobial Peptides Improve Nonstandard Amino Acid Incorporation in Wild-type Bacterial Cells. *ACS Chem Biol* **2020,** *15* (7), 1852-1861.
25. Wan, W.; Tharp, J. M.; Liu, W. R., Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **2014,** *1844* (6), 1059-1070.
26. Hao, B.; Gong, W.; Ferguson, T. K., et al., A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **2002,** *296* (5572), 1462-1466.
27. Srinivasan, G.; James, C. M.; Krzycki, J. A., Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **2002,** *296* (5572), 1459-1462.
28. Polycarpo, C. R.; Herring, S.; Berube, A., et al., Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **2006,** *580* (28-29), 6695-700.
29. Blight, S. K.; Larue, R. C.; Mahapatra, A., et al., Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **2004,** *431* (7006), 333-335.
30. Dumas, A.; Lercher, L.; Spicer, C. D., et al., Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **2015,** *6* (1), 50-69.
31. Johnson, D. B.; Xu, J.; Shen, Z., et al., RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat Chem Biol* **2011,** *7* (11), 779-786.
32. Schmied, W. H.; Elsasser, S. J.; Uttamapinant, C., et al., Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* **2014,** *136* (44), 15577-15583.
33. Owens, A. E.; Grasso, K. T.; Ziegler, C. A., et al., Two-Tier Screening Platform for Directed Evolution of Aminoacyl-tRNA Synthetases with Enhanced Stop Codon Suppression Efficiency. *ChemBioChem* **2017,** *18* (12), 1109-1116.
34. Sharma, V.; Zeng, Y.; Wang, W. W., et al., Evolving the N-Terminal Domain of Pyrrolysyl-tRNA Synthetase for Improved Incorporation of Noncanonical Amino Acids. *ChemBioChem* **2018,** *19* (1), 26-30.
35. Cubitt, A. B.; Heim, R.; Adams, S. R., et al., Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci* **1995,** *20* (11), 448-455.
36. Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J., et al., Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (21), 12111-12116.
37. Craggs, T. D., Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **2009,** *38* (10), 2865-2875.
38. Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V., Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *J. Am. Chem. Soc.* **2017,** *139* (30), 10239-10249.
39. Bartkiewicz, M.; Kazazić, S.; Krasowska, J., et al., Non–fluorescent mutant of green fluorescent protein sheds light on the mechanism of chromophore formation. *FEBS Lett.* **2018,** *592* (9), 1516-1523.

5. General Conclusions, Discussion and Future Work

5.1 General Conclusion and discussion

The strive to generate new biocatalysts is a promising avenue of research to increase the use of sustainable chemical processes. Exploiting Nature's catalysts to supplant and circumvent environmentally damaging and wasteful chemical transformations has become a main stay of academic and industrial endeavours. As our understanding of enzyme catalysis has become more prevalent the modification, repurposing and design of enzymes has greatly increased the range of reactions and number of substrates that can be employed by these catalysts.

With the advent of state-of-the-art chemical and synthetic biology technologies such as structure determination, synthetic gene synthesis, computational modelling and directed evolution, the manipulation of proteins has become increasingly faster, easier and more precise. Analytical techniques that allow in depth analysis of protein manipulation and high throughput screening of desired protein properties has allowed for the design of new proteins and enzymes that can meet the needs of modern-day chemistry and synthetic biology.

This thesis has striven to increase the strategies that are used to make artificial enzymes. In this case the use of iminium ion catalysed reactions through secondary amine organocatalysts has been demonstrated. Two design methods were explored that impart wholly new reactivity to the proteins of interest.

The first strategy was the modification of penicillin binding proteins with carbapenems that bear secondary amines in their structure. During this investigation it was shown that these carbapenems can perform iminium ion catalysed reactions in water to afford Michael-nitro additions to carbonyl compounds. Anchoring of two of these carbapenems to the protein β-lactamase from *M. tuberculosis* resulted in some selectivity likely derived from the protein backbone. However, this strategy proved poor as a method of enzyme design. Although there are many penicillin-binding proteins and further protein backbones may be explored the labelling first needs to be optimised, followed by the catalysis. This is not a trivial task and can be difficult to generate artificial enzymes. Nevertheless, it is a novel strategy. It may be prudent in the future to utilise the natural mechanism of other enzymes to covalently modify a protein backbone with a catalytic compound.

The introduction of unnatural amino acids was used as the second strategy to afford an artificial enzyme. Here it was shown that the use of secondary amine unnatural amino acids can be used to introduce new reactivities based on iminium ion catalysis. This thesis highlighted the benefit of using genetic code expansion over other methods of enzyme design. The synergistic use of a protein's natural traits such as internal hydrophobicity and natural binding sites and the novel catalytic potential of unnatural amino acids can allow the transfer of a reaction from one protein to another. This greatly expands the protein space that may be exploited for non-natural transformations. The artificial enzymes were able to facilitate selective, biocompatible iminium ion-based reduction of α,β-unsaturated aldehydes.

In addition, the mechanism of the enzymes produced through this method allow for future designs based on this strategy. Mechanistic and kinetic isotope effect studies strongly suggest an iminium ion-based mechanism that can be broadly applicable beyond hydride transfer catalysis. The use of NADPH and the coupling of the artificial protein to a natural enzyme also demonstrates the usefulness of this strategy in creating biocompatible organocatalysts.

Finally, the use of genetic code expansion is not confined to biocatalysts but has far ranging applications. Amber codon suppression using pyrrolysyl-tRNA synthetases is by far the most common method used in genetic code expansion. The more we learn about theses enzymes, the more we will be able to generate more robust unnatural amino acid incorporation in the future. The final study in this thesis was concerned with the N-terminus of the pyrrolysyl-tRNA synthetase enzyme from *M. bakeri*. It was shown that the modification of two homologs from the different species can have very different effects on unnatural amino acid incorporation efficiency. Detailed investigation of the enzymes used to afford unnatural amino acid incorporation needs to be taken in the future to determine the true underlying mechanisms of these highly important enzymes.

To conclude, this project generated an artificial enzyme through covalent modification that could perform a Michael nitro addition reaction. The selective reduction α,β-unsaturated aldehydes was performed using a second artificial enzyme design that was generated through genetic code expansion. Both of these systems were able to afford these small organocatalytically inspired reactions, under biocompatible conditions. This thesis has contributed to knowledge in enzyme design and it is hoped that the strategy and findings here will aid in generating next generation biocatalysts.

5.2 Future work

The most promising findings of this thesis are concerned with the introduction of secondary amine unnatural amino acids to afford iminium ion catalysed hydride transfer. Following preliminary studies with LmrR, the incorporation of the secondary amine into EcDHFR allowed the exploitation of the protein's natural mechanism. The catalysis worked, but it can be improved. One way to do this is to move the unnatural amino acid to new positions in the protein. A second is to directly evolve the protein using high-throughput assays that can evaluate the catalytic performance. In addition, the enantioselectivity of both LmrR Phe93DPK and EcDHFR Ala7DPK may be explored to determine if they can perform chiral reactions.

The unnatural amino acids used here are not confined to solely performing hydride transfer catalysis. With the possibility of incorporating these amino acids into any protein backbone many other organocatalytic reactions may be performed. For example, the formation of iminium ions in enzymes such as tryptophan synthase of tyrosine phenol lyase may afford novel biocatalytic aromatic substitution reactions. In addition, it may be prudent to pursue enamine catalysed reactions. Here, an artificial fluorinase or bio-orthogonal aldol reactions could be realised.

Introduction of new unnatural amino acids could also be attempted. With robust strategies in the literature, to evolve tRNA synthetases to introduce novel amino acids, it could be possible to introduce amino acids that are catalytically competent beyond those already proven. Examples may be secondary amine unnatural amino acids where the prolyl groups are β-proline instead of classic proline. This may increase the nucleophilicity and generate faster reactions by forming iminium ions quicker.

Natural enzyme catalysis may also be exploited further. Coupling of the artificial enzymes with the natural proteins catalysis may be prudent to explore to generate cascade reactions to produce industrially relevant and useful intermediates.

6. Materials and Method

6.1 Molecular biology

6.1.1 LB Media

NaCl (10 g), tryptone (10 g) and yeast extract (5 g) were dissolved in 1 L of dH_2O and sterilised by autoclave for 20 minutes at 121 °C. After cooling an appropriate antibiotic (if required) was added.

6.1.2 LB Agar

NaCl (10 g), tryptone (10 g), yeast extract (5 g) and agar (15 g) were dissolved in 1 L of dH_2O and sterilised by autoclave for 20 minutes at 121 °C. After cooling an appropriate antibiotic was added (if required) and the solution was poured into sterile petri dishes and left to solidify at room temperature.

6.1.3 Chemically Competent *E. coli* Cells

E. coli cells used for gene expression were of the strain BL21 (DE3). *E. coli* cell lines used for plasmid cloning and propagation were either DH5a, Stbl3 or MDS42. To make *E. coli* cells chemically competent, an aliquot of the appropriate cell line was plated onto LB agar plates and incubated overnight at 37 °C. One colony was picked and used to inoculate 100 mL of LB media. The cells were grown to an OD_{600} of 0.4. The media was removed by centrifugation (4,000 rpm, 4 °C, 20 min) and the cells chilled on ice for 10 minutes. The cells were resuspended in 50 mL of ice cold $CaCl_2$ (100 mM) solution and incubated on ice for 30 minutes. The cells were pelleted by centrifugation (4000 rpm, 4 °C, 20 min) and the supernatant discarded. The cells were resuspended in 10 mL of ice cold $CaCl_2$ (100 mM) 15% glycerol (v/v) solution and incubated on ice for 30 minutes. Finally, the cell suspension was aliquoted (100 µL) into sterile micro centrifuge tubes and frozen immediately at -80 °C.

6.1.4 Bacterial Transformation

Transformation of bacterial cells for both gene expression and plasmid propagation were performed as follows. Chemically competent *E. coli* cells were thawed on ice for

15 minutes. 50 ng/μL of DNA was added to the cells under sterile conditions. The cells were incubated for 15 minutes on ice then placed in a water bath at 42 °C for 45 seconds. The cells were then placed on ice for 2 minutes and LB media (900 μL) was added under sterile conditions. The cells were recovered at 37 °C for 45 minutes with constant agitation. The cells were then plated onto LB agar plates containing the appropriate antibiotic under sterile conditions. The plates were dried at 37 °C for 30 minutes and then flipped upside down and left to incubate overnight at 37 °C. In the case of double transformation for co-expression see **section 6.3**.

6.1.5 Plasmid construction

Plasmid construction was performed using the Gibson assembly method unless otherwise stated. NEBuilder DNA assembly master mix (2x) was mixed with the appropriate linearized vector and insert and incubated at 50 °C for 1 hour and 2 μL of the resultant mixture was transformed into chemically competent *E. coli* cells. Ratio of vector to insert were determined using the NEB Gibson Assembly online calculator.

6.1.6 Polymerase chain reaction (PCR)

PCR was performed using TaKaRa PrimeStar Max in 50 μL reaction volumes. The template DNA and primers were mixed with 25 μL PrimeStar Max master mix (2x) on ice and the volume made up to 50 μL with dH_2O and then placed in a thermocycler. The reaction was performed using the parameters in **Table 7** for 33 cycles.

Table 7. Polymerase chain reaction conditions.

Stage	Temperature (°C)	Time (s)
Denaturation	98	10
Annealing	55 – 65 (Dependant on primer T_m)	5
Extension	72	10/kbp

To the PCR product mixture was added fast digest buffer 10x (6 μL) followed by the restriction enzyme DpnI (Thermofisher Scientific) (1 μL). The mixture was incubated at 37 °C for 15 minutes. The products were purified by agarose gel electrophoresis.

6.1.7 Agarose Gel Electrophoresis

1% agarose gels were made by dissolving agarose (0.5 g) in TAE buffer (50 mL) and heated by microwave until all the solids were dissolved. To the hot solution was added 10,000x SyBr safe DNA stain (5 µL) and the solution was poured into a gel cast to set. Gel electrophoresis was performed at 120 V for 45 minutes. Gels were visualised on a Bio-rad gel imager. The DNA marker used was Thermo Gene ruler 1 kb plus (Thermo Scientific).

6.1.8 Sodium Dodecyl Sulphate Polyacryalmide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE gels were manually cast using the bio-rad mini protean kits. The recipes used are outlined in **Table 8**.

Table 8. SDS-PAGE recipe buffers.

Gel Layer	Resolving (12%)	Resolving (15%)	Stacking
dH_2O (mL)	3.4	2.4	2.9
Buffer (mL)	2.5	2.5	1.25
Bis-Acrylamide (mL)	4.0	5.0	0.83
10% APS (µL)	100	100	50
TEMED (µL)	10	10	5

Gels were stained using Coomassie brilliant blue and de-stained in dH_2O. Gels were visualised on a Bio-rad gel imager. The protein marker used was unstained protein marker (Thermo scientific).

6.1.9 Plasmid extraction and purification

Plasmid purification was performed using the Qiagen miniprep kit. In brief, following transformation, 1 colony was picked and used to inoculate 10 mL of LB. The culture was incubated overnight at 37 °C. The following day the culture was centrifuged (4,000 rpm, 4 °C, 20 min) and the plasmid DNA was purified following the manufacturers protocol. DNA concentration was determined using UV absorbance at 230 nm on a nanodrop instrument. Plasmids were stored at -20 °C for short term storage, and at -80 °C for long term storage.

6.1.10 Agarose Gel Electrophoresis PCR Product Purification

Following gel electrophoresis, the required DNA fragment was excised using a sharp sterile scalpel. To the gel pieces was added GQ buffer (3 x volume/weight of gel pieces) and the suspension placed in a thermoshaker (37 °C, 500 rpm) until the agarose gel pieces had fully dissolved. To the solution was added 2-propanol (1 x volume/weight of gel pieces) and incubated at room temperature for 5 minutes. The solution was then loaded onto a solid phase DNA extraction column. The column was centrifuged (13,000 rpm, room temperature, 1 min) the flow through was discarded and the column washed first with PB buffer (750 µL), followed by PE buffer (750 µL), from the Qiagen miniprep gel extraction kit. Finally, the DNA was eluted with 30 µL of dH_2O and the concentration was determined at 230 nm on a nanodrop instrument.

6.2 General Buffer Solutions and Reagents

6.2.1 Tris-Acetate EDTA Buffer for Agarose Gel Electrophoresis (TEA)

To 900 mL of dH_2O was added Tris-base (242 g), Glacial acetic acid (57 mL) and Ethylenediaminetetraacetic acid (18 g). The solution was adjusted to pH 8.2 and made up to 1 L with dH_2O. The buffer was stored at room temperature. For use the solution was diluted 50 times to the working concentration with dH_2O.

6.2.2 SDS-PAGE Resolving Buffer

To 200 mL of dH_2O was added Tris-base (46.5 g) and SDS (1.0 g). The solution was adjusted to pH 8.8 with HCl and made up to 250 mL with dH_2O. The solution was stored at 4 °C.

6.2.3 SDS-PAGE Stacking Buffer

To 200 mL of dH_2O was added Tris-base (15 g) and SDS (1.0 g). The solution was adjusted to pH 6.8 with HCl and made up to 250 mL with dH_2O. The solution was stored at 4 °C.

6.2.4 SDS-PAGE Running Buffer

To 1 L of dH_2O was added Tris-base (30.3 g), glycine (144.1 g) and SDS (10 g). The solution was diluted 10 x with dH_2O when used and stored at room temperature.

6.2.5 Coomassie Brilliant Blue SDS-PAGE Stain

To 1 L of dH_2O was added Coomassie Brilliant Blue – G (50 mg) followed by concentrated HCl (3.4 mL). The solution was stirred for 1 hour and stored at room temperature. For staining, SDS-PAGE gels were washed twice by submerging in dH_2O and microwaving for 1 minute. The gel was then submerged in the stain and microwaved for 1 minute and left to stain for 1 hour with constant agitation. To de-stain the gels, the dye was removed, and the gel washed with fresh dH_2O twice. The gel was then left in dH_2O for 2 hours with constant agitation.

6.2.6 SDS-PAGE Loading Buffer (2x) and Sample Preparation

SDS-PAGE Loading buffer (2x) was made by mixing SDS-PAGE stacking buffer (2.5 mL), glycerol (2.5 mL), 1% bromophenol blue (200 µL) and β-mercapto-ethanol (300 µL). The solution was made up to 10 mL with dH_2O and vortexed vigorously until homogenous. The solution was stored at room temperature.

For proteins to be extracted from cells (expression gels), 500 µL of the cells were taken and pelleted by centrifugation (13,000 rpm, 4 °C, 1 min) and the supernatant discarded. The cells were resuspended in 200 µL of dH_2O and to the suspension was added 200 µL of SDS-PAGE loading buffer. The samples were vortexed vigorously and heated for 5 minutes at 98 °C. The samples were centrifuged (13,000 rpm, 4 °C, 3 min) and 10 µL was loaded onto the SDS-PAGE gel.

For cell free samples, 20 µL of the sample was mixed with 20 µL of the SDS-PAGE loading buffer and heated for 3 minutes at 98 °C. The samples were centrifuged (13,000 rpm, 4 °C, 1 min) and 10 µL was loaded onto the SDS-PAGE gel.

6.2.7 Ammonium persulfate (APS)

To 200 µL of dH_2O was added APS (200 µg). The solution was stored at 4 °C.

6.2.8 GQ Buffer

To 40 mL of dH_2O was added Tris-base (0.12 g) and guanidine thiocyanate (32.5 g). The solution was adjusted to pH 6.5 with HCl and made up to 50 mL with dH_2O. The solution was stored at room temperature.

6.2.9 Antibiotic Solutions

Table 9. Antibiotic concentrations.

Antibiotic	Stock Concentration (mg/mL)	Working Concentration (µg/mL)
Ampicillin	100	100
Kanamycin	50	37.5
Spectinomycin	50	50

Antibiotic stock concentrations were made up in dH_2O and filter sterilised through 0.22 µm filters under sterile conditions then stored at -20 °C.

6.2.10 Isopropyl β-D-1-thiogalactopyranoside (IPTG)

To 100 mL of dH2O was added IPTG (12 g) and the solution was filter sterilised through 0.22 µm filters under sterile conditions, aliquoted into volumes of 1 mL, and stored at –20 °C.

6.3 Cloning, Expression and Purification of Recombinant Proteins

6.3.1 *Mycobacterium tuberculosis* β-lactamase (BlaC)

The gene encoding for the wild-type *Mycobacterium tuberculosis* β-lactamase (BlaC) without the 40-amino acid leader sequence and 20 bases either side complimentary to the vector was purchased as a double-stranded fragment (GeneArt, Invitrogen). The gene was cloned between NdeI and BamHI of a digested pET28a vector by

Gibson assembly to yield the wild type gene with an N-terminal His-tag originating from the vector. The gene was confirmed by DNA sequencing.

The plasmid was transformed into chemically competent BL21 (DE3) cells and grown on LB agar plates supplemented with kanamycin (37.5 µg/ml). One colony from the plate was picked to inoculate a 10 mL LB starter culture containing kanamycin (37.5 µg/mL) and incubated overnight at 37 °C. The starter culture was diluted into 1 L of fresh LB media containing kanamycin (37.5 µg/mL). The cells were grown at 37 °C until they reached an OD_{600} of 0.8, and IPTG was added to reach a final concentration of 0.5 mM. The cells were then incubated overnight at 20 °C. The cultures were harvested by centrifugation (4,000 rpm, 4 °C, 30 min) and the dry pellet was stored at –20 °C.

The pellet was resuspended in sodium phosphate buffer (35 mL per litre pellet, 50 mM NaP_i, 100 mM NaCl, 10 mM imidazole, pH 8.0). For protein labelling either 5 mg of meropenem or doripenem were added to the re-suspended cells. The cells were sonicated for 5 minutes (5 s on, 10 s off). The solids were separated by centrifugation (20,000 rpm for 30 min, 4 °C). The supernatant was filtered through 22 µm filters and loaded onto a Ni-NTA column equilibrated with the above buffer. The column was washed twice with wash buffer (25 mL, 50 mM NaP_i, 100 mM NaCl, 25 mM imidazole, pH 8.0) and then the protein was eluted in phosphate buffer (10 mL, 50 mM NaP_i, 100 mM NaCl, 300 mM imidazole, pH 8.0). 1 mg of doripenem or meropenem was added to the elution fraction, if required and incubated for 30 minutes at room temperature. Fast protein liquid chromatography (FPLC) was performed on an ÄKTA purifier system (GE Healthcare) at room temperature using a ProteoSEC size exclusion column (Generon, SEC-3-70-100 mL, 26 mm ID, 60 cm length, 3–70 kDa HR resin). Protein elution was monitored by UV absorbance at 280 nm and the elution buffer was 50 mM sodium phosphate buffer, 100 mM NaCl, pH 7.0. Fractions containing BlaC-carbapenem complexes were combined and concentrated using a 10 kDa cut off centrifugal concentrator (Millipore). Expression and purification were confirmed by SDS-PAGE.

6.3.2 *Mycobacterium tuberculosis* β-lactamase mutant (BlaC Glu166Ala)

The mutant protein BlaC Glu166Ala was expressed and purified in the same way as the wild type protein. The mutation was introduced using PCR with primers 1 and 2. The mutation was confirmed by DNA sequencing.

6.3.3 *Mycobacterium tuberculosis* D,D-Transpeptidase (Rv3330)

The gene encoding Rv3330 was a kind gift from the Alber lab. The gene was received in a pET-28c vector with a C-terminus His-tag.

The plasmid was transformed into chemically competent BL21 (DE3) cells and grown on LB agar plates supplemented with kanamycin (37.5 μg/ml). One colony from the plate was picked to inoculate a 10 mL LB starter culture containing kanamycin (37.5 μg/mL) and incubated overnight at 37 °C. The starter culture was diluted into 1 L of fresh LB media containing kanamycin (37.5 μg/mL). The cells were grown at 37 °C until they reached an OD_{600} of 0.6, and IPTG was added to reach a final concentration of 0.5 mM. The cells were then incubated overnight at 20 °C. The cultures were harvested by centrifugation (4,000 rpm, 4 °C, 30 min) and the dry pellet was stored at –20 °C.

The pellet was resuspended in Tris buffer (35 mL per litre pellet, 25 mM Tris, 50 mM NaCl, pH 8.0) with either 5 mg meropenem or doripenem added to the re-suspended cells. The cells were sonicated for 5 minutes (5 s on, 10 s off). The solids were separated by centrifugation (20,000 rpm, 4 °C, 30 mins). The supernatant was filtered through 22 μm filters and loaded onto a Ni-NTA column equilibrated with the above buffer. The column was washed twice with wash buffer (25 mL, 25 mM Tris, 50 mM NaCl, 25 mM imidazole, pH 8.0) and then the protein was eluted in Tris buffer (10 mL, 50 mM Tris, 100 mM NaCl, 300 mM imidazole, pH 8.0). 1 mg of doripenem or meropenem was added to the elution fraction and incubated for 30 minutes at room temperature. Fast protein liquid chromatography (FPLC) was performed on an ÄKTA purifier (GE Healthcare) system at room temperature using a ProteoSEC size exclusion column (Generon, SEC-3-70-100 ml, 26 mm ID, 60 cm length, 3 – 70 kDa HR resin). Protein elution was monitored by UV absorbance at 280 nm and the elution buffer was 50 mM sodium phosphate, 100 mM NaCl, pH 7.0. Fractions containing

Rv3330-carbapenem complexes were combined and concentrated using a 10 kDa cut off centrifugal concentrator (Millipore). Expression and purification were confirmed by SDS-PAGE.

6.3.4 Superfolder Green Fluorescent Protein (sfGFP Asp150TAG)

The gene encoding sfGFP was already present in our lab. The gene was digested using restriction enzymes NcoI and XhoI and ligated into a linearised pET-28a vector yielding the protein with a C-terminal His-tag. The gene was confirmed by DNA sequencing. The mutation was introduced by PCR with primers 3 and 4. The mutation was confirmed by DNA sequencing. The mutant plasmid was transformed using the procedure in section **6.3.11**.

Following transformation, the cells were grown at 37 °C until they reached an OD_{600} of 0.6, and IPTG was added to reach a final concentration of 0.5 mM. The cells were then incubated overnight at 20 °C. The cultures were harvested by centrifugation (4,000 rpm, 4 °C, 30 min) and the dry pellet was stored at –20 °C.

The pellet was resuspended in sodium phosphate buffer (35 mL per litre pellet, 50 mM NaP_i, 150 mM NaCl, 10 mM imidazole, pH 8.0). The cells were sonicated for 5 minutes (5 s on, 10 s off). The solids were separated by centrifugation (20,000 rpm , 4 °C, 30 min). The supernatant was filtered through 22 µm filters and loaded onto a Ni-NTA column equilibrated with the above buffer. The column was washed twice with wash buffer (25 ml, 50 mM NaP_i, 150 mM NaCl, 20 mM imidazole, pH 8.0) and then the protein was eluted in phosphate buffer (10 ml, 50 mM NaP_i, 150 mM NaCl, 300 mM imidazole, pH 8.0). Fast protein liquid chromatography (FPLC) was performed on an ÄKTA purifier (GE Healthcare) system at room temperature using a GE Healthcare BSD75 10/300 SEC column. Protein elution was monitored by UV absorbance at 280 nm and the elution buffer was 50 mM sodium phosphate buffer 150 mM NaCl pH 7.0. Fractions containing sfGFP were combined and concentrated using a 10 kDa cut off centrifugal concentrator (Millipore). Expression and purification were confirmed by SDS-PAGE.

6.3.5 *Lactococcus lactis* Multidrug Resistant Protein (LmrR)

The gene encoding for the wild-type LmrR was purchased as a double-stranded fragment (GeneArt, Invitrogen) with a C-terminal His-tag and two mutations Lys55Asn and Lys59Asp. The gene was codon optimised for expression in *E. coli*. The gene was cloned between NcoI and BamHI of a digested pET-28a vector by Gibson assembly to yield the wild-type gene. The plasmid was confirmed by DNA sequencing.

The plasmid was transformed into chemically competent BL21(DE3) cells and grown on LB agar plates supplemented with kanamycin (50 µg/mL). One colony from the plate was picked to inoculate a 10 mL LB starter culture containing kanamycin (37.5 µg/mL) and incubated overnight at 37 °C. The starter culture was diluted into 1 L of fresh LB media containing kanamycin (37.5 µg/mL). The cells were grown at 37 °C until they reached an OD_{600} of 0.8 – 1.0 and IPTG was added to reach a final concentration of 0.5 mM. The cells were then incubated overnight at 30 °C. The cultures were harvested by centrifugation (4,000 rpm, 4 °C, 30 min) and the dry pellet was stored at –20 °C.

The pellet was resuspended in PBS buffer (35 ml per litre pellet, 50 mM NaP_i, 150 mM NaCl, 10 mM imidazole, pH 8.0). DNAase (10 µg/mL) and $MgCl_2$ (100 µg/mL) were added along with PMSF (10 µg/mL). The cells were sonicated for 7 minutes (5 s on, 10 s off). The solids were separated by centrifugation (20,000 rpm, 4 °C, 30 min). The supernatant was loaded on to a Ni-NTA column equilibrated with the above buffer. The column was washed twice with wash buffer (25 ml, 50 mM NaP_i, 150 mM NaCl, 20 mM imidazole, pH 8.0) and then the protein was eluted in phosphate buffer (10 mL, 50 mM NaP_i, 150 mM NaCl, 300 mM imidazole, pH 8.0). Fast protein liquid chromatography (FPLC) was performed on an ÄKTA purifier (GE Healthcare) system at room temperature using a GE Healthcare BSD75 10/300 SEC column. Protein elution was monitored by UV absorbance at 280 nm and the elution buffer was 50 mM sodium phosphate buffer 150 mM NaCl pH 7.0. Fractions containing LmrR were combined and concentrated using a 10 kDa cut off centrifugal concentrator (Millipore). Expression and purification were confirmed by SDS-PAGE.

6.3.6 *Lactococcus lactis* Multidrug resistant regulator Protein Mutants (LmrR TAG)

The mutations for each mutant were introduced individually by PCR using primers 5 and 6 (Val15TAG), 7 and 8 (Asp19TAG), 9 and 10 (Met89) and 11 and 12 (Phe93). Each mutant was confirmed by DNA sequencing.

The mutant genes were transformed using the procedure in **section 6.3.10** and expressed following the wild type procedure with the exception of the addition of spectinomycin in the overnight and expression media and the addition of 1 mM of the relevant amino acid added up on induction.

For the preliminary screens, each mutant was purified by Ni-affinity chromatography as the wild-type was and then dialysed overnight into the reaction buffer PBS (50 mM NaP_i, 150 mM NaCl, pH 7.0). Following the identification of the active mutant LmrR Phe93DPK, this mutant was further purified using the same procedure as the wild-type protein (**Figure 80**).

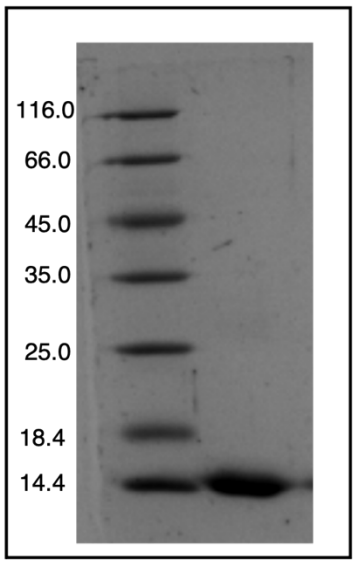


Figure 80. SDS-AGE analysis of LmrR Phe93DPK following size exclusion chromatography. Mass of LmrR Phe93DPK monomer approx. 14.5 kDa. Protein ladder in kDa.

6.3.7 *Escherichia coli* Dihydrofolate Reductase (EcDHFR)

The gene encoding wild type EcDHFR was already available in our lab. The gene was amplified by PCR using primers 13 and 14 and simultaneously a C-terminal His-

tag and 20 base pair overhangs complimentary to the vector were added. The gene was then cloned into a linearised pET-28a vector between restriction sites NcoI and BamHI using Gibson assembly to yield the wild-type plasmid. The plasmid was confirmed by DNA sequencing.

The plasmid was transformed into chemically competent BL21(DE3) cells and grown on LB agar plates supplemented with kanamycin (50 µg/mL). One colony from the plate was picked to inoculate a 10 mL LB starter culture containing kanamycin (37.5 µg/mL) and incubated overnight at 37 °C. The starter culture was diluted into 1 L of fresh LB media containing kanamycin (37.5 µg/mL). The cells were grown at 37 °C until they reached an OD_{600} of 0.6 – 0.8 and IPTG was added to reach a final concentration of 0.5 mM. The cells were then incubated overnight at 20 °C. The cultures were harvested by centrifugation (4,000 rpm, 4 °C, 30 min) and the dry pellet was stored at –20 °C.

The dry pellet was resuspended in sodium phosphate buffer (35 mL per litre pellet, 50 mM NaPi, 150 mM NaCl, 10 mM imidazole, pH 8.0). The cells were sonicated for 7 minutes (5 s on, 10 s off). The solids were separated by centrifugation (20,000 rpm, 4 °C, 30 min). The supernatant was loaded onto a Ni-NTA column equilibrated with the above buffer. The column was washed twice with wash buffer (25 mL, 50 mM NaP_i, 150 mM NaCl, 20 mM imidazole, pH 8.0) and then the protein was eluted in phosphate buffer (10 mL, 50 mM NaP_i, 150 mM NaCl, 300 mM imidazole, pH 8.0). Fast protein liquid chromatography (FPLC) was performed on an ÄKTA purifier (GE Healthcare) system at room temperature using a GE Healthcare BSD75 10/300 SEC column. Protein elution was monitored by UV absorbance at 280 nm and the elution buffer was 50 mM sodium phosphate buffer, 150 mM NaCl, pH 7.0. Fractions containing EcDHFR were combined and concentrated using a 10 kDa cut off centrifugal concentrator (Millipore). Expression and purification were confirmed by SDS-PAGE.

6.3.8 *Escherichia coli* Dihydrofolate Reductase Mutants (EcDHFR TAG)

The mutations were introduced individually using PCR with primers 15 and 16 (Ala7TAG), 17 and 18 (Phe31TAG) and 19 and 20 (Ser49TAG). Each mutant was confirmed by DNA sequencing.

The mutant genes were transformed using the procedure in **section 6.3.10** and expressed following the wild type procedure with the exception of the addition of spectinomycin in the overnight and expression media and the addition of 1 mM DPK added up on induction.

For the preliminary screens, each mutant was purified by Ni-affinity chromatography as the wild type was and then dialysed overnight into the reaction buffer PBS (50 mM NaP_i, 150 mM NaCl, pH 7.0). Following the identification of the active mutant EcDHFR Ala7DPK, this mutant was further purified using the same procedure as the wild type protein (**Figure 81**).

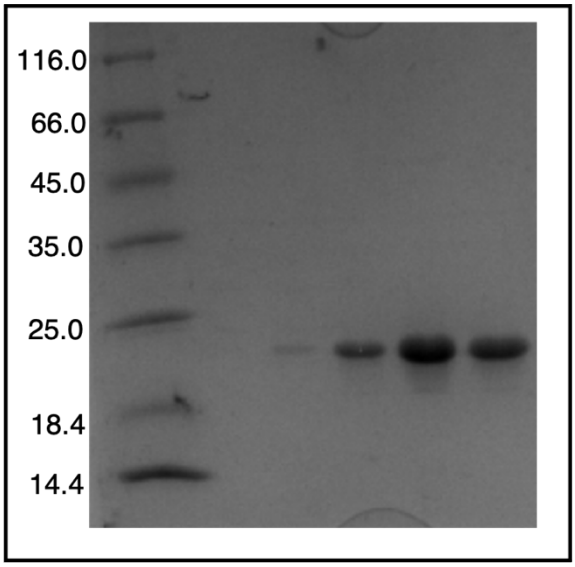


Figure 81. SDS-AGE analysis of LmrR Phe93DPK following size exclusion chromatography. Mass of LmrR Phe93DPK monomer approx. 19.5 kDa. Protein ladder in kDa.

6.3.9 *Thermoanaerobacter brockii* Alcohol Dehdrogenase (TbADH)

A pET-11c plasmid encoding the gene for TbADH under the T7 promoter was a kind gift from the Allemann lab. The plasmid was transformed into chemically competent BL21(DE3) cells and grown on LB agar plates supplemented with ampicillin (100 µg/mL). One colony from the plate was picked to inoculate a 10 mL LB starter culture containing ampicillin (100 µg/mL) and incubated overnight at 37 °C. The starter culture was diluted into 1 L of fresh LB media containing ampicillin (100 µg/mL). The cells were grown at 37 °C until they reached an OD_{600} of 0.6 – 0.8 and IPTG was added to reach a final concentration of 0.5 mM. The cells were then incubated

overnight at 30 °C. The cultures were harvested by centrifugation (4,000 rpm, 4 °C, 30 min) and the dry pellet was stored at –20 °C.

The dry pellet was resuspended in Tris buffer (30 mL per litre pellet, 20 mM, pH 8.0). The cells were sonicated for 7 minutes (5 s on, 10 s off) and the solids were separated by centrifugation (20,000 rpm, 30 min, 4 °C). The supernatant was then diluted with a further 30 mL of Tris buffer (20 mM, pH 8.0) and heated to 80 °C in a water bath for 30 minutes. Upon cooling to room temperature, the precipitate was removed by centrifugation (20,000 rpm, 30 min, 4 °C), and the supernatant filtered through 22 µm syringe filters. The protein was further purified by FPLC.

Fast protein liquid chromatography (FPLC) was performed on an ÄKTA purifier (GE Healthcare) system at room temperature using a GE Healthcare Q-sepharose anion-exchange column. Protein elution was monitored by UV absorbance at 280 nm and gradient elution was performed with a concentration from 0-1 M NaCl Tris buffer (10 mM, pH 8.0). Fractions containing TbADH were combined and concentrated using a 10 kDa cut off centrifugal concentrator (Millipore). Expression and purification were confirmed by SDS-PAGE (**Figure 82**). The TbADH was aliquoted into concentrations of 2 mg/mL and lyophilised overnight, the dry enzyme was stored at -80 °C until needed.

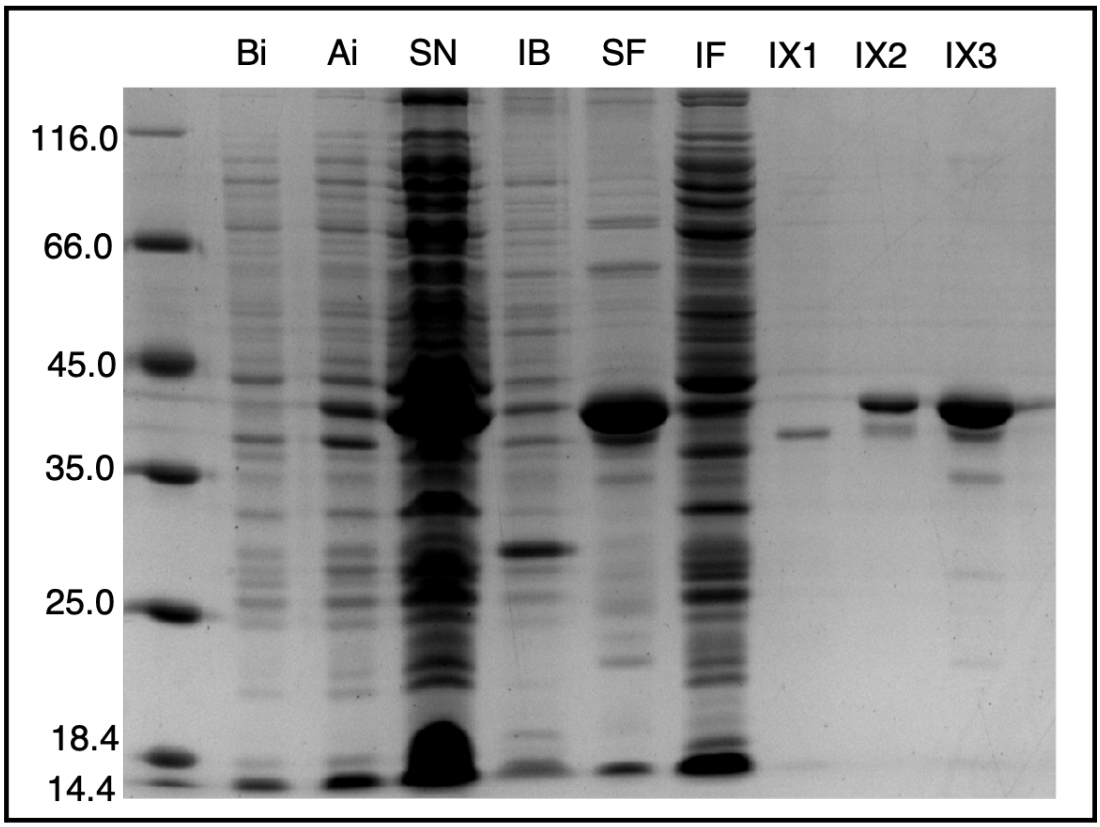


Figure 82. SDS-PAGE analysis for the expression and purification of TbADH. Bi = Before induction of gene expression by IPTG. Ai = After induction of gene expression by IPTG. SN = Supernatant following lysis by sonication. IB = Inclusion bodies following lysis by sonication. SF = Soluble fraction following thermal precipitation. IF = Insoluble fraction following thermal precipitation. IX1 = Ion exchange elution 1. IX2 = Ion exchange elution 2. IX3 = Ion exchange elution 3. Mass of TbADH approx. 37 kDa. Protein ladder in kDa.

6.3.10 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase and tRNA (MbPylRS)

The gene encoding the wild type MbPylRS and tRNA was provided by the Chin lab in a pCDF vector. MbPylRS was under the constitutive promoter glutamine tRNA synthetase promoter and the tRNA under the strong lpp promoter. The plasmid was co-transformed with the required gene into BL21 (DE3) cells along with the gene of interest and following recovery in LB media (37 °C, 1 hour) the cells were directly diluted into 10 mL of fresh LB media supplemented with kanamycin (50 µg /mL) and spectinomycin (50 µg/mL). The cultures were then expressed as per each protocol for the gene of interest.

6.3.11 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase C-terminal mutants (ThzKRS, AcKRS, PrKRS and PCCRS)

The mutant MbPylRS ThzKRS was generated by four point mutations introduced by two rounds of PCR. Firstly, using primers 21 and 22 (Asp267Ser), then using primers 23 and 24 (Cys313Val, Met315Phe and Asp344Gly). The mutant was confirmed by DNA sequencing.

The plasmids encoding the mutant synthetases AcKRS, PrKRS and PCCRS were already available in our lab. They consisted of the mutations Asp76Gly, Ser124Gly, Leu266Met, Leu270Ile, Tyr271Phe, Leu274Ala and Cys313Phe (AcKRS) Tyr271Phe and Cys313Thr (PrKRS), Asp311Glu, Cys313Ala and Val366Met (PCCRS).

The mutant synthetases were co-expressed with their respective genes of interest using the same procedure as the wild-type MbPylRS.

Table 10. C-terminal mutations aligned for each residue compared to the wild-type MbPylRS.

Residue No.	76	123	266	270	271	274	311	313	366
MbPylRS	D	S	L	L	Y	L	N	C	V
MbAcKRS	G	G	M	I	F	A	N	F	V
MbPrKRS	D	S	L	L	F	L	N	T	V
MbPCCRS	D	S	L	L	Y	L	Q	A	M

6.3.12 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase N-terminal mutants.

To introduce the N-terminal mutations of each *Mb* pyrrolysyl-tRNA synthetase gene, PCR was performed on the respective plasmids using primers 25 and 26. Each gene was confirmed by DNA sequencing. Expression of the MbPylRS N-terminal mutants was performed in the same way as the wild-type MbPylRS.

6.3.13 Primer Table

Table 11. Primer table.

Primer No.	F/R	Primer Sequence
1	F	CTGGATGCAGAAGCACCGGAACTGAATC
2	R	GATTCAGTTCCGGTGCTTCTGCATCCAG
3	F	AATATAATTTCAACAGCCATTAGGTGTATATTACCGCCGATAAACAG
4	R	ATGGCTGTTGAAATTATATTCCAGTTTATGACC
5	F	TGCGTGCTCAAACCAATTAGATCCTGCTGAATGTCCTGAAAC
6	R	TCGCCTTGTTTCAGGACATTCAGCAGG
7	F	AACCAATGTCATCCTGCTGTAGGTCCTGAAACAAGGCGATAAC
8	R	TGCCATACACATAGTTATCGCCTTGTTTCAGG
9	F	ATGAAAACTAGCGCCTGGCGTTCGAATCC
10	R	ACTCCAGGATTCGAACGCCAGG
11	F	TGCGCCTGGCGAAAGAATCCTGGAGTCGTGTG
12	R	TCAATGATTTTGTCCACACGACTCCAGG
13	F	GTTTAACTTTAAGAAGGAGATATACATATGATCAGTCTGATTGCGGC GTTAGC
14	R	CGGAGCTCGAATTCGTTAGCCGCTGCTGTGATGATGATGATGATGC TGCTGCCCCGCCGCTCCAGAATCTCAAAGC
15	F	TACATATGATCAGTCTGATTGCGTAGTTAGCGGTAG
16	R	CGCAATCAGACTGATCATATGTATATCTCC
17	F	TCGCCTGGTAGAAACGCAACACCTTAAATAAACC
18	R	CCATAATCACGGGTTTATTTAAGGTGTTGC
19	F	TTATGGGCCGCCATACCTGGGAATAGATCGGTCGTCC
20	R	AGGTATGGCGGCCCATAATCACGG
21	F	CCGACCCTGTATAACTATCTGCGTAAACTGGATCG
22	R	GTTATACAGGGTCGGGCTCAGCATCGGACGCAGG
23	F	AGGAATTCTTTGATCAGCGCTTCCAGGTTTTCACGGGTGCAGCCGCT GCCAAATTGCACAAAGTTAACCATGGTGAATTCTTCCAG
24	R	GCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAAAT TGTGGGCGGTAGCTGCATGGTGTATGG
25	F	TGGATGAGCCATACCGGCACCCTGCATAAAATCAAACATCGTGAAGT GAG
26	R	TGCCGGTATGGCTCATCCACAGGCCGGTCGCGCTAATCAGCACATC

6.3.14 Determination of protein concentration

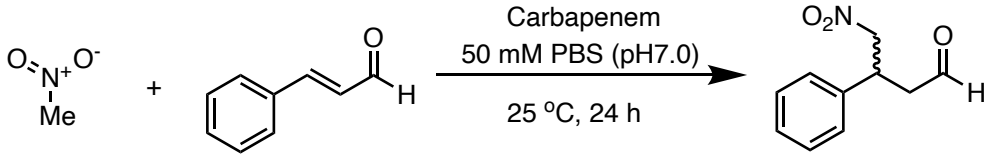
Protein concentration was determined using the Beer-Lambert law (**equation 1**). The molar extinction coefficient of each recombinant protein was determined using the Expasy Protoparam online software by inserting the proteins primary structure. Measurements were performed on nanodrop instrument at a wavelength of 280 nm.

Equation 1.

$$A = \varepsilon c \iota$$

6.4 Procedure for Small Molecule and Enzyme Catalysed Reactions and for the Determination of Enantioselectivity

6.4.1 Michael Nitro Addition Reaction (Carbapenems)



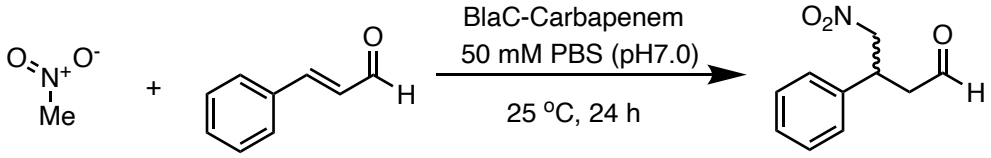
For the buffer screen, a concentration of 10 mM at pH 7.0 was used for each of the following: (4-(2-hydroxyethyl)-1- piperazineethanesulfonic acid (HEPES), potassium phosphate (KP_i), sodium phosphate (NaP_i), or phosphate-buffered saline (PBS, i.e. sodium phosphate and sodium chloride). For doripenem reactions, to a solution of doripenem (Glentham Life Sciences #GA9212) (1.00 mg, 2.60 µmol, 1 eq) in the indicated buffer (495 µL) was added cinnamaldehyde (1.57 mg, 11.9 µmol, 5 eq) in MeOH (5 µL) and nitromethane (1.27 µl, 23.8 µmol, 10 eq). For meropenem reactions, to a solution of meropenem trihydrate (Ark Pharm Inc, #AK161987) (1.14 mg, 2.38 µmol, 1 eq) in the indicated buffer (495 µl) was added cinnamaldehyde (1.72 mg, 13.0 µmol, 5 eq) in MeOH (5 µL) and nitromethane (1.39 µL, 26.0 µmol, 10 eq). For ertapenem reactions, to a solution of ertapenem sodium salt (Glentham Life Sciences, #GA8176) (1.04 mg, 2.10 µmol, 1 eq) in the indicated buffer (495 µl) was added cinnamaldehyde (1.39 mg, 10.5 µmol, 5 eq) in MeOH (5 µl) and nitromethane (1.12 µL, 21.0 µmol, 10 eq). The reactions were then placed in a thermoshaker at 25 °C and shaken at 800 rpm for 24 hours.

For reactions carried out at pH 7.5 and pH 8.0, the buffers were adjusted with 1 M NaOH, and the pH measured using a pH meter. For the co-solvent screen, the

reactions were performed by first adding 245 µL of the indicated buffer followed by 250 µL of the indicated solvent.

For reactions in buffer alone, the reaction mixtures were extracted with 700 µL of $CDCl_3$ and analysed by ^1H NMR spectroscopy. For those reactions with methanol or acetonitrile as the co-solvent, the solvents were evaporated under reduced pressure, and the crude mixture was extracted with 700 µL of $CDCl_3$. Reactions containing benzene or chloroform as the co-solvent were performed using the corresponding deuterated solvents and subjected directly to NMR analysis. Substrate to product conversion was estimated from integration of signals arising from cinnamaldehyde, the product and side product. ^1H NMR spectra were recorded in $CDCl_3$ or C_6D_6.

6.4.2 Michael Nitro Addition Reaction (BlaC-Carbapenem Artificial Enzymes)



To a microcentrifuge tube was added 250 µL of the enzyme solution (10 mg, 324 nmol, 0.2 eq) and to this was added 245 µL of reaction buffer (50 mM NaP_i, 100 mM NaCl, pH 7.0). To the enzyme was added 5 µl of a stock solution of cinnamaldehyde in methanol, (213 µg, 1.6 µmol, 1 eq) followed by 0.17 µL (195.2 µg, 3.2 µmol, 2 eq) of neat nitromethane. The reactions were shaken at 50 rpm, 25 °C for 24 h. 700 µL of $CDCl_3$ was added to the reactions when finished and the samples spun down in a micro centrifuge. The organic fraction was removed and subjected to ^1H NMR analysis. The control reactions carried out in tandem with meropenem and doripenem were performed as follows: For doripenem reactions, to a solution of doripenem (136 µg, 324 nmol, 1 eq) in PBS buffer (495 µl) was added cinnamaldehyde (213 µg, 1.6 µmol, 5 eq) in MeOH (5 µL) and nitromethane (3.2 µmol, 2 eq). For meropenem reactions, to a solution of meropenem trihydrate (124 µg, 324 nmol, 1 eq) in PBS buffer (495 µL) was added cinnamaldehyde (213 µg, 1.6 µmol, 5 eq) in MeOH (5 µl) and nitromethane 0.17 µL (3.2 µmol, 2 eq).

6.4.3 Determination of Enantioselectivity for the Nitro Michael Addition reaction (Carbapenems and BlaC-Carbapenem Artificial Enzymes)

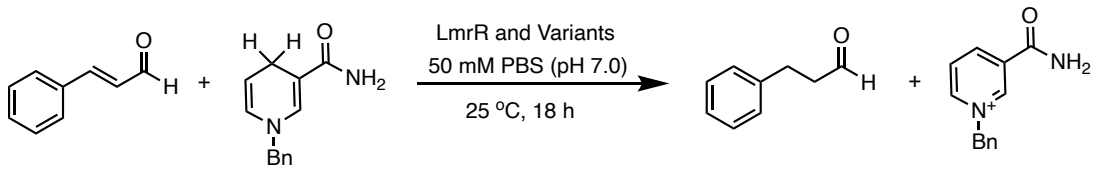
The reactions were scaled up 10 times with respect to the above screening conditions for the reactions with meropenem and doripenem in pH 7.0 PBS buffer and with 50% methanol and performed in 10-ml round bottom flasks with continuous stirring at room temperature for 24 h.

The reaction mixture was then extracted with DCM (10 mL × 3), and the organic fractions were combined, dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure to ca. 1 mL before purification by preparative TLC (EtOAc: hexane = 25:75). The silica gel was scraped from the plate and stirred in 1% MeOH:DCM (10 mL). The suspension was filtered and evaporated under reduced pressure. The resulting aldehyde was dissolved in methanol (5 mL), and to this was added ca. 5 equivalents of $NaBH_4$. The reaction was stirred overnight at room temperature.

The reaction mixture was then neutralized with 1 M HCl(aq), and then extracted with DCM (10 mL x 3). The organic fractions were combined, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The resulting alcohol was purified by preparative TLC (EtOAc:hexane = 35:65). The silica was scraped from the plate and stirred in 1% MeOH:DCM (10 mL). The solution was filtered and evaporated under vacuum. The purified alcohol was dissolved in 1 ml of 20% 2-propanol in hexane (HPLC grade). 20 µL of each sample was injected onto the HPLC.

The reactions catalysed by the artificial enzymes BlaC-meropenem and BlaC-doripenem were scaled up 10x with respect to the small-scale reactions above. Enantioselectivity was then determined using the same procedure as the small molecule catalysts.

6.4.4 Hydride Transfer Screening Reactions (LmrR and mutants)



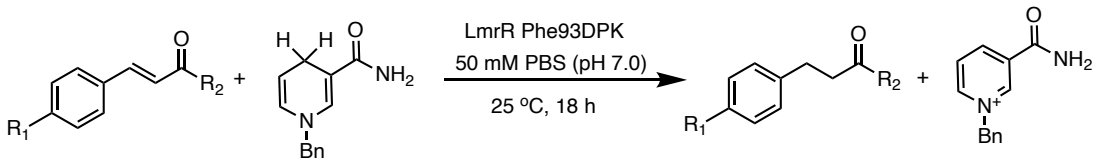
Catalysis screening reactions were performed with the purified enzyme variants. The enzymes were concentrated to 2 mg/mL. To 50 µL (100 µg, 6.8 nmol, 0.1 eq) of enzyme solution in the reaction buffer (50 mM NaP_i, 150 mM NaCl, pH 7.0) in a microcentrifuge tube was added a 5 µL solution of cinnamaldehyde (8.9 µg, 68 nmol, 1 eq) in methanol, followed by 5 µL solution of BNAH (28 µg, 136 nmol, 2 eq) in methanol and made up to 100 µL the reaction buffer. The reactions were placed in a thermomixer at 25 °C and 500 rpm and incubated for 18 hours. The reactions were halted by adding 200 µL of DCM and vortexed vigorously, then centrifuged to separate the layers (13,000 rpm, room temperature, 3 min) and 100 µL of the organic layer was removed and subjected to analysis by GC-MS. Substrate to product conversion was determined using a standard calibration curve of the product hydro-cinnamaldehyde. In addition to the enzyme variants, catalyst free controls and reactions with the wild type enzyme were also performed. The catalytic amino acid D-prolyl-L-lysine (DPK) was also used with the same catalyst loading as the enzymes.

Table 12. Triplicate conversion of substrate to the product hydro-cinnamaldehyde by LmrR and variants determined by comparison of the peak area against a standard calibration curve by GC-MS.

Catalyst	Conversion 1 (%)	Conversion 2 (%)	Conversion 3 (%)	Mean (%)	Standard Deviation (±)
V15-DPK	31	14	17	21	9.1
V15-LPK	15	11	11	12	2.5
V15-THZ	7	8	8	7	0.8
N19-DPK	18	17	15	17	1.8
N19-LPK	4	3	3	4	0.7
N19-THZ	33	33	28	31	3.0
M89-DPK	12	8	1	7	5.6
M89-LPK	5	4	5	5	0.6
M89-THZ	9	11	9	10	1.0
F93-DPK	68	57	49	58	9.4
F93-LPK	3	2	2	2	0.3
F93-THZ	13	10	13	12	1.9

Wt	2	2	2	2	0.2
DPK	20	21	18	20	1.8
-Ve	0	0	0	0	0.0

6.4.5 LmrR Phe93DPK Substrate Scope



All substrates were purchased from commercial retailers. The substrate scope for LmrR Phe93DPK, was performed as follows. To a microcentrifuge tube was added 500 µL of a stock solution of the enzyme (1 mg, 68 nmol, 1 eq) in PBS (50 mM NaP_i, 150 mM NaCl, pH 7.0). To this was added 400 µL of PBS buffer and then 50 µL of a stock solution of the α,β-unsaturated carbonyl substrate (680 nmol, 10 eq) in methanol followed by 50 µL of a stock solution of BNAH (725 µg, 3.4 mmol, 50 eq) in methanol. The reactions were placed in a thermoshaker and shook at 500 rpm for 48 hours at 25 °C.

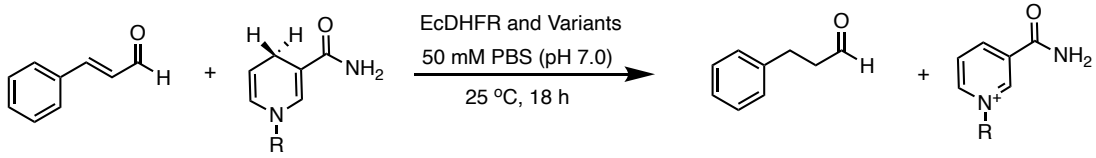
Following the completion of the reactions, DCM was added (500 µL) and the sample vortexed vigorously. The phases were separated by centrifugation (13,000 rpm, room temperature, 3 min). The organic phase was removed and placed in a new microcentrifuge tube. This was repeated a second time and the two extractions combined. The solvent was removed under nitrogen and the remaining residue re-dissolved in $CDCl_3$ (600 µL). Each sample was then subjected to ^1H NMR spectroscopy and the substrate to product conversion determined by integration of the aldehyde protons.

Table 13. Conversion of products determined by ^1H NMR spectroscopy using LmrR Phe93DPK as the catalyst.

Substrate	Product Conversion 1 (%)	Product Conversion 2 (%)	Product Conversion 3 (%)	Product Conversion Mean (%)
R_1 = Cl, R_2 = H	90	92	72	85
R_1 = F, R_2 = H	86	76	86	83
R_1 = Br, R_2 = H	92	100	100	97
R_1 = OMe, R_2 = H	100	100	95	98
R_1 = H, R_2 = H	79	77	61	72
R_1 = Me, R_2 = H	80	95	80	85

$R_1 = NO_2$, $R_2 = H$	81	100	86	89
$R_1 = Cl$, $R_2 = Me$	0	0	0	0

6.4.6 Hydride Transfer Screening Reactions (EcDHFR and mutants)

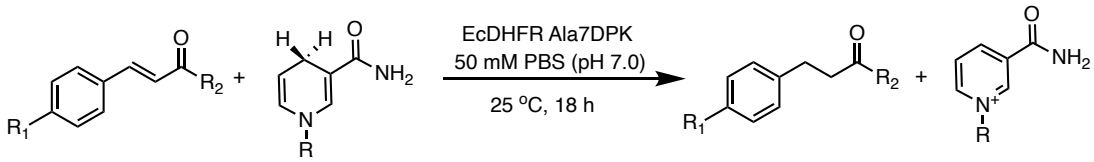


Catalysis screening reactions were performed with the purified enzyme variants. NADPH was purchased as the tetrasodium salt (Merck Life Sciences) and the concentration determined by measuring the UV absorbance at 340 nm using the extinction coefficient of 6220 M cm^{-1}. The enzymes were concentrated to 2 mg/mL. To 50 µL (100 µg, 5.12 nmol, 0.1 eq) of enzyme solution in the reaction buffer (50 mM NaP_i, 150 mM NaCl, pH 7.0) in a microcentrifuge tube was added a 5 µL solution of cinnamaldehyde (8.9 µg, 52 nmol, 1 eq) in methanol, followed by 5 µL solution of NADPH (77 µg, 104 nmol, 2 eq) in the reaction buffer and the reactions made up to 100 µL in the reaction buffer. The reactions were placed in a thermomixer at 25 °C and 500 rpm and incubated for 18 hours. The reactions were halted by adding 200 µL of DCM and vortexed vigorously, then centrifuged to separate the layers (13,000 rpm, room temperature, 3 min) and 100 µL of the organic layer was removed and subjected to analysis by GC-MS. Reaction yield was determined using a standard calibration curve of the product hydro-cinnamaldehyde. In addition to the enzyme variants, catalyst free controls and reactions with the wild type enzyme were also performed. The catalytic amino acid DPK was also used with the same catalyst loading as the enzymes.

Table 14. Triplicate substrate to product conversion of hydro-cinnamaldehyde by EcDHFR and variants determined by comparison of the peak area against a standard calibration curve by GC-MS.

Catalyst	Conversion 1 (%)	Conversion 2 (%)	Conversion 3 (%)	Mean (%)	Standard Deviation (±)
Ala7-DPK	89	88	99	92	6.3
Phe31-DPK	18	18	18	18	0.3
Ser49-DPK	0	0	0	0	0.0
Wt	0	0	0	0	0.0
DPK	0	0	0	0	0.0

6.4.7 EcDHFR Ala7DPK Substrate Scope



All substrates were purchased from commercial retailers. The substrate scope for EcDHFR Ala7DPK, was performed as follows. To a microcentrifuge tube was added 500 µL of a stock solution of the enzyme (100 µg, 5.12 nmol, 0.1 eq) in PBS (50 mM NaP_i, 150 mM NaCl, pH 7.0). To the enzyme was added 400 µL PBS buffer and then 50 µL of a stock solution of the α,β-unsaturated carbonyl substrate (512 nmol, 10 eq) in methanol followed by 50 µL of a stock solution of NADPH (725 µg, 3.4 mmol, 50 eq) in PBS. The reactions were placed in a thermoshaker and shook at 500 rpm for 48 hours at 25 °C.

Following the completion of the reactions, DCM was added (500 µL) and the sample vortexed vigorously. The phases were separated by centrifugation (13,000 room temperature, 3 min). The organic phase was removed and placed in a new microcentrifuge tube. This was repeated a second time and the two extractions combined. The solvent was removed under nitrogen and the remaining residue re-dissolved in $CDCl_3$ (600 µL). Each sample was then subjected to ^1H NMR spectroscopy and the substrate to product conversion determined by integration of the aldehyde protons.

Table 15. Conversion of products determined by ^1H NMR spectroscopy using EcDHFR Ala7DPK as the catalyst.

Substrate	Product Conversion 1 (%)	Product Conversion 2 (%)	Product Conversion 3 (%)	Product Conversion Mean (%)
R_1 = Cl, R_2 = H	33	60	54	49
R_1 = F, R_2 = H	30	18	25	24
R_1 = Br, R_2 = H	40	60	49	50
R_1 = OMe, R_2 = H	25	28	28	27
R_1 = H, R_2 = H	38	37	35	37
R_1 = Me, R_2 = H	33	43	36	37

$R_1 = NO_2$, $R_2 = H$	36	24	37	32
$R_1 = Cl$, $R_2 = Me$	0	0	0	0

6.5 Enzyme Kinetic Analysis

6.5.1 Determination of Vmax, k_{cat} and K_M

Product concentration was determined by measuring the peak area of the product from the GC-FID chromatogram and using a linear equation to solve for x (**equation 2**).

Equation 2.

$$Y = mx + c$$

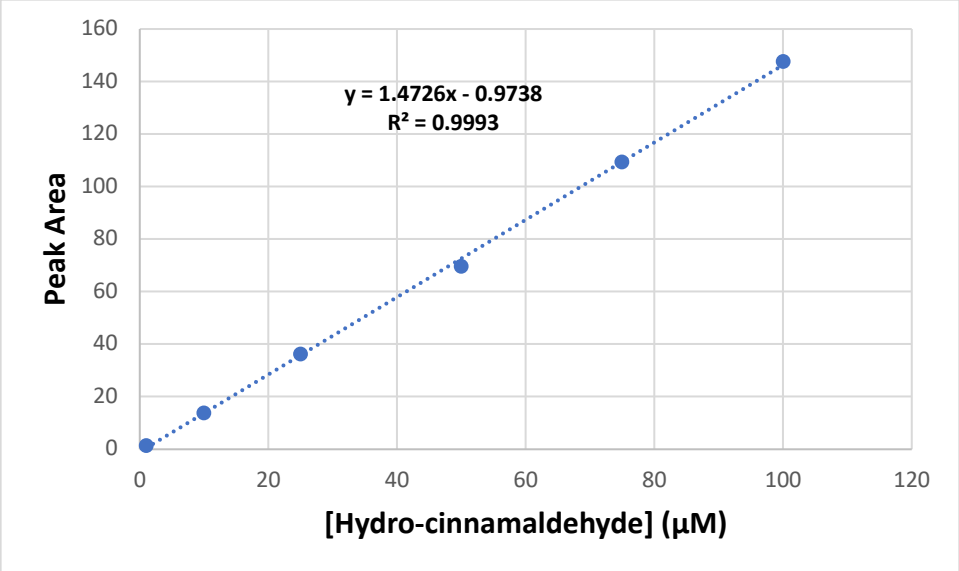


Figure 83. GC-FID hydro-cinnamaldehyde standard calibration curve for the determination of product concentration.

Vmax was determined using **equation 3**.

Equation 3.

$$Vmax = \frac{[Product]\ (\mu M)}{Time\ (s)}$$

k_{cat} was determined using **equation 4**.

Equation 4.

$$k_{cat} = \frac{Vmax\ (\mu M\ s^{-1})}{[Ez]\ (\mu M)}$$

K_M was determined using Prizm non-linear regression Michaelis-Menten output and was determined in μM.

To determine the catalytic efficiency k_{cat}/K_M **equation 6** was used.

Equation 6.

$$Catalytic\ efficiency = \frac{k_{cat}(s^{-1})}{K_M\ (M)}$$

6.5.3 Kinetic Analysis of LmrR Phe93DPK

To measure the enzyme kinetics of the artificial enzyme LmrR Phe93DPK saturating concentrations of the substrates cinnamaldehyde and BNAH were first determined. To 50 μM of the enzyme was added increasing concentrations of 0.5, 1.0, 1.5, 2.0 and 2.5 mM of cinnamaldehyde dissolved in 5 μL of methanol. To each reaction was added 5 mM of BNAH dissolved in 5 μL of methanol. The reactions were placed in a thermoshaker at 25 °C with agitation at 500 rpm for 2 hours. The reactions were halted with the addition of 200 μL of DCM and the organic phase was analysed using GC-FID to determine the product concentration. A second set of reactions was performed using 0.5 mM of cinnamaldehyde whilst varying the concentration of BNAH, 2.5, 5.0 and 10.0 mM. The reactions were again performed and analysed as above. With respect to an enzyme concentration of 50 μM the saturating concentrations of cinnamaldehyde and of BNAH were determined to be 1 mM and 5 mM respectively.

To determine the initial reaction rate in the linear phase, 50 μM of LmrR was incubated with 1 mM of cinnamaldehyde and 5 mM of BNAH. The products at each time point were analysed as above. The top of the linear phase (initial rate) was determined to be 2 hours.

Table 16. Time course for the reaction catalysed by LmrR-Phe93-DPK. Conversion of the substrate to product determined using GC-FID.

Time (h)	Mean conversion (%)
0.25	11 ± 0.7
0.50	16 ± 0.4
0.75	22 ± 1.9
1.25	30 ± 1.7
1.00	24 ± 1.1
1.50	33 ± 1.9
1.75	36 ± 2.4
2.00	41 ± 3.4
3.00	48 ± 1.7
4.00	53 ± 1.2
5.00	58 ± 2.8

After determining the saturating concentration of the substrates and the maximum rate, the k_{cat} and the $K_{M(BNAH)}$ of the enzyme was determined. To 50 µM of the enzyme in the reaction buffer (90 µL), was added 1 mM of cinnamaldehyde dissolved in 5 µL of methanol. The reaction was then initiated by adding BNAH dissolved in 5 µL of methanol in increasing concentrations. The reactions were performed for 2 hours at 25 °C with a shaking of 300 rpm. Three repeats of each reaction were performed. The products were extracted and analysed by GC-FID, see **section 6.14.** The mathematical analysis software Prizm was used to fit the data to a Michaelis-Menten curve. Mean rate and standard deviation were determined from three repeats.

Table 17. Rate determination of the reaction catalysed by LmrR Phe93DPK dependent on concentration of BNAH. Rates are an average of three repeats.

[BNAH] (µM)	Mean Rate x10^{-4} (s^{-1})
200	1.7 ± 0.18
400	2.6 ± 0.31
600	3.0 ± 0.94
800	4.6 ± 0.75
1000	5.5 ± 0.41
2000	7.3 ± 0.26
3000	10.0 ± 0.90
4000	10.1 ± 0.13
5000	9.2 ± 0.97

6.5.3 Kinetic Analysis of EcDHFR Ala7DPK

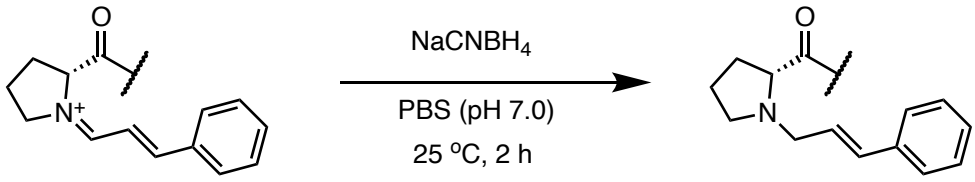
To measure the enzyme kinetics of the artificial enzyme EcDHFR Ala7DPK saturating concentrations of the substrates cinnamaldehyde and NADPH were first determined. 20 µM of the enzyme was incubated using the above protocol with 0.5, 1.0, 1.5, 2.0 and 2.5 mM of cinnamaldehyde dissolved in 5 µL of methanol. To each reaction was added 5 mM of NADPH dissolved in 5 µL of methanol. Reaction conditions and analysis of the substrate to product conversion were performed following the GC-MS protocols for the screening reactions above. A second set of reactions was performed using 1 mM of cinnamaldehyde whilst varying the concentration of NADPH, 0.1, 0.2 and 0.5 mM. The reactions were again performed and analysed as above. With respect to an enzyme concentration of 50 µM the saturating concentrations of cinnamaldehyde and of NADPH were determined to be 1 mM and 0.2 mM respectively.

After determining the saturating concentration of the substrates and the maximum rate, the k_{cat} and the $K_{0.5(NADPH)}$ of the enzyme were determined. To 20 µM of the enzyme in the reaction buffer (90 µL) was added 1 mM of cinnamaldehyde dissolved in 5 µL of methanol. The reaction was then initiated by adding NADPH dissolved in 5 µL of the reaction buffer in increasing concentrations. The reactions were performed for 2 hours at 25 °C with a shaking of 300 rpm. Three repeats of each reaction were performed. The products were extracted and analysed by GC-FID, see **section 6.14.** The mathematical analysis software Prizm was used to fit the data to a Michaelis-Menten curve. Mean rate and standard deviation were determined from three repeats.

Table 18. Rate determination of the reaction catalysed by EcDHFR Ala7DPK dependant on concentration of NADPH. Rates are an average of three repeats.

[NADPH] (µM)	Mean Rate x10^{-4} (s^{-1})
40	0.84 ± 0.44
50	1.68 ± 0.20
60	1.88 ± 0.45
75	2.29 ± 0.24
100	3.20 ± 0.27
125	3.51 ± 0.61
150	3.74 ± 0.36
250	4.61 ± 0.30
300	4.41 ± 0.41

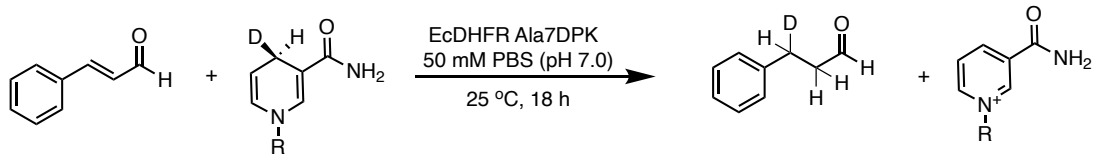
6.6 Determination of Iminium Ion Intermediates



To determine the formation of an iminium ion, the intermediate was trapped by covalent reduction of the carbonyl to the secondary amine forming a tertiary amine and subsequently analysed by mass spectrometry. The reaction was performed as follows. To a micro centrifuge tube was added 100 µg of enzyme in PBS (50 mM NaP_i, 150 mM NaCl, pH 7.0) and to this was added 20 eq of cinnamaldehyde (1 mM) in 10 µL of methanol. The reaction was placed in a thermoshaker for 2 hours at 25 °C with agitation at 500 rpm. After 2 hours 50 eq of sodium cyanoborohydride was added dissolved in methanol. The reactions were left overnight under the above reaction conditions.

The following day, the protein samples were buffer exchanged through micro centrifuge ultrafiltration columns, simultaneously removing the small molecules and exchanging into Tris buffer (50 mM, pH 8.0). 10 µL of the sample was removed and analysed by protein mass spectrometry. The remaining samples were then digested with 1:20 (chrymotrypsin:protein substrate) overnight at 25 °C. The digested samples were analysed by LC-MS

6.7 Isotope Incorporation Study



The incorporation of deuterium into cinnamaldehyde was performed as follows. NADPD concentration was determined by measuring the UV absorbance at 340 nm using the extinction coefficient of 6220 M cm^{-1}. To 50 µL (100 µg, 5.12 nmol, 0.1 eq) of enzyme solution in the reaction buffer (50 mM NaP_i, 150 mM NaCl, pH 7.0) in a microcentrifuge tube was added a 5 µL solution of cinnamaldehyde (8.9 µg, 68 nmol, 1 eq) in methanol, followed by 5 µL solution of NADPD (28 µg, 136 nmol, 2 eq) in the reaction buffer the reactions were made up to 100 µL with the reaction buffer. The reactions were placed in a thermomixer at 25 °C and 500 rpm and incubated for 18 hours. The reactions were halted by adding 200 µL of DCM and vortexed vigorously,

then centrifuged to separate the layers (13,000 rpm, room temperature, 3 min) and 100 µL of the organic layer was removed and subjected to analysis by GC-MS.

6.8 Kinetic Isotope Effect Experiments

To determine if there is a kinetic isotope effect at the hydride transfer step of the reaction the oxidation of NADPH and NADPD was measured comparatively using a UV plate reader assay. The concentrations of NADPH and NADPD were determined by UV absorbance at 340 nm and calculated using the extinction coefficient of 6220 M cm^{-1}. To 90 µL of the enzyme (final concentration 25 µM) on a 96 well plate, in the reaction buffer (NaP_i 50 mM, NaCl 150 mM, pH 7.0) was added 5 µL of NADPH or NADPD in the dissolved in the same buffer (final concentration 250 µM). The samples were placed in the plate reader (BMG Labtech FLUOstar OPTIMA microplate reader) at a fixed temperature of 25 °C for 5 min. After the allotted time the plate was ejected and 5 µL of cinnamaldehyde (final concentration 1 mM) in methanol was added to each reaction. The oxidation of NADPH/D was monitored at a wavelength of 340 nm for two hours. The reactions were performed in triplicate. Measurements were taken every minute for 2 hours.

Following completion of the reaction the slope was determined for both NADPH and NADPD. The kinetic isotope effect, if any, was determined using **equation 8**.

Equation 8.

$$KIE = \frac{Slope\ NADPH}{Slope\ NADPD}$$

6.9 Cofactor Recycling Reaction

To recycle the cofactor a scheme was devised where the NAD+ dependant enzyme glucose-6-phosphate dehydrogenase was coupled to the hydride transfer reaction catalysed by EcDHFR Ala7DPK. To a microcentrifuge tube was added 100 µM of the enzyme in the reaction buffer, 2 nM of glucose-6-phosphate dehydrogenase, 2 mM of glucose-6-phosphate and 1 mM of cinnamaldehyde in methanol. NADPH was then added in varying concentrations. The final reaction volume was made up to 100 µL with the reaction buffer. The reactions were placed in a thermomixer at 25 °C and 500 rpm and incubated for 18 hours. The reactions were halted by adding 200 µL of

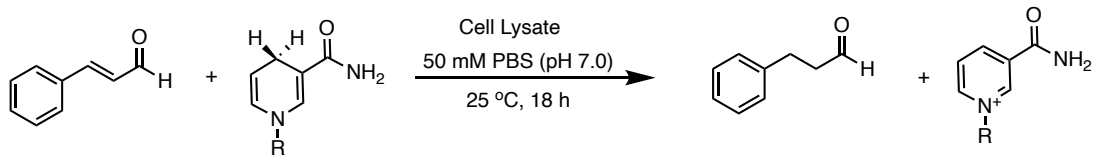
DCM and vortexed vigorously, then centrifuged to separate the layers (13,000 rpm, room temperature, 3 min) and 100 µL of the organic layer was removed and subjected to analysis by GC. Substrate to product conversion and TTN were determined from an average of three repeats.

The total turnover number (TTN) was determined using **equation 9**.

Equation 9.

$$TTN = \frac{Yield\ of\ product\ in\ moles}{Moles\ of\ cofactor\ added\ to\ the\ reaction}$$

6.10 Whole Cell Lysate Catalysis



To determine if the artificial enzyme EcDHFR Ala7DPK could catalyse the selective reduction of cinnamaldehyde using endogenous NADPH, 1 mL of cells with proven expression of EcDHFR Ala7DPK and BL21 (DE3) cells were normalised to an OD_{600} nm of 1.0. The cells were centrifuged (4,000 rpm, 4 °C, 5 min) and resuspend in PBS buffer (50 mM NaP_i, 150 mM NaCl, pH 7.0) 10 µg of lysozyme was added and the cells incubated at 37 °C for 1 h. 200 µg of cinnamaldehyde was then added to the lysed cells in 50 µL of MeOH. The reactions were incubated overnight at 25 °C and shaking at 500 rpm. The reactions were heated to 60 °C for 30 minutes and then centrifuged to remove the solids (13,000 rpm, room temperature, 10 min).

The supernatant was removed and to this was added 500 µL of DCM and vortexed vigorously, then centrifuged to separate the layers (13,000 rpm, room temperature, 3 min) and 100 µL of the organic layer were removed and subjected to analysis by GC-MS.

6.11 Unnatural Amino Acid Incorporation Efficiency Fluorescence Assays

Acetyl lysine, propionyl lysine, thioacetyl lysine and photocaged cysteine were already available in our laboratory. Cyclopropene lysine and Boc-lysine were purchased from commercial vendors. D-propyl-lysine and thiazolidine lysine were synthesised as outlined in **section 6.12**.

Chemically competent *E. coli* BL21(DE3) cells were transformed with the reporter plasmid, pET sfGFP(150TAG), and the appropriate PylRS plasmid. The transformation was conducted by incubation with plasmids on ice for 10 min, heat shock at 43 °C for 45 sec, incubation on ice for 2 min, and recovery in fresh LB media at 37 °C for 1 h with constant agitation. Cells were added into fresh LB media (10 mL) containing kanamycin (50 µg/mL; selection marker of the reporter plasmid) and spectinomycin (50 µg/mL, selection marker of the PylRS plasmid). Cells were cultured at 37 °C overnight with constant agitation.

The overnight culture was diluted into fresh LB media (25 mL) containing kanamycin (50 µg/ml) and spectinomycin (50 µg/mL). This culture was incubated at 37 °C with constant agitation until OD_{600} = 0.6 – 0.8, and isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to reach the final concentration of 1 mM. The culture was split into 50 mL falcon tubes (5 mL culture per tube), and unnatural amino acid was added, if required, to reach the final concentration of 5 mM (except for BocK, final concentration = 1 mM). Cultures were incubated at 25 °C or 37 °C for 18 h with constant agitation for protein production.

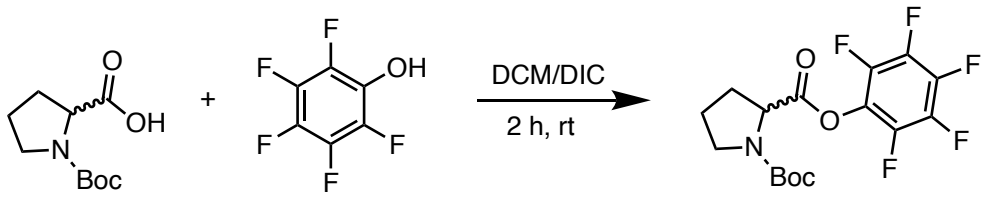
After incubation, OD_{600} of the cultures were recorded, and samples (1 mL) were taken from each culture. The samples were then centrifuged at 20 °C and (13,000 rpm, 20 °C, 5 min). Supernatants were removed, and the pellets were resuspended in PBS. The samples were centrifuged again at (13,000 rpm, 20 °C, 5 min) and the supernatants removed. The pellets were resuspended in PBS to OD_{600} = 1.0. The resuspension (200 µL) was placed into a 96-well plate, and the fluorescence was measured by a BMG Labtech FLUOstar OPTIMA microplate reader (excitation wavelength = 485 nm, emission wavelength = 520 nm, 20 flashes per well, 0.2 position delay, gain = 1000, shaking for 30 sec before plate reading at 25 °C). Three

biological repeats of each expression condition and two technical repeats of each fluorescent measurement were performed.

6.12 Synthesis of Unnatural amino acids and hydride donors

6.12.1 Synthesis of Secondary Amine Unnatural Amino Acids

D-prolyl-L-lysine (DPK), L-prolyl-L-lysine (LPK) and Thiazolidine lysine (ThzK) were synthesised using the same procedure using either D- or L- *N*-Boc protected proline as reagent **1** in the synthesis. ThzK was already available in our lab. The synthesis for ThzK was developed by a postdoctoral researcher Dr Xuifei Li and myself, following the synthesis of DPK and LPK.

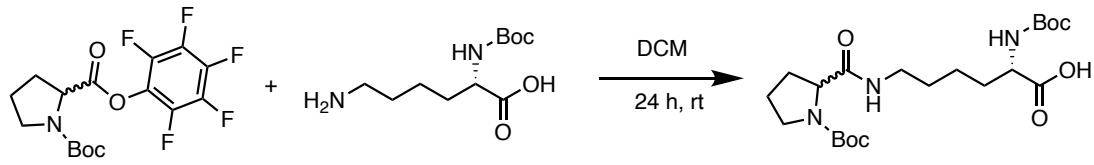


1 **2**

1 (1 g, 4.65 mmol, 1 eq) was dissolved in 20 mL of dichloromethane (DCM) in a round bottomed flask. To the solution was added pentafluorophenol (PFP) (0.94 g, 5.12 mmol, 1.1 eq) followed by Diisopropylcarbodiimide (DIC) (0.8 mL, 5.12 mmol, 1.1 eq). The mixture was left to stir at room temperature for 2 hours. The reaction mixture was then filtered under gravity and the solvent removed under vacuum. The resulting crude products were dissolved in 25% ethyl acetate in n-hexane and the solids removed by filtration. The solution was purified by flash silica chromatography (5% ethyl acetate in *n*-hexane) and the solvent was removed to produce **2** as a white crystalline solid. (Yield 75 – 80%)

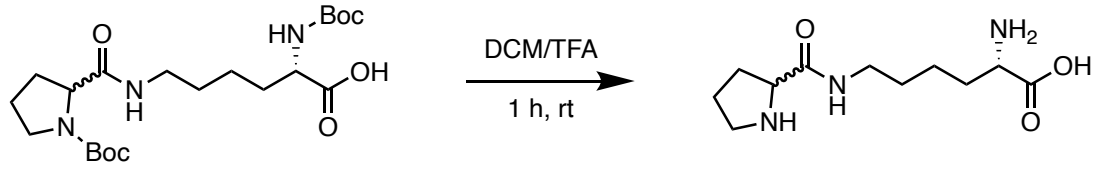
D-2 = ^{1}H NMR (600 MHz, $CDCl_3$): δ 4.70 – 4.54 (m, 1H), 3.70 – 3.88 (m, 2H), 2.53 – 2.31 (m, 1H), 2.30 – 2.13 (m, 1H), 2.12 – 1.93 (m, 2H) 1.50 – 1.41 (s, 9H) ppm. HRMS (ESI-(+)): m/z: calculated for $C_{16}H_{16}F_5NO_4Na$: 404.0897 $[M+Na]^+$; found, 404.0901

L-2 = ^{1}H NMR (600 MHz, $CDCl_3$): δ 4.70 – 4.54 (m, 1H), 3.70 – 3.88 (m, 2H), 2.53 – 2.31 (m, 1H), 2.30 – 2.13 (m, 1H), 2.12 – 1.93 (m, 2H) 1.50 – 1.43 (s, 9H) ppm. **HRMS** (ESI-(+)): m/z: calculated for $C_{16}H_{16}F_5NO_4Na$: 404.0897 $[M+Na]^+$; found, 404.0897



2 **3**

2 (1 g, 2.6 mmol, 1 eq) was dissolved in 100 mL of DCM in a round bottomed flask. To the solution was added *N*α-Boc-Lys-OH (642 mg, 2.6 mmol, 1 eq). The reaction was left to stir at room temperature overnight. The volatiles were then removed under vacuum and the resulting crude products were dissolved in DCM. The product was purified by flash silica chromatography using a step gradient. Firstly 5% methanol in DCM (5 CV) followed by 20% methanol in DCM (5 CV). The solvent was removed under vacuum to yield **3** as a thick colourless oil. Removal of PFP was determined by TLC and the product **3** was used directly in the next step.



3 **4**

3 was dissolved in 50 mL of DCM in a round bottomed flask and to this was added 50 mL of Trifluoroacetic acid (TFA). The reaction was left to stir for 1 hour. The volatiles were removed under vacuum to yield the crude products as a pale-yellow oil. The oil was dissolved in a minimal volume of methanol and 50 mL of dry ice-cold diethyl ether was added rapidly to precipitate the solids. The ether was poured off and any remaining solvents were removed under vacuum to yield **4** as an amino acid-TFA salt. **4** was dissolved in dH_2O, flash frozen in liquid N_2 and lyophilised overnight to yield a white powder. (yield 90%)

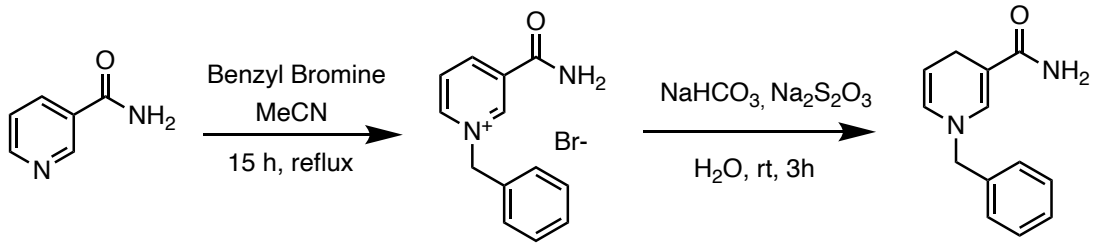
D-4 = ^{1}H NMR (600 MHz, D_2O): δ = 4.17 (t, *J* = 7.2 Hz, 1H), 3.88 (t, *J* = 6.3 Hz, 1H), 3.36 – 3.12 (m, 4H), 2.38 – 2.33 (m, 1H) 2.02 – 1.76 (m, 5H), 1.58 – 1.32 (m, 4H). **^{13}C NMR** (150 MHz, $CDCl_3$): d = 170.6, 168.4, 59.8, 54.4, 45.9, 38.8, 29.8, 28.3, 28.1 23.6, 21.9 ppm. **HRMS** (ESI-(+)): m/z calculated for $C_{11}H_{22}N_3O_3$: 244.1661 $[M+H]^+$; found: 244.1656

L-4 = ^{1}H NMR (600 MHz, D_2O): δ = 4.26 (t, *J* = 7.2 Hz, 1H), 3.97 (t, *J* = 6.3 Hz, 1H), 3.48 - 3.25 (m, 4H) 2.5 – 2.4 (m, 1H) 2.15 – 1.85 (m, 5H), 1.68 – 1.45 (m, 4H). **^{13}C NMR** (150 MHz, $CDCl_3$): d = 170.5, 168.3, 61.1, 53.8, 47.3, 40.2, 31.1, 31.0, 29.7,

25.1, 23.2 ppm. **HRMS** (ESI-(+)): m/z calculated for $C_{11}H_{22}N_3O_3$: 244.1661 $[M+H]^+$; found: 244.1662

The amount of amino acid present in the salt, was determined by mixing 5 mg of the unnatural amino acid salt with 5 mg of *N*-boc proline and measuring the integrals of the chiral protons. This allowed a rough estimate of how much amino acid to TFA was in the samples prior to addition to the cell cultures.

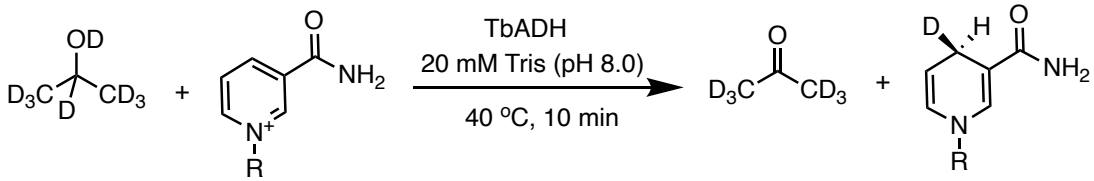
6.12.2 Synthesis of Benzyl-nicotinamide (BNAH)



Benzyl-nicotinamide was synthesised using a literature procedure. The final product was confirmed by ^{1}H NMR spectroscopy.

1H NMR (500 MHz, CDCl3): δ = 7.30 – 7.20 (m, 5H), 7.08 (d, *J* = 1.6 Hz, 1H), 5.66 (dq, *J* = 8, 3.5 Hz, 1H), 5.30 (s br, 2H), 4.68 (dt, *J* = 8, 3.5 Hz, 1H), 4.22 (s, 2H), 3.09 (dd, J = 3.3, 1.6 Hz, 2H)

6.12.3 Enzymatic Synthesis of [4*R*-^{2}H]-Nicotinamide Adenine Dinucleotide Phosphate (NADPD)



The lyophilised TbADH (in 20 mM Tris, pH 9.0) was reconstituted in 500 µL of dH_2O in a microcentrifuge tube. To the enzyme was added 7 mg of $NADP^+$ sodium salt dissolved in 500 µL of dH_2O. The reaction was placed in a water bath at 40 °C for 2 minutes and then 200 µL of D8 isopropyl alcohol was added. The reaction was returned to the water bath for 10 minutes. Following the end of the reaction the UV absorbance of the sample was measured at 340 nm. If the absorbance was at maximum of the UV detector the sample was immediately purified by strong anion exchange chromatography.

Purification of the NADPD was performed on a BioRad NGC chromatography system using a Dionex ProPac SAX-10 (22 x 250 mm) ion exchange preparative scale column. Buffer A was 20 mM Tris base (pH 9.0) and buffer B was 20 mM Tris base, 1 M NaCl, (pH 9.0). The entire sample was injected via loop injection and the sample was eluted using a gradient of 0 – 40% buffer B over 30 minutes and then 40 – 100% buffer B over 12 minutes. Elution was monitored at 340 nm. Fractions that were detected at 340 nm were pooled together and the pH modified to pH 7.5. The solvent was removed by lyophilisation and the sample redissolved in 1 mL of dH_2O. The concentration of NADPD was determined using UV spectroscopy my measuring the absorbance at 340 nm using the extinction coefficient of 6220 M cm^{-1}. Isotopic enrichment was determined using high resolution mass spectrometry.

6.13 NMR Spectroscopy

NMR Spectroscopy was performed on either a Bruker 600 MHz Avance equipped with a He cooled cryoprobe platform, a Bruker 500 MHz Avance equipped with a Prodigy N_2 cryoprobe or a Bruker 300 MHz Fourier NMR spectrometer. Data was analysed using Bruker TopSpin.

6.14 Mass Spectrometry

6.14.1 Protein Liquid Chromatography-Mass Spectrometry

Protein liquid chromatography-mass spectrometry (LC-MS) was acquired on a Waters Acquity H-Class UPLC system coupled to a Waters Synapt G2-Si quadrupole time of flight mass spectrometer. The column used was a Waters Acquity UPLC Protein C4 BEH column 300 Å, 1.7 μm (2.1 x 100 mm) held at 60 °C. The flow rate was 0.2 mL/min and the gradient employed is highlighted in the table below. Mass spectrometry data was collected in positive electrospray ionisation mode and the data analysed using Waters MassLynx 4.1. Deconvoluted mass spectra were generate using the maximum entropy 1 (MaxEnt 1) software.

Table 19. Protein LC-MS chromatography parameters.

Time (min)	A% (H_2O 0.1% CHOOH)	B% (ACN 0.1% CHOOH)
0	95	5
3	95	5
50	35	65
52	3	97
54	3	97
56	95	5
60	95	5

6.14.2 High Resolution Mass Spectrometry for Characterisation (HR-MS)

HR-MS for characterisation of small molecules was performed on a Waters Xevo G2-XS quadrupole time of flight mass spectrometer. The samples were injected directly into the mass spectrometer and the data collected in positive electrospray ionisation mode (ESI +ve). The data was analysed using Waters MassLynx 4.1.

6.14.3 Gas Chromatography Mass Spectrometry (GC-MS)

GC-MS was performed on a Perkin Elmer Clarus 680 GC system coupled to a Perking Elmer Clarus SQ 8C quadrupole mass spectrometer in electron impact ionisation mode. The column used was a Perkin Elmer Elite-1 30 m (0.25 mm x 0.25 μm). The inlet temperature was set to 150 °C with a split ratio of 19:1. An injection volume of 1 μL was used. The temperature programme started at 50 °C and held for 1 minute, then ramped up to 220 °C at 15 °C /min and held for 2 minutes. Data was analysed using Perkin and Elmer TurboMass software.

6.14.4 Liquid Chromatography-Mass Spectrometry (LC-MS)

Liquid chromatography-mass spectrometry (LC-MS) was acquired on a Waters Acquity H-Class UPLC system coupled to a Waters Synapt G2-Si quadrupole time of flight mass spectrometer. The column used was a Waters Acquity UPLC C18 BEH column 75 Å, 1.7 μm (2.1 x 100 mm) held at 40 °C. The flow rate was 0.3 mL/min and the gradient employed is highlighted in the table below. Mass spectrometry data was

collected in positive electrospray ionisation mode and the data analysed using Waters MassLynx 4.1.

Table 20. LC-MS chromatography parameters.

Time (min)	A% (H_2O 0.1% CHOOH)	B% (ACN 0.1% CHOOH)
0	97	3
1	97	3
30	40	60
31	5	95
33	5	95
34	97	3
40	97	3

6.15 Gas Chromatography Flame Ionisation Detection (GC-FID)

GC-FID was performed on an Agilent 7890A GC system equipped with a flame ionisation detector (FID). The column used was a Restek RE-bDEXsm 30 m (0.32 µm x 0.25 µm). The inlet temperature was set to 200 °C with a split ratio of 10:1. An injection volume of 3 µL was used. The temperature programme started at 80 ºC and held for 2 minutes, then ramped up to 200 °C at 15 °C /min and held for 2 minutes. Data was analysed using Agilent Chem Station software.

6.16 Analytical Size Exclusion Chromatography

Analytical size exclusion chromatography was performed on an Agilent infinity 1260 HPLC. The column used was an Agilent Bio SEC-3 150 A, 3 mm (4.6 x 300 mm) held at 20 °C. The elution was isocratic using PBS (50 mM NaP_i, 150 mM NaCl, pH 7.0). The flow rate was 1 mL/min. Detection was performed at 280 nm.

6.17 Chiral High-Performance Liquid Chromatography (Chiral HPLC)

Chiral separations were performed on an Agilent infinity 1260 HPLC. The column used was a Phenomonex Cellulose Lux-1 analytical chiral column held at 20 °C. The

elution was isocratic 75:25 *n*-hexane:2-propanol, the flow rate was 0.5 mL/min and detection was performed at a wavelength of 210 nm.

7. Appendix

7.1 Nucleotide and Amino Acid Sequences

7.1.1 *Mycobacterium tuberculosis* β-lactamase (BlaC)

Nucleotide Sequence:

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGGTGC
AGATCTGGCAGATCGTTTTGCAGAACTGGAACGTCGTTATGATGCACGTCTGGGTGTTTATGTTCCGG
CAACCGGCACCACCGCAGCAATTGAATATCGTGCAGATGAACGTTTTGCATTTTGCAGCACCTTTAAA
GCACCGCTGGTTGCAGCCGTTCTGCATCAGAATCCGCTGACACATCTGGATAAACTGATTACCTATAC
CTCCGATGATATTCGTAGCATTAGTCCGGTTGCACAGCAGCATGTTCAGACCGGTATGACCATTGGTC
AGCTGTGTGATGCAGCAATTCGTTATAGTGATGGCACCGCAGCCAATCTGCTGCTGGCAGATTTAGGT
GGTCCTGGTGGTGGTACAGCAGCCTTTACCGGTTATCTGCGTAGCCTGGGTGATACCGTTAGCCGTCT
GGATGCAGAAGAACCGGAACTGAATCGTGATCCGCCTGGTGATGAACGTGATACCACCACACCGCATG
CCATTGCACTGGTTCTGCAGCAGCTGGTTCTGGGTAATGCACTGCCTCCGGATAAACGTGCACTGCTG
ACCGATTGGATGGCACGTAATACCACCGGTGCCAAACGTATTCGTGCAGGTTTTCCGGCAGATTGGAA
AGTTATTGATAAAACCGGTACGGGTGATTATGGTCGTGCAAATGATATTGCAGTTGTTTGGAGCCCGA
CCGGTGTTCCGTATGTTGTTGCAGTTATGAGCGATCGTGCCGGTGGTGGCTATGATGCCGAACCGCGT
GAAGCACTGCTGGCGGAAGCAGCAACCTGTGTTGCCGGTGTTCTGGCATAA

Amino Acid Sequence:

MGSSHHHHHHSSGLVPRGSHMGADLADRFAELERRYDARLGVYVPATGTTAAIEYRADERFAFCSTFK
APLVAAVLHQNPLTHLDKLITYTSDDIRSISPVAQQHVQTGMTIGQLCDAAIRYSDGTAANLLLADLG
GPGGGTAAFTGYLRSLGDTVSRLDAEEPELNRDPPGDERDTTTPHAIALVLQQLVLGNALPPDKRALL
TDWMARNTTGAKRIRAGFPADWKVIDKTGTGDYGRANDIAVVWSPTGVPYVVAVMSDRAGGGYDAEPR
EALLAEAATCVAGVLA

7.1.2 *Mycobacterium tuberculosis* β-lactamase (BlaC Glu166Ala)

Nucleotide Sequence:

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGGTGC
AGATCTGGCAGATCGTTTTGCAGAACTGGAACGTCGTTATGATGCACGTCTGGGTGTTTATGTTCCGG
CAACCGGCACCACCGCAGCAATTGAATATCGTGCAGATGAACGTTTTGCATTTTGCAGCACCTTTAAA
GCACCGCTGGTTGCAGCCGTTCTGCATCAGAATCCGCTGACACATCTGGATAAACTGATTACCTATAC
CTCCGATGATATTCGTAGCATTAGTCCGGTTGCACAGCAGCATGTTCAGACCGGTATGACCATTGGTC
AGCTGTGTGATGCAGCAATTCGTTATAGTGATGGCACCGCAGCCAATCTGCTGCTGGCAGATTTAGGT
GGTCCTGGTGGTGGTACAGCAGCCTTTACCGGTTATCTGCGTAGCCTGGGTGATACCGTTAGCCGTCT
GGATGCAGAAGCACCGGAACTGAATCGTGATCCGCCTGGTGATGAACGTGATACCACCACACCGCATG
CCATTGCACTGGTTCTGCAGCAGCTGGTTCTGGGTAATGCACTGCCTCCGGATAAACGTGCACTGCTG
ACCGATTGGATGGCACGTAATACCACCGGTGCCAAACGTATTCGTGCAGGTTTTCCGGCAGATTGGAA
AGTTATTGATAAAACCGGTACGGGTGATTATGGTCGTGCAAATGATATTGCAGTTGTTTGGAGCCCGA
CCGGTGTTCCGTATGTTGTTGCAGTTATGAGCGATCGTGCCGGTGGTGGCTATGATGCCGAACCGCGT
GAAGCACTGCTGGCGGAAGCAGCAACCTGTGTTGCCGGTGTTCTGGCATAA

Amino Acid Sequence:

MGSSHHHHHHSSGLVPRGSHMGADLADRFAELERRYDARLGVYVPATGTTAAIEYRADERFAFCSTFK
APLVAAVLHQNPLTHLDKLITYTSDDIRSISPVAQQHVQTGMTIGQLCDAAIRYSDGTAANLLLADLG
GPGGGTAAFTGYLRSLGDTVSRLDAEAPELNRDPPGDERDTTTPHAIALVLQQLVLGNALPPDKRALL
TDWMARNTTGAKRIRAGFPADWKVIDKTGTGDYGRANDIAVVWSPTGVPYVVAVMSDRAGGGYDAEPR
EALLAEAATCVAGVLA

7.1.3 *Mycobacterium tuberculosis* D, D – Transpeptidase (Rv3330)

Nucleotide Sequence:

ATGGGTCATATGCCGTATAAGGTTAGTACCCCGCCGGCAGTGGATAGTAGCGAAGTTCCTGCAGCCGG
CGAACCTCCTTTACCTTTAGTTGTTCCGCCGACCCCGGTTGGTGGTAATGCATTAGGTGGCTGCGGCA
TTATTACCGCCCCGGGTAGCGCACCTGCACCTGGTGACGTGAGCGCAGAGGCATGGTTAGTGGCTGAT
CTGGATAGCGGTGCCGTGATTGCCGCAAGAGATCCTCATGGTCGCCATCGTCCGGCTAGTGTTATTAA
AGTTCTGGTTGCAATGGCCAGCATTAATACCCTGACCCTGAATAAAAGCGTGGCCGGCACCGCTGATG
ATGCTGCAGTGGAGGGTACCAAAGTGGGTGTGAATACCGGCGGCACCTATACCGTGAATCAGCTGCTG
CATGGCCTGCTGATGCATAGCGGTAATGATGCAGCATATGCATTAGCCCGCCAGCTGGGTGGTATGCC
TGCAGCATTAGAAAAAATTAACCTGCTGGCCGCAAAGCTGGGTGGTAGAGATACCCGTGTTGCCACCC
CTAGTGGTCTGGATGGTCCTGGTATGAGCACCAGTGCATATGATATTGGTCTGTTTTACCGCTACGCA
TGGCAGAATCCGGTGTTTGCTGATATTGTGGCAACCCGTACCTTTGATTTTCCGGGCCACGGCGATCA
TCCGGGTTATGAACTGGAAAATGATAACCAGCTGCTGTATAACTACCCGGGCGCCTTAGGTGGCAAAA
CCGGTTATACCGATGATGCCGGTCAGACCTTTGTTGGTGCAGCCAATCGCGATGGTCGTCGTCTTATG
ACCGTGCTGCTGCATGGCACCCGTCAACCTATTCCGCCTTGGGAACAGGCCGCACATCTGTTAGATTA
TGGCTTTAATACCCCGGCAGGTACCCAGATTGGCACCTTAATTGAACCGGATCCGAGTCTGATGAGCA
CCGATCGTAATCCGGCAGATCGTCAGCGTGTGGATCCGCAGGCAGCAGCACGTCTCGAGCACCACCAC
CACCACCACTGA

Amino Acid Sequence:

MGHMPYKVSTPPAVDSSEVPAAGEPPLPLVVPPTPVGGNALGGCGIITAPGSAPAPGDVSAEAWLVAD
LDSGAVIAARDPHGRHRPASVIKVLVAMASINTLTLNKSVAGTADDAAVEGTKVGVNTGGTYTVNQLL
HGLLMHSGNDAAYALARQLGGMPAALEKINLLAAKLGGRDTRVATPSGLDGPGMSTSAYDIGLFYRYA
WQNPVFADIVATRTFDFPGHGDHPGYELENDNQLLYNYPGALGGKTGYTDDAGQTFVGAANRDGRRLM
TVLLHGTRQPIPPWEQAAHLLDYGFNTPAGTQIGTLIEPDPSLMSTDRNPADRQRVDPQAAARLEHHH
HHH

7.1.4 Superfolder Green Fluorescent Protein (sfGFP Asp150TAG)

Nucleotide Sequence:

ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGTGATGT
GAATGGCCATAAATTTAGCGTTCGTGGCGAAGGCGAAGGTGATGCGACCAACGGTAAACTGACCCTGA
AATTTATTTGCACCACCGGTAAACTGCCGGTTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGC
GTTCAGTGCTTTAGCCGCTATCCGGATCATATGAAACGCCATGATTTCTTTAAAAGCGCGATGCCGGA
AGGCTATGTGCAGGAACGTACCATTAGCTTCAAAGATGATGGCACCTATAAAACCCGTGCGGAAGTTA
AATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGTATTGATTTTAAAGAAGATGGCAAC
ATTCTGGGTCATAAACTGGAATATAATTTCAACAGCCATTAGGTGTATATTACCGCCGATAAACAGAA
AAATGGCATCAAAGCGAACTTTAAAATCCGTCACAACGTGGAAGATGGTAGCGTGCAGCTGGCGGATC
ATTATCAGCAGAATACCCCGATTGGTGATGGCCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACC
CAGAGCGTTCTGAGCAAAGATCCGAATGAAAAACGTGATCATATGGTGCTGCTGGAATTTGTTACCGC
CGCGGGCATTACCCACGGTATGGATGAACTGTATAAAGGCAGCCACCATCATCATCACCATTAA

Amino Acid Sequence:

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYG
VQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGN
ILGHKLEYNFNSH*VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLST
QSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

7.1.5 *Lactococcus lactis* multidrug resistant regulator protein (LmrR)

Nucleotide Sequence:

ATGGGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATGTCATCCTGCTGAATGTCCTGAA
ACAAGGCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAAC
TGAATGAAGCCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGG
GGTGATGAAAGTCAAGGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACATGCG
CCTGGCGTTCGAATCCTGGAGTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTG
AAGCGATCAAAGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MGAEIPKEMLRAQTNVILLNVLKQGDNYVYGIIKQVKEASNGEMELNEATLYTIFDRLEQDGIISSYW
GDESQGGRRKYYRLTEIGHENMRLAFESWSRVDKIIENLEANKKSEAIKGSSHHHHHHSSG

7.1.6 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Val15TAG)

Nucleotide Sequence:

ATGGGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATTAGATCCTGCTGAATGTCCTGAA
ACAAGGCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAAC
TGAATGAAGCCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGG
GGTGATGAAAGTCAAGGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACATGCG
CCTGGCGTTCGAATCCTGGAGTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTG
AAGCGATCAAAGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MGAEIPKEMLRAQTN*ILLNVLKQGDNYVYGIIKQVKEASNGEMELNEATLYTIFDRLEQDGIISSYW
GDESQGGRRKYYRLTEIGHENMRLAFESWSRVDKIIENLEANKKSEAIKGSSHHHHHHSSG

7.1.7 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Asp19TAG)

Nucleotide Sequence:

ATGGGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATGTCATCCTGCTGTAGGTCCTGAA
ACAAGGCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAAC
TGAATGAAGCCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGG
GGTGATGAAAGTCAAGGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACATGCG
CCTGGCGTTCGAATCCTGGAGTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTG
AAGCGATCAAAGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MGAEIPKEMLRAQTNVILL*VLKQGDNYVYGIIKQVKEASNGEMELNEATLYTIFDRLEQDGIISSYW
GDESQGGRRKYYRLTEIGHENMRLAFESWSRVDKIIENLEANKKSEAIKGSSHHHHHHSSG

7.1.8 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Met89TAG)

Nucleotide Sequence:

ATGGGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATGTCATCCTGCTGAATGTCCTGAA
ACAAGGCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAAC
TGAATGAAGCCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGG
GGTGATGAAAGTCAAGGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACTAGCG
CCTGGCGTTCGAATCCTGGAGTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTG
AAGCGATCAAAGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MGAEIPKEMLRAQTNVILLNVLKQGDNYVYGIIKQVKEASNGEMELNEATLYTIFDRLEQDGIISSYW
GDESQGGRRKYYRLTEIGHEN*RLAFESWSRVDKIIENLEANKKSEAIKGSSHHHHHHSSG

7.1.9 *Lactococcus lactis* multidrug resistant regulator protein (LmrR Phe93TAG)

Nucleotide Sequence:

GGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATGTCATCCTGCTGAATGTCCTGAAACA
AGGCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAACTGA
ATGAAGCCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGGGGT
GATGAAAGTCAAGGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACATGCGCCT
GGCGTAGGAATCCTGGAGTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTGAAG
CGATCAAAGGCAGCAGCCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MGAEIPKEMLRAQTNVILLNVLKQGDNYVYGIIKQVKEASNGEMELNEATLYTIFDRLEQDGIISSYW
GDESQGGRRKYYRLTEIGHENMRLA*ESWSRVDKIIENLEANKKSEAIKGSSHHHHHHSSG

7.1.10 *Escherichia coli* Dihydrofolate Reductase (EcDHFR)

Nucleotide Sequence:

ATGATCAGTCTGATTGCGGCGTTAGCGGTAGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAA
CCTGCCTGCCGATCTCGCCTGGTTTAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATA
CCTGGGAATCAATCGGTCGTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTACG
GACGATCGCGTAACGTGGGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAAT
CATGGTGATTGGCGGCGGTCGCGTTTATGAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGC
ATATCGACGCAGAAGTGGAAGGCGACACCCATTTCCCGGATTACGAGCCGGATGACTGGGAATCGGTA
TTCAGCGAATTCCACGATGCTGATGCGCAGAACTCTCACAGCTATTGCTTTGAGATTCTGGAGCGGCG
GGGCAGCAGCCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MISLIAALAVDRVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRKNIILSSQPGT
DDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESV
FSEFHDADAQNSHSYCFEILERRGSSHHHHHHSSG

7.1.11 *Escherichia coli* Dihydrofolate Reductase (EcDHFR Ala7TAG)

Nucleotide Sequence:

ATGATCAGTCTGATTGCGTAGTTAGCGGTAGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAA
CCTGCCTGCCGATCTCGCCTGGTTTAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATA
CCTGGGAATCAATCGGTCGTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTACG
GACGATCGCGTAACGTGGGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAAT
CATGGTGATTGGCGGCGGTCGCGTTTATGAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGC
ATATCGACGCAGAAGTGGAAGGCGACACCCATTTCCCGGATTACGAGCCGGATGACTGGGAATCGGTA
TTCAGCGAATTCCACGATGCTGATGCGCAGAACTCTCACAGCTATTGCTTTGAGATTCTGGAGCGGCG
GGGCAGCAGCCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MISLIA*LAVDRVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRKNIILSSQPGT
DDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESV
FSEFHDADAQNSHSYCFEILERRGSSHHHHHHSSG

7.1.12 *Escherichia coli* Dihydrofolate Reductase (EcDHFR Phe31TAG)

Nucleotide Sequence:

ATGATCAGTCTGATTGCGGCGTTAGCGGTAGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAA
CCTGCCTGCCGATCTCGCCTGGTAGAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATA
CCTGGGAATCAATCGGTCGTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTACG
GACGATCGCGTAACGTGGGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAAT
CATGGTGATTGGCGGCGGTCGCGTTTATGAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGC
ATATCGACGCAGAAGTGGAAGGCGACACCCATTTCCCGGATTACGAGCCGGATGACTGGGAATCGGTA
TTCAGCGAATTCCACGATGCTGATGCGCAGAACTCTCACAGCTATTGCTTTGAGATTCTGGAGCGGCG
GGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MISLIAALAVDRVIGMENAMPWNLPADLAW*KRNTLNKPVIMGRHTWESIGRPLPGRKNIILSSQPGT
DDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESV
FSEFHDADAQNSHSYCFEILERRGSSHHHHHHSSG

7.1.13 *Escherichia coli* Dihydrofolate Reductase (EcDHFR Ser49TAG)

Nucleotide Sequence:

ATGATCAGTCTGATTGCGGCGTTAGCGGTAGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAA
CCTGCCTGCCGATCTCGCCTGGTTTAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATA
CCTGGGAATAGATCGGTCGTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTACG
GACGATCGCGTAACGTGGGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAAT
CATGGTGATTGGCGGCGGTCGCGTTTATGAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGC
ATATCGACGCAGAAGTGGAAGGCGACACCCATTTCCCGGATTACGAGCCGGATGACTGGGAATCGGTA
TTCAGCGAATTCCACGATGCTGATGCGCAGAACTCTCACAGCTATTGCTTTGAGATTCTGGAGCGGCG
GGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCTAA

Amino Acid Sequence:

MISLIAALAVDRVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWE*IGRPLPGRKNIILSSQPGT
DDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESV
FSEFHDADAQNSHSYCFEILERRGSSHHHHHHSSG

7.1.14 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCACCCT
GCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTATAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTTGCCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG

CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFCQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.15 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (ThzKRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCACCCT
GCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGGTGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGGGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGAGCCCGACCCTGTATAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTGTGCAATTTGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGGTAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVGAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLSPTLYN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFVQFGSGCTRENLEALIKEFLDYLEIDFE
IVGGSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.16 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (AcKRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCACCCT
GCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGGTGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGGGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA

GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGATGGCTCCGACCATTTTTAAC
TATGCTCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTTTTCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVGAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMMAPTIFN
YARKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFFQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.17 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (PrKRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCACCCT
GCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTTTAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTACCCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLFN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFTQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.18 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (PCCRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCACCCT
GCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTATAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTcagTTTgcgCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAgGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGatgGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGtggATTGGCGCGg
ggTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTG

Amino Acid Sequence:

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVQFAQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAMVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.19 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCATACCGGCACCCT
GCATAAAATCAAACATCGTGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTATAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTTGCCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSHTGTLHKIKHREVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFCQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.20 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg ThzKRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCATACCGGCACCCT
GCATAAAATCAAACATCGTGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGAGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGAGCCCGACCCTGTATAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTGTGCAATTTGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGGTAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSHTGTLHKIKHREVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVGAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLSTLYNY
LRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFVQFGSGCTRENLEALIKEFLDYLEIDFEI
VGGSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRSE
SYYNGISTNL

7.1.21 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg AcKRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCATACCGGCACCCT
GCATAAAATCAAACATCGTGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA

GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGATGGCTCCGACCATTTTTAAC
TATGCTCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTTTTCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSHTGTLHKIKHREVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVGAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMMAPTIFN
YARKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFFQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.22 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg PrKRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCATACCGGCACCCT
GCATAAAATCAAACATCGTGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC
TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTTTAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTACCCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSHTGTLHKIKHREVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLFN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFTQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.23 *Methanosarcina bakeri* Pyrrolysyl-tRNA Synthetase (Arg19His/His29Arg PCCRS)

Nucleotide Sequence:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCATACCGGCACCCT
GCATAAAATCAAACATCGTGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATC

TGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACC
TGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCGAAAGCAAAAA
CAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTG
CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGC
CCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCA
GCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTC
GTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGT
GAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAAT
TAAAAGCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGA
GCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTATAAC
TATCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCG
CAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTCAGTTTGCGCAAATGGGCAGCG
GCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAGGAATTCCTGGATTATCTGGAAATCGACTTCGAA
ATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGATGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGG
GGTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCCGTAGC
GAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

Amino Acid Sequence:

MDKKPLDVLISATGLWMSHTGTLHKIKHREVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKT
CKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPS
PAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDR
EDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYN
YLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVQFAQMGSGCTRENLEALIKEFLDYLEIDFE
IVGDSCMVYGDTLDIMHGDLELSSAMVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
ESYYNGISTNL

7.1.24 *Thermoanaerobacter brockii* Alcohol Dehydrogenase (TbADH)

Nucleotide Sequence:

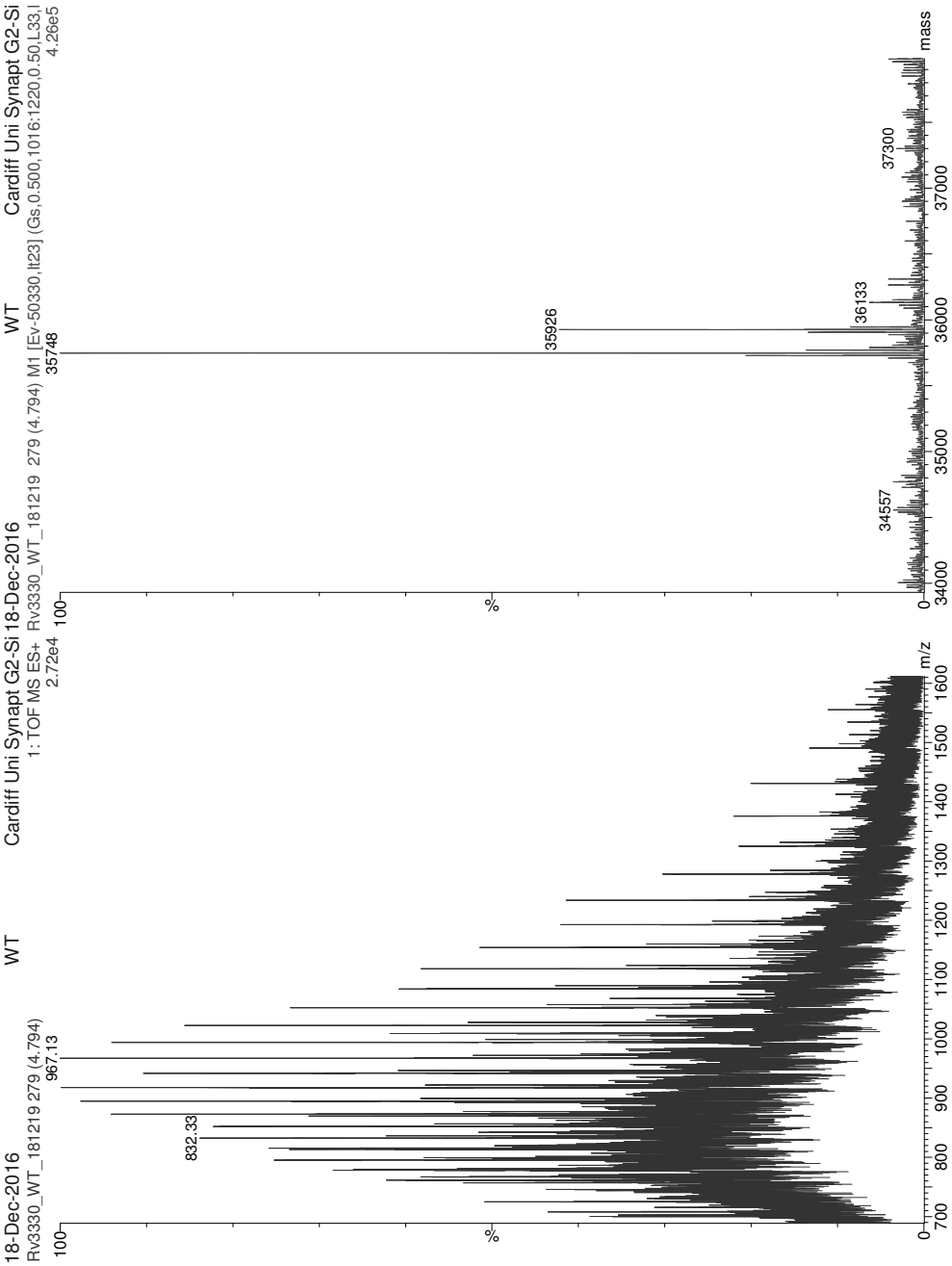
ATGAAAGGTTTTGCAATGCTCAGTATCGGTAAAGTTGGCTGGATTGAGAAGGAAAAGCCTGCTCCTGG
CCCATTTGATGCTATTGTAAGACCTCTAGCTGTGGCCCCTTGCACTTCGGACATTCATACCGTTTTTG
AAGGCGCCATTGGCGAAAGACATAACATGATACTCGGTCACGAAGCTGTAGGTGAAGTAGTTGAAGTA
GGTAGTGAGGTAAAAGATTTTAAACCTGGTGATCGCGTTGTTGTGCCAGCTATTACCCCTGATTGGCG
GACCTCTGAAGTACAAAGAGGATATCACCAGCACTCCGGTGGAATGCTGGCAGGCTGGAAATTTTCGA
ATGTAAAAGATGGTGTTTTTGGTGAATTTTTTCATGTGAATGATGCTGATATGAATTTAGCACATCTG
CCTAAAGAAATTCCATTGGAAGCTGCAGTTATGATTCCCGATATGATGACCACTGGTTTTCACGGAGC
TGAACTGGCAGATATAGAATTAGGTGCGACGGTAGCAGTTTTGGGTATTGGCCCAGTAGGTCTTATGG
CAGTCGCTGGTGCCAAATTGCGTGGAGCCGGAAGAATTATTGCCGTAGGCAGTAGACCAGTTTGTGTA
GATGCTGCAAAATACTATGGAGCTACTGATATTGTAAACTATAAAGATGGTCCTATCGAAAGTCAGAT
TATGAATCTAACTGAAGGCAAAGGTGTCGATGCTGCCATCATCGCTGGAGGAAATGCTGACATTATGG
CTACAGCAGTTAAGATTGTTAAACCTGGTGGCACCATCGCTAATGTAAATTATTTTGGCGAAGGAGAG
GTTTTGCCTGTTCCTCGTCTTGAATGGGGTTGCGGCATGGCTCATAAAACTATAAAAGGCGGGCTATG
CCCCGGTGGACGTCTAAGAATGGAAAGACTGATTGACCTTGTTTTTTATAAGCGTGTCGATCCTTCTA
AGCTCGTCACTCACGTTTTCCGGGGATTTGACAATATTGAAAAAGCCTTTATGTTGATGAAAGACAAA
CCAAAAGACCTAATCAAACCTGTTGTAATATTAGCATAA

Amino Acid Sequence:

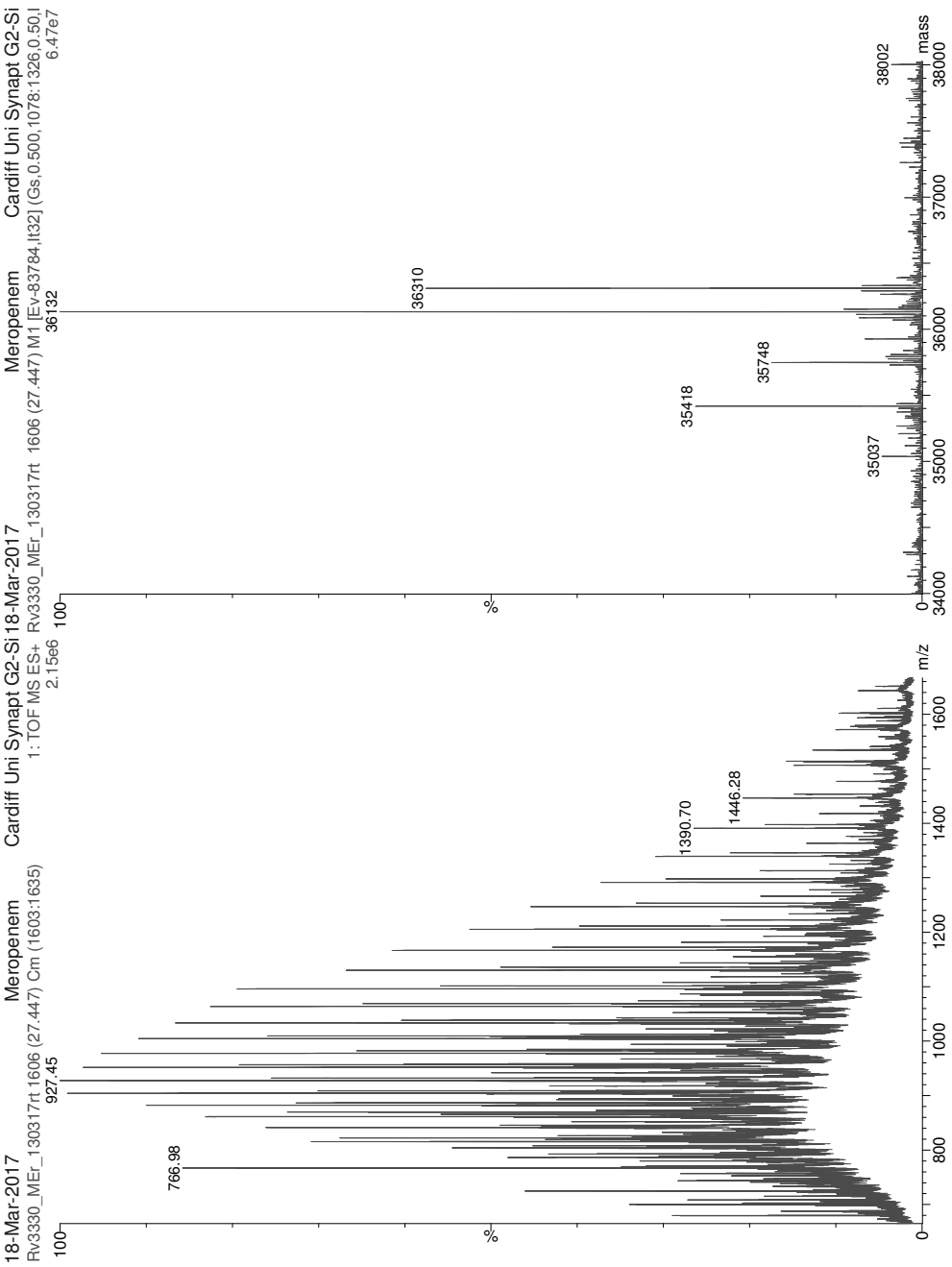
MKGFAMLSIGKVGWIEKEKPAPGPFDAIVRPLAVAPCTSDIHTVFEGAIGERHNMILGHE
AVGEVVEVGSEVKDFKPGDRVVVPAITPDWRTSEVQRGYHQHSGGMLAGWKFSNVKDGVF
GEFFHVNDADMNLAHLPKEIPLEAAVMIPDMMTTGFHGAELADIELGATVAVLGIGPVGL
MAVAGAKLRGAGRIIAVGSRPVCVDAAKYYGATDIVNYKDGPIESQIMNLTEGKGVDAAI
IAGGNADIMATAVKIVKPGGTIANVNYFGEGEVLPVPRLEWGCGMAHKTIKGGLCPGGRL
RMERLIDLVFYKRVDPSKLVTHVFRGFDNIEKAFMLMKDKPKDLIKPVVILA

7.2 Protein Mass Spectra

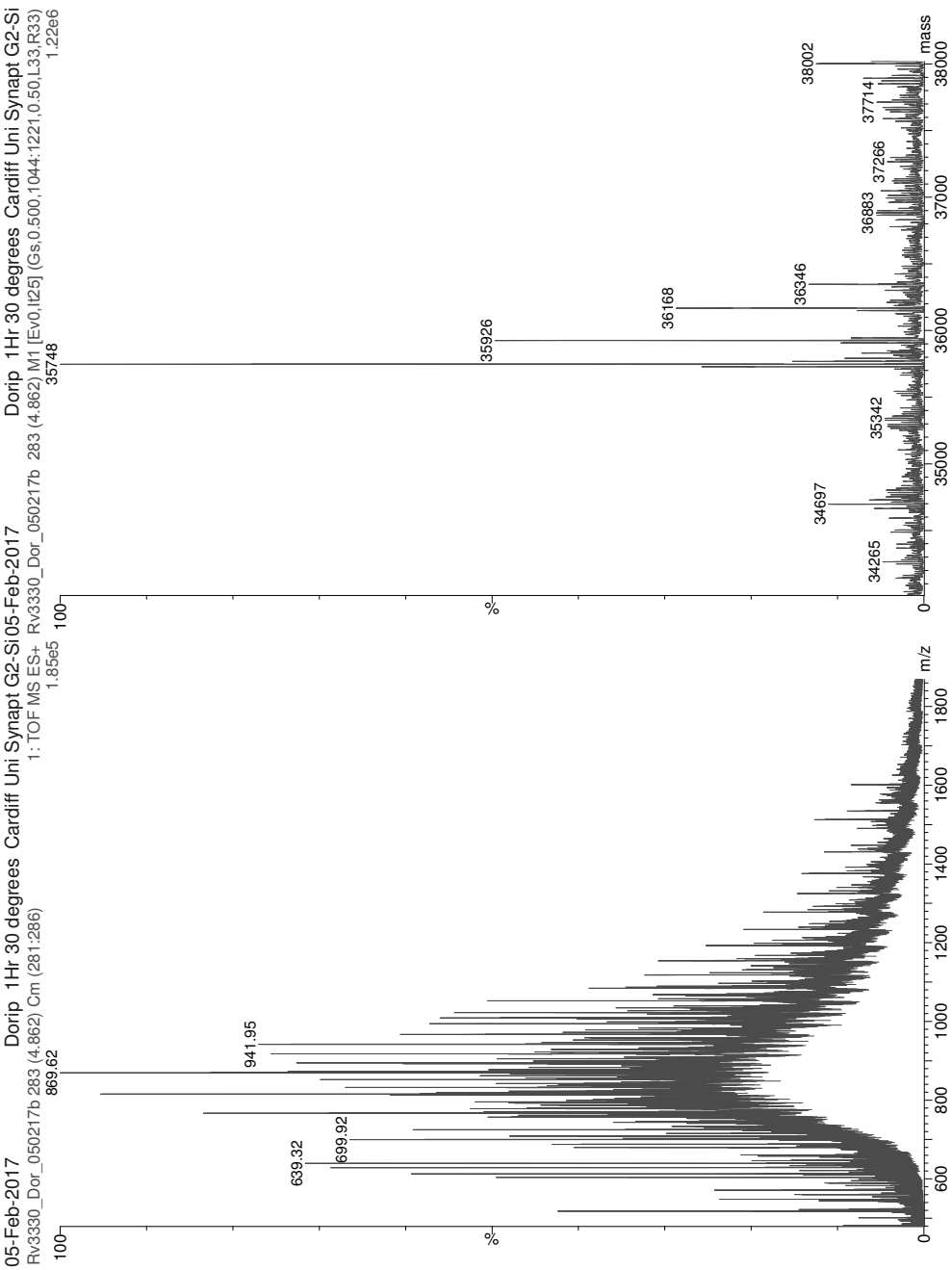
Wild-type Rv3330



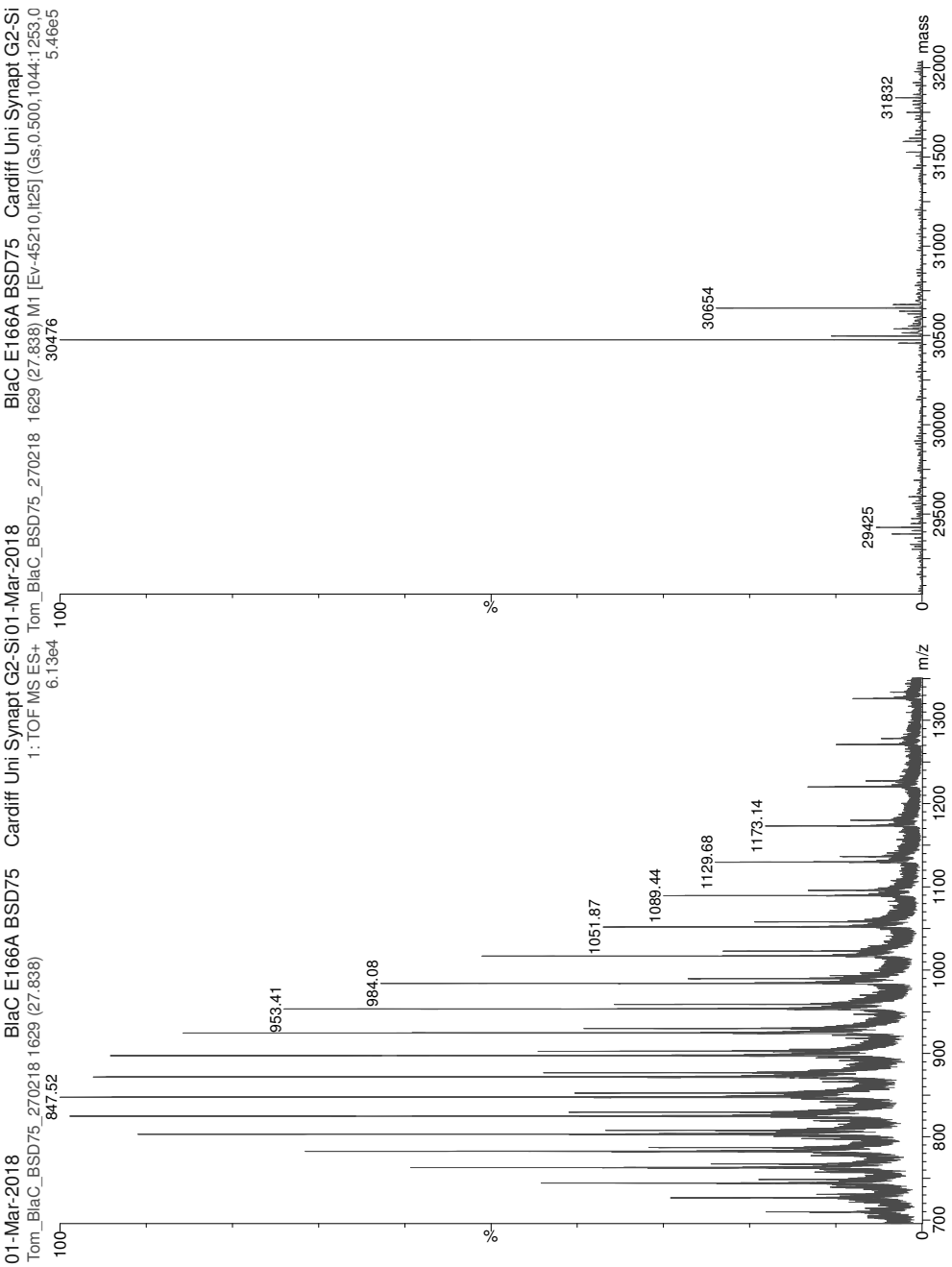
Rv3330-Meropenem



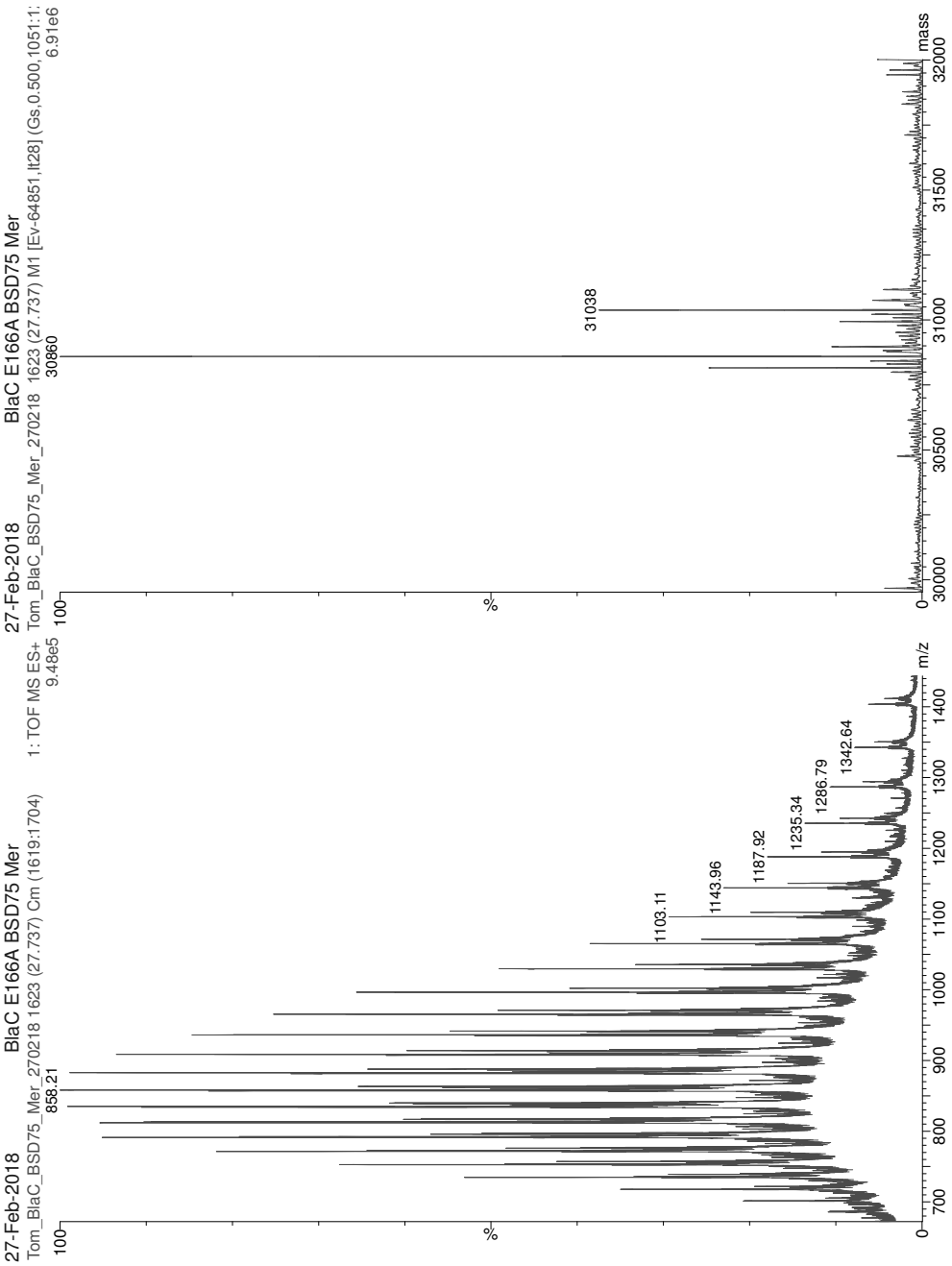
Rv3330-Doripenem



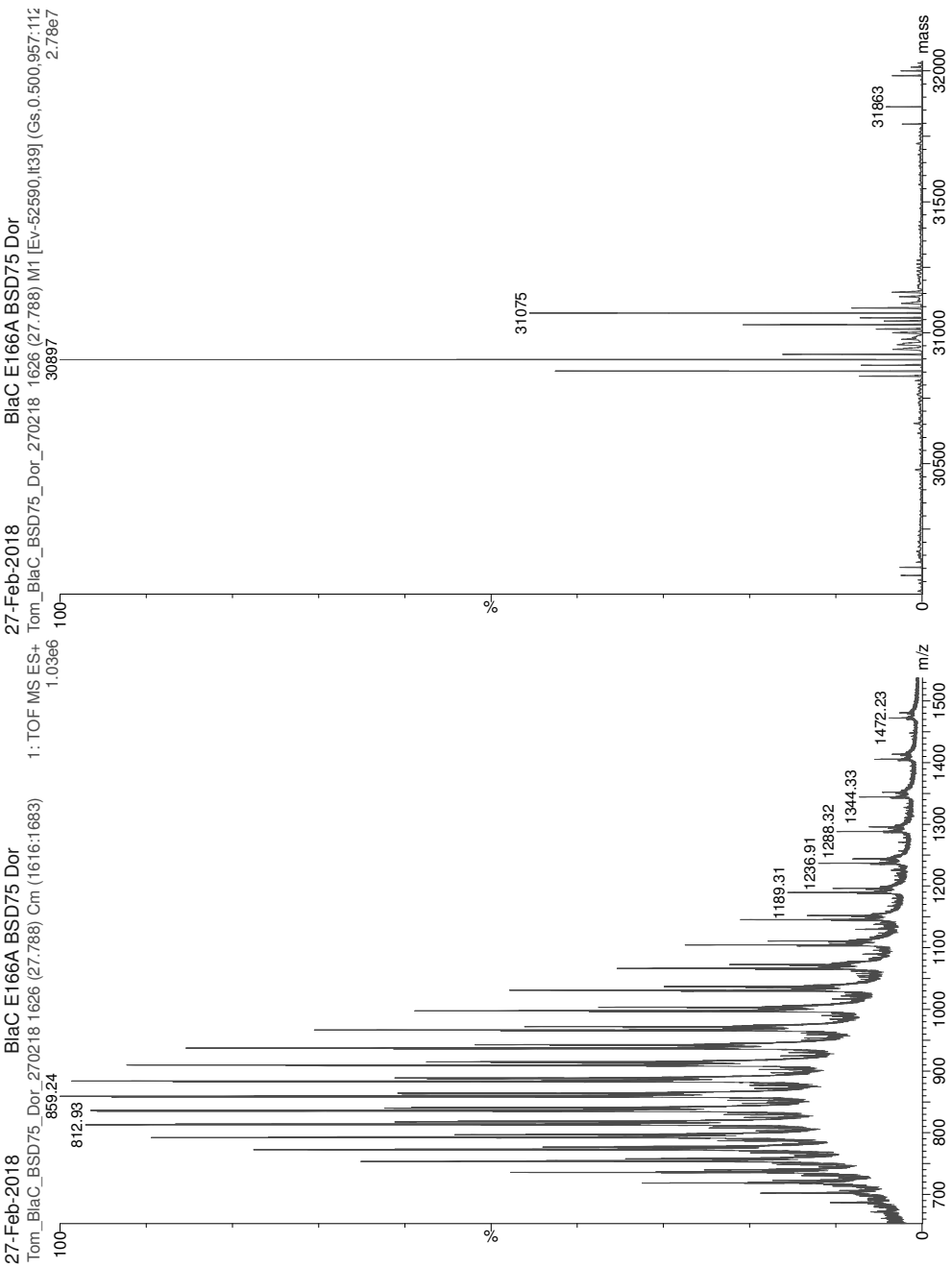
BlaC (E166A)


01-Mar-2018
Tom_BlaC_BSD75_270218 1629 (27.838)
BlaC E166A BSD75
Cardiff Uni Synapt G2-Si01-Mar-2018
1: TOF MS ES+ Tom_BlaC_BSD75_270218 1629 (27.838)
6.13e4
BlaC E166A BSD75
Cardiff Uni Synapt G2-Si
1629 (27.838) M1 [Ev-45210,It25] (Gs,0.500,1044:1253.0
5.4665
847.52
953.41
984.08
1051.87
1089.44
1129.68
1173.14
30476
30654
29425
31832
m/z
mass
%

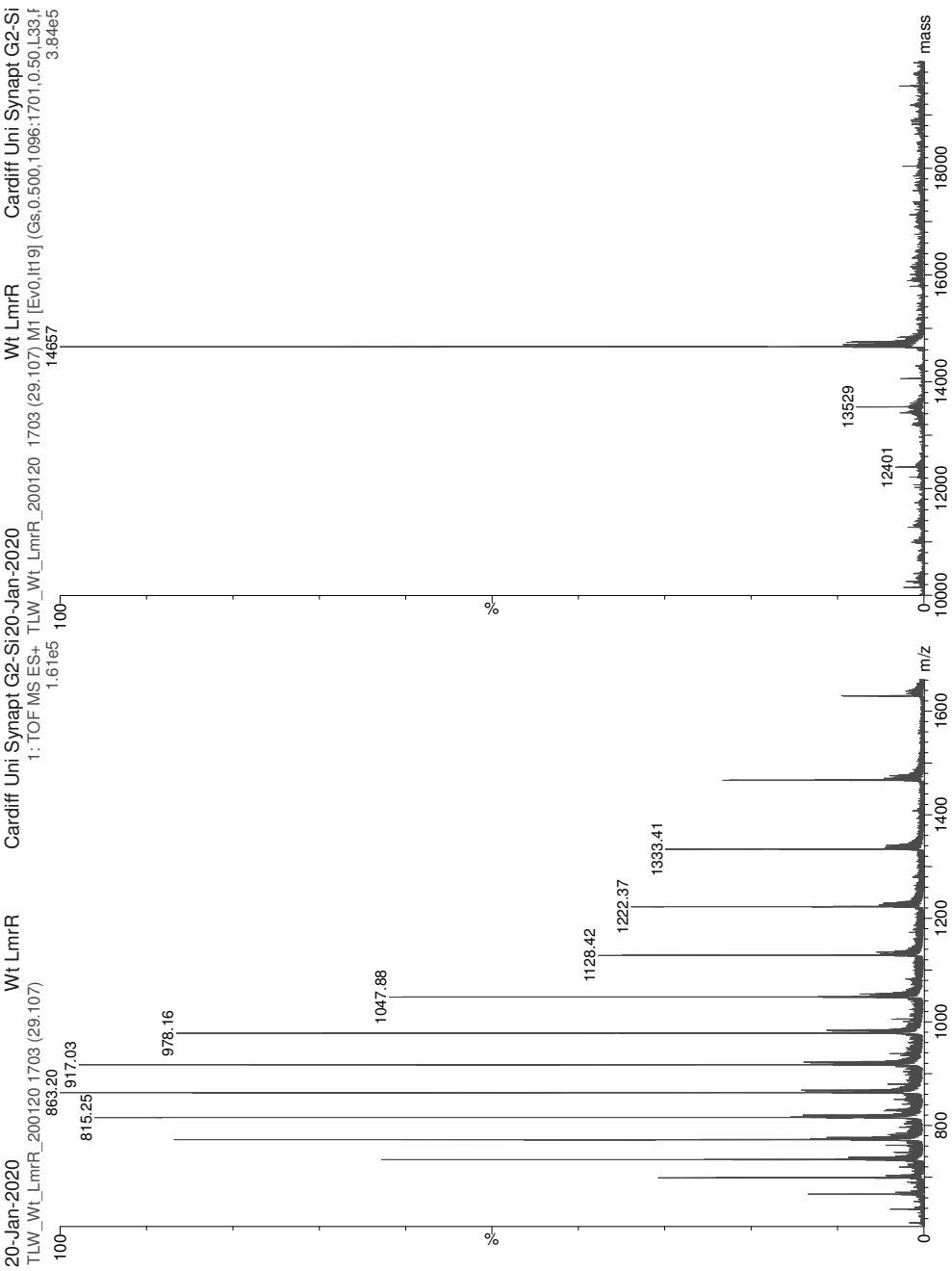
BlaC (E166A) - Meropenem



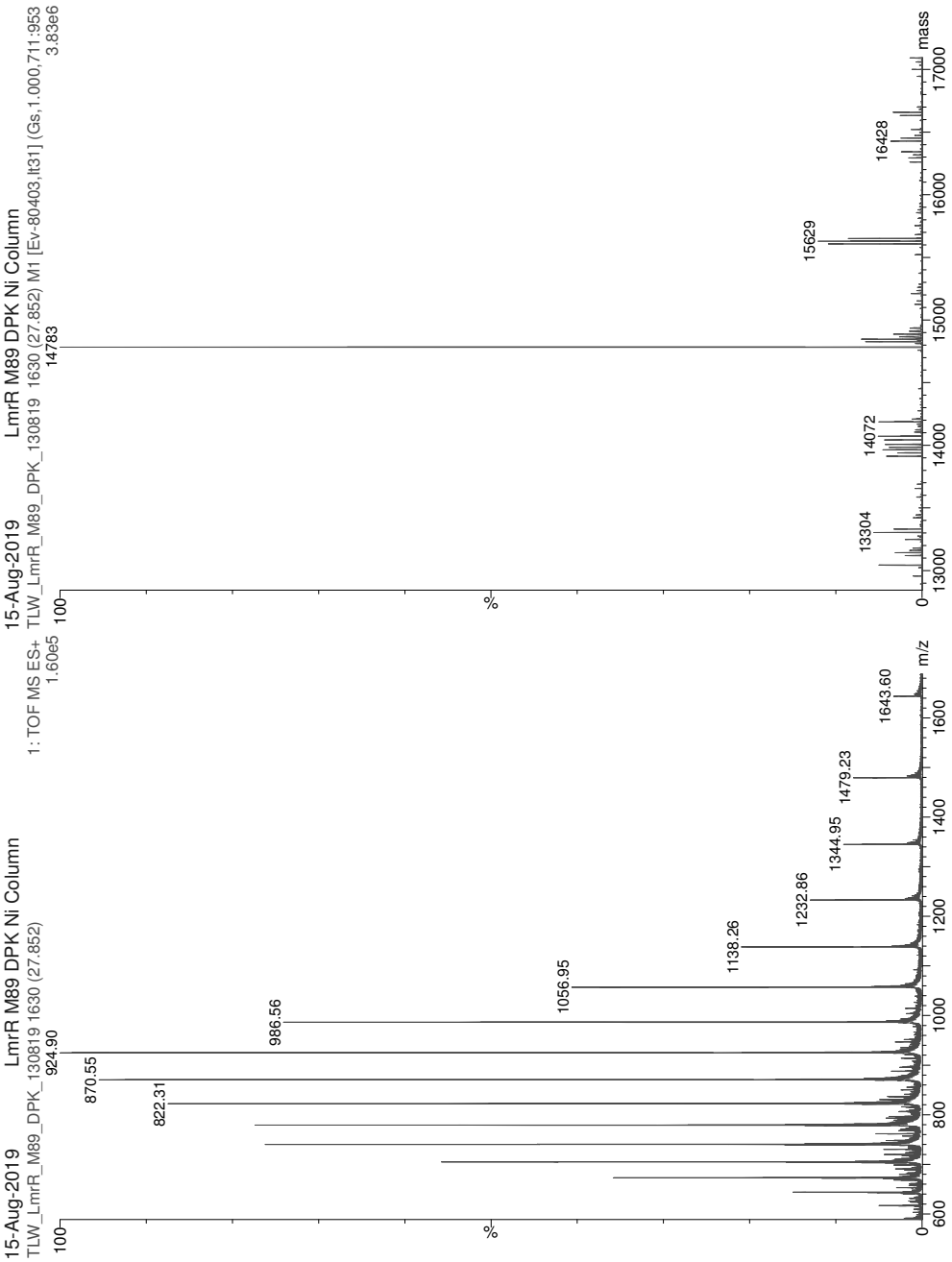
BlaC (E166A) – Doripenem



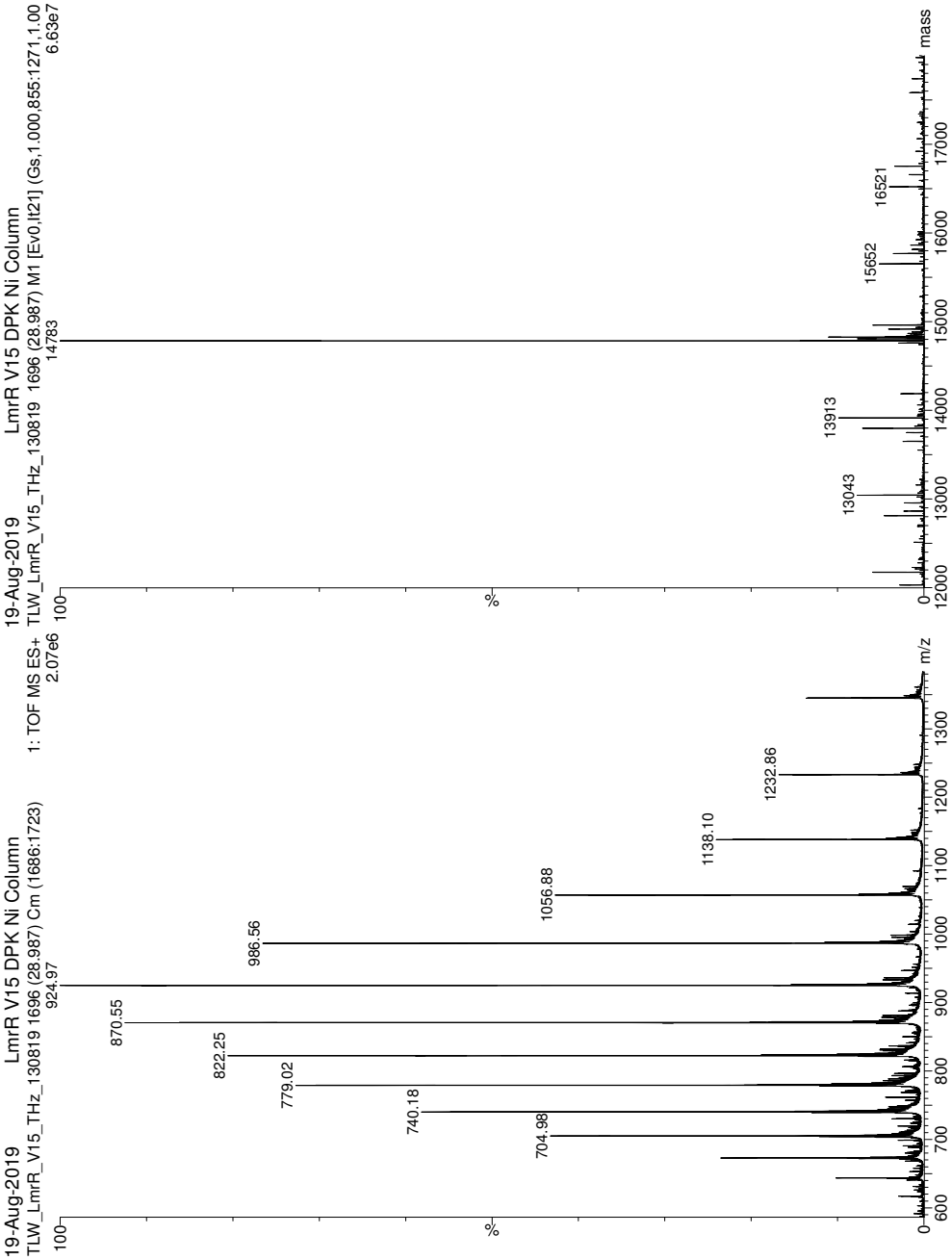
Wild-type LmrR



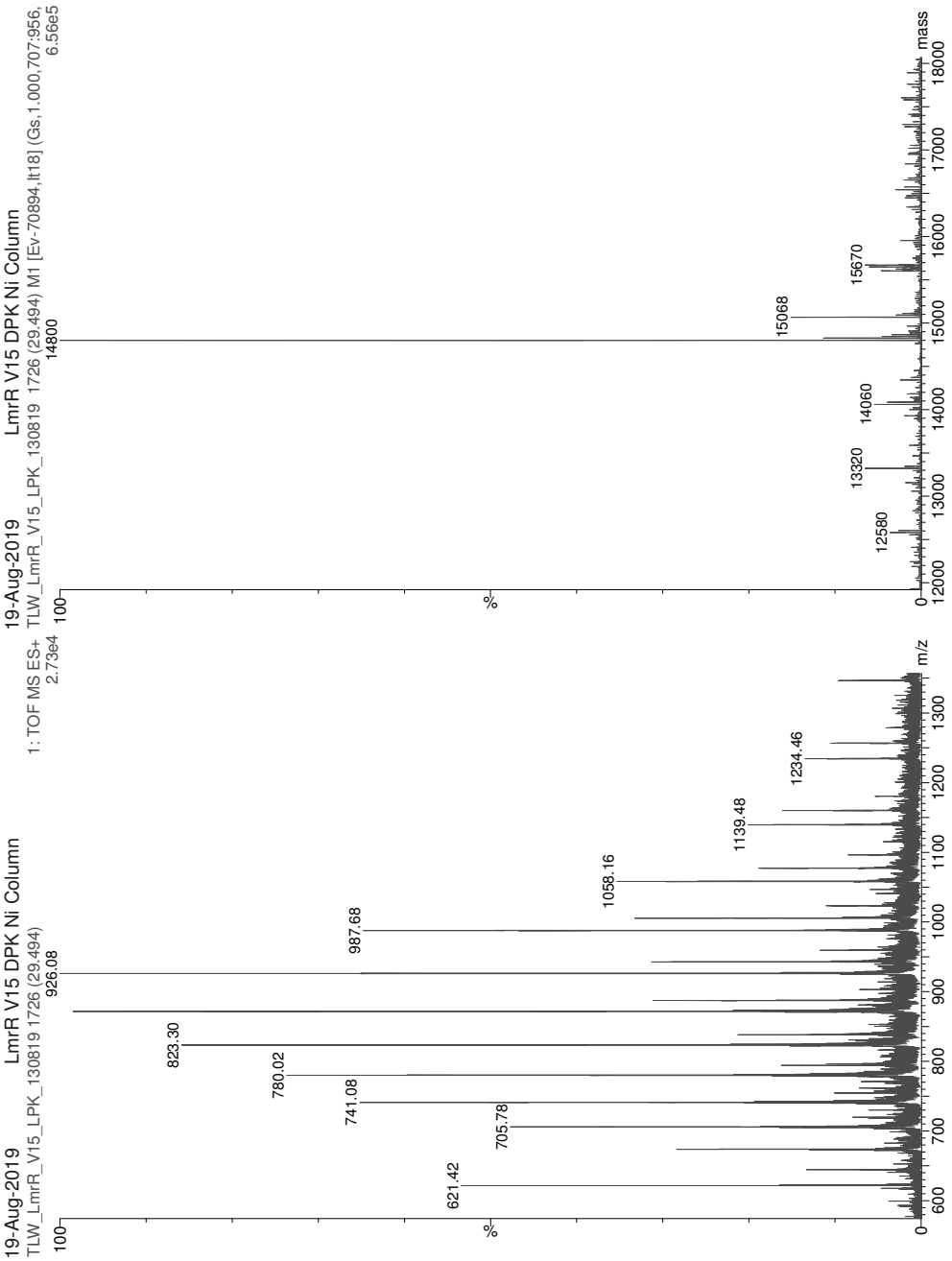
LmrR (V15DPK)


15-Aug-2019
TLW_LmrR_M89_DPK_130819 1630 (27.852)
LmrR M89 DPK Ni Column
1: TOF MS ES+
1.60e5
100
%
924.90
870.55
822.31
986.56
1056.95
1138.26
1232.86
1344.95
1479.23
1643.60
m/z
15-Aug-2019
TLW_LmrR_M89_DPK_130819 1630 (27.852) M1 [Ev-80403,It31] (Gs,1.000,711:953)
LmrR M89 DPK Ni Column
3.83e6
14783
15629
16428
13304
14072
mass

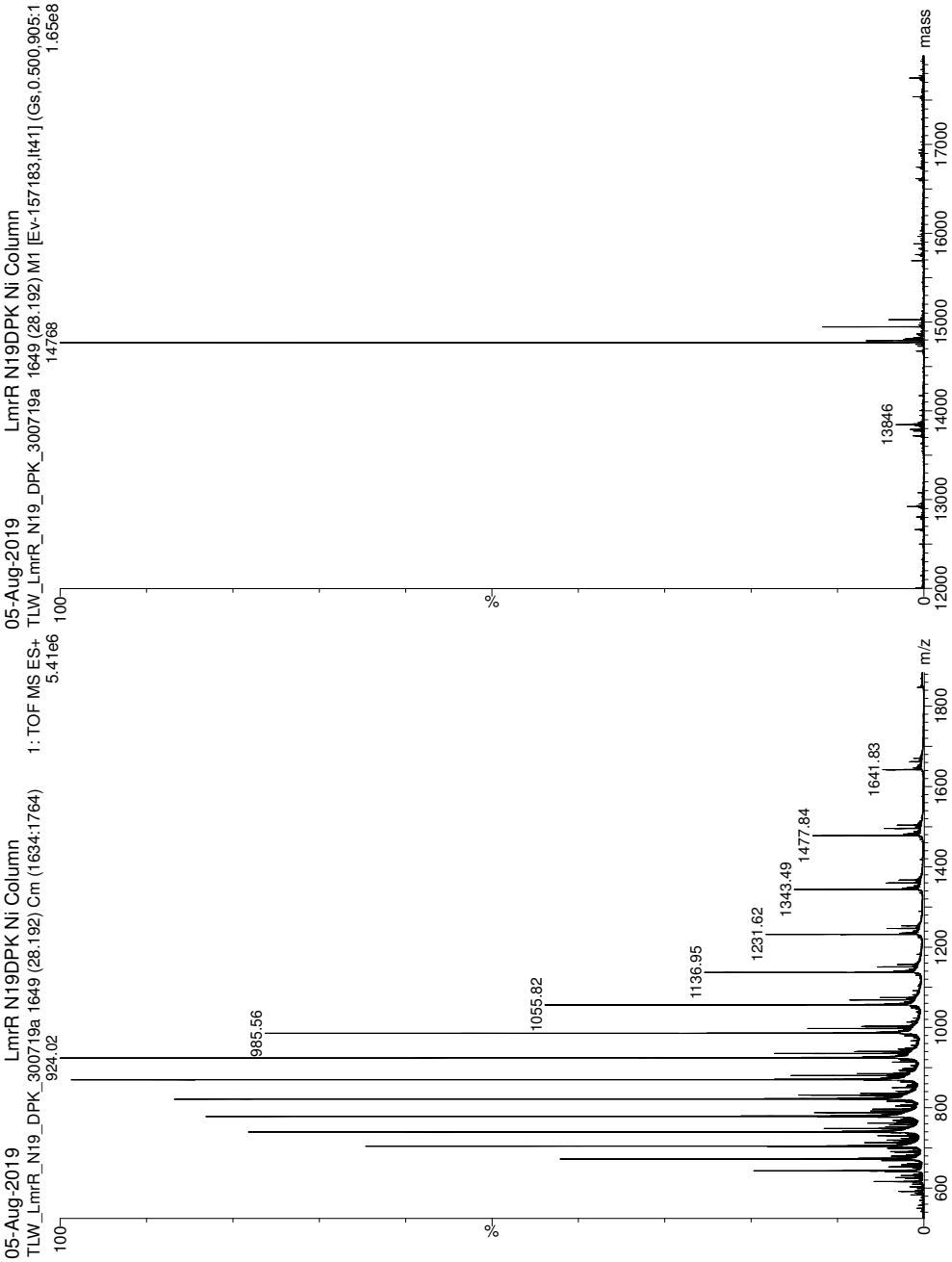
LmrR (V15LPK)


19-Aug-2019
LmrR V15 DPK Ni Column
TLW_LmrR_V15_THz_130819 1696 (28.987) Cm (1686:1723)
1: TOF MS ES+
2.07e6
100
%
0
600
700
800
900
924.97
870.55
822.25
779.02
740.18
704.98
986.56
1056.88
1138.10
1232.86
1000
1100
1200
1300
m/z
19-Aug-2019
LmrR V15 DPK Ni Column
TLW_LmrR_V15_THz_130819 1696 (28.987) M1 [Ev0,It21] (Gs,1.000,855:1271,1.00
6.63e7
14783
13913
13043
15652
16521
12000
13000
14000
15000
16000
17000
mass

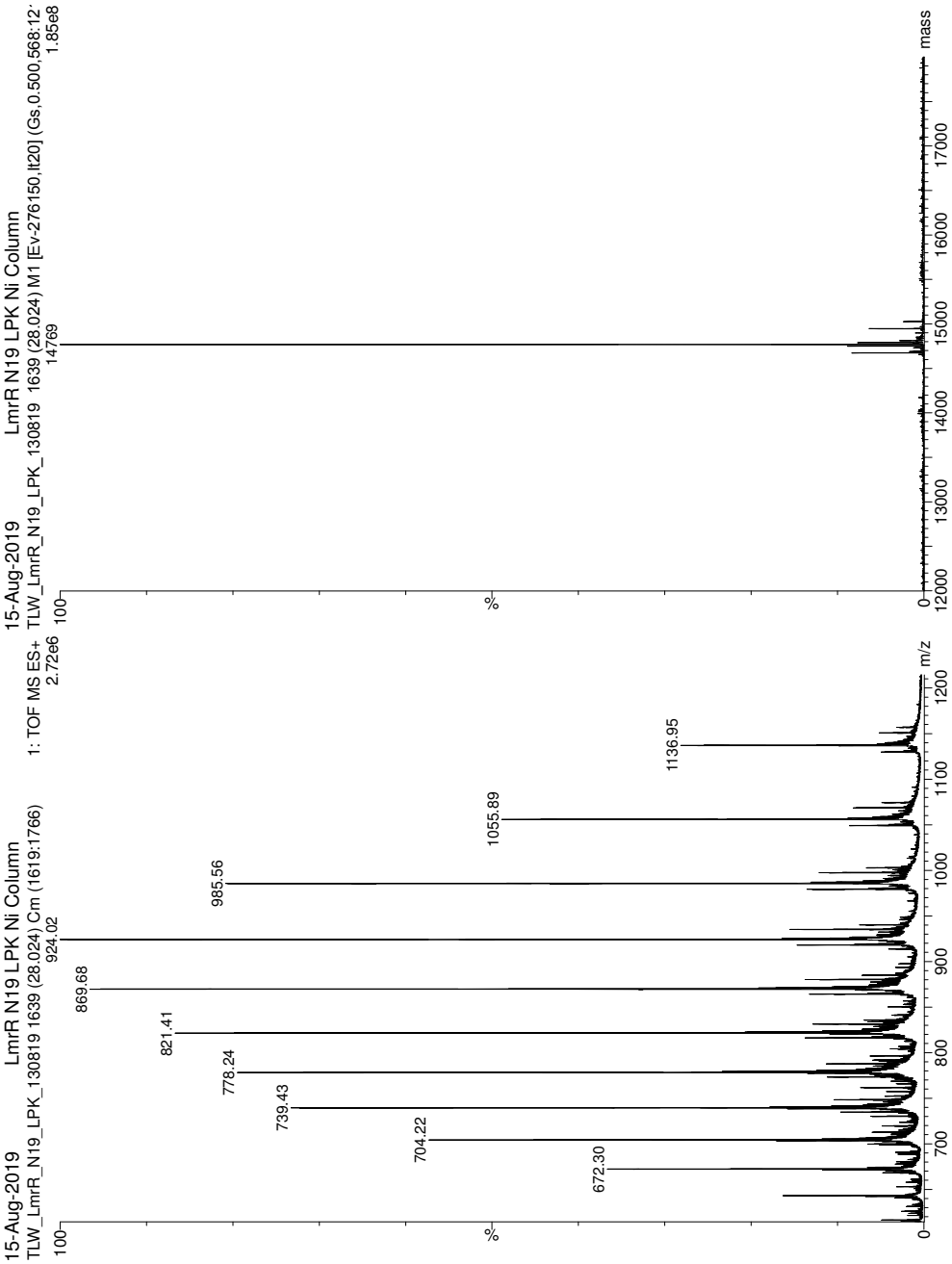
LmrR (V15ThzK)



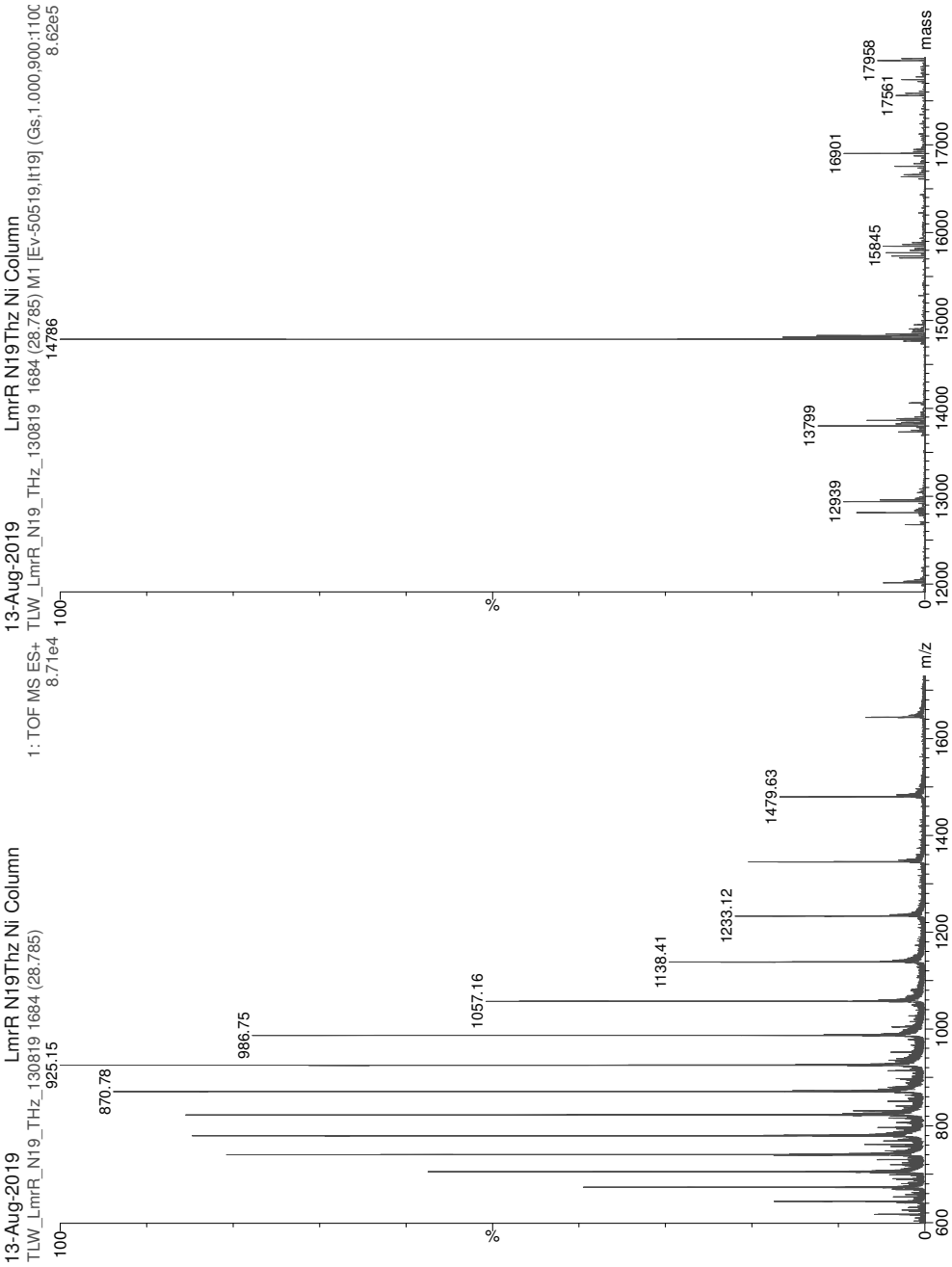
LmrR (N19DPK)


05-Aug-2019
TLW_LmrR_N19_DPK_300719a
LmrR N19DPK Ni Column
1649 (28.192) Cm (1634:1764)
1: TOF MS ES+
5.41e6
924.02
985.56
1055.82
1136.95
1231.62
1343.49
1477.84
1641.83
m/z
05-Aug-2019
TLW_LmrR_N19_DPK_300719a
LmrR N19DPK Ni Column
1649 (28.192) M1 [Ev-157183,It41] (Gs,0.500,905:1
1.65e8
14768
13846
mass

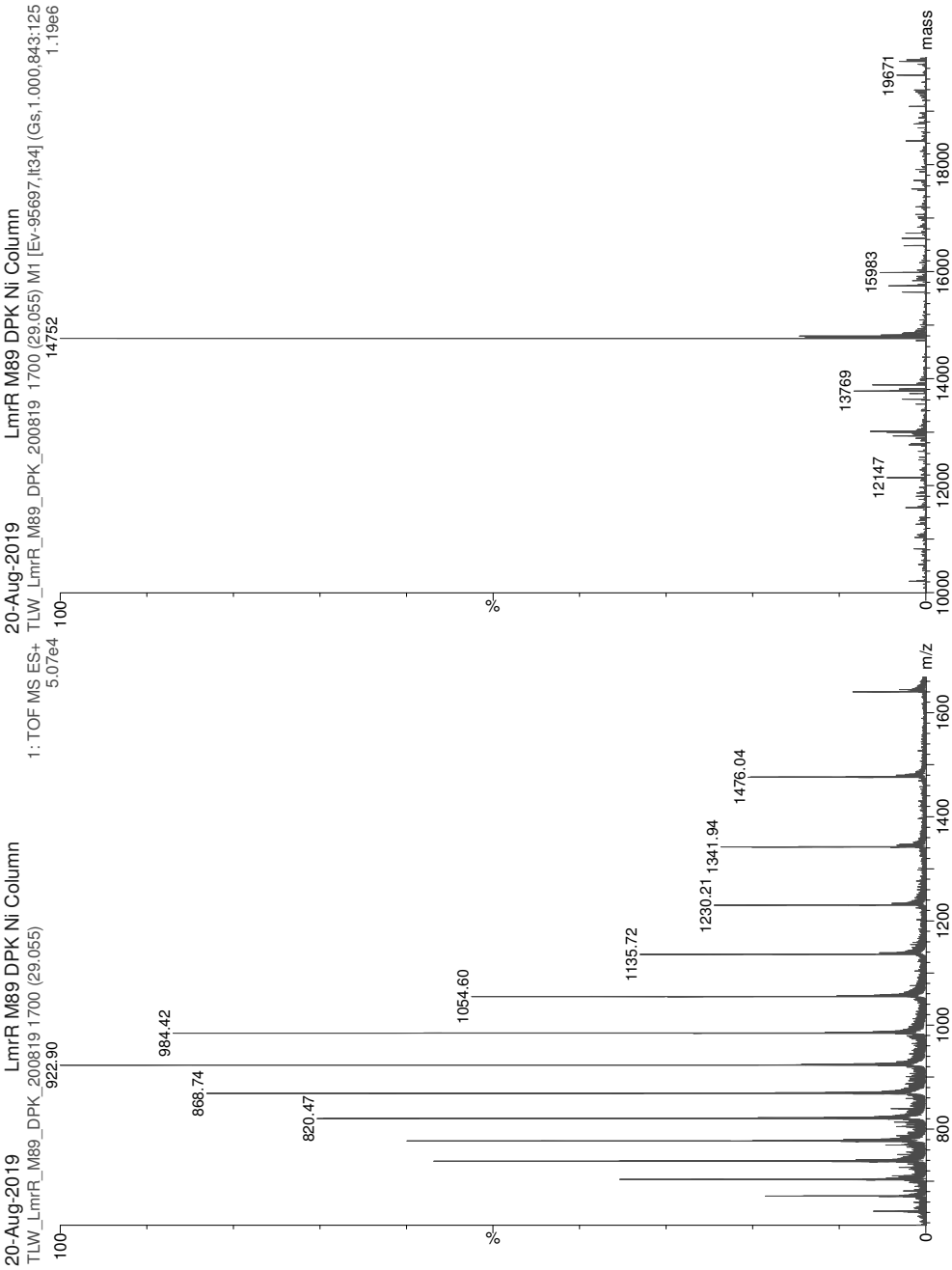
LmrR (N19LPK)


15-Aug-2019
TLW_LmrR_N19_LPK_130819 1639 (28.024) Cm (1619:1766)
LmrR N19 LPK Ni Column
1: TOF MS ES+
2.72e6
100
%
0
924.02
869.68
821.41
778.24
739.43
704.22
672.30
985.56
1055.89
1136.95
700
800
900
1000
1100
1200
m/z
15-Aug-2019
TLW_LmrR_N19_LPK_130819 1639 (28.024) M1 [Ev-276150,It20] (Gs,0.500,568:12
LmrR N19 LPK Ni Column
1.85e8
14769
12000
13000
14000
15000
16000
17000
mass

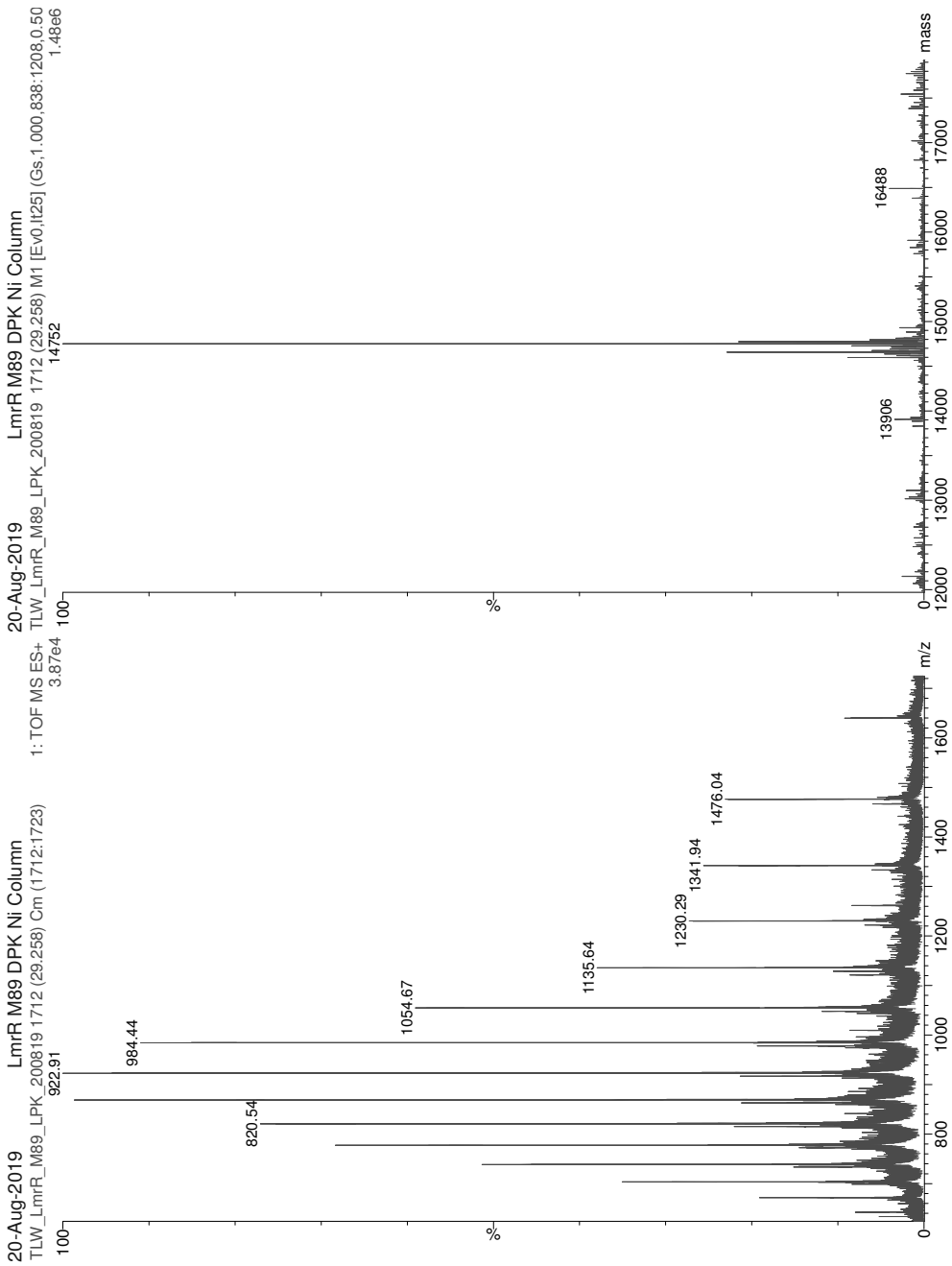
LmrR (N19ThzK)



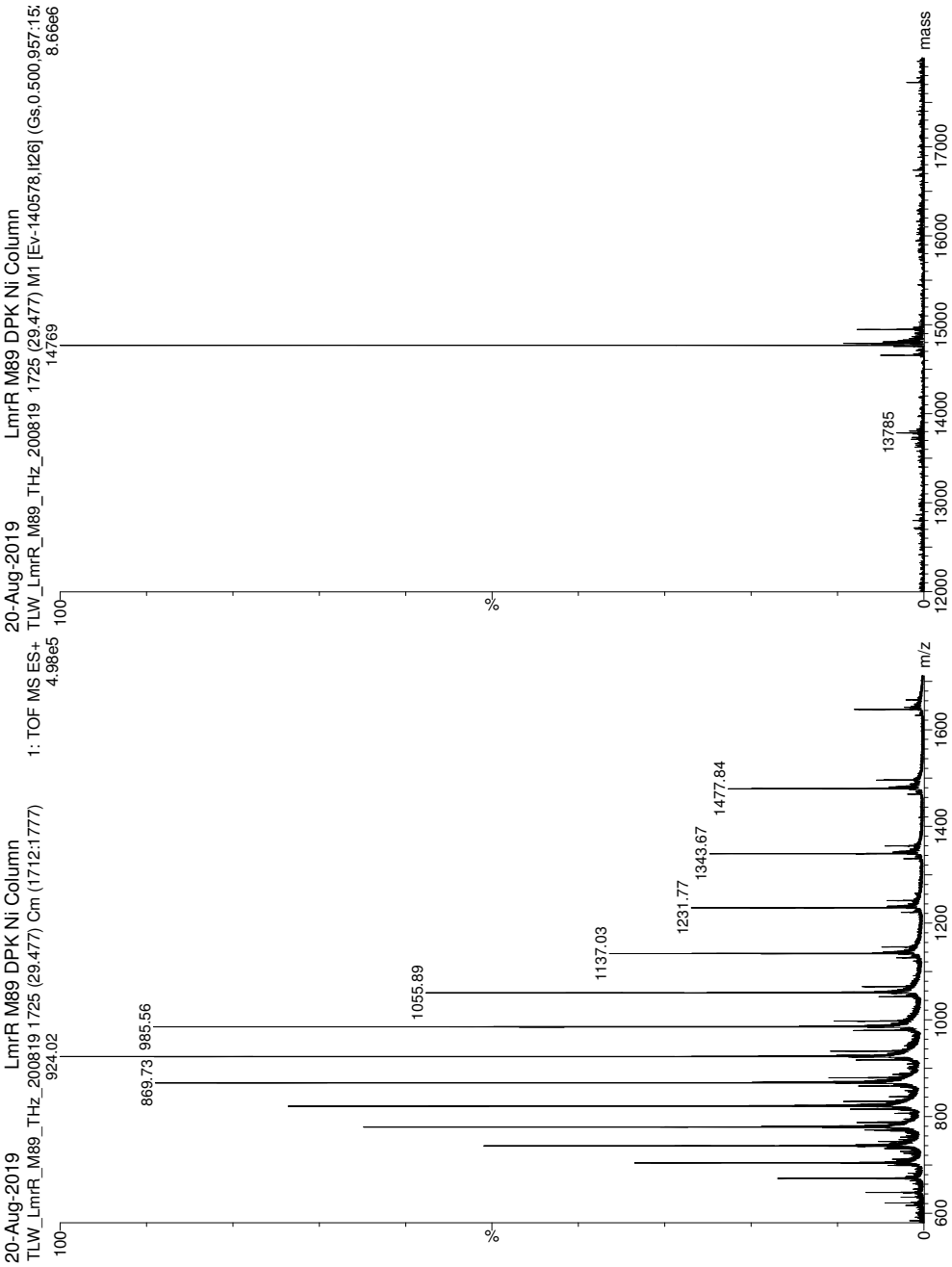
LmrR (M89DPK)


20-Aug-2019
TLW_LmrR_M89_DPK_200819 1700 (29.055)
LmrR M89 DPK Ni Column
1: TOF MS ES+
5.07e4
922.90
984.42
868.74
820.47
1054.60
1135.72
1230.21
1341.94
1476.04
m/z
20-Aug-2019
TLW_LmrR_M89_DPK_200819 1700 (29.055) M1 [Ev-95697,It34] (Gs,1.000,843:125
LmrR M89 DPK Ni Column
1.19e6
14752
15983
13769
12147
19671
mass

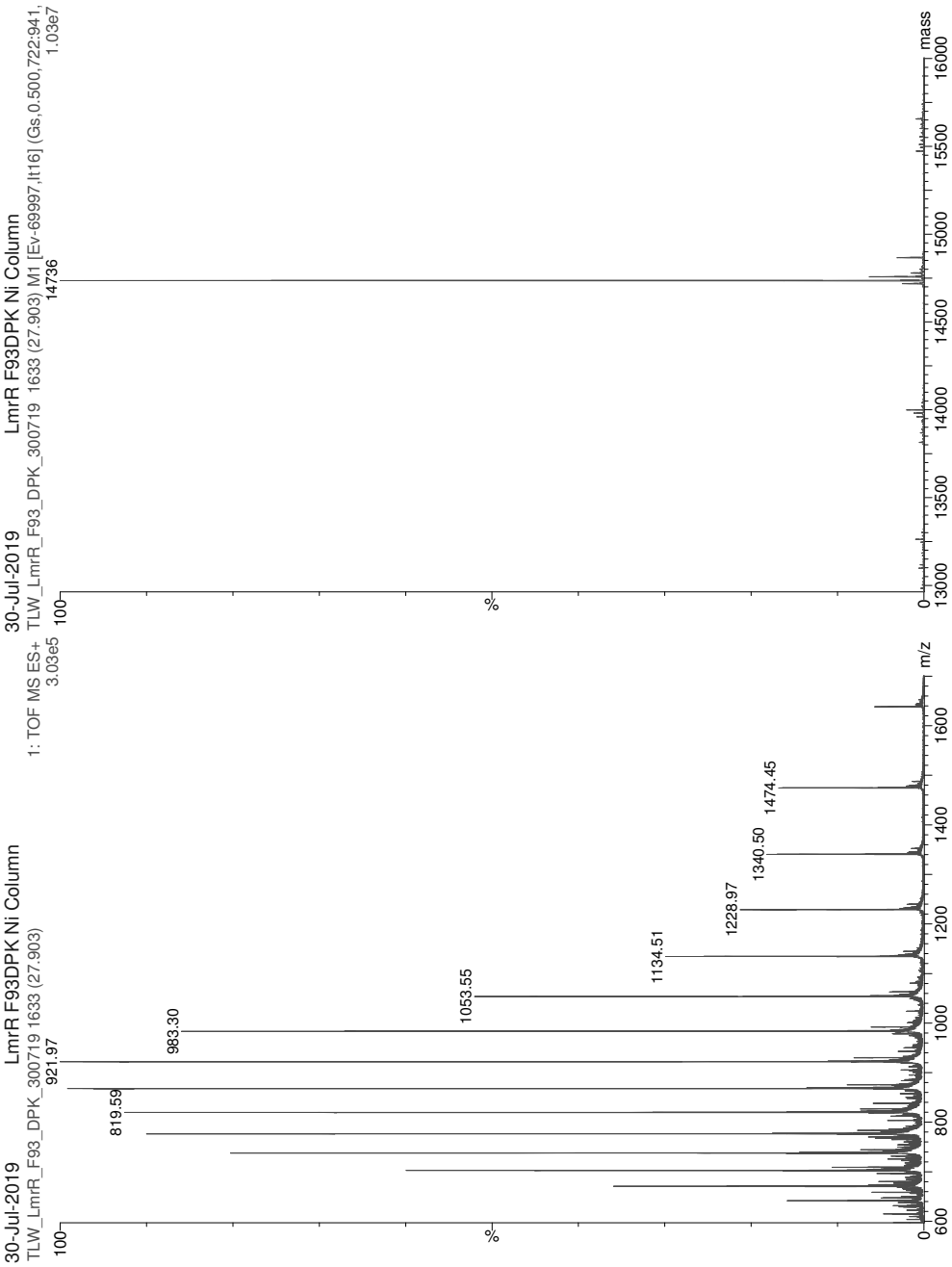
LmrR (M89LPK)


20-Aug-2019
TLW_LmrR_M89_LPK_200819 1712 (29.258) Cm (1712:1723)
1: TOF MS ES+
3.87e4
LmrR M89 DPK Ni Column
100
%
922.91
984.44
820.54
1054.67
1135.64
1230.29
1341.94
1476.04
0
800
1000
1200
1400
1600
m/z
20-Aug-2019
TLW_LmrR_M89_LPK_200819 1712 (29.258) M1 [Ev0,It25] (Gs,1.000,838:1208,0.50)
1: TOF MS ES+
1.48e6
LmrR M89 DPK Ni Column
100
%
14752
13906
16488
0
12000
13000
14000
15000
16000
17000
mass

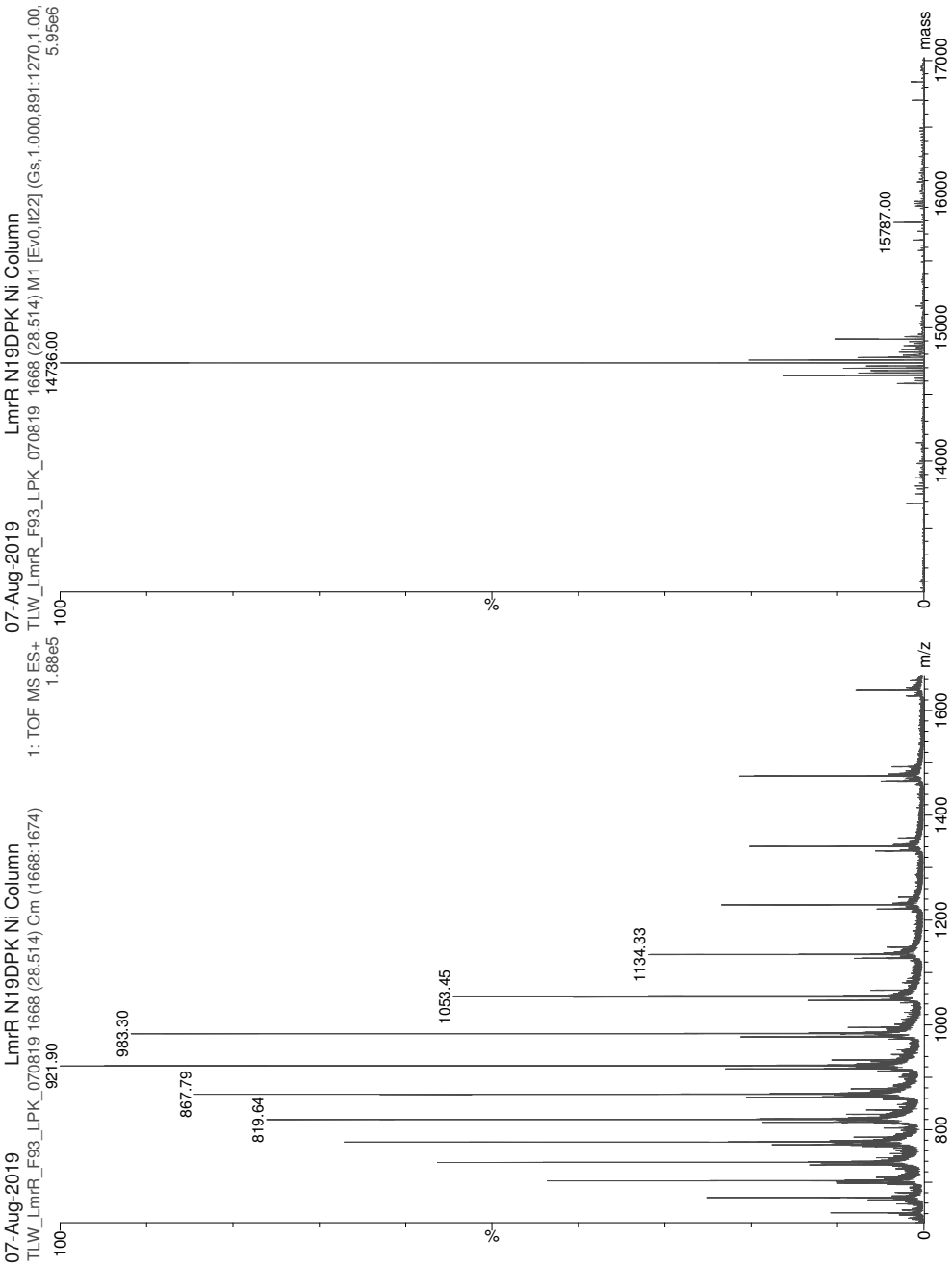
LmrR (M89ThzK)



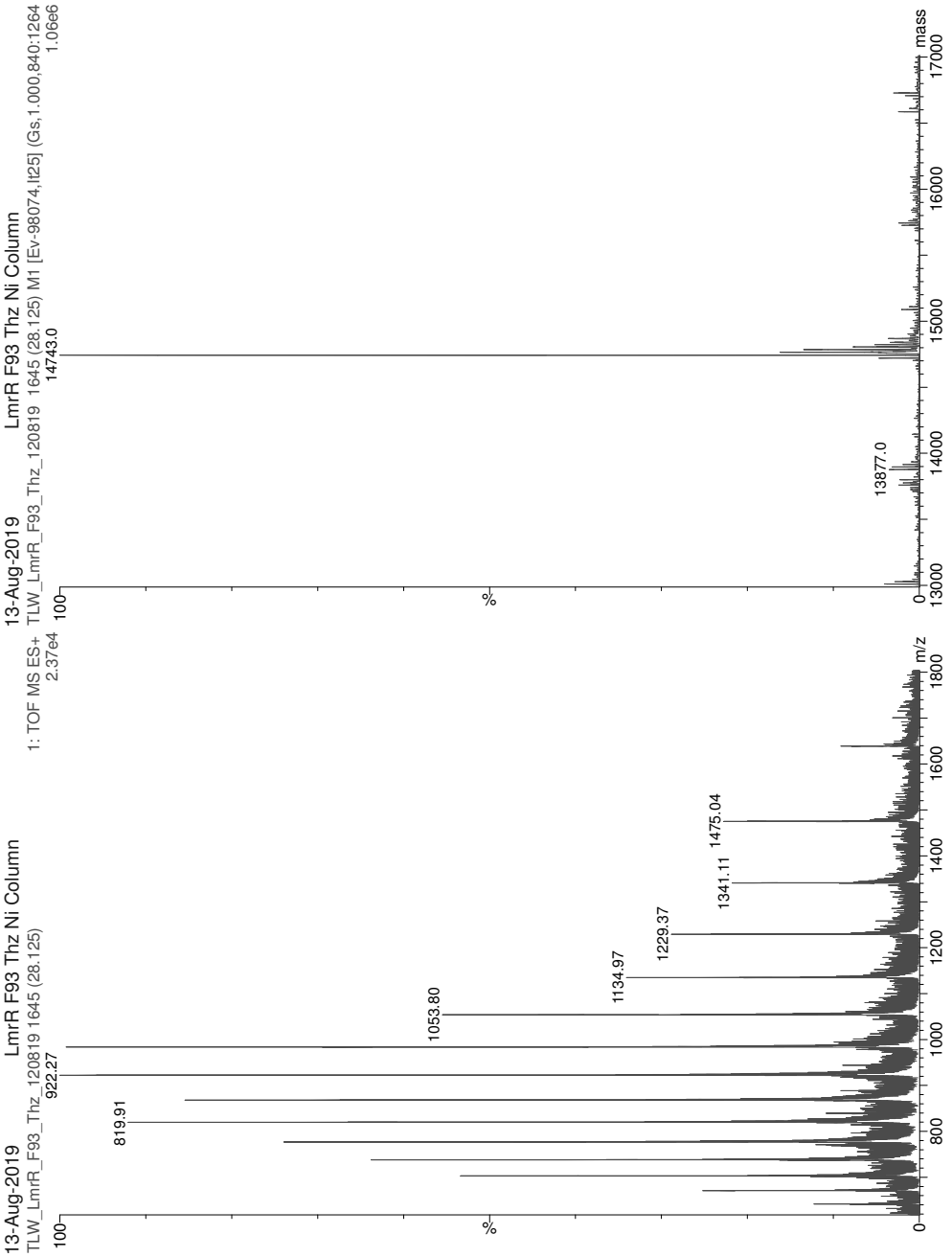
LmrR (F93DPK)


30-Jul-2019
TLW_LmrR_F93_DPK_300719 1633 (27.903)
LmrR F93DPK Ni Column
1: TOF MS ES+
3.03e5
921.97
819.59
983.30
1053.55
1134.51
1228.97
1340.50
1474.45
m/z
30-Jul-2019
TLW_LmrR_F93_DPK_300719 1633 (27.903) M1 [Ev-69997,It16] (Gs,0.500,722:941,
LmrR F93DPK Ni Column
1.03e7
14736
mass

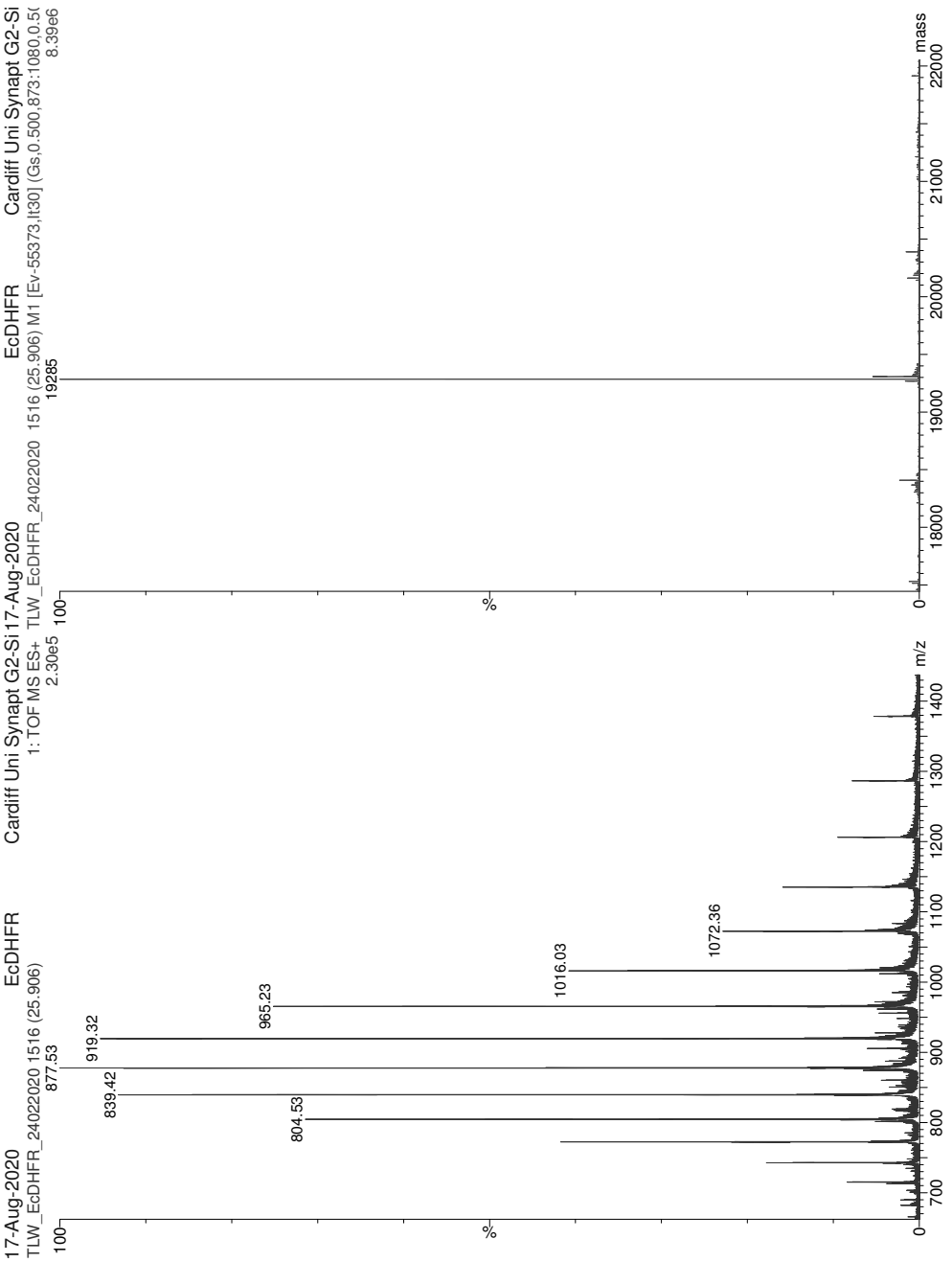
LmrR (F93LPK)


07-Aug-2019
TLW_LmrR_F93_LPK_070819 1668 (28.514) Cm (1668:1674)
LmrR N19DPK Ni Column
921.90
983.30
867.79
819.64
1053.45
1134.33
m/z
07-Aug-2019
1: TOF MS ES+ 1.88e5
TLW_LmrR_F93_LPK_070819 1668 (28.514) M1 [Ev0,It22] (Gs,1.000,891:1270,1.00, 5.95e6
LmrR N19DPK Ni Column
14736.00
15787.00
mass

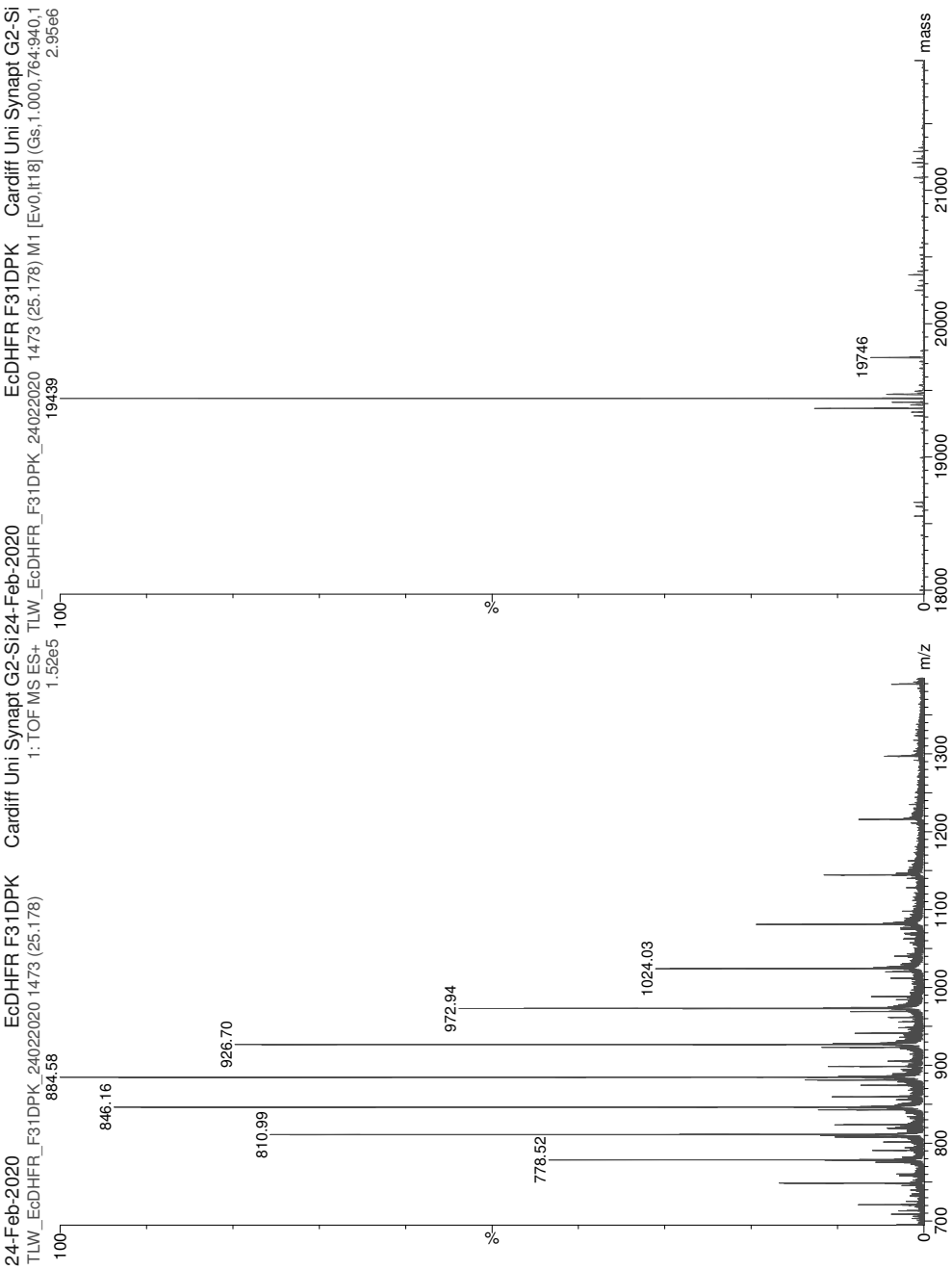
LmrR (F93ThzK)



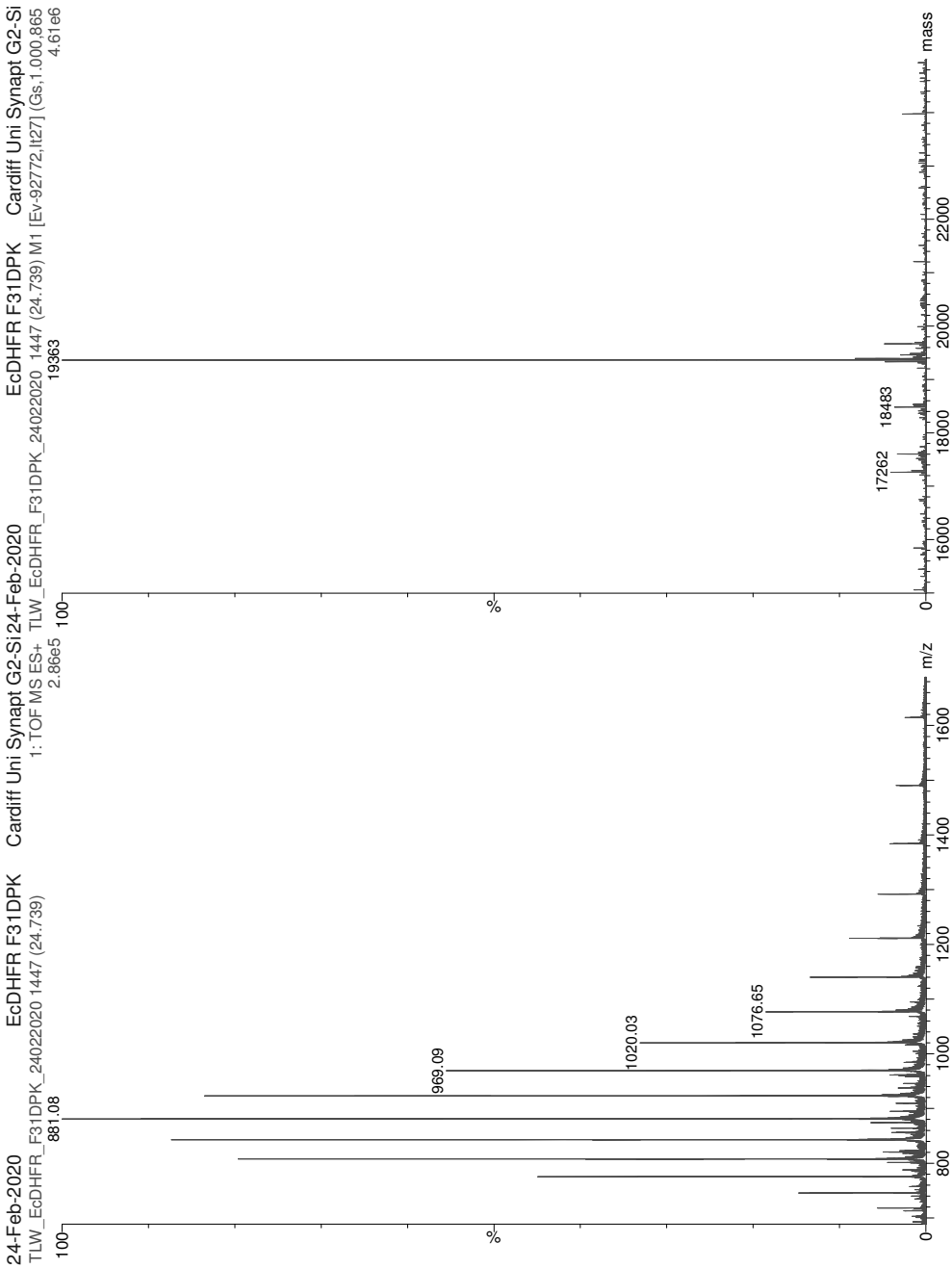
Wild-type EcDHFR



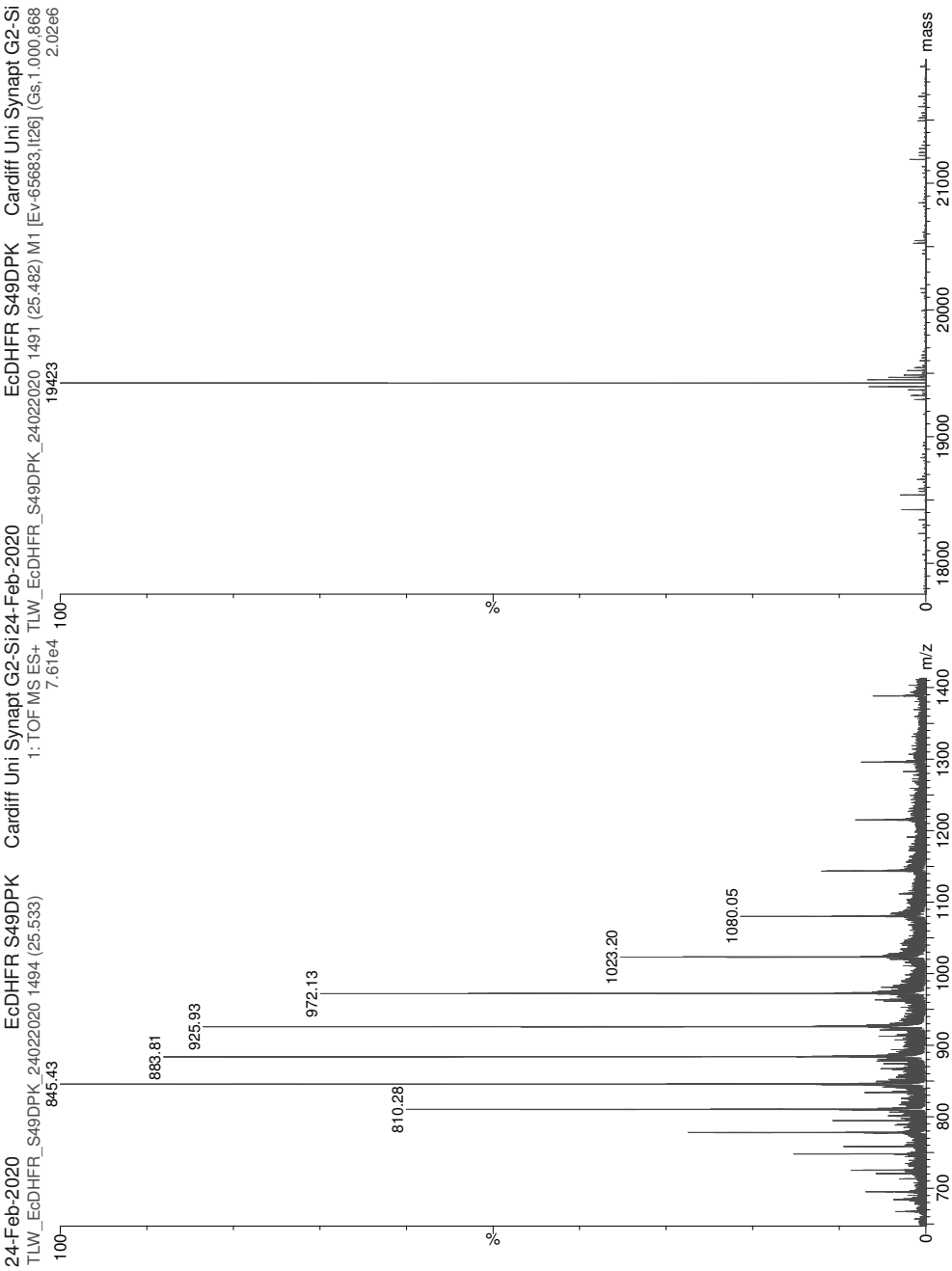
EcDHFR (A7DPK)



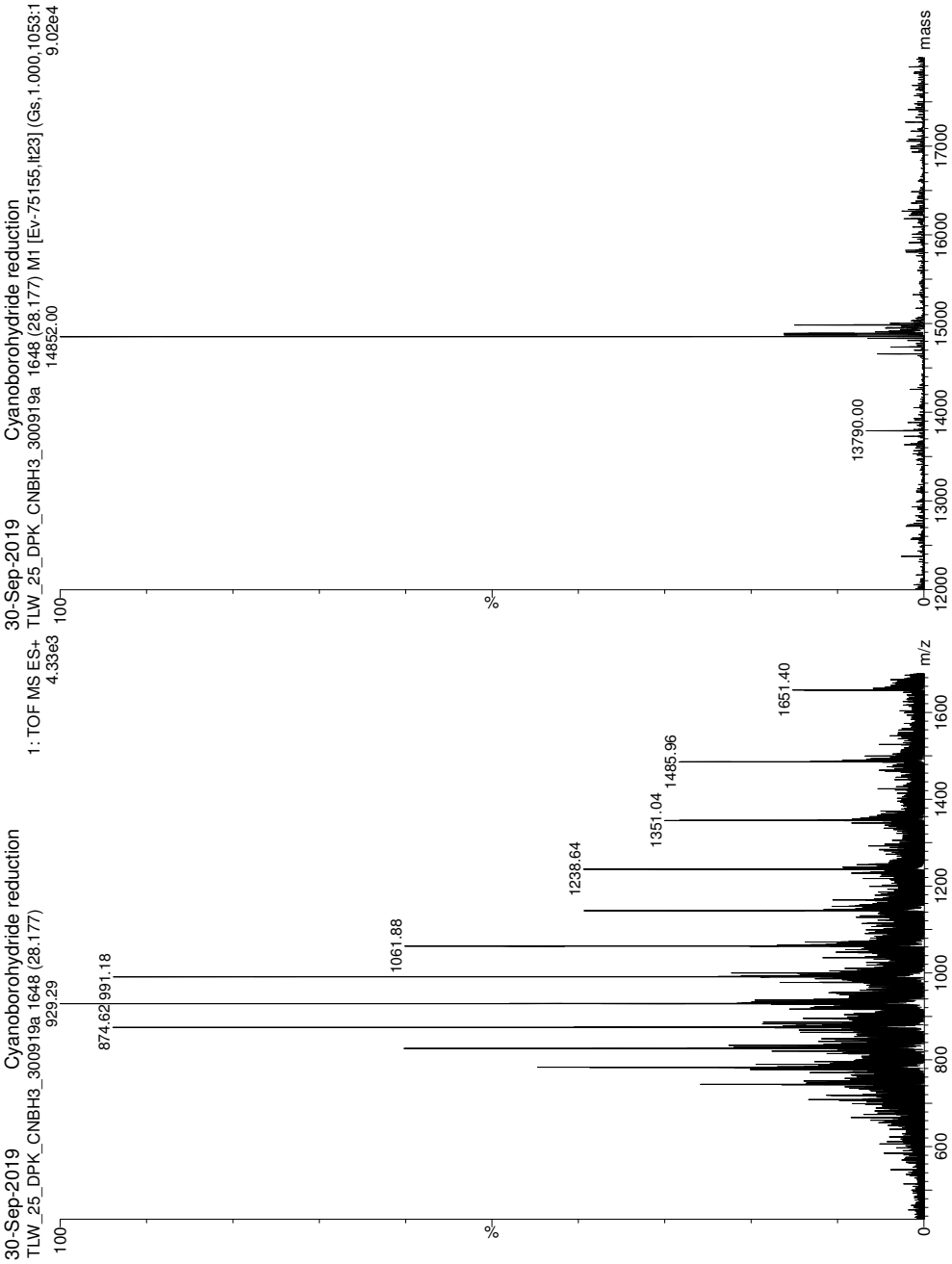
EcDHFR (F31DPK)



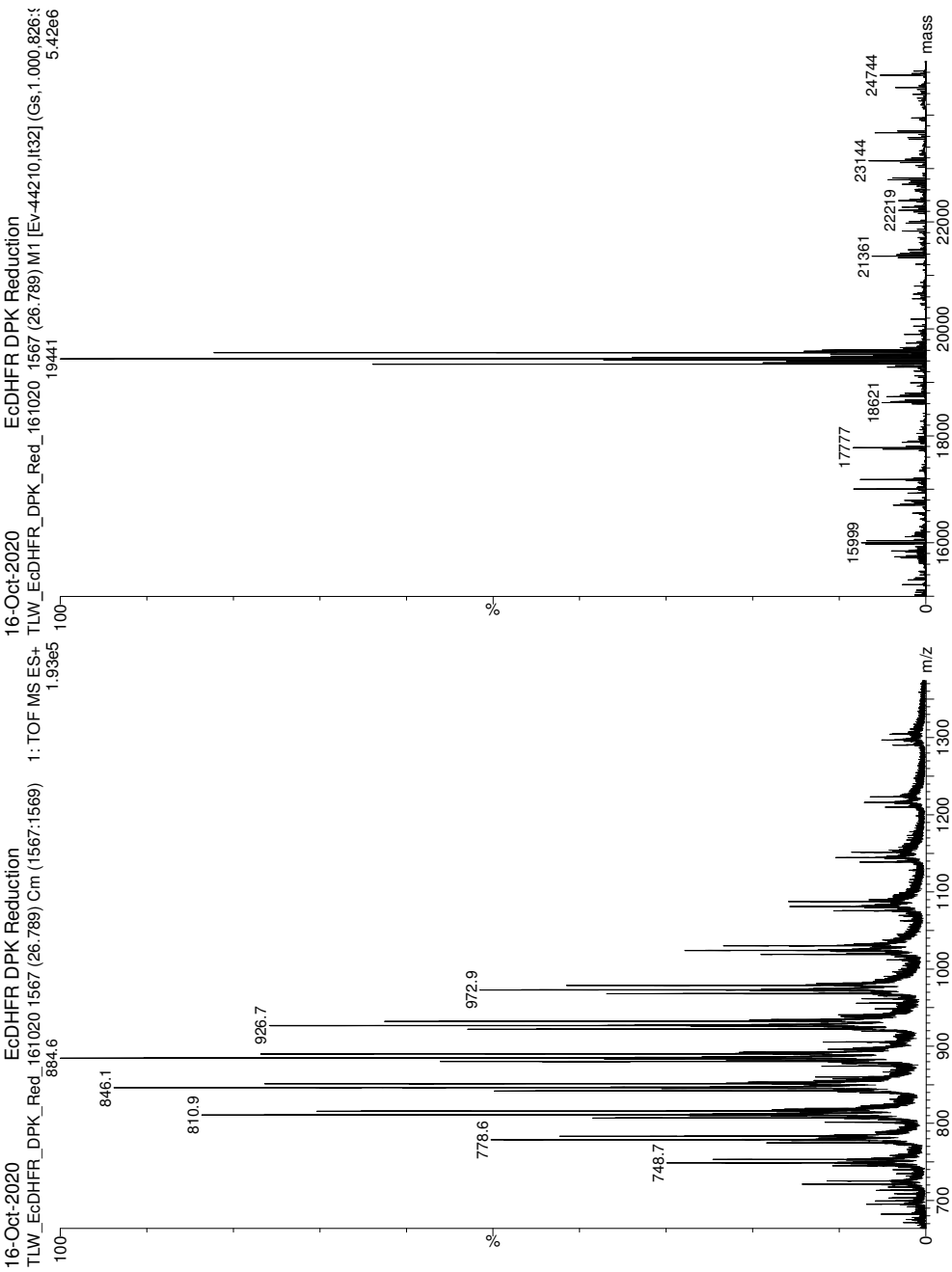
EcDHFR (S49DPK)



LmrR (F93DPK) Cyanoborohydride Reduction

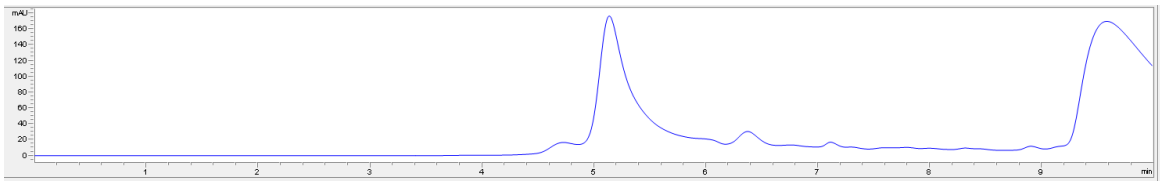


EcDHFR (Ala7DPK) Cyanoborohydride Reduction

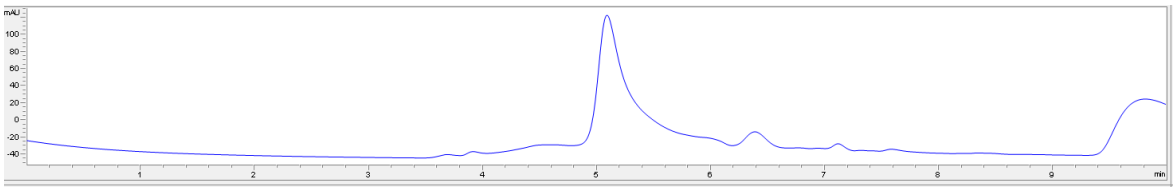


7.3 Analytical Size Exclusion Chromatograms (LmrR)

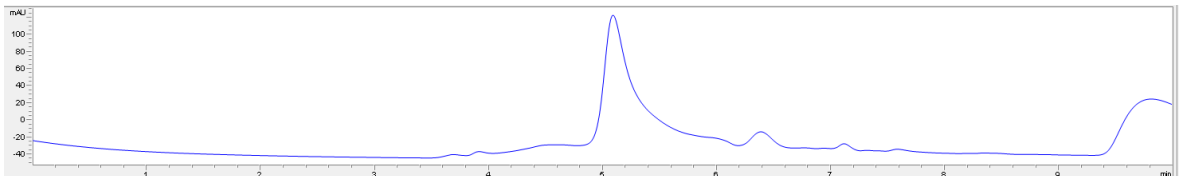
Wt-LmrR



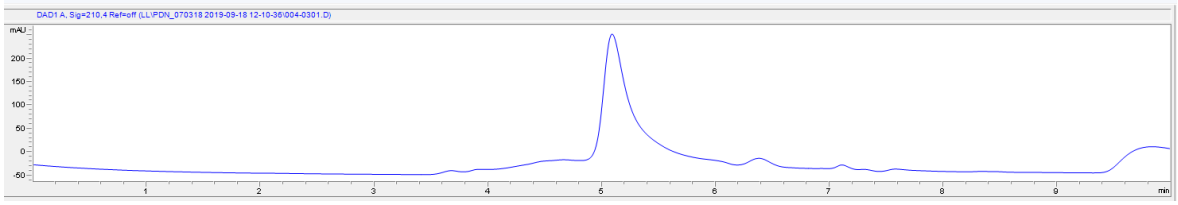
V15DPK



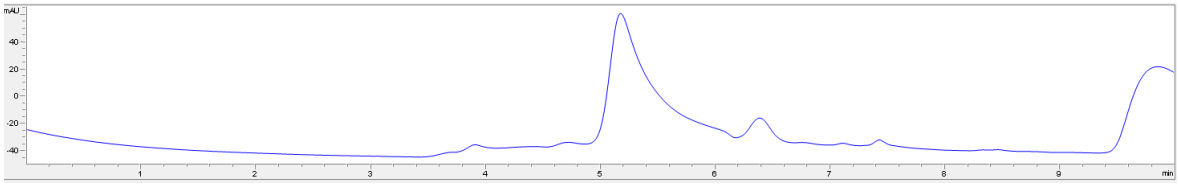
V15LPK



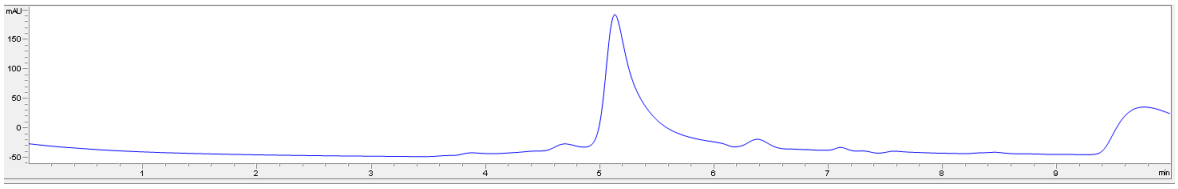
V15Thz



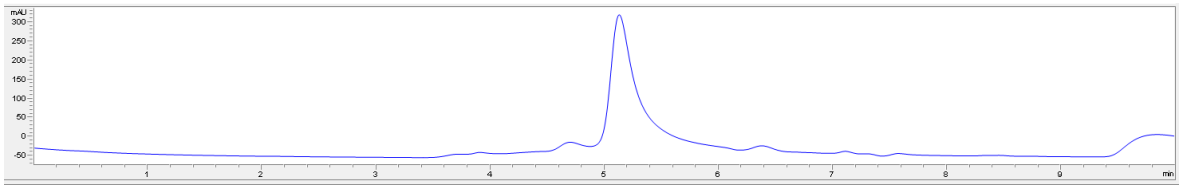
N19DPK



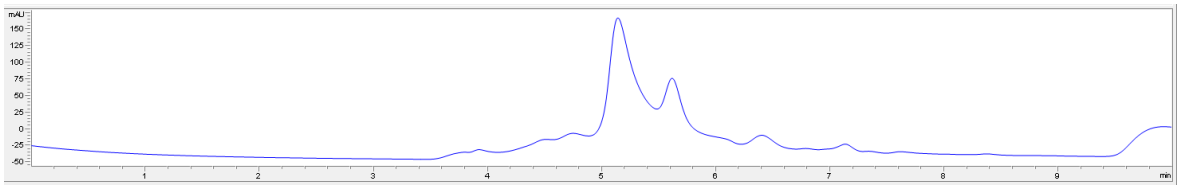
N19LPK



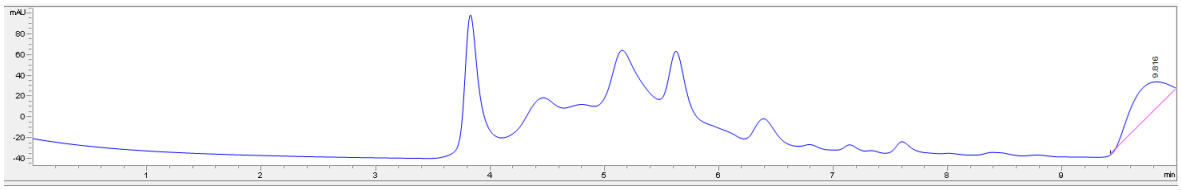
N19Thz



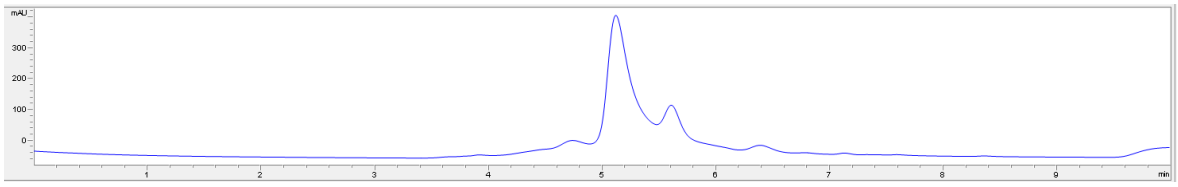
M89DPK



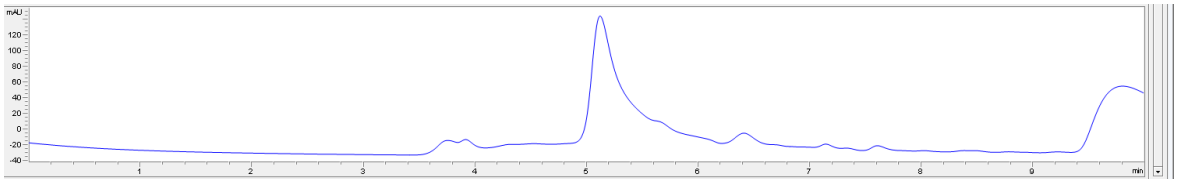
M89LPK



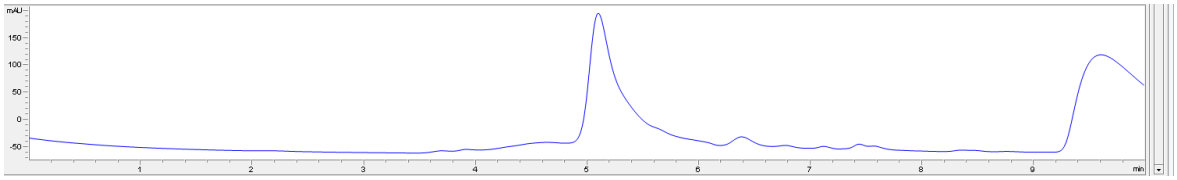
M89Thz



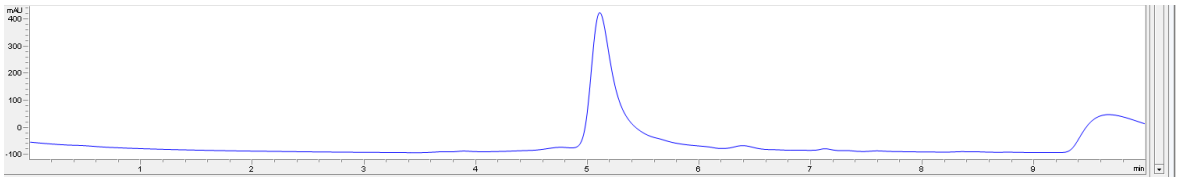
F93DPK



F93LPK

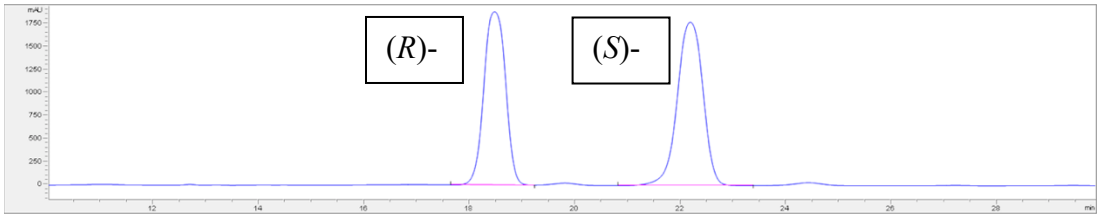


F93Thz (**3**)



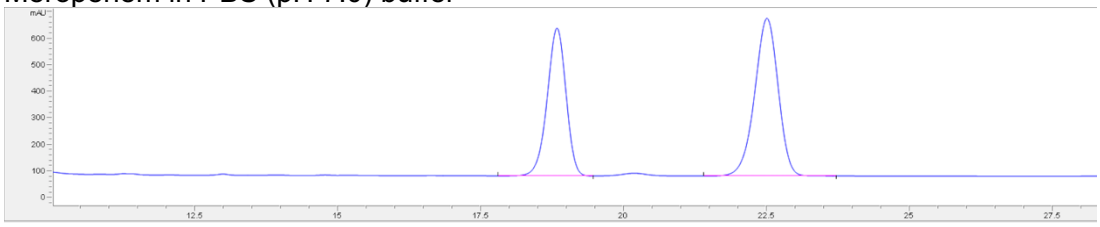
7.4 Chiral HPLC Chromatograms

Racemic standard



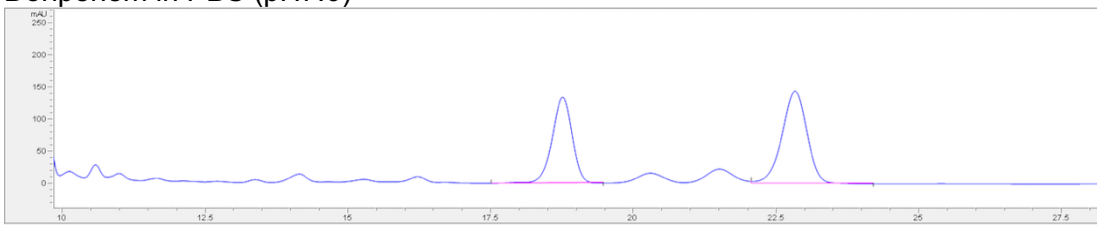
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.5	53631.9	47.2
2	22.2	60200.8	52.8

Meropenem in PBS (pH 7.0) buffer



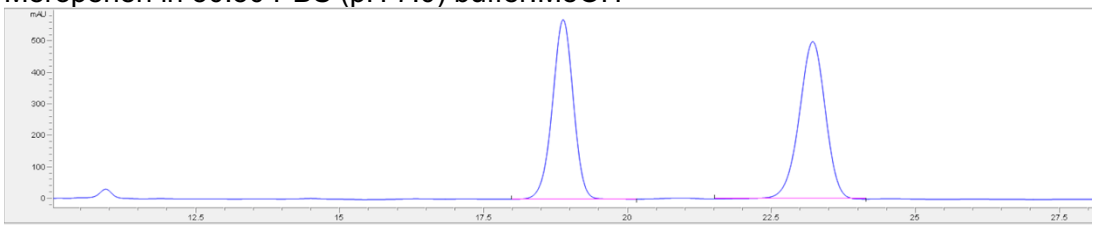
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.8	12791.9	42.9
2	22.5	17005.7	57.1

Doripenem in PBS (pH7.0)



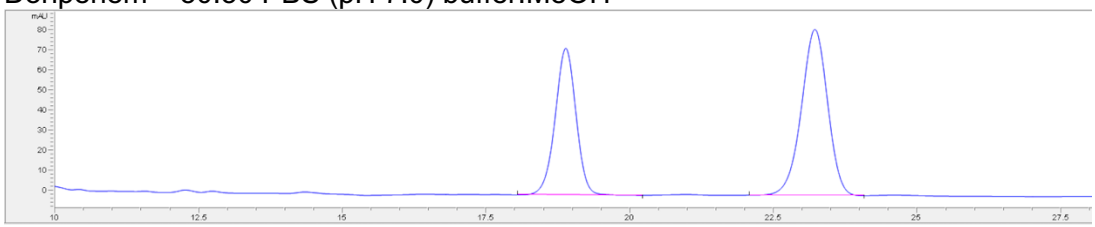
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.8	3277.4	42.7
2	22.8	4339.2	57.3

Meropenen in 50:50 PBS (pH 7.0) buffer:MeOH



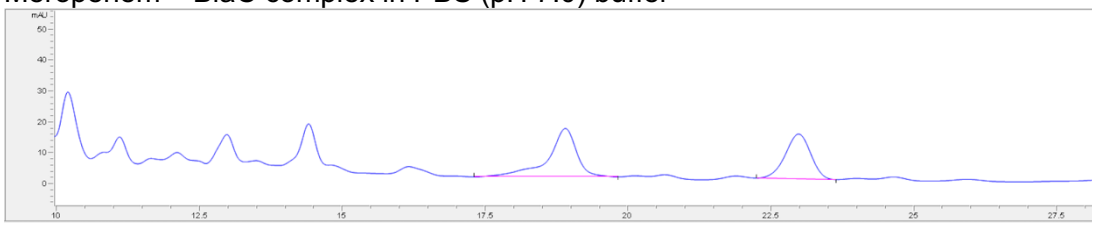
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.9	1855.8	41.3
2	23.1	2638.8	58.7

Doripenem – 50:50 PBS (pH 7.0) buffer:MeOH



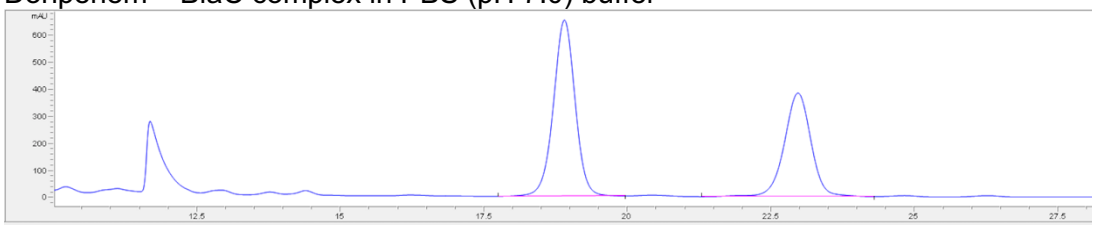
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.9	14448.3	47.4
2	23.1	16012.0	52.6

Meropenem – BlaC complex in PBS (pH 7.0) buffer



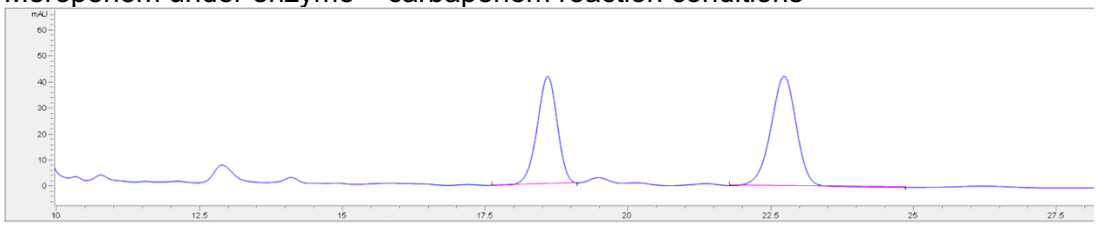
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.9	549.3	55.1
2	23.0	447.7	44.9

Doripenem – BlaC complex in PBS (pH 7.0) buffer



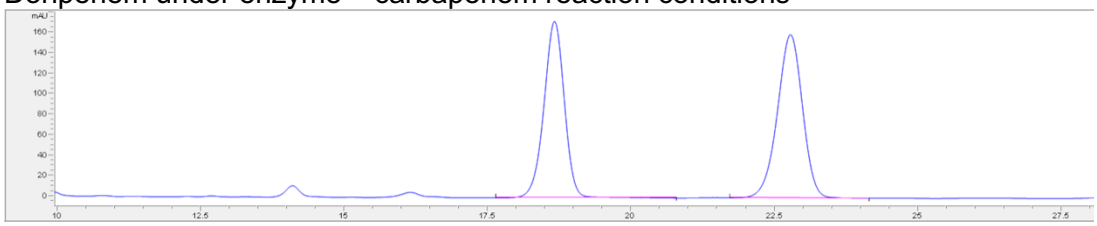
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.9	17109.7	58.0
2	23.1	12368.4	42.0

Meropenem under enzyme – carbapenem reaction conditions



Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.6	1009.3	43.4
2	22.7	1318.9	56.6

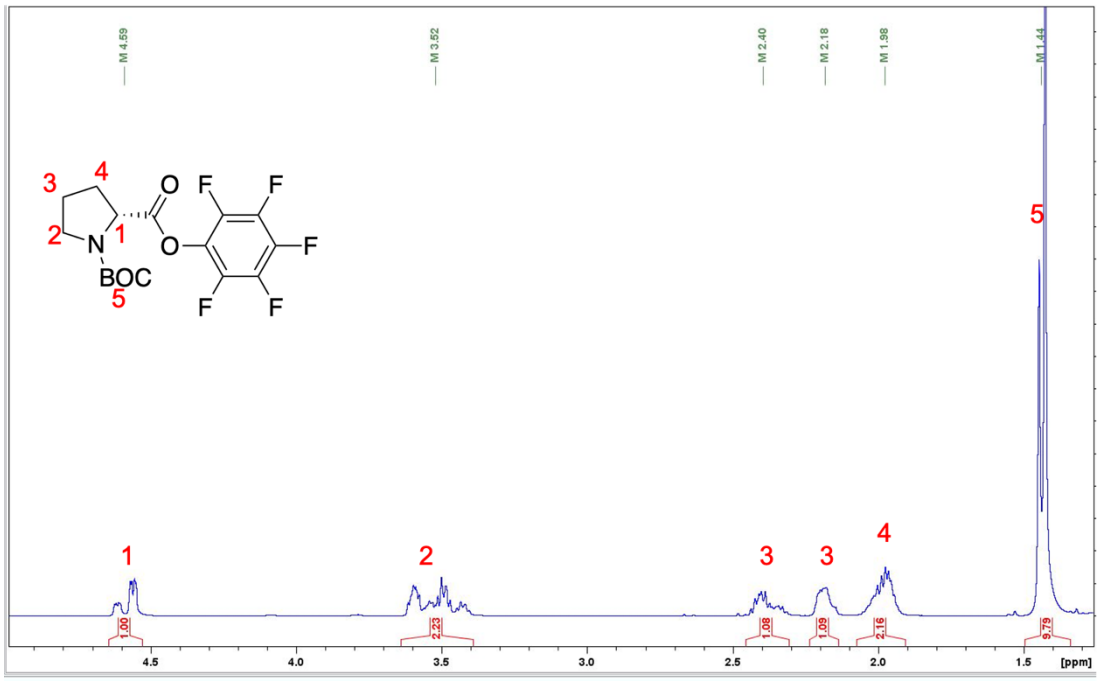
Doripenem under enzyme – carbapenem reaction conditions



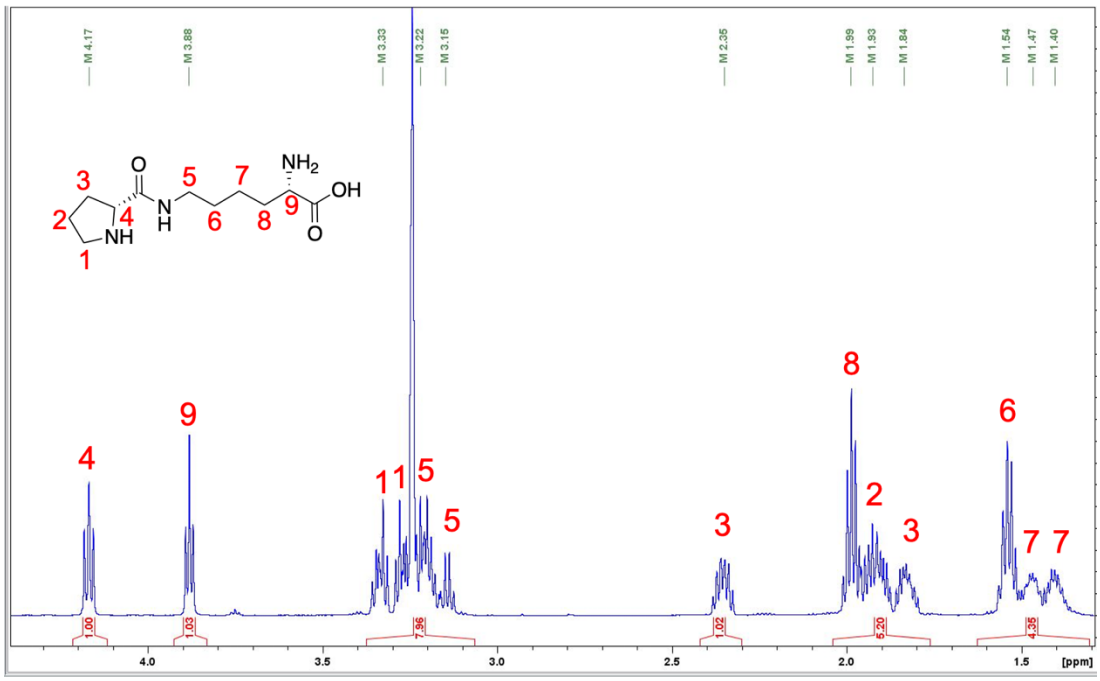
Peak	Rt (Min)	Area (mAU*s)	Area (%)
1	18.7	4314.2	46.3
2	22.8	4877.5	53.7

7.5 NMR Spectra

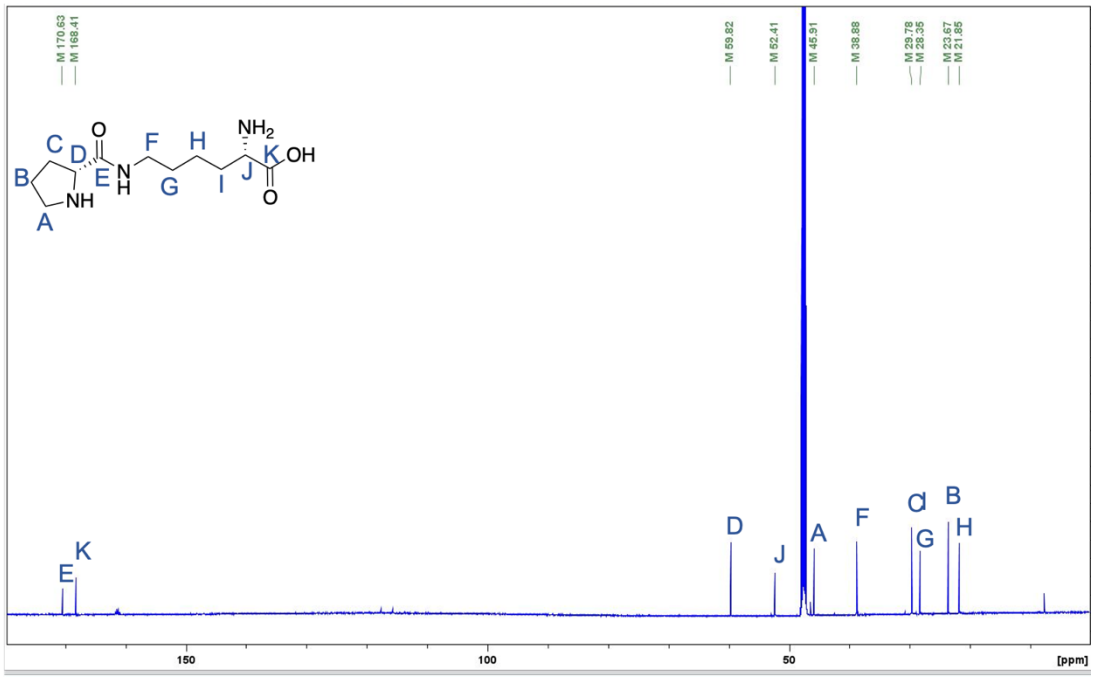
DPK-PFP ester ^{1}H NMR spectrum



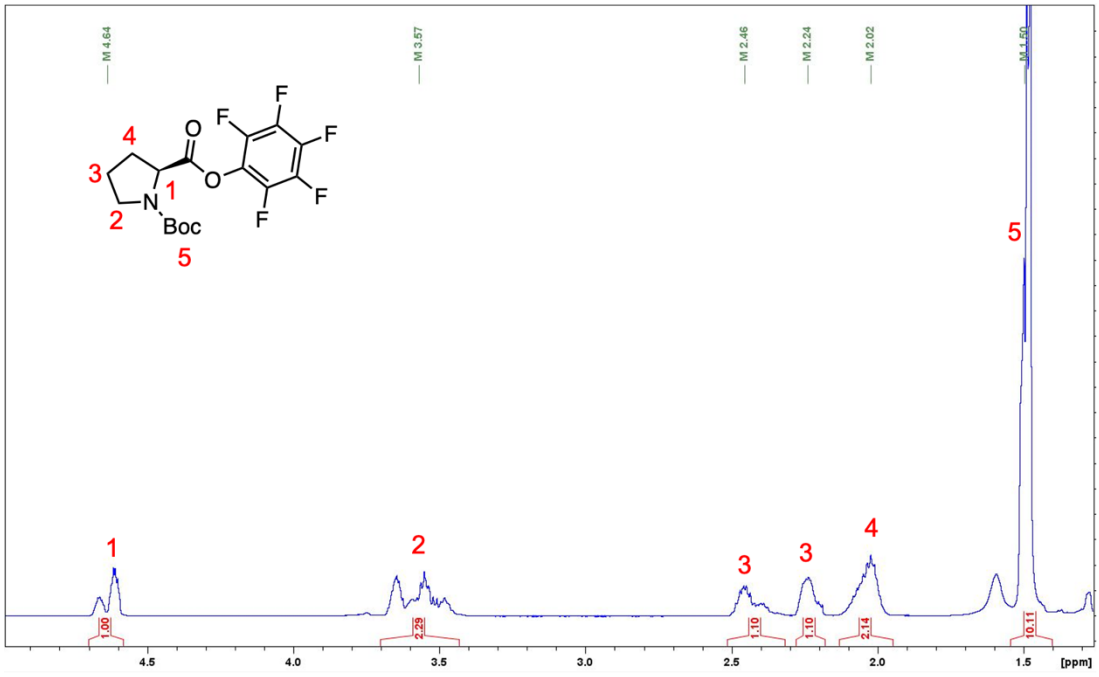
DPK ^{1}H NMR spectrum



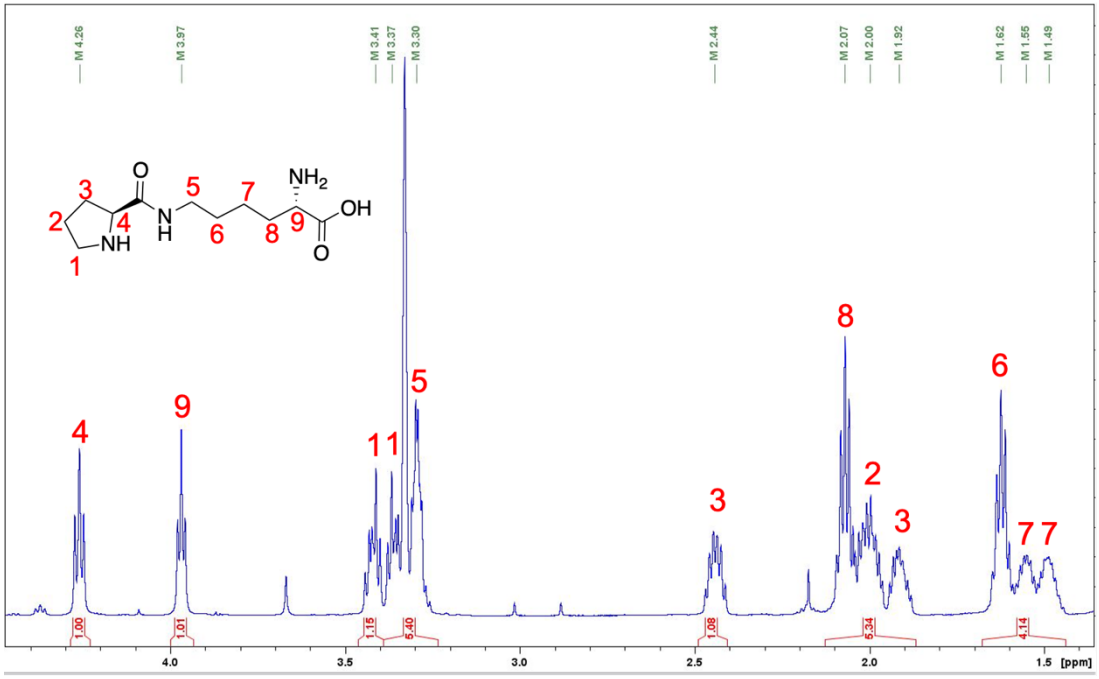
DPK ^{13}C NMR spectrum



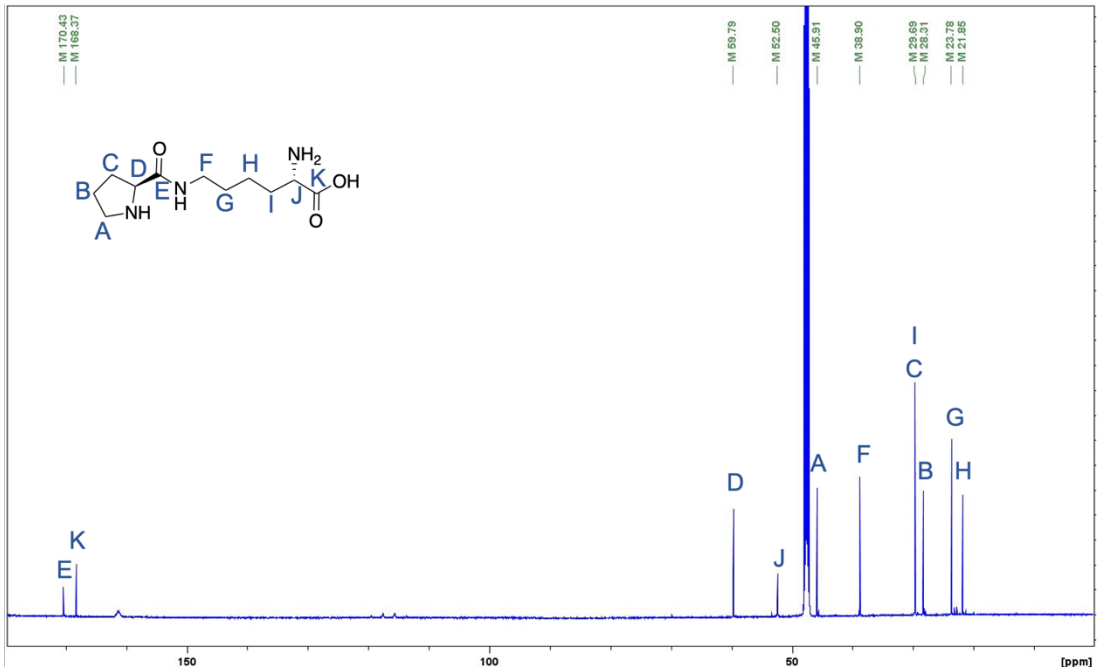
LPK-PFP ester ^{1}H NMR spectrum



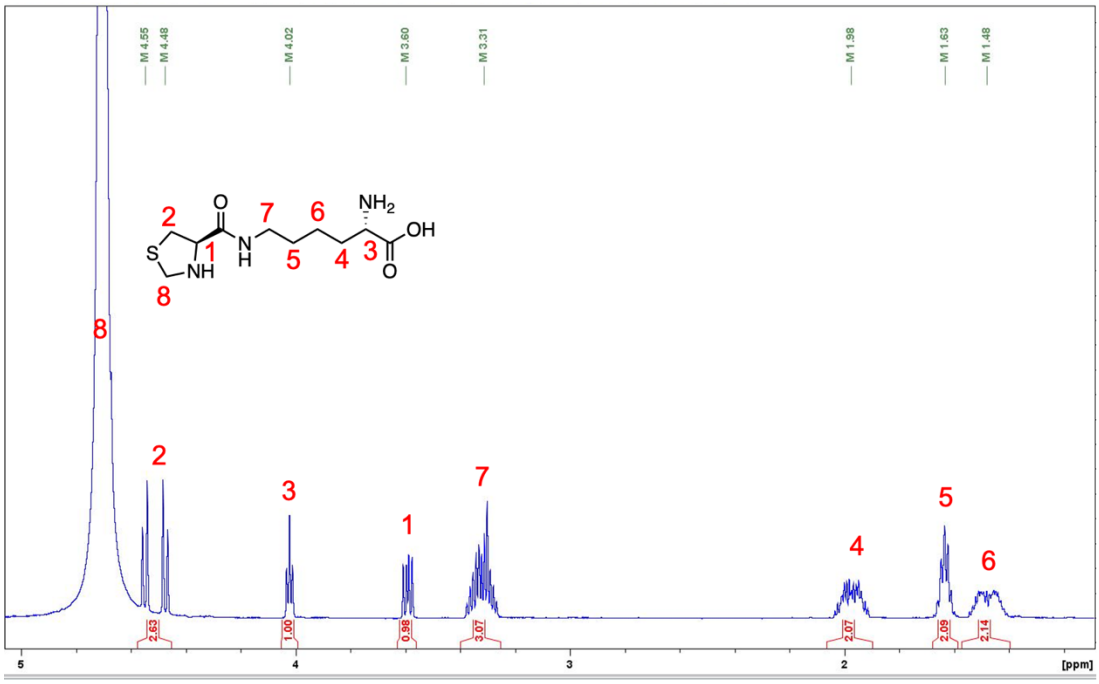
LPK ^{1}H NMR spectrum



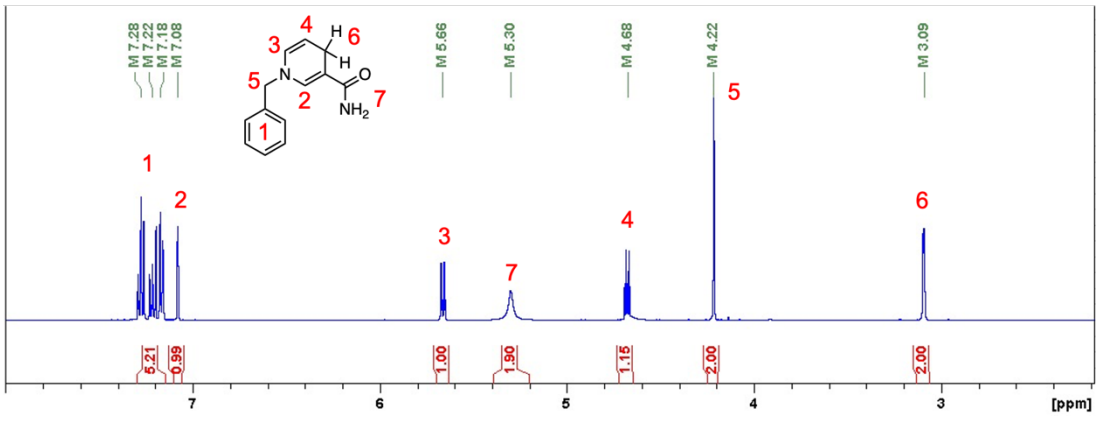
LPK ^{13}C NMR spectrum



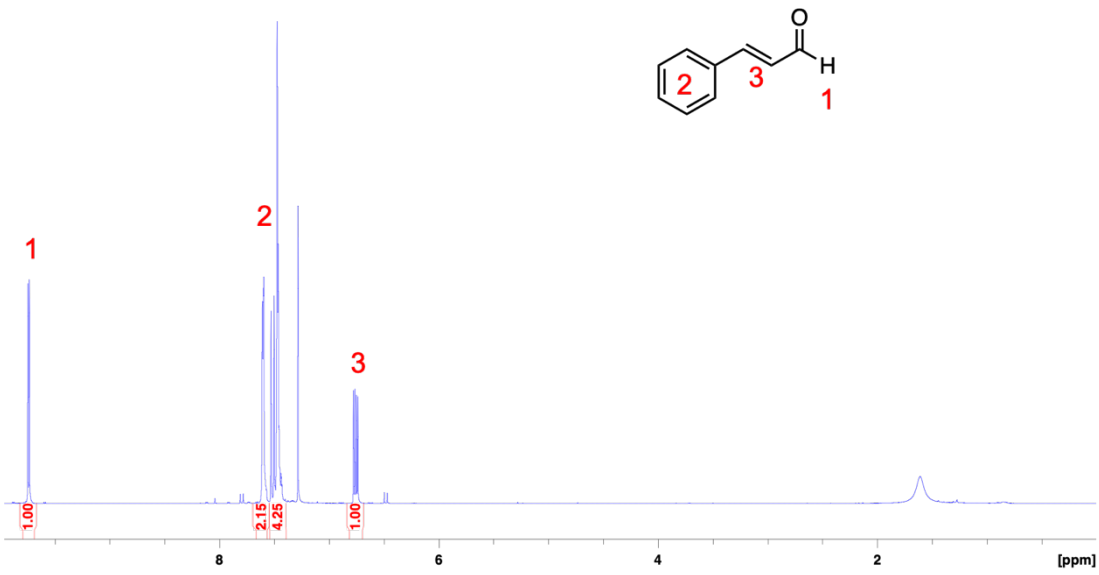
ThzK ^{1}H NMR spectrum



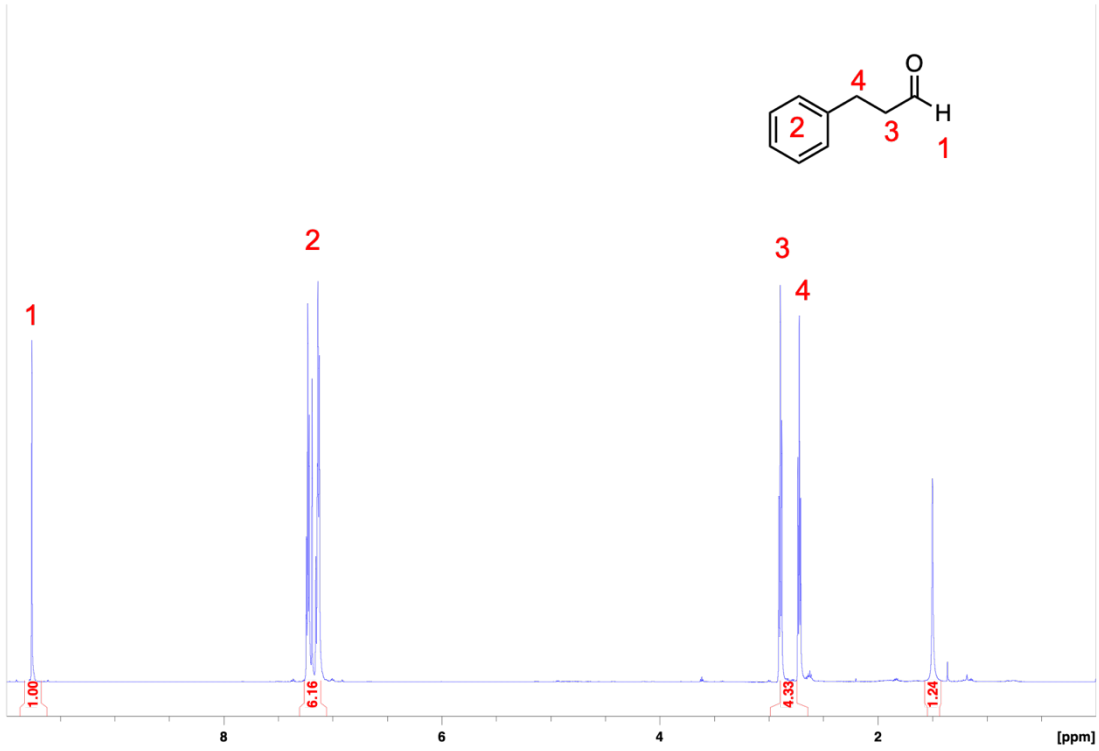
BNAH ^{1}H NMR Spectrum



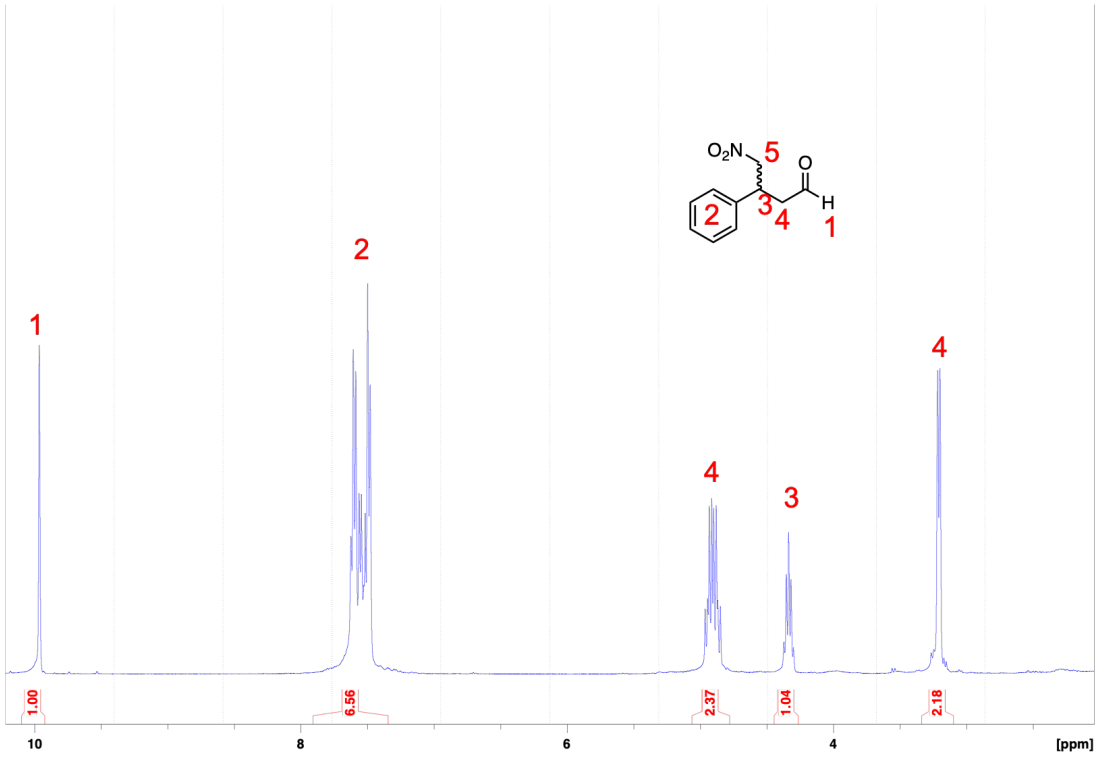
Cinnamaldehyde ^{1}H NMR Spectrum



Hydro-Cinnamaldehyde ^{1}H NMR Spectrum

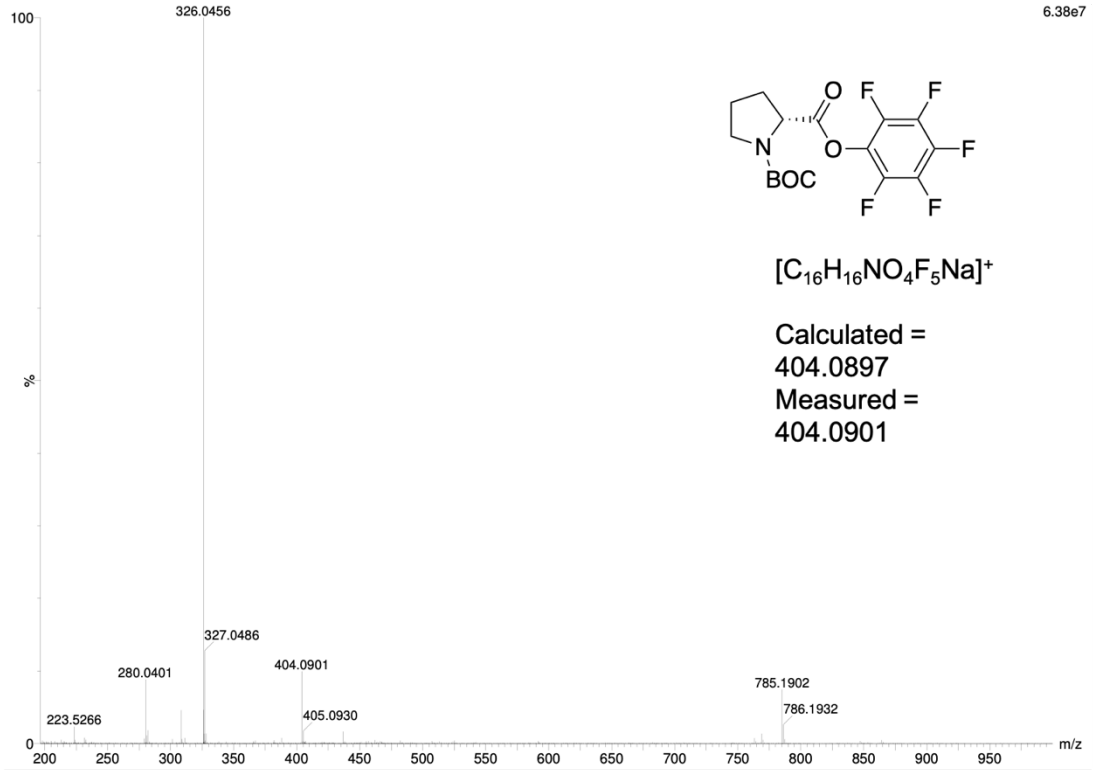


Nitro-Michael addition ^{1}H NMR Spectrum

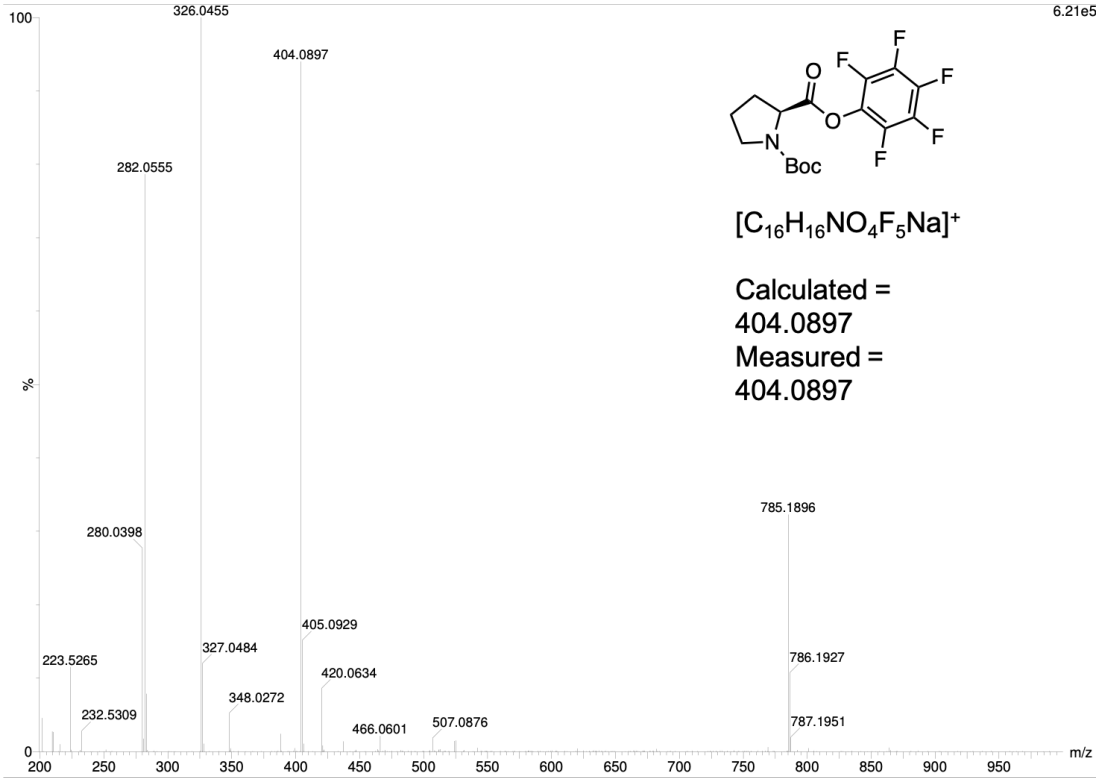


7.6 Mass spectra

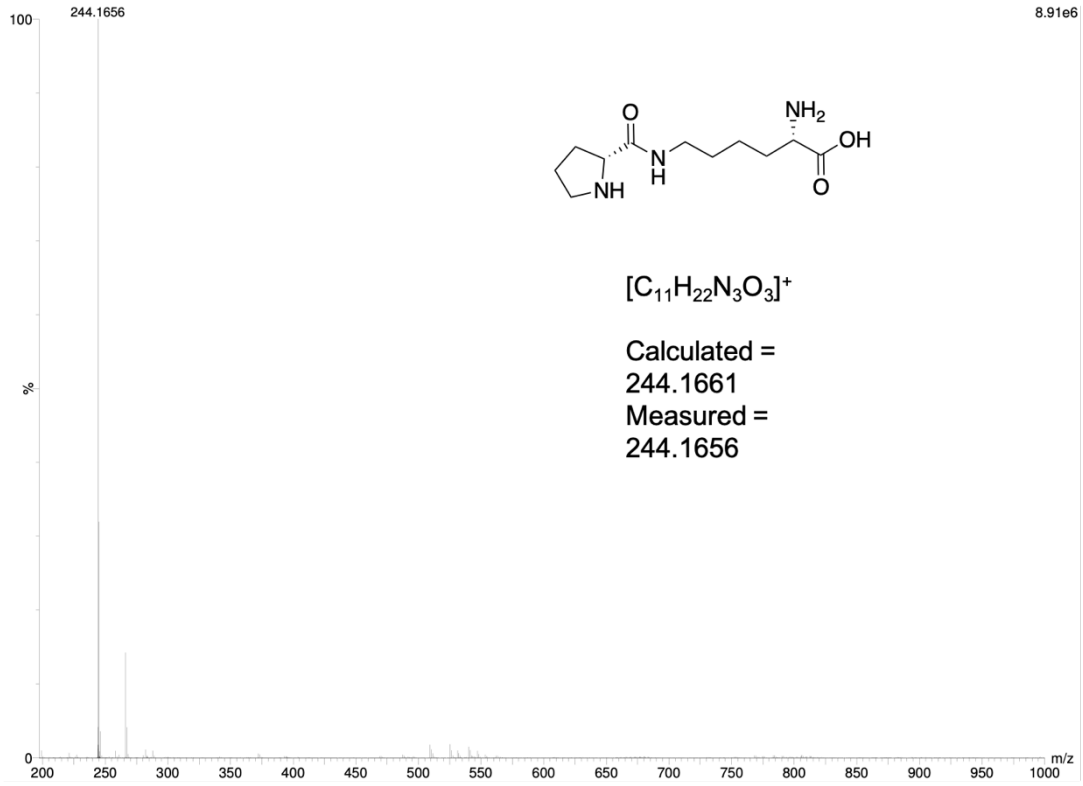
D-prolyl-PFP-ester High Resolution Mass spectrum



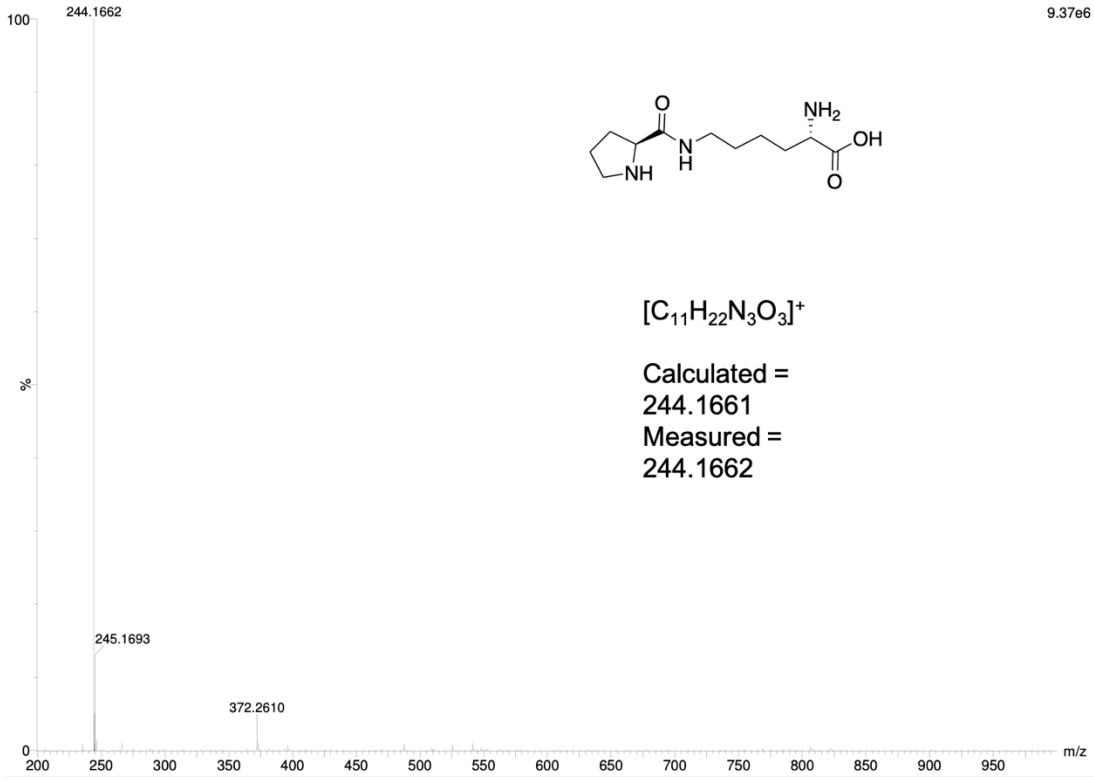
L-prolyl-PFP-Ester High Resolution Mass spectrum



DPK High Resolution Mass spectrum



LPK High Resolution Mass spectrum



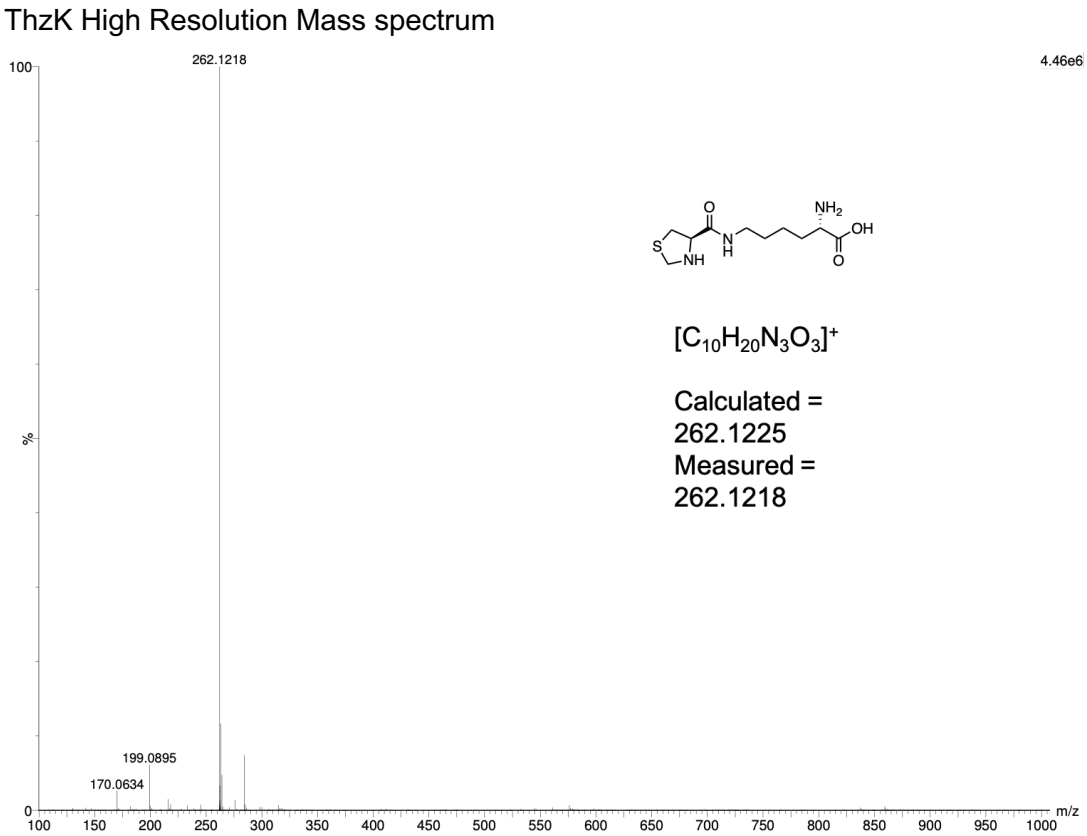


FIGURE 1. **A)** INDUCED FIT MODEL DESCRIBES THE PREORGANISED ACTIVE SITE CONFORMATION OF AN ENZYME FOR A SPECIFIC SUBSTRATE. THE ENZYME BINDS THE SUBSTRATE TO FORM AN ENZYME:SUBSTRATE COMPLEX (E:S) TO CATALYSE THE REACTION FOLLOWED BY PRODUCT RELEASE. **B)** ENZYME CATALYSIS ENERGY DIAGRAM. THE ACTIVATION ENERGY (E_A) OF THE ENZYME CATALYSED REACTION IS LOWER THAN THAT OF THE UNCATALYSED REACTION. BLUE LINE IS THE UNCATALYSED REACTION TRAJECTORY. GREEN LINE IS THE ENZYME CATALYSED REACTION TRAJECTORY. 3
FIGURE 2. EXAMPLES OF ENZYMES APPLIED IN INDUSTRIAL SETTINGS AS BIOCATALYSTS.[8, 9] 4
FIGURE 3. A) STEREOSELECTIVE REDUCTION OF ACETOPHENONE TO THE ALCOHOL CATALYSED BY WILD-TYPE CARBONYL REDUCTASE (CR) USING NADH AS THE HYDRIDE SOURCE. **B)** EFFECT OF MUTATIONS ON SUBSTRATE SCOPE AND STEREOCHEMISTRY ON 24 KETONE SUBSTRATES. **C)** REDUCTION OF ONE SUBSTRATE BY TWO VARIANTS OF THE ENZYME TO YIELD EITHER THE *R* OR *S* ENANTIOMER. **D)** CRYSTAL STRUCTURE OF THE CARBONYL REDUCTASE FROM *C. PARAPSILOSIS* WITH NAD^+ BOUND IN THE ACTIVE SITE (**PDB: 3WLE**) USED TO RATIONALLY RE-ENGINEER THE BINDING SITE. 7
FIGURE 4. A) CHEMICAL SYNTHESIS OF SITAGLIPTIN USING A CHIRAL RHODIUM CATALYST TO REDUCE THE ENAMINE USING HIGH PRESSURE H_2. **B)** BIOCATALYTIC ROUTE USING A TRANSAMINASE AND PLP WITH THE 2-PROPYLAIME AS THE AMINE DONOR. 9
FIGURE 5. DIRECTED EVOLUTION OF A TRANSAMINASE FOR THE BIOCATALYTIC SYNTHESIS OF SITAGLIPTIN...10
FIGURE 6. THE REPURPOSING OF HAEM BEARING ENZYMES FOR NITRILE AND CARBENE CHEMISTRY. FOLLOWING SELECTION, THE ENZYMES ARE IMPROVED THROUGH DIRECTED EVOLUTION TO GENERATE NEW ENZYME ACTIVITIES.....12
FIGURE 7. A) ALKYLATION AT SP^3 CARBON-HYDROGEN BONDS USING A REPURPOSED AND EVOLVED CYTOCHROME P450 ENZYME. **B)** SUBSTRATE SCOPE OF THE ENZYME USING AMINE BEARING SUBSTRATES AND DIFFERENT ALKYL SUBSTITUENTS. **C)** CHEMOENZYMATIC SYNTHESIS OF (+)-LYNGBIC ACID UTILISING THE REPURPOSED ENZYME TO FORM THE CHIRAL CENTRE. **D)** CHEMOENZYMATIC SYNTHESIS OF (-)-CUSPAREINE UTILISING THE REPURPOSED ENZYME TO FORM THE CHIRAL CENTRE.....13
FIGURE 8. A) THE STRUCTURE OF D-BIOTIN. **B)** A RUTHENIUM (RU) CATALYST COMPLEX HOSTED BY TETRAMERIC STREPTAVIDIN (**PDB:1STP**) FOR THE SELECTIVE HYDRATION OF KETONES. **C)** AN IRIDIUM (IR) CATALYST COMPLEX HOSTED BY DIMERIC STREPTAVIDIN (**PDB: 6S4Q**) FOR THE REDUCTION OF

CYCLIC IMINES. **D)** A RHODIUM (RH) CATALYST COMPLEX HOSTED BY MONOMERIC STREPTAVIDIN (**PDB:4JNJ**) FOR THE SELECTIVE SYNTHESIS OF Δ-LACTAMS. 15
FIGURE 9. A MAMMALIAN GENE SWITCH DEVISED BY USING A PURPOSEFULLY DESIGNED ARTIFICIAL METALLOENZYME. **A)** TWO BIOTIN CONJUGATES ONE FOR ABIOTIC CATALYSIS AND THE SECOND FOR CELL PENETRATION AND REPORTING. **B)** TETRAMERIC STREPTAVIDIN WITH FOUR LIGAND BINDING POCKETS ALLOWS DUAL OCCUPATION. **C)** SUBSTRATE TARGET WHICH ACTS AS A GENE REPRESSOR. **D**) ASSEMBLY OF THE ARM FOLLOWED BY DELIVERY TO THE CELL AND ACTIVATION OF BIOLUMINESCENT LUCIFERIN. 16
FIGURE 10. THE USE OF DIMERIC LMRR FOR THE DESIGN OF ARMS. **A)** SUPRAMOLECULAR ASSEMBLY TO GENERATE ARTIFICIAL ENZYMES CAPABLE OF CATALYSING FRIEDEL-CRAFT AND CYCLOPROPANATION REACTIONS. **B)** COVALENT MODIFICATION TO AFFORD DIELS-ALDER REACTIONS. **C)** METAL BINDING UNNATURAL AMINO ACID INCORPORATION THROUGH GENETIC CODE EXPANSION TO PERFORM FRIEDEL-CRAFTS AND HYDROXYLATION REACTIONS. 17
FIGURE 11. EXAMPLES OF SMALL MOLECULE ORGANOCATALYSTS. **A)** AMINE BASED ORGANOCATALYSTS THAT CAN ACTIVATE CARBONYL COMPOUNDS THROUGH THE FORMATION OF IMINIUM ION INTERMEDIATES. **B)** (THIO)UREA BASED ORGANOCATALYSTS THAT CAN ACTIVATE CARBONYLS THROUGH NON-COVALENT HYDROGEN BONDING. 18
FIGURE 12. FOUR METHODS OF DESIGNING AND GENERATING ORGANOCATALYTIC ARTIFICIAL ENZYMES. 19
FIGURE 13. THE USE OF CYSTEINE AS A REACTIVE THIOL NUCLEOPHILE TO ATTACH A CATALYTICALLY ACTIVE MOLECULE SITE-SPECIFICALLY TO A PROTEIN BACKBONE. 19
FIGURE 14. THREE EXAMPLES OF COVALENT MODIFICATION OF PROTEINS WITH CO-FACTORS TO IMPART NEW REACTIONS TO THE PROTEIN BACKBONES. **A)** COVALENT MODIFICATION OF PAPAIN WITH FLAVIN FOR THE OXIDATION OF BNAH. **B)** COVALENT MODIFICATION OF PAPAIN WITH THIAMINE TO PERFORM CARBON – CARBON BOND FORMING REACTIONS. **C)** COVALENT MODIFICATION OF ADIPOCYTE BINDING PROTEIN WITH PYRIDOXAMINE FOR THE TRANSAMINATION OF KETO ACIDS TO AMINO ACIDS. 20
FIGURE 15. ENZYMES AND ENZYME SECTIONS CAN BE COMPUTATIONALLY DESIGNED USING RELEVANT SOFTWARE. THESE DESIGNS CAN THEN BE PRODUCED THROUGH RECOMBINANT EXPRESSION AND TESTED EXPERIMENTALLY. 21
FIGURE 16. THE KEMP ELIMINATION REACTION. ABSTRACTION OF THE PROTON BY A BASE, RESULTS IN RING OPENING. THE TRANSITION STATE RESULTS IN NEGATIVE CHARGE BUILD UP AT THE PHENOLIC OXYGEN. 22
FIGURE 17. A) THE DIELS-ALDER CYCLOADDITION REACTION BETWEEN THE TARGET SUBSTRATES OF THE ENZYME DESIGN. **B)** DESIGN CONSIDERATIONS FOR INCLUDING HYDROGEN BOND ACCEPTOR AND DONOR TO NARROW THE ENERGY GAP BETWEEN THE LUMO AND HOMO OF THE SUBSTRATES BY CHARGE STABILISATION AND DESIRED GEOMETRIC POSITIONING TO REDUCE ENTROPIC CONTRIBUTIONS. **C)** TYROSINE 121 WAS SHOWN TO BE THE HYDROGEN BOND DONOR AND GLUTAMINE 195 THE HYDROGEN BOND ACCEPTOR. **D)** CRYSTAL STRUCTURE OF THE ARTIFICIAL DIELS-ALDERASE RESULTING FROM THE COMPUTATIONAL DESIGN BASED ON A B-PROPELLER PHOSPHATASE SCAFFOLD (**PDB: 3I1C**). 24
FIGURE 18. **A)** THE MORITA-BAYLISS-HILLMAN REACTION THAT PROCEEDS THROUGH TWO INTERMEDIATES. **B)** CRYSTAL STRUCTURE OF BH25 WITH THE CATALYTIC DYAD CYS39 AND LYS286 HIGHLIGHTED IN BLUE **PDB: 3UW6**). **C)** CRYSTAL STRUCTURE OF BH32 WITH THE CATALYTIC DYAD HIS23 AND GLU46 HIGHLIGHTED IN BLUE (**PDB: 3U26**). 26
FIGURE 19. A) CRYSTAL STRUCTURE OF THE COMPUTATIONALLY DESIGNED RETRO-ALDOLASE RA 95.0 (GREEN) (**PDB: 4A29**) AND RA 95.5 (YELLOW) (**PDB: 5AN7**) **B)** POSITIONS OF THE CATALYTIC LYS210 (RA 95.0) AND LYS83 (RA 95.5) BOUND TO KETONE INHIBITORS. **C)** PROPOSED CATALYTIC CYCLE FOR THE RETRO-ALDOL REACTION OF METHODOL. 27
FIGURE 20. REACTION SCOPE OF THE RA95 FAMILY OF COMPUTATIONALLY DESIGNED RETRO-ALDOLASES. **A)** THE ORIGINAL RETRO ALDOL REACTION USED FOR INITIAL SCREENING, THE CLEAVAGE OF THE STARTING PRODUCT (METHODOL) RESULTS IN A CHANGE IN FLUORESCENCE**. B)** HENRY ADDITION OF NITROMETHANE TO AROMATIC ALDEHYDES. **C)** CONJUGATE ADDITION OF ACETONE TO NITROSTYRENE**.** ENAMINE CATALYSED CONJUGATE ADDITION. **D)** CONJUGATE ADDITION OF CARBON NUCLEOPHILES. **E)** CONJUGATE OF NITROMETHANE TO A,B-UNSATURATED ALDEHYDES. **F)** KNOEVENAGEL CONDENSATIONS OF CARBON NUCLEOPHILES WITH A,B-UNSATURATED AROMATIC ALDEHYDES. 28
FIGURE 21. CARTOON REPRESENTATION OF THE COMPONENTS FOR NON-COVALENT TETHERING THAT CONSISTS OF A PROTEIN HOST, A TIGHT BINDING LIGAND AND A CATALYTIC MOIETY. 29

FIGURE 22. DECARBOXYLATIVE ALKYLATION REACTION PERFORMED BY THE NDI Π-ANION BASED STREPTAVIDIN CATALYST. 30
FIGURE 23. **A)** FIVE Π-ANION NDI BASED CATALYSTS CONJUGATED TO BIOTIN. **B)** CRYSTAL STRUCTURE OF TWO BIOTIN MONOMERS WITH BIOTIN ANCHORED INSIDE. RESIDUES IN BLUE WERE CRUCIAL FOR INSIGHTS INTO THE HYBRID CATALYSTS ACTIVITY (**PDB: 1MK5**). 31
FIGURE 24. MICHAEL-ALDOL DOMINO REACTION PERFORMED BY THE NDI Π-ANION BASED STREPTAVIDIN CATALYST. 31
FIGURE 25. ADDITION OF THE UNNATURAL AMINO ACID TO THE CULTURE MEDIUM ALLOWS INCORPORATION INTO THE PROTEIN BACKBONE BY AN ORTHOGONAL TRNA SYNTHETASE/TRNA 33
FIGURE 26. A) CRYSTAL STRUCTURE OF LMRR (**PDB: 3F8F**). **B)** CLOSE UP OF THE DIMER INTERFACE. THE RESIDUES IN ORANGE WERE THOSE TARGETED FOR MUTATION. **C)** REDUCTION OF *P*-AZIDOPHENYLALANINE TO *P*-AMINOPHENYLALANINE WITH TCEP. **D)** HYDRAZONE AND OXIME LIGATION PERFORMED BY THE UNNATURAL AMINO ACID *P*-AMINOPHENYLALANINE AT POSITION V15. 34
FIGURE 27. CRYSTAL STRUCTURE OF LMRR WITH THE UNNATURAL AMINO ACID *P*-AMINOPHENYLALANINE AT POSITION 15 (ORANGE) (**PDB: 6I8N**). RED RESIDUES THAT SURROUND THE ACTIVE SITE WERE TARGETED FOR DIRECTED EVOLUTION. 35
FIGURE 28. **A)** CRYSTAL STRUCTURE OF OA1.3 DERIVED FROM THE PROTEIN BACKBONE BH32. RESIDUES IN RED ARE THOSE RESULTING FROM DIRECTED EVOLUTION (**PDB: 6Q7R**). **B)**. FLUORESCEIN ESTER HYDROLYSIS CATALYSED BY METHYL-HISTIDINE. **C)** UNNATURAL AMINO ACID METHYL-HISTIDINE BOUND TO THE INHIBITOR BROMOACETOPHENONE. **D)** STABLE INTERMEDIATE FORMED BY HIS23 AND REACTIVE INTERMEDIATE FORMED BY M-HIS23. 36
FIGURE 29. THREE COMMON CYCLIC SECONDARY AMINE ORGANOCATALYSTS. **A)** L-PROLINE. **B)** FIRST GENERATION IMIDAZOLIDINONE (MACMILLAN) CATALYST. **C)** JORGENSEN-HAYASHI CATALYST. 37
FIGURE 30. A) IMINIUM ACTIVATION BY CYCLIC SECONDARY AMINES. **B)** DIAGRAM REPRESENTING THE REDUCTION OF THE ENERGY OF THE LUMO DUE TO THE IMINIUM SPECIES WITH RESPECT TO THE PARENT CARBONYL ION. **C)** ENAMINE ACTIVATION BY CYCLIC SECONDARY AMINES. **D)** DIAGRAM REPRESENTING THE INCREASE IN ENERGY OF THE HOMO OF THE ENAMINE SPECIES WITH RESPECT TO THE PARENT CARBONYL. 38
FIGURE 31. A) CRYSTAL STRUCTURE OF 4-OT WITH ONE MONOMER HIGHLIGHTED IN DARK BLUE. **B**) THE ENZYME'S NATIVE REACTION, A BASE CATALYSED TAUTOMERISATION. 40
FIGURE 32. REACTIONS CATALYSED BY 4-OT TAUTOMERASE. **A)** INTERMOLECULAR ALDOL REACTION. **B)** INTRAMOLECULAR ALDOL REACTION. **C)** ENAMINE CATALYSED CONJUGATE ADDITION. **D)** NITRO-MICHAEL ADDITION. 41
FIGURE 33. CHEMOENZYMATIC SYNTHESIS OF PREGABALIN AND THREE ANALOGUES. 42
FIGURE 34. A) THE STRUCTURE OF D-BIOTIN AND EIGHT SECONDARY AMINE ORGANOCATALYSTS THAT WERE TESTED. **B)** TETRAMERIC STRUCTURE OF STREPTAVIDIN (**PDB: 6GH7**) WITH BIOTIN PYRROLIDINE CATALYST (*R*) BOUND IN THE HYDROPHOBIC POCKET **C)** THE NITRO-MICHAEL ADDITION REACTION BETWEEN NITROMETHANE AND A RANGE OF A,B-UNSATURATED ALDEHYDES. 43
FIGURE 35. A) THE GENERAL MECHANISM OF THE PENICILLIN BINDING PROTEIN D-ALA-D-ALANINE TRANSPEPTIDASE THAT FACILITATES THE CROSS LINKING OF PEPTIDOGLYCAN IN BACTERIAL CELL WALL BIOSYNTHESIS (PG = PEPTIDOGLYCAN). **B)** THE GENERAL MECHANISM OF B-LACTAM HYDROLYSIS BY CLASS A B-LACTAMASES. BOTH ENZYMES REQUIRED A CATALYTIC SERINE (SHOWN IN BLUE). 54
FIGURE 36. A) GENERAL STRUCTURE OF THE PENICILLINS. **B)** GENERAL STRUCTURE OF THE CEPHALOSPORINS. **C)** GENERAL STRUCTURE OF THE CARBAPENEMS. **D)** THE STRUCTURES OF THE THREE CARBAPENEMS USED IN THIS STUDY. 55
FIGURE 37. A) CRYSTAL STRUCTURE OF D,D-TRANSPEPTIDASE RV3330 FROM *MYCOBACTERIUM TUBERCULOSIS* (**PDB: 4PPR**) WITH MEROPENEM COVALENTLY BOUND IN THE ACTIVE SITE. **B)** CLOSE UP OF THE MEROPENEM ACYL-ENZYME INTERMEDIATE BOUND TO THE NUCLEOPHILIC SERINE (RED) OF RV3330. **C)** CLOSE UP OF THE MEROPENEM ACYL-ENZYME INTERMEDIATE BOUND TO THE NUCLEOPHILIC SERINE (RED) OF BLAC. **D)** CRYSTAL STRUCTURE OF B-LACTAMASE BLAC FROM *MYCOBACTERIUM TUBERCULOSIS* (**PDB: 3DWZ**) WITH MEROPENEM COVALENTLY BOUND IN THE ACTIVE SITE. **E)** CHEMICAL STRUCTURE REPRESENTATION OF AS AN ESTER ADDUCT OF A NUCLEOPHILIC SERINE BOUND TO MEROPENEM. 56
FIGURE 38. THE NITRO MICHAEL ADDITION BETWEEN NITROMETHANE (**1**) AND CINNAMALDEHYDE (**2**) TO YIELD THE 1, 4 ADDITION PRODUCT (**3**) OR THE 1, 2 ADDITION SIDE PRODUCT (**4**). 57
FIGURE 39. SDS-PAGE ANALYSIS OF RECOMBINANT RV3330. CALCULATED MASS = 35879 DA. **A)** EXPRESSION ANALYSIS OF RV3330, 1 – BEFORE INDUCTION, 2 – AFTER INDUCTION WITH IPTG. **B**) NI-NTA

PURIFICATION ANALYSIS OF RV3330, 1 – FLOW THROUGH, 2 – WASH, 3 – ELUTION, 4 – INCLUSION BODIES. PROTEIN LADDERS IN KDA. TWO BANDS ARE PRESENT IN THE ELUTION FRACTION. THE MONOMER OF THE PROTEIN RV3330 AT APPROX. 36 KDA AND THE DISULPHIDE DIMER AT APPROX. 72 KDA. 60
FIGURE 40. DECONVOLUTED ESI-MS OF PURIFIED RV3330 ENZYMES. **A)** ADDUCT FORMATION BETWEEN THE ACTIVE SITE SERINE AND THE CARBAPENEMS. **B)** RV3330 REFERENCE MASS = 35748 DA (35926 DA = RV3330 + FORMYL-MET). **C)** RV3330 + MEROPENEM = 36132 DA (36310 DA = RV3330 + MEROPENEM + FORMYL-MET). **D)** RV3330 + DORIPENEM = 36167 DA (36345 DA = RV3330 + DORIPENEM + FORMYL-MET). 61
FIGURE 41. SDS-PAGE ANALYSIS OF RECOMBINANT BLAC(GLU166ALA). CALCULATED MASS = 30608 DA. **A)** EXPRESSION ANALYSIS OF BLAC(E166A), 1 – BEFORE INDUCTION, 2 – AFTER INDUCTION WITH IPTG. **B)** NI-NTA PURIFICATION ANALYSIS OF BLAC(E166A), 1 – FLOW THROUGH, 2 – WASH 1, 3 – WASH 2, 4 – ELUTION, 5 – INCLUSION BODIES. PROTEIN LADDER IN KDA. 62
FIGURE 42. DECONVOLUTED ESI-MS OF PURIFIED BLAC ENZYMES. **A)** RETRO ALDOL PRODUCT OF THE CARBAPENEMS FOLLOWING REACTION BETWEEN THE NUCLEOPHILIC SERINE AND THE CARBAPENEM. **B)** BLAC REFERENCE MASS = 30476 DA (30654 DA = BLAC + FORMYL-MET). **C)** BLAC + MEROPENEM = 30859 DA (31038 DA = BLAC + MEROPENEM + FORMYL-MET, 30816 DA = RETRO ALDOL PRODUCT OF MEROPENEM BOUND TO BLAC). **D)** BLAC + DORIPENEM = 30897 DA (31075 DA = BLAC + DORIPENEM + FORMYL-MET, 30853 DA = RETRO ALDOL PRODUCT OF DORIPENEM BOUND TO BLAC). 62
FIGURE 43. LC-MS ANALYSIS OF BLAC FOLLOWING 48 HOURS AFTER LABELLING WITH THE CARBAPENEMS. **A)** THE MEROPENEM-BLAC ARTIFICAL ENZYME. **B)** THE DORIPENEM-BLAC ARTIFICIAL ENZYME. THE MASS INDICATED AT 30476 IS THE CALCULATED MASS OF BLAC WITH NO CARBAPENEM ATTACHED. 64
FIGURE 44. THREE SECONDARY AMINE UNNATURAL AMINO ACIDS. **A)** D-PROLYL-L-LYSINE (DPK). **B)** L-PROLYL-L-LYSINE. **C)** THIAZOLIDINE LYSINE (THZK). 68
FIGURE 45. CRYSTAL STRUCTURE OF LMRR (**PDB:3F8F**). THE FOUR RESIDUES TARGETED FOR MUTATION ARE IN CYAN. 69
FIGURE 46. A) THE CRYSTAL STRUCTURE OF ECDHFR WITH NADPH BOUND IN THE ACTIVE SITE (**PDB: 1RA1**); RESIDUES TARGETED FOR MUTATION ARE IN YELLOW. **B)** CLOSE UP OF ECDHFR'S ACTIVE SITE. 70
FIGURE 47. SYNTHESIS OF THE UNNATURAL AMINO ACIDS DPK (*R*) (R = CH_2), LPK (*S*) (R = CH_2) AND THZK (*S*) (R = S). AN ACTIVATED ESTER IS FORMED BETWEEN THE CARBOXYLIC ACID OF THE PROLYL GROUP AND PENTAFLUOROPHENOL USING DIISOPROPYLCARBODIIMIDE AS THE COUPLING REAGENT. *N*A-BOC-L-LYSINE IS THEN COUPLED THROUGH A PEPTIDE BOND BY NUCLEOPHILIC SUBSTITUTION BETWEEN THE FREE AMINE AND THE ESTER OF THE ACTIVATED PROLYL-GROUP. THE DI-BOC PROTECTED DIPEPTIDE IS THEN DEPROTECTED USING TRIFLUORACETIC ACID. 71
FIGURE 48. SDS-PAGE ANALYSIS FOR THE INCORPORATION OF DPK, LPK AND THZK INTO THE REPORTER PROTEIN SFGFP (ASP150TAG). MASS OF SFGFP APPROX. 28 KDA. PROTEIN LADDER IN KDA. 72
FIGURE 49. SDS-PAGE ANALYSIS OF THE PURIFIED LMRR MUTANTS CONTAINING THE INDICATED UNNATURAL AMINO ACID AT THE SPECIFIED POSITION. AVERAGE MASS OF THE CALCULATED PROTEIN MONOMER = 14757 DA. PROTEIN LADDER IN KDA. 73
FIGURE 50. DECONVOLUTED ESI MASS SPECTRA OF WILD-TYPE LMRR AND EACH MUTANT, DETERMINED FROM THE PROTEIN MASS SPECTROMETRY CHARGED STATE. 74
FIGURE 51. A) NATURAL HYDRIDE DONOR NADH AND SYNTHETIC HYDRIDE DONOR BNAH. **B)** SELECTIVE REDUCTION OF CINNAMALDEHYDE BY BNAH. 75
FIGURE 52. SYNTHESIS OF THE HYDRIDE DONOR BENZYL-NICOTINAMIDE (BNAH) 75
FIGURE 53. A) SELECTIVE REDUCTION OF CINNAMALDEHYDE BY BNAH TO HYDRO-CINNAMALDEHYDE. **B)** PEAK AREA OF THE HYDRIDE TRANSFER PRODUCT HYDRO-CINNAMALDEHYDE. FOR EACH OF THE LMRR BASED ARTIFICIAL ENZYMES, THE WILD-TYPE LMRR, DPK AND THE SUBSTRATES ALONE UNDER THE SAME REACTION CONDITIONS. REACTION CONVERSION DETERMINED BY COMPARISON TO A STANDARD CALIBRATION CURVE FOR THE PRODUCT HYDRO-CINNAMALDEHYDE. EACH CHART ENTRY IS AN AVERAGE OF THREE REPEATS. ERROR BARS ARE A REPRESENTATION OF THE STANDARD DEVIATION. 76
FIGURE 54. SDS-PAGE ANALYSIS OF THE PURIFIED ECDHFR MUTANTS CONTAINING THE INDICATED UNNATURAL AMINO ACID AT THE SPECIFIED POSITION. AVERAGE MASS OF THE PROTEIN IS 19378 DA. PROTEIN LADDER IN KDA. 77
FIGURE 55. DECONVOLUTED ESI MASS SPECTRA OF WILD-TYPE ECDHFR AND EACH MUTANT, DETERMINED FROM THE PROTEIN MASS SPECTROMETRY CHARGED STATE. 78

FIGURE 56. A) SELECTIVE REDUCTION OF CINNAMALDEHYDE BY NADPH TO HYDRO-CINNAMALDEHYDE. **B)** PEAK AREA OF THE HYDRIDE TRANSFER PRODUCT HYDRO-CINNAMALDEHYDE. FOR EACH OF THE ECDHFR BASED ARTIFICIAL ENZYMES, THE WILD-TYPE ECDHFR, DPK AND THE SUBSTRATES ALONE UNDER THE SAME REACTION CONDITIONS. REACTION CONVERSION WAS DETERMINED BY COMPARISON TO A STANDARD CALIBRATION CURVE FOR THE PRODUCT HYDRO-CINNAMALDEHYDE. EACH CHART ENTRY IS AN AVERAGE OF THREE REPEATS. ERROR BARS ARE A REPRESENTATION OF THE STANDARD DEVIATION. 79

FIGURE 57. TIME COURSE CHART MEASURING THE REACTION OF LMRR-PHE93-DPK AS A CATALYST CONVERTING CINNAMALDEHYDE TO HYDRO-CINNAMALDEHYDE. 50 MM OF LMRR WAS INCUBATED WITH 1 MM OF CINNAMALDEHYDE AND 5 MM OF BNAH. THE REACTIONS WERE QUENCHED AT EACH TIME POINT WITH THE ADDITION OF DCM AND THE CONVERSION OF SUBSTRATE TO PRODUCT DETERMINED USING GC-FID. EACH TIME POINT WAS PERFORMED THREE TIMES AND THE MEAN REPORTED. ERROR BARS ARE A REPRESENTATION OF THE STANDARD DEVIATION. 80

FIGURE 58. KINETIC CHARACTERISATION OF LMRR PHE93DPK AND ECDHFR ALA7DPK. ASSAYS WERE PERFORMED USING 50 MM OF LMRR AND 20 MM OF ECDHFR. THE ASSAYS WERE PERFORMED USING SATURATING CONCENTRATIONS OF CINNAMALDEHYDE (1 MM) AND VARYING CONCENTRATIONS OF BNAH FOR LMRR AND NADPH FOR ECDHFR. THE RATE CONSTANTS FOR LMRR WERE FITTED TO A MICHAELIS-MENTEN CURVE AND VMAX, K_{CAT}, K_M AND K_{CAT}/K_M DETERMINED USING THE SOFTWARE PRIZM. THE RATE CONSTANTS FOR ECDHFR WERE FITTED TO AN ALLOSTERIC SIGMOIDAL CURVE RELEVANT TO THE HILL EQUATION USING THE SOFTWARE PRIZM. 81

FIGURE 59. SUBSTRATE SCOPE OF THE ARTIFICIAL ENZYMES LMRR PHE93DPK AND ECDHFR ALA7DPK. CONVERSION DETERMINED BY ^{1}H NMR BY COMPARISON OF THE INTEGRALS ARISING FROM THE ALDEHYDE PROTONS OF THE SUBSTRATES AND PRODUCTS AND NORMALISED PER MOL% OF ENZYME USED (LMRR = 10 MOL% AND ECDHFR = 1 MOL%) IN THE REACTION WITH RESPECT TO THE CARBONYL SUBSTRATE. REACTIONS WERE PERFORMED IN 50 MM PBS BUFFER (PH 7.0) FOR 48 HOURS AT 25 °C WITH CONSTANT AGITATION. EACH ENTRY IS AN AVERAGE OF THREE REPEATS. 82

FIGURE 60. A) COFACTOR RECYCLING SCHEME EMPLOYED TO REGENERATE NADPH *IN SITU*, DURING THE IMINIUM ION CATALYSED HYDRIDE TRANSFER REACTION CATALYSED ECDHFR ALA7DPK. **B)** CONVERSION OF CINNAMALDEHYDE TO HYDRO-CINNAMALDEHYDE WITH AND WITHOUT THE COFACTOR RECYCLING SCHEME. 83

FIGURE 61. GC-MS CHROMATOGRAMS PRODUCED FROM THE EXTRACTION OF THE ORGANIC COMPOUNDS FROM THE WHOLE CELL LYSATE CATALYSIS USING ECDHFR ALA7DPK. THE DESIRED PRODUCT HYDRO-CINNAMALDEHYDE HAS A RETENTION TIME OF 7.21 MINUTES. BL21 (DE3) CONTROL CELLS EXTRACTED IDENTIFIED PRODUCT COMPOUNDS ARE CINNAMYL ALCOHOL AT RETENTION TIME 8.03 MINUTES AND HYDRO-CINNAMYL ALCOHOL AT 8.80 MINUTES. THE MAJOR EXTRACTABLE COMPOUND IS CINNAMALDEHYDE AT RETENTION TIME 8.37 MINUTES. THE MAJOR EXTRACTABLE PRODUCT IN THE CELLS EXPRESSING ECDHFR ALA7DPK WAS THE SUBSTRATE CINNAMALDEHYDE. 85

FIGURE 62. LC-MS ANALYSIS OF THE CHYMOTRYPSIN DIGEST OF LMRR PHE93DPK. EXTRACTION OF THE PREDICTED PEPTIDE MASS THAT CONTAINS THE UNNATURAL AMINO FROM THE PEPTIDE DIGESTS RESULTED IN EXACT MASS IDENTIFICATION OF THE PREDICTED PEPTIDE MASS WITH DPK MODIFIED WITH CINNAMALDEHYDE. 86

FIGURE 63. LC-MS ANALYSIS OF THE CHYMOTRYPSIN DIGEST OF ECDHFR ALA7DPK. EXTRACTION OF THE PREDICTED PEPTIDE MASS THAT CONTAINS THE UNNATURAL AMINO FROM THE PEPTIDE DIGESTS RESULTED IN EXACT MASS IDENTIFICATION OF THE PREDICTED PEPTIDE MASS WITH DPK MODIFIED WITH CINNAMALDEHYDE. 87

FIGURE 64. NATURAL MECHANISM OF ECDHFR. THE SATURATED IMINE OF THE PTERIN RING OF DIHYDROFOLATE IS REDUCED TO FORM TETRAHYDROFOLATE USING THE PRO-*R* HYDRIDE OF NADPH...88

FIGURE 65. A) HR-MS ANALYSIS OF THE ENZYMATICALLY SYNTHESISED NADPD. CALCULATE M/Z FOR NADPD = 747.1052 **B)** THE ENZYMATIC REACTION USED TO INSTALL A DEUTERIUM ATOM AT THE PRO-*R* HYDRIDE OF NADPH. 89

FIGURE 66. GC-MS ANALYSIS OF THE PRODUCTS OF THE REACTION BETWEEN CINNAMALDEHYDE AND NADPH AND NADPD. INCORPORATION OF A DEUTERIUM ATOM RESULTS IN A MASS INCREASE OF 1 DA. FRAGMENT ANALYSIS OF THE RADICAL FORMED UPON ELECTRON IMPACT IONISATION SUGGESTS THAT THE DEUTERIUM IS TRANSFERRED TO THE GAMMA CARBON OF THE SUBSTRATE. 90

FIGURE 67. SELECTIVE TRANSFER OF THE PRO-*R* HYDRIDE OF NADPH TO THE GAMMA-CARBON OF CINNAMALDEHYDE BY ECDHFR ALA7DPK. 90

FIGURE 68. PROPOSAL OF THE MECHANISM OF SELECTIVE REDUCTION OF A,B-UNSATURATED CARBONYL COMPOUNDS BY ECDHFR ALA7DPK USING NADPH AS THE COFACTOR**. A)** FORMATION OF THE IMINIUM ION BETWEEN THE CARBONYL COMPOUND AND THE SECONDARY AMINE UNNATURAL AMINO ACID. **B)** TRANSFER OF THE PRO-*R* HYDRIDE OF NADPH TO THE GAMMA-CARBON OF THE A,B-UNSATURATED ALDEHYDE AND FORMATION ON AN ENAMINE. **C)** PROTONATION OF THE ENAMINE AND HYDROLYSIS OF THE RESULTING IMINIUM ION PRODUCING THE SATURATED PRODUCT AND OXIDISED $NADP^+$. 91
FIGURE 69. RATE CURVES DISPLAYING THE OXIDATION OF NADPH (BLUE) AND NADPD (RED) IN THE REACTION CATALYSED BY ECDHFR ALA7DPK. UV ABSORPTION AT 340 NM WAS USED TO MONITOR PRODUCT TURNOVER. TABLE DISPLAYING THE DETERMINATION OF THE KINETIC ISOTOPE EFFECT (KIE).....93
FIGURE 70. A SELECTION OF UNNATURAL AMINO ACIDS THAT HAVE BEEN SUCCESSFULLY INCORPORATED INTO A PROTEIN USING GENETIC CODE EXPANSION. **A)** BOC LYSINE OFTEN USED FOR INITIAL STUDIES. **B)** CYCLOPROPENE LYSINE AND **C)** THIAZOLIDINE LYSINE USED FOR SITE-SELECTIVE BIOCONJUGATION **D)** D-PROLYL LYSINE AN ANALOGUE OF PYRROLYSINE AND CATALYTICALLY COMPETENT. **E)** ACETYL LYSINE**, F)** PROPIONYL LYSINE AND **G)** THIOACETYL LYSINE USED FOR MIMICKING POST-TRANSLATIONAL MODIFICATIONS. **H)** PHOTOCAGED CYSTEINE USED IN PROTEIN ACTIVATION AND CONTROL. 98
FIGURE 71. A) CHARGING OF TRNA WITH AN AMINO ACID BY TRNA SYNTHETASE. **B)** RIBOSOMAL TRANSLATION OF MRNA INTO PROTEIN FROM THE N-TERMINUS TO C-TERMINUS. **C)** THE FULLY FORMED POLYPEPTIDE RELEASED FROM THE RIBOSOME FOLLOWING TERMINATION.....100
FIGURE 72. THE GENERAL MECHANISM OF TYPE II AMINO-ACYL TRNA SYNTHETASES. **A)** NUCLEOPHILIC ATTACK OF THE DEPROTONATED CARBOXYL GROUP TO THE A-PHOSPHATE FORMS A PENTAVALENT INTERMEDIATE, THIS COLLAPSES TO FORM AN ADENYLATE ESTER AND DIPHOSPHATE. **B)** THE AMINO ACID IS THEN CHARGED ONTO THE TRNA BY NUCLEOPHILIC SUBSTITUTION OF THE ACTIVATED ESTER. 101
FIGURE 73. A) CHARGING OF AN ORTHOGONAL TRNA WITH A NON-CANONICAL UNNATURAL AMINO ACID BY AN ORTHOGONAL TRNA SYNTHETASE. B**)** RIBOSOMAL TRANSLATION OF THE UNNATURAL AMINO ACID INTO THE PROTEIN BACKBONE USING A NON-CODING CODON. **C)** THE FULLY TRANSLATED POLYPEPTIDE WITH THE UNNATURAL AMINO ACID SITE-SPECIFICALLY INCORPORATED INTO TO THE PROTEIN. 102
FIGURE 74. A) CRYSTAL STRUCTURE OF MMPYLRS (**PDB: 6ABM**), C-TERMINAL DOMAIN IN RED AND N-TERMINAL DOMAIN IN BLUE. **B)** SEQUENCE ALIGNMENT OF MMPYLRS AND MBPYLRS N-TERMINUS. **C)** THE STRUCTURE OF THE 22^{ND} PROTEINOGENIC AMINO ACID PYRROLYSINE. 104
FIGURE 75. A) POST-TRANSLATIONAL CHROMOPHORE MATURATION OF GREEN FLUORESCENT PROTEIN PROPOSED MECHANISM.[37-39] **B)** TRANSLATION OF THE NATIVE PROTEIN PRODUCES GREEN FLUORESCENCE, A TAG MUTATION RESULTS IN PREMATURE TERMINATION RESULTING IN NO FLUORESCENCE, ADDITION OF AN UNNATURAL AMINO ACID AND THE CORRESPONDING MACHINERY RESULTS IN FLUORESCENCE CORRESPONDING TO THE INCORPORATION EFFICIENCY. **C)** CRYSTAL STRUCTURE OF SFGFP WITH THE ASN150 HIGHLIGHTED IN RED (**PDB: 4LQT**).....105
FIGURE 76. OUTLINE OF GFP REPORTER ASSAY TO DETERMINE INCORPORATION EFFICIENCY OF THE TRNA SYNTHETASES. CO-TRANSFORMATION OF THE SYNTHETASE/TRNA PLASMID WITH THE REPORTER PLASMID GFP. CULTURE TO AN OD_{600} OF 0.6. INDUCE GFP EXPRESSION WITH IPTG (0.5 MM) AND ADD THE UNNATURAL AMINO ACID (5 MM). SAMPLE THE CULTURE. NORMALISE THE OD_{600} TO 1.0 AND MEASURE THE RELATIVE FLUORESCENCE WITH A PLATE READER. 106
FIGURE 77. COMPARISON OF THE INCORPORATION EFFICIENCY OF THE FIVE TRNA SYNTHETASES WITH RESPECT TO THEIR UNNATURAL AMINO ACIDS. REPORTED AS NORMALISED FLUORESCENCE RELATIVE TO THE MOST INTENSE SIGNAL MEASURED FROM THE GFP PLATE READER ASSAY. MEAN OF THREE BIOLOGICAL REPLICATES ± STANDARD DEVIATION (S.D.). RAW DATA TABLE AND STATISTICAL P VALUES FOR EACH COMPARISON DETERMINED FROM TWO TAILED T TEST ASSUMING UNEQUAL VARIANCE. SIGNIFICANCE LEVEL >0.05. CULTURES PERFORMED AT 20 °C FOR 18 HOURS FOLLOWING REPORTER GENE EXPRESSION RESULTING FROM THE ADDITION OF 0.5 MM OF IPTG AND 5 MM OF THE RELEVANT UNNATURAL AMINO ACID. 107
FIGURE 78. COMPARISON OF THE INCORPORATION EFFICIENCY OF MBPYLRS (LEFT) WITH TWO NEW SUBSTRATES CYPK AND DPK AND ACKRS (RIGHT) WITH THE SUBSTRATES PK AND TACK. REPORTED AS NORMALISED FLUORESCENCE RELATIVE TO THE MOST INTENSE SIGNAL MEASURED FROM THE GFP PLATE READER ASSAY. MEAN OF THREE BIOLOGICAL REPLICATES ± STANDARD DEVIATION (S.D.). RAW DATA TABLE AND STATISTICAL P VALUES FOR EACH COMPARISON DETERMINED FROM TWO TAILED T TEST ASSUMING UNEQUAL VARIANCE. SIGNIFICANCE LEVEL >0.05. CULTURES PERFORMED AT 20 °C FOR 18 HOURS FOLLOWING REPORTER GENE EXPRESSION RESULTING FROM THE ADDITION OF 0.5 MM OF IPTG AND 5 MM OF THE RELEVANT UNNATURAL AMINO ACID. 108

FIGURE 79. COMPARISON OF THE INCORPORATION EFFICIENCY OF MBPYLRS WITH BOCK AND ACKRS WITH RESPECT TO EXPRESSION TEMPERATURES 20 °C AND 37 °C. REPORTED AS NORMALISED FLUORESCENCE RELATIVE TO THE MOST INTENSE SIGNAL MEASURED FROM THE GFP PLATE READER ASSAY. MEAN OF THREE BIOLOGICAL REPEATS. MEAN OF THREE BIOLOGICAL REPLICATES ± STANDARD DEVIATION (S.D.). RAW DATA TABLE AND STATISTICAL P VALUES FOR EACH COMPARISON DETERMINED FROM TWO TAILED T TEST ASSUMING UNEQUAL VARIANCE. SIGNIFICANCE LEVEL >0.05. CULTURES PERFORMED AT 20 °C OR 37 °C FOR 18 HOURS FOLLOWING REPORTER GENE EXPRESSION RESULTING FROM THE ADDITION OF 0.5 MM OF IPTG AND 5 MM OF THE RELEVANT UNNATURAL AMINO ACID. 109
FIGURE 80. SDS-AGE ANALYSIS OF LMRR PHE93DPK FOLLOWING SIZE EXCLUSION CHROMATOGRAPHY. MASS OF LMRR PHE93DPK MONOMER APPROX. 14.5 KDA. PROTEIN LADDER IN KDA. 128
FIGURE 81. SDS-AGE ANALYSIS OF LMRR PHE93DPK FOLLOWING SIZE EXCLUSION CHROMATOGRAPHY. MASS OF LMRR PHE93DPK MONOMER APPROX. 19.5 KDA. PROTEIN LADDER IN KDA. 130
FIGURE 82. SDS-PAGE ANALYSIS FOR THE EXPRESSION AND PURIFICATION OF TBADH. BI = BEFORE INDUCTION OF GENE EXPRESSION BY IPTG. AI = AFTER INDUCTION OF GENE EXPRESSION BY IPTG. SN = SUPERNATANT FOLLOWING LYSIS BY SONICATION. IB = INCLUSION BODIES FOLLOWING LYSIS BY SONICATION. SF = SOLUBLE FRACTION FOLLOWING THERMAL PRECIPITATION. IF = INSOLUBLE FRACTION FOLLOWING THERMAL PRECIPITATION. IX1 = ION EXCHANGE ELUTION 1. IX2 = ION EXCHANGE ELUTION 2. IX3 = ION EXCHANGE ELUTION 3. MASS OF TBADH APPROX. 37 KDA. PROTEIN LADDER IN KDA. 132
FIGURE 83. GC-FID HYDRO-CINNAMALDEHYDE STANDARD CALIBRATION CURVE FOR THE DETERMINATION OF PRODUCT CONCENTRATION. 142