The self-assembly of nanoarchitectures via protein-ligand interactions

Thomas Reuben Branson

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

The University of Leeds
School of Chemistry

September 2013

The great book, always open and which we should make an effort to read, is that of Nature.

Antoni Gaudí

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

1 An introduction, includes content from the publication: “Bacterial toxin inhibitors based on multivalent scaffolds” Chemical Society Reviews, 2013, volume 42, pages 4613-4622, T. R. Branson and W. B. Turnbull. The manuscript was composed by the candidate and W. B. Turnbull.

2 Building blocks, includes content from the publication: “Towards a Structural Basis for the Relationship Between Blood Group and the Severity of El Tor Cholera” Angewandte Chemie International Edition, 2012, volume 51, pages 5143-5146, P. K. Mandal, T. R. Branson, E. D. Hayes, J. F. Ross, J. A. Gavín, A. H. Daranas and W. B. Turnbull. The candidate created the I47T El Tor CTB mutant and performed ITC experiments with the oligosaccharides. Work carried out by the other authors has been noted within the thesis.

Acknowledgments

Firstly I would like to thank all those that I have shared the lab with over the last few years. James, Tom, Ivona, Kat, Kristian, Phil, Dan, Chadamas, Diana, Darren, Heather, Jeff and Ed have made my working days a joy.

I thank Dr Martin Fascione, Dr Tom M^{c}Allister, Dr Pintu Mandal and James Ross for providing me with lots of toys to play with. An excellent group effort.

The cooperation of the technical staff in the School of Chemistry and the Astbury Centre is also warmly acknowledged, with special thanks to Martin Huscroft. I would like to thank Dr Stuart Warriner, Dr Iain Manfield, Dr Mike Webb and Dr Simon Connell for their help and expert opinions during the project.

I would like to thank my family for their never ending support, especially in the last few weeks.

Ik wil Lieke bedanken voor haar liefde, begrip en wat perspectief.

Finally, I thank Dr Bruce Turnbull for his help and guidance, long discussions and constant encouragement.[1]

Abstract

Nature has evolved proteins that can spontaneously self-assemble to create complex structures such as virus particles and molecular motors. The fields of bionanoscience and synthetic biology are based on the concept that by combining biological building blocks with synthetic molecules it will be possible to construct novel nanoarchitectures and machines that can do useful work. This thesis describes strategies that have been developed using protein-ligand interactions to construct nanoscale assemblies.

The B-subunit of cholera toxin (CTB) has been site-specifically modified at the N-teminus, by oxime ligation, with carbohydrate ligands and the assembly of these modified proteins was observed. A W88E non-binding mutant of CTB was made and modified with GM1os ligands. The interaction of this pentavalent protein-based ligand with wild-type CTB has been investigated and showed the formation of a protein heterodimer.

The CTA2-subunit of the cholera toxin AB_5 complex was also modified at the N-terminus enzymatically with depsipeptides via sortase ligation and by oxime ligation. Biotin ligands have been covalently attached and the formation of a 2:1 complex with streptavidin was observed.

The structures and assemblies demonstrated herein have been analysed and characterised by a range of analytical and biophysical techniques including dynamic light scattering, analytical ultracentrifugation and isothermal titration calorimetry.

Abbreviations

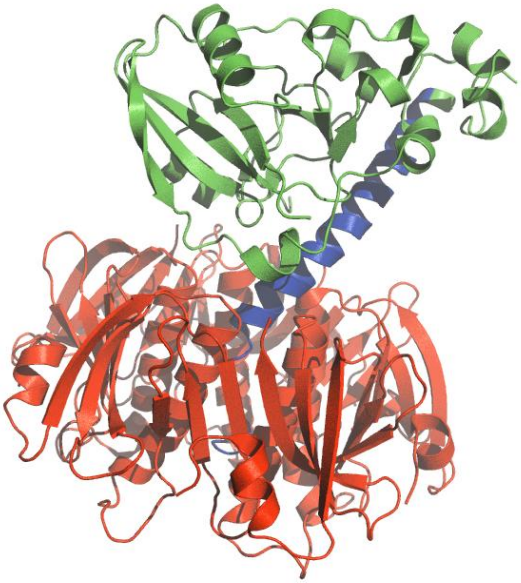
AFM	Atomic force microscopy
APS	Ammonium persulfate
AUC	Analytical ultracentrifugation
Boc	tert-Butoxylcarbonyl
BPSA	Bathophenanthrolinedisulfonic acid
BSA	Bovine serum albumin
CID	Chemical inducer of dimerisation
CIP	Calf-intestinal phosphatase
CMC	Critical micelle concentration
ConA	Concanavalin A
CT	Cholera toxin
CTA1	Cholera toxin A-subunit 1
CTA2	Cholera toxin A-subunit 2
CTB	Cholera toxin B-pentamer
CuAAC	Copper-assisted azide-alkyne cycloaddition
Dde	1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl
DIPEA	N,N-Diisopropylethylamine
DLS	Dynamic light scattering
DMAP	4-Dimethylaminopyridine
DMF	N,N-dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DOPC	1,2-Dioleoyl-sn-glycero-3-phosphocholine
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E. coli	Escherichia coli
ELISA	Enzyme-linked immunosorbent assay
ELLA	Enzyme-linked lectin assay
EM	Electron microscopy
ESI-MS	Electrospray ionisation-mass spectrometry
ESMS	Electrospray mass spectrometry
Fmoc	Fluorenylmethyloxycarbonyl

FPLC	Fast protein liquid chromatography
Gal	Galactose
GM1os	GM1 oligosaccharide
h	Hours
HCTU	2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HEPES	2-[4-(2-Hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HFIP	Hexafluoroisopropanol
HPLC	High Performance Liquid Chromatography
HRMS	High resolution mass spectrometry
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Lac	Lactose
LB	Lysogeny broth
LT	Heat-labile toxin
LTB	Heat-labile toxin B-pentamer
MBL	Mannose binding lectin
MBP	Maltose-binding protein
min	Minutes
MLV	Multilamellar vesicle
MNPG	*meta*-nitrophenyl α-D-galactopyranoside
MS	Mass spectrometry
OPD	o-Phenylenediamine
PBS	Phosphate-buffered saline
PCA	Polymer cycling assembly
PCR	Polymerase chain reaction
PFU	DNA polymerase (from Pyrococcus furiosus)
PPI	Protein-protein interaction
PWO	DNA polymerase (from Pyrococcus woesei)
RNA	Ribonucleic acid
rt	Room temperature
SAP	Serum amyloid P component
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SPPS	Solid phase peptide synthesis

Stx	Shiga-like toxin
StxB	Shiga-like toxin B-pentamer
SUV	Small unilamellar vesicles
TBTA	Tris-(Benzyltriazolylmethyl)amine
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
Tris	Tris(hydroxymethyl)aminomethane
TX-100	Triton X-100
VT	Verotoxin
w / v	Weight per volume

Contents

1 An introduction



Part I:

From the wilds of Nature to the confines of the lab

In this introduction, the broad discipline of self-assembling structures will be outlined with particular focus on protein assemblies mediated by protein-ligand interactions. The discussion will range from the wonders of the natural world to the humble ideas and designs created by man.

1.1 Nature the great architect

Taking a minute to think about the great variety of life on this planet can only leave you with a feeling of awe. The multitude of complex and beautiful forms that have covered our world seem to be innumerable in their differences. On the nanoscale these differences can become even more apparent. Even within a single species many different tissues are present and each may comprise many different types of cell.

There is structure to everything. On every scale a new level of appreciation can be found, from the straight, solid trunks of huge redwood trees, to the flexible muscle tissue allowing our movement and to the delicate, helical DNA that encodes all existence. The huge variety of natural structures is easy to see around us and on the nanoscale bioarchitecture can be just as impressive.

It took billions of years for the natural universe to be shaped around us, yet it is only in the last hundred years or so that humans have truly began to understand the fundamental principles that hold this world together. We are starting to better understand the complex mechanisms that govern how all these systems interact and scientists can now use this information to change and add to what nature has already created.

A central goal of synthetic biology is to learn from nature and take inspiration. This knowledge can then be used to create bioinspired and novel artificial architectures and systems that meet our needs, demands and are limited only by our own imagination.[2]

1.2 The rise of synthetic biology

Synthetic biology is an emerging field of science that focuses on the design and construction of biological materials and re-design of natural biological systems for purposes useful to us.[3] Whether this involves redesigning proteins, building novel biological networks or creating new life, synthetic biology combines ideas from a range of disciplines such as engineering, chemistry, biology and computing.

The term synthetic biology was first coined in 1912 by Stéphane Leduc with the publication of ‘La Biologie Synthétique’.[4] It has been redefined over the years; from the production of synthetic life forms, to a more recent definition of the design, construction and re-design of biological systems.[3] One thing that is clear is the overall goal of a greater understanding of nature and how this knowledge can be used to help us. By mimicking nature and gaining more information on the associations between different natural systems, it will become possible to assemble our own new programmed semi-biological systems.

The standardisation of biological parts is a major aim of synthetic biology. As has been done with electronics and building circuit boards, it is hoped that the building blocks of biological systems can be standardised to allow ‘plug and play’ design for non-natural functionality from a combination of natural and synthetic components. The Registry of Standard Biological Parts has already started along this path and hopes to provide a resource to make biology simpler to engineer.[5] An annual challenge to students is The International Genetically Engineered Machines competition (iGEM) where a kit of standard parts is used, to design and create new living biological systems.[6] Other sources such as the well established Protein Data Bank also add to this collective knowledge of standard parts.[7]

There has been a recent increase in attention from the media concerning synthetic biology. A BBC Horizon episode broadcast in January 2012 provocatively titled ‘Playing God’ was one of many programs and articles hoping to inform the public of this new branch of science.[8] As this area gains more interest, more people see the benefits and ethical questions start to be raised. Many organisations are working towards a better public understanding through outreach programmes, conferences and a wealth of internet based information.[3, 9-12]

Richard Feynman’s frequently quoted statement that, “What I cannot create, I do not understand”, holds true in synthetic biology and nanotechnology. Nature has achieved a remarkable array of different structures and the goal of scientists working in this area is to mimic and ultimately improve biological systems by combining genetically and structurally altered natural molecules.[13] A whole range of new materials could be constructed which would be used to solve problems in areas such as drug delivery and tissue engineering.[14]

The diverse creations that come from synthetic biology all have one thing in common, and that is self-assembly.[15] Whether new genes are being read, or artificial proteins are folding up, these processes happen spontaneously without human input. When creating new systems we must take advantage of the strategies already in place, one of the most powerful being self-assembly.

1.3 A lot of recognition for self-assembly

The key concept in both natural and synthetic multi-component systems is the ability of the materials to self-assemble into larger, more useful structures. For this result to occur, the components involved must have the ability to recognise each other and then stick together. The principles of self-assembly are often anthropomorphised through the use of terms such as “intelligent molecules” that “recognise” one another, which can easily distract the reader from the true wonder of these systems that have no intelligence at all. All processes are driven by diffusion and Brownian motion, and simply happen without choice. Nevertheless with a little external design and a push in the right direction, using self-assembly could trivialise the creation of large complicated structures.

There are many examples of self-assembly in nature, many of which are essential for life, including protein folding, the formation of cell membranes and DNA double helix formation. The human body is essentially a self-assembled system, with each cell constructing itself to enable life to exist. This ability to spontaneously organise also provides us with the tools to construct novel nanoscale materials.

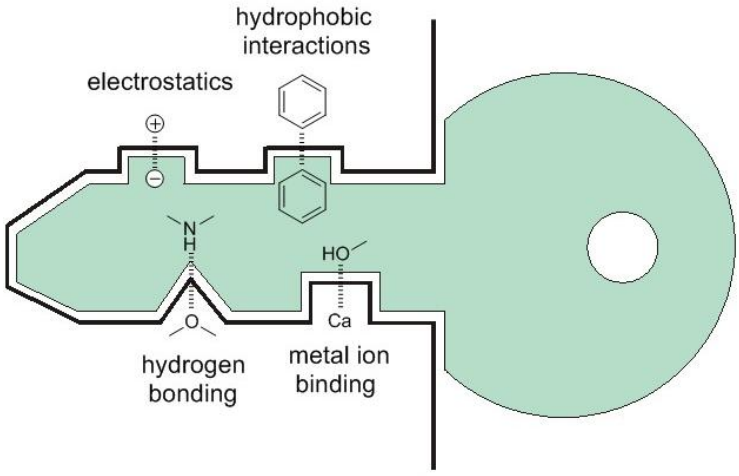


Fig 1.1 The concept of Emil Fischer’s ‘lock and key’ model with common supramolecular forces.

Scientists that boast the most well understood hold over self-assembly work in supramolecular chemistry as this field is reliant on this concept when creating structures. Supramolecular chemistry focuses on the use of non-covalent bonds including hydrogen bonds, hydrophobic interactions, and electrostatic interactions including metal binding and π-π stacking. With only a few simple recognition motifs, elaborate constructions can be created. The concept of molecular recognition has come a long way since Emil Fischer proposed a lock and key model for an enzyme reaction in 1890 (Fig 1.1). In 1987, Cram, Lehn and Pedersen shared the Nobel Prize for chemistry for their work on host-guest interactions and properly established the field of supramolecular chemistry.[16] It is beyond the scope of this introduction to discuss all the possibilities supramolecular chemistry affords and the area has been extensively reviewed.[17-20]

Nature has evolved using these weak interactions to induce the spontaneous organisation of many complex biological creations. For example, the cell membrane is formed by lipids clustering their hydrophobic tails; actin proteins polymerise into filaments to stabilise the structure of cells; and histones interact with DNA, winding it up into chromatin in the cell nucleus. Looking at the elegance of the interactions employed by biological systems, with a knowledge of the forces involved, has allowed scientists to explore the use of biomolecules as supramolecular building blocks.[21] So far, the manipulation of DNA and proteins has been the main focus of research in this area.

1.3.1 DNA: Boxes full of origami

The DNA double helix comprises only four different nucleotides (adenosine, thymidine, cytidine and guanosine) and stores all the information needed for life. This simplicity makes it an alluring material for the scientist interested in self-assembly. The double helix structure is held together via hydrogen bonds between each base pair and this complementarity allows DNA to be easily manipulated. The helix of DNA naturally forms with a right-hand rotation but it has been seen that under high salt conditions certain sequences can exist as left-handed. Recent research has also shown that pH can influence the handedness of DNA with guest molecules bound and so great control over these self-assembling superstructures can be achieved.[22]

Ned Seeman first showed the possibilities of what could be achieved using DNA [23] by constructing the first example of a polyhedron made from DNA. Here a cube was

created with each face of the object being a single-stranded cyclic piece of DNA, which was doubly catenated to four other strands and therefore holding the cube together (Fig 1.2). This groundbreaking work was performed to create structures that could be used to template larger assemblies, however it also opened up the potential for DNA to be used to create designed self-assembled structures for a variety of other uses.[24]

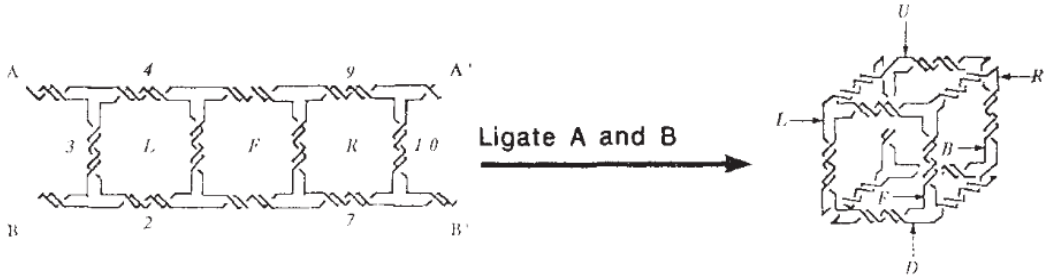


Fig 1.2 Ned Seeman's design of DNA catenated and ligated to form a cube.[23]

Since then so-called DNA origami has led to the creation of intricate two-dimensional assemblies and then more complicated three-dimensional structures. Here a long strand of DNA is used as a scaffold whilst smaller staple strands bind and fix the complete assembly into the desired shape. DNA origami has seen the construction of a vast array of structures from a 100 nm smiley face[25] to a megadalton-sized DNA box with an interior large enough to hold a ribosome[26] (Fig 1.3). This box even had a lid that could be 'unlocked' by the addition of 'key' oligonucleotides.

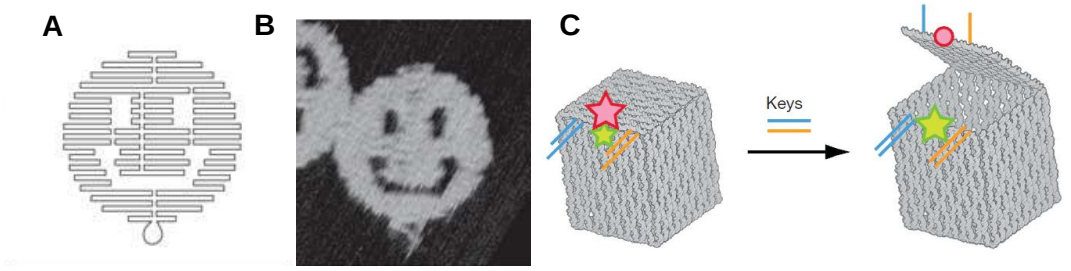


Fig 1.3 From (A) the DNA origami design of a smiley face, to (B) the AFM image of the assembled structure.[25] (C) A lockable box from DNA origami.[26]

Furthermore, DNA polyhedra have been produced that use a smaller number of different DNA strands so as to simplify the synthesis.[27] These strands then form three-point motifs that join together to form symmetrical three-dimensional objects such as a tetrahedron, a dodecahedron and a buckyball (truncated icosahedron). The objects created were dependent on concentration and the number of bases in the central loop of the motif (Fig 1.4). Polyhedra have also been made from tRNA

which takes advantage of the natural tRNA(Ser) structural motif which enables a rigid, stable nanostructure to be formed.[28] These structures are not only beautiful examples of the control that scientists can exert over biomolecules but they can also potentially be useful as scaffolds or nanoscale containers. Nucleotide manipulation has come so far now that almost any two or three dimensional architecture can be designed and constructed with extremely high precision.[29, 30]

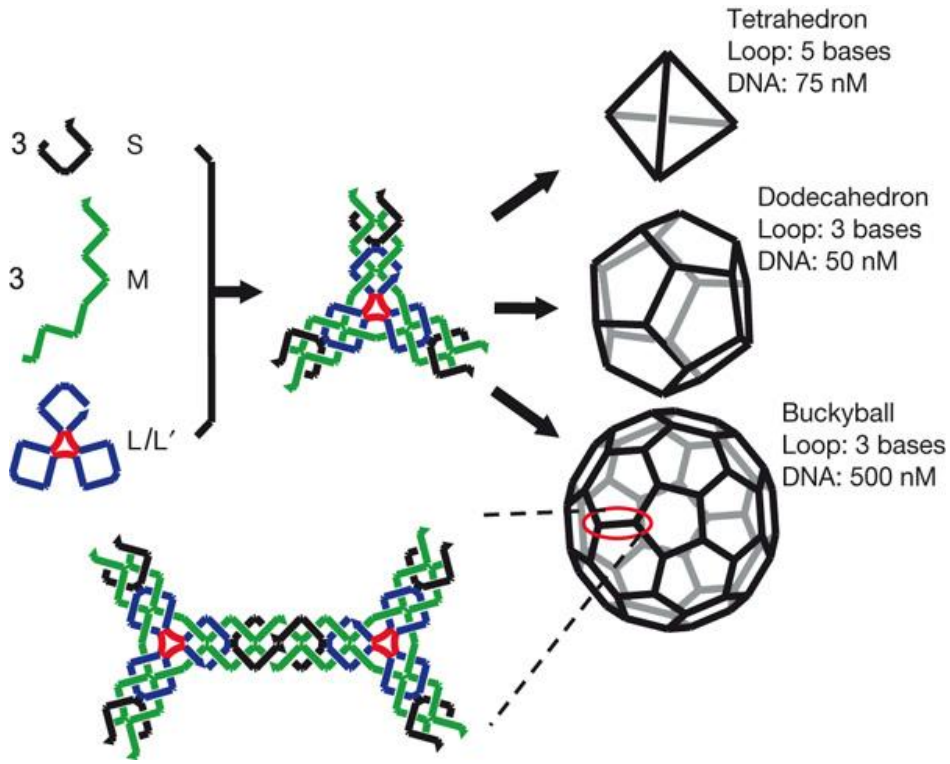


Fig 1.4 Self-assembling DNA stands into different three-dimensional structures.[27]

1.3.2 From peptide filaments to protein cages

Proteins represent a step up in complexity from DNA as we move to using a possible 20 natural amino acids in these biopolymers. Protein folding is a complicated phenomenon that is still not fully understood.[31] Hydrophobic interactions, polar contacts and salt bridges contribute to creating the stable, and often complicated, tertiary structures of proteins.

Proteins can interact with one another to give rise to the formation of many elegant architectures in nature. Research into protein-protein interactions (PPIs) has tended to focus on the self-assembly of peptides as these components are easier to make chemically than full proteins.[32] Coiled coils are a natural assembly and have been researched extensively.[33] They exist when two or more peptide α-helices interact

to coil around each other. They have hydrophobic ‘knobs in holes’ interactions between the strands and salt bridges also help to hold the structures together. The Woolfson group is leading the way in this area and has designed and created some complex coiled-coil structures.[34, 35] A recent study by Gradišar *et al.* used coiled coils to create a self-assembling three-dimensional tetrahedron from a single polypeptide chain (Fig 1.5).[36] This work showed that the design and manipulation of coiled coils has the potential to become as standardised as DNA origami.

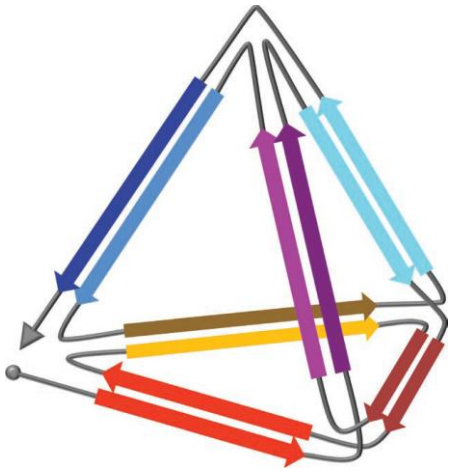


Fig 1.5 Schematic representation of the polypeptide self-assembling tetrahedron.[36]

Amyloid is another natural peptide supramolecular assembly. The development of this type of structure has been linked closely to Alzheimer’s disease and so these structures are the subject of much research.[37] Amyloids are created when proteins misfold into cross-β sheets which then form insoluble fibrous aggregates.[38] A further understanding of the formation of amyloids has been used as a guide to the design of β-sheet proteins. One example of recent work has seen an artificial β-barrel created from peptides designed to self-assemble into β-sheets.[39]

Viruses offer a plethora of self-assembled protein structures; from the simple icosohedral structure of the rhinovirus capsid, to the tubular arrangement of the Tobacco mosaic virus coat proteins and the almost alien appearance of bacteriophages. The viral proteins self-organise around nucleic acids to form symmetrical structures of varying shape and size ranging from a modest 30 nm for the rhinovirus up to a 440 nm diameter for the megavirus[40]. These constructions rely on the strength of many PPIs over large surface areas to give them great stability.

Building on this knowledge of control over peptides and PPIs there has been much research to manipulate these interactions. A non-natural peptide assembly has been created by Ghadiri and co-workers.[41] They used cyclic D,L-peptides that then self-organise to form nanotubes. Alternating D- and L-amino acids are used because in this configuration they can stack with backbone-backbone hydrogen bonds (Fig 1.6). The internal diameter of the nanotubes can be controlled simply by altering the size of the peptide rings and these structures could be employed as ion channels or closed reaction chambers.

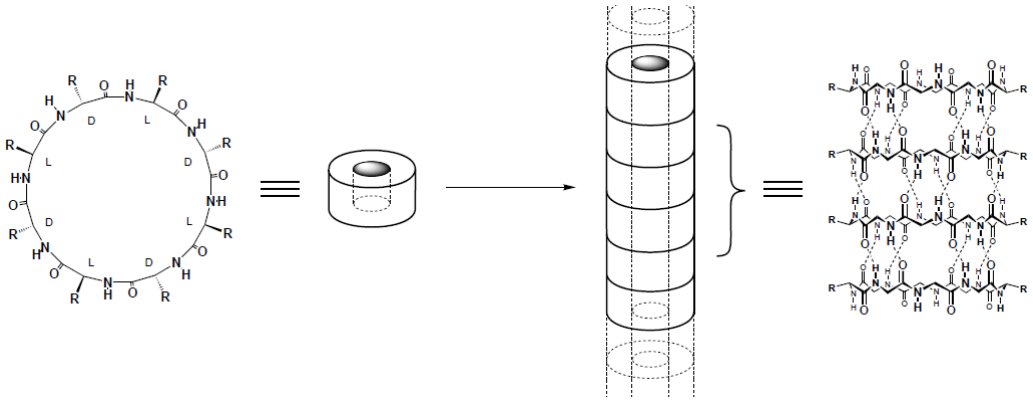


Fig 1.6 Ghadiri's cyclic structures of alternating D- and L-amino acids formed self-assembled nanotubes held together by hydrogen bonding.[41]

Fusion proteins have been used by Padilla *et al.* in a new strategy for the design of large nanoscale assemblies.[42] Two different proteins that each want to naturally assemble into oligomers were used; a natural dimer and a natural trimer. The different protein monomers were fused together and then oligomerisation allowed for the self-assembly of a nanoscale cage or a flat array depending on the geometry of the fusion protein (Fig 1.7). This work was quite elegant in the way that prior design led to the desired supramolecular complexes. Tetrahedra, fibres and two-dimensional lattices have also been made by similar techniques.[43, 44]

Breakthrough work from the Baker group showed the design of a novel protein-protein interface that facilitated the construction of a 24-subunit cage with octahedral symmetry.[45] In this study computational docking was used to predict which mutations in the proteins would lead to the desired interactions. Crystal structures of the protein constructions revealed atomic level accuracy to the original computational design. Many other structures that have been created through PPIs have been reviewed extensively.[46, 47]

Another characteristic of some proteins is their ability to bind small molecules. When this feature is exploited for initiating self-assembly, a whole new range of designs can be accessed.

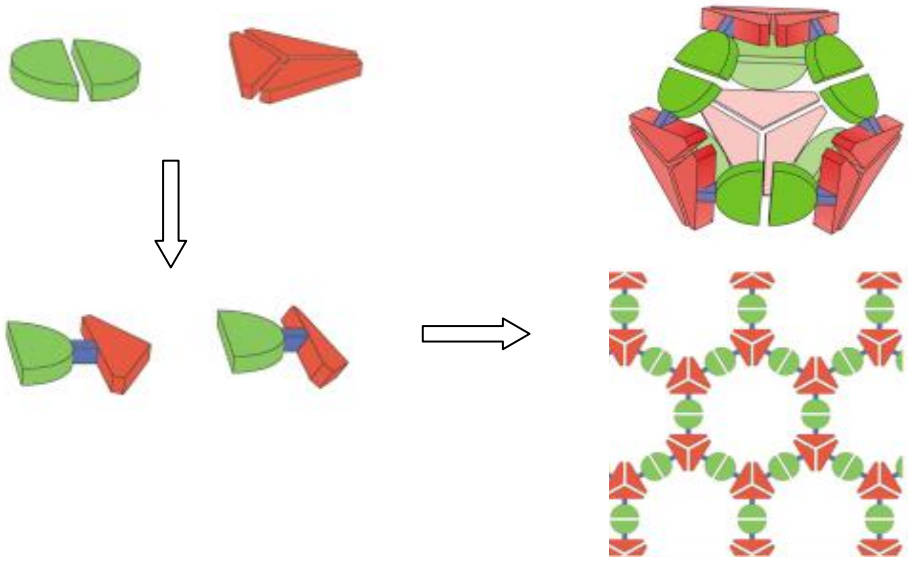


Fig 1.7 The strategy for creating flat networks as well as three-dimensional protein assemblies from fusion proteins of two natural oligomeric proteins.[42]

1.4 Ligands take hold of assembly

Just as DNA has its sequence recognition with base pairs and PPIs have evolved to specifically choose one interface out of a mixture, so too can protein-ligand interactions be precise and unambiguous. Protein-ligand interactions can thus be used as another means to create nanostructure assemblies. However, ligand mediated protein assembly is a concept that has been less widely researched and there are only a few examples of ligands being used to access large complexes. Here we define a ligand as a small molecule able to bind specifically to a larger protein. As with supramolecular chemistry, the same types of non-covalent interactions are relevant here, the ligand and protein binding site are now the recognition motifs.

Bringing two proteins together to initiate or increase their activity has been achieved using chemical inducers of dimerisation (CIDs). This dimerisation technique has been shown to be a powerful tool when investigating cellular processes.[48] Recent work from the Bertozzi group provides a nice example of CIDs at work. In this study the catalytic (cat) and localisation (loc) domains of a glycosyltransferase were split and fused to two different binding proteins.[49] The binding proteins FKBP and a

bacterial dihydrofolate reductase were used and a heterodivalent ligand was developed that could dimerise these proteins. Without this CID the cat domain was secreted from the cell but when the CID was employed it brought both halves together and so activated the full protein (Fig 1.8). Bringing two proteins together that were previously far apart is known as chemically induced proximity and has been shown to be useful not just in the dimerisation of proteins, but also in the formation of larger structures.[50]

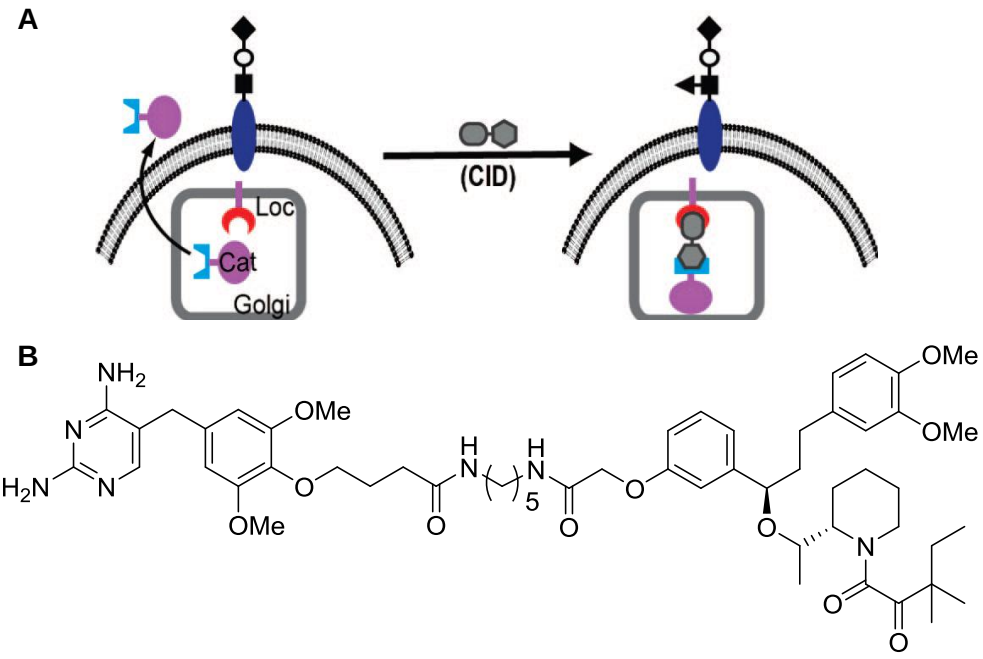
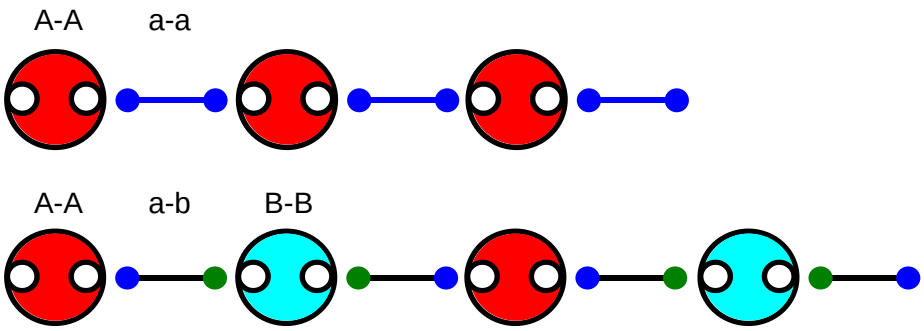


Fig 1.8 Activation of a glycotransferase using a CID. A) Only in the presence of the CID do the two domains associate activating the enzyme and in B) the structure of the CID is shown.[49]

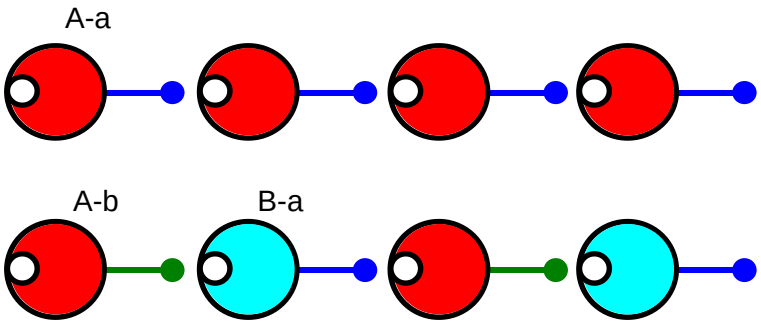
To create larger assemblies different tactics must be used and there are several different strategies for using protein-ligand interactions to create larger assemblies. The use of non-covalent interactions gives reversibility to the system and the possibility of self-correction which can aid the build up of ordered structures.

A homodimeric protein A-A with two binding sites could bind to a homodivalent ligand a-a which in turn would bind to another A-A and so on. Secondly homodimeric A-A could bind to B-B with a heterodivalent ligand a-b and a three membered system builds up (Scheme 1.1).



Scheme 1.1 Simple assembly systems of divalent ligands binding to proteins.

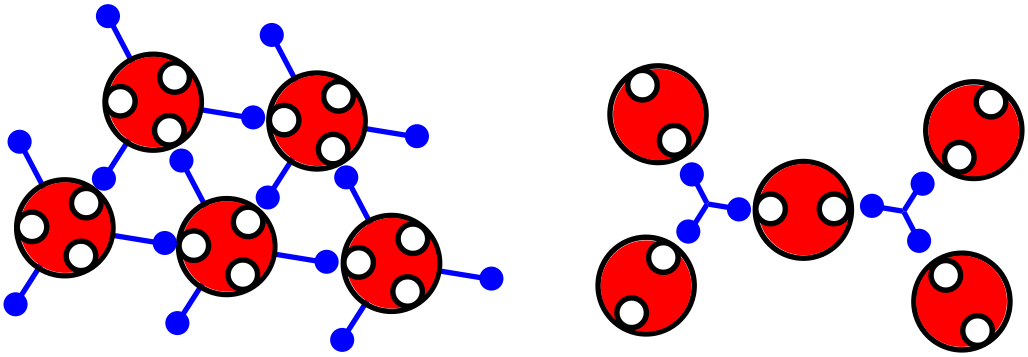
A different method would involve a covalent bond between protein A and ligand a (Scheme 1.2). This A-a unit then binds to a similar one and continues. Using the same technique but with a different ligand b attached to protein A, a dimeric system will be created.



Scheme 1.2 Assembly system for proteins with covalently attached ligands.

Structures can become increasingly complex when the valency is increased or when the valency of each protein and ligand are mixed (Scheme 1.3). Branching points can be created giving rise to two- and three-dimensional structures. With the right ligand design, proteins can be brought together in a well ordered manner.[51] Depending on the flexibility and arrangement of the ligand, proteins could be held in a specific way predesigned in the system to create discrete structures.

To covalently attach ligands to proteins many methods are available. Traditional methods such as targeting amine groups can modify multiple amino acid side chains and are thus not selective. Modifying cysteine residues can be a better strategy as it is more selective because cysteine is relatively rare in proteins. In recent years methods to specifically modify N- or C-termini have developed using native chemical ligation, sortase and oxime formation.[52, 53] Unnatural amino acids can also be introduced for modification with click chemistry.[54]



Scheme 1.3 Assembly systems with an increased valency and a mixed valency between ligands and proteins resulting in more complex structures.

1.4.1 Linear assemblies - it all starts to add up

Linear protein assemblies have been prepared by a number of groups. One example used a calix[4]arene core with carbohydrate ligands to polymerise the lectin LecA.[55] Two galactose moieties from the ligand bound to two sites on one side of LecA with each tetravalent ligand bringing two proteins together. AFM was performed on the structures and showed very straight assemblies with rare branching points attributed to conformational changes in the calixarene based ligand (Fig 1.9). Another group took streptavidin which has four binding sites.[56] The proteins were linked by bis-biotinylated ligands that bound to two binding sites on one side of streptavidin. These ligands were attached to terpyridine organised around an iron ion and the structures were deemed metal-organic protein networks.

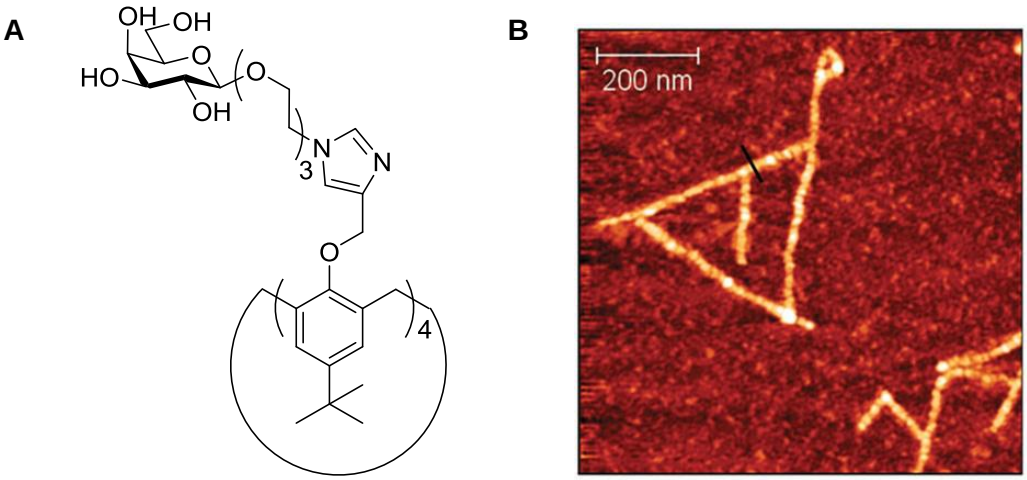


Fig 1.9 A) Representation of the galactose ligand. .B) An AFM image of the assembled protein polymers showing straight lines with occasional branching points.[55]

One of the first examples of controlled self-assembly by protein-ligand interactions of discrete nanostructures was by Carlson *et al*.[57] Building on their work with CIDs,

they expressed *E. coli* dihydrofolate reductase (DHFR) dimers connected by a flexible peptide linker with the aim of producing protein nanorings. Using a divalent methotrexate derivative (MTX_2-C_9) as a ligand to mediate the assembly, toroidal structures were created with diameters ranging from 8 to 20 nm depending on the length of the peptide linker (Fig 1.10).

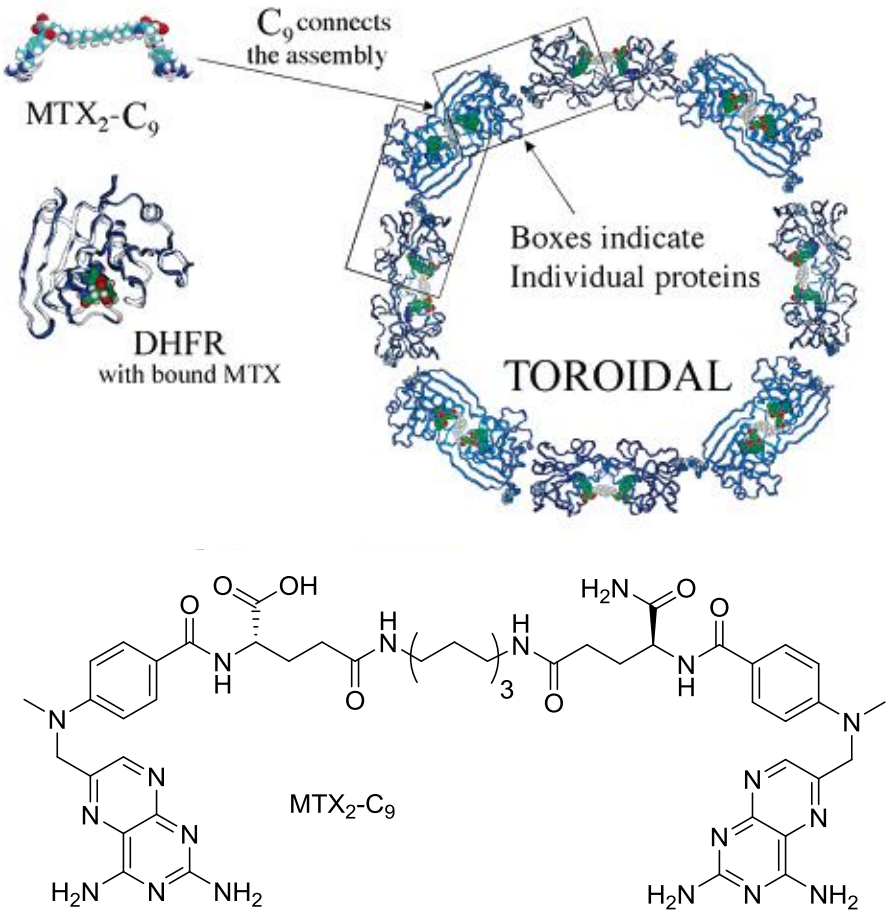


Fig 1.10 Representation of the cyclic structure created by DHFR dimer proteins linked with MTX_2-C_9.[57]

Other protein nanorings have been produced by the Meijer group.[58] Ribonuclease A was digested with a protease to yield the S-protein and the much smaller S-peptide which have a nanomolar affinity for each other. These fragments were then chemically joined with a flexible linker to form an A-a type system. The structures were studied at varying concentrations and it was found that the protein-peptide structures formed larger rings as the concentration was increased. The absence of linear constructs was attributed to the flexibility of the linkers and the propensity for cyclisation as the effective concentration of binding sites is increased when the proteins are brought together.

The Hayashi group have employed another type of ligand for their supramolecular protein polymers. Linear assemblies of a hemoprotein, cytochrome b562, conjugated to heme have been constructed.[59] A cysteine residue was introduced into the protein at a specific site, opposite to the heme binding pocket and then a heme modified with an iodoacetaamide group was attached. After acid denaturing and renaturing the protein to remove native heme, the protein-ligand interactions allowed polymerisation (Fig 1.11). Oligomers greater than 100-mers were observed using AFM, with lengths around 350 nm.

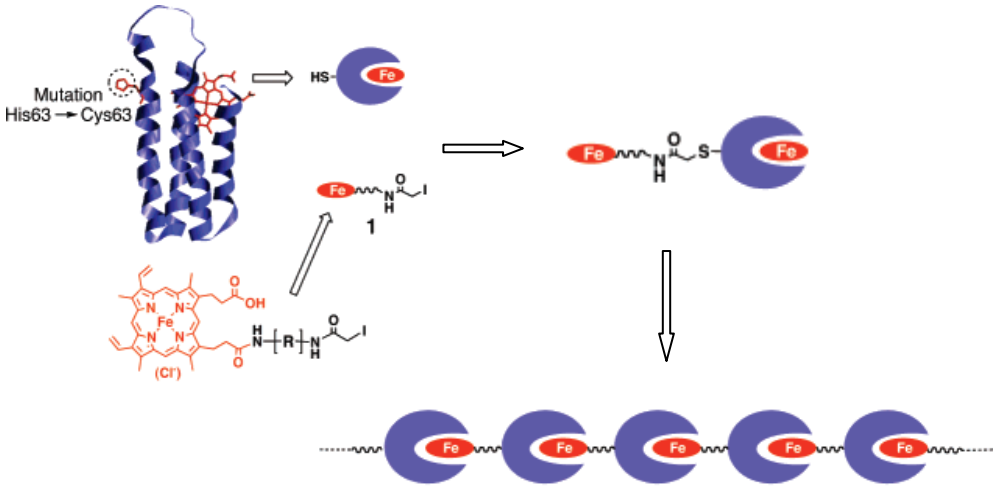


Fig 1.11 Representation showing heme attached to a heme binding protein and the one and two-dimensional hemoprotein networks that were formed.[59]

The polymers were further characterised by SEC and CD and it was shown that the system was under thermodynamic control due to the reversible associations.[60] Again it was found that the length of the assemblies was concentration dependent. This work was then extended to a more complex hemoprotein, myoglobin, to assess its physiochemical properties in the polymerised state.[61] Most recently the same group have created polymers of alternating myoglobin and streptavidin proteins.[62] For this they had to dimerise myoglobin by using a disulfide bond at the opposite side of the protein to the heme binding pocket, which enabled the protein to bind two heme units. When a heme-bisbiotin ligand was used, polymerisation could then occur (Fig 1.12). This is one of the only examples of a heterotropic linear protein assembly being formed.

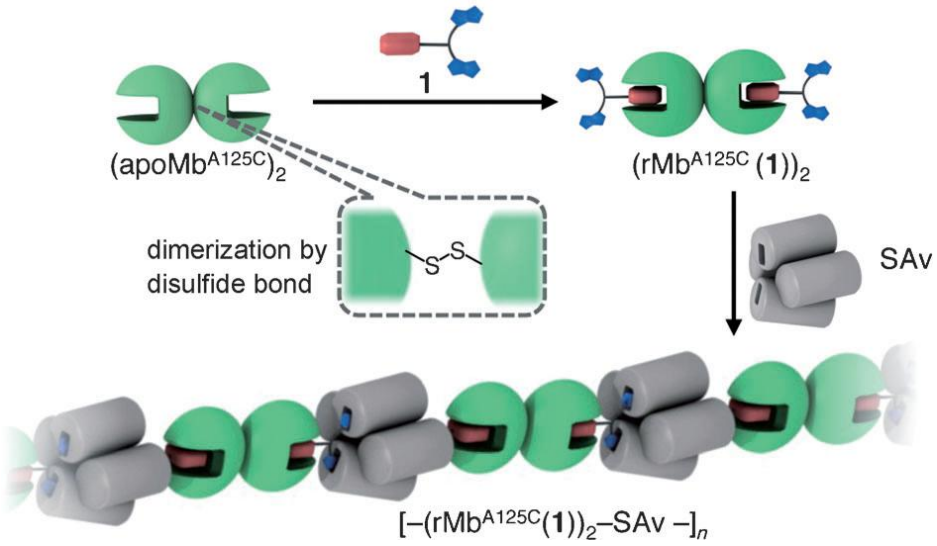


Fig 1.12 The assembly of the alternating copolymer made of streptavidin bound to dimerised myoglobin with heterovalent ligands.[62]

1.4.2 Making more contacts and building networks

There are relatively few examples in the literature of the use of protein-ligand interactions to create multi-dimensional nanoscale networks. The first such example was the production of a two-dimensional quadratic network by Ringler and Schulz.[63] A C_4-aldolase tetramer and streptavidin were combined to make the cross-linked network. Streptavidin has two biotin binding sites on each of its long faces, therefore the aldolase was mutated at selected sites to allow two biotin ligands to be tethered to each side of the protein, such that they had the correct spacing allowing a divalent interaction with streptavidin (Fig 1.13).

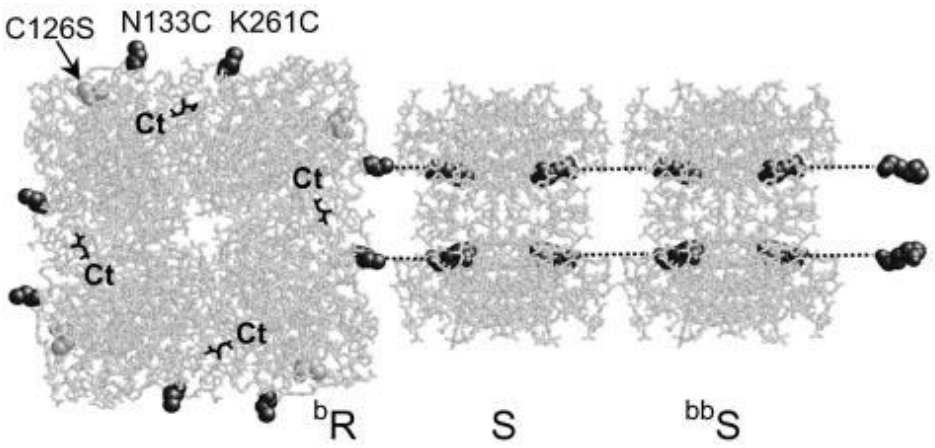


Fig 1.13 C_4-aldolase tetramer (^{b}R) with tethered biotin bound to streptavidin (S) which is then bound to another streptavidin (^{bb}S) when a divalent biotin ligand is added. Point mutations that allowed the ligand to be bound are highlighted on one subunit.[63]

As expected the streptavidin bound to the biotin-labelled aldolase, thus allowing the blocks to self-assemble in solution. Divalent-biotin ligands were also used to induce connections between the streptavidin and so a combination of covalently attached ligands and free ligands was used. This research showed order could be achieved with protein-ligand interactions and defined structures could be made. The reversibility of the networks was limited by the exceptional strength of the biotin-streptavidin interaction. This is one of the strongest protein-ligand interactions known and therefore there was no propensity for self-healing of irregular points in the network.

Other two-dimensional assemblies were created by Hayashi and co-workers using their hemoprotein polymers.[64] A trivalent heme ligand gave rise to branching points in the polymers and allowed the dimensionality of the networks to be controlled by altering the ratio of the heme-conjugated proteins to the branching molecules. Another two-dimensional network was produced by the same group using a combination of the same hemoproteins with gold nanoparticles as the branching points (Fig 1.14).[65] The heterotropic structures were visualised by TEM and AFM, which showed that the nanoparticles were assembled by the protein-ligand interactions.

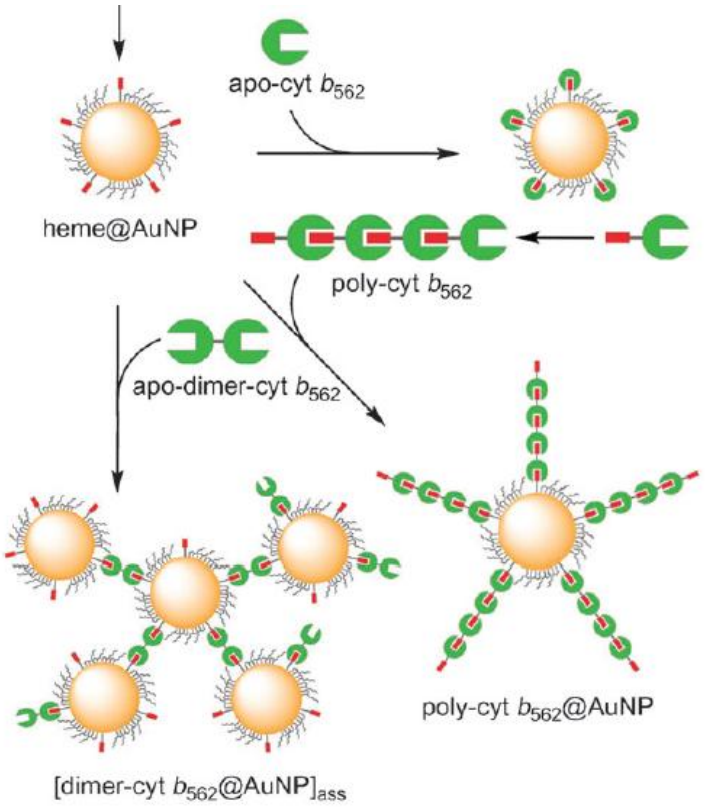


Fig 1.14 The assembly of hemoprotein-nanoparticle structures.[65]

1.4.3 To the third dimension and beyond

Each increase in dimensionality brings new challenges and greater control is always needed. One study by Mori *et al.* found that their protein assemblies, formed from streptavidin binding to a biotin labelled phosphatise, created uncontrolled aggregates and precipitated. This was due to too many ligands per receptor being used and it was only when they had the correct ratio that discrete protein structures started to form.[66] Another investigation by Ma and Bong saw streptavidin proteins brought together by mixed bivalent ligands.[67] The ligands had a biotin moiety at one end for streptavidin binding and then either a Tris-functionalized cyanuric acid (TCA) or melamine (TM) group at the other end. These two groups have a strong affinity for each other and so when mixed together with streptavidin, protein-ligand and supramolecular interactions began to build up large aggregates (Fig 1.15). The hydrogen bonding motifs could be dissociated thermally, breaking up the aggregates. The system was then cooled before being allowed to form again. The size and shape of the networks was, however, uncontrolled.

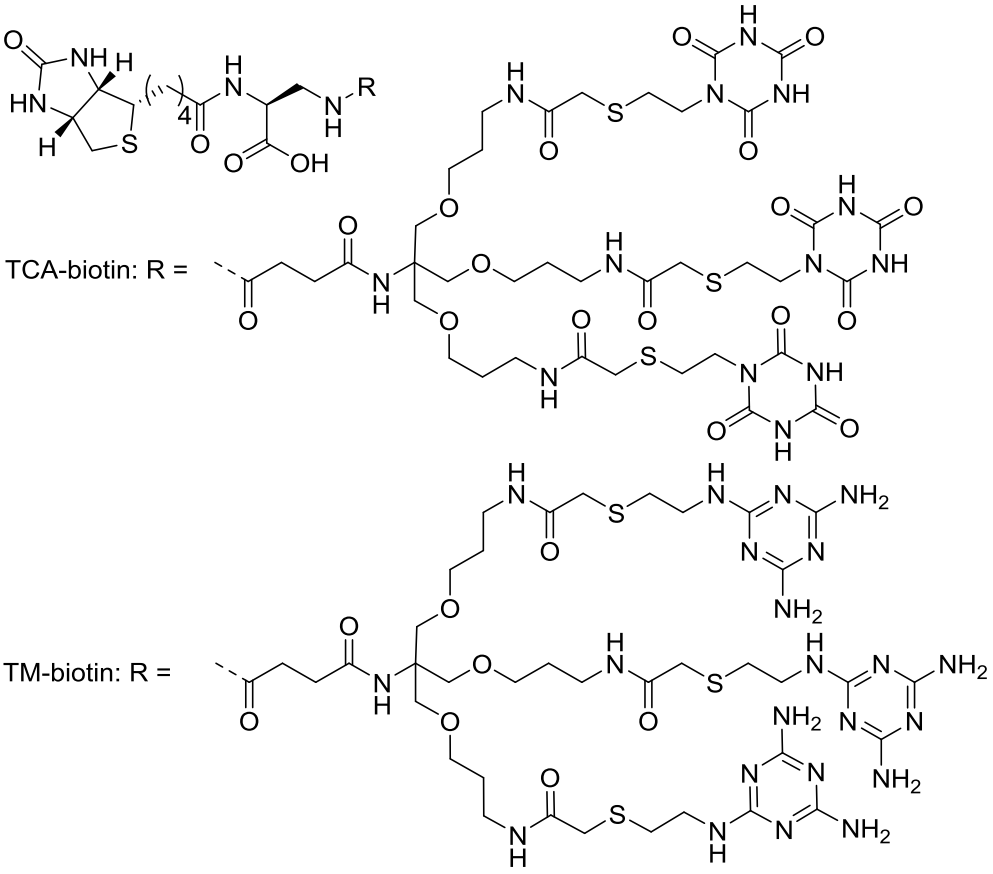


Fig 1.15 The structure of the complementary tris-functionalized cyanuric acid (TCA) and melamine (TM) ligands. When biotin was attached to the ligands, self-assembled aggregates with streptavidin were formed.

The first example of the controlled assembly of a predesigned three-dimensional lattice was shown by Dotan *et al.* with the cross-linking of the tetrameric lectin concanavalin A.[68] This work used molecular modelling to guide the design of the divalent ligands needed to create the self-assembling structures. When the divalent mannose ligand was added at 2:1 stoichiometry to the proteins, the tetrahedral lectins became non-covalently cross-linked into a diamond-like structure that crystallised out of solution (Fig 1.16). These protein crystals grew up to 100 nm in size, were highly stable to pH change across the range 4.5-8.0 and were not disassembled by addition of a competing mannose ligand, even up to a concentration of 1.5 M. This example provided the first demonstration of the great stability non-covalent protein assemblies can achieve and the assistance that computational design can offer. However, the lattices assembled were insoluble and their size could not be controlled with great precision.

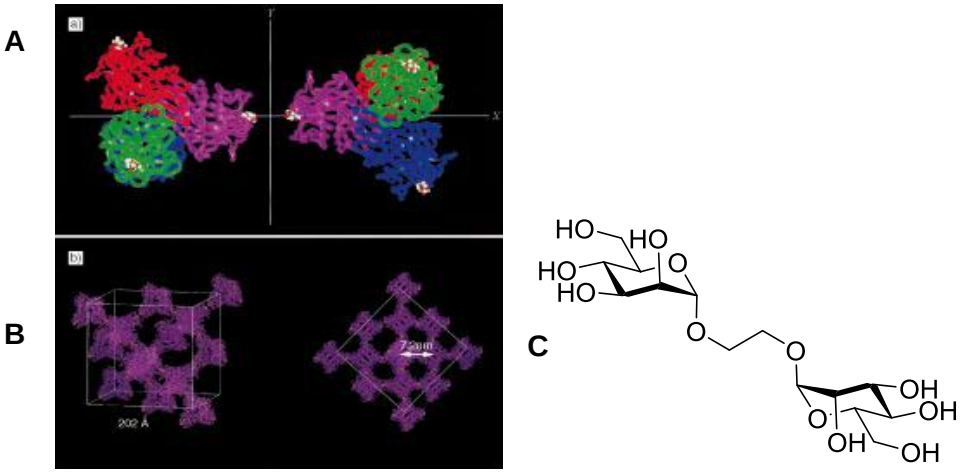


Fig 1.16 A) model of two tetrahedral lectins in a staggered orientation. B) Two views of the predicted diamond structured lattice. C) Structure of the divalent mannose ligand.[68]

A discrete three-dimensional structure was realised by the Whitesides group. They presented work in which a trivalent hapten ligand bound divalent antibodies in a thermodynamically and kinetically stable aggregate.[69] The reversibility of non-covalent interactions meant that the initially disordered system could eventually rearrange into stable structures leading to three IgG proteins being bound to two trivalent ligands (Fig 1.17). These structures again showed the enhanced stability that can be achieved through multivalency, as the complex had a 225 fold increase in stability compared to a single bound ligand.

Whilst there are a few examples of discrete three-dimensional structures formed by protein-ligand interactions, nobody has yet achieved the sophistication and

standardisation that has been seen with DNA. Nevertheless, these studies demonstrate the intricacies that must be fully understood in the design process before being able to manipulate protein self-assembly into the exact, organised, unnatural configuration that is desired. Another important point shown here is the inherent stability that multivalent assemblies possess and this desirable property makes these types of assembly attractive for many applications in nanotechnology.

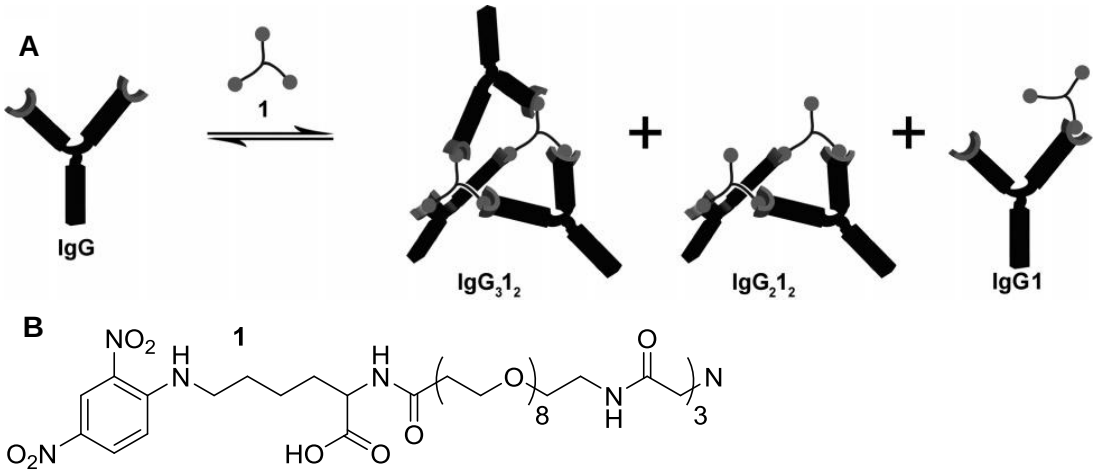


Fig 1.17 A bivalent antibody (IgG) and a trivalent hapten ligand (1) form stable aggregates in a 3:2 stoichiometry.[69]

1.4.4 The promise of greater things

Between the simple binding motifs of supramolecular chemistry and the complexity of biological systems, a common ground is being formed. The rise of synthetic biology has seen researchers playing with natural systems to bring their own ideas to life. By combining the strategies of supramolecular chemists with the building blocks supplied by nature we will create opportunities to shape new devices and structures inspired by nature but of our own design and for our own use.

Self-assembly is an important concept to be used for the goal of creating large ordered structures. Using carefully designed small molecules as the mortar to hold natural protein bricks together, has the potential to yield limitless extraordinary architectures. Nature's toolkit has been thoroughly ransacked by scientists looking to create more sophisticated structures with increasing precision and ease. Thanks to years of scientific advances there is plenty to build upon but more importantly there is still plenty to achieve.

An Introduction

Part II:
The power of multivalency

A review based on the following section has been published: T. R. Branson and W. B. Turnbull, “Bacterial toxin inhibitors based on multivalent scaffolds, *Chem. Soc. Rev.,* 2013, 42, 4613”.[70]

1.5 The power of multivalency

As seen with many of the assemblies discussed so far, it is often the combined effect of more than one interaction that gives strength and stability to the structures. This cooperative binding effect is known as multivalency.[71-73] Supramolecular chemistry takes full advantage of multivalency as individual non-covalent interactions can be quite weak but by combining many interactions, an overall more powerful force may be achieved. When ligands are arranged on a multivalent scaffold, their effective concentration increases and prearrangement can aid the interactions.[51]

Noncovalent multivalent interactions are prevalent in many biological processes and lead to the formation of countless larger complex structures. Improving our understanding of multivalency can give us a better insight into natural processes.[74] Carbohydrate interactions throughout nature are typically weak and therefore many biological processes use multivalency to overcome this problem.[75] Bacterial toxins are a good example of where the weak carbohydrate interactions and binding efficiency are greatly enhanced with multivalency.[73, 76] Work undertaken to find inhibitors for the toxins has tended to focus on multivalent binding. Some multivalent ligands have even led to the creation of protein assemblies which, whilst not being the ultimate goal of the projects, opened up other interesting possibilities.

1.5.1 Bacterial toxin inhibition with multivalent scaffolds

A significant portion of diarrhoeal diseases can be attributed to bacteria that produce protein toxins. These cases continue to pose a serious threat to human health, with an estimated two million deaths per annum, most of which are in children under five years old.[77] The most widely studied of these toxins are cholera toxin (CT), the closely related *E. coli* heat-labile toxin (LT)[78, 79] and shiga-like toxin (Stx; also known as verotoxin, VT).[80] Together, they belong to a family of AB_5 toxins that are comprised of a single toxic A-subunit associated with a non-toxic B-pentamer. The B-pentamer is a carbohydrate-binding protein enabling the toxin to enter cells (Fig 1.18).[81]

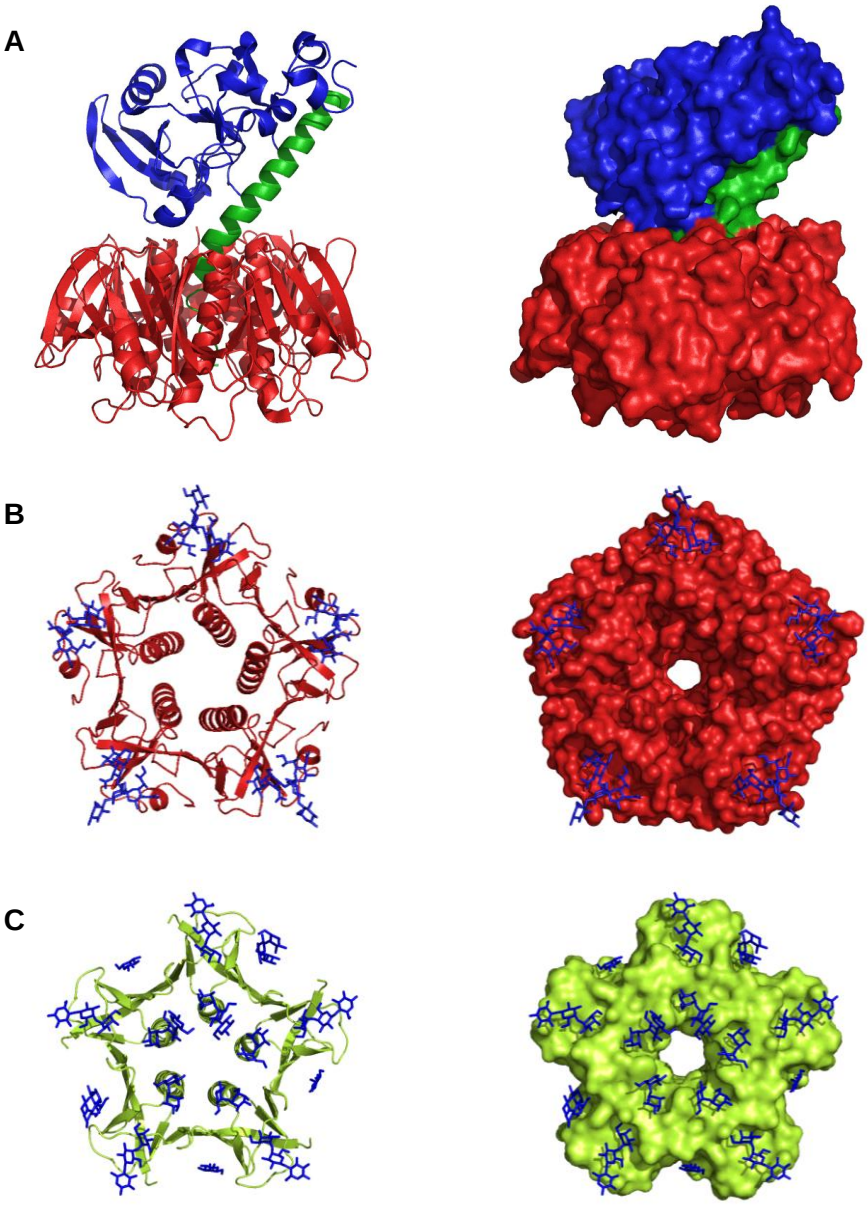


Fig 1.18 Cartoon and surface representations of bacterial toxins. (A) Cholera toxin AB_5 protein with CTA1 (blue), CTA2 (green) and CTB (red) (1S5F.pdb). (B) Cholera toxin B-pentamer (CTB) with GM1 ligand **1** depicted as a stick structure (3CHB.pdb); (C) Shiga-like toxin B-pentamer (VTB/StxB) with Gb_3 ligand **2** depicted as a stick structure (1BOS.pdb).

Following initial adhesion of the B-pentamer to specific glycolipids that are present on the surface of epithelial cells that line the intestine, the toxins enter the cells by endocytosis. They are then transported through the cell to the endoplasmic reticulum where the toxic A-subunit is released into the cytosol where it has its cytotoxic effect.[78, 82] In the case of CT and LT, the toxin ADP-ribosylates the Gs_α protein which leads to a rise in cAMP concentration in the cell leading to a complex series of events that result in release of water into the intestine. In the case of Stx,

the toxic A-subunit is an N-glycosidase that removes purine bases from ribosomal RNA, thus inhibiting protein synthesis and causing cell death.[83]

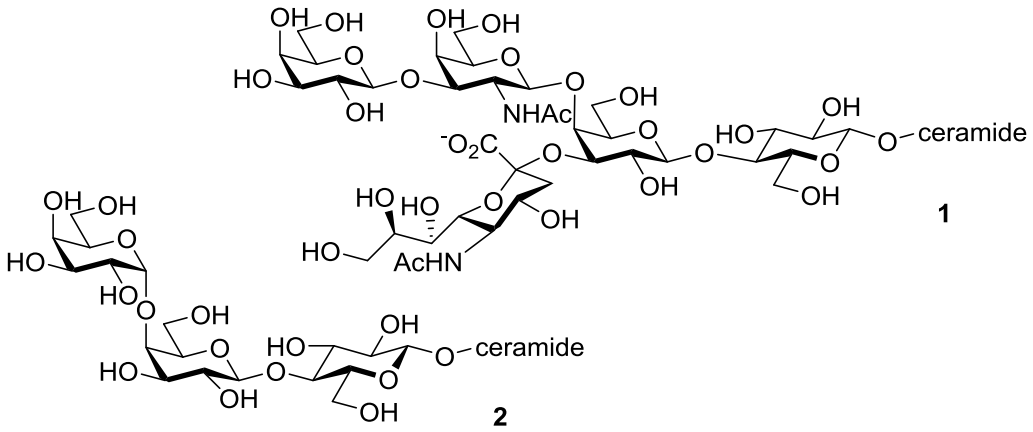


Fig 1.19 GM1 ligand **1** for CTB and Gb_3 ligand **2** for StxB.

The B-subunits of cholera and heat-labile toxins (CTB and LTB) are 80% identical, but share essentially no sequence homology with the shiga-like toxin B-subunit (StxB).[81] This observation is surprising as all three toxins have the same protein fold and have evolved to perform the same function, i.e., to enable toxin endocytosis by binding to cell surface glycolipids. This function is facilitated by having their glycolipid binding sites arranged on the same face of the protein so that they can engage multiple copies of the carbohydrate ligands at the same time.

The similarities and differences in protein sequences of CTB, LTB and StxB are reflected in their carbohydrate-binding specificities. CTB and LTB both recognise ganglioside GM1 **1**,[84] while StxB binds to globotriaosyl ceramide (Gb_3) **2** (Fig 1.19).[85] However, CTB and LTB are not identical, and certain variants of these proteins can differentiate between blood group oligosaccharides in a secondary binding site on the circumference of the pentamer.[86-88] The most striking difference between CTB/LTB and StxB lies in the affinities for their glycolipid ligands. The CTB/LTB-GM1 interaction is among the highest affinity protein-carbohydrate interactions known (K_d = 10-40 nM for a monovalent interaction),[89] whereas StxB binds only weakly to individual Gb_3 oligosaccharides (K_d = 1 mM).[90] Therefore, StxB relies on multivalency (i.e., simultaneous binding to multiple copies of its ligand) to achieve a functionally useful interaction with a cell surface. Up to 15 copies of the Gb_3 ligand can bind simultaneously to a StxB pentamer leading to a sub-nM dissociation constant.[85] Although CTB/LTB can bind tightly to an individual GM1 molecule, multivalent interactions are still important in facilitating endocytosis.[91]

1.5.2 Bacterial toxin inhibitors

As protein-carbohydrate interactions are an essential prerequisite for cell entry and toxin activity, the development of inhibitors for these interactions has attracted much interest over recent years. In general it is very challenging to make low weight inhibitors of carbohydrate-binding proteins as the binding sites are typically shallow and highly solvated. In the case of CTB/LTB, substantial changes to the GM1 structure usually result in a loss of affinity (Fig 1.20).[92] For example, methyl β-galactoside **3** has a 15 mM dissociation constant;[89] GM1 mimic **4** has a K_d value of 10 mM;[93] and aromatic α-galactosides **5** and **6** have K_d values of 125 and 12 mM, respectively.[94] As even the best of these compounds still bind around 1000 times more weakly than GM1 **1**, researchers have instead focussed their attention on multivalent inhibitors that can engage multiple carbohydrate binding sites simultaneously.[95]

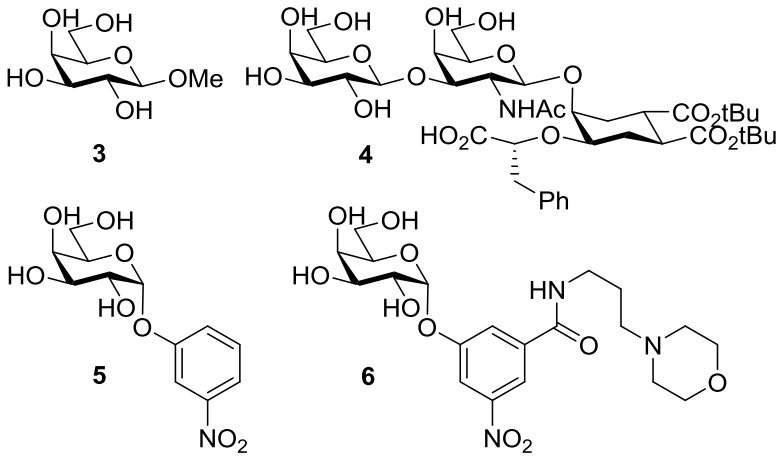


Fig 1.20 Monovalent galactoside inhibitors for CTB/LTB

Multivalent binding can be accomplished in many ways and so there are a wide range of designs employed as inhibitors. Glycopolymers, glycodendrimers, tailored glycoclusters and inhibitors exploiting templated assembly have all been developed to bind the bacterial toxins (Fig 1.21).

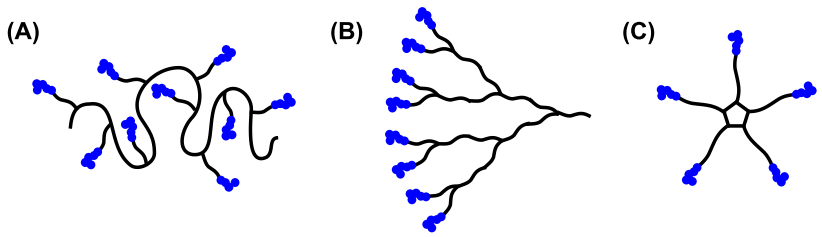


Fig 1.21 Multivalent bacterial toxin inhibitor designs. (A) glycopolymer, (B) glycodendrimer and (C) glycocluster.

1.5.3 Stringing together the inhibitors

To take advantage of the benefits of multivalency, a suitable method must be found to connect multiple ligands together. The most simple way to organise multiple copies of a ligand is to string them out along a chain. Polymers can be used to achieve this arrangement; their relative ease of synthesis and variability of structure and length make them ideal for use as a general architecture for multivalent presentation.

The benefits of using multivalent inhibitors for bacterial toxins were shown in an early example by Schengrund and Ringler.[96] In this study, reductive amination of the GM1 oligosaccharide (GM1os) to the free amino groups on the polylysine scaffold resulted in polymer **7** (Fig 1.22) with an average of eight oligosaccharides. It was found that this polymer was 1000-fold more effective than GM1os for inhibiting cholera toxin from adhering to GM1 coated plates.

Gb_3-polyacrylamide conjugates have been used to neutralise Stx-1 in human ACHN cells by Dohi *et al*.[97] They concluded that the clustering effect of multiple ligands presented on polymer **8** must be the reason for successful inhibition as the individual affinity of one copy of the Gb_3 ligand was too low to have any effect in cells. Analogous polyacrylamide glycopolymers having varying degrees of substitution with the Gb_3 trisaccharide were reported by Gargano *et al*.[98] They found that polymer **9** gave 5000 times enhancement of the inhibition of Stx-1 over the monomeric carbohydrate. The degree of ligand substitution (between 10% and 30%) did not have a substantial effect on inhibition. In contrast, other studies with similar polymers have shown that the density of the carbohydrate ligands on the polymer chain can affect the binding affinity to the Stx-2 isoform of the toxin, as can the length of spacer from the backbone chain.[99, 100] For example, a higher density of Gb_3 along polymer chain **10** provided higher affinity binding to Stx-2 but had little effect on binding to Stx-1.[99] The density-dependence of binding to Stx-2 is significant as the Stx-2 isoform of the toxin is considered to have greater clinical importance than Stx-1.

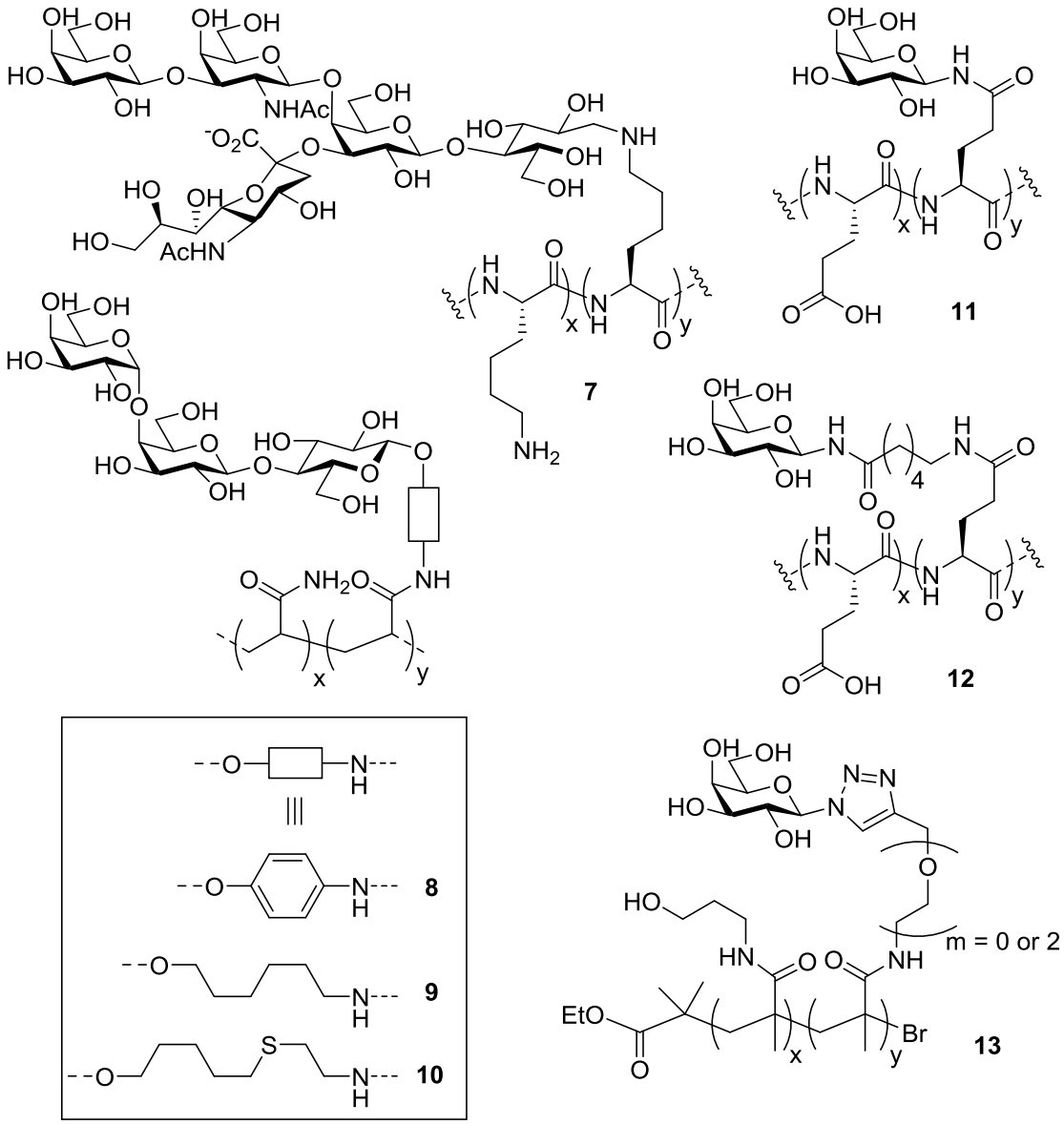


Fig 1.22 Glycopolymer inhibitors of bacterial toxins.

The Kiick group have extensively investigated the effect on inhibition of carbohydrate density; the specific distance between carbohydrates on a polypeptide chain; and the spacer length from the polypeptide backbone. Using galactosyl polymers **11** and **12** as inhibitors for CTB, it was found that as the density of ligands increases, so the inhibitory effect decreases.[101] These results are in contrast to the previous findings by Watanabe *et al.* with Stx-2.[99] Kiick and co-workers concluded that the best inhibition was achieved when the spacing between ligands on the polymer chain was matched to the distance between binding pockets of the toxin; this distance being 35 Å for CTB. When the density was sufficiently high that spacing between carbohydrates was smaller than the distance between binding sites, steric hindrance created by unbound ligands decreased overall binding. The spacer length from the polymer backbone chain told a similar story.[101] Inhibition

was greater when the spacer matched the natural ligand length of GM1 (16 Å). The longer ligand **12** could therefore fully penetrate the binding site as opposed to a reduced accessibility when a shorter spacer **11** was used. Matching the ligand spacing to the binding site dimensions improves the effectiveness of the ligands, giving rise to high inhibition with these well designed polymers.

Richards *et al.* had similar findings with their polymethacrylamide polymers **13** presenting galactosyl ligands for CTB binding.[102] This study went further to suggest that there is a varying relationship between ligand density and inhibition. At a high density of carbohydrates along the polymer chain, a high rate of statistical rebinding is achieved giving good inhibition. At only 10% density there is lower steric hindrance and a better fit to the binding site. Similar inhibition results were found per galactose moiety for polymers substituted 100% and 10% with ligand groups. Between these values, inhibition decreases as the balance between the competing effects worsens.

Further studies were performed by the Kiick group on the composition of polypeptide backbones. They found that random coil backbones were better suited than those with restricted alpha helical conformations, as flexibility of the polymers allows more accessible ligand groups.[103] They showed that electrostatic repulsion gave a larger hydrodynamic radius for a polymer with negatively charged residues than for a neutral chain.[104] The resulting larger chain dimensions led to better inhibition, again due to the accessibility of the ligands. The charge of the peptide backbone was also shown to be of importance as negative glutamic acid residues aided inhibition, neutral glycine residues were acceptable, but positively charged lysine residues were detrimental to the inhibitory properties of the glycopolymers.[105] There are positively charged residues present on the surface of CTB around the binding pocket and so complementing these charges should improve binding. A negatively charged backbone gave an IC_{50} value almost three times better than that for a neutral backbone, whereas introducing positive charges led to an IC_{50} value four times worse than a neutral backbone.

Bundle and co-workers have reported a strategy to identify optimal ligand groups for multivalent display by synthesising libraries of glycomimetics on polyacrylamide or aminated dextran backbones.[106] They found that screening ligand groups for CTB inhibition in a multivalent format made it easier to identify optimal ligands than if they had studied the analogous low affinity monovalent compounds.

Without prior knowledge of the valency or structure of the target protein, polymers are a good starting point for building multivalent inhibitors. However, defining the distance between ligand groups and the length of the spacer to the polymer backbone can result in much improved binding when these parameters can be matched with the dimensions of the protein binding sites. While linear polymers with pendant carbohydrate groups are relatively easy to make, they are also frequently heterogeneous in their distribution of the ligand groups. Therefore only a subset of ligand groups will be in the optimal arrangement to bind to the target protein. Multivalent scaffolds that provide greater homogeneity or restrict the ligand groups into a favourable orientation for binding could potentially provide improved inhibition.

1.5.4 Branching out into glycodendrimers

Glycodendrimers are monodisperse, branched tree-like structures that, in principle, combine the advantages of homogeneous small molecular inhibitors with the dimensions and high valencies of glycopolymers.[107, 108] Schengrund's group was the first to use PAMAM **14** and poly(propylene imine) **15** dendrimers for CT inhibition, building on their knowledge of multivalent polymers.[109, 110] These dendrimers (Fig 1.23), with an average of seven GM1 ligands attached, provided inhibition against CT binding to cell surfaces. The results were found to be similar to those for their polymers and provided 1000-fold increased inhibition relative to GM1 oligosaccharide.

The Pieters group have created a set of dendrimers, synthesised by a convergent approach, with a variety of carbohydrate end groups as CTB inhibitors (Fig 1.24).[111] Their first scaffolds were prepared with two, four and eight arms terminating in lactose sugars (**16a**-**c**).[112] As expected, it was found that the multivalent dendrimers gave an increase in inhibition with the octavalent ligand having the strongest binding with a K_d of 33 μM as measured by a fluorescence titration assay. More surprisingly however, were the results for the monovalent and divalent ligands. The monovalent lactosyl head group had a K_d of 248 μM, 73 times more potent than simple lactose. This increase in affinity could partially be explained by additional interactions being created between the aglycone and the protein. The divalent ligand **16a** gave an affinity only slightly stronger and it was proposed that this was because the linker length was not sufficient to allow both carbohydrates to bind simultaneously to adjacent subunits. It is therefore important

to note that binding site spacing needs to be taken into account when designing dendrimers for multivalent inhibition, as with glycopolymers.

A more systematic study then followed from the same group using the same dendrimer scaffold as before, but with increased linker lengths to improve the reach of their ligands. SPR binding studies demonstrated that dendrimers **17a-c** bearing an (*R*)-lactic GM1 mimic all achieved improved binding.[113] The monovalent ligand had an IC_{50} value of 97 µM, while that for divalent compound **17a** was 13 µM. The tetra- and octavalent inhibitors **17b** and **17c** both gave IC_{50} = 0.5 µM, however, the detection limit for the assay had been reached. An ELISA assay confirmed the octavalent dendrimer to be the most potent.

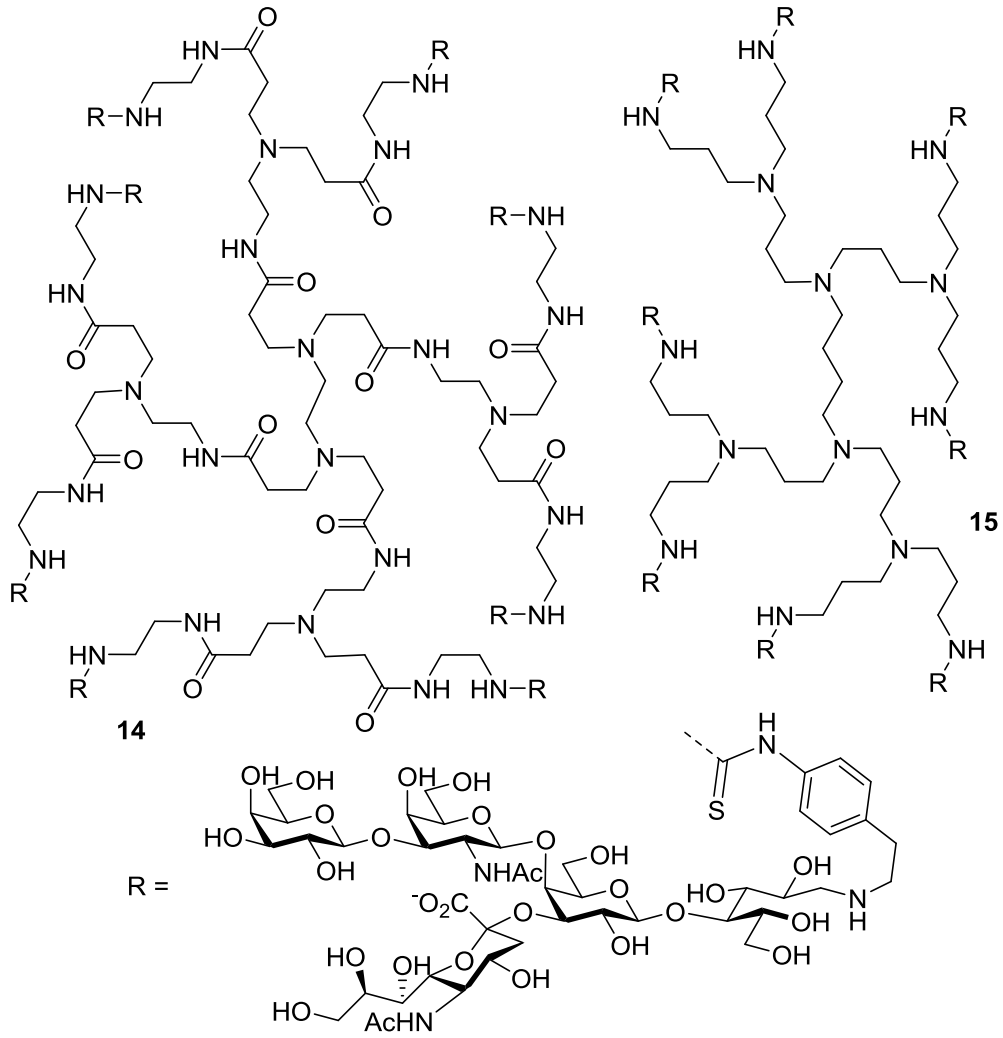


Fig 1.23 PAMAM and poly(propylene imine) glycodendrimers.

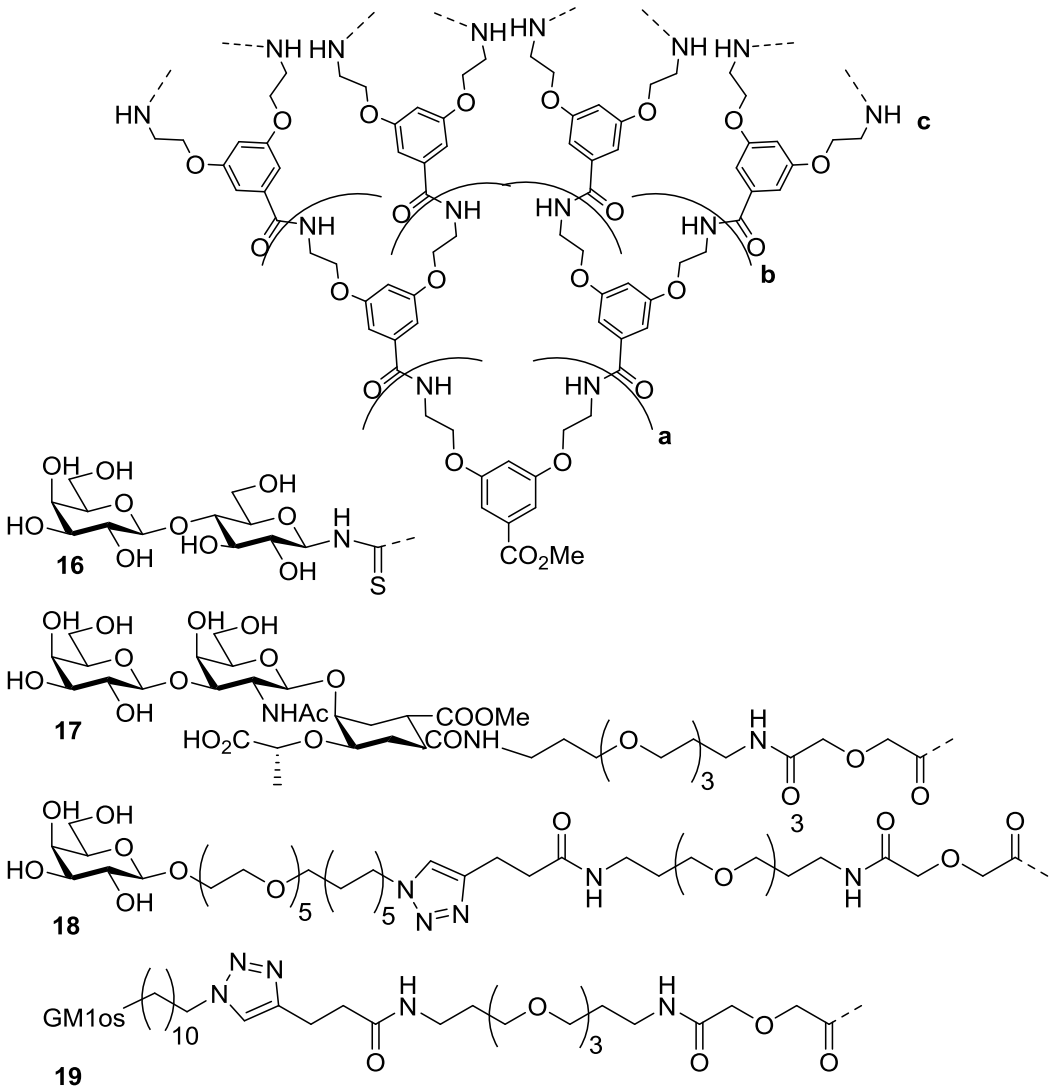


Fig 1.24 Glycodendrimer inhibitors synthesised by Pieters and co-workers.

Galactose ligands on a long PEG spacer were also studied.[114] The PEG spacer was introduced to mimic the hydrophilicity and lipophilicity of GM1 and was introduced onto the dendrimers via “click” chemistry. Divalent compound **18a** had an IC_{50} value of 130 µM in an ELISA assay, compared to 80 mM for the monovalent ligand. The best dendrimer was found to give similar activity to the natural GM1 oligosaccharide ligand (GM1os). This was octavalent ligand **18c** which had an IC_{50} of 12µM, i.e., 2500 times more potent per sugar group than the monovalent galactosyl compound. The strongest inhibitory potency was found when GM1 oligosaccharides were attached to dendritic scaffolds **a-c**.[115] Octavalent ligand **19c** was found to have an IC_{50} of 50 pM, almost 50,000 times stronger per carbohydrate group than the GM1 oligosaccharide.

Sisu *et al.* used a combination of analytical ultracentrifugation, dynamic light scattering and atomic force spectroscopy to show that Pieters’ di- and tetravalent

dendrimers (**19a** and **19b**) inhibited toxin adhesion through an aggregation mechanism.[116] Some dimerisation of LTB was shown to be induced by the divalent compounds but this was not seen for the tetravalent compound. While the multivalent ligands could have potentially formed discrete protein assemblies, in this case uncontrolled aggregation was observed. It was proposed that the random aggregation occurred as a consequence of a mismatch between the valencies of the ligands and the toxin.

The best glycodendrimers are better than the best of the linear polymers. However, the preference for having long flexible linkers to the ligand groups indicates that the densely packed globular shape of higher generation dendrimers is not desirable for inhibition. The glycodendrimers synthesised to date do not take advantage of the 5-fold symmetry of bacterial toxins, which suggests that further advantage could be gained through the design of tailor-made inhibitors.

1.5.5 Reaching for the stars

Concurrent independent studies by Bundle[117] and Fan[118] led to the creation of star-shaped inhibitors that take advantage of the pentagonal symmetry of bacterial toxins. Theoretically, a five armed star could sit on a toxin pentamer with each arm reaching out to a separate binding site. While the Bundle group focussed on inhibitors of Stx, Fan, Hol and co-workers designed inhibitors for LT and CT.

The Fan group attached galactose onto the arms of a scaffold radiating from a pentacyclen core (Fig 1.25).[118] A modular approach was adopted to the design and synthesis of the linker arms so that they could be varied in length. The use of squaric acid diesters as linking agents allowed two different amines to be coupled consecutively in an efficient manner. The effective length of long flexible linkers is actually much shorter than the extended conformation and it was found that the inhibitor was most potent when the linkers were of a size that presented the carbohydrate ligands efficiently at the binding sites. This finding confirmed the concept proposed previously by Kramer and Karpen that matching the effective dimensions of a ligand with flexible linkers and the binding site distribution, gives the greatest binding affinity.[119] The multivalent effect of the optimal pentavalent scaffold **20** (n = 4) gave a 100,000-fold enhancement in the IC_{50} value over a monovalent galactoside. It was also found that the pentavalent ligands formed 1:1 complexes with the toxins and no large scale aggregates were observed.

Higher affinity inhibitors were produced by incorporating m-nitrophenyl α-galactoside as the ligand group, and introducing a guanidine-bridged water soluble linker. The resulting compound **21** had an IC_{50} of 6 nM, which was about 3 times more potent than the GM1 oligosaccharide.[120]

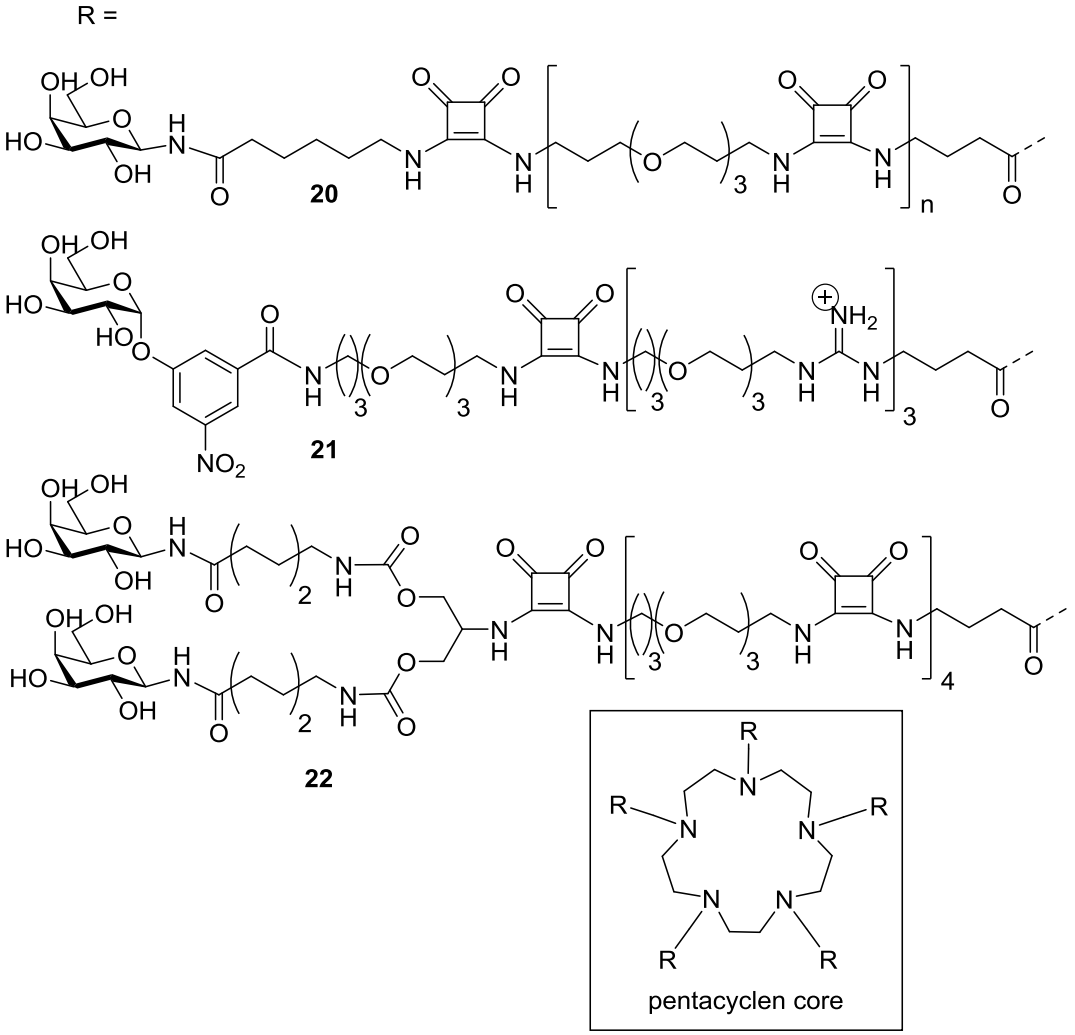


Fig 1.25 Star-shaped LT/CT inhibitors reported by Fan, Hol and co-workers.

Recently a study produced pentavalent ligands for CTB based on a calix[5]arene core.[121] GM1os and galactosyl ligands attached to the core structure were analysed by ELISA experiments which showed 450 pM and >1 mM IC_{50} values for the respective pentavalent ligands. The result for the galactosyl ligand was surprising compared to the similar Fan structures that showed strong inhibition with galactosyl ligands. This discrepancy may have been due to the lower flexibility of the calixarene or some supramolecular aggregation of the core structure. The pentavalent GM1os ligand showed an impressive 20,000 fold increase in inhibitory potency per arm compared to the monovalent GM1os ligand. The limit of detection for the ELISA may, however, have been reached. The assay went down to where

the concentration of inhibitor was lower than that of the CTB. If 1:1 binding was achieved, then at these lower concentrations it would not be possible to detect inhibition.

A concurrent study produced star-shaped inhibitors based on a corannulene core.[122] GM1os ligands were attached to the pentavalent core using microwave assisted CuAAC, “click”, chemistry and ELISA experiments were performed. The corannulene core structure exhibited an IC_{50} value of 5 nM after optimisation of the PEG chain arms. However, structures modified with simple galactose ligands failed to show any inhibition, which was thought to be due to the supramolecular aggregation of the corannulene core. This supramolecular assembly may also have contributed to the IC_{50} value being lower than what might have been expected from similar pentameric structures. Despite the tailor-made nature of the pentavalent structures, they are yet to reach the inhibitory power of the dendrimer structures with a higher valency. Octavalent dendrimer **19c** remains the most potent CTB inhibitor which may in part be simply due to unoptimised assay results.

1.5.6 Targeting multiple copies of the toxins

When using multivalent ligands to inhibit protein binding it has been seen that tailoring the design of the ligand can greatly improve its binding strength. However, it could prove more efficient to be able to complex more than one toxin protein with each inhibitor. Targeting multiple proteins at once can be achieved by using chemical-inducers of dimerisation (CIDs).[48] If the multivalency strategy were to be combined with CIDs then self-templating, self-assembled structures could potentially produce better inhibition of the toxins.

The Bundle group was the first to make a large step towards this goal with their Starfish ligand **23** (Fig 1.26).[117] Two carbohydrate ligands were arranged at the end of each arm of a pentavalent glucose core. Starfish **23** gave over a million fold increase in inhibition relative to the monovalent Gb_3 ligand. The decavalent structure was originally designed to bind to two separate binding sites on each sub-unit of the shiga-like toxin. However, a crystal structure of the complex revealed that Starfish bound only to one carbohydrate site in each of the toxin subunits, while using the remaining five Gb_3 groups to complex binding sites on a second StxB pentamer, thus holding two pentamers together. These unexpected dimeric structures were proposed to be thermodynamically favourable and were thought to avoid potential strain that would arise if the ligands were to chelate two binding sites

within a single StxB protomer. A later study indicated that divalent binding to a single StxB protomer was accompanied by a significant entropic penalty for restricting the dynamics of the protein.[123] Improving on Starfish, was another inhibitor from the same group nicknamed Daisy **24**.[124] This decavalent structure used the same Gb_3 trisaccharide but linked through the reducing terminus rather than two position of the central sugar of Gb_3. The new ligand with slightly longer linker spacing between the oligosaccharide groups was found to better protect mice against Stx-1 and Stx-2.

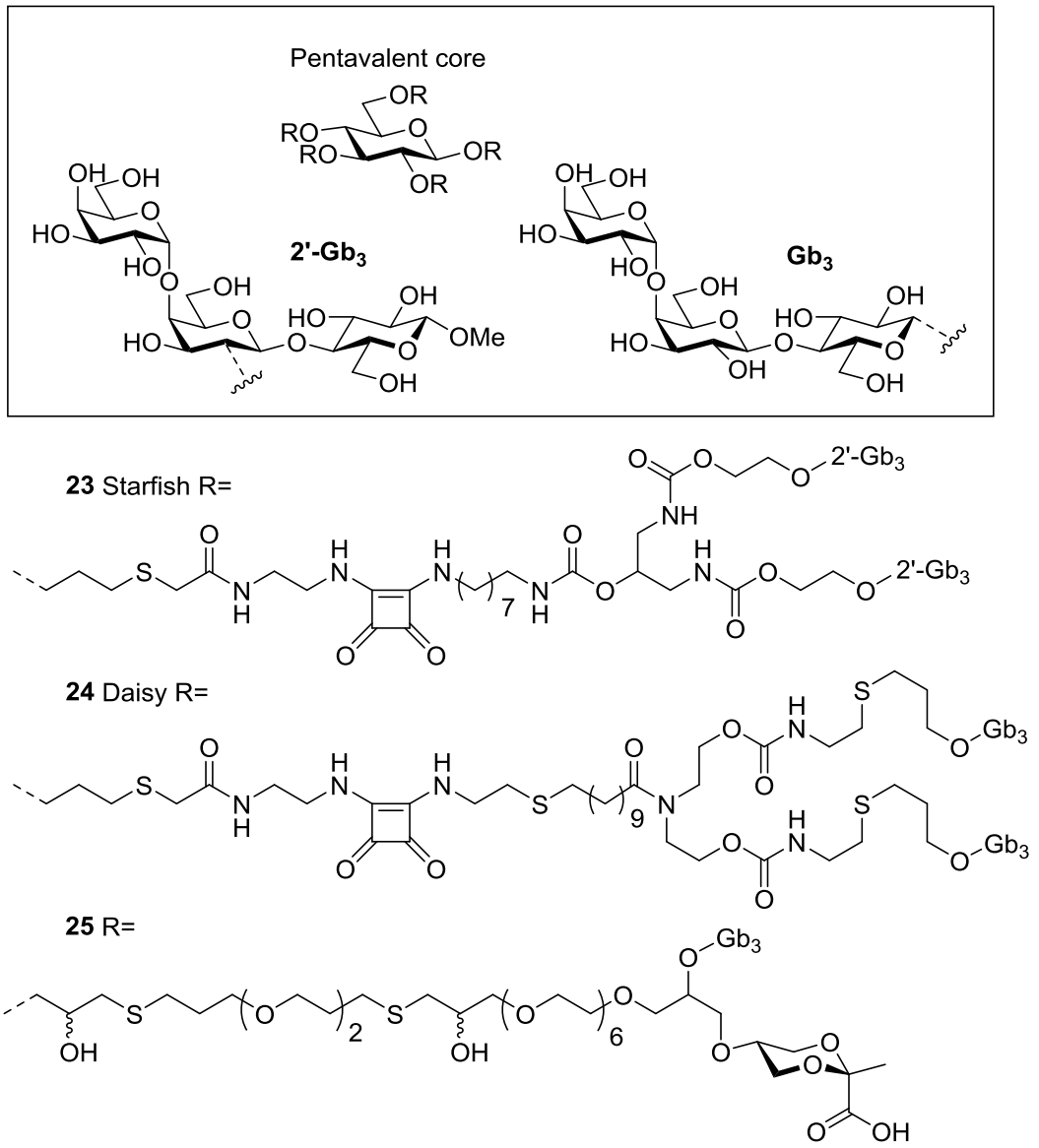


Fig 1.26 Star-shaped Stx inhibitors reported by Bundle and co-workers.

Fan's group also incorporated divalent ligands into their pentameric scaffolds to make the inhibitors decavalent (Fig 1.25).[125] These, as with the Bundle ligands, bound to two toxin pentamers in a face to face dimer. Decavalent ligand **22** with

optimised linker lengths gave a one million-fold decrease in the IC_{50} value compared to the simple monovalent ligand, again showing the power of multivalency in these carefully designed structures. Importantly, decavalent ligand **22** was over 10 times more potent than the equivalent pentavalent structure **20** (where n = 4).

1.5.7 Inhibitors using templated assembly

Templated assembly is another strategy for inducing protein dimerisation that was pioneered by Pepys *et al*. in their work on serum amyloid P component (SAP).[126] SAP is a pentameric protein in the pentraxin family that plays a role in the human innate immune system. Pepys *et al.* found that simple divalent proline derivatives, **26**, are able to bring two SAP pentamers together to form a face-to-face dimer (Fig 1.27) (Scheme 1.4 A).

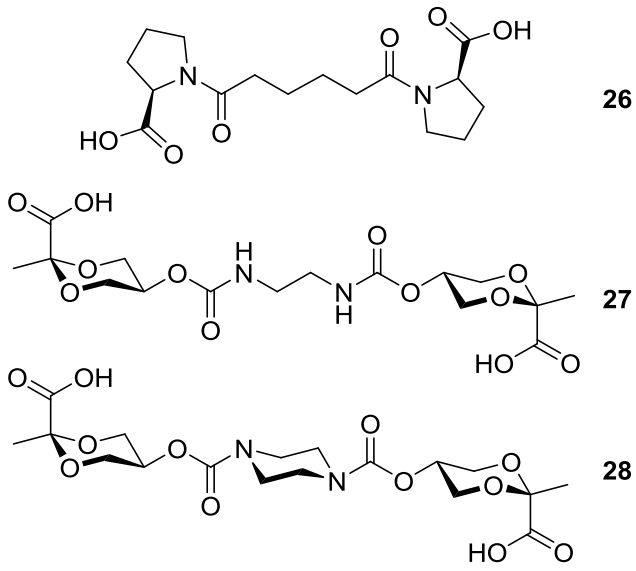


Fig 1.27 Homodimeric ligands for SAP

SAP shares the pentameric structure that is common to the bacterial toxins and their binding sites are conveniently spaced for a simple divalent ligand to bridge the binding sites in both proteins (Fig 1.28). Fan and co-workers made heterodivalent ligand **29**, which combines a proline derivative with a nitrophenyl galactoside, and showed that it could complex SAP with CTB.[127] Dynamic light scattering measurements confirmed the formation of the two protein complex and no significant higher aggregates were observed. Inhibition studies on CTB showed the divalent ligand to have an IC_{50} of 620 μM, but when SAP was included this value was reduced to 0.98 μM. The templating effect of the SAP protein thus provided a great enhancement in inhibition.

In a similar way, the Bundle group have used homodivalent, and heterodivalent ligands to create dimers of SAP[128] and to form complexes with StxB[129]. In this case they employed a cyclic ketal of pyruvate and glycerol as the SAP ligand. The flexible short ligand **27** had an IC_{50} value of 3.8 μM but a more rigid linker **28** gave a greater potency with an IC_{50} of 0.12 μM. The rigid unit was thought to minimise the loss of conformational entropy upon binding. Complexes of SAP with StxB were also formed by use of heterodivalent ligands comprising a pyruvate acetal and the Gb_3 oligosaccharide (Scheme 1.4 B).[129] As before, flexibility in the linker was detrimental to its inhibitory potency; rigid ligand **30** was about 50-fold better than the more flexible ligand **31**. They found that these smaller ligands were just as potent as the Starfish ligand in a cytotoxity assay. However, *in vivo* trials showed a poor performance for the heterodivalent inhibitors as a consequence of their rapid clearance when compared to the much larger Starfish compound **23**.

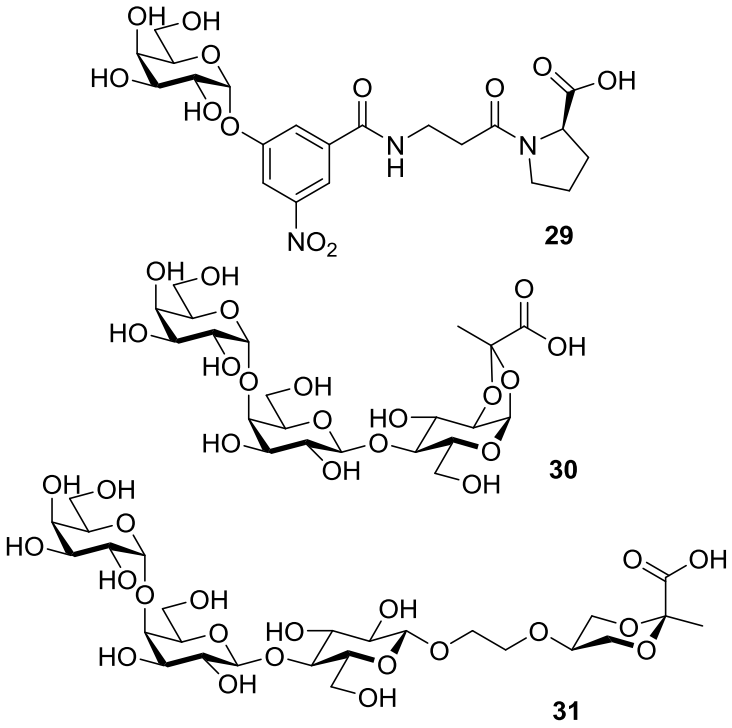
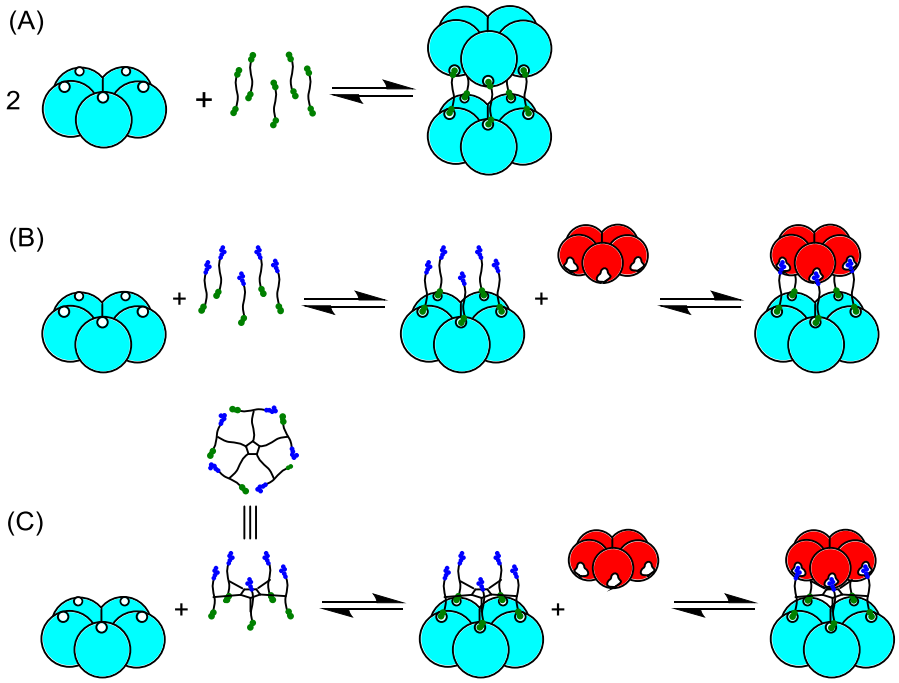


Fig 1.28 Heterodimeric ligands for SAP and CTB/LTB/StxB

Heterodivalent ligands comprising a cyclic pyruvate ketal of glycerol and the Gb_3 oligosaccharide were also incorporated onto modified a Starfish scaffold.[130] Ligand **25** (Fig 1.26), now bifunctional, was able to bring together the two different proteins as expected (Scheme 1.4 C). The inhibitor had an IC_{50} value of 140 μM but in the presence of SAP this value was improved by a factor of 35. A combination of ligand prearrangement and templating was also demonstrated using the same

heterodivalent ligands on a polymer backbone.[131] The polymers themselves had modest inhibitory power against Stx, similar to other polymeric inhibitors. But when combined with SAP, the protein templating effect for the ligand resulted in a 100,000 fold improvement on the IC_{50} value. These multimeric inhibitors performed much better than Starfish **23** in a cytotoxicity assay and trials *in vivo*.



Scheme 1.4 Templated assembly of protein dimerisation using (A) homodimeric ligands for SAP, (B) heterodimeric ligands for SAP and CTB/LTB/StxB and (C) heterodimeric ligands on a pentameric scaffold (e.g. compound **25**, Fig 1.26)

1.5.8 Conclusions

The studies outlined in this review show us that weak interactions can be greatly enhanced through multivalency. A wide variety of different structures have been created for use as multivalent inhibitors. This review is not intended to be a comprehensive review of all scaffolds investigated to date and other scaffolds, e.g., calixarenes and glyconanoparticles have been discussed elsewhere.[132, 133] Instead the aim has been to highlight some of the general principles of multivalent inhibitor design for the bacterial toxins by considering some of the major classes of multivalent scaffolds. Investigations using glycopolymers and glycodendrimers have led to the realisation that multivalent ligands that have been designed to have an improved fit to the carbohydrate binding sites can display a significant increase in affinity. The ligand spacing and length of spacer connecting the sugar to the scaffold are important to achieve optimal binding. However, better inhibition is not

necessarily achieved by maximising the number of ligand groups as increased steric crowding can prevent efficient interactions from forming.

An understanding of the structure of the bacterial toxins has led researchers to move away from simple polymeric inhibitors to more sophisticated constructions. Having a prior knowledge of the target protein is advantageous in aiding the design of the most suitable ligands and it has been seen that prearrangement of ligands to precisely fit the binding sites increases their inhibitory potential. The five-fold symmetry of the toxins has directed work towards star shaped inhibitors that present their carbohydrate ligands at precise positions for binding. Matching the valency of the inhibitor to that of the target protein can prevent unwanted aggregation. These designs can be further improved by creating self-assembling complexes of proteins. The templating effect created by the first protein can pre-organise the ligand for engaging with a second protein molecule. Templating can also have a powerful effect when used with simple small divalent ligands. Binding interactions that individually have unimpressive affinities, can still give rise to strong inhibition when assembling proteins.

In the search for higher affinity ligands and the improved inhibition of bacterial toxins, there are many different multivalent scaffolds to choose from. The smartest designs, including a combination of templating and pre-organisation of ligand groups, have the potential to produce the most potent inhibitors. Future studies should aim to combine these principles to develop scalable, economically viable multivalent materials that can ultimately be applied in the clinic.

An Introduction

Part III:
What it's all about

1.6 Outline of the project

The project was an investigation to develop general strategies for the construction of nanoarchitecture by exploiting protein-ligand interactions. As seen in the previous section, the use of protein-ligand interactions can bring about new and complex nanoarchitecture, and this area of research is still young and expanding. Protein-carbohydrate interactions are generally weak and so could be used to create a self-correcting, reversible system. CTB has five carbohydrate binding sites and there are a range of ligands with varying affinities to choose from.

The main aim of the project was to create a nanoscale virus-like particle from CTB protein building blocks that do not naturally have any affinity for forming such a structure. Pentagons do not tessellate flat but instead may fold round to form a dodecahedron. If twelve of the CTB pentamers could be brought together they may form a three-dimensional dodecahedral structure, held together via protein-carbohydrate interactions (Fig 1.29). Three main strategies were investigated towards this goal. The developments, pitfalls and other structures created on the path to this goal are discussed in the work presented here.

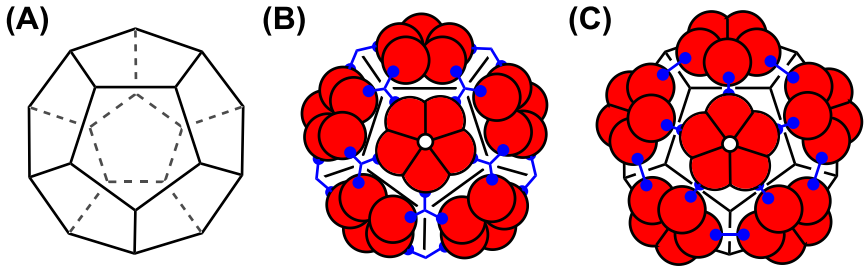
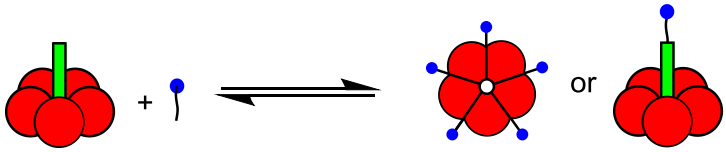


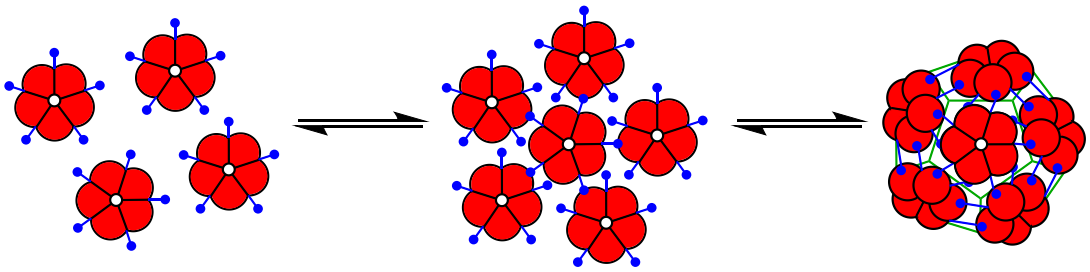
Fig 1.29 (A) Dodecahedron structure and the proposed arrangement of CTB proteins (red) to form such a structure either with (B) trivalent ligands (blue) or (C) divalent ligands (blue).

Site-specific modification of the CTB or AB_5 protein was also studied. Methods for covalent attachment of small molecules, such as carbohydrate ligands, at either the CTB or CTA2 N-terminus were investigated (Scheme 1.5).



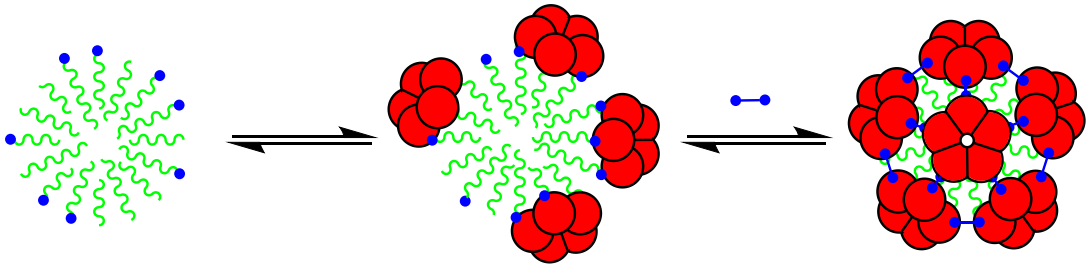
Scheme 1.5 Potential covalent modification of CTB (red) or CTA2 (light green) with ligands (blue).

The first strategy for protein assembly involved direct covalent modification of the CTB pentamers with carbohydrate ligands. The aim was that the modifications would allow the self-assembly of the proteins into discrete dodecahedral structures (Scheme 1.6). Each pentamer would be held to another via two complementary noncovalent interactions, with the N-terminus attached ligand of one pentamer linking to the carbohydrate binding site of another pentamer.



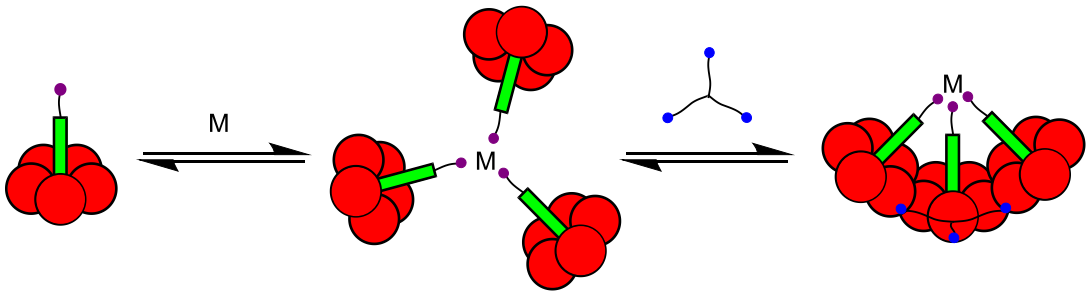
Scheme 1.6 Strategy for the self-assembly of CTB (red) pentamers via covalent attachment of carbohydrate ligands (blue). Pentamers are brought together via carbohydrate binding and arrange into discrete dodecahedral particles.

The second strategy was to use a separate template to preorganise the CTB proteins into the desired conformation (Scheme 1.7). A micelle template modified with carbohydrate ligands was used for this goal. Once assembled, the aim was that the protein units could be linked together through direct modification of the CTB pentamers with carbohydrate ligands or by the introduction of multivalent ligands.



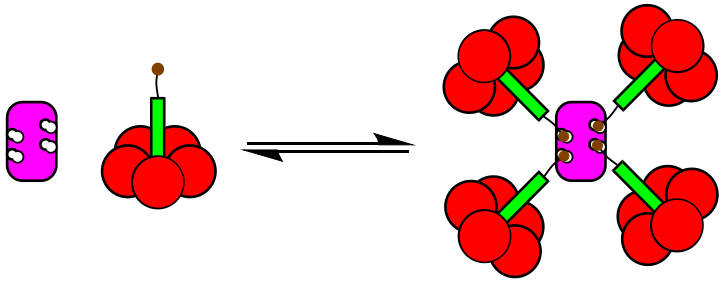
Scheme 1.7 Strategy using a micelle template (green) to first preorganise the CTB pentamers (red). Multivalent ligands (blue) are then introduced to cross-link the proteins.

Finally the AB_5 structure of cholera toxin could be used. The CTA2 peptide could be covalently modified with a metal binding group for coordination of three AB_5 proteins (Scheme 1.8). This protein cluster could then be cross-linked with multivalent ligands which may also bring together multiple clusters forming larger species.



Scheme 1.8 Using the AB_5 protein. Modification of the CTA2 peptide (light green) with metal binding ligands (purple) to cluster the proteins before cross-linking with multivalent ligands (blue).

To investigate other protein-ligand interactions, the CTA2 peptide could also be modified with a protein binding ligand. If this ligand could bind to a separate protein with multiple binding sites, such as biotin with streptavidin, then protein clusters could be formed through the interaction of this protein with modified AB_5 (Scheme 1.9). This small cluster could again then be cross-linked with multivalent ligands.



Scheme 1.9 Using the AB_5 protein, modification of the CTA2 peptide with biotin (brown) for binding to streptavidin (pink).

2 Building blocks

The analysis and characterisation of CTB

The cholera toxin B-subunit pentamer (CTB) has been identified as a potential building block for creating nanoscale assemblies using protein-carbohydrate interactions. To investigate the nanoscale structures, a range of analytical and biophysical techniques are at our disposal. This chapter outlines the complementary use of a number of these techniques to analyse the CTB protein and its interactions with carbohydrate ligands.

2.1 CTB expression and purification

To use CTB as a protein building block, the protein itself first needed to be made. CTB was expressed in a high salt growth medium using a *Vibrio sp.60* clone containing the pATA13 plasmid[134] (see appendices for protein and gene sequences). This marine vibrio species, which requires extra salt for optimal growth, was used because it directly exports the protein of interest out of the cells, thus simplifying purification. The protein can be purified from the media by precipitation. Initial studies were performed to optimise the procedure.

Initially the method of Fontana *et al.* was employed.[135] This method involved adjusting the pH to 3.5, adding sodium hexametaphosphate and storing at 4 °C. As CTB is known to dissociate into monomers below pH 3.9[136], precipitation was also tested at pH 4.5. The addition of ammonium sulfate is another common method for precipitating proteins. Stirring for 1 hour in 60% w/v ammonium sulfate was investigated.

In all cases, the precipitated protein was resuspended in *phosphate buffer* and purified on a lactosyl affinity column. CTB binds to lactosyl groups on the resin and so can be separated from any other material present. The CTB protein was eluted with a high concentration of lactose (300 mM). SDS-PAGE was used to determine the purity and size of CTB isolated from the affinity column (Fig 2.1). Comparison with the protein standards showed CTB had an apparent mass of around 40 kDa. This value is lower than that expected for the CTB pentamer (58 kDa) and is due to the compact, folded shape of the protein causing it act on the gel as if it is a smaller protein. Many proteins denature under the conditions of electrophoresis but CTB is stable enough as a pentamer that it mostly retains its tertiary structure on the gel with some dissociating into monomers. When boiled, the pentamer fully dissociates into monomers which then migrate on the gel at their expected mass of around 12 kDa.

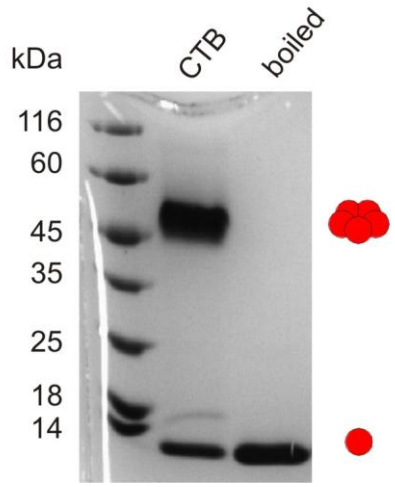


Fig 2.1 SDS-PAGE showing that CTB mostly migrates as a stable pentamer on the gel, with an apparent mass of around 40 kDa. When boiled, the protein dissociates into the monomers which migrate to the bottom of the gel.

The yield of each different purification method was calculated (Table 2.1). It was found that precipitation by ammonium sulfate followed by purification on a lactose affinity column produced a high yield of 30 mg L^{-1}. The yield was reduced by a third when sodium hexametaphosphate was used at pH 3.5 and when the pH was raised to 4.5 a further decrease in yield was found.

Table 2.1 Yields of CTB from various precipitation methods.

Precipitation method	Yield / mg L^{-1}
Sodium hexametaphosphate pH 3.5	10
Sodium hexametaphosphate pH 4.5	2
Ammonium sulfate 60% w/v	30

Mass spectrometry (MS) was used to confirm that it was indeed CTB being produced from the bacteria. The pentamer did not survive intact in the mass spectrometer due to the denaturing conditions of the solvents used in the electrospray process and so the spectrum showed only peaks for the individual monomeric B-subunits. A range of ions were observed for different charge states of the monomer and these could be deconvoluted to calculate the molecular weight. The mass measured by ESMS for CTB was 11642.9 Da (calculated 11642.9 Da) (Fig 2.2 A).

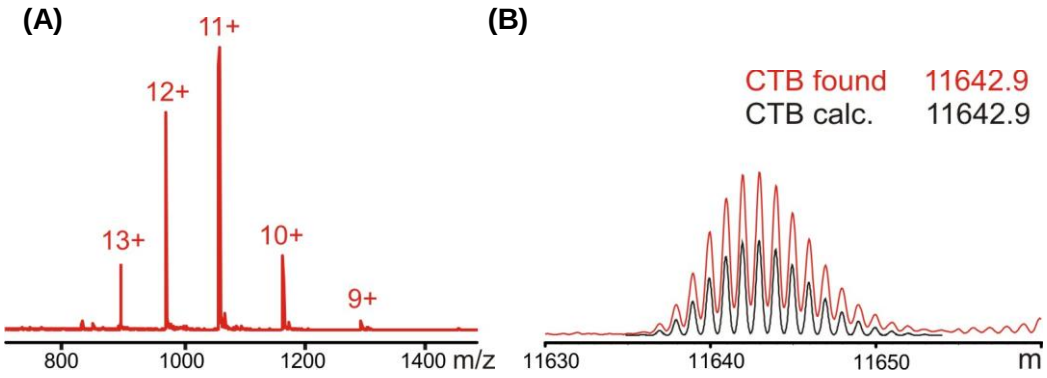


Fig 2.2 (A) Mass spectrum of the monomeric B-subunit showing a range of charge states and (B) the deconvoluted high resolution spectrum.

2.2 CTB analysis

The MS method used here does not give information on protein oligomerisation and SDS-PAGE can overestimate the amount of monomer present due the denaturing conditions. Therefore other techniques are needed to confirm the size of the proteins and any assemblies that will be made.

2.2.1 Size exclusion chromatography (SEC)

SEC can be used to separate proteins of different masses and thus reveal how many different species are present in a mixture. The components are separated based on their size and shape by using a bed of porous beads. The stationary phase contains tunnels of different sizes that will allow the passage of different sized molecules. Therefore samples with a smaller size than the pore size will have a longer path and retention volume, than larger particles that cannot enter all the pores. However, samples that are too large to enter any pores will elute at the void volume and cannot be analysed by SEC.

To calibrate the SEC column (superdex™ 200 using a Pharmacia ÄKTA FPLC system) a number of calibrants of different masses were first passed through the column (Fig 2.3 A). The known mass of each calibrant was plotted against the ratio of its retention volume (V_e) to the void volume (V_o) of the column and an exponential function was fitted to the data (Fig 2.3 B). The void volume was calculated as the retention volume for Blue Dextran which has a molecular weight of 2000 kDa and is unretained by the column. The R^2 value for the trend line of the graph was 0.98 and therefore shows a good correlation between the masses and their retention volumes. Subsequent analysis of proteins or their assemblies during SEC purification allowed estimation of the mass of the assembly. The CTB pentamer was found to have a SEC retention volume of 17.4 mL (Fig 2.3 C). When placed on

the calibration graph, the actual retention volume for CTB compares well to the expected value.

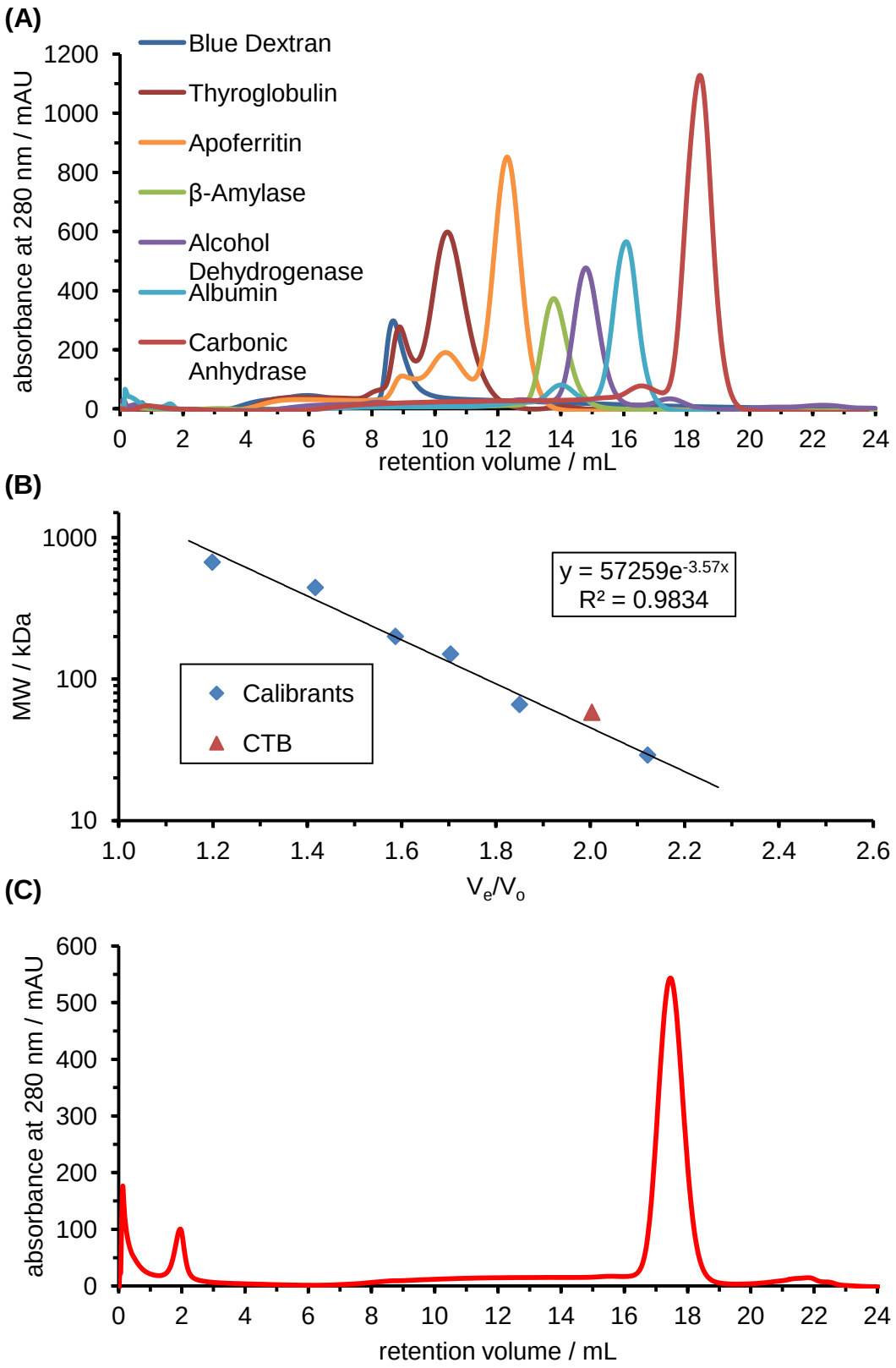


Fig 2.3 (A) Traces for each calibrant for SEC. (B) A calibration graph for the SEC column showing the molecular weights of the calibrants plotted against the ratio of retention volume (V_e) to void volume (V_o). (C) Chromatogram for CTB showing a retention volume of 17.4 mL for the pentamer.

2.2.2 Analytical Ultracentrifugation (AUC)

Sedimentation velocity AUC experiments were conducted to obtain mass information and assess the oligomerisation state of CTB.

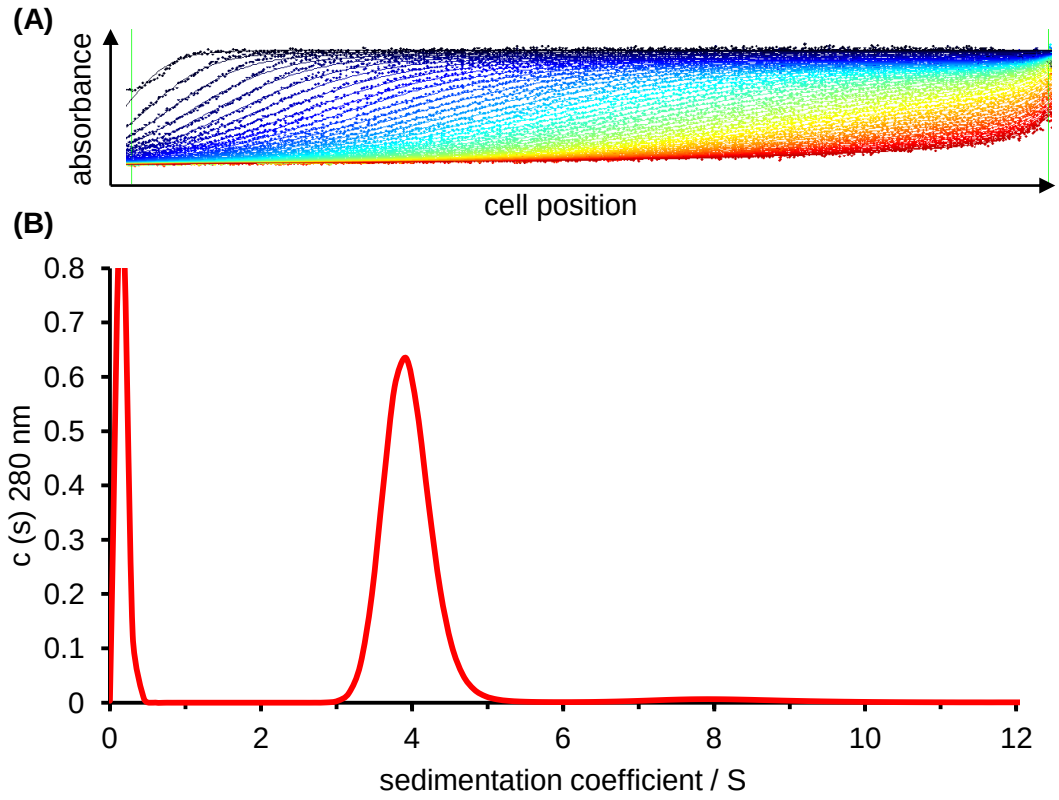


Fig 2.4 AUC sedimentation coefficient distribution of CTB performed at a rotor speed of 35,000 rpm, rotor temperature of 20 °C and scanned at 280 nm. (A) The absorbance data measured across the cell length. Measurements taken over time, from blue to red, show the gradual sedimentation of the absorbing species. (B) The c (s) plot shows a single species at 4.1 S which corresponds to a mass of 50 kDa for the CTB pentamer.

AUC measures the absorbance, and thus protein concentration, across the sample cell as it is centrifuged. Early in the experiment the protein sample is evenly distributed across the cell but as the experiment progresses a concentration gradient is built up from the sedimentation and diffusion of the particles (Fig 2.4 A). The position of the concentration boundary is monitored as a function of time, as the sample becomes sedimented. The rate of sedimentation gives information on the mass and the absorbance gradient gives information on particle shape. The Svedberg equation can be used to calculate size and shape information of the particles (Equation 2.1). The Lamm equation describes the sedimentation and diffusion of a particle under ultracentrifugation (Equation 2.2).

Equation 2.1
$$S = \frac{M(1-v\rho)}{N_A f} = \frac{MD(1-v\rho)}{RT} = \frac{u}{\omega^2 r}$$

Equation 2.2
$$\frac{\partial c}{\partial t} = \frac{1}{r}\frac{\partial c}{\partial r}\left[rD\frac{\partial c}{\partial r} - cs\omega^2 r^2\right]$$

where *s* is the sedimentation coefficient, *M* is the mass of the molecule, *v* is the partial specific volume (buoyancy), *ρ* is the density of the solution, N_A is Avogadro's number, *f* is the frictional coefficient, *D* is the diffusion coefficient, *T* is the temperature, *R* is the gas constant, *u* is the radial velocity of the particle, *r* is the radius, *w* is the angular velocity, c is the particle concentration and *t* is the time.

The Sedfit software[137] fits the parameters from the Svedberg and Lamm equations to create a c (s) plot (Fig 2.4 B) which shows the relative concentration of species with a given sedimentation coefficient, and from this calculates the mass of the particles. A sample of CTB was analysed at a monomer concentration of 800 µM and a single species was seen with a sedimentation coefficient of 4.1 S which corresponds to a mass of 50 kDa (Fig 2.4).

2.2.3 Dynamic light scattering (DLS)

DLS allows the hydrodynamic diameter to be calculated to give another measurement for the size of the protein. In a DLS experiment the sample is illuminated by a laser and the intensity of the scattered light fluctuations is detected at a given angle. Light scattered by different particles can be in phase or out of phase leading to higher or lower intensity. Particles diffuse in a liquid due to Brownian motion and the rate at which the scattering intensity fluctuates is related to the speed of motion and therefore the size of the particles (Fig 2.5).

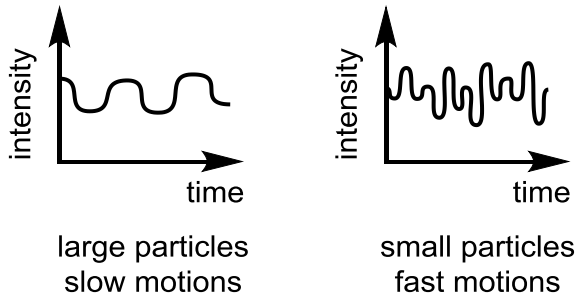


Fig 2.5 The different fluctuations in DLS scattering intensities over time for large, slow moving particles or small, fast moving particles.

The hydrodynamic diameter is calculated from the velocity of the Brownian motion for the particle defined by its translational diffusion coefficient, *D*. The hydrodynamic diameter can be calculated using the Stokes-Einstein equation (Equation 2.3). The

measurement obtained is for the diameter of a sphere that has the same translational diffusion coefficient as the particle and so may lead to inaccuracies for non-spherical particles and proteins.

Equation 2.3 $$d(H) = \frac{kT}{3\pi\eta D}$$

where *d(H)* is the hydrodynamic diameter, *k* is the Boltzmann constant, T is the temperature, η is the viscosity of the medium and *D* is the translational diffusion coefficient.

An auto correlator measures intensity fluctuations. The signal is compared with itself at varying time intervals to construct the correlation function, G(τ) (Fig 2.6). This function can be described as an exponential decay function dependent on the translational diffusion coefficient of the particle in solution (Equation 2.4). As the signal changes more slowly the correlation persists for longer with large particles compared to small particles.

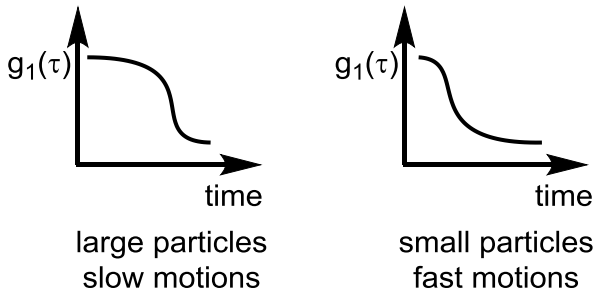


Fig 2.6 The difference in correlation decay for large and small particles over time.

Equation 2.4 $$G(\tau) = A(1 + Be^{-2\Gamma\tau})$$

where *A* is the baseline of the function, *B* is the intercept, τ is the time difference of the correlator and *Γ* is the decay rate (Equation 2.5).

Equation 2.5 $$\Gamma = Dq^2$$

where *q* is the wave vector (Equation 2.6).

Equation 2.6 $$q = \frac{4\pi n}{\lambda} sin\frac{\theta}{2}$$

where *n* is the refractive index of the dispersant, λ is the wavelength of the laser and θ is the scattering angle.

A sample of pure CTB at a concentration of 100 μM was filtered (400 nm filter) and analysed by DLS. The result showed a single peak corresponding to a particle with

a diameter of 5.6 nm (Fig 2.7). As CTB is not completely spherical this measurement compares well with the known dimensions of CTB of roughly 6.5 × 3.5 nm determined by x-ray crystallography.[138]

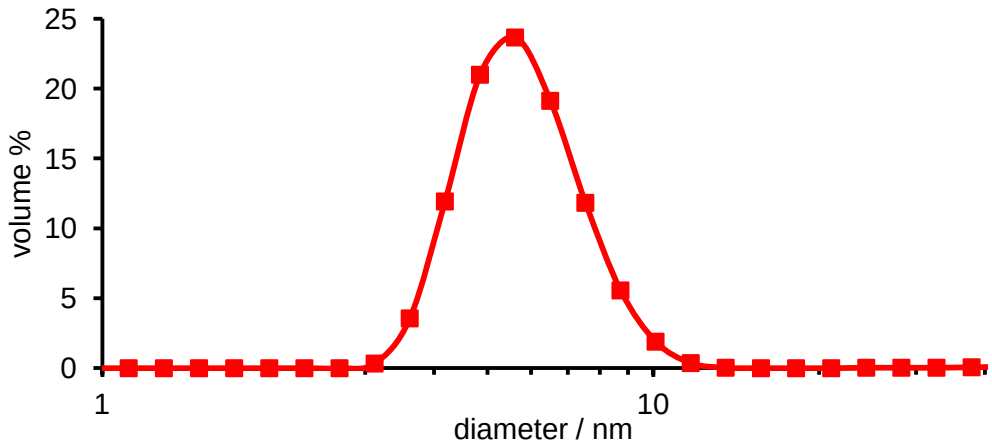


Fig 2.7 DLS measurement for CTB showing a single peak of diameter 5.6 nm.

2.2.4 Atomic Force Microscopy (AFM)

Whilst the biophysical techniques discussed above can give a good estimate of the size of particles in solution, being able to visualise the proteins directly would be much preferable.

In an AFM tapping mode experiment, a cantilever is made to vibrate at a resonant frequency and the amplitude of the vibration is measured via a laser (Fig 2.8). As the cantilever moves over the surface and encounters bumps, the amplitude of oscillation decreases and a feedback system adjusts the height of the cantilever to maintain a constant amplitude. This feedback allows a height representation of the surface to be constructed. Details of the size and shape of particles on a surface can be seen down to the low nanometre scale. Measuring the size of proteins pushes AFM to its detection limits, but it should be possible to resolve larger protein assemblies and structures.

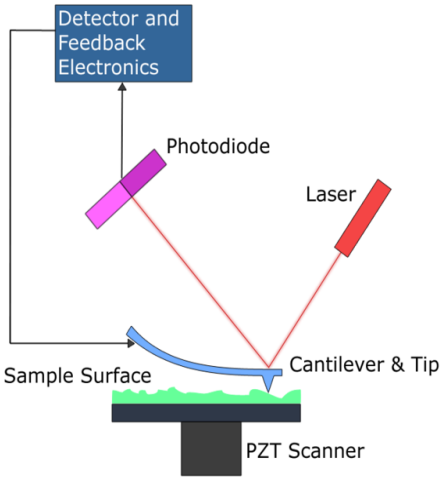


Fig 2.8 The AFM set-up showing the feedback system from a laser reflecting off the cantilever.[139]

It has been shown previously that CTB will spontaneously adhere to a mica surface under aqueous conditions which allowed AFM experiments to be conducted to image the protein.[140] CTB protein in a *citrate phosphate buffer* (pH 5.0) was incubated for 60 min and imaged. Scanning at pH 5 meant that the sample was two pH units away from the isoelectric point of CTB, so the protein would have a net positive charge and therefore interact more strongly with the negatively charged mica surface. Closely packed proteins were seen at a reasonable resolution (Fig 2.9 A). Expanding the image revealed the round shape of the CTB pentamer with some particles exhibiting a small hole at the centre (Fig 2.9 B & C).

Bilayers can be laid down on a mica surface for AFM imaging. GM1 glycolipids were incorporated into bilayers and used to bind CTB to the surface. This different way of capturing CTB on a surface would guarantee all the proteins were of the same orientation and could also show any assemblies formed by interactions with the top or sides of the proteins. Flat networks or vertical constructions could potentially be visualised by anchoring a CTB unit to the surface in this way. These networks could even be manipulated on the surface and imaged as the structures are assembled and disassembled.

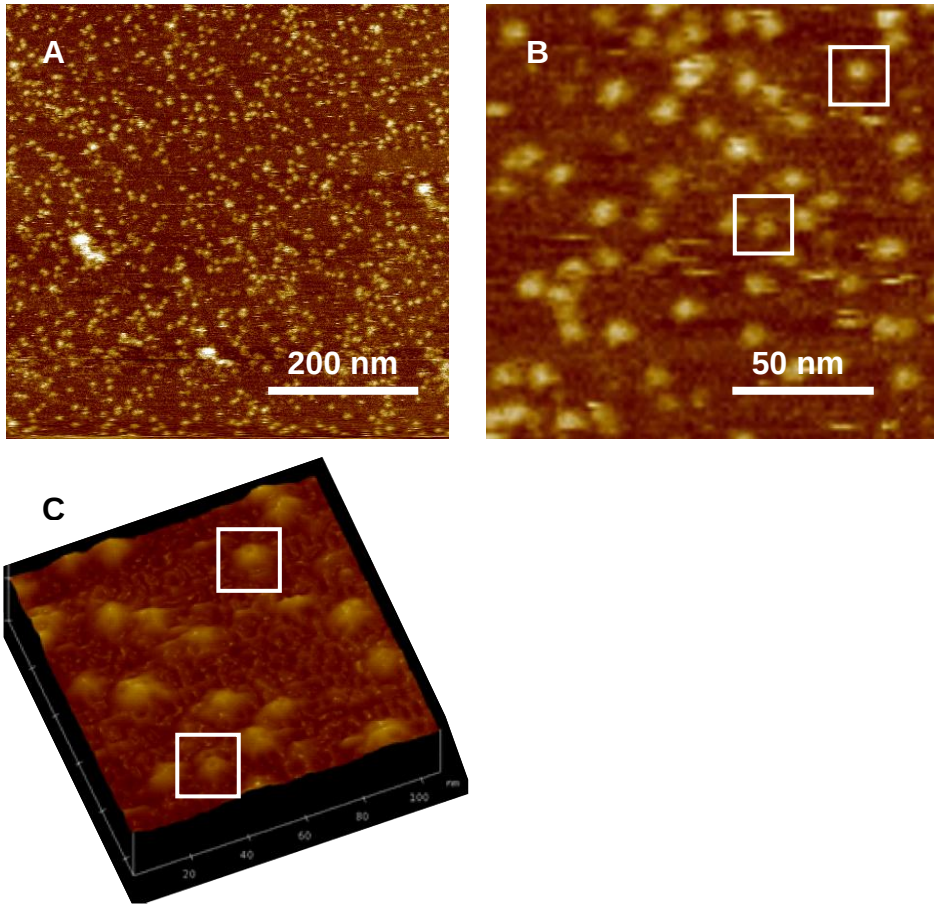


Fig 2.9 AFM image of CTB in *citrate-phosphate buffer*, pH 5, on a mica surface. (A) The pentamers are closely packed and reasonably resolved. (B) An expanded region of the image reveals the round shapes more clearly. (C) The expanded image cross section allows the depth of the structures to be seen (boxed) with a small recess in the centre and of appropriate size for a CTB pentamer (~6.5 nm).

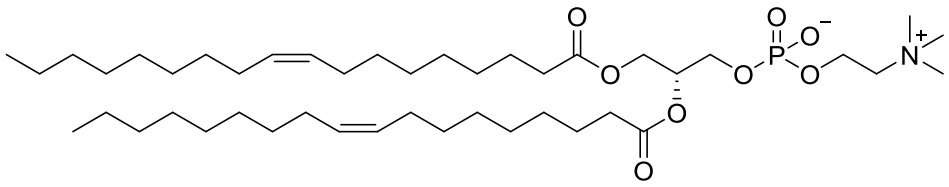


Fig 2.10 Structure of the DOPC phospholipid.

A lipid bilayer of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) (Fig 2.10) was prepared in PBS buffer, applied to a mica surface and imaged in tapping mode. A smooth single bilayer was observed with few defects (Fig 2.11 A).

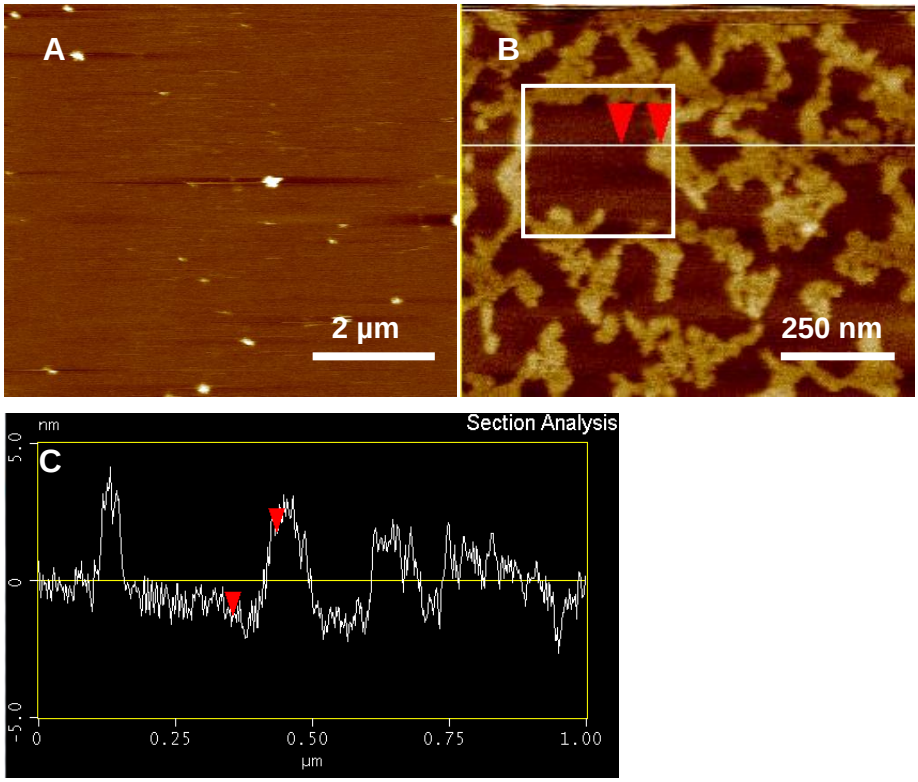


Fig 2.11 AFM images of (A) a flat lipid bilayer of DOPC and (B) a bilayer containing GM1, with CTB bound to the surface. A region had been cleared of CTB by applying force to the surface (boxed region). (C) Height profile for line across the scan in B. The height difference between the points marked with red triangles is 3.04 nm, which corresponds to the height of a CTB protein flat on the surface.

Another lipid bilayer of DOPC was prepared on a mica surface again in PBS buffer but with 1% GM1 incorporated. CTB was applied to the surface in PBS buffer and bound to the GM1 on the surface. A surface was observed with some areas ~3 nm higher. The higher areas could be moved by zooming in and applying force in an area. After zooming out, a square shaped region of the lower surface could be seen, where the excessive force had been applied (Fig 2.11 B). The height profile analysis showed a height difference of 3.04 nm between the points marked. The higher layer was consistent with CTB bound on the surface, but the resolution was not sufficient for the individual pentamers to be observed clearly. Although the individual protein pentamers could not be imaged clearly by this method, larger structures should prove easier to visualise with AFM.

2.2.5 Characterisation of protein assemblies

SEC can separate a mixture and give a rough estimate of the size of proteins. AUC gives more accurate mass information and can show the number of different size species in solution. The size of particles can also be found using DLS which is a

faster analysis and uses a smaller sample than the other techniques. These three techniques are complementary and a powerful array of techniques for analysing individual protein structures and higher assemblies in solution.

MS is vital to obtain a quick and accurate mass for CTB, to show modifications made and to what extent these reactions have proceeded. From SDS-PAGE the size of the structures can be seen and can also provide an indication of whether the protein has denatured or whether larger aggregates are beginning to form. SEC also gives another measure of the size of structures present and can be used to purify CTB or larger supramolecular particles. Finally, AFM allows imaging of the protein structures and so should present ultimate proof of the size and shape of the nanoscale protein assemblies.

2.3 Binding site studies with isothermal titration calorimetry (ITC)

CTB has 10 natural carbohydrate binding sites. If we wish to take advantage of this ligand binding ability then these sites need to be fully understood. Five blood group oligosaccharides bind to the top-side face of CTB and five GM1os groups bind to the bottom face (Fig 2.12). Both sets of carbohydrates are equally spaced around the five-fold symmetry axis of the protein (Fig 2.13).

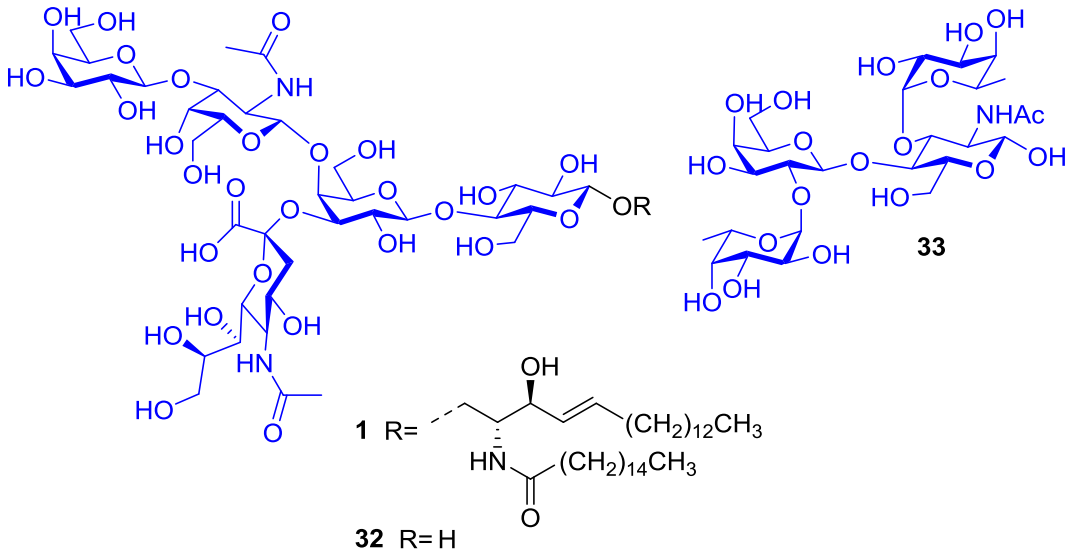


Fig 2.12 CTB binding ligands. The ganglioside GM1 **1** and the GM1os **32** bind to the bottom face of CTB whereas blood group O (lewis y) **33** binds to the top-side of CTB.

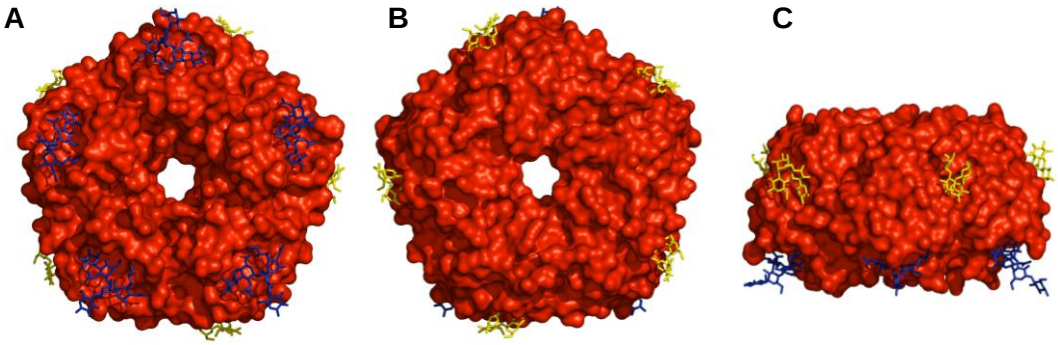


Fig 2.13 (A) Bottom, (B) Top and (C) side view of CTB with carbohydrates bound. Five GM1 oligosaccharides **32** (blue) bind to the bottom face and five blood group O structures **33** (yellow) bind the the top side evenly distributed around the five-fold symmetry of the protein.

Isothermal titration calorimetry (ITC) can be used to measure the affinity of the carbohydrate interactions. From a single titration a binding affinity, K_d, the enthalpy change, ΔH°, the entropy change, ΔS°, and the stoichiometry of a binding interaction can be found. The calorimeter has two cells, a reference cell and a sample cell. The cells are controlled by independent heaters and when samples are mixed during the titration heat is released for which the heaters compensate by supplying less power to the sample. The raw data shown is the difference in power between the heaters which is then integrated to give the Wiseman binding isotherm.

The shape of the curve describes the change in enthalpy per mole of ligand titrated into the protein and from this the binding affinity can be found. For a simple single-site model (Equation 2.7) the fitting parameters are varied in the Wiseman isotherm to find the best fit for the curve (Equation 2.8).[141] The equation relates the stepwise change in heat of the system for a given change in ligand concentration to the ratio of ligand to protein concentration. The shape of the curve depends upon the c value which is a ratio of protein concentration to the K_d (Equation 2.9). Low c value curves possess a hyperbolic curve and high c value curves show a sigmoidal shape (Fig 2.14).

Equation 2.7

$$X + M \rightleftharpoons XM$$

Equation 2.8

$$\frac{dQ}{d[X]_t} = \Delta H^{\circ} V_0 \left[\frac{1}{2} + \frac{1 - ([X]_t/[M]_t) - (K_d/[M]_t)}{2\sqrt{1 + [([X]_t/[M]_t) + (K_d/[M]_t)] - 4([X]_t/[M]_t)}}\right]$$

Equation 2.9

$$c = \frac{1}{K_d/[M]_t} = \frac{[M]_t}{K_d} = K_a[M]_t$$

where Q is the heat of the system, $[X]_t$ is the concentration of ligand after injection t and $[M]_t$ is the protein concentration after injection t. $\Delta H°$ is the enthalpy change, V_0 is the effective volume of the cell, K_d is the dissociation constant and c is a ratio of the protein concentration to the dissociation constant.

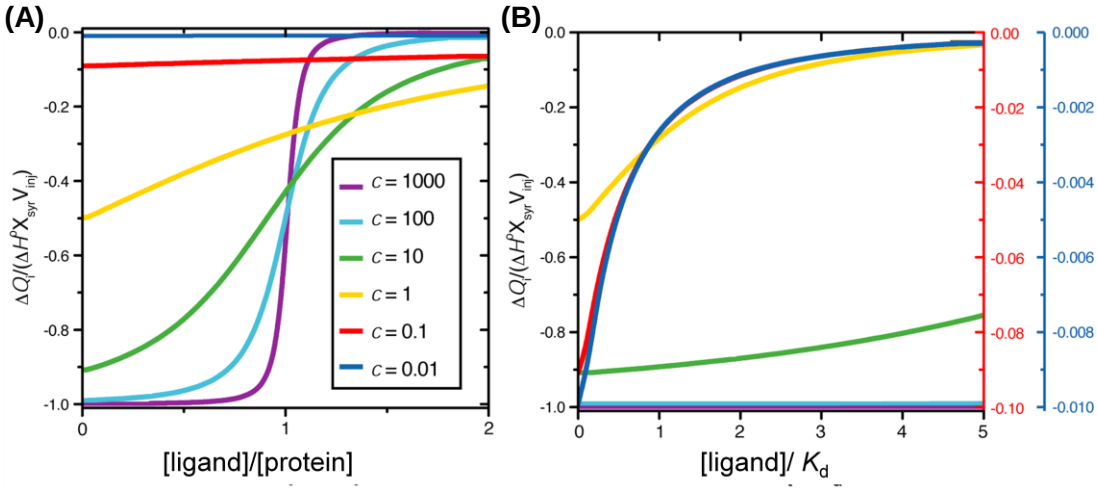


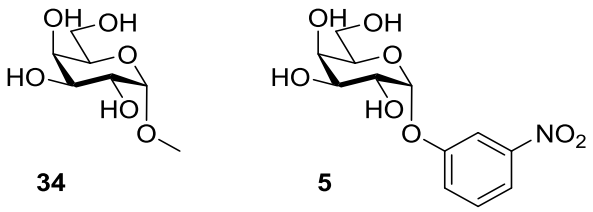
Fig 2.14 (A) The shape of the titration curve varies with the c value. (B) If the isotherm is depicted as a ratio of ligand concentration and dissociation constant then the hyperbolical curve for low c value systems is shown more clearly. The red and blue curves are magnified to highlight the similarity in curve shapes.[141]

2.3.1 GM1os derivatives

GM1 adhesion is the mechanism by which CTB binds to cells in the body. The five binding sites are evenly spaced around one face of the protein and a single GM1os interaction has been shown to have a high affinity with a K_d of 43 nM.[89] Such a high affinity would provide an excellent basis for the binding of other units or multivalent ligands to CTB.

To confirm the affinity of GM1os for CTB, the oligosaccharide first had to be purified. Ceramide glycanase (Takara Bio Inc) was used to cleave the ceramide chain from GM1 **1** enzymatically resulting in the GM1os **32**. The oligosaccharide was then separated from the ceramide lipid by reverse phase chromatography. Analysis by LC-MS showed the pure GM1os product **32** to have a mass of 997.6 Da (calculated 997.3 Da).

Isothermal titration calorimetry (ITC) was used to check the CTB protein was functional. The GM1os (50 µM) ligand was titrated into a solution of CTB (5 µM) in the cell (Fig 2.16). The binding affinity was calculated and the interaction was found to have a K_d of 22 nM which agreed well with the literature value of 43 nM[89] (Table 2.2). The GM1os binds with a very high affinity for a protein-carbohydrate interaction, but fragments of GM1os are known to bind with a much reduced affinity, with the terminal galactose found to be the most important moiety in the GM1os interaction[89] and so ITC was performed with α-methyl galactoside **34** (200 mM) titrated into CTB (350 µM). This monosaccharide gave a much reduced affinity with a K_d of only 12 mM, again in close agreement with the millimolar affinity reported previously.[89]



34 **5**

Fig 2.15 Galactose derivatives for ITC analysis.

Another galactose derivative, *meta*-nitrophenyl α-D-galactopyranoside (MNPG) **5** (15 mM), was titrated into CTB (170 µM) and a K_d of 550 µM was found (Fig 2.16) showing a 20 times increase in affinity from the simple galactoside **34**. This result for MNPG **5** was within the same order of magnitude as the K_d of 175 µM, which was found by ITC for **5** binding to LTB.[94]

Table 2.2 ITC data for GM1os and galactose derivatives binding to CTB. [a]Stoichiometry was set at 1.00 as it could not be varied in the fitting model.[141]

ligand	K_d / nM	ΔH° / kcal mol^{-1}	n
GM1os **32**	22.0 ± 2.61	-18.20 ± 0.13	1.01
34	(12.2 ± 0.69) × 10^6	-15.39 ± 1.71	1.00[a]
5	(550 ± 10.0) × 10^3	-7.12 ± 0.06	1.00[a]

The galactose ligands tested here, along with GM1os, showed a wide range of possible affinities that can be achieved by ligands for CTB. In this project we will take advantage of the variation in the binding affinities to design ligands with different properties. Lower affinity ligands will give a greater potential for reversibility and self-correction in the self-assembling systems. Furthermore, it may be possible

to use lower affinity ligands to assemble structures initially and displace them with higher affinity ligands to induce a change or reorganisation of the assembly.

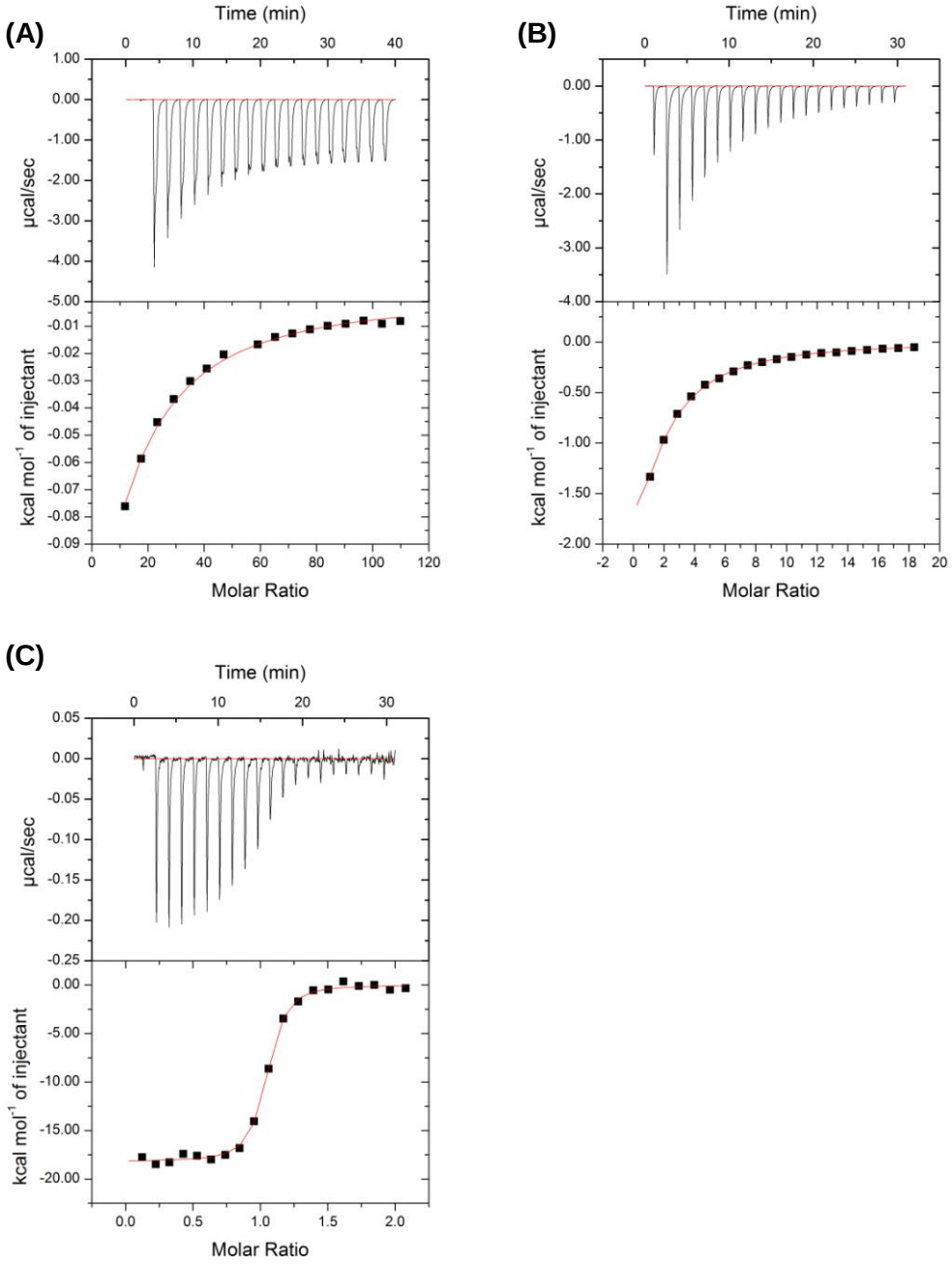


Fig 2.16 ITC titrations of CTB binding ligands. (A) The weak interaction of α-methyl galactoside **34** (200 mM) titrated into CTB (350 µM), (B) the stronger interaction of MNPG **5** (15 mM) titrated into CTB (170 µM) and (C) the very strong interaction of GM1os **32** (50 µM) titrated into CTB (5 µM).

2.3.2 Blood group oligosaccharides

A manuscript based on the following section has been published: P. K. Mandal, T. R. Branson, E. D. Hayes, J. F. Ross, J. A. Gavín, A. H. Daranas and W. B. Turnbull, “Towards a Structural Basis for the Relationship Between Blood Group and the Severity of El Tor Cholera, *Angew. Chem. Int. Ed.*, 2012, **51**, 5143-5146”.[88]

It is known that the severity of cholera caused by the El Tor biotype of *V. cholerae* is blood group dependent, but the reason for this observation is unknown.[142] Those with blood group O are affected more severely than those in A or B. However, there is no clear dependence for the O1 classical biotype of *V. cholerae*.[143] Krengel and co-workers published a crystal structure showing LTB in complex with a blood group A-Lewis-y oligosaccharide[144] and predictions were made that O- and B-blood group derivatives of the Lewis-y oligosaccharide would not bind to LTB.[145] Two oligosaccharides were synthesised to test this hypothesis, the O- and B-blood group derivatives **35** and **36** respectively (Fig 2.17) (synthesised by Dr Pintu Mandal). The B-blood group was chosen as it had been suggested that this oligosaccharide has a more pronounced affect on the severity of El Tor Cholera than the A-blood group.

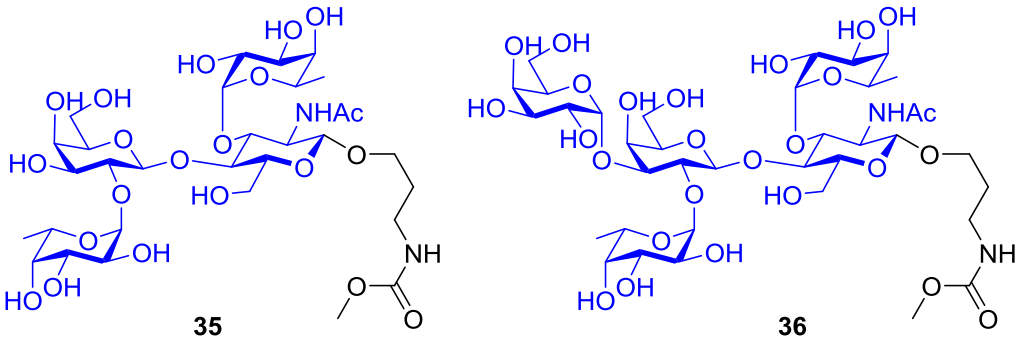


Fig 2.17 Blood group tetrasaccharide derivative O **35** and pentasaccharide B **36**.

A combination of ITC and saturation transfer difference (STD) NMR was used to confirm the binding of each carbohydrate structure with LTB. Similar affinities of 7.5 mM and 5.0 mM were found for **35** and **36** respectively (ITC performed by Dr Pintu Mandal and STD-NMR performed by Dr Antonio Daranas). It was found that El Tor CTB bound to blood group O **35** but not to blood group B **36**. STD-NMR was used to confirm which ligands were binding (Fig 2.21). The technique involves irradiating the nuclei on the protein and observing the magnetisation transferred to the ligand. The technique is particularly good for weakly binding ligands that are in fast exchange on the NMR timescale.

There are differences in the amino acid sequence, found in the carbohydrate binding pocket, between LTB and El Tor CTB. There are four differences between LTB and El Tor CTB in this region; S44N, N94H, S4N and T47I (Fig 2.18). It was proposed that the T47I difference was responsible for the lack of binding to blood group B with El Tor CTB. This change would result in the loss of a hydrogen bond from the threonine hydroxyl group to the 2-NHAc/OH group of carbohydrate residue E (Fig 2.18) and introduce a potential steric clash with the ethyl group of the isoleucine.

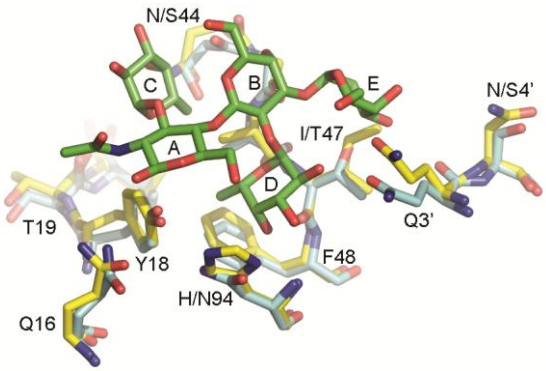


Fig 2.18 Overlay of LTB structure (light blue; 2O2L.pdb) and a model of El Tor CTB (yellow) constructed using the structure of classical biotype CTB mutant (3CHB.pdb) by introducing appropriate mutations at positions 18, 47 and 94 (model realised using pymol). The position of blood group B (green) is based on the coordinates of analogue 3b bound to LTB (2O2L.pdb). Q3' and N/S4' are residues in the neighbouring B-subunit.[88]

My role in the project was to test the key hypothesis that the T47I mutation was responsible for determining if the oligosaccharides bound to El Tor CTB. Plasmid pSAB2.2T (see appendix) contains the CTB gene for expression in *E. coli*. Site-directed mutagenesis was performed on this plasmid to introduce a threonine residue at position 47, creating plasmid pTRB-I47T. The binding of the CTB mutant produced from this plasmid could be compared to the wild-type El Tor CTB. *E. coli XL10* competent cells were transformed with the new plasmid, a plasmid prep was then performed to purify the DNA for sequencing and for transformation into *E. coli C41* cells for over-expression. The new protein was expressed by induction with IPTG and growth at 30 °C for 24 h. After purification by ammonium sulfate precipitation of the media, a nickel affinity column and then SEC, a yield of 2.9 mg L^{-1} was obtained. SDS-PAGE was performed to check the protein expression and purification (Fig 2.19).

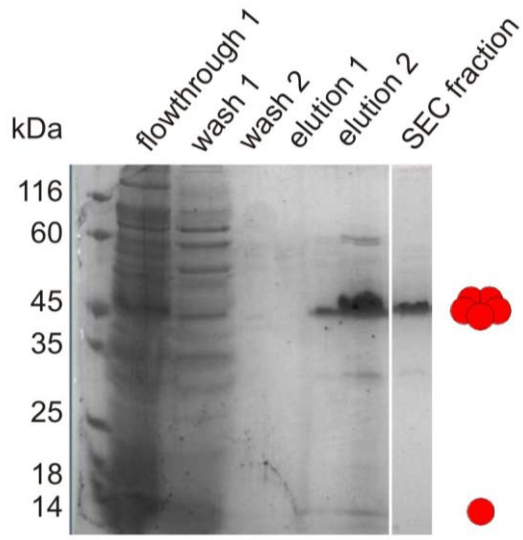


Fig 2.19 SDS-PAGE showing the expression of CTB I47T and purification by Ni affinity chromatography and SEC.

ITC titrations were performed with oligosaccharide **35** or **36** (32 mM) titrated into the I47T CTB mutant (120 µM) (Fig 2.20). The expected binding with **35** was found to have a K_d of 1.0 mM and an enthalpy change of -2.3 kcal mol^{-1}, similar to the interaction with wild-type CTB (Table 2.3). A slight buffer mismatch was seen in the titration due to the presence of D_2O from lyophilising the compound from an NMR sample, therefore a control titration was performed with the corresponding amount of D_2O which accounted for the large heat of dilution in the data. The binding of pentasaccharide **36** was inconclusive with direct titrations, possibly due to a low enthalpy for the interaction. Therefore a competition titration was performed with pentasaccharide **36** premixed with CTB I47T. Tetrasaccharide **35** was titrated into the mixture and a binding interaction was observed with a decreased enthalpy change indicative of competitive binding.

Table 2.3 ITC results for oligosaccharides **35** and **36** binding to LTB, El Tor CTB and CTB I47T.

Protein	Ligand	K_d / mM	ΔH° / kcal mol^{-1}
LTB	O **35**	7.5 ± 1.9	-5.8 ± 0.5
CTB I47T	O **35**	1.0 ± 0.3	-2.3 ± 0.8
El Tor CTB	O **35**	1.8 ± 0.2	-1.3 ± 0.2
LTB	B **36**	5.0 ± 0.5	-8.5 ± 0.9
El Tor CTB	B **36**	ẋ	ẋ
CTB I47T	B **36**	Binding detected	Binding detected

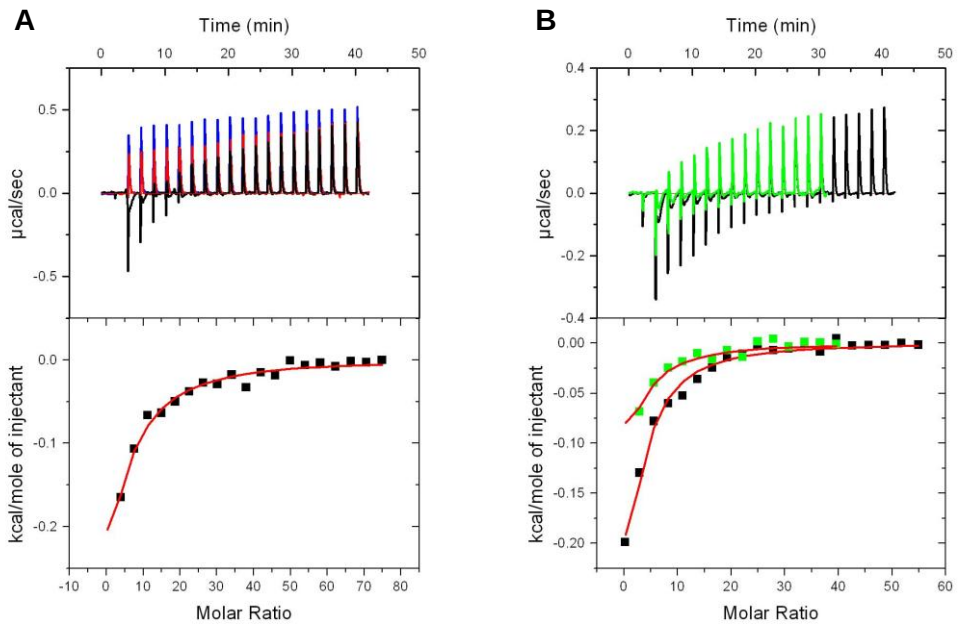


Fig 2.20 ITC titrations with I47T CTB mutant. (A) titration of tetrasaccharide **35** into CTB I47T (black) with corresponding dilution experiment (red) and control titration of 180 mM D2O into water (blue). (B) Titration of tetrasaccharide **35** into CTB I47T (black) overlaid with a competition titration experiment in which tetrasaccharide **35** was titrated into a mixture of I47T CTB premixed with 5.5 mM pentasaccharide **36** (green).[88]

Although it was not possible to obtain a binding affinity for the interaction of the I47T mutant with oligosaccharide **36** by ITC, STD-NMR analysis (performed by Dr Antonio Daranas) confirmed that the I47T mutant could bind to both tetrasaccharide **35** and pentasaccharide **36** (Fig 2.21). It was concluded that the structural basis for the selective binding by El Tor CTB was due to the isoleucine residue at position 47. This mutation prevents binding of the protein to blood group B, whereas with blood group O the toxins could be concentrated on the cell surface by blood group binding, which may act as a precursor to endocytosis via GM1os binding. Whilst these individual interactions are quite weak, multivalent presentation may give rise to a strong and important binding, which leads to the difference in the severity of cholera, from the El Tor biotype.

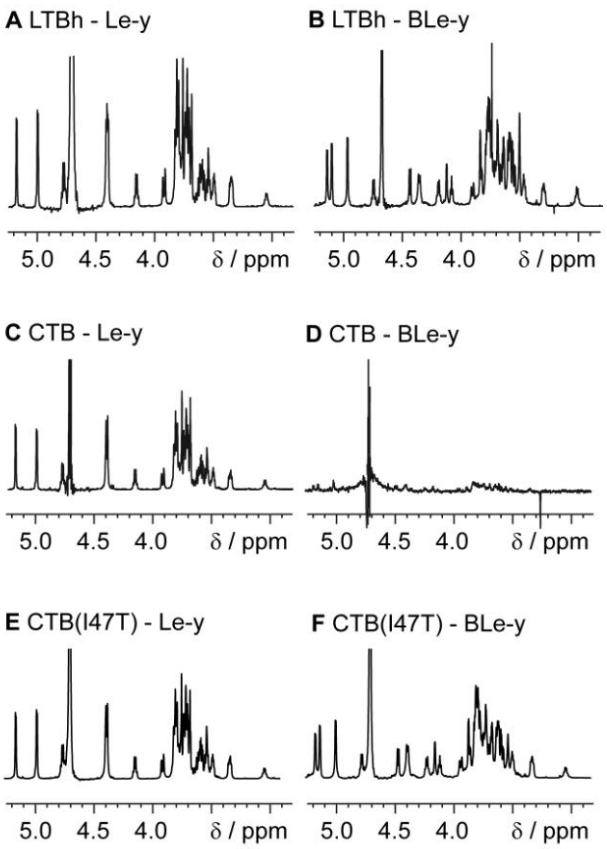


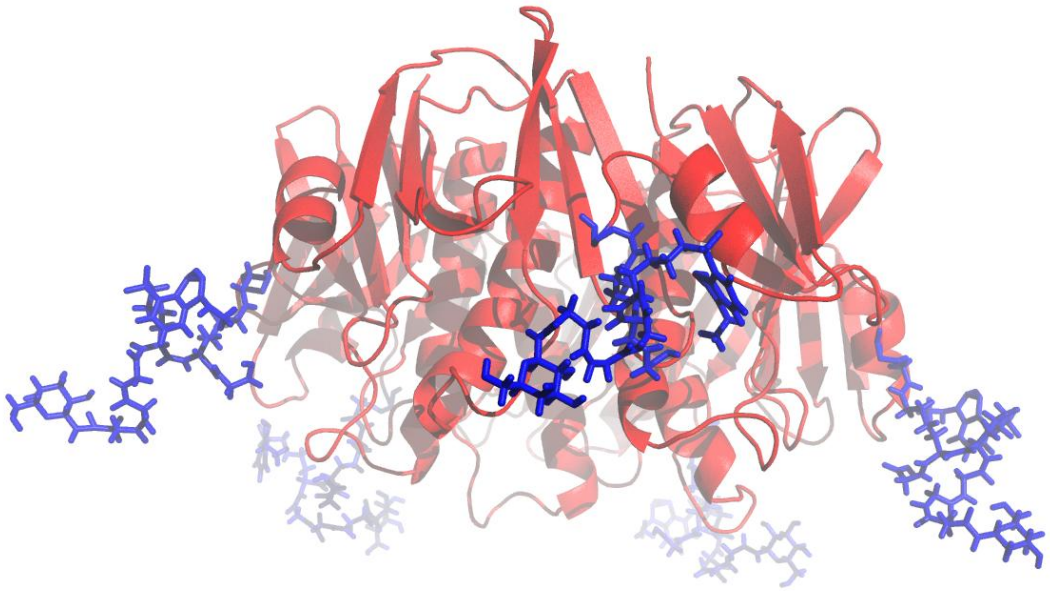
Fig 2.21 STD-NMR data showing the binding of O-blood group **35** to (A) LTB, (C) CTB and (E) CTB I47T. The binding of B-blood group **36** was seen with (B) LTB and (F) CTB I47T but not with (D) CTB.

2.4 Conclusions

The CTB protein can be expressed in a high yield and its size can be analysed by a variety of biophysical techniques. The carbohydrate binding sites have been well studied and the GM1os site was found to be especially suitable for potential ligand binding due to its location on the protein and high affinity interaction.

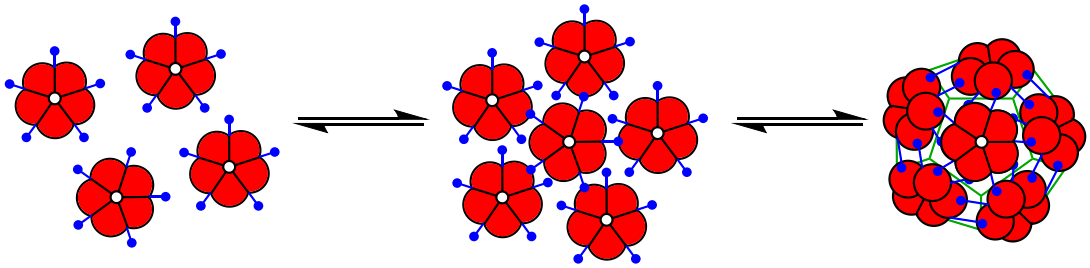
A combination of MS, SDS-PAGE, SEC, AUC, DLS, AFM and ITC can be used to characterise the size and shape of protein structures and ligand interactions.

3 Connecting the parts



CTB modification with carbohydrate ligands

The first strategy for assembling the proteins was to attach a self-complementary ligand at a suitable position on CTB. The ligands could cross-link adjacent pentamers via non-covalent interactions, with the ligand group attached to one pentamer binding to the carbohydrate binding site of another pentamer which in turn could provide another ligand to bind back (Scheme 3.1). Each of these divalent interactions should provide a constraint on the assembly process which could direct the assembly into the desired structure.



Scheme 3.1 Formation of a dodecahdron via protein-ligand interactions with modified CTB.

3.1 Protein modification

The N-terminus of CTB is situated on the protein surface towards the bottom of the side and almost directly between two of the GM1 binding sites (Fig 3.1). It has been shown that modifications at this position on CTB do not affect ligand binding to the protein.[146] It was therefore decided that this position would be perfect for modification as it provides five identical sites, equally spaced around the edge of CTB. If a galactose or larger GM1 oligosaccharide was attached at this position it could bind to another CTB unit and bring multiple copies together to form a larger protein complex. The N-terminal amino acid in CTB is a threonine residue which bears a unique amino alcohol group. Vicinal diols and, even more so, amino alcohols are easily oxidised with periodate to give an aldehyde.[147, 148] This strategy was previously applied by Chen *et al.* for the N-terminal modification of CTB.[146] The resultant aldehyde could then be reacted with an aminooxy functional group forming an oxime bond. Therefore any ligand of interest could be synthesised with an aminooxy group to be attached to the N-terminus of CTB.

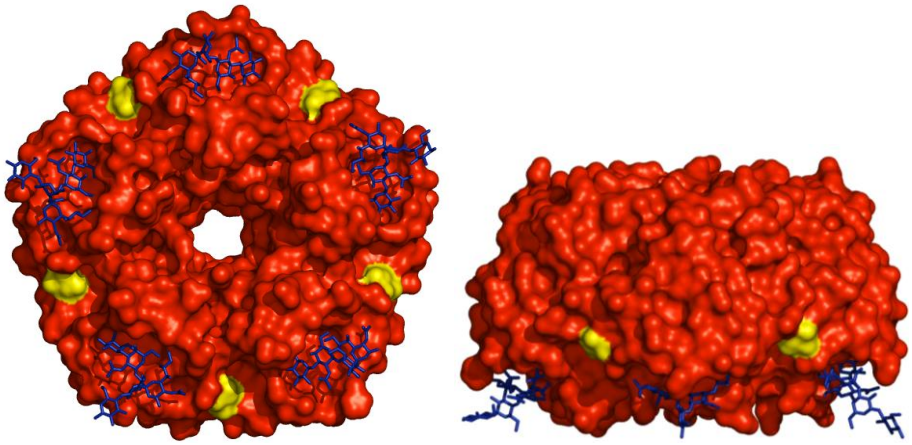


Fig 3.1 A surface representation of CTB with the position of the N-termini (yellow) on CTB seen between the binding sites with GM1os (blue).

3.1.1 Oxidation of the N-terminus

The oxidation reaction was performed on CTB at a concentration of 200 µM using $NaIO_4$. The reaction was found to proceed to completion after only 5 minutes with 5 equivalents of $NaIO_4$. The oxidised CTB protein (CTBox) was then purified on a PD10 G-25 minitrap column.

MS was used to confirm the successful reaction. The product was found to have a mass of was 11616.8 Da (calculated 11616.2 Da) indicating a mass loss of 27 Da which corresponded to the hydrated form of the CTBox product (Fig 3.2) (Scheme 3.2). The aldehyde was prone to hydration in aqueous solution due to the neighbouring carbonyl group which makes the aldehyde carbon more electrophilic. Fortunately, hydration is an equilibrium process and therefore reversible so that the aldehyde was still accessible for further reactions.

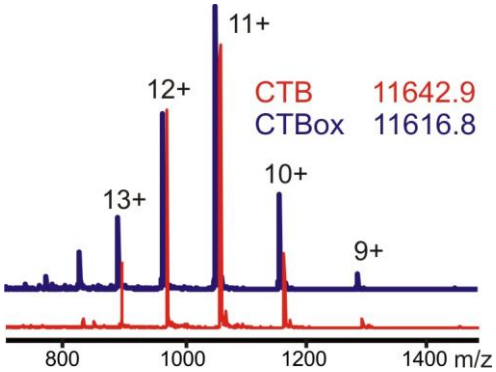
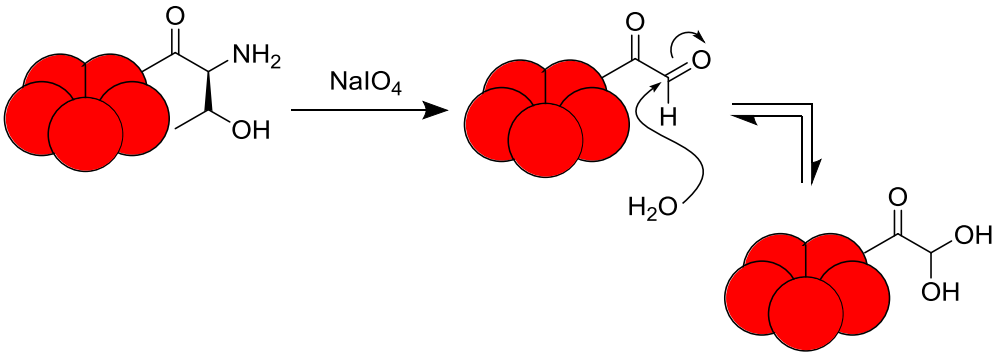
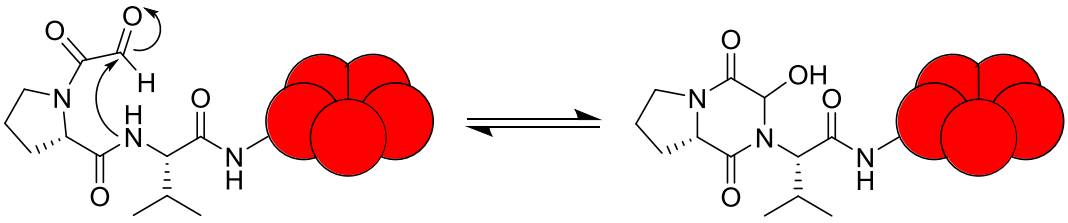


Fig 3.2 Mass spectrum of CTB and CTBox in the hydrated form showing a small change for each charge state observed by ESMS.



Scheme 3.2 Oxidation reaction on the N-terminus of CTB and the hydration of the aldehyde group. For simplicity only one subunit is shown in the reaction, whereas in reality all five subunits react.

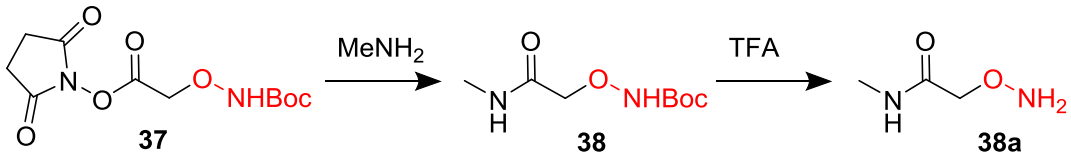
It has been reported that once the N-terminus of CTB has been oxidised to an aldehyde group, a slow cyclisation process can lead to the formation of an unreactive hemiaminal.[149] The proline residue at position two in the protein adopts a conformation that brings the aldehyde back into close proximity with the amide nitrogen of the third residue (Scheme 3.3). The cyclised form has an identical mass to the oxidised protein and so it is impossible to differentiate between these two species by MS. But this problem can be avoided if further reactions are performed immediately.



Scheme 3.3 The cyclisation of the N-terminus of CTBox.

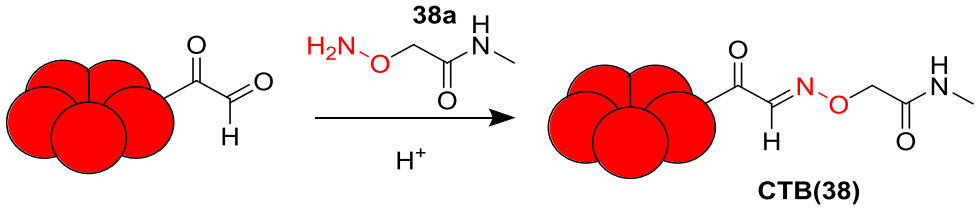
3.1.2 Oxime formation on the N-terminus

An aldehyde functional group on the N-terminus of CTBox can react with an aminooxy containing molecule to form an oxime group which is hydrolytically stable.[150] A simple aminooxy compound was synthesised to test this reaction on the protein. N-hydroxysuccinimide ester **37** was reacted with methyl amine to give amide **38** which was deprotected using trifluoroacetic acid (TFA) to yield oxyamine **38a** (Scheme 3.4). The compound is labelled with an **a**, after removal of the Boc group, to represent the deprotected version. This nomenclature will be used for all subsequent oxyamine compounds.



Scheme 3.4 Synthesis of amide **8**.

CTB was oxidised as before and then immediately in the same reaction vessel without any further purification, amide **38a** (15 equivalents to B-subunit) was added to the solution of CTBox at 200 µM and the pH was lowered to 2.5 by the addition of acetic acid (Scheme 3.5). After 30 min a sample was analysed by ESMS which gave a mass of 11684.5 Da (calculated 11684.3 Da). No oxidised protein was observed by ESMS and therefore the reaction was taken to have gone to completion with the successful attachment of the small amide to form protein **CTB(38)** (Fig 3.3). This reaction was simple, quick and the oxidation and oxime formation could be done consecutively without intermediate purification.



Scheme 3.5 The modification of CTBox with ligand **38a** to form an oxime bond

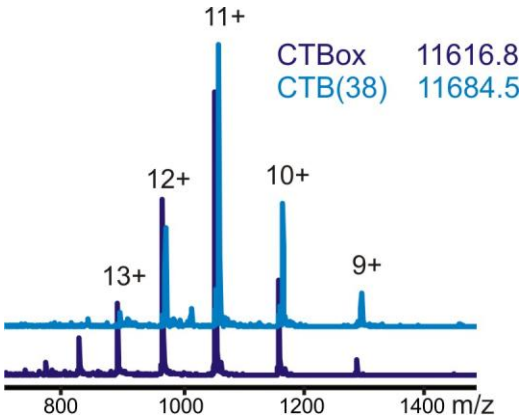


Fig 3.3 Mass spectrum of the new modified protein **CTB(38)** compared to CTBox.

SDS-PAGE was performed to determine the size of the modified protein (Fig 3.4). The gel showed that the protein had fully denatured into monomers rather than the usual pentamer due to the acidic conditions of the reaction. SDS-PAGE is performed under slightly denaturing conditions and so the result of only monomers

being observed may not be wholly representative of the sample in solution. However, it has been seen that CTB denatures into monomers below pH 4.[136] There are known methods for renaturing the protein from acidic conditions including a quick buffer change to a neutral pH and incubation at this new pH[151]. The original study by the Rose group to modify CTB with an oxime bond used SEC to purify their modified proteins.[146] Before purification, guanidinium hydrochloride was added, which denatures and solubilises the proteins. This study showed that it was possible to recover the pentamers with SEC even after deliberately denaturing them. However, it would be preferable to keep the pentamers intact as without stable modified pentamers there could be no interaction available with the GM1os binding sites and no formation of larger protein assemblies. Hirst and co-workers showed that denaturing CTB for any significant amount of time had a detrimental effect on the pentamer yield after renaturing.[152] This is due to a proline residue at the protomer interface that can isomerise when CTB is in the monomeric form and then prevents the pentamer reforming.

With the introduction of carbohydrate ligands, protein-ligand interactions should occur as soon as they are attached. However, if the proteins are in a denatured state they will be unable to bind ligands, thus preventing any possible interactions and assemblies. The stability of the pentamers was paramount to the project goal; without stable modified pentamers there could be no interactions possible with the GM1os binding sites and therefore no formation of larger protein assemblies.

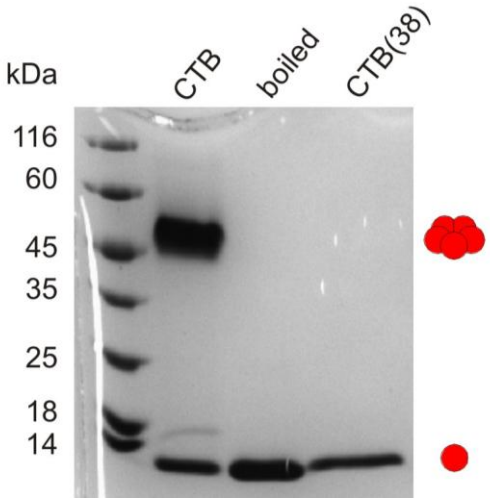
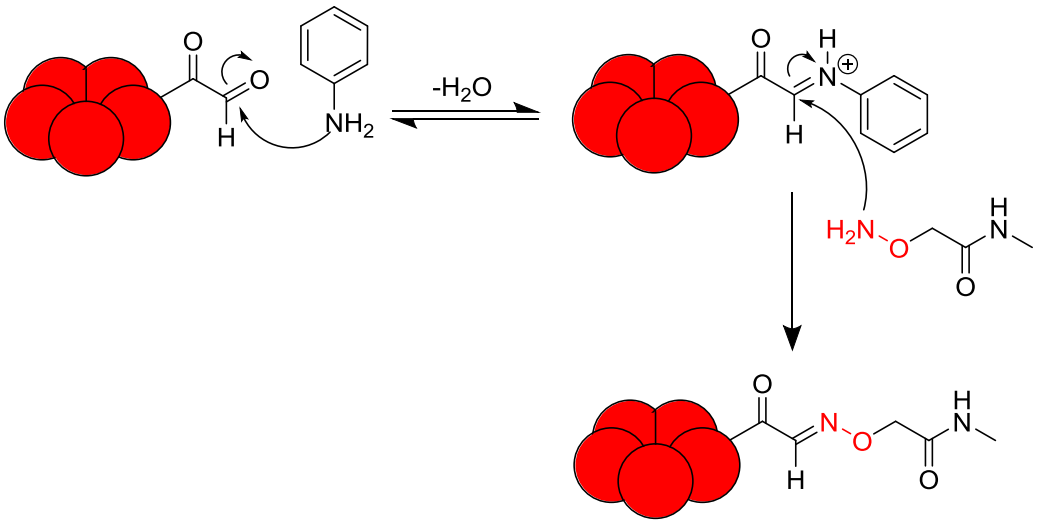


Fig 3.4 SDS-PAGE of the modified protein showing that following the acidic reaction conditions the protein was no longer stable as a pentamer on the gel.

To keep the CTB pentamer intact, the pH of the oximation reaction was raised to pH 4. CTB was oxidised as before at pH 7 and then transferred into a sodium acetate buffer at pH 4 via a centrifugal concentrator. Amide **38a** (15 equivalents) was added to this solution and the reaction was monitored by ESMS. After 4 hours it was found that the reaction had progressed only minimally and so it was concluded that the conditions were not acidic enough for the reaction to proceed successfully.

The modification needed to be performed at a more neutral pH to be sure that CTB did not denature, but another procedure had to be found. Oxime bond formation can be enhanced by the addition of aniline, which has been shown to catalyse the reaction.[52] Dirksen *et al.* demonstrated that in aqueous solution at pH 4.5 the rate of ligation of two peptides increased 400-fold when aniline was present and at pH 7 a moderate increase of 40-fold could be achieved. Under acidic conditions the oxime reaction progresses by nucleophilic attack by the aminooxy group on the protonated carbonyl followed by dehydration. The concentration of the protonated carbonyl is however extremely low at neutral pH due to its low pK_a in the range -4 to -10. An aminooxy amine has a pK_a, of 4.6 due to the α-effect, which is similar to aniline. The pK_a of the Schiff base formed by the aniline is only 2 units lower whereas that of the imine formed by an aminooxy is 5-6 units lower. This means that the aniline Schiff base is more significantly protonated at a given pH and can provide a more reactive intermediate for the oxyamine to attack. This reaction leads to an unprotonated oxime which is unreactive to hydrolysis above pH 2 (Scheme 3.6).



Scheme 3.6 Aniline catalysis of the oxime bond formation on the N-terminus of CTB.

The oxime reaction was performed at pH 7 with amide ligand **38a**. CTB was oxidised as before and amide ligand **38a** was added (15 equivalents). The sample was split in two and aniline was added to one batch to a final concentration of 100 mM. After 3 h in the aniline solution, ESMS showed a mass of 11684.0 Da (calculated 11684.3 Da) for **CTB(38)** and so the protein had fully reacted. The sample at pH 7 with no aniline present showed only a mass of 11616.6 Da (calculated 11616.2 Da) for CTBox indicating that it had not reacted at all. Although the reaction was a little slower, oxime formation could be achieved effectively at pH 7 with the aniline catalyst present.

The reaction was analysed by SEC (Fig 3.5) and the fractions collected were submitted to analysis by SDS-PAGE to confirm their size (Fig 3.6). Electrophoresis confirmed that CTB stayed mostly as a pentamer during the reaction and SEC showed the pentamer could be isolated. The protein was also separated from any unreacted ligand by SEC, which meant this method could be used not only for analysis but also purification of the modified proteins.

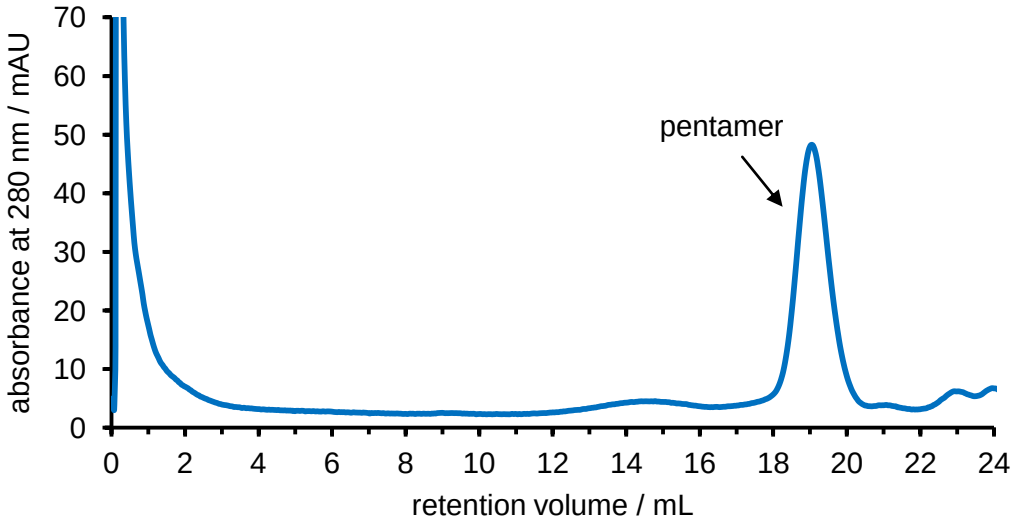


Fig 3.5 Chromatogram showing **CTB(38)**. The protein appears to remain as a stable pentamer under the new conditions.

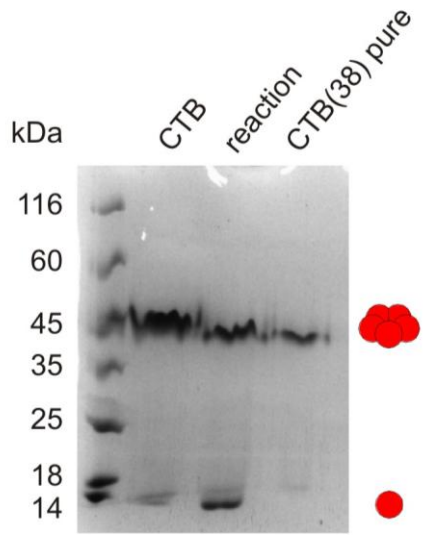
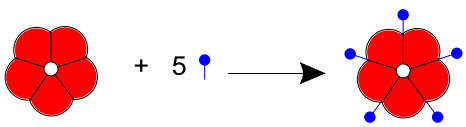


Fig 3.6 SDS-PAGE showing **CTB(38)**, modified at pH 7 with aniline as a catalyst, remained mostly as pentamer during the reaction and after SEC purification the pentamer was isolated.

3.1.3 Introducing carbohydrate ligands

A carbohydrate ligand was synthesised for attaching to the protein N-terminus. For the initial studies a galactose moiety was chosen as this has a low affinity for CTB and therefore could provide a reversible interaction. Low affinity ligands should be less prone to inducing protein dimerisation and should allow reversible self-correction of undesirable interactions that could lead to aggregation. Galactose ligands could also be displaced from CTB by monovalent GM1os and so the disassembly of any structures formed should be possible in this way. With five carbohydrate groups attached to CTB, these ligands would be able to make five interactions with other proteins whilst the binding sites would accept another five ligands from other proteins. Many simultaneous interactions would take place forming a strong protein complex. The individual components would be both pentavalent acceptors and donors (Scheme 3.7).



Scheme 3.7 Modified CTB with five carbohydrate groups.

A chromophore was included in the initial ligand design allowing the attachment of the ligand to be monitored by UV spectroscopy and as an extra tag to allow detection of only the components containing this label (Fig 3.7). A dansyl group was chosen that absorbed at 330 nm, different to the 280 nm absorption of the protein.

An amino acid linker section was included which could be altered in a modular fashion to create longer or shorter ligands.

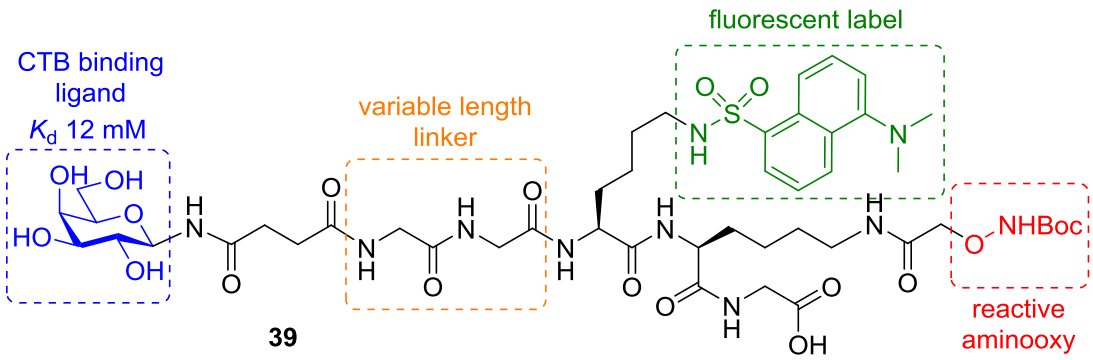


Fig 3.7 Structure of galactose ligand **39**.

Ligand **39** was constructed by solid phase peptide synthesis (SPPS) (synthesised by Dr Martin Fascione). Following removal of the Boc group, galactose ligand **39a** was then reacted with CTB to see if the carbohydrate groups could encourage assembly of the proteins into larger structures. The reaction conditions at neutral pH (100 mM aniline at pH 7) were used to react galactose ligand **39a** (10 equivalents) with CTBox at 200 µM monomer concentration. The oxime formation was monitored by ESMS and appeared to proceed more slowly, but after 24 h ESMS showed a mass of 12593.4 Da (calculated 12593.2 Da) indicating a complete reaction to form **CTB(39)** (Fig 3.8 B). The reaction solution had however turned slightly cloudy with some precipitation and so the mixture was centrifuged at 13,000 rpm for 1 min to remove the insoluble fraction. SDS-PAGE was then performed on the remaining solution (Fig 3.8 A) which showed the modified protein was still mostly in its pentameric form.

A sample of the soluble fraction was also was taken for analysis by SEC (Fig 3.9) and the trace showed pentameric modified protein. The dansyl group on the ligand absorbs at 330 nm and so this wavelength was also monitored by SEC. A peak with an absorbance at 330 nm was seen that coincided with the 280 nm peak confirming attachment of the ligand. A peak with a high absorbance at 330 nm was also seen in the SEC trace around a 30 mL retention volume, which was attributed to the excess unreacted ligand.

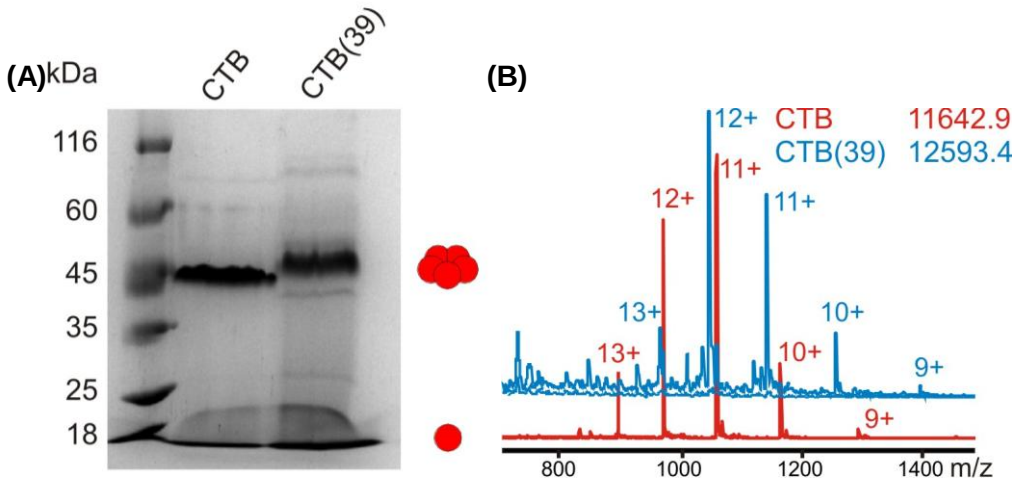


Fig 3.8 (A) SDS-PAGE showing that CTB modified with galactose ligand **39a** remained as a pentamer during the reaction. (B) Mass spectrum of CTB modified with galactose ligand **39a**.

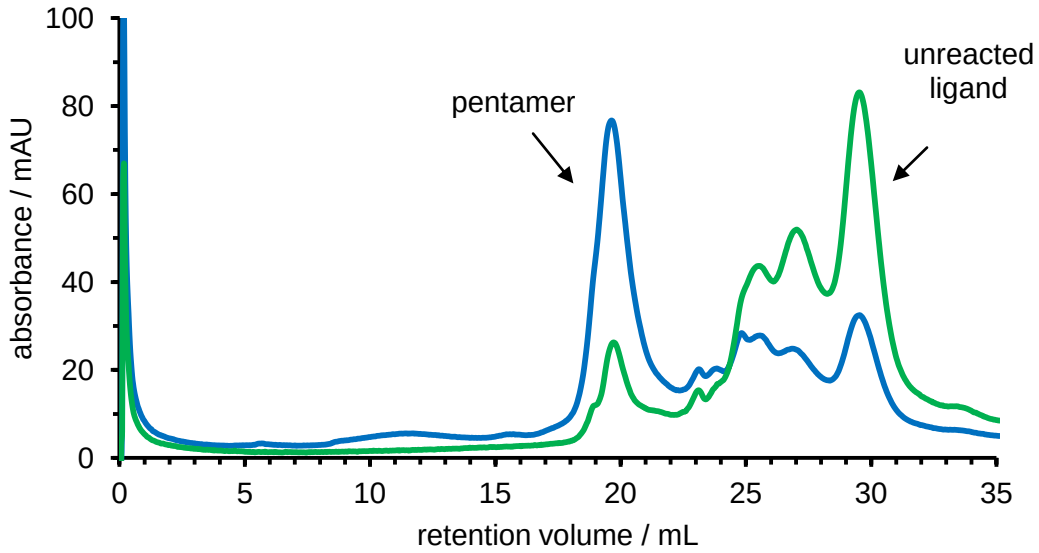


Fig 3.9 Chromatogram showing CTB modified with galactose ligand **39**, UV absorbance measured at 280 nm (blue) and 330 nm (green). The purified protein appeared as a pentamer with a strong absorbance from the dansyl group at 330 nm confirming ligand attachment. There are also peaks from the excess unreacted ligand giving a strong absorbance at 330 nm due to the dansyl group.

AUC was also used to evaluate the size of the modified protein **CTB(39)** and to check for larger assemblies. The pentamer fraction from SEC was concentrated to 10 µM and subjected to AUC analysis (Fig 3.10). The AUC results showed a peak at 4.2 S confirming the protein to be a pentamer but showed no formation of any larger structures.

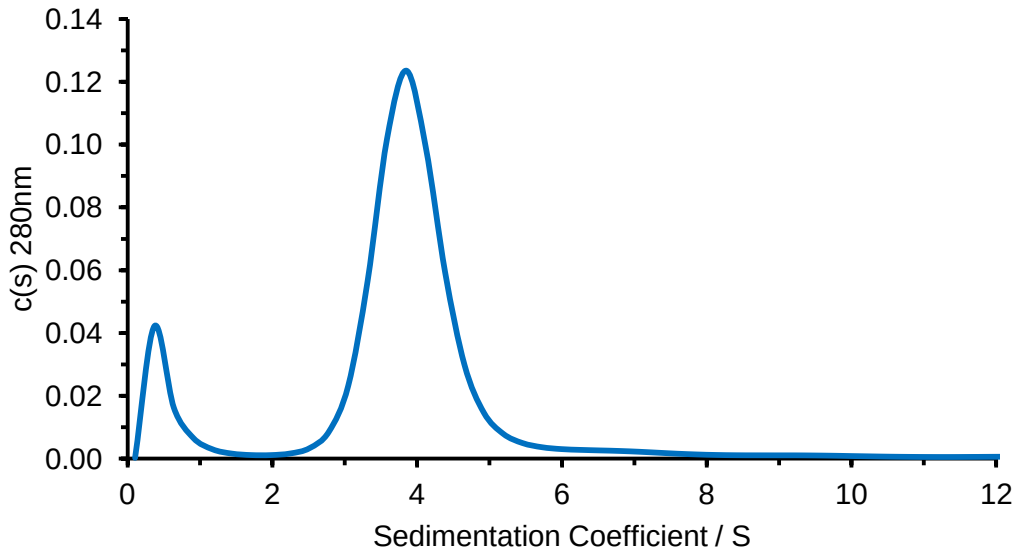


Fig 3.10 AUC sedimentation coefficient distribution of **CTB(39)** showing a species at 4.2 S which corresponds to a mass of 68 kDa for the modified protein pentamer.

3.1.4 The effects of concentration on protein assembly

Galactose ligand **39a** had been successfully attached to CTB but no protein assemblies had been observed in solution. To investigate whether the formation of larger complexes was concentration dependent, the CTB concentration was increased in the oximation reaction. CTB was oxidised as before but now at a higher protein concentration of 600 µM subunit concentration. Again, galactose ligand **39a** was attached via an oxime bond at pH 7 in *phosphate buffer* with 100 mM aniline present. This time the reaction progressed more quickly, as expected, and after only 1 h ESMS gave a mass of 12593.7 Da (calculated 12593.2 Da) indicating a complete reaction. During this reaction, a much larger amount of precipitation was observed than before when the CTB concentration was 200 µM. The reaction solution was analysed by SDS-PAGE (Fig 3.11) and most of the protein appeared as pentamer but a number of higher mass bands were also observed. SDS-PAGE is, however, not necessarily representative of what protein structures are present in solution. The intensity of the pentamer band on the gel does not imply that this is the dominant species in solution as other larger stable structures may have partially dissociated on the gel. But the gel does show that there are larger assemblies being formed and the precipitation observed was possibly from these larger insoluble protein aggregates (Scheme 3.8).

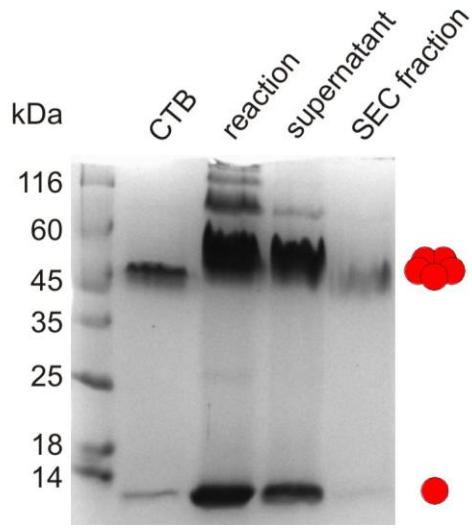


Fig 3.11 SDS-PAGE showing **CTB(39)** forming larger structures during the oximation reaction. After centrifugation these larger aggregates were removed and only pentamer and a small amount of dimer were observed in the supernatent. After purification by SEC only pentamer was seen.

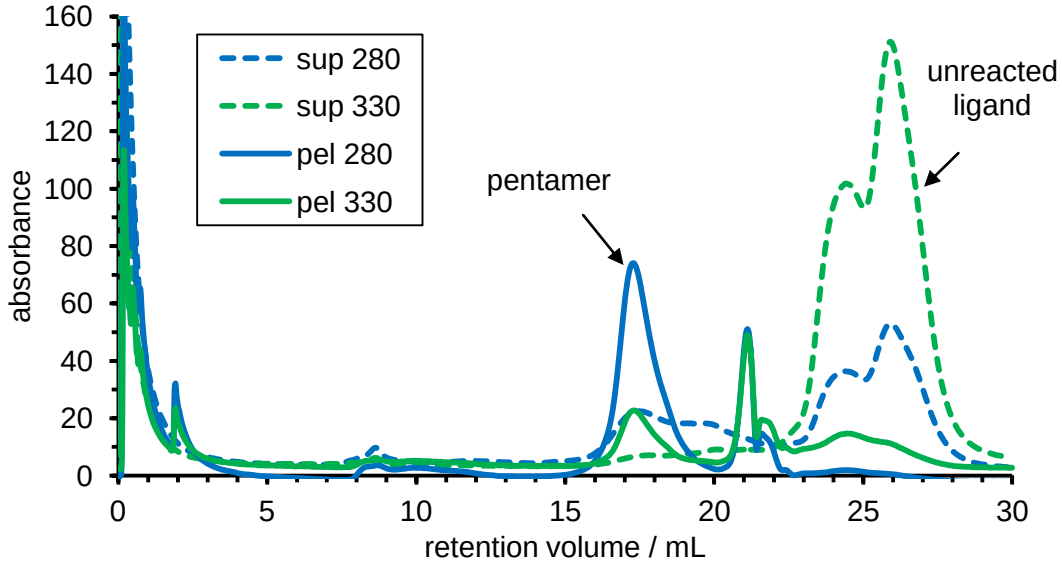


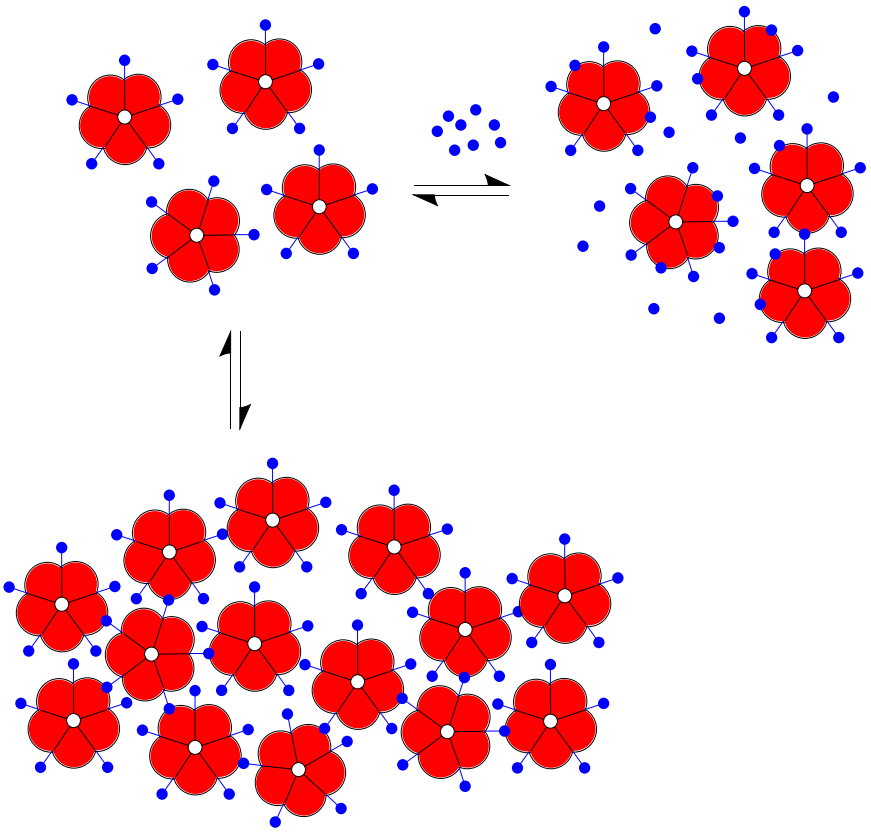
Fig 3.12 Chromatogram showing CTB modified with galactose ligand **39a**, after centrifugation of a sample to remove the insoluble precipitate. The UV absorbances were measured at 280 nm (blue) and 330 nm (green) showing minimal protein was soluble in the supernatant ("sup") and much more protein was recovered from the insoluble pellet ("pel").

The sample was centrifuged at 13 krpm for 1 min to remove the insoluble precipitate. The supernatant was analysed by SEC to find what remained in solution and *phosphate buffer* with 6 M guanidine hydrochloride was added to the pellet to redissolve the insoluble fraction before analysis by SEC (Fig 3.12). The SEC trace for the supernatant showed only a very small amount of protein pentamer and a large amount of the excess ligand. But much more protein was recovered from the pellet and much less excess ligand was observed. It was concluded that the

precipitate from the reaction did indeed contain the larger protein aggregates that were previously seen by SDS-PAGE and only minimal pentamer was stable in solution. Attempts were made to visualise the aggregates by using electron microscopy (EM), but no structures were observed.

3.2 Introduction of a competing ligand

The modified proteins were assembling into larger structures but this was uncontrolled leading to insoluble aggregates. At 200 µM CTB concentration there was minimal aggregation or precipitation but when this was increased to 600 µM most of the protein was seen to aggregate. It was of course necessary to gain a better control on the assembly process if the desired structures were to be achieved. A method was proposed to prevent all interactions with the introduction of a competing ligand (Scheme 3.8). If this ligand was present from the start at a high concentration it should bind to the CTB proteins and thus prevent the modified proteins from interacting.



Scheme 3.8 Uncontrolled aggregation of the modified proteins and the introduction of a competing ligand to prevent protein assembly.

The modification of CTB was again performed with galactose ligand **39a**, with CTB at a concentration of 600 µM which had previously led the proteins to precipitate. This time α-methyl galactoside **34** was included at 200 mM as a competing ligand to block the CTB carbohydrate binding sites. This simple galactoside has a K_d of 12 mM and so its concentration needed to be greater than the K_d to ensure complete binding. After 1.5 h ESMS showed the reaction to be complete and the mixture was analysed once again by SDS-PAGE (Fig 3.13). This time the analysis showed a band for the pentamer, a very light band for a larger structure that could have been a dimer of pentamers and another small band for monomers. Therefore, the gel indicated that the presence of the competing ligand prevented the protein from aggregating and had instead remained mainly as the pentamer.

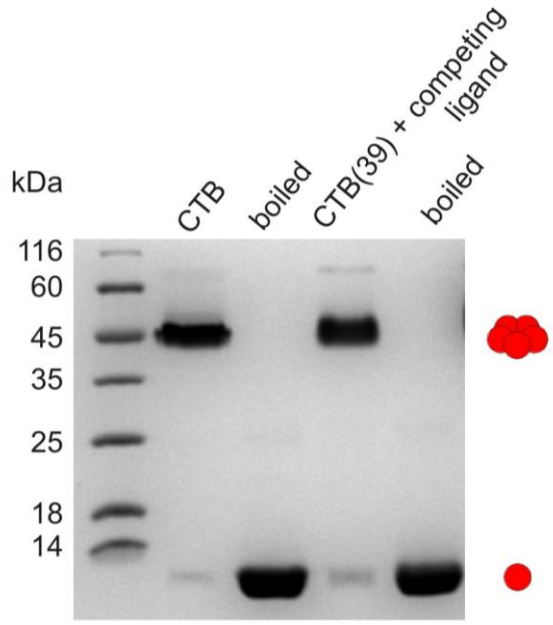


Fig 3.13 SDS-PAGE showing the modification of CTB with ligand **39a** remaining as a pentamer due to the presence of a competing ligand.

The concentration of the competing ligand was varied in an attempt to find a minimal concentration needed to prevent aggregation. The modification reaction was carried out at the protein concentration known to cause aggregation (600 µM) and the competing ligand concentration was varied from 0 to 200 mM in parallel reactions.

After completion of the reaction, samples were centrifuged at 13 krpm for 1 min and large pellets from the precipitated protein could be seen in the samples with 10 mM or less ligand and a very small pellet was observed in the other samples. The precipitate could not be analysed further but again showed how the concentration of

competing ligand was important for the proteins to remain in solution. After the samples were centrifuged, SDS-PAGE was performed to assess what remained in solution (Fig 3.14). The gel could also be visualised under a UV lamp which would highlight the dansyl groups of the ligand. This UV analysis of the gel was more revealing as it was more sensitive. It could be seen that with no competing ligand present most of the protein had precipitated and there was very little protein left. The amount of soluble protein increased with the concentration of competing ligand but small amounts of higher aggregates still showed. This observation confirmed that the competing ligand was necessary to keep the modified proteins in solution at a high protein concentration and inhibit the large scale aggregation. However, even with a competing ligand above 50 mM, some aggregation did still occur which was uncontrolled and tended to precipitate. The complementary nature of the binding sites and ligands was so strong that it was possible to overcome the great excess of free competing ligand.

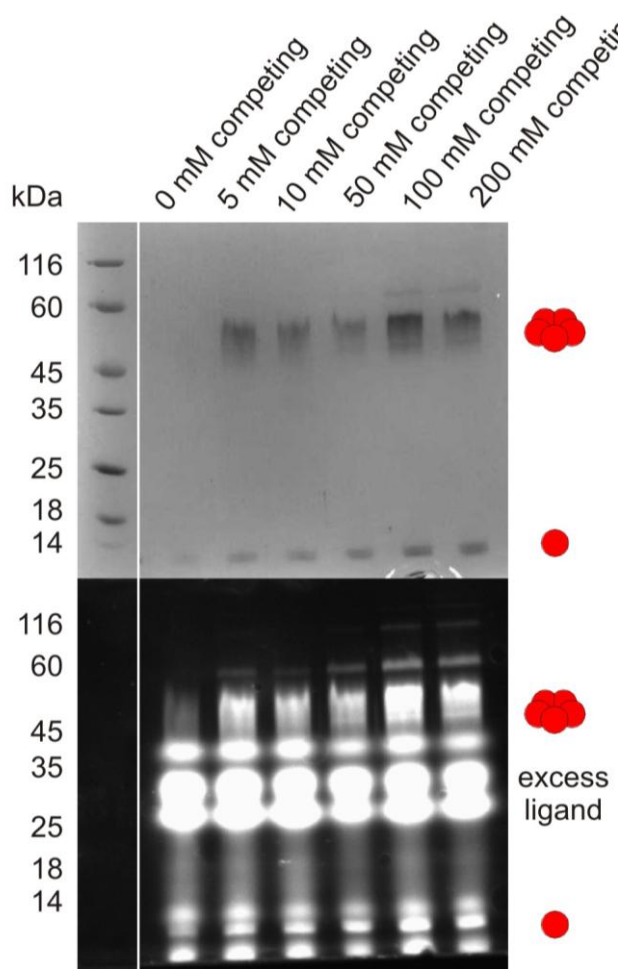


Fig 3.14 SDS-PAGE of the supernatants after centrifugation of the reactions. More protein is seen to remain in solution with more competing ligand but some larger assemblies are also still seen. Ligand alone in the absence of protein was seen to run at the same high position as shown above.

As it was shown that a competing ligand could inhibit most of the aggregation, it should also be possible to initiate the interactions between proteins by removing this competing factor. CTB was again modified at 600 µM, with the competing ligand **34** at 200 mM and no precipitation was observed. The sample was then dialysed extensively with four buffer changes over four days to completely remove competing ligand **34** and excess reacting ligand **39a**. It was found that much of the protein had precipitated during the dialysis and the concentration of soluble protein had reduced from 600 µM to 112 µM which was a loss of 81% of the total protein due to precipitation.

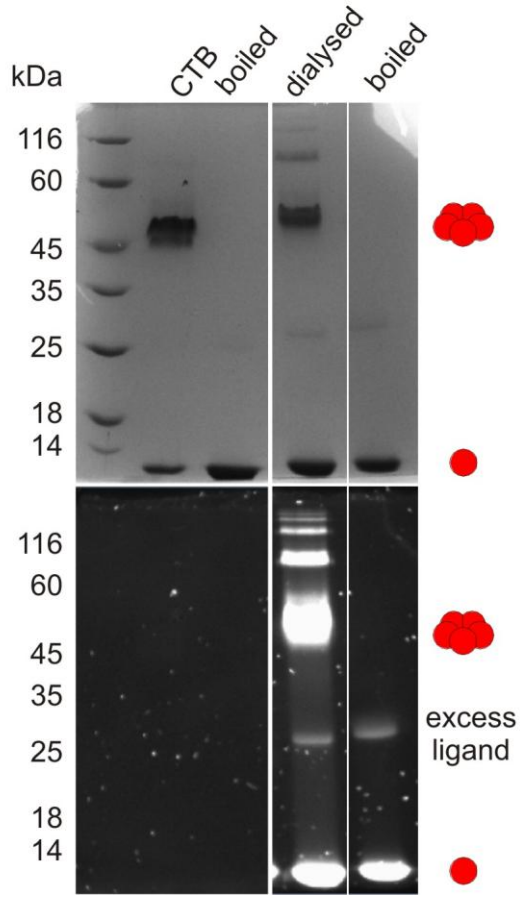


Fig 3.15 SDS-PAGE of **CTB(39)** modified with competing ligand and then dialysed extensively to remove the competing ligand and excess ligand **39a**. The protein was seen to aggregate after the removal of the ligand.

The sample was analysed by SDS-PAGE after the dialysis which now showed the modified protein pentamer, a number of larger complexes and minimal excess ligand remaining (Fig 3.15). Over the time course of the dialysis the **CTB(39)** proteins had been able to interact more and more resulting in the aggregation. It is

likely that this aggregation continued until the protein concentration was sufficiently low that the galactose binding interactions were no longer able to bring the separate pentamers together.

It was postulated that ligand **39a** might be long and flexible enough to reach round and bind back on the CTB protein it originated from. This self-binding may have been an entropically favourable interaction that would have had the consequence of preventing some aggregation of the pentamers (Fig 3.16). This binding may also explain why not all of the protein precipitated from solution when no competing ligand was present. The effective concentration of the attached ligands would be much higher than the overall protein concentration and so they may self-bind and when bound act themselves as competing ligands. ITC was performed with the sample of **CTB(39)** after extensive dialysis to see if any intra pentamer interactions could be detected. GM1os was titrated into the modified protein and any inter- or intra-ligand interactions would have competed with the GM1os.

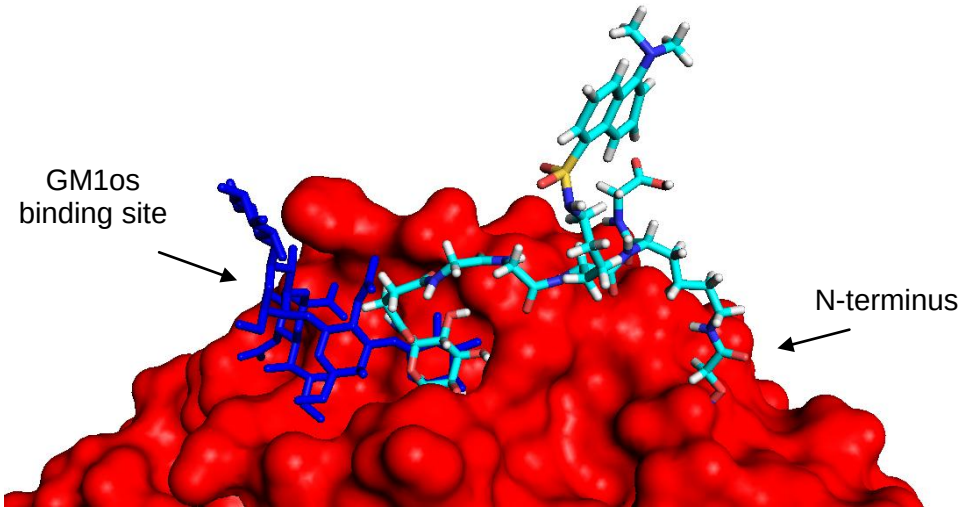


Fig 3.16 Model of **CTB(39)** with attached ligand **1**(cyan) adopting a possible conformation to self-bind to the nearest binding site on the same CTB pentamer (red) that it originates from. GM1os (blue) is also shown in the binding site for clarity.

The ITC titration showed little difference from the titration with wild-type CTB, with a K_d of 87.8 nM and Δ°H of -16.9 kcal mol^{-1}, and so no competition with the GM1os binding was observed (Fig 3.17). The titration was performed at a concentration of 3.5 μM for **CTB(39)** and therefore at this low protein concentration it could be concluded that the ligands did not have a high enough affinity to form inter-protein interactions. The result also showed that no intra-protein interactions were taking place as this would be independent of the overall protein concentration. Self-binding

interactions do not occur and the precipitation, or lack of, is due solely to the affinity of the galactose ligands for other CTB units and the self-complementary cross-linking ability of **CTB(39)** at the higher concentrations.

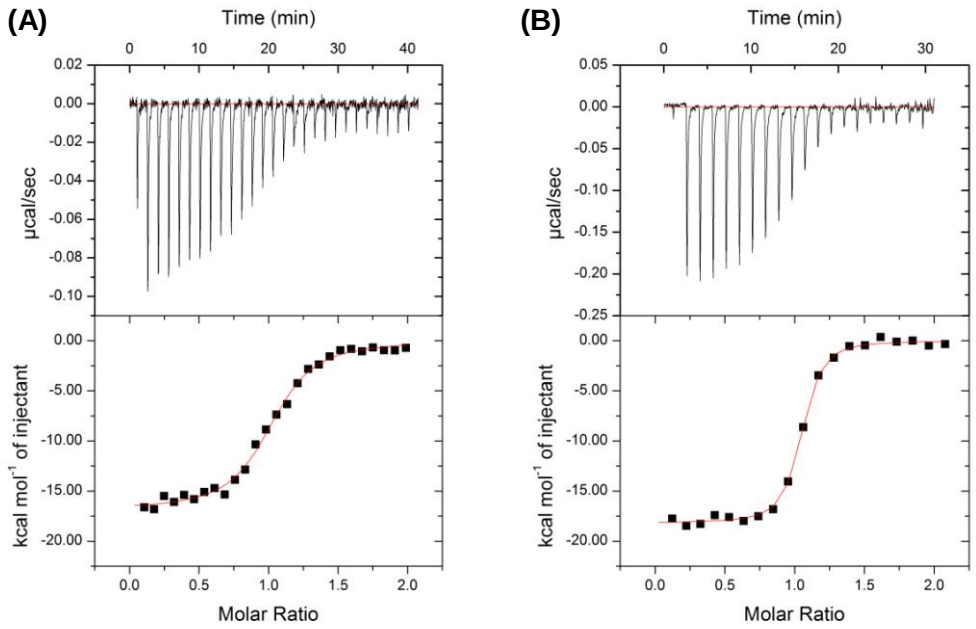


Fig 3.17 (A) ITC titration of GM1os (50 µM) into **CTB(39)** (3.5 µM). The K_d of 87.8 nM and Δ°H of -16.9 kcal mol^{-1} were similar to the wild-type CTB and GM1os interaction indicating that no intra- or inter-protein interactions were being made by the attached ligands. (B) The wild-tye CTB titration shown for comparrison.

3.3 The importance of ligand length

The length of the ligand may be important in determining how the proteins can interact. A range of ligands were synthesised to assess how the ligand length might be affecting the aggregation (Fig 3.18). With a longer flexible ligand, the CTB pentamers could come together in many different orientations giving rise to an unrestricted, uncontrolled aggregation as seen with ligand **39a**. However, if the ligands were too short then there may not be enough room for the proteins to interact in the desired arrangement. To achieve the goal of creating a dodecahedron from CTB proteins, there has to be some constraint in how the ligands are presented to bind proteins, resulting in only the desired complex structure.

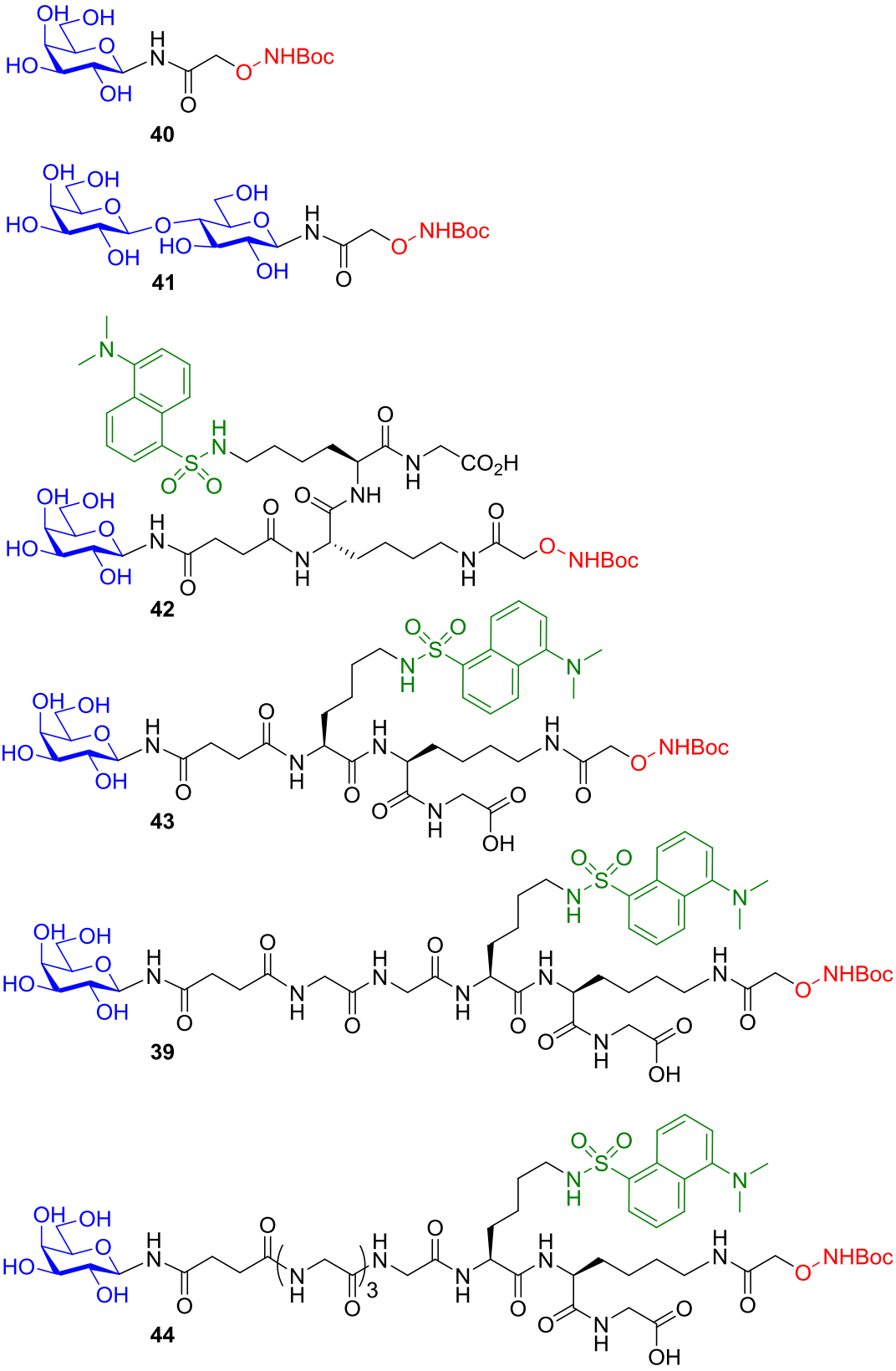
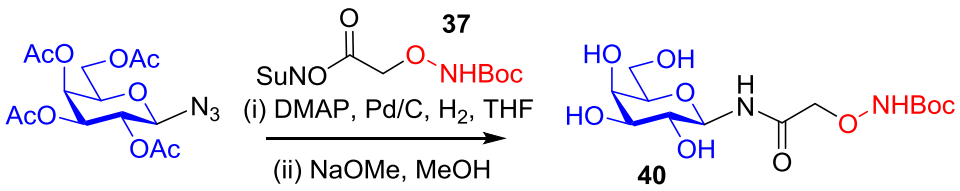


Fig 3.18 Different length galactose ligands.

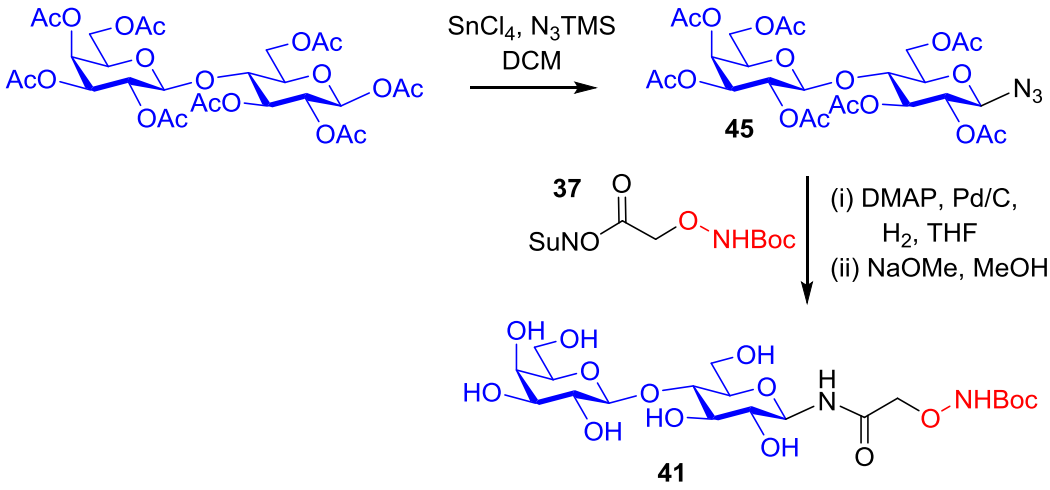
The shortest ligand investigated comprised of the amino oxy group attached via an amide bond to a galactose sugar. Acetylated galactosyl azide (provided by Dr. Martin Fascione) was hydrogenated using Pd/C and H_2 (g), in the presence of

N-hydroxysuccinimide ester **37** and DMAP (Scheme 3.1). The resulting galactosyl amide was then deacetylated under Zemplén conditions to yield the short galactose ligand **40**.



Scheme 3.1 synthesis of ligand **40**

A step up in length was achieved by using a lactose unit contributing an extra monosaccharide to the length. Acetylated lactose was reacted with TMS azide in DCM using $SnCl_4$ as a lewis acid to activate the anomeric acetate and produce lactosyl azide **45**. This was then hydrogenated using Pd/C and H_2 (g), in the presence of N-hydroxysuccinimide ester **37** and DMAP. Zemplén conditions were used for deacetylation to yield lactose ligand **41** (Scheme 3.2).



Scheme 3.2 Synthesis of ligand **41**.

Peptide ligand **43** was constructed as a short peptide using SPPS on 2-chlorotrityl resin preloaded with glycine for ease of synthesis. The resin was chosen as it is very acid labile and so the peptide ligand could be selectively cleaved without losing the Boc group from the oxyamine. Firstly Dde-protected lysine **46** was attached to the glycinyl resin and the Fmoc protecting group was removed with piperidine/DMF. Next dansyl lysine **47** was attached and again the Fmoc protection for the terminal amine was removed using piperidine/DMF. Succinic anhydride was then added to the linker followed by galactosylamine **48**. The Dde protecting group was removed

selectively with NH_2NH_2/DMF before the N-hydroxysuccinimide ester **37** was attached to the lysine side chain. Finally the ligand was cleaved from the resin using HFIP/DCM, before purification by SEC resulting in galactose ligand **43**.

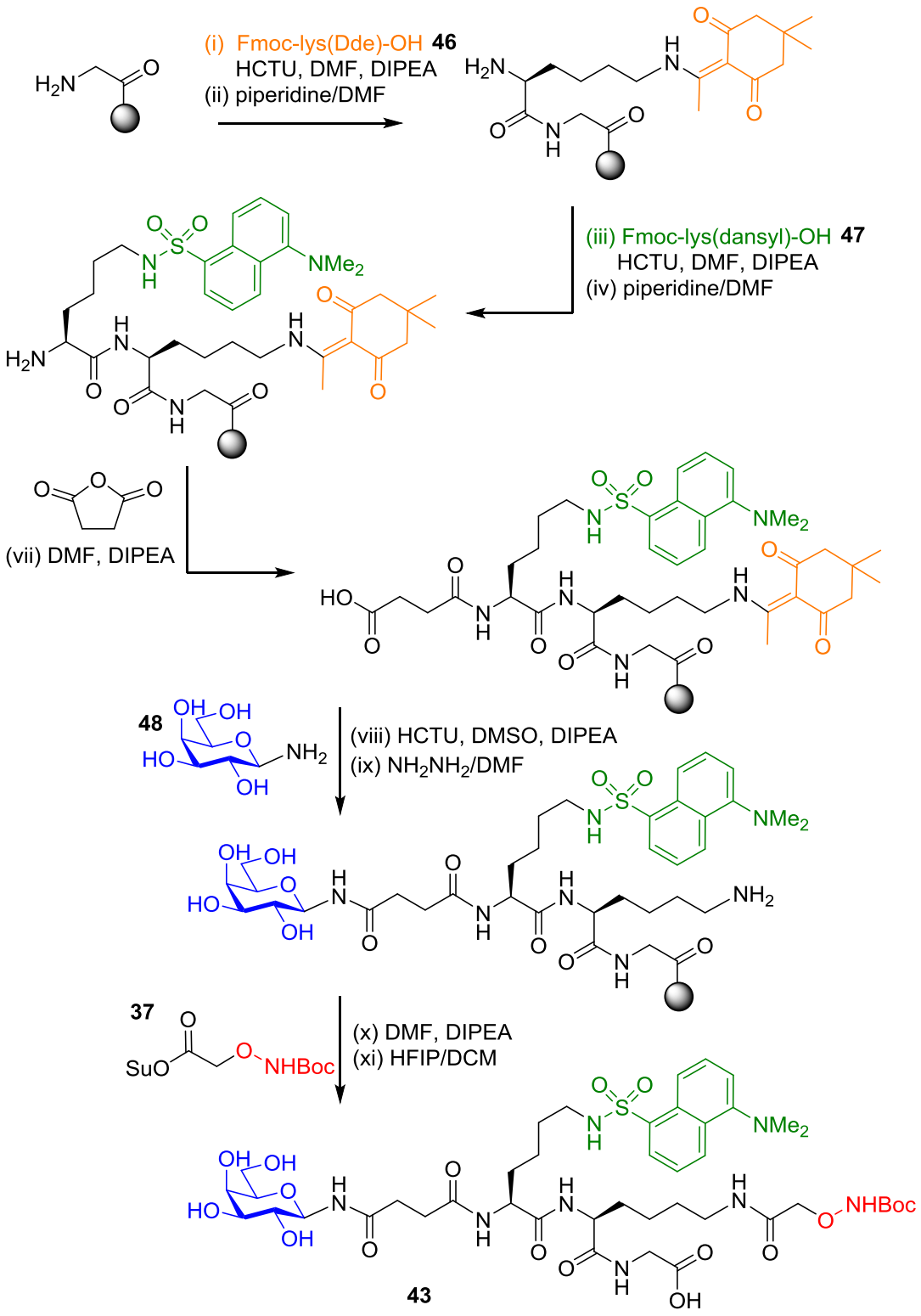


Fig 3.19 Synthesis of galactose ligand **43** by SPPS.

Ligands **42** and **44** (synthesised by Dr. Martin Fascione) were constructed using SPPS in the same way used for ligand **43**. The change in length from ligand **42** to ligand **43** involved a switch in the order of the dansyl lysine and lysine Dde amino acids to give an extra amide bond between the galactose moiety and the aminooxy group. Two glycine residues were then included for ligand **39** and a further two glycine residues for ligand **44**.

3.3.1 Varying the ligand length

Short galactosyl ligand **40a** was successfully attached to CTB at pH 7 with aniline catalysis at 600 µM protein concentration without a competing ligand present. ESMS analysis showed a mass of 11829.7 Da (calculated 11832.4 Da) which was within the error for the ESMS, for the completed reaction yielding **CTB(40)** (Fig 3.20 A).

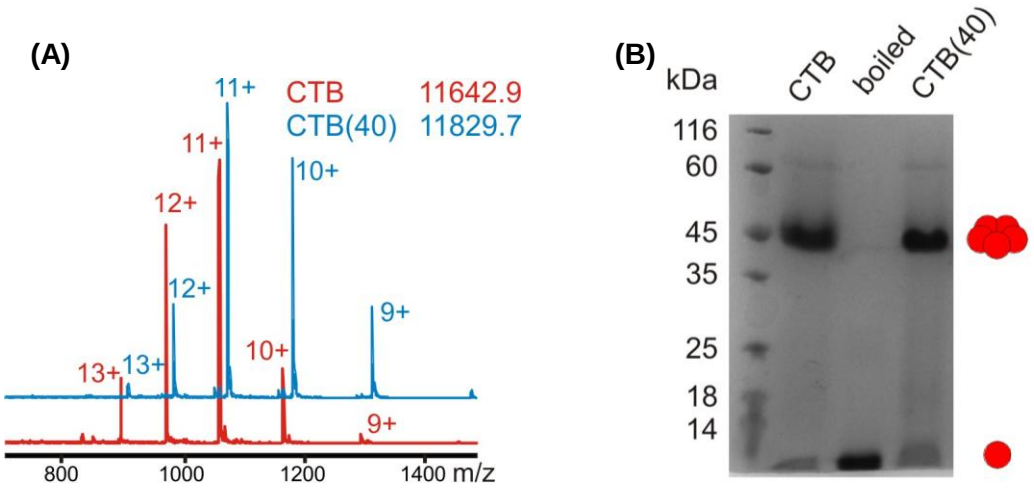


Fig 3.20 (A) Mass spectrum showing CTB modified with short galactose ligand **40a**. (B) SDS-PAGE of **CTB(40)** showing no aggregation with the short galactose ligand.

The reaction solution was analysed by SDS-PAGE and most of the protein appeared as pentamers (Fig 3.20 B). At the high CTB concentration where previously aggregation had been seen, only pentamer was visible with this ligand. To confirm that no larger species were being formed, AUC was performed on **CTB(40)** and showed the protein to be a pentamer with a mass of 54 kDa (Fig 3.21). As **CTB(40)** remained as pentamer it was concluded that the ligand was too short for the proteins to interact effectively and so did not form larger structures.

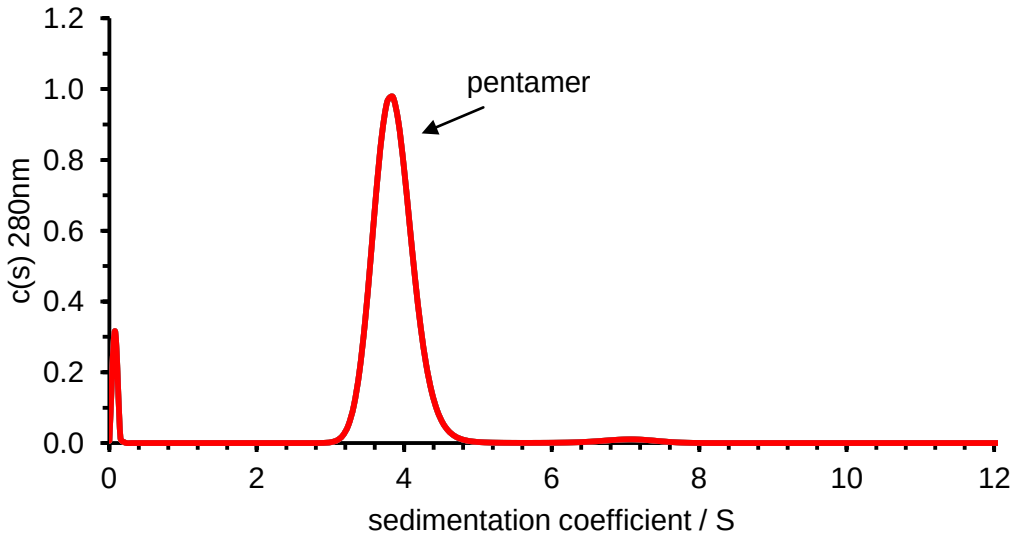


Fig 3.21 AUC sedimentation coefficient distribution of **CTB(40)**. The plot shows a species at 3.9 S which corresponds to a mass of 54 kDa for the modified CTB pentamer.

The next step up in length was to lactose ligand **41**. This ligand was attached to CTB at 600 µM with and without competing ligand. ESMS analysis gave a mass of 11992.1 Da (calculated 11994.5 Da) for the modified protein **CTB(41)** (Fig 3.22). The samples were analysed by SDS-PAGE which showed mostly CTB pentamer and analysis of a centrifuged sample showed that the protein did not precipitate and stayed in solution as stable pentamers (Fig 3.23). Again it was concluded that the ligand was not long enough to make substantial interactions and so no larger complexes were formed.

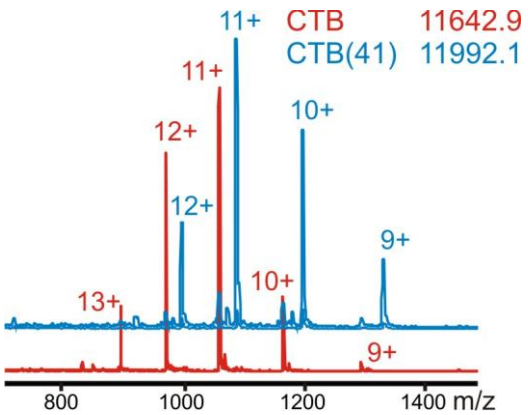


Fig 3.22 Mass spectrum showing CTB modified with lactose ligand **41a**.

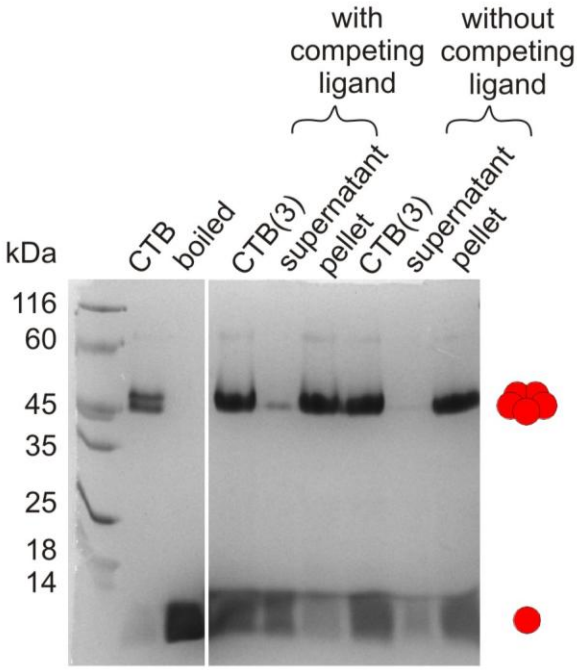


Fig 3.23 SDS-PAGE of modified **CTB(41)** modified with and without a competing ligand showing that all the protein stayed in solution and therefore did not aggregate.

The longer peptide based galactose ligand **42a** was attached to CTB. ESMS analysis showed the formation of **CTB(42)** with the mass of 12477.0 Da (calculated 12479.1 Da) for the modified protein (Fig 3.24). The reaction was performed at high (600 µM) and low (200 µM) protein concentrations to assess if this ligand would act in a similar way to ligand **39** at different concentrations. The reactions were also performed with and without competing ligand α-methyl galactoside at 200 mM.

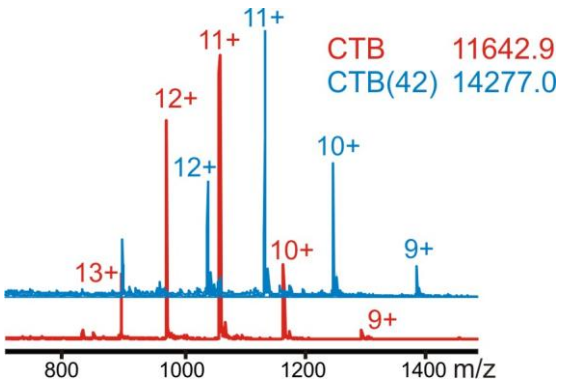


Fig 3.24 Mass spectrum showing CTB modified with longer galactose ligand **42a**.

The reaction samples were analysed by SDS-PAGE and samples were also centrifuged for 1 min at 13 krpm to assess the stability of the species in solution (Fig 3.25). It was seen that at 200 µM CTB monomer concentration with competing ligand, all protein was kept in solution as stable pentamers. When the competing ligand was not present at the lower protein concentration then a small amount

precipitated and was seen on the gel as monomers. Increasing the protein concentration to 600 µM increased the interactions between proteins and even with competing ligand present a small amount of the protein still precipitated and was seen in the pellet fraction on the gel. When the competing ligand was not present at 600 µM protein concentration, the majority of protein precipitated and was seen on the gel as monomers from the pellet. The gel showed how much protein was present in solution but did not give an accurate representation of the size of particles in solution and did not show any larger aggregates for the precipitated samples although this may have been due to only a small amount of sample loaded on the gel.

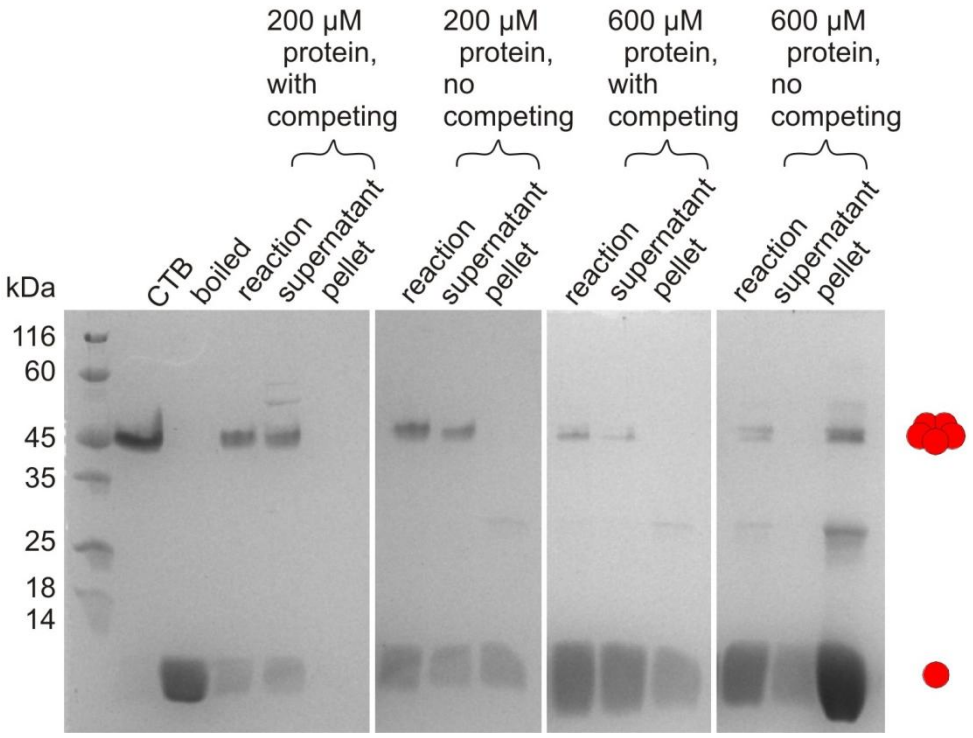


Fig 3.25 SDS-PAGE of CTB modified with ligand **42a**. Removing the competing ligand was shown to result in more precipitate in the pellet fraction especially at 600 µM protein.

The pellets from the different concentration and competing ligand conditions were again analysed by SDS-PAGE, only that this time a larger sample was applied to the gel in an attempt to get stronger bands (Fig 3.26). Electrophoresis now showed much more aggregation for the high concentration of **CTB(42)** without competing ligand similar to that seen by **CTB(39)**. A fair amount of protein had aggregated in the high protein concentration sample even with competing ligand and a similar situation was seen at the lower concentration with no competing ligand leading to more precipitation although not as much as at the higher concentration.

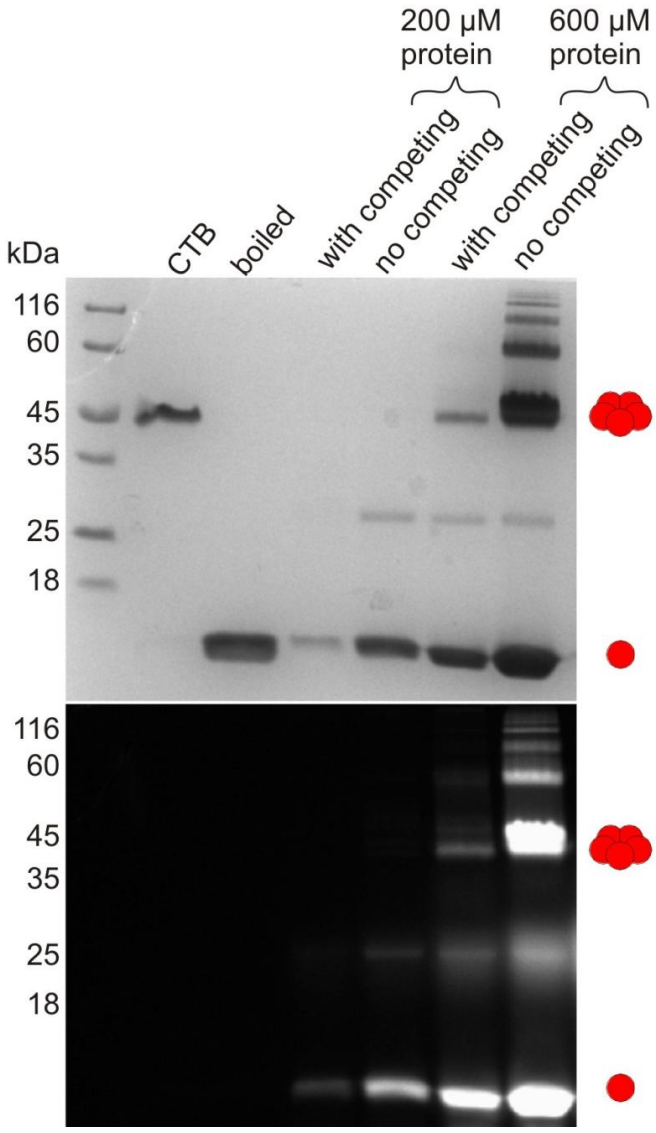


Fig 3.26 SDS-PAGE of pellets from **CTB(42)**. At the lower protein concentratio not much precipitated even without a competing ligand but at the higher protein concentration many larger complexes were seen on the gel.

Investigations were also performed with ligands **43** and **44**, but no differences were seen in the ability of the modified proteins to assemble compared to with ligand **42**. It appeared that with ligands as long or longer than ligand **42**, assembly of the proteins occurred but tended to form insoluble aggregates.

3.4 Conclusions

CTB can be site-specifically modified at the N-terminus first by oxidation and then by the formation of a stable oxime bond under acidic conditions or at pH 7 with an

aniline catalyst. CTB is stable at the neutral pH and so this improvement in the reaction allows for the attachment of carbohydrate ligands.

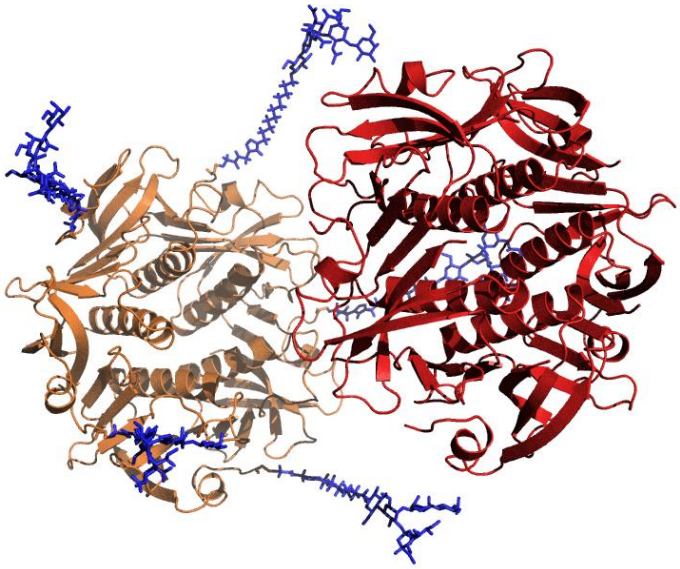
Protein assembly was shown via the attachment of a galactose ligand to CTB. This assembly is dependent somewhat on the concentration of the protein and can be mostly prevented with the introduction of a competing ligand. The length of the ligand was also found to be important as when it was too short, there was not enough space or flexibility for the interactions to take place. Unfortunately the structures formed have been uncontrolled aggregates that precipitated from solution. Any structural analysis of the larger structures proved difficult as only the stable pentamer at low concentration could be isolated from solution.

The initial strategy of creating a building block that has both multiple ligands and multiple binding sites may be too complicated to control. To form stable discrete particles, a “Goldilocks” set of conditions may need to be found. The correct concentration, the precise length of ligand and a suitable amount of competing ligand may all need to be varied systematically for the modified proteins to be controlled. The proteins were certainly assembling, but there was no control in this system and no constraints on what structures could form.

If the system could be created from two components, each with either the ligands or the binding sites, then this may be easier to control. Each component could be created individually before being combined in a specific way. Attempts to control the system are discussed in chapter 4.

Another method for encouraging the proteins to assemble into the desired structure would be to introduce a template. The proteins could then be prearranged before the ligands are added to crosslink the structure. This strategy should potentially give much more control over the size of the structures formed and prevent unwanted aggregation. Investigations into templating the system are discussed in chapter 5.

4 Controlling the system



The creation of a binding site mutant, its modification, characterisation and interactions with wild-type CTB

Creating a system that combined binding sites and complementary ligands resulted in uncontrolled aggregation. A new strategy needed to be devised to introduce an element of control into the system. The separation of ligands and binding sites into separate building blocks could provide this control. To achieve this goal, a mutant CTB protein had to be created that could not bind the carbohydrate ligands (Fig 4.1).

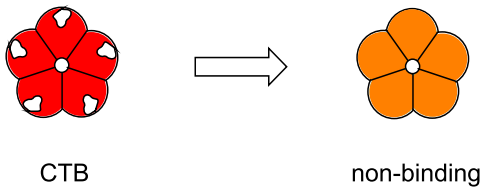
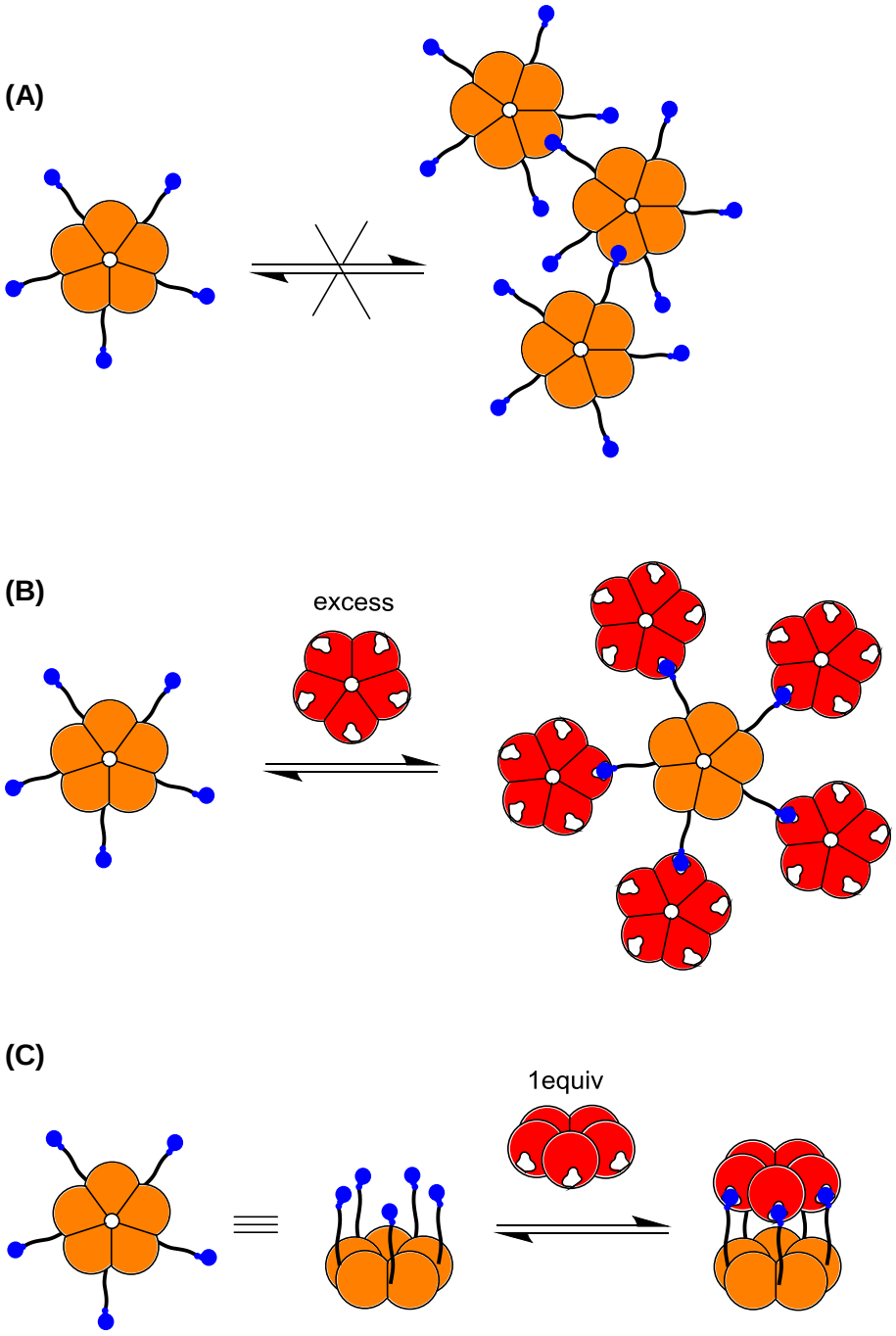


Fig 4.1 Representation of a mutant non-binding CTB protein.

This mutant CTB could then be modified with ligands in the same way as before (see chapter 3) resulting in a pentavalent protein based ligand. This ligand would not be able to bind to itself or others at the time of attachment but could still potentially interact with wild-type CTB (Scheme 4.1 A). Introduction of the wild-type protein with binding sites still intact, would then initiate the formation of larger protein complexes.

The creation of the mutant pentavalent ligand allows the possibility for controlling assembly when wild-type CTB is added. If an excess of the binding component was added to the ligand then it might be expected that five CTB units could bind around a single pentavalent ligand (Scheme 4.1 B). Or if an equal ratio of the ligand and receptor were used then one-to-one binding might be achieved (Scheme 4.1 C). The binding strength of the ligand groups could be varied to provide a more or less reversible system. The length of the linkers could also be altered to constrain the orientation in which the proteins interact and so favour one type of complex formation.

This chapter outlines the creation of a non-binding CTB mutant, its modification with a range of carbohydrate ligands and their interactions with wild-type CTB.



Scheme 4.1 (A) The non-binding mutant protein modified with ligands cannot form interactions by itself. (B) The assembly of five wild-type CTB proteins around a modified non-binding mutant protein. (C) The assembly of a 1:1 protein heterodimer.

4.1 Removing the binding capacity of CTB

Jobling and Holmes previously identified a number of mutations that change the binding properties of CTB.[153] A tryptophan residue (W88) sits in the GM1 binding pocket and forms a stacking interaction with the terminal galactose residue of GM1 (Fig 4.2). It was found that if this residue was mutated to a lysine or a glutamic acid residue then CTB no longer bound to GM1. Both these mutations introduce a

charged residue into the binding site and disrupt the stacking of the carbohydrate with CTB. It was therefore decided that these mutations would be introduced into CTB to create a non-binding protein.

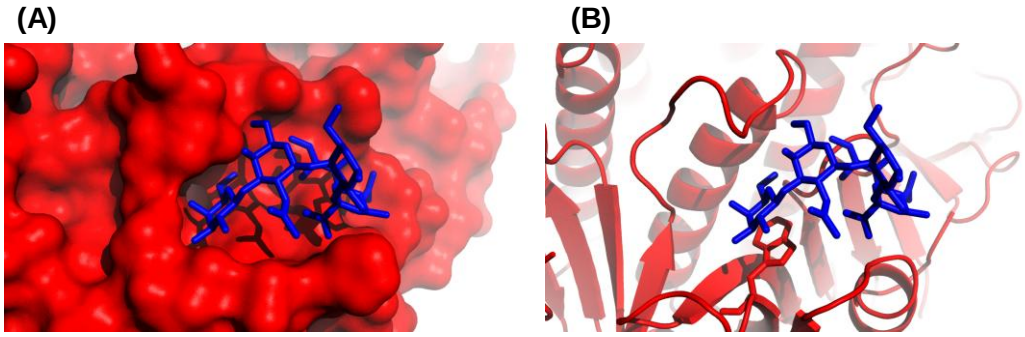


Fig 4.2 The carbohydrate binding site of CTB (red) shown with GM1os (blue) bound (A) with the surface representation and (B) without the surface and with W88 shown.

4.1.1 Making the mutants

The standard method for expressing wild-type CTB in the Turnbull lab used *Vibrio sp.60* bacteria with the gene encoded in the plasmid pATA13.[134] It was difficult to use this plasmid for introducing a mutation due to its large size and low copy number and so an *E. coli* system was used.

A pMAL-p5X plasmid was originally obtained from Genscript before an MBP-CTA2 fusion was introduced along with the gene for CTB, to produce plasmid pSAB2.1A. The MBP-CTA2 gene was then removed to produce plasmid pSAB2.2A containing only the CTB gene for expression in *E. coli*. A W88K mutation was then introduced to remove the carbohydrate binding capacity of the protein. However, the N-terminus of this CTB protein was an alanine residue. This CTB protein was encoded in plasmid pSAB2.2-T1AW88K (DNA manipulation performed by Mr James Ross).

Site-directed mutagenesis was performed on plasmid pSAB21.2-T1AW88K so that a threonine residue would be at position one of the protein and be suitable for N-terminal modification. This mutation produced plasmid pTRB-W88K. The binding site is in close proximity to the N-terminus and it was thought that a positive charge may affect reactions on the N-terminus. Therefore, a glutamic acid residue was also introduced at another position (Q61E), at the edge of the binding site, with the original W88K mutation still present to neutralise the overall charge and stabilise the protein (Fig 4.3). Plasmid pTRB-W88KQ61E encoded this new mutant protein. The

second mutation would only be a small structural change but should neutralise the overall charge, thus stabilising the mutant protein. A final mutation was introduced, into pTRB-W88K, to alter residue 88 and create a different mutant protein with a glutamic acid residue (W88E), producing plasmid pTRB-W88E. This change yielded a mutant protein with a negative charge instead of the positive charge from the lysine mutant.

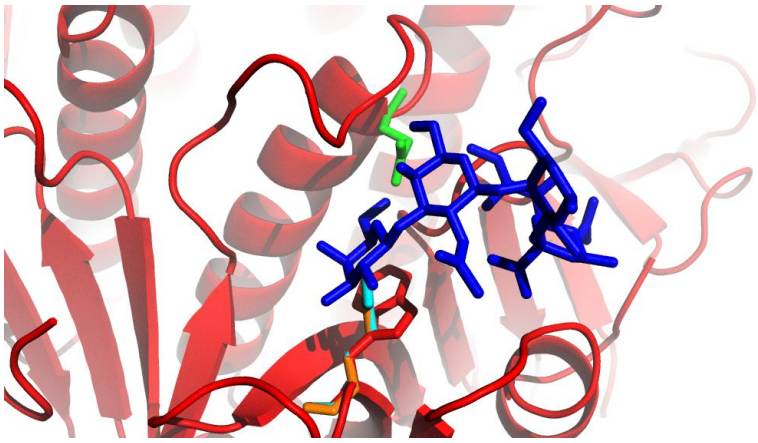


Fig 4.3 Representation of the CTB (red) binding site with GM1os (blue) bound. The tryptophan residue at position 88 in the wild-type is shown with the mutations W88K (light blue) and W88E (orange) superimposed. Mutation Q61E (green) is shown in close proximity in the binding site.

The amino acid sequences are shown below for all the CTB wild-type protein and the mutants described. Mutations from the wild-type sequence are underlined and highlighted in red:

	1	11	21	31
wild-type	TPQNITDLCA	EYHNTQIYTL	NDKIFSYTES	LAGKREMAII
W88K T1A	**A**PQNITDLCA	EYHNTQIYTL	NDKIFSYTES	LAGKREMAII
W88K	TPQNITDLCA	EYHNTQIYTL	NDKIFSYTES	LAGKREMAII
W88K Q61E	TPQNITDLCA	EYHNTQIYTL	NDKIFSYTES	LAGKREMAII
W88E	TPQNITDLCA	EYHNTQIYTL	NDKIFSYTES	LAGKREMAII

	41	51	61	71
Wild-type	TFKNGAIFQV	EVPGSQHIDS	QKKAIERMKD	TLRIAYLTEA
W88K T1A	TFKNGAIFQV	EVPGSQHIDS	QKKAIERMKD	TLRIAYLTEA
W88K	TFKNGAIFQV	EVPGSQHIDS	QKKAIERMKD	TLRIAYLTEA
W88K Q61E	TFKNGAIFQV	EVPGSQHIDS	**E**KKAIERMKD	TLRIAYLTEA
W88E	TFKNGAIFQV	EVPGSQHIDS	QKKAIERMKD	TLRIAYLTEA

	81	91	101
wild-type	KVEKLCVWNN	KTPHAIAAIS	MAN
W88K T1A	KVEKLCV**K**NN	KTPHAIAAIS	MAN
W88K	KVEKLCV**K**NN	KTPHAIAAIS	MAN
W88K Q61E	KVEKLCV**K**NN	KTPHAIAAIS	MAN
W88E	KVEKLCV**E**NN	KTPHAIAAIS	MAN

Gel electrophoresis was performed to assess the product of the mutagenesis reactions and showed that the mutant plasmids had been amplified successfully (Fig 4.4). *E. coli XL10* competent cells were transformed with the new plasmids. Plasmid preps were then performed to isolate each plasmid DNA for sequencing and for transformation into *E. coli C41* cells for expression.

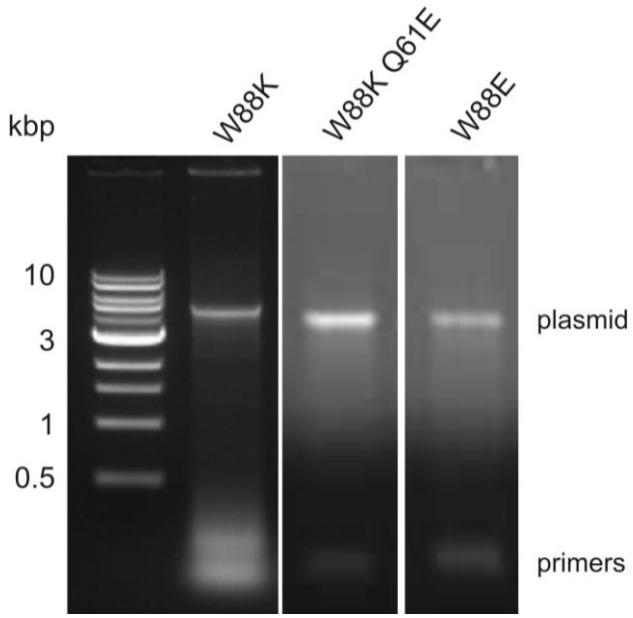


Fig 4.4 Gel electrophoresis analysis of pTRB-W88K, pTRB -W88KQ61E and pTRB-W88E containing the genes for the new mutant proteins.

Each new protein was expressed in *E. coli* by induction with IPTG and growth at 30 °C for 24 h. After purification by ammonium sulfate precipitation, nickel affinity chromatography and then SEC, mutant proteins W88K, W88K-Q61E and W88E were isolated with yields of 2.6 mg L^{-1}, 1.8 mg L^{-1} and 2.0 mg L^{-1} respectively. Mass spectrometry was used to confirm the mutations in the CTB protein. W88K was seen with a mass of 11583.9 Da (calculated 11584.3 Da), W88K-Q61E as 11584.4 Da (calculated 11584.9 Da) and the W88E mutant as 11585.0 Da (calculated 11584.9 Da) (Fig 4.5). A range of proteins were now available and could be tested to see which was most suitable for further reactions.

The mutations were accommodated so that the proteins were all stable pentamers in solution, as evident from the Ni affinity chromatography purification. Purification by this method requires the protein to be in pentameric form to bind to the resin. There are five histidine residues on the bottom face of CTB, one from each monomer, that together provide the binding interaction with the resin.

Alternative purification methods were attempted to improve the yield of the proteins obtained from the *E. coli*. Osmotic shock was used to lyse the cells and liberate protein from the periplasm. Nickel affinity chromatography was used to isolate CTB from the resultant mixture but no protein was found. Cells were also lysed using the cell disruption system and again nickel affinity chromatography was used to purify the protein. This time the mutant protein was obtained but after the affinity column it was still very impure. SEC was employed to purify the sample but it proved too impure to be separated easily in one run. Due to the laborious nature of this purification method and the low yields attainable, the procedure was abandoned and for future preparations only the original ammonium sulfate precipitation method was used.

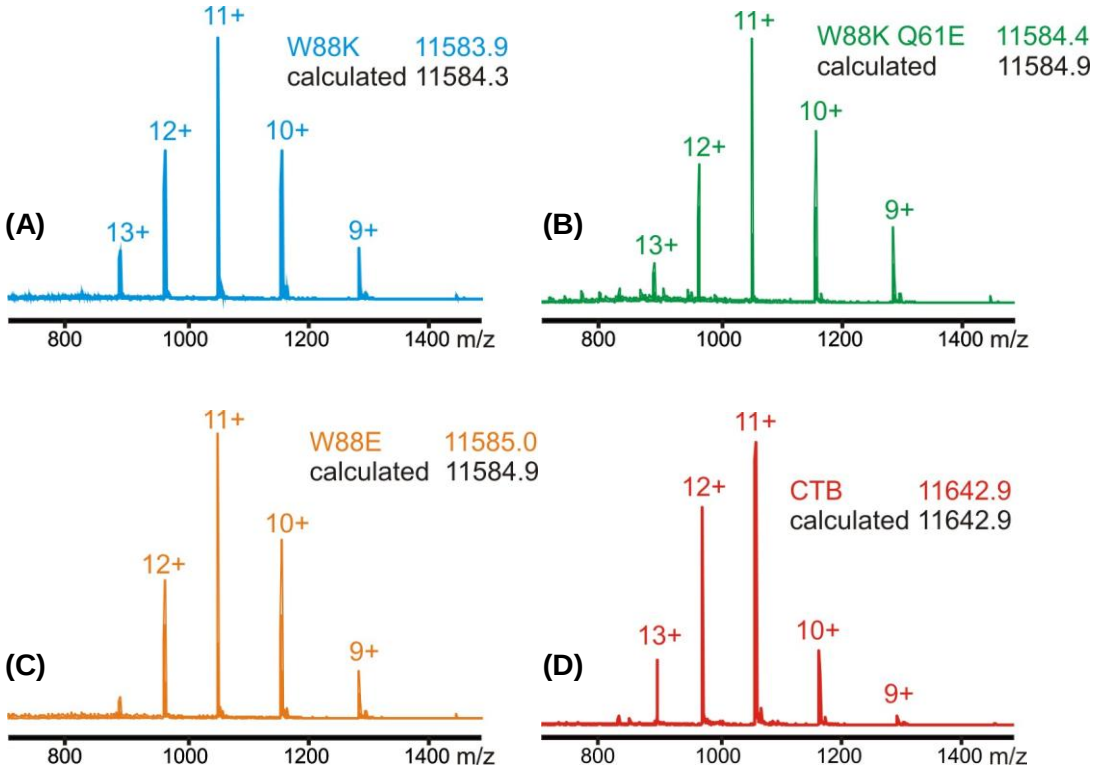


Fig 4.5 Mass spectra of the three different mutant CTB proteins; (A) W88K, (B) W88K-Q61E and (C) W88E. (D) The mass spectrum of wild-type CTB for comparison

4.1.2 Analysing the mutant proteins

The mutant proteins all gave stable pentamers and expressed with similar yields. They were therefore analysed to see which would be most suitable to use for modification. SDS-PAGE was performed to confirm the pentamer stability under the conditions of electrophoresis and showed quite a difference between the mutants

(Fig 4.6). W88K was less stable than wild-type CTB and most of the protein denatured appearing as monomers on the gel. W88K-Q61E runs much higher than the expected pentamer; possibly due to the charge difference created by the mutations or the protein may be forming some dimers of pentamers that are quite stable. W88E was just as stable as the wild-type and appeared at the expected pentamer position on the gel.

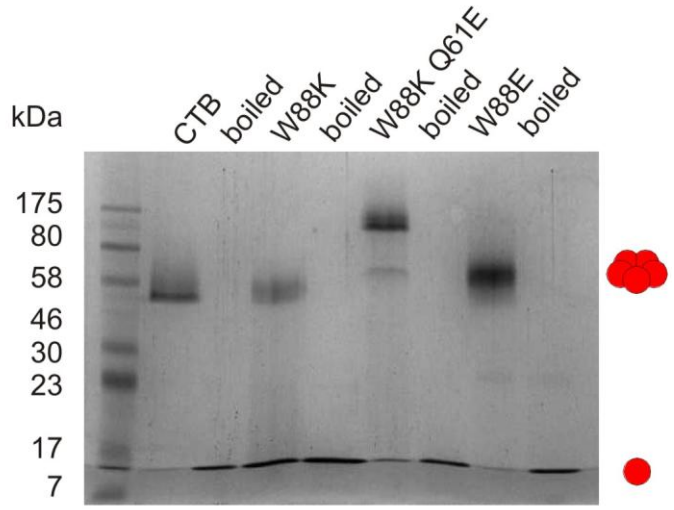


Fig 4.6 SDS-PAGE of all the CTB mutants, showing the variation in stability and position on the gel.

ITC was performed to confirm whether the mutations introduced into CTB did actually prevent the protein from binding carbohydrates. GM1os (125 µM) was titrated into a solution of each mutant CTB (10 µM), as expected no binding interaction with the pentasaccharide was observed (Fig 4.7). Therefore, the mutations that had been introduced were sufficient to prevent GM1os binding. ITC did not, however, prove that there is no binding at all. The pentasaccharide may not be able to bind with the same high affinity as with wild-type CTB but if the interaction had been reduced to a millimolar affinity then this would not have been detected under the conditions used. This low affinity is perhaps unlikely as the mutation should disrupt the terminal galactose binding, which is the most important residue for the pentasaccharide interaction.[89]

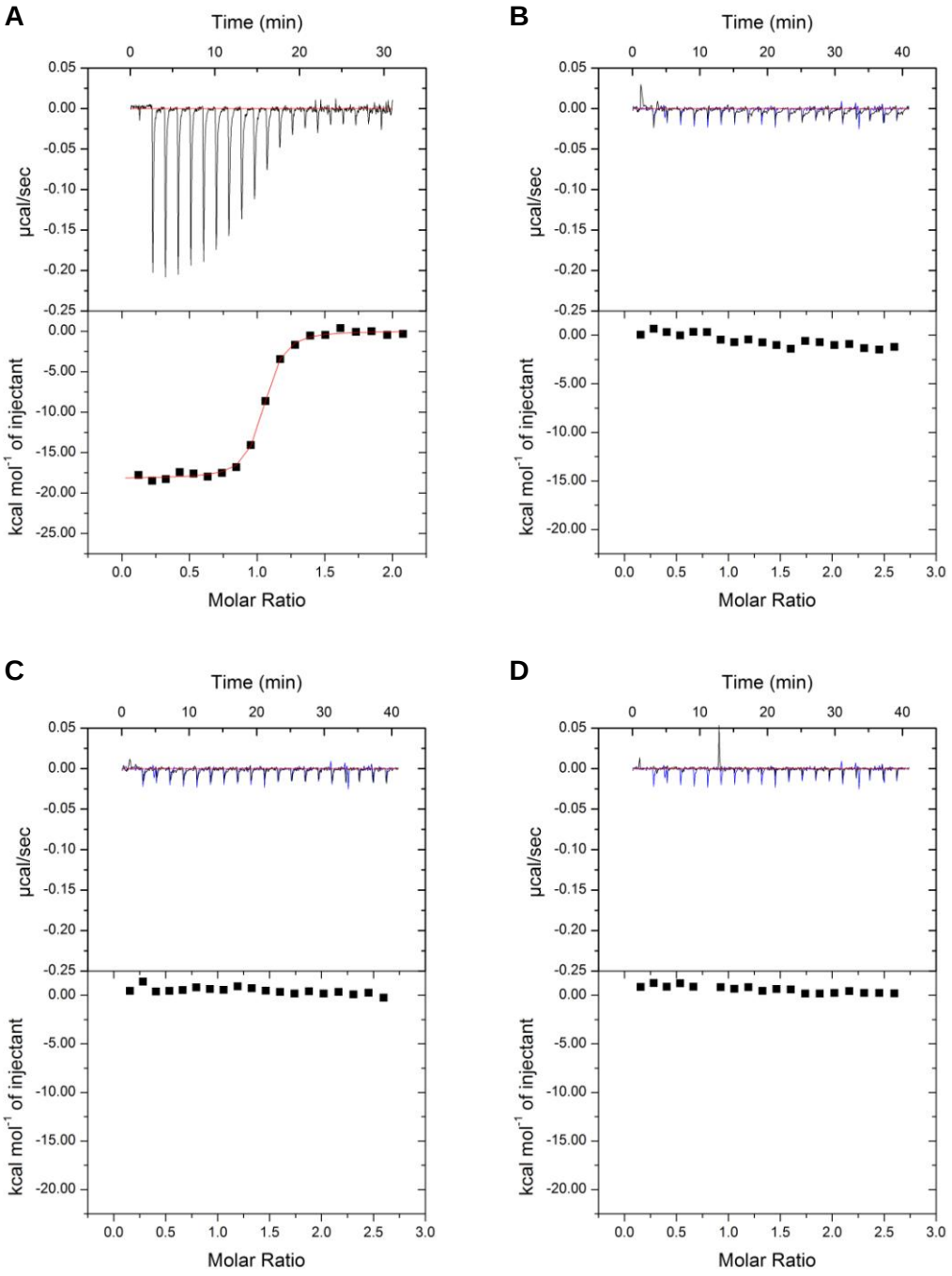


Fig 4.7 ITC plots for GM1os binding to (A) wild-type CTB and showing no interaction between GM1os and the mutant proteins (B) W88K, (C) W88E and (D) W88K-Q61E.

The mutant proteins showed a varying stability as pentamers when seen by SDS-PAGE. But under the conditions of the N-terminal oxime reaction any subtle difference in stability may be important for efficient modification. To compare the thermal stability of the proteins mutants, differential scanning fluorimetry (DSF) was performed (Fig 4.8). DSF can measure the melting temperature of proteins. As the temperature in the experiment is increased, proteins undergo unfolding and expose hydrophobic core regions. Sypro Orange is mixed with the protein samples, and its

fluorescence is quenched in aqueous solution. Upon binding to the exposed hydrophobic regions of proteins the fluorescence is no longer quenched. This fluorescence is measured in real time over the temperature range and so the gradual unfolding of proteins can be observed indirectly.[154] The midpoint of the unfolding transition is defined as the melting temperature for the protein. Sometimes proteins may exhibit multiple local maxima due to various aspects of the tertiary structure unfolding separately. The data is plotted as the change in fluorescence per degree against the temperature, resulting in a curve with a peak that corresponds to the midpoint of the unfolding transition, the melting temperature (T_m).

The T_m for each protein showed quite considerable variation given there are only one or two small differences in the amino acid sequences (Table 4.1). The wild-type CTB had a melting temperature of 81 °C which is slightly higher than the literature value of 74 °C as measured by DSC.[155] W88E had a very similar T_m at 79 °C. The W88K mutant had a lower thermal stability, melting at 72 °C. The second mutation, Q61E, did provide a small stabilising effect to the W88K protein, increasing the melting temperature to 74 °C. However, the W88K-containing mutants were noticeably less stable than wild-type CTB and the W88E mutant.

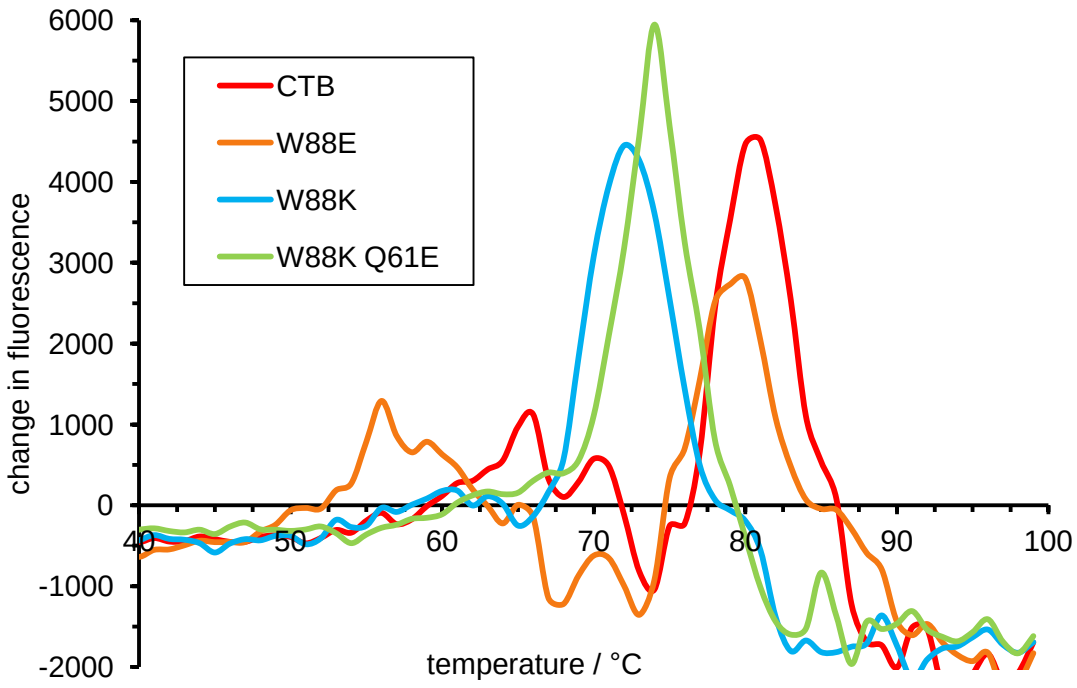


Fig 4.8 DSF showing the change in fluorescence of Sypro Orange with the melting of the CTB mutants. The peaks of the curves indicate the T_m.

Table 4.1 Melting temperatures for CTB and the mutant proteins measured by DSF.

Protein	Melting temperature / °C
CTB	81
W88E	79
W88K	72
W88K Q61E	74

4.1.3 Modification of the mutant proteins

Modification of the N-terminus of all the three mutant proteins was tested with galactose ligand **42** (Fig 4.9). The mutants were first oxidised with $NaIO_4$ and both the W88K and W88E mutants reacted within five minutes, as seen previously for wild-type CTB. The W88K-Q61E mutant, however, failed to oxidise fully even after 1 h. An oxime bond was then formed between galactose ligand **42a** and the oxidised mutant proteins using aniline as a catalyst at pH 7. ESMS confirmed a successful reaction with W88K and with W88E, and SDS-PAGE was used to assess the stability of the modified mutants (Fig 4.10).

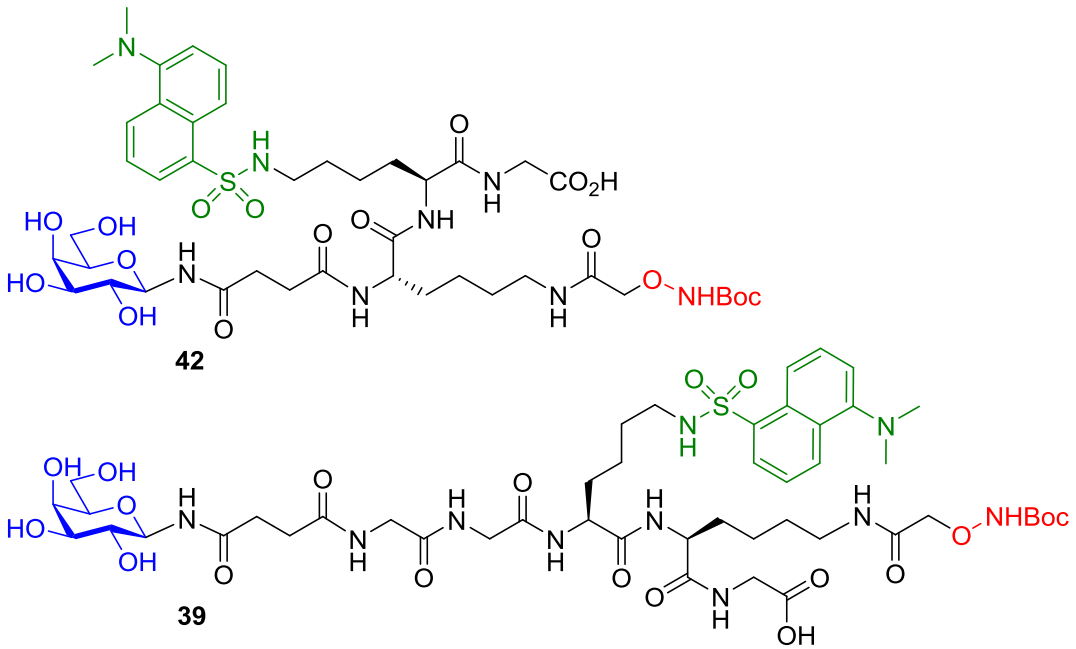


Fig 4.9 The structure of galactoside ligand **42** and the extended galactose ligand **39**.

It was seen that W88E had reacted with the fluorescent ligand and was a stable pentamer on the gel (Fig 4.10). W88K was seen to dissociate into monomers on the gel and so the reaction conditions led to further destabilisation or the modified mutant itself was less stable. The more thermally stable W88K-Q61E mutant also appeared as a stable pentamer on the gel, but had not reacted with the fluorescent oxyamine. W88E had the best stability of the three mutants developed and also behaved as well as CTB in the oximation reaction. It was therefore decided that W88E would be used in all further studies.

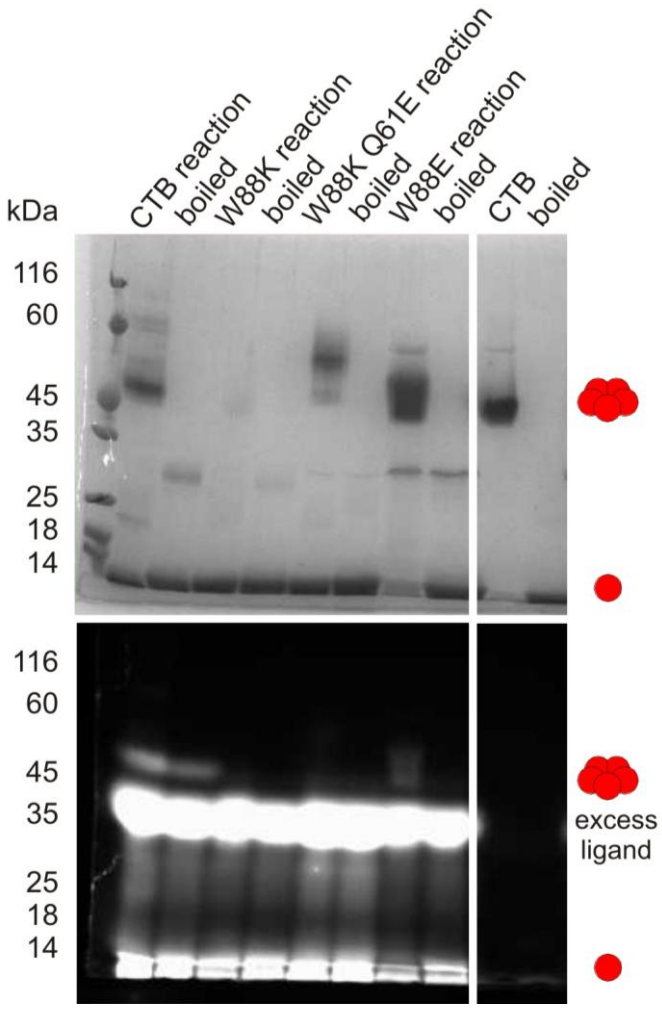


Fig 4.10 SDS-PAGE showing the reaction of the mutants with ligand **42**, stained with Coomassie Blue and under visualised under UV.

4.1.4 Modification of W88E with galactose ligands

W88E was modified at the N-terminus with ligand **42a** and also with the longer ligand **39a**. The longer ligand was used to see if the length of the linker was important for the interactions. Ligand **39** has two additional glycine residues and the

dansyl lysine at a later position in the structure leading to a three amino acid extension compared to ligand **42**. ESMS showed that the products had masses of 12419.7 Da (calculated 12422.0 Da) and 12536.4 Da (calculated 12536.1 Da) for **W88E(42)** and **W88E(39)** respectively (Fig 4.11). The non-binding mutants now presented five galactose moieties and so could be thought of as pentavalent ligands. These ligands, **W88E(42)** and **W88E(39)**, should be able to bind to CTB either with each carbohydrate binding to a different CTB pentamer or with all five forming a multivalent interaction with a single CTB pentamer (Scheme 4.1). These ligands resembled previous star-shaped multivalent ligands, **20** and **21**, produced by Fan and Hol[118], with the essential difference that the central scaffold of the ligands presented here were proteins rather than a synthetic molecule.

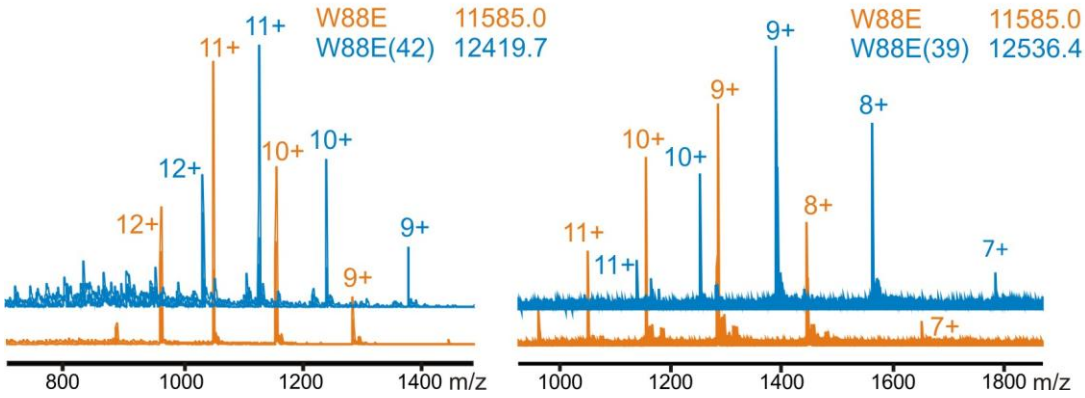


Fig 4.11 Mass spectra of W88E and the modified mutants **W88E(42)** and **W88E(39)**.

4.1.5 Analysis of interactions with wild-type CTB

A pentavalent ligand had now been created bearing five galactose moieties. The individual affinity of each carbohydrate was quite low (12 mM for a simple galactoside), therefore a high concentration of the ligand would be needed to bind to five different CTB pentamers. However, it was thought that the multivalent effect created by having five carbohydrates should give an overall rise in affinity for binding to CTB.

The modified mutant proteins, **W88E(42)** and **W88E(39)**, were mixed with CTB and the protein solution was analysed by DLS (Fig 4.12). Both **W88E(42)** and **W88E(39)** were seen to have a hydrodynamic diameter of 6.5 nm which was slightly larger than CTB with a hydrodynamic diameter of 5.6 nm. The extra size of the attached ligands on the mutant protein may have contributed to the increased size seen by DLS. When **W88E(42)** was mixed with CTB in a 1:1 ratio at a pentamer

concentration of 27 µM, a peak at 5.8 nm was observed indicating no change in the size of the particles. The amount of CTB was increased to give a ratio of 1:5, with **W88E(42)** at 12 µM and CTB at 58 µM pentamer concentrations. A peak at 6.1 nm was observed for this ratio indicating that again there was no significant assembly into larger structures. The difference in size between CTB and **W88E(42)** was too little for the analysis to resolve two separate peaks and so the mixtures gave a single hydrodynamic diameter between those of CTB and **W88E(42)**. Similar results were seen in the DLS experiments with **W88E(39)**. At the 1:1 ratio a peak at 5.6 nm was observed and at with an increase of CTB at a ratio of 1:5 at peak at 6.5 nm was seen. This result showed that a small increase in the ligand length offered no change in the ability of the ligand to bind to CTB and form larger assemblies.

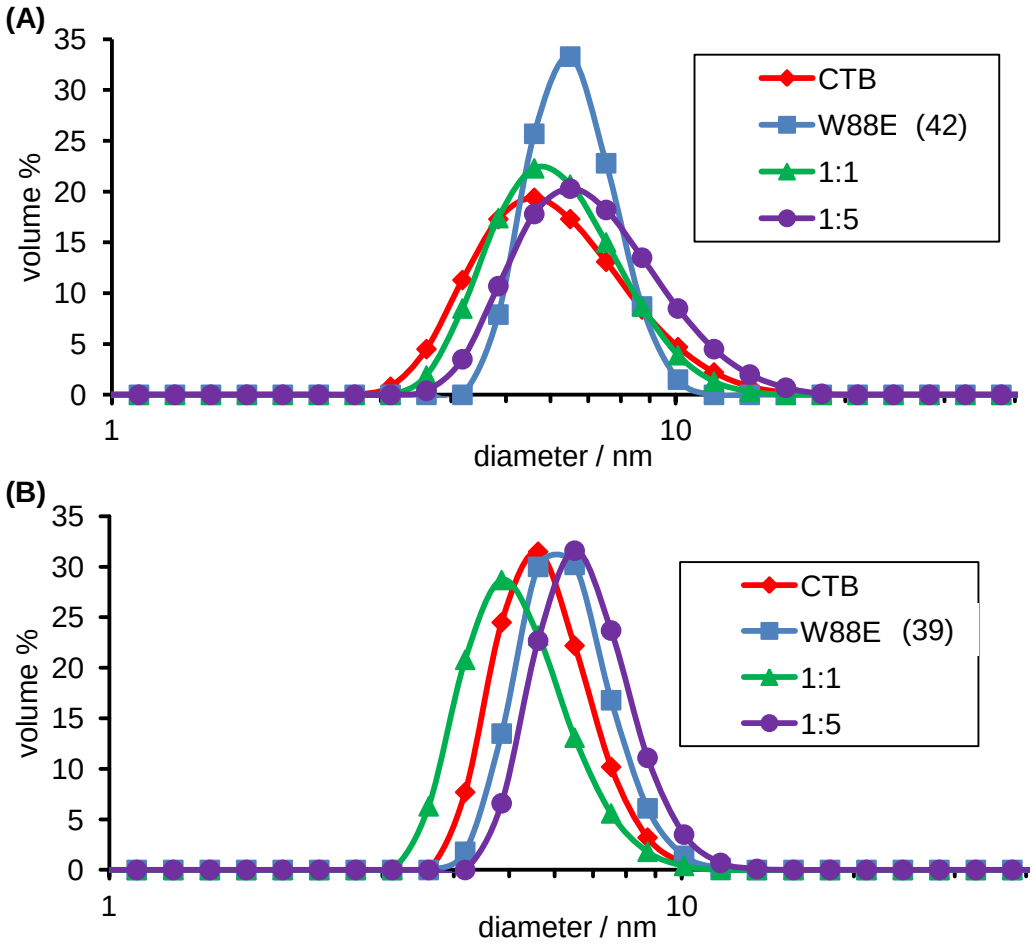


Fig 4.12 DLS plot of (A) **W88E(42)** mixed with CTB and (B) **W88E(39)** mixed with CTB. At a ratio of 1:1 with both components at 27 µM pentamer concentration no increase in size was observed for either modified mutant and at a ratio of 1:5 with the modified mutant at 12 µM and CTB at 58 µM pentamer concentrations no increase in size was also observed.

SEC was also performed on the mixtures of CTB and **W88E(42)** and no change in particle size was observed (Fig 4.13). The protein mixtures were monitored at 330 nm wavelength as this is where the dansyl group on the ligands absorbs. Using this

wavelength meant that only species containing **W88E(42)** would be seen and showed that CTB and the mixtures all eluted around the same retention volume of 17.5 mL.

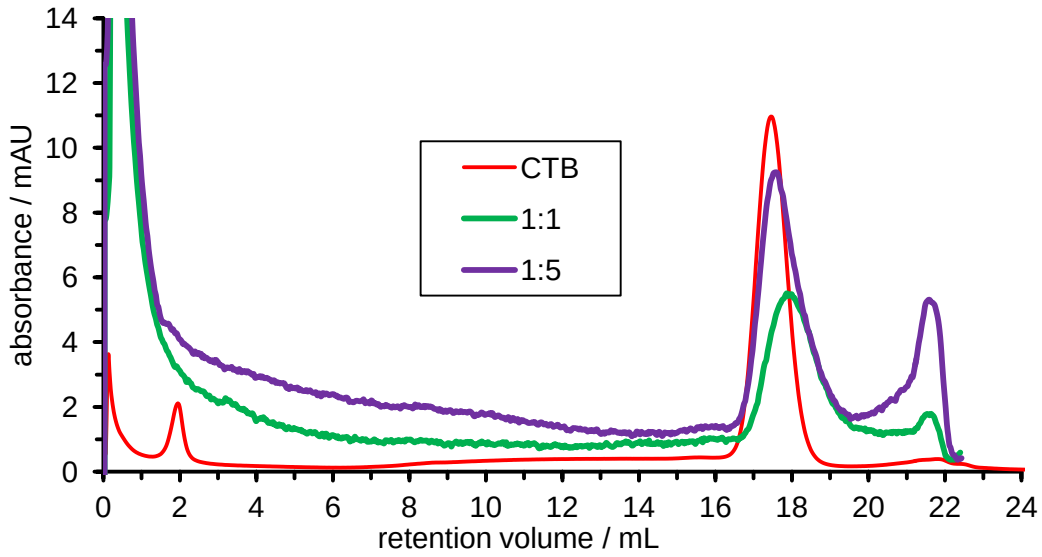


Fig 4.13 Chromatogram for CTB (red) and mixtures of **W88E(42)** and CTB with ratios of 1:1 (green) and 1:5 (purple). CTB was monitored at 280 nm, while the mixtures were monitored at 330 nm to show only the absorbance from the fluorescent ligand.

As the length of the ligand was unlikely to be the source of the problem for these ligands, ITC was performed to assess the binding affinity of the pentavalent galactose ligand. A titration was performed in which CTB (240 µM) was added into **W88E(39)** (5 µM) to reach 10 equivalents. The titration finished with CTB at a concentration of 40 µM and ligand **W88E(39)** at 4 µM, but no interaction was observed (Fig 4.14). It was concluded that the affinity of the galactose moieties on the ligand were too low for binding to be observed at the concentrations accessible with the proteins in ITC. The individual K_d of each galactose ligand on **W88E(39)** was likely to have been in the millimolar range, similar to the monovalent galactose affinity. No great increase in affinity was observed due to a multivalent effect and no binding of this ligand was observed with CTB.

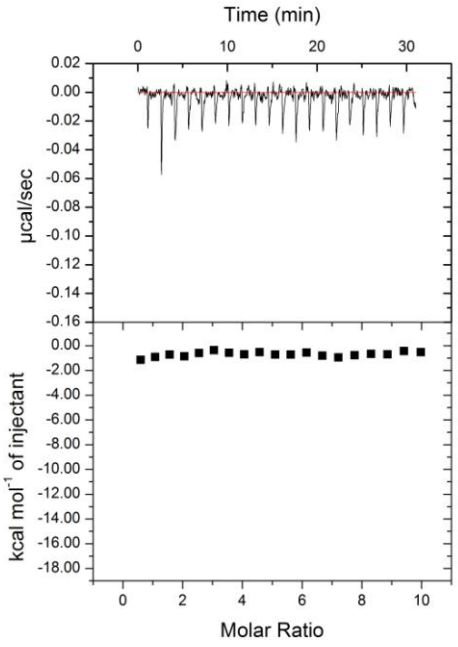


Fig 4.14 ITC trace showing no interaction as CTB (240 µM) was injected into a solution of **W88E(39)** (5 µM).

4.2 Increasing the ligand affinity

There was no sign of the proteins assembling when millimolar affinity carbohydrate ligands were used. The binding enhancement expected from a multivalent system was not observed. Pentavalent galactose ligand **20**, produced by Fan *et al.,* did show binding interactions with CTB[118]. The divalent, tetravalent and octavalent galactose dendrimers, **18a-c**, from the Pieters lab also showed low micromolar IC_{50} values.[114] But a recent study from the Zuilhof lab gave contrasting results.[121] Using a pentavalent galactose compound on a calixarene core, the Zuilhof lab were not able to detect an IC_{50} value as it was above the limit of the assay and so was only calculated as >1 mM. The arms of these compounds were of a comparable length to the Fan ligands, which makes it surprising that such different results were seen. The differences seen with multivalent galactose ligands could be due to the length and flexibility of the ligands. The protein based system presented here may suffer from a reduced flexibility and accessibility of the ligand groups as they are constrained around a large rigid scaffold.

Fan *et al.* did not report whether their compounds were forming 1:1 assemblies or binding to multiple proteins but investigations with the higher affinity pentavalent MNPG **21** ligand did yield 1:1 complexes as shown by DLS and crystallisation.[156]

Similarly the Zuilhof lab reported strong binding when pentavalent GM1os ligands were used.[121] These studies suggest that whilst low affinity ligands are not always ideal for building complexes, even with multivalent systems, increasing the size and flexibility of the ligands can aid assembly and moving to a higher affinity ligand will certainly aid the binding.

4.2.1 Click chemistry on the protein

The GM1 pentasaccharide, which has a nanomolar K_d for CTB, would give a significant increase in affinity over the simple galactoside. Therefore a pentavalent scaffold with the GM1os ligand would be expected to be able to bind to one or multiple CTB proteins.

The Zuilhof lab in Wageningen provided a GM1os molecule with an 11 carbon chain and azide functional group **49**. Taking advantage of the azide group, an alternative way of attaching ligands to CTB was envisaged using a copper-catalysed azide-alkyne cycloaddition (CuAAC), known as “click chemistry”.[157, 158] An alkyne would need to be introduced on to the N-terminus of CTB, which could then be “clicked” with an azide on the ligands. Attaching an alkyne on to CTB via oxime chemistry could be done in bulk and this modified protein could be stored and used when necessary with a variety of ligands containing azides, including the GM1os **49**. Another azide with a lactose group **51** was also provided by the Zuilhof lab to provide a comparison with a low affinity carbohydrate (K_d ca. 15 mM, similar to galactose) (Fig 4.15). Click chemistry is well studied and has become a well used, flexible tool with many conditions known for different reagents.

In order to introduce an alkyne function onto CTB, alkyne **52** (synthesised by Dr. Tom McAllister) was used containing an aminooxy function that could be attached to the CTB N-terminus using oxidation and oxime chemistry (Fig 4.15).

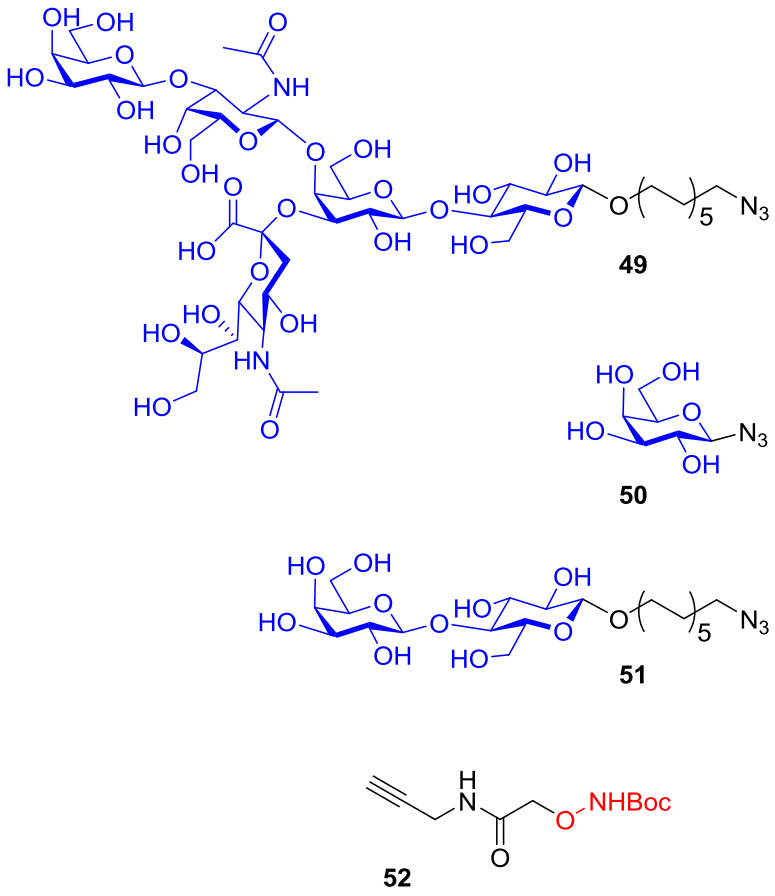
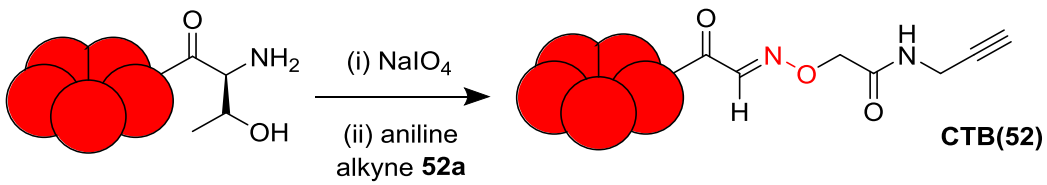


Fig 4.15 Carbohydrate azide derivatives and alkyne counterpart for attachment to the CTB N-terminus.

Ultimately the alkyne group would be needed on the W88E mutant but for ease of handling and issues with quantity of protein, trial reactions were carried out on wild-type CTB. The N-terminus of CTB was oxidised before successful modification with alkyne **52a** via oxime ligation (Scheme 4.2) and then the modified protein was purified on a PD G-25 minitrap column (GE Healthcare). ESMS was used to confirm that the modified protein **CTB(52)** had a mass of 11709.8 Da (calculated 11710.4 Da) (Fig 4.16). SDS-PAGE was performed to check the stability of the protein and showed the alkyne-labelled protein to be a stable pentamer (Fig 4.17).



Scheme 4.2 Modification of CTB with alkyne **52a**.

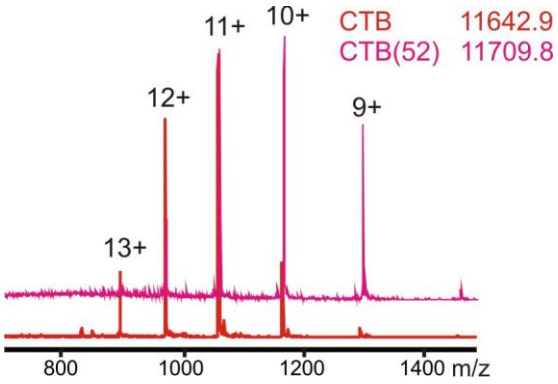


Fig 4.16 Mass spectrum of the alkyne modified CTB.

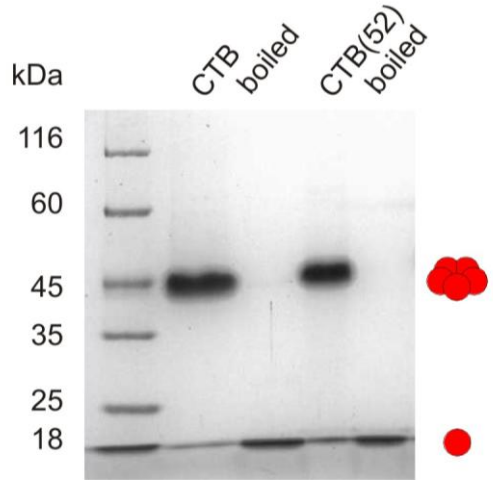


Fig 4.17 SDS-PAGE of CTB plus alkyne ligand **52a.** The gel shows that CTB modified with a small molecule containing a terminal alkyne is still stable as a pentamer.

The CuAAC was tested using alkyne-modified **CTB(52)** in a reaction with the lactose azide **51**. Protein **CTB(52)** was mixed with lactose azide **51**, the catalyst $CuSO_4$ and reducing agent tris(2-carboxyethyl)phosphine (TCEP) in *phosphate buffer*. Copper ligand tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) in DMSO was added to stabilise the copper(I) oxidation state and the mixture was left to react at room temperature. The reaction conditions used were a modification of the procedure reported by Wang *et al.*[159] DMSO was used to initially solubilise the TBTA rather than t-butanol, as reported by Wang *et al.*, as it was found to be completely insoluble in both t-butanol and H_2O. However after addition of the TBTA, the mixture still turned a cloudy white. After 48 h the reaction was centrifuged at 13 krpm for 1 min to remove the white precipitate, before purification. The reaction was not entirely successful as the modification only proceeded to around 60% conversion as estimated from the mass spectrum (Fig 8.2 see experimental). The product had a mass of 12247.5 Da (calculated 12247.7 Da). The incomplete reaction was thought to be due to the insolubility of the TBTA in water, which may

also have also caused some of the protein to precipitate. It was important for this reaction to reach completion as there is no way to purify the product from unreacted protein and so only a semi-modified protein could be purified which is neither reproducible nor useful for this project.

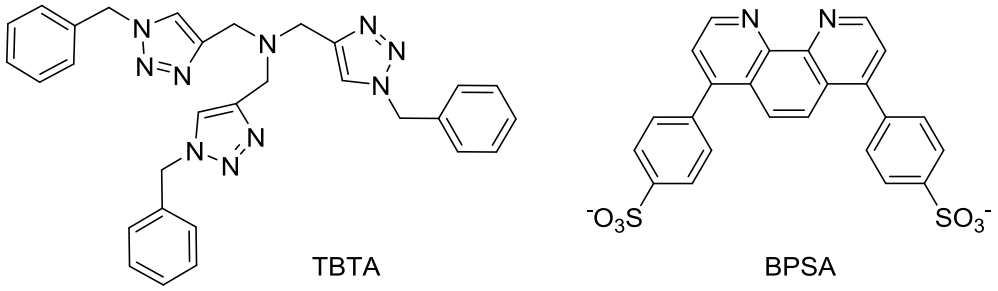


Fig 4.18 The structures of the copper ligands for CuAAC reactions.

The number of equivalents of azide was also found to be important. In the above reaction a conversion of ~60% was seen when there were 50 equivalents of azide present but when this number was reduced to 5 equivalents the conversion was reduced to ~10%. Other attempts were made to increase the conversion by keeping the TBTA in solution. TBTA was found to be more soluble at lower concentrations and so the CuAAC reaction was investigated following conditions closer to those reported by Heal *et al.* with all reagents at a lower concentration (TBTA now 20 × more dilute at 100 μM).[160] However, this proved to be unsuccessful and no product was obtained after 48 h with all the protein remaining as the alkyne. The low concentrations, whilst aiding the TBTA solubility, probably contributed to the failed reaction.

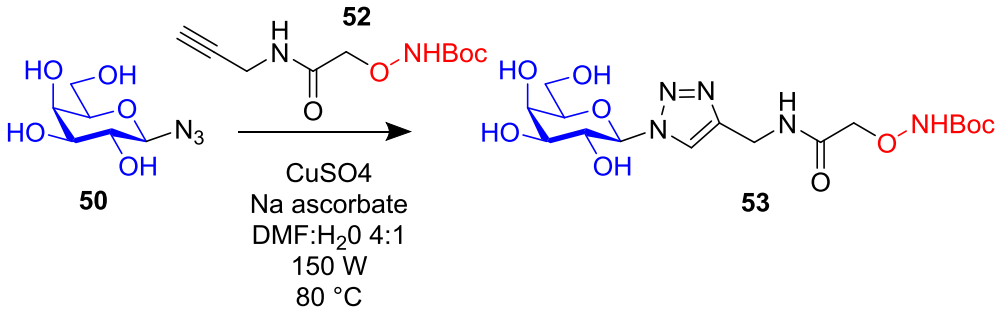
With the insolubility of TBTA proving to be a problem, a new copper ligand, bathophenanthrolinedisulfonic acid (BPSA), was used (Fig 4.18). This ligand is water soluble and had been reported as better than the TBTA for protein CuAAC reactions.[161] The CuAAC reaction was again performed on the alkyne modified protein under similar conditions to those that had previously led to a 60% conversion, but now with everything in aqueous solution.[162] However, reactions employing this new ligand proved unsuccessful and no product was seen.

The same reaction was performed on a simpler azide in order to assess whether the reaction was failing due to the nature of lactose azide **51** or the reaction conditions. It was thought that the carbon chain of lactose azide **51** may have formed micelles and so a simple galactosyl azide **50** (synthesised by Mr Daniel

Williamson) was used in parallel reactions with TBTA or BPSA and alkyne **CTB(52)**. The reactions with TBTA progressed only as well as before, again not reaching completion after 48 h. Similarly, no product was seen when BPSA was used and so this copper-chelating ligand was not investigated further.

4.2.2 Microwave-assisted click chemistry

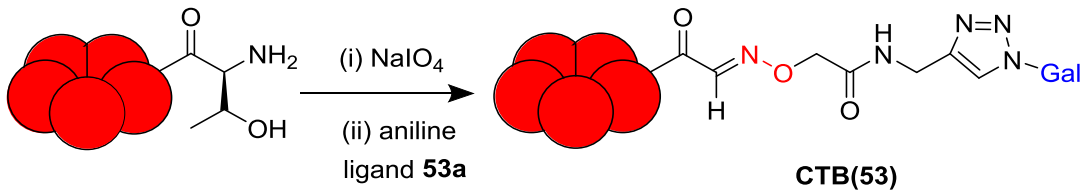
After CuAAC reactions directly on the protein proved difficult, it was thought that the reaction could be done first in organic solution and then the triazole-containing product could be attached to the protein by oximation. Conditions used were from a study by the Pieters lab for the microwave-assisted reaction of carbohydrates.[163] Alkyne **52** plus 1.5 equivalents of galactosyl azide **50**, 30 mol% $CuSO_4$ and 60 mol% sodium ascorbate were mixed together in DMF:H_2O 4:1 and reacted in a microwave for 20 min at 80 °C, 150 W power (Scheme 4.3). After the reaction had cooled, TLC showed no alkyne remaining and mass spectrometry confirmed the triazole containing product **53** with a mass of 433.2 Da (calculated 433.2 Da). The mixture was aliquoted out and lyophilised. No alkyne remained in the reaction mixture and so the product was not purified further. The only molecule, with an aminooxy group, that could react with the protein in further experiments would be the triazole product.



Scheme 4.3 Microwave-assisted click chemistry.

An aliquot of **53** was deprotected with TFA to give **53a** before reaction with CTBox, at 200 µM to avoid any potential aggregation, in the presence of aniline as a catalyst (Scheme 4.4). The reaction went to completion overnight with ESMS confirming the mass of the product **CTB(53)** to be 11915.4 Da (calc 11915.6 Da) (Fig 4.19). SDS-PAGE was performed and showed that the modified protein was stable as a pentamer during the reaction and after purification (Fig 4.20). A sample was taken before purification and centrifuged to ascertain if any protein had precipitated. This analysis is shown on the gel as the “pellet” fraction and it can be seen that the majority of the protein was stable in the “supernatant”. This method of

first forming the triazole before the oxime formation with the protein proved successful and a much better alternative to performing the CuAAC directly on the protein.



Scheme 4.4 Protein modification with ligand **53a**.

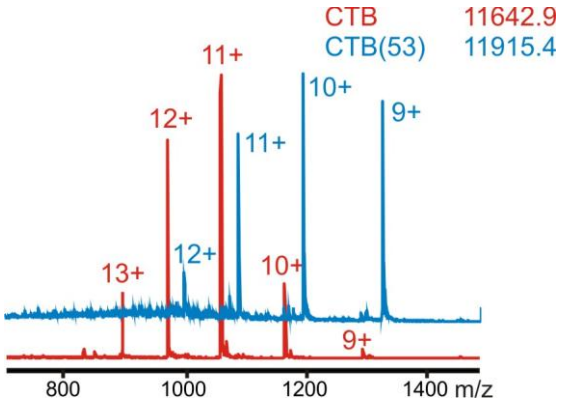


Fig 4.19 Mass spectrum of the modification of CTB with ligand **53a**.

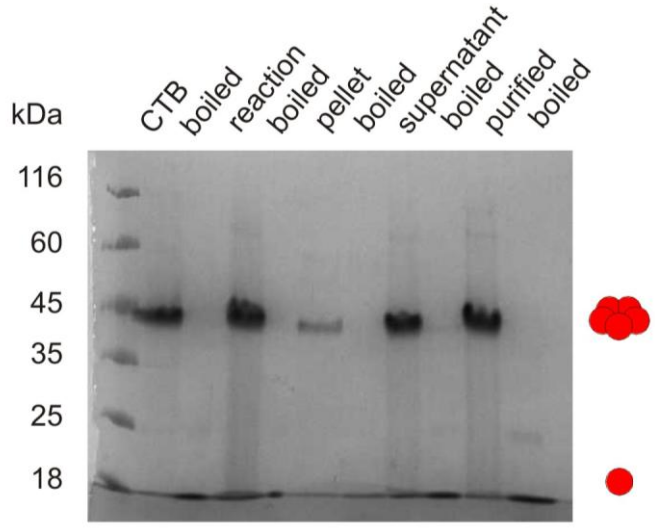
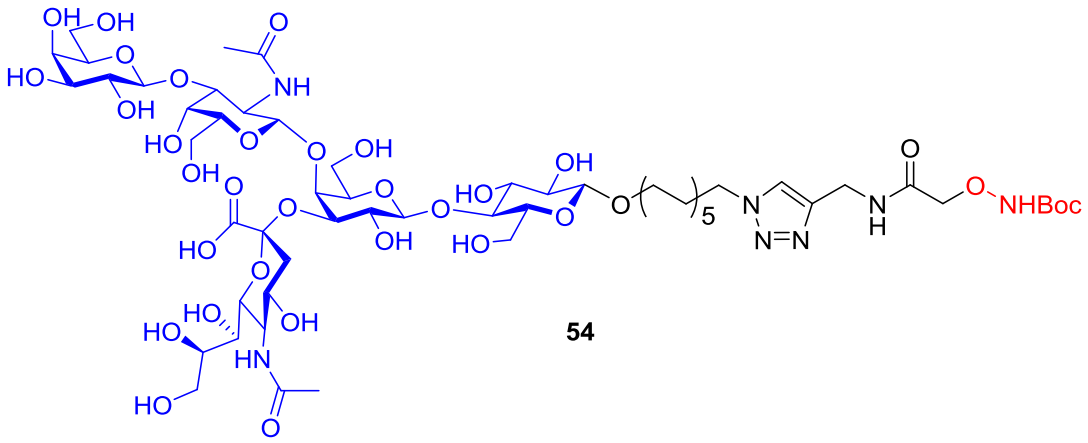


Fig 4.20 SDS-PAGE of CTB plus galactose ligand **53a**. The gel shows the modified protein stable as a pentamer after purification. During the reaction the pentamer was stable with very little precipitating (pellet) and most of the protein still present in the supernatant. The protein was isolated as a stable pentamer with no evidence for larger aggregates forming.

After a successful reaction with the simple galactosyl azide **50**, the microwave reaction was attempted on the larger lactose compound **51**. Similar conditions were

used as before but with all concentrations being reduced due to limitations with the minimum volumes possible in the microwave reaction vessel. Reactions proved unsuccessful probably due to the limitations on the concentrations of the reaction mixture.

In a parallel study, Dr. Tom McAllister performed the CuAAC reaction using a different procedure to connect GM1 azide **49** and alkyne **52**. This CUAAC reaction was successful at room temperature with stirring for two days (see appendix). After purification on a reverse phase column, the correct product GM1 **54** was isolated and could now be used in oxime reactions with the protein (Fig 4.21).



54

Fig 4.21 Structure of GM1 ligand **54** for protein modification.

This ligand was now used with the mutant protein in order to create the pentavalent ligand. The mutant protein was first oxidised and then W88Eox was purified using a PD G-25 minitrap column to remove the $NaIO_4$ that might oxidise the diol on the sialic acid residue of the GM1os in the subsequent reaction. GM1 **54a** was successfully attached to W88Eox to give pentavalent ligand **W88E(54)**, with ESMS confirming the mass for the product of 12844.51 Da (calculated 12844.41 Da) (Fig 4.22). The modified protein was purified using a PD G-25 minitrap column and SDS-PAGE confirmed it to be present as the pentamer with no aggregation taking place (Fig 4.23).

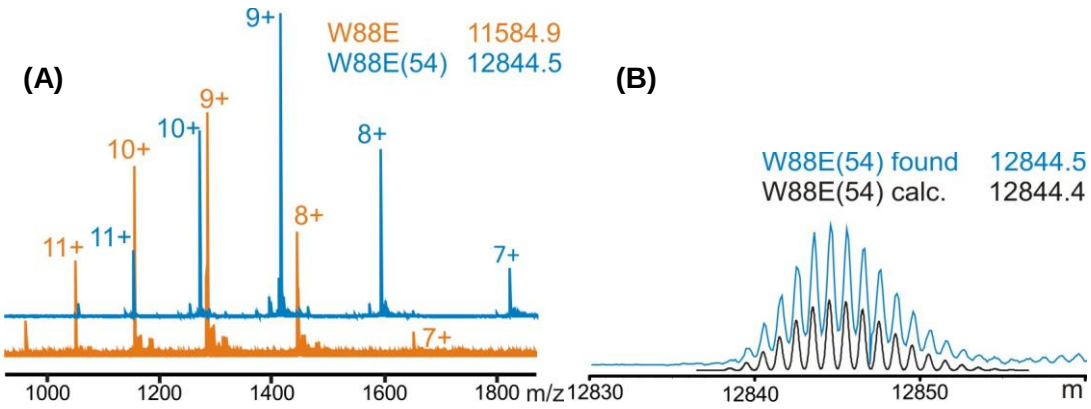


Fig 4.22 (A) Mass spectrum of W88E modified with the GM1 ligand **54a** and (B) the deconvoluted spectrum for **W88E(54)**.

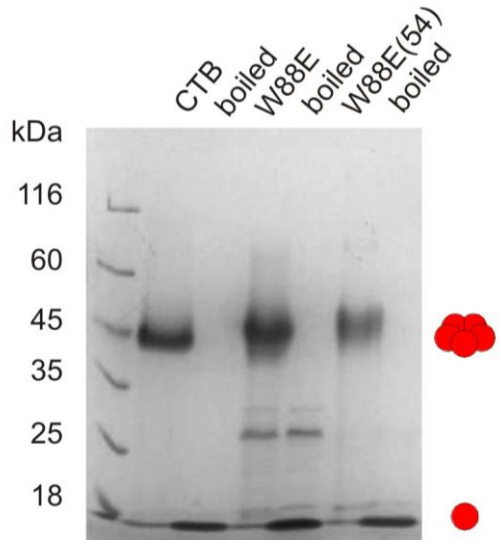


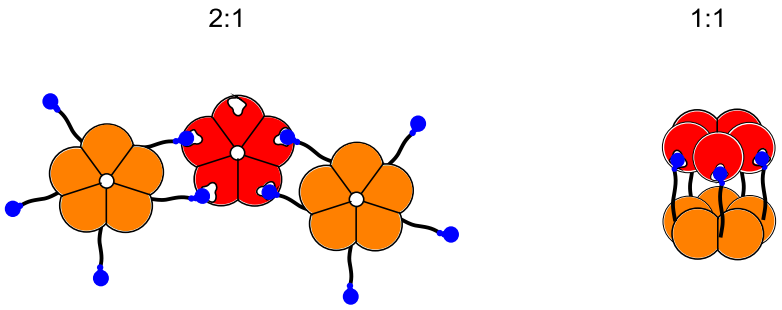
Fig 4.23 SDS-PAGE of the protein modified with GM1 ligand **54a**. The gel shows CTB, W88E and W88E modified with GM1 all as pentamers.

4.2.3 Pentavalent GM1os interactions with wild-type CTB

Now that GM1 was attached to a non-binding CTB protein, a pentavalent ligand had been created with potential for high affinity binding. The affinity of pentavalent ligand **W88E(54)** for wild-type CTB was therefore tested by ITC (Fig 4.24, Table 4.2). With CTB in the cell (4 µM) and ligand **W88E(54)** injected (43 µM) from the syringe a K_d was detected of 30 nM that indicated a slightly stronger interaction than for the monovalent interaction of GM1 azide **49** which had a K_d of 56 nM. This affinity was surprisingly similar to the monovalent interaction as the multivalent effect of binding five ligands on one scaffold was expected to give rise to a significant increase in affinity. The stoichiometry of 1.1 for the titration indicated a binding model in which the protein-based ligand forms a 1:1 complex with CTB. The ΔH° for the binding interaction was a little smaller than that of the monovalent

ligand. The prearrangement for the second interaction after a first GM1 unit has bound may mean that the enthalpy change upon binding is less for subsequent binding interactions.

Performing the ITC experiment but with the components in the cell and syringe reversed, so that CTB (50 µM) was injected into ligand **W88E(54)** (5 µM), gave a different result. The K_d was much the same at 32 nM and again within the same range as the monovalent interaction. The enthalpy value was higher than the reverse titration and similar to the monovalent interaction. However, this time a stoichiometry of 0.54 was observed. This stoichiometry of less than 1 implies that now there is not 1:1 binding but instead suggests that one CTB injected is using up the ligands of two full **W88E(54)** pentamers. The increasing ratio of components may play a role in the stoichiometry as initially there is a large excess of the ligand which may bind to more than one pentamer (Scheme 4.5). The build up of large aggregates may then occur that results in not all the binding sites being accessible to the ligands. For the stoichiometry to be 0.5, this 2:1 arrangement must be stable and prevent the binding of the other ligands to different CTB units. This finding is unexpected as the reverse titration showed 1:1 binding and can simply be explained with the formation of face-to-face dimers of pentamers.



Scheme 4.5 Proposed 2:1 and 1:1 binding models observed by ITC.

Table 4.2 ITC results for ligand **W88E(7)** binding to CTB compared to GM1os.

Cell component	Syringe component	K_d / nM	ΔH° / kcal mol^{-1}	n
CTB	**W88E(54)**	29.6 ± 7.5	-10.48 ± 0.98	1.11
W88E(54)	CTB	32.4 ± 10.7	-15.39 ± 1.71	0.54
CTB	GM1 **49**	55.6 ± 6.4	-14.92 ± 0.15	0.91

The Whitesides group found that trivalent ligands and antibodies initially formed aggregates before rearrangement over time into discrete particles.[69] In the ITC

experiment presented here, the same type of rearrangement may happen to form discrete dimers of pentamers. The timescale of the ITC experiment may be too short to see this rearrangement, whereas AUC experiments that are performed over a larger timescale may show stable structures.

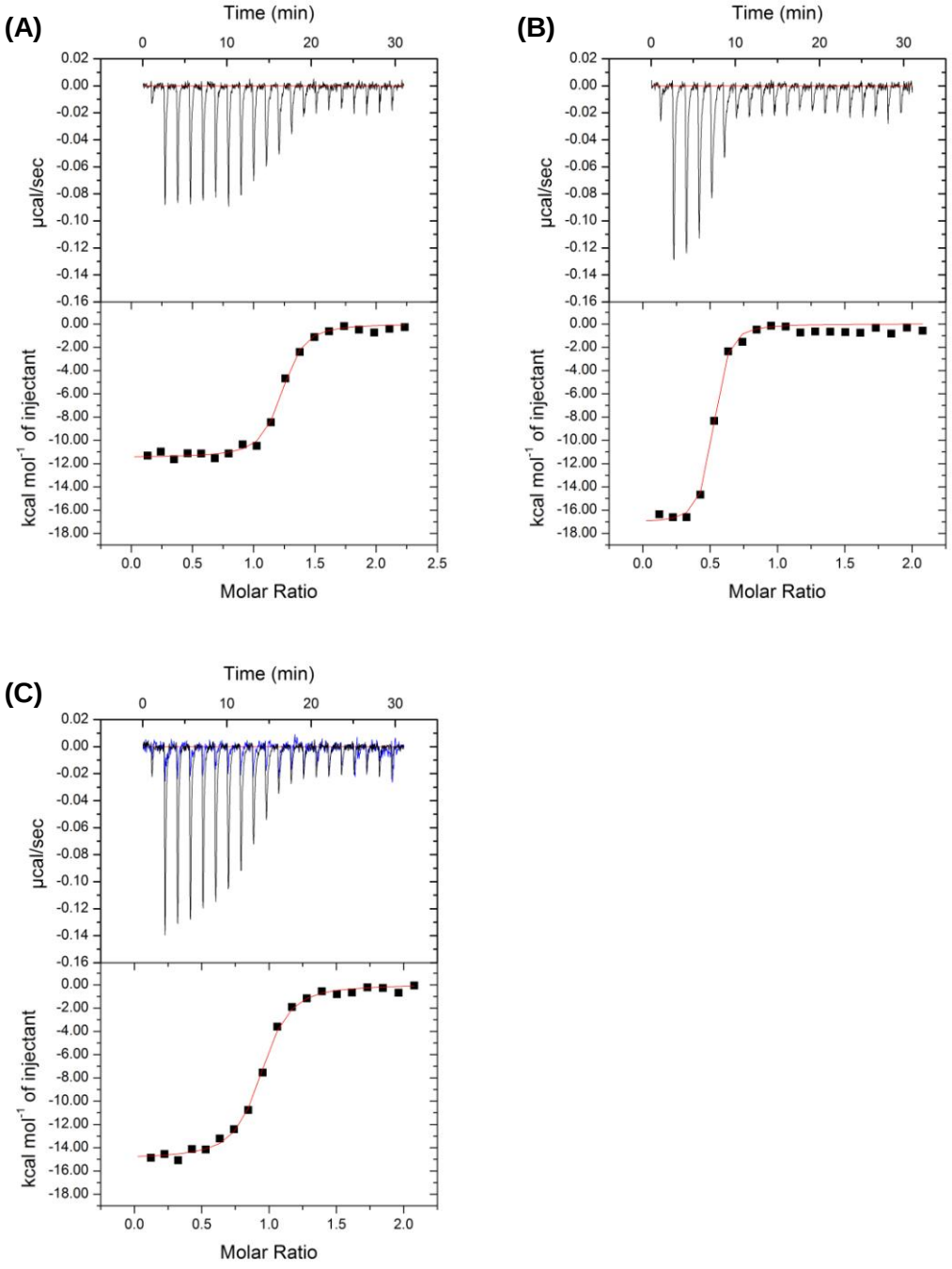


Fig 4.24 ITC traces of the binding between CTB and ligand **W88E(54)**. The ITC results show the strong interaction between CTB and ligand **W88E(54)**, and the differences observed when the components in the syringe and the cell are reversed. A) **W88E(54)** injected into CTB, B) CTB injected into **W88E(54)** and C) GM1 azide **49** injected into CTB.

AUC was performed at varying ratios of CTB and pentavalent ligand **W88E(54)** (Fig 4.25). At a ratio of 1:1 with both components at 20 µM pentamer concentration, the AUC showed a large peak at 6.3 S and a smaller peak at 3.6 S resulting in masses of 113 kDa and 49 kDa respectively. These peaks corresponded to a dimer of pentamers and for the single protein pentamer. The smaller peak was likely to be caused by a slight inaccuracy with the concentrations of the two components not being exactly equal. When the concentration of ligand **W88E(54)** was decreased to 4 µM, so the ratio with CTB was now 1:5, the pentamer peak was greatly increased. This was not surprising as there was now an excess of CTB. The peak for a dimer of pentamers also decreased, which was because now there was less **W88E(54)** present and so less to form dimers. Sisu *et al.* used AUC to analyse the structures formed by GM1os dendrimers and showed similar results for dimers of pentamers. The dendrimers gave a mismatched valency with CTB which caused aggregation of the protein.[116] Crucially in the work presented here, there were no other larger peaks observed. There was also little change to the concentration of proteins in the cells following AUC, indicating that no protein was lost from the formation and pelleting of large aggregates. Matching of the valencies led to the formation of only dimers of pentamers. The formation of a six component complex or any larger structures was not observed.

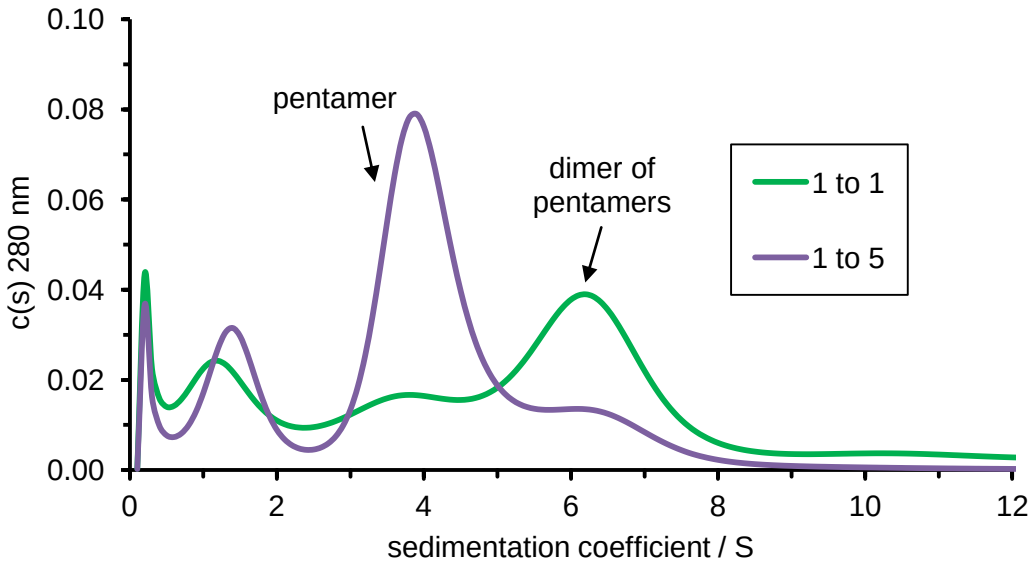


Fig 4.25 c (s) plot of CTB with pentavalent ligand **W88E(54)**. When ligand **W88E(54)** and CTB were mixed in equal quantities, a dimer of proteins was observed in the AUC (green). When the ratio of the components was altered to 1 to 5 so more wild-type CTB was present (purple), some dimer was still observed but more protein pentamer was seen due to the excess of CTB not binding to any ligand.

SDS-PAGE performed on the AUC samples confirmed the AUC results; a band for a dimer was seen in the 1:1 sample whilst a stronger band for pentamer as well as the dimer was seen in the 1:5 sample (Fig 4.26).

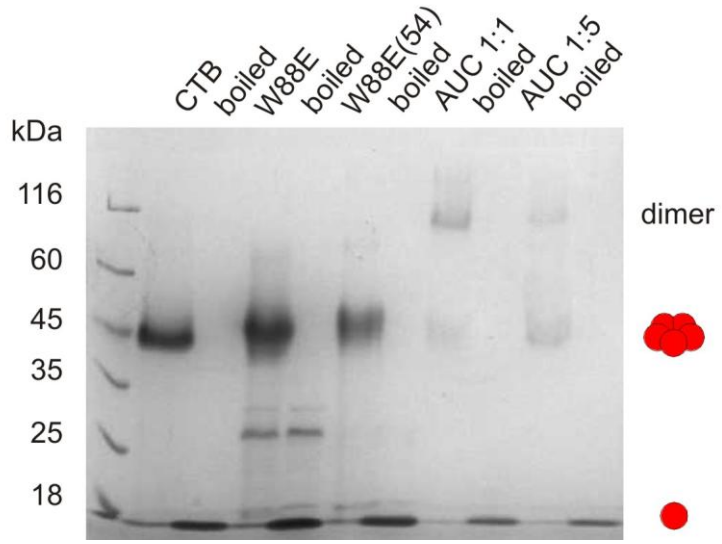


Fig 4.26 SDS-PAGE of AUC samples. The gel shows dimer formation in the 1:1 sample and some dimer in the 1:5 sample although due to the excess CTB there is a stronger pentamer band.

DLS was also used to assess the size of the assemblies (Fig 4.27). The samples from the AUC experiment were filtered before measurement by DLS, which confirmed the formation of species larger than CTB. At a 1:1 ratio, the DLS showed a peak at 8.4 nm, and at 1:5 ratio the peak was slightly shifted down to 7.8 nm indicating the formation of protein dimers in both cases. At the 1:5 ratio the peak was slightly lower, possibly due to the presence of the excess CTB with a size of 5.7 nm causing the peak to shift as the analysis could not differentiate well between the two species of similar size. The crystal structure of CTB indicates a diameter of 6.5 nm, however, the protein is not spherical. It has a depth of 3.5 nm and so it could be envisaged that a face-to-face dimer would have dimensions of 6.5 nm by at least 7 nm. The diameter measured by DLS was therefore in agreement with the expected result for a dimer of pentamers.

The combination of AUC, SDS-PAGE and DLS confirms that protein based pentavalent ligand **W88E(54)** is mostly binding to CTB in a 1:1 ratio causing the formation of protein heterodimers of around 8.4 nm diameter. Little evidence for larger aggregates was seen indicating that this ligand does not bind to multiple CTB proteins. The formation of a six pentamer complex that would be half a dodecahedron was not observed; only face-to-face protein heterodimers (Scheme 4.6). The GM1 ligands are the highest affinity ligands to which we have access and

so there are no other ligands that could be used to inhibit these interactions. Therefore this system with the high affinity ligands will not be useful for building structures larger than dimers of pentamers or more complex protein assemblies.

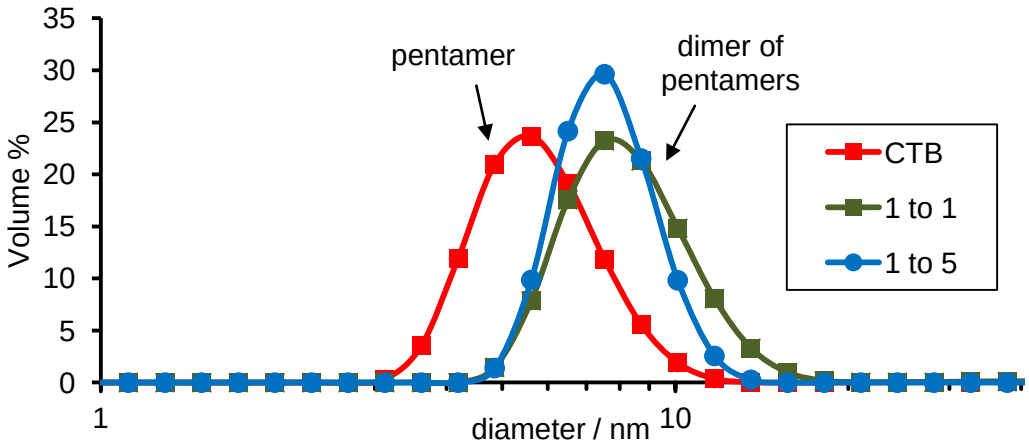
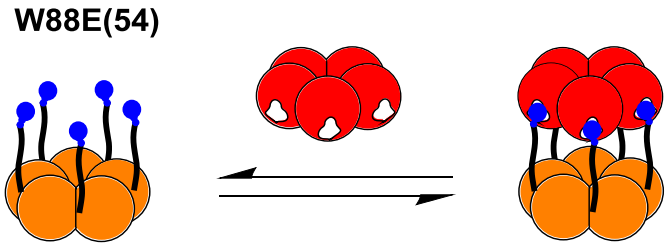


Fig 4.27 DLS results for CTB plus ligand **W88E(54)**. The DLS results show that mixing the ligand with wild-type CTB causes an increase in the size of the particles in solution compared to wild-type CTB (red). At a ratio of 1:1 (green), 8.4 nm particles are observed indicating dimer formation. When the ratio is increased (blue), no larger species are seen and the peak slightly shifts to a lower value possibly due to the presence of more pentamers of a smaller size to the dimers.



Scheme 4.6 Formation of protein heterodimers with ligand **W88E(54)** and CTB.

Whilst this assembly was perhaps a dead end for producing a dodecahedron, it did however present an alternative opportunity. As pentavalent ligand **W88E(54)** binds 1:1 to CTB with high affinity, there was potential for it to be used as a novel protein-based cholera toxin inhibitor.

4.2.4 Inhibition of CTB binding

The modified mutant protein is the perfect size for binding to CTB. The individual GM1 ligand groups are perfectly placed with the correct spacing for binding to CTB, which could make ligand **W88E(54)** a very powerful inhibitor of cholera toxin. Other inhibitors of the AB_5 bacterial toxins have varied in design from polymers and dendrimers to more tailor-made star structures (see chapter 1 part II).[70] The star-shaped structures with 5-fold symmetry represent some of the best inhibitors

due to a precise match of the ligand groups and binding sites. ITC had already showed that ligand **W88E(54)** has a strong binding affinity for CTB, but no increase in affinity from the monovalent interaction was seen. It was seen by Sisu *et al.* that multivalent GM1os dendrimers gave no increase in the K_d for the binding affinity as measured by ITC, but enzyme-linked immunosorbent assays (ELISA) showed a much increased inhibition of CTB.[116] Therefore further experiments were needed to show the inhibitory potential of pentavalent ligand **W88E(54)**.

An Enzyme-linked lectin assay (ELLA) was performed to determine the inhibitory potential of ligand **W88E(54)**. The inhibitor was premixed with horse-radish peroxidase-conjugated CTB (HRP-CTB) and then added to the plate. Unbound HRP-CTB was washed off the plate before a colorimetric assay with ortho-phenylenediamine (OPD) to determine the amount of HRP-CTB bound to the plate. The assay therefore tests the ability of ligands to inhibit the binding of CTB to ganglioside GM1 adsorbed on to the surface of the microtitre plate. Pentavalent ligand **W88E(54)** showed a very low IC_{50} value of 104 pM. The monovalent GM1 ligand **49** was also tested and a much higher IC_{50} value of 1.49 μM was found (Fig 4.28). Comparing the difference in potency from a mono- to pentavalent ligand showed a 14 thousand times increase, which gave an impressive relative inhibitory potency of 3 thousand times per GM1os ligand group (Table 4.3). The unmodified W88E protein was used as a control and as expected showed no inhibition of CTB. This result compared well with other pentavalent ligands such as GM1os-calix[5]arene that was recently shown to have an IC_{50} of 450 pM.[121]

However, the assay used a CTB-HRP concentration of 340 pM and so, with a 1:1 binding model, when the concentration of inhibitor went below the concentration of CTB-HRP then there would be excess CTB-HRP that could bind to the plate and give an absorbance signal. Therefore the limit of detection for this assay was reached with the **W88E(54)** as an inhibitor. If the concentration of CTB-HRP could be lowered then the assay could be performed again to better determine the true inhibitory potential of the pentavalent ligand.

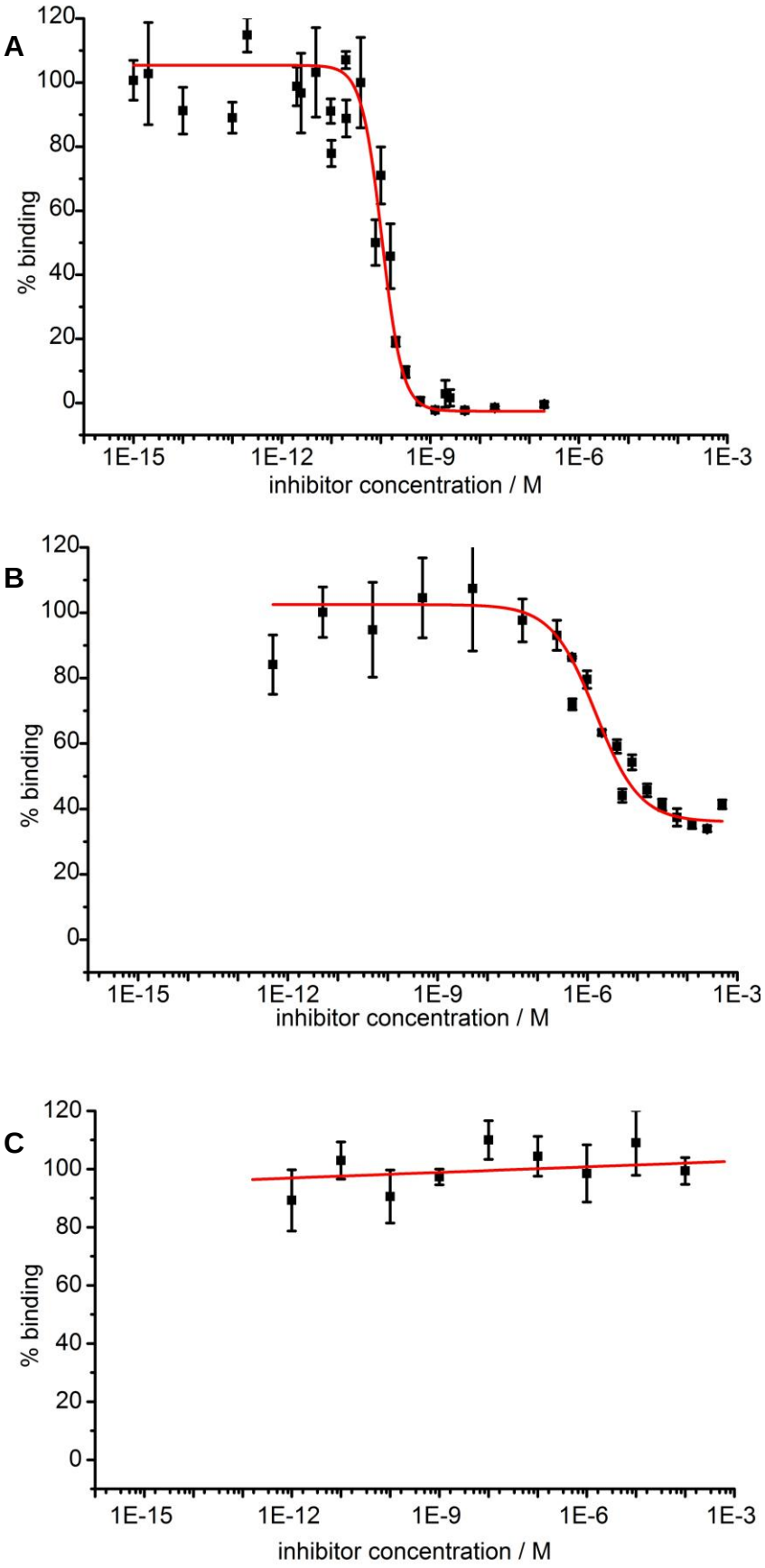


Fig 4.28 Fitted curves of the ELLA inhibition data for (A) pentavalent **W88E(54)**, (B) monovalent GM1 azide **49** and (C) the W88E protein.

The octavalent GM1os ligand **19c** from the Pieters group was tested using a similar assay and was found to have an IC_{50} of 50 pM which is the most successful inhibitor of CTB produced so far.[115] With a more sensitive assay, **W88E(54)** could be seen to rival this octavalent inhibitor. The protein based inhibitor also has the advantage that it is comparably easier to synthesise. The protein can be produced easily in large quantities and the reaction to modify and purify the protein is simple. Whilst the synthesis of the carbohydrate ligand itself is not trivial[164], overall a much more accessible structure can be produced compared to the calixarene or dendrimer inhibitors.

Table 4.3 Potency of the inhibitors measured by an ELLA. [a] No inhibition was detected. [b] Potency was measured relative to the monovalent ligand **49**.

Inhibitor	Valency	log(IC_{50})	IC_{50} / nM	Relative potency (per GM1os)[b]
GM1 azide **49**	1	-5.83 ± 0.11	1490	1 (1)
W88E(54)	5	-9.98 ± 0.08	0.104	14,300 (2,860)
W88E	0	- [a]	- [a]	-

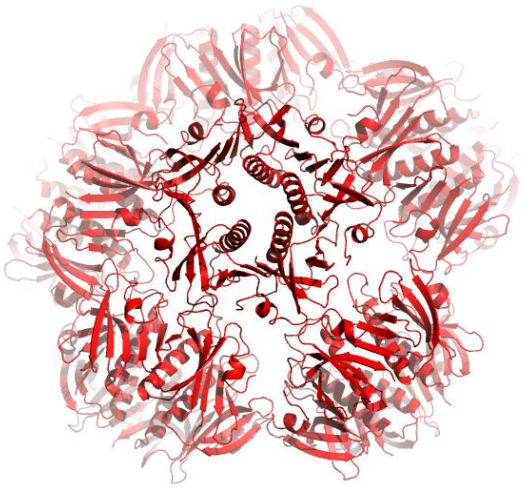
4.3 Conclusions

A mutation was introduced into the GM1os binding site of CTB which prevented any binding with the oligosaccharide. This mutant was modified with carbohydrate ligands and it was found that a simple galactoside ligand did not have a high enough affinity to bind to other proteins. A GM1 ligand was used which did induce interactions and binding between the modified protein and wild-type CTB. Dimers of pentamers were formed and no larger structures were seen. Whilst this observation meant that a dodecahedron or other larger structures were not accessible with these components, it nevertheless provided a potential CTB inhibitor. An IC_{50} value for GM1 modified mutant protein, **W88E(54)**, was found as 104 pM. This value is in the best of any pentavalent CTB inhibitors produced so far and in the same range as the best dendrimer structures.

Due to the strength of the five GM1os interactions, the reversibility of pentavalent ligand **W88E(54)** is limited. To allow the formation of the complex to be more reversible, lower affinity ligands could be used. It was already seen that galactoside ligands were not strong enough but studies have shown MNPG ligands to be suitable for protein binding.[156] MNPG ligands have a micromolar K_d and may be strong enough to bind to CTB but still lower in affinity than GM1, with a nanomolar

K_d. Use of the MNPG ligands could provide some reversibility to the system. The linker length of the GM1 ligands could also be shortened which may affect which interactions can take place. With shorter ligands it may not be possible for all five ligands to reach around and form a dimer of pentamers with CTB. Instead they could bind one CTB pentamer on each GM1 arm of the ligand. These studies could be performed to access a larger range of structures with this two component system.

5 Templating the construction



Investigations into the use of a template to preorganise the proteins

Uncontrolled aggregation occurred when ligands were attached directly to CTB at the N-terminus. The lack of control over the assembly process meant the structures were free to form any arrangement and so the creation of discrete particles was not observed. This represented the major challenge for the project; guiding the self-assembly process towards discrete complexes and away from disordered aggregates.

A new strategy for the formation of a dodecahedron from CTB was to use a scaffold to prearrange the proteins. Viral particles are often seen to be templated by their RNA genome[165] and supramolecular chemistry also often makes use of templates to give rise to complex architectures[73]. To arrange CTB proteins into a dodecahedral configuration, a spherical template was required.

This chapter will discuss the suitability and use of a micelle template for the prearrangement of CTB proteins.

5.1 Micelles

Triton X-100 (TX-100) **55** is a detergent that forms micelles of 7-10 nm in diameter. The size and the aggregation number (N_{agg}) can be varied with temperature.[166] A micelle of this size would fit inside the target protein dodecahedron and therefore could be used to template the formation of these structures. Tyloxapol **56** has a similar structure to TX-100 but is known to form more stable micelles with a lower critical micelle concentration (CMC) (Fig 5.1).[167] TX-100 and tyloxapol were investigated to assess their suitability for providing a template with tuneable size. If the micelles were modified with carbohydrate ligands then they would act as scaffolds for the proteins to assemble around and the degree of modification could be varied to adjust the valency of the micelles. Multivalent ligands could then be introduced to cross-link the pentamers on the micelles (Fig 5.2). Alternatively, the pentamers themselves could be modified directly with ligands at the N-terminus after assembly around the micelles and due to the prearrangement from the template, the pentamers should bind in the desired arrangement.

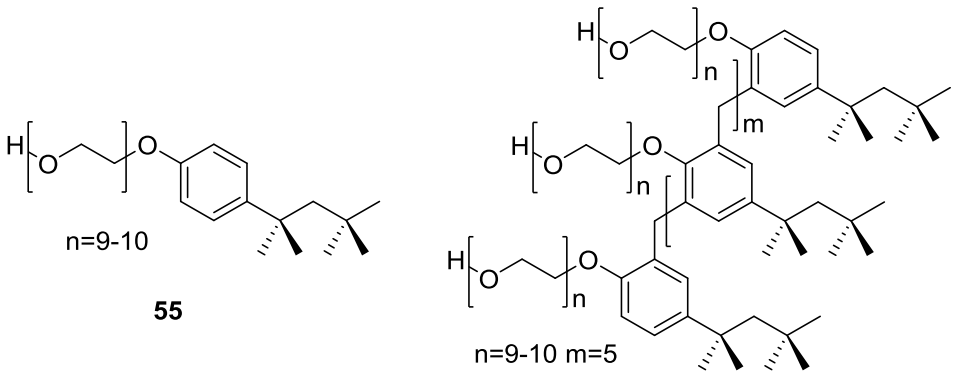


Fig 5.1 Structures of TX100 **55** and tyloxapol **56**.

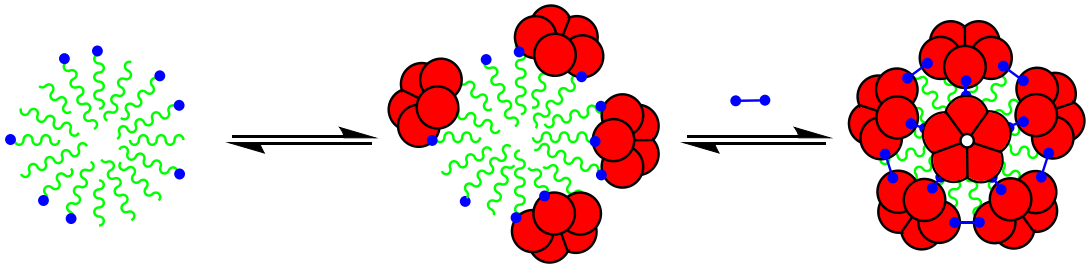


Fig 5.2 Templating strategy with micelles.

5.1.1 Analysis and modification of the micelles

The size of the micelles formed by TX-100 **55** and tyloxapol **56** was measured by DLS at a detergent concentration of 1 mg ml^{-1} (Fig 5.3). The data clearly show a slight difference between the two detergents. The size of the micelles measured for TX-100 **55** were 6.5 nm and for tyloxapol **56** were 7.6 nm at 20 °C which compare well with literature values.[166]

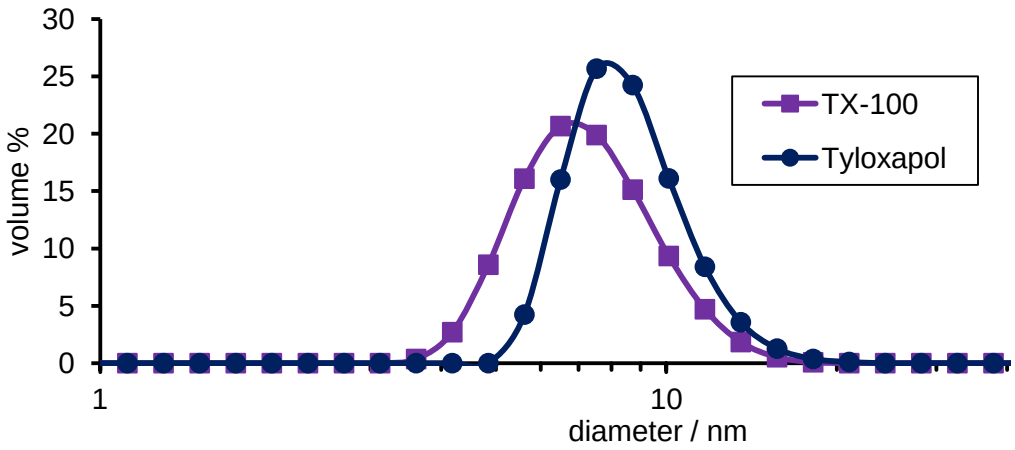


Fig 5.3 DLS showing TX-100 **55** (purple) and tyloxapol **56** (dark blue) with diameters of 6.5 nm and 7.6 nm respectively measured at 20 °C.

Both TX-100 **55** and tyloxapol **56** were modified with carbohydrate ligands. TX-100 galactoside **57**, tyloxapol galactoside **59** and TX-100 lactoside **58** were prepared (synthesised by Dr. Martin Fascione) (Fig 5.4). These modified detergents should present their carbohydrate ligands around the surface and so allow the binding of CTB. The attachment of a carbohydrate group on the detergents was not entirely successful with only 60% of the detergent molecules becoming modified. However, despite the incomplete conversion of the detergents, at 60% there would be approximately 60 carbohydrates per micelle because the N_{agg} of TX-100 is around 100 at 20 °C.[166] A micelle with 60 carbohydrate ligands would provide one ligand for each of the binding sites of the 12 pentamers that would be needed to form a dodecahedron. If it was found that binding to all the potential binding sites was not required or if it proved difficult to displace the template ligands then the percentage of modification could easily be lowered by dilution with unmodified TX-100 **55** or tyloxapol **56**.

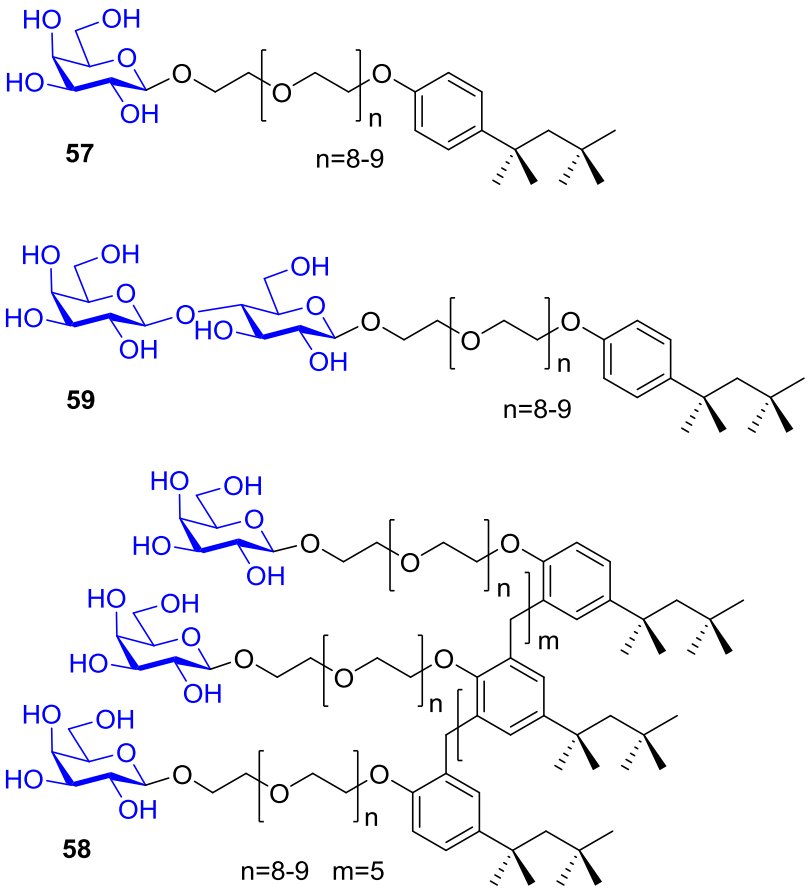


Fig 5.4 Structures of the detergents modified with carbohydrates.

To determine the CMC of the original and modified detergents, ITC titrations were performed. During the experiment the detergent at a concentration above its predicted CMC was titrated into water and heat was released as micelles dissociated.[168] Once the concentration of the detergent in the cell became greater than the CMC, then increasing the concentration resulted in no further heat change and, thus allowing the CMC to be calculated (Fig 5.5). The CMC for TX-100 **55** was found to be 0.27 mM which was in good agreement with a literature value of 0.31 mM found previously by ITC.[168] The ΔH° of -2.26 kcal mol^{-1} was also in good agreement with the literature value of -2.06 kcal mol^{-1}.[168] The CMC values for the TX-100 detergents 50% modified with carbohydrates **57** and **58** were found to be higher than the original detergent (Table 5.1). Galactoside **57** had a CMC of 0.40 mM and lactoside **58** was 0.59 mM which suggested that the modifications provided a slight hindrance to the micelle formation. The ΔH° for the glycosylated detergents was also reduced compared to TX-100. When the micelles formed the carbohydrates may not have packed as well into the structures and so ΔH° for the interactions was reduced. Tyloxapol **56** was known to have a CMC at least 10-fold lower than TX-100 **55**, with a literature value of 8.4 μM.[167] The ITC titration was performed at a lower concentration for this detergent, but there is no obvious heat change upon crossing the CMC value and so an accurate value for the CMC could not be resolved. Tyloxapol galactoside **59** was not analysed as it was thought that ITC would again not be able to give an accurate value.

Table 5.1 CMC values from ITC titrations

Detergent	CMC / mM	ΔH° / kcal mol^{-1}
TX-100 **55**	0.27	-2.26 ± 0.04
TX-100 gal **57**	0.40	-1.64 ± 0.02
TX-100 lac **58**	0.59	-1.00 ± 0.01

The N_{agg} value of 100 at 20 °C for TX-100 **55** meant that at a concentration of 1 mM, crucially above the CMC, the concentration of actual micelles would be 10 μM. To arrange 12 CTB pentamers around each micelle, the CTB (monomer) concentration would need to be 600 μM, which was within an easily accessible range for the protein.

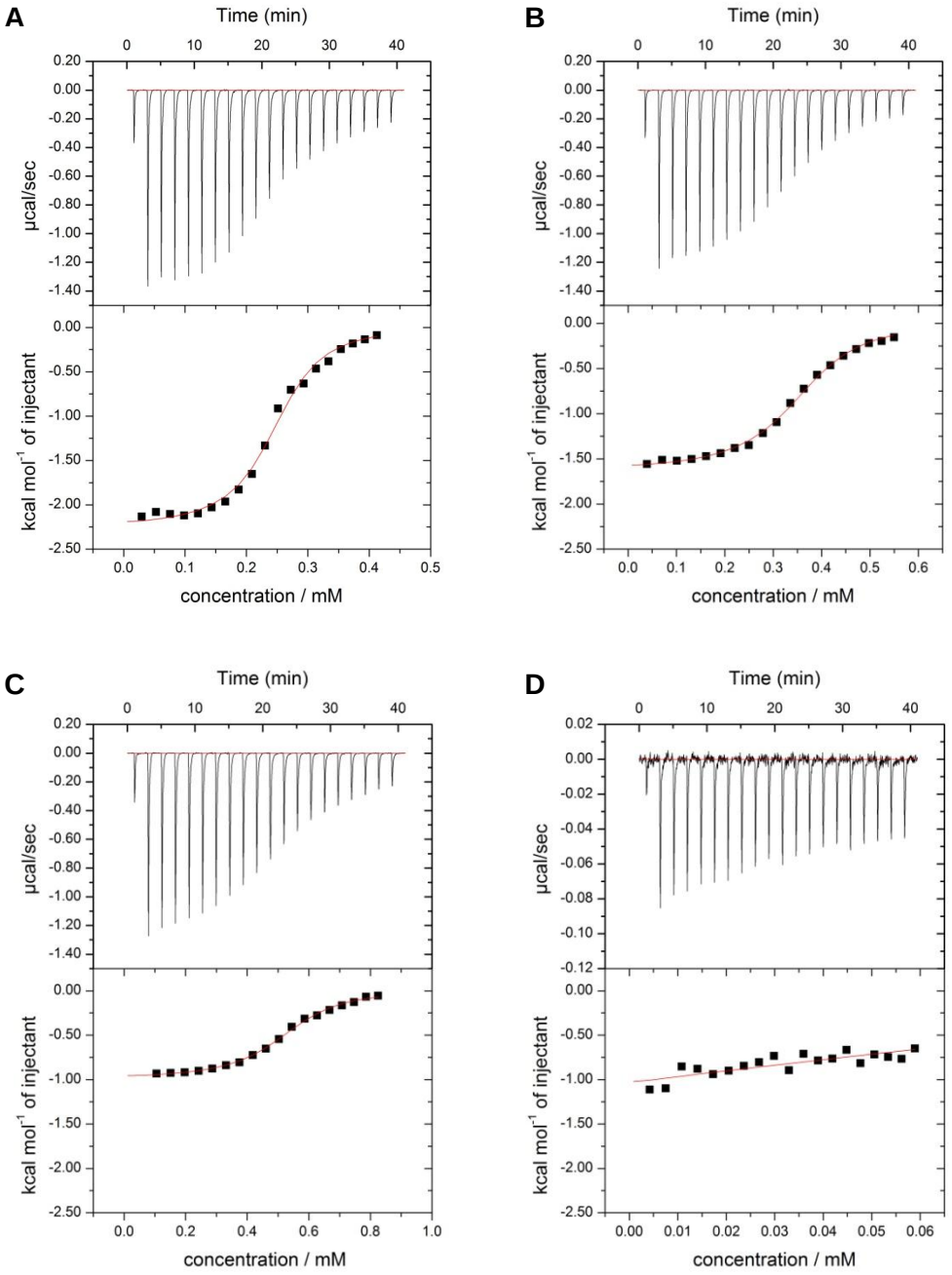


Fig 5.5 ITC traces to find the CMC of the detergents A) TX-100 **55**, B) TX-100 galactoside **57**, C) TX-100 lactoside **58** and D) tyloxapol **56**.

The sizes of the modified detergents were analysed by DLS (Fig 5.6). The samples were analysed at 1 mM concentration so that they were above the CMC for each detergent, and the percentage of carbohydrate components was varied to see if their addition would alter the size of the micelles formed.

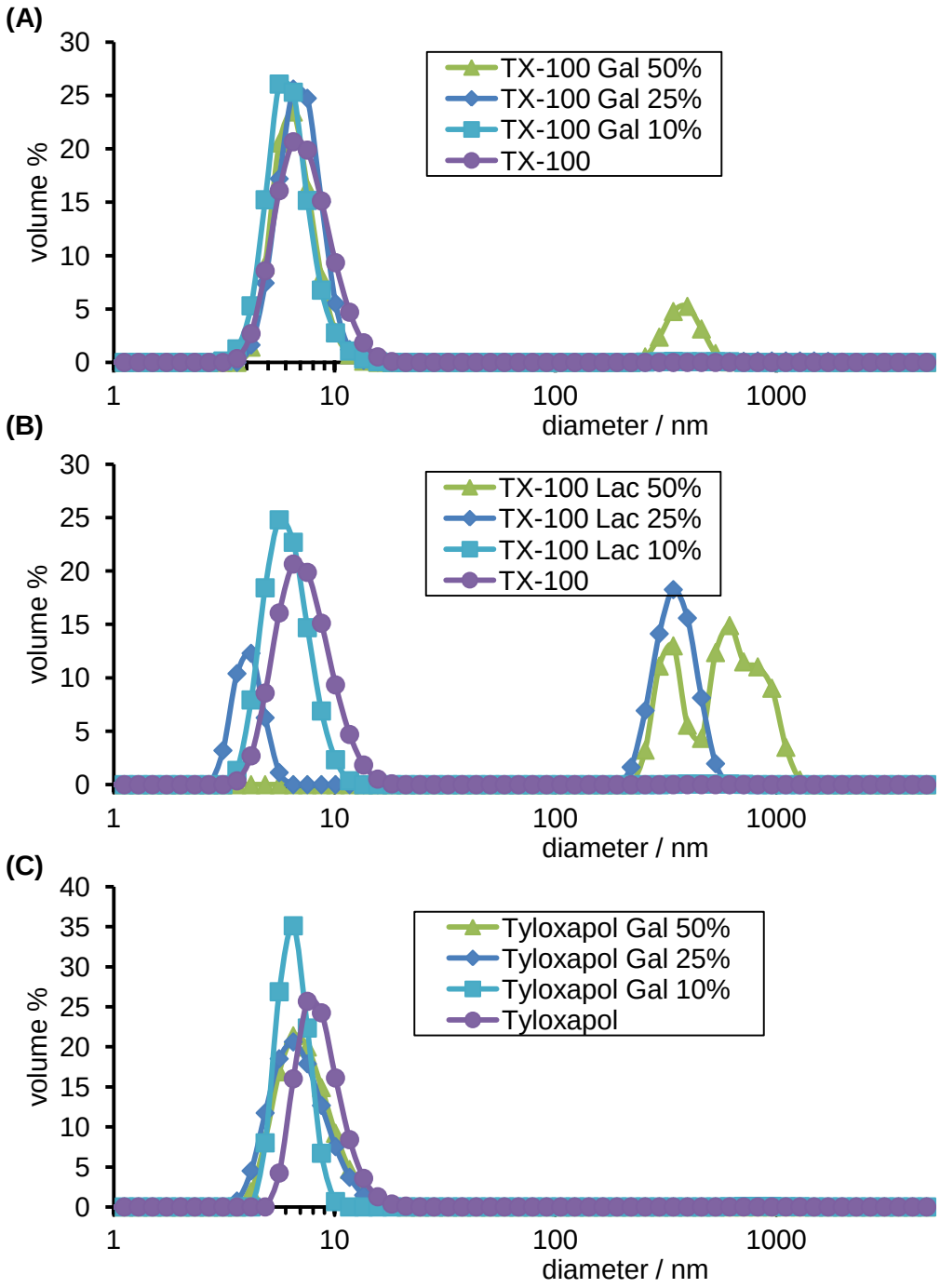


Fig 5.6 DLS of modified detergents at different percentages. (A) TX-100 galactoside **57**, (B) TX-100 lactoside **58** and (C) tyloxapol galactoside **59**.

It was found that for TX-100 galactoside **57** with 50% modified detergent, particles with a diameter of 380 nm were formed as well as the small micelles at 6.6 nm. When the amount of carbohydrates used was reduced to 25% or lower, this higher species disappeared and only the smaller micelles at 6.5 nm were observed. This effect was more pronounced for TX-100 lactoside **58**. A variable amount of large particles ranging 310-850 nm were observed with the higher percentage of modification. When the amount of carbohydrate was reduced to 10%, only 6.8 nm micelles were seen. Conversely, tyloxapol galactoside **59** was much more

representative of its parent detergent as, at 50% carbohydrate, only micelles of 7.0 nm were observed. The stability of tyloxapol **56** is known to be greater than TX-100 **55** and so it was not surprising that this detergent still formed stable micelles once modified. The glycosylated TX-100 detergents, however, were not as stable and led to the formation of large particles of unknown shape. The bulky carbohydrate groups presumably disrupted the hydrophilic interactions at the surface of the micelles or caused steric clashes when too many were present. This interference may have led to a rearrangement of the structures into much larger particles.

Reducing the amount of carbohydrate present to only 10% of the detergent concentration meant that there would be only around 10 of these groups present per micelle. This low amount may not be enough to bind multiple copies of CTB around its surface. The system however may reorganise to bind one CTB pentamer with the carbohydrates forming clusters in the micelles. This arrangement could be enough to start the templating of other CTB pentamers around the micelles with the introduction of additional divalent ligands. Ideally the amount of carbohydrates would need to be at least 12% so that each micelle could still potentially bind 12 pentamers.

5.1.2 CTB binding to the micelles

TX-100 galactoside **57** was mixed with increasing concentrations of CTB and DLS was used to observe any changes in the size of particles. If even a single CTB pentamer were to bind to the surface of a micelle, the particle size observed should be larger than that of the micelle or CTB. With more CTB units binding the change in size should be easily distinguishable. 10% TX-100 galactoside **57** (1 mM) was used which gave a micelle concentration of 10 μM, assuming an N_{agg} of 100. This amount of modified detergent was used to be sure that larger structures did not form and only particles of 7.2 nm were observed. CTB was then titrated into the solution at an initial ratio of 8 pentamers per micelle; this ratio was increased to 12 and then 16 (Fig 5.7). For tyloxapol galactoside **59**, 50% carbohydrate was used as the problems with micelle sizes did not seem to affect the tyloxapol based system. Again at 1 mM detergent, CTB was mixed at ratios of 8, 12 and 16 to the micelles. With the addition of CTB, no larger structures were observed by DLS with either the TX-100 or tyloxapol-based micelles. The analysis showed only one species present at a size equal to the CTB pentamer as there was now excess CTB present the signal shifted to show only this size particle. With both tyloxapol and TX-100 mixed with CTB, the two components did not appear to interact to form larger structures.

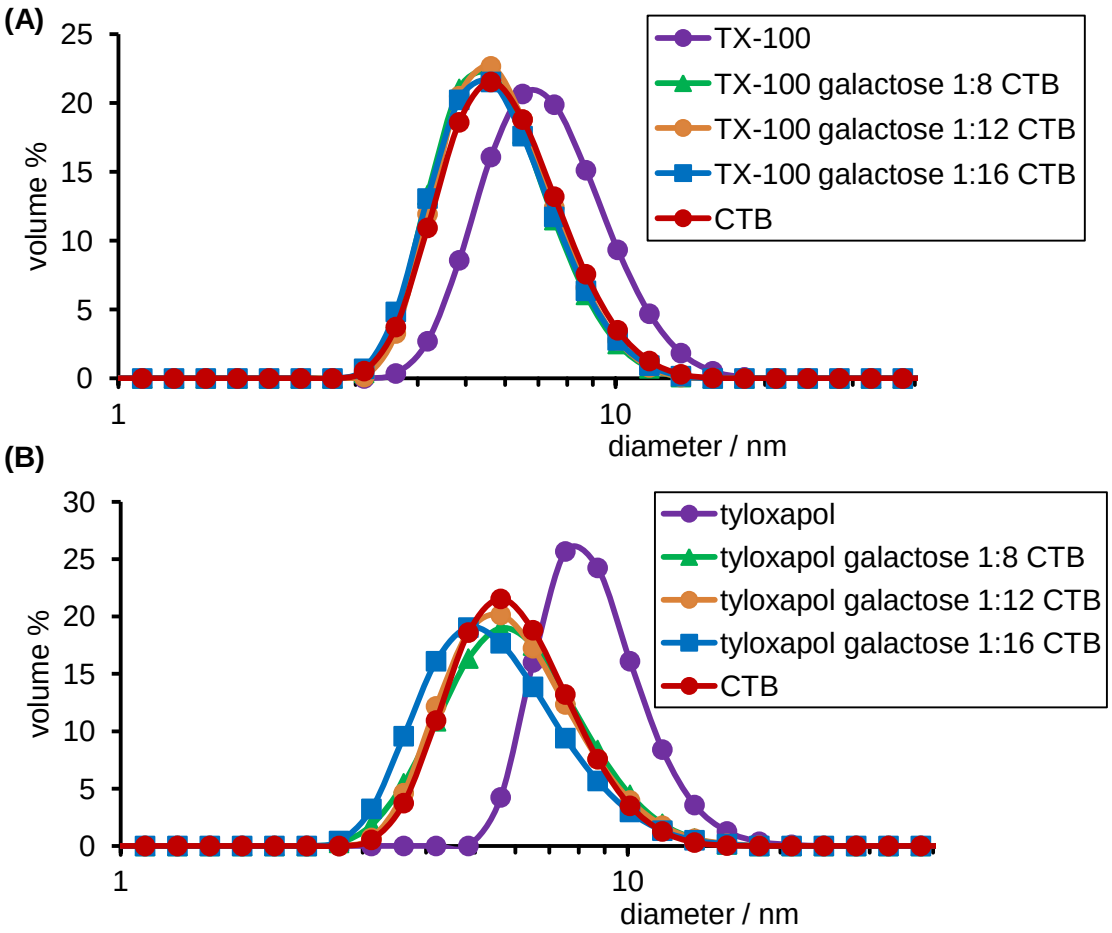


Fig 5.7 DLS of 1 mM detergents at 10% modification mixed with CTB at varying ratios. (A) TX-100 galactoside **57** and (B) tyloxapol galactoside **59**.

As no interactions had been observed by DLS, further analysis was needed to show whether the glycosylated micelles were capable of binding to CTB. Competition ITC titrations were therefore performed to further understand the binding interactions between the modified detergents and CTB. CTB was premixed with the modified detergents before GM1os was titrated in. Interactions between the detergents and CTB would lead to an apparent reduction in the GM1os interaction. Using a competing detergent concentration of 3.2 mM and the carbohydrate modification at 50% of detergent molecules gave a competing ligand concentration of 1.6 mM. The detergents were mixed with CTB (12 μM monomer concentration) and GM1os (140 μM) was titrated into the solution (Fig 5.8).

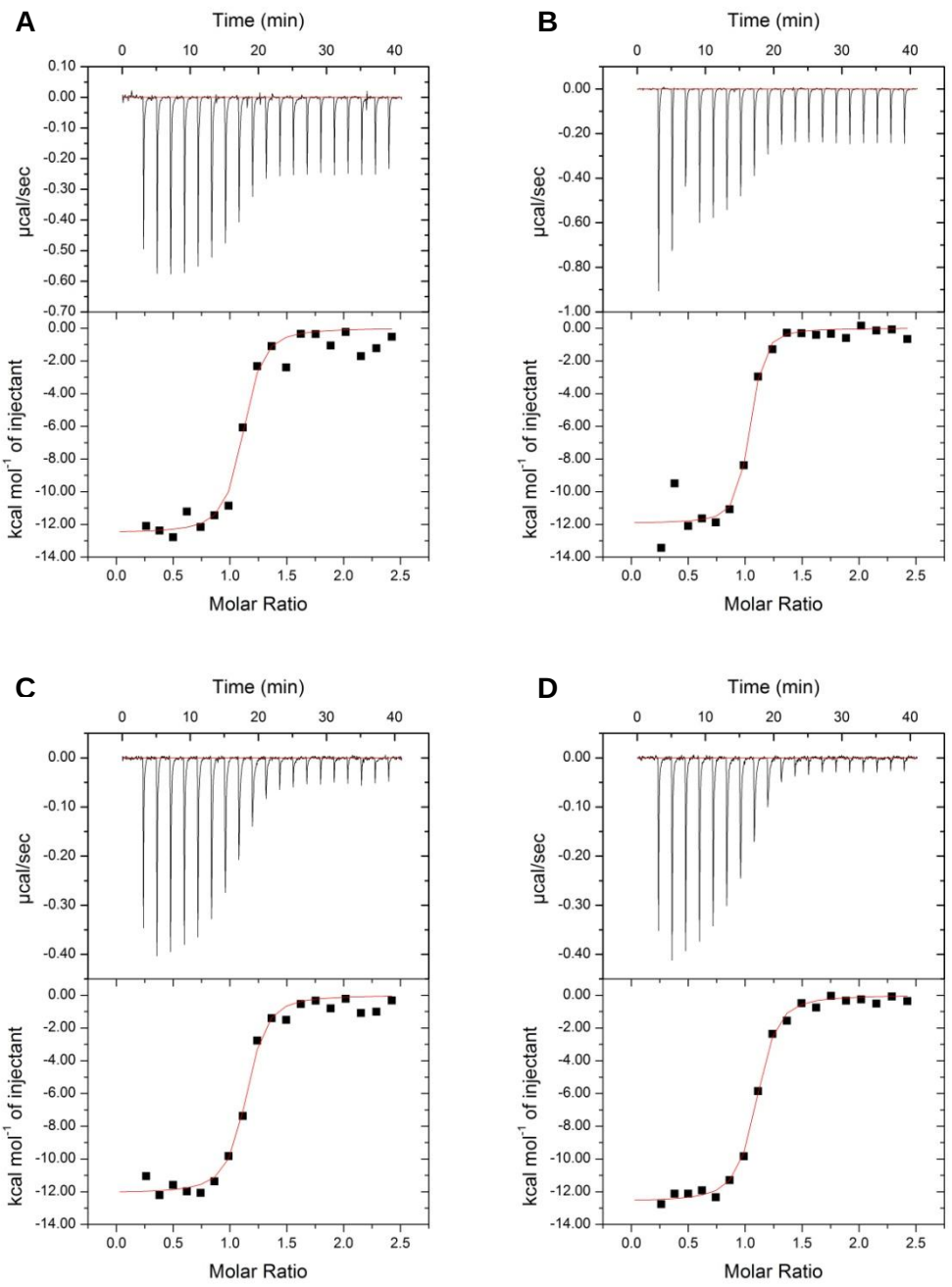


Fig 5.8 ITC traces of competition titrations with A) Tyloxapol galactose **59**, B) TX-100 galactose **57**, C) Tx100 lactose **58** and D) only the GM1os titration.

Titrations with all three modified detergents, **57**, **58** and **59**, showed no change to the usual GM1os binding (Table 5.2). The observed K_d for each titration was within the same range as the direct GM1os titration and there was no change in the enthalpy for the interaction which is often seen in competition titrations.[89] The ITC titrations were performed with the concentration of competing ligand at 1.6 mM which was around the expected K_d for galactose (12 mM). Whilst this concentration was not greater than the K_d it was hoped that a multivalent effect of having multiple

ligands fairly closely packed on the micelles may have contributed towards an enhanced binding affinity.

Table 5.2 ITC data for the titrations with premixed detergents and for the direct titration with GM1os.

Competing ligand	K_d / nM	ΔH° / kcal mol^{-1}	n
TX-100 galactoside **57**	37 ± 19.2	-11.9 ± 0.4	0.99
TX-100 lactoside **58**	72 ± 33.3	-12.5 ± 0.4	1.06
Tyloxapol galactoside **59**	90 ± 29.5	-12.8 ± 0.1	0.99

Ligand	K_d / nM	ΔH° / kcal mol^{-1}	n
GM1os	81 ± 15.0	-12.6 ± 0.2	1.05

The micelle concentration was only 32 µM, 100 times less than the overall detergent concentration due to the N_{agg}. Therefore the ligands were clustered in groups of around 10 galactose units at the concentration of 32 µM. This concentration is well below the K_d for galactose and so it is perhaps not surprising that no binding was observed by ITC. The micelles would present a higher effective concentration but in solution the ligands would be more disperse. The nature of the weak galactose interactions restricts the concentration range that is workable for the system and may have contributed to the failure of the modified micelles to bind to CTB.

The same problem was likely the cause for failure in the DLS experiments; the concentrations used were simply too low. But it would not have been practical to increase the concentrations enough to approach the K_d of galactose. The enhanced binding of a multivalent system previously seen with polymers and dendrimers was not observed with these micelles.[102, 114]

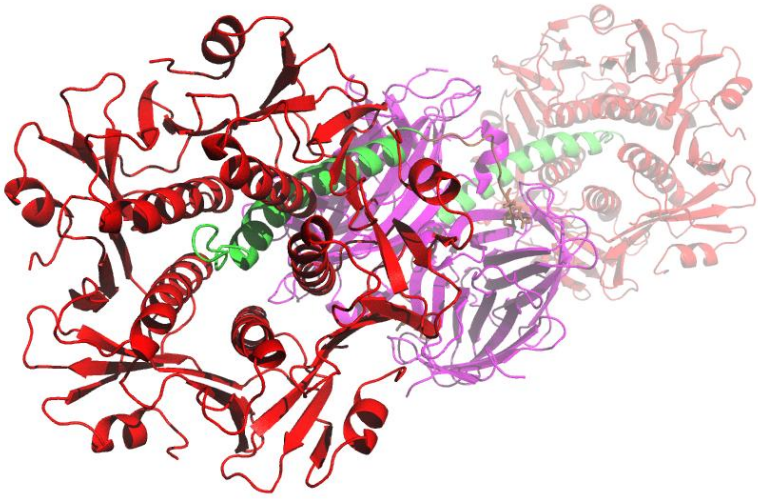
5.1.3 Conclusions

The TX-100 and tyloxapol micelles are potentially the perfect size to be templates for CTB. However, the unpredictable micelle structure when high percentages of carbohydrates were incorporated onto the detergents made their use with TX-100 unfeasible. Tyloxapol, however, provided a more stable structure with a lower CMC value. The low affinity of the carbohydrate interactions was proved to be an issue

and was likely the deciding factor in why no interactions were seen between the micelles and CTB.

A solution to the low affinity problem could be to use detergents modified with GM1os ligands. Whilst this is a larger carbohydrate and so may prove more of a problem for forming stable micelle structures, the use of tyloxapol could overcome these problems. The GM1os ligand has a much higher affinity for CTB and could be used at lower concentrations which did not work for the galactose-based system. Less ligands may be needed as well as a single GM1os should be strong enough to hold a CTB unit to the surface of the micelle. However, using GM1os ligands brings the problem of reduced reversibility of the interactions. There are no ligands available that could out-compete the GM1os ligands and so CTB could become stuck to the micelle surfaces without the subsequent options to cross-link them, to remove them or to rearrange the structures.

6 Tying the assembly together



Using the AB_5 complex to preorganise assembly

Cholera toxin exists as an AB_5 protein. After removal of the toxic subunit (CTA1), the CTA2 α helix extends from CTB out of the opposite face to the GM1os binding sites (Fig 6.1 A). The N-terminus of CTA2 is positioned at the end of this α helix and is an attractive target for site-specific modification. The N-terminus of CTB has been modified with ligands (chapter 2) and so the CTA2 N-terminus could also be targeted in a similar way (Fig 6.1 B).

The CTA2 peptide, whilst existing as an α helix when the full CTA unit is present, may actually become an unstructured peptide chain once cleaved. This unstructured nature may be beneficial for modification as it should add more flexibility to the peptide. If multiple copies of the AB_5 protein could be grouped together by clustering their CTA2 peptides, then the CTB units would all be presented facing outward (Scheme 6.1). With the CTB proteins now in close proximity, they could be linked together with multivalent ligands and there would be no need for a separate template.

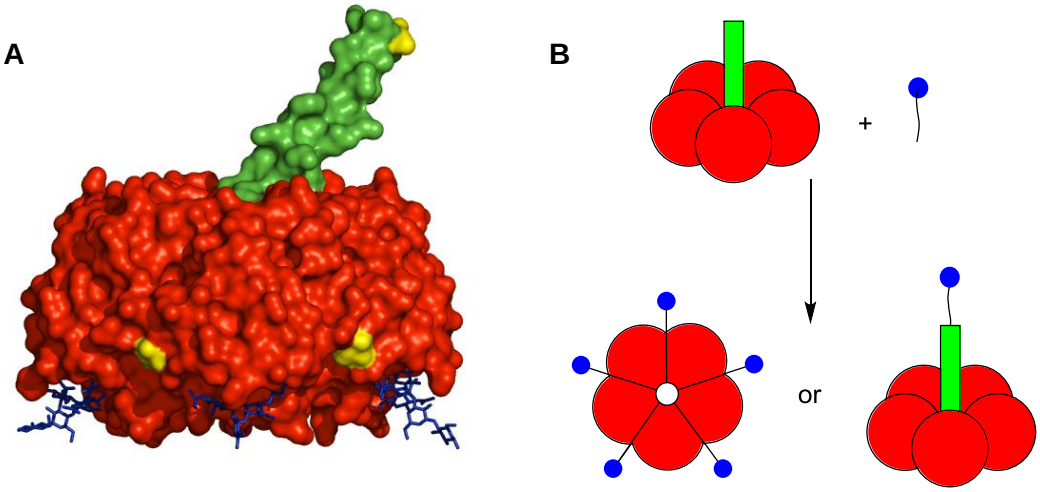


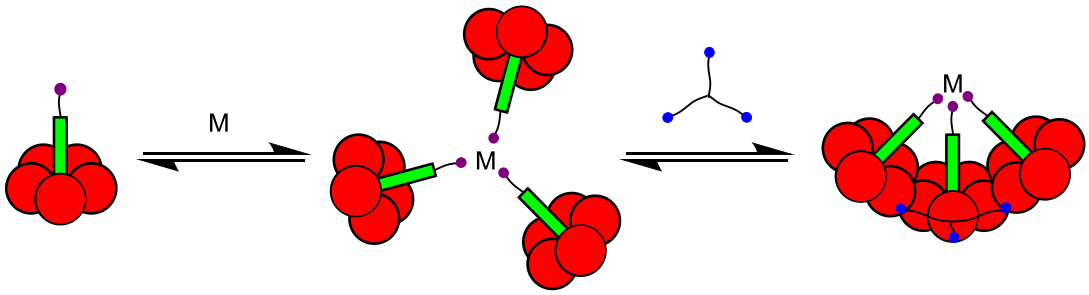
Fig 6.1 (A) Surface representation of the cholera toxin AB_5 after the removal of toxic CTA1, with the N-termini of both CTB and CTA2 highlighted (yellow) and the GM1os binding shown (blue). (B) Cartoon model for the potential modifications of the N-temini of CTB or CTA2 with ligands.

This chapter outlines the steps taken to modify the CTA2 peptide with ligands with the aim of bringing together multiple copies of the AB_5 protein. The use of a number of ligands is discussed and the different methods of their attachment onto CTA2.

6.1 Making use of metal complex formation

A strategy was needed to modify the CTA2 N-terminus with a ligand or functional group that could be used to join the proteins together. One method for achieving

this arrangement would be to attach a metal binding group to the CTA2 proteins and then coordinate multiple copies around a metal centre (Scheme 6.1). The use of metal coordination has been shown recently in a number of studies to be a useful method for the design of protein assemblies.[169, 170] A metal favouring an octahedral geometry binds three bidentate ligands. Once three AB_5 proteins were clustered around a metal centre then a trimeric ligand could be used to bind to the CTB subunits and crosslink the proteins. The arrangement would be constrained into a trimer of pentamers that would be effectively a quarter of a dodecahedron made up of AB_5 proteins. To achieve a full dodecahedron, further ligands could be introduced to cluster four of the trimers together to form a dodecamer. The flexibility of the CTA2 peptides would be of great importance as they would have to be able to bind to the metal and still bend around so the CTB moeities were in close proximity. The length and flexibility of the trimeric ligand would also have to be quite specific; being long enough to reach three CTB pentamers but not so long or flexible as to bind two or more binding sites on the same pentamer.



Scheme 6.1 The use of metal (M) binding to arrange three AB_5 proteins with bidentate ligands in an octahedral geometry. A trimeric carbohydrate ligand could then be used to constrain the proteins.

6.1.1 Investigating the metal binding

Fe^{2+} exhibits an octahedral coordination geometry with bound ligands. The ligand 2,2'-bipyridine (bpy) was chosen as a bidentate ligand which binds to Fe^{2+} in a 3:1 ratio. The binding of bpy derivatives produces a deep red colour with an absorbance maximum at 544 nm and an extinction coefficient of 8650 M^{-1} cm^{-1} (Fig 6.2). The absorbance was sufficient that, if in a complex with proteins at high micromolar concentrations, the red colour change and 544 nm peak would still be observable at the concentrations used in our studies. The bpy coordination to Fe^{2+} has also been shown to be dynamic and can adjust the spatial arrangement of the ligands to fit the needs of the system.[171]

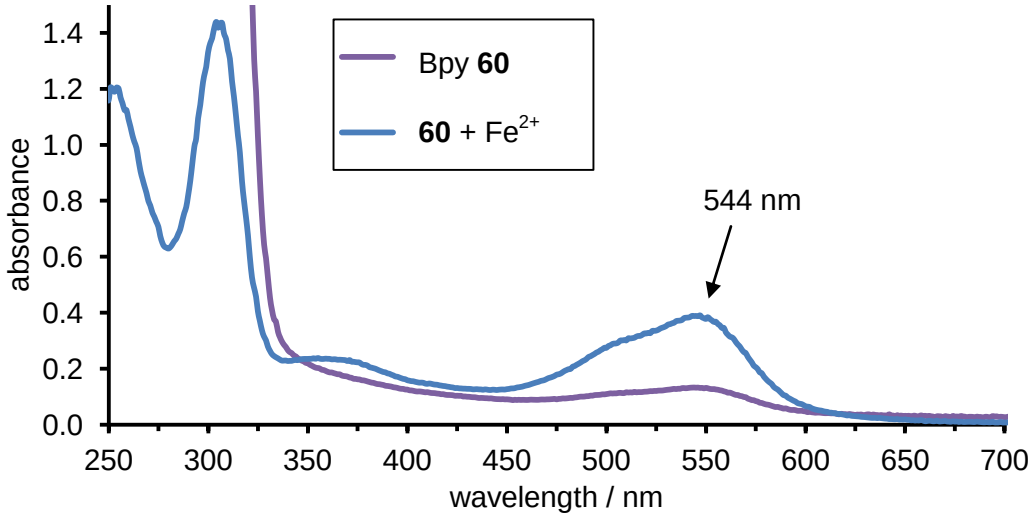
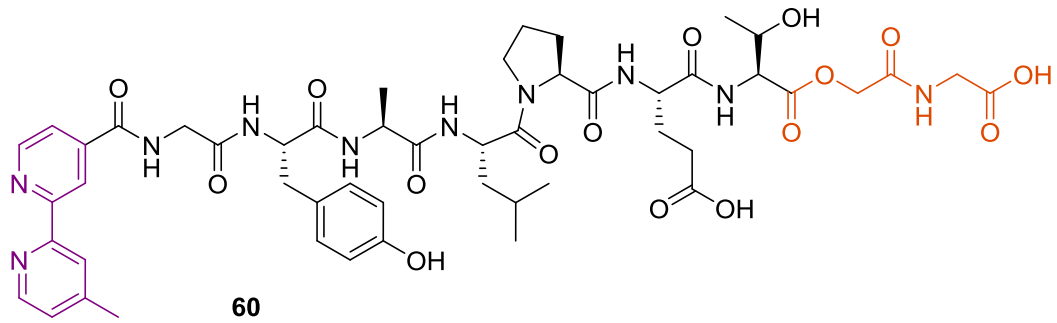


Fig 6.2 UV spectrum of depsipeptide **60** (150 µM) binding to Fe^{2+} (50 µM).



60

Fig 6.3 Structure of bpy containing depsipeptide **60**.

Fe^{2+} is incompatible with many buffers as it precipitates, but even in the presence of many buffer salts (HEPES, TRIS, NaCl, PO_4 etc) the complex with bpy is still formed and remains stable in solution. If AB_5 proteins were modified with bpy-depsipeptide **60** (Fig 6.3), they should be able to complex the metal in the same way, bringing three proteins together stable in a buffered solution.

The sortase enzyme was used to attach the metal binding bpy group to CTA2. The enzyme is a trans-peptidase from *S. aureus* that can be used for ligating peptides that contain a C-terminal LPETG sequence and an N-terminal glycine residue. A cysteine residue in the sortase enzyme becomes covalently attached to the LPET motif forming a thioester. The N-terminal glycine motif can then react with this intermediate to form an amide bond.

It had been seen previously by work in the Turnbull lab that depsipeptides containing ester linkages rather than amide bonds can enhance the reaction of an LPETG motif to GGG with the sortase enzyme.[172] The use of the ester linkage in a

depsipeptide, rather than an amide bond, means that once the ester bond was cleaved from the LPETG sequence then the hydroxyacetyl byproduct is not a substrate for the reverse reaction with the enzyme.[172] The enzymatic reaction is therefore no longer in equilibrium and can reach completion without a large excess of substrate. Hydrolysis of the ester can be a problem over time, but this can be overcome by using more equivalents in the reactions. The short depsipeptide **60** was synthesised by SPPS containing the bpy function and the C-terminus ending with an LPETG sequence (synthesised by Miss Fiona Ng) (Fig 6.3).

After the purification of depsipeptide **60** the UV spectrum showed a very small residual absorbance at 544 nm indicating a small amount of metal was bound that could not be removed by purification (Fig 6.2). $FeCl_2$ was titrated against bpy **60** to analyse the absorbance and find the amount of metal already bound to the ligand (Fig 6.4). The absorbance increased as Fe^{2+} was titrated in to the bpy **60** solution. The absorbance then reached a maximum when all the Fe^{2+} was bound to the ligand at a ratio of 0.22. This was below the expected ratio of 0.33 and showed that 34% of the ligand was already bound to a metal. As depsipeptide **60** did bind Fe^{2+} it was thought that the impurity would not significantly interfere with further studies.

The strength of the binding interaction in the metal complex was investigated by ITC. $FeCl_2$ (500 µM) was titrated into bpy depsipeptide **60** (50 µM) and an apparent affinity with a K_d of 2.9 µM was found for the interaction (Fig 6.5). A stoichiometry of 0.72 was observed which was difficult to interpret as a stoichiometry of 0.33, or lower if some of the ligand was already bound to metal, was expected due to the 3:1 ratio of ligand to metal binding. As excess iron enters the system, the complexes may adjust to form not only $Fe(bpy)_3$ but also $Fe(bpy)_2$ or even Fe(bpy) which would mean that the curve no longer represents the binding stoichiometry. However, the low micromolar K_d showed that the interaction would be strong enough and accessible at the micromolar concentrations normally used for protein reactions and other biophysical analysis.

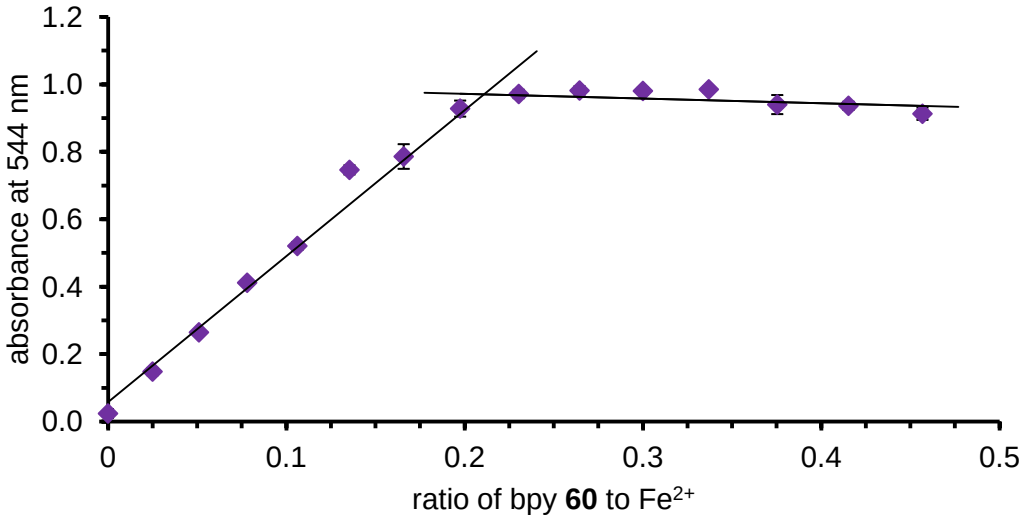


Fig 6.4 Titration of $FeCl_2$ against depsipeptide **60** showing the absorbance at 544 nm.

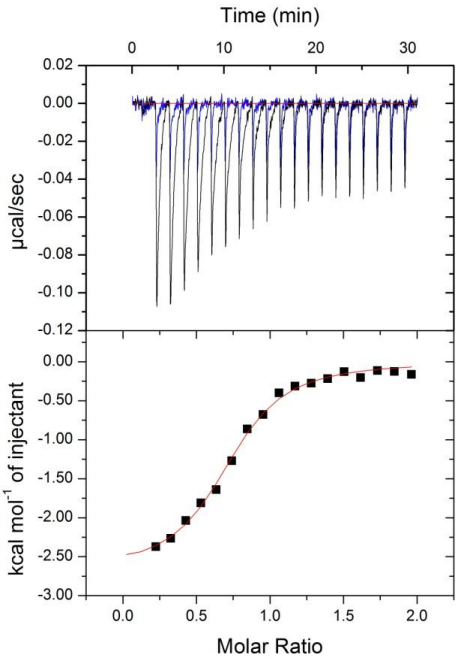


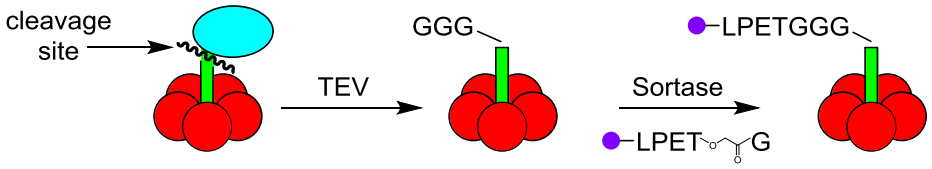
Fig 6.5 ITC trace for Fe^{2+} binding to bpy depsipeptide **60**.

6.1.2 AB_5 purification and modification

A suitable AB_5 protein needed to be made for attachment of bpy depsipeptide **60** to CTA2. The AB_5 protein was expressed as the CTB protein and a maltose binding protein fusion with CTA2 (MBP-CTA2). The two proteins were encoded in the same plasmid and expressed together forming the AB_5 in the bacteria before being exported into the growth medium. The MBP-CTA2 fusion was constructed to aid the solubility of CTA2 and help the purification process. The MBP protein would first

have to be cleaved before later reactions could be performed on the CTA2 N-terminus.

The tobacco etch virus (TEV) protease is specific to the recognition sequence of ENLYFQG with cleavage occurring between the Q and G residues of the protein. A TEV cleavage site had been incorporated into the sequence between MBP and CTA2 in the fusion protein and so after cleavage would reveal a GGG motif on the N-terminus of CTA2 (Scheme 6.2). This motif would then be accessible for the sortase enzyme to append a peptide containing a LPETG C-terminal sequence, such as depsipeptide **60**. TEV cleavage can occur at unspecific sequences in the protein after a long incubation period or if a large mol% of TEV is used. Care was needed to optimise the cleavage conditions for the use with the MBP-CTA2 protein.



Scheme 6.2 TEV cleavage of MBP-CTA2 followed by the attachment of a short depsipeptide to the revealed N-terminus of CTA2 via sortase ligation. The proteins are CTB (red), CTA2 (green) and MBP (cyan).

The MBP-AB_5 protein was encoded in plasmid pSAB2.1T. The N-terminus of the CTB protein encoded by this plasmid was a threonine residue (DNA constructed by Mr James Ross) so that oxime chemistry may also be possible with the CTB subunit. The protein complex was expressed in *E.coli* by induction with IPTG and growth at 30 °C for 24 h. After purification by ammonium sulfate precipitation, nickel affinity chromatography and then SEC, a yield of 2.5 mg L^{-1} was obtained (Fig 6.7). The concentration of the MBP-AB_5 protein was measured by UV spectroscopy at 280 nm, using a theoretical molar extinction coefficient of 128,745 M^{-1} cm^{-1}, calculated from the sequence by ExPASy Proteomics. ESMS showed two proteins with masses 48831.2 Da and 11643.2 Da for MBP-CTA2 and CTB respectively (calculated 48823.9 Da and 11642.9 Da) (Fig 6.8). The accuracy of the mass measured by ESMS reduced as the size of the proteins increased. The increased broadness of the high charge states led to a less precise measurement, but this was still within the right range for the expected mass of MBP-CTA2.

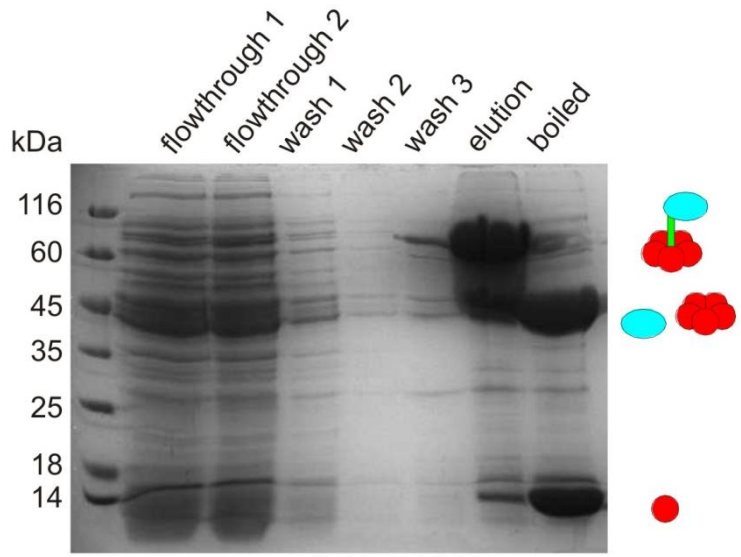


Fig 6.6 SDS-PAGE of AB_5 expression and purification in a Ni affinity column.

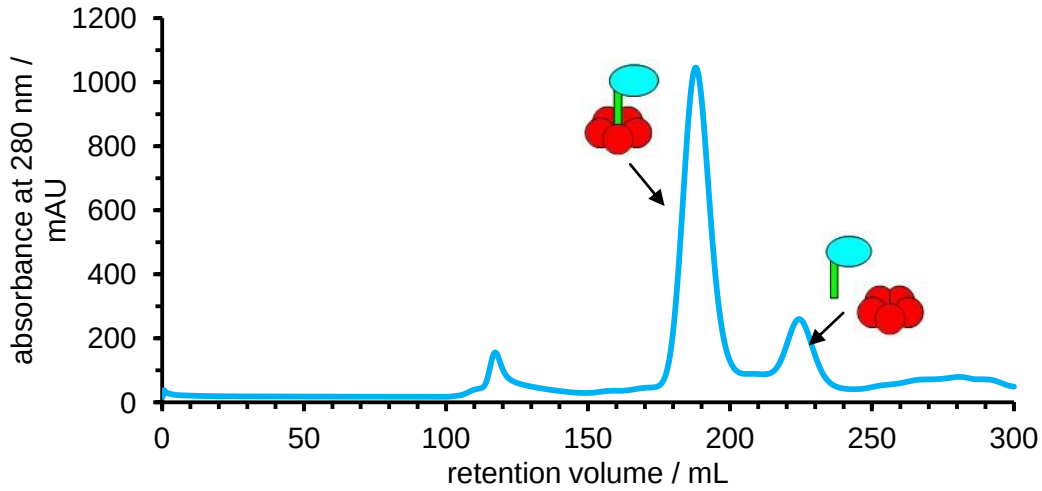


Fig 6.7 Chromatogram showing purification of MBP-AB_5 after expression and initial purification by Ni affinity chromatography.

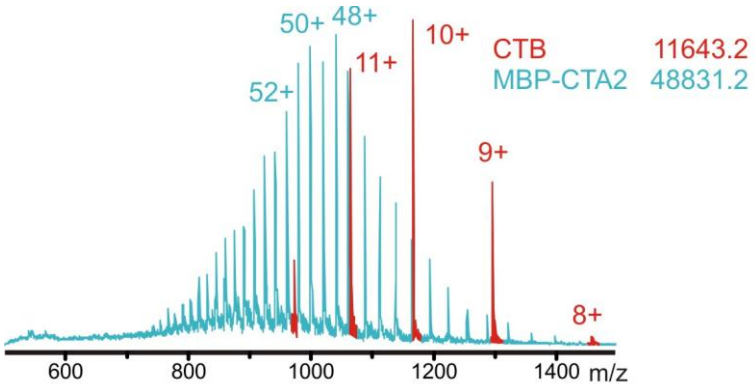


Fig 6.8 Mass spectrum of the purified AB_5 showing the separate MBP-CTA2 and CTB.

The MBP-AB_5 protein was found to be fairly unstable at room temperature or after incubation at 4 °C for more than a few days. The MBP-CTA2 fusion was seen to

denature and no longer stayed as the MBP-AB_5 complex. AUC was performed on a freshly purified sample and a sample that had been stored at 4 °C for 7 days, to analyse the size of the protein and its stability (Fig 6.9). The fresh sample showed a peak at 7.6 S which corresponded to a mass of 92.7 kDa matching fairly well with the expected mass of the MBP-AB_5 of 107 kDa. But the “old” sample showed a single peak at 5.8 S corresponding to a mass 45.6 kDa showing that the MBP-AB_5 was no longer stable in solution and had dissociated to yield the CTB and MBP-CTA2 separately.

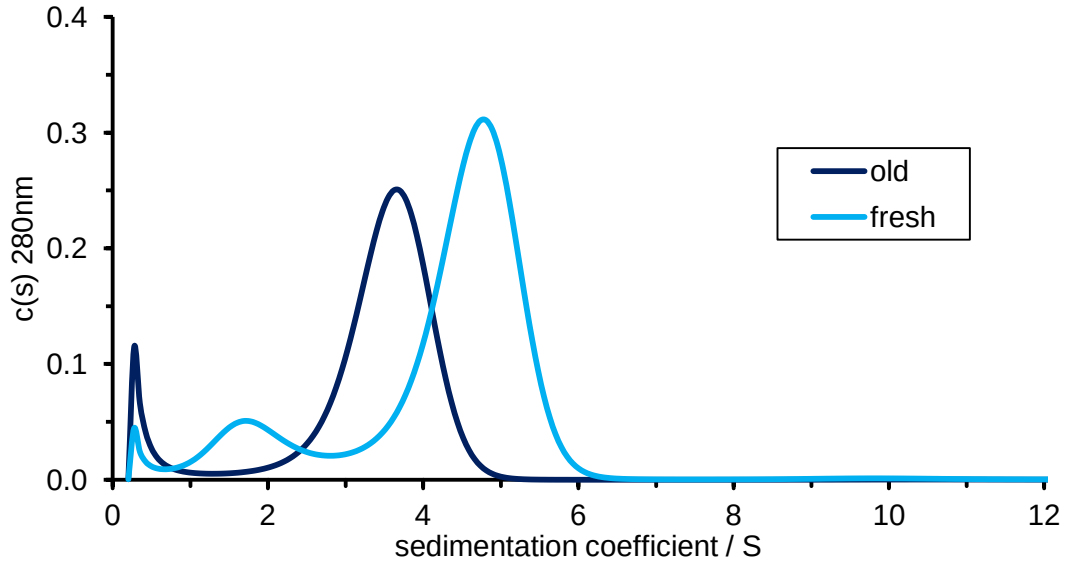


Fig 6.9 AUC c (s) plot of MBP-AB_5 performed at 30 krpm. A “fresh” sample showed clearly as the MBP-AB_5 protein and an “old” sample incubated at 4 °C for 7 days showed only a smaller species around the size of both CTB and MBP-CTA2.

The AUC samples were further analysed by SEC which confirmed the difference between the samples (Fig 6.10). For the fresh sample SEC showed a large peak with a retention volume of 15.8 mL for the AB_5 protein and a smaller peak that showed some of the tertiary structure had began to dissociate. The old sample showed a single peak with retention volume 18.5 mL for the CTB and MBP-CTA2 moeities.

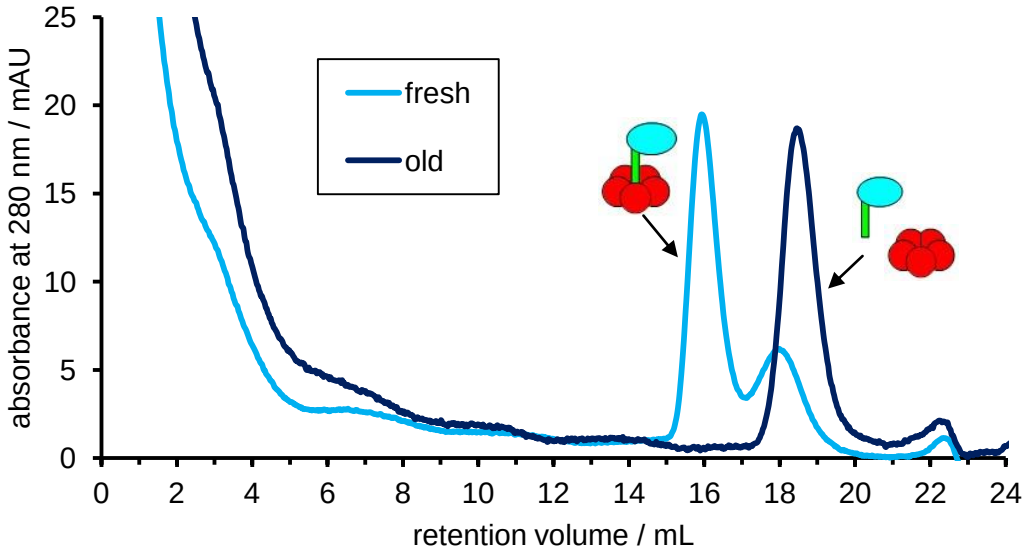


Fig 6.10 Chromatogram of the samples from the AUC experiment, showing the degradation of the MBP-AB_5 protein over time.

TEV (5 mol%) was incubated with MBP-AB_5, to cleave the protein, at room temperature in *HEPES buffer* and after 1.5 h the cleaved AB_5 was purified on a Ni affinity column. This affinity column captured CTB, and therefore the associated CTA2, bound to the resin and the cleaved MBP and TEV protease were removed. The AB_5 protein was eluted with 500 mM imidazole and ESMS analysis showed a mass of 5056.4 Da (calculated 5055.5 Da) for the cleaved CTA2 and the still intact CTB with a mass of 11642.9 (calculated 11642.9) (Fig 6.11).

6.1.3 The complexation of bpy-AB_5

After MBP-AB_5 was cleaved and purified to reveal the CTA2 N-terminus, the AB_5 protein was then incubated with sortase (10 mol%) at 37 °C and bpy depsipeptide **60** (5 equivalents). After 4 h the reaction product was purified by SEC and ESMS gave a result of 5984.3 Da (calculated 5983.5 Da) for the modified CTA2 (Fig 6.11) as part of the **A(60)B_5** complex.

A(60)B_5 (25 µM) was mixed in a 3:1 ratio with $FeCl_2$ with the aim to form a complex around the metal with three AB_5 proteins. DLS was used to analyse the mixture but no complexation was observed (Fig 6.12). **A(60)B_5** showed a peak corresponding to a diameter of 8.1 nm which was larger than the 5.8 nm observed for CTB. But when $FeCl_2$ was added no change in the size distribution was observed.

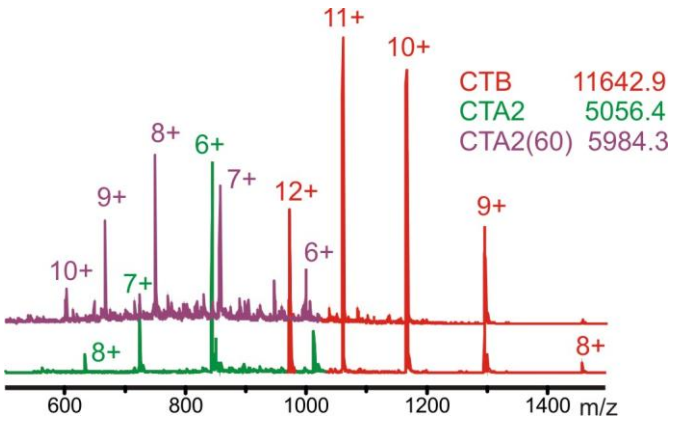


Fig 6.11 Mass spectrum of **A(60)B$_5$** showing complete modification of CTA2 and unmodified CTB.

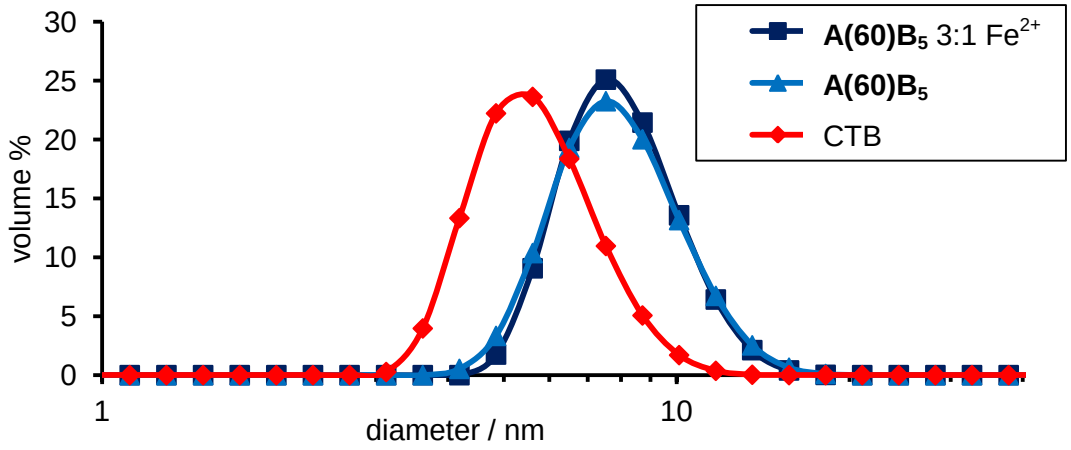


Fig 6.12 DLS results for **A(60)B$_5$** (blue) giving a diameter of 8.1 nm compared to CTB (red) at 5.8 nm but no increase in size shown on addition of $FeCl_2$ to **A(60)B$_5$** (dark blue).

The metal complex formation was investigated using UV spectroscopy. The 544 nm peak could be clearly seen when complexation occurred between the free bpy ligand **60** and $FeCl_2$. However when **A(60)B$_5$** was mixed with $FeCl_2$ at the same concentration (25 µM) no peak at 544 nm was observed and so the metal and protein were presumably not interacting (Fig 6.13). There was also no change in absorbance seen when the amount of Fe^{2+} was in excess at 10 equivalents.

As the metal binding was not seen to take place once the bpy ligand was attached to the protein, it was thought that the $Fe(bpy)_3$ complex could be formed first and then ligated to three CTA2 peptides simultaneously. The bpy depsipeptide **60** was premixed with $FeCl_2$ at a ratio of 3:1, resulting in the solution turning an intense red colour indicating that metal complexation was successful. Sortase was used to react the AB_5 protein with this metal complex and after 1 h the red colour had been lost to leave a colourless solution. After incubation for 4 h at 37 °C ESMS indicated that no reaction had taken place and only unreacted AB_5 was present. The

disappearance of the red colour suggested that the metal complex was being disrupted and an interaction between AB_5 and either the bpy group or the metal was taking place. This interaction must be of higher affinity than the $Fe(bpy)_3$ complex in order to disrupt it.

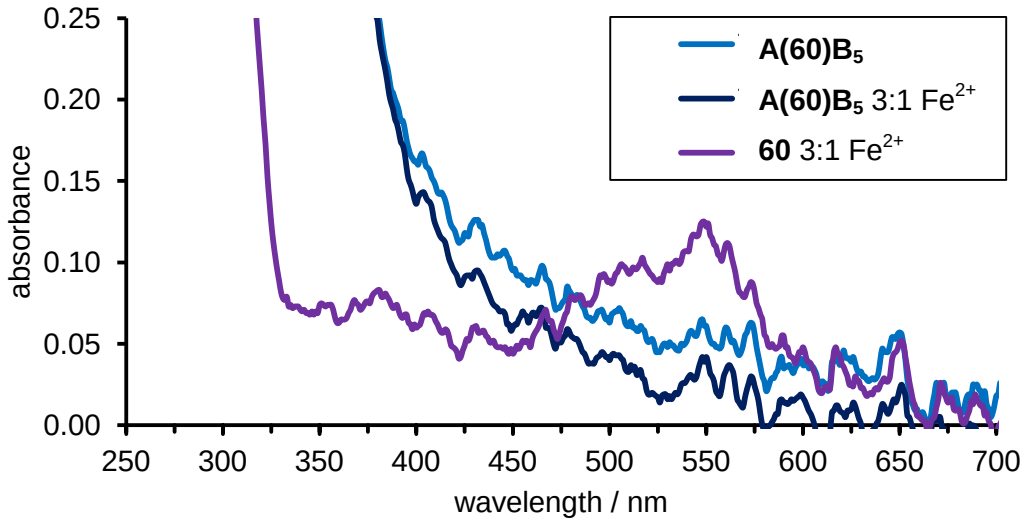


Fig 6.13 UV spectrum showing the 544 nm peak for bpy depsipeptide **60** binding $FeCl_2$ (purple) (25 µM). No peak was observed when **A(60)B$_5$** was mixed at a 3:1 ratio with Fe^{2+} (dark blue) or with excess Fe^{2+} (light blue).

UV spectroscopy was used to analyse conditions under which the metal complex formed. The $Fe(bpy)_3$ complex formation was tested in the presence and absence of the CTB and AB_5 proteins (Fig 6.14). It was found that the complex was formed as normal in the presence of CTB but not in the presence of AB_5 or when bpy **60** was attached to the AB_5 as **A(60)B$_5$**. Even increasing the Fe^{2+} concentration to 10 equivalents of Fe^{2+} per bpy **60** failed to form a complex in the presence of the AB_5 protein. These results indicated that Fe^{2+} was not complexing to the CTB protein and this protein did not inhibit the metal complexation. Therefore it was concluded that it was the CTA2 peptide that was interacting with the bpy group and preventing the formation of the metal complex. This interaction may be due to the sequence and unstructured nature of the CTA2 peptide that could be binding to the bpy group preventing any interactions with the metals.

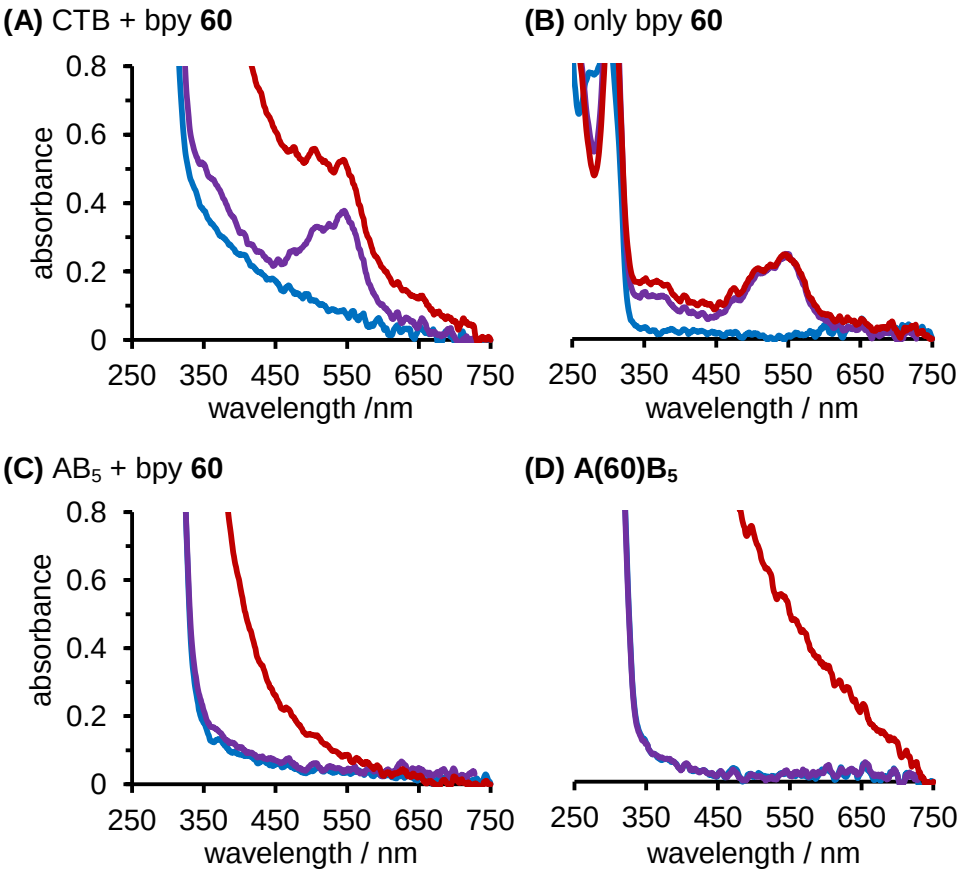


Fig 6.14 UV spectra of the proteins mixed with bpy **60** in the absence of Fe^{2+} (blue), at a ratio of 3:1 with Fe^{2+} (purple) and with an excess of the metal at 1:10 Fe^{2+} (red). The 544 nm peak was present in (A) CTB and bpy **60**, and (B) no protein and bpy **60**. But no absorbance was seen with (C) AB_5 and bpy **60**, or (D) **A(60)B$_5$**.

6.1.4 The creation of a new CTA2 sequence

The bpy metal binding motif failed to bind to Fe^{2+} in the presence of the AB_5. It was therefore postulated that changing the amino acid sequence of the CTA2 peptide might encourage binding. A polyasparagine chain was chosen as an appropriate sequence as it should form an extended chain giving plenty of space for the metal complex to form and UV spectroscopy showed that a polyasparagine sequence did not prevent the bpy **60** complex formation with Fe^{2+}.

A new gene was designed to encode the CTA2 peptide with the ENLYFQGGG sequence still intact for TEV cleavage, followed by five asparagine amino acids. All subsequent amino acids were removed up until the point at which the CTA2 peptide started to interact with the CTB pentamer and penetrate the centre of the pentamer (Fig 6.15). It was hoped that these alterations to the original sequence should

remove as much of the peptide sequence as possible and so prevent the binding to bpy **60** whilst still keeping the oligoglycine sortase tag and not affecting the interactions between CTA2 with CTB.

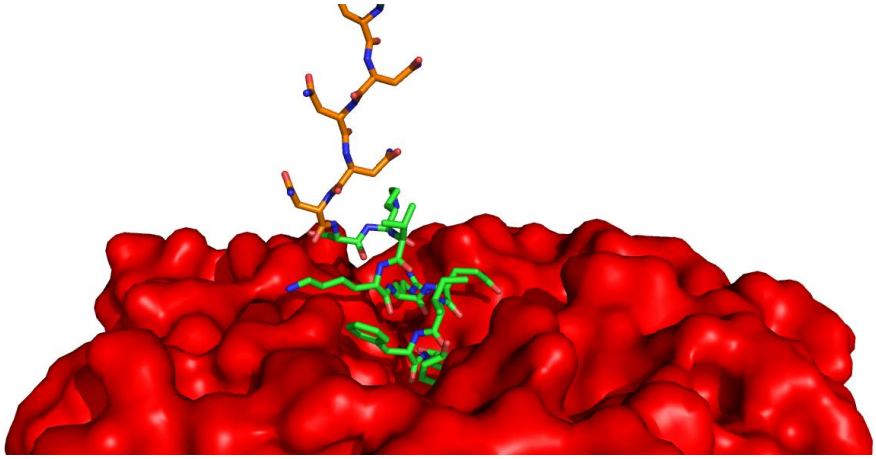


Fig 6.15 Representation of the interface where CTA2 penetrates through the centre of CTB. A surface is shown on CTB (red) and CTA2 is shown as a stick representation of the amino acids (green) with the polyasparagine mutations (orange).

The new gene was constructed from four short DNA parts by polymerase cycling assembly (PCA). After a successful PCA reaction, the original CTA2 gene was cut from plasmid pSAB2.1T before the new gene was ligated in to form the new plasmid pTRBAB5-D4N. *E. coli XL10* cells were transformed with the new plasmid before a plasmid prep was performed to purify the DNA for sequencing and for transformation into *E. coli BL21gold* cells for expression. The new plasmid encoded the MBP-CTA2 fusion protein with the new polyasparagine CTA2 sequence and the CTB protein.

The new MBP-AB_5 (D4N) protein was expressed in *E. coli BL21gold* containing the new pTRBAB5-D4N plasmid by induction with IPTG and growth at 30 °C for 24 h. Purification by ammonium sulfate precipitation and nickel affinity chromatography was shown by SDS-PAGE (Fig 6.16). SEC was used to isolate the MBP-AB_5 (D4N) protein and ESMS was performed that gave a mass of 47529.9 Da (calculated 47488.3 Da) for the MBP-CTA2 (D4N) fusion and a mass of 11647.1 Da (calculated 11643.3 Da) for CTB (Fig 6.17).

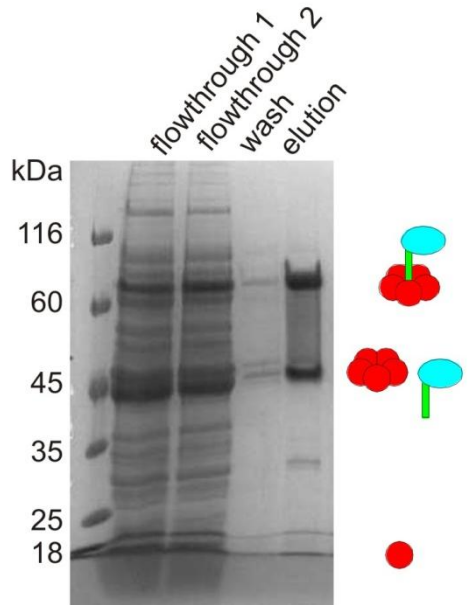


Fig 6.16 SDS-PAGE of the expression of AB_5 (polyasparagine CTA2) and purification by Ni affinity chromatography.

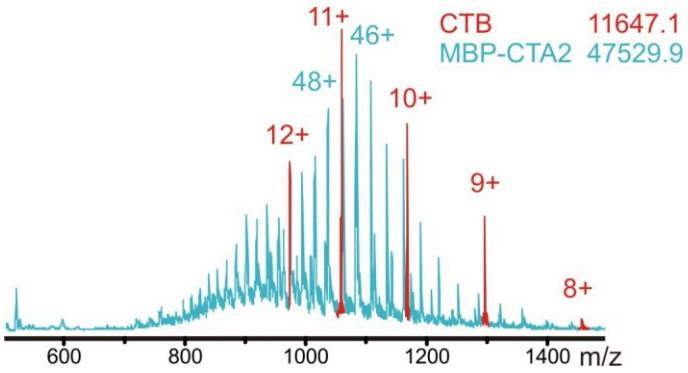


Fig 6.17 Mass spectrum of MBP-AB_5 (D4N) showing the two separate subunits MBP-CTA2 (D4N) and CTB.

After the new AB_5 protein was purified, the MBP-CTA2 (D4N) fusion was cleaved with TEV protease. ESMS showed the new CTA2 (D4N) peptide with a mass of 3723.2 Da (calculated 3719.9 Da) (Fig 6.18). The metal binding bpy depsipeptide **60** was conjugated on the CTA2 (D4N) N-terminus using the sortase enzyme to form **$A(60)B_5$ (D4N)**.The resultant modification was confirmed by ESMS showing the mass of 4651.2 Da (calculated 4647.9 Da) for the modified CTA2 (D4N) and an unaltered mass of 11643.2 Da (calculated 11642.9 Da) for CTB (Fig 6.18).

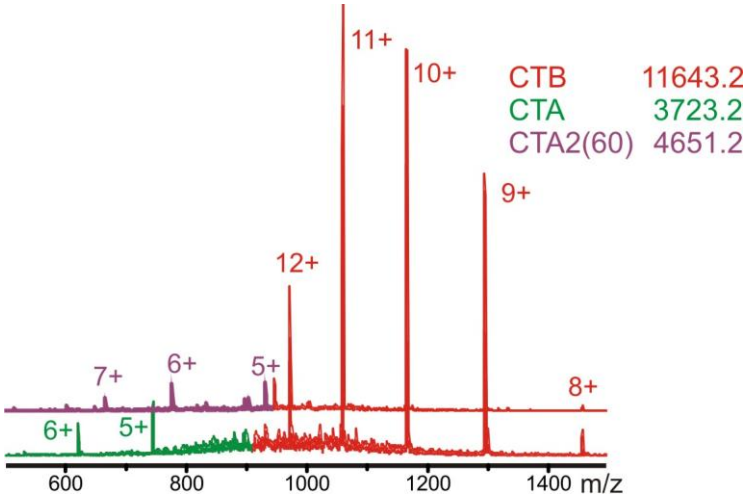


Fig 6.18 Mass spectrum of the new AB_5 (D4N) protein with the polyasparagine containing CTA2 peptide cleaved and after modification with bpy **60**.

Formation of the metal complex was again investigated using UV spectroscopy. The new **A(60)B_5 (D4N)** protein was mixed at 3:1 with $FeCl_2$ but just as before no absorbance was seen at 544 nm (Fig 6.19). The bpy ligand and the Fe^{2+} were again not able to form the complex in the presence of the protein. The new CTA2 (D4N) peptide had the same effect as before, preventing the metal binding. When bpy **60** and AB_5 (D4N) were mixed together with $FeCl_2$ at a ratio of 3:3:1 there was no absorbance seen at 544 nm as before. When the amount of free bpy was doubled again no change was seen but when the amount was tripled there was a sudden change in the absorbance profile. At three equivalents of bpy **60** to AB_5 (D4N), the ligand was able to bind to Fe^{2+} resulting in an absorbance peak at 544 nm. This observation suggested that the AB_5 protein was able to block two bpy ligands preventing their binding to the metal but addition of a third allows the metal complex formation.

This lack of metal complexation may have been the result of the bpy ligand binding to an unstructured peptide sequence as thought with the original CTA2 sequence. Polyasparagine sequences are used in the multiple cloning site of MBP fusion proteins to provide a spacer section between the MBP and protein of interest, and so it was thought that this new polyasparagine-containing sequence would provide the same elongated structure. The metal binding was seen to occur in the presence of a larger polyasparagine sequence and so the new sequence was not expected to cause these problems. Most of the peptide sequence had been changed and so this negative result may have been due to the small amount of CTA2 that remained as before. The section of the CTA2 sequence that begins to interact with and penetrate CTB may be the site of bpy binding. If this section of the CTA2 peptide is

to blame for the inactivity of the bpy groups towards Fe^{2+} binding then there is little more that can be changed. Altering this section of the peptide might seriously affect how the tertiary structure of AB_5 forms or if it forms at all.

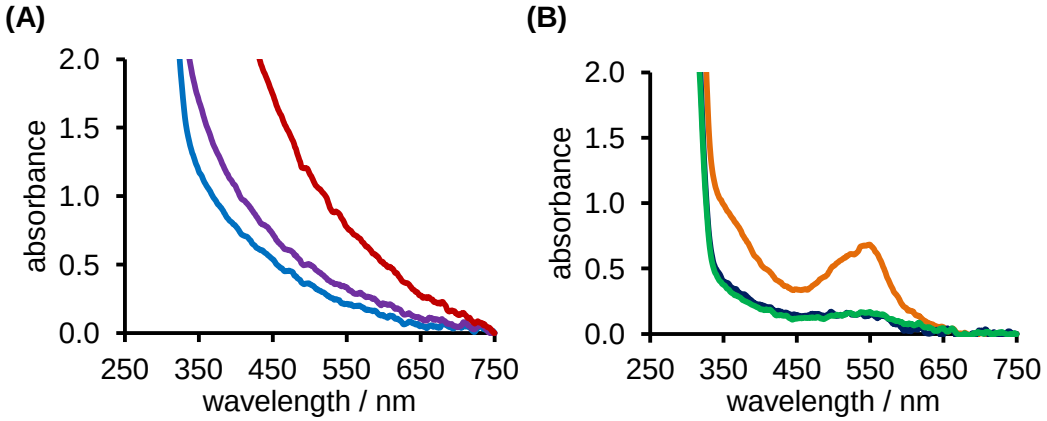


Fig 6.19 UV spectroscopy plots. (A) The polyasparagine CTA2 in **A(60)B$_5$** in the absence of Fe^{2+} (blue), with a ratio of 3:1 Fe^{2+} (purple) and with excess metal at 1:10 Fe^{2+} (red). (B) The polyasparagine CTA2 in the AB_5 protein mixed with bpy **60** with Fe^{2+} at ratios of 3:3:1 (dark blue), 3:6:1 (green) and 3:9:1 (orange).

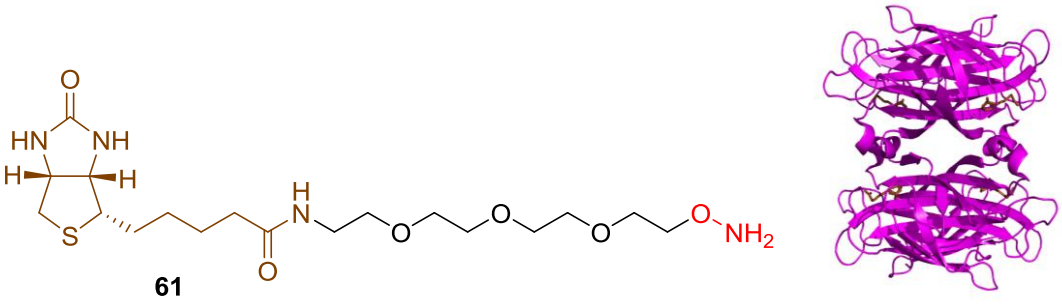
The CTA2 (D4N) sequence did, however, provide a more robust protein for TEV cleavage. Unspecific cleavage was seen with the original MBP-CTA2 sequence when extended incubation times and a large mol% of TEV were used. This unspecific cleavage was not observed with the MBP-CTA2 (D4N) protein, due to the alternative CTA2 sequence. This new protein could therefore be used in place of the original for further studies when TEV cleavage was required.

6.2 Protein-ligand interactions for assembly of the AB$_5$ complex

Another strategy considered for using the AB_5 protein to initiate assembly of the proteins was to attach a ligand for binding to a different protein. If the second protein had multiple binding sites then it could bring the multiple copies of AB_5 into close proximity for crosslinking with multivalent ligands. As the AB_5 protein would present only one ligand, a strong interaction would be needed to cluster the proteins.

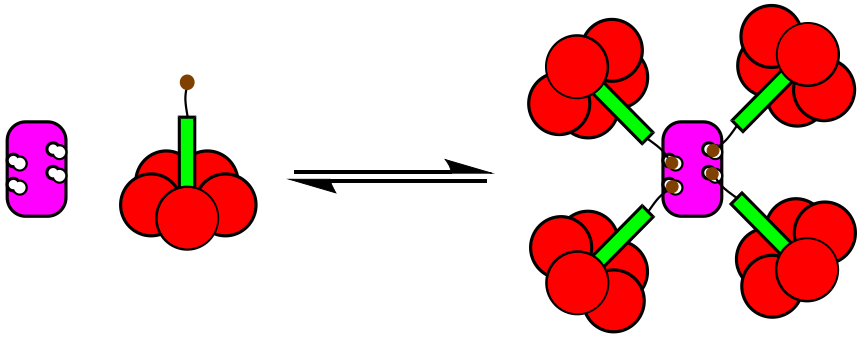
An obvious protein-ligand system to use was the streptavidin-biotin interaction which has a K_d of 10^{-14} M.[173] This affinity is exceptionally high for a protein-ligand interaction and easily high enough for the needs of this project. Streptavidin is a tetramer resembling a cuboid with dimensions of 6 × 5 × 4 nm with biotin binding sites in pairs on opposite faces of the protein (Fig 6.20).[174] If the AB_5 protein was

modified with biotin then potentially four AB_5 proteins could be bound to streptavidin (Scheme 6.3). The AB_5 proteins brought together in this large complex could then be used to seed the assembly of larger structures. A similar strategy to this was demonstrated recently by Mori *et al.* with the incorporation of biotin and bis-biotin ligands onto an alkaline phosphatase.[175] This work showed that the positioning and flexibility of the ligands could control the nature of the protein complexes formed with streptavidin.



61

Fig 6.20 The structure of biotin oxyamine **61** and the crystal structure of streptavidin (pink).



Scheme 6.3 The proposed assembly of streptavidin (pink) and a modified AB_5.

6.2.1 More AB_5 mutations

A biotin oxyamine **61** was commercially available (Thermo Scientific) and was suitable to be ligated onto the CTA2 peptide of the AB_5 protein. Oxime chemistry could be used, in the same way as previous reactions on CTB, which would expand the options for modification on AB_5 complexes. Before the ligation could be achieved, two changes to the AB_5 sequence needed to be made. The N-terminus of CTA2 had to be changed to a threonine or serine residue so that it was suitable for oxidation and the N-terminal threonine residue of CTB needed to be changed so that the reaction would be site-specific for only CTA2.

Plasmid pSAB2.1A contains the genes for MBP-CTA2 and CTB for expression in *E. coli*. The original plasmid encoded for CTB with an alanine residue at the N-terminus and so only the CTA2 peptide needed to be altered. Site-directed mutagenesis was performed to introduce a serine residue into the CTA2 gene so that the protein was suitable for N-terminal modification. Serine was chosen rather than threonine as the change from a glycine residue only required one mutated base pair (GGT to AGT) and still provided the amino alcohol group needed for oxidation. Gel electrophoresis was performed to assess the PCR and showed that the reaction had been successful to create plasmid pTRBAB5-G1S. *E. coli XL10* competent cells were transformed with the new plasmid. A plasmid prep was then performed to isolate the DNA for sequencing and for transformation into *E. coli C41* cells for expression.

The protein was then expressed from the *E.coli* by inducing with IPTG and growth at 30 °C for 24 h. After purification by ammonium sulfate precipitation of the media, a nickel affinity column and then SEC, a yield of 5.5 mg L^{-1} was obtained. SDS-PAGE confirmed the successful expression and the purification of MBP-AB_5 (G1S) (Fig 6.21) although SEC was needed to fully purify the MBP-AB_5 complex from excess uncomplexed CTB (Fig 6.22). MBP-AB_5 appears mostly as the complete complex on the gel but when boiled it fully dissociates into the MBP-CTA2 (G1S) fusion and CTB.

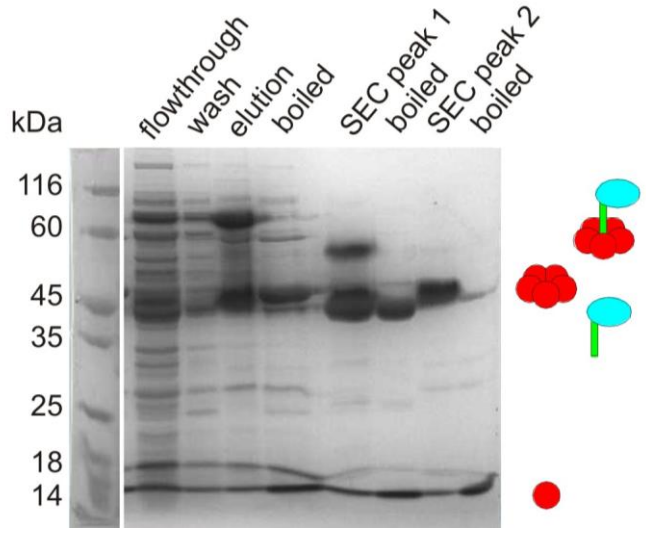


Fig 6.21 SDS-PAGE for MBP-AB_5 (G1S) nickel affinity column and SEC purification.

The concentration of the MBP-AB_5 (G1S) protein was measured by UV spectroscopy at 280 nm, using a theoretical molar extinction coefficient of 128,745 M^{-1} cm^{-1}, calculated from the sequence by ExPASy Proteomics. ESMS

showed two proteins with masses 48881.2 Da and 11613.2 Da for MBP-CTA2 and CTB respectively (calculated 48853.9 Da and 11611.9 Da).

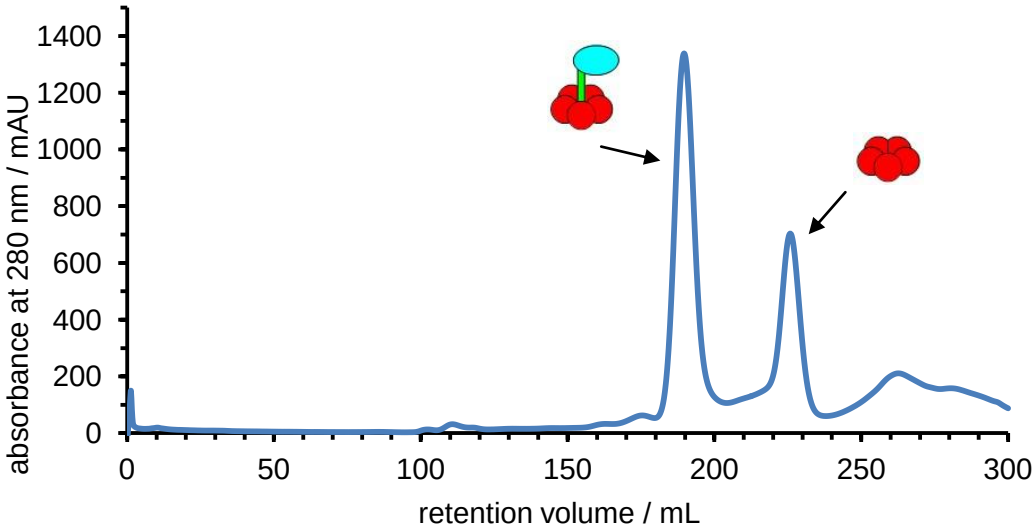


Fig 6.22 Chromatogram of MBP-AB_5 (G1S) purification.

6.2.2 AB_5 modifications

The MBP-CTA2 (G1S) fusion had to be cleaved to reveal the serine on the N-terminus of CTA2 before an oxime ligation was performed. This was achieved with TEV protease in the same way as previously done to reveal the sortase tag. The TEV protease is specific to the recognition sequence of ENLYFQG or ENLYFQS, working with the same efficiency for either sequence. Therefore it could still be used to cleave the MBP-CTA2 (G1S) fusion even after the mutation to introduce the serine residue.

The MBP-AB_5 (G1S) protein was mixed with TEV and the cleaved protein was purified by nickel affinity chromatography which removed the protease. SDS-PAGE was performed which shows MBP-AB_5 (G1S) before cleavage and after when the MBP has been removed (Fig 6.23). Analysis by ESMS showed the cleaved CTA2 (G1S) with a mass of 5085.6 Da (calculated 5085.5 Da) and CTB with a mass of 11612.0 Da (calculated 11611.9 Da).

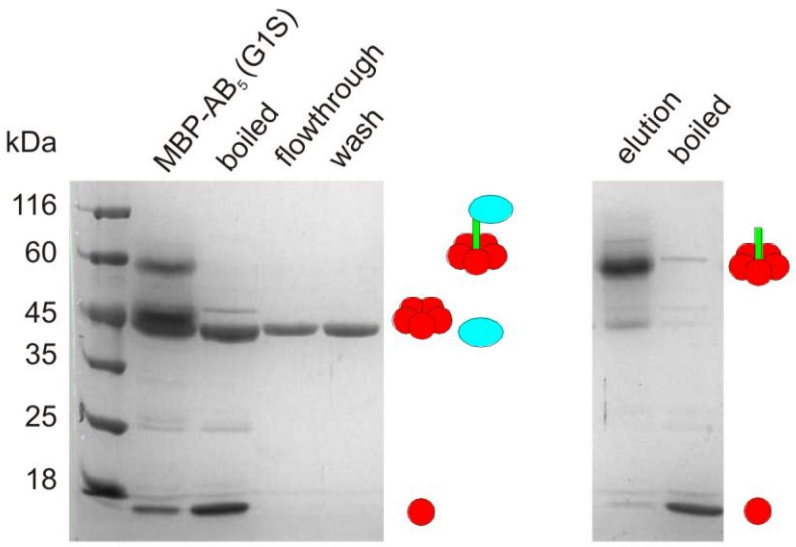
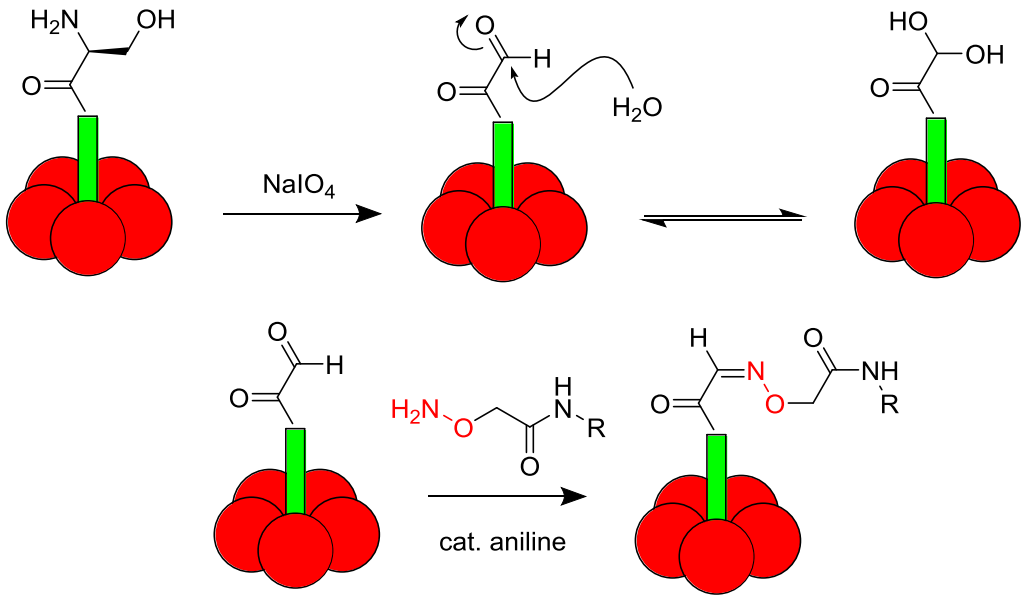


Fig 6.23 SDS-PAGE for the TEV cleavage of MBP-AB_5.(G1S). After cleavage and purification by Ni affinity chromatography no MBP remains.

The N-terminal serine residue revealed after TEV cleavage could now be oxidised using the same method as for CTB N-terminal oxidation using $NaIO_4$. After five minutes ESMS confirmed the site-specific oxidation of CTA2 (G1S) with a mass of 5076.4 Da for the hydrated form (calculated 5072.5 Da).



Scheme 6.4 Modification of the CTA2 N-terminus in the AB_5 complex by oxidation and then oxime formation with an aminooxy ligand at neutral pH using an aniline catalyst.

CTA2ox in the AB_5 complex was then modified with biotin ligand **61** in a reaction catalysed with aniline that went to completion overnight to produce ligand **A(61)B_5** (Scheme 6.4). ESMS confirmed that a successful reaction had taken place with modified CTA2 having a mass of 5470.7 Da (calc 5470.7 Da) (Fig 6.24).

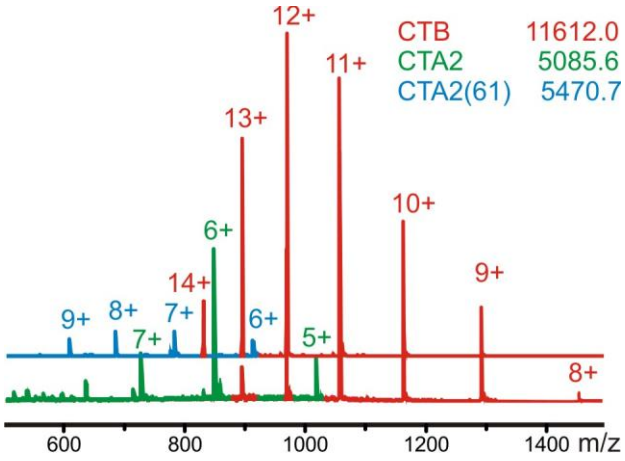


Fig 6.24 Mass spectrum of the AB_5 protein showing the two separate proteins CTB (red) and CTA2 (G1S) (green). Selective modification with the biotin ligand yields an unmodified CTB and **CTA2(61)** (blue).

6.2.3 Analysis of the interactions with streptavidin

The affinity of the biotin ligand for streptavidin was measured to assess the interaction and determine whether the PEG linker affected the affinity. ITC was performed with biotin **61** (30 µM) titrated into streptavidin (3 µM) and showed a very high affinity, with an apparent K_d of 1.72 nM (Table 6.1). This affinity however had a large error and the steepness of the curve in the ITC analysis suggested that the actual affinity was much higher (Fig 6.25 A). The c value for this isotherm is too high, because the transition is too steep and so an accurate K_d could not be determined. However, accurate values for the enthalpy and the stoichiometry are still available. A stoichiometry of 1.01 showed that biotin **61** was filling all the binding sites.

ITC was performed on the modified protein **A(61)B_5** to see the effects of having the biotin ligand attached to a protein and determine whether the high affinity interaction was still taking place. The **A(61)B_5** protein (30 µM) was titrated into streptavidin (5 µM). The apparent K_d for this interaction was found to be very similar to the interaction with biotin ligand **61**, at 0.70 nM (Table 6.1) which may have been because the limit of detection for the technique had been reached. Again the curve was steep and the error in the affinity was large, indicating a potentially stronger interaction (Fig 6.25 B). The enthalpy for the protein binding was much higher than for just the biotin ligand which may have been due to extra interactions between the proteins. An interesting point from this ITC result was the stoichiometry of 0.5,

which meant that only half the binding sites on streptavidin were being filled. Two **A(61)B$_5$** ligands were binding to the streptavidin tetramer and presumably the bulk of the attached protein blocked the interaction of a second **A(61)B$_5$** ligand binding to that side of streptavidin (Scheme 6.5).

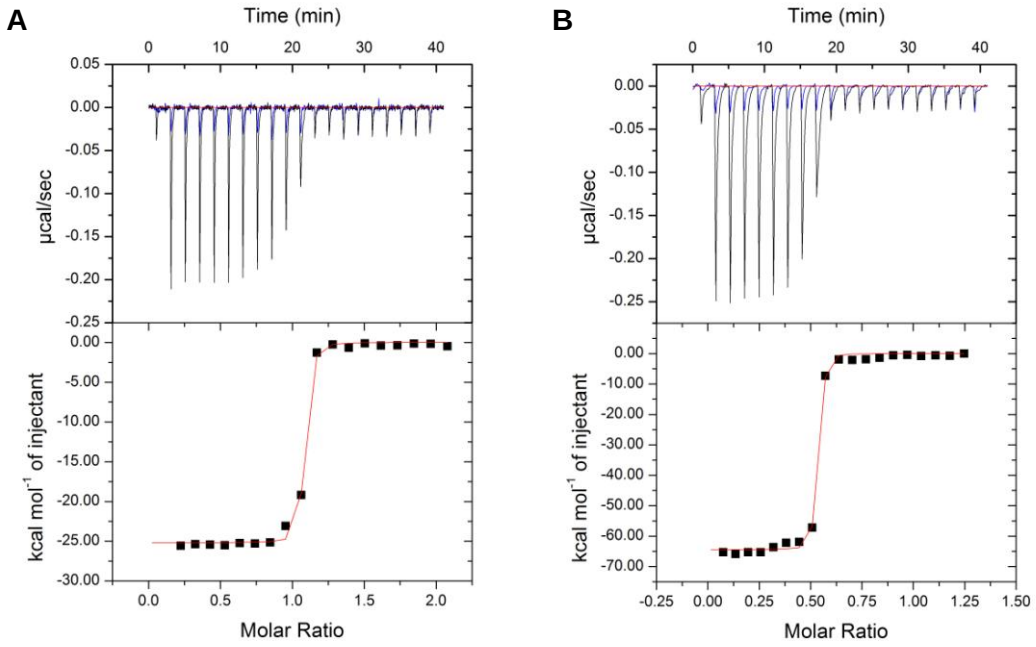
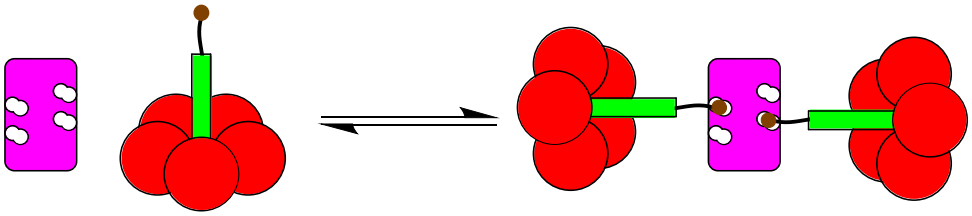


Fig 6.25 ITC trace for biotin binding to streptavidin. (A) Biotin ligand **61** (30 µM) injected into streptavidin (3 µM). (B) **A(61)B$_5$** (30 µM) injected into streptavidin (5 µM). Both titrations had a high affinity but titrating in the **A(61)B$_5$** protein resulted in an increased enthalpy and a stoichiometry of 0.5.

Table 6.1 ITC results for the interaction with streptavidin and biotin ligand **61** and **A(61)B$_5$**.

ligand	K_d / nM	ΔH° / kcal mol^{-1}	n
biotin **61**	1.72 ± 1.00	-25.74 ± 0.52	1.01
A(61)B$_5$	0.70 ± 0.20	-64.48 ± 0.15	0.51



Scheme 6.5 The proposed formation of a 2:1 complex between **A(61)B$_5$** and streptavidin

To determine the size of the protein assemblies being formed, SEC was performed. Streptavidin had a retention volume of 17.1 mL and the slightly larger **A(61)B$_5$** had

a retention volume of 16.1 mL. A mixture of the two proteins was prepared at a ratio of 2:1 with **A(61)B$_5$** at 20 µM and streptavidin tetramer at 10 µM. The SEC trace of the mixture showed a main peak with retention volume of 13.0 mL which could be attributed to the 2:1 protein complex (Fig 6.26). There was also a small peak at 15.7 mL indicating a small amount of excess **A(61)B$_5$** due to inaccuracies in the concentration measurement. Analysis of the chromatogram peak with the calibration plot for the column revealed that a model of a 3:1 complex was the best fit for the protein assembly (Fig 6.27). However, due to the nonspherical nature of the protein assembly, the model is likely to overestimate the actual size of the complex being formed and in reality it may behave as a much larger structure when analysed by SEC.

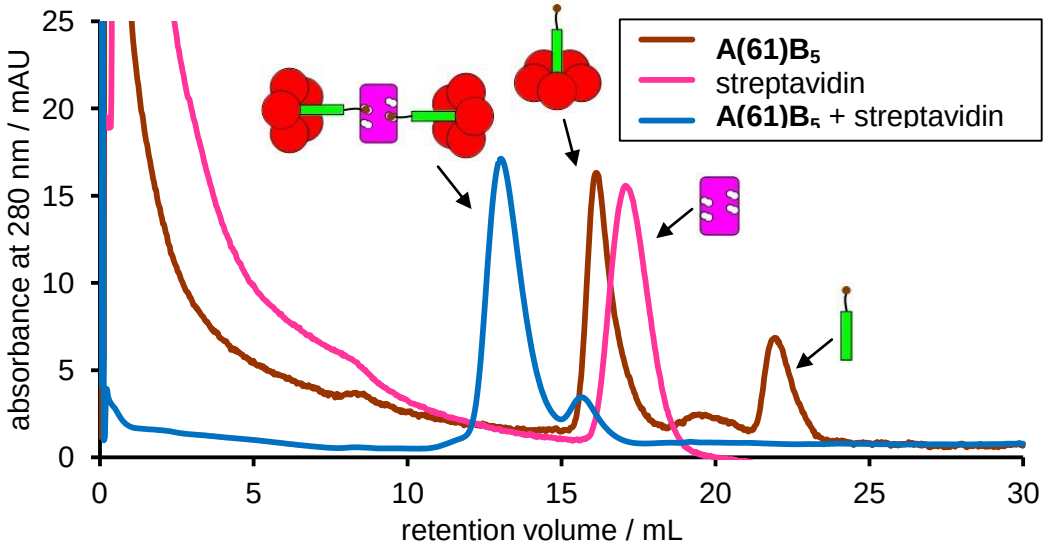


Fig 6.26 Chromatogram of the assembly of streptavidin and **A(61)B$_5$**.

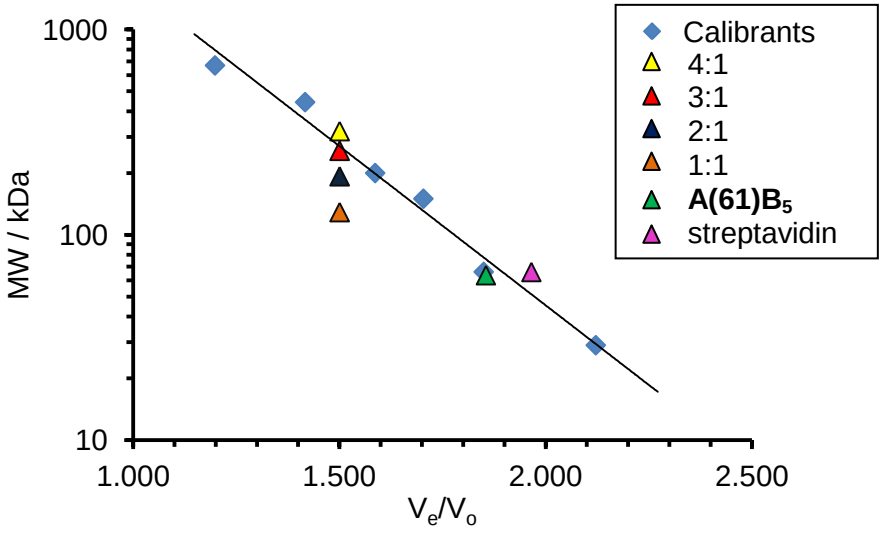


Fig 6.27 Analysis of the SEC calibration with the potential assembly products. The assembly comes at 1.5 V_e/V_o and the diferent possible assemblies of **A(61)B$_5$** and streptavidin are compared.

Another method was needed to confirm the size of the particles being formed and so AUC was performed. **A(61)B$_5$** was mixed with a fluorescently labelled streptavidin protein (Alexa Fluor 555 Streptavidin, Life Technologies) that absorbed at 555 nm. It was possible by using this extra wavelength to analyse only the assemblies involving streptavidin. A solution of pure streptavidin was seen to contain a single species with a sedimentation coefficient of 3.7 S corresponding to the expected mass of 65 kDa for the tetramer. The AUC was performed at different ratios of **A(61)B$_5$** ligand to streptavidin and showed that, as the two proteins were mixed, a larger species emerged (Fig 6.28). Increasing additions of **A(61)B$_5$** led to the increased formation of the new species with a sedimentation coefficient of 6.7 S corresponding to a mass of 150 kDa. This mass was between the expected mass of 130 kDa for a 1:1 complex and 193 kDa for a 2:1 complex. Again, the nonspherical shape of the complex may not have been described well by the fitting model leading to a less accurate estimate of the mass.

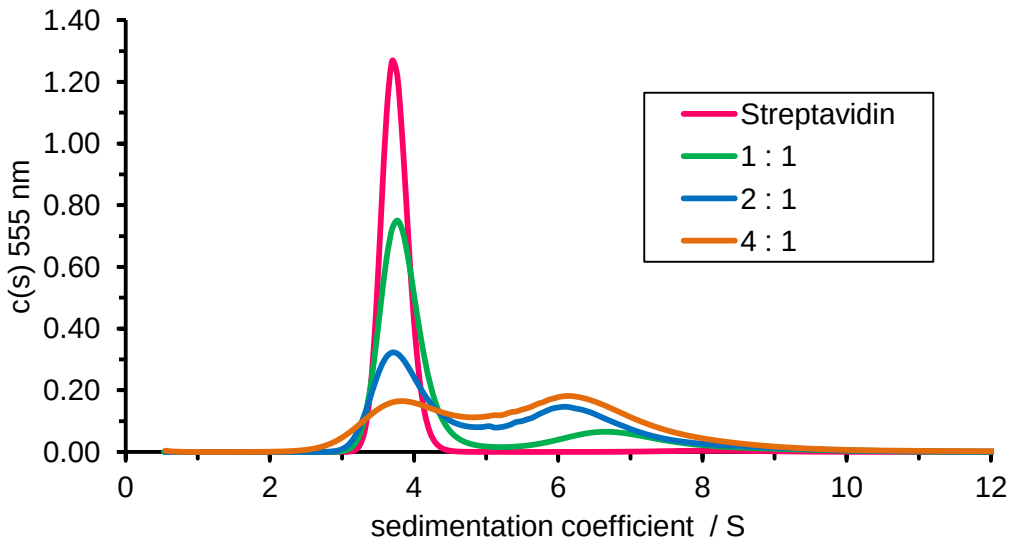


Fig 6.28 AUC c (s) plot of fluorescent streptavidin mixed with the biotinylated **A(61)B$_5$** protein measured at 555 nm and performed at 30 krpm. Only the assemblies with the fluorescent streptavidin are shown at this wavelength. The emergence of a second peak at 6.7 S was seen with increasing amounts of **A(61)B$_5$**.

Not all of the streptavidin was converted to the larger size particle even after the addition of 4 equivalents of **A(61)B$_5$**. This observation could suggest that there was a substantial error in the concentrations, with the **A(61)B$_5$** concentration being overestimated and so not enough **A(61)B$_5$** had been added. Another explanation would be the complex was not as stable as first thought. **A(61)B$_5$** may bind to streptavidin but then the CTA2 peptide could come loose from the CTB pentamers

leaving two different species, CTB and streptavidin with two CTA2-biotin peptides bound.

It can be concluded from a combination of the ITC, SEC and AUC data that a 2:1 complex with two AB_5 proteins facing out opposite sides of streptavidin is the most likely outcome for the species being formed. AFM could be used to visualise the assembly and so provide further evidence to support this hypothesis.

6.2.4 Conclusions and further work

Enzymatic modification of the CTA2 N-terminus was achieved by the sortase ligation of depsipeptides. A bpy group for metal binding was attached to CTA2 but no complexation with Fe^{2+} was observed. The CTA2 sequence was almost completely altered but still no binding was observed. It was thought that the CTA2 peptide was interacting with the bpy ligand at the point that CTA2 penetrates into CTB. If a different metal ligand could be used with a more stable binding interaction then this may overcome the ligand binding to the CTA2 peptide. A phenathroline ligand could be used which is known to bind Fe^{2+} in a 3:1 ratio and provides a stronger binding due to the restricted conformation of the molecule with the two nitrogen lone pairs fixed in position ready for metal binding (Fig 6.29).[176]

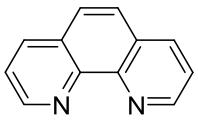


Fig 6.29 Structure of phenanthroline.

Strong protein-ligand interactions were demonstrated to be able to bring different proteins together. A biotin-modified AB_5 protein bound to streptavidin resulted in the formation of a 2:1 assembly which was a significant initial step towards the construction of larger complex structures.

The streptavidin-**A(61)B_5** system presented here would assemble two CTB units pointing in opposite directions. These CTB proteins have five GM1os binding sites each and so divalent GM1 ligands could potentially crosslink two separate CTB proteins producing a -A-B-A-C- polymeric system with three components (Fig 6.30). The **A(61)B_5** proteins could also bind to a GM1 coated surface and vertical constructions could then be made. There are many possibilities for using this basic

two protein system to build larger structures. The step wise introduction of other multivalent ligands and proteins could be used to make the system more complex.

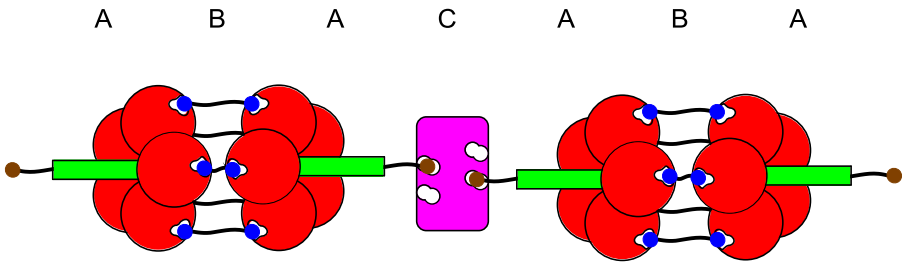
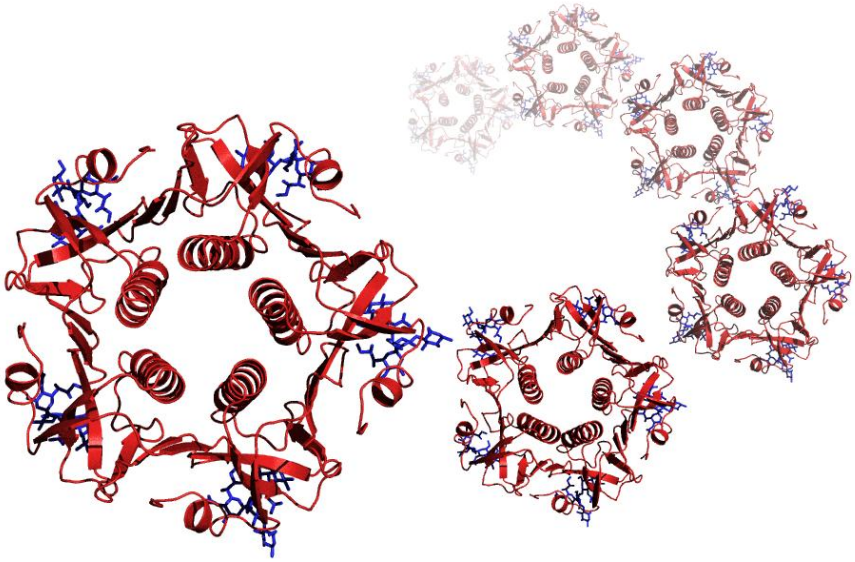


Fig 6.30 Potential polymeric assembly of “A” **A(61)B$_5$**, “B” divalent GM1 and “C” streptavidin.

7 Where do we go from here?



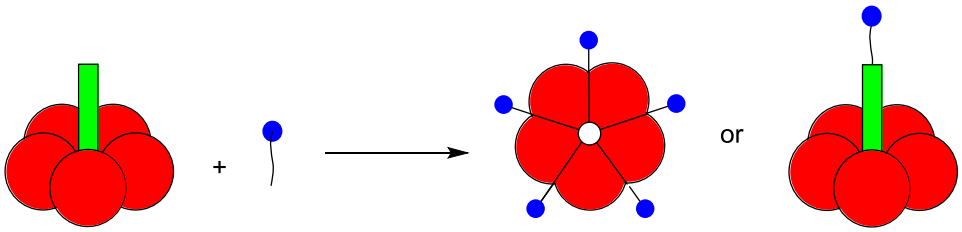
Conclusions and future work

7.1 Conclusions from the project

A range of protein assemblies have been created via protein-ligand interactions and these assemblies can be analysed by a variety of biophysical techniques. ESMS allowed the precise modification of proteins to be seen. ITC has been used to investigate the binding interactions of the ligands and gave information on the assembly stoichiometry. DLS, SEC and AUC were used to assess the size and the mass of the complexes formed.

7.1.1 Initial investigations

The site-specific modification of proteins was achieved at the N-terminus of both CTB and CTA2 (Scheme 7.1). Improvements on the original reaction conditions[146] to form an oxime bond were achieved by performing the reaction at pH 7 with an aniline catalyst[52].



Scheme 7.1 Modification of the N-temini of CTB or CTA2 with ligands.

Protein assembly was achieved via the attachment of a galactose ligand to the N-terminus of wild-type CTB. The assembly created with these modified proteins was uncontrolled and led to aggregation and precipitation of the proteins. The aggregation could be mostly prevented with the introduction of a simple galactoside as a competing ligand at 200 mM and by maintaining the protein at a low concentration of 200 µM . The length of the ligand was also found to be important; too short and there was not enough space for binding interactions to take place. The exact analysis of the large aggregates proved difficult as only the stable pentamer could be isolated from solution.

This initial strategy of creating a building block that has both multiple ligands and multiple binding sites was difficult to control. The correct concentration, the precise length of ligand and a suitable amount of competing ligand may all need to be varied systematically for the modified proteins to be controlled.

7.1.2 Two component systems

A number of mutations were introduced into the GM1os binding site of CTB which prevented binding with the GM1 oligosaccharide. W88E was chosen as the best mutant and so was modified with carbohydrate ligands. A simple galactoside ligand did not have a high enough affinity to bind to other proteins, but when a GM1 ligand was used, interactions with the wild-type CTB did occur. A combination of ITC, AUC and DLS led to the conclusion that heterodimers of pentamers were being formed and no larger structures (Fig 7.1).

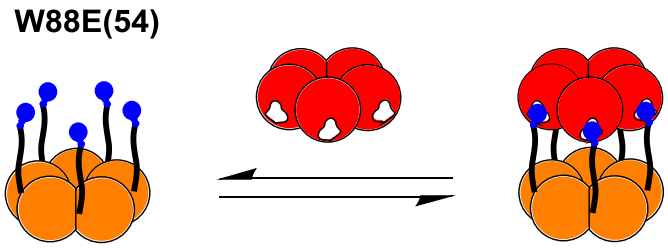


Fig 7.1 A 1:1 heterodimer of pentamers formed by **W88E(54)** binding to CTB.

Whilst this observation meant that a dodecahedron or other larger assemblies were not accessible with these components, it did provide a potential cholera toxin inhibitor. Inhibition studies showed an IC_{50} value for the GM1 modified mutant protein, **W88E(54)**, of 104 pM which was the best pentavalent CTB inhibitor reported thus far in the literature.

An investigation into the amino acid composition on the bottom face of the W88E protein may reveal positions that could be mutated to provide favourable interactions with the wild-type cholera toxin. These mutations may contribute to a more favourable protein-protein interaction when the two components come into close proximity. This effect could provide better overall binding and an enhanced affinity of the inhibitor.

If the length of the GM1 ligands was shortened this may affect the conformation that the ligand can take and therefore what interactions can take place. Shorter ligands may restrict each GM1os group to bind to a separate CTB protein, assembling five around the central modified mutant. A larger range of assembly structures could then be accessed with this system.

7.1.3 Templated assemblies

TX-100 and tyloxapol micelles modified with simple galactose and lactose groups were analysed. The introduction of the carbohydrates led to unpredictability in the micelle sizes. The amount of carbohydrate groups incorporated was therefore limited to around 10%. Mixtures of the micelles with CTB were analysed with ITC and DLS but no interactions were observed. The limitation on carbohydrate inclusion and the strength of the galactose interactions were likely to be the deciding factors as to why no interactions were seen between the micelles and CTB.

If detergents were modified with GM1os ligands, more success may be seen with binding CTB. This larger oligosaccharide may prove more of a problem when forming stable micelle structures but the use of tyloxapol may alleviate this problem. GM1os ligands have a much higher affinity for CTB and lower concentrations could be accessed at which the galactose based system was seen to not work. With these ligands CTB may be prearranged around a micelle surface before cross-linking of the proteins.

The AB_5 structure of cholera toxin was used to tie the proteins together by their CTA2 peptide tails. Enzymatic modification of the CTA2 N-terminus with depsipeptides was achieved with sortase. A bpy group for metal binding was attached to the CTA2 peptide but was not seen to bind to Fe^{2+}. Even after the CTA2 sequence was almost completely altered, no binding was observed. It was thought that the CTA2 peptide was interacting with the bpy ligand at the point that CTA2 penetrates into CTB. This part of the peptide sequence however, could not be changed without affecting the AB_5 assembly.

The use of a different metal ligand with a more stable binding interaction may overcome the problems of binding to the CTA2 peptide. A phenathroline ligand could be used which binds Fe^{2+} in a 3:1 ratio but with a stronger interaction due to the restricted conformation of the molecule.[176] The two nitrogens are fixed in this molecule, in a conformation more optimised for metal binding (Fig 6.29).

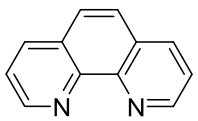


Fig 7.2 Structure of phenanthroline, with the nitrogen atoms constrained by the ring structure.

A mutation was made that allowed the CTA2 peptide to be also modified by oxidation and then oxime formation. A biotin ligand was attached to CTA2 and binding of the modified AB_5 protein was seen with streptavidin. A combination of ITC, SEC and AUC confirmed formation of a 2:1 assembly. Not all the streptavidin binding sites were filled due to the steric bulk of the AB_5 protein preventing interactions with both biotin binding sites on either side of the streptavidin protein. The binding of two AB_5 proteins resulted in two of the binding sites becoming inaccessible for other ligands. Less bulky biotin ligands could be used to fill these extra sites and build upon this structure. If a long flexible linker was used, then the excluded binding sites could be accessed, before attachment to another protein.

There are many possibilities of using this basic protein system to build larger structures. The **A(61)B_5** proteins could bind to a GM1 coated surface. Vertical constructions could then be made with streptavidin binding before a second **A(61)B_5** unit or a different biotinylated protein building the structure upwards (Fig 7.3). AFM could be used to visualise the step by step assembly.

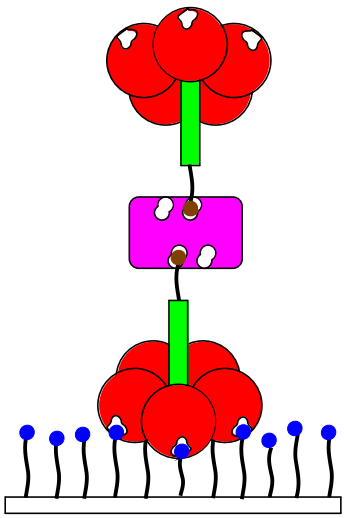


Fig 7.3 Surface bound construction of **A(61)B_5** and streptavidin.

7.1.4 Further assembly

The step wise introduction of other multivalent ligands and proteins could also be used to introduce another level of complexity into the system. The streptavidin **A(61)B_5** system presents the CTB proteins with five GM1os binding sites each pointing in opposite directions. If divalent GM1 ligands were synthesised then they

could potentially bind two separate CTB proteins producing a -A-B-A-C- polymeric system with three components (Fig 7.4 A).

The **W88E(54)** protein could also be used in combination with **A(61)B_5** and streptavidin. Mixing of the three components would result in a closed system (Fig 7.4 B). Again, with the introduction of divalent GM1 ligands at the right ratio, an oligomeric protein structure could be formed with the **W88E(54)** protein capping the construction.

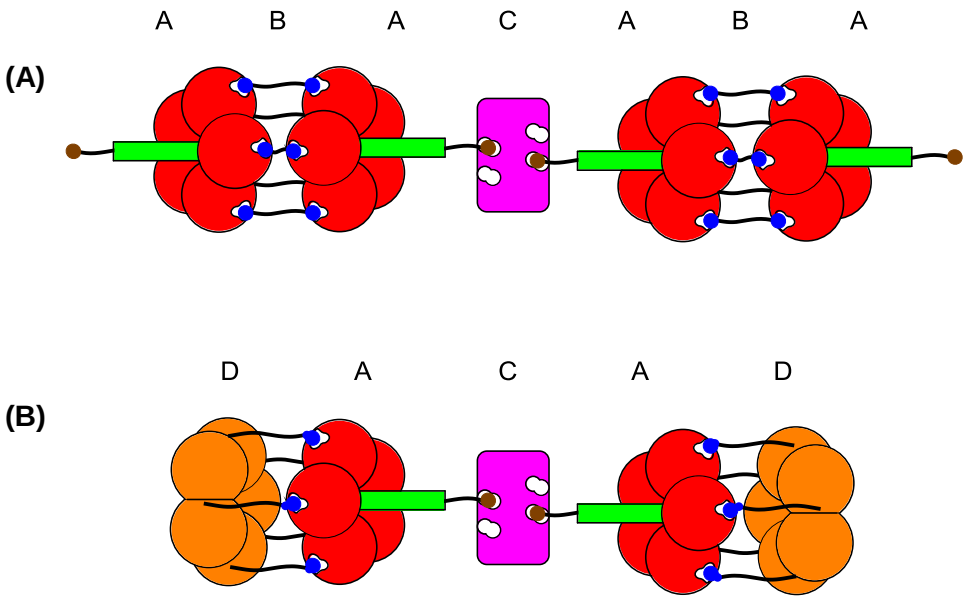
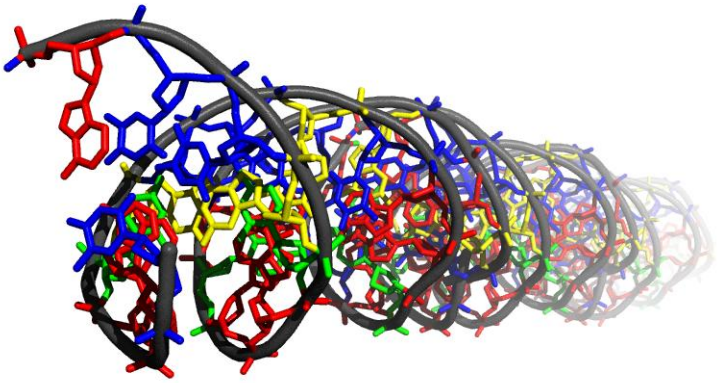


Fig 7.4 Potential (A) polymeric and (B) oligomeric assemblies of "A" **A(61)B_5**, "B" divalent GM1os, "C" streptavidin and "D" **W88E(54)**.

There are other proteins with different shapes and ligand binding abilities that could be used to create a wider variety of structures. The trimeric mannose binding lectin (MBL), with three mannose binding sites, could be combined with the pentameric CTB in a new strategy to form VLPs. The tetrameric concanavalin A (ConA) has four mannose binding sites in a tetrahedral geometry and so could provide access to whole new range of complex protein structures.

The creation of protein structures was shown with this work that expands the overall understanding of protein-ligand interactions. General strategies for the assembly of biomolecules have been investigated and can be further developed to advance the construction of nanoarchitecture.

8 How it all came about



Materials, methods and experimental

8.1 Molecular Biology and Protein Expression

8.1.1 Instrumentation and materials

Media and glassware were sterilised in a Prestige Medical bench top autoclave or in a LTE Touchclave-R autoclave. Bacterial cultures were grown using a Kuhner ShakerX shaker incubator. Cultures on agar plates were grown in a Gallenkamp incubator. Centrifugation was performed using a Beckman Coulter™ Avanti™ J-30I Centrifuge or a Heraeus multifuge 3 S-R centrifuge or a Heraeus pico centrifuge. Protein purification was performed with a GE Pharmacia ÄKTA FPLC system. Spectrophotometric readings were taken in a WPA Biowave II spectrophotometer or a NanoDrop 2000 spectrophotometer. SDS-PAGE was performed using a Bio-Rad mini protean 3 apparatus and visualised on a Gel Doc system.

Analytical grade reagents were supplied by Sigma-Aldrich, Fisher Scientific, Melford laboratories and VWR International. Pwo enzyme (DNA polymerase) (from *Pyrococcus woesei*) was made in the Turnbull lab. Restriction enzymes, T4 DNA ligase, DpnI and CIP were purchased from NEB. Competent cells were purchased from Stratagene or made in the Turnbull Lab. Primers and DNA parts for gene synthesis were purchased from IDT.

DNA primer, gene synthesis parts and full plasmid sequences are given in the appendix. Protein sequences for CTB, MBP-CTA2 and mutants thereof are given in the appendix.

8.1.2 Buffer solutions

All solutions were made up to 1 L with 18 MΩ water purified using an Elga purelab system unless otherwise stated. Buffer solutions were adjusted to the correct pH using NaOH or HCl.

Standard buffers

Phosphate buffer (pH 7.0, 0.1 M): $Na_2HPO_4.2H_2O$ (10.27 g), $NaH_2PO_4.2H_2O$ (6.60 g), 0.1 M NaCl (5.84 g).

*Phosphate Buffered Saline (*PBS (pH 7.4)): 137 mM NaCl (8.0 g), 2.7 mM KCl (0.2 g), 10 mM $Na_2HPO_4.2H_2O$ (1.78 g), 1.76 mM KH_2PO_4 (0.24 g).

Sodium Acetate Buffer (pH 4.0, 0.1 M): NaOAc (1.26 g), acetic acid (4.84 mL).

HEPES buffer (pH 7.5): 50 mM HEPES (14.1 g), 150 mM NaCl (8.8 g), 5 mM $CaCl_2$ (0.55 g).

TEAN Buffer (pH 7.0, 50 mM): 50 mM Tris (6.06 g), 200 mM NaCl (11.69 g), 3 mM NaN_3 (0.195 g), 1 mM EDTA (0.372 g).

TRIS Buffer (pH 7.4, 50 mM): 50 mM Tris (6.06 g), 200 mM NaCl (11.69 g).

Citrate-phosphate buffer (pH 5.0, 2 mM): 2 mM $Na_2HPO_4.2H_2O$ (0.36 g), 2 mM $NaC_6H_7O_7$ (0.43 g), 2 mM NaCl (0.12 g).

Bacterial growth media

LB growth media (LB): LB Broth (25 g premixed; 10 g Tryptone, 5 g Yeast extract, 10 g NaCl), Ampicillin (100 mg). For high salt media; NaCl (10 g) was added.

Agar Gels: agar (1.5 g) added to 100 mL of LB.

Protein analysis solutions

5 x SDS Running Buffer: 125 mM Tris (15.15 g), 960 mM glycine (72.1 g), 0.5% w/v SDS (5 g).

SDS Loading Buffer: Tris (0.5M, pH 6.8, 1.25 mL), SDS (0.3 g), DTT (0.47 g), glycerol (1 mL), bromophenol blue (~5 crystals), water (7.2 mL).

Coomassie stain: coomassie G250, methanol (400 mL), acetic acid (100 mL).

Coomassie destain: methanol (400 mL), acetic acid (100 mL)

DNA manipulation and analysis buffers

50 x TAE Buffer: 2 M Tris (242 g), acetic acid (57.1 mL), EDTA (0.5 M, pH 8, 100 mL).

10 x PFU Buffer: 2 M Tris-HCl (pH 8.8), 1 M $(NH_4)_2SO_4$, 1 M KCl, 10 mg mL^{-1} BSA, 10% (v/v) TX-100, 200 mM $MgSO_4$.

Ligase Buffer: 50mM Tris-HCl (pH 7.5), 10mM $MgCl_2$, 1mM ATP, 10mM DTT.

8.1.3 Gene synthesis and mutagenesis

Polymerase Chain Reaction (PCR)

PCR was used for site-directed mutagenesis. The materials used are summarised in Table 8.1. The mixture of plasmid (~100 µg mL^{-1}), primers (100 µM), dNTPs (25 mM) and PWO enzyme were taken through 22 cycles of 95 °C for 30 s, 54 °C for 30 s (or temperature adjusted to 3 °C below expected annealing temperature for primers used), 72 °C for 14 min before a final extension of 30 min at 72 °C. Following PCR, DpnI (1 µL) was added and incubated at 37 °C for 2 h to digest the original plasmid leaving only the new plasmid with the incorporated mutation.

Table 8.1 Summary of materials used for PCR.

Substance	Amount / µL
H_2O	42.9
PFU buffer 10X	5.0
dNTPs (25 mM)	0.4
PWO enzyme	0.5
Plasmid	0.2
Two primers	0.5

Polymer Cycling Assembly (PCA)

PCA was used to assemble a gene from four smaller DNA parts. The materials used are summarised in Table 8.2. The mixture of DNA parts (100 µM), dNTPs (25 mM) and PWO enzyme were taken through 30 cycles of 95 °C for 30 s, 54 °C for 30 s (3 °C below expected annealing temperature of DNA parts), 70 °C for 1 min before a final extension of 3 min at 70 °C. The whole cycle was repeated with the addition of only parts A and D (0.5 µL) to ensure amplification of the complete gene.

Table 8.2 Summary of materials used for PCA.

Substance	Amount / µL
H_2O	41.4
PFU buffer 10X	5.0
dNTPs (25 mM)	0.4
PWO enzyme	0.2
Parts B and C	1.0
Parts A and D	0.5

Double digests

Analytical double digests were performed to check the inclusion of restriction sites. The materials used are summarised in Table 8.3. The reagents were incubated together for one hour at 37°C.

Table 8.3 Summary of materials used for analytical double digests.

substance	volume / µL
H_2O	7.0
NEB 4 buffer 10X	1.0
plasmid	1.0
NdeI	0.5
XhoI	0.5

Full double digests were performed to prepare DNA for ligation. For full double digests Table 8.4 gives the materials used. All reactants were added, except for the calf-intestinal phosphatase (CIP), before incubation at 37 °C for one hour. CIP was then added to the digested plasmid and incubated for a further 1 h at 37 °C. Digested DNA was then purified by gel extraction following the procedure from the Qiagen gel extraction kit.

Table 8.4 Summary of materials used in double digests.

Substance	Amount / µL
H_2O	5.0
NEB 4 buffer 10X	3.0
Plasmid or gene	20.0
NdeI	1.0
XhoI	1.0
CIP	1.0

DNA ligation

After a double digest of the plasmid and gene of interest, the two DNA fragments were ligated together to form one new plasmid. Plasmid (50 ng), 10X ligase buffer (1 µL), T4 DNA ligase (1 µL), H_2O (up to 10 µL) and the gene of interest (3 equivalents to plasmid) were mixed together and incubated at 4 °C overnight or incubated for 30 min at room temperature. The new plasmid was then transformed into *E. coli XL10* competent cells.

Transformations

Plasmids were transformed into suitable competent cells by adding DNA (1 µl, ~100 µg mL^{-1}) to the competent cells (10 µl), and incubating on ice for 10 min. The cells were heat shocked for 45 s at 42 °C, and then put back on ice for 10 min. LB (800 µl) was added and the cell suspension was incubated at 37 °C for 1.5 h. An aliquot (100 µl) was spread on to an agar plate and incubated at 37 °C overnight.

A single colony was picked to inoculate LB (5 ml) with ampicillin (100 mg L^{-1}) antibiotic present and the culture was incubated overnight at 37 °C. An aliquot (0.5 ml) of culture was mixed with 80% glycerol (0.5 ml) and frozen at -80 °C in the freezer as a cell stock.

Plasmid preparations

The equipment and procedure for the Qiagen Miniprep kit was followed to extract DNA from a cell miniculture. UV spectroscopy gave the concentration and yield of the DNA (Equation 8.1).

Equation 8.1

$$concentration = (A_{260} - A_{320}) \times 50 \mu g mL^{-1}$$
$$yield = concentration \times volume$$

where A_{260} and A_{280} were the absorbances at 260 nm and 280 nm respectively.

Agarose gels

Agarose gels were used to assess DNA purity and confirm the size of DNA parts. The gels (1.2% w/v) were prepared by adding agarose (380 mg) to TAE buffer (40 mL) and heating in a microwave for 1.5 min. After heating, the solution was allowed to cool before adding ethidium bromide (1 µL, 10 mg mL^{-1}). The solution was poured into a mould and a comb added, removing any bubbles before the gel was left to set. Each DNA sample (5 µL) was mixed with loading buffer (1 µL) and then an aliquot (5 µL) was loaded on to the gel. Electrophoresis was performed at 100 V for approximately 20 min in TAE buffer. Gels were then visualised under UV light using the Gel Doc system.

8.1.4 Protein expression and purification

CTB expression from *Vibrio sp.60*

A stab from a stock of *Vibrio sp.60* containing the pATA13 plasmid[134] encoding the CTB protein was used to inoculate a flask of high salt LB growth media (2 x 100 mL) containing ampicillin (100 µg mL^{-1}) and this starter culture was incubated at 30 °C for 20 h. Starter culture (20 mL) was added to 10 x 1 L high salt LB growth media so the optical density at 600 nm (OD_{600}) was 0.05. This was then incubated at 30 °C and after 3 h the OD_{600} had reached ~0.5 and IPTG (240 mg) was added to each flask to induce protein over-expression. The incubation was continued for 24 h before isolating the cells by centrifugation at 17,000 x g for 25 min. The bacterial pellet was discarded and the supernatant was retained and purified by one of the two following methods.

Original precipitation method:[135] sodium hexametaphosphate (2.5 g L^{-1}) was added to the supernatant and the pH was adjusted to 3.5 by addition of HCl. The solution was then stored at 4 °C for 3 days to allow the protein to precipitate. The suspension was centrifuged at 16,000 x g for 25 min, before discarding the supernatant and the protein pellet was resuspended in *TRIS buffer* (20 mL). This solution was centrifuged at 5,000 x g for 10 min to remove any insoluble material, the pellet was again resuspended in *TRIS buffer* and centrifuged at 100,000 x g for 25 min. The supernatant was collected and filtered through a 0.45 µm filter (Sartorius Minisart) before the protein solution was purified on a lactose affinity column (resin from sigma) (~20/200) and eluted with *TEAN buffer* containing 300 mM lactose. The protein solution was dialysed into *phosphate buffer* using SnakeSkin® pleated dialysis tubing (Thermo Scientific) with 7000 MWCO. The buffer solution was changed three times to ensure effective dialysis.

After optimization: solid ammonium sulfate was added to the supernatant (final concentration 60% w/v) and stirred for 1 h. The solution was centrifuged at 17,000 x g for 25 min and the supernatant discarded. The protein pellet was resuspended in *phosphate buffer* and centrifuged once more at 5,000 x g for 10 min to remove any insoluble material. The supernatant was filtered through a 0.45 µm filter (Sartorius Minisart), purified on a lactose affinity column (resin from sigma) (~20/200) and eluted with *phosphate buffer* containing 300 mM lactose.

The protein solution was dialysed into *phosphate buffer* using SnakeSkin® pleated dialysis tubing (Thermo Scientific) with 7000 MWCO. The purity of the protein was

determined by SDS-PAGE. The concentration of CTB protein was measured by UV spectroscopy at 280 nm.

CTB mutant expression from *E. coli*

A stab from a stock of *E. coli C41* containing the pSAB2.2 plasmid or derivative (see appendix for plasmid and protein sequences) encoding the CTB protein was used to inoculate LB growth media (5 mL) containing ampicillin (100 µg mL^{-1}) and this starter culture was incubated at 37 °C for 20 h. Starter culture (1 mL) was added to LB growth media (4 x 1 L) which was then incubated at 37 °C and monitored until the OD_{600} had reached ~0.5 before IPTG (240 mg) was added to each flask to induce protein over-expression. The incubation was continued for 24 h at 30 °C before isolating the cells by centrifugation at 10,000 x g for 10 min. The bacterial pellet was discarded and the supernatant was retained.

Solid ammonium sulfate was added to the supernatant (final concentration 60% w/v) and stirred for 1 h. The solution was centrifuged at 17,000 x g for 25 min and the supernatant discarded. The protein pellet was resuspended in *phosphate buffer* and centrifuged once more at 5,000 x g for 10 min to remove any insoluble material. The supernatant was filtered through a 0.45 µm filter (Sartorius Minisart), purified on a nickel affinity column, washed with *phosphate buffer* containing 20 mM imidazole and eluted with *phosphate buffer* containing 500 mM imidazole.

The purity of the isolated protein was determined by SDS-PAGE and the concentration of CTB mutant protein was measured by UV spectroscopy at 280 nm.

MBP-AB_5 expression from *E. coli*

A stab from a stock of *E. coli BL21 gold* containing the pSAB2.1 plasmid or derivative (see appendix for plasmid and protein sequences) encoding a MBP-CTA2 fusion and a CTB protein was used to inoculate LB growth media (5 mL) containing ampicillin (100 µg mL^{-1}) and this starter culture was incubated at 37 °C for 20 h. Starter culture (1 mL) was added to LB growth media (4 x 1 L) which was then incubated at 37 °C, monitored until the OD_{600} had reached ~0.6 before IPTG (120 mg) was added to each flask to induce protein over-expression. The incubation was continued for 24 h at 30 °C before isolating the cells by centrifugation at 10,000 x g for 10 min. The bacterial pellet was discarded and the supernatant was retained.

Solid ammonium sulfate was added to the supernatant (final concentration 60% w/v) and stirred for 1 h. The solution was centrifuged at 10,000 x g for 25 min and the supernatant discarded. The protein pellet was resuspended in *HEPES buffer* (or *phosphate buffer* for the G1S mutant) and centrifuged once more at 5,000 x g for 10 min to remove any insoluble material. The supernatant was filtered through a 0.45 µm filter (Sartorius Minisart), purified on a nickel affinity column, washed with *HEPES buffer* (or *phosphate buffer* for the G1S mutant) containing 20 mM imidazole and eluted with *HEPES buffer* (or *phosphate buffer* for the G1S mutant) containing 500 mM imidazole. The eluted protein was concentrated and further purified by SEC.

The purity of the isolated protein was determined by SDS-PAGE and the concentration of MBP-AB_5 was measured by UV spectroscopy at 280 nm.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).
Protein purity and size was assessed by SDS-PAGE using BioRad tetragel apparatus. A resolving gel (10%) was prepared using the materials listed in Table 8.5. The TEMED solution was added to the mixture after all the other solutions had been mixed and immediately before pouring the gel into the plates. 0.1% w/v SDS was poured on top of the resolving gel to ensure a flat surface.

The stacking gel (5%) was then prepared using the materials in Table 8.5, also with the TEMED being left until last. The 0.1% w/v SDS was removed from the top of the set resolving gel, the stacking gel mixture poured on and the comb inserted. Samples for analysis were prepared by mixing 10 µL of protein sample mixed with 10 µL loading buffer. After the gel had set, protein samples were loaded into the wells. When required, samples were heated to 95 °C prior to loading on the gel; these samples are described as “boiled” on the gel figures. Electrophoresis was performed at 180 V in SDS running buffer for approximately 45 min or until after sufficient time that the blue protein loading buffer had reached the bottom of the gel. Gels were stained with Coomassie stain overnight before incubation in Coomassie destain for approximately 3 hours and storage in water until the gel was imaged. Or the samples were stained with Instant Blue Stain (TripleRed) for 30 min before being imaged.

Table 8.5 Summary of the materials used for preparing SDS-PAGE gels.

Material	10% Resolving Gel (mL)	5% Stacking Gel (mL)
H_2O	5.885	3.112
1.5 M TRIS-HCl pH 8.8	3.800	-
0.5 M TRIS-HCl pH 6.8	-	0.945
10% w/v SDS	0.150	0.050
40% Acrylamide	5.000	0.833
15% w/v APS	0.150	0.050
TEMED	0.015	0.010

Protein concentration analysis

Protein concentrations were measured using UV absorption at 280 nm. The concentration was determined using the Beer-Lambert law (Equation 8.2).

Equation 8.2 $$A = \varepsilon c l$$

Where *A* is the absorbance, ε is the extinction coefficient in $mol^{-1}\ dm^3\ cm^{-1}$, *c* is the concentration in $mol\ dm^{-3}$ and *l* is the pathlength in cm.

Theoretical extinction coefficients for the CTB, CTA2 and MBP-CTA2 proteins were obtained from ExPASy ProtParam (Table 8.6).[177]

Table 8.6 Extinction coefficients for the proteins.

Protein	extinction coefficient / $mol^{-1}\ dm^3\ cm^{-1}$
CTB (wild-type)	11585
CTB W88K	6085
CTB W88K Q61E	6085
CTB W88E	6085
CTB I47T	11585
CTA2 (wild-type)	2980
CTA2 (D4N)	1490
CTA2 (G1S)	2980
MBP-CTA2 (wild-type)	70820
MBP-CTA2 (D4N)	69330
MBP-CTA2 (G1S)	70820

8.1.5 Protein characterisation

8.1.5.1 Mass spectrometry (MS)

Using a Bruker HCT ultra machine, samples were first diluted to 50-10 µM with H_2O before analysis. Samples for analysis using a Maxis Impact electrospray were then also mixed 1:1 with MeOH containing 1% formic acid before analysis. Processing and deconvolution of multiple charge states was performed using Data Analysis software.

8.1.5.2 Size exclusion chromatography (SEC)

Proteins samples for purification or analysis (≤ 250 µL) were injected into the injection loop of the ÄKTA FPLC system (GE healthcare) before purification on a superdex™ 200 10/300 GL column. The flow rate was set at 0.4 mL min^{-1}, fractions (400 µL) were collected and their UV absorbance measured at 280 nm (or 330 nm for dansyl group measurement or 555 nm for fluorescent streptavidin measurements). Samples for purification after expression (≤ 2.5 mL) were injected into the injection loop before purification on a superdex™ 200 HiLoad™ 26/60 column. The flow rate was set at 1.0 mL min^{-1}, fractions were collected (5 mL) and their UV absorbance at 280 nm was measured.

8.1.5.3 Differential scanning fluorimetry (DSF)

Protein samples (250 µM final concentration in *phosphate buffer*) were mixed with Sypro Orange dye (10 x concentrate solution, Bio-Rad) (25 µL total volume) and added to the wells of a 96-well polystyrene plate. The plates were heated in a Real Time PCR Detection System (Bio-Rad) between 25-100 °C in increments of 1 °C. Fluorescence changes in the wells were measured by reading the excitation at 488 nm.

The change in fluorescence per degree is plotted against the temperature, resulting in a curve with a peak that corresponds to the midpoint of the unfolding transition, the melting temperature (T_m).

8.2 Chemical synthesis and protein modification

Compounds prepared by T. R. Branson are described in this section. Other compounds provided by colleagues in the Turnbull lab and external collaborators are described in the appendix.

8.2.1 General methods

Commercial reagents were used without purification, unless otherwise stated. Unless stated otherwise in the experimental section, analytical grade reagents were supplied by Sigma-Aldrich, Fisher Scientific, Melford laboratories and VWR International. All concentrations were performed *in vacuo*, unless otherwise stated. All reactions were carried out at room temperature unless stated otherwise.

^{1}H and ^{13}C NMR spectra were recorded either on a Bruker Avance 500 instrument (at 500 MHz and 125 MHz respectively) or on a Bruker DPX300 instrument (at 300 MHz and 75 MHz respectively) at 25 °C. ^{1}H NMR and ^{13}C NMR spectra were referenced using their residual solvent signals,[178] tetramethylsilane, and ethanol as internal standards. Chemical shifts are given in parts per million downfield from tetramethylsilane. Signals were assigned using a combination of COSY, HMQC and DEPT135 experiments. The following abbreviations were used to explain the signal multiplicities or characteristics: s, singlet; d, doublet; dd, double doublet; t, triplet; m, multiplet.

Electrospray ionisation mass spectra were obtained on a Bruker HCT Ultra Ion Trap Mass Spectrometer. High resolution electrospray ionisation mass spectra were obtained on Bruker Daltonics MicroTOF or Maxis Impact electrospray mass spectrometers or a WATERS GCT Premier Mass Spectrometer.

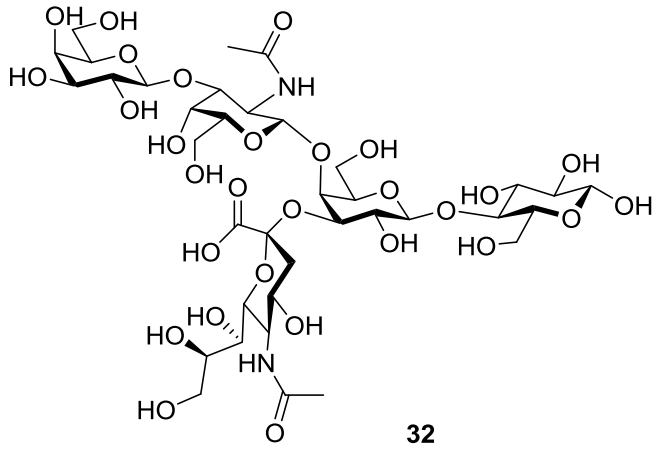
Flash chromatography was performed with silica gel 60 (Merck).

Infra-red spectra were recorded on a Perkin-Elmer Spectrum One FT-IR spectrometer.

Optical rotations were measured at the sodium D-line with an Optical Activity AA-100 polarimeter. $[\alpha]_D$ values are given in units of 10^{-1} deg cm^2 g^{-1}.

8.2.2 Ligand synthesis

β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-[5-acetamido-3,5-dideoxy-α-D-*glycero*-D-*galacto*-nonulopyranulosonyl-(2→3)]-β-D-galactopyranosyl-(1→4)-β-D-glucopyranose (GM1 oligosaccharide; GM1os) 32 [141]



32

The ceramide chain from GM1 (Avanti polar lipids Inc.) was enzymatically cleaved to give the GM1 oligosaccharide **32** (GM1os). Endo-glycosyl ceramidase (Takara Bio Inc) (EGC; 0.6 µL, 0.1 U) was added to a solution of GM1 (4.8 mg), NaOAc (10 mM, pH 5, 10 mL) containing bovine serum albumin (0.1 mg) and triton X-100 (2 mg). This mixture was incubated at 37 °C and after 8 days more EGC (24.4 µL, 4 U) was added. The cleavage was monitored by TLC (CH_3Cl : MeOH : H_2O solvent system, 30 : 40 : 4) and after nine weeks the mixture was extracted twice with ether (500 µL) to remove the triton X-100 and the aqueous layer was applied to a C18 column (Waters Sep-Pak® Vac 6cc). The column was washed with water (5 x 2.5 mL) and then methanol (3 x 2.5 mL). The first water fraction contained GM1os and so was passed over an Amberlite IRC-50 resin and the flow-through was lyophilised. Analysis by LC-MS showed GM1os to be pure and the mass measured by ESI-MS (negative ion mode) for GM1os was 997.6 Da (calculated $[M-H]^-$ 997.3 Da) (Fig 8.1). From the ^1H NMR spectrum it was seen that GM1os was an equal mixture of α and β anomers. The NMR spectrum was in accord with previous data.[141]

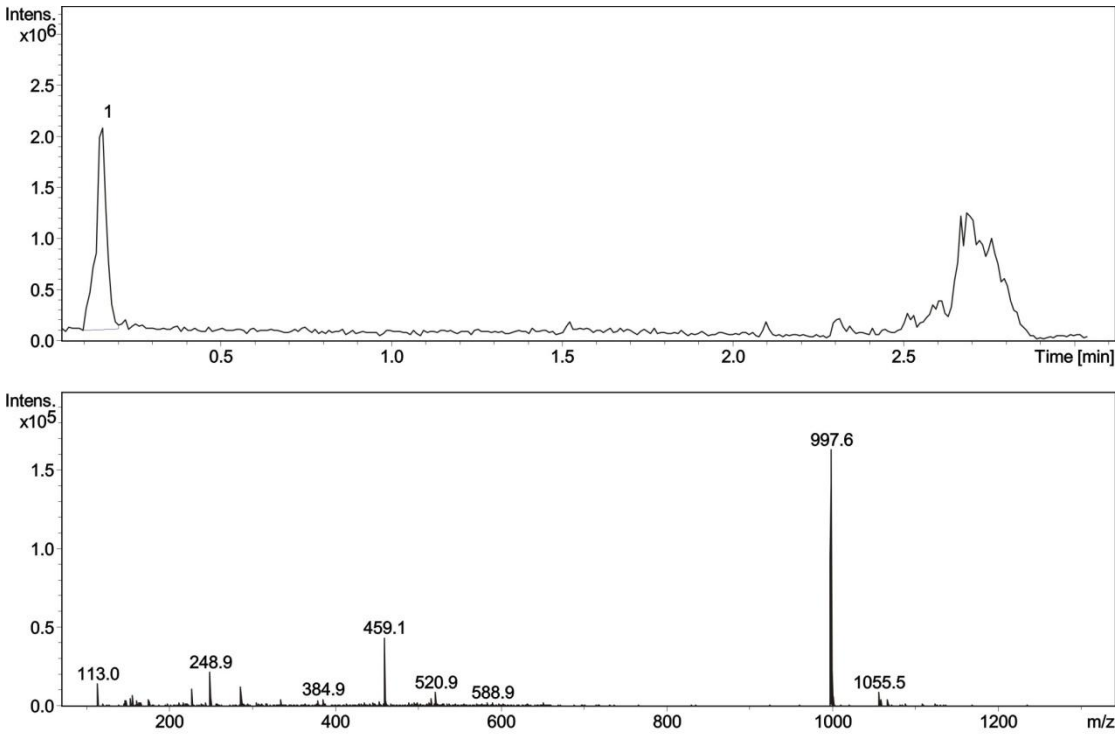
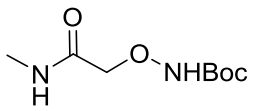


Fig 8.1 Mass spectrum showing the purified GM1os **32** (peak 1). The mass measured by LC-MS for peak 1 was 997.6 Da (calculated 997.3).

NMR: δ_H (500 MHz, D_2O); selected peaks: 5.15 (d, 1H, J 1.9 Hz, βDGlc H-1α), 4.60 (d, 1H, J 8.0 Hz, βDGlc H-1β a), 4.47 (d, 1H, J 8.0 Hz, βDGal(terminal) H-1), 4.47 (d, 1H, J 8.0 Hz, βDGal H-1), 2.60 (m, 1H, αNeu5Ac H-3eq), 1.97 (s, 3H, αNeu5Ac C$\underline{H}_3$), 1.94 (s, 3H, βDGalNAc CH_3), 1.87 (m, 1H, αNeu5Ac H-3ax).

N-methyl 2-(*tert*-butoxycarbonylaminooxy)acetamide 38

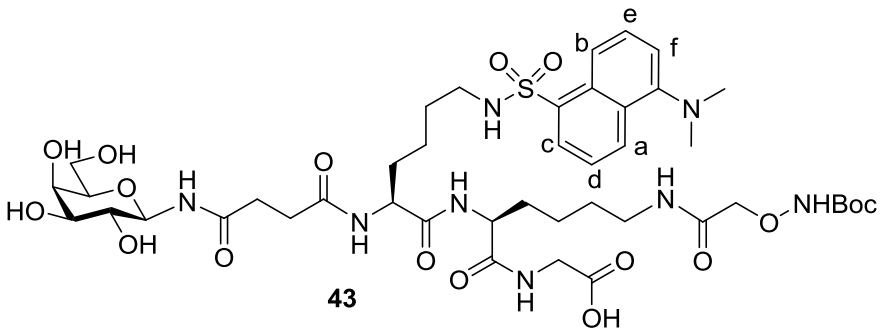


38

$MeNH_2$ (15 µL) was added to N-hydroxysuccinimide ester **37**[179] (25 mg) dissolved in methanol and stirred overnight before being concentrated. The product was purified by column chromatography (silica gel) and eluted with DCM/methanol 9:1. The pure fractions were combined and concentrated to afford the amide ligand **38**.

δ_H (500 MHz, MeOD); 4.27 (s, 2H, C$\underline{H}_2$), 2.82 (s, 3H, NHC$\underline{H}_3$), 1 50 (s, 9H, C(C$\underline{H}_3$)$_3$); δ_C (75 MHz, MeOD); 83.14 (C(CH_3)$_3$), 76.55 (CH_2), 28.57 (s, 3C, C(CH_3)$_3$), 25.95 (NHCH_3); **IR** (ν_{max}/cm^{-1}): 1724 (C=O), 1659 (C=O), 2978 (NH), 2931 (NH).

Galactose peptide ligand 43



43

Galactose Ligand **43** was synthesised on 2-chlorotrityl resin preloaded with glycine (200 mg, 0.8 mmol g^{-1}) by solid phase peptide synthesis (SPPS). The resin was swelled in DCM (35 ml) for 30 min. To the resin was added Fmoc Dde-protected lysine **46**[180] (256 mg, 0.48 mmol), HCTU (192 mg, 0.46 mmol), DMF (3.5 mL) and DIPEA (139 µL, 0.8 mmol) before being shaken for one hour. The resin was filtered and washed with DMF (3 mL, 3 x 2 min), then with piperidine/DMF (20%) (3 mL, 3 x 2 min) to remove the Fmoc group and then again with DMF (3 mL, 3 x 2 min).

Fmoc dansyl lysine **47**[181] (480 mg, 0.8 mmol) was mixed with HCTU (324 mg, 0.78 mmol), DMF (3.5 mL) and DIPEA (278 µL, 1.6 mmol) before being added to the resin and shaken for one hour. Again the resin was filtered and washed with DMF (3 mL, 3 x 2 min), then with piperidine/DMF (20%) (3 mL, 3 x 2 min) to remove the Fmoc group and then again with DMF (3 mL, 3 x 2 min).

Succinic anhydride (80 mg, 0.8 mmol) was added with DMF (3.5 mL) and DIPEA (278 µL, 1.6 mmol) before being shaken for 30 min. The resin was filtered and washed with DMF (3 mL, 3 x 2 min). A solution of galactosylamine (Sigma Aldrich, 91 mg, 0.5 mmol) in DMSO (3.5 mL) with HCTU (199 mg, 0.48 mmol) and DIPEA (41.7 µL, 0.24 mmol) was added before being shaken for one hour. Again the resin was filtered and washed with DMF (3 mL, 3 x 2 min), then with hydrazine/DMF (2%) (3 mL, 3 x 2 min) to remove the Dde protecting group and again with DMF (3 mL, 3 x 2 min).

A solution of N-hydroxysuccinimide ester **37**[179] (138 mg, 0.48 mmol) in DMF (3.5 mL) and DIPEA (139 µL, 0.8 mmol) was added before being shaken for 30 min. The resin was filtered and washed with DMF (3 mL, 3 x 2 min). The peptide was cleaved from the support by treating the resin with HFIP/DCM (30%) (3.5 mL) for 30 min;

this procedure was then repeated twice. The resin was subsequently washed with DCM (3 x 2 mL) and the combined washing solutions were combined and precipitated in cold ether, centrifuged and then the pellet was lyophilised from MeCN/water. The product was purified on an LH20 column equilibrated with methanol. The pure fractions were combined, concentrated and again lyophilised from MeCN/water to afford ligand **43** as a pale green foam. 19 mg, 12% yield. LC-MS showed the molecule to be 92.4% pure (figure 7.4.2.2).

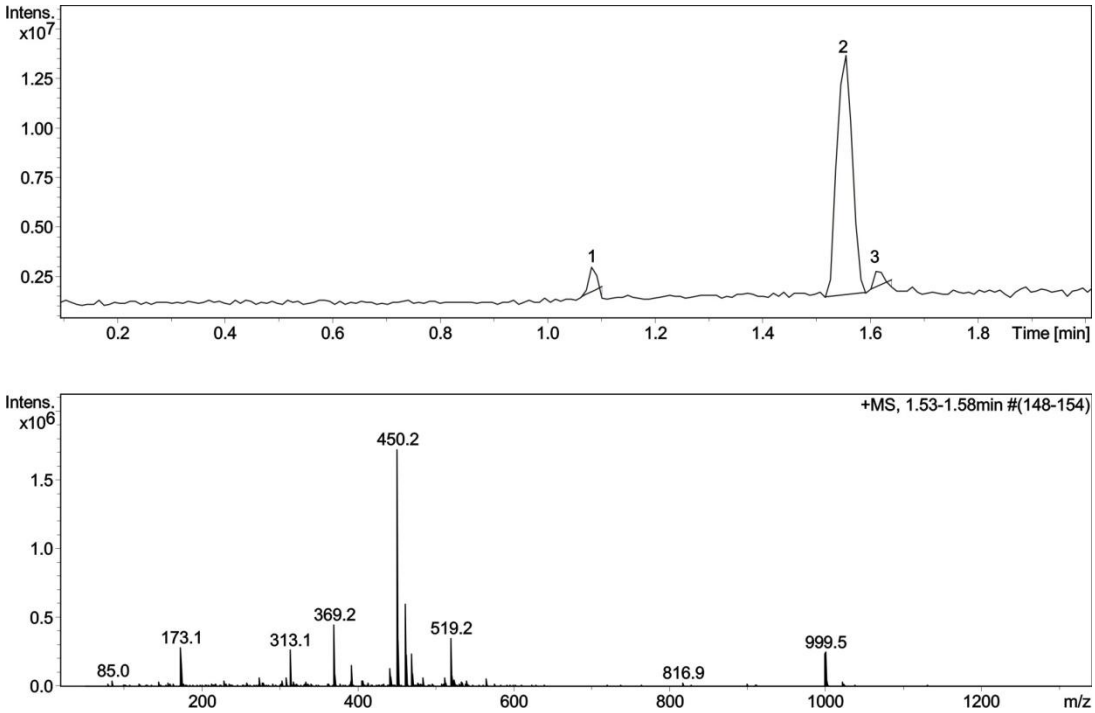
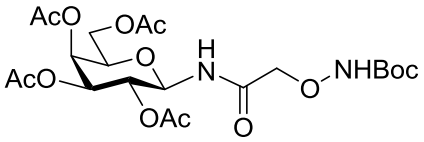


Fig. 7.4.2.2. mass spectrum showing the purity of galactose ligand **43**. The mass found for peak 2 by ESI-MS was 999.5 (calculated 999.4) and was 92.4% pure.

NMR δ_H (500 MHz, MeCN/D_2O 9:1); common peaks: 9.05 (d, 1H, *J* 8.6 Hz, Ar$\underline{H}_a$), 8.78 (d, 1H, J 8.7 Hz, Ar$\underline{H}_b$), 8.68 (dd, 1H, J 1.0 Hz, 7.3 Hz, Ar$\underline{H}_c$), 8.13 (m, 2H, Ar$\underline{H}_{d,e}$), 7.81 (d, 1H, J 7.5 Hz, Ar$\underline{H}_f$), 3.37 (s, 6H, N(C$\underline{H}_3$)$_3$); **IR** (v_{max}/cm^{-1}): 3377 (O-H), 1719 (C=O), 1648 (C=O); **HRMS**: found $[M+H]^+$ 999.4339, $C_{43}H_{67}N_8O_{17}S$ requires 999.4292.

***N*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-2-(*tert*-butoxycarbonylaminooxy)acetamide 62**

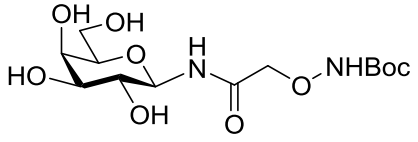


62

Anhydrous THF (2.5 mL) was added to 2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl azide[182] (100 mg, 0.27 mmol), DMAP (2 crystals), N-hydroxysuccinimide ester **37**[179] (88 mg, 0.294 mmol) and Pd/C (10 mg) before being placed under a H_2 atmosphere and left stirring overnight at room temperature. The reaction mixture was filtered through Celite and purified by flash column chromatography (silica gel; ethyl acetate/hexane 1:1). The pure fractions were concentrated to afford acetylated, Boc-protected ligand **62** as a colourless foam (75 mg, 54%).

$[\alpha]_D$ - 42.4 (c 0.5, $CHCl_3$); **NMR** δ_H (500 MHz, $CDCl_3$): 5.45 (s, 1H, H-4), 5.29 (dd, 1H, $J_{1,2}$ 9.3 Hz, $J_{1,NH}$ 9.3 Hz, H-1), 5.22 (dd, 1H, $J_{1,2}$ 9.3 Hz, $J_{2,3}$ 9.6 Hz, H-2), 5.15 (d, 1H, $J_{2,3}$ 9.6 Hz, H-3), 4.38 (d, 1H, $J_{CH2,CH2'}$ 16.5 Hz, CH_2), 4.33 (d, 1H, $J_{CH2,CH2'}$ 16.5 Hz, CH'_2) 4.15-4.03 (m, 3H, H-6, H-6', H-5), 2.17, 2.07, 2.04, 2.00 (4 × s, 12H, 4 × $COCH_3$), 1.51 (s, 9H, Boc); δ_C (75 MHz, $CDCl_3$): 171.4, 170.4, 170.1, 169.9, 169.6 (5 × C=O), 157.2 (C=O Boc), 77.9 (C-1), 77.2 ($\underline{C}(CH_3)_3$), 75.6 (CH_2), 72.4 (C-5), 71.0 (C-2), 68.4 (C-3), 67.1 (C-4), 61.2 (C-6), 28.1 ($C(\underline{C}H_3)_3$) 20.8, 20.7, 20.6, 20.6 ($C(O)(\underline{C}H_3)_3$); **IR** (ν_{max}/cm^{-1}): 2980 (C-H), 1750 (C=O); **HRMS**: found $[M+Na]^+$ 543.1822, $C_{21}H_{32}N_2O_{13}Na$ requires 543.1797.

***N*-β-D-galactopyranosyl 2-(*tert*-butyloxycarbonylaminoxy)acetamide 40**



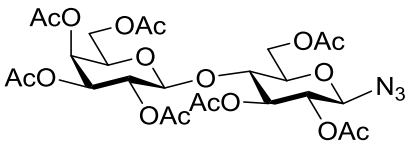
40

The acetyl protecting groups of compound **62** (25 mg) were removed by the addition of NaOMe (0.5 M, 77 µL) in anhydrous MeOH (240 µL) under N_2 overnight.

The product was concentrated and lyophilised from H_2O to afford ligand **40** as a colourless foam.

NMR: δ_H (500 MHz, CD_3OD): 4.90 (m, 1H, H-1), 4.28, 4.24 (2 × d, 2H, $J_{CH2,CH2'}$ 16.2 Hz, CH_2ON), 3.91 (d, 1H, $J_{3,4}$ 3.3 Hz, H-4), 3.62-3.77 (m, 3H, H-2, H-5, H-6a, H-6b), 3.57 (1H, (dd, 1H, $J_{2,3}$ 9.5 Hz $J_{3,4}$ 3.3 Hz, H-3), 1.46 (s, 9H, Boc).

2,3,6-tri-*O*-acetyl-4-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-(1→4)-β-D-glucopyranosyl azide 45 [183]

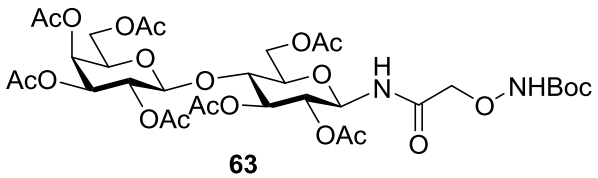


45

Anhydrous DCM (25 mL) was added to acetylated lactose (3 g, 4.16 mmol) and trimethylsilyl azide (0.66 mL, 4.99 mmol) under N_2. $SnCl_4$ (25 mL, 1 M in DCM) was added and the reaction was stirred overnight at room temperature. The reaction mixture was washed with H_2O twice and then saturated $NaHCO_3$, dried with magnesium sulfate and purified by flash column chromatography (silica gel; ethyl acetate/hexane 5:2). The pure fractions were concentrated to afford the acetylated lactosyl azide **45** as a colourless foam (1.49 g, 51%).

$[\alpha]_D$ - 22.4 (c 1, $CHCl_3$; lit. -22.0[183]); **NMR** δ_H (500 MHz, $CDCl_3$): 5.35 (d, 1H, $J_{3',4'}$ 3.4 Hz, H-4'), 5.21 (dd, 1H, $J_{2,3}$ 9.3 Hz, $J_{3,4}$ 9.4 Hz, H-3), 5.11 (dd, 1H, $J_{2',3'}$ 10.4 Hz, $J_{1',2'}$ 8.1 Hz, H-2'), 4.96 (dd, 1H, $J_{2',3'}$ 10.4 $J_{3',4'}$ Hz 3.4 Hz, H-3'), 4.86 (dd, 1H, $J_{1,2}$ 8.9 Hz $J_{2,3}$ 9.3 Hz, H-2), 4.63 (d, 1H, $J_{1,2}$ 8.9 Hz, H-1), 4.53 (m, 1H, H-6b), 4.50 (d, 1H, $J_{1',2'}$ 8.1 Hz, H-1'), 4.07-4.15 (m, 3H, H-6a, H-6a', H-6b'), 3.88 (dd, 1H, $J_{5',6a'}$ 6.8 Hz, $J_{5',6b'}$ 6.8 Hz, H-5'), 3.82 (dd, 1H, $J_{3,4}$ 9.4 Hz, $J_{4,5}$ 9.5 Hz, H-4), 3.71 (ddd, 1H, $J_{4,5}$ 9.5 Hz, $J_{5,6a}$ 6.6 Hz, $J_{5,6b}$ 4.9 Hz, H-5), 2.15, 2.14, 2.07, 2.07, 2.05, 2.05, 1.97 (7 × s, 21H, 7 × $COCH_3$) ; δ_C (75 MHz, $CDCl_3$): 170.4, 170.3, 170.1, 170.1, 169.6, 169.5, 169.1 (7 × C=O), 101.1 (C-1'), 87.7 (C-1), 75.8 (C-4), 74.8 (C-5), 72.5 (C-3), 71.0 (C-3'), 70.9 (C-5'), 70.7 (C-2), 69.0 (C-2'), 69.6 (C-4'), 61.7 (C-6), 60.7 (C-6'), 21.1, 20.8, 20.7, 20.7, 20.6, 20.5 (7 × C(O)$\underline{C}H_3)_3$); **IR** (ν_{max}/cm^{-1}): 2120 (N_3), 1750 (C=O); **HRMS**: found $[M+Na]^+$ 684.1890, $C_{26}H_{35}N_3O_{17}Na$ requires 684.1859.

***N*-[2,3,6-tri-*O*-acetyl-4-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-(1→4)-β-D-glucopyranosyl] 2-(tert-butyloxycarbonylaminoxy)acetamide 63**

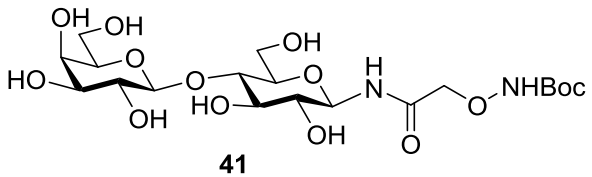


63

Anhydrous THF (2.5 mL) was added to acetylated lactosyl azide **45** (136 mg, 0.21 mmol), DMAP (2 crystals), N-hydroxysuccinimide ester **37**[179] (67.5 mg, 0.23 mmol) and Pd/C (10 mg) before being placed under a H_2 atmosphere and left stirring overnight. The reaction mixture was filtered through celite and purified by flash column chromatography (silica gel; ethyl acetate/hexane 3:1). The pure fractions were concentrated to afford acetylated, Boc protected ligand **63** as a colourless foam (80 mg, 48%).

$[\alpha]_D$ 21.2 (c 1, $CHCl_3$); **NMR δ_H** (500 MHz, $CDCl_3$): 5.28 (d, 1H, $J_{3',4'}$ 3.4 Hz, H-4'), 5.22 (dd, 1H, $J_{2,3}$ 9.3 Hz, $J_{3,4}$ 9.2 Hz, H-3), 5.20 (d, 1H, $J_{1,2}$ 9.3 Hz, H-1), 5.03 (dd, 1H, , $J_{1',2'}$ 7.9 Hz, $J_{2',3'}$ 10.4 Hz, H-2'), 4.90 (dd, 1H, $J_{2',3'}$ 10.4 Hz $J_{3',4'}$ 3.4 Hz, H-3'), 4.90 (dd, 1H, $J_{1,2}$ 9.3 Hz, $J_{2,3}$ 9.3 Hz,H-2), 4.43 (d, 1H, $J_{1',2'}$ 7.9 Hz H-1'), 4.37 (d, 1H, $J_{6a,6b}$ 10.9 Hz, H-6b), 4.29 (d, 2H, $J_{CH2,CH2'}$ 16.4 Hz, CH_2), 4.24 (d, 2H, $J_{CH2',CH2}$ 16.4 Hz, CH_2'), 4.10-4.00 (m, 3H, H-6a, H-6a', H-6b'), 3.84 (dd, 1H, $J_{5',6a'}$ 6.7 Hz, $J_{5',6b'}$ 9.5 Hz, H-5'), 3.76 (dd, 1H, $J_{3,4}$ 9.2 Hz, $J_{4,5}$ 9.8 Hz, H-4), 3.70 (m, 1H, H-5), 2.08, 2.04, 2.00, 1.99, 1.97, 1.95, 1.89 (7 × s, 21H, 7 × $COCH_3$), 1.41 (s, 9H, Boc) ; **δ_C** (75 MHz, $CDCl_3$): 171.4 (C=O), 170.9, 170.6, 170.4, 170.3, 170.0, 169.8, 169.2 (C=O), 157.7 (C=O), 101.1 (C-1'), 83.1 ($\underline{C}(CH_3)_3$), 77.6 (C-1), 76.1 (C-4), 76.0 (CH_2), 74.7 (C-3), 73.0 71.2 70.8 69.2 (C-2, C-2', C-3', C-5, C-5'), 66.8 (C-4'), 62.3 (C-6), 61.1 (C-6'), 28.0 ($C(\underline{C}H_3)_3$) 20.9, 20.7, 20.7, 20.6, 20.5, 20.5, 20.4 ($C(O)(\underline{C}H_3)_3$); **IR** (ν_{max}/cm^{-1}): 2981 (C-H), 1750 (C=O); **HRMS**: found $[M+Na]^+$ 831.2666, $C_{33}H_{48}N_2O_{21}Na$ requires 831.2642.

***N*-[4-*O*-β-D-galactopyranosyl-β-(1→4)-D-glucopyranosyl] 2-(tert-butyloxycarbonylaminoxy)acetamide[184] 41**

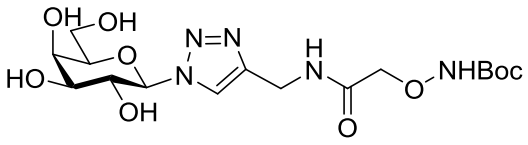


41

The acetyl protecting groups of compound **63** (50 mg) were removed by the addition of NaOMe (0.5 M, 173 µL) in anhydrous MeOH (300 µL) under N_2 overnight. The product was concentrated and lyophilised from H_2O to afford ligand **41** as a colourless foam.

NMR: δ_H (500 MHz, D_2O): 5.02 (d, 1H, $J_{1,2}$ 9.2 Hz, H-1), 4.45 (dd, 1H, $J_{1',2'}$ 7.7 Hz, H-1'), 4.25 (m, 2H, CH_2ON), 3.92-3.94 (m, 2H, H-4', H-6a), 3.64-3.83 (m, 8H, H-3, H-3', H-4, H-5, H-5', H-6a', H-6b, H-6b'), 2.85 (2H, (m, 1H, H-2, H-2'), 1.36 (s, 9H, Boc).

(4-[(tert-butoxycarbonyl)aminooxyacetamidyl]methyl)-([1,2,3]triazol-1-yl)-1-β-D-galactopyranoside 53 [185]



53

$CuSO_4$ (9.75 µL, 0.5 M in H_2O, 30 mol% final), sodium ascorbate (19.5 µL, 0.5 M in H_2O, 60 mol% final) and galactosyl azide **50**[182] (5.0 mg, 24.4 µmol) were added to 2-[*N*-(*t*-butoxycarbonyl)aminooxy]-*N*-(prop-2-ynyl)acetamide **52**[185] (3.7 mg, 16.2 µmol) with H_2O (25 µL) and DMF (271 µL) in a microwave tube and the reaction was subjected to 150 W, 80 °C, with stirring for 20 min to afford ligand **53**. The reaction mixture was lyophilised for direct use in protein reactions without further purification. **ESI-MS:** found $[M+H]^+$ 434.2 $C_{16}H_{27}N_5O_9$ requires 434.18.

8.2.3 Buffer solutions

All solutions were made up to 1 L with 18 MΩ water purified using an Elga purelab system unless otherwise stated. Buffer solutions were adjusted to the correct pH using NaOH or HCl.

Standard buffers

Phosphate buffer (pH 7.0, 0.1 M): $Na_2HPO_4.2H_2O$ (10.27 g), $NaH_2PO_4.2H_2O$ (6.60 g), 0.1 M NaCl (5.84 g).

HEPES buffer (pH 7.5): 50 mM HEPES (14.1 g), 150 mM NaCl (8.8 g), 5 mM $CaCl_2$ (0.55 g).

8.2.4 N-terminal oxidation and oxime formation

The concentrations stated for CTB refer to the monomer. AB_5 concentrations are given for the full complex. MS results stated are for the two most prominent charge states observed for each modified protein.

N-terminal oxidation of CTB to make CTBox

General procedure: The N-terminus of the CTB protein was modified by oxidation according to the procedure of Chen *et al*.[146] Methionine (15 µL, 200 mM, 10 equivalents, in *phosphate buffer*) was added to CTB (500 µL, 600 µM, in *phosphate buffer*), followed by the addition of $NaIO_4$ (30 µL, 50 mM, 5 equivalents, in *phosphate buffer*). After 5 min at room temperature the reaction was complete, yielding CTBox. **ESI-MS**: hemihydrate found $[M+11H]^{11+}$ 1057.06, $[M+10H]^{10+}$ 1162.65, $C_{514}H_{817}N_{139}O_{157}S_5$ requires $[M+11H]^{11+}$ 1056.99, $[M+10H]^{10+}$ 1162.59.

Oxime ligation at the CTB N-terminus

Ligands that were to be attached to the N-terminus of CTB via oxime formation were synthesised with a Boc-protected aminooxy group. The Boc group was then removed by the addition of TFA and stirring for 30 min before the TFA was evaporated under N_2. The product was redissolved in *phosphate buffer* for use in protein modification experiments or mixed directly with the protein. As the unprotected aminooxy group is highly reactive, it was used immediately without further characterisation. Masses quoted for the compounds used in oxime formation with CTBox or CTA2ox refer to the mass before removal of the Boc group, for

simplicity. After the removal of the Boc group the compound number gains an **a** to represent the deprotected version (as shown in previous chapters).

General procedure after optimisation: CTBox (500 µL, 600µM) and aniline (4.5 µL, to give a final concentration of 100 mM) were added to aminooxy-containing compound (10 equiv) in *phosphate buffer* and incubated at room temperature. Reactions were monitored by LC-MS until seen to be complete. The resulting modified protein was purified two times by SEC on a PD G-25 minitrap column to ensure complete removal of the unreacted oxyamine and aniline.

Acid catalysed amide ligand 38a attachment : CTB(38)

The amide ligand **38a** (10 µL, 0.16 M, 15 equiv) in *phosphate buffer* and acetic acid (50 µL) were added to CTBox (520 µL, 200 µM in *phosphate buffer*). The reaction went to completion after 30 min at room temperature, yielding **CTB(38)**. **ESI-MS**: found $[M+11H]^{11+}$ 1063.23, $[M+10H]^{10+}$ 1169.50, $C_{517}H_{821}N_{141}O_{157}S_6$ requires $[M+11H]^{11+}$ 1163.18, $[M+10H]^{10+}$ 1169.39.

Attachment of galactose peptide ligand 39a in the presence of a competing ligand: CTB(39)

The galactose ligand **39a** (1.07 mg, 10 equiv, in 50 µL *phosphate buffer*), aniline (1.5 µL, to give a final concentration of 100 mM) and methyl α-D-galactopyranoside (6.5 mg, 200 mM) as a competing ligand were added to CTBox (116.5 µL, 865 µM in *phosphate buffer*). After 90 min at room temperature the reaction went to completion, yielding **CTB(39)**. **ESI-MS**: found $[M+11H]^{11+}$ 1145.90, $[M+10H]^{10+}$ 1260.38, $C_{556}H_{877}N_{149}O_{172}S_6$ requires $[M+11H]^{11+}$ 1145.75, $[M+10H]^{10+}$ 1260.23.

Oxime ligation of small galactose ligand 40a: CTB(40)

Galactose ligand **40a** (0.58 mg, 27 equiv) in *phosphate buffer* (33 µL) and aniline (1.2 µL, to give a final concentration of 100 mM) were added to CTBox (100 µL, at a final concentration of 600 µM). The reaction went to completion after 24 h at room temperature, yielding **CTB(40)**. **ESI-MS**: found $[M+11H]^{11+}$ 1076.44, $[M+10H]^{10+}$ 1183.99, $C_{522}H_{829}N_{141}O_{162}S_5$ requires $[M+11H]^{11+}$ 1076.63, $[M+10H]^{10+}$ 1184.20.

Oxime ligation of lactose ligand 41a with or without a competing ligand: CTB(41)

Lactose ligand **41a** (1.34 mg, 35 equiv) in *phosphate buffer* (68 µL) and aniline (1.1 µL, to give a final concentration of 100 mM) were added to CTBox (55 µL, at a final concentration of 600 µM) with or without the presence of the competing ligand methyl α-D-galactopyranoside (4.5 mg, to a concentration of 200 mM) and these reactions went to completion after 3 h at room temperature, yielding **CTB(41)**. **ESI-MS**: found $[M+11H]^{11+}$ 1091.20, $[M+10H]^{10+}$ 1200.25, $C_{528}H_{839}N_{142}O_{167}S_5$ requires $[M+11H]^{11+}$ 1091.37, $[M+10H]^{10+}$ 1200.40.

Oxime ligation of galactose peptide ligand 42a with or without a competing ligand: CTB(42)

Galactose ligand **42a** (1.2 mg, 18 equiv) in *phosphate buffer* (64 µL) and aniline (1 µL, to give a final concentration of 100 mM) were added to CTBox (43.5 µL, to give a final concentration of 600 µM) with and without the presence of the competing ligand methyl α-D-galactopyranoside (4.2 mg, to a concentration of 200 mM). Galactosyl ligand **42a** (0.4 mg, 18 equiv) in *phosphate buffer* (94 µL) was also added with aniline (1 µL, to give a final concentration of 100 mM) to CTBox (14.6 µL, to give a final concentration of 200 µM) with and without the presence of a competing ligand methyl α-D-galactopyranoside (4.2 mg, to give a concentration of 200 mM). These reactions all went to completion after 24 h at room temperature, yielding **CTB(42)**. **ESI-MS**: found $[M+11H]^{11+}$ 1135.30, $[M+10H]^{10+}$ 1248.64, $C_{552}H_{871}N_{147}O_{170}S_6$ requires $[M+11H]^{11+}$ 1135.39, $[M+10H]^{10+}$ 1248.82.

N-terminal oxidation of CTB mutants: W88Kox and W88Eox

General procedure: The N-terminus of each of the CTB mutant proteins (W88K, W88K Q61E and W88E) was modified by oxidation according to the procedure of Chen *et al*.[146] Methionine (15 µL, 200 mM, 10 equivalents, in *phosphate buffer*) was added to the CTB mutant (500 µL, 600 µM, in *phosphate buffer*), followed by the addition of $NaIO_4$ (30 µL, 50 mM, 5 equivalents, in *phosphate buffer*). After 5 min at room temperature the reaction was complete, yielding W88Kox and W88Eox. The double mutant W88K Q61E failed to oxidise.

ESI-MS: W88Kox hemihydrate found $[M+11H]^{11+}$ 1051.51, $[M+10H]^{10+}$ 1156.56, $C_{509}H_{819}N_{139}O_{157}S_5$ requires $[M+11H]^{11+}$ 1051.66, $[M+10H]^{10+}$ 1156.73.

ESI-MS: W88Eox hemihydrate found $[M+11H]^{11+}$ 1051.60, $[M+10H]^{10+}$ 1156.68, $C_{508}H_{814}N_{138}O_{159}S_5$ requires $[M+11H]^{11+}$ 1051.71, $[M+10H]^{10+}$ 1156.79.

Oxime ligation of galactose peptide ligand 42a to W88K: W88K(42)

W88Kox (31 µL, 241 µM) and aniline (0.4 µL, to give a final concentration of 100 mM) were added to galactose **42a** (0.23 mg, 30 equiv) in *phosphate buffer* (6.5 µL). The reaction went to completion after 24 h at room temperature, yielding **W88K(42)**. **ESI-MS**: found $[M+10H]^{10+}$ 1242.75, $[M+11H]^{11+}$ 1129.88, $C_{547}H_{873}N_{147}O_{170}S_6$ requires $[M+10H]^{10+}$ 1242.96, $[M+11H]^{11+}$ 1130.06.

Oxime ligation of galactose peptide ligand 42a to W88E: W88E(42)

W88Eox (465 µL, 216 µM) and aniline (2.2 µL, to give a final concentration of 100 mM) were added to galactose **42a** (1.0 mg, 10 equiv) in *phosphate buffer* and this reaction went to completion after 24 h at room temperature. The product was purified by SEC, yielding **W88E(42)**. **ESI-MS**: found $[M+10H]^{10+}$ 1242.84, $[M+11H]^{11+}$ 1129.98, $C_{546}H_{868}N_{146}O_{172}S_6$ requires $[M+10H]^{10+}$ 1243.02, $[M+11H]^{11+}$ 1130.11.

Oxime ligation of galactose peptide ligand 39a to W88E: W88E(39)

W88Eox (110 µL, 520 µM) and aniline (1.0 µL, to give a final concentration of 100 mM) were added to galactose **39a** (0.5 mg, 8 equiv) in *phosphate buffer* and this reaction went to completion after 20 h at room temperature. The product was purified by SEC, yielding **W88E(39)**. **ESI-MS**: found $[M+8H]^{8+}$ 1567.92, $[M+9]^{9+}$ 1393.82, $C_{550}H_{874}N_{148}O_{174}S_6$ requires $[M+8]^{8+}$ 1567.78, $[M+9H]^{9+}$ 1393.70.

Oxime ligation of alkyne ligand 52a to CTB: CTB(52)

CTBox (419 µL, 305 µM) and aniline (0.4 µL, to give a final concentration of 100 mM) were added to alkyne **52a** (0.30 mg, 10 equiv) in *phosphate buffer* and this reaction went to completion after 20 h at room temperature. The modified protein was purified using a PD G-25 minitrap column and then concentrated using an Amicon Ultra-15 (30,000 MWCO) to yield **CTB(52)**. **ESI-MS**: found $[M+11H]^{11+}$ 1065.36, $[M+10H]^{10+}$ 1171.79, $C_{519}H_{821}N_{141}O_{157}S_5$ requires $[M+11H]^{11+}$ 1065.52, $[M+10H]^{10+}$ 1171.96.

Oxime ligation of galactose triazole ligand 53a to CTB: CTB(53)

CTBox (206 µL, 204 µM) and aniline (1.9 µL, to give a final concentration of 100 mM) were added to ligand **53a** (3.25 µmol, 76 equiv) in *phosphate buffer* and this reaction went to completion after 20h at room temperature. The resultant modification was purified by a PD G-25 minitrap column to yield **CTB(53)**. **ESI-MS**: found $[M+11H]^{11+}$ 1084.26, $[M+10H]^{10+}$ 1192.52, $C_{525}H_{832}N_{144}O_{162}S_5$ requires $[M+11H]^{11+}$ 1084.00, $[M+10H]^{10+}$ 1192.30.

Oxime ligation of GM1 ligand 54a to W88E: W88E(54)

Oxidised W88E protein (100 µL, 400 µM) and aniline (0.9 µL, to give a final concentration of 100 mM) were added to GM1 ligand **54a** (0.64 mg, 11 equiv) in *phosphate buffer*. After 20 h at room temperature, the modified protein was purified by a PD G-25 minitrap column, yielding **W88E(54)**. **ESI-MS**: found $[M+10H]^{10+}$ 1285.46, $[M+9H]^{9+}$ 1428.17, $C_{561}H_{901}N_{145}O_{188}S_5$ requires $[M+10H]^{10+}$ 1285.34, $[M+9H]^{9+}$ 1428.05.

N-terminal oxidation of CTA2 (G1S): CTA2ox

The N-terminus of CTA2 (G1S) was modified in a similar way to CTB by oxidation according to the procedure of Chen *et al*.[146] Methionine (1.42 µL, 200 mM, 10 equivalents, in *phosphate buffer*) was added to AB_5 (810 µL, 35 µM, in *phosphate buffer*), followed by the addition of $NaIO_4$ (2.84 µL, 50 mM, 5 equivalents, in *phosphate buffer*). After 5 min at room temperature the reaction was complete, yielding CTA2ox. **ESI-MS**: CTA2ox hemiacetal found $[M+7H]^{7+}$ 727.64, $[M+6H]^{6+}$ 848.74, $C_{237}H_{371}N_{67}O_{80}S_1$ requires $[M+7H]^{7+}$ 727.57, $[M+6]^{6+}$ 848.66. CTB found $[M+10H]^{10+}$ 1162.21, $[M+9H]^{9+}$ 1291.23, $C_{515}H_{820}N_{140}O_{155}S_5$ requires $[M+10H]^{10+}$ 1162.19, $[M+9H]^{9+}$ 1291.21.

Oxime ligation of biotin 61 to CTA2 (G1S): A(61)B_5

CTA2ox as part of the AB_5 complex (814 µL, 35 µM) and aniline (7.4 µL, to give a final concentration of 100 mM) were added to biotin ligand **61** (3.0 mg, 240 equiv) in *phosphate buffer* and this reaction proceeded overnight before purification two times by PD-10 G-25 minitrap column, yielding **A(61)B_5**. **ESI-MS**: found $[M+6H]^{6+}$

912.79, $[M+5H]^{5+}$ 1095.15, $C_{551}H_{878}N_{146}O_{174}S_5$ requires $[M+6H]^{6+}$ 912.77, $[M+5H]^{5+}$ 1095.13.

8.2.5 Protein CuAAC

Copper-catalysed azide-alkyne cycloaddition (CuAAC) between CTB(52) and lactosyl azide 51 in the presence of TBTA

Alkyne-modified CTB, **CTB(52)**, (142 µL, 305 µM, 80 µM final concentration) was mixed with $CuSO_4$ (5 µL, 100 mM, 2 mM final concentration), TCEP (12.5 µL, 100 mM, 5 mM final concentration) and lactosyl azide **51** (0.375 mg, 78 µL, 2.8 mM final concentration) all in *phosphate buffer* (at pH 8). TBTA was added to the solution (12.5 µL, 100 mM, 2 mM final concentration) in DMSO and the mixture was vortexed at room temperature for 48 h.[159] After centrifugation at 13 krpm for 1 min, the supernatant was purified using a PD G-25 minitrap column. ESI-MS showed an incomplete reaction with most of the product being the desired lactose modification but also a significant amount of alkyne starting material (Fig 8.2). Other products identified included CTB modified with a galactosyl ligand, indicating partial cleavage of the lactose moiety, and another product with only the aglycone chain attached suggesting some full cleavage of the carbohydrate section.

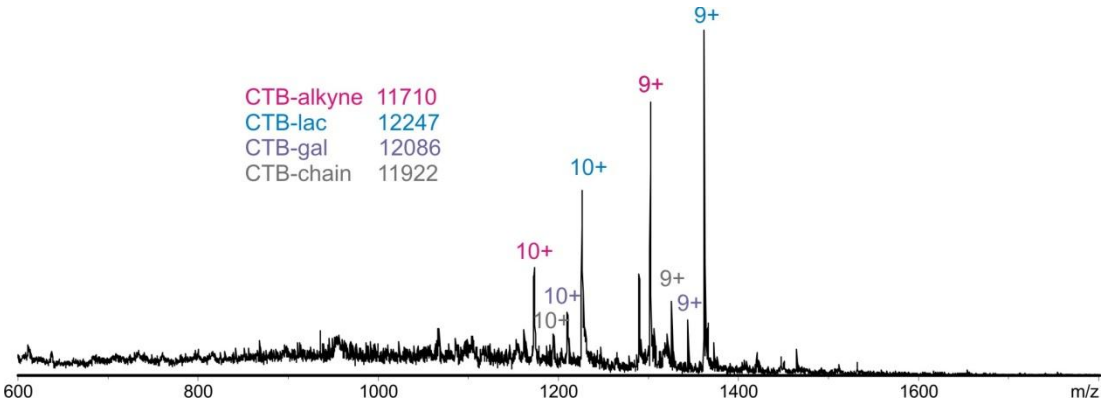


Fig 8.2 Mass spectrum of CuAAC reaction showing incomplete conversion.

CuAAC between CTB(52) and lactosyl azide 51 in the presence of BPSA

Alkyne **CTB(52)** (37 µL, 540 µM, 100 µM final concentration) was mixed with $CuSO_4$ (4 µL, 100 mM, 2 mM final concentration), TCEP (8 µL, 100 mM, 4 mM final concentration) and BPSA (8 µL, 100 mM, 4 mM final concentration) all in *phosphate buffer* (pH 8). Lactose azide **51** (0.1 mg, 143 µL, 5 mM final concentration) was added to the solution and the mixture was vortexed at room temperature for 48

h.[162] ESI-MS showed no reaction had taken place and all the protein was still the alkyne derivative.

CuAAC between CTB(52) and galactosyl azide 50 in the presence of either BPSA or TBTA

Alkyne **CTB(52)** (46.3 µL, 540 µM, 100 µM final concentration) was mixed with $CuSO_4$ (5 µL, 100 mM, 2 mM final concentration), TCEP (5 µL, 100 mM, 5 mM final concentration) and galactosyl azide **50** (32 mM, 39 µL, 5 mM final concentration) all in *phosphate buffer* (pH 8). BPSA or TBTA (12.5 µL, 40 mM, 2 mM final concentration) was added to the solution, in *phosphate buffer* or DMSO respectively, with extra *phosphate buffer* (142.2 µL, pH 8) and the mixture was vortexed at room temperature for 24 h.[159] Analysis by ESI-MS showed that no reaction had taken place with BPSA and an incomplete reaction had taken place with TBTA (Fig 8.3). The mass spectrum also showed an incomplete alkyne addition as some CTBox was still present, but this did not affect the progress of the CuAAC reaction.

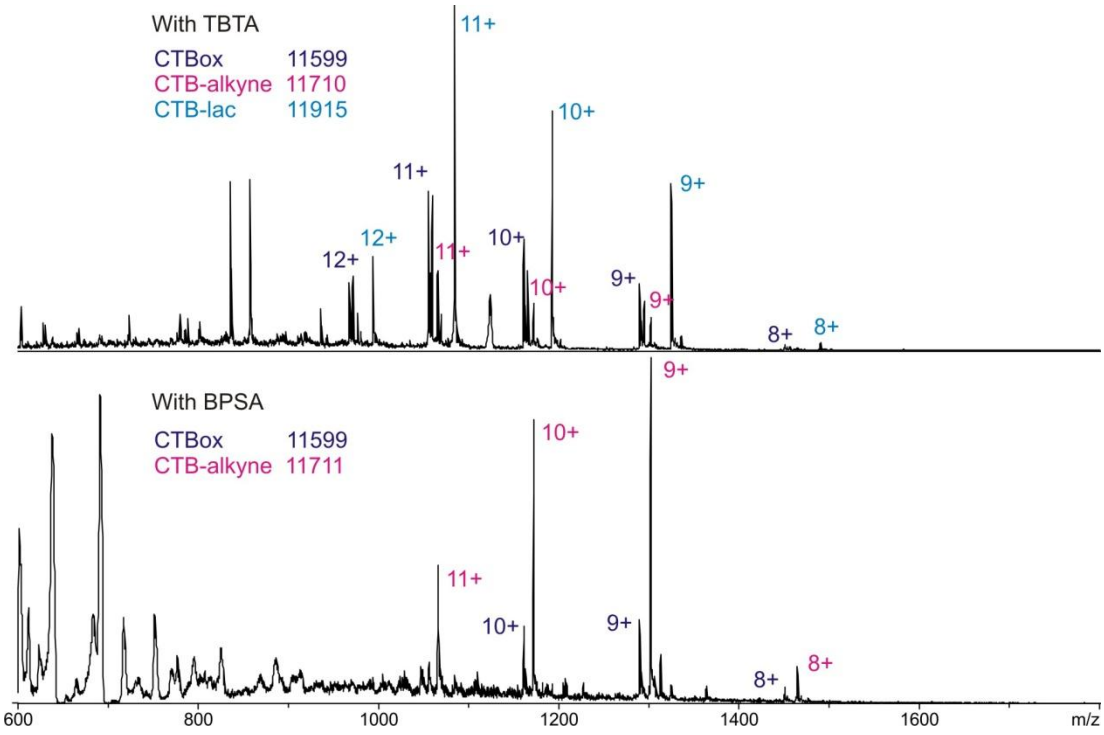


Fig 8.3 Mass spectrum of CuAAC reaction showing incomplete conversion.

8.2.6 Enzymatic cleavage and ligations

Cleavage of MBP-AB_5 using Tobacco Etch Virus (TEV) protease

General procedure: TEV protease (30 µL, 63 µM, 5 mol%) was added to MBP-AB_5 (750 µL, 50 µM) in *HEPES buffer* and incubated at room temperature for 90 min. The cleaved AB_5 protein was purified by Ni affinity chromatography, eluting with 500 mM imidazole in *HEPES buffer*.

ESI-MS: CTA2 (wild-type) found $[M+7H]^{7+}$ 723.48, $[M+6H]^{6+}$ 843.73, $C_{219}H_{342}N_{64}O_{74}$ requires $[M+7H]^{7+}$ 723.22, $[M+6H]^{6+}$ 843.59. Found for CTB $[M+11H]^{11+}$ 1059.33, $[M+10H]^{10+}$ 1165.14, $C_{516}H_{822}N_{140}O_{156}S_5$ requires $[M+11H]^{11+}$ 1059.45, $[M+10H]^{10+}$ 1165.29.

ESI-MS: CTA2 (D4N) found $[M+6H]^{6+}$ 621.56, $[M+5H]^{5+}$ 745.67, $C_{155}H_{248}N_{52}O_{55}$ requires $[M+6H]^{6+}$ 620.99, $[M+5H]^{5+}$ 744.98. Found for CTB $[M+11H]^{11+}$ 1059.75, $[M+10H]^{10+}$ 1165.51, $C_{516}H_{822}N_{140}O_{156}S_5$ requires $[M+11H]^{11+}$ 1059.45, $[M+10H]^{10+}$ 1165.29.

ESI-MS: CTA2 (G1S) found $[M+7H]^{7+}$ 727.37, $[M+6H]^{6+}$ 848.43 $C_{220}H_{344}N_{64}O_{75}$ requires $[M+7H]^{7+}$ 727.50, $[M+6H]^{6+}$ 848.58. Found for CTB (T1A) $[M+10H]^{10+}$ 1162.21, $[M+9H]^{9+}$ 1291.34, $C_{515}H_{820}N_{140}O_{155}S_5$ requires $[M+10H]^{10+}$ 1162.19, $[M+9H]^{9+}$ 1291.21.

Sortase ligation of bpy depsipeptide 60 and CTA2 (wild-type): $A(60)B_5$

AB_5 (wild-type) (120 µL, 225 µM) was mixed with sortase (7.1 µL, 380 µM, 10 mol%) and bpy depsipeptide **60** (27 µL, 5 mM, 5 equiv). The reaction was incubated at 37 °C for 3 h before purification by SEC to yield **$A(60)B_5$**. **ESI-MS**: found $[M+8H]^{8+}$ 749.11, $[M+7H]^{7+}$ 855.91, $C_{265}H_{399}N_{73}O_{86}$ requires $[M+8H]^{8+}$ 748.87 $[M+7H]^{7+}$ 855.70.

Attempted sortase ligation of Fe(bpy 60)$_3$ to CTA2 (wild-type)

$FeCl_2$ (3.4 µL, 2 mM) and bpy depsipeptide **60** (10.2 µL, 2 mM) in *HEPES buffer* were mixed together resulting in a red solution. AB_5 (wild-type) (85 µL, 218 µM) was mixed with sortase (9.8 µL, 380 µM, 20 mol%) and added to the Fe(bpy **60**)$_3$ complex (1 equiv bpy **60** to AB_5). After incubation at 37 °C for 1 h the red colour was no longer present and after 4 h analysis by ESI-MS showed no ligation reaction had taken place.

Sortase ligation of bpy depsipeptide 60 and CTA2 (D4N): A(60)B_5 (D4N)

AB_5 (D4N) (280 µL, 193 µM) was mixed with sortase (18 µL, 600 µM, 20 mol%) and bpy depsipeptide **60** (54 µL, 10 mM, 10 equiv). The reaction was incubated at 37 °C for 4 h before purification by Ni affinity chromatography to yield **A(60)B_5 (D4N)**. **ESI-MS**: found $[M+7H]^{7+}$ 665.85, $[M+6H]^{6+}$ 776.65, $C_{201}H_{305}N_{61}O_{67}$ requires $[M+7H]^{7+}$ 664.90, $[M+6H]^{6+}$ 775.55.

8.3 Biophysical analysis of binding interactions

8.3.1 Atomic force microscopy (AFM)

AFM experiments were carried out using a Nanoscope III Multimode AFM and Veeco NP-S cantilevers (spring constant ~0.3 N m^{-1}). Experiments were conducted in buffers listed below using a fluid cell and images were acquired in tapping mode at a frequency of ~9 kHz. Images reported here are raw, unfiltered data and in height mode were collected with 512 × 512 points at a scanning rate of 1-2 Hz.

A mica disk was freshly cleaved with tape before samples of protein solution (20-80 µL, 10^{-7} M) were added to the mica surface and incubated for 15-60 min before diluting with 0-2 x 30 µL of 2 mM *citrate-phosphate buffer* pH 5 before being imaged.

For three component bilayers 5 mM stock solutions of DOPC, cholesterol and sphingomyelin were prepared in $CHCl_3$. A 1 mg ml^{-1} 1:1:1 lipid solution was created by adding each component (125 µL). The $CHCl_3$ was evaporated under N_2, placed in a desiccator overnight and the lipids redissolved in H_2O to give a cloudy solution as multilamellar vesicles (MLVs) are formed. After sonication for 30 min, the solution became clear indicating that the lipids had been broken up into small unilamellar vesicles (SUVs). Lipid solution (100 µL) was applied to a freshly cleaved mica surface and incubated for 30 min. The mica was then rinsed with H_2O and left for a further hour before being imaged in contact mode.

DOPC lipid bilayers were prepared as above but were dissolved in PBS buffer. GM1 was incorporated into the lipid solution at 1 mg ml^{-1} after sonication and incubated overnight at 4 °C. CTB (40 µL, 10^{-7} M) was added after the bilayers were applied to a mica surface and incubated for 30 min before imaging in tapping mode.

8.3.2 Analytical ultracentrifugation (AUC)

Samples (0.41 mL) were centrifuged in 1.2 cm pathlength 2-sector aluminium centrepiece cells (sample in right hand sector, reference buffer in left hand sector) built with sapphire windows in an 8-place An50Ti or 6-place An60Ti analytical rotor running in an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Inc., Palo Alto, California 94304) at 35 krpm (unless otherwise stated) and at a temperature of 20 °C. Changes in solute concentration were detected by 150 absorbance scans measured at 280 nm (or other appropriate wavelength) over a period of 5-15 hours.

Analysis and fitting of the data was performed using the software SedFit.[137] A continuous c (s) distribution model was fitted to the data, taking every 2nd scan. The resolution was set at 200 over a sedimentation coefficient range of 0.1-15.0 S. Parameters were set for the partial specific volume as 0.73654, the buffer density of 1.04910 and the buffer viscosity at 0.01410. The frictional coefficient, the baseline and the raw data noise were floated in the fitting. The meniscus and bottom of the cell path were also floated after initial estimations from the raw data.

8.3.3 Dynamic light scattering (DLS)

Disposable polystyrene cuvettes were used for light scattering experiments performed using a Zetasizer Nano ZS (Malvern) at 20 °C measuring the scattering at an angle of 173° and using a laser wavelength of 633 nm. Protein samples (50-200 µL, 50-500 µM) and detergent samples (200 µL, 1-2 mM) in the appropriate buffer (either *phosphate buffer* or *HEPES buffer*) were filtered through a 0.8 µm filter (Millipore membrane filters) prior to analysis to ensure no dust particles were present which would skew the experiment. Readings of 10 scans over 10 seconds each were taken in triplicate with the average calculated. Measurement data was analysed using volume % which accounts for the volume of the objects when calculating relative quantities with a certain diameter. This measurement causes the data to not become skewed by the high scattering of larger particles.

8.3.4 Enzyme-linked Lectin assay (ELLA)

96-well polystyrene plates (NUNC Maxisorp) were prepared for the simultaneous testing of multiple samples. Ganglioside GM1 (100 µL, 1.3 µM in ethanol) was coated onto each well of a 96-well microtitre plate and the solvent was left to evaporate. The plate was washed with PBS (3 × 200 µL) to remove any unattached GM1 and any remaining free binding sites in the wells were blocked with BSA by

incubating with a PBS solution containing 1% (w/v) BSA (100 µL) for 30 min at 37 °C. The wells were then washed again with PBS (3 × 200 µL). A series of inhibitor samples spanning a range of concentrations were prepared in PBS containing 0.1% BSA and 0.05% Tween 20. Each sample was mixed with CTB-HRP in the same buffer to give a final CTB-HRP concentration of 25 ng mL^{-1}. These mixtures of inhibitor and toxin were incubated at room temperature for 2 h before 200 µL was transferred to each GM1 coated well. The limits of detection were determined by control samples containing only the CTB-HRP component and no inhibitor, which gave the highest response, and a blank sample of only buffer which gave the lower limit. These two measurements were used to find the maximum and minimum optical density values for CTB-HRP binding to the GM1-coated wells. The inhibitor-toxin mixtures were incubated at room temperature for 30 min before the plate was washed with PBS, 0.1% BSA, 0.05% Tween 20 (3 × 200 µL), to remove the unbound CTB-HRP-inhibitor complexes. An OPD solution was freshly prepared (25 mg OPD.2HCL, 7.5 mL 0.1 M citric acid, 7.5 mL 0.1 M sodium citrate and 6 µL of a 30% H_2O_2 solution, pH 6.0) and added to each well (100 µL). The solution was allowed to react in the dark, at room temperature, for 25 min before the oxidation reaction was quenched by the addition of 1 M H_2SO_4 (50 µL). The absorbance at 490 nm was measured within 5 min.

The samples were analysed in triplicate, initially over a large concentration range using a 10-fold dilution. A more accurate 2-fold dilution was then performed around the expected IC_{50} values. Anomalous data points were removed in a few cases leaving duplicate results. The error of each sample was calculated as the standard error (Equation 8.3).

Equation 8.3

$$standard\ error = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}}{\sqrt{N}}$$

where N is the size of the sample, x_i is the observed value of each sample and $\bar{x}$ is the mean value of the sample.

The absorbance of each sample was converted to percentage binding by comparison with the maximum absorbance and minimum absorbance obtained from the control samples. These values were then plotted against the concentration of inhibitor for each sample and curve fitting was performed using the standard dose-response model in Origin (Equation 8.4). Fitting was performed with weighted

data points according to the standard error. Concentration values were converted to log(concentration) before curve fitting.

Equation 8.4

$$y = A1 + \frac{A2 - A1}{1 + 10^{(logx0 - x)p}}$$

Where *A1* is y-value for the bottom plateau, *A2* is the y-value for the top plateau, *x0* is the IC_{50} value, *x* is log(inhibitor concentration) and *p* is the Hill slope parameter.

8.3.5 Isothermal titration calorimetry (ITC)

ITC titrations were performed using a Microcal ITC_{200} calorimeter (GE healthcare) with a cell volume of 0.2028 mL and data processed and fitted using Origin with the one site model.[186] Protein samples were dialysed into *phosphate buffer* (SnakeSkin® pleated dialysis tubing, Thermo Scientific, with 7000 MWCO) prior to analysis. The ligands were then dissolved in the dialysis buffer to ensure an exact match for the buffers during the titration. The reference cell was filled with water and the analysis cell was filled with the protein solution and both were allowed to reach thermal equilibrium at 25 °C before titrations were performed. Titrations typically comprised 20 injections of 2 µL at 2 minute intervals. Separate titrations of the ligand into buffer were used to subtract the heat of dilution from the integrated data prior to curve fitting. The binding stoichiometry (n) was fixed at 1.0 for low c-value titrations as described by Turnbull and Daranas.[141]

9 Appendix

Protein and DNA sequences, and yet more synthetic data

9.1 Protein sequences

Introduced mutations in the amino acid sequences are marked in red and underlined. Differences from the wild-type El Tor biotype of CTB in the CTB classical biotype and in LTB are marked in green. The TEV cleavage site in the MBP-CTA2 sequence is marked in blue. The sequence for CTA2 begins at G399 in the MBP-CTA2 sequence

CTB

```
                  1           11          21          31
wt (El Tor)       TPQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
W88K              TPQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
W88K T1A          APQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
W88K Q61E         TPQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
W88E              TPQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
T1A               APQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
I47T              TPQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII
CTB classical     TPQNITDLCA EYHNTQIHTL NDKIFSYTES LAGKREMAII
LTB               APQSITELCS EYHNTQIYTI NDKILSYTES MAGKREMVII

                  41          51          61          71
wt (El Tor)       TFKNGAIFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
W88K              TFKNGAIFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
W88K T1A          TFKNGAIFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
W88K Q61E         TFKNGAIFQV EVPGSQHIDS EKKAIERMKD TLRIAYLTEA
W88E              TFKNGAIFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
T1A               TFKNGAIFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
I47T              TFKNGATFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
CTB classical     TFKNGATFQV EVPGSQHIDS QKKAIERMKD TLRIAYLTEA
LTB               TFKSGATFQV EVPGSQHIDS QKKAIERMKD TLRITYLTET

                  81          91          101
wt (El Tor)       KVEKLCVWNN KTPHAIAAIS MAN
W88K              KVEKLCVKNN KTPHAIAAIS MAN
W88K T1A          KVEKLCVKNN KTPHAIAAIS MAN
W88K Q61E         KVEKLCVKNN KTPHAIAAIS MAN
W88E              KVEKLCVENN KTPHAIAAIS MAN
T1A               KVEKLCVWNN KTPHAIAAIS MAN
I47T              KVEKLCVWNN KTPHAIAAIS MAN
CTB classical     KVEKLCVWNN KTPHAIAAIS MAN
LTB               KVEKLCVWNN KTPNSIAAIS MEN
```

MBP-CTA2

```
                  1           11          21          31
wild-type         KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
D4N               KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
G1S               KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
```

```
                41          51          61          71
wild-type       DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT
D4N             DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT
G1S             DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT

                81          91          101         111
wild-type       PDKAFQDKLY PFTWDAVRYN GKLIAYPIAV EALSLIYNKD
D4N             PDKAFQDKLY PFTWDAVRYN GKLIAYPIAV EALSLIYNKD
G1S             PDKAFQDKLY PFTWDAVRYN GKLIAYPIAV EALSLIYNKD

                121         131         141         151
wild-type       LLPNPPKTWE EIPALDKELK AKGKSALMFN LQEPYFTWPL
D4N             LLPNPPKTWE EIPALDKELK AKGKSALMFN LQEPYFTWPL
G1S             LLPNPPKTWE EIPALDKELK AKGKSALMFN LQEPYFTWPL

                161         171         181         191
wild-type       IAADGGYAFK YENGKYDIKD KYDIKDVGVD NAGAKAGLTF
D4N             IAADGGYAFK YENGKYDIKD VGVDNAGAKA GLTFLVDLIK
G1S             IAADGGYAFK YENGKYDIKD VGVDNAGAKA GLTFLVDLIK

                201         211         221         231
wild-type       NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
D4N             NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
G1S             NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV

                241         251         261         271
wild-type       NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL
D4N             NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL
G1S             NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL

                281         291         301         311
wild-type       ENYLLTDEGL EAVNKDKPLG AVALKSYEEE LVKDPRIAAT
D4N             ENYLLTDEGL EAVNKDKPLG AVALKSYEEE LVKDPRIAAT
G1S             ENYLLTDEGL EAVNKDKPLG AVALKSYEEE LVKDPRIAAT

                321         331         341         351
wild-type       MENAQKGEIM PNIPQMSAFW YAVRTAVINA ASGRQTVDEA
D4N             MENAQKGEIM PNIPQMSAFW YAVRTAVINA ASGRQTVDEA
G1S             MENAQKGEIM PNIPQMSAFW YAVRTAVINA ASGRQTVDEA

                361         371         381         391
wild-type       LKDAQTNSSS NNNNNNNNNN LGIEGRISHM GSENLYFQGG
D4N             LKDAQTNSSS NNNNNNNNNN LGIEGRISHM GSENLYFQGG
G1S             LKDAQTNSSS NNNNNNNNNN LGIEGRISHM GSENLYFQSG

                401         411         421         431
wild-type       GDEKTQSHGV KFLDEYQSKV KRQIFSGYQS DIDTHNRIKD
D4N             GNNNNN---- -------SKV KRQIFSGYQS DIDTHNRIKD
G1S             GDEKTQSHGV KFLDEYQSKV KRQIFSGYQS DIDTHNRIKD

                441
wild-type       EL
D4N             EL
G1S             EL
```

9.2 DNA plasmid sequences

The genes encoding the proteins CTB, MBP and CTA2 are highlighted in red, blue and green respectively.

9.2.1 pATA13

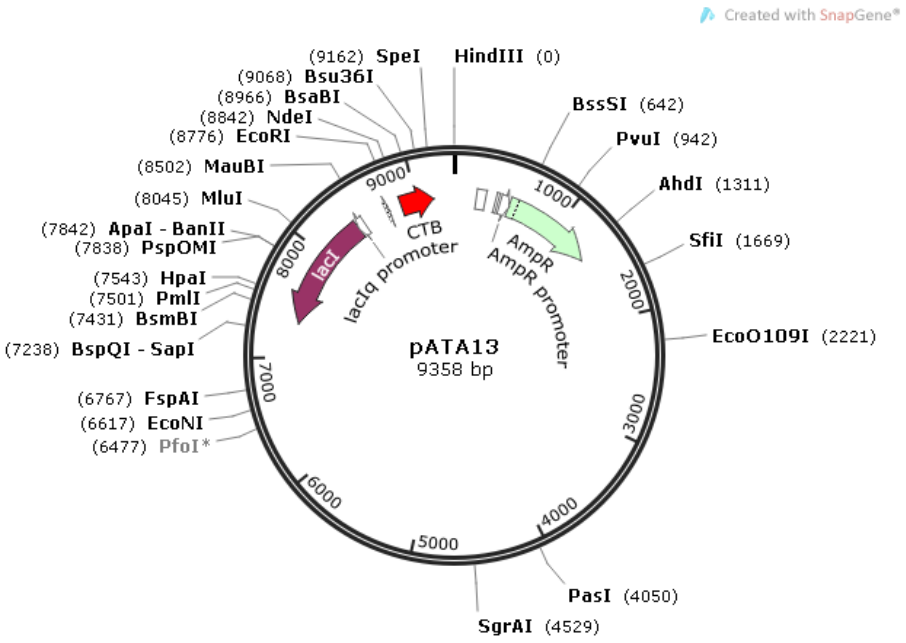


Fig 9.1 Plasmid map for pATA13

The CTB plasmid for expression in *Vibrio sp.60* is pATA13 (Fig 9.1).[187] It is derived from pTRH64[188], which in turn is derived from pMMB68[189], which originates from plasmid PMMB66EH.[190]

Full sequence:

AGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGC
GGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTC
AGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGG
CATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAA
CGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGT
GGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGC
CTTTTTGCGTTTCTACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA
GACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGT
GTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAA
AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTA
AGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT
GGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAA
TGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTAT
GCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCG
AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGA
GCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGC
GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCG
GATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGG
AGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCG

TAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGT
GCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTCTG
AAAGCGACCAGGTGCTCGGCGTGGCAAGACTCGCAGCGAACCCGTAGAAAGCCATGCTCCAGCCGCCC
GCATTGGAGAAATTCTTCAAATTCCCGTTGCACATAGCCCGGCAATTCCTTTCCCTGCTCTGCCATAA
GCGCAGCGAATGCCGGGTAATACTCGTCAACGATCTGATAGAGAAGGGTTTGCTCGGGTCGGTGGCTC
TGGTAACGACCAGTATCCCGATCCCGGCTGGCCGTCCTGGCCGCCACATGAGGCATGTTCCGCGTCCT
TGCAATACTGTGTTTACATACAGTCTATCGCTTAGCGGAAAGTTCTTTTACCCTCAGCCGAAATGCCT
GCCGTTGCTAGACATTGCCAGCCAGTGCCCGTCACTCCCGTACTAACTGTCACGAACCCCTGCAATAA
CTGTCACGCCCCCCTGCAATAACTGTCACGAACCCCTGCAATAACTGTCACGCCCCCAAACCTGCAAA
CCCAGCAGGGGCGGGGGCTGGCGGGGTGTTGGAAAAATCCATCCATGATTATCTAAGAATAATCCACT
AGGCGCGGTTATCAGCGCCCTTGTGGGGCGCTGCTGCCCTTGCCCAATATGCCCGGCCAGAGGCCGGA
TAGCTGGTCTATTCGCTGCGCTAGGCTACACACCGCCCCACCGCTGCGCGGCAGGGGGAAAGGCGGGC
AAAGCCCGCTAAACCCCACACCAAACCCCGCAGAAATACGCTGGAGCGCTTTTAGCCGCTTTAGCGGC
CTTTCCCCCTACCCGAAGGGTGGGGGCGCGTGTGCAGCCCCGCAGGGCCTGTCTCGGTCGATCATTCA
GCCCGGCTCATCCTTCTGGCGTGGCGGCAGACCGAACAAGGCGCGGTCGTGGTCGCGTTCAAGGTACG
CATCCATTGCCGCCATGAGCCGATCCTCCGGCCACTCGCTGCTGTTCACCTTGGCCAAAATCATGGCC
CCCACCAGCACCTTGCGCCTTGTTTCGTTCTTGCGCTCTTGCTGCTGTTCCCTTGCCCGCTCCCGCTG
AATTTCGGCATTGATTCGCGCTCGTTGTTCTTCGAGCTTGGCCAGCCGATCCGCCGCCTTGTTGCTCC
CCTTAACCATCTTGACACCCCATTGTTAATGTGCTGTCTCGTAGGCTATCATGGAGGCACAGCGGCGG
CAATCCCGACCCTACTTTGTAGGGGAGGGCGCACTTACCGGTTTCTCTTCGAGAAACTGGCCTAACGG
CCACCCTTCGGGCGGTGCGCTCTCCGAGGGCCATTGCATGGAGCCGAAAAGCAAAAGCAACAGCGAGG
CAGCATGGCGATTTATCACCTTACGGCGAAAACCGGCAGCAGGTCGGGCGGCCAATCGGCCAGGGCCA
AGGCCGACTACATCCAGCGCGAAGGCAAGTATGCCCGCGACATGGATGAAGTCTTGCACGCCGAATCC
GGGCACATGCCGGAGTTCGTCGAGCGGCCCGCCGACTACTGGGATGCTGCCGACCTGTATGAACGCGC
CAATGGGCGGCTGTTCAAGGAGGTCGAATTTGCCCTGCCGGTCGAGCTGACCCTCGACCAGCAGAAGG
CGCTGGCGTCCGAGTTCGCCCAGCACCTGACCGGTGCCGAGCGCCTGCCGTATACGCTGGCCATCCAT
GCCGGTGGCGGCGAGAACCCGCACTGCCACCTGATGATCTCCGAGCGGATCAATGACGGCATCGAGCG
GCCCGCCGCTCAGTGGTTCAAGCGGTACAACGGCAAGACCCCGGAGAAGGGCGGGGCACAGAAGACCG
AAGCGCTCAAGCCCAAGGCATGGCTTGAGCAGACCCGCGAGGCATGGGCCGACCATGCCAACCGGGCA
TTAGAGCGGGCTGGCCACGACGCCCGCATTGACCACAGAACACTTGAGGCGCAGGGCATCGAGCGCCT
GCCCGGTGTTCACCTGGGGCCGAACGTGGTGGAGATGGAAGGCCGGGGCATCCGCACCGACCGGGCAG
ACGTGGCCCTGAACATCGACACCGCCAACGCCCAGATCATCGACTTACAGGAATACCGGGAGGCAATA
GACCATGAACGCAATCGACAGAGTGAAGAAATCCAGAGGCATCAACGAGTTAGCGGAGCAGATCGAAC
CGCTGGCCCAGAGCATGGCGACACTGGCCGACGAAGCCCGGCAGGTCATGAGCCAGACCAAGCAGGCC
AGCGAGGCGCAGGCGGCGGAGTGGCTGAAAGCCCAGCGCCAGACAGGGGCGGCATGGGTGGAGCTGGC
CAAAGAGTTGCGGGAGGTAGCCGCCGAGGTGAGCAGCGCCGCGCAGAGCGCCCGGAGCGCGTCGCGGG
GGTGGCACTGGAAGCTATGGCTAACCGTGATGCTGGCTTCCATGATGCCTACGGTGGTGCTGCTGATC
GCATCGTTGCTCTTGCTCGACCTGACGCCACTGACAACCGAGGACGGCTCGATCTGGCTGCGCTTGGT
GGCCCGATGAAGAACGACAGGACTTTGCAGGCCATAGGCCGACAGCTCAAGGCCATGGGCTGTGAGCG
CTTCGATATCGGCGTCAGGGACGCACCCACCGGCCAGATGATGAACCGGGAATGGTCAGCCGCCGAAG
TGCTCCAGAACACGCCATGGCTCAAGCGGATGAATGCCCAGGGCAATGACGTGTATATCAGGCCCGCC
GAGCAGGAGCGGCATGGTCTGGTGCTGGTGGACGACCTCAGCGAGTTTGACCTGGATGACATGAAAGC
CGAGGGCCGGGAGCCTGCCCTGGTAGTGGAAACCAGCCCGAAGAACTATCAGGCATGGGTCAAGGTGG
CCGACGCCGCAGGCGGTGAACTTCGGGGGCAGATTGCCCGGACGCTGGCCAGCGAGTACGACGCCGAC
CCGGCCAGCGCCGACAGCCGCCACTATGGCCGCTTGGCGGGCTTCACCAACCGCAAGGACAAGCACAC
CACCCGCGCCGGTTATCAGCCGTGGGTGCTGCTGCGTGAATCCAAGGGCAAGACCGCCACCGCTGGCC
CGGCGCTGGTGCAGCAGGCTGGCCAGCAGATCGAGCAGGCCCAGCGGCAGCAGGAGAAGGCCCGCAGG
CTGGCCAGCCTCGAACTGCCCGAGCGGCAGCTTAGCCGCCACCGGCGCACGGCGCTGGACGAGTACCG
CAGCGAGATGGCCGGGCTGGTCAAGCGCTTCGGTCATGACCTCAGCAAGTGCGACTTTATCGCCGCGC
AGAAGCTGGCCAGCCGGGGCCGCAGTGCCGAGGAAATCGGCAAGGCCATGGCCGAGGCCAGCCCAGCG
CTGGCAGAGCGCAAGCCCGGCCACGAAGCGGATTACATCGAGCGCACCGTCAGCAAGGTCATGGGTCT
GCCCAGCGTCCAGCTTGCGCGGGCCGAGCTGGCACGGGCACCGGCACCCCGCCAGCGAGGCATGGACA
GGGGCGGGCCAGATTTCAGCATGTAGTGCTTGCGTTGGTACTCACGCCTGTTATACTATGAGTACTCA
CGCACAGAAGGGGGTTTTATGGAATACGAAAAAAGCGCTTCAGGGTCGGTCTACCTGATCAAAAGTGA
CAAGGGCTATTGGTTGCCCGGTGGCTTTGGTTATACGTCAAACAAGGCCGAGGCTGGCCGCTTTTCAG
TCGCTGATATGGCCAGCCTTAACCTTGACGGCTGCACCTTGTCCTTGTTCCGCGAAGACAAGCCTTTC
GGCCCCGGCAAGTTTCTCGGTGACTGATATGAAAGACCAAAAGGACAAGCAGACCGGCGACCTGCTGG
CCAGCCCTGACGCTGTACGCCAAGCGCGATATGCCGAGCGCATGAAGGCCAAAGGGATGCGTCAGCGC
AAGTTCTGGCTGACCGACGACGAATACGAGGCGCTGCGCGAGTGCCTGGAAGAACTCAGAGCGGCGCA
GGGCGGGGGTAGTGACCCCGCCAGCGCCTAACCACCAACTGCCTGCAAAGGAGGCAATCAATGGCTAC
CCATAAGCCTATCAATATTCTGGAGGCGTTCGCAGCAGCGCCGCCACCGCTGGACTACGTTTTGCCCA
ACATGGTGGCCGGTACGGTCGGGGCGCTGGTGTCGCCCGGTGGTGCCGGTAAATCCATGCTGGCCCTG

CAACTGGCCGCACAGATTGCAGGCGGGCCGGATCTGCTGGAGGTGGGCGAACTGCCCACCGGCCCGGT
GATCTACCTGCCCGCCGAAGACCCGCCCACCGCCATTCATCACCGCCTGCACGCCCTTGGGGCGCACC
TCAGCGCCGAGGAACGGCAAGCCGTGGCTGACGGCCTGCTGATCCAGCCGCTGATCGGCAGCCTGCCC
AACATCATGGCCCCGGAGTGGTTCGACGGCCTCAAGCGCGCCGCCGAGGGCCGCCGCCTGATGGTGCT
GGACACGCTGCGCCGGTTCCACATCGAGGAAGAAAACGCCAGCGGCCCCATGGCCCAGGTCATCGGTC
GCATGGAGGCCATCGCCGCCGATACCGGGTGCTCTATCGTGTTCCTGCACCATGCCAGCAAGGGCGCG
GCCATGATGGGCGCAGGCGACCAGCAGCAGGCCAGCCGGGGCAGCTCGGTACTGGTCGATAACATCCG
CTGGCAGTCCTACCTGTCGAGCATGACCAGCGCCGAGGCCGAGGAATGGGGTGTGGACGACGACCAGC
GCCGGTTCTTCGTCCGCTTCGGTGTGAGCAAGGCCAACTATGGCGCACCGTTCGCTGATCGGTGGTTC
AGGCGGCATGACGGCGGGGTGCTCAAGCCCGCCGTGCTGGAGAGGCAGCGCAAGAGCAAGGGGGTGCC
CCGTGGTGAAGCCTAAGAACAAGCACAGCCTCAGCCACGTCCGGCACGACCCGGCGCACTGTCTGGCC
CCCGGCCTGTTCCGTGCCCTCAAGCGGGGCGAGCGCAAGCGCAGCAAGCTGGACGTGACGTATGACTA
CGGCGACGGCAAGCGGATCGAGTTCAGCGGCCCGGAGCCGCTGGGCGCTGATGATCTGCGCATCCTGC
AAGGGCTGGTGGCCATGGCTGGGCCTAATGGCCTAGTGCTTGGCCCGGAACCCAAGACCGAAGGCGGA
CGGCAGCTCCGGCTGTTCCTGGAACCCAAGTGGGAGGCCGTCACCGCTGAATGCCATGTGGTCAAAGG
TAGCTATCGGGCGCTGGCAAAGGAAATCGGGGCAGAGGTCGATAGTGGTGGGGCGCTCAAGCACATAC
AGGACTGCATCGAGCGCCTTTGGAAGGTATCCATCATCGCCCAGAATGGCCGCAAGCGGCAGGGGTTT
CGGCTGCTGTCGGAGTACGCCAGCGACGAGGCGGACGGGCGCCTGTACGTGGCCCTGAACCCCTTGAT
CGCGCAGGCCGTCATGGGTGGCGGCCAGCATGTGCGCATCAGCATGGACGAGGTGCGGGCGCTGGACA
GCGAAACCGCCCGCCTGCTGCACCAGCGGCTGTGTGGCTGGATCGACCCCGGCAAAACCGGCAAGGCT
TCCATAGATACCTTGTGCGGCTATGTCTGGCCGTCAGAGGCCAGTGGTTCGACCATGCGCAAGCGCCG
CAAGCGGGTGCGCGAGGCGTTGCCGGAGCTGGTCGCGCTGGGCTGGACGGTAACCGAGTTCGCGGCGG
GCAAGTACGACATCACCCGGCCCAAGGCGGCAGGCTGACCCCCCCCACTCTATTGTAAACAAGACATT
TTTATCTTTTATATTCAATGGCTTATTTTCCTGCTAATTGGTAATACCATGAAAAATACCATGCTCAG
AAAAGGCTTAACAATATTTTGAAAAATTGCCTACTGAGCGCTGCCGCACAGCTCCATAGGCCGCTTTC
CTGGCTTTGCTTCCAGATGTATGCTCTTCTGCTCCCGAACGCCAGCAAGACGTAGCCCAGCGCGTCGG
CCAGCTTGCAATTCGCGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCC
AGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGA
GAGTTGCAGCAAGCGGTCCACGTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGG
CGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCA
GCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTG
GGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTC
CCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCG
CCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCC
ACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATC
AAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGAT
AGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACG
CCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGC
GACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGC
CCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCC
CGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGC
GCTATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCAACGTAAATGCCGCTTCGCCTT
CGCGCGCGAATTGCAAGCTGATCCGGGCTTATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGC
AGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCG
CACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAG
CTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATACAATTTCACACAGGAA
ACAGAATTCGGCCCCCATAAAAAGTAAATTAAAATTTGGTGTTTTTTTTACAGTTTTACTATCTTCAG
CATATGCACATGGAACACCTCAAAATATTACTGATTTGTGTGCAGAATACCACAACACACAAATATAT
ACGCTAAATGATAAGATATTTTCGTATACAGAATCTCTAGCTGGAAAAAGAGAGATGGCTATCATTAC
TTTTAAGAATGGTGCAATTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTCACAAAAAAAAG
CGATTGAAAGGATGAAGGATACCCTGAGGATTGCATATCTTACTGAAGCTAAAGTCGAAAAGTTATGT
GTATGGAATAATAAAACGCCTCATGCGATTGCCGCAATTAGTATGGCAAACTAGTTTGCTTTAAAAGC
ATGTCTAATGCTAGGAACCTATATAACAACTACTGTACTTATACTAATGAGCCTTATGCTGCATTTGA
AAAGGCGGTAGAGGATGCAATACCGATCCTTAAACTGTAACACTATAACAGCTTCCACTACAGGGAGC
TGTTATAGCAAACAGAAAAAACTAAGCTAGGCTGGGGGGGCA

9.2.2 pSAB2.1A

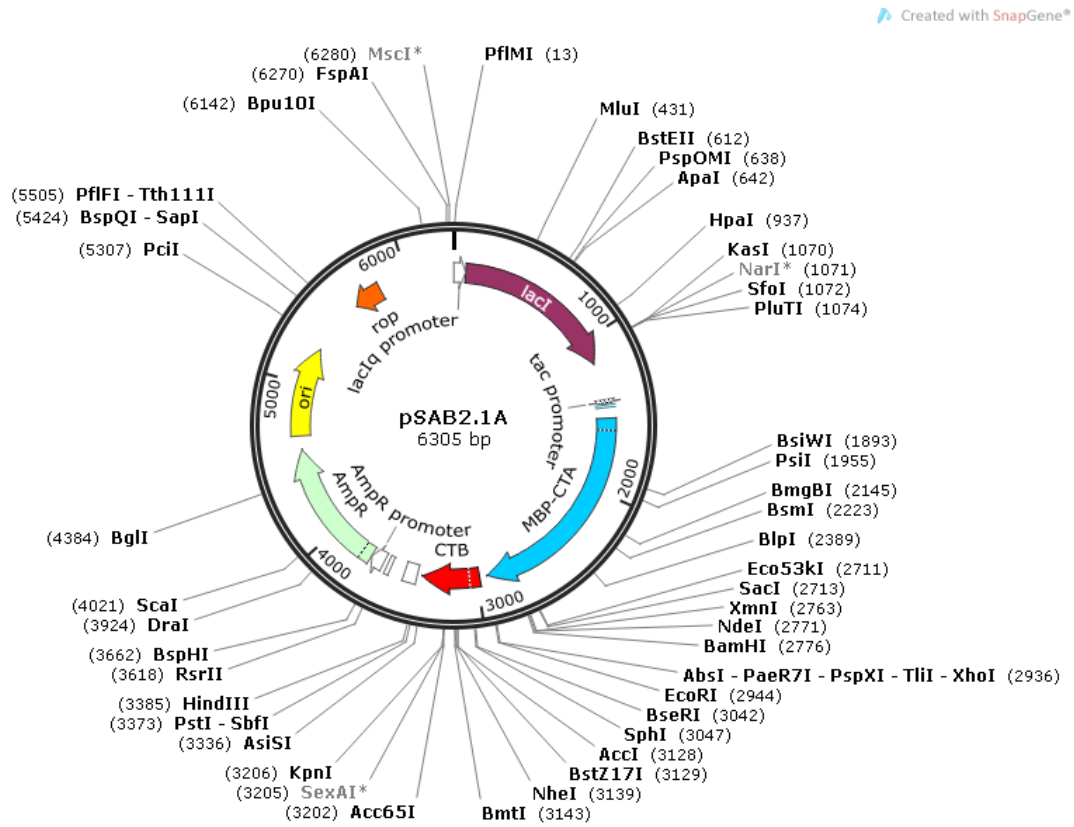


Fig 9.2 Plasmid map for pSAB2.1A

The MBP-AB_5 plasmid for expression in *E.coli* is pSAB2.1A (Fig 9.2). It is derived from pMAL-p5X obtained from Genscript. An MBP-CTA2 fusion was introduced along with the gene for CTB (1A), with the LTIIB efficient periplasmic targeting leader sequence. This resulted in plasmid pSAB2.1A (preparation of the plasmid performed by Mr James Ross).

Derivatives of plasmid pSAB2.1A include:
pSAB2.1T, pTRBAB5-D4N and pTRBAB5-G1S

Full sequence:

```
CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTC
AGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGAC
CGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGA
TGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATT
GGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGC
CGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGG
CGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGAT
GCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACC
CATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGG
GTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGC
TGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCAT
```

GTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCA
ACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATT
TCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACA
GGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGA
AGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACC
GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG
GCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCT
TATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTG
TGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTT
GCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGT
ATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCCAGTCCGTTTAGGTGTTTTC
ACGAGCAATTGACCAACAAGGACCATAGATTATGAAAATAAAAACAGGTGCACGCATCCTCGCATTAT
CCGCATTAACGACGATGATGTTTTCCGCCTCGGCTCTCGCCAAAATCGAAGAAGGTAAACTGGTAATC
TGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGG
AATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCG
ATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCT
GAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAA
CGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGC
CGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGC
GCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGC
GTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGG
GTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCA
GAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGA
CACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCG
TTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAA
AACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCT
GAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCGCCACTATGGAAAACGCCCAGAAAG
GTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAAC
GCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAA
CAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCACATATGGGATCCGAAAACC
TGTACTTTCAGGGTGGCGGTGATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTTGACGAATACCAA
TCTAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA
TGAATTATGACCTCGAGGTGAATTCACGAGCAATTGACCAACAAGGACCATAGATTATGAGCTTTAAG
AAAATTATCAAGGCATTTGTTATCATGGCTGCTTTGGTATCTGTTCAGGCGCATGCAGCTCCTCAAAA
TATTACTGATTTGTGCGCAGAATACCACAACACACAAATATATACGCTAAATGATAAGATCTTTTCGT
ATACAGAATCGCTAGCGGGAAAAAGAGAGATGGCTATCATTACTTTTAAGAATGGTGCAATTTTTCAA
GTAGAGGTACCAGGTAGTCAACATATAGATTCACAAAAAAAAGCGATTGAAAGGATGAAGGATACCCT
GAGGATTGCATATCTTACTGAAGCTAAAGTCGAAAAGTTATGTGTATGGAATAATAAAACGCCTCATG
CGATCGCCGCAATTAGTATGGCAAACTAAGTTTTCCCTGCAGGTAATTAAATAAGCTTCAAATAAAAC
GAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGT
AGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACG
CCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTC
TACAAACTCTTTCGGTCCGTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAA
TAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGC
CCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAA
AAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATC
CTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC
GGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACT
TGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGT
GCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGA
GCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAA
CTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAA
AGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCG
GTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTT
ATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTC
ACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTCCTTAGGA
CTGAGCGTCAACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG

CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG
ATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATAAGGTGCACTGTGACTGGG
TCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCA
TCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACC
GAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCT
GTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCC
ATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGG
GGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGT
TACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAG
GGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGAT
GCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAAC
CGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCG
CGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAG
GAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAAATT

9.2.3 pSAB2.2A

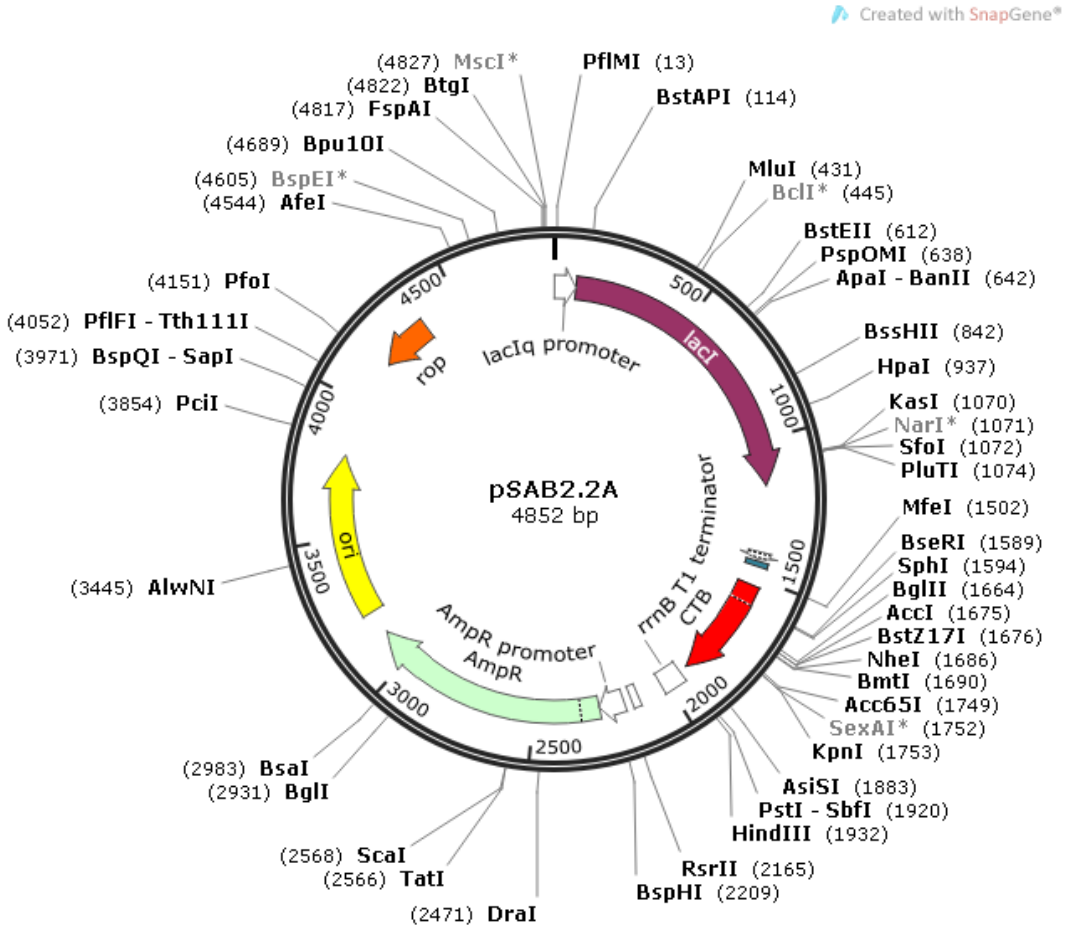


Fig 9.3 Plasmid map for pSAB2.2A

The CTB plasmid for expression in *E.coli* is pSAB2.2A (Fig 9.3). It is derived from pMAL-p5X obtained from Genscript. An MBP-CTA2 fusion was introduced along

with the gene for CTB (1A), with the LTIIB efficient periplasmic targeting leader sequence. This resulted in plasmid pSAB2.1A. The MBP-CTA2 gene was then removed to produce plasmid pSAB2.2A containing only the CTB gene (preparation of the plasmid performed by Mr James Ross).

Derivatives of plasmid pSAB2.2A include:
pSAB2.2T, pTRB-I47T, pSAB2.2-T1AW88K, pTRB-W88K, pTRB-W88E and pTRB-W88KQ61E

Full sequence:

CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTC
AGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGAC
CGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGA
TGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATT
GGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGC
CGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGG
CGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGAT
GCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACC
CATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGG
GTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGC
TGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCAT
GTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCA
ACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATT
TCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACA
GGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGA
AGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACC
GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG
GCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCT
TATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTG
TGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTT
GCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGT
ATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCCAGTCCGTTTAGGTGTTTTC
ACGAGCAATTGACCAACAAGGACCATAGATTATGAGCTTTAAGAAAATTATCAAGGCATTTGTTATCA
TGGCTGCTTTGGTATCTGTTCAGGCGCATGCAGCTCCTCAAAATATTACTGATTTGTGCGCAGAATAC
CACAACACACAAATATATACGCTAAATGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAAG
AGAGATGGCTATCATTACTTTTAAGAATGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACATA
TAGATTCACAAAAAAAAGCGATTGAAAGGATGAAGGATACCCTGAGGATTGCATATCTTACTGAAGCT
AAAGTCGAAAAGTTATGTGTATGGAATAATAAAACGCCTCATGCGATCGCCGCAATTAGTATGGCAAA
CTAAGTTTTCCCTGCAGGTAATTAAATAAGCTTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC
CTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATT
TGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAA
ATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTCGGTCCGTTGTTT
ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTT
TGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGC
ACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAAC
GTTTCCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGG
CAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA
AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACA
CTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATG
GGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCG
TGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC
TAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCG
GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCAT

TGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAA
CTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCA
GACCAAGTTTACTCATATATACTTTAGATTGATTTCCTTAGGACTGAGCGTCAACCCCGTAGAAAAGA
TCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCG
CTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAG
CAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTG
TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCG
TGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGG
TTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTAT
GAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACA
GGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCA
CCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCA
ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCC
CCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATC
TGTGCGGTATTTCACACCGCATATAAGGTGCACTGTGACTGGGTCATGGCTGCGCCCCGACACCCGCC
AACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCG
TCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAA
AGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAG
TTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTT
TGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAG
AGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACA
ACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATA
CAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAG
GGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCA
GGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAAC
CAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGC
CAGGACCCAACGCTGCCCGAAATT

9.3 DNA primers and parts

9.3.1 Primers for site-directed mutagenesis

Mutations are highlighted in red.

CTB mutation I47T

5′ CTT TTA AGA ATG GTG CAA **C**TT TTC AAG TAG AGG TAC C 3′
3′ GAA AAT TCT TAC CAC GTT **G**AA AAG GGC ATC TCC ATG G 5′

Predicted T_m = 62.2 °C

CTB mutation A1T

5′ GTT CAG GCG CAT GCA AC**T** CCT CAA AAT ATT ACT G 3′
3′ CAA GTC CGC GTA CGT TG**A** GGA GTT TTA TAA TGA C 5′

Predicted T_m = 62.2 °C

CTB mutation K88E

5′ CGA AAA GTT ATG TGT A**G**A GAA TAA TAA AAC GCC 3′
3′ GCT TTT CAA TAC ACA T**C**T CTT ATT ATT TTG CGG 5′

Predicted T_m = 56.5 °C

CTB mutation Q61E

```
5′ GTC AAC ATA TAG ATT CAG AAA AAA AAG CGA TTG AAA GG  3′
3′ CAG TTG TAT ATC TAA GTC TTT TTT TTC GCT AAC TTT CC  5′
```

Predicted T_m = 58.5 °C

CTA2 mutation G1S

```
5′ CCT GTA CTT TCA GAG TGG CGG TGA TG  3′
3′ GGA CAT GAA AGT CTC ACC GCC ACT AC  5′
```

Predicted T_m = 61.0 °C

9.3.2 DNA parts for gene synthesis

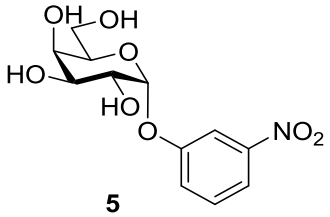
The sequences for the four DNA parts (Parts A, B, C and D) for the synthesis of the new CTA2 gene, containing five asparagine residues, are shown below. The NdeI restriction site is marked in blue and the XhoI restriction site is marked in purple.

```
A & C: 5′  GAT TTC ACA TAT GGG ATC CGA AAA CCT GTA CTT TCA AGG TGG
B & D: 3′                                                GT TCC ACC

A & C:     TGG CAA CAA C               CTC TAA GGT GAA GCG TCA GAT
B & D:     ACC GTT GTT GTT ATT GTT GAG ATT CCA CTT CGC AGT C

A & C:     ATT TTC AGG CTA TCA ATC TGA TAT TGA TAC CCA CAA CC
B & D:                          G ACT ATA ACT ATG GGT GTT GGC ATA

A & C:                                                  3′
B & D:     GTT CCT ACT TAA TAC TGG AGC TCC ACT TAA      5′
```

9.4 Other compounds used in the project

The following compounds which were used in the project were provided by colleagues in the Turnbull group or external collaborators. All literature compounds had analytical properties in agreement with those reported previously.

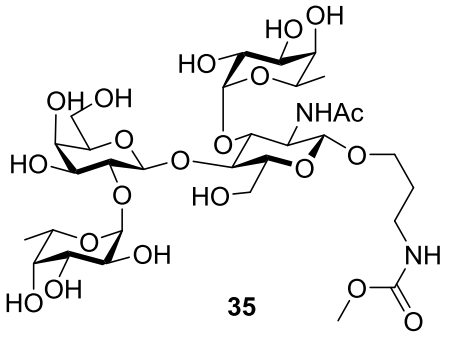
***meta*-nitrophenyl α-D-galactopyranoside (MNPG) 5**



5

Synthesised by Dr Ed Hayes.[191]

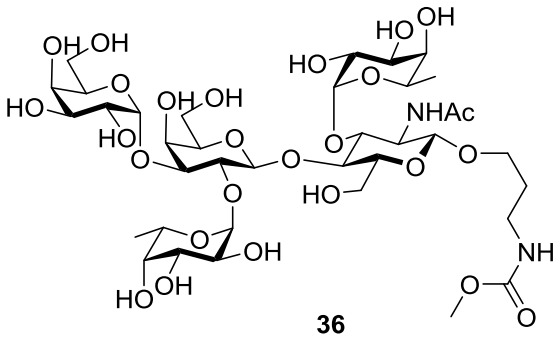
3-(Methyloxycarbonylamino)propyl α-L-fucopyranosyl-(1→2)-β-D-galactopyranosyl-(1→4)-[α-L-fucopyranosyl-(1→3)]-2-acetamido-2-deoxy-β-D-glucopyranoside 35



35

Synthesised by Dr Pintu Mandal[88]

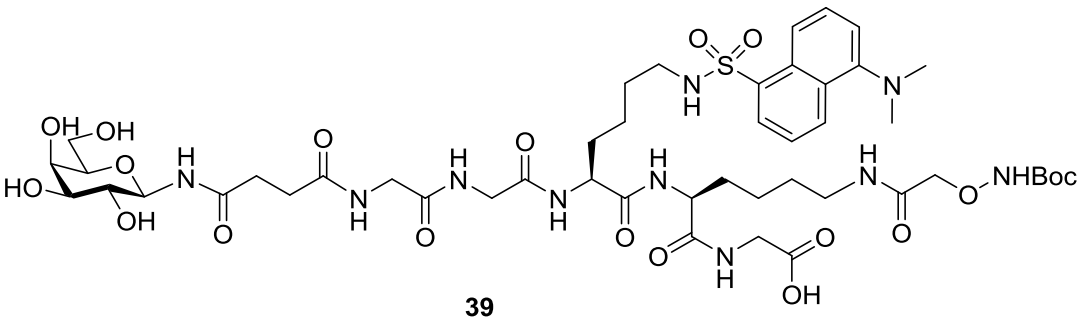
3-(Methyloxycarbonylamino)propyl) α-D-galactopyranosyl-(1→3)-[α-L-fucopyranosyl-(1→2)]-β-D-galactopyranosyl-(1→4)-[α-L-fucopyranosyl-(1→3)]-2-acetamido-2-deoxy-β-D-glucopyranoside 36



36

Synthesised by Dr Pintu Mandal.[88]

Galactose peptide ligand 39

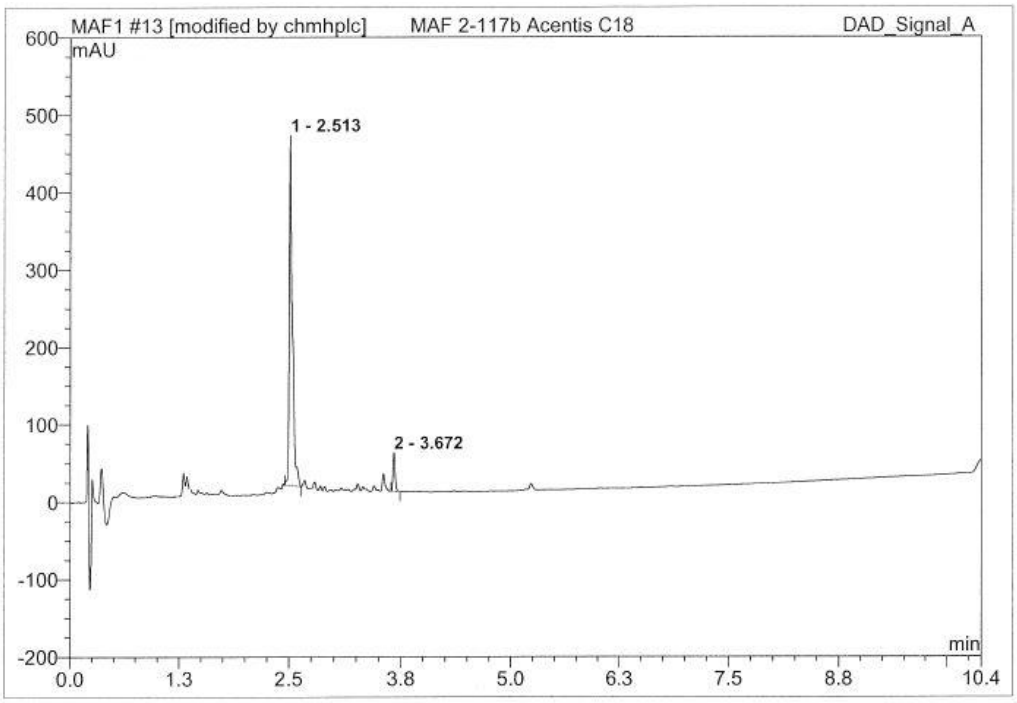


39

Synthesised by Dr Martin Fascione using SPPS.

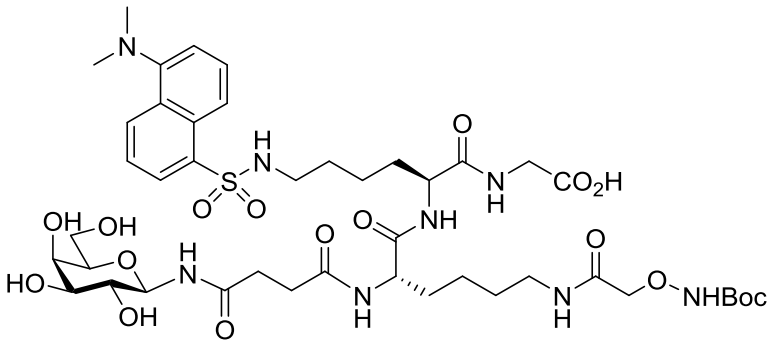
HRMS: Found $[M-H]^-$ 1111.4659, $C_{47}H_{71}N_{10}O_{19}S$ requires 1111.4623.

HPLC:



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	2.51	n.a.	452.029	18.382	92.58	n.a.	BMB
2	3.67	n.a.	49.939	1.474	7.42	n.a.	BMB
Total:			501.969	19.856	100.00	0.000	

Galactose peptide ligand 42

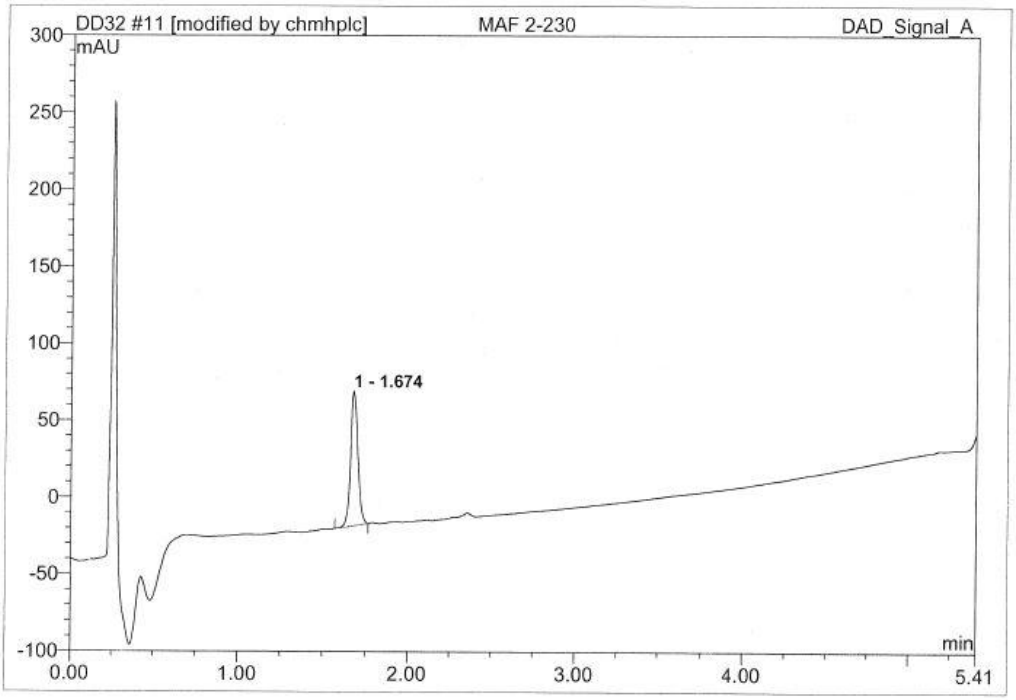


42

Synthesised by Dr Martin Fascione using SPPS.

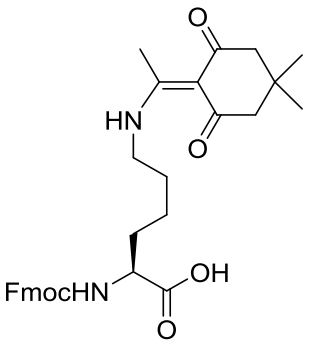
$[\alpha]_D^{25}$ -4.2 (*c* 0.12, MeOH); **HRMS**: Found $[M+H]^+$ 999.4328, $C_{43}H_{67}N_8O1_7S_1$ requires 999.4339; **IR** (v_{max}/cm^{-1}) 3368 (OH), 2912 (C=C), 1650 (C=O).

HPLC:



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	1.67	n.a.	87.725	4.413	100.00	n.a.	BMB*
Total:			87.725	4.413	100.00	0.000	

Fmoc-Lys(Dde)-OH 46[192]



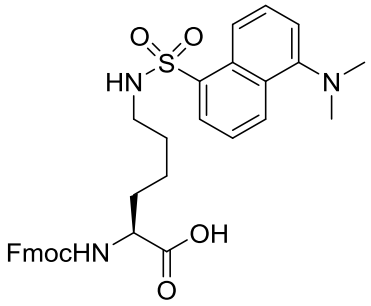
46

Synthesised by Dr Martin Fascione.

NMR: δ_H (300 MHz, $CDCl_3$): Dde: 8.56 (1H, d, 8.5 Hz, Ar-H), 8.37 (1H, d, 8.5 Hz, Ar-H), 8.21 (1H, d, 6.8 Hz, Ar-H), 7.58-7.54 (2H, m, Ar-H), 7.24 (d, 7.7 Hz, Ar-H), 2.87 (6H, s, $N(C\underline{H}_3)_2$) Fmoc: 7.80-7.77 (2H, m, Ar-H), 7.68 (1H, d, 7.7 Hz, Ar-H), 7.66 (1H, d, 7.7 Hz, Ar-H), 7.40-7.36 (2H, m, Ar-H), 7.32 (1H, d, 7.7 Hz, Ar-H), 7.29 (1H, d, 7.7 Hz, Ar-H), 4.38 (2H, dd, 6.8 Hz, 1.7 Hz, $C\underline{H}_2$-CH), 4.3 (1H, dd, 6.8 Hz, 6.8 Hz, CH_2-C<u>H</u>) Lys**:** 3.98-3.95 (1H, m, H_α), 1.64-1.20 (6H, m, $H_{\beta 2/3}$, $H_{\gamma 2/3}$, $H_{\delta 2/3}$), 2.85-2.82 (2H, m, $H_{\varepsilon 1/2}$); ^{13}C NMR (75 MHz, $CDCl_3$) δ 198.27, 174.81, 174.07, 156.20, 143.88, 141.26, 129.04, 128.23, 127.72, 127.06, 125.16, 120.00, 107.87,

67.04, 53.56, 52.44, 47.12, 43.29, 31.96, 30.16, 28.33, 22.44, 21.49, 18.21; **ESI-MS:** found [M+Na$^+$] 555.2 $C_{31}H_{36}N_2O_6Na$ requires 555.25.

Fmoc-Lys(dansyl)-OH 47[181]

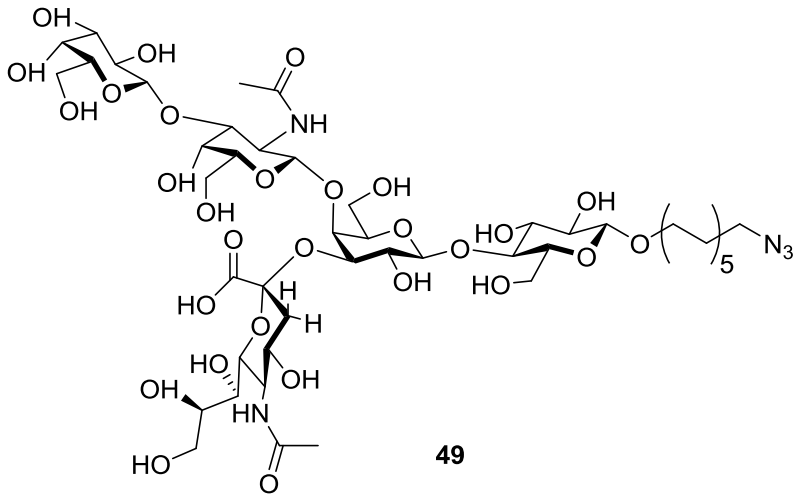


47

Synthesised by Dr Martin Fascione.

NMR: δ_H (500 MHz, CD_3OD): Dansyl: 8.56 (1H, d, 8.5 Hz, Ar-H), 8.37 (1H, d, 8.5 Hz, Ar-H), 8.21 (1H, d, 6.8 Hz, Ar-H), 7.58-7.54 (2H, m, Ar-H), 7.24 (d, 7.7 Hz, Ar-H), 2.87 (6H, s, N(C$\underline{H}_3$)$_2$) Fmoc: 7.80-7.77 (2H, m, Ar-H), 7.68 (1H, d, 7.7 Hz, Ar-H), 7.66 (1H, d, 7.7 Hz, Ar-H), 7.40-7.36 (2H, m, Ar-H), 7.32 (1H, d, 7.7 Hz, Ar-H), 7.29 (1H, d, 7.7 Hz, Ar-H), 4.38 (2H, dd, 6.8 Hz, 1.7 Hz, C$\underline{H}_2$-CH), 4.3 (1H, dd, 6.8 Hz, 6.8 Hz, CH$_2$-C$\underline{H}$) Lys: 3.98-3.95 (1H, m, H_α), 1.64-1.20 (6H, m, $H_{\beta2/3}$, $H_{\gamma2/3}$, $H_{\delta2/3}$), 2.85-2.82 (2H, m, $H_{\varepsilon1/2}$); **ESI-MS:** found [M+H$^+$] 602.2 $C_{33}H_{36}N_3O_6S$ requires 602.23.

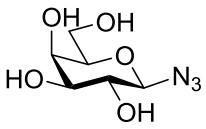
11-azidoundecyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-[5-acetamido-3,5-dideoxy-α-D-*glycero*-D-*galacto*-nonulopyranulosonyl-(2→3)]-β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside 49



49

Provided by the Zuilhof lab.[193]

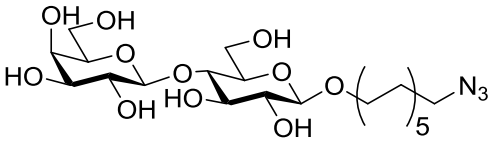
β-D-galactopyranosyl azide 50



50

Synthesised by Mr Dan Williamson.[182]

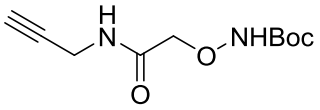
11-azidoundecyl β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside 51



51

Provided by the Zuilhof lab.[193]

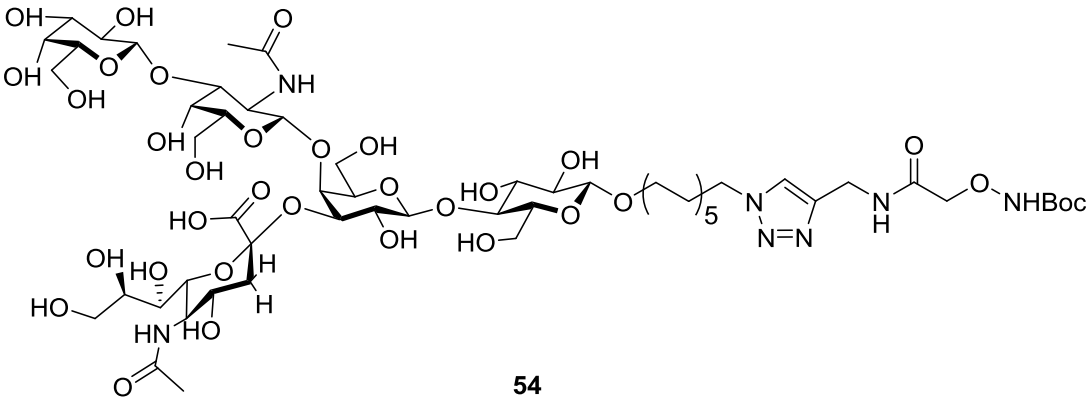
2-[*N*-(*t*-butoxycarbonyl)aminooxy]-*N*-(prop-2-ynyl)acetamide[185] 52



52

Synthesised by Dr Tom M^{c}Allister.

4-{*N*-[2-(*tert*-butoxycarbonylaminooxy)acetamidyl]-methyl}-1-{11-[β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-[5-acetamido-3,5-dideoxy-α-D-*glycero*-D-*galacto*-nonulopyranulosonyl-(2→3)]-β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyloxy]undecyl}-[1,2,3]triazole 54



54

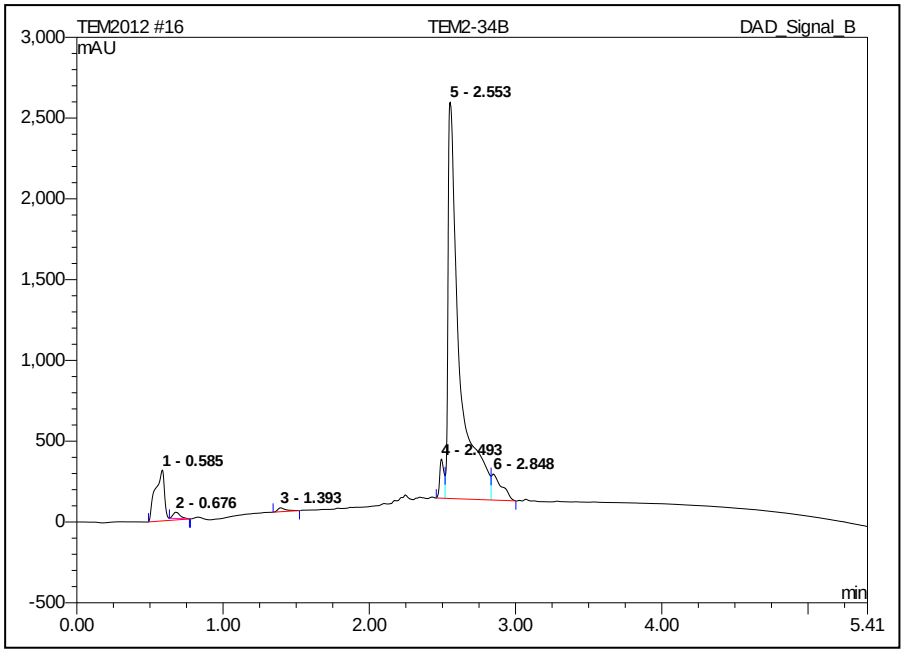
Synthesised by Dr Tom M^{c}Allister.

11-azidoundecyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-[5-acetamido-3,5-dideoxy-α-D-*glycero*-D-*galacto*-nonulopyranulosonyl-(2→3)]-β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside **49** (5.2 mg, 4.4 μmol) was dissolved in H_2O (100 μL) in a 1.5 mL Eppendorf tube. A solution of TBTA in THF (10 mg/mL, 70 μL, 0.7 mg, 1 μmol) was added to 2-[*N*-(*t*-butoxycarbonyl)aminooxy]-*N*-(prop-2-ynyl)acetamide **52** (2.0 mg, 9 μmol) and the resultant solution added to the Eppendorf. The vessels were rinsed with THF (2 × 115 μL) and the washes also added to the Eppendorf. $CuSO_4$ (1 mg/mL in H_2O, 100 μL, 0.1 mg, 0.6 μmol) and sodium-ascorbate (10 mg/mL in H_2O, 40 μL, 0.4 mg, 2 μmol) were added, followed by H_2O (60 μL), and the mixture was stirred at rt for 16 h. At this point, analysis of the reaction mixture by TLC and LCMS showed no product formation so further portions of $CuSO_4$ (10 mg/mL in H_2O, 5 μL, 0.5 mg, 3.1 μmol) and sodium-ascorbate (100 mg/mL in H_2O, 10 μL, 1 mg, 5 μmol) were added and stirring continued. After a further 16 h, analysis by LCMS suggested complete conversion to the expected product. The reaction was diluted with H_2O to a total volume of 12 mL and passed down a sep-pak column (pre-washed with dioxane and then equilibrated with H_2O). The column was then eluted with a step-wise gradient of aqueous dioxane (0-100% in 10% steps, 1.1 mL per wash). The column flow-through and 0-40% fractions were identified as containing the product by LCMS and TLC though the 30% and 40% fractions also contained TBTA. The flow-

through, 0, 10 and 20% fractions were combined and lyophilised to leave triazole **54** as a colourless amorphous solid (3.15 mg, 2.2 μmol, 50%)

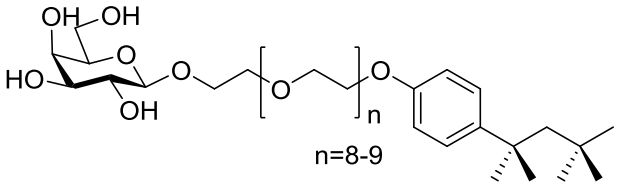
NMR: δ_H (500 MHz, D_2O); selected peaks: 7.4-7.9 (m, 1H, triazole H-5), 4.48 (d, 1H, J 7.6 Hz, GalNAc-C H-1), 4.46 (d, 1H, J 7.3 Hz, Gal-D H-1), 4.41 (d, 1H, J 8.0 Hz, Gal-B H-1), 4.33-4.35 (m, 2H, Glc-A H-1, Gal-B H-4), 2.60 (m, 1H, αNeu5Ac H-3eq), 1.97 (s, 3H, Neu5Ac CH_3), 1.94 (s, 3H, GalNAc-C CH_3), 1.87 (m, 1H, αNeu5Ac H-3ax); 1.34 (s, 9H, Boc); **HRMS**: found $[M+Na]^+$ 1444.6143, $C_{58}H_{99}N_7NaO_{33}$ requires 1444.6176.

HPLC:



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %
1	0.58	n.a.	314.805	23.766	8.88
2	0.68	n.a.	38.331	2.145	0.80
3	1.39	n.a.	23.906	1.546	0.58
4	2.49	n.a.	241.931	8.012	2.99
5	2.55	n.a.	2454.481	219.092	81.83
6	2.85	n.a.	161.202	13.188	4.93
Total:			3234.656	267.749	100.00

β-D-galactopyranosyl Triton X-100 57



57

Synthesised by Dr Martin Fascione

Trimethylsilyl trifluoromethanesulfonic anhydride (33 µL, 1.788 mmol) was added to a solution of TX-100 **55** (174 µL, 0.298 mmol), acetylated α-galactosyl trichloroacetimidate (440 mg, 0.896 mmol) and 4 Å MS (440 mg) in dichloromethane (2 mL) at −78 °C. The reaction mixture was warmed to r.t. over 30 min, and stirred for 1 h before quenching with aq. $NaHCO_3$ (2 mL), diluted with dichloromethane (10 mL) and washed with further aq. $NaHCO_3$ (10 mL) and aq. NaCl (10 mL), before being concentrated. The resulting crude residue was dissolved in methanol (5 mL) and sodium methoxide (39 mg, 0.716 mmol) was added, the reaction mixture was stirred for 1 h 30 min and then neutralised with Amberlite IRC 50 H^+ resin, filtered and concentrated. The crude product was purified by gel filtration (Sephadex LH-20 in methanol, flow rate 0.37 mL/min, 12 min fractions) and fractions 30-40 combined and concentrated to afford **57** (120 mg, 51% recovery, ~80-85% galactosylated).

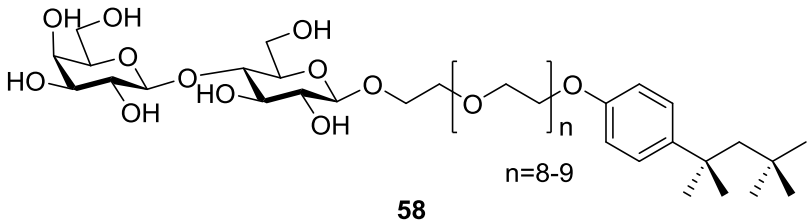
Selected peaks for TX-100 55:

NMR: δ_H (500 MHz, $CDCl_3$) 7.25 (d, *J* = 8.7 Hz, 2H, ArH), 6.82 (d, *J* = 8.7 Hz, 2H, ArH), 4.11 (t, *J* = 5.0 Hz, 2H), 3.84 (t, *J* = 5.0 Hz, 2H), 3.72 (m, 5H), 1.33 (s, 6H, CH_3), 0.71 (s, 9H, CH_3).

Selected new peaks for 57:

NMR: δ_H (500 MHz, $CDCl_3$) δ 4.29 (d, $J_{1,2}$ = 7.4 Hz, 1H (80-85%, measured against tert-butyl peaks at δ 0.71), 4.06 – 3.96 (m, 2H), 3.52 (m, 1H);

β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyl Triton X-100 58



58

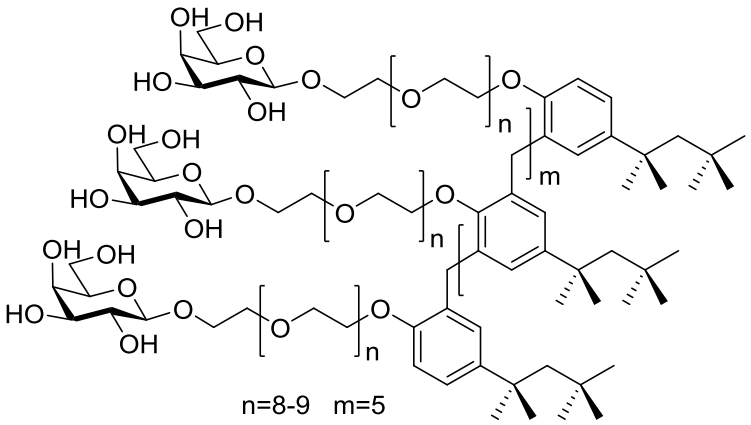
Synthesised by Dr Martin Fascione.

Trimethylsilyl trifluoromethanesulfonic anhydride (33 μL, 1.788 mmol) was added to a solution of TX-100 **55** (174 μL, 0.298 mmol), acetylated α-lactosyl trichloroacetimidate (660 mg, 0.890 mmol) and 4 Å MS (400 mg) in dichloromethane (2 mL) at −78 °C. The reaction mixture was warmed to r.t. over 30 min before quenching with aq. $NaHCO_3$ (2 mL), diluted with dichloromethane (10 mL) and washed with further aq. $NaHCO_3$ (10 mL) and aq. NaCl (10 mL), before being concentrated. The resulting crude residue was dissolved in 2:1:1 methanol-tetrahydrofuran-water (8 mL) and sodium methoxide (100 mg, 1.84 mmol) was added, the reaction mixture was stirred for 15 h and then neutralised with Amberlite IRC $50H^+$ resin, filtered and concentrated. The crude product was purified by gel filtration (Sephadex LH-20 in water, flow rate 0.35 mL/min, 12 min fractions) and fractions 56-70 combined and lyophilized to afford **58** (100 mg, 35% recovery, ~60-65% lactosylated, ~17% orthosester);

Selected peaks for 58:

NMR: δ_H (500 MHz, MeOD) 5.16 (d, *J* = 3.3 Hz, H-1α orthoester (17% of 1H)), 4.52 (d, *J* = 7.8 Hz, H-1β orthoester (17% of 1H)), 4.45 (d, *J* = 7.6 Hz, H-1β lactoside (60-65% of 1H)).

β-D-galactopyranosyl Tyloxapol 59



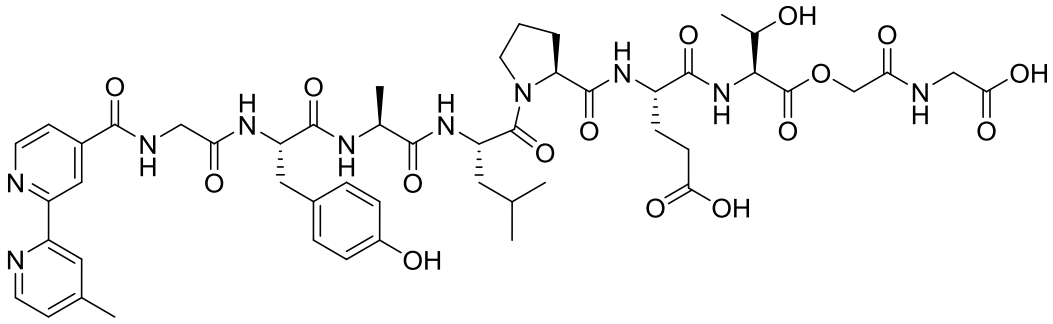
59

Synthesised by Dr Martin Fascione.

Trimethylsilyl trifluoromethanesulfonic anhydride (15 µL, 0.815 mmol) was added to a solution of Tyloxapol **56** (288 µL, 0.068 mmol, based on average m.w.),[167] acetylated α-galactosyl trichloroacetimidate (200 mg, 0.407 mmol) and 4 Å MS (200 mg) in dichloromethane (1 mL) at −78 °C. The reaction mixture was warmed to r.t. over 40 min before quenching with aq. $NaHCO_3$ (1 mL), diluted with dichloromethane (10 mL) and washed with further aq. $NaHCO_3$ (10 mL) and aq. NaCl (10 mL), before being concentrated. The resulting crude residue was dissolved in methanol (2 mL) and sodium methoxide (19.5 mg, 0.358 mmol) was added, the reaction mixture was stirred for 3 h and then neutralised with Amberlite IRC $50H^+$ resin, filtered and concentrated. The crude product was purified by gel filtration (Sephadex LH-20 in methanol, flow rate 0.4 mL/min, 12 min fractions) and fractions 18-24 combined and concentrated to afford **59** (298 mg, 94% recovery, ~78% galactosylated) as a colourless syrup; Selected anomeric peak for **59:**

NMR: δ_H (500 MHz, $CDCl_3$) 4.31 (m, H-1β (78% of 1H)).

Bpy depsipeptide 60



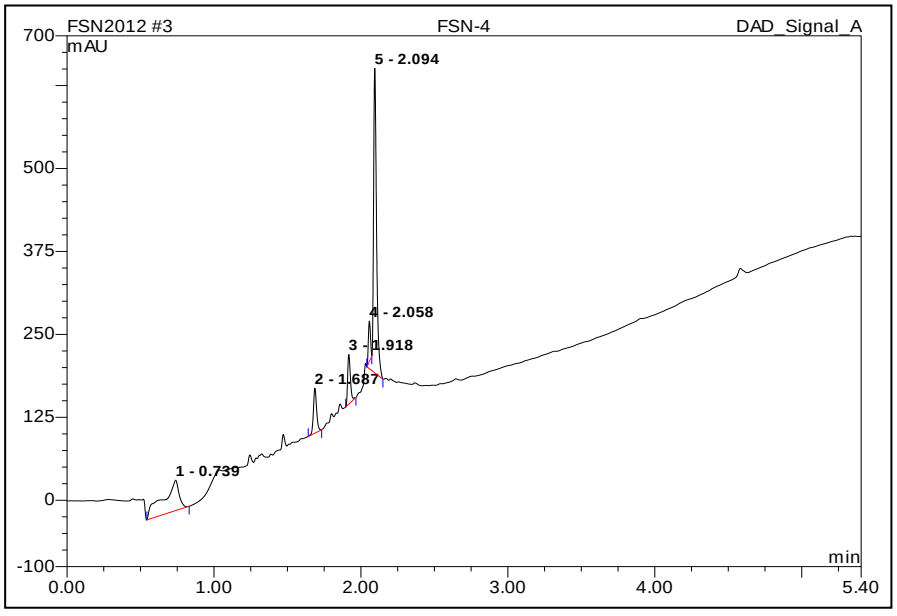
60

Synthesised by Miss Fiona Ng.

Synthesised by SPPS. A colourless lyophilisate yielded **60** (44 mg, 38%).

HRMS: Found [M+Cu^{2+}-H^{+}] 1122.3714, $C_{50}H_{63}CuN_{10}O_{16}$ requires 1122.3711.

HPLC:



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %
1	0.74	n.a.	45.774	5.924	27.72
2	1.69	n.a.	67.764	1.689	7.90
3	1.92	n.a.	74.887	1.621	7.59
4	2.06	n.a.	60.962	1.006	4.71
5	2.09	n.a.	458.796	11.129	52.08
Total:			708.184	21.370	100.00

10 References

[1] W. B. Turnbull (wbturnbull). "@BRSM_blog @TRBranson Keep running! I think the bear is tiring…", 23 September 2013, 9.48 a.m. Tweet.

[2] A. E. Rawlings, J. P. Bramble, S. S. Staniland, *Soft Matter* **2012**, *8*, 2675.

[3] http://syntheticbiology.org/, **Access Date 27/09/2013**.

[4] S. Leduc, *La Biologie Synthétique, étude de biophysique*, Poinat, A. Ed., Paris, **1912**.

[5] http://parts.igem.org/Main_Page, **Access Date 27/09/2013**.

[6] http://igem.org/Main_Page, **Access Date 27/09/2013**.

[7] http://www.rcsb.org/pdb/home/home.do, **Access Date 27/09/2013**.

[8] http://www.bbc.co.uk/programmes/b01b45zh, **Access Date 27/09/2013**.

[9] http://www.synbioproject.org/, **Access Date 27/09/2013**.

[10] http://www.parliament.uk/documents/post/postpn298.pdf, **Access Date 27/09/2013**.

[11] http://www.bristol.ac.uk/scn/syntheticbiology/, **Access Date 27/09/2013**.

[12] http://www.youtube.com/watch?v=Of289G6l2_A, **Access Date 27/09/2013**.

[13] K. Channon, E. H. C. Bromley, D. N. Woolfson, *Curr. Opin. Struct. Biol.* **2008**, *18*, 491.

[14] P. Ball, *Nanotechnol.* **2005**, *16*, R1.

[15] S. G. Zhang, *Nat. Biotechnol.* **2003**, *21*, 1171.

[16] http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1987/, **Access Date 27/09/2013**.

[17] J.-M. Lehn, *Angew. Chem. Int. Ed.* **1990**, *29*, 1304.

[18] C. A. Hunter, *Chem. Soc. Rev.* **1994**, *23*, 101.

[19] M. Fujita, *Chem. Soc. Rev.* **1998**, *27*, 417.

[20] D. B. Amabilino, J. F. Stoddart, *Chem. Rev.* **1995**, *95*, 2725.

[21] G. M. Whitesides, M. Boncheva, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 4769.

[22] P. G. A. Janssen, A. Ruiz-Carretero, D. Gonzalez-Rodriguez, E. W. Meijer, A. Schenning, *Angew. Chem. Int. Ed.* **2009**, *48*, 8103.

[23] J. Chen, N. C. Seeman, *Nature* **1991**, *350*, 631.

[24] N. C. Seeman, *Nature* **2003**, *421*, 427.

[25] P. W. K. Rothemund, *Nature* **2006**, *440*, 297.

[26] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S.

Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, J. Kjems, *Nature* **2009**, *459*, 73.

[27] Y. He, T. Ye, M. Su, C. Zhang, A. E. Ribbe, W. Jiang, C. D. Mao, *Nature* **2008**, *452*, 198.

[28] I. Severcan, C. Geary, A. Chworos, N. Voss, E. Jacovetty, L. Jaeger, *Nat. Chem.* **2010**, *2*, 772.

[29] E. Stulz, *Chem. Eu. J.* **2012**, *18*, 4456.

[30] N. C. Seeman, *Nano Lett.* **2010**, *10*, 1971.

[31] C. M. Dobson, *Nature* **2003**, *426*, 884.

[32] S. G. Zhang, D. M. Marini, W. Hwang, S. Santoso, *Curr. Opin. Chem. Biol.* **2002**, *6*, 865.

[33] E. H. C. Bromley, K. Channon, E. Moutevelis, D. N. Woolfson, *ACS Chem. Biol.* **2008**, *3*, 38.

[34] C. T. Armstrong, A. L. Boyle, E. H. C. Bromley, Z. N. Mahmoud, L. Smith, A. R. Thomson, D. N. Woolfson, *Faraday Discuss* **2009**, *143*, 305.

[35] A. L. Boyle, E. H. C. Bromley, G. J. Bartlett, R. B. Sessions, T. H. Sharp, C. L. Williams, P. M. G. Curmi, N. R. Forde, H. Linke, D. N. Woolfson, *J. Am. Chem. Soc.* **2012**.

[36] H. Gradišar, S. Božič, T. Doles, D. Vengust, I. Hafner-Bratkovič, A. Mertelj, B. Webb, A. Šali, S. Klavžar, R. Jerala, *Nat Chem Biol* **2013**, *advance online publication*.

[37] J. Hardy, D. J. Selkoe, *Science* **2002**, *297*, 353.

[38] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, D. Eisenberg, *Nature* **2005**, *435*, 773.

[39] Y. B. Lim, M. Lee, *Journal of Materials Chemistry* **2011**, *21*, 11680.

[40] D. Arslan, M. Legendre, V. Seltzer, C. Abergel, J. M. Claverie, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17486.

[41] D. T. Bong, T. D. Clark, J. R. Granja, M. R. Ghadiri, *Angew. Chem. Int. Ed.* **2001**, *40*, 988.

[42] J. E. Padilla, C. Colovos, T. O. Yeates, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2217.

[43] T. O. Yeates, J. E. Padilla, *Curr. Opin. Struct. Biol.* **2002**, *12*, 464.

[44] J. C. Sinclair, K. M. Davies, C. Venien-Bryan, M. E. M. Noble, *Nat. Nanotechnol.* **2011**, *6*, 558.

[45] N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates, D. Baker, *Science* **2012**, *336*, 1171.

[46] D. Papapostolou, S. Howorka, *Mol. Biosyst.* **2009**, *5*, 723.

[47] S. Howorka, *Curr. Opin. Biotechnol.* **2011**, *22*, 485.

[48] D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **1993**, *262*, 1019.

[49] J. L. Czlapinski, M. W. Schelle, L. W. Miller, S. T. Laughlin, J. J. Kohler, V. W. Cornish, C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130*, 13186.

[50] A. Fegan, B. White, J. C. T. Carlson, C. R. Wagner, *Chem. Rev.* **2010**, *110*, 3315.

[51] J. E. Gestwicki, C. W. Cairo, L. E. Strong, K. A. Oetjen, L. L. Kiessling, *J. Am. Chem. Soc.* **2002**, *124*, 14922.

[52] A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem. Int. Ed.* **2006**, *45*, 7581.

[53] C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem. Int. Ed.* **2008**, *47*, 10030.

[54] N. Stephanopoulos, M. B. Francis, *Nat. Chem. Biol.* **2011**, *7*, 876.

[55] D. Sicard, S. Cecioni, M. Iazykov, Y. Chevolot, S. E. Matthews, J.-P. Praly, E. Souteyrand, A. Imberty, S. Vidal, M. Phaner-Goutorbe, *Chem. Commun.* **2011**, *47*, 9483.

[56] S. Burazerovic, J. Gradinaru, J. Pierron, T. R. Ward, *Angew. Chem. Int. Ed.* **2007**, *46*, 5510.

[57] J. C. T. Carlson, S. S. Jena, M. Flenniken, T. F. Chou, R. A. Siegel, C. R. Wagner, *J. Am. Chem. Soc.* **2006**, *128*, 7630.

[58] M. M. C. Bastings, T. F. A. de Greef, J. L. J. van Dongen, M. Merkx, E. W. Meijer, *Chem. Sci.* **2010**, *1*, 79.

[59] H. Kitagishi, K. Oohora, H. Yamaguchi, H. Sato, T. Matsuo, A. Harada, T. Hayashi, *J. Am. Chem. Soc.* **2007**, *129*, 10326.

[60] H. Kitagishi, K. Oohora, T. Hayashi, *Biopolymers* **2009**, *91*, 194.

[61] K. Oohora, A. Onoda, H. Kitagishi, H. Yamaguchi, A. Harada, T. Hayashi, *Chem. Sci.* **2011**.

[62] K. Oohora, S. Burazerovic, A. Onoda, Y. M. Wilson, T. R. Ward, T. Hayashi, *Angew. Chem. Int. Ed.* **2012**, n/a.

[63] P. Ringler, G. E. Schulz, *Science* **2003**, *302*, 106.

[64] H. Kitagishi, Y. Kakikura, H. Yamaguchi, K. Oohora, A. Harada, T. Hayashi, *Angew. Chem. Int. Ed.* **2009**, *48*, 1271.

[65] A. Onoda, Y. Ueya, T. Sakamoto, T. Uematsu, T. Hayashi, *Chem. Commun.* **2010**, *46*, 9107.

[66] Y. Mori, K. Minamihata, H. Abe, M. Goto, N. Kamiya, *Org. Biomol. Chem.* **2011**, *9*, 5641.

[67] M. M. Ma, D. Bong, *Org. Biomol. Chem.* **2011**, *9*, 7296.

[68] N. Dotan, D. Arad, F. Frolow, A. Freeman, *Angew. Chem. Int. Ed.* **1999**, *38*, 2363.

[69] B. Bilgiçer, D. T. Moustakas, G. M. Whitesides, *J. Am. Chem. Soc.* **2007**, *129*, 3722.

[70] T. R. Branson, W. B. Turnbull, *Chem. Soc. Rev.* **2013**, *42*, 4613.

[71] P. I. Kitov, D. R. Bundle, *J. Am. Chem. Soc.* **2003**, *125*, 16271.

[72] Y. C. Lee, R. T. Lee, *Acc. Chem. Res.* **1995**, *28*, 321.

[73] J. D. Badjic, A. Nelson, S. J. Cantrill, W. B. Turnbull, J. F. Stoddart, *Acc. Chem. Res.* **2005**, *38*, 723.

[74] M. Mammen, S.-K. Choi, G. M. Whitesides, *Angew. Chem. Int. Ed.* **1998**, *37*, 2754.

[75] J. C. Sacchettini, L. G. Baum, C. F. Brewer, *Biochemistry* **2001**, *40*, 3009.

[76] R. J. Pieters, *Med. Res. Rev.* **2007**, *27*, 796.

[77] M. Kosek, C. Bern, R. L. Guerrant, *Bulletin of the World Health Organization* **2003**, *81*, 197.

[78] L. de Haan, T. R. Hirst, *Mol. Membr. Biol.* **2004**, *21*, 77.

[79] D. Vanden Broeck, C. Horvath, M. J. S. De Wolf, *Int. J. Biochem. Cell Biol.* **2007**, *39*, 1771.

[80] M. Muniesaa, J. A. Hammerlb, S. Hertwigb, B. Appelb, H. Brüssowc, *Appl Environ Microbiol.* **2012**, *78*, 4065.

[81] E. A. Merritt, W. G. J. Hol, *Curr. Op. in Struct. Bio.* **1995**, *5*, 165.

[82] K. Sandvig, B. van Deurs, *Gene Ther.* **2005**, *12*, 865.

[83] Y. Endo, K. Tsurugi, T. Yutsudo, Y. Takeda, T. Ogasawara, K. Igarashi, *European Journal of Biochemistry* **1988**, *171*, 45.

[84] E. A. Merritt, P. Kuhn, S. Sarfaty, J. L. Erbe, R. K. Holmes, W. G. J. Hol, *Journal of Molecular Biology* **1998**, *282*, 1043.

[85] H. Ling, A. Boodhoo, B. Hazes, M. D. Cummings, G. D. Armstrong, J. L. Brunton, R. J. Read, *Biochemistry* **1998**, *37*, 1777.

[86] Å. Holmner, G. Askarieh, M. Ökvist, U. Krengel, *Journal of Molecular Biology* **2007**, *371*, 754.

[87] J. E. Heggelund, E. Haugen, B. Lygren, A. Mackenzie, Å. Holmner, F. Vasile, J. J. Reina, A. Bernardi, U. Krengel, *Biochemical and Biophysical Research Communications* **2012**, *418*, 731.

[88] P. K. Mandal, T. R. Branson, E. D. Hayes, J. F. Ross, J. A. Gavín, A. H. Daranas, W. B. Turnbull, *Angew. Chem. Int. Ed.* **2012**, *51*, 5143.

[89] W. B. Turnbull, B. L. Precious, S. W. Homans, *J. Am. Chem. Soc.* **2004**, *126*, 1047.

[90] P. M. St. Hilaire, M. K. Boyd, E. J. Toone, *Biochemistry* **1994**, *33*, 14452.

[91] A. A. Wolf, M. G. Jobling, D. E. Saslowsky, E. Kern, K. R. Drake, A. K. Kerworthy, R. K. Holmes, W. I. Lencer., *Infect. Immun.* **2008**, 1476.

[92] E. D. Hayes, W. B. Turnbull, in *Synthesis and Biological Applications of Multivalent Glycoconjugates* (Eds.: O. Renaudet, N. Spinelli), Bentham Science Publishers, **2011**, pp. 78.

[93] D. Arosio, S. Baretti, S. Cattaldo, D. Potenza, A. Bernardi, *Bioorganic & Medicinal Chemistry Letters* **2003**, *13*, 3831.

[94] J. C. Pickens, E. A. Merritt, M. Ahn, C. Verlinde, W. G. J. Hol, E. K. Fan, *Chem. Biol.* **2002**, *9*, 215.

[95] M. E. Ivarsson, J.-C. Leroux, B. Castagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 4024.

[96] C. L. Schengrund, N. J. Ringler, *J. Biol. Chem.* **1989**, *264*, 13233.

[97] H. Dohi, Y. Nishida, M. Mizuno, M. Shinkai, T. Kobayashi, T. Takeda, H. Uzawa, K. Kobayashi, *Bioorg. Med. Chem.* **1999**, *7*, 2053.

[98] J. M. Gargano, T. Ngo, J. Y. Kim, D. W. K. Acheson, W. J. Lees, *J. Am. Chem. Soc.* **2001**, *123*, 12909.

[99] M. Watanabe, K. Matsuoka, E. Kita, K. Igai, N. Higashi, A. Miyagawa, T. Watanabe, R. Yanoshita, Y. Samejima, D. Terunuma, Y. Natori, K. Nishikawa, *J.Inf. Dis.* **2004**, *189*, 360.

[100] P. Neri, S. I. Nagano, S. Yokoyama, H. Dohi, K. Kobayashi, T. Miura, T. Inazu, T. Sugiyama, Y. Nishida, H. Mori, *Microbiol. Immun.* **2007**, *51*, 581.

[101] B. D. Polizzotti, K. L. Kiick, *Biomacromolecules* **2005**, *7*, 483.

[102] S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton, M. I. Gibson, *Angew. Chem., Int. Ed.* **2012**, n/a.

[103] S. Liu, K. L. Kiick, *Macromolecules* **2008**, *41*, 764.

[104] B. D. Polizzotti, R. Maheshwari, J. Vinkenborg, K. L. Kiick, *Macromolecules* **2007**, *40*, 7103.

[105] R. Maheshwari, E. A. Levenson, K. L. Kiick, *Macromol.Biosci.* **2010**, *10*, 68.

[106] H.-A. Tran, P. I. Kitov, E. Paszkiewicz, J. M. Sadowska, D. R. Bundle, *Org. Bio. Chem.* **2011**, *9*, 3658.

[107] W. B. Turnbull, J. F. Stoddart, *Rev. Mol. Biotech.* **2002**, *90*, 231.

[108] Y. M. Chabre, R. Roy, in *Advances in Carbohydrate Chemistry and Biochemistry, Vol. Volume 63* (Ed.: H. Derek), Academic Press, **2010**, pp. 165.

[109] J. Thompson, C.-L. Schengrund, *Glycoconj. J.* **1997**, *14*, 837.

[110] J. P. Thompson, C.-L. Schengrund, *Biochem. Pharm.* **1998**, *56*, 591.

[111] R. J. Pieters, *Org. Bio. Chem.* **2009**, *7*, 2013.

[112] I. Vrasidas, Nico J. de Mol, Rob M. J. Liskamp, Roland J. Pieters, *Eur. J. Org. Chem.* **2001**, *2001*, 4685.

[113] D. Arosio, I. Vrasidas, P. Valentini, R. M. J. Liskamp, R. J. Pieters, A. Bernardi, *Org. Bio. Chem.* **2004**, *2*, 2113.

[114] H. M. Branderhorst, R. M. J. Liskamp, G. M. Visser, R. J. Pieters, *Chem. Commun.* **2007**, 5043.

[115] A. V. Pukin, H. M. Branderhorst, C. Sisu, C. A. G. M. Weijers, M. Gilbert, R. M. J. Liskamp, G. M. Visser, H. Zuilhof, R. J. Pieters, *ChemBioChem* **2007**, *8*, 1500.

[116] C. Sisu, A. J. Baron, H. M. Branderhorst, S. D. Connel, C. Weijers, R. de Vries, E. D. Hayes, A. V. Pukin, M. Gilbert, R. J. Pieters, H. Zuilhof, G. M. Visser, W. B. Turnbull, *ChemBioChem* **2009**, *10*, 329.

[117] P. I. Kitov, J. M. Sadowska, G. Mulvey, G. D. Armstrong, H. Ling, N. S. Pannu, R. J. Read, D. R. Bundle, *Nature* **2000**, *403*, 669.

[118] E. Fan, Z. Zhang, W. E. Minke, Z. Hou, C. L. M. J. Verlinde, W. G. J. Hol, *J. Am. Chem. Soc.* **2000**, *122*, 2663.

[119] R. H. Kramer, J. W. Karpen, *Nature* **1998**, *395*, 710.

[120] Z. Zhang, J. C. Pickens, W. G. J. Hol, E. Fan, *Org. Lett.* **2004**, *6*, 1377.

[121] J. Garcia-Hartjes, S. Bernardi, C. A. G. M. Weijers, T. Wennekes, M. Gilbert, F. Sansone, A. Casnati, H. Zuilhof, *Org. Biomol. Chem.* **2013**, *11*, 4340.

[122] M. Mattarella, J. Garcia-Hartjes, T. Wennekes, H. Zuilhof, J. S. Siegel, *Org. Biomol. Chem.* **2013**, *11*, 4333.

[123] A. Yung, W. B. Turnbull, A. P. Kalverda, G. S. Thompson, S. W. Homans, P. Kitov, D. R. Bundle, *J. Am. Chem. Soc.* **2003**, *125*, 13058.

[124] G. L. Mulvey, P. Marcato, P. I. Kitov, J. Sadowska, D. R. Bundle, G. D. Armstrong, *J.Inf. Dis.* **2003**, *187*, 640.

[125] Z. S. Zhang, E. A. Merritt, M. Ahn, C. Roach, Z. Hou, C. Verlinde, W. G. J. Hol, E. Fan, *J. Am. Chem. Soc.* **2002**, *124*, 12991.

[126] M. B. Pepys, J. Herbert, W. L. Hutchinson, G. A. Tennent, H. J. Lachmann, J. R. Gallimore, L. B. Lovat, T. Bartfai, A. Alanine, C. Hertel, T. Hoffmann, R. Jakob-Roetne, R. D. Norcross, J. A. Kemp, K. Yamamura, M. Suzuki, G. W. Taylor, S. Murray, D. Thompson, A. Purvis, S. Kolstoe, S. P. Wood, P. N. Hawkins, *Nature* **2002**, *417*, 254.

[127] J. Y. Liu, Z. S. Zhang, X. J. Tan, W. G. J. Hol, C. Verlinde, E. K. Fan, *J. Am. Chem. Soc.* **2005**, *127*, 2044.

[128] J. G. S. Ho, P. I. Kitov, E. Paszkiewicz, J. Sadowska, D. R. Bundle, K. K.-S. Ng, *J. Biol. Chem.* **2005**, *280*, 31999.

[129] P. I. Kitov, T. Lipinski, E. Paszkiewicz, D. Solomon, J. M. Sadowska, G. A. Grant, G. L. Mulvey, E. N. Kitova, J. S. Klassen, G. D. Armstrong, D. R. Bundle, *Angew. Chem., Int. Ed.* **2008**, *47*, 672.

[130] D. Solomon, P. I. Kitov, E. Paszkiewicz, G. A. Grant, J. M. Sadowska, D. R. Bundle, *Org. Lett.* **2005**, *7*, 4369.

[131] P. I. Kitov, G. L. Mulvey, T. P. Griener, T. Lipinski, D. Solomon, E. Paszkiewicz, J. M. Jacobson, J. M. Sadowska, M. Suzuki, K. I. Yamamura, G. D. Armstrong, D. R. Bundle, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16837.

[132] F. Sansone, G. Rispoli, A. Casnati, R. Ungaro, in *Synthesis and Biological Applications of Multivalent Glycoconjugates* (Eds.: O. Renaudet, N. Spinelli), Bentham Science Publishers, **2011**, pp. 36.

[133] M. Marradi, F. Chiodo, I. Garcia, S. Penades, in *Synthesis and Biological Applications of Multivalent Glycoconjugates* (Eds.: O. Renaudet, N. Spinelli), Bentham Science Publishers, **2011**, pp. 164.

[134] A. T. Aman, S. Fraser, E. A. Merritt, C. Rodigherio, M. Kenny, M. Ahn, W. G. J. Hol, N. A. Williams, W. I. Lencer, T. R. Hirst, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8536.

[135] M. F. Fontana, R. Manetti, V. Giannelli, C. Magagnoli, A. Marchini, R. Olivieri, M. Domenighini, R. Rappuoli, M. Pizza, *Infect. Immun.* **1995**, *63*, 2356.

[136] L. W. Ruddock, S. P. Ruston, S. M. Kelly, N. C. Price, R. B. Freedman, T. R. Hirst, *J. Biol. Chem.* **1995**, *270*, 29953.

[137] P. Schuck, *Biophys. J.* **2000**, *78*, 1606.

[138] J. Yang, L. K. Tamm, T. W. Tillack, Z. Shao, *J. Mol. Biol.* **1993**, *229*, 286.

[139] http://en.wikipedia.org/wiki/Atomic_force_microscope, **Access Date 27/09/2013**.

[140] J. X. Mou, J. Yang, Z. F. Shao, *J. Mol. Biol.* **1995**, *248*, 507.

[141] W. B. Turnbull, A. H. Daranas, *J. Am. Chem. Soc.* **2003**, *125*, 14859.

[142] Å. Holmner, A. Mackenzie, U. Krengel, *FEBS Letters* **2010**, *584*, 2548.

[143] J. D. Clemens, D. A. Sack, J. R. Harris, J. Chakraborty, M. R. Khan, S. Huda, F. Ahmed, J. Gomes, M. R. Rao, A.-M. Svennerholm, J. Holmgren, *Journal of Infectious Diseases* **1989**, *159*, 770.

[144] A. Holmner, G. Askarieh, M. Oekvist, U. Krengel, *J. Mol. Biol.* **2007**, *371*, 754.

[145] L. Balanzino, J. Barra, E. Galváin, G. Roth, C. Monferran, *Mol Cell Biochem* **1999**, *194*, 53.

[146] J. H. Chen, W. G. Zeng, R. Offord, K. Rose, *Bioconjugate Chem.* **2003**, *14*, 614.

[147] R. Fields, H. B. Dixon, *Biochem J* **1968**, *108*, 883.

[148] A. Amore, K. Wals, E. Koekoek, R. Hoppes, M. Toebes, T. N. M. Schumacher, B. Rodenko, H. Ovaa, *ChemBioChem* **2013**, *14*, 123.

[149] K. Rose, J. H. Chen, M. Dragovic, W. G. Zeng, D. Jeannerat, P. Kamalaprija, U. Burger, *Bioconjugate Chem.* **1999**, *10*, 1038.

[150] K. Rose, *J. Am. Chem. Soc.* **1994**, *116*, 30.

[151] S. O. Hatic, J. A. McCann, W. D. Picking, *Anal. Biochem.* **2001**, *292*, 171.

[152] C. Lesieur, M. J. Cliff, R. Carter, R. F. L. James, A. R. Clarke, T. R. Hirst, *J. Biol. Chem.* **2002**, *277*, 16697.

[153] M. G. Jobling, R. K. Holmes, *Mol. Microbiol.* **1991**, *5*, 1755.

[154] U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, P. Nordlund, *Anal. Biochem.* **2006**, *357*, 289.

[155] B. Goins, E. Freire, *Biochemistry* **1988**, *27*, 2046.

[156] E. A. Merritt, Z. S. Zhang, J. C. Pickens, M. Ahn, W. G. J. Hol, E. K. Fan, *J. Am. Chem. Soc.* **2002**, *124*, 8818.

[157] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.

[158] C. W. Tornøe, C. Christensen, M. Meldal, *The Journal of Organic Chemistry* **2002**, *67*, 3057.

[159] Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125*, 3192.

[160] W. P. Heal, M. H. Wright, E. Thinon, E. W. Tate, *Nat. Protocols* **2012**, *7*, 105.

[161] W. G. Lewis, F. G. Magallon, V. V. Fokin, M. G. Finn, *J. Am. Chem. Soc.* **2004**, *126*, 9152.

[162] S. S. Gupta, K. S. Raja, E. Kaltgrad, E. Strable, M. G. Finn, *Chem. Commun.* **2005**, *0*, 4315.

[163] J. A. F. Joosten, N. T. H. Tholen, F. Ait El Maate, A. J. Brouwer, G. W. van Esse, D. T. S. Rijkers, R. M. J. Liskamp, R. J. Pieters, *Eur. J. Org. Chem.* **2005**, *2005*, 3182.

[164] A. V. Pukin, C. A. G. M. Weijers, B. van Lagen, R. Wechselberger, B. Sun, M. Gilbert, M.-F. Karwaski, D. E. A. Florack, B. C. Jacobs, A. P. Tio-Gillen, A. van Belkum, H. P. Endtz, G. M. Visser, H. Zuilhof, *Carbohydrate Research* **2008**, *343*, 636.

[165] S. B. Larson, A. McPherson, *Curr. Opin. Struct. Biol.* **2001**, *11*, 59.

[166] M. Li, Y. Rharbi, X. Huang, M. A. Winnik, *Journal of Colloid and Interface Science* **2000**, *230*, 135.

[167] H. Schott, *J. Colloid Interface Sci.* **1998**, *205*, 496.

[168] S. Dai, K. C. Tam, *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2003**, *229*, 157.

[169] N. J. M. Sanghamitra, T. Ueno, *Chem. Commun.* **2013**, *49*, 4114.

[170] J. D. Brodin, X. I. Ambroggio, C. Tang, K. N. Parent, T. S. Baker, F. A. Tezcan, *Nat. Chem.* **2012**, *4*, 375.

[171] S. Sakai, Y. Shigemasa, T. Sasaki, *Tetrahedron Let.* **1997**, *38*, 8145.

[172] D. J. Williamson, M. A. Fascione, M. E. Webb, W. B. Turnbull, *Angew. Chem. Int. Ed.* **2012**, *51*, 9377.

[173] P. C. Weber, J. J. Wendoloski, M. W. Pantoliano, F. R. Salemme, *J. Am. Chem. Soc.* **1992**, *114*, 3197.

[174] W. A. Hendrickson, A. Pähler, J. L. Smith, Y. Satow, E. A. Merritt, R. P. Phizackerley, *Proceedings of the National Academy of Sciences* **1989**, *86*, 2190.

[175] Y. Mori, R. Wakabayashi, M. Goto, N. Kamiya, *Org. Biomol. Chem.* **2013**, *11*, 914.

[176] R. Belcher, *The application of chelate compounds in analytical chemistry, Vol. 34*, Pure and Applied Chemistry, **1973**.

[177] http://web.expasy.org/protparam/, **Access Date 27/09/2013**.

[178] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *The Journal of Organic Chemistry* **1997**, *62*, 7512.

[179] S. Foillard, M. O. Rasmussen, J. Razkin, D. Boturyn, P. Dumy, *The Journal of Organic Chemistry* **2008**, *73*, 983.

[180] S. R. Chhabra, B. Hothi, D. J. Evans, P. D. White, B. W. Bycroft, W. C. Chan, *Tetrahedron Let.* **1998**, *39*, 1603.

[181] Y.-X. Chen, S. Koch, K. Uhlenbrock, K. Weise, D. Das, L. Gremer, L. Brunsveld, A. Wittinghofer, R. Winter, G. Triola, H. Waldmann, *Angew. Chem. Int. Ed.* **2010**, *49*, 6090.

[182] L. M. Artner, L. Merkel, N. Bohlke, F. Beceren-Braun, C. Weise, J. Dernedde, N. Budisa, C. P. R. Hackenberger, *Chem. Commun.* **2012**, *48*, 522.

[183] D. Dunstan, L. Hough, *Carbohydr. Res.* **1972**, *23*, 17.

[184] E. C. Rodriguez, L. A. Marcaurelle, C. R. Bertozzi, *The Journal of Organic Chemistry* **1998**, *63*, 7134.

[185] E. Trévisiol, E. Defrancq, J. Lhomme, A. Laayoun, P. Cros, *Eur. J. Org. Chem.* **2000**, *2000*, 211.

[186] T. Wiseman, S. Williston, J. F. Brandts, L.-N. Lin, *Anal. Biochem.* **1989**, *179*, 131.

[187] A. T. Aman, S. Fraser, E. A. Merritt, C. Rodigherio, M. Kenny, M. Ahn, W. G. J. Hol, N. A. Williams, W. I. Lencer, T. R. Hirst, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8536.

[188] T. O. Nashar, H. M. Webb, S. Eaglestone, N. A. Williams, T. R. Hirst, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 226.

[189] M. Sandkvist, T. R. Hirst, M. Bagdasarian, *J. Bacteriol.* **1987**, *169*, 4570.

[190] J. P. Fürste, W. Pansegrau, R. Frank, H. Blöcker, P. Scholz, M. Bagdasarian, E. Lanka, *Gene* **1986**, *48*, 119.

[191] V. M. Sokolov, V. I. Zakharov, E. P. Studentsov, *Russ. J. Gen. Chem.* **2002**, *72*, 806.

[192] B. W. Bycroft, W. C. Chan, S. R. Chhabra, N. D. Hone, *J. Chem. Soc., Chem. Commun.* **1993**, 778.

[193] A. V. Pukin, C. A. G. M. Weijers, B. van Lagen, R. Wechselberger, B. Sun, M. Gilbert, M.-F. Karwaski, D. E. A. Florack, B. C. Jacobs, A. P. Tio-Gillen, A. van Belkum, H. P. Endtz, G. M. Visser, H. Zuilhof, *Carbohydr. Res.* **2008**, *343*, 636.