Investigation into the Chemical Modification of Active Site Residues of an Aldolase Enzyme

Alun J. Myden

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

The University of Leeds
School of Chemistry

November 2014

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.

Acknowledgements

I would like to thank my supervisors Prof. Adam Nelson and Prof. Alan Berry for giving me this opportunity and for their guidance and support throughout the four years.

Thank you to all past and present members of both the Nelson and Berry groups for making the past four years a pleasant experience. Specifically, thanks to Francesco Marchetti, George Karageorgis, Giorgia Magnatti and Steven Kane for keeping me sane and becoming really close friends throughout the years. I would like to thank all members of Team Aldolase, especially Nicole Timms for laying the foundations of the project and Claire Windle for providing a good introduction into the biochemistry lab and giving me the best mug in the world!

For their extensive research into NAL, thank you Amanda Bolt, Gavin Williams, Thomas Harman and Thomas Woodhall.

Big thanks to George Preston for his (always positive) outlook on life, James Henkelis for his excellent spotting ability, Dave Yeo (for being Dave Yeo) and both Dan William*3*on and James Murray for our 11 am lunches and trips out to REDS and Cattle Grid.

I would like to thank the following people for both their invaluable technical support and interesting stories/chats: Simon Barret, Tanya Marinko-Covell, Martin Huscroft, Chi Trinh, James Ault, Nasir Khan, Norman Chan, Chris Empson and Francis Billinghurst.

Special thanks goes to Laura Cross for putting up with me and providing huge amounts of support! *She has been my rock – 'she's always there, she's an odd shape and she's cold to the touch'!*

A final mention goes to my family for always being there when I need them.

Abstract

It has previously been demonstrated that the biological incorporation and chemical modification of cysteine residues within the active site of the aldolase *N*-acetylneuraminic acid lyase (NAL) can lead to active chemically modified NAL variants.

This thesis reports the exploitation of this joint chemical/biological method in the incorporation of several unnatural amino acids into the active site of *Staphylococcus aureus* NAL (*Sa*NAL). This work was performed in an attempt to broaden the substrate specificity of the protein, to allow for the catalysis of the retro-aldol reaction of analogues of *N*-acetylneuraminic acid (the natural substrate of NAL).

This method was used to incorporate a range of unnatural amino acids into the active site and kinetic parameters of these proteins were assessed with several varied *N*-acetylneuraminic acid analogues:
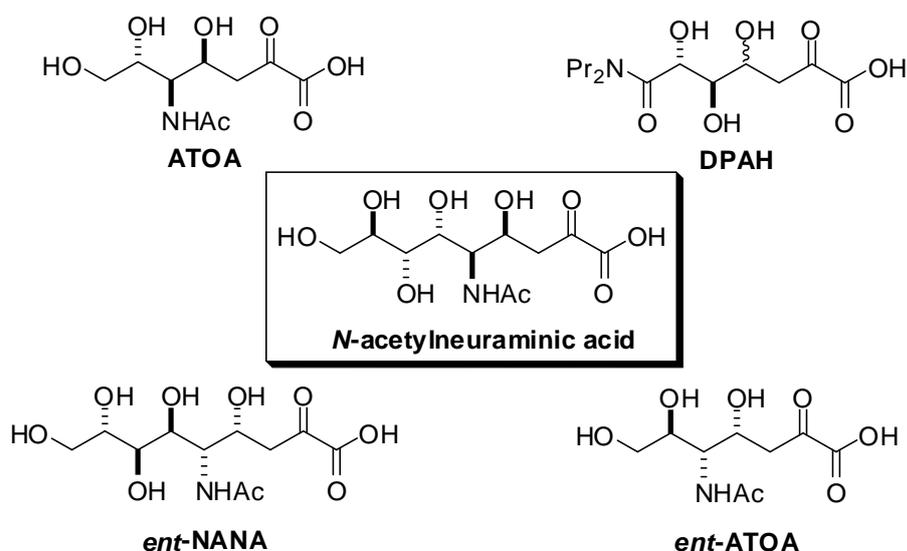
The residues F190 and E192 of *Sa*NAL were chosen for chemical modification as they have previously been shown to be important for the catalytic function of the protein.

A coupled enzyme assay was used to assess the kinetic parameters of the chemically modified variants of *Sa*NAL with the analogues of *N*-acetylneuraminic acid. As a result, it was found that a range of chemically modified amino acids at position 192 of *Sa*NAL would allow for improved specific activity (k_{cat}/K_M) of the retro aldol reaction with the substrate **DPAH**, when compared to wild-type *Sa*NAL with **DPAH**. An interesting relationship was also observed between the functionality of chemically modified amino acids at position 190 and the specific activity of the resulting protein in the retro-aldol of the substrate *N*-acetylneuraminic acid.

Contents

Abbreviations

Å	angstroms (10^{-10} m)
A_{280}	absorbance at 280 nm
A_{340}	absorbance at 340 nm
AA's	amino acids
Ac	acetyl
ATOA	(4*R*,5*R*,6*R*)-5-acetamido-4,6,7-trihydroxy-2-oxoheptanoic acid
COSY	correlation spectroscopy
d	doublet
Dha	dehydroalanine
DHAP	dihydroxyacetone phosphate
DHOB	(2*R*,3*S*)-2,3-dihydroxy-4-oxo-N,N-dipropylbutanamide
DiBr	2,5-dibromohexan-1,6-diamide
DMF	*N*,*N*-dimethylformamide
DNA	deoxyribonucleic acid
DPAH	(5*R*,6*R*)-7-(dipropylamino)-4,5,6-trihydroxy-2,7-dioxoheptanoic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
ent	enantiomer
ent-ATOA	the enantiomer of ATOA
ent-NANA	the enantiomer of *N*-acetylneuraminic acid
ep-PCR	error prone polymerase chain reaction
ESI-MS	electrospray ionisation mass spectrometry

FBP	fructose-1,6-biphosphate
Fur	furanose
× g	standard gravity
GFP	green fluorescent protein
Grubbs	Grubbs Catalyst, 2nd Generation
h	hour
HG II	Hoveyda-Grubbs Catalyst 2nd Generation
HMBC	heteronuclear multiple bond correlation
HMQC	heteronuclear multiple-quantum correlation
HOBt	1-hydroxybenzotriazole
IPTG	isopropyl β-D-1-thiogalactopyranoside
k_{cat}	catalytic constant
K_M	Michaelis constant
LDH	lactate dehydrogenase
ManNAc	*N*-Acetyl-D-mannosamine
mg	milligrams
min	minute
mM	millimolar
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide reduced form
NAL	*N*-acetylneuraminic acid lyase
NANA	*N*-acetylneuraminic acid
NCAA's	non-canonical amino acids
Neu5Ac	*N*-acetylneuraminic acid

PCR	polymerase chain reaction
PDB	protein data bank
RT	room temperature
s	seconds
S. aureus	*Staphylococcus aureus*
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sept	septuplet
sex	sextuplet
TEMED	*N,N,N'N'*-tetramethylethylenediamine
TFA	trifluoroacetic acid
TPPTS	3,3',3''-Phosphinidynetris(benzenesulfonic acid) trisodium salt
Tris	tris(hydroxymethyl) aminomethane
tRNA	transfer ribonucleic acid
UV	ultraviolet

Amino acid abbreviations:

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Unnatural amino acid abbreviations:

Letter code	Unnatural amino acid	Name
ACEC		S-(2S-amino-2-carboxyethyl) cysteine
AEC		S-(2-aminoethyl) cysteine
AMCEC		S-(2S-amino-2-methoxycarbonylethyl) cysteine
CHC		S-cyclohexyl cysteine
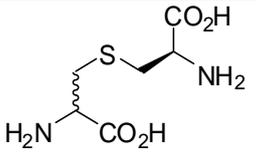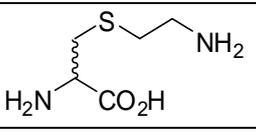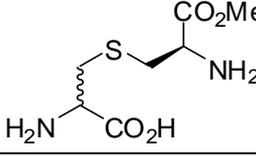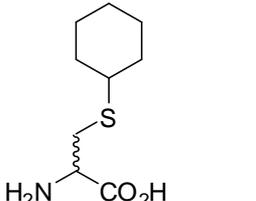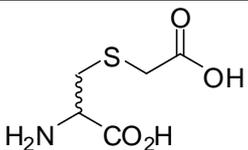
CMC		*S*-carboxymethyl cysteine
CPC		*S*-cyclopentyl cysteine
DHPC		*S*-(2,3-dihydroxypropyl) cysteine
HBC		*S*-(4-hydroxybutyl) cysteine
HEBC		*S*-(1-(2-hydroxyethyl)-butyl) cysteine
HPC		*S* -(2-hydroxypropyl) cysteine
MCMC		*S*-methoxylcarbonylmethyl cysteine
PC		*S*-propenyl cysteine
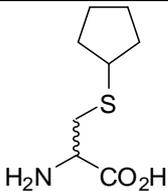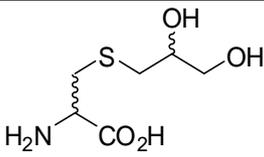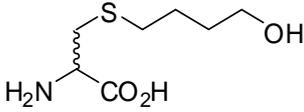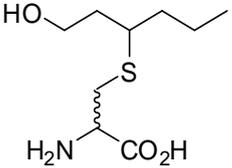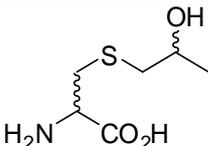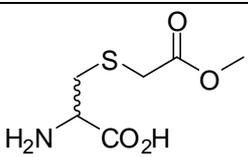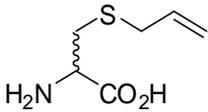
Introduction

1.1 Enzymatic Catalysis

Enzymes are highly selective and efficient catalysts essential to life, some capable of increasing the rate of reactions up to 10^{18}-fold.[1,2] Catalysis can be achieved by re-orientation of substrates, the formation of reactive intermediates and reduction of the transition state energy barrier.[3] Enzymes can also maintain their catalytic activity with substrate concentrations in the range of nanomolar to micromolar; much lower than those typically used in synthetic chemistry (micromolar to molar).[4] Enzymes have the ability to hold the substrates in the correct conformation *via* specific interactions between substrates and active site amino acid residues.[5] Enzyme-substrate interactions include covalent bonding, hydrogen bonding, and hydrophobic interactions.[6] Residues within the active site perform roles such as donating or accepting protons, facilitating metal coordination, and stabilising reactive intermediates. Generally, enzyme mechanisms involve the binding of substrate(s) within the active site, the enzyme-substrate complex undergoing a chemical reaction, and then the immediate release of the newly formed molecule(s).[7] Enzyme-catalysed reactions are often reversible, and the rate determining step of an enzyme reaction is frequently the diffusion of the product from the enzyme.[6] Enzymes are often highly specific and perform their catalytic function with a small range of substrates.[6]

1.1.1 Serine Proteases

Proteases are enzymes capable of cleaving amide bonds within proteins. Serine proteases make up one third of all proteases[8,9] and provide a useful example of enzyme catalysis, illustrating many of the hallmarks of enzymatic catalysis. Serine proteases have a triad of catalytic enzymatic residues – serine, histidine and aspartic acid – which are intimately involved in the catalysis, whilst other residues hold the substrate in place through non-covalent interactions (Scheme 1).[10]
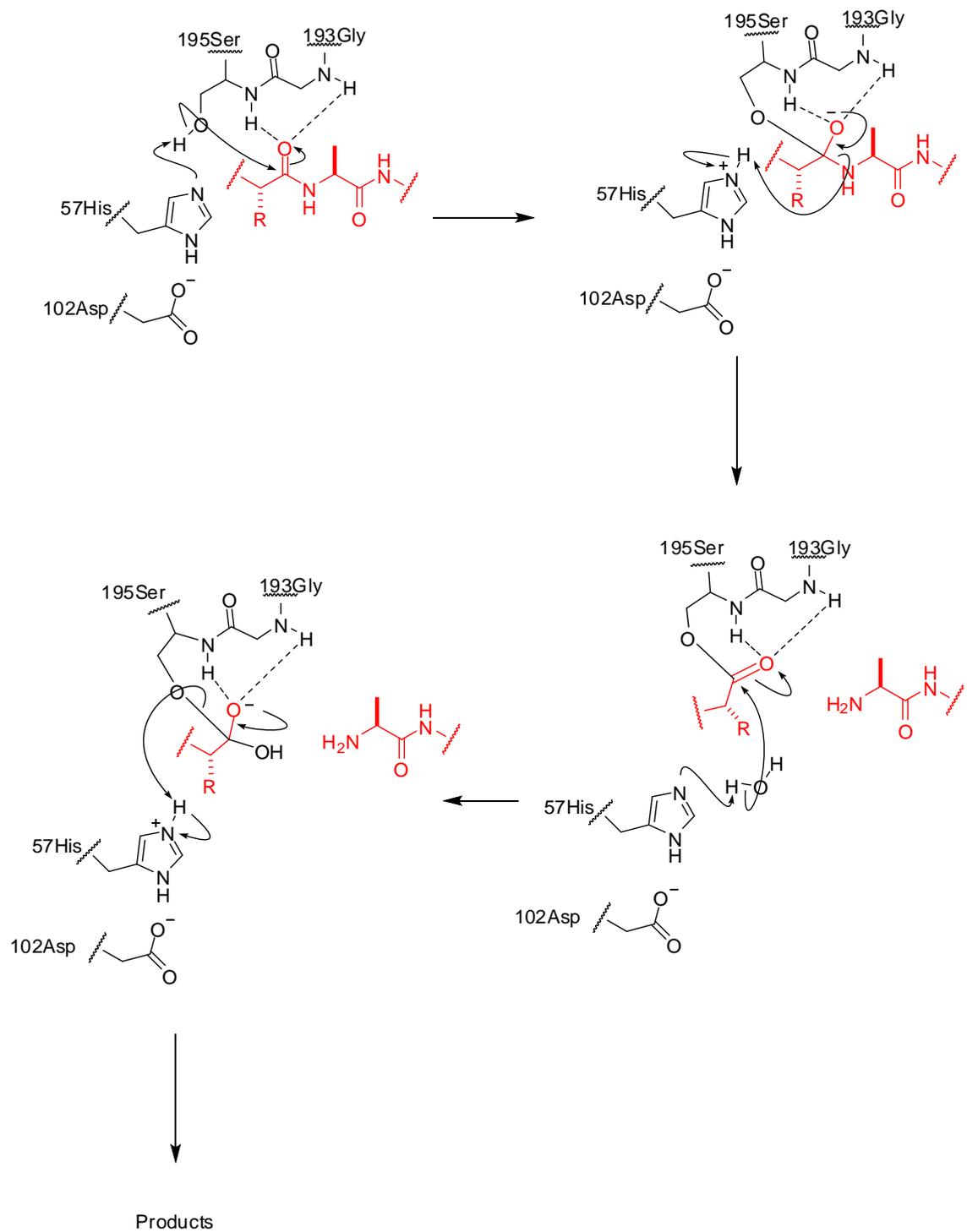
Scheme 1 | General mechanism for chymotrypsin-like serine proteases, adapted from [8,11]. The protein substrate (red) is bound by hydrogen bonding with Ser195 and with Gly193, whilst the His57 molecule acts as a general base, stabilised by Asp102. The residues are numbered according to *Bos taurus* chymotrypsinogen.

Residues Ser195 and Gly193 (numbered according to *Bos taurus* chymotrypsinogen) hold the substrate in place *via* N-H hydrogen bonding, and in doing so they aid the formation of an oxyanion. These interactions facilitate the cleavage of a functional group which is usually very unreactive.[8] The catalytic triad assists the cleavage because serine alone, with a pK_a of ~13, is not ionised under physiological conditions. This catalytic triad is composed of His57, Asp102 and Ser195. His195 (pK_a ~7) is stabilised *via* electrostatic interactions with the deprotonated Asp102 and is therefore able to deprotonate Ser195. This deprotonation leaves Ser195 suitably nucleophilic to attack the polarised amide bond of the protein substrate.[6,8]

1.1.2 Aldolase Enzymes.

The most relevant family of enzymes to this project are the aldolases, enzymes which catalyse aldol reactions. There are two types of aldolase enzymes which are classified by their mechanism. The mechanism of Class I aldolases proceeds *via* an enamine from a Schiff base formed between a catalytic lysine residue and the relevant substrate, which then performs the aldol reaction by attacking an aldehyde (Scheme 2).[12] The mechanism of Class II aldolases, however, proceeds *via* an enolate, stabilised by the coordination of a divalent metal, allowing the aldol reaction to occur (Scheme 3).[13]
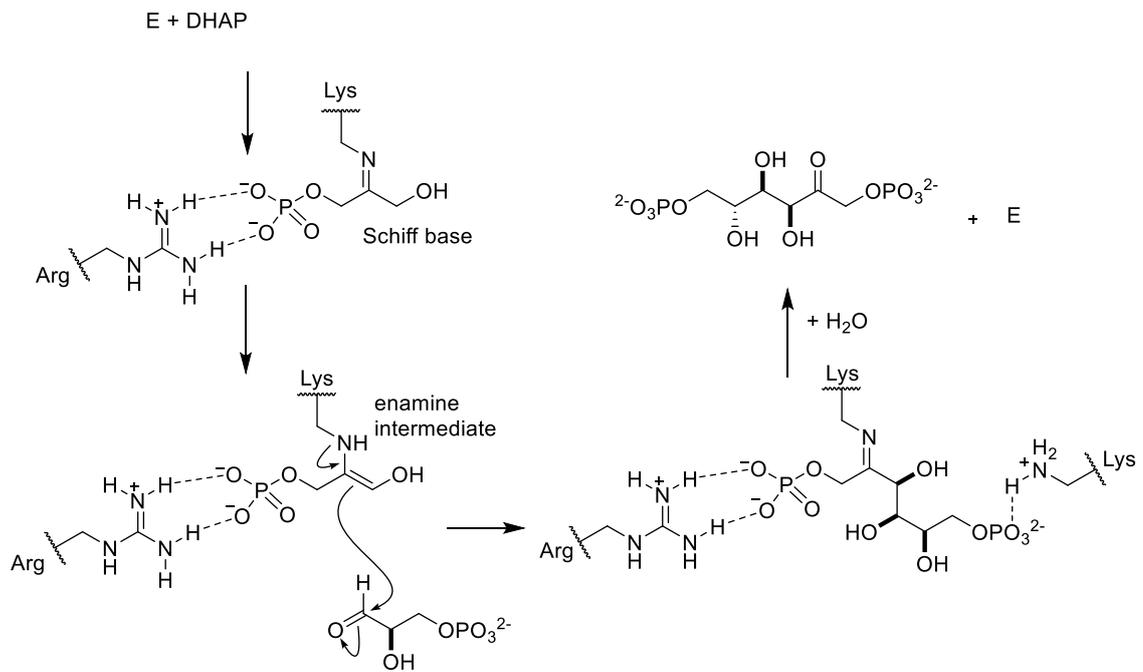
Scheme 2 | The general mechanism for a Class I aldolase, illustrated with fructose-1,6-biphosphate (FBP) aldolase. The substrates are dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate. The catalytic lysine residue forms a Schiff base with DHAP which leads to the formation of the reactive enamine which then attacks the glyceraldehyde molecule. Mechanism adapted from [14].
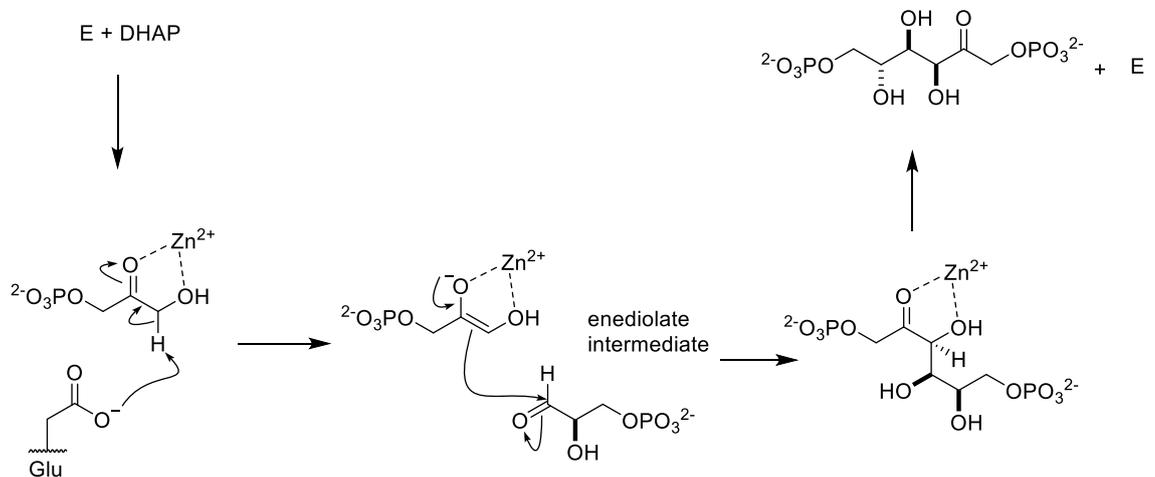
Scheme 3 | The general mechanism for a Class II aldolase exemplified using fructose-1,6-biphosphate (FBP) aldolase. The substrates are dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate. A glutamate residue is able to deprotonate the DHAP, forming an enediolate intermediate stabilised by a metal ion held by the enzyme (E). Mechanism adapted from [14].

1.1.3 Catalytic mechanism of *N*-acetylneuraminic acid lyase (NAL)

This project will focus on the aldolase *N*-acetylneuraminic acid lyase (NAL), a Class I aldolase which catalyses the reversible aldol condensation between *N*-acetyl-D-mannosamine (ManNAc) and pyruvate to yield the sialic acid, *N*-acetylneuraminic acid (Neu5Ac) (Scheme 4). NAL has commercial importance as it provides an efficient way to synthesise pure *N*-acetylneuraminic acid, a key intermediate in the synthesis of the influenza uptake inhibitor, Relenza.[14b]
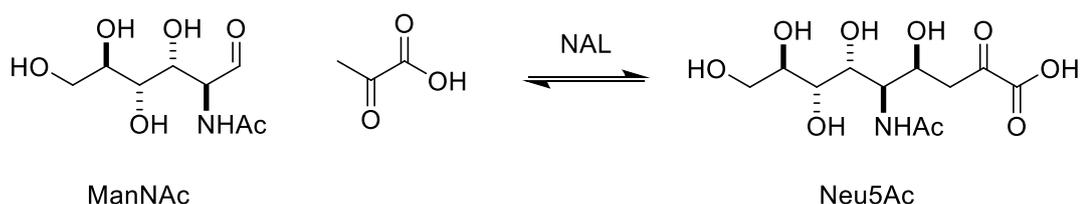
Scheme 4 | The reversible aldol reaction catalysed by the enzyme NAL. ManNAc and **Neu5Ac** are represented in the open chain form for clarity.

In-depth characterisation of the mechanism of this aldolase has been achieved through mutagenesis, crystal structure analysis and quantum mechanics molecular modelling (QM/MM) calculations.[15,16] The enzyme catalyses a bi-uni ordered condensation reaction in which pyruvate first binds to a catalytic lysine residue to form the reactive Schiff base. The substrate ManNAc is recognised by the enzyme through hydrogen bonding from the four alcohols of ManNAc to the residues Glu192, Asp191, Ser208 and Gly189 of *E. Coli* NAL (Figure 1, numbered according to *E. Coli* NAL).
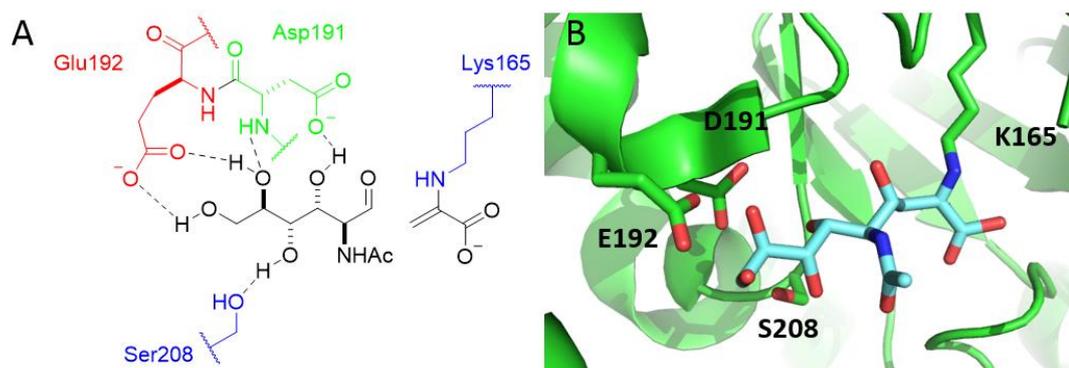
Figure 1 | Structural insight into *E. Coli* NAL. A) Hydrogen bond stabilisation to ManNAc is achieved through the side chains Glu192, Asp191, and Ser208. **B)** Crystal structure of the Y137A *E. Coli* NAL variant (green) covalently bound to full length *N*-acetylneuraminic acid (pale blue). The residues K165, D191, E192 and S208 have been highlighted. Numbering of residues is in accordance with *E. coli* NAL.[15,16]

The mechanism for the aldol reaction between *N*-acetyl-D-mannosamine (ManNAc) and pyruvate catalysed by NAL is shown in Scheme 5. In the forward direction, the enamine of the Schiff base (formed from the attack of pyruvate by Lys165) then undergoes nucleophilic attack onto the aldehyde of the substrate ManNAc. The nucleophilic enamine attacks the *si* face of ManNAc to produce the 4*S* – configured product. The stereoselectivity is achieved through the hydrogen bonding between the aldehyde group of ManNAc and the hydroxyl group of each of the residues Thr167 and Tyr137. During the forward aldol reaction, ManNAc gains a proton from Tyr137. QM/MM calculations suggest that this proton transfer is asynchronous and occurs when the distance between the anion of *N*-acetylneuraminic acid and the proton of Tyr137 is 1.7 Å or smaller. The resulting phenolate at residue 137 is stabilised through hydrogen bonding from the alcohol groups on Ser47 and Tyr110. The triad of amino acids – here, Ser47, Tyr110 and Tyr137 – is a common feature in several enzymes.[17-19] Once the aldolase reaction is complete, hydrolysis of the Schiff base reveals *N*-acetylneuraminic acid. Site-directed mutagenesis studies complement the QM/MM, as the mutants Y137F and Y137A exhibit decreased kinetic parameters.[15,16] The mutant Y137F exhibited ~ 0.2 % of activity of *E. coli* NAL with *N*-acetylneuraminic acid and the variant Y137A exhibited no detectable activity.[16]
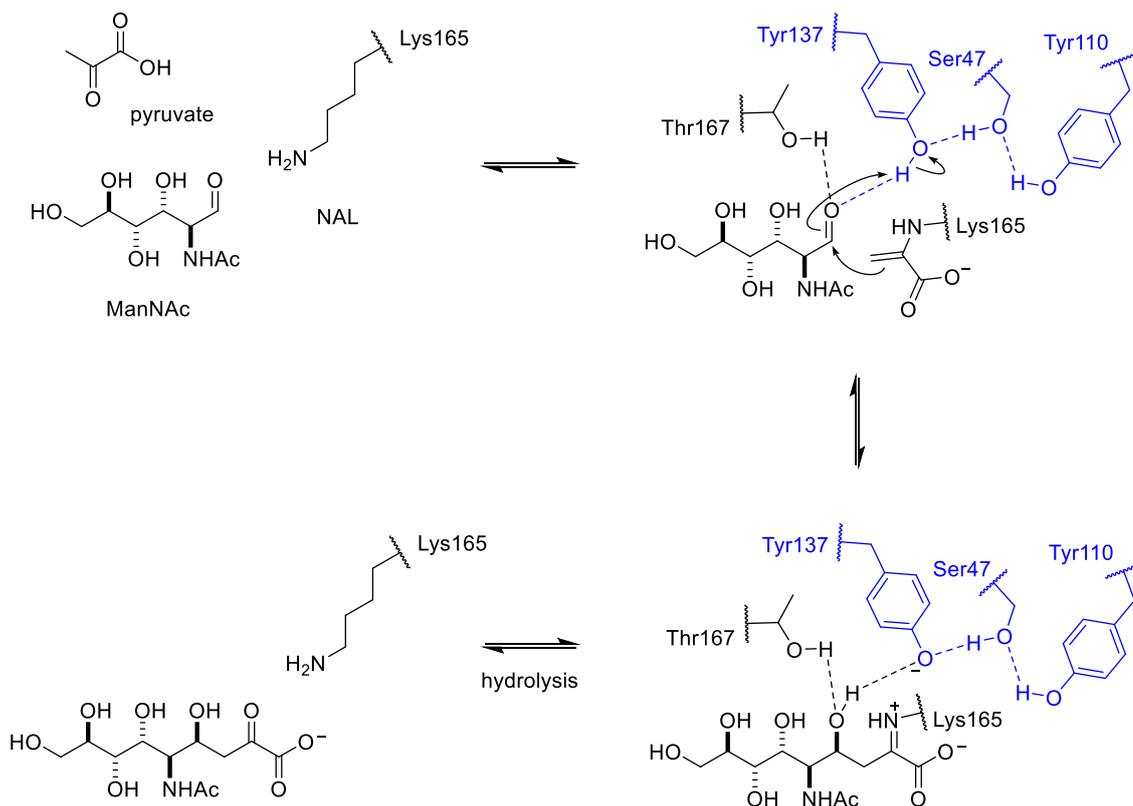
Scheme 5 | Mechanism of the forward aldol reaction between the substrates pyruvate and ManNAc catalysed by *E. coli* NAL. Key steps: 1) Lys165 attacks pyruvate, resulting in a Schiff base. **2)** The Schiff base attacks the *si* face of the substrate ManNAc which is then protonated by Tyr137, stabilised through a triad of residues Tyr137, Ser47 and Tyr110. **3)** Hydrolysis of the resulting iminium yields *N*-acetylneuraminic acid. Figure adapted from [15].

1.2 Enzymes with naturally modified catalytic residues

Post-translational modifications are reactions which occur at specific sites within a protein once the protein has been expressed. Post-translational modifications are important processes which are central to a wide range of critical biological mechanisms.[20] The post-translational modification of a protein can affect the protein in many ways, including: controlling the tertiary structure of a protein, stabilising the protein, and altering the functionality of the protein. For instance, lipidation can help orient groups of proteins into the appropriate membrane within a cell.[21] Post-translational modifications can also act as a switch to turn the function of the protein off or on; protein kinases and phosphatases add or remove phosphate (respectively) from nucleophilic functional groups such as serine, threonine and tyrosine.[20,22,23] Over two hundred types of post-translational modifications are known and almost all of them are catalysed by dedicated enzymes. Only very few proteins undergo automodifications to achieve their post-translational modifications, for example green fluorescent protein (GFP).[20]

Cofactors are non-peptidic species which aid the catalytic function of a protein. An example of this is the zinc ion found in the class II aldolase, discussed in section 1.1.2, Scheme 3.[24] There are several examples of the covalent incorporation of cofactors into the active site of enzymes by post-translational modifications.[25-28] The covalently incorporated cofactor can then act as a key catalytic group essential to the function of the protein. Often the cofactor will act as an electrophile, extending the reactivity of proteins beyond that seen with the canonical twenty amino acid side chains.[25-28] An example of this is the cofactor involved in the catalytic functions of histidine ammonia-lyase and phenylalanine ammonia lyase.[27,29]

1.2.1 Histidine ammonia-lyase and phenylalanine ammonia-lyase

Histidine and phenylalanine ammonia-lyase are enzymes which perform the elimination of ammonia from histidine and phenylalanine respectively.[27,29] Both histidine ammonia-lyase and phenylalanine ammonia-lyase have been found to perform their catalytic functions through a covalently bound methylidene-imidazoline cofactor.[27]

It has been proposed that the methylidene-imidazoline moiety in histidine ammonia-lyase is formed from an autocatalytic cyclic condensation of the amide of residue of Gly144 onto the carbonyl of residue of Ala142, followed by the elimination of the alcohol of Ser143 (Scheme 6).[27,30] Throughout this Chapter BH+ represents a general acid and B represents a general base.
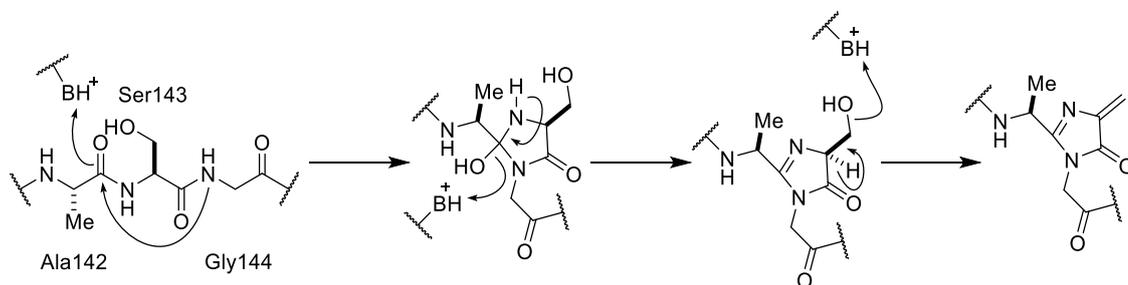
Scheme 6 | Proposed mechanism for the post translational formation of a methylidene-imidazoline cofactor in histidine ammonia-lyase (on the basis of the crystal structure of *Pseudomonas putida* histidine ammonia lyase). The cyclisation involves residues Gly144 and Ala142, followed by the elimination of a primary alcohol from Ser143. Mechanism adapted from [27,30].

The mechanisms of histidine ammonia-lyase and phenylalanine ammonia-lyase are believed to be similar. The proposed mechanism of histidine ammonia-lyase is depicted in Scheme 7.[29] The reaction begins with the attack of the Michael acceptor of methylidene-imidazoline cofactor by the imidazole moiety of histidine (step 1). A proton alpha to the imidazole moiety of histidine is then removed by a general base (step 2). The imidazole of histidine then facilitates the elimination of ammonia (step 3) and the α,β-unsaturated acid product is released, regenerating the methylidene-imidazoline cofactor (step 4).[29]
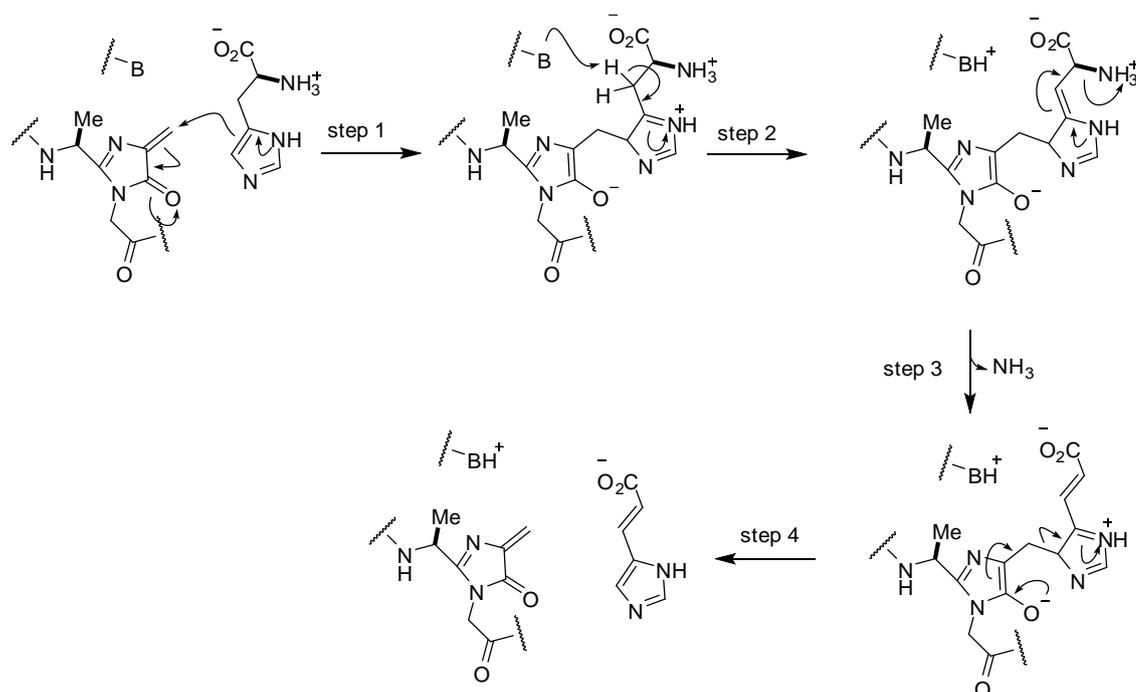
Scheme 7 | Proposed mechanism of histidine ammonia-lyase adapted from [29]. The methylidene-imidazoline cofactor acts as an electron sink, facilitating the elimination of ammonia.

1.2.2 Histidine decarboxylase

Decarboxylation is a process fundamental to many catabolic and anabolic pathways such as glycolysis and the Krebs cycle, providing a source of energy for biological cells and organisms.[6] Broadly speaking, decarboxylases utilise cofactors to facilitate the breakdown of α- and β- keto acids, amino acids and carbohydrates. Examples of these cofactors include: thiamine pyrophosphate for α-keto acid decarboxylation, and pyridoxal 5'-phosphate for amino acid decarboxylation. This is not the case for all decarboxylases, as some obtain their catalytic ability through post-translational modification.[20,26]

Histidine decarboxylase is an example of an enzyme with a post-translationally modified electrophilic pyruvoyl functionality, integral to the catalytic process of the enzyme.[30] The pyruvoyl functionality is formed from an intramolecular attack on a neighbouring amide by the alcohol of a serine residue, followed by cleavage and hydration (Scheme 8). Histidine decarboxylase was one of the first pyruvoyl enzymes to be discovered.[6,30,31] Other pyruvoyl post-translational modifications which catalyse α-decarboxylation of amino acids include *S*-adenosylmethionine, phosphatidyl serine, and aspartate α-decarboxylases.[6]
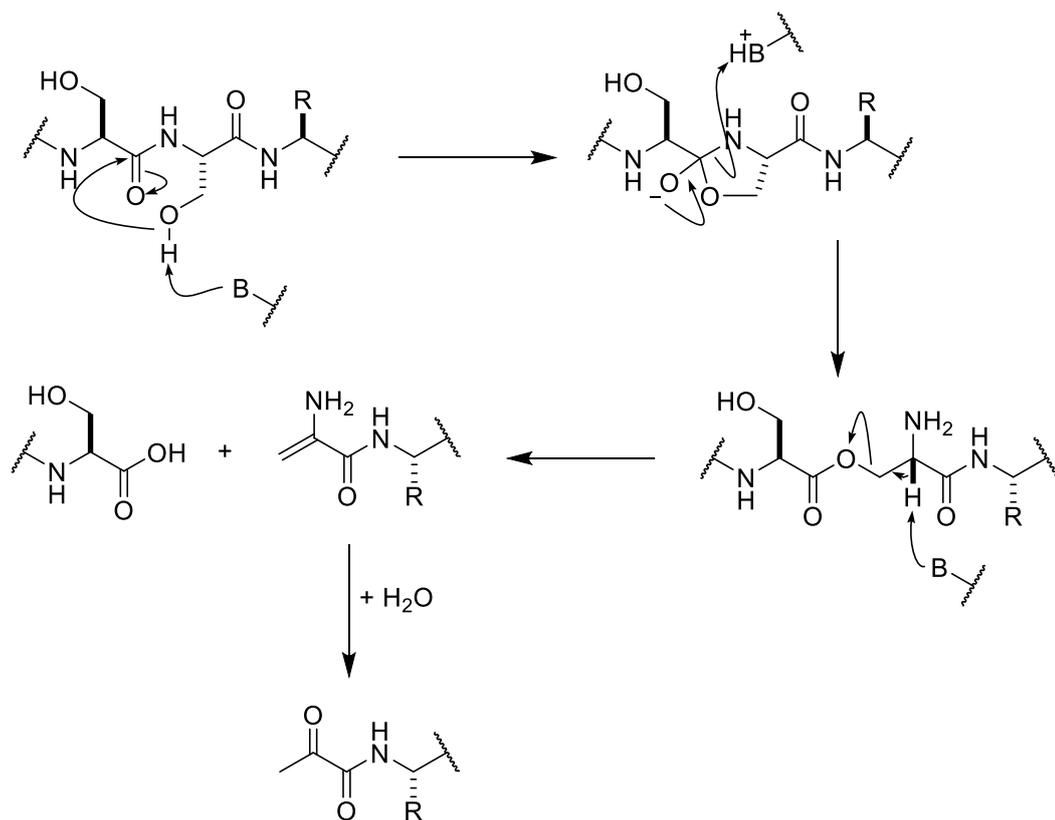
Scheme 8 | Proposed formation of α-ketoamide functionality of histidine decarboxylase, initiated by an intramolecular nucleophilic attack by a serine residue. Scheme adapted from [27].

Histidine decarboxylase catalyses the breakdown of histidine to histamine and carbon dioxide. Scheme 9 shows the proposed mechanism for the cleavage of the carboxylic acid of histidine.[27] The mechanism begins with the amine of histidine attacking the ketone to generate an iminium intermediate (steps 1 and 2). In step 3, the carboxylic acid of histidine is cleaved, releasing a molecule of CO_2. This is facilitated by the unnatural α-ketoamide functionality which acts as an electron sink, as well as the proton on the glutamic acid at position 197 (numbered according to *Lactobacillus 30a* histidine decarboxylase) providing electrostatic destabilisation. In steps 4-6, hydrolysis of the covalent intermediates leads to release of the product.[27,32]
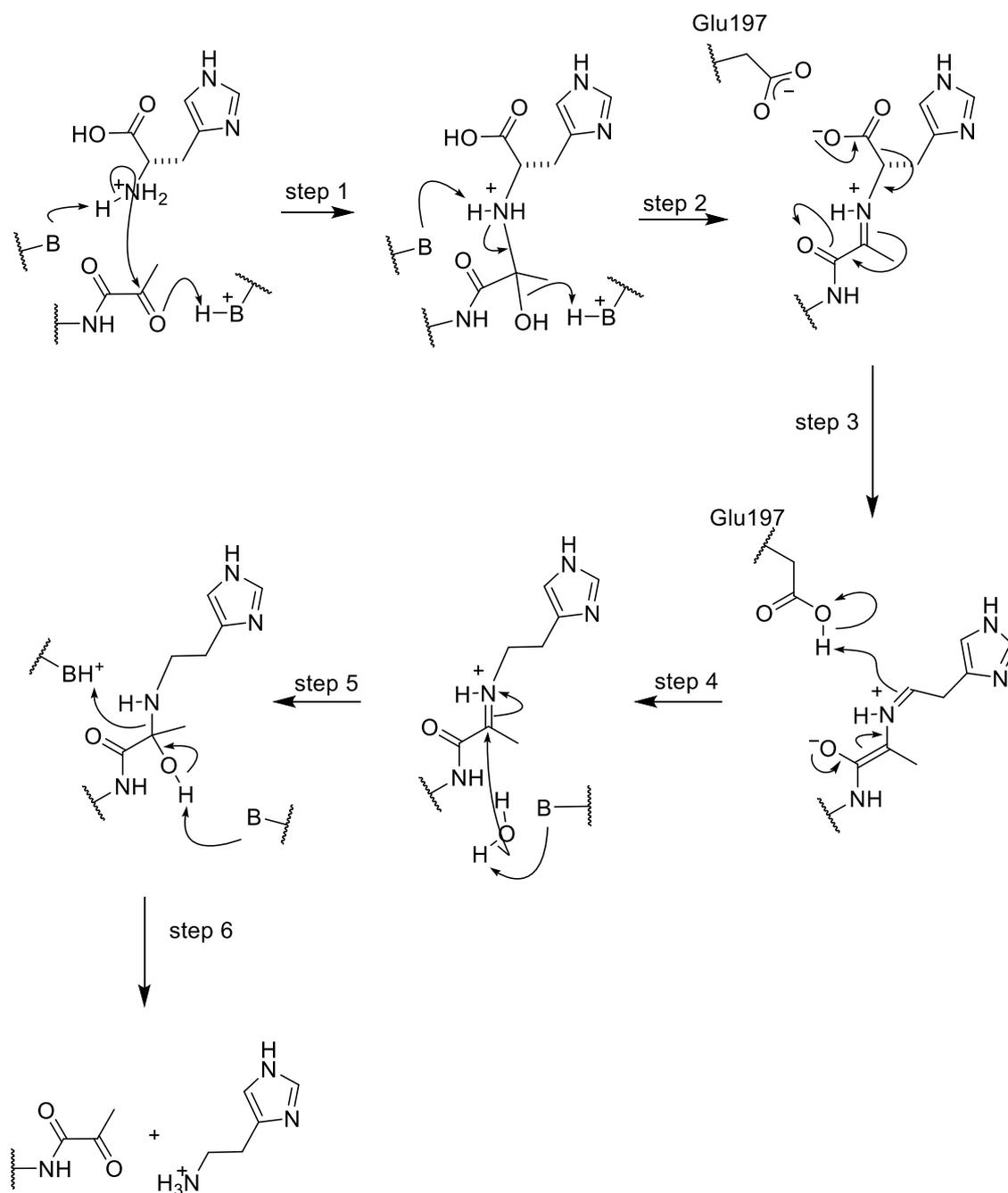
Scheme 9 | Proposed histidine decarboxylase mechanism, adapted from [27,33]. The post-translational α-ketoamide acts as an electron sink which facilitates the decarboxylation of histidine. (numbered according to *Lactobacillus 30a* histidine decarboxylase).

Both histidine ammonia-lyase and histidine decarboxylase utilise unnatural amino acids to catalyse reactions by acting as an electron sink. Such enzymatic mechanisms would not be possible without utilising cofactors or post-translational modifications.

1.3 Preparation of proteins containing unnatural amino acids

Chemists and biologists have developed many varied techniques to introduce new functionality into proteins with the goal of probing or altering function. The techniques span both chemical reactions with amino acid side chains and biological techniques that expand the genetic code. Often, the use of both pre- and post-translational modifications are necessary, for example, incorporating an amino acid into the genetic code prior to translation and then chemically modifying the incorporated amino acid side chain.[34]

This section will describe several methods, both chemical and biological, available to incorporate unnatural amino acids into proteins. Two commonly used methods of protein modification omitted from this section are native chemical ligation [35,36] and also solid phase peptide synthesis [37], as generally these techniques are not exploited for the modification of a single residue. Chemical ligation is largely exploited for protein-protein coupling,[38,39] and solid phase peptide synthesis is exploited for the synthesis of relatively short peptide chains.[37] These two techniques have been widely reviewed elsewhere.[36,37]

1.3.1 Schultz's methodology for the incorporation of unnatural amino acids

Over the past two decades Schultz and co-workers have developed methods to expand the genetic code. The method involves manipulating one of three stop codons (usually amber) into coding for the incorporation of an unnatural amino acid.[40-43] Inspiration for this may have come from the rare 21st and 22nd amino acids, selenocysteine and pyrrolysine, found in natural species and coded by the two stop codons amber and opal respectively.[44-46] The method relies on creating a 'codon-tRNA pair' with a corresponding tRNA synthetase which, during the process of translation, will insert an unnatural amino acid when confronted with an amber stop codon.[40,47] Crucially, the new tRNA must not act as a substrate for any other aminoacyl tRNA synthetase. The synthetase which is compatible with the unnatural amino acid must also be very selective. Many novel amino acids have been incorporated in this way, especially substituted phenylalanine analogues. Figure 2 shows a few examples of unnatural amino acids which have been incorporated in this fashion. This method of incorporating unnatural amino acids is most frequently used for aromatic residues, though other unnatural residues have also been incorporated.[48]
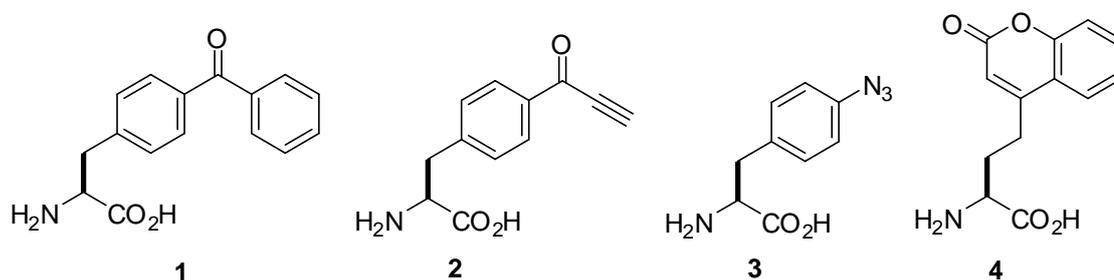
Figure 2 | Examples of unnatural amino acids which have been successfully added to the genetic code.[40,47,49,50] A broad range of aromatic amino acids has been incorporated into peptides using this technique.[48,51,52]

1.3.2 Chemical modification of nucleophilic amino acid residues

Analogous to post-translational modifications, chemists have been able to develop methodologies to allow the chemical modification of specific amino acid side chains within a protein.[53,54] Several factors need to be taken into account when performing chemical modifications of proteins. The reactions must be selective for a specific residue, and reactions must be capable of proceeding in a buffered aqueous environment at relatively mild pH and temperature.[55]

Nucleophilic residues such as lysine and cysteine can attack electrophiles directly to produce unnatural amino acids.[53,56,57] As the modifications are made after the protein has been expressed, the selectivity depends on the number and accessibility of nucleophilic residues. Residues on the surface of proteins are more readily available for chemical reactions than those that are buried within the folded protein. Cysteine residues are often utilised due to their relatively low occurrence in proteins and also their high nucleophilicity.[34] A wide array of modifications has been implemented with the aim of altering stability or reactivity, or to probe function. Again, the incorporation of the unnatural side chains relies on the tolerance of the protein for the incorporated functionality (Scheme 10 and Scheme 11).[58,59] Examples of the chemical modification of other amino acid side chains include the reactions of tyrosine residues with iminium ions,[60] cyclisation of arginine residues by reaction with 2-oxopropanal, and reaction of histidine residues with epoxides.[53]
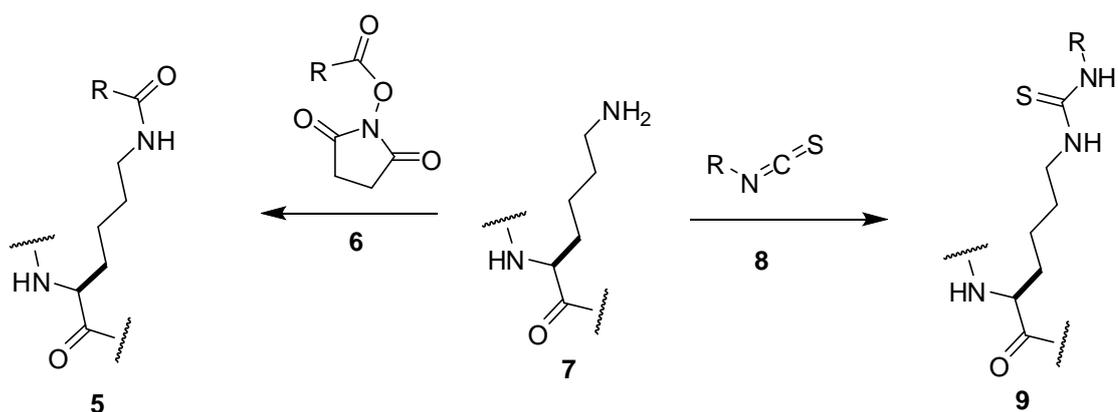
Scheme 10 | Chemical modifications of lysine residues.[59,61] Residue **5** results from nucleophilic attack onto the *N*-hydroxysuccinimide ester **6** and the residue **9** results from nucleophilic attack onto the isothiocyanate **8.**
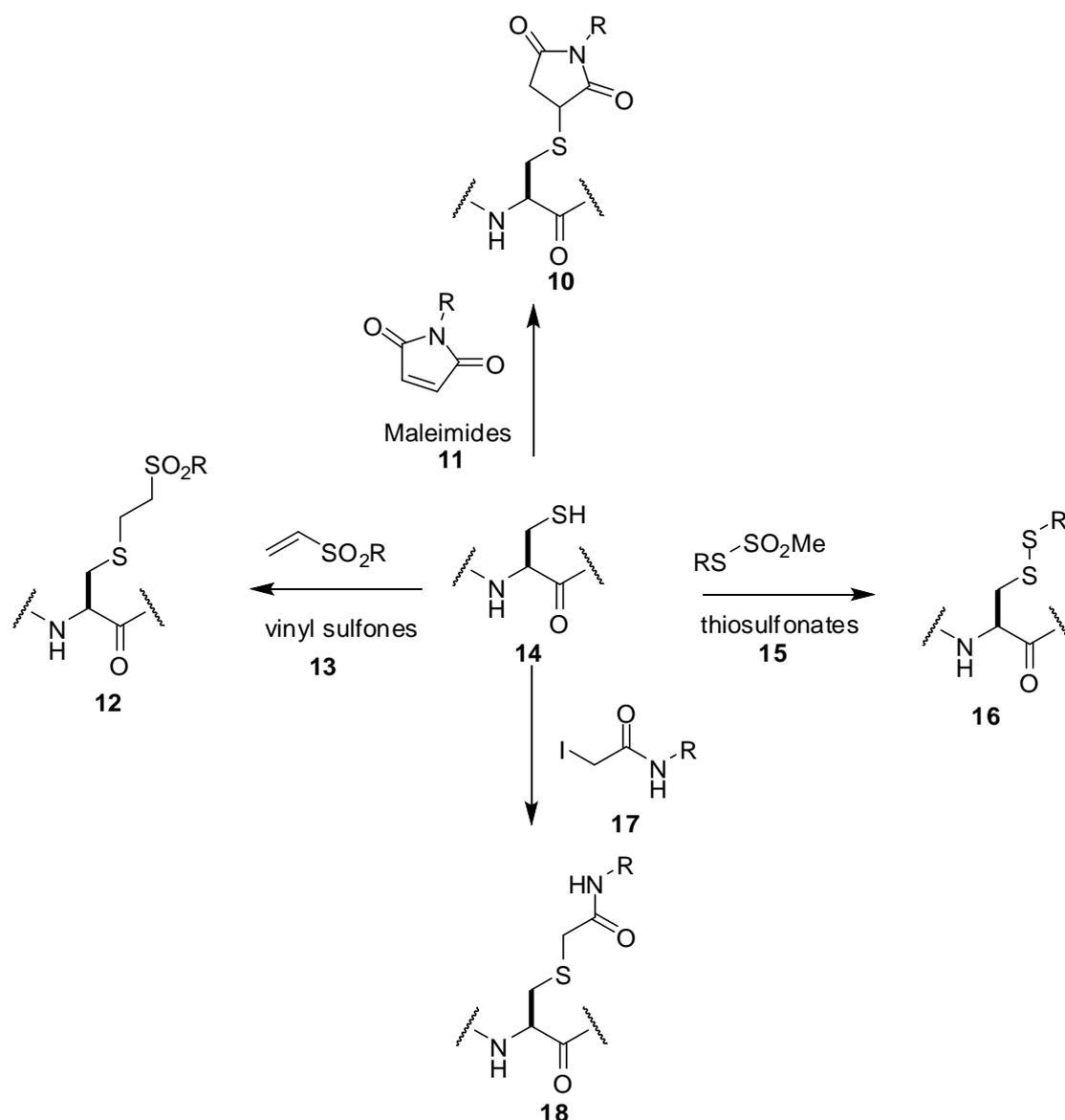
Scheme 11 | Chemical modifications of cysteine residues.[59,61,62] Examples shown are the nucleophilic attack onto maleimides, thiosulfonates, vinyl sulfones and alkylhalides.

There are also several examples of metal-mediated chemical modifications of nucleophilic amino acid residues. This includes the iridium-mediated reductive aminations of lysine, π-allyl palladium-mediated attack of allylic acetates by tyrosine, and rhodium-mediated carbene reactions with tryptophan. Metal-mediated methods for the chemical modification of amino acid side chains have been reviewed elsewhere.[53]

1.3.3 Chemical modification to yield dehydroalanine residues, and subsequent reaction

Dehydroalanine residues are known to form naturally within enzymes, such as bacterial phosphothreonine lyases *via* the post-translational oxidative elimination of cysteine or serine.[63,64] With the ability to act as a Michael acceptor, the dehydroalanine functionality can serve as an effective intermediate in the synthesis of unnatural amino acids.[65] Selective formation of dehydroalanine could therefore be a useful synthetic tool for the incorporation of unnatural amino acids (Scheme 12).
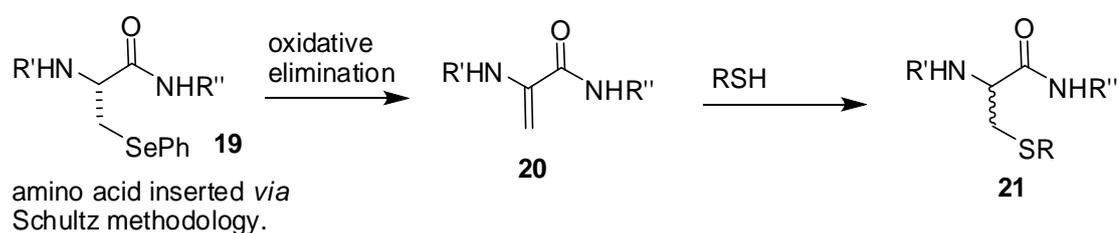
Scheme 12 | Michael addition of a thiol, *via* dehydroalanine formed from phenylselenocysteine.[65] In this example, phenylselenocysteine has been incorporated into the genetic code using Schultz's methodologies.

One method involves the insertion of a phenylselenocysteine residue followed by subsequent chemical modification (Scheme 12). After introduction of the unnatural residue **19**, selective elimination can yield the dehydroalanine residue **20**. Michael addition of a nucleophilic thiol can then result in the incorporation of an unnatural amino acid. A disadvantage to this method of chemical modification is the possibility of poor stereocontrol upon Michael addition of the thiol.[58,63] There have been several varied examples of unnatural thioether-containing amino acid side chains formed from a dehydroalanine intermediate. These modified side chains include alkene chains, amines, sugars, and phosphorus-containing species (Scheme 13).
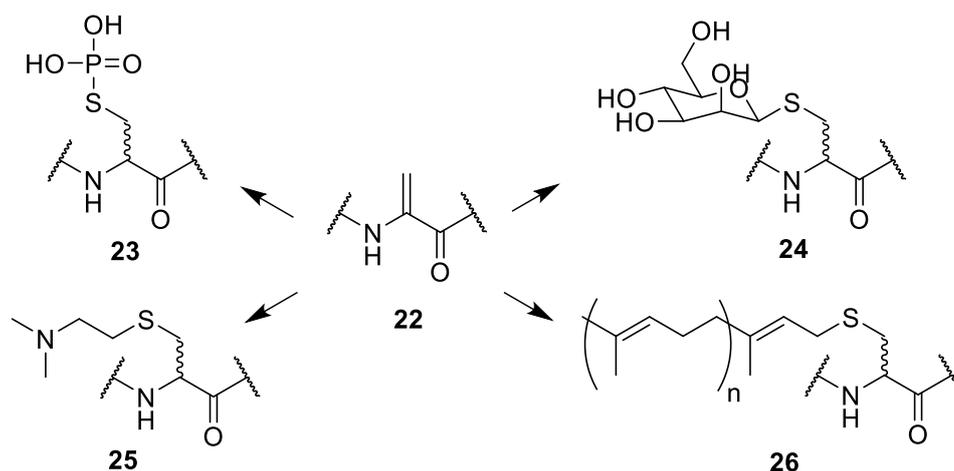
Scheme 13 | Examples of thioether-containing amino acids introduced through the Michael addition of a thiol to dehydroalanine.[55,63,65,66]

Alternatively, cysteine residues may be reacted with the dibromide **28** to produce a dehydroalanine residue (**30**) (Scheme 14).[67] Again, the resulting dehydroalanine can then be attacked by a nucleophilic thiol in a Michael addition, generating an unnatural amino acid. Providing there is only one cysteine available to react, the formation of dehydroalanine can be chemoselective.
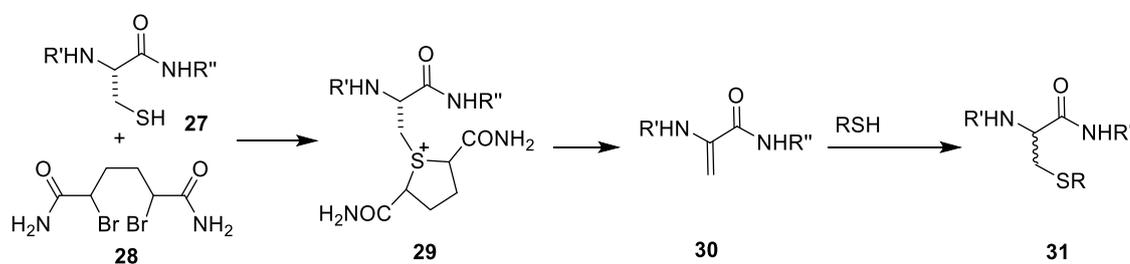
Scheme 14 | Utilising the dibromide 28 to produce dehydroalanine.[67] After the dehydroalanine residue **30** has been produced, subsequent reaction with a thiol can lead to the unnatural amino acid **31**.

18

1.4 Directed evolution of NAL

Sialic acids are a family of sugars characterised by a C-1 carboxylic acid and C-4 nitrogen substituent. *N*-Acetylneuraminic acid **49** is the most abundant sialic acid (Figure 3) and, as a result, *N*-acetylneuraminic acid is often referred to as sialic acid.[68] Sialic acids can be found in many higher organisms, often as terminal residues of glycan chains, and are involved in many biological functions, such as cellular and molecular recognition.[68] A key example of this is the involvement of sialic acids in the recognition of the influenza virus by cells.[69]
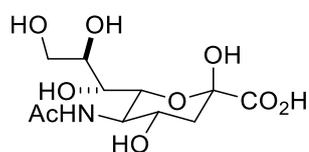
Figure 3 | *N*-Acetylneuraminic acid, the most common sialic acid.[68]

E. coli N-acetylneuraminic acid lyase (NAL) is a class I aldolase that catalyses the reversible condensation of pyruvate with *N*-acetylmannosamine (ManNAc), yielding *N*-acetylneuraminic acid (Neu5Ac) (Section 1.1.3).[14,16] NAL has a high degree of substrate specificity with each of the four alcohols of ManNAc involved in key hydrogen bonding interactions with the enzyme. As a result of the very specific hydrogen bonding between ManNAc and NAL, the enzyme is only capable of accepting a narrow range of five and six carbon ManNAc derivatives. NAL is unable to catalyse the condensation of other more varied substrates efficiently.[16,70]

1.4.1 Broadening the substrate specificity of NAL

The carboxamide **32** (Figure 4) has been shown to inhibit influenza sialidases.[16,71] Wild-type NAL is a poor catalyst of the aldol reaction for analogues of *N*-acetylneuraminic acid with the general structure **33**, which are possible precursors of the carboxamide **32**. Gavin Williams (University of Leeds) exploited directed evolution to broaden the substrate specificity of *E. coli* NAL in an attempt to allow the enzyme to accept molecules with the general structure **33** as substrates.[72]
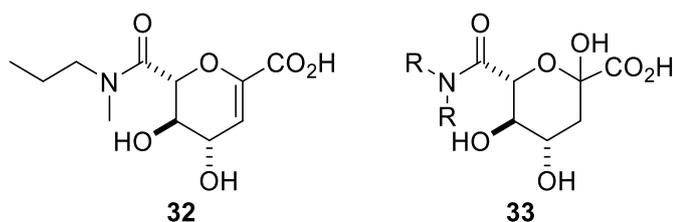
Figure 4 | An inhibitor of influenza A sialidase 32 and *N*-acetylneuraminic acid analogues 33.[14,72-75] Wild type NAL is a poor catalyst for the synthesis of molecules such as **33**, a possible precursor in the synthesis of **32**.[72]

With the knowledge of the active site in mind, broadening of the substrate specificity of *N*-acetylneuraminic acid lyase involved saturation mutagenesis of the Glu192, Asp191 and Ser208 residues.[76] After screening the resulting library of NAL variants, those altered at position 192 were found to accept several compounds of the general structure **33** as substrates more readily than wild-type *E. coli* NAL. This study led to the discovery of variant E192N, capable of accepting substrates of the general structure **33**.[14] The variant E192N was particularly efficient at catalysing the aldol reaction of the *N*-acetylneuraminic acid analogue (5*R*,6*R*)-7-(Dipropylamino)-4,5,6-trihydroxy-2,7-dioxoheptanoic acid (**DPAH**), which is a compound with general structure **33**. E192N was capable of performing the retroaldol reaction on **DPAH** ~50 times more efficiently than wild type *E.coli* NAL (Scheme 15). The structure of an analogue of DHOB, the precursor to **DPAH**, in complex with the variant E192N has been determined crystallographically.[16] The structural basis of the ability of the mutant to promote the reaction has been determined through crystal structure analysis of the *E. coli* NAL variant E192N in complex with a DHOB analogue. The crystal structure suggests that the propyl chains of the DHOB analogue orientate themselves into a hydrophobic pocket of the active site (see section 3.3 for more detail).[16] These interactions enable the recognition of the unnatural substrates and subsequent catalysis of the reversible aldol condensation.
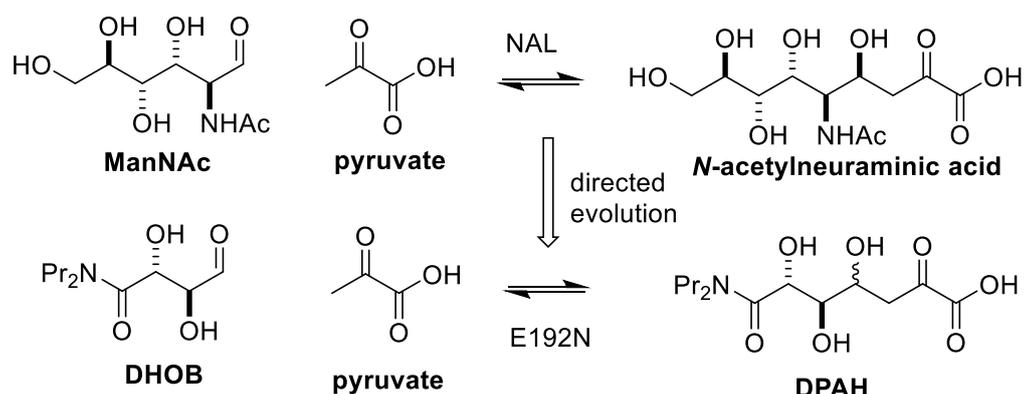
Scheme 15 | Directed evolution of NAL resulting in the discovery of the E192N variant. [71] This variant is capable of accepting and performing the reversible aldol reaction with a broad range of amide analogues of *N*-acetylneuraminic acid. The compounds **ManNAc**, ***N*-acetylneuraminic acid** and **DPAH** have been drawn in the open chain forms for clarity.

1.4.2 Engineering of the stereocontrol of an NAL variant

Though the E192N variant of NAL demonstrated broadened substrate specificity, poor stereocontrol was observed in the formation of the product (Scheme 16) limiting its utility in organic synthesis. Directed evolution was further utilised with NAL to engineer NAL variants that would act as complementary catalysts for the synthesis of two diastereomeric products. Using the E192N variant of NAL as a starting point, two further variants of NAL were developed for the synthesis of the two diastereomeric products.[77]
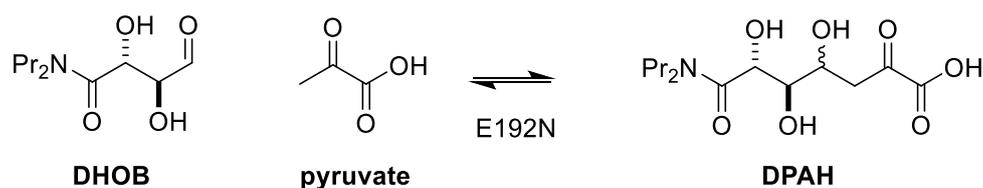
Scheme 16| E192N exerts poor stereocontrol in the formation of the substrate DPAH.

The two variants of NAL produced through the directed evolution of E192N NAL were (E192N/T167G) and (E192N/T167V/S208V). These variants provided complementary pairs which had favourable selectivity of roughly 40-fold for each of the diastereomers (Scheme 17).[77]
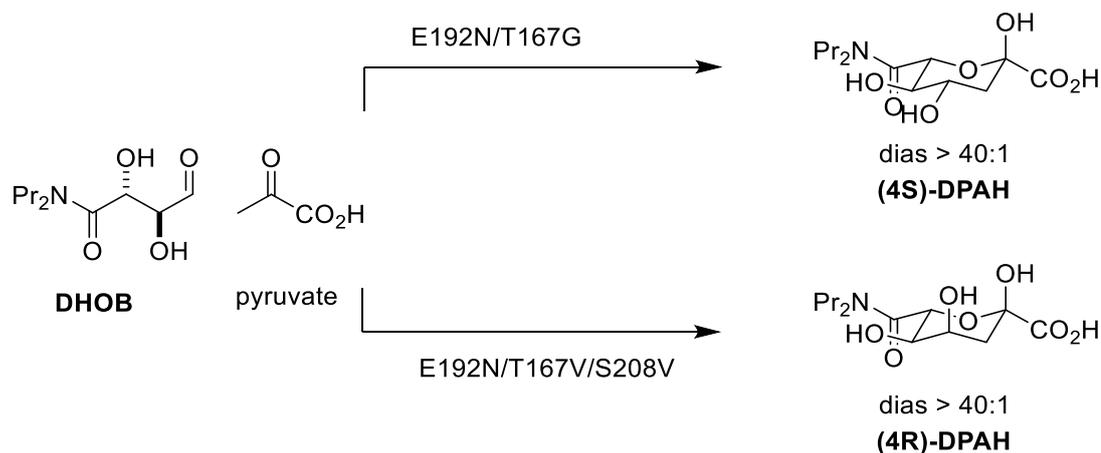
Scheme 17 | Complementary 4*S*- and 4*R*- selective enzyme variants of *E. coli* NAL. These variants are capable of performing the aldol condensation stereoselectively with > 40:1 diastereoselectivity.[77]

1.4.3 Development of a method for the incorporation of unnatural amino acids into the active site of *Sa*NAL

Wild type *S. aureus* NAL contained no natural cysteine residues, so selective incorporation of unnatural amino acids could be achieved by first incorporating a cysteine through site-directed mutagenesis (step 1, Scheme 18). Recently, chemical modification of a cysteine residue has resulted in the incorporation of a thia-lysine analogue at the residue 165.[78] Initially a cysteine residue was incorporated into residue 165 of *Sa*NAL. This variant was then treated with the dibromide **28** to form the dehydroalanine residue **dha** (step 2, Scheme 18). Subsequent Michael addition of a thiol, cysteamine, resulted in the formation of the K165-γ-thialysine variant of *Sa*NAL (step 3, Scheme 18).
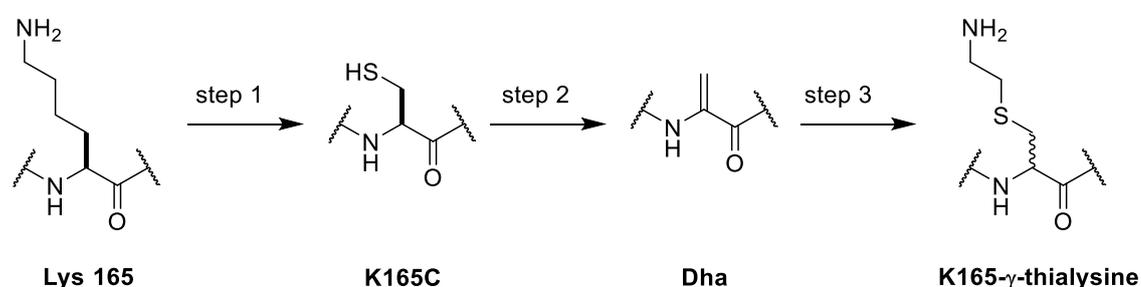
Scheme 18 | Modification of Lys165 *Sa*NAL to K165- γ-thialysine. Either epimer of K165-γ-thialysine may have formed. However, the crystal structure of the variant K165-γ-thialysine reveals only the amino acid with L-configuration.[78]

During the development of this methodology it was found that unfolding of the protein (facilitated by 6M urea) was necessary for the successful chemical modification of the active site residues. After chemical modification is complete the 6M urea is removed, allowing for the protein to refold (Scheme 19). The crystal structure of the SaNAL variant K165-γ-thialysine shows only the L-stereoisomer of the thialysine 165 residue (Figure 5). As both epimers of K165-γ-thialysine could result from the Michael addition of cysteamine, it is hypothesised that the R-stereoisomer was unable to refold after chemical modification. This may be a result of steric clash caused by the unnatural geometry of the incorporated side chain.
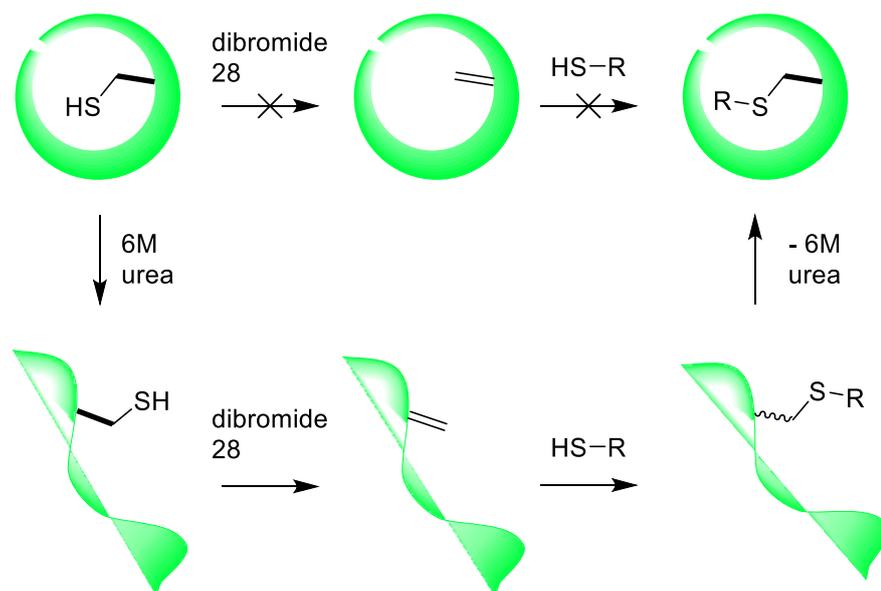
Scheme 19 | Schematic of the chemical modification process of K165C. This shows the necessity of unfolding the protein in order to access active site residues for chemical modification. Protein is represented in green.
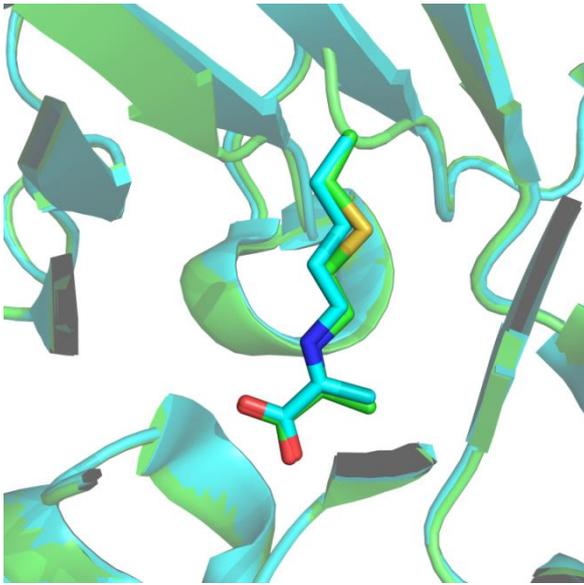
Figure 5 | Comparison of the catalytic residue 165 of wild type *Sa*NAL (blue) to the chemically modified variant, K165-γ-thialysine (green). Both crystal structures have pyruvate bound to residue 165. The two residues show a high degree of overlap, despite the thioether bond in the chemically modified variant K165-γ-thialysine. Only the L-stereochemistry was observed in the crystal structure of K165-γ-thialysine. This figure was generated using PyMOL (PDB file 4AH7).[78]

To gauge the efficiency of the modified protein, kinetic parameters were determined for the variant K165-γ-thialysine and the substrate *N*-acetylneuraminic acid. The unnatural protein was still found to catalyse the reverse aldol reaction of *N*-acetylneuraminic acid, though the efficiency of the protein had reduced to roughly 30% of the wild type *Sa*NAL protein.[78] It was also found that the optimal pH for the chemically modified protein (pH 6.8) was lower than the wild type *Sa*NAL (pH 7.4).[78] This shift in optimum pH was a direct result of the introduction of the thio-ether functionality.

The preparation of this chemically modified protein required the establishment of a robust method for the chemical modification of active site *Sa*NAL residues. This method could be exploited to incorporate a wide range of chemically modified proteins into the active site of *Sa*NAL, which may lead to the broadening of the range of substrates accepted by *Sa*NAL.

1.4.4 Development of a high throughput screen to identify chemically modified NALs with novel activity

The activity of chemically modified *Sa*NAL variants is currently being investigated within the Berry and Nelson groups. This research has involved screening the activity of chemically modified variants of *Sa*NAL with readily available substrates. Several residues within the active site of *Sa*NAL have been selected and replaced with cysteine through site-directed mutagenesis. These variants have been chemically modified by treatment with the dibromide **28** and reacted with a range of readily-available thiols. Once the chemical modifications were complete, the chemically modified variants of *Sa*NAL were screened with a range of readily available substrates in which an end point assay (thiobarbituric acid assay) was used to detect product formation.[79,80] This is a relatively fast method to screen for novel activity and has been used to find the *Sa*NAL variant F190-DHPC (see section 3.1.4). This approach, however, is limited to carbohydrates with α-hydroxy or α-NHAc aldehydes.

1.5 Conclusion and aim of the project

Enzymes are incredibly efficient catalysts capable of dramatically catalysing reactions involving a small range of substrates. The reactivity of enzymes can be broadened further (beyond the canonical 20 amino acids) by the utilisation of cofactors. Taking inspiration from this, biologists and chemists have developed methods of tailoring the reactivity of proteins, enabling the catalysis of reactions that would not be possible with natural substrates.

The aim of this project was to use chemical modification of specific residues within the active site of *Sa*NAL to expand substrate specificity of the enzyme. Preparation of chemically modified variants will be performed as per section 2.2. This method is appropriate as it exploits the lack of cysteine residues within wild-type *Sa*NAL and has previously been employed in the chemical modification of the active site residue K165 (section 1.4.3). The catalytic properties of these chemically modified variant enzymes will be investigated with analogues of *N*-acetylneuraminic acid (see section 2.1 for substrate design and synthesis).

Major considerations for the project will include the selection of active site *Sa*NAL residues to target for the chemical modification . Thiols that will be used for the chemical modification of these residues must be considered and selected, and potential substrates will be designed for testing with the chemically modified *Sa*NAL variants in kinetic parameter assays.

This project may yield insight into how chemical modification of *Sa*NAL could expand the catalytic repertoire of the enzyme.

Results and Discussion: Substrate and protein preparation

2.1 General overview of substrate synthesis

This Chapter describes the synthetic routes to the four potential substrates (4*R*,5*R*,6*R*)-5-acetamido-4,6,7-trihydroxy-2-oxoheptanoic acid (**ATOA**), (5*R*,6*R*)-7-(dipropylamino)-4,5,6-trihydroxy-2,7-dioxoheptanoic acid (**DPAH**), the enantiomer of *N*-acetylneuraminic acid (***ent*-NANA**) and the enantiomer of (4*R*,5*R*,6*R*)-5-acetamido-4,6,7-trihydroxy-2-oxoheptanoic acid (***ent*-ATOA**) (Figure 6). All of these compounds contain a common substructure, which is shown in red in Figure 6. Despite the common substructure, it was decided to develop three alternative synthetic approaches, due to the diverse stereochemistry and functionality of the compounds. These compounds were chosen as they are analogues of *N*-acetylneuraminic acid, so contain a 1,3 alcohol ketone relationship which is necessary for an aldol reaction. The potential substrates (Figure 6) were synthesised in order to explore the reactivity of chemically modified *Sa*NAL variants by varying length of substrate, stereochemistry and functionality. More detailed reasoning for the selection of each substrate is discussed in sections 3.1.1 (**ATOA**), 3.1.6 (**ent-ATOA**), 3.2.4 (***ent*-NANA**) and 3.3 (**DPAH**).
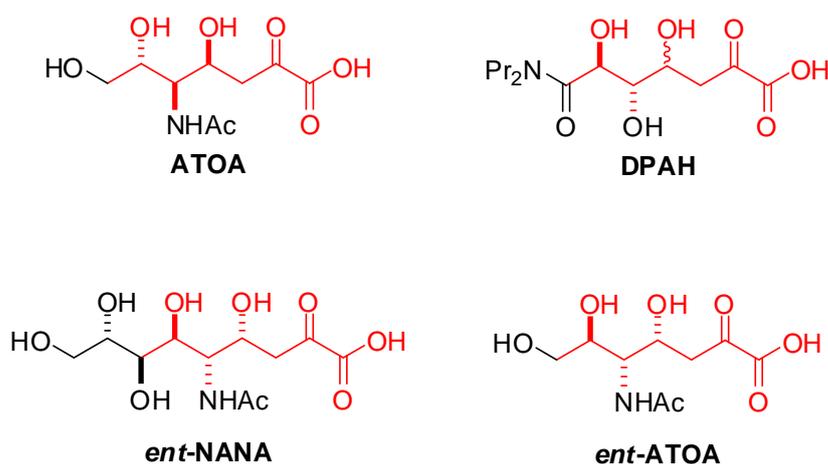
Figure 6 | Four potential substrates for NAL and NAL variants. The common substructure found in each compound is highlighted in red, note the configuration within the common substructure vary. For clarity the potential substrates are shown in the open chain forms.

Scheme 20 outlines the three alternative synthetic approaches to the potential substrates. Approach A would involve the truncation of the full length natural substrate, *N*-acetylneuraminic acid, to give the potential substrate **ATOA**. A beneficial aspect of this synthetic approach is that all three stereocentres of **ATOA** are already defined within the starting material, therefore removing the potentially problematic issue of stereocontrol.
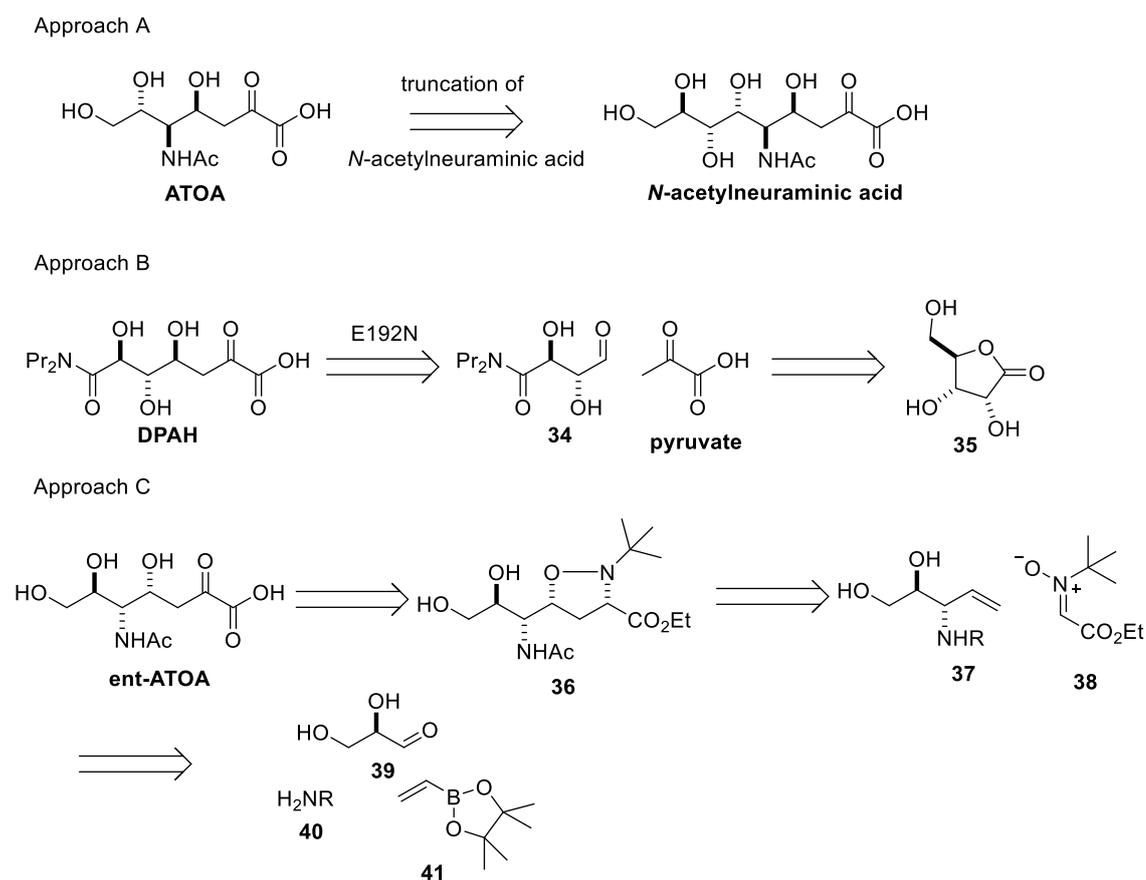
Scheme 20 | Three alternative approaches to the potential substrates. Approach A would involve the truncation of *N*-acetylneuraminic acid.[82] **Approach B** would utilise a chemoenzymatic strategy.[71] **Approach C** would involve a more direct chemical synthetic strategy.[81] The compounds **ATOA**, *N*-acetylneuraminic acid, **DPAH**, and **ent-ATOA** have been drawn in the open chain forms.

Approach B is a chemoenzymatic approach and would exploit an enzyme variant previously tailored by directed evolution to enable the synthesis of **DPAH** (Scheme 20). Initially the aldehyde **34** would be prepared from ribonolactone **35** and would contain most of the functionality of **DPAH**. The E192N variant of *E. coli* NAL (section 1.4.1) could then be utilised in conjunction with pyruvate to convert aldehyde **34** into the final product **DPAH**.

Finally, Approach C would exploit a more direct chemical synthetic strategy and would be employed for the synthesis of both *ent*-**NANA** and *ent*-**ATOA**. This approach would be based on three key steps. Initially a Petasis reaction would be performed involving an aldehyde (e.g. **39**), an amine **40** and a boronic ester **41** to give an alkene **37**. The boronic ester **41** would then be reacted with a nitrone **38** in a 1,3-dipolar cycloaddition to produce the compound **36**, a process which would result in the formation of two new stereocentres. Once compound **36** was synthesised, an elimination step would lead to the compound *ent*-**ATOA** (Scheme 20). An analogous approach could be used for *ent*-**NANA**. The success of this synthetic strategy would rely on the precedented high degree of stereocontrol.[81]

2.1.1 Synthesis of the truncated *N*-acetylneuraminic acid analogue, ATOA

Initially *N*-acetylneuraminic acid was treated with Dowex 50W H^+ ion exchange resin in methanol at 70 °C; after crystallisation, the protected sugar **42** was obtained in 50% yield. 500 MHz ^{1}H NMR spectroscopic analysis confirmed that only one anomer of compound **42** had been formed (Scheme 21). The compound **42** was treated with 2.5 equivalents of periodic acid in water, resulting in the formation of the hydrated aldehyde **43** which was purified by basic ion exchange column chromatography.
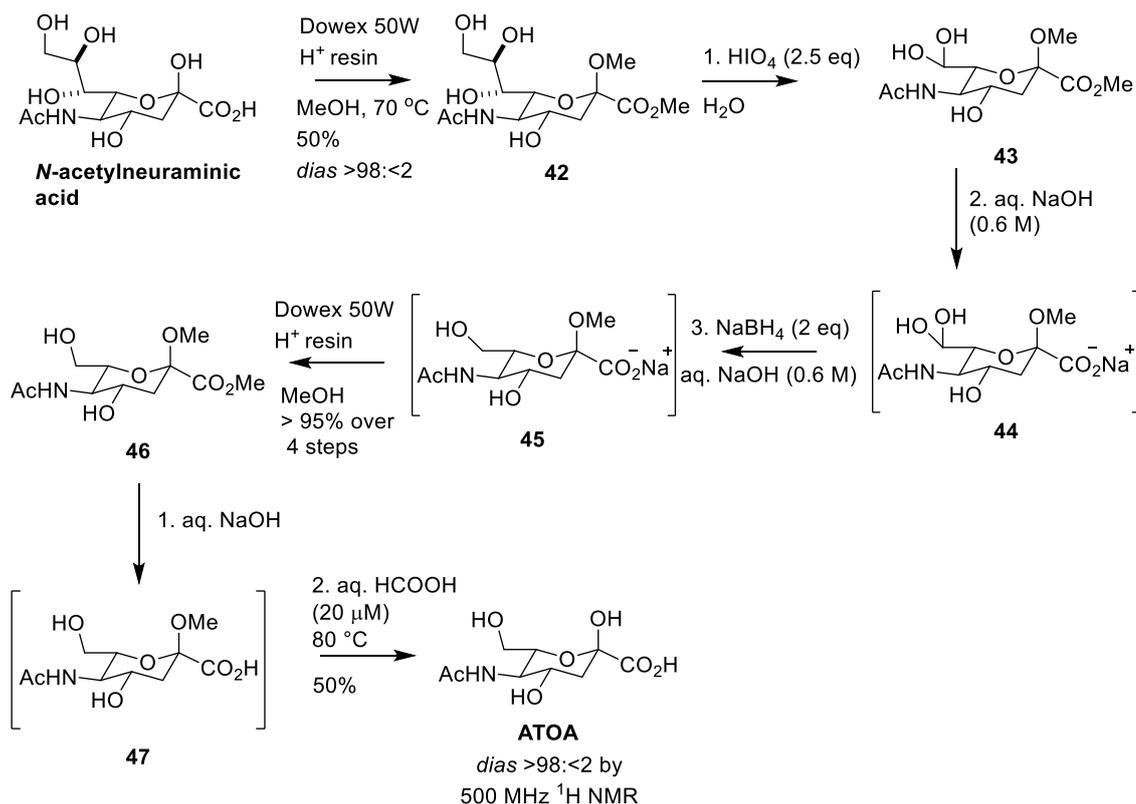
Scheme 21 | Synthesis of truncated *N*-acetylneuraminic acid analogue ATOA. The intermediates in parenthesis were not purified and characterised. Route adapted from patent [82].

Initial attempts at the reduction of the hydrated aldehyde **43** using 2.5 equivalents of sodium borohydride in methanol at 0 °C led to the over-reduced sugar **48** (Scheme 22). The over-reduction could be circumvented by the conversion of the methyl ester **43** into the less reactive carboxylic acid salt **44**. Thus the methyl ester **43** was suspended in 0.6 M aqueous sodium hydroxide until complete formation of the salt **44** was observed by LCMS. The carboxylate **44** was then treated with 2.5 equivalents of sodium borohydride to give compound **45** (which was observed by LCMS). For the purpose of purification, compound **45** was resuspended in methanol containing Dowex 50W H⁺ resin resulting in the formation of sugar **46** which was isolated *via* ion exchange chromatography. This four step process from protected *N*-acetylneuraminic acid **42** consistently gave the truncated sugar **46** in >95 % yield.
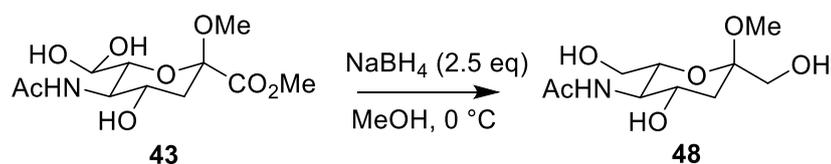
Scheme 22 | Over reduction of the sugar 43 to the sugar 48. Reduction with NaBH$_4$ in methanol at 0 °C led to over reduction of sugar **43**.

The truncated sugar **46** was treated with aqueous sodium hydroxide to give the intermediate **47** (observed by LCMS). After neutralisation with Dowex 50W H$^+$ ion exchange resin and concentration, the carboxylic acid intermediate **47** was treated with aqueous formic acid at 80 °C to provide **ATOA** in 50% yield over two steps after reverse phase column chromatography. The overall yield of the synthesis of **ATOA** was 25% from *N*-acetylneuraminic acid.

2.1.2 Synthesis of the substrate DPAH

The substrate **DPAH** was prepared by the chemoenzymatic approach B (Section 2.1, Scheme 20). The synthesis began with the treatment of the ribonolactone **35** with acetone and conc. H$_2$SO$_4$ to give the acetonide **49** in 70% yield (Scheme 23). Treatment of the alcohol **49** with iodine, triphenylphosphine and imidazole led to the formation of the unstable iodolactone **50** in 68% yield. The iodolactone **50** was treated with zinc/copper couple in refluxing 4:1 acetone– water; the carboxylic acid **51**, which was used without any purification, was obtained in 77% yield. The carboxylic acid **51** was treated with EDC, HOBt and dipropylamine to form the amide **52** in 69% yield. The amide **52** was then subjected to 9:1 TFA–water to give the diol **53** in 92% yield.
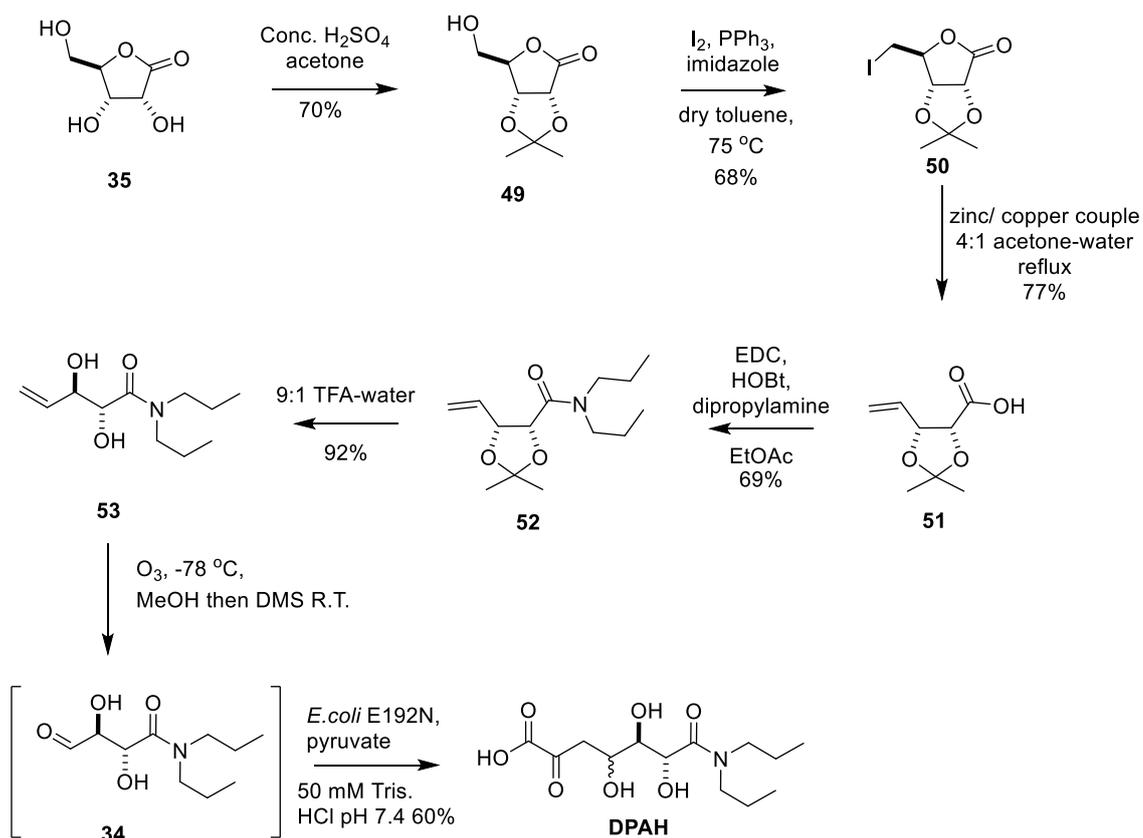
Scheme 23 | Chemoenzymatic approach to the synthesis of the substrate DPAH.[71]

The alkene **53** was treated with ozone at −78 °C until the ozonolysis was deemed to be complete (reaction mixture had turned blue), at which point dimethyl sulfide was added to reduce the intermediate ozonide. The intermediate aldehyde **34** was concentrated in *vacuo* and used without any further purification. The final steps in the synthesis of **DPAH** involved the resuspension of the intermediate aldehyde **34** in 50 mM Tris HCl buffer at pH 7.4; the intermediate aldehyde **34** was then treated with a five-fold excess of sodium pyruvate (in 50 mM Tris HCl buffer at pH 7.4) and the enzyme E192N (in 50 mM Tris HCl buffer at pH 7.4). After 24 hours the reaction was stopped by treatment with 2M aqueous formic acid and 2M aqueous ammonium hydroxide, followed by decomposition of the excess pyruvate *via* treatment with bakers' yeast. **DPAH** was isolated from the aldolase-catalysed step in 60% yield after purification by ion exchange chromatography. The distereoselectivity of the aldolase-catalysed step was consistent with that previously reported (~ 75:25).[72,76,83]

2.1.3 Synthesis of the enantiomers of *N*-acetylneuraminic acid and ATOA
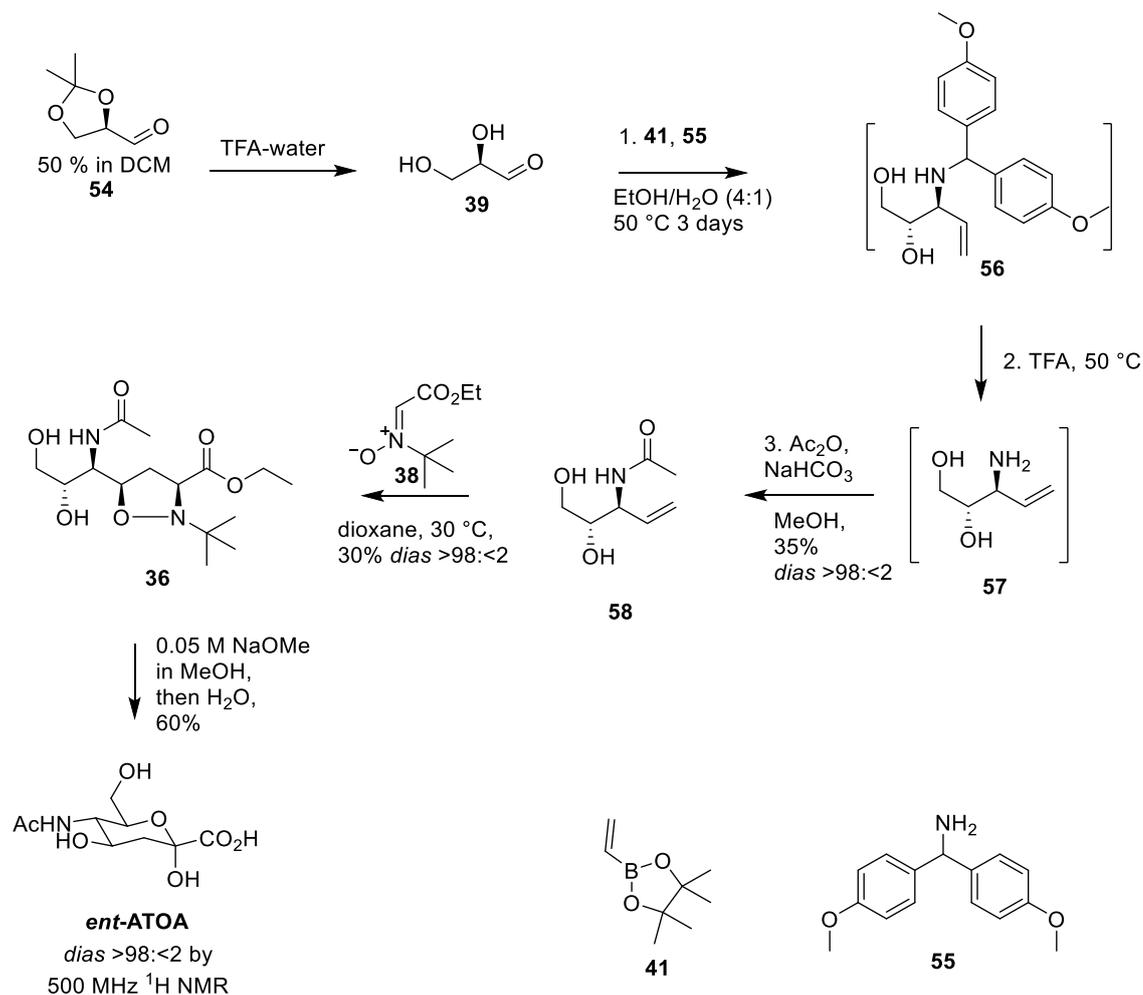
Scheme 24 | Synthetic route used for *ent*-ATOA. This synthesis involved a Petasis reaction, a 1,3 dipolar cycloaddition and a deprotection reaction.[81]

The synthesis of ***ent*-ATOA** (Scheme 24) began with the deprotection of the acetonide **54** by treatment with TFA in water, followed by immediate concentration of the reaction mixture to allow purification by reverse phase column chromatography to give a crude aldehyde **39** which was used without further purification (Scheme 24). The crude aldehyde **39** was treated with the vinyl boronic ester **41** and the amine **55** in 4:1 ethanol-water for 3 days at 50 °C to give the intermediate alkene **56** (which was detected by LCMS). TFA was then added to the reaction mixture and the solution was stirred at 50 °C until complete formation of the primary amine intermediate **57** was observed by LCMS; the solution was then reduced *in vacuo* and the primary amine **57** used without any further purification. The primary amine intermediate **57** was resuspended in methanol containing 2 equivalents of sodium bicarbonate and treated

with 1.1 equivalents of acetic anhydride for 1 h at room temperature. The amide **58** was obtained in 35% yield after purification. 500 MHz ^{1}H NMR spectroscopic analysis of the amide **58** showed that only one stereoisomer had been obtained. The stereoselectivity of the reaction may be rationalised in terms of coordination of the boronic ester **60** to the α-hydroxyl group of the imine intermediate **59** (Scheme 25, Newman projections **61** and **62**).
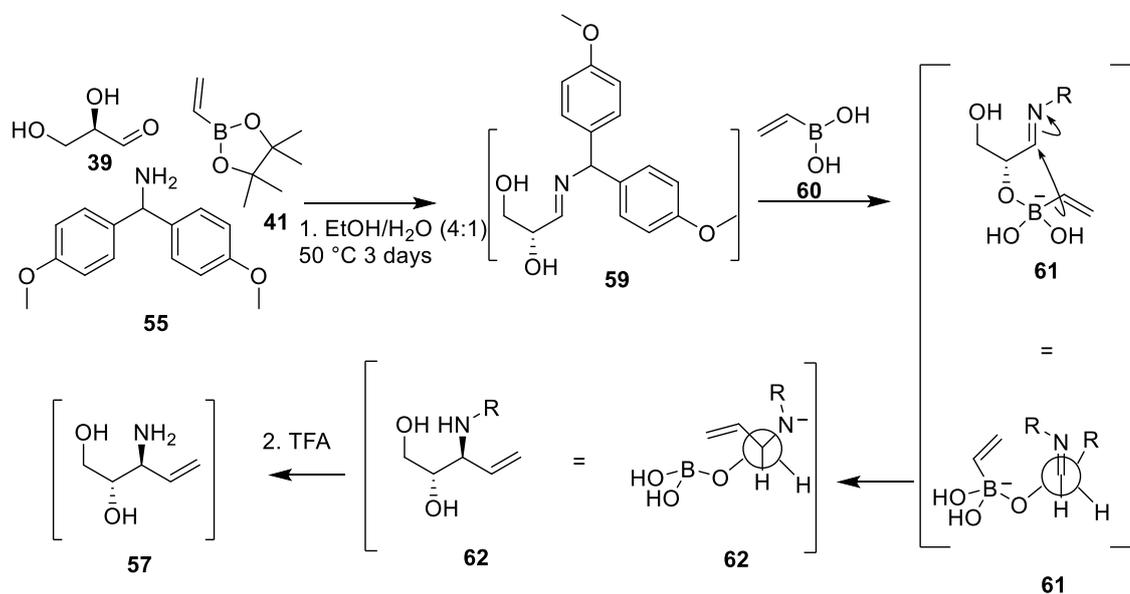
Scheme 25 | Theoretical rationalisation of the stereochemical outcome of the petasis reaction of the aldehyde 39, depicting the coordination of boronic ester to the α-hydroxyl of the imine intermediate **61**.

The alkene **58** was treated with the nitrone **38** in dioxane at 30 °C for two weeks, after which *iso*oxazolidine **36** was isolated in 30% yield as a single regio- and stereoisomer (judged by 500 MHz ^{1}H NMR spectroscopy). The stereochemical outcome can be rationalised through the minimisation of steric clash between the alkene substituent **58** and the nitrone **38** (transition state diagram in Scheme 26, transition state **63**).
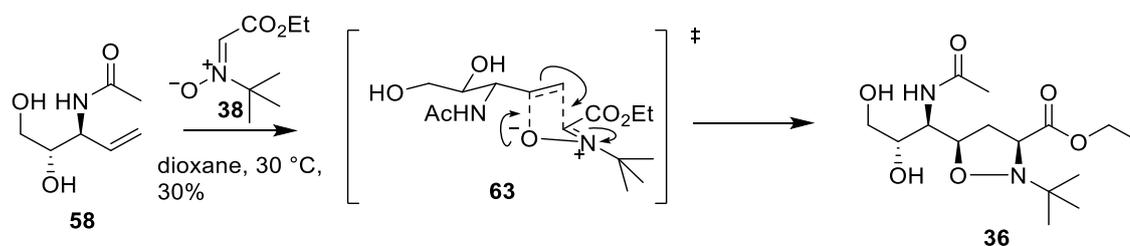
Scheme 26 | Proposed transition state for the 1,3-dipolar cyclisation of the alkene 58 and the nitrone 38.

Compound **36** was then treated with sodium methoxide in methanol, followed by the addition of water, resulting in the elimination to give ***ent*-ATOA** , isolated in 60% yield after purification *via* ion exchange chromatography eluting with 0–2M formic acid.

The synthesis of the compound ***ent*-NANA** was performed using the same strategy as the synthesis of the compound ***ent*-ATOA**, and can be seen in Scheme 27. The synthesis began with the compound **64** (L-arabinose) which, after Petasis reaction, deprotection and amide formation gave the alkene **65** in 50% yield; once again, only one stereoisomeric product was observed by 500MHz ^{1}H NMR spectroscopy. Subsequently, the 1,3-dipolar cycloaddition between the alkene **65** and the oxazole **38** gave the *iso*oxazolidine **66** in 55% yield. Treatment of the *iso*oxazolidine **66** with sodium methoxide in methanol, and then water, led to the compound ***ent*-NANA** as a single diastereomer in 55% yield after ion exchange chromatography.

Scheme 27 | Route used for the synthesis of *ent*-NANA.[81]

2.2 Preparation and characterisation of enzymes

This section describes the approach used to incorporate unnatural amino acids into the active site of *Sa*NAL. To illustrate the approach, the chemical modification of the E192C variant (Scheme 28) is described in detail. Methods for the expression, purification, chemical modification and characterisation of the modified enzymes are described. The process began with the glycerol stock solutions of XL10-Gold cells, containing the gene encoding either the E192C or F190C variant, and led to the production of chemically modified proteins. An example of this overall process – leading to the variant E192-DHPC – is shown in Scheme 28. The positions E192 and F190 were both chosen as modifications at these positions have been shown to broaden the substrate specificity of NAL (see sections 1.4.1 and 3.1.4 for more details on the position E192 and F190 respectively).
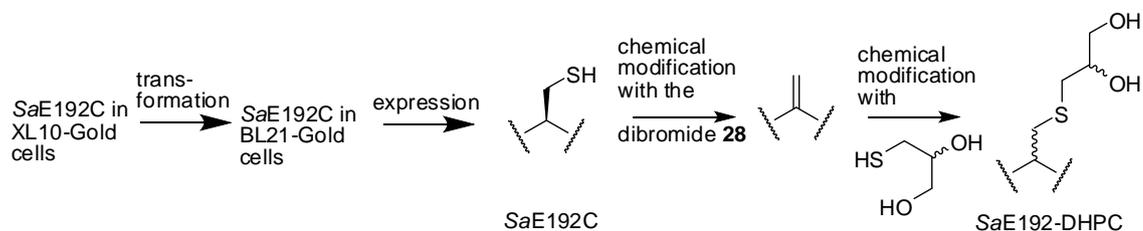
Scheme 28 | An overview of the integrated biological and chemical approach used to modify *Sa*NAL to yield a protein containing an unnatural amino acid. In this example the thiol, thioglycerol, was used to produce the E192-DHPC *Sa*NAL variant.

2.2.1 Expression and purification of the *Sa*NAL variants E192C and F190C

XL10-Gold cells containing the genes encoding either the E192C or F190C *Sa*NAL were provided by Nicole Timms and Claire Windle (University of Leeds). The plasmid DNA from the XL10 gold cells was isolated (Section 4.1.10) and analysed by DNA gel electrophoresis (Figure 7 and Section 4.1.9). Sequencing of the isolated DNA confirmed that the correct gene sequences were present. The DNA was then retransformed into BL21-Gold (DE3) cells and stored at -80 °C ready for protein expression.
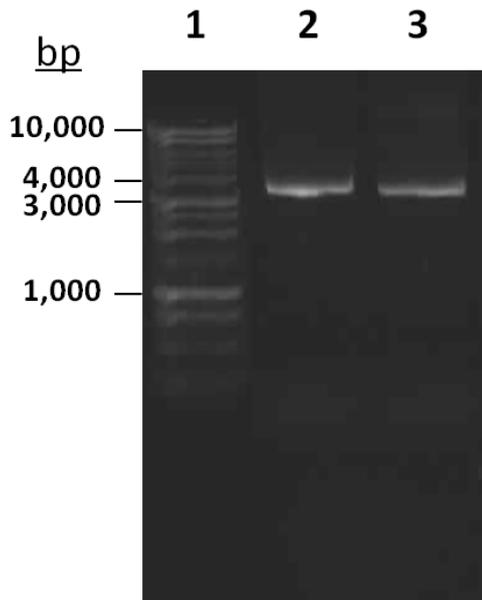
Figure 7 | An example agarose gel showing the plasmid pKK223-3 containing either the wild-type *Sa*NAL gene (lane 2), or the *Sa*NAL E192C gene (lane 3). Promega 1kb DNA ladder is shown in lane 1. DNA was isolated from XL10-Gold cells.

The E192C and F190C variants of *Sa*NAL were expressed on a 1-6 L scale using the BL21-gold (DE3) cells (Section 4.1.7). The expressed proteins, which contain hexahistidine tags, were purified using nickel affinity purification (Section 4.1.15). SDS-PAGE analysis (Section 4.1.14) of E192C at different stages of purification can be seen in Figure 8, and a single band at ~34 kDa in lane 9 of the SDS-PAGE gel indicates isolated purified E192C protein.
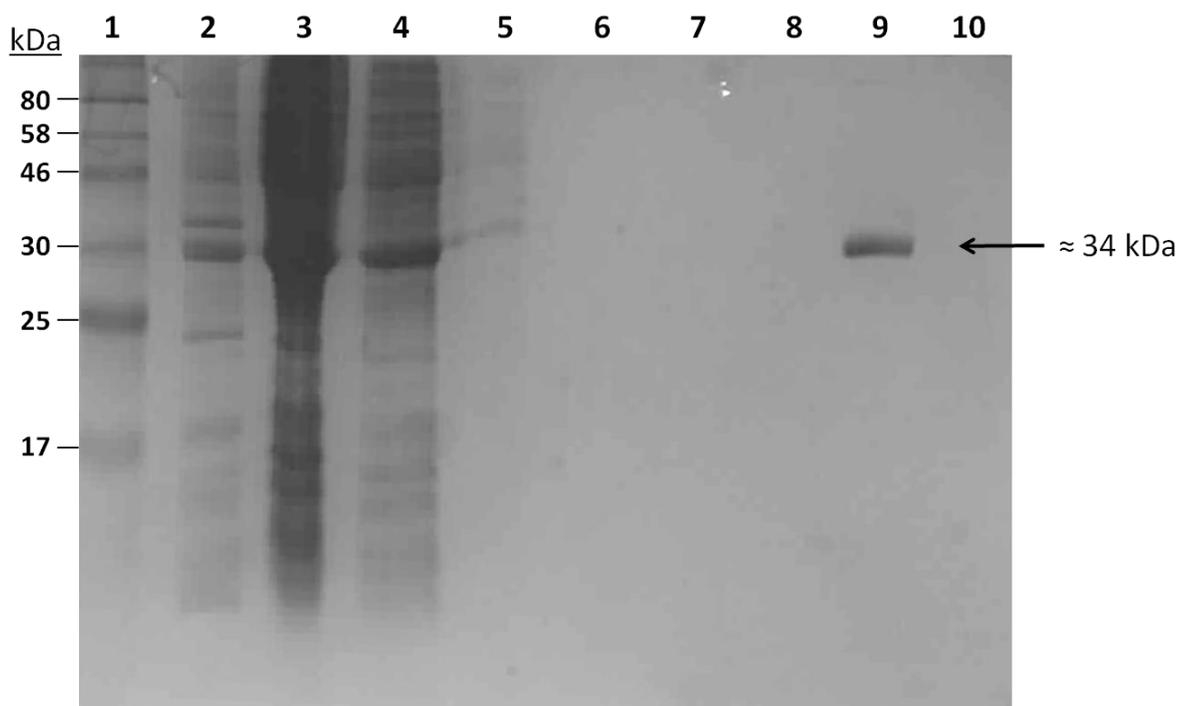
Figure 8 | An example SDS-PAGE of the stages of the nickel affinity purification of *Sa*NAL E192C protein. Lane 1) NEB prestained protein marker, broad range (7-175 kDa). 2) IPTG induced BL21-Gold (DE3) cells. 3) pellet after cell lysis. 4) supernatant after cell lysis. 5-8) Four tris buffer washes of the nickel resin after the His$_6$-tagged *Sa*NAL E192C protein had been bound. 9-10) Two elutions using 0.5 M imidazole in buffer, which facilitated the displacement of the *Sa*NAL E192C protein from the nickel resin. A single band, at approximately 34 kDa, is visible in lane 9 which corresponds to pure protein.

Following purification, protein samples were dialysed into 50 mM ammonium acetate buffer (Section 4.1.16), and the concentration measured (Section 4.1.17). The protein was aliquoted into 5-10 mg samples, which were then lyophilised. Lyophilised samples were characterised by mass spectrometry (Section 4.1.19). As an example, Figure 9 shows both the multiple mass-charge state spectrum for E192C, as well as the deconvoluted spectrum of the same sample. The multiple mass-charge state spectrum of E192C is the primary data obtained and shows the multiple charge states of the protein, as proteins have many sites that can be protonated. These charge states are related to the mass of the protein through the m/z (mass-to-charge) ratio.[84,84b] The peak with a charge state of 31+ has been highlighted in the multiple charge-state spectrum as an example (Figure 9). Deconvolution of the multiple mass-charge state spectrum was routinely performed by James Ault (University of Leeds) using software provided with the mass spectrometer equipment (Section 4.1.19).

The yield of E192C or F190C after purification was typically 50-70 mg per litre of expression culture. Lyophilised purified protein samples were stored at −20 °C until they were needed for chemical modification.
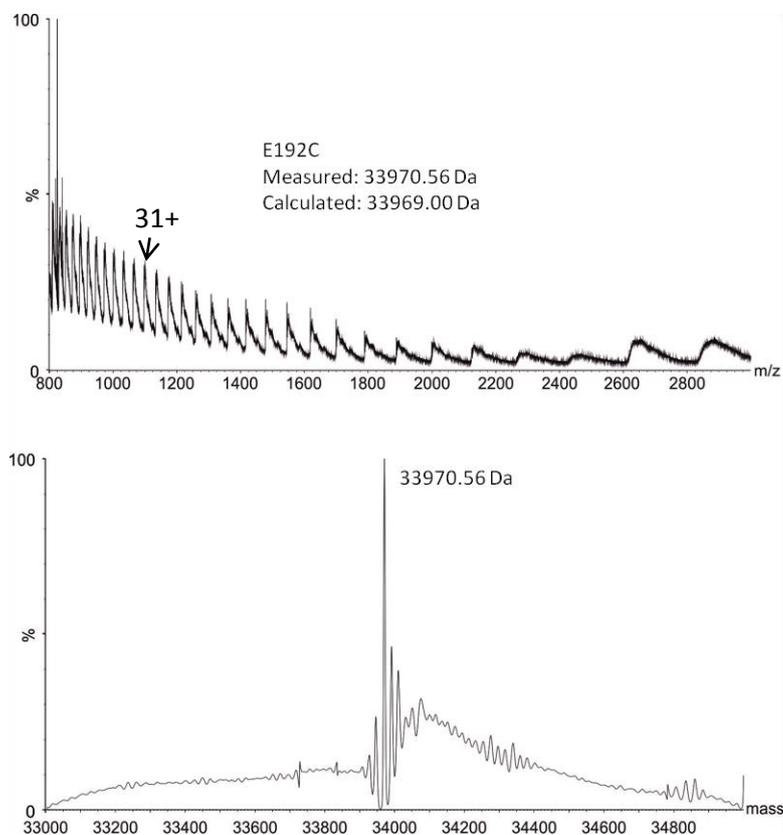
Figure 9 | ESI-MS spectra of *Sa*NAL E192C. (Top) showing mass charge states (Bottom) showing deconvoluted spectrum. Measured mass = 33970.559 Da. Calculated mass = 33969.00 Da.

2.2.2 Preparation of a chemically modified *Sa*NAL variant

The methodology used for the chemical modification of *Sa*NAL had already been established within the Berry group and had been successfully used to modify cysteine variant *Sa*NAL proteins. The chemical modification of either E192C or F190C began with the resuspension of the lyophilised protein (2 mg/mL) in 50 mM sodium phosphate buffer pH 8.0 containing 6 M urea. Urea-facilitated unfolding of the protein was necessary to allow accessibility of the active site cysteines to the chemical reagents. To a solution of unfolded protein, the dibromide compound **28** in DMF (12.8 mg in 97 µL per 2mg of protein) was added and the solution was

incubated for 1.5 h at 37 °C. At this point, a 70 µL sample of the reaction mixture could be buffer exchanged into 50 mM ammonium acetate (Section 4.1.19) and used for mass spectrometric analysis to allow for the observation of the completion of the reaction. The conversion was deemed to be complete when all the cysteine in the protein has been converted to dha. An observed decrease in mass of 34 Da would be consistent with the conversion of cysteine containing protein to dha containing protein.

Upon complete conversion of either E192C to E192dha, or F190C to F190dha, the protein was reacted with a thiol reagent. In this example, thioglycerol in 1.5 M Tris/HCl pH 8.8 (36 mg in 114 µL per 2 mg of protein), was added and the buffered solution containing E192dha was re-incubated for 2 h at 37 °C. At this point 70 µL of the protein reaction mixture could again be buffer exchanged (into 50 mM ammonium acetate) and used for mass spectrometric analysis. The resulting modified *Sa*NAL protein variant was then dialysed into sodium phosphate pH 8.0 buffer containing 6M urea, followed by dialysis into sodium phosphate buffer pH 8.0 and then two rounds of dialysis into 50 mM Tris HCl buffer pH 7.4 (section 4.1.23). Nicole Timms (University of Leeds) had previously demonstrated, using circular dichroism, that this procedure resulted in refolding of related chemically modified proteins at positions 165 of *Sa*NAL.[84b]

Finally the modified protein was concentrated to 5 mL (section 4.1.18) and purified by size exclusion chromatography (Section 4.1.24), a process in which the unfolded protein could be separated from the refolded modified protein (Figure 10). The protein collected after size exclusion chromatography was then concentrated (Section 4.1.18), and assessed using mass spectrometry (Figure 11).

Figure 10 | Size exclusion column trace of the purified *Sa*NAL E192-DHPC protein. The first peak represents unfolded protein and the second peak represents the refolded modified protein. The red markings show where fractions (2 mL per fraction) were collected.
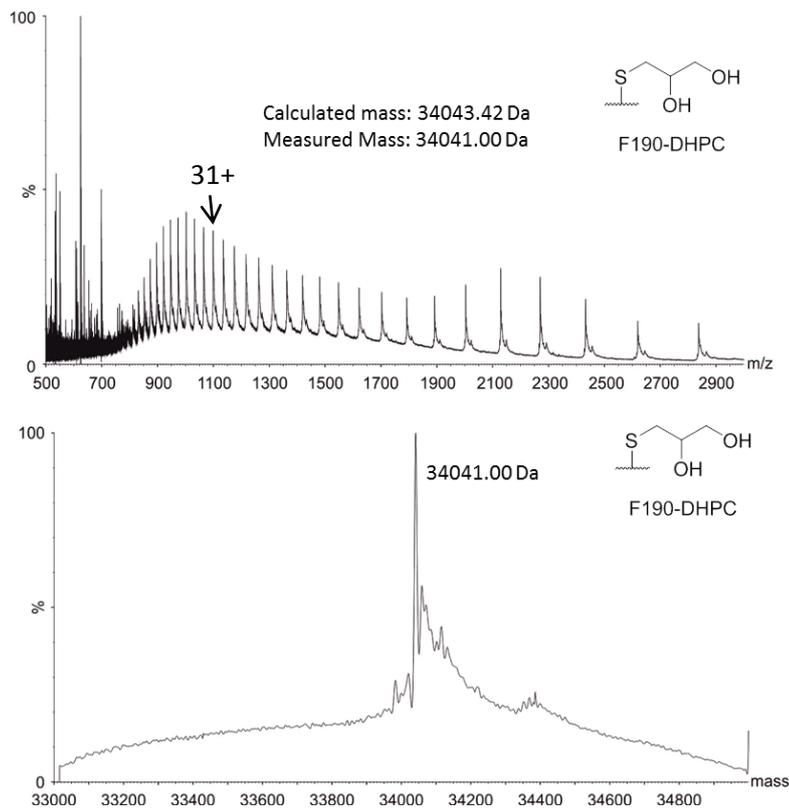
Figure 11 | ESI-MS spectra of *SaNAL* E192-DHPC (Top) showing mass charge state (Bottom) showing deconvoluted spectrum. Measured mass = 34041.00 Da. Calculated mass = 34043.42.
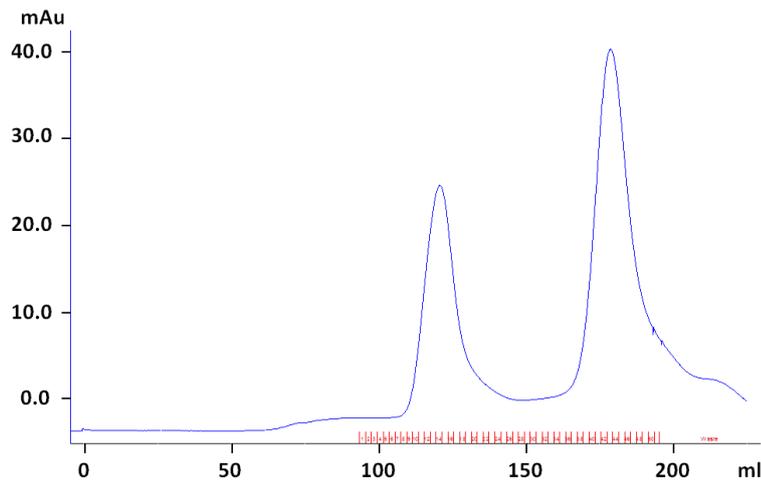
2.2.3 Scope of the chemical modification of *Sa*NAL

Twelve successful large scale (20-40 mg) modifications were performed at position E192C, and an additional four modifications were performed at the position F190C (see appendix I). Figure 12 shows the successful modifications performed at position E192C. The modifications cover a range of functional groups and include straight chain, branched and cyclic species.

Four modifications resulted in unnatural amino acids with carbonyl based functionality, E192-CMC (an analogue of glutamic acid), E192-MCMC, E192-ACEC and E192-AMCEC. Interestingly, E192-ACEC and E192-AMCEC made use of cysteine as the nucleophile. There were several modifications which resulted in amino acids with alcohol based functionality, such as E192-DHPC, E192-HPC (an analogue of threonine) and E192-HBC. The use of aminoethanethiol as the nucleophile resulted in the lysine analogue E192-AEC. Several of the modifications were hydrocarbon-based, resulting in non-polar unnatural side chains such as E192-PC, E192-CPC and E192-CHC. Finally E192-HEBC combines both a greasy hydrocarbon chain and a terminal alcohol side chain.
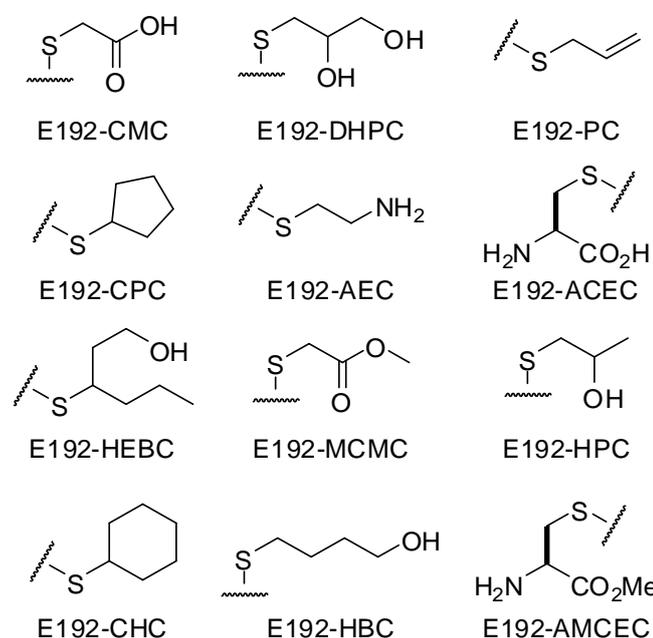
Figure 12 | Overview of the successful chemical modifications of *Sa*NAL at position E192C. Each structure represents the unnatural side chain incorporated through chemical modification into the position E192.

Not all of the modifications attempted at position E192C were successful. Attempting to modify E192C with compounds which had a thiol directly attached to an aromatic ring such the thiols **68** and **69** (Figure 13) did not result in any detectable product upon mass spectrometric analysis of the reaction mixture. The thiol **67** (Figure 13) represents an example of a modification which successfully incorporated into position E192C (detected by mass spectrometric analysis of the reaction mixture) though the protein continuously precipitated out of solution upon dialysis of the reaction mixture.
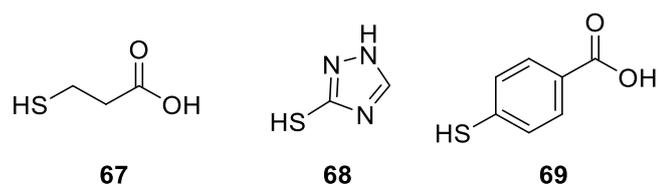
Figure 13 | The thiols used in failed modifications attempts using the modified *Sa*NAL protein E192C.

The four modifications performed at position F190C can be seen in Figure 14. The majority of these modifications contained alcohol-based functionality, F190-DHPC, F190-HBC and F190-HPC. F190-PC provides a hydrocarbon based functional group.
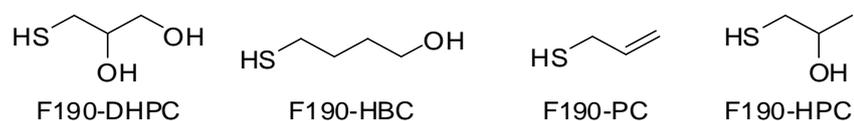
Figure 14 | Thiols used to chemically modify position F190C of *Sa*NAL.

2.3 Conclusion

The four potential substrates **ATOA**, **DPAH**, *ent*-**NANA** and *ent*-**ATOA** were synthesised through three different approaches (Section 2.1). Together with the chemically modified proteins, the effect of chemical modification of *Sa*NAL on substrate acceptance could then be measured *via* kinetic screening.

Results and Discussion: Kinetic Characterisation

Kinetic parameters of all *Sa*NAL variants with any of the *N*-acetylneuraminic acid analogues were determined using an established coupled enzyme assay (Scheme 29 and Section 4.1.21).[72,85] This assay measures the initial rate of the reverse aldol reaction through a change in absorbance at 340 nm. The change in absorbance is a result of the consumption of NADH, a co-factor of L-lactate dehydrogenase utilised in the reduction of pyruvate (a product of the retro aldol reaction). Assuming steady-state kinetics, the rate of consumption of NADH is equal to the rate of turnover of substrate by the aldolase.
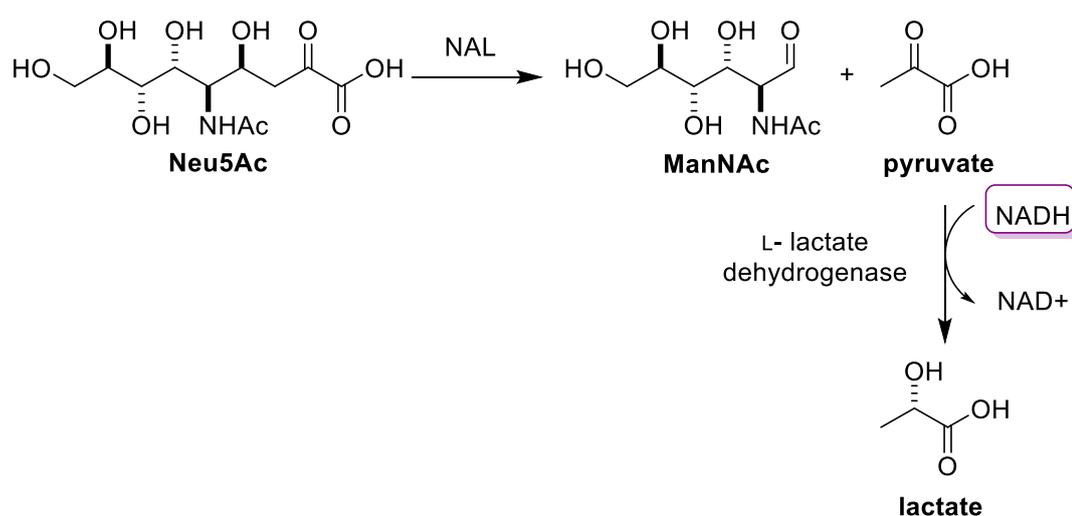
Scheme 29 | Representation of the coupled enzyme assay to determine the kinetic parameters of all *Sa*NAL variants. The cleavage of the full length substrate (in this case *N*-acetylneuraminic acid) is catalysed by an NAL enzyme resulting in pyruvate and an aldehyde. The reduction of pyruvate, catalysed by L-lactate dehydrogenase, results in a decrease in the concentration of NADH, which absorbs at 340 nm. [72,85] For clarity the carbohydrates **Neu5Ac** and **ManNAc** are shown in the open chain forms.

Each measurement was performed in a Tris/ HCl pH 7.4 buffer at 37 °C and the change in absorbance recorded over a period of 1 minute. Each assay consisted of 0.5 units of lactate dehydrogenase, 0.2 mM of NADH, between 0.03 and 0.1 mg/mL of *Sa*NAL variant, and a varied concentration of *N*-acetylneuraminic acid or *N*-acetylneuraminic acid analogue.

To obtain the kinetic parameters k_{cat}, K_M and k_{cat}/K_M for an aldolase variant, several measurements were performed using a fixed concentration of aldolase and a varied concentration of substrate. The specific activity for each concentration of substrate could be calculated from the change in absorbance measured, and a plot of the specific activity versus concentration of substrate could then be used to calculate kinetic parameters for the enzymes i.e. the k_{cat}, K_M and k_{cat}/K_M values.

Before the catalytic parameters of the unnatural amino acid-containing SaNAL variants were determined, kinetic parameters for the enzymes SaNAL and the SaNAL variant E192N with the substrates N-acetylneuraminic acid and the N-acetylneuraminic acid analogue **DPAH** were determined. The kinetic parameters measured for these aldolase enzymes would provide benchmarks to which other SaNAL variants could be compared. Determining the kinetic parameters for the enzyme SaNAL with the substrate N-acetylneuraminic acid was an obvious choice as SaNAL is the wild type enzyme and N-acetylneuraminic acid is the natural substrate. The SaNAL mutant E192N with the N-acetylneuraminic acid variant **DPAH** as the substrate was chosen, as the E. coli NAL variant E192N had previously been engineered to accept this substrate.[16] The kinetics of the SaNAL variant of E192N with the N-acetylneuraminic acid analogue **DPAH** had not been previously evaluated.

Kinetic parameters for each of the four enzyme/substrate combinations were established by first measuring the change in absorbance using a fixed amount of protein and an increasing concentration of substrate (up to 15 mM). The specific activity for each concentration of substrate was then calculated from these values. Graphs of specific activity against substrate concentration were plotted in order to establish the Michaelis-Menten kinetic parameters (Figure 15).
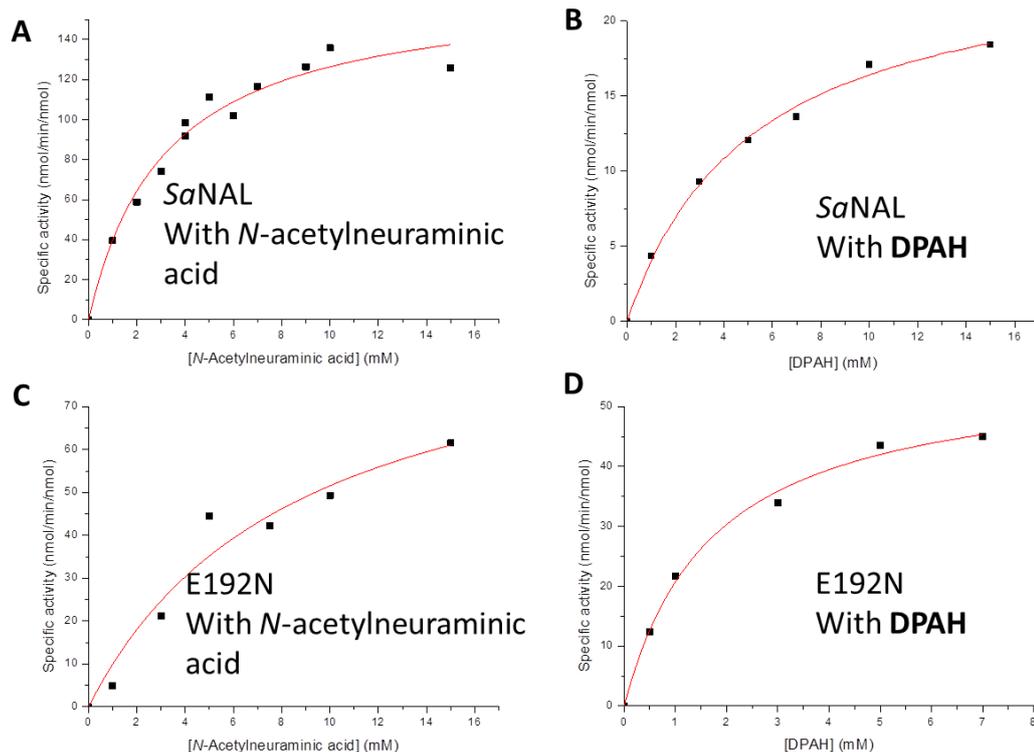
Figure 15 | Activity plots of four *Sa*NAL variants with *N*-acetylneuraminic acid and DPAH as substrates. A) Wild-type *Sa*NAL with *N*-acetylneuraminic acid; **B)** Wild-type *Sa*NAL with the *N*-acetylneuraminic acid analogue **DPAH**; **C)** The *Sa*NAL variant E192N with *N*-acetylneuraminic acid; **D)** The *Sa*NAL variant E192N with the *N*-acetylneuraminic acid analogue DPAH.

Table 1 shows the kinetic parameters (k_{cat}, K_M and k_{cat}/K_M) for all four of the *Sa*NAL aldolase/ substrate combinations examined, as well as those for the *E. coli* NAL combinations for comparison. The k_{cat}, K_M and k_{cat}/K_M values for *Sa*NAL with the substrate *N*-acetylneuraminic acid are comparable to others obtained within the Berry lab. As previously observed with *N*-acetylneuraminic acid [78], the k_{cat} is lower for *S. aureus* NAL compared to that of *E. coli* NAL. The K_M for *Sa*NAL is also lower than that of *E. coli* NAL. This has resulted in a k_{cat}/K_M of 52.1 min^{-1} mM^{-1}, similar to the k_{cat}/K_M of *E. coli* NAL, with *N*-acetylneuraminic acid as the substrate. Similarly, *Sa*NAL with **DPAH** (as a substrate) suffers a decrease in k_{cat} as well as a decrease in K_M, when compared to k_{cat} and K_M of *E. coli* NAL with **DPAH**. This, again, results in a small overall decrease in specific activity (k_{cat}/K_M).

Table 1 | The kinetic parameters observed for *Sa*NAL variants with either *N*-acetylneuraminic acid or DPAH as a substrate. Also included are published results for the *E. coli* NAL variants with *N*-acetylneuraminic acid and DPAH for comparison.[16] As previously stated (Section 1.4.3), *S. aureus* NAL was used throughout this investigation as it contains no natural cysteines.

		E. coli NAL [16]	*S. aureus* NAL	*E. coli* E192N [16]	*S. aureus* E192N
N-acetyl neuraminic acid	k_{cat} / min^{-1}	260 ± 6	166.7 ± 9.8	170 ± 10	96.5 ± 20.1
	K_M / mM	4.4 ± 0.3	3.2 ± 0.5	38 ± 5	8.7 ± 3.8
	k_{cat}/K_M / min^{-1} mM^{-1}	59	52.1	4.4	11.1
DPAH	k_{cat} / min^{-1}	73 ± 4	24.8 ± 1.1	130 ± 3	56.5 ± 2.4
	K_M / mM	11 ± 2	5.17 ± 0.6	0.4 ± 0.04	1.7 ± 0.2
	k_{cat}/K_M / min^{-1} mM^{-1}	7	4.8	340	33.2

Kinetic parameters with the *Sa*NAL variant E192N and the *N*-acetylneuraminic acid analogue **DPAH** had not previously been performed. Table 1 shows a comparison between the kinetic parameters observed with *Sa*NAL E192N and the published results for *E. coli* NAL E192N with **DPAH**. The difference in k_{cat} between the two proteins follows the same trend as *E. coli* NAL and *Sa*NAL, exhibiting a decrease in rate constant. There is, however, a ~4-fold increase in K_M from 0.4 mM with *E. coli* NAL E192N to 1.7 mM in *Sa*NAL E192N. This increase in K_M results in a ~10-fold decrease in k_{cat}/K_M for *Sa*NAL E192N when compared to *E. coli* NAL E192N with **DPAH**.

The greatly increased K_M value for the *Sa*NAL E192N variant in comparison to the *E. coli* NAL E192N variant can be rationalised by comparing the crystal structures of the two proteins (Figure 16). It has previously been established that the two propyl side chains of the *N*-acetylneuraminic acid analogue **DPAH** sit within a hydrophobic pocket of the active site of the E192N variant of *E. coli* NAL.[16]
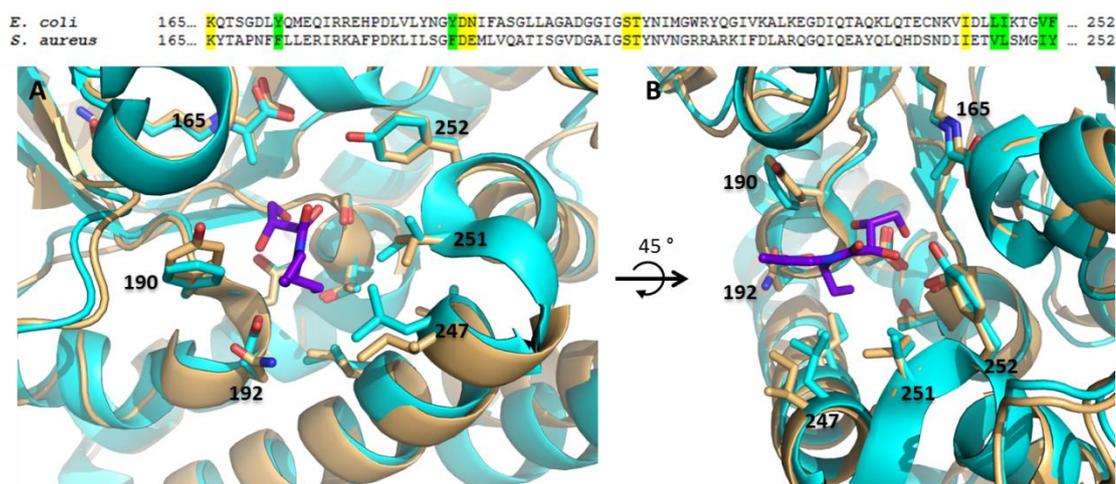
Figure 16 | Representation of the differences in the active sites of the *E. coli* NAL variant E192N and the wild-type *Sa*NAL active site. Top) Sequence alignment of *E. coli* NAL E192N and wild-type *Sa*NAL from residues 165 to 252. The highlighted residues were shown to be important to the recognition of the substrate **DPAH**,[16] residues highlighted in yellow are identical between proteins, and residues highlighted in green differ between proteins. **Bottom)** Structural alignment of the crystal structures of *E. coli* NAL E192N (light orange) in complex with pyruvate and the **DPAH** analogue **THB** (purple), and *Sa*NAL (cyan) in complex with pyruvate. This Figure was generated using PyMOL (PDB file 4AH7 and 2WPB).[16,78]

When overlaying the crystal structure of the *E. coli* NAL variant E192N (in complex with a **DPAH** analogue) with the crystal structure for *Sa*NAL, it becomes clear that several of the residues which make up the hydrophobic pocket are different. The residues 247, 251 and 252 are particularly interesting because in the wild-type *Sa*NAL structure the residues appear to reduce the size of the pocket in which the two propyl chains of **DPAH** would rest. A smaller hydrophobic pocket in the active site, and therefore possibly less efficient **DPAH** binding, may be the reason for the increased K_M value observed with *Sa*NAL E192N when compared to the *E. coli* NAL variant.

Despite the reduction in efficiency between the *Sa*NAL E192N variant when compared to the *E. coli* variant, the E192N variant of *Sa*NAL still exhibits a ~6 fold improvement in k_{cat}/K_M when compared to the wild type *Sa*NAL and the substrate **DPAH**. The study of the kinetic parameters of *Sa*NAL and the variant E192N provide two benchmarks in activity with which to compare the chemically modified variants of *Sa*NAL.

3.1 Kinetic characterisation of chemically modified *Sa*NAL variants using truncated *N*-acetylneuraminic acid analogues as potential substrates

Both site-directed and saturation mutagenesis of the active site of *N*-acetylneuraminic acid lyase has led to an increased level of understanding of the mechanism of catalysis.[15] For instance, mutagenesis and quantum mechanical molecular modelling (QM/MM) studies have shown that the residue Y137 is integral to the catalytic activity of the enzyme, acting as a proton donor in the forward aldol reaction (Figure 17).[15,86] The variant Y137F resulted in a dramatic reduction in activity with *N*-acetylneuraminic acid as a substrate compared to the wild-type protein. Furthermore, the variant Y137A has shown no detectable activity with *N*-acetylneuraminic acid as a substrate.[15] More interestingly, studies of modifications of position E192 have resulted in the discovery of the *N*-acetylneuraminic acid lyase variant E192N which demonstrated switched substrate specificity and allowed the protein to carry out the retro-aldol reaction with the unnatural substrate **DPAH** (k_{cat}/K_M = 336 $min^{-1}mM^{-1}$) more efficiently than cleavage of *N*-acetylneuraminic acid by wild type *E.coli* NAL (k_{cat}/K_M = 59 min^{-1} mM^{-1}).[72]

The positions targeted in this project needed to be relevant to the recognition of the substrate but not integral to the catalytic mechanism. It was therefore envisaged that chemical modification at position 192 might improve the catalysis of the cleavage of a range of *N*-acetylneuraminic acid analogues.
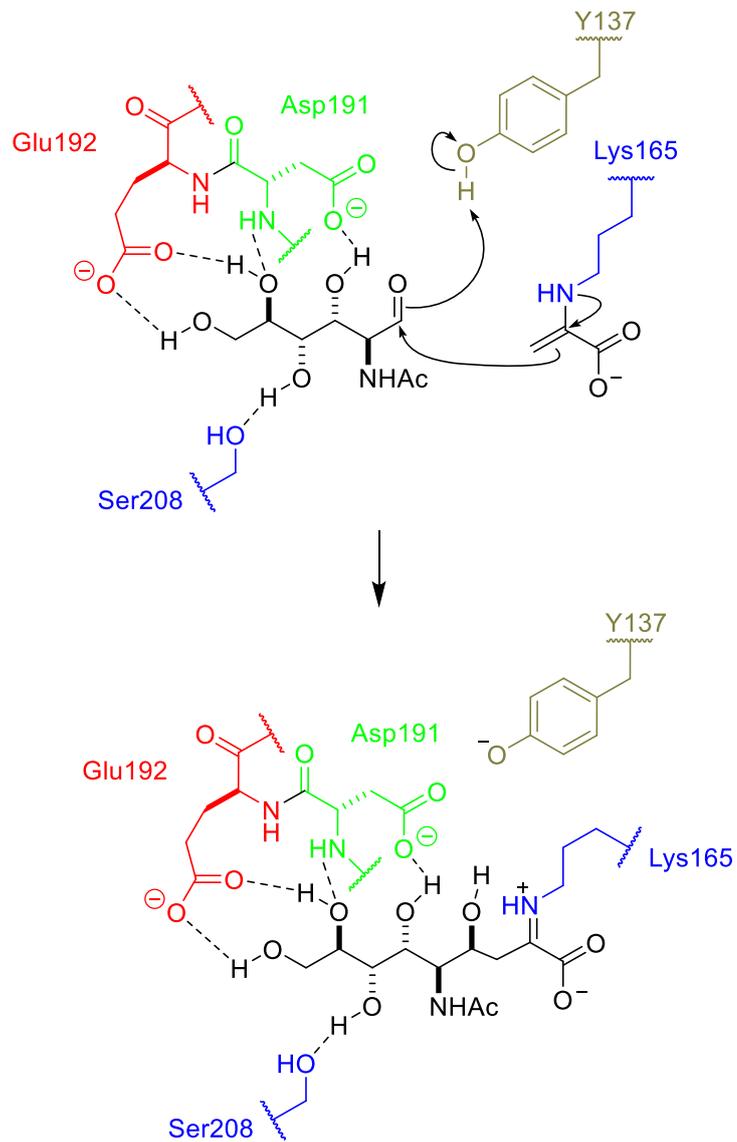
Figure 17 | Schematic of the active site of *E. coli* NAL with ManNAc and pyruvate bound showing H bonding between ManNAc and Glu192, Asp191 and Ser208. The donation of a proton from Y137 occurs during the formation of *N*-acetylneuraminic acid. Adapted from [72,87].

3.1.1 Previous modifications of position E192 of *E. coli* NAL with the aim of catalysing the retroaldol reactions of truncated *N*-acetylneuraminic acid analogues

Research performed by Amanda Bolt (University of Leeds) aimed to produce an aldolase capable of performing the retro-aldol reaction with the truncated *N*-acetylneuraminic acid derivatives **70** and **71** (Figure 18). The compounds **70** and **71** are structurally related to *N*-acetylneuraminic acid and **ATOA**, and the common functionality is highlighted in red in Figure 18. Each of the four compounds had the same sterochemical configuration at C-4, C-5 and C-6. The compounds **70** and **ATOA** differ only in the substitution at position C-5.
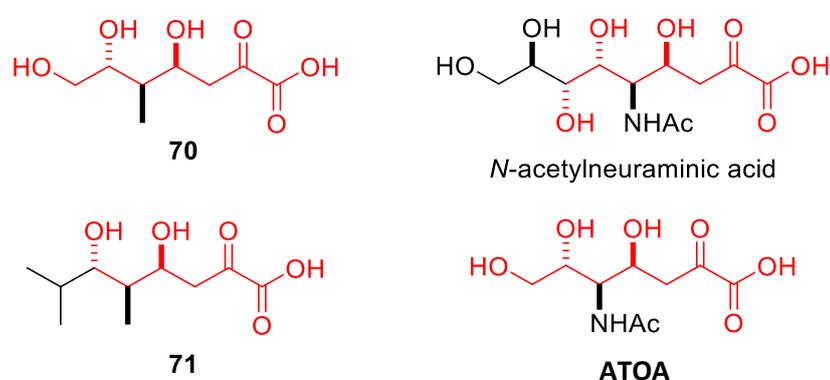
Figure 18 | The truncated sialic acid analogues 70 and 71 used by Amanda Bolt (University of Leeds), and both *N*-acetylneuraminic acid and the truncated analogue ATOA. Features common to all compounds are highlighted in red.

A saturation mutagenesis library at residue 192 (E192X) was created and then tested for catalytic activity using a single concentration of each of the *N*-acetylneuraminic acid analogues **70** and **71**. Each assay used 0.1 mM of substrate and an E192X variant.[87] The results from the E192X library with potential substrates **70** and **71** can be seen in Figure 19 and show an almost complete lack of activity. It was concluded that no E192X variant was capable of catalysing the aldol reaction of the truncated *N*-acetylneuraminic acid analogues **70** and **71**.
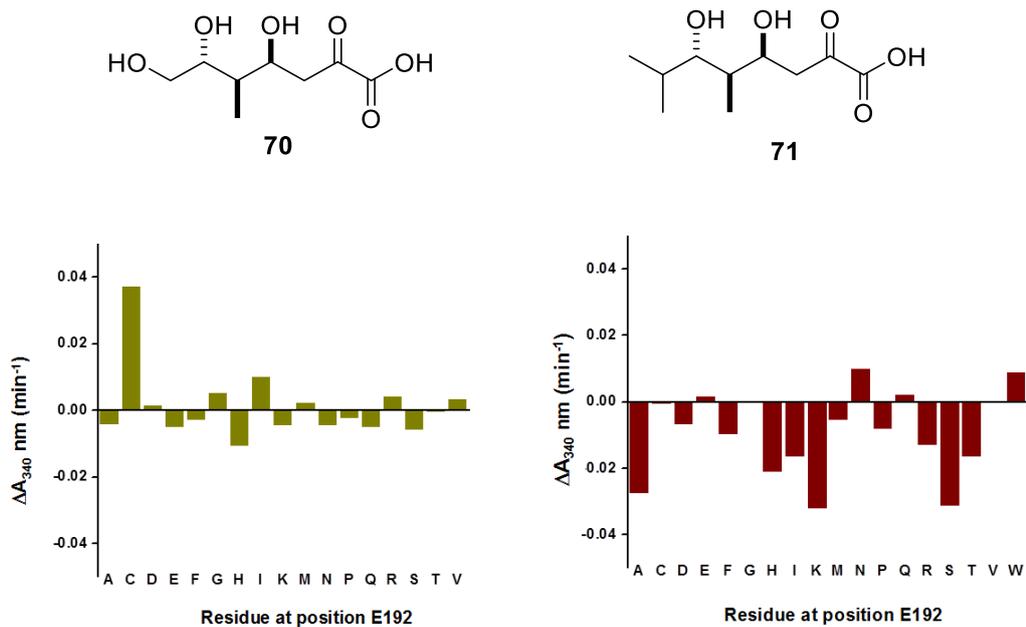
Figure 19 | Catalysis of E192X variants observed with 0.1 mM of substrate and 0.1 mg/mL of enzyme. Left) The rate of change in absorbance observed with compound **70**. Right) The rate of change in absorbance observed with compound **71**. The results shown have been background corrected, whereby measurements conducted in the absence of substrate were deducted from the measurements with substrate. Figure adapted from [87].

3.1.2 Chemical modification at position 192 of *Sa*NAL with the aim of catalysing the retro-aldol reaction of the truncated sialic acid analogue ATOA

One rationale for the lack of activity found with the truncated *N*-acetylneuraminic acid analogues **70** and **71** with the variant E192X library (Section 3.1.1, Figure 19) may be that the side chains at position 192 are too distant from the truncated analogues **70** and **71** to allow effective recognition of the substrate (Figure 20).
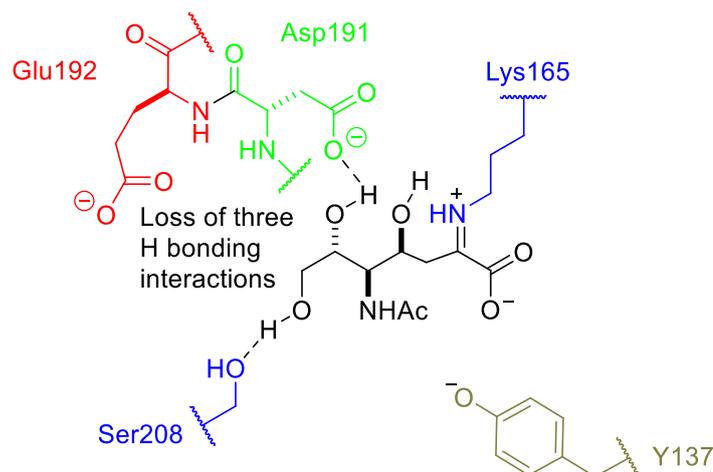
Figure 20 | Representation of a truncated *N*-acetylneuraminic acid analogue in the active site of NAL. The residue at position 192 may be too far from the truncated *N*-acetylneuraminic acid analogue to allow recognition of the truncated substrates **70** and **71**. The representation in the figure assumes that truncated substrates may maintain interactions with Ser208 and Asp191 observed with *N*-acetylneuraminic acid, while no interaction with 192 may be possible.

An investigation into whether the chemical modification of *Sa*NAL variants could enable catalysis of the retro-aldol reaction of truncated *N*-acetylneuraminic acid analogues had not previously been performed. The aim was to examine the kinetic parameters of the *Sa*NAL variants chemically modified at position 192 in the reaction with the truncated *N*-acetylneuraminic acid analogue **ATOA**. The side chains of the chemically modified E192 variants may provide unnatural functionality to the enzyme, and possibly promote activity with the truncated *N*-acetylneuraminic acid analogue **ATOA**. An example of how unnatural amino acids might increase the catalysis of the cleavage of truncated *N*-acetylneuraminic acid analogues can be seen in Figure 21. It is possible that the absence of interactions between the alcohols of the truncated *N*-acetylneuraminic acid analogue **ATOA** and the E192 residue could be responsible for the lack of catalysis (Figure 21 B). If residue 192 was to be lengthened

through chemical modification (e.g. to give the chemically modified *Sa*NAL variant E192-CMC), an interaction between the truncated *N*-acetylneuraminic acid analogue **ATOA** and the side chain might be restored (Figure 21 C). Figure 21 is however speculative and a thorough investigation into the effects of all the chemical modification of side chains at position E192 was needed.
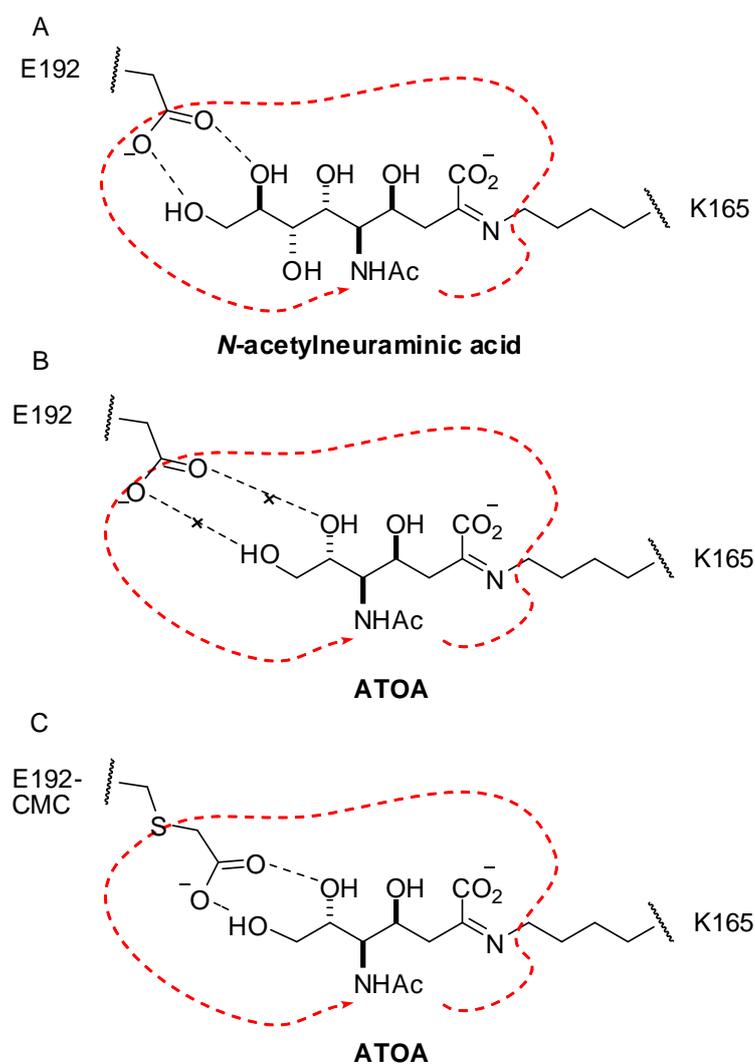
Figure 21 | A graphical representation of the active site of NAL in complex with truncated *N*-acetylneuraminic acid analogue ATOA and an unnatural amino acid at position E192. A) The hydrogen bonding between *N*-acetylneuraminic acid and E192 of wild-type *Sa*NAL. **B)** The truncated substrate **ATOA** may not be able to make such hydrogen bonding interactions. **C)** Incorporation of an unnatural lengthened glutamic acid analogue into position E192 may allow hydrogen bonding between the truncated substrate and the modified enzyme.

3.1.3 Evaluation of the catalytic ability of *Sa*NAL variants, chemically modified at position 192, with the truncated *N*-acetylneuraminic acid analogue ATOA as a substrate

Single concentration kinetic measurements were performed with all of the variants of *Sa*NAL that had been chemically modified at position 192. Each measurement was performed using a high (10 mM) concentration of the truncated *N*-acetylneuraminic acid analogue **ATOA** and 0.1 mg/mL of enzyme. The results of these experiments can be seen in Figure 22, as well as the background uncorrected measurements and the background measurements.
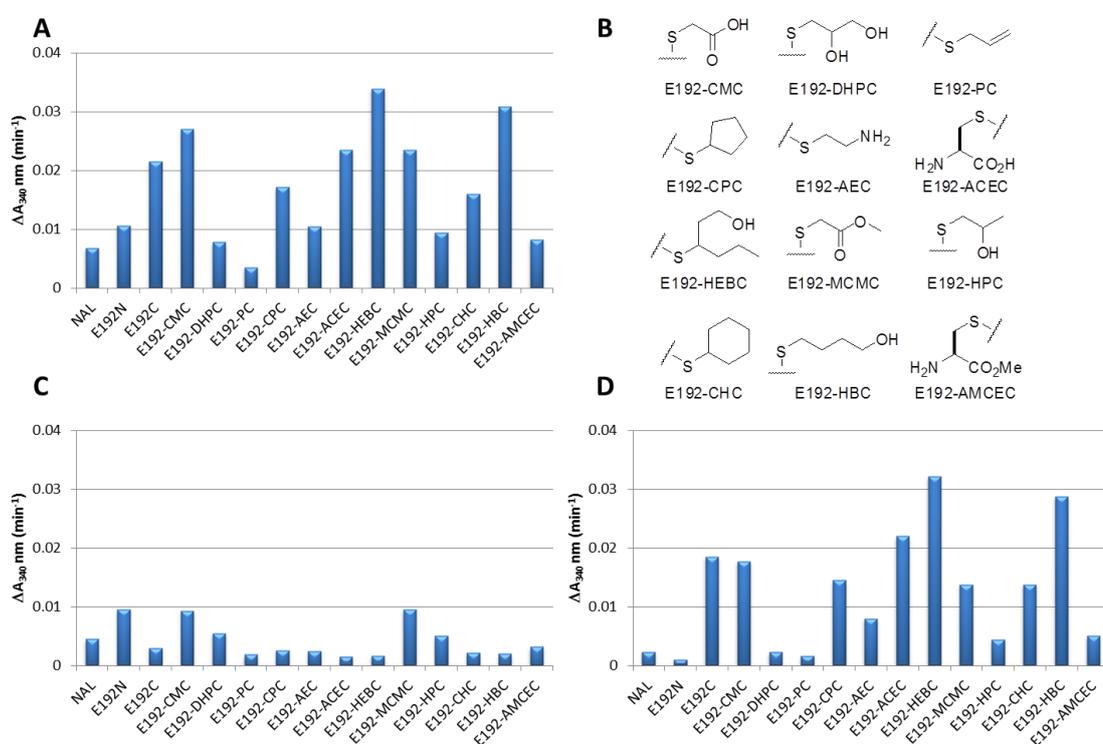
Figure 22 | The single point kinetic measurements performed using the truncated *N*-acetylneuraminic acid analogue ATOA. A) The uncorrected change in absorbance observed over the period of 1 minute. Each measurement was performed using 10 mM of substrate and 0.1 mg/mL of protein. **B)** The unnatural amino acid side chains of chemically modified proteins used in the assay. **C)** The background change in absorbance observed over 1 min. These measurements are the sum of two control experiments: (i) with 10 mM of substrate and no aldolase, and (ii) with 0.1 mg/mL of aldolase and no substrate. **D)** The background corrected change in absorbance observed over one minute.

The activities even at the high concentration of the compound **ATOA** proved modest and in some cases the background activity is almost as high as the uncorrected measurement (i.e. E192N). Although modest, there was significant increase in activity of some of the chemically modified variants (i.e. E192-HEBC and E192-HBC) compared to the wild-type protein.

The majority of the E192 variants were tested with higher concentrations of **ATOA** in an attempt to evaluate their kinetic parameters in greater detail. A maximum concentration of 30 mM of substrate was used to test the activity with the substrate **ATOA**. Even at these high concentrations of substrate, saturation of the enzyme was not approached. The specific activity (k_{cat}/K_M) could, however, be estimated for each enzyme. Figure 23 illustrates the activity of E192-ACEC and E192-PC in response to an increasing concentration of substrate **ATOA** (see appendix II for all other plots).
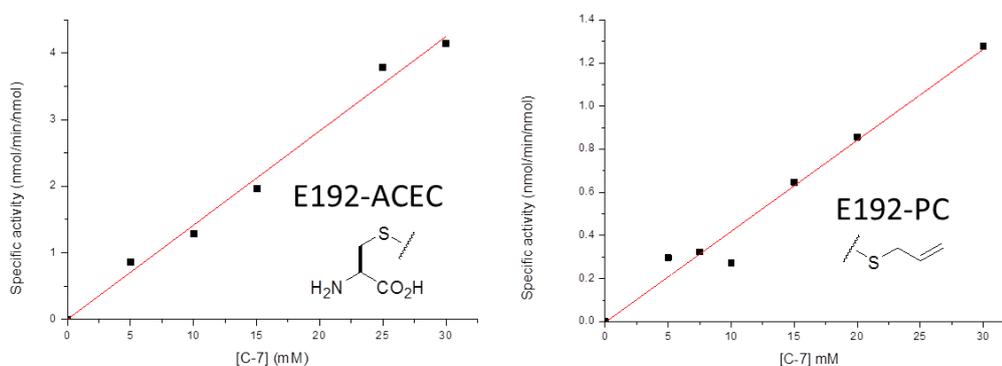
Figure 23 | Two examples of the assessment of the kinetic parameters of the *N*-acetylneuraminic acid analogue ATOA and *Sa*NAL variants chemically modified at residue 192. Left) The specific activity observed with E192-ACEC. **Right)** The specific activity observed with E192-PC. Even at concentrations of 30 mM of **ATOA**, saturation was not reached.

Figure 24 shows all of the estimated k_{cat}/K_M values calculated. When compared to the k_{cat}/K_M of either *N*-acetylneuraminic acid with wild type *Sa*NAL (52 min^{-1}mM^{-1}) or **DPAH** with the *Sa*NAL variant E192N (33 min^{-1} mM^{-1}) the results are clearly low. However, it would appear that subtle effects on the catalysis of the cleavage of truncated *N*-acetylneuraminic acid analogue **ATOA** has resulted from chemical modification of residue 192, though none of the chemically modified variants were more active than E192C. Out of the four best chemically modified variants of *Sa*NAL, E192-CMC, E192-ACEC and E192-HBC possess hydrogen donor and acceptor properties which may be capable of forming the predicted interactions seen in Figure 21.

Unexpectedly, the variant E192-CPC also exhibited a specific activity (k_{cat}/K_M) that was comparable to E192-CMC, E192-ACEC and E192-HBC.
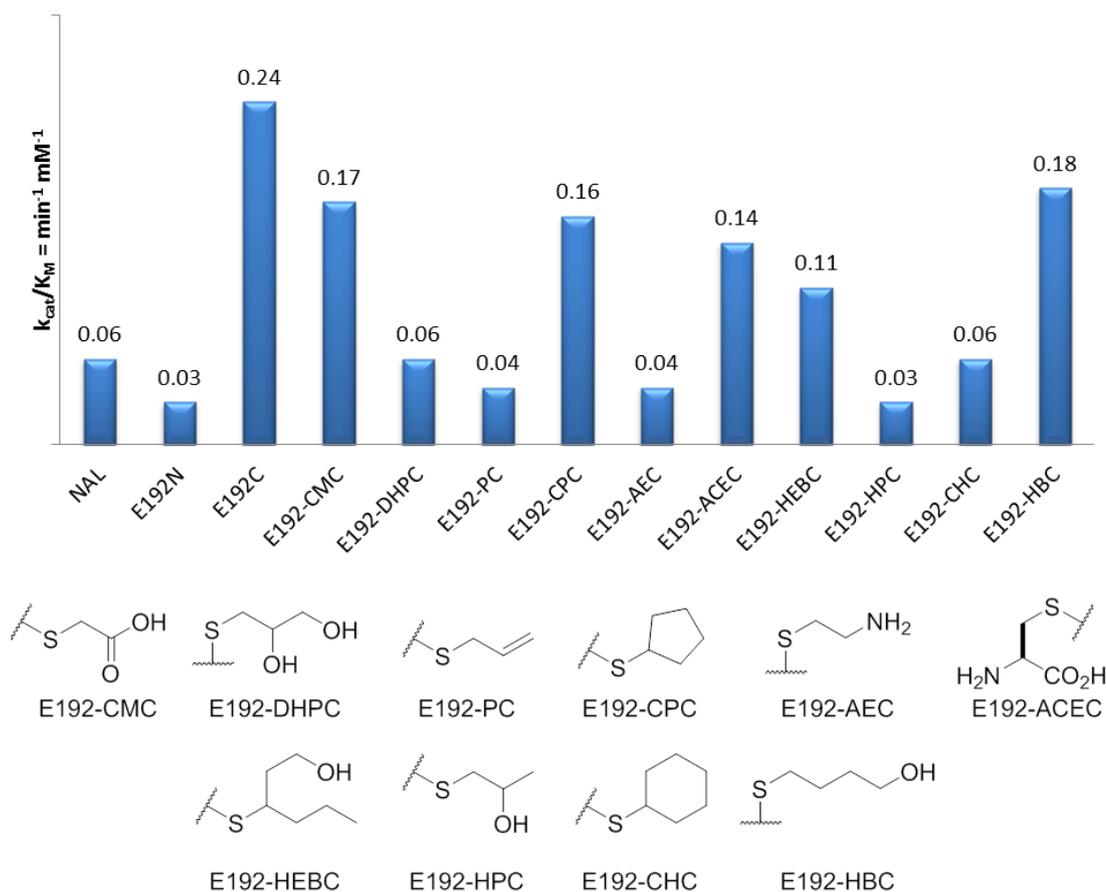
Figure 24 | Estimated k_{cat}/K_M values for some of the E192X variants. The values are estimated by consideration of the gradient of specific activity against concentration of **ATOA**. Each measurement taken for obtaining these results was performed using 0.1 mg/mL of protein and a concentration of **ATOA** ranging from 0 mM to 30 mM.

These studies have shown that the modified active site represented in Figure 21 is an oversimplification of the effects of chemical modification of active site residues on the recognition of the truncated *N*-acetylneuraminic acid analogue **ATOA**. However, the chemical modification of other residues within the active site of *Sa*NAL may have a greater impact on the catalysis of the truncated *N*-acetylneuraminic acid analogue **ATOA**.

3.1.4 Kinetic characterisation of *Sa*NAL variants, chemically modified at position 190, with a truncated *N*-acetylneuraminic acid analogue as a substrate

Parallel research into the effects of the chemical modification of the active site of *Sa*NAL by Claire Windle (University of Leeds) has resulted in the *Sa*NAL variant F190-DHPC. This variant is capable of catalysing the forward aldol reaction between erythrose and pyruvate to yield the sugar **72** in a more efficient manner than the wild type protein, *Sa*NAL (Scheme 30). In these studies, the concentration of product formed was determined using an established end-point assay.[79,80]
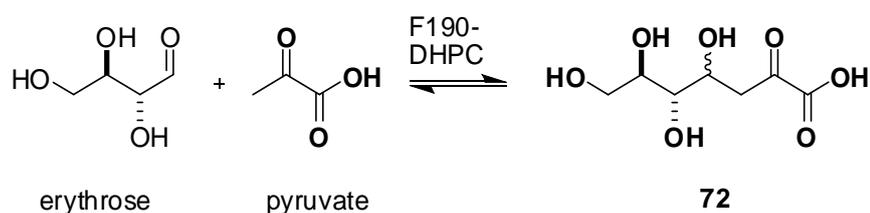
Scheme 30 | An F190-DHPC modified *Sa*NAL protein capable of the catalysis of the forward aldol reaction of erythrose and pyruvate. The variant F190-DHPC was determined to be more efficient than *Sa*NAL at catalysing the forward aldol reaction with these substrates. The concentration of product formed was established using an end point assay.[79,80]

The *N*-acetylneuraminic acid analogue **72** is comparable to the *N*-acetylneuraminic acid analogue **ATOA** in both length and functionality (Figure 25). Both sugars are seven carbons long although the configuration and substitution at C-5 is different.
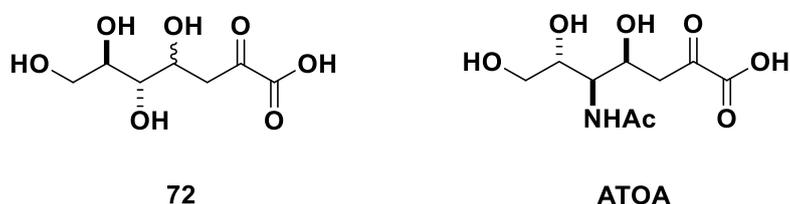
Figure 25 | The truncated *N*-acetylneuraminic acid analogue 72 compared to the *N*-acetylneuraminic acid analogue ATOA. The two sugars differ in functionality at carbon C-5 and configuration. Unlike compound **72**, **ATOA** contains a single epimer at position C-3.

The activity observed with the chemically modified variant F190-DHPC in the condensation of erythrose and pyruvate results from the replacement of an aromatic residue at position F190 (Figure 26) with an unnatural residue possessing polar functionality. Figure 26 illustrates the *Sa*NAL enzyme with pyruvate covalently bound at residue K165. Also highlighted are the residues F190 and E192. Upon chemical modification of F190, newly introduced hydrogen bonding may stabilise the erythrose molecule, allowing recognition and cleavage of the truncated substrate **72**.
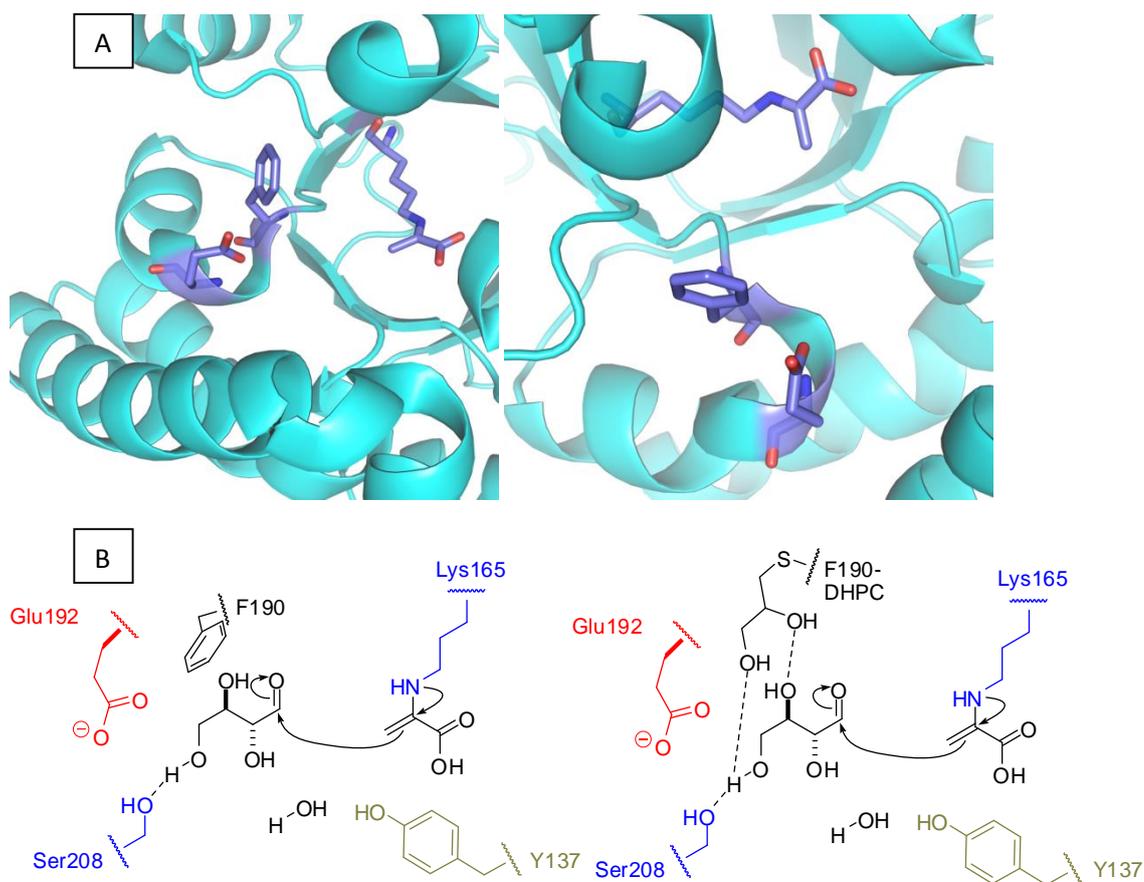
Figure 26 | Proposed basis for the catalysis of the aldol reaction of a truncated *N*-acetylneuraminic acid analogue by a chemically modified aldolase. A) Structure of the active site of *Sa*NAL (from two viewpoints) highlighting positions K165, covalently bound with pyruvate, F190 and E192. **B)** Proposed representation of the active site of *Sa*NAL in complex with erythrose (left) of the F190-DHPC variant of *Sa*NAL in complex with erythrose (right). These models assume that erythrose binds to the active site of F190-DHPC in a similar manner to *N*-acetylneuraminic acid binding to *E. coli* NAL. **(A)** was generated using PyMOL (PDB file 4AH7).[78]

3.1.5 Evaluation of the catalytic ability of *Sa*NAL variants, chemically modified at position 190, with the truncated *N*-acetylneuraminic acid analogue ATOA as a substrate

As activity of the variant F190-DHPC was observed with a truncated *N*-acetylneuraminic acid analogue **72**, four *Sa*NAL variants chemically modified at position 190 were prepared (Figure 27). The variants F190-HBC and F190-HPC were chosen as they contain hydroxyl groups similar to the variant F190-DHPC. The variant F190-PC was also chosen as an alternative to the hydroxyl functionality. These variants were then used in single concentration kinetic measurements with the *N*-acetylneuraminic acid analogue **ATOA** as substrate.

F190-DHPC   F190-HBC   F190-PC   F190-HPC
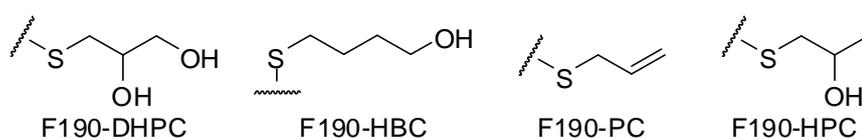
Figure 27 | The side chain of *Sa*NAL variants chemically modified at position 190. The variants F190-DHPC, F190-PC and F190-HPC contain hydroxyl groups. The variant F190-PC was chosen as a non-polar alternative to this functionality.

Kinetic measurements were taken over a one minute period and were performed using 0.1 mg/mL of modified protein and 5 mM of the *N*-acetylneuraminic acid analogue **ATOA**. The results of these experiments can be seen in Figure 28.
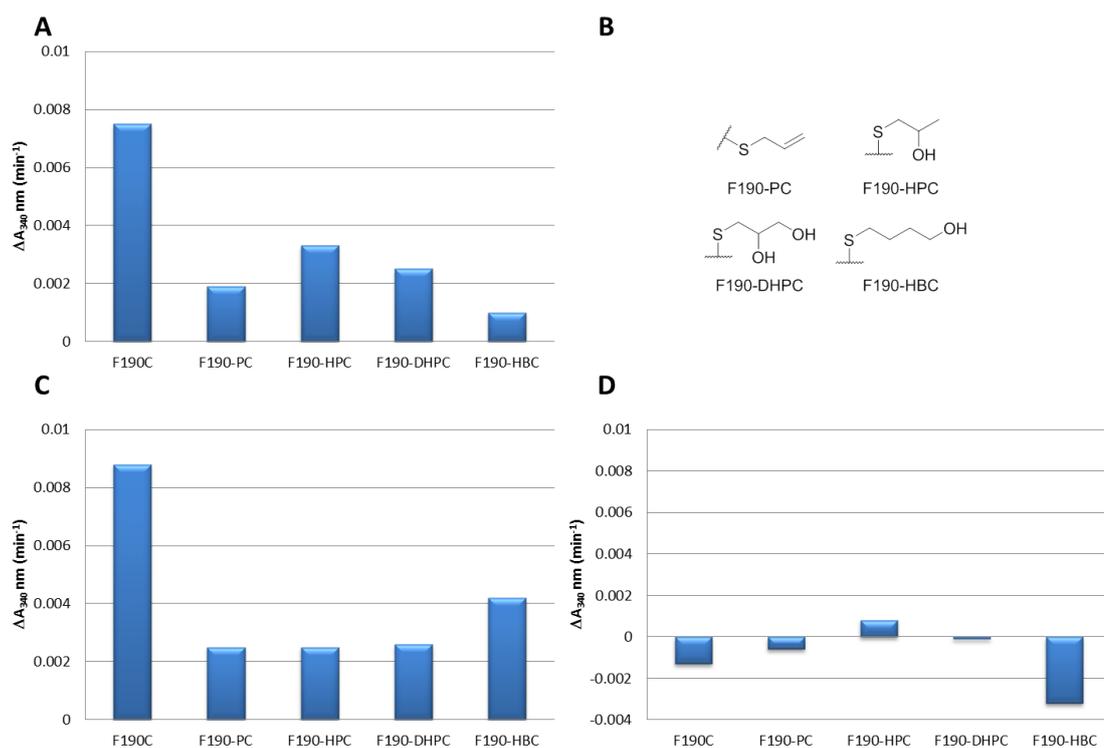
Figure 28 | The single concentration kinetic measurements performed using the truncated *N*-acetylneuraminic acid analogue ATOA. A) The uncorrected change in absorbance observed over the period of 1 minute. Each measurement was performed using 5 mM of **ATOA** and 0.1 mg/mL of modified protein. **B)** The unnatural amino acid side chains of chemically modified proteins used in the assay. **C)** The background change in absorbance observed over 1 min. These measurements are the sum of two control experiments: (i) with 5 mM of substrate in the cell and no aldolase, and (ii) with 0.1 mg/mL of protein and no substrate. **D)** The background corrected change in absorbance observed over one minute.

Unfortunately, no significant activity was observed with any of the four chemically modified enzymes. This result contrasts with the observation that the *Sa*NAL variant F190-DHPC is able to catalyse the forward aldol reaction between pyruvate and erythrose to give the truncated *N*-acetylneuraminic acid analogue **72** (Claire Windle, University of Leeds). It can be concluded that the difference in activity observed between the *N*-acetylneuraminic acid analogues **ATOA** and **5** could stem from the difference in configuration of the substrates, the functionality at C-5 or a combination of both factors.

3.1.6 Kinetic characterisation of *Sa*NAL variants, chemically modified at position 190, with an enantiomeric truncated *N*-acetylneuraminic acid analogue as a substrate

The molecule ***ent*-ATOA** (Figure 29) was synthesised to investigate the effects of configuration of the molecule on catalysis of the retro-aldol reaction. If the stereochemical configuration of the molecule was a major determinant in the turn-over of substrate then the molecule, ***ent*-ATOA** might be cleaved by at least one of the *Sa*NAL variants chemically modified at residue 190.

ATOA   **72**   *ent*-**ATOA**

Figure 29 | Comparison of three truncated *N*-acetylneuraminic acid analogues. ATOA contains the same stereochemical configuration and functionality as *N*-acetylneuraminic acid. The analogue **72** (the product of the forward reaction with erythrose and pyruvate catalysed by the *Sa*NAL variant F190-TG (see section 3.1.4) differs in stereochemical configuration and functionality at C-5. The analogue ***ent*- ATOA** is the enantiomer of the compound **ATOA**.

Single kinetic measurements were performed with ***ent*-ATOA** using 0.1 mg/mL of modified protein and 5 mM of the truncated *N*-acetylneuraminic acid analogue ***ent*-ATOA**. The results of these measurements can be seen in Figure 30.

Figure 30 | The single concentration kinetic measurements performed using the truncated *N*-acetylneuraminic acid analogue *ent*-ATOA. A) The uncorrected change in absorbance observed over the period of 1 minute. Each measurement was performed using 5 mM of ***ent*-ATOA** and 0.1 mg/mL of modified protein. **B)** The unnatural amino acid side chains of chemically modified proteins used in the assay. **C)** The background change in absorbance observed over 1 min. These measurements were taken using 0.1 mg/mL of aldolase and no substrate. **D)** The background corrected change in absorbance observed over one minute.

As with the measurements performed with the truncated *N*-acetylneuraminic acid analogue **ATOA**, there is no substantial activity with any of the *Sa*NAL variants tested. The conclusion, therefore, is that the lack of catalysis seen with this molecule relative to results seen with the truncated *N*-acetylneuraminic acid analogue **72** and the chemically modified *Sa*NAL variant F190-DHPC (Claire Windle, University of Leeds) is not only a result of the stereochemical configuration of the molecule. The catalysis of the truncated *N*-acetylneuraminic acid analogue must therefore have a strong relationship with the C-5 hydroxyl. The C-5 hydroxyl substituent in the substrate **72** must be a key determinant in the catalysis of the cleavage of the truncated substrate.
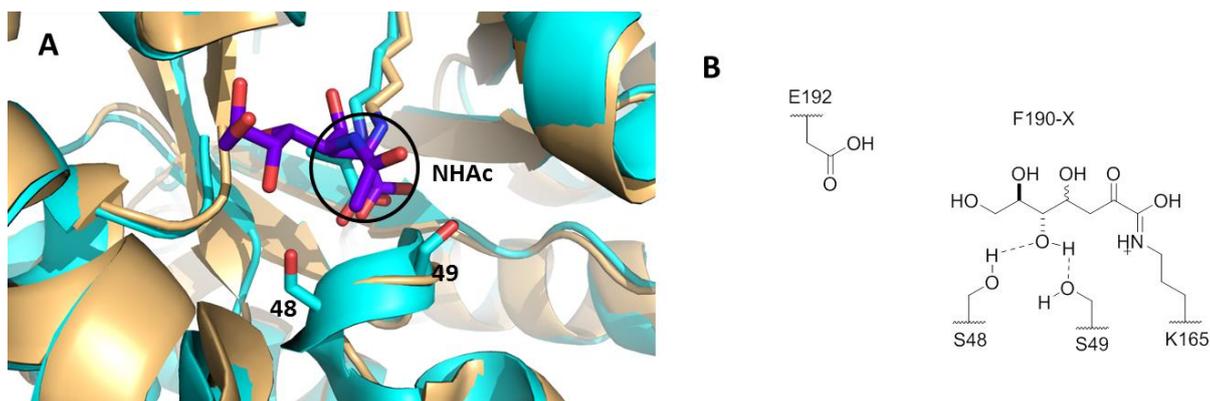
Figure 31 | Proposed interactions of the *N*-acetylneuraminic acid analogue 72. A) The binding site of *Sa*NAL (cyan) overlaid with the *E. coli* NAL variant Y127A (light orange) in complex with *N*-acetylneuraminic acid (purple). Highlighted are the residues S48 and S49 along with the C-5 NHAc functional group of *N*-acetylneuraminic acid. **B)** A schematic of possible hydrogen bonding between an *Sa*NAL variant and the truncated *N*-acetylneuraminic acid analogue **72.** Hydrogen bonding may occur between the residues S48, S49 and the alcohol at C-5. This hydrogen bonding may not be possible with the NHAc group of ***ent*-ATOA**. This figure assumes the same binding conformation of *N*-acetylneuraminic acid. The alcohol C-5 of the substrate **72** may be around 4 Å from the alcohol residues S48 and S49. **(A)** was generated using PyMOL (PDB files 4AH7 and 4BWL).[15,78]

The crystal structure in Figure 31 shows two serine side chains at positions 48 and 49. It may be possible that the side chains at residue 48 and 49 are able to form stabilising interactions with the C-5 hydroxyl group of the *N*-acetylneuraminic acid analogue **72** which are not possible with the truncated analogues **ATOA** and ***ent*-ATOA**. [15,78]

3.2 Kinetic characterisation of the *Sa*NAL variants, chemically modified at residue 192, with the substrate *N*-acetylneuraminic acid as a substrate

An obvious line of investigation was the study of the effects of chemical modification at position 192 of *Sa*NAL on activity with the natural substrate, *N*-acetylneuraminic acid. Previous studies of the E192X saturation mutagenesis library against *N*-acetylneuraminic acid using single concentration kinetic measurements had been performed (Amanda Bolt, University of Leeds), Figure 32. [87] The results of these measurements highlighted residue 192 as a highly relevant residue for the catalysis of the cleavage of *N*-acetylneuraminic acid. Changes at residue E192 resulted in a dramatic reduction in the turnover of *N*-acetylneuraminic acid with the exception of the mutant E192Q, Figure 32. It is therefore clear that the length and nature of the side chain at position 192 is crucial in the recognition of the substrate.
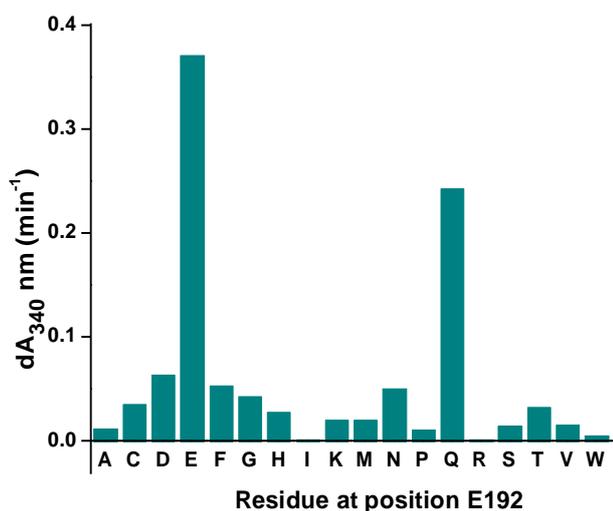
Figure 32 | Catalysis of E192X variants observed with 0.5 mM of the substrate *N*-acetylneuraminic acid and 0.03 mg/ml of enzyme. The change in absorbance was measured over 1 minute. Adapted from [87].

The observation that E192 is an important residue for the catalysis of *N*-acetylneuraminic acid can be rationalised structurally. The crystal structure of *Haemophilus influenza* NAL in complex with the inhibitor 4-oxo-sialic acid implicates hydrogen bonding between the two oxygen atoms of the side chain of the residue E192 and the alcohols on C-8 and C-9 of 4-oxo-sialic acid.[88] This could also explain the observation of catalysis with E192Q and *N*-acetylneuraminic acid (Figure 33). The conclusion drawn was that catalysis of the cleavage of *N*-acetylneuraminic acid by NAL relies heavily on hydrogen bonding between the substrate and the side chain at

position 192. Interactions at position 192 that lead to the cleavage of *N*-acetylneuraminic acid appear to require either an acid or an amide of the correct length (Figure 32). The shorter side chain variants of *E. coli* NAL, E192D and E192N, exhibited relatively little turnover of the reverse-aldol reaction of *N*-acetylneuraminic acid.
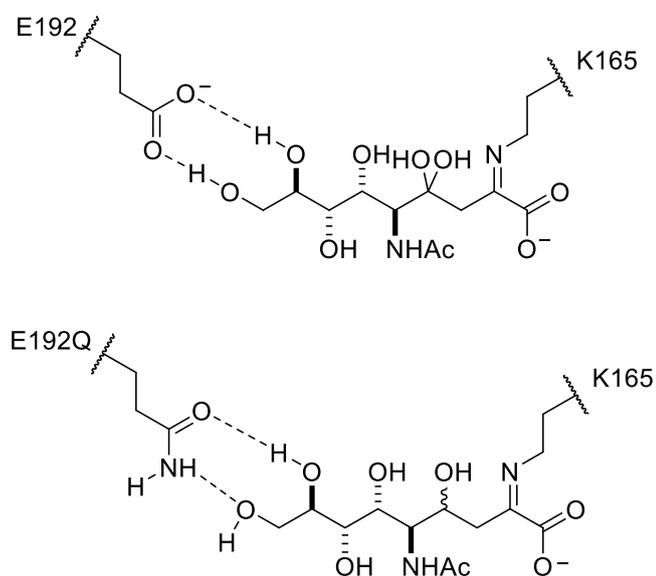

Figure 33 | Complexes of aldolase variants. Top) Structure of *E.coli* NAL in complex with the inhibitor 4-oxo-sialic acid, showing hydrogen bonding observed between the alcohols C-8 and C-9 of 4-oxo-sialic acid and the residue E192 and **Bottom)** Proposed interactions between the alcohols C-8 and C-9 of *N*-acetylneuraminic acid and the residue E192Q. These interactions may account for the activity of the variant enzyme E192Q with *N*-acetylneuraminic acid. Adapted from [87].

The decision was made to broaden the investigation of the effects of modification of residue 192 on the catalysis of *N*-acetylneuraminic acid by using unnatural amino acid side chains. Based on previous results, the hypothesis was that any chemical modification at residue 192 of *Sa*NAL would likely result in a great reduction in the ability of the protein to catalyse the retro-aldol reaction of *N*-acetylneuraminic acid.

3.2.1 Evaluation of the catalytic ability of *Sa*NAL variants, chemically modified at position 192, with the substrate *N*-acetylneuraminic acid

Due to the abundance of the substrate *N*-acetylneuraminic acid, it was decided that full kinetic evaluation of each of the chemically modified proteins with this substrate was viable. The kinetic parameters k_{cat}, K_M and k_{cat}/K_M were therefore determined for all of the *Sa*NAL variants that had been chemically modified at residue 192. The results of these experiments could then be compared to the kinetic parameters of *Sa*NAL with *N*-acetylneuraminic acid, and the impact of the modifications at position 192 determined.

The kinetic parameters for each variant were established by first measuring the rate of change in absorbance using a fixed amount of protein (0.1 mg/mL) and an increasing concentration of *N*-acetylneuraminic acid (up to 20 mM). The specific activity of the enzyme was calculated for each concentration of *N*-acetylneuraminic acid. Michaelis-Menten enzyme kinetics could then be established from the graph of the specific activity against substrate concentration (Figure 34).
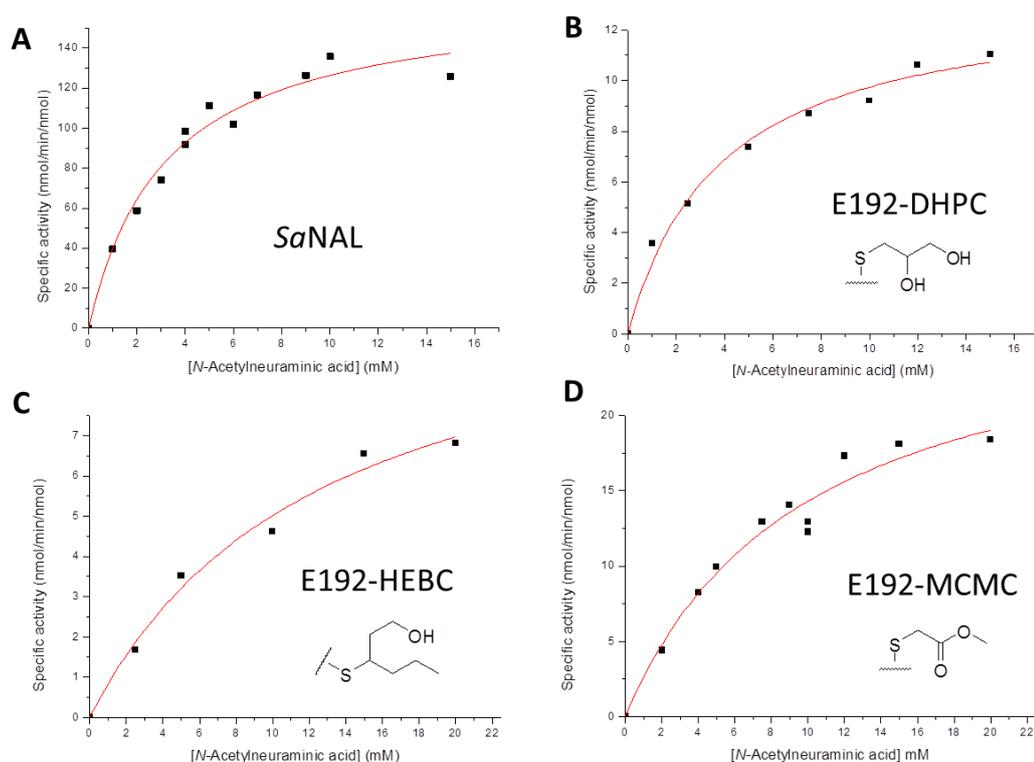
Figure 34 | Activity plots of four *Sa*NAL variants with the substrate *N*-acetylneuraminic acid. A) the wild-type protein *Sa*NAL, **B)** the chemically modified protein E192-DHPC, **C)** the chemically modified protein E192-HEBC, **D)** the chemically modified protein E192-MCMC.

Table 2 shows the k_{cat}, K_M and k_{cat}/K_M values of *N*-acetylneuraminic acid with all *Sa*NAL variants chemically modified at residue 192 in addition to wild-type *Sa*NAL, and the two mutants E192N and E192C. It is clear that the k_{cat} values of all of the *Sa*NAL variants chemically modified at position 192 are very low in comparison to the wild-type enzyme and also show a greatly reduced level of activity compared to E192N and E192C. The k_{cat} values for the chemically modified variants range from 8.8 min^{-1} (E192-CHC) to 28.4 min^{-1} (E192-MCMC) i.e. ~ 5 - 20% of that of *Sa*NAL (166.7 min^{-1}). In all but two cases (E192-DHPC and E192-HPC), the K_M is at least double that of wild-type *Sa*NAL (3.2 mM), with the highest K_M value resulting from the chemically modified variant E192-AEC (17.8 mM). The low k_{cat} values and the high K_M values together result in a greatly reduced specific activity of all the chemically modified proteins in comparison to wild type *Sa*NAL. These results are consistent with the prior knowledge that size and functional group at position 192 is important for recognition of the substrate *N*-acetylneuraminic acid.

Table 2 | Table of the kinetic parameters determined for each of the variants of *Sa*NAL (chemically modified at residue 192) with *N*-acetylneuraminic acid. Also included are the kinetic parameters determined for the wild type protein *Sa*NAL and the variants E192N and E192C**.** The errors for k_{cat} and K_M are fitting errors.

		Kinetic parameters		
		k_{cat} = min^{-1}	K_M = mM	k_{cat}/K_M = min^{-1} mM^{-1}
Protein variant	NAL	166.7 + 9.8	3.2 + 0.5	52.1
	E192N	96.5 + 20.1	8.7 + 3.8	11.1
	E192C	41.1 + 6.8	8.6 + 2.8	4.8
	E192-CMC	16.7 + 1.1	11.1 + 1.6	1.5
	E192-DHPC	13.4 + 0.8	3.8 + 0.7	3.5
	E192-PC	9.04 + 1.6	9.5 + 3.2	1.1
	E192-CPC	11.16 + 1.5	7.6 + 2.1	1.5
	E192-AEC	25.7 + 3.6	17.8 + 4.3	1.4
	E192-ACEC	14.0 + 1.9	12.8 + 2.2	1.1
	E192-HEBC	11.4 + 1.5	12.8 + 3.4	0.9
	E192-MCMC	28.4 + 3.3	9.9 + 2.4	2.9
	E192-HPC	16.34 + 2.0	5.7 + 1.7	2.9
	E192-CHC	8.8 + 0.8	9.5 + 1.9	0.9
	E192-HBC	20.6 + 2.5	9.6 + 2.6	2.1
	E192-AMCEC	11.1 + 0.5	7.8 + 1.0	1.4

When comparing the k_{cat} values of the unnatural variants it would appear that E192-AEC, E192-MCMC and E192-HBC provide the best results. The lowest K_M values were achieved by E192-DHPC and also E192-HPC. This resulted in the three best unnaturally modified variants for catalysis of *N*-acetylneuraminic acid being determined as E192-DHPC, E192-MCMC and E192-HPC.

The results of the k_{cat}/K_M can be seen in Figure 35. The results achieved were as expected as replacement of the glutamic acid residue with any of the unnatural residues resulted in a ~50 fold reduction in efficiency when compared to wild-type *Sa*NAL.
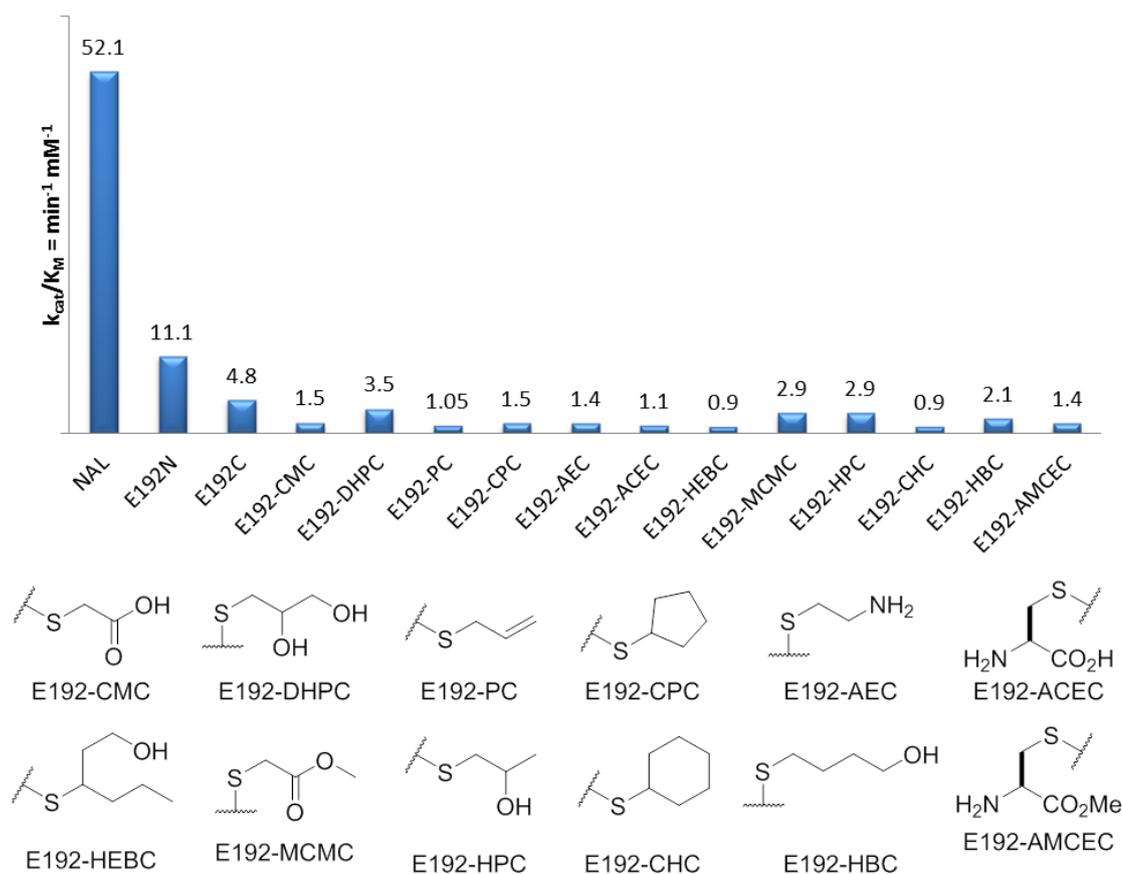

Figure 35 | k_{cat}/K_M (min^{-1} mM^{-1}) for the *Sa*NAL variants (chemically modified at residue 192) with the substrate *N*-acetylneuraminic acid.

The results gained from investigation of the effects of the chemical modification of residue 192 on the catalysis of the cleavage of *N*-acetylneuraminic acid confirm the importance of size on activity. All of the unnatural residues are large in comparison with E and Q. The large residues at position 192 may prevent efficient binding and cleavage of the substrate in the active site.

3.2.2 Kinetic characterisation of *Sa*NAL variants, chemically modified at residue 190, with the substrate *N*-acetylneuraminic acid

An investigation into the impact of varying residue F190 on the catalysis of the cleavage of *N*-acetylneuraminic acid had not previously been performed. An assertion could be made that the phenyl group at position 190 has minimal interactions with the numerous hydrogen bond donors on *N*-acetylneuraminic acid (Figure 36), as the hydrophobic phenyl group resides above the plane of the bound *N*-acetylneuraminic acid. Though residue F190 most likely does not hydrogen bond with the substrate *N*-acetylneuraminic acid, it may provide a key role in defining the active site of *Sa*NAL.

If, unlike residue E192, residue F190 is not integral to the identification and catalysis of the cleavage of *N*-acetylneuraminic acid, chemical modification of residue 190 may not drastically diminish activity. Therefore, full kinetic parameters were determined for *N*-acetylneuraminic acid with the four *Sa*NAL variants (F190-DHPC, F190-HPC, F190-HBC and F190-PC previously discussed in section 3.1.6) chemically modified at position 190.
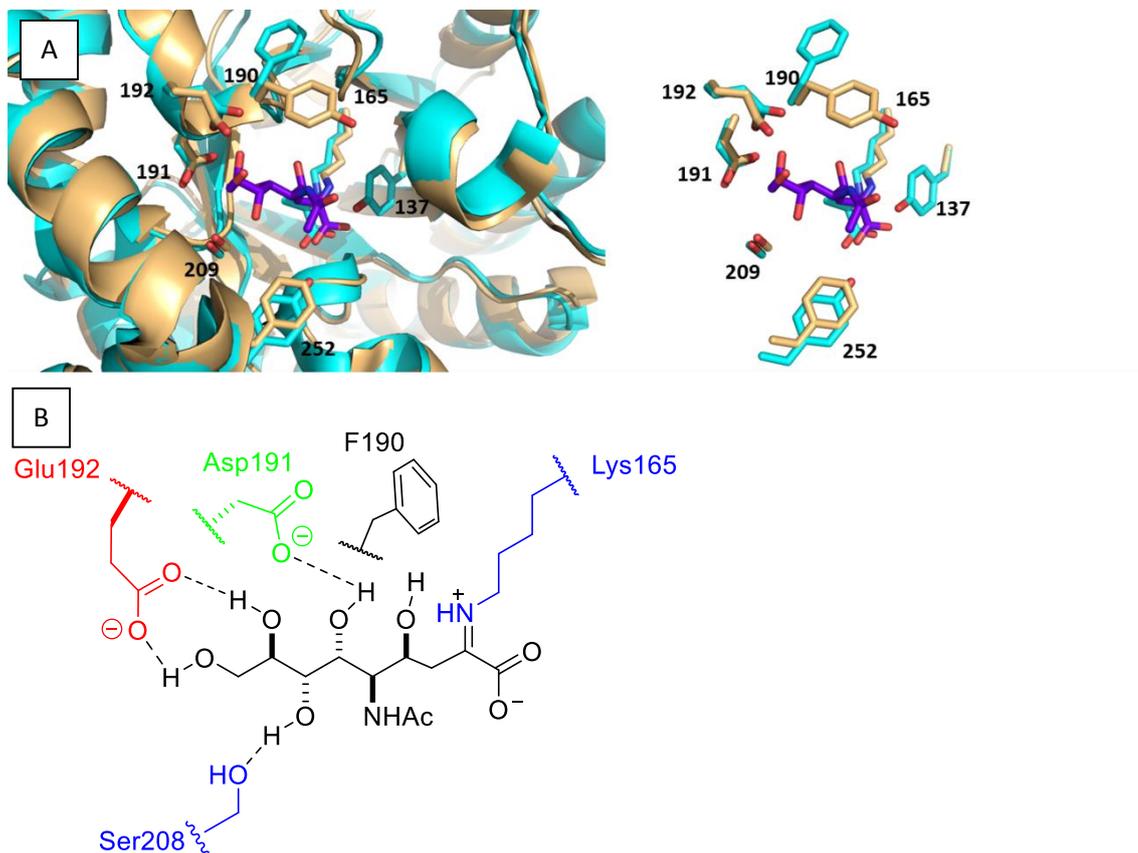
Figure 36 | Proposed binding of *N*-acetylneuraminic acid in the active site of *Sa*NAL. A) The crystal structure of the *E. coli.* NAL Y137A variant (light orange) with full length *N*-acetylneuraminic acid (purple) covalently bound to K165 was overlaid with the active site of *Sa*NAL (cyan) and these crystal structures were used to propose the binding of *N*-acetylneuraminic acid to the active site of *Sa*NAL. **B)** Representation of some important interactions between the substrate *N*-acetylneuraminic acid and the *Sa*NAL active site. It would appear that residue F190 does not directly interact with the substrate but may play a key role in defining the active site of *Sa*NAL. **(A)** was generated using PyMOL (PDB file 4AH7 and 4BWL).[15,78]

3.2.3 Evaluation of the catalytic ability of *Sa*NAL variants, chemically modified at position 190, with the substrate *N*-acetylneuraminic acid

Full kinetic parameters for the *Sa*NAL variants chemically modified at residue 190 together with F190C were established using *N*-acetylneuraminic acid as the substrate. Measurements were performed using 0.1 mg/mL of protein and an increasing concentration of *N*-acetylneuraminic acid ranging from 0 to 25 mM. Michaelis-Menten enzyme kinetics could then be established from the graph of the specific activity against substrate concentration (Figure 37).
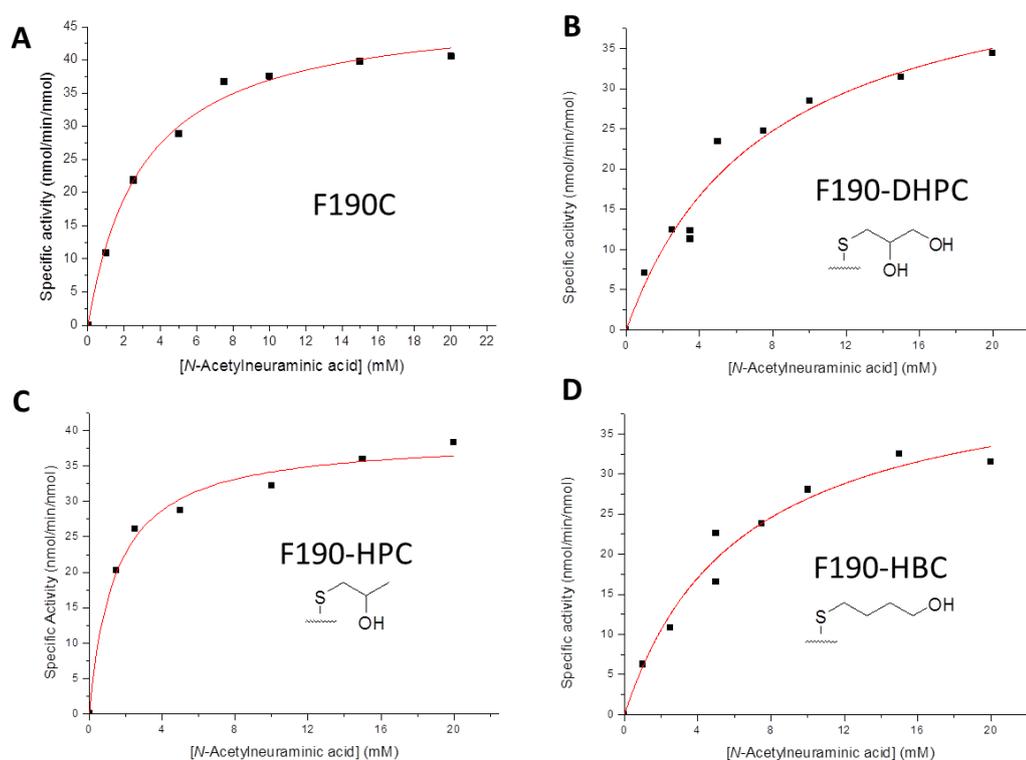
Figure 37 | Activity plots of four *Sa*NAL variants with the substrate *N*-acetylneuraminic acid. A) The site directed *Sa*NAL variant F190C **B)** The chemically modified variant F190-DHPC **C)** The chemically modified variant F190-HPC **D)** The chemically modified variant F190-HBC.

The results from these measurements can be seen in Table 3. All k_{cat}/K_M values were decreased in comparison to the wild-type protein but it would appear the larger F190-DHPC and F190-HBC retarded catalysis more than the slightly shorter residues F190-PC and F190-HPC. The k_{cat} parameters for the variants F190-DHPC and F190-HBC (48.4 ± 5.2 min^{-1} and 44.0 ± 4.0 min^{-1} respectively) are roughly a third of the wild-type k_{cat} value (166.7 min^{-1}). Along with the

decreased k_{cat} values, the K_M values for both of the variants F190-DHPC (7.6 $\pm$ 1.8 mM) and F190-HBC (6.3 $\pm$ 1.4 mM) are $\geq$ 2-fold that of *Sa*NAL (3.2 $\pm$ 0.5 mM). These factors result in a ~10-fold reduction in specific activity when compared to wild-type *Sa*NAL. The variants F190-DHPC and F190-HBC are also less efficient in the catalysis of the cleavage of *N*-acetylneuraminic acid when compared to the variant F190C. The comparable reduction in specific activity observed with the variants F190-DHPC and F190-HBC may stem from steric factors (being the two longest chemical modifications at this position) or perhaps from the electronic properties of the unnatural amino acid side chains, with both containing a terminal primary alcohol. This can be inferred as the two variants F190-DHPC and F190-HBC are ~3 fold more active than F190-PC and F190-HPC and, as well as being shorter, they contain no terminal alcohol.

Both variants F190-PC and F190-HPC exhibit an improvement on catalysis when compared to the variant F190C. The variant F190-PC maintains the K_M of 3.2 mM shown by wild-type *Sa*NAL (3.2 mM) and suffering a relatively small decrease in k_{cat} (72.4 $\pm$ 5.4 min^{-1} compared with 166.7 $\pm$ 9.8 min^{-1} of wild-type *Sa*NAL). On the other hand, the variant F190-HPC suffers a relatively large reduction in k_{cat} (39 $\pm$ 1.35 min^{-1}) but actually exhibits an improvement in K_M (1.4 $\pm$ 0.23 mM) when compared to the wild-type *Sa*NAL protein (3.2 $\pm$ 0.5 mM). This results in a maintenance of ~ 50% of the specific activity of wild-type enzyme. The >2-fold decrease in K_M may be a result of the secondary alcohol of F190-HPC providing extra stability to the *N*-acetylneuraminic acid substrate in complex with the active site.

Table 3 | Table of the kinetic parameters determined for each of the variants of *Sa*NAL chemically modified at residue 190 with *N*-acetylneuraminic acid. Also included are the kinetic parameters determined for the wild type protein *Sa*NAL and the variants E192N and E192C. The errors for k_{cat} and K_M are fitting errors.

		kinetic parameters		
		k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} mM^{-1})
*Sa*NAL variant	NAL	166.7 $\pm$ 9.8	3.2 $\pm$ 0.5	52.1
	E192N	96.5 $\pm$ 20.1	8.7 $\pm$ 3.8	11.1
	F190C	48.14 $\pm$ 1.6	3.0 $\pm$ 0.4	16.0
	F190-DHPC	48.4 $\pm$ 5.2	7.6 $\pm$ 1.8	6.4
	F190-HBC	44.0 $\pm$ 4.0	6.3 $\pm$ 1.4	7.0
	F190-PC	72.4 $\pm$ 5.4	3.18 $\pm$ 0.7	22.6
	F190-HPC	39.0 $\pm$ 1.4	1.4 $\pm$ 0.2	27.3

Figure 38 summarises the k_{cat}/K_M values for the *Sa*NAL variants chemically modified at position 190, and the substrate *N*-acetylneuraminic acid. Broadly speaking it would appear that the effect of modification of residue 190 on the catalysis of the cleavage of *N*-acetylneuraminic acid is less dramatic than modification of the residue 192. Though the effect is less dramatic, it may appear that the larger residues observed with the modifications F190-DHPC and F190-HBC may retard catalysis more than the smaller F190-PC and F190-HPC. It is apparent that upon replacement of the residue F190 with the small hydrophobic group (F190-PC), activity is still somewhat maintained. Also interesting is the large change in catalysis observed between the variants F190-DHPC and F190-HPC as the variation between the two is relatively small, with only the terminal alcohol of F190-DHPC being the variant factor.
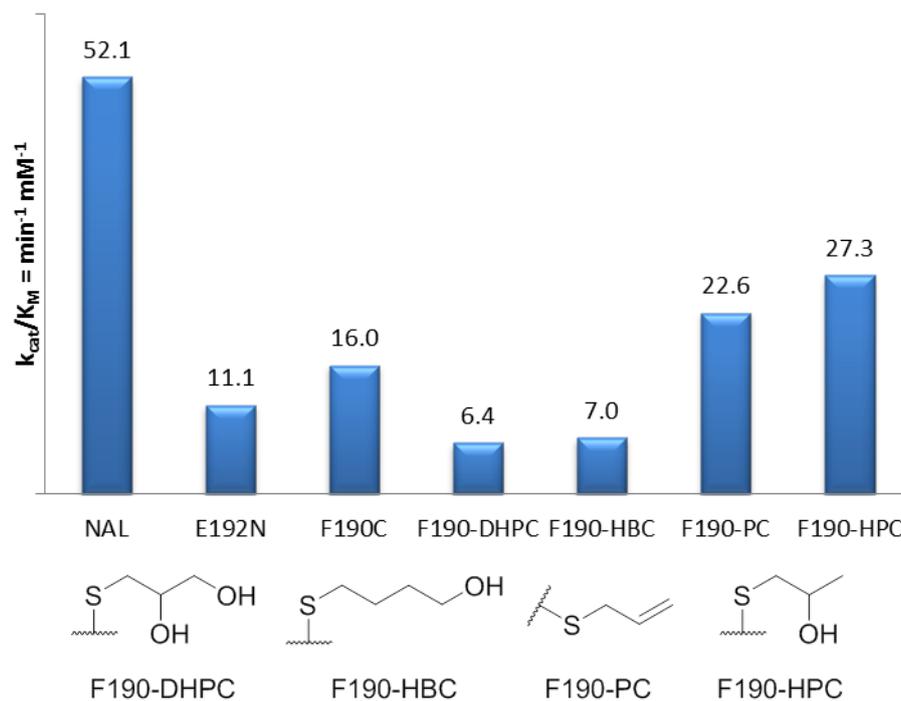
Figure 38 | k_{cat}/K_M (min^{-1} mM^{-1}) for the *Sa*NAL variants chemically modified at residue 190 with the substrate *N*-acetylneuraminic acid.

The preliminary results shown in Figure 38 indicate that further investigation into the chemical modification at position 190 of *Sa*NAL may result in the introduction of interesting catalytic properties to the enzyme. These studies have supported the theory that modifications to the residue 190 would have a less detrimental effect on catalytic function of *Sa*NAL than modification of the residue 192. Figure 39 shows possible reasons for the activities observed with the *Sa*NAL variants F190-HPC and F190-PC.
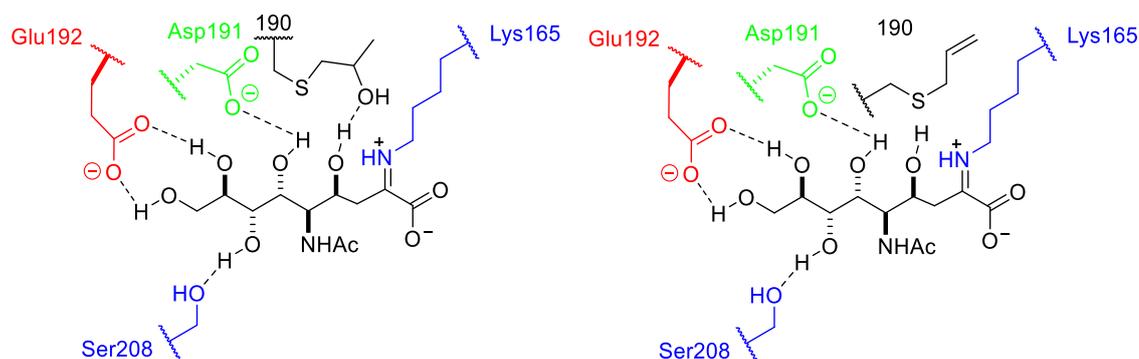
Figure 39 | Hypothetical proposed structure of the active site of the SaNAL variants F190-HPC and F190-PC in complex with the full length substrate *N*-acetylneuraminic acid. The active site has been adapted from the crystal structure of the *E. coli* NAL variant Y137A and the amino acid at residue 190 has been replaced with the unnatural amino acid variants HPC and PC. Additional hydrogen bonding to the full length *N*-acetylneuraminic acid provided by the alcohol of the unnatural amino acid HPC may provide additional recognition of the substrate to the active site, resulting in a decrease in K_M from wild type *Sa*NAL. Conversely the unnatural amino acid PC has similar non-polar properties to the natural F190 residue.

3.2.4 Kinetic characterisation of *Sa*NAL variants, chemically modified at position 190, with the enantiomeric *N*-acetylneuraminic acid analogue as a substrate

Twinned with the investigation into the effects of the configuration of substrate upon catalytic function of the enzyme seen in section 3.1.6, it was decided to investigate the catalytic function of the *Sa*NAL variants with the enantiomer of *N*-acetylneuraminic acid ***ent*-NANA**. The assumption in this case was that catalytic function may be diminished due to the well-established hydrogen bonding between the amino acids with *Sa*NAL and the side chains of *N*-acetylneuraminic acid (Figure 40).

N-acetylneuraminic acid   ***ent*-NANA**
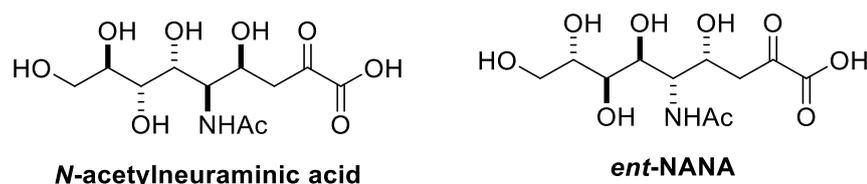
Figure 40 | Comparison of *N*-acetylneuraminic acid with the enantiomer *ent*-NANA. It may be expected that the *Sa*NAL variants would be less efficient at catalysing the retro-aldol reaction with the enantiomer ***ent*-NANA**, as the change in configuration is too great.

Single concentration kinetic measurements were performed with **ent-NANA** using 0.1 mg/mL of modified protein and 0.5 mM of the truncated *N*-acetylneuraminic acid analogue **ent-NANA**. The results of these measurements can be seen in Figure 41.
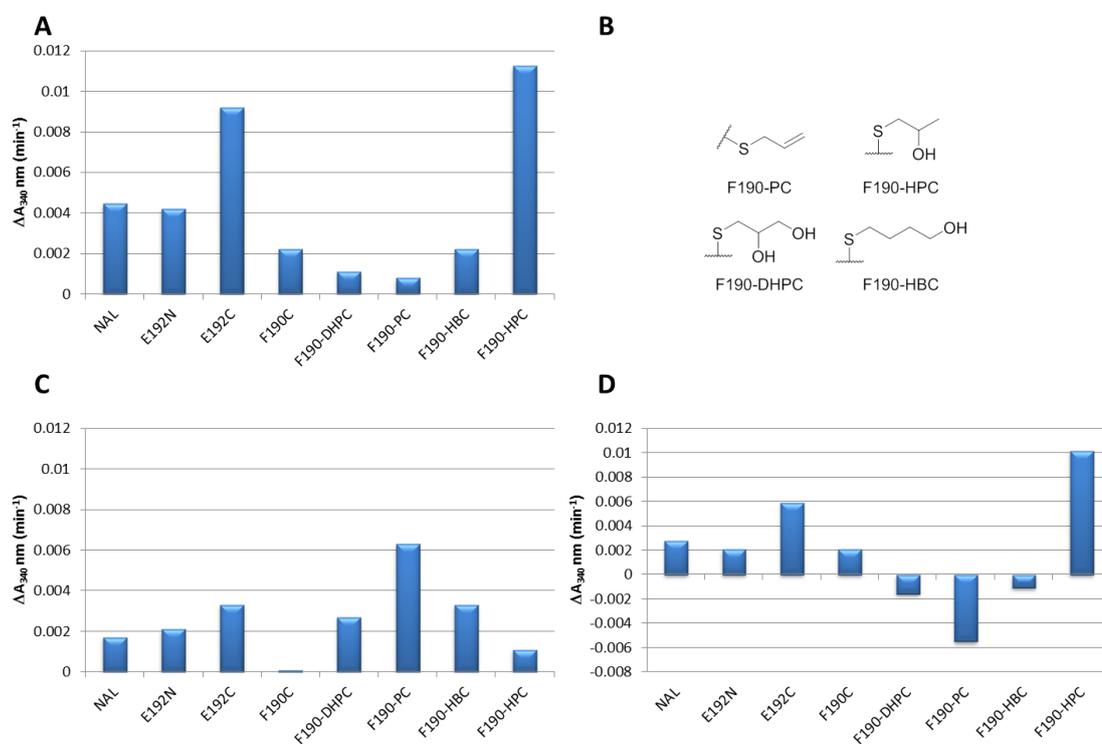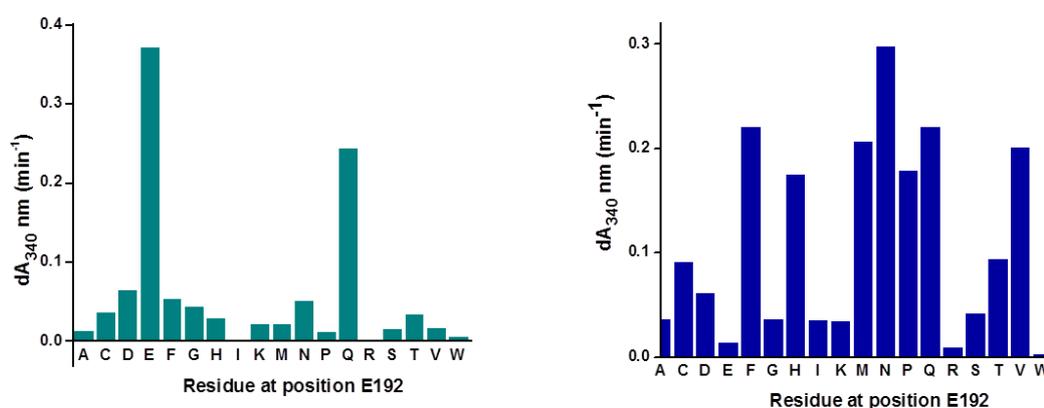

Figure 41 | Single concentration kinetic measurements performed using the enantiomeric *N*-acetylneuraminic acid analogue, *ent*-NANA. A) The uncorrected change in absorbance observed over the period of 1 minute. Each measurement was performed using 0.5 mM of **ent-NANA** and 0.1 mg/mL of modified protein. **B)** The unnatural amino acid side chains of chemically modified proteins used in the assay. **C)** The background change in absorbance, this measurement was taken using 0.1 mg/mL of aldolase and no substrate. **D)** The background corrected change in absorbance observed over one minute.

There is little activity with any of the *Sa*NAL variants with the enantiomer of *N*-acetylneuraminic acid (**ent-NANA**). It is likely that such a drastic change in stereochemistry of the substrate would require much more protein engineering than changing one specific residue within the active site of a protein. The lack of activity is made more obvious when comparing the activity to any of the chemically modified *Sa*NAL variants with *N*-acetylneuraminic acid as the substrate. For instance F190-DHPC with *N*-acetylneuraminic acid as the substrate produced a change in absorbance of 0.4333 over the period of one minute using 0.1 mg/mL of F190-DHPC and 0.5 mM of *N*-acetylneuraminic acid.

3.3 Kinetic characterisation of *Sa*NAL variants, chemically modified at residue 192, with the substrate DPAH

Previous studies of the E192X saturation mutagenesis library against the *N*-acetylneuraminic acid analogue **DPAH** using single concentration kinetic measurements had been performed, Figure 42.[16] The change in absorbance was measured over 1 min with 0.5 mM of DPAH and 0.03 mg of protein. Figure 42 demonstrates that the protein has a high tolerance for modification at position E192 with 7 out of the 18 variants producing a change in absorbance of > 0.1 over the 1 min period. The ability of several variants to accept the substrate contrasts sharply with the *E. coli* NAL E192X library and the substrate *N*-acetylneuraminic acid.

Figure 42 | Single concentration kinetic with an E192X mutant library using 0.5 mM of substrate and 0.03 mg/mL of aldolase. Left) measurements observed with *N*-acetylneuraminic acid. Right) measurements observed with the *N*-acetylneuraminic acid analogue **DPAH**. Figure taken from [16].

The increased promiscuity shown with the *N*-acetylneuraminic acid analogue **DPAH** highlights residue 192 as less integral to the recognition of the substrate than it is to the substrate *N*-acetylneuraminic acid. This can be understood through the properties of the sugar **DPAH**. The substrate *N*-acetylneuraminic acid relies on hydrogen bonding between the carboxylic acid residue of E192 and the substrate. The *N*-acetylneuraminic acid analogue **DPAH** replaces the C-8 and C-9 alcohols of the wild-type substrate with a much less polar di-propyl-amide. The di-propyl functionality, therefore, is unable to form hydrogen bonding interactions with the carboxylic acid of E192. Instead (as determined by analysis of the crystal structure of the E192N variant of *E. coli* NAL in complex with a DPAH analogue) the hydrophobic propyl side chains orientate towards non-polar amino acid side chains (Figure 43). This lack of direct interaction may account for the promiscuity of **DPAH** with variants at position 192.
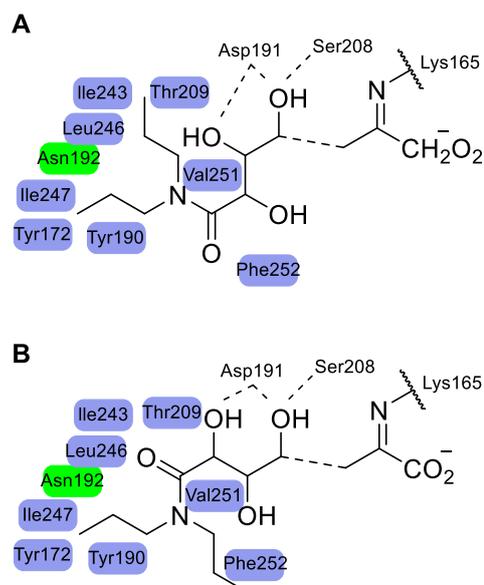
Figure 43 | The active site of the *E. coli* NAL variant E192N in complex with a DPAH analogue.
A) Structure found in three of the four subunits of E192N. **B)** Structure found in one of the four subunits of the crystal structure of the E192N variant of *E. coli* NAL. Figure adapted from [16].

It was therefore perceived that the unnatural *Sa*NAL variants, chemically modified at position 192 may exhibit the same promiscuous features demonstrated by the *E. coli* NAL saturation mutagenesis studies.

3.3.1 Evaluation of the catalytic ability of *Sa*NAL variants, chemically modified at residue 192, with the substrate DPAH

The full kinetic parameters of all of the *Sa*NAL variants at position 192 using **DPAH** as a substrate were determined. The measurements were performed using a fixed amount of protein (0.1 mg/mL) and a range of **DPAH** concentrations (up to 10 mM). The specific activity for each measurement could be calculated and from this the Michaelis-Menten kinetic parameters could be established (Figure 44).
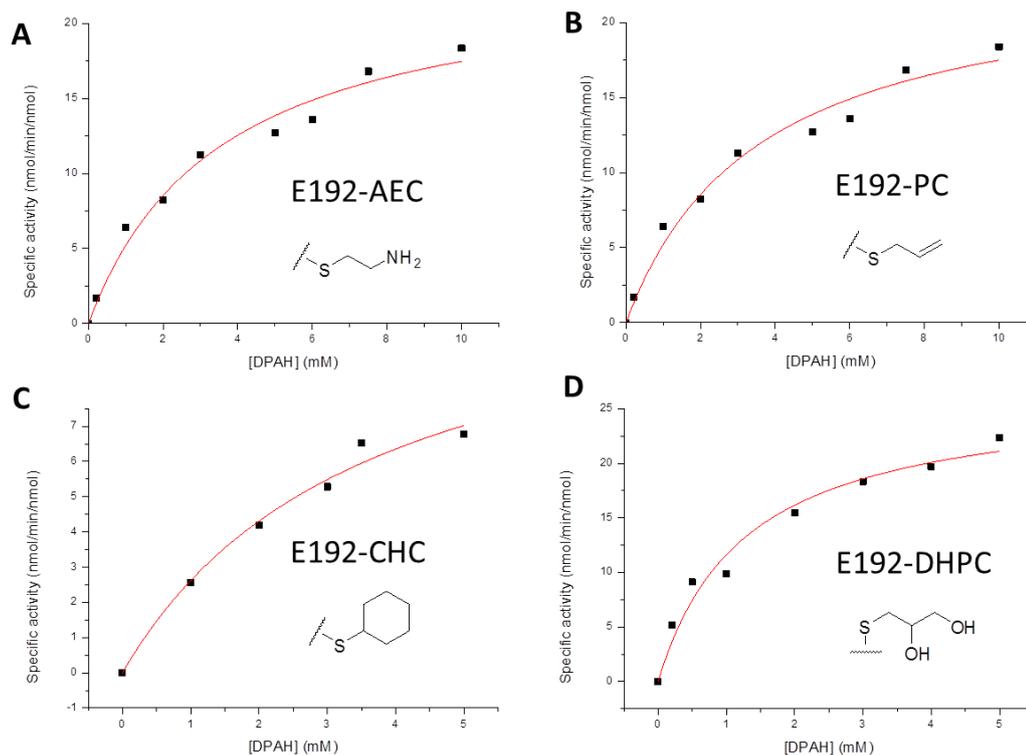
Figure 44 | Activity plots for four chemically modified *Sa*NAL variants with the substrate DPAH. A) E192-AEC **B)** E192-PC **C)** E192-CHC **D)** E192-DHPC.

From these activity plots (Figure 44), it was possible to obtain the k_{cat}, K_M and k_{cat}/K_M values (Michaelis-Menten parameters) for all variants of *Sa*NAL (chemically modified at position 192) and compare them to wild-type *Sa*NAL, along with the variants E192N and E192C. Table 4 shows these kinetic parameters. The most notable observation from the table is that the k_{cat}/k_M results are, although lower than E192N in all cases, highly variable. The results for k_{cat}/k_M ranging from 2.2 $min^{-1}\,mM^{-1}$ for the variant E192-CMC to 20.4 $min^{-1}\,mM^{-1}$ for the variant E192-DHPC.

Table 4 | Table of the kinetic parameters determined for each of the variants of *Sa*NAL chemically modified at residue 190 with the substrate DPAH. Also included are the kinetic parameters for *Sa*NAL and the variants E192N and E192C for comparison. The errors for k_{cat} and K_M are fitting errors.

		Kinetic parameters		
		k_{cat} = min^{-1}	K_M = mM	k_{cat}/K_M = min^{-1} mM^{-1}
*Sa*NAL variant	NAL	24.8 ± 1.1	5.2 ± 0.6	4.8
	E192N	56.5 ± 2.4	1.7 ± 0.2	33.2
	E192C	28.2 ± 1.4	2.3 ± 0.4	12.6
	E192-CMC	12.9 ± 0.6	5.9 ± 0.7	2.2
	E192-DHPC	26.5 ± 2.2	1.3 ± 0.3	20.4
	E192-PC	23.7 ± 2.1	3.5 ± 0.8	6.8
	E192-CPC	15.3 ± 1.4	1.5 ± 0.4	10.2
	E192-AEC	23.7 ± 2.1	3.6 ± 0.8	6.6
	E192-ACEC	24.7 ± 4.7	8.6 ± 2.9	2.9
	E192-HEBC	16.4 ± 2.1	1.9 ± 0.6	8.6
	E192-MCMC	13.6 ± 0.4	1.5 ± 0.2	9.1
	E192-HPC	18.3 ± 3.1	2.6 ± 1.0	7.0
	E192-CHC	12.1 ± 1.9	3.6 ± 1.1	3.4
	E192-HBC	17.8 ± 2.7	1.2 ± 0.5	14.8
	E192-AMCEC	12.2 ± 0.6	2.4 ± 0.3	5.1

It was assumed that *Sa*NAL with the *N*-acetylneuraminic acid analogue **DPAH** as the substrate would be tolerant to chemical modifications at position E192C. Table 4 confirms this, with many of the modifications maintaining catalysis with the *N*-acetylneuraminic acid analogue **DPAH**. The variants E192-DHPC in particular is able to maintain 60% of the specific activity (k_{cat}/K_M) of E192N and E192-HBC is able to maintain ~50% specific activity of E192N (Figure 45).
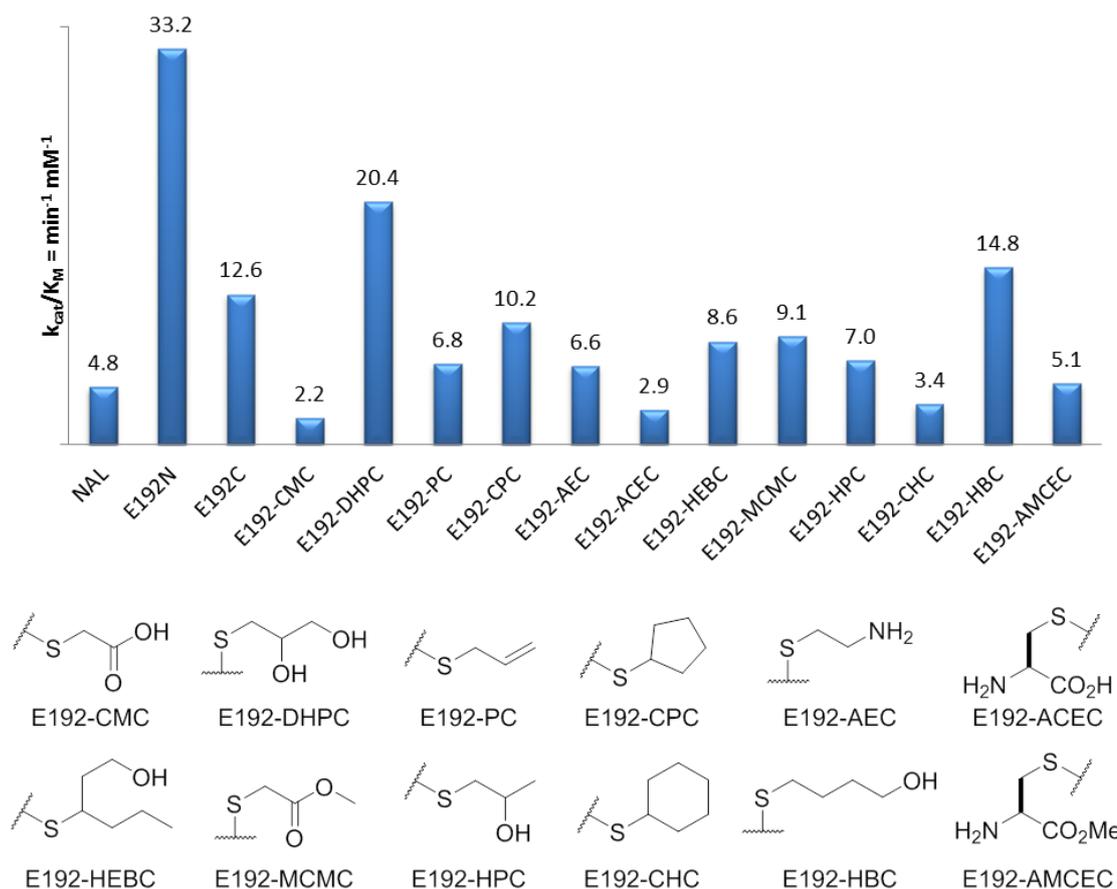
Figure 45 | The k_{cat}/K_M (min^{-1} mM^{-1}) values for all variants of *Sa*NAL chemically modified at position 192 with the substrate DPAH. For comparison the k_{cat}/K_M Values for *Sa*NAL and the variants E192N and E192C are also included.

Most of the unnatural variants of *Sa*NAL chemically modified at position 192 exhibit higher k_{cat}/K_M values compared to the wild-type protein. When compared to wild-type *Sa*NAL, the variants F190-HBC and F190-DHPC exhibit a 3- and 4-fold increase in specific activity with **DPAH** (respectively). Both these variants contain a relatively long side chain with a terminal primary alcohol, which may contribute to the increase in activity.

It is possibly unsurprising that E192-CMC and E192-ACEC exhibit poor kinetic parameters with the substrate **DPAH** as they both contain a formal negative charge at pH7.4 much like the wild type *E. coli* NAL and the *E. coli* NAL variant E192D, tested in single point kinetic measurements (Figure 42). The side chains of E192-CMC and E192-ACEC are lengthened analogue of the amino acids D and E. It would appear that removing this negative charge improves the kinetic parameters of the protein towards the reverse aldol reaction of **DPAH** as shown by the methyl ester forms of E192-CMC and E192-ACEC. E192-MCMC (k_{cat}/K_M = 9.1 min^{-1} mM^{-1}) exhibits a >4

fold improvement in specific activity than the acid variant E192-CMC (k_{cat}/K_M = 2.2 min^{-1} mM^{-1}). Similarly the variant E192-AMCEC (k_{cat}/K_M = 5.1 min^{-1} mM^{-1}) exhibits just under a 2 fold increase in specific activity compared to E192-ACEC (k_{cat}/K_M = 2.9 min^{-1} mM^{-1}).

Interestingly the variant E192-CPC (k_{cat}/K_M = 10.2 min^{-1} mM^{-1}) exhibits a ~ 3-fold increase in specific activity relative to the variant E192-CHC (k_{cat}/K_M = 3.4 min^{-1} mM^{-1}) containing a similar side chain. The increased size of E192-CHC may reduce the binding efficiency of DPAH to the NAL variant.

Although the *Sa*NAL variants chemically modified at residue 192 maintained catalytic function with the substrate **DPAH**, the functionality of the unnatural amino acids used may not be the best for the turnover of **DPAH**. For instance there are no amides in the library and two of the three polar side chains are cyclic, restricting movement and possibly blocking the substrate **DPAH**. Further research into the area may require the addition of more polar amino acid side chains at position 192. An interesting experiment would be the incorporation of the thiol 2-Mercaptoacetamide into position 192, which would lead to a lengthened analogue of the amino acids N and Q.

3.4 Conclusion

This research has shown that the integrated biological and chemical methods of modification of *Sa*NAL is a viable route for the incorporation of unnatural amino acid side chains into the protein. These chemically modified *Sa*NAL variants exhibited varied catalytic activity in the retro-aldol reaction with the substrate *N*-acetylneuraminic acid and the four analogues **ATOA**, **DPAH**, *ent*-**ATOA** and *ent*-**NANA**.

Chemical modifications at position 192 resulted in a sharp decrease in the catalysis of the retro-aldol reaction of the wild-type substrate *N*-acetylneuraminic acid. This result was unsurprising as Amanda Bolt (University of Leeds) had previously demonstrated both the importance of functionality and size of residue 192, of *E.coli* NAL, to the catalysis of the aldol reaction of *N*-acetylneuraminic acid. Interestingly, the chemical modification of residue 190 resulted in more varied kinetic parameters with *N*-acetylneuraminic acid. Both the variants F190-HPC and F190-PC resulted in an enzyme capable of withholding catalytic function with the substrate *N*-acetylneuraminic acid.

As with previous findings (Amanda Bolt, University of Leeds), many of the variants of *Sa*NAL (chemically modified at residue 192) displayed improved catalytic parameters in the retro-aldol reaction of the *N*-acetylneuraminic acid analogue **DPAH**, relative to the wild-type *Sa*NAL protein. This improvement in activity is likely a result of the replacement of the natural negative charge at residue 192.

Unfortunately, no chemically modified variant of *Sa*NAL was able to efficiently catalyse the retro-aldol of the truncated *N*-acetylneuraminic acid analogues **ATOA** and ***ent*-ATOA**. The catalysis of these substrates may be improved through the chemical modification of other residues within the active site of *Sa*NAL.

The engineering of aldolase enzymes able to accept unnatural substrates previously proved to be challenging. Now this work has shown that, using the methods developed within the Berry group (see Section 1.4.3), it is possible to incorporate a range of unnatural functionality into a specific position within the active site of *Sa*NAL. With further investigation, this method may allow for the development of an enzyme capable of catalysing the retro-aldol reaction of unnatural substrates such as truncated substrates.

3.5 Future work

Class I aldolases carry out their function through an active-site catalytic lysine residue which, in the presence of a ketone, is able to form a reactive enamine species. Enamine species are also exploited in organic synthesis by organocatalysts such as proline (Scheme 31).[89-91] MacMillan's imidazolidine catalysts are also utilised in the catalysis of reactions with α,β-unsaturated carbonyl compounds, through reactive iminium intermediates (Scheme 32).[92,93] There are no known examples of such iminium species in natural enzymes, however enzymes are able to exploit iminium species through post-translational modifications (section 1.2).
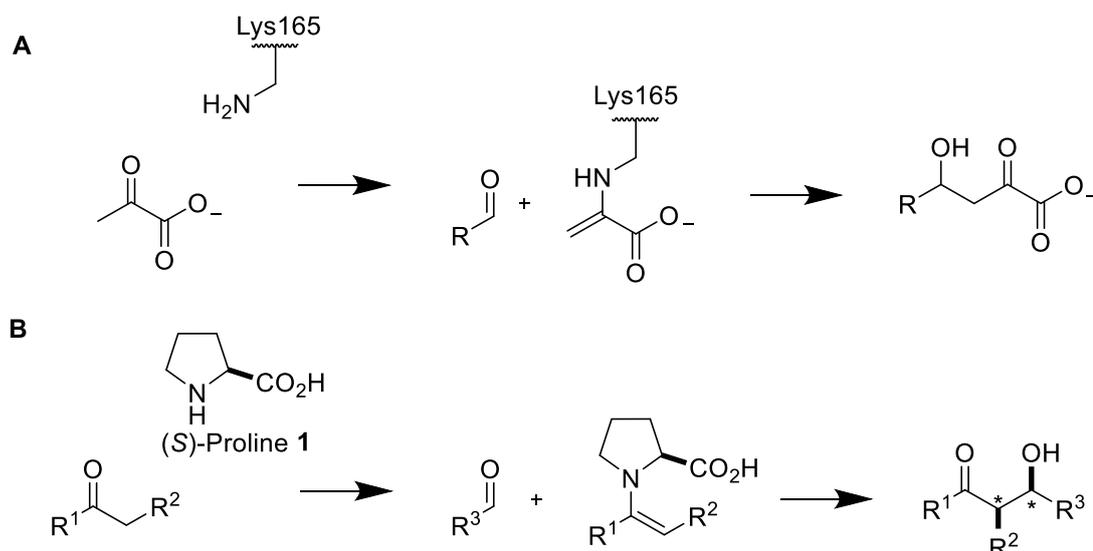
Scheme 31 | Two catalysts that exploit reactive enamine species. A) The Class I aldolase *Sa*NAL. **B)** The amino acid proline can be used in organic synthesis to catalyse aldol reaction. Proline does this by forming a reactive enamine species and as well as providing a proton (from the carboxylic acid). [94]

An ambitious objective, exploiting chemically modified *Sa*NAL variants, would be the alteration of the catalytic function of the protein, away from catalysis of the aldol reaction and toward unnatural iminium catalysis, such as Michael addition or Diels–Alder reactions (Scheme 32). One possible way to achieve this would be by replacing the catalytic lysine residue with a secondary amine derivative. This, twinned with the synthesis of α,β-unsaturated carbonyl compounds, has the potential to result in an iminium species such as Scheme 32. If successful, this approach could be utilised for the catalysis of Diels–Alder and Michael addition reactions.
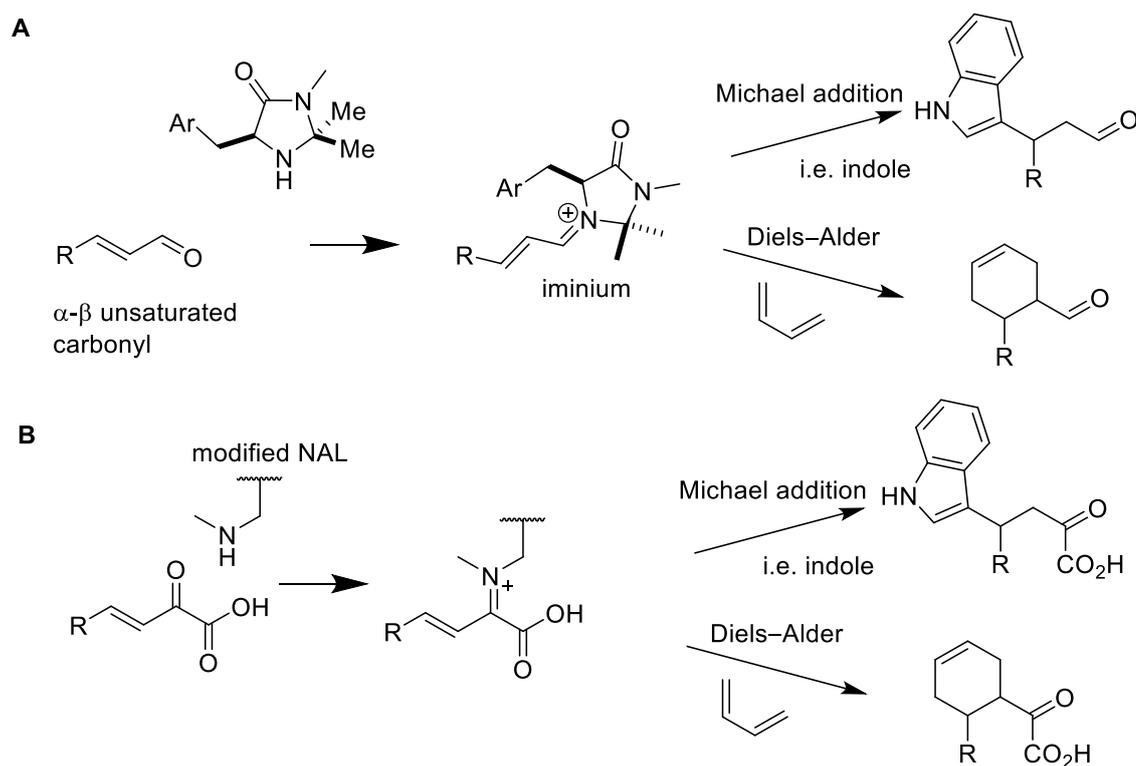
Scheme 32 | Reactions of α,β-unsaturated carbonyl compounds through an iminium intermediate. A) MacMillan organocatalysts have been utilised in the catalysis of reactions with α,β-unsaturated carbonyl compounds.[93,95-97] **B)** Hypothetical modified NAL capable of activating α,β-unsaturated carbonyl compounds towards reactions such as the nucleophilic addition of indole or Diels–Alder reactions. Such reactions could be performed intermolecularly (as illustrated) or intramolecularly.

Initial design of the α,β-unsaturated carbonyl compounds **73**, **74** and **75** was based on the *N*-acetylneuraminic acid analogue **DPAH** (Figure 46). As DPAH is able to tolerant changes in the active site of *Sa*NAL, taking this general structure and using it as a scaffold for the synthesis of α,β-unsaturated carbonyl compounds appeared to be a plausible starting point.
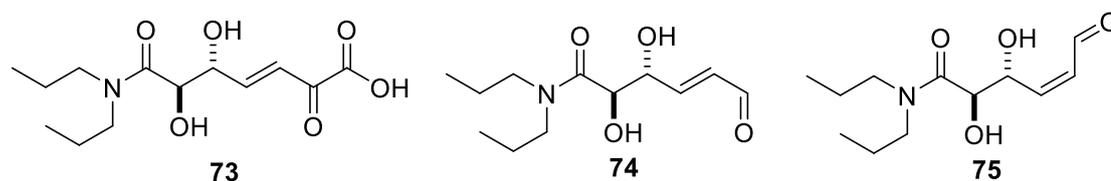
Figure 46 | Three α,β-unsaturated analogues of DPAH.

A variant of *Sa*NAL, K165C/E192N, could be exploited to produce *Sa*NAL variants containing a secondary amine at position of 165. The *Sa*NAL variant E192N would also be included as it would allow for the exploitation of the hydrophobic pocket by the α,β-unsaturated carbonyl compounds. The variant K165C could be exploited by chemical modification for the introduction of amino acid side chains containing secondary amines (Scheme 33).
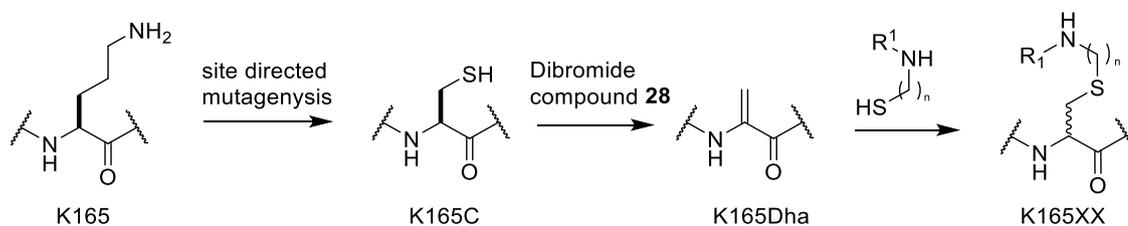
Scheme 33 | Proposed method for the incorporation of secondary amine side chains into residue 165 of *Sa*NAL.

Incubation of the α,β-unsaturated carbonyl compounds in the presence of a 2nd amine containing variant of *Sa*NAL followed by NaBH$_4$ reduction and mass spectrometry analysis could be used to evaluate the presence of a Schiff base (Scheme 34).
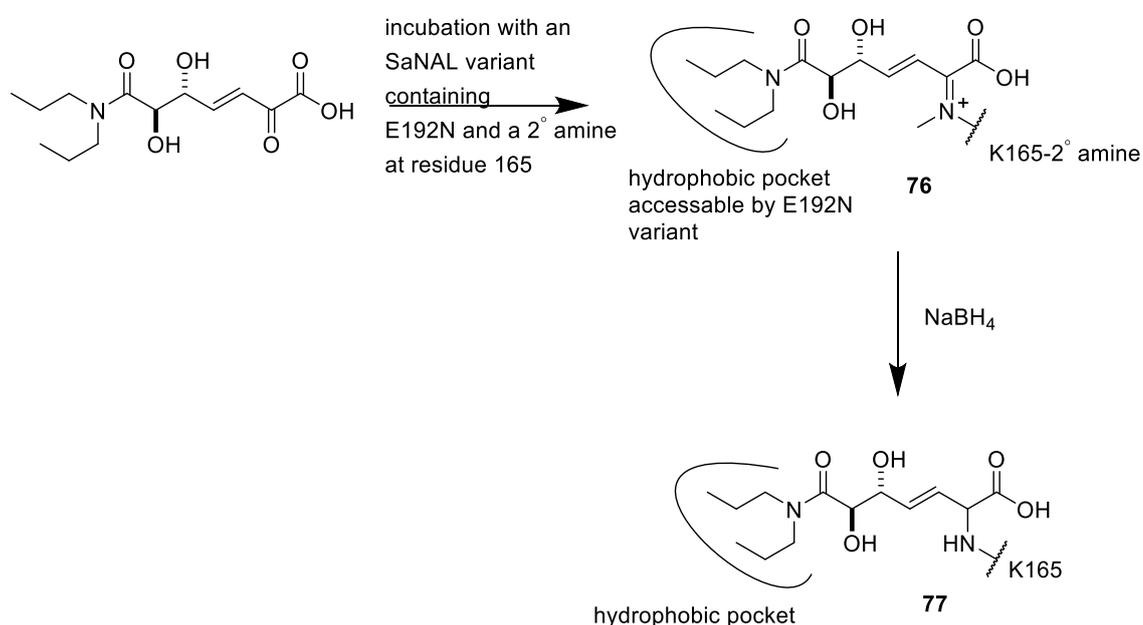
Scheme 34 | Proposed experiment to detect formation of an iminium species covalently bound to the active site. The incubation of an α,β-unsaturated carbonyl compound with a 2° amine variant of *Sa*NAL, followed by treatment with sodium borohydride and mass spectrometry analysis could be used to detect the formation of an iminium species.

After the proof of principle reaction (Scheme 34), many further experiments could be undertaken, possibly resulting in unnatural enzymes being exploited for iminium catalysis. Synthesis of α,β-unsaturated carbonyl compounds **74** and **75** (Figure 46) has begun as well as the synthesis of some thiols with secondary amine components which could be exploited in this future work. For further detail on these routes of synthesis please see Appendix III.

Experimental

4.1 General Biology experimental

4.1.1 Bacterial strains and plasmids

XL10-Gold (Stratagene)

Genotype: Tet^{r}Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F′ *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r) Amy Cam^r]

BL21-Gold (DE3) (Agilent Technologies)

Genotype: *E. coli* B F$^−$ *ompT hsdS*(rB$^−$ mB$^−$) *dcm$^+$* Tet^r *gal* λ(DE3) *endA* Hte

The pK nanA_S.aureus-His$_6$ (Amp^r) plasmid was supplied by Nicole Timms (University of Leeds). The genes for the mutants E192C and F190C were supplied by Claire Windle (University of Leeds) in the form of glycerol stock solutions of XL10-Gold cells.

4.1.2 Chemicals

All chemicals purchased were of analytical grade, and (unless stated otherwise) were used without further purification. 2-Aminoethane thiol HCl, 2-mercapto ethanol, disodium phosphate, hydrogen chloride 37% solution, L-cysteine methyl ester HCl, mercaptoacetic acid salt and monosodium phosphate were purchased from Acros Organics. Ammonium acetate, ampicillin sodium salt, *N,N*-dimethylformamide (DMF), ethanol, glycerol, L-cysteine, methanol, sodium chloride and tris(hydroxymethyl)aminomethane (tris base) were purchased from Fisher Scientific. 1-Mercapto-2-propanol, 1-thioglycerol, 2-furamethane thiol, 4-mercapto-1-butanol, benzyl mercaptan, cyclohexane thiol, cyclopentane thiol, N,N,N′,N′-tetramethylethylenediamine (TEMED), and nicotinamide adenine dinucleotide (NADH) were purchased from Sigma. 3-Mercapto-1-hexanol and sodium hydroxide were purchased from Alfa Aesar. 2-Propene-1-thiol, acrylamide: N,N′-methylenebisacrylamide 24:1 solution and sodium dodecyl sulfate (SDS) were purchased from Fluka. 3-Mercaptopropanoic acid and methyl thioglycolate were purchased from Aldrich. Lactate dehydrogenase was purchased from Roche. DNA size markers were purchased from Promega. Chelating Sepharose Fast Flow™ was purchased from GE Healthcare. Dithiothreitol (DTT) was purchased from

ForMedium. *N*-Acetylneuraminic acid was purchased from CarboSynth. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Generon. Broad range protein marker was purchased from NEB. Agarose was purchased from Melford.

The substrates **ATOA**, **DPAH**, ***ent*-ATOA**, ***ent*-NANA** and the dibromide compound **28** were synthesised by myself (section 4.2).

Stock solution of ampicillin (50 mg/mL) was made up using distilled water, and was sterile-filtered, then stored in 500 µL aliquots at -20 °C. Stock solution of IPTG (50 mg/mL) was made up using distilled water, and was sterile-filtered, then stored in 1 mL aliquots at -20 °C.

4.1.3 Aseptic techniques and sterilisation

Aseptic techniques were employed where applicable, media was autoclaved at 121 °C and 20 psi for 20 minutes before use. Filter-sterilisation was performed using 0.22 µm MiniSart® disposable filters (Sartourius AG).

4.1.4 Bacterial media

2×TY
Tryptone (1.6% w/v), yeast extract (1% w/v), NaCl (0.5% w/v).

2×TY-agar plates
Tryptone (1.6% w/v), yeast extract (1% w/v), NaCl (0.5% w/v), agar (1.5% w/v).

4.1.5 pH measurements

pH measurements were performed using a Jenway 3020 pH meter calibrated according to the manufacturer's instructions.

4.1.6 Centrifugation

Centrifugation of solutions > 1.5 mL were performed using an Avanti® J-26XP centrifuge (Beckman Coulter). Centrifugation of solutions ≤ 1.5 mL were performed using a desktop MSE micro centaur centrifuge (MSE).

4.1.7 Culture growth

DNA from glycerol stock solutions of XL10-Gold cells containing the pK nanA_S.aureus-His$_6$ (Amp^r) plasmid (wild type and mutants) was transformed into BL21-Gold (DE3) cells for protein expression (section 4.1.10 followed by section 4.1.13).

BL21-Gold (DE3) glycerol stock solutions of either SaNAL or the variants E192C, E192N and F190C were used to inoculate 5 mL 2×TY starter cultures supplemented with 50 μg/mL ampicillin. The starter cultures were incubated overnight in an orbital shaker-incubator (37 °C, 200 rpm). 5 μL of the starter culture was then used to inoculate 5 mL 2×TY media supplemented with 50 μg/mL ampicillin. This was left to incubate for 8 h in an orbital shaker-incubator (37 °C, 200 rpm). 10 mL of this 100 mL culture was then used to inoculate 1 L of 2xTY media supplemented with 50 μg/mL ampicillin. When the optical density of the culture at 600 nm (OD$_{600nm}$) reached 0.6, IPTG (100 μM final concentration) was added to induce protein expression.

4.1.8 Glycerol stock solutions

0.5 mL of a 5 mL starter culture and was added to 0.5 mL of sterile glycerol in a Nunc CryoTube™ vial (Thermo Fisher Scientific) and stored at -80 °C.

Glycerol stock solutions were also donated by Claire Windle and Nicole Timms (University of Leeds).

4.1.9 Agarose gel electrophoresis

Nucleic acids > 1.5 kb were resolved using 0.7 % (w/v) agarose gel electrophoresis and electrophoresed alongside a 1 kb DNA ladder. A photo of the gel could then be taken (using UV light) with an InGenius gel imager (Syngene).

TAE buffer (1L of 10 x stock)

48.4 g Tris base

11.4 mL glacial acetic acid (17.4 M)

3.7 g EDTA, disodium salt

deionized water to 1L

4.1.10 Plasmid DNA purification

Wizard® plus SV minipreps DNA purification was used to purify DNA and was used in accordance with the manufacturers' instructions.

4.1.11 Site-Directed Mutagenesis

Site-directed mutagenesis was carried out using a QuickChange® Lightning Mutagenesis Kit (Agilent) according to manufacturers' instructions. Primers had been previously designed, by Nicole Timms (University of Leeds), and can be seen in the appendix. Following mutagenesis, and dpn1 digest of methylated DNA, transformation of the mutagenized DNA into XL10-Gold cells was performed. Purification of the plasmid DNA and sequencing (Beckman Coulter Genomics) confirmed that mutagenesis was successful.

4.1.12 DNA sequencing

DNA sequencing was provided by Beckman Coulter Genomics, sequencing using the primers 'pKK For' and 'pKK Rev' (primers were provided by Beckman Coulter Genomics).

4.1.13 Transformation of XL10-Gold Ultracompetent Cells

A 45 µL aliquot of XL10-Gold ultracompetent cells was thawed on ice. The cells were transferred to a pre-chilled 14 mL BD Falcon polypropylene round-bottom tube. 2 µL β-mercaptoethanol was added to the cells, the contents swirled, then incubated for 10 mins. 1.5 µL of DNA was then added to the cells and the solution incubated for 30 mins. The cells were then heat-pulsed at 42 °C for 30 seconds, followed by immediate incubation on ice for 2 mins. Preheated (42 °C) 2×TY media (0.5 mL) was added to the cells and the solution incubated for 1 h at 37 °C with agitation at 225 rpm. The transformed cells were then plated on 2xTY-agar plates containing 50 µg/mL ampicillin, and incubated overnight at 37 °C. Single colonies were picked and grown in selection media overnight. These cells were then used to make up glycerol stock solutions.

Transformation into BL21-Gold (DE3) cells follows the same procedure, with two amendments. No β-mercaptoethanol was used, and the cells were heat-pulsed for only 20 seconds.

4.1.14 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The SDS-PAGE gels were made up of both a 'resolving gel' and a 'stacking gel'. The gels were set between two glass plates. The 'resolving gel' was added to the plates first and covered with isopropanol until the gel had polymerised. At this point the isopropanol was removed and the polymerised 'resolving gel' was topped with 'stacking gel' and a 10- or 15-well comb. Upon complete polymerisation the gel could be used immediately or stored in moist lab roll and cling film at 4 °C for up to 1 week.

Resolving gel:
7,500 µL 30% (w/v) acrylamide
3,750 µL 1.5 M Tris/HCl pH 8.8
150 µL 10% (w/v) SDS
3,500 µL H_2O
50 µL 25% (w/v) ammonium persulphate (freshly made and added last)
5 µL TEMED (added last)

Stacking gel:
625 µL 30% (w/v) acrylamide

625 µL 1.0 M Tris/HCl pH 6.9

50 µL 10% (w/v) SDS

3,650 µL H_2O

50 µL 25% (w/v) ammonium persulphate (freshly made and added last)

5 µL TEMED (added last)

<u>SDS running buffer:</u>

3 g Tris base

14.4 g glycerol

1 g SDS

140 µL β-mercaptoethanol

1 L with H_2O

<u>Sample buffer:</u>

308.5 mg dithiothreitol

400 mg SDS

20 mg bromophenol blue

2 mL glycerol

50 mM Tris/HCl pH 6.8 up to 20 mL

Protein samples were mixed with sample buffer 1:1 ratio and boiled for 5 min before loading onto a gel. One lane would be loaded with an NEB broad range protein marker. Electrophoresis was performed at 30-60 mA for 1 h. Samples were then stained with coomassie for a minimum of 1 h and then destained in deionised water for a minimum of 1 h. A photo of the gel could then be taken using an InGenius gel imager (Syngene).

4.1.15 Nickel affinity Purification of His$_6$-tagged proteins from *E. coli*

Expressed His$_6$-tagged proteins were purified by Chelating Sepharose™ fast flow resin, loaded with 0.2M NiCl$_2$.

Centrifugation of 1 L bacterial culture (12,000 × g for 20 min) provided a cell pellet. Growth medium was decanted off and the pellet re-suspended in 30 mL washing buffer *via* the use of a homogeniser. The re-suspended pellet was lysed using a TS series Benchtop Cell Disruptor (Constant Systems Ltd.). Centrifugation of the lysed cells (40,000 × g for 45 min at 4 °C) was followed by loading of the supernatant (containing the soluble protein) onto 5 mL of settled

Ni-NTA resin. The solution was agitated at 4 °C for 1h followed by centrifugation (2,200 × g for 6 min at 4°C) and the supernatant decanted. The resin was then washed 4 times with 40 mL of washing buffer, each wash required centrifugation (2,200 × g for 6 min at 4 °C) and removal of the resulting supernatant. 20 mL of elution buffer was then added to the resin and the solution incubated for 30 min at 4 °C with agitation, followed by centrifugation (2,200 × g for 6 min at 4 °C). The purified protein was then dialysed (section 4.1.16) in 50 mM Tris/HCl, followed by two rounds of dialysis into 20 mM ammonium acetate. The protein concentration was measured (section 4.1.17) and the protein split into 2.5, 5 or 10 mg aliquots. These aliquots were flash frozen (liquid nitrogen), then lyophilised, and the protein stored at -20 °C.

Washing buffer:

50 mM Tris/HCl pH7.4
20 mM Imidazole
0.5 M NaCl

Elution buffer:

50 mM Tris/HCl pH7.4
0.5 M Imidazole
0.5 M NaCl

4.1.16 Protein dialysis

Proteins were dialysed in 20 mM Tris/HCl buffers. Dialysis tubing was supplied by Fisher Scientific Limited and had a molecular weight cut off of 10 kDa. The volume of dialysis buffer was 1000 times that of the solution being dialysed. Each dialysis involved two rounds of gentle stirring of the dialysis buffer at 4 °C for 4 h.

4.1.17 Determination of protein concentration

Protein concentration was determined using the Beer–Lambert law measuring at 280nm, or using a Bradford assay which involved incubating ($\geq$ 5 mins) 980 µL Bradford reagent with 2µL protein solution and 18 µL buffer, and measuring the absorbance at 595 nm.

A Kontron Uvicon 930 spectrophotometer was used to measure absorbance values. Bovine serum albumin (BSA) was used as a standard for Bradford assays.

4.1.18 Increasing protein concentration

Protein concentration could be increased using Vivaspin 20 mL centrifugal concentrators (according to the manufacturer's instruction). The concentrator tubes were equilibrated with water followed by the appropriate buffer before use (according to the manufacturer's instruction).

4.1.19 Mass spectrometry

Mass spectrometry samples were suspended in 20 – 50 mM ammonium acetate. Buffer exchange of samples into ammonium acetate was achieved using Micro Bio-spin^{TM} chromatography columns (Bio-Rad Laboratories). The columns were first equilibrated according to manufacturers' instructions. Protein mass spectrometry was then performed on these samples by Dr. James Ault (University of Leeds) according to the following specification:

Nanoelectrospray mass spectrometry was performed using a quadrupole-ion mobility spectrometry-orthogonal time-of-flight (TOF) mass spectrometer (Synapt HDMS, Water UK Ltd.). Positive electrospray ionisation was used. The mass spectrometer was operated in positive TOF 'V' mode using a capillary voltage of 1.2 kV, cone voltage of 50 V, nanoelectrospray nitrogen gas pressure of 0.1 bar and backing pressure of 1.78 mbar. The source and desolvation temperatures were set at 80 °C and 150 °C, respectively. Nitrogen was used as buffer gas at a pressure of 8.0×10^{-3} mbar in the trap and transfer regions, and 3.6×10^{-4} mbar in the ion-mobility spectrometry (IMS) cell. Mass calibration was performed by a separate injection of sodium iodide at a concentration of 2 mg/mL in acetonitrile:water (1:1; v/v). Data processing, including deconvolution of multiple mass-charge state spectra, was performed using MassLynx v4.1 software (provided with the mass spectrometer).

4.1.20 Lyophilisation

Solutions were flash-frozen using liquid nitrogen. Lyophilisation was achieved using a Heto PowerDry PL3000 Freeze Dryer, and a ThermoSavant VLP120 ValuPump.

4.1.21 Lactate dehydrogenase coupled enzyme assay

Kinetic measurements of the aldol cleavage reaction were performed using a standard lactate dehydrogenase (LDH) coupled reaction (Scheme 35).
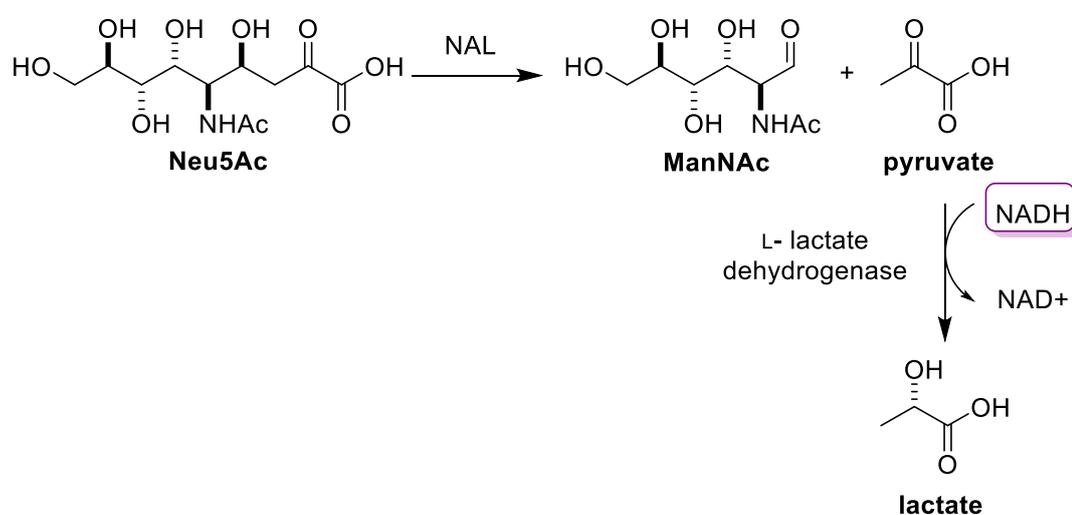
Scheme 35 | Coupled enzyme assay used for measuring the kinetic parameters of SaNAL.

The measurements were undertaken in a Tris/ HCl pH7.4 buffer at 30 °C. Each solution consisted of 0.5 units of LDH, 0.2 mM NADH, 0.03 or 0.1 mg/mL of aldolase enzyme, 0.0 – 4.5 mM of substrate and the Tris/ HCl buffer making up the solution to 1 mL. Upon transfer to the 1 mL cuvette the sample was inverted 4 times and transferred immediately to the absorbance spectrometer. Absorbance was measured at 340 nm over the period of 1 min. Kinetic parameters k_{cat}, K_M and k_{cat}/K_M were determined by non-linear regression analysis.

4.1.22 General Method for the chemical modification of the cysteine containing *Sa*NAL variants E192C or F190C

Lyophilised cysteine-containing variants of *Sa*NAL (2.5 mg – 10 mg) were dissolved in 50 mM sodium phosphate buffer (pH 8.0) containing 6 M urea, resulting in a protein concentration of 2 mg/mL. This solution was vortexed for 10 seconds. A 0.44M solution of the dibromide compound **28** in DMF was added (60 eq.) to the protein solution, and this solution was vortexed for 5 seconds and incubated with agitation (200 rpm) at 37 °C for 90 mins. A 0.85 M solution of thiol (100 eq) in 2 M Tris HCl, 0-7.5 % DMF (to aid solubility) was added to the reaction mixture and the solution was vortexed for 10 seconds, followed by incubation with agitation (200 rpm) at 37 °C for 120 mins. The reaction mixture was then dialysed (section 4.1.16), into 50 mM sodium phosphate with 6 M urea (2 rounds of dialysis), 50 mM sodium phosphate (two rounds of dialysis), followed by at least 2 rounds of dialysis into 50 mM Tris HCl pH 7.4. The protein was then concentrated to a volume of $\geq$ 2.5 mL (section 4.1.18), and concentrated protein was purified using size exclusion chromatography (50 mM Tris HCl buffer pH 7.4). The eluent was then concentrated (Section 4.1.18), the protein concentration was determined (section 4.1.17) and the protein stored at 4 °C ready for use in kinetic measurements. Protein was characterised by HR-MS (characterisation performed by Dr. James Ault, University of Leeds). Typically from a 10 mg conversion of protein you could expect ~2-4 mg recovery of chemically modified protein.

4.1.23 Refolding of modified variants

During the chemical modification process the protein is unfolded due to the presence of 6 M urea. To refold the protein the solution is first dialysed twice against a buffer containing 50 mM sodium phosphate pH 8.0, 6 M urea for 4 h at RT, and then twice more in a urea-free buffer.

4.1.24 Size Exclusion Chromatography

Gel filtration of proteins was performed using a Superdex S200 column attached to an ÄKTA prime at 4°C. The protein was eluted in 2 mL fractions using (degassed) 50 mM Tris phosphate (pH7.4) buffer.

4.2 Chemistry Experimental

All non-aqueous reactions were carried out under an atmosphere of nitrogen. Water-sensitive reactions were performed in oven-dried glassware cooled under nitrogen before use. Solvents were removed under reduced pressure using a Büchi rotary evaporator and a Vacuubrand PC2001 Vario diaphragm pump. The removal of water was achieved by lyophilises using a VirTis benchtop K freeze dryer and a Vacuubrand PC2001 Vario diaphragm pump.

Dry solvents were obtained using a solvent purification system (Innovative Technology Pure Solve[TM]). Ether refers to diethyl ether and petrol refers to petroleum spirit (b.p. 40-60 °C). *N*-Bromosuccinimide was recrystallised from water and dried under vacuum prior to use. All other solvents and reagents were of analytical grade and used as supplied. Commercially available starting materials were obtained from Sigma–Aldrich, Lancaster, Alfa Aesar, Novabiochem, Carbosynth or Fluorous Technologies.

Flash column chromatography was carried out using silica (35-70 μm particles) according to the method of Still, Kahn and Mitra. [98] Thin layer chromatography was carried out on commercially available pre-coated aluminium plates (Merck silica Kieselgel 60F$_{254}$).

Analytical LC-MS was performed using an Agilent 1200 series LC system comprising of a Bruker HCT Ultra ion trap mass spec, a high vacuum degasser, a binary pump, a high performance autosampler and micro well plate autosampler, an autosampler thermostat, a thermostated column compartment and diode array detector. The system used two solvent systems: MeCN/H$_2$O + 0.1% Formic acid with a Phenomenex Luna C18 50 x 2mm 5 micron column or. MeCN/H$_2$O with a Phenomenex Luna C18 50 x 2mm 5 micron column.

Proton and carbon NMR spectra were recorded on a Bruker Advance DPX300, Advance 500 or DRX500 spectrophotometer using an internal deuterium lock. Carbon NMR spectra were recorded with composite pulse decoupling using the waltz 16 pulse sequence. DEPT, COSY, HMQC and HMBC pulse sequences were routinely used to aid the assignment of spectra. Chemical shifts are quoted in parts per million downfield of tetramethylsilane, and coupling constants (*J*) are given in Hz. NMR spectra were recorded at 300 K unless otherwise stated.

Melting points were determined on a Reichert hot stage microscope and are uncorrected.

Nominal and accurate mass spectrometry using electrospray ionisation was carried out by staff in the School of Chemistry at the University of Leeds, using either a Micromass LCT-KA111 or Bruker MicroTOF mass spectrometer.

Optical activity measurements were recorded at room temperature on an Electron Schmidt + Haensch Polartronic H532 polarimeter; units for $[\alpha]_D$ are 10^{-1} deg cm^2 g^{-1} and are omitted.

Infrared spectra were recorded on a Perkin Elmer spectrum One FT-IR spectrophotometer.

Ion exchange chromatography was performed using DOWEX 1X8, 200-400 mesh resin.

DOWEX resins were activated (according to the suppliers guide lines) before use.

Ozone was generated using a Welsbach T-816 generator (0.4-0.6 psi), and the reaction was purged before and after the reaction with oxygen gas (0.4-0.6 psi) for a minimum of 20 mins.

Experimental Data:

2,5-Dibromohexanediamide 28.[99]
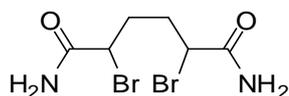
Adipic acid (5.23 g, 35.8 mmol) was added to thionyl chloride (15.0 mL, 207 mmol) and heated at reflux for 1.5 h, cooled to room temperature and CCl_4 (20 mL), NBS (15.0 g, 84.3 mmol) and 48% aqueous HBr (one drop) were added and solution heated at reflux for a further 4 h. The solution was then stirred at 0 °C for 30 mins, filtered, and the filtrate reduced *in vacuo* to give a crude red oil. The crude oil was added drop wise at 0 °C over a period of 20 min to a solution of 35 % aqueous NH_4OH (40 mL) and left to stir for 1 h, filtered and reduced *in vacuo*. The crude solid was crystallised from H_2O–MeOH to yield the dibromide **28** (5.2 g, 50%) as large flake crystals; m.p. 192-194 °C (from H_2O–MeOH) [lit.[100] 196 °C]; υ_{max}/cm^{-1} (solid) 3188, 2947, 2800 and 1666; δ_H (500 MHz; DMSO) 7.71 (2H, s, NH_2), 7.33 (2H, s, NH_2), 4.38-4.28 (2H, m, 2 × CHBr), 2.08-1.75 (4H, m, CH_2CH_2); δ_C (75 MHz; DMSO) 169.81 (CO), 169.75 (CO), 48.45 (CHBr), 48.17 (CHBr), 32.54 (CH_2), 32.41 (CH_2); m/z (ES) MNa^+ 322.0 (50%), 324.9 (100%), 326.9 (50%); HRMS Found M+Na: 324.8993. $C_6H_{10}Br_2O_2$ requires M+Na, 324.8986.

(*E*)-*N*-(2-Ethoxy-2-oxoethylidene)-2-methylpropan-2-amine oxide 38. [81,101]
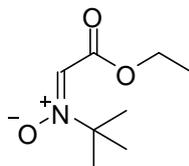
N-alkylhydroxylamine HCl (0.63 g, 5 mmol) and, 50% in toluene, ethyl glyoxylate solution (1.02 g, 5 mmol,) were added to a solution of sodium bicarbonate (0.84 g 10 mmol) in toluene (10 mL) and the reaction mixture was stirred at room temperature overnight. The solid was filtered and the filtrate was concentrated *in vacuo* to yield **38** (0.85 g, 98 %), with no further purification, as a pale yellow oil. R_f: 0.87 (9:1, DCM–MeOH); υ_{max}/cm^{-1} (film) 3392.17, 1724.07 and 1624.55; δ_H (500 MHz; CD$_3$OD) 7.25 (1H, s, 1-H), 4.23 (2H, q, *J* 7.1, ethyl 1-H), 1.51 (9H, s, ^{t}butyl 2-H$_3$), 1.29 (3H, t, *J* 7.1, ethyl 2-H$_3$); δ_C (75 MHz; CD$_3$OD) 160.7 (2-C), 121.3 (1-C), 74.6 (ethyl 1-C), 60.7 (^{t}butyl 1-C), 28.1 (^{t}butyl 2-C), 14.1 (ethyl 2-C); HRMS Found M+H: 174.1126. C$_8$H$_{15}$NO$_3$ requires M+H, 174.1125.

Methyl-5-acetamido-4-hydroxy-2-methoxy-6-((7*R*,8*R*)-7,8,9-trihydroxypropyl)tetrahydro-2*H*-pyran-2-carboxylate 42. [82]
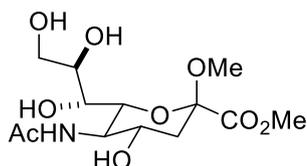
N-Acetylneuraminic acid (10 g, 32 mmol) and Dowex H$^+$ resin (23 g) were stirred in MeOH (40 ml) at reflux for 48 h. The reaction mixture was cooled to room temperature, filtered, reduced *in vacuo* to give the crude product which was recrystallised from MeOH–ether to yield the sugar **42** (5.2 g, 50 %) as large plate crystals, m.p. 59-60 °C (EtOAc–petrol); R_f: 0.67 (0.8:2:1, H$_2$O–^{i}PrOH–EtOAc); $[\alpha]_D^{20}$: –37.6 (*c*. 0.7 in MeOH); υ_{max}/cm^{-1} (solid) 3600-3050, 2935, 1634 and 1036; δ_H (500 MHz; D$_2$O) 3.97 (1H, app. dt, *J* 10.2 and 5.1, 4-H), 3.87-3.74 (7H, m, 5-H, 6-H, 7-H, 9-H$_A$, and methyl ester Me), 3.59 (1H, dd, *J* 12.0 and 5.7, 9-H$_B$), 3.51 (1H, app. d, *J* 9.6, 8-H), 3.18 (3H, s, methyl ether 1-H$_3$), 2.32 (1H, dd, *J* 13.1 and 5.1, 3-H$_{eq}$), 1.97 (3H, s, acetyl Me), 1.71 (1H, dd, *J* 13.1 and 10.2, 3-H$_{ax}$); δ_C (75 MHz; D$_2$O) 174.7 (methyl ester CO), 170.3 (acetyl CO), 99.1 (2-C), 70.4, 69.7, 67.8 (8-C), 66.3 (4-C), 63.2 (9-C), 53.4 (methyl ester Me), 51.6, 50.9

(methyl ether Me), 39.1 (3-C), 22.0 (acetyl Me); HRMS Found M+Na: 360.1276, $C_{13}H_{23}NO_9$ requires M+Na, 360.1265.

(2*R*,4*S*,5*R*,6*R*)-Methyl-5-acetamido-4-hydroxy-6-(hydroxymethyl)-2-methoxytetrahydro-2*H*-pyran-2-carboxylate 46.[82]
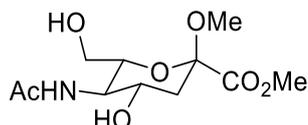
A solution of periodic acid (444 mg, 1.95 mmol) in water (25 mL) was added dropwise to a solution of Sugar **42** (250 mg, 0.74 mmol) in water (25 mL) over 10 min. The solution was stirred for a further 20 min at room temperature. The reaction mixture was then passed down a dowex 1X8 ion exchange column eluting with water (25 mL). The eluent was then reduced *in vacuo* and the resulting solid dissolved in methanol (5 mL) and cooled to 0 °C. NaBH₄ (64.7 mg, 1.71 mmol) was added over 5 min and the solution stirred for 20 min. After which dowex H⁺ resin was added to achieve pH 6, the solution was filtered, washed with methanol (5 mL) and the filtrate was then passed through a dowex 1X8 resin eluting with methanol (20 mL). The eluent was reduced *in vacuo* to yield the alcohol **46** (190.0 mg 95%) as a pale yellow glass solid, $[\alpha]_D^{23}$: −39.4 (*c*. 0.9 in MeOH); υ_{max}/cm^{-1} (solid) 3700-3100, 2895 and 1638; δ_H (500 MHz; D₂O) 3.93 (1H, ddd, *J* 11.6, 9.9 and 5.2, 4-H), 3.79 (3H, s, methyl ester Me), 3.74-3.51 (4H, m, 5-H, 6-H and 7-H), 3.16 (3H, s, methyl ether Me), 2.33 (1H, dd, *J* 13.2 and 5.2, 3-H$_{eq}$), 1.96 (3H, s, *J* 1.96, acetyl Me), 1.71 (1H, dd, *J* 13.2 and 11.6, 3-H$_{ax}$); δ_C (75 MHz; D₂O) 174.7 (methyl ester CO), 170.0 (acetyl CO), 98.8 (2-C), 73.1 (6-C), 66.0 (4-C), 60.6 (7-C), 53.4 (methyl ester Me), 51.8 (5-C), 50.7 (methyl ether Me), 39.0 (3-C), 21.9 (acetyl Me); HRMS Found M-H: 276.1099. $C_{11}N_{19}NO_7$ requires M-H, 276.1089.

(2*R*,4*S*,5*R*,6*R*)-5-Acetamido-2,4-dihydroxy-6-(hydroxymethyl)tetrahydro-2*H*-pyran-2-carboxylic acid (ATOA).[82]
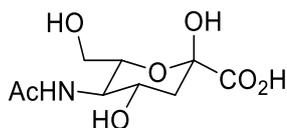
Sugar **107** (131 mg, 0.47 mmol) was dissolved in an aqueous solution of 0.06 M NaOH (10.5 ml) and stirred for 3 h at room temperature. The solution was neutralised with Dowex H⁺ resin,

filtered, and the filtrate reduced *in vacuo*. The solid was then dissolved in an aqueous solution of 20 μM formic acid (2mL) and heated to 80 °C for 12 h. The solution was cooled and reduced *in vacuo*. The resulting solid was purified *via* mass directed chromatography to yield **ATOA** (55 mg 61%) as a light yellow glass solid. $[\alpha]_D^{25}$: −36.4 (*c.* 0.1 in MeOH); υ_{max}/cm^{-1} (Solid) 3600-3050 and 1601; δ_H (500 MHz; D$_2$O) 3.93 (1H, ddd, *J* 11.4, 9.9 and 4.8, H-4), 3.72 (1H, ddd, *J* 10.6, 4.8, and 2.7, H-6), 3.66 (1H, app. d, *J* 10.6, H-5), 3.59 (1H, app. d, *J* 2.7, H-7$_A$), 3.58 (1H, app. d, *J* 4.8, H-7$_B$), 2.17 (1H, dd, *J* 13.0 and 4.8, H-3$_{eq}$), 1.98 (3H, s, acetyl 2-H$_3$), 1.75 (1H, dd, *J* 13.0 and 11.4, H-3$_{ax}$); δ_C (75 MHz; D$_2$O) 176.4 (carboxyl CO), 170.6 (acetyl CO), 96.2 (2-C), 72.7 (6-C), 67.0 (4-C), 61.1 (7-C), 52.5 (5-C), 29.5 (3-C), 22.1 (acetyl Me); HRMS Found M-H: 248.0776. C$_9$H$_{15}$NO$_7$ requires M-H, 248.0776.

2,3-*O*-Isopropylidene-D-ribono-1,4-lactone 49. [76,102]
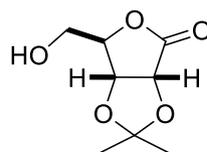
D-Ribono-1,4-lactone (15.0 g, 101 mmol) was added to a solution of conc. H$_2$SO$_4$ (12 mL) in acetone (750 mL) and stirred at room temperature for 12 h. The mixture was neutralised with solid sodium carbonate (60.0 g), dried (MgSO$_4$), filtered and concentrated *in vacuo* to give a crude product which was recrystallised from EtOAc–petrol to yield the lactone **49** (13.4 g, 70%) as colourless plates; m.p. 136-138 °C (from EtOAc–petrol), [lit.[76] 137-139 °C]; R_f: 0.72 (1:1 EtOAc–petrol); $[\alpha]_D^{20}$: −60.9 (*c.* 1.2 in MeOH) [lit.[76] $[\alpha]_D^{20}$: −67.6 (*c.* 1.0 in CHCl$_3$)]; υ_{max}/cm^{-1} (solid) 3501, 2990, 2950 and 1770; δ_H (500 MHz; CDCl$_3$) 4.84 (1H, d, *J* 5.6, 2-H), 4.79 (1H, d, *J* 5.6, 3-H), 4.64 (1H, app. t, *J* 1.8, 4-H), 4.01 (1H, ddd, *J* 12.3, 5.5, and 2.2, 5-H$_A$), 3.82 (1H, ddd, *J* 12.3, 5.5 and 1.7, 5-H$_B$), 2.58 (1H, t, *J* 5.5, OH), 1.48 (3H, s, ^{i}propyl Me$_A$), 1.39 (3H, s, ^{i}propyl Me$_B$); δ_C (75 MHz; CDCl$_3$) 175.4 (1-C), 113.1 (^{i}propyl 2-C), 83.3 (4-C), 78.3 (3-C), 75.7 (2-C), 61.8 (5-C), 26.7 (^{i}propyl Me$_A$), 25.4 (^{i}propyl Me$_B$); HRMS Found M+Na: 211.0579. C$_8$H$_{12}$O$_5$N requires M+Na, 211.0577.

5-Deoxy-5-iodo-2,3-*O*-isopropylidene-D-ribono-1,4-lactone 50. [76,102]
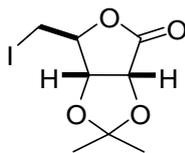
The Lactone **49** (13.0 g, 69.2 mmol) was added to a solution of imidazole (14.3 g, 210 mmol) and triphenylphosphine (64.8 g, 199 mmol), in toluene (240 mL) and heated to 75 °C. Iodine (35.1 g, 138 mmol) was added in small portions over 10 min and the mixture left to stir for a further 30 min. The reaction was then cooled to room temperature, quenched with 1M aqueous sodium thiosulfate solution (120 mL), and extracted with ethyl acetate (3 × 120 mL). The combined organic layers were washed with brine (120 mL), dried (MgSO$_4$), filtered and reduced *in vacuo* to give a crude product. Purification by column chromatography (gradient elution 5:95 → 50:50 EtOAc–petrol) gave the lactone **50** (12.9 g, 68%) as colourless needles, m.p. 96-97 °C (EtOAc–Petrol), [lit.[76] 89-91 °C]; R_f: 0.36 (1:9 EtOAc–Petrol); $[\alpha]_D^{20}$: −20.4 (*c.* 1.7 in MeOH), [lit.[76] −31.0 (*c.* 0.9 in CHCl$_3$)]; υmax/cm^{-1} (solid) 2992 and 1770; δ_H (500 MHz; CDCl$_3$) 4.98 (1H, d, *J* 6.1, 2-H), 4.63 (1H, dd, *J* 5.3 and 3.2, 4-H), 4.61 (1H, d, *J* 6.1, 3-H), 3.44 (1H, dd, *J* 11.2 and 3.2, 5-H$_A$), 3.39 (1H, dd, *J* 11.2 and 5.3, 5-H$_B$), 1.48 (3H, s, ^{i}propyl Me$_A$), 1.40 (3H, s, ^{i}propyl Me$_B$); δ_C (75 MHz; CDCl$_3$) 173.1 (1-C), 114.1 (^{i}propyl 2-C), 80.9 (4-C), 80.4 (3-C), 77.2 (2-C), 27.0 (^{i}propyl Me$_A$), 26.5 (^{i}propyl Me$_B$), 5.6 (5-C); m/z (ES) 618.9 (100%, 2M+Na).

(2*R*, 3*R*)-2,3-*O*-Isopropylidene-pent-4-enoic acid 51. [76,102]
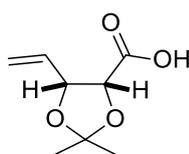
Zinc/copper couple (22.7 g) was added to a solution of the lactone **50** (13.0 g, 47.3 mmol) in 4:1 acetone–water (100 mL) and the resulting mixture was stirred at reflux for 2 h. The reaction mixture was cooled, filtered through celite, washing with acetone, and the filtrate concentrated *in vacuo.* The residue was redissolved in chloroform (80 mL) containing formic acid (2.2 mL) and washed with brine (65 mL). The aqueous phase was re-extracted with chloroform (4 × 35 mL), the combined organic layers were dried (NaSO$_4$), filtered and evaporated *in vacuo* to give the acid **51** (6.3 g, 77%) as a colourless oil. R_f: 0.5 (99:1 EtOAC–AcOH); $[\alpha]_D^{25}$: −31.3 (*c.* 0.8 in MeOH), [lit.[76] −24.6 (*c.* 1.56 in CHCl$_3$)]; υmax/cm^{-1} (film) 3600-2800 and 2990; δ_H (500 MHz; CDCl$_3$) 9.68 (1H, s, CO$_2$H), 5.47 (1H, ddd, *J* 17.2, 10.5 and 7.0, 4-

H), 5.48 (1H, d, J 17.2, 5-H$_{trans}$), 5.32 (1H, d, J 10.5, 5-H$_{cis}$), 4.86 (1H, app t, J 7.2, 3-H), 4.70 (1H, d, J 7.2, 2-H), 1.63 (3H, s, ^{i}propyl Me$_A$), 1.42 (3H, s, ^{i}propyl Me$_B$); δ_C (75 MHz; CDCl$_3$) 174.6 (1-C), 131.6 (4-C), 120.0 (5-C), 111.4 (^{i}propyl 2-C), 78.6 (3-C), 77.2 (2-C), 26.8 (^{i}propyl Me$_A$), 25.4 (^{i}propyl Me$_B$); HRMS Found M+Na: 195.0636, C$_8$H$_{12}$O$_4$ requires M+Na, 195.0628.

(2R, 3R)-2,3-O-Isopropylidenepent-4-enoic acid dipropylamide 52. [76,102]
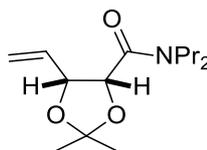
The carboxylic acid **51** (3.00 g, 17.3 mmol) was added to a solution of EDC (4.99g, 26.0 mmol), HOBt (3.51 g, 26.0 mmol) and dipropylamine (3.56 mL, 26.0 mmol) in EtOAc (600 mL), and the resulting mixture stirred at room temperature for 16 h. The reaction mixture was washed with water (400 mL), the aqueous phase extracted with EtOAc (2 × 300 mL), dried (MgSO$_4$) and concentrated *in vacuo* to give a crude product. Purification by column chromatography (1:1 EtOAc–petrol) gave the dipropylamide **52** (6.3 g, 69%) as a pale yellow oil. R_f: 0.31 (1:2 EtOAc–petrol); $[\alpha]_D^{20}$: −13.8 (*c*. 0.9 in MeOH) [lit.[76] −28.6 (c.0.91 in CHCl$_3$)]; υ_{max}/cm^{-1} (solid) 2965, 2936, 2876 and 1660; δ_H (500 MHz; CDCl$_3$) 5.80 (1H, ddd, J 17.4, 10.2 and 8.2, 4-H), 5.40 (1H, d, J 17.4, 5-H$_{trans}$), 5.25 (1H, d, J 10.2, 5-H$_{cis}$), 4.94 (1H, d, J 7.2, 2-H), 4.78 (1H, app. t, J 7.8, 3-H), 3.50 (1H, dt, J 13.4 and 7.7, propyl 1-H), 3.15 - 2.96 (3H, m, 3 × propyl 1-H), 1.66 (3H, s, Me$_A$), 1.65 - 1.74 (4H, m, 2 × propyl 2-H$_2$), 1.41 (3H, s, Me$_B$), 0.91 (3H, t, J 7.4, propyl 3-H$_3$), 0.88 (3H, t, J 7.4, propyl 3-H$_3$); δ_C (75 MHz; CDCl$_3$) 167.5 (1-C), 133.8 (4-C), 119.9 (5-C), 110.7 (^{i}propyl 2-C), 79.4 (3-C), 75.6 (2-C), 49.0 (propyl 1-C), 48.2 (propyl 1-C), 26.8 (^{i}propyl Me$_A$), 25.5 (^{i}propyl Me$_B$), 22.2 (propyl 2-C), 20.7 (propyl 2-C), 11.5 (propyl 3-C), 11.2 (propyl 3-C); HRMS Found M+H: 256.1910, C$_{14}$H$_{25}$NO$_3$ requires M+H, 256.1907.

(2R,3R)-2,3-Dihydroxy-N,N-dipropylpent-4-enamide 53. [76,102]
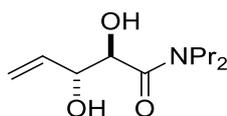
The dipropylamide **52** (3.0 g, 12 mmol) was dissolved in 9:1 trifluoroacetic acid–water (25 mL) and stirred for 2 min. The reaction mixture was concentrated *in vacuo* to give a crude product. Purification by column chromatography (2:3 EtOAc–petrol) gave the dihydroxyamide **53** (2.3 g,

92%) as a colourless amorphous solid; m.p. 75–77 °C (EtOAc–petrol) [lit.[76] 77–79 °C (CH$_2$Cl$_2$)]; R_f: 0.33 (1:1 EtOAc–petrol); $[\alpha]_D^{20}$: +9.8 º (c. 1.2 in MeOH) [lit.[76] $[\alpha]_D^{20}$ +17.1 (c. 0.84 in CHCl$_3$)]; υ_{max}/cm^{-1} (solid) 3325, 2966, 2877 and 1621; δ_H (500 MHz; CDCl$_3$) 5.81 (1H, ddd, J 17.0, 10.6 and 5.6, 4-H), 5.34 (1H, d, J 17.0, 5-H$_{trans}$), 5.26 (1H, d, J 10.6, 5-H$_{cis}$), 4.44 (1H, d, J 4.1, 2-H), 4.23 (1H, dd, J 5.6 and 4.1, 3-H), 3.58 (1H, ddd, J 13.6, 8.6 and 6.8, propyl 1-H), 3.36 (1H, ddd, J 15.6, 8.3 and 7.3, propyl 1-H), 3.16 (1H, ddd, J 14.8, 8.6 and 6.1, propyl 1-H), 3.05 (1H, ddd, J 15.6, 8.3 and 7.0, propyl 1-H), 1.70 - 1.50 (4H, m, 2 × propyl 2-H$_2$), 0.94 (3H, t, J 7.4, propyl 3-H$_3$), 0.90 (3H, t, J 7.4, propyl 3-H$_3$); δ_C (75 MHz; CDCl$_3$) 171.3 (1-C), 135.4 (4-C), 117.5 (5-C), 73.9 (3-C), 70.6 (2-C), 48.8 (propyl 1-C), 47.6 (propyl 1-C), 22.1 (propyl 2-C), 20.6 (propyl 2-C), 11.3 (propyl 3-C), 11.1 (propyl 3-C); m/z (ES) 238.1 (30.2%, M+Na).

(5R,6R)-7-(Dipropylamino)-4,5,6-trihydroxy-2,7-dioxoheptanoic acid (DPAH).[71,83]
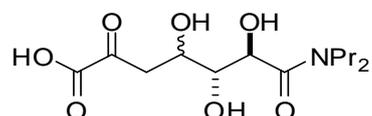
The alkene **53** (1.16 g, 5.3 mmol) was dissolved in methanol (20 mL) and cooled to -78 °C, oxygen was bubbled through the solution for 20 min and then O$_3$ was bubbled through the solution until a blue colour was observed, at which point oxygen was bubbled through the solution for 20 min. Dimethyl sulfide (0.85 mL, 25 mmol) was added to the reaction mixture and the mixture stirred at room temperature until a negative ozonide test was observed (starch iodine paper). The reaction was reduced *in vacuo*. The residue was re-suspended in 20 mM Tris HCl buffer pH 7.4 (37 mL). Sodium pyruvate (3.3 g, 30 mmol) was added, the pH corrected to pH 7.4, the enzyme E192N (3 mg/mL, 3 mL) was added and the reaction mixture stirred for 18 h. Aqueous formic acid (2M) was added to achieve pH 3 and the solution was stirred for 15 min, the pH was adjusted to pH 7 with aqueous ammonium hydroxide (2M) and baker's yeast was added (5 g) and the reaction stirred for 18 h. The emulsion was filtered through celite, pad washed with water (10 mL) and the filtrated was lyophilised. The crude product was re-suspended in water (3 mL) and purified *via* ion exchange chromatography to yield the sugar **DPAH** in (0.9 g, 60 %) as a colourless glass, R_f: 0.5 (5:2:2, EtOAc–AcOH–H2O); υ_{max}/cm^{-1} (solid) 3119, 2966, 2936, 2877, 1715 and 1621; δ_H (500 MHz; D$_2$O) 4.66 (1H, d, J 9.5, 6-H $_{(maj\,fur)}$), 4.01 (1H, ddd, J 11.8, 9.5 and 5.1, 4-H $_{(maj\,fur)}$), 3.74 (1H, t, J 9.5, 5-H $_{(maj\,fur)}$), 3.58-3.40 (2H, m, propyl 1-H$_2$), 3.38- 3.19 (2H, m, propyl 1-H$_2$), 2.20 (1H, dd, J 13.0 and 5.1, 3-H$_A$ $_{(maj\,fur)}$), 1.94 (1H, dd, J 13.0 and 11.8, 3-H$_B$ $_{(maj\,fur)}$), 1.72-1.64 (2H, m, propyl 2-H$_2$), 1.63-1.55 (2H, m, propyl 2-H$_2$), 0.95-0.87 (6H, m, 2 × propyl 3-H$_3$); HRMS Found M-H: 304.1399. C$_{13}$H$_{23}$NO$_7$ requires M-H, 304.1418.

The reaction produces two diastereoisomers at (4-*S*:4-*R* 75:25). The ratio of the diastereoisomers has been previously determined at 500 MHz ^{1}H NMR, 8 forms of the sugar **DPAH** (including four furanose and four pyranose rings) are present.[83,103] Full interpretation of the 8 products was not included but the spectra obtained were identical to the authentic sample.

N-((3*S*,4*S*)-4,5-Dihydroxypent-1-en-3-yl)acetamide 58.[81]
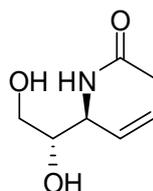
D-Glyceraldehyde (0.6 g, 6.7 mmol), di-*p*-anisylmethylamine (1.62 g, 6.7 mmol) and vinyl boronic pinicol ester (2.05 g, 13.3 mmol) were stirred in a 4:1 ethanol-water (30 mL) solution for 72 h at 50 °C. TFA (2 mL) was then added and the solution stirred for a further 16 h. The solvent was removed *in vacuo* and the crude solid re-suspended in methanol (17 mL). Sodium bicarbonate (1.12 g, 13.4 mmol) and Ac_2O (0.85 g, 8.4 mmol) was added and stirred for 1 h at room temperature. The solid was filtered and the filtrate reduced *in vacuo*. Crude product was purified *via* silica gel column chromatography (gradient elution CH_2Cl_2 → 90:10 CH_2Cl_2–MeOH) to give the alkene **58** (0.34 g, 35 %) as a colourless glass, R_f: 0.14 (9:1, DCM–MeOH), δ_H (500 MHz; CD_3OD) 5.97 (1H, ddd, *J* 17.2, 10.4 and 6.4, 2-H), 5.27 (1H, app. d, *J* 17.2, 1-H_A), 5.23 (1H, app. d, *J* 10.4, 1-H_B), 4.52 (1H, app. t, *J* 6.4, 3-H), 3.69-3.65 (1H, m, 4-H), 3.60 (1H, dd, *J* 11.3 and 4.5, 5-H_A), 3.54 (1H, dd, *J* 11.3 and 6.3, 5-H_B), 2.03 (3H, s, acetyl Me); δ_C (75 MHz; CD_3OD) 172.9 (acetyl CO), 135.6 (2-C), 117.5 (1-C), 74.6 (4-C), 64.4 (5-C), 55.3 (3-C), 22.8 (acetyl Me); HRMS Found M+Na: 182.0788. $C_7H_{13}NO_3$ requires M+Na, 182.0785.

(3*S*,5*R*)-Ethyl 5-((1*R*,2*S*)-1-acetamido-2,3-dihydroxypropyl)-2-tert-butylisoxazolidine-3-carboxylate 36.[81]
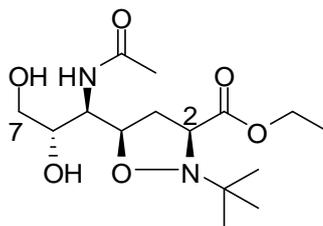
The alkene **58** (159 mg, 1.0 mmol) and the nitrone **38** (346 mg, 2.0 mmol) were stirred in dioxane (20 mL) at 30°C for 14 days. Reaction mixture was concentrated *in vacuo* and the crude product was purified *via* silica gel column chromatography (gradient elution: $CH_2Cl_2 \rightarrow$ 90:10 CH_2Cl_2–MeOH) to give the isoxazolidine **36** (99.6 mg, 30 %) as a colourless glass solid, R_f: 0.15 (9:1, CH_2Cl_2–MeOH); υ_{max}/cm^{-1} (solid); δ_H (500 MHz; CD$_3$OD) 4.67 (1H, app. t, *J* 7.5, H-4), 4.20 (2H, m, ethyl 1-H$_2$), 3.99 (1H, app. t, *J* 8.3, H-2), 3.87 (1H, app. d, *J* 9.3, H-5), 3.64-3.60 (1H, m, H-6), 3.58 (1H, dd, *J* 12.1 and 2.12, H-7$_A$), 3.47 (1H, dd, *J* 12.1 and 5.7, H-7$_B$), 2.67 (1H, app. dt, *J* 12.3 and 7.5, H-3$_A$), 2.17 (1H, app. dt, *J* 12.2 and 7.5, H-3$_B$), 2.04 (3H, s, acetyl Me), 1.27 (3H, q, *J* 7.1, ethyl 2-H$_3$), 1.13 (9H, s, ^{t}butyl 2-H$_3$); δ_C (75 MHz; CD$_3$OD) 174.12 (ethyl ester CO), 174.02 (acetyl CO), 77.11 (4-C), 72.86 (6-C), 64.77 (7-C), 62.20 (ethyl 1-C), 61.91 (2-C), 60.99 (^{t}butyl 1-C), 53.46 (5-C), 38.57, 25.67 (^{t}butyl 2-C), 22.51 (acetyl Me), 14.10 (ethyl 2-C); HRMS Found M+H: 333.2020. $C_{15}N_{28}N_2O_6$ requires M+H 333.2022.

(The signals in NMR spectra are labelled according to the numbering system for *N*-acetylneuraminic acid).

**5-Acetamido-2,4-dihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-carboxylic acid (*ent*-ATOA).[81]**
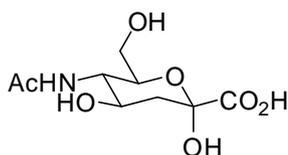
The Isoxazol **36** (77.6 mg, 0.23 mmol) was stirred in anhydrous MeONa (0.05 M, 10 mL) at room temperature for 12 h. The solution was then diluted with water (10 mL) and stirred for 24 h. The mixture was purified directly with ion-exchange chromatography (Dowex 1X8-100resin, formate form) with gradient elution 0 $\rightarrow$ 2 M formic acid, providing the carboxylic acid ***ent*-ATOA** as a colourless glass solid (34.4 mg, 60%), $[\alpha]_D^{20}$: +23.7, υ_{max}/cm^{-1} (solid) 3277,

2952, 2852 and 1682; δ_H (500 MHz; D_2O) 4.01 (1H, ddd, J 11.6, 9.9 and 4.8, H-4), 3.83 (1H, ddd, J 10.6, 4.8 and 2.7, H-6), 3.75 (1H, app. d, J 10.6, H-5), 3.64 (1H, d, J 2.7, H-7$_A$), 3.63 (1H, d, J 4.8, H-7$_B$), 2.22 (1H, dd, J 13.0 and 4.8, H-3$_{eq}$), 2.07 (3H, s, acetyl 2-H$_3$), 1.86 (1H, dd, J 13.0 and 11.6, H-3$_{ax}$); HRMS Found M-H: 248.0775. $C_9H_{15}NO_7$ requires M-H 248.0776.

N-((3*S*,4*S*,5*R*,6*S*)-4,5,6,7-Tetrahydroxyhept-1-en-3-yl)acetamide 65.[81]
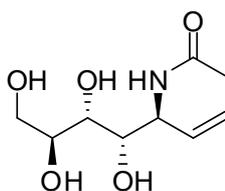
L-Arabinose (1.0 g, 6.7 mmol), di-*p*-anisylmethylamine (1.62 g, 6.7 mmol) and vinyl boronic pinicol ester (2.05 g, 13.3 mmol) were stirred in a 4:1 ethanol–water (30 mL) solution for 72 h at 50 °C. TFA (2 mL) was then added and the solution stirred for a further 16 h. The solvent was removed *in vacuo* and the crude solid re-suspended in methanol (17 mL). Sodium bicarbonate (1.12 g, 13.4 mmol) and Ac$_2$O (0.85 g, 8.4 mmol) was added and stirred for 1 h at room temperature. The solid was filtered and the filtrate reduced *in vacuo*. The crude product was purified *via* silica gel column chromatography (gradient elution: CH$_2$Cl$_2$ → 90:10 CH$_2$Cl$_2$–MeOH) to give the alkene **65** (0.74 g, 50 %) as a colourless glass, R_f: 0.06 (9:1, DCM–MeOH); υ_{max}/cm^{-1} (solid) 3346.28, 2940.66 and 1672.96; δ_H (500 MHz; CD$_3$OD) 6.06 (1H, ddd, J 16.8, 10.5 and 5.8, 2-H), 5.23 (1H, dd, J 16.8 and 2.9, 1-H$_A$), 5.18 (1H, dd, J 10.5 and 2.9, 1-H$_B$), 4.56 (1H, app. t, J 7.0, 3-H), 3.82 (1H, dd, J 11.0 and 3.5, 7-H$_A$), 3.76 (1H, d, J 8.6, 4-H), 3.72 (1H, app. dd, J 8.5, 3.5, 6-H), 3.66 (1H, dd, J 11.0 and 5.7, 7-H$_B$), 3.55 (1H, d, J 7.4, 5-H), 2.0 (3H, S, acetyl 1-H); δ_C (75 MHz; CD$_3$OD) 180.7 (acetyl CO), 136.9 (2-C), 116.8 (1-C), 72.4 (6-C), 72.0 (4-C), 71.2 (5-C), 64.5 (7-C), 54.9 (3-C), 22.8 (acetyl Me); HRMS Found M+H: 220.1175. $C_9H_{17}NO_5$ requires M+H, 220.1179.

(3*S*,5*R*)-Ethyl 5-((1*R*,2*S*,3*R*,4*S*)-1-acetamido-2,3,4,5-tetrahydroxypentyl)-2-tert-butylisoxazolidine-3-carboxylate 66.[81]
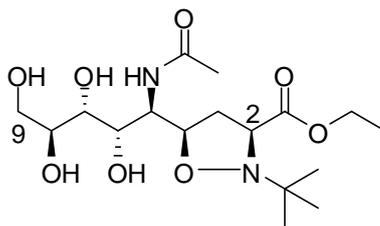
The alkene **65** (219 mg, 1.0 mmol) and the nitrone **38** (346 mg, 2.0 mmol) were stirred in dioxane (20 mL) at 30°C for 14 days. The reaction mixture was concentrated *in vacuo* and the crude product was purified *via* silica gel column chromatography (gradient elution: $CH_2Cl_2 \rightarrow$ 90:10 CH_2Cl_2–MeOH) to give the isoxazolidine **66** (215 mg, 55%) as a colourless glass, R_f: 0.18 (9:1, DCM-MeOH); υ_{max}/cm^{-1} (solid) 3399, 2924, 2851 and 1673; δ_H (500 MHz; CD$_3$OD) 4.77 (1H, ddd, *J* 8.4, 7.0 and 1.4, 4-H), 4.29-4.18 (2H, m, ethyl 1-H$_2$), 4.02 (1H, app. t, *J* 8.4, 2-H), 3.98 (1H, dd, *J* 10.2 and 1.4, 5-H), 3.91 (1H, app. d, *J* 10.2, 6-H), 3.92 (1H, dd, *J* 11.0 and 3.7, 9-H$_A$), 3.77 (1H, ddd, *J* 8.4, 5.3 and 3.7, 8-H), 3.69 (1H, dd, *J* 11.0 and 5.3, 9-H$_B$), 3.49 (1H, app. d, *J* 8.4, 7-H), 2.75 (1H, ddd, *J* 12.4, 8.2 and 7.0, 3-H$_A$), 2.24 (1H, dt, *J* 12.4 and 8.4, 3-H$_B$), 2.11 (3H, s, acetyl Me), 1.31 (3H, t, *J* 7.1, ethyl 2-H$_3$), 1.18 (9H, s, ^{t}butyl 2-H$_3$); δ_C (75 MHz; CD$_3$OD) 175.1 (ethyl ester CO), 174.4 (acetyl CO), 77.3 (4-C), 72.3 (8-C), 71.5 (7-C), 70.6 (6-C), 65.0 (9-C), 62.5 (ethyl 1-C), 62.2 (2-C), 61.2 (^{t}butyl 1-C), 39.0 (3-C), 26.0 (^{t}butyl 2-C), 22.9 (acetyl Me), 14.4 (ethyl 2-C); HRMS Found M+H: 393.2243. $C_{17}H_{32}N_2O_8$ requires M+H 393.2231.

(The signals in NMR spectra are labelled according to the numbering system for *N*-acetylneuraminic acid).

5-Acetamido-2,4-dihydroxy-6-((7*S*,8*S*)-7,8,9-trihydroxypropyl)tetrahydro-2*H*-pyran-2-carboxylic acid (*ent*-NANA).[81]
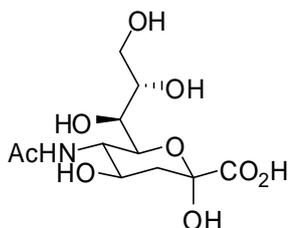
The Isoxazol **55** (76 mg, 0.25 mmol) was stirred in anhydrous MeONa (0.05 M, 10 mL) at room temperature for 12 h. The solution was then diluted with water (10 mL) and stirred for 24 h. The mixture was purified directly with ion-exchange chromatography (Dowex 1X8- 100resin,

formate form) with gradient elution 0 → 2 M formic acid, providing the carboxylic acid **ent-NANA** as a colourless glass solid (42.5 mg, 55%), $[\alpha]_D^{20}$: +18.3, υ_{max}/cm^{-1} (solid) 3325.4, 2931.48, 1732.42 and 1649.63; δ_H (500 MHz; D$_2$O) 4.03-4.00 (1H, m, 4-H), 3.99 (1H, dd, *J* 10.3 and 1.0, 6-H), 3.86 (1H, t, *J* 10.3, 5-H), 3.77 (1H, dd, *J* 11.8, 2.6, 9-H$_A$), 3.68 (1H, ddd, *J* 9.0, 6.3 and 2.6, 8-H), 3.55 (1H, dd, *J* 11.8 and 6.3, 9-H$_B$), 3.49 (1H, dd, J 9.0 and 1.0, 7-H), 2.25 (1H, dd, *J* 13.1 and 4.9, 3-H$_{eq}$), 1.98 (3H, s, acetyl 2-H$_3$), 1.82 (1H, dd, J 13.1 and 11.7, 3-H$_{ax}$); HRMS Found M-H: 308.0987. C$_{11}$H$_{19}$NO$_9$ requires M-H 308.0982.

(*R*)-2-(Furan-2-yl)-2-hydroxy-*N,N*-dipropylacetamide 85.
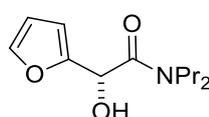
Hoveyda–Grubbs II (7.4 mg, 0.01 mmol, 5 mol%) was added to a solution of dihydroxyamide **53** (50 mg, 0.2 mmol) and acrolein (0.04 mL, 0.60 mmol) in MTBE (5 mL) and the resulting mixture was stirred at room temperature for 12 h. Tris(hydroxymethyl)phosphine (124 mg, 1.00 mmol) and NEt$_3$ (0.14 mL, 1.00 mmol) were added and stirred for 30 min, then silica (3 g) was added and stirred for 30 min. The slurry was filtered through Celite and concentrated to give a crude product. Purification by column chromatography (2:3 EtOAc–petrol) gave the furan **85** (20 mg, 38 %) as a dark yellow oil; R_f: 0.3 (2:3 EtOAc–petrol);$[\alpha]_D^{20}$: −26.6 (c. 1.2 in MeOH); υ_{max}/cm^{-1} (film) 3390, 2966, 2936, 2877 and 1646; δ_H (500 MHz; CDCl$_3$) 7.37 (1H, d, *J* 1.6, furanyl 5-H), 6.35 (1H, dd, *J* 3.3 and 1.6, furanyl 4-H), 6.30 (1H, d, *J* 3.3, furanyl 3-H), 5.29 (1H, d, *J* 6.8, 2-H), 4.68 (1H, d, *J* 6.8, OH), 3.49 (1H, ddd, *J* 15.1, 8.2 and 7.0, propyl 1-H), 3.23 (1H, ddd, *J* 15.1, 8.2 and 6.8, propyl 1-H), 3.11 (1H, ddd, *J* 15.0, 10.1 and 6.2, propyl 1-H), 3.00 (1H, ddd, *J* 15.0, 10.1 and 4.9, propyl 1-H), 1.65 - 1.54 (2H, m, 2 × propyl 2-H), 1.54-1.44 (1H, m, propyl 2-H), 1.25-1.10 (1H, m, propyl 2-H), 0.89 (3H, t, *J* 7.4, propyl 3-H), 0.79 (3H, t, *J* 7.4, propyl 3-H); δ_C (75 MHz; CDCl$_3$) 170.0 (1-C), 152.9 (furanyl 2-C), 143.0 (furanyl 5-C), 111.1 (furanyl 4-C), 108.4 (furanyl 3-C), 65.0 (2-C), 48.7 (propyl 1-H), 484 (propyl 1-H), 21.8 (propyl 2-H), 20.9 (propyl 2-H), 11.7 (propyl 3-H), 11.4 (propyl 3-H); HRMS Found M+Na: 248.1267, C$_{12}$H$_{19}$NO$_3$ requires M+Na, 248.1257.

(4*S*,5*S*,*E*)-Methyl-6-(dipropylamino)-4,5-dihydroxy-6-oxohex-2-enoate 90.
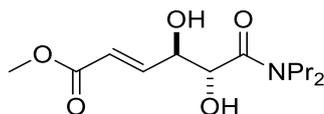
Grubbs II (49 mg, 0.06 mmol, 5 mol%), was added to a solution of dihydroxyamide **53** (250 mg, 1.2 mmol) and methyl acrylate (0.3 mL, 3.5 mmol) in MTBE (7 mL), and the resulting mixture was heated at reflux for 5 days. The reaction mixture was cooled to room temperature, tris(hydroxymethyl)phosphine (600 mg, 25.8 mmol) and NEt$_3$ (0.35 mL, 25.8 mmol) were added, stirred for 30 min, silica (6 g) added, stirred for 30 min, filtered through celite with EtOAc and concentrated under reduced pressure to give a crude product. Purification by column chromatography (2:3 EtOAc–petrol) give the methyl ester **90** (0.2 g, 63%,) as a colourless amorphous solid; m.p. 97–99 °C (petrol); R_f: 0.31 (1:2 EtOAc–petrol); $[\alpha]_D^{20}$: +5.9 (c. 0.8 in MeOH); υ_{max}/cm^{-1} (solid) 3500-3050, 2959 and 1716; δ_H (500 MHz; CDCl$_3$) 6.90 (1H, dd, *J* 15.5 and 4.6, 3-H), 6.13 (1H, dd, *J* 15.5 and 1.2, 2-H), 4.44 (1H, app. dd, *J* 7.5 and 4.6, 4-H), 4.41 (1H, app. s, 5-H), 3.80 (1H, d, *J* 8.5, OH), 3.74 (3H, s, methyl ester Me), 3.55 (1H, m, propyl 1-H), 3.45 (1H, d, *J* 4.5, OH), 3.34 (1H, m, propyl 1-H), 3.17 (1H, m, propyl 1-H), 3.09 (1H, m, propyl 1-H), 1.59 (4H, m, 2 × propyl 2-H), 0.93 (3H, t, *J* 7.42, propyl 3-H), 0.90 (3H, t, *J* 7.42, propyl 3-H); δ_C (75 MHz; CDCl$_3$) 171.2 (1-C), 166.8 (6-C), 145.6 (3-C), 123.0 (2-C), 73.0 (5-C), 70.5 (4-C), 52.1 (methyl ester CO), 49.4 (propyl 1-C), 48.2 (propyl 1-C), 22.6 (propyl 2-C), 21.0 (propyl 2-C), 11.7 (propyl 3-C), 11.5 (propyl 3-C); HRMS Found M+H: 274.1656. C$_{13}$H$_{23}$NO$_5$ requires M+H, 274.1649.

(*E*)-Methyl-3-((4*S*,5*S*)-5-(dipropylcarbamoyl)-2,2-dimethyl-1,3-dioxolan-4-yl)acrylate 91.
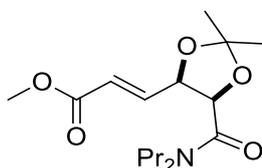
Methyl ester **90** (0.2 g, 0.7 mmol) and *p*TsOH (3.61 mg, 0.02 mmol) were added to a solution of 1:1 2,2-dimethoxypropane–acetone (10 mL) and stirred at room temperature for 12 h. Aqueous saturated NaHCO$_3$ (5 mL) was added, extracted with EtOAc (3 × 5 mL), dried (MgSO$_4$), filtered and concentrated under reduced pressure to give the methyl ester **91** (0.16 g, 72%) as a colourless oil; R_f: 0.31 (1:3 EtOAc–petrol); $[\alpha]_D^{20}$: +12.3 (*c*. 0.5 in MeOH); υ_{max}/cm^{-1} (solid) 2937, 2966, 1728 and 1651; δ_H (500 MHz; CDCl$_3$) 6.74 (1H, dd, *J* 15.6 and 6.4, 3-H), 6.04 (1H,

dd, *J* 15.6 and 1.2, 2-H), 4.95 (1H, d, *J* 7.5, 5-H), 4.89 (1H, ddd, *J* 7.5, 6.4 and 1.2, 4-H), 3.64 (3H, s, methyl ester Me), 3.35 (1H, dt, *J* 13.6 and 8.1, propyl 1-H), 3.12 (1H, ddd, *J* 15.5, 9.6 and 6.1, propyl 1-H), 3.03 (3H, ddd, *J* 15.5, 9.6 and 6.5, propyl 1-H), 2.96 (4H, dt, *J* 13.6 and 8.1, propyl 1-H), 1.58 (3H, s, ^{i}propyl Me$_A$), 1.57-1.48 (2H, m, propyl 2-H), 1.41 (2H, app. sextet *J* 7.6 propyl 2-H), 1.34 (3H, s, ^{i}propyl Me$_B$), 0.84 (3H, t, *J* 7.4, propyl 3-H), 0.79 (3H, t, *J* 7.4, propyl 3-H); δ_C (75 MHz; CDCl$_3$) 166.7 (C-6), 165.8 (C-1), 142.5 (C-3), 123.4 (C-2), 111.3 (^{i}propyl 2-C), 76.7 (4-C), 76.3 (5-C), 51.6 (methyl ester Me), 48.9 (propyl 1-C), 48.2 (propyl 1-C), 26.7 (Me$_B$), 25.3 (Me$_A$), 22.2 (propyl 2-C), 20.4 (propyl 2-C), 11.1 (propyl 3-C), 11.1 (propyl 3-C); HRMS Found M+H: 314.1911. C$_{16}$H$_{27}$NO$_5$ requires M+H, 314.1967.

Appendix I. Mass spectrometric data for the chemically modified proteins

Below are the multiple mass-charge state spectra of the *Sa*NAL variants used for kinetic measurements. These spectra were obtained through the methods shown in section 4.1.19. The multiple mass-charge state spectra (primary data) are shown and the 31+ peak is highlighted in each spectrum. This is related to the mass of the protein through the mass-to-charge ratio, m/z.[84] The primary data was routeinly deconvoluted by Dr. James Ault (University of Leeds) using MassLynx software to obtain the molecular mass of the protein. These samples have been taken after gel filtration. In some cases, adducts from side reactions were observed. This was potentially caused by over alkylation by the dibromide compound **28**.[104] Despite this, peaks were seen for the correct size of modified proteins and these samples were used for kinetic characterisation.

Calculated mass: 34043.42 Da
Measured Mass: 34041.00 Da

F190-DHPC

31+

E192-CPC
Calculated mass: 34037.45 Da
Measured mass: 34035.88 Da

E192-CPC

31+

E192-AEC

E192-AEC
Calculated mass: 34012.43 Da
Measured mass: 34011.41 Da

31+

E192-HPC
Calculated mass: 34027.43 Da
Measured mass: 34027.37 Da

E192-HPC

31+

Calculated mass: 34070.44 Da
Measured mass: 34075.10 Da
E192-AMCEC

31+

E192-HEBC
Calculated mass: 34069.48 Da
Measured mass: 34073.60 Da

31+

E192-MCMC
Calculated mass: 34041.41 Da
Measured mass: 34043.72 Da

31+

E192-PC
Calculated mass: 34009.54 Da
Measured mass: 34009.59 Da

31+

117

E192-ACEC
Calculated mass: 34056.42 Da
Measured mass: 34053.40 Da

E192-ACEC

Calculated mass: 34051.47 Da
Measured mass: 34050.23 Da

E192-CHC

Calculated mass: 34027.39 Da
Measured mass: 34026.82 Da

E192-CMC

E192-HBC

Calculated mass: 34041.45 Da
Measured mass: 34040.05 Da

F190C
Calculated mass: 33950.99 Da
Observed mass: 33950.89 Da

F190-DHPC
Calculated mass: 34025.94 Da
Observed mass: 34024.85 Da

F190-HBC
Calculated mass: 34023.97 Da
Measured mass: 34023.54 Da

F190-HPC
Calculated mass: 34009.95 Da
Measured mass: 34007.40 Da

F190-PC

Calculated mass: 33992.06 Da
Measured mass: 33990.9590 Da

Appendix II. Supplementary kinetic data

In order to assess the validity of the kinetic assay, several control measurements were performed, in which the decrease in concentration of NADH was observed over a period of 30 minutes. These measurements can be seen in Table 5. Each measurement was performed with one component of the assay missing, i.e. no protein was used when measuring each of the substrates effect on the assay; no substrate was used when measuring the effect of the protein on the assay. The background decrease in NADH concentration is therefore shown for each component in the assay. The substrate concentrations used in these measurements are the highest concentrations used when performing the kinetic assays. The enzyme concentration used is also typical of my standard assay at 0.1 mg/ml. From the changes in absorbance observed for each of the control reactions (Table 5) it could be inferred that no component would interfere with the kinetic assay measurements.

Table 5 | Background reactions performed to assess the reliability of the coupled enzyme assay.

|	50mM Tris (µL)	LDH (µL)	20mM NADH (µL)	100mM subs (µL)	E192-AMCEC (µL)	Δ A_{340}/30min
N-acetylneuraminic acid	788	2	10	200	0	**0.0001**
DPAH	938	2	10	50	0	**0.0011**
ATOA	688	2	10	300	0	**0.0019**
ent-ATOA	938	2	10	50	0	**0.0010**
ent-NANA	983	2	10	5	0	**0.0018**
enz	977.1	2	10	0	10.9	**0.0001**
no LDH-SA	779.1	0	10	200	10.9	**0.0017**

Graphs of specific activity used to determine kinetic parameters of SaNAL variants:

In this part of the appendix you will find the activity plots for all variants of *Sa*NAL with the substrates *N*-acetylneuraminic acid, DPAH and ATOA.
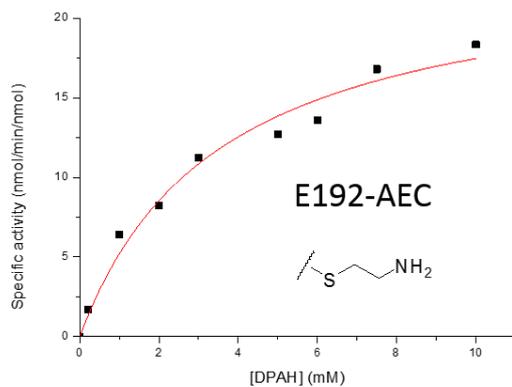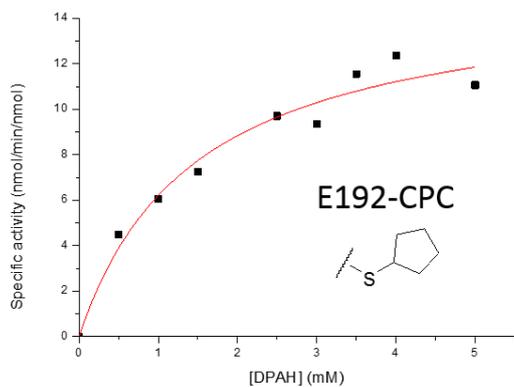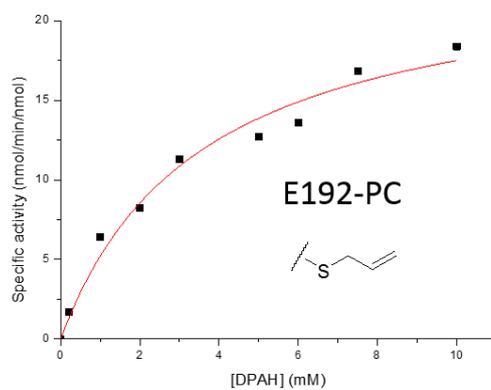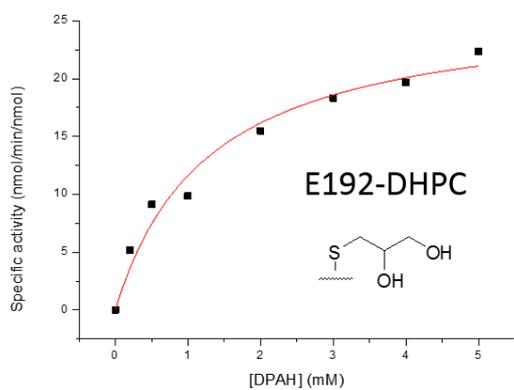
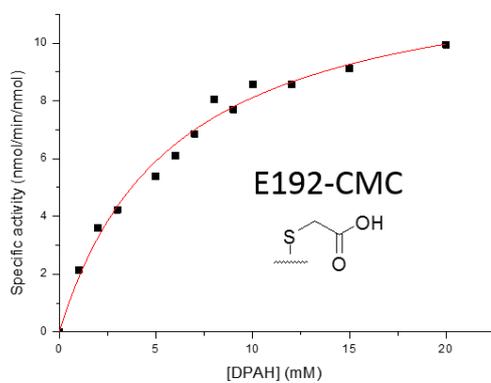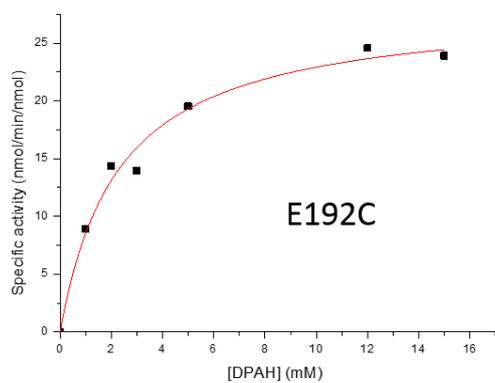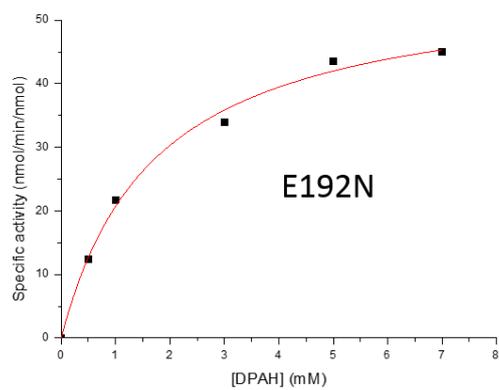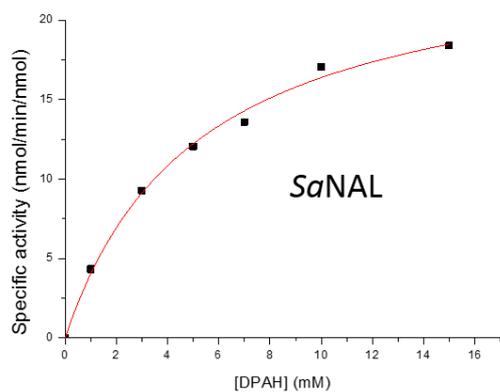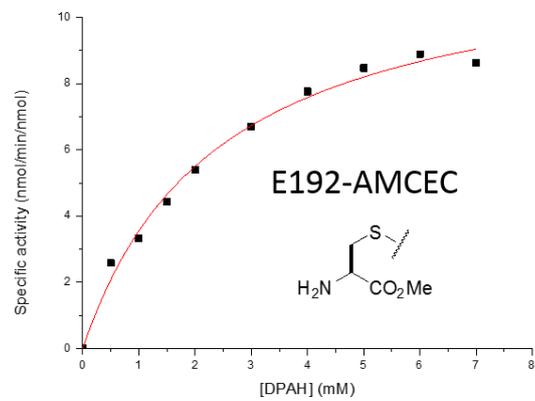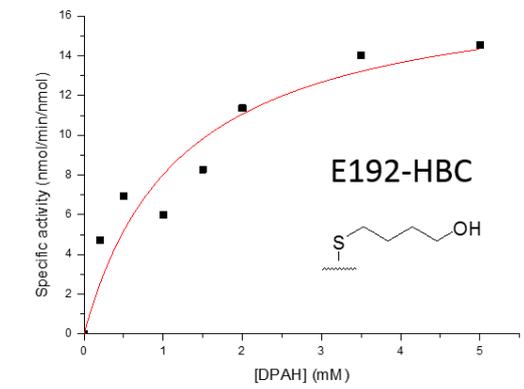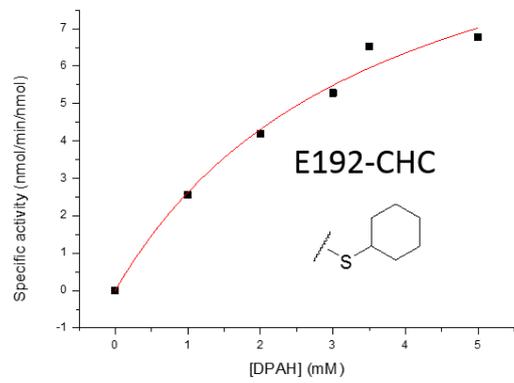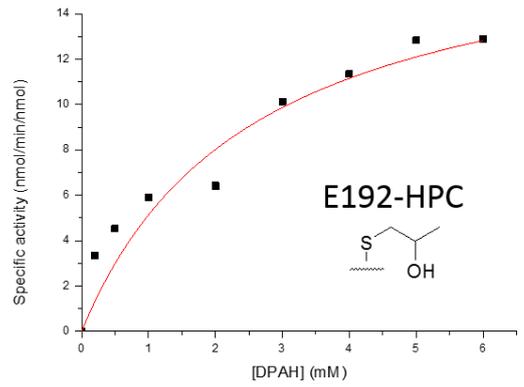## Activity plots of the SaNAL variants with the N-acetylneuraminic acid as a substrate:
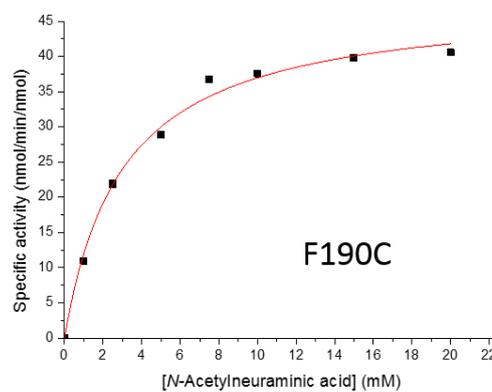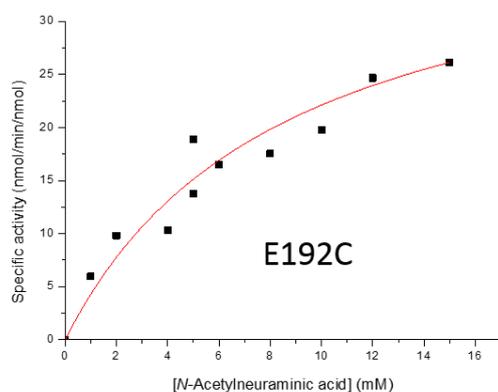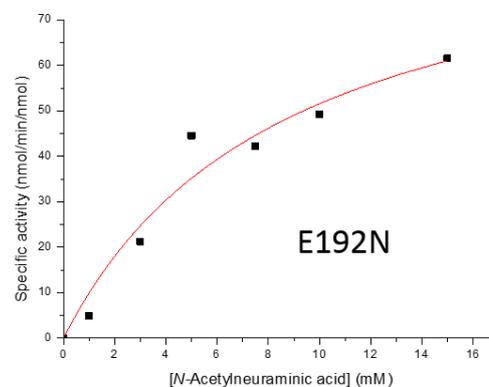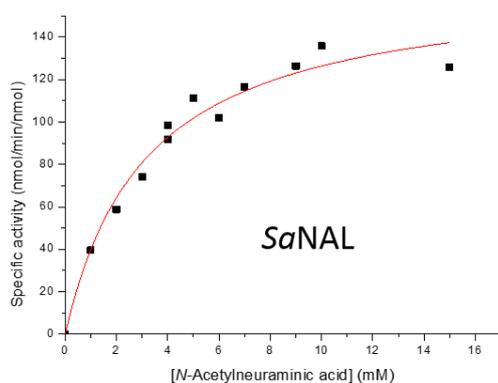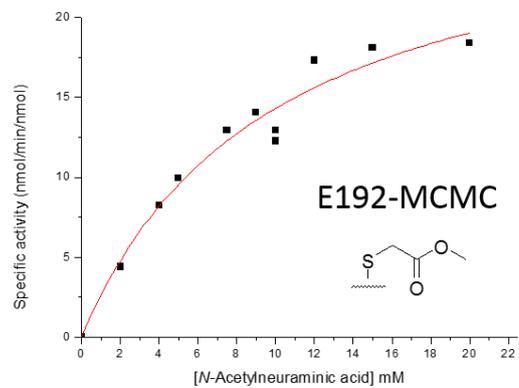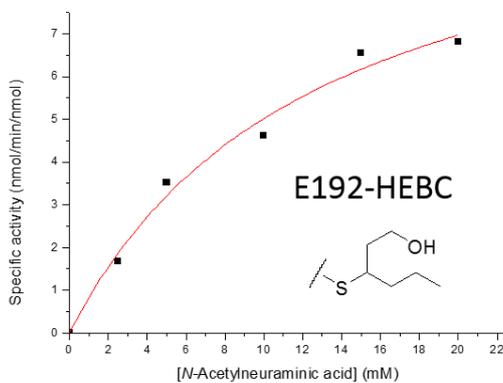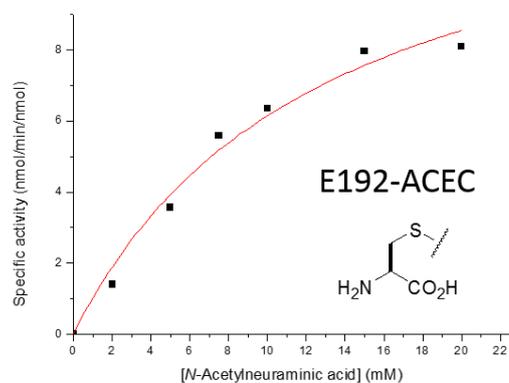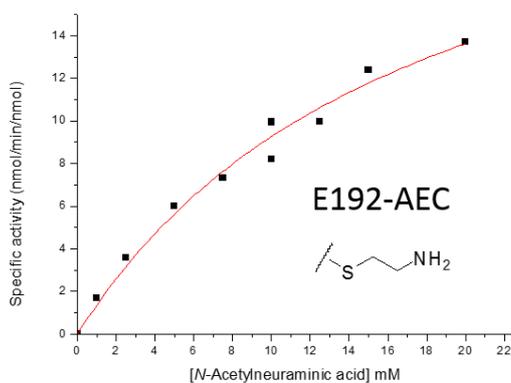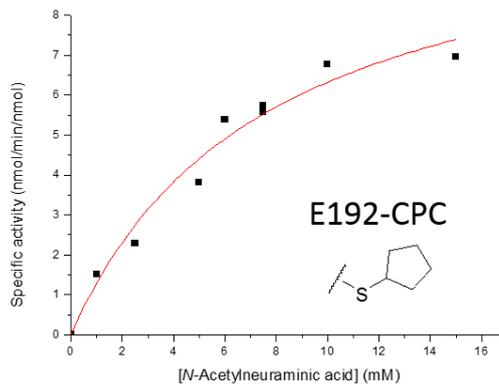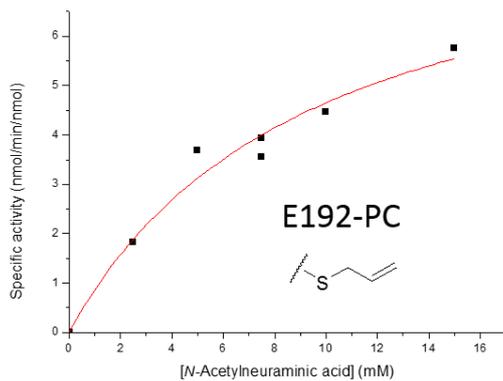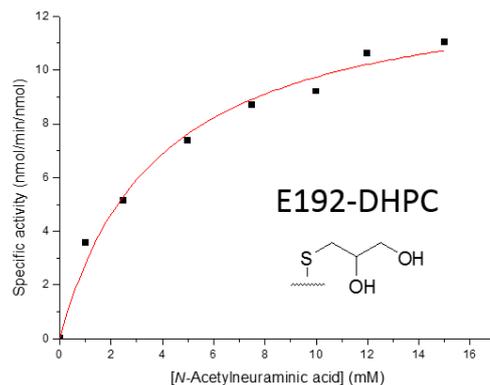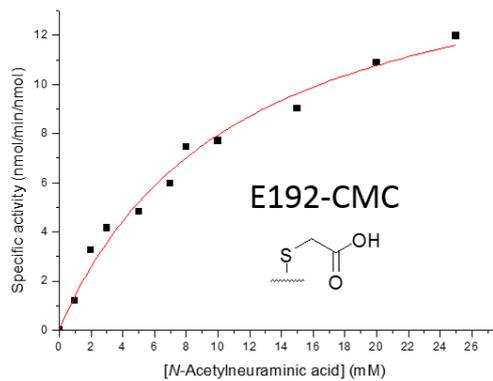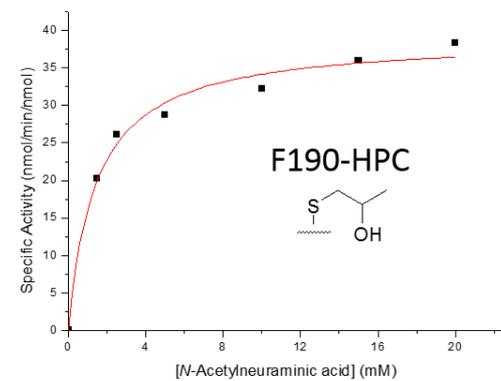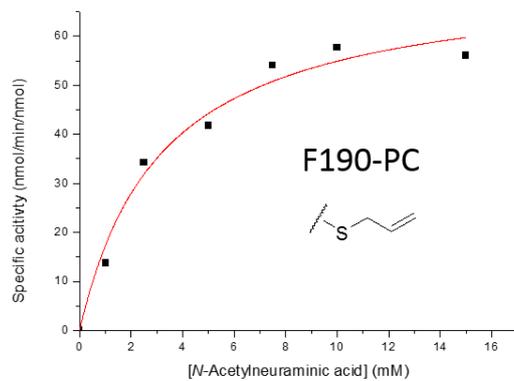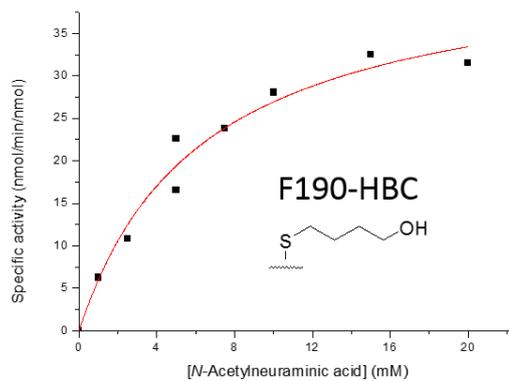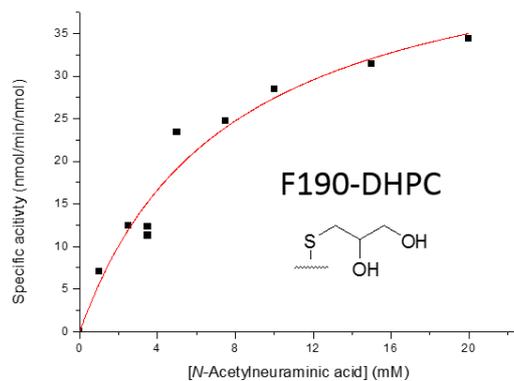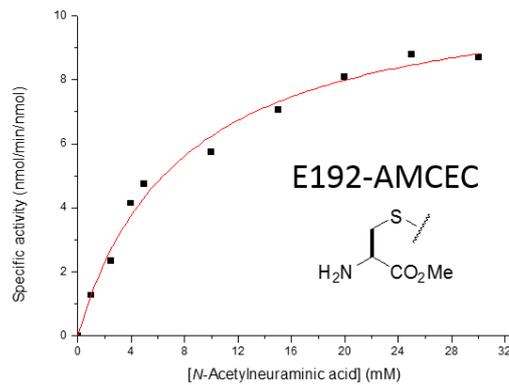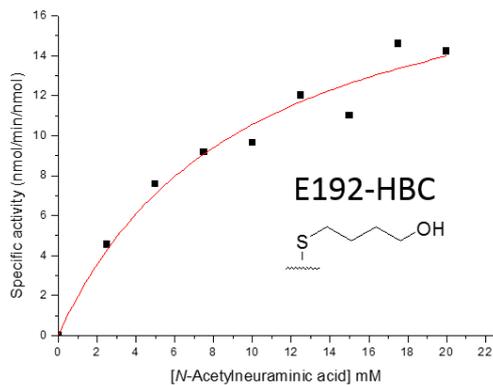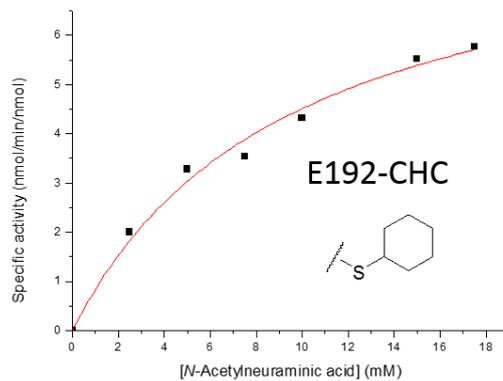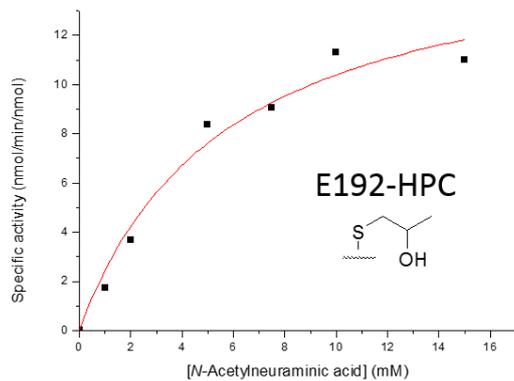
Activity plots of the SaNAL variants with the N-acetylneuraminic acid analogue DPAH as a substrate:
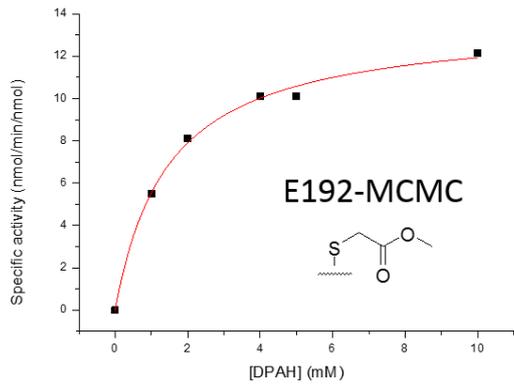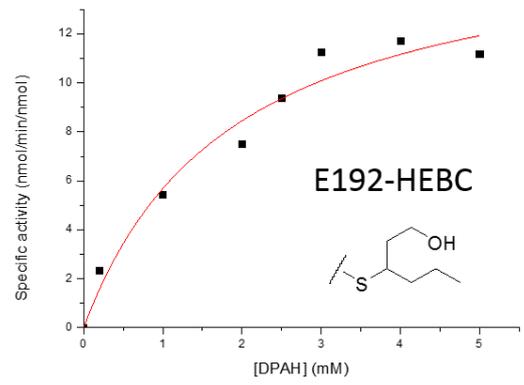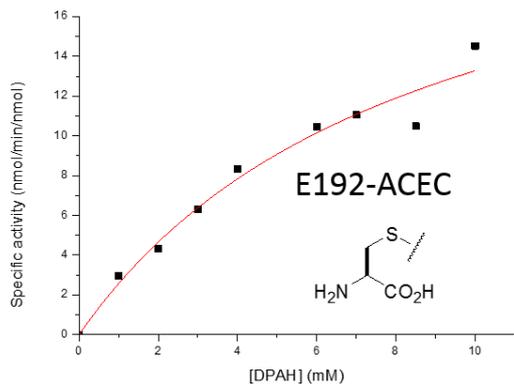
Activity plots of the SaNAL variants with the N-acetylneuraminic acid analogue ATOA as a substrate:
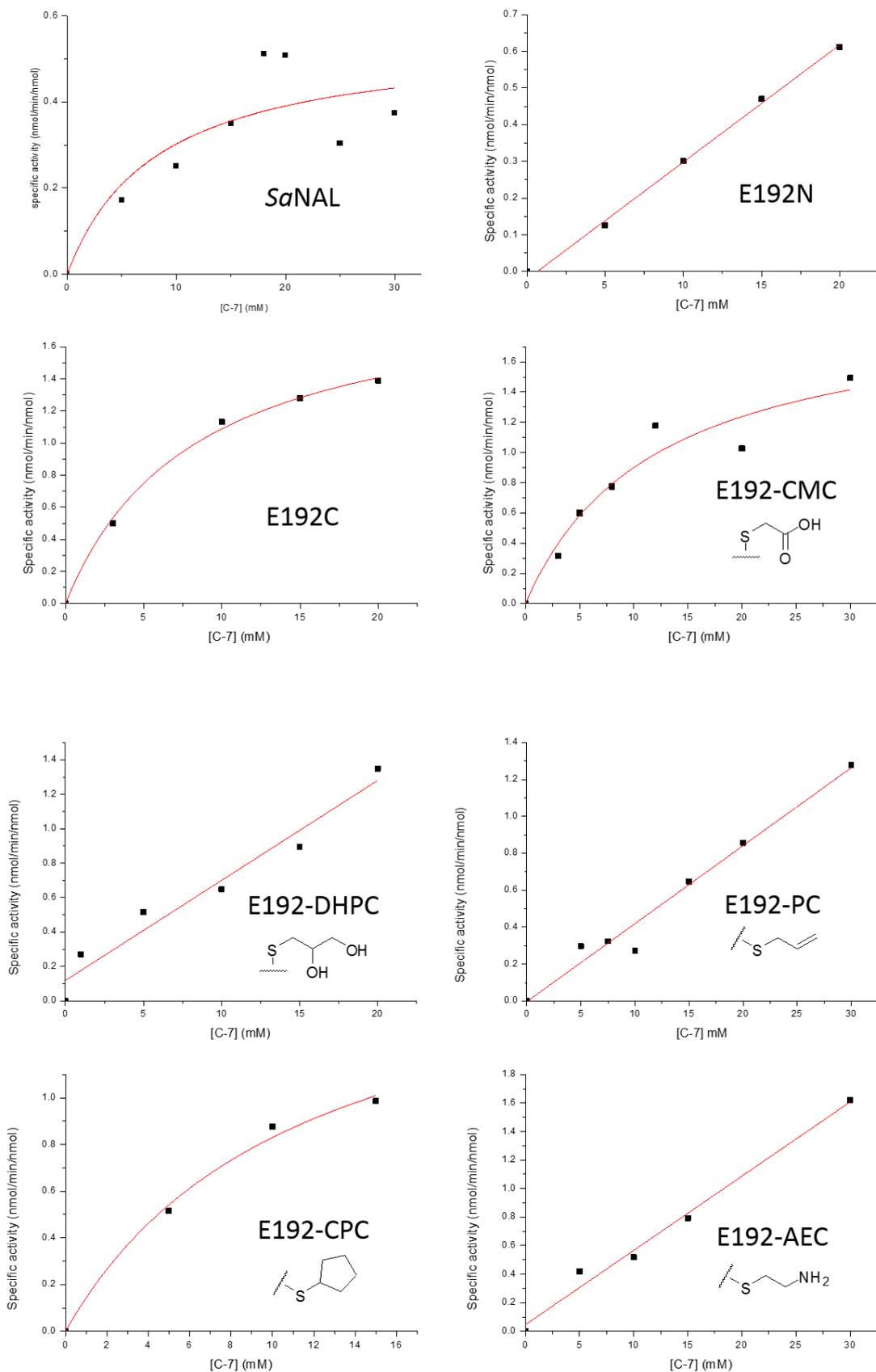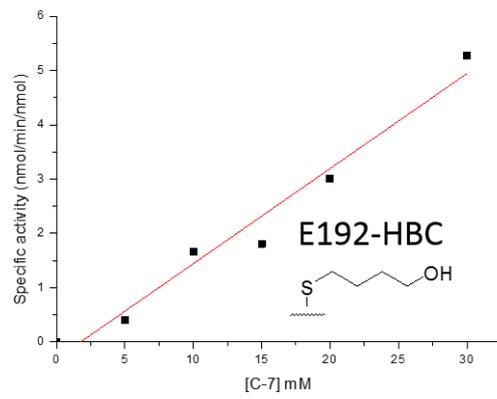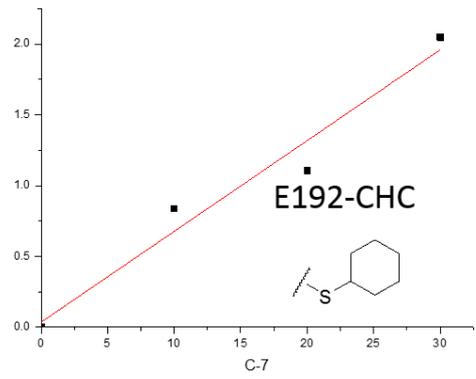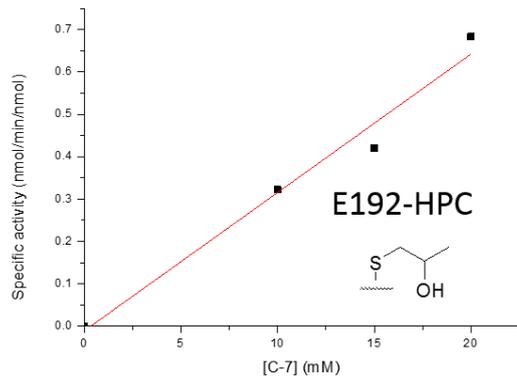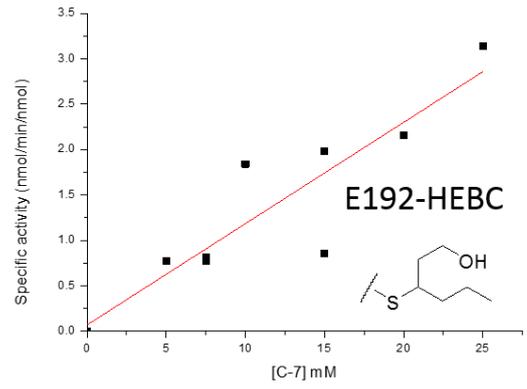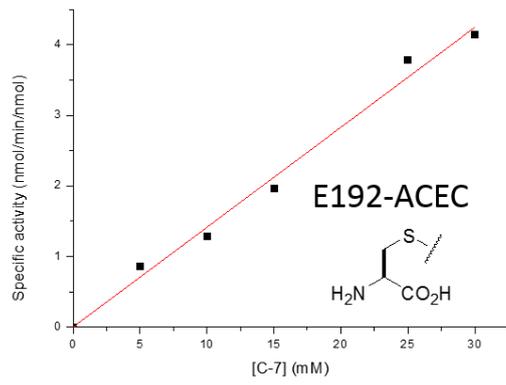
Appendix III. supplementary future work results and discussion

As discussed in section 3.5, the α,β-unsaturated carbonyl compounds **73**, **74** and **75** could potentially act as inhibitors of *Sa*NAL variants. Work towards the synthesis of the α,β-unsaturated carbonyl compounds **74** and **75** had begun. Compounds **73**, **74** and **75** (Figure 47) are to be synthesised as possible E192N aldolase inhibitors.
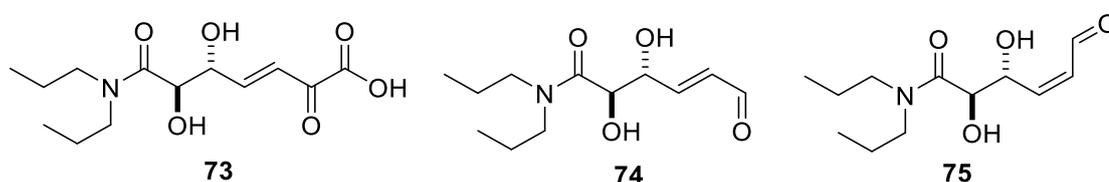
Figure 47 | Three α,β-unsaturated analogues of DPAH to be synthesised.

An obvious starting point for the synthesis of the compounds seen in (Scheme 36) would be the cross metathesis of the compound **53** with an alkene containing a carbonyl moiety such as compound **79**. This route was chosen as the synthesis of compound **53** had already been established.
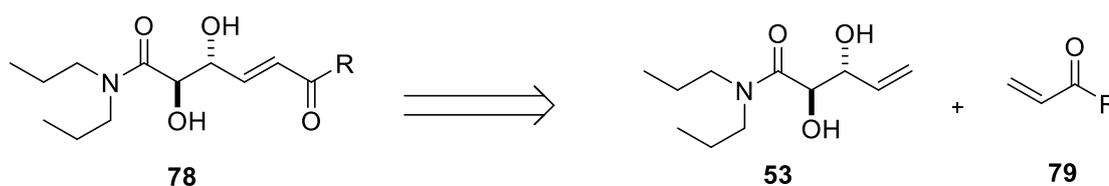
Scheme 36| proposed cross metathesis of compound 53 to yield compound 78.

Compound **78** would contain functionality which would hopefully (with little further chemical modification) lead to the synthesis of the compounds **73**, **74** and **75**. Therefore several cross metathesis reactions were attempted with the compound **53** and the alkenes **80, 81, 82** and **83** (Figure 48).
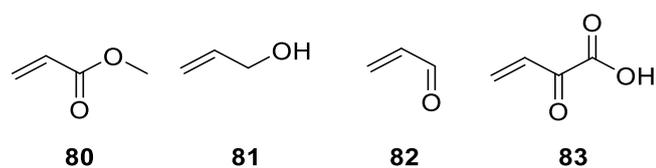
Figure 48 | Alkenes used in cross metathesis reactions with compound 53.

The metathesis reaction between compound **53** and **82**, was expected to give the compound **84**, but surprisingly gave the furan **85**, in 38% yield (Scheme 37).
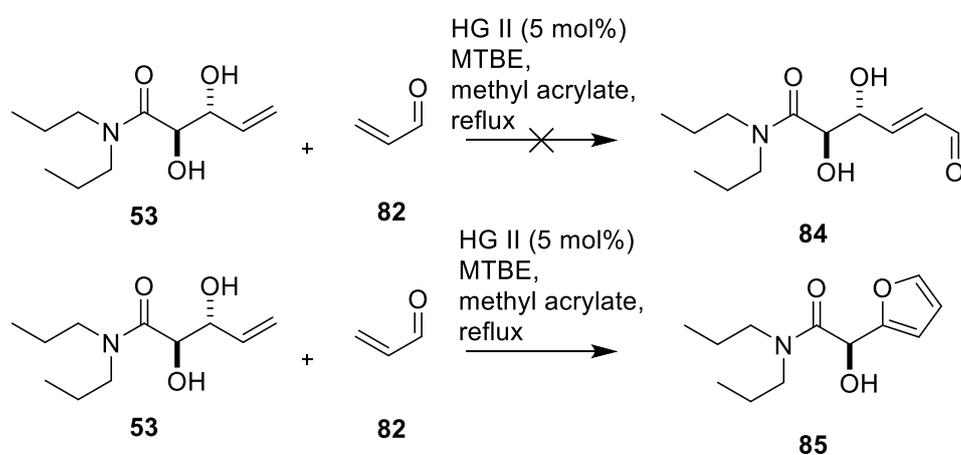
Scheme 37| cross metathesis reaction with acrolein 82 and the alkene 53. This reaction gave the furan **85.**

An Achmatowicz reaction could possibly be performed to open the furan ring and provide the α,β-unsaturated carbonyl **75** (Scheme 38).
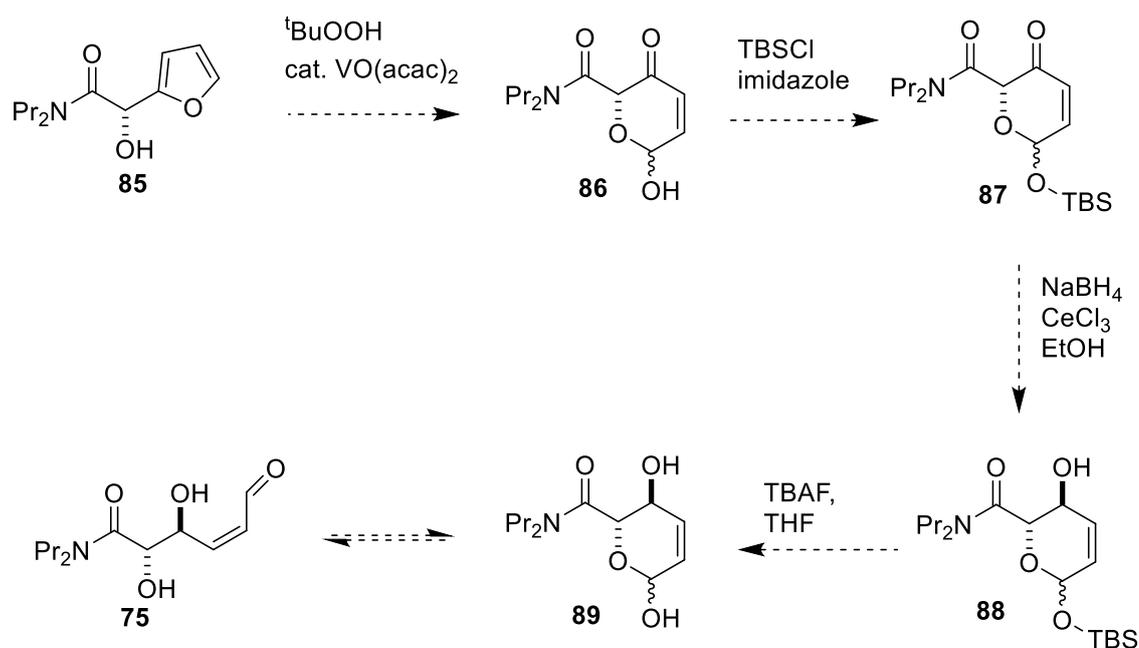
Scheme 38| Proposed Achmatowicz reaction leading to the α,β-unsaturated carbonyl compound 75.

An alternate route proposed for the synthesis of the α,β-unsaturated carbonyl compound **74** can be seen in Scheme 39. The route begins with the cross metathesis with the compound **53** and methyl acrylate **80** which gave the methyl ester **90** in 63% yield. After cross metathesis, the diol of compound **90** was protected to give the acetonide **91** in 72% yield. After this an oxidation, reduction and deprotection may lead to the compound **74**. A preliminary reduction of the methyl ester **91** with DIBAL-H has yielded a low quantity of the alcohol **92** (observed by 500 MHz ^{1}H NMR). With optimisation of this reaction, followed by oxidation and deprotection, **74** may result.
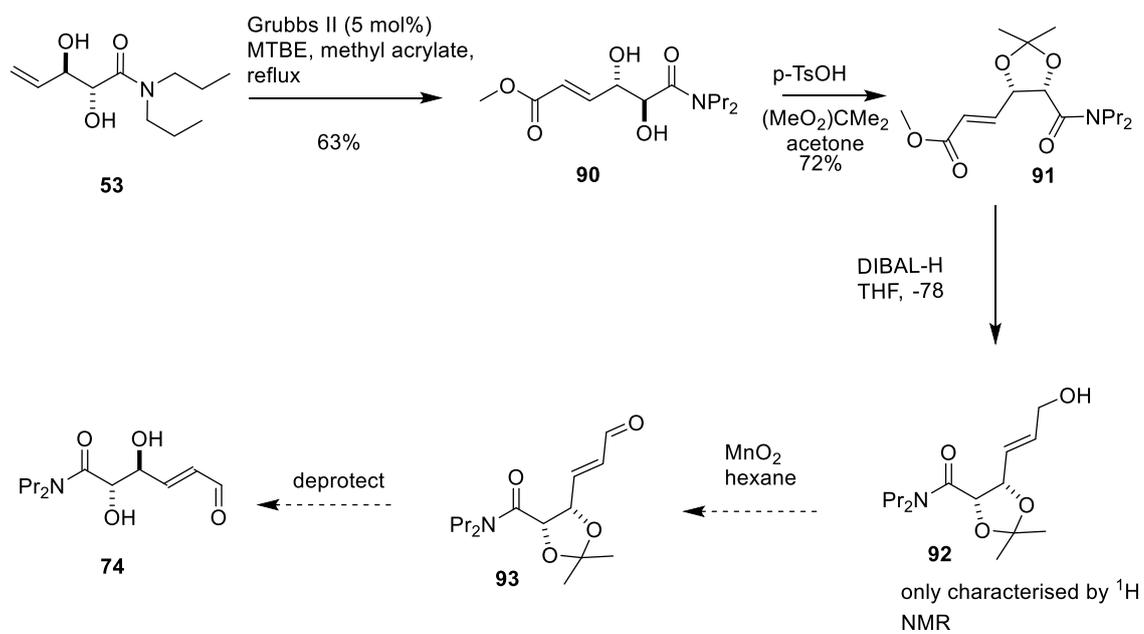
Scheme 39 | Proposed synthesis of the α,β-unsaturated carbonyl compounds 74.

References

(1) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme and Protein Folding.* W. H. Freeman & Co., New York, 1999.

(2) Ma, A., Hu, J., Karplus, M. and Dinner, A. R. Implications of alternative substrate binding modes for catalysis by uracil-DNA glycosylase: An apparent discrepancy resolved. *Biochemistry* **2006**, *45*, 13687-13696.

(3) Cech, T. R. Ribozymes, the first 20 years. *Biochem. Soc. Trans.* **2002**, *30*, 1162-1166.

(4) Li, X. Y. Liu, D. R. DNA-Templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem., Int. Ed.* **2004**, *43*, 4848-4870.

(5) Hammes-Schiffer, S.; Benkovic, S. J.: Relating protein motion to catalysis. In *Annu. Rev. Biochem.* 2006; Vol. 75; pp 519-541.

(6) Perry A. Frey, A. D. H.: *Enzymatic Reaction Mechanisms*; Oxford University Press, 2007.

(7) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Huebner, C. G.; Kern, D.: Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838-844.

(8) Hedstrom, L.: Serine protease mechanism and specificity. *Chem. Rev.* **2002**, *102*, 4501-4523.

(9) Kraut, J.: Serine Proteases - Structure and mechanism of catalysis. *Annu. Rev. Biochem.* **1977**, *46*, 331-358.

(10) Blow, D. M.: The tortuous story of Asp ... His ... Ser: structural analysis of alpha-chymotrypsin. *Trends Biochem. Sci.* **1997**, *22*, 405-408.

(11) Leskovac, V.; Trivic, S.; Pericin, D.; Popovic, M.; Kandrac, J.: Short hydrogen bonds in the catalytic mechanism of seirine proteases. *J. Serb. Chem. Soc.* **2008**, *73*, 393-403.

(12) Gefflaut, T.; Blonski, C.; Perie, J.; Willson, M.: Class I aldolases: Substrate specificity, mechanism, inhibitors and structural aspects. *Prog. Biophys. Mol. Biol.* **1995**, *63*, 301-340.

(13) Wang, W. J.; Mazurkewich, S.; Kimber, M. S.; Seah, S. Y. K.: Structural and Kinetic Characterization of 4-Hydroxy-4-methyl-2-oxoglutarate/4-Carboxy-4-hydroxy-2-oxoadipate Aldolase, a Protocatechuate Degradation Enzyme Evolutionarily Convergent with the HpaI and DmpG Pyruvate Aldolases. *J. Biol. Chem.* **2010**, *285*, 36608-36615.

(14) Bolt, A.; Berry, A.; Nelson, A.: Directed evolution of aldolases for exploitation in synthetic organic chemistry. *Arch. Biochem. Biophys.* **2008**, *474*, 318-330.

(14b) Patel, R.: *Biocatalysis in the Pharmaceutical and Biotechnology Industries*. CRC Press, 2006.

(15) Daniels, A. D.; Campeotto, I.; van der Kamp, M. W.; Bolt, A. H.; Trinh, C. H.; Phillips, S. E. V.; Pearson, A. R.; Nelson, A.; Mulholland, A. J.; Berry, A.: Reaction Mechanism of N-

Acetylneuraminic Acid Lyase Revealed by a Combination of Crystallography, QM/MM Simulation, and Mutagenesis. *Acs Chem. Biol.* **2014**, *9*, 1025-1032.

(16)   Campeotto, I.; Bolt, A. H.; Harman, T. A.; Dennis, C.; Trinh, C. H.; Phillips, S. E. V.; Nelson, A.; Pearson, A. R.; Berry, A.: Structural Insights into Substrate Specificity in Variants of N-Acetylneuraminic Acid Lyase Produced by Directed Evolution. *J. Mol. Biol.* **2010**, *404*, 56-69.

(17)   Dobson, R. C. J.; Perugini, M. A.; Jameson, G. B.; Gerrard, J. A.: Specificity versus catalytic potency: The role of threonine 44 in Escherichia coli dihydrodipicolinate synthase mediated catalysis. *Biochimie* **2009**, *91*, 1036-1044.

(18)   Dobson, R. C. J.; Valegard, K.; Gerrard, J. A.: The crystal structure of three site-directed mutants of Escherichia coli dihydrodipicolinate synthase: Further evidence for a catalytic triad. *J. Mol. Biol.* **2004**, *338*, 329-339.

(19)   Theodossis, A.; Walden, H.; Westwick, E. J.; Connaris, H.; Lamble, H. J.; Hough, D. W.; Danson, M. J.; Taylor, G. L.: The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in Sulfolobus solfataricus. *J. Biol. Chem.* **2004**, *279*, 43886-43892.

(20)   Walsh, C. T.: *Posttranslational Modification of Proteins. Expanding Nature's Inventory*; Roberts and Company, 2006.

(21)   Bijlmakers, M. J.; Marsh, M.: The on-off story of protein palmitoylation. *Trends. Cell. Biol.* **2003**, *13*, 32-42.

(22)   Shabb, J. B.: Physiological substrates of cAMP-dependent protein kinase. Chem. Rev. **2001**, *101*, 2381-2411.

(23)   Barford, D.: Molecular mechanisms of the protein serine threonine phosphatases. *Trends Biochem. Sci.* **1996**, *21*, 407-412.

(24)   Andrew Miller, J. T.: *Essentials of Chemical Biology*; Wiley, 2008.

(25)   Davidson, V. L.: Protein-derived cofactors. Expanding the scope of post-translational modifications. *Biochemistry* **2007**, *46*, 5283-5292.

(26)   Davidson, V. L.: Protein-derived Cofactors. In *eLS*; John Wiley & Sons, Ltd, 2001.

(27)   Okeley, N. M.; van der Donk, W. A.: Novel cofactors via post-translational modifications of enzyme active sites. *Chemistry & Biology* **2000**, *7*, R159-R171.

(28)   Heuts, D. P. H. M.; Scrutton, N. S.; McIntire, W. S.; Fraaije, M. W.: What's in a covalent bond? On the role and formation of covalently bound flavin cofactors. *Febs Journal* **2009**, *276*, 3405-3427.

(29)   Schwede, T. F.; Retey, J.; Schulz, G. E.: Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* **1999**, *38*, 5355-5361.

(30)   Vanpoelje, P. D.; Snell, E. E.: Pyruvoyl-dependent enzymes. *Annu. Rev. Biochem.* **1990**, *59*, 29-59.

(31) Recsei, P. A.; Snell, E. E.: Pyruvoyl enzymes. *Annu. Rev. Biochem.* **1984**, *53*, 357-387.

(32) Gallagher, T.; Rozwarski, D. A.; Ernst, S. R.; Hackert, M. L.: Refined structure of the pyruvoyl-dependent histidine-decarboxylase from lactobacillus-30A. *J. Mol. Biol.* **1993**, *230*, 516-528.

(33) Rose, I. A.; Kuo, D. J.: Role of CO_2 in proton activation by histidine-decarboxylase (pyruvoyl). *Biochemistry* **1992**, *31*, 5887-5892.

(34) Diaz-Rodriguez, A.; Davis, B. G.: Chemical modification in the creation of novel biocatalysts. *Curr. Opin. Chem. Biol.* **2011**, *15*, 211-219.

(35) Thapa, P.; Zhang, R.-Y.; Menon, V.; Bingham, J.-P.: Native chemical ligation: a boon to peptide chemistry. *Molecules (Basel, Switzerland)* **2014**, *19*, 14461-83.

(36) Malins, L. R.; Payne, R. J.: Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr. Opin. Chem. Biol.* **2014**, *22*, 70-78.

(37) Palomo, J. M.: Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *Rsc Advances* **2014**, *4*, 32658-32672.

(38) De Rosa, L.; Russomanno, A.; Romanelli, A.; D'Andrea, L. D.: Semi-Synthesis of Labeled Proteins for Spectroscopic Applications. *Molecules* **2013**, *18*, 440-465.

(39) Raibaut, L.; Ollivier, N.; Melnyk, O.: Sequential native peptide ligation strategies for total chemical protein synthesis. *Chem. Soc. Rev.* **2012**, *41*, 7001-7015.

(40) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G.: Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.

(41) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G.: Expanding the genetic code of Escherichia coli. *Science* **2001**, *292*, 498-500.

(42) Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G.: A general-method for site-specific incorporation of unnatural amino-acids into proteins. *Science* **1989**, *244*, 182-188.

(43) Xie, J. M.; Schultz, P. G.: Adding amino acids to the genetic repertoire. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548-554.

(44) Zhang, Y.; Gladyshev, V. N.: High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Res.* **2007**, *35*, 4952-4963.

(45) Stadtman, T. C.: Selenocysteine. In *Annual Review of Biochemistry*; Richardson, C. C., Ed., 1996; Vol. 65; pp 83-100.

(46) Atkins, J. F.; Gesteland, R.: Biochemistry - The 22nd amino acid. *Science* **2002**, *296*, 1409-1410.

(47) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. o acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **2003**, *125*, 11782-11783.

(48) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G.: Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **2015,** 6, 50-69.

(49) Wang, J.; Xie, J.; Schultz, P. G.: A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* **2006**, *128*, 8738-8739.

(50) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G.: An expanded eukaryotic genetic code. *Science* **2003**, *301*, 964-967.

(51) Liu, C. C.; Schultz, P. G.: Adding New Chemistries to the Genetic Code. In *Annual Review of Biochemistry, Vol 79*; Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., Thorner, J. W., Eds., 2010; Vol. 79; pp 413-444.

(52) Young, T. S.; Schultz, P. G.: Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.* **2010**, *285*, 11039-11044.

(53) Basle, E.; Joubert, N.; Pucheault, M.: Protein Chemical Modification on Endogenous Amino Acids. *Chemistry & Biology* **2010**, *17*, 213-227.

(54) Means, G. E.; Feeney, R. E.: Chemical Modifications of Proteins: History and Applications. *Bioconjugate Chem.* **1990**, *1*, 2-12.

(55) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G.: Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem.-Asian J.* **2009**, *4*, 630-640.

(56) Tawfik, D. S.: Side Chain Selective Chemical Modifications of Proteins. In *Protein Protocols Handbook, Third Edition*; Walker, J. M., Ed., 2009; pp 851-854.

(57) Lundblad, R.: Chemical modification of amino acid side-chains. Academic, 1996; pp 287-298.

(58) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G.: A "Tag-and-Modify" Approach to Site-Selective Protein Modification. *Acc. Chem. Res.* **2011**, *44*, 730-741.

(59) Takaoka, Y.; Ojida, A.; Hamachi, I.: Protein Organic Chemistry and Applications for Labeling and Engineering in Live-Cell Systems. *Angew. Chem., Int. Ed.* **2013**, *52*, 4088-4106.

(60) Joshi, N. S.; Whitaker, L. R.; Francis, M. B.: A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J. Am. Chem. Soc.* **2004**, *126*, 15942-15943.

(61) Hermanson, G. T.: *Bioconjugate techniques*, 1996.

(62) Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B.: Glycomethanethiosulfonates: powerful reagents for protein glycosylation. *Tetrahedron-Asymmetry* **2000**, *11*, 245-262.

(63) Wang, J.; Schiller, S. M.; Schultz, P. G.: A Biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem., Int. Ed.* **2007**, *46*, 6849-6851.

(64) Ke, Z.; Smith, G. K.; Zhang, Y.; Guo, H.: Molecular Mechanism for Eliminylation, a Newly Discovered Post-Translational Modification. *J. Am. Chem. Soc.* **2011**, *133*, 11103-11105.

(65) Zhu, Y.; van der Donk, W. A.: Convergent synthesis of peptide conjugates using dehydroalanines for chemoselective ligations. *Org. Lett.* **2001**, *3*, 1182-1192.

(66) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G.: Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: Versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* **2008**, *130*, 5052-5053.

(67) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G.: Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem., Int. Ed.* **2012**, *51*, 1835-1839.

(68) Traving, C.; Schauer, R.: Structure, function and metabolism of sialic acids. *Cell. Mol. Life. Sci.* **1998**, *54*, 1330-1349.

(69) Varki, A.: *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press. 1999.

(70) Wong, C.-H.; Halcomb, R. L.; Ichikawa, Y.; Kajimoto, T.: Enzymes in organic synthesis: application to the problems of carbohydrate recognition. Part 1. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 412-32.

(71) Woodhall, T.; Williams, G.; Berry, A.; Nelson, A.: Creation of a tailored aldolase for the parallel synthesis of sialic acid mimetics. *Angew. Chem., Int. Ed.* **2005**, *44*, 2109-2112.

(72) Williams, G. J.; Woodhall, T.; Nelson, A.; Berry, A.: Structure-guided saturation mutagenesis of N-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics. *Protein Eng. Des. Sel.* **2005**, *18*, 239-246.

(73) Sollis, S. L.; Smith, P. W.; Howes, P. D.; Cherry, P. C.; Bethell, R. C.: Novel inhibitors of influenza sialidase related to GG167 - Synthesis of 4-amino and guanidino-4H-pyran-2-carboxylic acid-6-propylamides; Selective inhibitors of influenza A virus sialidase. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1805-1808.

(74) Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Cobley, K. N.; Taylor, H.; Whittington, A. R.; Scicinski, J.; Bethell, R. C.; Taylor, N.; Skarzynski, T.; Cleasby, A.; Singh, O.; Wonacott, A.; Varghese, J.; Colman, P.: Novel inhibitors of influenza sialidases related to GG167 - Structure-activity, crystallographic and molecular dynamic studies with 4H-pyran-2-carboxylic acid 6-carboxamides. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2931-2936.

(75) Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Starkey, I. D.; Cobley, K. N.; Weston, H.; Scicinski, J.; Merritt, A.; Whittington, A.; Wyatt, P.; Taylor, N.; Green, D.; Bethell, R.; Madar, S.; Fenton, R. J.; Morley, P. J.; Pateman, T.; Beresford, A.: Dihydropyrancarboxamides related to zanamivir: A new series of inhibitors of influenza virus sialidases. 1. Discovery, synthesis, biological activity, and structure-activity relationships of 4-guanidino- and 4-amino-4H-pyran-6-carboxamides. *J. Med. Chem.* **1998**, *41*, 787-797.

(76) Woodhall, T.: Directed Evolution of Enzymes with Synthetically Useful Activity: Parallel Synthesis of Novel Sialic Acid Analogues. University of Leeds, 2004.

(77) Williams, G. J.; Woodhall, T.; Farnsworth, L. M.; Nelson, A.; Berry, A.: Creation of a pair of stereochemically complementary biocatalysts. *J. Am. Chem. Soc.* **2006**, *128*, 16238-16247.

(78)   Timms, N.; Windle, C. L.; Polyakova, A.; Ault, J. R.; Trinh, C. H.; Pearson, A. R.; Nelson, A.; Berry, A.: Structural Insights into the Recovery of Aldolase Activity in N-Acetylneuraminic Acid Lyase by Replacement of the Catalytically Active Lysine with gamma-Thialysine by Using a Chemical Mutagenesis Strategy. *Chembiochem* **2013**, *14*, 474-481.

(79)   Warren, L.: Thiobarbituric acid assay of sialic acids. *J. Biol. Chem.* **1959**, *234*, 1971-1975.

(80)   Warren, L.: Thiobarbituric acid assay of sialic acids. In *Methods in Enzymology*; Academic Press, 1963; Vol. Volume 6; pp 463-465.

(81)   Hong, Z. Y.; Liu, L.; Hsu, C. C.; Wong, C. H.: Three-step synthesis of sialic acids and derivatives. *Angew. Chem., Int. Ed.* **2006**, *45*, 7417-7421.

(82)   Veronesi, P. A.; Rodriguez, P. E. A.; Peschechera, E.; Veronesi, A.: Preparation of sialic acid analogs for use as prodrugs which inhibit neuraminidase, hemagglutinin and structural M2 protein bearing viruses.; Organisation. 2008, PCT Int. Appl. **2008**, WO 2008090 151 A1 20080731.

(83)   Harman, T.: PhD Thesis. University of Leeds, 2011.

(84)   Mann, M.; Meng, C. K.; Fenn, J.B.; Interpreting mass spectra of multiply charged ions. *Anal. Chem.*, **1989**, 61, 1702-1708.

(84b)   Timms, N.: PhD Thesis. University of Leeds, 2011.

(85)   Blostein, R.; Rutter, W. J.: Comparative studies of liver and muscle aldolase: II. Immunochemical and chromatographic differentiation. *J. Biol. Chem.* **1963**, *238*, 3280-3285.

(86)   Krueger, D.; Schauer, R.; Traving, C.: Characterization and mutagenesis of the recombinant N-acetylneuraminate lyase from Clostridium perfringens: Insights into the reaction mechanism. *Eur. J. Biochem.* **2001**, *268*, 3831-3839.

(87)   Bolt, A.: PhD Thesis. University of Leeds, 2008.

(88)   Barbosa, J.; Smith, B. J.; DeGori, R.; Ooi, H. C.; Marcuccio, S. M.; Campi, E. M.; Jackson, W. R.; Brossmer, R.; Sommer, M.; Lawrence, M. C.: Active site modulation in the N-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed Haemophilus influenzae enzyme. *J. Mol. Biol.* **2000**, *303*, 405-421.

(89)   Mukherjee, S.; Yang, J. W.; Hoffmann, S.; List, B.: Asymmetric Enamine Catalysis. *Chem. Rev.* **2007**, *107*, 5471-5569.

(90)   Pellissier, H.: Recent developments in asymmetric organocatalysis. *Adv. Synth. & Catal*. **2012**, *354,* 237-294.

(91)   Eder, U.; Sauer, G.; Weichert, R.: Total synthesis of optically active steroids .6. New type of asymmetric cyclization to optically active steroid cd partial structures. *Angew. Chem., Int. Ed.* **1971**, *10*, 496-497.

(92)   Huang, Y.; Walji, A. M.; Larsen, C. H.; MacMillan, D. W. C.: Enantioselective organo-cascade catalysis. *J. Am. Chem. Soc.* **2005**, *127*, 15051-15053.

(93)   Jen, W. S.; Wiener, J. J. M.; MacMillan, D. W. C.: New strategies for organic catalysis: The first enantioselective organocatalytic 1,3-dipolar cycloaddition. *J. Am. Chem. Soc.* **2000**, *122*, 9874-9875.

(94)   Xu, L.-W.; Lu, Y.: Primary amino acids: privileged catalysts in enantioselective organocatalysis. *Org. Biomol. Chem.* **2008**, *6*, 2047-2053.

(95)   Yang, J. W.; Fonseca, M. T. H.; List, B.: Catalytic asymmetric reductive Michael cyclization. *J. Am. Chem. Soc.* **2005**, *127*, 15036-15037.

(96)   Erkkila, A.; Majander, I.; Pihko, P. M.: Iminium catalysis. *Chem. Rev.* **2007**, *107*, 5416-5470.

(97)   A. Berkessel, H. G.: *Asymmetric Organocatalysis. From Biomimetic Concepts to Applications in Asymmetric Synthesis*; Wiley-VCH, 2005.

(98)   Still, W. C.; Kahn, M.; Mitra, A.: Rapid chromatographic technique for preparative separations with moderate resolution. *J. Org. Chem.* **1978**, *43*, 2923-2925.

(99)   Chooi, K. P.; Galan, S. R. G.; Raj, R.; McCullagh, J.; Mohammed, S.; Jones, L. H.; Davis, B. G.: Synthetic Phosphorylation of p38α Recapitulates Protein Kinase Activity. *J. Am. Chem. Soc.* **2014**, *136*, 1698-1701.

(100)   Bernton, A. W.; Ing, H. R.; Perkin, W. H.: Studies in the configuration of alpha alpha'-dibromodibasic acids Part II Derivatives of adipic acid. *J. Chem. Soc.* **1924**, *125*, 1492-1502.

(101)   Dicken, C. M.; Deshong, P.: Reactions at high-pressure - 3+2 dipolar cycloaddition of nitrones with vinyl ethers. *J. Org. Chem.* **1982**, *47*, 2047-2051.

(102)   Woodhall, T.; Williams, G.; Berry, A.; Nelson, A.: Synthesis of screening substrates for the directed evolution of sialic acid aldolase: towards tailored enzymes for the preparation of influenza A sialidase inhibitor analogues. *Org. Biomol. Chem.* **2005**, *3*, 1795-1800.

(103)   Woodhall, T.: PhD Thesis. University of Leeds, 2005.

(104)   Nathani, R.; Moody, P.; Smith, M. E. B.; Fitzmaurice, R. J.; Caddick, S.: Bioconjugation of Green Fluorescent Protein via an Unexpectedly Stable Cyclic Sulfonium Intermediate. *ChemBioChem* **2012**, *13*, 1283-1285.