Towards detailed structural understanding of α-ʟ-fucosidase: insights through X-ray crystallography and inhibitor-binding

Daniel Wayne Wright

University of York

Chemistry

September 2015

Abstract

Carbohydrates are one of the most abundant biomolecules and are fundamental to the correct function of many biological processes. The monosaccharide L-fucose is incorporated into biological polymers including oligo and polysaccharides, and glycoproteins. L-fucose is often appended to the end of a glycan chain, and as such is recognised by lectins in a number of molecular recognition events. Due to this, the monosaccharide plays a critical role in the immune response, the colonisation of bacteria in mammals, and cancer. Two enzymes regulate homeostasis of L-fucosylated biomolecules, GDP-L-fucosyltransferases append the sugar to nascent biomolecules while α-L-fucosidases catalyse its cleavage. Two α-L-fucosidases exist in the human genome. Deficiency of one of these enzymes causes the lysosomal storage disorder fucosidosis, and the enzyme is upregulated in a number of cancers. Meanwhile, the other enzyme has been shown to play a critical role in enabling the adhesion of the pathogen *Helicobacter pylori* to mammals. Thus, inhibition of α-L-fucosidase activity is clinically relevant. In this work, the 1.6 - 2.1 Å X-ray crystal structures of α-L-fucosidase inhibitors complexed with a bacterial α-L-fucosidase are presented and discussed. Of the inhibitors discussed, the majority comprise 5-membered iminocyclitols, a potent yet infrequently used framework for inhibition of glycoside hydrolases, and their mode of binding to the enzyme in an E_3 conformation is elaborated from crystal structures. Further, the crystallographic observation of the interaction between a 6-membered ring inhibitor comprising an aziridine moiety as an electrophilic trap and a glycoside hydrolase is reported for the first time. Finally, efforts towards the purification and crystallisation of α-L-fucosidases from *Homo sapiens* are documented. The results reported herein may aid in the rational design of more potent inhibitors of α-L-fucosidase in the future and may help direct future efforts towards the crystallisation and structure solution of the clinically important human enzymes.

List of Contents

List of Figures

ix

List of Tables

Acknowledgments

Please bear with me while I acknowledge the numerous people who helped and supported me during the duration of my Ph.D., without whom, I would be unlikely to be where I am today.

First and foremost, I would like to give a heartfelt thanks to Prof. Gideon Davies for providing me with both academic and pastoral support during the course of my Ph.D studies, and for being an excellent supervisor, teacher, disseminator of information and man. I would also like to express my gratefulness to Prof. Marek Brzozowski, for teaching me the basics of crystallogenesis and crystallography, for his excellent scientific direction, for recommending me as a Ph.D. student, and for being my Independent Panel Member during my Ph.D.

I have some amount of regret in finishing my Ph.D. due to the need to move on and leave the York Structural Biology Laboratory, as I feel I will be leaving family behind. All of the members of the laboratory radiate a warm nature, a feeling which was likely fostered by the late Prof. Guy Dodson, and I think it is fitting memory to his grace that this holds unto this day.

At the onset of my Ph.D. studies I had very little experience in techniques unique to the biochemistry laboratory, and I would like to give a warm thankyou to Wendy Offen for giving me an introduction to these, and for the numerous times she helped me thereafter. Additional thanks to Drs. Andrew Thompson and Glyn Hemsworth for their invaluable assistance in matters pertaining to the wet lab. During the latter part of my Ph.D., having been a newcomer to the world of tissue culture, I would like to thank Dr. Tim Ganderton for his assistance and patience in teaching me everything I know related to the field, and him and Dr. Cristina Viola for setting up and managing the YSBL Eukaryotic Expression Facility. I am also grateful to the facilities and staff at the Diamond Light Source and the European Synchrotron Research Facility for the provision of beamline time and support, and to Dr. Johan Turkenburg and Sam Hart for managing the X-ray laboratory and for their help in X-ray data collection and processing.

Special thanks to Sally Lewis and Simon Grist for their patience and steadfastness in managing the YSBL wet lab, for keeping the place running and being singularly excellent given the deluge of work thrown in their direction. Thanks also to Ruth Wilkinson, Gemma Moy, Catherine Foster and Jane Harrison for their great help in

administrative matters and to Tim Kirk for the provision of technical computer support.

Without the work carried out by a number of people both before and during my Ph.D., the work presented herein would be much more limited in scope, and I would like to thank all of the people who contributed effort towards this. To Dr. Alicia Lammerts van Bueren, who, prior to the commencement of my PhD, carried out a large body of work on *Bt*Fuc2970; for overexpressing the gene, purifying and crystallising the protein, yielding an initial model for structure solution which was used by myself many times, and for determining a number of ligand structures, illuminating the conformational itinerary for GH29 hydrolysis. To the numerous synthetic chemists working in the groups of I. Robina, J. Behr and H. Overkleeft, for synthesising the ligands and inhibitors used for inhibition studies.

All work and no play would make Dan a dull boy, and everyone involved in the prestigious yet unofficial YSBL board gaming society and curry night will always have a special place in my heart: Andrew, Ben, Chris, Katie, James, Jenni and Sophie. Beyond this, I would like to personally thank Katie Jameson and Sophie McKenna personally for being such amazing friends and such good sounding boards as they were navigating the ordeal of thesis writing at the same time as myself.

Finally, I would like to thank everyone involved in the provision of outreach activities which I had great involvement in during the course of my study, and to give a final and heartfelt thanks to everyone in the York Structural Biology Laboratory.

Author's Declaration

All work presented herein was conducted solely by the author excepting those collaborative works outlined below and stated explicitly throughout the main text. This document was written in its entirety by the author. The author declares that none of the original work presented herein was previously used for the fulfilment of a research degree at the University of York or any other university.

The following pieces of work which are discussed in the thesis are acknowledged along with the people who conducted them:

- Cloning, expression, crystallisation and X-ray crystal structure of native *Bt*Fuc2970 along with a number of inhibitor structures of the same enzyme was conducted and reported by Dr. Alicia Lammerts van Bueren prior to the commencement of this thesis.

- The design and synthesis of iminocyclitol compounds **1**, **8-9** was conducted by researchers under the supervision of Prof. Inmaculada Robina, Universidad de Sevilla. These compounds are described in chapters 2 and 3.

- The design and synthesis of iminocyclitol compounds **2-7**, **10-13** was conducted by Dr. Audrey Hottin and Prof. Jean-Bernard Behr, Université de Reims Champagne-Ardenne. These compounds are described in chapters 2 and 3.

- Diagrams showing the conformation of activity-based probes bound to *Bt*Fuc2970 plotted on the free-energy surface of the enzyme displayed in Chapter 4 were created by Prof. Carme Rovira.

- The design and synthesis of activity-based probes and related compounds **14-19**, which are discussed in chapter 4, was conducted by researchers under the supervision of Prof. Hermen Overkleeft, Universiteit Leiden, particularly Jianbing Jiang. Further, experiments relating to the probing of α-L-fucosidase activity by tissue culture-derived and Gaucher spleen-derived FUCA1, which were reported in a collaborative research article and discussed briefly in chapter 4 were performed by Wouter Kallameijn under the supervision of Prof. Johannes Aerts.

- The FUCA1 and FUCA2 gene transcripts described in chapter 5 were designed by Dr. Alicia Lammerts van Bueren, and were cloned and inserted into the pET-30a$^+$ vector by GenScript prior to the commencement of this thesis.

- Constructs for expression of the FUCA1 and FUCA2 genes in cultured mammalian cells as described in chapter 5 were received from the Complex Carbohydrate Research Centre through the Repository of Glyco-enzyme expression Constructs. This initiative designed and synthesised the constructs.

- Expression studies of FUCA1 using *Pichia pastoris* as an expression system was conducted by Dr. Jared Cartwright of the York Bioscience Technology Facility.

- Crystallisation trays for the optimisation of 8HisFUCA1:**2** crystals were designed and created by Wendy Offen

The peer-reviewed publications listed below were published during the course of this thesis. These publications are listed in order of publication date.

- Wright, D. W., Moreno-Vargas, A. J., Carmona, A. T., Robina, I. & Davies, G. J., Three dimensional structure of a bacterial α-L-fucosidase with a 5-membered iminocyclitol inhibitor, *Bioorgan Med Chem* 21, 4751-4754 (2013).

- Hottin, A. *et al.*, α-L-Fucosidase Inhibition by Pyrrolidine-Ferrocene Hybrids: Rationalization of Ligand-Binding Properties by Structural Studies, *Chem-Eur J* 19, 9526-9533 (2013).

- Moreno-Clavijo, E. *et al.*, Exploring a Multivalent Approach to alpha-L-Fucosidase Inhibition, *Eur J Org Chem*, 7328-7336 (2013).

- Hottin, A., Wright, D. W., Davies, G. J. & Behr, J. B., Exploiting the Hydrophobic Terrain in Fucosidases with Aryl-Substituted Pyrrolidine Iminosugars, *Chembiochem*, 277-283 (2014).

- Jiang, J., *et al.*, In vitro and in vivo comparative and competitive activity-based protein profiling of GH29 α-L-fucosidases, *Chem Sci* 6, 2782-2789 (2015).

Chapter 1: Introduction

Abstract

Carbohydrates are one of the most abundant biological polymers and are critical to many biological processes. Monosaccharides are catabolised as an energy source while polysaccharides have numerous important biological roles, both structural such as in the plant cell wall and functional *e.g.* in cell-cell interactions, signal transduction and protein folding and function. L-fucose is one of the series of L-configured monosaccharides and is unusual in its lack of a hydroxyl group at the C6 carbon; the sugar is incorporated into a number of biologically important polymers including those from plant and animal extracellular matrices, and many immunogens such as blood group ABO and Lewis antigens. GDP-L-fucosyltransferases and α-L-fucosidases catalyse the transfer of L-fucosyl moieties to nascent polysaccharides, and the cleavage in an *exo*-fashion of an α-L-fucoside from such a polymer, respectively. Deficiency in either of these enzymes causes disease phenotypes in mammals, reflecting the importance of L-fucosylated carbohydrates in nature. While deficiency in various GDP-L-fucosyltransferase activities leads to a range of outcomes due to the diverse roles of L-fucosyl moieties in biology, deficiency in α-L-fucosidase activity causes the lysosomal storage disorder fucosidosis, which causes severe neuronal and psychomotor defects and early death. Herein, the role of carbohydrates in nature, and particularly the role of the monosaccharide L-fucose, will be discussed.

1.1   *The role of carbohydrates in nature*

In the post-genomic revolution era, it is perhaps often overlooked how important carbohydrates are in the biochemical context. Carbohydrates are one of the most abundant species in nature, are found in every kingdom of life and are pivotal in many biological processes such as energy storage, the formation of durable biological structures such as the plant cell wall, and the immune response. It is been estimated that more than half of all mammalian proteins are glycoproteins[1] and it is thus no surprise that the attachment of carbohydrates to these proteins often serves a critical role in affecting the way they perform their functions[2]. A great number of disease states are also related to carbohydrate metabolism[2,3], and many pathogens effect their adhesion by exploitation of carbohydrate-mediated interactions[4-6].

Historically, the etymology of the term carbohydrate (also sugars, saccharides) was a literal chemical definition, 'hydrates of carbon' *i.e.* describing compounds that have the general chemical structure $C_x(H_2O)_y$. Many carbohydrates *e.g.* sialic acids, N-acetylhexosamines and deoxyribose, however, deviate from this general formula, and a more correct definition for a carbohydrate is a polyhydroxy aldehyde (aldose sugar) or ketone (ketose sugar). The simplest form of a carbohydrate is a monosaccharide, while polymerisation of monosaccharides with other carbohydrates leads to the formation of di, tri, oligo and polysaccharides (also glycans).

Carbohydrates are typified by their large number of chiral centres, and many biological systems have evolved to specifically recognise and interact with a single stereoisomer from the large number present in nature. The large number of chiral centres in these compounds poses a significant challenge for their chemical syntheses, or synthesis of related compounds *in vitro*. Thus, chemical syntheses of carbohydrate-related compounds typically involve the use of carbohydrates already containing the complex chirality required as a feedstock.

Carbohydrate monomers (monosaccharides) are named to represent their open-chain length, aldose or ketose nature, and stereochemistry. L- and D- nomenclature for all other than the shortest carbohydrate, glyceraldehyde, is not, as one would expect, based on the direction in which a carbohydrate epimer rotates the orientation of plane-polarised light, but is based on the chirality of the highest-order chiral centre in the open-chain form of the carbohydrate, and whether it is on the same side of the Fischer projection (Figure 1.1) as that of L- or D-glyceraldehyde.
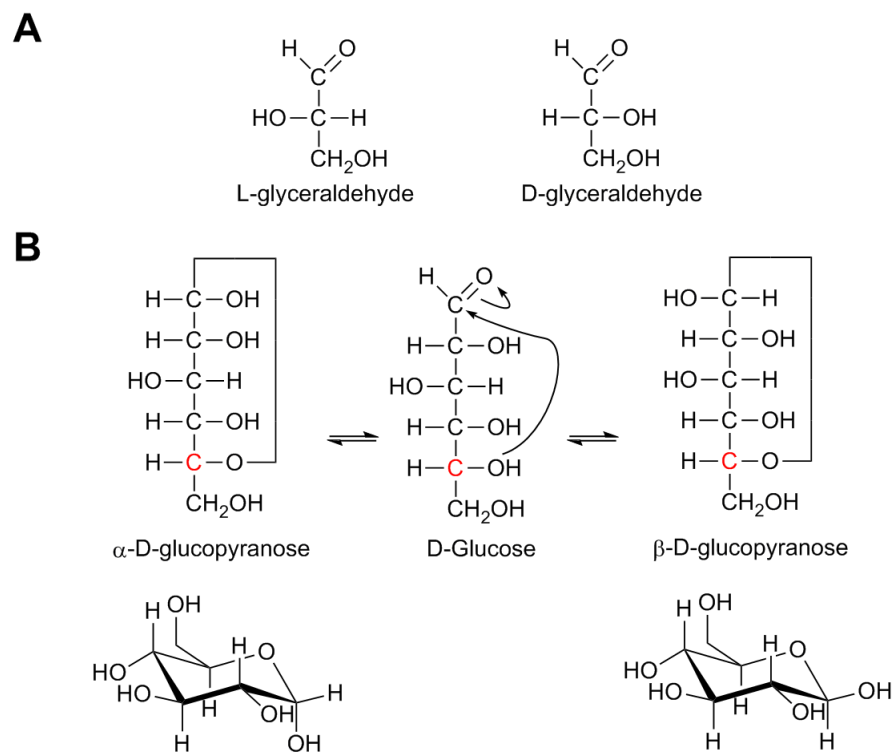
Figure 1.1. Carbohydrate nomenclature and ring-closing reactivity. Panel **A**: Both isomers of glyceraldehyde, the shortest chain carbohydrate, shown in Fischer projection. Panel **B**: Mutarotation of D-glucose. L- and D- nomenclature of carbohydrates other than glyceraldehyde is based on the highest number chiral centre (shown in red), and whether this is on the same side of the Fischer projection of L- or D- glyceraldehyde. α and β nomenclature is dictated by whether the endocyclic bond formed is on the same (α) or opposite (β) side of the Fischer projection as the newly formed, anomeric hydroxyl.

The carbonyl groups in carbohydrates are labile to nucleophilic attack from their internal hydroxyls and these compounds exist in equilibrium of their open-chain form with multiple closed ring forms (Figure 1.1 shows this for glucose). The ring strain of smaller heterocycles typically favours the formation of larger ring structures, up to the most favourable 6-cycles, which have the lowest strain energy. Thus, the most prevalent ring structures of carbohydrates are furanose and pyranose rings (5- or 6- membered ring structures, respectively, comprising a single O atom). Ring-closing reactivity can occur towards either the Re- or Si- face of the electrophilic carbonyl group leading to α- or β-products, with nomenclature being based on the chirality of the newly-formed (anomeric) chiral centre (Figure 1.1). Reaction is typically favoured from the face of the carbonyl which places hydroxyls in an equatorial conformation, in order to minimise trans-diaxial steric clashing.

Nomenclature dictates that the β-anomer of a carbohydrate is that where the anomeric hydroxyl has the same absolute stereochemistry as that of the highest-order chiral centre of the carbohydrate, whilst the α-anomer has different absolute stereochemistry between these two chiral centres.

1.1.1 Enzymes that act on carbohydrates

Further reactivity of carbohydrates occurs in nature through enzymatic catalysis. Glycosyltransferase enzymes catalyse the formation of glycosidic bonds between activated donor sugars and nucleophilic acceptor moieties to form polysaccharides, glycolipids or glycoproteins (Figure 1.2, **A**).[7] An important point of nomenclature is that of the reducing vs. the non-reducing end of a polysaccharide; reducing sugars contain free aldehyde groups or are able to form these moieties through mutarotation. While no free aldehyde group typically exists in biomolecules, the distinction of the reducing vs. non-reducing end of the polysaccharide chain is often made. Glycosyltransferases catalyse the appendage of saccharides to the non-reducing end of a nascent polysaccharide.

Hydroxyls of carbohydrates may further be sulfated or phosphorylated (Figure 1.2, **B-C**). The sulfation of oligosaccharides occurs by the action of sulfotransferases in the golgi compartment and drastically alters the physical and biological properties of the oligosaccharide.[8] In sulfated glycosaminoglycans such as heparin, sulfation is critical to biological function in mediating molecular interactions through electrostatics; sulfation patterns of this biomolecule have been shown to be important *e.g.* for its sequestration of growth factors[9], interaction with chemokines during leukocyte migration[10] and uptake by various cell types[11]. Oligosaccharide phosphorylation on the other hand is involved in enzyme trafficking. The incorporation of mannose-6-phosphate into N-glycans acts as a signalling sequence for transport of the nascent glycoprotein to the lysosome by its recognition with mannose-6-phosphate receptors and subsequent vesicular transport; whilst other mechanisms exist, this is the predominant means for lysosomal targeting[12].

A

B

C

D

R = saccharide, lipid or protein
R' = H or SO_3^-
R" = H or PO_4^{2-}
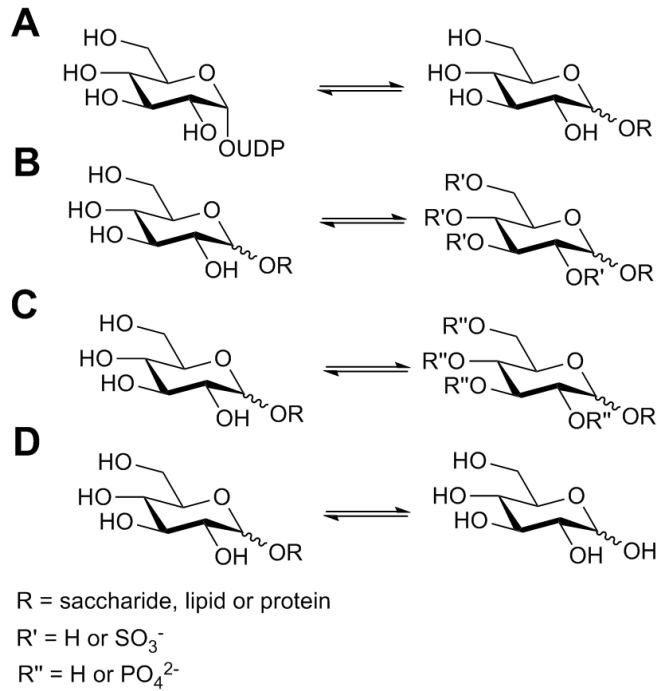
Figure 1.2. Schematic of some of the reactions catalysed by carbohydrate-processing enzymes. Glucose is used as a representative monosaccharide. Substrate specificities for these reactions may be slightly altered *i.e.* C2-C6 OH groups may already be modified. **A**: Glycosyltransferases catalyse the transfer of glycosyl moieties to acceptor moieties through activated donor sugars. **B**: Sulfotransferases (**B**) and kinases (**C**) may append sulfate or phosphate moieties to the free hydroxyls on glycosides. **D**: glycoside hydrolases catalyse the cleavage of the energetically stable glycoside bond.

Glycosidic bonds are incredibly stable, indeed even more so than other biological polymers such as DNA and proteins[13-15]; cleavage of the glycosidic bond is thus catalysed by glycoside hydrolase enzymes which enhance the rate of hydrolysis by as much as a factor of 10^{17} [13]. Glycoside hydrolases catalyse the hydrolysis of glycosylated structures such as polysaccharides, glycoproteins and glycolipids, to yield mono-, di- or oligosaccharides, and may act in either an *exo-* or *endo-* manner *i.e.* hydrolyse a terminal or non-terminal glycosidic bond respectively (Figure 1.2 **D**).

Carbohydrates have many biological roles, *inter alia*, in energy transfer[16], their incorporation in extracellular matrices such as the plant cell wall and animal mucin proteins[17], protein folding[18], their mediation of cell-cell interaction events and signal transduction[19].

1.2   Fucose

Fucose (6-deoxy-L-galactose, below) is a deoxyhexose sugar that is present in all domains of life, and belongs to the rare series of L-sugars. The α-anomer of fucose is the most thermodynamically favourable due to the anomeric effect; the C1 hydroxyl preferentially lies in an axial position (α-fucose) rather than an equatorial position (β-fucose).
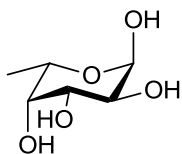
Fucose is incorporated into a number of polysaccharide structures throughout nature; often, the sugar is appended to the non-reducing end of glycans and forms part of a molecular recognition and signalling moiety, however a number of fucose-rich structural polysaccharide structures also exist.[20] As a structural polysaccharide, L-fucose is a component of the extracellular matrices of brown seaweed and some marine invertebrates[21,22], and is also incorporated as a side-chain component in the hemicellulosic portion of xyloglucan of some plants[23]. Fucose is also a component of both N-linked and O-linked glycans, and as such a number of fucosyltransferase enzymes are involved in appending L-fucosyl moieties to nascent oligosaccharides during the biosynthesis of these structures.[20] The homeostasis of fucosylated structures also requires the action of *exo*-acting α-L-fucosidases which cleave L-fucose residues from the non-reducing end of a polysaccharide chain. What follows is a description of some of the many places throughout nature that this monosaccharide may be found. A number of the more esoteric biological structures containing L-fucose will be discussed first, followed by a more in-depth discussion of the sugar's incorporation in carbohydrate structures important to human health.

1.2.1   L-Fucose as a component of plant and animal extracellular matrices

Holothurians, or sea cucumbers, incorporate fucosylated chondroitin sulfate (FucCS) into their body walls. These carbohydrates are glycosaminoglycans comprising branched L-fucose moieties that were first described in 1988, when extracted from the holothurian *Ludwigothurea grisea*.[24] The molecular architecture of FucCS has since been elucidated; the polysaccharide has a repeating main chain of β(1,4)-linked glucuronic acid and β(1,3)-linked N-acetyl galactosamine, with a large number of sulfated Fucα(1,3) branches appended to the main chain glucuronic

acid (Figure 1.3).[22,25] The analysis of FucCS extracts after chondroitinase treatment by Vieira *et al.* showed that the polysaccharide is incorporated into proteoglycans with a wide molecular weight range, however all products reported displayed a similar protein to carbohydrate ratio.[26] FucCS has been shown to have a number of beneficial effects to human physiology, including its activity as an anticoagulant and anti-thrombotic. The anticoagulant activity of FucCS is attributed to the sulfated fucose present in its structure, as de-sulfation or de-fucosylation of the carbohydrate abolishes its anticoagulant activity.[27] Likewise, the anti-thrombotic effectiveness of the polysaccharide is dependent on its branched fucosyl moieties.[28] Other described therapeutic effects of FucCS include, *inter alia*, its attenuation of tumour metastasis and inflammation through blocking the binding between selectins and sialyl-Le^x, *vide infra*[29], and its ability to ameliorate hyperglycemia in skeletal muscle through effects likely arising from activation of genes in the PKB/GLUT4 signalling pathway[30].

Brown seaweed, and a number of marine invertebrates, produce polysaccharides comprising mostly highly sulfated branched chains of L-fucose, termed fucoidans.[21,31] Following the trend of sulfated fucose-containing polysaccharides, these compounds have also been shown to have a number of beneficial effects to human physiology. As with FucCS, this includes their activity as anti-inflammatory agents[32], anti-coagulants[32,33], and anti-proliferatives against cancer cells[32,34], and due to this, the polysaccharide is marketed as a dietary supplement. Fucoidan has also been shown to have a marked inhibitory effect on human, hamster and guinea pig sperm binding to oocyte zona pellucida reflecting the roles of L-fucose in fertility, *vide infra*.[35]

Fucoidans from different organisms have diverse structures; linear L-fucose chains may be α(1,3) linked or built from repeating α(1,3) and α(1,4) linkages, different sulfation patterns exist between different organisms, acetylation may be present at positions 2, 3 or 4 of fucose residues, and some fucoidans also incorporate xylose, mannose, galactose, glucose, glucuronic acid and α(1,2)-linked fucose branches in their structures (Figure 1.3).[21,31]

L-Fucose is incorporated as a side-chain component in the pectic and hemicellulosic material of plant cell walls. In pectins, fucose is incorporated in the side-chains of rhamnogalactouronans I and II, though the sugar makes up a very small proportion of the overall carbohydrate content of these polysaccharides.[36] Xyloglucan comprises a large portion of the hemi-cellulosic biomass present in plant cell walls, and L-fucose is incorporated into xyloglucan side-chains in a genetically diverse subset of plants, *inter alia*, pines, legumes and onions, in XXFG and XLFG

subunits (Figure 1.4, for nomenclature see Fry *et al.* 1993[37]).[23] As the sugar is incorporated into xyloglucan in species which are evolutionarily unrelated, there was speculation about the importance of the sugar in plant development, leading to a number of studies to address this question.
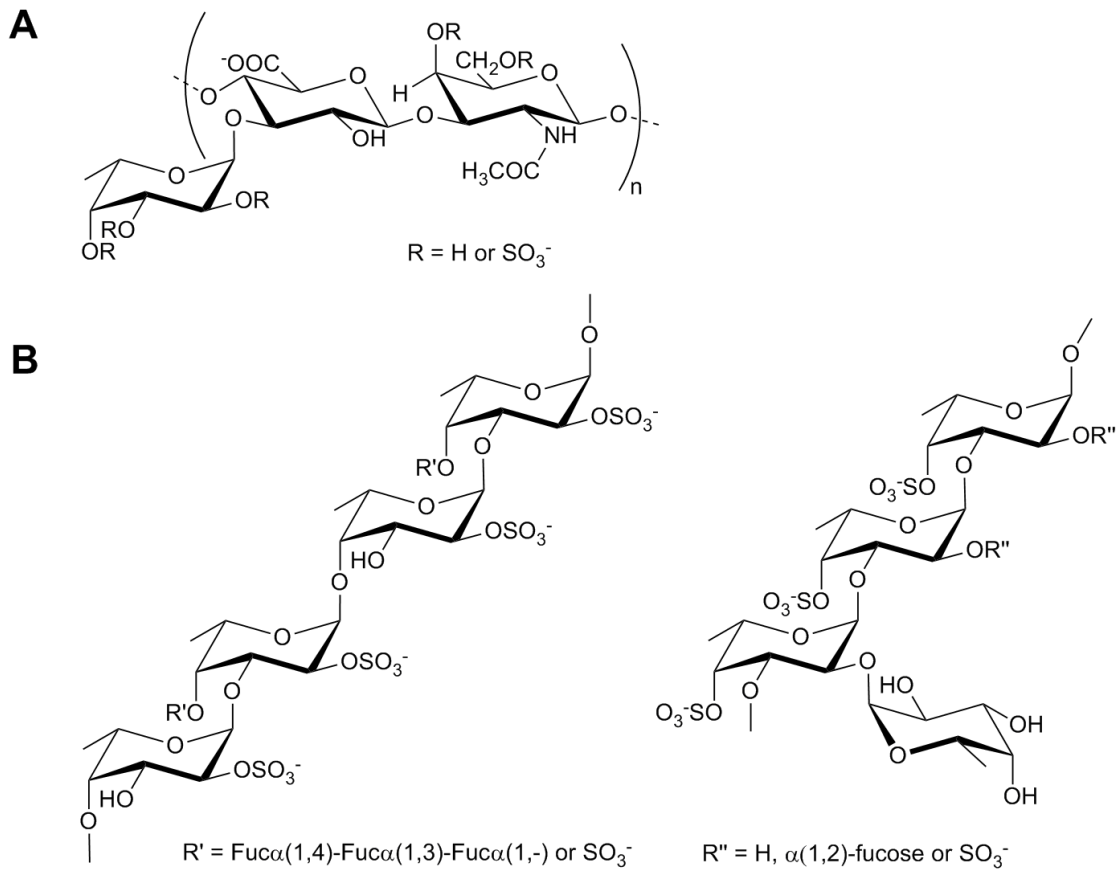
Figure 1.3. Strucures of fucose-rich carbohydrates from seaweed and sea cucumbers. Panel **A**: Structure of fucosylated chondroitin sulfate from the body wall of sea cucumbers, adapted from Pomin *et al.* 2014[22]. Panel **B**: Molecular architectures of selected fucoidans from brown seaweed and marine invertebrates to represent their marked structural heterogeneity, adapted from Ale *et al.* 2013[31].

Molecular dynamics simulations, supported by *in vitro* assays, of pea and nasturtium xyloglucans, differing in nasturtium xyloglucan's lack of α(1,2)-linked fucose residues, revealed that fucosylation of xyloglucan markedly increased its affinity for cellulose[38]; as the binding of xyloglucan to cellulose is thought to modulate cell growth, this result would seem to indicate a role of fucose in modulating plant growth.
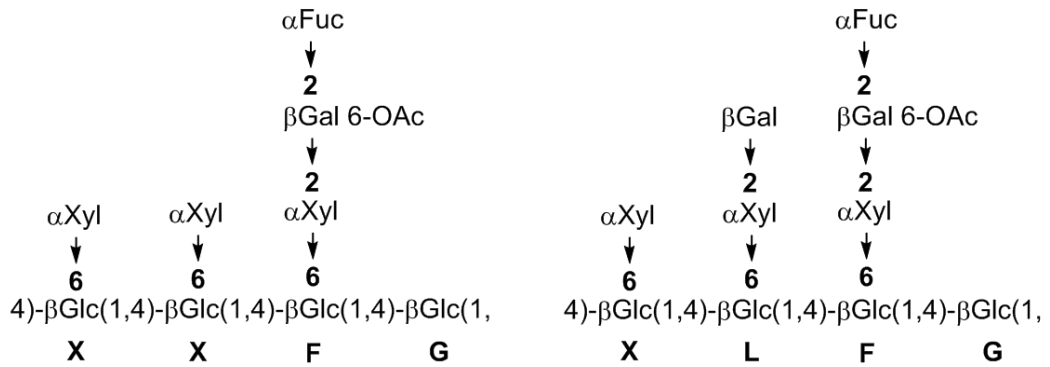
Figure 1.4. Fucose as a side-chain component of hemicellulose in the plant cell wall.

The characterisation of *mur1*, an *Arabidopsis thaliana* mutant strain completely deficient in L-fucose, revealed that the lack of L-fucose causes dwarfism and low cell wall strength *in planta*.[39] Later, an *A. thaliana mur2* mutant, lacking the $\alpha(1,2)$-fucosyltransferase that appends fucose to the Galβ(1,2)-Xylα(1,6)-Glc sidechain of xyloglucan (Figure 1.4) was characterised; this mutant strain has a 50 % reduction in fucose content of the plant's cell wall, indicating that half of the extracellular matrix fucose content in *A. thaliana* is from xyloglucan.[40] Whilst *mur1* mutant plants display low cell wall strength, *mur2* mutants were shown to have similar growth and cell wall strength to wild type plants under a variety of growth conditions, leading to speculation that the loss of fucose from fucosylated pectin is to be attributed for the lower growth and cell wall strength of the *mur1* mutant.[40] This may be explained by the fact that *mur2* mutant plants incorporate galactose where fucose would be appended in about one half of their xyloglucan side-chains[41]; the results of Levy and co-workers indicated that it is the presence of trisaccharide side-chains that influence the binding of xyloglucan to cellulose[38].

In 2003, a third *A. thaliana* mutant with influence on carbohydrate composition in the cell wall, *mur3*, was identified; this mutant is deficient in $\beta(1,2)$-galactosyl transferase activity, therefore the normally fucosylated side-chains of xyloglucan are lacking in both Fuc and Gal, and comprise only a single Xyl residue.[42] Tensile strength measurements taken on wild-type, *mur2*, and *mur3* mutants revealed that the galactosylation, rather than fucosylation, of xyloglucan is critical to cell wall strength.[43] It is thus uncertain what the role of $\alpha(1,2)$-fucosyl moieties in xyloglucan are.

In addition to its incorporation into structural components of the eukaryotic extracellular matrix, fucose is incorporated into both N-linked and O-linked glycan structures in animals, and N-linked glycans in plants.

1.2.2 Fucose in N-linked and O-linked glycans

In N-linked glycans and O-linked glycans, L-fucose is incorporated via Fucα(1,2)-Gal, Fucα(1,3)-GlcNAc, Fucα(1,4)-GlcNAc or Fucα(1,6)-GlcNAc linkages, or directly appended to protein serine or threonine residues.[20] For N-linked glycans, the monosaccharide represents a non-reducing end of the final oligosaccharide product. Thus, fucose is exposed to the biological microenvironment, and many cellular signalling processes involve the recognition of fucosylated glycans such as the ABO and Lewis antigens (Figure 1.5).

Lewis antigens are synthesised in various cell types by the action of numerous fucosyltransferase enzymes present in the human genome. These antigens have important roles in the immune response during leukocyte rolling and extravasation[44,45], and in the colonisation of xenobiotics spanning a spectrum of interactions with the host from symbiogenesis to pathogenicity[46]. The blood group antigens Le^x, H and Le^y are thought to be involved in carbohydrate-mediated cellular adhesion processes during signal transduction by type 1 glycosynapses.[47-49]

O-linked glycosylation involves the attachment of a sugar moiety or oligosaccharide to a protein serine or threonine residue. O-linked fucosylation occurs solely in eukaryotes, in the endoplasmic reticulum, and involves the addition of L-fucose or a fucosylated oligosaccharide to proteins containing a cysteine-rich motif.[50] Two fucosyltransferases are known to be involved in O-fucosylation, POFUT1 and POFUT2.
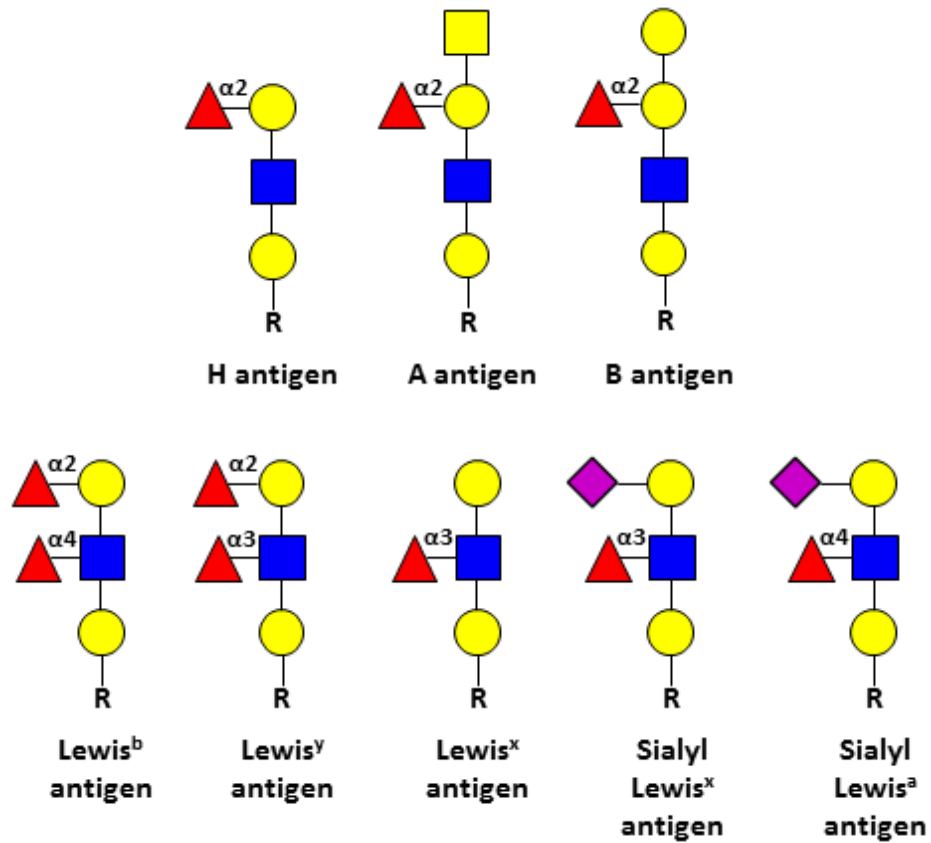
Figure 1.5. Examples of ABO and Lewis oligosaccharide structures containing L-fucose. Monosaccharides are displayed by the standard proposed by Varki *et al*. in 2009 (Yellow circle = galactose, yellow square = GalNAc, blue square = GlcNAc, purple diamond = sialic acid, red triangle = fucose).[51]

Human POFUT1 was first isolated and its biological function characterised in 2001.[52] The enzyme catalyses the addition of fucose to EGF repeats of, *inter alia*, Notch receptors. These receptors are large type I transmembrane proteins that are involved in intercellular signalling in a variety of cellular contexts during development and are conserved between metazoan species.[53] Genetic abnormalities leading to deficiency in the Notch pathway are linked to a number of developmental diseases such as Alagille syndrome, which is caused by deficiency in either the Notch ligand Jagged1[54] or the Notch2 receptor [55]. The Notch signalling mechanism is an elegant example of the use of regulated proteolysis by nature and was recently reviewed by Kopan and Ilagan[53]. Signal transduction by the mature Notch receptor is initiated by its binding to Notch ligands, which are often other transmembrane proteins, expressed on the surface of the signal-sending cell[56]. This binding event induces a conformational change in Notch, opening it to sequential proteolytic attack by ADAM metalloproteases and γ-secretase yielding the Notch intracellular

domain fragment.[57] This fragment is subsequently trafficked to the nucleus where it complexes with the DNA-binding protein CSL; in the absence of Notch this protein is involved in the formation of a co-repressor complex blocking gene transcription.[58] Finally, the CSL-Notch intracellular domain complex is recognised by Mastermind and the ternary complex formed causes transcriptional activation of target genes (Figure 1.6).[58]

Notch is O-fucosylated by POFUT1 between the second and third cysteines of its epidermal growth factor repeats in the consensus sequence Cys^2-X-X-X-X-(Ser/Thr)-Cys^3.[59,60] Further glycosylation of this core fucose occurs in mammalian Notch; extension of Notch core fucose occurs through the addition of GlcNAc by Fringe, and subsequent addition of galactose and sialic acid leading to the mature mammalian tetrasaccharide Siaα(2,3)-Galβ(1-4)-GlcNAcβ(1-3)-Fuc.[59,61,62] Studies have shown that *Pofut1*$^{-/-}$ mice display embryonic lethality[63], and that RNA interference of *Ofut1* in *Drosophila melanogaster* leads to loss of Notch function and drastic phenotypic defects[64], thus these fucosyltransferases are critical in Notch function and organism viability.
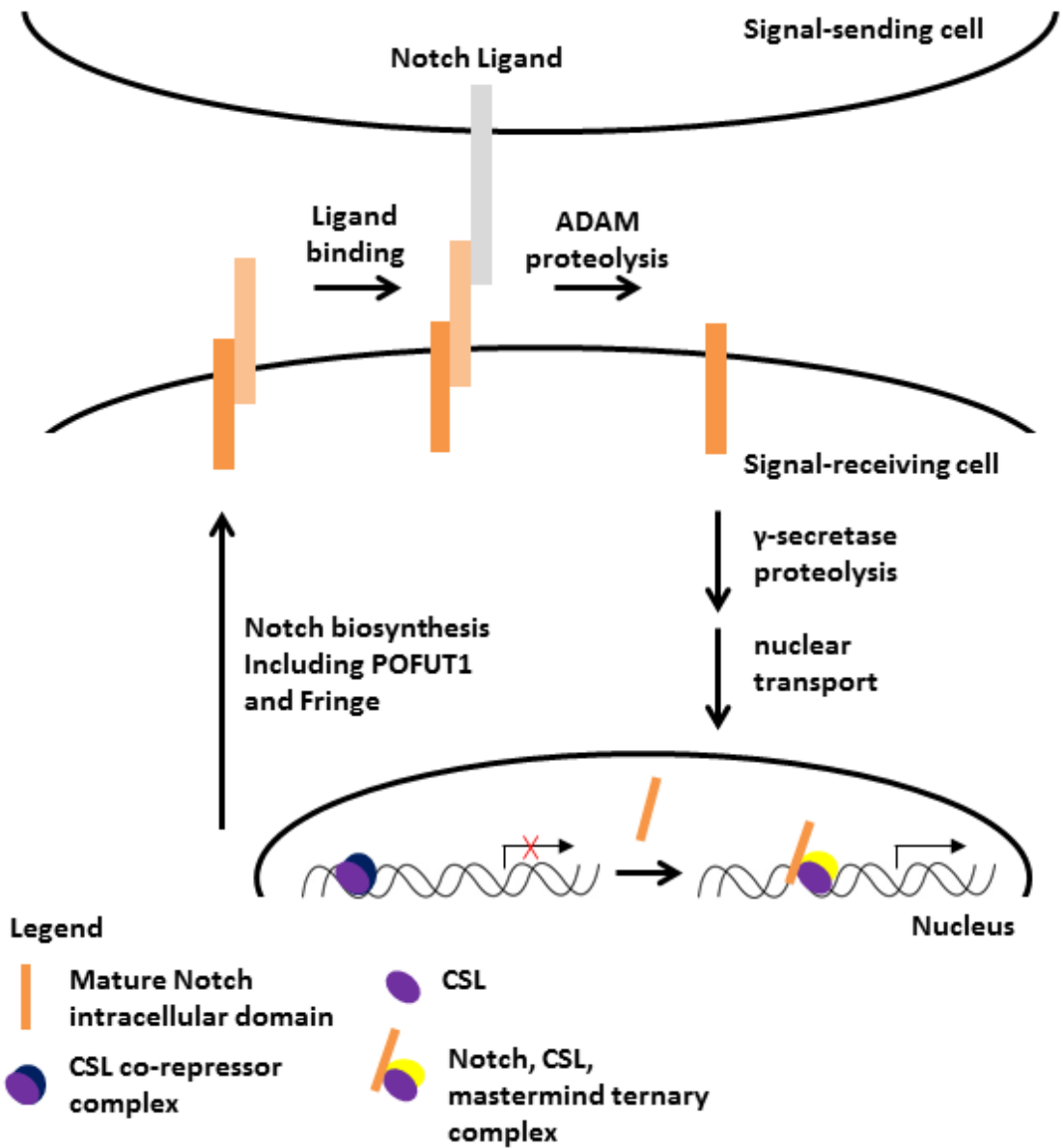
Figure 1.6. Mechanism of the Notch signalling pathway. Adapted from Kopan *et al*. 2009[53].

1.2.2.1   Fucose in the immune response

Lewis antigens (Figure 1.5) act as ligands for selectins and are involved in a number of molecular recognition events throughout the body, such as those occurring during leukocyte rolling, a critical part of the leukocyte adhesion cascade and the innate immune system.[44,45] For the correct localisation of leukocytes at a site of injury of infection, the immune cells must first escape the high shear forces exerted in the intravascular compartment before their extravasation across the endothelial cell membrane. During this first, 'slow rolling' stage of the leukocyte adhesion cascade, leukocytes tether with low affinity to the endothelial cell surface

by the interaction between C-type lectins called selectins and a number of selectin ligands, both of which are expressed on leukocytes, platelets, and the endothelial cell surface. A plethora of research undertaken in the early 1990s identified selectin ligands as glycoproteins that express the fucosylated sialyl-Le[x] tetrasaccharide.[65] Fucosylation is indeed critical to leukocyte function; genetic disruption of the α(1,3)-fucosyltransferase FUTVII has been shown to markedly impair leukocyte function *in vivo* in mice[66], and treatment of a chemokine-activated Monocytic cell line with α-ʟ-fucosidase was shown to reduce leukcoyte migration to a background level[67].

In 1992, a novel disease related to deficiency in leukocyte adhesion was characterised; tests run on samples from two unrelated boys displaying mental retardation, short stature, and recurrent contraction of bacterial infections showed that the patients did not correctly express Lewis antigens, along with having the non-Secretor and Bombay phenotypes.[68] As all of these characteristics represent a lack of fucosylated glycan production by distinct fucosyltransferase enzymes, the authors hypothesised deficiency in fucose metabolism is responsible for the disease, which is termed either leukocyte adhesion deficiency type II (LADII), or congenital disorder of glycosylation-IIc.[68]

More recently, genetic evidence has shown that LADII results from deficiency in the activity of a GDP-fucose specific transporter, which transports GDP-fucose across the cytosol and into the golgi compartment prior to glycosyltransferases incorporating the activated sugar into glycans.[69,70] The genetic abnormalities causing LADII can be due to either deficiency in activity of a properly localised GDP-fucose transporter, or its improper localisation in the endoplasmic reticulum.[71] In either case, this results in deficiency of sialyl-Le[x] and other antigens required for leukocyte rolling, disrupting the leukocyte adhesion cascade at an early stage.

1.2.2.2   *Fucose in cancer*

As it is approximated that more than half of all mammalian proteins are post-translationally glycosylated[1], it is perhaps not surprising that aberrant glycosylation is associated with cancer adhesion and metastasis. Of course, as the disease remains one of the leading causes of death in both developing and developed countries[72], a great deal of research has been aimed at the study of protein glycosylation in cancer. Changes in protein glycosylation, and in particular fucosylation, have been reported in various cancer types, predominantly with upregulation of cell surface glycan expression in neoplastic cell lines, however downregulation of some of these

species has also been reported in some cases, excellently reviewed by Christiansen et al.[73].

As many experiments studying the aberrant glycosylation of cancer cells utilise neoplastic cell lines in *in vitro* culture conditions, it is important to note that the glycosylation patterns displayed by cells may be influenced by the cell culture environment[74] and changes in glycosylation patterns may therefore not necessarily represent changes of neoplastic cells over healthy ones. Thus, it is important to support data gained from culture-based experiments with that from clinical cancer tissue samples. While significant correlations exist between the differences observed in both cell lines and tissues in lung cancer[75], this is not the case for all tumour types. For example, significant differences have been reported in the DNA methylation profile of ovarian cancer cell lines and tissues[76], and observations have shown that a panel of ovarian cancer cell lines from various origins more closely resemble themselves than the tissues they are supposed to model[77].

Nevertheless, upregulated core fucosylation *i.e.* $\alpha(1,6)$-fucosylation of interior GlcNAc residues on N-glycans appears to be a common feature of cancers excluding melanoma.[73] This is supported by the increased expression of the $\alpha(1,6)$-fucosyltransferase FUT8 in tissues from liver and breast cancers[78,79] and the increased intracellular GDP-fucose levels observed in hepatocellular carcinoma[80]. The core fucosylation of serum alpha-fetoprotein proves a potent biomarker for early diagnosis of hepatocellular carcinoma.[81] α-L-fucosidase also has potential as a biomarker for this cancer, both in diagnosis and prognosis, and when used in combination with alpha-fetoprotein the two biomarkers provide greater specificity for tracking the disease[82,83]. Serum α-L-fucosidase activity has also been shown to correlate with progression-free survival for breast cancer patients treated with trastuzumab, with a greater FUCA2 activity correlating with higher sensitivity to trastuzumab treatment.[84]

Fucosylation in ABO and Lewis type linkages may also be either up or downregulated in neoplastic cells.[73] Le^x and sialyl-Le^x expression has been observed to correlate with poor survival in patients with colon[85] and breast cancers[86]. Evidence suggests that FUT6 expression correlates with sialyl-Le^x expression and adhesion potential of neoplastic cells in a number of breast cancers[87] while a similar correlation has been shown on overexpression of FUT7 in colon cancer cell line HT-29[88]. The metastatic potential of human pancreatic tumour cell line BxPC-3 was inhibited after expression of *FUT3* antisense mRNA[89], indicating that this $\alpha(1,3/4)$-fucosyltransferase may be important in pancreatic cancer viability.

The Le[y] antigen appears to be upregulated in serous forms of ovarian cancer.[90] Transfection of the *FUT1* α(1,2)-fucosytransferase gene into ovarian cancers was further observed to lead to increased Le[y] expression and to promote neoplastic cell proliferation.[91,92] It has been postulated that the increased biosynthesis of selectin-associated Lewis antigens such as sialyl-Le[x] is important for cancer metastasis as the metastatic cascade may mirror the leukocyte adhesion cascade, requiring binding of cell-surface associated selectin ligands with selectins on a target cell surface for cancer metastasis.[93] Aberrant regulation of fucosyltransferases does not appear to be limited to their overexpression; extracts from hepatocellular carcinoma tissues have been observed to have downregulated α(1,2) and α(1,3)-fucosylation[94] and *FUT1* and *FUT4* mRNA were observed to be absent in almost 25% of over 100 tested melanoma cell lines[95].

As neoplastic cells typically require more L-fucose for proliferation than healthy cells, they uptake the monosaccharide with a greater efficiency. This may be exploited for the selective uptake of drugs to the site of neoplasia. For example, Yoshida *et al.* have reported the success of L-fucose bound liposomes containing the anti-cancer drug Cisplatin in inhibiting proliferation of pancreatic cancer *in vivo* in mice.[96] More recently, Babiuch and co-workers designed glycopolymers capable of self-assembling into nanoscale micelles.[97] These micelles were decorated with β-L-fucose and/or β-D-glucose moieties in varying ratios from 100% β-L-fucose to 100% β-D-glucose. The micelles richer in β-L-fucose were shown to have a greater uptake into pancreatic, lung and ovarian carcinoma cell lines, hence L-fucose may provide a valuable means for active targeting of these drug carriers. Although these compounds were not tested for drug transport into neoplastic cells, if this is possible they should provide a delivery mechanism with tuneable pharmacokinetic and pharmacodynamic properties along with low toxicity[98].

Interestingly, in 2001 a fucose-binding lectin was isolated from serum of the sea bass *Dicentrarchus labrax*[99]; this lectin was subsequently transfected into a variety of cancer cells through an adenovirus vector, and shown to have an anti-proliferative effect due to suppression of anti-apoptotic factors in the neoplastic cells[100]. This lectin may prove a novel means for cancer therapy.

1.2.2.3   Fucose in fertility

L-fucose and α-L-fucosidase have been shown to play critical roles in fertility across diverse species. The importance of L-fucose in this event was first described in 1982, when Huang *et al.* observed potent inhibition of human, hamster and guinea pig sperm-oocyte binding by the L-fucose rich polysaccharide fucoidan.[35] Human semen contains two α-L-fucosidase isoforms.[101] The majority of seminal α-L-fucosidase activity is found in the seminal fluid, less than *ca.* 3% of total activity is associated to the membranes of spermatozoa. Both seminal fluid and spermatozoa α-L-fucosidases from different donors have been observed to have a number of further isoforms (3-6 and 2, respectively) and these enzymes have different activity profiles; whilst seminal plasma samples have a broad activity profile with a number of maxima between pH 4.8 and 7.0, the spermatozoan isoform has a more acidic pH optimum between pH 3.4 and 4.0[101].

Seminal α-L-fucosidases have also been discovered in many mammals such as mice[102], hamsters[103] and also in diverse species such as *Unio elongatulus*[104], *Ascidiacea*[105]and *Drosophilidae*[106], suggesting a key role for the enzyme during fertilisation.

Analysis of the effects of α-L-fucosidase activity on spermatozoa fertility has been reported for hamsters[103] and mice[102]. In hamsters, pre-incubation of spermatozoa with the competitive α-L-fucosidase inhibitor deoxyfuconojirimycin (DFJ[107]) was observed not to inhibit sperm binding to oocyte zona pellucida, the glycoprotein layer surrounding the egg cell.[103] Pretreatment of spermatozoa with DFJ was, however, shown to reduce fertilisation success, preventing development of the embryo past the 2-cell stage. In mice, treatment of sperm with DFJ was shown not to inhibit binding to the zona pellucida or the oocyte membrane, and had no effect on fusion or binding. Pre-incubation of sperm with anti-fucosidase antibody, or pre-treatment of oocytes with purified human liver α-L-fucosidase, however, significantly impaired these events.

Thus, in mice, sperm α-L-fucosidase activity has little or no effect on fertility, however its glycan structures are critical. This is supported by the observation of Oh *et al.* that introduction of the neoglycoprotein BSA-fucose to mouse oocytes inhibits the interaction between their zona pellucida and sperm[108], thus it is likely that this binding event requires the interaction between fucosylated glycans present on mouse sperm α-L-fucosidase and zona pellucida lectins. The fucosylated immunodeterminant sialyl-Le[x] has been shown to be the highest abundance glycan expressed on human oocyte zona pellucida; present in both N-linked and O-linked

glycans, the epitope appears to be critical for spermatozoa-oocyte binding, as glycoconjugates bearing this motif or antibodies raised against it inhibit spermatozoa-oocyte binding.[109]

1.2.2.4   *Fucose in bacterial colonisation*

Many bacteria mediate their colonisation by decoration of their cell surfaces with lectins, allowing binding to glycan epitopes such as the fucosylated A, B, H, Le^a, Le^b, sialyl-Le^a, Le^x, Le^y and sialyl-Le^x antigens (Figure 1.5) expressed on cell surfaces in the gastrointestinal tract.[110]

Mucin proteins are thought to have a role in forming a protective barrier to invading bacteria; these proteins are the main macromolecular component of the mucus that creates a barrier between the eye, trachea, stomach and intestines and pathogens.[17] Mucins are repeat proteins comprising a large number of serine and threonine residues for O-glycan attachment, and are heavily glycosylated, typically comprising more than 70% O-glycans by mass. The O-glycans in mucins include fucosylated structures such as the ABO and Lewis antigens.[17] The role of mucins in defence against pathogens is supported by studies of Mucin knockout in mice, *Muc1*$^{-/-}$ and *Muc2*$^{-/-}$ have been shown to have susceptibility to infection by pathogens.[111,112]

An example of fucosylated antigens playing a role in bacterial adhesion is in the colonisation of the pathogen *Campylobacter jejuni* into the gastrointestinal tract. *C. jejuni* is a commensal in poultry, but is also the most common cause of gastroenteritis in the developed world, causing mild to severe diarrhea which may be concomitant with a fever[113]. The pathogen effects its colonisation with the host by interacting with cell surface glycans, including fucosylated Le^a and Le^x epitopes, and when grown under mammalian-like growth conditions, is a poor binder of non-fucosylated glycans[5]. Further evidence corroborates the importance of L-fucose to the viability of *C. jejuni* as a pathogen; L-fucose acts as a chemoattractant towards the organism[114], and the presence of L-fucose in its environment confers it a competitive advantage [115]. A gene encoding a fucose permease has been identified in *C. jejuni*[115], making it likely that the organism is able to utilise L-fucose as an energy source. The introduction of human milk oligosaccharides comprising fucosylated O-glycans abolishes *C. jejuni* adhesion to the mucosa, presumably through competitive inhibition of *C. jejuni* lectins, allowing clearance of the bacterium through the gastrointestinal tract.[116] Finally, mice deficient in Muc1 show susceptibility to *Campylobacter jejuni* infection.[112]

Fucosylated glycans also play an important role in the pathogenicity of *Pseudomonas aeruginosa*. This bacterium is an opportunistic pathogen that typically infects immunocompromised patients with chronic lung diseases, such as sufferers of cystic fibrosis, for whom infection with *P. aeruginosa* may be fatal. *P. aeruginosa* colonisation of the airway is partially dependent on its interaction with mucin glycans, particularly the sialyl-Lewis[x] antigen.[6] This interaction is effected by the *P. aeruginosa* fucose-binding lectin PAII-L, which has a micromolar affinity for free L-fucose.[117] The X-ray crystal structure of PAII-L complexed with L-fucose[118] explains the basis for this unusually high affinity of a lectin for its ligand; a pair of calcium ions in the fucose binding site aid in complexation of L-fucose through electrostatic interaction.

Some pathogens have adapted systems in order to take advantage of the host immune defence; an example of this is *Helicobacter pylori*, a stomach pathogen infecting approximately one half of the global human population. Colonisation density of this pathogen in chronic gastritis patients is related to the expression of Le[x/y] antigens on its cell surface, with strains expressing more Le[x/y] antigens having higher colonisation density.[119] An insightful study by Liu *et al.* found that *H. pylori* infection stimulates the host immune response, causing production of secreted human α-L-fucosidase (FUCA2).[4] The response of *H. pylori* is to utilise the L-fucose moieties thus released from gastric epithelial cells, both as an energy source and in order to further decorate their cell surface with Le[x/y] antigens for continued adhesion to epithelial cells; FUCA2 appears to be essential for *H. pylori* adhesion to human gastric cancer cells.[4]

Due to the importance of fucosylated glycan-mediated interactions to successful adhesion and colonisation of a number of pathogenic bacteria alluded to above, the use of fucosylated carbohydrates has been suggested as an alternative treatment for bacterial infection.[120] In this approach, fucosylated glycans would compete for host lectins, weakening the adhesion between pathogens and the host and allowing clearance of the pathogens through the gastrointestinal tract. This approach is already used by nature; human breast milk contains free L-fucose as well as a large number of fucosylated oligosaccharides which appear to supplement the nursing infant's resistance to pathogens.[121-123] For example, human milk oligosaccharides have been shown to provide resistance to the diarrheagenic symptoms after *Escherichia coli* heat-stable enterotoxin infection in suckling mice[124], and resistance to *Campylobacter jejuni* as discussed above[116]. The oligosaccharide profile of bovine milk is different to that of human milk, having higher levels of sialylation but lower levels of fucosylation.[121] Due to the lack of L-fucose, bovine

milk formulae may be less beneficial to the development of neonatal infants than human milk, and L-fucose has been proposed as a supplement to infant formulae.

The mucin MUC1 is expressed on normal and malignant T-cells, and has a modulatory role in function of these immune cells[125], further, high levels of serum MUC1 suppress the T cell response[126]. MUC1 co-localises with other mucin-type glycoproteins and signal transducer molecules on the cell surface of T-cell lymphomas[127]; these mucin-rich microdomains form type 2 glycosynapses where signal transduction by *e.g.* Src family kinases is effected after carbohydrate-protein mediated interaction likely between O-glycan linked structures such as sialyl-Le^x and sialyl-Le^a, and selectins[49].

It would be unfair to paint such a one-sided picture of host-microbe interactions; microbe interactions with their host in fact span a spectrum between pathogenicity and symbiosis, and while it is estimated that 90% of the cells in the human body are microbial in nature, the vast majority of these organisms have a commensal or symbiotic relationship with their host[46,128]. Of all of the mammalian organs, the gut has by far the highest microbial density[128], and many microbes that reside in the gut act as symbiotes, serving an important role in supplementing the host's enzymatic machinery for degradation of dietary plant polysaccharides[129]. Our own enzymatic machinery is indeed limited to utilisation of monosaccharides, some disaccharides, and starch. Residents of the gut microbiota thus allow us to digest plant polysaccharides such as xylan, pectin and alginate[130], with bacterial metabolism of otherwise indigestible dietary biomass generating short chain fatty-acids that account for 5-10% of the human daily caloric intake [131].

One of the most widely studied gut symbiotes is *Bacteroides thetaiotaomicron*, an obligate anaerobe and one of the most abundant microbial members of the healthy human intestinal tract.[132] *Bacteroides thetaiotaomicron* has numerous documented beneficial effects on its host. Angiogenesis of the intestinal tract is stimulated by the organism; germ-free adult mice, which display arrested capillary formation, were shown to have their capillary network restored within 10 days of colonisation with *Bacteroides thetaiotaomicron*.[133] *Bacteroides thetaiotaomicron* also promotes the development of a strong mucosal barrier, as colonisation of the organism in mice induces upregulation of *ang4*, which produces a ribonuclease that is bactericidal to some Gram-positive enteropathogens.[134]

In return for providing beneficial effects to its host, *Bacteroides thetaiotaomicron* ensures its own fitness by helping to shape the glycan profile of mucins in the gut to provide nutrition for its growth. *Bacteroides thetaiotaomicron* is able to digest a

number of glycans for growth[130], including fucosylated glycans, which are particularly important to the organism; of the 24 oligosaccharide transporters present in its genome, eight are fucose permeases[135]. Germ-free mice actually lack α1,2 and α1,6-fucosyltransferase activity, which can be reintroduced by their inoculation with *Bacteroides thetaiotaomicron* or extracts from conventionally grown mouse gut microflora.[136] The fucosyltransferase activity thus restored is due to microbiota affecting expression of host enzymes which are downregulated as a function of mammalian development after the weaning stage.[136] Hence, gut microbiota such as *Bacteroides thetaiotaomicron* alter expression levels of their host's enzymatic machinery in order to survive in the competitive environment of the gastrointestinal tract, both by stimulating the host to produce more nutrients required for their growth and by reducing the viability of pathogens that may compete for these nutrients.

Bacterial utilisation of glycans is effected by their expression of genes encoding glycoside hydrolases and glycan processing enzymes, in some bacteria, such as *Bacteroides thetaiotaomicron*, these are organised into polysaccharide utilisation loci (PULs). These PULs allow bacteria to efficiently forage for available nutrients and eliminate the need for constitutive expression of many of their enzymes. The archetypal PUL of *Bacteroides thetaiotaomicron* is the starch utilisation system (sus, reviewed by Martens *et al.* (Figure 1.7).[137] The sus gene locus contains eight genes, five of which transcribe outer membrane-associated proteins. SusDEF are carbohydrate binding proteins that interact with starch while SusG is an endo-acting α-amylase that breaks the polysaccharide down into smaller carbohydrates. These smaller glycans are transported across the outer membrane by the TonB-dependent receptor SusC and further degraded by glycoside hydrolases SusA and SusB, which have neopullulanase and α-glucosidase activities, respectively.[138] Finally, SusR is a hybrid two-component sensor/regulator that, in the presence of starch degradation products in the periplasm, transduces a signal for up-regulation of expression of *sus*ABCDEFG.[138] The presence of consecutive susC and susD-like genes are the minimum features needed to identify a sus-like PUL, and using these features, 88 sus-like PULs have been identified in *Bacteroides thetaiotaomicron*[139]. PULs likely provide an evolutionary advantage to *Bacteroides thetaiotaomicron* by allowing its adaptation to the rapidly changing environment of the distal intestine by coordination of the energy-intensive process of protein production with the available nutrient pool.

A sus-like system of *Bacteroides thetaiotaomicron* that is induced in response to free fucose was identified by Hooper *et al.*[140] This operon comprises 5 genes,

fucRPIAK. The fucPIAK genes encode, respectively, an L-fucose permease, L-fucose isomerase, L-fuculose-1-phosphate aldolase and L-fuculose kinase. In parallel to susR, fucR acts as a repressor of the *fucRIAK* genes until displaced by the presence of L-fucose, but does not regulate its own expression. Evidence also suggested that fucR acts as an L-fucose inducible co-repressor at another gene locus, which stimulates α(1,2)-fucosyltransferase activity by the host, corroborating the findings of Bry *et al*. discussed above[136]. Reflecting the evolution of *Bacteroides thetaiotaomicron* as a symbiote, the bacterium appears to ensure that there remains a source of dietary fucose in its environment, while also limiting the pressure it exerts on its host.
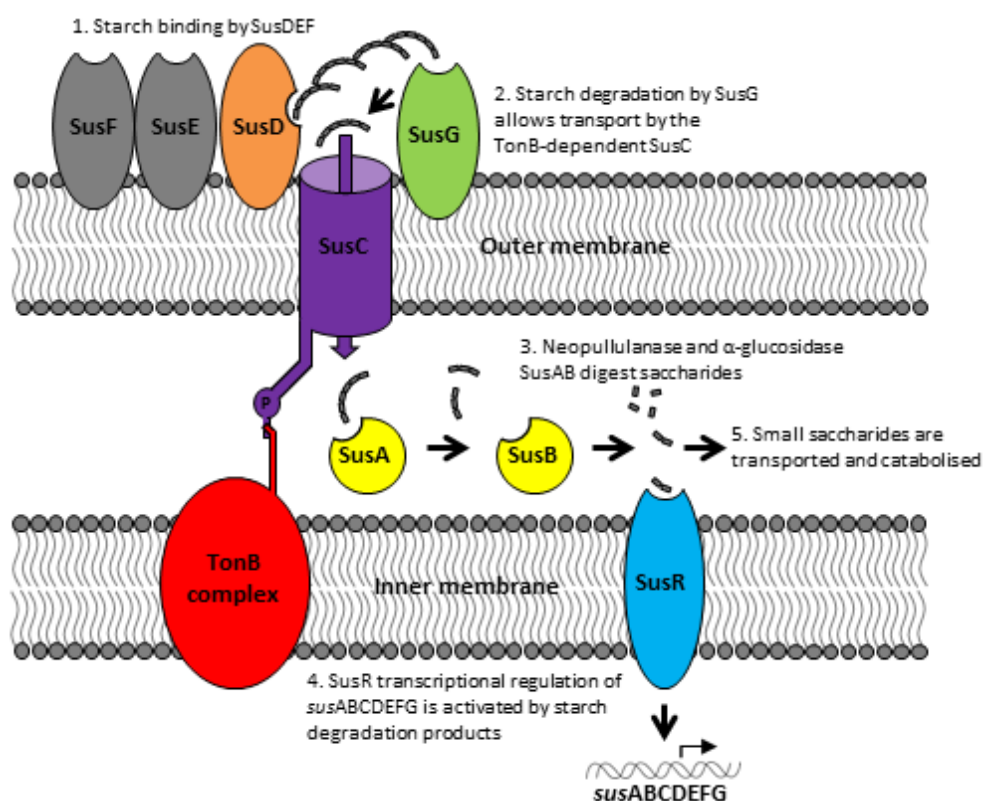
Figure 1.7. Schematic of the *Bacteroides thetaiotaomicron* starch utilisation system.

Interestingly, stress induced in mice through application of lipopolysaccharides has been correlated with a rapid increase in α(1,2)-fucosylaton of intestine epithelial cells by Fut2[141]; *Fut2*[-/-] mice were shown to require significantly more time for recovery of body weight after induced stress. Gnotobiotic mice under induced stress, or antibiotic-treated normal mice also showed slow recovery, indicating the importance of the microbiota in supplementing the host's immune system during systemic stress. This is supported by the observation that *Fut2*[-/-] mice subjected to a

non-lethal pathogenic bacterium before induction of stress showed significantly slower weight recovery than Fut2-sufficient mice. Taken together, these results would seem to indicate that up-regulation of α(1,2)-fucosylation is a measure taken by the host to maintain the viability of its gut commensals as a response to systemic stress.[141]

Due to the myriad roles of carbohydrates in nature, *vide supra*, it is clear that carbohydrate metabolism and homeostasis are critical biological processes. Correct homeostasis of glycan structures requires the concerted action of two key classes of enzymes; glycosyltransferases, which form glycosidic bonds, and glycoside hydrolases, which cleave those bonds. Both fucosyltransferases and fucosidases are of clinical importance due to various disease phenotypes which occur due to their malfunctioning or absence.

1.3 *Glycosyltransferases*

Glycosyltransferases catalyse the transfer of sugar donors, activated in a nucleoside phosphate, lipid phosphate, or un-substituted phosphate form, to acceptor moieties. Saccharides may in this way be appended to, *inter alia*, other sugars, proteins, and lipids.[7] The glycosyltransferase genes expressed by an organism, and their expression levels, are subject to change as a result of transition through developmental stages and environmental effects such as the onset of disease, and changes in the expression levels of glycosyltransferases contribute to the diversification of glycans.[142] Owing to their biologically critical nature, 1 % to 2 % of the genomes of an organism, regardless of domain, encode glycosyltransferase enzymes.[7]

Catalysis by glycosyltransferases leads to either inversion or retention of the chirality at the anomeric carbon of the donor sugar. While the mechanism of inverting glycosyltransferases is well understood and proceeds via an S_N2-like mechanism, that of retaining glycosyltransferases is less clear, but may proceed through an oxocarbenium ion-like intermediate with basic assistance of the donor phosphate group.[7]

1.3.1 Fucosyltransferases

Fucosyltransferases (EC numbers 2.4.1.65, 2.4.1.68, 2.4.1.69, 2.4.1.152, 2.4.1.214) catalyse the transfer of an activated GDP-fucose donor to an acceptor moiety, which can be a protein Ser or Thr residue through an α1-(S/T)-linkage, Galactose through an α(1,2)-linkage or N-Acetylglucosamine through an α(1,3), α(1,4) or α(1,6)-linkage.[143] A large repertoire of fucosyltransferases exists in the human genome due to the requirement to produce a diverse array of fucose-containing glycans. Each enzyme is thus adapted for the synthesis of a specific glycosidic bond from GDP-fucose to a specific or small subset of potential acceptor moieties, and expression in different cell types.[144]

Activation of fucose to a GDP-fucose form, which is required for fucosyltransferase reactivity, takes place through two pathways in mammals.[20,145] The major, *de novo* pathway accounts for more than 90 % of GDP-fucose biosynthesis; this pathway involves conversion of GDP-mannose to GDP-fucose through the action of two key enzymes and a GDP-4-keto-6-deoxymannose intermediate. The alternative pathway for GDP-fucose biosynthesis is termed the Salvage pathway and involves

phosphorylation of free fucose and subsequent condensation with GTP to yield GDP-fucose.

Currently, thirteen ORFs from the human genome have been annotated as coding for translation of proteins with fucosyltransferase activity. The majority of these proteins are grouped into glycosyltransferase family 10 according to the Carbohydrate Active Enzymes database, CAZy, available at http://www.cazy.org (GT10, FUT3-7, FUT9-11).[146,147]

FUT1 and 2 are α(1,2)-fucosyltransferases involved in biosynthesis of the blood group H antigen as described above.[148,149] The α(1,3)-fucosyltransferases FUT3-7 and FUT9 are differentially expressed in various cell types, leading to different patterns of fucosylated ligand biosynthesis in these cells; three distinct fucosylation patterns exist, myeloid-type patterns in leukocytes and brain, plasma-type patterns in plasma and liver, and Lewis-type patterns in bladder, kidney and milk.[144]

Fucosyltransferases 3, 5 and 6 only exist in higher-order mammals and evolved relatively recently; these proteins have a high sequence similarity with each other and are clustered together on chromosome 19 in humans.[150] These enzymes, along with fucosyltransferases 4, 7 and 9 all transfer L-fucose to GlcNAc in α(1,3) or α(1,4) linkages in various cell types.[151] FUT3 is also called the Lewis gene and has both α(1,3) and α(1,4)-fucosyltransferase activity; this enzyme is involved in the biosynthesis of Lewis antigens in epithelial cells.[152] FUT5 is also expressed in epithelial cells, but only has α(1,4)-fucosyltransferase activity. In gastric cancer cells, FUT3 has been shown to be involved in the synthesis of the Lewis b, a and y antigens, while FUT5 synthesises the Lewis x antigen; the combined action of these two enzymes appears to be necessary for biosynthesis of the sialylated antigens sialyl-Le^x and sialyl-Le^a [153].

FUT4 and FUT7 are α(1,3)-fucosyltransferases involved in the biosynthesis of selectin ligands in leukocytes and endothelial cells[154,155]; FUT7 deficiency has been shown to cause leukocytosis in mouse[66], and the enzyme only appears to catalyse the transfer of L-fucose to oligosaccharide acceptors already containing sialic acid[156]. FUT6 is a plasma-type α(1,3)-fucosyltransferase.[157]

FUT8 is unique in that it catalyses the transfer of L-fucose in an α(1,6)-linkage to the internal GlcNAc in biantennary glycans terminating with the same monosaccharide.[158,159] This fucosylation is also known as core fucosylation, and the expression of FUT8 appears to have an effect on tumour viability in liver[160] and prostate cancer[161] and indeed is upregulated in most cancers[73]. The substrate

25

specificity of FUT9 is different from the α(1,3)-fucosyltransferases FUT3-7, as the enzyme preferentially fucosylates the distal GlcNAc moiety of a polylactosamine acceptor, while FUT3-7 preferentially glycosylate an internal GlcNAc.[156] FUT10 may have a role in maintaining stem cell populations; overexpression of the α(1,3)-fucosyltransferase was shown to enhance the self-renewal of neural stem cells, whilst knockdown of the enzyme induced their differentiation.[162] To the author's knowledge, nothing is known about the substrate specificity or tissue distribution of FUT11.

Finally, POFUT1 and POFUT2 catalyse the transfer of L-fucosyl moieties to serine or threonine residues on the polypeptide repeat of proteins. POFUT1 is involved in the addition of fucose to epidermal growth factor type repeats of proteins such as Notch, *vide supra*, while POFUT2 transfers the unit to thrombospondin 1 type repeats.[163]

1.4   Glycoside hydrolases

Glycoside hydrolases effect the enzymatic cleavage of glycosidic bonds. The glycosidic bond is incredibly stable without catalytic aid, and is indeed more stable to degradation than the bonds present in such biological polymers such as DNA and proteins.[13-15] The rate enhancement provided by glycoside hydrolase involvement in cleavage of the glycosidic bond is huge, *e.g.* comparison of the rate of catalysis of starch by sweet potato β-amylase and the un-catalysed rate yields a rate enhancement of 10^{17}.[13]

Glycoside hydrolases, as with glycosyltransferases, may proceed with either inversion or retention of chirality at the sugar substrate's anomeric centre. Whilst other mechanisms exist, the majority of glycoside hydrolases utilise two catalytic carboxylates and proceed through either a single S_N2-like displacement (inverting) or two consecutive S_N2-like displacements (retaining) as originally proposed by Koshland.[164] In the canonical inverting mechanism, a basic carboxylate deprotonates water on the bottom of the glycoside ring (Figure 1.8 **A**). The activated hydroxyl then performs in-line nucleophilic attack on the anomeric carbon, whilst the leaving group is stabilised by protonation by the catalytic acid (Figure 1.8 **A**). The protonation states of the catalytic base and acid may then be regenerated through solvent. The product released thus has inverted stereochemistry compared to that of the substrate glycoside. A distance of *ca.* 10 Å typically exists between the catalytic base and acid residues, reflecting the steric requirements of allowing enough space to accommodate a water molecule on the bottom face of the glycoside ring.[165]

The canonical retaining glycoside hydrolase mechanism involves two S_N2-like displacements, and formation of a covalent enzyme intermediate. Here, the catalytic nucleophile itself performs in-line nucleophilic attack to a bound glycoside, with leaving group departure aided by the catalytic acid/base, leading to the formation of a covalent glycosyl-enzyme intermediate (Figure 1.8 **B**). Subsequently, the catalytic acid/base acts as a base, activating a water molecule as a hydroxyl nucleophile to act on the trapped glycoside and release the catalytic nucleophile (Figure 1.8 **B**). In retaining glycoside hydrolases with this mechanism, a distance of *ca.* 5.5 Å is observed between the catalytic residues, and the reaction product has the same stereochemistry as the substrate glycoside.
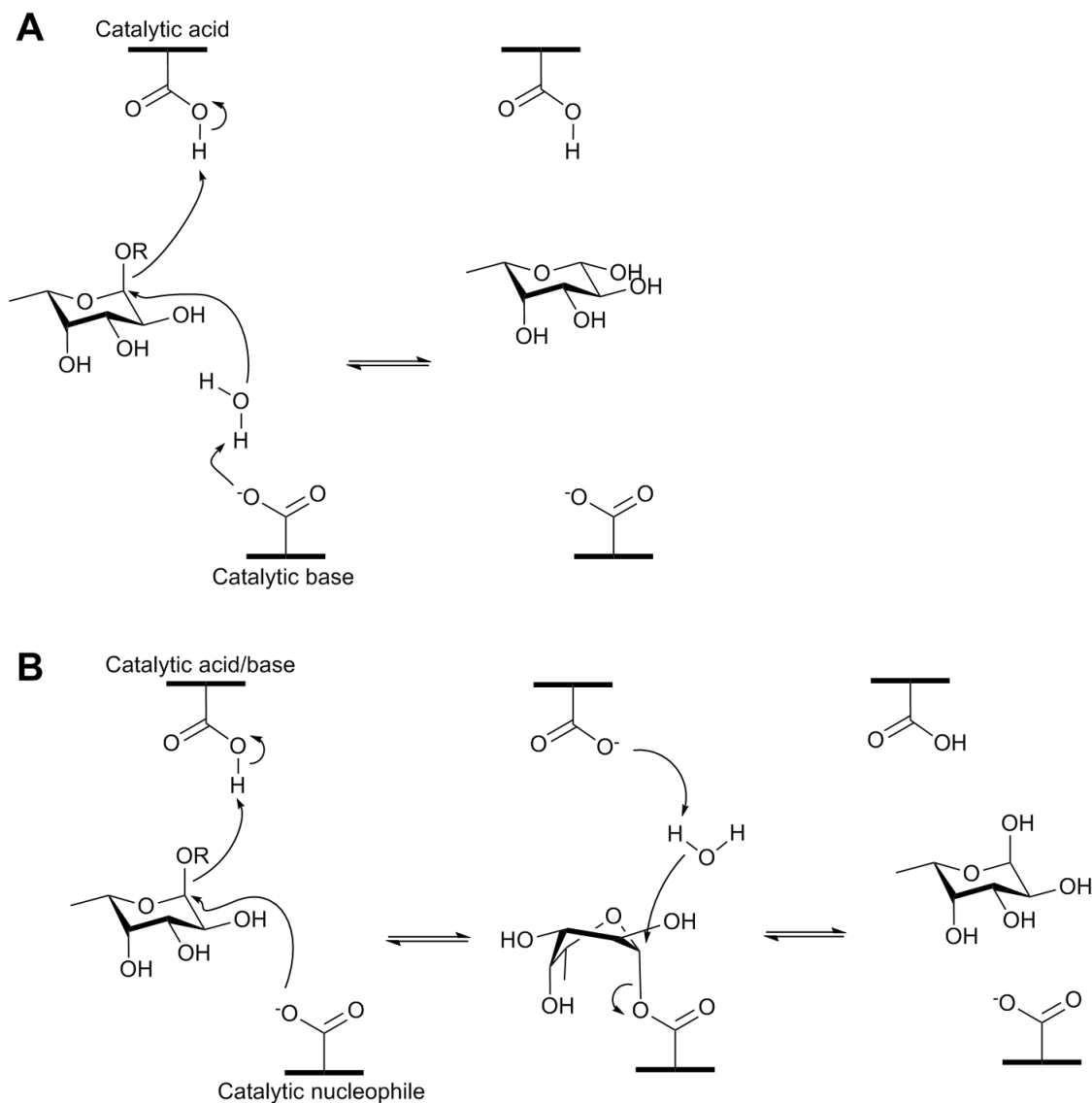
Figure 1.8. Canonical inverting (**A**) and retaining (**B**) mechanisms of glycoside hydrolases. Figures are drawn with the configuration of and conformation adopted by L-fucose.

Carbohydrates adopt a variety of conformations *in vitro*. For pyranose ring structures, a single chair form of the pyranose is typically the most energetically favourable and represents the conformational majority of the sugar in solution. The enzymatic transition states in glycoside hydrolases (there are two in retaining enzymes and one in inverting enzymes) contain susbstantial oxocarbenium ion-like character, with a degree of positive charge at the anomeric carbon due to cleavage of the glycosidic bond. The glycosyl carbocation is incredibly unstable in solution, it is believed that this species only exists through stabilisation of the oxocarbenium ion intermediate through electron donation by the endocyclic oxygen.[166] For this to occur, the glycosidic bond must be situated anti-periplanar to an sp^3 lone pair on

the endocyclic oxygen; this geometry requires co-planarity of the endocyclic oxygen, the anomeric carbon, and the atoms adjacent to these *i.e.* the C2 and C5 atoms. Thus, only four possible conformations whereby nucleophilic substitution may occur exist for pyranose sugars, namely the two boat and four half-chair conformations $^3H_4$, $^4H_3$, $^{2,5}B$ and $B_{2,5}$ (Figure 1.9).[166]
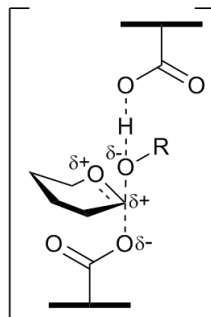
Figure 1.9. The transition state in glycoside hydrolases. The glycoside hydrolase transition state contains substantial oxocarbenium ion-like character and is stabilised by electron donation from the endocyclic oxygen sp^3 lone pair to the anomeric carbon. While only the glycosylation step of retaining glycoside hydrolysis is displayed, the electronic character of the deglycosylation step of this reaction and that of the inverting reaction are identical.

Glycoside hydrolases have evolved the capacity to stabilise conformations close to the enzymatic transition state by electronic and steric means, and the evolution of each glycoside hydrolase family appears to be directed towards the stabilisation of a particular reaction pathway involving a discrete $^3H_4$, $^4H_3$, $^{2,5}B$ or $B_{2,5}$ transition state.[167] As early as the 1940's, it was postulated by Pauling that the most potent inhibitors of an enzyme would be those that mimicked the 'strained activated complex' formed during catalysis.[168] Indeed, the distortion of the glycoside to be cleaved in glycoside hydrolase catalysis has been exploited both by nature and man through the generation of enzyme inhibitors that are able to mimic the geometrical and electronic properties of the enzymatic transition state.[169]

1.4.1 α-ʟ-fucosidases

α-ʟ-fucosidases (EC numbers 3.2.1.51, 3.2.1.63 and 3.2.1.111) catalyse the hydrolysis of the glycosidic bond of an α-ʟ-fucoside from the non-reducing end of a glycoconjugate. These enzymes primarily exist in the GH29 and GH95 families of the CAZy system for classification of carbohydrate active enzymes, which groups

carbohydrate-active enzymes into families based on their primary sequence.[170] Some α-ʟ-fucosidases have been reported which do not belong to GH29 or GH95 families, in 2002 an α-ʟ-fucosidase was identified from *A. thaliana* with activity against 2'-fucosyl-lactitol and xyloglucan XXFG polysaccharide but no homology to GH29 or GH95 enzymes, this enzyme is indeed currently 'non-classified' according to CAZy[171]. More recently a novel α-ʟ-fucosidase from *Paenibacillus thiaminolyticus* was characterised and reported to have no homology with GH29 or GH95, this enzyme does not currently have an entry in the CAZy database.[172]

Two α-ʟ-fucosidases exist in the human genome, FUCA1 and FUCA2. While FUCA1 is localised in the lysosome, FUCA2 is a secreted protein localised in plasma; both enzymes are classified in the GH29 family. Lack of α-ʟ-fucosidase (FUCA1) activity in the human lysosome can lead to the rare (*ca.* 100 cases known worldwide) storage disorder fucosidosis in man.[173] This disorder leads to accumulation of fucosylated N-glycan structures which cannot be broken down by the host's impaired catabolic pathway, the disease causes severe neurodegenerative symptoms, cachexia and low life expectancy[174]. Fucosidosis has also been reported in canines[175] and felines[176], and a wide spectrum of genetic FUCA1 mutations leading to fucosidosis in humans has been reported[177]. Gene therapy by retroviral insertion of FUCA1 into both canine and human fucosidotic fibroblasts was able to not only restore normal levels of α-ʟ-fucosidase function, but indeed such fibroblasts were able to hydrolyse radionuclide-labelled ʟ-[5,6-^{3}H]-fucosylated glycans with greater efficiency than non-fucosidotic fibroblasts.[178]

1.4.1.1 *GH95 α-ʟ-fucosidases*

The catalytic mechanism for GH95 enzymes is thought to be less canonical than that typical of inverting glycoside hydrolases. During the study of *Bifidobacterium bifidum* 1,2-α-ʟ-fucosidase, Nagae *et al.* found that while residue Glu^{566} was identified by X-ray crystal structure to lie in an appropriate position to act as the catalytic acid, and further confirmed by mutagenesis experiments, no carboxylate was observed in an appropriate position in X-ray crystal structure to act as the enzymatic base.[179] Rather, the authors concluded that a conserved structural motif creates an electrostatic relay in the enzyme to modulate the pK_a of Asn^{423} for its deprotonation of water (Figure 1.10). Further, this water molecule was shown to be held in the correct position and orientation for in-line attack on the anomeric carbon by hydrogen bonding interactions by the catalytic base Asn^{423} and another asparagine residue Asn^{421}.[179]
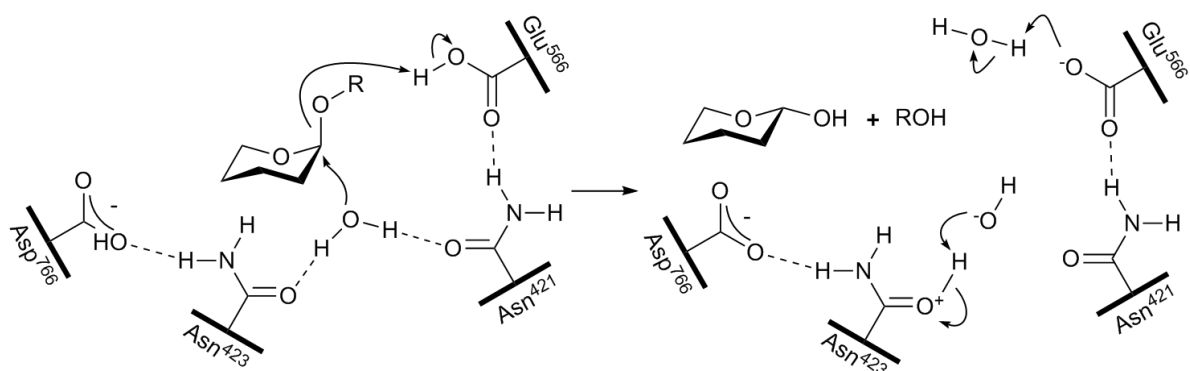
Figure 1.10. Schematic of the proposed electrostatic relay and catalytic mechanism postulated for GH95 enzymatic catalysis.[179] L-fucose ring substituents other than that at the anomeric carbon are omitted for clarity.

1.4.1.2 GH29 α-L-fucosidases

GH29 α-L-fucosidases act with net retention of the anomeric configuration after hydrolysis of a fucosylated substrate. In 2003, two research groups contemporaneously identified the catalytic nucleophile of α-L-fucosidases from *Thermotoga maritima* and the archaea *Sulfolobus solfataricus*.[180,181] This was achieved for the archaeal enzyme by chemical rescue of α-L-fucosidase activity of an inactive Asp^{242} mutant using sodium azide[181]. For the bacterial enzyme, identification of the catalytic nucleophile was achieved through covalent inactivation of the enzyme using an irreversible α-L-fucosidase inhibitor 2-deoxy-2-fluoro-α-L-fucosyl fluoride, enabling identification of the catalytic nucleophile Asp^{224} by peptic digest and mass spectrometry.[180] The fact that a covalent enzyme intermediate could be trapped is evidence that GH29 enzymes act through a classical Koshland double-displacement mechanism[164] involving, first, the attack of a nucleophilic carboxylate residue to form a covalent enzyme-substrate intermediate in the 'fucosylation' step (Figure 1.8 **B**). Aglycon departure is aided by a catalytic acid/base Glu residue. The catalytic acid/base Glu then activates water to regenerate fucose and the catalytic nucleophile in the 'defucosylation' step. Notably, the first X-ray crystal structure of an α-L-fucosidase was determined using the *Thermotoga maritima* enzyme (*Tm*GH29) in 2004; the catalytic, N-terminal domain of *Tm*GH29 adopts a $(\beta\alpha)_8$ TIM barrel fold and forms a hexamer in both crystal structure and solution (as determined by size exclusion chromatography, Figure 1.11).[182]
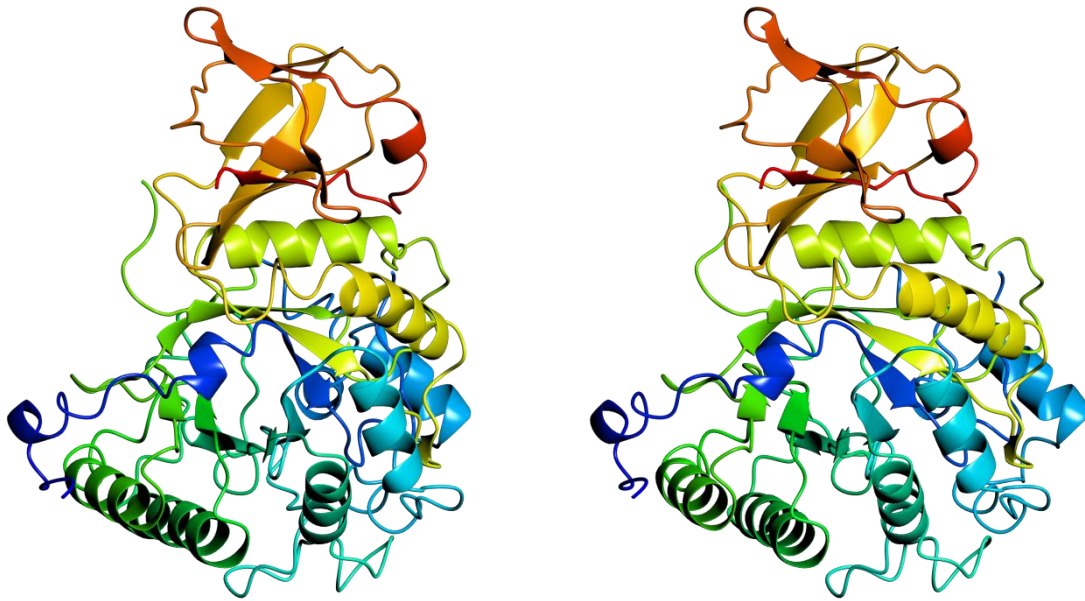
Figure 1.11. 3 dimensional fold of the GH29 α-ʟ-fucosidase from *Thermotoga maritima* as solved by X-ray crystal structure (PDB code 1HL8). The figure was drawn in wall-eye stereo with model colour blended through the protomer chain, from the blue N-terminus to the red C-terminus. The figure was drawn using CCP4MG[183].

Subsequently, a number of crystal structures from other members of the GH29 family have been reported, *i.e.* those from *Bacteroides thetaiotaomicron* (two have been reported, gene products from genes 2970[184] and 2192[185]), *Bifidobacterium longum* subsp. *infantis*[186,187], and the eukaryotic plant pathogen *Fusarium graminearum*[188]. Further, two GH29 crystal structures have been deposited in the protein databank but have not been supported by literature as of yet, *i.e.* those with PDB codes 4ZRX and 3GZA.

From the availability of GH29 crystal structures and related kinetic data, a number of observations can be made. First, the enzymes typically form dimers or higher order oligomers both in solution and in crystal structure (however the eukaryotic enzyme was observed not to form oligomers[188]). Second, while the catalytic nucleophile in GH29 enzymes is an Asp residue that is conserved by sequence, the catalytic acid/base Glu residue is not.[189] Third, the conserved catalytic $(\beta\alpha)_8$ α-ʟ-fucosidase domain appears at the N-terminus of the gene transcript, while other domains may or may not be present; *e.g.* *Tm*GH29 has a C-terminal β-sandwich domain[182] whilst *Fg*GH29 has a C-terminal βγ-crystallin domain[188] and *Bi*GH29 contains a number of C-terminal domains including a carbohydrate binding domain. Finally, the substrate specificity of this class of enzymes varies across its members,

while some enzymes are able to hydrolyse fucose in linkages to the 2,3,4 and 6-hydroxyls of reducing end sugars and work efficiently on the synthetic substrate *p*-nitrophenyl-α-L-fucopyranoside, others seem to act specifically on α(1,3/4)-linkages and have poor to no activity on *p*-nitrophenyl-α-L-fucopyranoside. This led to the designation of two subfamilies of GH29 α-L-fucosidases by Yamamoto and co-workers, where GH29-A enzymes have relaxed substrate specificity those from the GH29-B sub-family are α(1,3/4)-linkage specific.[190,191] The existence of two sub-families of GH29 is further evidenced by phylogenetic analysis, with two clades being comprised GH29 enzymes from animals, archaea and bacteria (GH29-A) and those from plants and bacteria (GH29-B).[190]

The structural basis for the different substrate specificity between the two GH29 sub-families has been probed by analysis of the crystal structures of two representative enzymes from *Bacteroides thetaiotaomicron*[191]. Sakurama *et al.* noted the absence of two residues involved in hydrogen bonding with the fucoside substrate in *Bt*Fuc2192 from GH29-B, while *Bt*Fuc2970 had a (conserved among the GH29-A subfamily) Arg residue that may provide favourable leaving group stabilisation for *p*-nitrophenyl-α-L-fucopyranoside hydrolysis. *Bt*Fuc2192 was observed to contain a carbohydrate binding pocket comprised five residues, four of which appear conserved in the GH29-B sub-family. The substrate specificity of *Bt*Fuc2192 further appeared to be limited to saccharides of length greater than 2, indicating a requirement for branched Gal moieties in the substrate for hydrolysis. Overall, the results would seem to indicate that GH29-A enzymes tightly bind a docking fucose moiety in their -1 sub-site, while enzymes from the GH29-B subfamily form weaker interactions with the -1 fucose residue but have enough space to accommodate a branched Gal residue[191]. The importance of the branched Gal residue in GH29-B enzymes is further evidenced by observation of clear electron density for the Fuc and Gal monosaccharides in a complex of a catalytically inactive mutant of *Bi*GH29 with lacto-N-fucopentaose II; electron density was poor for the GlcNAc moiety and absent for the reducing-end lactose disaccharide[187].

The *Bi*GH29 enzyme was observed crystallographically to undergo a large induced-fit conformational change of two loop regions on substrate binding.[187] The catalytic acid/base residue Glu[217], which is separated from the catalytic nucleophile by almost 21 Å in the substrate-free complex, displaces during a loop rearrangement on substrate binding to an active conformation, closing the active site pocket. The flexibility of loops around the enzymatic active site, including that containing the catalytic acid/base residue may be a conserved feature in GH29 fucosidases; in the

crystal structure of *Tm*GH29, no electron density is observed for a loop region which starts only 3 residues downstream of the catalytic acid/base[182]. Also, in the X-ray crystal structure of *Bt*Fuc2970 D229N mutant complexed with synthetic substrate *p*-nitrophenyl-α-ʟ-fucopyranoside (PDB code 2WVU) a conformational displacement of the catalytic acid/base E288 is observed compared to the *apo* structure.[184] It should be noted, however, that this residue lies in an inactive conformation as there is a *ca.* 6.5 Å distance between the glycosidic bond and the oxygen atoms of this carboxylate, and the rearrangement may be fortuitous and due to the hydrogen bonding interactions formed by the nitro-group of the substrate and Arg[262]. Finally, two discrete conformational states have been observed crystallographically for the GH29 enzyme from *Fusarium graminearum*.[188] The conformational dynamics in loop regions may be important for GH29 catalysis, and may further provide a structural basis for the difficulty in predicting the catalytic acid/base residues of this enzyme class *a priori*.[189]

The inhibition of human GH29 enzymes may have clinical potential. *H. pylori* adhesion has been shown to be correlated with expression of FUCA2 in the gut, and inhibition of this enzyme has been shown to abolish *H. pylori* virulence[4], thus inhibitors of this enzyme class may be useful in clearance of the pathogen. Compounds that inhibit α-ʟ-fucosidase may also have utility for chaperone therapy of fucosidosis, which, as discussed above, occurs due to deficiency in α-ʟ-fucosidase activity. In this approach, compounds that inhibit an enzyme are introduced into the cell and at low concentrations seem to activate these enzymes, presumably by promoting their correct folding. After transport of the chaperoned enzyme to their target compartment, the inhibitor is outcompeted by the enzymatic substrate and function is thus restored (Figure 1.12). This approach has been used for a number of different enzymes including both α- and β-galactosidases.[192,193] The approach has had some success in the treatment of Fabry disease[194] and is currently in phase 3 clinical trials[195]. A multitude of GH29 α-ʟ-fucosidase inhibitors have been described in the literature, which will be discussed later in this work.
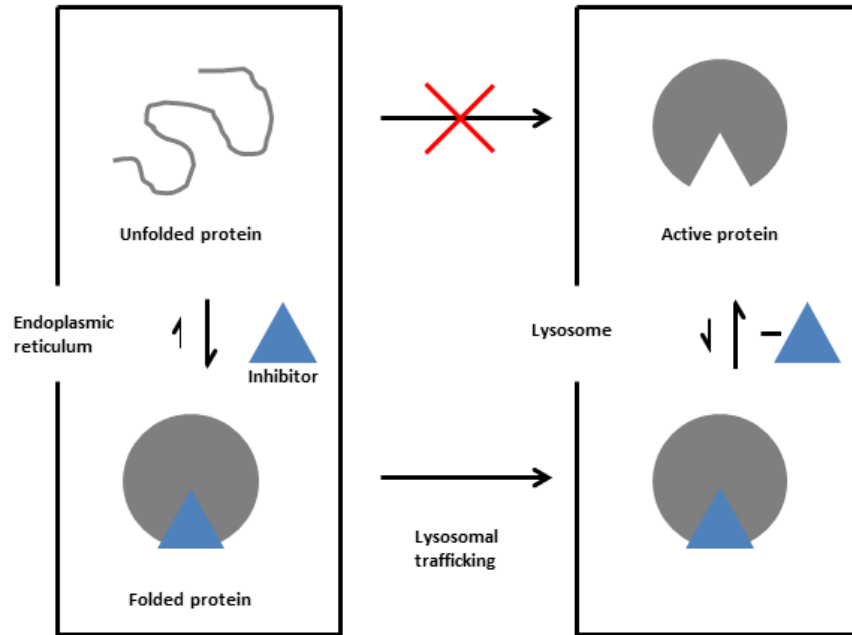
Figure 1.12. The mechanism postulated whereby chemical chaperone therapy may restore the function of misfolded lysosomal enzymes.

Unfortunately, despite the wealth of structural information that has been reported for α-ʟ-fucosidases from various sources, *vide supra*, no such information is available for the enzymes from man, FUCA1 and FUCA2.

1.5   Summary and thesis aims

Carbohydrates serve myriad critical roles in nature, encompassing *e.g.* fertility, development, diet, and the immune response. As such, genetic defect in carbohydrate processing leads to many diseases, and non-hereditary diseases often display altered profiles in constitutively expressed carbohydrate-containing structures.

Fucose, from the rare series of L-sugars, is incorporated into structural extracellular matrix carbohydrates, has roles in development due to its involvement in the Notch signalling pathway, and is an important monosaccharide in formation of the blood group antigens. As the sugar is incorporated into the branch termini of N- and O-linked glycans, it is a critical immuno-determinant. Regulation of the critical enzymes involved in formation and cleavage of L-fucosylated glycans (GDP L-fucosyltransferases and α-L-fucosidases) is differentiated as a result of a number of disease phenotypes notably including cancer.

The subject of the following work is the GH29 (CAZy) family α-L-fucosidases, with particular aim to characterise the structural and mechanistic features necessary for potent inhibition of the enzyme and, ultimately, gain insight which may helpful for treatment of disease phenotypes associated with aberrant regulation of the enzyme *e.g. H. pylori* infection, fucosidosis and cancer. Whilst a number of bacteria have many homologous copies of this enzyme class, only two enzymes exist in humans, the lysosomal FUCA1 and plasma FUCA2; unfortunately no crystallographic structural information is available for these critical enzymes. Due to this lack of structural information on human α-L-fucosidases, the GH29 enzyme from *Bacteroides thetaiotaomicron* (*Bt*Fuc2970) will feature heavily in this work as a surrogate for the human enzyme.

In the next chapter, the mechanism whereby five-membered iminocyclitols may inhibit GH29 enzymes will be studied by a combination of X-ray crystallography and kinetic studies. Subsequently, ferrocene-containing α-L-fucosidase inhibitors will be studied; due to the cytotoxic effects of the ferrocenium ion, these compounds may have utility as anti-cancer therapeutics. This third chapter will also study the potential for enhancements in binding affinity towards α-L-fucosidase by small molecule α-L-inhibitors possessing multiple fuco-configured moieties. In chapter 4, the mechanism of irreversible α-L-fucosidase inhibition by compounds containing aziridine moieties as an electrophilic trap will be studied; these compounds provide a template for inhibitors with utility as activity-based probes. Finally, efforts towards the generation of crystallisation-competent constructs of

human fucosidases FUCA1 and FUCA2 will be described before the final chapter wherein conclusions will be drawn from the work reported in this study.

Chapter 2: Inhibition of bacterial fucosidase *Bt*Fuc2970 by 5-membered iminocyclitols*

Abstract

GH29 α-ʟ-fucosidases are of importance due to their roles in maintaining homeostasis of N- and O-linked glycan structures and in fertility. Inhibitors of this enzyme class may prove clinically useful, for example for treatment of the lysosomal storage disorder fucosidosis and clearance of the stomach pathogen *Helicobacter pylori* from the gastrointestinal tract. Whilst a number of structural studies have been reported highlighting the interaction of GH29 enzymes with 6-membered inhibitors, this is not so for 5-membered inhibitors, that prove to be potent inhibitors of this enzyme class with inhibition constants in the µM-nM range. Herein, kinetic and structural parameters for the binding of seven 5-membered iminocyclitol inhibitors to a GH29 fucosidase from *Bacteroides thetaiotaomicron* (*Bt*Fuc2970) are reported. The compounds are shown to possess K_is for *Bt*Fuc2970 in the µM range. Crystal structures of the compounds at resolutions of 1.59 - 2.10 Å serve to highlight the mode of binding of this class of compounds to fucosidase through an E_3 conformation, mimicking the $^3H_4$ conformation for the transition state of catalysis. Ring contraction of these compounds compared to 6-membered inhibitors is shown to eliminate hydrogen bonding interactions. These crystal structures further serve to highlight where the aglycon moieties of natural substrates are likely to bind to GH29-A family enzymes. The insights gained in this work may aid in the design of a new series of more potent fucosidase inhibitors.

2.1 Introduction

GH29 α-ʟ-fucosidases are of critical biological importance, owing in part to their maintenance of a number of ʟ-fucosylated glycan structures such as the core fucose of Notch protein[196] and histo-blood group antigens[20], and also to their roles in fertility[109] (Figure 2.1). Deficiency in α-ʟ-fucosidase activity in mammals leads to the lysosomal storage disorder fucosidosis[3], while the enzyme is differentially regulated in a number of cancers[93,197]. Improved understanding of the biological activity of human GH29 enzymes would thus be clearly advantageous in the diagnostic and therapeutic contexts. Inhibitors of GH29 α-ʟ-fucosidases have already been shown to have the ability to modulate adhesion of the stomach pathogen *Helicobacter pylori*[4], and may have utility in treatment of fucosidosis through chaperone therapy[193]. Unfortunately, in spite of the potential benefit it may have on human health and disease, no crystal structure is currently available for human GH29 enzymes.

In this light, *Bt*Fuc2970 α-ʟ-fucosidase provides a model system for the study of α-ʟ-fucosidase inhibition. Whilst *Thermotoga maritima* α-ʟ-fucosidase (*Tm*GH29) has a greater primary sequence identity to human α-ʟ-fucosidases (Appendix 1), this enzyme contains an extended active site loop that is absent in FUCA1 and FUCA2 and is likely to affect substrate binding and aglycon specificity[182]. *Bt*Fuc2970 shares 27% and 28% sequence identity with human α-ʟ-fucosidases FUCA1 and FUCA2, respectively[198].

*Bt*Fuc2970 is on polysaccharide utilisation locus 44 of the *Bacteroides thetaiotaomicron* genome (Figure 2.2). The substrate that regulates the expression of genes on this locus is currently unknown. The locus contains open reading frames for 11 gene products. Five of these genes remain unannotated whilst the others transcribe proteins annotated as a family PL8 chondroitin AC lyase (*Bt*2964), a GH2 family enzyme (*Bt*2969), a GH29 α-ʟ-fucosidase (*Bt*2970), a SusC/D pair (*Bt*2958 and *Bt*2957 respectively) and a HTCS sensor/regulator (*Bt*2971)[199].
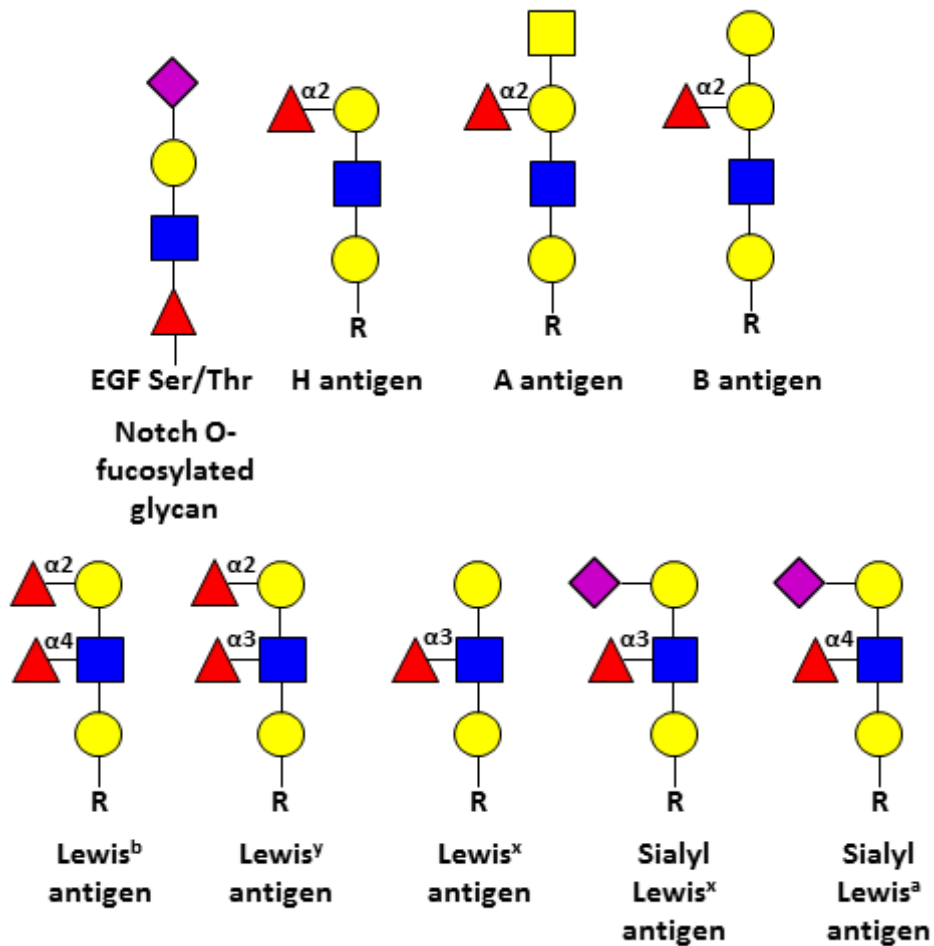
Figure 2.1 Some fucosylated structures that are important in biology. Monosaccharides are displayed by the standard proposed by Varki *et al*. in 2009 (Yellow circle = galactose, yellow square = GalNAc, blue square = GlcNAc, purple diamond = sialic acid, red triangle = fucose).[51]
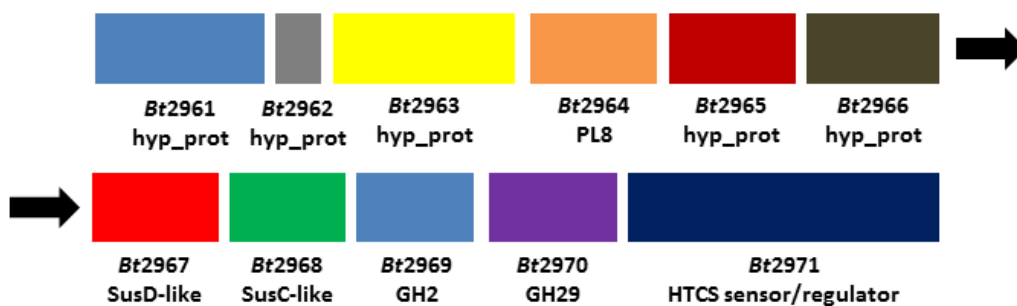
Figure 2.2. Polysaccharide utilisation locus 44 of the genome of *Bacteroides thetaiotaomicron* VPI-5482.

The genome of *Bacteroides thetaiotaomicron* contains a total of 9 genes currently annotated as encoding GH29 fucosidases; these enzymes are thought to be involved in the catabolism of exogenous fucosylated glycans. The organism indeed has a complex system for L-fucose metabolism thought to involve two discrete genetic loci[136,140]. The large number of GH29 enzymes possessed by *Bacteroides thetaiotaomicron* likely represents the need for the organism to digest the large number of glycosidic linkages and polysaccharides L-fucose is present in.

Up to the time of writing, two distinct α-L-fucosidases of *Bacteroides thetaiotaomicron* have been biochemically characterised. These enzymes have different substrate specificities[191]. *Bt*Fuc2192 belongs to the GH29-B subfamily and specifically acts to cleave α-(1→3/4)-fucosidic linkages (EC 3.2.1.111) while *Bt*Fuc2970 belongs to the GH29-A subfamily, has relaxed substrate specificity, and is able to cleave any fucosidic linkage (EC 3.2.1.51)[191].

2.1.1 Previous structural and mechanistic studies on *Bt*Fuc2970

The crystal structure of *Bt*Fuc2970 was solved previously[184], confirming assignment of the catalytic nucleophile (Asp229) and acid/base residues (Glu288) based on sequence predictions from the *Tm*GH29 enzyme.

Being a member of the GH29 family, *Bt*Fuc2970 adopts a $(\beta\alpha)_8$ TIM barrel fold[182] (Figure 2.3). The covalent intermediate for *Bt*Fuc2970 hydrolysis was trapped using 2-deoxy-2-fluoro-β-L-fucopyranose, revealing a $^3S_1$ conformation for the catalytic intermediate[184]. This, in addition to observation of the Michaelis complex with fucose (in a $^1C_4$ conformation) gave evidence to support the $^1C_4 \leftrightarrow ^3H_4 \leftrightarrow ^3S_1$ conformational itinerary postulated for GH29 enzyme catalysis based on previous work on *Tm*GH29[182]. Quantum mechanics and molecular dynamics simulations on ring distortion in fucose further highlighted that the $^1C_4 \leftrightarrow ^3H_4 \leftrightarrow ^3S_1$ conformational trajectory is energetically viable and indeed favourable[184].

The kinetics of *Bt*Fuc2970 hydrolysis of *p*-nitrophenyl-α-L-fucopyranoside and α-L-fucopyranosyl fluoride were subsequently determined[189]. The k_{cat} value of mutant *Bt*Fuc2970 E288A was approximately 2000-fold lower than that of the native enzyme, whilst no hydrolysis could be measured for the D229A mutant, confirming these residues as the catalytic acid/base and nucleophile, respectively[189].
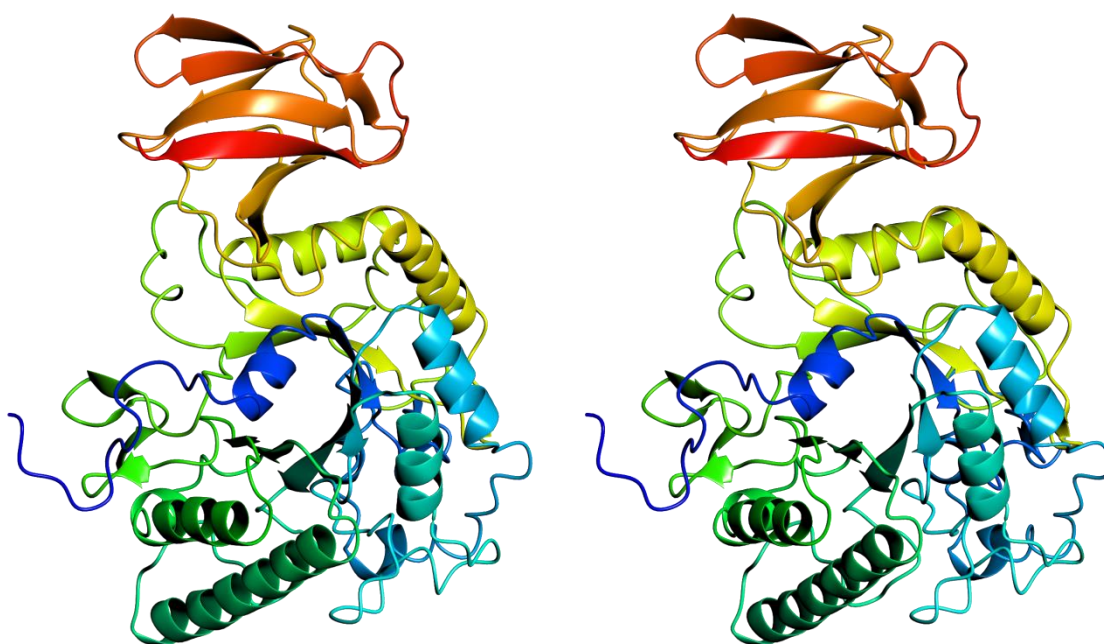
Figure 2.3. 3D fold of α-ʟ-fucosidase 2970 from *Bacteroides thetaiotaomicron* (*Bt*Fuc2970), PDB code 2WVV[184]. The main chain of a single protomer is displayed in wall-eye stereo as ribbons with colour scheme blended through the model from the blue N terminus to the red C terminus. Figure was drawn using CCP4MG[183].

The inhibition of *Bt*Fuc2970 by fuco-configured 6-membered inhibitors, displaying either a pyranose or piperidine ring system[184,200] has been probed. The 6-membered inhibitors studied displayed dissociation constants in the nM range and displayed tightest binding at pH 6.0. The pH optimum for *Bt*Fuc2970 catalysis has been determined at pH 7.4[200].

2.1.2 α-ʟ-fucosidase inhibition by 5-membered iminocyclitols

Interestingly, 5-membered iminosugars have been shown to prove potent inhibitors of a number of glycoside hydrolase families. These compounds appear to be exploited in nature, as many compounds with this functionality have been isolated as natural products from a variety of sources, including, for example, 2,5-dihydroxymethyl-3,4-dihydroxypyrrolidine from *Derris elliptica*[201], 1,4-dideoxy-1,4-imino-ᴅ-arabinitol from *Angylocalyx boutiqueanus*[202], 2-hydroxymethyl-3-hydroxypyrrolidine from *Castanospermum australe*[202] and mannostatins from *Streptoverticillium verticillus*[203]. These compounds have been shown to have

inhibitory activities towards a range of glycosidases including α-mannosidase, α-glucosidase and α-fucosidase[203-205].

Azasugars with appropriate configurations are thought to be potent inhibitors of those retaining glycoside hydrolases whose mechanism involves double-displacement and proceed through a covalent enzyme intermediate, as they mimic the positive charge build-up at the enzymatic transition state[169,206]. Further, 5-membered rings are more conformationally flexible than 6-membered rings, and the low-energy envelope conformation of 5-membered ring inhibitors mimics the high-energy boat or half-chair transition state conformations of 6-membered rings, as in both cases four adjacent atoms (the C5, O5, C1 and C2 atoms of natural substrates) have coplanarity. Hence, a great deal of research effort has been directed towards the synthesis and screening of 5-membered iminocyclitols as glycosidase inhibitors over the past few decades, the key findings of some studies pertaining to α-L-fucosidase will be summarised below.

Pyrrolidine-based α-L-fucosidase inhibitors were first assessed as early as 1993, when Wang *et al.* utilised aldolase enzymes for the chemoenzymatic synthesis of compounds **I-IV** (Table 2.1)[207]. These compounds had K_i values of 1.4 – 22 µM for bovine kidney α-L-fucosidase; it should be noted that this enzyme, which has significant sequence identity with α-L-fucosidases from *Homo sapiens* (Appendix 1), is the enzyme used for most kinetic studies against α-L-fucosidases described in the literature.

Table 2.1a. Some α-ʟ-fucosidase inhibitors described in the literature.
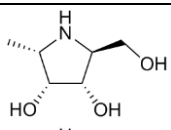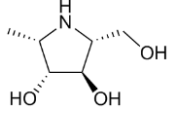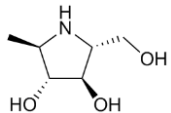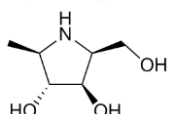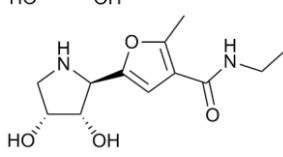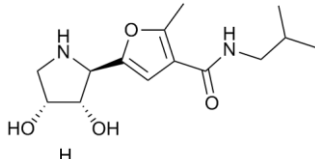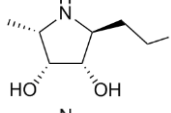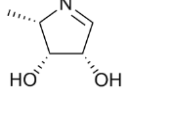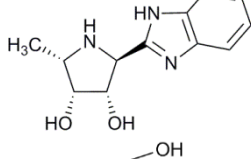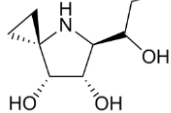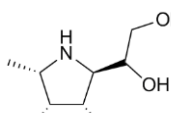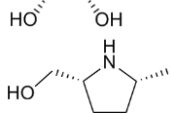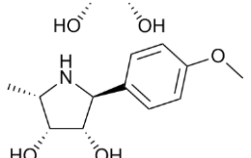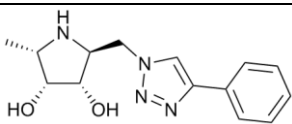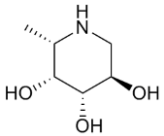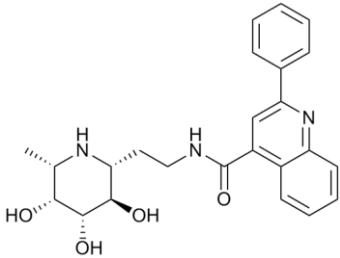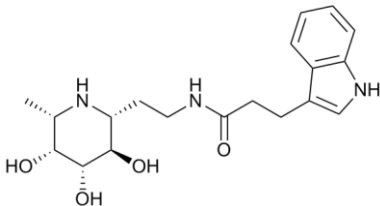
Compound number	Structural formula	K_i for bovine kidney α-ʟ-fucosidase (μM)
I		1.4
II		8
III		22
IV		4
V		3.2
VI		3.0
VII		0.008
VIII		0.010
IX		0.08
X		1.6
XI		0.2
XII		0.0049
XIII		0.010

Table 2.1b. Some α-L-fucosidase inhibitors described in the literature.

Compound number	Structural formula	K_i for bovine kidney α-L-fucosidase (µM)
XIV		0.024
XV		0.0048*
XVI		0.005
XVII		0.00105

* inhibition against bovine epididymis α-L-fucosidase

Research on the fully chemical syntheses of pyrrolidine α-L-fucosidase inhibitors started around the turn of the century. In 2001, Robina *et al.* noted that the currently available piperidine-based inhibitors of the enzyme class did not appear to be able to tolerate hydrophobic modifications which would be required for realisation of the compounds as therapeutic agents in a living system[208]. The authors proceeded to synthesise compounds **V-VI** (Table 2.1) using D-glucose as a starting material. Unfortunately, these compounds had limited utility as inhibitors, with mid-range K_i values of *ca.* 3 µM towards bovine kidney α-L-fucosidase. Further work on use of this scaffold did not improve inhibitory potency[209], and it was observed that for the realisation of more potent inhibitors such as the piperidine deoxyfuconojirimycin (DFJ)[210], an (*S*)-configured methyl group must be attached to the C4 carbon of these pyrrolidine inhibitors[209].

The chemical synthesis of compounds with an (*S*)-configured methyl group at the C4 carbon of pyrrolidine α-L-fucosidase inhibitors was first reported in 2004 by Chevrier *et al.*, who realised the synthesis of potent (K_i = 8-10 nM for bovine kidney) pyrrolidine inhibitors **VII-VIII** using D-ribose as a starting material[211](Table 2.1). A year later, Moreno-Vargas *et al.* reported the synthesis of

the benzimidazole compound **IX**, which displays 80 nM inhibition of bovine kidney α-L-fucosidase[212].

Attempts towards more potent inhibition by using transition-state mimicry were exploited in α-L-fucosidase inhibition in 2006 by Laroche *et al*. [213]. Here, the authors introduced spirocyclopropyl groups to the C4 carbon of a pyrrolidine scaffold e.g. compound **X**, Table 2.1; this functional group was postulated by the authors subsequently to flatten the heterocyclic ring causing the inhibitor to more closely resemble the half-chair conformation of the enzymatic transition state[214]. This approach was, however, a step back in terms of inhibitory potency compared to compounds **VII-IX** which were produced chronologically earlier, yielding compounds with K_i values only as high as 1.6 μM[213]. Further, the parent compound **XI** lacking a spirocyclopropyl moiety proved to be a more potent inhibitor (K_i = 0.2 μM, Table 2.1)[214].

Compounds isolated as natural products have also helped to identify and drive the synthesis of potent pyrrolidine α-L-fucosidase inhibitors. *Angylocalyx pynaertii*, for example, has been shown to produce a number of alkaloids including α-L-fucosidase inhibitor **XII**[215]. Subsequent chemoenzymatic synthesis and kinetic evaluation of this compound showed it to be a very potent inhibitor of the enzyme, with a K_i against bovine kidney α-L-fucosidase of 4.9 nM[216,217], however the compound was also shown to inhibit α- and β-galactosidases. Further, the compound 4-epi-(+)-codonopsinine **XIII** (Table 2.1), which was rationally designed based on a natural product from *Codonopsis clematidae*, has potent α-L-fucosidase inhibitory activity with K_i = 10 nM against bovine kidney α-L-fucosidase [218].

More recently, Elias-Rodruguez *et al*. used a combinatorial approach to synthesis of pyrrolidine α-L-fucosidase inhibitors with triazole-containing aglycon through use of the Cu(I) catalysed alkyne-azide cycloaddition reaction[219]. The most potent inhibitor synthesised in this way, **XIV**, has a K_i of 24 nM, showing that this technique has utility for the rapid generation of compounds containing diverse aglycon moieties.

A short mention should be given to 6-membered ring piperidine-based inhibitors of α-L-fucosidase. DFJ **XV** is a potent inhibitor of α-L-fucosidases, with a K_i of 4.8 nM for the bovine epididymis enzyme[210]. In 2003, Wu *et al*. developed a combinatorial library of GH29 α-L-fucosidase inhibitors using 6-membered ring azasugar (1R)-aminomethyl-1-deoxy-L-fuconojirimycin as a scaffold[220]. Interestingly, though the compounds thus created have a β-linkage at the anomeric carbon, the approach proved to be of great success, the most potent inhibitor **XVI** had an inhibition

potency of K_i = 5 nM. Further development of this inhibitor led to **XVII**, the most potent α-L-fucosidase inhibitor known (K_i = 1.05 nM) to the author's knowledge[221](Table 2.1). These studies highlight the importance of extended aromatic "aglycon" moieties in improving the activity of α-L-fuosidase inhibitors, which has been commented on many times throughout the literature.

Thus, while it is clear that correct chirality of the stereocentres in pyrrolidine inhibitors can help to improve inhibition potency, and that the introduction of aromatic aglycons can further improve inhibition potency, the structural features and minutiae of these interactions are still unclear. While a number of studies address the structural features of α-L-fucosidase by piperidines[184,200,221], structural information on inhibition of the enzyme by pyrrolidines is limited to docking studies on α-L-fucosidase from *Thermotoga maritima*[216].

To the author's knowledge, the first crystal structure of a 5-membered iminocyclitol complexed with a glycosidase enzyme was reported in 2005 against *Drosophila melanogaster* Golgi α-mannosidase II, whereby the substituents on a pyrrolidine inhibitor along with 5-membered rings containing other atoms were observed to have conformation dictated by the steric requirements for binding of the ring susbstituents in the enzyme rather than the energetic requirements of the 5-membered ring.[222] Subsequently, in 2007, the crystal structure of another 5-membered iminocyclitol bound to a glycoside hydrolase enzyme (GH5 endoglycoceramidase II) was reported[223]. In this structure, the iminocyclitol ring was observed to be 'flipped' relative to the orientation of 6-membered ring inhibitors, thereby allowing stabilisation of its binding through hydrogen bonding interactions (Figure 2.4). Additionally, the endocyclic nitrogen superimposed not on the endocyclic oxygen of 6-membered ring inhibitors, but on the anomeric carbon, presumably allowing its stabilisation by electrostatic interactions with the catalytic nucleophile and acid/base residues.
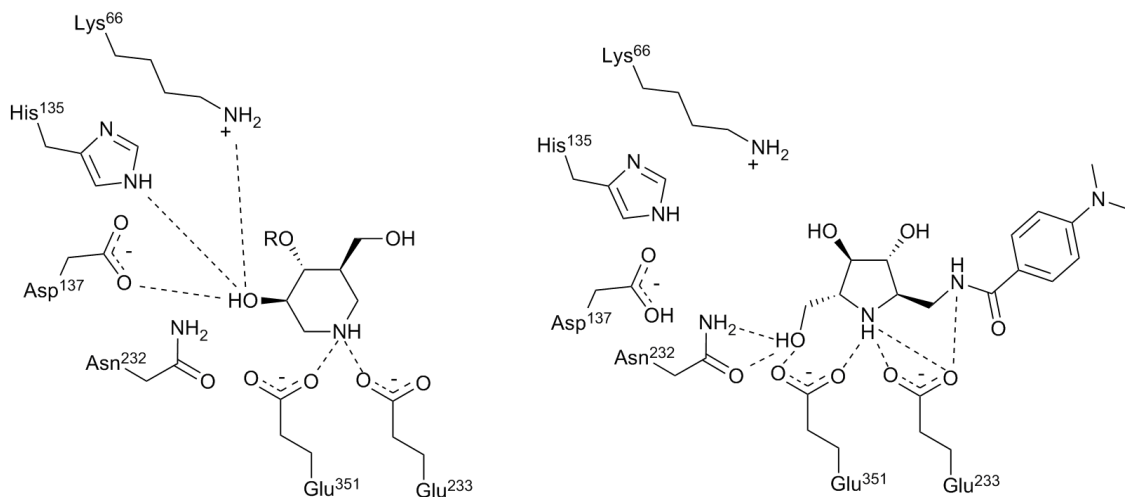
Figure 2.4. Binding mode of a 5-membered iminocyclitol to endoglycoceramidase II (PDB entry 2OYM), as contrasts with that of 6-membered ring azasugars. Adapted from Caines *et al.*[223].

Reported herein are a number of crystal structures of 5-membered iminosugars complexed with bacterial α-L-fucosidase *Bt*Fuc2970, which allow the dissection of their binding mode. Comparison with other GH29-A family enzymes enables insight into some conserved structural features in this subfamily. Further, the inhibition of *Bt*Fuc2970 by these compounds will be probed. The compounds described in this chapter were synthesised and evaluated as inhibitors of bovine kidney α-L-fucosidase by collaborators at the Universities of Reims and Seville, and their collaboration is acknowledged.

2.2 Materials and methods

2.2.1 Gene expression and protein purification

A plasmid (pET-YSBLIC3C) containing a gene encoding a truncated BtFuc2970 was available from previous work in the Davies group (Figure 2.5, the sequence of the truncated BtFuc2970 transcript is available in Appendix 2). An aliquot of this plasmid was transformed into chemically competent BL21 (DE3) cells, which were grown in LB media containing 30 mg L^{-1} kanamycin as a small scale (10 mL) culture (cells were incubated at 37 ºC overnight with aeration by shaking at 180 rpm). This culture was used to inoculate a larger (800 mL) volume of autoclaved LB media containing 30 mg L^{-1} kanamycin. Expression of the gene encoding BtFuc2970 was initiated by the addition of 1 mg L^{-1} isopropyl-β-D-1-thiogalactopyranoside (IPTG) when the culture had reached an A_{600} in the range 0.6 - 1.0. Cells were incubated at 30 ºC overnight; subsequently cells were harvested by centrifugation and resuspended in 20 mM HEPES buffer, 0.5 M NaCl, 20 mM imidazole (pH 7.4). Overexpression of BtFuc2970 was tentatively confirmed by the presence of a strong band at ca. 54 kDa molecular weight (the expected molecular weight of the BtFuc2970 gene product) on an SDS-PAGE gel.

BtFuc2970 was purified from cell lysate using a two-step procedure, comprising a Ni^{2+} affinity capture of the His_6-tagged protein and subsequent size-exclusion chromatographic separation. Cells were lysed by sonication, the resulting solution was centrifuged at 4416 RCF for 30 m and the supernatant decanted. This supernatant was applied to a 5 mL HiTrap column (GE Healthcare) preloaded with Ni^{2+} connected to an Äkta FPLC (GE Healthcare). 20 mM HEPES buffer, 0.5 M NaCl, 20 mM imidazole (pH 7.4) was passed through the column until the A_{280} response returned to the baseline. After this, an increasing concentration of imidazole was applied to the column as a gradient (to 500 mM imidazole over 50 mL). Fractions with a strong A_{280} response were run on a SDS-PAGE gel to determine the molecular weight of the protein species, and were concentrated by centrifugation at 5000 RCF using an Amicon Ultra-15 10K centrifugal filter. This protein was then applied to a Hi-Load 16/60 Superdex 200 prep grade size exclusion column (GE Healthcare) connected to an Äkta FPLC. A solution comprising 20 mM HEPES buffer, 0.1 M NaCl (pH 7.0) was run through the column and two major non-aggregated protein species were observed; both species were tentatively confirmed as being due to BtFuc2970 by SDS-PAGE. These species were attributed to be BtFuc2970 in monomeric (prevalent form) and dimeric oligomerisation states. Protein from the monomeric state was concentrated by

centrifugation at 5000 RCF using an Amicon Ultra-15 10K centrifugal filter and stored at -80ºC for subsequent experimentation.
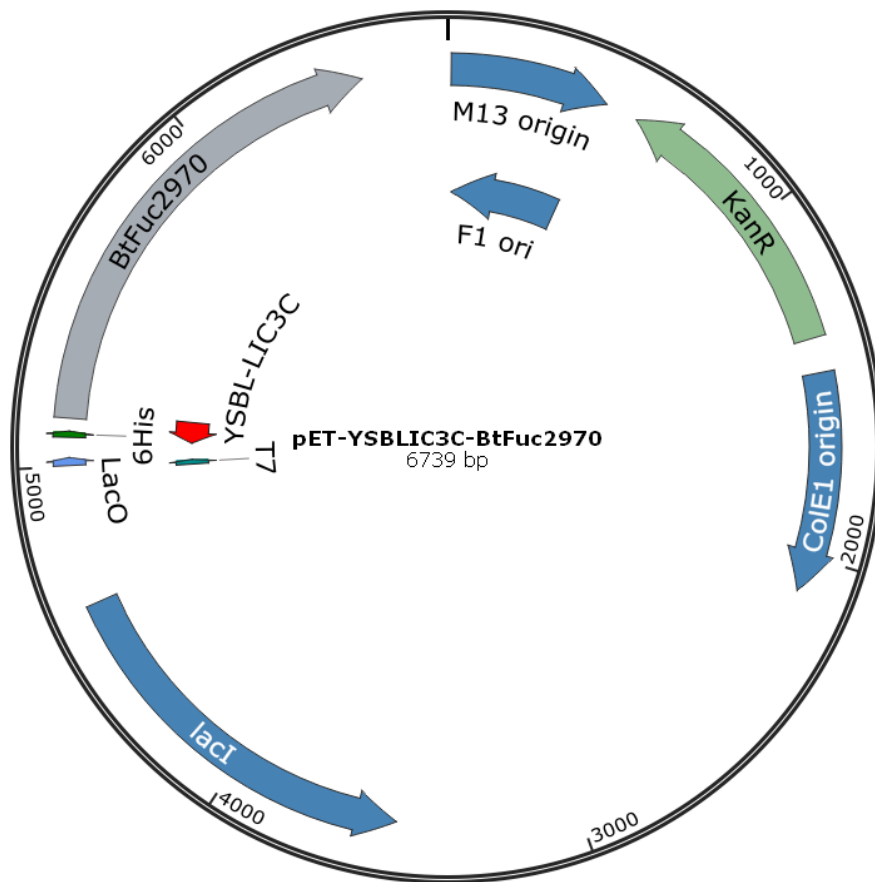
Figure 2.5. Plasmid map of pET-YSBLIC3C with *Bt*Fuc2970 cloned into the YSBL-LIC3C site. The ORF containing a gene encoding truncated *Bt*Fuc2970 is under control of the T7 promoter and begins transcription with an N-terminal 3C-cleavable His_6 affinity tag. The plasmid map was generated using SnapGene® Viewer.

2.2.2 Inhibitors

5-membered iminosugar compounds displaying inhibition of fucosidase enzymes were received from collaborators (research groups of Profs. Jean-Bernard Behr at the University of Reims and Inmaculada Robina at the University of Seville). These inhibitors display a 5-membered iminocyclitol core with stereochemical configurations matching those of fucose (excepting the lack of a C3 atom), and have 'aglycon' moieties attached to the C1 carbon (numbering scheme shown in Figure 2.6, as derived from that commonly used for pyranose sugars).
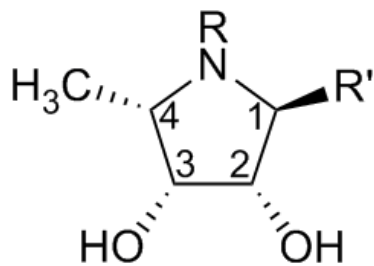
Figure 2.6. General form of 5-membered iminocyclitols described in Chapter 2.

Compounds **1-7**, which will be discussed in this chapter, along with references for their syntheses, are displayed in Table 2.2. All of the compounds display inhibitory activity towards bovine kidney fucosidase[212,218,224]. Compound **1** was tested for inhibitory activity towards other glycosidases and shown only to inhibit fucosidase[212].

Table 2.2. 5-membered iminocyclitols described in Chapter 2.

Compound number	IUPAC Name	Structural formula
1[212]	(2S,3S,4R,5S)-2-(1H-benzo[d]imidazol-2-yl)-5-methylpyrrolidine-3,4-diol	
2[218]	(2S,3S,4R,5S)-2-(4-methoxyphenyl)-1,5-dimethylpyrrolidine-3,4-diol	
3[218]	(2S,3S,4R,5S)-2-(4-methoxyphenyl)-5-methylpyrrolidine-3,4-diol	
4[218]	(2S,3R,4S,5S)-2-methyl-5-(4-methylphenyl)pyrrolidine-3,4-diol	
5[224]	(2S,3S,4R,5S)-3,4-dihydroxy-2-ethynyl-5-methylpyrrolidine	
6[224]	(2S,3S,4R,5S)-3,4-dihydroxy-2-[2'-phenyl]ethynyl-5-methylpyrrolidine	
7[224]	1-phenyl-4-[(2S,3S,4R,5S)-3,4-dihydroxy-5-methylpyrrolidin-2-yl]triazole	

2.2.3 Enzyme kinetics

The α-L-fucosidase activity of *Bt*Fuc2970 and its inhibition by compounds **1-3** and **5-7** were probed. Enzymatic assays were conducted using synthetic substrate 2-chloro,4-nitrophenyl-α-L-fucopyranoside (CNP-fucoside, CarboSynth Ltd.). The product of hydrolysis of this substrate (2-chloro,4-nitrophenol, CNP) absorbs visible radiation with a λ_{max} = 405 nm at pH above its pK_a (ca. 5.4). All experiments were carried out over a time-course of 5 m during which absorbance was measured at 405 nm using a Cintra 10 spectrophotometer (GBC Scientific Equipment Pty Ltd.). All solutions used to determine kinetic data were thermally equilibrated (37 °C).

2.2.3.1 Michaelis-Menten kinetics

The α-ʟ-fucosidase activity of purified *Bt*Fuc2970 was confirmed, and the Michaelis-Menten parameters of *Bt*Fuc2970 action upon CNP-fucoside determined. The rate of enzymatic hydrolysis was determined upon solutions of varying concentration of CNP-fucoside. Each solution (1 mL total volume) contained 50 mM HEPES buffer, 100 mM NaCl (pH 7.4) and a known concentration of substrate; 50 nM *Bt*Fuc2970 was added to this after thermal equilibration to initiate hydrolysis. The molar extinction coefficient of CNP (ε_{405} = 1.24 × 10^3 M^{-1}) was determined by measuring the absorption at 405 nm of a number of solutions of CNP-fucoside at known concentration that had been hydrolysed by excess *Bt*Fuc2970. The Michaelis-Menten parameters for CNP-fucoside hydrolysis (K_M = 0.49 ± 0.04 mM, V_{max} = 117 ± 2.6 min^{-1}, Figure 2.7) were calculated through direct fit to the Michaelis-Menten equation using the Origin graphing software (OriginLab, Northampton, MA).
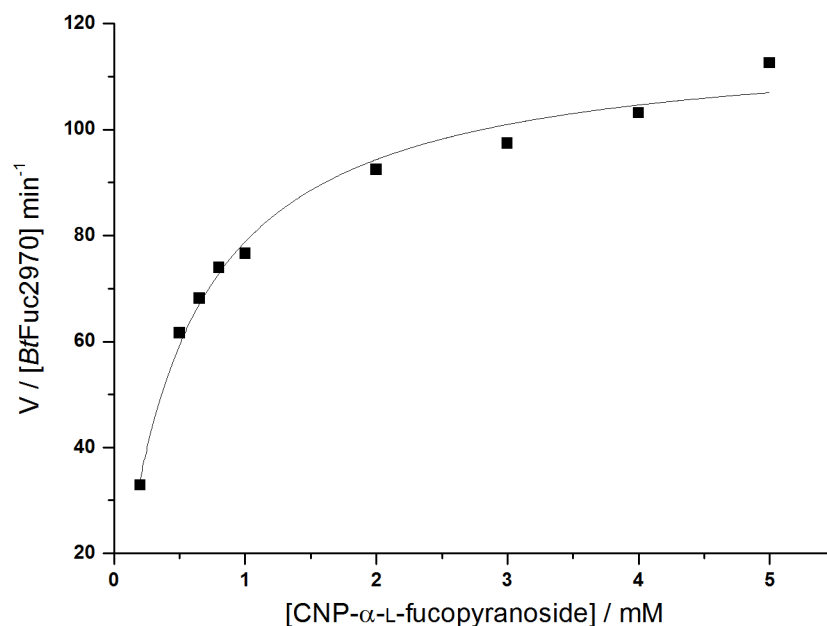
Figure 2.7. Michaelis-Menten kinetics of CNP-α-ʟ-fucopyranoside hydrolysis by *Bt*Fuc2970.

2.2.3.2 Enzyme inhibition

The inhibition of *Bt*Fuc2970 by compounds **1-3** and **5-7** was studied (Figure 2.8). Unfortunately an inadequate amount of compound **4** was available to determine its inhibition of *Bt*Fuc2970. The inhibition of *Bt*Fuc2970 by compounds **1-3** and **5-7** was studied by in each case determining non-inhibited rate and inhibited rate at various concentrations of inhibitor. Each solution contained (1 mL total volume) 50 mM HEPES buffer, 100 mM NaCl (pH 7.4), 250 nM *Bt*Fuc2970 and a known concentration of inhibitor; 50 µM CNP-fucoside was added to this to initiate hydrolysis.

Inhibition constants were calculated by determining enzymatic rates in the absence and presence of ligand and plotting their ratio against inhibitor concentration over a series of inhibitor concentrations. This plot should yield a straight line with gradient $= 1 / K_i$ and *y*-intercept of 1 according to Equation 2.1. K_is were thus determined by linear fit using the Origin (OriginLab, Northampton, MA). Inhibition plots used for the determination of K_i values are displayed in Figure 2.8.

Equation 2.1.

$$\frac{v_0}{v_i} = \frac{1}{K_i}[I] + 1$$
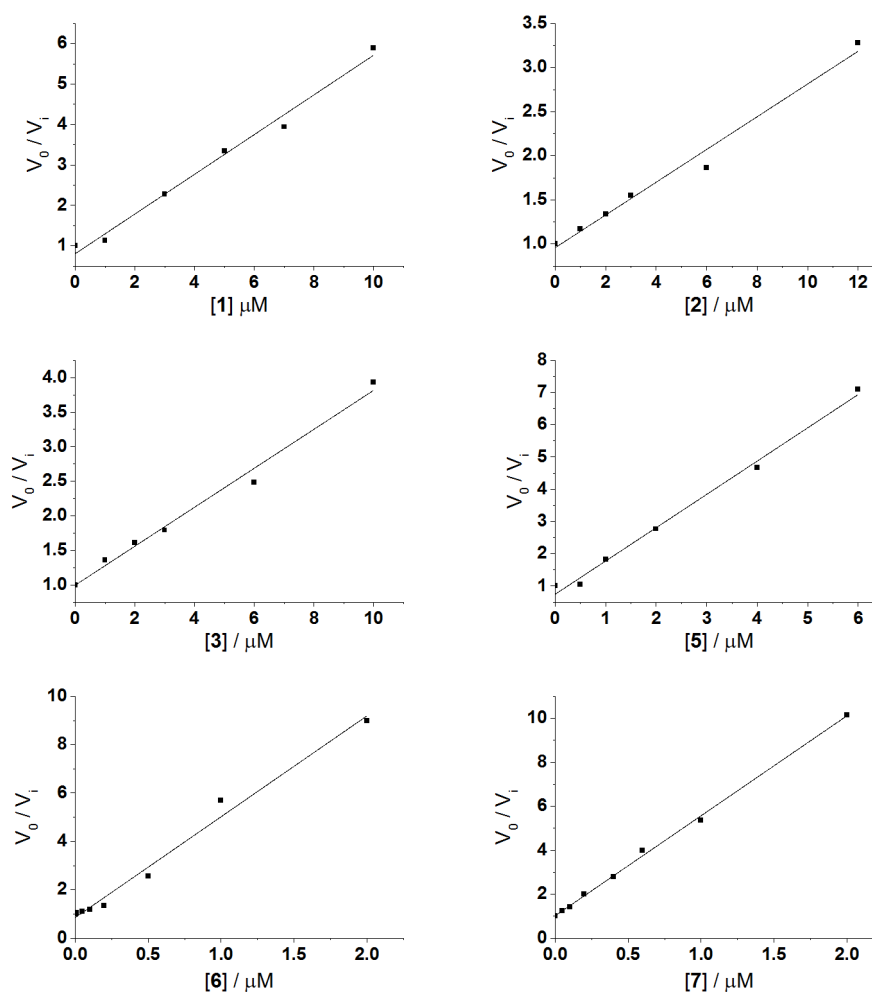
Figure 2.8. Plots showing inhibition of *Bt*Fuc2970 activity by compounds **1-3** and **5-7**.

2.2.4 Crystallisation and structure determination

2.2.4.1 Crystallisation

Crystals of *Bt*Fuc2970 were grown in either Greiner Bio One CELLSTAR 24-well culture plates (set up manually) using the hanging drop vapour diffusion method or MRC MAXI 48-well crystallisation plates (set up by use of a Mosquito liquid handling robot, TTP Labtech) using the sitting drop vapour diffusion method. Crystallisation trials were initiated based on previously successful conditions[184] and further optimised. *Bt*Fuc2970 was centrifuged for *ca*. 5 m before setting up crystallisation experiments.

For the determination of an *apo* crystal structure, a crystal of *Bt*Fuc2970 was transferred to a cryo-protectant solution containing mother liquor supplemented with 20% glycerol. Subsequently this crystal was cryo-cooled in liquid N_2.

Crystals of *Bt*Fuc2970 liganded with compounds **1-7** were obtained by either adding solid powder of each compound directly to a drop containing one or multiple crystals of *Bt*Fuc2970 using an acupuncture needle (compounds **1-4**), or by adding 1 µL compound dissolved at 4 mM in 50 mM HEPES 100 mM NaCl (pH 7.4) to the crystallisation drop and re-sealing it (compounds **5-7**). After incubation for at least *ca*. 30 m, these crystals were transferred to cryo-protectant solutions containing mother liquor supplemented with 20% glycerol and cryo-cooled using liquid N_2.

The crystal of *apo-Bt*Fuc2970 was sent to the European Synchrotron Research Facility for data collection, whilst those of *Bt*Fuc2970 complexed with each of inhibitors **1-7** were sent to Diamond Light Source.

2.2.4.2   Crystallisation conditions

Unless otherwise noted, crystallisation drops were set up in a 1:1 ratio of protein to mother liquor and *Bt*Fuc2970 was added as a 12 mg mL^{-1} solution (final concentration after dilution 6 mg mL^{-1}). All crystallisation experiments were performed at *ca*. 18 °C.

*apo-Bt*Fuc2970

Mother liquor: 17% w/v PEG 6K, 0.124 M ammonium sulfate, 0.1 M imidazole (pH 8) 2 parts protein: 3 parts mother liquor

*Bt*Fuc2970-**1**

Mother liquor: 10.45% w/v PEG 6K, 0.12 M ammonium sulfate, 0.095 M imidazole (pH 7)

*Bt*Fuc2970-**2**, *Bt*Fuc2970-**3**

Mother liquor: 12% w/v PEG 6K, 0.13 M ammonium sulfate, 0.1 M imidazole (pH 7)

*Bt*Fuc2970-**4**

Mother liquor: 20% w/vPEG 6K, 0.16 M ammonium sulfate, 0.1 M imidazole (pH 7)

*Bt*Fuc2970-**5**

Mother liquor: 16% w/v PEG 6K, 0.16 M ammonium sulfate, 0.1 M imidazole (pH 7)

*Bt*Fuc2970-**6**

Mother liquor: 14% w/v PEG 6K, 0.16 M ammonium sulfate, 0.1 M imidazole (pH 7)

*Bt*Fuc2970-**7**

Mother liquor: 20% w/v PEG 3350, 0.2 M ammonium sulfate, 0.1 M imidazole (pH 7)

2.2.4.3   Structure determination

Diffraction images for crystals of *apo-Bt*Fuc2970 and for each inhibitor complex were indexed and integrated using either XDS[225] or iMOSFLM[226]. POINTLESS[227] was used to determine the crystallographic space group in each case, and structure factor amplitudes were scaled and merged using SCALA[227] or AIMLESS[228].

The crystals of *Bt*Fuc2970 in an *apo* form and complexed with each of compounds **1**-**4** grew in a novel *P*2$_1$ space group, and these crystals were almost isomorphous to each other. The *apo*-crystal structure was solved by molecular replacement (MOLREP[229]) using a truncated version (A-chain of protein only, ligands and solvent removed) of the previously determined structure of *apo-Bt*Fuc2970 (PDB entry 2WVV[184]) as a search model. MOLREP was run using default parameters with model data to a maximum resolution of 3.0 Å. The molecular replacement solution was used to build a model of *apo-Bt*Fuc2970 using iterative cycles of model-building using COOT[230] and maximum-likelihood refinement using REFMAC5[231].

As the space groups of each inhibitor complex of *Bt*Fuc2970 liganded with compounds **1**-**4** were almost isomorphous to the *apo*-crystal, coordinates from the *apo*-crystal structure were used directly for determination of the inhibitor complex structures; R$_{free}$ sets were assigned as they had been for the *apo*-crystal structure to maintain the uniqueness and integrity of the cross-validation Free R set throughout. The crystals of *Bt*Fuc2970 complexed with each of compounds **5**-**7** grew in a space group almost isomorphous to PDB entry 2WVV[184]), and the same process as described above was used for determination of these structures, with 2WVV and its identical cross-validation set.

Coordinate sets and maximum likelihood refinement target values for each of compounds **1**-**7** were generated using CHEMDRAW3D and the PRODRG online server[232]. Inhibitor coordinates were initially added at 0.01 occupancies to prevent automatic addition of solvent into ligand density when using the COOT 'find

waters' function whilst minimising model bias. At late stages of refinement the ligand occupancy was increased to 1.0.

Final models for each of the compounds in this chapter were validated using MOLPROBITY[233] and deposited with the RCSB Protein Data Bank (www.rcsb.org[234], Table 2.3).

Table 2.3a. X-ray data collection and refinement statistics for the crystal structures discussed in Chapter 2.

	apo	*Bt*Fuc2970-**1**	*Bt*Fuc2970-**2**	*Bt*Fuc2970-**3**
Data collection				
Beamline/Date	ESRF ID23-2 21.11.2011	DLS I03 12.12.2011	DLS I04-1 8.3.2012	DLS I04-1 8.3.2012
Wavelength (Å)	0.8726	0.9763	0.9173	0.9173
Cell dimensions				
a, b, c (Å)	68.2, 96.6, 97.3	68.7, 95.6, 97.0	68.4, 95.8, 97.1	67.5, 94.6, 96.9
α, β, γ (°)	90, 91.3, 90	90, 91.2, 90	90, 90.8, 90	90, 90.9, 90
Resolution (Å)	30.7-1.59	29.0-1.73	97.1-1.58	29.3-2.10
R_{merge}	0.11(0.71)*	0.045(0.45)	0.10(0.45)	0.08(0.48)
I / σI	8.7(2.0)	12.8(1.8)	9.2(2.8)	13.3(2.6)
Completeness (%)	99.9(99.9)	98.3(94.5)	99.2(99.6)	98.9(99.7)
Redundancy	3.8	3.2	3.9	4.4
Wilson B value	15.1	26.4	20.5	32.2
Refinement				
Resolution (Å)	30.7-1.59	29.0-1.73	97.1-2.00	96.9-2.10
No. reflections	159881	128119	83899	70362
R_{work} / R_{free}	0.16/0.18	0.17/0.19	0.16/0.19	0.18/0.21
No. atoms				
Protein	7172	7083	7102	7045
Ligand/ion	69	96	100	112
Water	1059	729	852	405
B-factors (Å^2)				
Protein	14.2	33.5	24.9	39.4
Ligand/ion	37.3	52.1	40.0	63.5
Water	25.1	39.7	32.5	42.4
R.m.s. deviations				
Bond lengths (Å)	0.014	0.015	0.016	0.016
Bond angles (°)	1.5	1.5	1.6	1.6
Ramachandran Statistics (%)				
Preferred	96.0	95.6	95.8	95.0
Allowed	3.3	3.3	3.1	3.9
Outliers	0.7	1.1	1.1	1.1
PDB codes	4J27	4J28	4JFS	4JFT

*Values in parentheses are for highest-resolution shell.

Table 2.3b. X-ray data collection and refinement statistics for the crystal structures discussed in Chapter 2.

	*Bt*Fuc2970-**4**	*Bt*Fuc2970-**5**	*Bt*Fuc2970-**6**	*Bt*Fuc2970-**7**
Data collection				
Beamline/Date	DLS I03 12.12.2011	DLS I04-1 16.12.2012	DLS I04-1 16.12.2012	DLS I03 02.02.2014
Wavelength (Å)	0.9763	0.9200	0.9200	0.9763
Cell dimensions				
a, b, c (Å)	68.0,95.4,96.9	56.6,188.8,97.7	56.4,188.3,97.9	56.0,187.6,97.7
α,β,γ (°)	90, 90.8, 90	90,94.3,90	90,94.1,90	90,94.3,90
Resolution (Å)	55.4-1.66	38.5-2.10	39.3-1.77≠	55.9-1.95
R_{merge}	0.05(0.16)*	0.12(0.68)	0.11(1.72)	0.07(0.71)
I / σI	10.7(3.5)	7.2(1.8)	6.0(0.7)	8.8(1.7)
Completeness (%)	99.3(99.1)	93.5(88.8)	95.9(96.7)	98.2(97.4)
Redundancy	2.9	3.6	4.0	3.9
Wilson B value	20.5	19.6	22.4	31.9
Refinement				
Resolution (Å)	55.4-1.66	38.5-2.10	39.3-1.77	55.9-1.95
No. reflections	144771	104691	178165	135815
R_{work} / R_{free}	0.15/0.18	0.21/0.25	0.22/0.26	0.19/0.23
No. atoms				
Protein	7183	14052	14183	14304
Ligand/ion	100	90	124	136
Water	1098	579	588	892
B-factors				
Protein	24.3	32.9	35.8	38.3
Ligand/ion	41.8	55.5	50.6	55.6
Water	37.3	28.7	32.2	39.9
R.m.s. deviations				
Bond lengths(Å)	0.015	0.013	0.015	0.016
Bond angles(°)	1.5	1.4	1.5	1.5
Ramachandran Statistics (%)				
Preferred	96.0	95.6	95.6	95.7
Allowed	3.3	3.6	3.4	3.4
Outliers	0.7	0.8	1.0	0.9
PDB codes	4JFU	4PCT	4PCS	4PEE

*Values in parentheses are for highest-resolution shell.
≠Diffraction data were scaled inappropriately for the 4PCS dataset submitted to the RCSB. When scaled with a high resolution cut-off of 2 Å, statistics are as follows: R_{merge} = 0.705, I/σI = 1.6, Completeness = 96.3(87.0), Redundancy = 4.1(3.6), Wilson B value = 26.8, half-set correlation = 0.997(0.711).

2.3   Results and Discussion

2.3.1   Inhibition of *Bt*Fuc2970 by 5-membered iminocyclitols

K_is determined for inhibition of *Bt*Fuc2970 by compounds **1**-**7** are displayed in Table 2.4, along with those previously determined for inhibition of mammalian (bovine kidney) α-L-fucosidase[218,224,235] by collaborators.

Table 2.4. Inhibition of *Bt*Fuc2970 and mammalian fucosidase by compounds **1**-**7**.

Compound	K_i *vs*. *Bt*Fuc2970 / µM (values in parens: fitting error/ µM from linear fit in Origin)	K_i *vs*. bovine kidney fucosidase
1	2.0 (0.1)	0.08
2	5.4 (0.3)	0.81
3	3.5 (0.2)	0.01
4	N/D	0.01
5	1.0 (0.04)	0.07
6	0.24 (0.01)	0.014
7	0.22 (0.003)	0.005

All of the compounds studied display poorer binding affinities towards *Bt*Fuc2970 than towards mammalian enzymes. In general, the trends observed for inhibition of bovine kidney α-L-fucosidase are mirrored in inhibition of *Bt*Fuc2970, however the magnitude of inhibition is far lower; the potent, nanomolar affinities displayed by *e.g.* compound **7** against bovine kidney fucosidases are not apparent against *Bt*Fuc2970.

Kinetic determinations for inhibition of *Bt*Fuc2970 were carried out by continuous assay of hydrolysis of CNP-α-L-fucopyranoside at pH 7.4 – the previously determined pH optimum for enzymatic activity[200]; the tightest binding of 6-membered iminocyclitol inhibitors to *Bt*Fuc2970 was however previously observed to be at pH 6.0[200]. Inhibition constants against the bovine kidney α-L-fucosidase were determined by stopped assay of hydrolysis of *p*-nitrophenyl-α-L-fucopyranoside at pH 5.0 (**1**) [235] or 5.6 (**2**-**8**)[218,224]. The discrepancy observed between K_is against each enzyme may therefore either reflect electrostatic differences in the enzymes or inhibitors due to the differing experimental pH, or physical differences in the inhibitor binding sites of the two enzymes.

2.3.2 Structural analysis of the binding of 5-membered iminocyclitol inhibitors to *Bt*Fuc2970

2.3.2.1 X-ray data quality

Diffraction images for apo-*Bt*Fuc2970 were collected at 1.59 Å resolution, while those for *Bt*Fuc2970 complexed with compounds **1-7** were collected at 1.66-2.10 Å resolution (Table 2.3). X-ray data were scaled at resolution cut-offs based on criteria for R_{merge} and $I/\sigma(I)$ considerations. The dataset submitted to the RCSB for *Bt*Fuc2970-**6** (4PCS) was perhaps scaled to an inappropriate resolution, see Table 2.3. The resolution limit to which these X-ray data were useful would be better decided on the basis of half-dataset correlation co-efficients[228], however this technology was not available at the time of data scaling and merging.

Compounds **1-7** were unambiguously present in the crystal structures as evidenced by clear peaks in F_o-F_c maps after refinement of protein, solvent and other ions, but before the inclusion of compounds **1-7** in refinement (Figure 2.9).
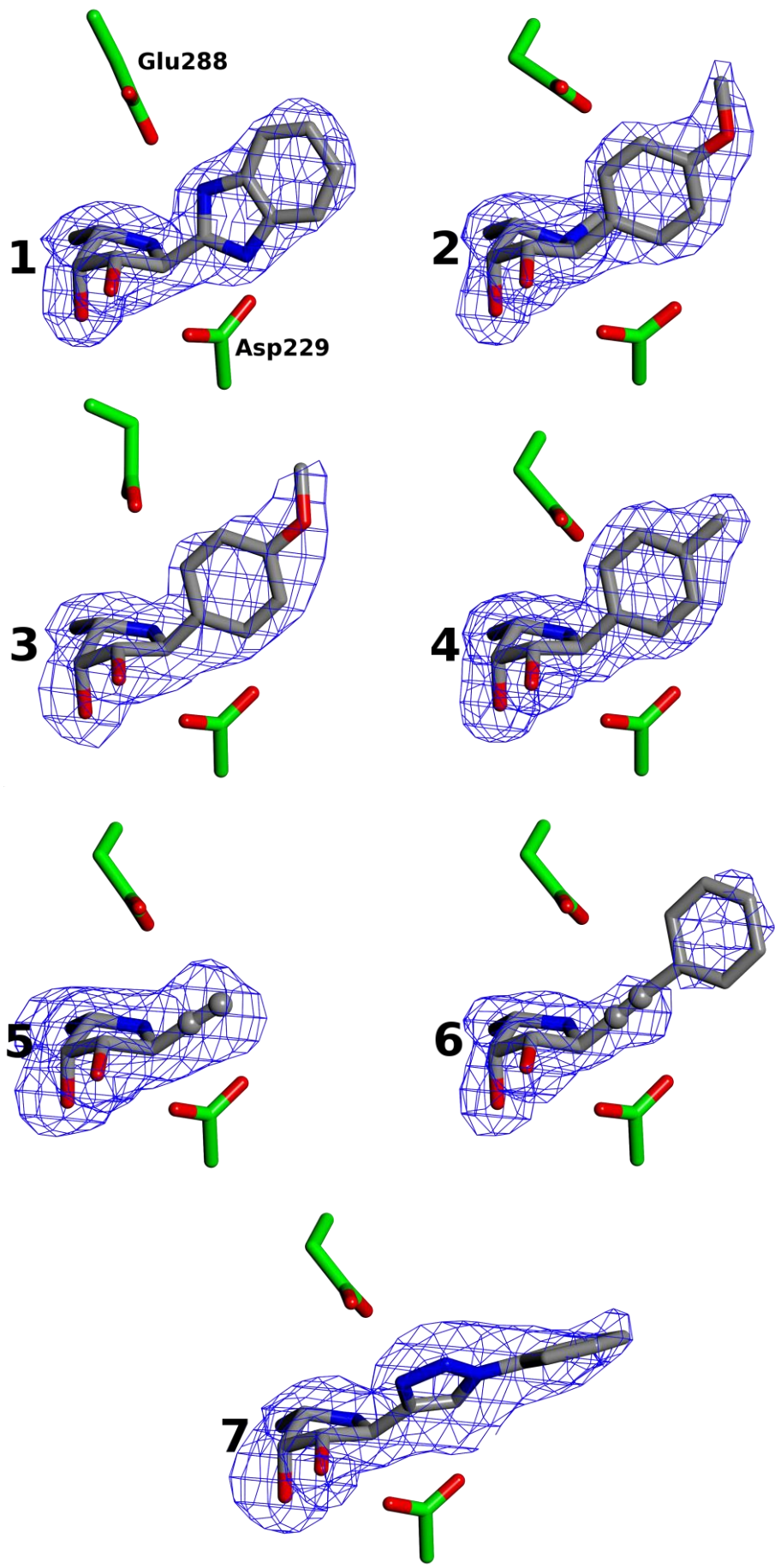
Figure 2.9. Figures of compounds **1**-**7**, respectively, lying in the active site of *Bt*Fuc2970. Atoms of compounds **1**-**7** are coloured by atom type (carbon in grey). The catalytic nucleophile Asp[229] and acid/base Glu[288] of *Bt*Fuc2970 are displayed coloured by atom type (carbon in green, residues are annotated for *Bt*Fuc2970:**1**). Carbon atoms displaying an alkyne bond in compounds **5** and **6** displayed as grey spheres. The maps displayed are F_o-F_c maps from before the incorporation of phases from compounds **1**-**7** in refinement, contoured at 5 σ (compounds **1**-**4**), 3 σ (**5**) or 2 σ (**6**, **7**). Figures were drawn using CCP4MG[183].

2.3.2.2 *Mode of binding of 5-membered iminocyclitols to BtFuc2970*

In the complex structures determined, the iminocyclitol cores of compounds **1**-**7** are observed to adopt an E_3 conformation (Figure 2.10). In this conformation, the ring nitrogen, C1, C2 and C4 atoms lie in a plane, and the C3 atom lies below this plane when viewed from the face where atom numbering increases in a clockwise fashion. This conformation mimics and gives evidence to support the postulated $^3H_4$ catalytic transition state for enzyme catalysis by GH29 fucosidases[182,184] on a geometrical basis.

The iminocyclitol cores of each of the inhibitors studied lie in the *Bt*Fuc2970 active site such that their C2 and C3 hydroxyls make hydrogen bonding interactions with active site residues (Figure 2.11). These hydrogen bonding interactions are the same as those made between the C3 and C4 hydroxyls of 6-membered inhibitors with *Bt*Fuc2970, and also with the *Tm*GH29 enzyme (Figure 2.12). 5-membered inhibitors lack the (*S*)-hydroxyl group that is attached at the C2 position of 6-membered inhibitors. This leads to loss of two hydrogen bonding interactions (between this hydroxyl and both Nε_1 of Trp[88] and Nε_2 of His[136]).

The conformation adopted by five-membered iminocyclitols in the enzymatic active site of *Bt*Fuc2970 contrasts greatly with that observed previously for this class of inhibitor against the GH5 family endoglycoceramidase[223]. Inhibition of this GH5 enzyme was shown to involve 'flipping' of the inhibitor *i.e.* a 180 ° rotation of the inhibitor ring such that the ring substituents pointing towards one direction of the active site now pointed the other direction (this can be seen graphically in Figure 2.3). The endocyclic amine of this inhibitor was observed to be positioned where the anomeric carbon of a natural substrate would be bound. Here, binding of the iminocyclitol inhibitor in a flipped conformation appears to be favoured due to electrostatic interactions between the endocyclic amine of the inhibitor and the

catalytic nucleophile and acid/base residues; further, a suitable hydrogen bonding network can still be formed even after 'flipping' of the inhibitor (Figure 2.3).
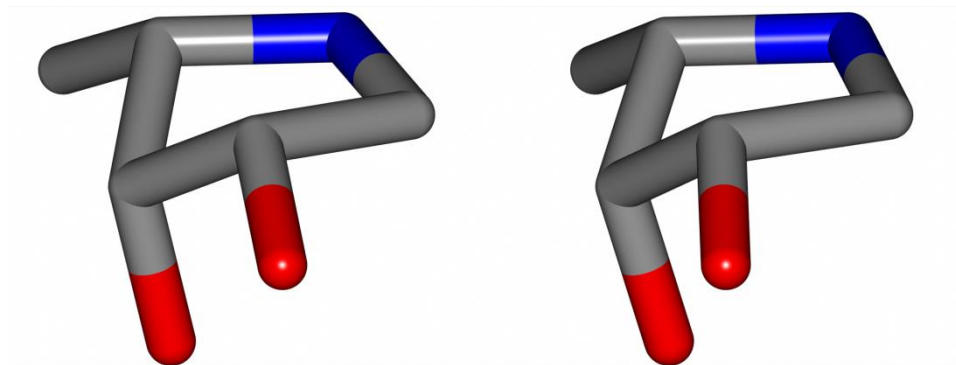
Figure 2.10. Conformation of the iminocyclitol core of five-membered iminocyclitol inhibitor **1** as observed in crystal structure bound to *Bt*Fuc2970. The figure was drawn in wall-eye stereo with aglycon atoms not displayed, using CCP4MG[183].
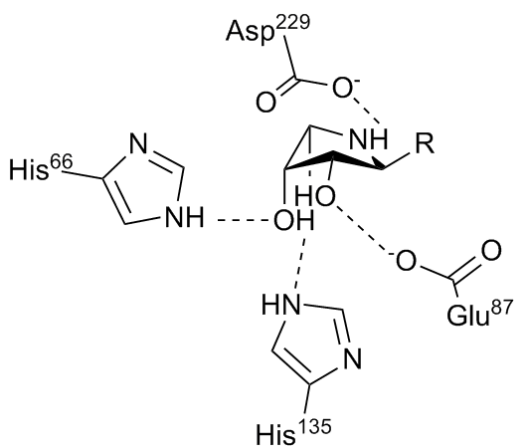
Figure 2.11. Schematic of the interactions between the iminocyclitol core of compounds **1**-**7** and *Bt*Fuc2970. Protonation states displayed are arbitrary. Hydrogen bonds < 3 Å shown as dashed lines.
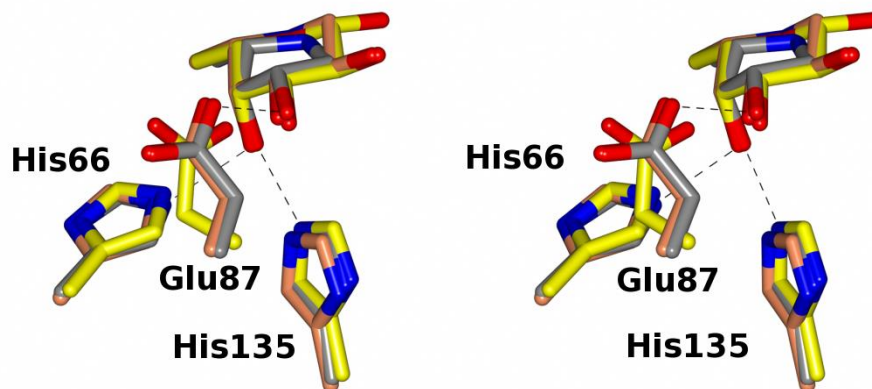
Figure 2.12. Superposition of the coordinates of 4J28 (*Bt*Fuc2970-**1**), 2XIB[200] and 1ODU[182] displayed in wall-eye stereo. The iminocyclitol or fucopyranose cores of inhibitors and residues that hydrogen bond to **1** are displayed in each case. Atoms are coloured by type; carbon atoms from 4J28 in grey, from 2XIB in coral and from 1ODU in yellow. Hydrogen bonds (2.6 – 2.8 Å) from **1** displayed as dashed bonds. Residues with numbering according to *Bt*Fuc2970 are annotated. The figure was drawn using CCP4MG[183].

For the case of fucose-mimetics such as **1-7** however, binding of the inhibitor in a flipped conformation would lead to loss of a number of hydrogen bonding interactions. The importance of the hydrogen bonds made by the hydroxyls of the C2, C3 and C4 carbon atoms of fucose-mimetics has previously been observed by Winchester *et al.* after analysis of structure activity relationships of α-ʟ-fucosidase inhibitors[107]. **1-7** have already lost the hydrogen bonding interactions made by their missing hydroxyl unit, and ring flipping would lead to loss of the other interactions thus leading to a significant energy barrier to inhibitor binding in this conformation.

The crystal structure of compound **3** has been determined[218]. The crystal structures of compound **3** in its enzyme free and enzyme bound forms were superposed (Figure 2.13). This superposition gives an RMSD of 0.18Å and shows that the conformation adopted by the compound as a ligand of *Bt*Fuc2970 is almost identical to the one it displays in solution. As it is expected that the conformation seen in small molecule crystal structure is the ground state conformation, it can be concluded that 5-membered iminocyclitols require negligible conformational change in binding to α-ʟ-fucosidases. This is likely reflected to some degree in the favourable binding of 5-membered iminocyclitols to *Bt*Fuc2970 despite the reduced number of hydrogen bonding interactions they make to the enzyme, compared to 6-membered inhibitors.
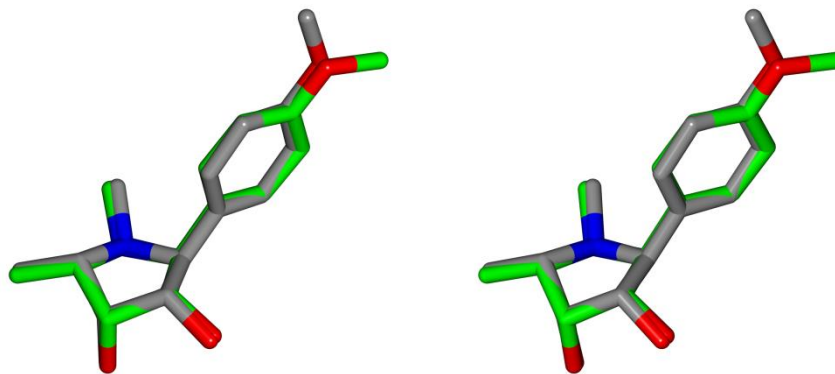
Figure 2.13. Superposition of the enzyme free and enzyme bound crystal structures of compound **3** displayed in wall-eye stereo. Atoms are coloured by type; carbon atoms from *Bt*Fuc2970**-3** in grey, from the small molecule crystal structure[218] in green.

2.3.2.3 *Positioning of the aglycon moieties of iminocyclitol inhibitors*

5-membered iminocyclitols lie in the *Bt*Fuc2970 active site in a way that satisfies the maximum number of hydrogen bonding interactions possible (Figure 2.12). The aglycon moieties of 5-membered inhibitors bound to *Bt*Fuc2970 are observed to occupy a different section of the hydrophobic landscape when compared to that occupied by the aglycons of 6-membered inhibitors bound to both *Bt*Fuc2970 and *Tm*GH29 (Figure 2.14). In this binding mode, the aglycon moieties of inhibitors **1**-**7** lay atop a hydrophobic ridge formed by residues Trp[88] and Trp[232]. Upon superposition of the available crystal structures of GH29-A family enzymes, this hydrophobic ridge appears to be conserved through 3D fold (Figure 2.15). Residue Trp[88] in *Bt*Fuc2970 is structurally conserved throughout GH29-A family enzymes whilst the residue corresponding to Trp[232] of *Bt*Fuc2970 invariantly presents a similarly sized van der Waal's radius (Figure 2.15). A hydrophobic region of GH29 family enzymes has previously been alluded to based on structure activity relationships *e.g.*[220], but has not to the author's knowledge been described structurally.

Initially, it was thought that ring contraction of 5-membered ring inhibitors compared to 6-membered ring inhibitors caused ring rotation to maintain the hydrogen bonding network of these compounds and thus oriented the aglycon moiety in a non-physiological direction *e.g.* [236]. However, after close inspection of the available crystal structures of GH29 enzymes it appears unlikely that this is the case.
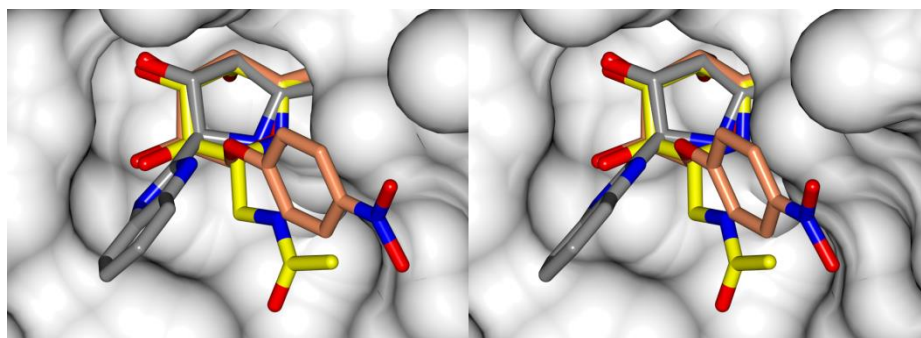
Figure 2.14. Superposition of the coordinates of 4J28 (*Bt*Fuc2970-**1**), 2WVU[184] and 2ZXA[221], shown in wall-eye stereo. Inhibitors are coloured by atom type; carbon atoms from **1** in grey, from *para*-nitrophenyl-α-ʟ-fucopyranoside (2WVU) in coral and from acetyl-fuconojirimycin (2ZXA) in yellow. Protein atoms of *Bt*Fuc2970 (4J28) displayed as a white surface. The figure was drawn using CCP4MG[183].
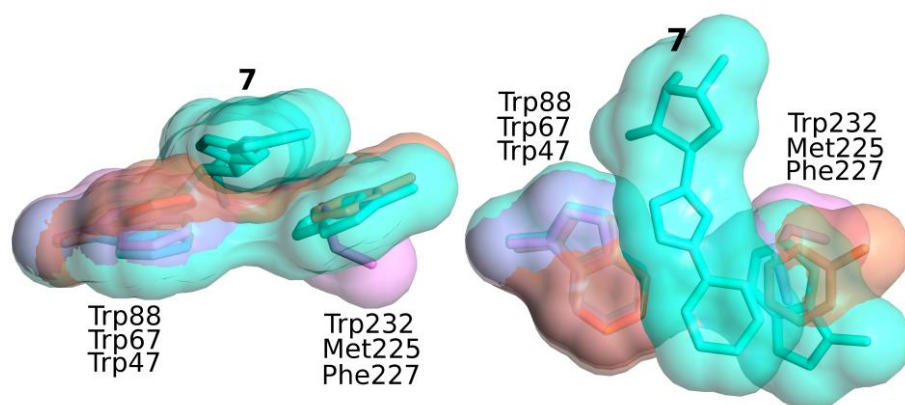
Figure 2.15. The hydrophobic ridge formed in GH29-A family fucosidase crystal structures. GH29-A fucosidase crystal structures from *Bt*Fuc2970 (liganded with **7**, side-chains of residues Trp[88] and Trp[232], and **7** displayed in cyan with cyan surfaces), *Thermotoga maritima* (PDB code 1HL8[182], side-chains of Trp[67] and Met[225] in pink with pink surfaces) and *Fusarium graminearum* (PDB code 4PSR[188], side-chains of Trp[47] and Phe[227] in coral with coral surfaces) were superposed. Orthogonal figures of the hydrophobic ridge were prepared using CCP4MG[183].

In the available crystal structures of α-ʟ-fucosidases (*Bt*Fuc2970 and *Tm*GH29) complexed with 6-membered ring compounds, the ligands display 'non-natural' features. Inhibitors complexed with *Tm*GH29 all display a β-linkage at C1, while in the crystal structure of a D229N mutant of *Bt*Fuc2970 complexed with 2,4-dinitrophenylfucopyranoside the catalytic acid/base E288 is flipped out of an active conformation. Comparison of the aglycon trace between 5-membered inhibitors of

*Bt*Fuc2970 and lacto-N-fucopentaose II complexed with *Bifidobacterium longuum* GH29-B fucosidase (PDB entry 3UET)[187], however, shows that the aglycons of compounds **1**-**7** lay in a similar position to that of lacto-N-fucopentaose II (Figure 2.16). As such, it is likely that the orientation observed for the aglycons of compounds **1**-**7** liganded with *Bt*Fuc2970 is physiologically relevant.
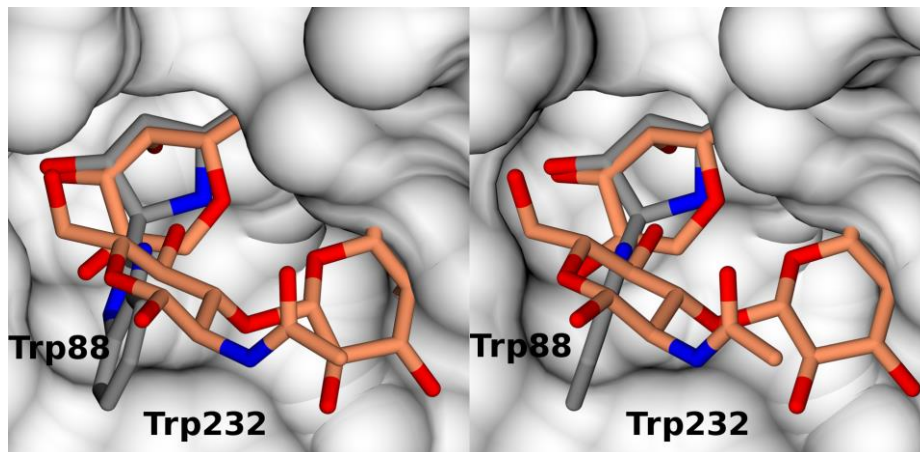
Figure 2.16. Superposition of the coordinates of 4J28 (*Bt*Fuc2970-**1**) and 3UET[187]. Inhibitors are coloured by atom type; carbon atoms from **1** in grey, from lacto-N-fucopentaose II in coral. Positions of residues Trp88 and Trp232 are indicated in the figure. Protein atoms of *Bt*Fuc2970 (4J28) displayed as a white surface. The figure was drawn using CCP4MG[183].

In all of the structures obtained throughout this chapter, sulfate ions were observed proximal to the enzymatic active site, in what appears to be an electrostatically positive binding site, binding of the sulfate is stabilised by a number of hydrogen bonds from residues Arg[262], Arg[271] and His[272] (Figure 2.17). Whilst sulfate ions were always present in crystallisation solutions, this is due to the preferential crystallisation of this enzyme in conditions containing sulfate.
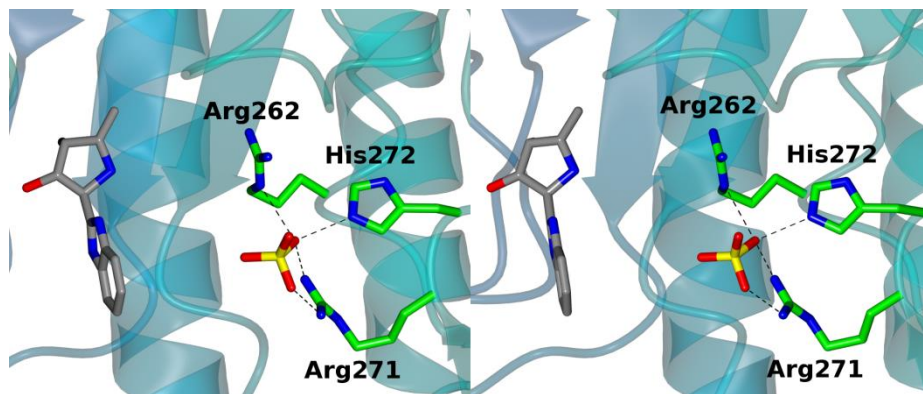
Figure 2.17. Position of sulfate ions proximal to the active site of *Bt*Fuc2970. The protein main chain of *Bt*Fuc2970-**1** (PDB entry 4J28) displayed as ribbons coloured in cyan. Atoms of **1**, the sidechains of residues Arg^{262}, Arg^{271} and His^{272} and sulfate moiety are displayed coloured by atom type (carbon atoms of **1** displayed in grey). Hydrogen bonds made between the sulfate moiety and the protein side-chains of *Bt*Fuc2970 are displayed as dashed lines. The figure was drawn using CCP4MG[183].

*Bt*Fuc2970 belongs to polysaccharide utilisation locus 44 of the genome of *Bacteroides thetaiotaomicron* VPI-5482; this locus also contains genes encoding a chondroitin AC lyase and a β-galactosidase. It is thus tempting to speculate that the substrate for this PUL is some form of fucosylated chondroitin sulfate. The glycosaminoglycan with this name, however, comprises main chain β1→4-linked glucuronic acid and β1→3-linked N-acetyl galactosamine, with sulfated Fucα1→3 branches appended to the main chain glucuronic acid[22,25]. The other currently annotated transcripts of this locus do not appear to have the necessary functionality to degrade this substrate, however they have not been biochemically characterised.

2.4 Conclusions

Kinetic and X-ray crystallographic studies were used to probe the interactions of a series of fuco-configured 5-membered iminocyclitols with GH29 α-L-fucosidase *Bt*Fuc2970 from *Bacteroides thetaiotaomicron*. X-ray crystal structures of these compounds complexed with *Bt*Fuc2970 elucidated their mode of binding; through an E_3 conformation that mimics the postulated $^3H_4$ transition state for catalysis[184]. This binding conformation was also observed to be the conformational ground state of these inhibitors based on analysis with the previously published small molecule crystal structure of a representative inhibitor. The loss of a single hydroxyl compared to 6-membered ring inhibitors was shown to cause disruption of the hydrogen bonding network which has previously been shown to be important for GH29 inhibition[220]. The crystal structures further served to highlight the likely orientation adopted by the aglycons of natural substrates in the GH29-A family and allowed the observation of a conserved hydrophobic ridge that is present in all GH29-A family members for which crystal structures are known. Kinetic studies of the inhibitory activity of these inhibitors on *Bt*Fuc2970, when compared with previous values against bovine kidney α-L-fucosidase, showed a discrepancy whereby inhibition was more potent against the mammalian enzyme than *Bt*Fuc2970; this may be due to differences in either the pH that inhibition experiments were carried out or the inhibitor binding sites in each enzyme.

*Bt*Fuc2970 belongs to polysaccharide utilisation locus 44 of the genome of *Bacteroides thetaiotaomicron* VPI-5482, which also contains a gene encoding a chondroitin AC lyase. It was observed that crystals of this enzyme preferentially form in sulfate-containing crystallisation solutions, and that sulfate is invariably located proximal to the enzymatic active site, in an area of the enzyme with a positive electrostatic potential. It is thus tempting to speculate that the natural substrate for this locus may be a form of fucosylated chondroitin sulfate. More detailed biochemical characterisation of the remaining gene products on this locus could prove helpful for determination of its biological substrate and function.

The results presented herein may aid in the design of more potent inhibitors to target GH29 α-L-fucosidases, which may have therapeutic relevance due to the myriad biological functions the enzyme is involved in. As no X-ray crystal structure is yet available for a human GH29 enzyme, it is uncertain to what extent structural insight made on this surrogate system is applicable to the aforementioned, however.

Chapter 3: Multivalency in fucosidase inhibition and targeting of neoplastic cells by ferrocenyl-iminocylitols*

Abstract

As GH29 α-ʟ-fucosidase is implicated, or differentially regulated in a number of disease phenotypes, there is a need for the generation of more potent inhibitors of this enzyme class which may be able to probe the biology of these diseases or treat them. In nature, multivalency effects often dramatically increase the affinity of interactions between carbohydrates and lectins, and it may be possible to harness this multivalency effect in the inhibition of carbohydrate processing enzymes. Since upregulation of human GH29 α-ʟ-fucosidase is a known determinant of oral, breast and liver neoplasias, the use of pharmacophore-containing α-ʟ-fucosidase inhibitors may provide a route for chemotherapy of these cancers. We report herein crystallographic structure determination of complexes of GH29 enzyme *Bt*Fuc2970 liganded with a monovalent α-ʟ-fucosidase inhibitor and its trivalent counterpart at 1.7 and 1.68 Å respectively, coupled with kinetic data for their inhibition of *Bt*Fuc2970. These data are used to conclude that multivalency affinity enhancements by chelation are geometrically impossible and those by stastistical rebinding are unlikely to be seen in this case due to the low valency of the compounds. Further, the 1.88 - 2.30 Å crystal structures of α-ʟ-fucosidase configured iminosugars coupled to ferrocene pharmacophores are reported. While these compounds are shown to bind in the GH29 active site and others have shown they have anti-proliferative action on the MDA-MB-231 breast cancer cell line, further experiments would be needed to assess their clinical usefulness.

3.1   Introduction

As GH29 α-ʟ-fucosidase is implicated or differentially regulated in a number of disease phenotypes, there is a need for the generation of more potent inhibitors of this enzyme class. In nature, multivalency effects are often found in the interactions of carbohydrates with their lectin receptors. These effects can dramatically (*ca.* 6 orders of magnitude) increase the affinity of a carbohydrate-lectin interaction,[237] and it may be possible to harness this multivalency in the inhibition of carbohydrate processing enzymes.[238]

3.1.1   Multivalency effects in carbohydrate recognition

Carbohydrates typically bind only weakly with their receptors, rarely with dissociation constants lower than 10^{-6} M[239]; this property can be of biological importance *e.g.* in the binding of selectins to their ligands in the inflammatory cascade[45]. Nevertheless, carbohydrate-receptor interactions are widespread and critical throughout biology, and some of the roles of carbohydrate interactions require tight binding between carbohydrates and their receptors.

It has been known for many years that mutlivalency or avidity effects are often exploited by nature in the interaction of carbohydrates with their receptors, ranging from species with low valency values to the high-order multivalency achieved by species such as viruses[239]. The availability of multivalent binding mechanisms for carbohydrates may lead to enhancement of the affinity of a particular interaction over what would be expected were the affinity to be corrected on a per subunit basis, hence providing access to the large affinities required for many biological processes. This effect is termed the 'cluster glycoside effect'.[240] Enhancements in binding affinity gained through this mechanism require a receptor lectin with multiple carbohydrate binding sites, and ligands containing multiple carbohydrates, such as the glycans expressed in the glycocalyx; expansion of oligosaccharides resembling N-acetyllactosamine glycans from monoantennary to triantennary has been shown to increase their affinity for mammalian hepatic lectins by 6 orders of magnitude[241].

During multivalent binding, rotational and translational entropic costs only need to be paid during the first binding event, thereby enhancing the affinity of the second sub-ligand binding event[239,242]; multivalency affinity enhancements achieved in this way can be as high as 10^3-10^6 fold[237]. A number of mechanisms exist whereby multivalency can effect the tighter overall binding of a ligand to its receptor

(Figure 3.1).[237] Chelation effects may occur when a multivalent ligand is able to bind simultaneously to multiple receptor sites on a single receptor, in this case a long enough linker region is required to allow individual subunits to span the distance between adjacent receptor binding sites. When linkers separating ligand subunits are too short for this, affinity enhancements may occur due to "statistical rebinding" or "proximity effects". These are caused by the slower dissociation of a ligand when multiple (sub)ligands are available for interaction with a receptor due to the higher local concentration of sub-ligands.[237] Clustering of receptors on cell surfaces is also common in nature, providing a high concentration of receptor moieties in a limited volume, to which multivalent inhibitors may bind. This clustering is a particularly prominent feature of lectins involved in cellular adhesion and signal transduction roles.
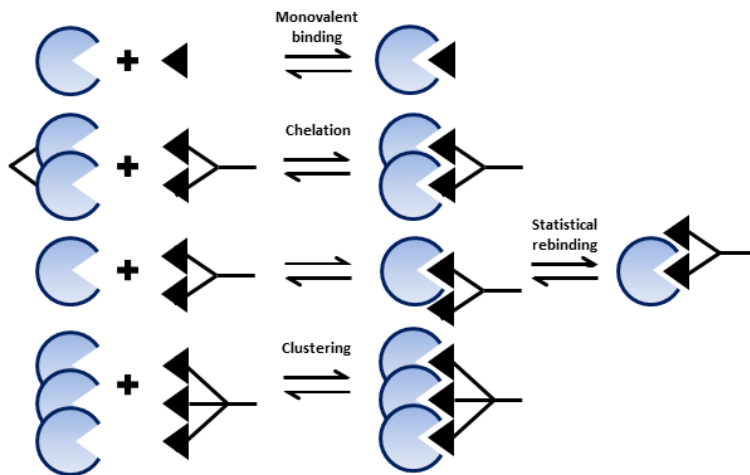
Figure 3.1. Mechanisms by which a ligand may interact through multivalency.

While multivalent ligands capable of chelating lead to very high multivalency affinity enhancements, which are optimal with relatively low valencies of under *ca.* 10, this is not the case for those ligands that may only garner "statistical rebinding" affinity enhancements; these ligands are able to achieve only weaker affinity enhancements of up to 2000-fold, and achieve these only with very high valency of the carbohydrate moiety.[237]

As mentioned previously, the adhesion and colonisation of a number of bacteria to their host is mediated by interactions between carbohydrates and their receptors. In some cases, multivalency effects are taken advantage of during the colonisation or pathogenicity of a prospective resident of the host. For example, the pathogenic

bacterium *Pseudomonas aueruginosa*, which was discussed briefly in Chapter 1, harbours tetrameric lectins, LecA and LecB, which bind to galactose and fucose, respectively[243,244]. The synthesis and evaluation of multivalent ligands for both of these lectins has been reported in the literature. For multivalent inhibition of LecB, Johansson *et al* produced two 15,625 member combinatorial libraries of amino-acid based dendrimer scaffolds appended with four fucose units by solid-phase peptide synthesis.[245] The compounds thus generated displayed a range of affinities towards LecB, with the best compounds showing a 100-fold affinity enhancement over fucose alone. Further, these compounds were able to inhibit *P. aeuruginosa* biofilm formation, with the most potent inhibitor showing complete inhibition at 50 μM, and disperse existing colonies of *P. aeuruginosa*.[245]

Interestingly, Brissonnet *et al*. recently reported the synthesis of iminosugar coated dextran polymers and their evaluation as multivalent ligands of carbohydrate processing enzymes, rather than lectins.[238] The authors generated dextran polymers of varying size functionalised with the iminosugar glycosidase hydrolase inhibitors deoxynojirimycin (DNJ) and deoxymannonojirimycin (DMJ). These compounds were then tested as inhibitors of a panel of glycoside hydrolases including the α-L-fucosidase from *Thermotoga maritima* and a mannoside phosphorylase from an uncultured *Bacteroides* strain. Surprisingly, while multivalency effects of up to 70-fold were reported, some of the enzymes tested even had their catalytic activities increased in the presence of DNJ or DMJ-coated dextrans, where it would be expected that the iminosugar inhibitors DNJ and DMJ bind in the active site and inhibit the enzymes. It is unclear what the basis is for the activation of these enzymes, but it may be due to non-specific allosteric factors or the compounds promoting alternative reaction pathways such as transglycosylation which compete for availability of active substrate.

3.1.2   Searching for determinants of neoplasia

Cancer remains one of the leading causes of death throughout the world,[72] and thus it is not at all surprising that the research field is both voluminous and rapidly growing (a search of PubMed for articles published in 2014 with title including 'cancer' had more than 55,000 hits, while the same search of previous years have successively fewer hits in a linear trend).

One of the most difficult hurdles to overcome in treatment of cancer is overcoming the remarkable phenotypic and functional heterogeneity possessed by neoplastic cells. It is believed that this heterogeneity is due to the differentiation of 'cancer stem cells' into mature cells with diverse phenotypes.[246,247] In the 'cancer stem cell' model, not all tumour cells are thought to be equal, with tumour growth being driven by relatively few tumourigenic stem cells that are capable of self-renewal and are incredibly resistant to stress, *i.e.* have features typical of normal stem cells. This model is popular as it helps to explain a number of the features associated with cancer in the clinical setting, *e.g.* that while therapy may be effective in the short-term, rarely does recurrence of the tumour not occur in the long term.

Due to the phenotypic diversity exhibited by neoplastic cells, common determinants (biomarkers) shared by many tumour cells from cancers of the same organ, or even multiple organs, are pivotal to the diagnosis, prognosis, and the decision of an appropriate therapeutic course for cancer.[248] One effective methodology for cancer chemotherapy is to find means to target cell disruption preferentially to neoplastic cells over healthy cells by utilising differences that exist between these two cell types.[249]

GH29 α-ʟ-fucosidase is known to be systematically upregulated in some cancers, including oral[197], breast[84] and liver[83] neoplasias. It may thus be envisaged that the use of compounds containing selectivity elements towards α-ʟ-fucosidase along with a pharmacophore warhead may allow the targeted disruption of tumour cells in patients with these cancers (Figure 3.2). This hypothesis does, however, hinge on the existence of a partitioning effect whereby the concentration of compounds of general form displayed in Figure 3.2 is higher around cells expressing higher levels of α-ʟ-fucosidase.

Ferrocenium salts are potent cytotoxic agents that are thought to act as such by their generation of reactive oxygen species (ROS) in cells[250]. Administration of ferrocenyl compounds into the body may lead to deleterious effects as ROS would have cytotoxic activities on all cells in a dose-dependent fashion, however neoplastic cells have an increased basal level of ROS and are particularly vulnerable to ROS-induced DNA damage mechanisms[251,252], making compounds containing this functionality particularly good pharmacophores for cancer cell disruption. Further, if directed by means of a selectivity moiety towards the biomarker α-ʟ-fucosidase (Figure 3.2), these species may provide a particularly efficient means for the disruption of oral, liver and breast cancers.

Figure 3.2. Schema of a targeted chemotherapeutic. The selectivity element allows targeting towards a particular neoplasia determinant while the pharmacophore effects cell disruption and the linker element joins these two functional elements together.

Herein, we aim to address the potential for a multivalent ligand configured for α-ʟ-fucosidase inhibitory activity to provide an affinity enhancement over its monovalent counterpart, by structural and kinetic means. Further, the crystal structures of three compounds designed as potential chemotherapeutics comprising an α-ʟ-fucosidase inhibitory scaffold and ferrocenyl aglycons are reported and the potential for the use of these compounds in chemotherapy is discussed. The syntheses of compounds described in this chapter, and their K_i values against bovine kidney α-ʟ-fucosidase were realised by collaborators at the Universities of Reims and Seville, and their collaboration is acknowledged.

3.2 Materials and methods
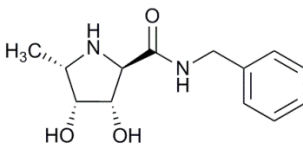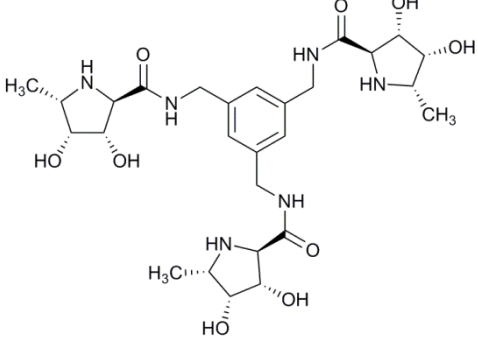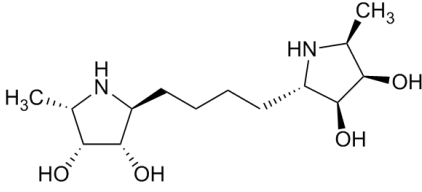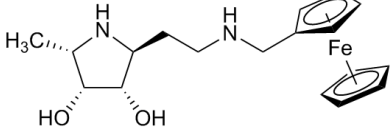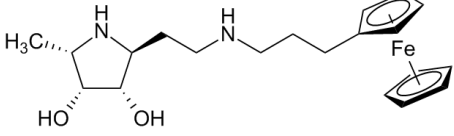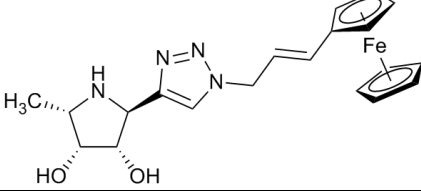
3.2.1 Gene expression and protein purification

Expression and purification of *Bt*Fuc2970 was effected as described previously in 2.2.1.

3.2.2 Inhibitors

Further 5-membered iminosugar compounds configured as inhibitors of α-L-fucosidase enzymes (see Chapter 2) were received from collaborators (research groups of Profs. Jean-Bernard Behr at the University of Reims and Inmaculada Robina at the University of Seville). The compounds discussed in this chapter are displayed in Table 3.1. The syntheses of compounds **8-9, 11-12**, along with their inhibition data towards bovine kidney α-L-fucosidase are documented (publications given in Table 3.1). Compounds **10** and **13** currently lie outside the scope of the scientific literature and thus do not have published syntheses. Compounds **8-9, 11-12** have been reported to display inhibition of bovine kidney fucosidases. Compounds **8** and **9** were further tested for inhibitory activity towards other glycosidases and shown only to inhibit α-L-fucosidase[212].

Table 3.1. 5-membered iminocyclitols described in chapter 3.

Compound number	IUPAC name	Structural formula
8[253]	(2*S*,3*S*,4*R*,5*S*)-2-Benzylcarbamoyl-5-methylpyrrolidine-3,4-diol	
9[253]	*N*,*N'*,*N''*-(1,3,5-phenylenetris(methylene))tris-[(2*S*,3*S*,4*R*,5*S*)-3,4-dihydroxy-5-methylpyrrolidine-2-carboxamide]	
10	1,4-bis[(2S,2'S,3S,3'S,4R,4'R,5S,5'S)-5-methyl-3,4-dihydroxypyrrolidin-2,2'-yl]-butane	
11[254]	(2*S*,3*S*,4*R*,5*S*)-2-[*N*-(methylferrocene)]aminoethyl-5-methylpyrrolidine-3,4-diol	
12[254]	(2*S*,3*S*,4*R*,5*S*)-2-[*N*-(propylferrocene)]aminoethyl-5-methylpyrrolidine-3,4-diol	
13	1-(3-ferrocenylprop-2-enyl)-4-[(2S,3S,4R,5S)-5-methyl-3,4-dihydroxypyrrolidin-2-yl]-triazole	

3.2.3 Enzyme kinetics

The α-ʟ-fucosidase activity of *Bt*Fuc2970 and its inhibition by compounds **8-13** were probed; enzymatic assays were conducted as described in **2.2.3**.

3.2.3.1   *Enzyme inhibition*

The inhibition of *Bt*Fuc2970 by compounds **8-13** was studied by in each case determining non-inhibited rate and inhibited rate at various concentrations of inhibitor. Each solution contained (1 mL total volume) 50 mM HEPES buffer, 100 mM NaCl (pH 7.4), 250 nM *Bt*Fuc2970 and a known concentration of inhibitor; 50 μM CNP-fucoside was added to this to initiate hydrolysis. Data analysis was carried out as described in *2.2.3.2*. Inhibition plots used for the determination of K_i values of **8-13** are displayed in Figure 3.3.
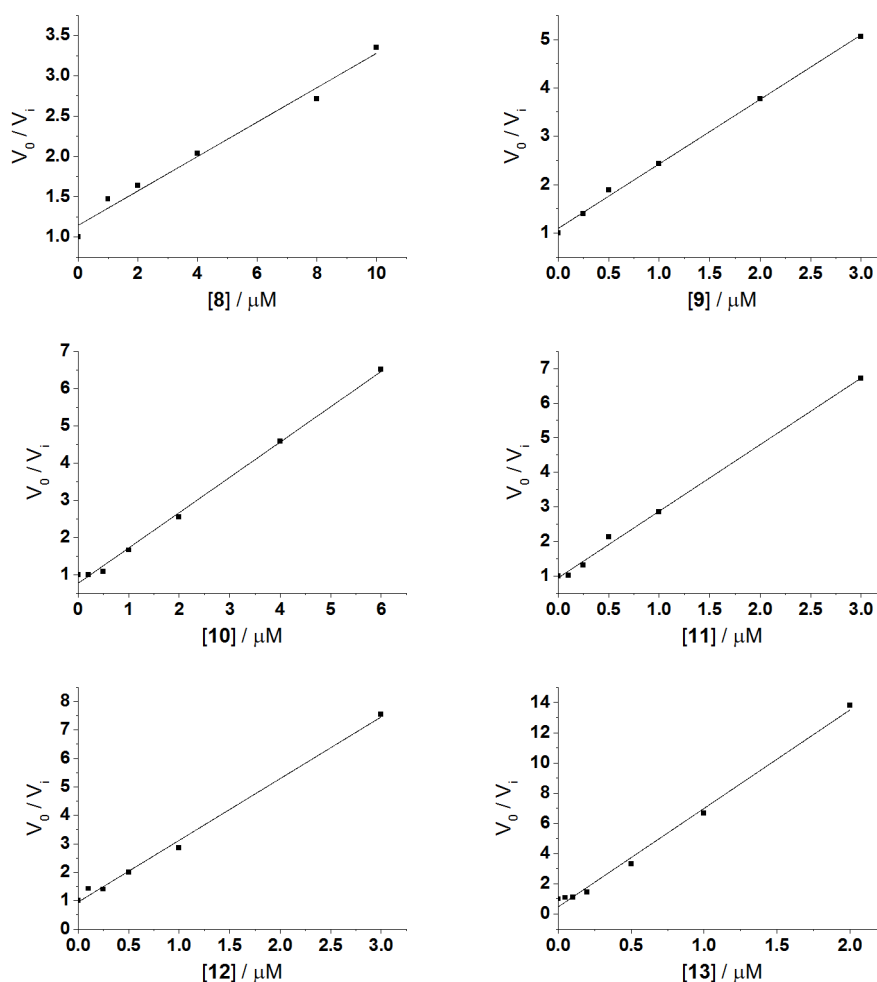
Figure 3.3. Plots showing inhibition of *Bt*Fuc2970 activity by compounds **8-13**.

3.2.4 Crystallisation and structure determination

3.2.4.1 Crystallisation

Crystallisation experiments were conducted as described previously in *2.2.4.1*.

Crystals of *Bt*Fuc2970 liganded with compounds **8-9** were obtained by adding solid powder of each compound directly to a drop containing one or multiple crystals of *Bt*Fuc2970 using an acupuncture needle.

Soaking crystals of *Bt*Fuc2970 with compounds **11** or **12** led to visible crystal cracking and poor diffraction quality. As such, co-crystallisation of *Bt*Fuc2970 with these compounds was effected. To this end, *Bt*Fuc2970 (24 mg mL^{-1}) was incubated with **11** or **12** (solutions of **11** and **12** dissolved at 4 mM in 20 mM HEPES, 100 mM NaCl, pH 7.0) in 1:1 stoichiometry for *ca*. 1 hour prior to setting up crystallisation experiments. Crystals obtained in this way were cryo-protected by transfer into solutions containing mother liquor supplemented with 20% glycerol and subsequently cryo-cooled in liquid N_2.

For compounds **10** and **13**, crystals were transferred to a new crystallisation drop (1 µL, 30% w/v PEG 3350, 0.2 M ammonium sulphate, 0.1 M HEPES, pH 7.0). To this solution, 1 µL of **10** or **13** (dissolved at 10 mM in 10 mM HEPES, 100 mM NaCl, pH 7.0) was added. After incubation for at least *ca*. 30 m, these crystals were transferred to cryo-protectant solutions containing mother liquor supplemented with 20% glycerol and cryo-cooled using liquid N_2.

Diffracting crystals of *Bt*Fuc2970 complexed with each of inhibitors **8-13** were sent to Diamond Light Source (**9-13**) or the European Synchrotron Research Facility (**8**) for data collection.

3.2.4.2 Crystallisation conditions

Unless otherwise noted, crystallisation drops were set up in a 1:1 ratio of protein to mother liquor and *Bt*Fuc2970 was added as a 12 mg mL^{-1} solution (final concentration after dilution 6 mg mL^{-1}). All crystallisation experiments were performed at *ca*. 18 °C.

*Bt*Fuc2970-**8**

Mother liquor: 13.5% w/v PEG 6K, 0.099 M ammonium sulfate, 0.09 M imidazole pH 7

*Bt*Fuc2970-**9**

Mother Liquor: 12.75% w/v PEG 6K, 0.1105 M ammonium sulfate, 0.085 M imidazole pH 7

*Bt*Fuc2970-**10**

Mother liquor: 20% w/v PEG 3350, 0.1 M ammonium sulphate, 0.1 M imidazole pH 7.0

*Bt*Fuc2970-**11**

Mother liquor: 14% w/v PEG 6K, 0.12 M ammonium sulfate, 0.1 M imidazole (pH 7)

Equal volumes of *Bt*Fuc2970 at 24 mg mL^{-1} and **10** (4 mM in 50 mM HEPES 100 mM NaCl (pH 7)) were incubated at room temperature for *ca*. 1 h prior to setting up the crystallisation experiment.

*Bt*Fuc2970-**12**

Mother liquor: 16% w/v PEG 6K, 0.13 M ammonium sulfate, 0.1 M imidazole (pH 7)

Equal volumes of *Bt*Fuc2970 at 24 mg mL^{-1} and **11** (4 mM in 50 mM HEPES 100 mM NaCl (pH 7)) were incubated at room temperature for *ca*. 1 h prior to setting up the crystallisation experiment.

*Bt*Fuc2970-**13**

Mother liquor: 20% w/v PEG 3350, 0.1 M ammonium sulphate, 0.1 M imidazole pH 7.0

3.2.4.3   Structure determination

Diffraction images for crystals of each inhibitor complex were indexed and integrated using either XDS[225] or iMOSFLM[226]. POINTLESS[227] was used to determine the crystallographic space group in each case, and structure factor amplitudes were scaled and merged using SCALA[227] or AIMLESS[228].

The crystals of *Bt*Fuc2970 complexed with compounds **8-9** grew in space groups similar to those described herein in *2.2.4.3* for *apo-Bt*Fuc2970 and *Bt*Fuc2970 complexed with compounds **1-4**. As the space group of each inhibitor complex of *Bt*Fuc2970 liganded with complexes **8-9** was almost isomorphous to the *apo*-crystal,

coordinates from the *apo*-crystal structure were used directly for determination of the inhibitor complex structures; R_{free} sets were assigned as they had been for the *apo*-crystal structure to maintain the uniqueness and integrity of the cross-validation Free R set. Coordinate sets and maximum likelihood refinement target values for each of compounds **8-9** were generated using CHEMDRAW3D and the PRODRG online server[232]. Inhibitor coordinates were initially added at 0.01 occupancy to prevent automatic addition of solvent into ligand density when using the COOT 'find waters' function whilst minimising model bias. At late stages of refinement the ligand occupancy was increased to 1.0.

The crystals for compounds **10-13** grew in a crystal form almost isomorphous with that previously reported on *apo-Bt*Fuc2970 (PDB code 2WVV[184]). In this case coordinates from the *apo*-crystal structure of PDB code 2WVV (non-protein atoms removed) were used directly for determination of the inhibitor complex structures; R_{free} sets were assigned as they had been for the *apo*-crystal structure to maintain the uniqueness and integrity of the cross-validation Free R set. Compounds **10-13** contain ferrocenylamine aglycon moieties that preclude dictionary creation and use in refinement using current versions of PRODRG and REFMAC. Maximum likelihood refinement target values and a coordinate set for each of these ligands were thus kindly generated manually by Garib Murshudov (Medical Research Council Laboratory of Molecular Biology, Cambridge).

Final models for compounds **8-13** were validated using MOLPROBITY.[233] X-ray data for compounds **8-9**, **11-13** were deposited with the RCSB Protein Data Bank (www.rcsb.org[234], Table 3.2).

Table 3.2a. X-ray data collection and refinement statistics for the crystal structures discussed in Chapter 3.

	BtFuc2970-**8**	BtFuc2970-**9**	BtFuc2970-**10**
Data collection			
Beamline/Date	ESRF ID23-2 21.11.2011	DLS I03 12.12.2011	DLS I03 02.02.2014
Wavelength (Å)	0.8726	0.9763	0.97625
Cell dimensions			
a, b, c (Å)	67.5, 95.4, 97.2	68.4, 96.0, 97.2	56.1,187.7,97.6
α, β, γ (°)	90, 90.1, 90	90, 91.3, 90	90,94.2,90
Resolution (Å)	97.1-1.57	48.7-1.68	67.6-1.83
R_{merge}	0.095(0.50)	0.096(0.54)	0.09(0.79)
I / σI	6.6(1.7)	9.7(2.5)	6.8(1.5)
Completeness (%)	99.0(99.6)	100(100)	99.8(99.8)
Redundancy	3.8	3.2	3.8(3.9)
Wilson B value	15.1	21.4	34.7
Refinement			
Resolution (Å)	97.1-1.7	48.7-1.68	67.6-2.10
No. reflections	135052	141053	110778
R_{work} / R_{free}	0.125/0.167	0.166/0.192	0.19/0.23
No. atoms			
Protein	7225	7187	14332
Ligand/ion	96	136	140
Water	1130	1055	773
B-factors (Å^2)			
Protein	16.0	23.7	44.0
Ligand/ion	31.7	44.3	60.7
Water	30.4	34.6	42.9
R.m.s. deviations			
Bond lengths (Å)	0.018	0.011	0.014
Bond angles (°)	1.7	1.3	1.5
Ramachandran Statistics (%)			
Preferred	96.3	96.5	96.1
Allowed	2.9	2.8	2.9
Outliers	0.8	0.7	1
PDB codes	4JL2	4JL1	Not yet deposited

*Values in parentheses are for highest-resolution shell.

Table 3.2b. X-ray data collection and refinement statistics for the crystal structures discussed in Chapter 3.

	*Bt*Fuc2970-**11**	*Bt*Fuc2970-**12**	*Bt*Fuc2970-**13**
Data collection			
Beamline/Date	DLS I04-1 8.3.2012	DLS I03 16.9.2012	DLS I03 02.02.2014
Wavelength (Å)	0.9173	1.200	0.97625
Cell dimensions			
a, b, c (Å)	55.5,187.0,97.7	56.5,189.0,97.5	56.5,188.8,98.1
α, β, γ (°)	90, 94.2, 90	90, 94.1, 90	90,93.9,90
Resolution (Å)	97.4-1.71	97.3-1.72	68.0-2.01
R_{merge}	0.07(0.43)	0.07(0.98)*	0.11(0.90)
$I / \sigma I$	10.9(2.8)	14.4(1.8)	5.9(1.2)
Completeness (%)	96.7(95.0)	99.9(99.9)	99.7(99.7)
Redundancy	4.0	6.3	3.7(3.7)
Wilson B value	23.2	42.1	36.8
Refinement			
Resolution (Å)	97.4-1.88	97.3-2.10	68.0-2.30
No. reflections	155246	112918	85655
R_{work} / R_{free}	0.192/0.222	0.21/0.24	0.20/0.24
No. atoms			
Protein	14138	13859	14259
Ligand/ion	172	155	153
Water	732	215	445
B-factors			
Protein	30.7	54.0	52.2
Ligand/ion	68.7	90.8	102.7
Water	30.2	45.5	44.1
R.m.s. deviations			
Bond lengths (Å)	0.017	0.013	0.013
Bond angles (°)	1.7	1.4	2.4
Ramachandran Statistics (%)			
Preferred	96.7	95.7	94.6
Allowed	2.4	3.4	4.0
Outliers	0.9	0.9	1.4
PDB codes	4JFV	4JFW	Not yet deposited

*Values in parentheses are for highest-resolution shell.

3.3 Results and Discussion

3.3.1 Multivalency effects in fucosidase inhibition

Inhibitory constant K_i values determined for inhibition of BtFuc2970 by compounds **8-10** in this work are displayed in Table 3.3, along with those reported for inhibition of bovine kidney fucosidase[253,255].

Table 3.3. Inhibition of BtFuc2970 and bovine kidney α-L-fucosidases by compounds **8-10**.

Compound	K_i vs. BtFuc2970 / μM (values in parens: fitting error from linear fit in Origin)	K_i vs. bovine kidney fucosidase / μM
8	4.7 (0.3)	2.1
9	0.75 (0.02)	0.3
10	1.1 (0.03)	0.023[255]

The inhibitory potencies of iminocyclitols **8** and **9** as calculated against both BtFuc2970 and bovine kidney α-L-fucosidases are in good agreement, with K_is being slightly weaker against BtFuc2970 by a factor of *ca.* 2.5. **10**, however, appears to be a far more potent inhibitor of bovine kidney α-L-fucosidase (23 nM) than BtFuc2970 (1.1 μM); it is uncertain what the cause of this discrepancy is.

3.3.1.1   X-ray crystallography and structural analysis

In the datasets where compounds **8-10** were soaked into crystals of BtFuc2970, the compounds were unambiguously present in the crystal structures as evidenced by clear peaks in their likelihood-weighted F_o-F_c maps after refinement of protein, solvent and other ions but before the inclusion of compounds **8-10** in refinement (Figure 3.4)
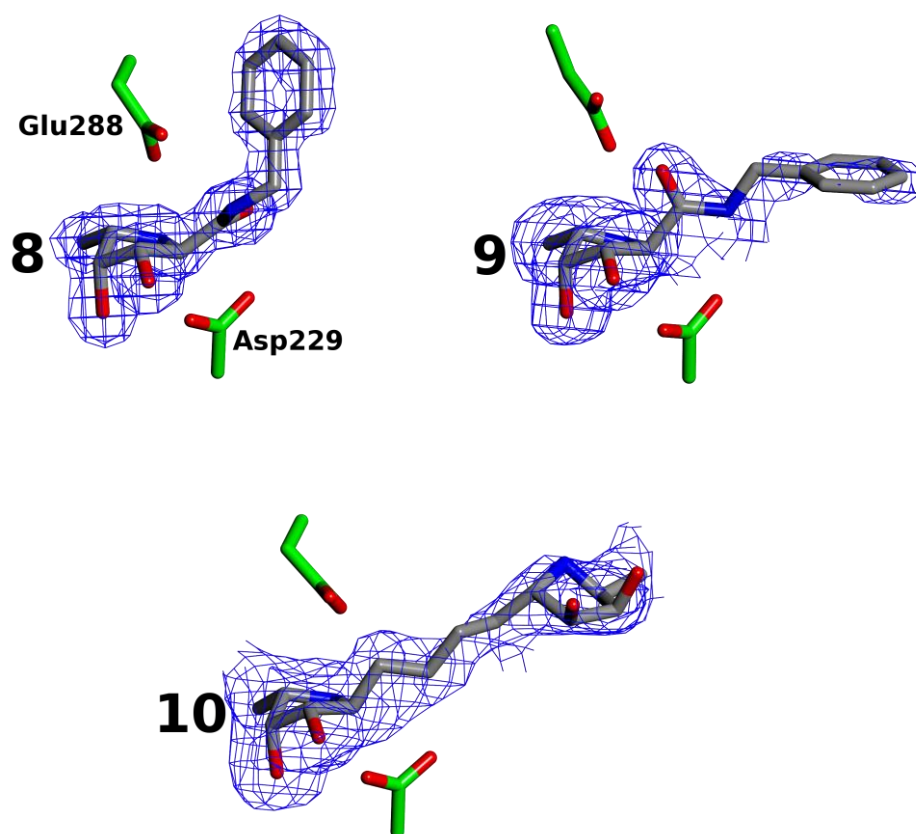
Figure 3.4. Figures of compounds **8-10**, respectively, lying in the active site of *Bt*Fuc2970. Atoms of compounds **8-10** are coloured by atom type (carbon in grey). The catalytic nucleophile Asp[229] and acid/base Glu[288] of *Bt*Fuc2970 are displayed coloured by atom type (carbon in green, residues are annotated for *Bt*Fuc2970:**8**). Electron density from likelihood-weighted F_o-F_c maps calculated before the incorporation of phases from compounds **8-10** in refinement is displayed, contoured at 5 σ (**8**), 3 σ (**9**), or 2 σ (**10**). Figures were drawn using CCP4MG[183].

The interactions made between the iminocyclitol cores of compounds **8-10** and the enzyme are identical to those observed for compounds **1-7**, and these inhibitors bind in an E_3 conformation as seen previously (see section *2.4.2.2*). The aglycons of **8** and **9** observed crystallographically are however poorly ordered. **9** displays two separate binding modes, in one of which the exocyclic amide carbonyl group forms hydrogen bonding interactions with Nε$_2$ of His[166] and Nε$_1$ of Trp[88]. The binding modes of **8** and **9** are displayed in Figure 3.5, highlighting the conformational flexibility displayed by the inhibitor aglycons. Each crystal structure comprises two independent observations of the ligand occupancy, and in each crystal structure, one of the independent observations is better resolved than the other. The aryl groups of **8** and **9** occupy different positions as observed in crystal structure (Figure 3.5).

87

This is thought to be a crystallographic artefact, however, as the 'better' observation of **8** is in close crystal contact with another protomer. The aryl moiety of **9** is observed to stack atop the hydrophobic ridge formed by *Bt*Fuc2970 residues Trp[88] and Trp[232] (this ridge is discussed in more detail in *2.4.2.3*). This strong hydrophobic interaction also helps to explain the more potent α-L-fucosidase inhibition of **10** over **8** and **9** (K_i = 0.023 *vs.* 2.1 and 0.3 μM respectively, Table 3.3).
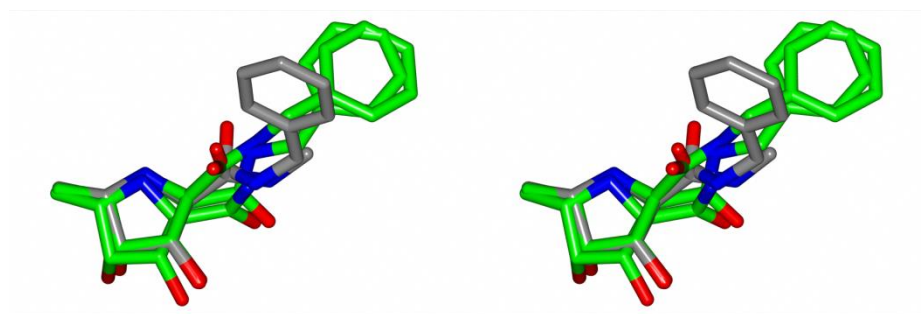
Figure 3.5. Monovalent **8** and trivalent **9** show disorder and alternate binding modes. The independent crystallographic observations of **8** and **9** were superposed, and are displayed as cylinders coloured by atom type (carbon atoms from **8** in grey, from **9** in green). The figure was drawn in wall-eye stereo using CCP4MG.[183]

Structural and kinetic data were analysed to determine the existence of a multivalency effect in the binding potency of α-L-fucosidase inhibitors. The closest distance between adjacent *Bt*Fuc2970 active sites in the crystal packing arrangement is *ca.* 45 Å. This is similar to the closest distance observed in the crystal structure of the GH29 enzyme from *Thermotoga Maritima*, which is slightly lower at *ca.* 40 Å. This large distance would preclude the binding of multiple fuco-configured moieties decorated around a single scaffold (as in **9**) through a chelation mechanism due to the short linker not allowing chelate binding.

Compound **9** is a *ca.* 7 times more potent fucosidase inhibitor than its monovalent counterpart **8**, against both bovine kidney fucosidase and *Bt*Fuc2970 (0.3 *vs.* 2.1 μM and 0.7 *vs.* 4.7 μM respectively, Table 2.). From these data it could be concluded that affininity enhancement due to a statistical rebinding effects was possible, however, a divalent analogue of **8** and **9** is a weaker inhibitor of bovine kidney α-L-fucosidase than its monovalent and trivalent counterparts (K_i = 4 μM vs. 0.3 μM and 2.1 μM, respectively[253]). Other sets of mono/multivalent compounds tested also seemed to show no significant improvement in binding affinity for multivalent

inhibitors over monovalent ones, or even poorer activity for the multivalent species.[253] It is thus perhaps more likely that the affinity enhancement seen in the case of compounds **8** and **9** is due to entropic effects rather than statistical rebinding effects. This is in agreement with the observation that affinity enhancements for multivalent ligands interacting through a statistical rebinding mechanism require there to be a large number of subunits to attain a multivalency affinity enhancement.[237] The expansion of this framework from a small molecule synthesis with low valency of α-L-fucosidase inhibitory moieties to that of larger compounds with much higher valency may however prove to be effective; this approach has been applied before with ligands not configured specifically towards α-L-fucosidase, which were intriguingly shown to activate *Thermotoga maritima* α-L-fucosidase rather than inhibit it[238].

Compound **10** is a divalent α-L-fucosidase inhibitor comprising two 5-membered iminocyclitols configured for α-L-fucosidase inhibitory activity and a short C_4 alkane linker. Interestingly, the entirety of the second iminocyclitol moiety *i.e.* that which is not binding to the enzymatic active site, is observed crystallographically, *cf.* the disorder seen for the aglycon of compound **9**. The crystallographic space groups for *Bt*Fuc2970 complexes with **8** and **9** are different to that for the complex with **10** (Table 3.3a), although as ample solvent channels are observed close to the active site in both structures, this should have no bearing on aglycon conformational flexibility. The second iminocyclitol moiety of **10** is however observed to lie atop Trp[232], one of the two tryptophan residues that form the conserved GH29 hydrophobic ridge (Figure 3.6). Non-specific interactions between carbohydrate moieties and aromatic amino acid side-chains are a common feature of carbohydrate binding, *e.g.* in the binding of sugars other than those flanking the scissile bond in endo-acting glycoside hydrolases with broad substrate specificity.[256] Further, the iminocyclitol ring, which should have partial positive charge at the pH of crystallogenesis and data collection, may make electrostatic interactions with the electron-rich Trp π-system. The binding of the second subunit of **9** to Trp[232] of *Bt*Fuc2970 thus potentially explains the observation of clear electron density for this moiety *cf.* that of the additional subunits of **9**, which may not bind in this position due to sterics.
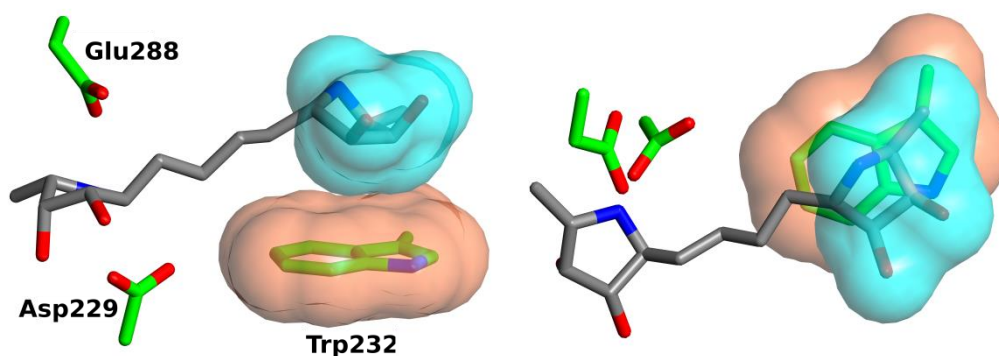
Figure 3.6. The second subunit of **10** binds above residue Trp[232]. Compound **10** and the side-chains of *Bt*Fuc2970 active site residues Asp[229] and Glu[288], and that of residue Trp[232] are displayed as cylinders coloured by atom type (carbon atoms from **10** in grey). Surfaces of Trp[232] and the atoms forming the second subunit of **10** are drawn, coloured in coral (Trp[232]) and cyan (**10**). Orthogonal images were drawn using CCP4MG.[183]

3.3.2   Pharmacophore targeting towards α-ʟ-fucosidase

K_is determined for inhibition of *Bt*Fuc2970 by compounds **11-13** are displayed in Table 3.4, along with those determined for inhibition of bovine kidney fucosidase[255,257].

Table 3.4.   Inhibition of *Bt*Fuc2970 and bovine kidney α-ʟ-fucosidases by compounds **11-13**.

Compound	K_i *vs.* *Bt*Fuc2970 / μM (values in parens: fitting error from linear fit in Origin)	K_i *vs.* mammalian fucosidase / μM
11	0.52 (0.01)	0.29
12	0.46 (0.02)	0.29
13	0.15 (0.005)	0.023[255]

The K_i values determined for inhibition of *Bt*Fuc2970 and bovine kidney α-ʟ-fucosidase by compounds **11-13** are in agreement; it has previously been noted that this class of compound are more potent inhibitors of bovine kidney fucosidases than *Bt*Fuc2970 (**2.3.1, 3.3.1**).

3.3.2.1   X-ray crystallography and structural analysis

In the datasets where compounds **11-13** were either co-crystallised with *Bt*Fuc2970 (**11-12**) or soaked into crystals of *Bt*Fuc2970 (**13**), the compounds were unambiguously present in the crystal structures as evidenced by clear peaks in their likelihood-weighted F_o-F_c maps after refinement of protein, solvent and other ions but before the inclusion of compounds **10-12** in refinement (Figure 3.7).
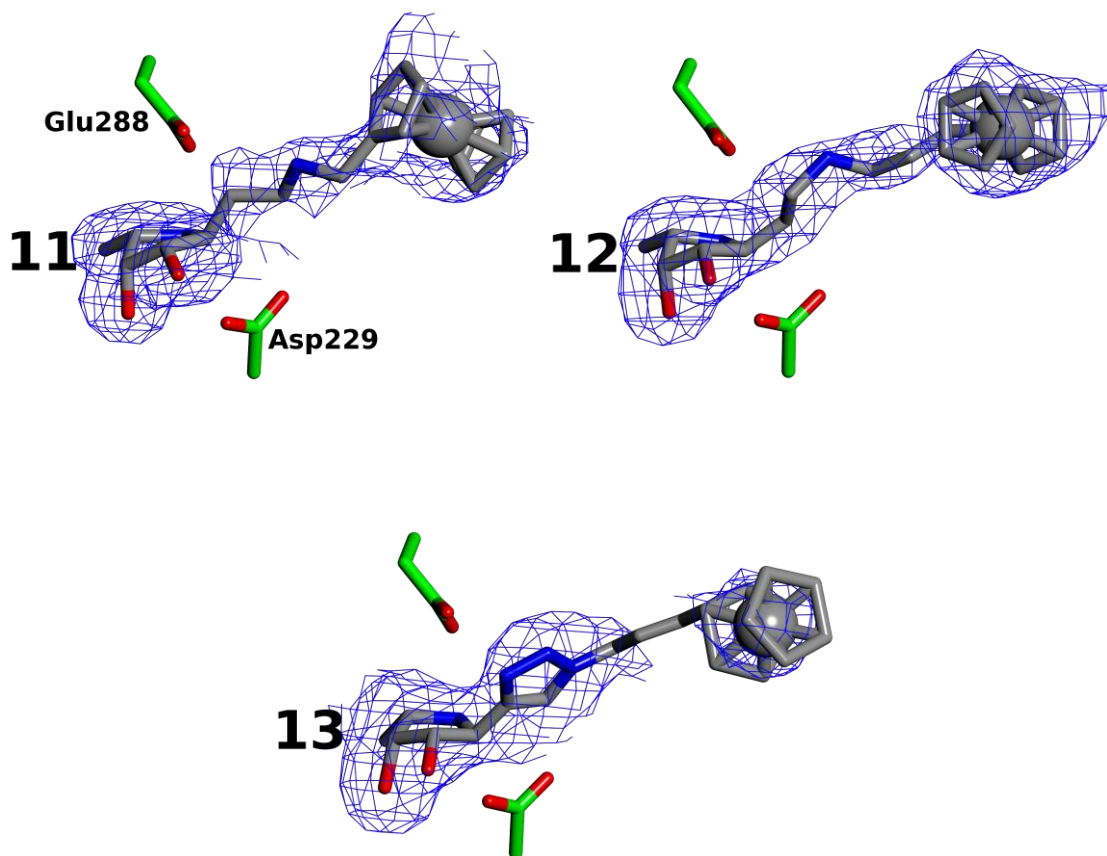
Figure 3.7. Compounds **11-13**, respectively, lying in the active site of *Bt*Fuc2970. Atoms of compounds **11-13** are coloured by atom type (carbon in grey). Fe atoms of compounds **11-13** are displayed as grey spheres. The catalytic nucleophile Asp[229] and acid/base Glu[288] of *Bt*Fuc2970 are displayed coloured by atom type (carbon in green, residues are shown for *Bt*Fuc2970:**11**). Electron density from likelihood-weighted F_o-F_c maps calculated before the incorporation of phases from compounds **11-13** in refinement is displayed, contoured at 3 σ (**11**), 4 σ (**12**) or 2 σ (compound **13**). Figures were drawn using CCP4MG[183].

The aglycons of compounds **11-13** are highly disordered. X-ray data for *Bt*Fuc2970 complexed with **12** were collected at 1.2 Å. At this wavelength, a significant anomalous scattering component is observed that is attributed to the ferrocenyl iron

atom, confirming its presence in the crystal structure (Figure 3.8). The high temperature factors observed for the aglycons of **11-13** represent their disorder, which is likely due to exposure to solvent. Only a small number of protein complexes containing ferrocenyl moieties are available in the PDB, in these structures the ferrocenyl moieties are all bound in buried hydrophobic regions of the protein[258,259] and it is thus no surprise that the ferrocenyl moieties are observed to be conformationally rigid. For the case of **11-13** however, it is reasonable to assume that the ligand aglycons are disordered as they are located far from the enzyme surface and thus solvent-exposed.

The only additional electrostatic interaction observed between compounds **11-13** and *Bt*Fuc2970 compared to compounds **1-10** is a hydrogen bond between the secondary amine moieties of **11-13** and a sulfate group likely abstracted from the crystallisation mother liquor (this can be seen in Figure 3.8).
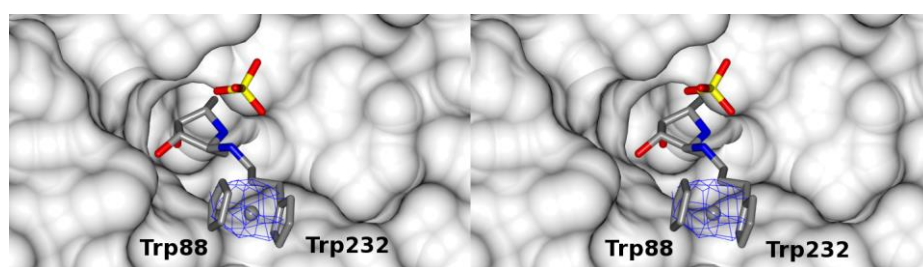
Figure 3.8. The position of the aglycon of compound **11** in the *Bt*Fuc2970 active site is supported using anomalous scattering. *Bt*Fuc2970 atoms are displayed as a white surface. **11** is displayed as cylinders coloured by atom type, with carbon atoms in grey. The electron density map shown is an anomalous difference map contoured at 5 σ (0.019 e Å^{-3}) and is averaged over the four molecules in the asymmetric unit. The figure was drawn in wall-eye stereo using CCP4MG[183].

Compounds **11-13** each display a ferrocenyl pharmacophore appended to the iminocyclitol core by an amine linkage. Ferrocenium salts are potent cytotoxic agents that are thought to act as such by their generation of reactive oxygen species (ROS) in cells, which are thought to induce apoptosis in the affected cell.[250,260] Compounds such as **11-13** were designed to test the hypothesis that the α-ʟ-fucosidase inhibitory moieties they possess may impart selective targeting of the ferrocenyl pharmacophores in these compounds towards cells upregulating α-ʟ-fucosidase (such as neoplastic cells from liver[83], breast[84] and oral cancers[197]). Neoplastic cells are more vulnerable to ROS-induced cell disruption than healthy

cells[251,252], however the selective targeting of a pharmacophore warhead towards cells overexpressing a biomarker such as α-L-fucosidase may further reduce the systemic toxicity imparted by chemotherapeutic agents.

Compounds **11** and **12** have been reported to show anti-proliferative activity towards the MDA-MB-231 breast cancer cell line[254,257], so clearly the ferrocenyl moieties in these compounds have their expected cytotoxic effect. This activity does not appear to be correlated to α-L-fucosidase inhibition, as an isomer of **12**, differing only in its stereochemistry at the anomeric carbon, displays more potent inhibition of MDA-MB-231 cells than **12**, though it exhibits a 25-fold decrease in inhibition of bovine kidney α-L-fucosidase (100% vs. 77% inhibition of MDA-MB-231 cell growth at 50 µM).[254,257]

It is, however, currently unknown whether compounds **11-13**, or any homologous ferrocene-containing α-L-fucosidase inhibitors show selectivity in their cytotoxic activity towards neoplastic cells over healthy cells. Needless to say, the potential for therapeutic use of compounds such as **11-13** would require there to be a significantly greater uptake of these compounds by neoplastic cells over healthy cells, as otherwise the application of many other ferrocene-containing compounds may provide a similar clinical outcome.

3.4   Conclusions

The crystal structures of a mono, di and trivalent compound complexed with the bacterial α-ʟ-fucosidase *Bt*Fuc2970 are reported. These compounds each contain a number of 5-membered iminocyclitol moieties configured towards inhibition of α-ʟ-fucosidase. Further, their inhibition of *Bt*Fuc2970 was evaluated by kinetic studies. Analyses utilising the crystal structures and kinetic data of these enzymes further indicate that a multivalency affinity enhancement is not possible with a chelation effect due to the distance between adjacent GH29 active sites, and that a 'statistical rebinding' multivalency affinity enhancement does not exist with low subunit (≤ 3) valencies, in agreement with the established literature. It may be possible that the multivalency effects often seen in carbohydrate binding interactions *e.g.* in lectins are also applicable to glycoside hydrolases however as, recently, a multivalent iminocyclitol functionalised dextran was shown to possess multivalent affinity enhancements to a number of carbohydrate processing enzymes including the α-ʟ-fucosidase from *Thermotoga maritima*.[238] It is unsure how these multivalent compounds interact, as surprisingly they proved to be activators rather than inhibitors of some of the enzymes tested.

α-ʟ-fucosidase inhibitors linked to ferrocene pharmacophores were complexed with *Bt*Fuc2970 and the structure of the resulting complexes were solved using X-ray crystallography. As α-ʟ-fucosidase is systematically upregulated in a number of cancers including liver, breast and oral cancer, α-ʟ-fucose configured iminocyclitols may provide a delivery method for cytotoxic pharmacophores. Compounds containing ferrocenyl groups are known cytotoxic agents, potentially through their induction of reactive oxygen species-mediated apoptosis. The ferrocene groups of the di-functional compounds were observed to point towards solvent and make no interactions with the protein itself, as such the aglycon moieties were observed to possess significant conformational flexibility; nonetheless, the position of the ferrocenyl Fe atom was in one case experimentally observed by its anomalous scattering component, providing evidence for the correct assignment of ligand position. These compounds have been shown to display anti-proliferative activity against a MDA-MB-231 breast cancer cell line. It is currently unknown, however, whether the iminocyclitol component effectively causes partitioning of these reagents *in vivo* towards neoplastic cells rather than healthy cells. Without the undertaking of this crucial experiment, it will not be known whether this approach has any utility for chemotherapy.

Chapter 4: Covalent trapping of α-ʟ-fucosidase by activity-based probes*

Abstract

Activity-based probes (ABPs) are a class of irreversible enzyme inhibitors which allow the specific trapping of active enzyme from a milieu of species both *in vitro* and *in vivo*. Incorporating, or being derivatised with reporter moieties, they allow the quantification, visualisation and tracking of enzyme activity in a variety of contexts. We report herein the 1.64-1.92 Å X-ray crystal structures of the bacterial α-ʟ-fucosidase *Bt*Fuc2970 covalently trapped with two ABP scaffolds at the enzymatic nucleophile, unequivocally showing that these ABPs react with only active α-ʟ-fucosidase at the enzymatic active site. This is the first time that a fuco-configured small molecule containing an electrophilic trap has been crystallographically observed bound to an α-ʟ-fucosidase enzyme and provides evidence for the postulated mechanism of these ABPs. Fuco-configured ABPs may have potential uses in tracking fertility events *in vivo* and for discovering novel enzymes with α-ʟ-fucosidase activity.

4.1 Introduction

4.1.1 Activity-based probing

The genomic revolution opened up a wealth of opportunities towards a more detailed understanding of human physiology and disease. In the past, classical techniques such as proteomics and transcriptomics have been useful in elucidating the biological roles of proteins and nucleic acids. To further our understanding of the roles of these species in physiology and disease, however, requires more information than is available by using these techniques, which are often applied to the understanding of global transcription or translation levels. This is due to the fact that the biological function of a given species of interest is correlated more firmly with activity than mere quantity. Many proteins, for example, require post-translational events to occur in order for them to carry out their intended function, such as their complexation with other proteins or nucleic acids, glycosylation[109] or phosphorylation[261], correct localisation in a target compartment, or the proteolytic cleavage of a pro-peptide into a mature form[262]. These events rely not entirely on the genome-level transcription of a protein, but rather on a complex array of factors which may be difficult to disentangle and prove intractable to mimic *in vitro*. Thus, there is a need for tools which can specifically track active enzyme activity *in vivo* and on a biological timescale, and with minimal perturbation of the system under study.

Molecular imaging has proven a powerful tool towards this aim, as the non-invasive technique allows the detailed spatiotemporal characterisation of specific targets *in vivo*, or of living cells *ex vivo*[263,264]. This technique often involves the use of molecular probes as reporter groups which have an output visible to microscopy techniques, and can be used to view biological processes in real time in intact cells. Molecular probes are used to effect detection of the biological target, and thus should have high specificity towards this target.

A number of strategies exist for the design of probes with specificity towards their biological target such as the inclusion of, *inter alia*, radiolabelled small molecules and monoclonal antibodies[264]. For the case of enzymes, known catalytic activity can be exploited in order to achieve this specificity; compounds that are designed in such a way are termed activity-based probes (ABPs) and act by covalently trapping an enzyme on the basis of its catalytic mechanism[265-267].

Activity-based probes contain a number of features allowing their effective use (Figure 4.1)[265]. Chemically reactive warheads allow the covalent and irreversible

trapping of an enzyme target. Specificity towards a particular enzyme class, and potentially to a particular subset of enzymes within a given class, is governed by specificity elements; the design of a specificity element allows tuning of reactivity to be either very specific, or broad, depending on intended function. Finally, the incorporation of a probe or reporter moiety in an ABP allows an output dependent on covalent binding of the ABP to its target, and may include the use of fluorogenic substrates for quantification or visualisation of active enzyme in the molecular imaging context and/or affinity tags for the isolation of active enzymes of a given class from a milieu of species.
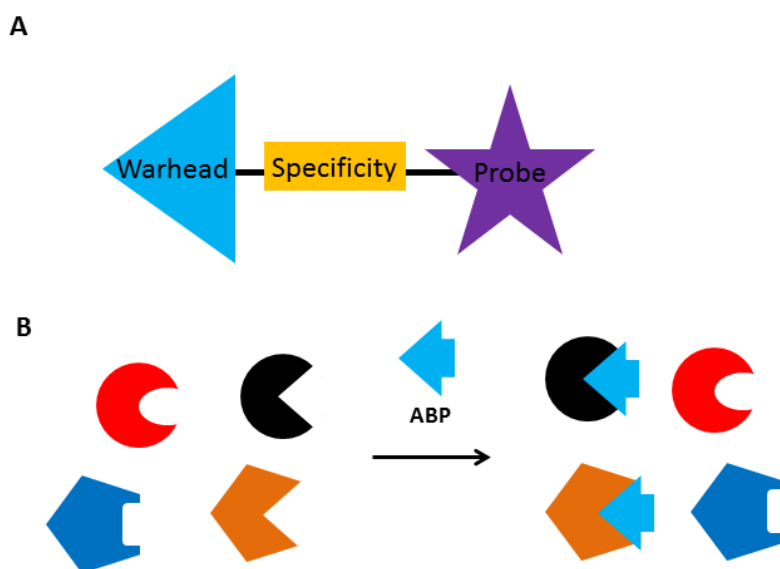
Figure 4.1. Structural features of activity-based probes (**A**). Warheads are designed to covalently and irreversibly trap an enzyme based on its mechanism. The specificity element of an ABP targets its reactivity towards active enzymes on the basis of reaction mechanism. ABP probes may be quantitative reporter elements or affinity tags, and may be introduced by derivatisation *in vivo*. Activity-based protein profiling shema (**B**). ABPs are introduced to a milieu of enzymes *in vitro* or *in vivo* but specifically label a target enzyme class based on activity.

The emergence of bio-orthogonal chemistry as a powerful tool available to the enterprising chemical biologist has had many implications on the field of activity-based protein profiling (ABPP)[267]. This technique can be used for the generation of two-step ABPs, that is ABPs that do not contain an intact probe, but contain a linker for functionalisation *in situ* after covalent binding towards their biological target (Figure 4.2). This strategy offers a number of advantages over the direct

inclusion of probe moieties in an ABP[267]. Perturbation to the biological system is likely to be weaker using a two-step approach, and binding of the ABP to its target may be more favourable as it more closely resembles a natural substrate. Binding of the ABP to off-targets may also be minimised by use of a two-step labelling regime, as the lipophilicity of a compound is thought to be linked to its non-specific binding[264]. Further, transport of an ABP across the cell membrane for *in vivo* experiments will likely be more favourable without the addition of hydrophobic reporter moieties. Examples of bio-orthogonal reactions that have been used in the ABPP field include the Staudinger ligation between azide and phosphine moieties[268,269], and both Cu(I) catalysed and Cu(I)-free azide-alkyne cycloaddition reactions[270,271].
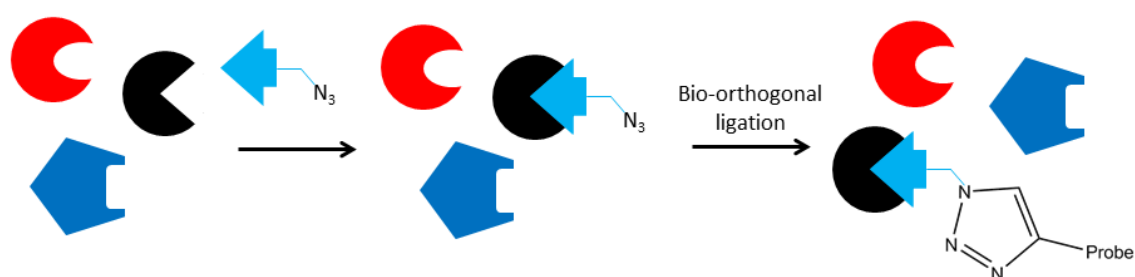
Figure 4.2. The mechanism of activity-based probing by a two-step process. Bio-orthogonal ligation using the azide-alkyne cycloaddition reaction is a representative example of methods which may be used.
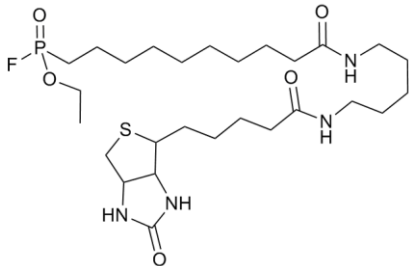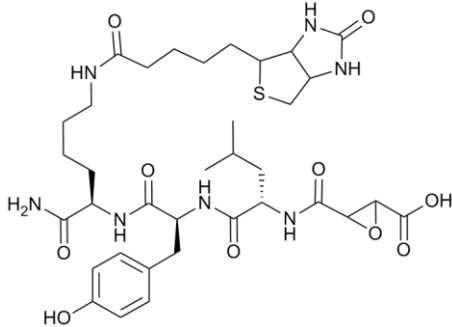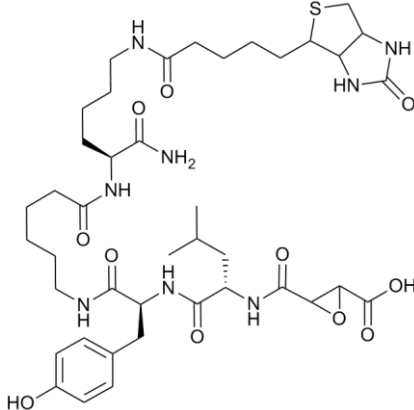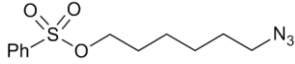
The potential scope of ABPP was highlighted by a seminal paper shortly before the completion of the human genome project. Liu and co-workers reported the synthesis of **XVIII** (Table 4.1), a compound designed from the fusion of the known serine protease inhibitor moiety fluorophosphonate with biotin, allowing the capture of trapped proteins through the use of streptavidin affinity chromatography. The authors further showed that this compound was able to label a number of serine proteases from rat testis. This selective labelling of a protein family in a complex medium containing many other species was concluded to be a powerful tool, allowing the interrogation of activity (rather than transcription) levels and dynamics thereof, of this enzyme class, which has a number of roles in physiology and disease.

Shortly after this, Greenbaum *et al*. reported the synthesis of two cysteine protease ABPs **XIX-XX** (Table 4.1), containing epoxide moieties as an electrophilic trap. These compounds, based on a natural product known to irreversibly inhibit the

enzyme class, contained both biotin moieties and an iodinatable phenol moiety. While **XIX** was shown to have activity against a broad panel of cysteine proteases, compound **XX** proved to be specific against the cathepsin family of proteases. Further, by using compound **XIX**, they were able to visualise the changes in cysteine protease activity as a function of disease progression in multiple tumour cell lines, illuminating the role of these enzymes in cancer development.

A two-step approach to ABPP was first exploited by Cravatt and co-workers in 2003, allowing, for the first time, application of the ABPP platform to the study of intact proteomes *in vivo*[270]. The authors demonstrated that azide-containing sulfonate **XXI** (Table 4.1) could be functionalised *in vivo* by Cu(I) catalysed Huisgen cycloaddition[272] with rhodamine alkyne. This two-step approach allowed the visualisation by fluorescence SDS-PAGE of glutathione *S*-transferases *in vitro*, in cultured COS cells and in heart tissue homogenates of mice treated with **XXI** and subsequently reacted with rhodamine alkyne. This coupling of the powerful tool of bio-orthogonal reactions with ABPP heralded its application across a variety of enzymatic and cellular contexts.

Table 4.1. Some activity-based probes described in the literature.

Compound number	Structural formula
XVIII	
XIX	
XX	
XXI	

Activity-based protein profiling was first brought into the glycoside hydrolase field a year later in 2004, when Vocadlo & Bertozzi demonstrated labelling of a number of β-galactosidases by a di-fluorosugar[269]. Their ABP, 6-azido-2,6-dideoxy-2-fluoro-β-D-galactosyl fluoride **XXII** (Table 4.2), was capable of irreversibly trapping β-galactosidase by stabilisation of the covalent enzyme intermediate via the 2-fluoro substituent (*vide supra*), and derivatisation of its pendant azide moiety *in situ* by the bio-orthogonal Staudinger ligation[268] resulted in a FLAG-tagged biochemical probe capable of labelling a number of β-galactosidase enzymes tested. The attachment of the sterically undemanding azide tag to the C6 position and subsequent derivatisation (*i.e.* use of a two-step approach) was posited by the authors to be critical for efficient labelling of the enzyme, due to the limited

hydrophobic cavities present in *exo*-acting glycoside hydrolases. Since this first-in-class, a number of studies reporting ABPP of glycosidases have been reported.
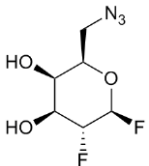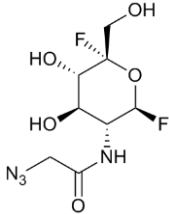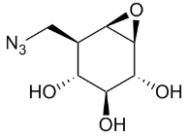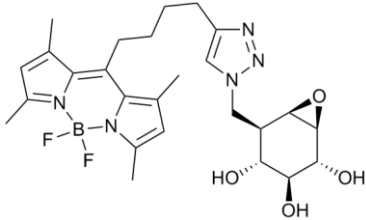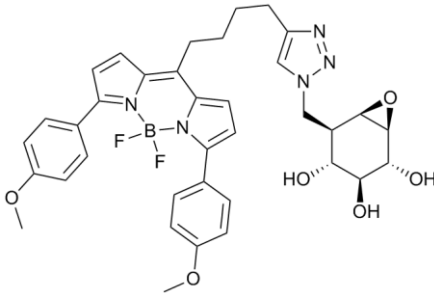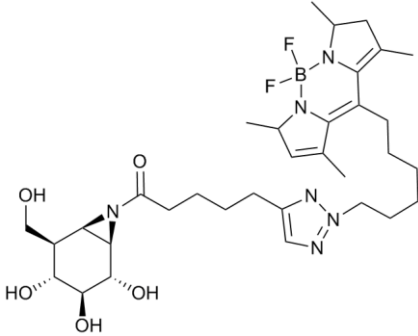
ABPP may be used to detect novel enzymes with a given activity, as was shown elegantly by Vocadlo and co-workers in 2008[273]. After the deposition of an inhibitor-bound crystal structure of a *Vibrio cholerae* GH3 family *exo*-glucosaminidase, the authors noted the presence of a large cavity around the inhibitor 2-acetamido group. Postulating that this cavity may allow attachment of a bio-orthogonal tag for two-step ABPP, the group synthesised compound **XXIII** (Table 4.2), which was indeed able to irreversibly inhibit the *Vibrio cholerae* GH3 enzyme. Functionalisation of this ABP was effected with both FLAG-peptide and biotin moieties. Further, application of **XXIII** to cultures of the opportunistic pathogen *Pseudomonas aeruginosa* and subsequent bio-orthogonal ligation with a biotin-alkyne allowed identification of a potentially clinically important novel GH3 family enzyme from this organism.

The use of epoxide containing, sugar configured probes by a number of research groups in the Netherlands[274] exemplifies the potential to use ABPP to track active enzyme dynamics *in vivo*. ABPs **XXIV-XXVI** (Table 4.2) reported by these groups are selective and incredibly potent irreversible inhibitors of glucocerebrosidase I, the major lysosomal enzyme that catalyses the hydrolysis of glucosylceramide; deficiency in activity of this enzyme causes the lysosomal storage disorder Gaucher disease[275]. Interestingly, use of a two-step ABPP experiment was shown to not be needed against this enzyme as, serendipitously, ABPs **XXV** and **XXVI** display more potent binding to the enzyme than their azido-counterpart **XXIV** (K_i = 44 nM for **XXIV** v. 7 and 8 nM for **XXV** and **XXVI** respectively), and the BODIPY ABPs were still able to enter the cellular compartment *in vivo*. By utilising the two BODIPY ABPs **XXV-XXVI**, which have different fluorescence emission spectra, the authors were able to conduct pulse-chase experiments on cultured fibroblasts. Incubation overnight with red-fluorescent **XXVI**, followed by the addition of green-fluorescent **XXV** at set time points allowed the authors to approximate the half-life of glucocerebrosidase I residency in the lysosome at 30 h, agreeing with literature values determined using conventional pulse-chase labelling[276]. Further, the authors were able to visualise the differences in glucocerebrosidase I activity in the fibroblasts of Gaucher patients with various mutant genes. This result in particular showed promise for **XXV-XXVI** as tools for the diagnosis of Gaucher disease.

One potential shortcoming of compounds **XXV-XXVII** as ABPs is that they are specific to glucocerebrosidase I, and have no activity against other retaining *exo*-β-glucosidases, glucocerebrosidases II and III and lactase/phlorizin hydrolase; these

enzymes may also have clinical relevance as glucocerebrosidase III is thought to be involved in the detoxification of plant glycosides[277] and lactase/phlorizin hydrolase deficiency is the cause of lactose intolerance[278]. Noting that all *exo*-β-glucosidases are able to hydrolyse 4-methylumbelliferyl-β-ᴅ-glucopyranoside, Overkleeft and co-workers developed aziridine ABPs such as **XXVII** (Table 4.2), which contain an electrophilic trap between the positions that would equate to the endocyclic oxygen and the anomeric carbon of a natural substrate[279]. Upon formation of the covalent enzyme intermediate, they expected that the pendant amide moiety would occupy approximately the position of the aglycon in natural substrates. Thus, the authors posited that this approach should have the ability to label not just glucocerebrosidase I, but also the other *exo*-β-glucosidases, which indeed proved to be the case. Further, as this approach exploits the presence of an enzymic cavity at the non-reducing end of the -1 sugar, which is a feature in common to all *exo*-glycosidases, this approach should have broad utility against this enzyme class[279].
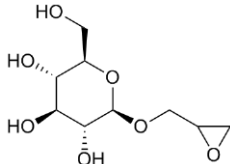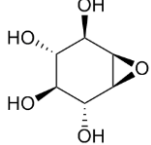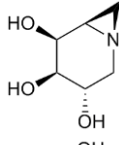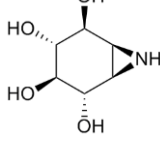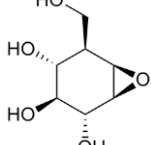
Table 4.2. Some activity-based probes active against glycoside hydrolases described in the literature.

Compound number	Structural formula
XXII	
XXIII	
XXIV	
XXV	
XXVI	
XXVII	

4.1.2 Glycosidase inhibition by ring-strained compounds

Ring-strained heterocycles are potent electrophiles and have use for the study of retaining glycoside hydrolases. Epoxyalkyl glycosides such as **XXVIII** (Table 4.3) have been used historically to elucidate the catalytic mechanism of glycoside hydrolases including, *inter alia*, lysozyme[280] and cellulase[281]. Enzymatic catalysis occurs preferentially to the epoxyalkane moiety, labelling either catalytic carboxylate. The covalently trapped glycosidase can further be observed crystallographically[282,283]. Unfortunately, labelling may be on either of the catalytic carboxylates so their assignment requires further experimental validation[282].

Table 4.3. Some ring-strained heterocycles displaying inhibition of glycoside hydrolases described in the literature.

Compound number	Structural formula
XXVIII	
XXIX	
XXX	
XXXI	
XXXII	

Prior to the availability of bioinformatics approaches to identify the catalytic nucleophile of glycoside hydrolases[170], conduritol B-epoxide **XXIX** (Table 4.3) was used for this purpose against a number of glucosidases such as the human enzymes sucrose-isomaltase[284] and α-glucosidase[285]. As a stable covalent intermediate is formed by nucleophilic attack of a glycoside hydrolase upon **XXIX**, peptic digestion and mass spectrometry allowed the identification of candidate enzymatic nucleophile residues.

Aziridines such as **XXX** (Table 4.3) offer advantages over epoxides in the study of glycoside hydrolases, as aziridine moieties are protonated at physiological pH and may form favourable electrostatic interactions with the negatively charged catalytic active site. The first use of an aziridine-based glycoside hydrolase inhibitor was reported in 1988 by Tong *et al*.[286], where the authors showed irreversible inhibition of coffee bean α-galactosidase by compound **XXX**. A year later, Caron and Withers rationalised that glycosidase inhibition potency should be increased by merging the reactivity of epoxy-glycosides with the benefits of a protonated nitrogen atom discussed above and reported the synthesis and evaluation of conduritol aziridine **XXXI** (Table 4.3)[287]. This compound did indeed prove to be a potent irreversible inhibitor of both α- and β-glucosidases.

A number of glycosidases complexed with ring-strained heterocyclic inhibitors have been observed by X-ray crystallography. In 2005, Sussman and co-workers determined the crystal structure of conduritol B-epoxide **XXIX** (Table 4.3) complexed with human glucocerebrosidase I[288]. The observation of a covalent adduct between residue Glu[340] and the anomeric carbon of **XXIX** in the 2.4 Å crystal structure confirmed assignment of the catalytic nucleophile and informed identification of two conformationally flexible loop regions. These regions, which were conformationally locked in the complex structure but showed flexibility in the apo structure of glucocerebrosidase I[289], were postulated to act as a lid controlling active site access; this finding is supported by Gaucher mutations in one of these loops.

The 1.9 Å X-ray crystal structure of cyclophellitol **XXXII** (Table 4.3) complexed with a GH1 family β-glucosidase from *Thermotoga maritima* was reported in 2007[290]. Cyclophellitol, a natural product first derived from the mushroom *Phellinus* sp.,[291] was observed covalently bound to the enzyme in a $^4C_1$ conformation.

ABP glycosides configured with ring substituents mimicking those of L-fucose may have utility *e.g.* as biomarkers for cancer[93,197] or for cell biology research, particularly in fertility, due to the known importance of α-L-fucosidase in this field[109]. Herein, the X-ray crystal structures of the bacterial α-L-fucosidase *Bt*Fuc2970 complexed with two such glycosides that have specificity towards GH29 retaining α-L-fucosidases are reported. In the enzyme-ligand complex structures (1.64 – 1.92 Å), covalent bonding is observed between the ligands and the enzymatic nucleophile, unequivocally showing that these ABPs react with only active α-L-fucosidase at the enzymatic active site. Surprisingly, one of the compounds appears to bind not in a low energy $^3S_1$ conformation, but in the $^3H_4$

conformation expected for the catalytic transition state. We would like to thank collaborators in the research group of Prof. Hermen Overkleeft, without whom this work would not have been possible.

4.2 Materials and methods

4.2.1 Gene expression and protein purification
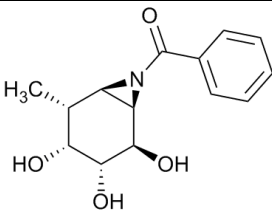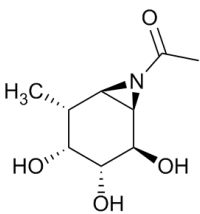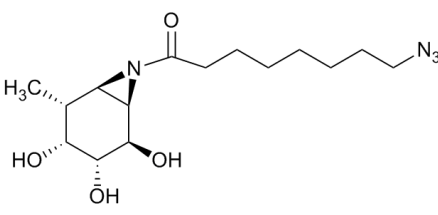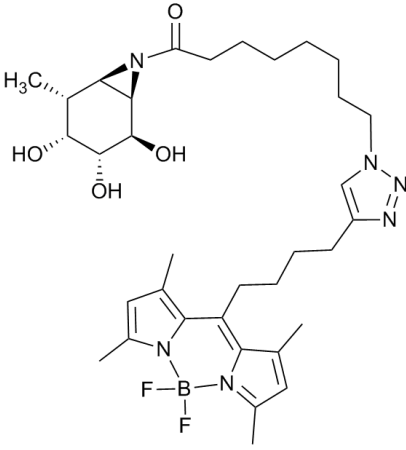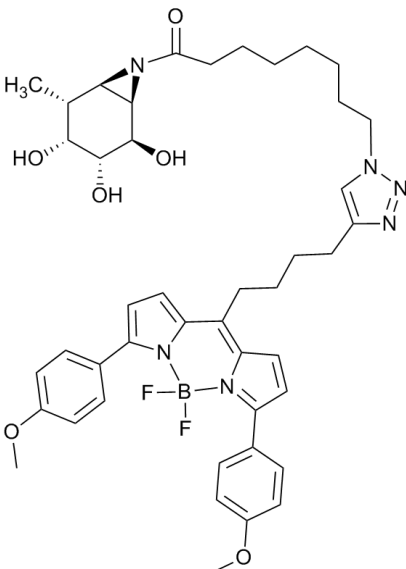
Expression and purification of *Bt*Fuc2970 was effected as described previously in **2.2.1**.

4.2.2 Activity-based probes and scaffolds

Compounds **14-19** (Table 4.4) were received from collaborators, namely the research group of Prof. Herman Overkleeft at the Leiden Institute of Chemistry, Leiden University. These compounds either display the characteristics of ABPs (**17-19**), or have a structure which represents the same reactivity of ABPs against an α-L-fucosidase target whilst not possessing a reporter moiety (**14-16**, **16** may however be derivatised by bio-orthogonal chemistry to yield **17-19** or other compounds bearing a reporter moiety). Syntheses of these compounds are detailed in [292].

Compounds **14-19** have aziridine moieties, located between the positions of a sugar substrate that would correspond to the endocyclic oxygen and C1. The stereochemistry of the positions which would equate to C2-C5 of a sugar substrate are the same as L-fucose. Thus, these compounds were designed to be specifically active towards α-L-fucosidase as a target.

Table 4.4 Activity-based probes and similar compounds described in chapter 4.

Compound number	IUPAC name	Structural formula
14	Phenyl((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)methanone	
15	1-((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)ethan-1-one	
16	8-Azido-1-((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)-octan-1-one	
17	8-(4-(4-(5,5-Difluoro-1,3,7,9-tetramethyl-5H-4l4,5l4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)-1-((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octan-1-one	
18	8-(4-(4-(5,5-Difluoro-3,7-bis(4-methoxyphenyl)-5H-4l4,5l4-dipyrrolo[1,2-c:2',1'-f]-[1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)-1-((1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octan-1-one	

| 19 | N-((1-(8-oxo-8-((1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)-6-(5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)hexanamide | |
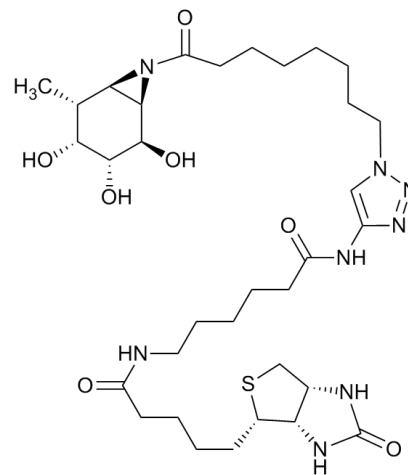
4.2.3 Crystallisation and structure determination

4.2.3.1 Crystallisation

General practice for crystallisation experiments was as described previously in *2.2.4.1*.

Crystals of *Bt*Fuc2970 liganded with compounds **14** and **15** were obtained by adding 1 μL compound dissolved at 5 mM in 10 mM HEPES, 100 mM NaCl, pH 7.0 (**14**) or 20 mM in 0.1 M HEPES, 0.2 M ammonium sulfate, 25% w/v PEG 3350, pH 7.5 (**15**) to drops containing crystals of *Bt*Fuc2970. After incubation for at least *ca.* 30 m, these crystals were transferred to cryo-protectant solutions containing mother liquor supplemented with 20% glycerol and cryo-cooled using liquid N$_2$.

Diffracting crystals of *Bt*Fuc2970 complexed with compounds **14** and **15** were sent to Diamond Light Source for data collection.

4.2.3.2 Crystallisation conditions

Unless otherwise noted, crystallisation drops were set up in a 1:1 ratio of protein to mother liquor and *Bt*Fuc2970 was added as a 12 mg mL^{-1} solution (final concentration after dilution 6 mg mL^{-1}). All crystallisation experiments were performed at *ca.* 18 °C.

*Bt*Fuc2970-**14**

Mother liquor: 25% w/v PEG 3350, 0.2 M ammonium sulfate, 0.1 M HEPES (pH 7.5)

*Bt*Fuc2970-**15**

Crystals were obtained through using the micro-seeding technique[293]. A seed stock was created by transferring micro-crystals of *Bt*Fuc2970 grown in 15% w/v PEG 3350, 0. M ammonium sulfate, 0.1 M HEPES (pH 7.5) to a microcentrifuge tube containing a seed bead (Hampton Research) and supplementing with 50 µL mother liquor. Microcrystals were then crushed by the use of a vortex mixer and diluted to concentrations of $10^{-1} - 10^{-3}$ in mother liquor.

Final crystals were obtained through seeding at a dilution of 10^{-3} into a solution containing mother liquor: 21% w/v PEG 3350, 0.2 M ammonium sulfate, 0.1 M HEPES (pH 7.5). The final solution in the crystallisation drop comprised a ratio of 0.5 protein: 0.4 mother liquor: 0.1 seed stock.

4.2.3.3 Structure determination

Diffraction images for crystals of each inhibitor complex were indexed and integrated using either IMOSFLM[226] (*Bt*Fuc2970-**14**) or XIA2[294] (*Bt*Fuc2970-**15**). AIMLESS[228] was used for data reduction in each case; for *Bt*Fuc2970-**15**, unmerged XIA2 data were merged and an R_{free} flag set using AIMLESS.

The crystals of *Bt*Fuc2970 complexed with compounds **14** and **15** grew in an almost isomorphous space group to PDB entry 4JFV; and protein atom co-ordinates from this entry were used directly to obtain starting phases for refinement. R_{free} sets were assigned as they had been for 4JFV to maintain the uniqueness and integrity of the cross-validation R_{free} set.

Coordinate sets and maximum likelihood refinement target values for compounds **14** and **15** were generated using CHEMDRAW3D and the PRODRG online server[232]. Inhibitor coordinates were initially added at 0.01 occupancy to prevent automatic addition of solvent into ligand density when using the COOT 'find waters' function whilst minimising model bias. At late stages of refinement the ligand occupancy was increased to 1.0. Dictionaries defining the link between the catalytic nucleophile of *Bt*Fuc2970 (Asp229) and compounds **14** and **15** were generated using JLIGAND[295].

Final models for *Bt*Fuc2970-**14** and *Bt*Fuc2970-**15** were validated using MOLPROBITY[233] and deposited with the RCSB Protein Data Bank (www.rcsb.org[234]).

4.2.3.4   *Conformations adopted by ABPs in crystal structures*

The conformations adopted by **14** and **15** in their crystal structures bound to *Bt*Fuc2970 were kindly determined and plotted by Dr. Carme Rovira at the University of Barcelona. Cremer-Pople parameters[296] were determined from PDB entries 4WSK and 4WSJ, averaged over the 3 or 4 ligands present in the PDB entry (for **14** and **15**, respectively). These parameters were then plotted on the conformational free energy landscape of *Bt*Fuc2970, which was previously calculated[184]; further methodological detail is available in the supporting information of this reference.

4.3 Results and Discussion

4.3.1 Crystal structure of *Bt*Fuc2970 covalently trapped by ABPs

Initially, a number of attempts were made to determine crystal structures of *Bt*Fuc2970 complexed with compounds **14**, **16-19**, by adding both solid powder of these compounds, and solubilised compounds to crystallisation drops containing crystals of *Bt*Fuc2970. Ligand binding was, however, only evident in one case, for compound **14**.

Diffraction images for this complex were collected at 1.73 Å resolution, however the data quality was anisotropic according to AIMLESS, and data were refined to 1.92 Å, as this was the recommended limit of data usefulness on the basis of half-dataset correlation[228](Table 4.5). **14** was observed to be bound to *Bt*Fuc2970 clearly in 3 of the 4 monomers in the asymmetric unit, before the introduction of phases from the ligand in refinement (Figure 4.3). The phenyl moiety of the aglycon of **14** was however poorly ordered and could only be modelled in 1 monomer of the 4 present in the asymmetric unit. The final refined R_{work} and R_{free} values for this complex structure are 0.18 and 0.23 respectively.
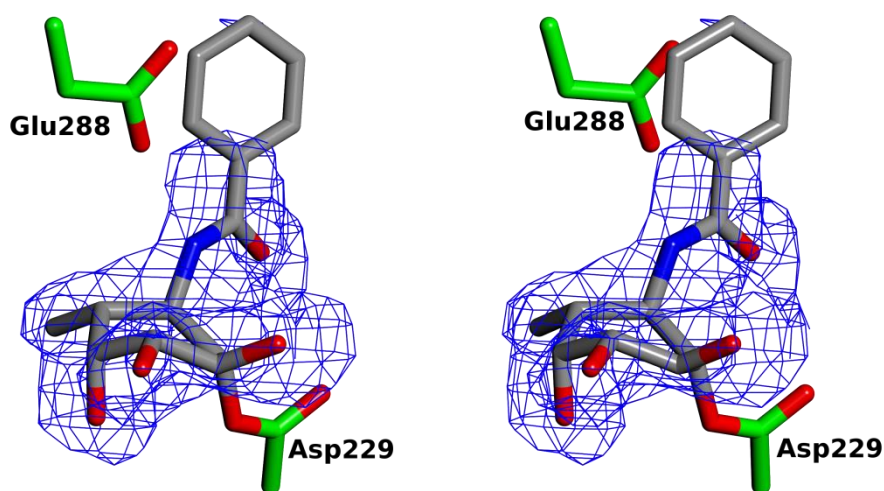
Figure 4.3. Compound **14** lying in the active site of *Bt*Fuc2970, displayed in wall-eye stereo. Atoms of compound **14** are coloured by atom type (carbon in grey). The catalytic nucleophile Asp[229] and acid/base Glu[288] of *Bt*Fuc2970 are displayed coloured by atom type (carbon in green). The map displayed is a likelihood-weighted F_o-F_c map from before the incorporation of phases from compound **14** in refinement, contoured at 2 σ (**14**). The figure was created using CCP4MG[183].

Table 4.5. X-ray data collection and refinement statistics for the crystal structures discussed in chapter 4.

	*Bt*Fuc2970-**14**	*Bt*Fuc2970-**15**
Data collection		
Beamline/Date	DLS i03 02/02/14	DLS i03 18/10/14
Wavelength (Å)	0.97625	0.97625
Cell dimensions		
a, b, c (Å)	55.6, 186.5, 98.2	55.5, 187.0, 98.2
α, β, γ (°)	90, 94.2, 90	90, 94.2, 90
Resolution (Å)	62.17-1.73	93.5-1.64
R_{merge}	0.095(0.86)*	0.058(0.62)
I / σI	7.0(1.8)	11.2(2.0)
Completeness (%)	96.9(96.7)	98.1(98.5)
Redundancy	4.0(3.7)	4.1(4.3)
Wilson B value	26.1	20.4
Refinement		
Resolution (Å)	97.9-1.92	97.9-1.64
No. reflections	292076	328440
R_{work} / R_{free}	0.18/0.23	0.16/0.19
No. atoms		
Protein	14267	14073
Ligand/ion	107	81
Water	1204	1618
B-factors (Å^2)		
Protein	31.8	25.1
Ligand/ion	41.9	24.3
Water	36.8	35.3
R.m.s. deviations		
Bond lengths (Å)	0.019	0.019
Bond angles (°)	1.81	1.78
Ramachandran Statistics (%)		
Preferred	95.8	96.7
Allowed	3.1	2.4
Outliers	1.1	0.9
PDB codes	4WSK	4WSJ

*Values in parentheses are for highest-resolution shell.

The postulated mechanism by which sugar-configured aziridine-containing compounds act to inhibit glycoside hydrolases is by trapping of the enzyme as a covalent intermediate, as has been shown previously for the epoxide-containing natural product cyclophellitol complexed with β-glucosidase[297]. S_N2-like nucleophilic attack of the catalytic nucleophile towards the anomeric position of a compound such as **14** would be expected to lead to trans-diaxial ring-opening of the aziridine and the formation of a stable covalent intermediate (Figure 4.4). The crystal structure of *Bt*Fuc2970-**14** clearly supports this mechanism, as clear electron density is observed for a covalent interaction between the catalytic nucleophile of *Bt*Fuc2970 and **14** (Figure 4.5), and since the covalent bond distances refine well to between 1.43 and 1.47 Å, as would be expected of an ester C-O bond.
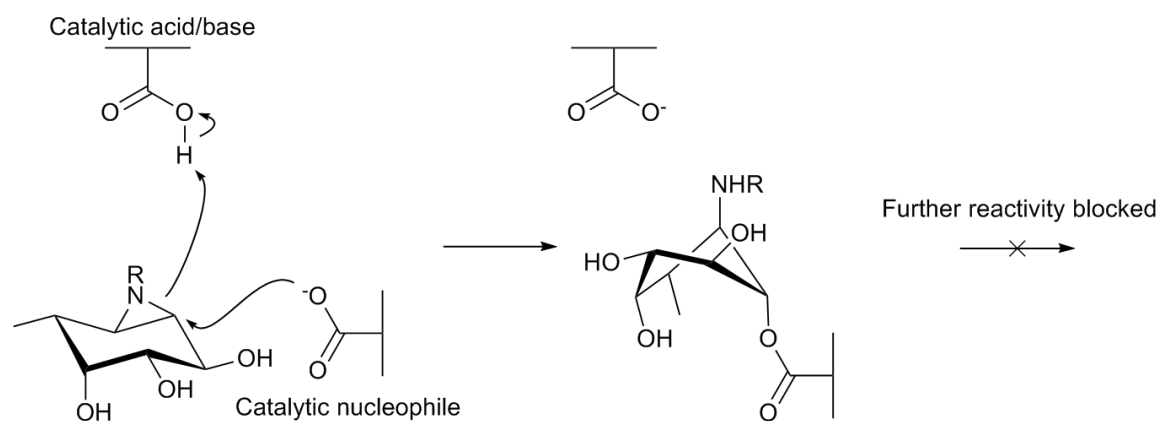
Figure 4.4. Proposed reaction schematic for ABP compounds towards α-L-fucosidase enzymes.

The interactions made between *Bt*Fuc2970 and **14** are displayed in Figure 4.6. The cyclohexane core of **14** is held tightly in place by hydrogen bonding interactions, and further stabilised by a hydrophobic interaction between Trp[316] and the C4, O4, C5, C6 atoms of **14** (numbering scheme as for L-fucose).
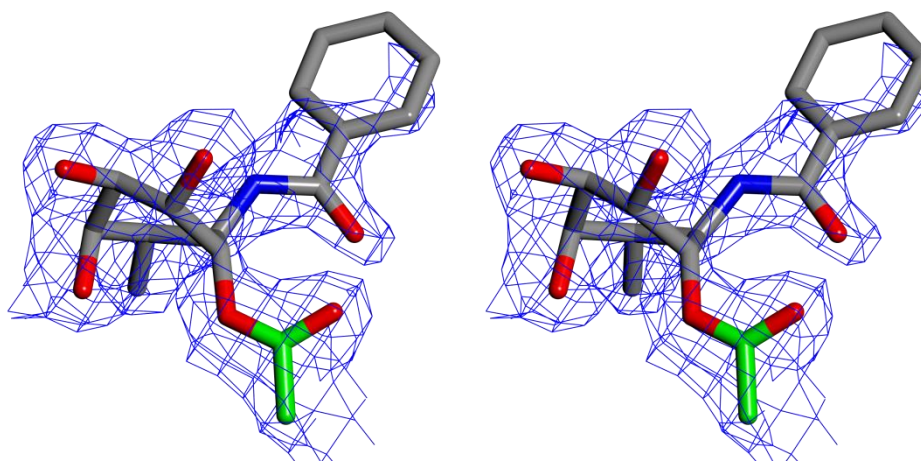
Figure 4.5. The covalent interaction between *Bt*Fuc2970 and **14**, displayed in wall-eye stereo. Atoms of compound **14** are coloured by atom type (carbon in grey). The catalytic nucleophile Asp[229] of *Bt*Fuc2970 is displayed coloured by atom type (carbon in green). Electron density displayed is calculated from likelihood-weighted $2F_o$-F_c from final models, contoured at 1 σ. The figure was created using CCP4MG[183].

The crystal structure of a covalent intermediate bound to *Bt*Fuc2970 was previously determined[184]. This result, PDB entry 2WVS, represents the irreversible binding of the difluoro fuco-configured inhibitor 2-fluoro-fucosylfluoride to the enzyme, and as such a direct comparison of the interactions formed by these two inhibitors is informative. In general, the active sites of these two structures have high homology, however there are a few differences which will be discussed.
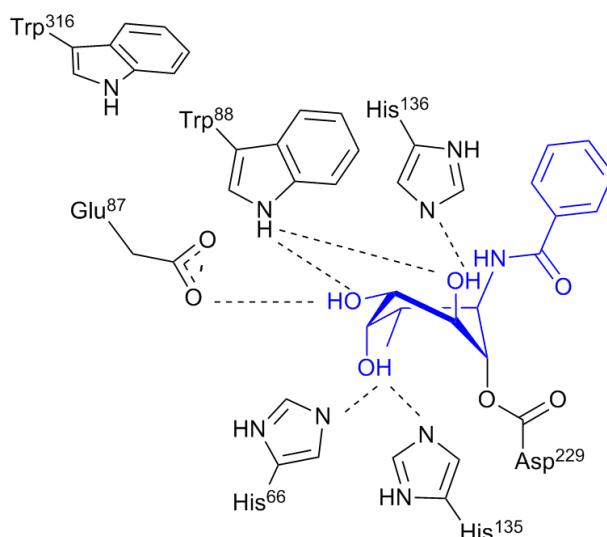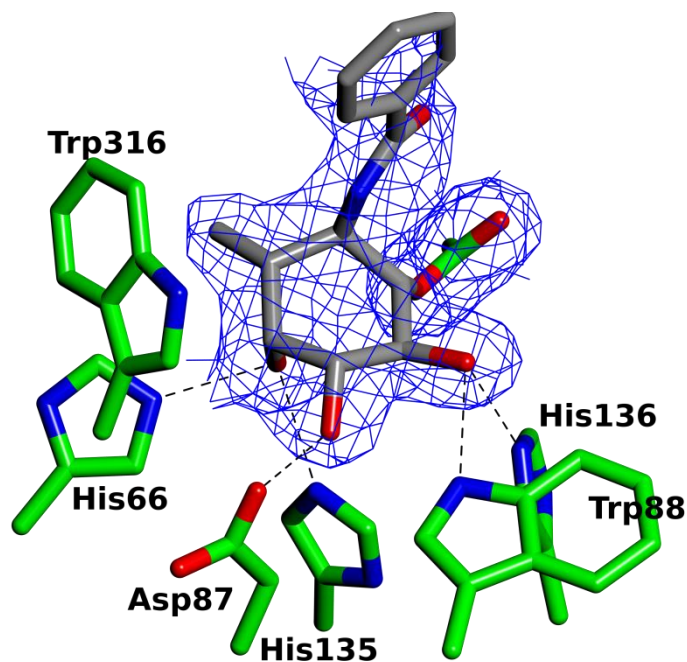
Figure 4.6. Interactions made between **14** and *Bt*Fuc2970. **A**: 3D representation drawn using CCP4MG[183]. Hydrogen bonds < 3.2 Å shown as dashed lines. Residue Trp[316] forms a *ca.* 3.7 Å hydrophobic interaction with **14**. **B**: Schematic representation drawn using CHEMDRAW. Protonation states displayed are arbitrary. Hydrogen bonds < 3.2 Å shown as dashed lines. Residue Trp[316] forms a *ca.* 3.7 Å hydrophobic interaction with **14**.

A minor shift in atomic coordinates of residue Trp[88] is observed between the two structures, likely causing weakening or loss of a hydrogen bonding interaction formed between this residue and the hydroxyl at the C2 position of the inhibitors (3.32 Å in 4WSK *vs.* 3.08 Å in 2WVS). A larger (1.3 Å) displacement is seen for residue Trp[232]. Residue Arg[262], which hydrogen bonds with both the catalytic

nucleophile and acid/base residue in 2WVS, is displaced in 4WSK, appearing in a conformation whereby hydrogen bonding occurs between this residue and the catalytic acid/base Glu[288] rather than the nucleophile residue Asp[229] (Figure 4.7). Residue 288 has the most striking conformational difference; in 4WSK, the enzymatic acid/base residue Glu[288] is positioned in an inactive conformation, located 6.8 Å away from the anomeric carbon of **14**, which is too far for Glu[288] to act as a proton donor to a pendant O at this position, Figure 4.7. In 2WVS, however, Gln[288] resides in what would be an active conformation were the residue not mutated (the shortest C1-Gln[288] distance is 5.5 Å).

As both *Bt*Fuc2970-**14** (4WSK) and 2WVS are covalent enzyme intermediates of the same enzyme, the inactive conformation for the catalytic acid/base observed in *Bt*Fuc2970-**14** is likely to be caused by the extra moieties present in **14** compared to 2-fluoro-fucosylfluoride. Upon ring-opening of the aziridine of **14**, its 'aglycon' moiety is attached to the position which would equate to that of the endocyclic oxygen of a natural substrate. It is thus postulated that *Bt*Fuc2970 does not tolerate appendages to this position well due to sterics. Indeed, the aglycon amide oxygen forms a close contact (*ca.* 3.4 Å) with the carbonyl in the enzymatic nucleophile A[229].

Based on the observations that the aglycon of **14** appears poorly ordered in its crystal structure (Figure 4.5) and the inactive conformation observed for the catalytic acid/base of *Bt*Fuc2970 when complexed with **14**, discussion was initiated with Prof. Overkleeft about the possibility of synthesising an ABP scaffold with a less sterically-demanding aglycon, which may have a higher on-rate and thus provide a clearer structure solution. Subsequent to this discussion, compound **15**, lacking the aryl group of **14**, was synthesised and soaked into crystals of *Bt*Fuc2970, yielding a crystal structure of the covalent adduct formed.
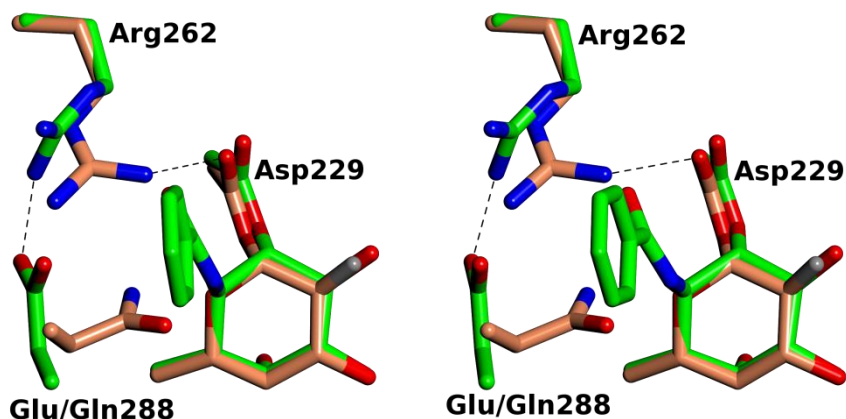
Figure 4.7. The active site of *Bt*Fuc2970 in PDB entries 4WSK (this work, *Bt*Fuc2970-**14**) and 2WVS (*Bt*Fuc2970 E288Q mutant covalently bound to 2-fluoro-fucosylfluoride[184]). Sidechain atoms of residues Asp[229], Arg[262], 288 (Glu for 4WSK, Gln for 2WVS) and ligand atoms are coloured by atom type (carbon atoms from 4WSK in green and from 2WVS in coral, fluorine from 2WVS in grey). Hydrogen bonds between Arg[262] and the nucleophile and residues 229 and 288 displayed as dashed bonds. The figure was drawn using CCP4MG[183].

Diffraction images for this dataset were collected and refined to 1.64 Å resolution, which was the extent of the data usefulness according to a half-dataset correlation[228]. This structure, with a less sterically-demanding aglycon, indeed more clearly showed interaction of the ligand aglycon with *Bt*Fuc2970, and all 4 copies of **15** in the asymmetric unit were observed to have clear electron density for the entirety of the ligand (Figure 4.8), including the covalent bond formed between the enzymatic nucleophile and **15**. The final refined R_{work} and R_{free} values for this complex structure are 0.16 and 0.19 respectively, again an improvement over *Bt*Fuc2970-**14**. As with *Bt*Fuc2970-**14**, the enzymatic acid/base residue Glu[288] is observed in an inactive conformation; this supports the hypothesis that this displacement is due to the presence of a pendant group to the carbon atom at which a natural substrate would have an endocyclic oxygen.
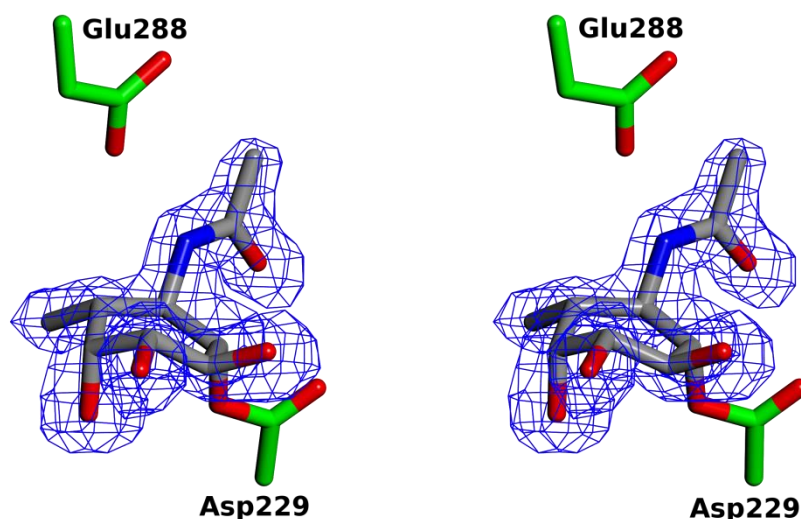
Figure 4.8. Compound **15** lying in the active site of *Bt*Fuc2970, displayed in wall-eye stereo. Atoms of compound **15** are coloured by atom type (carbon in grey). The catalytic nucleophile Asp[229] and acid/base Glu[288] of *Bt*Fuc2970 are displayed coloured by atom type (carbon in green). The map displayed is a likelihood-weighted F_o-F_c map from before the incorporation of phases from compound **15** in refinement, contoured at 3 σ (**15**). The figure was created using CCP4MG[183].

The conformations adopted by **14** and **15** in their crystal structures (4WSK and 4WSJ) were kindly determined by Dr. Carme Rovira at the University of Barcelona and plotted on the conformational free energy landscape of *Bt*Fuc2970 (Figure 4.9). **14** adopts a $^3S_1$ conformation, while **15** adopts a $^3H_4$ conformation; this is, interestingly, the conformation postulated for the catalytic transition state based upon the conformational trajectory previously shown for this enzyme[184]. The X-ray data quality used to generate 4WSJ (*Bt*Fuc2970-**15**) are better than that for 4WSK (*Bt*Fuc2970-**14**). In particular, the atomic coordinates of the C1 atom of **14** are uncertain due to poor electron density in this region; the conformation adopted by **14** is 4WSK is thus more influenced by mathematical refinement and geometric restraints than experimental observations. The electron density for **15**, however, is much clearer and thus it is concluded that both of these ABPs bind to this enzyme in a conformation more resembling that of the enzymatic transition state, however it is uncertain what the reason for this is.
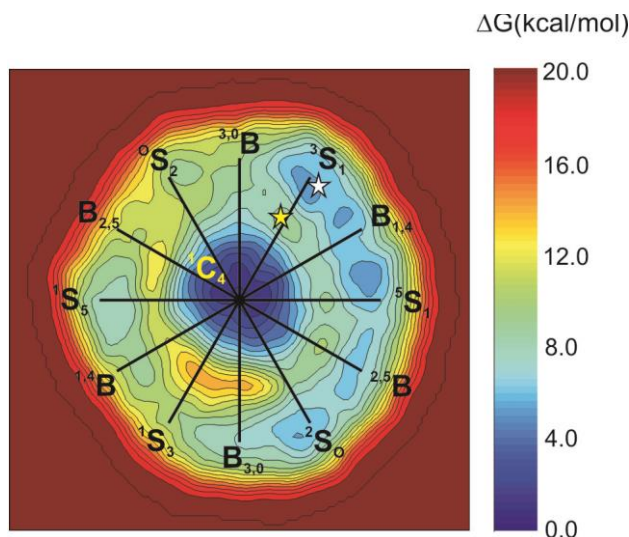
Figure 4.9. Conformations adopted by ligands in 4WSK and 4WSJ. Conformations are plotted as filled stars (Yellow: **15** from 4WSJ, White: **14** from 4WSK) on the calculated free energy landscape of *Bt*Fuc2970. Figure prepared by Carme Rovira at the University of Barcelona.

Catalytic nucleophile (D229N) and acid/base mutants (E288A) of *Bt*Fuc2970 were prepared in order to determine the crystal structures of these mutants complexed with ABPs. Catalytic nucleophile mutants are often used to observe a ligand complex after docking with the enzyme but before reactivity occurs *i.e.* the Michaelis complex. Acid/base mutants, on the other hand, reduce the rates of both the glycosylation and deglycosylation steps of the prototypical retaining glycoside hydrolase mechanism, hence allowing observation of a covalent intermediate when used in conjunction with reactive compounds with a good leaving group (these approaches were previously applied to *Bt*Fuc2970[184]).

Unfortunately, after soaking of both *Bt*Fuc2970 nucleophile and acid/base mutants with **15,** diffraction data representing complex formation could not be obtained; a high-resolution *apo* structure of D229N *Bt*Fuc2970 was however obtained after soaking of the enzyme with **15** for *ca.* 24 h.

4.4 Conclusions

We have crystallographically observed a covalent adduct between the catalytic nucleophile of a GH29 α-ʟ-fucosidase and ʟ-fucose-configured aziridine containing compounds, which were designed to have utility as activity-based probes. This crystallographic observation is the first documented between a glycoside hydrolase and sugar-configured aziridine containing compound, and serves to prove the postulated mechanism by which these compounds act to selectively target active enzyme.

ABPs able to covalently inhibit and label α-ʟ-fucosidase such as **17-19** have great potential in a variety of therapeutic, diagnostic and proteomics technologies. Compounds **17-18**, incorporating fluorescent probes for visualisation and quanitification of α-ʟ-fucosidase activity, may be useful as, *inter alia*, biomarkers of cancer progression, and for elucidating the field of cell biology, *e.g.* in the field of fertility. These compounds have already been demonstrated to have the ability to label GH29 enzymes both *in vitro* and *in vivo*[292]. Application of **19** to spleen lysates from a healthy individual and a sufferer of Gaucher disease allowed labelling of FUCA1/2 present in the sample (Figure 4.10), with a greater GH29 activity being evident for the Gaucher spleen; the upregulation of degradative enzymes is a known feature of Gaucher spleens[298]. Further, **19** was shown to bind to GH29 from mouse spleen, liver and kidney *in vivo* (Figure 4.10).
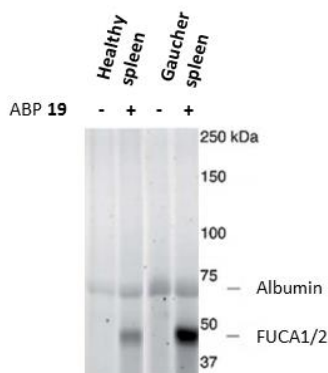
Figure 4.10. Labelling of GH29 α-ʟ-fucosidases by **19**. Adapted from [292].

If fertility events in humans are similar to those in mice, *i.e.* they require the presence of semen membrane-bound GH29 for its sialyl-Lewis[x] moieties to interact with oocyte zona pellucida and membranes, but not activity of the enzyme[102,108], compounds **17-18** may have use as biological probes for the *in vivo* tracking of fertility events.

121

As a number of cancer strains have upregulated α-L-fucosidase activity, there exists potential for chemotherapy using pharmacophores appended to covalent inactivators such as the ABPs discussed in this chapter; this approach may be more effective than that using covalent enzyme inhibitors such as **10-12** in the previous chapter. The biotinylated ABP **19** may be used for the screening of GH29 activity from culture supernatant allowing discovery of new enzymes in this family; irreversible binding of the ABP warhead to GH29 enzymes may be achieved through the application of supernatant over a streptavidin affinity column pre-incubated with **19**, for example. Elution with free biotin would allow the characterisation of novel GH29 enzymes *e.g.* as in the example of Vocadlo and co-workers[273].

As many of the potential purposes of fuco-configured ABPs are therapeutic in nature, understanding of the active site features of human GH29 enzymes is perhaps more pertinent than that of bacterial ones. Since, in the absence of a GH29 crystal structure, it is uncertain to what extent the active site of human GH29 enzymes is similar to that of *Bt*Fuc2970, elucidation of the crystal structure of human GH29 enzymes would be beneficial in allowing further, directed synthetic efforts to improve upon the existing ABP scaffold.

Chapter 5: Expression and purification of human fucosidases FUCA1 and FUCA2

Abstract

Human GH29 family α-ʟ-fucosidases I and II (FUCA1 and FUCA2) are critical to many biological events such as bacterial colonisation of the gastrointestinal tract, the immune response, and fertility. A multitude of inhibitors against this enzyme class has been reported in the literature; the inhibition of these proteins has clinical relevance *e.g.* for treatment of the lysosomal storage disorder fucosidosis and clearance of the stomach pathogen *Helicobacter pylori* from its host. The 3D crystal structures of these enzymes may aid in inhibitor design and potentially further illuminate their biological roles, however no such crystal structures are yet available. Described herein are attempts made towards the expression, purification and crystallisation of FUCA1 and FUCA2. The expression of these proteins in heterologous systems had only limited success as FUCA1 could be expressed and purified to a yield of 3.5 mg enzyme per litre culture volume while FUCA2 purification was unsuccessful. Pure FUCA1 and FUCA2 could however be obtained by expression of genes encoding the enzymes in both cultured HEK-293T and CHO-K1 cells; expression levels of FUCA1 were higher than that of FUCA2, and transient transfection of HEK-293T cells yielded more expressed protein than from CHO-K1 cells. Preliminary crystallisation experiments using a purified FUCA1 construct described herein are reported, along with Michaelis Menten kinetics for the enzyme (K_M = 0.099 ± 0.01 mM, V_{max} = 1305 ± 47 min^{-1}). This work may be built upon for the realisation of the crystal structures of FUCA1 and FUCA2.

5.1 Introduction

The physiological roles of mammalian α-ʟ-fucosidase have been discussed elsewhere (see Chapter 1 and section *1.4.1.2* in particular) and will not be repeated. Briefly, the enzyme has important roles in maintaining homeostasis of a number of biologically critical fucosylated glycans[20,196] which are often differentially regulated in neoplasia[93,197], and its deficiency causes the neurovisceral storage disorder fucosidosis[3]. Further, the enzyme plays an important role in fertility[109]. What follows is a review of the literature which stemmed particularly from the potential to learn more about and treat fucosidosis, namely the discovery, characterisation, and purification of α-ʟ-fucosidase in humans. Both the lysosomal enzyme FUCA1 and the less well-studied secreted α-ʟ-fucosidase FUCA2 will be discussed.

The existence of the α-ʟ-fucosidase enzyme was first documented in 1961 by Levvy and McAllan[299]. Due to the known presence of ʟ-fucose in mucins, blood group antigens[300] and milk oligosaccharides[301] at the time, the authors postulated the existence of ʟ-fucosidase enzyme/s with unknown anomeric specificity. To assess this, the authors first synthesised both *p*-nitrophenyl-α-ʟ-fucopyranoside and *p*-nitrophenyl-β-ʟ-fucopyranoside[302]. Tissue homogenates were collected from a range of mammalian species, gender and tissue types, and their action upon each substrate was probed.[299] While no tissue type released *p*-nitrophenolate from the β-linked substrate, a wide range of tissues released the product from *p*-nitrophenyl-α-ʟ-fucopyranoside thus proving the existence of an α-ʟ-fucosidase enzyme. Further, the authors examined the kinetic profile of the enzymes from rat epididymis and ox liver, and determined their pH optima at pH 6.1 and 5.6, respectively and Michaelis constants of 2.1 mM and 2.2 mM, respectively. A further observation made was the apparent abundance of α-ʟ-fucosidase in epididymal homogenates, hinting at a role for the enzyme in insemination which would be elucidated many years later.

The existence of two 'forms' of α-ʟ-fucosidase (α-ʟ-fucosidases I and II) in mammals was first demonstrated in pig kidney.[303] These forms could be distinguished by their elution profiles on size-exclusion chromatography, thermostability, pH dependence for catalytic activity and their ability to hydrolyse a variety of different fucosylated substrates. Subsequently, two groups contemporaneously showed the existence of these two distinct forms in human liver[304] and a variety of tissue types including the same[305]. Here, both groups demonstrated that the two forms isolated from a single organ could be separated by size-exclusion chromatography, while one group also achieved their separation

using ion-exchange chromatography[304]. The two forms were further characterised by their differing pH stability, thermostability, and substrate specificities. In each case, one form was observed not to bind to Sephadex G-200 and appeared in the void volume of the elution profile. While Robinson and Thorpe concluded that this enzyme was a higher molecular weight species than the other, and may be an aggregated form of it[304], the other authors postulated that the two forms are of similar molecular weights, and are only separated on Sephadex G-200 due to non-specific adsorption of one form onto Sephadex[305]. Numerous isoforms of human α-L-fucosidase have since been observed by isoelectric focusing of the enzyme from a variety of human tissue types e.g.[306-308].

Alhadeff et al. investigated the comparative α-L-fucosidase activities between normal and fucosidotic livers.[306] Activity assays were carried out using novel substrate 4-methylumbelliferyl-fuco-L-pyranoside. While the pH optima determined for healthy and fucosidotic liver were essentially identical (5.4 and 5.3 respectively), Michaelis constants were marginally different (85 µM and 58 µM respectively) and the fucosidotic liver was observed to have only 4% α-L-fucosidase activity compared to the enzyme from healthy liver. The authors further observed 6 isoforms of the enzyme from healthy liver using isoelectric focusing, with isoelectric points between 5.5 and 6.9. Sialic acid residues in the enzyme were cleaved by neuraminidase treatment and the α-L-fucosidase activity of the two most acidic isoforms of the enzyme was observed to be reduced, suggesting sialic acid residues are present in at least some isoforms of the enzyme and are important for its activity. The loss of activity at acidic pH after neuraminidase treatment of the enzyme was also observed by other, independent researchers.[309,310] A subsequent study showed that treatment of the more neutral isoforms of human liver α-L-fucosidase with sialyltransferase could regenerate the acidic isoforms of the enzyme, confirming the role of sialic acid residues in its activity at acidic pH.[311]

The effect of deglycosylation of the enzyme has been further studied. By treating human liver α-L-fucosidase with N-glycanase, Piesecki and Alhadeff showed that deglycosylation of the enzyme caused negligible difference in its recognition or activity against p-nitrophenyl-α-L-fucopyranoside, 2'-fucosyllactose, Lacto-N-fucopentaose II or the blood group A trisaccharide.[312] The activity profile of the enzyme with respect to pH was altered upon deglycosylation, however, with a shift towards a more neutral pH optimum and loss of enzymatic activity in the acidic regions. Further, deglycosylation of the enzyme did not appear to affect the gross conformation of the enzyme as observed by circular dichroism. Thus, glycosylation

of the enzyme may be critical in allowing its function in the acidic environment of the lysosome but appears to have limited, if any, other effects.

5.1.1 Purification of α-L-fucosidase

α-L-fucosidase was first purified to apparent homogeneity in 1975[313]; previous attempts to purify α-L-fucosidase from various sources had not removed the activity of other glycosidases[305,314]. α-L-fucosidase was purified from homogenates of five human livers using a two-step chromatography procedure utilising agarose-ε-aminocaproyl-fucosamine. By capturing the enzyme by affinity with a moiety resembling its substrate, purification was effected with high yield (66% by activity) and the purified protein was free from contaminating glycoside hydrolase activity. Electrophoresis of the purified enzyme using SDS-PAGE revealed a single protein band with molecular weight ca. 50 kDa. Analytical size exclusion chromatography and sedimentation equilibrium ultracentrifugation were also carried out on the enzyme and yielded molecular weights of *ca*. 175 kDa and *ca.* 230 kDa, respectively. Kinetic parameters of the purified enzyme's hydrolysis of 4-methylumbelliferyl-α-L-fucopyranoside and *p*-nitrophenyl-α-L-fucopyranoside substrates yielded K_M and V_{max} values of 0.22 mM, 14.1 µmol mg^{-1} and 0.43 mM, 19.6 µmol mg^{-1} respectively.

In 1978, the same author reported the purification of the enzyme from human serum using a similar method.[315] Isoelectric focusing of the enzyme revealed seven isoforms for the enzyme, but the predominant band was the most acidic with an isoelectric point around 5. Interestingly, treatment of the enzyme with neuraminidase (a glycoside hydrolase acting in an *exo*-fashion on sialic acid) drastically altered this pattern, resulting in isoforms with isoelectric points around pH 6-7. The oligomeric state of the enzyme was also greatly influenced by neuraminidase treatment; size exclusion chromatography revealed what was likely an oligomer of molecular weight 296 kDa, which reduced significantly to *ca*. 37 kDa on treatment of the enzyme with neuraminidase. Thus, sialic acid was concluded to be incorporated into the enzyme and to alter its oligomeric state. In this study, the authors ran SDS-PAGE gels of both serum and liver enzymes; two bands were observed for each enzyme, however these were of different molecular weights (56 and 54 kDa for the serum enzyme and 53 and 49 kDa for the liver enzyme). The authors concluded that post-translational modifications of serum α-L-fucosidase, such as sialylation, may be responsible for its properties, which are

126

different to those of fucosidases from liver and brain characterised by the same group.[316]

The complete FUCA1 open reading frame was sequenced from chromosomal DNA from human liver, placenta and colon in 1989[317]. The mature protein comprises 439 residues after cleavage of a 22 residue long hydrophobic signal peptide. The gene transcript contains three potential glycosylation sites based on the presence of three Asn-X-(Ser/Thr) consensus sequences.

Sequencing of the FUCA1 gene and isolation of cDNA clones was first carried out in 2003, allowing recombinant expression of the enzyme.[318] The enzyme was inserted from cDNA[317] into the *Bam*H1 restriction endonuclease site of expression vector pGEX-2T to yield a gene encoding FUCA1 as a glutathione *S*-transferase fusion protein[318]. The resulting gene was expressed in *Escherichia coli* BL21 cells. Most of the protein produced was located in the pellet fraction of the cells after lysis by sonication, indicating that it was sequestered in cellular inclusion bodies. The insoluble protein was, however, purified with high yield (65 mg purified protein from 1 L culture media), and showed a broad activity profile against substrate 4-methylumbelliferyl-α-ʟ-fucopyranoside. The pH profile of the enzyme was however dissimilar to FUCA1 in the neutral region, lacking an activity maximum between pH 6.0 and 7.0 that is seen for human fucosidases from many sources[313,315,316]. The authors explained this observation by the inability of *Escherichia coli* to glycosylate the protein.

In 2009, Liu and co-workers successfully expressed soluble FUCA1 using the *Escherichia coli* BL21(DE3) cell strain.[319] The authors inserted the *fuca1* sequence, less the first 22 amino acids which form a signal peptide, into the pET22b(+) vector. Cells bearing the plasmid were incubated in LB media and gene expression was effected by the addition of 1 mM IPTG; expression was carried out for 8 hours at 37 °C before harvesting and resuspension in 20 mM sodium acetate buffer pH 5.5. Purification of the enzyme was carried out by ammonium sulfate precipitation, followed by cation exchange chromatography, anion exchange chromatography, and finally size exclusion chromatography. It was observed that growth in LB media with pH 6.0-6.5 was essential for activity, as when pH was above 7.0 the protein expressed showed no activity but maintained solubility.

This study further investigated the catalytic nucleophile and acid/base residues of FUCA1. At this time, the only α-ʟ-fucosidase with a reported X-ray crystal structure was that from *Thermotoga maritima* (*Tm*GH29)[182]. The authors used multiple sequence alignment with FUCA1, *Tm*GH29 and a number of GH29

enzymes from other organisms to identify a conserved carboxylate corresponding to the catalytic nucleophile of TmGH29. No such conservation was observed for the catalytic acid/base of TmGH29, however. Liu et al. created mutant enzymes for each Asp or Glu residue of FUCA1 that appears to be conserved between GH29 enzymes and performed activity assays on the mutant enzymes using a range of synthetic substrates with varying pK_a values in order to generate Brønsted plots[319]. Only mutations to residues D225 and E289 of FUCA1 had any noticeable effect on k_{cat} or K_m values. The relative reaction rates of hydrolysis for the E289G mutant on substrates with a range of pK_a values revealed a change in mechanistic rate determining step when the pK_a of the substrate approached the solvent pH, from defucosylation to fucosylation, consistent with the absence of acid/base assistance.

Chemical rescue with sodium azide was further used in order to investigate the roles of D225 and E289 in catalysis. When enough space is available in a mutant glycoside hydrolase active site, $e.g.$ after mutation of enzymatic nucleophile or acid/base residue to glycine, azide is able to 'rescue' the activity of these enzymes as it can act as a nucleophile or general base (Figure 5.1).[180,320] The addition of sodium azide to E289G and D225G was observed to increase the rate of p-nitrophenyl-α-L-fucopyranoside hydrolysis by these enzymes 30 and 6-fold, respectively. Further, the authors observed α-fucosyl azide and β-fucosyl azide as reaction products on rescue of E289G and D225G respectively by nuclear magnetic resonance spectroscopy, providing more evidence for the assignment of D225 as the catalytic nucleophile and E289 as the catalytic acid/base residue.
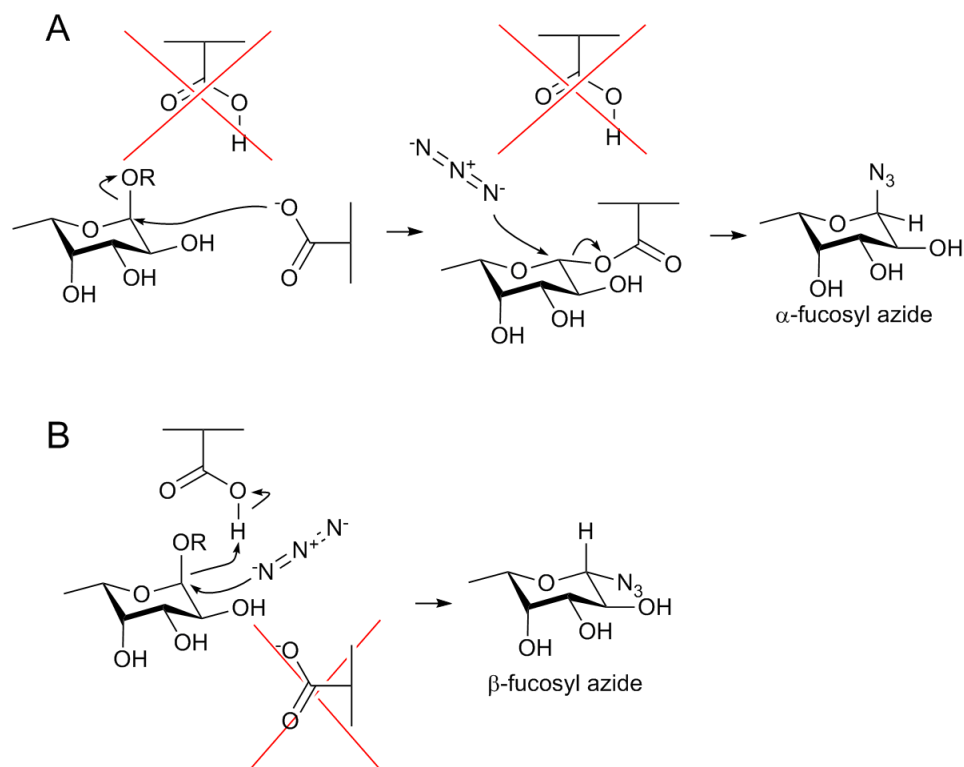
Figure 5.1. Chemical rescue of retaining α-ʟ-fucosidase activity by azide. Rescue of deficient acid/base activity leads to formation of an α-fucosyl product (**A**) while rescue of deficient nucleophile activity leads to formation of a β-fucosyl product (**B**).

While the assignment of D225 as the catalytic nucleophile of this enzyme is agreed within the field, there is contention over the assignment of E289 as the catalytic acid/base. In 2010, Bueren *et al.* noted that "on the basis of sequence alone, neither the catalytic acid observed structurally nor the human residue proposed through kinetic analysis of variants is conserved within the family".[184] From the structural point of view, residue D275 of FUCA1 appears more likely to be the catalytic acid/base as it lays more close to the active site observed in the crystal structure of every other GH29 enzyme currently available (Figure 5.2), although when this residue was mutated to glycine by Liu *et al.*, no difference in enzyme activity was observed[319]. Further, the position corresponding to that of residue E289 in a homology model of FUCA1 lies buried in the protein, not at all proximal to the enzymatic active site observed in other GH29 structures. While conformational flexibility has now been observed for many enzymes in the GH29 family including those from *Thermotoga maritima*[221] and the eukaryote *Fusarium graminearum*[188], such a large conformational change as that which would be envisioned to be

required for displacement of residue E289 into a catalytically active position is unprecedented in protein science.
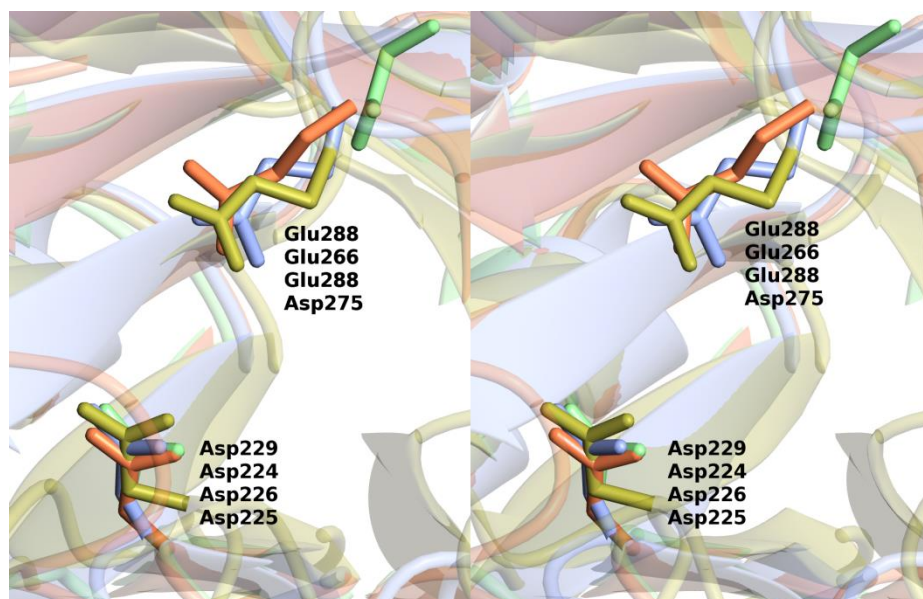
Figure 5.2. The assignment of the catalytic acid/base residue of human α-ʟ-fucosidase is still under contention. The protein chains of crystallographically observed α-ʟ-fucosidases from *Bacteroides thetaiotaomicron* (*Bt*Fuc2970, from PDB entry 4JL2), *Thermotoga maritima* (*Tm*GH29, from PDB entry 1OUD) and *Fusarium graminearum* (*Fg*GH29, from PDB entry 4NI3) along with that from a homology model of FUCA1 prepared using the Phyre2 server[321] are displayed as transparent ribbons coloured through the chain in ice blue (*Bt*Fuc2970), gold (*Fg*GH29), coral (*Tm*GH29) and light green (FUCA1). Further, the side-chains of the catalytic residues of *Bt*Fuc2970 (Asp[229] and Glu[288]), *Tm*GH29 (Asp[224] and Glu[266]) and *Fg*GH29 (Asp[226] and Glu[288]) are displayed as cylinders in the same colours at their respective protein chains, and the sidechains of the residues Asp[225] and Asp[275] from a homology model of FUCA1 are displayed in light green; these residues are further annotated on the figure in the order described above. The figure was drawn in wall-eye stereo using CCP4MG[183].

Contrary to FUCA1, there is a dearth of research dedicated to the secreted α-ʟ-fucosidase FUCA2.[322] The enzyme was first discovered in 1984[323], and its gene was subsequently assigned to chromosome 6[324]. In 2009, a role for the enzyme in cleavage of fucosylated glycan structures as a response to the invasion of stomach pathogen *Helicobacter pylori* was reported.[4] Expression of the enzyme in the neoplastic Capan 1 cell line was shown to be effected by the presence of

Helicobacter pylori. Further, knockdown studies showed that FUCA2 expression allowed the pathogen to incorporate exogenous fucose into its cell surface glycans, forming host-type Le[x] antigens. During this study, the authors successfully expressed and purified FUCA2 inserted into expression vector pCMV-Tag2B (containing a C-terminal FLAG tag) in 293T cells; purification of the enzyme was achieved through application of lysed cell supernatant over anti-FLAG antibody affinity beads. The colonisation of many pathogenic bacteria involves the interaction between fucosylated host endothelial glycans such as Le[x] and lectins expressed on the cell surface of the colonist.[5,6] This enzyme may thus be evolved to provide an innate defence mechanism against the invasion of pathogens which produce these host-like glycans, as host release of L-fucose may allow clearance of pathogens by out-competing their interaction with endothelial fucosylated glycans.

5.1.2   Managing the glycosylation status of glycoproteins

The decoration of proteins with N-linked glycans poses a significant hurdle in structural biological studies.[325] This is because, while the glycosylation of native glycoproteins is often crucial to their correct folding, the heterogenous glycosylation trees present in mammals make crystallisation difficult; as such, simplification of glycan trees in a protein sample is often necessary for the successful crystallogenesis of glycoproteins. This may be effected in three ways; through protein expression in cell lines that do not express critical carbohydrate processing enzymes involved in the N-glycosylation pathway, by inhibiting the activity of these enzymes during transient expression, or by enzymatic deglycosylation of the enzyme post-expression.[325]

Glycosylation deficient strains of both CHO[326,327] and HEK[328,329] cell lines have been discovered.[327] Many of these cell lines seem to have evolved by mutation offering an evolutionary advantage against plant lectin induced stress. While the currently available glycosylation deficient CHO cell lines are known to have mutations in a wide range of carbohydrate processing enzymes[327], the two glycosylation deficient HEK cell lines known to the author have deficiency in the N-acetylglucosaminyl transferase I[328] and α-mannosidase II[329] genes. The carbohydrate processing enzyme inhibitors kifunensine[330] and swainsonine[331] may also be used for simplification of N-linked glycosylation in proteins grown in tissue culture, when added to culture media these inhibitors inhibit cellular α-mannosidases I and II respectively and thereby prematurely halt N-glycosylation before the generation of complex glycans.[325]

Complex glycans pose a particular problem in the expression of glycoproteins destined for downstream crystallographic studies.[325] Complex glycans do not act as substrates for the endoglycosidase EndoH, and glycosylation trees cleaved by the enzyme retain their most interior GlcNAc moiety. This residue is known to be more important in mediating interactions with the hydrophobic region of the protein to which it is bound than the more exterior monosaccharides[332], and while proteins cleaved by EndoH to yield glycan trees containing just this sugar tend to have similar polydispersities to those containing the full length tree, loss of this residue has been shown to promote protein aggregation[325,333]. The premature termination of the N-glycosylation pathway before formation of complex glycans is thus extremely useful for the production of glycoproteins for crystallisation studies; these glycoproteins can be deglycosylated with EndoH rather than PNGase, increasing their propensity to crystallise.

A plethora of studies, including ones presented in this thesis, investigate the inhibition of GH29 fucosidases in order to harness the potential medical benefits that potent inhibitors of fucosidase may have, particularly for treatment of fucosidosis[3] and *Helicobacter pylori* infection[4]. Structural insights may play a large role in the design of potent fucosidase inhibitors however the crystal structures of both human fucosidases are currently unresolved.

Herein, efforts made towards the gene expression and purification of FUCA1 and FUCA2 are described, to finally shed light on the structural features of these important enzymes. Initially, expression tests using the heterologous *Escherichia coli* and *Pichia pastoris* expression systems will be described, followed by those using cultured mammalian cells as an expression system.

5.2 Materials and methods

5.2.1 Heterologous expression of FUCA1 and FUCA2

The genes encoding fucosidases FUCA1 and FUCA2 were codon optimised for expression in *Escherichia coli*, and were synthesised and inserted into pET-30a$^+$ vectors between *NcoI* and *XhoI* restriction endonuclease sites by GenScript (Appendix 3); these samples were available in the lab from previous work conducted by Dr. Alicia Lammerts van Bueren. These plasmids encode FUCA1 and FUCA2 with an N-terminal His$_6$ affinity tag and S-tag. Aliquots of these plasmids were transformed into chemically competent BL21(DE3), pLEMO and C41 cells, which were grown in LB media containing 30 mg L^{-1} kanamycin as small scale (10 mL) cultures (cells were incubated at 37 °C overnight with aeration by shaking at 180 rpm).

FUCA1 and FUCA2 constructs were also sub-cloned into both the pETite® NHis SUMO (EXPRESSO T7 SUMO cloning and Expression System, Lucigen) and pET-22b$^+$ vectors. Oligonucleotide primers were designed, and were synthesised by Eurofins MWG Operon.

For sub-cloning into the pETite® NHis SUMO vector, Phusion® polymerase (New England Biolabs) was used to amplify gene products from FUCA1 and FUCA2 templates in the pET-30a$^+$ vector by polymerase chain reaction. After treatment with *DpnI* to remove template DNA (*ca.* 1 hour at room temperature), the amplified reaction product was used to transform Hi-Control 10G chemically competent cells (EXPRESSO T7 SUMO cloning and Expression System, Lucigen) according to manufacturer's instructions. Diluted transformed cells were plated on LB Agar plates containing 30 mg L^{-1} Kanamycin. A number of colonies that grew were sequenced (GATC Biotech) and a colony containing the non-mutated insert was used to grow a 10 mL LB overnight culture containing 30 mg L^{-1} Kanamycin. Plasmid DNA was extracted and purified using a QIAprep Spin Miniprep Kit (Qiagen) and used to transform chemically competent BL21 (DE3) cells. These cells were plated on LB Agar plates containing 30 mg L^{-1} Kanamycin and a single colony was used for all subsequent experiments.

For sub-cloning into the pET-22b$^+$ vector, Phusion® polymerase (New England Biolabs) was used to amplify gene products from FUCA1 and FUCA2 templates in the pET-30a$^+$ vector by polymerase chain reaction. The amplified DNA obtained and pET-22b$^+$ vector were treated with the *NcoI* and *XhoI* restriction endonucleases separately to yield DNA segments with complementary sticky ends. Cleaved pET-

22b$^+$ vector was then treated with Antarctic phosphatase (37 °C, 2 hours) in order to prevent re-ligation of the vector with itself. The enzyme mix in the solution containing the treated vector was then heat inactivated at 80 °C for 20 minutes. Reaction products for both the insert and vector were purified using a PCR Purification kit (Qiagen) and were ligated using T4 ligase; a ratio of 1 part vector to 3 parts insert (by molarity) was used for the ligation reactions which were incubated at 16 °C overnight. The product of this reaction was then used to transform electrocompetent XL1 Blue cells (Agilent) and diluted transformed cells were plated on LB Agar plates containing 50 mg L^{-1} Ampicillin. A number of colonies that grew were sequenced (GATC Biotech) and a colony containing the non-mutated insert was used to grow a 10 mL LB overnight culture containing 50 mg L^{-1} Ampicillin. Plasmid DNA was extracted and purified using a QIAprep Spin Miniprep Kit (Qiagen) and used to transform chemically competent BL21 (DE3) cells. These cells were plated on LB Agar plates containing 50 mg L^{-1} Ampicillin and a single colony was used for all subsequent experiments.

5.2.1.1   Gene expression and protein purification

For small scale expression tests, 1 mL samples of 10 mL overnight cultures in LB media taken before induction with isopropyl-β-D-1-thiogalactopyranoside (IPTG) and after expression (at 37 °C for *ca.* 4 hours or at 16 °C overnight) were lysed to determine expression profiles. Samples were centrifuged and the media supernatant was removed. Subsequently lysis was performed chemically by the addition of 200 µL BugBuster® Protein Extraction Reagent (Novagen) over 10 minutes. Soluble fractions were obtained by centrifuging and removing the supernatant from thus treated samples, whilst insoluble fractions were obtained by the treatment of the resulting cell pellet with 50 µL 4% SDS over 10 minutes. Expression test samples were run on SDS-PAGE gels to determine recombinant gene expression and solubility profiles.

For large scale expression runs, 10 mL overnight cultures in autoclaved LB media containing either 30 mg L^{-1} Kanamycin or 50 mg L^{-1} Ampicillin, depending on the resistance gene present in the plasmid, were used to innoculate a larger (1 L) volume of autoclaved LB media containing the same concentration of antibiotic. Expression was initiated by the addition of 1 mg L^{-1} IPTG when the culture had reached an A$_{600}$ in the range 0.6 - 1.0. Cells were incubated at 16 °C overnight before harvesting by centrifugation. The buffers these samples were re-suspended in were comprised HEPES buffer (10-50 mM), 150-500 mM NaCl and 20 mM

imidazole (pH 7-8). Resuspended cell pellets were stored at -20 °C until purification was attempted.

For protein purification, cell pellets were thawed at room temperature and lysed by sonication. The resulting solutions were centrifuged at 4416 RCF for 30 m and the supernatant decanted. This supernatant was applied to a 5 mL HiTrap column (GE Healthcare) preloaded with Ni^{2+} connected to an Äkta FPLC (GE Healthcare). Low imidazole buffers (as above) were passed through the column until the A_{280} response returned to the baseline. After this, an increasing concentration of imidazole was applied to the column as a gradient (to 500 mM imidazole over 50 mL). Fractions with a strong A_{280} response were run on a SDS-PAGE gel to determine the molecular weight and purity of the protein species present in the eluate. For expression of FUCA1 in the pET-30a+ vector, the fractions deemed to have the highest purity were concentrated by centrifugation at 5000 RCF using an Amicon Ultra-15 10K centrifugal filter. This protein was then applied to a Hi-Load 16/60 Superdex 200 prep grade size exclusion column (GE Healthcare) connected to an Äkta FPLC. A solution comprising 50 mM HEPES buffer, 150 mM NaCl (pH 8.0) was run through the column and samples were collected from the eluate.

FUCA1 expression was also attempted using *Pichia pastoris* as an expression vector, by Dr. Jared Cartwright (Bioscience Technology Facility, University of York). FUCA1 template in pET-30a$^+$ vector was handed over to Dr. Jared Cartwright, who designed constructs containing a *Saccharomyces cerevisiae* prepro-alpha-factor signal peptide for secretion into the media and N-terminal His_6 tag for easy screening of expression by Western blotting, cloned the constructs and tested their expression levels on a small and large scale.

5.2.2 Expression of FUCA1 and FUCA2 in mammalian tissue culture

5.2.2.1 General Considerations

Tissue culture consumables were purchased as sterile packaged goods; Serological pipettes from Sarstedt, flasks from Corning, roller bottles from Greiner and culture media from Gibco.

All tissue culture experiments were conducted in a laminar flow hood with an airflow of *ca*. 0.75 m s^{-1}. At least 10 minutes UV irradiation was applied both prior to and after tissue culture experiments. The flow hood was wiped with 70% ethanol both prior to and after tissue culture experiments, and additionally with anti-bacterial spray after experiments. All items entering the hood were sprayed and

wiped with 70% ethanol before entering the hood. Tissue culture flasks were incubated at 37 °C in a 5% CO_2 atmosphere. Roller bottles were rolled at *ca.* 0.5 RPM.

Adherent HEK-293T and CHO-K1 cell lines originally purchased from Sigma Aldrich and subsequently maintained in-house were used for all mammalian cell culture experiments. Stock T75 flasks containing HEK-293T and CHO-K1 cells were maintained throughout the course of tissue culture experiments and were passaged regularly every 3 or 4 days. HEK-293T cells were cultured in DMEM media supplemented with 10% FBS while CHO-K1 cells were cultured in 1:1 DMEM/Ham's F10 media supplemented with 10% foetal bovine serum (FBS). Passaging was conducted in the following manner. Media was aspirated and the flask washed with 10 mL phosphate-buffered saline (PBS) which was then aspirated. Cells were released from the flask surface by the addition of 2 mL Trypsin/EDTA solution (Gibco) and administration of slight percussive force. Cells were diluted with 8 mL media containing 10% FBS and centrifuged at 200 RCF for 5 minutes. Subsequently, supernatant was decanted and cells were re-suspended in 10 mL media containing 10% FBS. Typically a 1 in 10 dilution of cells was effected into a fresh T75 flask into which 9 mL media supplemented with 10% FBS and 1 mL re-suspended cells was added. The remaining 9 mL of re-suspended cells were discarded.

5.2.2.2   *Cell expansion into larger culture volume*

Expansion of cells into larger culture volume was performed at the same time points as cell passage. The remaining 9 mL of re-suspended cells were transferred to T175 flasks containing 30 mL of media supplemented with 10% FBS, typically at a 1 in 10 or 20 dilution.

5.2.2.3   *FUCA1 and FUCA2 genes and plasmid amplification*

Constructs for the expression of FUCA1 and FUCA2 in mammalian cell culture were obtained through the Glyco-Enzyme Repository (http://glycoenzymes.ccrc.uga.edu/). Five constructs for FUCA1 and four constructs for FUCA2 were received (Figure 5.3). Constructs were received as glycerol stocks of live cultures of *Escherichia coli* DH5 Alpha cell lines containing the constructs in Figure 5.3 cloned into a modified mammalian Gateway DEST vector (for vector maps and sequences see the web-server at

http://glycoenzymes.ccrc.uga.edu/). While all constructs were available for FUCA1, the pGEC2 construct was not available for FUCA2.

pGEN1

NH_2–[TCM]–[His$_8$]–[Strep-II]–[TEV]–[FUCA1/2]–CO_2H

pGEN2

NH_2–[TCM]–[His$_8$]–[Avi]–[GFP]–[TEV]–[FUCA1/2]–CO_2H

pGEN3

NH_2–[TCM]–[His$_8$]–[Avi]–[GFP]–[Fc]–[TEV]–[FUCA1/2]–CO_2H

pGEC1

NH_2–[TCM]–[FUCA1/2]–[TEV]–[His$_8$]–[Strep-II]–CO_2H

pGEC2

NH_2–[TCM]–[FUCA1/2]–[TEV]–[GFP]–[Avi]–[His$_8$]–CO_2H

Legend
TCM   *Trypanosoma cruzi* α-mannosidase signal sequence
Strep-II Strep-Tag-II®
TEV   TEV protease cleavage site
Avi   AviTag™
GFP   Green fluorescent protein coding region
Fc   Antibody Fc fragment

Figure 5.3. FUCA1 and FUCA2 constructs obtained from the Glyco-Enzyme Repository.

Glycerol stocks of FUCA1 and FUCA2 constructs in the *E. coli* DH5 Alpha cell strain were used to inoculate 10 mL LB media containing 50 mg μL^{-1} Ampicillin and incubated at 37 °C with aeration at 180 rpm overnight. These cultures were subsequently used to inoculate 200 mL LB media containing 50 mg μL^{-1} Ampicillin which was incubated at 37 °C with aeration at 180 rpm overnight. Plasmid DNA was purified using a Macherey-Nagel NucleoBond® Xtra Midi EF Midiprep kit as per manufacturer's instructions.

FUCA1 and FUCA2 constructs containing minimal His$_8$ tags upstream of the gene encoding the α-L-fucosidase were designed. Three constructs were prepared, containing minimal, non-cleavable His-Tags either directly upstream of the FUCA1 coding sequence (8His_FUCA1), or containing one (8HisSG_FUCA1) or two (8HisSGSG_FUCA1) SG dipeptides before start of the FUCA1 coding sequence. Oligonucleotide primers were used to remove the existing N-terminal sequence, introduce the new sequence and amplify the FUCA1 gene from FUCA1 pGEN1 by polymerase chain reaction to yield linear plasmid (Appendix 4); polymerase chain

reaction was effected by using Q5 polymerase (New England Biolabs) as per the manufacturer's instructions, using the thermocycling conditions below (Table 5.1).

Table 5.1. Thermocycling conditions for generation of linear plasmid for FUCA1 constructs with minimal octahidstidine affinity tags.

Temperature / °C	Duration	Number of cycles
98 °C	30 s	1
98 °C	10 s	
67 °C	30 s	25
72 °C	3 m	
72 °C	2 m	1
4 °C	HOLD	1

Amplified products were treated with the KLD enzyme mix from a Q5 site directed mutagenesis kit (New England Biolabs) as per the manufacturer's instructions. The product of this reaction was used to transform NEB 5-Alpha cells (New England Biolabs). Transformed cells were applied neat and at 1 in 10 dilution with LB to LB Agar Ampicillin plates which were grown overnight at 37 °C. Colonies were picked from the plates after incubation overnight and grown in LB media containing 50 mg L^{-1} ampicillin as a small scale (10 mL) culture (cells were incubated at 37 °C overnight with aeration by shaking at 180 rpm).

DNA sequences of the products were checked using the CMV forward promoter primer (CGCAAATGGGCGGTAGGCGTG) and a reverse promoter specific to the C-terminus of the pGEN and pGEC vectors (TGACAACGGGCCACAACTCCTC). Glycerol stocks derived from media inoculated by a single colony of each of the three constructs were used for all subsequent experiments; DNA used for transfection testing was isolated using an identical method to FUCA1 pGEN1, *vide supra*.

5.2.2.4 *Transfection of adherent mammalian cell lines*

Transfections were performed using either linear polyethylenimine or polyethylenimine "Max" as transfection reagents. The quantity of plasmid DNA used for transfection was standardised to flask surface area with 0.28 µg DNA used per cm^2 flask surface area. Ratios of transfection reagent to plasmid DNA quoted, *vide infra*, are described by the relationship x mL of a 1 mg mL^{-1} stock of transfection reagent per µg plasmid DNA required for transfection where the ratio is x:1.

For transfections, the relevant volume of transfection reagent was added to a quantity of serum free DMEM and incubated for *ca.* 10-15 minutes. Plasmid DNA was then added and incubated for another *ca.* 10-15 minutes before administering to cells. Transfections were typically carried out between 50-70% confluency of cultured cells as is commonly quoted in transfection protocols.

5.2.3 Purification of cell culture derived FUCA1 and FUCA2 and assays

Conditioned media harvested from mammalian cell culture flasks was centrifuged for *ca.* 30 mins at 4416 RCF and the supernatant containing secreted proteins was retained. Supernatant was subsequently filtered through glass wool to remove further cell debris.

When tangential flow filtration was used for purification, a hollow fibre mPES MiniKros® Sampler Filter Module (Spectrum Labs) with a 10 kDa molecular weight cut-off was connected to a KrosFlo® Research Tangential Flow Filtration System (Spectrum Labs). A protease inhibitor tablet (cOmplete, EDTA free (Roche)) was added to clarified conditioned media and media was concentrated by tangential flow. Concentrated media was further clarified by centrifugation for *ca.* 10 mins at 4416 RCF and decanting the supernatant.

5.2.3.1   Immobilised metal affinity chromatography

Clarified media concentrated using tangential flow filtration was applied to a 1 mL HisTrap excel column (GE Healthcare) using a peristaltic pump. Conditioned media was applied once through the column fully, and subsequently recirculated through the column with both inlet and outlet tubing running to the same solution of conditioned media. In some experiments, tangential flow filtration was not used and instead clarified conditioned media was directly loaded onto the column using the procedure outlined above.

The lines on an ÄKTA HPLC (GE Healthcare) were equilibrated with buffer containing a low concentration of imidazole (50 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 8.0. Subsequently, the pre-loaded 1 mL HisTrap excel column was introduced to the system. Low imidazole buffer was run through the system for 5-10 mL before application of a linear gradient (over 20 mL) to buffer containing 500 mM imidazole but otherwise identical to the low imidazole buffer. Eluate fractions were collected using a 96-well plate.

Fractions with an A_{280} response were tested on SDS-PAGE gels and those containing bands at *ca.* 55 kDa which appeared to be suitably pure were concentrated using a 3 kDa molecular weight cut-off Microcon concentrator (Millipore).

5.2.3.2   *Size exclusion chromatography*

Concentrated protein fractions were applied to an S200 16-60 or 16-600 Superdex size exclusion column (GE Healthcare) pre-equilibrated with 10 mM Tris 150 mM NaCl (pH 8.0), connected to an ÄKTA HPLC (GE Healthcare). The same buffer used for column equilibration was passed through the column and fractions were collected using a 96-well plate. Fractions with an A_{280} response were tested on SDS-PAGE gels and those containing bands at *ca.* 55 kDa which appeared to be suitably pure were concentrated using a 3 kDa molecular weight cut-off Microcon concentrator (Millipore).

5.2.3.3   *Enzymatic assays on purified FUCA1*

A number of enzymatic assays were conducted on purified samples of the FUCA1 pGEN1 construct. Deglycosylation of the enzyme was tested using both the EndoH and PNGase enzymes. Further, TEV protease-mediated cleavage of the FUCA1 pGEN1 construct was tested.

Deglycosylation of FUCA1 pGEN1 was tested by use of EndoH$_f$ (New England Biolabs) using the manufacturer's recommended protocol; both native and denaturing conditions were tested. Deglycosylation assays using PNGase F (New England Biolabs) was also conducted according to the manufacturer's recommended protocol in both native and denaturing conditions. Proteolysis assays of FUCA1 pGEN1 were also conducted using AcTEV protease (Life Technologies), as per the manufacturer's recommended protocol.

5.2.3.4 Enzyme Kinetics

The α-L-fucosidase activity of a FUCA1 construct (8His_FUCA1, *vide infra*) was determined. Enzymatic assays were conducted using synthetic substrate 2-chloro,4-nitrophenyl-α-L-fucopyranoside (CNP-fucoside, CarboSynth Ltd.). The product of hydrolysis of this substrate (2-chloro,4-nitrophenol, CNP) absorbs visible radiation with a λ_{max} = 405 nm at pH above its pK_a (ca. 5.4). Experiments to determine the Michaelis-Menten parameters for catalysis by this construct were carried out over a time-course of 2 m during which absorbance was measured at 405 nm using a Cary 100 UV-Vis Spectrophotometer (Agilent Technologies). All solutions used to determine kinetic data were thermally equilibrated (37 ℃).

The rate of hydrolysis of CNP-fucoside by purified 8His_FUCA1 was determined upon solutions of varying concentration of CNP-fucoside. Each solution (1 mL total volume) contained 50 mM HEPES buffer, 100 mM NaCl (pH 7.4) and a known concentration of substrate; 250 nM FUCA1 was added to this after thermal equilibration to initiate hydrolysis. The molar extinction coefficient of CNP (ε_{405} = 1.24×10^3 M^{-1}) was determined by measuring the absorption at 405 nm of a number of solutions of CNP-fucoside at known concentration that had been hydrolysed by excess *Bt*Fuc2970 (See section 2.2.3.1). The Michaelis-Menten parameters for CNP-fucoside hydrolysis were calculated through direct fit to the Michaelis-Menten equation using the Origin graphing software (OriginLab, Northampton, MA).

5.2.3.5 Crystallisation of 8His_FUCA1

Purified 8His_FUCA1 (7.7 mg mL^{-1}) was incubated with 10 mM compound **2** (see **2.2** and Table 2.2) in a 1:1 ratio on ice for ca. 1 hour and centrifuged for ca. 5 min before setting up crystallisation experiments. All crystallisation experiments were performed at *ca.* 18 ℃. INDEX (Hampton Research) and PACT (Molecular Dimensions) crystallisation screens were set up; 120 nL aliquots of 8His_FUCA1:**2** were dispensed along with 120 nL aliquots of mother liquor from these screens in MRC 96-well crystallisation plates. Liquid handling was carried out using a Mosquito liquid handling robot, TTP Labtech. An optimisation tray (MRC MAXI 48-well crystallisation plate) was designed and dispensed by Wendy Offen from the most successful crystal 'hit'.

Crystals were transferred to cryo-protectant solutions containing mother liquor supplemented with 20% glycerol and cryo-cooled using liquid N$_2$. Diffraction was

tested in-house using a Rigaku micromax-007HF X-ray generator, used in conjunction with an Actor robotic sample changer.

5.3   Results and Discussion

A number of means towards the expression of active human fucosidases were carried out. Initially, heterologous expression of FUCA1 and FUCA2 was attempted in *Escherichia coli*, following the reported expression and purification of these enzymes by Liu & Li in 2009[334]. These will be discussed briefly.

5.3.1   Expression of FUCA1 and FUCA2 in heterologous organisms

Initially, FUCA1 and FUCA2 genes in pET-30a^{+} vectors which were codon optimised for expression in *E. coli*, lacked their native signal sequences and contained an N-terminal hexa-histidine affinity tag were tested for expression in BL21 (DE3), pLEMO and C41 cells on a small scale; expression in C41 cells seemed to yield a larger protein band in the soluble phase on SDS gel electrophoresis at the *ca.* 55 kDa range (not shown) so this cell line was used for 200 mL expression and purification tests. Most of the protein obtained was present in inclusion bodies in the insoluble fraction of the cells (Figures 5.4 and 5.5, **A**). Cells were lysed by sonication and soluble protein was purified by Ni^{2+}-affinity chromatography (Figures 5.4 and 5.5). For FUCA1, the protein band observed at *ca.* 55 kDa (Figure 5.4, **C**) was observed to be active against *p*-nitrophenyl-α-L-fucopyranoside on bench tests, causing a colour change in the reaction solution due to formation of a yellow *p*-nitrophenolate product, while this was not the case for FUCA2 (Figure 5.5, **C**). FUCA1 was further purified by size exclusion chromatography (Figure 5.4, **D-E**) with a total yield of *ca.* 3.5 mg L^{-1} culture volume. A second, higher (*ca.* 60 kDa) molecular weight species was observed to co-purify with the *ca.* 55 kDa band assigned as the FUCA1/2 gene transcript in eluates from the Ni^{2+} affinity chromatography steps (Figures 5.4 and 5.5, **B-C**). The size exclusion chromatography purification procedure for FUCA1 however allowed removal of this contaminant (Figure 5.4, **E**). While 3.5 mg L^{-1} is potentially a tractable yield of protein for subsequent X-ray crystallographic studies, further methods were explored to improve protein yield before starting crystallisation trials.
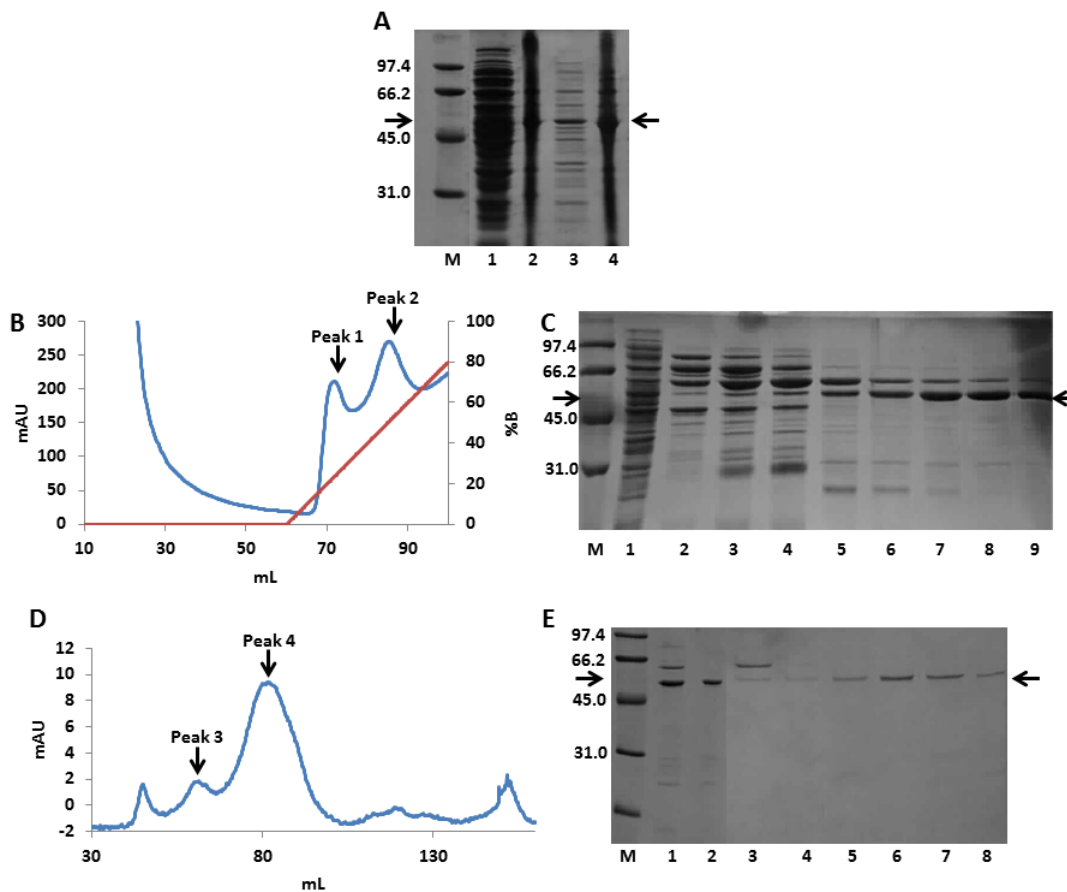
Figure 5.4. Expression and purification of FUCA1 in pET-30a$^+$ vector from transformed C41 cells. The sample run on the first lane of each gel figure (**A**, **C**, **E**) is a Biorad broad range molecular weight marker (**M**), with molecular weights of protein standards displayed. The black arrow on expression gels indicates the position of the band attributed to FUCA1. **A**: expression of FUCA1 in C41 cells; lanes **1-2** post-induction samples from expression at 16 °C (**1**: soluble fraction, **2**: insoluble fraction) lanes **3-4** post-induction samples from expression at 37 °C (**3**: soluble fraction, **4**: insoluble fraction). **B**: Immobilised metal affinity purification of recombinant FUCA1 in C41 cells; A$_{280}$ response is shown in blue while the percentage of high imidazole in the mixed buffer sample is shown in orange. **C**: Immobilised metal affinity purification of FUCA1 expressed in C41 cells; lane **1** shows a sample of the protein loaded onto the affinity column, the samples run on lanes **2-9** are from successively higher run volumes (samples **2-4** are from peak 1, while samples **5-9** are from peak 2, see **B**). **D**: Size exclusion chromatography of FUCA1 expressed in C41 cells; A$_{280}$ response is shown in blue. **E**: Size exclusion chromatography of FUCA1 expressed in C41 cells; lane **1** is a sample taken from the protein loaded onto the column, lanes **3-8** are eluate fractions from successively higher run volumes (samples **3-4** are from peak 3, while samples **5-8** are from peak 4, see **D**) while lane **2** is concentrated protein pooled from eluate fractions from peak **4**.

144
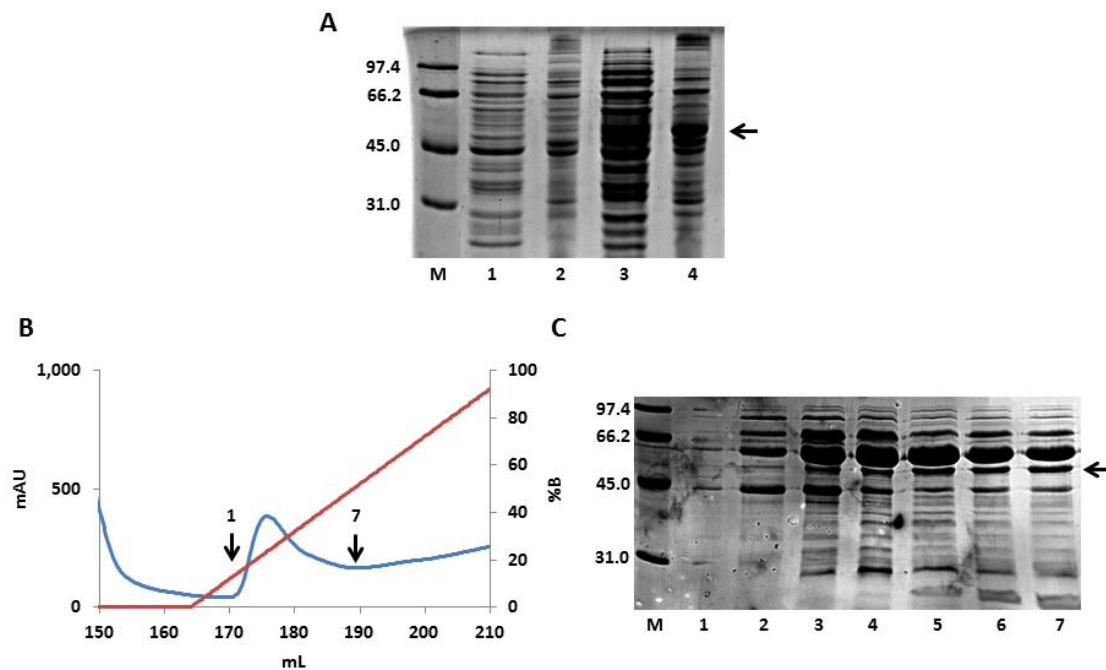
Figure 5.5. Expression and purification of FUCA2 in pET-30a$^+$ vector from transformed C41 cells. The sample run on the first lane of each gel figure (**A**, **C**) is a Biorad broad range molecular weight marker (**M**), with molecular weights of protein standards displayed. The black arrow on expression gels indicates the position of the band attributed to FUCA1. **A**: expression of FUCA2 in C41 cells; lanes **1-2** pre-induction samples (**1**: soluble fraction, **2**: insoluble fraction) lanes **3-4** post-induction samples from expression at 16 °C (**3**: soluble fraction, **4**: insoluble fraction). **B**: Immobilised metal affinity purification of recombinant FUCA2 in C41 cells; A$_{280}$ response is shown in blue while the percentage of high imidazole in the mixed buffer sample is shown in orange. **C**: Immobilised metal affinity purification of FUCA2 expressed in C41 cells; the samples run on lanes **1-7** are from successively higher run volumes (between fractions **1** to **7** as shown on the trace, see **B**).

It was assumed that the co-purified protein which runs on SDS-PAGE at *ca.* 60 kDa may be the chaperonin GroEL, however no experimentation either confirmed or denied this. Chaperonins such as GroEL act to aid in the correct folding of unfolded proteins, so it was expected that improving the folding capacity of the proteins may improve protein purity and yield. Attachment of the yeast small ubiquitin-like modifier (SUMO) peptide gene to proteins has been shown to improve protein solubility and folding, even for protein targets which are difficult to express in a soluble form.[335-337] This method was attempted for expression of FUCA1 and FUCA2; the genes encoding these proteins were sub-cloned into the pETite® N-His

SUMO Kan vector and used to transform BL21 (DE3) cells. Unfortunately, no significant difference in the yield of soluble protein was observed from SDS-PAGE (Figure 5.6) compared to expression in the pET-30a$^+$ vector (Figures 5.4 and 5.5, **A**) so other methods were explored for the expression of these enzymes.
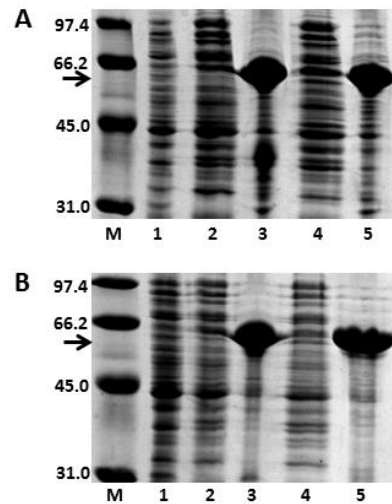
Figure 5.6. Expression of FUCA1/2 genes cloned in pETite® N-His SUMO Kan vector from transformed Hi-Control BL21 (DE3) cells. The sample run on the first lane of each figure is a Biorad low range molecular weight marker (**M**), with molecular weights of protein standards displayed. The black arrow on expression gels indicates the position of the band attributed to FUCA1/2. **A**: expression of FUCA1 in Hi-Control BL21 (DE3) cells. **B**: expression of FUCA2 in Hi-Control BL21 (DE3) cells. Lanes in each gel (**A, B**): lane **1** pre-induction sample from the soluble phase, lanes **2-3** post-induction samples from expression at 18 °C (**2**: soluble fraction, **3**: insoluble fraction), lanes 4-5 post-induction samples from expression at 37 °C (**4**: soluble fraction, **5**: insoluble fraction).

The FUCA1 and FUCA2 protein transcripts were analysed more closely. Use of the DIANNA disulfide bond prediction web-server[338] predicted two (FUCA1) or three (FUCA2) disulfide bonding pairs for the enzymes, and in homology models of the human α-ʟ-fucosidases prepared using iTASSER[339] the cysteine residues predicted by DIANNA appear to be close enough in space for disulfide bonding to occur. The formation of disulfide bonds in eukaryotes is achieved through the protein disulfide isomerase enzyme, which is localised in the endoplasmic reticulum and thus encountered during protein folding[340,341]; in bacterial cells however it is achieved by a number of enzymes in the perisplasmic space, which are not encountered during protein folding or encountered at all by proteins which do not enter this

146

compartment[342]. The pelB leader sequence of *Erwinia carotovora*[343] can be introduced into bacterial cells to stimulate compartmentalisation of a nascent protein towards the periplasmic space, thereby allowing for the formation of disulfide bonds[344].

FUCA1 and FUCA2 were thus sub-cloned into the pET-22b[+] plasmid, which contains a pelB leader sequence for protein trafficking to the periplasmic compartment. BL21 (DE3) cells were then transformed by these plasmids and expression levels determined (Figure 5.7), however again the solubility of expressed FUCA1 did not appear to be greater than that from the pET-30a[+] vector (Figure 5.4 and 5.5, **A**).
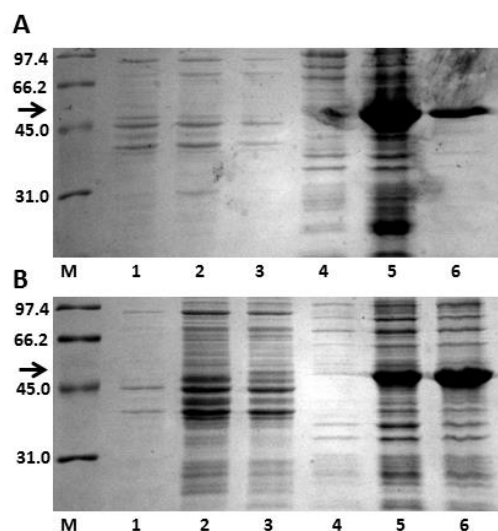
Figure 5.7. Expression of FUCA1/2 genes cloned in pET22b[+] from transformed BL21 (DE3) cells. The sample run on the first lane of each figure is a Biorad broad range molecular weight marker (**M**), with molecular weights of protein standards displayed. The black arrow on expression gels indicates the position of the band attributed to FUCA1/2. **A**: expression of FUCA1 in BL21 (DE3) cells. **B**: expression of FUCA2 in BL21 (DE3) cells. Lanes in each gel (**A**, **B**): lane **1** pre-induction sample from the soluble phase, lanes **2-3** post-induction samples from the soluble phase (**2**: expression at 16 °C, **3**: expression at 37 °C), lane **4** pre-induction sample from the insoluble phase, lanes **5-6** post-induction samples from the insoluble phase (**2**: expression at 16 °C, **3**: expression at 37 °C).

Pichia pastoris was further used as an alternative vector for expression of FUCA1. These experiments were conducted by Dr. Jared Cartwright of the Biosciences Technology Facility, University of York. While expression of FUCA1 in *Pichia*

pastoris did lead to production of some soluble protein, it was deemed that the low yield obtained was not worth following up.

At around this time, facilities for protein expression in both baculovirus-infected insect cells and mammalian cells were being developed at the York Structural Biology Laboratory. Since limited progress had been made with the recombinant expression of FUCA1 and even less with FUCA2, and in order to learn further skills, it was decided to pursue the expression of these enzymes in a more native environment through mammalian tissue culture (see **5.3.2**). FUCA1 and FUCA2 are known glycoproteins[306,309,310,312] and have predicted disulfide bonds (*vide supra*). As such, expression of the enzymes in a tissue culture environment should improve the solubility of the protein produced as the cellular environment and enzymatic machinery of cultured mammalian cells more closely resembles that in heterologous expression. This expression system is particularly suited to the expression of glycoproteins, proteins comprising disulfide bonds and secreted proteins, and typical yields of purified proteins produced from the HEK-293T cell line, for example, range between 1-80 mg L^{-1}.[345]

5.3.2   Expression of FUCA1 and FUCA2 in mammalian tissue culture

5.3.2.1   *Constructs from the CCRC*

9 constructs obtained from the complex carbohydrate research centre, five harbouring the gene for human α-ʟ-fucosidase FUCA1 and four harbouring that for FUCA2, were tested for expression levels in cultured HEK-293T and CHO-K1 adherent cells. Small scale expression tests were run of both secreted protein and protein harvested from cells (Figure 5.8).

Cells transfected with constructs containing green fluorescent protein were visibly green and the conditioned media from these samples also appeared green. While caution should be taken in interpreting differences in expression level between different membranes using Western blots, expression in CHO-K1 cells seemed to be uniformly lower than that in HEK-293T cells. The secretion of all of the constructs also seemed to be efficient, as most of the protein observed was from the soluble phase. The FUCA1 pGEN1 construct was chosen for further study as it provided a large quantity of protein and only a single band was observed on western blot; it was expected that the multiple bands observed for other constructs were due to degradation of the protein.
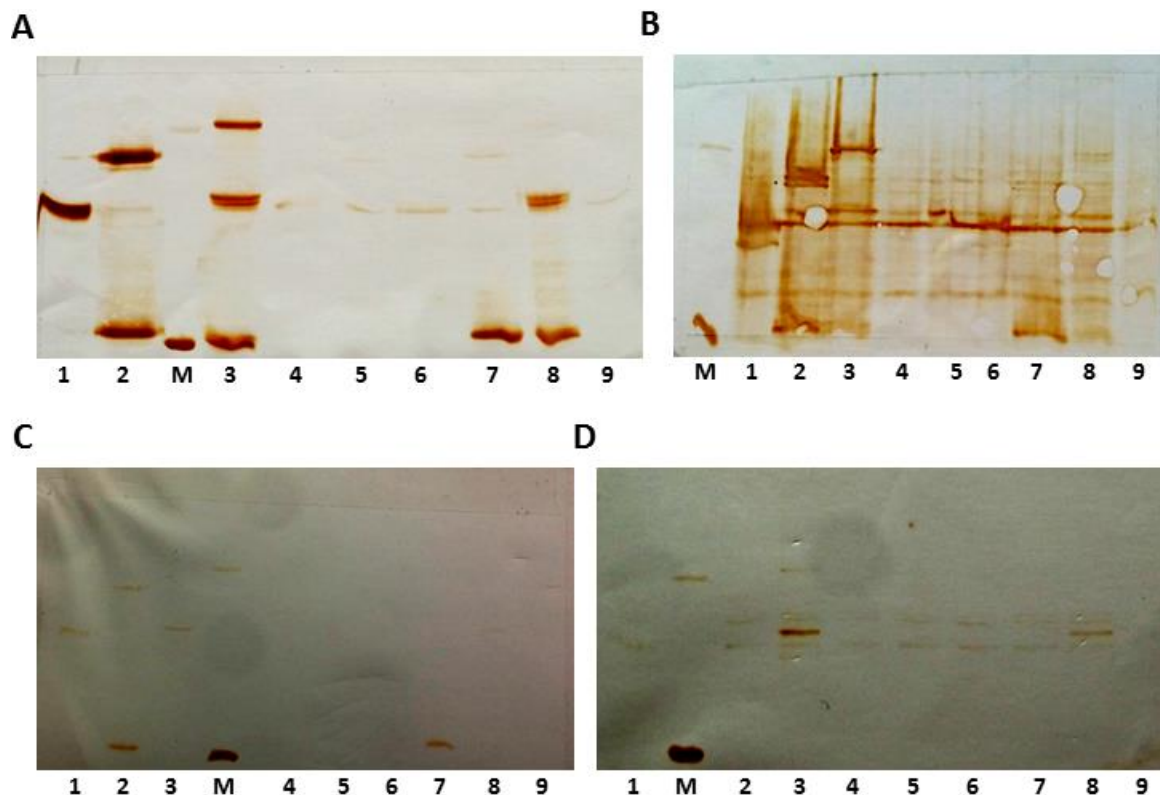
Figure 5.8. Expression levels of FUCA1 and FUCA2 constructs in HEK-293T and CHO-K1 cells. Expression levels were determined by Western blotting using an anti-His:HRP monoclonal antibody. Protein from both the soluble and insoluble phases was visualised; the samples run were of secreted protein from HEK-293T (**A**) and CHO-K1 (**C**) and protein isolated from cells of HEK-293T (**B**) and CHO-K1 (**D**). **M** is a low range molecular weight marker (Biorad). Samples **1-9** are, respectively, pGEN1, pGEN2, pGEN3, pGEC1 and pGEC2 constructs of FUCA1 and pGEN1, pGEN2, pGEN3 and pGEC1 constructs of FUCA2.

Subsequently, expression of the FUCA1 pGEN1 construct was carried out using different transfection reagents (linear polyethylenimine, polyethylenimine "Max") at a range of transfection reagent:DNA ratios, in both HEK-293T and CHO-K1 cells in order to determine the optimal transfection reagent (data not shown). Linear polyethylenimine was chosen for subsequent experiments as this transfection reagent and ratio seemed to give the highest yield of FUCA1 as observed by western blot.

Expression of the FUCA1 pGEN1 construct in HEK-293T cells was scaled up to *ca.* 450 mL in 15 T175 flasks. Two harvests were conducted, at 3 and 5 days after cell transfection. Conditioned media was concentrated by tangential-flow filtration and protein was purified using Ni^{2+} affinity chromatography followed by size exclusion

chromatography to yield 1 mg of protein from a total volume of 900 mL conditioned media (Figure 5.9). This purified protein was observed to be active against 2-chloro,4-nitrophenyl-α-ʟ-fucopyranoside on bench tests, causing a colour change in the reaction solution due to formation of a yellow 2-chloro,4-nitrophenolate product. The purified protein produced by cultured mammalian cells was observed to run higher on SDS-PAGE gels than that produced recombinantly by *Eshcerichia coli*, indicating the presence of N-linked glycans (Figure 5.9, **D** *vs. e.g.* Figure 5.4, **E**). Curiously, when expression of this construct in otherwise similar conditions (the same HEK-293T cell line albeit with a higher passage number, similar cell confluency at the time of transfection etc.) but using roller bottles for cell culturing, a *ca.* 10-fold lower yield of 100 µg L^{-1} purified protein was obtained.
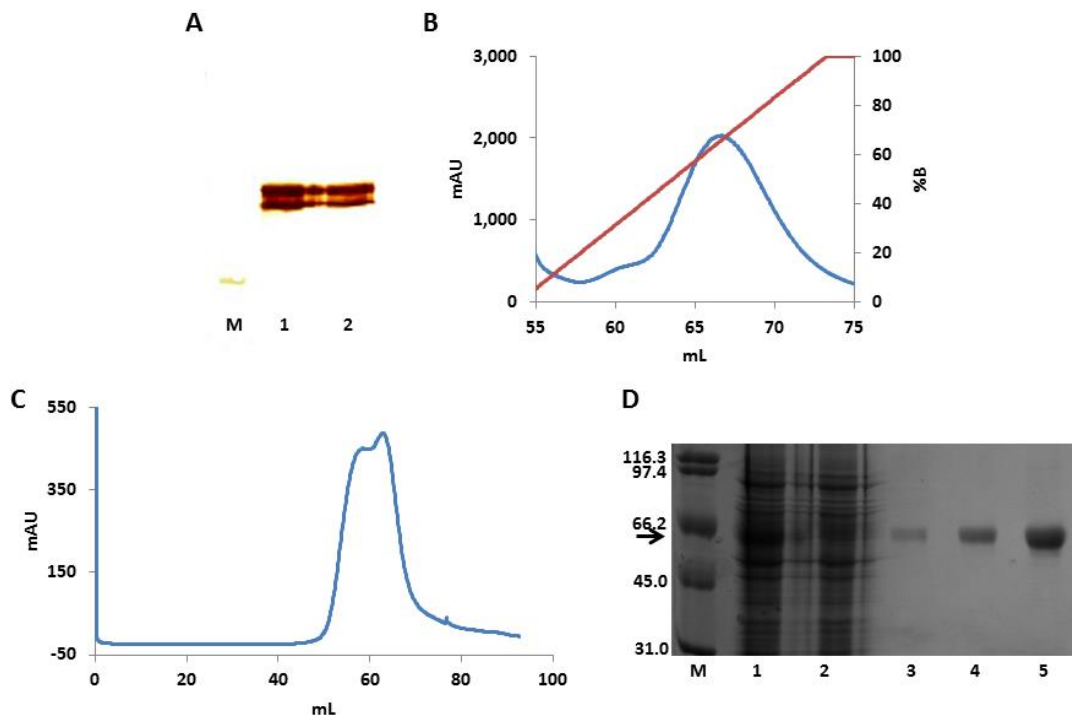
Figure 5.9. Expression and purification of FUCA1 pGEN1 construct in cultured HEK-293T cells. **A**: Western blots of FUCA1 expression, His-tagged proteins were visualised using an anti-His:HRP monoclonal antibody (**M**: biorad broad range molecular weight marker **1**: first 450 mL harvest, 3 days after transfection **2**: second 450 mL harvest, 5 days after transfection). **B**: Immobilised metal affinity purification of FUCA1 pGEN1; A_{280} response is shown in blue while the percentage of high imidazole in the mixed buffer sample is shown in orange. **C**: Size exclusion chromatography of FUCA1 pGEN1; A_{280} response is shown in blue. **D**: Samples from purification were visualised on an SDS-PAGE gel (**M**: biorad broad range molecular weight marker, **1**: Protein concentrated by tangential-flow filtration. **2**: sample from protein that had been run through the 1 mL Hi-Trap Excel column a single time, **3-4**: samples from Ni^{2+}-affinity chromatography purification of FUCA1 pGEN1, **5**: Concentrated FUCA1 pGEN1 sample loaded on the size exclusion column). Unfortunately, while samples from the size exclusion purification of FUCA1 pGEN1 were loaded on the gel, the intensity of the protein bands was too low to visualise.

5.3.2.2   *Assays on purified FUCA1 pGEN1*

Enzymatic assays were run on FUCA1 pGEN1 after Ni^{2+}-affinity and size exclusion purification. As the pGEN1 construct contains an N-terminal His_8 affinity tag that is cleavable by TEV protease, cleavage of the affinity tag was tested using TEV

protease (Figure 5.10); further, assays using the deglycosylating enzymes EndoH and PNGase were conducted (Figure 5.10).

Treatment of FUCA1 pGEN1 with TEV protease yielded no visible gel shift (Figure 5.10, **1**, **5-7**), making it unlikely that proteolytic cleavage was catalysed by this enzyme. The protein bands on treatment with EndoH using both standard and denaturing reaction conditions appear similar (Figure 5.10, **2-4**, **8**), and show the presence of heterogeneous species. This stands to reason as EndoH is not capable of accepting complex glycans as substrate; these species comprise a proportion of the glycans produced by mammalian cells. Treatment with PNGase however (Figure 5.10, **1**, **9-10**) yielded a visible gel shift towards an apparently homogenous species which is perhaps the fully deglycosylated enzyme (less its innermost GlcNAc residues).
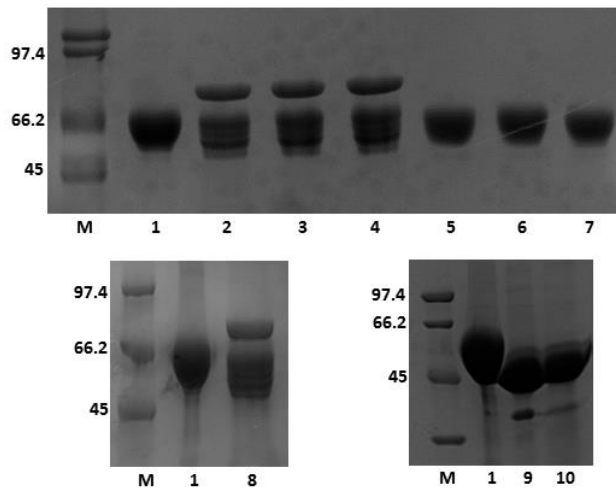
Figure 5.10. Enzymatic assays on purified FUCA1 pGEN1. Gels displayed are 12 % SDS-PAGE gels. Molecular weight ladders (Biorad broad range, **M**) are shown along with the molecular weight of the protein standards in kDa. Lane 1: native FUCA1. Lanes **2-4**: FUCA1 treated with native $EndoH_f$ assays for 1, 2 and 4 hours, respectively. Lanes **5-7**: FUCA1 treated with AcTEV protease for 1, 2 and 4 hours respectively. Lane **8**: FUCA1 treated with $EndoH_f$ in a denaturing assay. Lane **9**: FUCA1 treated with PNGase F in a native assay. Lane **10**: FUCA1 treated with PNGase F in a denaturing assay.

Expression of the FUCA1 pGEN1 construct in HEK-293T cells allowed the purification towards seemingly pure protein (Figure 5.9, **B-D**). The pGEN1 construct however contains a long (32 residues) region comprising a His_8 affinity

tag, Strep tag, attB1 sequence and TEV proteolysis site prior to the start of the FUCA1 gene product (Figure 5.3); long linker regions tend to hinder crystallogenesis and must typically be cleaved prior to crystal formation. As proteolysis of FUCA1 pGEN1 could not be achieved using TEV protease (Figure 5.10, **1**, **5-7**), and in order to reduce the number of steps taken for purification of a crystallisable construct, constructs were designed with minimal octahistidine tags at the N-terminus of the FUCA1 gene product.

5.3.2.3 *FUCA1 constructs containing minimal octahistidine affinity tags*

Three FUCA1 constructs were designed containing minimal, non-cleavable His-Tags either directly upstream of the FUCA1 coding sequence (8His_FUCA1), or containing one (8HisSG_FUCA1) or two (8HisSGSG_FUCA1) SG dipeptides before start of the FUCA1 coding sequence (for methodological details, see *5.2.2.3*).

Small scale expression tests were conducted on each of the constructs (Figure 5.11) in both HEK-293T and CHO-K1 cells, using both linear polyethylenimine and polyethylenimine "Max" as transfection reagents. As only a single protein band was observed on anti-His western blot for the 8His_FUCA1 construct, while multiple bands were seen for the 8HisSG_FUCA1 and 8HisSGSG_FUCA1 constructs (Figure 5.11), subsequent scale-up was carried out using 8His_FUCA1.
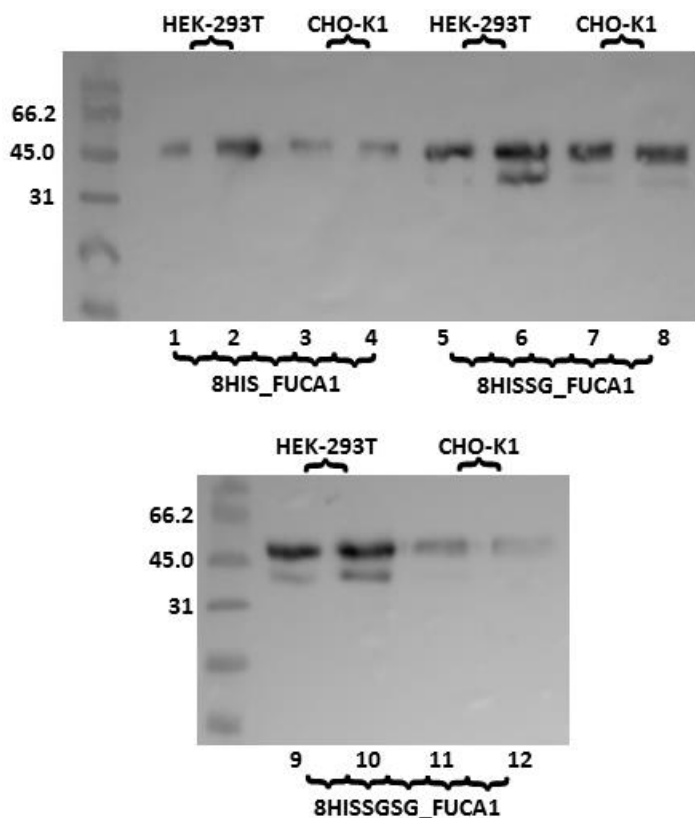
Figure 5.11. Small scale expression testing of FUCA1 constructs with minimal His-Tags. Molecular weight ladders (**M**) are shown along with the molecular weight of the protein standards in kDa. Constructs and cell lines are as displayed in braces. Odd lane numbers correspond to expression after transfection with a ratio of 1.7:1 linear polyethylenimine while even lane numbers correspond to expression after transfection with a ratio of 3.5:1 polyethylenimine "Max". Pairs of samples (**1** and **2**, **3** and **4** *etc.*) represent duplicate expression tests with plasmid DNA isolated from unique colonies formed after transformation of bacterial cells; two colonies were tested for each octahistidine-tagged construct.

Expression was thus scaled up to 90 mL (three T175 flasks). Unlike previously with FUCA1 pGEN1, transfection was effected when cells had reached *ca.* 90 % confluency (for previous tissue culture experiments, transfection was carried out at *ca.* 50-70 % confluency). Two cell harvests were conducted, each of 90 mL conditioned media. Harvested protein was clarified and purified using immobilised metal affinity chromatography, yielding 1.2 mg of 8His_FUCA1 in total; conditioned media from the first and second harvests were purified separately and less protein was purified from the second harvest (Figure 5.12).
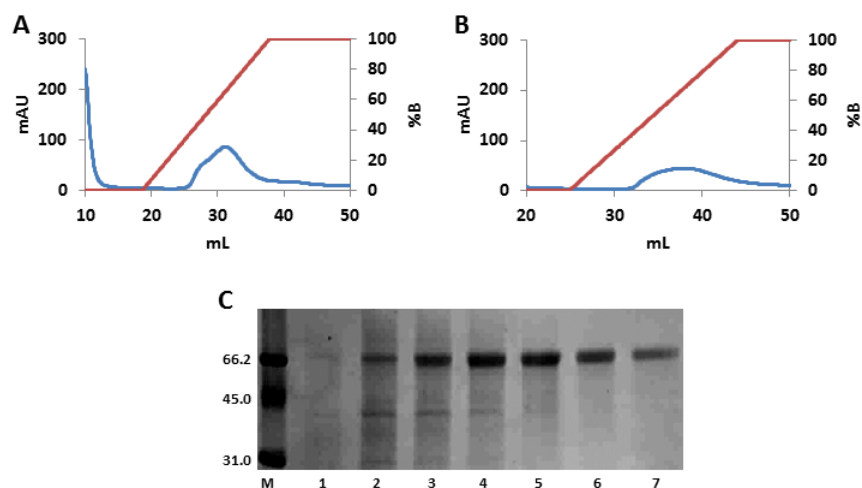
Figure 5.12. Expression and purification of 8His_FUCA1 expressed in cultured HEK-293T cells. The sample run on the first lane of each gel figure (**C**, **E**) is a Biorad broad range molecular weight marker (**M**), with molecular weights of protein standards displayed. **A** and **B**: Immobilised metal affinity purification of 8His_FUCA1; A_{280} response is shown in blue while the percentage of high imidazole in the mixed buffer sample is shown in orange (**A** is from the first 90 mL harvest while **B** is from the second). **C**: SDS-PAGE gel of samples from Ni^{2+} purification of the first harvest of 8His_FUCA1 (**A**, also representative of the second harvest, **B**). **M**: Biorad broad range molecular weight marker (molecular weights of protein standards are displayed), **1-7**: samples from increasing run volume, encompassing each fraction across the peak in **A**.

Expression was scaled up to 450 mL (fifteen T175 flasks), with transfection at *ca.* 50-70 % confluency. Purification of the 450 mL clarified conditioned media by immobilised metal affinity chromatography only yielded a further 1.4 mg 8His_FUCA1, however. This dramatic difference in expression yield may be due to the confluency of cultured cells at the time of transfection. While a number of transfection protocols cite confluencies for optimal transfection efficiency in the approximate range 40-80%, Aricescu *et al.* noted that "optimal transfection efficiency is achieved when adherent cells reach about 90% confluency".[346] Nevertheless, the two samples of 8His_FUCA1 were pooled, concentrated and further purified by gel filtration chromatography to yield the enzyme at high purity (Figure 5.13).
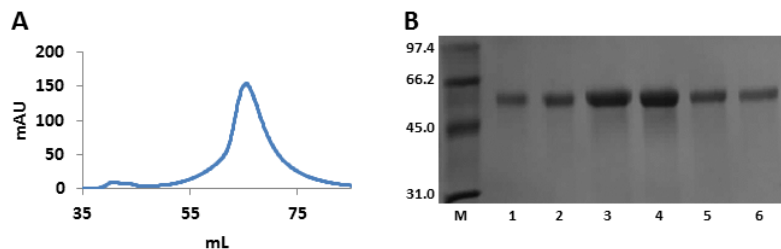
Figure 5.13. Purification of 8His_FUCA1 by size exclusion chromatography. **A**: Size exclusion chromatography of 8His_FUCA1; A_{280} response is shown in blue. **B**: SDS-PAGE gel of sample from size-exclusion chromatography of 8His_FUCA1. **M**: Biorad broad range molecular weight marker (molecular weights of protein standards are displayed), **1-6**: samples from increasing run volume across the large peak in **A**.

5.3.2.4   Assays and crystallisation of minimal octahistidine FUCA1 construct

The Michaelis-Menten parameters for CNP-fucoside hydrolysis by purified 8His_FUCA1 were determined; $K_M = 0.099 \pm 0.01$ mM, $V_{max} = 1305 \pm 47$ min^{-1}, Figure 5.14. These values correlate well with those determined previously by Liu *et al.* using purified FUCA1 expressed recombinantly in *Escherichia coli* ($K_M = 0.23$ mM, $V_{max} = 17.1$ s^{-1}).[334] The K_M value observed for 8His_FUCA1 is slightly lower; this could be due to the difference in substrate (2-chloro,4-nitrophenyl-α-L-fucopyranoside rather than 4-nitrophenyl-α-L-fucopyranoside), or due to differences in the experimental pH and buffer system. It is unlikely that the presence of glycans in the tissue-culture derived sample has any effect on substrate binding as treatment of human liver derived FUCA1 with N-glycanase has previously been shown not to affect its K_M value.[312]
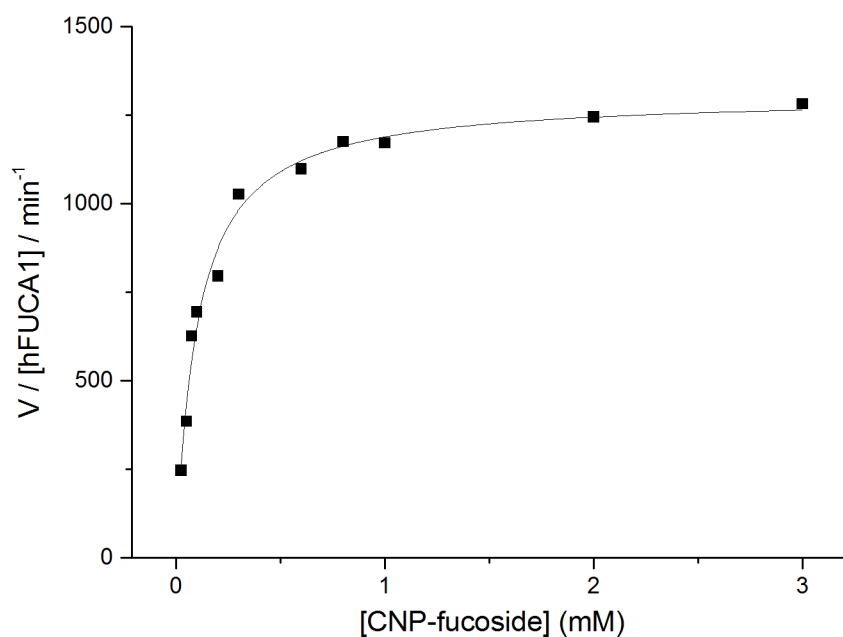
Figure 5.14. Michaelis-Menten kinetics of CNP-α-ʟ-fucopyranoside hydrolysis by 8His_FUCA1.

Crystallisation screens were also set up with purified 8His_FUCA1, using the sitting drop vapour diffusion method. In order to improve the propensity for the enzyme to crystallise, it was incubated with α-ʟ-fucosidase inhibitor **2** (see **2.2** and Table 2.2) for *ca*. 1 hour before crystallisation experiments. Initial crystallisation screens were kindly optimised by Wendy Offen to yield crystals which appeared visually promising (Figure 5.15); these crystals were tested for diffraction in-house. Unfortunately, diffraction of these crystals resulted in only a few, highly intense diffraction spots, interpreted as being due to the diffraction of small molecule crystals. As the crystallisation mother liquor contained only buffer and polyethylene glycol (optimised from PACT condition D1, Molecular Dimensions; this condition contains 0.1 M ᴅʟ-malic acid, MES monohydrate, Tris buffer system pH 4.0, 25 % w/v PEG 1500), it seems unlikely that the spots are due to mother liquor, but may be due to crystals of **2** that formed in the crystallisation drop.

157

Figure 5.15. Crystals formed in optimisation trays from the incubation mixture of 8His_FUCA1 and inhibitor **2**. On diffraction studies, these crystals appeared to be salt crystals.

The α-ʟ-fucosidase sample was not deglycosylated prior to crystallogenesis experiments. The presence of N-glycans in a protein sample is known to interfere with crystallisation due to the structural and conformational heterogeneity N-glycans introduce to a protein, and few crystal structures have been determined of proteins containing extensive N-glycosylation.[325] Therefore, these proteins are typically deglycosylated enzymatically prior to crystallisation experiments. [325] The deglycosylating enzyme PNGase, which accepts as substrate proteins with all types of N-glycan structure, converts the Asn residue forming an N-glycosidic bond to an Asp residue, thereby altering the surface charge of the enzyme, and this is also known to interfere with crystallogenesis and the polydispersity of the enzyme sample, causing its aggregation[325]. Unfortunately, the EndoH enzyme, which leaves a single GlcNAc monosaccharide at each glycosylation site, does not alter the surface charge of the protein, interfere with crystallogenesis or cause aggregation, does not act to cleave the complex-type N-glycans which comprise a portion of the N-glycans produced during the mammalian N-glycan biosynthetic pathway. It was shown previously that EndoH-catalysed deglycosylation of FUCA1 pGEN1 led to a heterogenous reaction product (*5.3.2.2*, Figure 5.10).

EndoH is thus the structural biologists' preferred deglycosylating enzyme, however means must be taken to avoid the biosynthesis of complex-type N-glycans during protein expression. Knowledge of the N-glycan biosynthetic pathway can be exploited to yield N-glycosylated proteins lacking complex-type modifications. For example, the α-mannosidase inhibitors kifunensine or swainsonine can be introduced to media in order to inhibit glycan N-glycan biosynthesis, limiting N-

glycan biosynthesis at the stage of oligomannose-type or hybrid-type N-glycans respectively.[325] Alternatively, a glycosylation deficient cell line such as HEK-293S GnTI[-/-][328] could be used for the expression of human α-ʟ-fucosidases with a higher propensity to crystallise.

5.4 Conclusions

The α-ʟ-fucosidases FUCA1 and FUCA2 were expressed recombinantly in *Escherichia coli*; FUCA1 was further purified by immobilised metal affinity chromatography and size-exclusion chromatography with a yield of 3.5 mg L^{-1} active protein, however this was not the case for FUCA2, which could not be purified. For both enzymes, most of the expressed protein was present in inclusion bodies. Attempts to further solubilise the proteins by using constructs both with an N-terminal SUMO module and pelB leader peptide to facilitate solubility and disulfide bond formation did not appear to improve protein solubility. While 3.5 mg L^{-1} purified enzyme is not a high yield for recombinant expression of a soluble protein in *Escherichia coli*, it may be possible to use this method to yield enough protein for crystallisation studies.

FUCA1 and FUCA2 were also expressed in cultured CHO-K1 and HEK-293T cells. A number of constructs containing the gene transcripts for FUCA1 and FUCA2 received from the Moremen Lab (Complex Carbohydrate Research Centre, University of Athens, Georgia) were expressed. Transfection of the shortest N-terminal construct, comprising both octahistidine and Strep-II affinity tags, into cultured CHO-K1 and HEK-293T was observed to allow expression of the most stable α-ʟ-fucosidase transcripts for both proteins from those tested, while the level of FUCA2 expression was lower than that of FUCA1. FUCA1 was further purified from conditioned media using a combination of immobilised metal affinity chromatography and size-exclusion chromatography to yield high purity protein which was active against the synthetic substrate 2-chloro,4-nitrophenyl-α-ʟ-fucopyranoside, however at low yield (*ca*. 1.1 mg L^{-1}).

For FUCA1 an additional construct comprising just an octahistidine affinity tag at the N-terminus of the FUCA1 gene transcript was cloned and used to transfect both CHO-K1 and HEK-293T cells. Transfection of highly confluent (*ca*. 90 %) HEK-293T cells using this construct allowed purification of active α-ʟ-fucosidase at greater than 6 mg L^{-1} yield. The Michaelis-Menten parameters of this purified enzyme were determined (K_M = 0.099 mM, V_{max} = 1305) and are in agreement with literature values. Crystallisation screens were set up with this purified enzyme complexed with a competitive α-ʟ-fucosidase inhibitor, however the only crystals that formed from these experiments were likely formed by the inhibitor itself. It is unsurprising that crystals of the enzyme did not form as at the time of crystallogenesis, the sample still contained a heterogenous pattern of N-glycans, which is known to interfere with crystallogenesis. Further attempts towards the

realisation of the crystal structure of human GH29 enzymes from a tissue culture source would benefit from either expression in glycosylation deficient cell lines or the introduction of inhibitors of the N-glycan biosynthetic pathway to the culture media during transfection. The yield of purified FUCA1 from transfected cultured HEK-293T cells was observed to be maximised when transfection was conducted at high (*ca.* 90%) confluency, and when expression was carried out in T175 flasks rather than by using the less labour-intensive roller bottle system.

FUCA2, which could not be purified in an active form by expression in a heterologous system, could be expressed and purified in cultured mammalian cells, however at lower yield than FUCA1. While FUCA1 perhaps has more significance in the clinical setting, the structural characteristics of FUCA2 may be advantageous in the study and treatment of host-pathogen interactions due to its documented role in *Helicobacter pylori* virulence.[4]

While the transient transfection system employed herein can be employed rapidly for the testing of new plasmid DNA clones, it is not the most ideal for large scale protein expression, as is likely to be required for structural studies. The generation of stable clones expressing GH29 fucosidases with relevant affinity tags for purification may provide a better long-term alternative for protein production; as it is likely that many studies downstream of an initial structure determination would be carried out, the realisation of a robust protein expression system for these enzymes would be ideal.[347]

Hopefully, the work documented herein will pave the way for the successful structural characterisation of these important enzymes in the near future.

Chapter 6: Conclusions and future perspectives

L-fucose has vital importance to the correct functioning of a number of biological species in a wide range of contexts, *inter alia*, in the formation of durable extracellular matrices, molecular adhesion and signalling events, fertility and development. α-L-fucosidases act to cleave L-fucose residues from these biological species and these enzymes have a relationship with disease phenotypes in mammals. Genetic factors leading to lack of lysosomal α-L-fucosidase activity cause the neurodegenerative lysosomal storage disorder fucosidosis. Further, the enzymes have a relationship with cancer, as many neoplastic cells differentially express fucosylated glycans such as the sialyl-Lewis^x antigen on their cell surface and may also upregulate expression of α-L-fucosidase. Finally, molecular adhesion and infectivity by the stomach pathogen *Helicobacter pylori* has been shown to correlate with plasma α-L-fucosidase activity. α-L-fucosidases are thus critically important enzymes, and it may be envisioned that the inhibition of these enzymes may provide clinical routes to treatment of the illnesses discussed above. As such, there has been generous research interest in developing inhibitors of α-L-fucosidase. As enzyme function is largely related to the 3-dimensional structure adopted in solution, the design of potent enzyme inhibitors may be aided by prior knowledge of the 3-dimensional structure of the enzyme in question along with understanding of its reaction mechanism and transition-state(s). Unfortunately, while the 3-dimensional structures of a number of bacterial α-L-fucosidases were known at the onset of this work, no structure for either mammalian α-L-fucosidase (lysosomal or plasma) was known.

Throughout this work, the 3-dimensional basis for the inhibition of GH29 family α-L-fucosidase enzymes by 5-membered iminosugars was probed by a combination of X-ray crystallography and kinetic methods. The conformation adopted by these inhibitors when complexed with a bacterial α-L-fucosidase from *Bacteroides thetaiotaomicron* (*Bt*Fuc2970) was observed crystallographically to be E_3 (equating to E_4 for a pyranoside), supporting the catalytic mechanism of this class of enzyme, which proceeds through a $^3H_4$ transition state; this conformation was concluded to occur due to hydrogen bonds made to key residues in the enzymatic active site. Through these studies, a hydrophobic ridge was also observed in GH29-A subfamily α-L-fucosidases which appears to provide a hydrophobic interaction motif for glycosidic aglycon moieties.

5-membered iminocyclitols have proven a versatile system for the inhibition of glycoside hydrolases, particularly of GH38 golgi α-mannosidase-II[222,348,349], and α-L-fucosidase as presented in this work. These heterocycles have the advantage that their conformation is dictated not by ring energetics but by the ring substituents, allowing greater design of substituents in order to fill pockets in the enzymatic active site to yield inhibitors which may be less promiscuous than 6-membered ring inhibitors. While these compounds have much potential for glycoside hydrolase inhibition in the clinical setting, they have only been applied to a small number of glycoside hydrolase families, and their more widespread use may provide access to powerful, selective inhibitors for treatment of a range of conditions including lysosomal storage disorders and cancer.

Irreversible enzyme inhibitors can be powerful tools for molecular imaging. Such compounds which comprise fluorophores or other moieties may be utilised in a process known as activity-based protein profiling, whereby the catalytic mechanism of an enzyme is exploited to yield highly specific and bio-orthogonal molecular imaging reagents. The 3-dimensional crystal structures of compounds upon which α-L-fucosidase activity-based probes can be appended were determined when complexed with *Bt*Fuc2970. These structures serve to show unequivocally the covalent bond formed between the enzymatic nucleophile residue and the fucose-mimic ring, after trans-diaxial ring-opening of the aziridine moiety present in these compounds, and represent the first crystal structures reported of a glycoside hydrolase ligated with such an aziridine compound. Compounds such as these may have utility in discovering novel enzymes with α-L-fucosidase activity, and as reporters of this activity, most prominently in studies related to fertility.

Activity-based probes like those discussed in the current work provide a powerful system for the study of enzyme activity and the dynamics of enzyme biosynthesis, trafficking and degradation. This technique has been applied in the past to a range of glycoside hydrolases such as glucocerebrosidase and other β-glucosidases, [274,350], β-galactosidases[269] and hexosaminidases[273], and will likely be expanded to other retaining glycoside hydrolases in the coming years, likely including clinically important enzymes. One shortcoming is that the covalent trapping of glycoside hydrolases to report on their activity has thus far only been applied against retaining glycoside hydrolases. Given a suitable synthetic chemistry framework, it should be possible to apply the techniques developed for retaining glycoside hydrolases to their inverting counterparts. For example, recent research reported the synthesis of N-bromoacetylglycosylamines and bromoketone-C-glycosides to target

inverting endo-xyloglucanases[351], and it may be envisaged that these compounds could be synthesised to include reporter moieties for use as activity-based probes.

Another potential use for activity-based probes such as those described herein is in the labelling of glycosyltransferases. While the glycosylation occurring during *N*-glycan biosynthesis has previously been tracked using labelled nucleoside sugars,[268,352] individual glycosyltransferases involved in this pathway or other glycosyl transfers have not yet been visualised by activity-based protein profiling. Many glycosyltransferases have clinical relevance, such as oligosaccharyltransferase which transfers a dolichol-linked precursor N-glycan to nascent glycoproteins on the luminal side of the endoplasmic reticulum and is thus critical for the correct maturation and folding of many proteins[353] and O-GlcNAc transferase, which transfers GlcNAc in a β-linkage to many proteins in the nuclear compartment; dysregulation of this enzyme is linked to diabetes[354], cancer[355] and neurodegenerative disorders[356]. Fucosyltransferases, meanwhile, are important due to their generation of fucosylated glycans which have been linked to cancer[73], the immune response[44,45] and host-pathogen interactions[5,6]. The development of activity-based probing techniques for trapping of glycosyltransferases thus has exciting prospects for improving our understanding of glycosylation in disease.

The results discussed above further our understanding of the enzymatic mechanism of α-L-fucosidases and may further aid in the design of more potent inhibitors of the enzyme class, however they do nothing to address a critical lack of understanding by the field, namely that the 3-dimensional structures of human α-L-fucosidases are still unknown. As these are the enzymes that are important to human health, this represents a bottleneck to growth of the field and the potential development of therapeutics for fucosidosis, cancer and infection by *Helicobacter pylori*. As such, a number of methods were carried out to attempt the expression of human fucosidases FUCA1 and FUCA2. While these enzymes have previously been expressed and purified to apparent homogeneity, the yield of purified enzyme obtained using these methods was not suitable for crystallographic experiments.

Expression of both human α-L-fucosidases was attempted using *Escherichia coli*, *Pichia pastoris*, and cultured mammalian cells as expression vectors. The most success was observed when the GH29 enzymes were obtained by transient transfection of highly confluent human embryonic kidney cells. For FUCA1 at least, the expression yield and purity after a two-step purification procedure leaves the author optimistic that a structural solution to this clinically important enzyme is on the horizon.

The human α-ʟ-fucosidases are each clinically important, however it is currently unknown to what extent the GH29 inhibitors described in the literature inhibit each enzyme, or what the structural features which may allow more selective inhibition of each enzyme may be. This may be clinically important however as, *e.g.*, whilst inhibition of FUCA2 may allow clearance of *Helicobacter pylori* from the stomach, the inhibition of FUCA1 *in vivo* is likely to cause a number of side-effects due to the importance of the enzyme in glycan catabolism. The availability of crystal structures for these enzymes could potentially inform the design of selective inhibitors for each human α-ʟ-fucosidase.

Appendix 1: Alignment of α-ʟ-fucosidase sequences from various organisms

```
              1         10        20        30        40        50        60
1. Bt2970    MKNTIIYRMK SRLITLLLLL ITVSSVT-FA QEAKKEIPLK YGATNEGKRQ DPAMQKFRDN
2. TmGH29                          MISMKPR YKPDWESLRE HTVPKWFDKA
3. HsFUCA1   MRSR-PAGPA LLLLLLFLGA AESVRR---- ----AQPPRR YTPDWPSLDS RPLPAWFDEA
4. HsFUCA2   MRPQELPRLA FPLLLLLLLL LPPPPCP--- ----AHSATR FDPTWESLDA RQLPAWFDQA
5. BtaGH29   MRSW-VVGAR LLLLLQLVLV LGAVRLPPCT DPRHCTDPPR YTPDWPSLDS RPLPAWFDEA

              70        80        90        100       110       120
1. Bt2970    RLGAFIHWGL YAIPGG---- ---------- ---------- EWNGKVYGG- ---AAEWLKS
2. TmGH29    KFGIFIHWGI YSVP--GWAT PTGELGKVPM DAWFFQNPYA EWYENSLRIK ESPTWEYHVK
3. HsFUCA1   KFGVFIHWGV FSVP--AWGS ---------- ---------- EWFWWHWQGE GRPQYQRFMR
4. HsFUCA2   KFGIFIHWGV FSVP--SFGS ---------- ---------- EWFWWYWQKE KIPKYVEFMK
5. BtaGH29   KFGVFVHWGV FSVP--AWGS ---------- ---------- EWFWWHWQGE KLPQYESFMK

              130       140       150       160       170       180
1. Bt2970    WAKVPADEWL KLMDQWNPTK FDAKKWAKMA KEMGTKYVKI TTKHHEGFCL WPSKYTK-YT
2. TmGH29    T-YGENFEYE KFADLFTAEK WDPQEWADLF KKAGAKYVIP TTKHHDGFCL WGIKYTD-FN
3. HsFUCA1   DNYPPGFSYA DFGPQFTARF FHPEEWADLF QAAGAKYVVL TTKHHEGFTN WPSPVSWNWN
4. HsFUCA2   DNYPPSFKYE DFGPLFTAKF FNANQWADIF QASGAKYIVL TSKHHEGFTL WGSEYSWNWN
5. BtaGH29   ENYPPDFSYA DFGPRFTARF FNPDSWADLF KAAGAKYVVL TTKHHEGYTN WPSPVSWNWN

              190       200       210       220       230       240
1. Bt2970    VANTPYKRDI LGELVKAYND E-GIDVHFYF S-VMDWSNPD --YRYDIKSK EDSIAFSRFL
2. TmGH29    SVKRGPKRDL VGDLAKAVRE A-GLRFGVYY SGGLDWRFTT EPIRYPEDLS YIRPNTYEYA
3. HsFUCA1   SKDVGPHRDL VGELGTALRK R-NIRYGLYH S-LLEWFHP- ---LYLLDKK NGFKTQHFVS
4. HsFUCA2   AIDEGPKRDI VKELEVAIRN RTDLRFGLYY S-LFEWFHP- ---LFLEDES SSFHKRQFPV
5. BtaGH29   SKDVGPHRDL VGELGTAIRK R-NIRYGLYH S-LLEWFHP- ---LYLRDKK NGFKTQYFVN

              250       260       270       280       290       300
1. Bt2970    EFTDNQLKEL ATRYPTVKDF WFDGTWDASV KKNGWWTAHA EQMLKELVPG VAINSRLRAD
2. TmGH29    DYAYKQVMEL VDLY-LPDVL WNDMGWPE-K GKE-----DL KYLFAYYYNK HPEG---SVN
3. HsFUCA1   AKTMPELYDL VNSY-KPDLI WSDGEWECPD TYW-----NS TNFLSWLYND SPVKDEVVVN
4. HsFUCA2   SKTLPELYEL VNNY-QPEVL WSDGDGGAPD QYW-----NS TGFLAWLYND SPVRGTVVTN
5. BtaGH29   AKTMPELYDL VNRY-KPDLI WSDGEWECPD TYW-----NS TDFLAWLYND SPVKDEVVVN

              310       320       330       340       350       360
1. Bt2970    DK-GKRHFDS NGRLMGDYES GYERRLPDPV KDLKVTQWDW EACMTIPENQ WGYHKDWSLS
2. TmGH29    DRWGVPHWDF KTAEYHVN-- -YPGDLPG-- ------YKW EFTRGIG-LS FGYNRNEGPE
3. HsFUCA1   DRWGQNCSCH HGGYYNCEDK FKPQSLPD-- -------HKW EMCTSIDKFS WGYRRDMALS
4. HsFUCA2   DRWGAGSICK HGGFYTCSDR YNPGHLLP-- -------HKW ENCMTIDKLS WGYRREAGIS
5. BtaGH29   DRWGQNCSCH HGGYYNCKDK FQPETLPD-- -------HKW EMCTSIDQRS WGYRRDMEMA

              370       380       390       400       410       420
1. Bt2970    YVKTPIEVID RIVHAVSMGG NMVVNFGPQA DGDFRPEEKA MATAIGKWMN RYGKAVYACD
2. TmGH29    HMLSVEQLVY TLVDVVSKGG NLLLNVGPKG DGTIPDLQKE RLLGLGEWLR KYGDAIYGTS
3. HsFUCA1   DVTEESEIIS ELVQTVSLGG NYLLNIGPTK DGLIVPIFQE RLLAVGKWLS INGEAIYASK
4. HsFUCA2   DYLTIEELVK QLVETVSCGG NLLMNIGPTL DGTISVVFEE RLRQMGSWLK VNGEAIYETH
5. BtaGH29   DITNESTIIS ELVQTVSLGG NYLLNVGPTK DGLIVPIFQE RLLAVGKWLS INGEAIYASK

              430       440       450       460       470       480
1. Bt2970    YAG--FEKQD WGYYTRGK-- NDEVYMVVFN QPYSERLIVK TPKGIT-VEK ATLLTTGEDI
2. TmGH29    VWERCCAKTE DGTEIRFTRK CNRIFVIFLG IPTGEKIVIE DLNLSAGTVR HFLTGERLSF
3. HsFUCA1   PWRVQWEKNT TSVWYTSK-- GSAVYAIFLH WPENGVLNLE SPITTS-TTK ITMLGIQGDL
4. HsFUCA2   TWRSQNDTVT PDVWYTSKPK EKLVYAIFLK WPTSGQLFLG HPKAILGATE VKILGHGQPL
5. BtaGH29   PWRVQSEKN- -SVWYTSK-- GLAVYAILLH WPEYGILSLI SPIATS-TTK VTMLGIQKDL

              490       500       510       520       525
1. Bt2970    TVVETTR-NE YNVSVPKKNP GEPYVIQLKV RAAKGTKSIY RDALT
2. TmGH29    KN-------- VGKNLEITVP KKLLETDSIT LVLEAVEE
3. HsFUCA1   KWSTDPD-KG LFISLPQLPP SAVPAEFAWT IKLTGVK
4. HsFUCA2   NWISLEQ-NG IMVELPQLTI HQMPCKWGWA LALTNVI
5. BtaGH29   KWSLNPSGKG LLVFLPQLPP AALPTEFAWT IKLTGVK
```

Appendix 1.1. Alignment of α-ʟ-fucosidase protein sequences. Protein sequences of *Bacteroides thetaiotaomicron* 2970 (Bt2970), *Thermotoga maritima* GH29 (TmGH29), *Homo sapiens* FUCA1 (HsFUCA1) and FUCA2 (HsFUCA2) and *Bos Taurus* GH29 (BtaGH29) are aligned. Amino acids are coloured according to sequence similarity; 100% similarity in green, 80-10% in light green, 60-80% in yellow. Alignment was performed using Geneious R8 (http://www.geneious.com, Kearse et al., 2012).[357]

Appendix 2: *Bt*Fuc2970 gene sequence and transcript in pET-YSBLIC3C vector

A2.1 BtFuc2970 gene sequence in pET-YSBLIC3C

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGAAGTTCTGT
TCCAGGGACCAGAAGCAAAGAAGGAAATTCCTTTGAAATATGGAGCTACGA
ATGAAGGTAAACGGCAAGACCCTGCCATGCAGAAGTTCCGTGACAATCGTTT
GGGTGCCTTCATCCACTGGGGACTGTATGCTATCCCCGGGGGAGAGTGGAAT
GGAAAAGTATATGGCGGAGCTGCCGAATGGCTGAAAAGCTGGGCAAAAGTA
CCTGCTGATGAATGGCTGAAACTGATGGATCAATGGAACCCTACGAAATTTG
ATGCGAAGAAATGGGCAAAGATGGCCAAAGAAATGGGTACTAAGTATGTCA
AGATTACAACGAAACATCATGAAGGCTTCTGTCTGTGGCCTAGTAAGTATAC
TAAATATACCGTAGCAAATACCCCATATAAGCGTGATATATTGGGCGAGTTG
GTGAAAGCCTATAATGACGAAGGAATTGATGTACACTTCTATTTCTCAGTGA
TGGACTGGAGTAATCCGGATTATCGTTATGATATAAAATCCAAAGAAGATAG
CATCGCCTTCAGCCGTTTCCTTGAATTTACCGACAATCAACTGAAAGAACTG
GCAACACGTTACCCGACCGTTAAGGACTTCTGGTTTGATGGTACGTGGGATG
CCAGCGTTAAAAAGAATGGTTGGTGGACAGCTCATGCAGAACAAATGTTGA
AGGAACTCGTTCCGGGTGTTGCCATCAATAGCCGCTTACGTGCTGATGACAA
AGGAAAGCGACATTTTGATAGCAATGGTCGTCTGATGGGTGACTACGAATCC
GGCTACGAACGCCGCTTGCCCGATCCGGTGAAAGATCTCAAAGTTACACAGT
GGGACTGGGAAGCCTGCATGACTATACCCGAAAATCAATGGGGATATCACA
AAGACTGGTCATTGAGCTATGTGAAAACTCCGATTGAAGTCATTGACCGCAT
TGTACACGCTGTTTCCATGGGTGGAAACATGGTTGTCAACTTCGGGCCTCAG
GCAGATGGTGATTTCCGTCCCGAAGAGAAAGCAATGGCTACAGCGATTGGTA
AGTGGATGAATCGTTACGGAAAAGCTGTTTATGCTTGCGATTATGCCGGATT
TGAAAAACAAGACTGGGGATATTATACACGTGGTAAAAACGATGAAGTTTAT
ATGGTAGTATTCAATCAGCCTTATAGTGAACGGTTGATTGTAAAGACTCCGA
AAGGCATTACAGTAGAAAAGCCACTTTGCTGACTACCGGTGAAGATATCAC
TGTTGTTGAGACAACCCGCAATGAATATAACGTATCTGTTCCTAAAAAGAAT
CCGGGTGAACCTTATGTAATTCAGCTTAAAGTTCGTGCAGCTAAAGGAACAA
AAAGTATTTATCGAGATGCTTTAACATAA

167

A2.2 BtFuc2970 transcript in pET-YSBLIC3C

MGSHHHHHHSSGLEVLFQGPAEAKKEIPLKYGATNEGKRQDPAMQKFRDNRLG
AFIHWGLYAIPGGEWNGKVYGGAAEWLKSWAKVPADEWLKLMDQWNPTKFD
AKKWAKMAKEMGTKYVKITTKHHEGFCLWPSKYTKYTVANTPYKRDILGELV
KAYNDEGIDVHFYFSVMDWSNPDYRYDIKSKEDSIAFSRFLEFTDNQLKELATR
YPTVKDFWFDGTWDASVKKNGWWTAHAEQMLKELVPGVAINSRLRADDKGK
RHFDSNGRLMGDYESGYERRLPDPVKDLKVTQWDWEACMTIPENQWGYHKD
WSLSYVKTPIEVIDRIVHAVSMGGNMVVNFGPQADGDFRPEEKAMATAIGKWM
NRYGKAVYACDYAGFEKQDWGYYTRGKNDEVYMVVFNQPYSERLIVKTPKGI
TVEKATLLTTGEDITVVETTRNEYNVSVPKKNPGEPYVIQLKVRAAKGTKSIYR
DALT

Appendix 3: FUCA1 and FUCA2 gene sequences and transcripts in pET30a$^+$ vector

A3.1 FUCA1 gene sequence in pET30a$^+$

ATGCACCATCATCATCATCATTCTTCTGGTCTGGTGCCACGCGGTTCTGGTAT
GAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCAGA
TCTGGGTACCGACGACGACGACAAGGCCATGGGCGGTGCCGCCGAATCAGT
GCGTCGTGCCCAACCGCCGCGTCGCTACACCCCGGATTGGCCGTCGCTGGAT
AGTCGCCCGCTGCCGGCTTGGTTTGATGAAGCAAAATTTGGCGTGTTCATTC
ATTGGGGTGTCTTTAGCGTGCCGGCTTGGGGCTCTGAATGGTTCTGGTGGCA
CTGGCAGGGTGAAGGTCGTCCGCAGTATCAACGTTTTATGCGCGATAACTAT
CCGCCGGGCTTCAGCTACGCAGACTTTGGTCCGCAGTTCACCGCTCGCTTTTT
CCATCCGGAAGAATGGGCCGACCTGTTTCAAGCCGCGGGTGCAAAATACGTG
GTTCTGACCACGAAACATCACGAAGGTTTCACCAACTGGCCGAGCCCGGTTT
CTTGGAACTGGAATTCCAAAGATGTGGGTCCGCATCGTGACCTGGTTGGCGA
ACTGGGTACGGCGCTGCGTAAACGCAATATTCGCTATGGCCTGTACCATTCT
CTGCTGGAATGGTTTCACCCGCTGTATCTGCTGGATAAGAAAAACGGTTTTA
AAACCCAGCACTTCGTTAGTGCCAAAACGATGCCGGAACTGTATGATCTGGT
CAATAGTTACAAACCGGATCTGATCTGGTCCGACGGCGAATGGGAATGCCCG
GACACCTATTGGAACAGCACGAATTTCCTGTCTTGGCTGTACAACGATAGTC
CGGTGAAAGACGAAGTCGTGGTTAACGATCGTTGGGGTCAGAATTGCTCCTG
TCATCACGGCGGTTACTACAACTGCGAAGATAAATTCAAACCGCAATCACTG
CCGGACCATAAATGGGAAATGTGTACCAGTATTGACAAATTCTCCTGGGGCT
ATCGTCGCGATATGGCACTGTCGGACGTTACCGAAGAATCAGAAATTATCTC
GGAACTGGTTCAGACGGTCAGCCTGGGCGGTAACTACCTGCTGAATATCGGC
CCGACCAAAGATGGTCTGATTGTCCCGATCTTTCAAGAACGTCTGCTGGCCG
TGGGCAAATGGCTGTCAATTAACGGTGAAGCAATCTATGCTTCGAAACCGTG
GCGCGTCCAGTGGGAGAAAAACACCACGTCTGTGTGGTATACCAGTAAAGG
CTCCGCGGTTTACGCCATTTTTCTGCACTGGCCGGAAAACGGTGTGCTGAATC
TGGAATCACCGATCACCACGTCGACCACGAAAATTACCATGCTGGGCATCCA
GGGTGATCTGAAATGGAGCACGGATCCGGACAAAGGTCTGTTCATCTCTCTG
CCGCAACTGCCGCCGAGTGCTGTTCCGGCAGAATTTGCTTGGACGATTAAAC
TGACGGGTGTGAAATGA

A3.2 *FUCA1 transcript in pET30a*[+]

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMGGAAES
VRRAQPPRRYTPDWPSLDSRPLPAWFDEAKFGVFIHWGVFSVPAWGSEWFWW
HWQGEGRPQYQRFMRDNYPPGFSYADFGPQFTARFFHPEEWADLFQAAGAKY
VVLTTKHHEGFTNWPSPVSWNWNSKDVGPHRDLVGELGTALRKRNIRYGLYHS
LLEWFHPLYLLDKKNGFKTQHFVSAKTMPELYDLVNSYKPDLIWSDGEWECPD
TYWNSTNFLSWLYNDSPVKDEVVVNDRWGQNCSCHHGGYYNCEDKFKPQSLP
DHKWEMCTSIDKFSWGYRRDMALSDVTEESEIISELVQTVSLGGNYLLNIGPTK
DGLIVPIFQERLLAVGKWLSINGEAIYASKPWRVQWEKNTTSVWYTSKGSAVY
AIFLHWPENGVLNLESPITTSTTKITMLGIQGDLKWSTDPDKGLFISLPQLPPSAV
PAEFAWTIKLTGVK

A3.3 FUCA2 gene sequence in pET30a$^+$

ATGCACCATCATCATCATCATTCTTCTGGTCTGGTGCCACGCGGTTCTGGTAT
GAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCAGA
TCTGGGTACCGACGACGACGACAAGGCCATGGGCGCTCACTCAGCGACCCGT
TTCGACCCGACCTGGGAATCACTGGATGCGCGTCAACTGCCGGCGTGGTTCG
ATCAAGCAAAATTTGGCATTTTTATCCATTGGGGCGTGTTTAGCGTCCCGTCA
TTCGGTTCGGAATGGTTTTGGTGGTATTGGCAGAAAGAAAAAATCCCGAAAT
ACGTGGAATTCATGAAAGATAACTATCCGCCGAGCTTTAAATACGAAGACTT
TGGTCCGCTGTTCACCGCGAAATTTTTCAACGCAAATCAGTGGGCTGATATCT
TCCAAGCGAGTGGCGCCAAATATATTGTTCTGACCTCCAAACATCACGAAGG
CTTTACGCTGTGGGGTAGCGAATATTCTTGGAACTGGAATGCGATTGATGAA
GGTCCGAAACGTGACATCGTTAAAGAACTGGAAGTCGCCATTCGTAATCGCA
CCGATCTGCGCTTCGGCCTGTATTACTCTCTGTTTGAATGGTTCCATCCGCTG
TTTCTGGAAGACGAAAGCTCTAGTTTCCACAAACGTCAGTTTCCTGTGAGTA
AAACCCTGCCGGAACTGTATGAACTGGTGAACAATTACCAACCGGAAGTTCT
GTGGAGCGATGGTGATGGCGGTGCACCGGATCAGTATTGGAACAGCACGGG
TTTCCTGGCTTGGCTGTACAATGAATCTCCGGTTCGTGGCACCGTGGTTACGA
ACGATCGCTGGGGCGCGGGTAGTATCTGCAAACATGGCGGTTTTTATACCTG
TTCCGACCGCTACAACCCGGGTCATCTGCTGCCGCACAAATGGGAAAATTGC
ATGACGATTGATAAACTGTCATGGGGCTATCGTCGCGAAGCCGGTATTTCGG
ACTACCTGACCATCGAAGAACTGGTGAAACAACTGGTGGAAACGGTTAGCTG
TGGCGGTAACCTGCTGATGAATATCGGCCCGACCCTGGATGGTACGATTTCA
GTCGTGTTTGAAGAACGTCTGCGCCAAATGGGCTCGTGGCTGAAAGTTAACG
GTGAAGCAATTTATGAAACCCACACGTGGCGTAGCCAGAATGATACCGTCAC
GCCGGACGTGTGGTATACCTCTAAACCGAAAGAAAAACTGGTCTACGCTATC
TTCCTGAAATGGCCGACCAGTGGCCAGCTGTTTCTGGGTCATCCGAAAGCGA
TTCTGGGTGCCACGGAAGTCAAACTGCTGGGCCACGGTCAGCCGCTGAACTG
GATTAGCCTGGAACAAAATGGCATCATGGTGGAACTGCCGCAGCTGACGATT
CATCAAATGCCGTGTAAATGGGGCTGGGCTCTGGCTCTGACCAATGTCATCT
AA

A3.4 *FUCA2 transcript in pET30a$^+$*

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMGAHSAT
RFDPTWESLDARQLPAWFDQAKFGIFIHWGVFSVPSFGSEWFWWYWQKEKIPK
YVEFMKDNYPPSFKYEDFGPLFTAKFFNANQWADIFQASGAKYIVLTSKHHEGF
TLWGSEYSWNWNAIDEGPKRDIVKELEVAIRNRTDLRFGLYYSLFEWFHPLFLE
DESSSFHKRQFPVSKTLPELYELVNNYQPEVLWSDGDGGAPDQYWNSTGFLAW
LYNESPVRGTVVTNDRWGAGSICKHGGFYTCSDRYNPGHLLPHKWENCMTIDK
LSWGYRREAGISDYLTIEELVKQLVETVSCGGNLLMNIGPTLDGTISVVFEERLR
QMGSWLKVNGEAIYETHTWRSQNDTVTPDVWYTSKPKEKLVYAIFLKWPTSG
QLFLGHPKAILGATEVKLLGHGQPLNWISLEQNGIMVELPQLTIHQMPCKWGW
ALALTNVI

Appendix 4: Generation of FUCA1 constructs with minimal N-terminal octahistidine affinity tags

A4.1 Oligonucleotide primers for generation of plasmids with only N-terminal octahistidine affinity tags

8His_FUCA1

Forward primer   CAGCCTCCGCGCCGCTAC

Reverse primer   GTGATGATGGTGATGGTGGTG

8HisSG_FUCA1

Forward primer   CAGCCTCCGCGCCGCTAC

Reverse primer   GCCGCTGTGATGATGGTGATGGTGGTG

8HisSGSG_FUCA1

Forward primer   CAGCCTCCGCGCCGCTAC

Reverse primer   GCCGCTGCCGCTGTGATGATGGTGATGGTGGTG

Abbreviations

µg	Microgram
µL	Microlitre
µM	Micromolar
µmol	Micromole
A. thaliana	*Arabidopsis thaliana*
ABP	Activity-based probe
ABPP	Activity-based protein profiling
ADAM	A disintegrin and metalloproteinase
Arg	Arginine
Asn	Asparigine
Asp	Aspartic acid
Bi	*Bifidobacterium longum subsp. Infantis*
BODIPY	Boron-dipyrromethene
BSA	Bovine serum albumin
Bt	*Bacteroides thetaiotaomicron*
C. jejuni	*Campylobacter jejuni*
C	Celsius
CAZy	Carbohydrate Active Enzymes database
cDNA	Chromosomal deoxyribonucleic acid
CHO	Chinese Hamster Ovary (cell strain)
CMV	Cytomegalovirus
CNP	2-chloro,4-nitrophenol
CNP-fucoside	2-chloro,4-nitrophenyl-α-ʟ-fucopyranoside

Da	Dalton
DFJ	Deoxyfuconojirimycin
DLS	Diamond Light Source
DMEM	Dulbecco's Modified Eagle's Medium
DMJ	Deoxymannonojirimycin
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EndoH	Endoglycosidase H
ESRF	European Synchrotron Research Facility
FBS	Foetal Bovine Serum
Fg	*Fuasrium graminearum*
FPLC	Fast protein liquid chromatography
Fuc	Fucose
FUCA1	α-ʟ-fucosidase 1 (*Homo sapiens*)
FUCA2	α-ʟ-fucosidase 2 (*Homo sapiens*)
FucCS	Fucosylated chondroitin sulfate
fucR	ʟ-Fucose operon activator
FUT1-11	GDP-α-ʟ-fucosyltransferases 1-11
Gal	Galactose
GalNAc	N-acetyl-galactosamine
GDP	Guanine disphosphate
GH	Glycoside hydrolase family
GH29-A	Subfamily A from glycoside hydrolase family 29
GH29-B	Subfamily B from glycoside hydrolase family 29
GlcNAc	N-acetyl-glucosamine

Gln	Glutamine
Glu	Glutamic acid
GTP	Guanine triphosphate
h	Hours
HEK	Human embryonic kidney (cell strain)
H. pylori	*Helicobacter pylori*
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HTCS	Hybrid two-component sensor/regulator
IPTG	Isopropyl β-ᴅ-1-thiogalactopyranoside
K_a	Acid dissociation constant
kDa	Kilodalton
K_i	Inhibition constant
K_M	Michaelis constant
L	Litre
LADII	Leukocyte adhesion deficiency type II
LB	Lysogeny Broth
m	Minutes
M	Molar
MES	2-(N-morpholino)ethanesulfonic acid
mg	Milligram
mL	Millilitre
mM	Millimolar
mPES	Modified polyethersulfone
mRNA	Messenger ribonucleic acid

MUC1	Cell surface-associated mucin 1
nL	Nanolitre
nm	Nanometre
nM	Nanomolar
P. aeruginosa	*Pseudomonas aeruginosa*
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PDB	Protein databank
PEG	Polyethylene glycol
PEI	Polethylenimine
pelB	Pectate lyase B
PL	Polysaccharide lyase family
PNGase	Peptide-N-glycosidase
POFUT1-2	Protein-O-fucosyltransferases 1-2
PUL	Polysaccharide utilisation locus
RCF	Relative centrifugal force
RCSB	Research Collaboratory for Structural Bioinformatics
RMSD	Root-mean-square deviation
ROS	Reactive oxygen species
RPM	Revolutions per minute
s	Seconds
SDS	Sodium dodecyl sulfate
Ser	Serine
Sia	Sialic acid
SUMO	Small ubiquitin-like modifier

Sus	Starch utilisation system
TE	Trypsin-EDTA 1x working stock, Life Technologies
TEV	Tobacco etch virus
Thr	Threonine
TIM	Triosephosphate isomerase
Tm	*Thermotoga maritima*
Tris	Tris(hydroxymethyl)aminomethane
Trp	Tryptophan
V_{max}	Maxmial enzyme velocity
w/v	Weight / volume
Xyl	Xylose

References

1   Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Bba-Gen Subjects* 1473, 4-8 (1999).

2   Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12, 43r-56r (2002).

3   Michalski, J. C. & Klein, A. Glycoprotein lysosomal storage disorders: alpha- and beta-mannosidosis, fucosidosis and alpha-N-acetylgalactosaminidase deficiency. *Biochim Biophys Acta* 1455, 69-84 (1999).

4   Liu, T. W. *et al.* Role for alpha-L-fucosidase in the control of Helicobacter pylori-infected gastric cancer cells. *P Natl Acad Sci USA* 106, 14581-14586 (2009).

5   Day, C. J. *et al.* Differential Carbohydrate Recognition by Campylobacter jejuni Strain 11168: Influences of Temperature and Growth Conditions. *Plos One* 4, e4927 (2009).

6   Scharfman, A. *et al.* Pseudomonas aeruginosa binds to neoglycoconjugates bearing mucin carbohydrate determinants and predominantly to sialyl-Lewis x conjugates. *Glycobiology* 9, 757-764 (1999).

7   Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem* 77, 521-555 (2008).

8   Honke, K. & Taniguchi, N. Sulfotransferases and sulfated oligosaccharides. *Med Res Rev* 22, 637-654 (2002).

9   Ishai-Michaeli, R. *et al.* Importance of size and sulfation of heparin in release of basic fibroblast growth factor from the vascular endothelium and extracellular matrix. *Biochemistry-Us* 31, 2080-2088 (1992).

10   Singh, A. *et al.* The Interaction of Heparin Tetrasaccharides with Chemokine CCL5 Is Modulated by Sulfation Pattern and pH. *J Biol Chem* 290, 15421-15436 (2015).

11   Raman, K., Mencio, C., Desai, U. R. & Kuberan, B. Sulfation patterns determine cellular internalization of heparin-like polysaccharides. *Mol Pharm* 10, 1442-1449 (2013).

12   Ni, X., Canuel, M. & Morales, C. R. The sorting and trafficking of lysosomal proteins. *Histol Histopathol* 21, 899-913 (2006).

13   Wolfenden, R., Lu, X. D. & Young, G. Spontaneous hydrolysis of glycosides. *J Am Chem Soc* 120, 6814-6815 (1998).

14   Radzicka, A. & Wolfenden, R. A Proficient Enzyme. *Science* 267, 90-93 (1995).

15   Radzicka, A. & Wolfenden, R. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J Am Chem Soc* 118, 6105-6109 (1996).

16   Jequier, E. Carbohydrates as a Source of Energy. *Am J Clin Nutr* 59, 682s-685s (1994).

17   Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunol* 1, 183-197 (2008).

18   Lederkremer, G. Z. Glycoprotein folding, quality control and ER-associated degradation. *Curr Opin Struc Biol* 19, 515-523 (2009).

19   Hakomori, S. The glycosynapse. *P Natl Acad Sci USA* 99, 225-232 (2002).

20   Becker, D. J. & Lowe, J. B. Fucose: biosynthesis and biological function in mammals. *Glycobiology* 13, 41r-53r (2003).

21   Li, B., Lu, F., Wei, X. & Zhao, R. Fucoidan: structure and bioactivity. *Molecules* 13, 1671-1695 (2008).

22   Pomin, V. H. Holothurian Fucosylated Chondroitin Sulfate. *Marine Drugs* 12, 232-254 (2014).

23   Hoffman, M. *et al.* Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. *Carbohydr Res* 340, 1826-1840 (2005).

24   Vieira, R. P. & Mourao, P. A. S. Occurrence of a Unique Fucose-Branched Chondroitin Sulfate in the Body Wall of a Sea-Cucumber. *J Biol Chem* 263, 18176-18183 (1988).

25   Vieira, R. P., Mulloy, B. & Mourao, P. A. S. Structure of a Fucose-Branched Chondroitin Sulfate from Sea-Cucumber - Evidence for the Presence of 3-O-Sulfo-Beta-D-Glucuronosyl Residues. *J Biol Chem* 266, 13530-13536 (1991).

26   Vieira, R. P., Pedrosa, C. & Mourao, P. A. S. Extensive Heterogeneity of Proteoglycans Bearing Fucose-Branched Chondroitin Sulfate Extracted from the Connective-Tissue of Sea-Cucumber. *Biochemistry-Us* 32, 2254-2262 (1993).

27   Mourao, P. A. S. *et al.* Structure and anticoagulant activity of a fucosylated chondroitin sulfate from echinoderm - Sulfated fucose branches on the polysaccharide account for its high anticoagulant action. *J Biol Chem* 271, 23973-23984 (1996).

28   Mourao, P. A. S., Guimaraes, M. A. M., Mulloy, B., Thomas, S. & Gray, E. Antithrombotic activity of a fucosylated chondroitin sulphate from echinoderm: sulphated fucose branches on the polysaccharide account for its antithrombotic action. *Brit J Haematol* 101, 647-652 (1998).

29   Borsig, L. *et al.* Selectin blocking activity of a fucosylated chondroitin sulfate glycosaminoglycan from sea cucumber - Effect on tumor metastasis and neutrophil recruitment. *J Biol Chem* 282, 14984-14991 (2007).

30   Hu, S. W. *et al.* Fucosylated chondroitin sulfate from Acaudina molpadioides improves hyperglycemia via activation of PKB/GLUT4 signaling in skeletal muscle of insulin resistant mice. *Food Funct* 4, 1639-1646 (2013).

31   Ale, M. T. & Meyer, A. S. Fucoidans from brown seaweeds: an update on structures, extraction techniques and use of enzymes as tools for structural elucidation. *Rsc Adv* 3, 8131-8141 (2013).

32   Cumashi, A. *et al.* A comparative study of the anti-inflammatory, anticoagulant, antiangiogenic, and antiadhesive activities of nine different fucoidans from brown seaweeds. *Glycobiology* 17, 541-552 (2007).

33   Wang, J., Zhang, Q., Zhang, Z., Song, H. & Li, P. Potential antioxidant and anticoagulant capacity of low molecular weight fucoidan fractions extracted from Laminaria japonica. *Int J Biol Macromol* 46, 6-12 (2010).

34   Ale, M. T., Maruyama, H., Tamauchi, H., Mikkelsen, J. D. & Meyer, A. S. Fucoidan from Sargassum sp and Fucus vesiculosus reduces cell viability of lung carcinoma and melanoma cells in vitro and activates natural killer cells in mice in vivo. *Int J Biol Macromol* 49, 331-336 (2011).

35   Huang, T. T. F., Ohzu, E. & Yanagimachi, R. Evidence Suggesting That L-Fucose Is Part of a Recognition Signal for Sperm-Zona Pellucida Attachment in Mammals. *Gamete Res* 5, 355-361 (1982).

36   Ridley, B. L., O'Neill, M. A. & Mohnen, D. A. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* 57, 929-967 (2001).

37   Fry, S. C. *et al.* An Unambiguous Nomenclature for Xyloglucan-Derived Oligosaccharides. *Physiol Plantarum* 89, 1-3 (1993).

38   Levy, S., Maclachlan, G. & Staehelin, L. A. Xyloglucan sidechains modulate binding to cellulose during in vitro binding assays as predicted by conformational dynamics simulations. *Plant J* 11, 373-386 (1997).

39   Reiter, W. D., Chapple, C. C. S. & Somerville, C. R. Altered Growth and Cell-Walls in a Fucose-Deficient Mutant of Arabidopsis. *Science* 261, 1032-1035 (1993).

40   Vanzin, G. F. *et al.* The mur2 mutant of Arabidopsis thaliana lacks fucosylated xyloglucan because of a lesion in fucosyltransferase AtFUT1. *Proc Natl Acad Sci U S A* 99, 3340-3345 (2002).

41   Zablackis, E. *et al.* Substitution of L-fucose by L-galactose in cell walls of Arabidopsis mur1. *Science* 272, 1808-1810 (1996).

42   Madson, M. *et al.* The MUR3 gene of Arabidopsis encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. *Plant Cell* 15, 1662-1670 (2003).

43   Pena, M. J., Ryden, P., Madson, M., Smith, A. C. & Carpita, N. C. The galactose residues of xyloglucan are essential to maintain mechanical strength of the primary cell walls in Arabidopsis during growth. *Plant Physiol* 134, 443-451 (2004).

44   Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nature Reviews Immunology* 7, 678-689 (2007).

45   Langer, H. F. & Chavakis, T. Leukocyte-endothelial interactions in inflammation. *J Cell Mol Med* 13, 1211-1220 (2009).

46   Hooper, L. V. & Gordon, J. I. Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. *Glycobiology* 11, 1R-10R (2001).

47   Eggens, I. *et al.* Specific Interaction between Lex and Lex Determinants - a Possible Basis for Cell Recognition in Preimplantation Embryos and in Embryonal Carcinoma-Cells. *J Biol Chem* 264, 9476-9484 (1989).

48   Zhu, Z. M., Kojima, N., Stroud, M. R., Hakomori, S. & Fenderson, B. A. Monoclonal antibody directed to Le(y) oligosaccharide inhibits implantation in the mouse. *Biol Reprod* 52, 903-912 (1995).

49   Hakomori, S. Glycosynapses: microdomains controlling carbohydrate-dependent cell adhesion and signaling. *An Acad Bras Cienc* 76, 553-572 (2004).

50   Lira-Navarrete, E. *et al.* Structural Insights into the Mechanism of Protein O-Fucosylation. *PLoS One* 6, e25365 (2011).

51   Varki, A. *et al.* Symbol nomenclature for glycan representation. *Proteomics* 9, 5398-5399 (2009).

52   Wang, Y. *et al.* Modification of epidermal growth factor-like repeats with O-fucose - Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase. *J Biol Chem* 276, 40338-40345 (2001).

53   Kopan, R. & Ilagan, M. X. G. The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. *Cell* 137, 216-233 (2009).

54   Li, L. H. *et al.* Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature Genetics* 16, 243-251 (1997).

55   Kamath, B. M. *et al.* Notch2 Mutations in Alagille Syndrome. *Journal of Hepatology* 54, S12-S13 (2011).

56   D'Souza, B., Miyamoto, A. & Weinmaster, G. The many facets of Notch ligands. *Oncogene* 27, 5148-5167 (2008).

57   Bray, S. J. Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Bio* 7, 678-689 (2006).

58   Kovall, R. A. More complicated than it looks: assembly of Notch pathway transcription complexes. *Oncogene* 27, 5099-5109 (2008).

59   Shao, L., Moloney, D. J. & Haltiwanger, R. Fringe modifies O-fucose on mouse Notch1 at epidermal growth factor-like repeats within the ligand-binding site and the abruptex region. *J Biol Chem* 278, 7775-7782 (2003).

60   Shao, L. & Haltiwanger, R. S. O-fucose modifications of epidermal growth factor-like repeats and thrombospondin type 1 repeats: unusual modifications in unusual places. *Cell Mol Life Sci* 60, 241-250 (2003).

61   Moloney, D. J. *et al.* Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem* 275, 9604-9611 (2000).

62   Chen, J. H., Moloney, D. J. & Stanley, P. Fringe modulation of Jagged1-induced Notch signaling requires the action of beta 4galactosyltransferase-1. *P Natl Acad Sci USA* 98, 13716-13721 (2001).

63   Shi, S. L. & Stanley, P. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *P Natl Acad Sci USA* 100, 5234-5239 (2003).

64   Okajima, T. & Irvine, K. D. Regulation of notch signaling by O-linked fucose. *Cell* 111, 893-904 (2002).

65   Springer, T. A. & Lasky, L. A. Cell adhesion. Sticky sugars for selectins. *Nature* 349, 196-197 (1991).

66   Maly, P. *et al.* The alpha(1,3)Fucosyltransferase Fuc-TVII controls leukocyte trafficking through an essential role in L-, E-, and P-selectin ligand biosynthesis. *Cell* 86, 643-653 (1996).

67   Ali, S. *et al.* Leukocyte extravasation: An immunoregulatory role for alpha-L-Fucosidase? *J Immunol* 181, 2407-2413 (2008).

68   Etzioni, A. *et al.* Recurrent Severe Infections Caused by a Novel Leukocyte Adhesion Deficiency. *New Engl J Med* 327, 1789-1792 (1992).

69   Lubke, T. *et al.* Complementation cloning identifies CDG-IIc, a new type of congenital disorders of glycosylation, as a GDP-fucose transporter deficiency. *Nature Genetics* 28, 73-76 (2001).

70   Luhn, K., Wild, M. K., Eckhardt, M., Gerardy-Schahn, R. & Vestweber, D. The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. *Nature Genetics* 28, 69-72 (2001).

71   Helmus, Y. *et al.* Leukocyte adhesion deficiency II patients with a dual defect of the GDP-fucose transporter. *Blood* 107, 3959-3966 (2006).

72   Jemal, A. *et al.* Global Cancer Statistics. *Ca-Cancer J Clin* 61, 69-90 (2011).

73   Christiansen, M. N. *et al.* Cell surface protein glycosylation in cancer. *Proteomics* 14, 525-546 (2014).

74   Butler, M. Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. *Cytotechnology* 50, 57-76 (2006).

75   Gazdar, A. F., Gao, B. N. & Minna, J. D. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? *Lung Cancer* 68, 309-318 (2010).

76   Houshdaran, S. *et al.* DNA Methylation Profiles of Ovarian Epithelial Carcinoma Tumors and Cell Lines. *Plos One* 5, e9359 (2010).

77   Gillet, J. P. *et al.* Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. *P Natl Acad Sci USA* 108, 18708-18713 (2011).

78   Potapenko, I. O. *et al.* Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Mol Oncol* 4, 98-118 (2010).

79   Blomme, B., Van Steenkiste, C., Callewaert, N. & Van Vlierberghe, H. Alteration of protein glycosylation in liver diseases. *Journal of Hepatology* 50, 592-603 (2009).

80   Miyoshi, E., Moriwaki, K. & Nakagawa, T. Biological function of fucosylation in cancer biology. *J Biochem-Tokyo* 143, 725-729 (2008).

81   Debruyne, E. N. & Delanghe, J. R. Diagnosing and monitoring hepatocellular carcinoma with alpha-fetoprotein: New aspects and applications. *Clin Chim Acta* 395, 19-26 (2008).

82   Wang, J. J. & Cao, E. H. Rapid kinetic rate assay of the serum alpha-L-fucosidase in patients with hepatocellular carcinoma by using a novel substrate. *Clin Chim Acta* 347, 103-109 (2004).

83   Fawzy Montaser, M., Amin Sakr, M. & Omar Khalifa, M. Alpha-l-fucosidase as a tumour marker of hepatocellular carcinoma. *Arab J Gastroenterol* 13, 9-13 (2012).

84   Matsumoto, K. *et al.* Identification of Predictive Biomarkers for Response to Trastuzumab Using Plasma FUCA Activity and N-Glycan Identified by MALDI-TOF-MS. *Journal of Proteome Research* 8, 457-462 (2009).

85   Nakagoe, T. *et al.* Expression of Lewis(a), sialyl Lewis(a), Lewis(x) and sialyl Lewis(x) antigens as prognostic factors in patients with colorectal cancer. *Can J Gastroenterol* 14, 753-760 (2000).

86   Ura, Y. *et al.* Quantitative Dot Blot Analyses of Blood-Group-Related Antigens in Paired Normal and Malignant Human Breast Tissues. *Int J Cancer* 50, 57-63 (1992).

87   Matsuura, N. *et al.* Gene expression of fucosyl- and sialyl-transferases which synthesize sialyl Lewis(x), the carbohydrate ligands for E-selectin, in human breast cancer. *Int J Oncol* 12, 1157-1164 (1998).

88   Yue, L. L., Yu, H. T., Zhang, C. J. & Liu, J. C. The Effect of Overexpression Of alpha (1,3)- upregulating VII on Adhesive Capability of Human Colon Carcinoma HT-29 Cells to HUVECs. *Adv Mater Res-Switz* 345, 250-256 (2012).

89   Aubert, M. *et al.* Peritoneal colonization by human pancreatic cancer cells is inhibited by antisense FUT3 sequence. *Int J Cancer* 88, 558-565 (2000).

90   Federici, M. F., Kudryashov, V., Saigo, P. E., Finstad, C. L. & Lloyd, K. O. Selection of carbohydrate antigens in human epithelial ovarian cancers as targets for immunotherapy: Serous and mucinous tumors exhibit distinctive patterns of expression. *Int J Cancer* 81, 193-198 (1999).

91   Liu, J. J. *et al.* Lewis(y) antigen stimulates the growth of ovarian cancer cells via regulation of the epidermal growth factor receptor pathway. *Oncol Rep* 23, 833-841 (2010).

92   Iwamori, M. *et al.* Alterations in the glycolipid composition and cellular properties of ovarian carcinoma-derived RMG-1 cells on transfection of the alpha 1,2-fucosyltransferase gene. *Cancer Sci* 96, 26-30 (2005).

93   Listinsky, J. J., Siegal, G. P. & Listinsky, C. M. The emerging importance of alpha-L-fucose in human breast cancer: a review. *Am J Transl Res* 3, 292-322 (2011).

94   Hutchinson, W. L., Du, M. Q., Johnson, P. J. & Williams, R. Fucosyltransferases - Differential Plasma and Tissue Alterations in Hepatocellular-Carcinoma and Cirrhosis. *Hepatology* 13, 683-688 (1991).

95   Laidler, P. *et al.* Characterization of glycosylation and adherent properties of melanoma cell lines. *Cancer Immunol Immun* 55, 112-118 (2006).

96   Yoshida, M. *et al.* Targeting anticancer drug delivery to pancreatic cancer cells using a fucose-bound nanoparticle approach. *PLoS One* 7, e39545 (2012).

97   Babiuch, K., Dag, A., Zhao, J., Lu, H. & Stenzel, M. H. Carbohydrate-specific Uptake of Fucosylated Polymeric Micelles by Different. *Biomacromolecules* (2015).

98   Abeylath, S. C., Ganta, S., Iyer, A. K. & Amiji, M. Combinatorial-Designed Multifunctional Polymeric Nanosystems for Tumor-Targeted Therapeutic Delivery. *Accounts Chem Res* 44, 1009-1017 (2011).

99   Cammarata, M., Vazzana, M., Chinnici, C. & Parrinello, N. A serum fucolectin isolated and characterized from sea bass Dicentrarchus labrax. *Bba-Gen Subjects* 1528, 196-202 (2001).

100   Wu, L. Q., Yang, X. Y., Duan, X. M., Cui, L. Z. & Li, G. C. Exogenous expression of marine lectins DlFBL and SpRBL induces cancer cell apoptosis possibly through PRMT5-E2F-1 pathway. *Sci Rep-Uk* 4, 4505 (2014).

101   Alhadeff, J. A. *et al.* Characterization of human semen alpha-L-fucosidases. *Mol Hum Reprod* 5, 809-815 (1999).

102   Phopin, K. *et al.* Roles of mouse sperm-associated alpha-L-fucosidases in fertilization. *Mol Reprod Dev* 80, 273-285 (2013).

103   Venditti, J. J., Swann, J. M. & Bean, B. S. Hamster Sperm-Associated Alpha-L-Fucosidase Functions During Fertilization. *Biol Reprod* 82, 572-579 (2010).

104   Focarelli, R., Cacace, M. G., Seraglia, R. & Rosati, F. A nonglycosylated, 68-kDa alpha-L-fucosidase is bound to the mollusc bivalve Unio elongatulus sperm plasma membrane and differs from a glycosylated 56-kDa form present in the seminal fluid. *Biochem Bioph Res Co* 234, 54-58 (1997).

105   Matsumoto, M., Hirata, J., Hirohashi, N. & Hoshi, M. Sperm-egg binding mediated by sperm alpha-L-fucosidase in the ascidian, Halocynthia roretzi. *Zool Sci* 19, 43-48 (2002).

106   Intra, J., Cenni, F. & Perotti, M. E. An alpha-L-fucosidase potentially involved in fertilization is present on Drosophila spermatozoa surface. *Mol Reprod Dev* 73, 1149-1158 (2006).

107   Winchester, B. *et al.* Inhibition of Alpha-L-Fucosidase by Derivatives of Deoxyfuconojirimycin and Deoxymannojirimycin. *Biochem J* 265, 277-282 (1990).

108   Oh, Y. S., Ahn, H. S. & Gye, M. C. Fucosyl neoglycoprotein binds to mouse epididymal spermatozoa and inhibits sperm. *Andrologia* 45, 363-368 (2013).

109   Pang, P. C. *et al.* Human Sperm Binding Is Mediated by the Sialyl-Lewis(x) Oligosaccharide on the Zona Pellucida. *Science* 333, 1761-1764 (2011).

110   Adlerberth, M. C., Isabelle Poilane, Agnes Wold, Anne Collignon, Ingegerd. Mechanisms of Colonisation and Colonisation Resistance of the Digestive Tract Part 1: Bacteria/host Interactions. *Microb Ecol Health D* 12, 223-239 (2000).

111   Zarepour, M. *et al.* The mucin Muc2 limits pathogen burdens and epithelial barrier dysfunction during Salmonella enterica serovar Typhimurium colitis. *Infect Immun* 81, 3672-3683 (2013).

112   McAuley, J. L. *et al.* MUC1 cell surface mucin is a critical element of the mucosal barrier to infection. *J Clin Invest* 117, 2313-2324 (2007).

113   Coker, A. O., Isokpehi Rd Fau - Thomas, B. N., Thomas Bn Fau - Amisu, K. O., Amisu Ko Fau - Obi, C. L. & Obi, C. L. Human campylobacteriosis in developing countries. *Emerg Infect Dis* 8, 237-244 (2002).

114   Hugdahl, M. B., Beery, J. T. & Doyle, M. P. Chemotactic Behavior of Campylobacter-Jejuni. *Infect Immun* 56, 1560-1566 (1988).

115   Stahl, M. *et al.* L-Fucose utilization provides Campylobacter jejuni with a competitive advantage. *P Natl Acad Sci USA* 108, 7194-7199 (2011).

116   Ruiz-Palacios, G. M., Cervantes, L. E., Ramos, P., Chavez-Munguia, B. & Newburg, D. S. Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J Biol Chem* 278, 14112-14120 (2003).

117   Garber, N., Guempel, U., Gilboagarber, N. & Doyle, R. J. Specificity of the Fucose-Binding Lectin of Pseudomonas-Aeruginosa. *Fems Microbiol Lett* 48, 331-334 (1987).

118   Mitchell, E. *et al.* Structural basis for oligosaccharide-mediated adhesion of Pseudomonas aeruginosa in the lungs of cystic fibrosis patients. *Nat Struct Biol* 9, 918-921 (2002).

119   Heneghan, M. A., McCarthy, C. F. & Moran, A. P. Relationship of blood group determinants on Helicobacter pylori lipopolysaccharide with host Lewis phenotype and inflammatory response. *Infect Immun* 68, 937-941 (2000).

120   Ofek, I., Hasty, D. L. & Sharon, N. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol Med Microbiol* 38, 181-191 (2003).

121   Nwosu, C. C. *et al.* Comparison of the Human and Bovine Milk N-Glycome via High-Performance Microfluidic Chip Liquid Chromatography and Tandem Mass Spectrometry. *Journal of Proteome Research* 11, 2912-2924 (2012).

122   Bode, L. Recent advances on structure, metabolism, and function of human milk oligosaccharides. *J Nutr* 136, 2127-2130 (2006).

123   Newburg, D. S., Ruiz-Palacios, G. M. & Morrow, A. L. Human milk glycans protect infants against enteric pathogens. *Annu Rev Nutr* 25, 37-58 (2005).

124   Newburg, D. S., Pickering, L. K., Mccluer, R. H. & Cleary, T. G. Fucosylated Oligosaccharides of Human-Milk Protect Suckling Mice from Heat-Stable Enterotoxin of Escherichia-Coli. *J Infect Dis* 162, 1075-1080 (1990).

125   Agrawal, B., Krantz, M. J., Parker, J. & Longenecker, B. M. Expression of MUC1 mucin on activated human T cells: Implications for a role of MUC1 in normal immune regulation. *Cancer Res* 58, 4079-4081 (1998).

126   Agrawal, B., Krantz, M. J., Reddish, M. A. & Longenecker, B. M. Cancer-associated MUC1 mucin inhibits human T-cell proliferation, which is reversible by IL-2. *Nature Medicine* 4, 43-49 (1998).

127   Handa, K., Jacobs, F., Longenecker, B. M. & Hakomori, S. Association of MUC-1 and PSGL-1 with low-density microdomain in T-lymphocytes: A preliminary note. *Biochem Bioph Res Co* 285, 788-794 (2001).

128   Savage, D. C. Microbial Ecology of Gastrointestinal-Tract. *Annu Rev Microbiol* 31, 107-133 (1977).

129   Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* 6, 121-131 (2008).

130   Salyers, A. A., Vercellotti, J. R., West, S. E. H. & Wilkins, T. D. Fermentation of Mucin and Plant Polysaccharides by Strains of Bacteroides from Human Colon. *Appl Environ Microb* 33, 319-322 (1977).

131   Mcneil, N. I. The Contribution of the Large-Intestine to Energy Supplies in Man. *Am J Clin Nutr* 39, 338-342 (1984).

132   Moore, W. E. C. & Holdeman, L. V. Human Fecal Flora - Normal Flora of 20 Japanese-Hawaiians. *Appl Microbiol* 27, 961-979 (1974).

133   Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *P Natl Acad Sci USA* 99, 15451-15455 (2002).

134   Hooper, L. V., Stappenbeck, T. S., Hong, C. V. & Gordon, J. I. Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol* 4, 269-273 (2003).

135   Xu, J. & Gordon, J. I. Honor thy symbionts. *P Natl Acad Sci USA* 100, 10452-10459 (2003).

136   Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273, 1380-1383 (1996).

137   Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *J Biol Chem* 284, 24673-24677 (2009).

138   D'Elia, J. N. & Salyers, A. A. Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *Journal of Bacteriology* 178, 7173-7179 (1996).

139   Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host & Microbe* 4, 447-457 (2008).

140   Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *P Natl Acad Sci USA* 96, 9833-9838 (1999).

141   Pickard, J. M. *et al.* Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature* 514, 638-641 (2014).

142   Varki, A. Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology* 3, 97-130 (1993).

143   Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* 126, 855-867 (2006).

144   Mollicone, R., Gibaud, A., Francois, A., Ratcliffe, M. & Oriol, R. Acceptor Specificity and Tissue Distribution of 3 Human Alpha-3-Fucosyltransferases. *Eur J Biochem* 191, 169-176 (1990).

145   Tonetti, M. *et al.* The metabolism of 6-deoxyhexoses in bacterial and animal cells. *Biochimie* 80, 923-931 (1998).

146   Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37, D233-D238 (2009).

147   Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42, D490-D495 (2014).

148   Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G. & Lowe, J. B. Sequence and Expression of a Candidate for the Human Secretor Blood-Group Alpha(1,2)Fucosyltransferase Gene (Fut2) - Homozygosity for an Enzyme-Inactivating Nonsense Mutation Commonly Correlates with the Non-Secretor Phenotype. *J Biol Chem* 270, 4640-4649 (1995).

149   Larsen, R. D., Ernst, L. K., Nair, R. P. & Lowe, J. B. Molecular-Cloning, Sequence, and Expression of a Human Gdp-L-Fucose - Beta-D-Galactoside 2-Alpha-L-Fucosyl-Transferase Cdna That Can Form the H-Blood Group Antigen. *P Natl Acad Sci USA* 87, 6674-6678 (1990).

150   Costache, M. *et al.* Evolution of fucosyltransferase genes in vertebrates. *J Biol Chem* 272, 29721-29728 (1997).

151   de Vries, T., Knegtel, R. M. A., Holmes, E. H. & Macher, B. A. Fucosyltransferases: structure/function studies. *Glycobiology* 11, 119R-128R (2001).

152   Kukowska-Latallo, J. F., Larsen, R. D., Nair, R. P. & Lowe, J. B. A cloned human cDNA determines expression of a mouse stage-specific embryonic antigen and the Lewis blood group alpha(1,3/1,4)fucosyltransferase. *Genes Dev* 4, 1288-1303 (1990).

153   Padro, M., Cobler, L., Garrido, M. & de Bolos, C. Down-regulation of FUT3 and FUT5 by shRNA alters Lewis antigens expression and reduces the adhesion capacities of gastric cancer cells. *Bba-Gen Subjects* 1810, 1141-1149 (2011).

154   Weston, B. W., Nair, R. P., Larsen, R. D. & Lowe, J. B. Isolation of a Novel Human Alpha(1,3)Fucosyltransferase Gene and Molecular Comparison to the Human Lewis Blood-Group Alpha(1,3/1,4)Fucosyltransferase Gene - Syntenic, Homologous, Nonallelic Genes Encoding Enzymes with Distinct Acceptor Substrate Specificities. *J Biol Chem* 267, 4152-4160 (1992).

155   Lowe, J. B. Selectin ligands, leukocyte trafficking, and fucosyltransferase genes. *Kidney Int* 51, 1418-1426 (1997).

156   Nishihara, S. *et al.* Alpha1,3-fucosyltransferase 9 (FUT9; Fuc-TIX) preferentially fucosylates the distal GlcNAc residue of polylactosamine chain while the other four alpha1,3FUT members preferentially fucosylate the inner GlcNAc residue. *Febs Lett* 462, 289-294 (1999).

157   Weston, B. W., Smith, P. L., Kelly, R. J. & Lowe, J. B. Molecular-Cloning of a 4th Member of a Human Alpha(1,3)Fucosyltransferase Gene Family - Multiple Homologous Sequences That Determine Expression of the Lewis-X, Sialyl Lewis-X, and Difucosyl Sialyl Lewis-X Epitopes. *J Biol Chem* 267, 24575-24584 (1992).

158   Voynow, J. A., Kaiser, R. S., Scanlin, T. F. & Glick, M. C. Purification and Characterization of Gdp-L-Fucose-N-Acetyl-Beta-D-Glucosaminide-Alpha-1-]6fucosyltransferase from Cultured Human Skin Fibroblasts - Requirement of a Specific Biantennary Oligosaccharide as Substrate. *J Biol Chem* 266, 21572-21577 (1991).

159   Miyoshi, E. *et al.* The alpha 1-6-fucosyltransferase gene and its biological significance. *Bba-Gen Subjects* 1473, 9-20 (1999).

160   Ji, J. *et al.* Expression of alpha 1,6-fucosyltransferase 8 in hepatitis B virus-related hepatocellular carcinoma influences tumour progression. *Digest Liver Dis* 45, 414-421 (2013).

161   Wang, X. C. *et al.* Overexpression of alpha (1,6) fucosyltransferase associated with aggressive prostate cancer. *Glycobiology* 24, 935-944 (2014).

162   Kumar, A. *et al.* The Lewis X-related alpha 1,3-Fucosyltransferase, Fut10, Is Required for the Maintenance of Stem Cell Populations. *J Biol Chem* 288, 28859-28868 (2013).

163   Chen, C. I. *et al.* Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation. *Embo J* 31, 3183-3197 (2012).

164   Koshland, D. E. Stereochemistry and the Mechanism of Enzymatic Reactions. *Biol Rev* 28, 416-436 (1953).

165   Mccarter, J. D. & Withers, S. G. Mechanisms of Enzymatic Glycoside Hydrolysis. *Curr Opin Struc Biol* 4, 885-892 (1994).

166   Sinnott, M. L. Catalytic Mechanisms of Enzymatic Glycosyl Transfer. *Chem Rev* 90, 1171-1202 (1990).

167   Davies, G. J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Acc Chem Res* 45, 308-316 (2012).

168   Pauling, L. *Chem. Eng. News*, 1375-1377 (1946).

169   Gloster, T. M. & Davies, G. J. Glycosidase inhibition: assessing mimicry of the transition state. *Org Biomol Chem* 8, 305-320 (2010).

170   Henrissat, B. A Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities. *Biochem J* 280, 309-316 (1991).

171   de la Torre, F., Sampedro, J., Zarra, I. & Revilla, G. AtFXG1, an Arabidopsis gene encoding alpha-L-fucosidase active against fucosylated xyloglucan oligosaccharides. *Plant Physiol* 128, 247-255 (2002).

172   Benesova, E. *et al.* Alpha-L-Fucosidase Isoenzyme iso2 from Paenibacillus thiaminolyticus. *BMC Biotechnol* 15 (2015).

173   Michalski, J. C. & Klein, A. Glycoprotein lysosomal storage disorders: alpha- and beta-mannosidosis, fucosidosis acid alpha-N-acetylgalactosaminidase deficiency. *Bba-Mol Basis Dis* 1455, 69-84 (1999).

174   Willems, P. J. *et al.* Fucosidosis Revisited - a Review of 77 Patients. *Am J Med Genet* 38, 111-131 (1991).

175   Kelly, W. R., Clague, A. E., Barns, R. J., Bate, M. J. & Mackay, B. M. Canine Alpha-L-Fucosidosis - a Storage Disease of Springer Spaniels. *Acta Neuropathol* 60, 9-13 (1983).

176   Arrol, L. P., Kerrins, A. M., Yamakawa, Y. & Smith, P. M. Fucosidosis in a domestic shorthair cat. *Journal of Feline Medicine and Surgery* 13, 120-124 (2011).

177   Willems, P. J., Seo, H. C., Coucke, P., Tonlorenzi, R. & O'Brien, J. S. Spectrum of mutations in fucosidosis. *Eur J Hum Genet* 7, 60-67 (1999).

178   Occhiodoro, T., Hopwood, J. J., Morris, C. P. & Anson, D. S. Correction of Alpha-L-Fucosidase Deficiency in Fucosidosis Fibroblasts by Retroviral Vector-Mediated Gene-Transfer. *Hum Gene Ther* 3, 365-369 (1992).

179   Nagae, M. *et al.* Structural basis of the catalytic reaction mechanism of novel 1,2-alpha-L-fucosidase from Bifidobacterium bifidum. *J Biol Chem* 282, 18497-18509 (2007).

180   Tarling, C. A. *et al.* Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from Thermotoga maritima through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis. *J Biol Chem* 278, 47394-47399 (2003).

181   Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M. & Moracci, M. Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from Sulfolobus solfataricus via chemical rescue of an inactive mutant. *Biochemistry-Us* 42, 9525-9531 (2003).

182   Sulzenbacher, G. *et al.* Crystal structure of Thermotoga maritima alpha-L-fucosidase - Insights into the catalytic mechanism and the molecular basis for fucosidosis. *J Biol Chem* 279, 13119-13128 (2004).

183   McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D* 67, 386-394 (2011).

184   van Bueren, A. L. *et al.* Analysis of the Reaction Coordinate of alpha-L-Fucosidases: A Combined Structural and Quantum Mechanical Approach. *J Am Chem Soc* 132, 1804-1806 (2010).

185   Guillotin, L., P., L. & Daniellou, R. Unraveling the substrate recognition mechanism and specificity of the unusual. *Biochemistry-Us* 53, 1447-1455 (2014).

186   Sela, D. A. *et al.* Bifidobacterium longum subsp infantis ATCC 15697 alpha-Fucosidases Are Active on Fucosylated Human Milk Oligosaccharides. *Appl Environ Microb* 78, 795-803 (2012).

187   Sakurama, H. *et al.* 1,3-1,4-alpha-L-Fucosynthase That Specifically Introduces Lewis a/x Antigens into Type-1/2 Chains. *J Biol Chem* 287, 16709-16719 (2012).

188   Cao, H., Walton, J. D., Brumm, P. & Phillips, G. N., Jr. Structure and substrate specificity of a eukaryotic fucosidase from Fusarium graminearum. *J Biol Chem* 289, 25624-25638 (2014).

189   Shaikh, F. A., van Bueren, A. L., Davies, G. J. & Withers, S. G. Identifying the Catalytic Acid/Base in GH29 alpha-L-Fucosidase Subfamilies. *Biochemistry-Us* 52, 5857-5864 (2013).

190   Ashida, H. *et al.* Two distinct alpha-l-fucosidases from Bifidobacterium bifidum are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* 19, 1010-1017 (2009).

191   Sakurama, H. *et al.* Differences in the Substrate Specificities and Active-Site Structures of Two alpha-L-Fucosidases (Glycoside Hydrolase Family 29) from Bacteroides thetaiotaomicron. *Biosci Biotech Bioch* 76, 1022-1024 (2012).

192   Matsuda, J. *et al.* Chemical chaperone therapy for brain pathology in G(M1)-gangliosidosis. *P Natl Acad Sci USA* 100, 15912-15917 (2003).

193   Fan, J. Q., Ishii, S., Asano, N. & Suzuki, Y. Accelerated transport and maturation of lysosomal alpha-galactosidase A in Fabry lymphoblasts by an enzyme inhibitor. *Nature Medicine* 5, 112-115 (1999).

194   Frustaci, A. *et al.* Improvement in cardiac function in the cardiac variant of Fabry's disease with galactose-infusion therapy. *N Engl J Med* 345, 25-32 (2001).

195   Ishii, S. Pharmacological chaperone therapy for Fabry disease. *P Jpn Acad B-Phys* 88, 18-30 (2012).

196   Stanley, P. Regulation of Notch signaling by glycosylation. *Curr Opin Struc Biol* 17, 530-535 (2007).

197   Shah, M., Telang, S., Raval, G., Shah, P. & Patel, P. S. Serum fucosylation changes in oral cancer and oral precancerous conditions - alpha-L-fucosidase as a marker. *Cancer* 113, 336-346 (2008).

198   Martens, E. C. *et al.* Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *Plos Biol* 9, e1001221 (2011).

199   Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299, 2074-2076 (2003).

200   van Bueren, A. L., Popat, S. D., Lin, C. H. & Davies, G. J. Structural and Thermodynamic Analyses of alpha-L-Fucosidase Inhibitors. *Chembiochem* 11, 1971-1974 (2010).

201   Welter, A., Jadot, J., Dardenne, G., Marlier, M. & Casimir, J. 2,5-Dihydroxymethyl-3,4-Dihydroxypyrrolidine in Leaves of Derris-Elliptica. *Phytochemistry* 15, 747-749 (1976).

202   Fleet, G. W. J. *et al.* Potent Competitive-Inhibition of Alpha-Galactosidase and Alpha-Glucosidase Activity by 1,4-Dideoxy-1,4-Iminopentitols - Syntheses of 1,4-Dideoxy-1,4-Imino-D-Lyxitol and of Both Enantiomers of 1,4-Dideoxy-1,4-Iminoarabinitol. *Tetrahedron Lett* 26, 3127-3130 (1985).

203   Aoyagi, T. *et al.* Mannostatin-a and Mannostatin-B - New Inhibitors of Alpha-D-Mannosidase, Produced by Streptoverticillium-Verticillus Var Quintum Me3-Ag3 - Taxonomy, Production, Isolation, Physicochemical Properties and Biological-Activities. *J Antibiot* 42, 883-889 (1989).

204   Elbein, A. D., Mitchell, M., Sanford, B. A., Fellows, L. E. & Evans, S. V. The Pyrrolidine Alkaloid, 2,5-Dihydroxymethyl-3,4-Dihydroxypyrrolidine, Inhibits Glycoprotein Processing. *J Biol Chem* 259, 2409-2413 (1984).

205   Asano, N. *et al.* Novel alpha-L-fucosidase inhibitors from the bark of Angylocalyx pynaertii (Leguminosae). *Eur J Biochem* 268, 35-41 (2001).

206   Bols, M. 1-aza sugars, apparent transition state analogues of equatorial glycoside formation/cleavage. *Accounts Chem Res* 31, 1-8 (1998).

207   Wang, Y. F., Dumas, D. P. & Wong, C. H. Chemoenzymatic Synthesis of 5-Membered Azasugars as Inhibitors of Fucosidase and Fucosyl-Transferase - an Issue Regarding the Stereochemistry Discrimination at Transition-States. *Tetrahedron Lett* 34, 403-406 (1993).

208   Robina, I. *et al.* New leads for selective inhibitors of alpha-L-fucosidases. Synthesis and glycosidase inhibitory activities of [(2R,3S,4R)-3,4-dihydroxypyrrolidin-2-yl]furan derivatives. *Bioorg Med Chem Lett* 11, 2555-2559 (2001).

209   Moreno-Vargas, A. J., Robina, I., Demange, R. & Vogel, P. Synthesis and glycosidase inhibitory activities of 5-(1',4'-dideoxy-1',4'-imino-D-erythrosyl)-2-methyl-3-furoic acid (=5-[3S, 4R)-3,4-dihydroxypyrrolidin-2-yl]-2-methylfuran-3-carboxylic acid) derivitives: New leads as selective alpha-L-fucosidase and beta-galactosidase inhibitors. *Helv Chim Acta* 86, 1894-1913 (2003).

210   Fleet, G. W. J., Shaw, A. N., Evans, S. V. & Evans, S. V. Synthesis from D-Glucose of 1,5-Dideoxy-1,5-Imino-L-Fucitol, a Potent Alpha-L-Fucosidase Inhibitor. *J Chem Soc Chem Comm*, 841-842 (1985).

211   Chevrier, C., LeNouen, D., Neuburger, M., Defoin, A. & Tarnus, C. Nitrone in L-lyxose series: cycloaddition way for the synthesis of new C-alpha-fucosides. *Tetrahedron Lett* 45, 5363-5366 (2004).

212   Moreno-Vargas, A. J., Carmona, A. T., Mora, F., Vogel, P. & Robina, I. Stereoselective synthesis of (2S,3S,4R,5S)-5-methylpyrrolidine-3,4-diol derivatives that are highly selective alpha-L-fucosidase inhibitors. *Chem Commun*, 4949-4951 (2005).

213   Laroche, C. *et al.* Spirocyclopropyl pyrrolidines as a new series of alpha-L-fucosidase inhibitors. *Bioorgan Med Chem* 14, 4047-4054 (2006).

214   Behr, J. B., Pearson, M. S. M., Bello, C., Vogel, P. & Plantier-Royon, R. Synthesis and L-fucosidase inhibitory potency of a cyclic sugar imine and its pyrrolidine analogue. *Tetrahedron-Asymmetr* 19, 1829-1832 (2008).

215   Yasuda, K. *et al.* New sugar-mimic alkaloids from the pods of Angylocalyx pynaertii. *J Nat Prod* 65, 198-202 (2002).

216   Calveras, J. *et al.* Dihydroxyacetone Phosphate Aldolase Catalyzed Synthesis of Structurally Diverse Polyhydroxylated Pyrrolidine Derivatives and Evaluation of their Glycosidase Inhibitory Properties. *Chem-Eur J* 15, 7310-7328 (2009).

217   Ak, A., Prudent, S., LeNouen, D., Defoin, A. & Tarnus, C. Synthesis of all-cis 2,5-imino-2,5-dideoxy-fucitol and its evaluation as a potent fucosidase and galactosidase inhibitor. *Bioorg Med Chem Lett* 20, 7410-7413 (2010).

218   Kotland, A., Accadbled, F., Robeyns, K. & Behr, J. B. Synthesis and Fucosidase Inhibitory Study of Unnatural Pyrrolidine Alkaloid 4-epi-(+)-Codonopsinine. *J Org Chem* 76, 4094-4098 (2011).

219   Elias-Rodriguez, P., Moreno-Clavijo, E., Carmona, A. T., Moreno-Vargas, A. J. & Robina, I. Rapid discovery of potent alpha-fucosidase inhibitors by in situ screening of a. *Org Biomol Chem* 21, 5898-5904 (2014).

220   Wu, C. Y., Chang, C. F., Chen, J. S. Y., Wong, C. H. & Lin, C. H. Rapid diversity-oriented synthesis in microtiter plates for in situ screening: Discovery of potent and selective alpha-fucosidase inhibitors. *Angew Chem Int Edit* 42, 4661-4664 (2003).

221   Wu, H. J. *et al.* Structural Basis of alpha-Fucosidase Inhibition by Iminocyclitols with K(i) Values in the Micro- to Picomolar Range. *Angew Chem Int Edit* 49, 337-340 (2010).

222   Kuntz, D. A., Ghavami, A., Johnston, B. D., Pinto, B. M. & Rose, D. R. Crystallographic analysis of the interactions of Drosophila melanogaster Golgi alpha-mannosidase II with the naturally occurring glycomimetic salacinol and its analogues. *Tetrahedron-Asymmetr* 16, 25-32 (2005).

223   Caines, M. E. C. *et al.* The structural basis of glycosidase inhibition by five-membered iminocyclitols: The clan a glycoside hydrolase endoglycoceramidase as a model system. *Angew Chem Int Edit* 46, 4474-4476 (2007).

224   Hottin, A., Wright, D. W., Davies, G. J. & Behr, J. B. Exploiting the Hydrophobic Terrain in Fucosidases with Aryl-Substituted Pyrrolidine Iminosugars. *Chembiochem*, 277-283 (2014).

225   Kabsch, W. Xds. *Acta Crystallogr D* 66, 125-132 (2010).

226   Leslie, A. G. W. & Powell, H. R. Processing diffraction data with MOSFLM. *Nato Sci Ser Ii Math* 245, 41-51 (2007).

227   Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D* 62, 72-82 (2006).

228   Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D* 69, 1204-1214 (2013).

229   Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* 30, 1022-1025 (1997).

230   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D* 66, 486-501 (2010).

231   Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* 67, 355-367 (2011).

232   Schuttelkopf, A. W. & van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D* 60, 1355-1363 (2004).

233   Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* 66, 12-21 (2010).

234   Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* 28, 235-242 (2000).

235   Moreno-Clavijo, E. *et al.* Synthesis of novel pyrrolidine 3,4-diol derivatives as inhibitors of alpha-L-fucosidases. *Org Biomol Chem* 7, 1192-1202 (2009).

236   Wright, D. W., Moreno-Vargas, A. J., Carmona, A. T., Robina, I. & Davies, G. J. Three dimensional structure of a bacterial alpha-L-fucosidase with a 5-membered iminocyclitol inhibitor. *Bioorgan Med Chem* 21, 4751-4754 (2013).

237   Pieters, R. J. Maximising multivalency effects in protein-carbohydrate interactions. *Org Biomol Chem* 7, 2013-2025 (2009).

238   Brissonnet, Y. *et al.* Polymeric iminosugars improve the activity of carbohydrate-processing enzymes. *Bioconjug Chem* 26, 766-772 (2015).

239   Lundquist, J. J. & Toone, E. J. The cluster glycoside effect. *Chem Rev* 102, 555-578 (2002).

240   Lee, Y. C. & Lee, R. T. Carbohydrate-Protein Interactions - Basis of Glycobiology. *Accounts Chem Res* 28, 321-327 (1995).

241   Lee, Y. C. *et al.* Binding of Synthetic Oligosaccharides to the Hepatic Gal Galnac Lectin - Dependence on Fine-Structural Features. *J Biol Chem* 258, 199-202 (1983).

242   Jencks, W. P. On the Attribution and Additivity of Binding-Energies. *P Natl Acad Sci-Biol* 78, 4046-4050 (1981).

243   Gilboaga, N., Garber, N. & Mizrahi, L. Purification of Galactose-Binding Hemagglutinin of Pseudomonas-Aeruginosa by Affinity Column Chromatography Using Sepharose. *Febs Lett* 28, 93-95 (1972).

244 Gilboa-Garber, N., Katcoff, D. J. & Garber, N. C. Identification and characterization of Pseudomonas aeruginosa PA-IIL lectin gene and protein compared to PA-IL. *Fems Immunol Med Mic* 29, 53-57 (2000).

245 Johansson, E. M. V. *et al.* Inhibition and Dispersion of Pseudomonas aeruginosa Biofilms by Glycopeptide Dendrimers Targeting the Fucose-Specific Lectin LecB. *Chem Biol* 15, 1249-1257 (2008).

246 Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* 414, 105-111 (2001).

247 Clevers, H. The cancer stem cell: premises, promises and challenges. *Nature Medicine* 17, 313-319 (2011).

248 Sawyers, C. L. The cancer biomarker problem. *Nature* 452, 548-552 (2008).

249 Brannon-Peppas, L. & Blanchette, J. O. Nanoparticle and targeted systems for cancer therapy. *Adv Drug Deliver Rev* 64, 206-212 (2012).

250 Osella, D. *et al.* On the mechanism of the antitumor activity of ferrocenium derivatives. *Inorg Chim Acta* 306, 42-48 (2000).

251 Schumacker, P. T. Reactive oxygen species in cancer cells: Live by the sword, die by the sword. *Cancer Cell* 10, 175-176 (2006).

252 Trachootham, D., Alexandre, J. & Huang, P. Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach? *Nat Rev Drug Discov* 8, 579-591 (2009).

253 Moreno-Clavijo, E. *et al.* Exploring a Multivalent Approach to alpha-L-Fucosidase Inhibition. *Eur J Org Chem* 2013, 7328-7336 (2013).

254 Hottin, A. *et al.* Iminosugar-ferrocene conjugates as potential anticancer agents. *Org Biomol Chem* 10, 5592-5597 (2012).

255 Unpublished results, Jean-Bernard Behr.

256 Cuskin, F. *et al.* How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *P Natl Acad Sci USA* 109, 20889-20894 (2012).

257 Hottin, A. *et al.* alpha-L-Fucosidase Inhibition by Pyrrolidine-Ferrocene Hybrids: Rationalization of Ligand-Binding Properties by Structural Studies. *Chem-Eur J* 19, 9526-9533 (2013).

258   Heine, A. *et al.* An antibody exo Diels-Alderase inhibitor complex at 1.95 angstrom resolution. *Science* 279, 1934-1940 (1998).

259   Salmon, A. J., Williams, M. L., Hofmann, A. & Poulsen, S. A. Protein crystal structures with ferrocene and ruthenocene-based enzyme inhibitors. *Chem Commun (Camb)* 48, 2328-2330 (2012).

260   Mooney, A. *et al.* Structure--activity relationship and mode of action of N-(6-ferrocenyl-2-naphthoyl) dipeptide ethyl esters: novel organometallic anticancer compounds. *J Med Chem* 55, 5455-5466 (2012).

261   Jia, J. H., Tong, C., Wang, B., Luo, L. P. & Jiang, J. Hedgehog signalling activity of Smoothened requires phosphorylation by protein kinase A and casein kinase I. *Nature* 432, 1045-1050 (2004).

262   Steiner, D. F. & Oyer, P. E. The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. *Proc Natl Acad Sci U S A* 57, 473-480 (1967).

263   Weissleder, R. & Ntziachristos, V. Shedding light onto live molecular targets. *Nature Medicine* 9, 123-128 (2003).

264   James, M. L. & Gambhir, S. S. A Molecular Imaging Primer: Modalities, Imaging Agents, and Applications. *Physiol Rev* 92, 897-965 (2012).

265   Heal, W. P., Dang, T. H. T. & Tate, E. W. Activity-based probes: discovering new biology and new drug targets. *Chem Soc Rev* 40, 246-257 (2011).

266   Grammel, M. & Hang, H. C. Chemical reporters for biological discovery. *Nat Chem Biol* 9, 475-484 (2013).

267   Willems, L. I. *et al.* Bioorthogonal Chemistry: Applications in Activity-Based Protein Profiling. *Accounts Chem Res* 44, 718-729 (2011).

268   Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* 287, 2007-2010 (2000).

269   Vocadlo, D. J. & Bertozzi, C. R. A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angew Chem Int Edit* 43, 5338-5342 (2004).

270   Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne [3+2] cycloaddition. *J Am Chem Soc* 125, 4686-4687 (2003).

271   van der Linden, W. A. *et al.* Two-step bioorthogonal activity-based proteasome profiling using copper-free click reagents: a comparative study. *Bioorg Med Chem* 20, 662-666 (2012).

272   Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Edit* 41, 2596-2599 (2002).

273   Stubbs, K. A. *et al.* Synthesis and use of mechanism-based protein-profiling probes for retaining beta-D-glucosaminidases facilitate identification of Pseudomonas aeruginosa NagZ. *J Am Chem Soc* 130, 327-335 (2008).

274   Witte, M. D. *et al.* Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat Chem Biol* 6, 907-913 (2010).

275   Bennett, L. L. & Mohan, D. Gaucher Disease and Its Treatment Options. *Annals of Pharmacotherapy* 47, 1182-1193 (2013).

276   Jonsson, L. M. V. *et al.* Biosynthesis and Maturation of Glucocerebrosidase in Gaucher Fibroblasts. *Eur J Biochem* 164, 171-179 (1987).

277   Gopalan, V., Pastuszyn, A., Galey, W. R. & Glew, R. H. Exolytic Hydrolysis of Toxic Plant Glucosides by Guinea-Pig Liver Cytosolic Beta-Glucosidase. *J Biol Chem* 267, 14027-14032 (1992).

278   Lloyd, M. *et al.* Regulation of Intestinal Lactase in Adult Hypolactasia. *J Clin Invest* 89, 524-529 (1992).

279   Kallemeijn, W. W. *et al.* Novel Activity-Based Probes for Broad-Spectrum Profiling of Retaining beta-Exoglucosidases In Situ and In Vivo. *Angew Chem Int Edit* 51, 12529-12533 (2012).

280   Thomas, E. W., Mckelvy, J. F. & Sharon, N. Specific and Irreversible Inhibition of Lysozyme by 2',3'-Epoxypropyl Beta-Glycosides of N-Acetyl-D-Glucosamine Oligomers. *Nature* 222, 485-486 (1969).

281   Legler, G. & Bause, E. Epoxyalkyl Oligo-(-]4)-Beta-D-Glucosides as Active-Site-Directed Inhibitors of Cellulases. *Carbohyd Res* 28, 45-52 (1973).

282   Havukainen, R., Torronen, A., Laitinen, T. & Rouvinen, J. Covalent binding of three epoxyalkyl xylosides to the active site of endo-1,4-xylanase II from Trichoderma reesei. *Biochemistry-Us* 35, 9617-9624 (1996).

283   Keitel, T., Simon, O., Borriss, R. & Heinemann, U. Molecular and Active-Site Structure of a Bacillus 1,3-1,4-Beta-Glucanase. *P Natl Acad Sci USA* 90, 5287-5291 (1993).

284   Quaroni, A., Gershon, E. & Semenza, G. Affinity Labeling of Active-Sites in Sucrase-Isomaltase Complex from Small-Intestine. *J Biol Chem* 249, 6424-6433 (1974).

285   Hermans, M. M. P., Kroos, M. A., Vanbeeumen, J., Oostra, B. A. & Reuser, A. J. J. Human Lysosomal Alpha-Glucosidase - Characterization of the Catalytic Site. *J Biol Chem* 266, 13507-13512 (1991).

286   Tong, M. K. & Ganem, B. A Potent New Class of Active-Site-Directed Glycosidase Inactivators. *J Am Chem Soc* 110, 312-313 (1988).

287   Caron, G. & Withers, S. G. Conduritol Aziridine - a New Mechanism-Based Glucosidase Inactivator. *Biochem Bioph Res Co* 163, 495-499 (1989).

288   Premkumar, L. *et al.* X-ray structure of human acid-beta-glucosidase covalently bound to conduritol-B-epoxide - Implications for Gaucher disease. *J Biol Chem* 280, 23815-23819 (2005).

289   Dvir, H. *et al.* X-ray structure of human acid-beta-glucosidase, the defective enzyme in Gaucher disease. *Embo Rep* 4, 704-709 (2003).

290   Gloster, T. M., Madsen, R. & Davies, G. J. Structural basis for cyclophellitol inhibition of a beta-glucosidase. *Org Biomol Chem* 5, 444-446 (2007).

291   Atsumi, S. *et al.* Production, Isolation and Structure Determination of a Novel Beta-Glucosidase Inhibitor, Cyclophellitol, from Phellinus Sp. *J Antibiot* 43, 49-53 (1990).

292   Jiang, J. *et al.* In vitro and in vivo comparative and competitive activity-based protein profiling of GH29 [small alpha]-l-fucosidases. *Chem Sci* 6, 2782-2789 (2015).

293   Bergfors, T. Seeds to crystals. *J Struct Biol* 142, 66-76 (2003).

294   Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43, 186-190 (2010).

295   Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* 68, 431-440 (2012).

296   Cremer, D. & Pople, J. A. General Definition of Ring Puckering Coordinates. *J Am Chem Soc* 97, 1354-1358 (1975).

297   Gloster, T. M., Madsen, R. & Davies, G. J. Structural basis for cyclophellitol inhibition of a beta-glucosidase. *Org Biomol Chem* 5, 444-446 (2007).

298   Moran, M. T. *et al.* Pathologic gene expression in Gaucher disease: up-regulation of cysteine proteinases including osteoclastic cathepsin K. *Blood* 96, 1969-1978 (2000).

299   Levvy, G. A. & McAllan, A. Mammalian fucosidases. 2. alpha-L-Fucosidase. *Biochem J* 80, 435-439 (1961).

300   Watkins, W. M. Enzymes of Trichomonas-Foetus - Action of Cell-Free Extracts on Blood-Group Substances and Low-Molecular-Weight Glycosides. *Biochem J* 71, 261-274 (1959).

301   Kuhn, R., Baer, H. H. & Gauhe, A. Fucosido-Lactose, Das Trisaccharid Der Frauenmilch. *Chem Ber-Recl* 88, 1135-1146 (1955).

302   Levvy, G. A. & McAllan, A. Mammalian fucosidases. 1. The synthesis of substrates and inhibitors. *Biochem J* 80, 433-435 (1961).

303   Wiederschain, G. Y. & Rosenfeld, E. L. 2 Forms of Alpha-L-Fucosidase from Pig Kidney and Their Action on Natural Oligosaccharides. *Biochem Bioph Res Co* 44, 1008-1014 (1971).

304   Robinson, D. & Thorpe, R. Human Liver Alpha-L-Fucosidases. *Clin Chim Acta* 47, 403-407 (1973).

305   Wiederschain, G. Y. & Prokopen, A. A. Various Forms of Alpha-L-Fucosidase and Other Glycosidases in Porcine Kidneys. *Biochemistry-Ussr+* 38, 113-120 (1973).

306   Alhadeff, J. A., Miller, A. L., Wenger, D. A. & Obrien, J. S. Electrophoretic Forms of Human Liver Alpha-L-Fucosidase and Their Relationship to Fucosidosis (Mucopolysaccharidosis F). *Clin Chim Acta* 57, 307-313 (1974).

307   Turner, B. M., Turner, V. S., Beratis, N. G. & Hirschhorn, K. Polymorphism of Human Alpha-Fucosidase. *American Journal of Human Genetics* 27, 651-661 (1975).

308   Johnson, S. W. & Alhadeff, J. A. Mammalian Alpha-L-Fucosidases. *Comp Biochem Phys B* 99, 479-488 (1991).

309   Thorpe, R. & Robinson, D. Isoelectric Focusing of Isoenzymes of Human Liver Alpha-L-Fucosidase. *Febs Lett* 54, 89-92 (1975).

310   Turner, B. M., Beratis, N. G., Turner, V. S. & Hirschho.K. Isoenzymes of Human Alpha-L-Fucosidase Detectable by Starch-Gel Electrophoresis. *Clin Chim Acta* 57, 29-35 (1974).

311   Alhadeff, J. A., Cimino, G. & Janowsky, A. Isoenzymes of Human Liver Alpha-L-Fucosidase - Chemical Relationship, Kinetic Studies, and Immunochemical Characterization. *Mol Cell Biochem* 19, 171-180 (1978).

312   Piesecki, S. & Alhadeff, J. A. The Effect of Carbohydrate Removal on the Properties of Human Liver Alpha-L-Fucosidase. *Biochim Biophys Acta* 1119, 194-200 (1992).

313   Alhadeff, J. A., Miller, A. L., Wenaas, H., Vedvick, T. & Obrien, J. S. Human Liver Alpha-L-Fucosidase - Purification, Characterization, and Immunochemical Studies. *J Biol Chem* 250, 7106-7113 (1975).

314   Carlsen, R. B. & Pierce, J. G. Purification and Properties of an Alpha-L-Fucosidade from Rat Epididymis. *J Biol Chem* 247, 23-32 (1972).

315   Alhadeff, J. A. & Janowsky, A. J. Human-Serum Alpha-L-Fucosidase. *Clin Chim Acta* 82, 133-140 (1978).

316   Alhadeff, J. A. & Janowsky, A. J. Purification and Properties of Human-Brain Alpha-L-Fucosidase. *J Neurochem* 28, 423-427 (1977).

317   Occhiodoro, T., Beckmann, K. R., Morris, C. P. & Hopwood, J. J. Human alpha-L-fucosidase: complete coding sequence from cDNA clones. *Biochem Biophys Res Commun* 164, 439-445 (1989).

318   de Carlos, A. *et al.* Purification of human alpha-L-fucosidase precursor expressed in Escherichia coli as a glutathione S-transferase fusion protein. *J Chromatogr B Analyt Technol Biomed Life Sci* 786, 7-15 (2003).

319   Liu, S.-W. *et al.* Identification of Essential Residues of Human alpha-L-Fucosidase and Tests of Its Mechanism. *Biochemistry-Us* 48, 110-120 (2009).

320   Vocadlo, D. J., Wicki, J., Rupitz, K. & Withers, S. G. A case for reverse protonation: Identification of Glu160 as an acid/base catalyst in Thermoanaerobacterium saccharolyticum ss-xylosidase and detailed kinetic analysis of a site-directed mutant. *Biochemistry-Us* 41, 9736-9746 (2002).

321   Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845-858 (2015).

322   Intra, J., Perotti, M. E., Pavesi, G. & Horner, D. Comparative and phylogenetic analysis of alpha-L-fucosidase genes. *Gene* 392, 34-46 (2007).

323   Eiberg, H., Mohr, J. & Nielsen, L. S. Linkage of Plasma Alpha-L-Fucosidase (Fuca2) and the Plasminogen (Plg) System. *Clin Genet* 26, 23-29 (1984).

324   Murray, J. C., Sadler, E., Eddy, R. L., Shows, T. B. & Buetow, K. H. Evidence for Assignment of Plasminogen (Plg) to Chromosome-6, Not Chromosome-4. *Cytogenetics and Cell Genetics* 40, 709-709 (1985).

325   Chang, V. T. *et al.* Glycoprotein structural genomics: Solving the glycosylation problem. *Structure* 15, 267-273 (2007).

326   Stanley, P. Chinese-Hamster Ovary Cell Mutants with Multiple Glycosylation Defects for Production of Glycoproteins with Minimal Carbohydrate Heterogeneity. *Mol Cell Biol* 9, 377-383 (1989).

327   Patnaik, S. K. & Stanley, P. Lectin-resistant CHO glycosylation mutants. *Methods Enzymol* 416, 159-182 (2006).

328   Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *P Natl Acad Sci USA* 99, 13419-13424 (2002).

329   Crispin, M. *et al.* A Human Embryonic Kidney 293T Cell Line Mutated at the Golgi alpha-Mannosidase II Locus. *J Biol Chem* 284, 21684-21695 (2009).

330   Elbein, A. D., Tropea, J. E., Mitchell, M. & Kaushal, G. P. Kifunensine, a Potent Inhibitor of the Glycoprotein Processing Mannosidase-I. *J Biol Chem* 265, 15599-15605 (1990).

331   Elbein, A. D., Solf, R., Dorling, P. R. & Vosbeck, K. Swainsonine - an Inhibitor of Glycoprotein Processing. *P Natl Acad Sci-Biol* 78, 7393-7397 (1981).

332   Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14, 103-114 (2004).

333   Davis, S. J. *et al.* Ligand-Binding by the Immunoglobulin Superfamily Recognition Molecule Cd2 Is Glycosylation-Independent. *J Biol Chem* 270, 369-375 (1995).

334   Liu, S.-W. & Li, Y.-K. Expression, Purification and Characterization of Human alpha-L-Fucosidase. *J Chin Chem Soc-Taip* 56, 850-858 (2009).

335   Marblestone, J. G. *et al.* Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci* 15, 182-189 (2006).

336   Zuo, X. *et al.* Expression and purification of SARS coronavirus proteins using SUMO-fusions. *Protein Expres Purif* 42, 100-110 (2005).

337   Zuo, X. *et al.* Enhanced expression and purification of membrane proteins by SUMO fusion in Escherichia coli. *J Struct Funct Genomics* 6, 103-111 (2005).

338   Ferre, F. & Clote, P. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res* 33, W230-232 (2005).

339   Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5, 725-738 (2010).

340   Goldberger, R. F., Epstein, C. J. & Anfinsen, C. B. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *J Biol Chem* 238, 628-635 (1963).

341   Hatahet, F. & Ruddock, L. W. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxid Redox Signal* 11, 2807-2850 (2009).

342   Denoncin, K. & Collet, J. F. Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead. *Antioxid Redox Signal* 19, 63-71 (2013).

343   Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J. & Wilcox, G. Characterization of the Erwinia carotovora pelB gene and its product pectate lyase. *J Bacteriol* 169, 4379-4383 (1987).

344   de Marco, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb Cell Fact* 8, 26 (2009).

345   Aricescu, A. R. & Owens, R. J. Expression of recombinant glycoproteins in mammalian cells: towards an integrative approach to structural biology. *Curr Opin Struc Biol* 23, 345-356 (2013).

346   Aricescu, A. R., Lu, W. X. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D* 62, 1243-1250 (2006).

347   Wilke, S. *et al.* Streamlining Homogeneous Glycoprotein Production for Biophysical and Structural Applications by Targeted Cell Line Development. *Plos One* 6, e27829 (2011).

348   Fiaux, H. *et al.* Functionalized pyrrolidine inhibitors of human type II alpha-mannosidases as anti-cancer agents: Optimizing the fit to the active site. *Bioorgan Med Chem* 16, 7337-7346 (2008).

349   Kuntz, D. A. *et al.* Structural Investigation of the Binding of 5-Substituted Swainsonine Analogues to Golgi alpha-Mannosidase II. *Chembiochem* 11, 673-680 (2010).

350   Witte, M. D. *et al.* Activity-Based Profiling of Retaining beta-Glucosidases: A Comparative Study. *Chembiochem* 12, 1263-1269 (2011).

351   Fenger, T. H. & Brumer, H. Synthesis and Analysis of Specific Covalent Inhibitors of endo-Xyloglucanases. *Chembiochem* 16, 575-583 (2015).

352   Laughlin, S. T., Baskin, J. M., Amacher, S. L. & Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* 320, 664-667 (2008).

353   Bieberich, E. Synthesis, Processing, and Function of N-glycans in N-glycoproteins. *Advances in neurobiology* 9, 47-70 (2014).

354   Issad, T., Masson, E. & Pagesy, P. O-GlcNAc modification, insulin signaling and diabetic complications. *Diabetes Metab* 36, 423-435 (2010).

355   Ma, Z. Y. & Vosseller, K. O-GlcNAc in cancer biology. *Amino Acids* 45, 719-733 (2013).

356   Lazarus, B. D., Love, D. C. & Hanover, J. A. O-GlcNAc cycling: Implications for neurodegenerative disorders. *Int J Biochem Cell B* 41, 2134-2146 (2009).

357   Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649 (2012).