

**Exploration of the active site of *Staphylococcus aureus*
Neuraminic acid Lyase using non-canonical amino acids and
their effects on substrate specificity.**

Robert Peter Smith



The University of Leeds
Astbury Centre for Structural Molecular Biology

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

September 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given 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 the thesis may be published without proper acknowledgement.

© 2017 The University of Leeds and Robert Peter Smith

Acknowledgements

First and foremost, thank you to my supervisors, Alan, Adam and Joe. Without your help, guidance and wisdom, my project would not have been as successful as it was. Also, this thesis would not have been possible without the input and guidance of Colin from GSK, I truly appreciate the time you took to guide my computational experiments.

In addition, this work would not have been possible without James Ault and Rachel George for performing mass spectrometry on the large number of samples required for this project, and Chi Trinh for the many hours spent shooting crystals and refining structures with me at all hours of the day.

Thank you to my wonderful girlfriend Michelle, who provided support throughout my PhD. Without your unfailing love and support I very much doubt I would have made it this far. I am indescribably thankful for your presence over these years.

I am so grateful for coming into such a great lab, and for it continuing to be filled with so many brilliant people. The Berry and Radford labs, and visits to the pub on Fridays, deserve a lot of credit for helping me through these four years. An honourable mention goes to Claire for showing me the ropes when I started and helping me out when I got stuck.

Also, thanks to those over in Chemistry who helped me out, including: Alun, Rich, Rong, James and Sam. I hope I didn't get in the way too much over there!

Cheers to the Monday night gaming people, past and present, for all the good times. Because of you all I have always looked forward to Mondays, they were the highlight of the week.

Thanks go to Lucia, Marie and Rakesh for making my time in Stevenage actively enjoyable!

Finally, a massive thank you to my parents and grandad, who have supported me these 28 years and always encouraged me to pursue my passions.

Abstract

The cost of chemical synthesis of pharmaceuticals contributes significantly to their final price and part of this cost is incurred due to use of extreme temperatures and pressures required by some traditional catalysts. The interest in catalysis using enzymes, or biocatalysis, from industry has been growing recently, due to enzymes' ability to work at room temperature and pressure, and the reduction in toxic solvent waste produced from an enzyme reaction compared to a traditionally catalysed reaction. The specificity of enzymes, while useful in product formation, can make applying them to synthetic chemistry challenging due to the restriction this causes in substrates that each enzyme accepts. This can often be avoided by amino acid mutagenesis, but when this is performed genetically, only 20 different amino acids can be used. Non-canonical amino acids (ncAAs) have the potential to enhance properties of enzymes, such as enzyme stability and substrate specificities, to hitherto unseen extremes, due to the massive diversity of amino acids outside the canonical 20.

900 enzyme-aldehyde pairs were screened for activity, and Y252Lanthionine was found to catalyse the aldol reaction between pyruvate and glucuronolactone better than the wild type enzyme for the same reaction. Upon crystallisation, this enzyme was found to be a mixture of both L- and D-stereoisomers at the protein backbone where the ncAA was inserted. Computational experiments were performed to assess the substrate binding capability of the modified enzyme and the wild-type enzyme. The modified side chain holds the substrate more tightly than the wild-type side chain, contributing to increased residence time in the active site.

Table of Contents

Acknowledgements	iii
Abstract	iv
Table of Contents	v
List of Figures	ix
List of Abbreviations	xiv
Chapter 1 - Introduction	1
1.1 Enzymes in action	1
1.1.1 History of enzymes.....	1
1.1.2 Uses of enzymes.....	2
1.1.3 Power of enzymes: Catalysis on a protein scaffold.	5
1.2 Aldolases and NAL	11
1.2.1 Aldolases in industry	11
1.2.2 Neuraminic acid lyase as an aldolase	13
1.2.3 Mutagenesis of NAL for non-natural substrate catalysis ...	16
1.3 Structural features of <i>Staphylococcus aureus</i> Neuraminic acid lyase.....	20
1.3.1 Overall structure	20
1.3.2 Mechanism of action of Neuraminic acid lyase	22
1.3.3 Chemical modification of saNAL	24
Chapter 2 - Materials and Methods	32
2.1 Materials	32
2.2 General methods.....	33
2.2.1 2xTY media production.....	33
2.2.2 Agar plate production	33
2.2.3 Glycerol stock generation.....	33
2.3 DNA methods	34
2.3.1 DNA extraction and purification	34
2.3.2 Site-Directed Mutagenesis	34
2.3.3 Agarose gel electrophoresis	34
2.3.4 Sequencing	34
2.4 Protein methods	35
2.4.1 Protein expression	35
2.4.2 Protein purification	35

2.4.3 Saturation library protein expression	36
2.4.4 Saturation library protein purification	36
2.4.5 Bradford assay	36
2.4.6 SDS-PAGE gel electrophoresis	37
2.4.7 Molecular mass measurement by Liquid chromatography-mass spectrometry.	37
2.4.8 Protein Dialysis	38
2.4.9 Protein concentration	38
2.4.10 Size exclusion chromatography	38
2.4.11 Freeze drying	39
2.4.12 Modification used during screening	39
2.4.13 Modification as optimised for Y252Lni	39
2.4.14 Thiobarbituric acid assay - initial	40
2.4.15 Thiobarbituric acid assay - enhanced signal	40
2.4.16 Protein crystallisation	41
2.4.17 X-Ray crystallography	41
2.4.18 Circular Dichroism	42
2.5 Computational methods	43
2.5.1 Protein chain alignment	43
2.5.2 Structure preparation	43
2.5.3 GRID surface construction	43
2.5.4 Dynamics Generalised protocol	44
2.5.5 ManNAc molecular dynamics	44
2.5.6 Glucuronolactone molecular dynamics	45
2.5.7 Binding pose steered dynamics	46
2.5.8 Full length product molecular dynamics	47
2.5.9 Full length product steered dynamics	47
2.5.10 Dendrogram generation	47
2.6 Chemical synthesis of 2,5-dibromohexanediamide	49
Chapter 3 - Discovery of a chemically modified aldolase with a novel activity	50
3.1 Introduction	50
3.2 Assay development	51
3.2.1 Assay signal enhancement	51
3.2.2 Assay variability	56
3.3 Screening set-up and results	60

3.3.1 Choice of mutation positions	60
3.3.2 Choice of ncAAs to be incorporated	62
3.3.3 Choice of substrates	64
3.3.4 Screening results	65
3.4 Hit confirmation	69
3.4.1 Re-modification and control reaction	69
3.4.2 Removal of unfolded protein and initial kinetic assessment	74
3.5 Product confirmation	79
3.5.1 ¹³ C-NMR	79
3.5.2 ¹ H-NMR	80
3.5.3 Mass spectrometry	82
3.6 Kinetics	84
3.6.1 Plate layout	84
3.6.2 Kinetic parameters	85
3.7 Summary	88
Chapter 4 - Studying substrate interactions with position 252 of <i>S.</i> <i>aureus</i> NAL	89
4.1 Saturation mutagenesis at position 252	89
4.1.1 Saturation library construction	89
4.1.2 Saturation library purification and activity	90
4.2 X-ray crystallographic studies of Y252Lni <i>sa</i> NAL	95
4.2.1 Crystal formation, data collection and refinement	95
4.2.2 Description of the modified protein structure	97
4.2.3 Side-chain structural features	100
4.3 Computational studies of the active site of wild-type and Y252Lni <i>sa</i> NAL	105
4.3.1 Preliminary investigation	105
4.3.2 Substrate binding from bulk solvent	108
4.3.3 Substrate binding in the active site	109
4.3.4 Substrate retention in the active site	120
4.3.5 Removing the full length product	122
4.4 Summary	124
Chapter 5 - Conclusion	126
5.1 Summary	126
5.2 Future directions	129

5.3 Concluding remarks	131
Chapter 6 - Bibliography	132

List of Figures

Figure 1.1 Examples of industrial sectors that use enzymes	3
Figure 1.2 The reaction performed by orotidine 5'-phosphate decarboxylase.	5
Figure 1.3 Schematic representations of three different hypotheses of how enzymes achieve specific catalysis.	7
Figure 1.4 Reaction mechanism of carbonic anhydrase.....	9
Figure 1.5 The sequential addition reactions performed by DERA when run in the synthesis direction and Structures of two examples of commercially available statin molecules.....	12
Figure 1.6 Reaction performed by Neuraminic acid lyase.	14
Figure 1.7 Table summarising published kinetic data of wild-type NAL enzymes breaking Neu5Ac down into pyruvate and ManNAc from different species.	14
Figure 1.8 Table summarising published kinetic data of wild-type NAL enzymes from different species' activity for substrates other than Neu5Ac.....	16
Figure 1.9 Summary of publications in which residues have been mutated in the active site of E. coli NAL.	17
Figure 1.10 View of the active site of E. coli NAL showing the mutated residues listed in figure 1.9.	18
Figure 1.11 View of the saNAL tetramer.	20
Figure 1.12 Two views of a single monomer of saNAL, rotated by 180° around a vertical axis.	21
Figure 1.13 View of the active site of saNAL with and without pyruvate bound to K165.	22
Figure 1.14 Competing hypotheses of the mechanism of NAL.	23
Figure 1.15 The mechanism by which the ncAAs are formed in situ on the protein backbone.	26
Figure 1.16 Activity of wild-type saNAL compared to K165thialysine (ThI) in buffers at different pH.	28
Figure 1.17 The modification from phenylalanine to 2,3-dihydroxypropyl cysteine (Dpc) performed on saNAL.	30
Figure 1.18 Hydrogen bonding network cause by insertion of a Dpc side-chain at position 190 of saNAL.	31
Figure 3.1 Reactions required for visible product using the TBA assay....	52

Figure 3.2 Standard curve of commercially purchased N-acetylneuraminic acid when analysed by the TBA assay.	53
Figure 3.3 Graph showing the effect of periodate incubation time on the absorbance produced in the TBA assay.	54
Figure 3.4 The difference in signal from overnight reactions between different assay protocols.	55
Figure 3.5 Standard curve of commercially purchased N-acetylneuraminic acid when analysed by the TBA assay using both the original and newly optimised assay conditions.	56
Figure 3.6 The average absorbance of each row or column of an empty plate.	57
Figure 3.7 Average standard deviation of each well in each row or column of an empty plate.	58
Figure 3.8 Chart showing the absorbance produced by control reactions between pyruvate and different aldehydes.	59
Figure 3.9 View of the active site of wild-type <i>S. aureus</i> N-acetylneuraminic acid lyase showing the positions to be targeted in the screen.	61
Figure 3.10 Translated DNA sequencing of the residues to be targeted in the screen.	62
Figure 3.11 Structural diagrams of the nCAAs produced when modified using the thiols selected for the screen.	63
Figure 3.12 Aldehyde substrates used in the screen.	65
Figure 3.13 Example plate layout diagram of the arrangement for the activity screen.	66
Figure 3.14 Summary table of the top ten hits found in the screen.	68
Figure 3.15 Chemical structural diagrams for the discovered hit.	70
Figure 3.16 Mass spectrum analysis of a modification of Y252C using L-cysteine as the thiol to confirm Y252Lni production.	71
Figure 3.17 Mechanistic diagram of the chemical modification procedure to produce nCAAs	72
Figure 3.18 Mass spectrum analysis of Y252Lni produced using the optimised chemical modification procedure.	73
Figure 3.19 Summary of the control reactions performed for the reaction between pyruvate and glucuronolactone with Y252Lni.	74
Figure 3.20 Size exclusion chromatography gel filtration elution profile of Y252Lni.	75
Figure 3.21 Mass spectrometer data of each peak from the size exclusion chromatography gel filtration of Y252Lni.	76

Figure 3.22 Circular dichroism (CD) scans of wild-type and Y252Lni enzymes.	77
Figure 3.23 Increase of absorbance over time in the TBA assay.....	78
Figure 3.24 ^{13}C DEPT-135 NMR of a mixture of pyruvate, labelled with ^{13}C at the methyl group, and glucuronolactone alone and with enzyme..	79
Figure 3.25 ^1H - ^{13}C HMQC NMR of a mixture of natural abundance pyruvate, glucuronolactone and enzyme after 14 days.....	81
Figure 3.26 ^1H - ^1H COSY NMR of a mixture of natural abundance pyruvate, glucuronolactone and enzyme after 14 days.....	81
Figure 3.27 Possible products of the reaction between pyruvate and glucuronolactone.....	82
Figure 3.28 Mass spectrum analysis of a mixture of natural abundance pyruvate, glucuronolactone and enzyme after 14 days.	83
Figure 3.29 Cleland diagram representing the sequence of events during NAL catalysis.....	84
Figure 3.30 Table showing the generalised layout of the kinetic analysis plates.	85
Figure 3.31 Average rate of reaction at different glucuronolactone concentrations for both wild-type and Y252Lni. Summary table of the fit statistics of the average results of the kinetic characterisation. ..	86
Figure 3.32 Summary table of the fit statistics of the average results of the kinetic characterisation.	86
Figure 4.1 Translated DNA sequencing of all members of the saturation mutagenesis library at position 252.	89
Figure 4.2 Composite SDS-PAGE gel image of all 20 members of the canonical amino acid site-directed mutagenesis library at position 252.	90
Figure 4.3 Photographs of the TBA assay plates showing the increase in colour over time of all the members of the saturation library at position 252.	92
Figure 4.4 Fold-difference in activity from wild-type while performing the aldol reaction between pyruvate and glucuronolactone.	93
Figure 4.5 Fold-difference from wild-type as in figure 4.4, compared to the molecular weight of the amino acid at position 252.....	93
Figure 4.6 Y252Lni crystals examined using X-Ray crystallography.	96
Figure 4.7 The modelled Y252A structure and the $2F_{\text{obs}}-F_{\text{cal}}$ electron density map of the Y252Lni crystal overlaid.	97
Figure 4.8 Cartoon depiction of the wild-type and Y252Lni modified enzymes.	98

Figure 4.9 Cartoon depiction of one monomer of wild-type and Y252Lni overlaid.	98
Figure 4.10 Overlaid monomers of wild-type and Y252Lni structures, Y137 and K165.	99
Figure 4.11 Overlaid monomers of wild-type and Y252Lni structures, F190.	100
Figure 4.12 The lanthionine side chains at 252 overlaid.	101
Figure 4.13 Y252Lni structure showing residues within 5 Å of the modified side chain Y252Lni.	101
Figure 4.14 Y252Lni side chain, with the distances measured as being potential hydrogen bonds shown as dashed yellow lines.	102
Figure 4.15 Diagrams of the amino acids at 252.	103
Figure 4.16 L-Lni is not a good fit in the electron density maps of some monomers.	104
Figure 4.17 Chemical structure of glucuronolactone, the catalytic species, and the non-catalytic analogue used in crystal trials.	105
Figure 4.18 Overlaid subunits of wild-type and Y252Lni saNAL.	107
Figure 4.19 GRID calculations using a hydroxyl group as the probe on wild-type, Y252L-Lni and Y252D-Lni enzymes.	109
Figure 4.20 Glucuronolactone.	110
Figure 4.21 Final orientations of simulations using the wild-type structure and glucuronolactone.	111
Figure 4.22 Representative poses during the preliminary simulations of Y252Lni with glucuronolactone.	112
Figure 4.23 Positions of the substrate in the Y252D-Lni rotation simulations before and after energy minimisation.	113
Figure 4.24 The different groups and subgroups of binding that were observed.	114
Figure 4.25 Schematic diagram showing the purpose of the distance measurements taken from the substrate binding simulations.	115
Figure 4.26 Ten point running average data for each of the five distances measured.	116
Figure 4.27 Dendrogram showing the relationship between different poses identified from examination of the running average distance data. .	117
Figure 4.28 Table showing total simulation time spent in each group by each enzyme.	119
Figure 4.29 Graphed percentage data from figure 4.28.	119
Figure 4.30 Maximum weight constant applied that did not disrupt the binding of each group.	121

Figure 4.31 The average maximum weight required to disrupt binding..	122
Figure 4.32 Example of the hydrogen bonding network attaching the full length product to the active site.	123
Figure 5.1 Chemical diagram of Lysine forming a Schiff base and N-methyl-thialysine forming an iminium ion in the same reaction. ..	130

List of Abbreviations

Cec	(carboxyethyl)cysteine
Chc	(cyclohexanyl)cysteine
COSY	Correlation spectroscopy
D ₂ O	Deuterium oxide
D-Ara	D-Arabinose
DEPT	Distortionless enhancement by polarization transfer
DERA	2-Deoxy-D-ribose 5-phosphate aldolase
D-Gal	D-Galactose
D-GalNAc	<i>N</i> -Acetyl-D-galactosamine
D-Glc	D-Glucose
D-Glc acid	D-Glucuronic acid
D-GlcN	D-Glucosamine
D-GlcNAc	<i>N</i> -Acetyl-D-glucosamine
DiBr	2,5-Dibromohexandiamide
D-Lyx	D-Lyxose
D-Man	D-Mannose
DPA	Dipropylamide analogue
Dpc	2,3-Dihydroxypropyl cysteine
D-Rib	D-Ribose
D-Xyl	D-Xylose
EDTA	Ethylenediaminetetraacetic acid
Fec	(2,2,2-Trifluoroethyl)cysteine
Fmc	(2-Fuylmethyl)cysteine
H ₂ O	Water
Hec	(2-Hydroxyethyl)cysteine
Hhc	(1-Hydroxy-3-hexanyl)cysteine
HMQC	Heteronuclear multiple-quantum correlation spectroscopy

L-Ara	L-Arabinose
L-Fuc	L-Fucose
Lni	Lanthionine
ManNAc	<i>N</i> -Acetyl- <i>D</i> -mannosamine
Mic	(2-(3,5-Dimethyl-4-isoxazolyl)ethyl)cysteine
Moc	(2-Methoxy-2-oxoethyl)cysteine
MOE	Molecular Operating Environment
NAL	Neuraminic acid lyase
ncAA(s)	Non-canonical amino acid(s)
Neu5Ac	<i>N</i> -Acetyl- <i>D</i> -neuraminic acid
OMP	Orotidine 5'-phosphate
OPD	Orotidine 5'-phosphate decarboxylase
PTM	Post translational modification
QM/MM	Quantum mechanics/molecular mechanics
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
TBA	Thiobarbituric acid
TCA	Trichloroacetic acid
Thl	Thialysine
TIM	Triose phosphate isomerase
UMP	Uridine 5'-phosphate
WT	Wild-type

Chapter 1 - Introduction

1.1 Enzymes in action

1.1.1 History of enzymes

Humans have been using enzymes to improve their lives for thousands of years, either using the enzymes in isolation or using the enzyme as part of a whole organism. A classic example of enzymes in action is during fermentation. There is evidence of humans actively attempting to ferment fruit into alcoholic beverages over 9000 years ago (McGovern et al., 2004), employing the enzymes contained within naturally occurring yeasts and other microbes to turn sugars into ethanol. Indeed, even the raising of livestock such as goats and cattle, which has been occurring for 10 000 years (Zeder and Hesse, 2000), could be viewed as utilising organic bio-reactors to ferment indigestible long-chain carbohydrates into edible food. A more direct use of enzymes is in the production of cheese. The initial step in cheese making employs an extract of the enzymes from the stomach of a ruminant to curdle milk into curds and whey and this process has been occurring for at least 7000 years (Salque et al., 2013).

As effective as these practices may have been at providing food and drink, serious study of enzymes only began to occur in the 19th century. Earlier discoveries by scientists such as Réaumur had determined that the process of digestion in birds is a chemical process and not a mechanical one, but they had not yet discovered the active components of digestion (de Réaumur, 1752). Years later, it was during study related to the millennia old process of brewing that the first enzyme was discovered. In 1833, Anselme Payen discovered the enzyme diesterase, recently more commonly known as amylase, after isolating it from malted barley (Payen and Persoz, 1833). The barley malt was ground and extracted with water, which upon addition of alcohol caused an amylase complex to precipitate out of solution. This precipitate could be dried and, when re-dissolved, was still capable of catalysis. In 1877 the term “enzyme” was first coined by Wilhelm Kühne to describe cell extracts capable of catalysis, distinguishing them from the catalytic or fermentation abilities of living cells (Kühne, 1877).

The actual identity of the molecules that perform catalysis was still uncertain, however. The impurity of the enzyme isolations to this point had meant that, while protein was the major constituent of catalytic preparations, the possibility existed that it was an inert partner that was in some way attached to or supported the catalytic species. This was mostly put to rest by experiments performed by John Northrop in 1930, in which he purified pepsin via seven sequential crystallisations in different solvents, showing that in all likelihood, the resulting catalytically active crystals were composed of a single species, and that species was a protein (Northrop, 1930). As such, the stage was set for the modern scientific study of enzymes as we know it.

1.1.2 Uses of enzymes

The current uses of enzymes still include the ancient, with the fermentation and dairy industries still relying heavily on enzyme preparations, but have expanded to include more modern industries as well. A table of examples can be seen in figure 1.1. Modern enzymes are often expressed in microbiological hosts, regardless of the enzyme's original source, as this allows for optimisable expression that can be scaled up or down with minimal notice.

Enzymes have been so widely adopted as they have a number of advantages over traditional catalysts. Firstly, if the reaction to be performed is already being performed by Nature, e.g. fermentation and the denaturing of milk proteins for cheese production are both performed in nature but have been co-opted by humans, then Nature will almost certainly have already developed an enzyme to perform that function. This enzyme will often be an efficient catalyst for this reaction, and it may prove easier to simply use the natural catalyst than to try to develop novel catalysts for that reaction.

Secondly, even if the reaction is to take place in conditions that are very different to those that are present in Nature, but the reaction is the same as a natural reaction, e.g. biological laundry detergents using enzymes at high temperatures, enzymes can usually be modified to maintain functionality under a wide range of conditions. Directed evolution describes the intentional evolution of an enzyme, in this case, to adapt it for use under non-natural conditions. This has allowed enzymes to be adapted for high temperatures, high solvent concentrations and increased activity at extreme pH (Giver et al., 1998, Reetz et al., 2010, Cherry et al., 1999).


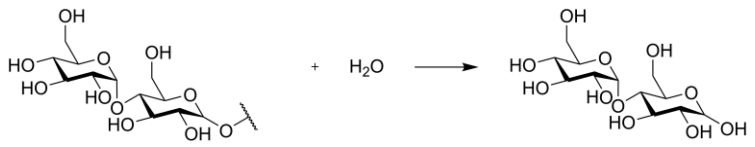
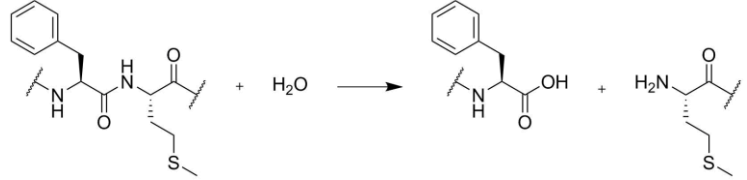
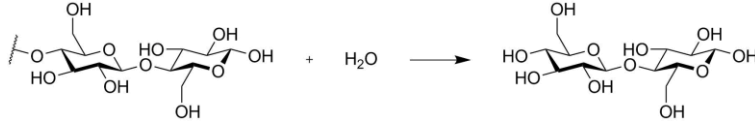
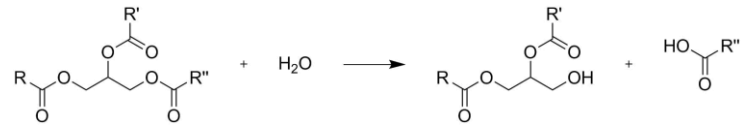
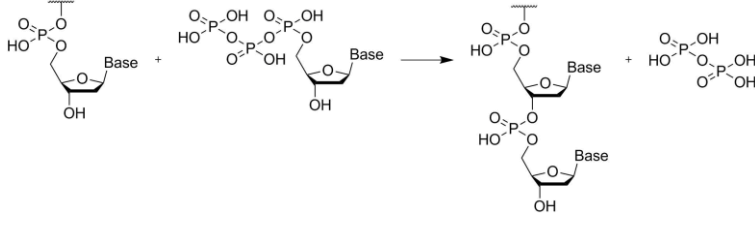
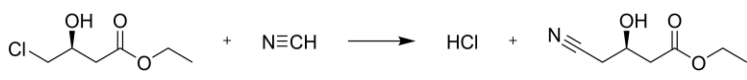
Industry	Example enzyme - use	Reaction catalysed
Cooking	Papain - tenderise meat	 <p style="text-align: center;">R' = not Valine</p>
Brewing	Amylase (diastase) - starches into fermentable sugars	
Dairy	Chymosin - curdling milk	
Biofuels	Cellulase - cellulose to fermentable sugars	
Detergents	Lipase - removing fatty stains	
Molecular biology	DNA polymerase - amplifying DNA in PCR reactions	
Chemical manufacture	Halohydrin dehalogenase - synthesising intermediate in atorvastatin manufacture	

Figure 1.1 Examples of industrial sectors that use enzymes, an enzyme that is used in that sector, and the reaction that each enzyme performs.

For all reactions, enzymes have evolved to function in aqueous solution, whereas traditional catalysts have most often been used in organic solvents. Even some common organic solvents, such as dichloromethane, are toxic however, and as such reducing the use of these solvents is beneficial for the health and safety of the people who use them. Another factor to consider when using organic solvents is that even the non-toxic organic solvents need to be disposed of via a specialised waste stream, which incurs its own cost. Using water as the solvent in which a reaction is conducted circumvents both these concerns with organic solvents, and enzymes have the advantage of being ready-made catalysts for conducting reactions in aqueous solutions.

Additionally, the majority of enzymes have evolved to act on chiral substrates such as amino acids and sugars, and as a consequence, enzymes are often able to easily distinguish between different chiral substrates and act selectively to process or produce one stereoisomer alone. Traditional catalysts are often less selective and have less control over the stereochemistries they produce. This can often mean an increase in waste if the wrong stereochemistry is discarded, or an increase in process complexity if it is resolved further down the production line.

Also, traditional catalysts often require extreme temperatures or pressures to achieve the desired reaction outcomes, either in terms of yield or reaction rate. Most studied enzymes have evolved to perform reactions at room temperature and atmospheric pressure, and as such an enzyme catalysed reaction can often be conducted at a temperature and pressure much closer to ambient than the same reaction performed using traditional catalytic methods.

Finally, enzymes can be made using very cheap starting materials with the expression organism performing the complicated chemical synthesis of the catalyst. Even with the addition of trace elements and recyclable co-factors, the cost of enzymes as catalysts can be quite low. When this is compared to some traditional catalysts which may include precious metals or complicated molecules that need to be synthesised, potentially costing hundreds of pounds per gram or more, the cost of enzymes becomes quite reasonable.

1.1.3 Power of enzymes: Catalysis on a protein scaffold.

As we have seen, enzymes are used widely across industrial sectors and even in the home due to their ability to increase the rates of chemical reactions. Some enzymes display astonishing enhancements of the rate of reaction when compared to the uncatalysed reaction. One of the most extreme examples of this rate enhancement comes from the enzyme orotidine 5'-phosphate decarboxylase (OPD). This enzyme catalyses the decarboxylation of orotidine 5'-phosphate (OMP) to form uridine 5'-phosphate (UMP) (figure 1.2), uridine being one of the four bases in RNA.

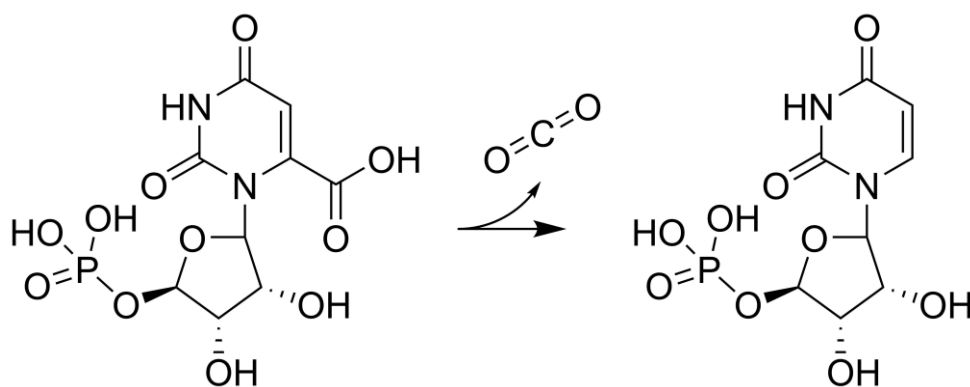


Figure 1.2 The reaction performed by orotidine 5'-phosphate decarboxylase (OPD). Orotidine 5'-phosphate has one carbon dioxide molecule removed from it to form uridine 5'-phosphate, one of the bases in RNA. This reaction is very slow without catalysis, taking millions of years per turnover, but is much faster when catalysed with OPD, taking only milliseconds.

Yeast OPD was examined performing the decarboxylation of a related substrate, 1-methyl orotic acid, and its rate was compared to the reaction occurring in buffer without a catalyst (Radzicka and Wolfenden, 1995). The uncatalysed reaction was calculated to take 113 million years per turnover when conducted in neutral buffer at room temperature. In comparison, when conducted in the same conditions with enzyme, the turnover rate was one molecule in 26 milliseconds, meaning the enzyme catalysed reaction has a turnover rate approximately 10^{17} fold higher than the uncatalysed reaction, giving this enzyme one of the greatest rate enhancements known. The enzyme accomplishes this prodigious rate increase without the use of any co-factors or metal ions, making this feat even more remarkable.

Carbonic anhydrase is one of the fastest enzymes known. It catalyses the reversible conversion of carbon dioxide and water into bicarbonate and a proton, and with a turnover rate of 10^6 per second, the enzyme turns over

the substrate at the rate that substrate can diffuse into the active site. The overall direction in which the reaction occurs will depend on the region of the body that the enzyme is in at the time, e.g. in the lungs the enzyme will be mostly responsible for removing bicarbonate and creating carbon dioxide to be removed from the body; in the deeper tissues, the interconversion of carbon dioxide and bicarbonate can be used to control the local pH; and in highly respiring tissues carbon dioxide produced during respiration will be turned into bicarbonate for more efficient transport in the blood. The uncatalysed rate of this reaction is much faster than for OPD, taking only 7.7 seconds to turnover, making the rate enhancement of the enzyme lower than OPD despite the astonishingly fast catalysed rate. Most carbonic anhydrases utilise an inorganic cofactor, a zinc ion, in the active site, making this a metalloprotein.

Serine proteases are some of the best-studied enzymes and as such their mechanism has been examined in detail. They catalyse a cleavage reaction that takes place at a specific place in the protein sequence, i.e. only on the C-terminal side of lysines and arginines for trypsin, and mostly function in digestion to break down proteins into smaller polypeptide chains. This reaction takes place approximately 10^{10} times faster when catalysed by the enzyme than when it is uncatalysed, and this with the aforementioned sequence specificity.

This specificity of enzymes was notable from an early stage in enzyme research, eventually leading to the proposal of the lock and key model of enzyme selectivity (Fischer, 1894). This model suggested that both the enzyme and the substrate were rigid and the enzyme achieved its specificity by being a complementary shape to the substrate, allowing it to bind in the correct orientation for catalysis (figure 1.3 A). While this theory could describe a number of enzyme reactions quite well, a later theory was proposed called the induced fit theory (Koshland, 1958). This suggests that some parts of an enzyme are much more mobile and change conformation depending on whether a molecule is bound to them (figure 1.3 B). As such, the binding of a substrate may change the structure of the enzyme in a way that induces catalysis. Similar molecules binding to the protein that are not substrates may bind as well, but they do not induce the correct changes in conformation required for a productive reaction.

Again, the induced fit theory of enzyme catalysis could also explain most enzyme reactions well, and later research, especially X-ray crystallography structures of enzymes bound to substrate mimics, have shown it to be

accurate in that some proteins do significantly change structure when enzymes bind. However, it does not fully describe reality. More recent hypotheses suggest that during an enzymatic reaction, both the substrate changes the shape of the enzyme, and the enzyme changes the shape of the substrate. This is called substrate destabilisation.

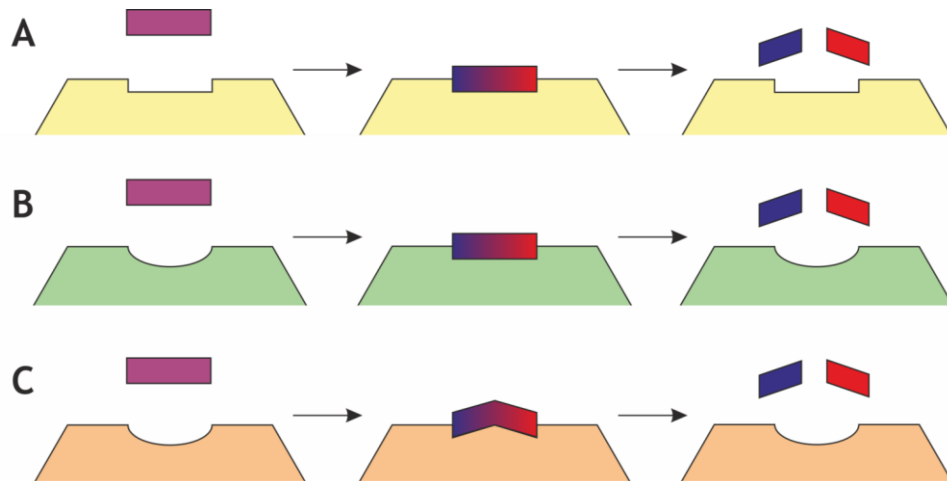


Figure 1.3 Schematic representations of three different hypotheses of how enzymes achieve specific catalysis. **A** Shows the lock-and-key theory proposed by Fischer in which both enzyme and substrate are rigid. **B** Shows the induced fit theory proposed by Koshland in which the enzyme reacts upon binding of the correct substrate to give a complementary active site. **C** Shows induced fit theory combined with transition state binding, in which the enzyme and substrate both change shape, the enzyme changing shape upon substrate binding and having highest affinity for the transition state of the reaction, and the substrate moving into the transition state in response.

In essence, the enzyme will change the substrate, either by Van Der Waals or electrostatic interactions, into a conformation that is more energetically active, e.g. straining the bond that is to be broken, modulating the electrostatic charge on a functional group that is to be removed, or forcing two atoms that are to be bonded close together. This often causes the substrate to more closely resemble the transition state of the reaction. The enzyme will have a higher affinity for the transition state of the reaction than for the substrate in its native conformation, allowing it to hold onto the molecule firmly at this point. The transition state of the reaction is therefore stabilised. This is demonstrated in the activity of catalytic antibodies, antibodies which have been engineered to be able to perform a chemical reaction. Antibodies can be made that bind the transition state of a reaction and when the substrate of the particular

reaction is supplied to these antibodies, a low but observable rate of catalysis occurs (Belogurov et al., 2009).

The low reaction rate of catalytic antibodies is partially due to the fact that substrate destabilisation can also happen in other ways, for example a charge on a reaction intermediate could be neutralised by corresponding opposing charges on side chains in the active site. Once the reaction occurs, the transition state is no longer present, and the enzyme loses affinity for the molecule, allowing it to leave the enzyme active site (figure 1.3 C).

Another method by which enzymes enhance catalysis is through providing an alternate mechanism for the reaction to take. For example, while a reaction in solution may at some point have to abstract a proton directly from water, the enzyme may provide this proton via a side chain such as histidine at a pK_a that much more readily donates the proton. Alternatively, a short-lived covalent intermediate may form between a substrate and the enzyme, changing the reaction mechanism required to produce the desired products to a mechanism that has a much lower activation energy.

Orotate decarboxylase uses a combination of these methods to achieve its dramatic rate enhancement. It is thought that the primary rate enhancement strategy is via transition state stabilisation, but it also utilises destabilisation of the ground state of the substrate to initiate the reaction (Fujihashi et al., 2015). Upon binding to the enzyme, OMP causes a flexible loop to close over the active site, isolating it from the solution (Harris et al., 2002). An aspartate residue interacts with the carboxylate group to be removed and pushes the carboxylate group out of the plane of the ring, destabilising the substrate (Chan et al., 2009). The distorted carboxylate group then detaches from the ring, forming a carbanion on the pyrimidine ring. This carbanion transition state is stabilised partially by the distribution of the electrons through the conjugated ring system, but mostly by interaction with a lysine group on the enzyme providing a positive charge (Toth et al., 2007).

Trypsin also uses transition state stabilisation to enable its catalytic function. Simply, the active site consists of a triad of residues, serine, histidine and aspartate, and also a region called an oxyanion hole, where a number of backbone nitrogen groups face into the active site. When the substrate binds, the histidine removes a proton from the serine group, which can then attack the carbonyl carbon of the amide bond. The newly

formed charge on the histidine is stabilised by the negative charge on the aspartate. The attack of the serine onto the carbonyl carbon forms a negatively charged oxyanion on a tetrahedral carbon intermediate, which is stabilised by the oxyanion hole in the active site. The oxyanion then attacks the tetrahedral carbon, reforming the planar carbonyl and breaking the amide bond, releasing one half of the substrate. Water is then used to reform a tetrahedral intermediate, which is again stabilised by the oxyanion hole. This second tetrahedral intermediate is removed from the enzyme using the same attack by the oxyanion as previously (Polgár, 2005).

Carbonic anhydrase primarily enhances the rate of the reaction it catalyses by providing an alternate hydroxyl source for the carbon dioxide, or an alternate proton source for the bicarbonate (figure 1.4). When carbon dioxide is present it binds to the backbone amine of threonine in the active site. The carbon dioxide carbon is then attacked by a hydroxyl group that is bound to a zinc atom in the active site. This forms the bicarbonate, which then leaves the active site, meaning the substrate has been converted to product, but the active site is not yet in a catalysis ready state.

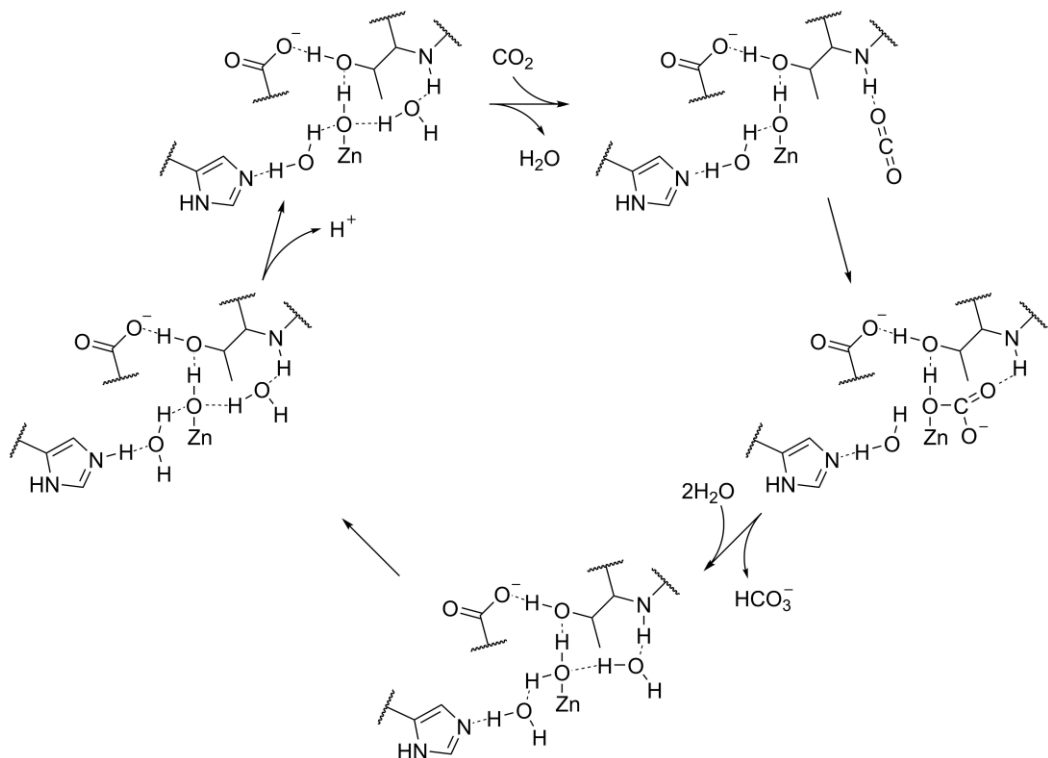


Figure 1.4 Reaction mechanism of carbonic anhydrase, adapted from Lindskog, 1997, with arrows representing the CO₂ to bicarbonate reaction.

Two water molecules then enter the active site, one binding to the zinc atom, and the other hydrogen bonding to the backbone amine of the threonine residue. One of the protons is abstracted from a water molecule associated with a histidine residue, protonating the histidine. The water molecule associated with the histidine concurrently abstracts a proton from the zinc-bound water, leaving a water molecule associated with the histidine, and a hydroxyl bound to the zinc atom. The proton is then lost from the histidine to a buffer in the solution, returning the active site to a catalytic state (Lindskog, 1997).

Interestingly, carbonic anhydrase makes use of a prosthetic group, a tightly bound non-polypeptide species, in the form of a zinc atom in the active site to conduct catalysis. Many enzymes make use of prosthetic groups and covalently bound post-translational modifications (PTMs), to increase the diversity of chemical groups that proteins are able to access, making up for deficiencies in the 20 canonical amino acids, and as such can aid in all methods of activity enhancement mentioned above. PTMs can be considered natural non-canonical amino acids if they are covalently bound to the protein at an alpha carbon or on the side chain of a canonical amino acid.

1.2 Aldolases and NAL

Aldolases perform a reversible synthetic reaction between an aldehyde group and another carbonyl group, forming a larger molecule, now with an alcohol group attached to the carbon in place of the aldehyde. This not only forms a molecule that has a longer carbon chain via formation of a carbon-carbon bond but also has the potential to form two new stereocenters in the product molecule. Carbon-carbon bond forming reactions can be quite difficult to conduct using traditional catalysts, making enzymes that are able to perform this reaction valuable. Additionally, non-enzymatic chemical reactions often suffer from poor stereoselectivity, while aldolase catalysed reactions can be exceptionally stereoselective. These properties make aldolase enzymes potentially very useful as catalysts in chemical synthesis reactions.

1.2.1 Aldolases in industry

While aldolases have great potential as biocatalytic enzymes due to their carbon-carbon bond forming ability, the narrow substrate scope and lack of stereochemically pure product of some aldolases have however held back the use of aldolases in industry more widely. Efforts have been made to improve aldolases, modifying their activity to make their industrial use more viable.

For example, transaldolase has been modified to enhance production of bioethanol in *Pichia stipitis* (Chen et al., 2012). Transaldolase was found to be one of the limiting enzymes in the pathway that produces ethanol from xylose, a sugar abundant in the hemicellulose that makes up part of the cell walls of woody plants. This chain of reactions would allow by-products of the wood industry to be converted into liquid fuel. Transaldolase catalyses the conversion of sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate into erythrose-4-phosphate and fructose-4-phosphate. Error-prone PCR was performed on the transaldolase gene to produce mutant genes, which were then expressed and assessed for activity in a high throughput screen. Two mutants, Q263R and K190M, were discovered which displayed enhanced transaldolase activity. When these modified genes were inserted into the natural host, *P. stipitis*, the ethanol production per cell was significantly enhanced in the cultures with the modified genes. This example shows that not only are aldolases important enzymes in manufacture, but they are amenable to mutations that improve their abilities.

A widely studied aldolase for industrial applications is the enzyme 2-deoxy-D-ribose 5-phosphate aldolase (DERA), used in the production of statin drugs. Statins are chemically varied drugs designed to inhibit the action of 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase, the first enzyme in the pathway that produces cholesterol, therefore reducing the level of cholesterol in the blood. High levels of cholesterol in the blood have been connected to an increase in the risk of cardiovascular disease, giving statins the potential to reduce the chance of this disease occurring.

DERA catalyses the reversible addition of acetaldehyde to D-glyceraldehyde 3-phosphate to produce 2-deoxy-D-ribose 5-phosphate, but is also able to add acetaldehyde onto a relatively wide range of other aldehyde molecules. DERA is also interesting as the products of the reaction also contain an aldehyde group in a terminal position, allowing them to have a subsequent acetaldehyde molecule added onto the carbon backbone of the molecule. From these properties, DERA is able to produce long carbon chains from relatively short precursor molecules, these carbon chains have regularly placed, chiral features along the carbon chain, and one end of the molecules produced can display a variety of chemical groups (figure 1.5A). The diverse array of molecules that DERA can produce can then be used as building blocks in the construction of the larger statin molecules (figure 1.5B).

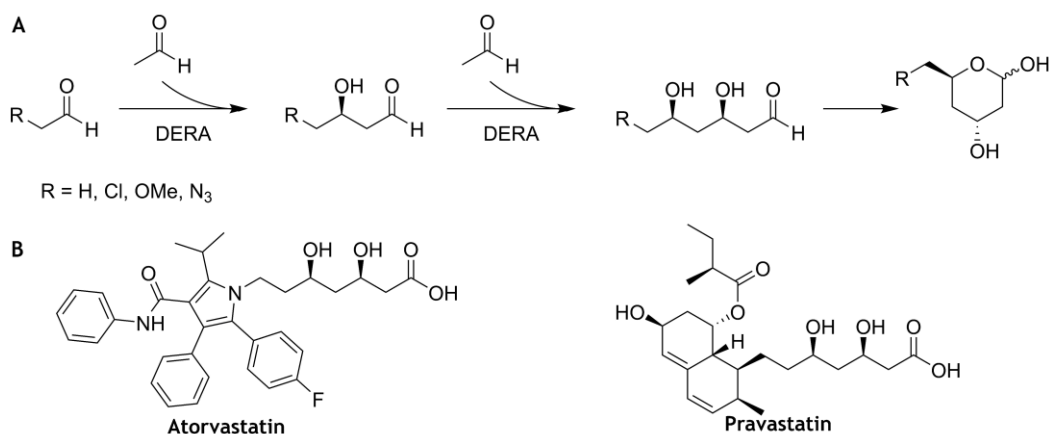


Figure 1.5 A The sequential addition reactions performed by DERA when run in the synthesis direction. The final lactonisation, turning a linear product into a cyclic one, occurs spontaneously in solution and does not require the enzyme. Adapted from Greenburg et al., 2004. B Structures of two examples of commercially available statin molecules, Atorvastatin and Pravastatin. The structure homologous to that produced by DERA can be seen on the right hand side of each molecule.

DERA has been subjected to directed evolution in an effort to optimise it to produce industrially useful building blocks, such as (3*R*, 5*S*)-6-chloro-2,4,6-trideoxyhexapyranoside (Ošljaj et al., 2013). Error prone PCR was used to generate a diverse library of DERA variants, which were screened for resistance to chloroacetaldehyde. These were then combined to give variants which had increased stability when subjected to concentrations of chloroacetaldehyde up to 500 mM and 1 M concentrations of acetaldehyde, industrially relevant concentrations of the reagents (Jennewein et al., 2006).

Even without directed evolution, very efficient variants of DERA can be discovered. In an effort to find versions of DERA with improved activity and improved tolerance to high substrate concentrations compared to the *E. coli* protein, DNA was isolated from diverse environmental samples and screened for activity. After process optimisation, an enzyme was found that, under the conditions tested, could produce 93 g/l of product over 3 hours with 2% (w/w) of DERA added, compared to 76 g/l of product over 3 hours with 4.8% (w/w) of *E. coli* DERA (Greenberg et al., 2004).

There are a number of other aldolases that have been adapted to make potentially useful industrial catalysts. For example, threonine aldolases naturally produce β -hydroxy- α -amino acids, which are found in a number of compounds such as antibiotics and immunosuppressants (Gutierrez et al., 2008, Liu et al., 2000). Another example of an aldolase that has been adapted to make potentially useful compounds is neuraminic acid lyase.

1.2.2 Neuraminic acid lyase as an aldolase

Neuraminic acid lyase is an aldolase that catalyses the reversible addition of a three-carbon molecule, pyruvate, onto a six-carbon *N*-acylated amino sugar, *N*-acetyl-*D*-mannosamine (ManNAc), to form the nine-carbon molecule *N*-acetyl neuraminic acid (Neu5Ac) (figure 1.6). The enzyme's likely function *in vivo* is to break down Neu5Ac into pyruvate and ManNAc for use as a carbon source and to regulate the intracellular concentration of Neu5Ac. However, when provided with pyruvate and ManNAc, the enzyme can synthesise Neu5Ac. As such, it provides a catalytic route to high complexity sugar-derived molecules.

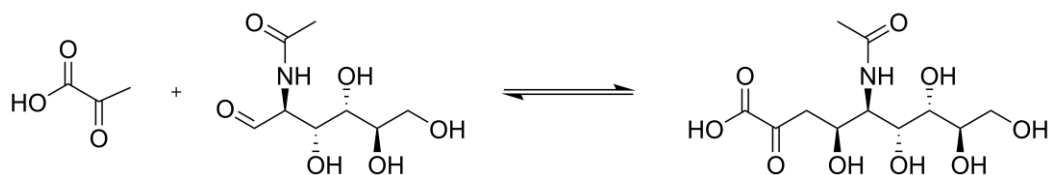


Figure 1.6 Reaction performed by Neuraminic acid lyase (NAL). NAL performs an aldol reaction, attaching one molecule of pyruvate to one molecule of *N*-acetyl-D-mannosamine, creating one molecule of *N*-acetyl-D-neuraminic acid.

Studies have been conducted on a number of bacterial and eukaryotic NAL variants and the catalytic rates of NAL from different organisms display a range of activities, a sample of which are summarised in figure 1.7.

Organism	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($\text{min}^{-1} \text{mM}^{-1}$)	Source
<i>Escherichia coli</i>	260 ± 6	4.4 ± 0.3	59	Williams et al., 2005
	600 ± 20	2.5 ± 0.3	240	Li et al., 2008
	0.81	2.6	0.311	Wada et al., 2003
	11 ± 1.8	2.6 ± 0.2	4.0	Hsu et al., 2005
<i>Pasturella multocida</i>	960 ± 60	4.9 ± 0.7	200	Li et al., 2008
<i>Staphylococcus aureus</i>	250 ± 5	2.2 ± 0.1	114	Timms et al., 2013
<i>Clostridium perfringens</i>		1.85		Schauer et al., 1971
<i>Sus scrofa</i>		3.7		Schauer and Wember, 1996

Figure 1.7 Table summarising published kinetic data of wild-type NAL enzymes breaking Neu5Ac down into pyruvate and ManNAc from different species. Blank cells in the table represent data that were not in the specified publication.

By far the most studied variant is that derived from *E. coli*. When the breakdown reaction, from Neu5Ac to ManNAc and pyruvate, is considered, the K_m , or the substrate concentration at which the rate is half that of the

maximum rate, of the *E. coli* NAL (*ecNAL*) is generally around 3 mM. The k_{cat} , or the turnover rate of this particular enzyme-substrate complex, of this reaction has much more variable published measurements, from 0.8 per minute to 600 per minute, a difference of approximately 3 orders of magnitude. This could be due to differences in measurement techniques or differences in the quality of the enzyme preparations. The low K_m of NAL highlights its potential as a biocatalytic tool in synthesis reactions, as it will reach its highest rate at what would be relatively low concentrations of substrate for a chemical reaction in an industrial setting.

Neu5Ac analogues are on the market as anti-influenza medication (Hayden et al., 1999, Yamashita et al., 2009, von Itzstein et al., 1993), as the flu virus uses Neu5Ac on the surface of cells as an anchor to attach itself to. The Neu5Ac analogues act to bind to these Neu5Ac recognition sites on the virus particles, preventing virus binding and entry. As such, NAL represents an industrially interesting enzyme, providing easy access to Neu5Ac homologues. However, the utility of NAL in this regard is dependent upon it accepting a wide range of substrates. NAL has been studied with a wide range of potential substrates, and a selection of studies are summarised in figure 1.8.

In general, the more similar a molecule is to Neu5Ac, the better it will be catalysed by NAL. Small modifications to Neu5Ac, e.g. the replacement of a hydrogen atom with a chlorine atom, are generally tolerated well and the further a modification is from the aldehyde end of Neu5Ac, the better the change is tolerated. Even large changes, such as the addition of a sugar group via a glycosidic bond can lead to substrates with fairly high product yields if the sugar is attached in the 5 or 6 position (Huang et al., 2007). The range of substrates that NAL can accept is therefore limited to those molecules mostly chemically similar to Neu5Ac.

Organism	Substrate (s)	Rate (% Neu5Ac)	Ref
<i>E. coli</i>	7-Diproylamide-ketoseptose and pyruvate	12	Williams et al., 2005
	<i>N</i> -Acetyl-L-neuraminic acid	1.0	
	3-Deoxy-D-manno-oct-2-ulosonic acid	2.3	
	3-Deoxy-L-manno-oct-2-ulosonic acid	3.7	
	<i>N</i> -Acetyl-L-neuraminic acid	0.0013	Hsu et al., 2005
	3-Deoxy-D-manno-oct-2-ulosonic acid	2.2	
	3-Deoxy-L-manno-oct-2-ulosonic acid	3.8	
<i>C. perfringens</i>	<i>N</i> -Glycoloylneuraminic acid	93	Schauer et al., 1971
	<i>N</i> -Acetyl-4- <i>O</i> -acetylneuraminic acid	22	
	<i>N</i> -Acetyl-5- <i>O</i> -acetylneuraminic acid	97	
	<i>N</i> -Monochloroacetylneuraminic acid	106	
<i>S. scrofa</i>	<i>N</i> -Glycoloylneuraminic acid	55	Schauer and Wember, 1996
	<i>N</i> -Acetyl-9- <i>O</i> -acetylneuraminic acid	32	
<i>S. aureus</i>	Erythrose and pyruvate	65	Windle et al., 2017

Figure 1.8 Table summarising published kinetic data of wild-type NAL enzymes' activity for substrates other than Neu5Ac. Activity is given as a percentage of the wild-type substrate from each publication (not shown).

1.2.3 Mutagenesis of NAL for non-natural substrate catalysis

A number of studies have been performed on NAL in which residues around the active site were mutated, and the effects these had on catalysis were studied. Both randomised and targeted methods of mutation have been

employed in mutational studies of NAL, and these have targeted a number of different amino acids around the active site (figure 1.9 and 1.10).

Residue	Mutated to	Substrate(s)	Source
T48	A	7-diproylaminde-ketoseptose and pyruvate	Williams et al., 2006
L142	R	L-aspartate- β -semialdehyde and pyruvate	Joerger et al., 2003
T167	G/V	7-diproylaminde-ketoseptose and pyruvate	Williams et al., 2006
D191	All	7-diproylaminde-ketoseptose and pyruvate	Williams et al., 2005
E192	All	7-diproylaminde-ketoseptose and pyruvate	Williams et al., 2005
S208	All	7-diproylaminde-ketoseptose and pyruvate	Williams et al., 2005
V251	I	<i>N</i> -Acetyl-L-neuraminic acid, 3-Deoxy-D-manno-oct-2-ulosonic acid, 3-Deoxy-L-manno-oct-2-ulosonic acid	Wada et al., 2003

Figure 1.9 Summary of publications in which residues have been mutated in the active site of *E. coli* NAL, resulting in a change in substrate specificity, and the substrates for which these mutations were made. Other mutations were made in some of these publications, but these are not included here as they were not in the active site.

In one study, error-prone PCR was used in an attempt to switch NAL from accepting *N*-acetyl-D-neuraminic acid to accepting *N*-acetyl-L-neuraminic acid. The activity of wild-type NAL for the L-isomer was too low to allow a screen to be performed, so the screening was performed using D- and L-KDO, for which wild-type NAL displays a low level of activity, 2.3 and 3.7% respectively of the k_{cat}/K_m for Neu5Ac. Three mutations were identified during the rounds of mutagenesis, but only one mutation, V251I, was actually in the active site. This mutation was found in combination with two other mutations, Y98H and F115L. The resultant enzyme displayed a reduced activity for D-Neu5Ac, 61%, but also a small activity for L-Neu5Ac,

0.2%, though this was mostly due to the approximately 100-fold higher K_m . This shows that NAL can be altered to accept substrates with different stereoisomeric features, albeit with a low activity in the mutated enzyme (Wada et al., 2003).

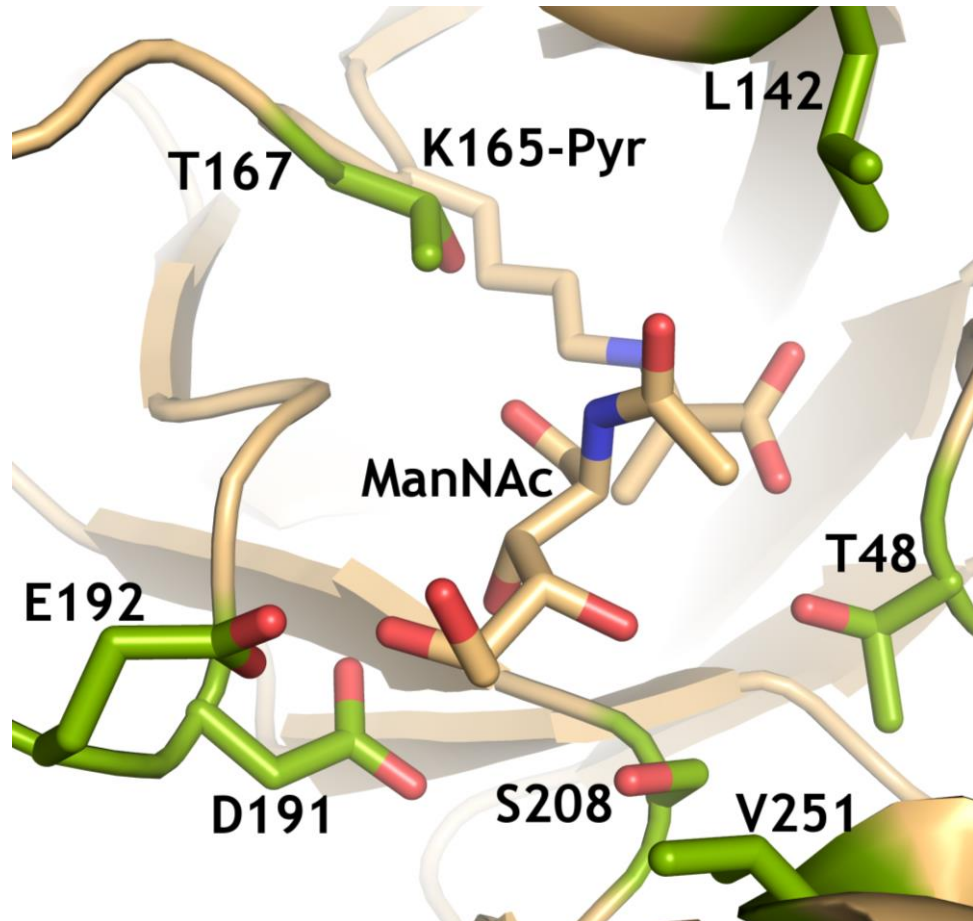


Figure 1.10 View of the active site of *E. coli* NAL showing the mutated residues listed in figure 1.9. Mutations in these residues have all resulted in some change in substrate scope of NAL. The wild-type substrate, ManNAc, is labelled and shown in a binding conformation. The protein backbone is shown as a cartoon in beige, with the mutated residues shown as sticks in green. (PDB: 4BWL).

Mutations to NAL have also increased the activity of the enzyme for dihydrodipicolinate synthesis from pyruvate and L-aspartate- β -semialdehyde. A leucine residue at 142 was identified as conserved in NAL sequences from two different organisms, while the homologous residue in the dihydrodipicolinate synthase sequence of two different organisms is an arginine. Therefore, this mutation was made in the NAL sequence and the activity of L142R NAL for dihydrodipicolinate synthesis was measured to be 19 times higher than that of the wild-type NAL for the same substrate, as measured by comparing the k_{cat}/K_m for L-aspartate- β -semialdehyde. This

activity was high enough to rescue a dihydrodipicolinate synthase activity negative strain but was well below the wild-type activity of dihydrodipicolinate synthase, which was 14,000-fold higher than that of wild-type NAL (Joerger et al., 2003).

In an effort to create a partially biocatalytic route to sialidase inhibitors, a number of residues were identified that could create a hydrophobic pocket capable of accommodating a dipropylamide group in an analogue of a sialidase inhibitor related to Zanamivir (Williams et al., 2005). Saturation mutagenesis was performed on each of the residues D191, E192 and S208, giving three separate libraries. Only mutations at E192 had a major effect on the activity of NAL on the dipropylamide analogue (DPA). Almost every E192 mutant displayed almost no activity for Neu5Ac, while some mutants showed an approximately 9-fold increase in activity for DPA. The protein variant that displayed the highest activity towards DPA was E192N. When the $k_{\text{cat}}/K_{\text{m}}$ values of the wild-type and E192N variant were examined, E192N had a $k_{\text{cat}}/K_{\text{m}}$ value approximately 13-fold lower than wild-type for the natural substrate, Neu5Ac, and a $k_{\text{cat}}/K_{\text{m}}$ value approximately 50-fold higher for DPA. This enzyme variant was also more widely applicable to other substrates with different hydrophobic groups on the amide group.

These data show that NAL is an adaptable biocatalyst with significant applied functionality to the industrial scale synthesis of molecules of interest. Modification of NAL using non-canonical amino acids may allow an even greater range of molecules to be synthesised with this enzyme.

1.3 Structural features of *Staphylococcus aureus* Neuraminic acid lyase

1.3.1 Overall structure

Staphylococcus aureus Neuraminic acid lyase (saNAL) is very similar in structure to that of the well-studied *E. coli* enzyme. The folded, active protein is composed of four identical monomers of 34 kDa forming a homotetramer of 136 kDa in roughly the shape of a torus, doughnut or diamond of monomers with a hole in the middle of the tetramer (figure 1.11).

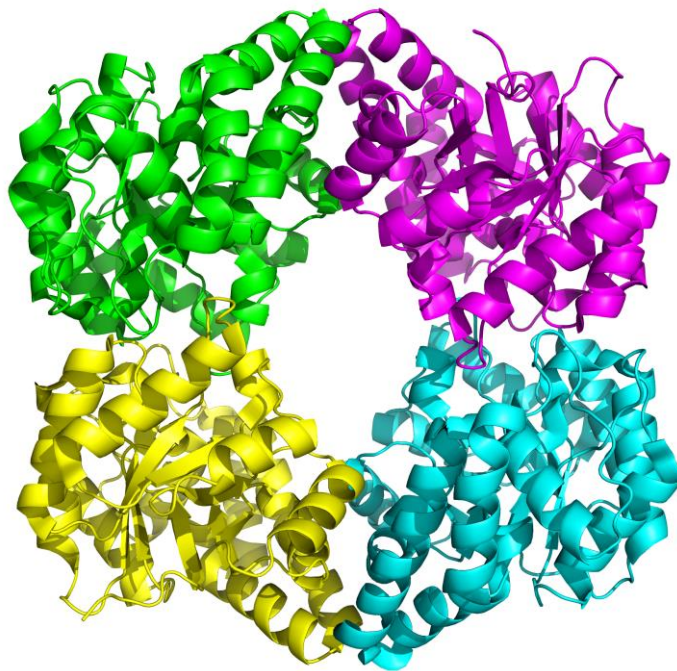


Figure 1.11 View of the saNAL tetramer. Each monomer of the tetramer is differently coloured. This view highlights the hole in the ring of monomers

The core of each monomer is an α/β triosephosphate isomerase-type barrel, consisting of 8 α -helices and 8 parallel β -strands. The α/β -barrels are roughly in line with the plane of the ring, with the lumen of each monomer directed at an angle of approximately 45° from the plane of the ring (figure 1.12). Outside of this core structure, towards the C-terminal end of the amino acid sequence, there is a bundle of three helices, providing residues to both the interaction interfaces of the monomers and also to the active site (figure 1.12). The active site opening of each monomer points towards the hole in the centre of the ring of tetramers. An active site lysine projects into the lumen of the α/β -barrel, towards the centre of the ring of monomers, presenting the lysine side chain amine

group at the inside edge of the barrel, aligned with the ends of the β -strands.

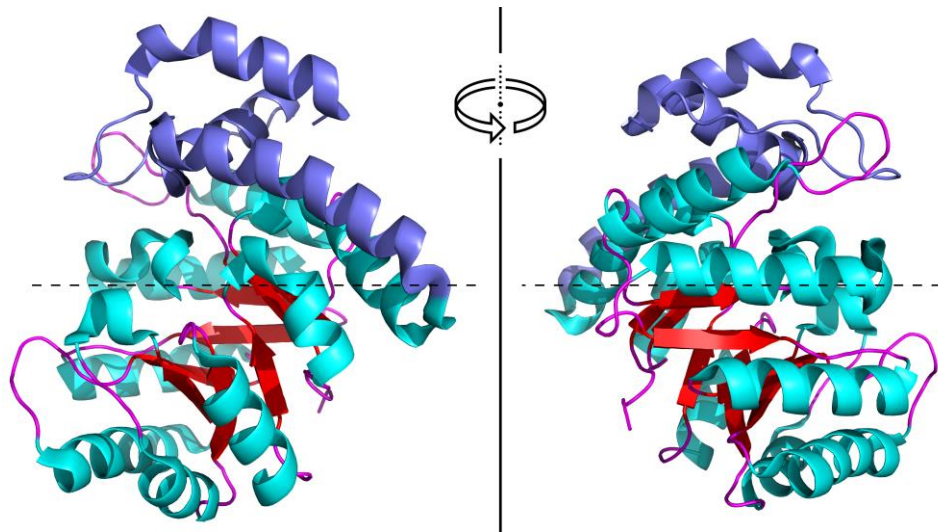


Figure 1.12 Two views of a single monomer of *saNAL*, rotated by 180° around a vertical axis. The light blue, magenta and red regions of the structure represent the α -helices, loops and β -sheets of the α/β -barrel respectively, while the dark blue region highlights the three helix bundle on the outside of this barrel structure. The dashed line across each structure represents the plane of the ring of monomers.

The hole in the tetrameric ring between the monomers is approximately 22 \AA wide, but the width of the active site changes upon pyruvate binding. Before the pyruvate binds, residues 139 to 146 lack any conserved structure, giving the active site a relatively large opening. However, after the pyruvate binds, the active site becomes more narrow, as this unstructured loop in the protein from residues 139 to 146 becomes rigid and partially occludes the active site, making the opening approximately 10 \AA by 10 \AA wide (figure 1.13).

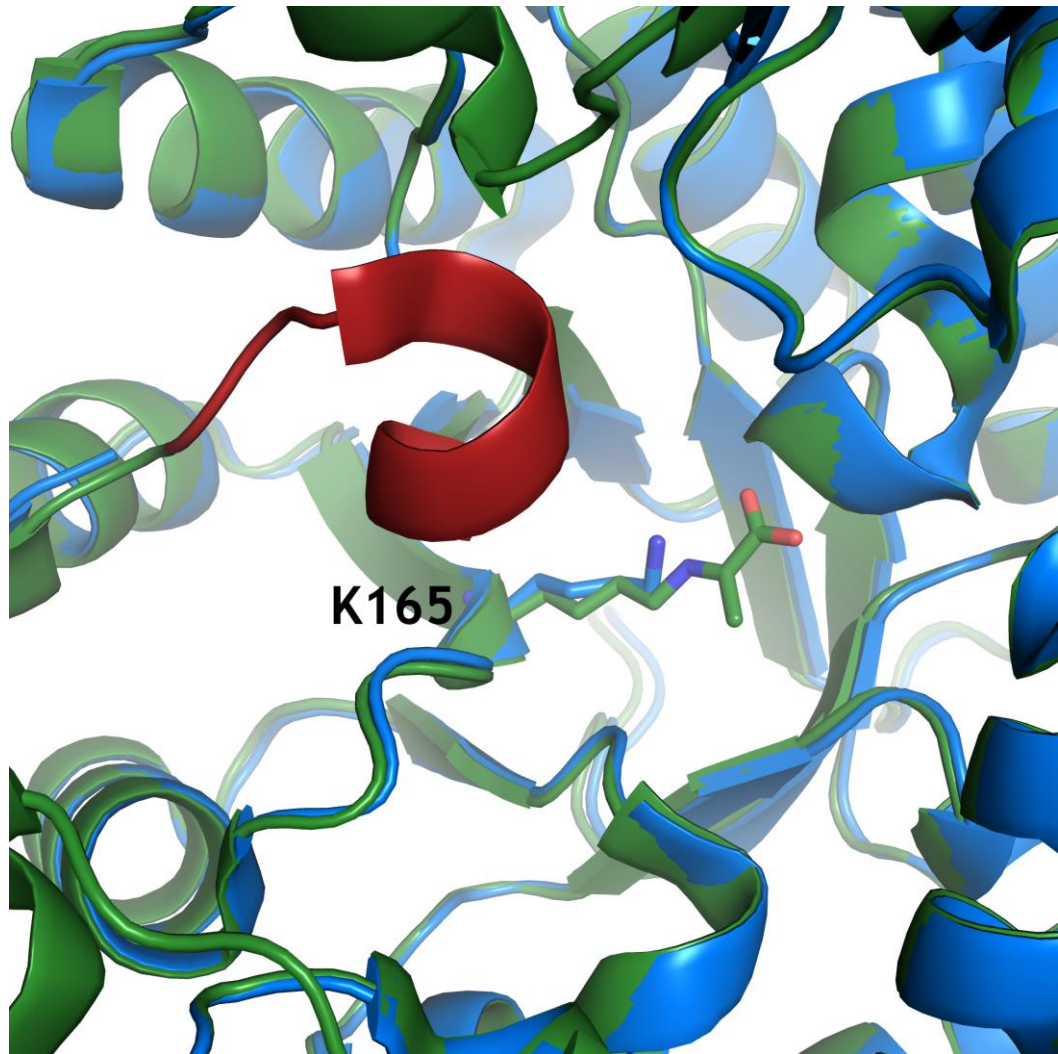


Figure 1.13 View of the active site of saNAL with and without pyruvate bound to K165. In blue is a cartoon representation of the protein backbone without pyruvate bound, in green is a cartoon representation of the protein backbone with pyruvate bound, and in red is the section of the protein backbone that becomes rigid upon pyruvate binding. The active site lysine, at the base of the active site, is shown as sticks and coloured green or blue corresponding to whether it does or does not, respectively, have pyruvate bound.

1.3.2 Mechanism of action of Neuraminic acid lyase

The mechanism of action of NAL has been the subject of some previous study, most notably the *Haemophilus influenzae* and *E. coli* variants. Two main hypotheses have been suggested for the mechanism by which NAL catalyses the reversible aldol reaction between *N*-acetyl-D-mannosamine and pyruvate. Due to the lack of a histidine residue in the active site to act as a general base, one of the hypotheses suggested that formation of the Schiff base with the full-length substrate allowed the carboxylate group of

the substrate to attack the aldol proton on the fourth carbon (figure 1.14 A) (Smith et al., 1999). An alternative was also proposed in which the highly conserved Y137 residue serves to aid in the abstraction of the proton during catalysis. This was mostly based on X-ray crystallography data with substrate analogues in the active site (figure 1.14 B) (Barbosa et al., 2000).

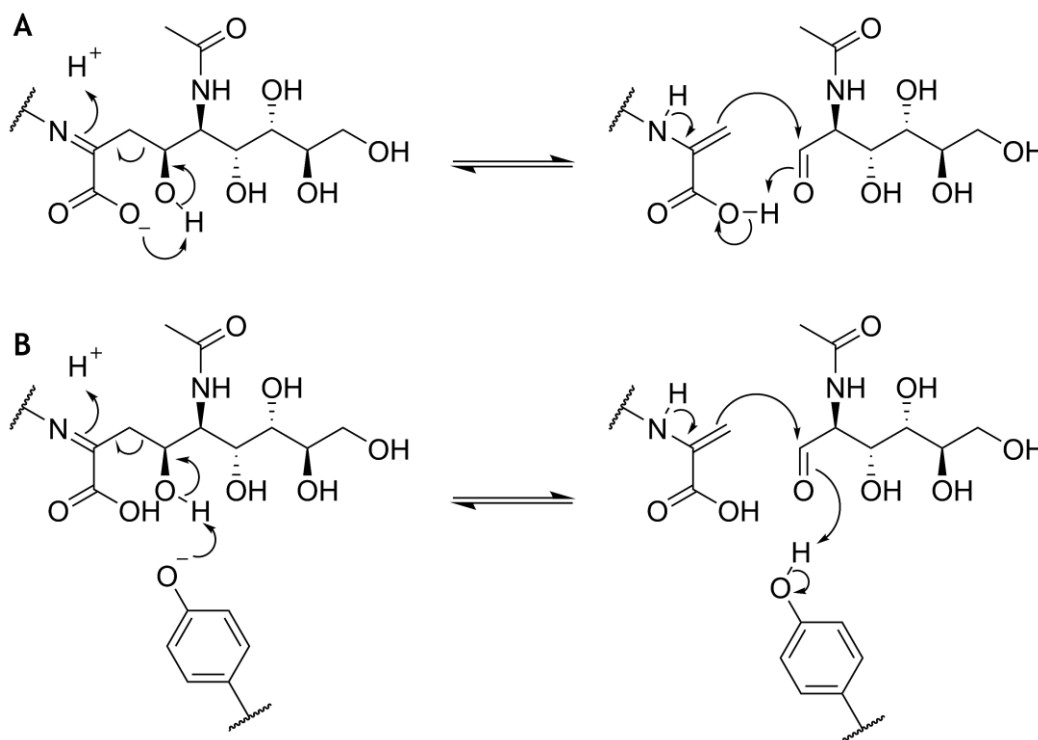


Figure 1.14 Competing hypotheses of the mechanism of NAL. **A** The substrate catalysed mechanism by which the pyruvate carboxylate group is able to catalyse the reaction after formation of the Schiff base. **B** The protein catalysed method by which Y137 is instrumental in the catalysis, either as a proton donor or acceptor.

More recently, quantum mechanics/molecular mechanics (QM/MM) studies in conjunction with a fortuitous crystallography experiment have shed more light on the mechanism of NAL (Daniels et al., 2014). The highly conserved tyrosine, Y137 in *ecNAL* was mutated to alanine, inactivating the enzyme as far as could be measured. When this mutant was crystallised and Neu5Ac soaked in, electron density was found connected to the active site lysine in all monomers. In three of the four monomers, the density fit the shape of Neu5Ac, but in one of the monomers, the density connected to the enzyme was the size and shape of pyruvate, with a region of density beyond the pyruvate matching the shape of ManNAc. Therefore, in one structure was represented the structure of both the substrates and product just before and just after the reaction, respectively.

The ManNAc structure was used in QM/MM simulations to determine proton movements during catalysis. QM/MM studies are *in silico* experiments using two different sets of equations to describe the atoms. Firstly, a small region around the bond-formation region is simulated using quantum mechanical equations, allowing atoms to move and for bonds to form and break if energetically favourable to do so. A larger region around this is simulated with molecular mechanical equations, which allow atoms to move and bonds to stretch, but not break. This is done as quantum mechanical simulations are much more costly in terms of processing power than molecular mechanical simulations, and so a balance is struck between the number of atoms simulated by the more accurate quantum mechanical equations and the number of atoms simulated by the less computationally expensive molecular mechanical equations.

ManNAc, the pyruvate-Schiff base and the tyrosine side chain were simulated quantum mechanically, meaning that each of the previously suggested hypotheses could occur, either substrate-assisted catalysis or catalysis involving the conserved tyrosine. In all simulations, the proton from the tyrosine side chain transferred onto the aldehyde oxygen of the ManNAc, showing that this residue is the proton donor in the wild-type NAL reaction. The mechanism of the synthetic aldol reaction was thus found to start by attack of the enamine form of the pyruvate-Schiff base onto the aldehyde group of ManNAc. This forms a negatively charged oxyanion transition state which is stabilised by the hydroxyl group of T167. Y137 then donates its hydroxyl proton to the oxyanion of the substrate. Deprotonated Y137 is subsequently stabilised by a hydrogen bonding network consisting of the hydroxyl groups from S47 and Y110.

1.3.3 Chemical modification of *sa*NAL

Given the ample study of NAL that has taken place and been described here, previous work from our group has explored the potential of non-canonical amino acids (ncAAs) in the active site of NAL. Side chains with different chemical groups than those found in the canonical 20 amino acids should allow a greater range of activities to be performed using NAL derivatives as an enzyme, with the ncAA acting as an artificial PTM. Insertion of ncAAs around the active site could allow for more exotic substrates to be catalysed or may allow for a greater rate enhancement for some molecules for which NAL already has some background activity. Alternatively, if the ncAA is inserted in place of the catalytic lysine residue at position 165 then the mechanism of the reaction the enzyme performs

could even be modified, changing the enzyme from an aldolase to another enzyme.

A chemical modification method was chosen for insertion of the ncAAs into the protein, based on the method discovered in the Davis lab (Chalker et al., 2012), and has been further developed since. The method first requires insertion of a cysteine residue in the position to be modified. All cysteine residues will be modified in the final protein, so a cysteine-free version of the protein is required. Some work was performed previously in the lab in an attempt to remove the cysteine residues from the well-studied *E. coli* NAL to enable its use in modification experiments, but the removal of all the cysteine residues rendered the protein insoluble when purified. A naturally cysteine-free variant of NAL was found in *Staphylococcus aureus* and due to this, all modification experiments have been performed on *sa*NAL. This variant is structurally highly homologous to *ec*NAL (RMSD=1.43 Å), especially around the active site, so results from studies performed on the *E. coli* protein should still be relevant for the *S. aureus* protein.

Once the cysteine has been inserted in the desired position, the protein can be expressed and purified. For most positions on the enzyme, especially in the active site, the protein will need to be unfolded prior to modification to allow the modification reagent adequate access to the amino acid. *Sa*NAL has been shown to tolerate folding and unfolding very well (Timms et al., 2013), allowing this enzyme to be used in ncAA modification. The protein is then treated using 2,5-dibromohexandiamide, which performs a bis-alkylation-elimination reaction to produce a dehydroalanine residue in place of the cysteine. This dehydroalanine residue can then be selectively targeted by a thiol, performing a Michael addition onto the protein. This forms an ncAA that is a cysteine derivative with a side chain that has a carbon atom at the β position, a sulphur atom at the γ position, and whatever was attached to the thiol used to perform the Michael addition at positions further down the chain (figure 1.15).

This chemical incorporation method for ncAAs was chosen over the genetic incorporation method for a number of reasons. Firstly, genetic incorporation of a ncAA requires an associated tRNA/tRNA synthetase pair to be evolved. For this evolution to take place the ncAA of interest must be homologous enough to a canonical amino acid for the tRNA synthetase pair to be evolvable. A recent review of the literature regarding genetic incorporation of ncAAs into enzymes can be found here: (Agostini et al.,

2017). As such, most ncAAs inserted using this method are homologues of canonical amino acids (Wang et al., 2006). Evolving a tRNA/tRNA synthetase pair takes time, and requires knowledge of which ncAAs will produce the desired effects before studying them. This is ideal for investigations into what a specific side chain will do in many different positions, but it far from ideal if a number of side chains are to be tested and their potential function in the protein unknown. Each ncAA would need its own tRNA/tRNA synthetase pair to be evolved, and most or all of these ncAAs could be uninteresting, which would mean a lot of wasted effort.

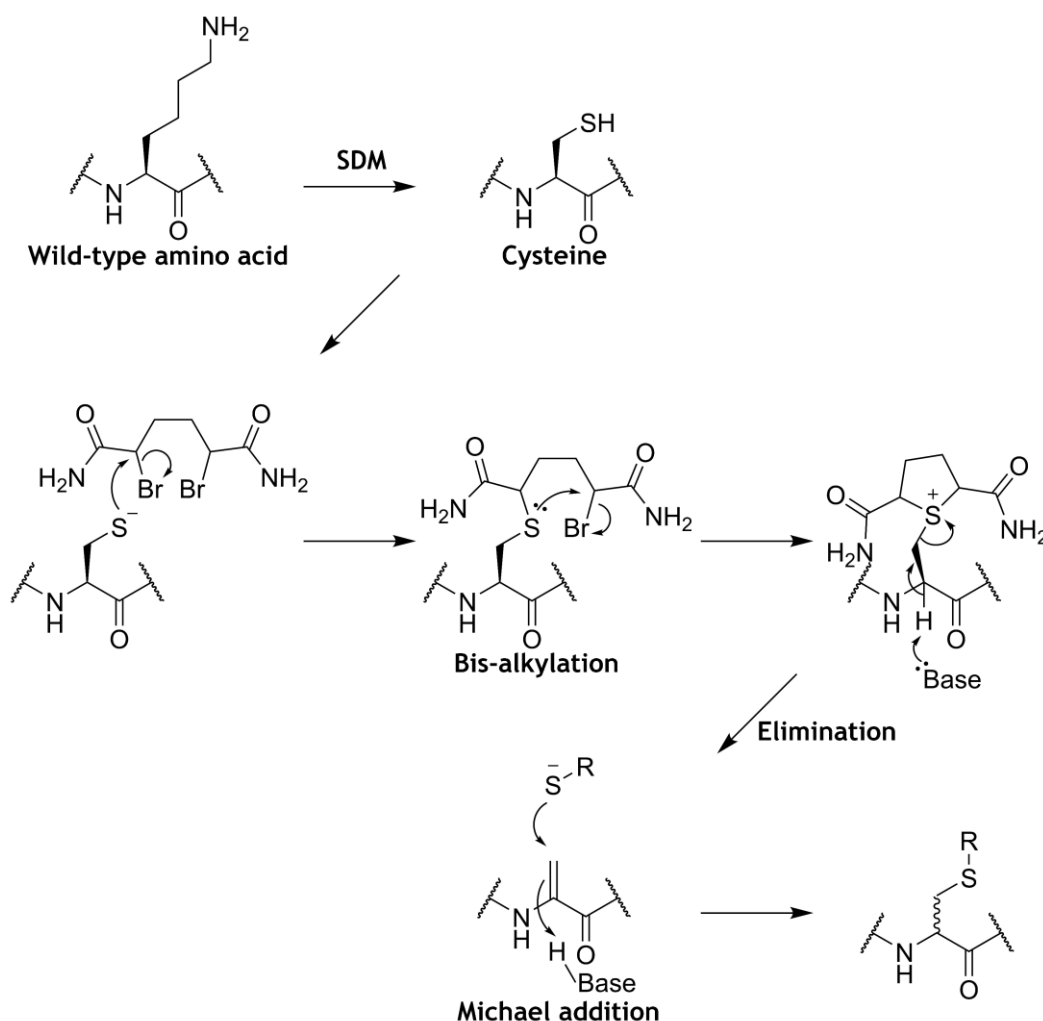


Figure 1.15 The mechanism by which the ncAAs are formed *in situ* on the protein backbone. The wild-type amino acid shown is lysine, but this same method can be used on any of the canonical amino acids. SDM in the figure stands for site-directed mutagenesis. The stereochemistry at the α -carbon of the resulting amino acid is uncertain as there is no reason to think that the Michael addition will be stereochemically selective in solution on unfolded protein.

Additionally, *saNAL* expresses quite well, giving yields of 60 mg per litre of culture medium, or higher, without complicated expression and purification methods. This allows large amounts of protein to be generated, which can then be modified into large amounts of ncAA-containing enzyme. In contrast to this, due to the presence of both ncAA-encoding tRNA and stop-encoding tRNA with the same codon sequence in some organisms used in genetic incorporation, a fraction of each initiated translation will have a stop tRNA inserted at the ncAA position, resulting in a truncated protein (Zhang et al., 2004). There has however been some effort recently to mitigate this by developing strains of *E. coli* in which every single amber stop codon, and the release factor for the amber stop-codon, have been removed from the genome (Lajoie et al., 2013).

Modifying the protein to contain the ncAA only after purification leads to another advantage of the chemical modification method. Some ncAAs will be toxic to living cells (Fowden et al., 1967), either due to their chemical activities or, if similar enough to canonical amino acids, their acting as inhibitors for cellular processes requiring the canonical amino acids (Coggin and Martin, 1965). When using genetic incorporation, ncAAs are most often fed to the cells, meaning these toxic ncAAs would effectively be inaccessible using this method. However, if chemical incorporation is used, live cells never encounter the ncAAs, so their toxicity is not important. This increases the chemical groups available for use on ncAA side chains.

A drawback of using chemical incorporation, including the method used in this thesis, is that there will be some restrictions to the ncAAs incorporated due to the method of incorporation. The method used here requires the reinsertion of a sulphur in the γ -position of the side chain as a thiol is used to react with the dehydroalanine. Other methods of chemical incorporation do not include this sulphur but have other restrictions on the types of side chains possible. For example, a related method to the one used here also uses dehydroalanine as its reactive residue, but inserts the ncAA side chain using a zinc catalysed reaction between the alkene and an iodo-molecule (Wright et al., 2016). The iodine is lost in solution, meaning the ncAA produced consists of alanine directly attached to the group that was attached to the iodine. While being less restrictive in the ncAAs that could be made with this method, the requirement for a low oxygen environment necessary to prevent unwanted side reactions is onerous and less practical than the method used in this thesis.

Interestingly, a recent development combines the genetic and chemical ncAA incorporation methods (Yang et al., 2016). This method uses genetic incorporation to introduce a phosphoserine into the protein backbone. This can then be dephosphorylated to a dehydroalanine under relatively mild conditions, and the resulting dehydroalanine residue can be chemically converted into another ncAA using the methods described above.

Initial work in the lab was to show that active enzyme could be recovered even after modification. As mentioned, work was done to show that the wild-type *saNAL* enzyme successfully refolded after unfolding with urea, which allowed residues deep in the active site to be targeted for modification. Subsequently, site-directed mutagenesis (SDM) was performed on *saNAL* to change the active site lysine at position 165 into a cysteine. This was then modified with aminoethanethiol to form a sulphur-containing lysine mimic, γ -thialysine (Thl). The activity of this K165Thl enzyme was measured at the wild-type enzymes optimal pH, 7.4 and at other pH values to find a new optimal pH of 6.8 (figure 1.16).

pH	Enzyme	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($\text{min}^{-1}\text{mM}^{-1}$)	% wild-type
7.4	Wild-type	250 ± 5	2.2 ± 0.1	114	100
	K165C	0.08 ± 0.004	0.8 ± 0.1	0.100	0.088
	K165Thl	26 ± 0.9	1.4 ± 0.2	18.6	16
6.8	Wild-type	260 ± 6	2.4 ± 0.2	108	95
	K165Thl	29 ± 0.8	0.9 ± 0.1	32.2	28

Figure 1.16 Activity of wild-type *saNAL* compared to K165thialysine (Thl) in buffers at different pH. The optimum pH of the wild-type enzyme was 7.4, while the optimum of the modified enzyme was determined to be 6.8. The modified enzyme is approximately 2-fold more active at pH 6.8 than 7.4, while the wild-type enzyme's activity is only slightly reduced by the same pH change.

The modified enzyme had a lower activity than the wild-type enzyme, and a shifted pH optimum. Both these effects were hypothesised to be due to the sulphur in the γ -position of the side chain having an effect on the ionisation state of the terminal amino group. This effect is thus very site-specific, replacing any other amino acid in the active side with a γ -thia-analogue would not be expected to have the same effect on activity, as the effect of the sulphur on the electrostatics would have

different consequences for activity based on the side chain and its position in the active site.

The modified enzyme was also crystallised, showing that it is able to fold and have a stable structure. The structure itself was similar to that of the wild-type enzyme, with the amino group of K165Thl in a very similar position to that of 165. The carbon-sulphur bonds in the side chain are longer than the carbon-carbon bonds in the canonical side chain, but the angle between the bonds is more acute making the overall carbon-sulphur-carbon distance almost identical to that of the carbon-carbon-carbon distance in the wild-type (Timms et al., 2013).

This showed that active enzyme could be produced using the chemical modification method, and the observed reduction in activity seen would not necessarily occur when the modifications were made at different positions around the enzyme. Indeed, if a different substrate and a different residue in the active site is selected, the same effect that sulphur had on the electrostatics of the side chain that caused a reduction in activity at K165 may cause an increase in the activity when compared to an amino acid without the sulphur.

A number of different positions were identified around the active site of *saNAL* by first identifying residues in close proximity to the substrate in a structure of an inactive variant of *ecNAL* with *N*-acetylneuraminic acid bound (Windle et al., 2017). The residues homologous to these in the *saNAL* structure were changed to cysteine residues using SDM. These mutants were all modified with a range of different thiols to give a large number of different enzymes, each with a single nCAA in the active site. These were tested against a range of aldehyde substrates, in combination with pyruvate, to see if any had an effect on activity.

Multiple active enzymes were found, and a number had activities for a substrate higher than the wild-type enzyme had for that substrate, with the best enzyme-aldehyde combination being F190 modified into a 2,3-dihydroxypropyl cysteine (Dpc) (figure 1.17) performing the synthetic aldol reaction between pyruvate and erythrose to make 3-deoxy-2-heptulosonic acid. This nCAA is not similar to any of the canonical amino acids and is very different from phenylalanine, the wild-type amino acid at that position. Additionally, erythrose is a four carbon sugar compared to the 6 carbon amino-acylated sugar that is the wild-type substrate.

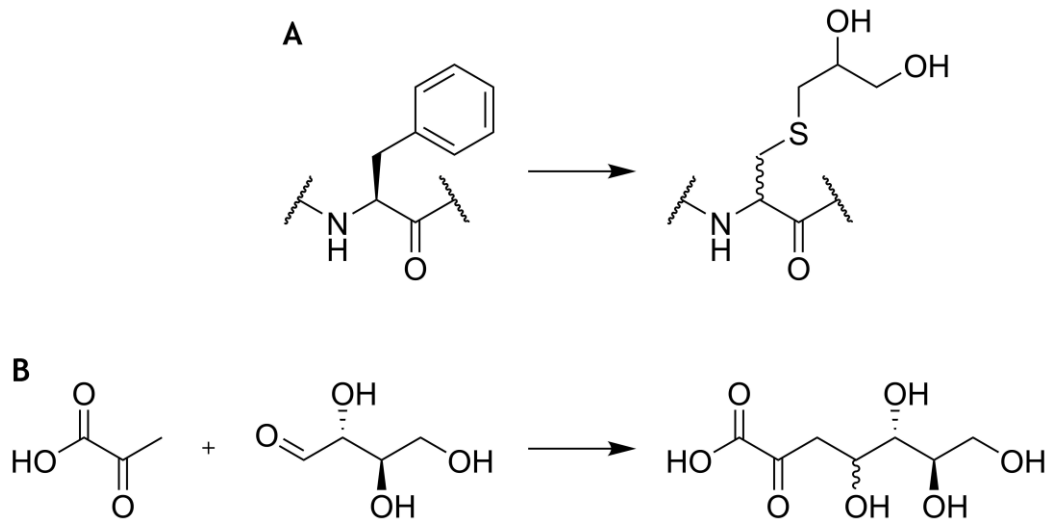


Figure 1.17 A The modification from phenylalanine to 2,3-dihydroxypropylcysteine (Dpc) performed on *sa*NAL to produce the enhanced rate of the reaction shown in B, between pyruvate and erythrose making 3-deoxy-2-heptulosonic acid.

This enzyme was also crystallised, again showing that ncAAs produced using this method can produce stable, foldable enzymes, and it was found that the modified side chain formed a hydrogen bonding network between E192 and D141, both on opposing sides of the side chain (figure 1.18). This kind of interaction would not be possible with any of the canonical amino acids, and as such, all of the canonical amino acids were inserted into the active site at 190 and none of them came close to the activity of the ncAA.

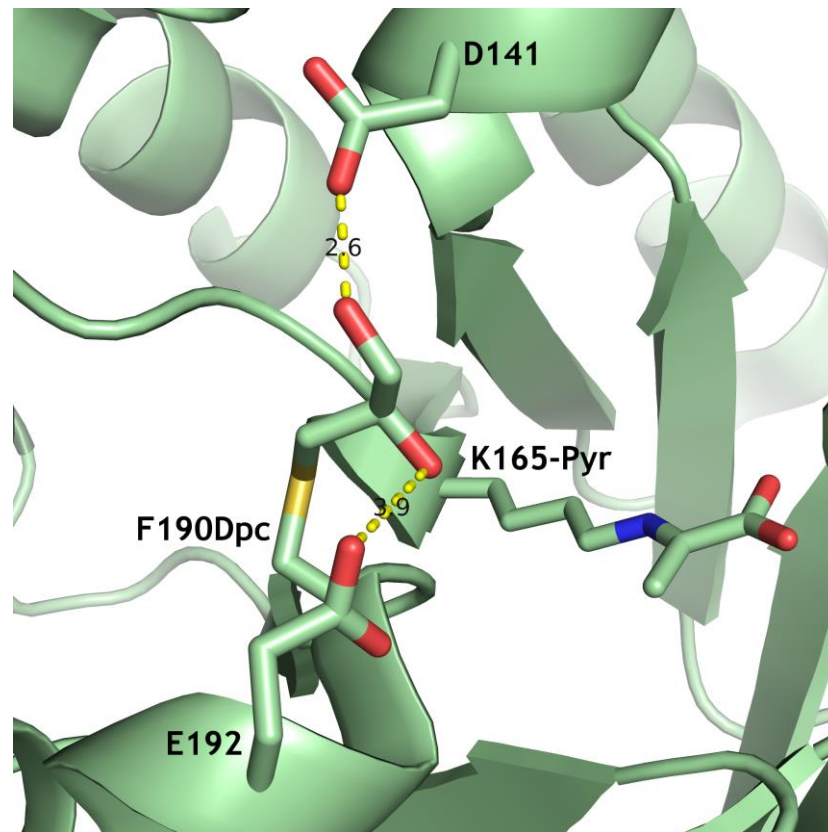


Figure 1.18 Hydrogen bonding network cause by insertion of a 2,3-dihydroxypropyl cysteine (Dpc) side-chain at position 190 of *saNAL*. The residues involved in the network have their side chains shown, as does K165, for orientation. The hydrogen bonds are shown as yellow dashed lines, and their distances in Ångstroms are given in the figure.

Using *in silico* simulations of the modified and wild-type active site, it was found that the nCAA-induced hydrogen bonding network had a bracing effect on the erythrose molecule, holding it in the active site and closer to the pyruvate-Schiff-base at K165. The wild-type enzyme, unable to form the hydrogen bonding network, did not provide a brace or block for the erythrose, allowing it to move more freely and so spend less time in the active site next to the pyruvate (Windle et al., 2017).

Due to these findings in the lab, it was decided that the modification method shown above would continue to be used to assess the effect of nCAAs on the substrate specificity of *saNAL*. The specific aims of the project were to use a number of thiols to modify multiple positions around the active site of *saNAL* and then to use all the enzymes produced to screen for synthetic activity between pyruvate and a diverse array of aldehydes. Any hits would then be taken forward, have their activity probed in further detail and be structurally examined to determine the causes behind the activity enhancement.

Chapter 2 - Materials and Methods

2.1 Materials

The bacterial strain used for expression was *E. coli* BL21(DE3) {B⁻ F⁻ *dcm ompT hsdS*(r_B⁻m_B) *gal* λ(DE3)}. The bacterial strain used for DNA storage was *E. coli* XL10-Gold {*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte* Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 tet^R F⁺[*proAB lacI^qΔM15 Tn10*(Tet^R) Amy Cam^R]}.

2,5-Dibromohexanediamide was synthesised during the course of this project by the author, according to the method given in section 2.5. Protein molecular weight markers, TEMED, acrylamide, yeast extract, tryptone, aminoethanethiol, dibasic sodium phosphate, monobasic sodium phosphate, urea and N-acetylneuraminic acid were purchased from Sigma Chemicals Limited (Dorset, UK). Methanol, glacial acetic acid, ethanol, glycerol, Tris base, imidazole, β-mercaptoethanol, glycine, SYBRsafe DNA gel stain, ampicillin, ammonium acetate, sodium chloride, dialysis tubing and Zeba spin desalting columns were purchased from Thermo Fisher Scientific Limited (Loughborough, UK). Hydrochloric acid was purchased from Acros Organics (Geel, Belgium). IPTG was purchased from Melford Laboratories Limited (Suffolk, UK). Sodium dodecyl sulphate was purchased from BDH Chemicals Limited (Dorset, UK) Ammonium persulphate was purchased from Amersham Biosciences (Buckinghamshire, UK). Bradford reagent was purchased from Bio-Rad (California, USA). Bovine Serum Albumin (BSA) was purchased from Thermo scientific (Massachusetts, USA). Chelating sepharose resin was purchased from GE Healthcare (Illinois, USA). Instant Blue quickstain was purchased from Expedeon (California, USA). Amicon Ultra-15 concentrators were purchased from Merck Millipore (Massachusetts, USA). Miniprep kits were purchased from Promega (Wisconsin, USA). Quikchange II kits were purchased from Agilent (California, USA). Primers were ordered from Integrated DNA technologies (Iowa, USA) Sequencing was performed by Genewiz (New Jersey, USA). Cell disruption was performed using a Cell Disruptor from Constant Cell Disruption Systems (Northamptonshire, UK).

2.2 General methods

2.2.1 2xTY media production

2xTY media was made by the following method: 16 g of tryptone, 10 g of yeast extract and 5 g of sodium chloride was added to a container and made up to 1 l with RO water. Media was then autoclaved at 121 °C for a minimum of 20 minutes

2.2.2 Agar plate production

The agar plates used were 2xTY agar plates and made by addition of 16 g of tryptone, 10 g of yeast extract, 5 g of salt and 15 g agar to a container at least 2 l in volume, and made up to 1 l with RO water. Media was then autoclaved at 121 °C for a minimum of 20 minutes, and the solution left to cool to 55 °C. A number of plates were poured to provide a non-selective growth sample if needed. Ampicillin was then added to the remaining solution create a final concentration of 50 µg/ml. The resulting antibiotic containing agar solution was then poured into plates to form selective growth substrate.

2.2.3 Glycerol stock generation

Glycerol stocks were generated by picking a single colony from a plate and incubating it in 5 ml 2xTY media at 37 °C with shaking at 200 rpm for 16 hours. From the resulting solution, 500 µl was taken and added to a sterile cryo-vial containing 500 µl of 50% glycerol. The cryo-vial was mixed thoroughly and stored at -20 °C.

2.3 DNA methods

2.3.1 DNA extraction and purification

DNA was extracted from cultures of *E. coli* cultures that were started either by picking from a single colony, or by taking 5 µl of glycerol stock and adding the cells to 5 ml 2xTY media. The media was incubated for 16 hours at 37 °C with shaking at 200 rpm. Subsequently, a mini-prep was performed according to the manufacturer's instructions (Wizard Plus SV minipreps). DNA stocks were stored at -20 °C.

2.3.2 Site-Directed Mutagenesis

DNA template used was produced using the method given in section 2.3.1. Primers were designed using a tool on the Agilent website (<http://www.genomics.agilent.com/primerDesignProgram.jsp>) as appropriate for the Quikchange II kit and to produce the desired mutation and ordered from Integrated DNA Technologies. Codons were chosen as the most used codon in *E. coli* for that amino acid. The primers and Quikchange II kit was used according to manufacturer's instructions. Resulting colonies were grown up, stored as glycerol stocks (section 2.2.3) and sequenced (section 2.3.4) to confirm the expected mutation had been inserted.

2.3.3 Agarose gel electrophoresis

Agarose gels were made by addition of 1 g of agarose to 100 ml of 40 mM Tris base, 20 mM acetic acid and 1 mM ethylenediaminetetraacetic acid (EDTA). This was then microwaved to heat the solution until the agarose was fully dissolved. The solution was then allowed to cool to 60 °C and 10 µl SYBR safe DNA gel stain was added and mixed in thoroughly. The gel was poured as needed and left to cool. The gel was loaded into an electrophoresis tank according to manufacturer's procedures, samples loaded and run at 10 V/cm for a time appropriate for the size of sample being studied.

2.3.4 Sequencing

Samples of 15 µl of DNA solution at 100 ng/µl in DNase free water were sent to Genewiz (NJ, USA) and sequencing was performed by Genewiz.

2.4 Protein methods

2.4.1 Protein expression

Glycerol stocks or single colonies of *E. coli* BL21 (DE3) strains were used to inoculate 5 ml 2xTY starter cultures supplemented with 50 µg/ml ampicillin. Starter cultures were incubated for 16 hours at 37 °C and shaken at 200rpm in an orbital incubator. 100 µl of this culture was then used to inoculate 100 ml 2xTY media supplemented with 50 µg/ml ampicillin and left to incubate for 8 hours at 37 °C, 200 rpm in an orbital incubator. The 100 ml culture was used to inoculate 2xTY media supplemented with 50 µg/ml ampicillin, with 10 ml of culture used for each litre of media. When the culture had reached an optical density at 600 nm (OD_{600nm}) of 0.6, IPTG was added to a final concentration of 0.1 mM to induce protein expression, and the cells were grown for a further 14 hours.

2.4.2 Protein purification

Hexahistidine tagged proteins were purified using a batch purification method with Chelating Sepharose™ fast flow resin. Cell pellet was obtained from 1 l of bacterial culture by centrifugation at 9 000 g for 20 min. 40 ml of washing buffer (50 mM Tris.HCl, 20 mM imidazole, 0.5 M NaCl, pH 7.4) was used to re-suspend the cell pellet and the pellet was homogenised using a glass homogeniser. The re-suspended cell pellet was lysed as per manufacturers guidelines using a Cell Disruptor. The lysed cells were then centrifuged at 29 400 g for 45 minutes at 4 °C. The supernatant containing the soluble protein fraction was then loaded onto 5 ml of Chelating Sepharose™ fast flow resin. The tagged protein was allowed to bind for 30 minutes with gentle agitation at 4 °C. Centrifugation at 3 000 g for 5 minutes at 4 °C produced supernatant containing non-bound proteins, which was discarded. The resin was then washed with 40 ml wash buffer four times, with thorough mixing and then centrifugation at 3 000 g for 6 min at 4 °C after which the supernatant was removed. Elution of the 6xHis tagged protein was achieved by addition of 20 ml elution buffer (50 mM Tris.HCl, 0.5 M imidazole, 0.5 M NaCl, pH 7.4) and incubation for 20 minutes with agitation at 4 °C. The protein was then isolated from the resin by centrifugation at 3 000 g for 5 minutes at 4 °C followed by removal of the supernatant. A further 20 ml of elution buffer was added and the resin incubated at 4 °C for a further 20 minutes. The remaining protein was then isolated from the resin by centrifugation at

3 000 g for 5 minutes at 4 °C, and the supernatant was removed once again. Both elution steps were then dialysed (section 2.4.8) separately against 50 mM ammonium acetate pH 7.0 and freeze dried (section 2.4.11).

2.4.3 Saturation library protein expression

Glycerol stocks of *E. coli* BL21 (DE3) strains were used to inoculate 5 ml 2xTY starter cultures supplemented with 50 µg/ml ampicillin. Starter cultures were incubated for 16 hours at 37 °C and shaken at 200rpm in an orbital incubator. 500 µl of this culture was then used to inoculate 50 ml 2xTY media supplemented with 50 µg/ml ampicillin and left to incubate at 37 °C, 200 rpm in an orbital incubator until the culture reached an OD₆₀₀ of 0.6, approximately 2 hours. IPTG was added to a final concentration of 0.1 mM to induce protein expression, and the cells were grown for a further 6 hours. The 50 ml cultures were then spun down in a falcon tube for 20 min at 3500 g, the supernatant poured off, and the pellets were frozen at -20°C.

2.4.4 Saturation library protein purification

4 ml lysis buffer (50 mM sodium phosphate, 100 mM sodium chloride, 2 mM magnesium chloride, 0.2 mg/ml lysozyme, 0.05 mg/ml DNase, pH 8.0) was added to the frozen pellet from section 2.4.3, and was resuspended and incubated on ice for 1 hour. The solution was then sonicated for 40 s five times to complete lysis. Lysed cells were then spun for 60 min at 4900 g. The soluble fraction was loaded onto a disposable, gravity elution column packed with 1 ml nickel bound Chelating Sepharose™ fast flow resin and incubated for 15 min with agitation every 5 min. The resin was washed with 30 ml washing buffer (50 mM Tris.HCl, 20 mM imidazole, 0.5 M NaCl, pH 7.4). Once all wash buffer had run through, 1 ml elution buffer (50 mM Tris.HCl, 0.5 M imidazole, 0.5 M NaCl, pH 7.4) was added and left to incubate for 5 min. The elution was allowed to run out of the column and collected. Another 1 ml of elution buffer was added and the process repeated 3 times to give a total of 4 fractions. The protein concentration of each fraction was measured (see section 2.4.5) and the highest concentration fractions were dialysed (section 2.4.8) into 50 mM sodium phosphate, pH 7.4.

2.4.5 Bradford assay

Bradford assay standard curve was constructed using a Bovine Serum Albumin standard solution provided at 2 mg/ml. This was diluted to a

range of concentrations using 50 mM Tris.HCl buffer at pH 7.4. Three cuvettes were prepared containing 980 μ l Bradford reagent and 18 μ l 50 mM Tris.HCl at pH 7.4 for each protein concentration of the standard and for blanks. Each cuvette then had 2 μ l of the appropriate sample added, was inverted 4 times and incubated for ten minutes. After which the OD₅₉₅ was recorded for each sample. A standard curve was constructed to find the equation of the line of absorbance against protein concentration, and this equation was rearranged to allow future calculation of protein concentration without running a contemporaneous standard curve. The equation used is given below:

$$\text{protein conc. (mg/ml)} = \frac{\text{Sample (OD}_{595}) - \text{Control (OD}_{595})}{0.16356}$$

2.4.6 SDS-PAGE gel electrophoresis

Gels were made up containing the components listed in Table 2.1. Protein samples were prepared for SDS PAGE by adding an equal volume of 2x loading buffer (100 mM Tris.HCl, 140 mM sodium dodecyl sulphate (SDS), 3 mM bromophenol blue, 200 mM dithiothreitol, 2.75 M glycerol, pH 6.8) and incubating at 95°C for 5 minutes before being loaded onto the gel. The gel was surrounded by running buffer (190 mM glycine, 25 mM Tris base, 3.5 mM SDS, 2 mM 2-mercaptoethanol) and initially subjected to a voltage of 90V until the protein sample had fully entered the separating gel, at which point the voltage was increased to 400V and remained as such until the dye front from the samples had reached the bottom of the gel. Gels were then stained for 1 hour using quickstain and subsequently the stain was removed by two consecutive 2 hour incubations with 20 ml RO water.

2.4.7 Molecular mass measurement by Liquid chromatography-mass spectrometry.

Zeba spin desalting Columns were used for all sample preparation. Column was washed five times using 500 μ l of 50 mM ammonium acetate each time, and the columns spun at 1000 g for one minute each time. After all five washes of 50 mM ammonium acetate, 75 μ l of protein solution was added to the column, the column was spun as above, and the eluate was submitted to the University of Leeds Mass Spectrometry Facility. The method used is as follows:

Protein desalting and mass analysis was performed by LC-MS using an M-class ACQUITY UPLC (Waters UK, Manchester, UK) interfaced to a Synapt G2S Q-IMT-TOF mass spectrometer (Waters UK, Manchester, UK). 1 μ l of 5 μ M sample was loaded onto a MassPREP protein desalting column (Waters

UK, Manchester, UK) washed with 10 % solvent B in A for 5 min at 25 $\mu\text{L min}^{-1}$. After valve switching, the bound protein was eluted by a gradient of 2-40 % solvent B in A over 1 min at 25 $\mu\text{L min}^{-1}$. The column was subsequently washed with 95 % solvent B in A for 6 min before re-equilibration at 5 % solvent B in A ready for the next injection. Solvent A was 0.1 % formic acid in water, solvent B was 0.1 % formic acid in acetonitrile. The column eluant was directed in to the mass spectrometer via a Z-spray electrospray source. The MS was operated in positive TOF mode using a capillary voltage of 3.0 kV, sample cone of 40 V and source offset of 80 V. Backing pressure was 7.9 mbar and trap bias 4.0 V. The source temperature was 80 °C and desolvation was 100 °C. Argon was used as the buffer gas at a pressure of 9.1×10^{-3} mbar in the trap and transfer regions of the TriWave device. Mass calibration was performed by a separate injection of [Glu]-fibrinopeptide b at a concentration of 250 fmol μL^{-1} in MS/MS mode and a CID voltage (trap region) of 32 V. Data processing was performed using the MassLynx v4.1 suite of software supplied with the mass spectrometer.

2.4.8 Protein Dialysis

Proteins were dialysed into the appropriate buffer for the use of the protein. Dialysis tubing was used with a molecular weight cut off of 10 kDa. The volume of dialysis buffer used was at least 1 000 \times the volume of the protein sample to be dialysed. Dialysis was performed over a minimum of 2 hours (room temperature) or 4 hours (4 °C) with constant stirring; the dialysis buffer was then replaced with fresh buffer and at least another 2 or 4 hours (as above) of dialysis carried out.

2.4.9 Protein concentration

Amicon Ultra-15 concentrators with a 10 kDa cut off were used to concentrate protein samples. The membrane was first washed with distilled water and centrifuged at 3 000 g for 10 minutes. The concentrator membrane was then equilibrated with the same buffer the protein was kept in. The protein sample was then applied to the concentrator and spun at 3 000 g at 4°C until the desired concentration was achieved.

2.4.10 Size exclusion chromatography

At 4 °C, a HiLoad 26/600 Sephadex 200 column was equilibrated with degassed 50 mM sodium phosphate buffer at pH 7.4. 5 ml of the same buffer was applied to the injection loop 3 times while the AKTA was in "load" configuration. Less than 5 ml of protein at higher than 5 mg/ml was

applied to the injection loop, and the injection rate was set to 0.8 ml/min. The AKTA was changed to “inj” to apply the contents of the injection loop to the column, and the rate of buffer flow was changed to 2.0 ml/min. Fractions were loaded onto a fraction collector, and 2 ml fractions were collected starting from approximately 40 min after injection. Aggregated protein eluted after approximately 50 min and correctly folded protein eluted after 70 min. After approximately 90 minutes had elapsed, the fractions were collected and each peak was pooled separately. The column was first washed with 200 ml degassed RO water and then with 400 ml 20% ethanol solution.

2.4.11 Freeze drying

Protein samples were first dialysed (section 2.4.8) into 20 mM ammonium acetate pH 7.0. The protein concentration of the sample was determined and then appropriate aliquots were made in pre-weighed containers. The samples were then flash frozen in liquid nitrogen before being lyophilised using a Thermo Electron Corporation Heto PowerDry PL300 freeze dryer. After freeze drying, the weight of the containers with freeze dried protein was measured for an exact mass of protein in each container.

2.4.12 Modification used during screening

The protein modification method was modified from (Windle et al., 2017). Lyophilised protein was resuspended in 50 mM sodium phosphate at pH 8 with 6 M urea, 2 mg of protein in 1 ml of buffer. 2,5-dibromohexanediamide was made up at a concentration of 15.2 mg in 115 µl of DMF, and 100 µl of this solution was added to each millilitre of resuspended protein. The mixture was then vortexed briefly and incubated for 1.5 h at 37°C, 200 rpm in an orbital shaker. 40 µl of 0.1 mg/µl of thiol in 1.5 M Tris.HCl pH 8.8 was then added for each ml of resuspended protein, the mixture vortexed briefly again, and incubated for 2 h at 37°C, 200 rpm. The sample was inserted into dialysis tubing and left to dialyse in 5 litres of 50 mM sodium phosphate at pH 8 with 6 M urea overnight (section 2.2.5). The samples were transferred to another 5 litres of solution and left to dialyse for 2 h. The samples were then put through two rounds of dialysis for 2 h each in 50 mM Tris HCl buffer pH 7.5.

2.4.13 Modification as optimised for Y252Lni

Lyophilised protein was resuspended in 50 mM sodium phosphate at pH 8 with 6 M urea, 2 mg of protein in 1 ml of buffer. 2,5-dibromohexanediamide was made up at a concentration of 0.06 mg/µl in

DMF, and 50 µl of this solution was added to each millilitre of resuspended protein. The mixture was then vortexed briefly and incubated for 90 min at 37°C, 200 rpm in an orbital shaker and then transferred to 50°C without shaking for 60 min. The protein was then buffer exchanged into fresh 50 mM sodium phosphate at pH 8 with 6 M urea using PD10 columns (GE Healthcare, Illinois, USA) according to manufacturer's instructions. 60 µl of 0.1 mg/µl of thiol in 1.5 M Tris.HCl pH 8.8 was then added for each millilitre of buffer exchanged protein, the mixture vortexed briefly, and incubated for 2 h at 37°C, 200 rpm. The sample was inserted into dialysis tubing and left to dialyse in 5 litres of 50 mM sodium phosphate at pH 8 with 6 M urea overnight (section 2.2.5). The samples were then put through two rounds of dialysis for 2 h each in 50 mM sodium phosphate at pH 7.4.

2.4.14 Thiobarbituric acid assay - initial

Samples were set up as given in the text, but in all cases, a 200 µl aqueous sample potentially containing a TBA active molecule was taken and applied to a 96-well deep well plate (containing 12 µl 12% (w/v) trichloroacetic acid if the assay is a kinetic assay, as opposed to an endpoint assay). Once all samples had been taken, 11 µl of 0.2 M sodium periodate, 9 M phosphoric acid solution was added to each well with sample. The plate was shaken, spun down for 5 minutes at 2 000g and incubation at room temperature for 20 minutes. After this incubation 45 µl of 10% (w/v) sodium arsenite, 0.5 M sodium sulphate, 50 mM sulphuric acid solution was added, and the plate was gently agitated until a brown colour appeared and also disappeared in every well being assayed. Subsequently, 135 µl of 0.6% (w/v) thiobarbituric acid (TBA), 0.5 M sodium sulphate solution was added to each well. The plate was shaken, spun down for 5 minutes at 2 000g and incubated at 65 °C for 30 minutes. This was spun down at 3 000g for 5 minutes, and 85 µl of each well was transferred to a 96-well flat, clear-bottom reading plate.

2.4.15 Thiobarbituric acid assay - enhanced signal

Samples were set up as given in the text, but in all cases, a 100 µl aqueous sample potentially containing a TBA active molecule was taken and applied to a 96-well deep well plate (containing 12 µl 12% (w/v) trichloroacetic acid if the assay is a kinetic assay, as opposed to an endpoint assay). Once all samples had been taken, 22 µl of 0.2 M sodium periodate, 9 M phosphoric acid solution was added to each well with sample. The plate was shaken, spun down for 5 minutes at 2 000g and incubation at room

temperature for 20 minutes. After this incubation 90 µl of 10% (w/v) sodium arsenite, 0.5 M sodium sulphate, 50 mM sulphuric acid solution was added, and the plate was gently agitated until a brown colour appeared and also disappeared in every well being assayed. Subsequently, 270 µl of 0.6% (w/v) thiobarbituric acid (TBA), 0.5 M sodium sulphate solution was added to each well. The plate was shaken, spun down for 5 minutes at 2 000g and incubated at 65 °C for 30 minutes. This was spun down at 3 000g for 5 minutes, and 85 µl of each well was transferred to a 96-well flat, clear-bottom reading plate.

2.4.16 Protein crystallisation

Size exclusion chromatography was performed (see section 2.4.10) on a sample of Y252Lni modified protein and fractions containing the peak corresponding to correctly folded monomer were collected and pooled. This was concentrated (section 2.4.9) to 9 mg/ml for crystallisation. Crystals were grown in a total of 24 conditions, each containing 200 mM sodium chloride, 100 mM Tris.HCl, either 18, 20, 22, 24, 26 or 28% PEG 3350, and at either pH 7.0, 7.5, 8.0 or 8.5. The crystals were grown on a 24-well plate using the hanging drop method, each well contained one mother liquor solution and 3 protein crystallisation drops hanging from the coverslip consisting of either 1 µl mother liquor and 2 µl protein solution, 2 µl mother liquor and 2 µl protein solution, or 2 µl mother liquor and 1 µl protein solution, for a total volume of 3 or 4 µl in each drop. Three replicate plates were produced, and crystals had formed on all plates after 14 days.

2.4.17 X-Ray crystallography

Crystals were prepared for exposure to X-ray radiation by soaking in a solution of mother liquor containing PEG 400 as a cryoprotectant. Each crystal was soaked in a solution containing 15% PEG 400 and 100 mM sodium pyruvate, followed by a solution containing 20% PEG 400 and 100 mM sodium pyruvate, and finally a solution containing 25% PEG 400 and 100 mM sodium pyruvate. After this final soak, the crystal was flash frozen in liquid nitrogen and stored at -196 °C. Crystals were shot on the i24 beamline at the Diamond Light Source in Didcot, Oxfordshire, UK. The wavelength of the beamline was 0.9686 Å, the beam had a size of 2 500 µm² and the transmission was 75% of maximum. 1900 images were taken, each 0.1° rotation around a straight line distance drawn through the crystal, taking into account the individual features of the crystal, giving a

total rotation of 190° around the crystal. Data analysis was performed as described in the text (section 4.2.1)

2.4.18 Circular Dichroism

Proteins were prepared at 0.2 mg/ml in a 50 mM sodium phosphate solution at pH 7.4 from solutions that had undergone size exclusion chromatography (section 2.4.10). CD spectra were obtained using a 1 mm quartz cell on a Chirascan CD spectrophotometer over a range of 180-260 nm at 1 nm intervals.

2.5 Computational methods

2.5.1 Protein chain alignment

Chain alignment was performed using Molecular Operating Environment (MOE) from Chemical Computing Group (Montreal, Canada). Both the wild-type structure (PDB ID: 4ah7) and the structure generated in this work were loaded into MOE. All chains were considered individually, and not grouped with any others. Chains were aligned based on all residues' sequence and structural factors. For the superposition, all residues' α atoms were then considered, using the previous alignment.

2.5.2 Structure preparation

Structures were prepared using the Protonate 3D function in MOE, at a temperature of 300 K, pH 7 and an ion concentration of 100 mM. Protonate 3D treats the backbones of pAAs and npAAs differently, it gives the nitrogen atoms in npAAs a charge of +1 while the charge of backbone nitrogens of pAAs have a charge of 0, and the backbone carbonyl adjacent to the npAA nitrogen is converted into a hydroxyl group. As such, the backbones of K165 and Y252Lni residues were corrected to be identical to the pAA backbones. Additionally, the Schiff base between K165 and pyruvate in the *ecNAL* wild-type structure and the *saNAL* Y252D-Lni structure were corrected from a single bond to a double bond between the nitrogen and the pyruvate.

2.5.3 GRID surface construction

GRID surfaces were calculated and visualised using MOE (Goodford, 1985). Initially, a decanoic acid was built in MOE to give a molecule that had regions that were both polar and apolar. A GRID surface was then calculated using aliphatic hydroxyl group as a probe. The interaction energy of the probe was modified, the resulting surfaces examined and an interaction energy of -3.6 kcal/mol was chosen as this produced small green surfaces around the carboxylate group while leaving the rest of the molecule uncovered. To ensure that pictures could be taken from an identical perspective for all structures, structures were aligned as follows. One wild-type structure and two Y252Lni structures, chains A-D, were loaded. For the wild-type structure chain B was considered and all others were ignored, for one of the Y252Lni structures, chain B, the Y252L-Lni structure, was also considered while the others were ignored. For the other Y252Lni structure, chain D was considered while the rest were ignored, giving us the Y252D-Lni structure. The sequences were aligned

according to sequence and structural factors. Sequences were then superposed using this alignment, and, due to the combination of considered/ignored chains, each active site that has been selected as representative of each structure was aligned. This allowed a view to be chosen that displayed all active sites in the exact same orientation. As such, the interaction potential surface of aliphatic hydroxyl groups for each enzyme was individually calculated and displayed.

2.5.4 Dynamics Generalised protocol

All molecular dynamics were performed using MOE. In general, molecular dynamics simulations were run as follows. A molecular structure was loaded into MOE, and a number of different regions were defined. Active regions were considered during calculations, while inactive regions had no effect on the simulation, and mobile regions were allowed to move, while immobile regions were held in place. Unless otherwise stated, an active, mobile region approximately 9 Å radius sphere from the substrate or active site centre; an active, immobile region from approximately 9 Å to 13.5 Å radius from the substrate or active site centre; and an inactive, immobile region consisting of the rest of the structure, were defined. Each structure was energy minimised before starting the simulation to reduce the forces on each atom and make sure they are in a natural position. The simulation usually had two phases, an equilibration phase and an unrestricted dynamics phase. In the equilibration phase, the heavy atoms have light restraints on them to prevent significant random deviation from their starting positions. During the unrestricted phase these restraints are removed and the atoms are allowed to move freely. The AMBER 10 forcefield (Case et al., 2005, Case et al., 2008, Pearlman et al., 1995) was used to represent the protein structures while the Extended Hückel Theory forcefield (Hoffmann, 1963) was used for small molecule structures. The equation of motion used was the Nosé-Poincaré-Andersen formulation (Sturgeon and Laird, 1999, Bond et al., 2002). Structures were recorded every 0.5 ps for review after the simulation, other than for video production, where structures were recorded every 0.1 ps for increased frames per second and smoother video.

2.5.5 ManNAc molecular dynamics

As a control, the first simulations were run using the ecNAL structure with ManNAc bound as the substrate. These were taken from PDB ID: 4BWL (Daniels et al., 2014). This served as a test of the forcefields being used and how the input structures were interpreted by these forcefields.

Dynamics were performed as given in the generalised protocol in section 2.5.4 but the sphere of motion was only 4.5 Å radius for this initial test, and the corresponding sphere of immobile activity was between 4.5 and 9 Å. The equilibration phase lasted 100 ps and the unrestricted dynamics phase lasted a subsequent 500 ps.

A second test was using saNAL, PDB ID: 4ah7 (Timms et al., 2013), with ManNAc as the substrate. ManNAc was floated into the active site in a similar position to that observed in the ecNAL structure. Dynamics were performed as given in the generalised protocol in section 2.5.4 but the sphere of motion was only 4.5 Å for this simulation and the corresponding sphere of immobile activity was between 4.5 and 9 Å. The equilibration phase lasted 100 ps and the unrestricted dynamics phase lasted a subsequent 500 ps.

2.5.6 Glucuronolactone molecular dynamics

In all the enzymes, wild-type, L-Lni and D-Lni, the glucuronolactone substrate was placed within hydrogen bonding distance of the terminus of the amino acid at the 252 position, with the terminal aldehyde group pointing towards the pyruvate bound to the active site lysine. It was then rotated into two different positions, one with the ring hydroxyl groups pointed towards the residue at 252 and one with the ring hydroxyls pointing away. A tether was made to restrict the aldehyde carbon to within 3.5 and 4.5 Å of the pyruvate methyl, both to ensure that the substrate stayed in the active site and to bias the simulations to show active site interactions. Dynamics were performed as given in the generalised protocol. The equilibration phase lasted 100 ps and the unrestricted dynamics phase lasted a subsequent 500 ps. To enable smoother video, the starting positions of these simulations were used to generate a second set of simulations, and dynamics were performed as given in the generalised protocol in section 2.5.4. The only phase was a 100 ps unrestricted dynamics phase and structures were saved every 0.1 ps.

A second series of simulations were run in all enzymes where the glucuronolactone substrate was placed within hydrogen bonding distance of the amino acid at the 252 position with the terminal aldehyde group pointing towards the pyruvate bound to the active site lysine. It was then rotated into 6 different positions, each separated by approximately 60 degrees. Dynamics were performed as given in the generalised protocol in

section 2.5.4. The equilibration phase lasted 100 ps and the unrestricted dynamics phase lasted a subsequent 500 ps.

Once all of the simulations had been analysed, the movements were classified into 3 broad classes. A second set of simulations were run for all Y252Lni containing structures, moving the substrate further from the non-proteinogenic side chain. These were rotated, energy minimised, simulated and analysed in the same way as the previous set. An extra class was found in these simulations, increasing the overall number to 4 classes.

2.5.7 Binding pose steered dynamics

To assess the strengths of each binding pose, a representative structure of each class was chosen (figure 4.21). A methane molecule was placed outside of the active site so that there were no atoms between the substrate and the methane molecule. A distance restraint was placed between the carbon atom on the methane and the closest heavy atom on the substrate that was not involved in any hydrogen bonds. The restraint had a lower bound of 2 Å and an upper bound of 4 Å. The methane atom was then moved down the vector of the restraint to a distance of 16 Å. The weighting of the restraint was varied from 0 to 0.3 in increments of 0.02. The restraint energy, which forms the pulling force on the tether, is determined by equation 2.1.

Equation 2.1:

$$E_d = w[p(L - r) + p(r - U)]$$

Where E_d is the distance restraint energy, w is the weight value, L is the lower bound of the restraint, r is the distance between the atoms involved, U is the upper bound of the restraint and $p(x)$ is a function defined in equation 2.2.

Equation 2.2:

$$p(t) = \begin{cases} 0 & \text{if } t < 0 \\ t^3(6 - 8t + 3t^2) & \text{if } t \in [0,1] \\ t & \text{if } t > 1 \end{cases}$$

Where t is the result of the equation inside the brackets in equation X.1.

This gives a force profile where the force increases with distance the further away from the restraint boundaries the atoms lie. All restraints were set to the same initial distance and identical upper and lower bounds to ensure that the actual forces are as similar as possible. As the specific force applied at any one point of the simulation varies by distance, and the distance changes all the time as the atoms oscillate during the simulation,

it would not be feasible to measure the actual force required to disrupt binding.

Dynamics were performed as given in the generalised protocol in section 2.5.4. Each simulation was run for 200 ps of unrestricted dynamics, and the time at which the substrate was out of position was recorded. The substrate was defined as out of position when the only hydrogen bonds remaining were those between the ring hydroxyls and the carboxylate group or when the substrate had moved to the extent that it was a different class.

As a proxy measure of binding force, the highest weight applied that did not disrupt the binding pose was assessed.

2.5.8 Full length product molecular dynamics

Full length product was built in MOE. R and S stereoisomers at the aldol carbon were made, and the pentameric ring was rotated into two positions for each of the stereoisomers, giving a total of four full length products analysed. Pyruvate was removed from the structures used for simulations previously, and each product was floated in so that the atoms that previously formed the pyruvate were in a position as if the Schiff base had just broken. Dynamics were performed as given in the generalised protocol in section 2.5.4. The equilibration phase lasted 100 ps and the unrestricted dynamics phase lasted a subsequent 500 ps.

2.5.9 Full length product steered dynamics

Full length product pulling was performed on the structures set up in section 2.5.8 (full length) using the same method as given in section 2.5.7 (binding pose strength). The range of weights used ranged from 0.1-0.7.

2.5.10 Dendrogram generation

Five key distances were chosen to describe the binding poses found in the rotation simulations: from the aldehyde carbon on the substrate to the methyl carbon of the pyruvate; from the aldehyde carbon on the substrate to the oxygen of the hydroxyl on Y137; from the oxygen of the C5 hydroxyl to the carbon of the carboxylate of D141; from the oxygen of the C5 hydroxyl to the carbon of the carboxylate of E192; and from the oxygen of the C5 hydroxyl to either the carbon bonded to the hydroxyl in Y252, or the carbon of the carboxylate group in lanthionine. The distance measurements were made using database tools in MOE. The distance measurements were extracted into Excel and when the raw data was

analysed using a 10 point running average, a number of conformation shifts were observed. To assess these conformations individually, a heat map was applied to the most diagnostic distance or distances for each run individually to easily identify when the distance shifted significantly. Runs were classed as stable when the ten-point running average data was approximately stable in all distances for greater than 30 ps. Then, all the raw data points over the stable run were extracted into individual spreadsheets, which were then averaged and exported to a tab delimited .txt file.

In R (version 3.1.1 with the fastcluster package installed), the .txt file was converted into a plot-able matrix using the code:

```
matrix_name <- as.matrix(read.table("file.txt",  
header=TRUE, sep = "\t",  
row.names = 1,  
as.is=TRUE))
```

Where `matrix_name` is a variable name for the matrixed data and `file.txt` is the text file exported. The matrix was then plotted into a dendrogram using the code:

```
matrix_clust = hclust(dist(matrix_name, method =  
"euclidean"), method = "complete", members = NULL)  
plot(matrix_clust, hang = -1)
```

Where `matrix_clust` is the clustered `matrix_name` data.

During extraction, the number of picoseconds each pose was stable for was recorded, which allowed further data analysis.

2.6 Chemical synthesis of 2,5-dibromohexanediamide

The 2,5-dibromohexanediamide synthesis was conducted similarly to the procedure given in (Chalker et al., 2012) with reduced masses. Adipic acid (5.00 g) was added to a 100 ml round bottom flask and suspended in thionylchloride (15 ml). The flask was equipped with a condenser and heated to reflux (bath temperature 80 °C) open to air. Reaction was heated for 30 minutes to allow adipic acid to dissolve and then for a further 60 minutes. The reaction was then cooled to room temperature and CCl₄ (20 ml) was added to the reaction followed by *N*-bromosuccinimide (14.62 g). The reaction was stirred vigorously and 2 drops of HBr were added by pipette. The reaction was heated to reflux open to air, and the colour of the reaction changed from red to black. After the reaction mixture had turned completely black the reaction was cooled to RT and then to 0 °C, the mixture was stirred at 0 °C to ensure all succinimide had precipitated, and this precipitate was then removed by filtration. Et₂O (10 ml) was used to rinse and complete the filtration. The filtrate was concentrated *in vacuo* to give a dark red liquid. In a 100 ml round bottom flask, 40 ml of NH₄OH (25% aqueous) was cooled to 0 °C. The crude acid chloride was added dropwise over 20 minutes to the ammonia solution with rapid stirring. After the addition was complete, the reaction was stirred vigorously at 0 °C for 1 h. The 2,5-dibromohexanediamide precipitated from the reaction mixture as a dark solid. This was isolated by filtration and partially dried. The product was purified by trituration: the dark solid was suspended in 40 ml of 1:1 MeOH:H₂O and heated to 60 °C. The mixture was stirred vigorously at 60 °C until the liquid in the vessel had changed colour from clear to orange, and for at least 30 minutes. The mixture was then cooled to RT and the white solid was isolated by filtration and washed with MeOH. The product was dried under high vacuum, with a yield of 25%. ¹H NMR (500 MHz, D₂O): δ = 1.76-2.25 (4H, m, CH₂CH₂), 4.35 (2H, m, 2 × CHBr), 7.31 (2H, s), 7.70 (2H, s) (2 × NH₂).

Chapter 3 - Discovery of a chemically modified aldolase with a novel activity

3.1 Introduction

Developments in the past few years have allowed protein engineers to build enzyme catalysts *de novo*, predicting which amino acids are required in which position on a protein scaffold for the desired function, generally using one of the rosetta software packages (Siegel et al., 2010, Khersonsky et al., 2011, Jiang et al., 2008). However, these experiments can still give unpredictable results, such as complete abandonment of the designed active site in favour of another region on the enzyme (Giger et al., 2013). Non-canonical amino acids (ncAAs) are much less well studied than the 20 canonical amino acids, and so *in silico* predictions involving ncAAs would give even less reliable results than the *in silico* predictions using canonical amino acids. With the chemical modification method of introducing ncAAs (described in section 1.3.3) a large number of different side chains can be inserted into a range of different positions across the active site of an enzyme, provided the enzyme to be studied contains no cysteine residues (either naturally or by mutagenesis). This allows a brute force solution, trying as many different combinations of side chain, position and substrate as possible to find useful or interesting activities. This type of combinatorial search is not possible using genetically encoded amino acids, or would at least be a lot more labour intensive, as each different ncAA would have to be expressed in a different genotype of *E. coli*, and each position that the ncAA was required in would need separate protein expression experiments.

Since a large number of enzyme-aldehyde combinations were to be tested, a reliable assay was needed, not only to be used as a screen to sample the widest chemical space of both side chains and substrates, but also to be used to measure kinetic parameters.

3.2 Assay development

It was proposed to use a long established assay that uses thiobarbituric acid (TBA) reacting with 1,3-dicarbonyls to form a chromophore, called the TBA assay. This assay was originally designed to detect neuraminic acids (Warren, 1959), but has been previously re-purposed to assay the concentration of a wider range of aldol reaction products in a high throughput format (Windle et al., 2017).

The assay relies on the conversion of the aldehyde carbon to a hydroxyl group upon catalysis (figure 3.1A). Given the right substrate selection, this forms a 1,2 vicinal diol; or a 1-hydroxy-2-aminoalcohol derivative. Any of these structures can be attacked by a periodate anion which can cleave the carbon-carbon bond between the newly created hydroxyl carbon and the carbon attached to the hydroxyl/amine/amide. Since pyruvate is the other substrate, a 1,3-dicarbonyl is formed (figure 3.1B) which can then react with two TBA molecules to give a pink coloured molecule, and this can be detected by a spectrophotometer via absorbance at a wavelength of 550 nm.

The tolerance of the assay towards different users, one of the factors determining the robustness of an assay (Vander Heyden et al., 2001), had not been assessed however, and as such the assay was examined for operator robustness, variability and signal-to-noise ratio

3.2.1 Assay signal enhancement

Initial tests with the assay resulted in a signal-to-noise ratio that was quite low, and was mostly because the signal was a lot weaker than had been shown previously in the lab. A low signal-to-noise ratio would hinder hit identification in the final screen and increase the number of false negatives. The low signal proved to be repeatable and a number of hypotheses were tested in an effort to increase the signal.

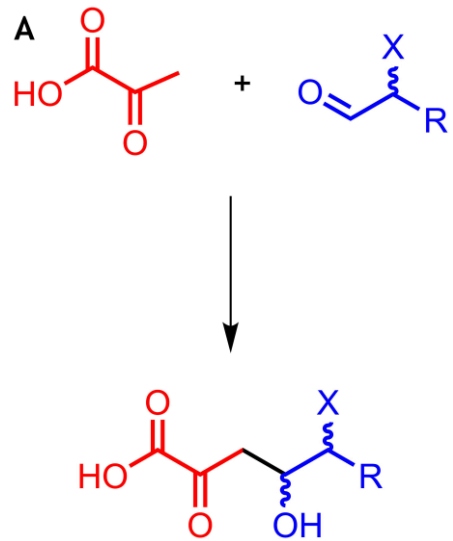
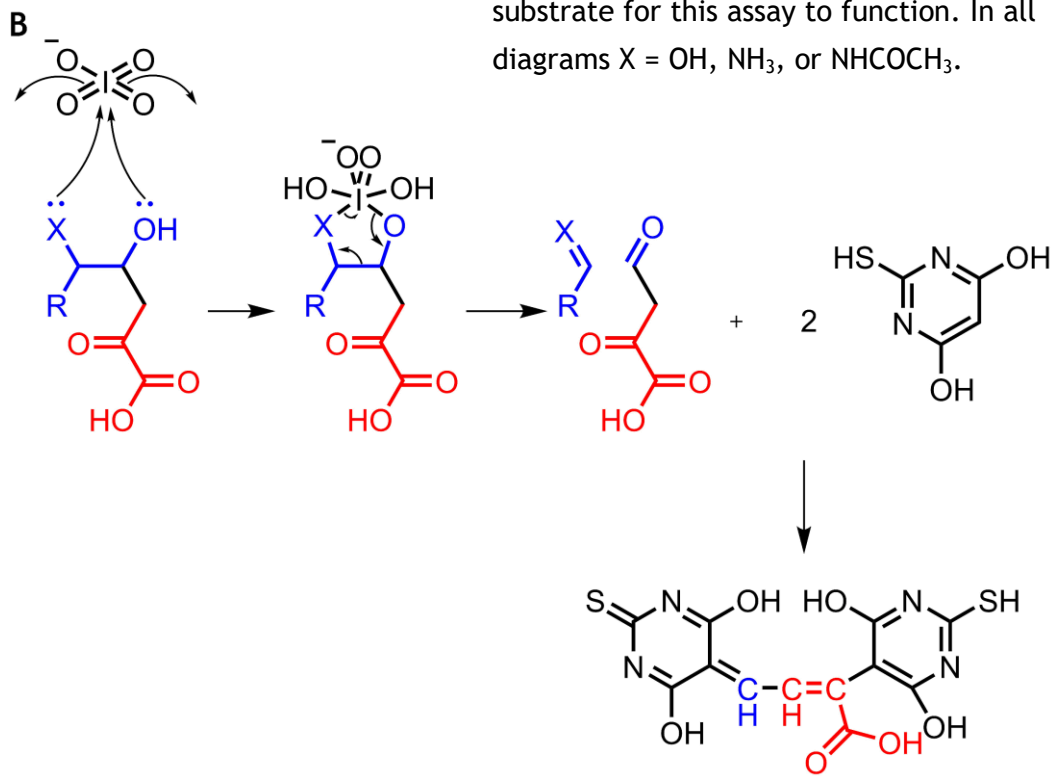


Figure 3.1 Reactions required for visible product using the TBA assay **A** Generalised reaction scheme for aldol reactions that *N*-acetylneuraminic acid lyase can perform that maybe assayed using the TBA assay. **B** The reaction by which the thiobarbituric acid assay forms a chromophore when aldol product is present, highlighting the requirement for a hydroxyl/amine/amide group adjacent to the aldehyde group in the substrate for this assay to function. In all diagrams X = OH, NH₃, or NHCOCH₃.



The positive control used was a pure, commercially-sourced sample of the product of the wild-type reaction between pyruvate and *N*-acetyl-D-mannosamine (ManNAc): *N*-acetylneuraminic acid (Neu5Ac). A standard curve was analysed, decreasing from 250 nmole Neu5Ac. The amount of variability observed in the points meant that the points below 63 nmole all displayed ranges of absorbance that overlapped with other points below 63 nmole. Only the points with the two highest concentrations showed absorbance values that were outside the range of absorbance values displayed by any other points. This means that any activity that resulted in less than 125 nmoles of product would be less accurately measured (figure 3.2).

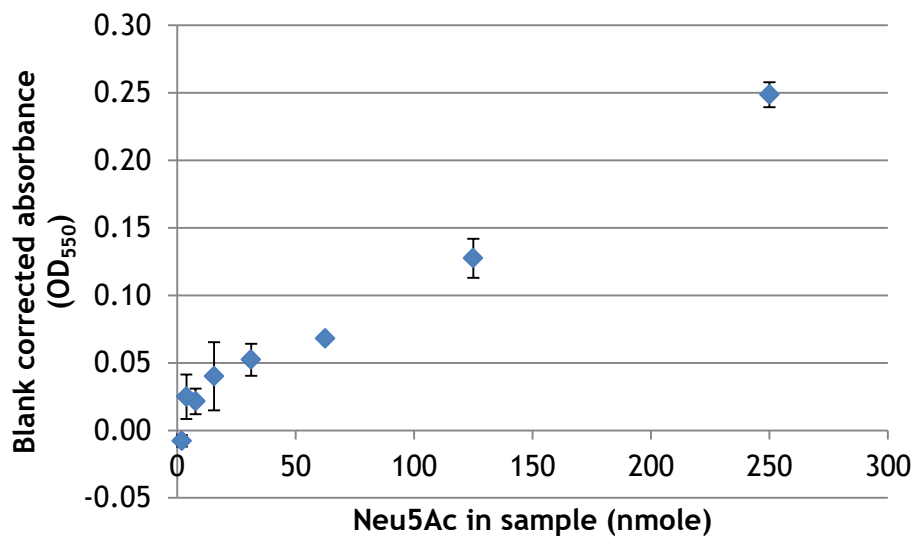


Figure 3.2 Standard curve of commercially purchased *N*-acetylneuraminic acid when analysed by the TBA assay. Samples were run in triplicate and blanked against the average absorbance of 8 blank wells. Error bars show standard error of the blank corrected samples.

One possible explanation for the relatively low signal could be incomplete cleavage of the product upon addition of periodate. To address this, a reaction between ManNAc (5 mM) and pyruvate (100 mM) catalysed by wild-type *N*-Acetylneuraminic acid lyase was set up and incubated overnight. Samples were taken in duplicate and periodate added to all. The periodate reaction was quenched with arsenite at the appropriate time, and after all samples had been quenched, the rest of the assay was conducted according to section 2.4.14. The resulting absorbance values are shown in figure 3.3.

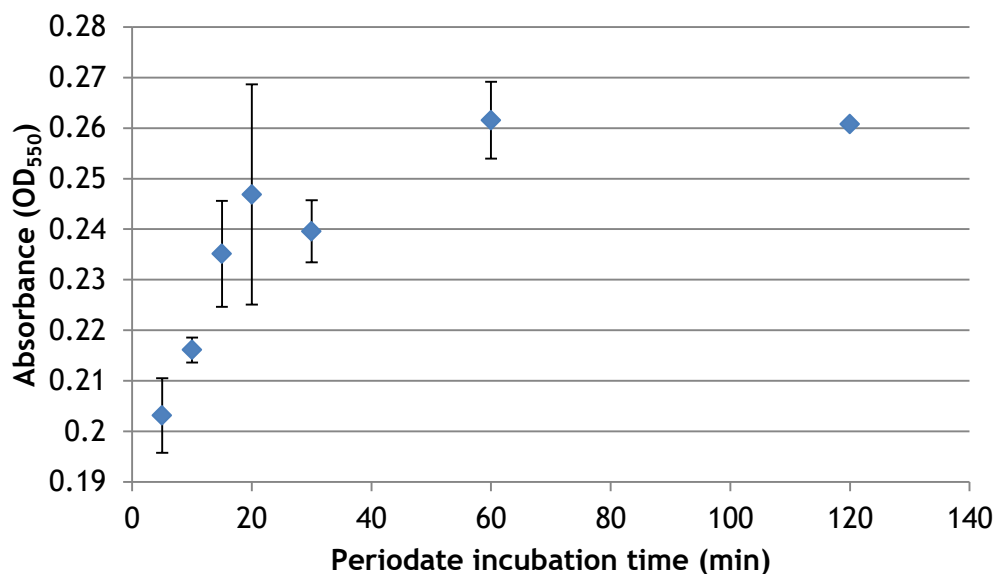


Figure 3.3 Graph showing the effect of periodate incubation time on the absorbance produced in the TBA assay from a reaction with pyruvate, ManNAc and wild-type enzyme after overnight incubation. Samples were run in duplicate and were not blanked. Error bars shown are the standard error.

The reaction between ManNAc and pyruvate gives a much lower absorbance than pure Neu5Ac. The absorbance increased with the time of periodate incubation until 60 minutes, but further incubation had no effect. Additionally, the data shows that after 20 minutes incubation, which is the incubation time used in the protocol to this point, the reaction produced 75% of the absorbance of the 60 minute incubation. Subsequently, it was decided that increasing the incubation time from 20 min would not give an increase in absorbance worth the extra incubation time.

Another alternative for increasing the assay signal was increasing the amount of periodate and TBA added to the reaction. This should increase the signal by moving the chemical reaction equilibrium further towards the chromophore product, or increase the rate at which the colour develops. Two reactions were set up using pyruvate, wild-type enzyme and either ManNAc or mannose, and left to incubate overnight. In this case mannose was included as the screen to be conducted subsequent to assay development would include aldehydes with both hydroxyl groups and amine/amide groups at the 2 position. Samples were taken of these reactions and either treated as normal (see section 2.4.14); had double the volume of periodate and arsenite added; had double the volume of TBA

added; or had double the volume of periodate, arsenite and TBA added. The results can be seen in figure 3.4.

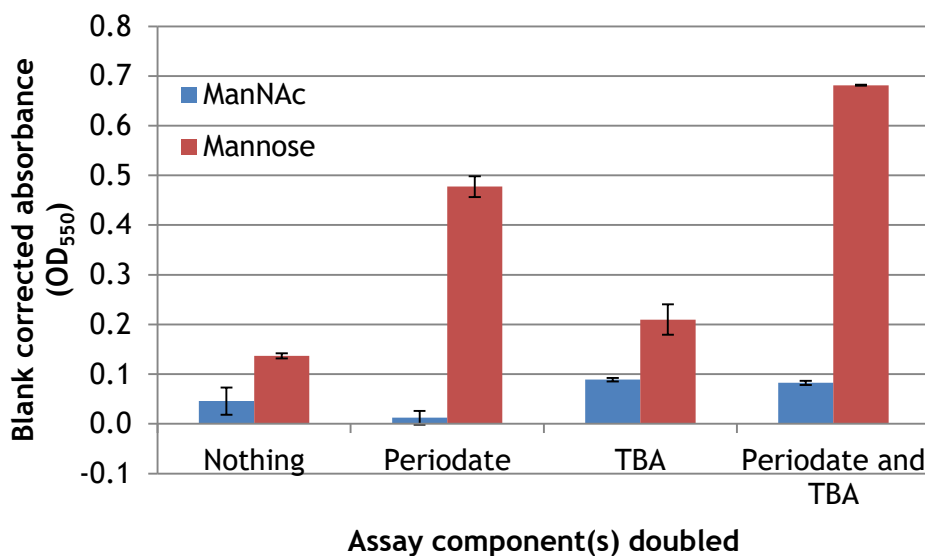


Figure 3.4 The difference in signal from overnight reactions incubated with wild-type enzyme, pyruvate and either ManNAc or mannose, between different assay protocols in which one or both reagents responsible for signal production have been doubled in amount. Arsenite acts to neutralise the unreacted periodate, so while the amount of arsenite added was increased, it was only increased in the presence of increased pyruvate, and so is not mentioned on the graph for brevity. Absorbance produced by ManNAc reaction shown by blue bars, absorbance produced by mannose reaction shown by red bars. Error bars are standard error.

Firstly, we can see that the mannose gives much higher absorbance readings in all assay conditions, producing an absorbance reading up to 0.6 absorbance units greater than that of ManNAc when both assay components are doubled. Additionally, the two reactions, ManNAc and mannose, reacted differently to the different conditions. The absorbance produced by the reaction between mannose and pyruvate increases 3.5-fold when periodate is increased, 1.5-fold when TBA is increased and increases 5-fold when both components were increased. The absorbance produced by the reaction between ManNAc and pyruvate increases by 1.9-fold when TBA alone is doubled, and 1.8-fold when both periodate and TBA are increased but decreases to only 0.3-fold the original absorbance when only additional periodate is added. The best conditions were therefore considered to be an increase of both periodate and arsenite, and TBA, as this should increase the absorbance of substrates hydroxylated at position 2 and substrates with an amino/amide group at position 2.

When the new assay parameters (section 2.4.15) were used on another standard curve of Neu5Ac (figure 3.5), the concentration of Neu5Ac required to produce a signal distinguishable from background was lower, at 31 nmoles of Neu5Ac compared to 125 nmoles using the previous method. This was due to a reduction in variability in the samples with lower concentration of Neu5Ac, along with an increase in signal at the higher concentrations.

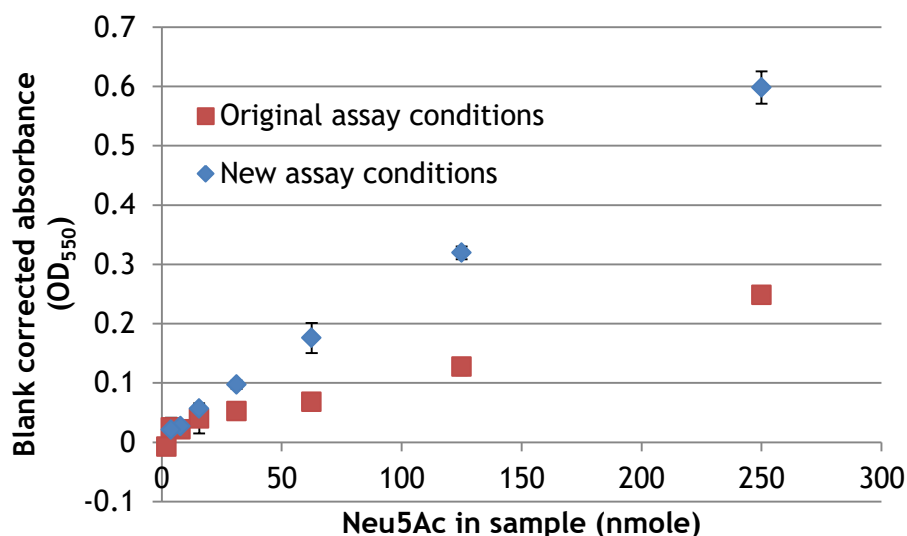


Figure 3.5 Standard curve of commercially purchased *N*-acetylneuraminic acid when analysed by the TBA assay using both the original and newly optimised assay conditions. Red shows the original assay conditions, as seen in figure 3.2 and blue shows the newly optimised assay conditions. Samples were blanked against the average absorbance of at least 4 blank wells. Error bars show standard error of the blank corrected samples.

3.2.2 Assay variability

It is important to minimise or at least characterise the variability of an assay so that the assay can be used in the most optimal way, i.e. if an assay has very little variability, then points may be run singly, if the variability is higher, then duplicates or triplicates may be required.

Initially, to exclude any systematic variation due to the plate reader, an empty plate was read at 550 nm five times to ascertain both the read-to-read variability, how much the absorbance readings change when the plate is read, and cross plate variability, how much the readings alter across the plate.

Firstly, the average absorbance of each well over the five reads was taken, and then the average value of each column and row of the plate was examined (figure 3.6). There appears to be no systematic variability in absorbance across the columns or down the rows, each column and row is approximately as variable as another, and the variability is small compared to the expected absorbance differences.

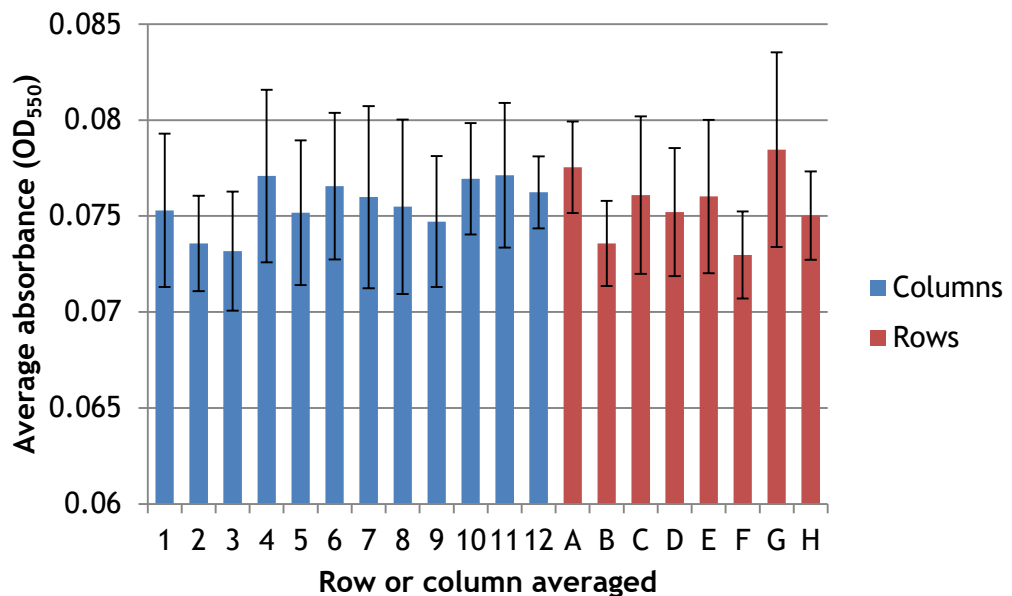


Figure 3.6 The average absorbance of each row or column of an empty plate over 5 reads. Columns are shown in blue and rows in red. Error bars are the standard deviations of the averages for each well in that row or column.

The standard deviation of each well over the five reads was also taken and a summary graph can be seen in figure 3.7. This shows that while the average absorbance across the plate does not vary systematically, the variance in the points does.

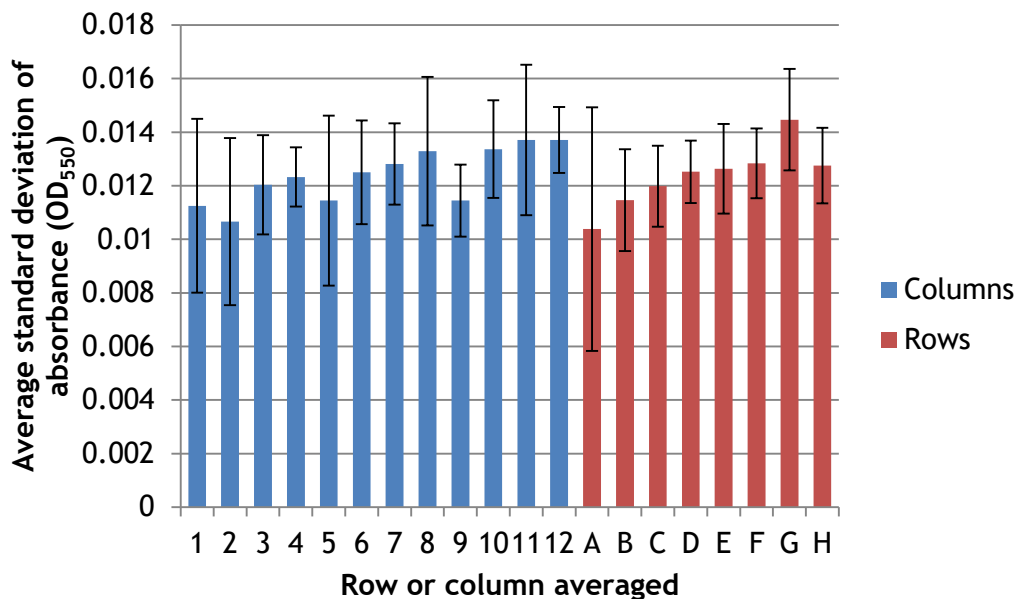


Figure 3.7 Average standard deviation of each well in each row or column of an empty plate over 5 reads. Columns are shown in blue and rows in red. Error bars are the standard deviations of the average for each row or column.

As the wells get further from A1, i.e. as wells further to the right and bottom of the plate are considered, the variability of the wells increases. As the average values of the wells showed no systematic variation, this should not make much of a difference to the final results if it is taken into account. As such, following these results, all plates were read 3 times and the resulting absorbance values were averaged across all three reads to reduce the influence of this variability.

Plate reader induced variability is not the only source, there is a significant source of variability in the assay procedure itself. A selection of aldehydes were mixed with pyruvate, equivalent to negative controls, and were assayed as usual to gauge the variability of the assay under normal use (figure 3.8). All the aldehydes produced an absorbance below 0.2 and the variability was small. This suggests that any activity should be easily distinguishable from the negative controls run on the same plate, and there does not necessarily need to be duplicate negative control wells per aldehyde.

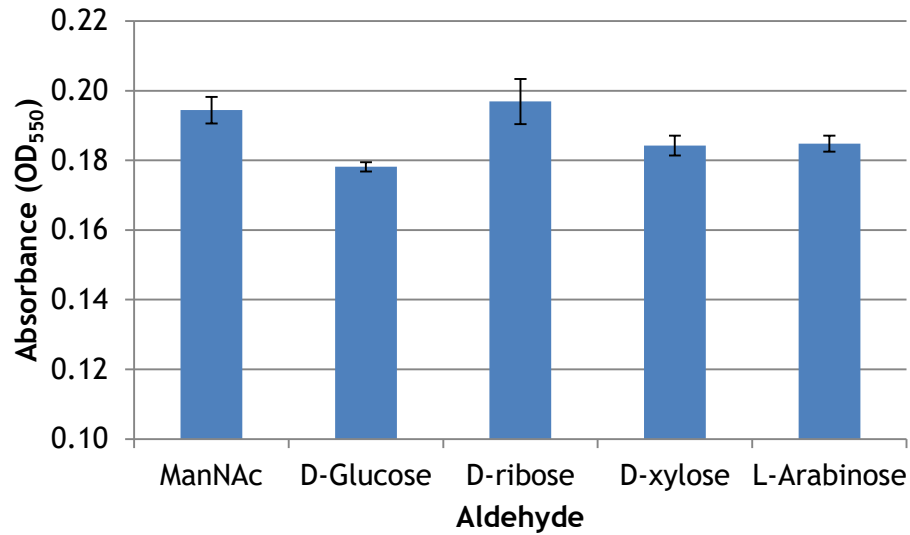


Figure 3.8 Chart showing the absorbance produced by reactions between pyruvate and different aldehydes without enzyme when analysed by the TBA assay. Error bars shown are standard error.

As can be seen in figure 3.4 and 3.5, the variability of known positive reactions with the new assay conditions are also quite low, and so this assay should be suitable for screening.

3.3 Screening set-up and results

3.3.1 Choice of mutation positions

A number of positions around the active site of NAL have previously been identified as affecting the substrate specificity of the enzyme (Williams et al., 2005, Devenish and Gerrard, 2009, Krüger et al., 2001, Campeotto et al., 2010). These have mostly focused on the *E. coli* protein, but the *S. aureus* protein is highly structurally homologous to the *E. coli* (RMSD = 1.43Å), especially around the active site, so these positions were expected to be important in the *S. aureus* enzyme as well. Some of these positions have previously been studied in the lab using the method of modification used here, yet in those modifications they had not shown any significant increases in activity (Windle et al., 2017). The positions targeted in the original study were selected by taking the highly homologous *E. coli* structure with *N*-acetylneuraminic acid bound (Daniels et al., 2014), and selecting all residues within 5 Å of the substrate. These residues were then recorded and their homologues in the *S. aureus* structure were selected for mutagenesis to cysteine residues. As previous work has already been done in this area, the residues used in this report were those that had not been extensively studied in that work. The residues selected for modification were I139, D191, S208, I243, I251 and Y252 (figure 3.9).

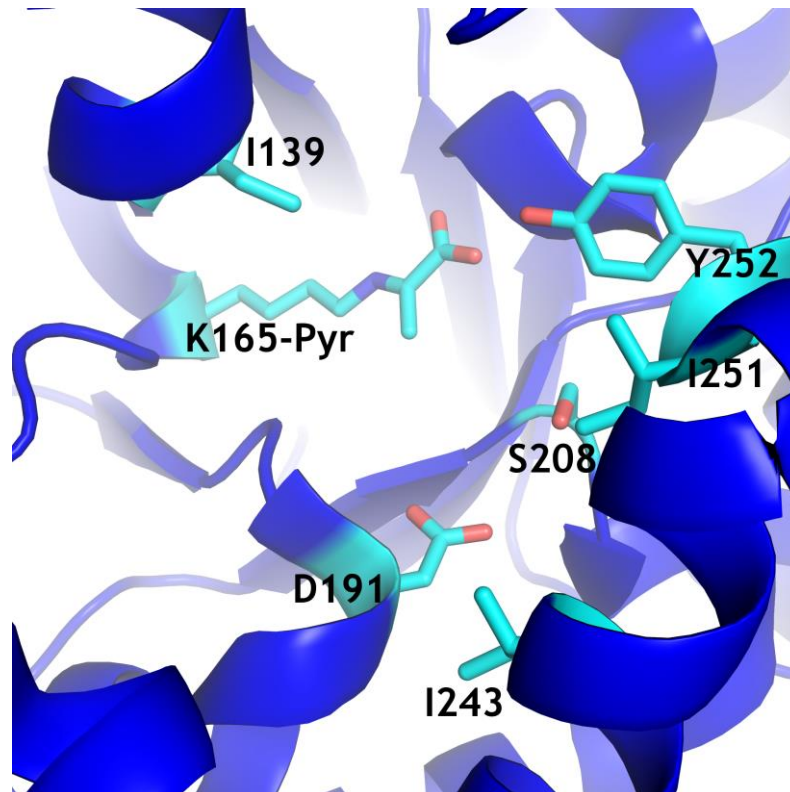


Figure 3.9 View of the active site of wild-type *S. aureus* N-acetylneuraminic acid lyase showing the positions to be targeted in the screen. This view looks down into the active site from the centre of the tetrameric ring of monomers that comprise active NAL, where substrate would enter to bind catalytically. Side chains of targeted residues and the catalytic lysine residue forming a Schiff-base with pyruvate are shown in cyan and the backbone of the protein is shown in blue.

These residues were arranged around the active site, and at different distances from the pyruvate bound to the catalytic lysine, residue 165. Placing the same modifications around the active site in this manner should allow any interesting modifications to be identified no matter what position they need to be in to display enhanced activity.

The cysteine mutants for all these have previously been made in the lab (Windle, 2015) and were re-cloned into E. coli genotypes BL21 (DE3) for expression and XL10 Gold for storage of the DNA. The sequencing for each of these showed the expected sequence (figure 3.10).

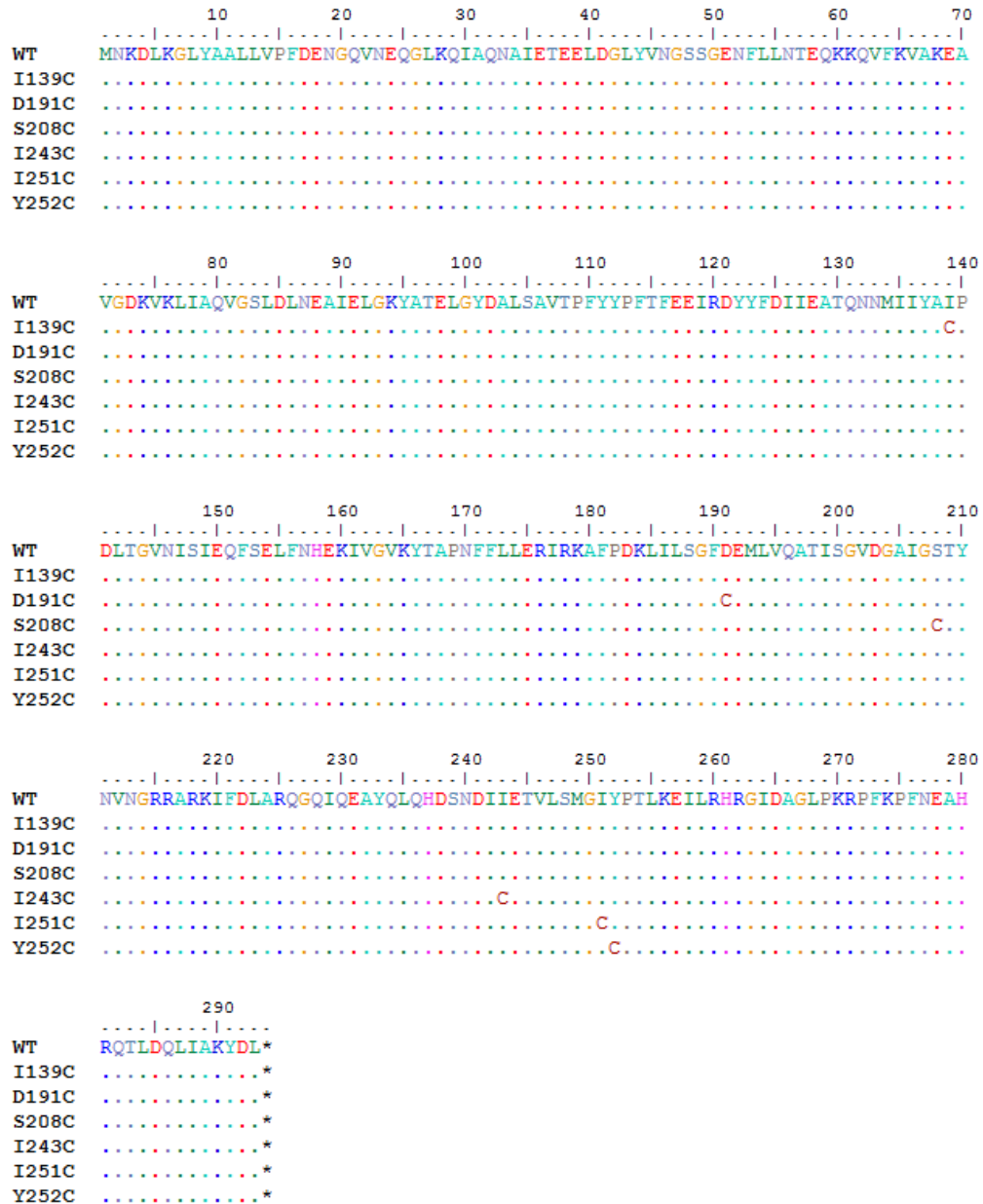


Figure 3.10 Translated DNA sequencing of the residues to be targeted in the screen showing successful mutation in all cases of the desired residue to cysteine, the first step required for performing chemical modification.

3.3.2 Choice of ncAAs to be incorporated

While theoretically any thiol molecule could be incorporated using the protein modification method described, there were a number of restrictions and factors to be considered that limited the thiols to those

that were actually useful and interesting. Firstly, previous experiments have shown that modifications using a thiol with a conjugated ring system directly adjacent to the thiol group, e.g. as in thiophenol, fail to produce the expected nCAA. This may be due to the conjugated ring system tending to draw the thiolate electrons into the delocalised ring, making the Michael addition unsuccessful under the conditions that were used here. As such, thiols of this type had to be discounted from the pool of screening side chains. Additionally, if the thiol is very large and considering the location of the active site in NAL, i.e. in the centre of an α/β barrel, the resulting nCAA may cause the protein to misfold which would most likely result in an inactive enzyme. A diverse set of side chains would also give the screening assay the greatest chance of finding any potential activity, so a range of functional groups was desired. Finally, cost had to be considered, as a not insignificant amount of thiol was required for each screen, and six screens were to be performed in total.

After considering these factors in combination, a set of thiols were selected that would produce the amino acids shown in figure 3.11.

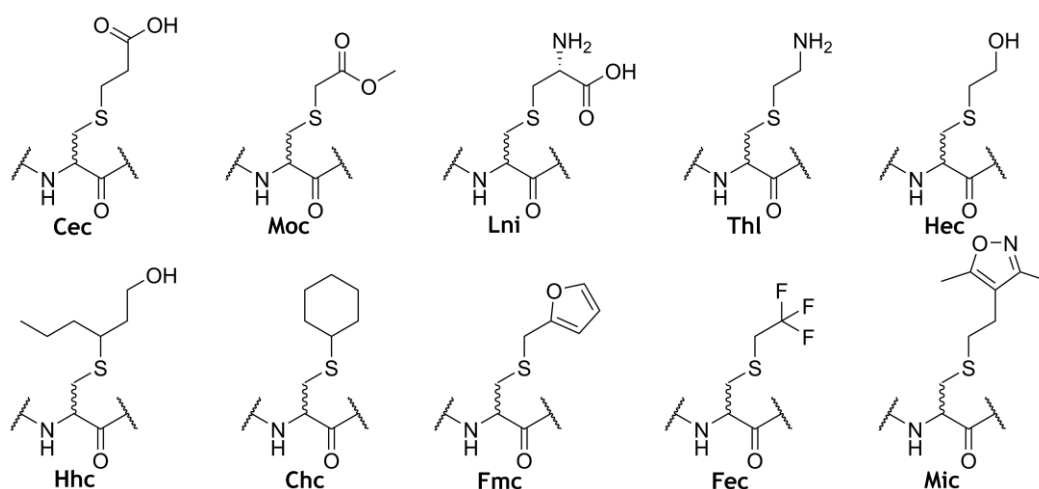


Figure 3.11 Structural diagrams of the nCAs produced when modified using the thiols selected for the screen. Top row, from left: (carboxyethyl)cysteine (Cec), (2-methoxy-2-oxoethyl)cysteine (Moc), lanthionine (Lni), thialysine (Thl), and (2-hydroxyethyl)cysteine (Hec). Bottom row, from left: (1-hydroxy-3-hexanyl) cysteine (Hhc), (cyclohexanyl)cysteine (Chc), (2-Fuylmethyl)cysteine (Fmc), (2,2,2-trifluoroethyl)cysteine (Fec), and (2-(3,5-Dimethyl-4-isoxazolyl)ethyl)cysteine (Mic).

The lowest molecular mass that a modified amino acid produced using this method in these experiments could have is high in comparison to all the canonical amino acids. If a position in the active site is chosen that has a

large volume for the side chain to inhabit, utilising chemical modification at this position allows for a range of different ncAAs to be inserted, potentially with a diverse set of functional groups in the side chains. However, if the position targeted has a limited amount of space for the side chain, the large size of chemically modified amino acid side chains may restrict the amount of modifications that could be performed at this position while still producing protein capable of folding correctly.

The thiols chosen had a diversity of potential charge, from hydrophobic cyclohexanethiol to zwitterionic cysteine and varied in size from producing amino acids ranging from 164 Da in the case of thia-lysine to 244 Da in the case of 2-(3,5-dimethylisoxazol-4-yl) ethyl cysteine. This enabled a range of activities to potentially be detected, with a limited number of amino acids produced.

3.3.3 Choice of substrates

Substrate selection involved a balancing act between two opposing factors; increased activity for substrates similar to the wild-type substrate should be easier to discover, but the more different a substrate is from the wild-type substrate the more interesting the hit is. Therefore a set of potential substrates were chosen that were both very similar to the natural substrate, e.g. *N*-acetyl-*D*-glucosamine, and quite different from the natural substrate e.g. *L*-arabinose and γ -lactone-*D*-gucuronic acid. The full set can be seen in figure 3.12. These substrates have both 5-carbon and 6-carbon members, both *L*- and *D*-sugars, and one cyclic sugar acid. Amino- and amino-acylated sugars are also included.

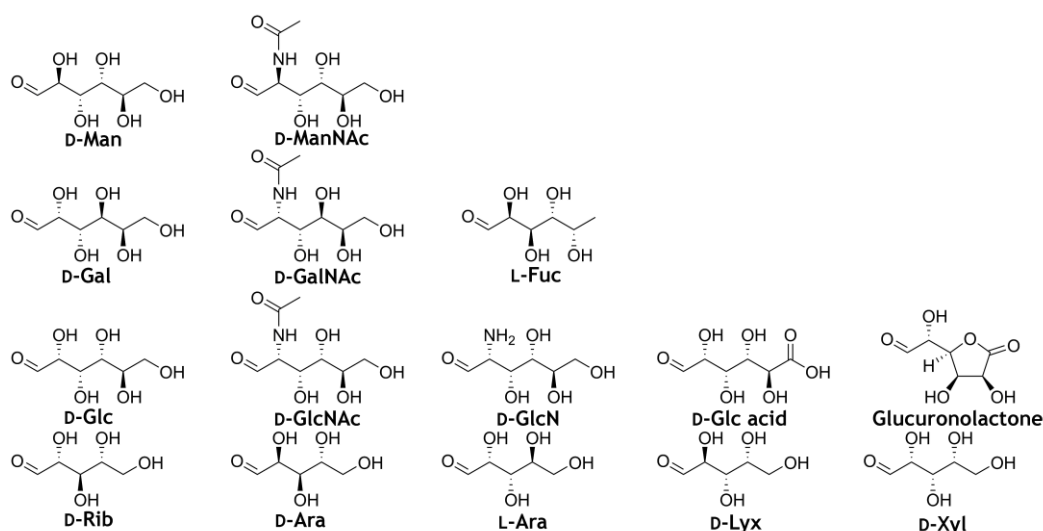


Figure 3.12 Aldehyde substrates used in the screen. Aldehyde displayed in open chain conformation because NAL accepts substrates in this conformation. Top row, from left: D-mannose (D-Man), *N*-acetyl-D-mannosamine (D-ManNAc). Second row, from left: D-galactose (D-Gal), *N*-acetyl-D-galactosamine (D-GalNAc), L-fucose (L-Fuc). Third row, from left: D-glucose (D-Glc), *N*-acetyl-D-glucosamine (D-GlcNAc), D-glucosamine (D-GlcN), D-glucuronic acid (D-Glc acid), glucuronolactone. Bottom row, from left: D-ribose (D-Rib), D-arabinose (D-Ara), L-arabinose (L-Ara), D-lyxose (D-Lyx), D-xylose (D-Xyl).

3.3.4 Screening results

The screens were laid out as shown in figure 3.13. This set up required four plates per position mutated, because each plate could contain half the modified enzymes and half the aldehydes, giving a total of 24 plates for all the positions. Reactions involving ncAAs were set up in adjacent duplicates while control reactions, i.e. reactions which used wild-type enzyme or had no enzyme at all, were set up in singlet on at least two plates each. Every plate contained a row of D-mannose to act as a positive control when in combination with wild-type. As such the control reactions and especially the D-mannose reaction, enabled the variability across plates to be compared.

	1	2	3	4	5	6	7	8	9	10	11	12
D-mannose A	Enzyme 1	Enzyme 2	Enzyme 3	Enzyme 4	Enzyme 5	Wild-type enzyme	No enzyme					
Aldehyde 1 B												
Aldehyde 2 C												
Aldehyde 3 D												
Aldehyde 4 E												
Aldehyde 5 F												
Aldehyde 6 G												
Aldehyde 7 H												

Figure 3.13 Example plate layout diagram of the arrangement for the activity screen. This represents only one quarter of each position screened as only half the enzymes for a position and half the aldehydes are included. Modified enzymes were placed in duplicate in two adjacent columns while aldehydes were placed in every well of a row. Wild-type enzyme and a no enzyme control were placed in single columns at the end of the plate.

Each position was assayed separately, and the enzyme concentration used in each experiment was determined by the lowest protein concentration at the particular position that was being considered, e.g. all enzymes with modifications at position 139 were diluted to be the same concentration as the modified enzyme with the lowest concentration of the enzymes modified at 139. This ensured that the raw absorbance values could be compared across all four plates and minimised the space on the plate taken up by controls as only one wild-type control was needed per aldehyde; opposed to one wild-type control per enzyme concentration per aldehyde. This in turn increased throughput and reduced reagent usage and toxic waste production. While the concentration of the enzymes varied by position modified, the other assay components were as followed; 5 mM aldehyde substrate, 100 mM sodium pyruvate, 50 mM Tris.HCl pH7.4.

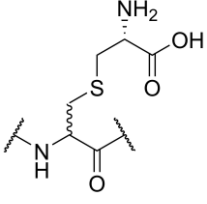

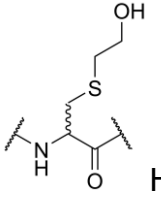

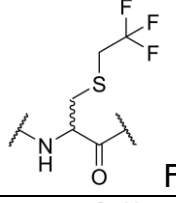

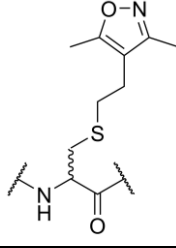


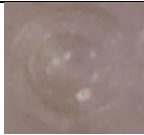


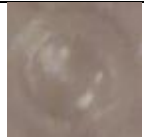

In total, 900 different combinations of enzymes and aldehydes were tested for activity. Of those, 443 reactions displayed some activity that was higher than the no enzyme control meaning that 49% of the enzymes made had some activity. Additionally, 65 enzyme/aldehyde combinations displayed an absorbance higher than the wild-type for that aldehyde, or 7% of the reactions. However, 30 out of the 60 enzymes created, 50%, were more active than the wild-type for at least one aldehyde.

Both the ratio of the activity of the modified enzyme over the activity of the wild-type enzyme and the absolute difference in absorbance between the activity of the modified enzyme and the activity of the wild-type

enzyme were considered. These methods both isolated the same top ten enzyme/aldehyde combinations, and in roughly the same order. These 10 hits are shown in figure 3.14. Most of the top ten involved glucosamine, either *N*-acetylated or not, and (2-hydroxyethyl) cysteine (Hec) is the most numerous side chain represented. However, when the plate photos are taken into account, the only hit in the top ten that produced any significant pink colour was the hit at position 252 with the ncAA Lni and the substrate glucuronolactone.

These apparent top ten hits that did not have any observable pink colour also had high absorbance values at a wavelength of 600 nm, a wavelength commonly used to determine the amount of particulate matter in a sample that would contribute to light scattering. This scattering by larger particles would appear as absorbance at any visible wavelength tested, and is probably the cause of the high absorbance values shown here. What the cause of the large particles however is more difficult to say. It has been noted previously in the lab however that *N*-acetyl-D-glucosamine sometimes gave an absorbance in the TBA assay that did not correspond to the visual appearance of the wells (unpublished data). Therefore it was decided to focus on Y252Lni and the reaction between pyruvate and glucuronolactone for the remainder.

(Next page) Figure 3.14 Summary table of the top ten hits found in the screen. Structures of the ncAAs and their three letter codes are shown. Lni, Lanthionine; Hec, (2-hydroxyethyl)cysteine; Fec, (2,2,2-trifluoroethyl)cysteine; Mic, (2-(3,5-Dimethyl-4-isoxazolyl)ethyl)cysteine. The modified positions are listed in the order with which they appear along the protein chain. The modifications and substrate hits are listed primarily according to the hit positions and secondarily by their positions on the plate, from top left to bottom right. The Δ Absorbance values are the absorbance values from the hit with absorbance of the wild-type for that substrate in that experiment subtracted from it. The well photos are cut out from a photo of the whole plate, and the left well of each duplicate was chosen for display here.

Position	Modification	Aldehyde	Δ Absorbance (rank)	Plate photo
I139	 Lni	N-acetyl-D-glucosamine	0.041 (9)	
	 Hec		0.108 (2)	
	 Fec		0.042 (7)	
	 Mic		0.100 (3)	
D191	Hec	N-acetyl-D-glucosamine	0.038 (10)	
	Mic		0.109 (1)	
S208	Hec	N-acetyl-D-glucosamine	0.089 (4)	
I243	No hits present in the top ten			
I251	Hec	D-glucosamine	0.048 (6)	
		N-acetyl-D-glucosamine	0.042 (7)	
Y252	Lni	Glucuronolactone	0.050 (5)	

3.4 Hit confirmation

3.4.1 Re-modification and control reaction

The most promising hit discovered in the screen was Y252C modified using cysteine as the thiol, when catalysing the reaction between pyruvate and D-glucurono-3,6-lactone, or glucuronolactone to form the product (5S)-5-[(2R,3R,4S)-3,4-dihydroxy-5-oxoxolan-2-yl]-4,5-dihydroxy-2-oxopentanoic acid (DHOPA) (figure 3.15 C). This hit is interesting as NAL has not previously been reported to have activity with a cyclic substrate. ManNAC in solution forms an equilibrium between cyclic and linear conformations, and the cyclic conformation does make up the majority of the molecules. However, X-Ray crystallography data shows that the substrate only enters the enzyme in a linear conformation (Daniels et al., 2014). The side chain formed when cysteine attacks dehydroalanine in this way resembles a lanthionine residue (Horn et al., 1941, Paul and Van der Donk, 2005) that has been included in the protein chain (figure 3.15 A and B).

Firstly, the hit was confirmed across multiple experiments to ensure that the difference in activity was repeatable and consistent. The activity of the modified enzyme was fairly consistent at approximately 1.7 times the absorbance produced when blank corrected and compared to wild-type under the conditions used, i.e. 100 mM pyruvate, 5 mM aldehyde, 16 hour incubation.

Initially, the modification procedure as used for the screen did not cause all reactions to occur with 100% success. As the screen would only be used to identify potentially interesting modifications and their reactions and would not be used to derive any numerical values, incomplete modification was deemed acceptable. However, an incompletely modified enzyme is unsuitable for any kinetic assessments or for crystallisation studies, so the modification procedure had to be examined.

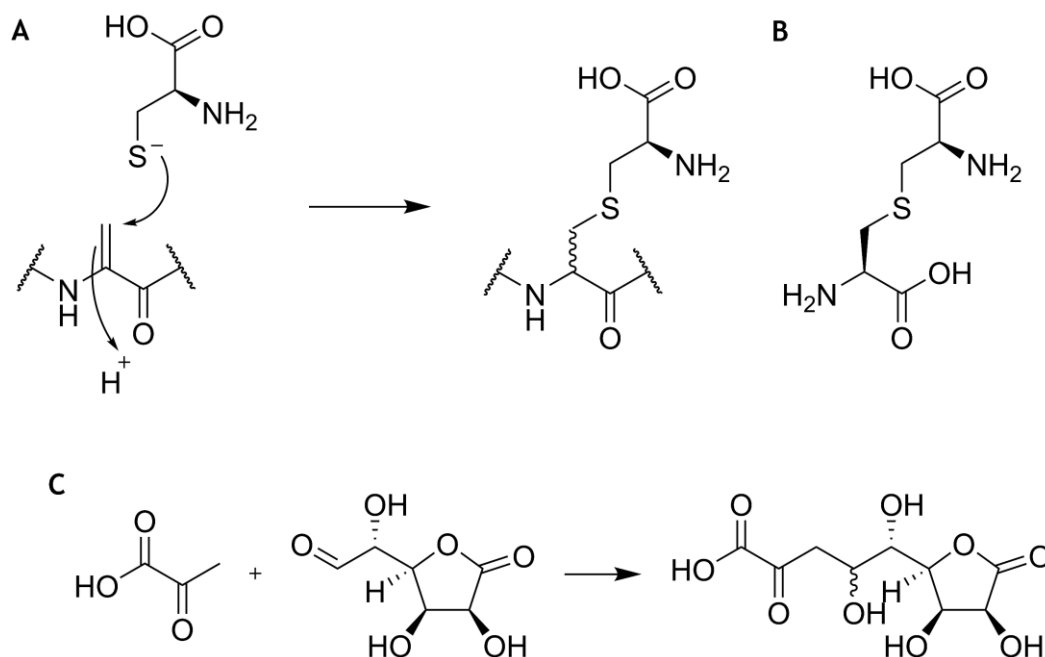


Figure 3.15 Chemical structural diagrams for the discovered hit. **A** Diagram of the mechanism of the modification reaction between Y252-Dehydroalanine (Dha) and cysteine to create Y252-Lanthionine (Lni). **B** Structure of the free amino acid L-lanthionine (Paul and Van der Donk, 2005) for comparison to the modified side chain. **C**, Reaction between pyruvate and glucuronolactone to give the anticipated product.

Mass spectrometric analysis showed that after modification of Y252C using cysteine as the thiol, only approximately 40% of the protein was the mass expected for the correctly modified protein (Figure 3.16). Other than this, approximately 30% of the protein was of a mass that corresponds to Y252Dha, i.e. the thiol modification has been unsuccessful; and 30% of the protein was 174 Da larger than the correctly modified protein should be. The 174 Da mass increase corresponds to the presence of a bis-alkylation elimination intermediate remaining attached to the protein. Both of these undesired species are created by a different failure during the modification procedure.

The modification procedure used during screening can be broken down into four distinct actions performed on the protein. Firstly, the protein is dissolved in pH 8.0 buffer to deprotonate the thiolate on the cysteine side chain. Second, the cysteine is converted into dehydroalanine using 2,5-dibromohexandiamide (DiBr). The dehydroalanine is then converted into the side chain of choice by addition of thiol. Finally the protein is subjected to dialysis into fresh reaction buffer to stop the modification and then dialysis into storage buffer.

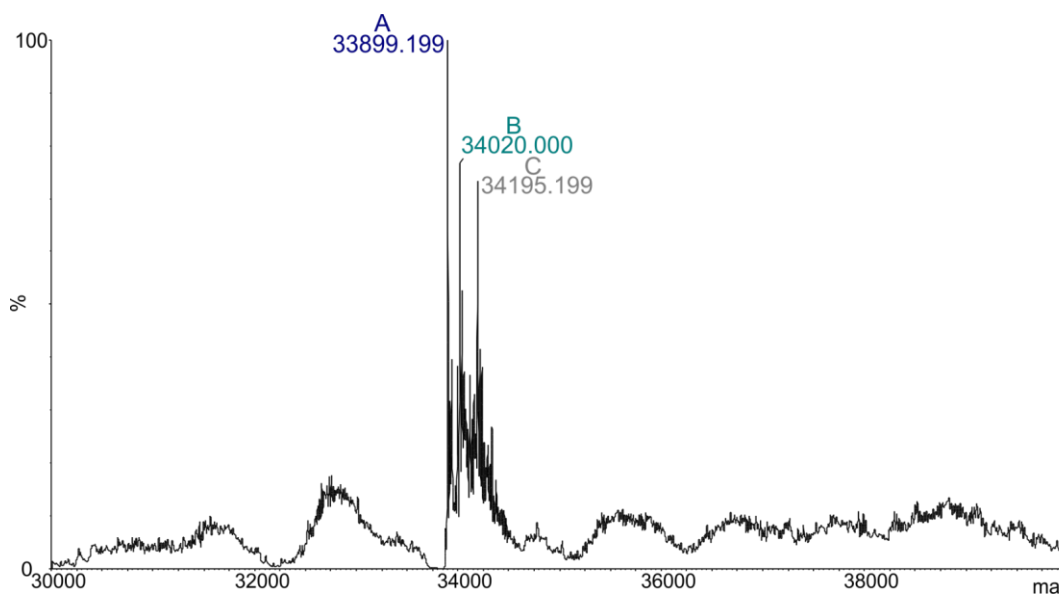


Figure 3.16 Mass spectrum analysis of a modification of Y252C using L-cysteine as the thiol to confirm Y252Lni production. Expected mass of Y252Lni: 34022.6 Da, expected mass of Y252Dha: 33901.4 Da, error of measurement ± 3.5 Da.

The 174 Da mass increase is hypothesised to occur in two potential ways. Firstly, the DiBr reagent may not eliminate from the cysteine, remaining on the protein as a sulfonium ion. This sulfonium ion can then be attacked by the thiol (figure 3.17, route A) when added, to give a protein of a mass 174 Da larger than the correctly modified protein. Alternatively, the introduced thiol could attack the released thiolanyl molecule in solution, forming a thiol consisting of the desired thiol and the remaining modification reagent (figure 3.17, route B). The first mechanism could be solved by heating the protein to a higher temperature, causing the sulfonium ion to be eliminated at a greater rate, forming the dehydroalanine residue. The second mechanism can be discouraged by moving the protein into buffer that does not contain the reagent, once the reaction has taken place. As such, a 50°C incubation step was added to cause release of the cyclised modification reagent, and a buffer exchange step was introduced to get rid of any spent or active modification reagent in solution. Additionally, the amount of DiBr applied was decreased to reduce the likelihood of it remaining in solution after the buffer exchange step.

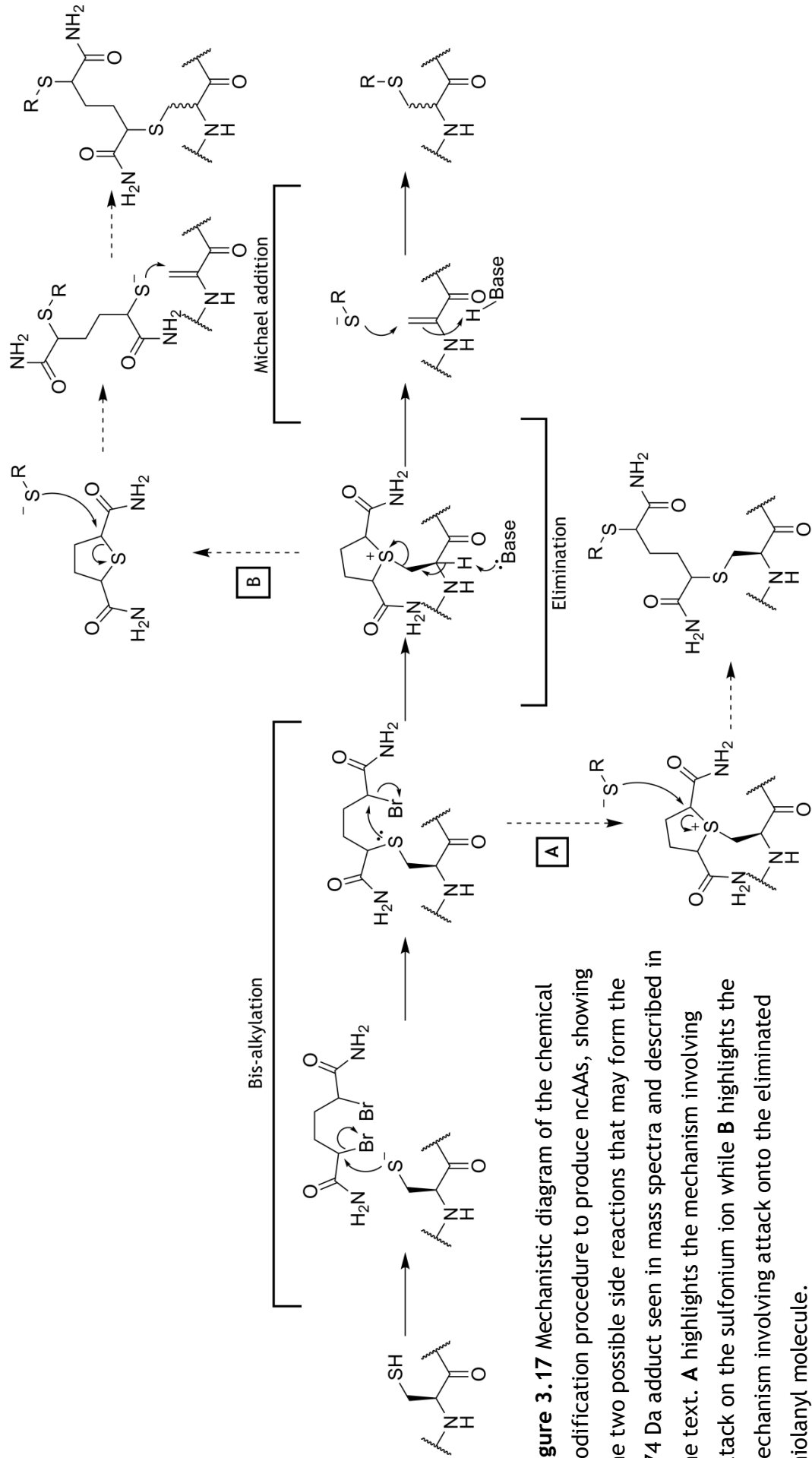


Figure 3.17 Mechanistic diagram of the chemical modification procedure to produce ncAAs, showing the two possible side reactions that may form the 174 Da adduct seen in mass spectra and described in the text. **A** highlights the mechanism involving attack on the sulfonium ion while **B** highlights the mechanism involving attack onto the eliminated thiolanyl molecule.

The other undesired species, remaining Y252Dha, is caused by incomplete modification of the protein by the thiol, and this was prevented by increasing the amount of thiol added to the protein solution. This procedure, together with the procedures mentioned above, caused the proportion of correctly modified protein to increase from 40 to 60% (figure 3.18). This was deemed sufficient for further experiments.

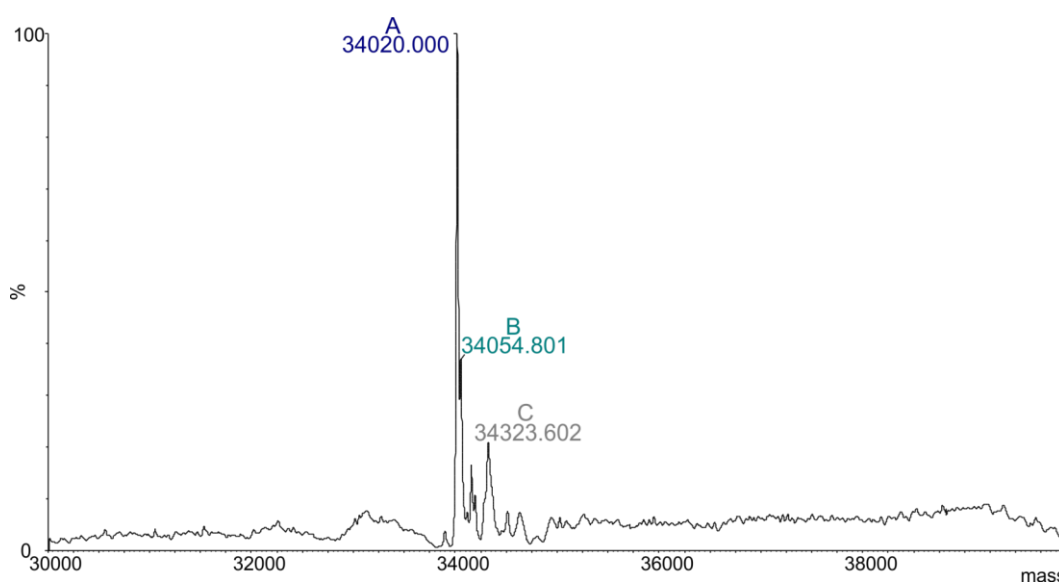


Figure 3.18 Mass spectrum analysis of Y252Lni produced using the optimised chemical modification procedure. Expected mass of Y252Lni: 34022.6 Da, error of measurement ± 3.5 Da.

To confirm that the activity observed was due to an aldolase reaction occurring, a number of controls needed to be run. While the screening reaction had a no enzyme control, this would not account for a non-specific reaction between the enzyme and either of the substrates. All possible controls were assembled on a plate using freshly modified enzyme and incubated for 24 hrs. Duplicate wells of all combinations of pyruvate, Y252C modified with cysteine, and aldehyde were created and then assayed using the enhanced TBA assay (section 2.4.15). The only pair of wells that gave significant signal were the wells in which the pyruvate, glucuronolactone and enzyme were all incubated together (figure 3.19). There was a slightly higher absorbance for control wells including glucuronolactone when the enzyme was not present when compared to control reaction with enzyme present. This suggests that a small background reaction took place, and this reaction is inhibited by presence

of the enzyme. The enzyme may therefore interact with the aldehyde even in the absence of pyruvate. The blank corrected absorbance of the control samples was no more than 8.5% that of the samples with all assay components and so this activity's contribution to the absorbance is likely not significant.




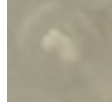



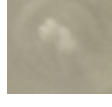
		Glucuronolactone				
		✓		✗		
Pyruvate	✓		1.660		0.181	✓
			0.331		0.222	✗
			0.231		0.183	✓
	✗		0.346		0.204	✗
						Y252Lni

Figure 3.19 Summary of the control reactions performed for the reaction between pyruvate and glucuronolactone with Y252Lni as a catalyst as analysed by the TBA assay. For each entry in the table the picture shown is of one of the duplicate wells for each condition, and the number is the average absorbance value after the TBA assay of the duplicate wells. Ticks and crosses indicate the presence or absence respectively of the specified reaction component. If present, the components had a concentration of: pyruvate, 100 mM; glucuronolactone, 10 mM; Y252Lni enzyme, 0.5 mg/ml. The buffer used in the assay and used as a replacement for any missing assay components was 50 mM sodium phosphate pH 7.4.

3.4.2 Removal of unfolded protein and initial kinetic assessment

When size exclusion chromatography is performed upon modified enzymes, we have previously noticed (Windle, 2015) that a significant proportion of the protein post-modification elutes at a volume that suggests that it is unfolded. As such, before any kinetic data was gathered, the protein was gel filtered.

The protein was gel filtered according to section 2.4.10 and the resulting chromatogram is shown in figure 3.20. As previously noticed, there is a large amount of unfolded protein, as well as correctly folded tetramer. There also appears to be a peak corresponding to folded monomer, which has not previously been seen (Figure 3.20, peak 3, purple). Approximately 2.0 mg (20%) came off in the first peak (red), 6.3 mg (64%) eluted in the second peak (green) and the third peak (purple) contained 1.5 mg (15%).

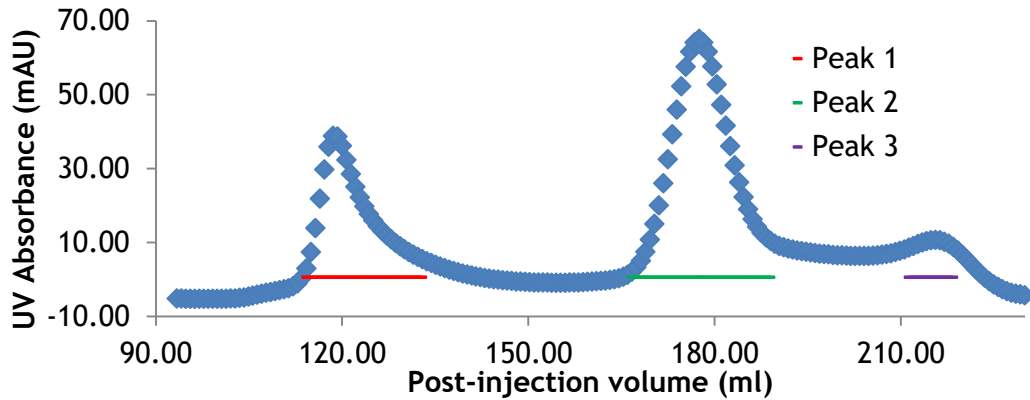


Figure 3.20 Size exclusion chromatography gel filtration elution profile of Y252Lni. The elution fractions were pooled over the volumes highlighted with lines underneath the absorbance values, showing peak 1 in red, peak 2 in green and peak 3 in purple.

To confirm the identities of each peak, mass spectrum analysis was performed on a sample from each peak, and the results are shown in figure 3.21. The three peaks display approximately equal proportions of each species shown previously, so there can be little correlation between protein species and the stability of folding, i.e. the correctly modified protein is not more or less likely to fold correctly than the improperly modified protein species.

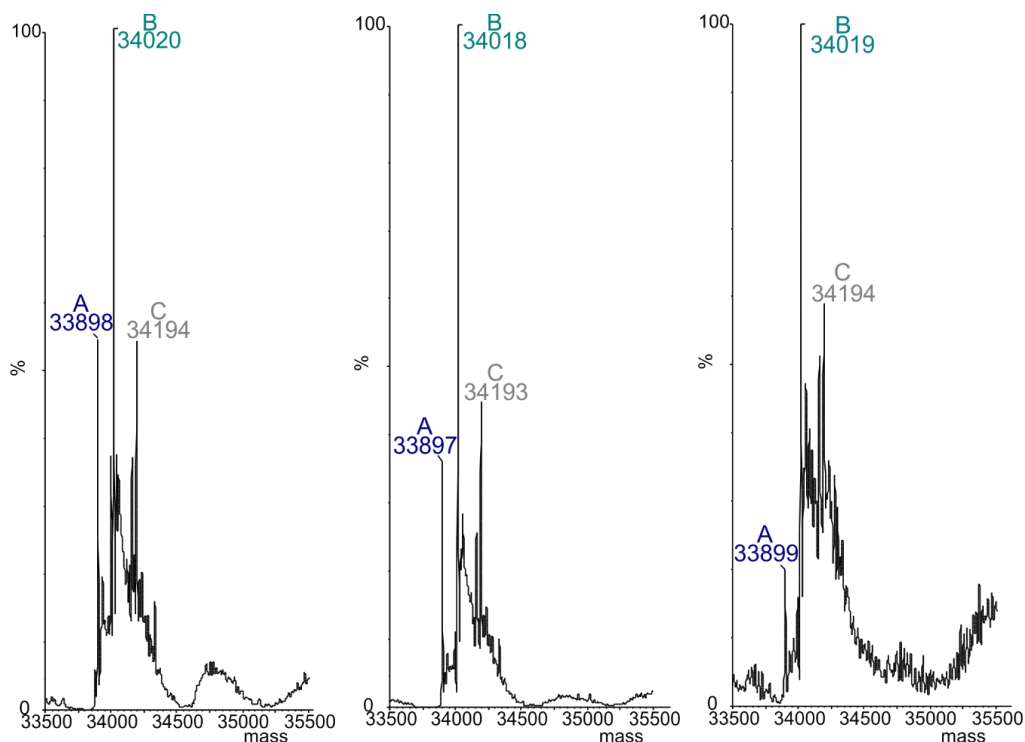


Figure 3.21 Mass spectrometer data of each peak from the size exclusion chromatography gel filtration of Y252Lni. **Left**, Peak 1 (unfolded); **Centre**, peak 2 (folded tetramer); **Right**, peak 3 (folded monomer). Blue, A, peak is Y252Dha (expected mass 33901 Da; Turquoise, B, peak is Y252Lni (expected mass 34022 Da); Grey, C, peak is Y252Lni +172 Da.

To ensure that the modified enzyme was folded in the same way as the wild-type, and not merely the same size as shown in figure 3.20, circular dichroism (CD) scans were performed on both the modified enzyme and the wild-type. CD scans used polarised ultraviolet light at different wavelengths to detect different secondary structures in proteins, i.e. β -sheets produce a peak at around 195 nm and a trough at around 217 nm, while α -helices produce a peak below 190 nm and two troughs at 208 and 222 nm.

Both modified and wild-type enzymes had CD scans performed according to section 2.4.18. The resulting spectra are shown in figure 3.22. The modified enzyme has a spectrum almost identical to that of the wild-type, so is most likely folded very similarly.

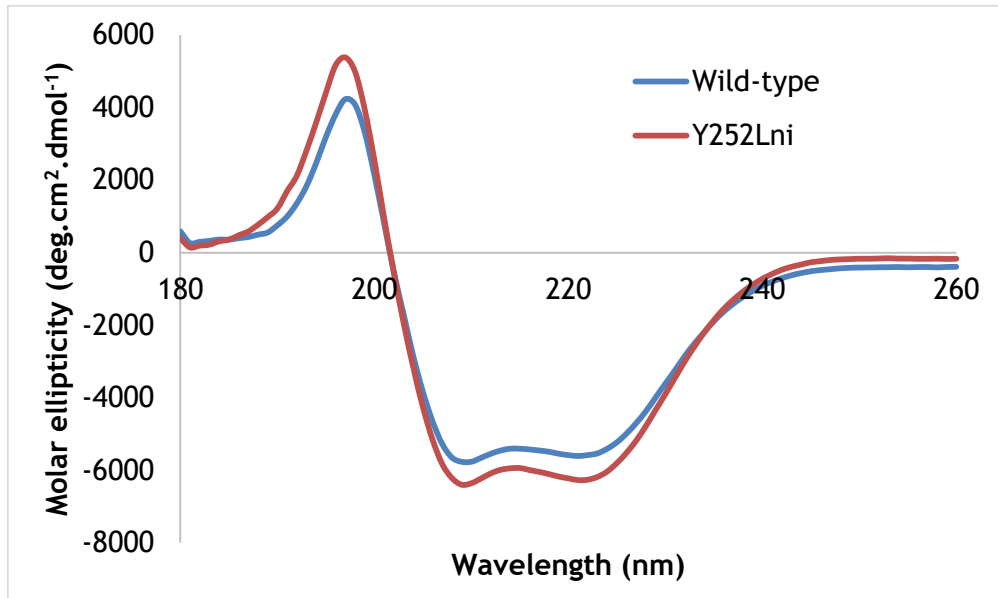


Figure 3.22 Circular dichroism (CD) scans of wild-type and Y252Lni enzymes. Wild-type is shown in blue and Y252Lni is shown in red.

Since the reaction is definitely catalytic and certainly requires each of the enzyme, pyruvate and the aldehyde to work, initial kinetic experiments were performed to find a time span in which the enzyme displays initial rate kinetics. As such 4 reactions were set up with 100 mM sodium pyruvate, 10 mM glucuronolactone, 50 mM sodium phosphate pH 7.4 and 0.25 mg/ml of either no enzyme, wild-type enzyme, gel filtered Y252Lni or non-gel filtered Lni, and samples were taken for 5 hours. The no-enzyme reaction showed no significant change in absorbance over the lifetime of the experiment, and the non-gel filtered Y252Lni increased in absorbance at a faster rate than the wild-type (figure 3.23). Additionally, the reaction was linear throughout the time span examined, so this length of time should be sufficient for the study of initial rate kinetics.

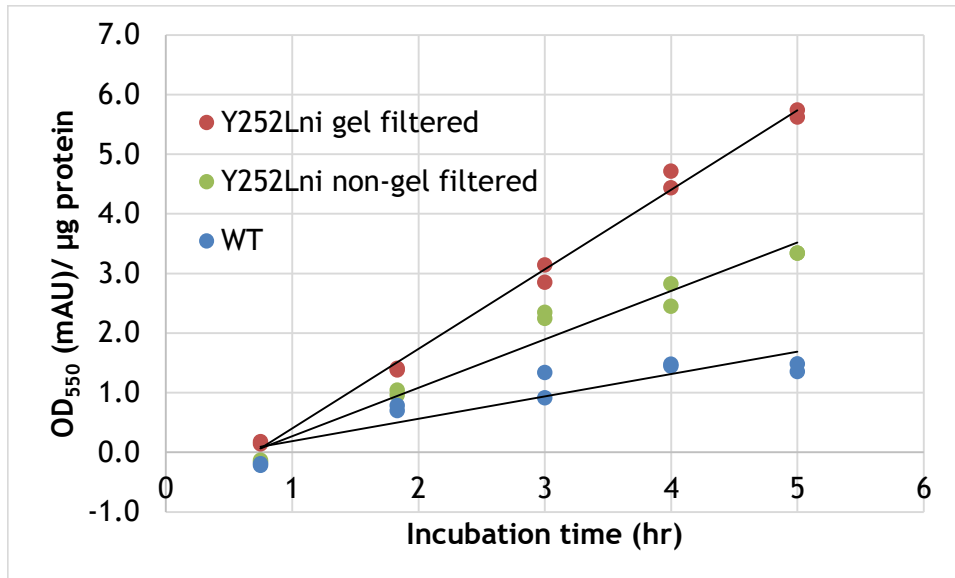


Figure 3.23 Increase of absorbance over time in the TBA assay of 3 reactions between pyruvate and glucuronolactone as catalysed by gel filtered Y252Lni (red) non-gel filtered Y252Lni (green) and wild-type enzyme (blue).

All the reactions were blanked with a time specific blank reaction OD value, and were quantified as per µg of total protein and as observed previously, the wild-type enzyme did display some activity for the reaction above the blank. Most interestingly though, the gel filtered Y252Lni had a higher reaction rate than the other protein samples; the non-gel filtered modified enzyme has a rate here approximately 2 times that of the wild-type, while the gel filtered enzyme displays a rate approximately 3.5 times that of wild type. The non-gel filtered protein displays approximately 60% the activity of the gel filtered protein. This is roughly in-line with the proportions observed during elution from the size exclusion chromatography, where 64% of the protein came off the column in a volume corresponding to correctly folded tetramer. These data taken together suggest that the active enzyme species is the fully folded tetramer, and the unfolded protein and folded monomer observed eluting from the column are not active.

3.5 Product confirmation

It was important to ascertain what the enzyme was producing to cause the positive TBA reaction, i.e. that the enzyme is performing the expected aldol reaction shown in figure 3.15 C, and that there is not a non-specific reaction occurring.

3.5.1 ^{13}C -NMR

Initially, pyruvate enriched with ^{13}C at the methyl group was used as a substrate in the reaction. This should produce two large peaks that are easily visible on ^{13}C NMR, corresponding to the ketone and hydrate form of pyruvate found in solution (Damitio et al., 1992). Two reactions were set up, both had ^{13}C -pyruvate at 100 mM and glucuronolactone at 10 mM in 50 mM sodium phosphate at pH 7.4 and with a D_2O probe. Additionally, to one reaction was added Y252Lni in 50 mM sodium phosphate at pH 7.4 to start the reaction. Early spectra showed identical peaks in both samples at 25 and 27 ppm. After five days, Distortionless enhancement by polarisation transfer-135 experiments (DEPT-135) showed a clear difference between the reactions with and without enzyme (figure 3.24).

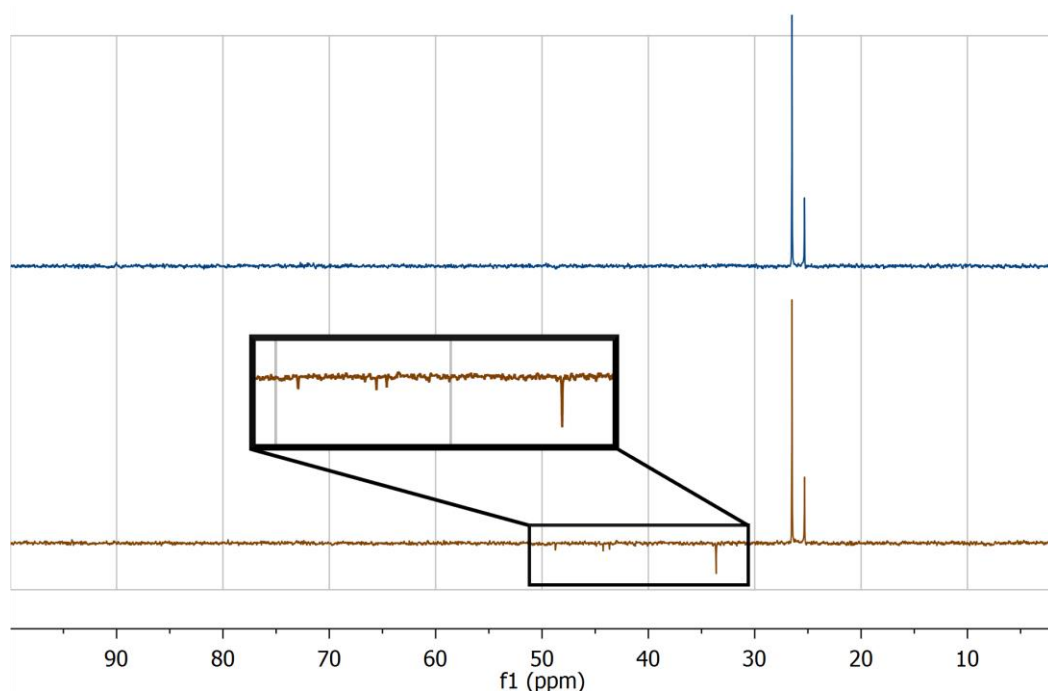


Figure 3.24 ^{13}C DEPT-135 NMR of a mixture of pyruvate, labelled with ^{13}C at the methyl group, and glucuronolactone alone (blue) and when incubated with enzyme (dark yellow) for 5 days. Four downwards pointing CH_2 peaks are present in the reaction with enzyme, corresponding the CH_2 groups that belong to the product.

The two peaks associated with pyruvate are still present, as would be expected from the large excess of pyruvate in the sample, but an additional 4 peaks at 33.5, 43.5, 44.0 and 48.0 ppm appeared in the NMR sample including enzyme. Not only were these peaks shifted downfield of the pyruvate peaks, as would be expected from the pyruvate undergoing the anticipated aldol reaction due to the carbon atoms now being in the proximity of more oxygen atoms, but these peaks were also pointing downwards. In a DEPT-135 experiment, carbons in CH₂ groups point downwards and carbons in CH₁ and CH₃ groups point upwards. This shows that the CH₃ group in pyruvate has been turned into four different CH₂ groups. This could be the formation of a number of different isoforms of the expected product, e.g. different anomers, diastereomers and potentially lactone formations.

3.5.2 ¹H-NMR

To obtain more information about the products, another reaction was set up as above, using natural abundance pyruvate but without the addition of a D₂O probe. After 14 days the TBA assay was used to confirm that the reaction had turned over almost all the glucuronolactone into product. Subsequently, the reaction was freeze dried to remove the H₂O, and then re-dissolved in 10% the original volume using D₂O, increasing all the concentrations by 10-fold to increase the intensity of the ¹H signal. Two NMR experiments were run on the sample, a Heteronuclear multiple-quantum correlation spectroscopy (HMQC) and a ¹H¹H-Correlation spectroscopy (COSY) experiment. The HMQC allowed correlation of the ¹H and ¹³C spectra, and in this data acquisition one proton peak correlated with a carbon peak that had a chemical shift identical to that of one of the product peaks found in the ¹³C enriched NMR spectra, 33.5 ppm (figure 3.25). With this the proton peak at 1.9 ppm could be assigned to the CH₂ group formed from the methyl group on pyruvate. This should have allowed the carbon chain to be traced from one end of the molecule to the other using the ¹H¹H-COSY experiment (figure 3.26), but the data from this experiment produced a complex spectrum that made it difficult to definitively trace a single molecule through.

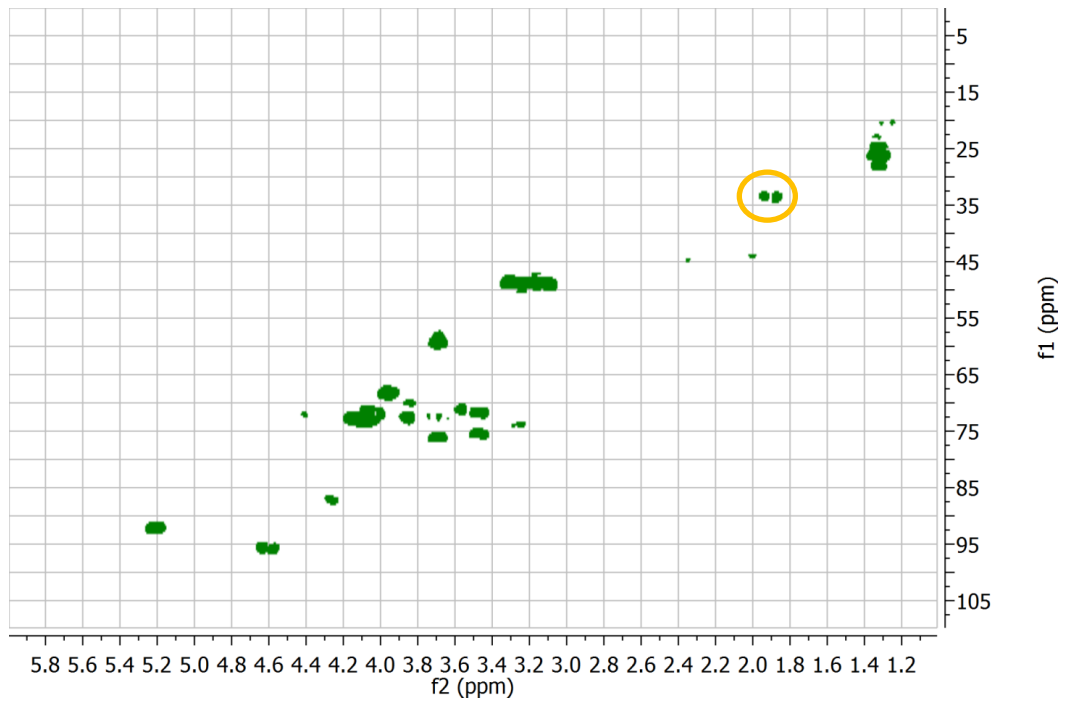


Figure 3.25 ^1H - ^{13}C HMQC NMR of a mixture of natural abundance pyruvate, glucuronolactone and enzyme after 14 days. The peak circled in orange has the same ppm as the most prominent product peak from the DEPT 135, and so the proton peak at 1.9 ppm must correspond to the CH_2 on the most abundant product.

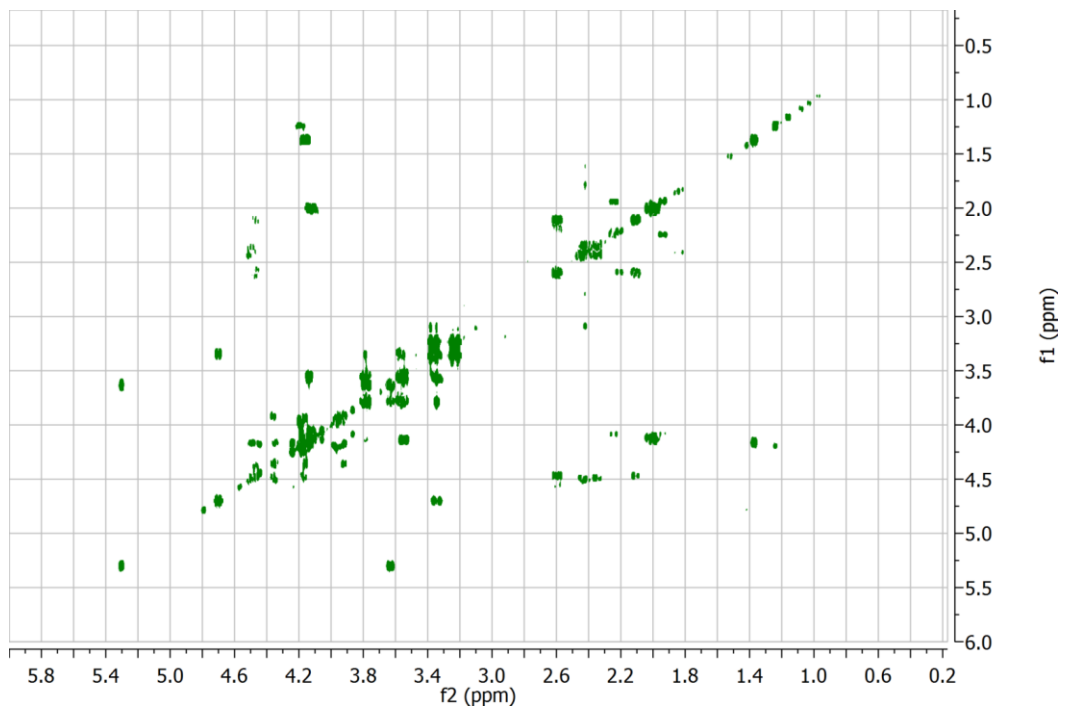


Figure 3.26 ^1H - ^1H COSY NMR of a mixture of natural abundance pyruvate, glucuronolactone and enzyme after 14 days.

All the product peaks on either the carbon or the proton spectra were in the anticipated regions for the respective atoms in the chemical environments they should be for the expected product. Whilst the exact forms of the product, e.g. bicyclic/monocyclic and what type of cycle etc., could not be identified, these results provide support to show the enzyme is performing the anticipated aldol reaction (figure 3.15 C). Some potential products are shown in figure 3.27.

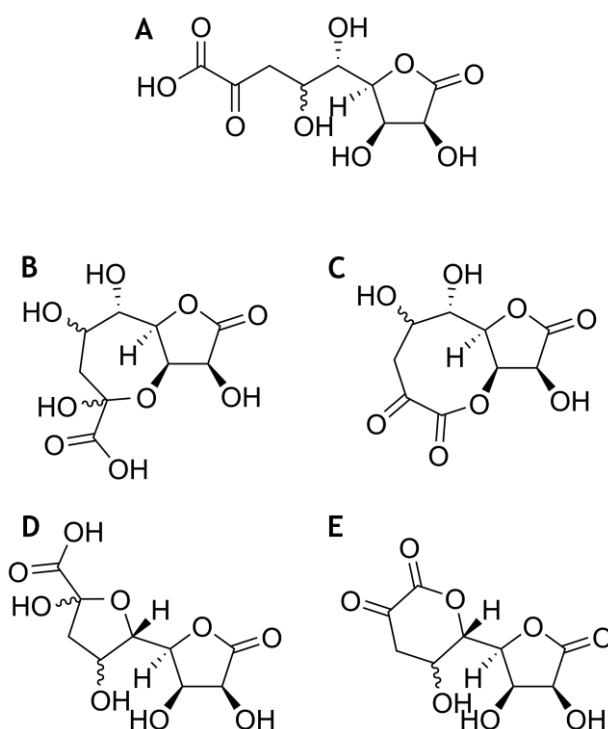


Figure 3.27 Possible products of the reaction between pyruvate and glucuronolactone. **A** The full length product immediately after catalysis. **B - E** are all potential conformations of **A** that would be caused by different ring formation mechanisms at different positions. **B** and **C** display rings formed on the hydroxyl of carbon 7, while **D** and **E** display rings formed on the hydroxyl of carbon 5. **B** and **D** are both ring-chain tautomers, while **C** and **E** involve a second lactone ring formation.

3.5.3 Mass spectrometry

Mass spectrometric analysis was also conducted on the freeze dried reaction used in section 3.5.2, (experiments performed by Sam Liver). The analysis had a number of different masses present, most notably 352.98, 462.98 and 572.98 Da (figure 3.28). The most intense of these peaks, 352 Da, equates to the expected mass of product, 264 Da, plus that of pyruvate, 88 Da. As there is a large excess of pyruvate in the freeze dried

sample, approximately 9:1 excess, it is not surprising that the product would be detected with pyruvate as an adduct. Additionally, the other, larger peaks equate to the sequential addition of one or more sodium pyruvate (110 Da) adducts onto the product-pyruvate complex. These data confirm the turnover of the product, and indicate that it is most likely the structure expected.

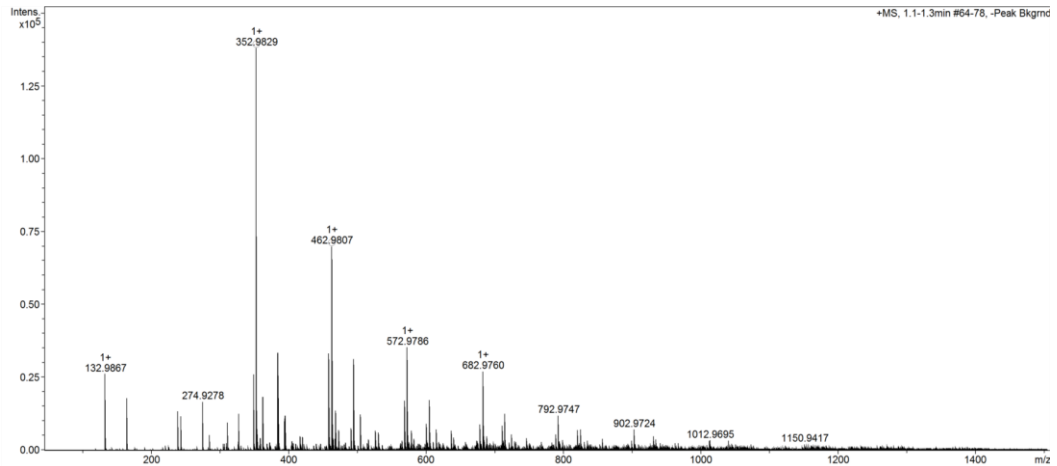


Figure 3.28 Mass spectrum analysis of a mixture of natural abundance pyruvate, glucuronolactone and enzyme after 14 days. 352.98 Da corresponds to the expected product mass + pyruvate, most likely flying as a pyruvate adduct.

3.6 Kinetics

The reaction performed by NAL is a sequential ordered bi-uni reaction. The pyruvate binds in the active site, forming a Schiff base with lysine at position 165. This is then followed by binding of the aldehyde substrate, allowing catalysis to occur (figure 3.29). Binding of pyruvate is necessary for catalysis to take place and the K_M of pyruvate in the wild-type enzyme is below 10 mM in most literature sources (Ferrero et al., 1996), well below the concentrations that have been used previously in this thesis. Therefore, only the kinetics of glucuronolactone were assessed, as the position of the modification should make little difference to the binding of pyruvate. As such, varying the concentration of pyruvate around the concentrations that have been used up to this point would make very little difference to the rate, and to greatly change the concentration of pyruvate would inevitably change the equilibrium position of the reaction considerably. The concentration of pyruvate used in these kinetics will be approximately 10-fold greater than K_M , meaning effectively all the enzyme molecules will have pyruvate bound.

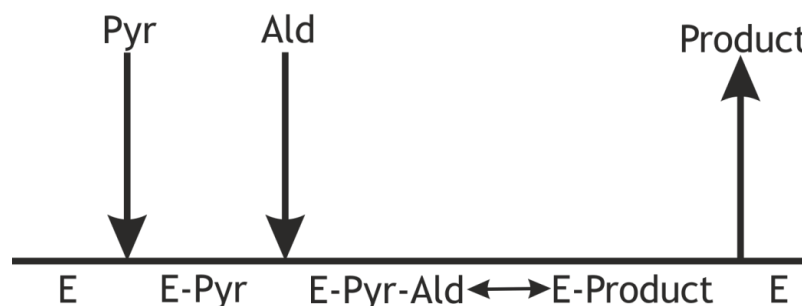


Figure 3.29 Cleland diagram representing the sequence of events during NAL catalysis. Arrows downwards represent binding events, arrows upwards represent unbinding events. E, enzyme; Pyr, pyruvate; Ald, aldehyde.

3.6.1 Plate layout

The plate layout was designed to maximise the amount of data that could be extracted from each plate. For both wild-type and Y252Lni enzymes, 8 different concentrations of glucuronolactone were used, from 0 to 15 mM, and these reactions were set up in a plate. Samples of 100 μ l were taken in duplicate, to increase the reliability of the data, and added to 11 μ l of 12% trichloroacetic acid (TCA) in a separate plate for each enzyme. Samples were taken at 5 points throughout the 6 hour incubation to allow an accurate rate to be determined. The two final columns of the plate were

used for duplicate standard curves to allow calculation of concentration from absorbance. The generalised plate layout can be seen in figure 3.30.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.5 hr		1 hr		2 hr		4 hr		6 hr		4000 μ M Std.	
B	0.5 hr		1 hr		2 hr		4 hr		6 hr		2000 μ M Std.	
C	0.5 hr		1 hr		2 hr		4 hr		6 hr		1000 μ M Std.	
D	0.5 hr		1 hr		2 hr		4 hr		6 hr		500 μ M Std.	
E	0.5 hr		1 hr		2 hr		4 hr		6 hr		250 μ M Std.	
F	0.5 hr		1 hr		2 hr		4 hr		6 hr		125 μ M Std.	
G	0.5 hr		1 hr		2 hr		4 hr		6 hr		62.5 μ M Std.	
H	0.5 hr		1 hr		2 hr		4 hr		6 hr		0 μ M Std.	

Figure 3.30 Table showing the generalised layout of the kinetic analysis plates. Times shown are the times that sample was taken, intensity of blue represents the concentration of glucuronolactone in each reaction, from 0 mM in white at the top to 15 mM in dark blue at the bottom. Concentration and position of the *N*-acetylneuraminic acid standard samples are shown in columns 11 and 12.

3.6.2 Kinetic parameters

As seen before, Y252Lni was better at catalysing the reaction between glucuronolactone and pyruvate than the wild-type enzyme (figure 3.31). However, all kinetic runs showed significant substrate inhibition above 10 mM of glucuronolactone with both enzymes. This complicates the data fitting as the classic Michaelis-Menten equation is unsuitable for use with substrate inhibition. Fortunately, there is a derivation of the Michaelis-Menten equation that accounts for substrate inhibition. Equation (3.1) shows the normal Michaelis-Menten equation while equation (3.2) shows the substrate inhibition equation.

(3.1)

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

(3.2)

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

Kinetics were performed five times in a single week in an attempt to increase the reproducibility of the data. The average data from the five repeats are shown in figure 3.31.

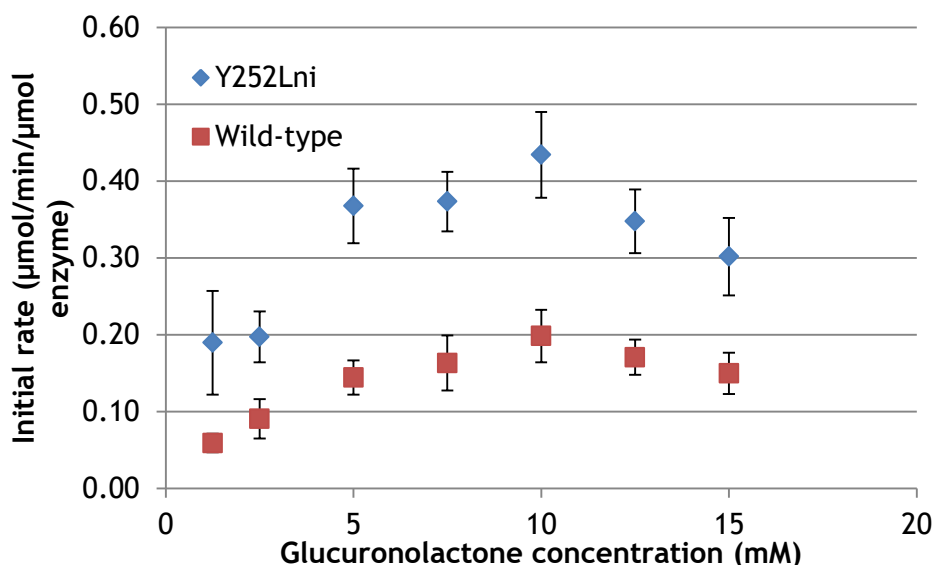


Figure 3.31 Average rate of reaction at different glucuronolactone concentrations for both wild-type and Y252Lni as calculated from the TBA assay. Rate with Y252Lni is shown in blue and wild-type is shown in red. Error bars are standard error, n=5.

As has been seen previously, Y252Lni displayed a consistently higher rate than the wild-type at all concentrations of substrate over all assays. There was however significant day-to-day variation in the activities observed, while the general theme was maintained in almost all non-blank samples i.e., that the modified enzyme displayed a greater rate than the wild-type, little else was consistent between days.

When these data are fit using the substrate inhibition formula, the parameters given in figure 3.32 are produced.

	Y252Lni	Wild-type
$K_{m \text{ app}}$ (mM)	38 (± 190)	80 (± 500)
$k_{\text{cat app}}$ (min^{-1})	4.3 (± 20)	3.3 (± 20)
K_i (mM)	1.6 (± 8.0)	1.1 (± 7.0)

Figure 3.32 Summary table of the fit statistics of the average results of the kinetic characterisation. All values are apparent for glucuronolactone at a pyruvate concentration of 100 mM. Errors given in brackets are the standard errors of the fit statistics.

The results suggest that the modified enzyme has a higher maximum rate and reaches that maximum rate at a lower concentration of substrate than the wild-type enzyme, but the errors on the fit are very high, so firm conclusions are not possible. Ideally, these experiments would have been repeated in an attempt to obtain less variable data, but time constraints meant that this was not possible.

3.7 Summary

Initially, a pre-existing assay was assessed for feasibility and optimised for use as a screening assay. This was successful, and resulted in a large increase in maximum possible absorbance from the assay. Subsequently, 900 different enzyme/aldehyde combinations were screened for enhanced activity over wild-type for each particular aldehyde. A number of positive hits were found according to plate reader data, but when the plates were visually inspected, only one of the top ten hits in the plate reader data also appeared pink by eye, and as such, this hit was chosen for further characterisation.

The hit, Y252C modified to produce a lanthionine residue, catalysing a reaction between pyruvate and glucuronolactone, was re-made and the modification procedure was optimised for production of pure Y252Lni. This allowed the enhanced activity of Y252Lni for glucuronolactone over the wild-type to be confirmed, and it gave a rate approximately 3.5x higher than that of the wild-type enzyme, when using 100 mM pyruvate and 10 mM glucuronolactone.

Reactions were set up, with both pyruvate enriched with ^{13}C at the methyl and a large scale reaction with regular pyruvate. This allowed NMR experiments to be performed which confirmed turn over of the substrates into products, and also allowed initial NMR characterisation of one of the most abundant products.

In an effort to more completely characterise the kinetic properties of the modified enzyme, apparent glucuronolactone kinetics were performed, but these gave variable results, with the only firm conclusion being what was found previously, that Y252Lni performs the reaction between pyruvate and glucuronolactone at a faster rate than the wild-type enzyme.

Nevertheless, this position was deemed interesting, and the reaction between pyruvate and glucuronolactone was unique in that it uses a cyclic substrate, where NAL has been shown to only accept sugars in their linear forms. As such, further research was conducted into this position and modification in an attempt to elucidate the effect of the modification on the reaction.

Chapter 4 - Studying substrate interactions with position 252 of *S. aureus* NAL

While accurate kinetics could not be obtained, chapter 3 showed that the mutation of a tyrosine to a lanthionine at position 252 in *sa*NAL increases the rate of catalysis of the aldol reaction between pyruvate and glucuronolactone. To find out if any of the natural side chains could mimic this effect on the rate, and to try to gain more molecular insights into the change in substrate preference, a saturation library was constructed.

4.1 Saturation mutagenesis at position 252

4.1.1 Saturation library construction

The saturation library of primers was designed using a tool available on the Agilent website (www.genomics.agilent.com/primerDesignProgram.jsp) to predict the sequence of primers that will incorporate the mutation desired. The Quikchange II kit was used according to manufacturer's instructions and *E. coli* strains expressing *sa*NAL with all 20 possible canonical amino acids at position 252 were obtained (Figure 4.1) (for method see section 2.3.2).

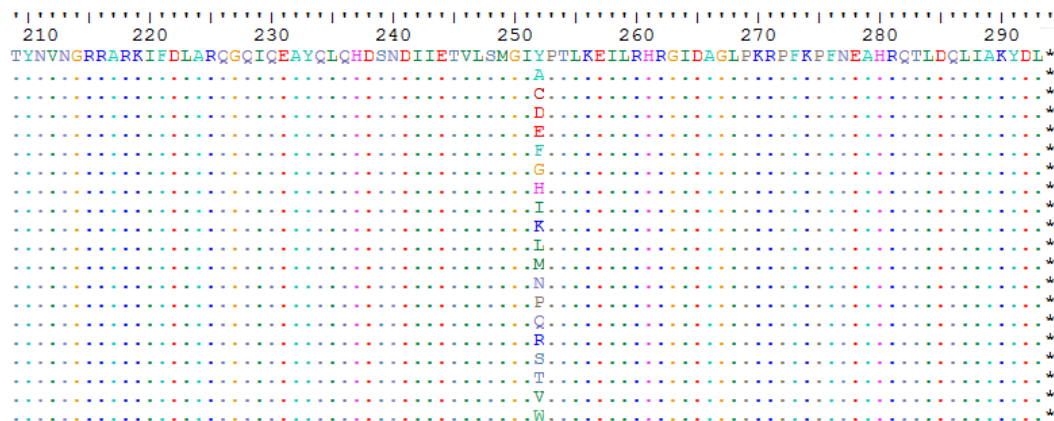


Figure 4.1 Translated DNA sequencing of all members of the saturation mutagenesis library at position 252, with wild-type, Y252, at the top, and all the other canonical amino acid mutants ordered alphabetically by their single-letter codes. Only a part of the sequencing results are shown for brevity, but there were no other changes in sequence in the remainder of the protein.

4.1.2 Saturation library purification and activity

With all members of the saturation library transformed into *E. coli* capable of high levels of expression (for genotype see section 2.1) the activity of each mutant was tested. The library was split into two groups for ease of handling during the expression and purification; A-L and M-Y. These two groups were expressed on different days and the spun-down pellet of cells were stored at -80 °C, after which each group of expressions, A-L and M-Y, was purified on consecutive days (see sections 2.4.3 and 2.4.4 for methods). This was done to reduce the amount of time that the first group of variants that were purified spent in the fridge, as their stability is unknown, while still allowing all 20 variants to be tested for activity at the same time. The purifications were analysed by SDS-PAGE and the resulting solutions were found to be pure (figure 4.2).

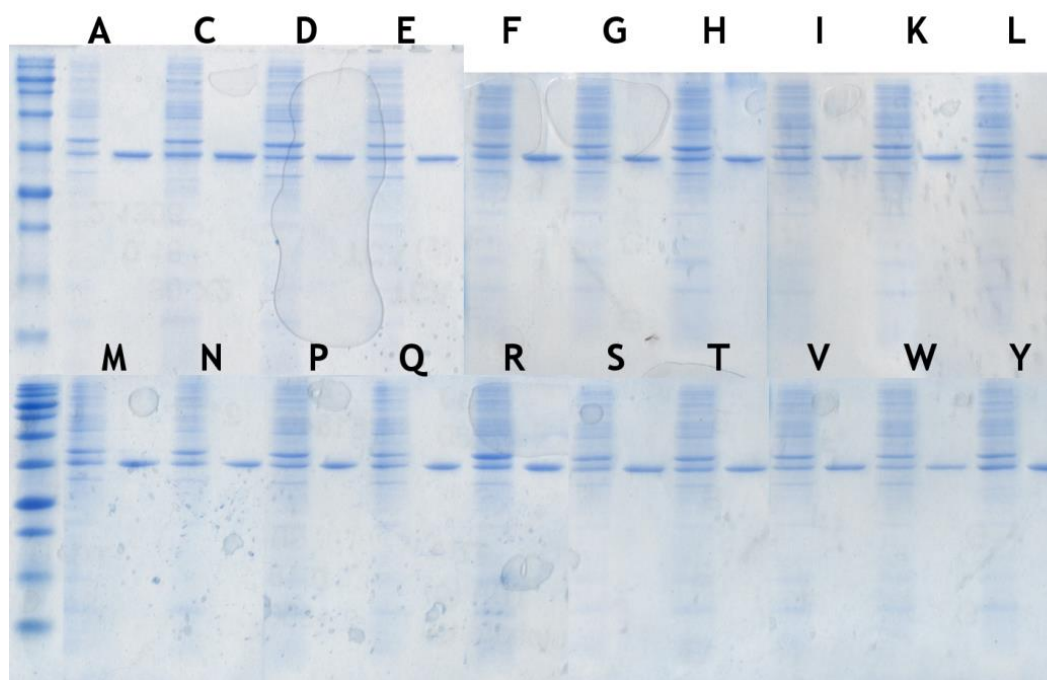


Figure 4.2 Composite SDS-PAGE gel image of all 20 members of the canonical amino acid site-directed mutagenesis library at position 252. Letters above the pairs of lanes represent the amino acid at position 252, the left lane of each pair is the soluble fraction of lysed cells of each expressed protein variant, diluted 50-fold, while the right lane of each pair of lanes is the protein solution after all purification steps have been performed, diluted 40-fold.

The concentrations of each purified protein were measured and all enzymes were diluted to 0.5 mg/ml to allow comparison of the reactions directly. To reduce the day-to-day variability seen in previous experiments (section 3.6.2) all 20 enzymes were tested on the same day. As such, each

enzyme was only assessed at a single concentration of pyruvate and glucuronolactone to save space on the assay plate. The final concentrations of the reactions were 0.25 mg/ml enzyme, 10 mM glucuronolactone, 100 mM sodium pyruvate, 50 mM sodium phosphate, pH 7.4. The reactions were run for 5 hours, duplicate samples were taken every 75 minutes and stopped with TCA. Full methods are given in section 2.4.15. Samples were also taken from a blank reaction to account for any non-enzymatic rate of change in absorbance, and a standard curve of Neu5Ac was added to calculate the concentration of product in each well.

This experiment allowed a rate of product formation to be calculated for each variant and compared to the rate of product formation by wild-type. This difference in rate compared to wild-type could then be used along with data from section 3.4.3, performed using an identical method, to compare the nCAA to the canonical amino acids.

The most obvious result from the data gathered is that the wild-type enzyme, Y252, is amongst the least active canonical amino acids at this position for the reaction between glucuronolactone and pyruvate. The difference in activity between the proteins is large and can be seen from visual inspection of the plate (figure 4.3). When the plate reader data is examined, the rate observed in the wild-type enzyme of 13 mAU/hr is similar to that observed in previous experiments, and so the rest of the data is likely to be as accurate as the wild-type i.e. the wild-type activity in this experiment is comparable to previous experiments. The mutant proteins (figure 4.4, blue bars) can be split up into three groups based on their activity for glucuronolactone and pyruvate. Enzymes with a high activity, more than 3-fold more active than wild-type, were Ala, Cys, Glu, Gly, His, Ile, Met, Asn, Gln, Ser, Thr, Val and Lni. Enzymes with a moderate activity, between 3-fold and the same activity as wild-type, were Asp, Phe, Leu, Pro and Trp. Finally, enzymes with a low activity, equal to or below that of wild-type, were Lys, Arg and Tyr.

This puts Y252Lni, with an activity 3.5-fold higher than wild-type, towards the top of the list when compared to the canonical amino acids. To further examine the structural implication of the changes at position 252, the activities of the canonical amino acid variants were compared to the molecular weight of the amino acid at position 252. A rough trend can be observed, showing that the larger amino acids tend to show lower activities (figure 4.5, blue dots). While this correlation is not very strong ($r^2=0.40$), it could be explained via removal of bulk in the active site.

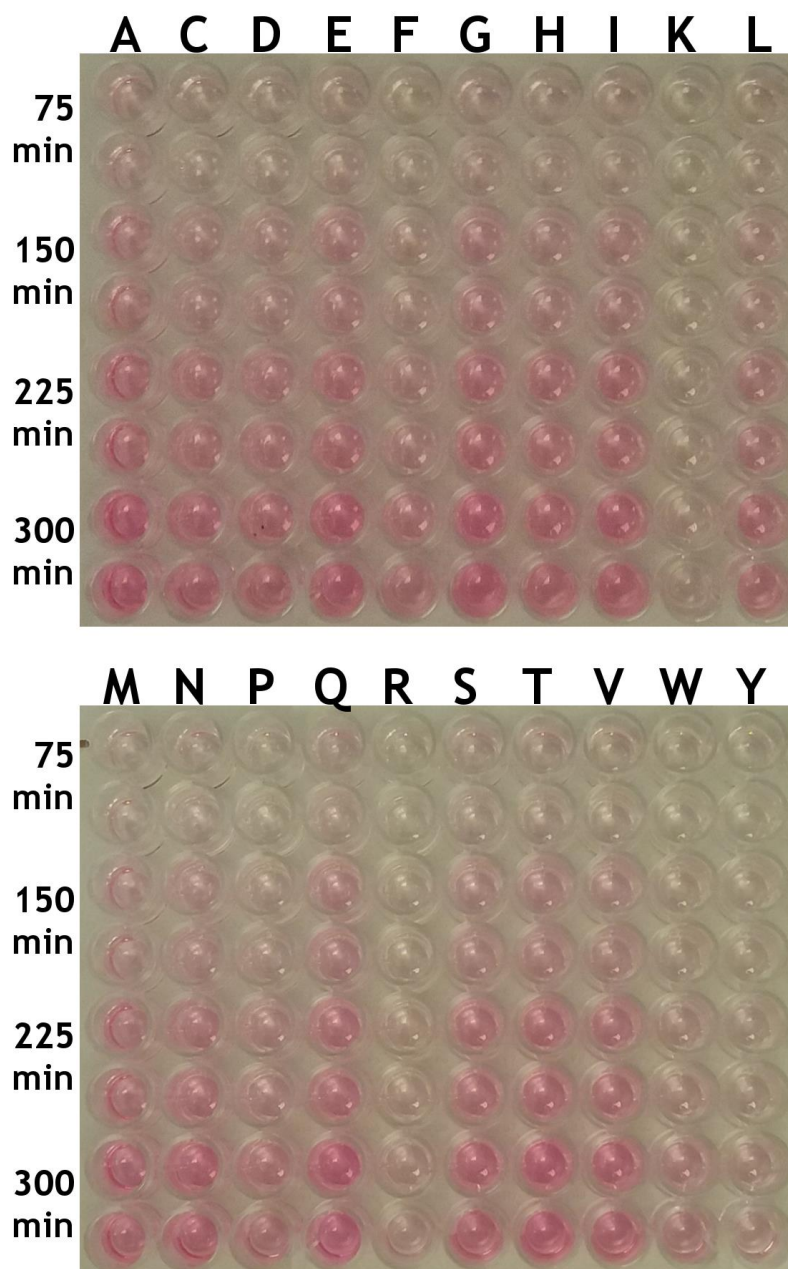


Figure 4.3 Photographs of the TBA assay plates showing the increase in colour over time of all the members of the saturation library at position 252 performing the synthetic aldol reaction between pyruvate and glucuronolactone. Each column represents the samples from one reaction, each using a different enzyme. The enzyme used in each reaction is shown above each column. Samples were taken in duplicate at the times indicated to the left of the photographs, and as such each time label refers to the two rows to which it is adjacent.

Lower mass amino acids will in general take up less space than higher mass amino acids. This removal of bulk from the active site would create more space, potentially allowing the cyclic substrate more room to enter and move into a catalytic conformation. However, when the activity of Lni is

compared on this graph, it does not fit the general trend (figure 4.5, red dot) having both a relatively high activity and a high mass side chain.

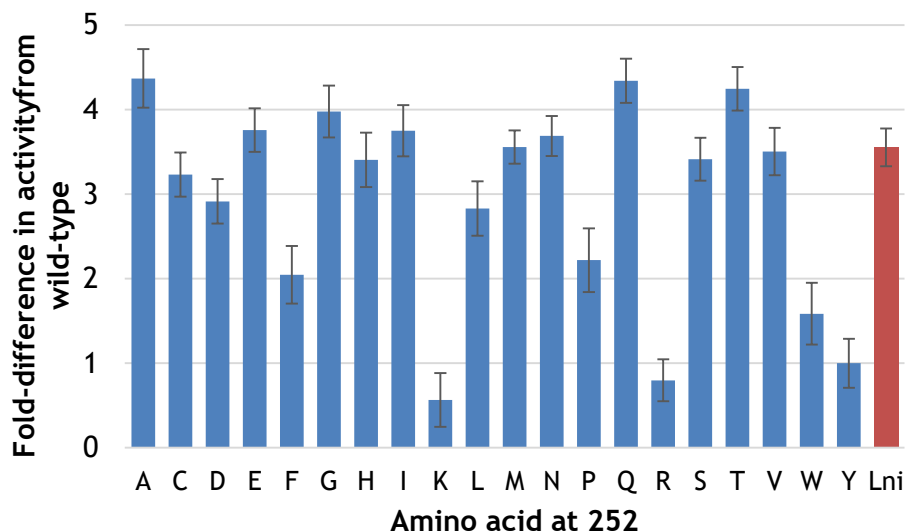


Figure 4.4 Fold-difference in activity from wild-type while performing the aldol reaction between pyruvate and glucuronolactone for each of the canonical amino acid variants at position 252 (blue) and the nAA lanthionine at position 252 (red) as found in the TBA assay. Error bars shown are the standard error of the rate.

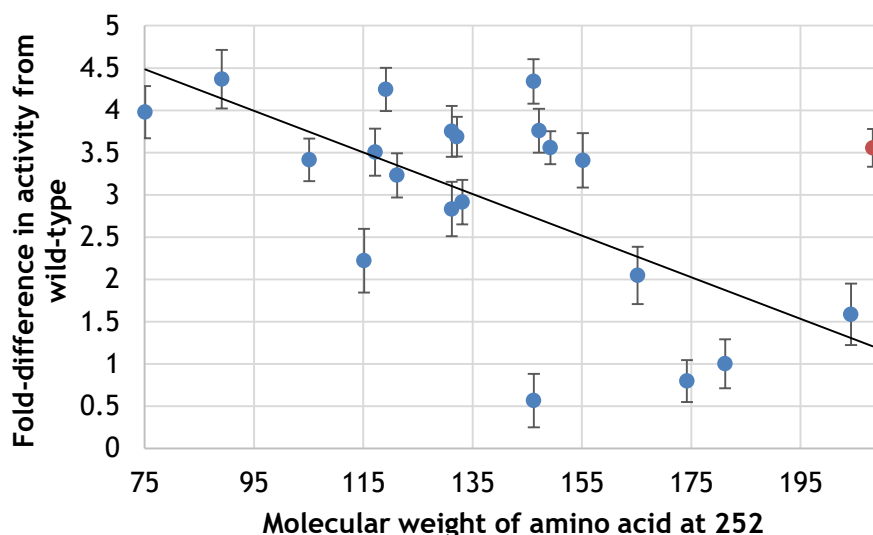


Figure 4.5 Fold-difference from wild-type as in figure 4.4, compared to the molecular weight of the amino acid at position 252. Canonical amino acids at 252 are shown as blue dots and the nAA lanthionine is shown in red. Error bars are the standard error of the rate.

The saturation library data shows that the Y252Lni enzyme is likely not a candidate for the industrial or large-scale catalysis of the synthesis reaction between pyruvate and glucuronolactone, as canonical amino acid

variants are easier to produce and in some cases have a higher activity. That the Y252Lni enzyme still displays enhanced activity compared to the wild-type, even with its larger side chain mass, suggests that the mechanism by which it achieves the rate enhancement is different to that of the canonical amino acids. Therefore, a structural investigation into the mechanism by which Y252Lni enhances the aldol reaction between could reveal useful mechanistic insights.

4.2 X-ray crystallographic studies of Y252Lni *sa*NAL

The structure of the Y252Lni variant was investigated to provide an explanation for its activity in the reaction between pyruvate and glucuronolactone.

4.2.1 Crystal formation, data collection and refinement

Crystals were formed using a published method for crystallisation of wild-type and chemically modified *sa*NAL (Timms et al., 2013, Windle et al., 2017) (see section 2.4.16 and 2.4.17). Briefly, size exclusion chromatography was performed on a sample of Y252Lni modified protein and the peak corresponding to correctly folded tetramer was collected to ensure a folded, homogeneous population of protein. This was concentrated to 9 mg/ml for crystallisation. The crystallisation conditions used were 200 mM sodium chloride, 100 mM Tris.HCl, 18-28% PEG 3350, pH 7-8.5. The crystals were grown on a 24-well plate using the hanging drop method, each well contained 3 protein crystallisation drops containing a total volume of 3 or 4 μ l, the protein solution: mother liquor ratios used were 2:1, 2:2 and 1:2. Three replicate plates were produced, and crystals had formed on all plates after 14 days.

The enzyme structure in complex with pyruvate would provide more relevant information with respect to aldehyde binding in a catalytically useful conformation compared to the structure without pyruvate, as the pyruvate binding must occur before the aldehyde binds. As such, the crystals selected (figure 4.6) were soaked in a cryoprotectant containing pyruvate. The crystals were first soaked in a cryoprotectant consisting of mother liquor with an additional 15% PEG 400 and 100 mM sodium pyruvate, then the crystals were transferred to a solution of mother liquor with 20% PEG 400 and 100 mM sodium pyruvate, then finally to a solution of mother liquor, 25% PEG 400 and 100 mM sodium pyruvate. Each soak lasted for approximately 30 s, and crystals were subsequently flash cooled using liquid nitrogen.

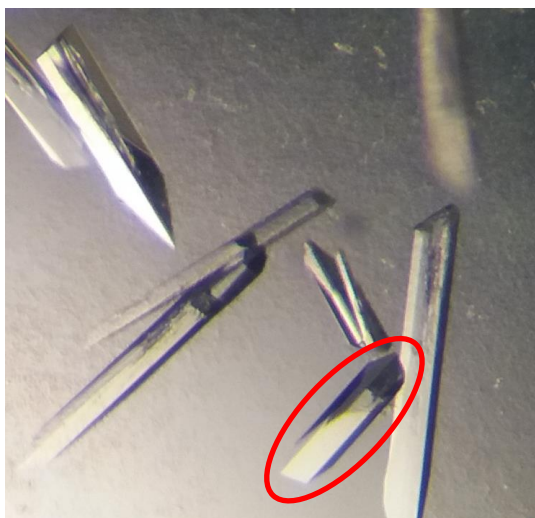


Figure 4.6 Y252Lni crystals examined using X-Ray crystallography. Circled in red is the crystal from which the dataset used to produce the Y252Lni structure was derived. Crystal growth conditions were 200 mM sodium chloride, 100 mM Tris.HCl, 24% PEG 3350, pH 8.5.

Data collection was performed at Diamond Light Source (Oxfordshire, UK) on beamline I24 at 100 K and the structure was derived from a single crystal (figure 4.6, circled in red) and the resulting diffraction was at a moderately good resolution, 1.95 Å. The dimensions of the unit cell were 81.05 Å by 124.27 Å by 179.54 Å with angles of 90.00, 97.77, and 90.00 and the space group of the cell was group P2₁. The Matthews Coefficient (Matthews, 1968) estimates the solvent content of the unit cell using the unit cell dimensions and the molecular weight of the protein. When calculated for this structure, the Matthews coefficient indicated that the unit cell consisted of 12 monomers, and as such, 12 monomers were searched for during molecular replacement. Molecular replacement was performed with PHASER (McCoy et al., 2007) using a monomer of the wild-type structure with pyruvate bound (PDB code: 4ah7) as the search model. Molecular replacement showed that the protein was a tetramer, the same as the wild-type and previously solved structures of modified *sa*NAL (Timms et al., 2013, Windle et al., 2017).

The CCP4 suite program REFMAC5 (Murshudov et al., 2011) was used to perform all rounds of refinement. The starting model used was the wild-type structure with pyruvate bound and the residue at 252 modelled as alanine to prevent bias in the model. Rigid body refinement was carried out initially, followed by iterative rounds of restrained refinement and model building using COOT (Emsley et al., 2010). The $2F_{\text{obs}}-F_{\text{cal}}$ and $F_{\text{obs}}-F_{\text{cal}}$ electron density difference maps show that the alanine modelled into

position 252 does not accurately describe the electron density found in any of the 12 monomers in the unit cell. In 11 out of the 12 monomers, the electron density is clearly larger than the alanine model, but in 1 of the monomers there was no evidence of electron density beyond the alpha carbon of the modelled alanine. In those monomers which included density at 252, there was a large atom in the γ -position of the side chain, which is consistent with a sulphur atom present in the nCAA side chain (figure 4.7).

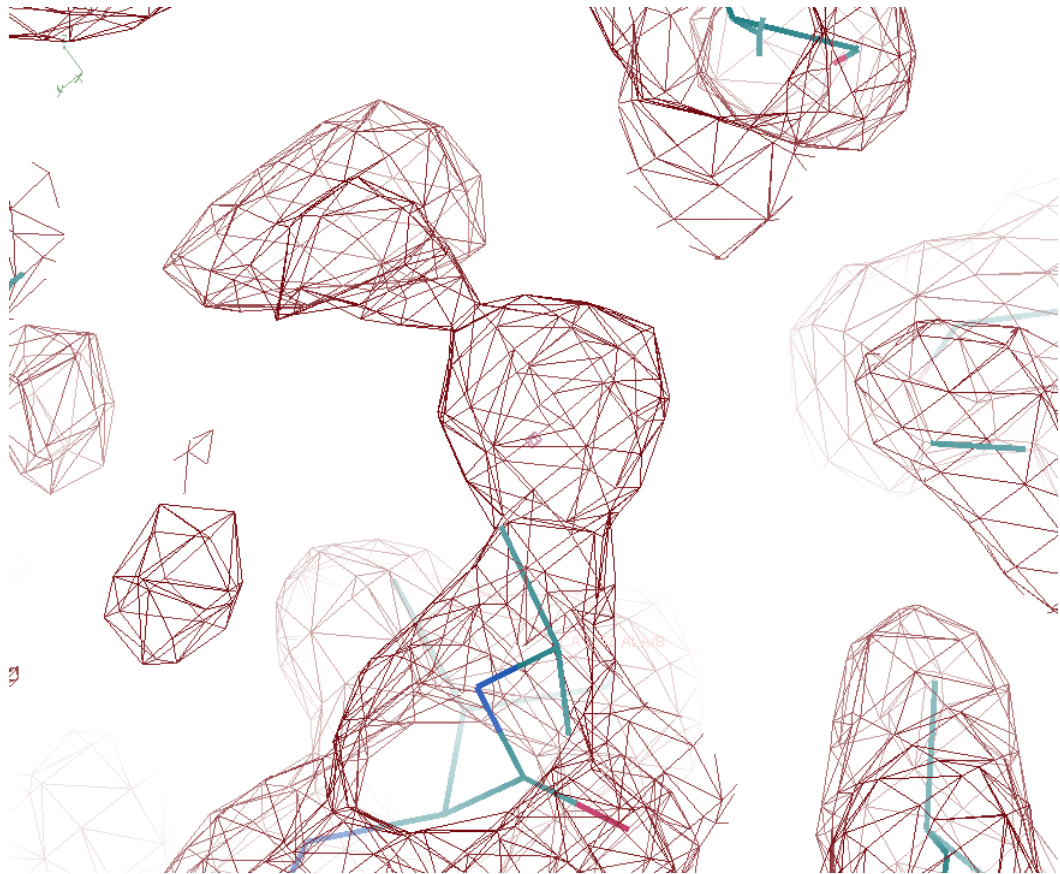


Figure 4.7 The modelled Y252A structure and the $2F_{\text{obs}} - F_{\text{cal}}$ electron density map of the Y252Lni crystal overlaid. The modelled structure is shown as blue sticks and the electron density map is a red grid. Taken from WinCOOT (Emsley et al., 2010).

4.2.2 Description of the modified protein structure

The structure of Y252Lni saNAL is a homotetramer, each monomer displaying the same α/β TIM-barrel fold observed in the wild-type protein (Timms et al., 2013). The overall structures of the modified and wild-type proteins are almost identical (RMSD = 0.620 Å) (figure 4.8) showing that incorporation of the nCAA at this position has very little effect on the fold of the protein, and this similarity is even clearer in the active site, bar the amino acid at 252 (figure 4.9).

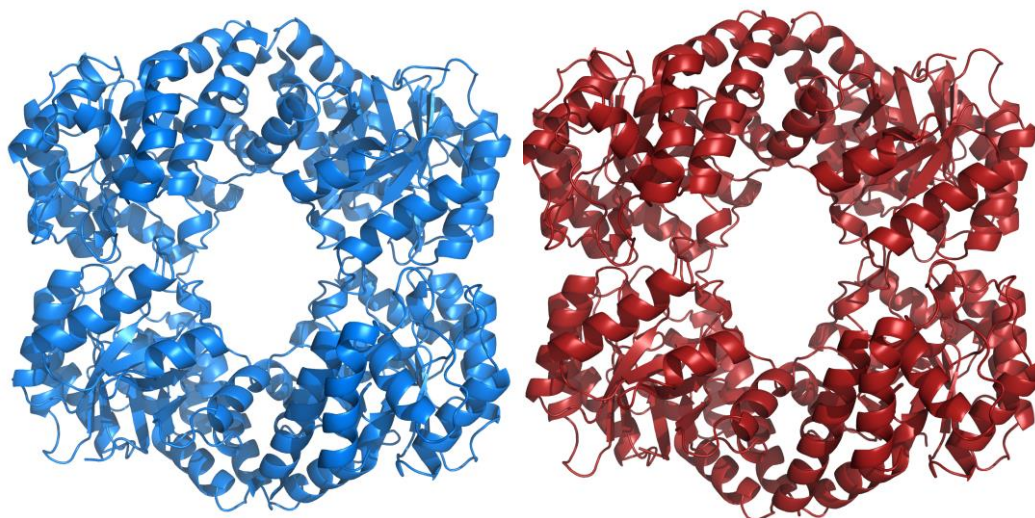


Figure 4.8 Cartoon depiction of the wild-type (blue) and Y252Lni modified (red) enzymes, showing all four monomers in a complete homotetramer. The structures are very similar, but some differences can be seen around the edges of each structure where the amino acids are less tightly held in position.

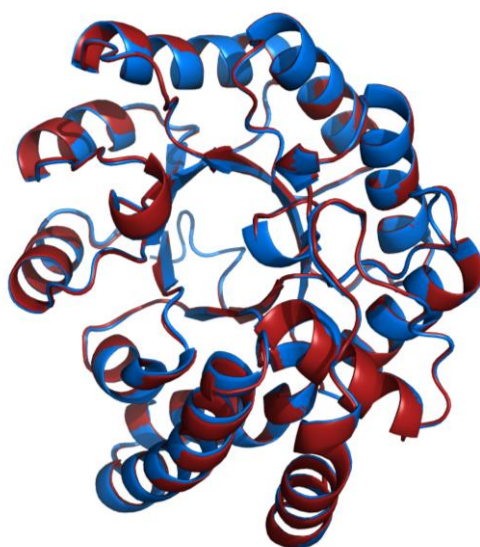


Figure 4.9 Cartoon depiction of one monomer of wild-type and Y252Lni overlaid (blue and red). The two monomers overlay almost perfectly (RMSD = 0.235 Å).

The positions of the main catalytic side chains are also very similar between the wild-type and the modified protein. The catalytic lysine in complex with pyruvate is present in all monomers, and in the same position as in the wild-type enzyme. Residues Y137 and T167, which are proton sinks and sources during catalysis (Daniels et al., 2014), are also positioned similarly in both the wild-type and the modified structures. Y137 and the pyruvate attached to K165 are in slightly different positions while the T167 in both structures overlays almost perfectly. In the modified structure, Y137 is shifted by about 1 Å from the wild-type position, in the direction away from the pyruvate. However, this change in position coincides with a change in position in the pyruvate in the modified

structure, in the same direction, by 0.7 Å (figure 4.10). As such, these two changes in position occur in the same direction and by roughly the same distance, and so probably compensate for each other.

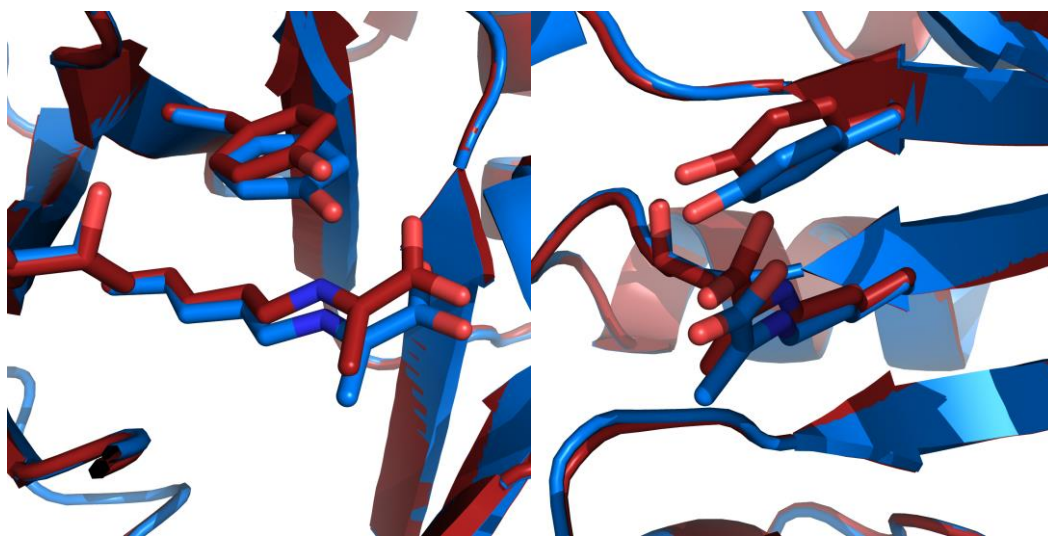


Figure 4.10 Overlaid monomers of wild-type and Y252Lni structures. Residue side chains of Y137, K165-pyruvate and T167 are shown as sticks while the rest of the monomers are shown as backbones represented by cartoons. The structure on the left was rotated 90° to the left to produce the picture on the right. Light blue shows wild-type monomer and residue carbon atoms, while dark red shows Y252Lni monomer and carbon atoms, other atoms are coloured according to their identity: Dark blue, nitrogen; light red, oxygen.

There is only one residue, F190, in a significantly different position in the active site when the wild-type enzyme is compared to Y252Lni. This amino acid occurs in two different rotamers in the two structures (figure 4.11). In the wild-type the phenylalanine side chain projects into the active site, occluding it slightly, whilst in the modified enzyme the phenylalanine side chain points into the bulk of the protein, leaving the active site more open. This rotamer difference means the wild-type active site is 5 Å narrower than the modified active site. The increased volume available in the active site could help the modified enzyme accommodate a more bulky substrate, like glucuronolactone.

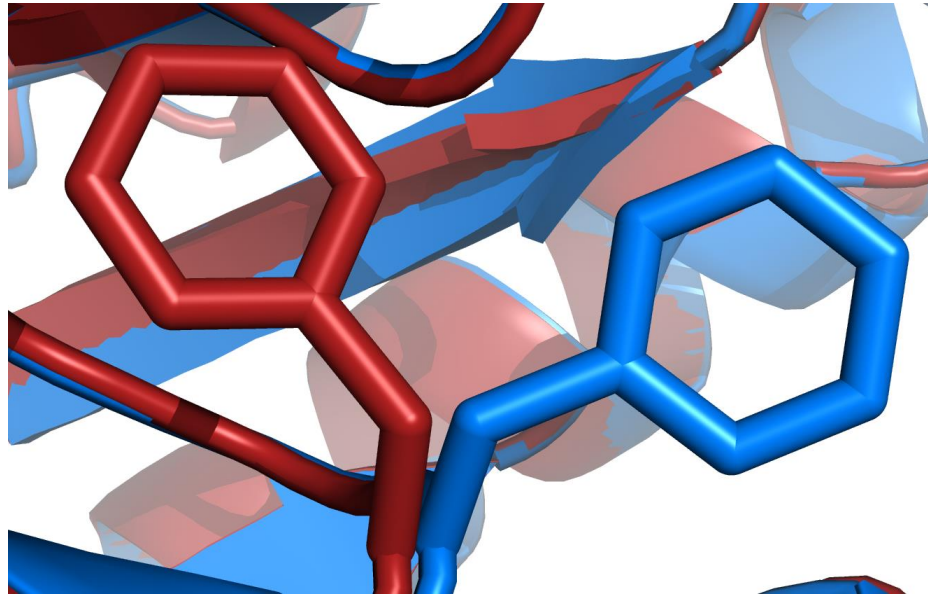


Figure 4.11 Overlaid monomers of wild-type and Y252Lni structures. Residue side chains of F190 are shown as sticks while the rest of the monomers are shown as backbones represented by cartoons. Light blue shows wild-type monomer and residue carbon atoms, while dark red shows Y252Lni monomer and carbon atoms.

4.2.3 Side-chain structural features

The lanthionine side chain displayed good density in all the monomers in which it was found, 11 out of the 12 monomers and was in a similar position in all 11 (figure 4.12). This suggests that it is kept fairly rigidly in place by interactions with the other amino acids in the active site. The temperature factor of the side-chains at 252 was 45.45 on average, compared to an average temperature factor of 29.89 across the entire structure. The modified amino acid, lanthionine, has a number of features that should allow it to make a number of different polar interactions and hydrogen bonds using the carboxylate and amino groups, and also the thioether group. However, there are no clear polar groups in close proximity to the side chain (figure 4.13), so there is likely no electrostatic interaction between the side chain and the rest of the active site.

The side chain is most likely held in place by repulsion from the hydrophobic residues surrounding it, which are well ordered. The presence of a hydrophobic pocket in the enzyme at this region could be anticipated due to the identity of the wild-type amino acid at this position being either tyrosine, in the *S. aureus* protein, or phenylalanine, in the *E. coli* protein, both of which have large hydrophobic regions and would need a hydrophobic pocket to hold them stable in the active site.

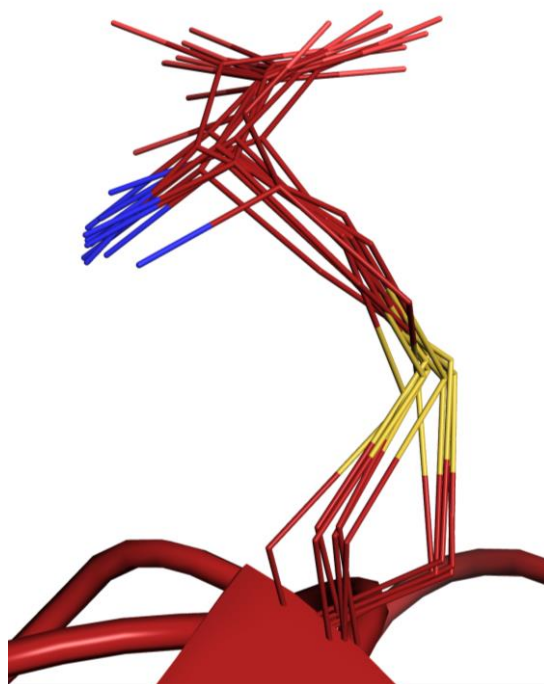


Figure 4.12 The lanthionine side chains at 252 overlaid. The side chains of every monomer overlay fairly well, showing the side chain is well structured.

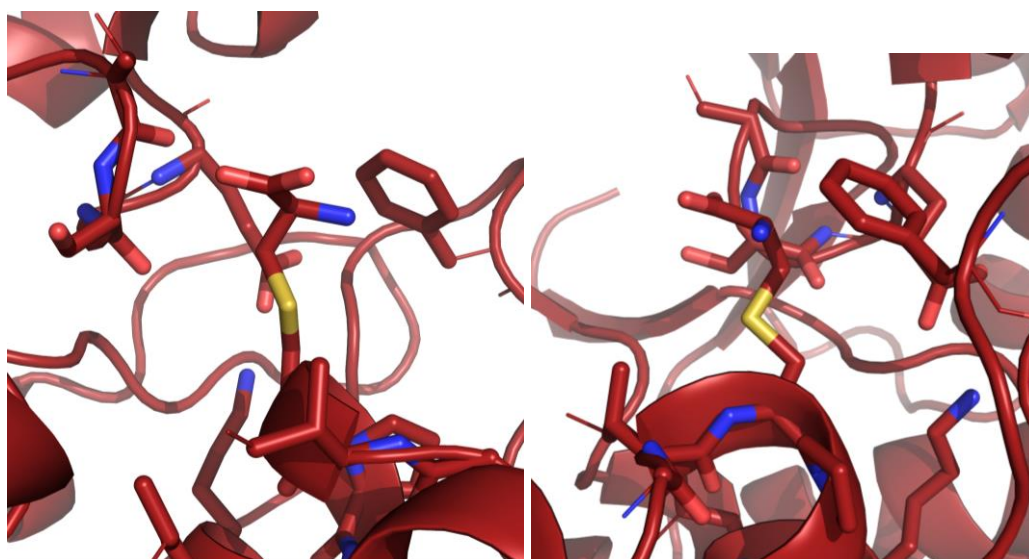


Figure 4.13 Y252Lni structure showing residues within 5 Å of the modified side chain Y252Lni. Atoms of Y252Lni and atoms within 5 Å of Y252Lni are shown as thick rods, atoms further than 5 Å from Y252Lni in a residue with one or more atoms within 5 Å of the modified side chain are shown as thin sticks, atoms in residues that have no atoms within 5 Å of Y252Lni are not shown. The structure on the left was rotated 90° to the left to produce the picture on the right. Non carbon atoms are coloured according to their identity: Dark blue, nitrogen; light red, oxygen; yellow, sulphur.

The side chain could be held in shape by intra-molecular hydrogen bonds between the carboxylate group, the amino group and the sulphur atom. The distances between these groups are 2.6 Å from carboxylate oxygen to side chain amine nitrogen, 2.9 Å from side chain amine nitrogen to the sulphur atom and 2.9 Å from the sulphur atom to the backbone nitrogen of the same amino acid (figure 4.14). These distances are within hydrogen bonding distance and the charges along the side chain, from the carboxylate end, alternate from negative to positive to slightly negative; carboxylate to amine to sulphur atom. The potential for hydrogen bonding and the complementary charges could both work together to keep the amino acid side chain in place, explaining the similarity observed between all the side chains.



Figure 4.14 Y252Lni side chain, with the distances measured as being potential hydrogen bonds shown as dashed yellow lines. The structure on the left was rotated 90° to the left to produce the picture on the right. Non carbon atoms are coloured according to their identity: Dark blue, nitrogen; light red, oxygen; yellow, sulphur.

The lanthionine side chain inserted into the active site had L-stereochemistry in the side chain, since L-cysteine was used in the modification. It was also built into the electron density with L-stereochemistry at the C α position, partly as previous structures of chemically modified enzymes have shown modified amino acids to adopt L-stereochemistry at the protein backbone following the Michael addition, which in theory could generate either D- or L-stereochemistries (figure 4.15).

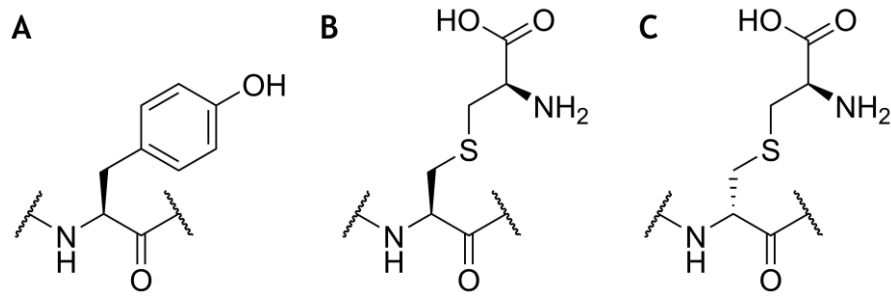
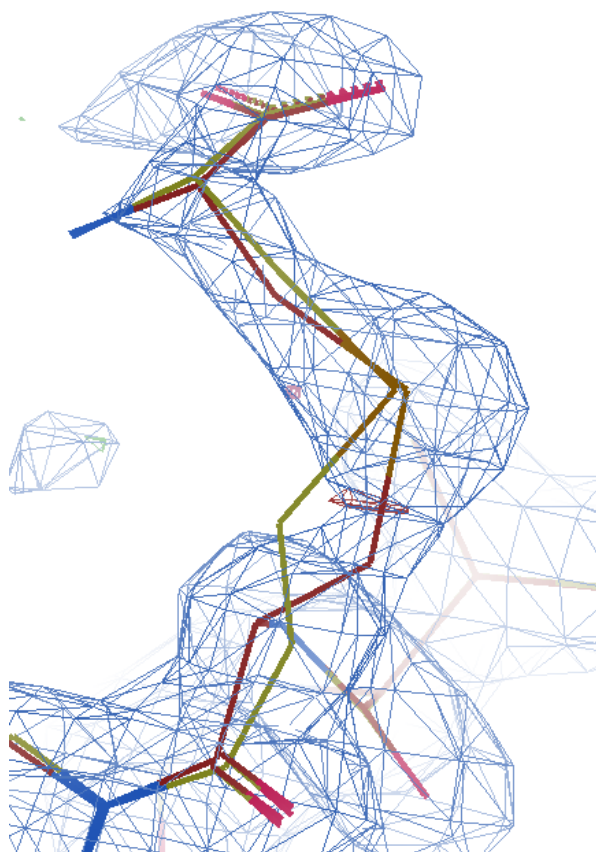


Figure 4.15 Diagrams of the amino acids at 252. **A** Tyrosine, the wild-type residue at 252. **B** and **C** D- and L-stereoisomers, respectively, of lanthionine, the ncAA inserted at 252 in this structure.

However, close examination of the modelled side chain compared to the density on some of the monomers showed that the modelled L-stereochemistry did not accurately describe the density, and that a D-stereochemistry at the protein backbone provided a more accurate description (figure 4.16). When all the monomers are examined in this way, in 6 of the 11 monomers in which the lanthionine is present, an L-stereochemistry at the protein backbone better described the density, while in 5 of the 11 monomers a D-stereochemistry at the protein backbone better described the density. The crystal densities are all averages of the atoms present in multiple copies of protein. If a mixture of different atomic structures are present in a single crystal, these will be averaged together. As such, some of the side-chains were more clearly one stereochemistry, showing that at this position in the crystal the side chains were predominantly of a single stereoisomer, while some were relatively ambiguous, showing that at this position the side chains were a more even mixture of stereoisomers.

There is no reason to think that the modification reaction should be stereospecific, and as such it would be expected that the L- and D-isomers form at equal proportions (Timms et al., 2013). The reason why D-amino acids were never observed in previous crystal structures was hypothesised to be because the D-isomer may be unable to refold, and would therefore be removed when size-exclusion chromatography is performed (Windle et al., 2017). However, given an unfolded protein and a suitably permissive position on the protein backbone, there is not necessarily any reason that the protein would not fold with a D-amino acid in its backbone.

A



B

	L-Lni	D-Lni
Bonds	1.228	0.979
Angles	2.116	1.235
Planes	0.691	0.243
Chirals	3.848	1.157
Non-bonded	0.026	0.047

Figure 4.16 L-Lni is not a good fit in the electron density maps of some monomers. **A** The modelled Y252Lni structure and the $2F_{\text{obs}}-F_{\text{cal}}$ electron density map (RMSD = 1.09 Å) of two different amino acids at position 252. The electron density is shown as a blue mesh, L-Lni is shown as green sticks and D-Lni is shown as red sticks. **B** Table showing the fit statistics of the amino acids shown in **A**, where larger numbers represent a worse fit.

4.3 Computational studies of the active site of wild-type and Y252Lni saNAL

A number of attempts were made to soak in both pyruvate and either glucuronolactone or a non-catalytic substrate analogue, with a hydroxyl group in place of the catalytic aldehyde group (figure 4.17). The same method was used as previously (section 2.4.16 and 2.4.17), but the final soak also included glucuronolactone or its non-catalytic analogue. However, a large number of crystals fragmented upon introduction to the final soak and those that remained to be frozen did not diffract coherently when inserted into the X-ray beam.

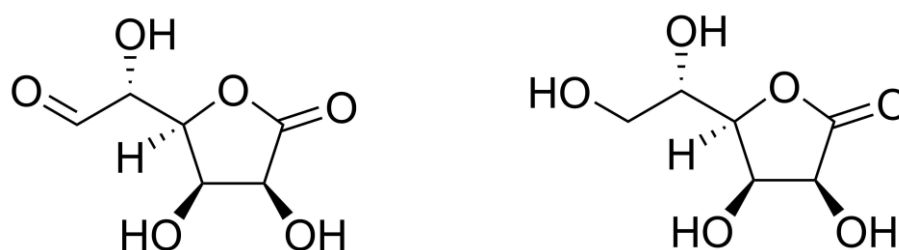


Figure 4.17 Chemical structure of glucuronolactone, the catalytic species (left), and the non-catalytic analogue used in crystal trials (right)

As the crystal structure of the pyruvate bound enzyme was of a high enough resolution (1.95 Å) to position the side chains in the active site with high confidence, molecular dynamic simulations of the protein interacting with the substrate were considered. While these simulations would only present theoretical, *in silico* predictions of how the substrate may interact, there are a number of advantages from dynamics simulations over crystal structures. Firstly, crystal structures represent an average position over a large number of protein molecules, which might cause less popular binding positions to be missed. Also, X-ray structures display a single, fixed position in which the substrate is bound, and this will not be representative of the action of the enzyme in solution. When done appropriately, molecular dynamics allow a number of potential binding positions to be examined, and also allow movements of the substrate in the active site to be examined. As such, molecular dynamics was used as a tool to elucidate the actions of the substrate in the active site.

4.3.1 Preliminary investigation

As an initial test of the force fields and simulation settings a structure with substrate bound in a catalytic conformation was simulated. Fortunately, a structure of mutant *E. coli* NAL with both pyruvate and ManNAc bound in

the active site has previously been published (PDB ID: 4bwl; Daniels et al., 2014). Molecular dynamics were used on this structure first to ensure the molecular system was represented in a realistic way. The ManNAc and all residues with at least one atom within 4.5 Å of the ManNAc were simulated for 600 ps as described in section 2.5.5. The ManNAc remained stably bound in the active site for the duration of the simulation. This behaviour is expected as enzymes should have at least some affinity for their substrates, even if they have a higher affinity for the transition state. This suggests that the simulation parameters were appropriate for the molecular system over the timescale simulated.

To ensure that these settings were also appropriate for *sa*NAL, ManNAc was inserted *in silico* into the active site of *sa*NAL with pyruvate bound (PDB ID: 4ah7; Timms et al., 2013), in a position similar to that of the *ec*NAL crystal structure. Molecular dynamics were then performed on the *sa*NAL structure with ManNAc in the active site using the same method as the *ec*NAL. Again, the ManNAc remained bound in the active site, showing that the dynamics parameters are appropriate for the wild-type *sa*NAL protein as well.

Running simulations on the active sites of all four chains in the tetrameric wild-type structure as well as all 12 chains of the 3 tetramers in the unit cell of the Y252Lni structure would greatly increase the amount of simulations to be run and the amount of data to be analysed, without necessarily increasing the quality of the data. Ideally, one subunit of each of wild-type, Y252D-Lni and Y252L-Lni would be selected to be used in all simulation experiments. Therefore, to assess the extent of similarity of each chain of the crystal structure, all subunits of the Y252Lni structure and all subunits of the *sa*NAL wild-type structure were overlaid, and the orientation of the side chains at position 252 was examined (figure 4.18 A). The tyrosine side chains in the wild-type structure are well ordered (figure 4.18 B), as are the L-Lni side chains (figure 4.18 C). The D-Lni side chains are less well ordered however (figure 4.18 D), which could be anticipated as a D-amino acid in the protein backbone is a more unnatural situation than a non-canonical L-amino acid in the backbone. As such, the tetramer composed of chains A-D of the Y252Lni structure was chosen for use in molecular dynamics, this contained one chain with D-Lni, two chains with L-Lni and one chain that did not display any side chain density at the 252 position. In further simulations, chain B of the wild-type structure was chosen to represent the wild-type active site, chain B was chosen from the Y252Lni structure to represent Y252L-Lni-containing active sites, and chain

D was chosen to represent D-Lni-containing active sites (shown in blue, red and green respectively in figure 4.18).

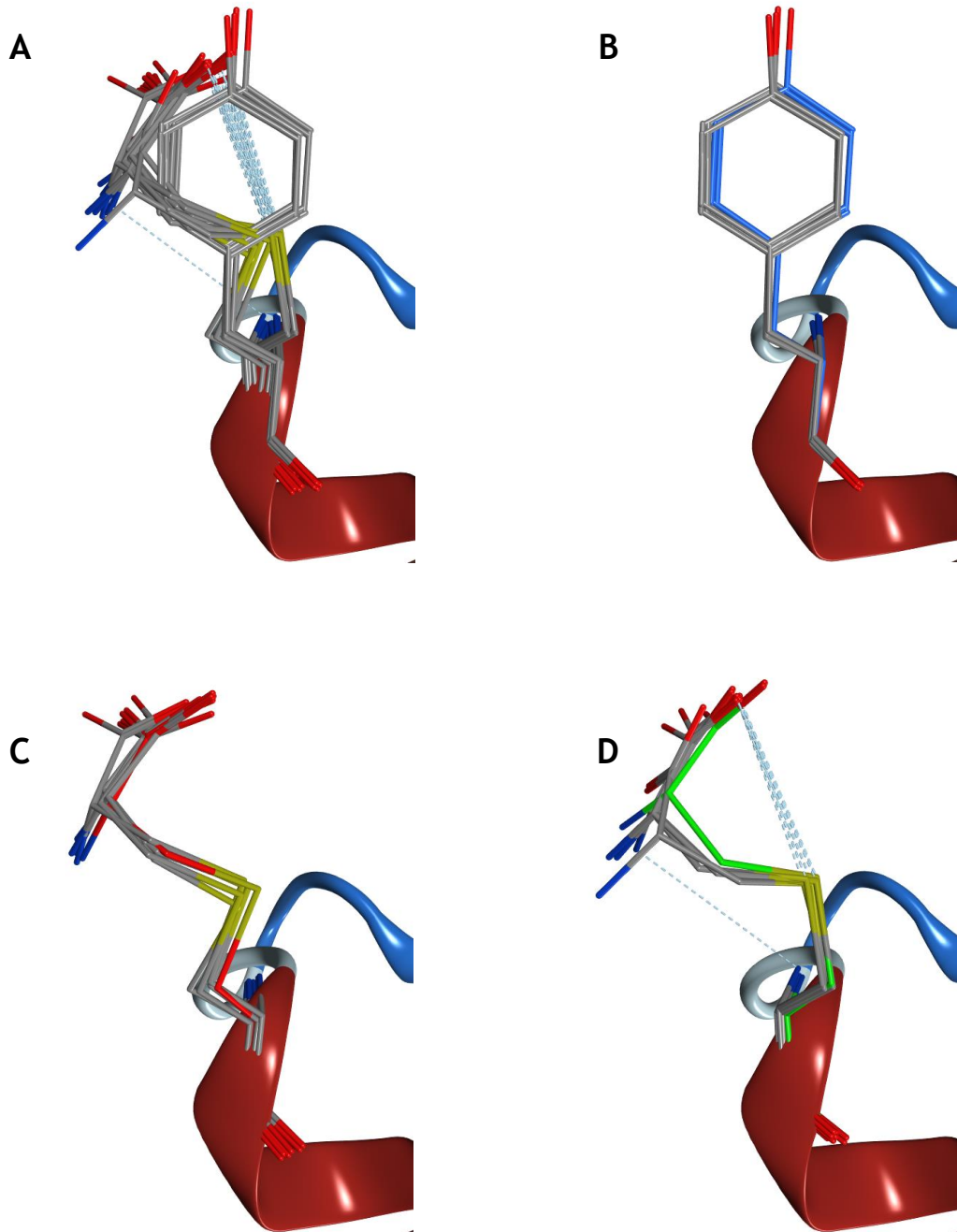


Figure 4.18 Overlaid subunits of wild-type and Y252Lni *saNAL*. **A** All wild-type, Y252L-Lni and Y252D-Lni subunits; **B** all four wild-type subunits with subunit B shown in blue; **C** Y252L-Lni with subunit B shown in red and **D** Y252 D-Lni with subunit D shown in green. A much greater amount of variability is seen in the chains containing D-Lni. In all images dotted blue lines connect atoms likely to form hydrogen bonds

Three potential causes for enhanced catalytic activity were assessed:

- the lanthionine-containing enzyme is better at causing the substrate to bind than the wild-type enzyme (section 4.3.2).
- the lanthionine-containing enzyme is better at holding onto the substrate in a catalytic orientation than the wild-type (section 4.3.3 and 4.3.4).
- the lanthionine-containing enzyme is better at removing the full length product than the wild-type (section 4.3.5)

4.3.2 Substrate binding from bulk solvent

A simple method to estimate binding is using the GRID method (Goodford, 1985). This calculates a likely energy of binding of a functional group to all points on the surface of a molecule, in this case a protein. Where the binding energy is greater than a certain threshold value, a green layer is added to the surface of the protein. As such, the regions on a protein where certain functional groups are likely to bind can easily be visualised.

A total of three hydroxyl groups occur on glucuronolactone making a hydroxyl group a good probe to assess binding potential in the active site. GRID was therefore performed on wild-type, Y252L-Lni and Y252D-Lni using an aliphatic hydroxyl group as the functional group probe to determine the binding surface.

The green region represents where hydroxyl groups are likely to bind (figure 4.19). A large green region can be seen around the lanthionine side chain's carboxylate, whilst this green region is absent in the wild-type structure. This suggests that there is more surface area available for binding in the Lni-containing active site than in the wild-type active site.

One of the best ways to assess binding of substrate from bulk solvent is to place a substrate molecule outside of the active site and see which enzyme binds the molecule in a catalytic pose most often or for the longest period of time. However, simulating binding to an active site from bulk solvent is difficult over feasible simulation timescales as the substrate is more likely to move away from the enzyme than towards it when allowed to float free, and if it does bind to the enzyme it is unlikely to find its way to the active site, binding onto hydrophilic surface residues elsewhere. This was attempted but, even with a tether holding the substrate near the protein so it could not diffuse away, substrate attached non-specifically to the exterior of the protein (data not shown).

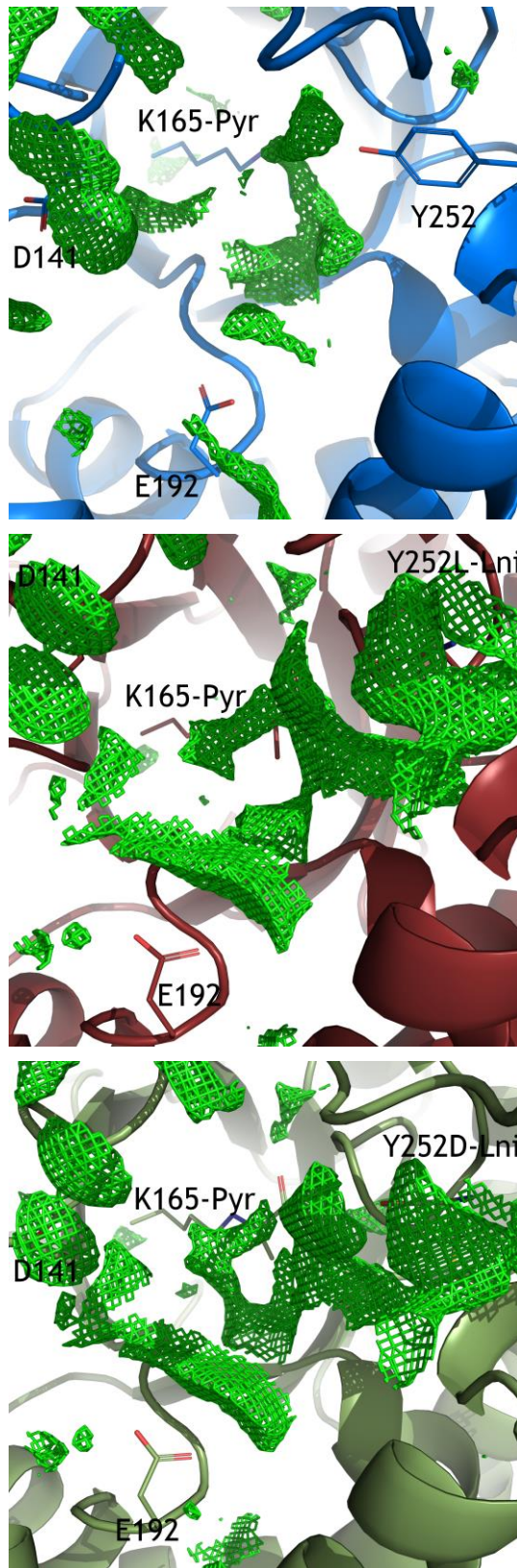


Figure 4.19 GRID calculations using a hydroxyl group as the probe on wild-type, Y252L-Lni and Y252D-Lni enzymes. GRID surface indicating likely binding location is shown as a green mesh **Top** wild-type, **Middle** Y252L-Lni, **Bottom** Y252D-Lni. Amino acids are shown as sticks with atoms coloured by identity, amino acid backbones are shown as a cartoon and coloured according to enzyme, wild-type in blue, Y252L-Lni in red and Y252D-Lni in green. GRID surface shown at an interaction energy of - 3.6 kcal/mol. D141, K165, E192 and the residue at 252 are shown and labelled.

4.3.3 Substrate binding in the active site

On glucuronolactone, all the hydroxyl groups point in the same direction after energy minimisation, giving the molecule significantly different hydrogen bonding potential depending on which side of the substrate is

considered (figure 4.20). Initially, two molecular dynamics simulations were run for each enzyme; wild-type, Y252L-Lni and Y252 D-Lni, one with the hydroxyl groups pointed towards the amino acid at position 252, and one with the hydroxyl groups pointing away. The substrate was inserted in the active site with the aldehyde group pointing towards the key catalytic residues, K165 and Y137, to encourage the simulation to produce potentially catalytic conformations, and a tether was made between the pyruvate methyl group and the aldehyde carbon to restrict the two groups to between 3.5 and 4.5 Å of each other to ensure that the substrate remained in the active site and in an area that was catalytically interesting.

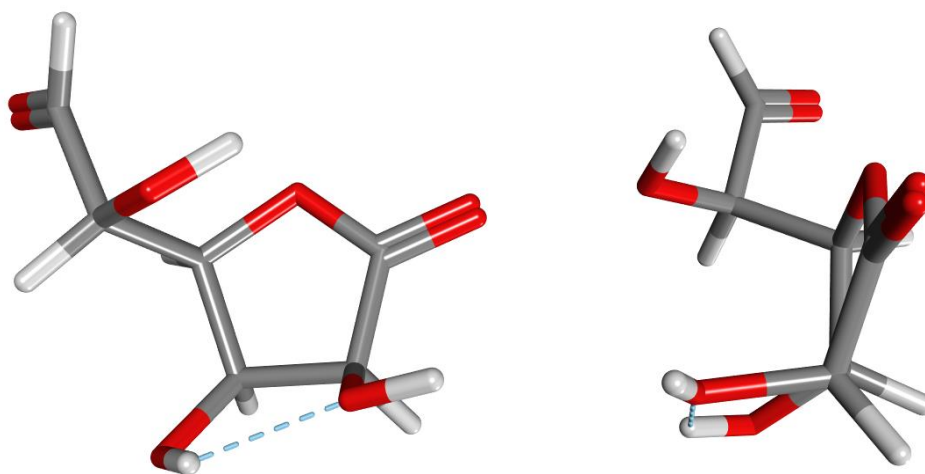


Figure 4.20 Glucuronolactone. Hydroxyl groups are all positioned on the same face of the lactone ring when energy minimised (**Left**) shown face on, and (**Right**) side on to highlight that the hydroxyl groups all point towards the same face of the molecule when energy minimised. Atoms are coloured by identity.

In the wild-type structure, the substrate moved so that the ring hydroxyls were taking part in a hydrogen bonding network with E192 within 100 ps in both simulations (figure V.1, V.2 and 4.21), suggesting that E192 likely plays a key role in glucuronolactone binding. The modified enzyme still has this residue, and so we could speculate that E192 probably plays some role in catalysis in the Y252Lni enzyme also. Additionally, Y252 can be seen to play very little role in the binding of glucuronolactone in the wild-type enzyme.

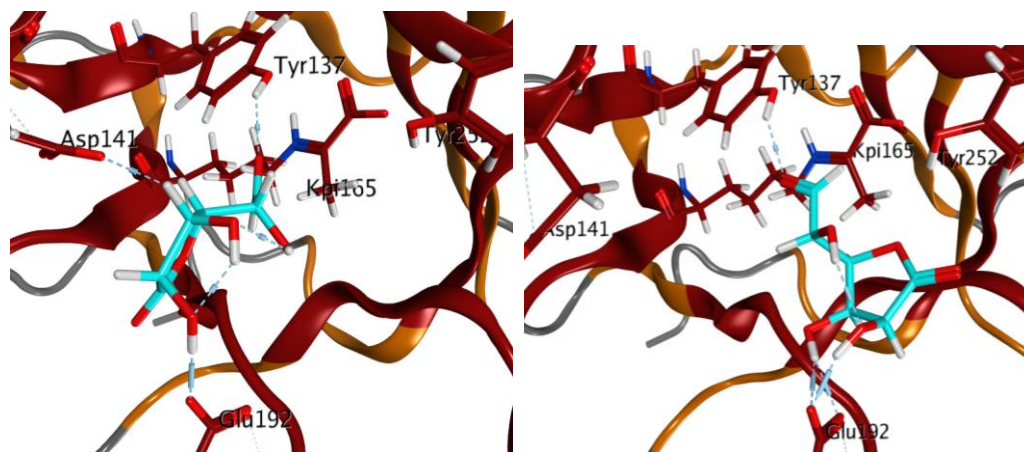


Figure 4.21 Final orientations of simulations using the wild-type structure and glucuronolactone. **Left** the substrate ring hydroxyls starting pointing towards the bottom of the frame, and **Right** towards the top of the frame. Both simulations resulted in the substrate interacting with the carboxylate group on E192. Turquoise atoms show the substrate, Red atoms are part of the protein, Red ribbons, atoms and turquoise atoms were all active and mobile in the simulations, orange ribbons were active but not mobile, and grey ribbons were neither active nor mobile. Y137, D141, K165, E192 and Y252 are shown and labelled. See also Video figures V.1 and V.2. **Figure V.1** shows a simulation that was started with the hydroxyl groups of the substrate pointed away from the tyrosine at residue 252, while **figure V.2** shows a simulation that was started with the hydroxyl groups pointed towards Y252. Both simulations resulted in substrate binding to E192.

When the results from the modified enzymes are examined, each orientation of substrate displays different behaviour. When the hydroxyl groups are pointed towards Y252Lni, they form a hydrogen bonding network with the nAA and remain bound (figure 4.22), but when the ring hydroxyl groups are pointed away from the nAA, the substrate forms no significant bonding with the active site at all (figure 4.22). This shows that, for the timescales being simulated, initial orientation of the

substrate in the active site is very important, and so the effects of substrate orientation will be examined in more depth.

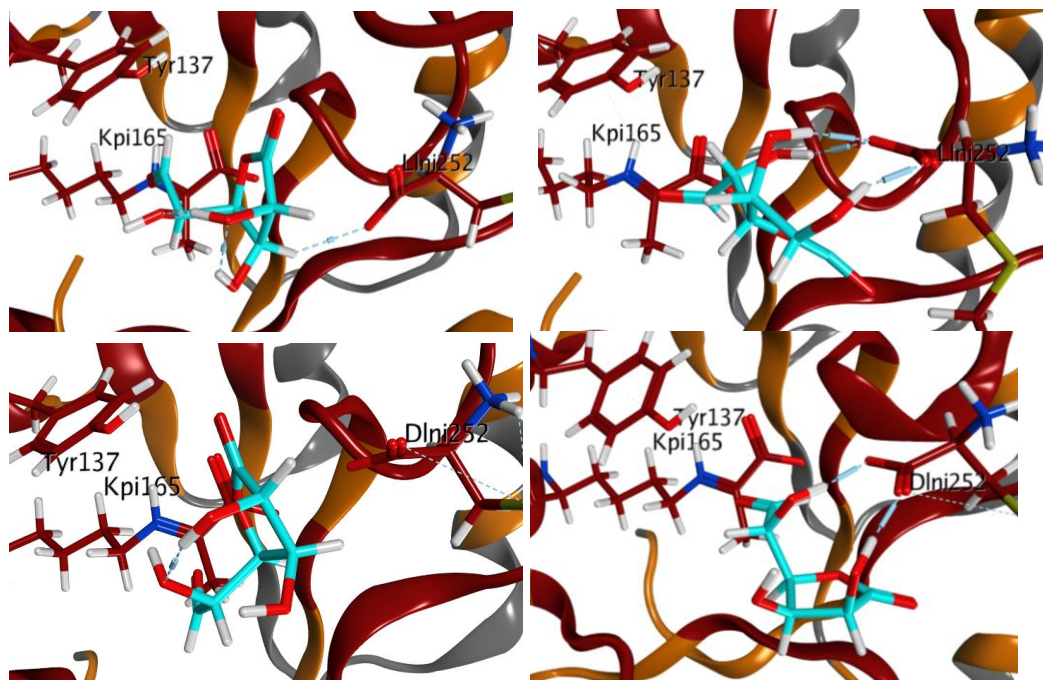


Figure 4.22 Representative poses during the preliminary simulations of Y252Lni with glucuronolactone. **Top** Simulations using Y252L-Lni, **Bottom** using Y252D-Lni. **Left** simulations started with hydroxyls pointed away from Lni and **Right** with hydroxyls pointed towards Lni. Turquoise atoms show the substrate, Red atoms are part of the protein, Red ribbons, atoms and turquoise atoms were all active and mobile in the simulations, orange ribbons were active but not mobile, and grey ribbons were neither active nor mobile. Y137, K165 and Y252Lni are shown and labelled.

As such, the structures of each of the protein had the substrate inserted in the active site, next to the residue at position 252 with the hydroxyl groups facing away from the side chain of the residue at 252. A tether was not applied in these experiments so that the substrate could sample a wider set of binding poses if it was energetically favourable for it to do so. The substrate was then rotated around an axis from the aldehyde group to the lactone ring to maintain the proximity of the aldehyde to the pyruvate. This encourages catalytically interesting binding of the substrate, as the binding positions that have an effect on catalysis are those in which the aldehyde group is close to the pyruvate. The substrate was rotated by 60° 6 times to give a full 360° set of substrate starting positions (figure 4.23 Left). This allows a larger sampling of possible binding events than if one simulation was run for six times longer. Data shown in the figure is for

Y252D-Lni, but a similar outcome occurred with both wild-type and Y252L-Lni.



Figure 4.23 Positions of the substrate in the Y252D-Lni rotation simulations before (**left**) and after (**right**) energy minimisation. Only the C6 carbonyl group and C5 carbon are shown for five of the six substrates before minimisation to decrease the complexity of the figure. The C6 carbonyl, C5 and C5 hydroxyl are shown after minimisation to give an idea of which direction the ring is facing in each position. Turquoise, orange, green, blue, yellow and purple atoms show the substrate each from a simulation started with the substrate at a different angle. Red atoms are part of the protein, Red ribbons, atoms and all substrate atoms were all active and mobile in the simulations, orange ribbons were active but not mobile, and grey ribbons were neither active nor mobile. K165 and Y252D-Lni are shown and labelled.

After energy minimisation, the substrates were still in three or four distinct positions (figure 4.23 Right), giving a good indication that there is not one lowest energy state for the substrate in the active site. If the active site was highly selective for a specific substrate, we could expect that there would only be one energy minimum in the active site, and so all six starting positions should converge on one position after energy minimisation. That there are many energy minima displayed by this substrate is unsurprising as this enzyme has not naturally evolved to accept it. Each of these simulations was then run for 600 ps according to the method given in section 2.5.6.

It was observed that during the simulations with Y252Lni the substrate remained next to the nCAA, even if it made few major contacts. This was most likely because the active site is fairly open in the modified enzyme (figure 4.11) meaning there are no residues to interact with on the face away from 252. Consequently, there will be no attracting force, and the

timescale simulated was not long enough for significant stochastic movement away from the weak interacting forces that were made with the enzyme. As such, to sample the substrate's binding potential across the whole of the active site, a second set of rotations were performed with the substrate on the side of the modified enzyme's active site, away from the residue at 252, energy minimised, and simulated as previously. The wild-type enzyme has an active site that is more narrow (figure 4.11), and so did not require a second set of simulations.

From visual examination of the behaviour of the substrate during the resulting dynamics runs, a number of binding groups, where substrates in multiple different runs behaved similarly, could be identified and these are described in Figure 4.24.

Binding group	Defining feature
Group I	Ring hydroxyls bound to the opposite side of the active site from the residue at position 252
I.1	Ring hydroxyls strongly bound to D141
I.2	Ring hydroxyls strongly bound to E192
I.3	Substrate bound to both D141 and E192
Group II	Ring hydroxyls bound to the modified residue
II.1	Ring hydroxyls bound to Y252Lni with an aldehyde - pyruvate distance of less than 5 Å
II.2	Ring hydroxyls bound to Y252Lni with an aldehyde - pyruvate distance of more than 5 Å
Group III	Substrate interacts with both the acidic residues from group I and the residue at position 252
Group IV	No significant or long lasting interaction

Figure 4.24 The different groups (in bold) and subgroups of binding that were observed, and how these groups and subgroups were defined.

In an effort to categorise the simulations in a more systematic manner, a number of key distances were extracted from the simulations at 0.5 ps intervals for each simulation run. Two sets of distances were recorded. Firstly, from the aldehyde carbon to the methyl carbon of the pyruvate, and from the aldehyde carbon to the oxygen of the hydroxyl on Y137, to

give a proxy of catalytic ability, i.e. the smaller both these measurements are the more likely it is that a substrate bound in that position would be able to react with the pyruvate to give a product (Figure 4.25 A). The second set of measurements were all from the oxygen of the C5 hydroxyl to; the carbon of the carboxylate of D141; the carbon of the carboxylate of E192; and to either the carbon bonded to the hydroxyl in Y252, or the carbon of the carboxylate group in lanthionine, depending on which enzyme was being measured. The second set of measurements enable triangulation of the substrate in the active site (Figure 4.25 B).

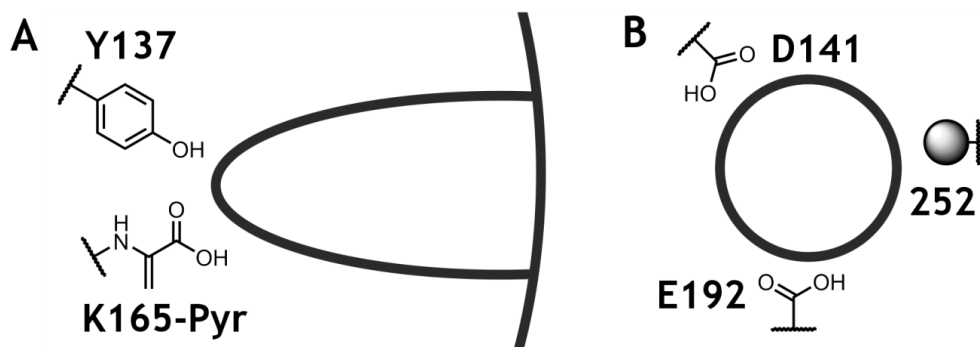


Figure 4.25 Schematic diagram showing the purpose of the distance measurements taken from the substrate binding simulations. **A** Side schematic of the active site showing how measuring from Y137 and K165 to the substrate should give an approximate measure of how far into the active site the substrate is. **B** Schematic view down into the active site showing how measuring the distance of the substrate from D141, E192 and the amino acid at 252 allows triangulation of the substrate in the active site. Amino acids where known are shown as their terminal side chain structures, the enzyme is represented by black lines to show the pocket of the active site.

When the 10 point running average data for each of these distance parameters is examined for each run, significant shifts in some of the measured distances can be seen over the lifetime of the simulation. Some of these are an apparently random oscillation of the substrate in the active site (figure 4.26 A), others are fast switches of binding conformation (figure 4.26 B). The running average data for each distance over each molecular dynamics run was examined for any pose that was stable for 30 ps or longer. As an example, the run shown in figure 4.26 A was not included as it never achieved a stable conformation, while the run shown in figure 4.26 B was split into 3 separate runs, as explained in the figure legend.

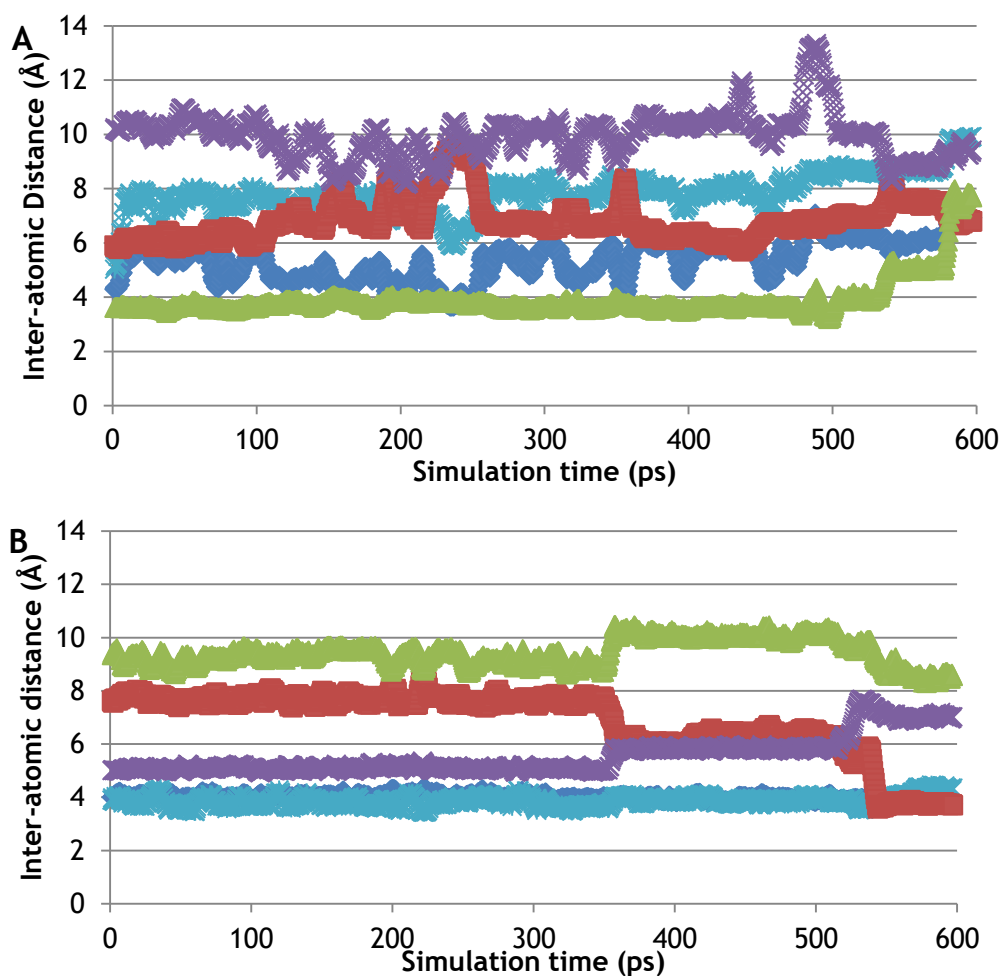


Figure 4.26 Ten point running average data for each of the five distances measured. A substrate that is not stably bound is shown, (A, L-Lni_2_180deg) and below it a substrate displaying significant shifts in conformation can be seen (B, D-Lni_2_300deg). Conformation shifts are seen as sharp changes in a number of measurements, e.g., at approximately 360 ps D-Lni_2_300deg switches from group IV to group II.1 and switches again around 530 ps to group III. Substrate aldehyde to pyruvate methyl distance in dark blue, aldehyde to Y137 hydroxyl oxygen in light blue, C5 hydroxyl oxygen to D141 carboxylate carbon in green, C5 hydroxyl oxygen to E192 carboxylate carbon in red and C5 hydroxyl oxygen to Y252Lni carboxylate carbon in purple.

Each pose was examined using a representative time point and categorised according to the groups given in figure 4.24. The raw distance data for each of the conformations stable for >30 ps was averaged, and the average values were examined using a clustering algorithm (section 2.5.10) and a dendrogram was produced to visualise the similarities (figure 4.27).

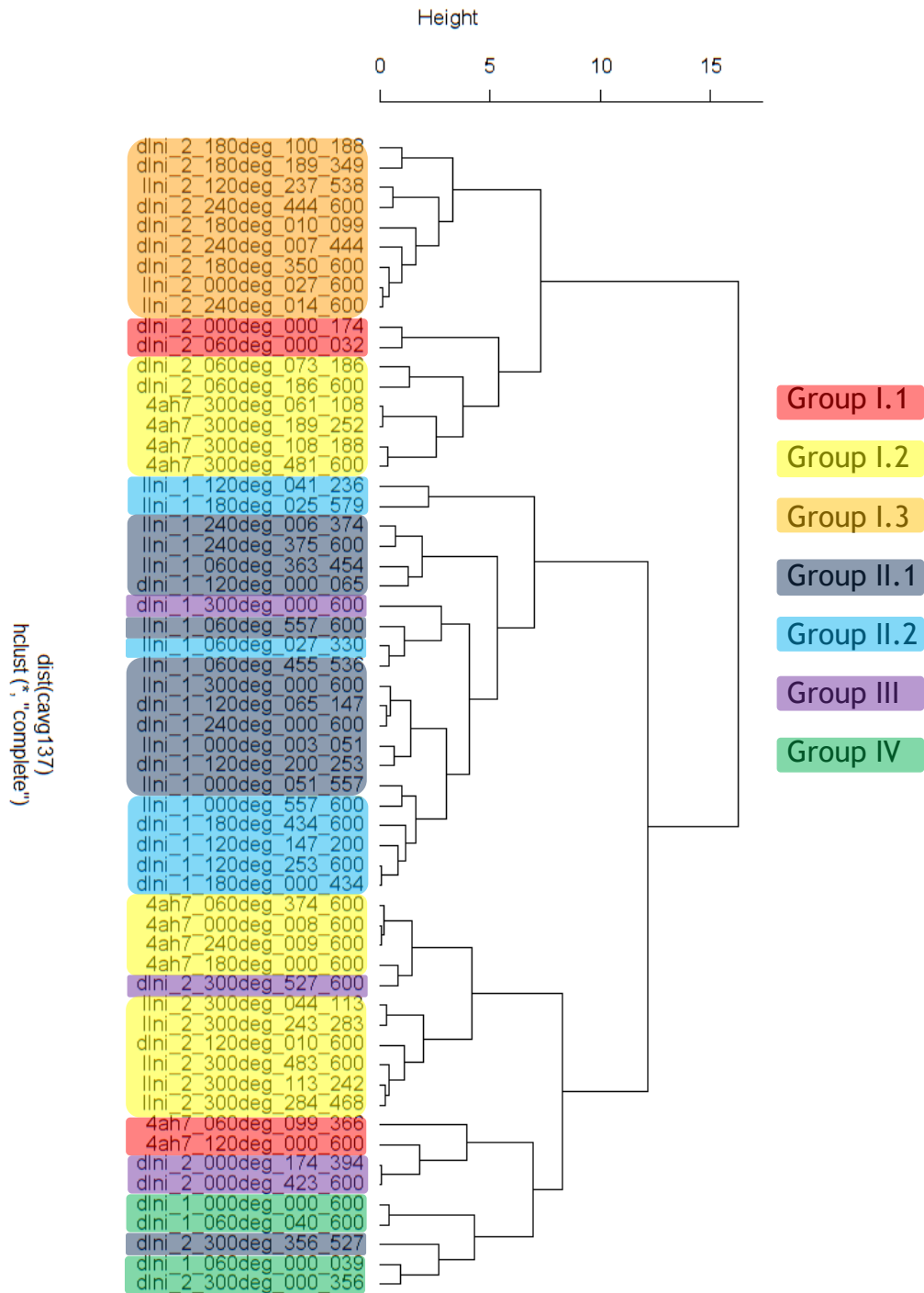


Figure 4.27 Dendrogram showing the relationship between different poses identified from examination of the running average distance data. The lower the Height where two groups split, the more closely related they are. Poses are labelled as “enzyme_(experiment number)_angle of rotation_start of pose (ps)_end of pose (ps)” The enzymes used are wild-type; “4ah7”, Y252L-Lni; “llni”, and Y252D-Lni; “dlni.” The experiment numbers in the modified enzyme data correspond to; 1, substrate placed next to 252; and 2, substrate placed away from 252. Poses are coloured according to group (shown in key).

Grouping in the dendrogram would show that a set of poses are similar, and the dendrogram shows the groups assigned using visual identification are mostly accurate, but also reveals greater granularity than found during visual inspection. It shows a significant separation between the group I forms with large enzyme-aldehyde distances at the top of the dendrogram, and the group I forms with small enzyme-aldehyde distances, which are more similar to group II binding forms and grouped with them towards the bottom of the dendrogram.

This increased the confidence in the groups, but the distance data gathered allowed more quantitative analysis to be performed. The number of picoseconds each pose was sustained for during analysis was recorded. As such, the proportion of time that the substrate spent in each group could be calculated and analysed by enzyme (figure 4.28 and figure 4.29).

In the wild-type enzyme the group that persisted for the longest amount of simulation time was group I.2 and it was present for 64% of the time of simulation. The wild-type enzyme only displayed one other pose, group I.1, which was present for 24% of the simulation time. In all rows of the table, percentages do not add up to 100% as there was some time in almost all simulations where the substrate was not stably bound.

In the Lni-containing structures there were a much greater range of groups of binding observed. For L-Lni, the group that the substrate was in for the longest amount of time was II.2 with 22% of simulation time, and for D-Lni it was group IV, also with 22% of simulation time. Substrate was only observed binding stably to the residue at 252 when that residue was lanthionine, stable binding to the tyrosine residue at 252 in the wild-type enzyme was never observed. Subsequently, groups II and III are only observed in the modified enzymes. Additionally, group III and IV are only observed in Y252D-Lni enzymes. The reasons for this are unclear, as the starting position of the side chain is very similar in both D- and L-Lni containing enzymes. For group III binding, the presence of this group in D-Lni only may be due to the fact that the D-isomer at the backbone points the side chain towards the active site more, which allows it to come further into the active site and form hydrogen bonds with the substrate, even when the substrate is also hydrogen bonded to the other side of the active site. Overall, there was a much wider range of groups present in the D-Lni simulations than either the L-Lni or the wild-type enzyme, potentially due to lower stability of the D-Lni enzyme, which could be expected of an enzyme containing a single D-amino acid in the peptide backbone.

Group	I.1	I.2	I.3	II.1	II.2	III	IV	Total
Wild-type (ps)	867	2318	0	0	0	0	0	3185
Wild-type (%)	24.1	64.4	0.0	0.0	0.0	0.0	0.0	88.5
Y252L-Lni (ps)	0	539	1460	1456	1598	0	0	5053
Y252L-Lni (%)	0.0	7.5	20.3	20.2	22.2	0.0	0.0	70.2
Y252D-Lni (ps)	206	1117	1180	971	1000	1070	1555	7099
Y252D-Lni (%)	2.9	15.5	16.4	13.5	13.9	14.9	21.6	98.60

Figure 4.28 Table showing the amount of simulation time spent in each group by each enzyme, and what percentage of total possible simulation time for that enzyme that amount of time corresponds to. Note that percentages do not add up to 100 as there was some time during most simulations when the substrate was not stably bound.

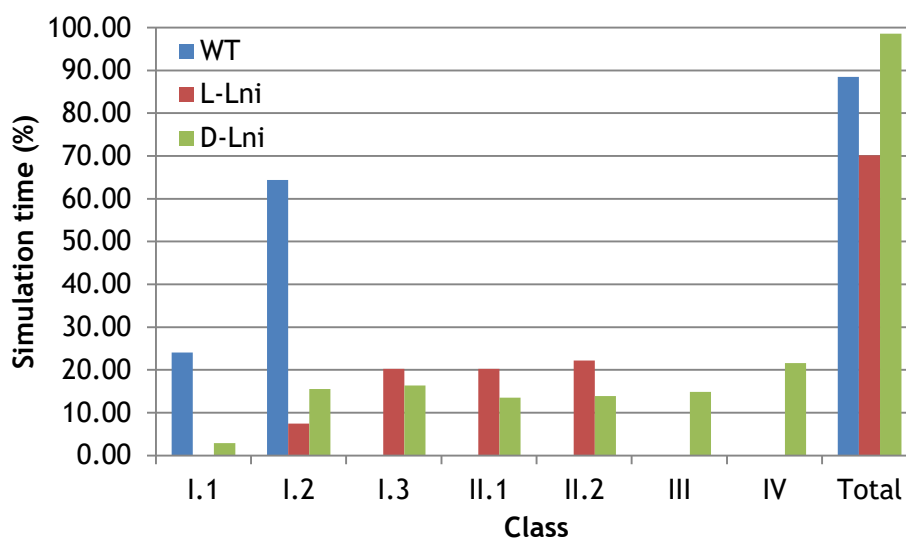


Figure 4.29 Graphed percentage data from figure 4.28. Percentage of time that each enzyme had substrate bound in each group is shown, along with total simulation time that each enzyme had substrate stably bound in any group. Wild-type enzyme is shown in blue, L-Lni in red and D-Lni in green.

While this data suggests that Y252D-Lni may be slightly more active than the wild-type, for approximately 20% of simulation time for Y252D-Lni, the substrate was bound in Group IV, i.e. displaying no significant and stable bonds with the active site. This binding group would likely be transient *in vitro*, and the presence of substrate in this group classified as stably bound likely represents an artefact in the simulation process. This artefact most

likely exists because of the significant difference between the timescale being simulated and the timescale over which the reaction takes place. The time period of each simulation was 600 ps, and the time that the enzymes were observed for in total was 3.6 ns for the wild-type and 7.2 ns for each enzyme containing lanthionine. When this is compared to the turnover of the wild-type enzyme of 20 s, it is clear that the difference between these timescales is vast, approximately 10^{11} -fold. Even if binding of substrate, which is what we are simulating, is 1,000-times faster than the reaction, there is still a vast difference between simulation time and binding time. Therefore, the simulations give us an idea of what poses may be favoured by the enzymes, but may not be an accurate measure of the proportion of time that the substrate in each enzyme stays in each pose.

4.3.4 Substrate retention in the active site

This distinction was investigated by measuring relative strength of each binding pose, as stronger poses should last for longer and be less susceptible to disrupting influences such as bulk solvent, buffer components and trace contaminants. A representative structure of each subgroup was taken, and the strength of each binding pose was assessed by performing steered dynamics, running a number of different simulations during which a force is applied to the substrate in a direction pulling it directly out of the active site.

The force is applied by means of a tether between the substrate and a carbon atom that was inserted outside of the active site. The point to which the tether was attached on the substrate was the heavy atom in the substrate that was closest to the entrance to the active site and also not involved in any hydrogen bonding. The carbon atom was inserted in a position such that a straight-line force acting to pull the substrate towards the carbon atom would pull the substrate out of the active site and would not cause the substrate to be pulled into any other atoms. The algorithm used to determine the exact force that is applied by the tether varies the force by distance and as such all the simulations were started with tethers of identical length. However, due to random oscillations in the substrate during the simulation, the direct force applied to the molecule will change throughout the lifetime of the simulation. The algorithm can be modified to either increase or decrease the force that is applied by use of a weight constant.

Each simulation used the force equations found in section 2.5.7 and had a weight constant applied of between 0 and 0.3. The weight constant

required to pull the substrate out of its binding pose was recorded for each group along with the amount of time this took. The substrate was defined as out of its binding pose when the only hydrogen bonds remaining were those between the ring hydroxyls and the carboxylate group, or when the substrate had moved so that it was binding as another group.

Figure 4.30 shows the maximum weight constant applied to the tether that did not disrupt the binding of the substrate, which functions as an approximate measure of relative overall binding force. Surprisingly, the subgroup that required the least amount of force to disrupt was I.1, while group IV, with no significant interactions in the active site, came thoroughly in the middle of the poses. The subgroup that required the strongest pulling force to extract was III, which was only observed in the modified enzyme.

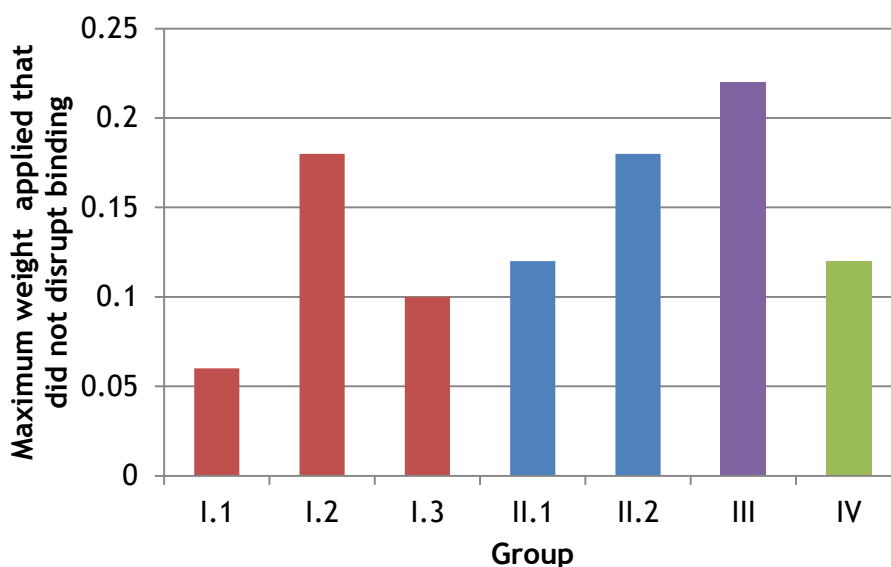


Figure 4.30 Maximum weight constant applied that did not disrupt the binding of each group. Binding groups involving wild-type residues, group I generally required a lower force to disrupt than those involving the modified residue, groups II and III.

The substrate displayed different behaviours when placed in each of the three enzymes. Each enzyme can be assigned two sets of binding groups; those in which the substrate was observed in any configuration; and those in which the substrate was observed with the aldehyde group in close proximity to the key catalytic residues. When the averages of these two classifications of binding groups are assessed for each enzyme, it can be seen that the both the L- and the D-stereoisomers of the modified amino acid display a slightly higher average force to disrupt (figure 4.31).

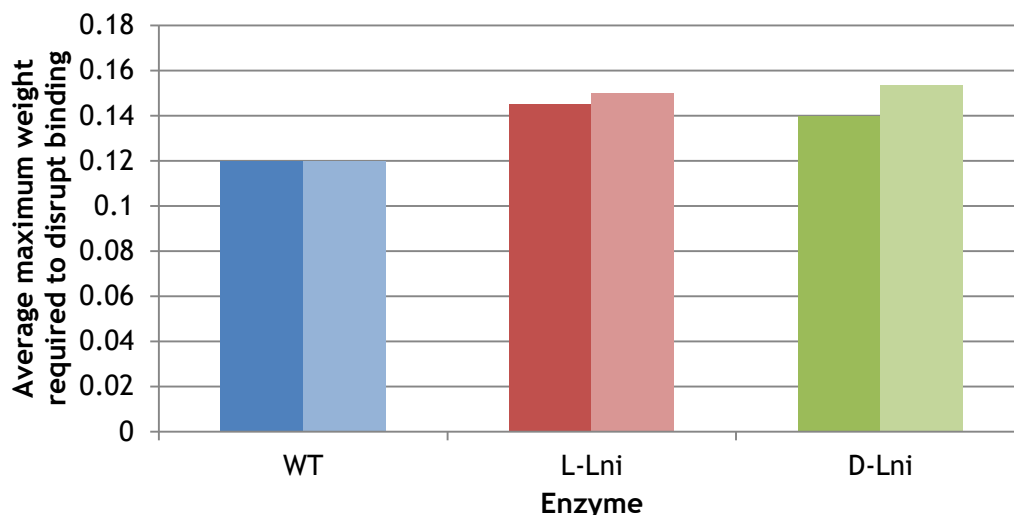


Figure 4.31 The average maximum weight required to disrupt binding for; those groups in which the substrate was observed in any configuration for each enzyme (**bold colours**); and those in which the substrate was observed with the aldehyde group in close proximity to the key catalytic residues for each enzyme (**faded colours**).

Taken together this data suggests that when the substrate binds in the groups seen in the wild-type enzyme it is held slightly less tightly and may be more easily disrupted than when it is bound in the groups seen in either of the modified enzymes.

4.3.5 Removing the full length product

Another explanation for increased activity could be due to an increased rate of product removal from the active site. If the product moves out of the active site faster in the modified enzyme than in the wild type, this would open the active site up for substrate binding more rapidly. To assess the ability of each enzyme to retain product, the pyruvate was removed from each structure to leave the unmodified lysine residue. Each structure was energy minimised and 4 products; (*R*) and (*S*) stereoisomers at the aldol hydroxyl carbon, each with the ring in two orientations; were inserted into the active site. Each of these situations was simulated for 600 ps (see section 2.5.8), and the product remained bound for the lifetime of the simulation in every combination of enzyme and product. This is mostly due to the strong network of hydrogen bonds holding the terminal acid group in place (figure 4.32).

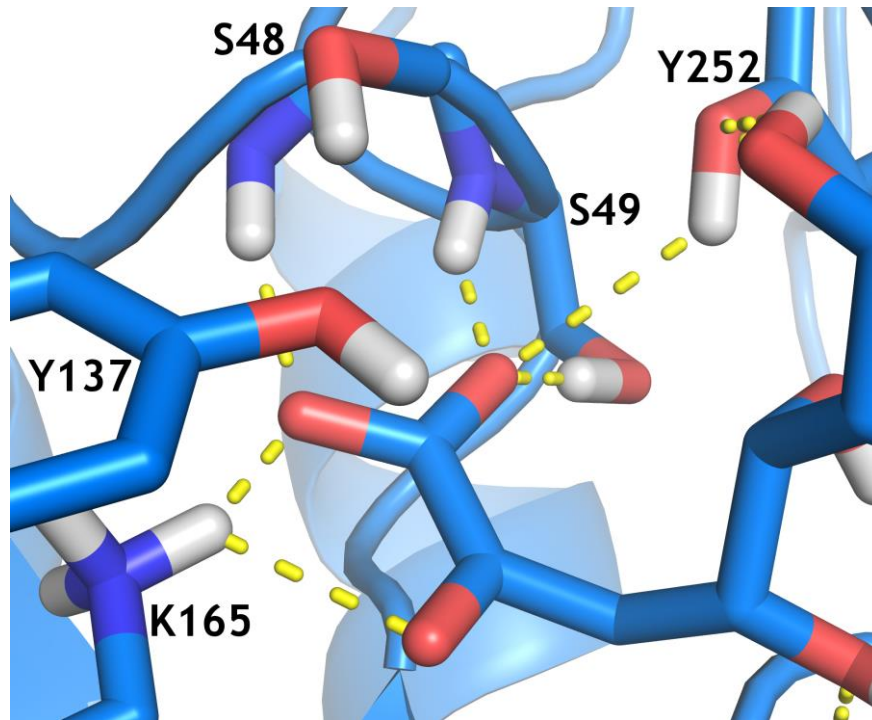


Figure 4.32 Example of the hydrogen bonding network attaching the full length product to the active site. This image depicts binding in the wild-type enzyme but the hydrogen bonding network was similar in all combinations of enzyme and product. Most important are the five hydrogen bonds (yellow dashed lines) connecting the carboxylate group on the product and S48, S49 and K165. The substrate and residues S48, S49, Y137, K165 and Y252 are shown and the residues are labelled.

The product was pulled out of the active site in the same way as during assessment of the strength of binding poses (section 2.5.9), but the results were less than conclusive. During the assessment of strength of binding poses, there was a clear time at which substrates were out of pose, i.e. when only the ring hydroxyls remained bound, or when the substrate changed group. However, when the full length product is pulled out of the active site, it is harder to find a clear point at which the product can be considered removed from the site. This is because the hydrogen bonds between the terminal carboxylate group and the enzyme tend to break individually and there no clear transition from in position to out of position. This would increase the subjectivity of the data gathered, reducing its value to an extent that renders it not informative.

4.4 Summary

The saturation mutagenesis shows that position 252 is important in determining the activity of *sa*NAL for the reaction between glucuronolactone and pyruvate. A number of the canonical amino acid mutants displayed high activity, some even showing activity as high as the ncAA-containing enzyme. There was a rough trend in the data for the canonical amino acids, in that the variants containing larger amino acids at position 252 generally showed a lower activity for glucuronolactone than the variants with smaller amino acids at 252. Y252Lni seemed to defy this trend by having both a high activity and a high mass side chain at 252. This anomalous activity warranted further inspection by X-ray crystallography in an attempt to elucidate the method by which this activity enhancement was achieved.

Crystallographic studies showed that the modified amino acid side chain is held in position by occupying the hydrophobic pocket that the tyrosine occupies in the wild-type enzyme, and is held in shape by interactions between atoms on the lanthionine residue itself. Additionally, these studies showed that the modified amino acid is present in both L- and D-stereoisomers at the protein backbone. This has not been seen previously in enzymes modified using the same method as was used here, and as such is the first example of a mixed backbone stereochemistry enzyme produced using this method. Attempts to soak in both the substrate and the substrate analogue were unsuccessful, and as such, observation of the binding of the glucuronolactone substrate to the enzyme active site was examined using *in silico* methods.

From the GRID data examining potential substrate binding (figure 4.19) we can see that there is a larger surface area for the substrate to bind in the active site of the modified enzyme compared to the wild-type, suggesting that the modified enzyme is more likely to bind substrate from bulk solvent than the wild-type. When this data is combined with the binding strength data, i.e. that the binding poses involving the modified side chain are harder to disrupt than the binding poses involving the wild-type residues, this suggests that when the lanthionine residue binds substrate, the substrate is less likely to be removed from the active site than if it bound the wild-type residues. The lanthionine is observed most often engaging the ring hydroxyl groups via the carboxylate group on the ncAA. The position of the side chain of the ncAA in the active site is stabilised via a hydrogen bonding network between the carboxylate group, the amine

group on the side chain, the sulphur atom, and the amine group in the main chain. This hydrogen bonding network keeps the carboxylate pointing towards the centre of the active site and available for productive hydrogen bonding.

While the proportion of time spent in proximity to the catalytic residues during simulation suggests that the wild-type enzyme would be more catalytic, this is at odds with the assay data, and could be due to the short time span of the simulations compared to the enzymatic rate. The full length removal data was inconclusive, suggesting either that the modified side chain has no effect on the rate of product removal, or it occurs over a much longer timespan than that simulated. Overall, it appears that the Lni-containing active site is more available for binding of the substrate and when the substrate is bound, substrate in poses involving the lanthionine residue are held more strongly than substrate interacting only with wild-type residues.

Chapter 5 - Conclusion

5.1 Summary

This work described the screening for and discovery of an enzyme containing a non-canonical amino acid (ncAA) in the active site of an enzyme, Neuraminic acid lyase (NAL). The resulting enzyme catalysed the aldol reaction between pyruvate and an aldehyde, glucuronolactone, better than the wild-type enzyme for the same reaction. The structure of this enzyme was then studied by X-ray crystallography and *in silico* methods in an effort to elucidate the mechanism by which the enzyme achieves this rate enhancement.

First, a long-established assay was assessed for its suitability for use in high throughput screen format. The variability of the assay was found to be acceptable, and the signal to noise ratio was increased to enable more reliable hit identification. For the screen, a diverse set of 15 aldehydes were chosen as potential substrates, with members of the selection representing different lengths of substrate, different stereochemical configurations, and even one cyclic molecule. Six positions in the active site were targeted, some of which have been shown in previous work to have an effect on the substrate specificity of NAL and some of which have not previously been shown to have any influence. The mutated residues were positioned around the active site and at different distances from the active site lysine so that any resulting ncAAs would be placed in varied positions across the active site. Ten thiols were used to create the ncAAs on the protein backbone and these were a diverse mixture that would produce both ncAAs that were very similar to canonical amino acids, e.g. thialysine, and also very dissimilar, e.g. (1-hydroxy-3-hexanyl) cysteine and (2-(3,5-dimethyl-4-isoxazolyl)ethyl)cysteine. The thiols also displayed a number of functional groups and ranged from hydrophobic cyclohexanethiol to zwitterionic and hydrophilic cysteine. These sets of variables produced 60 different enzymes to assess for activity, 10 thiols all in 6 different positions, and 15 aldehyde substrates to give a total of 900 different enzyme-aldehyde pairs that were screened for enhanced activity over the wild-type enzyme for each aldehyde.

The most significant hit that was discovered was Y252C mutant modified with cysteine as the thiol to produce Y252Lanthionine (Y252Lni), when

catalysing the aldol reaction between pyruvate and glucuronolactone. The use of this substrate is quite interesting, as NAL has previously been shown to accept sugars in their linear conformation, and glucuronolactone is cyclic. Y252Lni was found to have an activity approximately 3.5-fold higher than that of wild-type at 10 mM glucuronolactone and 100 mM pyruvate. Full apparent kinetics for glucuronolactone were performed, but a number of factors negatively impacted the data; there was a relatively large degree of variability between the repeats, and significant substrate inhibition occurred above 12 mM glucuronolactone in both enzymes, both of which made the data fitting more complex. Subsequently, the kinetic values for both Y252Lni and wild-type for this reaction had large errors. Time constraints unfortunately meant that these kinetics could not be repeated to gain higher quality data. However, even the low quality data still showed that Y252Lni was a better catalyst for the glucuronolactone and pyruvate reaction than the wild-type, and as such, the effect that this position and the nCAA in particular had on this reaction was examined more closely.

Attempts were made to identify the exact product being made by the enzyme, and while these showed that the reaction occurring was most likely the reaction that was anticipated, there were a number of different molecules present. These could be epimers at the aldol carbon or a number of different structural isoforms.

Saturation mutagenesis at position 252 was performed, and all of the variants were tested for activity in the reaction between glucuronolactone and pyruvate at a single concentration of each substrate, 10 mM glucuronolactone and 100 mM pyruvate. Of the 20 canonical amino acid variants, 8 were more active for the reaction than Y252Lni and 17 of the 19 non-wild-type canonical amino acid variants were more active than the wild-type enzyme. When the molecular weights of each of the amino acids at 252 were compared to their activity as recorded by the assay it was observed that there was a rough negative correlation between amino acid side chain size at 252 and activity. This suggests that removal of bulk at position 252 allows glucuronolactone to enter the active site more easily. The activity of Y252Lni enzyme flouts this trend by having a higher mass than any canonical amino acid, while also having a high activity. This suggests that the mechanism of activity enhancement used by Y252Lni may not be the same as that displayed in the canonical amino acid variants.

When the structure of Y252Lni was studied by X-ray crystallography, the resulting structure had reasonable resolution, 1.95 Å, and the ncAA side chain was found to be well ordered in the active site. There were 12 monomers in the unit cell, 11 of these had side chain density at position 252, and the side chain was in a similar position in all 11 of these monomers. In addition, the resulting ncAA had both D- and L-stereochemistry at the protein backbone in different monomers, in approximately 1:1 ratio. Unfortunately, efforts to obtain a crystal with either glucuronolactone or a non-catalytic analogue were unsuccessful, so substrate binding was examined *in silico*. These determined that Y252Lni had a larger surface area in the active site that was likely to bind the substrate compared to the wild-type enzyme, and molecular dynamics simulations suggested that when the substrate binds so that interacts with the ncAA side chain, it is held in the active site more strongly than when it binds and interacts with residues present in the wild-type enzyme. These data suggest that increased residence time in the active site of Y252Lni leads to the increased rate of catalysis.

This work therefore represents the creation of not only an ncAA-containing enzyme with enhanced activity for a reaction than wild-type, but the creation of a functional enzyme with a D-amino acid in the active site, the first such creation in the author's knowledge. The discovery that NAL is able to catalyse the reaction between glucuronolactone and pyruvate is also novel, and given previous evidence that NAL binds sugars in the open chain form and not the ring form (Daniels et al., 2014), represents a new class of molecules that NAL may be able to synthesise, as molecules derived from glucuronolactone could theoretically form bicyclic structures in solution.

5.2 Future directions

This work has opened up a number of avenues of further research which would be valuable to investigate further. The previously unrecognised potential of NAL as a biocatalyst for the synthesis of bicyclic lactones could be further explored. A directed evolution effort, for example error-prone PCR or CASTing (Reetz et al., 2006a, Reetz et al., 2006b) could be applied to NAL to optimise it for lactone molecules. These mutants could then be applied to the synthesis of a diverse library of glucuronolactone homologues.

The potential of ncAAs to create novel biocatalysts has so far been underexplored. Applications of ncAAs to enzymes have focused on aspects such as improving thermal stability, expanding or changing substrate scope and optimising the catalytic properties. While these aims are commendable, and showcase the utility of ncAAs in enhancing the properties of enzymes, much more ambitious changes could be made to enzymes using ncAAs.

Using an ncAA to insert a metal ion, either through insertion of a metal ion binding site or through insertion of a covalently bound metal, could allow metal ion-catalysed reactions to be performed on an enzyme scaffold. There exists the potential to insert not only iron or zinc ions, which are fairly common in enzymes, but also metals such as palladium, used in cross-coupling reactions (Seechurn et al., 2012), and titanium, used in α -olefin polymerisation (Kashiwa, 2004), giving rise to completely novel enzymes with catalytic abilities previously unknown in biochemistry.

Even without the use of metal ions, ncAAs can provide functional groups that do not appear in the canonical amino acids. For example, secondary amine groups can form a Schiff base in the same way as primary amines, such as that found in the active site of NAL. However, when a Schiff base forms on a secondary amine, a positively charged nitrogen species occurs, called an iminium ion. This iminium ion is an electrophilic species, which occurs only rarely in the canonical amino acids. The presence of an electrophile in the active site of an enzyme allows it to perform a range of reactions that are either rare or unknown in enzymes currently (Erkkilä et al., 2007).

This functionality is theoretically easy to introduce to NAL via the method used in this thesis. In chapter 3, one of the thiols used in the modification procedure was aminoethanethiol, creating a thialysine residue, a structural homologue of lysine with a sulphur atom at the γ -position on the side chain. If this thiol was replaced with *N*-methyl-aminoethanethiol, this would provide a secondary amine in the active site in the same position as the primary amine in the wild-type enzyme (figure 5.1). This should allow it to undergo the Schiff base reaction with pyruvate, forming an iminium ion in the active site.

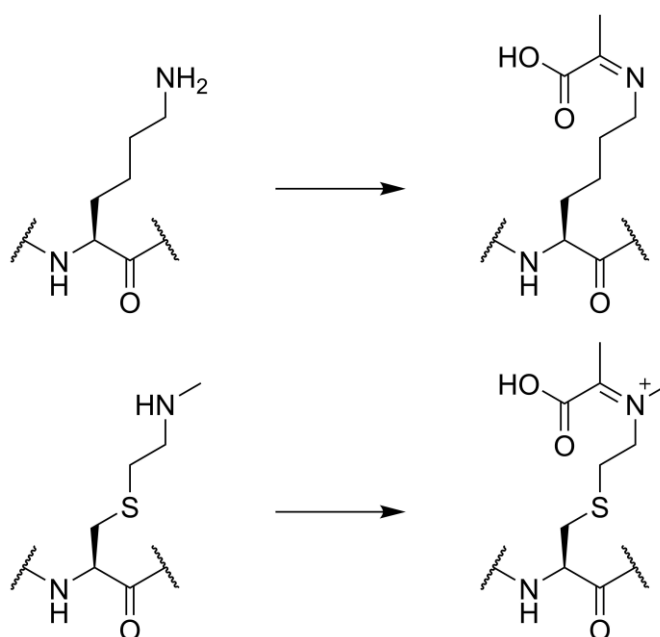


Figure 5.1 Chemical diagram of Lysine forming a Schiff base (**above**) and *N*-methyl-thialysine forming an iminium ion in the same reaction (**below**).

This thiol, *N*-methyl-aminoethanethiol, was synthesised during the course of the project, but successful protein modification could not be obtained when the synthesised thiol was used in the modification reaction. As such, the potential of this side chain to perform novel reactions in place of the catalytic lysine could not be tested. If the appropriate side chain could be inserted into the active site in this position, either using the method shown here or an alternative method, the resulting enzyme's potential catalytic abilities would be very interesting.

5.3 Concluding remarks

The study of biocatalysis is growing fast in the literature, and biocatalysts are being much more widely used in industry for the synthesis of bulk and fine chemicals and pharmaceuticals. As such, there is an increasing need for novel catalysts to supply the demand for green catalysis. Techniques like directed evolution and site-directed mutagenesis are powerful tools to adapt biocatalytic enzymes for industrial use, but they are limited to using the canonical 20 amino acids. Non-canonical amino acids have the potential to make up for the lack of side chain diversity seen in the canonical amino acids, supplying functional groups or structural components that may be lacking.

Even from a purely research perspective, that D-amino acids can be accommodated in some regions of enzyme backbones, and in positions that actively function during catalysis is exciting. This effectively doubles the amino acids available to experimenters performing ncAA mutagenesis, unlocking new avenues of research that have not been fully explored to date. Non-canonical amino acids offer an exciting and untapped resource in an enzyme engineers' toolbox, potentially allowing researchers to surpass the abilities of enzymes using canonical amino acids, expanding the abilities of enzymes further than is possible in Nature.

Chapter 6 - Bibliography

- Agostini, F., Völler, J.-S., Kocsch, B., Acevedo-Rocha, C. G., Kubyshkin, V. and Budisa, N. (2017) 'Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology', *Angewandte Chemie International Edition*, 56(33), pp. 9680-9703.
- Barbosa, J. A. R. G., Smith, B. J., DeGori, R., Ooi, H. C., Marcuccio, S. M., Campi, E. M., Jackson, W. R., Brossmer, R., Sommer, M. and Lawrence, M. C. (2000) 'Active site modulation in the *N*-acetylneuraminidase lyase sub-family as revealed by the structure of the inhibitor-complexed *Haemophilus influenzae* enzyme', *Journal of Molecular Biology*, 303(3), pp. 405-421.
- Belogurov, A., Kozyr, A., Ponomarenko, N. and Gabibov, A. (2009) 'Catalytic antibodies: balancing between Dr. Jekyll and Mr. Hyde', *BioEssays*, 31(11), pp. 1161-1171.
- Bond, S., Leimkuhler, B. and Laird, B. B. (2002) 'The Nosé-Poincaré Method for Constant Temperature Molecular Dynamics', *Journal of Computational Physics*, 151(1), pp. 114-134.
- Campeotto, I., Bolt, A. H., Harman, T. A., Dennis, C., Trinh, C. H., Phillips, S. E. V., Nelson, A., Pearson, A. R. and Berry, A. (2010) 'Structural Insights into Substrate Specificity in Variants of *N*-Acetylneuraminic Acid Lyase Produced by Directed Evolution', *Journal of Molecular Biology*, 404(1), pp. 56-69.
- Case, D. A., Cheatham, T. E., Darden, T. O. M., Gohlke, H., Luo, R. A. Y., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. (2005) 'The Amber Biomolecular Simulation Programs', *Journal of Computational Chemistry*, 26(16), pp. 1668-1688.
- Case, D. A., Darden, T. A., Cheatham, I., T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V. and Kollman, P. A. (2008) *AMBER 10*.
- Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. and Davis, B. G. (2012) 'Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications', *Angewandte Chemie International Edition*, 51(8), pp. 1835-1839.
- Chan, K. K., Wood, B. M., Fedorov, A. A., Fedorov, E. V., Imker, H. J., Amyes, T. L., Richard, J. P., Almo, S. C. and Gerlt, J. A. (2009) 'Mechanism of the Orotidine 5'-Monophosphate Decarboxylase-Catalyzed Reaction: Evidence for Substrate Destabilization', *Biochemistry*, 48(24), pp. 5518-5531.
- Chen, S. H., Hwang, D. R., Chen, G. H., Hsu, N. S., Wu, Y. T., Li, T. L. and Wong, C. H. (2012) 'Engineering transaldolase in *Pichia stipitis* to improve bioethanol production', *ACS Chemical Biology*, 7(3), pp. 481-6.

- Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A. and Pedersen, A. H. (1999) 'Directed evolution of a fungal peroxidase', *Nature Biotechnology*, 17(4), pp. 379-84.
- Coggin, J. H. and Martin, W. R. (1965) '6-Diazo-5-Oxo-l-Norleucine Inhibition of *Escherichia coli*', *Journal of Bacteriology*, 89(5), pp. 1348-1353.
- Damitio, J., Smith, G., Meany, J. E. and Pocker, Y. (1992) 'A comparative study of the enolization of pyruvate and the reversible dehydration of pyruvate hydrate', *Journal of the American Chemical Society*, 114(8), pp. 3081-3087.
- 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. and Berry, A. (2014) 'Reaction Mechanism of *N*-Acetylneuraminic Acid Lyase Revealed by a Combination of Crystallography, QM/MM Simulation, and Mutagenesis', *ACS Chemical Biology*, 9(4), pp. 1025-1032.
- de Réaumur, R. A. F. (1752) 'Observations sur la digestion des oiseaux', *Histoire de l'academie royale des sciences*, (1).
- Devenish, S. R. A. and Gerrard, J. A. (2009) 'The quaternary structure of *Escherichia coli* *N*-acetylneuraminate lyase is essential for functional expression', *Biochemical and Biophysical Research Communications*, 388(1), pp. 107-111.
- Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*, 66(Pt 4), pp. 486-501.
- Erkkilä, A., Majander, I. and Pihko, P. M. (2007) 'Iminium Catalysis', *Chemical Reviews*, 107(12), pp. 5416-5470.
- Ferrero, M. A., Reglero, A., Fernandez-Lopez, M., Ordas, R. and Rodriguez-Aparicio, L. B. (1996) '*N*-acetyl-D-neuraminic acid lyase generates the sialic acid for colominic acid biosynthesis in *Escherichia coli* K1', *Biochemical Journal*, 317(Pt 1), pp. 157-165.
- Fischer, E. (1894) 'Einfluss der Configuration auf die Wirkung der Enzyme', *Berichte der deutschen chemischen Gesellschaft*, 27(3), pp. 2985-2993.
- Fowden, L., Lewis, D. and Tristram, H. (1967) 'Toxic Amino Acids: Their Action as Antimetabolites', *Advances in Enzymology and Related Areas of Molecular Biology*: John Wiley & Sons, Inc., pp. 89-163.
- Fujihashi, M., Mnpotra, J. S., Mishra, R. K., Pai, E. F. and Kotra, L. P. (2015) 'Orotidine Monophosphate Decarboxylase - A Fascinating Workhorse Enzyme with Therapeutic Potential', *Journal of Genetics and Genomics*, 42(5), pp. 221-234.
- Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N. and Hilvert, D. (2013) 'Evolution of a designed retro-aldolase leads to complete active site remodeling', *Nature Chemical Biology*, 9(8), pp. 494-498.
- Giver, L., Gershenson, A., Freskgard, P.-O. and Arnold, F. H. (1998) 'Directed evolution of a thermostable esterase', *Proceedings of the National Academy of Sciences*, 95(22), pp. 12809-12813.
- Goodford, P. J. (1985) 'A computational procedure for determining energetically favorable binding sites on biologically important macromolecules', *Journal of Medicinal Chemistry*, 28(7), pp. 849-857.

- Greenberg, W. A., Varvak, A., Hanson, S. R., Wong, K., Huang, H., Chen, P. and Burk, M. J. (2004) 'Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates', *Proceedings of the National Academy of Sciences of the United States of America*, 101(16), pp. 5788-93.
- Gutierrez, M. L., Garrabou, X., Agosta, E., Servi, S., Parella, T., Joglar, J. and Clapés, P. (2008) 'Serine Hydroxymethyl Transferase from *Streptococcus thermophilus* and L-Threonine Aldolase from *Escherichia coli* as Stereocomplementary Biocatalysts for the Synthesis of β -Hydroxy- α,ω -diamino Acid Derivatives', *Chemistry - A European Journal*, 14(15), pp. 4647-4656.
- Harris, P., Poulsen, J.-C. N., Jensen, K. F. and Larsen, S. (2002) 'Substrate Binding Induces Domain Movements in Orotidine 5'-Monophosphate Decarboxylase', *Journal of Molecular Biology*, 318(4), pp. 1019-1029.
- Hayden, F. G., Atmar, R. L., Schilling, M., Johnson, C., Poretz, D., Paar, D., Huson, L., Ward, P., Mills, R. G. and Group, t. O. S. (1999) 'Use of the Selective Oral Neuraminidase Inhibitor Oseltamivir to Prevent Influenza', *New England Journal of Medicine*, 341(18), pp. 1336-1343.
- Hoffmann, R. (1963) 'An Extended Hückel Theory. I. Hydrocarbons', *The Journal of Chemical Physics*, 39(6), pp. 1397-1412.
- Horn, M. J., Jones, D. B. and Ringel, S. J. (1941) 'Isolation of a new sulfur-containing amino acid (lanthionine) from sodium carbonate-treated wool', *Journal of Biological Chemistry*, 138(1), pp. 141-149.
- Hsu, C.-C., Hong, Z., Wada, M., Franke, D. and Wong, C.-H. (2005) 'Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase', *Proceedings of the National Academy of Sciences of the United States of America*, 102(26), pp. 9122-9126.
- Huang, S., Yu, H. and Chen, X. (2007) 'Disaccharides as Sialic Acid Aldolase Substrates: Synthesis of Disaccharides Containing a Sialic Acid at the Reducing End', *Angewandte Chemie International Edition*, 46(13), pp. 2249-2253.
- Jennewein, S., Schurmann, M., Wolberg, M., Hilker, I., Luiten, R., Wubbolts, M. and Mink, D. (2006) 'Directed evolution of an industrial biocatalyst: 2-deoxy-D-ribose 5-phosphate aldolase', *Biotechnology Journal*, 1(5), pp. 537-48.
- Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L. and Baker, D. (2008) 'De Novo Computational Design of Retro-Aldol Enzymes', *Science*, 319(5868), pp. 1387-1391.
- Joerger, A. C., Mayer, S. and Fersht, A. R. (2003) 'Mimicking natural evolution *in vitro*: An N-acetylneuraminase lyase mutant with an increased dihydrodipicolinate synthase activity', *Proceedings of the National Academy of Sciences of the United States of America*, 100(10), pp. 5694-5699.
- Kashiwa, N. (2004) 'The discovery and progress of MgCl₂-supported TiCl₄ catalysts', *Journal of Polymer Science Part A: Polymer Chemistry*, 42(1), pp. 1-8.

- Khersonsky, O., Röthlisberger, D., Wollacott, A. M., Murphy, P., Dym, O., Albeck, S., Kiss, G., Houk, K. N., Baker, D. and Tawfik, D. S. (2011) 'Optimization of the *in silico* designed Kemp eliminase KE70 by computational design and directed evolution', *Journal of Molecular Biology*, 407(3), pp. 391-412.
- Koshland, D. E. (1958) 'Application of a Theory of Enzyme Specificity to Protein Synthesis', *Proceedings of the National Academy of Sciences of the United States of America*, 44(2), pp. 98-104.
- Krüger, D., Schauer, R. and Traving, C. (2001) 'Characterization and mutagenesis of the recombinant *N*-acetylneuraminase lyase from *Clostridium perfringens*', *European Journal of Biochemistry*, 268(13), pp. 3831-3839.
- Kühne, W. F. (1877) 'Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente', *Verhandlungen des naturhistorisch-medicinischen Vereins zu Heidelberg*, 1, pp. 190-193.
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M. and Isaacs, F. J. (2013) 'Genomically Recoded Organisms Expand Biological Functions', *Science*, 342(6156), pp. 357-360.
- Li, Y., Yu, H., Cao, H., Lau, K., Muthana, S., Tiwari, V. K., Son, B. and Chen, X. (2008) '*Pasteurella multocida* sialic acid aldolase: a promising biocatalyst', *Applied Microbiology and Biotechnology*, 79(6), pp. 963.
- Lindskog, S. (1997) 'Structure and mechanism of carbonic anhydrase', *Pharmacology & Therapeutics*, 74(1), pp. 1-20.
- Liu, J.-Q., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S. and Yamada, H. (2000) 'Diversity of microbial threonine aldolases and their application', *Journal of Molecular Catalysis B: Enzymatic*, 10(1), pp. 107-115.
- Matthews, B. W. (1968) 'Solvent content of protein crystals', *Journal of Molecular Biology*, 33(2), pp. 491-497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(Pt 4), pp. 658-674.
- McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nuñez, A., Butrym, E. D., Richards, M. P., Wang, C.-s., Cheng, G., Zhao, Z. and Wang, C. (2004) 'Fermented beverages of pre- and proto-historic China', *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), pp. 17593-17598.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D: Biological Crystallography*, 67(Pt 4), pp. 355-367.
- Northrop, J. H. (1930) 'Crystalline Pepsin I. Isolation and Tests of Purity', *The Journal of General Physiology*, 13(6), pp. 739-766.
- Ošlaj, M., Cluzeau, J., Orkić, D., Kopitar, G., Mrak, P. and Časar, Z. (2013) 'A Highly Productive, Whole-Cell DERA Chemoenzymatic Process for

- Production of Key Lactonized Side-Chain Intermediates in Statin Synthesis', *PLOS ONE*, 8(5), pp. e62250.
- Paul, M. and Van der Donk, W. A. (2005) 'Chemical and Enzymatic Synthesis of Lanthionines', *Mini-Reviews in Organic Chemistry*, 2(1), pp. 23-37.
- Payen, A. and Persoz, J.-F. (1833) 'Mémoire sur la diastase, les principaux produits de ses réactions et leurs applications aux arts industriels', *Annales de chimie et de physique*, 53, pp. 73-92.
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., Ferguson, D., Seibel, G. and Kollman, P. (1995) 'AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules', *Computer Physics Communications*, 91(1), pp. 1-41.
- Polgár, L. (2005) 'The catalytic triad of serine peptidases', *Cellular and Molecular Life Sciences CMLS*, 62(19), pp. 2161-2172.
- Radzicka, A. and Wolfenden, R. (1995) 'A proficient enzyme', *Science*, 267(5194), pp. 90-93.
- Reetz, M. T., Carballeira, J. D., Peyralans, J., Höbenreich, H., Maichele, A. and Vogel, A. (2006a) 'Expanding the Substrate Scope of Enzymes: Combining Mutations Obtained by CASTing', *Chemistry - A European Journal*, 12(23), pp. 6031-6038.
- Reetz, M. T., Soni, P., Fernandez, L., Gumulya, Y. and Carballeira, J. D. (2010) 'Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method', *Chemical Communications*, 46(45), pp. 8657-8658.
- Reetz, M. T., Wang, L.-W. and Bocola, M. (2006b) 'Directed Evolution of Enantioselective Enzymes: Iterative Cycles of CASTing for Probing Protein-Sequence Space', *Angewandte Chemie*, 118(8), pp. 1258-1263.
- Salque, M., Bogucki, P. I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M. and Evershed, R. P. (2013) 'Earliest evidence for cheese making in the sixth millennium BC in northern Europe', *Nature*, 493(7433), pp. 522-525.
- Schauer, R. and Wember, M. 1996. Isolation and Characterization of Sialate Lyase from Pig Kidney. *Biological Chemistry*.
- Schauer, R., Wember, M., Wirtz-Peitz, F. and Ferreira do Amaral, C. (1971) 'Studies on the substrate specificity of acylneuraminidase', *Hoppe-Seyler's Zeitschrift für Physiologische Chemie*, 352(8), pp. 1073-80.
- Seechurn, C. C. C. J., Kitching, M. O., Colacot, T. J. and Snieckus, V. (2012) 'Palladium-Catalyzed Cross-Coupling: A Historical Contextual Perspective to the 2010 Nobel Prize', *Angewandte Chemie International Edition*, 51(21), pp. 5062-5085.
- Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St.Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E. and Baker, D. (2010) 'Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction', *Science*, 329(5989), pp. 309-313.

- Smith, B. J., Lawrence, M. C. and Barbosa, J. A. R. G. (1999) 'Substrate-Assisted Catalysis in Sialic Acid Aldolase', *The Journal of Organic Chemistry*, 64(3), pp. 945-949.
- Sturgeon, J. B. and Laird, B. B. (1999) 'Symplectic algorithm for constant-pressure molecular dynamics using a Nose-Poincare thermostat', *The Journal of Chemical Physics*, 112(8), pp. 3474-3482.
- Timms, N., Windle, C. L., Polyakova, A., Ault, J. R., Trinh, C. H., Pearson, A. R., Nelson, A. and Berry, A. (2013) 'Structural insights into the recovery of aldolase activity in *N*-acetylneuraminic acid lyase by replacement of the catalytically active lysine with γ -thialysine by using a chemical mutagenesis strategy', *Chembiochem*, 14(4), pp. 474-81.
- Toth, K., Amyes, T. L., Wood, B. M., Chan, K., Gerlt, J. A. and Richard, J. P. (2007) 'Product Deuterium Isotope Effect for Orotidine 5'-Monophosphate Decarboxylase: Evidence for the Existence of a Short-Lived Carbanion Intermediate', *Journal of the American Chemical Society*, 129(43), pp. 12946-12947.
- Vander Heyden, Y., Nijhuis, A., Smeyers-Verbeke, J., Vandeginste, B. G. M. and Massart, D. L. (2001) 'Guidance for robustness/ruggedness tests in method validation', *Journal of Pharmaceutical and Biomedical Analysis*, 24(5), pp. 723-753.
- von Itzstein, M., Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M. and Penn, C. R. (1993) 'Rational design of potent sialidase-based inhibitors of influenza virus replication', *Nature*, 363(6428), pp. 418-423.
- Wada, M., Hsu, C.-C., Franke, D., Mitchell, M., Heine, A., Wilson, I. and Wong, C.-H. (2003) 'Directed evolution of *N*-acetylneuraminic acid aldolase to catalyze enantiomeric aldol reactions', *Bioorganic & Medicinal Chemistry*, 11(9), pp. 2091-2098.
- Wang, L., Xie, J. and Schultz, P. G. (2006) 'Expanding The Genetic Code', *Annual Review of Biophysics and Biomolecular Structure*, 35(1), pp. 225-249.
- Warren, L. (1959) 'The Thiobarbituric Acid Assay of Sialic Acids', *Journal of Biological Chemistry*, 234(8), pp. 1971-1975.
- Williams, G. J., Woodhall, T., Farnsworth, L. M., Nelson, A. and Berry, A. (2006) 'Creation of a Pair of Stereochemically Complementary Biocatalysts', *Journal of the American Chemical Society*, 128(50), pp. 16238-16247.
- Williams, G. J., Woodhall, T., Nelson, A. and Berry, A. (2005) 'Structure-guided saturation mutagenesis of *N*-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics', *Protein Engineering Design and Selection*, 18(5), pp. 239-246.
- Windle, C. L. (2015) *Altering enzyme activities using chemical modification*. PhD, University of Leeds.
- Windle, C. L., Simmons, K. J., Ault, J. R., Trinh, C. H., Nelson, A., Pearson, A. R. and Berry, A. (2017) 'Extending enzyme molecular recognition with an expanded amino acid alphabet', *Proceedings of the National Academy of Sciences of the United States of America*, 114(10), pp. 2610-2615.

- Wright, T. H., Bower, B. J., Chalker, J. M., Bernardes, G. J. L., Wiewiora, R., Ng, W.-L., Raj, R., Faulkner, S., Vallée, M. R. J., Phanumartwiwath, A., Coleman, O. D., Thézénas, M.-L., Khan, M., Galan, S. R. G., Lercher, L., Schombs, M. W., Gerstberger, S., Palm-Espling, M. E., Baldwin, A. J., Kessler, B. M., Claridge, T. D. W., Mohammed, S. and Davis, B. G. (2016) 'Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity', *Science*, 354(6312).
- Yamashita, M., Tomozawa, T., Kakuta, M., Tokumitsu, A., Nasu, H. and Kubo, S. (2009) 'CS-8958, a Prodrug of the New Neuraminidase Inhibitor R-125489, Shows Long-Acting Anti-Influenza Virus Activity', *Antimicrobial Agents and Chemotherapy*, 53(1), pp. 186-192.
- Yang, A., Ha, S., Ahn, J., Kim, R., Kim, S., Lee, Y., Kim, J., Söll, D., Lee, H.-Y. and Park, H.-S. (2016) 'A chemical biology route to site-specific authentic protein modifications', *Science*, 354(6312), pp. 623.
- Zeder, M. A. and Hesse, B. (2000) 'The Initial Domestication of Goats (*Capra hircus*) in the Zagros Mountains 10,000 Years Ago', *Science*, 287(5461), pp. 2254-2257.
- Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D. S. and Schultz, P. G. (2004) 'Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 101(24), pp. 8882-8887.