Protein based molecular probes by unnatural amino acid incorporation

Peter Crane

Linacre College

Oxford University

A Thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2018

Abstract

The “tag & modify” strategy for protein modification relies upon the genetic incorporation of an uncommon or unnatural amino acid into a protein backbone, followed by a chemo-selective modification to yield differentially modified proteins. This thesis describes the creation of a protein-based glycoconjugate tool for interrogating biological function.

In Chapter 2, the unnatural amino acid, azidohomoalanine was genetically incorporated into a library of distance defined Np276 proteins via a selective pressure incorporation. Methods to prevent the common post translational modification N-terminal gluconylation were identified, including preliminary work on a small molecule intervention. The proteins were subsequently characterised with respect to other members of the (limited) family of pentapeptide repeat protein and the key biophysical parameters (T_M, stability) with relate to it being a multivalent scaffold were investigated.

In Chapter 3, An initial investigation into obtaining a selective amine modification initially via N-hydroxysuccinimide esters, led to the discovery (and characterisation) of a clear selectivity for N-terminal proline Isothiocyanate modification. The dual modification of proteins via the N-term Pro & the ubiquitous (glyco) copper(I)-catalysed azide alkyne cycloaddition was subsequently used to generate homogenously dual modified Np276 scaffolds.

In Chapter 4, these proteins were then used in a FACS assay against a murine sialoadhesin - chinese hamster ovary cell line, the results showing promise for the further development of multivalent glycated probes of function.

Acknowledgements

I would like to take the opportunity to thank all my co-workers, supervisors, friends, family and colleagues for their support and guidance in developing and producing this thesis. It has been a real journey of personal discovery but one that I leave considerably stronger than when I entered.

To begin with, I would like to thank Professor Ben Davis (Ben). Ben has been a guiding light throughout the course of my D.Phil. providing scientific inspiration and guidance, as well as crucial life lessons on how to manage my time more effectively and to free my mind from the tedious tasks that were often distracting. These skills along with embracing a broader view of science are things I will take forward with me and will continue to use in the future.

Secondly to Ben, I would like to extend my warmest gratitude to Mr Matt Selby from UCB Celltech. Matt has been a rock throughout the whole of my studies. Providing guidance, support and help continuously. Matt's contribution cannot be underestimated, and I am forever grateful for his patience and ability.

Thirdly I would also like to thank: Prash, David, Alistair and Carl from UCB Celltech. In particular, to Prash and Carl for taking me under their wings in the protein expression laboratories, teaching me how to do my molecular biology in a more efficient and stress-free way and instructing me towards the joys of "owning a Porsche".

I have worked with many talented colleagues and fellow students during my time in Oxford. It would be impossible for me to thank them all here but I would like to extend my thanks to: Macarena, Feng, Bala, Yiqun, Thomas Parsons and Sylvain (Post-Docs) and my fellow students (Tom, Argyha, Leanne, Sonia, Redia, Mahima, Smita, Phin, Seb, Ritu, Rogen, Ben and others. I would also like to single out former group member Rob Simion and Bhaksar out

for their support, especially during the dark years of the “old ICL”. Within the department, I would also like to thank the MS team, the NMR team and Dr Chris Blackwell for all their support and guidance,

I would also like to extend my warmest thanks to Dr Catherine Walter, vice-principal of Linacre college Oxford, for her steady support, guidance and help during my time at Oxford. It has been a pleasure to know you, to discuss work and also “the Gym”.

Also, I would like to extend my thanks to my varied networks of friends outside the laboratory. From all the guys at Linacre (Sami, Dan, Willian), the guys in The Martin (you know who you are), the fabulous gentlemen of the FWS (especially Rev. Jim for his great pastoral advice), the chivalrous gentlemen of Apollo and finally Kelvin & David. It has been a truly pleasurable time and one where again I have learnt huge amounts

I would also like to thank all the many colleagues I have worked with in various organisations across Oxford & the world: Dan, Anna, John, Jon, Sophia, Justin, Susan, Marek, Bauke, Ajoeb, Danuta, Karolina, Lina, Mira, Maria, Anne, Adam and others.

Finally I would like to thank my long-suffering family, especially my mother and father, for their emotional (and financial to be fair) support over the long hard trek to submission. I would also especially like to thank my even longer suffering girlfriend Rachel, for showing extraordinary patience and calmness even when I was perhaps not.

Amino acid single letter code and abbreviations

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

Abbreviations

AA	Amino acid
AARS	Aminoacyl tRNA synthetase
aaRS	Aminoacyl tRNA synthetase
Aha	Azidohomoalanine
Bn	Benzyl
Boc	*tert*-Butyloxycarbonyl
CD	Circular dichroism
CD169	Siglec-1
CD22	Siglec-2
CuAAC	Copper catalysed alkyne azide cycloaddition
DBU	1,8-Diazabicycloundec-7-ene
DCM	Dichloromethane
Dha	Dehydroalanine
DSF	Differential scanning fluorimetry
E. coli	Escherichia coli
ELISA	Enzyme-linked immunosorbent assay
EPO	Erythropoietin
FA	Formic acid
FACS	Fluorescence-activated cell sorting
FITC	Fluorescein isothiocyanate
FPLC	Fast protein liquid chromatography
Hpg	Homopropargylglycine
IPA	2-Propanol
IPhe	*para*-Iodophenylalanine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared
LCT	Liquid chromatography time of flight mass spec
MeCN	Acetonitrile
MetRS	Methionyl-tRNA synthetase
mRNA	Messenger RNA
mSn	Murine sialoadhesin

MW	Molecular weight
MWCO	Molecular weight cut off
NaBu	Sodium butyrate
NCAA	Non-canonical amino acid
ncAA	Non-canonical amino acid
NCL	Native chemical ligation
NHS	N-hydroxy succinimde
NMR	Nuclear magnetic resonance
PITC	Phenylisothiocyanate
PRP	Pentapeptide repeat protein
PTM	Post translational modification
RMSD	Root mean square deviation
SDP	Sulfodicholorphenol
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sn	Sialoadhesin
SPI	Selective pressure incorporation
SPR	Surface plasmon resonance
TBDMS	tert-Butyldimethylsillyl
TBDPS	tert-Butyldiphenylsillyl
TCA	Trichloroacetamide
TFA	Trifluoroacetic acid
TFM	Trifluoromethionine
TolH	Toluene
tRNA	Transfer RNA
UAA	Unnatural amino acid

Contents Page

Chapter 1

Introduction

Contents Page

1.0 Overview

This thesis aims to develop a series of distance defined multivalent probes of the glycan-binding proteins: the Siglecs. To do this, protein constructs each decorated with a distance spaced pair of immunologically active glycans would need to be synthesised. These could be then used to interrogate the molecular basis of binding in carbohydrate binding proteins. By relating a distance between binding motifs (in this project: sugars) to a biological effect (such as binding), we draw the analogy to a physical measuring device: the ruler. Henceforth, throughout this project, this concept is referred to as a "molecular ruler"[*]. To construct a molecular ruler, it was necessary to first decide upon an appropriate scaffold. In this case of this project, we decided to investigate a pentapeptide repeat unit protein due to it having regular distance spaced 'coils', glycans could then be installed upon this scaffold at different discrete distances apart from each other. It was then necessary to develop a way to incorporate glycans onto the protein in a site specific manner (through UAA incorporation and bioconjugation chemistry), as well as introduce a means of following the subsequent biological assays (e.g. via a dye incorporation). This would give protein constructs with the glycan binding motifs separated by a distance as defined by the protein backbone. By installing glycans at different positions, different distances could be created and a family of distance spaced probes generated. These constructs could then be tested for biological activity, with the hope of elucidating a preferential binding distance for a receptor specific to the glycan used. This introduction provides a broad overview of the topics pertinent to construction and testing of the molecular ruler concept: multivalency, pentapeptide repeat proteins, unnatural amino acid incorporation, bioconjugation and finally the family of receptors under investigation: the Siglecs.

[*] Within this project we often abbreviate this to just "ruler" or in the case of protein expressions, "ruler proteins".

1.1 Multivalency

Whilst the binding of monovalent ligands to receptors may be individually weak ($K_d^{affinity}$~mM/μM), by simultaneously binding multiple ligands (attached to the same linker), the overall binding strength or avidity† is increased dramatically ($K_d^{avidity}$~nM).[1,2] It is this observation that is central to the design and use of multivalent ligands for protein receptors. Whilst there are many theoretical models to understand multivalent systems, a common equation, originally proposed by Whitesides *et al*, is illustrated in Figure 1.[3–5]

$$\Delta G_N^\circ(i) = i\Delta H_{affinity}^\circ - iT\Delta S_{affinity}^\circ + (i-1)T\Delta S_{trans+rot}^\circ + (i-1)\Delta H_{linker}^\circ - (i-1)T\Delta S_{conf}^\circ + (i-1)\Delta G_{coop}^\circ - RT\ln\left(\frac{\Omega_i}{\Omega_0}\right)$$

Figure 1. The theoretical contributions to the binding Gibbs free energy from a multivalent system with N ligands and with i receptor-bound ligands.[1]

In this model, the first two terms ($i(\Delta H_{affinity}^\circ - T\Delta S_{affinity}^\circ)$) represent the standard Gibbs free energy for monovalent binding multiplied by the number of bound ligands, i.[6,7] The second term,$(i-1)T\Delta S_{trans+rot}^\circ$, represents the un-favourable entropic contribution of the receptor and ligand binding.[8–10] The $(i-1)\Delta H_{linker}^\circ$ term accounts for any binding of the linker or construct to the receptor. The $(i-1)T\Delta S_{conf}^\circ$ term accounts for the conformational changes of the linker, the $(i-1)\Delta G_{coop}^\circ$ for the possibility of cooperativity (i.e. the ligand binding influences subsequent receptor interactions) and the final term accounts for the avidity entropy (statistical degeneracy).[11,12] This model doesn't accurately deal with the potential for receptor aggregation, nor self-aggregation of the ligand system prior to binding. From this model, it

† Avidity refers to the accumulated non-covalent interaction between a receptor and ligand. Affinity refers to a single interaction.

necessarily follows that a bivalent system of 2 ligands (i =2) joined by a rigid linker will minimise the loss of conformational entropy upon binding, [13] but will only show a binding interaction if the distance and geometry of ligand spacing are correct for receptor binding.[14–16] This concept is the guiding principle behind the idea of a 'molecular ruler', a defined distance spaced set of ligands on a protein scaffold. Since the initial studies in this project towards the goal of creating a functioning 'molecular ruler', other groups have developed analogous distanced spaced systems with the aim of demonstrating this concept.[17]

1.1.1 Defined distance multivalent probes

Zhang and Zhao, amongst others, independently looked at the use of polymeric nucleic acids as multivalent scaffolds.[18–21] Zhao used a ssDNA chain comprised of adenine to construct a polyvalent probe decorated with anti-CD20 antibodies. It was then demonstrated that this construct showed enhanced binding to CD20, as well as clustering effects.[22]

In late 2012, a system utilising an ankyrin repeat protein modified with mannose (via a maleimide linkage to a free cysteine) to form a bivalent probe was reported.[23] An assay to determine how mannose spacing affected concanavalin A aggregation[24] was conducted, but to date no follow-up studies have been conducted with other biologically relevant targets.

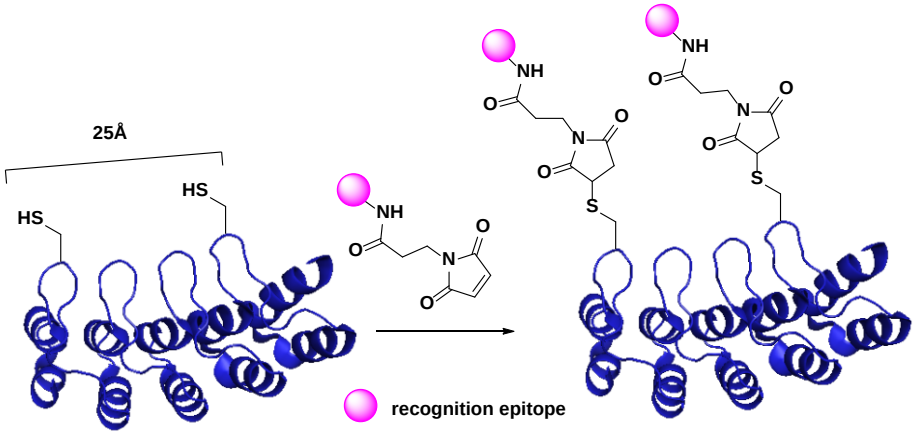


Figure 2. Designed ankyrin repeat protein (Darpin) for use as a multivalent scaffold. Figure created using pdb 2BKG based upon the work of Roberts *et al.*[23]

The most promising results have been in peptide-based multivalent scaffolds. Koktsch *et al* in 2015, showed that the modification of a peptide α-coiled-coil motif with galactose moieties could be used to probe the binding to asialoglycoprotein receptors on hepatocytes by flow cytometry.[25] Elegant work in 2013, by Wennemers *et al* also looked at using distance defined hybrid bombesin ligands with an optimised distance between agonist and antagonist to enable incredibly high tumour uptake and retention.[26] This work relied on the rigid polyproline motif previously investigated by many others in the literature.[27–29]

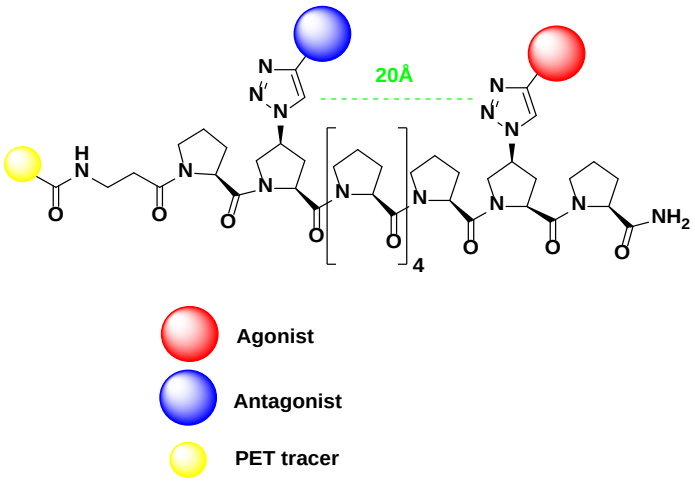


Figure 3. Hybrid bombesin multivalent ligands. A variety of distances (10-30Å) were trialled, it was found that a distance of 20Å between the elements exhibits the highest tumour uptake.

Polyproline was also used in 2015 by Lee *et al* to create rigid chondroitin sulfate modified glycopeptides. These polyproline peptides were conjugated to the sugar via a CuAAC reaction and allowed for the installation of the sugar as discrete distances. These glycopeptides were subsequently used to find a neuronal promoting agent for the NGF (nerve growth factor) mediated signalling pathway.[30] This example represents one of the most advanced examples of the use of a rigid multivalent peptide probe to interrogate sugar-protein binding.

1.1.2 General multivalent probes

Tthe examples listed previously illustrate examples of rigid distance spaced probes of biological function, there are numerous reports of the use of non-distance spaced probes of biological function.[31] These include nanoparticles,[32,33] polymers,[34] liposomes,[35–37] aptamers[38] and dendrimers.[39–42]

Whilst the general applications of this variety of probes is beyond the scope of this review, there are many examples of glycan modified probes being used to interrogate the biological functions of sugar binding proteins. In addition, small molecules are commonly used as the scaffold to construct multivalent probes of lectins. A selection of these (along with some macromolecular examples) are highlighted in Figure 4.[43]

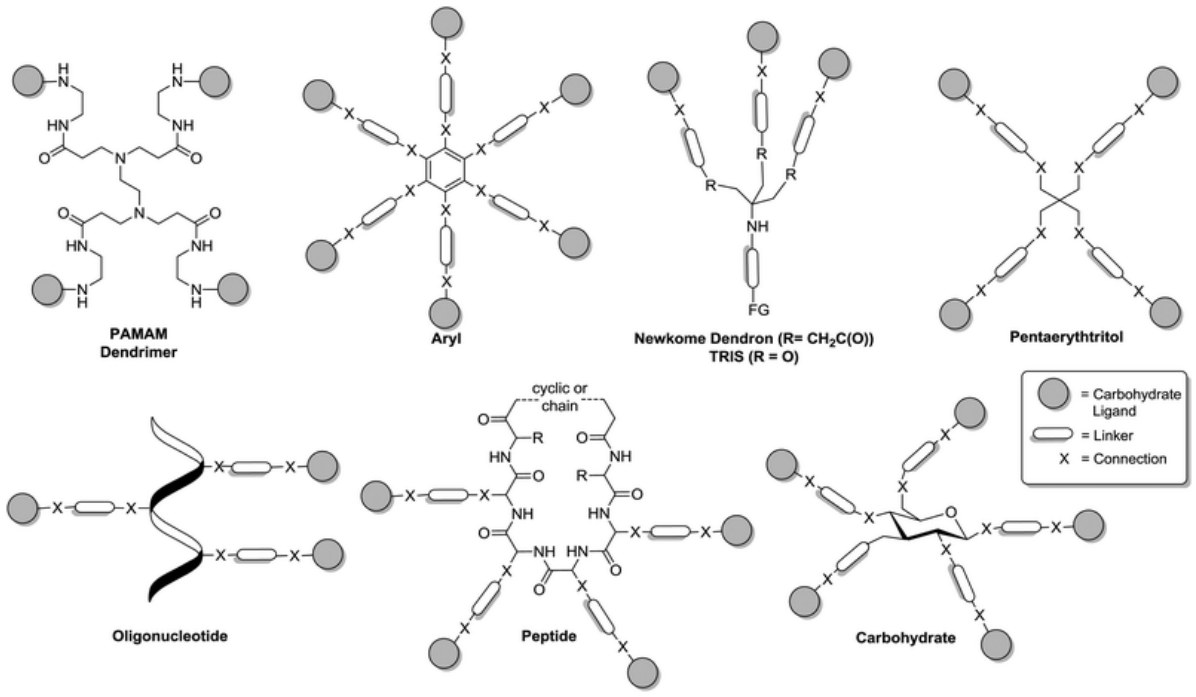


Figure 4. Examples of commonly used multivalent scaffolds, reproduced with permission from Müller, C.; Despras, G.; Lindhorst, T. K. Chem. Soc. Rev. 2016, 45 (11), 3275–3302. These small probes have been reviewed extensively and are well documented in the literature.

Individual glycan-protein interactions are typically characterised by a weak binding affinity. This weak interaction doesn't explain the multitude of biological roles that glycans play. Therefore, in 1995 Lee *et al* proposed the "Cluster Glycoside Effect".[44,45] This effect states that as the number of carbohydrate ligands increases, the binding interaction between the lectin and multivalent display of carbohydrates increases with a disproportional strength when compared to the individual monosaccharide. In order to exhibit a strong cluster glycoside effect, the lectin must have multiple glycan binding sites (or form multimers), and the glycan ligands must be in the correct orientation and spacing to efficiently bind (as well as be the correct glycan).[43]

Since this study, others have proposed the notion of primary, secondary and tertiary effects within the study of glycoclusters. The primary effect corresponding to the effect of the carbohydrate moiety, the effect of the linker and the nature of the scaffold. The secondary effects correspond to the orientation of the glycans on the scaffold, for example distance and orientation of 2 glycans on a protein scaffold. The tertiary effects are the dynamic and supramolecular effects that can occur on the cell surface, when in interaction with the multivalent probe. These could include receptor clustering or even receptor unmasking.[43]

1.1.3 Glycopolymers and materials based methods to study multivalent interactions

One area that has attracted much attention as a multivalent probe of glycan-binding proteins is in the creation of glycopolymers. In general, there are 2 synthetic approaches used to prepare precisely defined glycopolymers. One is the polymerisation of carbohydrate modified monomers (i.e. no deprotection is needed) and the other is the modification of functional polymers with carbohydrates (i.e. with a CuAAC reaction).[46] Synthesis methods for glycopolymers used in lectin assays have been reviewed in detail by Becer *et al*.[47] Many researchers including Becer[48] have synthesised precision glycopolymers and then subsequently

investigated the binding of this glycopolymer with a lectin. In one example, Haddleton *et el* was able to synthesise a glycopolymer comprised of di(ethylene glycol) ethyl ether acrylate and mannose. This polymer was subsequently shown to show a strong binding against the DC-SIGN lectin and was able to inhibit gp120-DC-SIGN binding. These glycopolymers could potentially have a therapeutic role in blocking viral adhesion to host receptors (anti-adhesion therapy).[48]

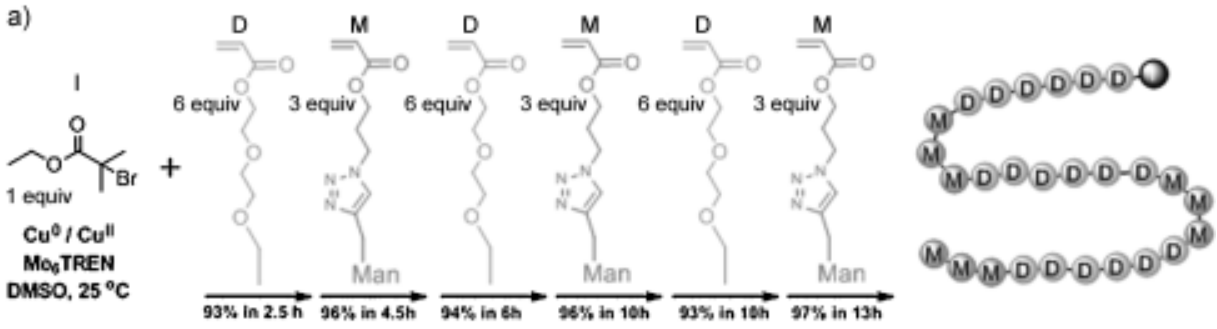


Figure 5. The synthesis of a sequence controlled di(ethylene glycol) ethyl ether acrylate –mannose hexablock copolymer. Reproduced with permission from Zhang, Q.; Collins, J.; Anastasaki, A.; Wallis, R.; Mitchell, D. A.; Becer, C. R.; Haddleton, D. M. *Angew. Chemie Int. Ed.* **2013**, *52* (16), 4435–4439.

Despite these advances, the polymer typically used within the field of glycobiology is still PAA.[49] While these glycopolymers typically show sufficient binding avidity to demonstrate bioactivity *in vitro,* the lack of control over the spatial orientation of the glycan motifs increases the potential for off-target effects. This is particularly important in the creation of new classes of more specific therapeutics. This has, in turn, necessitated the study of more controlled systems outlined previously and later in the discussion around the Siglec family of receptors.

1.2 Pentapeptide repeat proteins

In this project, a protein previously utilised within the Ben Davis group as a model for protein modification reactions was to be used as the multivalent scaffold. This protein, Np276, is a member of the pentapeptide repeat (PRP) unit family of proteins and was chosen as it shows a highly regular PRP coil – ideal for the creation of a multivalent scaffold.

The PRP unit family, which has more than 500 members [50] [51] across prokaryotic and eukaryotic kingdoms,[52] are a family of repeat proteins characterised by β-quadrilateral helix (a Rfr-fold).[53] These repeats are formed of consensus sequences of [S,T,A,V][D,N][L,F]-[S,T,R][G],[54] sometimes interrupted by non-canonical loops; a representative selection of these proteins are illustrated in Figure 6.[52] As each face is composed of 5 amino acids, the central amino acid is labelled *i*, the previous two i^{-1}, i^{+2} and the following two i^{+1}, i^{+2} (from N->C). The side chains *i* and i^{-2} are often interior facing hydrophobic residues whilst the other 3 (i^{-1}, i^{+1}, i^{+2}) are exterior facing. In total there are 4 pentapeptide faces making a single turn, this corresponds to 20 amino acids per turn, in the regular coils. Furthermore, the N and C-termini of PRPs are at opposite ends of the extended coil with many exhibiting a C-terminal dimerisation zone (α helix between 2 β-strands).[55] PRPs that target topoisomerases, and thus mimic DNA (PRPs are also negatively charged, which can also help prevent aggregation and also mimics DNA), all exhibit this dimer structure.[56]

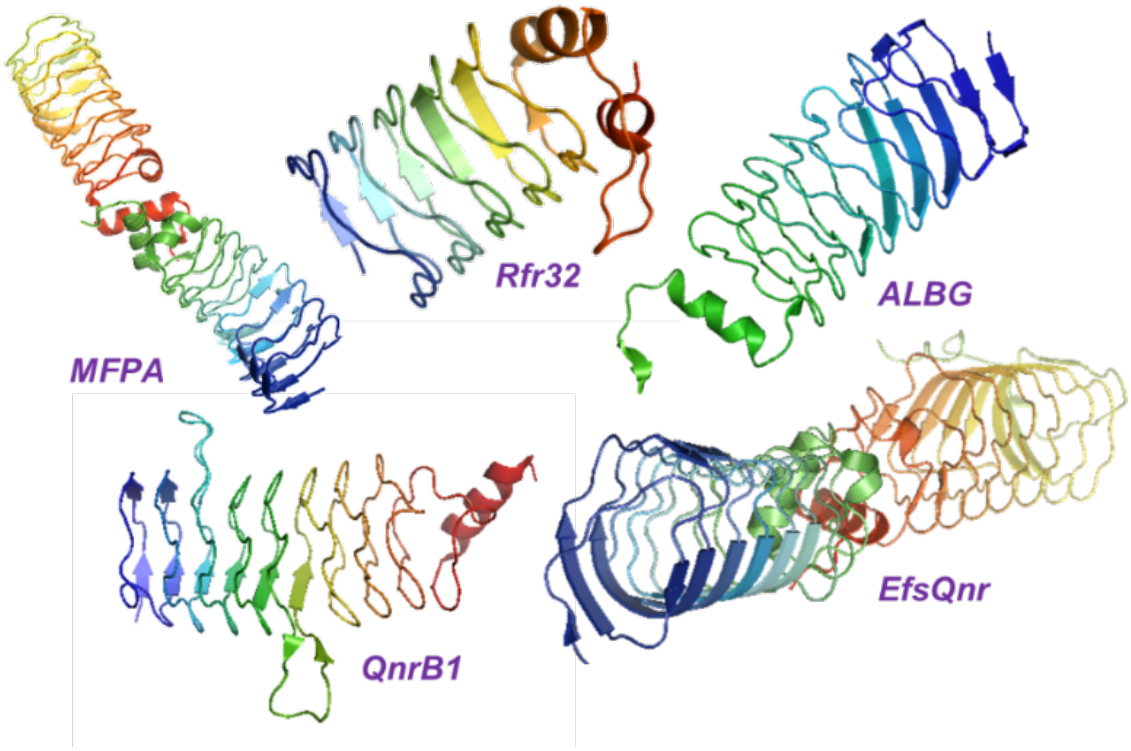


Figure 6. Structures of some PRP proteins. In all cases the β-quadrilateral helix that characterises these proteins can be clearly seen.[57–62]

1.2.1 The biological function of PRPs

Whilst the biological functions of these proteins are generally unknown,[63,64] for the nine Qnr proteins (QnrA1, QnrB1, QnrC, QnrD1, QnrS1 etc.) and others including AhQnr, EfsQnr, McbG, AlbG, and MfpA fluoro-quinolone resistance via DNA gyrase or topoisomerase inhibition has been reported.

DNA gyrase is a type II DNA topoisomerase,[65,66] consisting of GyrA and GyrB which together form an active A_2B_2 tetramer.[55] The mechanism for its action and binding with DNA and PRPs is shown in Figure 7. Aside from the mechanism shown below, other studies have reported that the loops on the Qnr proteins help to destabilise the cleavage complex, something supported by an observation that Mfpa functions with a second protein MfpB (which essentially functions

as the non-canonical loops). [67] Due to these interesting properties, PRPs have begun to attract increasing attention from the biochemical community.

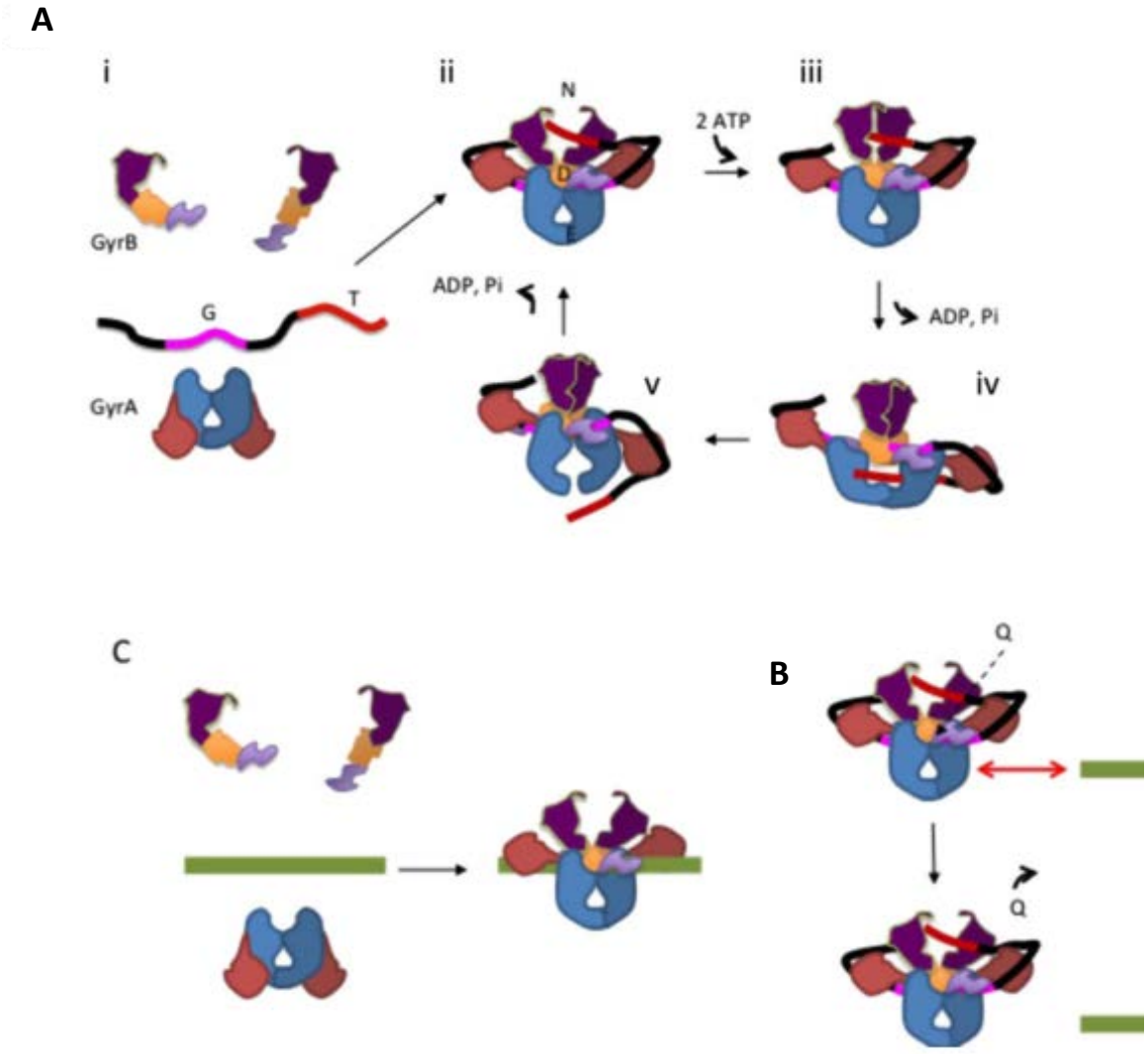


Figure 7. Adapted with permission from Heddle *et al.*[55] A). The mode of action of DNA gyrase. The GyrA and GyrB dimers assemble on the DNA, with the G segment binding to the active site of the DNA gate (B). The protein-DNA complex is formed, GyrB N-terminal regions dimerise and capture the T-segment upon ATP binding. The G segment is cleaved and then transported to the exit gate of the enzyme. C). The PRP (green rod) competes with DNA for binding to the G-segment. D). Fluoro-quinolones bind to the DNA-gyrase complex at the G-segment cleavage site. The PRP can recognise this conformation, bind to it, destabilise it and cause loss of the quinolone binding site thus restoring enzymatic activity.

1.2.2 Other proteins with a similar structure

The Antarctic bacterium, *Marinomonas primoryensis*, produces a AFP (antifreeze protein) MpAFP[68,69], with a molecular mass of ~34kDa.[70] This protein folds in a calcium-dependant fashion into a β- helix, the coils being spaced into approximately 4.6Å subunits (Np276 is 4.8Å). Whilst the PRP proteins haven't yet been implicated in antifreeze activity, the structural similarities raise the question whether these proteins also confer physical protection to their organisms instead of just antibiotic resistance.

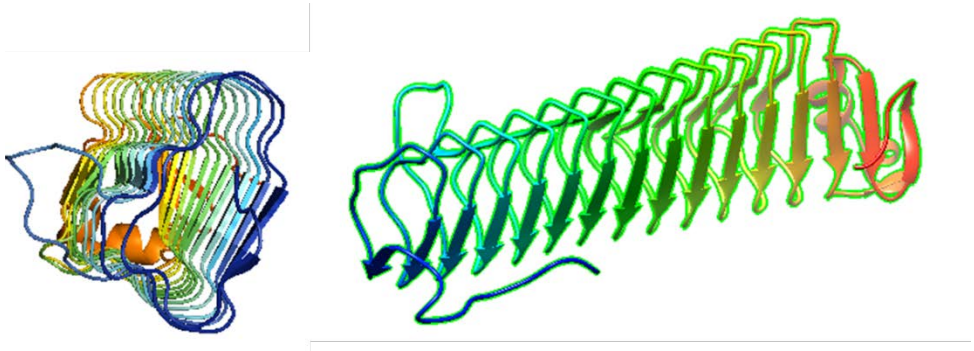


Figure 8. X-ray crystallographic MpAFP protein structure, the protein coils are spaced 4.6Å apart. Unlike PRPs this protein also binds a calcium ion in its coils.

Other repeat proteins of interest include the ankyrin proteins (Figure 9), which exhibit exceptional thermal stabilities and high expression levels (60-80°C T_M /200mgL^{-1} expression yields). Other proteins characterised as repeats include tetratricopeptide (TPR), leucine-rich (LRR), HEAT and armadillo repeats.[71]

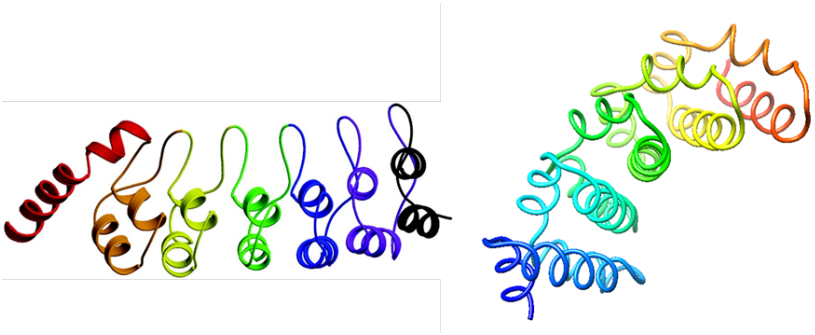


Figure 9. Structure of a model ankyrin protein (left) and a designed armadillo repeat protein (DArmRP) (4DB6). These protein types have also been used to create multivalent scaffolds as outlined in section 1.1.1.

1.2.3 Np276

Npun 275/276, an ancestral PRP protein from *Nostoc Punctiforme*, nitrogen-fixing cyanobacteria, was formed by the fusion of two adjacently encoded proteins *Np275* (98 AA) and *Np276* (75 AA). Cyanobacteria also appear to have a uniquely large amount of PRP containing proteins encoded in their genome; *Nostoc Punctiforme* having 40 such proteins.[72]

Like other pentapeptide repeat unit proteins, this protein is characterised by the regular pentapeptide β-quadrilateral helix but unlike other proteins of this class doesn't contain any non-canonical loops[52]. Furthermore, unlike other members of the family, it doesn't contain a C-terminal dimerisation domain; suggesting that this ancestral protein theoretically would exist as a monomer in solution, making Np276 an ideal model for a multivalent scaffold.

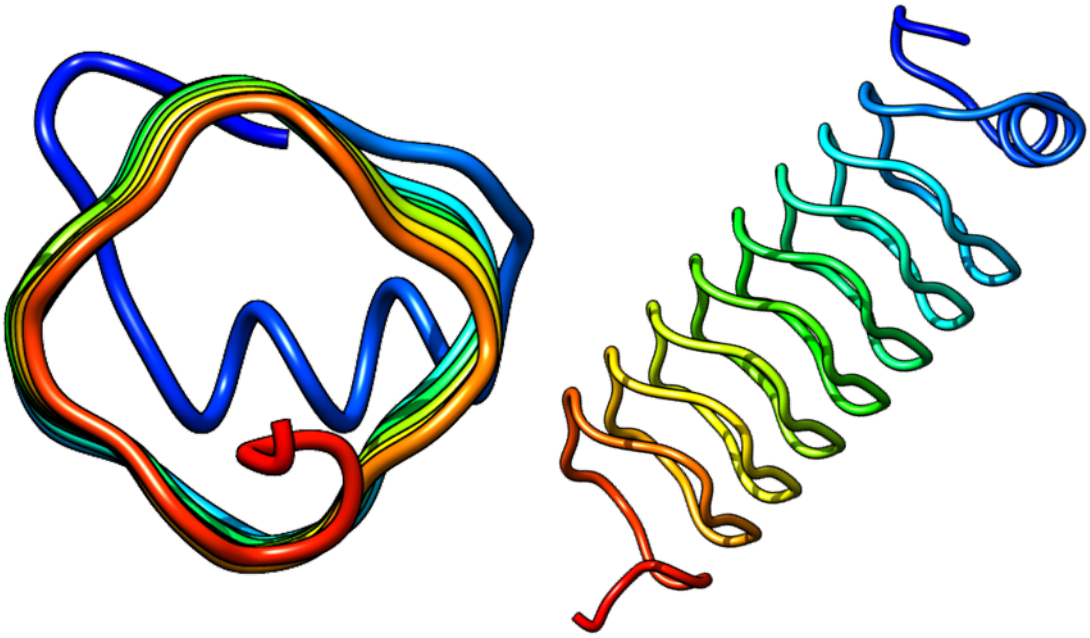


Figure 10. Structures of Np276 (PDB 2j8K) and sequence of a mutant. The protein contains an N-terminal helix (blue) but is open at the C-termini, indicating that it shouldn't dimerise. The Np276 helices are spaced 4.8Å apart from each other.

1.3 Unnatural amino acids in chemical biology

Chemical biology is a field in which the tools, methods and techniques of chemistry are applied directly to biology & medicine. One important area in the field is the chemical modification of complex biological macromolecules in selective or specific ways, commonly referred to as bio-orthogonal chemistry. Chemical methods for post-translational (after expression) protein modification are described later on in this chapter; however, many of these methods rely on the incorporation of unnatural amino acids that can act as reactive handles for modification. An approach pioneered by Davis *et al*, the "Tag and Modify" strategy, sought to move protein modification away from an unselective to a selective/specific strategy, in which proteins could be modified in a homogenous way and then used to interrogate biological function. Central to this approach is the incorporation of unnatural amino acids.[73,74]

For the remainder of this review, canonical amino acids (CAA) can be defined as naturally occurring proteinogenic L-α-amino acids. Amino acids such as selenocysteine, formyl-methionine and pyrrolysine are all proteinogenic but special canonical amino acids. Canonical natural amino acids are all characterised by a 3-letter code and a corresponding single letter code designation. Unnatural, non-canonical amino acids (NCAA) (not found in nature) are also characterised by a 3-letter code such as Aha (or AHA) for azidohomoalanine.[75]

There are 2 main approaches for the incorporation of unnatural amino acids: a residue specific approach (in which a global replacement of a CAA amino acid with an unnatural amino acid is observed) and a site specific approach (which can be broken down into stop codon suppression and non-triplet coding pairs). Within this project, the focus will predominately be on the residue specific incorporation of UAAs as reactive tags, and as such the majority of this review is dedicated to this incorporation methodology.

1.3.1 Protein synthesis

Protein synthesis requires 2 crucial steps: transcription and translation. The rate of translation and transcription are roughly equal in bacteria, with the whole process (protein expression) consuming as much as 30-50% of all energy in growth stage bacteria. After DNA is transcribed to mRNA, this new polynucleotide is captured by a ribosome and translated to a polypeptide. The information flowing in these processes from DNA via mRNA to a functional/structural unit, is the central dogma of molecular biology (Figure 11).[76]

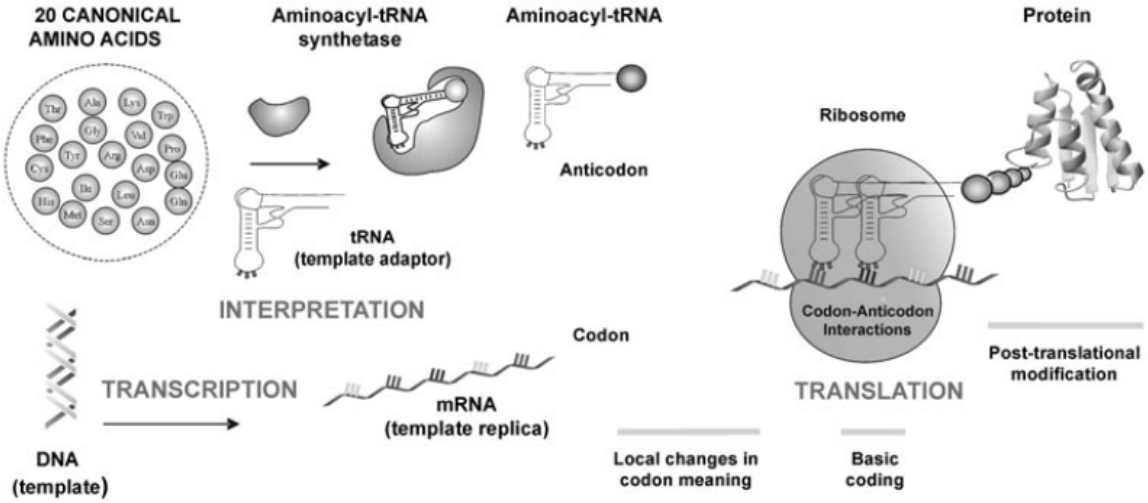


Figure 11. Protein expression begins with transcription. In transcription: RNA polymerase, ribonucleotides and transcription factors lead to the formation of a mRNA template. Concurrently, tRNA is charged with amino acids and then in the ribosome the mRNA and tRNA interact. Following an initiation event, elongation and termination the newly formed polypeptide chain (formed by a peptidyl transferase) is released from the ribosome. This newly formed protein can then be modified in a post-translational manner. Reproduced with permission from Angewandte Chemie International Edition, 2004, 43: 6426–6463.

The translation process is controlled by 2 molecular recognition processes: codon/anti-codon interactions between mRNA and tRNA, and site selection of amino acids by aminoacyl-tRNA synthetases (AARS).

1.3.2 AARS and codon/anti-codon interactions

The tRNA amino acylation reaction is conducted by a family of 20 cellular cytosolic (in prokaryotes) enzymes called AARS. This reaction occurs in a 2 step fashion in the active site of the enzyme, in the presence of Mg^{2+} and ATP. [77–80]

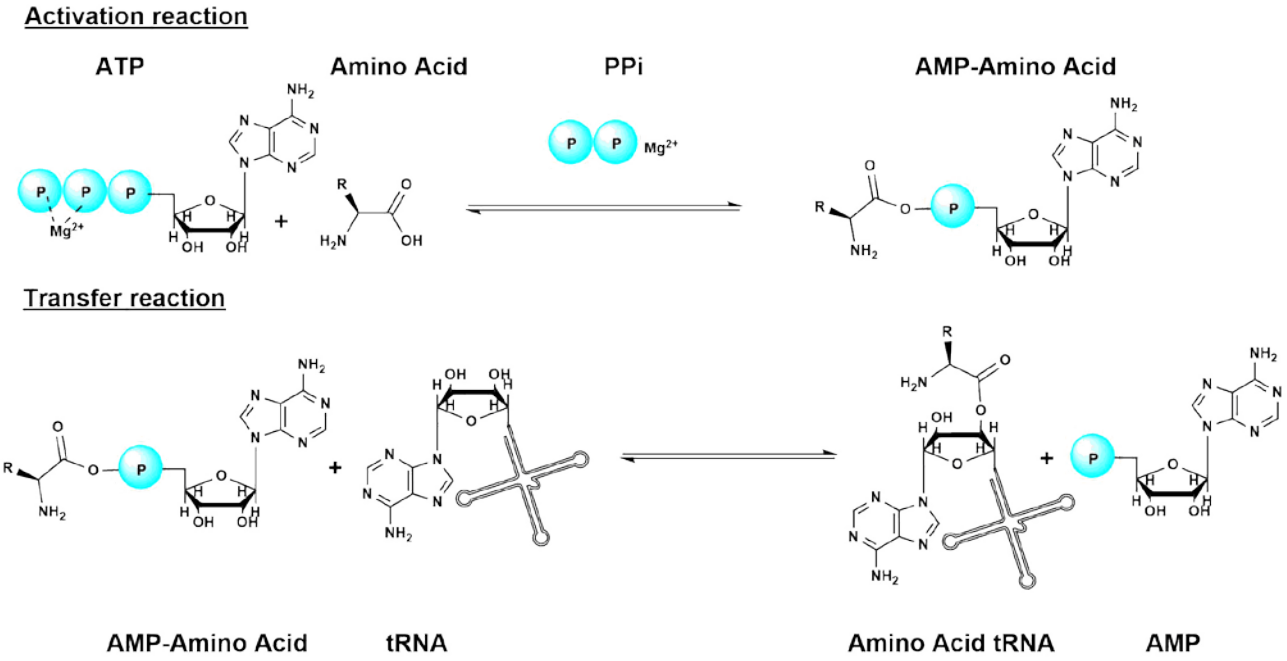


Figure 12. Chemical reactions involved in tRNA acylation. (Top) the activation reaction to the form the AMP-amino acid.(Bottom) The transfer reaction which charges the amino acid on the tRNA.

In the ribosome, there are 2 sites (A and P) where elongation occurs (between tRNA bound in the A sites and P sites). The first interaction (mRNA-tRNA hydrogen bonding) is a loose interaction (kinetic proofreading), the second interaction is the most important and the third interaction has some flexibility (wobble position). The error frequency in protein translation is about 0.03% due to steps that help to prevent mistranslation (amino acid selection, tRNA charging and the ribosomal synthesis). Fortunately, there is some substrate flexibility for the incorporation of structural analogues (e.g. residue specific), whereas for the incorporation of more unusual substrates an engineered AARS/tRNA system can be used (site specific).[76]

1.3.3 Residue specific incorporation

Residue specific incorporation of an NCAA is a method that leads to the global or partial (at every residue) replacement of a specific canonical amino acid residue with an unnatural analogue. Residue specific incorporation usually involves the use of auxotrophic bacterial strains, along with a media switch step in which the unnatural amino acid is added to the media (selective pressure incorporation). To date, numerous amino acids including fluorinated Met derivatives, Hpg (alkyne surrogate of Met), Aha (azide surrogate of Met), SeMet, halogenated and substituted tryptophans and substituted phenylalanines have been incorporated using this methodology.[75]

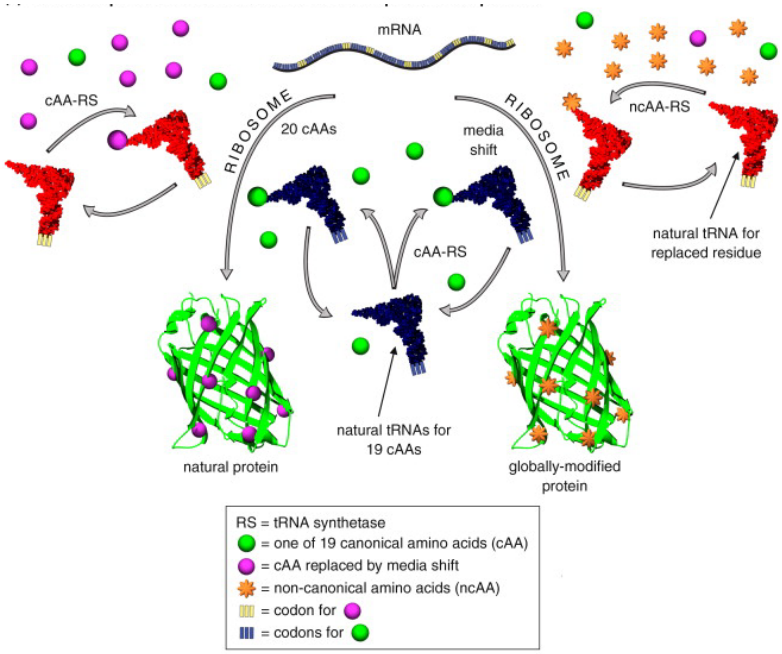


Figure 13. Overview of the residue specific approach showing how an UAA replaces an AA. Reproduced with permission from Current Opinion in Chemical Biology, 2010, 774-780.

The biochemical community first began investigating NCAAs in the early 1950s with the work of Richmond *et al*,[81] which was subsequently built upon by Cohen *et al*,[82] to incorporate SeMet (in place of Met) into proteins in a residue specific fashion.[83] Cohen used auxotrophic strains with close control of fermentation conditions, in order to affect a global incorporation of the

methionine analogue.[83] Since the 90's SeMet incorporation has become increasingly popular as a tool in protein X-ray crystallography (phasing). In order to achieve the successful incorporation of a non-canonical amino acid (NCAA) into a protein via a residue specific method, it is important that there is: efficient cellular uptake of the NCAA, high intracellular concentrations of the NCAA (for acylation/activation), chemical stability of the NCAA, a high rate of tRNA acylation and finally efficient translation into the polypeptide chain.[76,84,85]

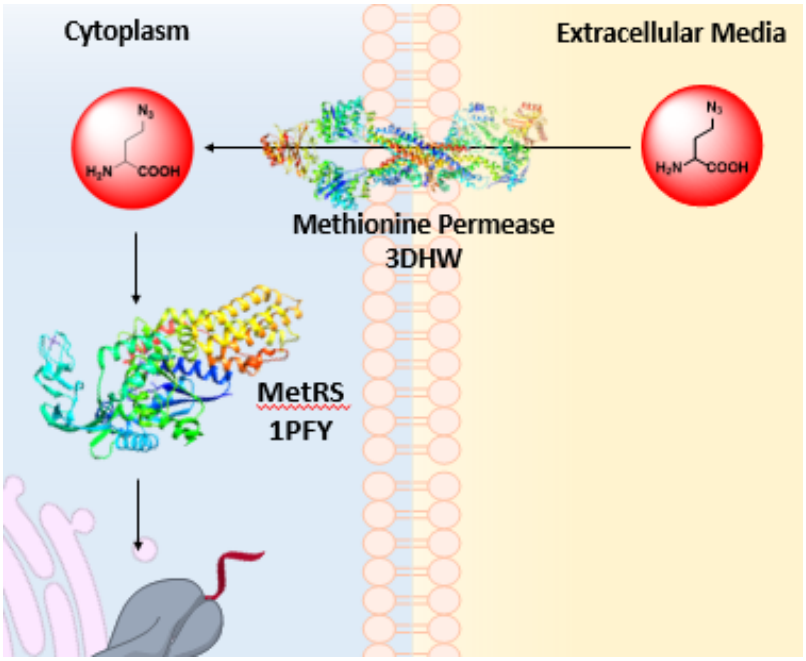


Figure 14. Residue specific incorporation of unnatural amino acids into *E. coli* expressed recombinant proteins relies upon both the effective uptake and also aminoacylation of the amino acid homologue. These two processes are governed with Met by a methionine tRNA synthetase (MetRS) (an AARS) and a membrane associated methionine permease (3DHW).

1.3.4 Applications of residue specific incorporation

1.3.4.1 FUNCAT and BONCAT

BONCAT (bio-orthogonal non-canonical amino acid tagging) and FUNCAT (fluorescent non-canonical amino acid tagging) are two functional applications for residue specific incorporation. Both methodologies rely on the global residue specific incorporation of a reactive handle (normally alkyne or azide: Aha or Hpg). In BONCAT these can then be modified by an enrichment tag; whilst, in FUNCAT these can be modified with a fluorophore.[86] These methodologies both allow for the proteome wide evaluation of protein expression in response to external stimuli.[87]

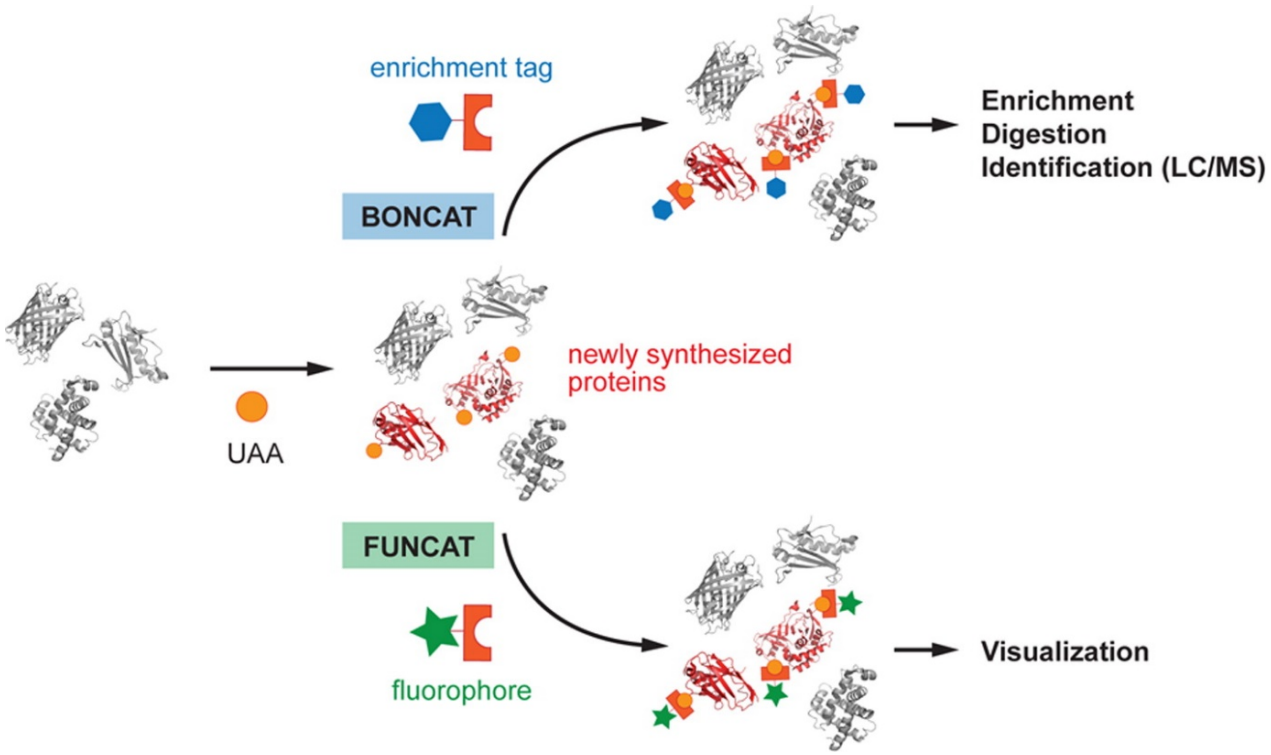


Figure 15. A comparison of the BONCAT and FUNCAT methodologies. An NCAA (orange) was incorporated in a residue specific manner into all proteins undergoing expression after addition of the NCAA. These newly synthesised proteins can then be labelled with an enrichment tag for identification or a fluorophore for microscopy. Reprinted (adapted) with permission from (Chem. Rev., 2014, 114 (9), pp 4764–4806). Copyright (2014) American Chemical Society.

BONCAT was first reported in 2006 by Tirrell *et al*. In this initial report, HEK293 cells were labelled with Aha before being reacted with a biotin containing alkyne probe.[88] Following an avidin purification and subsequent treatment with trypsin, the peptides were analysed by tandem mass spectrometry. 195 proteins were identified in this 2-hour labelling window, thus demonstrating the power of this method. Following this initial report, the method has been used in the labelling of *E. coli* lysates,[89] analysis of the secretome,[89] analysis of expression changes during T-cell activation (QuanCAT), [90] analysis of the hippocampal neuropil proteome[91] and in the analysis of protein expression in zebrafish larvae.[92]

FUNCAT, which involves the tagging of the newly synthesised proteins with a fluorophore before imaging via gel or microscopy, has been used in several applications in both mammalian and bacterial cells. Furthermore, by pulse labelling (add one NCAA for an allotted time, stop and then add another NCAA) cells with different amino acids (Hpg/Aha) dual labelling and visualisation can be conducted.[93,94]

1.3.4.2 Blue fluorescent proteins

In a study to generate inherently blue fluorescent proteins, human annexin A5 (anxA5) was labelled using tryptophan autotrophs and the NCAAs (4-Aza)Trp and (5-Aza)Trp.[76,95,96] The incorporation of these NCAAs resulted in proteins with greatly increased spectral qualities versus the wildtype (CAA) protein.[97] With (4-Aza)Trp a Stokes shift of approximately 130nm was observed. Furthermore, by pre-incubating L-serine with tryptophan synthase and the appropriate azaindole, the required (Aza)Trp amino acid could be generated *in situ* before being added into the expression culture. This approach eliminates the problems of synthesising the tryptophan analogues.[98] Other work has also been conducted by Deo *et al* on the residue specific incorporation of 3-amino-L-tyrosine or 3-fluoro-L-tyrosine into the fluorescent protein DsRed. In this case altered fluorescent properties were also observed.[99]

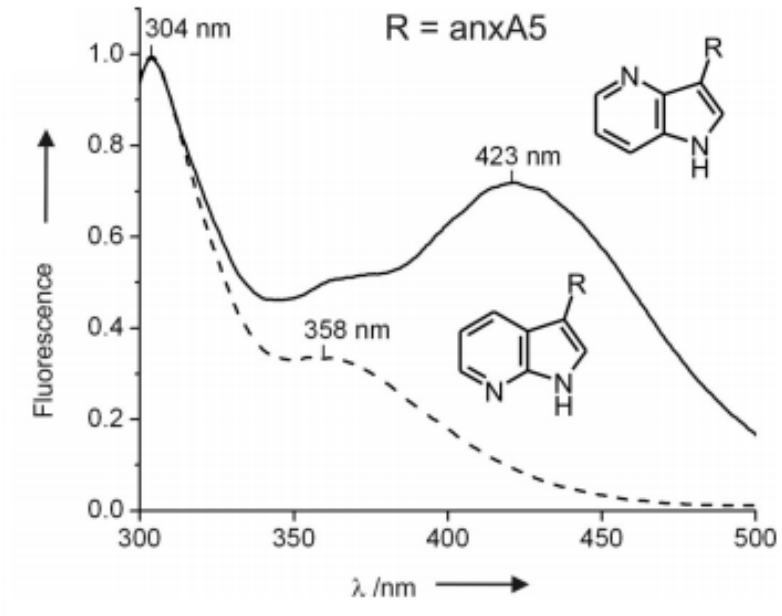


Figure 16. Spectral (emission profiles) properties for two different aza-tryptophan isomers incorporated in anxA5. The emission maximum of (4-Aza)Trp-anxA5 (solid line) is 65nm red-shifted compared to (7-aza)Trp-anxA5 (dashed line). The spectral shoulder of (4-aza)Trp-anxA5 (solid) is due (4-Aza) Trp residue being partially solvent exposed. Upon a calcium denaturation of the protein, only one peak is observed. Diagram replicated with permission from National Academy of Sciences © 2008.

1.3.4.3 Gold fluorescent proteins

GFP (a green fluorescent protein from *Aequoria victoria*) is a common protein used in cellular assays and molecular biology. GFP exhibits a strong fluorescence due to the presence of a (4-(*p*-hydroxybenzylidene)imidazolidin-5-one) chromophore that is formed post-translational between residues S65-G67. An aromatic residue at position 66 is also essential to fluorescence in GFP. By using an auxotrophic tryptophan *E. coli* strain along with selective pressure, the relaxed substrate specificity of the AARS can be used to incorporate non-natural tryptophan residues into GFP. ECFP (enhanced cyan fluorescent protein) a protein formed by a Trp57 and a Trp66 mutation. When incorporated with 4-amino-tryptophan showed a large Stokes shift, resulting in a new class of "gold fluorescent protein GdFP). Incorporation into EGFP (enhanced GFP) which only contains a Trp57, resulted in no spectral changes indicating that a Trp66 is essential to see altered spectral properties.[100]

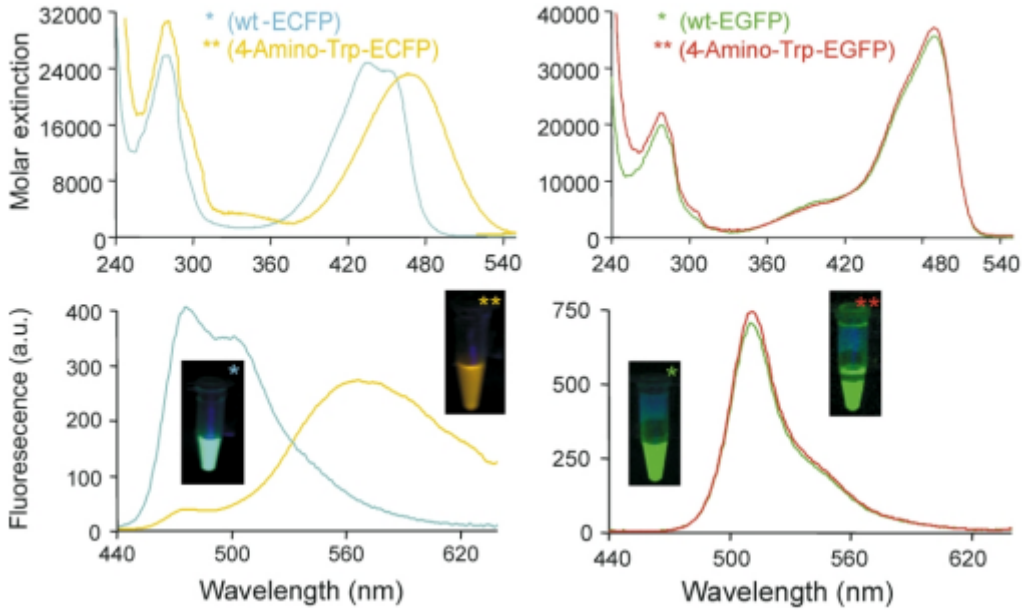


Figure 17. Spectral properties of ECFP, EGFP and the substituted variants. 4-amino-ECFP is also known as GdFP. The GdFP exhibits the different spectral profile due to the presence of both electron donating (amino group) and electron accepting groups. [100] Diagram replicated with permission from Journal of Molecular Biology, 2003, 1071-1081.

1.3.4.4 Applications in materials science

There are also potential future applications for the residue specific incorporation of NCAAs in material science. One area of particular interest is in the creation of conducting biomaterials, through incorporation of an electroactive amino acid. By crosslinking these systems, an extended conjugated system based on β-sheets could then be created. Electroactive amino acids are those that contain sulfur or selenium heteroatoms (with an unshared pair of electrons in a p orbital & vacant d orbitals) that can be conjugated to aromatic systems. Kothakota *et al* incorporated thienyl-alanine (via a Phe auxotrophic strain) into encoded repeats to create potentially new conducting materials. Despite showing incorporation, no electrochemical applications were shown.[101]

Fluorous amino acids have also been used to create "Teflon like" lamellar surfaces as determined by wide angle X-ray scattering. Through the residue specific incorporation of *p*-fluorophenylalanine, a global replacement of 98% was observed.[102,103]

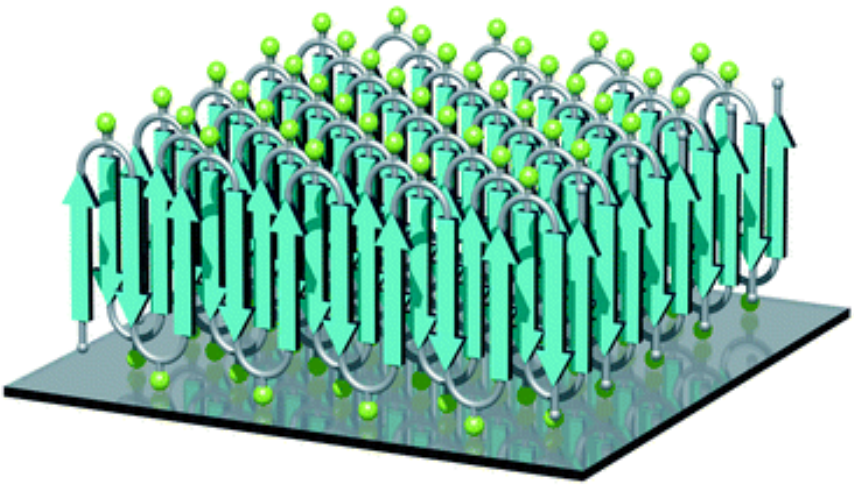


Figure 18. A lamellar phase formed by $((AG)_3FG)$ (F is the NCAA). The green spots in the diagram represent fluorine. Diagram reproduced with permission from the Royal Society of Chemistry, Chemical Communications, 2001.

1.3.4.5 Fluorinated amino acids as probes

Fluorinated aromatic amino acids have also been used to alter the spectral properties of proteins. One approach has been via the residue specific incorporation of fluorinated phenylalanine residues (*o*-fluorophenylalanine, *m*-fluorophenylalanine and *p*-fluorophenylalanine) into azurin and human annexin V. This creates a unique trace of 2 peaks (260nm/270nm) in the UV spectrum, known as fingers. These don't overlap with the residual UV spectrum of tyrosine and tryptophan residues in proteins providing a simple way of detecting modified proteins.[104]

Additional work has also looked at the incorporation of fluorinated tryptophan analogues into proteins in order to alter their fluorescence profiles. Of particular note is the 5-F(Trp) NCAA, which when incorporated in annexin-V produces a 14 nm red-shifted emission maxima. Conversely, 4-F(Trp) results in a loss of Trp fluorescence and instead highlights the (previously masked) contributions from both Phe and Tyr residues. 4-F(Trp) is therefore known as a silent fluorophore.[105]

^{19}F NMR on proteins is also a powerful method to interrogate protein conformational structure.[106] ^{19}F NMR (along with dynamic NMR) on GFP with fluorinated tryptophan residues allowed for distinguishing of two different conformational states in this protein.[107] Budisa *et al* have also looked at the incorporation of trifluoromethionine and difluoromethionine (low incorporation rates were observed) into proteins such as phage lysozyme and GFP, this was in order to conduct sensitive ^{19}F NMR measurements.[108] Fluorinated proline derivatives have also been used to study the rates of cis/trans proline isomerization in proteins and simple peptides.[109–111]

1.3.4.6 Fluorous proteins

Whilst a methyl to trifluoromethyl substitution causes a moderate increase in steric bulk (12Å), the most pronounced effect comes from the fluorous effect. The fluorous effect refers to the observation that perfluorinated compounds will preferentially interact with other perfluorinated groups. This makes fluorinated biomolecules resistant to solvent-mediated denaturation and stabilises the protein around a fluorous core. [112,113]

A common way of altering protein structure or stability, is to incorporate (in a global residue specific manner) polyfluorinated amino acids into the protein backbone. Common amino acids that have been targeted for replacement in this approach include: Leu, Val, Met and Ile. When Leu zipper proteins were incorporated with 5,5,5-trifluoroleucine or 5,5,5,5',5',5'-hexafluoroleucine by Tirrell *et al*, a T_M increase of 10°C and 20°C was observed respectively. An increased resistance to urea denaturation is also observed. These observations can again be ascribed to the fluorous effect .[114]

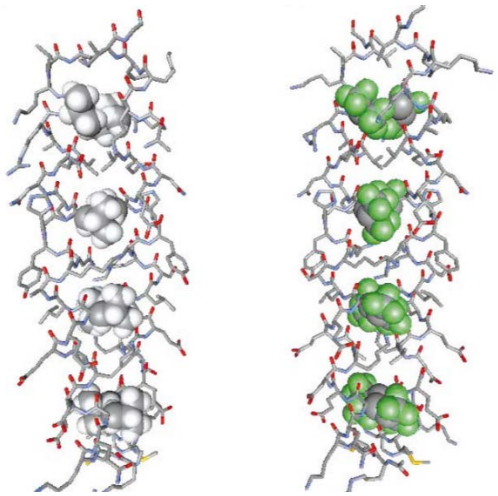


Figure 19. A natural leu zipper protein (left) and a leu zipper protein incorporated with perfluorinated amino acids (left). The fluorous core (green atoms) can be seen clearly. Figure reproduced with permission from Elsevier, Chemical and Biology, 2000, 153-157.

1.3.5 Site specific methods

1.3.5.1 Stop codon suppression

Expansion of the genetic code in an orthogonal manner to the proteinogenic amino acids is possible through non-sense suppression that utilises orthogonal tRNA systems. In this approach engineered AARS from another organism, (that is translationally orthogonal) is used to acylate the non-sense tRNA thus expanding the genetic code in a site specific (highly controlled) way.

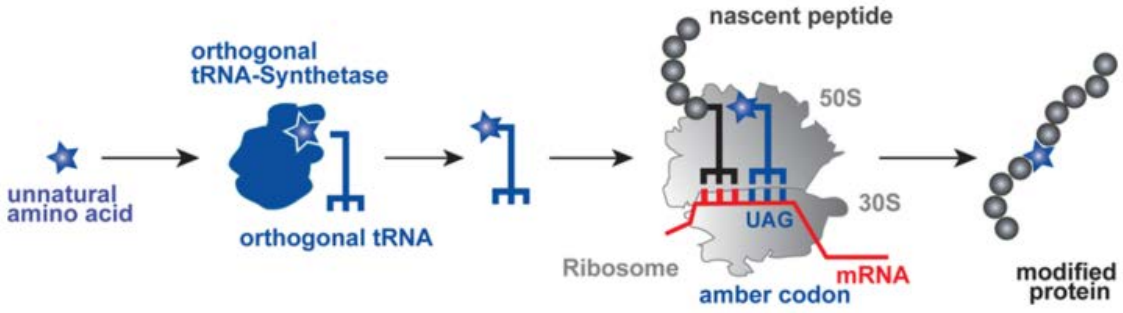


Figure 20. A schematic example showing the incorporation of an unnatural amino acid using an orthogonal tRNA/tRNA-synthetase pair. This approach allows for the site specific incorporation of unnatural amino acids utilising a stop codon (in this example in response to the amber non-sense codon). Reprinted (adapted) with permission from (Chem. Rev., 2014, 114 (9), pp 4764–4806). Copyright (2014) American Chemical Society

Stop codons which have been utilised in this approach include the amber (TAG), opal (TGA) and ochre (TAA) stop codons. Of these, the amber stop codon (TAG) is the most widely used as it is the rarest of all the *E. coli* codons. This methodology has also been extended into numerous other organisms such as yeast, zebrafish and CHO cells.[115]

1.3.5.2 *M. jannaschii* TyrRS/tRNA and other TyrRS systems

One of the most widely used orthogonal pairings used in *E. coli* expression is the *Methanococcus jannaschii* TyrRS/tRNA pair from *M. jannaschii*.[116] This pairing was first used to incorporate *p*-methoxyphenylalanine; however, since this first report rapid screening methods have been developed which have dramatically expanded the scope of NCAAs substrates that can be incorporated.[117] In general, most NCAAs incorporated via this system contain a *para*-functionalised phenyl ring with: allyl,[118] alkyne,[119] azides,[120] halides,[121] tetrazines,[122] and ketones[123] all having been genetically incorporated *in vivo*. By mutating the TyrRS enzyme the substrate scope can be relaxed further allowing the incorporation of meta-substituted phenyl rings and heteroaromatics.[124,125] Whilst the *M. jannaschii* TyrRS/tRNA pair is good for UAA incorporation into bacterial proteins, for incorporation into high order organisms other systems are required. For this purpose *E. coli* TyrRS/tRNA and E. coli TyrRS/Bs-tRNA. pairs can be used for incorporations into eukaryotes. [115]

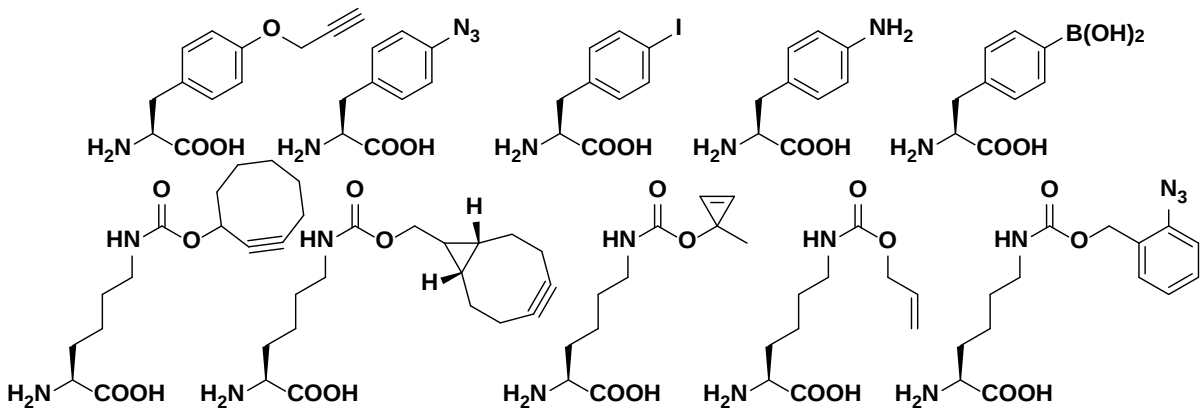


Figure 21. NCAA that have been incorporated in a site specific manner utilising non-sense suppression methods. These amino acid handles can be used as tags for metal mediated couplings, strain promoted reactions and other conjugations. The amino acids on the top line are incorporated via a TyrRS/tRNA pair, whereas those on the bottom can be incorporated via a PylRS/tRNA pair.

1.3.5.3 *E. coli* LeuRS/ tRNA

The *E. coli* LeuRS/ tRNA is an orthogonal system for the incorporation of NCAAs into higher organisms, in particular yeast. Using this pair, NCAAs containing a photo-caged aged cysteine and ferrocene rings have been incorporated *in vivo*. In general, this pairing has a relaxed substrate specificity but shows a preference to cysteine, alkyl or methionine based analogues.[126–128]

1.3.5.4 PylRS/tRNA

PylRS/tRNA is a naturally occurring AARS/tRNA pair found in archaebacteria (*M. mazei* and *M. barkeri*), where it was found that a TAG codon encoded for a 22nd (special) canonical amino acid: pyrrolysine.[129] The PylRS/tRNA is orthogonal in both *E. coli* and eukaryotic cells, and is able to incorporate a range of pyrrolysine derivatives (functionalised via an ε-amino-carbamate modified lysine). Examples of amino acids incorporated using this AARS/tRNA pairing include: Boc-protected lysine,[130] alkynes and the naturally occurring PTMs (mono and di-methyl-lysines).[131] Many PylRS mutants (both *M. mazei* (Mm) and *M. barkeri* (Mb) have subsequently been prepared and used for the incorporation of increasingly more complex NCAAs bearing functionalities such as: Se-alkylselenocysteines,[132] *N*-ε-benzyloxycarbonyl-L-lysine,[133] alloc-lys,[134] norbornene,[135] trans-cyclooctene,[136] *meta*-substituted phenylalanine,[137] unnatural histidines,[138] and spin labels.[139] This increasingly broad substrate scope along with the ability to directly translate the system from *E. coli* to eukaryotic cells makes this PylRS/tRNA increasingly powerful. Figure 22 highlights the cross organism orthogonality of the various AARS/tRNA pairs.

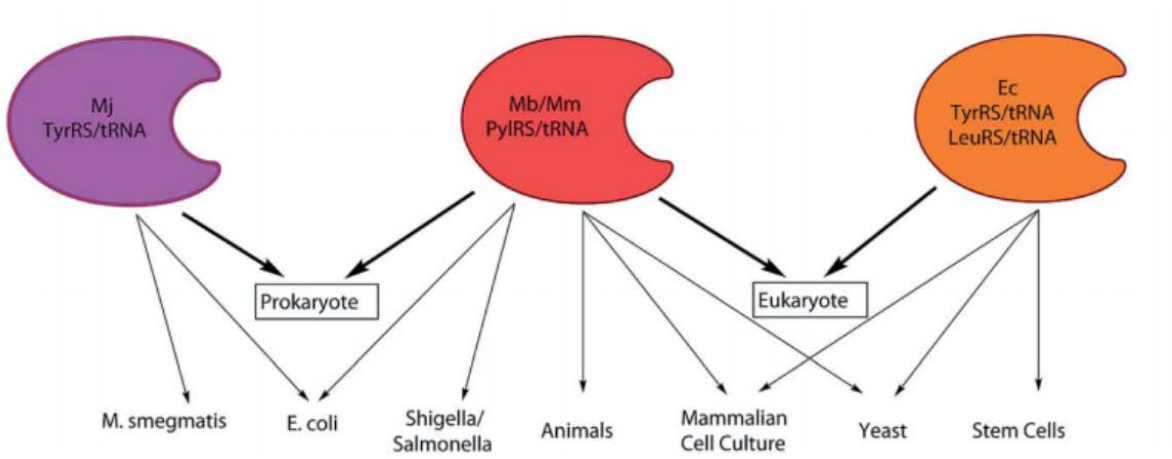


Figure 22. A schematic example showing the incorporation of an unnatural amino acid using an orthogonal tRNA/tRNA-synthetase pair. This approach allows for the site specific incorporation of unnatural amino acids utilising a stop codon. Reproduced with permission from Chem. Sci., 2015,6, 50-69.

1.3.5.5 Quadruplet coding

Quadruplet coding was first developed by Schultz *et al* using LysRS/tRNA in response to an AGGA quadruplet code.[140] Chin *et al* have further expanded this methodology, by engineering a new orthogonal translational pathway including the creation of orthogonal ribosomal machinery. This quadruplet coding potentially opens up 256 new blank codons that could be used for the incorporation of new UAAs. The use of quadruplet coding has expanded the scope of site specific protein modification, allowing for the incorporation of multiple orthogonal reactive handles. This was used by Chin *et al* to install a FRET pair on a protein, thus demonstrating the potential of this method to chemical biologists.[141]

1.3.6 Multiple NCAA incorporation

The incorporation of multiple NCAA has been realised by combining numerous methodologies into a single experiment. An initial approach towards a dual site specific protein involved the use of a TyrRS/tRNA (*para*-azidophenylalanine) with a mutated (ochre stop codon specific) PylRS/tRNA pair (alkyne) to create after a conjugation reaction, a FRET pair.[141] As previously described there has also been work investigating the use of orthogonal quadruplet coding to facilitate dual (and higher) labelling of proteins with NCAAs. Budisa and others have also developed polyauxotrophic bacterial strains for the multiple residue specific incorporation of NCAAs.[142,143] Preliminary reports have also looked at combining this traditional residue specific methodology with the site specific non-sense suppression; however, this latter approach is still in its infancy.[144]

1.4 An introduction to bioconjugation chemistry

The covalent modification of a bio-molecule (e.g. protein, cell membrane glycolipid, DNA strand), using reagents and techniques from chemistry, is commonly referred to as bio-conjugation. The goal of this method is to enhance or augment the properties of the biomolecule relative to the unmodified structure by covalently modifying a biomolecule with an additional reagent.[73,145,146] Whilst considerable work has been undertaken on the modification of DNA/RNA, viral capsids and other biomolecules[‡] *in vitro*, as well as the modification of these biomolecules *in vivo* (mammalian and *E. coli*),[147] this thesis shall only look at the field of protein bio-conjugation. Chemical biologists may seek many reasons to modify proteins; these include: to improve the stability of proteins, to endow them with favourable biophysical properties, to study proteins in a biological environment, to immobilise them, to increase their therapeutic potential or to use them as new protein-based materials.

There currently exist numerous methods for the modification of both natural proteins and proteins containing unnatural, orthogonal and amino acid 'tags'. Methods for the incorporation of UAAs are described in the previous section of this introduction, bio-conjugation methods utilising these unnatural amino acids can be grouped under the 'tag and modify' strategy of protein modification.[73] The tag and modify approach to the chemical post-translational modification of proteins relies upon the site specific incorporation[16] of reactive natural or unnatural chemical tags;[17] which can then, through the use of highly selective or specific chemical reactions[18], be discretely modified to generate complex unnatural proteins (Figure 23).

[‡] Many of the methods used to modify biomolecules have also been used in polymer and macromolecular chemistry.

Conversely, more traditional methods for protein modification often rely on the empirical determination of parameters that would, by exploiting the subtle nuances in protein structure, allow potentially selective reactions at one or more residues in comparison to other endogenous amino acids.

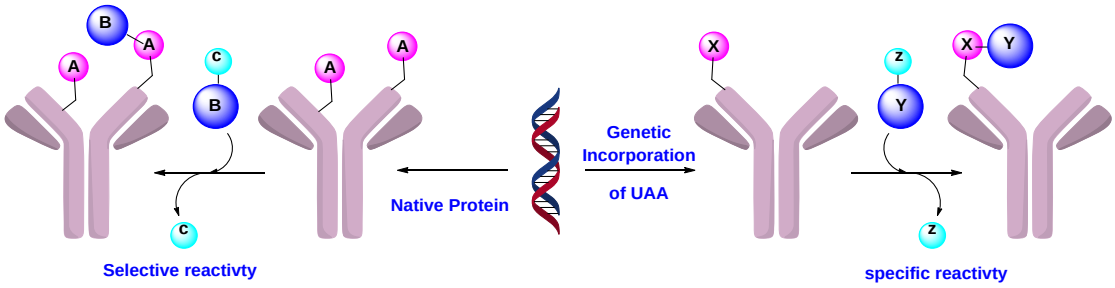


Figure 23. Modification of a hypothetical Antibody with site specific (UAA) and site selective reagents. In this example the sphere labelled X represents a UAA and A an natural amino acid. The species Y-Z/B-c is a site specific or site selective reagent.

The challenges in bio-conjugation are twofold: firstly the reactions must be incredibly specific (or selective) and secondly the reactions must be conducted under metabolic conditions (or thereabouts).[21] In this sense, the process of optimising protein bio-conjugations is maybe opposite to traditional small molecule methodology. In the latter, there is no perceived limit on conditions whereas in protein bioconjugations the chemistry is dictated by biocompatibility (pH 6-8, temperature < 37°C and aqueous solvents; perhaps analogous to "green chemistry").[146,148,149]

1.4.1 Reactions at natural residues

1.4.1.1 Cysteine.

Cysteine represents an important natural tag for post-translational modification due to its low natural abundance (1-2%) and high nucleophilicity in comparison to other residues. While the thiol group of cysteine has classically been used in maleimide coupling, alkylation and

disulfide formations/exchanges, new methods have also been investigated. Several classical and newer methods to modify cysteine residues are illustrated in Table 1.[145]

As well as directly modifying the cysteine residue, there are methods to convert this residue into a reactive dehydroalanine (dha) tag.[73] This dha tag can then serve as a reactive centre for thio-Michael conjugate additions, dramatically expanding the scope of PTMs that can be chemically incorporated into the protein. By using a milder bis-alkylation reagent, the intermediate could be formed rapidly and then used to introduce site specifically sugars onto a protein backbone.[150] Mechanistic studies by Caddick *et al* have isolated the intermediate cyclic sulfonium,[151] and the application of dha installation has further been extended to the creation of logic gated antibodies (Figure 24).[152]

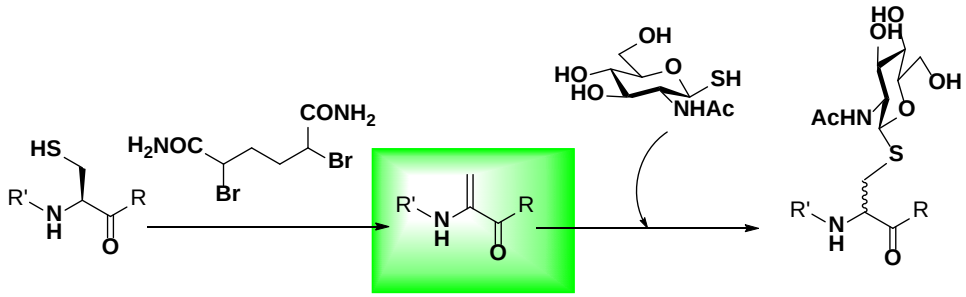
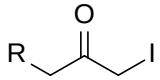
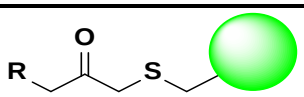
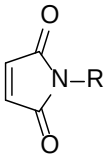
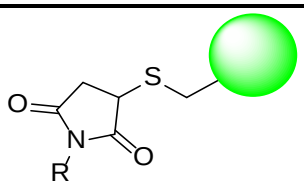
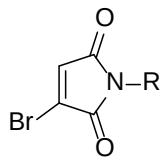
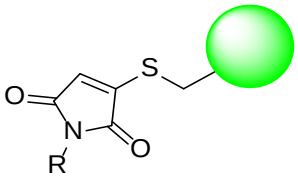
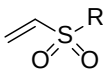
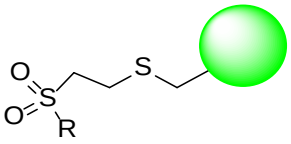
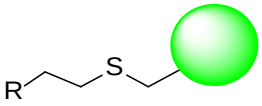
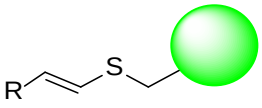
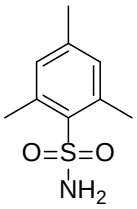
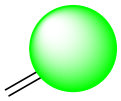
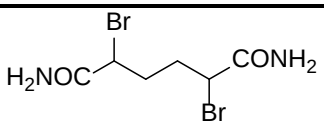
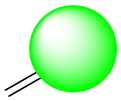
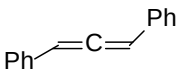
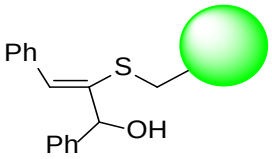


Figure 24. The reaction of a solvent exposed cysteine residue with the dibromomaleimide reagent leads to the formation of dha. This can then react with free thiols to generate new S-linked PTMs.

Table 1. A summary of commonly used cysteine modification reagents. In this example, the green sphere represents a protein. Disulfide bridging reagents are not shown, nor is Ellman's reagent.[153]

Reagent	Product	Reference
		154–156
		157–161

		162
		163
		164
		165
		166
		150
		167

There are also numerous reagents for disulfide re-bridging including dibromomaleimides,[168] dichloro-s-tetrazines[169] and organo arsenical compounds.[170] These are not covered here, but have found extensive use in the stabilisation (locking) of protein conformation.[171]

1.4.1.2 Lysine.

Lysine (K) conjugation, via the ε-amino group, is a traditional target for bio-conjugation reactions. By varying the pH of the reaction buffer, selectivity between differing lysine residues (standard pK_a 10 in water) and the α-amino group (pK_a 8) can often be obtained. At higher pH values, ε-amino groups are deprotonated and thus considerably more nucleophilic. Examples

of reagents used to modify lysine residues include: NHS esters,[172–174] isocyanates,[175] isothiocyanates,[176,177] perfluorinated nitrobenzenes,[178] sulfonyl halides,[179] iminothiolanes,[180] azetidinones,[181,182] dichlorotriazines,[183] α-halocarbonyls, acid anhydrides and also reductive amination via formation of the intermediate imine (Figure 25).[184–186]

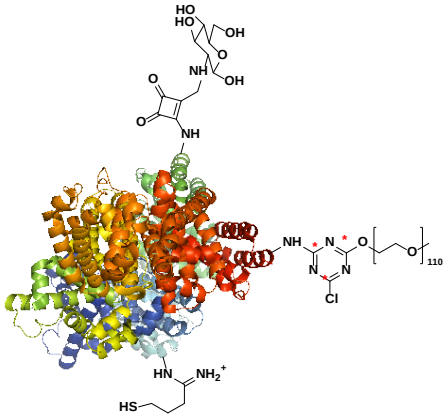


Figure 25. Illustration of more unusual lysine modifications affected using: a) A 2 step squaric acid diester glycosylation[187] b) dichlorotriazine (stars represent ^{14}C) modification[183] and c) reaction with Traut's reagent.[180]

Despite there being several reports of selective lysine modifications of proteins, cross reactivity with a more reactive group (inherent or due to microenvironment) such as a free cysteine is always a concern. Furthermore, the choice of buffer is also essential as TRIS containing buffers will preferentially react over the protein.[145]

1.4.1.3 Other natural residues

Serine and threonine are poor nucleophiles for bio-conjugation reactions, with their hydroxyl groups each having a pK_a>13. In some cases, with neighbouring groups such as histidine or arginine in the microenvironment, reactions with amine reactive reagents such as NHS esters are observed.[145] Histidine modifications are also highly dependent upon the microenvironment of the residue (histidine has a pKa in the physiological range). Isolated examples of histidine modification have been reported in the wider literature including the use of epoxides and

alkenals.[188,189] Arginine has also been used in isolated reports as a reactive handle, mostly via reaction with α-dicarbonyl compounds.[190,191] Aspartic acid and glutamic acids have also received scant attention, mostly being used in reactions with carbodiimide coupling reagents. Methionine isn't commonly used as a reactive handle, but reactions include oxidation or acidic alkylation.[145]

Tyrosine and also tryptophan have been widely used in bioconjugation reactions. Reactions at tyrosine are dependant of the microenvironment and often take advantage of the electron rich character of the aromatic ring or the reactivity of the phenolate hydroxyl. Reactions include acylation (using DMAP activated acyl transfer),[192] ortho-nitration with tetranitromethane,[145] diazonium reactions, Mannich reactions[193] and other organometallic reactions such as π-allylation or oxidative coupling.[194,195] Tryptophan has the second lowest abundance (after Cys) of all natural amino acids; however, at least 90% of proteins contain at least one. Common reagents to modify tryptophan include dialdehydes (before modification to a hydrazine),[196] chlorosulfonium reagents or vinylmetallocarbenoids. Two examples of tryptophan and tyrosine modifications are shown in Figure 26.[197,198]

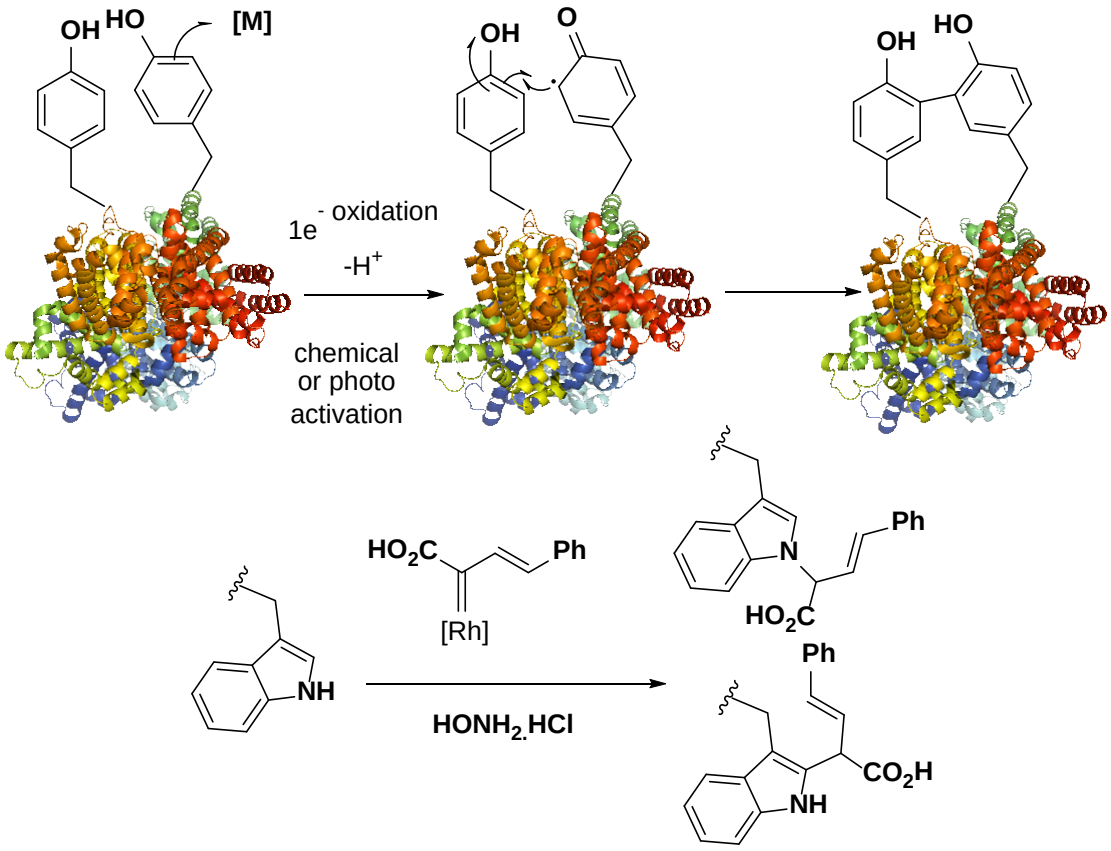


Figure 26. Oxidative coupling of tyrosine residues (top). The reaction of Trp residue with a vinyl rhodium carbenoid (bottom) can occur at 2 sites on the tryptophan indole ring.

1.4.1.4 N-terminal Modification

The selective modification of the N-terminal α-amino group is possible due to the anchimeric assistance of the neighbouring amide, which in turn lowers its pKa compared to that of ε-amino groups. This enables pH controlled selective modifications with acylation reagents (NHS, anhydrides) and other amine reactive reagents; however, in general, these reactions are non-general, as the reactivity is heavily dependent upon the microenvironment of the N-terminus.[145,148] Newer methods which help to alleviate these problems include ketene mediated modification (developed by Che *et al*) and transaminations. A transamination method utilising vitamin B6, was developed by Francis *et al*, which could be conducted under relatively mild

conditions in contrast to previous reports.[199] Later developments have yielded newer transamination reagents such as Rapport's salt,[200] 2PCA[201] and FHMDP[202] which are even more active. Whilst serine and threonine have few uses (in in-chain modification), they have found use as reactive handles at the N-terminus of proteins. Phosphate mediated ligation,[203] periodate oxidation (into a glyoxal)[204,205] and O->N acyl transfers are but a few of the methods available.[206] Tryptophan N-terminal modifications via sulfenylation[207] and Pictet Spengler reactions[208] have also been reported, as well as reactions on other amino acids. A usual N-terminal residue, proline has also been targeted by Francis *et al* in an oxidative coupling with *ortho*-amino phenols.[209] This report in 2015, marked the first targeted use of this residue as a reactive handle in the bio-conjugation literature.

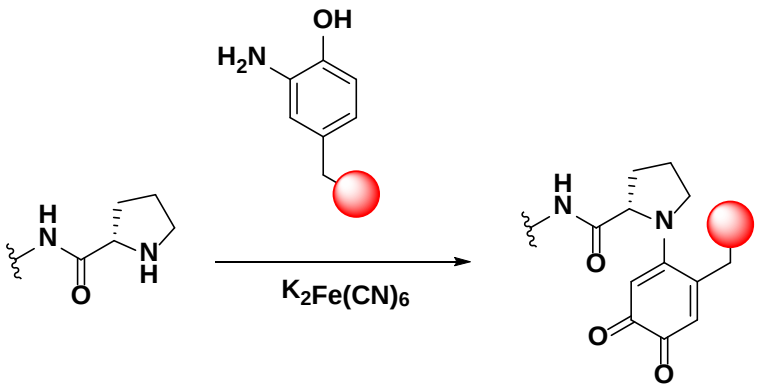


Figure 27. Oxidative N-terminal proline modification, as developed by Francis *et al.* This method uses o-aminophenols or o-catechols which are oxidisied in-situ to an active coupling species.

1.4.2 Native chemical ligation

An N-terminal cysteine was the starting point for the now widely used methodology, native chemical ligation (NCL).[210,211] This method allows for the selective modification of N-terminal residues in the presence of other in-chain nucleophilic groups (cysteine, lysine, tyrosine) and occurs rapidly in the presence of a reductant. In the reaction, an N-terminal cysteine thiolate reacts reversibly with a synthetic thioester, followed by an S->N acyl transfer to form the new peptide amide bond and the free cysteine. There have been several advances in this

methodology since the first reports by Kent *et al*,[212] with new and easier methods to form the thioester being reported, as well as more selective/new methods for the ligation.[213,214] These methods amongst others have been utilised in the formal synthetic synthesis of several proteins including EPO.[215]

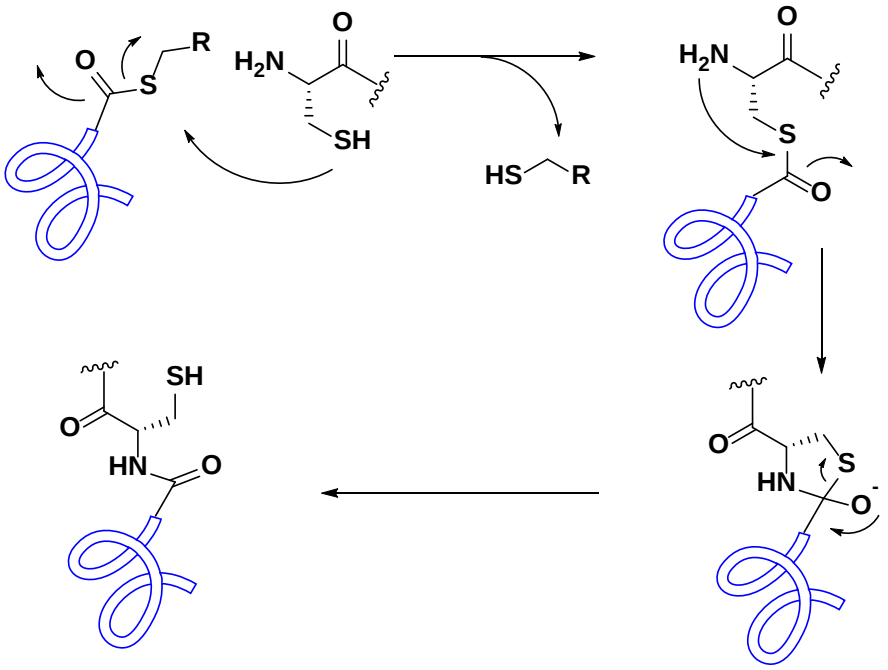


Figure 28. Mechanism for a classical native chemical ligation reaction. Whilst this mechanism is now well known, there is now a pletherea of chemical ligation methologies.

1.4.3 Organometallic protein modifications

Transition metal-mediated protein modification, whether by traditional organometallic reagents (or by metalloproteins) has become an increasingly important area of study in the bioconjugation community.[195,216] Organometallic (R-M), and perhaps more relevantly transition metal complexes, are typically used in the wider organic literature to affect transformation such as hydrogenation, C-C bond formation, C-H functionalisation and C-X formation. Metal catalysed reactions are often conducted in anhydrous organic solvents (such

as THF) and at high substrate concentrations of 100mM-1M. These conditions are clearly incompatible with complex biomolecules, such as proteins, which also contain an added complication of many competing Lewis basic side groups. A summary of common transition metal mediated reactions at incorporated and natural residues in a model protein are shown in Figure 29.

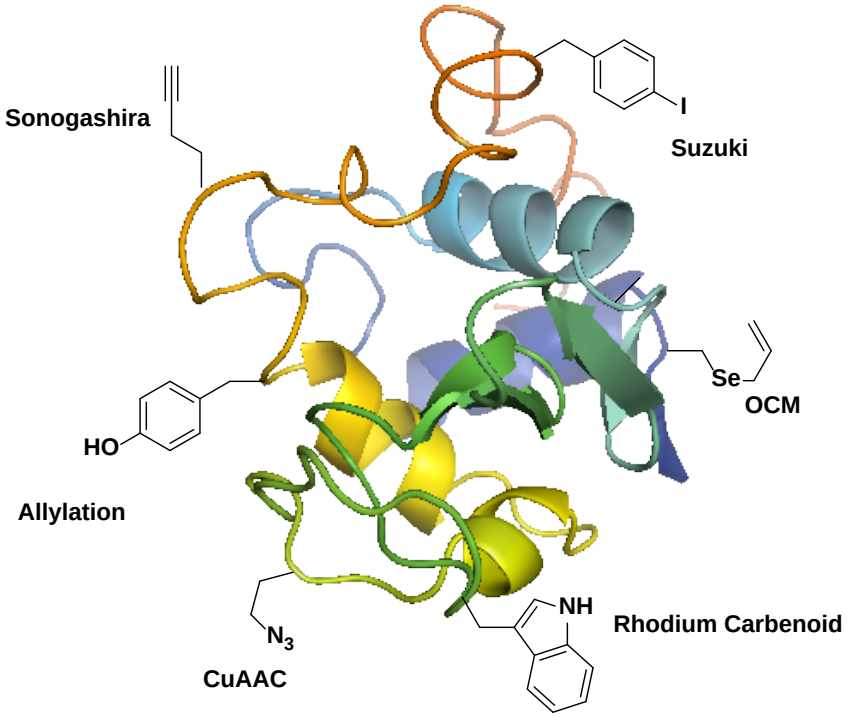


Figure 29. Examples of transition metal mediated reactions that have been demonstrated in the literature. Many of these reactions have also been demonstrated on the surface of living cells.

Aside from copper, the most widely used of the transition metals in protein modification is palladium. Reactions that have been conducted on proteins (containing an UAA tag) that use a palladium catalyst include the Suzuki-Miyori reaction,[217] the Mizoroki-Heck reaction and the Sonogashira reaction.[218] There are also other palladium mediated transformations that have found applications in the modification of natural amino acids such as tyrosine π-allylation.[194] Although there were earlier reports of both Suzuki and Sonogashira reactions on proteins, it wasn't until Chalker *et al* developed the hydroxy-pyridinamine catalysts that the field

developed rapidly.[219] Since, then other catalysts and methods as well as new applications including glycocaylx remodelling have been developed (Figure 30).[220,221]

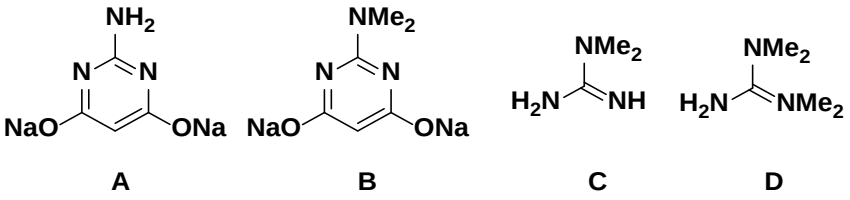


Figure 30. Catalysts B-D show a higher rate of reactivity than the original catalyst (A).

Elegant work by Davis *et al*, also developed ruthenium Grubbs-Hoyveda II catalysts for use in protein modification with allylic chalcogen residues.[222,223] This approach has been investigated mechanistically, where it was found a sulfur relay mechanism may be in operation. There are also extensive examples of rhodium carbenoids for both tryptophan (and cysteine) modifications.[197,198] This approach has been further extended to template/proximity driven modifications. There are also examples of chemical or photochemical cross linking of tyrosine residues on proteins by use of nickel, ruthenium and palladium catalysts. (See previous Figure 29sec).[195] Furthermore, there have also been examples of reductive alkylation of proteins using iridium catalysed transfer hydrogenation.[184]

1.4.4 Metal catalysed cycloadditions: the CuAAC.

Alkyne-azide cycloadditions (1,3-dipolar Husigen cycloadditions) catalysed by predominately copper (or ruthenium)[224] have revolutionised the field of bio conjugation.[225] There has been substantial work into elucidating a mechanism for the CuAAC reaction in-particular, a mechanism supported by kinetic studies, isolation of intermediates and computational modelling is illustrated in Figure 31.[226]

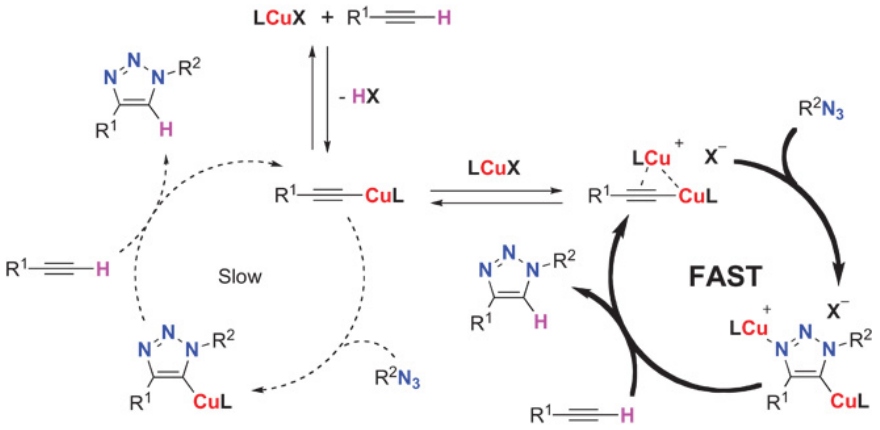


Figure 31. Bertrand *et al* were successful in isolating a key bis copper acetylide intermediate. This enabled them to show that of the 2 pathways, the bis-pathway 1 was the predominant one. This figure is reproduced with permission from Bertrand *et al*, Sci. Advances., 1, 5, 2015

The CuAAC reaction (commonly miscorrectly referred to as the "click") proceeds in often quantitative conversions, rapidly and in aqueous conditions. There are two main methods for employing this reaction in protein modification. The first utilises a source of copper (I) in an appropriate solvent such as MeCN, along with a stabilising ligand.[216,227] Whilst this method shows high rates of reaction the copper species is also prone to oxidation, and high purity copper must be used. The second method uses a copper (II) salt and an excess of reductant, such as ascorbate. For both methods there have been numerous ligands developed, which have further enhanced the scope of this versatile reaction.[228]

1.4.5 Reactions on genetically incorporated unnatural amino acids

In addition to the organometallic reactions outlined previously, due to the genetic incorporation (both residue specific and site specific) of unnatural amino acids, there now exists a plethora of potential conjugation methods. These include nitrone cycloadditions, Staudinger ligations, photoclick reactions (cycloadditions), inverse electron demand Diels Alder reactions, strain promoted azide-alkyne cycloadditions, halosydnones and many more.[229,230] Many of these methods show very high rates of reactions (orders of magnitude more than the CuAAC), use

no toxic metal catalysts and have already been applied in a variety of scenarios including in the labelling of live cells.[74]

1.5. Sialic acids in biology

The sialic acids are a family of biologically important sugars derived from N-acetylneuraminic acid (Neu5Ac (NANA)) and are found in approximately 50 naturally occurring forms.[231] All sialic acid derivatives are composed of a $^2C_5$ pyranose core and in naturally occurring sialosides exist in an α-configuration.[231] Furthermore, sialic acids are most commonly found at the terminal non-reducing ends of complex glycans and are thus implicated in many protein-glycan interactions.[232,233] They are also non-ubiquitous in nature and are mainly found in just the deuterostome lineage and in bacteria. In vertebrates, humans predominately express Neu5Ac and 5,(7),9-N,O-diacetylneuraminic acid (Neu5,(7)9Ac$_2$), whilst porcine cells (pig) are known to predominately express Neu5Gc and murine cells Neu5Ac. In addition, there are several rarer members of the family including: KDN and the sialic acid analogues KDO and pseudaminic acid.[231]

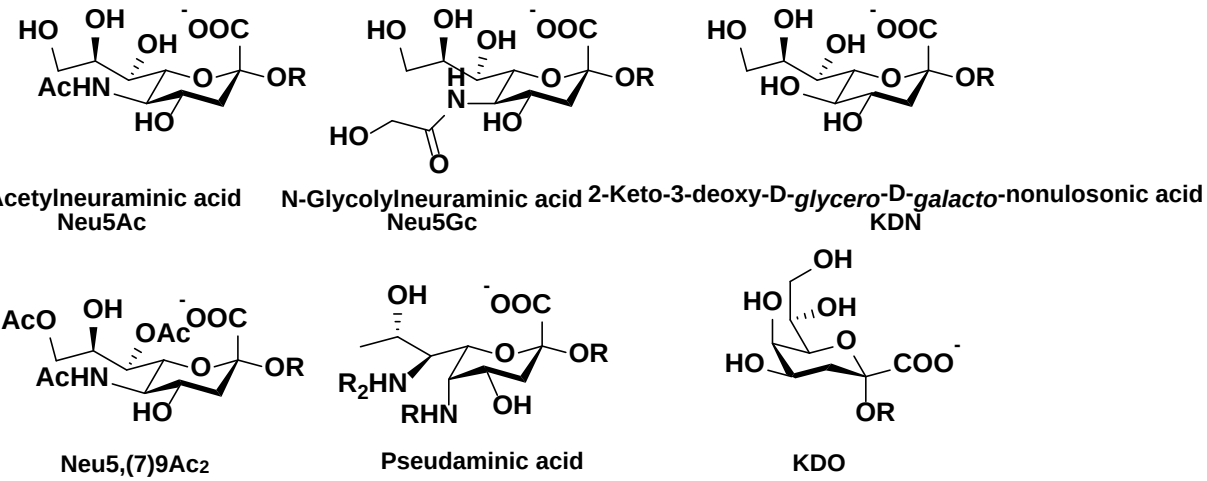


Figure 32. Members of the sialic acid family. N or O linked substituted derivatives of Neu5Ac. KDN glycans are de-aminated neuraminic acids, whereas KDO and pseudaminic acid aren't technically members of the family but are suggested to be synthesised via an analogous biosynthetic route. In all cases R represents another sugar.

1.5.1 Synthetic sialic acid analogues

Due to the weak nature of monovalent (natural) sialic acid-protein interaction, several groups have investigated higher affinity unnatural ligands for sialic acid binding proteins.[234,49] These ligands can then be coated onto multivalent constructs such as liposomes or nanoparticles and used for the specific delivery of cargo to cells.[37,235] Several groups have used structure guided ligand design to create more potent ligands for Siglec-1 (both by crystal structure and *in-silico* screens),[236,237] whilst other groups have used structure-activity relationship guided design to develop new ligands for Siglec-2/7/9/10.[238–240] Modifications at C9 (other positions commonly modified include the C5 amide and C4)[241] with bulky aromatic residues have provided some of the most promising results with several orders of magnitude increases in binding observed (Figure 33.).[242–244]

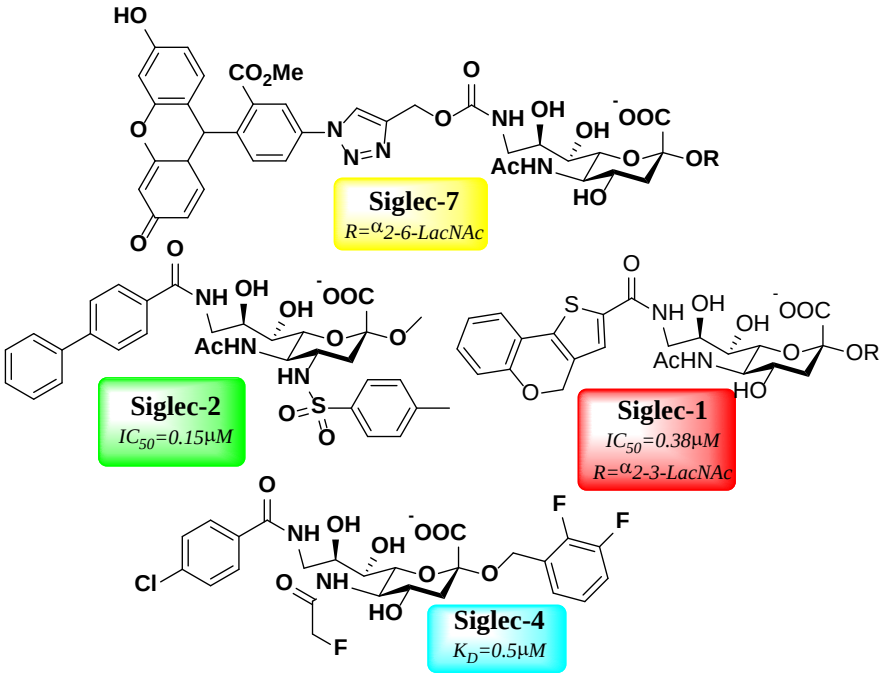


Figure 33. Improved sialic acids structures as higher affinity ligands for Siglecs.[234] Elegant approaches for the high-throughput discovery of Siglec binding ligands have been utilised by Paulson *et al*, in a click and pick methodology.[245]

1.5.2 Multivalent targeting of Siglecs

As briefly outlined above, there is also considerable work investigating the multivalent targeting of Siglec receptors. In one study, Paulson *et al* were able to construct a decavalent protein construct that was conjugated via a nanoparticle to a high affinity CD22 ligand.[246] Through the construction of this low valence multivalent scaffold, Paulson was able to demonstrate a strong binding affinity to CD22, even in the presence of *cis* ligands, due to the bridging of multiple CD22 receptors on B-cells.[247] Furthermore the work showed that this approach even worked for low valence (n=2, IgG) scaffolds validating the importance of balancing avidity with the optimal interaction distance/geometry.

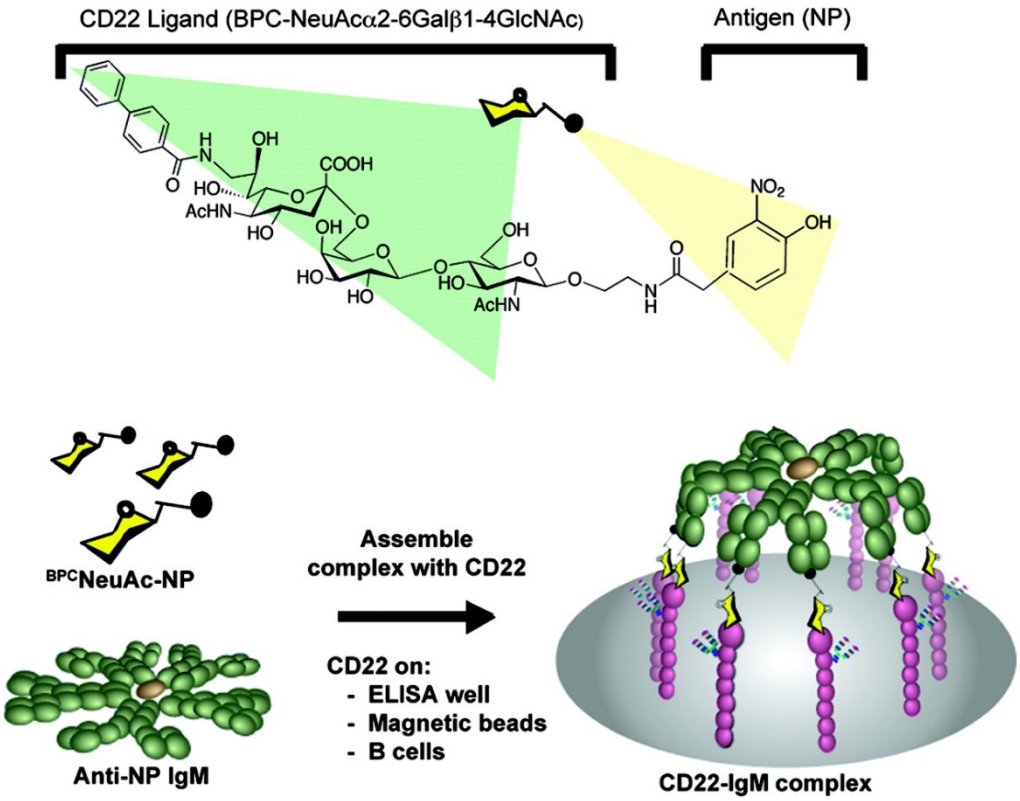


Figure 34. The ligand BPCNeuAc-AP was used in the ligand-driven self-assembly between the decavalent scaffold, anti-NP IgM and CD22. Reprinted with permission from O'Reilly, M. K.; Collins, B. E.; Han, S.; Liao, L.; Rillahan, C.; Kitov, P. I.; Bundle, D. R.; Paulson, J. C. J. Am. Chem. Soc. 2008, 130 (24), 7736–7745. Copyright 2008 American Chemical Society.

Low valency interactions in glycan binding had previously been noted by Bundle *et al* further providing impetuous for the construction of a distance spaced low valence, multivalent probe.[248]

Other multivalent systems used in the study of Siglecs include liposomes. Paulson and Crocker *et al*, have shown that liposomes targeting CD169 and CD22 can be used to deliver cargo to cell presenting these receptors.[36,37] Likewise, elegant work with viral capsids for the targeting of the CD22 receptor has also been demonstrated by Finn *et al*.[249]

The other approach traditionally used to study multivalency in Siglec biology is through the use of glycopolymers. These polymers have been synthesised in a variety of methods including via ruthenium-catalysed olefin metathesis polymerisation,[250] as well as with the use of polyacrylamide (PAA). The advantage of the use of polyacrylamide probes is their ease of synthesis and high avidity (Figure 35), however they are unable to provide detailed structural data regarding binding modes. The method for the use of PAA glycopolymer probes to interrogate a range of Siglecs has been extensively reviewed and experimentally detailed by Paulson *et al*.[49]

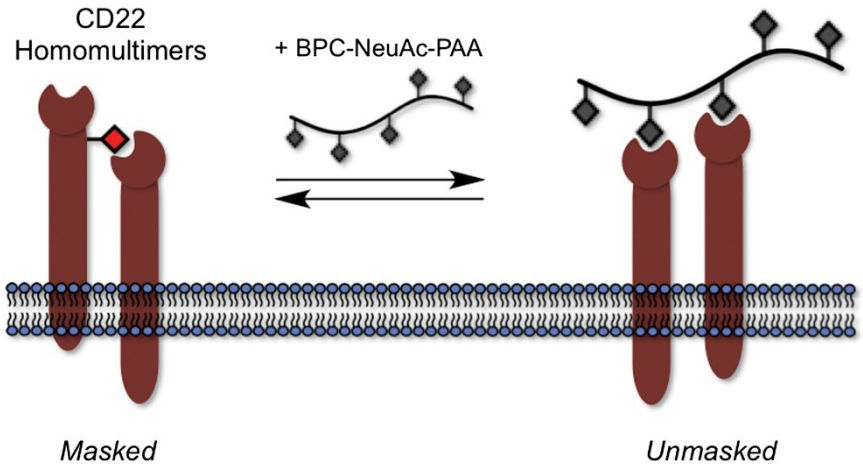


Figure 35. High avidity polymeric probes can unmask Siglecs. In this case CD22 is unmasked from homomultimer clusters in the presence of BPC-NeuAc-PAA. Figure reproduced with permission from Bertozzi *et al.*[44]

1.5.3 Sialic acid-protein interactions: the Siglecs

Sialic acid binding immunoglobin-like lectins (Siglecs) show a great specificity for sialosides and are the main focus of this chapter; however, there are many other proteins which also bind sialic acids.[231] These include the selectins, a family of three C-type lectins called E (CD62E), L (CD62L) and P (CD62P) selectin, that are expressed on endothelial cells, leucocytes and platelets respectively; all three being involved in leucocyte trafficking.[251–253] These selectins show a preference for sialylated and fucosylated glycans in particular siaLe^x (sialyl lewis X) and its 6-sulfated analogues.[254] Other proteins that bind sialic acids include Factor H (in vertebrates) and the influenza A virus hemagglutinin (in viruses).[255,256]

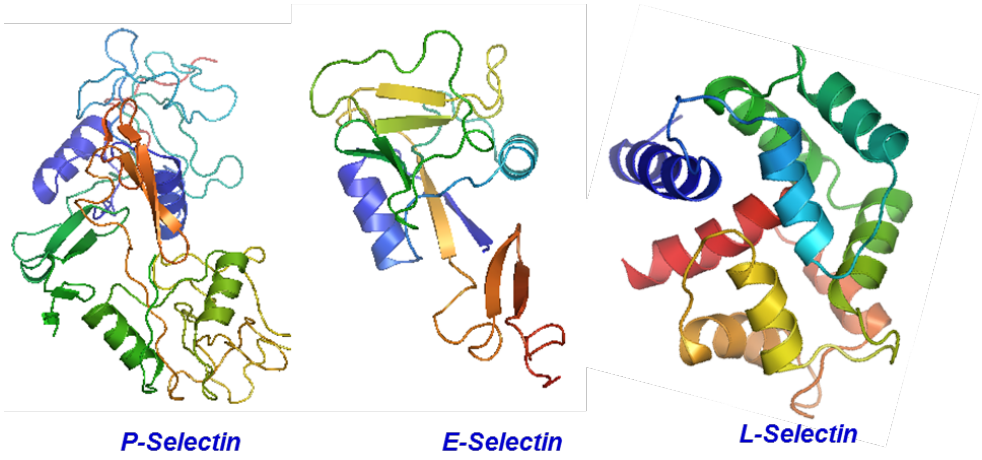


Figure 36. Protein crystal structures of the 3 selectins. P-selectin is expressed on activated endothelial cells and platelets, its expression can be induced by thrombin. E-selectin is expressed on activated endothelial cells and its expression induced by TNFα. L-selectin is expressed on leukocytes.

Siglecs are single pass (type I) membrane proteins, that have an extracellular N-terminal V-set immunoglobulin domain (a variable domain similar to that found on antibodies) and a variable number of C2-set immunoglobulin domains (similar in primary and tertiary structure to Ig constant domain 2). They are thought to be crucial in cell-cell interaction and are thus implicated in cellular communication within both the innate and adaptive immune system.[257,258]

When classifying Siglecs it is appropriate to group them into two families based upon sequence similarity and evolutionary behaviour.[259] One family of Siglecs consists of Siglec-1 (sialoadhesin), Siglec-2 (CD22), Siglec-4 (MAG) and Siglec-15. These 4 Siglecs all have clear orthologues in mammalian species and show 25-30% sequence identity, making them ideal targets to study with murine models. [260]

On the other hand there are the CD33 related Siglecs, which in humans are: Siglec-3 (CD33), Siglec-5 (CD170), Siglec-6 (CD327), Siglec-7 (CD328), Siglec-8, Siglec-9 (CD329), Siglec-10, Siglec-11, Siglec-14 and Siglec-16.[261] These rapidly evolving CD33 related Siglecs differ between species and show high sequence similarities between each other (50-99%) and often have retained intracellular tyrosine based signalling motifs for inhibitory cell signalling. There is also evidence that these CD33 Siglecs have undergone significant changes during human evolution by gene duplication, exon shuffling, exon loss and gene conversion.[262]

The binding of sialylated glycans to the N-terminal domains of sialoadhesin, Siglec-5 and Siglec-7 has demonstrated the basis of the wider Siglec sugar specificity. Whilst binding of sialic acid to a Siglec is as with all lectins, governed by predominantly hydrogen bonding and weaker hydrophobic interactions caused by the stacking of monosaccharides onto the side chain of an aromatic amino acid.[236,263–265] The Siglecs also exhibit a shallow pocket in the N-terminal Ig domain, which contains an arginine. This arginine forms a salt bridge to the carboxylic acid of the sialic acid and is thus a key factor in the sugar specificity of this protein family. Domain swapping experiments have also shown that the variable C-C' loop of the V set domain is a key factor in glycan specificity and has been shown to undergo significant conformational changes upon ligand binding.[266]

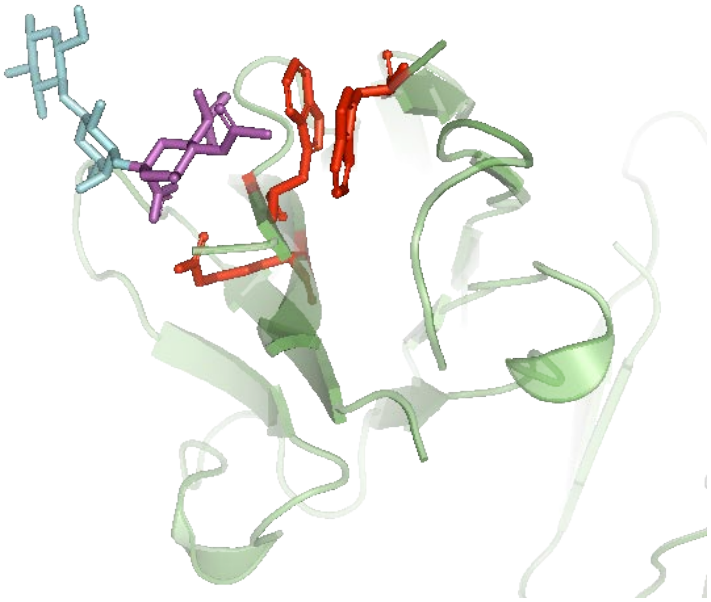


Figure 37. Sialoadhesin binding in the terminal domain with NeuAc. The salt bridge with the Arg96 can be seen. Furthermore, the C9 position can be seen to extend into the hydrophobic region (Trp) supporting the observation that hydrophobic residues at this position enhance binding. (1QFO)[267]

Siglecs are commonly 'masked' by glycan-Siglec interactions with cell surface sialic acids located on the same cell.[268] Given that cell surface sialic acid concentrations can be in excess of 100mM, it is thought that these *cis* interactions (on the same cell surface) help to modulate the function of the Siglec receptors. An exception is sialoadhesin, which due to its greater size (compared to the other Siglecs) is masked less by these *cis* interactions. Siglecs can also bind *trans* to glycans (not on the same cell), especially if the *trans* ligand is a higher affinity artificial ligand as previously described.[269,270] *Trans* binding can occur in the presence of *cis* ligands (*cis* ligands can downregulate this binding as it's in equilibrium) and this *cis/trans* interaction has been shown to be important in both B-cell activation (CD22)[271] and also in NK-cell cytotoxicity (Siglec-7).[272] Siglecs also exhibit a cluster glycoside effect (a high avidity interaction) of either the opposing *trans* counter receptors or of the Siglec receptor itself (such as in CD22). A common way to account for these complex effects from biological experiments is to remove cell-surface sialic acids using sialidase treatment.

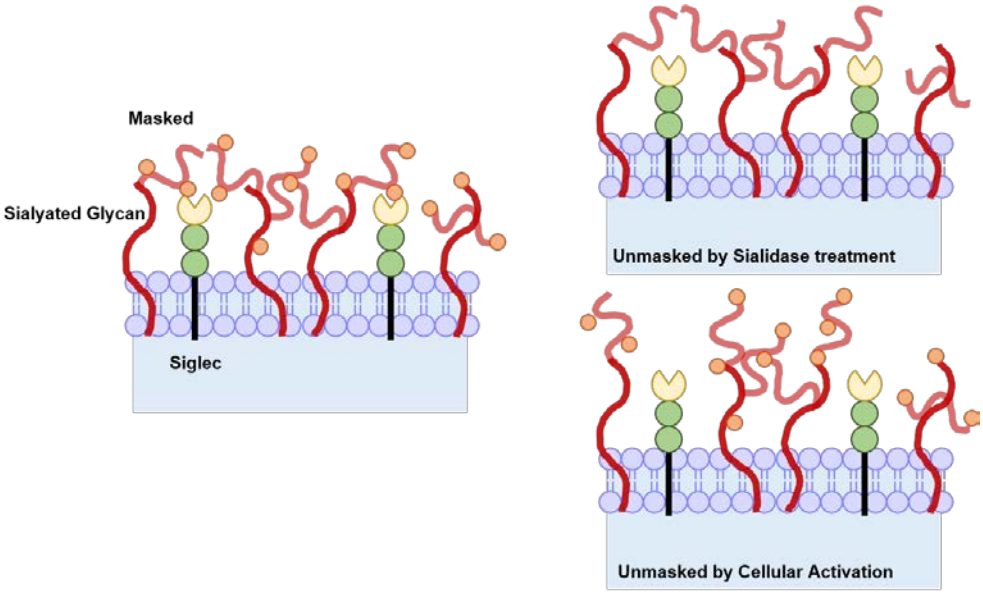


Figure 38. Siglecs are often masked at the cell surface by cis interactions with sialic acids on the cell surface. Upon exposure to sialidase or cellular activation, cis cleavage can occur. Unmasked Siglecs can bind in trans interactions, even *cis* masked Siglecs can bind trans if the binding affinity for *trans* is higher than for *cis*.

The cytoplasmic domains of many Siglecs (especially CD33 Siglecs) also contain many activatory and inhibitory domains that are involved in cell signalling. A lot of Siglecs (CD33/Siglec-2) contain immunoreceptor tyrosine-based inhibitory motifs (ITIMs) which are involved in negative signalling via the recruitment of SHP1 (PTPN6) and SHP2 (PTPN11).[273] In addition the Siglecs: Siglec-14,15 and 16 associate with immunoreceptor tyrosine-based activatory motif (ITAM) adaptor DAP12 indicating that these Siglecs are involved in activating receptors. Siglec-11 and Siglec-2, also contain a Grb2 binding motif (Growth factor receptor-bound protein 2) a positive effector for cellular activation.[260]

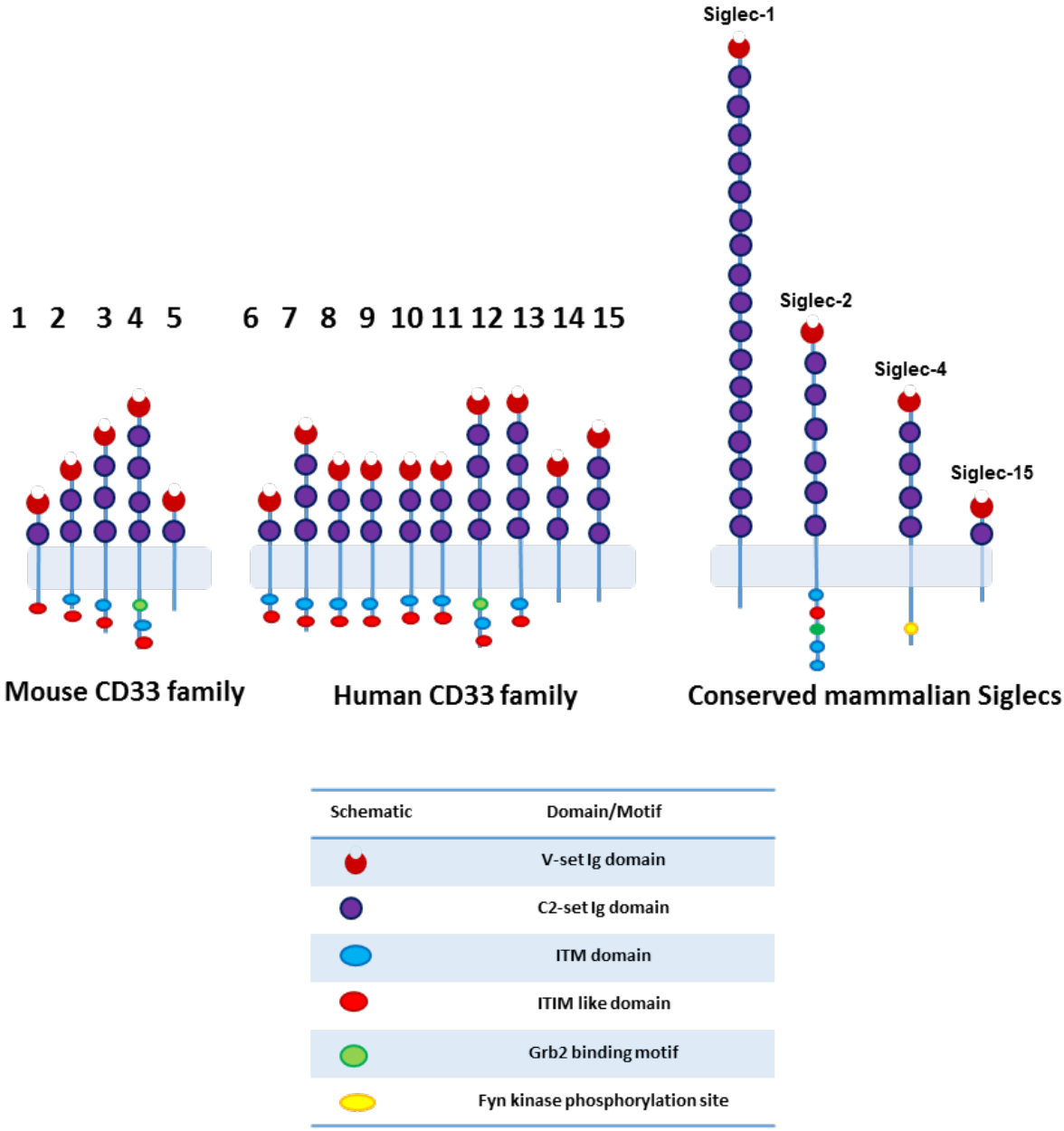


Figure 39. The range of Siglecs and their structures.[233,259] Key: 1.) CD33 2.) Siglec-E 3.) Siglec-F 4.) Siglec-G 5.) Siglec-H 6.) CD33 7.) Siglec-5 8.) Siglec-6 9.) Siglec-7 10.) Siglec-8 11.) Siglec-9 12.) Siglec-10 13.) Siglec-11 14.) Siglec-14 15.) Siglec-16

1.5.4 Siglec 1: Sialoadhesin: CD169

Sialoadhesin (Siglec-1 or Sn) is a cell adhesion molecule, predominately found upon macrophages (Mφ). It is a Siglec with 17 Ig domains (1 variable and 16 constant) and is a member of the Ig super family (IgSF), that doesn't need unmasking from *cis* interactions before binding.[274] It lacks the tyrosine signalling motif in the cytoplasmic tail, again implying it is not involved in cell signalling and shows a binding preference for α2-3 linked sialoasides.[275] Despite Sn being present in subscapular sinus Mφs in the lymph node (which suggests that Sn is involved in endocytosis), other studies showed that Sn was a non-phagocytic receptor for red blood cells. Subsequent studies with *Neisseria meningitide*s showed that Mφs would uptake bacteria dependent upon Sn.[276] In addition, porcine Sn has also been shown to undergo clathrin mediated endocytosis while liposomes displaying a high degree of sialylation were shown to be taken up in a Sn dependant manner.[277] Sn has been implicated in: HIV (cells infected showed 10 fold gene induction)[278] and in the modulation of T-cell activation and function, making it an interesting target for binding studies.[274,279,280]

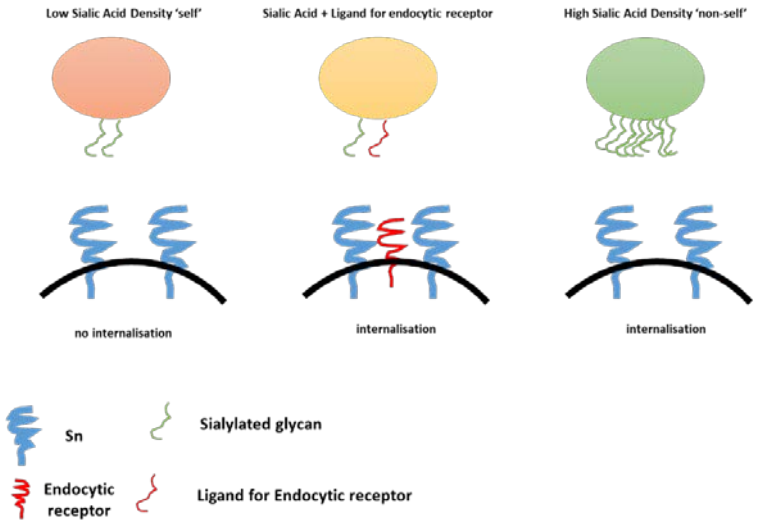


Figure 40. Possible Sn binding modes. With a low sialic acid density on the opposing particle no internalisation into a Mϕ is seen. When the opposing particle displays additional ligands, Sn may synergise and then drive internalisation. If the Sialic acid density on the opposing particle is high, phagocytosis by Mϕ may occur.

1.5.5 Siglec-2 CD22, Siglec-4 MAG & Siglec-15 (Siglec-H)

Siglec-2, CD22 is a regulator for B cell signalling, homeostasis and survival.[281–283] It is predominately involved with setting a threshold for the antigen-induced activation of B-cells via a strong binding preference for α2-6 linked sialosides.[36,235,257]

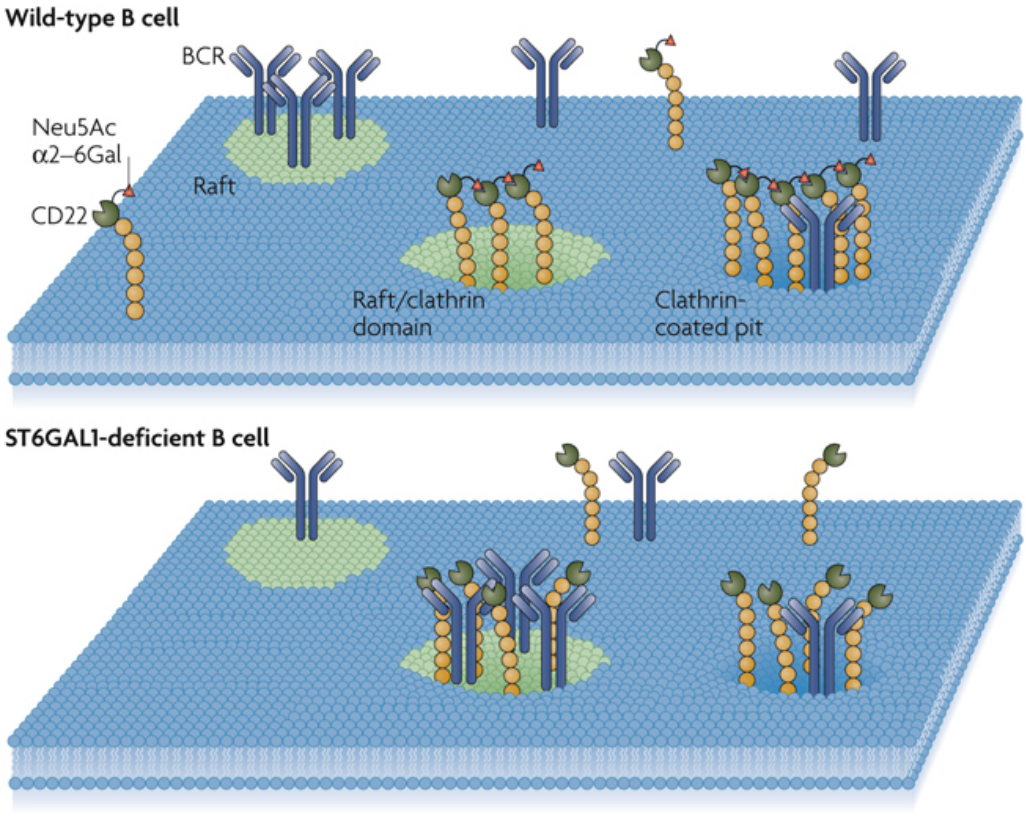


Figure 41. CD22 internalisation models adapted with permission from Nature Publishing Group.[258] In resting B-cells CD22 is found predominately within clathrin coated pits. When the cell surface cis interactions with glycans are removed, the B-cell receptor localises with CD22 in these clathrin coated pits.

Myelin-associated glycoprotein (MAG), or Siglec-4, is a membrane protein found on periaxonal Schwann cells and on the oligodendroglia membranes of myelin.[284] It helps to stabilise myelin-axon interactions by binding to α2-3 linked sialic acid gangliosides. The smallest Siglec, Siglec-15, has a very small cytoplasmic region and is mostly expressed in macrophages and dendritic cells. It preferentially binds to α2-3 sialosides but little is known of its exact role in immunity.[285]

1.5.6 The CD33 family of Siglecs: CD33, Siglec-5/14, Siglec-6, Siglec-7, Siglec-8 (F in mouse), Siglec-9 (E in mouse), Siglec-10 (G in mouse), Siglec-11 and Siglec-12/13.

CD33 is a myeloid specific surface protein on progenitor and leukemic cells and has been implicated in cellular proliferation or differentiation (during haematopoietic differentiation it is expressed earlier than other Siglecs). As well as myelomonocytic cells, it is also expressed on mast cells and blood basophils and like other Siglecs it binds with the highest affinity to α2-6 linked sialic acids. However *cis* interactions serve to mask it.[262,286] Siglec-5 is expressed on neutrophils, monocytes, basophils, mast cells and macrophages. Increased amounts of Siglec-5 expression have been found on alveolar and lymphatic (node) macrophages upon activation by bacterial or virial pathogens. One suggested role is in phagocytosis. Interestingly its specificity is almost identical to that of Siglec-14 (as is its structure) and it thus may be that previous work on Siglec-5 may have also been activating Siglec-14. The main difference between Siglec-14 and Siglec-5 is that Siglec-14 has 3 Ig domains whereas Siglec-5 has 4 (first 2 of each are almost identical).[258,259]

Siglec-6 was originally found to be a leptin binding protein (protein hormone involved in obesity) and a placental protein, however much of its biology regarding its role in fertility remains unclear. Siglec-7 is highly expressed on human NK (natural killer) cells and very weakly on monocytes.[287] The role of Siglec-7 has been found to be inhibitory and has been suggested as a possible target for therapeutic antibodies in the treatment of myeloid leukaemia.[266,288]

Siglec-8 is selectively expressed in eosinophils but weakly expressed in human mast cells and human basophils. Eosinophils when exposed to Siglec-8 antibodies *in vitro*, underwent apoptosis (programmed cell death), indicating that activated (by allergen) eosinophils may be

susceptible to Siglec-8 mediated cell death.[289] Mast cell dependent contraction of human airway rings *in vitro* was also inhibited by Siglec-8. It is important to note that Siglec-8 has evolved to recognise 6'-sulfated sialyl Lewis X. Siglec-9 is expressed predominately on neutrophils and monocytes. Immature neutrophils obtain Siglec-9 on their cell surfaces late in cell differentiation but before CD16 (cluster of differentiation 16) expression occurs. Siglec-9 antibodies have been linked to apoptosis.[286]

Siglec-10 is expressed mainly on monocytes, dendritic cells and weakly on eosinophils. The preference here is for an α2-6 linkage however functional studies are lacking.[258]

Siglec-11 is expressed on cells in tissues such as Kupffer cells and brain microglia, they bind well to α2-8 sialoconjugates but no functional studies have been reported. [258] Siglec-12 has 2 V-set regions and is present in humans, Siglec-13 is only in primates. Little is known of its biological functions.[260]

1.5.7 Siglecs as a therapy area.

As of 2015, there are currently 18 anti-Siglec antibodies and derivatives in development for the treatment of a range of diseases. 10 of these biologics are specific against CD22, 6 against CD33 and the remaining two against Siglec-4/Siglec-15 respectively. The antibody derivatives targeting CD33 (Siglec-3) are exclusively for the treatment for acute myeloid leukaemia. CD22 targeting therapeutics treat a range of cancers including acute lymphoid leukaemia, follicular lymphoma and non-Hodgkin lymphoma.[234]

1.6 Aims of this project

This project aims to construct a series of distance defined multivalent probes of the glycan-binding proteins: the Siglecs. By creating bivalent probes for the Siglec receptors, in particular, Sn, the aim is to correlate a binding distance to a biological effect such as internalisation. This could allow for the creation of more specific therapeutics. An additional goal of this work will be to demonstrate the most advanced use of the "tag and modify" strategy to date. The aim to demonstrate an application of the tag and modify methodology also meant that polymeric or polypeptide methodologies were not considered.

As the Np276 protein has been previously used within the Ben Davis group as a model protein for protein modification reactions, it was chosen as the scaffold. The regular spacing of coils in Np276 allows the distance defined presentation of 2 glycans. By distance spacing them, a physical distance (in this case defined by the X-ray structure) could be related to a preference for receptor binding distance, and then potentially a further biological effect.

A summary of the aims of this project are:

- Express Np276 protein in a clean and folded state including with characterisation, and also with residue specific incorporation of the reactive "tag", azidohomoalanine.
- Modify the protein with both glycans and also 1 equivalent of dye so that the reaction can be followed in binding studies. The dye modification should occur at one site and away from where the glycans are being presented on the scaffold.
- Demonstrate a binding effect initially with a model cell line (a CHO cell line) using a receptor that effectively binds without Siglec-glycan cis interactions.

- If a family of these rulers can be tested with varying distances and from this a preference for a particular distance elucidated this may form the basis for a more detailed exploration.

1.7 Bibliography

(1) Krishnamurthy, V. M.; Estroff, L. a; Whitesides, G. M. *Multivalency in Ligand Design*; 2006.

(2) Kitov, P. I.; Bundle, D. R. *J. Am. Chem. Soc.* **2003**, *125* (52), 16271–16284.

(3) Mammen, M.; Choi, S. K.; Whitesides, G. M. *Angew. Chemie - Int. Ed.* **1998**, *37* (20), 2754–2794.

(4) Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E. W.; Haag, R. *Angew. Chemie - Int. Ed.* **2012**, *51* (42), 10472–10498.

(5) Diestler, D. J.; Knapp, E. W. *J. Phys. Chem. C* **2010**, *114* (12), 5287–5304.

(6) Jencks, W. P. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (7), 4046–4050.

(7) Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci.* **1971**, *68* (8), 1678–1683.

(8) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. *Angew. Chemie - Int. Ed.* **2004**, *43* (48), 6596–6616.

(9) Dunitz, J. D. *Chem. Biol.* **1995**, *2* (11), 709–712.

(10) Gilli, P.; Ferretti, V.; Gilli, G.; Borea, P. A. *J. Phys. Chem.* **1994**, *98*, 1515–1518.

(11) Ercolani, G. *J. Am. Chem. Soc.* **2003**, *125* (51), 16097–16103.

(12) Hunter, C. A.; Anderson, H. L. *Angew. Chemie - Int. Ed.* **2009**, *48* (41), 7488–7499.

(13) Mammen, M.; Shakhnovich, E. I.; Whitesides, G. M. *J. Org. Chem.* **1998**, *3263* (11), 3168–3175.

(14) Joshi, A.; Vance, D.; Rai, P.; Thiyagarajan, A.; Kane, R. S. *Chem. - A Eur. J.* **2008**, *14* (26), 7738–7747.

(15) Kane, R. S. *Langmuir* **2010**, *26* (11), 8636–8640.

(16) Krishnamurthy, V. M.; Semetey, V.; Bracher, P. J.; Shen, N.; Whitesides, G. M. *J. Am. Chem. Soc.* **2007**, *129* (5), 1312–1320.

(17) Varner, C. T.; Rosen, T.; Martin, J. T.; Kane, R. S. *Biomacromolecules* **2015**, *16* (1), 43–55.

(18) Zhang, Z.; Eckert, M. A.; Ali, M. M.; Liu, L.; Kang, D. K.; Chang, E.; Pone, E. J.; Sender, L. S.; Fruman, D. A.; Zhao, W. *ChemBioChem* **2014**, *15* (9), 1268–1273.

(19) Englund, E. a.; Wang, D.; Fujigaki, H.; Sakai, H.; Micklitsch, C. M.; Ghirlando, R.; Martin-Manso, G.; Pendrak, M. L.; Roberts, D. D.; Durell, S. R.; Appella, D. H. *Nat. Commun.* **2012**, *3*, 614.

(20) Zhao, X.; Lis, J. T.; Shi, H. *Nucleic Acid Ther.* **2013**, *23* (3), 238–242.

(21) Ahmad, K. M.; Xiao, Y.; Tom Soh, H. *Nucleic Acids Res.* **2012**, *40* (22), 11777–11783.

(22) Chu, T.-W.; Kopeček, J. *Biomater. Sci.* **2015**, *3* (7), 908–922.

(23) Hollenbeck, J. J.; Danner, D. J.; Landgren, R. M.; Rainbolt, T. K.; Roberts, D. S. *Biomacromolecules* **2012**, *13* (7), 1996–2002.

(24) Vetri, V.; Carrotta, R.; Picone, P.; Di Carlo, M.; Militello, V. *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804* (1), 173–183.

(25) Zacco, E.; Hütter, J.; Heier, J. L.; Mortier, J.; Seeberger, P. H.; Lepenies, B.; Koksch, B. *ACS*

Chem. Biol. **2015**, *10* (9), 2065–2072.

(26) Kroll, C.; Mansi, R.; Braun, F.; Dobitz, S.; Maecke, H. R.; Wennemers, H. *J. Am. Chem. Soc.* **2013**.

(27) Moradi, M.; Babin, V.; Roland, C.; Darden, T. a; Sagui, C. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (49), 20746–20751.

(28) Schuler, B.; Lipman, E. a; Steinbach, P. J.; Kumke, M.; Eaton, W. a. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (8), 2754–2759.

(29) Ungar-Waron, H.; Gurari, D.; Hurwitz, E.; Sela, M. *Eur. J. Immunol.* **1973**, *3* (4), 201–205.

(30) Liu, P.; Chen, L.; Toh, J. K. C.; Ang, Y. L.; Jee, J.-E.; Lim, J.; Lee, S. S.; Lee, S.-G. *Chem. Sci.* **2015**, *6* (1), 450–456.

(31) Levine, P. M.; Carberry, T. P.; Holub, J. M.; Kirshenbaum, K. *Medchemcomm* **2013**, *4* (3), 493–509.

(32) Matassini, C.; Marradi, M.; Cardona, F.; Parmeggiani, C.; Robina, I.; Moreno-Vargas, A. J.; Penades, S.; Goti, A. *RSC Adv.* **2015**, 95817–95822.

(33) Davis, M. E.; Chen, Z. G.; Shin, D. M. *Nat. Rev. Drug Discov.* **2008**, *7* (9), 771–782.

(34) Sigal, G. B.; Mammen, M.; Dahmann, G.; Whitesides, G. M. *J. Am. Chem. Soc.* **1996**, *118* (16), 3789–3800.

(35) Kawasaki, N.; Vela, J. L.; Nycholat, C. M.; Rademacher, C.; Khurana, A.; van Rooijen, N.; Crocker, P. R.; Kronenberg, M.; Paulson, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (19), 7826–7831.

(36) Macauley, M. S.; Pfrengle, F.; Rademacher, C.; Nycholat, C. M.; Gale, A. J.; von Drygalski, A.; Paulson, J. C. *J. Clin. Invest.* **2013**, *123* (7), 3074–3083.

(37) Chen, W. C.; Kawasaki, N.; Nycholat, C. M.; Han, S.; Pilotte, J.; Crocker, P. R.; Paulson, J. C. *PLoS One* **2012**, *7* (6), e39039.

(38) Mallikaratchy, P. R.; Ruggiero, A.; Gardner, J. R.; Kuryavyi, V.; Maguire, W. F.; Heaney, M. L.; McDevitt, M. R.; Patel, D. J.; Scheinberg, D. A. *Nucleic Acids Res.* **2011**, *39* (6), 2458–2469.

(39) Rolland, O.; Turrin, C.-O.; Caminade, A.-M.; Majoral, J.-P. *New J. Chem.* **2009**, *33* (9), 1809.

(40) Mintzer, M. A.; Dane, E. L.; O'Toole, G. A.; Grinstaff, M. W. *Mol. Pharm.* **2012**, *9* (3), 342–354.

(41) Garcia-Vallejo, J. J.; Ambrosini, M.; Overbeek, A.; van Riel, W. E.; Bloem, K.; Unger, W. W. J.; Chiodo, F.; Bolscher, J. G.; Nazmi, K.; Kalay, H.; van Kooyk, Y. *Mol. Immunol.* **2013**, *53* (4), 387–397.

(42) Xie, J.; Wang, J.; Chen, H.; Shen, W.; Sinko, P. J.; Dong, H.; Zhao, R.; Lu, Y.; Zhu, Y.; Jia, L. *Sci. Rep.* **2015**, *5*, 9445.

(43) Müller, C.; Despras, G.; Lindhorst, T. K. *Chem. Soc. Rev.* **2016**, *45* (11), 3275–3302.

(44) Belardi, B.; Bertozzi, C. R. *Chem. Biol.* **2015**, *22* (8), 983–993.

(45) Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002**, *102* (2), 555–578.

(46) Yilmaz, G.; Becer, C. R. *Front. Bioeng. Biotechnol.* **2014**, *2* (October), 1–18.

(47) Becer, C. R. *Macromol. Rapid Commun.* **2012**, *33* (9), 742–752.

(48) Zhang, Q.; Collins, J.; Anastasaki, A.; Wallis, R.; Mitchell, D. A.; Becer, C. R.; Haddleton, D. M. *Angew. Chemie Int. Ed.* **2013**, *52* (16), 4435–4439.

(49) O'Reilly, M. K.; Paulson, J. C. *Multivalent ligands for siglecs.*, 1st ed.; Elsevier Inc, 2010; Vol. 478.

(50) Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E. L. L.; Studholme, D. J.; Yeats, C.; Eddy, S. R. *Nucleic Acids Res.* **2004**, *32* (Database issue), D138-41.

(51) Jacoby, G. a; Hooper, D. C. *Antimicrob. Agents Chemother.* **2013**, *57* (4), 1930–1934.

(52) Vetting, M.; Hegde, S.; Fajardo, E.; Fiser, A.; Roderick, S.; Takiff, H.; Blanchard, J. *Biochemistry* **2006**, *45* (1).

(53) Buchko, G. W.; Ni, S.; Robinson, H.; Welsh, E.; Pakrasi, H.; Kennedy, M. *Protein Sci.* **2006**, *15* (11), 2579–2595.

(54) Bateman, a; Murzin, a G.; Teichmann, S. a. *Protein Sci.* **1998**, *7* (6), 1477–1480.

(55) Shah, S.; Heddle, J. G. *Appl. Microbiol. Biotechnol.* **2014**, *98* (23), 9545–9560.

(56) Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (5), 2754–2759.

(57) Hegde, S. S.; Vetting, M. W.; Mitchenall, L. a; Maxwell, A.; Blanchard, J. S. *Antimicrob. Agents Chemother.* **2011**, *55* (1), 110–117.

(58) Gilchrist, C. A.; Gray, D. A.; Baker, R. T. *J. Biol. Chem.* **1997**, *272* (51), 32280–32285.

(59) Mitchenall, L. A.; Maxwell, A.; Takiff, H. E.; Blanchard, J. S. *Science (80-.).* **2005**, *308* (June), 1480–1484.

(60) Vetting, M. W.; Hegde, S. S.; Zhang, Y.; Blanchard, J. S. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2011**, *67* (Pt 3), 296–302.

(61) Buchko, G. W.; Robinson, H.; Ni, S.; Pakrasi, H. B.; Kennedy, M. a. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2006**, *62* (Pt 12), 1251–1254.

(62) Park, K. S.; Lee, J. H.; Jeong, D. U.; Lee, J. J.; Wu, X.; Jeong, B. C.; Kang, C.-M.; Lee, S. H. *Antimicrob. Agents Chemother.* **2011**, *55* (9), 4475–4478.

(63) Vetting, M. W.; Hegde, S. S.; Hazleton, K. Z.; Blanchard, J. S. *Protein Sci.* **2007**, *16* (4), 755–760.

(64) Tavío, M. M.; Jacoby, G. a; Hooper, D. C. *J. Antimicrob. Chemother.* **2014**, *69* (8), 2102–2109.

(65) Nöllmann, M.; Crisona, N. J.; Arimondo, P. B. *Biochimie* **2007**, *89* (4), 490–499.

(66) Bates, A. D. *Curr. Biol.* **2006**, *16* (6), R204-6.

(67) Tao, J.; Han, J.; Wu, H.; Hu, X.; Deng, J.; Fleming, J.; Maxwell, A.; Bi, L.; Mi, K. *Nucleic Acids Res.* **2013**, *41* (4), 2370–2381.

(68) Gilbert, J. a; Davies, P. L.; Laybourn-Parry, J. *FEMS Microbiol. Lett.* **2005**, *245* (1), 67–72.

(69) Garnham, C. P.; Gilbert, J. a; Hartman, C. P.; Campbell, R. L.; Laybourn-Parry, J.; Davies, P. L. *Biochem. J.* **2008**, *411* (1), 171–180.

(70) Garnham, C. P.; Campbell, R. L.; Davies, P. L. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (18), 7363–7367.

(71) Madhurantakam, C.; Varadamsetty, G.; Grutter, M. G.; Pluckthun, A.; Mittl, P. R. E. *Protein*

Sci. **2012**, *21* (7), 1015–1028.

(72) Meeks, J. C.; Elhai, J.; Thiel, T.; Potts, M.; Larimer, F.; Lamerdin, J.; Predki, P.; Atlas, R. *Photosynthesis Research*. 2001, pp 85–106.

(73) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G. *Acc. Chem. Res.* **2011**, *44* (9), 730–741.

(74) Sletten, E. M.; Bertozzi, C. R. *Angew. Chemie Int. Ed.* **2009**, *48* (38), 6974–6998.

(75) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Curr. Opin. Chem. Biol.* **2010**, *14* (6), 774–780.

(76) Budisa, N. *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (47), 6426–6463.

(77) Schimmel, P. *Annu. Rev. Biochem.* **1987**, *56* (1), 125–158.

(78) Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, *7* (6), 881–889.

(79) Ibba, M.; Söll, D. *Annu. Rev. Biochem.* **2000**, *69* (1), 617–650.

(80) Arnez, J. G.; Moras, D. *Trends Biochem. Sci.* **1997**, *22* (6), 211–216.

(81) Richmond, M. H. *Bacteriol. Rev.* **1962**, *26* (4), 398–420.

(82) Munier, R.; Cohen, G. N. *Incorporation d'analogues structuraux d'aminoacides dans les protéines bactériennes*; Elsevier, 1956; Vol. 21.

(83) Cowie, D. B.; Cohen, G. N. *Biochim. Biophys. Acta* **1957**, *26* (2), 252–261.

(84) Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. *Eur. J. Biochem.* **1995**, *230* (2), 788–796.

(85) Budisa, N.; Karnbrock, W.; Steinbacher, S.; Humm, a; Prade, L.; Neuefeind, T.; Moroder, L.; Huber, R. *J. Mol. Biol.* **1997**, *270* (4), 616–623.

(86) Lang, K.; Chin, J. W. *Chem. Rev.* **2014**.

(87) Ngo, J. T.; Tirrell, D. A. *Acc. Chem. Res.* **2011**, *44* (9), 677–685.

(88) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (25), 9482–9487.

(89) Eichelbaum, K.; Winter, M.; Diaz, M. B.; Herzig, S.; Krijgsveld, J. *Nat. Biotechnol.* **2012**, *30* (10), 984–990.

(90) Howden, A. J. M.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat. Methods* **2013**, *10* (4), 343–346.

(91) Hodas, J. J. L.; Nehring, A.; Höche, N.; Sweredoski, M. J.; Pielot, R.; Hess, S.; Tirrell, D. A.; Dieterich, D. C.; Schuman, E. M. *Proteomics* **2012**, *12* (15–16), 2464–2476.

(92) Hinz, F. I.; Dieterich, D. C.; Tirrell, D. A.; Schuman, E. M. *ACS Chem. Neurosci.* **2012**, *3* (1), 40–49.

(93) Beatty, K. E.; Tirrell, D. A. *Bioorg. Med. Chem. Lett.* **2008**, *18* (22), 5995–5999.

(94) Dieterich, D. C.; Lee, J. J.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Nat. Protoc.* **2007**, *2* (3), 532–540.

(95) Lepthien, S.; Hoesl, M. G.; Merkel, L.; Budisa, N. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (42), 16095–16100.

(96) Cash, M. T.; Schreiner, P. R.; Phillips, R. S. *Org. Biomol. Chem.* **2005**, *3* (20), 3701–3706.

(97) Gerke, V.; Moss, S. E. *Physiol. Rev.* **2002**, *82* (2).

(98) Broos, J.; Gabellieri, E.; Biemans-Oldehinkel, E.; Strambini, G. B. *Protein Sci.* **2003**, *12* (9), 1991–2000.

(99) Goulding, A.; Shrestha, S.; Dria, K.; Hunt, E.; Deo, S. K. *Protein Eng. Des. Sel.* **2008**, *21* (2), 101–106.

(100) Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2003**, *328* (5), 1071–1081.

(101) Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.* **1995**, *117* (1), 536–537.

(102) van Hest, J. C. M.; Tirrell, D. A. *Chem. Commun.* **2001**, *0* (19), 1897–1904.

(103) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27* (19), 5471–5475.

(104) Minks, C.; Huber, R.; Moroder, L.; Budisa, N. *Anal. Biochem.* **2000**, *284* (1), 29–34.

(105) Bronskill, P. M.; Wong, J. T. *Biochem. J.* **1988**, *249* (1), 305–308.

(106) Kitevski-LeBlanc, J. L.; Prosser, R. S. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *62*, 1–33.

(107) Seifert, M. H. J.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. *J. Am. Chem. Soc.* **2002**, *124* (27), 7932–7942.

(108) Budisa, N.; Pipitone, O.; Siwanowicz, I.; Rubini, M.; Pal, P.; Holak, T.; Gelmi, M. *Chem. Biodivers.* **2004**, *1* (10), 1465–1475.

(109) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. *Angew. Chemie Int. Ed.* **2001**, *40* (5), 923–925.

(110) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chem. Biol.* **1999**, *6* (2), 63–70.

(111) Jenkins, C. L.; Raines, R. T. *Nat. Prod. Rep.* **2002**, *19* (1), 49–59.

(112) Massimo Cametti, Benoit Crousse, Pierangelo Metrangolo, Roberto Milanici, Massimo Cametti, G. R. *Chem. Soc. Rev.* **2012**, *41* (1), 31–42.

(113) Lee, K. H.; Lee, H. Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43* (51), 16277–16284.

(114) Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123* (44), 11089–11090.

(115) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **2015**, *6* (1), 50–69.

(116) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *Journal of the American Chemical Society*. American Chemical Society 2000, pp 5010–5011.

(117) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292* (5516), 498–500.

(118) Zhang, Z.; Wang, L.; Brock, A.; Schultz, P. G. *Angew. Chemie Int. Ed.* **2002**, *41* (15), 2840–2842.

(119) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. *Biochemistry* **2009**, *48* (25), 5953–5962.

(120) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124* (31), 9026–9027.

(121) Wang, L.; Xie, J.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68* (1), 174–176.

(122) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. *J. Am. Chem. Soc.* **2012**, *134* (6), 2898–2901.

(123) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (1), 56–61.

(124) Lee, H. S.; Spraggon, G.; Schultz, P. G.; Wang, F. *J. Am. Chem. Soc.* **2009**, *131* (7), 2481–2483.

(125) Sakamoto, K.; Murayama, K.; Oki, K.; Iraha, F.; Kato-Murayama, M.; Takahashi, M.; Ohtake, K.; Kobayashi, T.; Kuramitsu, S.; Shirouzu, M.; Yokoyama, S. *Structure* **2009**, *17* (3), 335–344.

(126) Ai, H. W.; Shen, W.; Brustad, E.; Schultz, P. G. *Angew. Chemie - Int. Ed.* **2010**, *49* (5), 935–937.

(127) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126* (44), 14306–14307.

(128) Soma, A.; Himeno, H. *Nucleic Acids Res.* **1998**, *26* (19), 4374–4381.

(129) Blight, S. K.; Larue, R. C.; Mahapatra, A.; Longstaff, D. G.; Chang, E.; Zhao, G.; Kang, P. T.; Green-Church, K. B.; Chan, M. K.; Krzycki, J. A. *Nature* **2004**, *431* (7006), 333–335.

(130) Kobayashi, T.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *J. Mol. Biol.* **2009**, *385* (5), 1352–1360.

(131) Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. *J. Am. Chem. Soc.* **2009**, *131* (40), 14194–14195.

(132) Wang, Z. U.; Wang, Y.-S.; Pai, P.-J.; Russell, W. K.; Russell, D. H.; Liu, W. R. *Biochemistry* **2012**, *51* (26), 5232–5234.

(133) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371* (4), 818–822.

(134) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15* (11), 1187–1197.

(135) Kaya, E.; Vrabel, M.; Deiml, C.; Prill, S.; Fluxa, V. S.; Carell, T. *Angew. Chemie Int. Ed.* **2012**, *51* (18), 4466–4469.

(136) Plass, T.; Milles, S.; Koehler, C.; Szymański, J.; Mueller, R.; Wießler, M.; Schultz, C.; Lemke, E. A. *Angew. Chemie Int. Ed.* **2012**, *51* (17), 4166–4170.

(137) Wang, Y.-S.; Fang, X.; Chen, H.-Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R. *ACS Chem. Biol.* **2013**, *8* (2), 405–415.

(138) Xiao, H.; Peters, F. B.; Yang, P.-Y.; Reed, S.; Chittuluru, J. R.; Schultz, P. G. *ACS Chem. Biol.* **2014**, *9* (5), 1092–1096.

(139) Schmidt, M. J.; Borbas, J.; Drescher, M.; Summerer, D. *J. Am. Chem. Soc.* **2014**, *136* (4), 1238–1241.

(140) Niu, W.; Schultz, P. G.; Guo, J. *ACS Chem. Biol.* **2013**, *8* (7), 1640–1645.

(141) Wang, K.; Sachdeva, A.; Cox, D. J.; Wilf, N. W.; Lang, K.; Wallace, S.; Mehl, R. A.; Chin, J. W. *Nat. Chem.* **2014**, *6* (5), 393–403.

(142) Hoesl, M. G.; Budisa, N. *Angew. Chem. Int. Ed. Engl.* **2011**, *50* (13), 2896–2902.

(143) Mühlberg, M.; Hoesl, M. G.; Kuehne, C.; Dernedde, J.; Budisa, N.; Hackenberger, C. P. R. *Beilstein J. Org. Chem* **2015**, *11*, 784–791.

(144) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. *Nat. Chem. Biol.* **2011**, *7* (11), 779–786.

(145) Koniev, O.; Wagner, A. *Chem. Soc. Rev.* **2015**, *44* (15), 5495–5551.

(146) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, 4740.

(147) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. *Bioconjug. Chem.* **2010**, *21* (10), 1912–1916.

(148) Cal, P. M. S. D.; Bernardes, G. J. L.; Gois, P. M. P. *Biomacromolecules* **2015**, *115* (5), 2904–2910.

(149) Li, C.-J.; Trost, B. M. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (36), 13197–13202.

(150) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. *Chem. Sci.* **2011**, *2* (9), 1666.

(151) Nathani, R.; Moody, P.; Smith, M. E. B.; Fitzmaurice, R. J.; Caddick, S. *ChemBioChem* **2012**, *13* (9), 1283–1285.

(152) Gunnoo, S. B.; Finney, H. M.; Baker, T. S.; Lawson, A. D.; Anthony, D. C.; Davis, B. G. *Nat. Commun.* **2014**, *5*, 4388.

(153) Ellman, G. L. *Arch. Biochem. Biophys.* **1959**, *82* (1), 70–77.

(154) Kim, J. R.; Yoon, H. W.; Kwon, K. S.; Lee, S. R.; Rhee, S. G. *Anal. Biochem.* **2000**, *283* (2), 214–221.

(155) Macmillan, D.; Bill, R. M.; Sage, K. A.; Fern, D.; Flitsch, S. L. *Chem. Biol.* **2001**, *8* (2), 133–145.

(156) Swanwick, R. S.; Daines, A. M.; Tey, L. H.; Flitsch, S. L.; Allemann, R. K. *ChemBioChem* **2005**, *6* (8), 1338–1340.

(157) Zhang, Y.; Bhatt, V. S.; Sun, G.; Wang, P. G.; Palmer, A. F. *Bioconjug. Chem.* **2008**, *19* (11), 2221–2230.

(158) Ishii, Y.; Lehrer, S. S. *Biophys. J.* **1986**, *50* (1), 75–80.

(159) Wu, C. W.; Yarbrough, L. R. *Biochemistry* **1976**, *15* (13), 2863–2868.

(160) Fontaine, S. D.; Reid, R.; Robinson, L.; Ashley, G. W.; Santi, D. V. *Bioconjug. Chem.* **2015**, *26* (1), 145–152.

(161) Shen, B.-Q.; Xu, K.; Liu, L.; Raab, H.; Bhakta, S.; Kenrick, M.; Parsons-Reponte, K. L.; Tien, J.; Yu, S.-F.; Mai, E.; Li, D.; Tibbitts, J.; Baudys, J.; Saad, O. M.; Scales, S. J.; McDonald, P. J.; Hass, P. E.; Eigenbrot, C.; Nguyen, T.; Solis, W. a; Fuji, R. N.; Flagella, K. M.; Patel, D.; Spencer, S. D.; Khawli, L. a; Ebens, A.; Wong, W. L.; Vandlen, R.; Kaur, S.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Junutula, J. R. *Nat. Biotechnol.* **2012**, *30* (2), 184–189.

(162) Tedaldi, L. M.; Smith, M. E. B.; Nathani, R. I.; Baker, J. R. *Chem. Commun. (Camb).* **2009**, *67* (25), 6583–6585.

(163) Morales-Sanfrutos, J.; Lopez-Jaramillo, F. J.; Hernandez-Mateo, F.; Santoyo-Gonzalez, F. *J. Org. Chem.* **2010**, *75* (12), 4039–4047.

(164) Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G. **2009**, 7798–7802.

(165) Conte, M. Lo; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A. *Chem. Commun.* **2011**, *47* (39), 11086.

(166) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130* (15), 5052–5053.

(167) Chan, A. O.-Y.; Tsai, J. L.-L.; Lo, V. K.-Y.; Li, G.-L.; Wong, M.-K.; Che, C.-M. *Chem. Commun.* **2013**, *49* (14), 1428.

(168) Jones, M. W.; Strickland, R. A.; Schumacher, F. F.; Caddick, S.; Baker, J. R.; Gibson, M. I.; Haddleton, D. M. *J. Am. Chem. Soc.* **2012**, *134* (3), 1847–1852.

(169) Tucker, M. J.; Courter, J. R.; Chen, J.; Atasoylu, O.; Smith, A. B.; Hochstrasser, R. M. *Angew. Chemie - Int. Ed.* **2010**, *49* (21), 3612–3616.

(170) Wilson, P.; Anastasaki, A.; Owen, M. R.; Kempe, K.; Haddleton, D. M.; Mann, S. K.; Johnston, A. P. R.; Quinn, J. F.; Whittaker, M. R.; Hogg, P. J.; Davis, T. P. *J. Am. Chem. Soc.* **2015**, *137* (12), 4215–4222.

(171) Maruani, A.; Smith, M. E. B.; Miranda, E.; Chester, K. A.; Chudasama, V.; Caddick, S. *Nat. Commun.* **2015**, *6*, 6645.

(172) Kalkhof, S.; Sinz, A. *Anal. Bioanal. Chem.* **2008**, *392* (1–2), 1–8.

(173) Smith, S.; Miller, T. *Anal. Biochem.* **1991**, *253*, 247–253.

(174) Smith, G. P. *Bioconjug. Chem.* **2006**, *17* (2), 501–506.

(175) Norman, R.; Creech, J.; Norman, R. *J. Am. Chem. Soc* **1940**, *3396*, 1670–1673.

(176) Landel, A. M. *Anal. Biochem.* **1976**, *73* (2), 280–289.

(177) Butcher, E. C.; Scollay, R. G.; Weissman, I. L. *J. Immunol. Methods* **1980**, *37* (2), 109–121.

(178) Sutton, B. D. A.; Drewes, S. E.; Welz, U. *Biochem. J.* **1972**, *130*, 589–595.

(179) Lin, J.-K.; Chang, J.-Y. *Anal. Chem.* **1975**, *47* (9), 1634–1638.

(180) Traut, R. R.; Bollen, a; Sun, T. T.; Hershey, J. W.; Sundberg, J.; Pierce, L. R. *Biochemistry* **1973**, *12* (17), 3266–3273.

(181) Gavrilyuk, J. I.; Wuellner, U.; Barbas, C. F. *Bioorganic Med. Chem. Lett.* **2009**, *19* (5), 1421–1424.

(182) Gavrilyuk, J. I.; Wuellner, U.; Salahuddin, S.; Goswami, R. K.; Sinha, S. C.; Barbas, C. F. *Bioorganic Med. Chem. Lett.* **2009**, *19* (14), 3716–3720.

(183) Abuchowski, A.; van Es, T.; Palczuk, N. C.; Davis, F. F. *Journal* **1977**, *252* (11), 3578–3581.

(184) McFarland, J. M.; Francis, M. B. *J. Am. Chem. Soc.* **2005**, *127* (39), 13490–13491.

(185) Schwartz, B. A.; Gray, G. R. *Arch. Biochem. Biophys.* **1977**, *181* (2), 542.

(186) Lee, R. T.; Lee, Y. C. *Biochemistry* **1980**, *19* (1), 156–163.

(187) Wurm, F.; Steinbach, T.; Klok, H. A. *Chem. Commun. (Camb).* **2013**, *49* (71), 7815–7817.

(188) Li, X.; Ma, H.; Nie, L.; Sun, M.; Xiong, S. *Anal. Chim. Acta* **2004**, *515* (2), 255–260.

(189) Zamora, R.; Alaiz, M.; Hidalgo, F. J. *Chem Res Toxicol* **1999**, *5*, 654–660.

(190) Riordan, J. F. *Biochemistry* **1973**, *12* (20), 3915–3923.

(191) Gauthier, M. A.; Klok, H. A. *Biomacromolecules* **2011**, *12* (2), 482–493.

(192) Koshi, Y.; Nakata, E.; Miyagawa, M.; Tsukiji, S.; Ogawa, T.; Hamachi, I. *J. Am. Chem. Soc.* **2008**, *130* (1), 245–251.

(193) Joshi, N. S.; Whitaker, L. R.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126* (49), 15942–15943.

(194) Tilley, S. D.; Francis, M. B. *J. Am. Chem. Soc.* **2006**, *128* (4), 1080–1081.

(195) Antos, J. M.; Francis, M. B. *Curr. Opin. Chem. Biol.* **2006**, *10* (3), 253–262.

(196) Foettinger, A.; Melmer, M.; Leitner, A.; Lindner, W. *Bioconjug. Chem.* **2007**, *18* (5), 1678–1683.

(197) Antos, J. M.; McFarland, J. M.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2009**, *131* (17), 6301–6308.

(198) Antos, J. M.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126* (33), 10256–10257.

(199) Gilmore, J. M.; Scheck, R. a; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem. Int. Ed. Engl.* **2006**, *45* (32), 5307–5311.

(200) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C. H.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2013**, *135* (45), 17223–17229.

(201) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol.* **2015**, *11* (5), 326–331.

(202) Zhang, M.; Zhang, X.; Li, J.; Guo, Q. *Chin. J. Chem.* **2011**, *29*, 1715–1720.

(203) Thomas, G. L.; Payne, R. J. *Chem. Commun. (Camb).* **2009**, No. 28, 4260–4262.

(204) Geoghegan, K. F.; Stroh, J. G. *Bioconjug. Chem.* **1992**, *3* (2), 138–146.

(205) El-Mahdi, O.; Melnyk, O. *Bioconjug. Chem.* **2013**, *24* (5), 735–765.

(206) Liu, C. F.; Tam, J. P. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (14), 6584–6588.

(207) Malins, L. R.; Cergol, K. M.; Payne, R. J. *Chem. Sci.* **2014**, *5* (1), 260.

(208) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett.* **2000**, *41* (21), 4069–4073.

(209) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, *136* (27), 9572–9579.

(210) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (18), 10068–10073.

(211) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128* (20), 6640–6646.

(212) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science (80-.).* **1994**, *266* (5186), 776–779.

(213) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (June), 6705–6710.

(214) Townsend, S. D.; Tan, Z.; Dong, S.; Shang, S.; Brailsford, J. a; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134* (8), 3912–3916.

(215) Wang, P.; Dong, S.; Shieh, J.-H.; Peguero, E.; Hendrickson, R.; Moore, M. a S.; Danishefsky, S. J. *Science* **2013**, *342* (6164), 1357–1360.

(216) Yang, M.; Li, J.; Chen, P. R. *Chem. Soc. Rev.* **2014**.

(217) Spicer, C. D.; Triemer, T.; Davis, B. G. *J. Am. Chem. Soc.* **2012**, *134* (2), 800–803.

(218) Miyaura, N.; Suzuki, A. *Chemistry* **1995**, *95* (1), 2457–2483.

(219) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. *J. Am. Chem. Soc.* **2009**, *131* (45), 16346–16347.

(220) Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135*, 13612–13615.

(221) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2013**, *49* (27), 2747–2749.

(222) Lin, Y. a; Chalker, J. M.; Davis, B. G. *Chembiochem* **2009**, *10* (6), 959–969.

(223) Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135* (33), 12156–12159.

(224) Zhang, L.; Chen, X.; Xue, P.; Sun, H. H. Y.; Williams, I. D.; Sharpless, K. B.; Fokin, V. V.; Jia, G. *J. Am. Chem. Soc.* **2005**, *127* (46), 15998–15999.

(225) Meldal, M.; Tornoe, C. W. *Chem. Rev.* **2008**, *108* (8), 2952–3015.

(226) Jin, L.; Tolentino, D. R.; Melaimi, M.; Bertrand, G. *Sci. Adv.* **2015**, *1* (5), e1500304–e1500304.

(227) Jewett, J. C.; Bertozzi, C. R. *Chem. Soc. Rev.* **2010**, *39* (4), 1272.

(228) Kennedy, D. C.; McKay, C. S.; Legault, M. C. B.; Danielson, D. C.; Blake, J. A.; Pegoraro, A. F.; Stolow, A.; Mester, Z.; Pezacki, J. P. *J. Am. Chem. Soc.* **2011**, *133* (44), 17993–18001.

(229) Borrmann, A.; van Hest, J. *Chem. Sci.* **2014**, *5* (6), 2123-2134.

(230) Chin, J. *Chem. Commun.* **2013**, *49* (94), 11007.

(231) Angata, T.; Varki, A. *Chem. Rev.* **2002**, *102* (2), 439–469.

(232) Ghazarian, H.; Idoni, B.; Oppenheimer, S. B. *Acta Histochem.* **2011**, *113* (3), 236–247.

(233) Jandus, C.; Simon, H.-U.; von Gunten, S. *Biochem. Pharmacol.* **2011**, *82* (4), 323–332.

(234) Angata, T.; Nycholat, C. M.; Macauley, M. S. *Trends Pharmacol. Sci.* **2015**, *36* (10), 645–660.

(235) Chen, W. C.; Completo, G. C.; Sigal, D. S.; Crocker, P. R.; Saven, A.; Paulson, J. C. *Blood* **2010**, *115*, 4778–4786.

(236) Zaccai, N. R.; Maenaka, K.; Maenaka, T.; Crocker, P. R.; Brossmer, R.; Kelm, S.; Jones, E. Y. *Structure* **2003**, *11* (5), 557–567.

(237) Zaccai, N. R.; May, A. P.; Robinson, R. C.; Burtnick, L. D.; Crocker, P. R.; Brossmer, R.; Kelm, S.; Jones, E. Y. *J. Mol. Biol.* **2007**, *365* (5), 1469–1479.

(238) Rillahan, C. D.; Schwartz, E.; Rademacher, C.; McBride, R.; Rangarajan, J.; Fokin, V. V.; Paulson, J. C. *ACS Chem. Biol.* **2013**, *8* (7), 1417–1422.

(239) Attrill, H.; Takazawa, H.; Witt, S.; Kelm, S.; Isecke, R.; Brossmer, R.; Ando, T.; Ishida, H.; Kiso, M.; Crocker, P. R.; van Aalten, D. M. F. *Biochem. J.* **2006**, *397* (2).

(240) Schwardt, O.; Gäthje, H.; Vedani, A.; Mesch, S.; Gao, G.-P.; Spreafico, M.; von Orelli, J.; Kelm, S.; Ernst, B. *J. Med. Chem.* **2009**, *52* (4), 989–1004.

(241) Kelm, S.; Madge, P.; Islam, T.; Bennett, R.; Koliwer-Brandl, H.; Waespy, M.; von Itzstein, M.; Haselhorst, T. *Angew. Chemie Int. Ed.* **2013**, *52* (13), 3616–3620.

(242) Kelm, S.; Gerlach, J.; Brossmer, R.; Danzer, C.-P.; Nitschke, L. *J. Exp. Med.* **2002**, *195* (9).

(243) Mesch, S.; Lemme, K.; Wittwer, M.; Koliwer-Brandl, H.; Schwardt, O.; Kelm, S.; Ernst, B. *ChemMedChem* **2012**, *7* (1), 134–143.

(244) Rillahan, C. D.; Macauley, M. S.; Schwartz, E.; He, Y.; McBride, R.; Arlian, B. M.; Rangarajan, J.; Fokin, V. V.; Paulson, J. C. *Chem. Sci.* **2014**, *5* (6), 2398–2406.

(245) Rillahan, C. D.; Schwartz, E.; McBride, R.; Fokin, V. V.; Paulson, J. C. *Angew. Chemie - Int. Ed.* **2012**, *51* (44), 11014–11018.

(246) O'Reilly, M. K.; Collins, B. E.; Han, S.; Liao, L.; Rillahan, C.; Kitov, P. I.; Bundle, D. R.; Paulson, J. C. *J. Am. Chem. Soc.* **2008**, *130* (24), 7736–7745.

(247) Collins, B. E.; Blixt, O.; Han, S.; Duong, B.; Li, H.; Nathan, J. K.; Bovin, N.; Paulson, J. C. *J. Immunol.* **2006**, *177* (5), 2994 LP-3003.

(248) Kitov, P. I.; Sadowska, J. M.; Mulvey, G.; Armstrong, G. D.; Ling, H.; Pannu, N. S.; Read, R. J.; Bundle, D. R. *Nature* **2000**, *403*, 669.

(249) Eiton Kaltgrad, ; Mary K. O'Reilly,; Liang Liao, Shoufa Han, ; James C. Paulson, ; M. G. Finn, J. Am. Chem. Soc., **2008**, 130 (14), 4578–4579.

(250) Yang, Z.-Q.; Puffer, E. B.; Pontrello, J. K.; Kiessling, L. L. *Carbohydr. Res.* **2002**, *337* (18), 1605–1613.

(251) Schauer, R. *Curr. Opin. Struct. Biol.* **2009**, *19* (5), 507–514.

(252) Kneuer, C.; Ehrhardt, C.; Radomski, M. W.; Bakowsky, U. *Drug Discov. Today* **2006**, *11* (21), 1034–1040.

(253) Ley, K. *Trends Mol. Med.* **2003**, *9* (6), 263–268.

(254) Rodgers, S. D.; Camphausen, R. T.; Hammer, D. A. *Biophys. J.* **2000**, *79* (2), 694–706.

(255) Meri, S.; Pangburn, M. K. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (10), 3982–3986.

(256) Sauter, N. K.; Hanson, J. E.; Glick, G. D.; Brown, J. H.; Crowther, R. L.; Park, S. J.; Skehel, J. J.; Wiley, D. C. *Biochemistry* **1992**, *31* (40), 9609–9621.

(257) Crocker, P. R.; Varki, a. *Trends Immunol.* **2001**, *22* (6), 337–342.

(258) Crocker, P. R.; Paulson, J. C.; Varki, A. *Nat. Rev. Immunol.* **2007**, *7* (4), 255–266.

(259) Crocker, P. R. *Curr. Opin. Pharmacol.* **2005**, *5* (4), 431–437.

(260) Macauley, M. S.; Crocker, P. R.; Paulson, J. C. *Nat. Rev. Immunol.* **2014**, *14* (10), 653–666.

(261) Cao, H.; Crocker, P. R. *Immunology* **2011**, *132* (1), 18–26.

(262) Angata, T.; Margulies, E. H.; Green, E. D.; Varki, A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (36), 13251–13256.

(263) May, a P.; Robinson, R. C.; Vinson, M.; Crocker, P. R.; Jones, E. Y. *Mol. Cell* **1998**, *1* (5), 719–728.

(264) Alphey, M. S.; Attrill, H.; Crocker, P. R.; van Aalten, D. M. F. *J. Biol. Chem.* **2003**, *278* (5), 3372–3377.

(265) Attrill, H.; Imamura, A.; Sharma, R. S.; Kiso, M.; Crocker, P. R.; van Aalten, D. M. F. *J. Biol. Chem.* **2006**, *281* (43), 32774–32783.

(266) Yamaji, T.; Teranishi, T.; Alphey, M. S.; Crocker, P. R.; Hashimoto, Y. *J. Biol. Chem.* **2002**, *277* (8), 6324–6332.

(267) Bukrinsky, J. T.; St Hilaire, P. M.; Meldal, M.; Crocker, P. R.; Henriksen, A. *Biochim. Biophys. Acta* **2004**, *1702* (2), 173–179.

(268) Collins, B. E.; Blixt, O.; DeSieno, A. R.; Bovin, N.; Marth, J. D.; Paulson, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (16), 6104–6109.

(269) Freeman, S.; Kelm, S.; Barber, E.; Crocker, P. *Blood* **1995**, *85* (8).

(270) Hanasaki, K.; Varki, A.; Powell, L. D. *J. Biol. Chem.* **1995**, *270* (13), 7533–7542.

(271) Razi, N.; Varki, A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (13), 7469–7474.

(272) Nicoll, G.; Avril, T.; Lock, K.; Furukawa, K.; Bovin, N.; Crocker, P. R. *Eur. J. Immunol.* **2003**, *33* (6), 1642–1648.

(273) Ravetch, J. V.; Lanier, L. L. *Science (80-.).* **2000**, *290* (5489).

(274) O'Neill, A. S. G.; van den Berg, T. K.; Mullen, G. E. D. *Immunology* **2013**, *138* (3), 198–207.

(275) Klaas, M.; Crocker, P. R. *Semin. Immunopathol.* **2012**, *34* (3), 353–364.

(276) Jones, C.; Virji, M.; Crocker, P. R. *Mol. Microbiol.* **2003**, *49* (5), 1213–1225.

(277) Delputte, P. L.; Van Gorp, H.; Favoreel, H. W.; Hoebeke, I.; Delrue, I.; Dewerchin, H.; Verdonck, F.; Verhasselt, B.; Cox, E.; Nauwynck, H. J. *PLoS One* **2011**, *6* (2), e16827.

(278) van der Kuyl, A. C.; van den Burg, R.; Zorgdrager, F.; Groot, F.; Berkhout, B.; Cornelissen, M. *PLoS One* **2007**, *2* (2), e257.

(279) Ludewig, B.; Cervantes-Barragan, L. *Nat. Immunol.* **2012**, *13* (1), 13–14.

(280) Saunderson, S. C.; Dunn, A. C.; Crocker, P. R.; Mclellan, A. D. **2014**, *123* (2), 208–216.

(281) Müller, J.; Obermeier, I.; Wöhner, M.; Brandl, C.; Mrotzek, S.; Angermüller, S.; Maity, P. C.; Reth, M.; Nitschke, L. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (30), 12402–12407.

(282) Poe, J. C.; Tedder, T. F. *Trends Immunol.* **2012**, *33* (8), 413–420.

(283) Müller, J.; Nitschke, L. *Nat. Rev. Rheumatol.* **2014**, *10* (7), 422–428.

(284) Blixt, O.; Collins, B. E.; van den Nieuwenhof, I. M.; Crocker, P. R.; Paulson, J. C. *J. Biol. Chem.* **2003**, *278* (33), 31007–31019.

(285) Kameda, Y.; Takahata, M.; Mikuni, S.; Shimizu, T.; Hamano, H.; Angata, T.; Hatakeyama, S.; Kinjo, M.; Iwasaki, N. *Bone* **2015**, *71*, 217–226.

(286) Biedermann, B.; Gil, D.; Bowen, D. T.; Crocker, P. R. *Leuk. Res.* **2007**, *31* (2), 211–220.

(287) Angata, T.; Varki, A. *Glycobiology* **2000**, *10* (4), 431–438.

(288) Hudak, J. E.; Canham, S. M.; Bertozzi, C. R. *Nat. Chem. Biol.* **2014**, *10* (1), 69–75.

(289) Kiwamoto, T.; Kawasaki, N.; Paulson, J. C.; Bochner, B. S. *Pharmacol. Ther.* **2012**, *135* (3), 327–336.

Chapter 2

The expression and characterisation of a pentapeptide repeat unit library.

Contents Page

2.0 Introduction and aims

The aim of this chapter is to express a library of Np276 azidohomoalanine (AHA) containing proteins in high purity and with full biophysical data to support their use as multivalent scaffolds. Furthermore, there should be no post translational modifications perturbing the homogeneity of the samples as this could perturb later bioconjugation or biological testing experiments. As illustrated in Figure 1., published work by others in the Ben Davis group had shown that a high % of a post translational modification, N-terminal gluconylation, is often observed with the Np276 protein.[1]

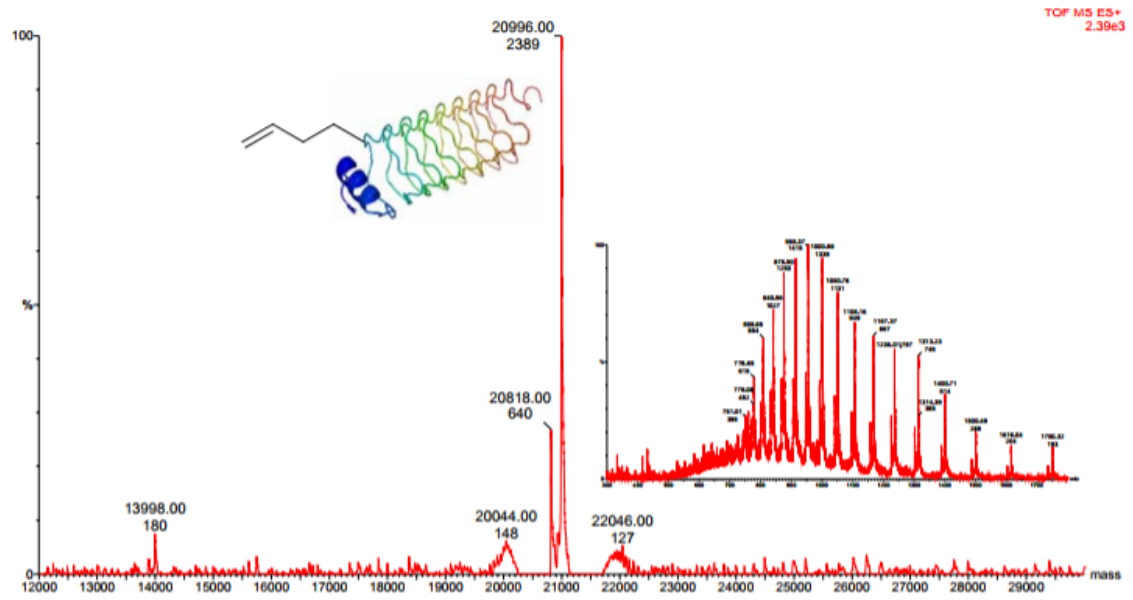


Figure 1. ESI-MS analysis of the expressed Np276 protein incorporated with a UAA showed up to 79% of N-terminal gluconylation. Reproduced with permission from Fernández-González, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. a.; Errey, J. C.; Davis, B. G. Chem. Sci. 2010, 1 (6), 709.

Dr Macarena Sanchez Navarro and Dr James Errey (former group members) had also undertaken expressions of Np276 (unpublished work) on the ruler protein mutants in Table 1. In these unpublished experiments, a high % of N-terminal gluconylation was again observed in the expressed proteins. In addition, no biophysical characterisation was conducted on these

proteins. These proteins all contained a hexa histidine tag at the N-terminal end (pet28d vector) and were expressed with full replacement of Met with AHA. In many of their expressions, degradation of the protein was also observed on SDS-PAGE gels as bands of a lower molecular weight.

Table 1. Table of Np276 mutants prepared by Dr James Errey and Dr Macarena Sanchez Navarro. In the table below the distances refer to the crystal structure distances between the methionine residues on the same face of the protein. Distances were measured using PyMol.

Construct	Mutations	Distance (Crystal Structure)/ Å
5 Å	M21I, K25/94/123A T81M, T201K	**5.2**
10 Å	M21I, K25/94/123A I101M, T201K	**9.9**
15 Å	M21I, K25/94/123A N121M, T201K	**14.7**
20 Å	M21I, K25/94/123A D141M, T201K	**19.6**
20 Å	M21/61I, K25/94/123A D41M, N141M T201K	**19.6**
25 Å	M21/61I, K25/94/123A, D41M, D141M, T201K	**24.6**
30 Å	M21/61I, K25/94/123A, D41M, D161M, T201K	**29.5**
Single	M21I	**N/A**

2.1 Results and Discussion

2.1.1 Theoretical analysis of the Np276 protein and comparison to other proteins of the PRP family.

Ramachandran analysis of the literature crystal structure of Np276 showed that the ψ angles are in the majority positive, indicating a clockwise rotation of the Cα-C bonds, and the majority of φ angles are negative. The areas of the Ramachandran plot correspond to areas consistent with a beta sheet, left-handed alpha helix and a right-handed alpha helix, secondary structures. The Ramachandran analysis of Np276 shows that the crystal structure model is good and thus will act as a good model for further studies.[2–4]

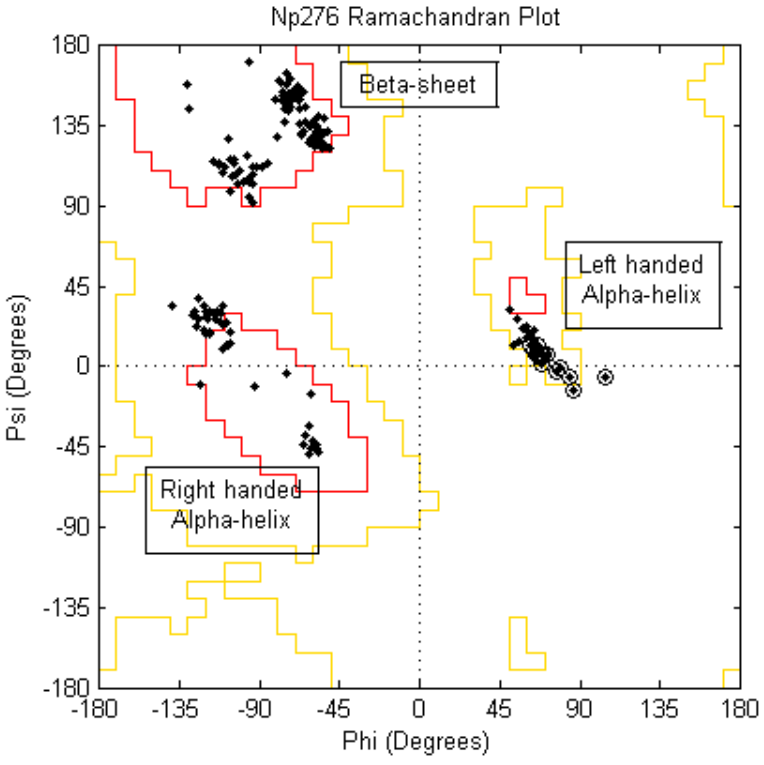


Figure 2. Np276 Ramachandran plot generated from pdb 2j8K. The red regions correspond to core regions (labelled) and the orange regions correspond to allowed regions.

Ramachandran analysis compares very favourably to that of the Rfr32 protein. The Rfr32 protein was previously illustrated in Chapter 1, Figure 6.

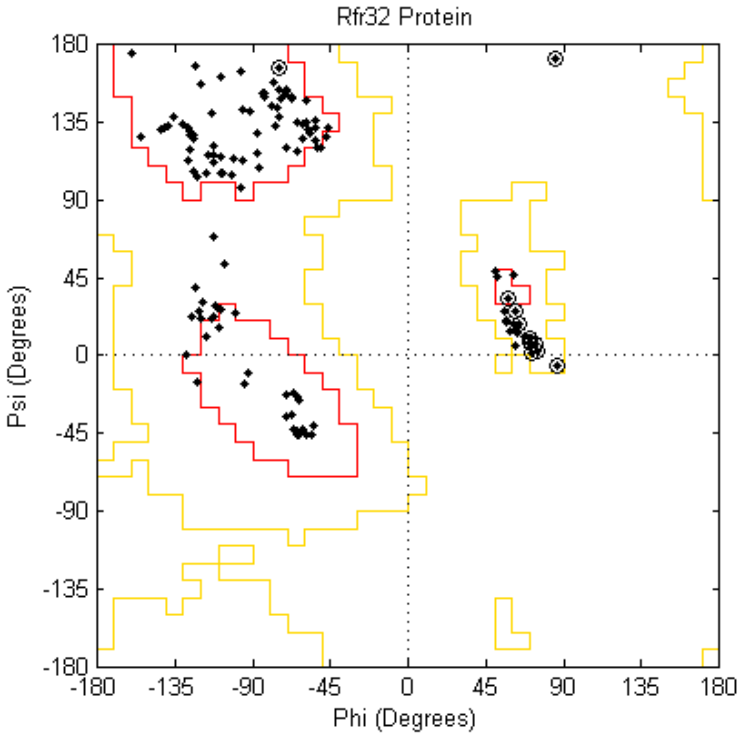


Figure 3. Ramachandran plot of Rfr32. Whilst 76% of residues are in the most favoured regions, it is noted that the *i-2, i-1* and *i+2* residues are all in the same regions, regardless of face.

In comparison to the Np276 protein the Rfr32 protein contains 21 PRP repeat sequences, corresponding to 5 right-handed beta-helical coils for a total length of ~19Å. Like Np276 the coil completes a full turn every 4.8Å or 20 residues; however, unlike Np276, Rfr32 also contains a disulphide linkage. It could be extrapolated from the Rfr32 plot that many of the residues listed as left handed alpha helix in Np276 plot are in fact the *i+2* residues. Importantly, as Rfr32 and Np276 are very similar in structure, the Rfr32 protein could act as a model with which to compare biophysical data.

2.1.2 AHA synthesis

Azidohomoalanine was synthesised according to the protocol described by Tirrell *et al.*[5] Whilst there are other protocols for the synthesis of AHA, the Tirrell protocol was chosen due to its low cost and lack of purification steps (a negative aspect being the inherent dangers of the azide transfer).[6]

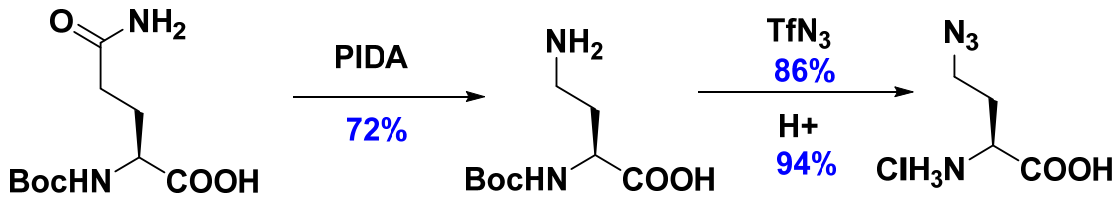


Figure 4. Synthesis of L-azidohomoalanine via a 3 step protocol.[7] This sequence could also be conducted on large scales.

Starting from commercially available N-Boc-L-glutamine, a Hoffman degradation with PIDA ((diacetoxyiodo)benzene) gave the intermediate, 4-amino-2-((tert-butoxycarbonyl)amino)butanoic acid, in 72% yield. Azide transfer with the *in situ* prepared, volatile and toxic reagent, trifyl azide (TfN_3) with a Lewis acid ($CuSO_4$*) gave the corresponding Boc-AHA. Acid mediated deprotection subsequently gave the AHA salt (HCl or TFA). If desired this could then be subjected to a DOWEX-H^+ column to give the pure amino acid as the free base.

* Other divalent first row metal ions such as Ni(II) and Zn(II) have also been shown to be effective. The metal acting as a Lewis Acid.[64]

2.1.3. Initial expression of Np276 constructs & comparisons to previous work.

Expressions of the Np276 proteins were initially conducted as previously outlined utilising the protocol developed in the group and outlined by Vijakrishna *et al*.[8] Whilst complete incorporation of azidohomoalanine was observed by ESI-TOF MS, a deleterious PTM was observed, characterised by a +178 Da adduct.[9] This is the same N-terminal gluconylation modification that was seen by Dr Macarena Sanchez Navarro and Dr James Errey in their expressions of these proteins. The 20Å protein was expressed in selenomet media for 16 hours (a minimal glucose media with all components except methionine that is normally used for SeMet incorporations) with approximately 60% N-terminal gluconylation being observed. In addition, poor protein purity was observed despite nickel affinity and anion exchange purification steps.

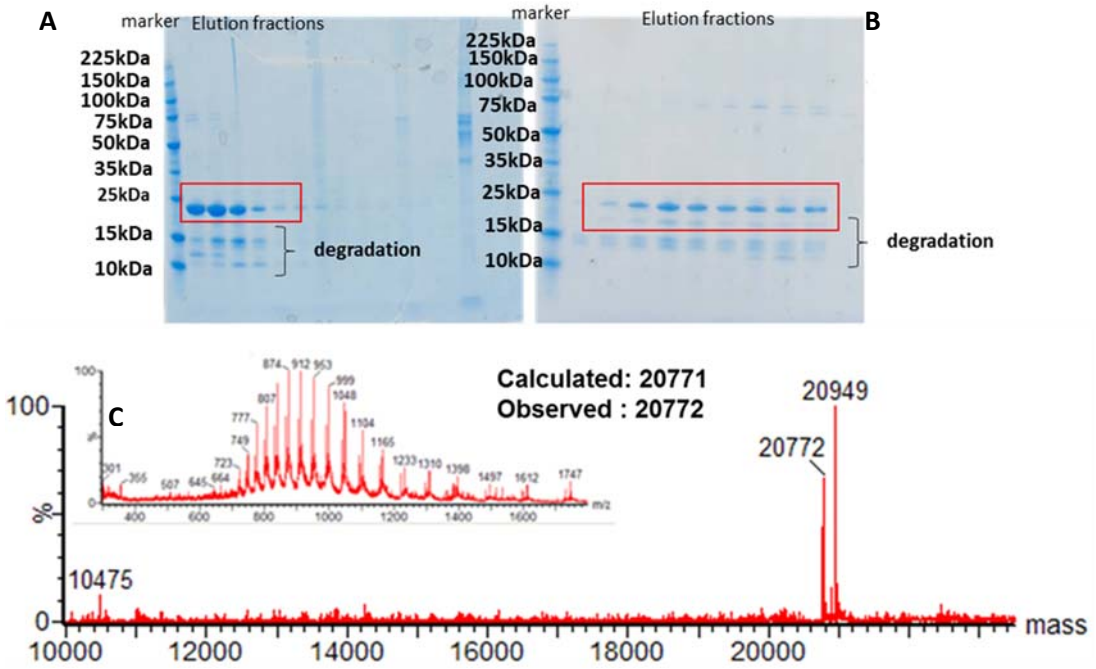


Figure 5. 20Å AHA 61-121 T201K protein expressed with AHA. A) SDS-PAGE analysis of the protein purified on a 5ml nickel HP column B) SDS-PAGE analysis of the fractions purified by Q-AEX column C) LCT-TOF MS analysis of the purified protein, exhibiting a high % (60%) of N-terminal gluconylation.

The 10Å Np276 proteins (Figure 6) was expressed at 30°C for 10 hours. Consequently, although by SDS-PAGE the amount of impurities were decrease, a high proportion of N-terminal gluconylation was observed again.

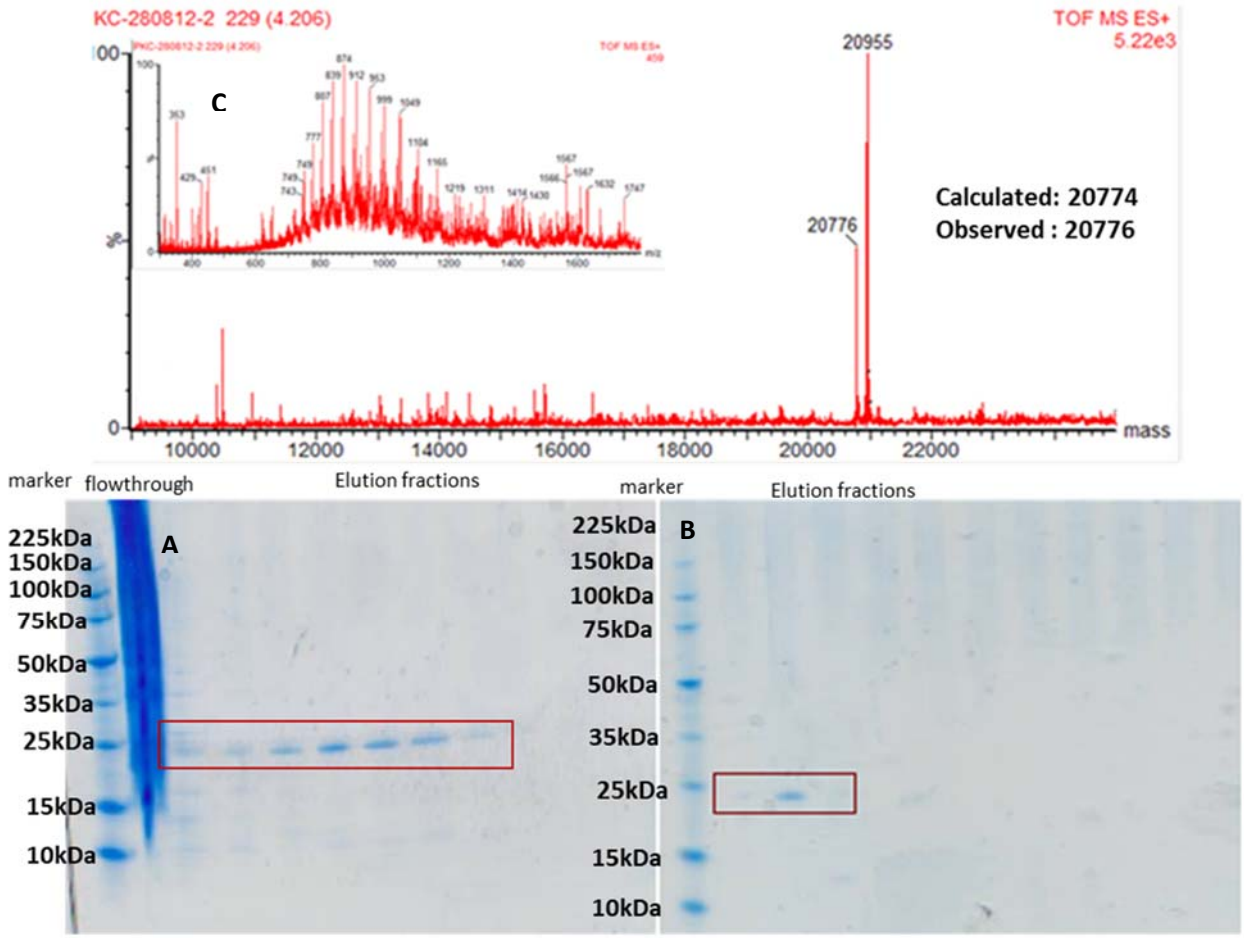


Figure 6. 10Å AHA 61-101 T201K protein expressed with AHA. A) SDS-PAGE analysis of the protein purified on a 5ml nickel HP column B) SDS-PAGE analysis of the fractions purified by Q-AEX column C) ESI-TOF MS analysis of the purified protein, exhibiting a high % of N-terminal gluconylation.

Before expressing a library of ruler proteins it was thus necessary to resolve this issue with the N-terminal gluconylation as this problem has affected both the work of previous group members and the initial attempts at expressing Np276 ruler proteins outlined here.

2.1.4. N-terminal gluconylation

N-terminal gluconylation is a common[†] non-enzymatic post translational modification[‡] that occurs at the N-terminus of (*E. coli* expressed) hexa-his-tagged proteins.[9]

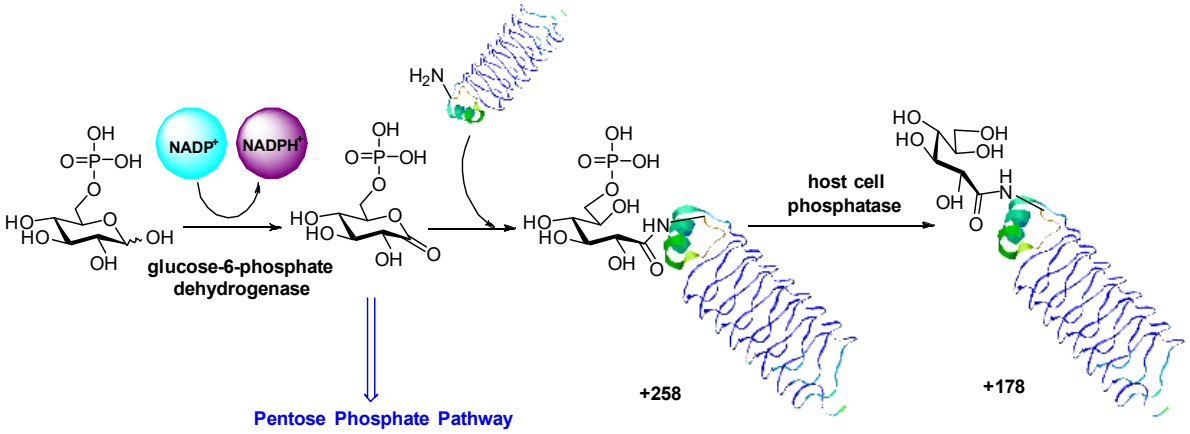


Figure 7. A proposed mechanism for the non-enzymatic post translational modification of Np276. The mechanism is expected to occur via the ring opening of the lactone, followed by subsequent dephosphorylation. Mechanistically the original authors propose a general base or nucleophilic mechanism.

The origins of this modification lie in the build-up of the highly reactive intermediate, glucose-6-phosphate lactone. The first step of the (oxidative) pentose phosphate pathway (PPP), leads to the formation of glucose-6-phosphate which is subsequently dehydrogenated to the reactive lactone.[10] The key enzymes involved in this pathway are G6PDH (encoded by the *zwf* gene) and 6-phosphogluconolactonase (encoded by *pgl*). Furthermore, it has been previously noted that N-terminal gluconylation is most prevalent in B strain *E. coli* (such as the parent B834 used in this SPI work) and is minimised in K strains.[11,12] Indeed in the BL21 strain, ^{13}C NMR

[†] (136 citations since 1999)

flux analysis has indicated that glucose-6-phosphate lactone hydrolysis is the major bottleneck of the PPP.[13–15]

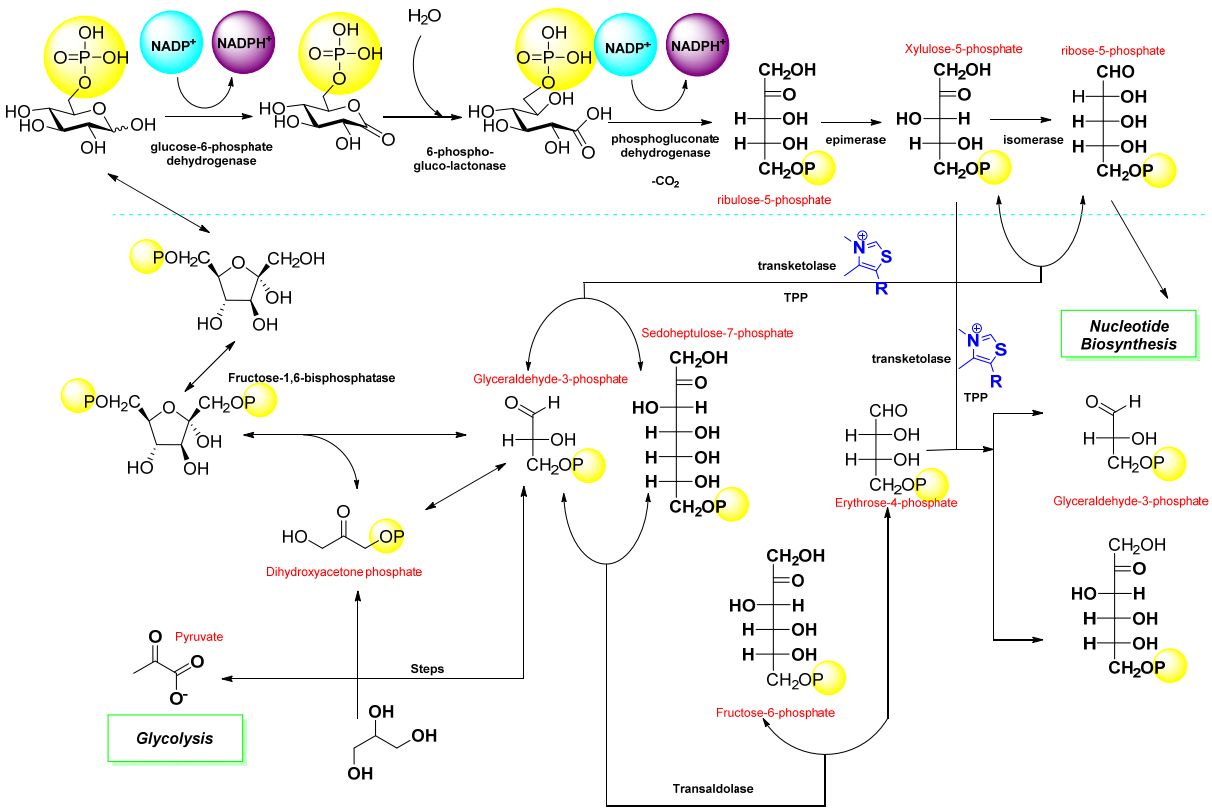


Figure 8. The pentose phosphate pathway. Key intermediates are represented in chair and Fischer projections, with key enzymes labelled.[10]

Preventing N-terminal gluconylation is an important, yet overlooked area of study for the biochemical community: as the presence of this modification has been shown to interfere with the effective crystallisation of proteins for X-ray crystallography, as well as reducing the homogeneity of protein samples.[13] Current approaches to solving the challenge posed by this modification include, via the genetic incorporation of mutations at the his-tag (a G2F) or by the alteration of genes in the PPP of the host *E. coli* strain.[9,13] In this study2 different areas were investigated, (the use of additives and SPI compatible mutations) in order to prevent this deleterious PTM, along with condition optimisation in each case.

2.1.5. Additives

The proposal for this approach was, by using an exogenous source of an additive could the PPP be affected, and hence the levels of the reactive intermediate 6-phosphogluconolactone, thus preventing N-terminal gluconylation? As the dehydrogenation of glucose-6-phosphate to the lactone is the rate determining step of the PPP, this step was initially investigated.[9]

Initially glutathione (GSH) was investigated, a common intracellular species in both eukaryotes and prokaryotes. In *E. coli*, the intracellular concentration depends on levels of free cysteine available, the aeration of the culture and the phase of growth. Roles of glutathione include maintenance of the K^+ pool and protecting against oxidants/electrophiles. More interestingly, given that GSH is in equilibrium with its oxidised disulfide form (GSSG), via the enzymes (glutathione reductase (*gor*)[16] and glutathione peroxidase *btuE*)[17] a process that generates NADPH, could the redox state of the cells be altered and thus affect the PPP? This idea was inspired by biomedical research conducted into G6PDH deficiency in higher organisms.[18]

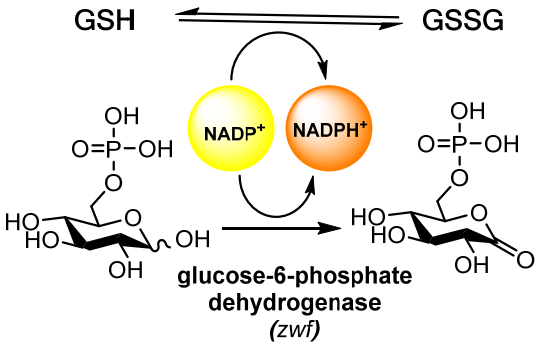


Figure 9. Proposed cycle, transcription rates of *zwf* is linked to both growth rate and superoxide concentration.

Using a 15Å, M21I, K25/94/123A N121M, T201K mutant which had shown a high % of N-terminal gluconylation during the previous azidohomoalanine incorporation experiments the hypothesis was tested with a range of additives, by growing *E. coli* B834 (DE3) in 2XTY

media, before then splitting and supplementing with the required additive at a concentration of 1mM. After inducing with 1mM IPTG, the samples were prepared and analysed simultaneously by a PhysNexus™ system and then XEVO-TOF-MS. The use of the robotic PhysNexus™ purification system controls for manipulation errors between samples, as human error is minimised (Figures 10 and 11). By starting from the same culture which was then split, it also ensured the experiment controlled for subtle variations between expressions.

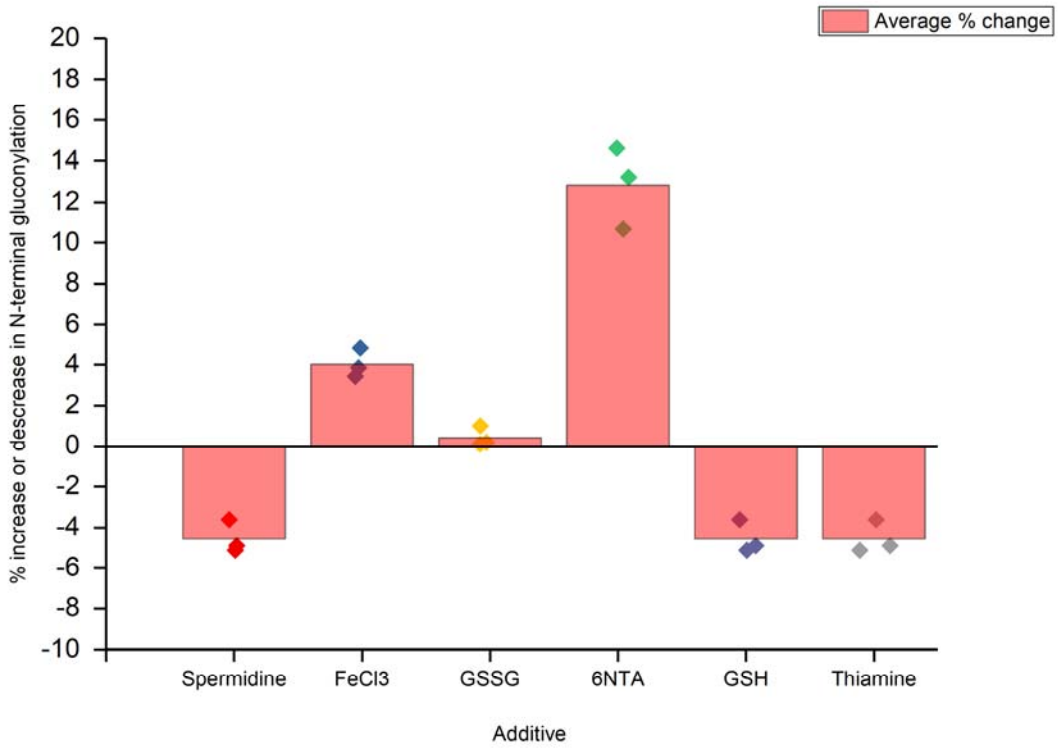


Figure 10. Percentage increase or decrease of N-terminal gluconylation for different additives versus a control sample (no additive). For samples run with spermidine, GSH and thiamine: an equal loss is observed due to the lack of any N-terminal gluconylation in these samples (e.g. the control showed ~4% gluconylation). Reactions were conducted in triplicate and were conducted concurrently using the same culture, flasks, shaking, volume and temperatures.

In the case of spermidine, GSH and thiamine: the N-terminal modification was no longer observed. Thus when compared to the control which showed 4% N-terminal gluconylation, a reduction is observed. Whilst the magnitude of this reduction is capped at 4% (as this was the

% of the modification the control showed), it does indicate that they were retarding the formation of the N-terminally gluconylated product. With $FeCl_3$ and GSSG small increases were seen. Strikingly, with a higher concentration of 6-amino-nicotinamide (6-NTA) a >12% of N-terminal gluconylation was observed, including the observation of the de-phosphorylated phospho-glucolactone. Investigating (not in triplicate) how these additives affected expression yields indicated that spermidine and thiamine led to an increase in expression yields. 6-NTA, along with GSH, GSSSG and $FeCl_3$ all led to a reduction in expression yield.

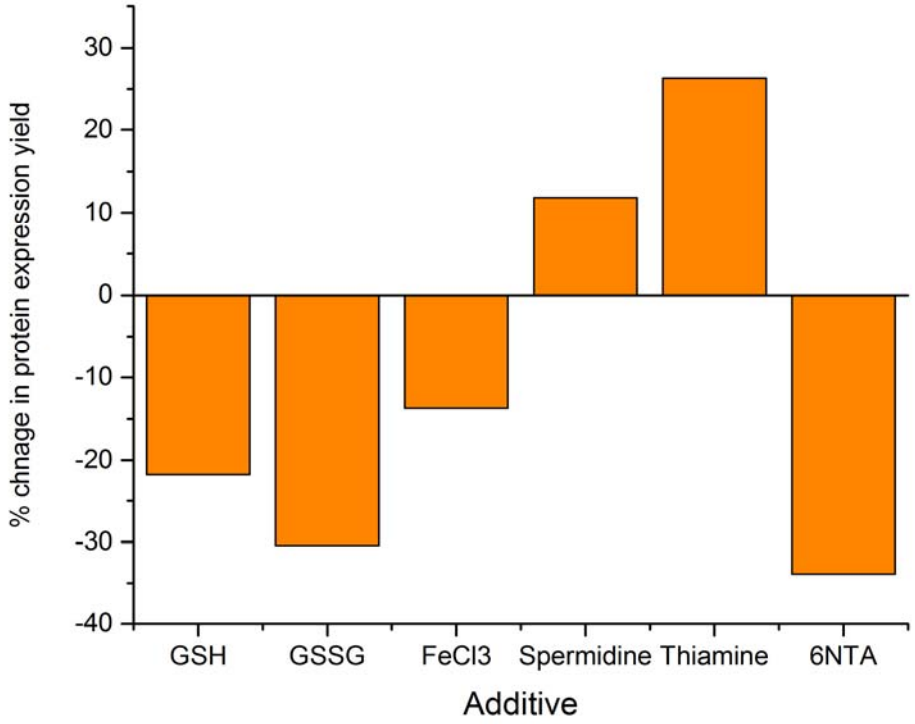


Figure 11. % increase or decrease of protein expression levels for different additives versus a control. Data is based on gel densitometry readings from ImageJ. This data wasn't conducted in triplicate and thus represents the result of the pooled expressions.

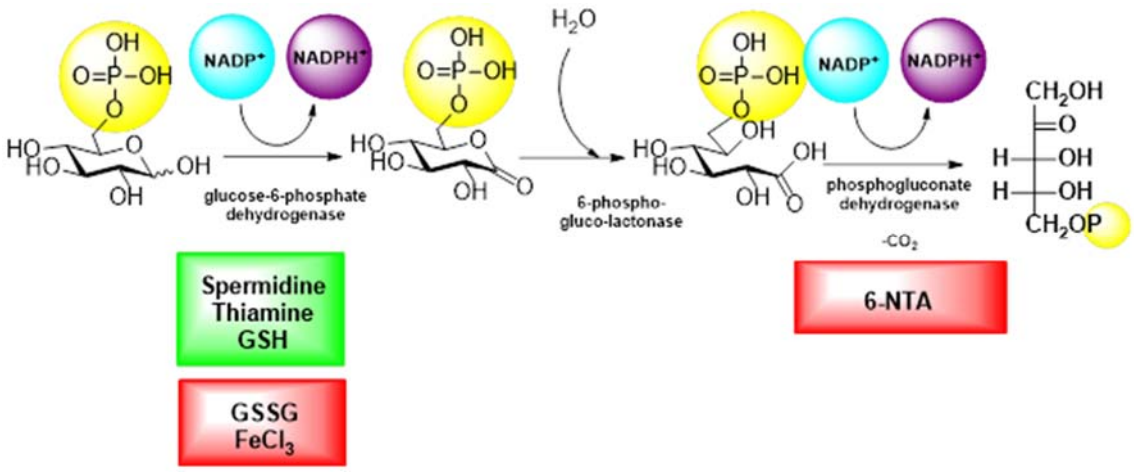


Figure 12. Proposed sites of actions for various additives. These could be on the actual enzymes or via upregulation of genes linked to these additives (i.e. oxidative stress).

The factors that affect the PPP and thus the build-up of the reactive lactone intermediate include: glucose concentration, activity of and level of G6PDH, activity of and level of G6PDH and activity of and level of PGDH. Spermidine and other polyamines (which enter *E.coli* via transport proteins expressed from the *potABCD* operon) have been found to protect *E. coli* cells from oxidative stress.[19,20] Oxidative stress has been found to activate the 2-stage (via SoxR -> SoxS) *SoxRS* regulon, which in turn activates the translation of *zwf*, the gene that encodes for G6PDH. Thus additives which increase the oxidative stress of the cell are likely to increase levels of this enzyme (and thus the reactive lactone); conversely additives which decrease oxidative stress could decrease levels of this enzyme. The result with spermidine (a decrease) and also thiamine support this. Thiamine has also been shown in *E. coli* to protect against oxidative damage to cells, via scavenging of the superoxide radical. In this way, less expression of the G6PDH enzyme is observed, as previously described.[21]

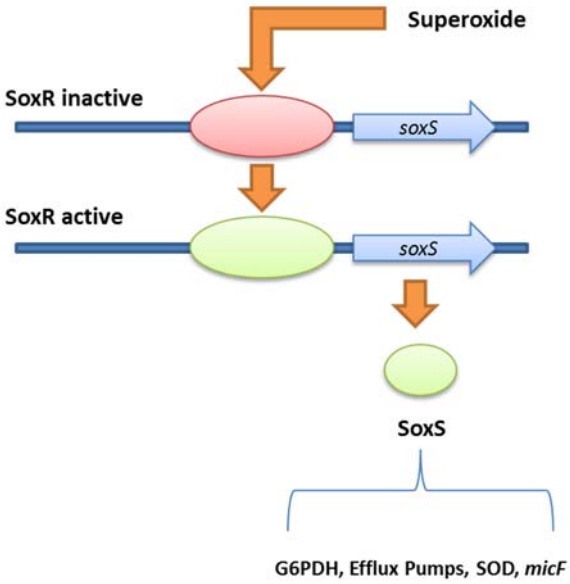


Figure 13. The *SoxRS* regulon for the protection of *E. coli* from oxidative stress. *micF* is a gene that controls expression of ompF. SOD is superoxide dismutase.[22]

Iron chloride, as a heavy metal, also contributes to oxidative stress in the bacterium.[23,24] Heavy metals generate reactive species via processes such as the Fenton reaction and iron responsive proteins such as Fur are also under the control of the *SoxRS* regulon.[23,25–29] Thus with an increased Iron concentration, an increase in N-terminal gluconylation would be expected. 6-aminonicotinamide has been shown to increase levels of the lactone in rats.[30] GSH can also help protect cells against oxidative stress in a mechanism that is perhaps based more on the oxidative state of the cell than NADPH levels, the excess of GSSG would confer little protection against oxidative species and this is corroborated by the negligible change in gluconylation observed.[16,31–34]

2.1.6 Expression conditions for gluconylation prevention

As the PTM is a non-enzymatic process it was essential to investigate the potential effects of temperature. By conducting the expression reactions (with the Np276 15Å, M21I, K25/94/123A N121M, T201K construct) at varying temperatures, a ~10% drop in the levels of modification was observed upon conducting the experiment at 20°C versus 30°C; indicating that lower expression yields would be beneficial.

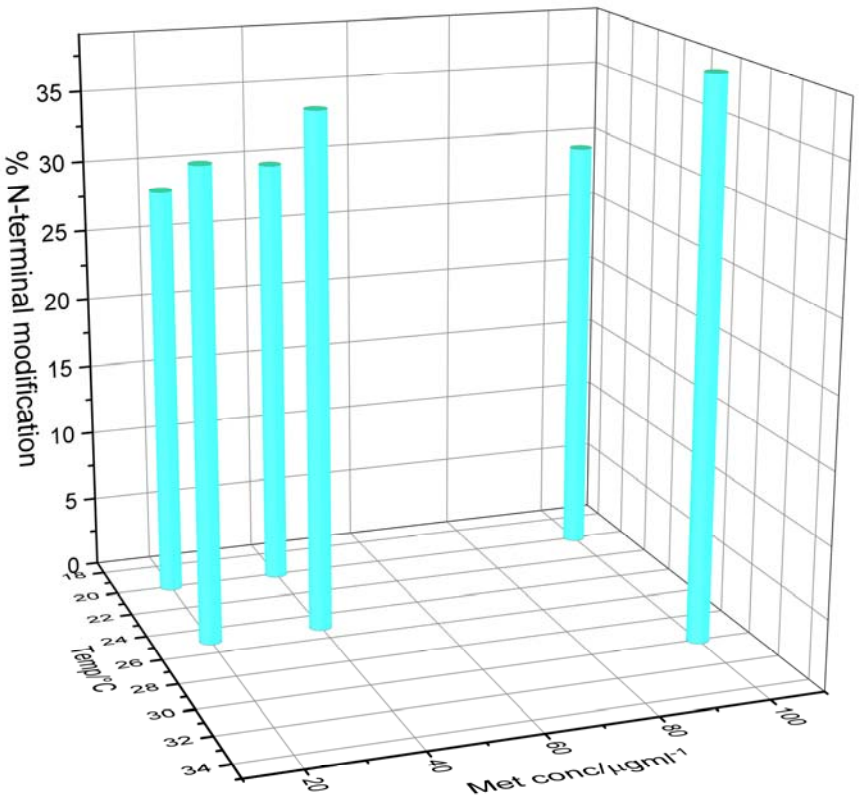


Figure 14. Percentage of N-terminal gluconylation (+178Da) versus temperature. This graph indicates that conducting the expression at a lower temperature would result in less N-terminal modification.

However, upon a subsequent SPI expression of the 10 and 20 Å rulers in selenomet media, significant percentages of the modification were observed despite the expressions being done at 18°C for 15 hours (Figure 15.).

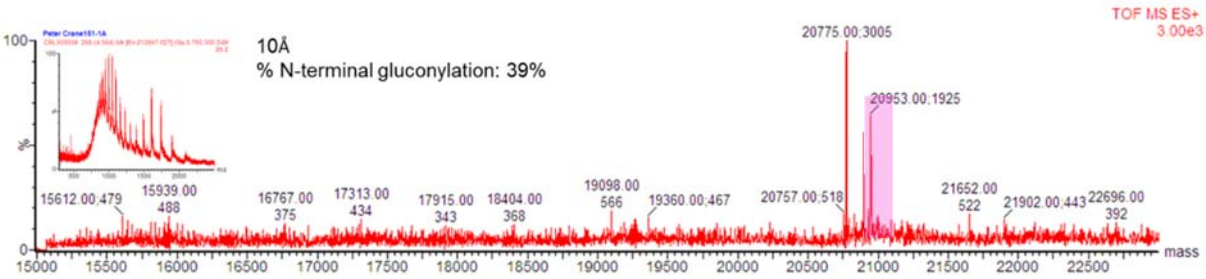


Figure 15. LCT-TOF-MS analysis for the 10Å protein expressed with AHA in selenomet media. The +178 modification is clearly evident and highlighted with a pink box.

Many of the previous experiments had been conducted using LB media; however in many cases higher proportions of the post translational modification were observed (approximately 45%/ND below) when the expression was conducted in selenomet media (Figure 16.).

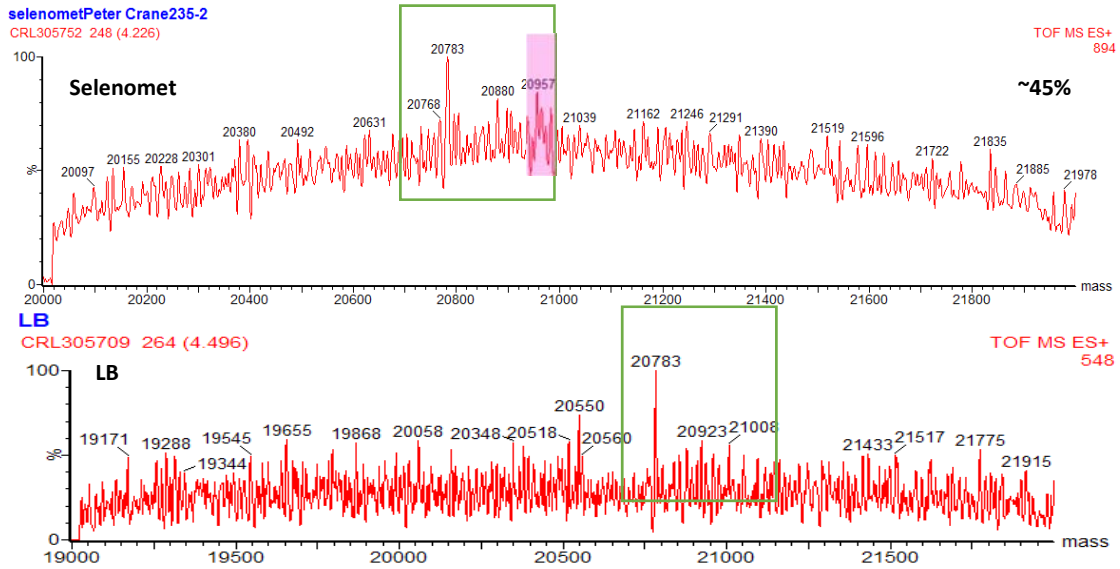


Figure 16. MS analysis of selenomet media expression versus an LB expressions. In the high glucose selenomet media a higher % of N-terminal gluconylation is observed. This is highlighted by a pink box.

LB media is a tryptone/yeast extract (complex nitrogen source) based media, whereas selenomet media is predominately glucose based (glucose-6-phosphate is the major inflow into the PPP). In line with a previous patent report,[35] the use of a complex nitrogen based medium

results in a lower proportion of N-terminal gluconylation. Unfortunately, a complex nitrogen source cannot be used in the SPI method due to the necessity for the removal of all endogenous methionine.

It is acknowledged that the data presented here is preliminary, however it points towards a new way of preventing PTMs especially N-terminal gluconylation. Future work is needed to verify this effect in other proteins, in more replicates, under other expression conditions and to ensure that the proteins remain functional. The use of expression optimisation also reduced the proportion of the gluconylation product but again not completely. More general methods such as mutations were thus trialled.

2.1.7 Mutations to prevent N-terminal gluconylation

The use of a G2F mutation had been previously considered in the literature, but unfortunately due to the bulkiness of this penultimate amino acid, N-terminal methionine excision (NME) was found not to occur.[9,36] This renders it unusable with the selective pressure incorporation method for the expression of proteins (i.e. an additional Met/Met surrogate will be present).

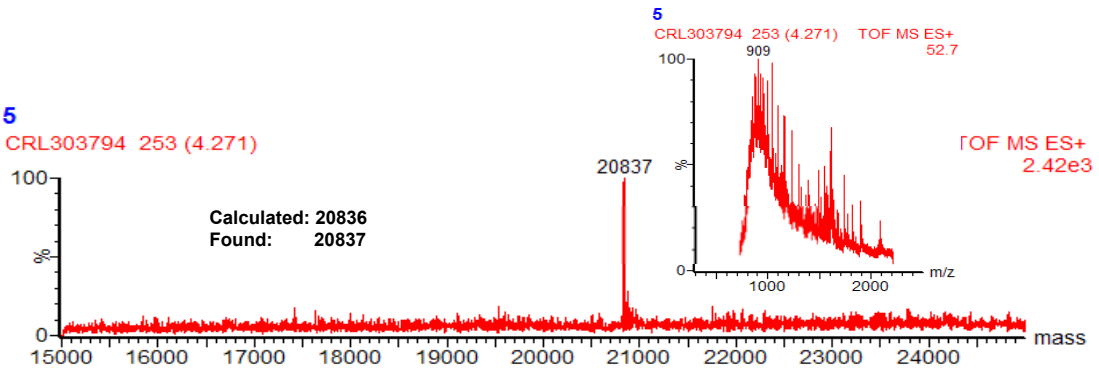
It has been noted by Tirrell *et al*, that N-terminal excision is unlikely to occur when Gln, His, Glu, Phe, Lys, Tyr, Trp, Arg and Met are at the 2 position[§]. Through the work of Budisa and Tirrell, NME Rules have been extended to UAAs including Aha, Hpg (homopropargyl glycine) and trifluoromethionine. Aha is both deformylated (by peptide deformylase) and Met excised efficiently in *E. coli*, conversely Hpg is inefficient in both of these processes. The N-terminal group also has a bearing upon the stability of the protein via the N-end rule, in which the smaller

[§] Even the third and later amino acids can have an effect on NME (interestingly the αC stereochemistry of the 2nd aa has an effect upon EcMetAP).

the terminal amino acid the more stable the subsequent protein,[37] this rule is different for proline in which the rule is specific for each protein.[37–39]

The effect of amino acids at the 2 position upon gluconylation was investigated in the original publication.[9] It was found that the modification was specific for the his-tag and that with Met-Phe-His6 and Met-Pro-His6 the modification was prevented. Out of these, proline would be the most suitable for the his-tag mutation as NME would still occur. The mutation here however differs from the original paper in that in this case the tag consists of a MXSSHHHHHH tag whereas in the original publication it was an MXHHHHHH tag. In order to introduce the G2P mutation, two successive rounds of mutagenesis were required; the first to give the CCC proline codon. This codon however, represents a rare codon** with the *E. coli* transcriptome and thus will retard expression. A subsequent round of mutagenesis gave the G2P with the CCA proline codon.

The expression of the 5 and 10Å ruler proteins was conducted at 30°C for 16 hours in LB media. The results from this expression indicated that no N-terminal gluconylation was observed; in corroboration with results previously obtained in this media (LB media leads to less modification than the glucose rich selenomet media).



** Rare codons are codons, which have a low amount of associated tRNA in *E. coli*. Rare codons can cause termination or retardation of translation and will severely affect the expression yields of expressed proteins.

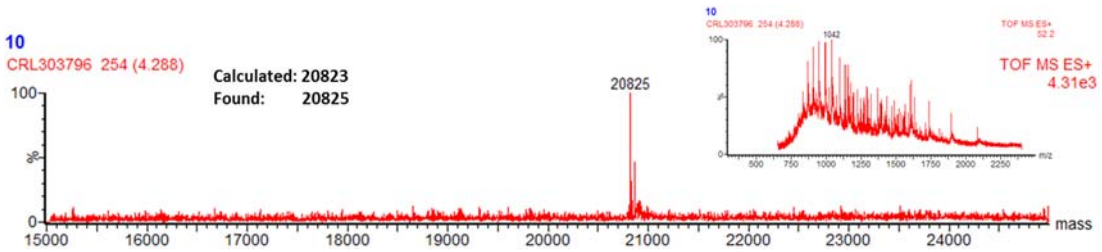


Figure 17. ESI-TOF-MS for the expressed (Met containing) Np276 5 and 10Å, G2P proteins. No N-terminal gluconylation was observed when the G2P mutation was used along with LB media.

2.1.8 Expression of AHA containing Np276 proteins

Following on from these Met/LB trials, the protein was also expressed in selenomet media using the SPI method to incorporate azidohomoalanine. Expression of AHA containing 5Å and 10Å ruler proteins utilising a lower temperature (18°C) led to less modification; however, traces were still observed (Figure 18).

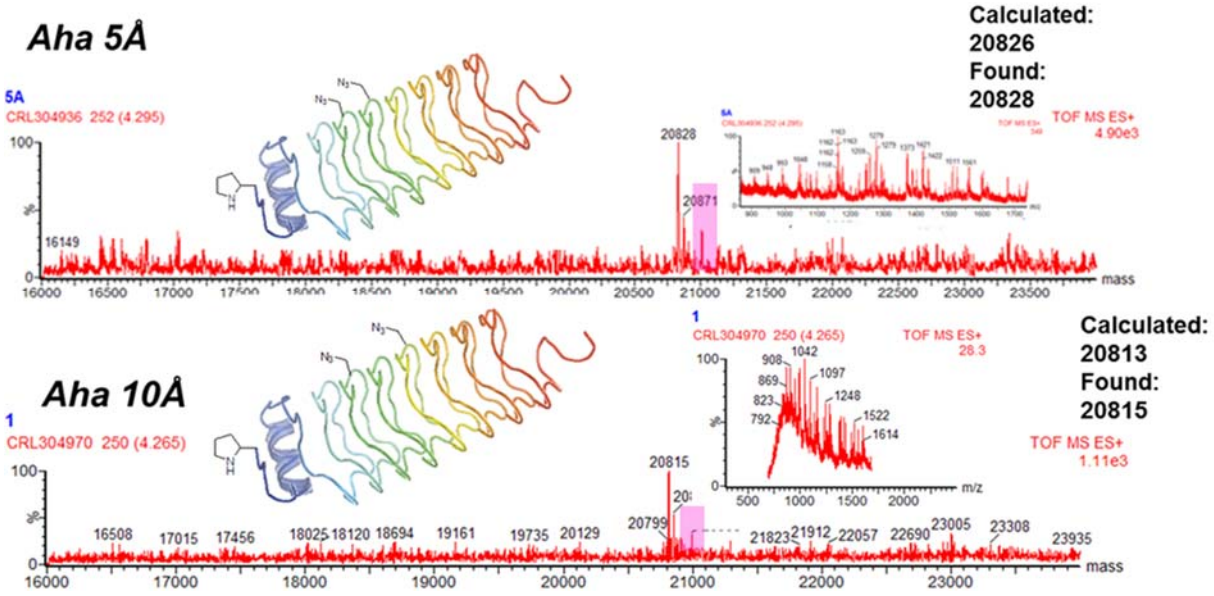


Figure 18. LCT-TOF-MS for the expressed (AHA containing) Np276 5 and 10Å, G2P mutants. In both cases small peaks corresponding to the presence of N-terminal gluconylation can be observed and are highlighted in pink.

Expression of an AHA ruler, 5Å, utilising a lower temperature (18°C) and an increase in media concentration (double nutrient mix) was also then undertaken to investigate whether increasing the glucose concentration with the mutation would fully help prevent the N-terminal modification. In this case, none of the modification was observed when compared to previous AHA expressions. A 15Å ruler protein was also expressed but without the G2P mutation to ensure that this suppression was as a direct result of the mutation. This expression gave a protein in which the phosphorylated N-terminal gluconyl protein modification was also observed in addition to the +178 modification.

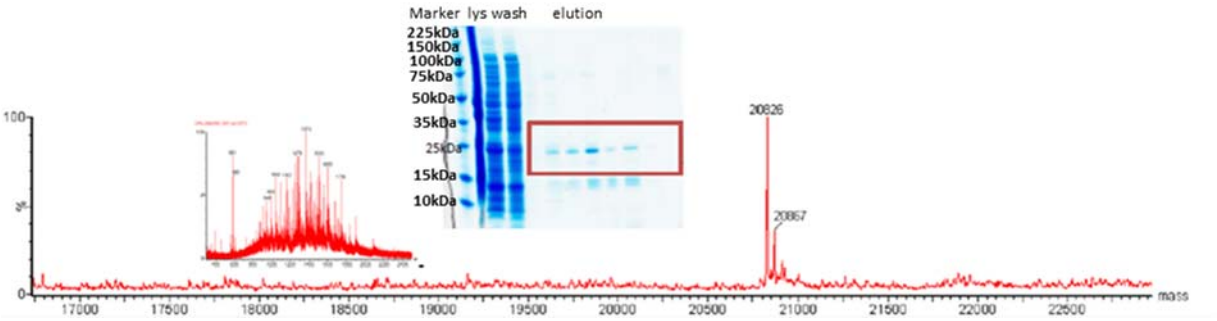


Figure 19. ESI-TOF-MS for the expressed (AHA containing) Np276 5 Å, G2P mutant. What can be observed is no trace of the +178Da modification.

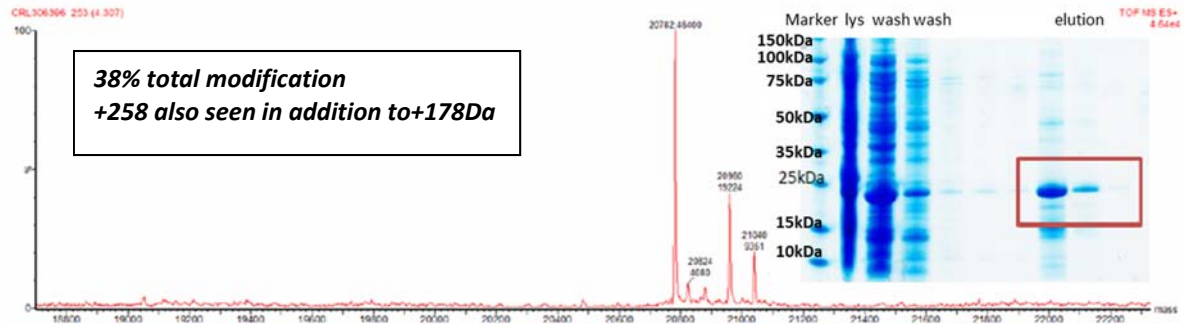


Figure 20. ESI-TOF-MS for the expressed (AHA containing) Np276 15 Å protein. What can be observed is the presence of both the phosphorylated and dephosphorylated gluconyl adducts.

A summary of all the prepared constructs is shown in Table 2. In conclusion, two methods were developed for the prevention of N-terminal gluconylation: an additive method and a

supplementation/mutational method. As the work on additives is only preliminary and further work is needed to expand its scope and use, the mutational approach was used due to its compatibility with SPI.

Table 2. Table of novel Np276 mutants prepared containing the G2P mutation, which along with the conditions previously outlined allowed for gluconylation free expression. The distance of the Met/AHA spacing remains unchanged as does the presence of the N-terminal histag.

Construct	Mutations	Distance (Crystal Structure)/ Å
5 Å	G2P, M21I, K25/94/123A T81M, T201K	**5.2**
10 Å	G2P, M21I, K25/94/123A I101M, T201K	**9.9**
15 Å	G2P, M21I, K25/94/123A N121M, T201K	**14.7**
20 Å	G2P, M21I, K25/94/123A D141M, T201K	**19.6**
20 Å	G2P, M21/61I, K25/94/123A D41M, N141M T201K	**19.6**
25 Å	G2P, M21/61I, K25/94/123A, D41M, D141M, T201K	**24.6**
30 Å	G2P, M21/61I, K25/94/123A, D41M, D161M, T201K	**29.5**
Single	G2P, M21I	**N/A**

2.1.9 Protein stability, characterisation and DSF assays

Protein stability is strongly related to the Gibbs free energy of unfolding, which in turn is a temperature dependant function. As the temperature of a solution increases, the proportion of unfolded and folded protein increases until at the melting temperature (T_M) 50% is unfolded and 50% is folded (i.e. $\Delta G = 0$).[40]

$$\Delta G_{unfolding} = -RTLnK_{eq}$$

$$K_{eq} = \frac{k_{unfolded}}{k_{folded}} = \frac{[Unfolded]}{[Folded]}$$

Figure 21. Thermodynamics of protein unfolding. The equilibria concentration of the unfolded and folded states can be expressed in terms of rate constants (for the equilibria) or the concentrations of species in solution. Whilst protein unfolding isn't always a reversible phenomenon, this model offers a basic framework.

Common ways of ascertaining this melting temperature include by differential scanning calorimetry (DSC),[41] differential static light scattering (DSLS)[42] CD or by a ThermoFluor assay. The ThermoFluor assay, also known as differential scanning fluorimetry (DSF), measures the thermal unfolding of a protein in the presence of a hydrophobic dye.[43] The dye used, SYPRO Orange, exhibits a high fluorescence in hydrophobic environments but is quenched in aqueous solutions. Upon binding a hydrophobic region, SYPRO Orange experiences a significant increase in quantum yield. Despite this, it's spectral properties (λex 470 nm /λem 570 nm) are compatible with real-time PCR instruments allowing the unfolding of the protein to be monitored by fluorescence emission.

In the experiments described herein, SYPRO Orange is used due to its high signal to noise ratio in a protocol outlined by Niesen *et al*.[44,45] By changing buffers, the effect of the buffer on the T_M can be determined and thus by extension the protein stability. Once the data is collected by

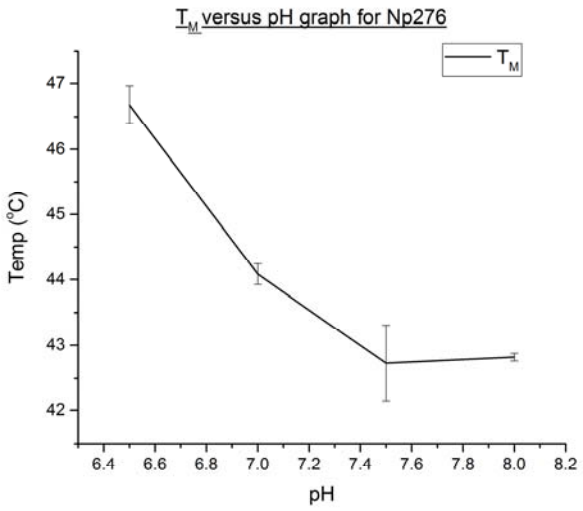
real-time PCR instrument, the data is plotted and then T_M determined by fitting with a Boltzmann sigmoidal fitting function as described in the original protocol.

$$y = \frac{A_1 - A_2}{1 + e^{(x - x_0)/dx}} + A_2$$

Figure 22. The Boltzmann sigmoidal curve is used to fit the raw data. A1 (emission maxima of the unfolding curve), A2 (emission minima at low temperature), x (temperature) and x_0 (value of temperature at 50% unfolding, the T_M).

In the DSF assay, initially a lower fluorescence emission is observed indicating that the protein is folded and that the hydrophobic core isn't solvent exposed. As the protein-dye solution is heated, more of the hydrophobic core can become solvent exposed, thus resulting in more dye binding and a higher fluorescence emission intensity. After the maxima is reached, with the hydrophobic core fully exposed, a decrease in fluorescence intensity is then seen as the protein begins to form aggregates.

Np276 protein was incubated in 10mM NaPi at a range of pH values from 6.5 to 8.0. Whilst there wasn't a strong trend, an increase in T_M was observed as the pH was lowered.



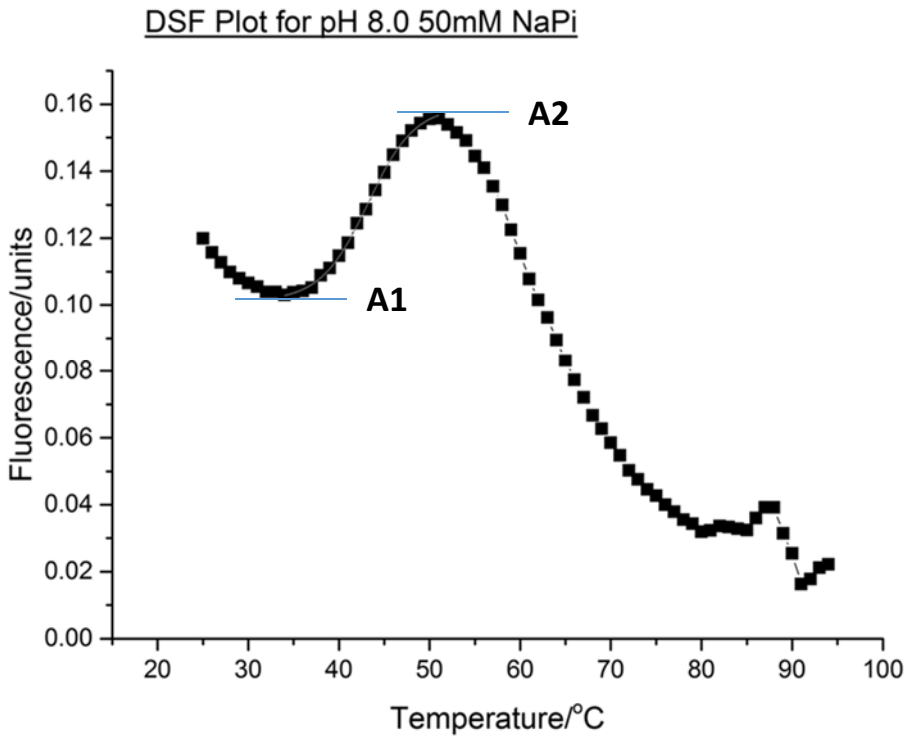


Figure 23. T_M versus pH graph for Np276 in different 10mM NaPi buffers (top) Raw DSF plot for Np276 in 50mM NaPi (bottom). A1 and A2 from Figure 22 (for use in the Boltzmann fit are shown on this for reference)

Despite this trend, a much clearer transition is observed when using 50mM NaPi. A clearer transition indicates there is a definite distinction between the folded and unfolded states and thus this buffer was used in subsequent studies. A key observation being that for the protein to remain folded is that it should be kept cold as much as possible.

A problem often observed on SDS-PAGE with the Np276 fusion protein, is the presence of bands of lower molecular mass. These were subsequently identified by LCT-MS analysis, a PAWS *(*Protein Analysis Work Sheet)/ExPASy Peptide cutter analysis of these cleavage sites indicated that this back-bone cleavage was non-enzymatic in nature. The only enzyme that could potentially have caused such cleavages would be Proteinase K, a broad spectrum serine protease. This enzyme isn't expressed in *E. coli*; furthermore, a double concentration of protease inhibitor and mild, cold, lysis (sonication) conditions were always used.[46]

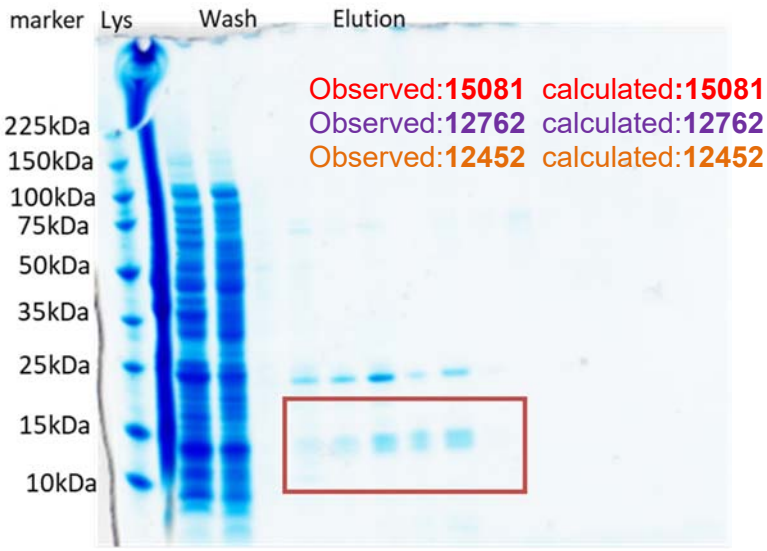


Figure 24. Np276 15Å ruler protein purification gel (Ni-NTA) and identification of the cleavage products. The PAWS and ExPASy analysis indicated that the degradation products were non-enzymatic in nature.

Subsequently, a degradation versus time assay was also conducted at ambient temperature. In this, it was found that the % of degradation products (those listed above) dramatically increased over time, further supporting the idea that this was a non-enzymatic degradation of the purified protein.

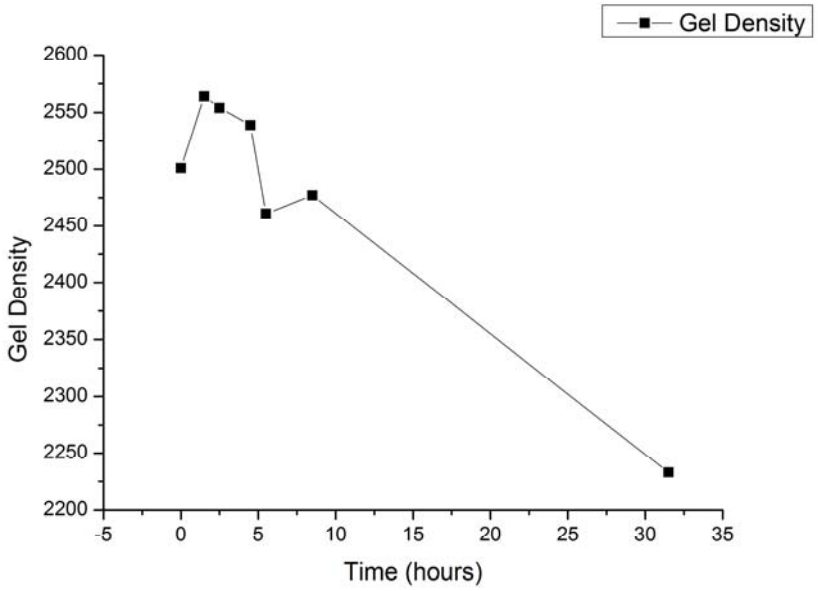


Figure 25. Gel densitometry as measured using ImageJ of the Np276 protein as a function of time. The protein was kept inside a sealed Eppendorf at ambient (25°C) temperature for the allotted time. Only after a significant period of time is a high level of protein loss observed by SDS-PAGE.

In addition to the studies conducted here, the protein was also incubated with a variety of potential laboratory contaminants to assess whether these were causing degradation of the backbone (these contaminants and conditions were found not to cause degradation). The removal of these lower molecular weight impurities became imperative and thus a range of purification methods were also investigated (see experimental). The results of these studies showed that these impurities were difficult to remove by anion exchange or gel filtration, thus it was best to isolate the pure protein. By minimising the handling of the protein (i.e. by ammonium sulfate cut) and later, by use of special resin/milder lysis conditions (see below), this deleterious degradation could be prevented.

2.1.10 Protein Expression using high density growth conditions

In order to improve yields subsequent work looked at using high density expression in tandem with the SPI method. An initial expression with 2XTY and Met was undertaken and the gel and mass spectrum for this is shown in Figure 26. Gratifyingly, the protein was expressed in a pure state (with no truncation observable on the SDS-PAGE Gel (note these expressions were undertaken in a different lab but same gel/prep as used in Oxford). In comparison to previous expressions, these expressions were carried out using both a cell disrupter (a milder form of lysis) and a nickel sepharose excel resin. This resin allows for the inclusion of 1mM EDTA into buffers (without causing Ni^{2+} leaching) helping to prevent any metal mediated (Lewis acid) backbone cleavage.

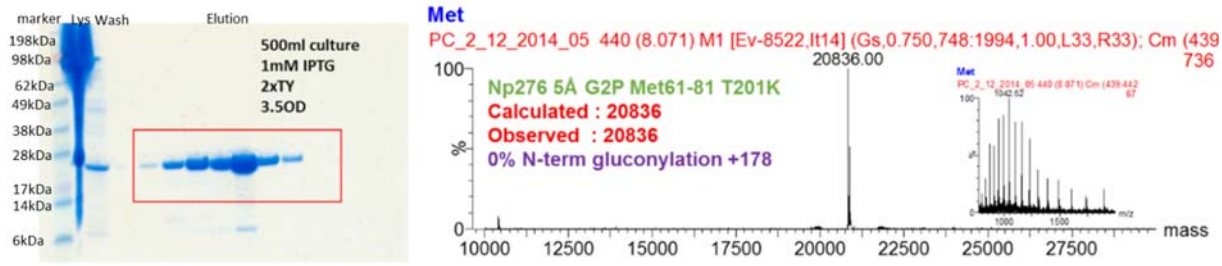


Figure 26. SDS-PAGE and mass spectra of the Met incorporated protein. 0% N-terminal gluconylation was observed as this protein was expressed in the nitrogen rich 2XTY media.

Following this, the protocol was expanded into the expression of the AHA incorporated protein. Again this yielded protein, without any truncation, which from mass spectral analysis showed azidohomoalanine incorporation. Unfortunately, a trace of approximately ~3.7% N-terminal gluconylation of this protein was observed by MS, in all cases a +41Da adduct was observed (Figure 27).

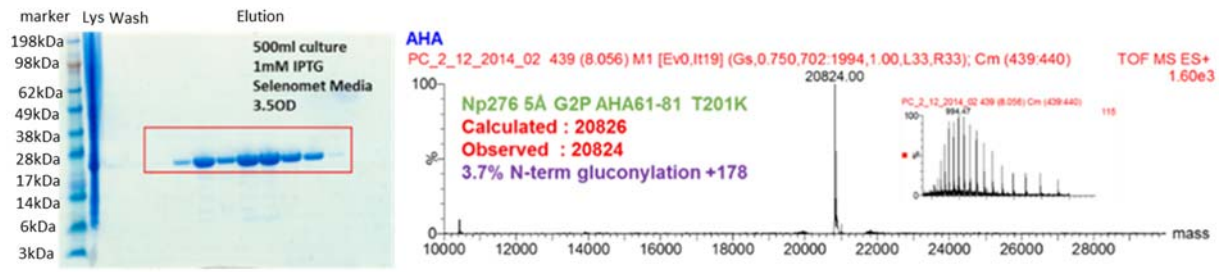


Figure 27. SDS-PAGE and mass spectra of the AHA incorporated protein. When conducting the expression with AHA 3.7% N-terminal gluconylation was observed. Compared to Figure 26, this involved a media shift to the high glucose selenomet media.

The experiment was subsequently repeated but instead lowered the temperature to 16°C and doubled the amount of nutrient mix (4X). This again decreased the % of this N-terminal modification, once more the +40 modification was observed (Figure 28).

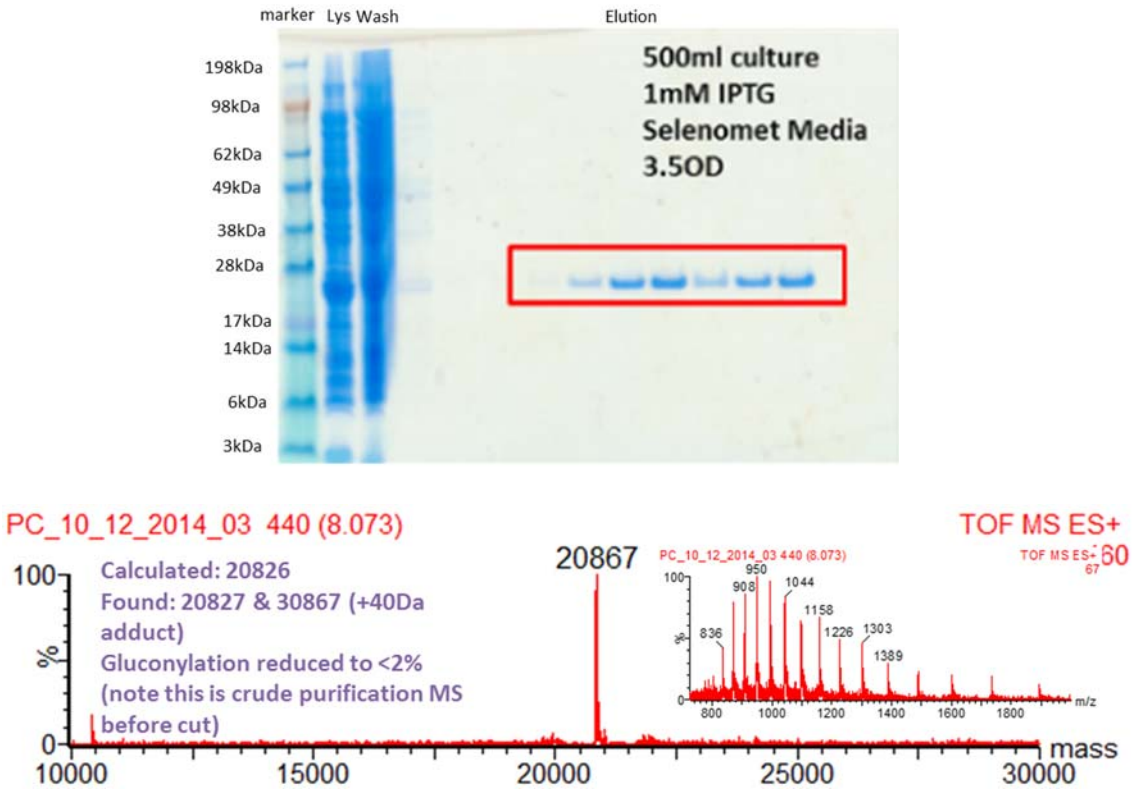
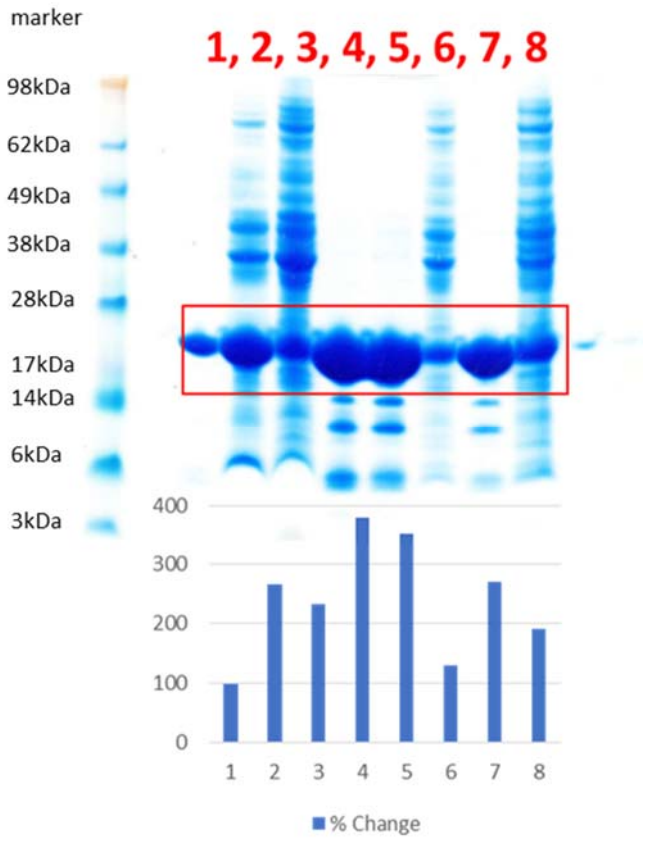


Figure 28. SDS-PAGE and Mass Spectra of the AHA incorporated protein. This result had also been shown previously, conducting the expressions at a lower temperature and with 4X nutrient mix (glucose concentration) led to less N-terminal gluconylation.

When a controlled expression trial was conducted using the PhysNexus™ and a range of expression conditions, it was found that conducting expressions at lower IPTG concentrations (0.3 >1mM) and with MOPS/glycerol present, led to the highest yields. Compared to the control, a ~350% increase in yield was observed for an expression utilising MOPS/glycerol and also 0.3mM IPTG.



Trial	OD (A^{600})	MOPS (Y/N)	IPTG (mM)
1	2.5	N	1
2	2.5	Y	1
3	2.5	N	0.3
4	2.5	Y	0.3
5	2.5	Y	0.1
6	2.5	N	0.1
7	1.25	Y	1
8	5	Y	1

Figure 29. SDS-PAGE and gel densitometry yields for an expression trial conducted with the 15Å Np276 G2P Met containing ruler protein. MOPs/glycerol supplementation supports the cells at higher OD allowing for a higher protein yield.

2.1.11 The physical characterisation of PRPs by CD

The (bio) physical characterisation of pentapeptide repeat proteins have received scant attention, in comparison to other classes of proteins. [47,48]

The circular dichroism of pentapeptide repeat proteins has been shown in the literature to register a poor correspondence to secondary structure calculations (SSC). SSC calculations with programs such as CDNN 2.1 (Applied Photophysics), compare the RMSD between the calculated and crystallographic CD spectra.[49] This means it is best to compare to reference spectra rather than calculate SSC for this class of proteins.[49] Np276 proteins were prepared in NaPi buffer (as chloride interferes with the spectra) and run from 200-260nm on a ChiraScan™ CD machine. Whilst, there have been studies on the MfpA dimer, the most structurally similar PRP protein to Np276 is the Rfr32 protein[50,51], its CD spectrum is shown below as it represents the best comparison to Np276. A similar structure was also obtained by Woolfson *et al* for the AhQnr PRP protein, however this protein shows less similarity to Np276 when compared to Rfr32[52]

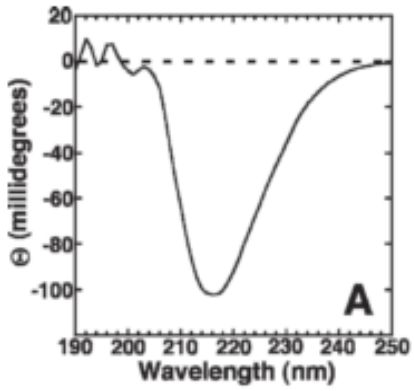


Figure 30. Literature, experimental CD data of the Rfr32 protein. Reproduced with permission. The characteristic dip at 210-225nm observed for the Np276 protein is also seen in Rfr32.

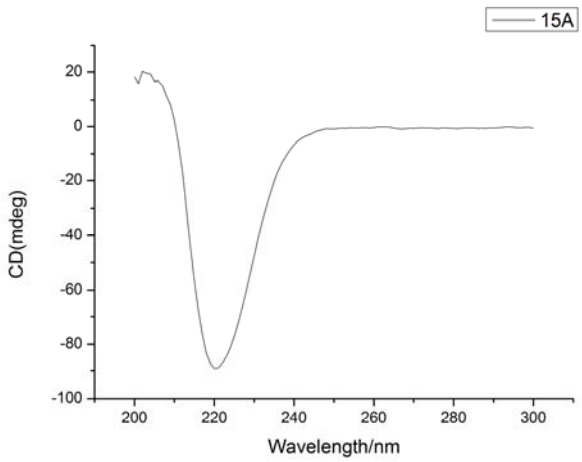


Figure 31. Np276 15Å, showing the characteristic CD shape for a PRP protein. The shape of the graph here corresponds favourably to that of the Rfr32 shown in Figure 30.

As can be seen from Figure 31, the Np276 protein compares favourably to the Rfr32 a protein. When the Np276 protein misfolds or denatures, it tends towards a characteristic α-structure as shown for a series of poorly expressed and misfolded proteins in Figure 32. [53–55]

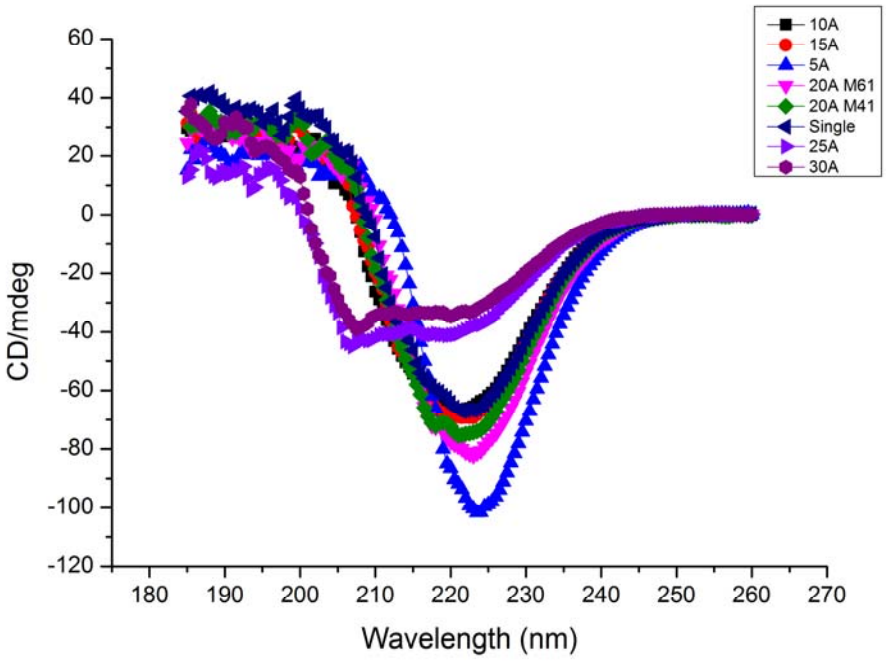


Figure 32. CD spectra for some poorly folded Np276 proteins. In particular, the 25Å and 30Å proteins have adopted an α structure.

A time versus protein stability assay was conducted with the Np276 (G2P, M21I, K25/94/123A T81M (Met)) but despite running the assay for 24 hours at 25°C, the CD indicated that little denaturation of the protein had occurred. Although this result is positive, it could be that a backbone cleavage was occurring but that the resulting fragments still retained the β-secondary structure giving the CD the characteristic dip. However, when compared to the SDS-PAGE/densitometry assay (Figure 25) where over a 10hr period little change was seen, this is indicative that the protein is moderately stable in buffer, when no external pressures (solvent/heat/agitation) are applied.

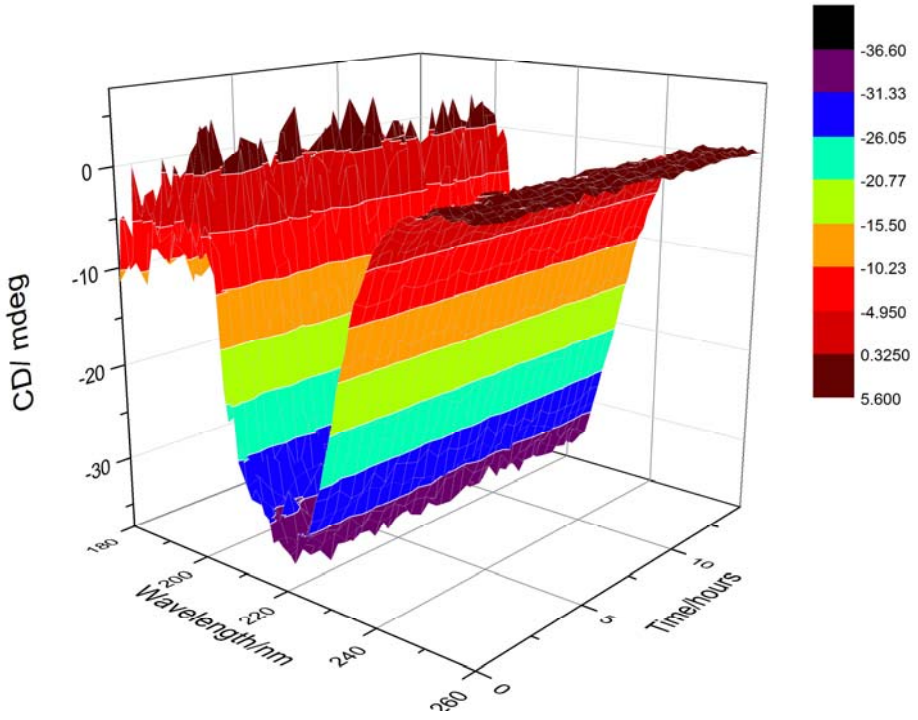


Figure 33. CD spectra for the Np276 5Å P2P, Lys mutant (Met). The CD assay was run at a constant 25°C for 10 hours.

To ascertain the thermal stability of the protein, a CD melt study on this mutant was conducted, the wavelength was recorded across the full range and is plotted at 220nm along with the full spectra (Figure 34).

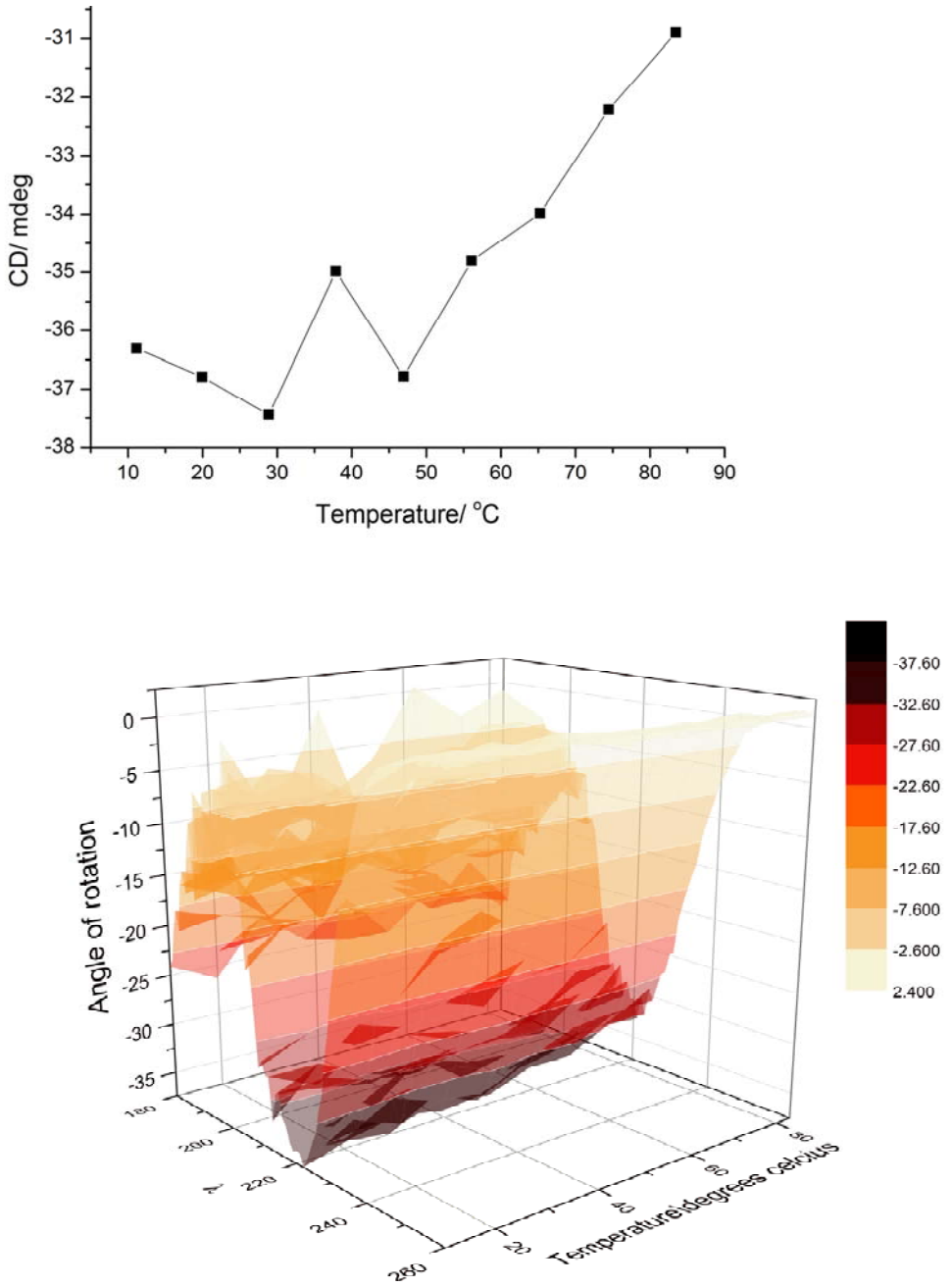


Figure 34. CD spectra for the Np276 5Å P2P, Lys mutant (Met). The CD assay was run versus temperature with spectra (repeat of 5) being conducted at each temperature (10-90°C).

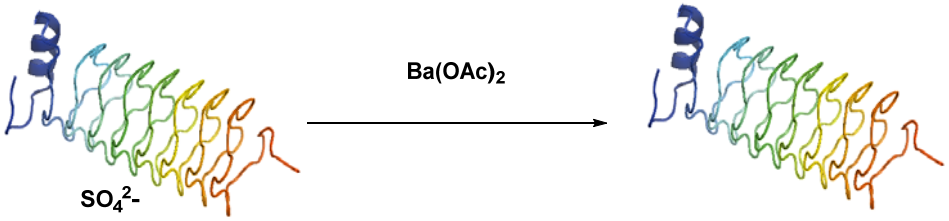
The data presented in Figure 34 showed that change in circular dichroism was only observed from around 40°C. The DSF assays also indicated that the protein unfolding began to occur around this temperature. Despite this the structure remains remarkably stable over the temperature range, something that doesn't correlate with the DSF assays, as the CD data at

90°C still shows β-structure. This indicates that the Np276 protein potentially doesn't unfold in response to temperature or is unfolding in a non-cooperative manner. One important difference between the DSF assay and the CD thermal shift assay is that the DSF assay includes the addition of an additional hydrophobic dye. As Np276 isn't a globular protein there is the potential for this dye to be perturbing the protein structure by interfering with the core of the β-quadrilateral helix, resulting in a lower T_M value from the DSF assay. It is also noted that the best results with the DSF assay are reported to be obtained with globular proteins.[56]

Importantly, as the protein modification reactions will use dyes (which are often hydrophobic) the protein solution should be kept as cold as possible to reduce the possibility of unfolding (chemical denaturation). In addition despite the discrepancy about temperature dependant unfolding, it has been shown that Np276 folds into a PRP coil (like Rfr32) and this remains stable at lower temperatures in both assays. Furthermore, it was also shown that the protein can also adopt an undesireable α-secondary structure.

2.1.12 Adduct formation on ESI-TOF MS

After an ammonium sulfate cut, a +98Da adduct was often observed by ESI-TOF MS, this corresponds to the sulfate or phosphate adducts as noted by Chowdhury *et al*.[57] In order to confirm this with Np276, the protein was incubated with barium acetate for 30 minutes. A white precipitate was observed (barium sulfate) and subsequently removed by centrifugation. Mass spectral analysis subsequently showed that the adduct peak was removed.



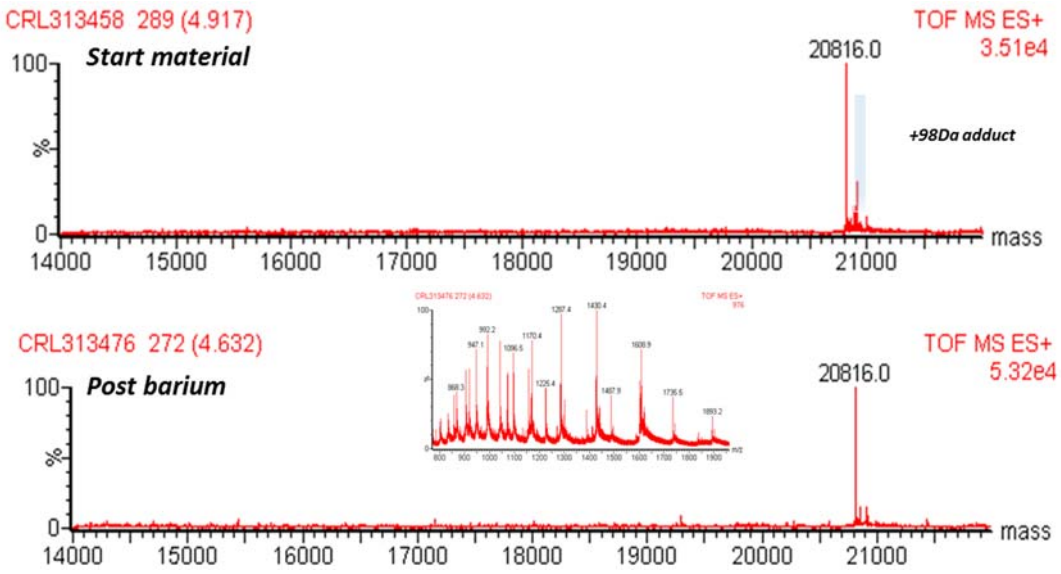


Figure 35. Removal of +98Da adducts from Np276 spectra by treatment with barium acetate. The +98Da adduct is highlighted with a blue box.

In addition, a common adduct of +42Da was observed on LCT-TOF-MS systems, the adduct being proposed as an acetonitrile adduct. Protein acetylation isn't commonly found in *E. coli*[††] and no urea was used (thus no opportunity for carbamylation).[58] Np276 G2P mutants appear to have a high propensity to displaying this adduct, that is until the N-terminus is blocked with the addition of a dye (chapter 3), at which point the adduct is no longer seen. Adduct formation at N-terminal proline has previously been reported within the literature (via urea).[59–61] Thus, it is reasonable to conclude that adducts seen are artefacts from ESI-TOF-MS analysis and thus inconsequential to the final expressed proteins.

[††] Due to a lack of N-α-terminal acetyltransferase complexes. In eukaryotic expression acetylation is a common PTM where it is in involved in protein stability.

2.1.13 MS/MS analysis of the Np276 protein with trypsin

Np276 proteins were initially digested with trypsin, a serine protease commonly used in bottom-up proteomics. Digestion of Np276 overnight at 37°C, followed by purification on a C18 ZipTip cartridge, gave the peptides which were submitted to the SYNAPT nanoflow system. A fit of found peptides to a theoretical digest indicated that a limited number of peptides were observed, many of the theoretical peptides being of a large size. Instead, other enzymes which might yield smaller peptides were investigated.[62]

Associated Datafile: PC_PKC15 (50 - 1990 amu)
Trypsin:/K-\P /R-\P

Frag#	Res#	Sequence	Theor(Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]	[M+5H]
T3	27-34	(R)QLYAAGER(D)	906.46	907.46	454.24	303.16	227.62	182.30
T4	35-42	(R)DFSIVDLR(G)	963.50	964.51	482.76	322.18	241.88	193.71
▪T2	17-26	(R)GSHIDVEALR(Q)	1095.57	1096.58	548.79	366.20	274.90	220.12
T6	74-87	(R)ADLSGATLNGADLR(G)	1372.69	1373.70	687.36	458.57	344.18	275.55
T1	1-16	(-)PSSHHHHHSSGLVPR(G)	1807.87	1808.88	904.94	603.63	452.98	362.58
T3-4	27-42	(R)QLYAAGERDFSIVDLR(G)	1851.95	1852.96	926.98	618.32	463.99	371.40
T2-3	17-34	(R)GSHIDVEALRQLYAAGE R(D)	1984.01	1985.02	993.01	662.35	497.01	397.81
T9	174-200	(R)ANLTGANLEDANLEGTI LEGGNNNLAK(-)	2725.35	2726.36	1363.68	909.46	682.35	546.08
T1-2	1-26	(-)PSSHHHHHSSGLVPRG SHIDVEALR(Q)	2885.43	2886.44	1443.72	962.82	722.37	578.09
T5	43-73	(R)GAVLENINLSGAILHGA XLDEANLQQANLSR(A)	3228.70	3229.70	1615.36	1077.24	808.18	646.75
T8	138-173	(R)EADLSEANLEAADLSGA DLAIADLHQANLHQAALER(A)	3757.04	3758.05	1879.53	1253.36	940.27	752.42
T4-5	35-73	(R)DFSIVDLRGAVLENINL SGAILHGAXLDEANLQQANL SR(A)	4176.66	4177.67	2089.34	1393.23	1045.17	836.34
T5-6	43-87	(R)GAVLENINLSGAILHGA XLDEANLQQANLSRADLSGA TLNGADLR(G)	4586.06	4587.07	2294.04	1529.69	1147.52	918.22
T7	88-137	(R)GANLSAADLSDAILDNA ILEGAILDEAVLNQAXLAAA NLEQAILSHANIR(E)	5096.69	5097.70	2549.35	1699.90	1275.18	1020.35
T6-7	74-137	(R)ADLSGATLNGADLRGAN LSAADLSDAILDNAILEGAI LDEAVLNQAXLAAANLEQAI	6452.16	6453.17	3227.09	2151.73	1614.05	1291.44

Figure 36. Tryptic digest fitted to the theoretical digest (0.1Da tolerance) showing a poor sequence coverage (37%). The digest was purified with a C18 ZipTip.

2.1.14 MS/MS analysis of the Np276 protein with AspN

AspN is a metalloprotease from an engineered *Pseudomonas fragi*. It has been reported to specifically cleave at the N-terminal side of aspartate residues (it also requires zinc to function) and offers a complementary or separate cleavage profile to Trypsin. In this case it was trialled, as a theoretical digest with ExPASy indicated that shorter peptides would be produced, these would then be more amenable to analysis via the SYNAPT system.

Associated Datafile: PC65C (50 - 1990 amu)
AspN:-/D

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
D1	1-20	(-) PSSHHHHHHSSGLVPRG SHI (D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I) DVEALRQLYAAGER (D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-39	(R) DFSIV (D)	579.29	580.30	290.65	194.10	145.83
D4	40-61	(V) DLRGAVLENINLSGAIL HGAJL (D)	2271.24	2272.25	1136.63	758.09	568.82
D5	62-74	(L) DEANLQQANLSRA (D)	1428.70	1429.70	715.36	477.24	358.18
D6	75-84	(A) DLSGATLNGA (D)	917.45	918.45	459.73	306.82	230.37
D7	85-94	(A) DLRGANLSAA (D)	986.51	987.52	494.27	329.85	247.64
D8	95-97	(A) DLS (D)	333.15	334.16	167.58	112.06	84.30
D9	98-101	(S) DAIL (D)	430.24	431.25	216.13	144.42	108.57
D10	102-111	(L) DNAILEGAIL (D)	1027.55	1028.56	514.79	343.53	257.90
D11	112-139	(L) DEAVLNQAJLAAANLEQ AILSHANIREA (D)	2970.53	2971.54	1486.27	991.18	743.64
D12	140-149	(A) DLSEANLEAA (D)	1031.48	1032.48	516.75	344.83	258.88
D13	150-154	(A) DLSGA (D)	461.21	462.22	231.61	154.75	116.31
D14	155-159	(A) DLAIA (D)	501.28	502.29	251.65	168.10	126.33
D15	160-182	(A) DLHQANLHQAALERANL TGANLE (D)	2498.26	2499.27	1250.14	833.76	625.57
D16	183-200	(E) DANLEGTILEGGNNNLA K(-)	1841.91	1842.92	921.96	614.98	461.49

Figure 37. AspN digest of the 15Å protein fitted to the theoretical digest (0.1Da tolerance) showing a good sequence coverage (79%). The digest was purified with a C18 ZipTip.

2.1.15 Library synthesis

The Np276 ruler library of 8 proteins outlined in Table 3 (with G2P) were expressed in selenomet minimal media, utilising the B834 (DE3) strain along with double nutrient mix, and azidohomoalanine. Expression was conducted at 16°C for 16 hours overnight in all cases with the proteins being purified by nickel affinity on an AKTA-FPLC system; in some cases an anion exchange step was used but in most an ammonium sulfate cut was used instead. These proteins were expressed prior to the use of the more reliable high yield method, but in most cases good protein purity was observed along with consistent secondary structure and clean mass spec (the +42Da adduct being that previously described). A summary of the results from these expressions is shown in Table 3, with the full purification and characterisation data contained in the experimental.

Table 3. Table of expressed novel Np276 mutants prepared containing the G2P mutation. The proteins were expressed without any N-terminal gluconylation and showed the required β structure.

Construct	Mutations	Yield/mgL^{-1}	Calculated Mass/Da	Observed Mass/Da
5 Å	G2P, M21I, K25/94/123A T81M, T201K	3.8	20826	20826
10 Å	G2P, M21I, K25/94/123A I101M, T201K	6.3	20813	20814
15 Å	G2P, M21I, K25/94/123A N121M, T201K	3.7	20812	20814
20 Å	G2P, M21I, K25/94/123A D141M, T201K	6.7	20811	20812
20 Å	G2P, M21/61I, K25/94/123A D41M, N141M T201K	7.8	20811	20814
25 Å	G2P, M21/61I, K25/94/123A, D41M, D141M, T201K	5.4	20800	20802
30 Å	G2P, M21/61I, K25/94/123A, D41M, D161M, T201K	4.9	20810	20812
Single	G2P, M21I	11.42	20873	20873

2.2 Conclusions

In conclusion an Np276 'ruler' library incorporated with the unnatural amino acid AHA was expressed and characterised by biophysical techniques. In order to express this library challenges including N-terminal gluconylation and purification/expression issues were surmounted. The N-terminal gluconylation work yielded a potential additive based method to prevent this modification; however, further proof of concept work would be needed to validate this and also to expand its scope. The mutational approach was also shown, when combined with an optimisation of expression conditions, to yield for the first time a reliable set of expression conditions that rigorously supressed the modification. These conditions were used to express a library of Np276 rulers. These ruler proteins were characterised by MS, MS/MS, SDS-PAGE gel and CD. Due to the incompatibility of the protein family with CD prediction algorithms, the obtained spectra were compared with literature protein spectra such as that from Rfr32. This showed a high degree of similarity between the 2 proteins, thus indicating that a β-structure was present and that the proteins were correctly folded. The work carried out here provides an excellent base to modify and subsequently use the Np276 proteins as a molecular probe of function, a distance spaced multivalent scaffold, a molecular ruler.

2.3 Biological experimental section

General expression protocol for Np276 proteins

1µL of the plasmid (pet28d) encoding for a Np276 ruler protein was transformed into 20µL of B834(DE3) cells according to the manufacturer's instructions. After plating on kanamycin LB:agar plates and incubating overnight at 37°C, a healthy colony was used to inoculate an overnight culture of 20ml of freshly autoclaved LB media supplemented with 50µgL^{-1} kanamycin. After shaking at 180rpm and 37°C for 16 hours, the overnight culture was added to 1L of freshly autoclaved selenomet media, supplemented with sterile 50µgL^{-1} kanamycin, 100ml filter sterilised selenomet nutrient mix (5.1g/50ml) and 40µgL^{-1} sterile filtered L-methionine. The culture was grown at 37°C and 180rpm until an optical density of 0.6-0.8 was reached, approximately 3 hours in bevelled flasks. The culture was pelleted with a JA10 rotor, 7000 rpm , 10 minutes, 4°C and washed thrice with 1L of freshly autoclaved room temperature, selenomet media supplemented with 50µgL^{-1} sterile kanamycin but absent of methionine or azidohomoalanine. The cells were subsequently re-suspended in 1L of pre-warmed (37°C) selenomet media supplemented with 50µgL^{-1} sterile kanamycin, 100µgL^{-1} of sterilised azidohomoalanine and 50ml of sterile nutrient mix. After incubation at 37°C for 30 minutes and 30°C (temperature specified in the main body of the text) for 30 minutes, expression was induced with solid IPTG (240mg, 1mmol) and the culture incubated for 16 hours at 30°C. (The temperature at which the expression is conducted is specified in the main text). The cells were then harvested, 7000rpm, 10 minutes JA10, 4°C and flash frozen in liquid nitrogen.

DNA preparation and manipulation of Np276 proteins

DNA analysis was conducted using the QIAGEN Mini-Prep kit. In this a 5ml overnight culture was grown and the resultant culture centrifuged (3750rpm, 10 min, 4°C, Beckmann Coulter X-12R table top centrifuge). The protocol described in the Mini-Prep kit was then followed. The resulting solution of plasmid was submitted to the Oxford Geneservice (Source Bioscience) utilising the primers, T7F and T7R (for Np276 proteins, pet28d vector). Site directed mutagenesis was carried out using a Stratagene QuikChange II XL site directed mutagenesis kit and the resultant plasmid DNA transformed into Xl-10 gold ultra-competent cells as per the manufacturer's instructions. For mutations of the Np276 proteins, primers were designed using the PrimerX service and the reactions prepared as required by the Stratagene QuikChange II XL site directed mutagenesis kit. The PCR cycle for successful mutations of the Np276 proteins is shown below. The full sequences for the proteins is shown in the data set for each protein.

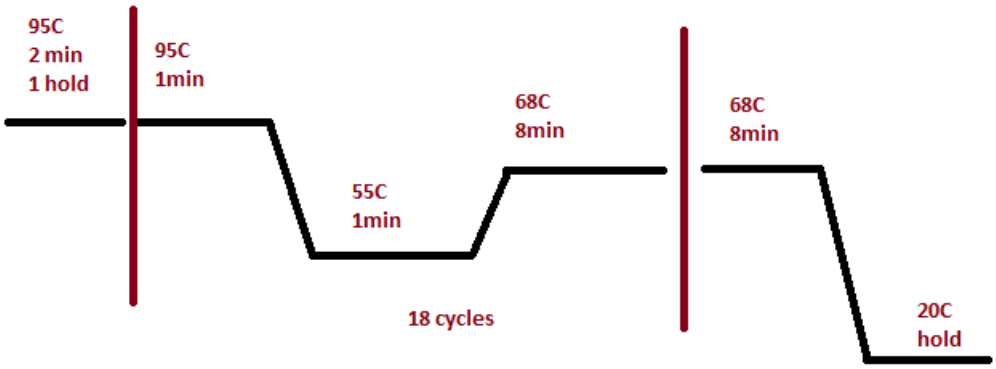


Figure E1. PCR cycle used to prepare the G2P Np276 mutants. This protocol was generated via PrimerX.

Primers for the mutagenesis were ordered from Sigma Aldrich.

Primer pair for CCC (Rare Codon, G2P mutation at the N-terminus of Np276).

Forward: 5' GAAGGAGATATACCATGCCCAGCAGCCATCATCATCATC 3'

Reverse: 5' GATGATGATGATGGCTGCTGGGCATGGTATATCTCCTTC 3'

Primer pair for CCA (G2P mutation at the N-terminus of Np276).

Forward: 5'-GAAGGAGATATACCATGCCAAGCAGCCATCATCATCATC

Reverse: 5'-GATGATGATGATGGCTGCTTGGCATGGTATATCTCCTTC

SDS-PAGE analysis of purified proteins

The protein fractions had a 8µL aliquot removed and 2µl of stain was added before the samples in micro-centrifuge tubes was spun on a table top centrifuge for approx. 5 seconds. These samples were then placed in a heating block at 80-90°C for 10 minutes. The SDS-PAGE gel (4-12% Bis-Tris Gel, Invitrogen) was loaded onto the holder and 800ml of 20X MES buffer added (760ml H_2O:40ml MES buffer). The standard (10-225kDa SharpStain Novex or SeeBlue Plus2) was added to well 1, followed by the fractions. The gel was then run at 200V, for approximately 30 minutes-1 hour, until the bands were clearly resolved. The gel was removed and placed into instant blue stain with shaking. It was shaken overnight before being washed and shaken thrice (30-1hour each) with dd. H_2O. The gel was then scanned into a PC on a digital flatbed scanner.

Protein analysis by LC-MS

Protein liquid chromatography MS (LC-MS) was conducted using a Waters LCT Classic (ESI-TOF-MS) connected to a Shimadzu 20 Series HPLC, using a Merck Chromolith C18 (50x2mm) column. Water (solvent A) and acetonitrile (solvent B) each with 0.1% formic acid were used as a mobile phase at a flow rate of 0.4ml/min. The gradient was set to 95% A (2 mins, isocratic), to 95% B after 8 mins. Then isocratic (95% B) for 3 minutes. The electrospray source parameters were: capillary voltage (3kV), cone voltage (250kV) with nitrogen as the nebuliser and solvation gases. Spectra were processed using the MaxEnt algorithim on Waters

MassLynx v4.1. For concentration versus signal studies the reader is directed to literature sources for Np276.[63]

Manual protein purification of Np276 proteins

The cell pellet was added to approximately 30ml of nickel affinity chromatography loading buffer (20mM tris.HCl, 40mM imidazole, 500mM NaCl, 5% glycerol, pH8.0), and 10mg of lysozyme, DNAse and 2 tablets of Roche EDTA free protease inhibitor were added. After stirring gently for 30 mins at 4°C, the cells were sonicated (40% power, on 30s, off for 1 minute each) a total of 5 times. The cells were then pelleted, 20000rpm, JA25.5, 4°C, 30min and the cell lysate loaded onto a freshly prepared column. Manual affinity chromatography was performed as follows. An empty plastic column was charged with a small quantity ~2-3 ml of the affinity resin slurry (Ni NTA). This was then washed with water, EDTA solution (20mM tris, 100mM EDTA, pH7.5), water, nickel solution (100mM nickel sulfate), water and binding buffer **(A)** (pH7.5, 500mM NaCl, 20mM Tris, 40mM imidazole) (all 30ml). The protein sample on ice was then loaded by gravity. The columned was eluted using the following gradient system: elution buffer **(B)** (pH7.5, 500mM NaCl, 500mM imidazole, and 20mM tris).

Table E1. Buffer conditions for manual purification. The buffer should be made up and kept cold and sterile prior to use.

	Volume A (binding)/ml	Volume B (elution)/ml
100% binding buffer	50	0
3% elution buffer	32	1
10% elution buffer	27	3
20% elution buffer	24	6
40% elution buffer	18	12
100% elution buffer	0	30

AKTA-FPLC Protein Purification for Np276 Proteins with a manually packed Ni-Excel (sepharose) column.

The cell pellet was added to approximately 30ml of nickel affinity chromatography loading buffer (20mM tris, 40mM imidazole, 500mM NaCl, 1mM EDTA, pH7.8), and a single tablet of EDTA free protease inhibitor was added. The cells were then passed through a cell homogeniser, pre-chilled to 4°C The cells were then pelleted, 15000rpm, JA25.5, 4°C, 30min and the cell lysate incubated overnight with 3ml of prewashed nickel sepharose Excel resin.

A key advantage of Ni sepharose excel resin (GE Healthcare) over other commercial resins is its ability to tolerate a high % of EDTA. Up to 10mM may be safely used without concern of leaching of the bound metal. Furthermore, the column doesn't need to be recharged between uses and can be carefully washed with dilute (0.1M) NaOH prior to use. For Np276 purifications fresh resin was used for each different protein. Before use, the resin (6ml of a slurry: 3ml resin) was centrifuged (1000g, 20min, 4°C) and washed twice with ice cold PBS (20ml), after decanting the buffer this resin was then incubated with the lysate for 16 hours at 4°C, with an additional tablet of protease inhibitor (per 50ml total volume). Instead of using a manual column or a NiNTA HP or FF column, a manual loading AKTA column was loaded by hand by carefully drawing through the lysate-resin which had incubated (rotated) overnight.

The manual column was loaded and then the column run on an AKTA-FPLC (Explorer system). Following elution with a 4 step gradient on the AKTA-FPLC system, fractions containing protein were combined and subjected to an 85% ammonium sulfate cut. After stirring gently overnight (16 hours) the protein was precipitated out of solution (20k, 30min, JA25.5,4°C) and the protein re-dissolved in a minimum of NaPi (50mM, pH8.0, 5% glycerol). The protein was then run though a PD-10 column. ESI-TOF-MS was recorded.

AKTA-FPLC protein purification for Np276 proteins with a Ni NTA 5ml column and an anion exchange HP 5ml Q column.

The cells were then pelleted, 20000rpm, JA25.5, 4°C, 30min and the cell lysate loaded onto a freshly prepared 5ml nickel Hi-Trap HP column. Following elution with a 4 step gradient on the AKTA-FPLC system (20CV 0% B, 5CV 5% B, 5CV 10% B, 5CV 15% B, 5CV 20% B, 10CV 100%), fractions containing protein were combined and subjected to an 85% Ammonium Sulfate cut. After stirring gently overnight (16 hours) the protein was precipitated out of solution (22k, 30min, JA25.5, 4°C) and the protein re-dissolved in a minimum of NaPi (50mM, pH8.0, 5% glycerol). The protein solution was then loaded onto a anion exchange 5ml HiTrap HP Q column and was eluted with a linear gradient (150mM NaCl, 20mM TRIS to 1M NaCl, 20mM tris, 20CV). SDS-PAGE analysis showed that protein was present. Fractions containing protein were combined and subjected to an 85% ammonium sulfate cut. After stirring gently overnight (16 hours) the protein was precipitated out of solution (22k, 30min, JA25.5, 4°C) and the protein re-dissolved in a minimum of NaPi (50mM, pH8.0, 5% glycerol). The protein was then dialysed overnight (4°C, 4L) against the same buffer.

High growth (2XTY) expression protocol

1µL of the plasmid encoding for the 5Å Np276 ruler was transformed into 20µL of B834(DE3) cells according to the manufacturer's instructions. After plating and incubating overnight, a healthy colony was used to inoculate an overnight culture of 200ml of freshly autoclaved 2XTY media supplemented with 50µgL^{-1} kanamycin. After shaking at 180rpm and 37°C for 16 hours, the overnight culture was added to 500ml of freshly autoclaved 2XTY media supplemented with sterile 50µgL^{-1} kanamycin. The culture was grown at 37°C and 180rpm until an A600 of 3.5 was reached, approximately 4.5 hours in bevelled flasks. The culture was pelleted with a JS4.2 rotor, 4000 rpm, 15 minutes, 4°C and washed thrice with 1L of freshly autoclaved room temperature, selenomet media supplemented with 50µgL^{-1} sterile kanamycin but absent of methionine or azidohomoalanine. The cells were subsequently re-suspended in 500mL of pre-warmed (37°C) selenomet media supplemented with 50µgL^{-1} sterile kanamycin, 100µgL^{-1} of sterilised azidohomoalanine (or 125µgL^{-1} of AHA.HCl) and n*25ml (for 500ml culture, for 2X=50ml, 4X100ml, 8X200ml) of sterile nutrient mix. After incubation at 37°C for 30 minutes and 18°C for 30 minutes, MOPS-glycerol (50mM MOPS buffer pH7.4, 1mM Mg salts, 2% glycerol) was added (25ml for 500mL culture, [stock solution: 1M MOPS, 20mM $MgCl_2$/$MgSO_4$, 40% glycerol]) and expression was then induced with solid IPTG (240mgL^{-1}, 1mmol) and the culture incubated for 16 hours at 18°C (or 16°C, see main text).The cells were then harvested, 4000rpm, 15 minutes JS4.2, 4°C and flash frozen in liquid nitrogen. Purification using the excel protocol was then undertaken.

All CD Spectra were recorded in a quartz cuvette (path length 1mm) using an Applied Photo physics ChiraScan™ machine. Data analysis was conducted as described in the literature.[54] In a large number of expressions poorly folded proteins were observed, sometimes even after handling. Unless the protein concentration is stated, all CD spectra were run at 0.5mgml^{-1} of protein concentration.

Stability studies: CD time

Using Np276 5Å G2P,M21I,K25/94/123A, diluted to a concentration of 0.466mgml^{-1} with NaPi, Initially a blank of pH8.0 50mM NaPi was run (200µL), following this the protein was loaded into the cuvette and at 25°C the stability studied overnight for 10 hours. 5 repeats per time point were run with wavelengths between 180nm-260nm.The control software automatically removes the baseline.

Stability studies: CD temperature

Using Np276 5Å G2P, M21I, K25/94/123A diluted to a concentration of 0.466mgml^{-1} with NaPi. Initially a blank of pH8.0 50mM NaPi was run (200µL), following this the protein was loaded into the cuvette and at 10°C the stability studied versus temperature (11°C, 20°C, 29°C, 38°C, 47°C, 56°C, 65°C, 74°C and 83°C). 5 repeats per temperature were run between 180-260nm. The control software automatically removes the baseline. The data was plotted as a function of temperature, wavelength and CD using Origin to obtain the surface.

Stability studies: SDS-PAGE/LCT-MS

An LCT-MS/SDS-PAGE study with some Np276 5Å G2P, M21I, K25/94/123A (200µL, 1.4mgml^{-1}). The sample was in a 1.5ml eppendorf,10µL aliquots were removed at the allotted

time points and snap frozen. All samples were subsequently run on a SDS-PAGE gel (2x buffer, MES (20X), 4-12% BT gel, Coomassie blue stain) before being loaded into the Geldock. The density of the bands was then estimated. LCT analysis (protein diluted to 0.05mgml^{-1}) was subsequently conducted on all samples.

Stability studies: conditions

Aliquots of protein were incubated for 30 minutes under conditions likely to cause truncation. These were: protein at ambient temperature, protein at 4°C (control), under various physical conditions, extreme heat and with residual trace Virkon added. Rotation was undertaken at 40rpm and centrifugation at 10,000g at ambient temperature. Ultrasound was conducted with continuous pulsing (1s on/1s off). A gel was then re-run, and the intensity of the degradation products determined by gel densitometry using Image-J software analysis. The outcome of this was that it indicated the protein was not degrading under the experimental conditions tested, suggesting that they were perhaps degrading during the actual expression (or being expressed truncated).

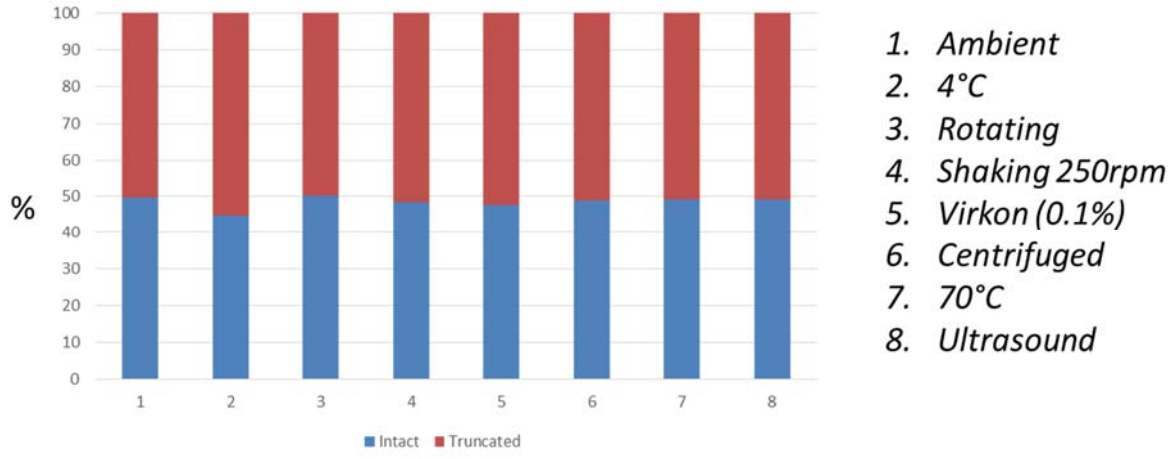


Figure E2. % truncation chart, calculated *via* gel densitometry. In all cases the protein samples were incubated at the stated conditions for 30 minutes, before a second gel was then run.

N-terminal gluconylation assays: temperature and methionine concentration.

1µL of the plasmid (confirmed by Mini-prep and gene sequencing) encoding for the 15Å ruler (Np276) was transformed into 20µL of B834(DE3) cells according to the manufacturer's instructions. After plating (at 90:10 cells to SOC media, 50:50 and 10:90) and incubating overnight, a healthy colony was used to inoculate an overnight culture of 20ml of freshly autoclaved LB media supplemented with 50µgL^{-1} kanamycin. After shaking at 220rpm and 37°C for 16 hours, 400µL of the o/n culture was added to 20ml of LB supplemented with 50µg/ml of kanamycin. After growing to an OD of 0.6, the cells were pelleted 3750rpm, 15 minutes, 4°C and suspended in 20ml of selenomet media supplemented with an appropriate concentration of methionine (20/40/100 µg/ml) and with 50µg/ml kanamycin. After shaking for 30 minutes at 37°C and then 30 minutes at 30°C the cells were induced with 1mM IPTG and shaken (220rpm) at the appropriate temperature for 16 hours.

The cell culture was taken and pelleted at 3750 rpm, 4°C for 20 minutes. To the cell lysate was then added 1.5ml of Bugbuster Master Mix and the mixture shaken at ambient temperature for 20 minutes. After centrifugation, 13K, 25 minutes, 1.5ml of equilibration buffer (20mM NaPi, 300mM NaCl, 10mM imidazole, pH7.4) was added and then 0.6ml of this incubated at ambient temperature for 30 minutes with the His-Pur Ni-NTA spin column. After centrifugation 800g, 2min, the remaining solution was subsequently loaded on sequentially until all had been added. The standard procedure as stated by the manufacturer's was then followed. The eluted His tagged protein solution was then added to a 10K MWCO vivaspin, concentrated to 100µL and buffer exchanged with approximately 3ml of 50mM NaPi pH 8 until again 100µL. A sample was then run on the LCT. The % N-terminal gluconylation was then determined via the ratio of intensities from the mass spectra. The assumption is that the ionisation for both is about the same.

Table E2. Temperature and methionine conditions used in a conditions screen for N-terminal gluconylation.

Trial	Temp/°C	Met conc/ µg ml	% N term
1	20	20	29
2	20	40	30.33
3	20	100	30
4	25	20	33
5	25	40	36.1
6	30	100	39

N-terminal gluconylation assays: general procedure for expressions.

Please refer to the general Np276 expression protocol.

N-terminal gluconylation assays: additive screen.

200ml of freshly autoclaved 2XTY media supplemented with 50µgL^{-1} kanamycin. After shaking at 180rpm and 37°C for 16 hours, the overnight culture was added to 1000ml of freshly autoclaved 2XTY media supplemented with sterile 50µgL^{-1} kanamycin. The culture was grown at 37°C and 180rpm until an A600 of 2.5 was reached, approximately 3.5 hours in bevelled flasks. The culture was pelleted with a JS4.2 rotor, 4000 rpm, 15 minutes, 4°C and washed thrice with 1L of freshly autoclaved room temperature selenomet media supplemented with 50µgL^{-1} sterile kanamycin but lacking methionine or azidohomoalanine. The cells were subsequently re-suspended in 1000mL of pre-warmed (37°C) selenomet media supplemented with 50µgL^{-1} sterile kanamycin, 40µgL^{-1} of methionine and 50ml of sterile nutrient mix. The cells were then split into 50ml cultures, each in a 250ml bevelled flask. The additive at a concentration of 1mM was then added. After incubation at 37°C for 30 minutes and 16°C for 30 minutes, MOPS-glycerol (50mM MOPS buffer pH7.4, 1mM Mg salts, 2% glycerol) was added (25ml for 500mL culture, [stock solution: 1M MOPS, 20mM $MgCl_2$/$MgSO_4$, 40%

glycerol]) and expression was then induced with solid IPTG (240mgL^{-1}, 1mmol) and the culture incubated for 16 hours at 16°C/200rpm.

A 20ml aliquot of each cell expression was then taken and harvested, 4000rpm, 15 minutes JS4.2, 4°C and flash frozen in liquid nitrogen. To the cell lysate was then added 1.5ml of Bugbuster master mix and the mixture shaken at ambient temperature for 20 minutes. After centrifugation, 13K, 25 minutes, 2 ml of equilibration buffer (20mM NaPi, 500mM NaCl, 40mM imidazole, 10mM EDTA, pH8.0) was added. After centrifugation 800g, 2min, 1ml of the clarified cell lysate was added to a PhysNexus™ plate. The PhysNexus™ equipped with excel tips was then run (2x1ml buffer A wash, elution into 100µL buffer B). A sample of each was then run on XEVO-TOF-MS to determine the % of N-terminal gluconylation. Expressions with each of the additives was conducted in triplicate, the control was run only once giving approximately 4% N-terminal gluconylation. Protein samples using the same additive were pooled and run on the same SDS-PAGE gel (Figure E3) before being analysed by ImageJ.

The above protocol was also used for expression screening, minus the addition of additives.

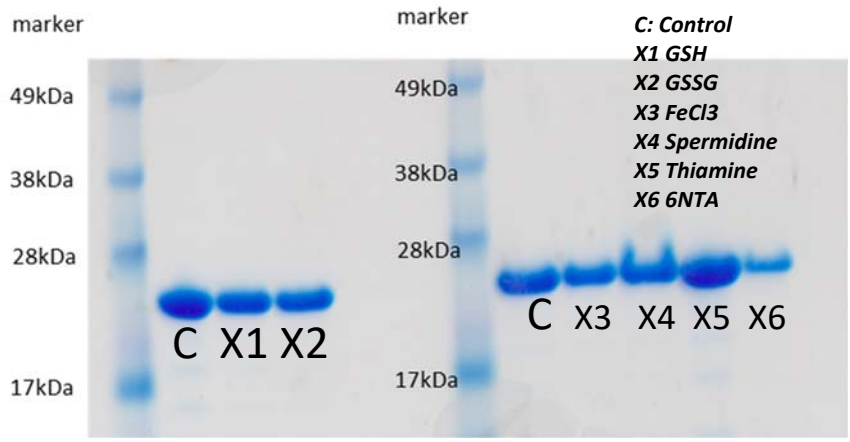
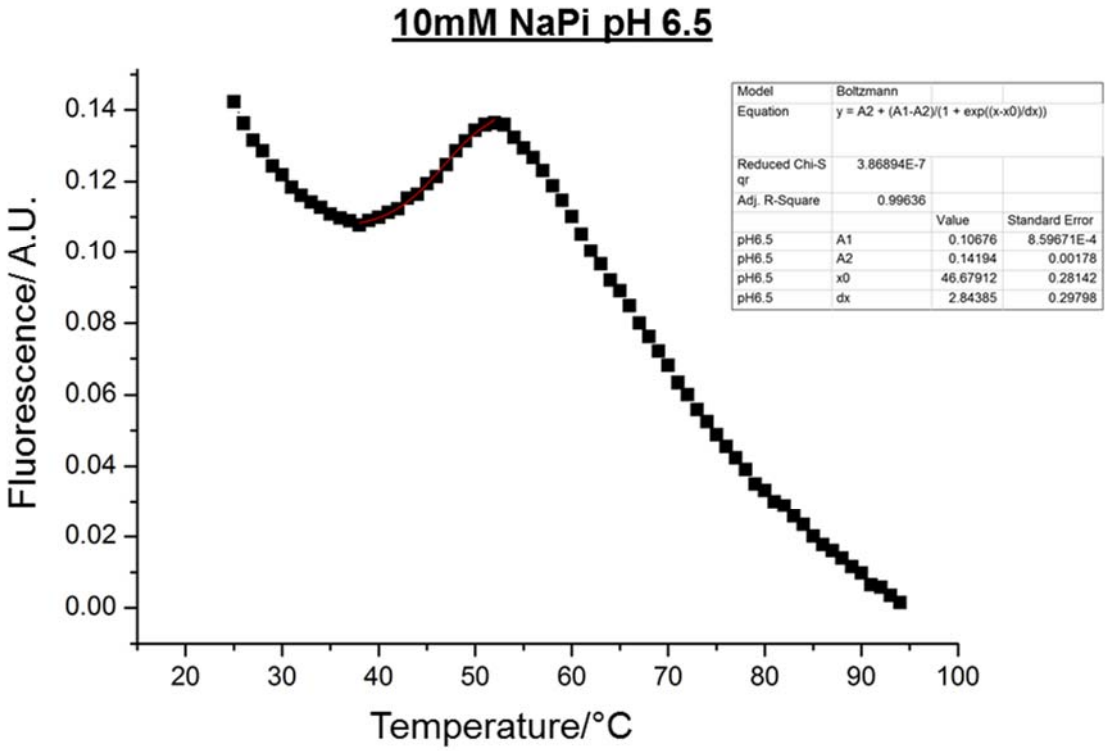


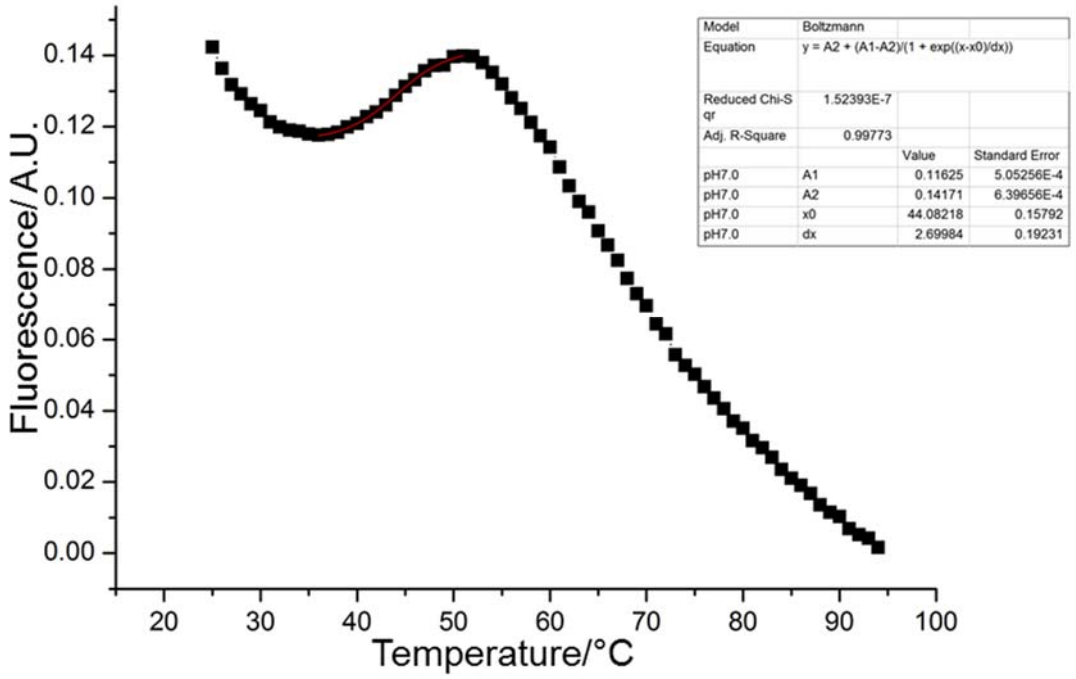
Figure E3 N-terminal gluconylation additive screen gels of the pooled fractions. The change in the expression yield was quantified by comparing the control lane to protein lanes on Image J (densitometry).

Experimental procedure for DSF assays

The protein, Np276 5Å G2P, M21I, K25/94/123A, was diluted to 10µM (200µL + 472µL NaPi 50mM pH8.0). Dye solutions in various buffers (NaPi 50mM pH8,7.5,7 and 6.0, 500mM imidazole, 20mM tris, 50mM KPi and pH7.4 10mM NaPi) were prepared (0.2µL Sypro Orange dye + 500µL buffer). To each well of the white PCR plate was added the protein solution (10µL) followed by the dye/buffer solution (40µL). The plate is sealed and then run on the BioRad real-time-PCR machine, with temperatures ranging from 25°C to 95°C (1°C/30s) This raw data is imported into Origin, and a Boltzmann function fitted to the data. From this the T_M value can be calculated.

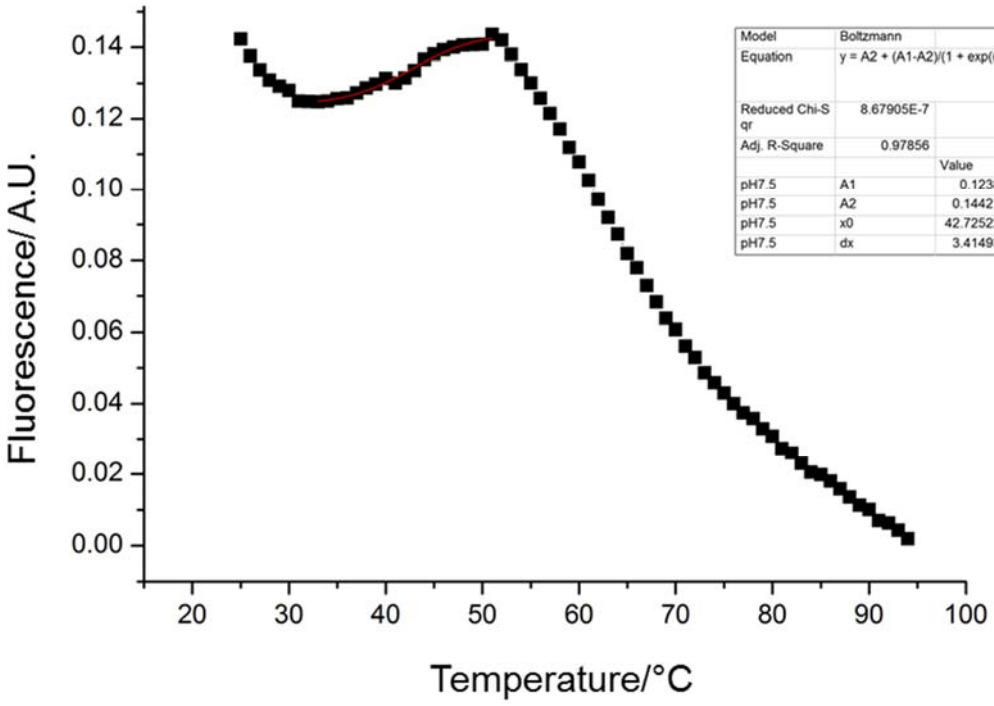


10mM NaPi pH 7.0



Model	Boltzmann		
Equation	y = A2 + (A1-A2)/(1 + exp((x-x0)/dx))		
Reduced Chi-Sqr	1.52393E-7		
Adj. R-Square	0.99773		
		Value	Standard Error
pH7.0	A1	0.11625	5.05256E-4
pH7.0	A2	0.14171	6.39656E-4
pH7.0	x0	44.08218	0.15792
pH7.0	dx	2.69984	0.19231

10mM NaPi pH 7.5



Model	Boltzmann		
Equation	y = A2 + (A1-A2)/(1 + exp((x-x0)/dx))		
Reduced Chi-Sqr	8.67905E-7		
Adj. R-Square	0.97856		
		Value	Standard Error
pH7.5	A1	0.1238	0.00124
pH7.5	A2	0.14421	0.00159
pH7.5	x0	42.72522	0.57564
pH7.5	dx	3.41493	0.72507

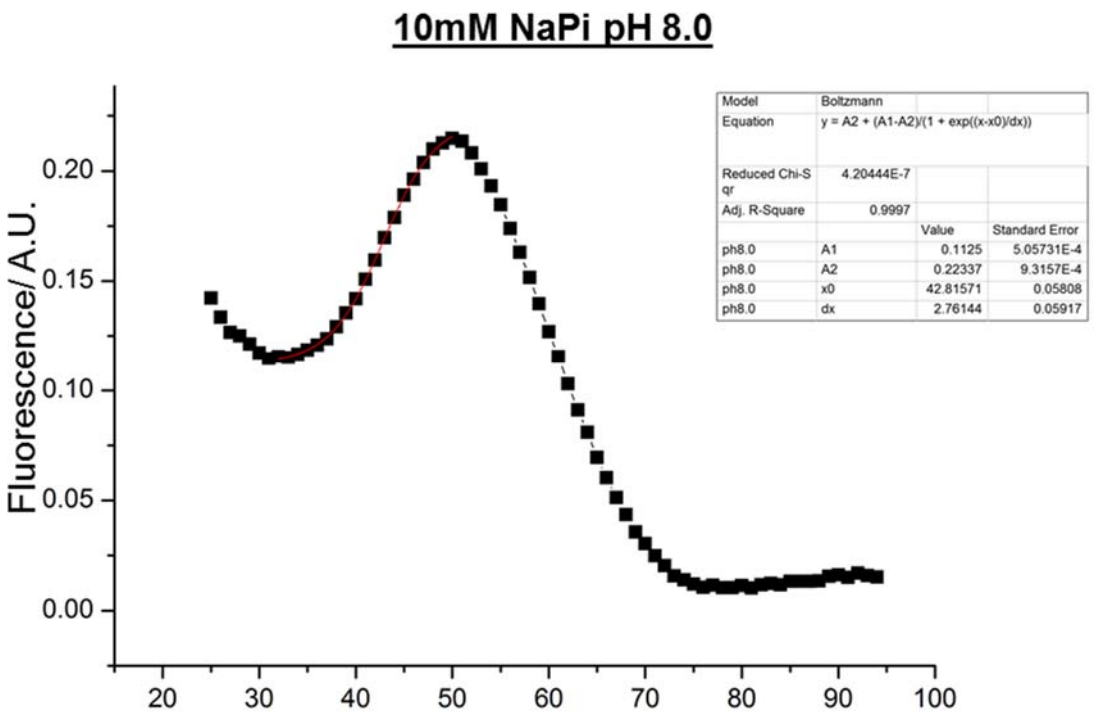


Figure E4. Raw plots for the DSF assay (average of 3 scans). To these plots the Boltzmann curve was fitted and the T_M calculated (X_O in the equation).

Standard MS/MS protocol

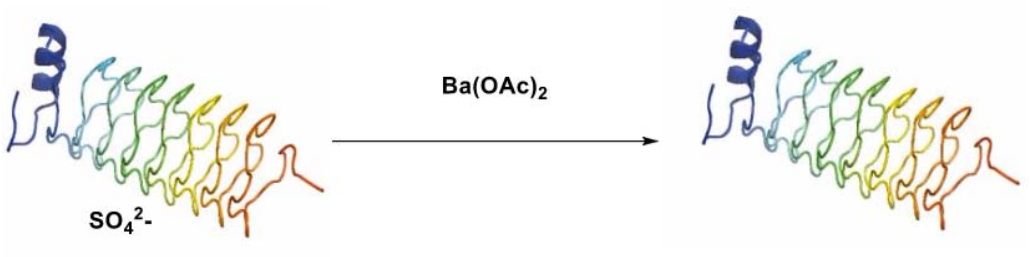
To 20µL of protein solution (0.3-0.6 $mgml^{-1}$) was added 30µL ambic (50mM, pH8.0), the reactions were heated to 95°C for 15 minutes and subsequently cooled to room temperature. Trypsin (0.1µgμL^{-1}) was added to the reaction mixture in a ratio of 1:50 (eq/eq) of protein. Reactions involving AspN used a 1:200 ratio (0.02µgμL^{-1}) of protease. The digestions were run overnight for 16 hours, at 37°C, with a rotation of 180rpm.

C18 ZipTip purification protocol

To the reaction was added 1µL of formic acid, and the reactions vortexed briefly. 10µL of the reaction mixture was loaded onto a pre-equilibrated ZIPTIP C18 tip, washed with wash buffer (H_2O + 0.1% formic acid) and finally eluted with 10µL of elution buffer (40-75% MeCN:60-

25% H_2O, 0.1% formic acid). The sample was diluted with wash buffer and submitted to MS/MS analysis.

Barium acetate reaction



To 100µL of the protein solution (10Å, P2P M61-M101 T201K, 50mM NaPi pH8.0) was added barium acetate (12.5mgml^{-1}) until a white precipitate was no longer observed to be formed after centrifugation in a table top centrifuge. MS was then run, indicating that the +98amu peak was a non-covalent sulfate adduct.

PAWS analysis and ExPASy peptide cutter

Analysis of the protein break points was conducted used the protein analysis workshop (PAWS) and ExPASy peptide cutter (http://web.expasy.org/peptide_cutter). This was based off both gel and LCT-MS data for the cleavage of the Np276 protein.

Np276 15A P2p no term met

[1-200] mass = 20822.9
mass to match = 15091.0 ± 10.7

P[1-144]A = 15090.7	H[8-152]S = 15084.7	P[15-160]D = 15093.7	G[17-162]H = 15090.7
Q[27-172]E = 15088.7	G[32-177]T = 15097.7	E[33-178]G = 15097.7	S[37-183]D = 15092.7
D[40-186]L = 15091.7	V[45-191]L = 15092.7		

1	P	S	S	H	H	H	H	H	H	S	S	G	L	V	P	R	G	S	H	I	D	V	E	A	L	R	Q	L	Y	A	30
31	A	G	E	R	D	F	S	I	V	D	L	R	G	A	V	L	E	N	I	N	L	S	G	A	I	L	H	G	A	M	60
61	L	D	E	A	N	L	Q	Q	A	N	L	S	R	A	D	L	S	G	A	T	L	N	G	A	D	L	R	G	A	N	90
91	L	S	A	A	D	L	S	D	A	I	L	D	N	A	I	L	E	G	A	I	L	D	E	A	V	L	N	Q	A	M	120
121	L	A	A	A	N	L	E	Q	A	I	L	S	H	A	N	I	R	E	A	D	L	S	E	A	N	L	E	A	A	D	150
151	L	S	G	A	D	L	A	I	A	D	L	H	Q	A	N	L	H	Q	A	A	L	E	R	A	N	L	T	G	A	N	180
181	L	E	D	A	N	L	E	G	T	I	L	E	G	G	N	N	N	L	A	K											200

[1-200] mass = 20822.9
mass to match = 12772.0 ± 9.1

P[1-122]A = 12772.2	H[8-130]I = 12763.2	L[13-135]N = 12780.3	V[14-136]I = 12780.3
I[38-162]H = 12774.2	M[60-183]D = 12778.0	A[64-188]G = 12774.0	A[74-199]A = 12775.0

1	P	S	S	H	H	H	H	H	H	S	S	G	L	V	P	R	G	S	H	I	D	V	E	A	L	R	Q	L	Y	A	30
31	A	G	E	R	D	F	S	I	V	D	L	R	G	A	V	L	E	N	I	N	L	S	G	A	I	L	H	G	A	M	60
61	L	D	E	A	N	L	Q	Q	A	N	L	S	R	A	D	L	S	G	A	T	L	N	G	A	D	L	R	G	A	N	90
91	L	S	A	A	D	L	S	D	A	I	L	D	N	A	I	L	E	G	A	I	L	D	E	A	V	L	N	Q	A	M	120
121	L	A	A	A	N	L	E	Q	A	I	L	S	H	A	N	I	R	E	A	D	L	S	E	A	N	L	E	A	A	D	150
151	L	S	G	A	D	L	A	I	A	D	L	H	Q	A	N	L	H	Q	A	A	L	E	R	A	N	L	T	G	A	N	180
181	L	E	D	A	N	L	E	G	T	I	L	E	G	G	N	N	N	L	A	K											200

Np276 15A P2p no term met

[1-200] mass = 20822.9
mass to match = 12457.0 ± 8.8

P[1-119]A = 12456.7	H[7-126]L = 12458.9	H[8-127]E = 12450.9	L[13-132]S = 12458.0
P[15-134]A = 12453.9	Q[27-146]L = 12448.9	L[28-147]E = 12449.8	F[36-157]A = 12458.8
R[42-163]Q = 12461.8	I[55-175]N = 12455.8	L[56-176]L = 12455.8	Q[68-189]T = 12448.6
A[79-200]K = 12459.7			

1	P	S	S	H	H	H	H	H	H	S	S	G	L	V	P	R	G	S	H	I	D	V	E	A	L	R	Q	L	Y	A	30
31	A	G	E	R	D	F	S	I	V	D	L	R	G	A	V	L	E	N	I	N	L	S	G	A	I	L	H	G	A	M	60
61	L	D	E	A	N	L	Q	Q	A	N	L	S	R	A	D	L	S	G	A	T	L	N	G	A	D	L	R	G	A	N	90
91	L	S	A	A	D	L	S	D	A	I	L	D	N	A	I	L	E	G	A	I	L	D	E	A	V	L	N	Q	A	M	120
121	L	A	A	A	N	L	E	Q	A	I	L	S	H	A	N	I	R	E	A	D	L	S	E	A	N	L	E	A	A	D	150
151	L	S	G	A	D	L	A	I	A	D	L	H	Q	A	N	L	H	Q	A	A	L	E	R	A	N	L	T	G	A	N	180
181	L	E	D	A	N	L	E	G	T	I	L	E	G	G	N	N	N	L	A	K											200

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	8	16 26 34 42 73 87 137 173
Asp-N endopeptidase	15	20 34 39 61 74 84 94 97 101 111 139 149 154 159 182
Asp-N endopeptidase + N-terminal Glu	29	20 22 32 34 39 46 61 62 74 84 94 97 101 106 111 112 126 137 139 142 146 149 154 159 171 181 182 186 191
CNBr	2	60 120
Chymotrypsin-high specificity (C-term to [FYW], not before P)	2	29 36
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	49	4 5 6 7 8 9 13 19 25 28 29 36 41 46 51 56 57 60 61 66 71 76 81 86 91 96 101 106 111 116 120 121 126 131 133 141 146 151 156 161 162 166 167 171 176 181 186 191 198
Clostripain	8	16 26 34 42 73 87 137 173
Formic acid	15	21 35 40 62 75 85 95 98 102 112 140 150 155 160 183
Glutamyl endopeptidase	14	23 33 47 63 107 113 127 138 143 147 172 182 187 192
Hydroxylamine	1	82
LysC	1	200
LysN	1	199
Pepsin (pH1.3)	65	12 24 25 27 35 40 41 45 46 50 51 55 56 60 61 65 66 70 71 76 80 81 85 86 90 91 95 96 100 101 105 106 110 111 115 116 120 121 125 126 130 131 140 141 145 146 150 151 155 156 160 161 165 166 170 171 176 180 181 185 186 190 191 197 198
Pepsin (pH>2)	66	12 24 25 27 29 35 40 41 45 46 50 51 55 56 60 61 65 66 70 71 76 80 81 85 86 90 91 95 96 100 101 105 106 110 111 115 116 120 121 125 126 130 131 140 141 145 146 150 151 155 156 160 161 165 166 170 171 176 180 181 185 186 190 191 197 198
Proteinase K	107	13 14 20 22 23 24 25 28 29 30 31 33 36 38 39 41 44 45 46 47 49 51 54 55 56 59 61 63 64 66 69 71 74 76 79 80 81 84 86 89 91 93 94 96 99 100 101 104 105 106 107 109 110 111 113 114 115 116 119 121 122 123

		124 126 127 129 130 131 134 136 138 139 141 143 144 146 147 148 149 151 154 156 157 158 159 161 164 166 169 170 171 172 174 176 177 179 181 182 184 186 187 189 190 191 192 198 199
Staphylococcal peptidase I	14	23 33 47 63 107 113 127 138 143 147 172 182 187 192
Thermolysin	72	12 19 24 27 29 30 37 38 43 44 45 48 50 53 54 55 58 59 60 65 68 70 73 78 80 83 88 90 92 93 99 100 103 104 105 108 109 110 114 115 118 119 120 121 122 123 125 128 129 130 133 135 145 148 153 156 157 158 163 165 168 169 170 173 175 178 180 185 189 190 197 198
Thrombin	1	16
Trypsin	9	16 26 34 42 73 87 137 173 200

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A/ M L A /A A N L E Q A I L S H A N I R E A D L S E A/ N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K

Figure E5. PAWS and ExPASy Peptide cutter analysis. This analysis indicated that the cleavage was non-enzymatic in nature.

Ramachandran analysis on Np276 and Rrf32

Ramachandran analysis of the Np276 and Rfr32 proteins were conducted using MatLab R2012a with the bioinformatics toolbox. The Ramachandran plot (shown for Np276) was computed using the following code. The resultant plot was manipulated manually.

```
>> Struct2j8k = getpdb('2j8k');

>> ramachandran(Struct2j8k,'glycine',true,'regions',true);
```

Protein library expression protocol

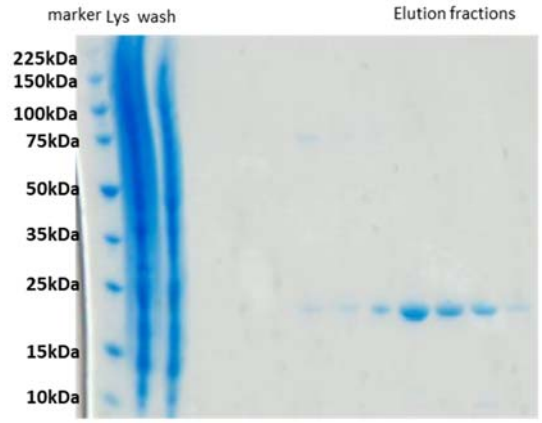
1μL of the plasmid (pet28d) encoding for the Np276 G2P ruler protein (exact mutant is shown in the data title below) was transformed into 20μL of B834(DE3) cells according to the manufacturer's instructions. After plating on kanamycin LB:agar plates and incubating overnight at 37°C, a healthy colony was used to inoculate an overnight culture of 20ml of freshly autoclaved LB media supplemented with 50μgL^{-1} kanamycin. After shaking at 180rpm and 37°C for 16 hours, the overnight culture was added to 1L of freshly autoclaved selenomet minimal media, supplemented with sterile 50μgL^{-1} kanamycin, 100ml filter sterilised selenomet media nutrient mix (5.1g/50ml) and 40μgL^{-1} sterile filtered L-methionine. The culture was grown at 37°C and 180rpm until an optical density of 1.2 was reached, approximately 4.5 hours in bevelled flasks. The culture was pelleted with a JA10 rotor, 7000 rpm , 10 minutes, 4°C and washed thrice with 1L of freshly autoclaved room temperature, selenomet media supplemented with 50μgL^{-1} sterile kanamycin but lacking methionine or azidohomoalanine. The cells were subsequently re-suspended in 1L of pre-warmed (37°C) selenomet media supplemented with 50μgL^{-1} sterile kanamycin, 100μgL^{-1} of sterilised azidohomoalanine and 100ml of sterile nutrient mix. After incubation at 37°C for 30 minutes and 18°C for 30 minutes, expression was induced with solid IPTG (240mg, 1mmol) and the culture incubated for 16 hours at 18°C.

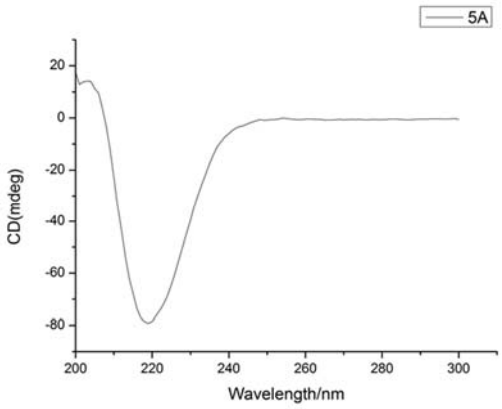
The cells were then harvested, 7000rpm, 10 minutes JA10, 4°C and flash frozen in liquid nitrogen. The cell pellet was added to approximately 30ml of nickel affinity chromatography loading buffer (20mM tris.HCl, 40mM imidazole, 500mM NaCl, 5% glycerol, pH8.0), and 10mg of lysozyme, DNAse and 2 tablets of Roche EDTA free protease inhibitor were added. After stirring gently for 30 mins at 4°C, the cells were sonicated (40% power, on 30s, off for 1

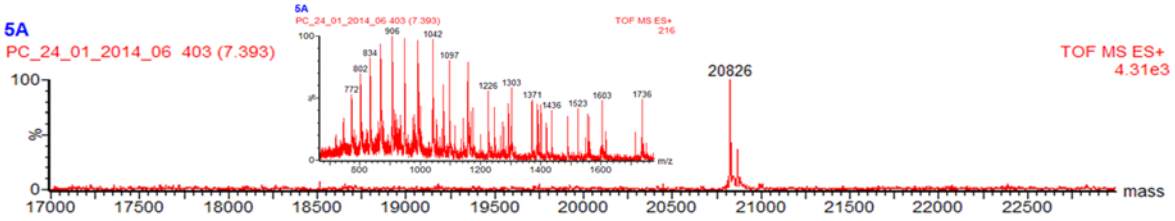
minute each) a total of 5 times. The cells were then pelleted, 20000rpm, JA25.5, 4°C, 30min and the cell lysate loaded onto a freshly prepared 5ml nickel Hi-Trap HP column. Following elution with a 4 step gradient on the AKTA-FPLC system (20CV 0% B, 5CV 5% B, 5CV 10% B, 5CV 15% B, 5CV 20% B, 10CV 100%), fractions containing protein were combined and subjected to an 85% ammonium sulfate cut. After stirring gently overnight (16 hours) the protein was precipitated out of solution (22k, 30min, JA25.5, 4°C) and the protein re-dissolved in a minimum of NaPi (50mM, pH8.0, 5% glycerol). The protein was then dialysed overnight (4°C, 4L) against the same buffer. ESI-TOF-MS, MS/MS and CD were all recorded, the protein filter sterilised (0.22μM) and flash frozen into aliquots.

5Å: G2P, M21I, K25/94/123A T81M, T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A M L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A N L A A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K







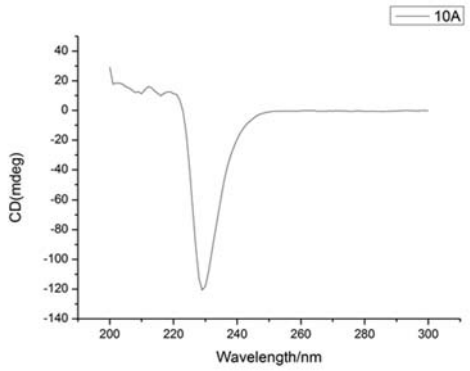
Associated Datafile: PC65A (50 - 1990 amu)
AspN:-/D

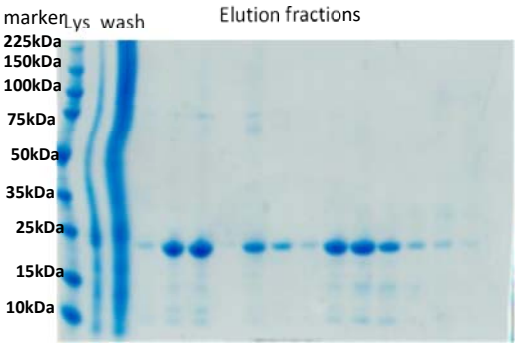
Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
D1	1-20	(-)PSSHHHHHHSSGLVPRGSHI(D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I)DVEALRQLYAAGER(D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-39	(R)DFSIV(D)	579.29	580.30	290.65	194.10	145.83
D4	40-61	(V)DLRGAVLENINLSGAILHGABL(D)	2271.24	2272.25	1136.63	758.09	568.82
D5	62-74	(L)DEANLQQANLSRA(D)	1428.70	1429.70	715.36	477.24	358.18
D6	75-84	(A)DLSGABLNGA(D)	942.45	943.46	472.23	315.16	236.62
D7	85-94	(A)DLRGANLSAA(D)	986.51	987.52	494.27	329.85	247.64
D8	95-97	(A)DLS(D)	333.15	334.16	167.58	112.06	84.30
D9	98-101	(S)DAIL(D)	430.24	431.25	216.13	144.42	108.57
D10	102-111	(L)DNAILEGAIL(D)	1027.55	1028.56	514.79	343.53	257.90
D11	112-139	(L)DEAVLNQANLAAANLEQAILSHANIREA(D)	2958.52	2959.52	1480.27	987.18	740.64
D12	140-149	(A)DLSEANLEAA(D)	1031.48	1032.48	516.75	344.83	258.88
D13	150-154	(A)DLSGA(D)	461.21	462.22	231.61	154.75	116.31
D14	155-159	(A)DLAIA(D)	501.28	502.29	251.65	168.10	126.33
D15	160-182	(A)DLHQANLHQAALERANLTGANLE(D)	2498.26	2499.27	1250.14	833.76	625.57
D16	183-200	(E)DANLEGTILEGGNNNLAK(-)	1841.91	1842.92	921.96	614.98	461.49

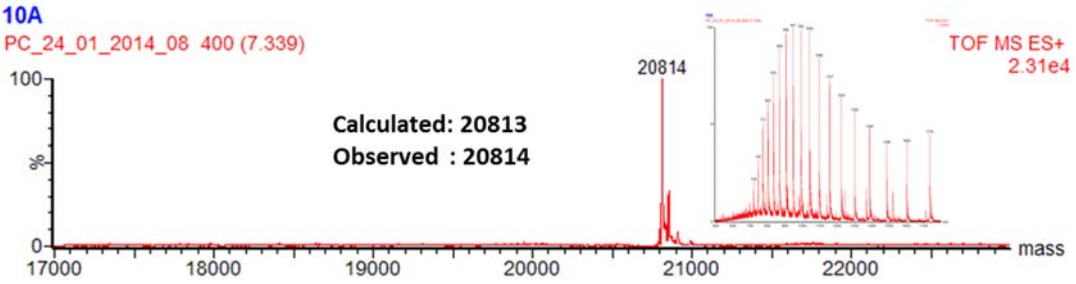
Figure E6. Data collected for the Np276 5Å AHA expression. This protein was expressed in a yield of 3.8mgL^{-1}.

10Å: G2P, M21I, K25/94/123A I101M, T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A M L D N A I L E G A I L D E A V L N Q A N L A A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K







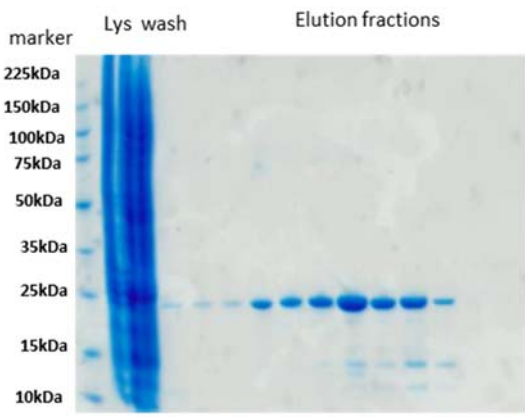
Associated Datafile: PC65B (50 - 1990 amu)
AspN:-/D

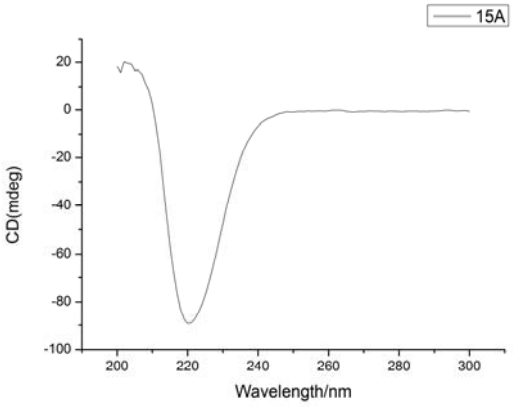
Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
▪D1	1-20	(-) PSSHHHHHHSSGLVPRG SHI (D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I) DVEALRQLYAAGER (D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-39	(R) DFSIV (D)	579.29	580.30	290.65	194.10	145.83
D4	40-61	(V) DLRGAVLENINLSGAIL HGABL (D)	2271.24	2272.25	1136.63	758.09	568.82
D5	62-74	(L) DEANLQQANLSRA (D)	1428.70	1429.70	715.36	477.24	358.18
D6	75-84	(A) DLSGATLNGA (D)	917.45	918.45	459.73	306.82	230.37
D7	85-94	(A) DLRGANLSAA (D)	986.51	987.52	494.27	329.85	247.64
D8	95-97	(A) DLS (D)	333.15	334.16	167.58	112.06	84.30
D9	98-101	(S) DABL (D)	443.21	444.22	222.61	148.75	111.81
D10	102-111	(L) DNAILEGAIL (D)	1027.55	1028.56	514.79	343.53	257.90
D11	112-139	(L) DEAVLNQANLAAANLEQ AILSHANIREA (D)	2958.52	2959.52	1480.27	987.18	740.64
D12	140-149	(A) DLSEANLEAA (D)	1031.48	1032.48	516.75	344.83	258.88
D13	150-154	(A) DLSGA (D)	461.21	462.22	231.61	154.75	116.31
D14	155-159	(A) DLAIA (D)	501.28	502.29	251.65	168.10	126.33
D15	160-182	(A) DLHQANLHQAALERANL TGANLE (D)	2498.26	2499.27	1250.14	833.76	625.57
D16	183-200	(E) DANLEGTILEGGNNNLA K(-)	1841.91	1842.92	921.96	614.98	461.49

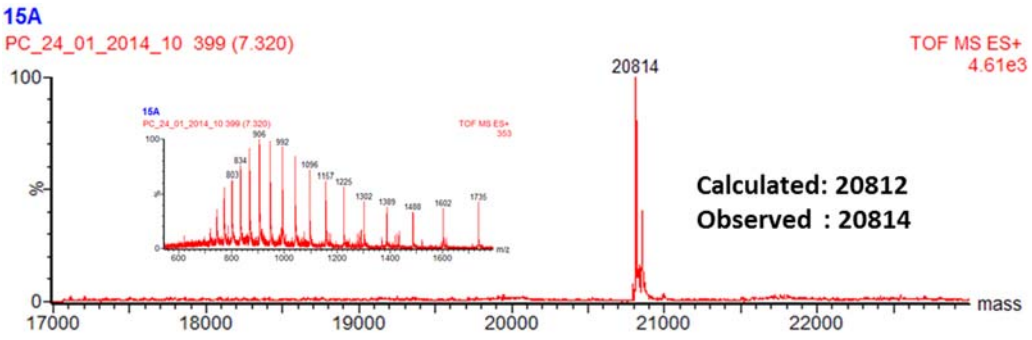
Figure E7. Data collected for the Np276 10Å AHA expression. This protein was expressed in a yield of 6.3mgL^{-1}.The unusual pattern seen on the gel is as a result of incorrect running of the AKTA FPLC.

15Å: G2P, M21I, K25/94/123A N121M, T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A M L A A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K







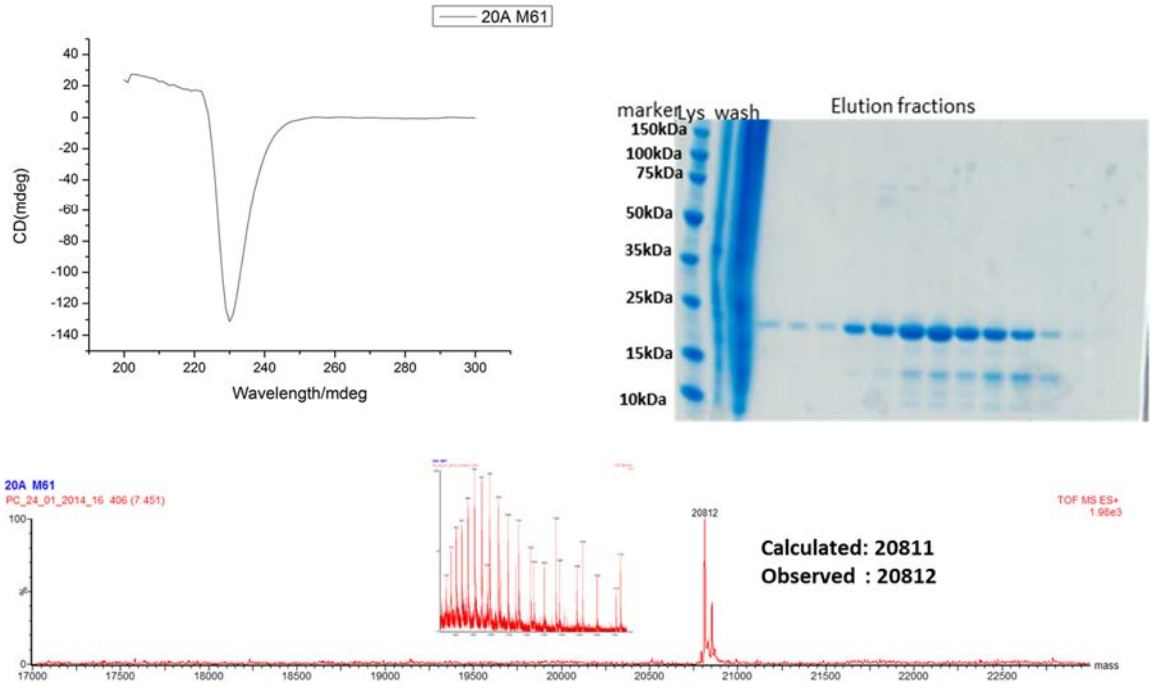
Associated Datafile: PC65C (50 - 1990 amu)
AspN:-/D

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
D1	1-20	(-) PSSHHHHHHSSGLVPRG SHI (D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I) DVEALRQLYAAGER (D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-39	(R) DFSIV (D)	579.29	580.30	290.65	194.10	145.83
D4	40-61	(V) DLRGAVLENINLSGAIL HGAJL (D)	2271.24	2272.25	1136.63	758.09	568.82
D5	62-74	(L) DEANLQQANLSRA (D)	1428.70	1429.70	715.36	477.24	358.18
D6	75-84	(A) DLSGATLNGA (D)	917.45	918.45	459.73	306.82	230.37
D7	85-94	(A) DLRGANLSAA (D)	986.51	987.52	494.27	329.85	247.64
D8	95-97	(A) DLS (D)	333.15	334.16	167.58	112.06	84.30
D9	98-101	(S) DAIL (D)	430.24	431.25	216.13	144.42	108.57
D10	102-111	(L) DNAILEGAIL (D)	1027.55	1028.56	514.79	343.53	257.90
D11	112-139	(L) DEAVLNQAJLAAANLEQ AILSHANIREA (D)	2970.53	2971.54	1486.27	991.18	743.64
D12	140-149	(A) DLSEANLEAA (D)	1031.48	1032.48	516.75	344.83	258.88
D13	150-154	(A) DLSGA (D)	461.21	462.22	231.61	154.75	116.31
D14	155-159	(A) DLAIA (D)	501.28	502.29	251.65	168.10	126.33
D15	160-182	(A) DLHQANLHQAALERANL TGANLE (D)	2498.26	2499.27	1250.14	833.76	625.57
D16	183-200	(E) DANLEGTILEGGNNNLA K(-)	1841.91	1842.92	921.96	614.98	461.49

Figure E8. Data collected for the Np276 15Å AHA expression. This protein was expressed in a yield of 3.7mgL^{-1}.

20Å M61: G2P, M21I, K25/94/123A D141M, T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A N L A A A N L E Q A I L S H A N I R E A M L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K



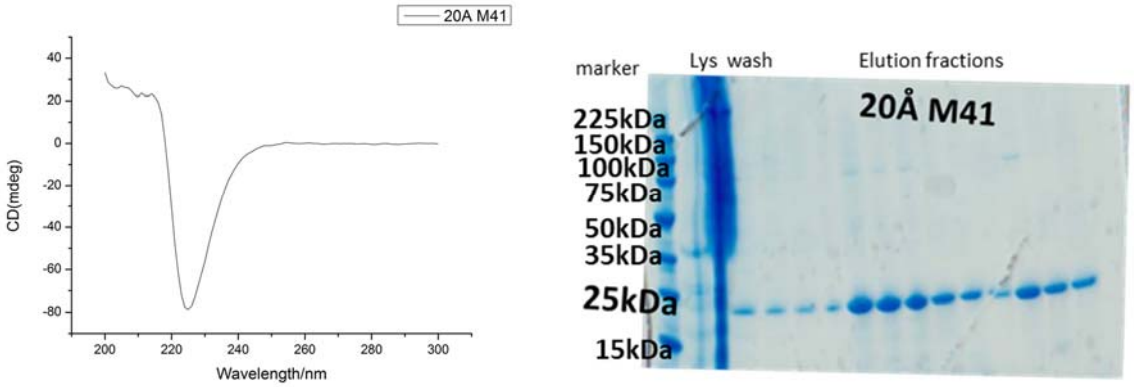
Associated Datafile: PC65D (50 - 1990 amu)
AspN:-/D

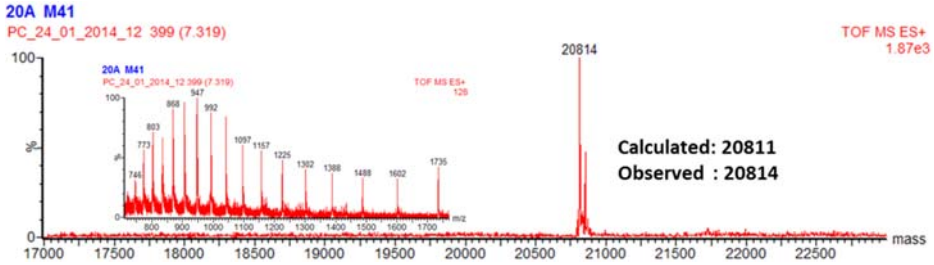
Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
▪D1	1-20	(-) PSSHHHHHHSSGLVPRG SHI (D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I) DVEALRQLYAAGER (D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-39	(R) DFSIV (D)	579.29	580.30	290.65	194.10	145.83
D4	40-61	(V) DLRGAVLENINLSGAIL HGABL (D)	2271.24	2272.25	1136.63	758.09	568.82
D5	62-74	(L) DEANLQQANLSRA (D)	1428.70	1429.70	715.36	477.24	358.18
D6	75-84	(A) DLSGATLNGA (D)	917.45	918.45	459.73	306.82	230.37
D7	85-94	(A) DLRGANLSAA (D)	986.51	987.52	494.27	329.85	247.64
D8	95-97	(A) DLS (D)	333.15	334.16	167.58	112.06	84.30
D9	98-101	(S) DAIL (D)	430.24	431.25	216.13	144.42	108.57
D10	102-111	(L) DNAILEGAIL (D)	1027.55	1028.56	514.79	343.53	257.90
D11	112-149	(L) DEAVLNQANLAAANLEQ AILSHANIREABLSEANLEA A (D)	3985.34	3986.35	1993.68	1329.46	997.34
D12	150-154	(A) DLSGA (D)	461.21	462.22	231.61	154.75	116.31
D13	155-159	(A) DLAIA (D)	501.28	502.29	251.65	168.10	126.33
D14	160-182	(A) DLHQANLHQAALERANL TGANLE (D)	2498.26	2499.27	1250.14	833.76	625.57
D15	183-200	(E) DANLEGTILEGGNNNLA K (-)	1841.91	1842.92	921.96	614.98	461.49

Figure E9. Data collected for the Np276 20Å 61 AHA expression. This protein was expressed in a yield of 6.7mgL^{-1}.

20Å M41: G2P, M21/61I, K25/94/123A D41M, N141M T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V M L R G A V L E N I N L S G A I L H G A I L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A M L A A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K





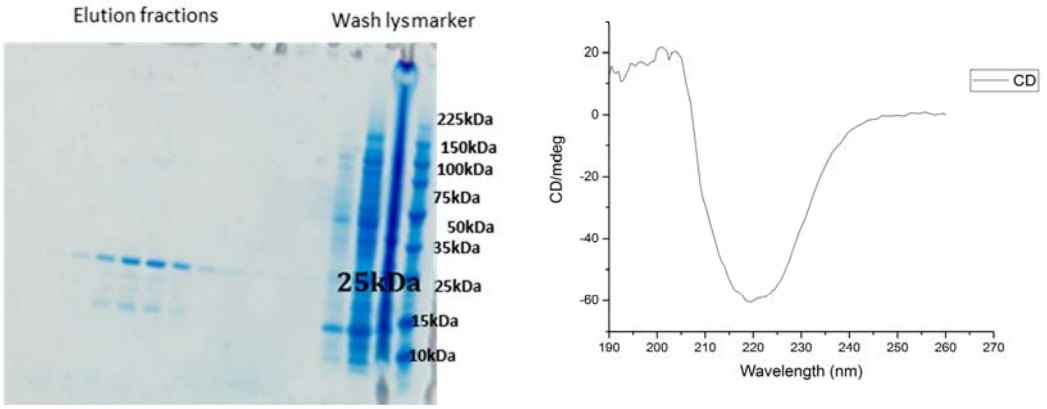
Associated Datafile: PC65E (50 - 1990 amu)
AspN:-D

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
D1	1-20	(-) PSSHHHHHHSSGLVPRG SHI (D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I) DVEALRQLYAAGER (D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-61	(R) DFSIVBLRGAVLENINL SGAILHGAIL (D)	2830.58	2831.59	1416.30	944.54	708.65
D4	62-74	(L) DEANLQQANLSRA (D)	1428.70	1429.70	715.36	477.24	358.18
D5	75-84	(A) DLSGATLNGA (D)	917.45	918.45	459.73	306.82	230.37
D6	85-94	(A) DLRGANLSAA (D)	986.51	987.52	494.27	329.85	247.64
D7	95-97	(A) DLS (D)	333.15	334.16	167.58	112.06	84.30
D8	98-101	(S) DAIL (D)	430.24	431.25	216.13	144.42	108.57
D9	102-111	(L) DNAILEGAIL (D)	1027.55	1028.56	514.79	343.53	257.90
D10	112-139	(L) DEAVLNQABLAAANLEQ AILSHANIREA (D)	2970.53	2971.54	1486.27	991.18	743.64
D11	140-149	(A) DLSEANLEAA (D)	1031.48	1032.48	516.75	344.83	258.88
D12	150-154	(A) DLSGA (D)	461.21	462.22	231.61	154.75	116.31
D13	155-159	(A) DLAIA (D)	501.28	502.29	251.65	168.10	126.33
D14	160-182	(A) DLHQANLHQAALERANL TGANLE (D)	2498.26	2499.27	1250.14	833.76	625.57
D15	183-200	(E) DANLEGTILEGGNNNLA K(-)	1841.91	1842.92	921.96	614.98	461.49
D1-2	1-34	(-) PSSHHHHHHSSGLVPRG	3776.11	3777.12	1889.06	1259.71	945.04

Figure E10. Data collected for the Np276 20Å 41 AHA expression. This protein was expressed in a yield of 7.8mgL^{-1}.

Single AHA: G2P, M21I

P S S H H H H H H S S G L V P R G S H I D V G K L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S K A D L S D A I L D N A I L E G A I L D E A V L N Q A N L K A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A T



Untitled
Associated Datafile: PKC86 (50 - 1990 amu)
AspN:-/D

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
D1	1-20	(-)PSSHHHHHHSSGLVPRGSHI(D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I)DVGKLRQLYAAGER(D)	1574.85	1575.86	788.43	525.96	394.72
D3	35-39	(R)DFSIV(D)	579.29	580.30	290.65	194.10	145.83
D4	40-61	(V)DLRGAVLENINLSGAILHGABL(D)	2271.24	2272.25	1136.63	758.09	568.82
D5	62-74	(L)DEANLQQANLSRA(D)	1428.70	1429.70	715.36	477.24	358.18
D6	75-84	(A)DLSGATLNGA(D)	917.45	918.45	459.73	306.82	230.37
D7	85-94	(A)DLRGANLSKA(D)	1043.57	1044.58	522.79	348.87	261.90
D8	95-97	(A)DLS(D)	333.15	334.16	167.58	112.06	84.30
D9	98-101	(S)DAIL(D)	430.24	431.25	216.13	144.42	108.57
D10	102-111	(L)DNAILEGAIL(D)	1027.55	1028.56	514.79	343.53	257.90
D11	112-139	(L)DEAVLNQANLKAANLEQAILSHANIREA(D)	3015.57	3016.58	1508.79	1006.20	754.90
D12	140-149	(A)DLSEANLEAA(D)	1031.48	1032.48	516.75	344.83	258.88
D13	150-154	(A)DLSGA(D)	461.21	462.22	231.61	154.75	116.31
D14	155-159	(A)DLAIA(D)	501.28	502.29	251.65	168.10	126.33
D15	160-182	(A)DLHQANLHQAALERANLTGANLE(D)	2498.26	2499.27	1250.14	833.76	625.57
D16	183-200	(E)DANLEGTILEGGNNNLAT(-)	1814.86	1815.87	908.44	605.96	454.72

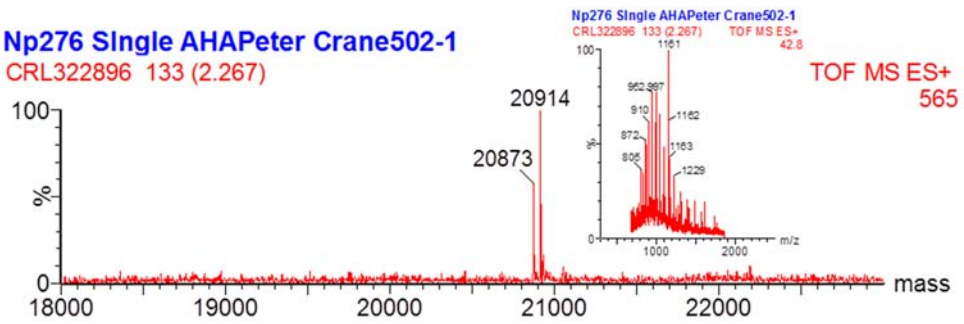
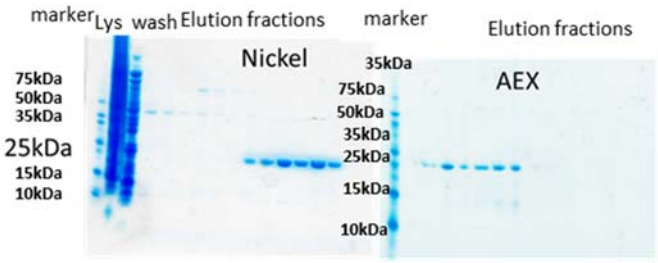
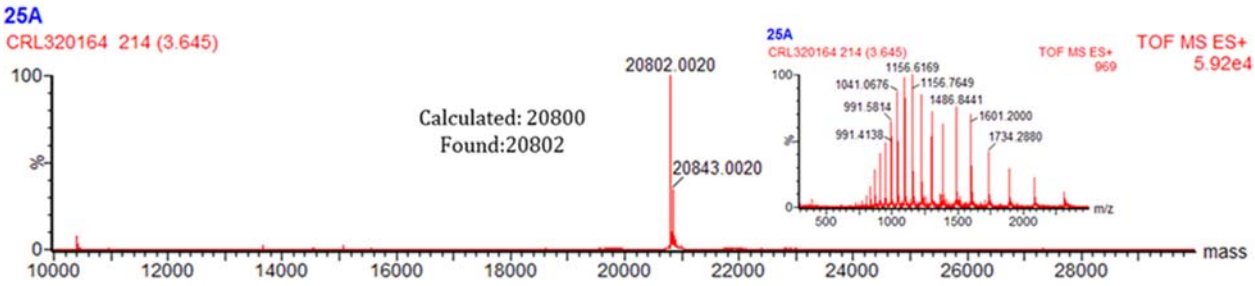


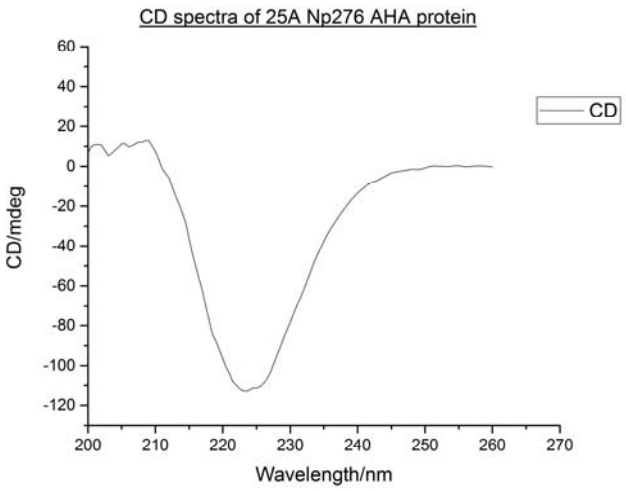
Figure E11. Data collected for the Np276 Single AHA expression. This protein was expressed in a yield of 11.42mgL^{-1}.

25Å: G2P, M21/61I, K25/94/123A, D41M, D141M, T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V M L R G A V L E N I N L S G A I L H G A I L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A N L A A A N L E Q A I L S H A N I R E A M L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K







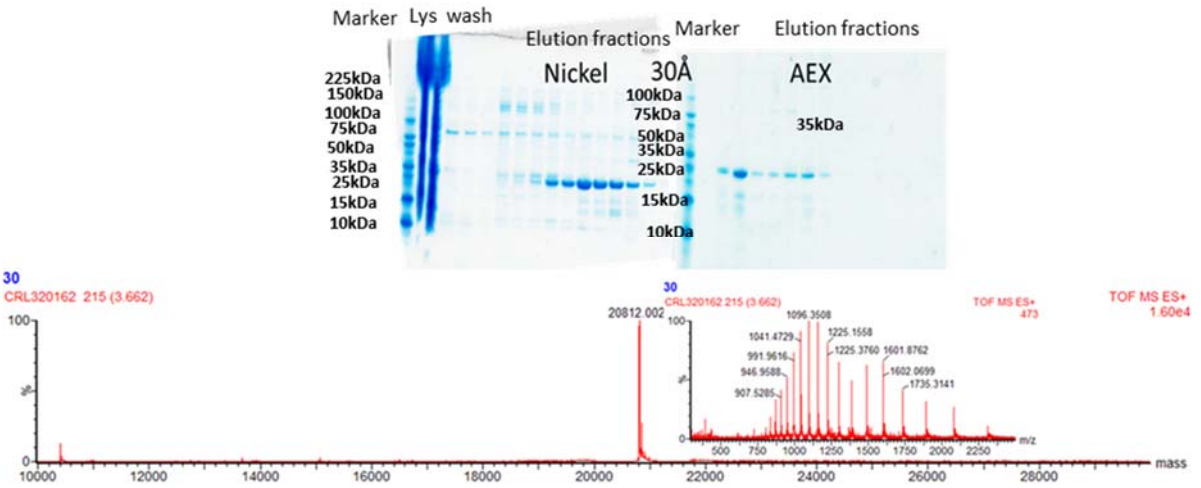
Associated Datafile: PC46F (50 - 1990 amu)
AspN:-/D

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
D1	1-20	(-) PSSHHHHHHSSGLVSRG SHI (D)	2192.05	2193.06	1097.03	731.69	549.02
D2	21-34	(I) DVEALRQLYAAGER (D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-61	(R) DFSIVBLRGAVLENINL SGAILHGAIL (D)	2830.58	2831.59	1416.30	944.54	708.65
D4	62-74	(L) DEANLQQANLSRA (D)	1428.70	1429.70	715.36	477.24	358.18
D5	75-84	(A) DLSGATLNGA (D)	917.45	918.45	459.73	306.82	230.37
▪D6	85-94	(A) DLRGANLSAA (D)	986.51	987.52	494.27	329.85	247.64
D7	95-97	(A) DLS (D)	333.15	334.16	167.58	112.06	84.30
D8	98-101	(S) DAIL (D)	430.24	431.25	216.13	144.42	108.57
D9	102-111	(L) DNAILEGAIL (D)	1027.55	1028.56	514.79	343.53	257.90
D10	112-149	(L) DEAVLNQANLAAANLEQ AILSHANIREABLSEANLEA A (D)	3985.34	3986.35	1993.68	1329.46	997.34
D11	150-154	(A) DLSGA (D)	461.21	462.22	231.61	154.75	116.31
D12	155-159	(A) DLAIA (D)	501.28	502.29	251.65	168.10	126.33
D13	160-182	(A) DLHQANLHQAALERANL TGANLE (D)	2498.26	2499.27	1250.14	833.76	625.57
D14	183-200	(E) DANLEGTILEGGNNNLA K (-)	1841.91	1842.92	921.96	614.98	461.49
D1-2	1-34	(-) PSSHHHHHHSSGLVSRG	3766.07	3767.08	1884.05	1256.37	942.53

Figure E12. Data collected for the Np276 25Å AHA expression. This protein was expressed in a yield of 5.4mgL^{-1}.

30Å:G2P, M21/61I, K25/94/123A, D41M, D161M, T201K

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V M L R G A V L E N I N L S G A I L H G A I L D E A N L Q Q A N L S R A D L S G A T L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A N L A A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A M L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K



Associated Datafile: PC46E (50 - 1990 amu)
AspN:-/D

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
■D1	1-20	(-)PSSHHHHHHSSGLVPRG SHI(D)	2202.07	2203.08	1102.04	735.03	551.53
D2	21-34	(I)DVEALRQLYAAGER(D)	1589.82	1590.82	795.92	530.95	398.46
D3	35-61	(R)DFSIVBLRGAVLENINL SGAILHGAIL(D)	2830.58	2831.59	1416.30	944.54	708.65
D4	62-74	(L)DEANLQQANLSRA(D)	1428.70	1429.70	715.36	477.24	358.18
D5	75-84	(A)DLSGATLNGA(D)	917.45	918.45	459.73	306.82	230.37
D6	85-94	(A)DLRGANLSAA(D)	986.51	987.52	494.27	329.85	247.64
D7	95-97	(A)DLS(D)	333.15	334.16	167.58	112.06	84.30
D8	98-101	(S)DAIL(D)	430.24	431.25	216.13	144.42	108.57
D9	102-111	(L)DNAILEGAIL(D)	1027.55	1028.56	514.79	343.53	257.90
D10	112-139	(L)DEAVLNQANLAAANLEQ AILSHANIREA(D)	2958.52	2959.52	1480.27	987.18	740.64
D11	140-149	(A)DLSEANLEAA(D)	1031.48	1032.48	516.75	344.83	258.88
D12	150-154	(A)DLSGA(D)	461.21	462.22	231.61	154.75	116.31
D13	155-182	(A)DLAIABLHQANLHQAAL ERANLTGANLE(D)	2992.56	2993.57	1497.29	998.53	749.15
D14	183-200	(E)DANLEGTILEGGNNNLA K(-)	1841.91	1842.92	921.96	614.98	461.49
D1-2	1-34	(-)PSSHHHHHHSSGLVPRG SHIDVEALRQLYAAGER(D)	3776.11	3777.12	1889.06	1259.71	945.04

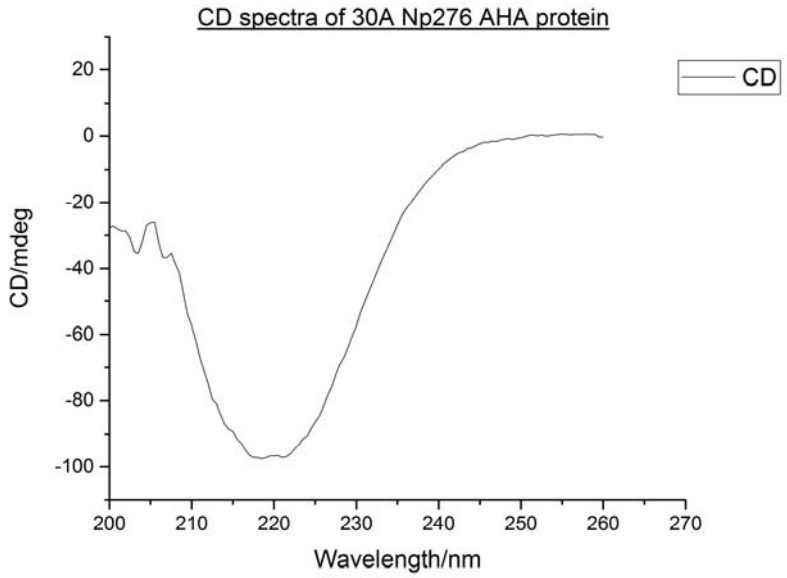


Figure E13. Data collected for the Np276 30Å AHA expression. This protein was expressed in a yield of 4.9mgL^{-1}.

2.4 Organic chemistry experimental

Chemicals were purchased from Sigma Aldrich and unless stated used as received. Solvents were purchased from Sigma Aldrich, Acros and/or Rathburn Chemicals. Dry solvents were purchased as anhydrous but dry alcohol-free DCM was obtained from an activated aluminium Grubbs still and stored under argon over 4Å molecular sieves. Air/moisture sensitive reactions were carried out, in furnace (450°C) dried glassware, under N_2, using dry disposable syringes and a manifold setup. Analytical TLC was run with silica gel, 60 F254, 0.25 mm pre-coated TLC plates (Merck) and visualised with UV_{254}, moly-dip (2.5g ammonium molybdate, 100 cm^3 water, 5.3cm^3 conc. H_2SO_4) or ninhydrin as stated. Flash column chromatography was carried out using silica gel (Merck, particle size 0.040–0.063 mm).

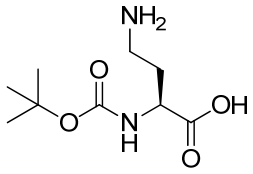
NMR spectra were recorded at 303K using a 400MHz (^{1}H), 300MHz (^{1}H) and 250 MHZ (^{1}H) and ^{13}C with 100MHz or 75MHz. Structural assignments are based upon a combination of DEPT, COSY, HMBC and HSQC. Signals are reported as singlet (s), broad singlet (br, s) doublet (d), doublet of doublets (dd), triplet (t), quartet (q) and multiplet (m). Coupling constants are all in Hz and peaks reported in ppm (δ) relative to residual solvent peaks.

LRMS: ESI-MS (positive and negative ion mode) were recorded on a Micromass platform spectrometer using MeOH as a solvent. IR was recorded on a Bio-Rad FTS-6000 FT spectrometer equipped with a DuraSamplIR II diamond ATR head (oils and solids were examined neat). Absorption maxima (ν_{max}) are recorded in wavenumbers (cm^{-1}).

Melting points were recorded using a Lecia Galen III hot stage microscope fitted with a Testo 720 thermocouple probe and are uncorrected. Optical rotations were measured on a Perkin Elmer 241 polarimeter at 589 nm (Na D line) with a path length of 1 dm and are reported with

implied units of 10^{-1} deg cm^2 g^{-1}. Concentrations (c) are given in g100 cm^{-3}. In this experimental, cm^3 is used as the standard unit for volume throughout the organic experimental, this is in contrast to the biological experimental wherein ml is used.

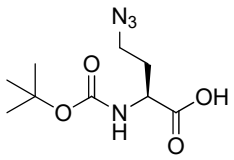
4-amino-2-((tert-butoxycarbonyl)amino)butanoic acid



To a vessel was charged N-Boc-glutamine (10.00g, 40.6mmol) and THF:H_2O ($80cm^3$:$20cm^3$) added. To this was added $PhI(OAc)_2$ (15.70g, 48.7mmol) over 10 minutes at 4°C. The reaction was stirred under argon for 16 hours, after which the reaction was concentrated *in vacuo*, H_2O ($30cm^3$) added and the aqueous layer washed thrice with EtOAc ($2x30cm^3$). The aqueous layer was then concentrated in *vacuo*, cooled on ice and washed with ice cold $CHCl_3$ ($30cm^3$) and Et_2O ($30cm^3$) before again being dried *in vacuo*. This gave the final compound as a pale yellow solid (6.32g, 71.6%).

LR-MS (ESI) 219.13 $[M+H]^+$, 241.12 $[M+Na]^+$; $[\alpha]^{20}_D$=-4.4 (c=0.224), 1:1 MeOH:H_2O), Lit. $[\alpha]^{20}_D$=-8.1 (c=1.0, H_2O)[1] R_F=0.833 (4:1:1, IPA:H_2O:AcOH); mp 208-210°C; IR (neat) ν_{max} (cm^{-1}): 3414 (N-H), 2977 (C-H), 1689 (C=O); ^{1}H NMR (400 MHz, D_2O, δ) δ 3.97 – 3.92 (br s, 1H, α-CH), 3.02 (t, *J* = 7.7 Hz, 2H, -*CH*$_2$$NH_2$), 2.14 – 2.03 (m, 1H, CH,), 1.97 – 1.87 (m, 1H,), 1.40 (s, 9H, -C*CH*$_3$); ^{13}C NMR (250 MHz, D_2O, δ) 178.35 (C=O), 157.95 (C=O), 81.68 (C, *C*$(CH_3)_3$), 53.9 (CH, Cα), 37.20(CH_2, CH_2NH_2), 30.17(2°, CH_2), 28 (CH_3, *C(CH*$_3$*)*$_3$).

Boc-azidohomoalanine

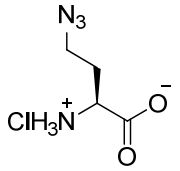


To a vessel was charged sodium azide (7.75g, 119.2mmol) and H_2O (13cm^3). DCM (26cm^3) was then added with vigorous stirring. The reaction was cooled to 0ºC and triflic anhydride (3.1cm^3 ,18.4mmol) added drop wise with vigorous stirring. The reaction was warmed to ambient temperature over 2 hours, before being extracted with DCM (2x26cm^3) and washed with saturated $NaHCO_3$ solution (43cm^3).

To a separate vessel was charged 4-amino-2-((tert-butoxycarbonyl) amino)butanoic acid (2.00g, 9.2mmol.) and H_2O (27cm^3) added. To this $CuSO_4.H_2O$ (23mg, 0.092mmol.) was added followed by MeOH (39cm^3) and K_2CO_3 (1.9g, 13.74mmol). With very vigorous stirring under a gentle flow of N_2, was added drop wise the DCM-triflic azide solution to the reaction in an ambient temperature water bath. The light green solution was then stirred at ambient temperature for 15 hours after which the reaction was carefully concentrated *in vacuo*, DCM (40cm^3) added and the layers separated. The aqueous layer was washed twice with more DCM (2x40cm^3); The organics combined, dried ($MgSO_4$), filtered (gravity) and concentrated *in vacuo* yielding Boc-azidohomoalanine as a yellow oil (1.92g, 85.7%).

$[\alpha]^{20}_D$=+19.1 (c=0.261, MeOH), Lit. $[\alpha]^{20}_D$=+19.0 (c=1.0, $CHCl_3$)[1], LR-MS (ESI) 267.10 $[M+Na]^+$ 511.20 $[2M+Na]^+$; R_F=0.88 (4:1:1, IPA:H_2O:AcOH); IR (neat) ν_{max} (cm^{-1}): 2980 (C-H), 2099 (N_3), 1691cm^{-1} (C=O); ^{1}H NMR (300 MHz, D_2O, δ) 4.28 – 4.14 (m, 1H, α-CH), 3.54 – 3.40 (m, 2H, CH_2N_3), 2.15 (m, 1H, CH), 1.95 (m,1H, CH), 1.45 (s, 9H). ^{13}C NMR (126 MHz, D_2O, δ) 176.93 (C=O), 158.40 (C=O), 82.30(C, $C(CH_3)_3$), 70.50(CH, Cα), 48.30(CH_2 , CH_2N_3), 30.28(CH_2, *CH_2*CH_2N_3), 28.27(CH_3, C(*CH_3*)$_3$);

Azidohomoalanine



To a vessel at 0ºC was charged the Boc-azidohomoalanine (1.55g, 6.35 mmol), DCM (10.1cm^3) and TFA (0.9cm^3, 11.75 mmol). The reaction was stirred at 0°C for 2 hours and then concentrated in *vacuo*. The off white solid formed washed repeatedly with dilute 1M HCl (10x10cm^3) yielding the HCL salt as an off white yellowy solid (1.07g, 94%).

m.p 228-230°C Lit. m.p 231-232°C $[\alpha]^D$=+14.8 (c=0.108, 1:1 MeOH:H_2O), Lit. $[\alpha]^{20}_D$=+6.6 (c=1.0, H_2O)[1] IR (neat) ν_{max} (cm^{-1}): 3400 (br) (NH_2) ~3050 (OH) 2109.78 (N_3), 1736.84 cm^{-1} (C=O). LR-MS (ESI) 167.07 $[M+Na]^+$ ^{1}H NMR (300 MHz, D_2O, δ) 4.16 (t, *J* = 6.3 Hz, 1H, Hα), 3.62 (t, *J* = 6.4 Hz, 2H, CH_2N_3), 2.33 – 2.08 (m, 2H, CH_2).^{13}C NMR (75 MHz, D_2O, δ) 171.70 (C=O), 51.03(CH, Cα), 47.01(CH_2 , *C*H_2N_3), (CH_2, *C*$H_2CH_2N_3$) 28.99.

2.5 Bibliography

(1) Fernández-González, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. a.; Errey, J. C.; Davis, B. G. *Chem. Sci.* **2010**, *1* (6), 709.

(2) Kleywegt, G. J.; Jones, T. A. *Structure* **1996**, *4* (12), 1395–1400.

(3) Hooft, R. W.; Sander, C.; Vriend, G. *Comput. Appl. Biosci.* **1997**, *13* (4), 425–430.

(4) Zhou, A. Q.; Hern, C. S.; Regan, L. *Protein Sci.* **2011**, *20* (7), 1166–1171.

(5) Link, a J.; Vink, M. K. S.; Tirrell, D. a. *Nat. Protoc.* **2007**, *2* (8), 1879–1883.

(6) Roth, S.; Drewe, W. C.; Thomas, N. R. **2010**, *5* (12), 1967–1973.

(7) Johansson, H.; Pedersen, D. S. *European J. Org. Chem.* **2012**, No. 23, 4267–4281.

(8) Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G. **2009**, 7798–7802.

(9) Geoghegan, K. F.; Dixon, H. B.; Rosner, P. J.; Hoth, L. R.; Lanzetti, a J.; Borzilleri, K. a; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, a S.; Kamath, a V; Stroh, J. G. *Anal. Biochem.* **1999**, *267* (1), 169–184.

(10) Stincone, A.; Prigione, A.; Cramer, T.; Wamelink, M. M. C.; Campbell, K.; Cheung, E.; Olin-Sandoval, V.; Grüning, N.-M.; Krüger, A.; Tauqeer Alam, M.; Keller, M. a.; Breitenbach, M.; Brindle, K. M.; Rabinowitz, J. D.; Ralser, M. *Biol. Rev.* **2014**, *90* (3), 927–963.

(11) Sprenger, G. A. *Arch. Microbiol.* **1995**, *164* (5), 324–330.

(12) Meier, S.; Jensen, P. R.; Duus, J. Ø. *Chembiochem* **2012**, *13* (2), 308–310.

(13) Aon, J. C.; Caimi, R. J.; Taylor, A. H.; Lu, Q.; Oluboyede, F.; Dally, J.; Kessler, M. D.; Kerrigan, J. J.; Lewis, T. S.; Wysocki, L. a; Patel, P. S. *Appl. Environ. Microbiol.* **2008**, *74* (4), 950–958.

(14) Zhao, J.; Shimizu, K. *J. Biotechnol.* **2003**, *101* (2), 101–117.

(15) Peskov, K.; Mogilevskaya, E.; Demin, O. *FEBS J.* **2012**, *279* (18), 3374–3385.

(16) Masip, L.; Veeravalli, K.; Georgiou, G. *Antioxid. Redox Signal.* **2006**, *8* (5–6), 753–762.

(17) Arenas, F. A.; Díaz, W. A.; Leal, C. A.; Pérez-Donoso, J. M.; Imlay, J. A.; Vásquez, C. C. *Biochem. Biophys. Res. Commun.* **2010**, *398* (4), 690–694.

(18) Tsouko, E.; Khan, A.; White, M.; Han, J.; Shi, Y.; Merchant, F.; Sharpe, M.; Xin, L.; Frigo, D. *Oncogenesis* **2014**, *3*.

(19) Jung, I. L.; Kim, I. G. *Cell Biol. Toxicol.* **2003**, *19* (1), 29–41.

(20) Jung, I. L.; Kim, I. G. **2003**, *301*, 915–922.

(21) Jung, I. L.; Kim, I. G. *Environ. Toxicol. Pharmacol.* **2003**, *15* (1), 19–26.

(22) Delihas, N.; Forst, S. *J. Mol. Biol.* **2001**, *313* (1), 1–12.

(23) Touati, D. *Arch. Biochem. Biophys.* **2000**, *373* (1), 1–6.

(24) Kershaw, C. J.; Brown, N. L.; Constantinidou, C.; Patel, M. D.; Hobman, J. L. *Microbiology*

2005, *151* (Pt 4), 1187–1198.

(25) Pham, a. N.; Xing, G.; Miller, C. J.; Waite, T. D. *J. Catal.* **2013**, *301*, 54–64.

(26) Manchado, M.; Michan, C.; Pueyo, C. *J. Bacteriol.* **2000**, *182* (23), 6842–6844.

(27) Nunoshiba, T.; Hidalgo, E.; Cuevas, C. F. A.; Demple, B. *J. Bacteriol.* **1992**, *174* (19), 6054–6060.

(28) Wu, J.; Weiss, B. *J. Bacteriol.* **1991**, *173* (9), 2864–2871.

(29) Chubiz, L. M.; Glekas, G. D.; Rao, C. V. *J. Bacteriol.* **2012**, *194* (18), 4867–4875.

(30) Tyson, R. L.; Perron, J.; Sutherland, G. R. *Neuroreport* **2000**, *11* (9), 1845–1848.

(31) Ferguson, G. P.; Booth, I. R. *J. Bacteriol.* **1998**, *180* (16), 4314–4318.

(32) Smirnova, G. V.; Muzyka, N. G.; Glukhovchenko, M. N.; Oktyabrsky, O. N. *Free Radic. Biol. Med.* **2000**, *28* (7), 1009–1016.

(33) Vergauwen, B.; Pauwels, F.; Vaneechoutte, M.; Van Beeumen, J. J. *J. Bacteriol.* **2003**, *185* (5), 1572–1581.

(34) Helbig, K.; Bleuel, C.; Krauss, G. J.; Nies, D. H. *J. Bacteriol.* **2008**, *190* (15), 5431–5438.

(35) Heath, B.; George, W.; Ann, A. Ep 1 127 605 b1 (12), 2006.

(36) Giglione, C.; Boularot, A.; Meinnel, T. *Cell. Mol. Life Sci.* **2004**, *61* (12), 1455–1474.

(37) Varshavsky, a. *Genes to cells* **1997**, *2* (1), 13–28.

(38) Varshavsky, a. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (22), 12142–12149.

(39) Schuenemann, V. J.; Kralik, S. M.; Albrecht, R.; Spall, S. K.; Truscott, K. N.; Dougan, D. a; Zeth, K. *EMBO Rep.* **2009**, *10* (5), 508–514.

(40) Tayyab, S.; Siddiqui, M. U.; Ahmad, N. *Biochem. Educ.* **1995**, *23* (3), 162–164.

(41) Privalov, P. L.; Dragan, A. I. *Biophys. Chem.* **2007**, *126* (1–3), 16–24.

(42) Kopec, J.; Schneider, G. *J. Struct. Biol.* **2011**, *175* (2), 216–223.

(43) Boivin, S.; Kozak, S.; Meijers, R. *Protein Expr. Purif.* **2013**, *91* (2), 192–206.

(44) Niesen, F. H.; Berglund, H.; Vedadi, M. *Nat. Protoc.* **2007**, *2* (9), 2212–2221.

(45) Faroongsarng, D.; Kongprasertkit, J. *AAPS PharmSciTech* **2014**, *15* (2), 465–471.

(46) Petrotchenko, E. V.; Serpa, J. J.; Hardie, D. B.; Berjanskii, M.; Suriyamongkol, B. P.; Wishart, D. S.; Borchers, C. H. *Mol. Cell. Proteomics* **2012**, *11* (7), M111.013524-M111.013524.

(47) Khrapunov, S.; Cheng, H.; Hegde, S.; Blanchard, J.; Brenowitz, M. *J. Biol. Chem.* **2008**, *283* (52), 36290–36299.

(48) Khrapunov, S.; Brenowitz, M. *Biophys. Chem.* **2011**, *159* (1), 33–40.

(49) Khrapunov, S. *Anal. Biochem.* **2009**, *389* (2), 174–176.

(50) Buchko, G. W.; Robinson, H.; Pakrasi, H. B.; Kennedy, M. a. *J. Struct. Biol.* **2008**, *162* (1), 184–192.

(51) Buchko, G. W.; Robinson, H.; Ni, S.; Pakrasi, H. B.; Kennedy, M. a. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2006**, *62* (Pt 12), 1251–1254.

(52) Xiong, X.; Bromley, E. H. C.; Oelschlaeger, P.; Woolfson, D. N.; Spencer, J. *Nucleic Acids*

Res. **2011**, *39* (9), 3917–3927.

(53) Johnson, W. C. *Proteins* **1990**, *7* (3), 205–214.

(54) Greenfield, N. J. *Nat. Protoc.* **2006**, *1* (6), 2876–2890.

(55) Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochim. Biophys. Acta* **2005**, *1751* (2), 119–139.

(56) Huynh, K.; Partch, C. L. *Curr. Protoc. protein Sci.* **2015**, *79* (February), 28.9.1-28.9.14.

(57) Chowdhury, S.; Katta, V.; Beavis, R.; Chait, B. *J. Am. Soc. Mass Spectrom.* **1990**, *1* (5), 382–388.

(58) Soppa, J. *Archaea* **2010**, *2010*.

(59) Kollipara, L.; Zahedi, R. P. *Proteomics* **2013**, *13* (6), 941–944.

(60) Olajuyigbe, F. M.; Demitri, N.; Ajele, J. O.; Maurizio, E.; Randaccio, L.; Geremia, S. *ACS Med. Chem. Lett.* **2010**, *1* (6), 254–257.

(61) Sun, S.; Zhou, J.-Y.; Yang, W.; Zhang, H. *Anal. Biochem.* **2014**, *446*, 76–81.

(62) Olsen, J. V; Ong, S.-E.; Mann, M. *Mol. Cell. Proteomics* **2004**, *3* (6), 608–614.

(63) Vallée, R.; Schombs, M. W.; Balaban, Z.; Colyer, J.; Davis, B. G. *Chem. Commun.* **2016**, *52* (14), 2–5.

(64) Alper, P. B.; Hung, S. C.; Wong, C. H. *Tetrahedron Lett.* **1996**, *37* (34), 6029–6032.

Chapter 3

Synthesis of carbohydrate ligands for use in bio-conjugation reactions

Contents Page

3.1 Introduction and aim

Carbohydrate chemistry as a field has existed for over 100 years, with the foundations of modern organic chemistry being laid upon the pioneering studies of Emil Fischer, Haworth and others who worked on the study of sugars. Despite this long history, there still exists many challenges within the field: the rapid synthesis of complex glycans on scale among them. One area of carbohydrate chemistry which continues to receive attention from the synthetic community due to its inherent synthetic challenge and the biological applications of the resultant oligosaccharides, is sialyation.[1]

The problems in chemical sialylation are multifold: the deactivation of the anomeric centre by the geminal acid, a propensity to eliminate to the glycal and the lack of neighbouring group participation to control selectivity.[2] Despite this, much work has been done on the synthesis of direct Neu-5-Ac donors, included in this range are thio-sialosides[3], chlorides[4], phosphites, xantphates[5] and imidates. Yields are highly sensitive to small changes in solvent (RCN for α), temperature, promoter and stoichiometry of donor/acceptor.[2,3] There are also newer indirect 5 substituted donors and C-3 modified donors, which in many cases provide higher levels of selectivity and yields.[2] The caveat being, that the number of steps to functionalise these donors outweighs the perceived synthetic advantages. Within the wider literature, the majority of sialosides are now synthesised by chemoenzymatic methods.[6] By combining enzymes such as NANA, Pd26ST, PMST and NMCSS a huge variety of unnatural and complex glycans have been synthesised. [7–15]

In this chapter, the synthesis of the complex carbohydrate ligand propargyl-(5-acetimido-3,5-dideoxy-d-glycero-D-galacto-2-nonulopyranosylonic acid-(α2-3)-D-galactopyranosyl-(1-4)-2-acetimido-2-deoxy-D-glucopyranoside (sialylactosamine or SiaLacNAc) is attempted through the use of a chemo-enzymatic methodology. This α-2,3 sialylactosamine binds

strongly to Siglec-1, the receptor to be probed with the molecular ruler and is what has motivated the choice of sugar. This work is based upon a variety of literature reports[16] and a synthetic route conducted by Kastren *et al*.[17]

3.2 Results and Discussion

3.2.1 Heyns rearrangement synthetic route

The Heyns rearrangement, an alpha ketol rearrangement analogous to the Amadori reaction has been previously used to prepare N-acetyl-D-lactosamine on significant scale, via a method which doesn't utilise any column chromatography steps.[18,19] Given the low cost of start materials this method was trialled.

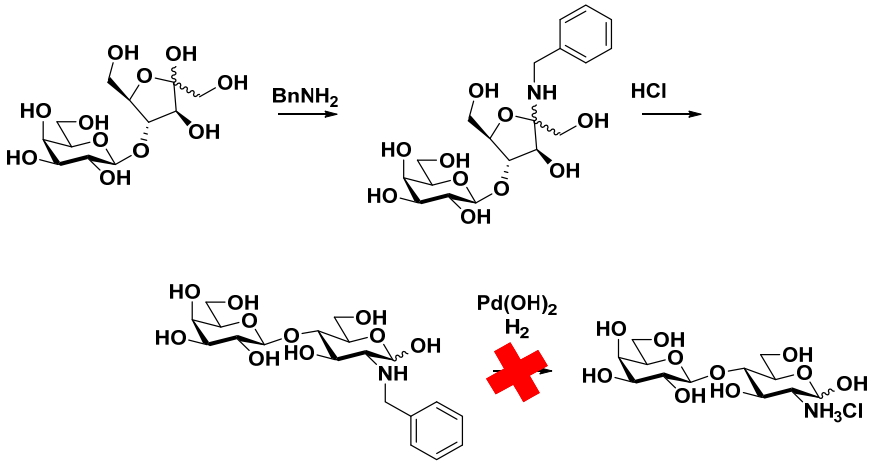


Figure 1. Synthesis of lactosamine via the Heyns rearrangement. MS analysis showed the formation of the intermediate product (Figure 2).

In the first stage of the reaction, lactulose was heated in the presence of neat benzylamine to give the hemiaminal intermediate as a viscous oil after separation; the removal of excess

benzylamine at this stage also proved extremely difficult. This was followed by the acid-mediated Heyns rearrangement (which was monitored by TLC) which after dehydrogenation with Pearlman's catalyst yielded traces of product by MS analysis.[20,21] Pearlman's catalyst was chosen over Pd/C as it displays higher reactivity towards N-debenzylation.[22] Despite numerous reactions, attempts to push this hydrogenation to completion were unsuccessful, as judged by MS analysis. This problem was also reported in the literature.

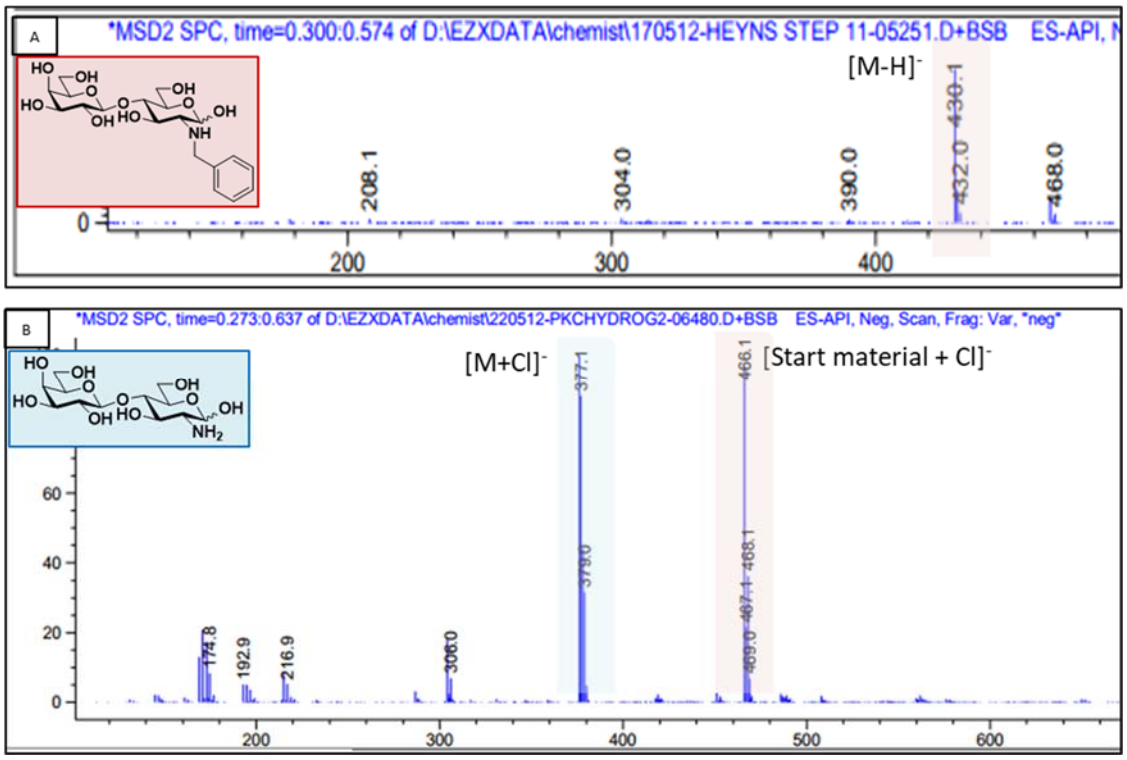
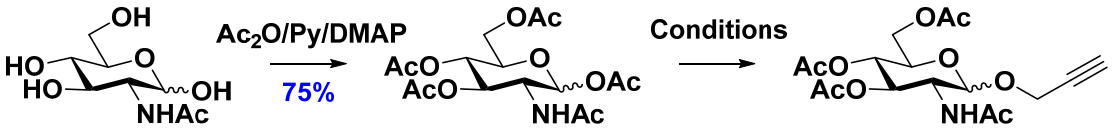


Figure 2. (A) The intermediate product formed during the Heyns rearrangement (shown via a negative mode MS and highlighted in red). (B) The failed debenzylation showing the presence of the product and the start material. As this reaction was run in HCl the $[M+Cl]^-$ ions are seen in the negative mode.

Due to the long reaction times for the deprotection step, the focus moved to alternative routes as described by Kastren *et al.*[17] After the conclusion of our studies, an improved procedure by Shan *et al* was also published that sought to solve many of the work-up problems encountered in this Heyns rearrangement synthetic route.[23]

3.2.2. Linear synthesis

The linear synthesis of SiaLacNAc was attempted starting from the constituent monosaccharides. After acetylation of N-acetyl glucosamine to give the full acetylated sugar as a 9:1 α/β ratio, the sulfuric acid-silica microwave glycosylation as reported by Roy *et al* was investigated. This gave rise to multiple products on TLC.[24] ^{1}H NMR indicated that the predominant product was the unreacted start material. An analogous method using rare earth triflates under microwave irradiation also gave rise to multiple products, as well as the product in around 46% yield.[25,26]



Method	Isolated Yield	Notes
Sulfuric acid-silica	0%	No reaction
$Yb(OTf)_3$/MW/DCM	46%	Many side products

Figure 3. Formation of the O-propargyl product from a crude 9:1 alpha to beta mixture of per-acetylated GlcNAc.

The β-anomer of per acetylated GlcNAc was thus prepared by a 4 step procedure via the *p*-methoxybenzyl imine (1-4) (which promotes formation of the β-derivative).[26] It was later found that by utilising montromotrilite K-10 clay, on 40g scale the β-anomer could be prepared in one step in an isolated yield of 27%.[27] With the β anomer in hand, the O-propargyl GlcNAc (5) derivative could be prepared by a rare earth triflate catalysed glycosylation followed by deacetylation under Zemplen conditions (6). The next step was the protection of the primary hydroxyl group with a bulky silyl protecting group leading to the selected acceptors (7) and (8).[28]

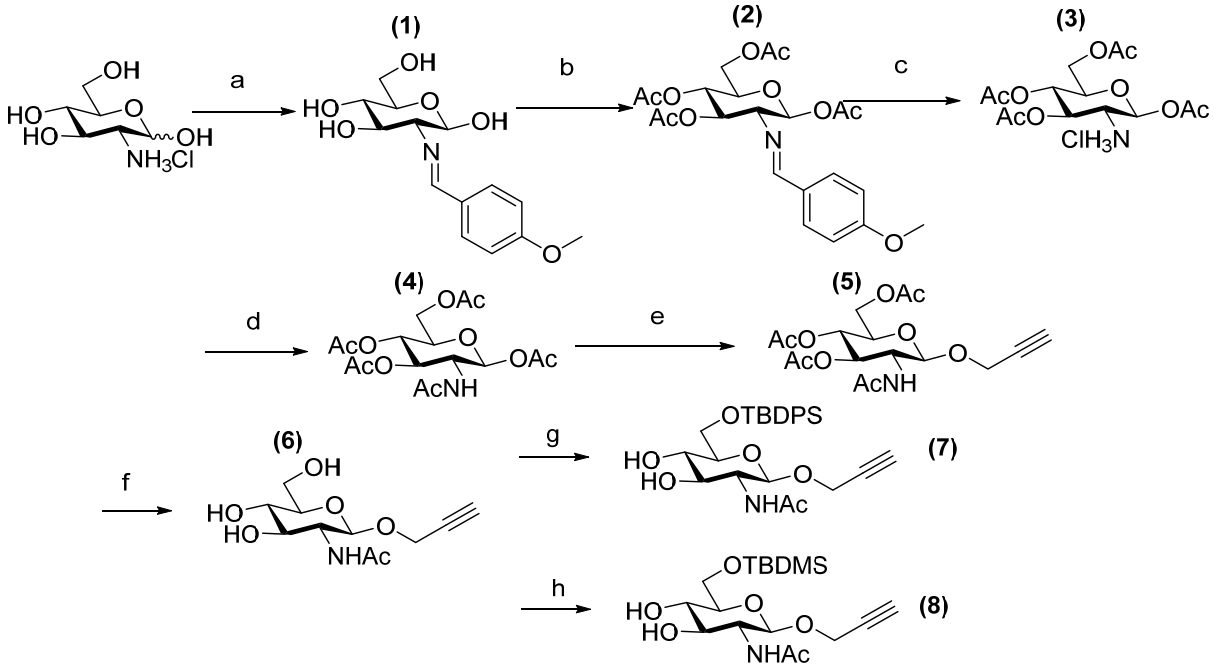


Figure 4. Synthesis of the 6 protected acceptor(s). a) NaOH (1M) p-anisaldehyde 87%, b) Ac_2O/Py 78% c) 5N HCl 88% d) Et_3N/Ac2O 86% e)$Yb(OTf)_3$, DCM, propargyl alcohol reflux 69% f) NaOMe/MeOH quant g) NaOMe/MeOH, TBDPSCl Py DMAP 78% two steps h) TBDMSCl Py/DMAP 67%

The trichloroacetamide galactose donor was synthesised in 2 steps from per acetylated galactose with BnH_2 followed by DBU/TCA, this approach negated the use of hydrazine (Figure 5).

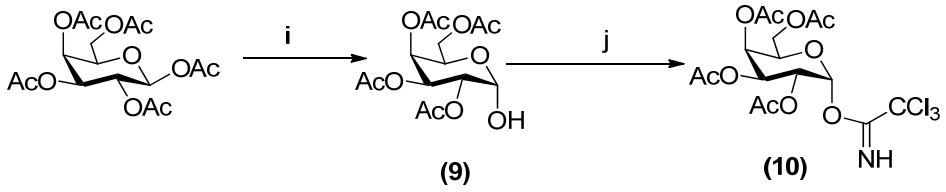


Figure 5. Synthesis of the 6 protected acceptor(s) i)BnH_2/THF 82% j) DBU, CCl_3CN,DCM 67%

The Lewis acid mediated glycosylations proceeded smoothly in the case of TBDPS giving product with yields of 55 % (12). The use of TBDMS, albeit on a larger scale, lead to a reduced yield (26%) (11) due to the formation of hydrolysis and cleavage products.

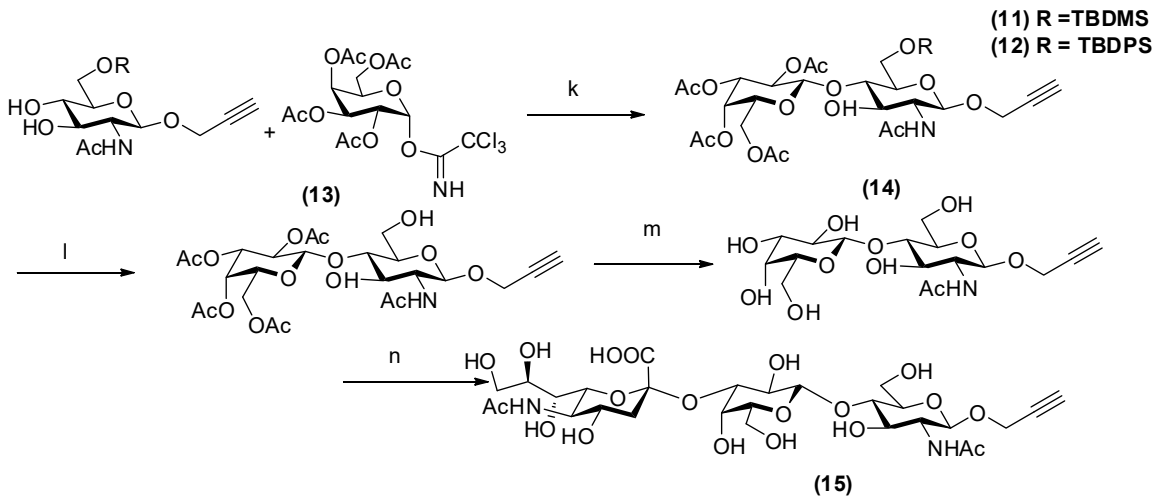


Figure 6. k)$BF_3.OEt_2$ DCM -42°C R=TBDMS 26% R=TBDPS 55% l) R=TBDPS, BF_3/MeCN 89% R=TBDMS, 3HF.Et_3N 42%m) NaOMe/MeOH 92% n)NMCSS, PMST, Sialic acid, CTP, pH 8.5 69%

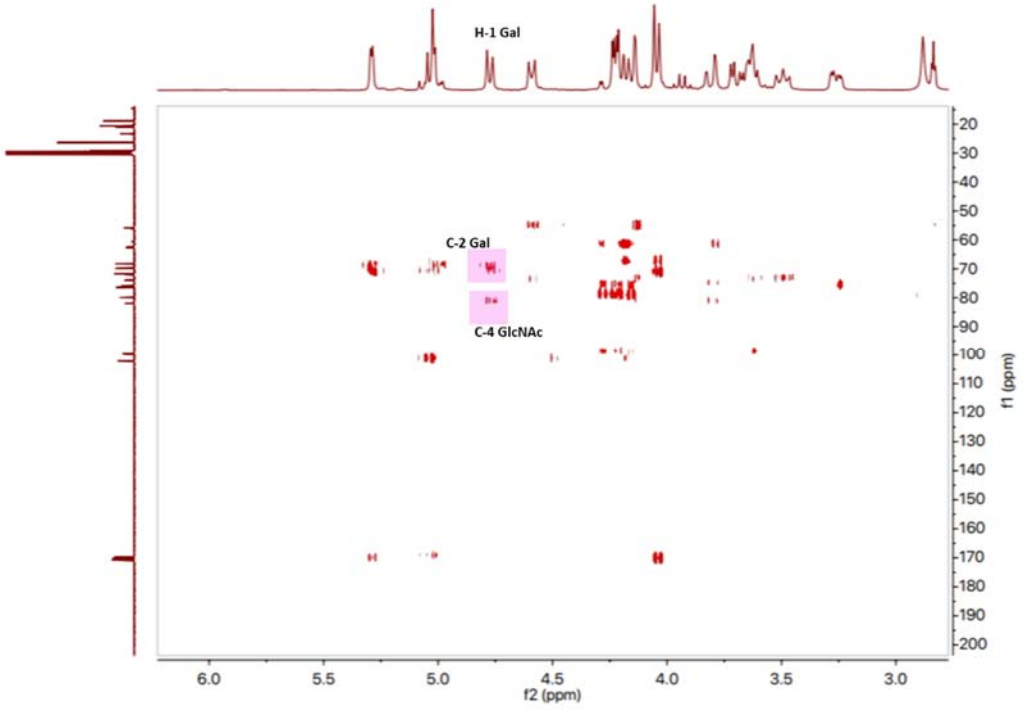


Figure 7. The HMBC cross peak between H-1 Gal and C-4 GlcNAc confirmed that the disaccharide had formed a 1->4 linkage (shown for TBDMS disaccharide). This HMBC coupling corresponds to what has previously demonstrated in the literature for this disaccharide (conducted in d6-acetone).[17]

HMBC was used to determine that the correct linkage was formed (1->4). The coupling constant on H1-Gal was found to be 8.2 Hz, also confirming that the linkage was beta in stereochemistry (Figure 7). Deprotection of the TBDPS group could be conducted with boron

triflouride in a few hours, however significant degradation was observed. When HF.Et_3N was used long reaction times (1 week) were needed for full conversion (13). On the other hand deprotection of the TBDMS protected sugar could be accomplished overnight with 3HF.Et_3N. After Zemplen deacetylation with NaOMe, the deprotected O-propargyl disaccharide was sialyated.[29]

3.2.3. Sialylation

The enzymes, NMCSS (*Neisseria Meningitides* CMP-sialic acid synthetase) and PMST (*Pasteurella multocida* alpha-2-3-sialyltransferase) were expressed by transformation into Bl21(DE3), before inducing expression with IPTG. After purification by nickel affinity chromaography the purified enzymes were obtained in quantities sufficient for the chemoenzymatic synthesis of the trisachcaride (Figure 8).

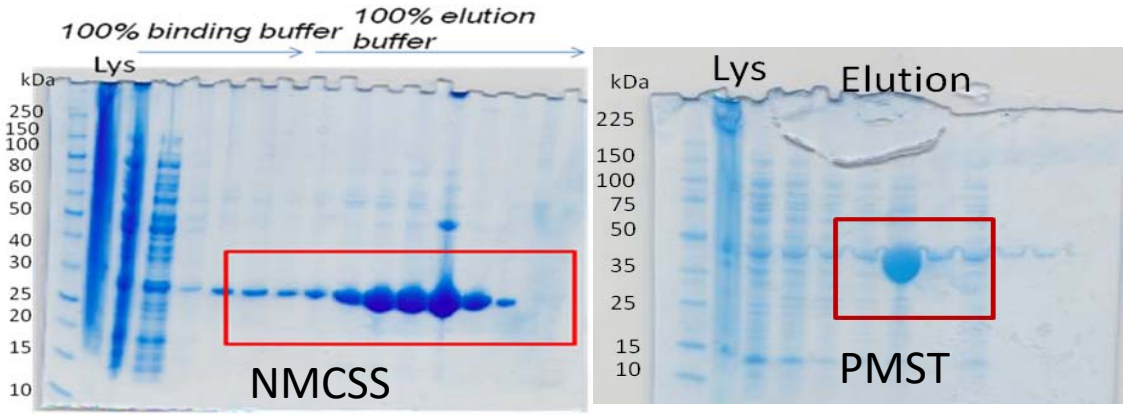


Figure 8. SDS-PAGE (4-12% Bis Tris, MES) analysis of purified NMCSS and purified PMST. NMCSS was expressed in a yield of 66mgL^{-1} whilst the PMST was expressed with a yield of 52mgL^{-1}.

In this enzymatic step, NMCSS generates the substrate CMP sialic acid for the PMST enzyme *in situ* (Figure 9).

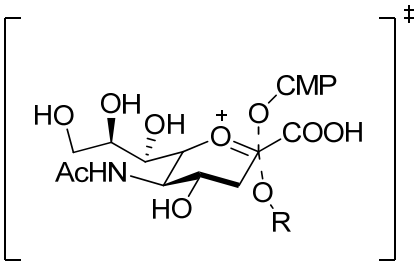


Figure 9. Transition state for the PMST catalysed sialylation reaction. This enzyme has been used by Chen *et al* for the efficient and selective one pot multienzyme preparation of complex mammalian glycans.[30]

The reaction proceeds efficiently in pH 8.6 buffer (with Mg^{2+}) and yielded after column chromatography the trisaccharide in high yields. Several parameters can be used for determining the stereochemistry of the glycosidic linkage formed, one of the most common approaches is to look at H-3eq. H-3eq protons are downfield shifted to approximately 2.67-2.72ppm for the α anomer, compared to the β anomer which (for unprotected sialosides) is observed between 2.25ppm-2.40ppm. In this case, the H-3eq was observed at 2.69ppm (Figure 10) confirming the alpha stereochemistry of the linkage. There also exist several other empirical rules for sialoside determination, but the H-3eq is the most general for unprotected sialosides.[2]

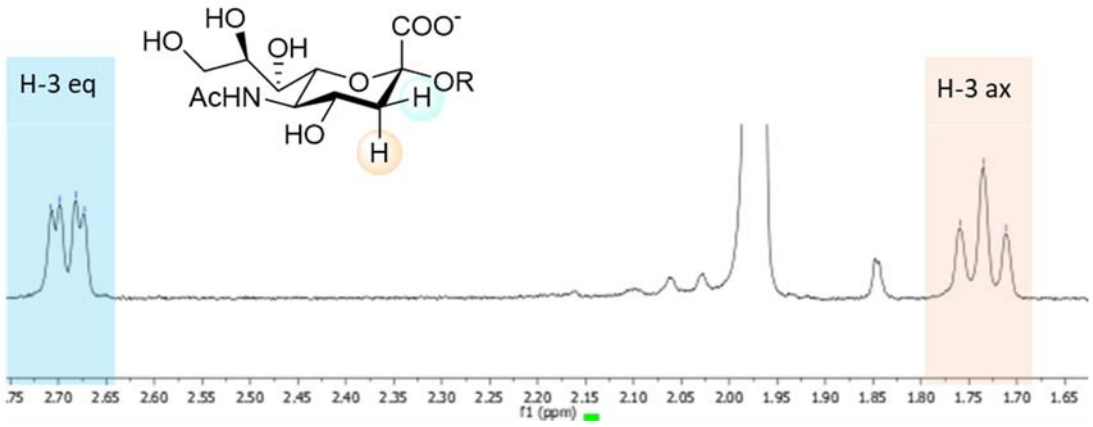


Figure 10. Based upon standard literature precedent for assigning the stereochemistry of sialic acid, the presence of a H-3eq at ~2.69ppm indicated the presence of the alpha anomer. Furthermore, no peaks are seen between 2.25-2.40pm indicating that only the alpha anomer was formed.

H-3eq and H-3ax were further confirmed to be the H-3 protons through strong cross peaks on the COSY spectrum (Figure 11). In addition to the use of H-3eq and H-3ax for stereochemistry determination, the H-4 peak can also be used to determine the stereochemistry of the anomeric centre, again based upon well documented empirical rules. In this case, the COSY was used to identify the H-4 (Figure 11), the presence of the H-4 at ~3.6 ppm again indicated that the linkage was α in configuration. For unprotected sialosides, 3.6-3.8 ppm is seen for α anomers and 3.9-4.2 ppm for β. [31]

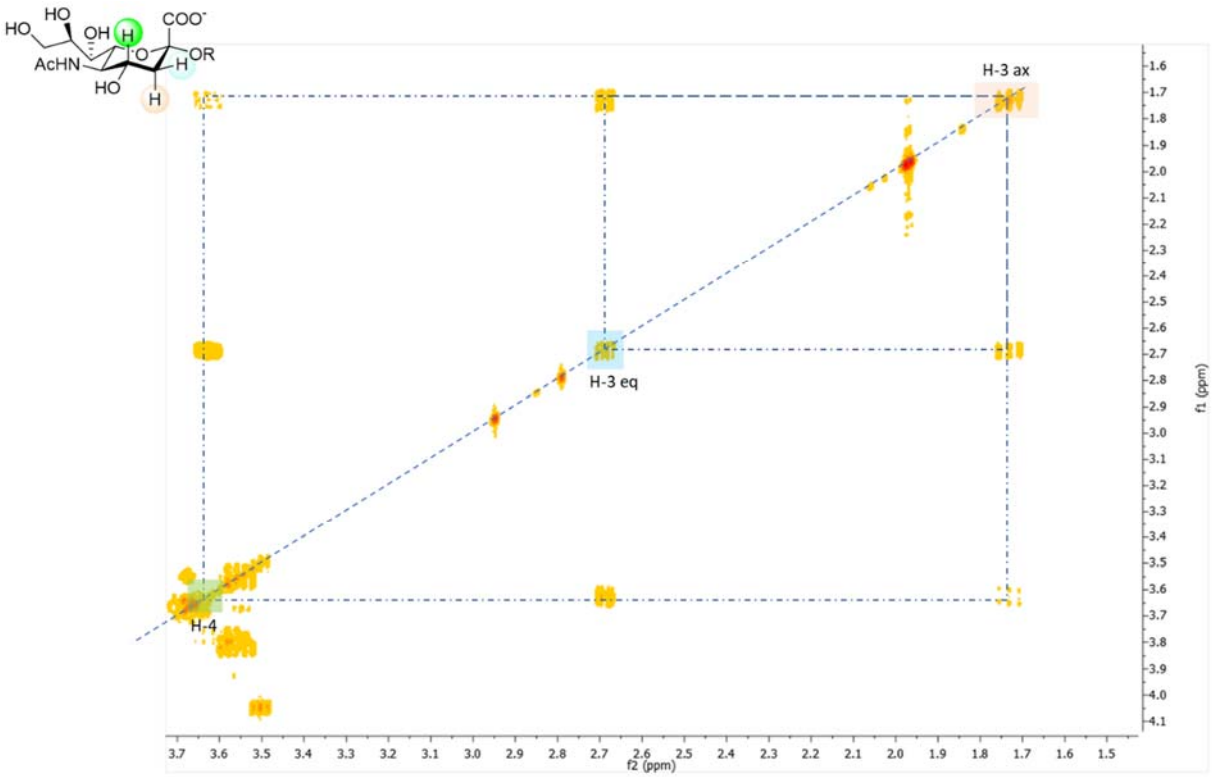


Figure 11. Based upon standard literature precedent for assigning the stereochemistry of sialic acid, the presence of a H-3eq at ~2.69ppm indicated the presence of the alpha anomer, whilst the presence of H-4 at 3.6 ppm also indicated that the alpha anomer was formed.

3.3 Conclusions

In conclusion, the synthesis of the trisaccharide, propargyl-(5-acetimido-3,5-dideoxy-d-glycero-D-galacto-2-nonulopyranosylonic acid-(α2-3)- -D-galactopyranosyl-(1-4)-2-acetimido-2-deoxy-D-glucopyranoside, was completed via a chemoenzymatic synthetic route in an overall yield of 16.7% (over 7 linear steps from 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-β-D-glucopyranose). This trisaccharide was consistent with data previously reported in the literature and subsequently taken forward into the bio-conjugation reactions in Chapter 4.

As a large quantity of the trisaccharide would be required for subsequent conjugation reactions, methods to rapidly synthesise the intermediate lactosamine on a larger scale were trialled. Despite literature promise, these methods (Heyns rearrangement or microwave assisted synthesis) didn't yield product in high enough yields or sufficient purity to outweigh the current linear synthetic route and were thus not investigated further.

Looking forward, within the Siglec family of receptors a large number of different sugars are bound selectivity by the different receptors.[32] Thus this route may not be general if the project scope is expanded and other synthesis methods may have to be adopted to enable access to other sugars e.g. sialyl Lewis-X. Furthermore, as outlined in Chapter 1 significant progress has been made in the design and synthesis of high affinity unnatural ligands for Siglec receptors. If these ligands were to be taken up in this project, the substrate specificity of the enzymes would need to be investigated as well as additional synthetic steps introduced to synthesise the unnatural sialic acids.[33]

3.4 Biological experimental

Protein expression

To 5ml of LB solution containing 5uL of 10 $mgml^{-1}$ syringe filtered ampicillin was added 5uL of the NMCSS or PMST glycerol stock (BL21(DE3)). After shaking at 220rpm/37°C for 16 hours, the cells were spun down at 3750 rpm for 10 minutes. QIAGEN Miniprep was run and the resulting plasmid DNA sequenced by the gene sequence service. The reagents used were the same as those used in Chapter 2 (See general biological experimental for details).

NMCSS

M M E K Q N I A V I L A R Q N S K G L P L K N L R K M N G I S L L G H T I N A A I S
S K C F D R I I V S T D G G L I A E E A K N F G V E V V L R P A E L A S D T A S S I S
G V I H A L E T I G S N S G T V T L L Q P T S P L R T G A H I R E A F S L F D E K I K
G S V V S A C P M E H H P L K T L L Q I N N G E Y A P M R H L S D L E Q P R Q Q L
P Q A F R P N G A I Y I N D T A S L I A N N C F F I A P T K L Y I M S H Q D S I D I D T
E L D L Q Q A E N I L N H K E S L E H H H H H H

PMST

M H H H H H H S S G V D L G T E N L Y F Q S M K T I T L Y L D P A S L P A L N Q
L M D F T Q N N E D K T H P R I F G L S R F K I P D N I I T Q Y Q N I H F V E L K D N
R P T E A L F T I L D Q Y P G N I E L N I H L N I A H S V Q L I R P I L A Y R F K H L D
R V S I Q Q L N L Y D D G S M E Y V D L E K E E N K D I S A E I K Q A E K Q L S H Y
L L T G K I K F D N P T I A R Y V W Q S A F P V K Y H F L S T D Y F E K A E F L Q P
L K E Y L A E N Y Q K M D W T A Y Q Q L T P E Q Q A F Y L T L V G F N D E V K Q
S L E V Q Q A K F I F T G T T T W E G N T D V R E Y Y A Q Q Q L N L L N H F T Q A
E G D L F I G D H Y K I Y F K G H P R G G E I N D Y I L N N A K N I T N I P A N I S F
E V L M M T G L L P D K V G G V A S S L Y F S L P K E K I S H I I F T S N K Q V K S
K E D A L N N P Y V K V M R R L G I I D E S Q V I F W D S L K Q L

To 20ml of LB solution containing 20uL of 10 $mgml^{-1}$ syringe filtered ampicillin was added 20uL of the NMCSS or PMST glycerol stock (BL21(DE3)). After shaking at 220rpm and 37°C

for 16 hours, the overnight culture was added to 1L of room temperature autoclaved LB with 1ml of 10 $mgml^{-1}$ ampicillin solution. The culture was shaken at 37°C, 220rpm until an OD of 0.69 was reached. 1ml of 1mM sterile IPTG solution was then added in one portion and the culture shaken at 220rpm, 30°C for 18 hours. The cells were then pelleted, 8000rpm, JLA9.1, 4°C and flash frozen in liquid nitrogen. This cell pellet was then added to approximately 20ml of Ni affinity chromatography loading buffer (pH7.5, 500mM NaCl, 20mM tris, 40mM imidazole), and 10mg of lysozyme, DNAse and EDTA free protease inhibitor were added. After stirring for 1 hour at 4°C, the cells were sonicated on/off for 30 seconds each, a total of 5 times. The cells were then pelleted, 20000rpm, JA25.5, 4°C, for 40min and the brown cell lysate loaded onto a prepared manual nickel affinity column. Following elution with high imidazole buffer (pH7.5, 500mM NaCl, 20mM tris, 500mM imidazole), fractions containing protein as determined by SDS-PAGE analysis were combined, dialysed against pH8, 50mM NaPi buffer and concentrated with vivaspin (5K MWCO). The resultant protein fractions for NMCSS were analysed by BCA assay ($66mgL^{-1}$), aliquoted into 1ml portions and snap frozen. The resultant protein fractions for PMST were analysed by BCA assay ($52mgL^{-1}$), aliquoted into 1ml portions and snap frozen.

3.5 Organic chemistry experimental

Chemicals were purchased from Sigma Aldrich and unless stated used as received. Solvents were purchased from Sigma Aldrich, Acros and/or Rathburn Chemicals. Dry solvents were purchased as anhydrous but dry alcohol-free DCM was obtained from an activated aluminium Grubbs still and stored under argon over 4Å molecular sieves. Air/moisture sensitive reactions were carried out, in furnace (450°C) dried glassware, under N_2, using dry disposable syringes and a manifold setup. Analytical TLC was run with silica gel, 60 F254, 0.25 mm pre-coated TLC plates (Merck) and visualised with UV_{254}, moly-dip (2.5g ammonium molybdate, 100 cm^3 water, 5.3cm^3 conc. H_2SO_4) or ninhydrin as stated. Flash column chromatography was carried out using silica gel (Merck, particle size 0.040–0.063 mm).

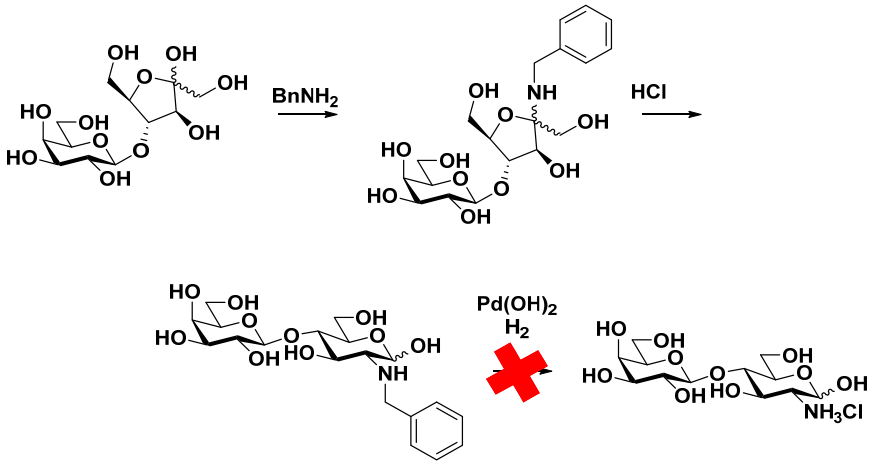
NMR spectra were recorded at 303K using a 400MHz (^{1}H), 300MHz (^{1}H) and 250 MHZ (^{1}H) and ^{13}C with 100MHz or 75MHz. Structural assignments are based upon a combination of DEPT, COSY, HMBC and HSQC. Signals are reported as singlet (s), broad singlet (br, s) doublet (d), doublet of doublets (dd), triplet (t), quartet (q) and multiplet (m). Coupling constants are all in Hz and peaks reported in ppm (δ) relative to residual solvent peaks.

LRMS: ESI-MS (positive and negative ion mode) were recorded on a Micromass platform spectrometer using MeOH as a solvent. IR was recorded on a Bio-Rad FTS-6000 FT spectrometer equipped with a DuraSamplIR II diamond ATR head (oils and solids were examined neat). Absorption maxima (ν_{max}) are recorded in wavenumbers (cm^{-1}).

Melting points were recorded using a Lecia Galen III hot stage microscope fitted with a Testo 720 thermocouple probe and are uncorrected. Optical rotations were measured on a Perkin Elmer 241 polarimeter at 589 nm (Na D line) with a path length of 1 dm and are reported with

implied units of 10^{-1} deg cm^2 g^{-1}. Concentrations (c) are given in g100 cm^{-3}. In this experimental, cm^3 is used as the standard unit for volume throughout the organic experimental, this is in contrast to the biological experimental wherein ml is used.

Heyns rearranagement synthesis of lactosamine

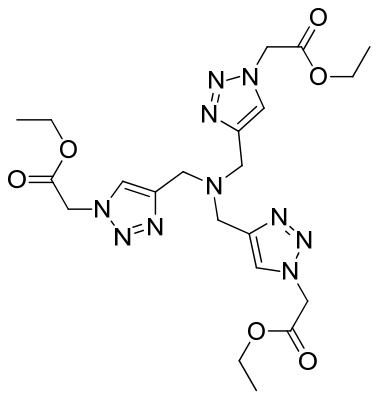


To a flask was charged lactulose (2.54g, 7.42mmol) and to this was added benzylamine (5.67cm^3, 57.88mmol) in one portion. The mixture was stirred under nitrogen at 40°C for 3 days before methanol (6.00cm^3) was added and the solution added to rapidly stirring ether (250cm^3). A yellow solid precipitated which was washed with PET-ether and concentrated i*n vacuo*, to give yellowish oil as a mixture of products with residual solvent (6.37g, 200%). This crude oil (5.23g) was taken up in glacial acetic acid (2.20 cm^3)/methanol (18.00cm^3) and stirred at ambient temperature for 2 hours (a red colouration was observed). This was subsequently poured into ether to give the product as a yellow solid after filtration, which would convert back to brown oil (4.94g, 94%) on standing. LRMS (ESI-) calculated for: C19H29NO10 (M - H) 430.18; found: 430.1 (M -H)

To the intermediate (4.07g) oil was added H_2O (10cm^3) followed by HCl until a pH of 1 was reached. $Pd(OH)_2$ (0.205g) was added and the flask backfilled with hydrogen. After stirring for

over 1 week with the continuous filling of the balloon, limited conversion was seen by TLC/MS. It was concluded that the catalyst was completely poisoned: reaction failed.

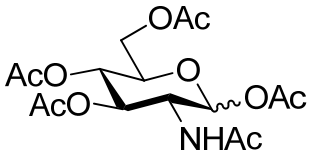
Ethyl-2,2',2''-(4,4',4''-(nitrilotris(methylene))tris(1H-1,2,3-triazole-4,1-diyl))triacetate



To a vessel was charged NaN_3 (2.31g, 35.5mmol), DMF (5cm^3) and then at ambient temperature with rapid stirring was added dropwise, ethylchloroacetate (3 cm^3, 28mmol).This was stirred for 24 hours. After which 150 cm^3 of ether was added and the solution washed with water (50 cm^3). The organic layer was separated, dried (sodium sulfate), filtered and concentrated to around 10 cm^3. This solution was added to another flask containing propargyl amine (0.7 cm^3, 4.6mmol) in MeCN (5 cm^3) under argon. The solution was cooled to 0°C and 2,6-lutidine (0.6 cm^3, 5.2mmol) and CuBr (20mg, 0.14mmol) were added in one portion. The reaction was stirred for 3 days and then concentrated *in vacuo*. The residue was then columned in 10% MeOH/EtOAc to give the product as a white solid (1.90g, 74%).

m.p =120-121°C. IR (neat) ν_{max} (cm^{-1}): 1743(C=O). LRMS (ESI-) calculated for: $C_{21}H_{30}N_{10}O_6$ (M -H) 517.23; found: 517.24 (M -H).^{1}H NMR (300 MHz, $CDCl_3$, δ) 7.85 (s, 3H, CH_{Ar}), 5.15 (s, 6H, CH_2COOEt), 4.24 (q, *J* = 7.1 Hz, 6H, $CO\underline{CH_2}CH_3$), 3.80 (s 6H, NCH_2Ar), 1.28 (t, *J* = 7.1 Hz, 9H, CH_3). ^{13}C NMR (75 MHz, $CDCl_3$) δ 166.43 (CO), 144.38(C, C_{Ar}), 125.45(CH, CH_{Ar}), 62.47(CH_2, $CO\underline{CH_2}CH_3$), 50.99(CH_2, $\underline{C}H_2COOEt$), 47.33(CH_2, NCH_2), 14.17(CH_3, CH_3).

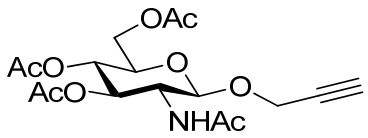
2-Acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-α/β-D-glucopyranose



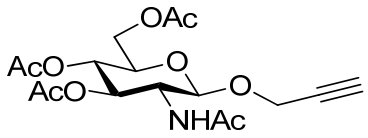
To a solution of N-acetyl-D-glucosamine (10.00g, 45.2mmol) under Ar in dry pyridine (100 cm^3) was added dropwise at 0°C, Ac_2O (100cm^3) and the reaction stirred at room temperature for 24 hours. TLC analysis indicated that the reaction had gone to completion and the reaction was concentrated *in vacuo*. The residue was redissolved in EtOAc (50 cm^3) and washed consecutively with 1M $HCL_{(aq)}$ (50 cm^3), sat. $NaHCO_3$ (50 cm^3) and sat. $brine_{(aq)}$(50 cm^3). The solution was then dried ($MgSO_4$), filtered (gravity) and concentrated in *vacuo* to yield a colourless oil, which upon drying gave a white solid (13.25, 75%).

IR (neat) ν_{max} (cm^{-1}): 1741.54 (C=O) 1670.47 (C=O, NHAc) . LR-MS (ESI+) 407.16 $[M+K]^+$. ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.13 (d, *J* = 3.6 Hz, 1H,H-1), 5.80 (d, *J*= 9.0 Hz, 1H, NH,), 5.24 – 5.15 (m, 2H, H-3,H-4), 4.51 – 4.40 (m, 1H,H-2), 4.21 (dd, *J*= 12.4, 4.0 Hz, 1H, H-6), 4.03 (dd, *J*= 12.6, 2.3 Hz, 1H, H-6), 4.00 – 3.93 (m, 1H, H-5), 2.16 (s, 3H, Me), 2.05 (s, 3H, Me), 2.01 (s, 3H, Me), 2.01 (s, 3H, Me), 1.90 (s, 3H, Me). ^{13}C NMR (75 MHz, $CDCl_3$) δ 171.52 (C=O), 170.63 (C=O), 170.03 (C=O), 169.08 (C=O), 168.65 (C=O), 90.62(CH, C1), 70.58(CH, C5), 69.66(CH, C3), 67.55(CH, C4), 61.54(CH_2, C6), 50.95(CH, C2), 22.92(Me), 20.85(Me), 20.63(Me), 20.50(Mex2):

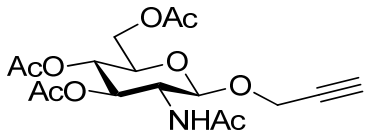
Propargyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranoside trials



To a vessel was charged 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-α,β-D-glucopyranose (0.60g, 1.54mmol), dry DCM (50cm^3), propargyl alcohol (0.3cm^3, 5.2 mmol), $Yb(OTf)_3$ (0.28g, 0.46mmol) and the reaction refluxed overnight. After 48 hours, the reaction was washed with $NaHCO_3$, H_2O and saturated brine before being evaporated and concentrated *in vacuo* yielding start material (0.53g).



To a vessel was charged dry silica (9.60g) and dry ether (48 cm^3) under argon. To this was added conc. H_2SO_4 (0.96 cm^3) and the slurry stirred for 5 minutes. The slurry was then concentrated and dried *in vacuo* for 3 hours. To a microwave vessel was charged the 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-α,β-D-glucopyranose (100mg, 0.25mmol, 1.5eq), anhydrous DCE (2cm^3) and the H_2SO_4-sillica (2mg), it was kept at ambient temperature for 1.5 hours. Propargyl alcohol (0.01cm^3, 0.17mmol, 1eq) was added and the reaction microwaved at 110°C, high stirring, 300W under N_2. TLC and ^{1}H NMR indicated that the start material was the sole product.

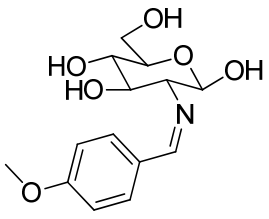


To a vessel was charged 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy- α,β-D-glucopyranose (1.60g, 4.1mmol), $Yb(OTf)_3$ (0.76g, 1.23mmol), propargylic alcohol (0.48cm^3, 8.2mmol) and

dry DCM (4.00 cm^3). The reaction was degassed with Ar and under nitrogen run on a MW at 100°C, 300W for 1 hour; washed with $NaHCO_3$, water, brine and then dried over Na_2SO_4 and filtered. After titrating with ice cold petroleum-ether (40-60) an off white solid was isolated (0.43g, 27%), which was found to be pure product.

m.p=194-196 °C. $[\alpha]^D$=+88.8 (c=0.058, MeOH) lit. $[\alpha]^D$=-39 (c=0.1, $CHCl_3$)[34] Rf =0.45 (EtOAc) LRMS (ESI+) calculated for: $C_{17}H_{23}NNaO_9$ (M + Na) 408.36; found: 408.11 (M + Na). IR (neat) ν_{max} (cm^{-1}): 3245.65 (NH), 1743.70 (CO), 1657.94 (CO, NHAc), 1555, 1375 ,^{1}H NMR (300 MHz, $CDCl_3$, δ) 5.46 (d, *J* = 8.9 Hz, 1H, NH), 5.28 (dd, *J* = 12.9, 7.0 Hz, 1H, H-3), 5.09 (t, *J* = 9.6 Hz, 1H, H-4), 4.85 (d, *J* = 8.4 Hz, 1H, H-1), 4.38 (d, *J* = 2.4 Hz, 2H, CH_2), 4.27 (dd, *J* = 12.3, 4.6 Hz, 1H, H-6), 4.14 (dd, *J* = 12.3, 2.4 Hz, 1H, H-6), 3.95 (dt, *J* = 10.4, 8.7 Hz, 1H, H-2), 3.72 (ddd, *J* = 9.9, 4.6, 2.5 Hz, 1H, H-5), 2.46 (t, *J* = 2.4 Hz, 1H, CCH), 2.09 (s, 3H, Me), 2.03 (s, 3H, Me), 2.03 (s, 3H, Me), 1.96 (s, 3H, Me). ^{13}C NMR (75 MHz, $CDCl_3$, δ) 171.04 (CO), 170.81 (CO), 170.46(CO), 169.48(CO), 98.47(C-1), 78.63, 75.49, 72.55(CH, C-5), 72.11(CH, C-3), 68.65(CH, C-4), 62.09(CH_2, C-6), 56.01(CH_2, *C*H_2C), 54.41(CH, C-2), 23.44(Me), 20.84(Me), 20.78(Me), 20.73(Me). This data was consistent with literature data for this compound and subsequently used as a benchmark for future experiments.

2-Deoxy-2-[*p*-methoxybenzylidene(amino)]-β-D-glucopyranose

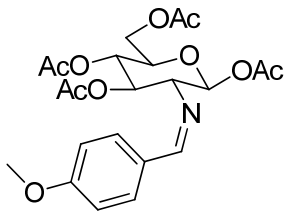


To a vessel was charged $NaOH_{(aq)}$ (1M, 120cm^3) and with vigorous stirring D-(+)-glucosamine hydrochloride (25.0 g, 0.116mol) was added carefully. The mixture was stirred for 10 minutes until a colourless solution formed. *p*-anisaldehyde (17cm^3, 0.140mol) was then added slowly

over 10 minutes with vigorous stirring. The reaction was cooled to 4°C and allowed to stand at this temperature for 2 hours. The white precipitate was filtered (suction), washed (200cm^3 ice cold H_2O), and then dried (suction and then high vacuum). This yielded the product as a white solid, 2-deoxy-2-[*p*-methoxybenzylidene(amino)]-β-D-glucopyranose (30.3g 87%).

m.p =188-190°C. $[\alpha]^D$=-21.2 (c=0.541, 5:3 MeOH:H_2O), Lit. $[\alpha]^D$ = - 65.2 (c = 1, DMSO)[26] IR (neat) ν_{max} (cm^{-1}) 3487.3 (NH) 3332 (br, OH) 3222.5 (br, OH). LRMS (ESI+) calculated for: $C_{14}H_{20}NO_6$ (M + H) 298.13; found: 298.11 (M + H) ^{1}H NMR (300 MHz, DMSO) δ 8.06 (s, 1H, HC=N), 7.63 (d, *J* = 8.8 Hz, 2H, ArH), 6.93 (d, *J* = 8.7 Hz, 2H, ArH), 6.48 (d, *J* = 6.7 Hz, 1H,-OH H-1), 4.87 (d, *J* = 5.2 Hz, 1H, -OH, H-4), 4.76 (d, *J* = 5.6 Hz, 1H, -OH, H-3), 4.64 (t, *J* = 7.2 Hz, 1H, H-1), 4.50 (t, *J* = 5.8 Hz, 1H, -OH, H-6)), 3.74 (s, 3H, OMe), 3.72 – 3.62 (dd, *J* = 10.8, 6.4 Hz, 1H, H-6), 3.48 – 3.37 (m, 2H, H-6, H-3),3.24 (dd, *J* = 9.6, 4.0 Hz, 1H, H-5), 3.16 (dd, *J* = 8.5, 5.3 Hz, 1H, H-4), 2.79 – 2.68 (m, 1H, H-2).^{13}C NMR (75 MHz, DMSO) δ 161.29 (CHN), (C-Ar), 131.83 (C-Ar), 129.66(C-Ar), 113.93(C-Ar),95.64(C-1), 78.21 (C-2), 76.87(C-5), 74.61 (C-3), 70.40 (C-4), 61.30 (C-6), 55.32(OCH_3).

1,3,4,6-Tetra-O-acetyl-2-deoxy-2-[*p*-methoxybenzylidene(amino)]-β-D-glucopyranose

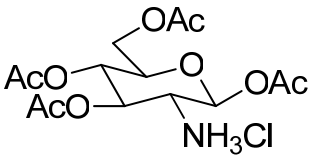


To a vessel was charged 2-deoxy-2-[*p*-methoxybenzylidene(amino)]-β-D-glucopyranose (15.00g, 50.45mmol) and pyridine (81cm^3). To this was added at 4°C with vigorous stirring acetic anhydride (45cm^3). The reaction was stirred at 4°C for 1 hour, the start material slowly goes into solution, so DMAP (cat.) was added at 0°C and the reaction stirred for a further 2

hours. The white solid dissolves very slowly into the solution, so extra pyridine 50cm^3 and Ac_2O 20cm^3 was added before the reaction was stirred for a further 5 hours. This yielded a colourless solution, which was poured into iced water (600cm^3), filtered (suction) and washed with cold water (800cm^3) to give a white solid. The product was still very wet, so was dissolved in DCM (200cm^3), washed with brine, dried with Na_2SO_4, filtered (gravity), concentrated *in vacuo* and finally dried under vacuum. This yielded a white solid, 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-[*p*-methoxybenzylidene(amino)]-β -D-glucopyranose (15.47g, 66%).

m.p=187-189°C. IR (neat) ν_{max} (cm^{-1}) 1748 (CO) 1736 (CO) 1646 (CO) 1605 (CO). $[\alpha]^D$=+100 (c=0.338, $CHCl_3$), lit. $[\alpha]^{23\ D}$ + 88.1 (c = 1, CHCl3)[26], LR-MS (ESI+) calculated for: $C_{22}H_{28}NO_{10}$ (M + H) 466.17; found: 466.16 (M + H). ^{1}H NMR (400 MHz, $CDCl_3$ δ) 8.15 (s, 1H, CH), 7.64 (t, *J* = 8.0 Hz, 2H, ArH), 6.91 (d, *J* = 8.7 Hz, 2H), ArH, 5.94 (d, *J* = 8.3 Hz, 1H, H-1), 5.42 (t, *J* = 9.6 Hz, 1H, H-3), 5.14 (t, *J* = 9.8 Hz, 1H, H-4), 4.38 (dd, *J* = 12.4, 4.5 Hz, 1H, H-6), 4.12 (dd, *J* = 12.4, 1.9 Hz, 1H, H-6), 3.97 (ddd, *J* = 10.1, 4.4, 2.0 Hz, 1H, H-5), 3.84 (s, 3H, OMe), 3.48 – 3.39 (m, 1H, H-2), 2.09 (s, 3H, Me), 2.03 (s, 3H, Me), 2.01 (s, 3H, Me), 1.87 (s, 3H, Me). ^{13}C NMR (75 MHz, $CDCl_3$, δ) 170.81 (CO), 170.02(CO), 169.67(CO), 168.89(CO), 164.40 (CH, CH), 162.47 (C-Ar), 130.41 (2xCH, Ar), 128.44 (C Ar), 114.21 (CxCH, Ar), 93.31 (CH, C-1), 73.41 (CH, C-3), 73.08 (CH, C-5), 72.92 (CH, C-2), 68.21 (CH, C-4), 61.98 (CH_2 C-6), 55.55 (CH_3, OMe), 20.94 (Me), 20.89(Me), 20.82 (Me), 20.64 (Me).

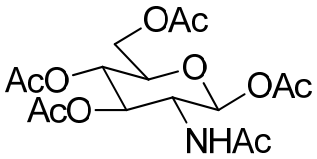
1,3,4,6-Tetra-O-acetyl-2-deoxy-β-D-glucosamine hydrochloride



To 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-[*p*-methoxybenzylidene(amino)]-β-D-glucopyranose (6.00g, 0.13mmol) was added analytical acetone (66cm^3) at r.t.p. To this with vigorous stirring was added at ambient temperature HCl (5N, 3.3cm^3) slowly and the reaction cooled to 4°C and fridged for 2 hours. A blue colour can be seen in the reaction mixture corresponding to the anisaldehyde. The white solid was filtered under vacuum and washed with 3x50cm^3 of ice cold ether, then put on high vac. This yielded a white solid 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-β-D-glucosamine hydrochloride (4.78g, 97%).

m.p=220-222°C. IR (neat) ν_{max} (cm^{-1}) 1767 (CO) 1744 (CO) $[\alpha]^D$=+24.1 (c=0.12, 1:1 MeOH:H_2O), lit. $[\alpha]^D$=+20 (c=0.5, $(CH_3)_2CO$)[35] LRMS (ESI+) calculated for: $C_{14}H_2NO_9$ (M+H) 348.33 Found: 348.09 (M+H). ^{1}H NMR (300 MHz, DMSO, δ) 8.64 (br s, 3H, NH_3^+), 5.84 (d, *J* = 8.7 Hz, 1H, H-1), 5.35 – 5.21 (m, 1H, H-4), 4.88 (t, *J* = 9.6 Hz, 1H, H-3), 4.15 (dd, *J* = 12.4, 4.3 Hz, 1H, H-6), 4.06 – 3.91 (m, 2H, H-6, H-5), 3.54 (t, *J* = 9.5 Hz, 1H, H-2), 2.12 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.95 (s, 3H, OAc), 1.93 (s, 3H, OAc). ^{13}C NMR (75 MHz, DMSO, δ) δ 169.95 (CO), 169.75 (CO), 169.29(CO), 168.62 (CO), 90.09 (CH, C-1), 71.59(CH, C-3), 70.33(CH, C-4), 67.80(CH, C-5), 61.25 (CH_2 C-6), 52.13(CH, C-2), 20.94(Me), 20.85(Me), 20.47(Me), 20.33(Me).

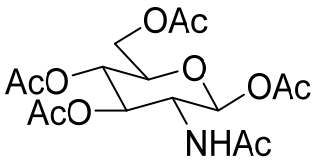
2-Acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-β-D-glucopyranose



To a vessel was charged the 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-β-D-glucosamine hydrochloride (4.03g, 10.42mmol) and DCM (70cm^3) was added at ambient temperature. To this was added Et_3N (6cm^3) and then Ac_2O (11.6cm^3) dropwise. The reaction was stirred at ambient temperature for 5 hours. MeOH (2cm^3) was added, and the organics washed with HCl (1M, 2x40cm^3), $NaHCO_3$ (1x60cm^3), Brine (1x40cm^3), dried with Na_2SO_4, filtered (gravity) and concentrated *in vacuo*. This yielded a white solid, 2-Acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-β-D-glucopyranose (3.48g, 85.7%).

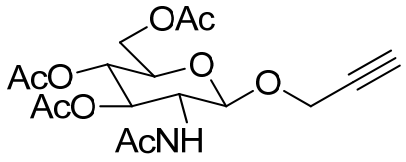
m.p=191-192°C. IR (neat) ν_{max} (cm^{-1}) 3251 (br NH) 1744 (CO) 1655 (CO, NHAc) $[\alpha]^D$=+23.4 (c=0.46, MeOH), lit. $[\alpha]^{D\ 23}$ + 2.9 (CHCl3, c = 1)[26], LR-MS (ESI+) calculated for: $C_{16}H_{23}NNaO_{10}$ (M + Na) 412.34; found: 412.11 (M + Na). Rf=0.3(EtOAc). ^{1}H NMR (300 MHz, $CDCl_3$) δ 5.69 (d, *J* = 8.8 Hz, 1H, H-1), 5.60 (d, *J* = 9.5 Hz, 1H, NHAc), 5.18 – 5.08 (m, 2H, H-4, H-3), 4.35 – 4.22 (m, 2H, H-2, H-6), 4.12 (dd, *J* = 12.5, 2.2 Hz, 1H, H-6), 3.79 (tdd, *J* = 20.2, 11.2, 9.0 Hz, 1H, H-5), 2.11 (s, 3H, OAc), 2.09 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc), 1.93 (s, 3H, NHAc). ^{13}C NMR (75 MHz, $CDCl_3$) δ 171.30 (CO), 170.77 (CO), 170.25(CO), 169.62 (CO), 169.41 (CO), 92.62 (C-1), 72.88(C-5), 72.74 (C-3), 68.06 (C-4), 61.83(C-6), 52.98(C-2), 23.24(Me), 20.99(Me), 20.81(Me), 20.76(Me), 20.67(Me).

2-Acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-β-D-glucopyranose (K-10 Clay Method)



To a large flask was added at 0°C, Ac_2O (500cm^3) and slowly over 20 minutes a mixture of the N-acetyl-D-(+)-glucosamine (40g, 0.10 mol) and montmorillonite K-10 (125g). After stirring at ambient temperature for 24 hours, the reaction was filtered through 20g of cellite and washed with methyl acetate (500cm^3). The orange solution was then concentrated *in vacuo* before being recrystallised from hot methanol, to yield after filtration the desired compound (19.2g, 27%). Data consistent with previously isolated compound, 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-β-D-glucopyranose, prepared via the anisaldehyde protections sequence and characterised on the previous page of this experimental. The data was also consistent with literature data for this compound.[36]

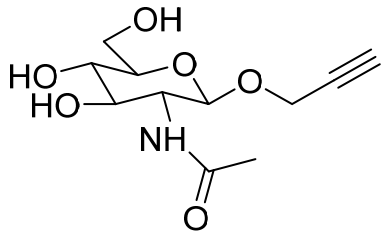
Propargyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranoside



To a dry vessel was charged the 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-β-D-glucopyranose (1.5g, 3.85mmol) with a small amount of crushed 3Å molecular sieves. To this was added DCM (120cm^3), and then propargyl alcohol (0.5cm^3, 8.6mmol) under nitrogen. To this was finally added $Yb(OTf)_3$ (0.726g, 1.17mmol) and the reaction refluxed for 24hours.

After washing with $NaHCO_3$ (50cm^3), brine (50cm^3) and drying (Na_2SO_4); the reaction mixture was columned with 100% ETOAc to give a white solid as the product, propargyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranoside (1.02g, 69%). Data consistent with previously isolated compound, propargyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranoside, prepared via the microwave reactions and characterised on page 170 of this experimental. The data was also consistent with literature data for this compound.[37]

Propargyl 2-acetamido-2-deoxy-β-D-glucopyranoside

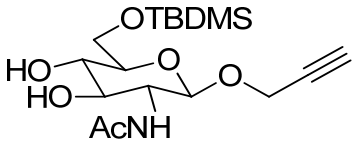


To the propargyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranoside (4.00g) was added dry methanol (150cm^3) and NaOMe solution (0.5M, 2cm^3). After stirring for 1 hour, DOWEX-H+ was added until the pH was approximately 7. The solution was filtered through a frit and concentrated *in vacuo* to yield a white solid propargyl 2-acetamido-2-deoxy-β-D-glucopyranoside (2.69g, 100%)

LRMS (ESI-) calculated for: $C_{11}H_{16}NO_6$ (M -H) 258.1; found: 258.0 (M - H). m.p 154.5-155.8 $[\alpha]^D$= -54.3 (c=1.1, MeOH). Lit. $[\alpha]^{D20}$-54° (c = 0.1 g cm-1 in $CHCl_3$) [34]; IR (neat) ν_{max} (cm^{-1}) 3291 (br OH) 3243 (br OH) 1657 (CO, NHAc). Rf =(0.0, EtOAc, 0.1 9:1 EtOAc:MeOH) ^{1}H NMR (300 MHz, MeOD) δ 4.63 (d, J = 8.4 Hz, 1H, H-1), 4.36 (dd, J = 15.9, 2.2 Hz, 2H, CH_2CC), 3.91 (d, J = 11.7 Hz, 1H, H-3), 3.77 – 3.62 (m, J = 17.3, 6.8 Hz, 2H, H-2, H-6), 3.58 – 3.44 (m, 1H, H-6), 3.38 – 3.27 (m, 2H, H-4,H-5), 2.87 (d, J = 2.2 Hz, 1H, CCH), 2.01 (s, 3H,

Me). ^{13}C NMR (75 MHz, MeOD) δ 173.82 (CO), 100.44 (CH, C-1), 80.04(CC), 78.03 (CH, C-5), 76.17(CC), 75.89(CH, C-3), 72.05 (CH, C-4), 62.73(CH_2, C-6), 57.07 (CH, C-2), 56.44(CH_2 ,CH_2CC), 22.99 (Me).

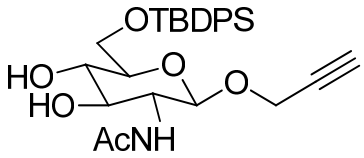
Propargyl-2-acetamido-3,4-tri-O-acetyl-6-O-tert-butyldimethylsilyl-2-deoxy-β–D-glucopyranoside.



To propargyl 2-acetamido-2-deoxy-β-D-glucopyranoside (1g, 2.6mmol) was added TBDMSCl (0.879g, 5.83mmol) and 1 crystal of DMAP. Dry pyridine (35cm^3) was then added and the reaction stirred for 10 hours. The reaction was quenched with MeOH, concentrated *in vacuo* and columned in 100% EtOAc to 10% MeOH. Following recrystallisation from DCM:TolH the product was isolated as a white solid, (0.97g, 67%).

m.p 156-157 $[\alpha]^D$= -29.5 (c=.541, MeOH). Lit. $[\alpha]^{D20}$= -63° (c = 0.1 g cm-1 in $CHCl_3$) [34]. IR (neat) ν_{max} (cm^{-1}) 3287 (br OH) 1656 (CO, NHAc).Rf =(0.125, EtOAC) LRMS (ESI-) calculated for: $C17H_{30}NO_6Si$(M -H) 372.18; found: 372.1 (M -H).^1H NMR (300 MHz, $CDCl_3$) δ 6.46 (d, *J* = 5.5 Hz, 1H, NH), 4.65 (d, *J* = 8.2 Hz, 1H, H-1), 4.36 (d, *J* = 1.7 Hz, 2H, CH_2CC), 3.93 (dd, *J* = 11.0, 3.5 Hz, 1H,H-6), 3.83 (dd, *J* = 11.0, 5.4 Hz, 1H, H-6), 3.74 – 3.64 (m, 1H,H-2), 3.62 – 3.53 (m, 1H, H-3), 3.47 (d, *J* = 8.6 Hz, 1H, H-4), 3.43 – 3.30 (m, 1H, H-5), 2.51 (t, *J* = 2.3 Hz, 1H, CCH), 2.05 (s, 3H, Me), 0.89 (s, 9H, $C(CH_3)_3$), 0.08 (s, 6H, 2x$SiCH_3$). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.74 (CO), 98.15 (CH, C-1), 78.77(CH, CCH), 75.75(CH, C-5), 75.47(CC), 75.10(CH, C-3), 72.09 (CH, C-4), 63.59(CH_2 C-6), 57.07 (CH, C-2), 55.60(CH_2, CH_2CC), 25.87(Me), 23.47(CH_3, $C(CH_3)_3$), 18.32(C, $C(CH_3)_3$), -5.31(CH_3, $SiCH_3$).

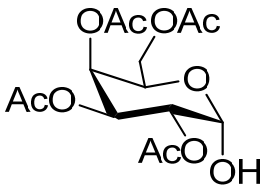
Propargyl-2-acetamido-3,4-tri-O-acetyl-6-O-tert-butyldiphenylsilyl-2-deoxy-β –D-glucopyranoside



To a vessel was charged the propargyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranoside (0.5050g, 1.31mmol) and MeOH (50cm^3). To this was added NaOMe (cat) and the reaction stirred for 1 hour before being quenched with DOWEX-H+ and filtered. After concentration *in vacuo*, the solid was dried with toluene and then TBDPSCl (0.52cm^3, 2mmol), DMAP (cat) and pyridine (13cm^3) were added. The reaction was stirred for 15 hours at ambient temperature, after which TLC indicated it had gone to completion. The reaction was quenched with MeOH and then columned on the Biotage Iscolera (100% EtOAc to 10% MeOH/EtOAc). This yielded a white solid, propargyl 2-acetamido-3,4-tri-*O*-acetyl-6-*O*-*tert*-butyldiphenylsilyl-2-deoxy-β –D-glucopyranoside as the product, (0.5099g, 78% two steps).

LRMS (ESI+) calculated for: $C_{27}H_{35}NNaO_6Si$ (M + Na) 520.65; found: 520.19 (M + Na).Rf=0.30 (EtOAc) m.p=101-103°C. IR (neat) ν_{max} (cm^{-1}) 3276.90 (br NH/OH) 1656 (CO, NHAc) $[\alpha]^D$=-4 (c=0.32, MeOH), lit. $[\alpha]^D$=-64(c=1, $CHCl_3$)[38] ^{1}H NMR (300 MHz, Acetone, δ) 7.82 – 7.76 (m, 4H, ArH), 7.47 – 7.42 (m, 6H, ArH), 7.20 (d, *J* = 7.4 Hz, 1H, NH), 4.74 (d, *J* = 8.0 Hz, 1H, H-1), 4.39 (qd, *J* = 15.6, 2.4 Hz, 2H, CH_2), 4.02 (ddd, *J* = 16.2, 11.0, 3.6 Hz, 2H, H-6), 3.61 (m, 3H,H-2,H-3,H-5), 3.43 (ddd, *J* = 9.5, 5.2, 2.0 Hz, 1H, H-4), 2.98 (t, *J* = 2.4 Hz, 1H, CCH), 1.93 (s, 3H, Me), 1.07 (s, 9H, ^{t}Bu). ^{13}C NMR (75 MHz, Acetone) δ 171.23 (CO), 136.52 (C_{Ar}), 136.43(C_{Ar}), 134.66(C_{Ar}), 134.52(C_{Ar}), 130.58(C_{Ar}), 128.61(C_{Ar}), 99.87(CH, C-1), 80.19 (CC), 77.82 (CH, C-5), 76.64(CC), 76.25(CH, C-3), 71.81(CH, C-4), 64.48(CH, C-6), 57.14 (CH, C-2), 55.62 (CH_2, CH_2CC), 27.22 (Me), 23.34 (CH_3, $C(CH_3)_3$), 19.92(C, $C(CH_3)_3$).

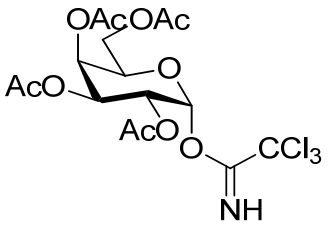
tetra-O-acetyl-α-D-galactose



To a vessel was charged under argon, 1,2,3,4,5-per-O-acetyl-galactose (3.48g, 8.90 mmol) and THF (30cm^3). To this was added a ambient temperature, benzylamine (1.45cm^3, 13.27 mmol). The reaction mixture was stirred for 21 hours. To the reaction mixture was added 1M HCl (40cm^3), and then EtOAc (50cm^3). After separation of the organics, drying (Na_2SO_4) and filtering (gravity), a red oil was obtained after concentration *in vacuo*. The reaction mixture was columned on silica with 1:1 pet ether:ETOAc to yield a single product tetra-*O*-acetyl-α-D-galactose (2.56g, 82.5%).

$[\alpha]^D$= 119.5 (c=.235, MeOH). IR (neat) ν_{max} (cm^{-1}) 3466 (br OH) 1638 (CO). Rf =(0.866, EtOAc) LRMS (ESI-) calculated for: $C_{14}H_{19}O_{10}$ (M -H) 347.10; found: 357.09 (M -H). ^{1}H NMR (300 MHz, $CDCl_3$) δ 5.44 (m, 3H, H-1,H-2, H-4), 5.14 (dd, *J* = 10.7, 3.5 Hz, 1H, H-3), 4.46 (t, *J* = 6.6 Hz, 1H, H-5), 4.11 (m, 2H, H-6), 2.13 (s, 3H, Me), 2.08 (s, 3H, Me), 2.04 (s, 3H, Me), 1.98 (s, 3H, Me). ^{13}C NMR (75 MHz, $CDCl_3$) δ 170.58 (CO), 170.41(CO), 170.26 (CO), 170.09 (CO), 90.64 (CH, C-1), 71.00 (CH, C-5), 68.20 (CH, C-3), 67.25(CH, C-4), 66.19 (CH, C-2), 61.80 (CH_2, C-6), 20.80 (Me), 20.69 (Me), 20.63 (Me), 20.62 (Me).

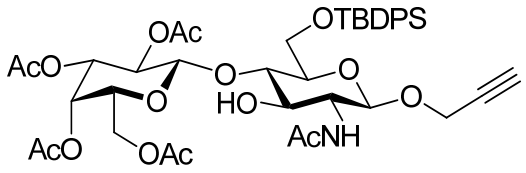
tetra-O-acetyl-α D-galactose trichloroacetimidate



To a dry vessel under argon was charged the tetra-*O*-acetyl-α-D-galactose (2.41g, 6.9mmol), DCM (20cm^3) and DBU (0.32cm^3, 2.14 mmol). To this was then added trichloroacetonitrile (2.1 cm^3, 20.94mmol) and the reaction mixture stirred at ambient temperature for 2 hours. After concentration *in vacuo*, 1M HCl (40cm^3) and DCM (40cm^3) were added. The organics were separated and washed with brine, before being concentrated and dried under vacuum. The red oil was then quickly loaded onto a silica column and columned in 2:1 PET ether: ETOAc yielding the tetra-*O*-acetyl-α-D-galactose trichloroacetimidate (2.30g, 67%).

m.p 123-125 $[\alpha]^D$= -118.9 (c=0.875, $CHCl_3$). Lit. $[\alpha]^D$= -6.9 (c=9, $CHCl_3$)[39] IR (neat) ν_{max} (cm^{-1}) 1743 (CO) 1677 (CO). Rf =(0.82, EtOAc) LRMS (ESI+) calculated for: $C_{16}H_{20}Cl_3NNaO_{10}$ (M+Na) 514.01; found: 514(M +Na). ^{1}H NMR (300 MHz, $CDCl_3$, δ) 8.66 (s, 1H, NH), 6.60 (d, *J* = 3.3 Hz, 1H, H-1), 5.56 (dd, *J* = 2.9, 1.2 Hz, 1H, H-4), 5.39 (m, H-2, H-3), 4.44 (dd, *J* = 7.4, 6.6 Hz, 1H, H-5), 4.21 – 4.03 (m, 2H, H-6), 2.16 (s, 3H, Me), 2.02 (s, 3H, Me), 2.01 (s, 3H, Me), 2.01 (s, 3H, Me).^{13}C NMR (75 MHz, $CDCl_3$, δ) 170.41(CO), 170.22 (CO), 170.19 (CO), 170.09 (CO), 161.10(*C*NHCCl$_3$), 93.70 (CH, C-1), 90.93 (C, CCl_3), 69.16 (CH, C-5), 67.66 (CH, C-3), 67.54(CH, C-4), 67.07(CH, C-2), 61.40(CH_2, C-6), 20.79(Me), 20.76(Me), 20.74(Me), 20.68(Me).

Propargyl-O-(2,3,4,6-tetra-O-acetyl--D-galactopyranosyl)-(β1-4)-2-acetamido-6-O-tertbutyldiphenylsilyl-2-deoxy--D-glucopyranoside

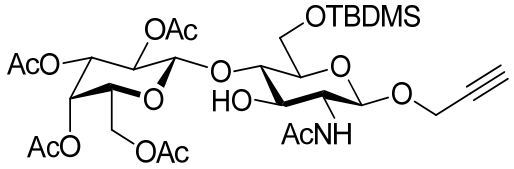


A Schlenk tube was dried overnight in a furnace (450°C), and then 4Å molecular sieves added. The flask was flame dried under vacuum, before being backfilled with dry argon. Once cooled, O-propargyl 2-acetamido-3,4-tri-*O*-acetyl-6-*O*-*tert*-butyldiphenylsilyl-2-deoxy-β–D-glucopyranoside (0.1306g ,0.263mmol, 1 equiv.) and the tetra-*O*-acetyl-α-D-galactose trichloroacetimidate (0.129g, 0.263mmol) were dissolved in anhydrous toluene and added by cannula to the Schlenk tube. The toluene was then evaporated under vacuum and the flask refilled with toluene. This was repeated 3 times and the brown paste dried overnight *in vacuo*. To this was added dry DCM (8cm^3) and the reaction stirred for 30 minutes under nitrogen. The reaction was cooled to -42°C and $BF_3.OEt_2$ (38μL, 0.316mmol) was added. The reaction was stirred for 4 hours before being quenched with $NaHCO_3$ (15cm^3), extracted with DCM (10cm^3), dried with Na_2SO_4, filtered and concentrated *in vacuo.* The reaction was then columned on the Biotage-Iscolera (25g, Si column) with a 2:1 PET ether:EtOAc to 10% MeOH:EtOAc gradient. This yielded the product, propargyl 2-acetamido-6-*O*-*tert*-butyldiphenylsilyl-2-deoxy-4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside as a white solid (0.11g, 50%).

LRMS (ESI+) calculated for: $C_{41}H_{54}NO_{15}Si$ (M +H) 828.32; found: 828.32 m.p 129-132˚C $[\alpha]^D$=-11.2 (c=0.48, $CHCl_3$). lit. $[\alpha]^D$=-6 (c=1, $CHCl_3$)[38] IR (neat) ν_{max}= 3501 (OH), 3279 (NH, NHAc), 2929, 2856, 1749 (CO), 1656 (CO, NHAc) Rf= 0.57 (EtOAc). ^{1}H NMR (500 MHz, $CDCl_3$, δ) 7.74 – 7.69 (m, 4H, H_{Ar}), 7.45 – 7.35 (m, 6H, H_{Ar}), 5.66 (d, *J* = 8.1 Hz, 1H, NH), 5.37 (d, *J* = 3.1 Hz, 1H, H-4 Gal), 5.20 (dd, *J* = 10.4, 8.1 Hz, 1H, H-2 Gal), 4.98 (dd, *J* = 10.4,

3.4 Hz, 1H, H-3 Gal), 4.87 (d, *J* = 8.2 Hz, 1H, H-1 GlcNAc), 4.73 (d, *J* = 8.1 Hz, 1H, H-1 Gal), 4.34 (ddd, *J* = 49.1, 15.7, 2.3 Hz, 2H, CH_2CC), 4.15 (d, *J* = 6.6 Hz, 2H, H-6 Gal), 4.03 – 3.96 (m, 1H, H-3 GlcNAc), 3.93 (m,2H, *H-5 Gal, H-6 GlcNAc*), 3.87 – 3.79 (m,2H, H-6 GlcNAc, H-4 GlcNAc), 3.65 – 3.55 (m, 1H, H-2 GlcNAc), 3.43 (d, *J* = 8.7 Hz, 1H, H-5 GlcNAc), 2.45 (t, *J* = 2.3 Hz, 1H, CH), 2.15 (s, 3H, Me), 2.06 (s, 3H, Me), 2.03 (s, 1H, Me), 1.98 (s, 1H, Me), 1.72 (s, 1H, Me), 1.07 (s, 9H, Me x3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.09(CO), 170.87(CO), 170.56(CO), 170.36(CO), 169.59(CO), 136.40(CH, Ar), 135.95(CH, Ar), 133.90(C, Ar), 132.93(C, Ar) 130.35 (CH, Ar), 128.34(CH, Ar), 128.12(CH, Ar), 101.42(CH, C-1 Gal), 98.09 (CH, C-1 GlcNAc), 80.28(CH, C-4 GlcNAc), 79.20,(C, CCH), 75.52 (CH, CCH), 75.06 (CH, C-5 GlcNAc), 72.04(CH, C-3 GlcNAc), 71.70(CH, C-5 Gal), 71.23(CH, C-3 Gal), 69.27(CH, C-2, Gal), 67.32(CH, C-4 Gal), 62.16 (CH_2, C-6 Gal) 61.64(CH_2, C-6 GlcNAc), 56.85(CH, C-2 GlcNAc), 55.67 (CH_2, CH_2CC), 27.27 (CH_3 x 3, $C(CH_3)_3$), 24.11 (Me), 21.06 (Me), 21.02 (Me), 20.97 (Me), 20.78 (Me), 19.76 (C, $C(CH_3)_3$).

Propargyl-O-(2,3,4,6-tetra-O-acetyl--D-galactopyranosyl)-(β1-4)-2-acetamido-6-O-tertbutyldimethylsilyl-2-deoxy--D-glucopyranoside



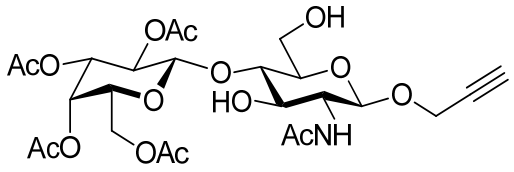
A Schlenk tube was dried overnight in a furnace (450°C), and then 4Å molecular sieves added. The flask was flame dried under vacuum, before being backfilled with dry argon. Once cooled, propargyl 2-acetamido-3,4-tri-*O*-acetyl-6-*O*-*tert*-butyldimethylsilyl-2-deoxy-β–D-glucopyranoside (0.402g, 1.08mmol) and the tetra-*O*-acetyl-α-D-galactose trichloroacetimidate (0.561g, 1.14mmol) were dissolved in anhydrous toluene and added by cannula to the Schlenk tube. The toluene was then evaporated under vacuum and the flask

refilled with toluene. This was repeated 3 times and the brown paste dried overnight *in vacuo*. To this was added dry DCM (25cm^3) and the reaction stirred for 30 minutes under nitrogen. The reaction was cooled to -42°C and $BF_3.OEt_2$ (140µL, 1.14mmol) was added. The reaction was stirred for 7.5 hours before being quenched with $NaHCO_3$, extracted with DCM (30cm^3), dried with Na_2SO_4, filtered and concentrated *in vacuo.* The reaction was then columned on silica with a 10% MeOH:EtOAc gradient. This yielded the product, propargyl 2-acetamido-6-*O*-*tert*-butyldiphenylsilyl-2-deoxy-4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside as a white solid (0.197g, 26%).

LRMS (ESI+) calculated for: $C_{31}H_{50}NO_{15}Si$ (M +H) 704.29; found: 704.28 Rf =0.42(EtOAc), IR (neat) vmax= 3532 (OH), 1731 (CO), 1663 (CO, NHAc), m.p =189.1-190, $[\alpha]^D$=-15.5 (c=0.4, $CHCl_3$).lit $[\alpha]^{D20}$=-26° (c = 0.1 g cm-1 in $CHCl_3$) [34]. ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.51 (d, *J* = 8.0 Hz, 1H, NH), 5.40 (d, *J* = 3.1 Hz, 1H, H-4 Gal), 5.23 (dd, *J* = 10.4, 8.1 Hz, 1H, H-2 Gal), 5.02 – 4.97 (m, 1H, H-3 Gal), 4.85 (d, *J* = 8.2 Hz, 1H, H-1 Gal), 4.63 (d, *J* = 8.1 Hz, 1H, H-1 GlcNAc), 4.35 (qd, *J* = 15.8, 2.4 Hz, 2H, CH_2CC), 4.17 – 4.11 (m, *J* = 9.7, 6.5, 2.7 Hz, 3H, H-6 Gal, H-5 Gal), 3.97 (m, 1H, H-3 GlcNAc), 3.85 (m *J* = 11.5, 1.6 Hz, 1H, H-6 GlcNAc), 3.74 (m, 1H, H-6 GlcNAc), 3.62 (t, *J* = 8.8 Hz, 1H, H-4 GlcNAc), 3.52 (dd, *J* = 18.0, 8.2 Hz, 1H, H-2 GlcNAc), 3.39 (d, *J* = 7.5 Hz, 1H, H-5 GlcNAc), 2.44 (t, *J* = 2.3Hz, 1H, CCH), 2.17 (s, 3H, Me), 2.08 (s, 3H, Me), 2.07 (s, 3H, Me), 2.03 (s, 3H, Me), 1.99 (s, 3H, Me), 0.92 (s, 9H, $C(CH_3)_3$), 0.09 (s, 3H, $SiCH_3$), 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (75 MHz, $CDCl_3$) δ 170.70 (CO), 170.45(CO), 170.12(CO), 169.99(CO), 169.28(CO), 101.44 (CH, C-1 Gal), 97.78 (CH, C-1 GlcNAc), 80.80 (CH, C-4 GlcNAc), 78.82 (C, CC), 75.14(CH, CC), 74.75 (CH, C-5 GlcNAc), 72.03 (CH, C-3 GlcNAc), 71.21(CH, C-5 Gal), 70.88(CH, C-3 Gal), 68.84 (CH, C-2 Gal), 66.90(CH, C-4 Gal), 61.53(CH_2, C-6 Gal), 61.41(CH_2, C-6 GlcNAc), 55.94 (CH, C-2 GlcNAc), 55.22 (CH_2, CH_2CC), 25.86(CH_3, $C(CH_3)_3$)), 23.54(Ac, CH_3), 20.70(Ac,

CH_3), 20.60(Ac, CH_3), 20.52(Ac, CH_3), 18.25(C, *C*$(CH_3)_3$)), -4.96(CH_3, $SiCH_3$), -5.22 (CH_3, $SiCH_3$).

2-acetamido-2-deoxy-4-O-(2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl)-1-propargyl-β-D-glucopyranoside



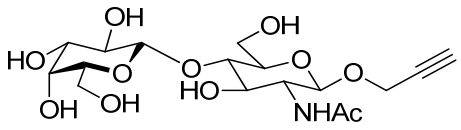
To a flask was charged the propargyl *O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-(1→4)-2-acetamido-6-*O*-*tert*butyldiphenylsilyl-2-deoxy-β-D-glucopyranoside:, (86.3mg, 0.12mmol) to this was added under N_2, dry THF (2cm^3) and 37% $Et_3N.3HF$ (0.11cm^3, .84mmol). The reaction was stirred for 1 week at ambient temperature until by TLC it was complete. The reaction was loaded directly onto silica and columned (10% MeOH/EtOAc) to yield a white solid as the product (26mg, 42%).

To a flask was charged propargyl *O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-(1→4)-2-acetamido-6-*O*-*tert*butyldiphenylsilyl-2-deoxy-β-D-glucopyranoside (44mg, 0.05mmol) and dry MeCN (1.5cm^3) added under argon. To this was added $BF_3.OEt_2$ (11μL, 0.09mmol) at ambient temperature and the reaction stirred until no further start material was observed (4 hours). After concentration *in vacuo*, the crude reaction mixture was loaded onto a silica column and eluted with 10% MeOH/EtOAc. This gave the product as a white solid (28mg, 89%).

To a flask was charged the propargyl *O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-(1→4)-2-acetamido-6-*O*-*tert*butyldimethyllsilyl-2-deoxy-β-D-glucopyranoside:, (84mg, 0.12mmol) to this was added under N_2, dry THF (2cm^3) and 37% $Et_3N.3HF$ (0.11cm^3, 0.84mmol). The reaction was stirred for 16 hours at ambient temperature. The reaction was loaded directly onto silica and columned (10% MeOH/EtOAc) to yield a white solid as the product (40mg, 42%).

LRMS (ESI+) calculated for: $C_{25}H_{36}NO_{15}$ (M +H) 590.2; found: 590.19. Rf =0.16 (EtOAc). m.p =107-110, IR (neat) ν_{max} (cm^{-1}) 3468 (NH) 3275 (OH), 1743 (CO, NHAc). $[\alpha]^D$=-21.5 (c=0.065, MeOH). Lit. $[\alpha]^{D20}$-19.5° (c = 0.1 g cm-1 in $CHCl_3$) [34] .^{1}H NMR (500 MHz, MeOD, δ) 5.44 (d, J = 2.5 Hz, 1H, H-4 Gal), 5.18 (m, H-3, H-2 gal, 2H), 4.86 (d, J = 8.9 Hz, 1H, H-1 Gal), 4.65 (d, J = 8.4 Hz, 1H, H-1 GlcNAc), 4.44 – 4.35 (m, 2H, CH_2CC), 4.29 – 4.15 (m, 3H, H-5 Gal, H-6a/6b Gal), 3.85 (dt, *J* = 8.5, 4.3 Hz, 1H, H-6 GlcNAc), 3.81 – 3.72 (m, 1H, H-2 GlcNAc), 3.69 – 3.57 (m, 3H, H-4, H-3, H-6 GlcNAc), 3.39 (ddd, *J* = 9.4, 4.5, 1.7 Hz, 1H, H-5, GlcNAc), 2.90 (t, *J* = 2.4 Hz, 1H), 2.18 (s, 3H, Me), 2.13 (s, 3H, Me), 2.06 – 2.06 (s, 3H, Me), 2.01 (s, 3H, Me), 1.98 – 1.97 (s, 3H, Me). ^{13}C NMR (126 MHz, MeOD) δ 174.03 (C=O), 172.64 (C=O), 172.29 (C=O),, 171.80 (C=O),, 171.72 (C=O), 102.83 (CH, C-1 Gal), 100.69 (CH, C-1, GlcNAc), 82.36 (CH,H-4 GlcNAc), 79.31 (C, CC) 77.27 (CH, CC) 76.78(CH,H-5 GlcNAc), 74.36(CH,H-3 GlcNAc), 72.74 (CH, C-5/C-3 Gal) 70.91 (CH, C-2 Gal), 69.19 (CH, C-4 Gal), 63.13(CH2, C-6, GlcNAc) 61.95 (CH_2, C-6 Gal), 56.93 (C-H, C-2 GlcNAc), 56.81 (CH_2, CH2CCH), 23.38 (CH3, Me), 21.16 (CH3, Me), 21.02 (CH_3, Me), 20.91 (CH3, Me x2).

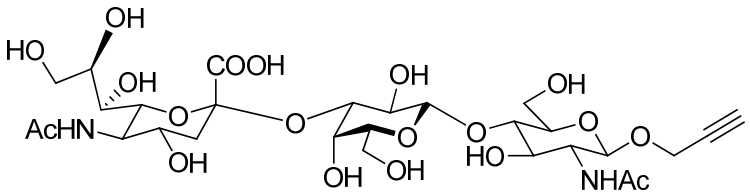
2-Acetamido-2-deoxy-4-O--D-galactopyranosyl-1-propargyl-D-glucopyranoside



To propargyl *O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-(1→4)-2-acetamido-2-deoxy-β-D-glucopyranoside (47.5mg, 80.6μmol) was added MeOH (6cm^3). To this was added at ambient temperature NaOMe solution (0.1cm^3, 0.5M). The reaction was stirred for 30 minutes, and then neutralised with DOWEX-H+. Upon reaching pH7 it was filtered through a pipette packed with glass wool and concentrated *in vacuo* to give fully deprotected product 2-acetamido-2-deoxy-4-O-β-D-galactopyranosyl-1-propargyl-D-glucopyranoside (31.3mg, 92%).

LRMS (ESI+) calculated for: $C_{17}H_{28}NO_{11}$ (M +H) 422.16; found: 422.16. Rf=0.62(1:2:2 WIPE). m.p = 239-242˚C $[\alpha]^D$=-83.3 (c=0.015, MeOH). Lit. $[\alpha]^{D20}$=-60° (c = 0.1 g cm^{-1} in MeOH) [34]. IR (neat) ν_{max}= 3362 (br OH), 3262 (br OH) 1737(CO, NHAc).^{1}H NMR (500 MHz, D_2O) δ 4.75 (d, *J* = 7.8 Hz, 1H, H-1 GlcNAc), 4.48 (d, *J* = 7.8 Hz, 1H, H-1 Gal), 4.42 (d, *J* = 2.2 Hz, 2H, CH2CC), 4.00 (dd, *J* = 12.3, 2.1 Hz, 1H, H-6a GlcNAc), 3.93 (d, *J* = 3.3 Hz, 1H, H-4 Gal), 3.84 (dd, *J* = 12.3, 5.1 Hz, 1H, H-6 GlcNAc), 3.80 – 3.71 (m, 6H, H-2,H-3,H-4 GlcNAc, H-5 Gal, H-6 Gal), 3.68 (dd, *J* = 10.0, 3.4 Hz, 1H, H-3 Gal), 3.65 – 3.59 (m, 1H, H-5 GlcNAc), 3.55 (dd, *J* = 9.9, 7.9 Hz, 1H, H-2 Gal), 2.92 (t, *J* = 2.3 Hz, 1H, CCH), 2.05 (s, 3H, Me). ^{13}C NMR (126 MHz, D_2O) δ 175.13 (CO), 103.25(CH, C-1 Gal), 99.70 (CH, C-1 GlcNAc), 79.12 (C-4, GlcNAc), 78.68, 76.53, 75.73, 75.24, 72.87, 72.75, 71.33, 68.92(CH, C-4 Gal), 61.40 (C-6 GlcNac), 60.35(CH_2, C-6 Gal), 57.08(CH_2, C*H*$_2$CC),, 55.23 (CH, C-2 GlcNAc),, 22.58 (CH_3).

Propargyl-(5-acetimido-3,5-dideoxy-d-glycero-D-galacto-2-nonulopyranosylonic acid-(α2-3)- -D-galactopyranosyl-(1-4)-2-acetimido-2-deoxy--D-glucopyranoside (SiaLacNAc-alkyne)



To a flask containing the lactosamine (19.58mg, 0.04649mmol) was added the CTP (36.75mg,0.076mmol), 100mM pH 8.5 $Mg(OAc)_2$ buffer and Sialic acid (21.56mg, 0.076mmol), the pH was then adjusted back to 8.5. NMCSS (138uL, 1.67mgml^{-1}) and PMST(8.81μL, 5.2mgml^{-1}) were added and the reaction shaken at 37°C for 3 hours. The reaction was subsequently concentrated and columned with 1:2:2 water:IPA:EtOAc (WIPE) to give 23. 1 mg of pure product (69%). m.p 150-155˚C. $[\alpha]^D$=+48.1 (c=0.027, 4:1 MeOH:H_2O). Lit. $[\alpha]^D$ +7.6 (c = 0.5 g cm^{-3} in H_2O:DMSO 5:1).[17] IR (neat) ν_{max}= 3329 (br OH), 1633 (CO, NHAc). Rf =0.32 (1:2:2 WIPE) LRMS (ESI-) calculated for: $C_{28}H_{44}N_2O_{19}$ (M-H) 711.25; found: 711.25 (M -H).. ^{1}H NMR (500 MHz, D_2O) δ 4.49 (d, *J* = 7.8 Hz, 1H, H-Gal-1), 4.35 (s, 2H, CH_2CC), 4.05 (d, *J* = 9.8 Hz, 1H, H-3 Gal), 3.97 – 3.87 (m, 2H), 3.87 – 3.74 (m, 4H), 3.75 – 3.47 (m, 12H), 2.85 (s, 1H, CCH), 2.69 (dd, *J* = 12.4, 3.8 Hz, 1H, Neu5Ac H-3), 2.00 – 1.93 (m, 6H, 2xMe), 1.74 (t, *J* = 11.8 Hz, 1H, Neu5Ac H-3). ^{13}C NMR (126 MHz, D_2O, δ) 175.02 (C=O), 174.76 (C=O), 173.90(C=O, Neu-5-Ac), 102.55 (CH, C-1 Gal), 99.81 (C, C-1, Neu-5-Ac), 99.38 (CH, C-1, GlcNAc) 78.75, 78.12, 76.18, 75.44, 75.15, 74.84, 72.86, 72.33, 71.77, 69.37, 68.34, 68.08, 67.48, 62.56, 61.02 (CH2, C-6, Gal) , 59.94 (CH2, C-6 GlcNAc) 56.73 (CH2, Ch2CCH), 54.84 (CH, C-2 GlcNAc), 51.65 (C-2, GlcNAc), 39.59, 22.22 (CH3), 22.06 (CH3).

3.6 Bibliography

(1) Angata, T.; Varki, A. *Chem. Rev.* **2002**, *102* (2), 439–469.

(2) Boons, G.-J.; Demchenko, A. V. *Chem. Rev.* **2000**, *100* (12), 4539–4566.

(3) Crich, D.; Li, W. *J. Org. Chem.* **2007**, *72* (20), 7794–7797.

(4) Pazynina, G.; Nasonov, V.; Belyanchikov, I.; Brossmer, R.; Maisel, M.; Tuzikov, A.; Bovin, N. *Int. J. Carbohydr. Chem.* **2010**, *2010*, 1–8.

(5) Martichonok, V.; Whitesides, G. M. *J. Org. Chem.* **1996**, *61* (5), 1702–1706.

(6) Rillahan, C. D.; Schwartz, E.; McBride, R.; Fokin, V. V.; Paulson, J. C. *Angew. Chemie - Int. Ed.* **2012**, *51* (44), 11014–11018.

(7) Sun, M.; Li, Y.; Chokhawala, H. a; Henning, R.; Chen, X. *Biotechnol. Lett.* **2008**, *30* (4), 671–676.

(8) Li, Y.; Yu, H.; Cao, H.; Muthana, S.; Chen, X. *Appl. Microbiol. Biotechnol.* **2012**, *93* (6), 2411–2423.

(9) Ding, L.; Yu, H.; Lau, K.; Li, Y.; Muthana, S.; Wang, J.; Chen, X. *Chem. Commun.* **2011**, *47* (30), 8691–8693.

(10) Li, Y.; Yu, H.; Cao, H.; Lau, K.; Muthana, S.; Tiwari, V. K.; Son, B.; Chen, X. *Appl. Microbiol. Biotechnol.* **2008**, *79* (6), 963–970.

(11) Cheng, J.; Huang, S.; Yu, H.; Li, Y.; Lau, K.; Chen, X. *Glycobiology* **2009**, *20* (2), 260–268.

(12) Muthana, S.; Yu, H.; Huang, S.; Chen, X. *J. Am. Chem. Soc.* **2007**, *129* (39), 11918–11919.

(13) Chokhawala, H. A.; Cao, H.; Yu, H.; Chen, X. *J. Am. Chem. Soc.* **2007**, *129* (35), 10630–10631.

(14) Ni, L.; Sun, M.; Yu, H.; Chokhawala, H.; Chen, X.; Fisher, A. J. *Biochemistry* **2006**, *45* (7), 2139–2148.

(15) Chen, J.; Liu, Y.; Li, X.; Wang, Y.; Ding, H.; Ma, G.; Su, Z. *Protein Expr. Purif.* **2009**, *66* (1), 82–90.

(16) Shengshu, H.; Hai, Y.; Xi, C. *Sci. China. Chem.* **2011**, *54* (1), 117–128.

(17) van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature* **2007**, *446* (7139), 1105–1109.

(18) Wrodnigg, T. M.; Stutz, A. E. *Angew. Chemie - Int. Ed.* **1999**, *38* (6), 827–828.

(19) Wrodnigg, T. M.; Lundt, I.; Stütz, A. E. *J. Carbohydr. Chem.* **2006**, *25* (1), 33–41.

(20) Frongia, A.; Secci, F.; Capitta, F.; Piras, P. P.; Sanna, M. L. *Chem. Commun. (Camb).* **2013**, *49* (78), 8812–8814.

(21) Hara, A.; Imamura, A.; Ando, H.; Ishida, H.; Kiso, M. *Molecules* **2013**, *19* (1), 414–437.

(22) Albers, P. W.; Möbus, K.; Wieland, S. D.; Parker, S. F. *Phys. Chem. Chem. Phys.* **2015**, *17* (7),

5274–5278.

(23) Shan, Y.; Oulaidi, F.; Lahmann, M. *Tetrahedron Lett.* **2013**, *54* (30), 3960–3961.

(24) Roy, B.; Mukhopadhyay, B. *Tetrahedron Lett.* **2007**, *48* (22), 3783–3787.

(25) Christensen, H.; Christiansen, M. S.; Petersen, J.; Jensen, H. H. *Org. Biomol. Chem.* **2008**, *6* (18), 3276–3283.

(26) Allman, S. A.; Jensen, H. H.; Vijayakrishnan, B.; Garnett, J. A.; Leon, E.; Liu, Y.; Anthony, D. C.; Sibson, N. R.; Feizi, T.; Matthews, S.; Davis, B. G. *ChemBioChem* **2009**, *10* (15), 2522–2529.

(27) Knapp, S.; Huhn, R.; Benjamin, A. *Org. Synth.* **2007**, *84* (September), 68.

(28) Pawlak, J. B.; Gential, G. P. P.; Ruckwardt, T. J.; Bremmers, J. S.; Meeuwenoord, N. J.; Ossendorp, F. A.; Overkleeft, H. S.; Filippov, D. V; Van Kasteren, S. I. *Angew. Chemie - Int. Ed.* **2015**, *54* (19), 5628–5631.

(29) van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G. *Nat. Protoc.* **2007**, *2* (12), 3185–3194.

(30) Yu, H.; Chen, X. *Org. Biomol. Chem.* **2016**, *14* (10), 2809–2818.

(31) Paulsen, H.; Tietz, H. *Angew. Chemie Int. Ed. English* **1982**, *21* (12), 927–928.

(32) Crocker, P. R. *Curr. Opin. Pharmacol.* **2005**, *5* (4), 431–437.

(33) Angata, T.; Nycholat, C. M.; Macauley, M. S. *Trends Pharmacol. Sci.* **2015**, *36*, 645–660.

(34) Daly, R.; Vaz, G.; Davies, A. M.; Senge, M. O.; Scanlan, E. M. *Chem. - A Eur. J.* **2012**, *18* (46), 14671–14679.

(35) Pérez, E. M. S.; Ávalos, M.; Babiano, R.; Cintas, P.; Light, M. E.; Jiménez, J. L.; Palacios, J. C.; Sancho, A. *Carbohydr. Res.* **2010**, *345* (1), 23–32.

(36) Andersen, S. M.; Heuckendorff, M.; Jensen, H. H. *Org. Lett.* **2015**, *17* (4), 944–947.

(37) Masania, J.; Li, J.; Smerdon, S. J.; Macmillan, D. *Org. Biomol. Chem.* **2010**, *8* (22), 5113.

(38) Rauthu, S. R.; Shiao, T. C.; André, S.; Miller, M. C.; Madej, É.; Mayo, K. H.; Gabius, H. J.; Roy, R. *ChemBioChem* **2015**, *16* (1), 126–139.

(39) Pilgrim, W.; Murphy, P. V. *J. Org. Chem.* **2010**, *75* (20), 6747–6755.

Chapter 4

Protein modification, at endogenous and unnatural sites, for the synthesis of a dual-modified protein library

Contents Page

4.1 Introduction and aims

Protein modification reactions can be conducted in a site selective or site specific fashion. Site selective reactions at endogenous residues exploit subtle nuances in the residue microenvironment in order to show selectivity for one site over another. Examples of site selective reactions include lysine NHS ester reactions and cysteine alkylation.[1] In contrast, specific reactions occur at a specific site, or reactive handle, due to the intrinsic chemical reactivity of the target site. Much of the work around site specific protein modification reactions has been developed by organic chemists leveraging the developments in UAA incorporation methodologies, and is reviewed in Chapter 1. Examples of site specific reactions include: the CuAAC reaction,[2] Suzuki reaction[3,4] and Sonogashira reaction.[5] The installation of reactive "tags" and their specific modification was the basis for the tag and modify strategy proposed by Davis *et al*.[6] If successful, the methodology developed here would represent an advanced application of this strategy. This chapter aims to prepare dual modified Np276 ruler constructs functionalised with both sugar and with dye moieties. The end goal is to create dual modified probes of sugar binding, which can be tracked by fluorescence (flow cytometry). To complete these modifications, a strategy for homogenous dye modification needed to be developed. Previous work (based on manufacturer's recommendations by Dr Navvarro and Dr Errey) hadn't yielded any homogenously modified constructs but instead a mix of products. Furthermore, no consideration had been given to the protein structure after these modifications. In deciding upon a strategy, the reactions could be conducted in a [1+2] sequence (selective dye conjugation followed by specific CuAAC reaction) or in a [2+1] sequence (specific CuAAC reaction followed by a selective dye conjugation). The chosen order will come down to both ease of synthesis, as well as whichever route minimises wastage of the more scarce or expensive reagent. In this chapter, both routes will be considered.

4.2 Results and Discussion

4.2.1 Protein dye labelling via endogenous amino acid reactions

4.2.1.1 Active ester labelling: NHS biotin model system

Using NHS biotin as a model system for an N-terminal NHS conjugation (due to its low cost vs an Alexa dye), Np276 5Å G2P protein was reacted with NHS biotin (in DMSO) at varying temperatures and equivalents. The results from these trials indicated that multiple modifications were observed in nearly all cases. Interestingly, despite the protein system only have 2 (ε/α) amino groups, up to 8 modifications were observed by LCT-ESI-TOF-MS (Figure 1.), the dye systems were thus trialled to observe whether this multiple modification was due to the biotin moiety.

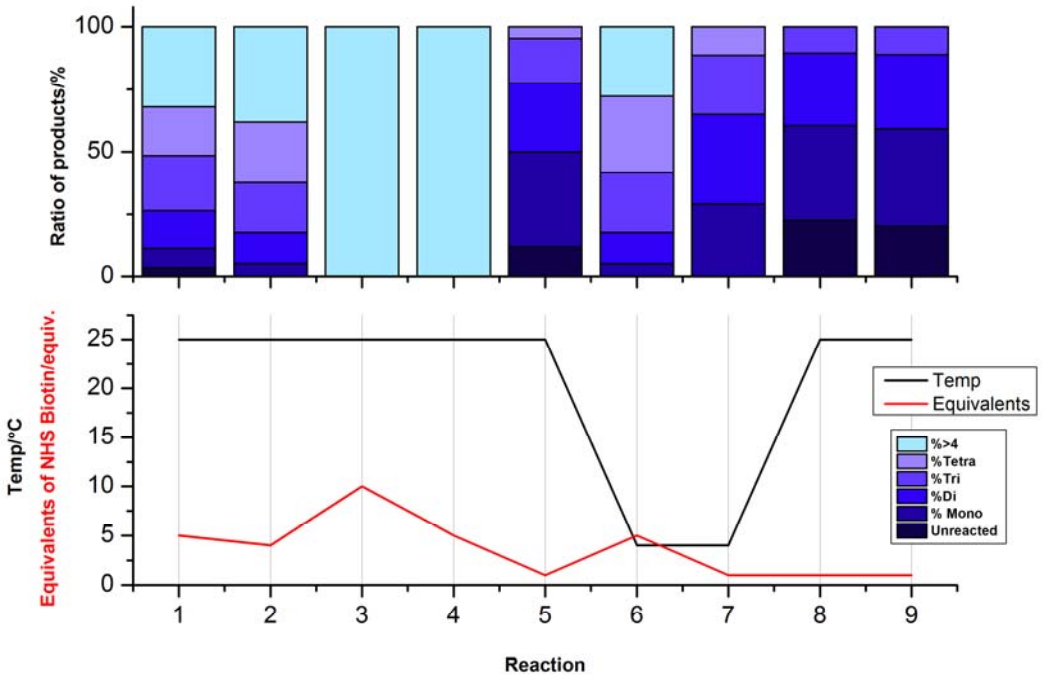


Figure 1. Labelling of Np276 with NHS biotin. (Top) A bar chart showing the distribution of products in the reaction. (Bottom) The experimental conditions used in these reactions. The reactions proved hard to control with multiple biotin modifications seen in all of the reactions.

4.1.1.2 Alexa-488 NHS and SDP ester modification

Alexa-488 NHS ester is a proprietary product of Life-Technologies, with enhanced brightness, photostability, and pH insensitivity when compared to fluorescein based derivatives. This would make it an ideal dye to use in cell-based studies, even though the cost is very high. Model labelling reactions were conducted with Np276 5Å G2P T201K M61-81, with the Alexa-488 being made up as a stock solution in anhydrous DMSO. The results of these modification trials are shown in Figure 2.

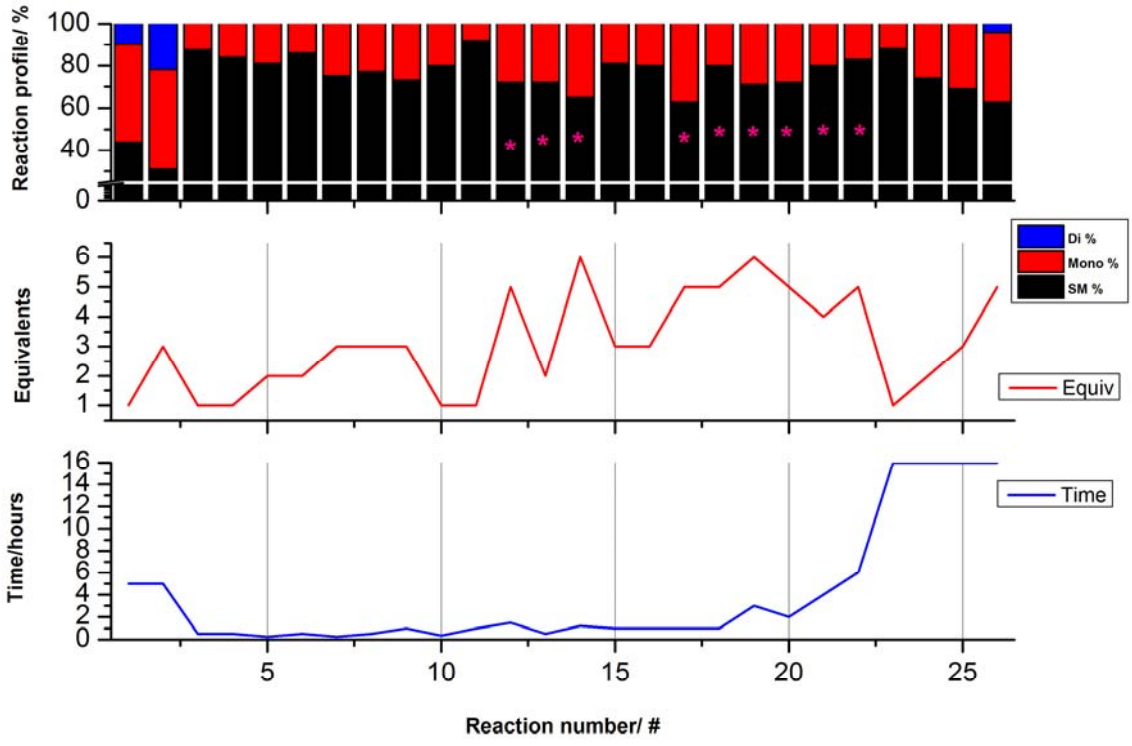


Figure 2. NHS ester reactions with NHS Alexa-488. In reactions with a red star portion-wise addition was used. In 3 cases di labelling was observed, however in most cases a significant proportion of start material remained. The top bar chart shows the proportion of reaction products observed by LCT-ESI-TOF-MS, whilst the bottom graph show the reaction conditions.

What was clear, as previously discussed, was that NHS esters offered little control in their site selectivity. Furthermore, the cost of reagents and instability of the reagents made them less

viable as reagents for the creation of dual modified libraries. Trials with the more stable SDP derivative also didn't lead to a homogeneously modified product, with a significant proportion of start material still being observed (Figure 3).

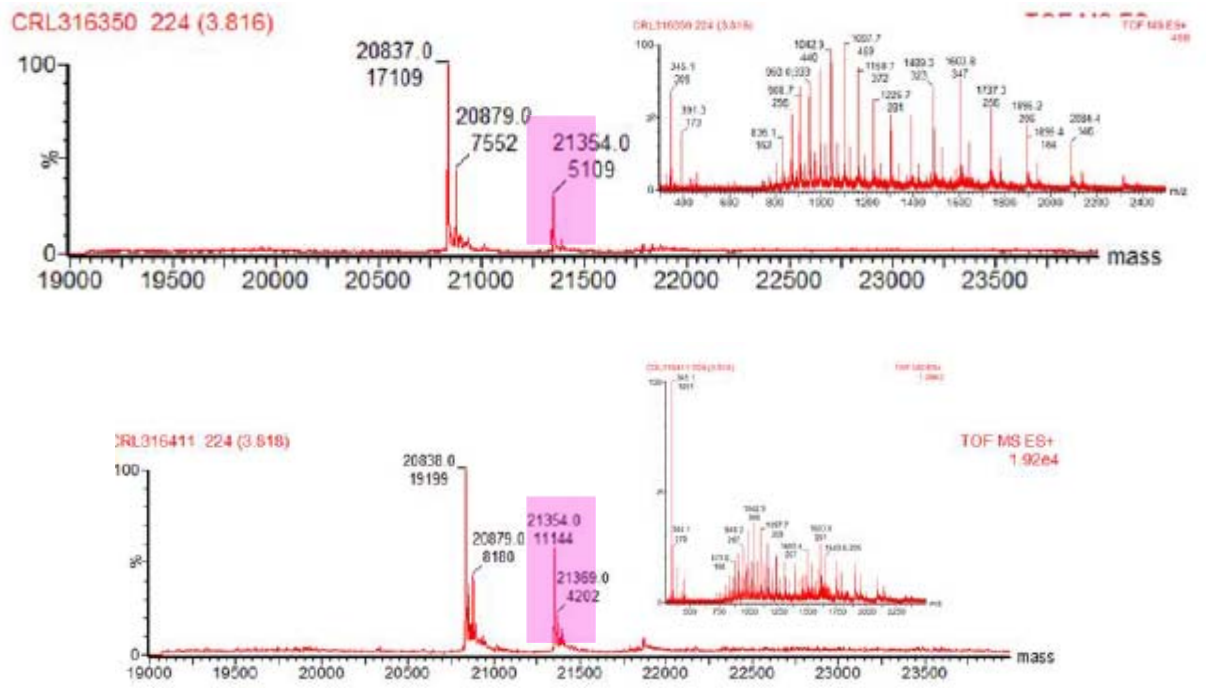


Figure 3. Alexa-488 SDP reactions (Top) 1 equivalent of Alexa-488 SDP ester was added at room temperature and the reaction rotated for 3 hours (17% conversion). (Bottom) 2 equivalents of Alexa-488 SDP ester was added at 4°C and the reaction rotated for 3 hours (29 % conversion). The SDP modified protein is highlighted in pink.

4.2.1.3 Isothiocyanate modification background

The modification of the N-terminus of proteins with FITC has been conducted in the classical bio-conjugation literature through the careful control of pH (with the limited examples of glycine, phenylalanine and aspartic acid).[1] In 2009, Brown *et al* investigated the modification of macrophage migration inhibitory factor (MIF) with dietary isothiocyanates.[7] In this report, it was found that the N-terminal proline is the selective modification site. The N-terminal proline, in this case, lies within the active site of the enzyme, Brown *et al* suggest this results in a lowering of its pKa from ~9.5 in the free form, to 5.6 in the protein (Figure 4).[8–11]

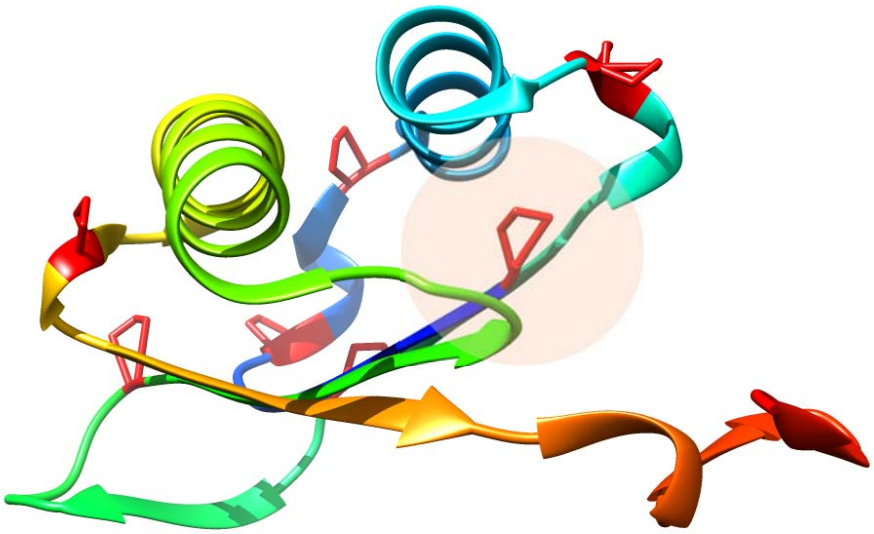


Figure 4. Crystal structure of MIF (3B9S). The N-terminal proline is at the end of the dark blue chain and is highlighted with an orange circle. The reactivity of this N-terminal proline towards dietary isothiocyanates has led to the design of irreversible isothiocyanate inhibitors.[10]

Whilst the Np276 N-terminal proline isn't found within an active site, proline does have precedent as a highly nucleophilic species (It is widely used as a nucleophilic organocatalyst in the organic literature).[12] Furthermore, the isothiocyanates bind the proline residue in an irreversible manner, which would make this an ideal target for modification.[10] To date there

are limited reports of other N-terminal proline modification (reported in last 3 years), the oxidative coupling developed by Francis *et al* a notable example.[13]

4.2.1.4 Isothiocyanate modification optimisation

The pH selectivity of the FITC modification was studied to ascertain an optimal pH for the reaction. Thankfully in pH 8.0 buffer, rapid modification was observed with only one modification being observed. By using pH 8.0, the buffer for purification and the CuAAC reaction could also be kept constant, minimising the need for additional buffer exchange steps.

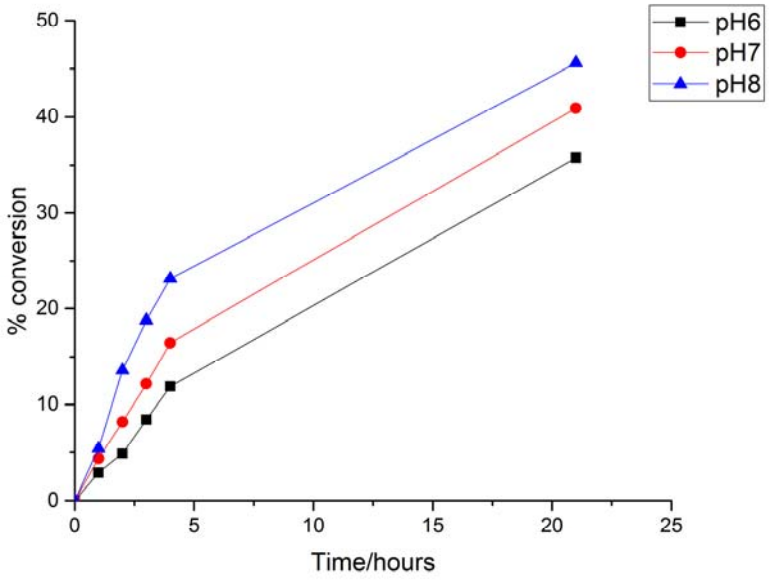


Figure 5. % conversion versus time assay for Np276 modification reactions using Np276/FITC. These reactions were conducted at a variety of pH buffers, in 50mM NaPi. The reaction proceeded the best in pH 8.0 buffer. Data points collected are shown with symbols with lines between for clarity.

The number of dye equivalents needed for efficient modification was investigated with 5Å G2P Met 61-81 T201K protein, at a concentration of 1.5mgml^{-1}. The reactions were rotated at ambient temperature (40rpm) and the ratio of products determined by MS intensity calculations. What can be seen from Figure 6 is that the reaction proceeds much more rapidly with higher dye concentrations. In all cases, no dual dye modification was seen (pH 8.0, 50mM

NaPi), based on these results 10-20 equivalents of dye were routinely used to modify the proteins often using multiple additions.

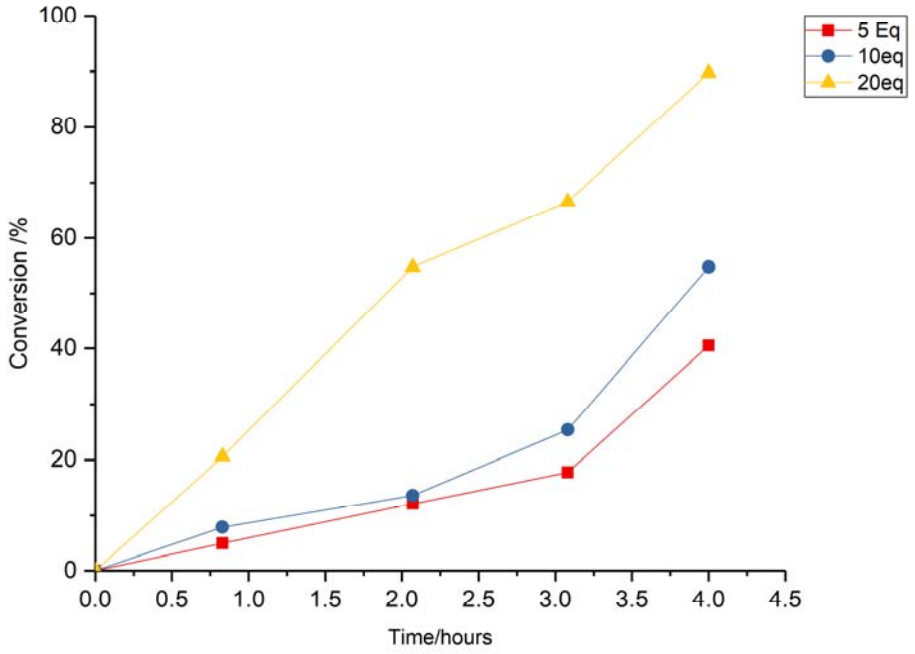


Figure 6. % conversion versus time assay for Np276 modification reactions using Np276/FITC. These reactions were conducted with varying quantities of FITC. A higher quantity of FITC resulted in a higher conversion. 20eq was shown to be much better than 10eq. Using 10eq gave only marginal gains in conversion versus 5eq. Data points collected are shown with symbols with lines between for clarity.

Additional studies (experimental), also showed that at higher protein concentrations a higher % conversion was observed. Furthermore, even faster rates of conversion could be observed by using a $NaHCO_3$ (higher pH) buffer. Despite this, the NaPi buffer system was used in subsequent reactions with a range of the Np276 ruler proteins due to its compatibility in the CuAAC/protein stability. In all cases, the protein reaction was monitored by LCT-ESI-TOF-MS until full conversion was observed (Figure 7).

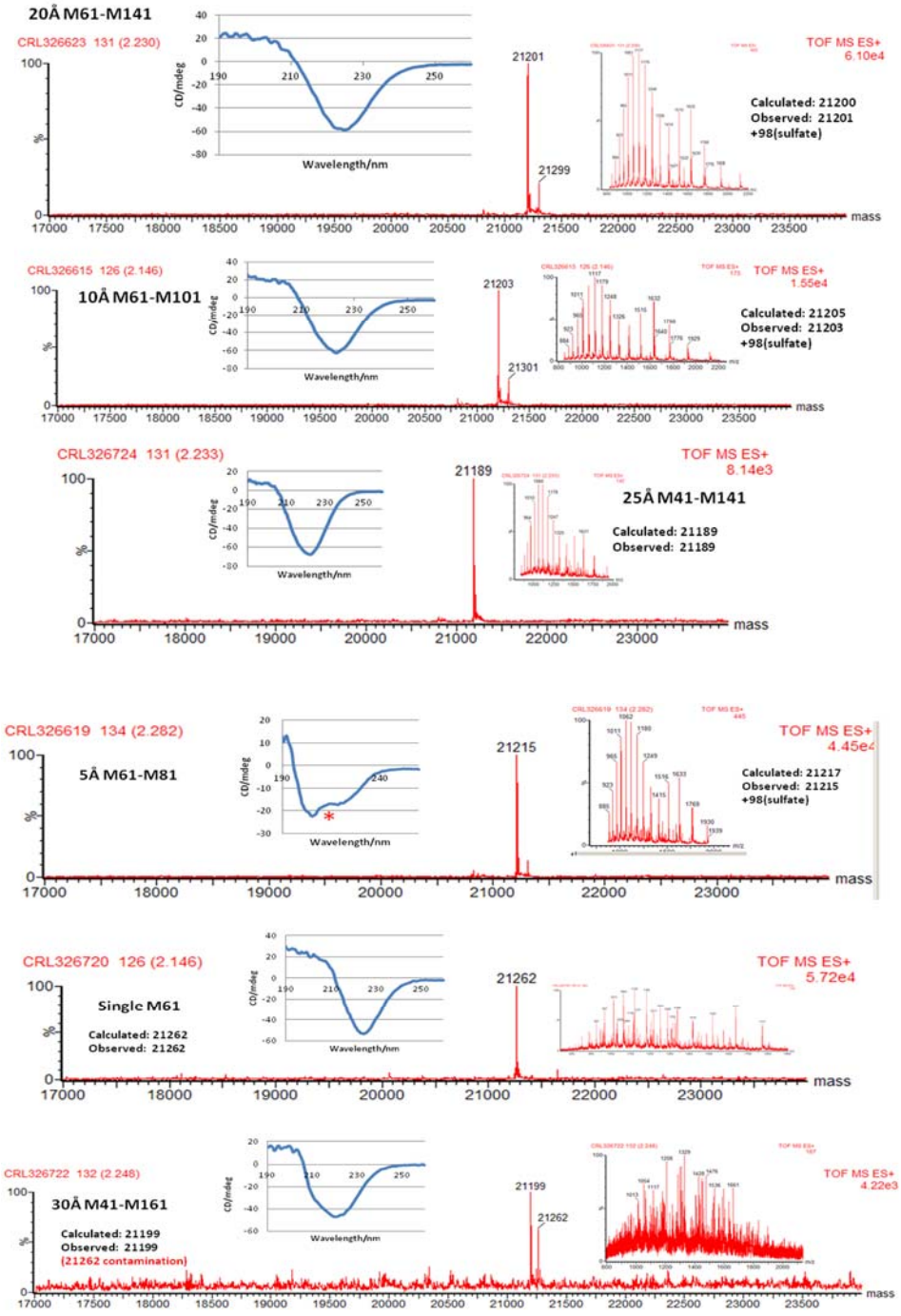


Figure 7. The homogeneous modification of Np276 proteins with FITC dye. The +98Da adduct was previously shown to be a sulfate adduct. The perturbation of the protein secondary structure (β-quadrilateral helix) into the α-structure can observed in the 5Å ruler experiment.

While in some cases the deleterious α-structure was observed, in the majority of examples, by keeping the reactions cold this could be prevented. Furthermore, the dye was rapidly removed by use of a PD-Miditrap and also Pierce dye removal resin, this ensures that little of the residual

dye was present in the subsequent CuAAC. Despite this modification only occurring once, there was no proof of the site selective nature of the modification. Thus MS/MS and MS studies were conducted.

4.2.2 Proteomics: Trypsin and AspN digests

The tryptic digest and subsequent ZipTip® clean-up led to no observation of the Lys-201 FITC modification, likewise there was no evidence of the Pro-2 FITC modification or of the Pro-2 start material However there was abundant evidence for the unmodified Lys-201 (Figure 8).

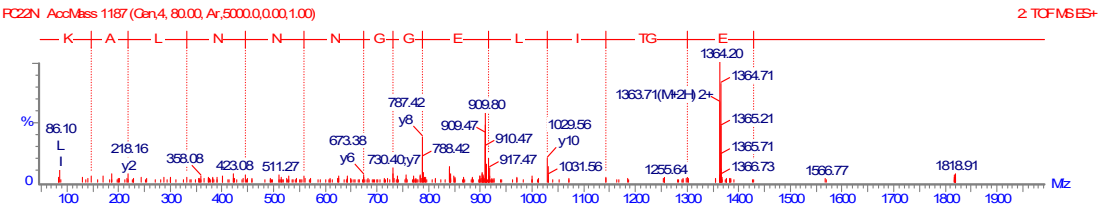


Figure 8. T9 (ANLTGANLEDANLEGTILEGGNNNLAK). Score 195, delta 0.05. No evidence could be found for the Pro-2 FITC modification, nor any evidence for the modification at Lys-201.

The AspN enzyme was thus used. The AspN/ ZipTip® procedure was analysed by MS/MS as well, in this case many more peptides were observed. The analysis of the D1 and D16 peptides was undertaken. The D16 peptide was observed as the free lysine (D16) (no FITC/formyl/acetyl etc), conversely the HisTag (D1) was not observed as a free peptide.

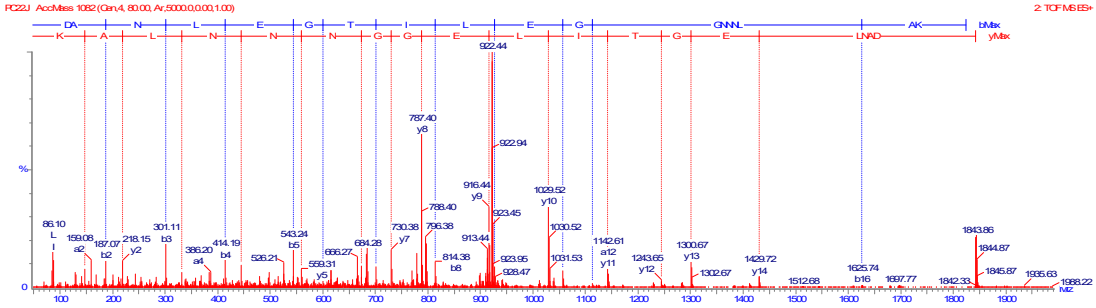


Figure 9. D16 (DANLEGTILEGGNNNLAK) Score 450. Delta 0.01. No-modification at the C-terminal lysine was observed with FITC.

Again for the D1 fragment, little fragmentation was observed. A $(M+H)^{3+}$ peak was seen for the formylated species and again no evidence for D1 (no formyl) or D1(FITC) was seen with the digestions conducted with AspN several variables were investigated to ascertain whether these were influencing the identification of the desired modified peptide. Various conditions were then trialled including concentrations of acetonitrile in the elution buffer (40-75%), digest times, purification methods (Sep-Pak light vs ZipTip®), acid or base quench and denaturation conditions. In all cases, FITC labelled peptides were not observed. Other methods including different enzymes and techniques to help in the identification of the modification site were then trialled.

4.2.2.1 Pepsin digest

A pepsin digest (non-specific enzyme) was subsequently undertaken at a low pH, and from this extensive fragmentation was observed. The theoretical digest is shown in Figure 10.

P S S H H H H H H S S G L V P R G S H I D V E A L R Q L Y A A G E R D F S I V D L R G A V L E N I N L S G A I L H G A M L D E A N L Q Q A N L S R A D L S G A M L N G A D L R G A N L S A A D L S D A I L D N A I L E G A I L D E A V L N Q A N L A A A N L E Q A I L S H A N I R E A D L S E A N L E A A D L S G A D L A I A D L H Q A N L H Q A A L E R A N L T G A N L E D A N L E G T I L E G G N N N L A K

Pepsin (pH 2)

12 24 25 27 29 35 40 41 45 46 50 51 55 56 60 61 65 66 70 71 76 80 81 85 86 90 91 95 96 100 101 105 106 110 111 115 116 120 121 125 126 130 131 140 66 141 145 146 150 151 155 156 160 161 165 166 170 171 176 180 181 185 186 190 191 197 198

Figure 10. Theoretical pepsin digest. Pepsin is a digestive tract (stomach) enzyme that shows maximum activity at pH 2 and cleaves preferentially between hydrophobic residues.

The pepsin digests showed no C-terminal Lys modification and no N-terminal modification. It did however, show some N-terminal formyl product. Nearly complete coverage was observed for the protein (Figure 11).

Untitled
Associated Datafile: PC55A (50 - 1990 amu)
Pepsin/F/Y/W/L

Frag#	Res#	Sequence	Theor(Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
M1	1-13	(-)PSSHHHHHHSSGL(V)	1455.65	1456.66	728.83	486.22	364.92
M2	14-25	(L)VPRGSHIDVEAL(R)	1291.69	1292.70	646.85	431.57	323.93
M3	26-28	(L)RQL(Y)	415.25	416.26	208.63	139.43	104.82
M4	29-29	(L)Y(A)	181.07	182.08	91.54	61.37	46.28
M5	30-36	(Y)AAGERDF(S)	764.35	765.35	383.18	255.79	192.09
M6	37-41	(F)SIVDL(R)	545.31	546.31	273.66	182.78	137.33
M7	42-46	(L)RGAVL(E)	514.32	515.33	258.17	172.45	129.59
M8	47-51	(L)ENINL(S)	601.31	602.31	301.66	201.44	151.33
M9	52-56	(L)SGAIL(H)	459.27	460.28	230.64	154.10	115.83
M10	57-61	(L)HGABL(D)	522.27	523.27	262.14	175.10	131.57
M11	62-66	(L)DEANL(Q)	560.24	561.25	281.13	187.76	141.07
M12	67-71	(L)QQANL(S)	572.29	573.30	287.15	191.77	144.08
M13	72-76	(L)SRADL(S)	560.29	561.30	281.15	187.77	141.08
M14	77-81	(L)SGABL(N)	472.24	473.25	237.13	158.42	119.07
M15	82-86	(L)NGADL(R)	488.22	489.23	245.12	163.75	123.06
M16	87-91	(L)RGANL(S)	529.30	530.31	265.66	177.44	133.33
M17	92-96	(L)SAADL(S)	475.23	476.24	238.62	159.42	119.81
M18	97-101	(L)SDAIL(D)	517.27	518.28	259.65	173.43	130.33
M19	102-106	(L)DNAIL(E)	544.29	545.29	273.15	182.44	137.08
M20	107-111	(L)EGAIL(D)	501.28	502.29	251.65	168.10	126.33
M21	112-116	(L)DEAVL(N)	545.27	546.28	273.64	182.76	137.33
M22	117-121	(L)NQANL(A)	558.28	559.28	280.15	187.10	140.58
M23	122-126	(L)AAANL(E)	458.25	459.26	230.13	153.76	115.57
M24	127-131	(L)EQAIL(S)	572.32	573.32	287.17	191.78	144.09
M25	132-141	(L)SHANIREADL(S)	1124.56	1125.57	563.29	375.86	282.15
M26	142-146	(L)SEANL(E)	532.25	533.26	267.13	178.42	134.07
M27	147-151	(L)EAADL(S)	517.24	518.25	259.63	173.42	130.32
M28	152-156	(L)SGADL(A)	461.21	462.22	231.61	154.75	116.31
M29	157-161	(L)AIADL(H)	501.28	502.29	251.65	168.10	126.33
M30	162-166	(L)HQANL(H)	581.29	582.30	291.65	194.77	146.33
M31	167-171	(L)HQAAL(E)	538.29	539.29	270.15	180.44	135.58

Untitled
Associated Datafile: PC55A (50 - 1990 amu)
Pepsin/F/Y/W/L

Frag#	Res#	Sequence	Theor(Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]
M32	172-176	(L)ERANL(T)	601.32	602.33	301.67	201.45	151.34
M33	177-181	(L)TGANL(E)	474.24	475.25	238.13	159.09	119.57
M34	182-186	(L)EDANL(E)	560.24	561.25	281.13	187.76	141.07
M35	187-191	(L)EGTIL(E)	531.29	532.30	266.65	178.10	133.83
M36	192-198	(L)EGGNNNL(A)	716.31	717.32	359.16	239.78	180.09
M37	199-200	(L)AK(-)	217.14	218.15	109.58	73.39	55.29
M1-2	1-25	(-)PSSHHHHHHSSGLVPRG SHIDVEAL(R)	2729.33	2730.34	1365.67	910.78	683.34
M2-3	14-28	(L)VPRGSHIDVEALRQL(Y)	1688.93	1689.94	845.47	563.99	423.24
M3-4	26-29	(L)RQLY(A)	578.32	579.33	290.17	193.78	145.59
M4-5	29-36	(L)YAAGERDF(S)	927.41	928.42	464.71	310.14	232.86
M5-6	30-41	(Y)AAGERDFSIVDL(R)	1291.64	1292.65	646.83	431.55	323.92
M6-7	37-46	(F)SIVDLRGAVL(E)	1041.62	1042.63	521.82	348.21	261.41
M7-8	42-51	(L)RGAVLENINL(S)	1097.62	1098.63	549.82	366.88	275.41
M8-9	47-56	(L)ENINLSGAIL(H)	1042.57	1043.57	522.29	348.53	261.65
M9-10	52-61	(L)SGAILHGABL(D)	963.53	964.53	482.77	322.18	241.89
M10-11	57-66	(L)HGABLDEANL(Q)	1064.50	1065.51	533.26	355.84	267.13
M11-12	62-71	(L)DEANLQQANL(S)	1114.53	1115.53	558.27	372.52	279.64
M12-13	67-76	(L)QQANLSRADL(S)	1114.57	1115.58	558.29	372.53	279.65
M13-14	72-81	(L)SRADLSGABL(N)	1014.52	1015.53	508.27	339.18	254.64
M14-15	77-86	(L)SGABLNGADL(R)	942.45	943.46	472.23	315.16	236.62
M15-16	82-91	(L)NGADLRGANL(S)	999.51	1000.52	500.76	334.18	250.89
M16-17	87-96	(L)RGANLSAADL(S)	986.51	987.52	494.27	329.85	247.64
M17-18	92-101	(L)SAADLSDAIL(D)	974.49	975.50	488.25	325.84	244.63
M18-19	97-106	(L)SDAILDNAIL(E)	1043.55	1044.56	522.78	348.86	261.90
M19-20	102-111	(L)DNAILEGAIL(D)	1027.55	1028.56	514.79	343.53	257.90
M20-21	107-116	(L)EGAILDEAVL(N)	1028.54	1029.55	515.28	343.85	258.14
M21-22	112-121	(L)DEAVLNQANL(A)	1085.54	1086.54	543.78	362.85	272.39
M22-23	117-126	(L)NQANLAAANL(E)	998.51	999.52	500.27	333.85	250.64
M23-24	122-131	(L)AAANLEQAIL(S)	1012.56	1013.56	507.29	338.53	254.15
M24-25	127-141	(L)EQAILSHANIREADL(S)	1678.86	1679.87	840.44	560.63	420.72

Figure 11. Pepsin digests of the Np276 FITC system. Despite the large number of peptides observed, there was no evidence for the presence of a N-terminal Pro-2 FITC or the modification of the Lys-201 FITC.

4.2.2.2 UV detection of peptides, Q-Exactive-OrbiTrap.

Samples were subsequently run on the Q-Exactive, fitted with a UV detector along with uHPLC. It was reasoned that by validating the MS1 results with absorption at 495nm, the FITC containing peptides should be able to be easily identified. The direct injection of digest reactions (from AspN) into the Q-Exactive-uHPLC system yielded some peaks that absorbed in the 495nm region, yet there were no clearly identifiable peaks. This was also the case with an attempted thrombin cleavage & thermolysin digest. The data for the thermolysin digest is shown in Figure 12, it has been speculated that the peak at 502 Da could correspond to a FITC-Pro fragmentation product.

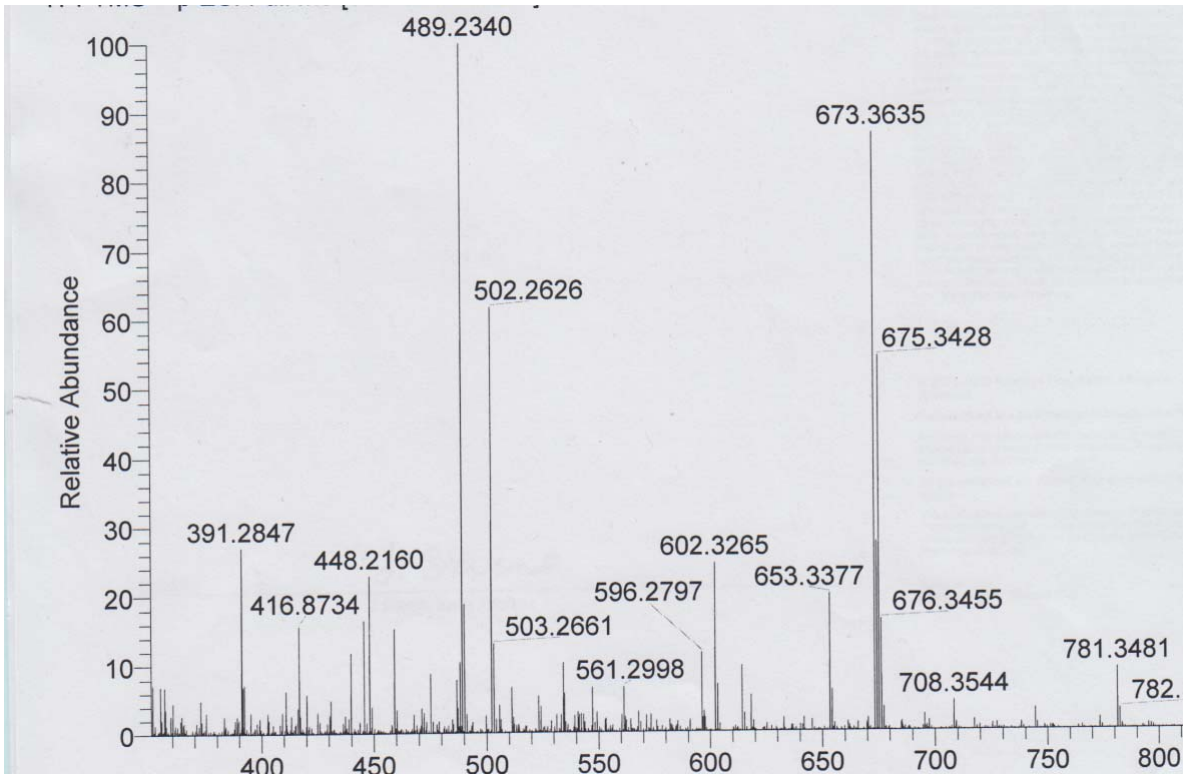


Figure 12. Mass spectra for peak displaying a 495nm absorbance on the UV. Ions don't correspond to expected products and thus don't provide evidence for Pro-2 FITC modification.

4.2.2.3 MALDI-TOF/TOF.

The protein sample was digested via an AspN digest before being run on a MALDI-TOF/TOF (different mode to the ESI methods). The results from this showed several of the peptides previously observed. However, no peptides corresponding to the N-terminal FITC were observed.

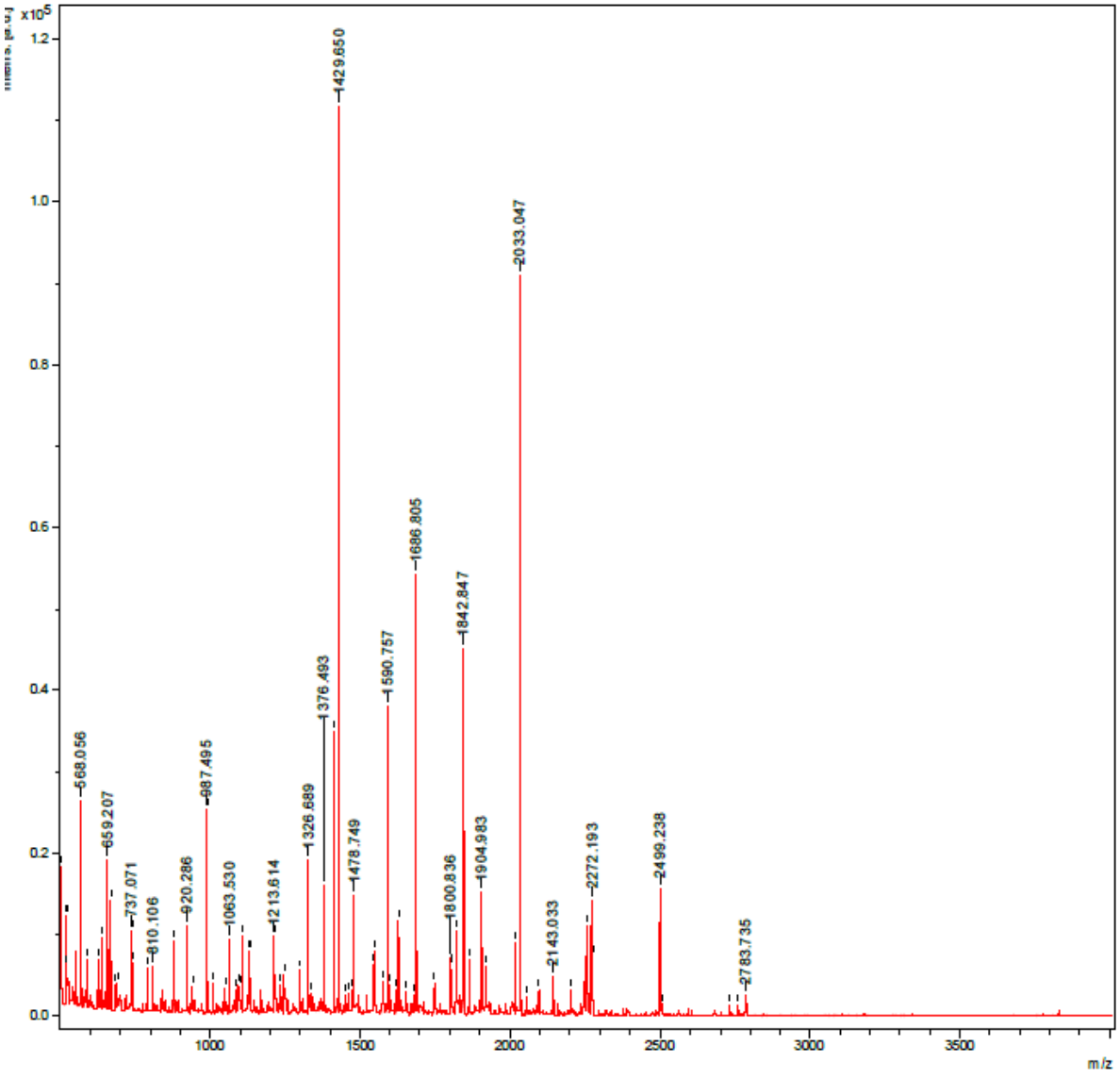


Figure 13. MALDI MS (MS1) of AspN digest. No peptides corresponding to a FITC modified N-terminal proline peptide were observed. The MALDI was trialled as it was hoped the use of a different ionization method may have allowed for identification of the desired Pro-2-FITC peptide.

4.2.2.4 Edman degradation of the FITC labelled protein

FITC forms a thiourea bond after reaction with the protein, whilst this bond is reported to be moderately stable under physiological conditions, it is expected to be unstable under conditions analogous to those commonly encountered in a bottom-up proteomics experiment (high temperature, pH). The FITC thiohydantoin product was not detected by MS.

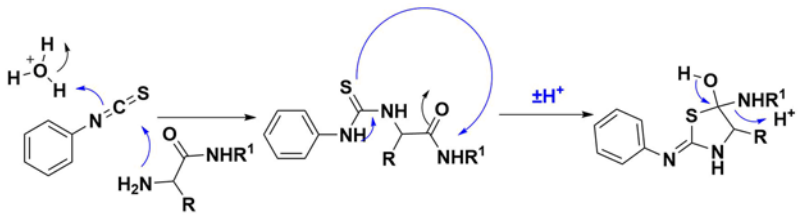


Figure 14. Mechanism for the Edman degradation. The N-terminal amine attacks the isothiocyanate in the presence of acid (usually trifluoroacetic) followed by a cyclisation.

PITC (phenyl isothiocyanate) was then investigated to observe whether it would undergo a similar transformation and if the peptide for this modification could be observed by MS/MS. The Np276 (5Å, G2P T201K, Met) mutant was reacted with PITC at ambient temperature, as with FITC mono modification of the protein ruler was seen indicating that this modification is specific to the isothiocyanate functionality. A subsequent AspN digest didn't yield any conclusive evidence for the site of modification, furthermore attempts to isolate the degraded products from an Edman degradation failed. In addition, a 100% FITC modified Np276 protein was reaction with PITC. No reaction was observed indicating that the modification was occurring at the same site.

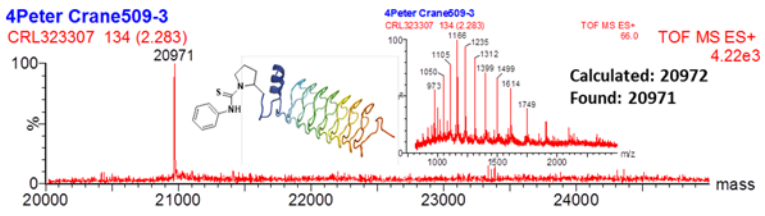


Figure 15. PITC reaction with 5Å M61-M81 G2P protein showing the formation of a single product, analogous to the result of the reaction with FITC.

Untitled
Associated Datafile: PC_4_PITC (50 - 1990 amu)
AspN:-/D

Frag#	Res#	Sequence	Theor(Bo)	[M+H]	[M+2H]	[M+3H]
D1	1-20	(-)PSSHHHHHHSSGLVPRGSHI(D)	2321.11	2322.11	1161.56	774.71
D2	21-34	(I)DVEALRQLYAAGER(D)	1589.82	1590.82	795.92	530.95
D3	35-39	(R)DFSIV(D)	579.29	580.30	290.65	194.10
D4	40-61	(V)DLRGAVLENINLSGAILHGAML(D)	2276.23	2277.24	1139.12	759.75
D5	62-74	(L)DEANLQQANLSRA(D)	1428.70	1429.70	715.36	477.24
D6	75-84	(A)DLSGAMLNGA(D)	947.44	948.45	474.73	316.82
D7	85-94	(A)DLRGANLSAA(D)	986.51	987.52	494.27	329.85
D8	95-97	(A)DLS(D)	333.15	334.16	167.58	112.06
D9	98-101	(S)DAIL(D)	430.24	431.25	216.13	144.42
D10	102-111	(L)DNAILEGAIL(D)	1027.55	1028.56	514.79	343.53
D11	112-139	(L)DEAVLNQANLAAANLEQAILSHANIREA(D)	2958.52	2959.52	1480.27	987.18
D12	140-149	(A)DLSEANLEAA(D)	1031.48	1032.48	516.75	344.83
D13	150-154	(A)DLSGA(D)	461.21	462.22	231.61	154.75
D14	155-159	(A)DLAIA(D)	501.28	502.29	251.65	168.10
D15	160-182	(A)DLHQANLHQAALERANLTGANLE(D)	2498.26	2499.27	1250.14	833.76
D16	183-200	(E)DANLEGTILEGGNNNLAK(-)	1841.91	1842.92	921.96	614.98
D1-2	1-34	(-)PSSHHHHHHSSGLVPRGSHIDVEALRQLYAAGER(D)	3895.24	3896.24	1948.63	1299.42
D2-3	21-39	(I)DVEALRQLYAAGERDFSIV(D)	2151.10	2152.10	1076.56	718.04
D3-4	35-61	(R)DFSIVDLRGAVLENINLSGAILHGAML(D)	2837.51	2838.52	1419.76	946.84
D4-5	40-74	(V)DLRGAVLENINLSGAILHGAMLDEANLQQANLSRA(D)	3689.17	3690.18	1845.59	1230.73
D5-6	62-84	(L)DEANLQQANLSRADLSGAMLNGA(D)	2358.12	2359.13	1180.07	787.05
D6-7	75-94	(A)DLSGAMLNGADLRGANLSAA(D)	1915.94	1916.95	958.98	639.66
D7-8	85-97	(A)DLRGANLSAADLS(D)	1301.66	1302.67	651.84	434.89
D8-9	95-101	(A)DLSDAIL(D)	745.39	746.39	373.70	249.47
D9-10	98-111	(S)DAILDNAILEGAIL(D)	1439.79	1440.79	720.90	480.94
D10-11	102-139	(L)DNAILEGAILDEAVLNQANLAAANLEQAILSHANIREA(D)	3970.41	3971.42	1986.21	1324.48
D11-12	112-149	(L)DEAVLNQANLAAANLEQAILSHANIREADLSEANLEAA(D)	3974.31	3975.32	1988.16	1325.78
D12-13	140-154	(A)DLSEANLEAADLSGA(D)	1474.68	1475.69	738.35	492.57
D13-14	150-159	(A)DLSGADLAIA(D)	944.48	945.49	473.25	315.83

Figure 16. PITC modification of the Np276 and subsequent MS/MS (AspN) digest. No evidence was found for the PITC modification of the Np276 protein.

4.2.2.5. Thrombin cleavage of the dye labelled protein

The Np276 protein contains an N-terminal HisTag, with the G2P mutation and a thrombin cleavage site. Thrombin as a widely used serine protease cleaved the HisTag fragment from the Np276 protein, resulting in a remaining fragment that showed none of the FITC modification. Thus the FITC tag can be determined to be bound at the non-PRP N-terminal domain most likely at the N-terminal proline. This cleavage reaction was conducted with the range of the Np276 FITC modified proteins and it was found to always occur off the PRP chain, and instead on the HisTag away from the sugar binding sites.

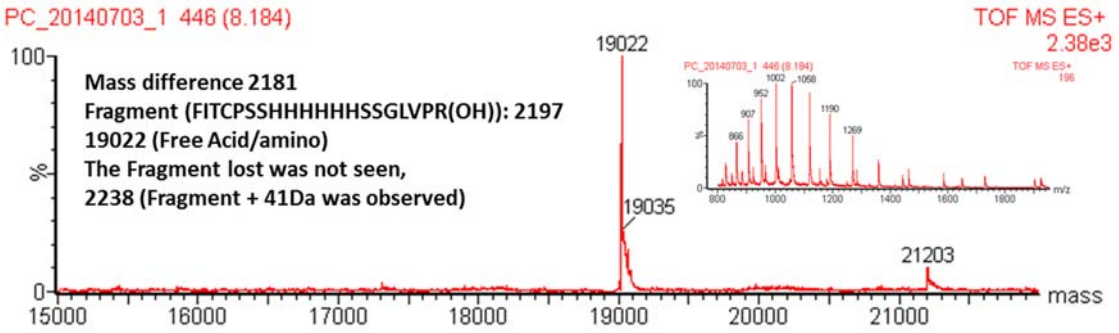


Figure 17. Thrombin digest of the Np276 FITC protein, showing the fragment with no FITC bound thus indicating that the FITC dye isn't bonded to the main protein chain.

Once FITC had been shown not to bind to the main PRP coils of Np276, this validated its use as a dye label and further supported a proposed [1+2] modification sequence.

4.2.3 CuAAC reactions and dual modification

The CuAAC reaction, using the catalytic system exploited previously with the group, was used to functionalise the proteins with glycans *via* a CuAAC reaction. This reaction proceeded cleanly and rapidly, giving the expected modified products in ~100% conversion as previously reported (Figure 18.). Furthermore, given that the surface of the Np276 scaffold is fully surface exposed, no regio-selectivity is observed in the reactions (as with SSβG) and complete modification is seen.[14] Gratifyingly it also appears from MS evidence that the CuAAC reaction is compatible with the presence of the FITC, as no thiourea copper complexation is observed nor degradation of the thiourea.

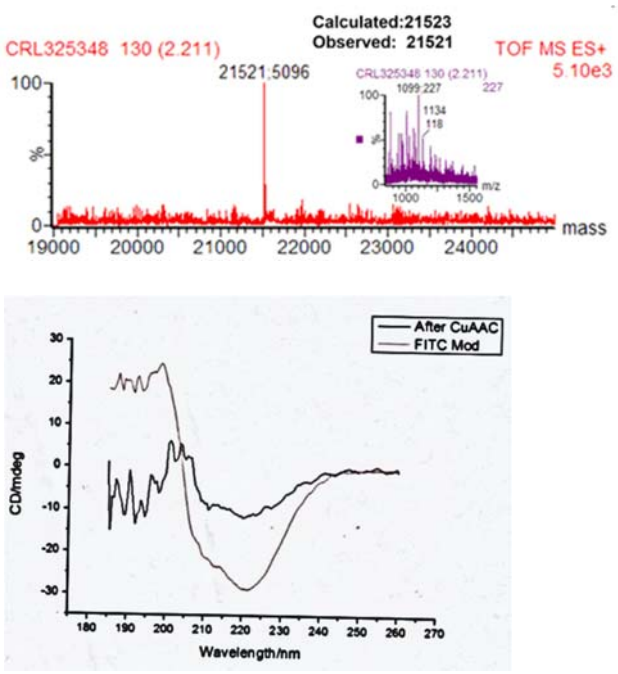


Figure 18. Dual modification of a single AHA ruler with a model GlcNAc sugar. This reaction showed the formation of a single product, the dual modified ruler protein (FITC and GlcNAc). CD spectra before and after the CuAAC reaction with O-propargyl GlcNAc. After the CuAAC reaction there was less secondary structure present.

Optimisation using a range of catalyst concentrations, sugar concentrations and temperatures found that the reaction proceeds well at lower temperatures (4°C) and at lower catalyst

concentrations (6 equivalents). The concentration of the sugar alkyne is however, important and should ideally be around 120eq/azide (experimental).

Despite the success of this chemistry in producing homogenous proteins, the structure of the protein appeared significantly perturbed as illustrated by the CD data (obtained after desalting *via* PD-10). The majority of examples of protein modification methods reported within the literature, do not contain biophysical (structural) evidence for the mildness of the reaction conditions on structure, aside from enzymatic activities. These reactions were then trialled with the SiaLacNAc trisaccharide previously synthesised, to create the dual modified constructs for an initial binding strength study (SiaLacNAc is charged so should repel the Np276 surface).

4.2.3.1. Dual modification: SiaLacNAC constructs.

Preliminary studies with the Np276 (no G2P) AHA incorporated proteins had demonstrated an efficient homogenous CuAAC reaction onto an N-terminally gluconylated protein that was un-modified with eye. A [2+1] dual modification with a CuAAC was thus conducted with G2P modified proteins giving di-sialylated protein. Attempted MS/MS also didn't elucidate either the site of or sugar coupling (sugar at azide due to orthogonality).The DiSiaLacNAc protein was subsequently treated with Alexa-488 SDP ester. Unfortunately, this didn't yield any of the desired product and instead the reactions appeared messy with large proportions of start material remaining. This approach ([2+1]) was not used for the FITC couplings as the FITC reaction could be conducted on scale first, with low-cost reagents and the resulting homogenously modified protein used in the more expensive (in terms of materials) CuAAC reaction. Proteins were thus modified by use of a FITC coupling, followed by a CuAAC reaction ([1+2]). These were validated with DiSiaLacNAc and FITC, giving homogenously modified [1+2] modified proteins, albeit it on a small scale (Figure 19).

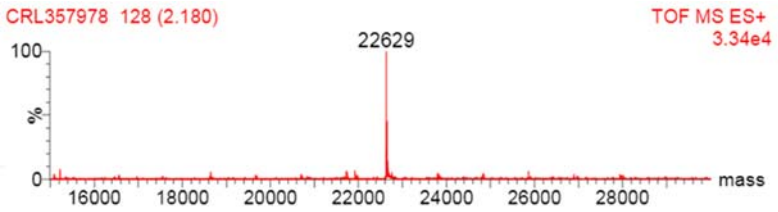
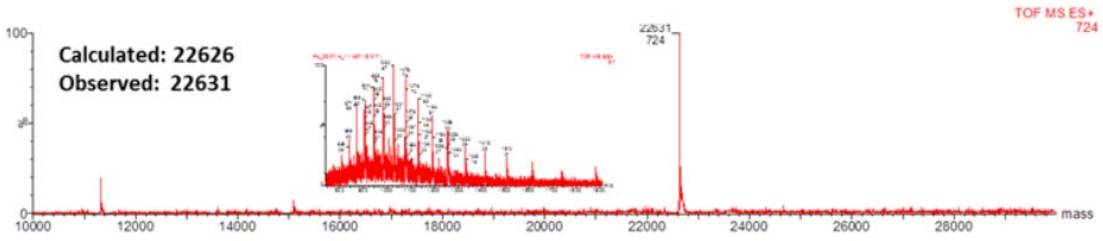


Figure 19. [1+2] DiSiaLacNAc, FITC modified 15Å protein. Calculated:22628 Observed:22629. The reaction was conducted on a small scale.

As many of these studies had been conducted on a small scale, or with protein concentration that would be too low for biological studies. Whilst, it was expected that the protein structure would be perturbed, it was essential to establish whether a dual (natural sugar) modified sugar construct would display any noticeable binding to the mSn receptor. In this respect, a result here would justify the creation of more or new dual (natural sugar) modified probes.

4.2.3.2. Dual modification: SiaLacNAC constructs.

The 30Å protein was also modified on larger scale and analysed by MS and CD. Whilst the protein again showed a relaxed structure in comparison to the pure expressed protein, there was evidence of some α/β-structure. Furthermore, the protein was homogeneously modified in a [1+2] fashion. Attempts to obtain MS/MS on this protein were unsuccessful. This protein was then used in a cell assays.



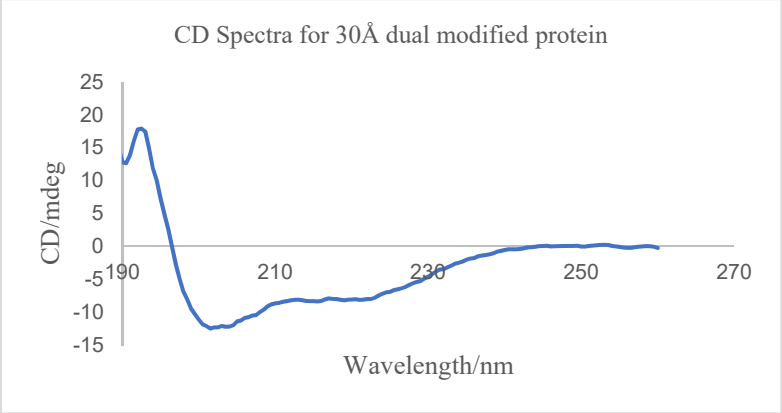


Figure 20. The dual modified protein used in cellular assays. Whilst the reaction formed a homogenous product, the CD showed an incorrect secondary structure (not the PRP structure).

4.3. Conclusions and discussion

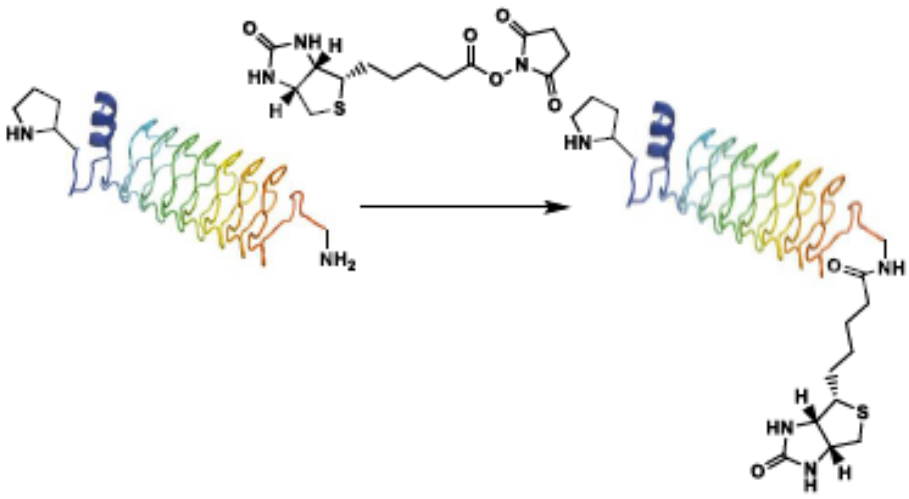
In conclusion, dual modified protein constructs were synthesised but due to on-going problems with the proteins' secondary structure/expression, a full ruler library wasn't generated. In order to synthesise these dual modified constructs, the reactivity of the N-terminal proline was utilised. Efficient and selective modification was observed when using isothiocyanate dyes, more so than with the Alexa-488 NHS esters (and more so than the biotin esters). Attempts to characterise the site of this site selective modification by MS/MS failed, despite a variety of techniques being utilised. Eventually by use of a thrombin cleavage, it was shown that the modification was isolated, selective to the cleavable portion of the N-terminal HisTag. There are no other ε amino groups in this sequence and furthermore many of the residues aren't observed to be particularly nucleophilic with respect to isothiocyanates. A further advantage of using FITC is its low cost in comparison to the proprietary Alexa-NHS dye family. This allowed for the dye modification to be undertaken on scale as a first step, with the CuAAC reaction being conducted in the same buffer sequentially. This represents a site selective and site specific [1+2] protein modification reaction sequence. In 2015, a [1+n] dual modification was reported by Budisa *et al* which utilised an N-terminal oxime coupling followed by the CuAAC reaction.[15] In the case of this study, there were several azide residues (n) and many were thus modified. The reports presented here represents one of the most controlled examples of a N-terminal/CuAAC modification sequence. This protein construct can be used to demonstrate binding affinity in the following cellular assays.

4.4 Experimental

LCT-ESI-TOF-MS

Protein liquid chromatography MS (LC-MS) was conducted using a Waters LCT classic (ESI-TOF-MS) connected to a Shimadzu 20 Series HPLC, using a Merck Chromolith C18 (50x2mm) column. Water (solvent A) and acetonitrile (solvent B) each with 0.1% formic acid were used as a mobile phase at a flow rate of 0.4ml/min. The gradient was set to 95% A (2 mins, isocratic), to 95% B after 8 mins. Then isocratic (95% B) for 3 minutes. The electrospray source parameters were: capillary voltage (3kV), cone voltage (250kV) with nitrogen as the nebuliser and solvation gases. Spectra were processed using the MaxEnt algorithm on Waters MassLynx v4.1. For concentration versus signal studies, the reader is directed to literature sources for Np276.[16]

NHS biotin reactions



To the protein solution pH 8.0, 50mM NaPi (P2P, M61-M81 Met T201K) (concentration and volume in the table below), was charged the appropriate quantity of the biotin NHS ester in DMSO. The reaction was then rotated at 45rpm for the required time. The reaction was then loaded onto a G-25 spin column and eluted (1000g, 2 min). LCT-ESI-TOF-MS was then run on resulting solution. 1-7 biotin additions in varying proportions were observed, often in a statistical distribution.

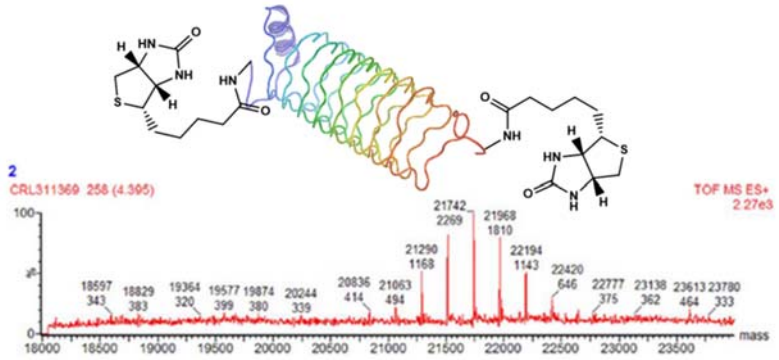
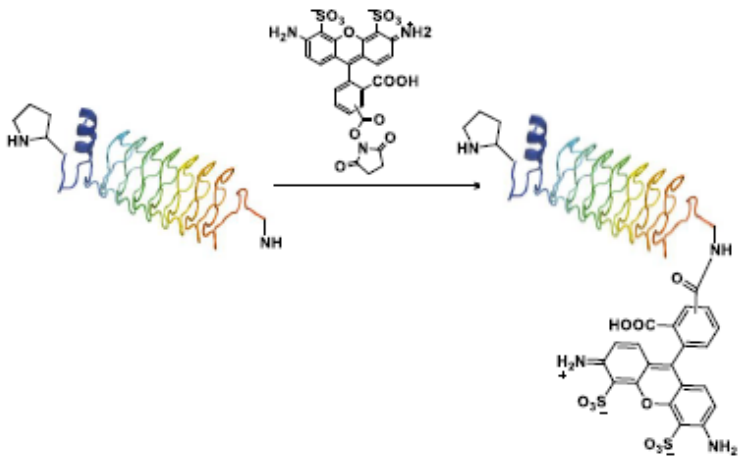


Figure E1. An example of a Np276 biotin NHS reaction showing a distribution of biotin additions. Only 2 biotins are shown on the image of the Np276 protein for clarity.

Table E1. A table showing all of the reaction conditions as well as the distribution of products from the reaction of the Np276 protein with NHS biotin.

Reaction	Protein (1.868mgml^{-1})/µL	Eq biotin NHS ester (1eq=0.306µL, 10mgml^{-1})	Time/ hours	Temp /°C	Yield mono /%	Yield di /%	Yield tri /%	Yield tetra /%	Yield >4 /%
1	100	5eq, 1 portion	6.5	25	8.1%	~15%	22%	19.6%	32%
2	100	1eq x 4 (per hr	6	25	5.2%	12.5%	~20%	24.3%	38%
3	100	10eq, 1 portion	5	25	-	-	-	-	100%
4	100	1eq x 5 (1 per hr)	5.5	25	-	-	-	-	100%
5	100	1eq, 1 portion	4.5	25	38%	37.4%	18%	4.6%	0%
6	100	5eq, 1 portion	4.5	4	5%	12.7%	24%	30.6%	31%
7	100	1eq, 1 portion	4.5	4	29%	36%	23.5%	11.4%	0%
8	100	1eq, 1 portion {1 mgml^{-1} reagent}	5	25	38.1%	29%	10.5%	0%	0%
9	100	1eq, slow addition	5	25	39%	29.7%	11.1%	0%	0%

NHS Alexa-488 reactions



To the protein solution pH8.0, 50mM NaPi (P2P, M61-M81 Met T201K) (concentration and volume in the table below), was charged the appropriate quantity of the Alexa-488 NHS ester

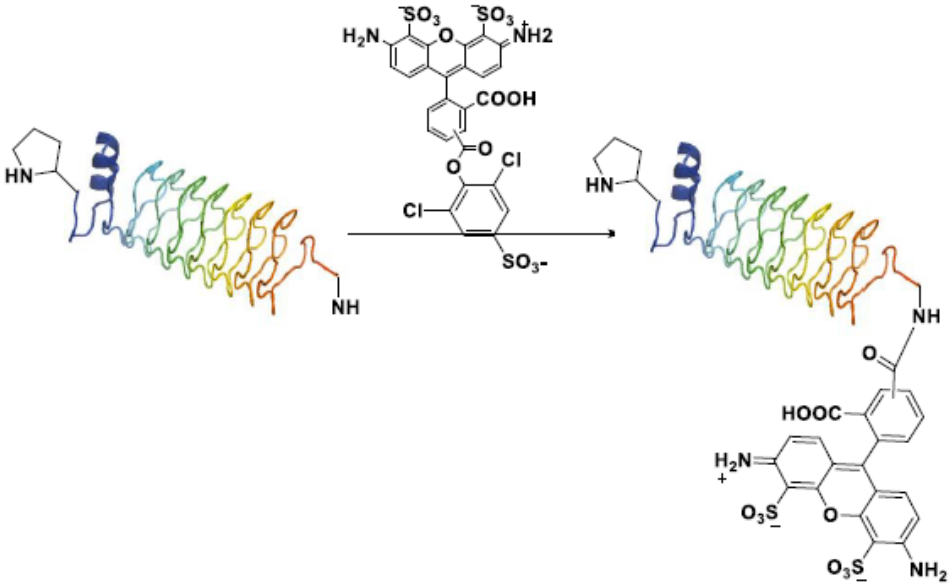
in DMSO (1mgml^{-1}). The reaction was then rotated at 45rpm for the required time. The reaction was then loaded onto a G-25 spin column and eluted (1000g, 2 min). LCT-ESI-TOF-MS was then run on resulting solution. 1-2 Alexa-488 additions in varying proportions were observed.

Table E2. A table showing all of the reaction conditions as well as the distribution of products from the reaction of the Np276 protein with Alexa-488 NHS ester.

Reaction	**Protein (1.868mg ml^{-1})/µL**	**Eq Alexa-488 NHS ester (1eq=5.76µL, 1mgml^{-1})**	**Time/ hours**	**Temp /°C**	**pH**	**Yield mono /%**
1	100	1	5	25	8	46%+ 10%Di
2	100	3	5	25	8	47% Mono+22%Di
3	100	1	0.5	25	8	12.5%
4	100	1	05	25	8	16%
5	100	2	0.25	25	8	19%
6	100	2	0.5	25	8	14%
7	100	3	0.25	25	8	25%
8	100	3	0.5	25	8	23%
9	100	3	1	25	8	27%
10	100	1	0.33	25	8	20%
11	100	1	1	25	8	8%
12	100	5 (1eq/15 min)	1.5	25	8	28%
13	100	2 (1eq/15min)	0.5	25	8	28%
14	100	6 (1eq, 10min)	1.2	25	8	35%
15	100	3	1	25	6.5	19%
16	100	3	1	25	7	20%
17	100	5 (1 per 10min)	1	25	6.5	37%
18	100	5 (1 per 10min)	1	25	7	20%
19	100	6 (1eq per 30	3	25	8	29%

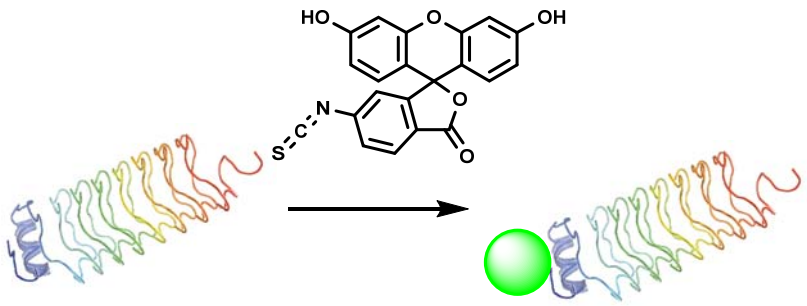
20	100	5eq (1 per 10min)	2	25	8	28%
21	100	4eq (1 per hour)	4	25	8	20%
22	100	5eq (1 per hour)	6	25	8	17%
23	100	1	16	4	8	12%
24	100	2	16	4	8	26%
25	100	3	16	4	8	31%
26	100	5	16	4	8	33% mono 4%di

SDP Alexa-488 reactions



To the Np276 protein solution pH8.0, 50mM NaPi (P2P, M61-M81 Met T201K)(20 uL, 1.868mgml^{-1}) , was charged the appropriate quantity of the Alexa-488 SDP ester (0.145uL = 1 eq in DMSO (10mgml^{-1}). The reaction was then rotated at 45rpm for the required time/temp. (3 hours). The reaction was then loaded onto a G-25 spin column and eluted (1000g, 2 min). LCT-ESI-TOF-MS was then run on resulting solution.

FITC reactions for modifications of Np276



To the protein solution (1eq, NaPi 50mM, pH8.0, 5% glycerol) was added FITC (100eq) in NaPi (50mM, pH8.0, 5% glycerol) at a concentration of 1mgml^{-1} or 5mgml^{-1}. In both cases FITC is sparingly soluble and must be vortexed vigorously before removing an aliquot for the reaction. FITC was added to the reaction mixture and in the dark the reaction rotated at ambient temperature on a lab rotisserie. The reaction was followed by LCT-ESI-TOF-MS and upon completion, passed through a pre-prepared size exclusion column (PD-10, G25 MidiTrap or G-25 MiniTrap) before being dialysed against NaPi 50mM pH8.0 buffer. If the reaction appears to be occurring slowly or has stalled, further equivalents of FITC can be added (100eq) portion-wise until completion. It is advised to avoid the use of DMSO/acetone in dissolving the FITC dye as this can lead to protein denaturation. After completion of the reaction the reaction is passed through a size exclusion column (PD-10, G-25 Midi/MiniTrap) and then dialysed against 4L of NaPi buffer. The product can be analysed by MS & CD spectroscopy.

Table E3. A table showing all of the reaction set-up, as well as the distribution of products for the reaction of the Np276 proteins with FITC.

Protein	Conc (mgml^{-1})	Vol/µL	nmol	100eq of FITC (1mgml^{-1})/µL	100eq of FITC (5mgml^{-1})/µL	Conversion
5Å	0.376	250	4.51	{95eq,}	-	100%, 3 addition 17 hour
10Å	0.6329	250	7.603	{95eq}	-	100% 3 addition 17 hour
15Å	0.374	250	4.49	174	-	100% (2 addition, 9 hr)
20Å M61	0.67	250	9.65	376	-	100% (2 addition, 9 hr)

20Å M41	0.776	380	13.08	{95eq}	-	100%, 3 addition 17 hour
25Å	0.537	900	23.2	-	180.67	100% (1 addition, 3 hr)
30Å	0.49	680	16	-	124.7	100% (1 addition, 3 hr)
Single AHA	1.42	600	40.818	-	317.88	100% (1 addition, 3 hr)

FITC reaction optimisation protocol

To 400µL of the protein solution in buffer (NaPi, 50mM, pH8.0 (or variable)) was added the FITC in NaPi and the Eppendorf (light safe Eppendorf) then placed on an end over end shaker (45rpm, ambient temperature (20°C or specified temperature). Reactions are monitored by ESI-MS with samples being quenched into 0.1M glycine buffer prior to analysis. The standard reaction conditions used for all trials are:

Table E4. Experimental set-up for all of the condition optimisation experiments conducted with FITC.

Protein	**Conc/mgml^{-1}**	**Volume/µL**	**Moles/nmol**	**Vol of FITC/µL (5mgml^{-1} suspension in NaPi)/10 equiv.**
5Å Np276	2.5	400	47.1	37.4 (10 equivalents)

Parameters tested: temperature

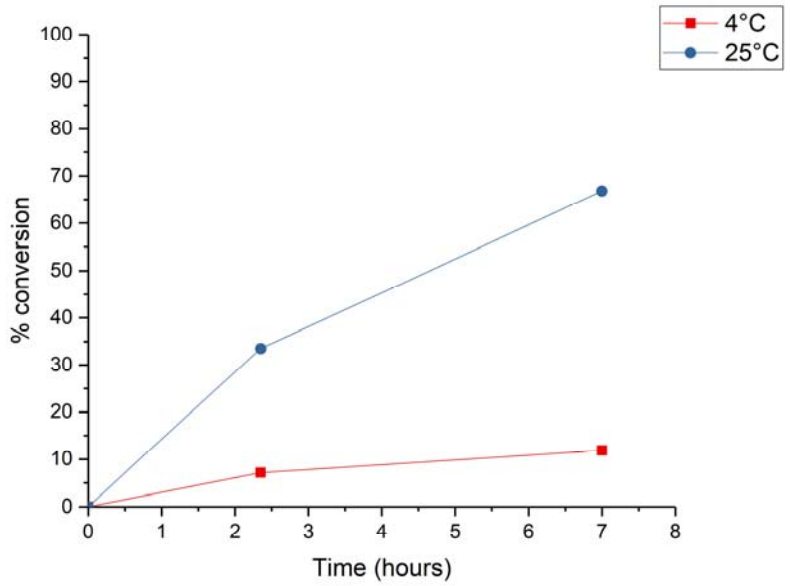


Figure E2. FITC optimisation reactions conducted at two different temperatures. The reaction gave higher conversions at higher temperatures.

Parameters tested: buffer

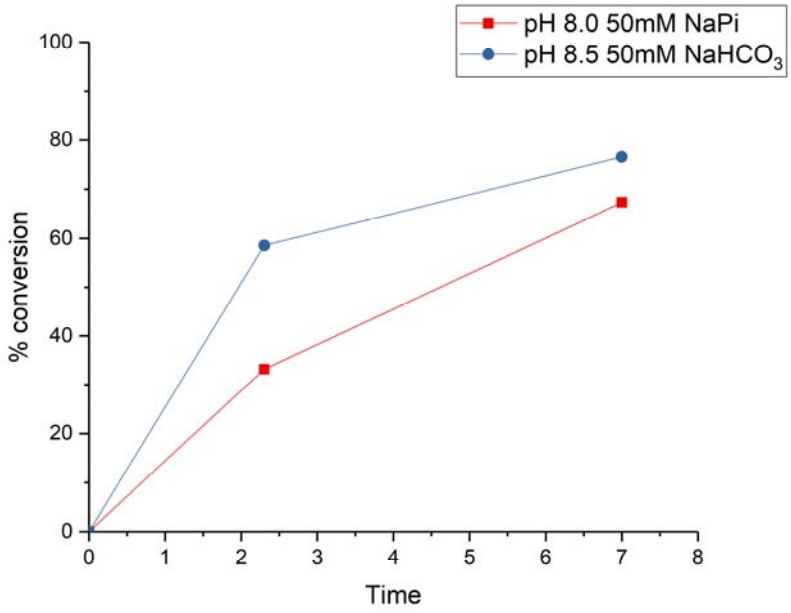


Figure E3. FITC optimisation reactions conducted with different buffers. $NaHCO_3$ was shown to be a better buffer for the reaction.

Standard MS/MS protocol

To 20µL of protein solution (0.3-0.6mgml^{-1}) was added 30µL ambic (50mM, pH 8.0), the reactions were heated to 95°C for 15 minutes and subsequently cooled to room temperature. Trypsin (0.1mgml^{-1}) was added to the reaction mixture in a ratio of 1:50 (eq/eq) of protein. Reactions involving AspN used a 1:200 ratio (0.02mgml^{-1}) of protease. The digestions were run overnight for 16 hours, at 37°C, with a rotation of 180rpm.

C18 ZipTip® purification protocol

To the reaction was added 1µL of formic acid, and the reactions vortexed briefly. 10µL of the reaction mixture was loaded onto a pre-equilibrated ZipTip® C18 tip, washed with wash buffer (H_2O + 0.1% FA) and finally eluted with 10µL of elution buffer (40-75% MeCN:60-25% H_2O, 0.1% formic acid). The sample was diluted with wash buffer and submitted to MS/MS analysis.

Sep-Pak lite C18 protocol

To the reaction was added 1µL of formic acid, and the reactions vortexed briefly. 50µL of the reaction mixture was loaded onto a pre-equilibrated Sep-Pak lite C18 column, washed with 10ml wash buffer (H_2O + 0.1% FA) and finally eluted with 2ml of elution buffer (40-75% MeCN:60-25% H_2O, 0.1% formic acid). After freeze drying overnight, the sample was diluted with wash buffer and submitted to MS/MS analysis.

Variation of variables in the MS/MS digests, for FITC identification.

Several trials were run comparing the different purification methods, it was found that the results obtained were similar in all cases. Likewise, other variables including: the elution buffers (% acetonitrile), time of digest, urea denaturation and heat were also investigated; this appears to have little effect upon the recovery of peptides as no N-terminal peptide was seen in these conditions.

Pepsin digest protocol

To 20µL of protein solution (0.3-0.6mgml^{-1}) was added pepsin (0.2µg, pH 2.0) along with 15µL H_2O & 15µL 0.1M HCl. pH~2, checked by pH paper. The reaction was heated at 37°C, 180rpm for approximately 3 hours. The sample was subsequently purified by a ZipTip® C18 tip and submitted for MS/MS analysis. Good coverage of the protein was obtained but there was no evidence of N-terminal peptide.

Q-exactive

The Q-exactive mass spectrometer fitted with a UV detector (495nm) was used to relate peptides observed after the digest to their UV absorbance. For this, protein solutions post protease digestion were simply injected rather than being run on the MS/MS.

MALDI-MS/MS (TOF/TOF)

The sample from the AspN digest, was provided crude to Dr. Holger Kramer (DPAG) and run on the MALDI TOF/TOF with LIFT. From this the expected peptides were not observed, however other peptides also seen on Synapt CID MS/MS were observed here as well.

Thermolysin protocol

To 30µL of protein solution (0.3-0.6mgml^{-1}) was added thrombin cleavage buffer (0.6µL, 500 mM tris-HCl, pH 8.0, 100 mM $CaCl_2$) and thermolysin (1µL, 1mgml^{-1}) the reactions were heated to 75°C for 60 minutes and subsequently cooled to room temperature. The sample was submitted neat for MS/MS analysis.

Edman degradation reactions

To 60µL of the starting protein, 10Å FITC modified (0.8nmol, 1eq), was added 5µL of PITC (41800eq). Along with 20µL 1:1:1 (water: MeOH: piperidine), after shaking for 1 hour, a MS was run which indicated that no reaction had occurred (FITC blocking it). Henceforth reaction occurring at the same site as FITC.

To 100µL of the starting protein, 5Å M61-M81 G2P unmodified (3.7nmol, 1eq) was added 8.8µL of PITC (200eq) and DMSO (10µL) with piperidine (5µL). After shaking overnight an MS was run which indicated that complete conversion had occurred. The sample was then submitted for MS/MS (AspN).

Thrombin protocol to identify the cleavage site

To 90µL of protein solution (0.3-0.6mgml^{-1}) was added thrombin cleavage buffer (9µL, 500 mM tris-HCl, pH 8.0, 100mM $CaCl_2$) and thrombin (9µL, 1mgml^{-1}). The reactions were shaken overnight at 4°C for 16 hours, or at room temperature for between 1-4 hours. The samples were then spun down in a bench top centrifuge and passed through a filter vial to remove the resin. They were then run directly on the LCT-ESI-TOF-MS.

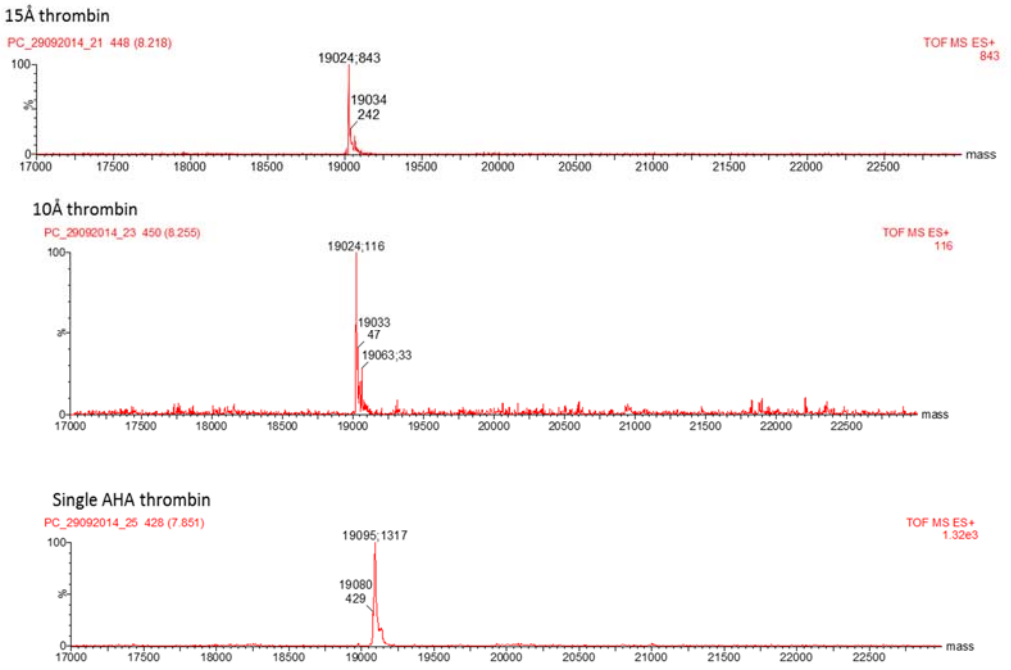
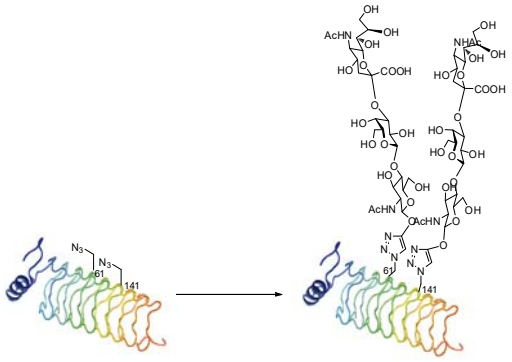


Figure E4 Examples of thrombin cleavage reactions.These illustrated that the FITC modifications were occurring on the N-terminus of the proteins and not on the main PRP chain.

Dual modification reactions

Protein modification on Np276 20Å ruler (gluconylated protein) approach.



To 500µL of protein solution (20Å, M21I, K25/94/123A D141M, T201K 0.5mg, 24nmol) in NaPi pH8.0 50mM buffer was added 2mg of SiaLacNAc (0.002mmol) dissolved into 20µL of water. Separately to an Eppendorf tube was charged 9.5mg of the click ligand and 75uL of MeCN. Separately CuBr (5mg) was dissolved in MeCN (500µL). 63µL of the ligand solution and 163µL of the CuBr solution were mixed immediately before use and 60µL of this solution was added to the protein-sugar solution. After rotating, at ambient temperature for 4 hours, the reaction was stopped by pipetting the reaction mixture onto a pre-prepped PD-10 MiniTrap (size exclusion column). After elution of the protein by centrifugation (1000gx2min) the protein was concentrated (VivaSpin, 10K MWCO) and analysed by ESI-TOF-MS mass spec and SDS-PAGE. MS indicated the reaction had been successful.

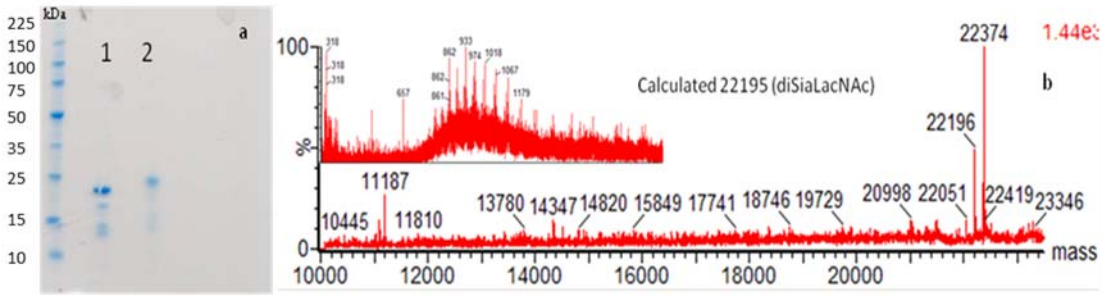
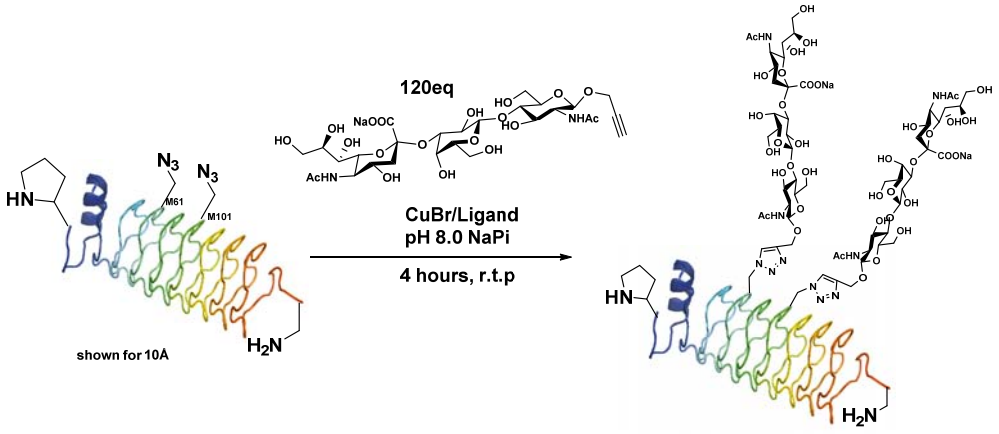


Figure E5 A) SDS-PAGE, 1) unmodified protein 2) modified protein b) LCT-ESI-TOF-MS analysis showing the formation of the modified glycosylated protein along with an intact gluconylation modification.

Protein CuAAC reactions for the preparation of [2+1] modified proteins.

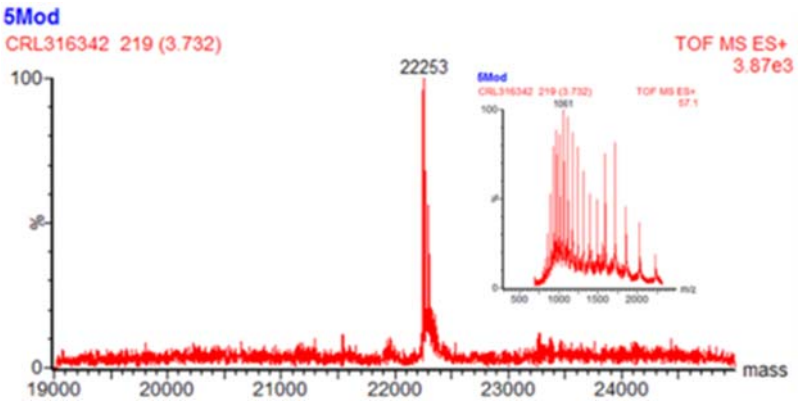


To the protein solution (pH 8.0, 50mM NaPi) was added in ddH_2O the SiaLacNAc sugar (120 eq). The protein solution was then charged with argon. In a separate Eppendorf was charged CuBr (5mg, 500µL MeCN) and separately the click ligand (9.5mg, 75µL). The active CuAAC reaction complex was prepared by mixing the CuBr solution (163µL) and the ligand solution (63µL) together. This was then used directly in the CuAAC reactions. All CuAAC reactions were rotated at 45rpm for 4 hours (25°C) before being loaded onto a pre-equilibrated MIDI-Trap spin column. The protein was eluted (1000g, 2 min) and analysed by ESI-TOF-MS analysis & MS/MS.

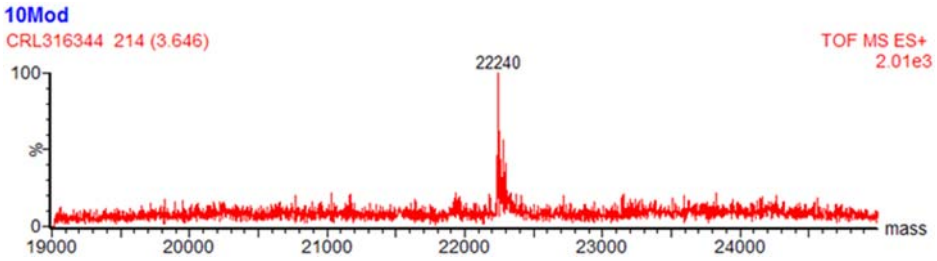
Table E5. Reaction set-up for the CuAAC reaction of the Np276 proteins with SiaLacNAc alkyne

Np276 G2P ruler protein	Conc/mgml^{-1}	Volume/µL	CuBr/Ligand Complex /µL	Conversion
5	0.579	150	30	100%
10	0.614	150	21	100%
15	0.235	400	21.42	100%
20	0.915	150	31.32	100%

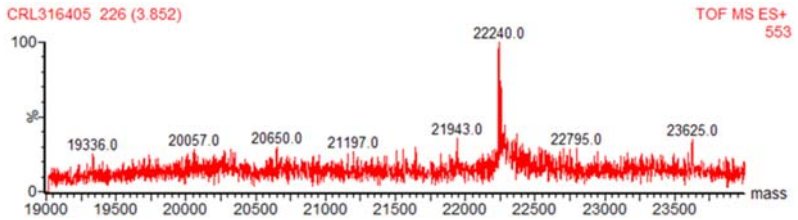
5Å



10Å



15Å



20Å

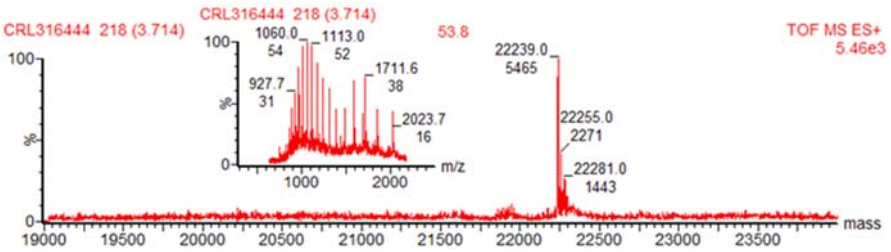


Figure E6 Towards a [2+1] modification approach table of reaction conditions for the CuAAC reactions. 5Å calculated:22250 observed:22253 10Å calculated: 22237 observed:22240 15Å calculated:22236 observed:22240 20Å calculated: 22235 observed: 22239

MS/MS (protocol with trypsin for SiaLacNAc modified proteins)

To 20µL of protein solution (0.3-0.6 $mgml^{-1}$) was added 30µL ambic (50mM, pH8.0), the reactions were heated to 95°C for 15 minutes and subsequently cooled to room temperature. Trypsin (0.1$mgml^{-1}$) was added to the reaction mixture in a ratio of 1:50 (eq/eq) of protein. To the reaction was added 1µL of formic acid, and the reactions vortexed briefly. 10µL of the reaction mixture was loaded onto a pre-equilibrated ZipTip® C18 tip, washed with wash buffer (H_2O + 0.1% FA) and finally eluted with 10µL of elution buffer (40-75% MeCN:60-25% H_2O, 0.1% formic acid). The sample was diluted with wash buffer and submitted to MS/MS analysis.

15Å MS/MS

MPSSHHHHHHSSGLVPRGSHIDVEALRQLYAAGERDFSIV
Associated Datafile: PC_PKCN6 (50 - 1990 amu)
Trypsin:/K-\P /R-\P

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]	[M+5H]
▪T1	1-16	(-)PSSHHHHHHSSGLVPR(G)	1807.87	1808.88	904.94	603.63	452.98	362.58
T2	17-26	(R)GSHIDVEALR(Q)	1095.57	1096.58	548.79	366.20	274.90	220.12
T3	27-34	(R)QLYAAGER(D)	906.46	907.46	454.24	303.16	227.62	182.30
T4	35-42	(R)DFSIVDLR(G)	963.50	964.51	482.76	322.18	241.88	193.71
T5	43-73	(R)GAVLENINLSGAILHGA JLDEANLQQANLSR(A)	3942.99	3944.00	1972.50	1315.34	986.76	789.61
T6	74-87	(R)ADLSGATLNGADLR(G)	1372.69	1373.70	687.36	458.57	344.18	275.55
T7	88-137	(R)GANLSAADLSDAILDNA ILEGAILDEAVLNQAJLAAA NLEQAILSHANIR(E)	5809.09	5810.10	2905.55	1937.37	1453.28	1162.83
T8	138-173	(R)EADLSEANLEAADLSGA DLAIADLHQANLHQAALER(A)	3757.04	3758.05	1879.53	1253.36	940.27	752.42
T9	174-200	(R)ANLTGANLEDANLEGTI LEGGNNNLAK(-)	2725.35	2726.36	1363.68	909.46	682.35	546.08

10Å MS/MS

MPSSHHHHHHSSGLVPRGSHIDVEALRQLYAAGERDFSIV
Associated Datafile: PC_PKCN7 (50 - 1990 amu)
Trypsin:/K-\P /R-\P

Frag#	Res#	Sequence	Theor (Bo)	[M+H]	[M+2H]	[M+3H]	[M+4H]	[M+5H]
T1	1-16	(-)PSSHHHHHHSSGLVPR(G)	1807.87	1808.88	904.94	603.63	452.98	362.58
T2	17-26	(R)GSHIDVEALR(Q)	1095.57	1096.58	548.79	366.20	274.90	220.12
T3	27-34	(R)QLYAAGER(D)	906.46	907.46	454.24	303.16	227.62	182.30
T4	35-42	(R)DFSIVDLR(G)	963.50	964.51	482.76	322.18	241.88	193.71
T5	43-73	(R)GAVLENINLSGAILHGA JLDEANLQQANLSR(A)	3942.99	3944.00	1972.50	1315.34	986.76	789.61
T6	74-87	(R)ADLSGATLNGADLR(G)	1372.69	1373.70	687.36	458.57	344.18	275.55
▪T7	88-137	(R)GANLSAADLSDAJLDNA ILEGAILDEAVLNQANLAAA NLEQAILSHANIR(E)	5810.04	5811.05	2906.03	1937.69	1453.52	1163.02
T8	138-173	(R)EADLSEANLEAADLSGA DLAIADLHQANLHQAALER(A)	3757.04	3758.05	1879.53	1253.36	940.27	752.42
T9	174-200	(R)ANLTGANLEDANLEGTI LEGGNNNLAK(-)	2725.35	2726.36	1363.68	909.46	682.35	546.08

Figure E7 MS/MS data for the DiSiaLacNAc modifed 15 and 10Å modified proteins, the 5 and 20Å were unsuccessful yielding no data.

Protein CuAAC reactions CD studies.

The study was conducted using the appropriate Np276 protein as stated above. Initially a blank of pH 8.0 50mM NaPi was run (200µL), following this the protein was loaded into the cuvette and at 25°C the CD spectra run. The control software automatically removes the baseline. The standard CD protocol is given in Chapter 2.

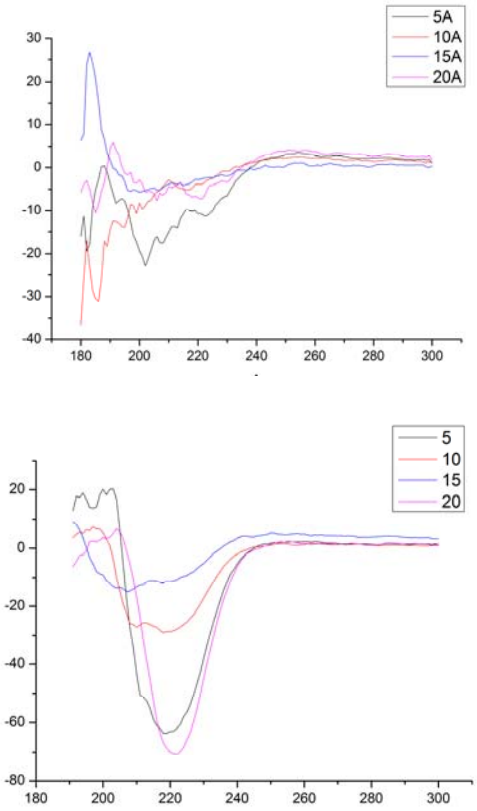


Figure E8. CD spectra for the modified and unmodified proteins. The CuAAC modified proteins (top) showed little structure, however the proteins prior to modification (bottom) also shown a variety of structures. The CuAAC reaction appeared to perturb the secondary structure of the proteins even further. The CD of the CuAAC modified proteins were all run at 0.1mgml^{-1} and for the unmodified at the concentrations in Table E5.

Alexa-488 SDP bio-conjugation reactions onto glycan modified protein.

To the protein solution pH 8.0, 50mM NaPi (5Å SiaLacNAC system in previous experiment), was charged the appropriate quantity of the Alexa-488 SDP ester (5 eq in DMSO (10mgml^{-1}). The reaction was then rotated at 45rpm for the required time/temp. (3 hours, rtp). The reaction was then loaded onto a G-25 Spin column and eluted (1000g, 2 min). LCT-ESI-TOF-MS was then run on resulting solution.

Table E6. The reaction set-up concentrations and volumes used for the reaction.

DiSiaLacNAc modified protein	Protein conc./mgml^{-1}	Volume/µL	Vol dye/µL
5Å	0.45	100	0.83

5Å SDP Modification

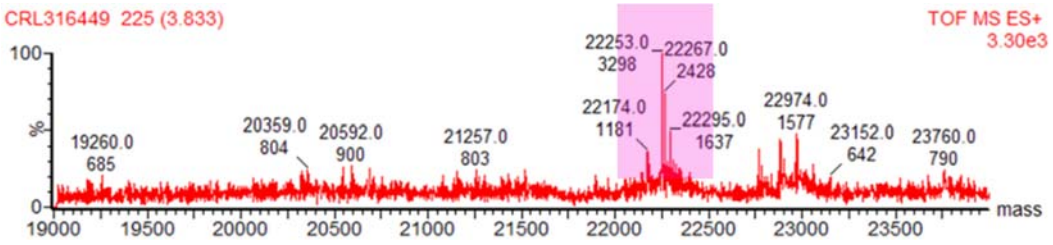


Figure E9. Alexa-488 SDP reactions with the 5Å diSiaLacNAc modified proteins gave weak MS data. The start material is highlighted in pink, with no clear product peak visible.

Protein CuAAC reactions for the preparation of [1+2] modified proteins with FITC

Optimisation Reactions with O-propargyl GlcNAc (no FITC)

The copper(I) catalyst was prepared *in situ* by reacting CuBr (163µL, 10mgml^{-1} in MeCN) with the tris-triazole click ligand (63µL, 126mgml^{-1} in MeCN) (Ethyl-2,2',2''-(4,4',4''-(nitrilotris(methylene))tris(1H-1,2,3-triazole-4,1-diyl))triacetate). The catalyst was then used immediately. To the start material, Np276 5A AHA61-81 G2P (25µL, 2.5mgml^{-1}) in NaPi (50mM, pH8.0) was added O-propargyl GlcNAc (in ddH_2O) and the CuAAC catalyst solution added. After shaking (45rpm) at 22°C for 0.3 hours, the reaction was submitted directly for MS analysis.

	Equivalents Sugar	Concentration Sugar (mg/ml)	Equiv. CuBr	Temp/°C	Vol Sugar (uL)	Vol CuBr.Ligand (uL)	t=20mins (% conversion from MS)
#1	120	5	60	25	51.33311321	9.850875391	100%
#2	120	100	60	25	2.56665566	9.850875391	87%
#3	60	5	60	25	25.6665566	9.850875391	85%
#4	60	10	60	25	12.8332783	9.850875391	81%
#5	60	100	60	25	1.28332783	9.850875391	76%
#6	120	10	6	25	25.6665566	0.985087539	100%
#7	120	10	12	25	25.6665566	1.970175078	74%
#8	120	10	60	25	25.6665566	9.850875391	85%
#9	120	10	60	4	25.6665566	9.850875391	100%

Figure E10. A table of reaction optimisations using the O-propargyl GlcNAc, lower sugars concentrations were shown to result in lower conversions.

CuAAC reaction with O-propargyl GlcNAc on FITC modified protein.

The copper(I) catalyst was prepared *in situ* by reacting CuBr (163µL, 10mgml^{-1} in MeCN) with the tris-triazole click ligand (63µL, 126mgml^{-1} in MeCN). The catalyst was then used immediately. To the start material, Np276 single AHA61 G2P FITC (100µL, 1.42mgml^{-1}) in NaPi (50mM, 5% glycerol, pH8.0) was added O-propargyl GlcNAc (0.207mg in 2µL ddH_2O). The protein solution was then diluted up to 300µL with NaPi and the CuAAC catalyst solution added (32.35µL). After shaking (45rpm) at 22°C for 4 hours, the reaction was run through a

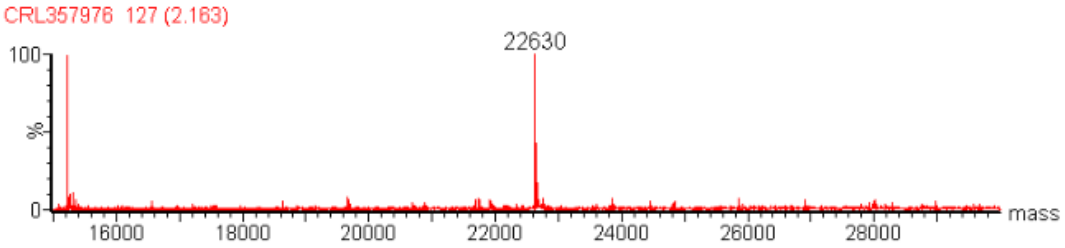
prepared PD-10 MiniTrap and then submitted for MS analysis and CD analysis (diluted to 0.5 $mgml^{-1}$).

Model CuAAC reaction protocol with explanation for reproducible application

The copper(I) catalyst was prepared in situ by reacting CuBr (163µL, $10mgml^{-1}$ in MeCN) with the tris-triazole click ligand (63µL, $126mgml^{-1}$ in MeCN). The catalyst was then used immediately. To the start material, Np276 in NaPi (50mM, pH 8.0) was added 120 eq of the SiaLacNAc ($10mgml^{-1}$) in water. The CuAAC catalyst solution was added in one portion and after shaking (45rpm) at 22°C for 2 hours, the reaction was run through a prepared PD-10 MiniTrap and then submitted for MS analysis. Before undertaking these reactions, the protein solutions were adjusted to the same concentration (with pH 8.0 50mM NaPi) and the specific volumes used shown in the table below.

Table E7. The reaction set-up concentrations and volumes used for the reaction.

FITC modified ruler protein	Protein Conc.	Volume	Volume SiaLacNAc (120 eq)/uL	CuBr/Ligand Complex /uL	Conversion
10Å	$1.25mgml^{-1}$	100uL	24	1.74	100%
15Å	$1.25mgml^{-1}$	100uL	24	1.74	100%
30Å	$1.25mgml^{-1}$	100uL	24	1.74	100%



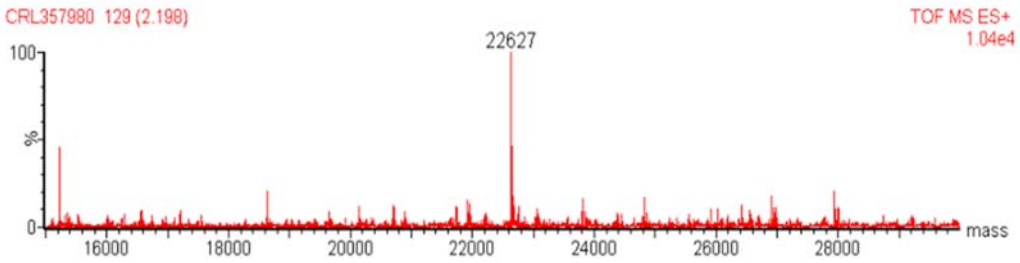


Figure E11. [1+2] modified proteins reaction conditions and results. Top (10Å: Calculated: 22626 Observed: 22630) and bottom (30Å: Calculated: 22626 Observed: 22627).

Notes to the method

The Copper should be at least 99.999% purity and fully in the Cu(I) state for the reaction to proceed efficiently. This catalyst-ligand solution should be colourless to a slight-off green when fresh.

Figure E12. Good quality Cu(I) is shown on the left as a light green to yellow solution. Poor quality Cu(I) insufficient for the reaction is shown on the right (dark green).

SiaLacNAc can be dissolved into a NaPi buffer, but given the difficulty in weighing this material; ddH_2O is perhaps better. The use of a pH8.0 buffer is important (basic pH), whilst chloride containing buffers should be avoided as they retard the reaction. PBS is the buffer to be used in cell assays, it also quenches the reactive copper species. It is essential to remove the CuL complex from the reaction before analysis.

Large-scale CuAAC reaction on FITC modified protein for use in cellular experiments.

The copper(I) catalyst was prepared in situ by reacting CuBr (163µL, 10mgml^{-1} in MeCN) with the tris-triazole click ligand (63µL, 126mgml^{-1} in MeCN). The catalyst was then used immediately. To the start material, Np276 AHA41-AHA161 G2P FITC (200µL, 1.32mgml^{-1}) in NaPi (50mM, 5% glycerol, pH8.0) was added SiaLacNAc (212µL, 10mgml^{-1}). The CuAAC catalyst solution was added (14.8µL) and after shaking (45rpm) at 22°C for 2 hours, the reaction was run through a prepared PD-10 MiniTrap and then submitted for MS analysis. The complete reaction was then dialysed thrice against PBS ready for testing. Attempts to obtain MS/MS were unsuccessful. CD was run at a concentration of 0.42mgml^{-1}.

4.5 Bibliography

(1) Koniev, O.; Wagner, A. *Chem. Soc. Rev.* **2015**, *44* (15), 5495–5551.

(2) Meldal, M.; Tornoe, C. W. *Chem. Rev.* **2008**, *108* (8), 2952–3015.

(3) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. *J. Am. Chem. Soc.* **2009**, *131* (45), 16346–16347.

(4) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2013**, *49* (27), 2747–2749.

(5) Li, J.; Lin, S.; Wang, J.; Jia, S.; Yang, M.; Hao, Z.; Zhang, X.; Chen, P. R. *J. Am. Chem. Soc.* **2013**, *135* (19), 7330–7338.

(6) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G. *Acc. Chem. Res.* **2011**, *44* (9), 730–741.

(7) Brown, K. K.; Blaikie, F. H.; Smith, R. a J.; Tyndall, J. D. a; Lue, H.; Bernhagen, J.; Winterbourn, C. C.; Hampton, M. B. *J. Biol. Chem.* **2009**, *284* (47), 32425–32433.

(8) Senter, P. D.; Al-Abed, Y.; Metz, C. N.; Benigni, F.; Mitchell, R. a; Chesney, J.; Han, J.; Gartner, C. G.; Nelson, S. D.; Todaro, G. J.; Bucala, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (1), 144–149.

(9) Swope, M.; Sun, H. W.; Blake, P. R.; Lolis, E. *EMBO J.* **1998**, *17* (13), 3534–3541.

(10) Ouertatani-Sakouhi, H.; El-Turk, F.; Fauvet, B.; Roger, T.; Le Roy, D.; Karpinar, D. P.; Leng, L.; Bucala, R.; Zweckstetter, M.; Calandra, T.; Lashuel, H. a. *Biochemistry* **2009**, *48* (41), 9858–9870.

(11) Stamps, S. L.; Fitzgerald, M. C.; Whitman, C. P. *Biochemistry* **1998**, *37* (28), 10195–10202.

(12) List, B.; Lerner, R. a; Barbas III, C. F. *J. Am. Chem. Soc.* **2000**, *122* (Copyright (C) 2010 American Chemical Society (ACS). All Rights Reserved.), 2395–2396.

(13) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, No. 136, 9572–9579.

(14) van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature* **2007**, *446* (7139), 1105–1109.

(15) Mühlberg, M.; Hoesl, M. G.; Kuehne, C.; Dernedde, J.; Budisa, N.; Hackenberger, C. P. R. *Beilstein J. Org. Chem* **2015**, *11*, 784–791.

(16) Vallée, R.; Schombs, M. W.; Balaban, Z.; Colyer, J.; Davis, B. G. *Chem. Commun.* **2016**, *52* (14), 2–5.

Chapter 5

The preliminary biological evaluation of the molecular ruler concept

Contents Page

5.1 Introduction and Aim

The aim of this chapter was to assess whether the 'molecular ruler' construct would bind to a model cell line, in this case, a CHO line over expressing mSn on the cell surface. Sialoadhesin (Sn) also known as CD169 or Siglec-1, shows a binding preference for α2-3 linked glycoconjugates (e.g. α2-3-SiaLacNAc) and has been studied by extensively by Crocker *et al*.[1–3] Murine Sn (mSn) is an ideal Siglec to trial the molecular ruler concept with as it doesn't suffer from the *cis* interactions that mask the Siglec to *trans* binding.[4,5] The need to obtain a high avidity interaction to 'out compete' the cis interactions (e.g. with CD22) has led to the wide adoption of high avidity but imprecise probes such as polymers or liposomes, in the study of Siglecs.[6,7] The Siglec family of receptors and multivalent methods to study them were previously outlined in in Chapter 1.

Given the challenges outlined in the previous chapter on protein modification, one dual modified protein (30Å) along with a FITC modified control of the same protein was chosen, to assess whether any binding was observed. Reports in the literature also normally utilised NaBu to increase cellular expression of the receptor and thus the response. Concerned about how this may perturb cell viability, the use of NaBu was initially investigated to determine suitability before conducting a binding assay.

5.2 Results and Discussion

5.2.1 Growth and maintenance of murine Sn d1-17 CHO cells for cellular experiments.

The mSn d1-17 CHO cells were probed with the commercially available 3D6.112 anti-mCD169 rat mAb, FITC conjugate which targets domain 1 of the Sn receptor. The other commercially available Ab, SER-4, targets 2 and 3 domain of the receptor.

The cells were probed with several dilutions of the antibody, ranging from the manufacturer's recommended dilution (1x, 0.01mgml^{-1}) in PBS through to an 8X (0.00125mgml^{-1}) dilution. After titrating the Ab, it was confirmed that the CHO cells were expressing the active mSn receptor on their surface (Figure 1).

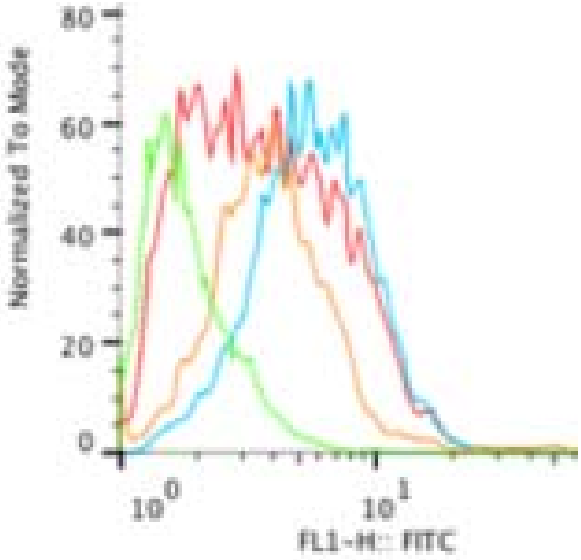
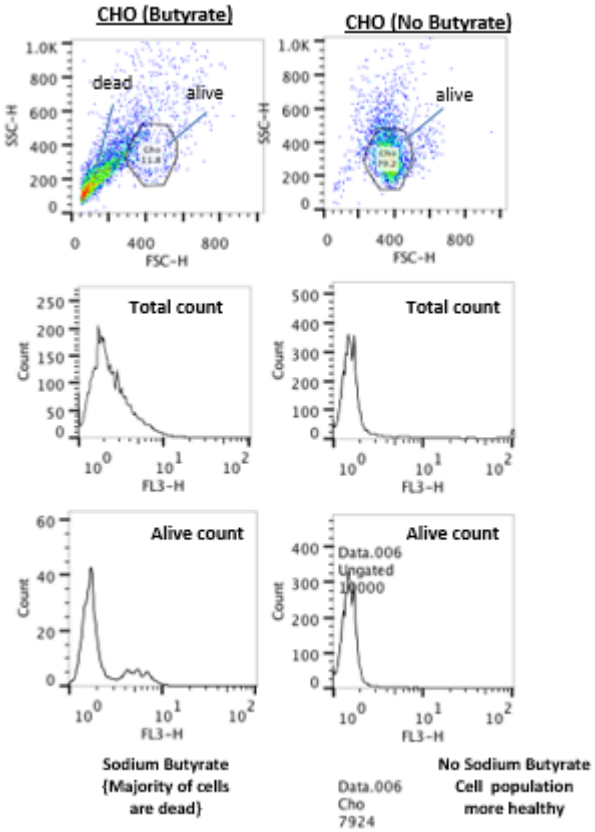


Figure 1. Expression of mSn on CHO-K1, all results conducted in triplicate. Red (1X concentration), Blue (2X concentration), Orange (4X concentration) and Green (8X concentration).

5.2.2 Sodium butyrate assays

The use and investigation of sodium butyrate (NaBu) as an additive for CHO expression has (and continues to be) been investigated within the literature, as a means to increase recombinant protein expression levels.[8] Whilst the exact mechanism for the increase in protein titre is yet to be elucidated, it is speculated to occur via a loosening of chromatin packaging (via H3/H4 histone hyperacetylation due to inactivation of histone deacetylase). The loosening of the chromatin packaging thus results in an increase in transcription and thus protein synthesis.[9,10]

The growth of the mSn d1-17 (domain 1-17) CHO cells was undertaken as per the literature report.[11] A culture was grown in the literature conditions (Ham's F-10, 5% FCS, P/S), as well as with media supplemented with the additive, sodium butyrate (10mM NaBu).



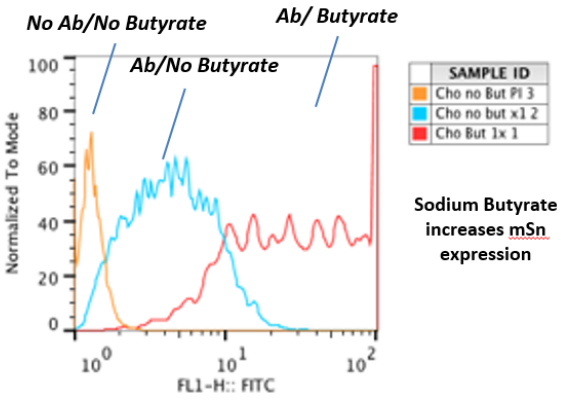


Figure 2. Expression of mSn on CHO-K1. When NaBu is used there is an up-regulation of the receptor on the cell surfaces. The cells were probed with anti-mSn(3D6) FITC conjugated Ab. However, there is also a clear perturbation of the healthy cell population.

From the initial experiments, it was clear that a higher expression of mSn on the CHO cell surface was observed with a higher concentration of sodium butyrate. However, a higher amount of cell death was also seen. A titration of NaBu (0,1,2,5mM) was subsequently conducted to ascertain whether a lower concentration of this additive could be used to improve the cell surface concentration of receptor and thus the data quality.

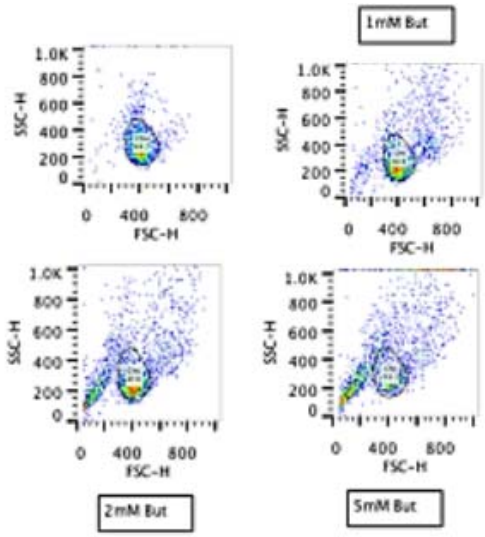


Figure 3. Side Scatter/Front Scatter plots for the varying concentrations of NaBu (0,1,2,5 from top left to bottom right). As the concentration of NaBu increases the quantity of dead and dis-formed cells can be seen to increase.

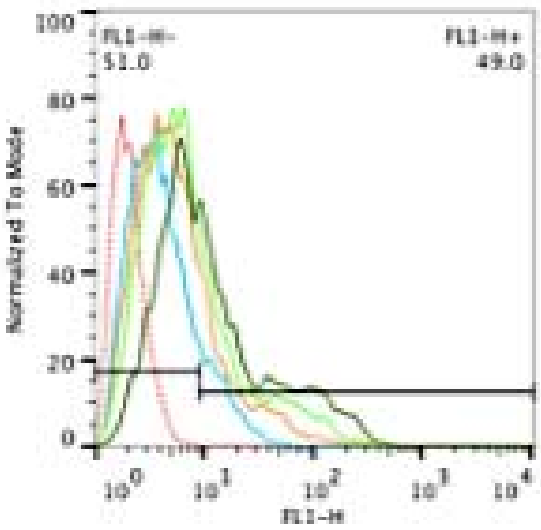


Figure 4. A histogram showing the intensity on the FITC channel (FL1-H). As can be clearly seen with increasing NaBu more mSn (Ab binding to it) can be seen. (Cells (Red), 0mM (Blue), 1mM(Orange), 2mM (Light Green) and 5mM (Dark Green).

Again, from this, the cells with no NaBu look very healthy but on the addition of 1mM NaBu significant cell death is observed. The cell death appears to increase with additional NaBu; however, more mSn is still observed with increasing concentrations of this additive as previously observed. Even at 1mM NaBu cell death and perturbation is still observed.

Based on these observations the use of sodium butyrate was not continued. Despite its use to improve receptor expression levels in more simplistic experiments it was felt that by causing significant cell death we could potentially perturb the binding of the ruler construct(s) to the cell surface receptors. Furthermore, there would also be more difficulty in accurately controlling for the % upregulation of the receptor. A further problem not considered in the literature, is that at higher receptor concentrations there is the potential for multiple receptor–construct binding.

5.2.3. Molecular Ruler binding assay

We subsequently decided to probe the cells with a 30Å FITC/DiSiaLacNAc modified ruler as well as the 30Å FITC ruler. The samples were prepared by PD MidiTrap G-25/dye removal column and dialysis thrice against PBS to fully remove any residual FITC.

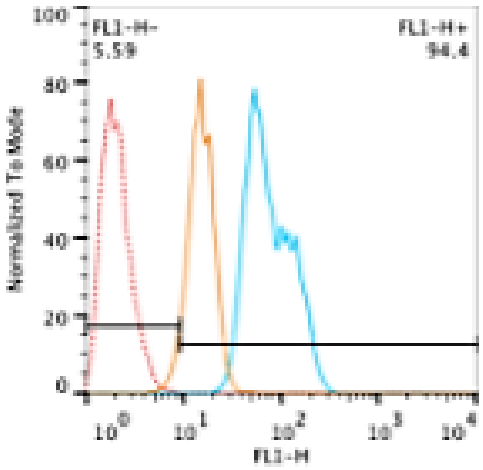


Figure 5. Histograms on the FITC channel for the Np276 DiSiaLacNAc ruler (Blue), FITC labelled ruler (Orange) and the cells with no Ab/ruler (Red). All run in triplicate, for 10,000 events and on the FL1-H channel.

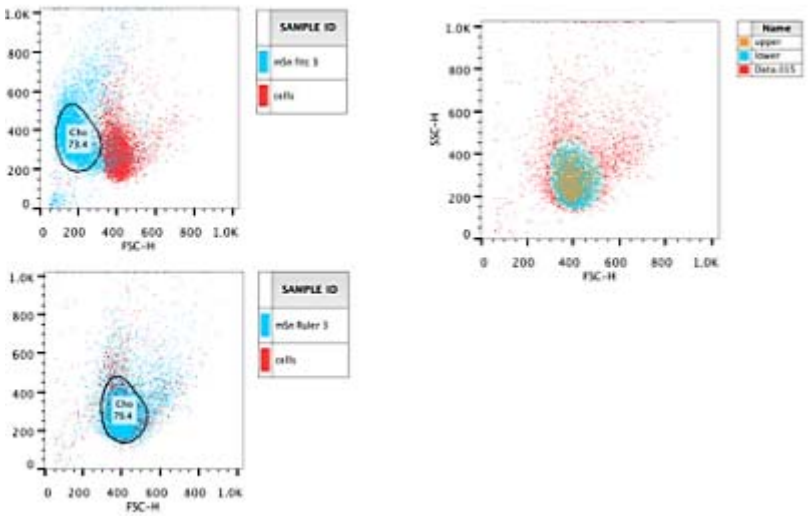


Figure 6. Side Scatter/Front Scatter plots for the ruler labelling experiments. The first plot (top left) shows that when the cells are incubated with just the FITC ruler, a strong change in cell morphology is observed. Whilst for

the DiSiaLacNAc ruler/cells, a splitting is seen on the histogram indicating these cells are from the same population (top right/bottom left)

The results obtained indicated that the cells were interacting associatively with both the glycosylated and the control ruler proteins. Despite these observations, the glycosylated multivalent probe showed a higher binding to the mSn expressing cells. The fact that cells with the FITC-ruler (no sugar) showed a high % of cell binding indicates that the protein linker (Np276 + FITC) were also interacting strongly with the backbone. We also noticed a splitting of the peak on the mSn/DiSiaLacNAc histogram (blue, Figure 5). When running this on the scatter plot we actually observed that this is still part of the same cell population and no morphological changes are observed.

5.3 Conclusions

In conclusion, the preliminary biological evaluation of a molecular ruler binding to a model mSn cell line was conducted. Sodium butyrate, as used in the literature, was evaluated but not used due to its negative effects upon cell health. Whilst a control protein labelled with dye exhibited an associative interaction with the cell surface (Np276 is negatively charged as well, whilst charge dominates the Sn-glycan interaction), the glycosylated protein bound more strongly to the Sn expressing cells. Further experiments would be necessary to ascertain whether the backbone (linker) of this multivalent system is interacting with the Sn; a factor the Whiteside's model of multivalency doesn't account for well. In addition, by using higher affinity glycans such as those outlined previously improved binding could also be obtained, which may improve the data quality. Finally, if the mSn receptor could be probed with the full glycosylated molecular library then subtle biophysical interactions may be elucidated which as highlighted previously could have important therapeutic implications.

5.4 Experimental

Cell maintenance and growth

All cell lines were received as DMSO stocks from Professor Paul Crocker (School of Life Sciences, Dundee). All experiments were conducted in triplicate and reagents purchased from Sigma, Lonza and BioRad. The antibody used was an anti-mSn 3D6.112 clone conjugated to FITC.

Cell lines used: mSn d1-17 CD169 (CHO-K1)

The cell DMSO stocks were thawed rapidly in a 37°C water bath before being added to 10ml of pre-warmed (37°C) Ham's F-10 media supplemented with 5% fetal calf serum, along with 100U/ml and 100ug/ml of pen-strep. The cells were centrifuged (3 minutes, 300g) and then suspended in 10ml of media (Ham's F-10 as above). 1ml of this cell suspension was then added to 40ml of media, in a T-175 flask. The cells were then grown and passaged (after approximately 1 day) before confluence. Cells were subsequently used to create further working stocks after 1 further passage.

Flow cytometry settings:

Flow cytometry was performed on a Becton Dickinson FACSCalibur (four colour, two laser) flow cytometer with excitation at 488 nm (Ar-laser) and fluorescence detected in the FL-1 channel. 10,000 events were collected for each experiment. Data was collected in CellQuest Pro and processed in FlowJo. Due to the low resolution images in the body of the text, coloured referenced captions are given to assist the reader.

Antibody assay of mSn

In the sodium butyrate trials and for FACS analysis a mSn d1-17 Ab conjugated to FITC for FACS analysis was used. The cells were grown until confluence before the media was removed. The media was then replaced (40ml) before NaBu in media (100mM stock solution) was added to a final concentration of 10mM. The cells were then incubated at 37°C, 5% CO_2 for 7 days before being trypsinised (5ml of trypsin EDTA).After incubating for 5 mins at 37°C and agitation the trypsin was quenched with 10ml of media. The cells were counted and plated at 4x10^5 cells a well on a 96 well plate. The cells were then centrifuged (300g, 3min, 4°C) and incubated with Ab (1X(36µl of 0.1mgml^{-1} in 600µL FACS buffer)) for 30mins at 4°C. The cells were probed with a 2x, 4x, and 8x dilution of this Ab. The cells were then washed (centrifuged 3min 300g, 4°C) thrice with 200µL x 3 of FACS buffer (2% FBS/PBS) and re-suspended in 200µL FACS buffer. The cells were subsequently analysed using the previously described flow cytometry settings.

mSnd1-17 NaBu growth trials (initial).

The cells were grown until confluence before the media was removed. The media was then replaced (40ml) before NaBu in media (100mM stock solution) was added to a final concentration of 10mM. The cells were then incubated at 37°C, 5% CO_2 for 7 days before being trypsinised (5ml of trypsin EDTA). After incubating for 5 mins at 37°C and agitation the trypsin was quenched with 10ml of media. The cells were counted and plated at 4x10^5 cells a well on a 96 well plate. The cells were then centrifuged (300g, 3min, 4°C) and incubated with Ab (1X) for 30mins at 4°C. The cells were then washed (centrifuged 3min 300g, 4°C) thrice with 200µL x 3 of FACS buffer (2% FBS/PBS) and re-suspended in 200µL FACS buffer. The cells were subsequently analysed using the previously described flow cytometry settings.

NaBu concentration titration.

The cells were grown until confluence before the media was removed and the cells split into 3 T75 flasks. The media was then replaced (15ml) before NaBu in media (100mM stock solution) was added to a final concentration of 1, 2 and 5mM. The cells were then incubated at 37°C, 5% CO_2 for 2 days before being trypsinised (5ml of trypsin EDTA). After incubating for 5 mins at 37°C and agitation the trypsin was quenched with 10ml of media. The cells were counted and plated at $2x10^{\wedge 5}$ cells a well on a 96 well plate. The cells were then centrifuged (300g, 3min, 4°C) and incubated with Ab (1X) for 30mins at 4°C. The cells were then washed (centrifuged 3min 300g, 4°C) thrice with 200µL x 3 of FACS buffer (2% FBS/PBS) and re-suspended in 200µL FACS buffer. The cells were subsequently analysed using the previously described flow cytometry settings.

Ruler Trials

The cells were grown until confluence before being trypsinised (5ml of trypsin EDTA). After incubating for 5 mins at 37°C and agitation the trypsin was quenched with 10ml of Media. The cells were counted and plated at $2x10^{\wedge 5}$ cells a well on a 96 well plate. The cells were then centrifuged (300g, 3min, 4°C) and incubated with 30Å DiSiaLAcNAc ruler ($0.42mgml^{-1}$, 19µM) and FITC ruler ($0.55mgml^{-1}$, diluted to $0.40mgml^{-1}$ with PBS, 19µM) for 30mins at 4°C. The cells were then washed (centrifuged 3min, 300g, 4°C) thrice with 200µL x 3 of FACS buffer (2% FBS/PBS) and re-suspended in 200µL FACS buffer before being placed on ice in the dark. The cells were subsequently analysed using the previously described flow cytometry settings. All manipulation was undertaken under foil to prevent bleaching and the results obtained in triplicate.

5.5 Bibliography

(1) Crocker, P. R.; Kelm, S.; Dubois, C.; Martin, B.; McWilliam, a S.; Shotton, D. M.; Paulson, J. C.; Gordon, S. *EMBO J.* **1991**, *10* (7), 1661–1669.

(2) Saunderson, S. C.; Dunn, A. C.; Crocker, P. R.; Mclellan, A. D. **2014**, *123* (2), 208–216.

(3) Klaas, M.; Crocker, P. R. *Semin. Immunopathol.* **2012**, *34* (3), 353–364.

(4) Crocker, P. R.; Paulson, J. C.; Varki, A. *Nat. Rev. Immunol.* **2007**, *7* (4), 255–266.

(5) Crocker, P. R. *Curr. Opin. Pharmacol.* **2005**, *5* (4), 431–437.

(6) O'Reilly, M. K.; Paulson, J. C. *Multivalent ligands for siglecs.*, 1st ed.; Elsevier Inc, 2010; Vol. 478.

(7) Kawasaki, N.; Vela, J. L.; Nycholat, C. M.; Rademacher, C.; Khurana, A.; van Rooijen, N.; Crocker, P. R.; Kronenberg, M.; Paulson, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (19), 7826–7831.

(8) Prasad, K. N.; Sinha, P. K. *In Vitro* **1976**, *12* (2), 125–132.

(9) Jiang, Z.; Sharfstein, S. T. *Biotechnol. Bioeng.* **2008**, *100* (1), 189–194.

(10) Chen, F.; Kou, T.; Fan, L.; Zhou, Y.; Ye, Z.; Zhao, L.; Tan, W.-S. *Biotechnol. Bioprocess Eng.* **2011**, *16* (6), 1157–1165.

(11) Kidder, D.; Richards, H. E.; Ziltener, H. J.; Garden, O. a; Crocker, P. R. *J. Immunol.* **2013**, *190* (6), 2593–2602.

Chapter 6

Conclusions and future work

Conclusions

In conclusion, an Np276 ruler protein library was incorporated with azidohomoalanine and characterised using biophysical techniques. Previous group members has expressed the Np276 proteins with no biophysical data and with the PTM (N-terminal gluconylation) present. An initial challenge was a high prevalence of a deleterious PTM: N-terminal gluconylation. In comparison to other proteins in the literature, this protein showed a high propensity towards this modification. Initially, a novel approach based around an intervention into the biochemical pathways of the *E. coli* was investigated. Preliminary work indicated this approach could in future represent an easily scalable approach to suppress PTMs; however much further work is needed to validate this method and to expand on its scope. By using fine-tuned conditions and a mutation, this N-terminal modification could be effectively suppressed. These conditions were used to express a library of Np276 rulers, multiple times. These ruler proteins were characterised by MS, MS/MS, gel and CD; and were characterised by comparison to literature proteins such as Rfr32 as *in-silico* structure prediction has been shown to be inaccurate with this protein class. The results of this analysis indicated that a well-folded library of UAA incorporated proteins had been synthesised in high purity and moderate yield.

These proteins were then modified in a [1+2] method utilising a site selective reaction at an endogenous amino acid, and 2 site specific reactions at the genetically incorporated azide tags. Previous work had looked at conducting dual modifications; however in all cases incomplete or mixed modified products were observed (never homogeneous) thus new strategies were needed. Initially, classical and commercial NHS/SDP esters were investigated (firstly using biotin due to cost) but poor or multiple modifications were observed. Isothiocyanates were thus investigated; in these cases it was found that a single, complete, homogeneous modification was observed. This was eventually characterised to be occurring on the cleavable His-tag (most

likely at the proline) but away from the sugar binding areas. Following this, proteins were modified in a CuAAC reaction to give a dual modified protein construct. In 2015, a [1+n] dual modification was reported by Budisa *et al*;[1] however the model used here is significantly more elegant in its design as no statistical distribution is observed. An immunologically active sugar SiaLacNAc O-alkyne was synthesised via a multistep procedure and then used in the CuAAC reactions. One homogenously modified 30Å construct was prepared but unfortunately this showed a poor secondary structure. It was decided to use this construct in a mSn binding assay, as due to the low affinity interactions of sugar-protein interactions there was concern no binding interaction would be observed.

To test initial binding of the sugar-protein construct and to assess if there was a high background binding, a 30 Å protein modified with SiaLacNAc/FITC and FITC were used in binding assays against mSn CHO cells. This showed a binding difference versus the FITC bound ruler (no sugar). Whilst this does indicate that there may be backbone interactions with the mSn receptors/cell surface the higher degree of binding for the sugar modified proteins indicated that this approach of using a dual modified protein as a molecular ruler may hold promise.

Whilst a dual modified protein library of distance spaced proteins (rulers) was not fully synthesised and tested in this system, due to issues in synthesis and modification of the proteins, this preliminary data and modification/expression protocol developed demonstrates a novel approach to utilising UAA incorporation into proteins.

Future work and perspectives

The initial expansion of this work will be to test the Sn receptor with a full range of modified ruler proteins; this will enable the subtle binding interactions and thus the distance preferences of this receptor to be elucidated. The "ruler" idea is particularly powerful as by the versatile CuAAC reaction (and changing the sugar alkyne) it can act as a platform technology to test other receptors such as other Siglecs, M6PR and the galectins.

From the expression and modification optimisation work, the use of an additives based approach shows promise. Further work that would be needed is: an extension to other proteins, other strains, more repeats, different conditions and protein activity assays. If these can be demonstrated then this approach could become a new tool for protein expression groups. From the protein modification work, the use of an N-terminal proline endogenous reactive handle was discovered and could be expanded to new approaches. Concurrently, others in the literature began looking at the unique properties of this residue.[2]

The concept of a molecular ruler can also act as a scaffold to aid the development of more advanced physical techniques such as DEER EPR and ^{19}F distance measurements. Both these methods were briefly trialled with the Np276 protein here (EPR problems were encountered with signal strength for DEER and with the ^{19}F NMR, we were unable to conjugate on a short chain fluorinated alkyne for the study), and in the future could be two exciting applications for this UAA ruler approach.[3,4]

Finally, looking forward other proteins with a regular structure but perhaps a more robust stability profile could also be investigated. This would save time, allow for the more rapid generation of libraries and also give more useful tools of function. A potential way to enact this would be through the incorporation of a fluorous core into the Np276 protein through use of residue specific incorporation (Chapter 1).

Perhaps the largest opportunity is in the incorporation of multiple UAAs into a single protein. As noted in Chapter 1, there are several powerful systems for this and by combining these methods, proteins with increasingly complex (but bespoke/unnatural) architectures can be created. Whilst, the community is focussed on creating new methods for these multiple incorporations there exists a need to demonstrate chemical orthogonality in biological systems. Questions such as, can an CuAAC and Suzuki reaction on a dual modified protein be done either way around without degradation of the opposing reactive handle? By investigating these questions, more complex systems for the interrogation of biological systems can be synthesised. Some early exploratory work was undertaken into combining the stop codon methodology utilised by Spicer *et al* (Suzuki chemistry) with that of the Aha residue specific incorporation. This method proved unsuccessful, however with more significant optimisation it could be developed into a more general approach. [5,6]

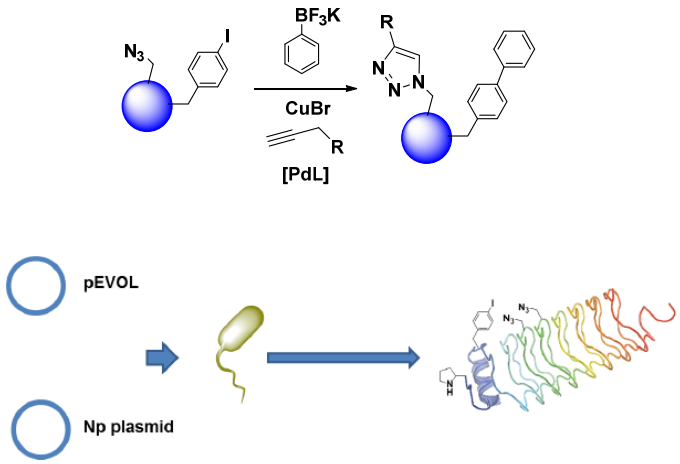


Figure 1. Dual incorporation of the Np276 plasmid and a pEVOL (iodo-phenylalanine incorporation) into an auxotrophic strain, should with significant optimisation result in a dual (UAA) modified protein

Bibliography

(1) Mühlberg, M.; Hoesl, M. G.; Kuehne, C.; Dernedde, J.; Budisa, N.; Hackenberger, C. P. R. *Beilstein J. Org. Chem* **2015**, *11*, 784–791.

(2) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, *136* (27), 9572–9579.

(3) Sale, K.; Song, L.; Liu, Y.-S.; Perozo, E.; Fajer, P. *J. Am. Chem. Soc.* **2005**, *127* (26), 9334–9335.

(4) Azarkh, M.; Okle, O.; Singh, V.; Seemann, I. T.; Hartig, J. S.; Dietrich, D. R.; Drescher, M. *Chembiochem* **2011**, *12* (13), 1992–1995.

(5) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2013**, *49* (27), 2747–2749.

(6) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, 4740.

Appendix

Reprint Permissions

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 04, 2016

This Agreement between Peter Crane ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3961850542319
License date	Oct 04, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Opinion in Chemical Biology
Licensed Content Title	Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications
Licensed Content Author	Jeremiah A Johnson,Ying Y Lu,James A Van Deventer,David A Tirrell
Licensed Content Date	December 2010
Licensed Content Volume Number	14
Licensed Content Issue Number	6
Licensed Content Pages	7
Start Page	774
End Page	780
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figure 1
Title of your thesis/dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Estimated size (number of pages)	300
Elsevier VAT number	GB 494 6272 12
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Total	0.00 GBP
Terms and Conditions	

INTRODUCTION

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Oct 04, 2016

This Agreement between Peter Crane ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3961961463624
License date	Oct 04, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire
Licensed Content Author	Nediljko Budisa
Licensed Content Date	Dec 1, 2004
Licensed Content Pages	38
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 2
Will you be translating?	No
Title of your thesis / dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Expected size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, United Kingdom Attn: Peter Crane
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with

Title: Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins

Author: Kathrin Lang, Jason W. Chin

Publication: Chemical Reviews

Publisher: American Chemical Society

Date: May 1, 2014

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Oct 05, 2016

This Agreement between Peter Crane ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	3962831391785
License date	Oct 05, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Reviews Immunology
Licensed Content Title	Siglecs and their roles in the immune system
Licensed Content Author	Paul R. Crocker,James C. PaulsonandAjit Varki
Licensed Content Date	Apr 1, 2007
Licensed Content Volume Number	7
Licensed Content Issue Number	4
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 5
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Estimated size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Billing Type	Invoice
Billing Address	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, United Kingdom Attn: Peter Crane
Total	0.00 GBP
Terms and Conditions	

Creative Commons License Deed

Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0)

This is a human-readable summary of (and not a substitute for) the license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

SPRINGER LICENSE
TERMS AND CONDITIONS

Oct 03, 2016

This Agreement between Peter Crane ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

License Number	3961630597037
License date	Oct 03, 2016
Licensed Content Publisher	Springer
Licensed Content Publication	Applied Microbiology and Biotechnology
Licensed Content Title	Squaring up to DNA: pentapeptide repeat proteins and DNA mimicry
Licensed Content Author	Shama Shah
Licensed Content Date	Jan 1, 2014
Licensed Content Volume Number	98
Licensed Content Issue Number	23
Type of Use	Thesis/Dissertation
Portion	Figures/tables/illustrations
Number of figures/tables/illustrations	4
Author of this Springer article	No
Order reference number	
Original figure numbers	Figure 4
Title of your thesis / dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Estimated size(pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Billing Type	Invoice
Billing Address	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, United Kingdom Attn: Peter Crane
Total	0.00 GBP
Terms and Conditions	

Introduction
The publisher for this copyrighted material is Springer. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Oct 03, 2016

This Agreement between Peter Crane ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	3961631105481
License date	Oct 03, 2016
Licensed content publisher	Oxford University Press
Licensed content publication	Nucleic Acids Research
Licensed content title	Structural insights into quinolone antibiotic resistance mediated by pentapeptide repeat proteins: conserved surface loops direct the activity of a Qnr protein from a Gram-negative bacterium:
Licensed content author	Xiaoli Xiong, Elizabeth H. C. Bromley, Peter Oelschlaeger, Derek N. Woolfson, James Spencer
Licensed content date	05/01/2011
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Protein based molecular probes by unnatural amino acid incorporation
Publisher of your work	n/a
Expected publication date	Dec 2016
Permissions cost	0.00 GBP
Value added tax	0.00 GBP
Total	0.00 GBP
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	GB125506730
Billing Type	Invoice
Billing Address	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, United Kingdom Attn: Peter Crane
Total	0.00 GBP
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Oct 03, 2016

This Agreement between Peter Crane ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3961631306579
License date	Oct 03, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Protein Science
Licensed Content Title	Characterization of two potentially universal turn motifs that shape the repeated five-residues fold—Crystal structure of a lumenal pentapeptide repeat protein from Cyanothece 51142
Licensed Content Author	Garry W. Buchko,Shuisong Ni,Howard Robinson,Eric A. Welsh,Himadri B. Pakrasi,Michael A. Kennedy
Licensed Content Date	Nov 1, 2006
Licensed Content Pages	17
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	9
Original Wiley figure/table number(s)	Figure 9
Will you be translating?	No
Title of your thesis / dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Expected size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, United Kingdom Attn: Peter Crane
Total	0.00 GBP
Terms and Conditions	

TERMS AND CONDITIONS

ROYAL SOCIETY OF CHEMISTRY LICENSE TERMS AND CONDITIONS

Oct 04, 2016

This Agreement between Peter Crane ("You") and Royal Society of Chemistry ("Royal Society of Chemistry") consists of your license details and the terms and conditions provided by Royal Society of Chemistry and Copyright Clearance Center.

License Number	3961951125473
License date	Oct 04, 2016
Licensed Content Publisher	Royal Society of Chemistry
Licensed Content Publication	Chemical Communications (Cambridge)
Licensed Content Title	Protein-based materials, toward a new level of structural control
Licensed Content Author	Jan C. M. van Hest,David A. Tirrell
Licensed Content Date	Sep 18, 2001
Licensed Content Issue Number	19
Type of Use	Thesis/Dissertation
Requestor type	academic/educational
Portion	figures/tables/images
Number of figures/tables/images	1
Format	print and electronic
Distribution quantity	3
Will you be translating?	no
Order reference number	
Title of the thesis/dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Estimated size	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Billing Type	Invoice
Billing Address	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, United Kingdom Attn: Peter Crane
Total	0.00 GBP

Terms and Conditions

This License Agreement is between {Requestor Name} (“You”) and The Royal Society of Chemistry (“RSC”) provided by the Copyright Clearance Center (“CCC”). The license consists of your order details, the terms and conditions provided by the Royal Society of Chemistry, and the payment terms and conditions.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 04, 2016

This Agreement between Peter Crane ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3961950885278
License date	Oct 04, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Chemistry & Biology
Licensed Content Title	Towards the nonstick egg: designing fluorous proteins
Licensed Content Author	E Neil,G Marsh
Licensed Content Date	1 July 2000
Licensed Content Volume Number	7
Licensed Content Issue Number	7
Licensed Content Pages	5
Start Page	R153
End Page	R157
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	5
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	
Title of your thesis/dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Estimated size (number of pages)	300
Elsevier VAT number	GB 494 6272 12
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 04, 2016

This Agreement between Peter Crane ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3961951322948
License date	Oct 04, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Molecular Biology
Licensed Content Title	Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins
Licensed Content Author	Jae Hyun Bae,Marina Rubini,Gregor Jung,Georg Wiegand,Markus H.J. Seifert,M.Kamran Azim,Jeong-Sun Kim,Andreas Zumbusch,Tad A. Holak,Luis Moroder,Robert Huber,Nediljko Budisa
Licensed Content Date	16 May 2003
Licensed Content Volume Number	328
Licensed Content Issue Number	5
Licensed Content Pages	11
Start Page	1071
End Page	1081
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	2
Title of your thesis/dissertation	Protein based molecular probes by unnatural amino acid incorporation
Expected completion date	Dec 2016
Estimated size (number of pages)	300
Elsevier VAT number	GB 494 6272 12
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Jan 30, 2018

This Agreement between Peter Crane ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4278880474031
License date	Jan 30, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Sequence-Controlled Multi-Block Glycopolymers to Inhibit DC-SIGN-gp120 Binding
Licensed Content Author	Qiang Zhang,Jennifer Collins,Athina Anastasaki,Russell Wallis,Daniel A. Mitchell,C. Remzi Becer,David M. Haddleton
Licensed Content Date	Mar 11, 2013
Licensed Content Pages	5
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	2
Will you be translating?	No
Title of your thesis / dissertation	Molecular probes by unnatural amino acid incorporation
Expected completion date	Jan 2018
Expected size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	EU826007151
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 29, 2018

This Agreement between Peter Crane ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4278381045549
License date	Jan 29, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Chemistry & Biology
Licensed Content Title	Chemical Lectinology: Tools for Probing the Ligands and Dynamics of Mammalian Lectins In Vivo
Licensed Content Author	Brian Belardi,Carolyn R. Bertozzi
Licensed Content Date	Aug 20, 2015
Licensed Content Volume	22
Licensed Content Issue	8
Licensed Content Pages	11
Start Page	983
End Page	993
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 4
Title of your thesis/dissertation	Molecular probes by unnatural amino acid incorporation
Expected completion date	Jan 2018
Estimated size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions

Designing logical codon reassignment – Expanding the chemistry in biology

A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chem. Sci.*, 2015, **6**, 50
DOI: 10.1039/C4SC01534G

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Feb 04, 2018

This Agreement between Peter Crane ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4282170953878
License date	Feb 04, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Sequence-Controlled Multi-Block Glycopolymers to Inhibit DC-SIGN-gp120 Binding
Licensed Content Author	Qiang Zhang,Jennifer Collins,Athina Anastasaki,Russell Wallis,Daniel A. Mitchell,C. Remzi Becer,David M. Haddleton
Licensed Content Date	Mar 11, 2013
Licensed Content Pages	5
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 2
Will you be translating?	No
Title of your thesis / dissertation	Molecular probes by unnatural amino acid incorporation
Expected completion date	Jan 2018
Expected size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	EU826007151
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Feb 04, 2018

This Agreement between Peter Crane ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4282171232060
License date	Feb 04, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Thiyl Glycosylation of Olefinic Proteins: S-Linked Glycoconjugate Synthesis
Licensed Content Author	Nicola Floyd,Balakumar Vijayakrishnan,Julia R. Koeppe,Benjamin G. Davis
Licensed Content Date	Sep 8, 2009
Licensed Content Pages	5
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure s4
Will you be translating?	No
Title of your thesis / dissertation	Molecular probes by unnatural amino acid incorporation
Expected completion date	Jan 2018
Expected size (number of pages)	300
Requestor Location	Peter Crane 126 Ferry Hinksey Road Oxford Oxford, other United Kingdom Attn:
Publisher Tax ID	EU826007151
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that