



Altering enzyme activities using chemical modification

Claire Louise Windle

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

September 2015

The candidate confirms that the submitted work is her own and that the appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgement.

© 2015 The University of Leeds and Claire Louise Windle

Jointly Authored Publications

Throughout this thesis the work directly attributable to the candidate is as follows:

- i) Literature research and compilation of the manuscript stated above.
- ii) The candidate performed all the experimental work and data analysis unless otherwise stated.

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Details of jointly authored publications and the contributions of other authors to these manuscripts:

Chapter 3 contains work from the following publication

Timms, N., Windle, C. L., Polyakova, A., Ault, J. R., Trinh, C. H., Pearson, A. R., Nelson, A. & Berry, A. (2013) Structural insights into the recovery of aldolase activity in n-acetylneuraminic acid lyase by replacement of the catalytically active lysine with gamma-thialysine by using a chemical mutagenesis strategy. *Chembiochem*, 14, 474-81

In this work, I carried out the chemical modification of the wild-type *S. aureus* NAL (*sa*NAL) into K165- γ -thialysine *sa*NAL and performed analysis of mass spectrometry data, size exclusion chromatography, kinetic assays, pH profiles and circular dichroism experiments, I also carried out crystallisation of the K165C *sa*NAL, data processing for K165C *sa*NAL, K165- γ -thialysine apo *sa*NAL and K165- γ -thialysine *sa*NAL in complex with pyruvate, model building and refinement for wild-type *sa*NAL, wild-type *sa*NAL in complex with pyruvate, K165C *sa*NAL, K165- γ -thialysine apo *sa*NAL and K165- γ -thialysine *sa*NAL in complex with pyruvate. Dr Nicole Timms developed and optimised the system for the chemical modification procedure for use in *E. coli* and *S. aureus* NAL, performed chemical modification of *sa*NAL and analysed mass

spectrometry data, carried out kinetic assays and refolding assays of the modified protein. Anna Polyakova performed optimisation of crystallisation conditions for the wild-type saNAL, crystallisation of wild-type saNAL, data collection and data processing for the wild-type apo saNAL and wild-type in complex with pyruvate saNAL. Dr James Ault performed the mass spectrometry of the chemical modification. Dr Chi Trinh supervised and aided with all X-ray crystallographic aspects of the project; crystallisation, data collection, data processing and refinement. Dr Arwen Pearson provided scientific discussion of the X-ray crystallographic aspects. Prof. Adam Nelson and Prof. Alan Berry provided the project concept and aided with scientific discussion, data interpretation and manuscript preparation.

Acknowledgements

Firstly I would like to thank my supervisors, Professor Alan Berry, Professor Adam Nelson and Professor Arwen Pearson for all their encouragement and guidance. Alan has been a great supervisor; he has always been enthusiastic about my project and provided continuous support. Thank you to Adam for his brilliant supervision and for explaining organic chemistry to me, and thank you to Arwen for the motivation and help with crystallography!

I would like to thank Dr James Ault for all the mass spectrometry work that he has carried out for me, and for always being friendly and having the time for a good chat. A huge thanks has to go to Dr Chi Trinh for all his endless help with crystallography, you definitely made data collection more bearable!

Over the last four years the most important thing I have learnt, is that the best way to get through a PhD is to make wonderful friends who will always make you laugh and take you for a drink when science seems to hate you. To the Berrys, Radfords and Mass spec you've all made the Astbury Centre an amazing place to work. Thanks to Sasha, Rachel, Sophie, Jan and Lydia for all the cocktails and dancing, and thanks to all the Monday night people Paul, Patrick, Tom and Rob for providing much needed distraction from thesis writing.

Finally I'd like to thank my family, my Mum and Dad, Sister and Grandparents without you I would not be the person that I am today. Thank you for always supporting the decisions that I make and encouraging me to do what makes me happy. I hope I have made you proud!

Abstract

In Nature there are twenty proteogenic amino acid 'building blocks', from which proteins and enzymes are constructed. These proteogenic amino acids confer activity to enzymes; however there are many instances where the chemistries provided by these 'building blocks' are expanded upon. Nature recruits an array of cofactors, post-translational modifications and post-translationally generated cofactors, all which help to provide function or activity. Until recently the protein engineer was restricted to the use of the twenty proteogenic amino acids, and so access to this increased chemical diversity was highly challenging.

In this thesis, chemical modification has been used to insert a variety of non-canonical amino acids (ncAAs) throughout the active site of the enzyme *N*-acetylneuraminic acid lyase (NAL). This modification method incorporates ncAAs site-specifically into a protein, via a dehydroalanine intermediate and conjugate addition with a thiol compound. Initial work using this method replaced the catalytic lysine at position 165 with the non-canonical analogue γ -thialysine. It was possible to obtain homogeneously modified protein in high yields for detailed kinetic and X-ray crystallographic studies, and therefore possible to elucidate the catalytic and structural consequences of this modification.

The work to replace Lys165 with a non-canonical analogue provided a starting point to expand the incorporation of ncAAs into NAL. A total of thirteen different non-canonical side chains were incorporated, individually, at thirteen different positions within the active site of NAL. These modified enzymes were then screened for activity with ten different substrates to determine the effects of ncAA incorporation. It was seen that the ncAAs were well tolerated by the enzyme, as active modified enzymes were produced. By incorporating ncAAs it was possible to alter the substrate specificity of the enzyme. The modified enzyme F190Dpc, containing a dihydroxypropyl cysteine side chain, was found to have an increased activity with an altered substrate, erythrose. This activity was higher than the wild-type enzyme with both the altered substrate and the wild-type substrate, and the non-canonical Dpc side chain outperformed any of the proteogenic amino acids when inserted at the same position in the protein, for the substrate erythrose.

This research begins to explore the possibilities of what may be achieved by use of ncAAs. Facile incorporation of ncAAs will allow protein engineers to take inspiration from Nature and expand the chemistries provided by the proteogenic amino acids, hopefully to engineer novel activities or catalysis.

Contents

Acknowledgements	iii
Abstract	iv
Contents	v
List of Figures	ix
Abbreviations	xiii
Chapter 1 Introduction	2
1.1 Amino acid library extension	5
1.1.1 21 st and 22 nd amino acids	5
1.1.2 Cofactors and post-translational modifications	7
1.1.3 Post-translationally generated cofactors	10
1.2 Incorporation of non-canonical amino acids.....	15
1.2.1 Auxotrophic incorporation	16
1.2.2 tRNA suppressor technology	18
1.2.3 Chemical modification of proteins	21
1.3 Catalysis with non-canonical amino acids.....	25
1.4 Aldolases.....	29
1.4.1 <i>N</i> -Acetylneuraminic acid lyase	34
1.4.2 Protein engineering of <i>N</i> -acetylneuraminic acid lyase.....	37
1.5 Thesis aims and objectives	40
1.5.1 To determine structural and functional effects of replacing the catalytic lysine at position 165 with the ncAA γ -thialysine.	40
1.5.2 Develop methods for high-throughput incorporation of ncAAs by chemical modification, and screening of modified variants.	40
1.5.3 To characterise modified enzymes with altered substrate specificities brought about by incorporation of a non-canonical side chain.	41
Chapter 2 Materials and Methods	42
2.1 Materials.....	42
2.1.1 Technical equipment	42
2.1.2 Chemicals.....	43
2.1.3 Media.....	45
2.2 General Methods.....	45

2.2.1	Centrifugation.....	45
2.2.2	pH measurements	45
2.2.3	Aseptic technique	46
2.2.4	Antibiotic stocks	46
2.2.5	Culture growth.....	46
2.2.6	Glycerol stocks.....	46
2.3	DNA Methods.....	47
2.3.1	Bacterial strains and plasmids	47
2.3.2	Plasmid purification.....	47
2.3.3	Agarose gel electrophoresis	47
2.3.4	DNA quantification	48
2.3.5	DNA sequencing	48
2.3.6	Site directed Mutagenesis	48
2.3.7	Transformation	48
2.4	Protein methods.....	48
2.4.1	Affinity chromatography purification of His ₆ -tagged proteins.....	48
2.4.2	Sodium dodecyl sulphate polyacrylamide gel electrophoresis	49
2.4.3	Protein dialysis.....	51
2.4.4	Protein concentration determination	51
2.4.5	Increasing protein concentration	51
2.4.6	Lyophilisation.....	51
2.4.7	Sample preparation for ESI mass spectrometry.....	52
2.4.8	ESI mass spectrometry of proteins.....	52
2.4.9	Conversion of cysteine to dehydroalanine by 2, 5-dibromohexan-1, 6-diamide (diBr)	52
2.4.10	Conversion of dehydroalanine to non-canonical amino acid side chains using sulphur nucleophiles.....	53
2.4.11	Refolding of non-canonical amino acid containing enzymes	57
2.4.12	Size exclusion chromatography	57
2.4.13	Lactate dehydrogenase coupled enzyme kinetic assay	57
2.4.14	pH profiles.....	58
2.4.15	Far UV circular dichroism spectroscopy	59
2.4.16	Thiobarbituric acid assay (TBA assay).....	60
2.4.17	Purification of DOH by anion exchange	61
2.4.18	Proton NMR	62

2.4.19	ESI Mass spectrometry of small molecules.....	62
2.5	Crystallographic methods.....	62
2.5.1	Protein crystallisation.....	62
2.5.2	Cryoprotection and production of enzyme-pyruvate complexes	62
2.5.3	Data collection.....	63
2.5.4	Data processing	63
2.5.5	Refinement	63
Chapter 3 Production and Characterisation of K165-γ-thialysine saNAL		64
3.1	Incorporation of γ -thialysine into saNAL.....	66
3.1.1	Characterisation of wild-type and K165C saNAL.....	67
3.1.2	Conversion of K165C saNAL into K165dha saNAL.....	70
3.1.3	Modification of K165dha saNAL into K165- γ -thialysine saNAL.....	71
3.1.4	Large scale conversion of K165C saNAL into K165- γ -thialysine saNAL ...	72
3.2	Characterisation of K165- γ -thialysine saNAL	73
3.2.1	Far UV circular dichroism and initial kinetic assays of K165- γ -thialysine	73
3.2.2	Kinetic comparison of wild-type saNAL and K165- γ -thialysine saNAL....	77
3.3	Crystallographic studies of K165- γ -thialysine saNAL	78
3.3.1	Crystallisation	79
3.3.2	Data collection and processing.....	80
3.3.3	Crystal structures.....	86
3.4	Further kinetic studies of K165- γ -thialysine saNAL.....	91
3.5	Summary	95
Chapter 4 Production and Screening of Modified Enzymes for Altered Activities		97
4.1	Selection of residues for modification	97
4.2	Production of cysteine mutants	99
4.3	Modifications to generate a variety of ncAAs.....	102
4.3.1	Conversion of cysteine variants to dha	102
4.3.2	Selection of ncAA side chains	106
4.3.3	Optimisation of Michael addition of thiols.....	107
4.4	Screening modified enzymes for activity	114
4.4.1	TBA assay for screening.....	115
4.4.2	Screening modified enzymes for activities.....	116
4.5	Investigation of hits from screening.....	119
4.5.1	F172Bc activity with GlcNAc and pyruvate.....	120

4.5.2	F190Dpc activity with erythrose and pyruvate	121
4.5.3	NMR analysis of DOH.....	125
4.6	Summary	133
Chapter 5	Characterisation of F190Dpc <i>sa</i>NAL.....	134
5.1	Kinetic characterisation of F190Dpc <i>sa</i> NAL	134
5.1.1	Aldol condensation kinetics.....	135
5.1.2	Kinetic comparison of wild-type <i>sa</i> NAL and F190Dpc <i>sa</i> NAL.....	137
5.2	F190X saturation library	141
5.2.1	Production of F190X saturation library	141
5.2.2	Kinetic comparison of the F190X library and F190Dpc <i>sa</i> NAL	142
5.3	Comparison of F190Dpc to directed evolution studies.....	145
5.4	Crystallographic studies of F190Dpc <i>sa</i> NAL	148
5.4.1	Crystallisation of F190Dpc <i>sa</i> NAL	148
5.4.2	Data collection, processing and refinement.....	149
5.4.3	Crystallographic comparison of wild-type <i>sa</i> NAL and F190Dpc <i>sa</i> NAL.	152
5.5	Modelling and energy minimisations	159
5.5.1	Modelling of DOH into the active site of F190Dpc <i>sa</i> NAL and wild-type <i>sa</i> NAL.....	159
5.5.2	Comparison of F190Dpc <i>sa</i> NAL and wild-type <i>sa</i> NAL models.....	161
5.6	Summary	165
Chapter 6	Summary, Future Perspectives and Conclusions	167
6.1	Summary	167
6.2	Future perspectives.....	169
6.2.1	Further alteration of substrate specificity.....	169
6.2.2	Exploitation of interactions not found in the proteogenic amino acids	171
6.2.3	Alteration of catalysis	171
6.3	Concluding remarks.....	173
Chapter 7	Appendix.....	175

List of Figures

Figure 1.1 The four nucleoside bases found in DNA and the twenty naturally occurring proteogenic amino acids.	3
Figure 1.2 Examples of some post-translationally generated cofactors and post-translational modifications (PTMs).	4
Figure 1.3 Structures of the 21 st and 22 nd amino acids selenocysteine and pyrrolysine.	6
Figure 1.4 Cofactors and mechanism of the pyruvate dehydrogenase complex (PDH).	9
Figure 1.5 Panel A shows the mechanism of backbone cleavage between Ser81 and Ser81 to form the pyruvoyl cofactor in histidine decarboxylase.	11
Figure 1.6 Formylglycine (FGly) structure and mechanism of action.	13
Figure 1.7 Structures of the post-translationally generated quinone cofactors.	14
Figure 1.8 Structure of the post-translationally generated cofactor 5-methylene-3,5-dihydroimidazol-4-one (MIO).	15
Figure 1.9 Non-canonical amino acids used in global replacement studies of a lipase. ...	17
Figure 1.10 Schematic of nonsense suppression.	19
Figure 1.11 Conversion of active site serine to cysteine in subtilisin.	22
Figure 1.12 Methyl lysine and acetyl lysine analogues produced from alkylating a cysteine residue.	24
Figure 1.13 Bis-alkylation-elimination mechanism for the conversion of a cysteine residue into a dehydroalanine residue.	25
Figure 1.14 Tyrosine analogues 7-hydroxycoumarinyl and 7-methylcoumarinyl inserted at Tyr309 in a bacterial phosphotriesterase.	26
Figure 1.15 Reaction of aspartate aminotransferase.	27
Figure 1.16 Lysine and glutamate non-canonical analogues inserted by alkylation of a cysteine.	28
Figure 1.17 Generalised Class I and Class II aldolase mechanisms.	30
Figure 1.18 Classification of aldolases by their donor substrate.	32
Figure 1.19 DERA catalysed reaction.	33
Figure 1.20 Wild-type <i>N</i> -acetylneuraminic acid lyase (NAL) catalysed aldol reaction.	34
Figure 1.21 Proposed mechanism of <i>N</i> -acetylneuraminic acid lyase.	35
Figure 1.22 Crystal structure of the <i>E. coli</i> NAL.	36
Figure 1.23 Active site of the Y137A <i>E. coli</i> NAL variant in complex with the product Neu5Ac.	37
Figure 1.24 Protein engineering studies on the <i>E. coli</i> NAL.	39
Figure 2.1 Schematic of lactate dehydrogenase (LDH) coupled enzyme assay.	58

Figure 3.1. Schematic of the chemical modification method used to convert a cysteine residue into a non-canonical amino acid (ncAA).	65
Figure 3.2. Comparison of the direct alkylation method and the bis-alkylation-elimination followed by a Michael addition method.....	67
Figure 3.3. Purification and ESI mass spectra of wild-type and K165C <i>sa</i> NAL.	69
Figure 3.4. ESI mass spectra of K165dha <i>sa</i> NAL.	71
Figure 3.5 ESI mass spectrum of K165- γ -thialysine <i>sa</i> NAL.	72
Figure 3.6. Schematic representation of chemical modification producing both D and L forms of the ncAA γ -thialysine.....	75
Figure 3.7. Size exclusion chromatography of K165- γ -thialysine.....	76
Figure 3.8. Far UV circular dichroism of wild-type and K165- γ -thialysine.....	77
Figure 3.9. K165- γ -thialysine <i>sa</i> NAL crystal formed from hanging drop vapour diffusion	79
Figure 3.10 Model building of the γ -thialysine side chain into chain C of the K165- γ -thialysine apo crystal structure.	83
Figure 3.11. Electron density maps for the side chains at position 165.	85
Figure 3.12. Structural comparison of wild-type and K165- γ -thialysine <i>sa</i> NAL.	87
Figure 3.13. Comparison of C α -NZ distances in lysine and γ -thialysine at position 165 in the wild-type and K165- γ -thialysine <i>sa</i> NAL structures.....	90
Figure 3.14. Comparison of the pyruvate binding site in the wild-type and K165- γ -thialysine crystal structures.	91
Figure 3.15. pH curves for wild-type and K165- γ -thialysine <i>sa</i> NAL.....	93
Figure 3.16. Schematic of pH titration curves.	95
Figure 4.1. Image of the wild-type <i>S. aureus</i> NAL active site	98
Figure 4.2 ESI mass spectra of cysteine variants.	101
Figure 4.3. ESI mass spectra, showing various modifications of cysteine variants into dehydroalanine.....	103
Figure 4.4. ESI mass spectra of I251C modifications to I251dha.....	105
Figure 4.5a ESI mass spectra of modifications of F190dha with various thiols using conditions stated in Table 4.3.....	111
Figure 4.6 ESI mass spectra of modifications of F190dha with various thiols using conditions stated in Table 4.3.....	112
Figure 4.7. Aldehydes used for screening modified enzymes for activity.	114
Figure 4.8. Schematic of the thiobarbituric acid (TBA) assay	115
Figure 4.9. Image of screening carried out at position 190 and analysed by TBA assay..	117
Figure 4.10. Heat map of absorbance data, at 550 nm, from screening carried out by TBA assay at position 190.....	118

Figure 4.11 Images of screening carried out at positions S208 and F172 which have been analysed by TBA assay.	119
Figure 4.12. Modified enzyme F172Bc showing an increased activity with GlcNAc and pyruvate, over that of the wild-type.	120
Figure 4.13. ESI negative ion mode mass spectrum of the product formed from reaction of GlcNAc and pyruvate with F172Bc.	121
Figure 4.14. Modified enzymes F190Hpc, F190Hbc and F190Dpc showing increased activity with erythrose and pyruvate, over that of the wild-type.	122
Figure 4.15. ESI mass spectrum showing correctly modified F190Dpc.	123
Figure 4.16. ESI negative ion mode mass spectrum of the product formed from reaction of erythrose and pyruvate with F190Dpc.	124
Figure 4.17. Anion exchange of 3-deoxy 2-heptulosonic acid (DOH).	125
Figure 4.18. Proton NMR spectrum of DOH purified by anion exchange.	126
Figure 4.19 Aldol condensation of pyruvate and D-erythrose to form deoxyheptulosonic acid.	127
Figure 4.20. Proton COSY and TOCSY NMR spectra of the product DOH.	128
Figure 4.21 Approximate coupling constants in pyranose rings.	129
Figure 5.1 Aldol condensation reactions of wild-type and F190Dpc <i>sa</i> NAL.	134
Figure 5.2 Time course of the aldol condensation between erythrose and pyruvate catalysed by F190Dpc and wild-type <i>sa</i> NAL.	136
Figure 5.3. Rate equations for use in kinetic analysis.	137
Figure 5.4 Kinetic comparison of F190Dpc and wild-type <i>sa</i> NAL with the substrates erythrose and pyruvate, and the substrates ManNAc and pyruvate.	139
Figure 5.5 Equation 5 used to calculate the specificity switch for F190Dpc <i>sa</i> NAL with the substrates erythrose and pyruvate.	141
Figure 5.6. Sequencing data for the 18 F190 mutants. The 18 different proteogenic amino can be seen to have been inserted at position 190.	142
Figure 5.7 Kinetic comparison of the F190X saturation library and F190Dpc for the aldol condensation of erythrose and pyruvate.	144
Figure 5.8. Sequencing and mass spectrometry of the I251R <i>sa</i> NAL variant.	146
Figure 5.9 Aldol condensation kinetics, for the condensation of erythrose and pyruvate, carried out with the I251R <i>sa</i> NAL variant.	147
Figure 5.10. Image of F190Dpc crystal mounted in a loop during data collection.	149
Figure 5.11 Structure of the dihydroxypropylcysteine (Dpc) side and electron density for the Dpc side chain found at position 190 in the F190Dpc crystal structure.	152
Figure 5.12 Crystal structures of F190Dpc and wild-type <i>sa</i> NAL both in complex with pyruvate.	153

Figure 5.13 Panel A shows a comparison of the active site of wild-type and F190Dpc crystal structures in complex with pyruvate.	155
Figure 5.14 Overlay of chain B of F190Dpc crystal structure in complex with pyruvate and chain B of the wild-type <i>sa</i> NAL crystal structure in complex with pyruvate. .	156
Figure 5.15 Distance matrix of the wild-type and F190Dpc crystal structures.....	157
Figure 5.16 F190Dpc and wild-type crystal structures, in complex with pyruvate, displayed as surface representations to show the differences in shape of the active sites.....	158
Figure 5.17 Aldol condensation of D-erythrose and pyruvate to form the product (4 <i>R</i> , 5 <i>S</i> , 6 <i>R</i>) DOH or (4 <i>S</i> , 5 <i>S</i> , 6 <i>R</i>) DOH.....	160
Figure 5.18 Active site models of F190Dpc and wild-type with the (4 <i>R</i>) and (4 <i>S</i>) DOH product.....	162
Figure 5.19 Overlays of the F190Dpc and wild-type models.	163
Figure 5.20 Comparison of the active sites of the F190Dpc and wild-type models in surface representation.	164
Figure 6.1 Wild-type and modified F190Dpc <i>sa</i> NAL structures and reactions.....	168
Figure 6.2 Potential ketone donor compounds for <i>sa</i> NAL.	170
Figure 6.3 Interaction of a thiourea containing <i>nc</i> AA with a carboxylic acid containing substrate.	171
Figure 6.4 Schematic of iminium ion catalysis.....	173
Figure 7.1 Raw data for the aldol condensation of erythrose and pyruvate to form deoxyheptulosonic acid determined by TBA assay	176
Figure 7.2 Standard curve of Neu5Ac.	176

Abbreviations

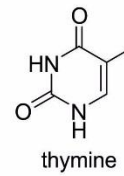
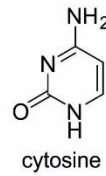
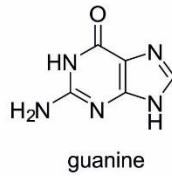
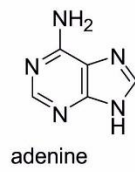
Å	angstrom (10^{-10} m)
A ₂₈₀	absorbance at 280 nm
A ₃₄₀	absorbance at 340 nm
BSA	bovine serum albumin
diBr	2,5-dibromohexan-1,6-diamide
DMF	dimethylformamide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
Dpc	dihydroxypropyl cysteine
<i>E. coli</i>	<i>Escherichia coli</i>
ecNAL	<i>N</i> -acetylneuraminic acid lyase from <i>E. coli</i>
EDTA	ethylenediaminetetraacetic acid
ESI-MS	electrospray ionisation mass spectrometry
far-UV CD	far-ultraviolet circular dichroism
<i>H. Influenza</i>	<i>Haemophilus influenzae</i>
hr	hour
IPTG	isopropyl β -D-1-thiogalactopyranoside
k_{cat}	catalytic constant
K_m	Michaelis constant
LDH	lactate dehydrogenase
ManNAc	<i>N</i> -acetyl-D-mannosamine
min	minute
NAD ⁺	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide reduced form
NAL	<i>N</i> -acetylneuraminic acid lyase
PCR	polymerase chain reaction
PDB	Protein Data Bank
RNA	ribonucleic acid
rpm	revolutions per minute
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
saNAL	<i>Staphylococcus aureus N</i> -acetylneuraminic acid lyase
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
s	second
TEMED	<i>N,N,N',N'</i> -tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
UV	ultraviolet

Chapter 1 Introduction

Enzymes have evolved as highly powerful biological catalysts which allow biological reactions to occur within cells. These biological catalysts can significantly increase reaction rates up to 10^{20} -fold, whilst operating under ambient conditions (Lad *et al.*, 2003, Stockbridge *et al.*, 2010). This catalytic power has been exploited for use in many industrial processes.

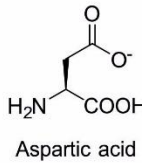
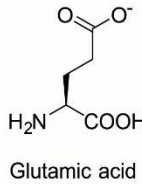
Enzymes are polypeptides and so are produced by the sequential addition of amino acids. This process is directed by information stored as DNA in the genetic code. The genetic code is a degenerate three-base code which allows biological information to be stored as DNA. It consists of 64 three base codons, made up from the four nucleoside bases guanine (G), cytosine (C), adenine (A) and thymine (T) (Figure 1.1). This genetic code programmes the production of proteins through the processes of transcription and translation, as each codon directs the insertion of one of the proteogenic amino acids. Degeneracy within the genetic code, whereby multiple codons will encode the same amino acid, accounts for the 64 codons coding for only twenty amino acids and three stop codons.

Nucleoside bases

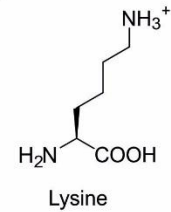
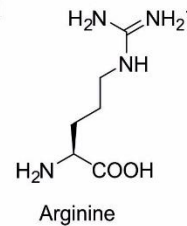


Proteogenic amino acids

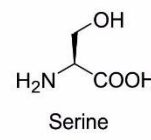
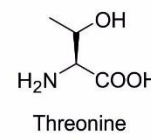
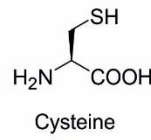
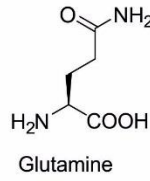
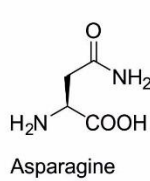
Negatively charged



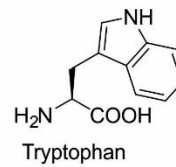
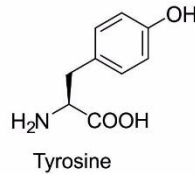
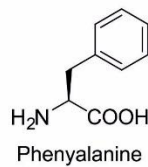
Positively charged



Polar



Aromatic



Aliphatic

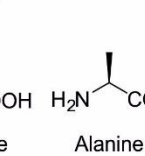
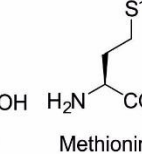
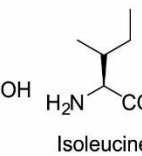
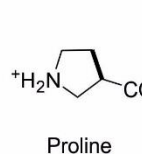
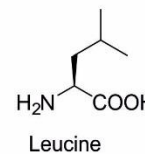
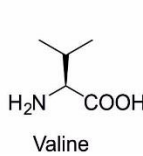
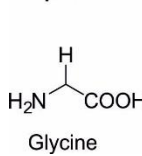


Figure 1.1 The four nucleoside bases found in DNA and the twenty naturally occurring proteogenic amino acids. The proteogenic amino acids have been grouped by the functionality of their side chains.

It is generally accepted that proteins are constructed from the twenty proteogenic amino acids. However, there are many instances where proteins are further processed after translation; post-translational modifications (PTMs) like acetylation, glycosylation and methylation further diversify the translated proteins and are all used to help regulate protein function within the cell (Seo and Lee, 2004). Enzymes are catalytically active proteins and so

are also constructed from the twenty proteogenic amino acids. The interplay of these amino acid side chains allow functionality by providing an environment primed for catalysis. Although the proteogenic amino acids provide a diverse range of side chains including aliphatic, aromatic, polar, positive and negatively charged side chains, they also lack certain functionalities. Therefore Nature has developed ways to circumvent these limitations. Many proteins require non-protein components for catalytic activity; metal ions, cofactors and prosthetic groups are all routinely recruited to enzymes and are essential for some catalysis (Broderick, 2001). In recent years it has also been seen that enzymes can utilise amino acids outside of the twenty proteogenic amino acids, and the 21st and 22nd amino acids selenocysteine and pyrrolysine (Chambers *et al.*, 1986, Hao *et al.*, 2002) were discovered in 1986 and 2002. The side chains of the proteogenic amino acids can also be post-translationally modified, and these modifications can form novel cofactors which provide functionality. These examples show the need to diversify the side chains of the proteogenic amino acids, and that by including more varied chemistries novel functionalities may be achieved.

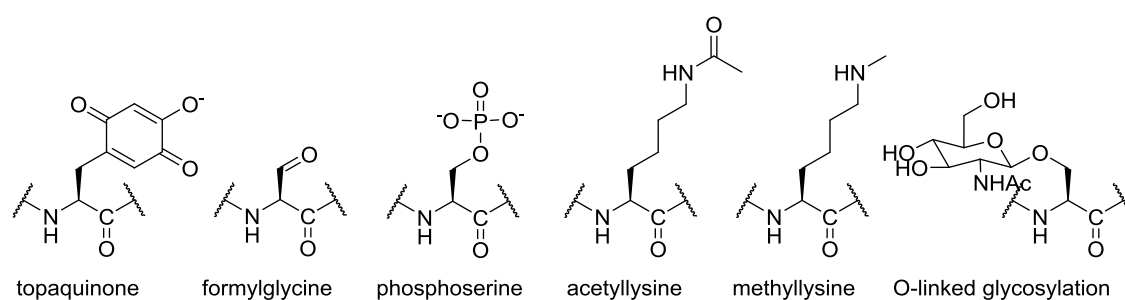


Figure 1.2 Examples of some post-translationally generated cofactors and post-translational modifications (PTMs). Topaquinone and formylglycine are post-translationally generated cofactors and produced from modification of tyrosine and cysteine or serine. Phosphoserine is a common PTM formed by the phosphorylation of serine. Acetyllysine and methyllysine are common PTMs generated by the acetylation or methylation of lysine residues. O-linked glycosylation is another common PTM and is shown here as a serine modified by an O-N-acetylglucosamine.

This thesis will focus on the incorporation of non-canonical amino acids (ncAAs) in enzyme active sites, and the subsequent effects of these ncAAs on catalytic activity. ncAAs will be defined as any amino acid which has been incorporated into the protein in an artificial manner, for example by chemical modification of a proteogenic amino acid (Basle *et al.*, 2010) or by engineering and repurposing of the natural translation machinery for ncAA insertion (Dumas *et al.*, 2015, Liu and Schultz, 2010).

1.1 Amino acid library extension

In Nature the protein backbone of an enzyme is made up from a library of 20 amino acids. These 20 amino acids allow a diversity of structures and activities to be achieved. However by the use of cofactors and post-translational modifications the number and diversity of activities achieved by the proteome can be increased and these extensions of the amino acid library can be essential for a number of biological processes.

1.1.1 21st and 22nd amino acids

Although it is generally accepted that there are twenty proteogenic amino acids, encoded by the genetic code, there are instances when Nature has expanded upon these. Two examples of this are selenocysteine and pyrrolysine which are commonly referred to as the 21st and 22nd amino acids.

Selenium is an important micronutrient, for both prokaryotes and eukaryotes, which can be incorporated into proteins in the form of selenocysteine (Schwarz and Foltz, 1957, Pinsent, 1954, Lin *et al.*, 2015). Selenocysteine is a structural analogue of the amino acid cysteine where the sulphur has been replaced with selenium. The discovery of selenocysteine in proteins began with the finding that selenium was essential for the production of an active formic dehydrogenase in some coliform organisms (Pinsent, 1954). Later, selenium was found to play an essential role in glutathione peroxidase where it was also found to be present in stoichiometric amounts (Flohe *et al.*, 1973). In 1986 the gene sequences of two selenoproteins; glutathione peroxidase (Chambers *et al.*, 1986) and formate dehydrogenase (Zinoni *et al.*, 1986) were determined. These sequences revealed that the selenocysteine was encoded by an 'in-frame' opal stop codon TGA. This co-translational insertion of selenocysteine has been found in all major domains of life (Hatfield and Gladyshev, 2002, Varlamova *et al.*, 2013). Clearly the TGA codon has two different functions; however this is not the first instance of a codon adopting dual functionality, the ATG codon functions to initiate

protein synthesis and to incorporate methionine at internal positions within a protein. To allow the dual function of the TGA codon there are auxiliary components which direct the insertion of selenocysteine, these components differ between organisms and are reviewed here by (Varlamova *et al.*, 2013, Hatfield and Gladyshev, 2002, Caban and Copeland, 2006, Driscoll and Copeland, 2003).

Pyrrolysine is another amino acid that is co-translationally inserted in a similar way to selenocysteine, and is dubbed the 22nd amino acid. Pyrrolysine was first discovered in proteins in 2002, from X-ray crystallographic studies where it was found in the active site of the methanogenic methylamine methyltransferase from *Methanosarcina barkeri* (Hao *et al.*, 2002). Pyrrolysine has been shown to be required for pathways which convert methylamines to methane (Mahapatra *et al.*, 2006, Gaston *et al.*, 2011), and is incorporated co-translationally into proteins at an amber stop codon, TAG.

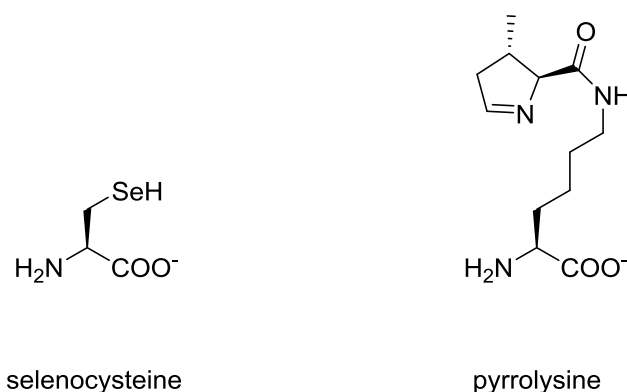


Figure 1.3 Structures of the 21st and 22nd amino acids selenocysteine and pyrrolysine.

Incorporation of selenocysteine and pyrrolysine allow the proteins to access distinctive properties that would otherwise not be provided by the first twenty proteogenic amino acids. The inclusion of the selenium atom in the side chain of selenocysteine provides unique chemistry to the side chain compared to cysteine. For example selenocysteine has a lowered pK_a, 5.4, compared to that of cysteine, 8.3 (Johansson *et al.*, 2005). Due to the lower pK_a, the selenol will be ionised at physiological pH whereas the thiol would be predominantly protonated (Kim and Gladyshev, 2005). Also the selenium produces a much more nucleophilic side chain than the cysteine which significantly enhances the chemical reaction rate with electrophiles (Arnér, 2010). Incorporation of selenocysteine and pyrrolysine and their

essential roles within proteins, highlights the need to expand on the chemistries provided by the twenty proteogenic amino acids in natural enzymes and suggests intriguing possibilities for the protein engineer.

1.1.2 Cofactors and post-translational modifications

The genetic code provides the information necessary for biological functions and for the production of proteins. However when comparing the amount of information that could be stored by the genetic code to the proteome of a single organism, there is a discrepancy. There are far more proteins produced than could be encoded by the identified protein coding genes (Walsh *et al.*, 2005). This discrepancy is due, in part, to the large amount of post-translational processing that occurs for many proteins. There are a wide variety of different post-translational modifications (PTMs) ranging from protein splicing to acetylation and glycosylation, which help to fine tune cellular events. These covalent post-translational modifications allow the chemistries of the twenty proteogenic amino acids to be altered and expanded.

Post-translational modifications (PTM) occur on nascent proteins and are often enzyme catalysed events. Modifications can take place on the backbone of the protein or on an amino acid side chain (Walsh *et al.*, 2005). PTMs can occur in two forms, firstly the cleavage of the protein backbone, and secondly the covalent addition of a chemical group usually to a nucleophilic side chain of an amino acid (Mann and Jensen, 2003, Walsh *et al.*, 2005). These modifications produce a wide range of effects; regulation of protein localisation, enzyme activity and protein-protein interactions (Seo and Lee, 2004).

Cofactors are also recruited by proteins to achieve biological functions. Cofactors are non-protein molecules and range from metal ions and iron-sulphur clusters to heme and flavins. One excellent example of an enzyme that recruits multiple different cofactors to achieve its biological function is the pyruvate dehydrogenase complex (PDH), one of the central enzymes in aerobic metabolism. The PDH links glycolysis and the tricarboxylic acid (TCA) cycle by converting pyruvate to acetyl-CoA (Patel and Korotchkina, 2003, Mande *et al.*, 1996). The PDH is a multi-enzyme complex comprised of multiple copies of three catalytic enzymes; pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and dihydrolipoamide dehydrogenase (E3) which work sequentially to catalyse the oxidative decarboxylation of pyruvate (Patel *et al.*, 2014) (Figure 1.4). To achieve this function the three catalytic enzymes are required along with five organic cofactors; thiamine pyrophosphate (TPP), lipoamide,

coenzyme A (CoA), flavin adenine dinucleotide (FAD), nicotinamide adenine dinucleotide (NAD⁺) and one inorganic cofactor; Mg²⁺. TPP forms a covalent link to E1 and provides a highly nucleophilic group that can attack the carbonyl carbon of pyruvate, and cleave CO₂ from the pyruvate. Lipoamide is covalently linked to E2 and the disulphide bond can be reduced to accept the acetyl group from TPP. The acetyl group is passed onto CoA, producing acetylCoA. E3 then uses FAD to oxidise the dihydrolipoamide reforming the disulphide bond producing FADH₂, which is in turn oxidised to FAD by NAD⁺ (Patel *et al.*, 2014)

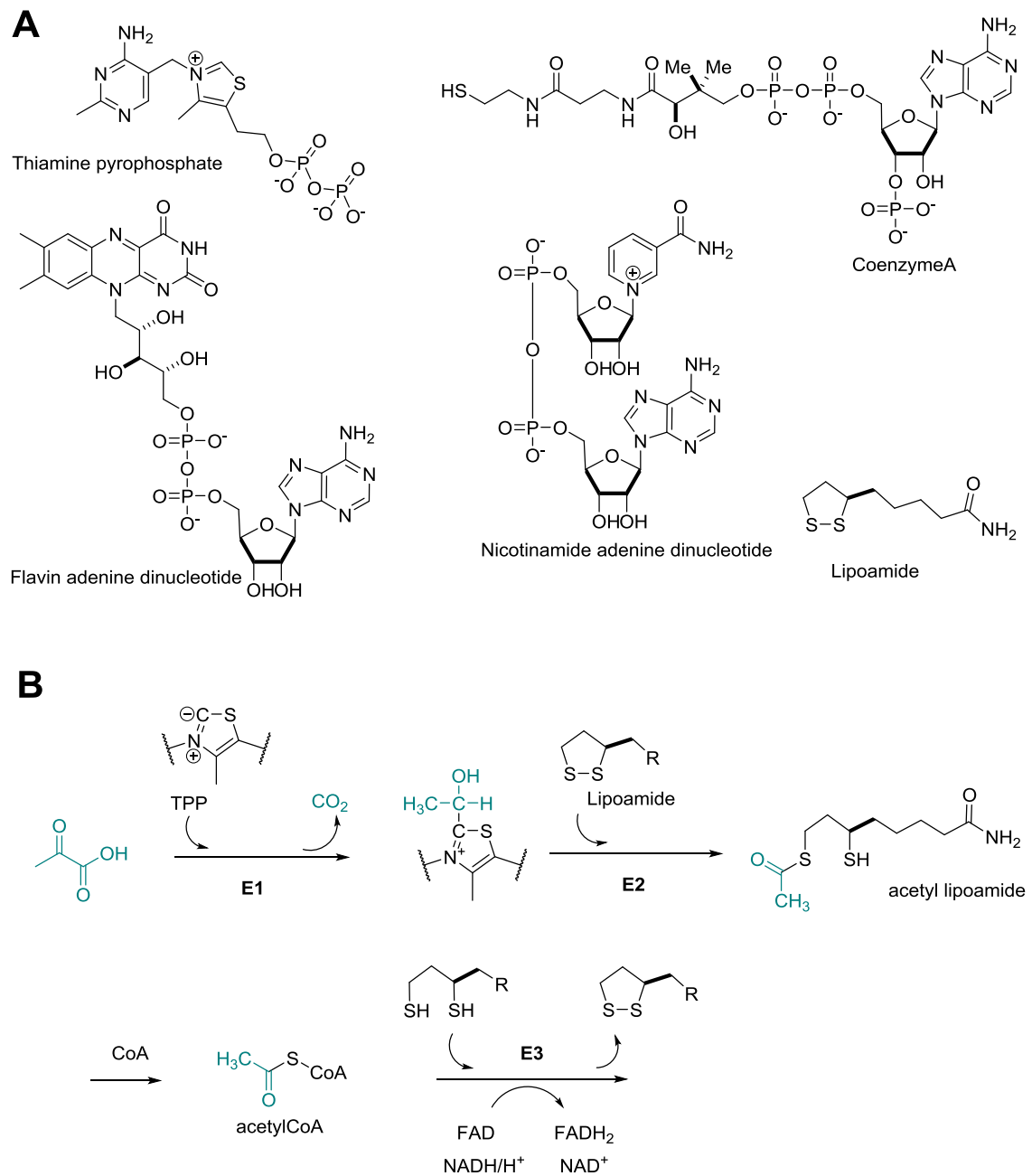


Figure 1.4 Cofactors and mechanism of the pyruvate dehydrogenase complex (PDH). Structures of the cofactors used by the PDH are shown in panel A; thiamine pyrophosphate (TPP) coenzyme A (CoA). Flavin adenine dinucleotide (FAD), nicotinamide adenine dinucleotide (NAD⁺) and lipoamide. Panel B shown the roles of the cofactors in the PDH, TPP provides a nucleophile to attack the carbonyl of pyruvate cleaving off CO₂ and producing hydroxyethyl-TPP, lipoamide becomes reduced and transfers the acetyl group to CoA producing acetylCoA. The lipoamide is subsequently reoxidised by FAD and NAD⁺.

1.1.3 Post-translationally generated cofactors

As discussed above, many enzymes require cofactors for their function. These cofactors are usually biosynthesised separately to the protein and subsequently recruited to the protein. However more recently, there have been a large number of studies which have revealed enzymes which have alternative routes to producing cofactors via post-translational modification events (Okeley and Van Der Donk, 2000, Davidson, 2011, Yukl and Wilmot, 2012). These novel catalytic centres are generally produced by modification of one or a number of amino acid side chains and can be generated in a number of ways; some occur from auto-catalytic events and others require accessory enzymes.

1.1.3.1 Pyruvoyl group

Histidine decarboxylase was one of the first enzymes that was identified to undergo one of these modification events. Many amino acid decarboxylases require the cofactor pyridoxal-5-phosphate, derived from vitamin B6, for their function (Figure 1.5 B). However it was found that histidine decarboxylase of *Lactobacillus* 30a did not require PLP for its function (Riley and Snell, 1968, Vanpoelje and Snell, 1990). Instead it was found that the enzyme contained an amino-bound pyruvoyl group which acts in a manner analogous to PLP; forming a Schiff base with the histidine substrate and facilitating decarboxylation (Okeley and Van Der Donk, 2000).

Histidine decarboxylase is produced as an inactive hexameric proenzyme, which undergoes an autocatalytic cleavage event to produce the α and β chains of the active enzyme (Figure 1.5 A). This cleavage event occurs between two serine residues, Ser81 and Ser82 (Vaaler *et al.*, 1982, Vanpoelje and Snell, 1990). The inactive proenzyme contains no pyruvate and it has been shown that the pyruvoyl group is generated, post-translationally, during the autocatalytic cleavage event (Recsei *et al.*, 1983). Ser82 becomes the pyruvoyl group on the amino terminus of the α chain, and Ser81 becomes the C-terminus of the β subunit. Several other enzymes have been identified which require the generation of a similar pyruvoyl group; arginine, aspartate and S-adenosylmethionine decarboxylases (Bale and Ealick, 2010) and D-proline and L-glycine reductases (Okeley and Van Der Donk, 2000), all of which use similar pathways to post-translationally generate the cofactor. Self-production of the pyruvoyl group means that these enzyme functions are no longer dependent on recruiting an externally synthesised cofactor.

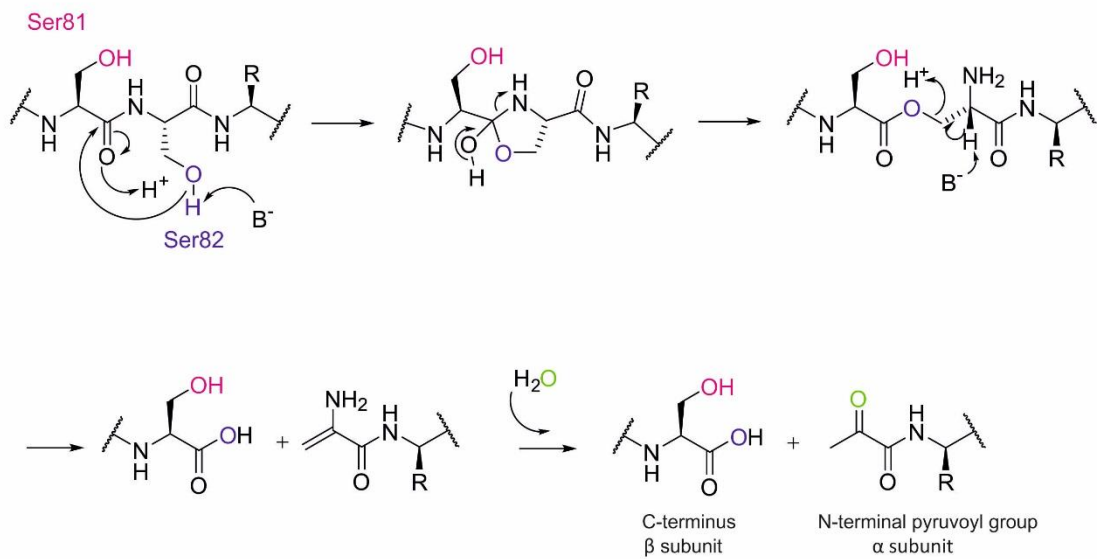
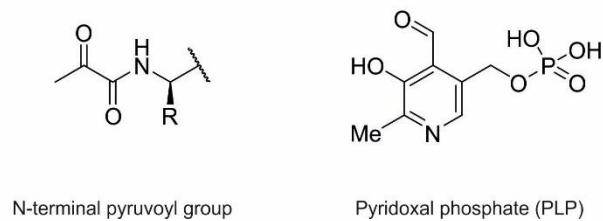
A**B**

Figure 1.5 Panel A shows the mechanism of backbone cleavage between Ser81 and Ser81 to form the pyruvoyl cofactor in histidine decarboxylase. The hydroxyl of Ser82 attacks the carboxyl of Ser81 producing tetrahedral intermediate that then forms an ester intermediate which undergoes elimination to release the C-terminal β subunit and hydrolysis of the dehydroalanine α subunit produces the N-terminal pyruvoyl group. Panel B shows a comparison of the structures of the post-translationally generated N-terminal pyruvoyl cofactor and pyridoxal phosphate which is commonly used instead.

1.1.3.2 Formylglycine

Type I sulfatases are a group of enzymes that catalyse the hydrolysis of sulphate esters. These enzymes contain a catalytically essential formylglycine (FGly) residue in their active site (Figure 1.6A), which is generated by post-translational modification (Bojarova and Williams, 2008, Appel and Bertozzi, 2015). FGly contains an aldehyde functional group which can form a geminal-diol by hydration, one of these hydroxyls can then act as the nucleophile for attack of the sulphur atom of the substrate (Lukatela *et al.*, 1998) (Figure 1.6B). This causes removal of the sulphate group and cleavage of the sulphate ester, producing a sulphated enzyme, the sulphate group can then be eliminated when the aldehyde is regenerated (Dierks *et al.*, 1998, Appel and Bertozzi, 2015).

The cofactor FGly is produced by oxidation of a cysteine, in prokaryotes and eukaryotes, or a serine, in prokaryotes, which is directed by a common peptide motif (C/S)XPXRXXXLTGR. This generation of FGly is performed by dedicated enzymes; formylglycine generating enzymes (FGEs) and anaerobic sulfatase maturing enzymes (anSMEs). FGEs are aerobic and operate through a redox mechanism that uses molecular oxygen (Dierks *et al.*, 2005), whereas anSMEs work in an anaerobic manner and require the cofactor S-adenosylmethionine (Bojarova and Williams, 2008). FGly sulfatases produce very impressive rate enhancements, for example $k_{\text{cat}}/k_{\text{uncat}}$ of up to 10^{26} for certain substrates (Edwards *et al.*, 2012, Appel and Bertozzi, 2015). This significant rate enhancement may be attributed to the unique catalytic properties of the FGly side chain.

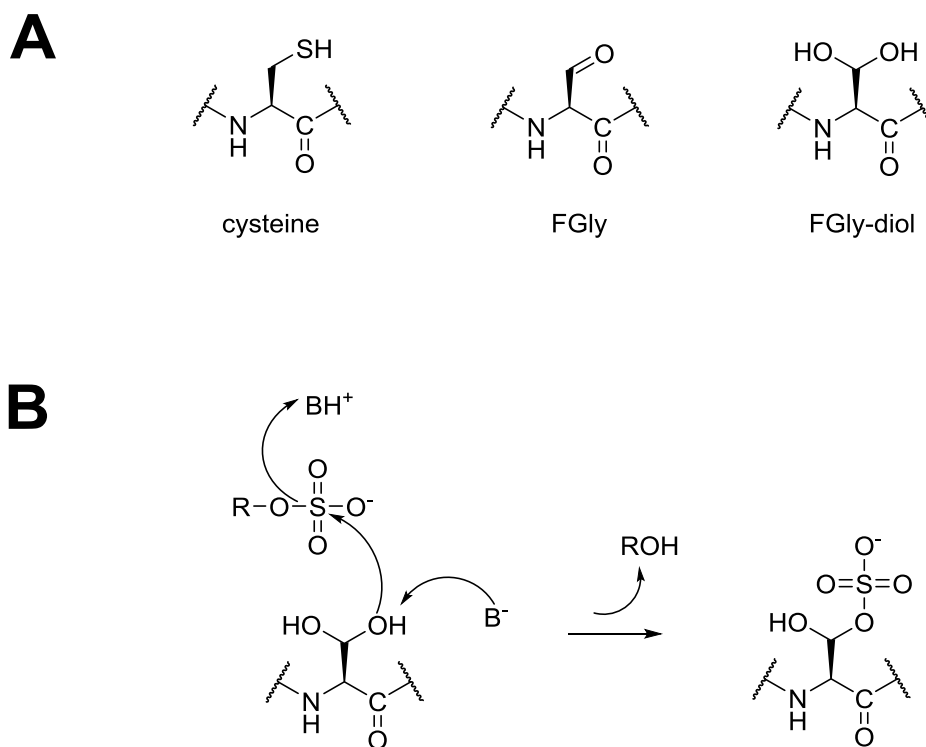


Figure 1.6 Formylglycine (FGly) structure and mechanism of action. Panel A shows the structures of cysteine, which is post-translationally converted into FGly, FGly and FGly-diol which is the hydrated diol form of FGly. Panel B shows the mechanism by which FGly cleaves the sulphate ester, one of the hydroxyls from the FGly-diol acts as a nucleophile to attack the sulphur atom, cleaving the sulphate ester bond and producing a sulphated FGly.

1.1.3.3 Post-translationally generated quinone cofactors

Redox active enzymes play vital roles in both prokaryotic and eukaryotic organisms. However due to the limited functionality provided by the proteogenic amino acids, these enzymes tend to require additional cofactors or metal ions to achieve their function. In recent years enzymes with catalytic redox-active centres derived from amino acid side chains have been discovered. Many of these enzymes produce quinone cofactors, which are highly versatile cofactors due to their strong electrophilic nature and their ability to undergo two reversible, one-electron reductions (Lang and Klinman, 2001, Davidson, 2011). There are four quinone cofactors which are derived from modification of amino acid side chains; topaquinone (TPQ), lysine tyrosylquinone (LTQ), cysteine tryptophanquinone (CTQ), and tryptophan tryptophylquinone (TTQ) (Figure 1.7). TPQ and LTQ are both produced auto-catalytically and derive from a tyrosine side chain, with LTQ being formed from the crosslinking of a lysine and tyrosine and

TPQ being formed by modification of a single tyrosine (Davidson, 2011). The side chain of a tyrosine residue can normally be involved in radical processes, by undergoing oxidation to the phenoxyl radical, as in ribonucleotide reductase. But, to be involved in a redox process, the tyrosine side chain is required to undergo modification to a quinone (Dubois and Klinman, 2005). CTQ and TTQ are both derived from tryptophan by oxidation of the aromatic ring. Production of TTQ requires accessory proteins and it is hypothesised that CTQ may also require accessory proteins (Davidson, 2011). These post-translationally derived cofactors provide electrophilic functionalities to enzymes which are otherwise absent from the side chains of the standard proteogenic amino acids.

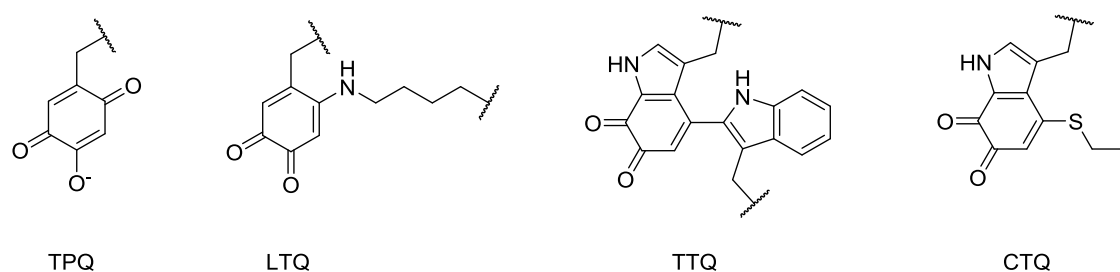


Figure 1.7 Structures of the post-translationally generated quinone cofactors topaquinone (TPQ), lysine tyrosylquinone (LTQ), tryptophan tryptophylquinone (TTQ) and cysteine tryptophanquinone (CTQ).

1.1.3.4 5-methylene-3,5-dihydroimidazol-4-one (MIO)

Ammonia lyases and amino mutases share sequence and structural similarities. Aromatic amino acids undergo degradation by a different mechanism than other amino acids. Non-oxidative deamination of the aromatic amino acids is carried out by histidine, phenylalanine and tyrosine ammonia lyases (Ravikiran and Mahalakshmi, 2014). Amino mutases catalyse the conversion of α -amino acids to β -amino acids and ammonia lyases catalyse the asymmetric addition of ammonia to a double bond (Turner, 2011). The aromatic amino acid lyases and amino mutases have been found to produce the electrophilic cofactor 5-methylene-3,5-dihydroimidazol-4-one (MIO) (Figure 1.8) which allows them to carry out their processes (Christianson *et al.*, 2007, Turner, 2011). Production of MIO occurs in an autocatalytic manner, where a conserved motif of Ala-Ser-Gly undergoes a self-catalysed condensation reaction to produce the electrophilic MIO cofactor (Christianson *et al.*, 2007).

The heterocyclic moiety found in MIO has enhanced electrophilic properties, over that of a standard dehydroalanine, which aids the mechanism of amino mutases and ammonia lyases.

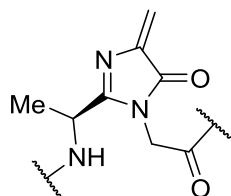


Figure 1.8 Structure of the post-translationally generated cofactor 5-methylene-3,5-dihydroimidazol-4-one (MIO).

1.2 Incorporation of non-canonical amino acids

The previous sections have highlighted the many ways in which Nature expands on the chemistries provided by the twenty proteogenic amino acids. Within a laboratory environment, advances in molecular biology and recombinant protein expression has made studies involving replacing or altering the proteogenic amino acids, within a protein, commonplace. However, access to the various modifications that Nature uses so frequently can be much more difficult.

Developing methods to add more diverse building blocks into proteins should give an excellent platform to study their effects on protein function. Also, by providing more varied chemistries, this may expand the range of functions that can be achieved by proteins and enzymes. Methods for the modification of proteins and incorporation of non-canonical amino acids have advanced rapidly in recent years. These methods can be separated into two general groups, firstly the methods that incorporate non-canonical residues using the genetic machinery of the cell and secondly methods that use the distinctive properties of some amino acid side chains to target them for chemical modification.

1.2.1 Auxotrophic incorporation

In the 1950s it was discovered that selenomethionine could be used to replace methionine within bacterial proteins (Cowie and Cohen, 1957). This replacement of a proteogenic amino acid for a non-canonical amino acid (ncAA) can be achieved by the use of auxotrophic strains of bacteria. Amino acid auxotrophs are deficient in the machinery required to produce one of the proteogenic amino acids, and therefore the amino acid must be supplemented for growth. By supplementing the media with a non-canonical analogue of the proteogenic amino acid it is sometimes possible to insert the non-canonical analogue at every instance that the proteogenic amino acid should have been incorporated.

This global replacement method can cause significant changes in the chemical and physical properties of proteins, and since all cellular proteins will be modified it provides a useful tool for proteomic studies (Johnson *et al.*, 2010, Link *et al.*, 2003). The global replacement of the hydrophobic amino acid leucine with fluorinated analogues has been shown to cause thermal stabilisation and stabilisation against chemical denaturation in an engineered leucine zipper (Tang and Tirrell, 2001). Replacement of proline with fluoroproline in collagen strands (ProProGly)⁷, to produce a protein containing two thirds non-canonical residues, has been shown to increase the thermal stability by over 50°C (Shoulders *et al.*, 2010). Multiple different ncAAs, up to 13, have been individually, globally incorporated into a lipase enzyme. Incorporation of each non-canonical analogue of methionine, proline, phenylalanine, and tyrosine (shown in Figure 1.9) was carried out to determine the physical and chemical effects (Hoesl *et al.*, 2011). These global replacements produced a variety of different effects including; increased enzyme activity, increased optimum temperature and increased tolerance to a variety of organic solvents, surfactants and denaturing agents (Acevedo-Rocha *et al.*, 2013, Hoesl *et al.*, 2011).

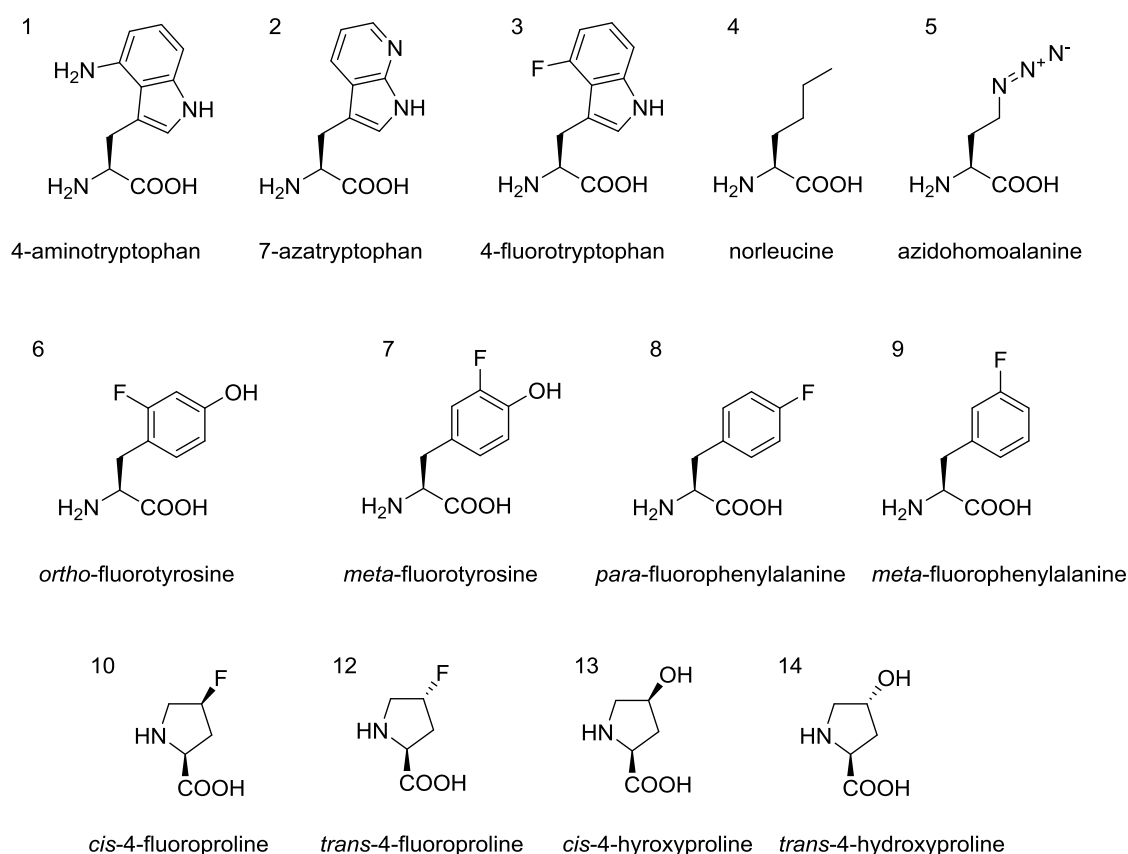


Figure 1.9 Non-canonical amino acids used in global replacement studies of a lipase. Each ncAA was inserted using auxotrophic methods. Tryptophan was replaced with 1-3 using a tryptophan auxotroph. Methionine was replaced with 4 and 5 using a methionine auxotroph. Tyrosine was replaced with 6 and 7 using a tyrosine auxotroph. Phenylalanine was replaced with 8 and 9 using a phenylalanine auxotroph. Proline was replaced with 10-14 using a proline auxotroph.

Using auxotrophic bacterial strains provides a method to insert ncAAs without the need to manipulate the genetic machinery of the cell. However this approach relies on the ability of the aminoacyl-tRNA synthetase to accept different substrates, and subsequently load them onto the appropriate tRNA. In some cases the loading of the ncAA by the aminoacyl-tRNA synthetase can be very slow, and so to produce the modified proteins the aminoacyl-tRNA synthetase must be overexpressed within the cell (Kiick *et al.*, 2000). Another drawback of this method is that the amino acid analogues can prove to be toxic to the cell, which can lead to reduced expression of the modified protein and sometimes less than 100% incorporation of the ncAA (Budisa *et al.*, 2004, Singh-Blom *et al.*, 2014). To circumvent these cytotoxic effects, cell-free protein synthesis has been developed to globally incorporate analogues of the amino acid tryptophan (Singh-Blom *et al.*, 2014). Using this cell-free system, which is depleted in

tryptophan it was possible to achieve 100% incorporation of the tryptophan analogue into target proteins.

These residue specific incorporation methods are very useful for studying the effects of global ncAA incorporation, or for producing proteins with amino acids that can be chemically modified at multiple positions. However to study more localised effects of ncAA incorporation at one position, for example inserting an ncAA into the active site of an enzyme and investigating the effects on catalysis, a site-specific method of incorporation is needed.

1.2.2 tRNA suppressor technology

To insert a non-canonical residue site-specifically *in vivo*, alterations to the translational machinery of the cell must be made. Amino acid residues are inserted into a growing polypeptide chain, via tRNA, in response to the triplet code of mRNA. tRNAs are aminoacylated with their corresponding amino acid by a tRNA synthetase. The tRNA/tRNA synthetase pairs that are encoded in the cell have already evolved to load a specific amino acid onto its complementary tRNA; therefore to insert an ncAA a new orthogonal tRNA/tRNA synthetase pair must be evolved. The evolved tRNA must not cross react with the host system and must efficiently incorporate the ncAA in response to a codon. The evolved tRNA synthetase must efficiently load the ncAA onto the new tRNA and must not recognise any of the host tRNAs. The most commonly used codon, at which the ncAA is incorporated, in this methodology is the amber stop codon UAG, due to this codon being a nonsense codon the method is commonly referred to as nonsense suppression (Figure 1.10).

The first instance of nonsense suppression was used to insert *O*-methyl-L-tyrosine into dihydrofolate reductase, in response to an amber stop codon UAG in *E. coli* (Wang *et al.*, 2001). To insert the ncAA an orthogonal tRNA/tRNA synthetase pair was evolved from the *Methanococcus jannaschii* (*M. jannaschii*) tyrosyl tRNA/tRNA synthetase pair. The machinery from *M. jannaschii* was a good target as an orthogonal pair as it already had a number of different elements from the tyrosyl tRNA tRNA synthetase pair found in *E. coli*. With further protein engineering, using rounds of positive and negative selection, it was possible to develop a tRNA/ tRNA synthetase pair that was fully orthogonal to the *E. coli* system, and which was used to insert the ncAA in response to the amber stop codon with a 99% fidelity (Wang *et al.*, 2001).

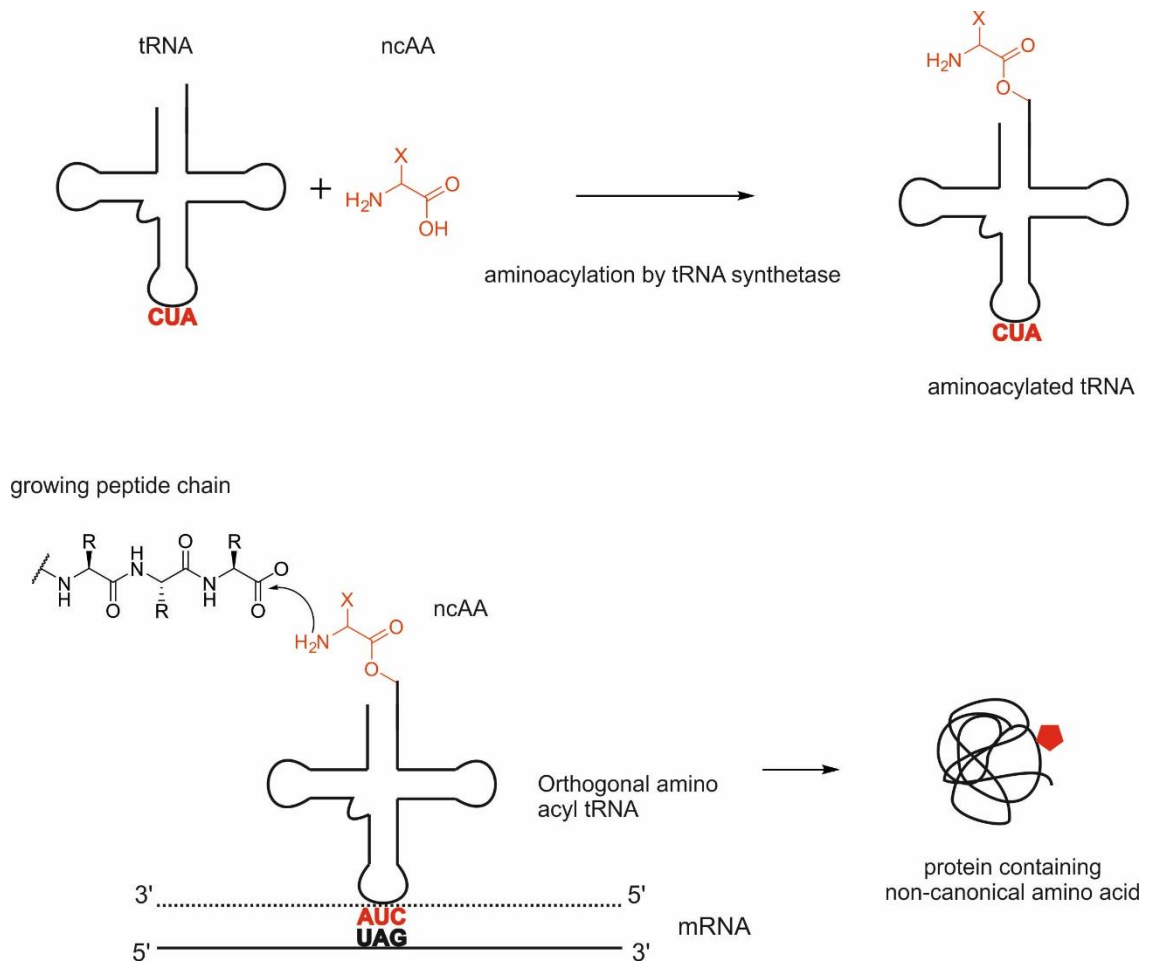


Figure 1.10 Schematic of nonsense suppression. The orthogonal tRNA synthetase aminoacylates the orthogonal tRNA with the non-canonical amino acid (ncAA) ready for use in translation. The anticodon loop of the orthogonal tRNA recognises the amber stop codon UAG and therefore inserts the ncAA into the growing peptide chain, rather than termination of translation occurring.

Over recent years this method has been extensively used and improved, allowing incorporation of over 150 different ncAAs into a variety of prokaryotic and eukaryotic systems (Dumas *et al.*, 2015, Lang and Chin, 2014, Tuley *et al.*, 2014, Young and Schultz, 2010). Approximately a third of the ncAAs to date are inserted using an orthogonal pair based on the original *M. jannaschii* tyrosyl tRNA/tRNA synthetase pair. However the *M. jannaschii* tyrosyl tRNA/ tRNA synthetase pair is not orthogonal in higher organisms like yeast, therefore to insert ncAAs the *E. coli* tyrosyl tRNA/tRNA synthetase pair was used (Chin *et al.*, 2003). Other commonly used orthogonal pairs are the leucine tRNA/tRNA synthetase pair from *E. coli* for use in higher organisms like yeast and the pyrrolysyl tRNA/tRNA synthetase pairs from *Methanosarcina barkeri* and *Methanosarcina maze*. The discovery of pyrrolysine in 2002 (as discussed in Section 1.1.1), found that pyrrolysine was inserted into the proteins of some organisms in response to a UAG codon, which would normally encode translation termination. It was found that pyrrolysine was loaded onto a corresponding tRNA, which recognises the UAG codon, as an intact amino acid by a pyrrolysyl tRNA synthetase (Blight *et al.*, 2004). This gave an excellent starting point for an orthogonal pair; due to its origins in archaeobacteria it has been found to be orthogonal in both prokaryotes and eukaryotes (Hao *et al.*, 2011). Also since the pyrrolysyl tRNA synthetase does not recognise one of the twenty proteogenic amino acids there is no need to remove its natural synthetase activity before evolving an activity for a new ncAA. The substrate scope of the wild-type pyrrolysyl tRNA synthetase was quite broad allowing the insertion of multiple different pyrrolysine analogues. However, with further engineering mutant pyrrolysyl tRNA synthetases have been developed to further expand the substrate scope, and over 80 different ncAAs can be inserted using the orthogonal pairs of wild-type and mutant pyrrolysyl tRNA/tRNA synthetases pairs (Dumas *et al.*, 2015). Using these orthogonal pairs it has been possible to insert ncAAs which are mimics of naturally occurring PTMs, for example the lysine modification N- ϵ -acetyllysine, a common histone modification was incorporated into myoglobin using an evolved pyrrolysyl tRNA/tRNA synthetases pair (Neumann *et al.*, 2008).

Expanding on this technology, tRNA/tRNA synthetase pairs have been developed that will decode a quadruplet codon and insert an ncAA in response. Using this method it was possible to incorporate two different ncAAs into the same protein by using the quadruplet codon to encode homoglutamine and insert it using a lysyl tRNA/synthetase pair derived from *Pyrococcus horikoshii*, and the also using the amber stop codon to insert *O*-methyl-L-tyrosine using the tyrosyl tRNA/synthetase pair from *M. jannaschii*. However efficiency of incorporation

was low (Anderson *et al.*, 2004). Low incorporation efficiency in response to the quadruplet codon is due to the larger quadruplet anti-codon of the tRNA being poorly accepted by the natural ribosome. Therefore, to increase efficiency the ribosome needs to be evolved to accept the quadruplet tRNA. To prevent this engineering affecting translation within the cell, a quadruplet decoding ribosome has been developed that works orthogonally to the cell's natural machinery. This engineered ribosome has been shown to efficiently recognise and decode orthogonal mRNA designated by an alternative Shine-Dalgarno sequence, the orthogonal mRNA has also been shown not to cross interact with the wild-type ribosome (Neumann *et al.*, 2010, Wang *et al.*, 2007).

This research shows the significant advances that have been made towards site-specific *in vivo* introduction of ncAAs. This method has provided an excellent way in which to site-specifically incorporate tags for modification, ways to study PTMs and incorporate biophysical and fluorescent probes (Davis and Chin, 2012). Due to the use of the pyrrolysyl tRNA synthetase, and its relaxed substrate specificity, it has been possible to considerably expand the range of ncAAs that can be incorporated in this manner. However the core structural motifs that can be incorporated are mainly phenylalanine, tyrosine or lysine analogues (Wan *et al.*, 2014, England, 2004, O'donoghue *et al.*, 2013, Davis and Chin, 2012, Dumas *et al.*, 2015). These methods are also reliant on the ability of a tRNA/tRNA synthetase pair to accept the desired ncAA. To incorporate a wide variety of different side chains into one protein, it would be necessary to evolve a specific tRNA/tRNA synthetase pair for each ncAA to be incorporated. Therefore more work towards discovery and engineering of new orthogonal pairs is needed to keep expanding the side chain chemistries that can be inserted. The majority of the ncAAs are inserted in this manner are in response to an amber stop codon, and so the natural function of the amber stop codon is in competition with the evolved function to insert the ncAA. This competition can reduce the efficiency of incorporation, resulting in low yields of modified protein, and can sometimes prevent incorporation entirely (O'donoghue *et al.*, 2013, Zhang *et al.*, 2013).

1.2.3 Chemical modification of proteins

Chemical modification of proteins provides a way to attempt to synthetically mimic the many varied modifications performed by Nature. Modifying proteins chemically can prove to be quite a challenge as proteins contain a variety of reactive groups; carboxylic acids, amines, amides, alcohols and thiols. Therefore the chemistry employed must be specific, usually to target the side chain of a specific amino acid, and the chemistry must be carried out under

conditions that will not irrevocably impair the protein structure or function, for example aqueous conditions, near neutral pH and ambient temperatures (Spicer and Davis, 2014, Basle *et al.*, 2010).

Pivotal experiments carried out in the 1960s used chemical methods to convert one amino acid into another. Firstly an *O*-*p*-toluenesulfonyl-L-serine residue, on a short peptide, was converted into a thiol containing cysteine residue by reaction with thioacetate (Zioudrou *et al.*, 1965). This methodology was extended for use in proteins; the serine in the active site of subtilisin was treated firstly with phenylmethanesulfonyl fluoride (PMSF) and then with either thioacetate or thiobenzoate to produce a cysteine residue (Polgar and Bender, 1966). These experiments, carried out in the era before the development of site-directed mutagenesis, were the first examples of amino acid replacements. Although genetic manipulation has now made exchanging amino acids simple and commonplace, the power of this chemical modification method must not be underestimated.

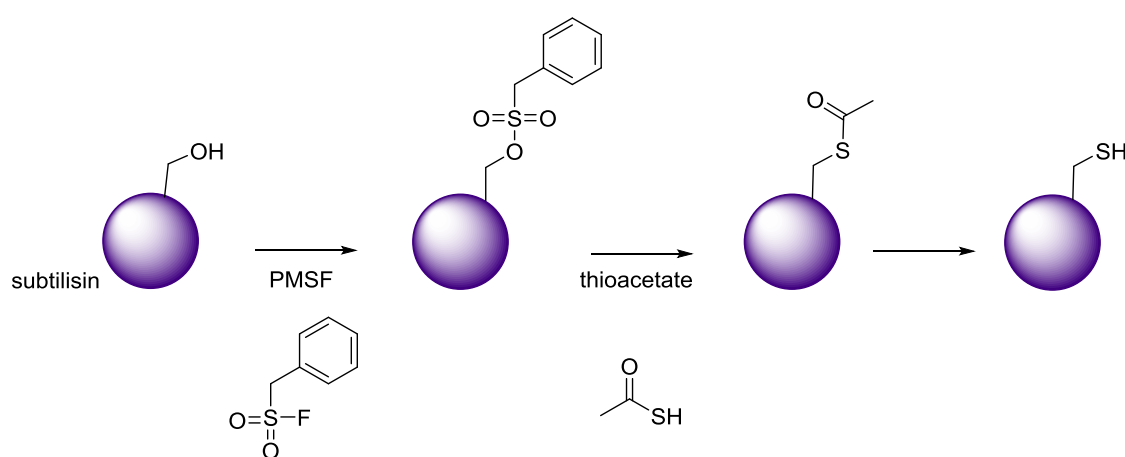


Figure 1.11 Conversion of active site serine to cysteine in subtilisin. Reaction with phenylmethanesulfonyl fluoride (PMSF) with the active site serine followed by displacement of the phenylmethanesulfonyl group by thioacetate followed by spontaneous hydrolysis to form cysteine.

The most common approaches for chemical modification of proteins takes advantage of unique chemistries found in the side chains of amino acids. Another powerful approach to selective chemical modification can now be achieved by using a tRNA/tRNA synthetase pair to genetically incorporate an ncAA with a selectively reactive group, which can then be subsequently modified (Spicer and Davis, 2014, Hao *et al.*, 2011). Modification using the proteogenic amino acids utilises the chemistries of the side chains, for example; the primary

amine group of lysine which can act as a nucleophile, the side chain of arginine can be modified to a pyrimidine derivative (Oya *et al.*, 1999), and histidine can be selectively modified with epoxides (Chen *et al.*, 2003). However the most common proteogenic side chain to be modified is cysteine. Cysteine has a number of advantages as the target for modification; firstly it is one of the most uncommon of the proteogenic amino acids, which may allow more selective modification, and secondly its side chain is a free thiol group which is highly nucleophilic and not found within the side chains of the other proteogenic amino acids (Chalker *et al.*, 2009).

1.2.3.1 Chemical modification at cysteine

Cysteine provides a unique chemical handle which can be modified in multiple different ways. The sulphur in the cysteine can interconvert between multiple oxidation states, which is how post-translational modification of cysteine plays a role cellular redox regulation (Wani *et al.*, 2014). Oxidation of cysteine also allows formation of disulphides and by modification via disulphide formation it has been possible to study glycosylation mimics (Gamblin *et al.*, 2004). The free thiol group of cysteine is the most highly nucleophilic of the proteogenic amino acid side chains and with control of pH, selective modification of cysteine over other nucleophilic side chains lysine and histidine, can be achieved (Chalker *et al.*, 2009, Bischoff and Schlüter, 2012). Cysteine is one of the least commonly occurring proteogenic amino acids, and with the use of site-directed mutagenesis to remove other cysteine residues and to place a cysteine at a desired position for modification, it can provide a highly selective handle for modification. Due to the nucleophilic nature of cysteine it can be targeted by electrophiles for alkylation, which can be achieved using electrophilic maleimides and α -halocarbonyls. Alkylation of cysteine can also be used to produce mimics of the post-translational modifications of lysine, methyl lysine and acetyl lysine within proteins (Huang *et al.*, 2010, Simon *et al.*, 2007)(Figure 1.12). Methyl lysine and acetyl lysine play important roles in the epigenetic control of the cell, and access to these structural mimics has provided important insight into these post-translational modifications.

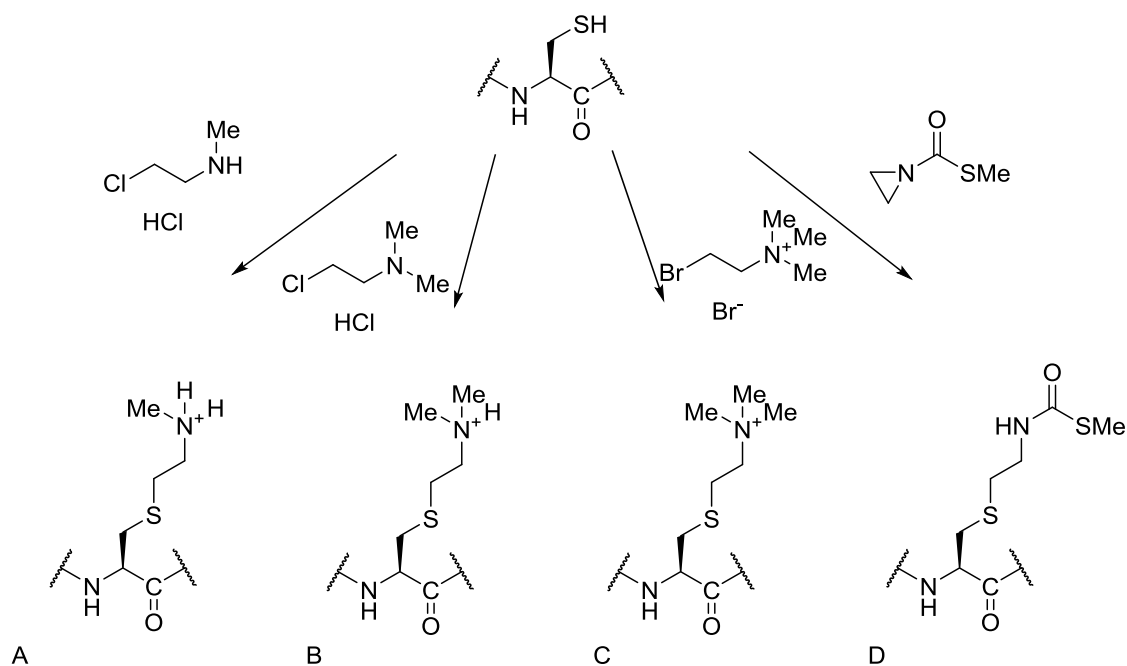


Figure 1.12 Methyl lysine and acetyl lysine analogues produced from alkylating a cysteine residue. A-C show methyl lysine analogues A) monomethyl- γ -thialysine formed by reaction of cysteine and (2-chloroethyl)-methylammonium chloride, B) dimethyl- γ -thialysine formed by reaction of cysteine and (2-chloroethyl)-dimethylammonium chloride, C) trimethyl- γ -thialysine formed by reaction of cysteine and (2-bromoethyl)-trimethylammonium bromide (Simon *et al.*, 2007). D is the acetyl lysine analogue, methylthiocarbonyl- γ -thialysine formed from reaction of cysteine with methylthiocarbonyl-aziridine (Huang *et al.*, 2010).

Cysteine can also act as a precursor for the production of the reactive electrophilic handle dehydroalanine (dha). Dha is a more selective target for modification, than any of the proteogenic amino acid side chains, as it is highly electrophilic, a feature which is lacking within the proteogenic amino acids. Synthetically dha can provide access to many different PTM analogues and ncAAs by a conjugate addition with a thiol; dha is electrophilic and therefore acts as a Michael acceptor for reactions with nucleophilic thiols.

Dehydroalanine is found in nature in compounds known as lantibiotics, short peptides which undergo post-translational modification, where it is produced from the enzymatic dehydration of serine (Willey and Van Der Donk, 2007). These modifications to produce unsaturated compounds are common within lantibiotic synthesis, and dehydration of threonine can also occur to produce dehydrobutyrine (Willey and Van Der Donk, 2007, Xie and Van Der Donk, 2004). There are multiple reported methods for the synthetic production of dha in proteins, including, elimination of active site serines, the use of bromomaleimides and the use of

nonsense suppression to insert a phenylselenocysteine residue followed by a subsequent reaction with hydrogen peroxide to form dha (Chalker *et al.*, 2011, Tedaldi *et al.*, 2009, Guo *et al.*, 2008, Neet and Koshland, 1966). Another method to produce dha, uses the compound 2,5-dibromohexan-1,6-diamide to target a cysteine residue which undergoes double alkylation to produce a sulfonium salt which is then eliminated to produce dha (Chalker *et al.*, 2011) (Figure 1.13). This method to produce dha has then been combined with the Michael addition of thiol compounds to produce analogues of lysine PTMs on synthetic histones (Chalker *et al.*, 2012), extended phosphorylated residues to investigate kinase activation (Rowan *et al.*, 2013), and modify enzyme active site residues (Timms *et al.*, 2013).

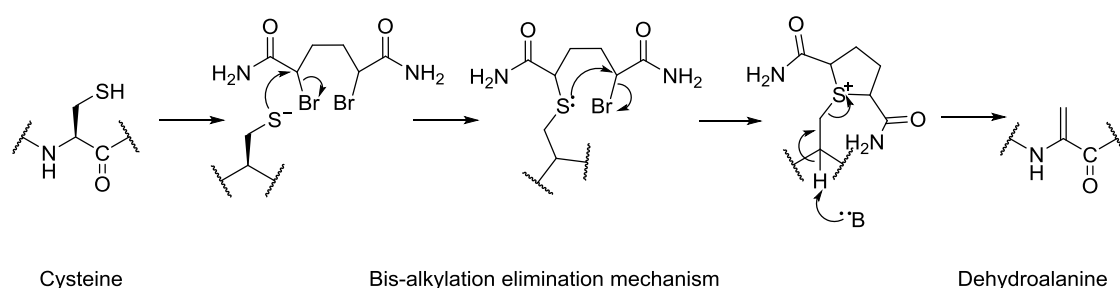


Figure 1.13 Bis-alkylation-elimination mechanism for the conversion of a cysteine residue into a dehydroalanine residue carried out by the compound 2,5-dibromohexan-1,6-diamide (Chalker *et al.*, 2011).

1.3 Catalysis with non-canonical amino acids

There are many different examples of ncAAs being used to mimic PTMs, as photo-crosslinkers, as biochemical and biophysical probes and labels, some of which have been mentioned previously and have been reviewed by (Dumas *et al.*, 2015, Davis and Chin, 2012, Basle *et al.*, 2010, Hao *et al.*, 2011). However, another interesting use of ncAAs is determining their effects on the activity of enzymes.

Global replacement of amino acids with a non-canonical analogue has been shown to cause increases in activity. For example replacement of methionine with the non-canonical norleucine, increased peroxxygenase activity in a cytochrome P450 monooxygenase (Cirino *et al.*, 2003). Replacement of tryptophan with 5-fluorotryptophan in glutathione transferase produced a 4-fold increase in k_{cat} by enhancing the product release (Parsons *et al.*, 1998). It has also been reported that site-specific introduction of ncAAs can be used to alter catalysis.

Studies into the nitroreductase enzyme from *E. coli*, which has found use as a prodrug activator for cancer therapeutics, revealed that Phe124 was a critical residue for substrate binding (Jackson *et al.*, 2006). Therefore mutagenesis to proteogenic amino acids was carried out with limited success in improving the activity, with lysine being identified as the best proteogenic amino acid. Therefore ncAAs were investigated to see if further improvements could be made. Eight different phenylalanine analogues were inserted at position 124 using nonsense suppression, with varying effects on catalysis, by inserting *p*-nitrophenylalanine a 30-fold improvement over the wild-type enzyme was seen and a 2.3-fold improvement over the proteogenic F124K variant was seen (Jackson *et al.*, 2006). A bacterial phosphotriesterase (PTE) activity has also been improved by the site-specific incorporation of ncAAs by nonsense suppression. This bacterial PTE was shown to hydrolyse toxic organophosphate pesticides like paraoxon, at rates thought to be at their evolutionary limits. From structural studies aromatic ring interactions between Tyr309, Phe132 and Trp131 were seen to stabilise the ring of the paraoxon substrate. The Tyr309 was replaced with the ncAAs 7-hydroxycoumarinyl and 7-methylcoumarinyl, and the 7-hydroxycoumarinyl residue increased the turnover rate (k_{cat}) by 8-fold and 11-fold at higher pH, however it did not affect the k_{cat}/K_m (Ugwumba *et al.*, 2011).

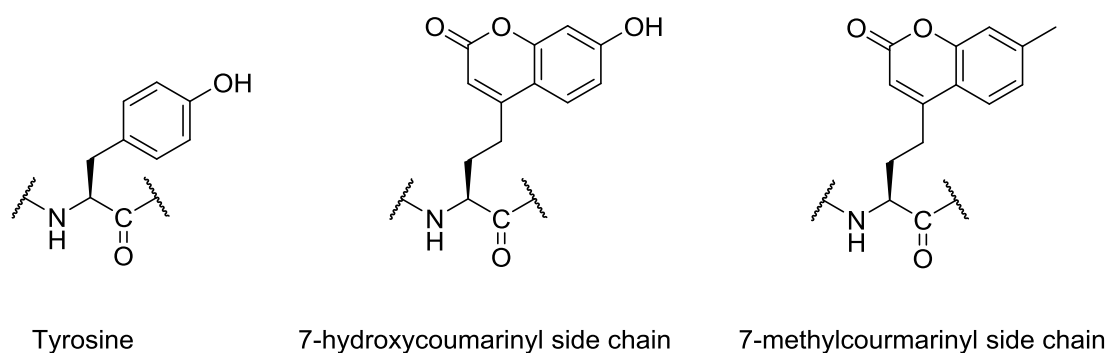


Figure 1.14 Tyrosine analogues 7-hydroxycoumarinyl and 7-methylcoumarinyl inserted at Tyr309 in a bacterial phosphotriesterase. Replacement of tyrosine with 7-hydroxycoumarinyl increase turnover rate of the toxic pesticide paraoxon by 8-fold.

Chemical modification has been used to alter enzyme active sites for decades, for example in the 1960s PMSF was found to modify the active site serine of serine proteases (Fahrney and Gold, 1963). In more recent years alkylation of cysteine residues has been carried out on a number of enzymes to investigate the effects on active site residues. The active site lysine (K107) of fructose-1,6-bisphosphate aldolase (FBPA) is critical for activity, conversion of this to a cysteine (K107C) and then subsequent reaction with bromoethylamine yielded the sulphur containing analogue γ -thialysine. The mutagenesis of the catalytic lysine into a cysteine residue reduced k_{cat}/K_m 700-fold, and subsequent modification to γ -thialysine recovered 80% of wild-type activity (Hopkins *et al.*, 2002). This reduction in activity, of the modified enzyme compared to the wild-type enzyme, was attributed to only 80% conversion of the cysteine residue into γ -thialysine. Therefore the modified γ -thialysine side chain was an excellent mimic of the wild-type lysine, simulating both the structure and function.

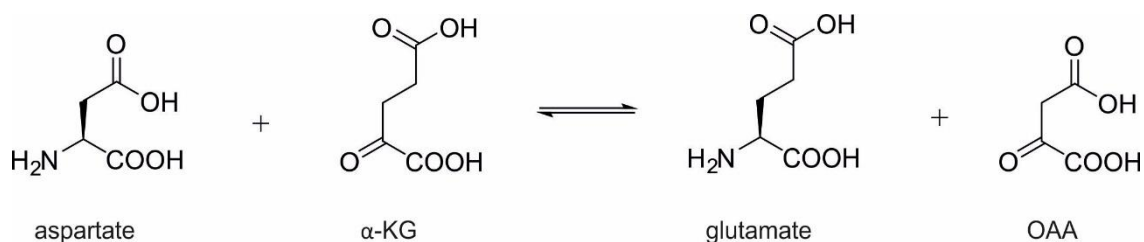


Figure 1.15 Reaction of aspartate aminotransferase. Aspartate amino transferase catalyses the transfer of the amino group from the amino acid aspartate onto the ketoacid α -ketoglutarate (α -KG) to produce the amino acid glutamate and oxaloacetate (OAA).

Replacement of lysine with γ -thialysine has also been carried out in other enzymes. Aspartate aminotransferase, which catalyses the transfer of an amino group from aspartate onto α -ketoglutarate (α -KG) producing glutamate and oxaloacetate (OAA) (Figure 1.15), has a catalytic lysine (K258) which was replaced with γ -thialysine (Gloss and Kirsch, 1995a). Again mutagenesis to the cysteine mutant resulted in a decrease in activity but modification to γ -thialysine only reconstituted 10% of wild-type activity. In this case, conversion of >95% to the γ -thialysine side chain was observed and so incomplete modification was not the cause of decreased activity. It was hypothesised that the change in pK_a of the γ -thialysine, which is 1.3 units lower than that of lysine, produces a less basic residue which is less effective at abstracting a proton from the C_α of the substrate (Gloss and Kirsch, 1995a). This work has

been further extended to insert a variety of lysine and glutamate analogues at position 258 and their effects on activity were determined (Figure 1.16) (Gloss and Kirsch, 1995b). Interestingly it was seen that carboxylate containing nCAAs, increased the binding affinity for the amino acid substrate up to 4-fold.

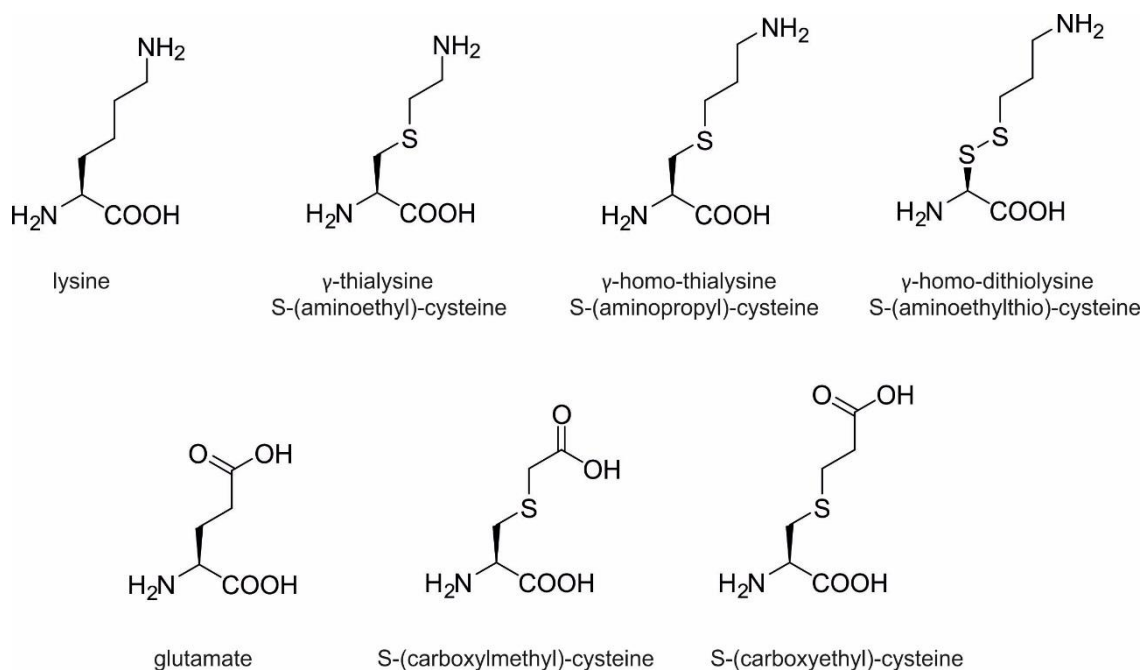


Figure 1.16 Lysine and glutamate non-canonical analogues inserted by alkylation of a cysteine in aspartate amino transferase (Gloss and Kirsch, 1995b).

The enzyme glutathione transferase is a dimeric protein which requires an aromatic phenylalanine or tyrosine from one subunit, to be wedged into a hydrophobic patch of the other subunit for activity. The aromatic Tyr50 was seen to be essential for activity in human glutathione transferase with mutation to alanine or cysteine inactivating the enzyme (Hegazy *et al.*, 2004). The Y50C variant was alkylated with twenty different haloalkanes to generate nCAAs at this position. For the majority of the modifications between 60 and 100 % conversion of the cysteine was achieved, and for all site-specific modification was achieved as assessed by mass spectrometry. Modification of Y50C with iodobutane to produce an S-butyl cysteine residue resulted in a 10-fold increase in k_{cat}/K_m than the wild-type (Y50) enzyme (Hegazy *et al.*, 2006). These modifications as well as being used to increase the activity of the enzyme have been used to investigate the allostery of the enzyme mechanism (Hegazy *et al.*, 2013). From these examples, clearly one of the advantages of using chemical modification to insert nCAAs is

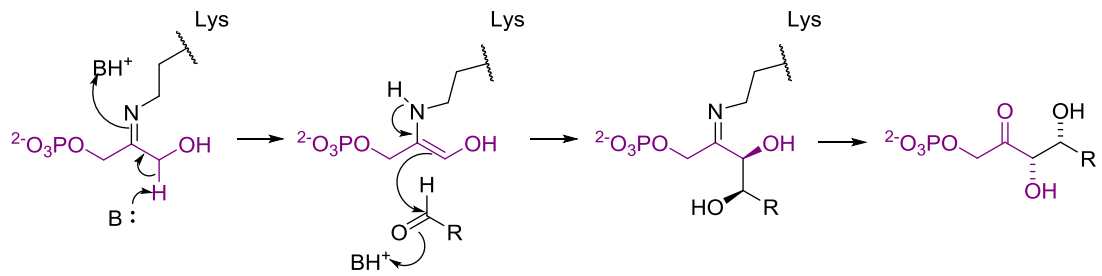
the ease with which multiple different non-canonical side chains can be accessed from one cysteine variant.

With some of these alkylation modification experiments, to allow site-specific modification of one cysteine, it has been necessary to remove cysteine residues that were not to be targeted (Planas and Kirsch, 1991). In other cases if the cysteine residues were buried within the protein they could not be modified and so did not need to be mutated (Hopkins *et al.*, 2002). These alkylation experiments can be hindered by incomplete conversion to the desired ncAA, as can the nonsense suppression method since efficiency of incorporation is often incomplete, resulting in a heterogeneous sample of modified and unmodified protein. However, now with the facile conversion of cysteine into dha and an addition of a thiol it is possible to routinely produce homogenous ncAA containing protein (Chalker *et al.*, 2011, Timms *et al.*, 2013). To alkylate a cysteine residue, it is necessary to have an electrophile with which to modify, this can restrict the side chains that can be produced through this method without dedicated organic synthesis of a range of alkylation reagents. Whereas by incorporating an electrophilic dha, nucleophilic thiol compounds are used to generate the ncAA side chain which allows commercial access to more diverse side chains, as will be demonstrated in this thesis.

1.4 Aldolases

The aldol reaction is one of the foremost reactions in organic chemistry. During the aldol condensation between two carbonyl compounds a new carbon-carbon bond is formed, to produce a new β -hydroxyl carbonyl compound (Machajewski and Wong, 2000). This reaction is of great interest, as aldol structural units are found throughout nature and during the condensation up to two new stereocentres may be formed. The aldol reaction can be catalysed enzymatically by the family of enzymes known as the aldolases. Aldolases are a type of lyase enzyme which can form and cleave carbon-carbon bonds, in the condensation reaction a aldehyde acceptor is added onto a ketone donor. The aldolase mechanism can be split into two different classes; the Class I aldolases form a Schiff bases between a catalytic lysine and the ketone donor, and Class II aldolases which use a zinc ion to stabilise the negative charge on the ketone donor intermediate (Marsh and Lebherz, 1992) (Figure 1.17).

Class I Mechanism



Class II Mechanism

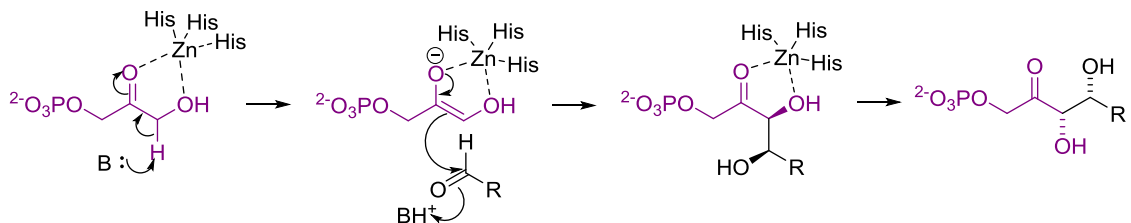
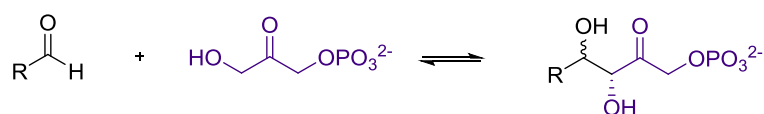


Figure 1.17 Generalised Class I and Class II aldolase mechanisms. Class I and Class II aldolase mechanisms for the condensation of the ketone donor dihydroxyacetone phosphate (DHAP) highlighted in purple, with an aldehyde acceptor. The Class I mechanism shows a Schiff base formed between a lysine and the DHAP ketone donor, The Class II mechanism shows negative charge on the DHAP ketone donor intermediate being stabilised by a zinc ion coordinated by histidine residues.

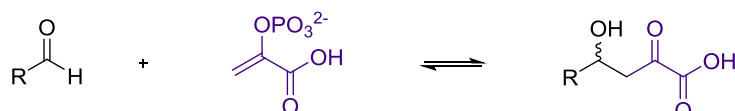
Aldolases can also be classified according to their ketone donor; dihydroxyacetone phosphate (DHAP)-dependent, pyruvate- or phosphoenolpyruvate-dependent, glycine-dependent and acetaldehyde-dependent (Sukumaran and Hanefeld, 2005) (Figure 1.18). DHAP-dependent aldolases catalyse the addition of DHAP onto an aldehyde acceptor substrate. One well known example of this type of aldolase is fructose-1,6-bisphosphate aldolase (FBPA) which plays an important role in glycolysis and gluconeogenesis, catalysing the reversible cleavage of fructose 1,6-bisphosphate into DHAP and glyceraldehyde 3-phosphate. Pyruvate- or phosphoenol-pyruvate dependent aldolases catalyse the addition of pyruvate (or phosphoenol pyruvate) onto an aldehyde acceptor. Glycine-dependent aldolases catalyse the addition of glycine onto an aldehyde acceptor. There are two types of glycine dependent aldolase; threonine aldolases which condense glycine with acetaldehyde to form threonine, and serine hydroxyl methyl transferases which condense glycine with N^5,N^{10} methylene tetrahydrofolate,

and both these enzymes require the cofactor PLP (Contestabile *et al.*, 2001). Acetaldehyde-dependent aldolases catalyse the addition of acetaldehyde onto an aldehyde acceptor. There is only one known example of an acetaldehyde-dependent aldolase: 2-deoxy-D-ribose-5-phosphate aldolase (DERA). DERA is an unusual aldolase, as both the donor and acceptor substrates are aldehydes and so can catalyse two sequential additions of acetaldehyde to the acceptor aldehyde (Dean *et al.*, 2007). Aldolases tend to be highly specific for their donor substrate; however they are far more promiscuous towards aldehyde acceptor substrates

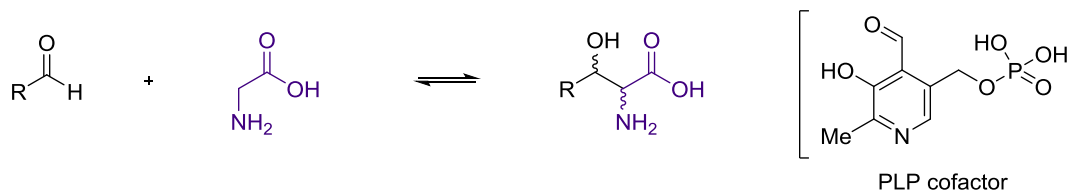
DHAP-dependent aldolases



Pyruvate-/Phosphoenol-pyruvate dependent aldolases



Glycine-dependent aldolases



Acetaldehyde-dependent aldolases (DERA)

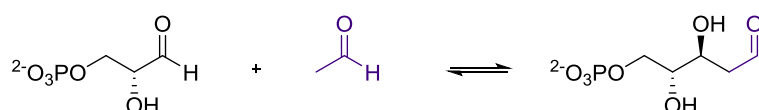


Figure 1.18 Classification of aldolases by their donor substrate. All the aldolases require an aldehyde acceptor substrate and their respective ketone donor; dihydroxyacetone phosphate (DHAP), pyruvate or phosphoenol pyruvate, glycine or acetaldehyde. Reactions for DHAP-, pyruvate- or phosphoenol pyruvate- and glycine-dependent aldolases have been shown with a generalised aldehyde acceptor, the acetaldehyde-dependent aldolase is shown with glyceraldehyde 3-phosphate as the aldehyde acceptor as this is the reaction catalysed by DERA the only known acetaldehyde dependent aldolase.

Aldolase catalysed reactions are an attractive option for the production of biologically significant compounds like carbohydrates, since traditional syntheses are difficult due to the fact these compounds can contain multiple functional groups and tend to be water-soluble (Fesko and Gruber-Khadjawi, 2013). Since there is not always an enzyme available for a desired aldol reaction, there has been significant work towards the evolution of aldolases for synthetically useful reactions has been carried out, reviewed by (Windle *et al.*, 2014, Bolt *et al.*, 2008, Dean *et al.*, 2007, Baker and Seah, 2012). One excellent example of aldolase engineering for synthetically useful compounds was performed with DERA. DERA was found to catalyse the

sequential addition of two equivalents acetaldehyde to chloroacetaldehyde (CAA), which produced a valuable precursor for the production of statin drugs, for example Atorvastatin (Figure 1.19). Unfortunately the enzymatic reaction was hampered by low affinity for CAA and low tolerance to CAA. Therefore a directed evolution approach was used to produce a variant with significantly increased affinity for CAA and 10-fold higher catalytic activity in the presence of 500 mM CAA than the wild-type (Greenberg *et al.*, 2004).

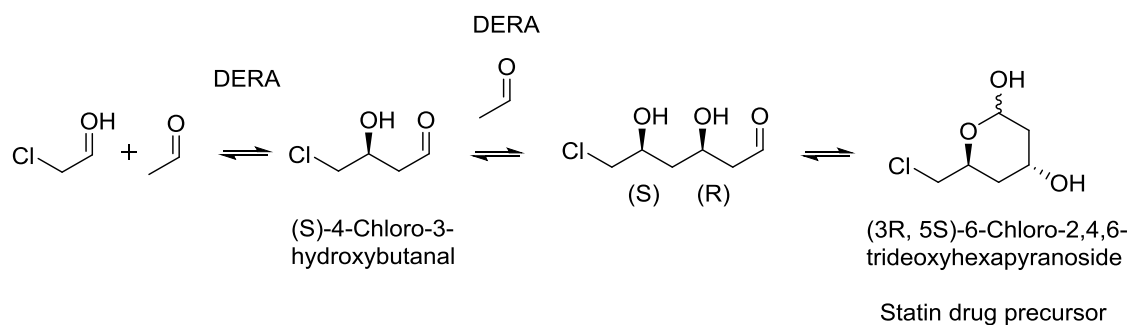


Figure 1.19 DERA catalysed reaction of two equivalents of chloroacetaldehyde to acetaldehyde to produce the statin drug precursor (3R, 5S)-6-Chloro 2,4,6-trideoxyhexapyranoside.

1.4.1 *N*-Acetylneuraminic acid lyase

N-Acetylneuraminic acid lyase (NAL) catalyses the reversible aldol condensation of *N*-acetyl-D-mannosamine (ManNAc) and pyruvate to form *N*-acetylneuraminic acid (Neu5Ac). Neu5Ac is the most abundant of the sialic acids; a group of 9-carbon sugars which modulate a variety of biological processes and are often found at the terminal position of glycoconjugates (Severi *et al.*, 2007, Ji *et al.*, 2015). Neu5Ac also plays a role in host-pathogen interactions and infectivity of the influenza virus, which has led to analogues of Neu5Ac, Tamiflu® and Relenza®, being developed as anti-influenza drugs.

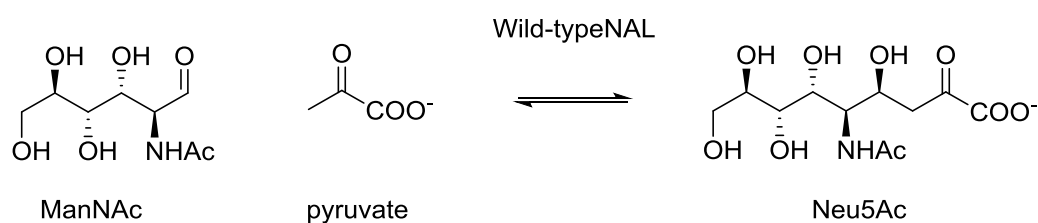


Figure 1.20 Wild-type *N*-acetylneuraminic acid lyase (NAL) catalysed aldol reaction of *N*-acetyl-D-mannosamine (ManNAc) and pyruvate to form *N*-acetylneuraminic acid (Neu5Ac).

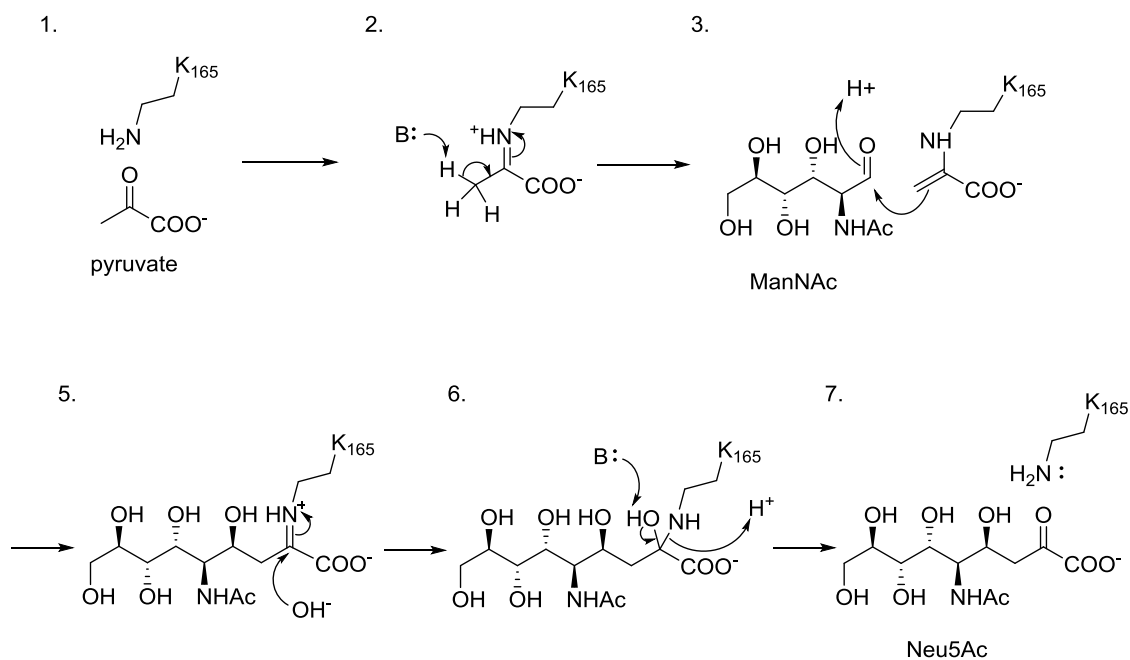


Figure 1.21 Proposed mechanism of *N*-acetylneuraminic acid lyase. Lysine 165 is the catalytic lysine which forms a Schiff base with the pyruvate donor. The Schiff base enamine then attacks the aldehyde carbon of *N*-acetyl-D-mannosamine (ManNAc). Schiff base hydrolysis then releases the product *N*-acetylneuraminic acid (Neu5Ac).

NAL is a Class I aldolase and has been shown to follow an ordered bi-uni aldol condensation kinetic mechanism (Groher and Hoelsch, 2012). Pyruvate binds in the active site forming a Schiff base with the catalytic lysine residue at position 165. The enamine Schiff base complex can then attack the aldehyde carbon of ManNAc to form the new carbon-carbon bond of the β -hydroxy compound. Subsequent hydrolysis of the Schiff base then releases the Neu5Ac product.

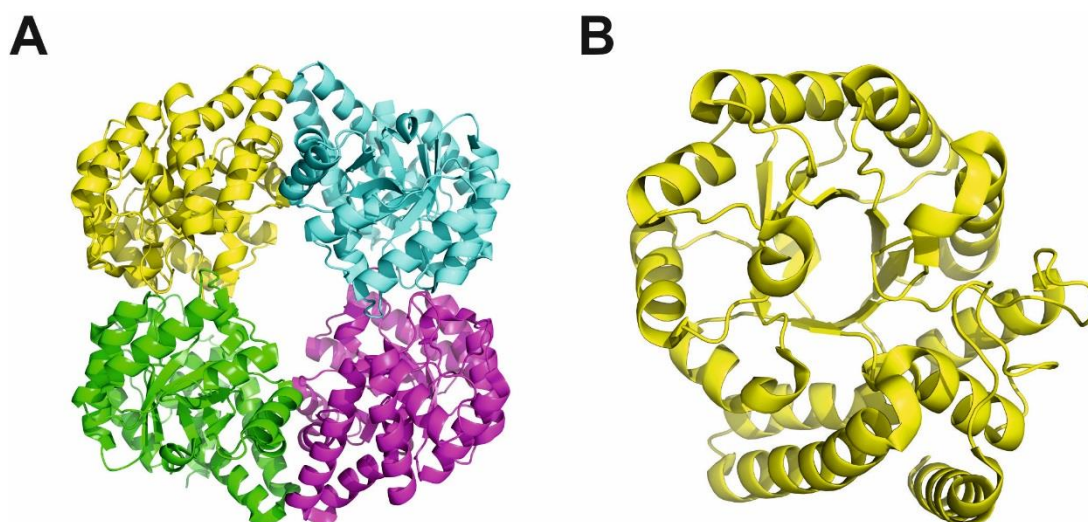


Figure 1.22 Crystal structure of the *E. coli* NAL (pdb code 4BWL) (Daniels *et al.*, 2014). **A** shows the homotetramer of NAL and **B** shows the $(\alpha/\beta)_8$ TIM barrel fold of each NAL monomer.

A number of studies have been carried out on NAL, with NALs from *Clostridium perfringens* and *E. coli* being cloned and expressed (Nees *et al.*, 1976, Uchida *et al.*, 1984). There have also been a number of X-ray crystallographic studies carried out; crystal structures of the *Haemophilus influenzae* and *E. coli* NALs have been solved (Barbosa *et al.*, 2000, Izard *et al.*, 1994). These studies have shown NAL to adopt a tetrameric $(\alpha/\beta)_8$ TIM barrel structure. X-ray crystallographic studies of NAL with the substrate pyruvate bound in the active site have shown the presence of the Lys165-pyruvate Schiff base in the active site (Campeotto *et al.*, 2010, Lawrence *et al.*, 1997). A variety of substrate analogues have also been solved in the crystal structures including; 4-oxo-Neu5Ac in the active site of the *Haemophilus influenzae* NAL (Barbosa *et al.*, 2000), and a dipropylamide analogue of ManNAc in the active site of the *E. coli* E192N variant NAL (Campeotto *et al.*, 2010), also recently the X-ray crystal structure of the *E. coli* Y137A variant NAL was solved in complex with Neu5Ac (Daniels *et al.*, 2014). These studies, and recent QM/MM studies have been able to highlight catalytically important residues within the active site of NAL (Daniels *et al.*, 2014). The hydroxyl of Tyr137, a residue which is catalytically essential, has been shown to be the proton donor for the ManNAc aldehyde oxygen in the wild-type *E. coli* NAL reaction (3. in Figure 1.21) (Daniels *et al.*, 2014, Barbosa *et al.*, 2000). The residues Tyr110 and Ser47 are thought to act as a proton shuttle to aid Tyr137 as the proton donor, and Thr167 is involved in stabilising the transition state by

forming hydrogen bonds with the aldehyde oxygen of ManNAc, which becomes the C4 hydroxyl of Neu5Ac (Daniels *et al.*, 2014).

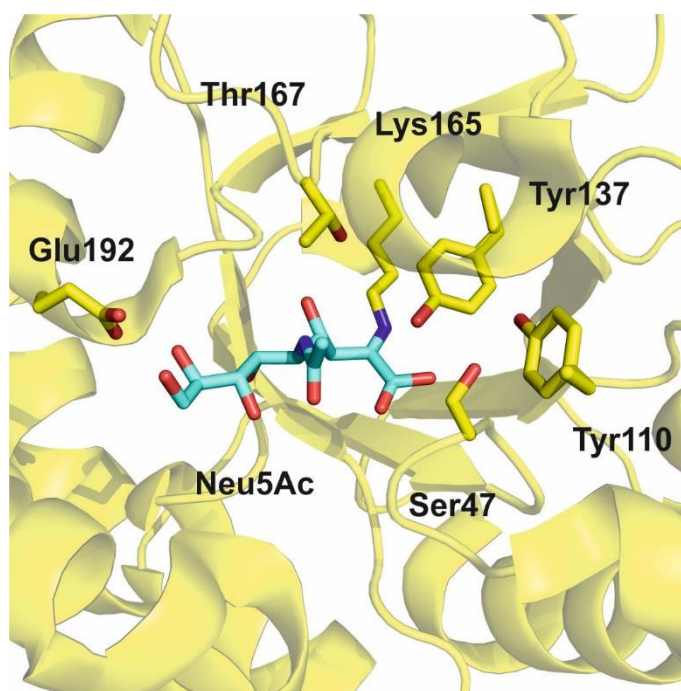


Figure 1.23 Active site of the Y137A *E. coli* NAL variant in complex with the product Neu5Ac. Important active site residues have been highlighted as sticks and the alanine at position 137 has been modelled back in as a tyrosine. Glu192 is involved in stabilising the glycerol end of Neu5Ac, Lys165 forms a Schiff base with the pyruvate end of Neu5Ac, Ser47 and Tyr110 provide a proton shuttle for the proton donor Tyr137 and Thr167 hydrogen bonds to the C4 hydroxyl and stabilises the transition state (Daniels *et al.*, 2014).

1.4.2 Protein engineering of *N*-acetylneuraminic acid lyase

Since there is a significant amount of structural and mechanistic knowledge of NAL, and the ability to form new carbon-carbon bonds is extremely desirable, it is an excellent target for protein engineering studies. NAL has been shown to be highly specific for the ketone donor, pyruvate, with only a few reports of it utilising 3-fluoropyruvate as well (Daniels *et al.*, 2014, Watts and Withers, 2004, Chokhawala *et al.*, 2007). However NAL has been shown to be far more promiscuous towards a variety of aldehyde acceptor substrates, accepting various C6, C5 and C4 monosaccharides (Gijzen *et al.*, 1996, Fitz *et al.*, 1995, Kim *et al.*, 1988). NAL has been the focus of multiple different protein engineering studies, changing both the substrate specificity and the stereochemistry of the enzyme (Hsu *et al.*, 2005, Wada *et al.*, 2003, Williams *et al.*, 2005, Williams *et al.*, 2006, Joerger *et al.*, 2003).

Rational redesign, using structural comparison, was used to switch the specificity of *E. coli* NAL from ManNAc and pyruvate to L-aspartate- β -semialdehyde and pyruvate, a reaction usually catalysed by the enzyme dihydrodipicolinate synthase, another class I aldolase (Blickling *et al.*, 1997). The variant L142R showed a 19-fold increase in activity with L-aspartate- β -semialdehyde over the wild-type NAL (Joerger *et al.*, 2003). The *E. coli* NAL has been shown to be highly selective for D-sialic acid, and so directed evolution has been used to relax the stereochemistry of the *E. coli* NAL. These experiments produced an enzyme capable of accepting L-sugars as well as D-sugars by the mutation V251I (Wada *et al.*, 2003). Insertion of other proteogenic amino acids at this position produced variants with enhanced activities for a variety of different substrates (Chou *et al.*, 2011). This work was further developed to produce an enzyme with high specificity towards L-sugars, both L-sialic acid and an analogue L-3-deoxy-manno-2-octulosonic acid (L-KDO), showing a 30-fold improvement in specificity to L-KDO over that of the wild-type NAL (Hsu *et al.*, 2005). One of the most notable examples of altering the substrate specificity of *E. coli* NAL was a semi-rational redesign experiment to increase the activity of NAL with a dipropylamide analogue of Neu5Ac, [4*R*,5*R*,6*R*]-7-(dipropylamino)-4,5,6-trihydroxy-2,7-dioxoheptanoic acid ([4*R*]-DPAH) (Williams *et al.*, 2005). Saturation libraries of residues along the aldehyde binding end of the active site; D191, E192 and S208 were produced and screened for activity with DPAH. The best variant, E192N, exhibited a 640-fold switch in k_{cat}/K_m from Neu5Ac to ([4*R*]-DPAH, this variant also had a k_{cat}/K_m for ([4*R*]-DPAH that was higher than that of the wild-type *E. coli* NAL with Neu5Ac (Williams *et al.*, 2005). The E192N variant was shown to exhibit poor stereocontrol at the C4 position in the aldol condensation of (2*R*,3*S*)-2,3-dihydroxy-4-oxo-*N,N*-dipropylbutanamide (DHOB) and pyruvate to form either [4*R*]-DPAH or [4*S*]-DPAH. Therefore directed evolution was used, with E192N as a starting point, to produce a pair of highly stereoselective variants (Williams *et al.*, 2006).

NAL has been the focus of many different studies, which has provided extensive mechanistic and structural knowledge of this enzyme. The use of enzymes for the synthetic production of carbon-carbons bonds is also highly desirable due to their high specificity and ability to work under ambient conditions. Since NAL has already been shown to be a highly malleable enzyme through traditional protein engineering experiments, it may be an ideal candidate for the incorporation of ncAAs and subsequent alteration of activity.

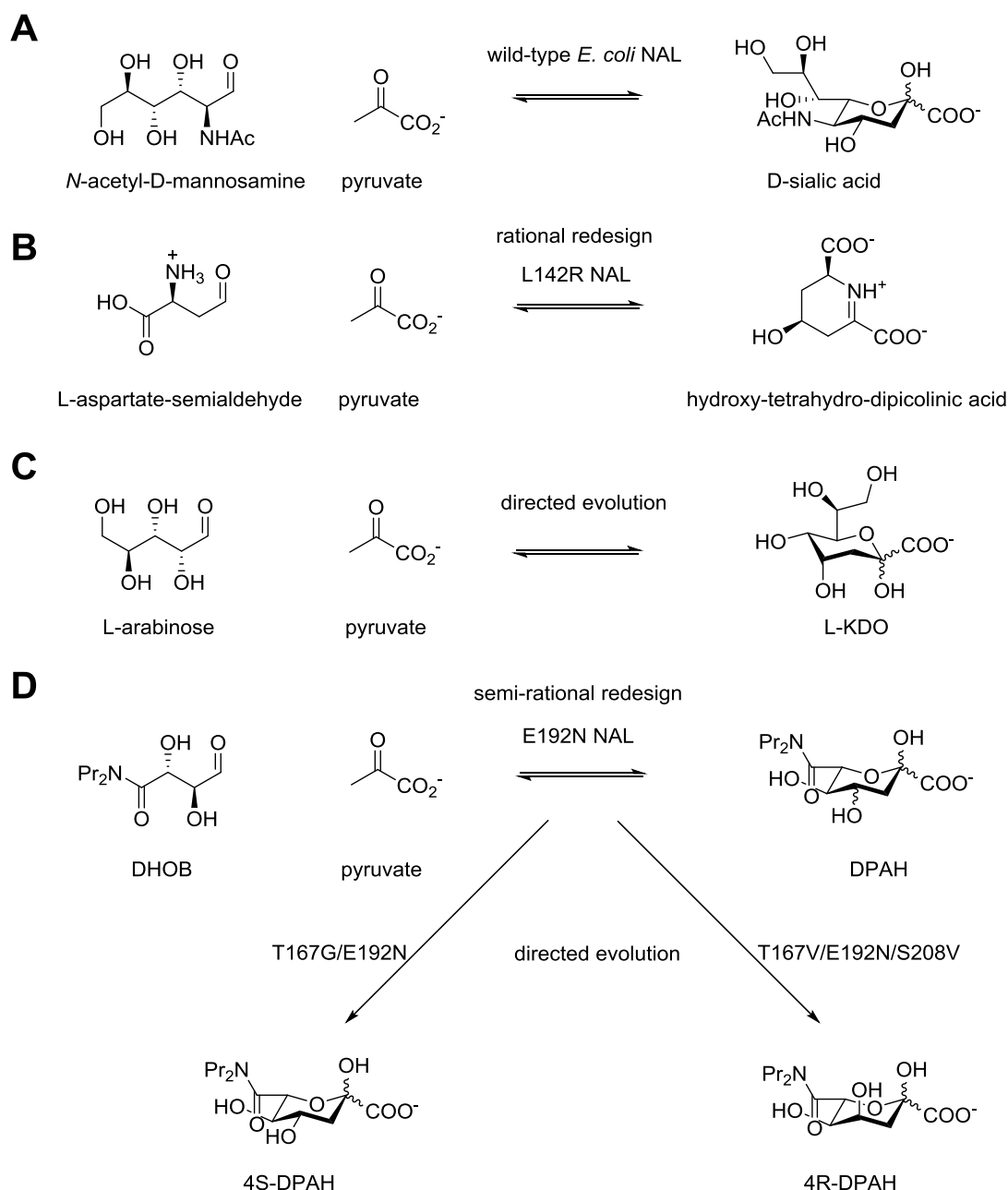


Figure 1.24 Protein engineering studies on the *E. coli* NAL. A shows the wild-type reaction of NAL, the reversible aldol condensation of *N*-acetyl-D-mannosamine and pyruvate to form *N*-acetylneuraminic acid or D-sialic acid (Neu5Ac). B shows the condensation of L-aspartate semialdehyde and pyruvate to form hydroxyl-tetrahydro-dipicolinic acid, this activity was developed in NAL by rational redesign to confer dihydrodipicolinate synthase activity onto NAL by the variant L142R (Joerger *et al.*, 2003). C shows the condensation of L-arabinose and pyruvate to form L-3-deoxy-*manno*-2-octulosonic acid (L-KDO), this activity was developed in NAL by directed evolution (Hsu *et al.*, 2005). D shows the condensation of (2*R*,3*S*)-2,3-dihydroxy-4-oxo-*N,N*-dipropylbutanamide (DHOB) and pyruvate to form DPAH a dipropylamide analogue of Neu5Ac. This activity was developed in NAL by screening saturation libraries of a variety of positions in the active site (Williams *et al.*, 2005), the best variant E192N was then subjected to directed evolution to produce a pair of stereocomplementary enzymes (Williams *et al.*, 2006).

1.5 Thesis aims and objectives

This thesis will report the use of a chemical method to site-specifically introduce ncAAs into the active site of the *Staphylococcus aureus* NAL (*sa*NAL). Subsequent use of this approach was used to examine the effects of ncAA incorporation and also to alter the substrate specificity of the enzyme. The objectives of this work were as follows;

1.5.1 To determine structural and functional effects of replacing the catalytic lysine at position 165 with the ncAA γ -thialysine.

Previous work within the group had identified the cysteine free *sa*NAL as the ideal system for the incorporation of ncAAs by a chemical modification method. This work had shown that by incorporation of a cysteine residue into *sa*NAL by site-directed mutagenesis, it was possible to achieve site-specific insertion of an ncAA via a dehydroalanine intermediate at a desired position (Nicole Timms, PhD thesis, University of Leeds). Conversion of the catalytic lysine at position 165 into γ -thialysine had been previously carried out in *sa*NAL. Therefore this chemical modification will be confirmed, and used to produce high yields of modified protein for detailed kinetic analysis to be carried out. High yield of chemically modified protein will also allow detailed structural analysis, using X-ray crystallographic methods, of the modified protein to be performed. Using this approach it will be possible to determine the structural and catalytic effects of replacing a catalytic proteogenic amino acid with a non-canonical analogue.

1.5.2 Develop methods for high-throughput incorporation of ncAAs by chemical modification, and screening of modified variants.

By using a chemical modification method to insert ncAAs, multiple different ncAAs will be individually incorporated site-specifically into the active site of *sa*NAL, producing many different modified variants. These modified enzymes can then be screened for activity with a variety of different substrates, to investigate the effect of ncAAs on substrate chain length, stereochemistry and functional group substitutions. This approach will allow a significant number of modified enzymes to be both produced and screened.

1.5.3 To characterise modified enzymes with altered substrate specificities brought about by incorporation of a non-canonical side chain.

Once modified ncAA-containing enzymes have been screened for altered activities, any 'hits' will be taken forward and their altered activity can be analysed. Kinetic analysis will allow the extent that the substrate specificity has been changed to be determined. Structural analysis may elucidate a potential mechanism by which the ncAA has altered the enzyme activity.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Technical equipment

Equipment

Manufacturer

Centrifuges

Avanti J-26 XP Centrifuge

Beckman Coulter, Brea, CA, USA

GenFuge 24D Centrifuge

Progen Scientific, London, UK

Micro Centaur

MSE, London, UK

Incubators, mixers & shakers

Gallenkamp Economy Incubator Size 1

Sanyo, Watford, UK

ORBISAFE Orbital Incubator

Sanyo, Watford, UK

Stuart Magnetic Stirrer SB161

CamLab, Cambridge, UK

Stuart Orbital Incubator S150

Bibby Scientific, Stone, UK

MS1 Minishaker

IKA®, Staufen, Germany

Vortex Genie 2

Scientific Industries, New York, USA

Gel electrophoresis equipment

Vari-Gel midi system

CamLab, Cambridge, UK

Mini-PROTEAN® Electrophoresis system

Bio-Rad laboratories, Hertfordshire,
UK

PowerPac™ Basic power supply

Bio-Rad laboratories, Hertfordshire,
UK

Protein purification equipment

ÄKTAprime plus

GE healthcare, Little Chalfont, UK

Superdex™ S200 26/60 gel filtration column

GE healthcare, Little Chalfont, UK

Spectrophotometer

UVIKON 930

Kontron Instruments, UK

Microplate reader

FLUOstar Galaxy plate reader

BMG Labtech, Aylesbury, Bucks, UK

Other equipment

Jenway 3020 pH meter

Bibby Scientific, Stone, UK

PTC-100 Programmable Thermal Controller

GMI-Inc, Minneapolis, MN, USA

Techne Dri-Block Heater DB-2A

Bibby Scientific, Stone, UK

Grant JB1 Unstirred Waterbath

Grant Instruments, Shepreth, UK

2.1.2 Chemicals

Agarose

Melford Laboratories, Suffolk, UK

Ammonium Acetate

Fisher Scientific, Loughborough, UK

Ammonium Bicarbonate

Acros Organics (Fisher Scientific)

D-Arabinose

Sigma-Aldrich, Dorset, UK

Agar

Melford Laboratories, Suffolk, UK

Agarose

Melford Laboratories, Suffolk, UK

Ammonium persulfate

Acros Organics (Fisher Scientific)

Acetone

Fisher Scientific, Loughborough, UK

Acrylamide, 30%

Severn Biotech, Kidderminster, UK

Ampicillin

Formedium, Norfolk, UK

N-Acetyl- D-Glucosamine

Sigma-Aldrich, Dorset, UK

N-Acetyl- D-mannosamine

Carbosynth, Berkshire, UK

N-Acetylneuraminic acid

Carbosynth, Berkshire, UK

Bovine serum albumin

Bio-Rad, Hertfordshire, UK

Bromophenol blue

Sigma-Aldrich, Dorset, UK

β -mercaptoethanol

Stratagene, Cambridge, UK

Coomassie Brilliant Blue

BDH Biochemical, Poole, UK

Benzyl Mercaptan

Sigma-Aldrich, Dorset, UK

Bradford reagent

Bio-Rad, Hertfordshire, UK

L-Cysteine

Sigma-Aldrich, Dorset, UK

D-Cysteine hydrochloride

MP Biomedicals, Cambridge, UK

Cyclohexanone	Sigma-Aldrich, Dorset, UK
<i>N,N</i> -Dimethylformamide	Sigma-Aldrich, Dorset, UK
Dimethyl sulphoxide, DMSO	Sigma-Aldrich, Dorset, UK
1,2-Dithiothreitol, DTT	Formedium, Norfolk, UK
Ethanol	Fisher Scientific, Loughborough, UK
Ethylenediamine tetra acetic acid, EDTA	Sigma-Aldrich, Dorset, UK
D-Erythrose	Sigma-Aldrich, Dorset, UK
2-Furanmethanethiol	Sigma-Aldrich, Dorset, UK
Formic acid	Sigma-Aldrich, Dorset, UK
D-Galactose	Fluka (Sigma-Aldrich)
Glycerol	Fisher Scientific, Loughborough, UK
Glycine	Fisher Scientific, Loughborough, UK
D-Glucose	Fisher Scientific, Loughborough, UK
Isopropyl β -D-1-thiogalactopyranoside	Generon, Berkshire, UK
Imidazole	Acros Organics (Fisher Scientific)
Lactate Dehydrogenase	Sigma-Aldrich, Dorset, UK
Lysozyme-from chicken egg white	Sigma-Aldrich, Dorset, UK
D-Lyxose	Sigma-Aldrich, Dorset, UK
Mannose	MP Biomedicals, Cambridge, UK
Mercaptopropionic acid	Sigma-Aldrich, Dorset, UK
MES monohydrate	Alfa Aesar, Lancashire, UK
2-Mercaptoethanol	Sigma-Aldrich, Dorset, UK
4-Mercapto-1-butanol	Sigma-Aldrich, Dorset, UK
1-Mercapto-2-propanol	Sigma-Aldrich, Dorset, UK
4-Mercapto-1-butanol	Sigma-Aldrich, Dorset, UK
Methylthioglycolate	Sigma-Aldrich, Dorset, UK
3-Mercaptopropionic acid	Sigma-Aldrich, Dorset, UK
β -Nicotinamide Adenine Dinucleotide reduced	Sigma-Aldrich, Dorset, UK
Orthophosphoric acid	Sigma-Aldrich, Dorset, UK
2-Propene-1-thiol	Sigma-Aldrich, Dorset, UK
Polyethylene glycol 3350	Sigma-Aldrich, Dorset, UK
Polyethylene glycol 400	Sigma-Aldrich, Dorset, UK
Propan-2-ol	Fisher Scientific, Loughborough, UK
D-Ribose	Fluka (Sigma-Aldrich)
Sodium chloride, NaCl	Fisher Scientific, Loughborough, UK

Sodium dodecyl sulphate, SDS	Sigma-Aldrich, Dorset, UK
Sodium hydroxide, NaOH	Fisher Scientific, Loughborough, UK
Sodium phosphate dibasic, Na ₂ HPO ₄	Sigma-Aldrich, Dorset, UK
Sodium phosphate dibasic, NaH ₂ PO ₄	Sigma-Aldrich, Dorset, UK
Tetramethylethylenediamine (TEMED)	Sigma-Aldrich, Dorset, UK
Tris Base	Fisher Scientific, Loughborough, UK
Tryptone	Fluka (Sigma-Aldrich)
Trichloroacetic acid	Sigma-Aldrich, Dorset, UK
2-Thiobarbituric acid	Sigma-Aldrich, Dorset, UK
Urea	Sigma-Aldrich, Dorset, UK
D-Xylose	Sigma-Aldrich, Dorset, UK
Yeast extract	Fluka (Sigma-Aldrich)

2,5-dibromohexan-1,6-diamide (diBr) was synthesised and supplied by Alun Myden (University of Leeds, UK) and Robert Smith (University of Leeds, UK).

2.1.3 Media

Unless otherwise stated all bacterial cultures were grown in 2 TY media which had been sterilised by autoclaving at 121°C for 20 min. 1 L of media was prepared using distilled water and contained 16g tryptone, 10 g yeast, 5 g NaCl, if solid phase media was required this was supplemented with 1.5% (w/v) agar.

2.2 General Methods

2.2.1 Centrifugation

Centrifugation of samples above 1.5 mL and 96 well plates was performed in an Avanti J-26 XP Centrifuge (Beckman Coulter, Brea, CA, USA) at a temperature of 4 °C unless otherwise stated.

Centrifugation of samples <1.5 mL was performed using either a GenFuge 24D Centrifuge (Progen Scientific, London, UK) or a SANYO Micro Centaur Centrifuge (MSE, London, UK)

2.2.2 pH measurements

pH measurements were performed using a Jenway 3020 pH meter (Bibby Scientific, Stone, UK) calibrated according to the Manufacturer's guidelines.

2.2.3 Aseptic technique

Aseptic technique was used throughout. All heat resistant materials and both solid and liquid phase media were sterilised by autoclaving at 121 °C for 20 min. Solutions which were heat labile were sterilised by filter sterilisation using Millex® Syringe Filter unit (EMD Millipore, Hertfordshire, UK) and equipment which could not be heat sterilised was cleaned down with 70 % (v/v) ethanol.

2.2.4 Antibiotic stocks

Ampicillin was the antibiotic used throughout at a final concentration of 50 µg mL⁻¹ in both solid and liquid phase media. Ampicillin was made to a stock concentration of 50 mg mL⁻¹ using distilled water; this stock was sterile filtered before use or before storage at -20 °C.

2.2.5 Culture growth

Single colonies were picked from 2 TY agar plates supplemented with 50 µg mL⁻¹ ampicillin and used to inoculate 5 mL starter cultures of 2 TY supplemented with 50 µg mL⁻¹ ampicillin. These starter cultures were incubated for 16 hours at 37°C, with shaking at 200rpm in an orbital incubator. 50 µL of starter culture was used to inoculate a 50 mL day culture (unless stated otherwise) and incubated for 8 hours at 37°C, with shaking at 200rpm in a 250 mL conical flask. 10 mL of day culture was used to inoculate 1 L of 2 TY supplemented with 50 µg mL⁻¹ ampicillin in a 2.5 L conical flask. Cultures were incubated at 37°C, with shaking at 200rpm until an optical density (O.D) of 0.6 at 600 nm was reached, and then protein expression was induced by addition of IPTG to a final concentration of 0.1 mM.

2.2.6 Glycerol stocks

Glycerol stocks were made from starter cultures that had been incubated for 16 hours at 37°C, with shaking at 200rpm. A 1:1 v/v solution of culture to sterile glycerol was made and stored in a sterile Nunc CryoTube™ (Thermo Fisher Scientific) at -80°C. Bacteria could then be streaked directly from glycerol stocks onto 2TY agar plates and single colonies picked for inoculation of starter cultures.

2.3 DNA Methods

2.3.1 Bacterial strains and plasmids

Bacterial strains used in DNA manipulation were;

***E. coli* XL1-Blue** (Agilent Technologies, Cheshire, UK)

recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' *proAB lacI^qΔM15 Tn10* (Tet^r)]

***E. coli* XL10-Gold** (Agilent Technologies, Cheshire, UK)

Tet^rΔ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte* [F' *proAB lacI^q ΔM15 Tn10*(Tet^r)AmyCam^r]

Bacterial strains used for protein expression were;

***E. coli* BL21-Gold (DE3)** (Agilent Technologies, Cheshire, UK)

(B⁻ F⁻ *ompT hsdS*(r_B⁻ m_B⁻) *dcm⁺ Tet^r gal λ*(DE3) *endA Hte*

The p*KnanA_S.aureus*-His₆ plasmid was provided by Professor A. Berry (University of Leeds, UK)

2.3.2 Plasmid purification

5mL starter cultures, of 2 TY media supplemented with 50 μg mL⁻¹ ampicillin, were inoculated with either single colonies or glycerol stocks of bacteria and grown for 16 hours at 37°C, 200rpm in an orbital incubator. Plasmid DNA was then purified from these starter cultures using Wizard® Plus SV Minipreps DNA Purification System (Promega, Southampton) according to manufacturer's guidelines.

2.3.3 Agarose gel electrophoresis

Agarose gel electrophoresis was performed with 0.7 % (w/v) agarose gel for nucleic acids <1.5 kb. Agarose gels were prepared by dissolving agarose in TAE buffer and run with a 1 kb ladder for determining the size of DNA fragments.

10 x TAE buffer

10 mM EDTA

200 mM Glacial acetic acid

400 mM Tris base

2.3.4 DNA quantification

DNA quantification was carried out spectrophotometrically at 260nm. An absorbance reading of 1.0 is equal to 50 ng μL^{-1} of double stranded DNA. Agarose gels were performed with an appropriate

2.3.5 DNA sequencing

DNA sequencing was carried out by Beckman Coulter Genomics (Essex, UK)

2.3.6 Site-directed Mutagenesis

Mutagenesis was carried out using QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies LDA UK Limited). Site-directed mutagenesis primers were designed as per the manufacturer's guidelines.

2.3.7 Transformation

Transformations were carried out using a heat shock method according to the method provided with the mutagenesis kit. Plasmid DNA was transformed into both *E. Coli* XL10-Gold Ultracompetant cells and *E. Coli* BL21-Gold (DE3) cells.

2.4 Protein methods

2.4.1 Affinity chromatography purification of His₆-tagged proteins

E. coli cells with the plasmid containing the target gene for protein expression were cultured as in Section 2.2.5. His₆-tagged proteins were purified using Chelating Sepharose™ fast flow resin, chelated with nickel ions, in a batch purification method.

Cell pellet was harvested from 1 L of bacterial culture by centrifugation at 12 000 g for 20 min. The cell pellet was re-suspended in 50 mL of washing buffer per 10 g of cells using a manual homogeniser. The re-suspended cells were then lysed at 20 KPSI using a Cell Disruptor from Constant Cell Disruption Systems (Northamptonshire, UK). Insoluble cell debris was separated from the soluble protein by centrifugation at 40 000 g for 45 min at 4 °C.

The supernatant containing the soluble protein fraction was then loaded onto 5 mL of Chelating Sepharose™ fast flow resin in a 50 mL falcon tube, and incubated for 30 min with agitation at 4 °C. Centrifugation at 4 000 g for 6 min at 4 °C produced a supernatant containing any non-bound proteins, which was discarded. The resin was then thoroughly re-suspended

using 40 mL of washing buffer and centrifuged at 4 000 g for 6 min at 4 °C after which the supernatant was discarded, this step was repeated three times. Elution of the His₆-tagged protein was achieved by addition of 20 mL elution buffer and incubation for 1 hour at 4 °C with agitation, after centrifugation at 4 000 g for 6 min at 4 °C the supernatant containing the His₆-tagged protein was collected. The protein was then dialysed (Section 2.4.3) either against 20 mM Tris/HCl then sterile filtered and stored at 4 °C, or against 20 mM ammonium acetate pH 7.0 and lyophilised (Section 2.4.6)

Washing buffer

50 mM Tris/HCl pH7.4

20 mM Imidazole

500 mM NaCl

Elution buffer

50 mM Tris/HCl pH7.4

500 mM Imidazole

500 mM NaCl

2.4.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) was performed with resolving gel containing 15% acrylamide and stacking gel containing 3.75% acrylamide. Gels were made up containing the components listed below, the two polymerisation initiators, APS and TEMED, were the final components to be added. Gels were set between two glass plates. The resolving gel was set first with a layer of isopropanol to prevent drying, once the resolving gel was set the isopropanol was discarded and the stacking gel added with a comb inserted to form sample wells.

Protein samples were prepared for SDS PAGE by addition of an equal volume of 2 x loading buffer (recipe below) incubating at 100 °C for 5 min and then centrifugation for 5 min at 13 000 g before being loaded onto the gel. Separation by electrophoresis was carried out at 30-60 mA for approximately 1-2 hour. SDS-PAGE gels were then stained for 1 hour using a methanol : acetic acid : water (5:1:1, v/v/v) solution containing 0.1 % (w/v) coomassie Brilliant Blue R250 solution and then destained using the same solution without the Coomassie Brilliant Blue R250. For more rapid development SDS-PAGE gels were also stained using Quick Coomassie Stain (Generon, Berkshire, UK).

Resolving gel:

7,500 μ L 30% (w/v) acrylamide

3,750 μ L 1.5 M Tris/HCl pH8.8

3,500 μ L H₂O

150 μ L 10% (w/v) SDS

50 μ L 25% (w/v) APS

5 μ L TEMED

Stacking gel:

625 μ L 30% (w/v) acrylamide

625 μ L 1.0 M Tris/HCl pH 6.9

3,650 μ L H₂O

50 μ L 10% (w/v) SDS

50 μ L 25% (w/v) APS

5 μ L TEMED

Running buffer:

14.4 g glycine

3 g Tris base

1g SDS

140 μ L β -mercaptoethanol

H₂O made up to 1L

2 x loading buffer for SDS PAGE:

2 mL 10% (w/v) SDS

200 μ L 0.2% (w/v) bromophenol blue in ethanol

154 mg DTT

1 mL glycerol

170 μ L 1 M Tris/HCl pH 6.9

163 μ L H₂O

2.4.3 Protein dialysis

Dialysis of proteins was carried out using BioDesign dialysis tubing (Thermo Fisher Scientific, Loughborough, UK) with a 14 kDa molecular weight cut off. Proteins were dialysed into an appropriate buffer for the downstream application of the protein. The volume of dialysis buffer used was at least 500 fold larger than the protein sample to be dialysed. Dialysis was carried out with stirring either at room temperature for 2 hours or 4 °C for 4 hours, then the dialysis buffer was replaced with fresh buffer and the process repeated.

2.4.4 Protein concentration determination

Protein concentration was determined by reading the protein sample absorbance at 280 nm. The protein concentration was then calculated using the Beer Lambert law.

$$A = \epsilon cl$$

Equation 1. Beer-Lambert law for calculation of protein concentration. A is the absorbance value at 280 nm, ϵ is the molar extinction coefficient of the protein, l is the path length and c is the concentration of the protein.

To determine the protein concentration of a protein with very few aromatic groups, or when imidazole was present, the Bradford assay was used. Bradford reagent (Bio-Rad Laboratories, Hertfordshire, UK) was used as per manufacturer's guidelines. A calibration curve was produced using a range of known concentrations of bovine serum albumin (BSA) and this was used to determine unknown protein concentrations.

2.4.5 Increasing protein concentration

Protein samples were concentrated using an appropriately sized Vivaspin Protein Concentrator (10 kDa molecular weight cut off) (Generon, Berkshire, UK) for the sample volume. Proteins were concentrated as per the manufacturer's guidelines

2.4.6 Lyophilisation

Proteins were first dialysed (Section 2.4.3) into 50 mM ammonium acetate pH 7.0. Once dialysis was complete the protein concentration was determined and aliquots containing between 2 and 20 mg of protein were pipetted into either 1.5 mL eppendorfs or 15 mL falcon tubes with pierced lids. The aliquots were then flash frozen in liquid nitrogen before being lyophilised using a Thermo Electron Corporation Heto PowerDry PL300 freeze dryer

2.4.7 Sample preparation for ESI mass spectrometry

Protein samples were prepared for mass spectrometry by desalting 70 μL of sample into 50 mM ammonium acetate pH 7.0. Samples were prepared using Micro Bio-Spin™ Chromatography Columns (Bio-rad Laboratories, Hertfordshire, UK); columns were equilibrated with 4 column volumes of 50 mM ammonium acetate pH 7.0 before use as per manufacturer's guidelines. For the preparation of protein samples which were taken from chemical modification reactions, 70 μL was desalted through two desalting columns equilibrated with 50 mM ammonium acetate pH 7.0.

2.4.8 ESI mass spectrometry of proteins

Mass spectrometry of proteins was performed by Dr James Ault using a quadrupole-ion mobility-orthogonal time of flight mass spectrometer (Synapt G2-S, Waters, Manchester, UK). Samples were analysed in acetonitrile : 1 % aqueous formic acid (50:50 v/v). Nano Flow electrospray ionisation was used and the mass spectrometer operated in positive mode time of flight (TOF). A capillary voltage of 1.2 kV and cone voltage of 50 V was used, nano-electrospray nitrogen gas pressure of 0.1 bar and backing pressure of 1.78 mbar. The source and desolvation temperatures were set at 80 °C and 150 °C, respectively. Nitrogen was used as buffer gas at a pressure of 8.0×10^{-3} mbar in the trap and transfer regions and 3.6×10^{-4} mbar in the ion mobility cell. Mass calibration was performed using sodium iodide ($2 \mu\text{g } \mu\text{L}^{-1}$) and data was analysed using Mass Lynx v4.1 software.

2.4.9 Conversion of cysteine to dehydroalanine by 2, 5-dibromohexan-1, 6-diamide (diBr)

2.5 mg of lyophilised cysteine-containing *s*αNAL was resuspended thoroughly in 1.25 mL prewarmed (37 °C) 50 mM sodium phosphate buffer pH 8.0 containing 6 M urea. A stock solution was made of 0.13 mg μL^{-1} or 0.08 mg μL^{-1} diBr in DMF, the lower concentration of diBr was used if alkylation adducts were seen in the modification as discussed in Section 4.3.1. 100 μL of diBr stock was added to the protein solution followed by vortexing for 30 s. This solution was then left to incubate for 1 hr 30 min at 37 °C, with shaking at 200rpm in an orbital incubator. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the cysteine to dehydroalanine had occurred; if full conversion had not occurred reactions were incubated for a further 20 min. This dehydroalanine-containing protein solution was used immediately in the Michael

addition thiol reaction. For the production of larger quantities of dehydroalanine-containing protein all concentrations and volumes in this procedure were scaled up accordingly, up to a 50 mg scale.

2.4.10 Conversion of dehydroalanine to non-canonical amino acid side chains using sulphur nucleophiles

Addition of sulphur nucleophiles was performed on dehydroalanine-containing protein (2 mg mL^{-1}) in 50 mM sodium phosphate pH 8.0 containing 6 M urea. All thiol compounds were prepared in the fume hood. For the production of larger quantities of ncAA-containing protein all concentrations and volumes in this procedure were scaled up accordingly.

2.4.10.1 Aminoethanthiol

A stock solution was made of $0.1 \text{ mg } \mu\text{L}^{-1}$ of aminoethanethiol in 2 M Tris/HCl pH 8.8 and the pH checked to ensure a pH of 8-9. $40 \mu\text{L}$ of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at $37 \text{ }^\circ\text{C}$ with shaking at 200 rpm. After incubation a $70 \mu\text{L}$ sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.2 Mercaptoacetic acid

A stock solution was made of $0.1 \text{ mg } \mu\text{L}^{-1}$ of mercaptoacetic acid in 2 M Tris/HCl pH 8.8 and the pH checked to ensure a pH of 8-9. $40 \mu\text{L}$ of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at $37 \text{ }^\circ\text{C}$ with shaking at 200 rpm. After incubation a $70 \mu\text{L}$ sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.3 Mercapto-1-butanol

A stock solution was made of 2:1:1 (v/v/v) mercapto-1-butanol:DMF:Tris/HCl (2 M, pH 8.8), and the pH checked to ensure a pH of 8-9. 40 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.4 Mercapto-2-propanol

A stock solution was made of 50:50 (v/v) mercapto-2-propanol:Tris/HCl (2 M, pH 8.8), and the pH checked to ensure a pH of 8-9. 40 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.5 Methylthioglycolate

A stock solution was made of 2:1:1 (v/v/v) methylthioglycolate:DMF:Tris/HCl (2 M, pH 8.8), and the pH checked to ensure a pH of 8-9. 40 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.6 Thioglycerol

A stock solution was made of 50:50 (v/v) thioglycerol:Tris/HCl (2 M, pH 8.8), and the pH checked to ensure a pH of 8-9. 40 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a

70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.7 Sodium mercaptopyruvate

A stock solution was made of 0.1 mg μL^{-1} of sodium mercaptopyruvate in 2 M Tris/HCl pH 8.8 and the pH checked to ensure a pH of 8-9. 80 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.8 Benzylmercaptan

A stock solution was made of 1:2 (v/v) benzylmercaptan:DMF, and the pH checked to ensure a pH of 8-9. 30 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.9 4-Pyridylethylmercaptan

A stock solution was made of 1:2 (v/v) pyridylethylmercaptan:DMF, and the pH checked to ensure a pH of 8-9. 30 μL of this solution was added to 1.25 mL of 2 mg mL^{-1} dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.10 2-Furanmethanethiol

A stock solution was made of 2:1:1 (v/v/v) furanmethanethiol:DMF:Tris/HCl (2 M, pH 8.8), and the pH checked to ensure a pH of 8-9. 40 μL of this solution was added to 1.25 mL of 2 mg mL⁻¹ dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.11 L-Cysteine

A stock solution was made of 0.13 mg μL^{-1} of L-cysteine in 2 M Tris/HCl pH 8.8 and the pH checked to ensure a pH of 8-9. 120 μL of this solution was added to 1.25 mL of 2 mg mL⁻¹ dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.12 D-Cysteine

A stock solution was made of 0.13 mg μL^{-1} of D-cysteine in 2 M Tris/HCl pH 8.8 and the pH checked to ensure a pH of 8-9. 120 μL of this solution was added to 1.25 mL of 2 mg mL⁻¹ dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a 70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.10.13 Propene-1-thiol

A stock solution was made of 2:1:1 (v/v/v) propene-1-thiol:DMF:Tris/HCl (2 M, pH 8.8), and the pH checked to ensure a pH of 8-9. 40 μL of this solution was added to 1.25 mL of 2 mg mL⁻¹ dehydroalanine-containing protein in 50 mM sodium phosphate pH 8.0 containing 6 M urea. The solution was incubated for 2 hours at 37 °C with shaking at 200 rpm. After incubation a

70 μL sample was taken and desalted into 20 mM ammonium acetate pH 7.0 (Section 2.4.7) for analysis by positive mode ESI mass spectrometry. Mass spectrometry was used to assess whether full conversion of the dehydroalanine to the sulphur containing side chain had occurred.

2.4.11 Refolding of non-canonical amino acid containing enzymes

Chemical modification of the proteins is carried out under denaturing conditions in buffer containing 6 M urea. After chemical modification the proteins were subsequently refolded by dialysis. Dialysis was first performed in buffer containing 6 M urea to allow removal of excess modifying reagents that would otherwise cause aggregation of the protein upon refolding. Dialysis (Section 2.4.3) was performed twice against 50 mM sodium phosphate pH 8.0 containing 6 M urea for between 2 and 4 hours at room temperature. Then dialysis against an appropriate urea free buffer was performed twice for between 2 and 4 hours at room temperature.

2.4.12 Size exclusion chromatography

Size exclusion chromatography was performed for further purify proteins and carried out using an ÄKTA Prime purification system (GE Healthcare Life Sciences) with a Superdex S200 column. Size exclusion chromatography was carried out in an appropriate buffer, usually 50 mM Tris/HCl pH 7.4, depending on the downstream application for the protein. Buffers were sterile filtered using Nalgene Sterile Mach V Supor Filter Unit, 0.2 μm pore size, 500 ml (Thermo Fisher Scientific, Loughborough, UK) and degassed, using a vacuum tap, before use in size exclusion chromatography. Protein was injected onto the column at a concentration of between 5-10 mg mL^{-1} and a flow rate of 2 mL min^{-1} .

2.4.13 Lactate dehydrogenase coupled enzyme kinetic assay

Kinetic parameters for the aldol cleavage of reaction were determined using a lactate dehydrogenase (LDH) coupled enzyme assay. This assay couples the production of pyruvate from the aldol cleavage, to the LDH catalysed conversion of pyruvate to lactate which utilises NADH which is oxidised to NAD^+ .

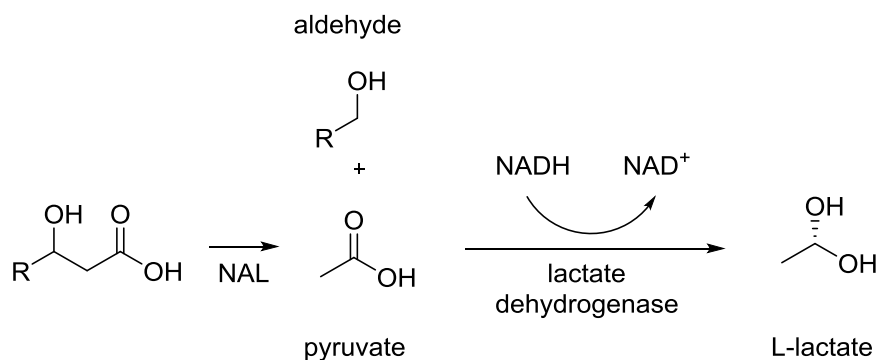


Figure 2.1 Schematic of lactate dehydrogenase (LDH) coupled enzyme assay. The aldol substrate is cleaved to form an aldehyde and pyruvate, the pyruvate is then converted to L-lactate by LDH which oxidises NADH to NAD⁺ and causes a drop in absorbance at 340 nm.

Reactions were performed at 30 °C unless otherwise stated. Reactions contained 0.5 units of LDH, 0.2 mM NADH, 2-300 μL of the substrate (100 mM stock in 1 M Tris/HCl pH 7.4), a volume of 50 mM Tris/HCl pH 7.4 to make a total reaction volume of 1 mL and reactions were initiated by addition of NAL or variant NAL (in 50 mM Tris/HCl pH 7.4). For reactions with wild-type *saNAL* 50 μL of 0.2 mg mL⁻¹ enzyme was added to initiate reactions, for reactions with K165C *saNAL* 200 μL of 2 mg mL⁻¹ enzyme was added to initiate reactions and for reactions with K165- γ -thialysine *saNAL* 50 μL of 2 mg mL⁻¹ enzyme was added to initiate reactions. Reactions were assembled and performed in a 1 mL quartz cuvette. Once the desired enzyme had been added to the reaction it was inverted to mix and the absorbance at 340 nm was measured every 6 s for 1-5 min. The absorbance change of the reaction was measured in a UVIKON 930 spectrophotometer (Kontron Instruments, UK). Specific activity in nmol of substrate cleaved per min per nmol of enzyme was calculated using the molar extinction coefficient of NADH (6620 M⁻¹ cm⁻¹) and the Beer-Lambert law (Equation 1). Kinetic parameters could then be determined by non-linear regression analysis.

2.4.14 pH profiles

The pH profiles were determined at 30 °C using a 3 component buffering system. The three component buffering system allows a pH range between 4 and 9 to be covered whilst maintaining a constant ionic strength (Ellis and Morrison, 1982). A stock of the three component buffer was made containing 0.75 M MES, 1.5 M Tris, and 0.75 M acetic acid. The substrate was made up in 1 M 3 component buffer and the pH adjusted to the desired pH using HCl or NaOH. Reactions were performed in a 1 mL quartz cuvette. The reaction cuvette contained 0.5 units LDH, 0.2 mM NADH, varying volumes of substrate between 2 μL and 300 μL

(stock concentration 50 mM) and 50 mM of the correct pH 3 component buffer system. 50 μ L of the desired enzyme (made up in 50 mM Tris/HCl pH 7.4) was added last followed by inversion and placed in the spectrophotometer. The absorbance change at 340 nm was measured over 1 min in a UVIKON 930 spectrophotometer (Kontron Instruments, UK). The pH was checked, using pH paper, before the enzyme was added and after the reaction had been measured over 1 min to ensure no change in pH had occurred. Specific activity in nmol of substrate cleaved per min per nmol of enzyme was calculated using the molar extinction coefficient of NADH ($6620 \text{ M}^{-1} \text{ cm}^{-1}$) and the Beer-Lambert law (Equation 1). The kinetic parameters for each protein at each pH were determined using non-linear regression analysis. Kinetic parameters, k_{cat}/K_m , were plotted against reaction pH and fitted using Equation 2.

$$\frac{k_{cat}}{K_m} = \frac{\left(k_{cat}/K_m\right)_{max}}{1 + 10^{(pK_a1-pH)} + 10^{(pH-pK_a2)}}$$

Equation 2 Equation used to fit the data shown in panels A and B, where k_{cat}/K_m is the specificity constant for *N*-acetylneuraminic acid cleavage, $(k_{cat}/K_m)_{max}$ is the theoretical maximum value for k_{cat}/K_m and pK_a1 and pK_a2 are the first and second ionisable groups

2.4.15 Far UV circular dichroism spectroscopy

CD spectra were obtained at room temperature using a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK). All spectra were recorded in a quartz cuvette with path length 1.0 mm. Protein samples were prepared in Tris/HCl (50 mM, pH 7.4) to a concentration of 6 μ M. Modified protein samples were prepared by chemical modification of the protein as in Sections 2.4.9 and 2.4.10 followed by refolding via dialysis. Modified protein samples were also analysed after size exclusion chromatography had been performed. Samples of non-modified proteins were prepared by dialysis into Tris/HCl (50 mM, pH 7.4). Samples of refolded non-modified proteins were first dialysed into Tris/HCl (50 mM, pH 7.4) containing 6 M urea and then dialysis into Tris/HCl (50 mM, pH 7.4).

2.4.16 Thiobarbituric acid assay (TBA assay)

To each sample, 11 μL of sodium periodate (0.2 M in 9 M phosphoric acid) was added, the samples were then vortexed for 30 s and centrifuged at 1000 g for 2 min at room temperature. Samples were incubated at room temperature for 20 min. After the incubation time 45 μL of sodium arsenite (10% w/v in 0.5 M Na_2SO_4 containing 0.05 M H_2SO_4) was added and samples were agitated to cause dissipation of any brown discolouration that had formed. Once all samples had turned clear 135 μL of thiobarbituric acid (TBA, 0.6% in 0.5 M Na_2SO_4) was added to each and then incubated at 70 °C for 30 min. After incubation samples were centrifuged at 4000 g for 5 min to separate any precipitate, 85 μL of each sample was then transferred to a shallow, flat bottomed 96 well plate and the absorbance of each well read at 550 nm in a FLUOstar Galaxy plate reader (BMG labtech, Germany).

2.4.16.1 Screening of modified proteins using TBA assay

Aldehydes used to screen for activity were ManNAc, GlcNAc, mannose, glucose, galactose, lyxose, xylose, ribose, arabinose, erythrose. All aldehydes and pyruvate were made up in 50 mM sodium phosphate pH 7.4. Reactions were set up containing 100 μL of each aldehyde (stock concentration 20 mM, final concentration 8 mM) and 100 μL of sodium pyruvate (stock concentration 200 mM, final concentration 80 mM). Reactions were set up in a, 96 deep well plate and incubated for 16 hours at room temperature with 50 μL of modified enzyme (stock concentration 1 mg mL^{-1} , final concentration 0.2 mg mL^{-1}). Each screen also contained reactions of each aldehyde and pyruvate incubated with wild-type *sa*NAL as a comparison and blanks containing each aldehyde and pyruvate and enzyme was replaced with an equal volume of buffer (50 mM sodium phosphate pH 7.4).

After the 16 hour incubation, the TBA assay was performed as in Section 2.4.16. Absorbance data was then analysed to determine whether activity had been increased or decreased over that of the wild-type with each aldehyde.

2.4.16.2 Standard curve of sialic acid by TBA assay

To obtain a standard curve of sialic acid twelve 100 μL samples were prepared containing between 0 and 0.08 μmoles of sialic acid. The TBA assay was performed as in Section 2.4.16 on each sample and then 85 μL of each sample was then transferred to a shallow, flat bottomed 96 well plate and the absorbance of each well read at 550 nm in a FLUOstar Galaxy plate reader (BMG labtech, Germany). This data was then plotted and the gradient of the line could be used to calculate the quantity of aldol product present from an unknown reaction.

2.4.16.3 Kinetic assays using the TBA assay

Kinetic assays were performed in 96 well deep well plates. Reactions were set up with varying concentrations of aldehyde (0.4 mM -15 mM final concentration) and a fixed pyruvate concentration (80 mM final concentration) made up in 50 mM sodium phosphate pH 7.4. 50 μ L of aldehyde and 50 μ L of pyruvate were added to each well and the reaction initiated by the addition of 25 μ L of enzyme. For each aldehyde concentration a blank was also carried out which contained the aldehyde and pyruvate but the enzyme was replaced with an equal volume of buffer (50 mM sodium phosphate pH 7.4). For wild-type kinetic assays a stock concentration of 0.7 mg mL⁻¹ was used in all assays giving a final concentration of 0.14 mg mL⁻¹. For F190Dpc kinetic assays with ManNAc as the aldehyde substrate a stock concentration of 0.7 mg mL⁻¹ was to give a final concentration of 0.14 mg mL⁻¹ and for F190Dpc kinetic assays with erythrose as the aldehyde substrate a stock concentration of 0.3 mg mL⁻¹ was used to give a final concentration of 0.06 mg mL⁻¹. After the enzyme had been added to each well the plate was vortexed for 30 s and then centrifuged at room temperature for 1 min at 1000 g. The reactions were then allowed to incubate for 2 hours at room temperature.

Once the incubation time was complete 100 μ L of each reaction was transferred to a new 96 deep well plate and 10 μ L of TCA (12 % w/v) was added to stop the reactions. The TBA assay was then performed as in Section 2.4.16. To convert the absorbance data into the amount of product formed a standard curve of sialic acid was used. The data was then analysed by non-linear regression.

2.4.17 Purification of DOH by anion exchange

Erythrose (500mg, 4.2 mmol) and sodium pyruvate (2.29g, 21 mmol) were dissolved in sodium phosphate buffer (50 mM pH 7.4 10mL) and F190TG (250 μ L, 0.8mg) was added. This reaction was incubated at room temperature for a minimum of 48 hours before being purified by anion exchange chromatography. Anion exchange chromatography was carried out on AG1x8 resin (HCO₃⁻, 100-200 mesh) and product was eluted using a 0-0.4M ammonium bicarbonate linear gradient (Auge and Gautheron-Le Narvor, 1997). Fractions containing product were identified using the TBA assay on a 20 μ L sample and then analysed by NMR. Four different forms of the product were identified in the proton NMR spectrum and 2D NMR methods, COSY and TOCSY, were used to assign the different forms.

Form 1 was assigned as δ H (500 MHz; D₂O) 3.99-3.90 (1H, m, 4-H), 3.78-3.74 (1H, m, 6-H), 3.66-3.62 (2H, m, 7-H), 3.45 (1H, appt, J 9.5, 5-H), 2.19 (1H, dd, J 13.1, 5.1, 3-Heq), 1.8 (1H, dd,

J 12.9, 11.6, 3-Hax. Form 2 was assigned as δ H (500 MHz; D₂O) 4.2 (1H, m, 4-H), 4.05 (1H, m, 6-H), 3.70 (1H, dd, J 10.3, 3.3, 5-H), 2.16 (1H, dd, J 14.8, 3.6, 3-H^A), 2.05 (1H, dd, J 14.7, 3.1, 3-H^B). Form 3 was assigned as δ H (500 MHz; D₂O) 4.53 (1H, m, 4-H), 2.38 (1H, dd, J 13.7, 7.0, 3-H^A), 2.32 (1H, dd, J 13.5, 5.3, 3-H^B). Form 4 was assigned as δ H (500 MHz; D₂O) 4.53 (1H, m, 4-H), 4.15 (1H, m, 5-H), 3.83 (1H, m, 6-H), 2.55 (1H, dd, J 14.3, 7.2, 3-H^A), 2.08 (1H, dd, J 14.4, 2.8, 3-H^B).

2.4.18 Proton NMR

Proton NMR spectra were recorded on a Bruker Advance 500 or a DRX500 with an internal deuterium lock. COSY and TOCSY pulse sequences were also used to aid the assignment of spectra

2.4.19 ESI Mass spectrometry of small molecules

Accurate mass spectrometry using electrospray ionisation was carried out on an automated system for either a Micromass LCT-KA111 (Waters, Manchester, UK) or Bruker MicroTOF (Bruker Massachusetts, USA) mass spectrometer.

2.5 Crystallographic methods

2.5.1 Protein crystallisation

Crystallisation conditions for *S. aureus* NAL were optimised from those previously found for the *E. coli* NAL (Campeotto *et al.*, 2010). Conditions used were; 100 mM Tris/HCl (pH 7.0-8.5), 200 mM NaCl, 18-28% (w/v) polyethylene glycol (PEG) 3350. These crystallisation conditions were also used for all the modified proteins as well. Crystals were grown by hanging drop vapour diffusion and yielded crystals in 7-10 days.

2.5.2 Cryoprotection and production of enzyme-pyruvate complexes

Cryoprotection of crystals was carried out by soaking in mother liquor with 15% PEG 400 for 1 min, before being sequentially transferred for 1 min into mother liquor with 5% increments of PEG 400. The final soak contains 25% PEG 400 then the crystals were flash cooled in liquid nitrogen. To produce the enzyme-pyruvate complexes cryoprotection was carried out as stated above but the mother liquor was also supplemented with 100 mM sodium pyruvate.

2.5.3 Data collection

Data collection was carried out at Diamond Light Source macromolecular crystallography beam lines I24, I04-1, I03 and I02. Datasets were collected from single crystals for all structures at 100 K.

2.5.4 Data processing

Integration and scaling of data was carried out using MOSFLM (Leslie, 2006) and SCALA (Evans, 2006). The wild-type NAL structure had been previously solved by molecular replacement in PHASER (McCoy, 2007, McCoy *et al.*, 2007) using the *H. influenza* structure (PDB code IF74) (Barbosa *et al.*, 2000). Structures that were found to be in the same space group as the wild-type *sa*NAL structure were solved using direct Fourier methods and structures found to be in different space groups than the wild-type *sa*NAL were solved using molecular replacement in either PHASER or MOLREP using the wild-type *sa*NAL structure (PDB code 4ahp)(Timms *et al.*, 2013).

2.5.5 Refinement

Refinement was carried out in REFMAC 5 (Murshudov *et al.*, 2011, Murshudov *et al.*, 1997) with subsequent rounds of model building carried out in COOT (Emsley *et al.*, 2010) after each cycle of refinement. For all structures initially 10 rounds of rigid body refinement was carried out followed by multiple rounds of restrained refinement and model building. If it was appropriate TLS regions were applied to structures in REFMAC 5 and TLS refinement was carried out during the cycles of refinement. To generate the co-ordinate and restraint files for model building of ligands and non-canonical amino acids, PRODRG (Schuttelkopf and Van Aalten, 2004) was used for earlier structures and for later structures the in-built ligand builder in COOT was used. Structures were validated using Molprobit (Chen *et al.*, 2010) and the PDB validation server.

Chapter 3 Production and Characterisation of K165- γ -thialysine

saNAL

The aim of this work was to utilise non-canonical amino acids (ncAAs) to alter enzyme substrate specificities. There have been many studies into introducing ncAAs into proteins using both global and site-specific incorporation methods (Link *et al.*, 2003, Davis and Chin, 2012, Dumas *et al.*, 2015). Due to the nature of this work it was necessary to use a site-specific method, so that the effect of introducing a single ncAA could be analysed. The two main methods of site-specific incorporation of ncAAs are; firstly the genetic method of using an orthogonal tRNA/tRNA synthetase pair to insert an ncAA at an amber stop codon (Liu and Schultz, 2010, Young and Schultz, 2010), and secondly a variety of chemical methods (Diaz-Rodriguez and Davis, 2011, Desantis and Jones, 1999). The further aim of this work was to investigate the effects of individually incorporating multiple different ncAAs, throughout an enzyme, on the enzyme activity. It was therefore necessary to have a method that could incorporate many different ncAAs, quickly and efficiently. Due to the length of time it would take to evolve each tRNA/tRNA synthase pair for every ncAA to be inserted into the protein, the chemical modification approach was taken. The method chosen was a two-step method, that involved targeting a cysteine residue and first converting to dehydroalanine (dha) via a bis-alkylation-elimination method using the compound 2, 5-dibromohexane-1,6-diamide (diBr) (Chalker *et al.*, 2011) . The dha is then converted into the ncAA side chain by a subsequent Michael addition of a thiol (Figure 3.1). Therefore the first step was to optimise this method for use on a desired protein.

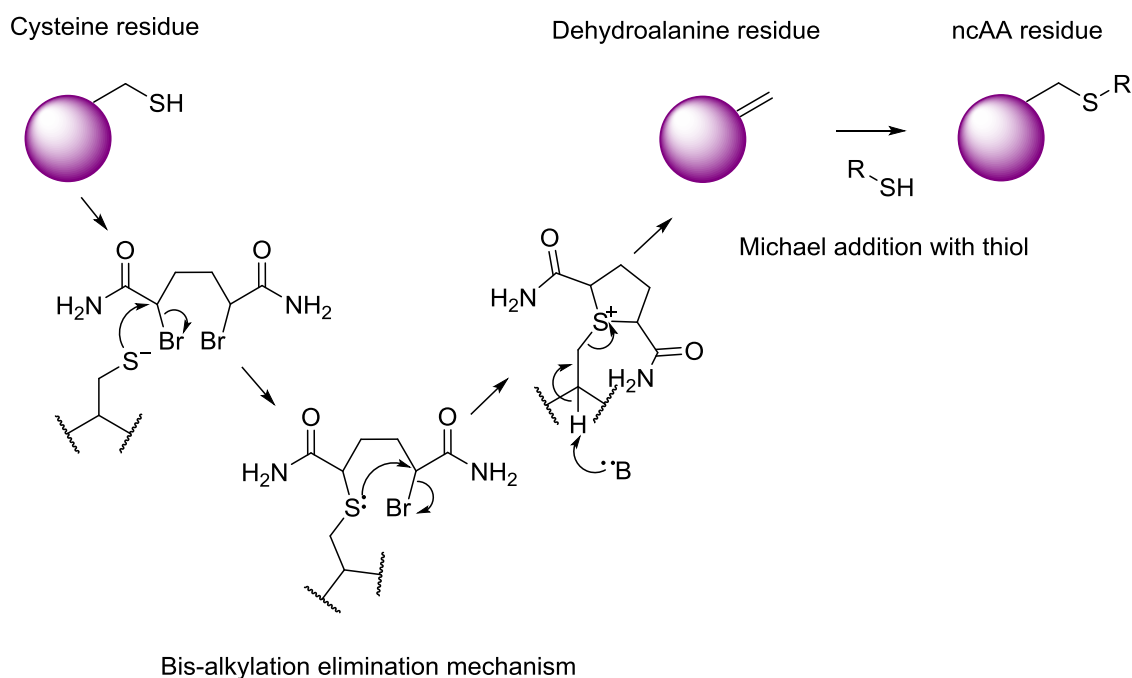


Figure 3.1. Schematic of the chemical modification method used to convert a cysteine residue into a non-canonical amino acid (ncAA). The first stage of the modification is to convert the cysteine residue into dehydroalanine using the compound 2, 5-dibromohexane-1,6-diamide (diBr). Subsequent incubation with a thiol compound should yield the side chain of the ncAA.

Previous work, carried out by Dr Nicole Timms, had identified the *S. aureus* variant of *N*-acetylneuraminic acid lyase (*sa*NAL) as an ideal candidate for incorporation of ncAAs by chemical modification (Nicole Timms, PhD thesis, University of Leeds)(Timms *et al.*, 2013). Due to *sa*NAL being a naturally cysteine-free variant of *N*-acetylneuraminic acid lyase it allows for site-specific incorporation of ncAAs. Using site-directed mutagenesis, a cysteine residue can be inserted at the desired position within *sa*NAL, and subsequent modification of the cysteine residue will incorporate the ncAA site-specifically. Surface-exposed cysteines, on folded proteins, have been shown to readily convert to ncAAs using this method (Chalker *et al.*, 2012, Rowan *et al.*, 2013). However, previous work had identified that residues located within the active site of *sa*NAL and were not accessible to the modification reagents (Nicole Timms, PhD thesis, University of Leeds). Therefore it was necessary to carry out the modification process under denaturing conditions. *sa*NAL was shown to be a highly robust protein that could undergo unfolding in 6 M urea with subsequent refolding without significant loss in activity, this allowed modifications to be carried out whilst the protein was in a denatured state (Timms *et al.*, 2013). The combination of being able to site-specifically introduce ncAAs into *sa*NAL and

being able to access any residue for modification, by denaturing the protein, meant that *s*NAL was an excellent system to evaluate the effects of ncAA incorporation. This chapter will describe the use of this chemical method to replace the catalytic lysine, in *s*NAL, with the ncAA γ -thialysine, and the kinetic and structural characterisation of this modified protein.

3.1 Incorporation of γ -thialysine into *s*NAL

The first residue targeted for conversion to an ncAA was the catalytic lysine at position 165 and this was to be converted into a thioether analogue of lysine; the ncAA γ -thialysine (Figure 3.2). This modification was selected as a 'proof of principle', since it had previously been shown that aldolase activity in fructose-1,6-bisphosphate aldolase (FBPA) could be knocked out by replacing the active site lysine with cysteine, and that subsequent direct alkylation with bromoethylamine, to produce γ -thialysine, restored activity (Hopkins *et al.*, 2002)(Figure 3.2). Direct alkylation with halides has been used extensively to modify cysteine residues and introduce ncAAs into proteins (Hegazy *et al.*, 2006, Gloss and Kirsch, 1995a, Hopkins *et al.*, 2002, Bochar *et al.*, 1999); however not only cysteine residues may be modified using this method. Nucleophilic side chains of lysine and histidine residues may also undergo alkylation (Heinrikson, 1966) and careful control of pH is needed to ensure preferential alkylation at cysteine residues (Hopkins *et al.*, 2005, Chalker *et al.*, 2009). Therefore the bis-alkylation-elimination method to produce a dehydroalanine is a more selective modification method as it produces an electrophilic centre that is not present in the twenty proteogenic amino acids, and can selectively targeted with a nucleophile.

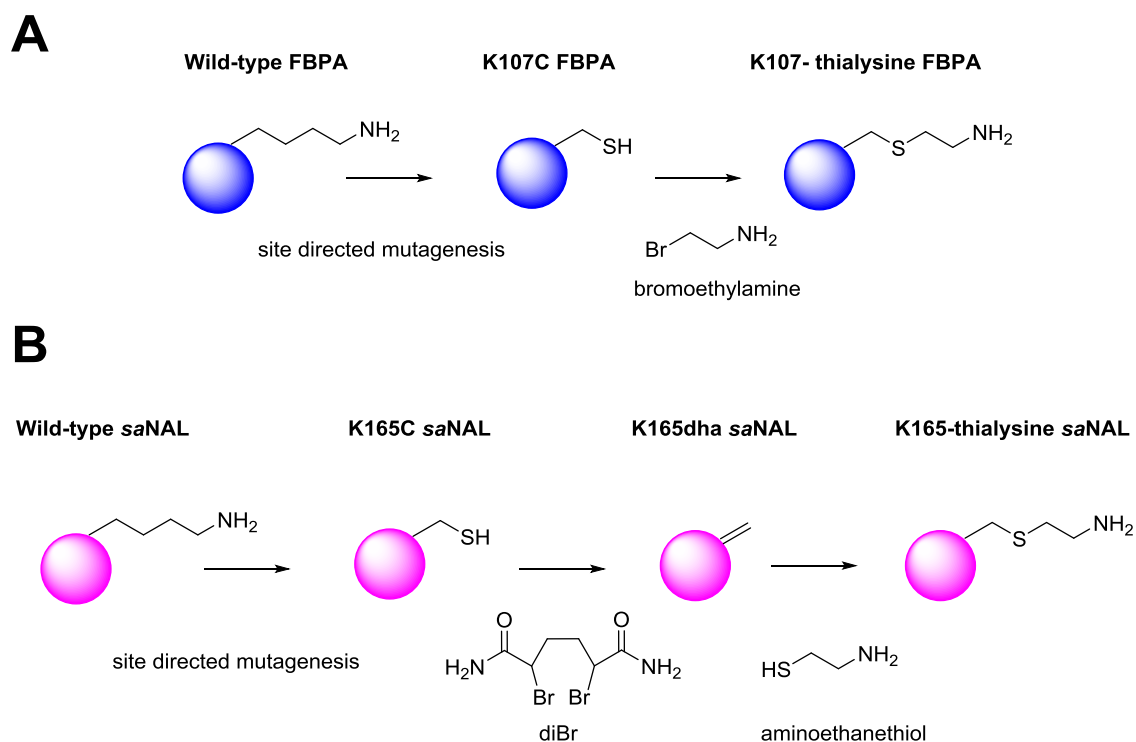


Figure 3.2. Comparison of the direct alkylation method and the bis-alkylation-elimination followed by a Michael addition method, to incorporation a γ -thialysine residue into a protein. A) Schematic of the conversion of the catalytic lysine at position 107 in fructose 1,6-bisphosphate aldolase (FBPA) into the non-canonical amino acid (ncAA) γ -thialysine. The first stage of the conversion is to produce a cysteine mutant at position 165 using site-directed mutagenesis and then direct alkylation with bromoethylamine produces the γ -thialysine. B) Schematic of the conversion of the catalytic lysine at position 165 in saNAL into the ncAA γ -thialysine. The first stage of the conversion is to produce a cysteine mutant at position 165 using site-directed mutagenesis. The cysteine variant can then be converted to dehydroalanine (dha) using the compound 2, 5-dibromohexan(1,6)diamide (diBr). Once the cysteine residue is fully converted to dha, aminoethanethiol can then be used to convert the dha residue into γ -thialysine

3.1.1 Characterisation of wild-type and K165C saNAL

To target K165 for modification, a cysteine mutant at position 165 was needed. The K165C mutant had previously been produced by site-directed mutagenesis, and both K165C and wild-type saNAL DNA were kindly provided by Prof. Alan Berry (University of Leeds, UK). Wild-type and K165C saNAL were expressed and purified using previously described methods (Williams *et al.*, 2005, Timms *et al.*, 2013). SDS PAGE was used to assess protein purity; in the elution fraction a clear band of protein could be seen at approximately 34 kDa indicating successfully purified protein (Figure 3.3). To further assess whether the proteins had been successfully

produced, positive mode ESI mass spectrometry was used. This identified the molecular mass of wild-type *sa*NAL to be 33992.5 Da and the molecular mass of the K165C variant to be 33969 Da, both in agreement with the masses predicted from the amino acid sequences, 33995 Da and 33970 Da respectively (Figure 3.3).

Once proteins had been purified they were either stored at 4°C where they were stable for up to three months, or they were lyophilised and stored at -80°C. Proteins were prepared for lyophilisation by first dialysing into a volatile buffer, usually 50 mM ammonium acetate pH 7.0, and then flash freezing in liquid nitrogen. Lyophilised proteins were stable for up to 9 months.

Kinetic parameters, k_{cat} , K_m and k_{cat}/K_m , were determined for wild-type *sa*NAL and for K165C *sa*NAL (shown in Table 3.1) using the LDH coupled assay described in Section 2.4.13. Kinetic characterisation of the wild-type *sa*NAL and K165C variant revealed that the k_{cat}/K_m for *N*-acetylneuraminic acid (Neu5Ac) cleavage was reduced 1000-fold in the K165C variant and the k_{cat} was also reduced 625-fold. This significant reduction in activity caused by replacement of the catalytic lysine with a cysteine has also been found in previous aldolase studies (Hopkins *et al.*, 2002)

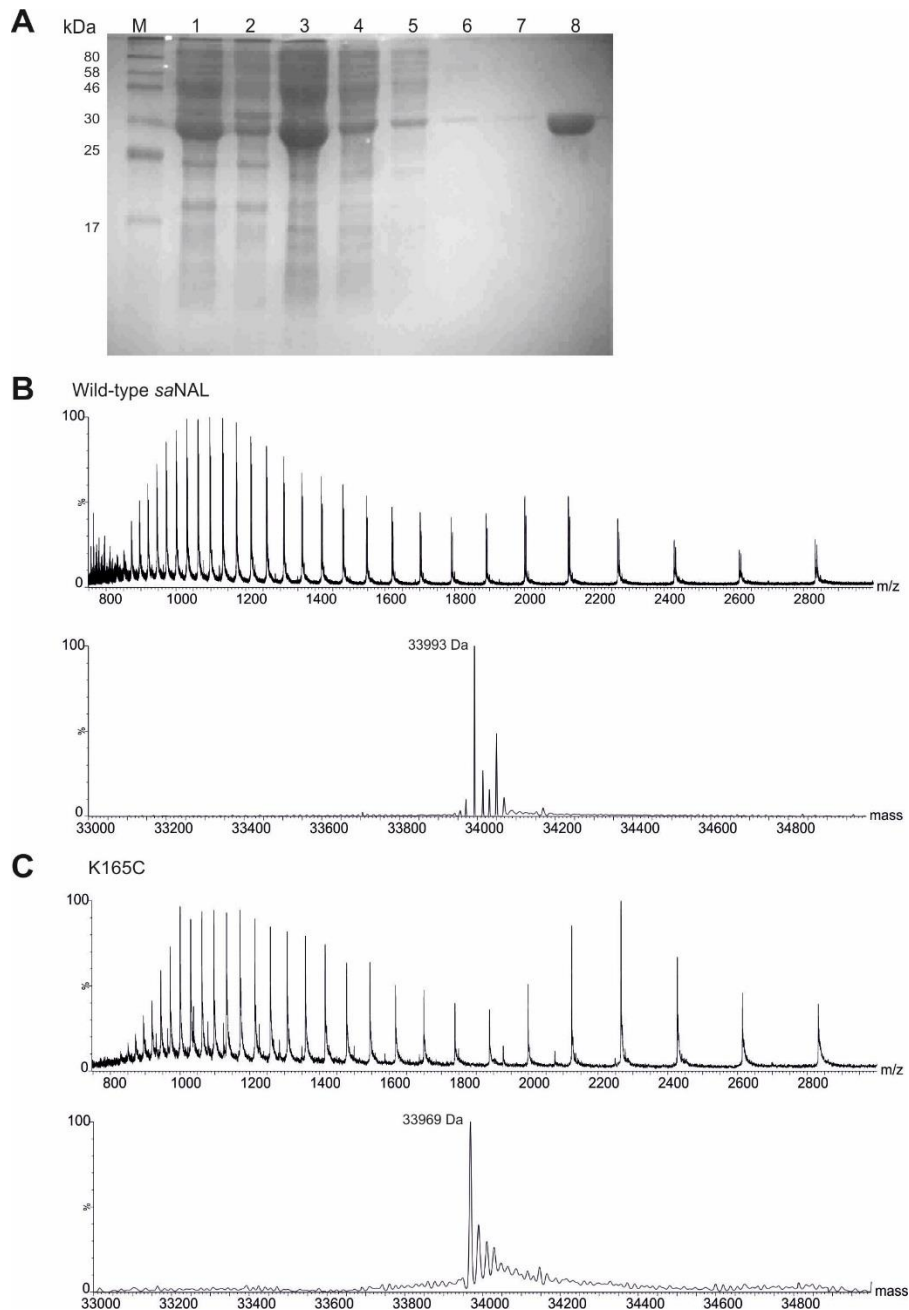


Figure 3.3. Purification and ESI mass spectra of wild-type and K165C *saNAL*. Panel A shows an example SDS-PAGE gel of the batch nickel affinity purification of wild-type *saNAL*. A broad range (7-175 kDa) molecular weight marker is shown in lane M. Whole cell lysate can be seen in lane 1 with a clear band of overexpressed protein at approximately 34 kDa. Insoluble and soluble fractions can be seen in lane 2 and 3 respectively, in the soluble fraction the overexpressed protein band is clearly visible. Samples from the load and 3 wash steps carried out are shown in lanes 4 to 7 and purified eluted protein is shown in lane 8. Panels B and C show ESI mass spectra of wild-type and K165C *saNAL*, raw data is shown above in each panel and deconvoluted spectra are shown below. The mass spectra show the molecular mass of the wild-type and K165C proteins to be 33993 Da and 33969 Da respectively which are in agreement with their predicted molecular masses of 33995 Da and 33970 Da, respectively.

Enzyme	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($\text{min}^{-1}.\text{mM}^{-1}$)
Wild-type	250 ± 4.6	2.2 ± 0.12	114
K165C	0.4 ± 0.014	3.8 ± 0.45	0.11

Table 3.1. Steady-state kinetic parameters for the cleavage of *N*-acetylneuraminic acid (Neu5Ac) by both the wild-type and K165C *s*NAL. Kinetic parameters were determined using the LDH coupled assay (Section 2.4.13).

3.1.2 Conversion of K165C *s*NAL into K165dha *s*NAL

The first step of the modification to introduce γ -thialysine at position 165 was to convert the cysteine variant, K165C, into K165dha. This conversion was carried out using the compound diBr, the diBr used in all experiments was synthesised by Robert Smith and Alun Myden (University of Leeds, UK). DiBr causes a bis-alkylation-elimination mechanism, with the cysteine residue, producing a dehydroalanine residue (Figure 3.1). Small scale modifications (2.5 mg) were initially carried out: 2.5 mg of lyophilised K165C protein was taken and resuspended in 1.25 mL of pre-warmed (37 °C) sodium phosphate buffer (50 mM, pH 8.0) containing 6 M urea. To this protein solution, 100 μL of diBr (0.13 mg μL^{-1} in DMF) was added (581 equivalents of diBr relative to protein), and then incubated for 1 hr 30 min at 37 °C with shaking at 200 rpm. After the incubation period, a 70 μL sample was taken and buffer exchanged into 50 mM ammonium acetate pH 7.0, using a Micro Bio-Spin™ Chromatography Column, for analysis by ESI mass spectrometry, which was performed by Dr James Ault (University of Leeds). Using ESI mass spectrometry it was possible to assess whether conversion from K165C into K165dha had occurred. The expected mass for K165dha was 33934 Da and the observed mass was 33936.6 Da (Figure 3.4) indicating conversion had occurred.

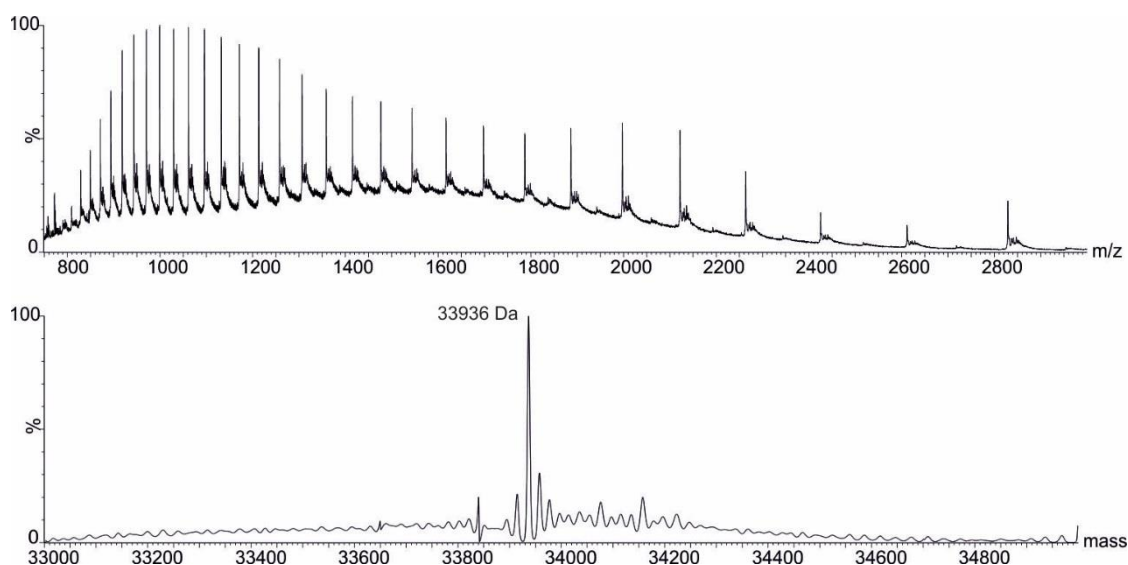


Figure 3.4. ESI mass spectra of K165dha $s\alpha$ NAL. The raw data is shown above and the deconvoluted spectrum below. Mass spectrum of K165dha was obtained after 1 hr 30 min incubation (37 °C, 200 rpm) of K165C $s\alpha$ NAL (1.25 mL, 2 mg mL⁻¹) with diBr (100 μ L, 0.13 mg μ L⁻¹). ESI-MS showed the molecular mass of K165dha to be 33936.6 Da which is in agreement with the expected mass of 33934 Da, the mass spectrum shows no evidence for any remaining K165C protein which has the molecular mass 33970 Da.

3.1.3 Modification of K165dha $s\alpha$ NAL into K165- γ -thialysine $s\alpha$ NAL

Once K165C had been successfully converted to K165dha, the next step of the modification was carried out. Aminoethanethiol was used to produce the ncAA γ -thialysine via a Michael addition onto the dehydroalanine residue at position 165. 40 μ L of aminoethanethiol (0.1 mg mL⁻¹ in 2 M Tris/HCl pH 7.4) was added to the 1.25 mL of denatured K165dha (2 mg mL⁻¹ in 50 mM sodium phosphate buffer, pH 8.0, containing 6 M urea) and incubated for 2 hours at 37 °C with shaking at 200rpm. To monitor the conversion, after 2 hours a 70 μ L sample was taken from the reaction mixture for analysis by ESI mass spectrometry, which was performed by Dr James Ault. The sample was first buffer exchanged into 50 mM sodium phosphate buffer, pH 8.0, containing 6 M urea to remove excess aminoethanethiol, as this had previously been shown to cause the protein to stick to the spin column matrix (Nicole Timms, PhD thesis, University of Leeds), and then desalted into 50 mM ammonium acetate pH 7.0. The observed mass of K165- γ -thialysine was 34011 Da - in agreement with the expected mass of 34012 Da.

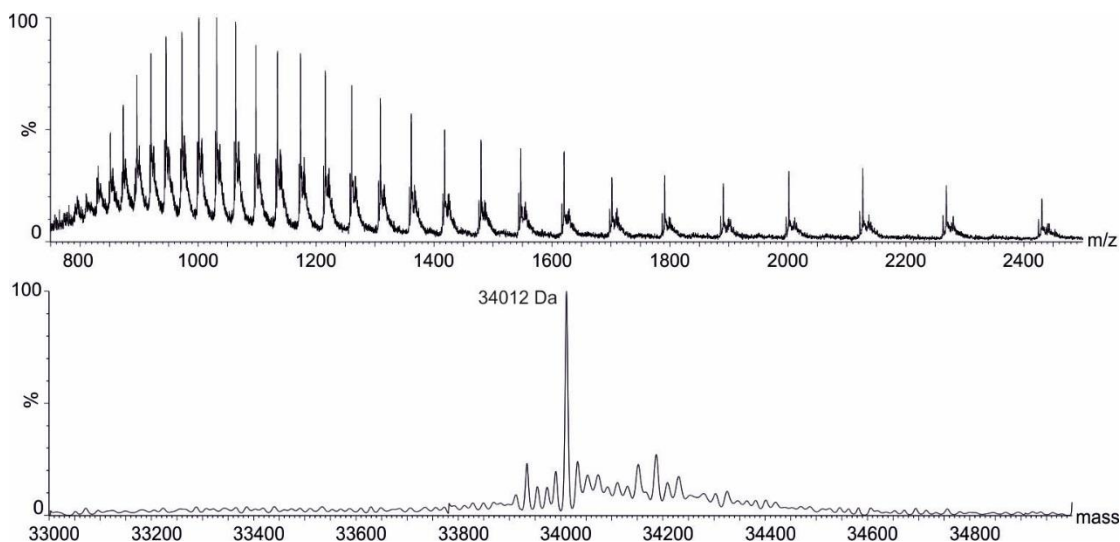


Figure 3.5 ESI mass spectrum of K165- γ -thialysine $s\alpha$ NAL. The raw data is shown above and the deconvoluted spectrum below. Mass spectrum of K165- γ -thialysine $s\alpha$ NAL was obtained after 2 hour incubation (37 °C, 200 rpm) of K165dha $s\alpha$ NAL (1.25 mL, 2 mg mL⁻¹) with aminoethanethiol (40 μ L, 0.1 mg μ L⁻¹). ESI mass spectrometry showed the molecular mass of K165- γ -thialysine to be 34012Da - in agreement with the expected mass of 34012 Da.

After the conversion to K165- γ -thialysine had been achieved, the protein was refolded by removing the urea by dialysis. As before, when preparing the protein for mass spectrometry, it was observed that excess aminoethanethiol caused the protein to precipitate on removal of the urea. Therefore the protein was first dialysed against sodium phosphate buffer (50 mM, pH 8.0) containing 6 M urea and then against the same buffer containing no urea.

3.1.4 Large scale conversion of K165C $s\alpha$ NAL into K165- γ -thialysine $s\alpha$ NAL

For further studies of the modified protein, K165- γ -thialysine, larger quantities of the protein were required. To produce these larger quantities scale up of the modification procedure was carried out; modifications were carried out on a 20 mg scale and up to a 50 mg scale if needed. The modification was carried out in the same manner as with the small scale modification, 20 mg of lyophilised K165C protein was resuspended in 10 mL of sodium phosphate buffer (50 mM pH 8.0) containing 6 M urea which had been pre-warmed to 37 °C. The protein solution was vortexed for 30 s to ensure the lyophilised protein was thoroughly redissolved, then 800 μ L of diBr (0.13 mg μ L⁻¹ in DMF; 581 equivalents), was added and the solution incubated for 1 hr 30 min at 37 °C, with shaking 200 rpm. Once conversion to K165dha was confirmed using ESI mass spectrometry, then 320 μ L of aminoethanethiol (0.1 mg μ L⁻¹ in 2 M Tris/HCl pH 8.8) was added and incubated at 37 °C, with shaking at 200 rpm for 2 hours.

3.2 Characterisation of K165- γ -thialysine *sa*NAL

ESI mass spectrometry indicated that full conversion of the variant K165C into K165- γ -thialysine had been successfully achieved. Now that correctly modified protein had been produced, characterisation of the nCAA-containing protein was carried out. Far UV circular dichroism (CD) was used to assess the fold of the modified protein and LDH coupled kinetics were used to obtain kinetic parameters.

3.2.1 Far UV circular dichroism and initial kinetic assays of K165- γ -thialysine

The modification of K165C *sa*NAL into K165- γ -thialysine had been performed under denaturing conditions and then the urea was removed by dialysis. Removal of urea in this manner had been shown to refold wild-type *sa*NAL successfully (Nicole Timms, PhD thesis, University of Leeds)(Timms *et al.*, 2013); to determine whether the modified protein had refolded correctly, far UV CD was used (Section 2.4.15).

Far UV CD was carried out on samples of refolded, modified K165- γ -thialysine enzyme and both wild-type *sa*NAL and wild-type *sa*NAL that had undergone refolding in the same manner as the modified protein. CD spectra indicated that the refolded modified protein did not have an identical CD spectrum to wild-type *sa*NAL (Figure 3.8). Therefore further investigation was carried out to determine whether refolding had been successful for K165- γ -thialysine *sa*NAL and whether this modified protein was active.

If the modified protein had refolded correctly it is likely that it would be more active compared to the K165C variant. Therefore, kinetic assays were carried out with the modified K165- γ -thialysine protein to determine if there was a regain of activity. LDH coupled kinetics were carried out as in Section 2.4.13. Kinetic assays showed that incorporation of the γ -thialysine residue had restored significant activity compared to the K165C variant, showing a 90-fold increase in k_{cat}/K_m (Table 3.2). However, kinetic assays also showed that the activity of the K165- γ -thialysine enzyme was considerably lower compared to that of the wild-type *sa*NAL, showing an 11-fold decrease in k_{cat}/K_m .

This restoration of activity indicated that, although the CD spectrum of the modified protein appeared not to be identical to the wild-type protein, it was a protein containing a non-canonical side chain that regained some wild-type enzyme activity. Therefore further investigation was carried out to assess whether the difference in CD spectra was a consequence of the modification procedure.

Enzyme	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($\text{min}^{-1}.\text{mM}^{-1}$)
Wild-type	250 ± 4.6	2.2 ± 0.12	114
K165C	0.40 ± 0.014	3.8 ± 0.45	0.11
K165- γ -thialysine	11 ± 0.23	1.1 ± 0.11	10

Table 3.2. Steady-state kinetic parameters for the cleavage of *N*-acetylneuraminic acid by both the wild-type, K165C and the nAA containing protein K165- γ -thialysine *sa*NAL. Kinetic parameters were determined using the LDH coupled assay.

During the modification procedure the dehydroalanine intermediate produced from reaction of the cysteine residue with diBr is planar. Therefore the Michael addition of the thiol, to produce the non-canonical side chain, could occur on either face of the double bond. This means that it is possible to generate either a D or L non-canonical amino acid from the dehydroalanine intermediate (Zhu and Van Der Donk, 2001, Chalker *et al.*, 2012). It was therefore hypothesised that any protein that had been modified could have produced a mixed population of both D and L non-canonical side chains, and potentially the protein that had modified to the D form of the non-canonical amino acid may not have refolded correctly (Figure 3.6).

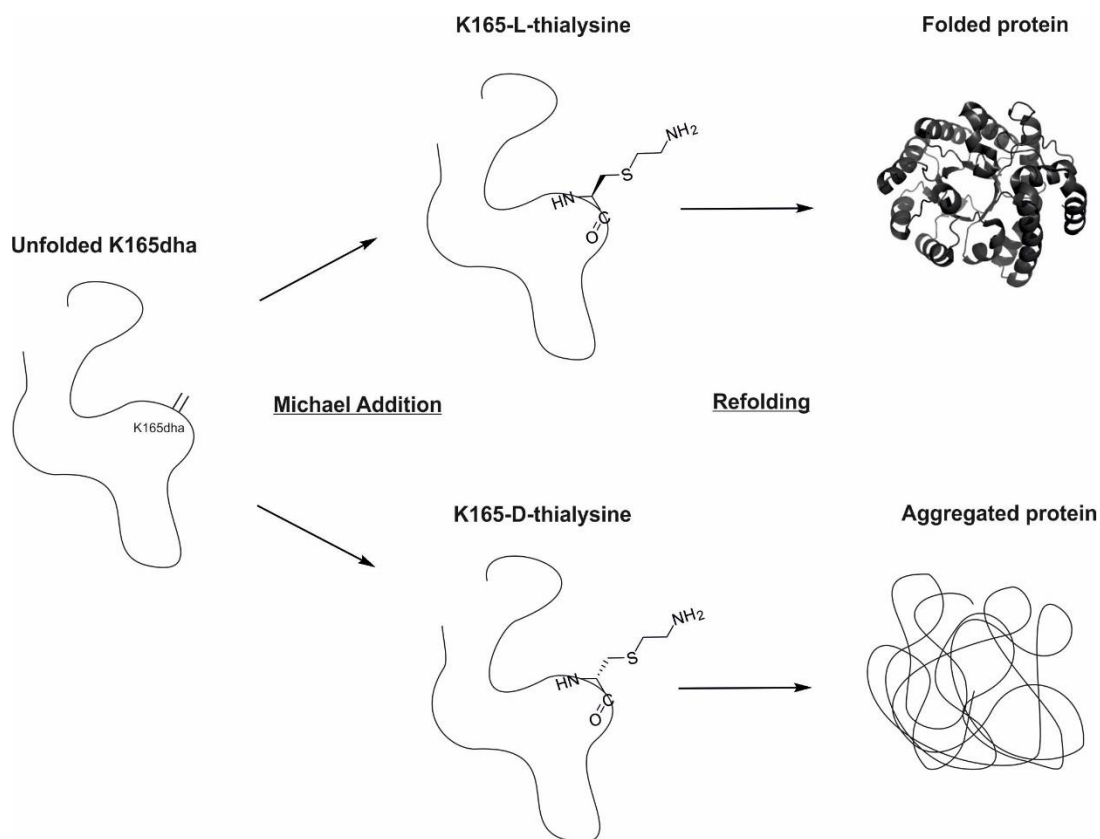


Figure 3.6. Schematic representation of chemical modification producing both D and L forms of the nCAA γ -thialysine. Michael addition of aminoethanethiol is shown occurring on either faces of the dehydroalanine (dha) leading to both D and L forms of γ -thialysine. It is hypothesised that protein containing L- γ -thialysine will refold whereas protein containing D- γ -thialysine will aggregate.

To see whether a mixed species of K165- γ -thialysine had been produced during the modification procedure, size exclusion chromatography was used. Correctly modified K165- γ -thialysine protein underwent size exclusion chromatography using a Superdex S200 preparative grade column (GE Healthcare Life Sciences). During gel filtration of the modified protein, the elution profile was significantly different to that of either wild-type or K165C *sa*NAL. These unmodified proteins eluted in one peak at approximately 180 mL whereas the modified protein was seen to elute in two peaks. In the elution profile of K165- γ -thialysine the first protein eluted in the void volume at approximately 110 mL and the second protein peak eluted at approximately 180 mL (Figure 3.7) - the elution volume expected for a tetramer made up of 34 kDa monomers, as in wild-type *sa*NAL. This indicated that the modification procedure produced a mixed species of nCAA-containing protein. From the ESI-MS analysis all

of the protein is seen to contain γ -thialysine however a proportion of the modified protein must not refold correctly and therefore elutes in the void volume.

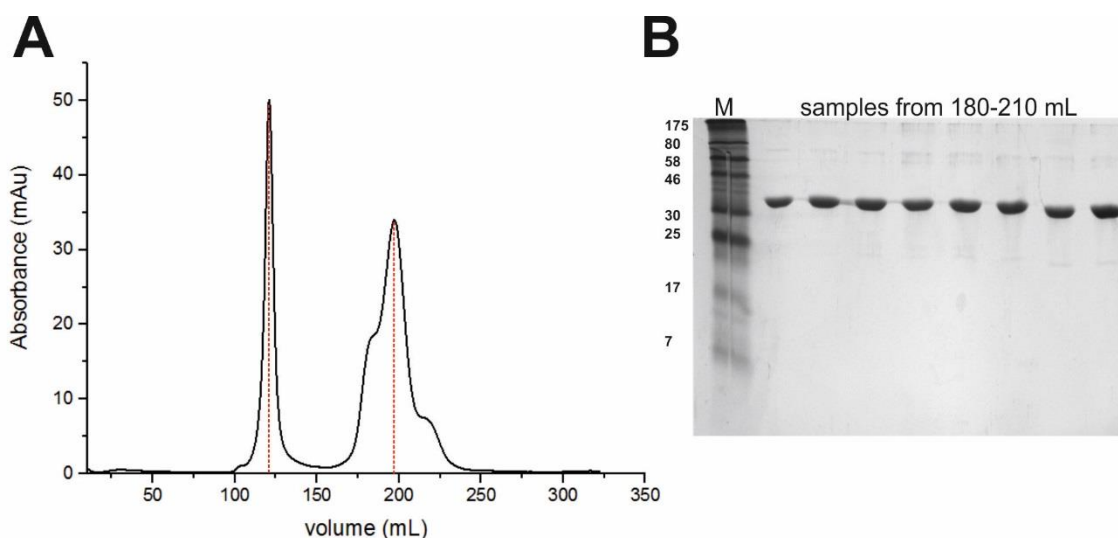


Figure 3.7. Size exclusion chromatography of K165- γ -thialysine. Panel A shows the A280 nm trace of the modified protein with the peak of the void volume at 110-125 mL and the peak for the modified protein at 180-210 mL. Panel B shows the SDS-PAGE gel of samples from the gel-filtration peak at 180-210 mL showing protein bands at approximately 34 kDa, broad range (7-175 kDa) molecular weight markers are shown in lane M.

Far UV CD was then used to analyse samples from both the void volume peak and the peak at 180 mL, which will be referred to as gel-filtered K165- γ -thialysine. The far UV CD spectrum for the void volume peak showed that it did not have the same CD spectrum as the wild-type protein. Whereas the sample for gel-filtered K165- γ -thialysine has a spectrum identical to that of the wild-type protein (Figure 3.8). This data shows that incorrectly folded protein is formed during the modification process, potentially due to production of a D amino acid at the modification position. However by using size exclusion chromatography this material can be removed to leave the correctly modified and correctly folded protein.

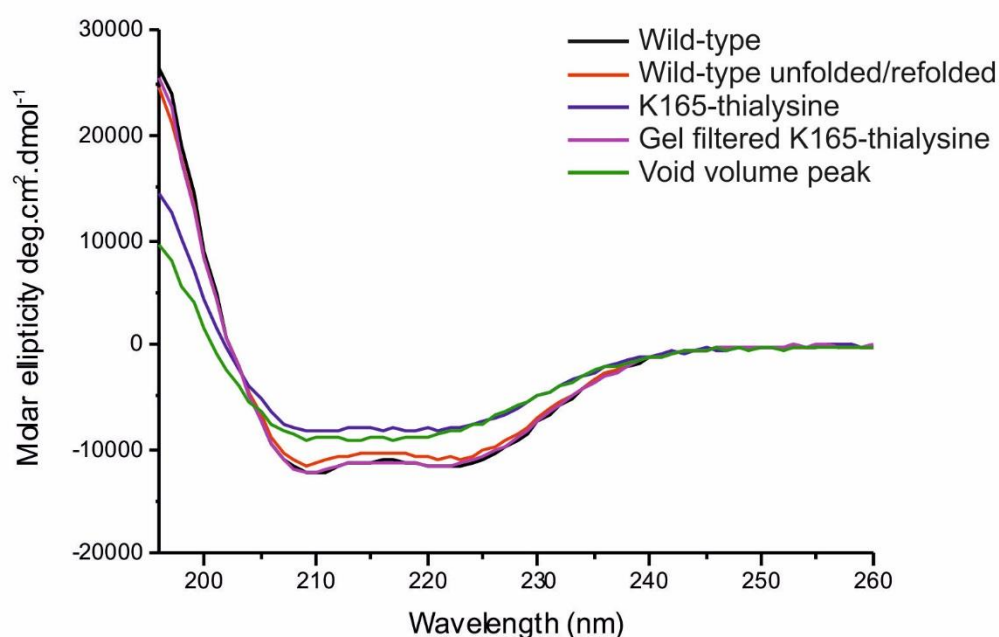


Figure 3.8. Far UV circular dichroism of wild-type and K165- γ -thialysine. CD spectra of wild-type *s* α NAL (black), wild-type *s* α NAL that has undergone unfolding in 6 M urea and subsequent refolding (red), K165- γ -thialysine *s* α NAL produced using the chemical modification procedure and therefore unfolded in 6 M urea and then refolded (blue) and K165- γ -thialysine *s* α NAL that has undergone gel-filtration: both the protein that eluted at the correct elution volume gel-filtered K165- γ -thialysine (purple) and the void volume (green).

3.2.2 Kinetic comparison of wild-type *s* α NAL and K165- γ -thialysine *s* α NAL

Kinetic analysis was carried out with the gel-filtered K165- γ -thialysine protein, using LDH couple kinetics in 50 mM Tris/HCl pH 7.4. This showed an approximate doubling in activity for both k_{cat} and k_{cat}/K_m over that of the modified protein before gel-filtration (Table 3.3). This data indicates that during the modification process although the protein is correctly modified to contain γ -thialysine (as is shown from the mass spectrometry data), some of the protein may not refold correctly. Using size exclusion chromatography this incorrectly folded material can be separated from correctly folded protein, and upon removal of the incorrectly folded material the modified protein regains more activity, indicating the incorrectly folded protein is either inactive or far less active than the correctly folded modified protein. Therefore the gel-filtered K165- γ -thialysine *s* α NAL was used in all further experiments.

Enzyme	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($\text{min}^{-1}.\text{mM}^{-1}$)
Wild-type	250 ± 4.6	2.2 ± 0.12	114
K165C	0.40 ± 0.014	3.8 ± 0.45	0.11
K165- γ -thialysine	11 ± 0.23	1.1 ± 0.11	10
pre gel-filtration			
Gel-filtered K165- γ -thialysine	26 ± 0.9	1.4 ± 0.2	19

Table 3.3. Steady-state kinetic parameters for the cleavage of *N*-acetylneuraminic acid by the wild-type *sa*NAL, K165C *sa*NAL and the ncAA-containing protein K165- γ -thialysine *sa*NAL before and after undergoing size exclusion chromatography. Kinetic parameters were determined using the LDH coupled assay.

This decrease in activity compared to the wild-type enzyme, but increase in activity compared to the cysteine variant K165C, brought about by the insertion of the ncAA γ -thialysine has also been seen in previous γ -thialysine studies, where direct alkylation of a cysteine was used to insert the ncAA (Gloss and Kirsch, 1995a, Hopkins *et al.*, 2002). The decrease in activity in the previous studies was attributed to incomplete conversion of the cysteine to the non-canonical side chain. However in this current study there is no evidence in the ESI-MS that there is any remaining dha or cysteine residue, at position 165. Therefore the decrease in activity may be due to structural differences between the lysine and γ -thialysine side chains.

3.3 Crystallographic studies of K165- γ -thialysine *sa*NAL

To determine the structural effects of inserting the non-canonical amino acid, γ -thialysine into the active site of *sa*NAL, the crystal structures of the K165- γ -thialysine, wild-type and K165C enzymes were solved. To investigate whether the incorporation of a sulphur atom into the lysine chain affected the structural binding of pyruvate in the active site, which could have led to the decrease in activity when compared to the wild-type enzyme, the structures of both modified and wild-type proteins were also solved in complex with pyruvate.

3.3.1 Crystallisation

Crystallisation conditions had been previously determined for the *E. coli* NAL enzyme (Campeotto *et al.*, 2009, Campeotto *et al.*, 2010), and so these conditions were optimised for wild-type *sa*NAL. Optimisation of crystallisation conditions of the wild-type *sa*NAL was carried out by Dr Chi Trinh (University of Leeds, UK) and Anna Polyakova (University of Leeds, UK). Crystallisation conditions used were 200 mM NaCl, 100 mM Tris/HCl (pH 7.0-8.5), and 18-28% (w/v) PEG 3350. These conditions were also used for crystallisation of K165- γ -thialysine and K165C *sa*NAL enzymes. Wild-type and K165C *sa*NAL were prepared for crystallisation by purification using nickel affinity chromatography (Section 2.4.1) and then size exclusion chromatography (Section 2.4.12). K165- γ -thialysine *sa*NAL was prepared for crystallisation by first purifying K165C protein by nickel affinity chromatography. The K165C protein was modified to K165- γ -thialysine which then underwent size exclusion chromatography, and correctly folded K165- γ -thialysine which eluted at approximately 180 mL was used for structural studies. Proteins were concentrated to 7-10 mg mL⁻¹ prior to crystallisation. Crystallisation trays were set up using hanging drop vapour diffusion, with drops containing 1:2, 2:2 and 2:1 of protein to mother liquor. Hanging drop vapour diffusion yielded crystals in 7-10 days and crystals usually formed in either hexagonal or rectangular blocks (Figure 3.9).

To produce cryo-cooled crystals for data collection, the cryoprotectant used was PEG 400. Crystals were soaked in mother liquor supplemented with increasing amounts of PEG 400. The first soak contained 15% PEG 400 (w/v) and increased in 5% increments to 25% PEG 400 (w/v). Crystals were soaked in each concentration of PEG400 for approximately 1 min and then flash cooled in liquid nitrogen ready for data collection.

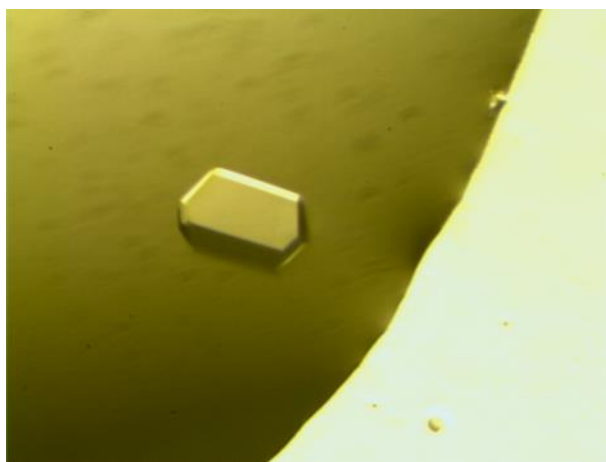


Figure 3.9. K165- γ -thialysine *sa*NAL crystal formed from hanging drop vapour diffusion in 200 mM NaCl, 100 mM Tris/HCl (pH 7.0-8.5), PEG 3350 (18-28% w/v).

The structures of the apo wild-type and K165- γ -thialysine *sa*NAL were solved, and the structures of their pyruvate bound complexes were also solved. To produce the complexes of K165- γ -thialysine and wild-type *sa*NAL with pyruvate, the crystals were soaked with sodium pyruvate. Soaking was carried as with the apo structures with mother liquor supplemented with 15-25% PEG400 (w/v) but soaks also contained 100 mM pyruvate (Campeotto *et al.*, 2009). Attempts were also made to solve the structure of the K165C *sa*NAL in complex with pyruvate; however despite soaking the crystals with 100 mM pyruvate the subsequent structures showed no evidence of pyruvate in the active site.

3.3.2 Data collection and processing

Data collection for all crystals was carried out at Diamond Light Source (Oxfordshire, UK), and data sets were produced from single crystals. Data sets for wild-type and K165- γ -thialysine *sa*NAL structures, both the apo and pyruvate soaked, were collected on beamline I02 and data for the K165C *sa*NAL structure were collected on beamline I04-1. Data processing for the wild-type *sa*NAL structures, both the apo and in complex with pyruvate, was carried out by Dr Chi Trinh and Anna Polyakova. Integration and scaling of all datasets was performed using the CCP4 programs MOSFLM and SCALA (Leslie, 2006, Evans, 2006) and full structural statistics can be seen in Table 3.4. The wild-type *sa*NAL apo structure was determined by molecular replacement methods using the program PHASER (McCoy, 2007, McCoy *et al.*, 2007), using the *H. influenza* NAL structure (pdb code 1F74) (Barbosa *et al.*, 2000) as the search model. Since the other structures have a very similar unit cell parameters to the wild-type *sa*NAL structure (see Table 3.4), and the same space group, $P2_12_12_1$, it was possible to determine these structures using the refined structure of wild-type *sa*NAL apo structure as a starting model for refinement and rebuilding into both $2F_{\text{obs}}-F_{\text{cal}}$ and $F_{\text{obs}}-F_{\text{cal}}$ maps using REFMAC (Murshudov *et al.*, 1997, Murshudov *et al.*, 2011) and COOT (Emsley *et al.*, 2010). All crystal structures had an upper resolution between 1.9 Å and 2.35 Å.

	Wild-type <i>sa</i> NAL	Wild-type <i>sa</i> NAL pyruvate	K165C <i>sa</i> NAL	K165C- γ -thialysine <i>sa</i> NAL	K165C- γ -thialysine <i>sa</i> NAL pyruvate
Diamond beamline	I02	I02	I04-1	I02	I02
PBD accession code	4ahp	4ah7	4ahq	4aho	4ama
Resolution (Å)	2.10	2.30	1.95	2.00	2.35
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
<i>a</i> (Å)	82.5	82.3	82.2	82.3	82.9
<i>b</i> (Å)	109.6	109.9	110.0	109.7	110.3
<i>c</i> (Å)	131.3	131.7	133.7	131.4	132.2
<i>R</i> _{merge} §*	0.075(0.350)	0.117(0.361)	0.067(0.47)	0.064(0.435)	0.11(0.498)
<i>R</i> _{pim} +*	0.044(0.20)	0.065(0.201)	0.03(0.225)	0.036(0.248)	0.065(0.289)
Observed reflections	258,615	212,634	499,678	320,723	193,152
Unique reflections	68,405	53,415	86,766	80,641	50,421
Completeness (%)*	97.7(92.8)	99.4(100)	97.8(93.4)	99.6(100)	98.5(99.0)
Multiplicity*	3.8(3.7)	4.0(4.0)	5.8(5.1)	4.0(4.1)	3.8(3.9)
<I/σ> I/σ *	10.8(3.4)	9.7(3.7)	14.9(3.2)	14.6 (3.5)	11.4(3.2)
Refinement					
<i>R</i> _{factor} (%)	0.19	0.20	0.19	0.19	0.18
<i>R</i> _{free} (%)†	0.24	0.25	0.23	0.23	0.24
No. of protein atoms	9015	9189	9222	8922	9261
No. of solvent molecules	381	130	368	362	356
No of ligand atoms	-	20	-	-	20
Average overall B-factor (Å ²)	25.8	17.9	31.4	33.1	27.7
Average ligand B-factor (Å ²)	-	13.5	-	-	26.6
Average solvent B-factor (Å ²)	26.8	14.3	35.8	31.9	22.8
RMS bond lengths (Å) ξ	0.013	0.013	0.015	0.013	0.013
RMS bond angles (Å) ξ	1.441	1.617	1.78	1.441	1.586
Ramachdran analysis, the percentage of residues in the regions of plot (%) ‡					
Most favoured	98.8	98.1	99.1	98.7	98.3
Outliers	0	0	1'	0	0

$$\S R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}$$

+ *R*_{pim} - precision-indicating (multiplicity-weighted) *R*_{merge}, relative to all I+ or I-.

* Values given in parentheses correspond to those in the outermost shell of the resolution range.

† *R*_{free} was calculated with 5% of the reflections set aside randomly.

ξ Based on the ideal geometry values of Engh & Huber (Engh and Huber, 1991).

‡ Ramachdran analysis using the program MolProbity (Lovell *et al.*, 2003)

'The side chain of Tyr111 is in close proximity to Leu142, Thr143 and Phe110 from an adjacent chain causing the phi and psi angles of Tyr111 to lie in an unfavoured region of the Ramachdran plot

Table 3.4. Structural data statistics for the X-ray crystallographic structures of wild-type and K165- γ -thialysine *sa*NAL both apo and in complex with pyruvate and K165C *sa*NAL.

The CCP4 program REFMAC (Murshudov *et al.*, 2011, Murshudov *et al.*, 1997) was used for refinement of all structures. Initially ten rounds of rigid body refinement were carried out, followed by multiple rounds of restrained refinement, and TLS (translation, libration, screw-motion) refinement was also used. After each round of refinement, model building was performed in COOT (Emsley *et al.*, 2010). For the refinement of the structures wild-type *sa*NAL, wild-type *sa*NAL in complex with pyruvate, K165C *sa*NAL, K165- γ -thialysine *sa*NAL and K165- γ -thialysine *sa*NAL in complex with pyruvate, a wild-type model with the lysine at position 165 modelled as an alanine was used as a starting structure to produce non-biased electron density around this position. The structure for K165C *sa*NAL, when initially refined with an alanine at position 165 showed extending electron density from the C β that fitted a cysteine residue. The structure for wild-type *sa*NAL, when initially refined with an alanine at position 165 showed extending electron density from the C β that fitted a lysine residue. The structure of the wild-type *sa*NAL in complex with pyruvate had positive electron density around position 165 that appeared to be a lysine residue; however there was continuous positive electron density extending from where the amine group of the lysine residue would be. This extended electron density indicated the lysine was covalently linked to another molecule, most likely the pyruvate group. Therefore the lysine residue in complex with pyruvate was modelled and refined. To build in this non-standard residue restraint and coordinate files for the pyruvate bound lysine had to be generated using the online PRODRG server (Schuttelkopf and Van Aalten, 2004). The files generated were compared to the standard restraint and coordinate files for a lysine residue, manually edited and given the HETATM code KPI. Once the residue was built, rounds of restrained refinement were carried out and the positive electron density around the 165 side chain was no longer present indicating that the lysine in complex with pyruvate was correctly modelled.

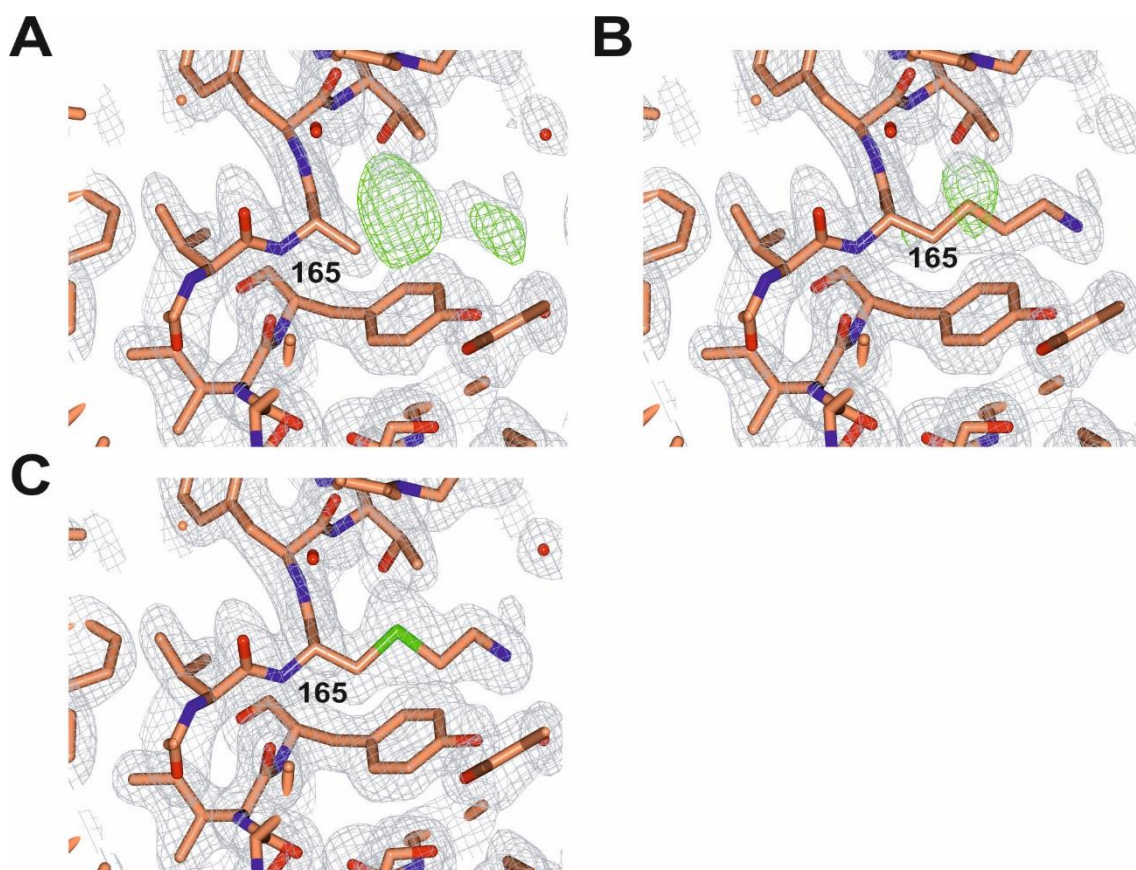


Figure 3.10 Model building of the γ -thialysine side chain into chain C of the K165- γ -thialysine apo crystal structure. The $2F_{\text{obs}} - F_{\text{cal}}$ map is shown in grey and the $F_{\text{obs}} - F_{\text{cal}}$ map is shown in green. Panel A shows the initial model with an alanine at position 165, electron density can be seen extending from the $C\beta$, and positive electron density is shown in green. Panel B shows the electron density at position 165 modelled as a lysine residue, positive electron density, shown in green, can be seen at the γ position indicating an atom that is more electron dense than a carbon. Panel C shows electron density at position 165 modelled as a γ -thialysine residue and no positive electron density can be seen, indicating the side chain is correctly modelled.

A starting structure with an alanine modelled at position 165 was used for the refinement of the nCAA-containing protein K165- γ -thialysine *sa*NAL both apo and in complex with pyruvate. For the K165- γ -thialysine *sa*NAL apo structure, clear positive electron density could be seen extending from the $C\beta$ of the alanine residue which looked to fit the side chain of a lysine, however when a lysine residue was built in and refined into the model there was still excess positive electron density around the γ position of the side chain (Figure 3.10). This increased electron density at the γ position indicated that it was not a carbon atom at this position and was in fact a more electron dense atom such as a sulphur (Figure 3.10). When refining the K165- γ -thialysine *sa*NAL structure in complex with pyruvate a similar process was carried out

as for the apo structure. It was seen that the atom at the γ position was more electron dense than a carbon atom, indicating a sulphur atom at this position, and there was extending electron density from the terminal amine indicating the residue was covalently linked to pyruvate. Therefore restraint and coordinate files for a γ -thialysine and γ -thialysine in complex with pyruvate were generated using the online server PRODRG and given the HETATM codes SLZ and KPY respectively (Schuttelkopf and Van Aalten, 2004). These files were compared to the standard coordinate files for cysteine and methionine to compare the $C\beta-S\gamma$ bond angles and bond lengths, and then manually edited before being used in model building and refinement. Once the γ -thialysine residue had been built into K165- γ -thialysine apo structure a round of refinement was performed, and the excess positive electron density around the γ position of the side chain was no longer present indicating the γ -thialysine residue was correctly modelled. The same process was used for the K165- γ -thialysine structure in complex with pyruvate (Figure 3.11).

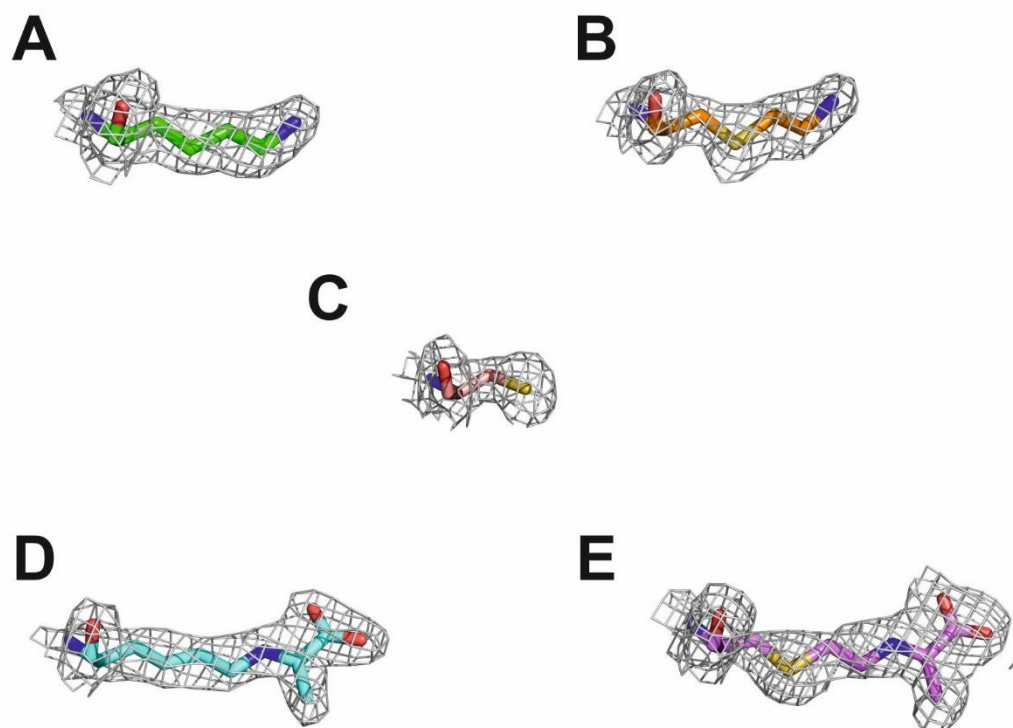


Figure 3.11. Electron density $2F_{\text{obs}}-F_{\text{cal}}$ maps for the side chains at position 165. Panel A shows electron density for the lysine at position 165 in the wild-type *saNAL* apo crystal structure. Panel B shows electron density for the γ -thialysine (HETATM SLZ) at position 165 in the K165- γ -thialysine *saNAL* apo crystal structure. Panel C shows electron density for the cysteine at position 165 in the K165C *saNAL* crystal structure. Panel D shows electron density for the lysine in complex with pyruvate (HETATM KPI) at position 165 in the wild-type *saNAL*, in complex with pyruvate, crystal structure. Panel E shows electron density for the γ -thialysine in complex with pyruvate (HETATM KPY) at position 165 in the K165- γ -thialysine *saNAL*, in complex with pyruvate, crystal structure.

All solved structures contained homotetramers in the unit cell with each monomer having an $(\alpha/\beta)_8$ TIM barrel fold (Figure 3.12). Excitingly in all four subunits for the nCAA containing protein K165- γ -thialysine *saNAL*, in both the apo and pyruvate bound structures, the density clearly showed the non-canonical side chain for γ -thialysine showing efficient and complete incorporation of γ -thialysine into the protein. Interestingly in all subunits there was no evidence for the D enantiomer of the γ -thialysine. This indicates that if the D γ -thialysine does form in the modification process, it is highly likely that this is the aggregated fraction that is removed during the gel filtration step.

3.3.3 Crystal structures

The crystal structures of wild-type *sa*NAL, K165C *sa*NAL and K165- γ -thialysine *sa*NAL were solved to determine the structural effects of incorporating the non-canonical side chain γ -thialysine. No major structural effects were seen from the incorporation of γ -thialysine into *sa*NAL as all structures were seen as homotetramers (panel A Figure 3.12) in the unit cell. Each monomer within the tetramer has the same fold of an $(\alpha/\beta)_8$ TIM barrel (panel B Figure 3.12), which is also seen in the *E. coli* and *H. influenza* NAL structures (Izard *et al.*, 1994, Barbosa *et al.*, 2000). In each TIM barrel monomer, the α helices form the outer wall of the monomer and enclose the parallel β sheets. Therefore neither the cysteine mutant nor the incorporation of γ -thialysine altered the overall fold of the protein. Alignment at C α atoms of the wild-type and K165- γ -thialysine *sa*NAL apo structures, gives an overall RMSD of only 0.13 Å and alignment of K165C and wild-type *sa*NAL gives an overall RMSD of 0.49 Å. The overall fold of the protein was also not affected when both wild-type and the modified enzymes were in complex with pyruvate, as when these structures were aligned by C- α atoms the overall RMSD was 0.31 Å.

The main structural differences could be seen when comparing the apo and pyruvate bound structures of both the wild-type and K165- γ -thialysine *sa*NAL. In both the apo structures there is a region of disordered electron density which corresponds to a loop (residues 138-146), however once pyruvate is bound in both the wild-type and K165- γ -thialysine structures this loop becomes ordered and can clearly be seen in the structure. This loop (residues 138-146) in both of the pyruvate bound structures lies over the entrance to the active site as can be seen in panel C of Figure 3.12. In both the wild-type and K165- γ -thialysine *sa*NAL structures in complex with pyruvate, this rearrangement of the loop region causes a significant change in the positioning of Tyr137, which can be seen in panel C of Figure 3.12. Tyr137 is a residue that has been shown to be essential for catalysis (Daniels *et al.*, 2014) and so this repositioning of Tyr137, where it is shifted towards the terminal amine end of both the lysine and γ -thialysine, may help to stabilise the pyruvate binding in the active site. Surprisingly, in the K165C structure the loop region between residues 138-146 has clear ordered electron density even though there is no evidence for pyruvate in the active site.

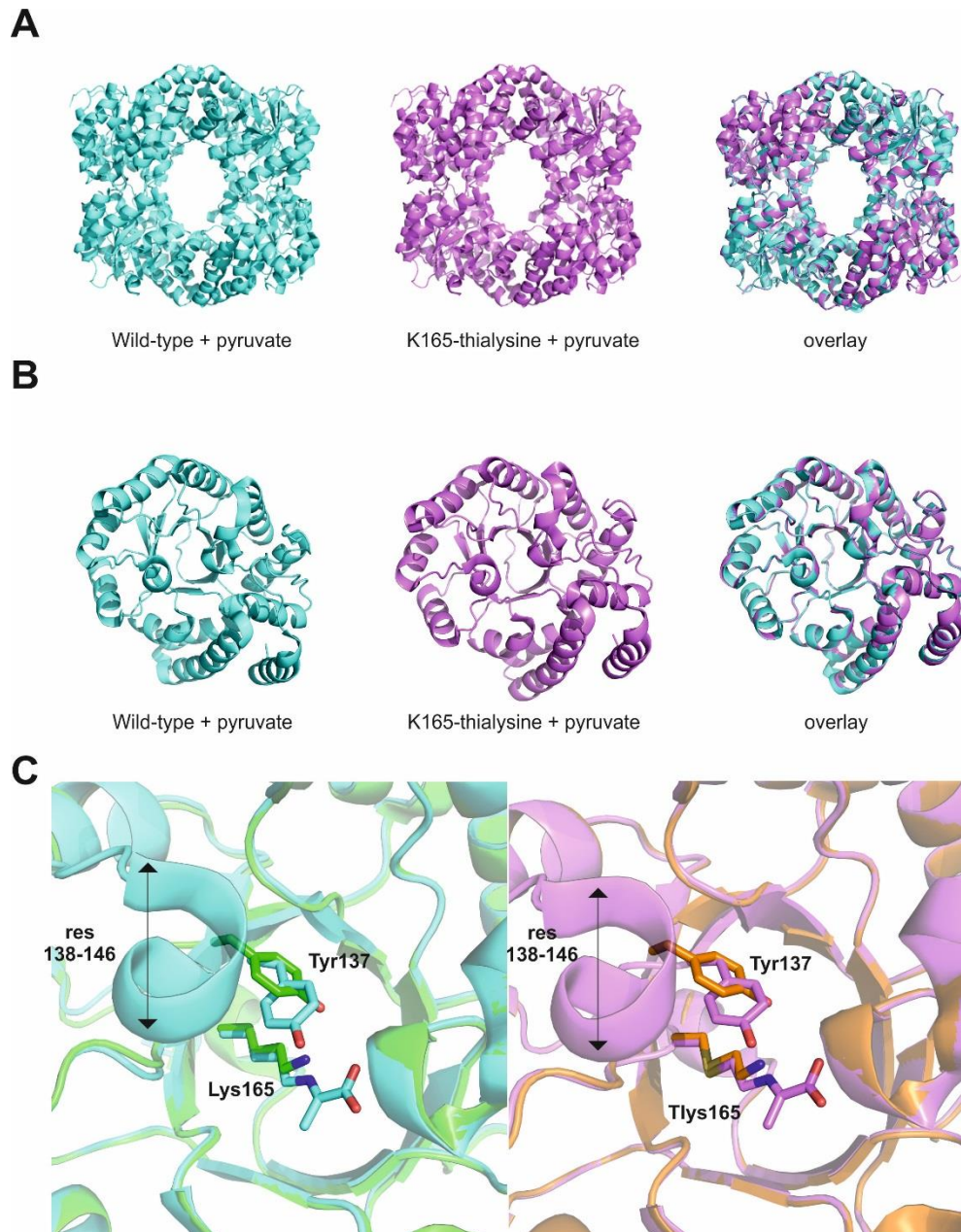


Figure 3.12. Structural comparison of wild-type and K165- γ -thialysine saNAL. Panel A shows the homotetramers for wild-type (cyan) and K165- γ -thialysine (purple) saNAL structures in complex with pyruvate. Both structures are shown separately and overlaid. Panel B shows the $(\alpha/\beta)_8$ TIM barrel fold of the monomer for both wild-type (cyan) and K165- γ -thialysine (purple) saNAL structures in complex with pyruvate, both structures are shown separately and overlaid. Panel C shows overlays of the apo and pyruvate complex structures for wild-type and K165- γ -thialysine saNAL structures; wild-type saNAL apo is shown in green and wild-type saNAL in complex with pyruvate is shown in cyan. The rearrangement of Tyr137 and the appearance of the loop region of residues 138-146 can be seen in the structure in complex with pyruvate. K165- γ -thialysine saNAL apo is shown in orange and K165- γ -thialysine saNAL in complex with pyruvate is shown in purple, again the rearrangement of Tyr137 and the appearance of the loop region of residues 138-146 can be seen in the structure in complex with pyruvate.

From the structural comparison, the incorporation of γ -thialysine at position 165 does not affect the overall fold of the protein. Also the structural changes that occur in the wild-type *sa*NAL structure when pyruvate binds also occur in the K165- γ -thialysine structure when pyruvate binds; the repositioning of Tyr137 and residues 138-146 forming an ordered loop over the entrance to the active site. This identical structural behaviour during the first stage of the catalytic mechanism, pyruvate binding, therefore provides no rationale for the reduction in activity of the ncAA containing enzyme.

The presence of the sulphur atom in the side chain of γ -thialysine may cause localised structural differences that affect the binding of pyruvate, and reduce activity in the modified enzyme compared to the wild-type enzyme. The sulphur at the gamma position in the side chain would affect both the bond lengths and angles around this position compared to those in the wild-type lysine side chain (Ammon *et al.*, 1991). In their respective structures, both the lysine and the γ -thialysine residues at position 165 are seen in the fully extended *anti* configuration, both when they have formed a Schiff base with pyruvate and when they have not. In the Lys165 residue, within the wild-type *sa*NAL structures, the C β -C γ and C γ -C δ bond lengths, and the C β -C γ -C δ bond angle are very similar to the values for the bond lengths and angle for free lysine (shown in Table 3.5). In the γ -thialysine residue, within the K165- γ -thialysine structures, the C β -S γ and S γ -C δ bond lengths, and the C β -S γ -C δ bond angle are also very similar to the values for free γ -thialysine (Ammon *et al.*, 1991). Comparison of the γ -thialysine residue to the lysine residue shows that the C β -S γ and S γ -C δ bond lengths are 0.3 Å longer than the respective C β -C γ and C γ -C δ in the lysine. This increase in bond length may cause the γ -thialysine residue to extend further into the active site of the enzyme. However the C β -S γ -C δ is a tighter bond angle than the respective C β -C γ -C δ found in the wild-type lysine residue, which counteracts the increase in the C β -S γ and S γ -C δ bond lengths. Comparison of the C α -NZ distances in the Lys165 and γ -thialysine 165 residues (Figure 3.13) show that the γ -thialysine residue is approximately 0.3 Å longer than the lysine residue, when comparing the apo structures and 0.3 Å longer when comparing the structures in complex with pyruvate. The overlays at C α of the γ -thialysine and lysine residues show that the increased bond lengths at C β -S γ and S γ -C δ does appear to be compensated by the decreased C β -S γ -C δ bond angle. In the overlay of the apo lysine and γ -thialysine the terminal amine in the γ -thialysine is positioned 0.5 Å (averaged value from all 4 subunits) away from the terminal amine of the lysine. In the pyruvate bound structures the terminal amine of the γ -thialysine

residue is positioned 0.3 Å (averaged value from all 4 subunits) away from the terminal amine of the lysine (Figure 3.13).

In the catalysis of NAL, Schiff base formation is critical for activity. The difference in bond angle and lengths in the γ -thialysine side chain may place the terminal amine into a position that is not optimal for Schiff base formation. Therefore this repositioning of the terminal amine may contribute to the decrease in activity seen in the modified enzyme.

	Wild-type (Lys165) <i>sa</i> NAL	Free lysine		K165 γ -thia- lysine <i>sa</i> NAL	Free γ -thia- lysine
C β - C γ bond length	1.52 Å	1.53 Å	C β - S γ bond length	1.84Å	1.82Å
C γ -C δ bond length	1.52 Å	1.54 Å	S γ -C δ bond length	1.80Å	1.81Å
C β -C γ -C δ bond angle	110.14 °	110°	C β -S γ -C δ bond angle	97.2°	102.9°

Table 3.5. Comparison of bond lengths and angles at the γ position of residue 165 in the wild-type and K165- γ -thialysine *sa*NAL crystal structures to the bond lengths and angles at the γ position in free lysine and γ -thialysine.

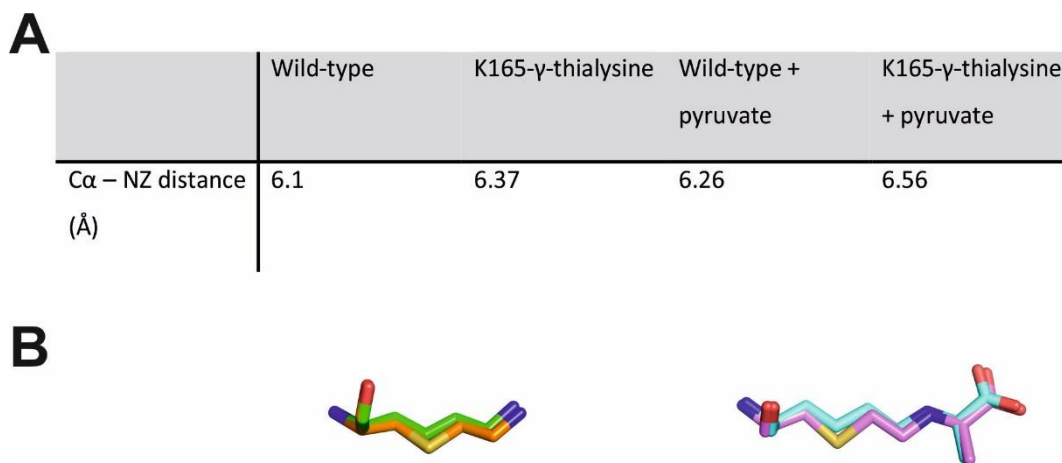


Figure 3.13. Comparison of $\text{C}\alpha$ -NZ distances in lysine and γ -thialysine at position 165 in the wild-type and K165- γ -thialysine *s*aNAL structures. The table in panel A shows the averaged $\text{C}\alpha$ -NZ distances from all four subunits for the wild-type *s*aNAL structures both apo and in complex with pyruvate and for the K165- γ -thialysine *s*aNAL structures both apo and in complex with pyruvate. Panel B shows overlays at $\text{C}\alpha$ of Lys165 (green) and γ -thialysine165 (orange) and overlays at $\text{C}\alpha$ of Lys165 in complex with pyruvate (cyan) and γ -thialysine165 in complex with pyruvate (purple).

To further investigate the effects of the 0.3 Å repositioning of the terminal amine found in the γ -thialysine residue, the pyruvate binding pocket was examined. In the structure of the wild-type *s*aNAL in complex with pyruvate the lysine forms a Schiff base with pyruvate and the carboxylate end of the pyruvate hydrogen bonds to both the back bone and side chain of Ser48 and Ser49 (Figure 3.14). The K165- γ -thialysine structure in complex with pyruvate is seen to make the same interactions, within the pyruvate binding pocket, as the wild-type. Therefore, the introduction of the non-canonical side chain does not appear to structurally affect the pyruvate binding.

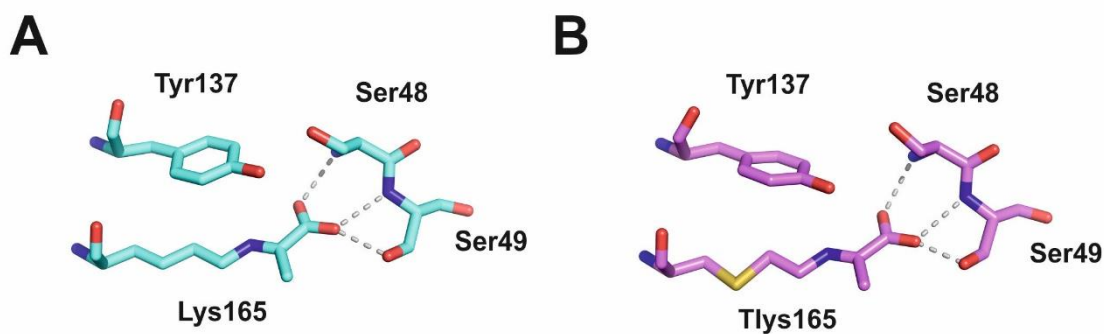


Figure 3.14. Comparison of the pyruvate binding site in the wild-type and K165- γ -thialysine crystal structures. A shows the pyruvate binding site in the wild-type *sαNAL* structure in complex with pyruvate (cyan). B shows the pyruvate binding site in the K165- γ -thialysine *sαNAL* structure in complex with pyruvate (purple). In both structures, the carboxylate end of the pyruvate is seen to form hydrogen bonds with the side chain and backbone of Ser48 and Ser49.

Through analysis of the wild-type and K165- γ -thialysine *sαNAL* crystal structures, it appears there is very little structural difference. Upon pyruvate binding, both wild-type and K165- γ -thialysine undergo the same structural changes and the interactions within the pyruvate binding site are very similar in both proteins. The main structural difference between the wild-type and K165- γ -thialysine is the 0.3 Å repositioning of the terminal amine of the γ -thialysine residue caused by the altered C β -S γ and S γ -C δ bond lengths and the C β -S γ -C δ bond angle. This repositioning of the amine group may contribute to the decrease in activity seen in the modified enzyme, as the repositioning could affect Schiff base formation. However, since the modified enzyme does not appear to have any other significant structural differences, it was investigated whether another property of the modified enzyme was also contributing to the decrease in activity.

3.4 Further kinetic studies of K165- γ -thialysine *sαNAL*

From initial kinetic studies of the modified protein it was seen that once the K165- γ -thialysine protein had been gel filtered and any incorrectly folded material removed, the activity of the modified protein was still reduced compared to the wild-type enzyme. The k_{cat}/K_m of the modified protein is 6-fold lower and the k_{cat} is approximately 10-fold lower than that of the wild-type enzyme, however the K_m of the two proteins differs very little. This decrease in activity could not be attributed to incomplete modification of the protein as the mass

spectrometry data shows no peak for any remaining unmodified K165C or K165dha protein. The crystal structures also show that, as expected, γ -thialysine is an excellent structural mimic of lysine. Structurally, the K165- γ -thialysine protein is extremely similar to the wild-type protein and there are no significant structural differences between the pyruvate binding interactions. The inclusion of a sulphur atom in the side chain of γ -thialysine causes slight structural differences between lysine and γ -thialysine which alters the position of the terminal amine by 0.3 Å, however this does not appear to affect the pyruvate binding site. Therefore the decrease in activity may be caused by the intrinsic chemical properties of having a sulphur atom in the side chain of the γ -thialysine.

The catalytic lysine at position 165 is critical for the Class I aldol mechanism of NAL. A Schiff base is formed between Lys165 and pyruvate; the amine of group of the lysine residue is deprotonated and acts as a nucleophile to attack the carbonyl of pyruvate, forming the Schiff base. The pK_a of free lysine is 10.5 and therefore would not be easily deprotonated at physiological pH, therefore the pK_a of Lys165 must be perturbed to achieve this activity. Perturbations in amino acid pK_a s are commonly seen in enzyme catalysis and can be achieved by the microenvironment of the enzyme active site (Harris and Turner, 2002, Isom *et al.*, 2011). To determine how much the pK_a of the lysine at position 165 is perturbed from free lysine the programme PROPKA was used (Bas *et al.*, 2008). PROPKA allows the pK_a of side chains within a crystal structure to be calculated. The pK_a of Lys165 was calculated to be 7.3; therefore the pK_a is 3.2 units lower than the pK_a of free lysine. The pK_a of free γ -thialysine is 9.5 (Hermann and Lemke, 1968); assuming that the same perturbation in pK_a occurs within the active site of the K165- γ -thialysine enzyme, the pK_a of γ -thialysine would be reduced to 6.3. Therefore, the lysine residue has a pK_a which is 1 unit higher than γ -thialysine. This difference in pK_a seen in the catalytic residue of the two enzymes, which is responsible for Schiff base formation, may cause the proteins to have different pH optimums. This may be the reason for reduced activity in the K165- γ -thialysine enzyme, as all assays to date have been carried out at pH 7.4.

To investigate whether the pH optimum of the modified enzyme has been altered compared to that of the wild-type pH curves for both wild-type and K165- γ -thialysine were measured. pH profiles were measured using LDH coupled kinetics in a three-component buffer system, to ensure constant ionic strength over the full pH range (Ellis and Morrison, 1982). The kinetic parameters were determined and the data were then fit using the equation shown in panel C of Figure 3.15 (Gloss and Kirsch, 1995a).

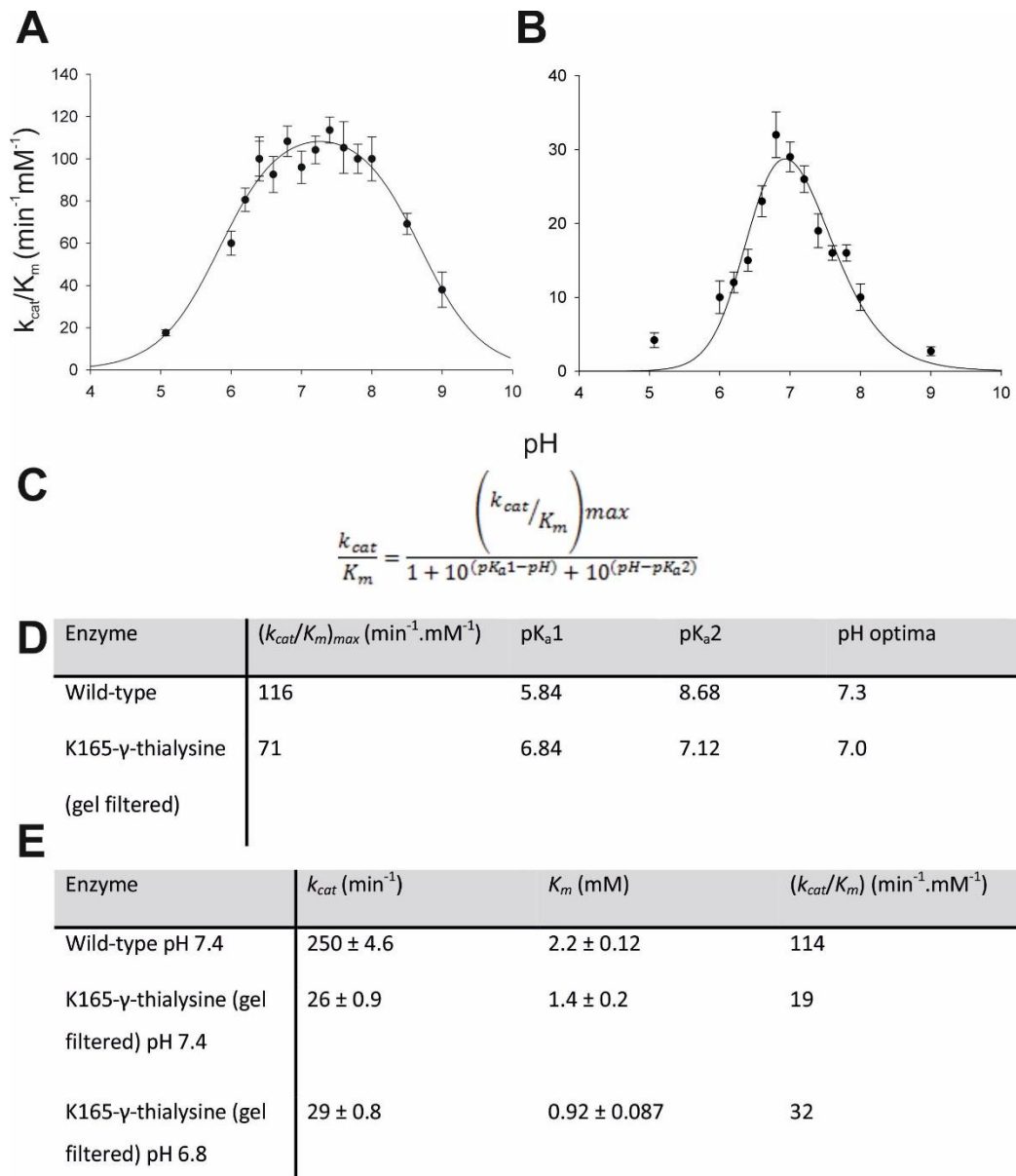


Figure 3.15. pH curves for wild-type and K165- γ -thialysine *sa*NAL. A shows the pH curve for wild-type *sa*NAL. B shows the pH curve for K165- γ -thialysine *sa*NAL. C shows the equation used to fit the data shown in panels A and B, where k_{cat}/K_m is the specificity constant for *N*-acetylneuraminic acid cleavage, $(k_{cat}/K_m)_{max}$ is the theoretical maximum value for k_{cat}/K_m and pK_{a1} and pK_{a2} are the first and second ionisable groups. D shows the theoretical values for $(k_{cat}/K_m)_{max}$, pK_{a1} and pK_{a2} determined from the data fits for both the wild-type and K165- γ -thialysine pH curves. E shows comparison of kinetics parameters k_{cat} , K_m and k_{cat}/K_m at the pH optima of wild-type *sa*NAL, 7.4 from the experimental data, and the pH optima of K165- γ -thialysine *sa*NAL, 6.8 from the experimental data. Kinetic parameters for all pH curves were determined from LDH coupled kinetics carried out in a three component buffer system to ensure constant ionic strength over the full pH range

The pH profiles for the wild-type and K165- γ -thialysine enzymes show that both the shape of the pH curve and the pH optimum of the modified enzyme are altered from that of the wild-type. The theoretical pH optimum, determined from the fit of the pH profiles, of the wild-type enzyme is 7.3 whereas the pH optimum for the modified enzyme is lowered to 7.0 (panels A, B and D Figure 3.15). Comparison of kinetic parameters for the wild-type and K165- γ -thialysine enzymes at their respective experimental pH optima shows that the k_{cat}/K_m of the wild-type is 3.5-fold greater than the k_{cat}/K_m of the modified enzyme, therefore the modified enzyme regains approximately 30 % of the wild-type activity.

The difference in pH optima is not the only difference between the two pH curves, the shape of the K165- γ -thialysine *sa*NAL is also narrower than that of the wild-type. The wild-type enzyme shows a bell shaped pH profile where the pK_a s of the two ionisable groups are 5.84 and 8.68, approximately 3 pH units apart, which should allow the theoretical $(k_{\text{cat}}/K_m)_{\text{max}}$ to essentially be achieved (panel A Figure 3.16). From the fit of the data the theoretical $(k_{\text{cat}}/K_m)_{\text{max}}$ is $116 \text{ min}^{-1}\text{mM}^{-1}$ which compares well with the observed (k_{cat}/K_m) of the wild-type at pH 7.4, which is $114 \text{ min}^{-1}\text{mM}^{-1}$. The pH profile for K165- γ -thialysine *sa*NAL is significantly narrowed, compared to the wild-type. This narrowing of the pH curve brings the pK_a s of the two ionisable groups into much closer proximity than in the wild-type (panel D Figure 3.15). The pK_a s of the two ionisable groups are 6.84 and 7.12, approximately 0.3 pH units apart. Therefore, in the case of the K165- γ -thialysine, the theoretical $(k_{\text{cat}}/K_m)_{\text{max}}$ of $71 \text{ min}^{-1}\text{mM}^{-1}$ can never be achieved since the pK_a s are much closer than 3.5 pH units apart and so the enzyme will never, fully, be in the correct ionisation state for catalysis. From the theoretical titration curve only 30% of the enzyme could be in the active form (Figure 3.16).

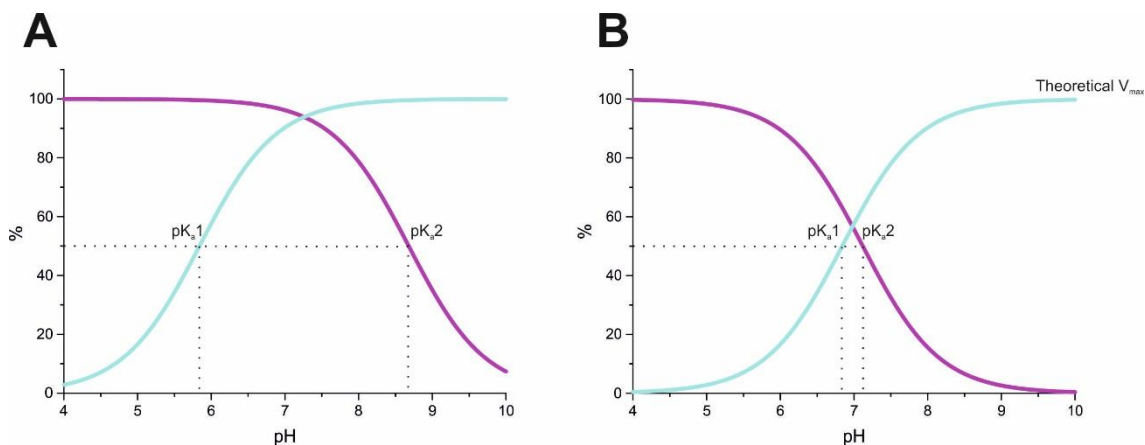


Figure 3.16. pH titration curves. Individual titration curves for two residues, with pK_a s determined in Figure 3.15D, are calculated and plotted using the Henderson-Hasselbach equation. In both curves the blue line represents the titration of the first ionisable group and the purple line represents the titration of a second ionisable group. **A)** Represents the wild-type *saNAL* where the pK_a of the first ionisable group is 5.84 and second is 8.68 therefore the pK_a s are approximately 3 pH units apart and almost all of the enzyme would be in the active form at the pH optima of 7.4. **B)** Represents K165- γ -thialysine *saNAL* where the pK_a of the first ionisable group is 6.84 and second is 7.12 therefore the pK_a s are 0.28 pH units apart and so only 30% of the enzyme would be in the active form.

From the theoretical fit of the K165- γ -thialysine *saNAL* pH curve, this enzyme should be an excellent catalyst with a theoretical $(k_{cat}/K_m)_{max}$ of $71 \text{ min}^{-1}\text{mM}^{-1}$. However due to the imbalance of the pK_a s in the active site K165- γ -thialysine *saNAL* can only achieve $32 \text{ min}^{-1}\text{mM}^{-1}$. The wild-type *saNAL* has evolved to have perfectly synergistic active site and replacing the catalytically essential residue with a structural mimic, which has a much lowered pK_a , has altered the catalysis.

3.5 Summary

The main aim of this chapter was to develop and optimise a method to incorporate ncAAs into the enzyme *saNAL*. Complete incorporation of the ncAA, γ -thialysine, was achieved at position 165, as assessed by mass spectrometry. The chemical modification to introduce γ -thialysine has been well tolerated by the enzyme *saNAL*, producing correctly folded and active protein. Excitingly the ncAA-containing protein can be produced on a scale of up to 50 mg which has allowed extensive structural and kinetic studies.

Initial kinetic studies showed the modified enzyme activity to be 11-fold lower than that of wild-type *saNAL*. However by using size exclusion chromatography it was possible to remove

incorrectly folded material, and by investigating the pH dependency of the enzyme it was possible to find that the pH optima had been shifted, from the wild-type at pH 7.4, to pH 6.8. These further studies showed that the modified K165- γ -thialysine activity is 3.5-fold lower than the wild-type, as judged by k_{cat}/K_m . From further structural and kinetic studies it was possible to attribute this decrease in activity to the perturbation in pK_a s in the active site caused by the sulphur atom in the side chain of γ -thialysine, and also the 0.3 Å shift in the position of the terminal amine group seen in the crystal structure. The perturbation in pK_a of the 165 side chain brings the pK_a s of the two ionisable groups within 3 pH units meaning that the theoretical V_{max} of the enzyme cannot be achieved.

Although the modified enzyme does not achieve the full wild-type activity, a robust method to incorporate ncAAs has been developed that can be carried out routinely. This work sets the scene to extend the incorporation of novel ncAAs throughout the active site of NAL and investigate the effects on activity.

Chapter 4 Production and Screening of Modified Enzymes for Altered Activities

In the previous chapter the incorporation of the ncAA γ -thialysine, via chemical modification, at position 165 in *sa*NAL was described. By carrying out this modification, as a proof of principle, it was possible to develop a robust incorporation method for ncAAs into *sa*NAL. Excitingly, using this method, full conversion to the ncAA was observed via mass spectrometry and ncAA-containing protein was obtained in high yield for both kinetic and crystallographic studies. This chemical method could now be used to introduce a variety of ncAAs individually throughout the active site of *sa*NAL to investigate whether they could be used to alter the activity of the enzyme.

This chapter will describe the use of the chemical ncAA incorporation method to introduce up to 13 different non-canonical side chains at a variety of different positions throughout the active site of *sa*NAL. It will also describe screening these ncAA-containing enzymes for altered substrate specificity with a variety of different aldehydes.

4.1 Selection of residues for modification

The aim of this work was to use ncAAs to alter substrate specificity, and so the first stage was to select the sites at which the non-canonical side chains were to be incorporated. There are many directed evolution studies that have shown residues away from the active site can have a significant effect on substrate specificity (Leferink *et al.*, 2014, Romero and Arnold, 2009, Shimotohno *et al.*, 2001). However, in this instance, to reduce the sequence space to be screened and since an X-ray crystal structure of wild-type *sa*NAL had been solved, we decided to look at replacing residues within the active site of the protein, and within close proximity (~ 5 Å) to the substrate. Unfortunately there is no crystal structure of *sa*NAL with the full length *N*-acetylneuraminic acid (Neu5Ac) product bound; however there is a structure of the *E. coli* NAL with Neu5Ac in the active site (Daniels *et al.*, 2014). Therefore using the *E. coli* NAL structure, residues within approximately 5 Å (Table 4.1) of the product were selected for conversion to cysteine, and subsequently modified. In Figure 4.1 the residues selected for mutagenesis to cysteine can be seen, and their proximity to the, enzyme-bound, full length Neu5Ac product is shown in Table 4.1. In total thirteen sites were selected to be converted into cysteine mutants by site-directed mutagenesis.

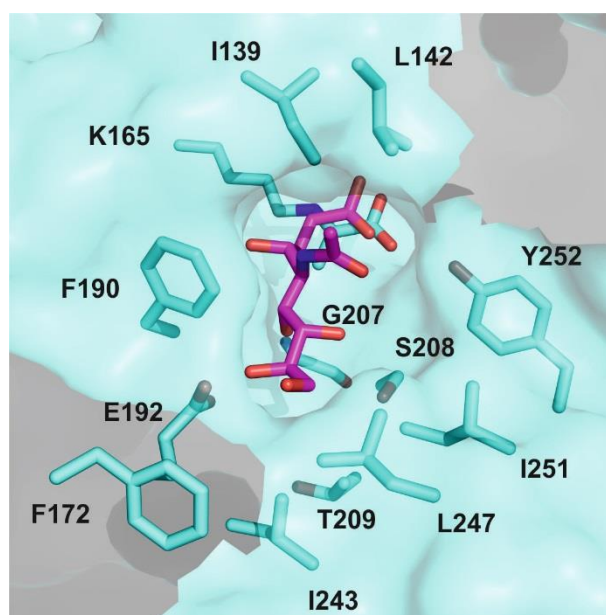


Figure 4.1. Image of the wild-type *S. aureus* NAL active site (PDB 4ah7 shown in cyan) (Timms *et al.*, 2013) with the full length N-acetylneuraminic acid (Neu5Ac) taken from the *E. coli* NAL in complex with Neu5Ac, overlaid (PDB 4bwl shown in magenta) (Daniels *et al.*, 2014). The residues that were targeted for site-directed mutagenesis into cysteine and subsequent modification into non-canonical amino acids are highlighted as sticks.

Residue	Distance to Neu5Ac (Å)	Residue	Distance to Neu5Ac (Å)
I139	3.0	S208	3.2
L142	4.2	T209	4.8
K165	-	I243	4.1
F172	5.7	L247	4.3
F190	5.3	I251	4.6
E192	2.9	Y252	3.5
G207	3.3		

Table 4.1 Residues in the *S. aureus* NAL selected for conversion to cysteine, and their relative distances from the product Neu5Ac in the active site. Distances were calculated by overlaying the *sa*NAL X-ray crystal structure, in complex with pyruvate, with the *E. coli* NAL in complex with Neu5Ac and the distances between the closest residue atom and product atom were measured. A distance is not shown for K165 as it forms a covalent link to the product.

Previous studies on NAL have shown that it is highly specific for the ketone donor, pyruvate, however NAL is far more promiscuous regarding the aldehyde acceptor (Fitz *et al.*, 1995, Kim *et al.*, 1988). Therefore the majority of sites selected for mutagenesis were towards the aldehyde end of the active site, as this end may be more easily manipulated. As can be seen from Figure 4.1 a wide variety of residues were selected to be exchanged for ncAAs, from very small residues like glycine to much bulkier side chains like tyrosine and phenylalanine. All the residues selected would be converted to non-canonical side chains using the chemical modification method described in the previous chapter.

4.2 Production of cysteine mutants

To allow incorporation of non-canonical side chains at the selected sites in the active site of *sa*NAL (shown in Figure 4.1.), it was necessary to produce cysteine mutants at these positions. Site-directed mutagenesis was used to produce the cysteine mutants, and the mutant genes were provided by Sam Hickman (University of Leeds).

Once the cysteine mutants were successfully generated they were expressed and purified in the same way as wild-type *sa*NAL using batch nickel affinity chromatography (Section 2.4.1).

All cysteine variants expressed at reduced yields compared to the wild-type. Wild-type *saNAL* expresses at approximately 100 mg per litre growth medium, whereas the cysteine variants expressed at between 30-80 mg per litre growth medium depending on the variant. All cysteine variants expressed in high enough yields for modifications to be carried out. Once the cysteine variants were expressed and purified, they were lyophilised (Section 2.4.6) before being modified to incorporate ncAAs. Before modifications were carried out, ESI mass spectrometry was performed by Dr James Ault to ensure that the cysteine variants were the correct mass.

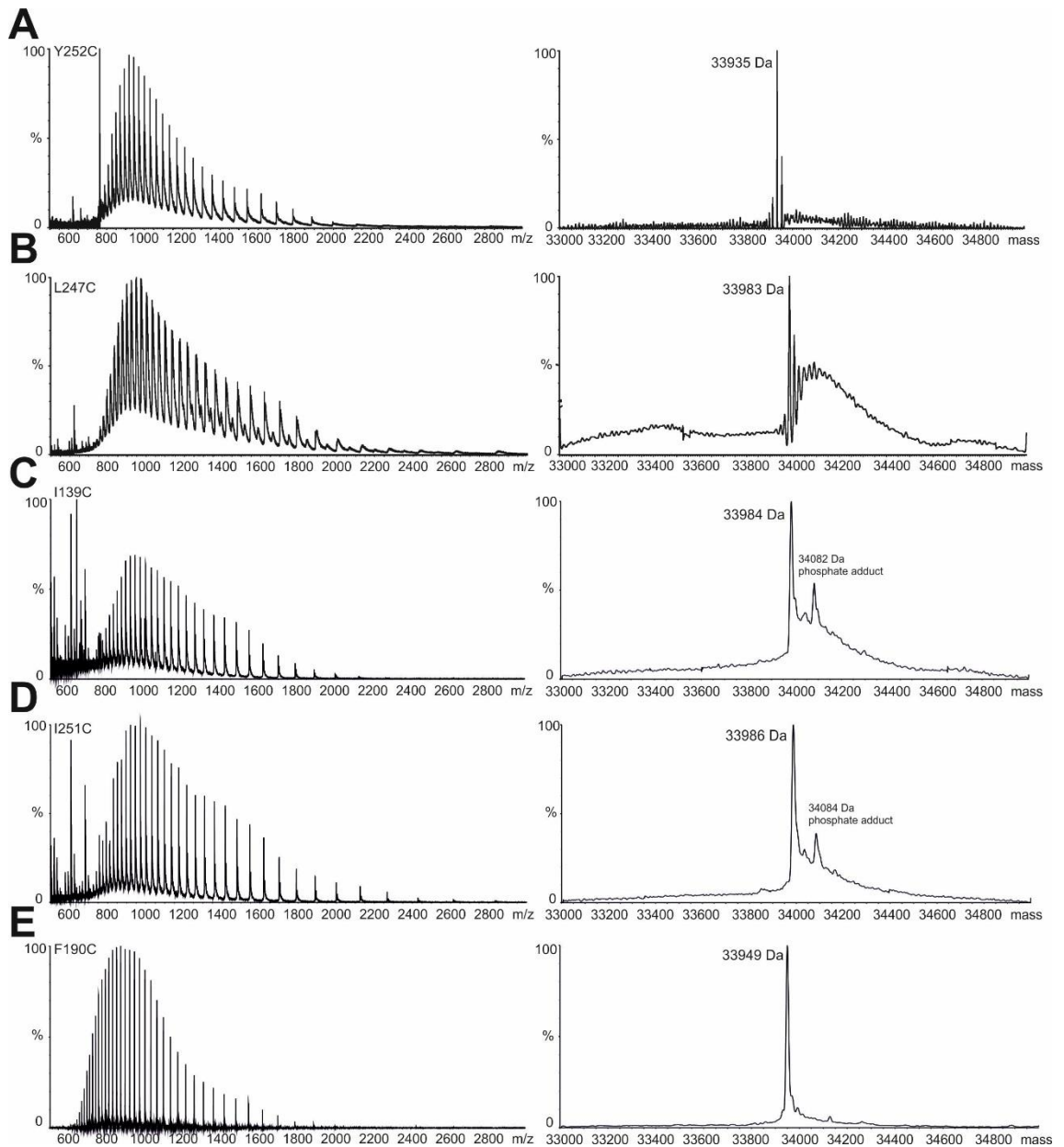


Figure 4.2 ESI mass spectra of cysteine variants. In all spectra the raw data is shown to the left and the deconvoluted data shown to the right. Panel A shows mass spectrum for Y252C, the expected mass was 33934 Da and the observed mass was 33935 Da. Panel B shows mass spectrum for L247C, the expected mass was 33984 Da and the observed mass was 33983 Da. Panel C shows mass spectrum for I139C, the expected mass was 33984 Da and the observed mass was 33984 Da a peak at 34082 Da can also be seen in the spectrum which is +98 Da and most likely corresponds to a phosphate adduct. Panel D shows mass spectrum for I251C, the expected mass was 33984 Da and the observed mass was 33986 Da a peak at 34084 Da can also be seen in the spectrum which is +98 Da and most likely corresponds to a phosphate adduct. Panel E shows mass spectrum for F190C, the expected mass was 33950 Da and the observed mass was 33949 Da.

4.3 Modifications to generate a variety of ncAAs

Once cysteine variant proteins had been expressed and purified, it had to be established that the modification protocols, developed to convert K165C into K165- γ -thialysine, would also work on these variants. Modifications to convert the cysteine variants into dha-containing proteins were carried out first to ensure the first stage of the modification worked. Once it was possible to produce dha-containing proteins then the conditions to produce a variety of ncAAs were developed.

4.3.1 Conversion of cysteine variants to dha

To determine whether the same conditions to convert K165C into K165dha were applicable for the other cysteine variants, the same procedure to convert the cysteine to dehydroalanine was used. 2.5 mg of lyophilised protein was taken up in 1.25 mL of pre-warmed (37°C) sodium phosphate buffer (50 mM, pH 8.0) containing 6 M urea. The protein solution was vortexed to ensure that the protein had fully redissolved, and then 100 μ L of the diBr (0.13 mg μ L⁻¹ in DMF) solution was added. The solution was then incubated at 37°C with shaking at 200 rpm for 1 hr 30 min.

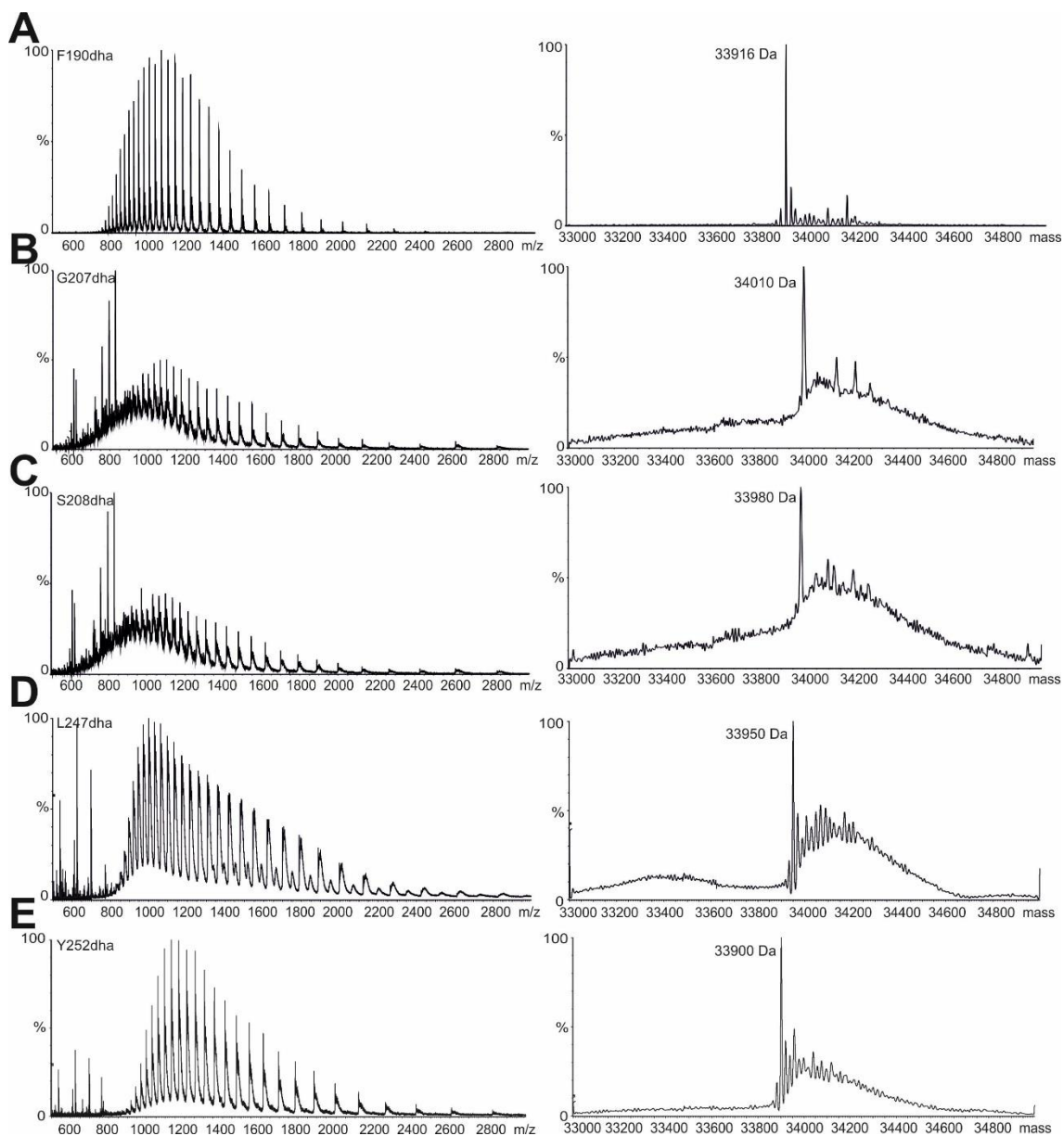


Figure 4.3. ESI mass spectra, showing various modifications of cysteine variants into dehydroalanine (dha) raw data is shown on the left and deconvoluted data on the right. All panels show mass spectra of dehydroalanine-containing proteins which have been modified using the conditions; 2.5 mg of cysteine variant dissolved in 1.25 mL of pre-warmed (37°C) sodium phosphate buffer (50 mM, pH 8.0) containing 6 M urea, then 100 μ L of diBr (0.13 mg μ L⁻¹ or 0.08 mg μ L⁻¹ in DMF) added and incubated for 1 hr 30 min. Panel A shows mass spectrum of F190dha, the expected mass was 33917 Da and the observed mass was 33916 Da. Panel B shows mass spectrum of G207dha, the expected mass was 34007 Da and the observed mass was 34010 Da. Panel C shows mass spectrum of S208dha, the expected mass was 33977 Da and the observed mass was 33980 Da. Panel D shows mass spectrum of L247dha, the expected mass was 33951 Da and the observed mass was 33950 Da. Panel E shows mass spectrum of Y252dha, the expected mass was 33901 Da and the observed mass was 33900 Da.

Once the reaction had been incubated a 70 μL sample was taken and analysed by ESI mass spectrometry. If a peak of correct molecular mass for the dha-containing protein was observed, the modification was considered successful. If no peak was observed for any remaining cysteine-containing protein, the modification was considered to have gone to completion. Due to samples for mass spectrometry being taken from the reaction mixture, containing 6 M urea, the sample preparation was sometimes not as efficient, reducing the intensity of the protein in the mass spectra, as can be seen from the noise in the deconvoluted spectra. However if there was a single main peak, of the correct molecular mass, the modification was considered successful.

In some of the modifications, two peaks were observed in the ESI-mass spectra, the first had the correct molecular mass for the dehydroalanine intermediate and a second was 174 Da larger in mass (Figure 4.4). It was determined that this extra mass was caused by an uncyclized sulfonium intermediate, that comes from the dehydroalanine not eliminating from the protein after it has reacted (Nathani *et al.*, 2012, Chalker *et al.*, 2011). Incubation at 37 °C is reported to remove this uncyclized intermediate (Chalker *et al.*, 2011). However, since the modifications had already been carried out at 37 °C, it was hypothesised that this 174 Da adduct could be caused by the excess of diBr in the reaction mixture. Modifications were therefore carried out under the same conditions as before but using 0.08 mg μL^{-1} of diBr. Conversion of the cysteine residue into dehydroalanine still went to completion after 1 hr 30 min but no evidence for the 174 Da adduct was seen, therefore the lower concentration of 0.08 mg μL^{-1} diBr was used in all subsequent modifications.

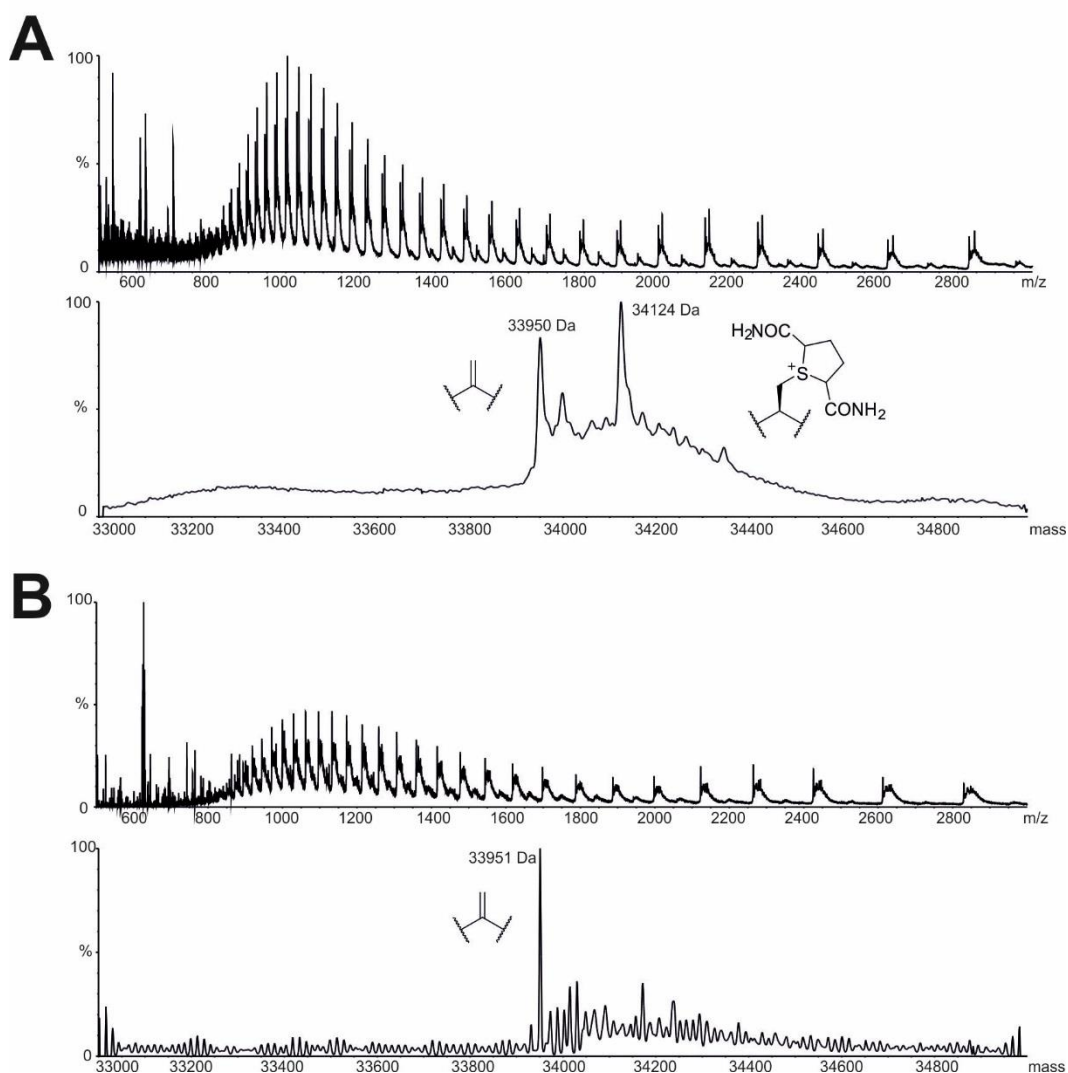


Figure 4.4. ESI mass spectra of I251C modifications to I251dha. In both spectra the raw data is shown above and the deconvoluted data shown below. Panel A shows mass spectra of I251dha which has been modified using the conditions; 2.5 mg of I251C protein dissolved in 1.25 mL of pre-warmed (37°C) sodium phosphate buffer (50 mM, pH 8.0) containing 6 M urea, then 100 μL of diBr (0.13 $\text{mg } \mu\text{L}^{-1}$ in DMF) added and incubated for 1 hr 30 min. Two peaks are clearly visible in the spectrum the first at 33950 Da corresponding to correctly modified I251dha which has an expected molecular mass of 33950 Da, and the second peak 174 Da larger than the first peak which corresponds to the uncyclized sulfonium ion. Panel B shows mass spectra of I251dha which has been modified using the conditions; 2.5 mg of I251C protein dissolved in 1.25 mL of pre-warmed (37°C) sodium phosphate buffer (50 mM, pH 8.0) containing 6 M urea, then 100 μL of diBr (0.08 $\text{mg } \mu\text{L}^{-1}$ in DMF) added and incubated for 1 hr 30 min. Only one distinct peak can be seen in this spectrum with a molecular mass of 33951 Da corresponding to the correctly modified I251dha which has an expected molecular mass of 33950 Da.

4.3.2 Selection of ncAA side chains

Having demonstrated it was possible to routinely produce dha-containing proteins from the cysteine variants, the dha-containing proteins could now be modified with thiols to produce non-canonical side chains. All thiols chosen for use in the modification procedure were commercially available, to allow easy and scalable modifications, as the process would not be dependent on the amount of thiol that could be synthesised. The thiols selected were chosen to produce a wide variety of non-canonical side chains (Table 4.2). Some thiols produced elongated versions of canonical amino acids and some produced side chains with chemistries not present, or uncommon, within the twenty proteogenic amino acids.

Cysteine variants	Thiols			
K165C S208C	1)	2)	3)	4)
L247C I251C				
L142C F190C	5)	6)	7)	8)
Y252C G207C				
E192C I139C	9)	10)	11)	12)
F172C N170C				
T209C I243C	13)	14)		

Table 4.2. Cysteine variants and thiols used in the chemical modification procedure. The first column shows the positions converted to cysteine for subsequent conversion to ncAAs. The second column shows the thiols used to produce the ncAA side chains. Thiols 1-14 are 3-mercaptopropionic acid, aminoethanethiol, benzylmercaptan, methylthioglycolate, sodium mercaptopyruvate, mercapto-2-propanol, D-cysteine, L-cysteine, 4-pyridylethylmercaptan, mercapto-1-butanol, mercaptoacetic acid, thioglycerol, 2-furanmethanethiol, 2-propene-1-thiol.

4.3.3 Optimisation of Michael addition of thiols

The next step was to convert the dha residue into a range of different ncAAs. The thiols had been chosen to produce a variety of different non-canonical side chains (Table 4.2). For production of the various ncAA-containing proteins, we investigated whether a single set of modification conditions could be found. However, due to the variation in the physical properties of the thiol compounds, for example some were provided as liquids, other solids, each had to have conditions optimised for use in the modification procedure. Optimised conditions for the conversion of dehydroalanine into the non-canonical side chains are shown in Table 4.4. The protein F190dha was chosen for use in the optimisation experiments. Optimisations were carried out on a 2.5 mg scale, 1.25 mL (2 mg mL^{-1}) of F190dha was taken and a Michael addition with a thiol compound was then carried out, this was performed with each thiol shown in Table 4.2.

For initial test modifications the conditions used to convert K165dha to K165- γ -thialysine were used as a starting point as they had been previously successful. In the K165- γ -thialysine study the aminoethanethiol was dissolved in Tris/HCl (2 M, pH 8.8) to a concentration of $0.1 \text{ mg } \mu\text{L}^{-1}$ before being added to the dha-containing protein; therefore this was used for the solid thiol compounds (aminoethanethiol (**2** Table 4.2), mercaptopyruvate (**5**), mercaptoacetic acid (**11**), L-cysteine (**8**) and D-cysteine (**7**)), which were all soluble in aqueous solvent. In the selection of thiol compounds a number of them were provided as liquids; from the literature undiluted β -mercaptoethanol was added to dehydroalanine-containing protein which produced conversion to the non-canonical side chain (Chalker *et al.*, 2011). However, to ensure a constant pH in the reactions with the liquid thiols, Tris/HCl (2 M, pH 8.0) was added to the thiol compounds, in a 1:1 ratio, before addition to the reaction. For the thiols that were soluble in aqueous solvent (mercapto-2-propanol (**6**) and thioglycerol (**12**)), addition of Tris/HCl and then incubation with the dehydroalanine containing protein worked well, but it was found that some of the thiol compounds were not soluble in aqueous solvent. Mercapto-1-butanol (**10**), methylthioglycolate (**4**), benzylmercaptan (**3**), 4-pyridylethyl mercaptan (**9**), 2-furanmethane thiol (**13**) and 2-propene-1-thiol (**14**) would not solubilise in the aqueous reaction mixture and would cause precipitate to form if they were added directly to the reaction. Therefore these compounds were first solubilised in DMF and then Tris/HCl (2 M, pH 8.0) added to ensure the pH was approximately 8.0 before being added to the reaction mixture. This approach worked well for mercapto-1-butanol, methylthioglycolate, 2-furanmethane thiol and 2-propene-1-thiol, but it was seen that benzylmercaptan and 4-pyridylethylmercaptan would

still cause some precipitate in the reaction mixture which was aggravated with the addition of Tris/HCl. Therefore these were solubilised in DMF, before being added to the dehydroalanine containing protein. Some precipitate still formed in the reactions with benzylmercaptan and 4-pyridylethylmercaptan but it could be easily removed by centrifugation, after the incubation time was completed, and did not affect the modification to produce the non-canonical side chain. The conditions used for each thiol Michael addition are summarised in Table 4.3

Thiol	Conditions	Incubation
Aminoethanethiol	40 μL (0.1 mg μL^{-1} Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Mercaptoacetic acid	40 μL (0.1 mg μL^{-1} Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Mercapto-1-butanol	40 μL (20 μL thiol, 10 μL DMF, 10 μL Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Mercapto-2-propanol	40 μL (20 μL thiol, 20 μL Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Methylthioglycolate	40 μL (20 μL thiol, 10 μL DMF, 10 μL Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Thioglycerol	40 μL (20 μL thiol, 20 μL Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Sodium mercaptopyruvate	40 μL (0.1 mg μL^{-1} Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
Benzylmercaptan	30 μL (10 μL thiol, 20 μL DMF)	2 hours, 37 °C, 200 rpm
4-Pyridylethylmercaptan	30 μL (10 μL thiol, 20 μL DMF)	2 hours, 37 °C, 200 rpm
2-Furanmethane thiol	40 μL (20 μL thiol, 10 μL DMF, 10 μL Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
L-cysteine	120 μL (0.13 mg μL^{-1} Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
D-cysteine	120 μL (0.13 mg μL^{-1} Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm
2-Propene-1-thiol	40 μL (20 μL thiol, 10 μL DMF, 10 μL Tris/HCl (2 M pH 8.0))	2 hours, 37 °C, 200 rpm

Table 4.3. Conditions for modification with thiol compounds to produce non-canonical side chains. The conditions shown are to convert 2.5 mg of dha-containing protein into 2.5 mg of nCAA containing protein.

Once the thiol compound had been added to the dha-containing protein the reaction was incubated at 37 °C with shaking at 200 rpm. To assess whether the Michael additions with the thiol compounds had been successful, ESI-mass spectrometry was used, performed by Dr James Ault. Samples were taken during the incubation period and it was seen that a 2 hour incubation time was optimal for the modification process to go to completion. These conditions for nAA incorporation were optimised using F190dha protein and the ESI mass spectra are shown for the successful modifications at this position (Figure 4.5 and Figure 4.6). It was seen that the conditions used for modification, with the various thiols, at position 190 could also be used for modification at the other positions selected. The successful reactions had produced nAAs in the protein, and so it was necessary to develop a naming convention for them. The names and abbreviations shown in Table 4.4 were used.

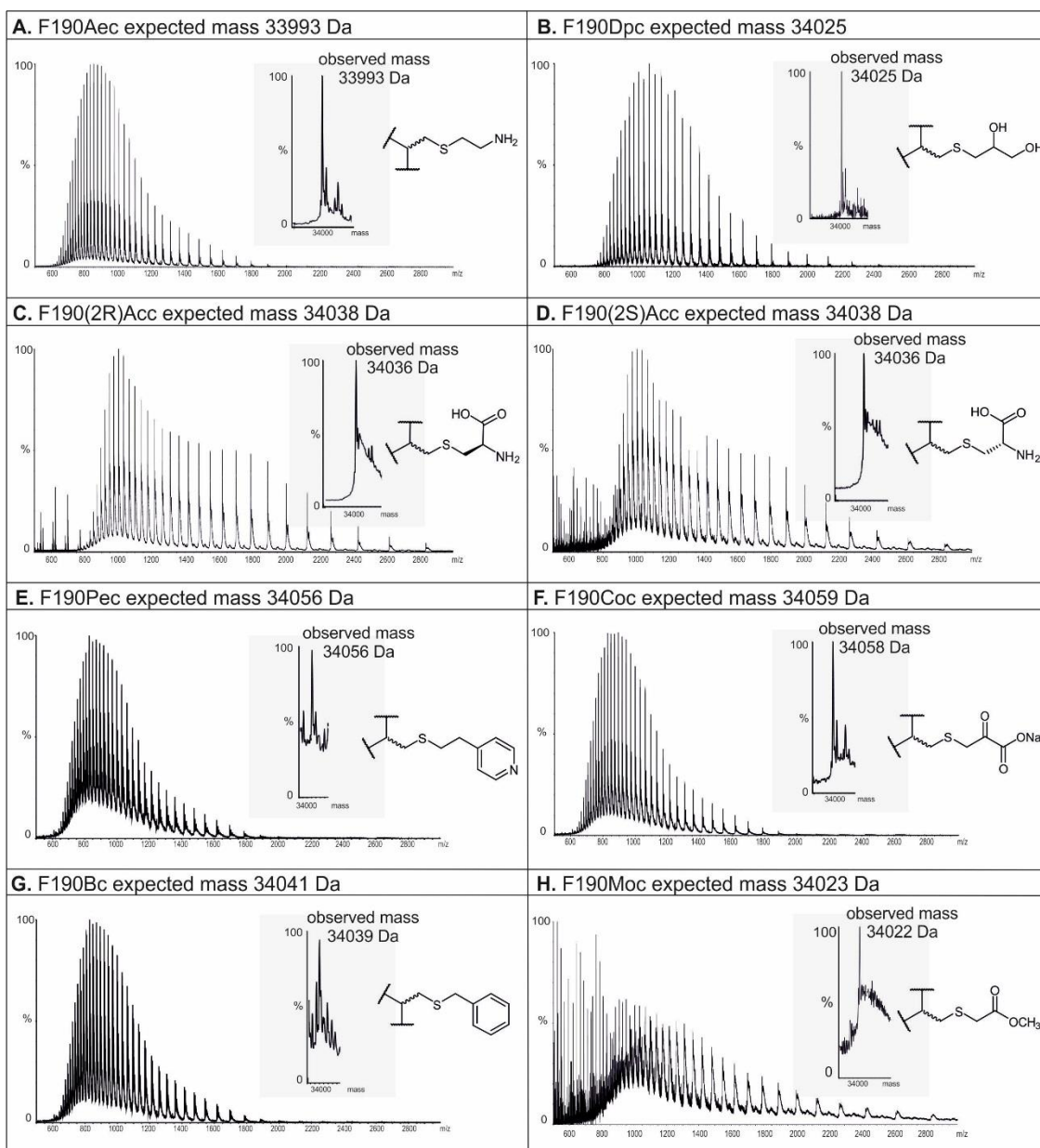


Figure 4.5 ESI mass spectra of modifications of F190dha with various thiols using conditions stated in Table 4.3. Each panel shows the raw mass spectrometry data with the deconvoluted data inset with a grey back ground, the nCAA side chain incorporated is also shown. Panels show mass spectra of F190dha modified with; A) aminoethanethiol to produce F190Aec, expected mass 33993 Da observed mass 33993 Da, B) thioglycerol to produce F190Dpc, expected mass 34025 Da observed mass 34025 Da, C) L-cysteine, to produce F190(2R)Acc, expected mass 34038 Da observed mass 34036 Da, D) D-cysteine to produce F190(2S)Acc, expected mass 34038 Da observed mass 34036 Da, E) pyridylethylmercaptan to produce F190Pec, expected mass 34056 Da observed mass 34056 Da, F) mercaptopyruvate to produce F190Coc, expected mass 34059 Da observed mass 34059 Da, G) benzylmercaptan to produce F190Bc, expected mass 34041 Da observed mass 34039 Da, H) methylthioglycolate to produce F190Moc, expected mass 34023 Da observed mass 34022 Da.

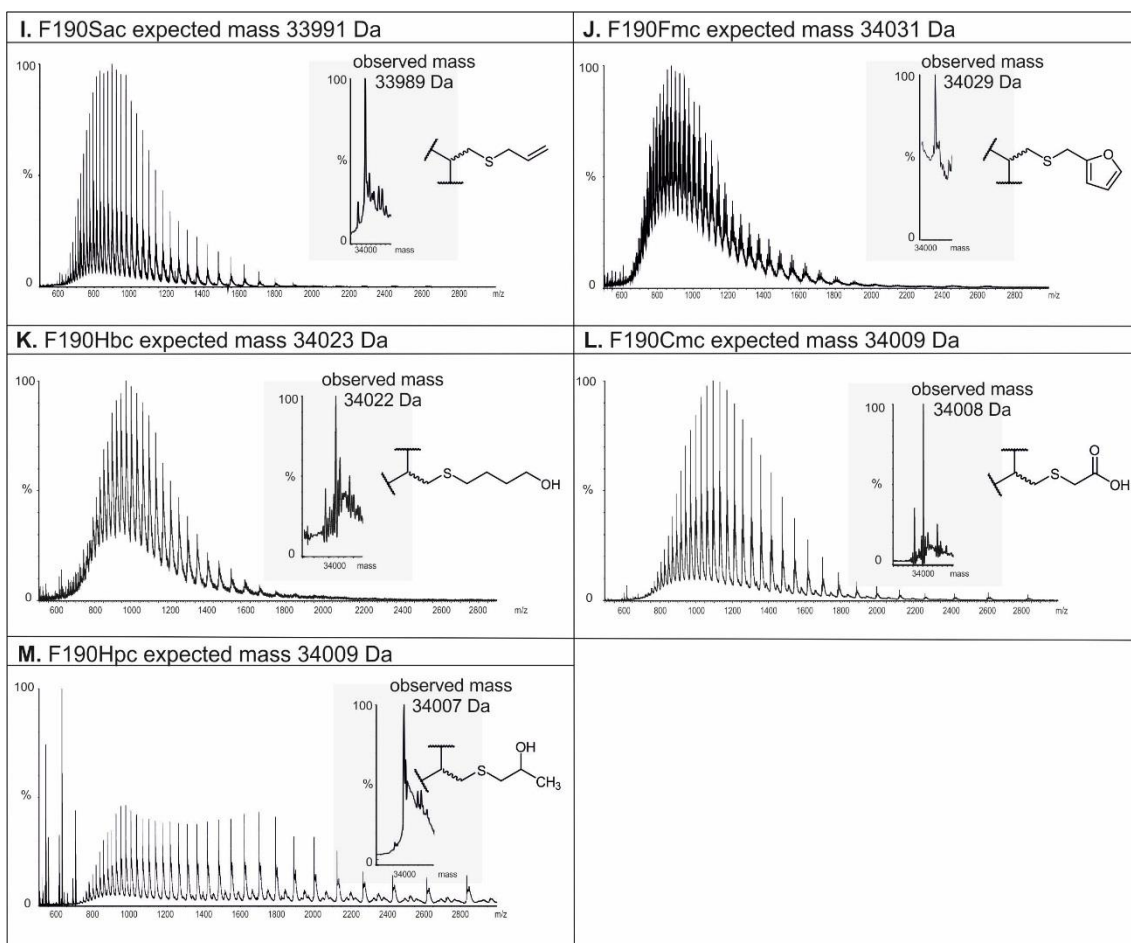


Figure 4.6 ESI mass spectra of modifications of F190dha with various thiols using conditions stated in Table 4.3. Each panel shows the raw mass spectrometry data with the deconvoluted data inset with a grey back ground, the nAA side chain incorporated is also shown. Panels show mass spectra of F190dha modified with; I) propene-1-thiol to produce F190Sac, expected mass 33991 Da observed mass 33989 Da, J) furanmethane thiol to produce F190Fmc, expected mass 34031 Da observed mass 34029 Da, K) mercapto-1-butanol to produce F190Hbc, expected mass 34023 Da observed mass 34022 Da, L) mercaptoacetic acid to produce F190Cmc, expected mass 34009 Da observed mass 34008 Da, M) mercapto-2-propanol to produce F190Hpc, expected mass 34009 Da observed mass 34007 Da.

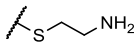
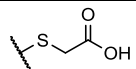
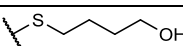
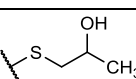
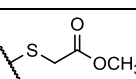
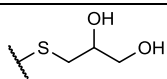
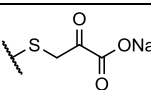
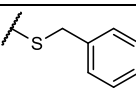
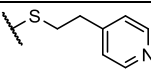
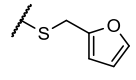
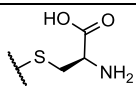
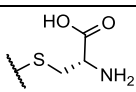
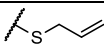
Thiol compound	ncAA side chain	ncAA name	Abbreviation
Aminoethanethiol		2-aminoethyl cysteine	Aec
Mercaptoacetic acid		carboxymethyl cysteine	Cmc
Mercapto-1-butanol		4-hydroxybutyl cysteine	Hbc
Mercapto-2-propanol		2-hydroxypropyl cysteine	Hpc
Methylthioglycolate		2-methoxy-2-oxoethyl cysteine	Moc
Thioglycerol		2, 3-dihydroxypropyl cysteine	Dpc
Mercaptopyruvate		2-carboxy-2-oxoethyl cysteine	Coc
Benzylmercaptan		benzylcysteine	Bc
4-Pyridylethyl mercaptan		2-(pyridin-4-yl)ethyl cysteine	Pec
Furanmethane thiol		furan-2-ylmethyl cysteine	Fmc
L-cysteine		(2R)-2-amino-2-carboxyethyl cysteine	(2R)Acc
D-cysteine		(2S)-2-amino-2-carboxyethyl cysteine	(2S)Acc
2-Propene-1-thiol		S-allyl cysteine	Sac

Table 4.4. Table showing ncAA side chains, names and abbreviations produced from modifications with thiol compounds.

4.4 Screening modified enzymes for activity

Having established methods for producing a variety of ncAAs at different positions in *saNAL*, these modified enzymes could be screened to determine the effects of introducing the non-canonical side chain on enzyme substrate specificity. It was decided that the enzymes would be screened for activity in the synthetic, aldol condensation direction as this would produce variations of sialic acid that may be of use in complex oligosaccharide syntheses. In the screening process, the donor in the aldol condensation would remain as pyruvate and modified enzymes would be screened for altered aldehyde acceptor activities. A variety of different aldehyde acceptors were chosen for screening which would allow the effect of ncAA on altered substrate length, stereochemistry and group substitutions to be assessed. Ten aldehydes in total were chosen (shown in Figure 4.7).

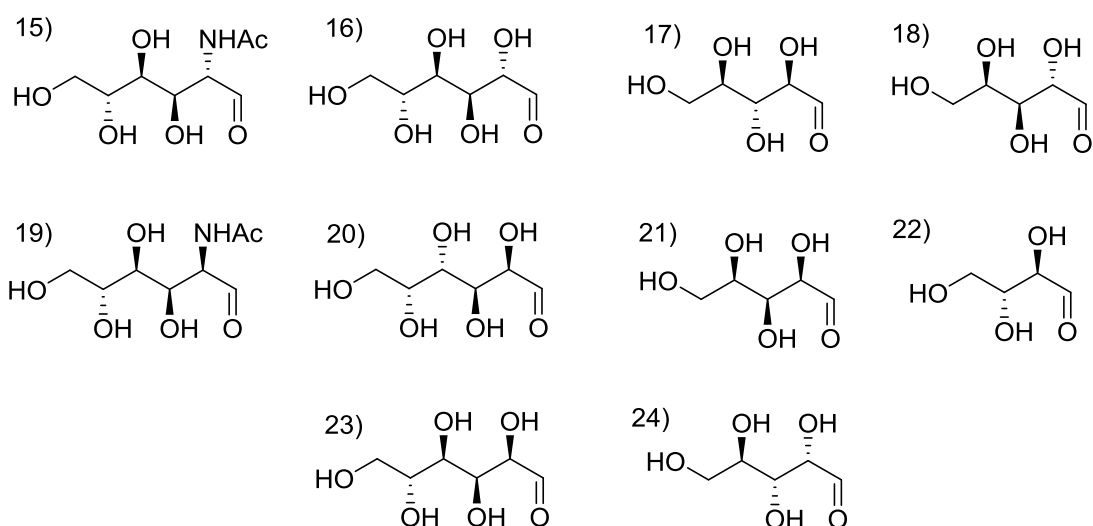


Figure 4.7. Aldehydes used for screening modified enzymes for activity. The aldehydes from 15-24 are; *N*-acetyl-D-mannosamine (15), D-mannose (16), D-ribose (17), D-lyxose (18), *N*-acetyl-D-glucosamine (19), D-galactose (20), D-xylose (21), D-erythrose (22), D-glucose (23), D-arabinose (24).

4.4.1 TBA assay for screening

To screen the modified enzymes for aldol condensation activity, it was necessary to have an assay that would detect the new carbon-carbon bond formation. The thiobarbituric acid assay (TBA assay) has been used for the detection and quantification of sialic acids (Warren, 1959). The TBA assay is a colorimetric assay that has previously been used to follow the enzyme catalysed aldol condensation of pyruvate and aldehyde (Wolterink-Van Loo *et al.*, 2009, Buchanan *et al.*, 1999), and so this assay was used to screen the modified enzymes for activity. In the screening procedure when an aldol condensation reaction has occurred, the TBA assay can be used to detect the aldol condensation product. The aldol condensation product is first cleaved using periodic acid to form a 1,3-dicarbonyl, which will then react with TBA to form a highly conjugated pink product (Figure 4.8). For the TBA assay to be viable the group attached to the C2 of the aldehyde acceptor (X in Figure 4.8) must be either a hydroxyl or an *N*-acetyl group, and so the aldehydes used in the screening were selected taking this fact into consideration.

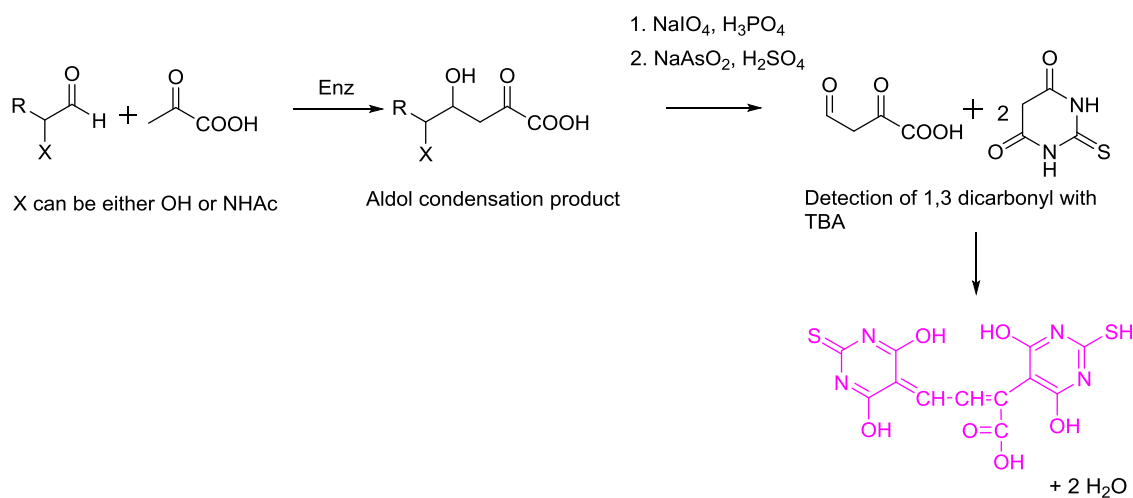


Figure 4.8. Schematic of the thiobarbituric acid (TBA) assay, used to assay modified variants activity with various aldehydes. The 1,3-dicarbonyl reacts with TBA to form a highly conjugated product (shown in pink) which turns pink and can be easily detected by eye or at 550 nm.

To allow screening to be carried out in a plate based format the volumes used in the assay needed to be significantly reduced from the original procedure. Previous work within our group had used the TBA assay in a plate format (Angela Kinnell, PhD thesis, University of Leeds) and so this method was used when screening the modified enzymes.

4.4.2 Screening modified enzymes for activities

When screening the modified enzymes for activity, reactions were carried out in deep well 96-well plates. Reactions were set up containing sodium pyruvate (100 μ L, final concentration 80 mM, in 50 mM sodium phosphate buffer, pH 7.4), aldehyde (100 μ L, final concentration 8 mM, in 50 mM sodium phosphate buffer, pH 7.4) and modified enzyme (50 μ L, 1.0 mg mL⁻¹). Reactions were performed for each modified enzyme with each one of the aldehydes shown in Figure 4.7. All screens contained blanks containing no enzyme, and reactions with each of the aldehydes was also performed containing wild-type *sa*NAL, which would provide a comparison to identify modified enzymes with increased activity. Once reactions were set up they were incubated for 16 hours at room temperature. After incubation the TBA assay was performed to assess activity; 11 μ L of sodium periodate (0.2 M in 9 M H₃PO₄) was added to each reaction, the plate was vortexed and then centrifuged for 2 min at 1000 g before being incubated at room temperature for 20 min. Periodate oxidation was terminated by addition of 45 μ L of sodium arsenite (10% w/v in 0.5M Na₂SO₄ and 0.05M H₂SO₄), this caused a brown discolouration to form, which was dissipated by vortexing the plate for 30 s. Once the brown discolouration had dissipated, 135 μ L TBA (0.6 % w/v in 0.5M Na₂SO₄) was added to each well, the plate was centrifuged for 2 min at 1000 g before being incubated in a heat block at 70 °C for 30 min.

Once the assay had been performed, hits were easily identifiable by eye, but for further corroboration absorbance data at 550 nm was also acquired. Samples (85 μ L), of each assayed reaction, were transferred into shallow flat bottomed 96-well plates so that the absorbance could be read in a plate reader. Unfortunately, initial attempts at measuring accurate absorbance data were not successful due to turbidity in the samples. It was found that by centrifuging the deep well plate (10 min 6000 g) after the assay had been performed it was possible to separate a clear layer that could then be analysed by absorbance spectroscopy in shallow 96-well plates. Figure 4.9 shows an example of a screen carried out at position 190, where the wild-type phenylalanine residue has been replaced with multiple different non-canonical amino acids. When analysing the screens, the activity of each modified variant with a specific aldehyde was compared to the wild-type with that same aldehyde. Absorbance

data at 550 nm could also be processed to show whether a modified variant had increased or decreased the activity with a specific aldehyde compared to the wild-type level. The processed data for Figure 4.9 is shown in Figure 4.10 where the non-canonical side chains have been grouped according to their chemistry.

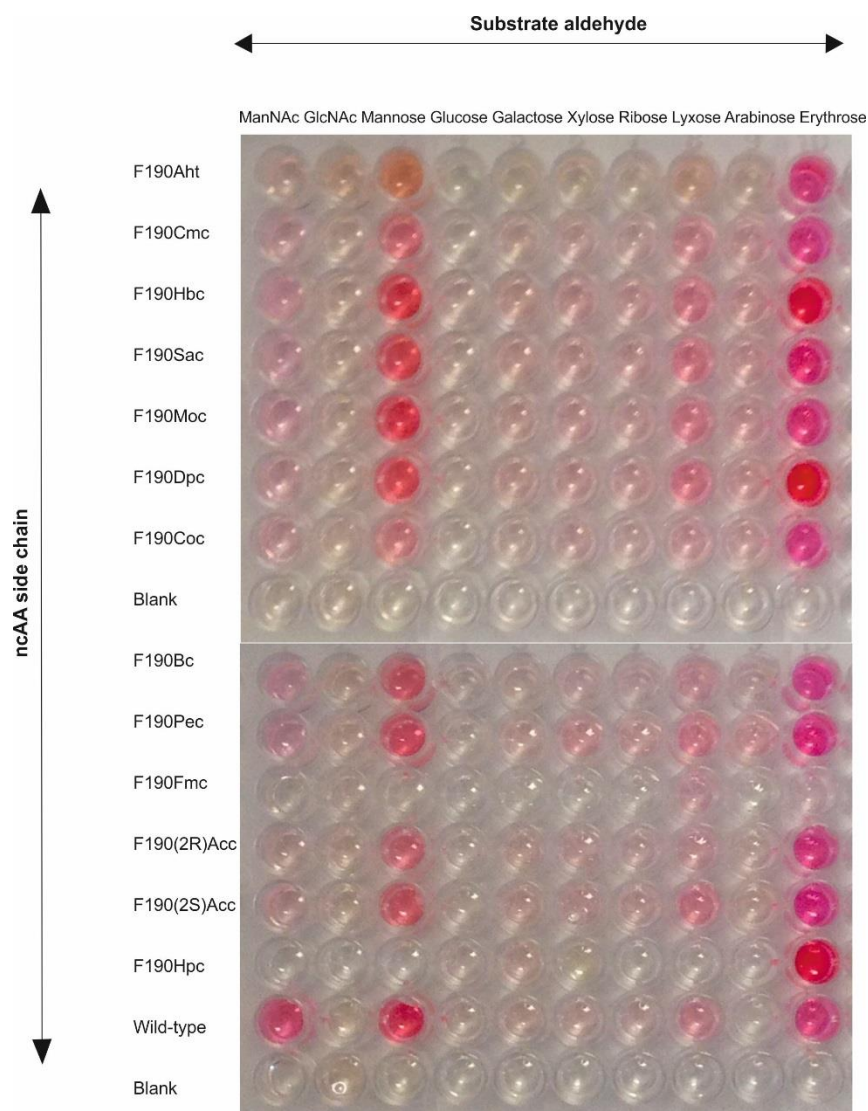


Figure 4.9. Image of screening carried out at position 190 and analysed by TBA assay. The phenylalanine at position 190 was replaced with ncAAs listed on the left-hand side of the image, and then screened for activity with the aldehydes shown along the top of the image. Screening reactions were set up containing sodium pyruvate (100 μ L, final concentration 80mM, in 50 mM sodium phosphate buffer, pH 7.4), aldehyde (100 μ L, final concentration 8mM, in 50 mM sodium phosphate buffer, pH 7.4) and modified enzyme (50 μ L, 1.0 mg mL⁻¹) and incubated for 16 hours at room temperature. Reactions were then assayed using the TBA assay (Section 2.4.16.1).

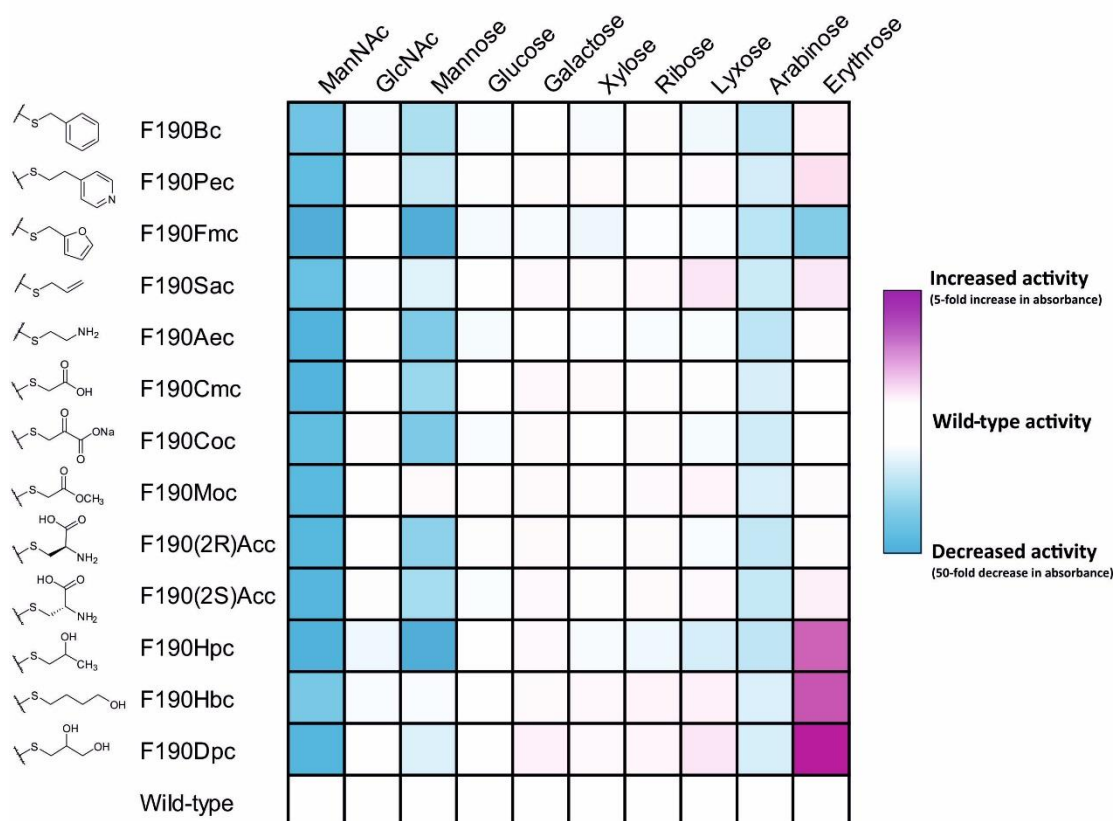


Figure 4.10. Heat map of absorbance data, at 550 nm, from screening carried out by TBA assay at position 190. ncAAs inserted at position 190 are shown down the left-hand side (grouped according to their side chain chemistry), and aldehydes screened with are shown along the top. Samples from reactions were analysed by TBA assay (Section 2.4.16.1) and absorbance data at 550 nm recorded. Absorbance data for each modified enzyme was then compared to the wild-type absorbance data for each aldehyde, and colour coded; blue indicating a decrease in activity and pink indicating an increase in activity compared to the wild-type.

From the screening carried out at position 190 it can be seen that the majority of modified variants either decreased activity with certain aldehydes or retained the same activity as wild-type *s*NAL, which was typical in many of the screens carried out, other examples of screens are shown in Figure 4.11. In the screen showed at position 190 there were three modified variants that appeared to significantly increase the activity with erythrose over that of the wild-type *s*NAL. Therefore any modified variants that showed a distinct increase in activity over the wild-type with a specific aldehyde were classed as ‘hits’. These ‘hits’ were then produced on a larger scale, size exclusion chromatography was performed to remove any incorrectly folded material and then further characterisation of the hits was carried out.

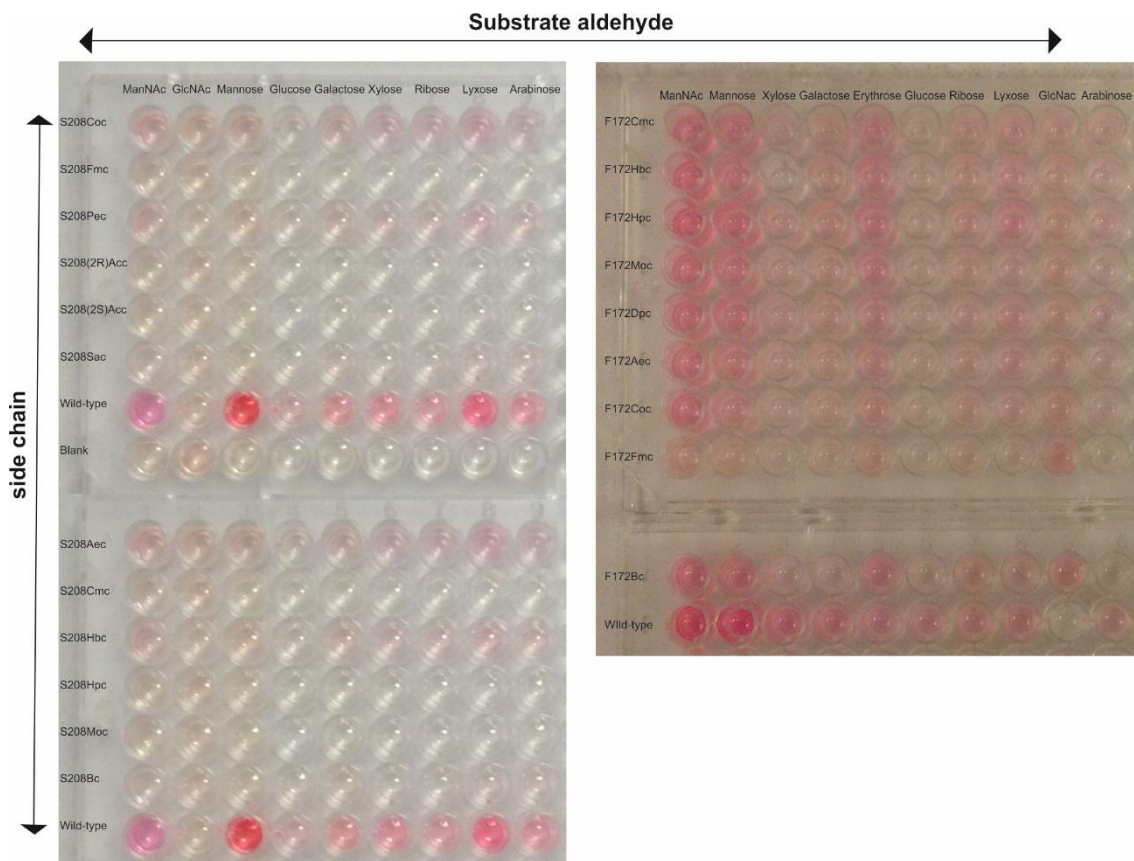


Figure 4.11 Images of screening carried out at positions S208 and F172 which have been analysed by TBA assay. The image on the left shows screening of modified enzymes that were produced by replacing the serine at position 208 with various ncAAs. The image on the right shows screening of modified enzymes that were produced by replacing the phenylalanine at position 172 with various ncAAs. The ncAAs which have been inserted at each position are listed down the left side of the plate images. The modified enzymes were screened for activity with the aldehydes shown along the top of the plate images. Screening reactions were set up containing sodium pyruvate (100 μ L, final concentration 80mM, in 50 mM sodium phosphate buffer, pH 7.4), aldehyde (100 μ L, final concentration 8mM, in 50 mM sodium phosphate buffer, pH 7.4) and modified enzyme (50 μ L, 1.0 mg mL⁻¹) and incubated for 16 hours at room temperature. Reactions were then assayed using the TBA assay (Section 2.4.16.1).

4.5 Investigation of hits from screening

During the screening a number of hits were found that increased the activity with a specific aldehyde over that of wild-type *sa*NAL, including the ‘hits’ found in the position 190 screening (Figure 4.9 and Figure 4.10). However some of these ‘hits’ only produced minor increases in activity and so were not further characterised.

4.5.1 F172Bc activity with GlcNAc and pyruvate

One 'hit' that was found in the screening at position 172, produced an increase in activity with the aldehyde *N*-acetyl-D-glucosamine (GlcNAc) by replacing the wild-type phenylalanine with a benzylcysteine side chain (F172Bc) (Figure 4.12). Although the activity seen with F172Bc and GlcNAc was modest, this activity was very interesting as it has been reported in the literature and from the results shown that wild-type NAL does not utilise GlcNAc as a substrate (Hsu *et al.*, 2005, Gijsen *et al.*, 1996).

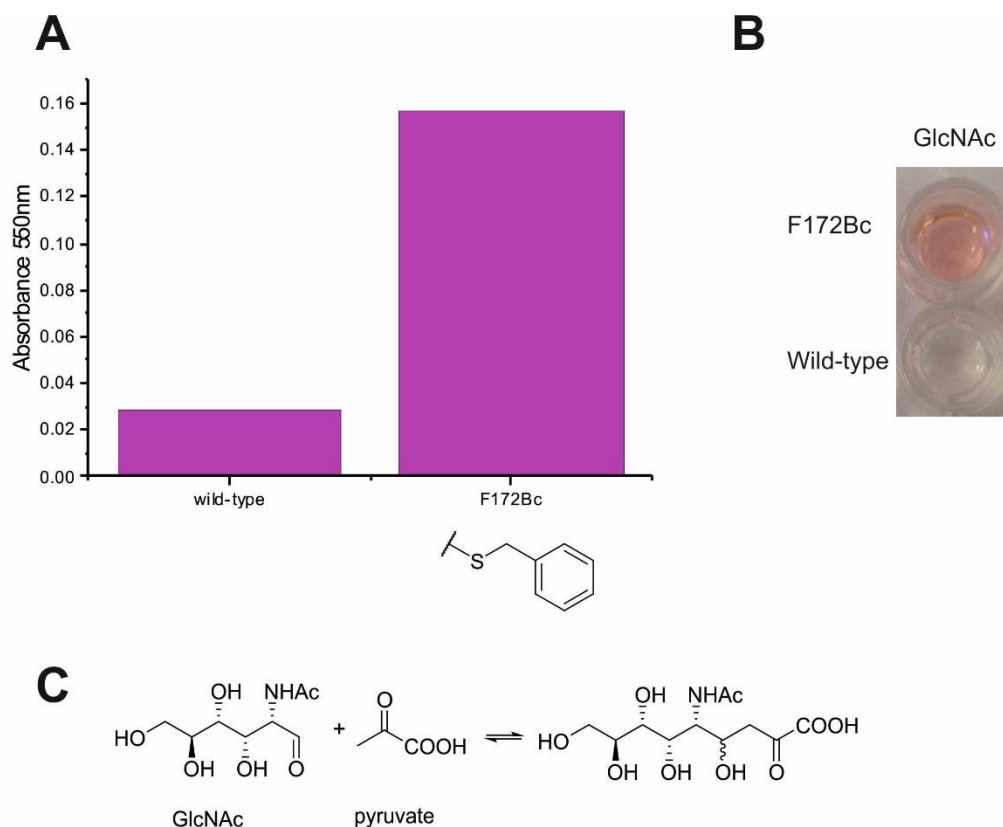


Figure 4.12. Modified enzyme F172Bc showing an increased activity with GlcNAc and pyruvate, over that of the wild-type. Panel A shows the absorbance data at 550 nm, from the TBA assay, for F172Bc and wild-type enzymes with GlcNAc and pyruvate. Panel B shows an image of TBA assayed samples from reactions of the F172Bc and wild-type enzymes, with GlcNAc and pyruvate. Panel C shows the reaction between GlcNAc and pyruvate to form the full length aldol product.

From the TBA assay there is a clear increase in the pink colour when F172Bc has been incubated with GlcNAc and pyruvate compared to wild-type *sa*NAL, under the same conditions, indicating an increase in activity. Once this hit had been identified, it was necessary to further confirm the activity with GlcNAc using mass spectrometry. ESI mass spectrometry was carried out on a sample from the reaction of F172BC with GlcNAc and pyruvate (Figure

4.13). The reaction was carried out using the same conditions as were used in the TBA assay and incubated for 16 hours at room temperature. Then a sample from the reaction was taken and ESI negative ion mode mass spectrometry was performed (Section 2.4.19). From the mass spectrum there was a clear peak at 308 Da which correlates to the molecular mass of the full length product, 309 Da, minus 1 hydrogen due to the negative ion mode.

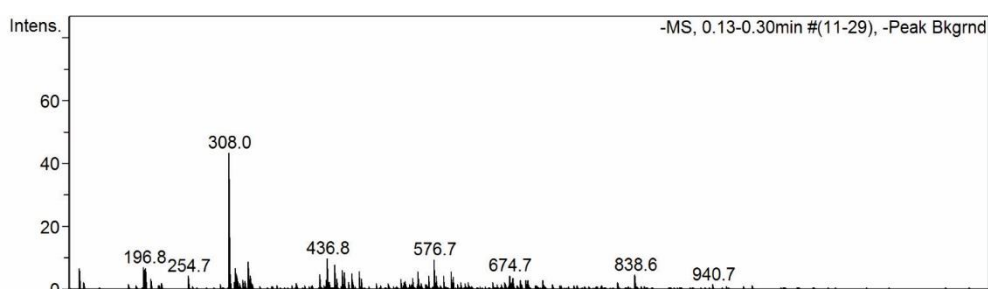


Figure 4.13. ESI negative ion mode mass spectrum of the product formed from reaction of GlcNAc and pyruvate with F172Bc. A peak at 308 Da can be seen; indicating the full length product has been formed, which has a molecular mass of 309 Da minus a hydrogen due to the negative ion mode.

It was possible to confirm the F172Bc activity with GlcNAc using both the TBA assay and ESI-negative ion mode mass spectrometry, however when proton NMR was used to characterise the product there was no evidence for the full length product in the spectrum. NMR analysis was carried out on larger scale reactions, 80 mg of GlcNAc and a 10-fold molar excess of pyruvate, in sodium phosphate buffer (50 mM, pH 7.4), with 3.0 mg of F172Bc enzyme, the reaction was incubated for 24 hours but there was still no evidence for the full length product. Therefore it was hypothesised that the reason that the full length product was not seen in NMR was due to the low yield of the reaction. It was possible to quantify the amount of product produced from the F172Bc reaction, using a sialic acid standard curve. Quantification revealed only 1.85 mg of the full length product was produced after a 24 hour incubation, and so further characterisation was not carried out.

4.5.2 F190Dpc activity with erythrose and pyruvate

Screening at position 190 revealed that there were three ncAA-containing enzyme variants which had increased the activity with the aldehyde erythrose and pyruvate to produce the full length product 3-deoxy-2-heptulosonic acid (DOH). Analysis by TBA assay showed by replacing the phenylalanine at position 190 with 2-hydroxypropyl cysteine, 4-hydroxybutyl cysteine or

2, 3-dihydroxypropyl cysteine, the activity with erythrose and pyruvate was increased over that of the wild-type with erythrose and pyruvate (Figure 4.14). By comparing the TBA assay data F190Dpc appeared to produce the greatest increase in activity with erythrose and so this enzyme was taken forwards to characterise.

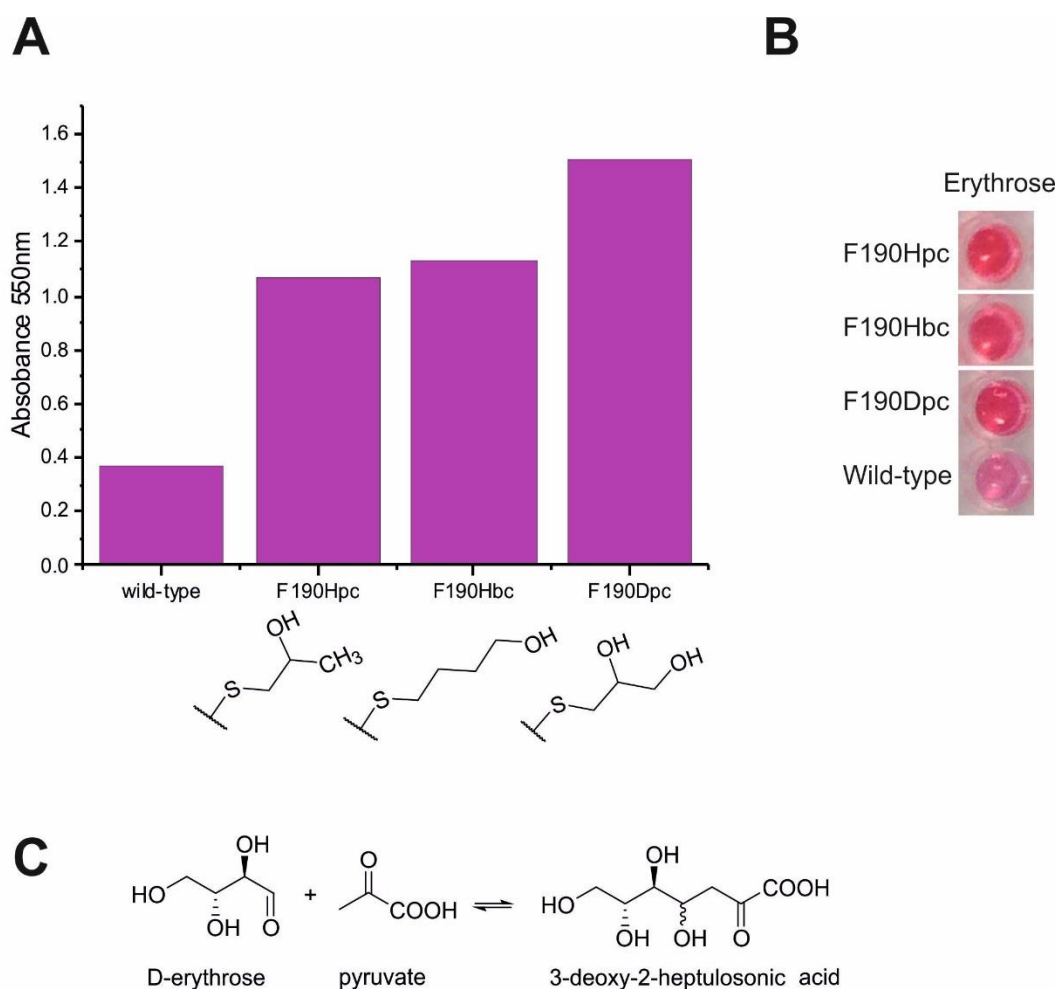


Figure 4.14. Modified enzymes F190Hpc, F190Hbc and F190Dpc showing increased activity with erythrose and pyruvate, over that of the wild-type. Panel A shows the absorbance data at 550 nm, from the TBA assay, for the F190Hpc, F190Hbc, F190Dpc and wild-type enzymes with erythrose and pyruvate. Panel B shows an image of TBA assayed samples from reactions of the F190Hpc, F190Hbc, F190Dpc and wild-type enzymes, with erythrose and pyruvate. Panel C shows the reaction between erythrose and pyruvate to form 3-deoxy-2-heptulosonic acid (DOH).

The F190Dpc enzyme was produced on a large scale for further characterisation to be carried out with it. Large scale modification was performed on a 20 mg scale in the same manner as with the small scale modification. 20 mg of lyophilised K165C protein was resuspended in

10 mL of sodium phosphate buffer (50 mM pH 8.0) containing 6 M urea which had been pre-warmed to 37 °C. The protein solution was vortexed for 30 s to ensure the lyophilised protein was thoroughly redissolved, then 800 μL of diBr (0.08 mg μL^{-1} in diBr), was added and the solution incubated for 1 hr 30 min at 37 °C, 200 rpm. Once conversion to F190dha was confirmed using ESI mass spectrometry, then 320 μL of thioglycerol (160 μL thiol, 160 μL 2 M Tris/HCl pH 8.0) was added and incubated at 37 °C, 200 rpm for 2 hours. Correctly modified protein (shown in Figure 4.15) was then refolded by dialysis and underwent size exclusion chromatography, to remove any incorrectly folded material.

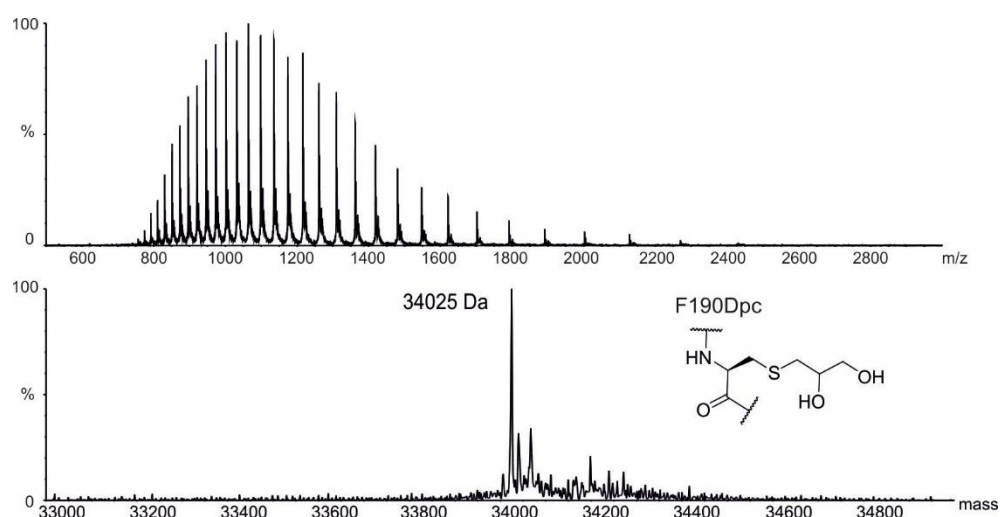


Figure 4.15. ESI mass spectrum showing correctly modified F190Dpc. The expected mass of F190Dpc is 34025 Da and the observed mass was 34025 Da.

To further confirm that the product being produced in the reaction of erythrose and pyruvate with F190Dpc was in fact DOH, ESI-negative ion mode mass spectrometry performed. A reaction was set up containing erythrose, pyruvate and F190Dpc; using the same conditions as in the TBA assay and then incubated for 16 hours at room temperature, before being analysed by mass spectrometry. The mass spectrum shows a clear peak at 207 Da which correlates to the molecular mass of the full length product, 208 Da, minus 1 hydrogen due to the negative ion mode (Figure 4.16). Both mass spectrometry and the TBA assay indicated that the reaction of the modified enzyme, F190Dpc, with erythrose and pyruvate was producing the expected product DOH. Therefore, using a sialic acid standard curve, the amount of DOH produced was quantified. A large scale reaction was set up with 60 mg of erythrose, a 10-fold molar excess of

pyruvate, in sodium phosphate buffer (50 mM, pH 7.4), and incubated with 0.5 mg of F190Dpc enzyme for 24 hours. 72 mg of DOH was produced in 24 hours showing an approximate 70 % yield.

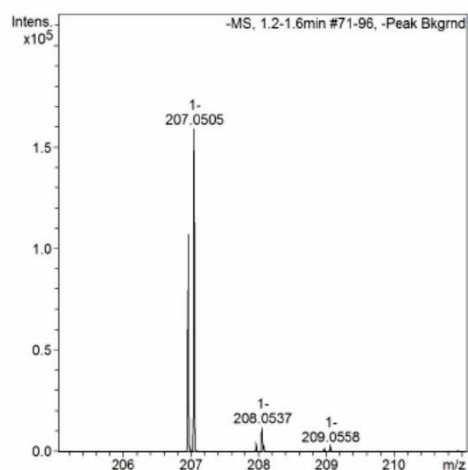


Figure 4.16. ESI negative ion mode mass spectrum of the product formed from reaction of erythrose and pyruvate with F190Dpc. A peak at 207 Da can be seen; indicating the product 3-deoxy 2-heptulosonic acid (DOH) has been formed, which has a molecular mass of 208 Da minus a hydrogen due to the negative ion mode.

To characterise the product from the F190Dpc reaction with erythrose and pyruvate proton NMR was used. Comparison of the proton NMR spectrum of the reaction mixture, to a standard of erythrose and pyruvate showed differences in chemical shifts and splitting patterns of some of the peaks. These differences indicated that a reaction had occurred and that an aldol condensation product was present. However it was not possible to characterise the product due to high levels of impurities, in the spectra, that were present from the erythrose starting material. To allow analysis by NMR the reaction mixture had to be purified. A large scale reaction was set up, erythrose (500 mg, 4.2 mmol) and sodium pyruvate (2.29 g, 21 mmol) in sodium phosphate buffer (50 mM, pH 7.4) were incubated with F190Dpc (250 μ L, 3.2 mg mL⁻¹) for 24 hours before the product was purified by anion exchange chromatography. Anion exchange chromatography was carried out on AG1x8 resin (HCO₃⁻, 100-200 mesh) and product was eluted using a 0-0.4 M ammonium bicarbonate linear gradient. Due to the product, DOH, having no distinctive wavelength absorbance, the fractions were analysed by TBA assay to find where the product had eluted. The product was seen to elute, by the TBA

assay, between 5 and 20 % of 0.4 M ammonium bicarbonate, and so these were analysed by NMR.

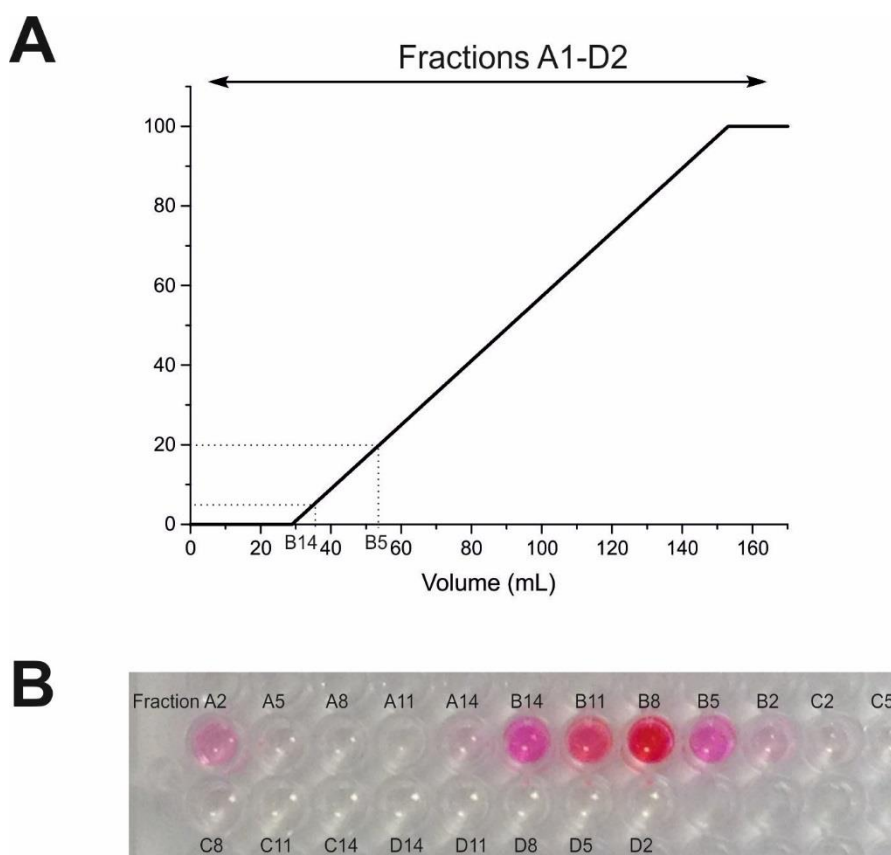


Figure 4.17. Anion exchange of 3-deoxy 2-heptulosonic acid (DOH). Panel A shows the linear gradient of $(\text{NH}_4)\text{HCO}_3$ used to elute DOH on a AG1x8 resin (HCO_3^- , 100-200 mesh) column. Panel B shows fractions collected from the anion exchange and analysed by TBA assay. DOH is seen to elute in fractions B14-B5 which corresponds to between 5 and 20 % $(\text{NH}_4)\text{HCO}_3$.

4.5.3 NMR analysis of DOH

Once the DOH product had been purified by anion exchange, 500 MHz ^1H NMR spectroscopy, in D_2O , was carried out on the purified fractions (Section 2.4.18). From the proton NMR diagnostic aldol condensation peaks could be seen in the region 1.95-2.65 ppm. When an aldol condensation occurs, the protons which are originally from the methyl group of pyruvate will undergo changes in chemical shift and splitting pattern, as they form a diastereotopic pair of protons at the new carbon-carbon centre, between C3 and C4. This produces the peaks at 1.95-2.65 ppm and indicates that an aldol condensation has occurred.

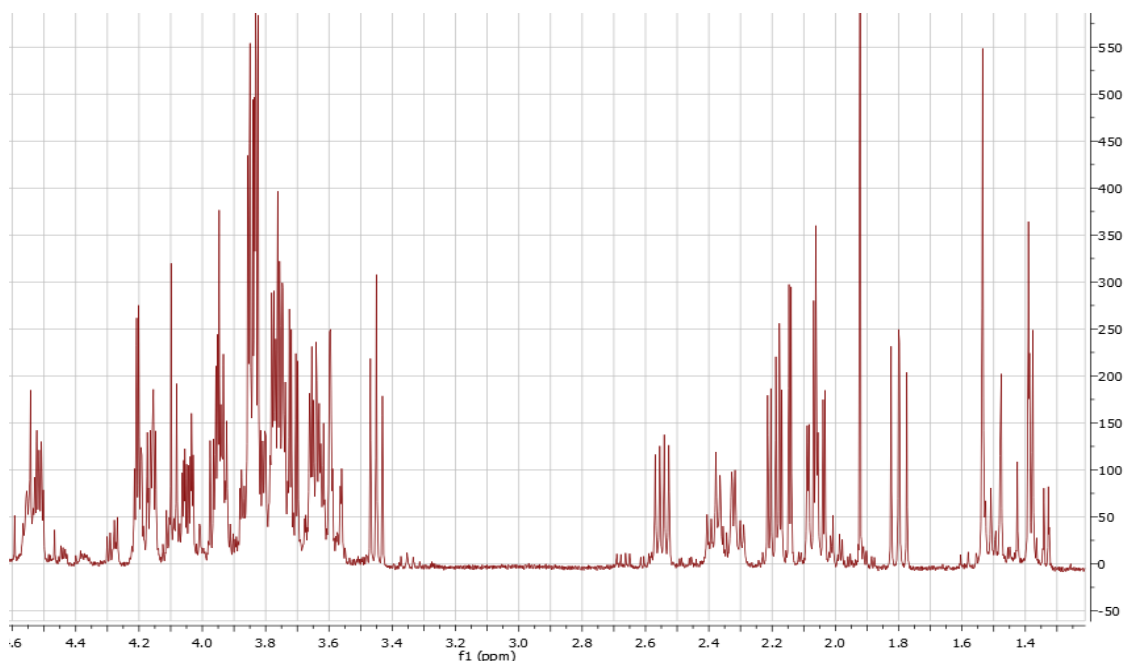


Figure 4.18. Proton NMR spectrum of DOH purified by anion exchange.

The peaks at 1.95-2.65 ppm in the proton NMR indicate that an aldol condensation between erythrose and pyruvate has occurred to form DOH. DOH is a seven carbon sugar and so it may cyclise to produce multiple different forms (shown in Figure 4.19). DOH has the potential to form both pyranose and furanose rings, each of which can adopt either the α or β anomer of the ring structure, also the stereochemistry at the new carbon-carbon centre of DOH could be in either the *R* or *S* configuration. Therefore it is unsurprising that the proton NMR spectrum looks highly complicated (Figure 4.18), peaks in the region 1.95-2.65 ppm indicated that four different forms of the product were present in the sample. Since four different forms of the product were found in the proton NMR it was necessary to use 2D NMR techniques to assign the different forms seen.

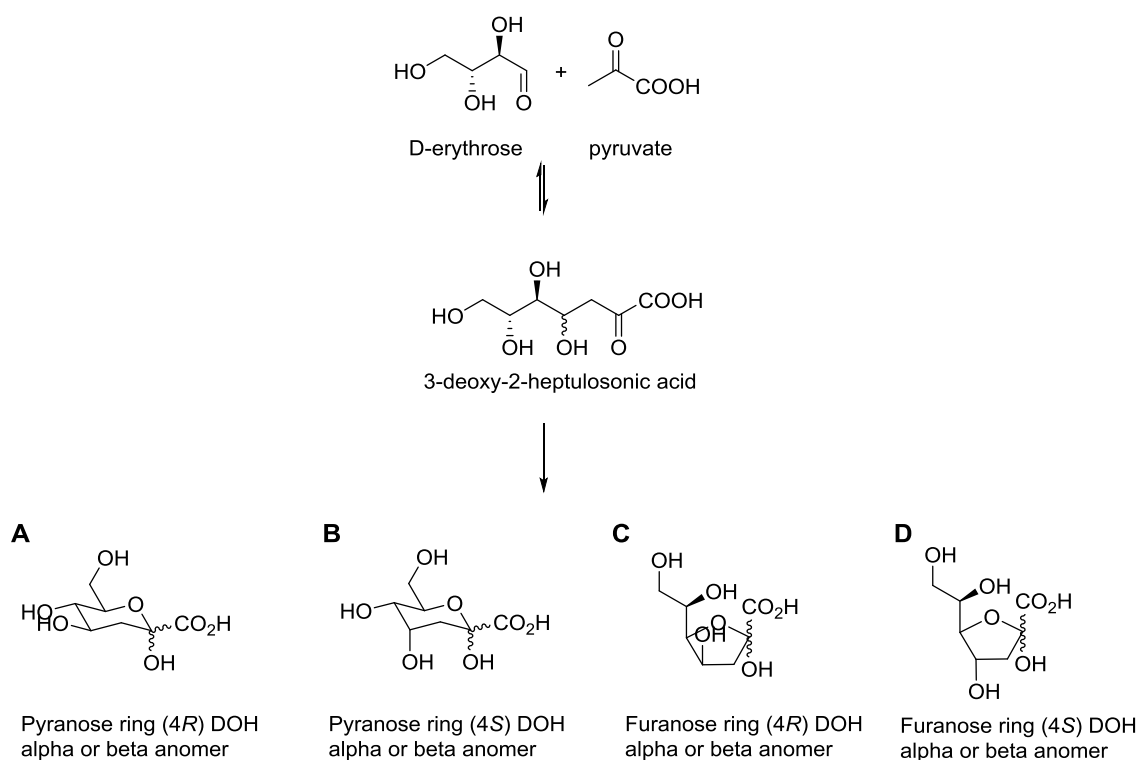


Figure 4.19 Aldol condensation of pyruvate and D-erythrose to form deoxyheptulosonic acid (DOH). DOH can then cyclise in multiple different ways, shown in structures A-H. A shows the α or β (4*R*)DOH pyranose, B shows the α or β (4*S*)DOH pyranose C shows the α or β (4*R*)DOH furanose, D shows the α or β (4*S*)DOH furanose.

To allow characterisation of the different forms of the product, both COSY (correlation spectroscopy) and TOCSY (total correlation spectroscopy) 2D NMR techniques were used. COSY allows the protons that are coupling to each other to be easily identifiable, and TOCSY allows protons, which may not be directly coupled to each other but are in the same spin system to be identified. In the 2D NMR spectra shown in Figure 4.20 the spin systems (shown in the TOCSY spectrum) and the coupled protons (in the COSY spectrum) can be seen for the four distinct forms seen of the product. NMR spectroscopic data is summarised in Table 4.5 however even using the 2D NMR techniques, it was not possible to assign coupling constants and chemical shifts for some of the protons in the different DOH forms due to overlapping peaks in the spectra.

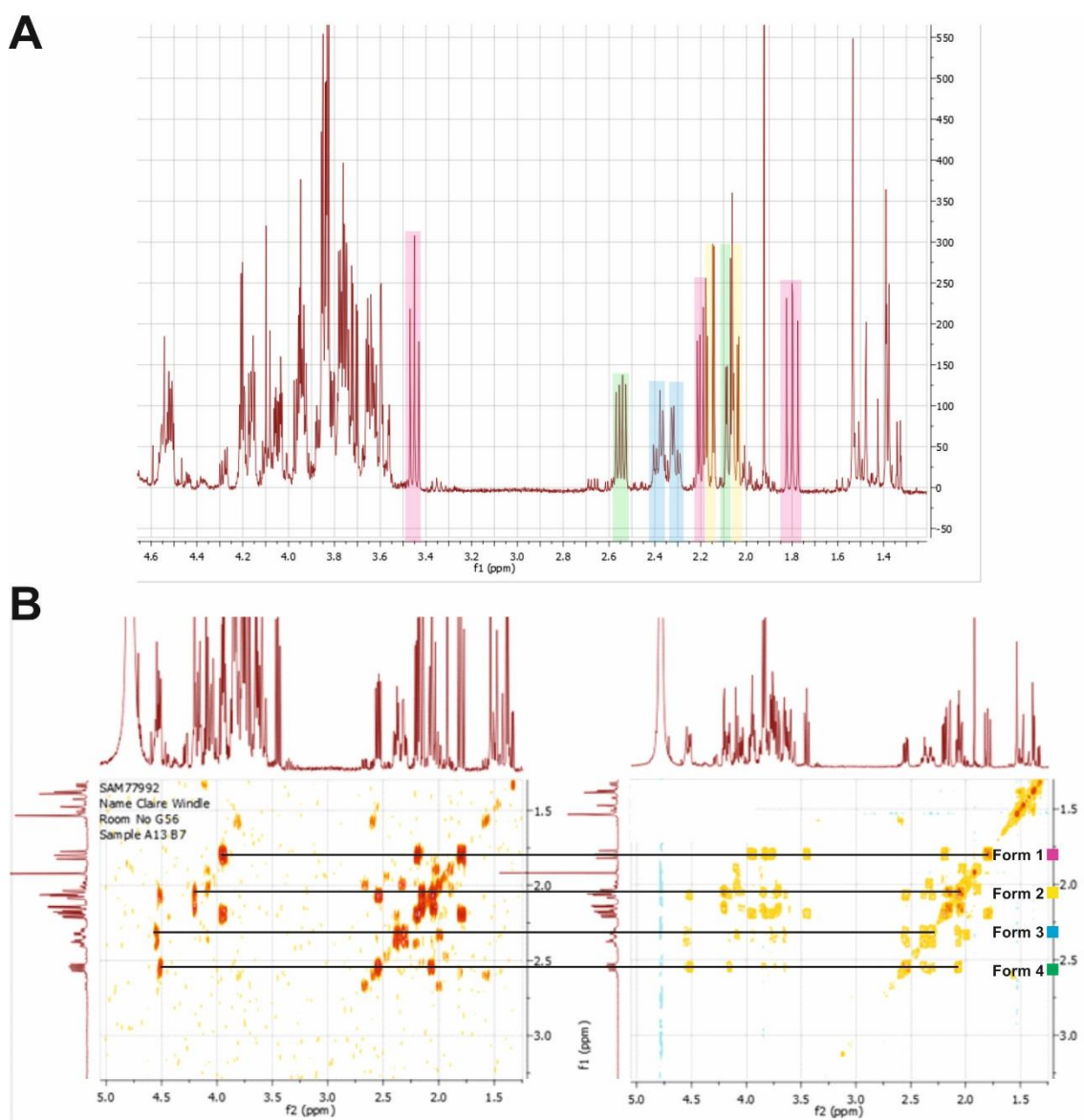


Figure 4.20. Proton COSY and TOCSY NMR spectra of the product DOH. Panel A shows the proton NMR spectrum of DOH with easily isolatable peaks corresponding to form 1 highlighted in pink, form 2 highlighted in yellow, form 3 highlighted in blue and form 4 highlighted in green. Panel B shows COSY spectrum on the left-hand side and TOCSY spectrum is shown on the right-hand side. The spin systems for each of the different forms of DOH have been highlighted by black lines.

Previously published data for the compound methyl 3-deoxy-L-arabino-2-heptulosonate (Fitz *et al.*, 1995) could be used as a comparison for the NMR data of DOH as it is the enantiomer of (4*R*) DOH, and the methyl ester group would not affect the NMR spectrum of the rest of the sugar. The NMR spectroscopic data for form 1 of the product corresponded to the previously published data, indicating that the product produced was DOH as expected (shown in Table 4.5 and Table 4.6).

Previously published NMR data of the (4*R*,5*R*,6*R*)- and (4*S*,5*R*,6*R*)-6-dipropylcarbamoyl-2-oxo-4,5,6-trihydroxy-hexanoic acid ((4*R*) and (4*S*) dipropyl sialic acid) also provides a comparison for the DOH forms found, since the ring structure is comparable to DOH (Woodhall *et al.*, 2005). All forms of DOH seen in the NMR correspond to NMR data seen for the different pyranose and furanose forms of (4*R*) and (4*S*) dipropyl sialic acid. NMR data for the forms of DOH seen in the NMR spectra is summarised in Table 4.5, and previously published data of structures similar to DOH that have been used for comparison is summarised in Table 4.6.

NMR spectroscopic data for form 1 of DOH correlates well with the pyranose (4*R*) dipropyl sialic acid indicating that this form of DOH may have adopted a pyranose ring (Table 4.5 and Table 4.6). Also the coupling constants between the protons on C3 and C4 may allow the relative configuration at the new carbon-carbon centre to be determined. From the data for form 1 of DOH the large 11.6 coupling constant is consistent with diaxial coupling in the pyranose ring, and the smaller 5.1 coupling constant is consistent with axial-equatorial coupling, and so the hydroxyl group at the C4 position should be equatorial (Figure 4.21). Therefore the stereochemistry at C4 is in the *R* configuration, which corresponds with the data for the pyranose (4*R*) dipropyl sialic acid.

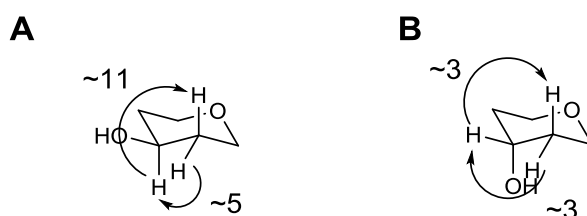


Figure 4.21 Approximate coupling constants in pyranose rings. Pyranose rings have been shown with groups only at two carbons for simplicity. A shows diaxial coupling which produces a large coupling constant ~ 11 , and equatorial axial coupling which produces a smaller coupling constant ~ 5 . B shows diequatorial and equatorial axial coupling which both produce small coupling constants ~ 3 . Coupling constants shown have been taken from the experimental data for form 1 and 2 of DOH.

Form 2 of DOH correlates well with the data for a pyranose (4*S*) dipropyl sialic acid, this may indicate that form 2 has also adopted a pyranose form (Table 4.5 and Table 4.6). The coupling constants for H3-H4 are both quite small 3.6 and 3.1 therefore there is no diaxial coupling between the protons at C3 and C4 in form 2 (Figure 4.21). No diaxial coupling between the

protons would indicate that the hydroxyl group at C4 is in the axial position producing the 4S configuration at this position.

Data for form 3 of DOH correlates with a furanose (4S) dipropyl sialic acid (Table 4.5 and Table 4.6), and the presence of coupling constant of 7 also indicates that form 3 has adopted a five membered furanose ring as this coupling constant cannot be present in a pyranose ring. Similarities in the coupling constants between form 3 of DOH and the furanose (4S) dipropyl sialic acid show that form 3 may have the (4S) configuration. NMR spectroscopic data for form 4 correlates with the data for the furanose (4S) dipropyl sialic acid which may indicate that form four has cyclised to a furanose ring. The coupling constants for form 4 also correlate very well with the coupling constants of a furanose (4S) dipropyl sialic acid (Table 4.5 and Table 4.6), and so form 4 may have the same (4S) stereochemistry. It is quite likely that both form 3 and form 4 are both furanose rings with (4S) configuration which may indicate that they are interconverting α and β anomers.

To calculate the amount of each form of DOH present an NMR spectrum with a known amount of methanol was collected. It was seen that form 1, which is most likely the only (4R) configuration present, was present at almost 100 %. Form 2, 3 and 4, which all have the (4S) configuration, were present at approximately 30 % each. This is consistent with the data from Woodhall *et al* where the (4R) dipropyl sialic acid was present in one pyranose form at 80 % and the (4S) dipropyl sialic acid was present in two pyranose and two furanose forms in the ratio 15:10:45:30.

From the NMR data and assignments, DOH is produced from the reaction of D-erythrose and pyruvate with F190Dpc *sa*NAL. It is highly likely, from the NMR spectra, that the DOH product is produced with both (4R) and (4S) stereochemistry therefore the aldol condensation of D-erythrose and pyruvate by F190Dpc is not stereospecific. Also from the NMR analysis it is likely that the DOH produced forms both pyranose and furanose rings.

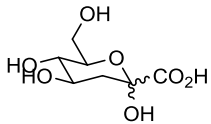
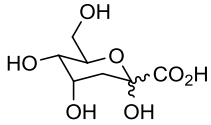
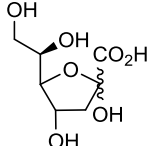
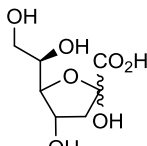
DOH form	% of form seen	δ /ppm H-3	δ /ppm H-4	δ /ppm H-5	δ /ppm H-6	δ /ppm H-7	J/Hz H3 ^A - H3 ^B	J/Hz H3-H4
<u>Form 1</u> 	<95	2.19, 1.8	3.94*	3.45	3.76*	3.64*	13.0	5.1, 11.6
<u>Form 2</u> 	30	2.16, 2.05	4.25*	3.75	4.05*	nd	14.8	3.6, 3.1
<u>Form 3</u> 	30	2.38, 2.32	4.55*	nd	nd	nd	13.6	7.0, 5.3
<u>Form 4</u> 	30	2.55, 2.08	4.53*	4.15*	3.83*	nd	14.4	7.2, 2.8

Table 4.5 NMR spectroscopic data for the four different forms of DOH observed in the NMR spectra. The colours underlining the different form of DOH relate back to Figure 1.16. * indicates a multiplet that has a chemical shift centred on this chemical shift.

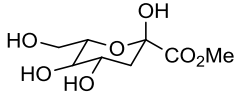
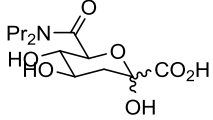
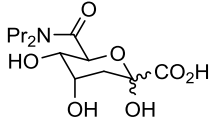
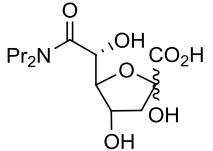
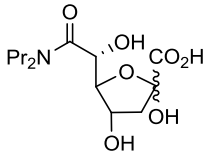
Compound	% of form seen	δ /ppm H-3	δ /ppm H-4	δ /ppm H-5	δ /ppm H-6	δ /ppm H-7	J /Hz H3 ^A -H3 ^B	J /Hz H3-H4
methyl 3-deoxy-L-arabino -2-heptulosonate 	<95	2.24, 1.82	3.93	3.39	3.82	3.77	13.2	5.1, 11.7
(4 <i>R</i>) dipropyl sialic acid Pyranose1 	80	2.23, 1.84	3.95	3.64	4.61	-	13.3	5.1, 11.5
(4 <i>S</i>) dipropyl sialic acid Pyranose1 	15	2.15, 2.11	4.2	3.86	4.95	-	15.0	3.4, 3.4
(4 <i>S</i>) dipropyl sialic acid Furanose2 	30	2.42, 2.39	4.57	4.10	4.54	-	15	5.8, 6.4
(4 <i>S</i>) dipropyl sialic acid Furanose1 	45	2.64, 2.09	4.52	4.25	4.46	-	14.5	6.9, 2.6

Table 4.6 Previously published NMR spectroscopic data of similar structures to DOH (Fitz *et al.*, 1995 Woodhall *et al.*, 2005).

4.6 Summary

This chapter has described the insertion of many ncAAs, individually throughout the active site of *sa*NAL, and the screening of these modified enzymes for altered activity. 13 different ncAAs have been individually incorporated into *sa*NAL, by chemical modification, at a variety of different positions. These modified enzymes were then screened for activity with 10 different aldehydes and pyruvate. Excitingly, using this approach it has been possible to show that ncAAs are well tolerated within the enzyme, with modified enzymes not only retaining activity but also showing increased activities.

Through this methodology it has been possible to bring about an increase in activity with GlcNAc, an aldehyde which is not accepted by the wild-type enzyme, by incorporation of a benzylcysteine at position 172. It has also been possible to increase the activity with erythrose by incorporation of a 2, 3-dihydroxy-propyl cysteine side chain at position 190. This activity produces the full length aldol product 3-deoxy-2-heptulosonic acid, which is a related compound to 3-deoxy-D-arabino-2-heptulosonate-7-phosphate (DAH7P). DAH7P is an important intermediate in the shikimate pathway, found in bacteria and plants, for the production of essential aromatic amino acids (Dondoni *et al.*, 1994, Maeda and Dudareva, 2012). Due to the similarity of DOH to DAH7P, DOH could prove useful as a starter unit for compounds with potential herbicidal and antimicrobial effects, there is also interest in DOH and other 3-deoxy-2-ulosonic acids due to their potential use in complex oligosaccharide syntheses (Hekking *et al.*, 2006, Pradhan *et al.*, 2013, Enders and Gasperi, 2007).

From this chapter it is clear that enzyme activity can be altered by using ncAAs. By incorporation of ncAAs into the active site of *sa*NAL, it has been possible to alter the activity towards various aldehyde substrates, some of which produce potentially synthetically useful compounds.

Chapter 5 Characterisation of F190Dpc *sa*NAL

The incorporation of a variety of ncAAs individually, throughout *sa*NAL, and the subsequent screening of these modified enzymes for altered substrate specificity, was described in the previous chapter. Identification and preliminary investigation of the enzyme F190Dpc, which showed an increase in activity with erythrose and pyruvate as substrates (Figure 5.1), was also described. This chapter will detail further characterisation of this activity, using both kinetic and X-ray crystallographic methods.

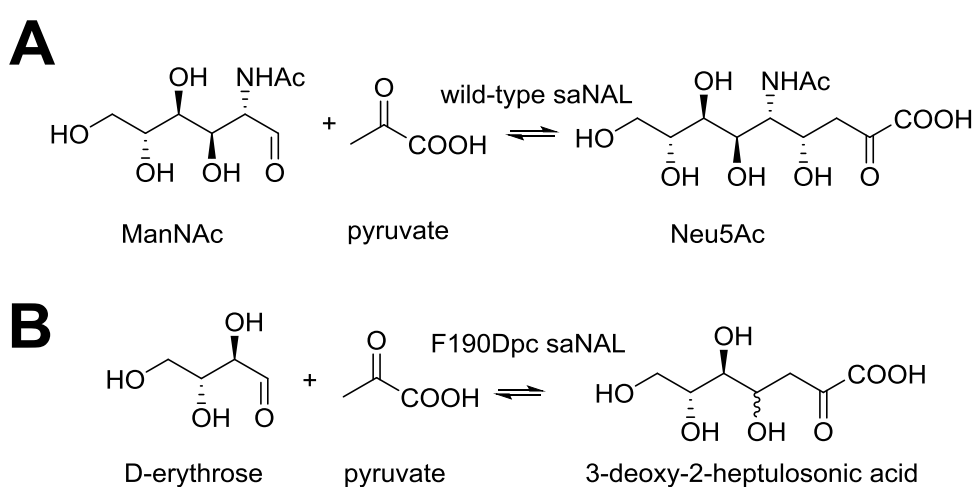


Figure 5.1 Aldol condensation reactions of wild-type and F190Dpc *sa*NAL. A) aldol condensation of *N*-acetyl-D-mannosamine (ManNAc) with pyruvate to form *N*-acetylneuraminic acid (Neu5Ac), catalysed by wild-type *sa*NAL. B) aldol condensation of erythrose and pyruvate to form 3-deoxy-2-heptulosonic acid, catalysed by F190Dpc *sa*NAL.

5.1 Kinetic characterisation of F190Dpc *sa*NAL

Screening experiments, shown in the previous chapter, identified that by replacing the phenylalanine side chain at position 190 with the ncAA 2,3-dihydroxypropyl cysteine, an increase in activity with the aldehyde erythrose was produced. To allow quantification of this increase in activity, kinetic parameters were determined for both the wild-type and F190Dpc enzymes. Kinetic parameters were determined for both enzymes with the substrates erythrose and pyruvate and, for comparison, with the wild-type substrates ManNAc and pyruvate.

5.1.1 Aldol condensation kinetics

Kinetic parameters were determined for the aldol condensation reaction. To follow this reaction the TBA assay was used, which can detect the amount of aldol product formed. However the TBA assay is an end point assay and so the kinetic parameters had to be determined in a discontinuous manner.

To determine kinetic parameters using a discontinuous assay, the initial rates at a variety of substrate concentrations must be determined. For each initial rate determination, samples at various time points must be taken and analysed. This approach would mean, using a minimum of 6 substrate concentrations and 6 time points, 36 assays must be performed. Given the fact that the TBA assay is fairly complex to perform, and the number of kinetic sets that were to be performed, a more streamlined approach was taken. The linear range for the reaction was determined and one time point selected. This time point could then be used as the initial rate, at each substrate concentration, as all absorbance values for time points below this would be linear.

The linear range was found by taking samples from reactions containing 8 mM erythrose and 80 mM pyruvate, with either 0.14 mg mL⁻¹ F190Dpc or wild-type *saNAL*, over a course of 4 hours. For each time point, the reaction was stopped using trichloroacetic acid (TCA) (12% w/v), each sample was then assayed using the TBA assay, and absorbance data at 550 nm was collected. The linear range was then determined by plotting the absorbance data against time (Figure 5.2). The linear range for the reaction between erythrose and pyruvate with the F190Dpc *saNAL* was found to be between 0 and 2 hours at room temperature, and so absorbance values below approximately 0.5 absorbance units were within the linear region. For the wild-type enzyme reaction with erythrose and pyruvate, the linear range was between 0 and 4 hours with absorbance values below 0.1 being in the linear range. Therefore all kinetic assays were carried out with an incubation time of 2 hours at room temperature.

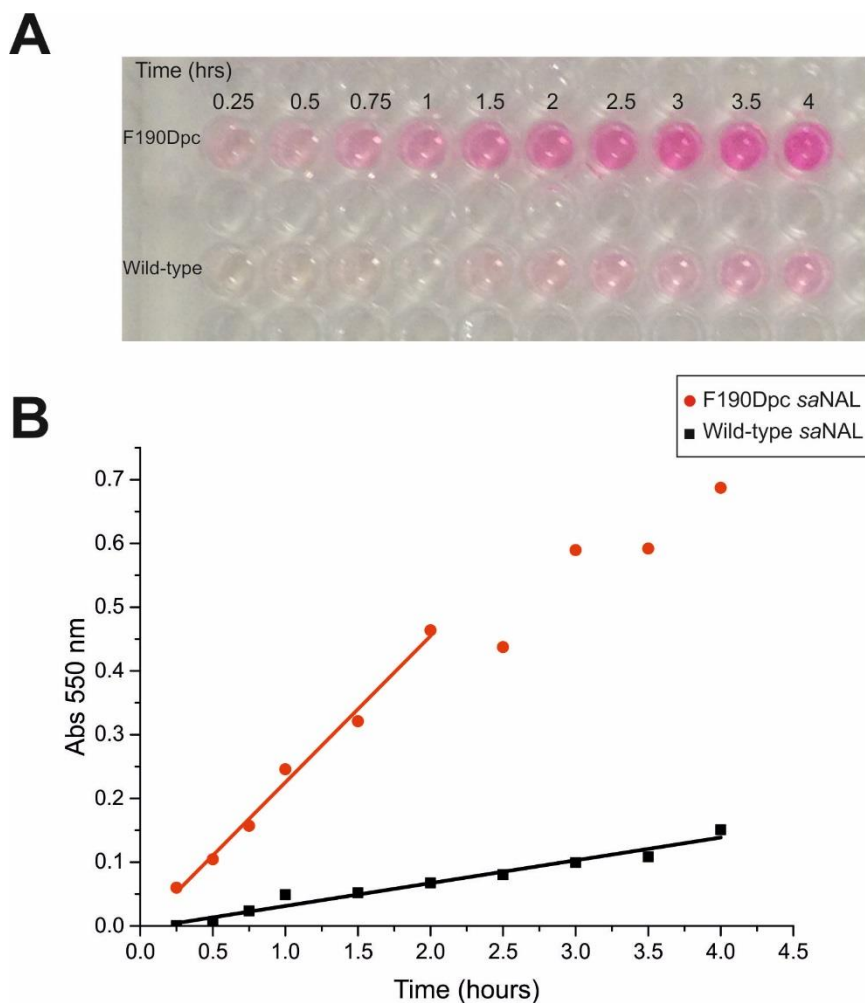


Figure 5.2 Time course of the aldol condensation between erythrose and pyruvate catalysed by F190Dpc and wild-type *saNAL*. Panel A shows an image of the TBA assayed samples taken from reactions containing 8 mM erythrose and 80 mM pyruvate, with either 0.14 mg mL⁻¹ F190Dpc or wild-type *saNAL*. Samples were taken at time points between 0.25 and 4 hours and stopped using trichloroacetic acid, before being analysed by the TBA assay. Panel B shows the plots of the absorbance data at 550 nm of the TBA assayed samples against time, the plot for wild-type *saNAL* is shown in black and the plot for F190Dpc *saNAL* is shown in red.

±The reaction for erythrose and pyruvate is a two-substrate reaction and therefore a 25-point matrix varying the concentrations of both substrates should be carried out and then fitted to Equation 1 (Figure 5.3). However, as the TBA assay is fairly complex to perform and the number of assays that would be needed, it was decided that these kinetics would be carried out with one, saturating concentration of pyruvate. Therefore the concentration of pyruvate was kept constant and saturating at 80 mM (Appendix Figure 7.1), and the aldehyde concentration varied. Using these conditions to carry out the kinetics allows Equation 1 (Figure 5.3) to be rearranged to the same form as the Michaelis-Menten equation (rearrangement shown in

Figure 5.3), and so the kinetic data could be fitted using non-linear regression to Equation 4 (Figure 5.3), to give k_{cat}^{app} , K_m^{app} and k_{cat}^{app}/K_m^{app} for the aldehyde substrate. Therefore this approach was used for all aldol condensation kinetics carried out.

Equation 1.

$$v = \frac{V_{max}[A][B]}{K_{iA}K_B + K_B[A] + K_A[B] + [A][B]}$$

Equation 2.

$$v = \frac{V_{max}[B]}{\frac{K_{iA}K_B}{[A]} + K_B + \frac{K_A[B]}{[A]} + [B]}$$

Equation 3.

$$v = \frac{V_{max}[B]}{\left(\frac{K_{iA}K_B}{[A]} + K_B\right) + \left(\frac{K_A}{[A]} + 1\right)[B]}$$

Equation 4.

$$v = \frac{V_{max}[S]}{K_m + [S]}$$

Figure 5.3. Rate equations for use in kinetic analysis. Equation 1 is the two substrate rate equation for an ordered reaction. K_A is the K_m for substrate A and K_B is the K_m for substrate B. Equations 2 and 3 show the rearrangement of the equation 1 into the same form as the Michealis-Menten equation by keeping the concentration of A constant. Under these conditions $\left(\frac{K_{iA}K_B}{[A]} + K_B\right)$ and $\left(\frac{K_A}{[A]} + 1\right)$ are constants, therefore equation 3 has the same form as the Michaelis-Menten equation (Equation 4) which describes a rectangular hyperbolae.

5.1.2 Kinetic comparison of wild-type *sa*NAL and F190Dpc *sa*NAL

Kinetic parameters were determined, using the TBA assay, for wild-type *sa*NAL and F190Dpc *sa*NAL with the substrates erythrose and pyruvate, and the wild-type substrates ManNAc and pyruvate. To convert the absorbance data from the TBA assay into the amount of aldol product formed a standard curve of sialic acid was used (Section 2.4.16.2 and Figure 7.2). Sialic acid can be used as a reference for any aldol product since the TBA assay works by cleaving any aldol product to produce a 1-3 dicarbonyl, which is then detected by TBA to form a highly

conjugated pink product. Therefore regardless of the chain length, group substitutions or stereochemistry, the compound that is detected in the TBA assay is the same 1-3 dicarbonyl.

Wild-type *sa*NAL for use in kinetic assays was purified by nickel affinity. F190Dpc for use in kinetic assays was produced by purification of F190C *sa*NAL, subsequent conversion to F190Dpc by chemical modification (Section 2.4.9 and 2.4.10.6) was carried out and then size exclusion chromatography was performed. To determine the kinetic parameters for wild-type and F190Dpc *sa*NAL, reactions were carried out in deep well 96-well plates containing 50 μ L sodium pyruvate (80 mM final concentration), 50 μ L of either erythrose or ManNAc (varied between 0.4 mM and 15 mM final concentration) and the reactions initiated by addition of 25 μ L of either wild-type or F190Dpc enzyme (0.7 mg mL⁻¹ stock concentration, to give final concentration 0.14 mg mL⁻¹). All reactions were carried out in duplicates and blanks were performed containing 25 μ L of buffer instead of enzyme. After a 2 hour incubation at room temperature, 100 μ L of each reaction was transferred into a new well containing 10 μ L TCA (12%) to stop the reaction, and then each reaction underwent the TBA assay, as previously described in Section 2.4.16.3. After the final stage of the TBA assay the plate was centrifuged at 4900 g for 6 min to separate any precipitate from the coloured solution. 85 μ L of each assayed reaction was then transferred into a flat-bottomed 96-well plate so that absorbance data at 550 nm could be acquired. Absorbance data for the F190Dpc enzyme, under these conditions, revealed that the absorbance values for the reactions with the highest concentrations of erythrose were above 1 absorbance unit. Therefore the kinetic assays were performed again using a lower concentration of F190Dpc (0.06 mg mL⁻¹ final concentration). Once absorbance data had been acquired for the condensation reactions, the amount of aldol product was quantified using a standard curve of sialic acid. The quantified data were then fitted using non-linear regression.

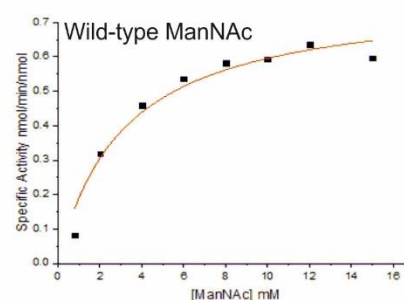
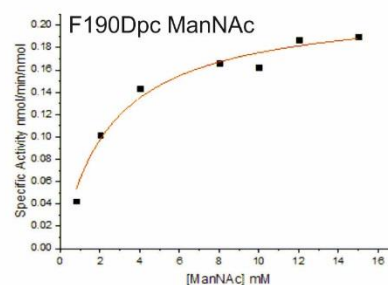
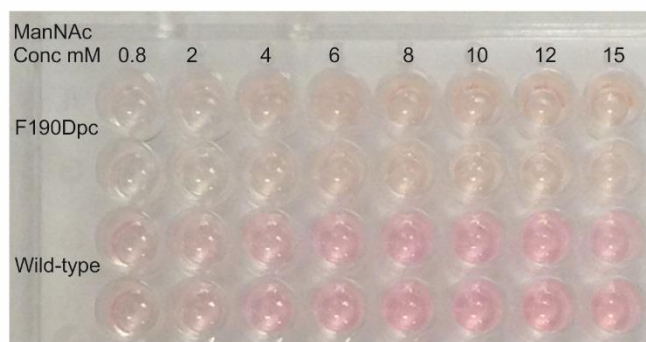
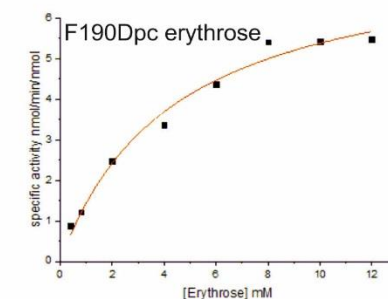
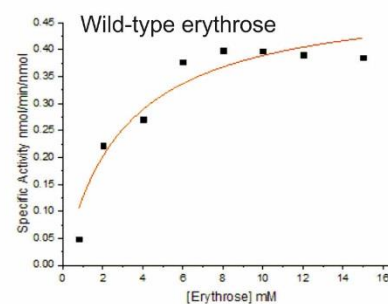
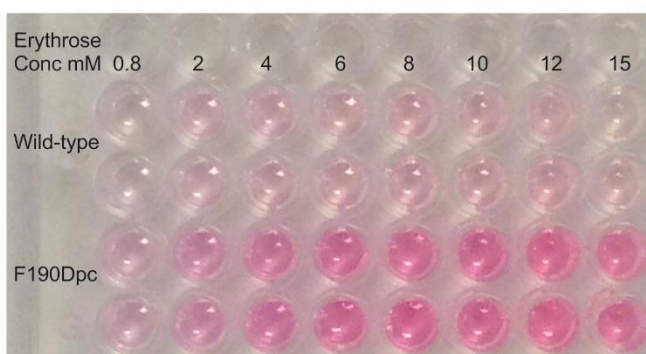
A**B**

Figure 5.4 Kinetic comparison of F190Dpc and wild-type *sa*NAL with the substrates erythrose and pyruvate, and the substrates ManNAc and pyruvate. Panel A shows an image of TBA assayed samples from condensation reactions containing; 80 mM pyruvate, between 0.8 and 15 mM ManNAc, 0.14 mg mL⁻¹ of wild-type or F190Dpc enzyme, and which had been incubated at room temperature for 2 hours. Absorbance data from the TBA assay was converted to specific activity using a standard curve to calculate the amount of product formed. Plots of specific activity against [ManNAc] for both enzymes are shown on the right-hand side. Panel B shows an image of TBA assayed samples from condensation reactions containing; 80 mM pyruvate, between 0.4 and 15 mM erythrose, 0.14 mg mL⁻¹ of wild-type enzyme or 0.06 mg mL⁻¹ F190Dpc enzyme, and which had been incubated at room temperature for 2 hours. Absorbance data from the TBA assay was converted to specific activity using a standard curve to calculate the amount of product formed. Plots of specific activity against [erythrose] for both enzymes are shown on the right-hand side.

Enzyme	Substrate	$k_{\text{cat}}^{\text{app}}$ (min^{-1})	$K_{\text{m}}^{\text{app}}$ (mM)	$k_{\text{cat}}^{\text{app}}/K_{\text{m}}^{\text{app}}$ ($\text{min}^{-1}.\text{mM}^{-1}$)
F190Dpc	Erythrose	7.6 ± 0.56	4.4 ± 0.81	1.7
F190Dpc	ManNAc	0.22 ± 0.01	2.5 ± 0.43	0.09
Wild-type	Erythrose	0.5 ± 0.05	3.0 ± 0.9	0.17
Wild-type	ManNAc	0.8 ± 0.05	3.1 ± 0.7	0.26

Table 5.1 Kinetic parameters for the aldol condensation of erythrose and pyruvate with both the wild-type and F190Dpc *sa*NAL, and the kinetic parameters for the aldol condensation of ManNAc and pyruvate with both the wild-type and F190Dpc *sa*NAL. Kinetic parameters were determined from the TBA assayed samples using a standard curve to determine the amount of aldol condensation product produced for each reaction, and then data was fitted using non-linear regression.

From Figure 5.4 it is obvious that F190Dpc *sa*NAL catalyses the reaction between erythrose and pyruvate much better than wild-type *sa*NAL. F190Dpc shows a 15-fold increase in $k_{\text{cat}}^{\text{app}}$ and a 10-fold increase in $k_{\text{cat}}^{\text{app}}/K_{\text{m}}^{\text{app}}$ compared to the wild-type enzyme with erythrose and pyruvate. Interestingly the F190Dpc reaction with erythrose and pyruvate has a 10-fold higher $k_{\text{cat}}^{\text{app}}$ and a 7-fold higher $k_{\text{cat}}^{\text{app}}/K_{\text{m}}^{\text{app}}$ compared to the wild-type reaction with ManNAc and pyruvate. Using Equation 5 shown in Figure 5.5, the specificity switch of the F190Dpc enzyme could be calculated, which showed a 30-fold switch in specificity toward the erythrose substrate. These results show that the modified F190Dpc *sa*NAL is not only better than the wild-type enzyme with the altered substrate erythrose, but is also better than the wild-type enzyme with the wild-type substrate ManNAc.

Equation 5.

$$\text{specificity switch} = \frac{\left(\frac{k_{cat}^{app}}{K_m^{app}} \text{F190Dpc erythrose} \right) / \left(\frac{k_{cat}^{app}}{K_m^{app}} \text{F190Dpc ManNac} \right)}{\left(\frac{k_{cat}^{app}}{K_m^{app}} \text{Wild - type erythrose} \right) / \left(\frac{k_{cat}^{app}}{K_m^{app}} \text{Wild - type ManNac} \right)}$$

Figure 5.5 Equation 5 used to calculate the specificity switch for F190Dpc *s*NAL with the substrates erythrose and pyruvate.

5.2 F190X saturation library

Clearly F190Dpc *s*NAL is a better catalyst than wild-type *s*NAL for the reaction of erythrose and pyruvate. However since this activity increase was achieved by use of an nCAA, we were interested to see whether it could be achieved by inserting one of the proteogenic amino acids at the same position in the enzyme. Therefore to compare the non-canonical Dpc side chain to all of the proteogenic amino acids, a saturation library at position 190 was produced.

5.2.1 Production of F190X saturation library

The DNA for the wild-type and F190C mutant had already been produced during the course of this work, and so there were 18 mutants at position 190 to be made. The F190X saturation library was produced by site-directed mutagenesis; primers were designed to insert each of the remaining 18 proteogenic amino acids at position 190. Sequencing was used to confirm that the mutations at position 190 had been successfully inserted (data shown in Figure 5.6) and that no other mutations had occurred in the DNA. The mutant plasmids were then transformed into BL21 competent *E. coli* cells for expression. Variant proteins were expressed on a 1 litre scale and then purified by batch nickel affinity chromatography (Section 2.4.1).

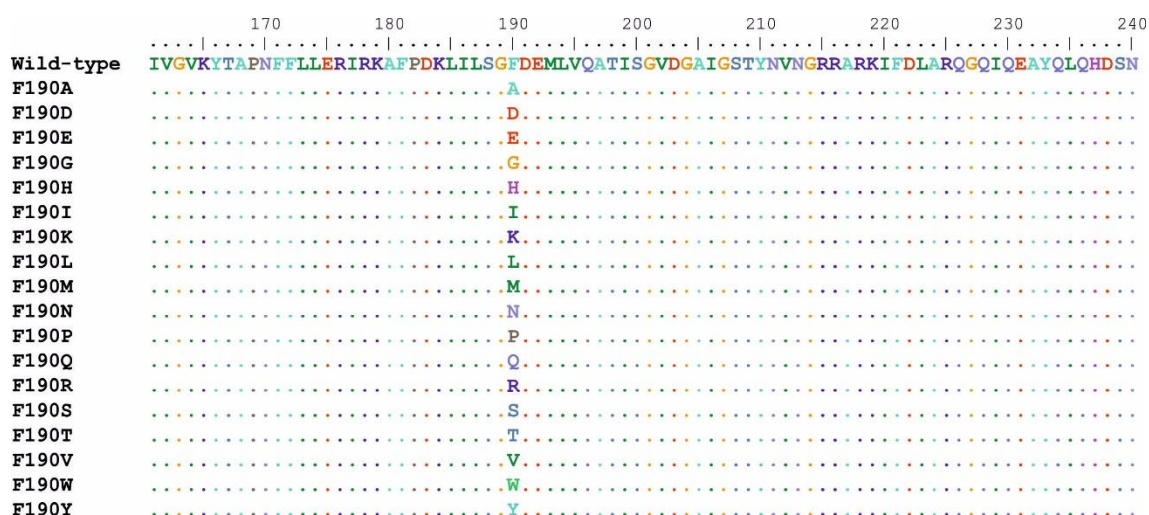


Figure 5.6. Sequencing data for the 18 F190 mutants. The 18 different proteogenic amino can be seen to have been inserted at position 190.

5.2.2 Kinetic comparison of the F190X library and F190Dpc *sa*NAL

Kinetic assays were carried out with the purified F190 variants. Kinetic assays were performed with 5 different aldehyde concentrations (from 0.8 mM to 15 mM final concentration) using the same TBA assay method as for the wild-type and F190Dpc enzymes. Kinetic parameters for all F190 variants and wild-type are shown in Table 5.2.

F190 Mutant	$k_{\text{cat}}^{\text{app}}$ (min^{-1})	$K_{\text{m}}^{\text{app}}$ (mM)	$k_{\text{cat}}^{\text{app}}/K_{\text{m}}^{\text{app}}$ ($\text{min}^{-1}.\text{mM}^{-1}$)
A	0.9 ± 0.2	7.8 ± 4.5	0.12
C	1.0 ± 0.16	7.2 ± 2.3	0.14
D	1.2 ± 0.06	2.8 ± 0.45	0.43
E	1.5 ± 0.3	4.7 ± 2.4	0.32
G	1.0 ± 0.34	5.0 ± 4.0	0.20
H	1.1 ± 0.09	4.0 ± 0.82	0.28
I	0.9 ± 0.07	2.9 ± 0.8	0.31
K	1.0 ± 0.05	2.0 ± 0.40	0.50
L	0.9 ± 0.02	2.3 ± 0.26	0.39
M	1.1 ± 0.05	2.0 ± 0.40	0.55
N	0.84 ± 0.07	3.3 ± 0.80	0.25
P	1.0 ± 0.11	4.2 ± 1.2	0.24
Q	0.89 ± 0.05	1.5 ± 0.40	0.6
R	1.0 ± 0.01	1.6 ± 0.08	0.63
S	1.1 ± 0.1	2.7 ± 1.1	0.41
T	1.2 ± 0.03	2.0 ± 0.4	0.6
V	1.1 ± 0.03	1.6 ± 0.3	0.69
W	1.1 ± 0.03	2.4 ± 0.24	0.46
Y	1.0 ± 0.05	1.3 ± 0.3	0.77
Wild-type (F)	0.5 ± 0.05	3.0 ± 0.9	0.17
F190Dpc	7.6 ± 0.56	4.4 ± 0.81	1.7

Table 5.2 Kinetic parameters for the aldol condensation of erythrose and pyruvate for the F190X saturation library, wild-type and the modified F190Dpc *sa*NAL. Kinetic parameters were determined by using the TBA assay (Section 2.4.16.3) to analyse samples from aldol condensation reactions containing; 80 mM pyruvate, between 0.8 and 15 mM erythrose, 0.14 mg mL⁻¹ of wild-type of F190X enzyme or 0.06 mg mL⁻¹ of F190Dpc which had then been incubated for 2 hours at room temperature. The amount of aldol condensation product (DOH) produced was then quantified using a standard curve and data fitted using non-linear regression.

From the kinetic parameters the F190Dpc enzyme has the highest turn-over rate for the substrates erythrose and pyruvate. The increase in activity brought about by F190Dpc can be seen in panel A of Figure 5.7, which shows an image of TBA assayed samples from the kinetic assays performed. Comparison of $k_{\text{cat}}^{\text{app}}$, shows that F190Dpc has a 5-fold higher $k_{\text{cat}}^{\text{app}}$ than the next best proteogenic variant which is F190E. Therefore this level of increased activity, with erythrose and pyruvate, could not have been achieved had we used a traditional protein engineering approach, which is restricted to the 20 proteogenic amino acids, at this position. Interestingly, the kinetics for the saturation library and F190Dpc show that the $K_{\text{m}}^{\text{app}}$ does not change significantly between all the variants and the increase in activity caused by the Dpc side chain is produced from a change in $k_{\text{cat}}^{\text{app}}$.

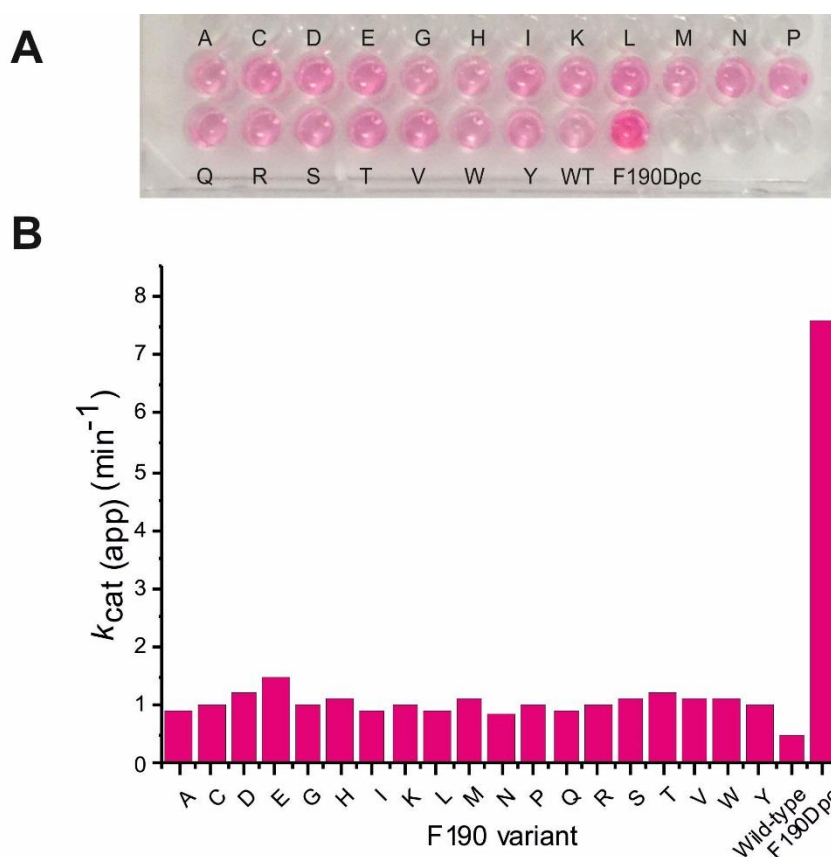
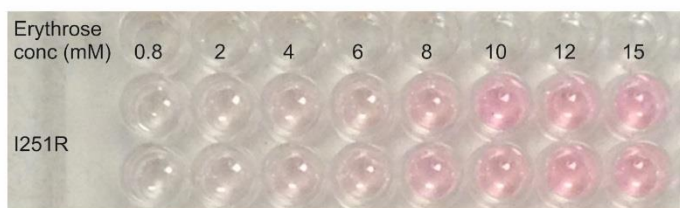


Figure 5.7 Kinetic comparison of the F190X saturation library and F190Dpc for the aldol condensation of erythrose and pyruvate. Panel A shows an image of the TBA assayed samples from condensation reactions containing; 80 mM pyruvate, 10 mM erythrose, 0.14 mg mL^{-1} of wild-type of F190X enzyme or 0.06 mg mL^{-1} of F190Dpc and had been incubated at room temperature for 2 hours. Panel B shows a bar chart representation of the $k_{\text{cat}}^{\text{app}}$ values for the F190X variants, wild-type and F190Dpc *sa*NAL.

5.3 Comparison of F190Dpc to directed evolution studies

The F190Dpc *sa*NAL proves to be a better catalyst for the reaction of erythrose and pyruvate than the wild-type *sa*NAL, and it also outperforms all the proteogenic amino acids at position 190 for this reaction. A previous directed evolution study on the *E. coli* NAL, carried out by the Wong group, produced a variant of the *E. coli* NAL with an increased activity with L-3-deoxy-manno-2-octulosonic acid (Hsu *et al.*, 2005), further studies on this enzyme showed that by altering the residue at position 251 the substrate specificity could be altered. By producing the variant V251R the enzyme had an increased specificity towards erythrose as the aldehyde acceptor (Chou *et al.*, 2011). Unfortunately due to the differences in experimental procedure for determining the kinetic parameters, it was not possible to directly compare the kinetic parameter values. However the *E. coli* NAL has very high sequence and structural similarity to *sa*NAL, therefore we decided to make the equivalent mutation in *sa*NAL (I251R) and determine the kinetic parameters for this variant.

The I251R mutation in *sa*NAL was made using site-directed mutagenesis and once sequencing had confirmed the mutation; the variant was expressed and purified. The I251R *sa*NAL variant was purified using nickel affinity chromatography and a sample was taken and analysed by ESI mass spectrometry. A peak with the mass 34036 Da was observed in the mass spectrum which correlated to the molecular mass of the I251R variant 34038 Da (Figure 5.8).

A**B**

Enzyme	Substrate	$k_{\text{cat}}^{\text{app}}$ (min^{-1})	$K_{\text{m}}^{\text{app}}$ (mM)	$k_{\text{cat}}^{\text{app}}/K_{\text{m}}^{\text{app}}$ ($\text{min}^{-1} \cdot \text{mM}^{-1}$)
F190Dpc	Erythrose	7.6 ± 0.56	4.4 ± 0.81	1.7
Wild-type	Erythrose	0.5 ± 0.05	3.0 ± 0.9	0.17
I251R	Erythrose	1.2 ± 0.2	20 ± 9.1	0.06

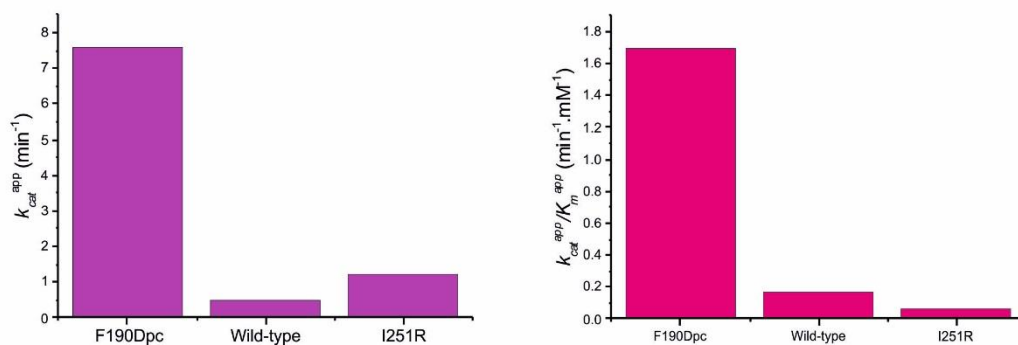
C

Figure 5.9 Aldol condensation kinetics, for the condensation of erythrose and pyruvate, carried out with the I251R *sa*NAL variant. Panel A shows an image of the TBA assayed samples from the condensation reactions which contained; 80 mM pyruvate, between 0.8 and 15 mM erythrose and 0.14 mg mL⁻¹ I251R enzyme and had been incubated for 2 hours at room temperature. Panel B shows the kinetic parameters for the aldol condensation of erythrose and pyruvate performed with F190Dpc, wild-type and I251R *sa*NAL. Kinetic parameters were determined from the TBA assayed samples by using a standard curve to quantify the amount of aldol condensation product (DOH) produced and then data was fitted using non-linear regression. Panel C shows bar chart representations to compare the $k_{\text{cat}}^{\text{appp}}$ and $k_{\text{cat}}^{\text{appp}}/K_{\text{m}}^{\text{appp}}$ for the aldol condensation of erythrose and pyruvate catalysed by F190Dpc, wild-type and the I251R variant *sa*NAL. F190Dpc shows a 6-fold higher $k_{\text{cat}}^{\text{appp}}$ than the I251R variant and a 30-fold higher $k_{\text{cat}}^{\text{appp}}/K_{\text{m}}^{\text{appp}}$ than the I251R variant.

These experiments show that the ncAA-containing-enzyme F190Dpc has an increased activity, with the substrates erythrose and pyruvate, over that of the wild-type enzyme. The Dpc side chain also out-performs all of the proteogenic amino acids at position 190. And the F190Dpc enzyme is also better than a variant developed through directed evolution, for this activity.

5.4 Crystallographic studies of F190Dpc *sa*NAL

Thorough kinetic analysis had been carried out on the F190Dpc enzyme to determine the switch in substrate specificity. Therefore, to try and determine the mechanism by which the Dpc side chain achieved this increase in activity with erythrose and pyruvate, X-ray crystallographic studies were carried out. The crystal structure of the F190Dpc enzyme was solved in complex with pyruvate, using the same methods as for the wild-type and K165- γ -thialysine enzymes.

5.4.1 Crystallisation of F190Dpc *sa*NAL

Crystallisation of F190Dpc was carried out using the same conditions as were optimised for the wild-type *sa*NAL enzyme (Timms *et al.*, 2013). The F190Dpc enzyme was produced on a 50 mg scale using the chemical modification procedure. Once the modification to the Dpc side chain had been confirmed using ESI negative ion mode mass spectrometry, the protein was re-folded by dialysis and further purified using size exclusion chromatography (Section 2.4.12). F190Dpc protein at approximately 7 mg mL⁻¹ was applied to a Superdex preparative S200 column (GE life sciences), and fractions containing protein which eluted at 180 mL were collected. Fractions containing the pure protein were concentrated to between 7-10 mg mL⁻¹ for crystallisation. Crystallisation conditions used were 200 mM NaCl, 100 mM Tris/HCl (pH 7.0-8.5) and 18-28% (w/v) PEG 3350, and crystal trays were prepared by hand using hanging drop vapour diffusion. Each well contained 3 drops of either 3 or 4 μ L, with the ratios 1:2 2:2 2:1 of protein to mother liquor. Using these methods single crystals formed in 7-10 days.

To produce the cryo-cooled enzyme-pyruvate complex for data collection, the crystals were soaked in mother liquor containing PEG 400 and sodium pyruvate. Crystals were first soaked in mother liquor supplemented with 15 % PEG 400 (w/v) and 100 mM sodium pyruvate, before being transferred to mother liquor supplemented with 20 % PEG 400 (w/v) and 100 mM sodium pyruvate and the final soak contained 25 % PEG 400 and 100 mM sodium pyruvate. Crystals were soaked in each of the different PEG 400 concentrations for approximately 30 s, and then flash cooled in liquid nitrogen ready for data collection.

5.4.2 Data collection, processing and refinement

Data collection for F190Dpc in complex with pyruvate was performed at Diamond Light Source (Oxfordshire, UK). The data set was recorded on beamline I0-4 at 100 K from a single crystal (Figure 5.10). Integration and scaling of the data set was performed using the CCP4 programs MOSFLM and SCALA (Leslie, 2006, Evans, 2006). Scaling was carried out with a resolution range of 1.7-75.52 Å and 1-1400 images were used. Scaling statistics and full structural statistical data for the F190Dpc structure, in complex with pyruvate, can be seen in Table 5.3.

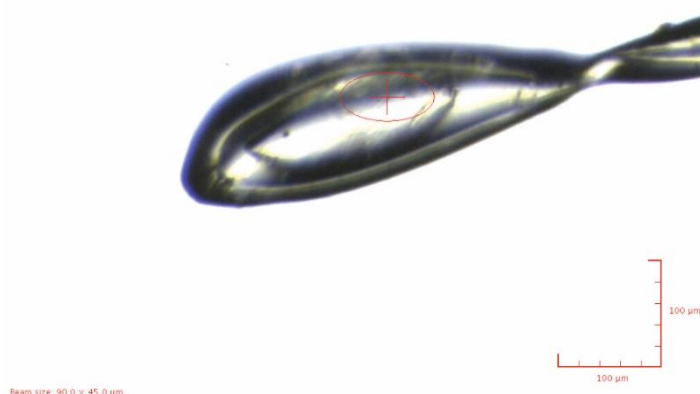


Figure 5.10. Image of F190Dpc crystal mounted in a loop during data collection.

After scaling had been performed, molecular replacement was used to determine the F190Dpc structure. Molecular replacement was carried out using PHASER (Mccoy, 2007), with the wild-type apo *sa*NAL structure (PDB code 4ahp) as the search model consisting of chain A of the homotetramer with the water molecules removed. By running the CCP4 program 'Matthews Coefficient' (Kantardjieff and Rupp, 2003, Matthews, 1968), which works to estimate the solvent content of the unit cell based on the molecular weight of the protein and the unit cell parameters, it was estimated that 4 monomers were present in the unit cell and so 4 copies were searched for during the molecular replacement. The unit cell parameters were $a=55.5$ Å, $b=134.3$ Å, $c=79.5$ Å, $\alpha=90$, $\beta=108$, $\gamma=90$ with space group $P2_1$ and an upper resolution of 1.7 Å. The molecular replacement solution showed the crystal structure to be a tetramer, as was seen with the wild-type and K165- γ -thialysine crystal structures.

F190Dpc saNAL in complex with pyruvate	
Diamond beamline	I0-4
Space group	$P2_1$
Resolution (Å)	1.7
a (Å)	55.5
b (Å)	134.3
c (Å)	79.5
$R_{\text{merge}} \ S^*$	0.071(0.42)
R_{pim}^{+*}	0.035(0.24)
Observed reflections	565,256
Unique reflections	120,355
Completeness (%)*	99.4(95.7)
Multiplicity*	4.7(4.0)
$\langle I/\sigma \rangle / \sigma^*$	11.2(3.0)
Refinement	
R_{factor} (%)	0.18
R_{free} (%)†	0.21
No. of protein atoms	9298
No. of solvent molecules	278
No of ligand atoms	20
Average overall B-factor (Å ²)	23.3
Average ligand B-factor (Å ²)	22.1
Average solvent B-factor (Å ²)	27
RMS bond lengths (Å) §	0.019
RMS bond angles (Å) §	2.04
Ramachandran analysis, the percentage of residues in the regions of plot (%) ‡	
Most favoured	98.0
Outliers	4'

$$\S R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}$$

+ Rpim - precision-indicating (multiplicity-weighted) Rmerge, relative to all I+ or I-.

* Values given in parentheses correspond to those in the outermost shell of the resolution range.

† R_{free} was calculated with 5% of the reflections set aside randomly.

§ Based on the ideal geometry values of Engh & Huber (Engh and Huber, 1991).

‡ Ramachandran analysis using the program MolProbity (Lovell *et al.*, 2003)

'The side-chain of Tyr111, in all subunits, is in close proximity to Leu142, Thr143 and Phe110 from an adjacent chain causing the phi and psi angles of Tyr111 to lie in an unfavoured region of the Ramachandran plot

Table 5.3. Structural statistics for the F190Dpc saNAL X-ray crystal structure in complex with pyruvate.

Once molecular replacement had been performed and a suitable model found, refinement was carried out. The CCP4 program REFMAC5 (Murshudov *et al.*, 2011) was used to carry out multiple rounds of refinement. For refinement the starting model was the wild-type apo *sa*NAL structure with the waters removed and the residues at position 165 and 190 modelled as alanine to prevent any bias in the model. Initial rigid body refinement was carried out, followed by iterative rounds of restrained refinement and model building carried out in REFMAC and COOT (Emsley *et al.*, 2010) respectively. During the restrained refinement, TLS refinement was also carried out. From both the $2F_{\text{obs}}-F_{\text{cal}}$ and $F_{\text{obs}}-F_{\text{cal}}$ electron density maps it was clear that the residue at position 165 did not fit an alanine residue, as there was excess positive electron density extending out from the C β that indicated a lysine in complex with pyruvate. Therefore the lysine in complex with pyruvate was built into the structure in the same way as for the wild-type *sa*NAL in complex with pyruvate discussed in Section 3.3.2. The lysine in complex with pyruvate was designated the same HET code, KPI, as before and the same library file for the bond lengths and angles was used.

Similarly at position 190 there was also excess positive electron density extending from the C β when the residue was modelled as an alanine. The positive electron density was quite large at the γ position indicating a more electron dense atom such as a sulphur atom. However, when the residue was modelled as a cysteine residue there was still excess density extending from the γ position. This excess density indicated the presence of the non-canonical side chain Dpc (panel A Figure 5.11), and so the Dpc side chain was modelled into all subunits. To generate the structure and library file for the Dpc side chain, the ligand builder extension in COOT was used. Once the structure file and the library file were produced they were checked and edited manually and then modelled into the F190Dpc structure. Subsequent rounds of refinement were carried out and the Dpc side chain was observed to be in multiple conformations in two out of the four subunits (Figure 5.11).

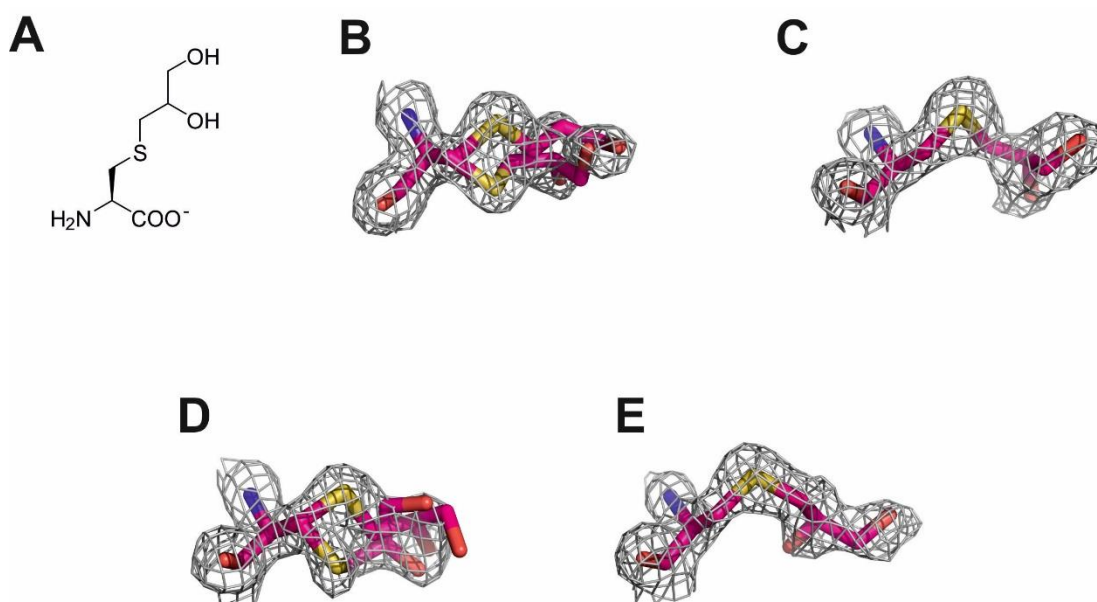


Figure 5.11 Structure of the dihydroxypropylcysteine (Dpc) side and electron density for the Dpc side chain found at position 190 in the F190Dpc crystal structure in complex with pyruvate. Panel A shows the structure of the Dpc side chain. Panel B shows the $2F_{\text{obs}}-F_{\text{cal}}$ density for the Dpc side chain at position 190 in chain A. Panel C shows the $2F_{\text{obs}}-F_{\text{cal}}$ density for the Dpc side chain at position 190 in chain B. Panel D shows the $2F_{\text{obs}}-F_{\text{cal}}$ density for the Dpc side chain at position 190 in chain C. Panel E shows the $2F_{\text{obs}}-F_{\text{cal}}$ density for the Dpc side chain at position 190 in chain D. In chains A and C the Dpc side chain is seen to adopt two conformations and in chains B and D it is seen to adopt one.

5.4.3 Crystallographic comparison of wild-type *sa*NAL and F190Dpc *sa*NAL

The crystal structure of F190Dpc was determined to try and elucidate the mechanism by which this non-canonical side chain produces the increase in activity, with the substrates erythrose and pyruvate. The crystal structure of F190Dpc in complex with pyruvate, is a tetramer made up of monomers with an $(\alpha/\beta)_8$ TIM barrel fold, where the α helices enclose the β sheets, which was also seen in the wild-type *sa*NAL, *E. coli* NAL and *H. influenza* NAL structures (Timms *et al.*, 2013, Barbosa *et al.*, 2000, Lawrence *et al.*, 1997). The F190Dpc structure in complex with pyruvate had the same ordered loop region from residues 137-146, as was observed in the wild-type structure in complex with pyruvate (Timms *et al.*, 2013). Alignment of the F190Dpc structure and the wild-type structure, both in complex with pyruvate has an RMSD of 0.5 Å. Therefore there was very little overall structural change caused by the incorporation of the Dpc side chain into the active site of *sa*NAL.

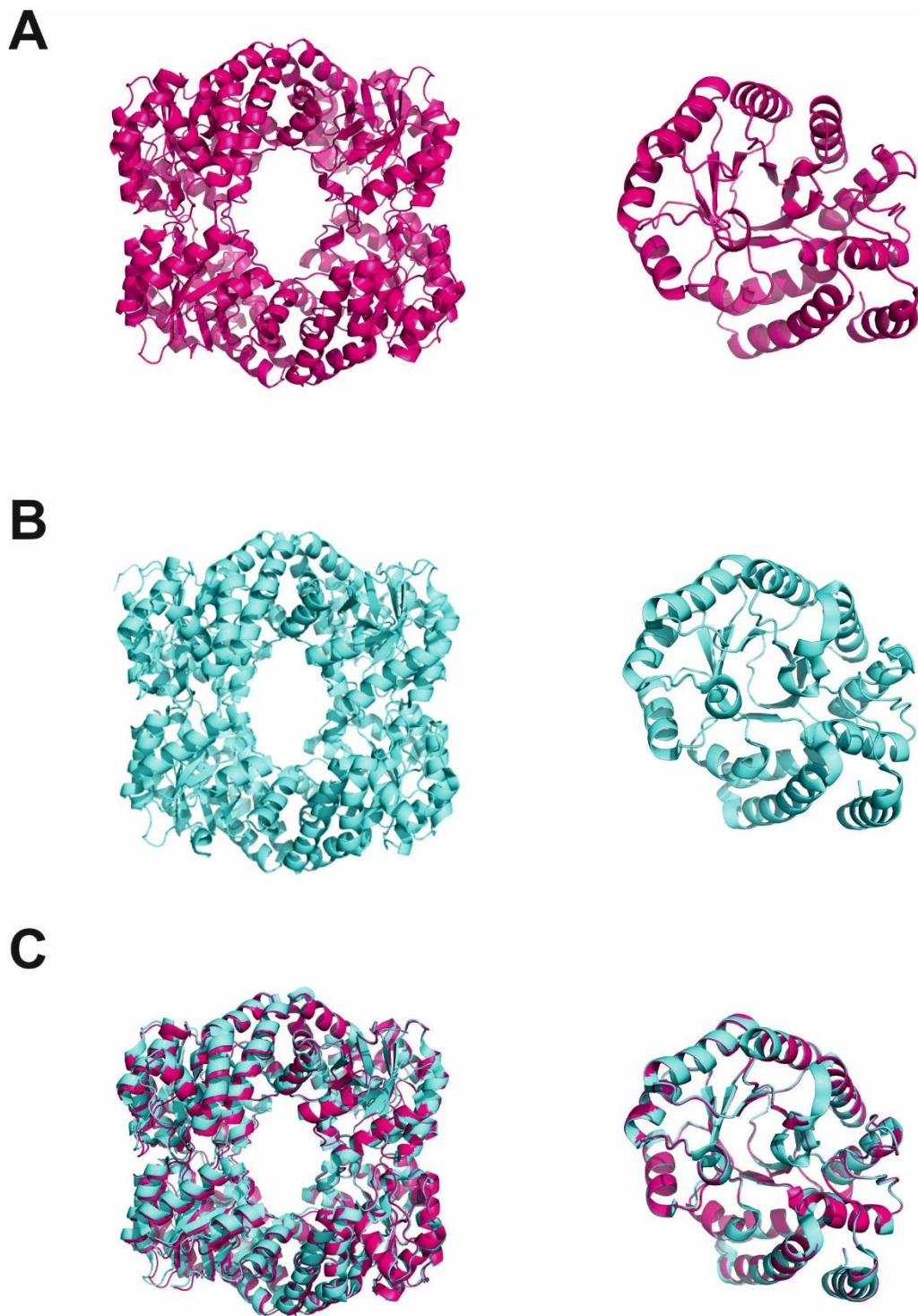


Figure 5.12 Crystal structures of F190Dpc and wild-type *saNAL* both in complex with pyruvate. Panel A shows the homotetramer and monomer of F190Dpc *saNAL* in complex with pyruvate. Panel B shows the homotetramer and monomer of wild-type in complex with pyruvate. Panel C shows overlays of the homotetramer and monomer of both wild-type *saNAL* and F190Dpc *saNAL* in complex with pyruvate.

Comparison of the active site of the F190Dpc structure and the wild-type structure, both in complex with pyruvate, shows that the pyruvate binding end of the active site appears to be unaffected by the incorporation of the ncAA at the other end of the active site. Comparison of the phenylalanine and Dpc side chains (Panel B Figure 5.13) shows the aromatic side chain has been exchanged for a more polar side chain which can adopt multiple conformations, as was seen in two out of the four subunits, and is also 0.8 Å longer than the phenylalanine. Panel A of Figure 5.13 highlights the positioning of residue 190 in the active site of both the wild-type *sNAL* and F190Dpc. From previous studies it is known that the full-length wild-type product Neu5Ac lies in the active site between the catalytic Lys165 and Glu192; Lys165 forms a Schiff base with the pyruvate end of the Neu5Ac and Glu192 hydrogen bonds to the glycerol end of Neu5Ac (Daniels *et al.*, 2014). The altered substrate specificity, with erythrose and pyruvate, which is achieved by the Dpc side chain, produces a shortened C7 product (DOH) compared to the wild-type C9 product (Neu5Ac). Therefore it was hypothesised that the positioning of the Dpc side chain at 190 meant that it could interact and potentially stabilise the binding of the product in the active site, whereas the aromatic phenylalanine could not.

Comparison of the region around position 190 (Panel C Figure 5.13), shows that the side chains of Asp141 and Glu192 are within close proximity to the phenylalanine side chain in the wild-type structure, and the Dpc side chain in the modified enzyme structure. Due to the two hydroxyl groups in the Dpc side chain, Dpc may be able to interact with Asp141 and Glu192 whereas the aromatic side chain of phenylalanine would not. In the F190Dpc crystal structure Asp141 is approximately 3.1 Å away from the Dpc side chain and Glu192 is approximately 3.3 Å from Dpc (approximate values shown as they are averaged between the four subunits).

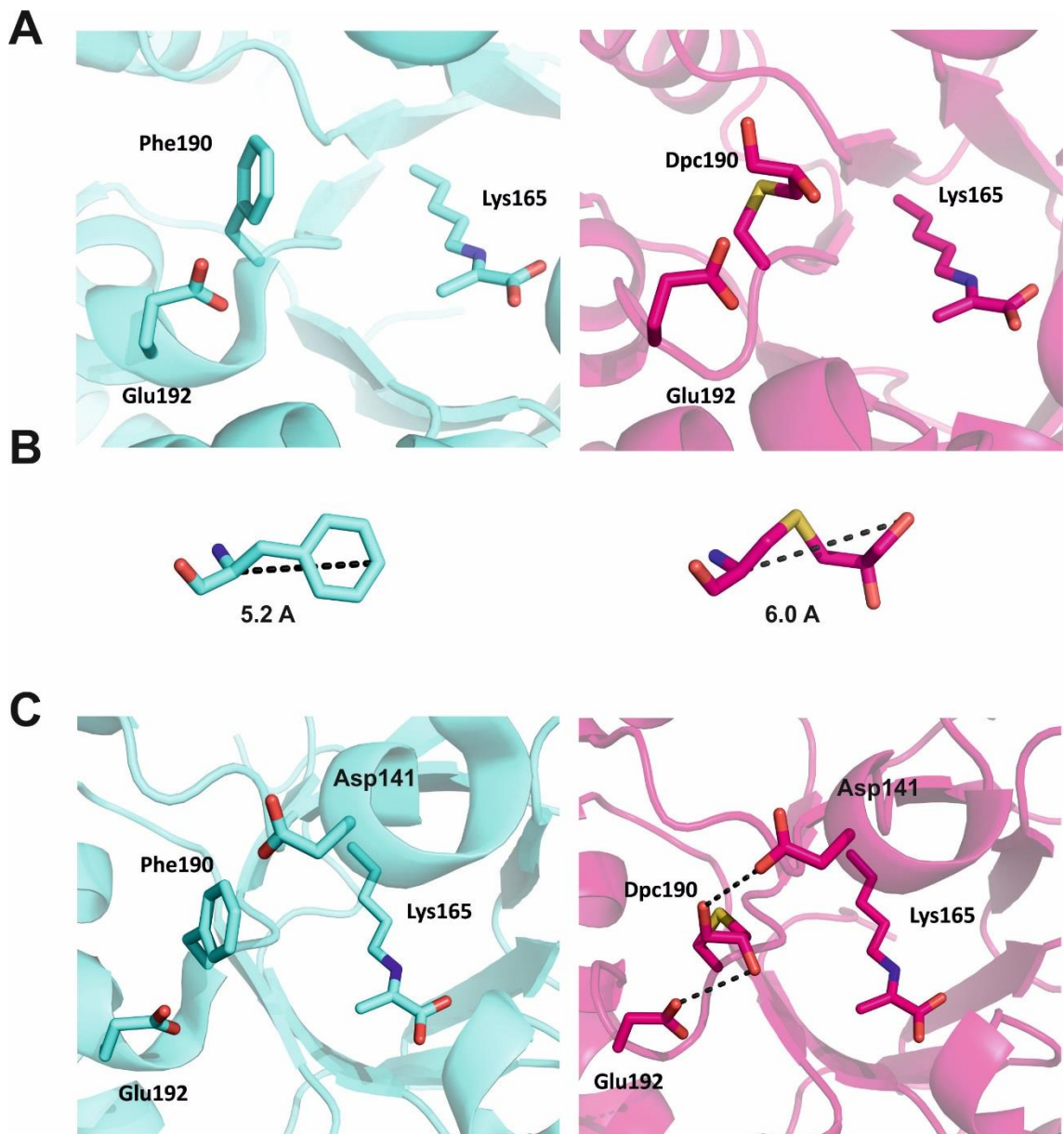


Figure 5.13 Panel A shows a comparison of the active site of wild-type and F190Dpc crystal structures in complex with pyruvate. Lys165 and Glu192 are highlighted as sticks to show where the full-length product would lie in the active site and the residue at 190 is also highlighted as sticks. Panel B shows a comparison of the phenylalanine side chain found at position 190 in the wild-type *sa*NAL and the dihydroxypropylcysteine side chain found at position 190 in the modified F190Dpc *sa*NAL, the residue lengths from the C α to the terminal are highlighted by dashed lines. Panel C shows comparison of residues around position 190, potential interactions between Dpc190, Asp141 and Glu192 are highlighted by dashed lines.

By overlaying the active sites of the wild-type and F190Dpc structures, in complex with pyruvate (shown in Figure 5.14), the structural differences in the active sites can be seen. The main difference is at position 167, where the threonine at this position in the F190Dpc structure lies further into the active site. The C α of Thr167 is position 1.0 Å closer to the Lys165

bound to pyruvate, in the F190Dpc structure than in the wild-type structure. Thr167 has previously been shown to stabilise the transition state of the wild-type reaction by stabilising the aldehyde of ManNAc which becomes the C4 hydroxyl of the full length Neu5Ac (Daniels *et al.*, 2014).

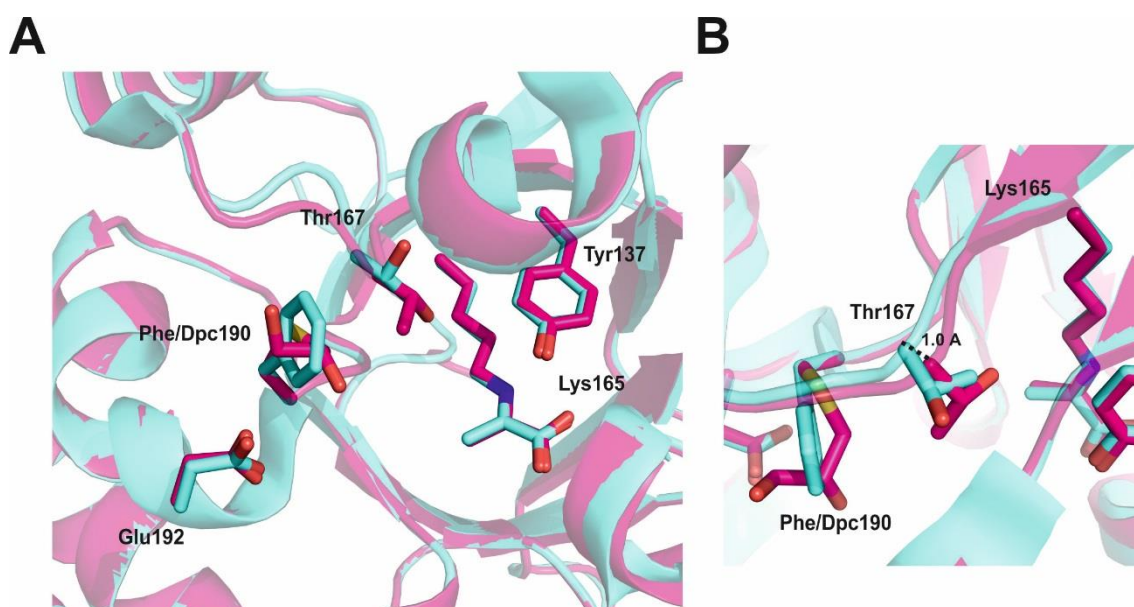


Figure 5.14 Overlay of chain B of F190Dpc crystal structure in complex with pyruvate and chain B of the wild-type *saNAL* crystal structure in complex with pyruvate. Panel A shows the active site of both the F190Dpc and the wild-type *saNAL* crystal structures with active site residues highlighted as sticks. Panel B shows Thr167 in F190Dpc and the wild-type *saNAL* crystal structures, with the difference in position of the C α atoms of Th167 highlighted by a dashed line.

By producing a distance matrix of both the wild-type and F190Dpc structures it was possible to highlight the regions most affected by the incorporation of the ncAA, Dpc (Figure 5.15). Thirteen residues which line the active site and which have side chains directed into the active site were selected; distances between each residue were taken in both the wild-type and F190Dpc structures. The wild-type values were subtracted from the F190Dpc values to highlight which distances had been reduced and which had been increased between the two structures. A pink colour in the distance matrix indicates distances that have decreased in the F190Dpc *saNAL* structure and a purple colour indicates distances that have increased in the F190Dpc *saNAL* structure, compared to the wild-type *saNAL* structure.

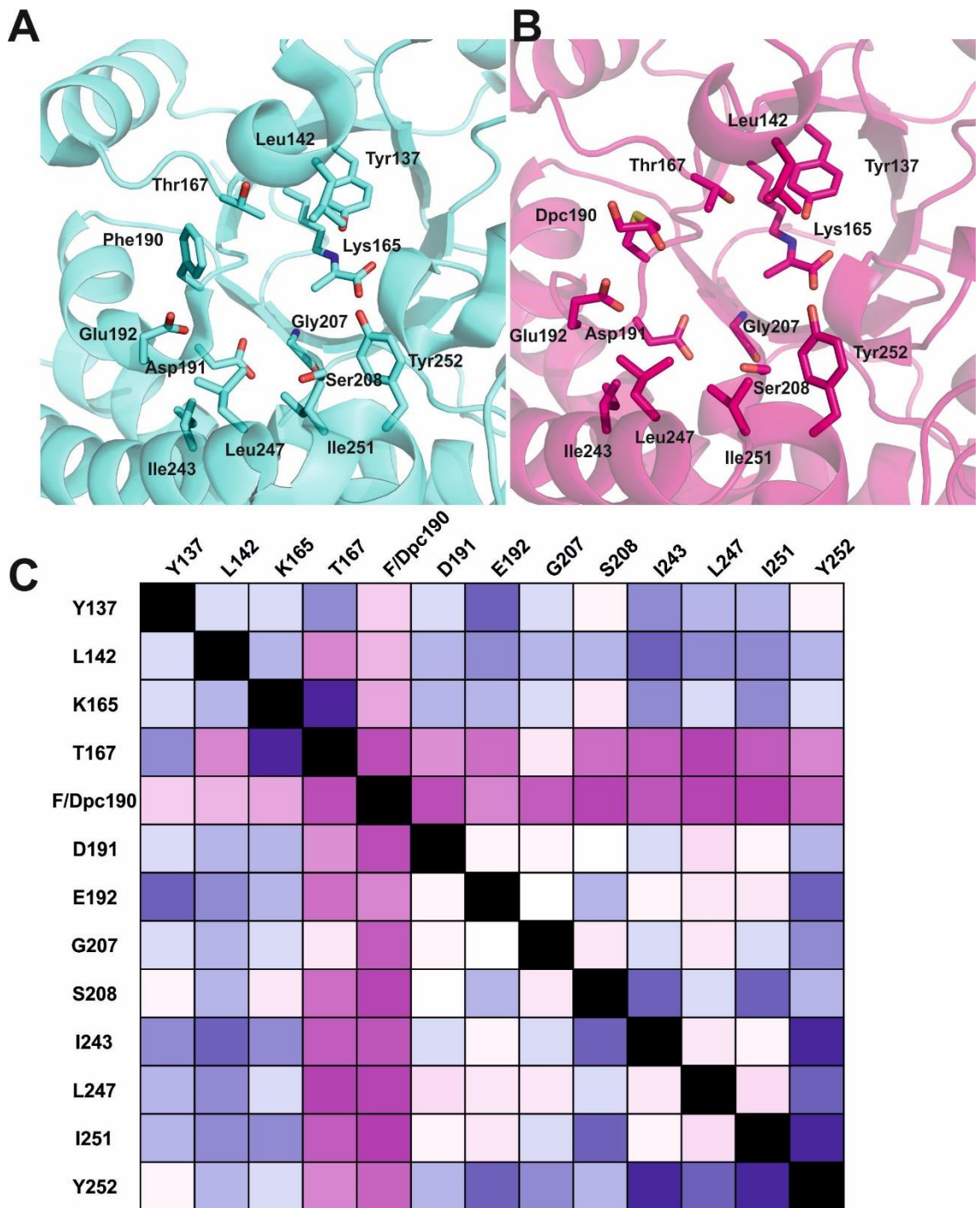


Figure 5.15 Distance matrix of the wild-type and F190Dpc crystal structures. Panels A and B show the active site of the wild-type (cyan) and F190Dpc (pink) crystal structures, with the residues that distances have been measured between highlighted as sticks. Panel C shows the distance matrix where the difference between the wild-type active site distances and the F190Dpc active site distances are displayed as follows; a purple colour shows distances that have increased in the F190Dpc structure compared to the wild-type structure, and a pink colour shows distances that have decreased in the F190Dpc *sa*NAL structure compared to the wild-type *sa*NAL structure.

The distance matrix shows that in the F190Dpc structure the distances between residue 190 and all the other 12 residues have decreased, presumably this has occurred as the Dpc side chain is a longer side chain than the wild-type phenylalanine. Interestingly, in the F190Dpc structure Thr167 moves away from the catalytic lysine at position 165 and moves closer to the residues that line the aldehyde end of the active site, this may bring Thr167 into closer proximity to the aldehyde substrate erythrose. The differences in the active site shape can also be seen in Figure 5.16 where the active sites of the wild-type and F190Dpc structures, in complex with pyruvate are shown as surface representations. The active site of the F190Dpc structure appears to be narrowed by the presence of the Dpc side chain compared to the wild-type, the narrowing of the active site may help the binding of the shortened aldehyde substrate erythrose.

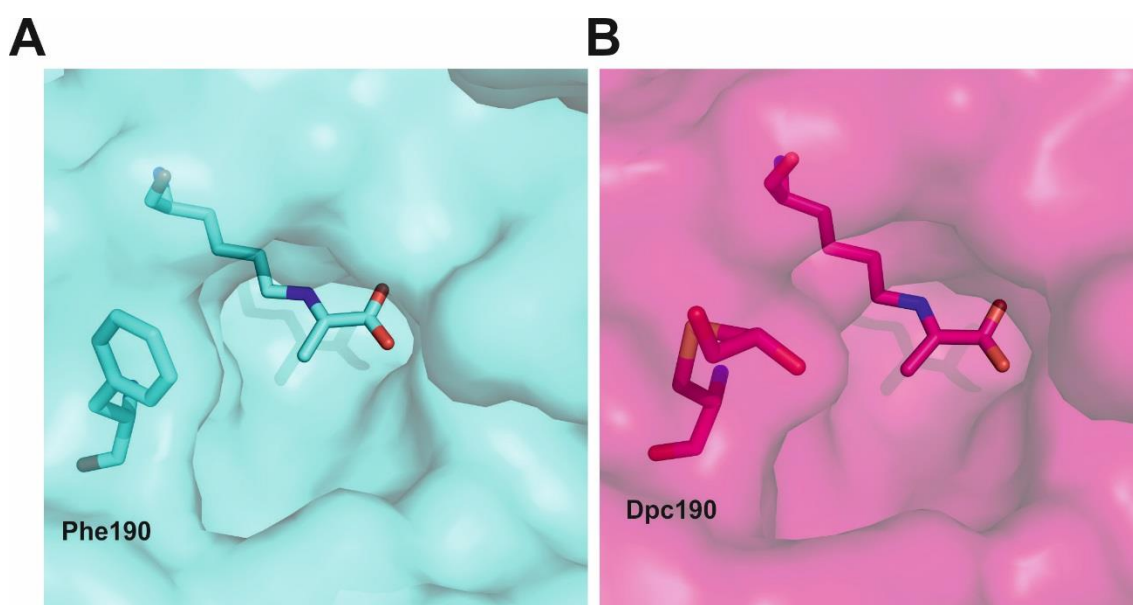


Figure 5.16 F190Dpc and wild-type crystal structures, in complex with pyruvate, displayed as surface representations to show the differences in shape of the active sites.

Although the crystal structure of F190Dpc, in complex with pyruvate, showed some structural differences from the wild-type, there was no clear reasoning for the increase in activity with erythrose. To try and further elucidate this and to investigate the binding of the product in the active site, attempts were made to determine the crystal structure in complex with the full-length DOH product. F190Dpc crystals were soaked with 100 mM DOH but no evidence for the full-length product was observed in the electron density maps. Previous studies used an

inactive Y137A variant to trap the full-length wild-type substrate Neu5Ac in the active site of the wild-type *E.coli* protein and successfully determine the crystal structure (Daniels *et al.*, 2014), therefore a Y137A/F190C double mutant was produced in *sa*NAL. Once the Y137A/F190C variant was purified the chemical modification procedure was carried out to produce the modified Y137AF190Dpc. This modified protein was then crystallised and soaks with 100 mM DOH and also soaks with 100 mM pyruvate and 100 mM erythrose were carried out. However, again, no evidence for the full-length DOH product was observed in the resultant electron density maps. Co-crystallisation experiments with the Y137A/F190Dpc protein were also carried out, where between 15 mM and 75 mM DOH was added into the crystallisation conditions. Crystals were obtained but again there was no evidence of the full-length product in the active site.

In the crystal structures of the Y137A/F190Dpc variant the loop region between residue 137 and 146 could be seen in the electron density, even though there was no evidence for pyruvate or anything bound in the active site. In the previous crystallographic studies this loop region became visible when a substrate was bound in the active site (Section 3.3.3) (Timms *et al.*, 2013). Therefore this rigidity of the 137-146 loop, which lies over the entrance of the active site, may have prevented any of the substrates entering the active site. Since it was not possible to determine the crystal structure with the full-length DOH product in the active site, molecular modelling and energy minimisation experiments were employed to investigate the binding of the full-length product in the active site.

5.5 Modelling and energy minimisations

Computational modelling and energy minimisations were carried out to try and highlight the differences in binding of the F190Dpc and wild-type enzymes with the product DOH, as it was not possible to do this by crystallographic methods. These experiments were carried out by Dr Katie Simmons (University of Leeds).

5.5.1 Modelling of DOH into the active site of F190Dpc *sa*NAL and wild-type *sa*NAL

For the energy minimisation experiments to be performed, models of both the wild-type and F190Dpc enzymes with the full-length DOH had to be produced. Models were created in the program Maestro. The *E.coli* NAL structure in complex with the full-length Neu5Ac was used as a template to build the full-length DOH into the active site of both the wild-type *sa*NAL and F190Dpc *sa*NAL active sites, since the *E.coli* NAL is an excellent structural mimic of *sa*NAL. The

C4 hydroxyl of DOH can adopt either the *R* or the *S* configuration (Figure 5.17) when the aldol condensation occurs, as was indicated by the NMR studies (Section 4.5.3). Therefore, both the (4*R*) and (4*S*) forms of DOH were modelled, individually, into the active sites of the F190Dpc and wild-type structures.

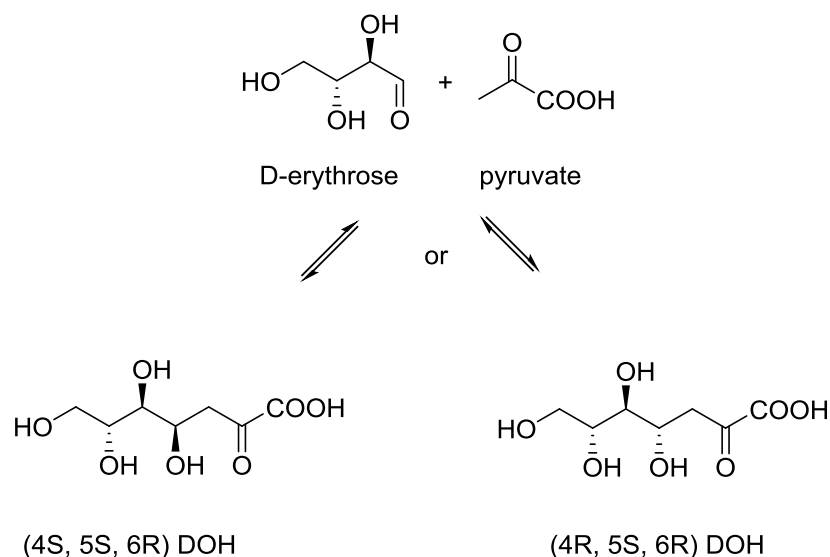


Figure 5.17 Aldol condensation of D-erythrose and pyruvate to form the product (4*R*, 5*S*, 6*R*) DOH or (4*S*, 5*S*, 6*R*) DOH.

Once the models of F190Dpc *sa*NAL with (4*R*) and (4*S*) DOH in the active site, and wild-type *sa*NAL with (4*R*) and (4*S*) DOH in the active site, had been produced they were subjected to energy minimisation. Energy minimisations were performed by selecting a region within 6 Å of the DOH product; the side chains in this area were allowed to minimise to their lowest energy state, whilst the back bone was kept rigid. The backbone was kept rigid as there were no significant backbone changes between the F190Dpc or wild-type structures, in complex with pyruvate, and the *E.coli* structure in complex with the full-length Neu5Ac product (Daniels *et al.*, 2014). The models underwent 500 rounds of energy minimisation and then the values were checked to see if they had converged. If the energy values had not converged the models underwent a further 500 rounds of energy minimisation, to ensure that the values had converged and produced the lowest energy model. These lowest energy models were then compared to elucidate potential mechanisms behind the increase in activity with erythrose and pyruvate, achieved by incorporation of the Dpc side chain.

5.5.2 Comparison of F190Dpc *sa*NAL and wild-type *sa*NAL models

The models generated from the computational energy minimisations revealed that the pyruvate binding region of the active site appears unchanged, and the glycerol end of the full-length product, DOH, is stabilised by Asp191 and Ser208. It is clear that the Dpc side chain does not directly interact and stabilise the shortened C7 product, as was hypothesised from the crystal structure. However the interactions between Asp141, Glu192 and Dpc190, which were identified in the F190Dpc crystal structure (Panel C of Figure 5.13), produce significant differences in the F190Dpc models compared to the wild-type models. In the F190Dpc models, the Dpc side chain forms a hydrogen bonding network between Asp141 and Glu192, whereas in the wild-type models the phenylalanine cannot form these interactions. These hydrogen bonding interactions are present in the F190Dpc models with either the (4*R*) or (4*S*) DOH in the active site (Figure 5.18).

The hydrogen bonding network between Dpc190, Asp141 and Glu192 also appears to reposition the Dpc side chain in the F190Dpc models, compared to the phenylalanine in the wild-type *sa*NAL models. The hydrogen bonding interactions serve to pull the Dpc side chain further into the active site (Figure 5.19), whereas the wild-type phenylalanine residue appears to be pushed out from the active site. In the F190Dpc model with (4*S*) DOH the Dpc side chain is 2.3 Å closer into the active site and in the model with (4*R*) DOH it is 1.8 Å closer into the active site. This positioning of the Dpc side chain helps to narrow the active site as can be seen in the surface representations in Figure 5.20.

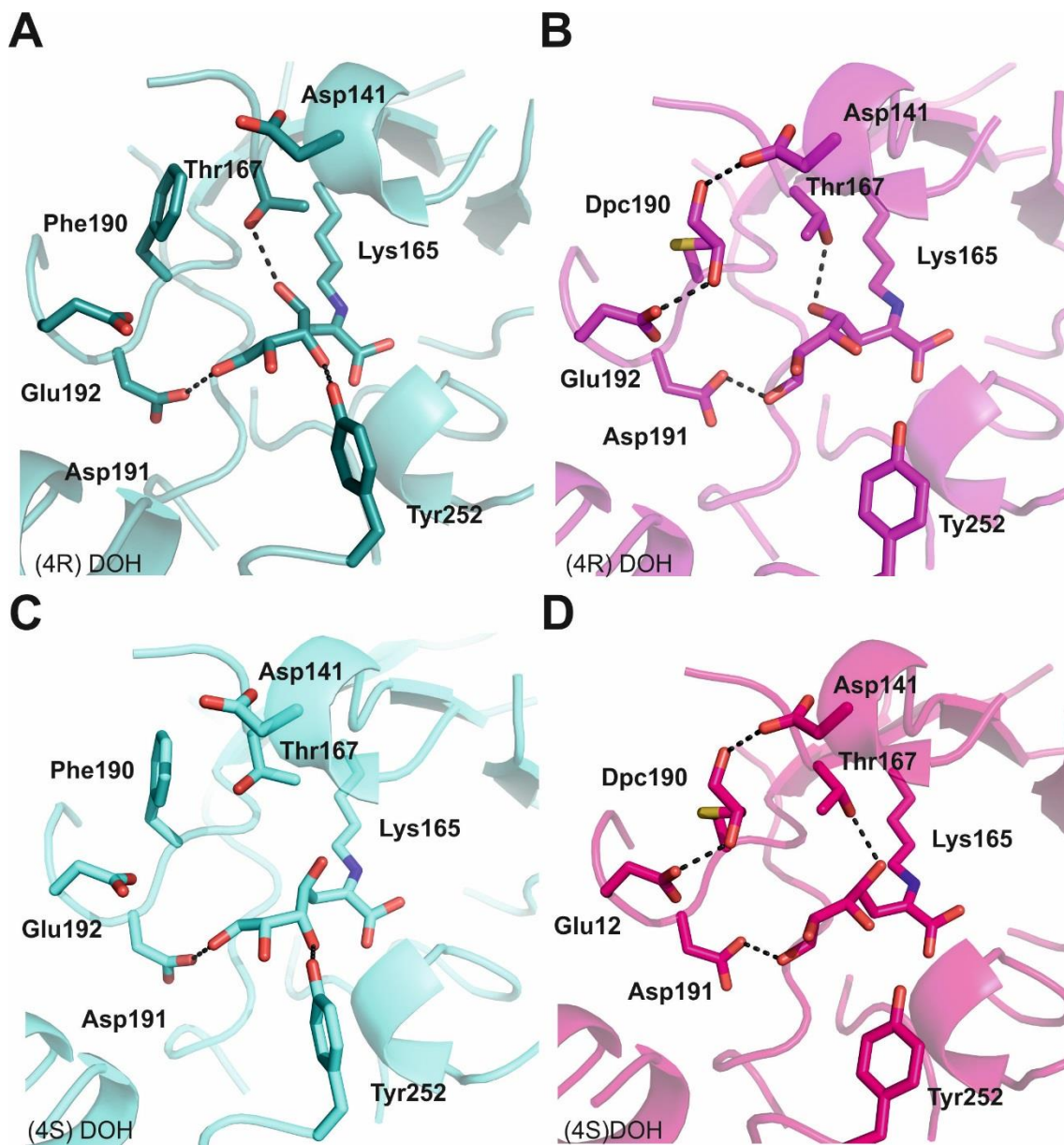


Figure 5.18 Active site models of F190Dpc and wild-type with the (4R) and (4S) DOH product. Models of wild-type and F190Dpc with the (4R) DOH product are shown in panels A and B in teal and magenta respectively. Models of wild-type and F190Dpc with the (4S) DOH product are shown in panels C and D in cyan and pink respectively. Active site residues are highlighted in sticks and hydrogen bonds are shown by dashed lines.

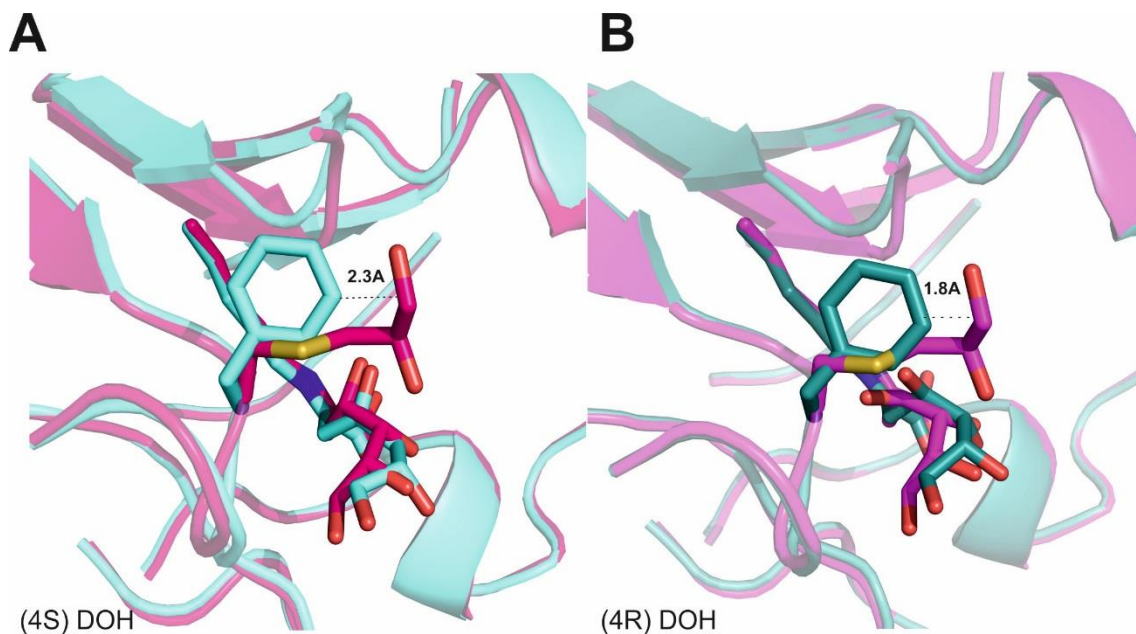


Figure 5.19 Overlays of the F190Dpc and wild-type models. Panel A shows the overlay of the models with the (4S) DOH product, F190Dpc is shown in pink and wild-type is shown in cyan. Panel B shows the overlay of the models with the (4R) DOH product, F190Dpc is shown in magenta and wild-type is shown in teal. The distance between the wild-type phenylalanine and F190Dpc dihydroxypropylcysteine side chain is highlighted by a dashed line in both of the overlays.

From the surface representations, both of the F190Dpc models have smaller, narrower active sites than the respective wild-type models (Figure 5.20). This narrowing of the active site was seen in the comparison of the crystal structures (Figure 5.16), but is more pronounced in the models. Narrowing of the active site is clearly due to the Dpc side chain, and may help to bind the shorter C7 DOH product over the longer, wild-type, C9 Neu5Ac. The wild-type models have much more open active sites, than the F190Dpc models, and this may allow the smaller C7 substrate to bind in more than one conformation (Kim *et al.*, 2014).

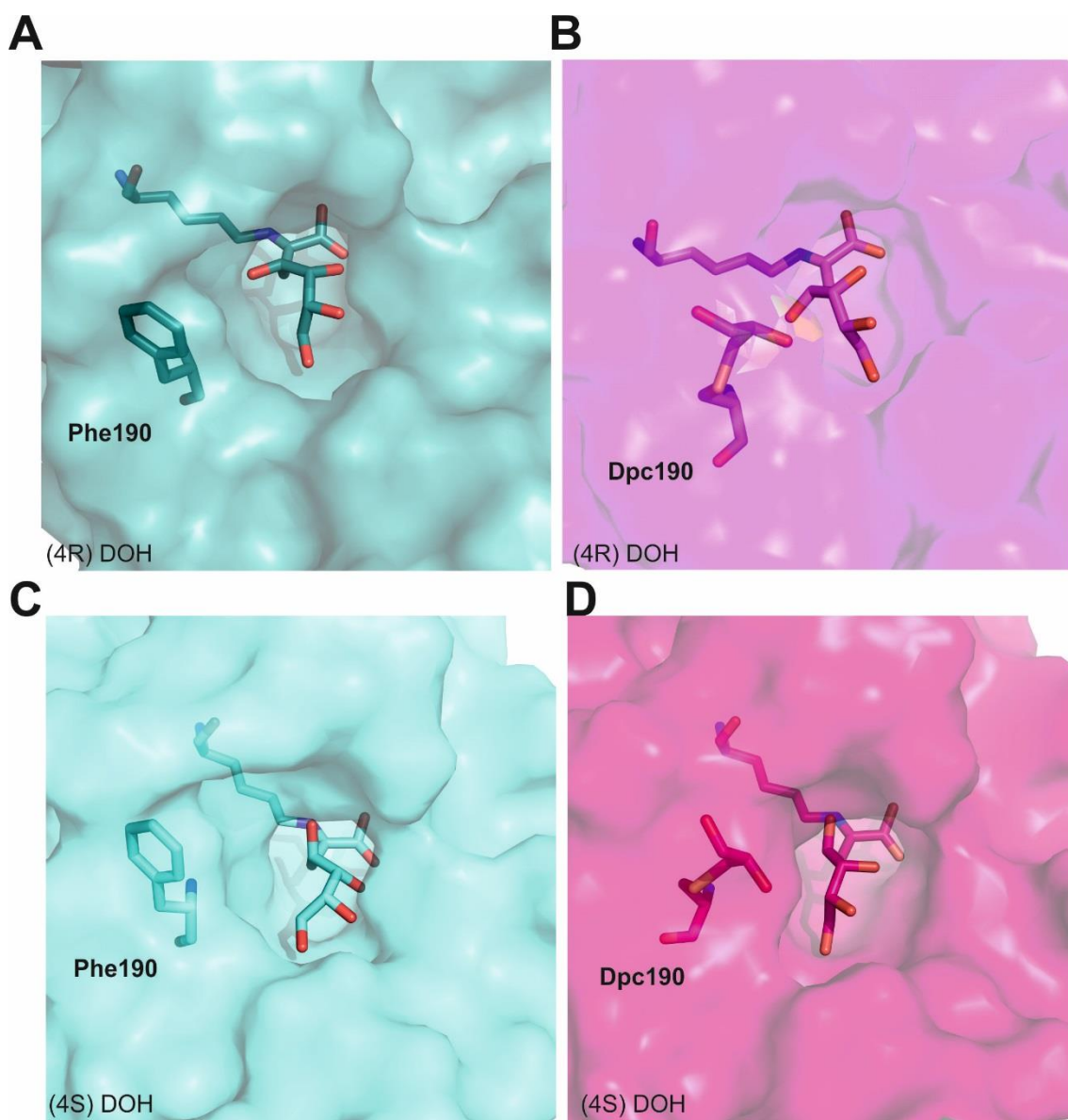


Figure 5.20 Comparison of the active sites of the F190Dpc and wild-type models in surface representation. Models of F190Dpc and wild-type with the (4R) DOH product are shown in panels A and B in magenta and teal respectively. Models of F190Dpc and wild-type with the (4S) DOH product are shown in panels C and D in cyan and pink respectively.

There are also several other significant differences at the aldehyde end of the active site in the F190Dpc models compared to the wild-type models. Figure 5.18 shows that in the wild-type models, Tyr252 is within hydrogen bonding distance, approximately 3 Å in both the (4R) and (4S) DOH models, of the C5 hydroxyl whereas in the F190Dpc it is much further away, approximately 4.5 Å in both the (4R) and (4S) DOH models, from the C5 hydroxyl. There are also differences in the positioning of Thr167, in the F190Dpc and wild-type models with the (4R) DOH Thr167 is positioned so it can hydrogen bond to the C4 hydroxyl in both the

wild-type and F190Dpc *sa*NAL models. However when modelled with the (4S) DOH only Thr167 in the F190Dpc model is with hydrogen bonding distance of the C4 hydroxyl. The hydrogen bonding interaction between the C4 hydroxyl and Thr167 has been previously shown to be important for transition state stabilisation (Daniels *et al.*, 2014). In the F190Dpc model with the (4S) DOH Thr167 is 2.8 Å from the C4 hydroxyl of the product whereas in the wild-type it is 4.6 Å from the C4 hydroxyl of the product, therefore the wild-type may not be able to stabilise the transition state to produce the (4S) DOH product. Since the increase in activity for erythrose and pyruvate with the F190Dpc enzyme comes from an enhanced k_{cat} , as shown from the kinetic analysis, it is likely that a transition state stabilisation is occurring.

Therefore from the energy minimised models it seems to be a combination of multiple factors that bring about the increase in activity, with the substrates erythrose and pyruvate, in the modified F190Dpc *sa*NAL. The combination of the narrowing of the active site brought about by the hydrogen bonding network that the Dpc side chain forms and the difference in positioning of Thr167, are probably the main reasons behind the activity increase.

5.6 Summary

By incorporating an ncAA into an enzyme it has been possible to alter the substrate specificity of the enzyme. By exchanging the phenylalanine at position 190, in *sa*NAL, for the non-canonical side chain dihydroxypropyl cysteine (Dpc), the k_{cat}^{app} for the substrates erythrose and pyruvate was increased 15-fold and k_{cat}^{app}/K_m^{app} was increased 10-fold. Interestingly, the F190Dpc *sa*NAL was also a better enzyme with the substrates erythrose and pyruvate than the wild-type *sa*NAL with the wild-type substrates ManNAc and pyruvate. Kinetic comparison of F190Dpc to a saturation library of all the twenty proteogenic amino acids at position 190, showed that the Dpc side chain outperformed all of the proteogenic amino acids at this position. F190Dpc *sa*NAL was also a better enzyme with erythrose and pyruvate than a variant produced from a previously published directed evolution study.

X-ray crystallographic studies, along with computational modelling and energy minimisations have been used to try and determine the mechanism by which F190Dpc *sa*NAL achieves the increase in activity with erythrose and pyruvate. These studies highlighted that the Dpc side chain causes the active site to become smaller, which may increase the activity with the shortened substrate. Also in the F190Dpc structure and models the positioning of the residue

Thr167 is altered compared to the wild-type, and may be altered in a way which allows it to more easily stabilise the transition state in the F190Dpc enzyme with the DOH product.

This work shows that both altering and improving catalysis is something that is achievable using ncAAs. It also shows that by further diversifying the side chain chemistries throughout an enzyme it might be possible to find activities that wouldn't have been achievable if only the proteogenic amino acids had been available.

Chapter 6 Summary, Future Perspectives and Conclusions

6.1 Summary

The thesis describes the use of a chemical modification method to insert a variety of ncAAs individually throughout the active site of *sa*NAL. It also describes the screening of these ncAA-containing enzymes for altered activities, and the characterisation of an ncAA-containing enzyme (F190Dpc *sa*NAL) with altered substrate specificity. Through the course of the work presented in this thesis; the optimisation and insertion of 13 different ncAAs at positions within *sa*NAL has been demonstrated, and multiple X-ray crystal structures have been solved, including the structures of two ncAA-containing enzymes, to help elucidate the effect of the ncAAs on the enzyme structure.

The introduction to this thesis describes previous examples, from the literature, where ncAAs have been used to alter catalysis. Enzyme redesign experiments, using site-specific incorporation of ncAAs by nonsense suppression, have been shown to increase catalytic efficiencies of a bacterial phosphotriesterase and a nitroreductase (Ugwumba *et al.*, 2011, Jackson *et al.*, 2006) and alkylation of a cysteine variant to produce an *S*-butyl cysteine increased the catalytic efficiency of a glutathione transferase (Hegazy *et al.*, 2006). However, this work describes the first instance of utilising ncAAs in a way analogous to directed evolution or semi-rational redesign experiments, to alter enzyme activities. Incorporation of a wide variety of different ncAAs has been used to produce many modified variants, screening of the modified variants for activity has been carried out and then the modified enzyme with an altered activity was identified and characterised. The chemical modification approach allowed a far more diverse range of ncAAs to be inserted at a cysteine residue, than would have been achievable had nonsense suppression been used, as the core structural motifs of the ncAAs incorporated in this manner are restricted to lysine, phenylalanine and tyrosine, and a new orthogonal tRNA/tRNA synthetase pair would be needed for each ncAA to be inserted (Wan *et al.*, 2014, O'donoghue *et al.*, 2013, Dumas *et al.*, 2015). The chemical modification approach also allowed for facile incorporation of many different ncAA side chains just by the production of one cysteine variant, without problems with incomplete incorporation of the ncAA which has hampered some nonsense suppression experiments (O'donoghue *et al.*, 2013, Zhang *et al.*, 2013). The use of the diBr compound to selectively modify cysteine to the electrophilic bio-orthogonal dha, allows a more specific chemical modification method, than the more

traditionally used nucleophilic attack onto a cysteine side chain, which could also modify histidine and lysine residues (Chalker *et al.*, 2009). Due to the electrophilic nature of dha it can undergo a Michael addition with a sulphur nucleophile, producing the side chain of the nAA, this reaction is highly selective due to the lack of highly electrophilic centres within the proteogenic amino acids.

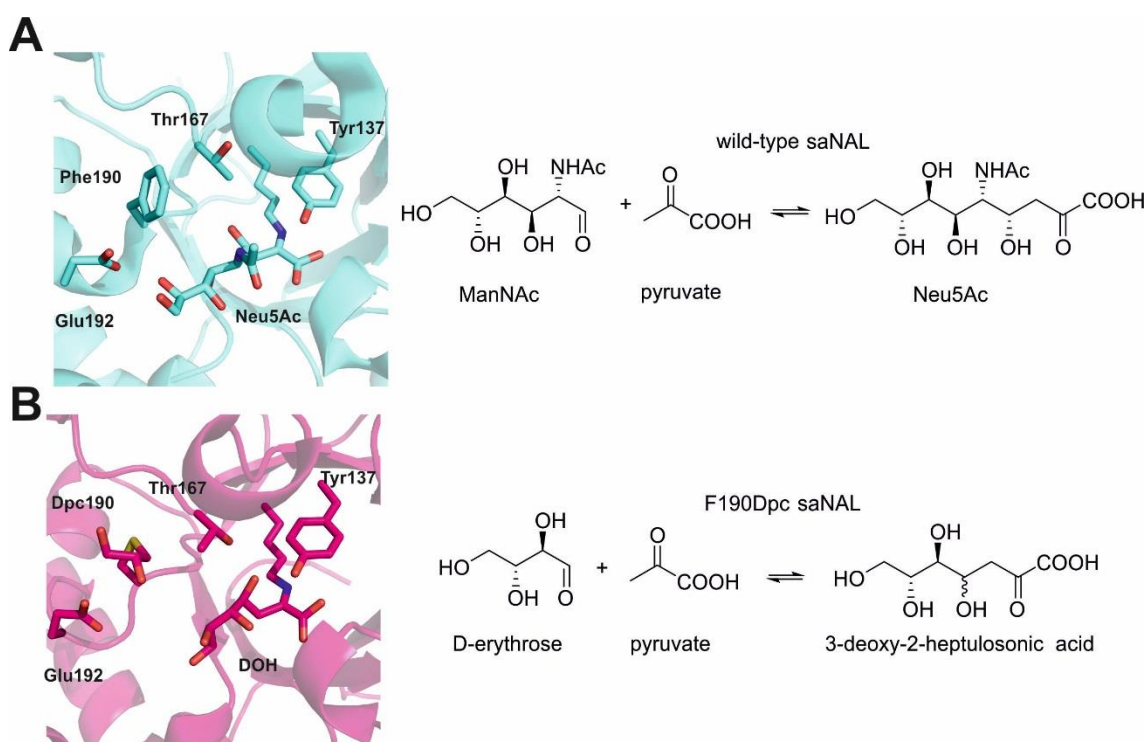


Figure 6.1 Wild-type and modified F190Dpc *sa*NAL structures and reactions. Panel A shows the crystal structure of the wild-type *sa*NAL with the product *N*-acetylneuraminic acid (Neu5Ac) modelled into the active site, and the wild-type *sa*NAL reversible aldol reaction of *N*-acetyl-D-mannosamine (ManNAc) and pyruvate to form Neu5Ac. Panel B shows the crystal structure of the modified F190Dpc *sa*NAL with the product (4*S*) 3-deoxy-2-heptulosonic acid (DOH) modelled into the active site, and the F190Dpc *sa*NAL reversible aldol reaction of D-erythrose and pyruvate to form DOH.

This work has shown that by using nAAs it has been possible to alter the activity of *sa*NAL. A modified nAA-containing enzyme, F190Dpc *sa*NAL, was found that had an increased activity with the shorter aldehyde substrate erythrose than the wild-type ManNAc aldehyde substrate. F190Dpc was shown to be a more efficient enzyme with erythrose and pyruvate as substrates than the wild-type enzyme, showing a 10-fold increase in k_{cat}^{app}/K_m^{app} and a 15-fold increase in

$k_{\text{cat}}^{\text{app}}$. F190Dpc was also shown to be a more efficient for the altered activity than the wild-type enzyme with the wild-type substrates, ManNAc and pyruvate, showing a ~10-fold increase in $k_{\text{cat}}^{\text{app}}$. Production and kinetic characterisation of a saturation library at position 190 also revealed that the non-canonical Dpc residue enhanced activity with erythrose by more than was achievable by any of the proteogenic amino acids, at that position.

6.2 Future perspectives

This work shows the use of ncAAs to alter enzyme substrate specificity; however it may be possible to further enhance these activities. Also the facile incorporation of ncAAs within an enzyme active site may provide a route to altering enzyme activities in other ways. For example by incorporating non-proteogenic side chains it may be possible to engineer and exploit interactions and chemistries not available in naturally produced proteins.

6.2.1 Further alteration of substrate specificity

The inclusion of the non-canonical Dpc side chain enhances activity with the substrates erythrose and pyruvate. However it may be possible to further enhance this activity by combining the chemical modification procedure with more traditional protein engineering approaches. For example by the addition of further proteogenic mutations at other sites within the enzyme the altered activity may be increased. Due to the necessity to retain the cysteine residue at position 190 and not to insert any other cysteine residues, which would prevent the modification procedure from being site-specific, it would be required to insert mutations in a targeted manner. Other residues within the active site could be targeted for saturation mutagenesis, due to the significant amount of structural data on both wild-type and F190Dpc *s*NAL, it may be possible to select mutations that would increase the binding efficiency with a smaller substrate. Also a CASTing (combinatorial active-site saturation test) approach could be used; sets of two or three residues, in close proximity to the active site, would be randomised and then screened for enhanced activity (Reetz *et al.*, 2005, Reetz *et al.*, 2006). Both of these approaches target specific residues and produce fairly small library sizes for screening, therefore would be applicable to combining with the chemical modification procedure. The variants with the cysteine at position 190 would be grown in 96 deep-well plates and purified using HisGrab™ 96 well plates (Pierce, Rockford, USA). The proteins could be unfolded and the chemical modification procedure carried out in plate format, followed by screening for enhanced activity by the TBA assay. Combination of traditional protein engineering and

incorporation of ncAAs by nonsense suppression has been carried out previously and shown to be successful in enhancing functionality. A glycan-binding antibody, engineered to incorporate a *p*-boronophenylalanine in response to a TAG codon underwent evolution of the complementarity determining region (CDR) loops and was evolved to bind an acyclic amino sugar (Liu *et al.*, 2009).

The work presented in this thesis has focussed on the alteration of the aldehyde acceptor substrate of *sa*NAL. This approach was taken as previous studies had shown the evolvability of the aldehyde binding pocket and that wild-type NAL had natural substrate promiscuity with regards to aldehyde acceptors (Williams *et al.*, 2005, Williams *et al.*, 2006, Campeotto *et al.*, 2010, Joerger *et al.*, 2003, Chou *et al.*, 2011, Hsu *et al.*, 2005, Gijzen *et al.*, 1996, Kim *et al.*, 1988). However it would also be interesting to attempt to alter the ketone donor substrate of NAL. The substrate scope of the ketone donor is far stricter than that of the aldehyde acceptor; 3-fluoropyruvate has been shown to be a donor for some pyruvate-dependent aldolases including the *E. coli* NAL (Daniels *et al.*, 2014, Chokhawala *et al.*, 2007), however 2-oxobutyrate, 3-bromo-pyruvate, 3-hydroxypyruvate are not accepted by NAL (Kim *et al.*, 1988). Therefore an approach using ncAAs could be taken to attempt to evolve activity with these donor substrates that would produce interesting biologically active compounds, for example production of compounds with two new chiral centres.

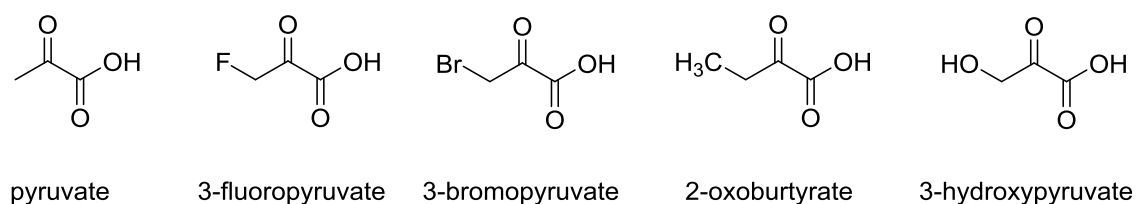


Figure 6.2 Potential ketone donor compounds for *sa*NAL. Pyruvate is the wild-type *sa*NAL ketone donor, 3-fluoropyruvate is reported to be accepted by *E. coli* NAL (Daniels *et al.*, 2014, Chokhawala *et al.*, 2007), 3-bromopyruvate, 2-oxobutyrate and 3-hydroxypyruvate are all reported to not be accepted by NAL.

6.2.2 Exploitation of interactions not found in the proteogenic amino acids

Along with using ncAAs to alter substrate specificity, it may also be possible to utilise ncAAs to engineer novel binding motifs into proteins and enzymes. Carboxylate groups are a commonly found functional group in biological compounds. By use of ncAAs, it may be possible to incorporate a thiourea containing side chain which would interact with the carboxylate group (Fitzmaurice *et al.*, 2007). By introducing interactions like the carboxylate-thiourea binding motifs it may be possible to alter substrate binding.

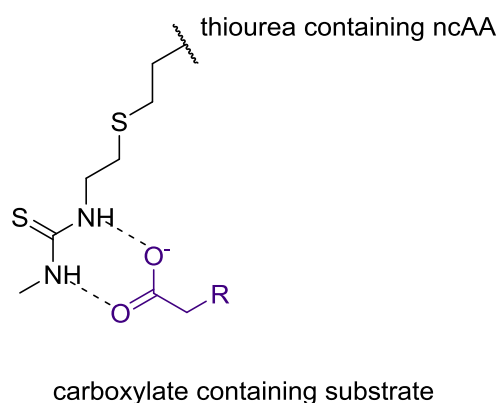


Figure 6.3 Interaction of a thiourea containing ncAA with a carboxylic acid containing substrate.

6.2.3 Alteration of catalysis

By utilising non proteogenic side chains it may be possible to expand the catalysis of enzymes. This principle has been recently demonstrated by the production of an artificial metalloenzyme, which by incorporation of the metal-binding ncAA (2,2'-bipyridin-5yl)alanine (BpyAla) catalyses an asymmetric Friedel-Crafts reaction (Drienovska *et al.*, 2015). This evolution of catalytic activity could be further extended to catalysis not generally found in enzymes.

NAL catalyses reactions through a reactive enamine species; however by incorporation of methyl-lysine analogues it may also be possible to access iminium ion chemistry. Iminium ion catalysis is exploited by reactions such as Diels-Alder reactions and Michael additions. By reaction of a methyl-lysine analogue with a novel synthesised α,β -unsaturated carbonyl compound, an iminium ion would be produced which could then be reacted with a nucleophile or diene to undergo a Michael addition or Diels-Alder reaction respectively. *sa*NAL would

provide an excellent scaffold for this work, as it has been optimised for the incorporation of the ncAAs that would be needed for the production of the iminium ion. Also *sa*NAL has an $(\alpha/\beta)_8$ -barrel fold, this fold is the most common throughout proteins and has been found in all except one of the enzyme classes; therefore it is highly versatile fold (Henn-Sax *et al.*, 2001). This $(\alpha/\beta)_8$ -barrel, fold found in many aldolases, has been shown to be highly malleable for catalysis. For example directed evolution studies of 2-keto-3-deoxy-6-phosphogluconate aldolase, repositioned the catalytic lysine within the active site of the enzyme, which produced an active variant with an altered substrate specificity (Wymer *et al.*, 2001). This repositioning of the catalytic lysine has also been demonstrated in a computationally designed retro aldolase, where directed evolution studies on the designed enzyme relocated the catalytic lysine which resulted in a more active variant (Giger *et al.*, 2013). These examples show the catalytic malleability of aldolases, and so an aldolase may provide an ideal scaffold for engineering iminium ion catalysis.

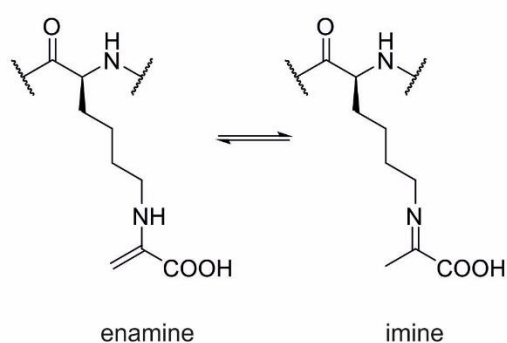
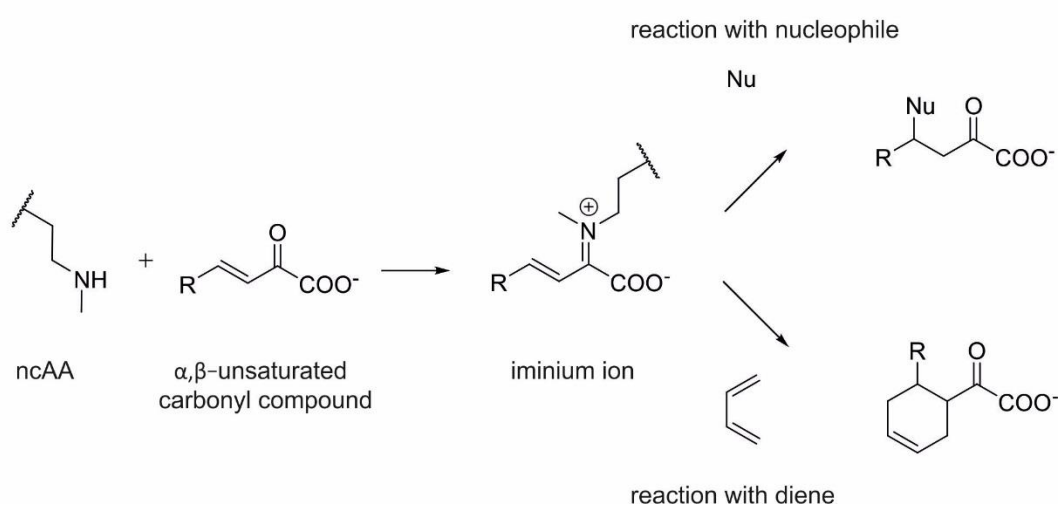
A**B**

Figure 6.4 Schematic of iminium ion catalysis. Panel A shows the enamine and imine formed during Schiff base formation between a lysine residue and pyruvate. Panel B shows potential iminium ion catalysis, reaction of the nCAA methyl-lysine analogue and an α,β -unsaturated carbonyl compound produces the iminium ion which can subsequently be reacted with nucleophiles or dienes.

6.3 Concluding remarks

Through this research it has been possible to use a chemical modification method to incorporate varied non-canonical side chain chemistries, individually throughout *saNAL*. It has also been possible to use this approach to alter the activity of the enzyme and produce an nCAA-containing enzyme with an increased activity towards a shortened substrate. The modified enzyme shows increased activity with the shortened substrate over that of the wild-type enzyme with the shortened substrate and also the wild-type substrate. The modified side chain outperforms any of the proteogenic amino acids for activity, with the shortened

substrate, when inserted at the same position within the protein. This work shows that by diversifying the side chains available to the protein engineer, it may be possible to develop activities above and beyond what would be achievable with the traditional proteogenic 'tool-kit'. This research provides an excellent platform, with the facile expansion of the amino acid alphabet, to extend the use of ncAAs to further alter enzyme activities and potentially develop novel catalysis in enzymes.

Chapter 7 Appendix

pH	Wild-type <i>sa</i> NAL			K165- γ -thialysine <i>sa</i> NAL		
	k_{cat} (min^{-1})	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ ($\text{min}^{-1} \text{mM}^{-1}$)	k_{cat} (min^{-1})	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ ($\text{min}^{-1} \text{mM}^{-1}$)
5.0	80 \pm 3	4.5 \pm 0.4	18 \pm 2	19 \pm 1.6	4.5 \pm 0.98	4.2 \pm 0.99
6.0	240 \pm 9	4.0 \pm 0.4	60 \pm 6	31 \pm 2.2	3.0 \pm 0.61	10 \pm 2.2
6.2	250 \pm 6	3.1 \pm 0.2	81 \pm 5	28 \pm 0.96	2.3 \pm 0.25	12 \pm 1.4
6.4	230 \pm 7	2.3 \pm 0.2	100 \pm 8	28 \pm 0.81	1.9 \pm 0.19	15 \pm 1.5
6.6	250 \pm 8	2.7 \pm 0.3	93 \pm 9	32 \pm 0.86	1.4 \pm 0.12	23 \pm 2.1
6.8	260 \pm 6	2.4 \pm 0.2	108 \pm 7	29 \pm 0.8	0.92 \pm 0.09	32 \pm 3.11
7.0	240 \pm 7	2.5 \pm 0.2	96 \pm 8	29 \pm 0.55	0.99 \pm 0.07	29 \pm 2.0
7.2	250 \pm 5	2.4 \pm 0.2	104 \pm 7	26 \pm 1.9	1.0 \pm 0.30	26 \pm 1.8
7.4	250 \pm 5	2.2 \pm 0.1	114 \pm 6	26 \pm 0.85	1.4 \pm 0.2	19 \pm 2.3
7.6	200 \pm 7	1.9 \pm 0.2	105 \pm 12	25 \pm 0.5	1.6 \pm 0.1	16 \pm 1.0
7.8	230 \pm 6	2.3 \pm 0.2	100 \pm 7	23 \pm 0.4	1.4 \pm 0.09	16 \pm 1.1
8.0	260 \pm 13	2.6 \pm 0.3	100 \pm 10	25 \pm 1.8	2.5 \pm 0.4	10 \pm 1.8
9.0	57 \pm 4	1.5 \pm 0.3	38 \pm 8	8.6 \pm 0.79	3.2 \pm 0.7	2.7 \pm 0.61

Table 7.1 Kinetic parameters for wild-type and K165- γ -thialysine *sa*NAL at various pH values.

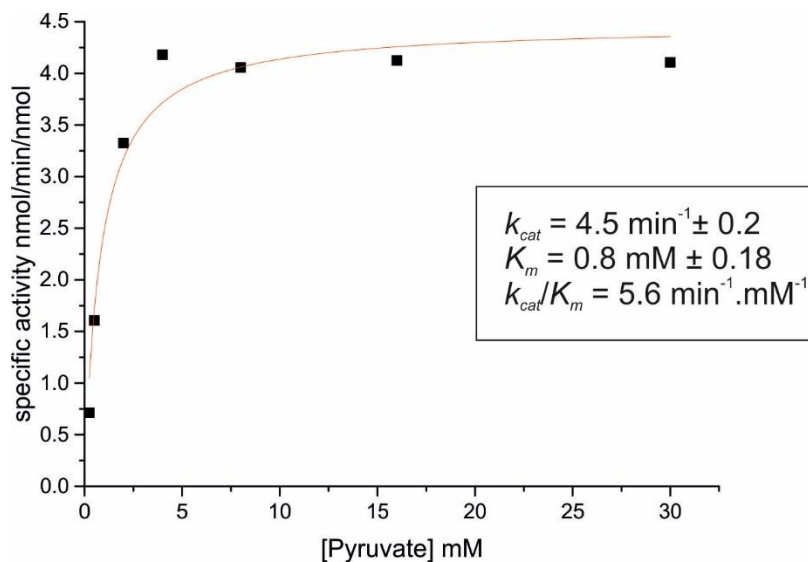


Figure 7.1 Raw data for the aldol condensation of erythrose and pyruvate to form deoxyheptulosonic acid determined by TBA assay (Section 2.4.16.3). Absorbance data at 550 nm was collected and converted to specific activity using a standard curve of sialic acid to determine the amount of product formed.

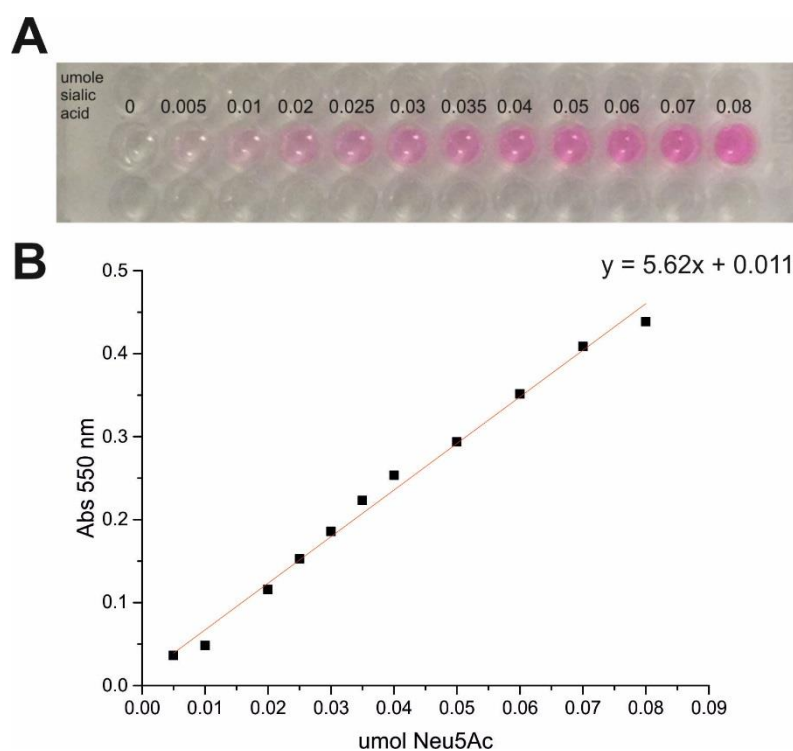


Figure 7.2 Standard curve of Neu5Ac. A shows an image of samples of known concentration of Neu5Ac which have been assayed by TBA assay (Section 2.4.16.2). B shows the absorbance data at 550nm of the assayed samples

Oligonucleotide primers used in this work

Oligonucleotide primer	Sequence (5'-3')
saNAL-F190A-For	-C CCA GAC AAA TTA ATC TTA TCT GGC GCT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190A-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC AGC GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190D-For	-C CCA GAC AAA TTA ATC TTA TCT GGC GAT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190D-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC ATC GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190E-For	-TTC CCA GAC AAA TTA ATC TTA TCT GGC GAG GAT GAA ATG TTA GTT CAA GCT ACG ATT-
saNAL-F190E-Rev	-AAT CGT AGC TTG AAC TAA CAT TTC ATC CTC GCC AGA TAA GAT TAA TTT GTC TGG GAA-
saNAL-F190G-For	-C CCA GAC AAA TTA ATC TTA TCT GGC GGT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190G-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC ACC GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190H-For	-C CCA GAC AAA TTA ATC TTA TCT GGC CAT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190H-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC ATG GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190I-For	-CA GAC AAA TTA ATC TTA TCT GGC ATT GAT GAA ATG TTA GTT CAA GCT-
saNAL-F190I-Rev	-AGC TTG AAC TAA CAT TTC ATC AAT GCC AGA TAA GAT TAA TTT GTC TG-
saNAL-F190K-For	-TTC CCA GAC AAA TTA ATC TTA TCT GGC AAG GAT GAA ATG TTA GTT CAA GCT ACG ATT-
saNAL-F190K-Rev	-AAT CGT AGC TTG AAC TAA CAT TTC ATC CTT GCC AGA TAA GAT TAA TTT GTC TGG GAA-
saNAL-F190L-For	-GAC AAA TTA ATC TTA TCT GGC TTA GAT GAA ATG TTA GTT CAA GCT AC-
saNAL-F190L-Rev	-GT AGC TTG AAC TAA CAT TTC ATC TAA GCC AGA TAA GAT TAA TTT GTC-
saNAL-F190M-For	-CCA GAC AAA TTA ATC TTA TCT GGC ATG GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190M-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC CAT GCC AGA TAA GAT TAA TTT GTC TGG-
saNAL-F190N-For	-C CCA GAC AAA TTA ATC TTA TCT GGC AAT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190N-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC ATT GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190P-For	-C CCA GAC AAA TTA ATC TTA TCT GGC CCT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190P-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC AGG GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190Q-For	-TTC CCA GAC AAA TTA ATC TTA TCT GGC CAG GAT GAA ATG TTA GTT CAA GCT ACG ATT-
saNAL-F190Q-Rev	-AAT CGT AGC TTG AAC TAA CAT TTC ATC CTG GCC AGA TAA GAT TAA TTT GTC TGG GAA-
saNAL-F190R-For	-C CCA GAC AAA TTA ATC TTA TCT GGC CGT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190R-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC ACG GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190S-For	-C CCA GAC AAA TTA ATC TTA TCT GGC AGT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190S-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC ACT GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190T-For	-C CCA GAC AAA TTA ATC TTA TCT GGC ACT GAT GAA ATG TTA GTT CAA GCT ACG-
saNAL-F190T-Rev	-CGT AGC TTG AAC TAA CAT TTC ATC AGT GCC AGA TAA GAT TAA TTT GTC TGG G-
saNAL-F190V-For	-CA GAC AAA TTA ATC TTA TCT GGC GTT GAT GAA ATG TTA GTT CAA GCT-
saNAL-F190V-Rev	-AGC TTG AAC TAA CAT TTC ATC AAC GCC AGA TAA GAT TAA TTT GTC TG-
saNAL-F190W-For	-CCA GAC AAA TTA ATC TTA TCT GGC TGG GAT GAA ATG TTA GTT CAA GCT ACG A-
saNAL-F190W-Rev	-T CGT AGC TTG AAC TAA CAT TTC ATC CCA GCC AGA TAA GAT TAA TTT GTC TGG-
saNAL-F190Y-For	-CCA GAC AAA TTA ATC TTA TCT GGC TAT GAT GAA ATG TTA GTT CAA G-
saNAL-F190Y-Rev	-C TTG AAC TAA CAT TTC ATC ATA GCC AGA TAA GAT TAA TTT GTC TGG-
saNAL-I251R-For	-CAG TGT TAT CAA TGG GGA GGT ATC CAA CAT TGA AAG-
saNAL-I251R-Rev	-CTT TCA ATG TTG GAT ACC TCC CCA TTG ATA ACA CTG-

References

- Acevedo-Rocha, C. G., Hoesl, M. G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., Antranikian, G. & Budisa, N. (2013) Non-Canonical Amino Acids as a Useful Synthetic Biological Tool for Lipase-Catalysed Reactions in Hostile Environments. *Catal. Sci. Technol.*, 3, 1198-1201.
- Ammon, H. L., Prasad, S. M. & Gerlt, J. A. (1991) Structure of Thialysine Hydrochloride *Acta Crystallogr. Sect. C-Cryst. Struct. Commun.*, 47, 1476-1478.
- Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S. & Schultz, P. G. (2004) An Expanded Genetic Code with a Functional Quadruplet Codon. *Proc. Natl. Acad. Sci. U. S. A.*, 101, 7566-7571.
- Appel, M. J. & Bertozzi, C. R. (2015) Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.*, 10, 72-84.
- Arnér, E. S. J. (2010) Selenoproteins—What Unique Properties Can Arise with Selenocysteine in Place of Cysteine? *Exp. Cell. Res.*, 316, 1296-1303.
- Auge, C. & Gautheron-le Narvor, C. (1997) Preparative Carbohydrate Chemistry. IN Hanessian, S. (Ed.). New York, Marcel Dekker, INC.
- Baker, P. & Seah, S. Y. K. (2012) Rational Approaches for Engineering Novel Functionalities in Carbon-Carbon Bond Forming Enzymes. *Comput. Struct. Biotechnol. J.*, 2, 1-10.
- Bale, S. & Ealick, S. E. (2010) Structural Biology of S-Adenosylmethionine Decarboxylase. *Amino acids*, 38, 451-60.
- Barbosa, J., Smith, B. J., DeGori, R., Ooi, H. C., Marcuccio, S. M., Campi, E. M., Jackson, W. R., Brossmer, R., Sommer, M. & Lawrence, M. C. (2000) Active Site Modulation in the N-Acetylneuraminase Lyase Sub-Family as Revealed by the Structure of the Inhibitor-Complexed *Haemophilus Influenzae* Enzyme. *J. Mol. Biol.*, 303, 405-421.
- Bas, D. C., Rogers, D. M. & Jensen, J. H. (2008) Very Fast Prediction and Rationalization of Pk(a) Values for Protein-Ligand Complexes. *Protein. Struct. Funct. Genet.*, 73, 765-783.
- Basle, E., Joubert, N. & Pucheault, M. (2010) Protein Chemical Modification on Endogenous Amino Acids. *Chem. Biol.*, 17, 213-227.
- Bischoff, R. & Schlüter, H. (2012) Amino Acids: Chemistry, Functionality and Selected Non-Enzymatic Post-Translational Modifications. *J. Proteomics*, 75, 2275-2296.
- Blickling, S., Renner, C., Laber, B., Pohlenz, H.-D., Holak, T. A. & Huber, R. (1997) Reaction Mechanism of *Escherichia Coli* Dihydrodipicolinate Synthase Investigated by X-Ray Crystallography and Nmr Spectroscopy. *Biochemistry*, 36, 24-33.
- Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K. & Krzycki, J. A. (2004) Direct Charging of Trn^{ca} with Pyrrolysine in Vitro and in Vivo. *Nature*, 431, 333-335.
- Bochar, D. A., Taberner, L., Stauffacher, C. V. & Rodwell, V. W. (1999) Aminoethylcysteine Can Replace the Function of the Essential Active Site Lysine of Pseudomonas Mevalonii 3-Hydroxy-3-Methylglutaryl Coenzyme a Reductase. *Biochemistry*, 38, 8879-83.
- Bojarova, P. & Williams, S. J. (2008) Sulfotransferases, Sulfatases and Formylglycine-Generating Enzymes: A Sulfation Fascination. *Curr. Opin. Chem. Biol.*, 12, 573-581.
- Bolt, A., Berry, A. & Nelson, A. (2008) Directed Evolution of Aldolases for Exploitation in Synthetic Organic Chemistry. *Arch. Biochem. Biophys.*, 474, 318-330.
- Broderick, J. B. (2001) Coenzymes and Cofactors. *Els.* John Wiley & Sons, Ltd.
- Buchanan, C. L., Connaris, H., Danson, M. J., Reeve, C. D. & Hough, D. W. (1999) An Extremely Thermostable Aldolase from *Sulfolobus Solfataricus* with Specificity for Non-Phosphorylated Substrates. *Biochem. J.*, 343 Pt 3, 563-70.

- Budisa, N., Pal, P. P., Alefelder, S., Birle, P., Krywcun, T., Rubini, M., Wenger, W., Bae, J. H. & Steiner, T. (2004) Probing the Role of Tryptophans in *Aequorea Victoria* Green Fluorescent Proteins with an Expanded Genetic Code. *Biol. Chem.*, 385, 191-202.
- Caban, K. & Copeland, P. R. (2006) Size Matters: A View of Selenocysteine Incorporation from the Ribosome. *Cell Mol. Life Sci.*, 63, 73-81.
- Campeotto, I., Bolt, A. H., Harman, T. A., Dennis, C., Trinh, C. H., Phillips, S. E. V., Nelson, A., Pearson, A. R. & Berry, A. (2010) Structural Insights into Substrate Specificity in Variants of N-Acetylneuraminic Acid Lyase Produced by Directed Evolution. *J. Mol. Biol.*, 404, 56-69.
- Campeotto, I., Carr, S. B., Trinh, C. H., Nelson, A. S., Berry, A., Phillips, S. E. V. & Pearson, A. R. (2009) Structure of an Escherichia Coli N-Acetyl-D-Neuraminic Acid Lyase Mutant, E192n, in Complex with Pyruvate at 1.45 Angstrom Resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 65, 1088-1090.
- Chalker, J. M., Bernardes, G. J. L., Lin, Y. A. & Davis, B. G. (2009) Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem.-Asian J.*, 4, 630-640.
- Chalker, J. M., Gunnoo, S. B., Boutureira, O., Gerstberger, S. C., Fernandez-Gonzalez, M., Bernardes, G. J. L., Griffin, L., Hailu, H., Schofield, C. J. & Davis, B. G. (2011) Methods for Converting Cysteine to Dehydroalanine on Peptides and Proteins. *Chem. Sci.*, 2, 1666-1676.
- Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. (2012) Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem. Int. Ed.*, 51, 1835-1839.
- Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W. & Harrison, P. R. (1986) The Structure of the Mouse Glutathione-Peroxidase Gene - the Selenocysteine in the Active-Site Is Encoded by the Termination Codon, Tga. *EMBO J.*, 5, 1221-1227.
- Chen, G., Heim, A., Riether, D., Yee, D., Milgrom, Y., Gawinowicz, M. A. & Sames, D. (2003) Reactivity of Functional Groups on the Protein Surface: Development of Epoxide Probes for Protein Labeling. *J. Am. Chem. Soc.*, 125, 8130-8133.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010) Molprobity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 12-21.
- Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z. & Schultz, P. G. (2003) An Expanded Eukaryotic Genetic Code. *Science*, 301, 964-7.
- Chokhawala, H. A., Cao, H., Yu, H. & Chen, X. (2007) Enzymatic Synthesis of Fluorinated Mechanistic Probes for Sialidases and Sialyltransferases. *J. Am. Chem. Soc.*, 129, 10630-10631.
- Chou, C.-Y., Ko, T.-P., Wu, K.-J., Huang, K.-F., Lin, C.-H., Wong, C.-H. & Wang, A. H. J. (2011) Modulation of Substrate Specificities of D-Sialic Acid Aldolase through Single Mutations of Val-251. *J. Biol. Chem.*, 286, 14057-14064.
- Christianson, C. V., Montavon, T. J., Festin, G. M., Cooke, H. A., Shen, B. & Bruner, S. D. (2007) The Mechanism of Mio-Based Aminomutases in Beta-Amino Acid Biosynthesis. *J. Am. Chem. Soc.*, 129, 15744-+.
- Cirino, P. C., Tang, Y., Takahashi, K., Tirrell, D. A. & Arnold, F. H. (2003) Global Incorporation of Norleucine in Place of Methionine in Cytochrome P450 Bm-3 Heme Domain Increases Peroxygenase Activity. *Biotechnol. Bioeng.*, 83, 729-34.
- Contestabile, R., Paiardini, A., Pascarella, S., di Salvo, M. L., D'Aguanno, S. & Bossa, F. (2001) L-Threonine Aldolase, Serine Hydroxymethyltransferase and Fungal Alanine Racemase. A

- Subgroup of Strictly Related Enzymes Specialized for Different Functions. *Eur. J. Biochem.*, 268, 6508-25.
- Cowie, D. B. & Cohen, G. N. (1957) Biosynthesis by *Escherichia Coli* of Active Altered Proteins Containing Selenium Instead of Sulfur. *Biochim. Biophys. Acta*, 26, 252-61.
- Daniels, A. D., Campeotto, I., van der Kamp, M. W., Bolt, A. H., Trinh, C. H., Phillips, S. E. V., Pearson, A. R., Nelson, A., Mulholland, A. J. & Berry, A. (2014) Reaction Mechanism of *N*-Acetylneuraminic Acid Lyase Revealed by a Combination of Crystallography, Qm/Mm Simulation, and Mutagenesis. *ACS Chem. Biol.*, 9, 1025-1032.
- Davidson, V. L. (2011) Generation of Protein-Derived Redox Cofactors by Posttranslational Modification. *Mol. Biosyst.*, 7, 29-37.
- Davis, L. & Chin, J. W. (2012) Designer Proteins: Applications of Genetic Code Expansion in Cell Biology. *Nat. Rev. Mol. Cell Biol.*, 13, 168-182.
- Dean, S. M., Greenberg, W. A. & Wong, C. H. (2007) Recent Advances in Aldolase-Catalyzed Asymmetric Synthesis. *Adv. Synth. Catal.*, 349, 1308-1320.
- DeSantis, G. & Jones, J. B. (1999) Chemical Modification of Enzymes for Enhanced Functionality. *Curr. Opin. Biotechnol.*, 10, 324-330.
- Diaz-Rodriguez, A. & Davis, B. G. (2011) Chemical Modification in the Creation of Novel Biocatalysts. *Curr. Opin. Chem. Biol.*, 15, 211-219.
- Dierks, T., Dickmanns, A., Preusser-Kunze, A., Schmidt, B., Mariappan, M., von Figura, K., Ficner, R. & Rudolph, M. G. (2005) Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme. *Cell*, 121, 541-552.
- Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M. A. & von Figura, K. (1998) Posttranslational Formation of Formylglycine in Prokaryotic Sulfatases by Modification of Either Cysteine or Serine. *J. Biol. Chem.*, 273, 25560-25564.
- Dondoni, A., Marra, A. & Merino, P. (1994) Installation of the Pyruvate Unit in Glycidic Aldehydes Via a Wittig Olefination Michael Addition Sequence Utilizing a Thiazole-Armed Carbonyl Ylid - a New Stereoselective Route to 3-Deoxy-2-Ulosonic Acids and the Total Synthesis of Dah, Kdn and 4-Epi-Kdn. *J. Am. Chem. Soc.*, 116, 3324-3336.
- Drienovska, I., Rioz-Martinez, A., Draksharapu, A. & Roelfes, G. (2015) Novel Artificial Metalloenzymes by in Vivo Incorporation of Metal-Binding Unnatural Amino Acids. *Chem. Sci.*, 6, 770-776.
- Driscoll, D. M. & Copeland, P. R. (2003) Mechanism and Regulation of Selenoprotein Synthesis. *Annu. Rev. Nutr.*, 23, 17-40.
- DuBois, J. L. & Klinman, J. P. (2005) Mechanism of Post-Translational Quinone Formation in Copper Amine Oxidases and Its Relationship to the Catalytic Turnover. *Arch. Biochem. Biophys.*, 433, 255-265.
- Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. (2015) Designing Logical Codon Reassignment - Expanding the Chemistry in Biology. *Chem. Sci*, 6, 50-69.
- Edwards, D. R., Lohman, D. C. & Wolfenden, R. (2012) Catalytic Proficiency: The Extreme Case of S-O Cleaving Sulfatases. *J. Am. Chem. Soc.*, 134, 525-531.
- Ellis, K. J. & Morrison, J. F. (1982) Buffers of Constant Ionic-Strength for Studying pH-Dependent Processes. *Methods in Enzymology*, 87, 405-426.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 486-501.
- Enders, D. & Gasperi, T. (2007) Proline Organocatalysis as a New Tool for the Asymmetric Synthesis of Ulosonic Acid Precursors. *Chem. Commun.*, 88-90.
- Engh, R. A. & Huber, R. (1991) Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement. *Acta Crystallogr. Sect. A*, 47, 392-400.

- England, P. M. (2004) Unnatural Amino Acid Mutagenesis: A Precise Tool for Probing Protein Structure and Function. *Biochemistry*, 43, 11623-11629.
- Evans, P. (2006) Scaling and Assessment of Data Quality. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 72-82.
- Fahrney, D. E. & Gold, A. M. (1963) Sulfonyl Fluorides as Inhibitors of Esterases. I. Rates of Reaction with Acetylcholinesterase, A-Chymotrypsin, and Trypsin. *J. Am. Chem. Soc.*, 85, 997-1000.
- Fesko, K. & Gruber-Khadjawi, M. (2013) Biocatalytic Methods for Cc Bond Formation. *ChemCatChem*, 5, 1248-1272.
- Fitz, W., Schwark, J.-R. & Wong, C.-H. (1995) Aldotetroses and C(3)-Modified Aldohexoses as Substrates for *N*-Acetylneuraminic Acid Aldolase: A Model for the Explanation of the Normal and the Inversed Stereoselectivity. *J. Org. Chem.*, 60, 3663-3670.
- Fitzmaurice, R. J., Gaggini, F., Srinivasan, N. & Kilburn, J. D. (2007) Carboxylate Binding in Polar Solvents Using Pyridylguanidinium Salts. *Org Biomol Chem*, 5, 1706-14.
- Flohe, L., Gunzler, W. A. & Schock, H. H. (1973) Glutathione Peroxidase - Selenoenzyme. *FEBS Lett.*, 32, 132-134.
- Gamblin, D. P., Garnier, P., van Kasteren, S., Oldham, N. J., Fairbanks, A. J. & Davis, B. G. (2004) Glyco-Ses: Selenenylsulfide-Mediated Protein Glycoconjugation—a New Strategy in Post-Translational Modification. *Angew. Chem. Int. Ed.*, 43, 828-833.
- Gaston, M. A., Zhang, L., Green-Church, K. B. & Krzycki, J. A. (2011) The Complete Biosynthesis of the Genetically Encoded Amino Acid Pyrrolysine from Lysine. *Nature*, 471, 647-650.
- Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N. & Hilvert, D. (2013) Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.*, 9, 494-8.
- Gijzen, H. J., Qiao, L., Fitz, W. & Wong, C. H. (1996) Recent Advances in the Chemoenzymatic Synthesis of Carbohydrates and Carbohydrate Mimetics. *Chem. Rev.*, 96, 443-474.
- Gloss, L. M. & Kirsch, J. F. (1995a) Decreasing the Basicity of the Active-Site Base, Lys-258, of *Escherichia Coli* Aspartate-Aminotransferase by Replacement with Gamma-Thialysine. *Biochemistry*, 34, 3990-3998.
- Gloss, L. M. & Kirsch, J. F. (1995b) Examining the Structural and Chemical Flexibility of the Active Site Base, Lys-258, of *Escherichia Coli* Aspartate Aminotransferase by Replacement with Unnatural Amino Acids. *Biochemistry*, 34, 12323-32.
- Greenberg, W. A., Varvak, A., Hanson, S. R., Wong, K., Huang, H., Chen, P. & Burk, M. J. (2004) Development of an Efficient, Scalable, Aldolase-Catalyzed Process for Enantioselective Synthesis of Statin Intermediates. *Proc. Natl. Acad. Sci. U. S. A.*, 101, 5788-93.
- Groher, A. & Hoelsch, K. (2012) Mechanistic Model for the Synthesis of *N*-Acetylneuraminic Acid Using *N*-Acetylneuraminic Lyase from *Escherichia Coli* K12. *J. Mol. Catal. B: Enzym.*, 83, 1-7.
- Guo, J. T., Wang, J. Y., Lee, J. S. & Schultz, P. G. (2008) Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chem. Int. Ed.*, 47, 6399-6401.
- Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A. & Chan, M. K. (2002) A New Uag-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science*, 296, 1462-1466.
- Hao, Z., Hong, S., Chen, X. & Chen, P. R. (2011) Introducing Bioorthogonal Functionalities into Proteins in Living Cells. *Acc. Chem. Res.*, 44, 742-51.
- Harris, T. K. & Turner, G. J. (2002) Structural Basis of Perturbed Pk(a) Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life*, 53, 85-98.
- Hatfield, D. L. & Gladyshev, V. N. (2002) How Selenium Has Altered Our Understanding of the Genetic Code. *Mol. Cell. Biol.*, 22, 3565-3576.

- Hegazy, U. M., Hellman, U. & Mannervik, B. (2006) Replacement Surgery with Unnatural Amino Acids in the Lock-and-Key Joint of Glutathione Transferase Subunits. *Chem. Biol.*, 13, 929-936.
- Hegazy, U. M., Mannervik, B. & Stenberg, G. (2004) Functional Role of the Lock and Key Motif at the Subunit Interface of Glutathione Transferase P1-1. *J. Biol. Chem.*, 279, 9586-96.
- Hegazy, U. M., Musdal, Y. & Mannervik, B. (2013) Hidden Allostery in Human Glutathione Transferase P1-1 Unveiled by Unnatural Amino Acid Substitutions and Inhibition Studies. *J. Mol. Biol.*, 425, 1509-14.
- Heinrikson, R. L. (1966) On the Alkylation of Amino Acid Residues at the Active Site of Ribonuclease. *J. Biol. Chem.*, 241, 1393-1405.
- Hekking, K. F. W., Moelands, M. A. H., van Delft, F. L. & Rutjes, F. P. J. T. (2006) An in-Depth Study on Ring-Closing Metathesis of Carbohydrate-Derived Alpha-Alkoxyacrylates: Efficient Syntheses of Dah, Kdo, and 2-Deoxy-Beta-Kdo. *J. Org. Chem.*, 71, 6444-6450.
- Henn-Sax, M., Hocker, B., Wilmanns, M. & Sterner, R. (2001) Divergent Evolution of (Betaalpha)8-Barrel Enzymes. *Biol Chem*, 382, 1315-20.
- Hermann, P. & Lemke, K. (1968) Ionization and Stability Constants of Copper(2) Complexes of Some Amino Acids and Their Sulfur-Containing Analogs *Hoppe-Seyler's Z. Physiol. Chem.*, 349, 390-&.
- Hoesl, M. G., Acevedo-Rocha, C. G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., Budisa, N. & Antranikian, G. (2011) Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem*, 3, 213-221.
- Hopkins, C. E., Hernandez, G., Lee, J. P. & Tolan, D. R. (2005) Aminoethylation in Model Peptides Reveals Conditions for Maximizing Thiol Specificity. *Arch. Biochem, Biophys.*, 443, 1-10.
- Hopkins, C. E., O'Connor, P. B., Allen, K. N., Costello, C. E. & Tolan, D. R. (2002) Chemical-Modification Rescue Assessed by Mass Spectrometry Demonstrates That Gamma-Thia-Lysine Yields the Same Activity as Lysine in Aldolase. *Protein Sci.*, 11, 1591-9.
- Hsu, C. C., Hong, Z., Wada, M., Franke, D. & Wong, C. H. (2005) Directed Evolution of D-Sialic Acid Aldolase to L-3-Deoxy-Manno-2-Octulosonic Acid (L-Kdo) Aldolase. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 9122-6.
- Huang, R., Holbert, M. A., Tarrant, M. K., Curtet, S., Colquhoun, D. R., Dancy, B. M., Dancy, B. C., Hwang, Y., Tang, Y., Meeth, K., Marmorstein, R., Cole, R. N., Khochbin, S. & Cole, P. A. (2010) Site-Specific Introduction of an Acetyl-Lysine Mimic into Peptides and Proteins by Cysteine Alkylation. *J. Am. Chem. Soc.*, 132, 9986-9987.
- Isom, D. G., Castaneda, C. A., Cannon, B. R. & Garcia-Moreno, B. E. (2011) Large Shifts in Pk(a) Values of Lysine Residues Buried inside a Protein. *Proc. Natl. Acad. Sci. U. S. A.*, 108, 5260-5265.
- Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G. & Colman, P. M. (1994) The 3-Dimensional Structure of *N*-Acetylneuraminase Lyase for *Escherichia Coli* *Structure*, 2, 361-369.
- Jackson, J. C., Duffy, S. P., Hess, K. R. & Mehl, R. A. (2006) Improving Nature's Enzyme Active Site with Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.*, 128, 11124-11127.
- Ji, W., Sun, W., Feng, J., Song, T., Zhang, D., Ouyang, P., Gu, Z. & Xie, J. (2015) Characterization of a Novel *N*-Acetylneuraminic Acid Lyase Favoring *N*-Acetylneuraminic Acid Synthesis. *Sci. Rep.*, 5.
- Joerger, A. C., Mayer, S. & Fersht, A. R. (2003) Mimicking Natural Evolution in Vitro: An *N*-Acetylneuraminase Lyase Mutant with an Increased Dihydrodipicolinate Synthase Activity. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 5694-9.
- Johansson, L., Gafvelin, G. & Arner, E. S. (2005) Selenocysteine in Proteins-Properties and Biotechnological Use. *Biochim. Biophys. Acta.*, 1726, 1-13.

- Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. (2010) Residue-Specific Incorporation of Non-Canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin.Chem. Biol.*, 14, 774-780.
- Kantardjieff, K. A. & Rupp, B. (2003) Matthews Coefficient Probabilities: Improved Estimates for Unit Cell Contents of Proteins, DNA, and Protein–Nucleic Acid Complex Crystals. *Protein Sci.*, 12, 1865-1871.
- Kiick, K. L., van Hest, J. C. M. & Tirrell, D. A. (2000) Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-Trna Synthetase Activity of a Bacterial Expression Host. *Angew. Chem. Int. Ed.*, 112, 2232-2236.
- Kim, H. S., Ha, S. H., Sethaphong, L., Koo, Y. M. & Yingling, Y. G. (2014) The Relationship between Enhanced Enzyme Activity and Structural Dynamics in Ionic Liquids: A Combined Computational and Experimental Study. *Phys. Chem. Chem. Phys.*, 16, 2944-53.
- Kim, H. Y. & Gladyshev, V. N. (2005) Different Catalytic Mechanisms in Mammalian Selenocysteine- and Cysteine-Containing Methionine-R-Sulfoxide Reductases. *PLoS Biol.*, 3, e375.
- Kim, M. J., Hennen, W. J., Sweers, H. M. & Wong, C. H. (1988) Enzymes in Carbohydrate Synthesis: *N*-Acetylneuraminic Acid Aldolase Catalyzed Reactions and Preparation of *N*-Acetyl-2-Deoxy-D-Neuraminic Acid Derivatives. *J. Am. Chem. Soc.*, 110, 6481-6486.
- Lad, C., Williams, N. H. & Wolfenden, R. (2003) The Rate of Hydrolysis of Phosphomonoester Dianions and the Exceptional Catalytic Proficiencies of Protein and Inositol Phosphatases. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 5607-10.
- Lang, A. & Klinman, J. P. (2001) Quinone Cofactors. *Els.* John Wiley & Sons, Ltd.
- Lang, K. & Chin, J. W. (2014) Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.*, 114, 4764-4806.
- Lawrence, M. C., Barbosa, J. A. R. G., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C. & Marcuccio, S. M. (1997) Structure and Mechanism of a Sub-Family of Enzymes Related to *N*-Acetylneuraminic Lyase1. *J. Mol. Biol.*, 266, 381-399.
- Leferink, N. G. H., Antonyuk, S. V., Houwman, J. A., Scrutton, N. S., Eady, R. R. & Hasnain, S. S. (2014) Impact of Residues Remote from the Catalytic Centre on Enzyme Catalysis of Copper Nitrite Reductase. *Nat. Commun.*, 5.
- Leslie, A. G. W. (2006) The Integration of Macromolecular Diffraction Data. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 48-57.
- Lin, J., Peng, T., Jiang, L., Ni, J. Z., Liu, Q., Chen, L. & Zhang, Y. (2015) Comparative Genomics Reveals New Candidate Genes Involved in Selenium Metabolism in Prokaryotes. *Genome Biol. Evol.*, 7, 664-76.
- Link, A. J., Mock, M. L. & Tirrell, D. A. (2003) Non-Canonical Amino Acids in Protein Engineering. *Curr. Opin. Biotechnol.*, 14, 603-609.
- Liu, C. C., Mack, A. V., Brustad, E. M., Mills, J. H., Groff, D., Smider, V. V. & Schultz, P. G. (2009) Evolution of Proteins with Genetically Encoded “Chemical Warheads”. *J Am Chem Soc*, 131, 9616-9617.
- Liu, C. C. & Schultz, P. G. (2010) Adding New Chemistries to the Genetic Code. IN Kornberg, R. D., Raetz, C. R. H., Rothman, J. E. & Thorner, J. W. (Eds.) *Annual Review of Biochemistry, Vol 79*. Palo Alto, Annual Reviews.
- Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003) Structure Validation by C Alpha Geometry: Phi,Psi and C Beta Deviation. *Protein. Struct. Funct. Genet.*, 50, 437-450.
- Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K. & Saenger, W. (1998) Crystal Structure of Human Arylsulfatase A: The Aldehyde Function and the

- Metal Ion at the Active Site Suggest a Novel Mechanism for Sulfate Ester Hydrolysis. *Biochemistry*, 37, 3654-64.
- Machajewski, T. D. & Wong, C. H. (2000) The Catalytic Asymmetric Aldol Reaction. *Angew. Chem. Int. Ed.*, 39, 1352-1374.
- Maeda, H. & Dudareva, N. (2012) The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annu. Rev. Plant Biol.*, 63, 73-105.
- Mahapatra, A., Patel, A., Soares, J. A., Larue, R. C., Zhang, J. K., Metcalf, W. W. & Krzycki, J. A. (2006) Characterization of a Methanosarcina Acetivorans Mutant Unable to Translate Uag as Pyrrolysine. *Mol. Microbiol.*, 59, 56-66.
- Mande, S. S., Sarfaty, S., Allen, M. D., Perham, R. N. & Hol, W. G. J. (1996) Protein-Protein Interactions in the Pyruvate Dehydrogenase Multienzyme Complex: Dihydrolipoamide Dehydrogenase Complexed with the Binding Domain of Dihydrolipoamide Acetyltransferase. *Structure*, 4, 277-286.
- Mann, M. & Jensen, O. N. (2003) Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.*, 21, 255-261.
- Marsh, J. J. & Lebherz, H. G. (1992) Fructose-Bisphosphate Aldolases: An Evolutionary History. *Trends Biochem. Sci.*, 17, 110-3.
- Matthews, B. W. (1968) Solvent Content of Protein Crystals. *J. Mol. Biol.*, 33, 491-7.
- McCoy, A. J. (2007) Solving Structures of Protein Complexes by Molecular Replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.*, 63, 32-41.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007) Phaser Crystallographic Software. *J. Appl. Crystallogr.*, 40, 658-674.
- Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011) Refmac5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D Biol. Crystallogr.*, 67, 355-367.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D Biol. Crystallogr.*, 53, 240-255.
- Nathani, R., Moody, P., Smith, M. E. B., Fitzmaurice, R. J. & Caddick, S. (2012) Bioconjugation of Green Fluorescent Protein Via an Unexpectedly Stable Cyclic Sulfonium Intermediate. *ChemBioChem*, 13, 1283-1285.
- Nees, S., Schauer, R. & Mayer, F. (1976) Purification and Characterization of *N*-Acetylneuraminase Lyase from *Clostridium Perfringens*. *Hoppe-Seyler's Z. Physiol. Chem*, 357, 839-53.
- Neet, K. E. & Koshland, D. E. (1966) Conversion of Serine at Active Site of Subtilisin to Cysteine - a Chemical Mutation. *Proc. Natl. Acad. Sci. U. S. A.*, 56, 1606-&.
- Neumann, H., Peak-Chew, S. Y. & Chin, J. W. (2008) Genetically Encoding *N*-Epsilon-Acetyllysine in Recombinant Proteins. *Nat. Chem. Biol.*, 4, 232-234.
- Neumann, H., Wang, K. H., Davis, L., Garcia-Alai, M. & Chin, J. W. (2010) Encoding Multiple Unnatural Amino Acids Via Evolution of a Quadruplet-Decoding Ribosome. *Nature*, 464, 441-444.
- O'Donoghue, P., Ling, J. Q., Wang, Y. S. & Soll, D. (2013) Upgrading Protein Synthesis for Synthetic Biology. *Nat. Chem. Biol.*, 9, 594-598.
- Okeley, N. M. & van der Donk, W. A. (2000) Novel Cofactors Via Post-Translational Modifications of Enzyme Active Sites. *Chem. Biol.*, 7, R159-R171.
- Oya, T., Hattori, N., Mizuno, Y., Miyata, S., Maeda, S., Osawa, T. & Uchida, K. (1999) Methylglyoxal Modification of Protein. Chemical and Immunochemical Characterization of Methylglyoxal-Arginine Adducts. *J. Biol. Chem.*, 274, 18492-502.
- Parsons, J. F., Xiao, G., Gilliland, G. L. & Armstrong, R. N. (1998) Enzymes Harboring Unnatural Amino Acids: Mechanistic and Structural Analysis of the Enhanced Catalytic Activity of

- a Glutathione Transferase Containing 5-Fluorotryptophan. *Biochemistry*, 37, 6286-6294.
- Patel, M. S. & Korotchkina, L. G. (2003) The Biochemistry of the Pyruvate Dehydrogenase Complex. *Biochem. Mol. Biol. Educ.*, 31, 5-15.
- Patel, M. S., Nemeria, N. S., Furey, W. & Jordan, F. (2014) The Pyruvate Dehydrogenase Complexes: Structure-Based Function and Regulation. *J. Biol. Chem.*, 289, 16615-16623.
- Pinsent, J. (1954) The Need for Selenite and Molybdate in the Formation of Formic Dehydrogenase by Members of the *Coli-Aerogenes* Group of Bacteria. *Biochem. J.*, 57, 10-16.
- Planas, A. & Kirsch, J. F. (1991) Reengineering the Catalytic Lysine of Aspartate-Aminotransferase by Chemical Elaboration of a Genetically Introduced Cysteine. *Biochemistry*, 30, 8268-8276.
- Polgar, L. & Bender, M. L. (1966) A New Enzyme Containing a Synthetically Formed Active Site. Thiol-Subtilisin1. *J. Am. Chem. Soc.*, 88, 3153-3154.
- Pradhan, T. K., Lin, C. C. & Mong, K. K. T. (2013) Formal Synthesis of 3-Deoxy-D-Manno-Octulosonic Acid (Kdo) and 3-Deoxy-D-Arabino-2-Heptulosonic Acid (Dah). *Synlett*, 24, 219-222.
- Ravikiran, B. & Mahalakshmi, R. (2014) Unusual Post-Translational Protein Modifications: The Benefits of Sophistication. *RSC Adv.*, 4, 33958-33974.
- Recsei, P. A., Huynh, Q. K. & Snell, E. E. (1983) Conversion of Prohistidine Decarboxylase to Histidine-Decarboxylase - Peptide-Chain Cleavage by Non-Hydrolytic Serinolysis. *Proc. Natl. Acad. Sci. U. S. A.*, 80, 973-977.
- Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. (2005) Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angewandte Chemie*, 117, 4264-4268.
- Reetz, M. T., Carballeira, J. D., Peyralans, J., Höbenreich, H., Maichele, A. & Vogel, A. (2006) Expanding the Substrate Scope of Enzymes: Combining Mutations Obtained by Casting. *Chem. Eur. J.*, 12, 6031-6038.
- Riley, W. D. & Snell, E. E. (1968) Histidine Decarboxylase of *Lactobacillus 30a*. Iv. The Presence of Covalently Bound Pyruvate as the Prosthetic Group. *Biochemistry*, 7, 3520-8.
- Romero, P. A. & Arnold, F. H. (2009) Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.*, 10, 866-876.
- Rowan, F. C., Richards, M., Bibby, R. A., Thompson, A., Bayliss, R. & Blagg, J. (2013) Insights into Aurora-a Kinase Activation Using Unnatural Amino Acids Incorporated by Chemical Modification. *ACS Chem. Biol.*, 8, 2184-2191.
- Schuttelkopf, A. W. & van Aalten, D. M. F. (2004) Prodrgr: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr. D Biol. Crystallogr.*, 60, 1355-1363.
- Schwarz, K. & Foltz, C. M. (1957) Selenium as an Integral Part of Factor-3 against Dietary Necrotic Liver Degeneration. *J. Am. Chem. Soc.*, 79, 3292-3293.
- Seo, J. & Lee, K. J. (2004) Post-Translational Modifications and Their Biological Functions: Proteomic Analysis and Systematic Approaches. *J. Biochem. Mol. Biol.*, 37, 35-44.
- Severi, E., Hood, D. W. & Thomas, G. H. (2007) Sialic Acid Utilization by Bacterial Pathogens. *Microbiol-SGM*, 153, 2817-2822.
- Shimotohno, A., Oue, S., Yano, T., Kuramitsu, S. & Kagamiyama, H. (2001) Demonstration of the Importance and Usefulness of Manipulating Non-Active-Site Residues in Protein Design. *J. Biochem.*, 129, 943-8.

- Shoulders, M. D., Satyshur, K. A., Forest, K. T. & Raines, R. T. (2010) Stereoelectronic and Steric Effects in Side Chains Preorganize a Protein Main Chain. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 559-64.
- Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J. & Shokat, K. M. (2007) The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell*, 128, 1003-1012.
- Singh-Blom, A., Hughes, R. A. & Ellington, A. D. (2014) An Amino Acid Depleted Cell-Free Protein Synthesis System for the Incorporation of Non-Canonical Amino Acid Analogs into Proteins. *J. Biotechnol.*, 178, 12-22.
- Spicer, C. D. & Davis, B. G. (2014) Selective Chemical Protein Modification. *Nat. Commun.*, 5.
- Stockbridge, R. B., Lewis, C. A., Jr., Yuan, Y. & Wolfenden, R. (2010) Impact of Temperature on the Time Required for the Establishment of Primordial Biochemistry, and for the Evolution of Enzymes. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 22102-5.
- Sukumaran, J. & Hanefeld, U. (2005) Enantioselective C-C Bond Synthesis Catalysed by Enzymes. *Chem. Soc. Rev.*, 34, 530-42.
- Tang, Y. & Tirrell, D. A. (2001) Biosynthesis of a Highly Stable Coiled-Coil Protein Containing Hexafluoroleucine in an Engineered Bacterial Host. *J. Am. Chem. Soc.*, 123, 11089-11090.
- Tedaldi, L. M., Smith, M. E. B., Nathani, R. I. & Baker, J. R. (2009) Bromomaleimides: New Reagents for the Selective and Reversible Modification of Cysteine. *Chem. Commun.*, 6583-6585.
- Timms, N., Windle, C. L., Polyakova, A., Ault, J. R., Trinh, C. H., Pearson, A. R., Nelson, A. & Berry, A. (2013) Structural Insights into the Recovery of Aldolase Activity in *N*-Acetylneuraminic Acid Lyase by Replacement of the Catalytically Active Lysine with Gamma-Thialysine by Using a Chemical Mutagenesis Strategy. *ChemBioChem*, 14, 474-81.
- Tuley, A., Wang, Y.-S., Fang, X., Kurra, Y., Rezenom, Y. H. & Liu, W. R. (2014) The Genetic Incorporation of Thirteen Novel Non-Canonical Amino Acids. *Chem. Commun.*, 50, 2673-2675.
- Turner, N. J. (2011) Ammonia Lyases and Aminomutases as Biocatalysts for the Synthesis of Alpha-Amino and Beta-Amino Acids. *Curr. Opin. Chem. Biol.*, 15, 234-240.
- Uchida, Y., Tsukada, Y. & Sugimori, T. (1984) Purification and Properties of *N*-Acetylneuraminic Lyase from *Escherichia Coli*. *J. Biochem.*, 96, 507-522.
- Ugwumba, I. N., Ozawa, K., Xu, Z. Q., Ely, F., Foo, J. L., Herlt, A. J., Coppin, C., Brown, S., Taylor, M. C., Ollis, D. L., Mander, L. N., Schenk, G., Dixon, N. E., Otting, G., Oakeshott, J. G. & Jackson, C. J. (2011) Improving a Natural Enzyme Activity through Incorporation of Unnatural Amino Acids. *J. Am. Chem. Soc.*, 133, 326-33.
- Vaaler, G. L., Recsei, P. A., Fox, J. L. & Snell, E. E. (1982) Histidine-Decarboxylase of *Lactobacillus-30a* - Comparative Sequences of the Beta-Chain from Wild-Type and Mutant Enzymes. *J. Biol. Chem.*, 257, 2770-2774.
- Vanpoelje, P. D. & Snell, E. E. (1990) Pyruvoyl-Dependent Enzymes *Annu. Rev. Biochem*, 59, 29-59.
- Varlamova, E. G., Goltyaev, M. V., Novoselov, S. V., Novoselov, V. I. & Fesenko, E. E. (2013) Selenocysteine Biosynthesis and Mechanism of Incorporation into Growing Proteins. *Mol. Biol.*, 47, 488-495.
- Wada, M., Hsu, C. C., Franke, D., Mitchell, M., Heine, A., Wilson, I. & Wong, C. H. (2003) Directed Evolution of *N*-Acetylneuraminic Acid Aldolase to Catalyze Enantiomeric Aldol Reactions. *Bioorg. Med. Chem.*, 11, 2091-2098.

- Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. (2005) Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem. Int. Ed.*, 44, 7342-7372.
- Wan, W., Tharp, J. M. & Liu, W. R. (2014) Pyrrolysyl-Trna Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta*, 1844, 1059-1070.
- Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. (2007) Evolved Orthogonal Ribosomes Enhance the Efficiency of Synthetic Genetic Code Expansion. *Nat. Biotech.*, 25, 770-777.
- Wang, L., Brock, A., Herberich, B. & Schultz, P. G. (2001) Expanding the Genetic Code of *Escherichia Coli*. *Science*, 292, 498-500.
- Wani, R., Nagata, A. & Murray, B. W. (2014) Protein Redox Chemistry: Post-Translational Cysteine Modifications That Regulate Signal Transduction and Drug Pharmacology. *Front Pharmacol*, 5.
- Warren, L. (1959) Thiobarbituric Acid Assay of Sialic Acids. *J. Biol. Chem.*, 234, 1971-1975.
- Watts, A. G. & Withers, S. G. (2004) The Synthesis of Some Mechanistic Probes for Sialic Acid Processing Enzymes and the Labeling of a Sialidase from *Trypanosoma Rangeli*. *Can. J. Chem.*, 82, 1581-1588.
- Willey, J. M. & van der Donk, W. A. (2007) Lantibiotics: Peptides of Diverse Structure and Function. *Annu. Rev. Microbiol.*, 61, 477-501.
- Williams, G. J., Woodhall, T., Farnsworth, L. M., Nelson, A. & Berry, A. (2006) Creation of a Pair of Stereochemically Complementary Biocatalysts. *J. Am. Chem. Soc.*, 128, 16238-16247.
- Williams, G. J., Woodhall, T., Nelson, A. & Berry, A. (2005) Structure-Guided Saturation Mutagenesis of *N*-Acetylneuraminic Acid Lyase for the Synthesis of Sialic Acid Mimetics. *Protein Eng. Des. Sel.*, 18, 239-246.
- Windle, C. L., Müller, M., Nelson, A. & Berry, A. (2014) Engineering Aldolases as Biocatalysts. *Curr. Opin. Chem. Biol.*, 19, 25-33.
- Wolterink-van Loo, S., Siemerink, M. A. J., Perrakis, G., Kaper, T., Kengen, S. W. M. & van der Oost, J. (2009) Improving Low-Temperature Activity of *Sulfolobus Acidocaldarius* 2-Keto-3-Deoxygluconate Aldolase. *Archaea*, 2, 233-239.
- Woodhall, T., Williams, G., Berry, A. & Nelson, A. (2005) Synthesis of Screening Substrates for the Directed Evolution of Sialic Acid Aldolase: Towards Tailored Enzymes for the Preparation of Influenza a Sialidase Inhibitor Analogues. *Org. Biomol. Chem.*, 3, 1795-800.
- Wymer, N., Buchanan, L. V., Henderson, D., Mehta, N., Botting, C. H., Pocivavsek, L., Fierke, C. A., Toone, E. J. & Naismith, J. H. (2001) Directed Evolution of a New Catalytic Site in 2-Keto-3-Deoxy-6-Phosphogluconate Aldolase from *Escherichia Coli*. *Structure*, 9, 1-9.
- Xie, L. & van der Donk, W. A. (2004) Post-Translational Modifications During Lantibiotic Biosynthesis. *Curr. Opin. Chem. Biol.*, 8, 498-507.
- Young, T. S. & Schultz, P. G. (2010) Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.*, 285, 11039-44.
- Yukl, E. T. & Wilmot, C. M. (2012) Cofactor Biosynthesis through Protein Post-Translational Modification. *Curr. Opin. Chem. Biol.*, 16, 54-59.
- Zhang, W. H., Otting, G. & Jackson, C. J. (2013) Protein Engineering with Unnatural Amino Acids. *Curr. Opin. Struct. Biol.*, 23, 581-587.
- Zhu, Y. & van der Donk, W. A. (2001) Convergent Synthesis of Peptide Conjugates Using Dehydroalanines for Chemoselective Ligations. *Org. Lett.*, 3, 1189-1192.
- Zinoni, F., Birkmann, A., Stadtman, T. C. & Bock, A. (1986) Nucleotide-Sequence and Expression of the Selenocysteine-Containing Polypeptide of Formate Dehydrogenase (Formate-

Hydrogen-Lyase-Linked) from *Escherichia Coli*. *Proc. Natl. Acad. Sci. U. S. A.*, 83, 4650-4654.

Zioudrou, C., Wilchek, M. & Patchornik, A. (1965) Conversion of the L-Serine Residue to an L-Cysteine Residue in Peptides*. *Biochemistry*, 4, 1811-1822.