Glycoside Hydrolases, Family GH73 and a Structural Characterization of GH73 Enzyme FlgJ

by

Patryk Zaloba

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

MASTER OF SCIENCE

Department of Microbiology

University of Manitoba

Winnipeg

Abstract

FlgJ belongs to Carbohydrate Active enZyme (CAZy) family GH73 and facilitates passage of the bacterial flagellum through the peptidoglycan (PG) layer by cleaving the glycosidic bonds within glycan strands of PG. In this thesis I present the structure of the GH73 enzyme FlgJ from bacterial pathogen *Salmonella typhimurium* (*St*FlgJ). The *St*FlgJ active site was found to be blocked by the C-terminus of a neighbouring symmetry mate. To investigate if the C-terminus of FlgJ inhibits enzymatic activity similarly to the N-terminus of GH73 enzyme Auto, the glycolytic activity of *St*FlgJ was measured with and without its C-terminus. The assays revealed *St*FlgJ activity to be unaffected by the presence of the C-terminal sequence. Removal of the C-terminus did, however, allow a crystal structure of the domain to be obtained where a β-hairpin known to accommodate critical catalytic residues was found capable of opening widely, which likely aids in substrate capture and turnover.

Acknowledgements

I would like to thank my advisory committee members Dr. Ayush Kumar, Dr. Mario Bieringer and Dr. Brian Mark for their guidance and invaluable assistance in my thesis research. I thank Ms. Veronica Larmour for technical assistance as well as Mr. Ben Bailey-Elkin for his assistance with X-ray crystallographic methods and software and Ms. Miriam Derksen for cloning the original FlgJ glycoside hydrolase domain from Salmonella *typhimurium*. This research was supported by the Natural Sciences and Engineering Research Council as well as the University of Manitoba Faculty of Graduate Studies who provided me with a Master's fellowship.

Table of Contents

List of Tables

List of Figures

List of Copyrighted Material for which Permission was Obtained

All figures have permission obtained for use in this thesis except for Figure 2 which is being used under the Fair Dealing Guideline of The Copyright Act.

Chapter 1
Glycoside Hydrolases- Introduction

Carbohydrates, in the form of polysaccharides, play a distinctive role in nature. Given that absolute possible number of hexasaccharide isomers results in trillions of variants (Laine, 1994), the functional and structural diversity of these molecules is far greater than what is possible with peptides or nucleic acids of equivalent size. Add to that the plenitude of non-carbohydrate substituents that can adorn these polysaccharides, and the diversity of complex sugars can truly be appreciated. Given their diversity, polysaccharides play a role in a vast number of biological processes such as structure, carbon reserves, energy storage, and highly selective intra- and intercellular signalling events due to their many, versatile, configurations. The enzymes hydrolysing, cleaving, synthesising, and modifying these saccharides; glycoside hydrolases, lyases, glycosyltransferases, and carbohydrate esterases are implicated in just as many biological processes (figure 1). These carbohydrate-active enzymes, coined CAZymes (www.cazy.org)(Henrissat, 1991), evolved to perform this massive array of functions from a finite number of possible protein folds (Cantarel et al., 2009). Such an enormous range of substrates and enzymes gives rise to challenges in experimental characterization and functional annotation in genomes. This introductory chapter aims to provide a glimpse into the world of CAZymes.

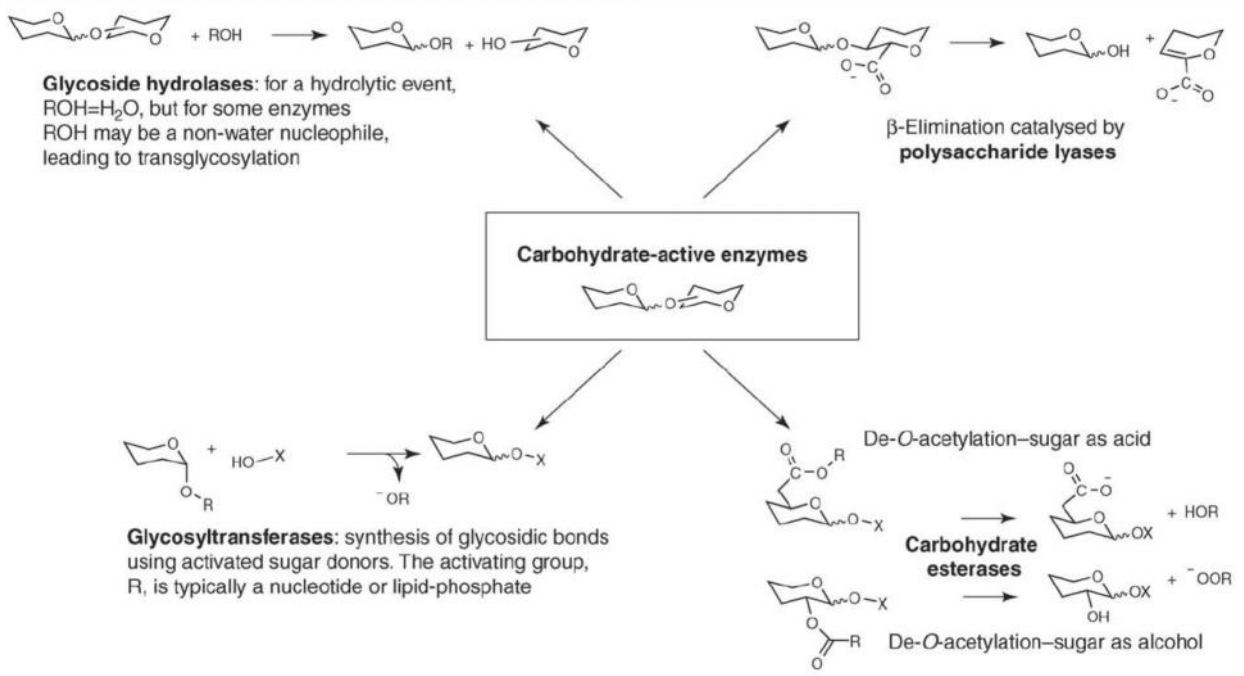


Figure 1. The different classes of carbohydrate active enzymes involved in the formation, modification and breakdown of glycosides (Figure reproduced with permission from Davies *et al*. 2005)

Glycan-degrading enzymes are composed of lyases and glycoside hydrolases (GH's) with hydrolytic glycoside hydrolases being the larger group. Glycan lyases are characterized typically through elimination reactions (Yip & Withers, 2006). The proteins encoded by glycoside hydrolase genes, which will be the primary topic of this thesis, are classified by amino acid sequence with bioinformatics analysis, particularly that of active site domains (Cantarel et al., 2009). Classification is made module by module as CAZymes are recurrently modular, with catalytic domains that can potentially harbour a number of other distinct modules that may or may not be catalytic (Lombard et al., 2014). Therefore a modular, full length CAZyme can be

assigned to more than one family if its multiple modules belong to different families (Cantarel et al., 2009). GH73 enzymes for example, which are the main focus of this thesis, have a high degree of modularity and are localized at the surface of bacteria, typically through membrane anchoring (Lipski et al., 2015). Examples that capture the degree of modularity include GH73 members that are bi-functional, having additional aminidase function with GH activity (Bourgeois et al., 2009; Foster et al., 1995; Rashid, et al. 1995; Yokoi et al., 2008). Another example is the display of repeating sequences on terminal ends of these enzymes which target peptidoglycan for binding to the cell wall, the primary substrate of GH73 enzymes, in a highly specific manner (Buist et al., 2008; Eckert et al., 2006; Mesnage et al., 2014). This modularity is, in some instances, even necessary for function as some GH73 domains are non-functional when expressed alone (Bai et al., 2014).

Glycoside Hydrolase- Classification

Glycoside hydrolase classification based upon amino acid sequence similarities began in 1991 and is still in use today (Henrissat, 1991). Given that there is a direct relationship between sequence and folding similarities (Chothia & Lesk, 1986), there is greater reason to classify the enzymes based on their structure similarities and protein sequence as opposed to function, which previous classification systems used

(Henrissat & Davies, 1997; Henrissat, 1991). Classification systems that group enzymes based on substrate or products do not necessarily address structural features and cannot take into account evolutionary events such as divergence or convergent evolution, i.e. different lineages of enzymes which end up catalyzing the same reaction (Henrissat & Davies, 1997). Ambiguity arose for enzymes showing broad specificity, i.e. those that act on multiple substrates (Nakajima et al., 1986; Raimbaud et al., 1989). Sequence and structural similarities allow grouping of GHs into 100-plus families and the resulting cataloguing system is known as CAZy, from CAZymes (www.cazy.org) (Henrissat, 1991). Currently there is an understanding of the mechanism of action of about 85 of the 100-plus families (Vocadlo & Davies, 2008). Enzymes in the same family are expected to share parallels in their mechanism of action so the prediction of the carbohydrate substrate in broad terms is possible in some cases with the CAZy classification system (Davies & Sinnott, 2008). However, a significant problem for the functional annotation of family related genes exists with the manifestation of enzymes that act on dissimilar substrates in the same family. By the grouping of members into subfamilies within families, this problem can sometimes be overcome (Stam et al., 2006), but the current knowledge base is still largely insufficient for such predictions (Cantarel et al., 2009).

Despite some inherent drawbacks, this system of classification has been validated time after time as sequence and 3D structure data of GHs have greatly increased over the last few decades with the development of new methodologies and technologies (Cantarel et al., 2009; Henrissat & Davies, 1997). Glycoside hydrolase enzymes originating from the same family, termed GH families, typically have the same catalytic mechanism (Cantarel et al., 2009). At its core, the same three-dimensional fold is expected to occur within each of the GH families (Divne et al., 1994; Törrönen et al., 1993). In some cases conservation of the catalytic mechanism has also been observed to go beyond one particular family to include gross similarities amongst members of related families. These related families arrange into what is termed GH-clans (Comfort et al., 2007).

The origins of the classification system was built on sequence comparison using hydrophobic cluster analysis (HCA) (Henrissat et al., 1995). This technique is based on the observation that hydrophobic amino acids form clusters, with these clusters regularly corresponding to secondary structures. The value of this technique arises in cases where only two residues may share an identity between two related proteins and conventional sequence alignments would fail due to too low sequence similarity (Henrissat & Davies, 1997). Explained briefly, amino acid sequences are redrawn in two

dimensions (Figure 2A), with a helical pitch, and hydrophobic amino acids (*Y, W, F, L, M, I, V*) are identified (colored). This theoretical 'helical wheel' is then cut parallel to its axis (figure 2B) and duplicated to "*restore the full environment that each amino acid has on the α-helix representation*" (figure 2C) (Callebaut et al., 1997). This resulting cluster map can provide insight into secondary structures: α-helices appear as long horizontal clusters and β-strands as shorter, vertical, arrays (Woodcock et al., 1992). This technique is most useful when comparing sequences with similarity so low as to escape detection by then conventional means, i.e. sequence alignment (Henrissat & Davies, 1997). As mentioned previously, the great diversity of GH enzymes makes HCA useful for comparing sequences which may be dissimilar at the primary sequence level, but which share gross structural similarities of the active site and overall 3D structure (Henrissat & Davies, 1997).

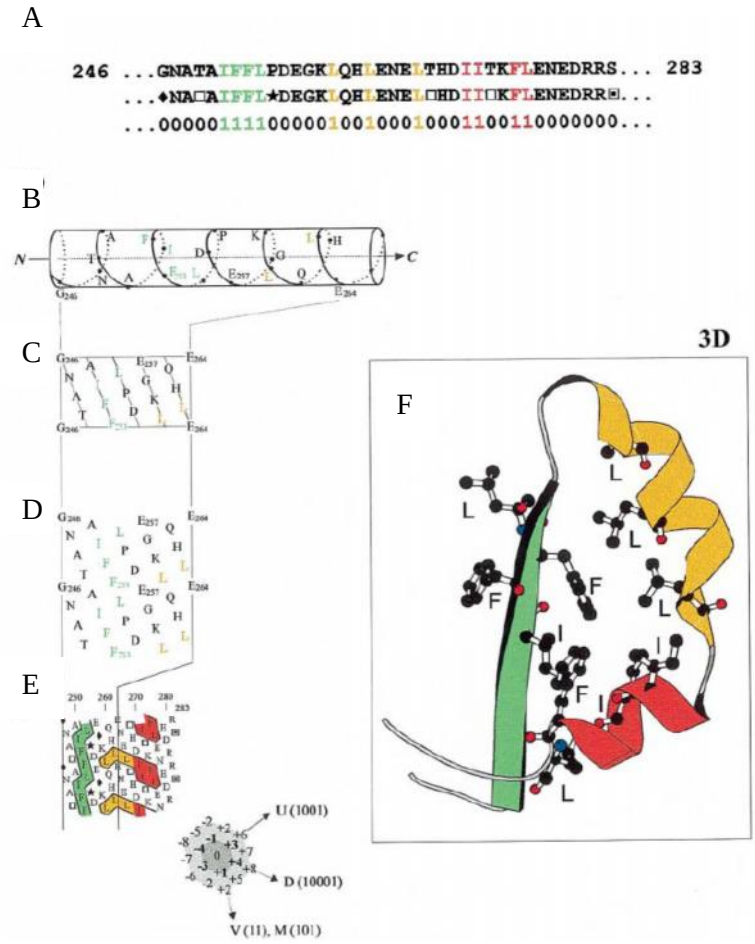


Figure 2. Example of the HCA method: A) a linear segment (1D) of human α1-antitrypsin, shown on top, has its hydrophobic amino acids coloured and translated into the HCA code. B) The sequence is then displayed on a cylinder as an α-helix. C) The cylinder is cut parallel to its axis and unrolled D) and duplicated. E) Hydrophobic amino acids are not randomly distributed and tend to form clusters. Vertical cluster (green) are linked with a β-strand whereas the horizontal ones (orange and yellow) matches to α-helices. F) Corresponding experimental 3-dimensional structure shows the predictive power of HCA (Figure reproduced from Callebaut *et al*, 1997)

With the ground work that has been done, there is predictive power that can be combined with the output of genome-sequencing projects to find new and novel GHs. These interactions are beginning to produce great numbers of potential GH sequences, and the existing classification has numerous times provided useful mechanistic and structural information (Henrissat & Davies, 1997). I will next cover the assorted nomenclature presently used to characterize GHs before moving on to describe the GH family 73 which is the subject of this thesis.

Glycoside Hydrolase Mechanism Basics

Advances in DNA sequencing and 3D structure determination methods have greatly expanded our knowledge of GH enzymes. In addition, more meticulous characterisation of individual enzymes is possible with the development of sophisticated methods for identifying specific catalytic residues (Macleod et al., 1994; Namchuk & Withers, 1995; Withers & Aebersold, 1995). Accompanying these developments, the enzymatic hydrolysis of glycosidic bonds has in recent years been considerably characterized as well, especially in the case of understanding the molecular reaction at the active site, through structural studies, enzyme inhibition, and physical organic chemistry experiments (Vocadlo & Davies, 2008). Of the 132 GH families that exist, three-dimensional structures now exist for members of over 75 families and catalytic mechanisms for 85 families have been proposed (Lombard et al., 2014).

To understand the catalysis of glycosidic linkages, one must first understand the types of reactions that occur, the different locations at which the reactions take place, stereochemistry of products that arise, and the kinds of amino acid residues that participate. Due to polysaccharide heterogeneity, there is a large variety of GH

enzymes, yet despite this, enzymatic hydrolysis of glycosidic bonds, in general, occurs with a general acid catalysis mechanism which typically necessitates two key residues: a proton donor and a nucleophile/base (Koshland, 1953; McCarter & Withers, 1994; Sinnott, 1990). Hydrolysis proceeds along one of two major mechanisms resulting in either retention or inversion of anomeric configuration (Figure 3A). When glycoside hydrolases catalyze a hydrolysis reaction, the departure of the saccharide leaving group is facilitated by a general acid. A general base then enables nucleophilic attack at the anomeric carbon and depending on the mechanism, this can result in a product where the stereochemistry of the anomeric carbon is either inverted or retained with respect to the substrate (Figure 3B, C).

Briefly, inverting mechanisms utilize a single-step reaction involving displacement of the leaving group by water. The active site of configuration inverting enzymes use two side-chain carboxyl groups, where one acts as an acid to protonate the glycosidic oxygen of the scissile bond, while the other acts as a base to activate an intervening water molecule that attacks the anomeric carbon of the glycosidic bond, resulting in bond cleavage and a product with inverted anomeric stereochemistry (Figure 3B) (Zechel & Withers, 2000). Water plays a major role in this mechanism,

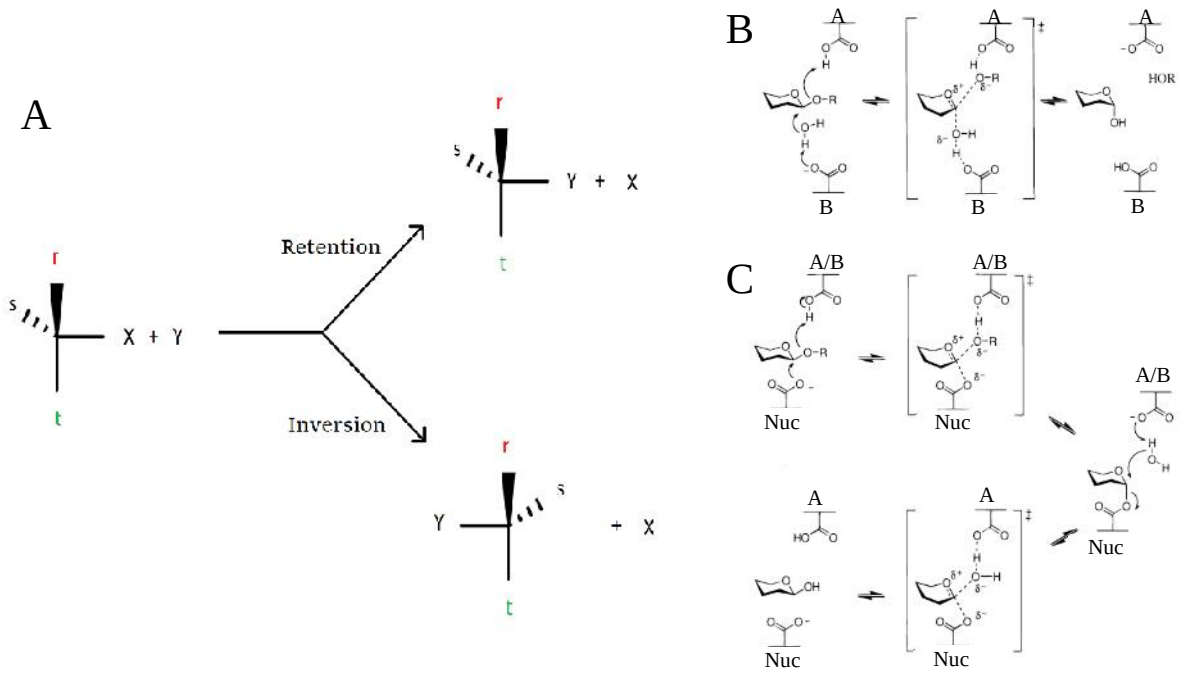


Figure 3. General Mechanisms of Glycoside Hydrolases A) Simplified explanation of the stereochemistry of inversion and retention. B) General mechanisms for inverting glycosidases. The glycosidic oxygen is protonated by the general acid (A) and the leaving of the aglycon group follows. This is followed closely by a general base-activated water molecule attack (B). This single step substitution results in a product with stereochemistry opposite to that of the substrate. C) Retaining glycosidases' mechanism which occurs with the glycosidic oxygen being protonated by the general acid (A), priming the anomeric centre attack by the nucleophile (Nuc). The ensuing glycosyl-enzyme is hydrolyzed by a water molecule in a second nucleophilic substitution, generating a product with the stereochemistry that is the same as the substrate. Transition states marked with ‡. (Figure reprinted with permission from Zechel *et al*, 2000. Copyright 2000 American Chemical Society)

acting as the major nucleophile upon being activated by the enzyme general base. As such, the distance between the general acid and base residues within the active site has been shown to be placed ~10.5 Å apart in inverting GH families where a water and substrate need to be positioned between the two catalytic residues (McCarter & Withers, 1994; Zechel & Withers, 2000). Alternatively, the configuration retaining mechanism utilizes a two-step, double-displacement reaction where, in the first step,

one enzymatic carboxyl group acts as an acid to protonate the glycosidic oxygen of the scissile bond while the other acts as a nucleophile that attacks the anomeric carbon of glycosidic bond (Figure 3C). This combined action cleaves the glycosidic bond and results in the formation of a glycosyl-enzyme intermediate (Zechel & Withers, 2000). In the second step, the carboxyl group that initially protonated the glycosidic oxygen now acts as a general base to activate an incoming water that attacks the anomeric center of the intermediate and yields a product with retained anomeric stereochemistry (Zechel & Withers, 2000). Given there are two inversions during the course of the double-displacement mechanism, the net result is retention of stereochemistry at C-1. Since the nucleophile residue interacts directly with the substrate, the distance has been observed to be a fairly consistent 5.5 A in retaining GHs (Wang et al., 1994). The transition states of both mechanisms occur with significant oxocarbenium ion character whereby nucleophilic attack causes a dissociation of C-1 to a planar sp^2 form and an alignment of lone pair oxygen electrons (McCarter & Withers, 1994; Sinnott, 1990; Zechel & Withers, 2000).

When describing the location of catalytic action within a polysaccharide chain, a subsite convention, developed by enzymologists, is used in order to describe the position of the different monosaccharide substituents (Davies et al., 1997). By convention, a -n, +n nomenclature is used whereby the cleavage takes place between -1 and +1 subsites. Any other residues toward either the non-reducing end are labelled sequentially into minus subsites and any residues toward the reducing end are labelled accordingly in increasing positive integers. Such a convention is necessary for providing reference within a polysaccharide chain and within the active sites within saccharide-metabolizing enzymes (Davies et al., 1997). The C-1 anomeric carbon described throughout belongs to the -1 subsite within a polysaccharide chain.

Characterizing Glycoside Hydrolases

A main feature of GH enzymatic reactions is the nature of the intermediate step of the catalytic reaction. Most retaining glycosidases possess a mechanism with a covalently bonded intermediate (Davies & Henrissat, 1995; McCarter & Withers, 1994; Wang et al., 1994), with certain studies having shown structural evidence of a glycosyl-enzyme intermediate with a covalently bonded substrate molecule (figure 4) (Uitdehaag et al., 1999; Vocadlo et al., 2001), though the first evidence was shown with the use of KIEs, or kinetic isotope effects (Sinnott & Souchard, 1973). This method requires two

near identical substrates for the particular enzyme being studied. One substrate is comprised of the common, light isotope, while the second substrate used is made with a different stable isotope, typically a less abundant 'heavy' isotope (^{2}H, ^{13}O, ^{15}N, or ^{18}O). Simply, KIEs are the differences in rates at which the enzyme processes the substrate under two near-identical conditions; one reaction rate measured with substrate comprised of the light isotope, divided by the rate at which substrate with the heavy isotope is processed (k_{light}/k_{heavy}= KIE). The slight difference in reaction rates has to do with different bond vibrational energies between the isotopes and the group they are bound to. Heavier atoms, especially ones in which the percentage difference in mass is greatest compared to the light isotope, as in deuterium/hydrogen, have decreased ground state vibrational frequencies relative to lighter atoms (figure 5) (Vocadlo & Davies, 2008). More energy is required to cleave a bond that contains a heavy isotope as opposed to one that contains only light isotopes, hence the difference in reaction rates (Vocadlo & Davies, 2008).

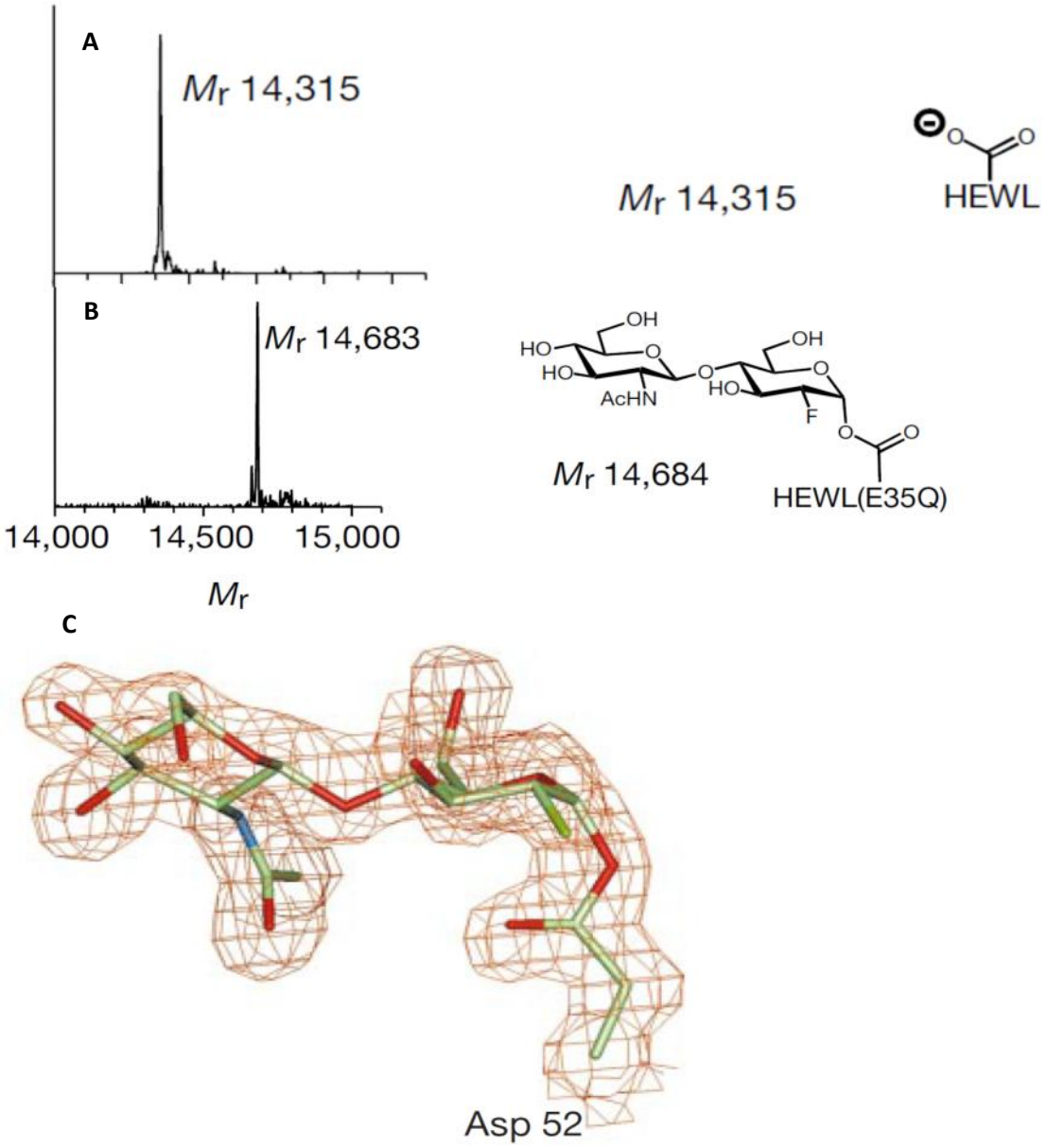


Figure 4 Enzyme-Substrate Complexes. ESI-MS mass spectra of HEWL (Hen Egg-White Lysozyme) and HEWL complexes. A) Wildtype HEWL. B) HEWL incubated with NAG2F, a substrate with fluorine in place of the 2-acetamido group which has a reduced rate of release, essentially a bound inhibitor (Barnett *et al.*, 1967). C) The corresponding crystal structure cartoon shows the covalently bound substrate to the acid residue. (figure reproduced with permission from Vocadlo *et al,* 2001)

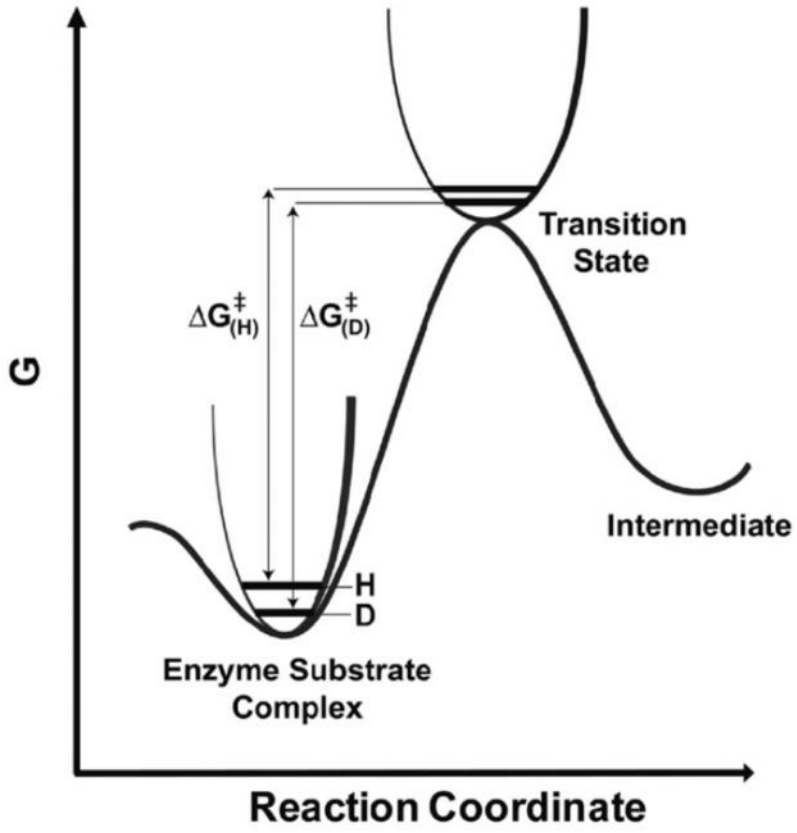


Figure 5. Kinetic Isotope Effect Diagram. Simplified reaction coordinate highlighting the difference in energies when heavy isotopes are used. A larger activation energy is seen with heavier atoms (D, deuterated) between ground state and transition state leading to a decrease in corresponding reaction rates which can be observed and measured. (figure reproduced with permission from Vocadlo *et al,* 2008)

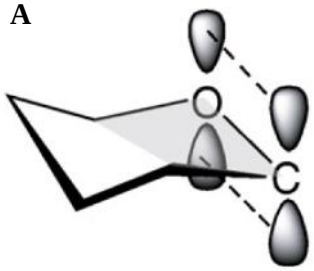
Observations involving glucosidases and deuterium labelled substrates show a larger KIE for deglycosylation than glycosylation and given the difference, it is suggestive of a covalently bonded intermediate and S_N2 oxocarbenium transition states (Kempton & Withers, 1992; Umezurike, 1988; Van Doorslaer et al., 1984). The dissociative oxocarbenium transition states have additional significance when considering the planar geometry the anomeric carbon has during its transition from binding to leaving group, to binding nucleophile (in both retaining and inverting

mechanisms). In particular, a sp^2-hybridized anomeric carbon has certain stereochemical characteristics that can give insight into mechanism, this will be discussed next.

Substrate Distortion During Hydrolysis

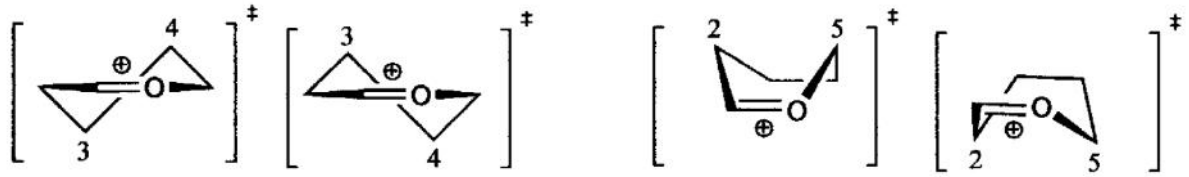
In addition to the nature of the catalytic intermediate and stereochemical outcome of a GH catalytic mechanism, a third feature of GH kinetics is the substrate distortion that occurs during catalysis. Within saccharides characterized by a six member hexose ring, the oxocarbenium-like transition states of the -1 subsite are stabilized by lone-pair electron donation from the endocyclic oxygen across the O-5-C-1 bond (figure 6A) (Vocadlo & Davies, 2008). This gives partial double-bond character to the structure which demands that the transition state C-5, O-5, C-1 and C-2 must be co-planar; those conformations are seen only for the six member pyranose ring in two half-chair ($^4H_3$ and $^3H_4$) and two classical boat ($^{2,5}B$ and $B_{2,5}$) conformations (figure 6B) (Davies et al., 1997). Additional consequences of the pyranose ring being distorted away from its relaxed conformation, besides permitting direct in-line nucleophilic attack, is the minimization of potential steric hindrances (Vocadlo & Davies, 2008).

A



$^4H_3$

B



$^4H_3$ $^3H_4$ $^{2,5}B$ $B_{2,5}$

Figure 6. Transition State Conformations. A) Activation of the anomeric center. Stabilized positive charge by the oxygen lone pair of the ring oxygen prefers the adoption of certain conformations with planar geometry. (Figure reproduced with permission from Vocadlo and Davies, 2008) B) Showing the transition states (‡) which possess strong oxocarbenium-ion-like character with planar geometry. Such transition states should display one of four possible conformations $^4H_3$ and $^3H_4$ half-chairs and $^{2,5}B$ and $B_{2,5}$ boats. (Figure reproduced with permission from Davies *et al*, 2003)

If the case where two species, substrate and product, are known in terms of their stereochemical conformation, one could potentially predict transition state structure conformations using Stoddard's classic carbohydrate stereochemistry (Stoddart, 1971) which relies on the condition of planar geometry for the intermediate due to delocalization across the O-5-C-1 bond (figure 7) (Davies et al., 2003). Based on classic carbohydrate stereochemistry, one structure can pseudorotate into another stable structure only through a strained, planar intermediate (Stoddart, 1971).

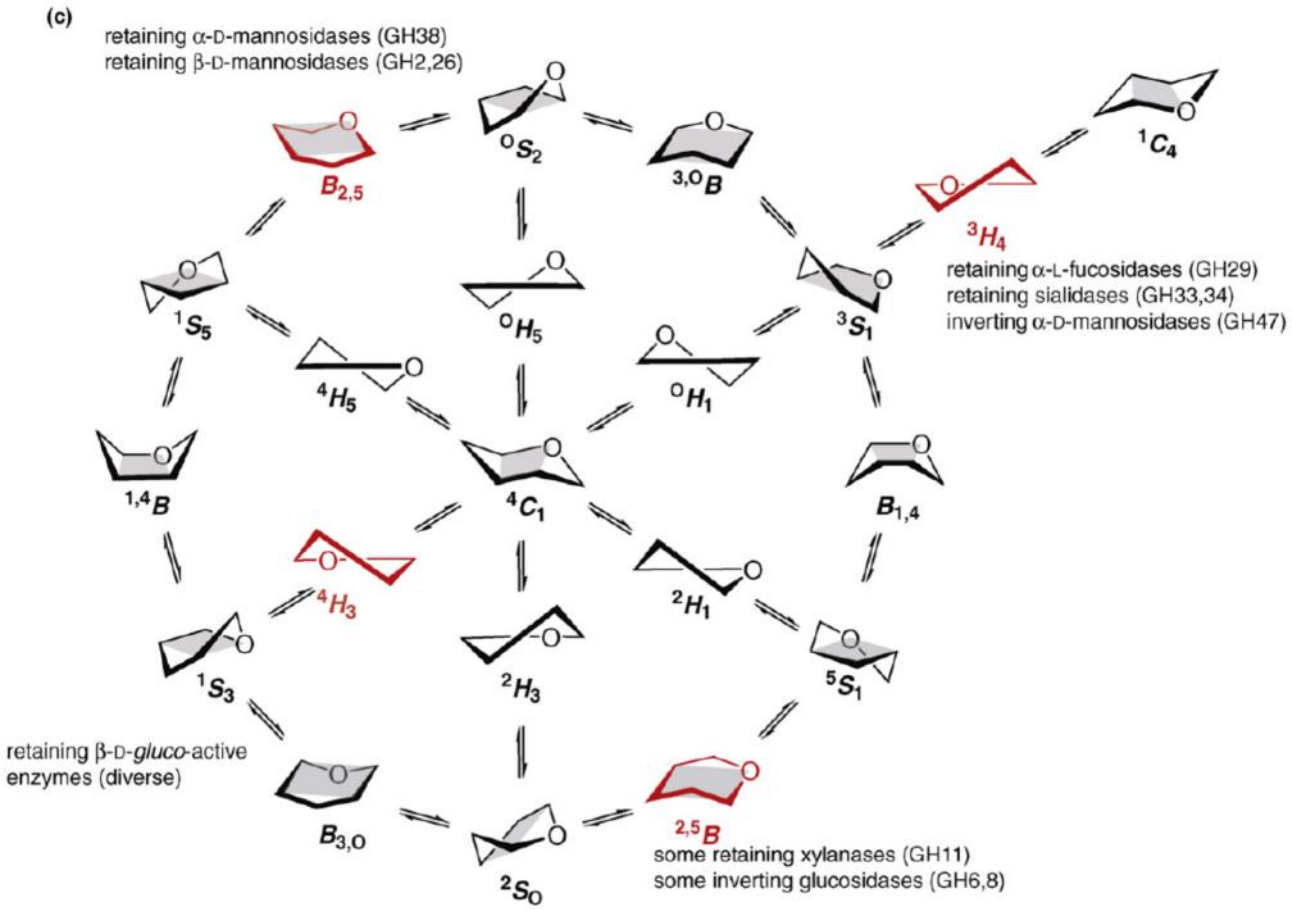


Figure 7. Inter-conversion itinerary with the four common transition state conformations highlighted in red and some currently known conformation pathways noted. (figure reproduced with permission from Vocadlo and Davies, 2008)

The observed substrate distortions can also be thought of as the anomeric center 'migrating' (Vocadlo & Davies, 2008). The conformational changes of the sugar during catalysis causes a shift of the anomeric center from one side of the plane with the leaving group, passing through the high energy sp^2 transition state before finally collapsing into the side of the plane with the a catalytic center of the glycoside hydrolase (figure 8) (Davies et al., 1998).

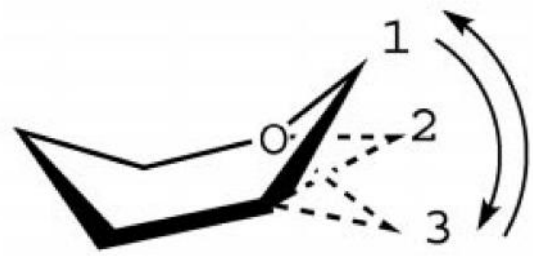


Figure 8. Anomeric Center Migration. The anomeric center migrates from a position of being bonded to the leaving group (1), and moves through the plane of the pyronose molecule. It is this halfway position that O-5, C-1 and C-1 are co-planar (2). This transition state then collapses to a position on the opposite side of the plane of the pyranose ring (3) (Figure reproduced with permission from Mark *et al,* 2001)

Other Glycoside Hydrolase Considerations

The nature of the catalytic nucleophile may also vary from the common acid/base carboxylate residue. The residue is most often an enzyme-derived carboxylate however certain families such as GH families 33, 34, and 83, called sialidases, all achieve catalysis with a retaining mechanism involving a tyrosine (figure 9) (Vocadlo & Davies, 2008). Mass spectrometry, site directed mutatgenesis, structural data of covalently bound fluoro-sugars, and KIE studies all provide good support that the reaction proceeds via a catalytically active tyrosine (Amaya et al., 2004; Watson et al., 2003; Watts et al., 2003; Yang et al., 2000), and on this basis those families have been grouped into a clan of hydrolases called GH-E (Vocadlo & Davies, 2008).

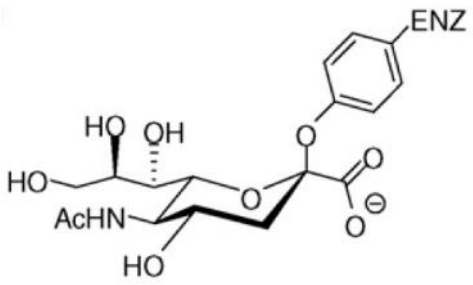


Figure 9 Trans-sialidases. Enzymic tyrosine forms a covalent glycosyl-enzyme intermediate. (Figure reproduced with permission from Vocadlo and Davies, 2008)

A third group of GHs with differing mechanisms are those that exploit a neighbouring acetamido group at C-2 on the substrate, which is positioned next to the anomeric center of the glycosidic bond to be hydrolyzed and is used as a nucleophile in a double displacement mechanism (Vocadlo & Davies, 2008). This mechanism is termed substrate assisted catalysis. The 2-acetamido group is polarized and oriented by the enzyme to react with the anomeric center, forming a bond between the carbonyl oxygen of the acetamido and the anomeric center effectively displacing the leaving group in the first step of a retaining mechanism with a resulting oxazoline intermediate (figure 10) (Macauley et al., 2005; Mark et al., 2001). One consequence of this mechanism is there is no covalent bond between substrate and enzyme at any point along the reaction coordinate though net retention of stereochemistry occurs (Macauley et al., 2005). These types of GHs are most researched in family 84 seeing as they have roles in the etiology of several diseases in humans; functioning in cleaving terminal N-acetylglucosamine residues from glycoconjugates (Liu et al., 2004; Vosseller et al., 2002). Families 18 and 20 also have the 2-acetamido group play the role of nucleophile and proceed through an oxazoline intermediate (Knapp et al., 1996; Mark et al., 2001; Terwisscha van Scheltinga et al., 1995). Typically the acetamido group is polarized and oriented by an enzymatic residue to attack the anomeric center

(Vocadlo & Davies, 2008). Also worth noting, an acetamido group adjacent to the anomeric center does not necessitate substrate-assisted catalysis. Hen egg white lysozyme from GH family 22 is one such GH enzyme (Vocadlo et al., 2001), as well as β-N-acetylglucosaminidases from family GH3 (Vötsch & Templin, 2000), which both proceed through a glycosyl-enzyme intermediate even though their substrates contain a suitable positioned 2-acetamido group.

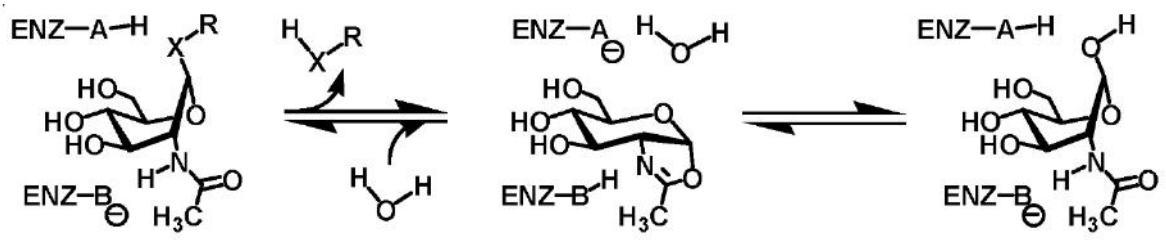


Figure 10 Substrate assisted catalysis. The 2-acetamido group forms a transient oxazoline intermediate. (Figure reproduced with permission from Macauley *et al*, 2005)

Glycoside Hydrolase Family 73

Glycosidase family GH73 hydrolases cleave the β-1,4 glycosidic linkage between the polymeric aminosugars N-acetylglucosaminyl (GlcNAc) and N-acetylmuramyl (MurNAc). Polymers of these two sugars form the glycan strands in bacterial cell wall peptidoglycan, also known as murein. Peptidoglycan (PG) is the major cell wall component in Gram negative and positive bacteria and is primarily responsible for

providing strength and rigidity to the bacterial cell. Up to 70% w/w of the cell wall of Gram positive bacteria is PG, whereas less than 10% of the cell wall of gram negative is comprised of PG (Schleifer & Kandler, 1972). The glycan stands of PG are crosslinked by short peptides, making the PG layer a single, flexible three-dimensional macromolecular structure that completely encases the microbe (Schleifer & Kandler, 1972). Many bacterial processes such as: growth, PG turnover, cell separation, protein secretion, formation of secretion systems, flagellar assembly, genetic competence, and sporulation require controlled PG remodelling by PG hydrolases, termed autolysins (Holtje, 1998; Neuhaus & Baddiley, 2003; Smith et al., 2000; Vollmer et al., 2008). Some processes requiring PG hydrolysis are carried out, in part, by GH73 autolysins (Cabanes et al., 2004; Camiade et al., 2010; Eckert et al., 2006; Nambu et al., 1999). Categorization of autolysins is made according to the type of PG bonds cleaved. N-acetylglucosaminidases, N-acetylmuramidases and lytic transglycosylases, endopeptidases, and amidases will all cleave bonds within the PG polymer with different specificities (Vollmer et al., 2008). GH73 enzymes are β-*N*-acetylglucosaminidases specifically, hydrolyzing the N-acetylglucosaminyl-β-1,4-N-acetylmuramyl linkage between GlcNAc and MurNAc (figure 11) (Bublitz et al., 2009; De Las Rivas et al., 2002; Herlihey et al., 2014; Lipski et al., 2015).

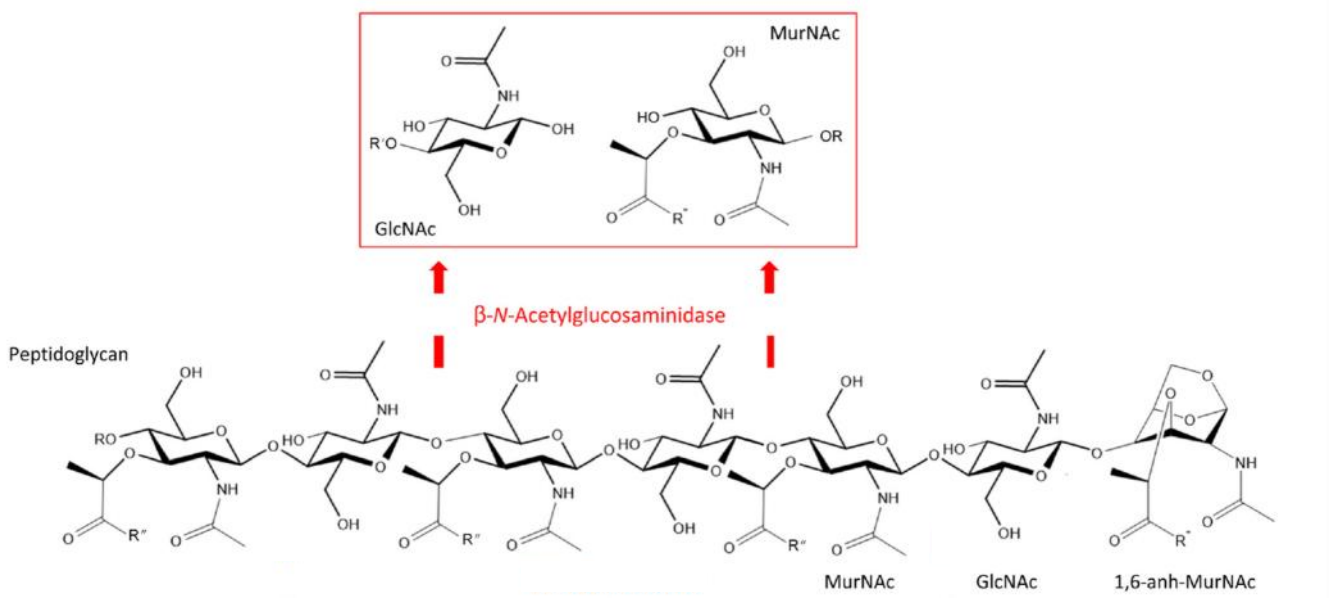


Figure 11 The β-*N*-acetylglucosaminidases hydrolyze the β-1,4 linkage between GlcNAc and MurNAc residues in PG (Figure reproduced with permission from Herlihey *et al*, 2014).

While the substrate specificity of many GH73 enzymes has been determined, details about their catalytic mechanism remain unclear with insufficient structural or kinetic data to support a defined mechanism of action. Structural studies of some GH73 enzymes have been carried out, and a handful of members have either known structures, biochemical data, kinetics data, and/or a catalytic mechanism proposed. Despite this, there is surprising diversity being uncovered in studying this family, namely what appears to be two possible mechanisms (inverting or substrate assisted catalysis)(Lipski et al., 2015). In addition, there are incomplete areas with regards to structural research on GH73 enzymes, particularly the isolation of a bound substrate

molecule in the active site, which would provide powerful insight into how the enzymatic reaction proceeds. Those GH73 members with the most characterization include: Auto, FlgJ, LytB, AcmA and Atl(WM), and TM0633. Each enzyme has structural or biochemical data that has contributed to a description of family GH73. The functions of those characterized members are: the surface-associated autolysin and virulence factor Auto from *Listeria monocytogenes* (Bublitz et al., 2009)*;* the flagellar protein FlgJ from *Sphingomonas sp.* (Hashimoto et al., 2009; Herlihey et al., 2014); LytB, a *Streptococcus pneumoniae* glycosidase functioning at the septum to separate the cell walls of two daughter cells (Bai et al., 2014), AcmA from *Lactococcus lactis*, responsible for lysis in the stationary phase and thought to be required for cell separation (Inagaki et al., 2009); Atl_{WM} encoded by *Staphylococcus warneri* M with unknown function (Yokoi et al., 2008) and most recently TM0633, a glucosaminidase from *Thermotoga maritima,* a hyperthermophilic bacterium (Lipski et al., 2015). Based on the recent work that has been done on the family, it has become apparent that the GH73 may contain some members which undergo what appears to be a classic single displacement configuration inverting mechanism, while others potentially use substrate assistance in a two-step configuration retaining mechanism (given that GlcNAc bound in the -1 subsite contains an appropriate 2-acetomido group). This unexplored aspect to the GH73 family grants justification for research, as well as the

fact that more than a few of the characterized GH73 enzymes were found to be critical for infection: LytB, for example, is critical for pneumococcal cell division and nasal colonization (Garcia et al., 1999; Gosink et al., 2000); Auto is essential for pathogenic listeriosis caused by *Listeria monocytogenes* (Cabanes et al., 2004); and FlgJ is involved in biogenesis and function of the flagellum which is a major virulence factor of motile pathogenic bacteria (Moens & Vanderleyden, 1996; Nambu et al., 1999).

GH73 enzymes are possibly best classified as being related based on their conserved catalytic active site residues. An invariant glutamate has been identified as the catalytic general acid (Bai et al., 2014; Bublitz et al., 2009; Herlihey et al., 2014; Inagaki et al., 2009; Lipski et al., 2015; Maruyama et al., 2010; Yokoi et al., 2008). The first in depth structural studies, two crystallographic papers published in 2009 for Auto and FlgJ, drew strong resemblances to other GH families. The overall structures around the active site was describe as a lysozyme-like fold, based on its first characterization of eukaryotic lysozymes within GH families 22 and 23 (Phillips, 1967; Weaver et al., 1995). This fold can be described as a mostly globular enzyme made up of at least 5 α-helices and a β-sheet hairpin subdomain in GH73. The β-hairpin forms a deep groove that completes the active site, especially so for certain family GH73 enzymes with β-hairpins that appear to be appreciably longer than related lysozyme-

like GH families (Bublitz et al., 2009; Maruyama et al., 2010). Strong fold similarities were noted to members of GH23, a family containing hydrolytic PG lysozymes acting as N-acetylmuramidases and lytic transglycosylases (cleaving the MurNAc-β (1,4)-GlcNAc glycosidic bond) as well as GH19 chitinases (which cleave between linked GlcNAc residues) (Bublitz et al., 2009; Hashimoto et al., 2009). It was particularly interesting for Auto and FlgJ to be structurally related to enzymes with such contrasting functions like soluble lytic transglycosylase 70 (Slt70) from *E. coli*, and eukaryotic goose egg white lysozyme (GEWL), both GH23 (Figure 12) (Bublitz et al., 2009; Maruyama et al., 2010). Analysis of GH73 function has shown that despite having a lysozyme-like fold, which is typically associated with muramidases, GH73 enzymes differed in function, instead targeting the GlcNAc-β(1,4)-MurNAc bond of PG as β-*N*-acetylglucosaminidases (Bublitz et al., 2009; Herlihey et al., 2014). Within the active site of Auto, residue Glu122, the putative catalytic acid/base residue, was seen to occupy a position analogous to the catalytic residues Glu478 in c-Slt70, and Glu73 in GEWL, and confirmed to be catalytically necessary with a knock-out variant (Bublitz et al., 2009). Thus, this Glu residue in Auto has been predicted to be the catalytic general acid; the lysozyme fold therefore seems to be shared across glycoside hydrolases that are functionally diverse. In addition to the conserved glutamate general acid/base residue, other active site restudies within GH73 enzymes include a set of three

aromatic residues. In Auto, two tyrosines and a third hydrophobic residue, phenylalanine, surround its general acid/base (Glu122) (in much the same way as the GH23 enzymes (c-Slt70 and GEWL) use a hydrophobic methionine) to create a hydrophobic environment that presumably increases the protonation of the catalytic glutamate carboxyl group through a rise in its pKa (Bublitz et al., 2009). The similarities of Auto and the GH23 enzyme residues end there however, as other analogous residues are not conserved. Nonetheless, a GH73 sequence alignment specifies that a second glutamate, Glu156, is conserved in these enzymes, located on a conserved β-hairpin (figure 13) (Bublitz et al., 2009). Located at a distance of ~13 Å from Glu122 in Auto, this second carboxylate residue was shown to be necessary for catalysis, possibly participating in a configuration inverting catalytic mechanism with Glu122 (Bublitz et al., 2009). The significance of this additional residue will be described in detail below for Auto and other GH73 enzymes.

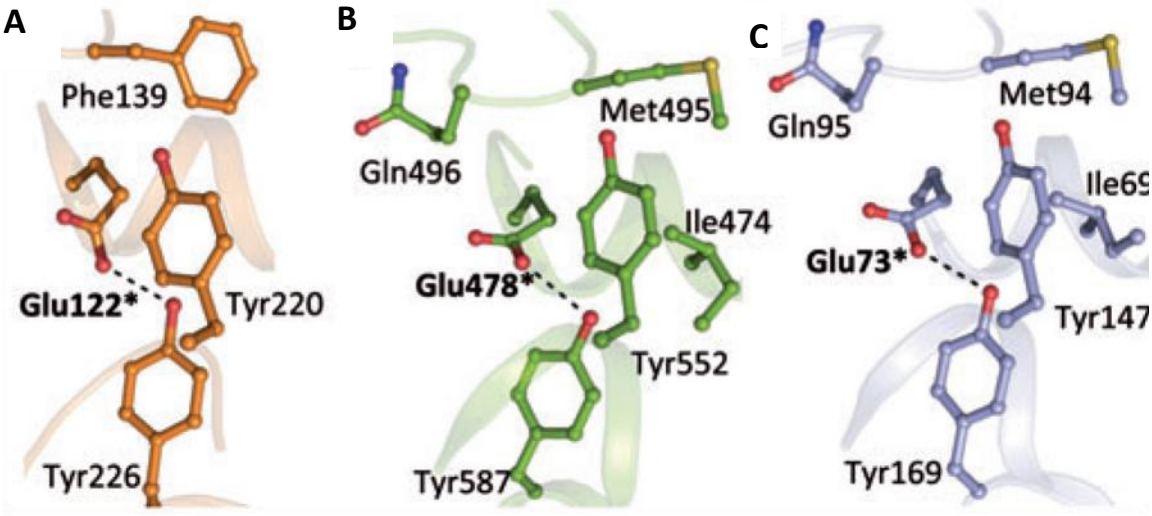


Figure 12. Spatial Conservation of Related GH Catalytic Residues. The active sites of A) Auto, B) c-Slt70 (lytic transglycosylase from *E. coli*) and C) goose egg-white lysozyme. The catalytic glutamate is designated with *, also seen are the conserved hydrophobic residues. (Figure reproduced with permission from from Bublitz *et al,* 2009).

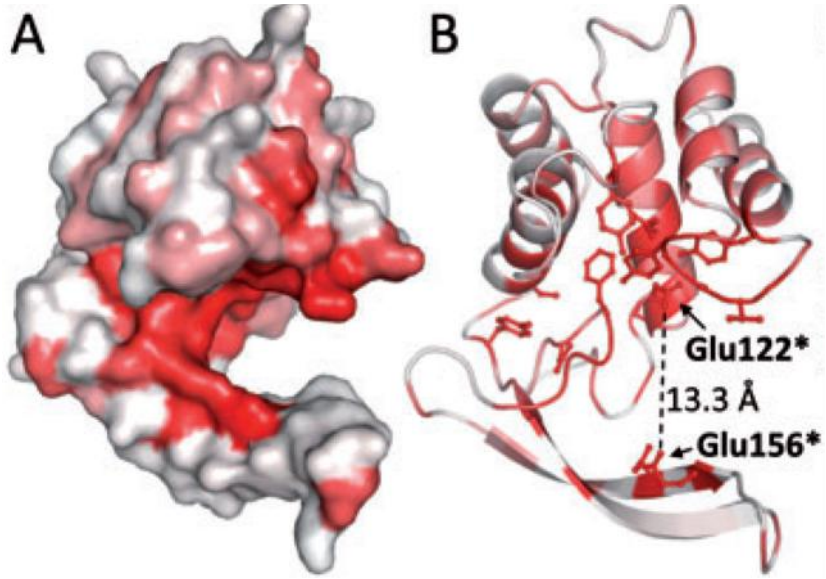


Figure 13. Conservation of Active Site Sequence in GH73 Enzymes. 458 family GH73 sequences were compared using ClustalW and residues of low (white), medium (pink), and high (red) sequence similarity were mapped over the cartoon structure of Auto. Catalytic glutamates designated with * (Figure reproduced with permission from Bublitz *et al,* 2009).

The second GH73 paper published in 2009, also claiming to be the first structural determination of a GH73 family peptidoglycan hydrolase, was of flagellar

protein FlgJ from the Gram-negative bacterium *Sphingomonas sp* (Hashimoto et al., 2009). This structural study again noticed a similarity to other GH families, especially a conservation of a glutamate general-acid residue (Glu185) and a long β-hairpin fold reminiscent of the distantly-related GH23 lysozymes and lytic transglycosylases (Hashimoto et al., 2009). The lysozyme fold therefore appears to support at three distinct but related catalytic functions (muralytic, chitinase and glucosaminidase function) and appears to be a co-evolved feature in eukaryotes and bacteria that acts as a scaffold for which to hydrolyze PG at various locations (Bublitz et al., 2009).

In a follow up study on FlgJ (Maruyama et al., 2010), point mutations of the conserved general-acid/base glutamate on the alpha-helical domain (Glu185), the conserved glutamate on the β-hairpin predicted to serve a role in catalysis as a general base in an inverting mechanism (Glu224), and a conserved tyrosine (Tyr281) located in the active site, were all shown to cause a drastic decrease in catalytic activity, further reinforcing their importance in catalysis and the use of an inverting mechanism by GH73 enzymes (Maruyama et al., 2010). Also noted was that the primary structure of the β-hairpin differed among different GH73 members. *S. warneri* M autolysin Atl_{WN} has a large deletion in the β-strands of the hairpin, making it much shorter than the analogous hairpin of Auto and FlgJ. This difference in β-hairpin structure was

proposed as an explanation as to why the corresponding conserved residue to FlgJ Glu224, Asp1275 in Atl_{WM}, retained its activity upon point mutation to a catalytically inert residue (Glu1275->Ala) (Maruyama et al., 2010; Yokoi et al., 2008). These findings, that there exists residual activity when a putative base/nucleophile is converted to a non-catalytic residue, point to potential catalysis involving assistance by the 2-acetamido group in a substrate-assisted mechanism within Atl_{WM} (Maruyama et al., 2010). It was also noted that the primary structure of the β-hairpin in Atl_{WM} was more similar to the GH73 autolysin Lyt-B from *S. pneumoniae* as oppose to FlgJ and Auto, which share a more extended β-hairpin that bears catalytically-necessary residues (Maruyama et al., 2010).

Recently, Lyt-B was characterized and proved to be characteristically different in some ways from previously structurally characterized GH73 enzymes (Bai et al., 2014). This enzyme, which acts on immature PG at cell-separation sites, showed marked similarities in the active site for the general acid glutamate and surrounding hydrophobic residues to FlgJ and Auto (figure 14).

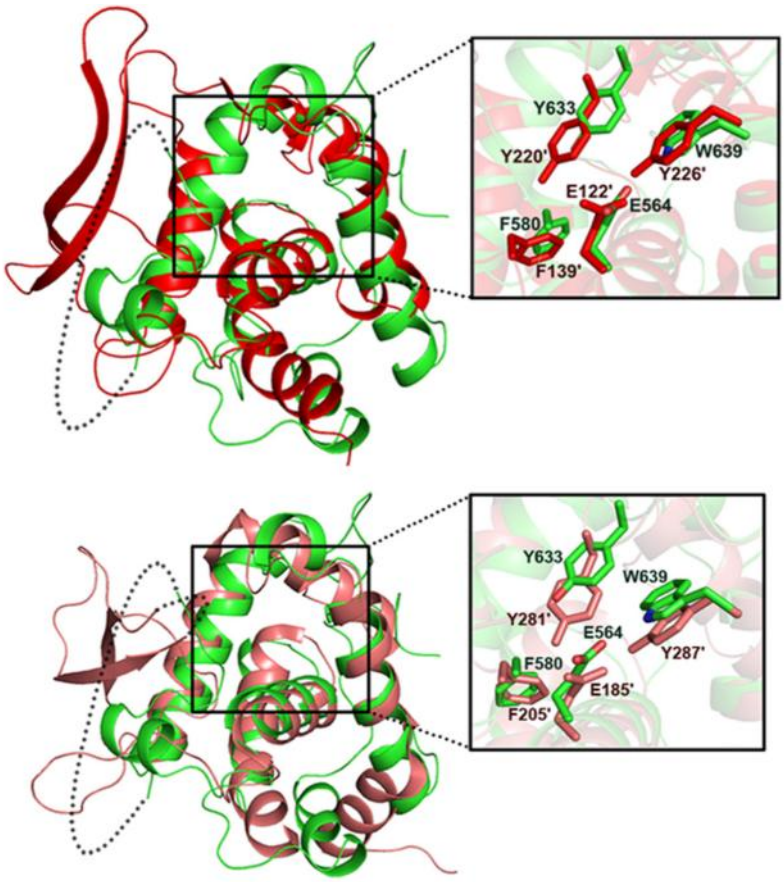


Figure 14. Superimposition of Lyt-B (green) over GH73 enzymes Auto (red) and FlgJ (pink) showing overall structure similarities and, more importantly, analogous location of key catalytic residues. The missing β-fold around the binding grove is denoted with a grey dotted connection.

(Figure reproduced with permission from Bai *et al*, 2014)

Unfortunately, the structure of the β-hairpin around the binding grove was not visible in the crystal structure due to its flexibility, causing it to be disordered with inadequate density for a structure. Distinct from other GH73 enzymes Lyt-B required its other modules in order to be catalytically active. Multiple module organization of GH73 enzymes is not uncommon however, though in other studies the GH73 catalytic modules alone were functional by themselves, either expressed unaccompanied by other modules or they naturally are non-modular (Bublitz et al., 2009; Lipski et al., 2015; Maruyama et al., 2010).

The β-Hairpin and the Identity of a GH73 Nucleophile

To further add diversity to the GH73 family, the identification of a catalytic nucleophile/general-base residue has been inconclusive with varying results obtained from different mutational studies. As already mentioned, studies in which a β-hairpin houses a conserved and necessary glutamic acid have supported a single displacement, inverting, mechanism of action; indeed, mutation of the β-hairpin-located carboxylates has yielded abrogation of catalysis (figure 15), particularly in Auto, FlgJ and TM06330 (Bublitz et al., 2009; Herlihey et al., 2014; Lipski et al., 2015; Maruyama et al., 2010). Other studies with Atl_{WM}, and AcmA have pointed to GH73 enzymes in which residual activity accompanies the same type of mutations (Inagaki et al., 2009; Yokoi et al., 2008). These findings point to a probable mechanism involving substrate assistance, due to a lack of a conserved residue which might play the role of a nucleophile.

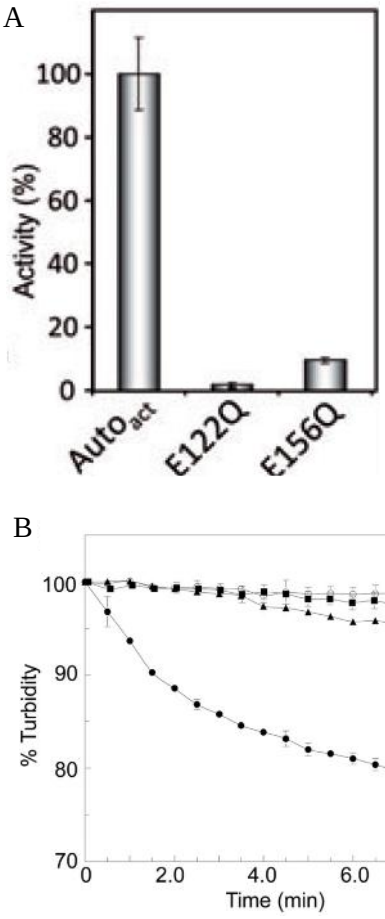


Figure 15 A) Auto mutational analysis showing abrogation of activity with a mutation to non-polar, non-carboxylate reside of two active site glutamate residues. E122 is the catalytic residue in the active site pocket and E156 is located on the β-hairpin. B) Similar assay performed on FlgJ with no enzyme added; open circle, wild-type FlgJC; circle, a mutation in the β-hairpin catalytic residue; square, and a mutation in the active site catalytic residue; triangle. (Figures reproduced with permission from Bublitz *et al*, 2009 and Herlihey *et al*, 2014)

Recently, a phylogenetic tree was constructed showing how the 2800 most recent theoretical GH73 sequences fit into five sequence clusters among the three different phyla (figure 16) (Lipski et al., 2015). Sequences belonging primarily to proteobacteria were assigned into clusters 1 and 5, clusters 2 and 4 contain sequences from Firmicutes, while cluster 3 belongs to Bacteroidetes. FlgJ, Auto and TM0633 belong to a group of three clusters: 1, 2 and 3, that conserve two carboxylate residues involved in a potential inverting acid/base mechanism. The other GH73 enzymes with

potential substrate assisted mechanisms, for which mutation of the a secondary conserved carboxylate residue does not decrease enzyme activity, characterized in AcmA and Atl_{WM}, grouped based on homology into clusters 2, 4 and 5. Cluster 2 does indeed seem to have sequences that may belong to either proposed mechanism. Molecular modeling of representative members' primary sequences from each cluster showed a variability in the region above the active site, in the β-hairpin (Lipski et al., 2015). It was suggested that the β-hairpin permits specific ligand binding and may modulate the catalytic mechanisms within the GH73 family with a variability in the length of the fold. Future studies will aim to further investigate similarities and differences between the two possible mechanisms of action, characterise how β-hairpin length functions in specificity, and produce a structure with ligand bound so that a concrete mechanism can be proposed for the different types of GH73 enzymes.

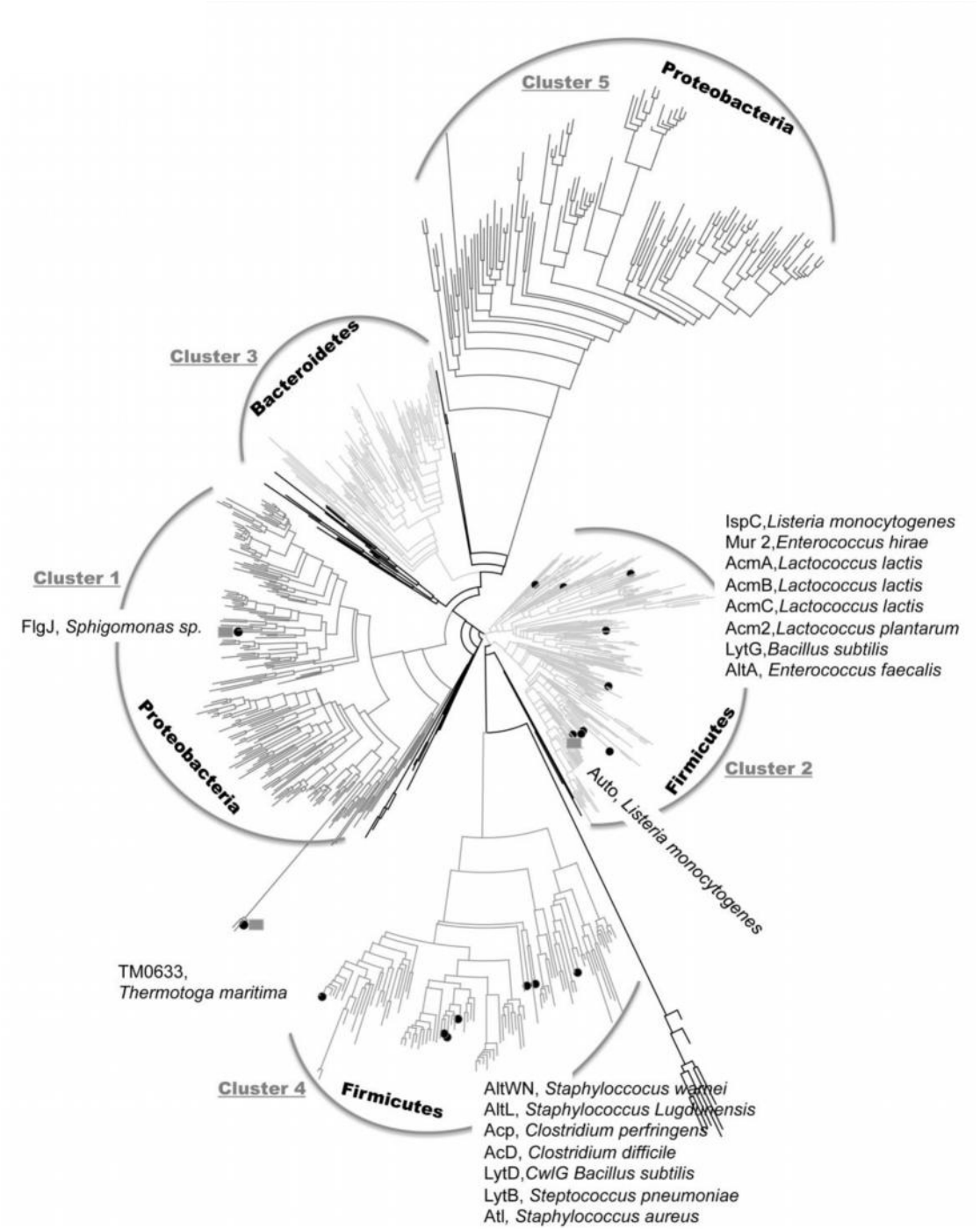


Figure 16. Phylogenetic Analysis of Family GH73. The branches for cluster 1 through 5 are shown. Within each cluster, sequences are organized by phylum. Enzymes which have been characterized are labeled and with a black circle and GH73 enzymes with structural characterization are labeled with grey rectangles. (Figures reproduced with permission from Lipski *et al*. 2014)

Structure Determination and Activity Investigation of *Salmonella typhimurium* Glycoside Hydrolase FlgJ

The primary work of my thesis is based on a structural characterization of FlgJ from *Salmonella typhimurium* (*St*FlgJ). The protein has been crystallized previously (Kikuchi et al., 2009), though no structure had been deposited into the Protein Data Bank and thus no structural information relating to this specific FlgJ protein was available. As mentioned previously, FlgJ is involved in biogenesis and function of the flagellum with the flagellum acting as a major virulence factor for motile pathogenic bacteria (Nambu et al., 1999). FlgJ was shown to be an essential protein for the formation of the flagellum. During synthesis, component proteins synthesized in the cytoplasm are transported through a central flagellar canal upon being added to the growing rod, hook, and filament structures. Formation of the rod relies upon penetration through the bacterial PG layer of the cell wall and thus remodelling enzymes like FlgJ are necessary (Moens & Vanderleyden, 1996). To explore the structural basis of the catalytic mechanism of FlgJ, I crystallized the GH domain of FlgJ from *Salmonella enterica* subsp. *typhimurium* and determined its molecular structure using an in-house SAD-phasing experiment with assistance from PhD student Ben Bailey-Elkin in the Mark Lab. The ultimate goal of my research was to determine the structure of the FlgJ GH domain bound to a natural substrate or substrate analogue to

understand how active site residues interact with substrate to promote catalysis. The true nature of catalysis among GH73 enzymes has yet to be established due in part to a lack of structural studies of these enzyme bound to substrate or mechanism-based inhibitors. Thus far, no structures have been published that have an exposed active site with ligand molecules bound, though groups have attempted to obtain such structures (Maruyama et al., 2010; Lipski et al., 2015). One limitation is the complex nature of the PG substrate. Cell wall PG, and the lytic products, are complex in terms of cross-linking, stem peptide structure, and glycan chain modifications, all of which contribute to substrate specificity for autolysins (Herlihey et al., 2014; Kempton & Withers, 1992). Thus the range of possible substrates recognized by GH73 enzymes is large and the size of PG fragment that is bound by GH73 enzymes remains unclear in many cases. Another obstacle encountered in some GH73 studies, including this one, is the occlusion of the enzyme active site by auto-inhibition loops or C-terminal structures of the GH domain itself, which have been found to block the active site of GH73 enzymes within the crystal lattice (Bublitz et al., 2009; Hashimoto et al., 2009). Consistent with these previous findings, the crystal structure of the full-length *St*FlgJ GH domain described here was found to have its active site occluded by the C-terminal α-helix of a neighbouring symmetry mate within the crystal lattice. Fortunately, I was able to recrystallize the domain after removing the C-terminal helix, and another

crystal structure of the domain was obtained by SAD-phasing. This second structure had occlusion of the active site as well due to a β-hairpin sitting within the active site. This truncated mutant did however have suitable electron density to build the β-hairpin, uncommon with GH73 enzymes in which the β-hairpin is often disordered. The position of the structurally ordered β-hairpin in the C-terminally truncated FlgJ structure lends support for a β-hairpin that has a high degree of flexibility and is involved in a highly dynamic catalytic mechanism that requires large conformational changes of the β-hairpin. The structures, though only achieving a penultimate goal, are the first FlgJ protein structure characterized for clinically relevant *S. typhimurium*. In addition they provide credit toward a tertiary structure exposition, namely of the flexible β-hairpin that may function to dynamically bind substrate.

Chapter 2
Materials and Methods

The open reading frame (ORF) for the C-terminal glycoside hydrolase domain of *St*FlgJ (residues 151-316, NCBI Gene ID: 1252700) was originally cloned into an expression vector by Ms. Miriam Derkson (a 2013 summer student in the Mark Lab) and provided to me to initiate my thesis research. The cloning procedure was carried out as follows: The *St*FlgJ GH domain ORF was amplified by PCR from genomic DNA isolated from *Salmonella enterica subsp. enterica serovar Typhimurium str. LT2* (B) using primers 5'-GATATACATATGGACAGTAAAGACTTTCTGGCC-3' and 5'-GATATAGGATCCAAAGAGATTGTCGAGATTCGC-3'. The primers introduced NdeI and BamHI restriction sites at the 5' and 3' ends of the PCR amplicon, respectively, to facilitate ligation of the DNA into a modified pET28b(+) vector that fuses a His_6 purification tag in-frame to the 3'-end of the GH domain open reading frame. The resulting expression vector, pET28-*St*$FlgJ_{GH(151-316)}$, was used to transform *Escherichia coli* BL21 Gold (DE3) cells. The cells were then grown at 37˚C to an optical density (OD_{600nm}) of 0.4-0.6 in 500 ml LB medium supplemented with 35 µg/mL kanamycin. The culture temperature was subsequently lowered to 28 ˚C and recombinant protein expression induced with the addition of 1 mM isopropyl-D-thiogalactoside (IPTG) for 4 hours with shaking. Cells were then pelleted by centrifugation (4000 rpm, 20 min at

4˚C) and either stored at -80˚C or immediately re-suspended in 20 mL ice cold lysis buffer (1M NaCl, 50 mM Tris-HCl pH 8.0 with 0.1 mM phenylmethylsulfonyl fluoride) for cell disruption. Cells were lysed via French press (Aminco) and the soluble fraction was recovered by centrifugation (20000 rpm, 30 min at 4˚C). *St*FlgJ$_{GH(151-316)}$ was isolated by affinity chromatography using Ni-NTA resin (Qiagen) pre-equilibrated with the lysis buffer. Recombinant protein was bound to the resin over a 20 min incubation period with agitation on ice, and then transferred to a column for thorough washing with binding buffer, the same buffer used for lysis. *St*FlgJ$_{GH(151-316)}$ was then eluted using 750 mM NaCl, 37.5 mM Tris-HCl pH 8.0, 250 mM imidazole. The eluted protein was dialyzed overnight against 800 mM, NaCl 10 mM Tris-HCl pH8.0; and once more the following day against 500 mM NaCl 10 mM Tris-HCl pH 8.0. The dialyzed protein was then concentrated and loaded onto a HiLoad 16/60 Superdex 75 size exclusion column (GE Healthcare) equilibrated with 500 mM NaCl 10 mM Tris-HCl pH 8.0. Fractions containing *St*FlgJ$_{GH(151-316)}$ were pooled, concentrated and subjected to crystallization trials.

To remove the 15 amino acids from the C-terminus if the *St*FlgJ$_{GH(151-316)}$ GH domain that were found to block the domain active site *in trans*, I redesigned the plasmid pET28-*St*FlgJ$_{GH(151-316)}$ using a new primer set (5'GATATACATATGGACAGTAAAGACTTTCTGGCC-3' and 5'-

GATATAGGATCCACTCATCGCTTTCAACTGCTG-3', sense and antisense, respectively), which amplified a sequence excluding the last 15 codons of the full-length domain and placed a BamHI site at the 3'end. The modified *St*FlgJ amplicon, encoding residues 151-301 of full-length FlgJ, was ligated into the pET28b(+) vector described above to generate pET28-*St*FlgJ$_{GH(151-301)}$, from which protein was expressed and purified as describe above for *St*FlgJ$_{GH(151-316)}$.

Crystallization of FlgJ$_{GH(151-316)}$ catalytic domain was performed at 20˚C using the hanging drop vapor diffusion method. Crystal were grown within 72 hours using 2 µL of protein solution at 35 mg/mL in 500 mM NaCl 10 mM Tris-HCl pH 8.0 mixed with an equal volume of mother liquor consisting of 18-22% (w/v) polyethylene glycol 3350 and 250-350 mM NaI. The drops were then equilibrated by vapor diffusion against 50-150 µL of mother liquor. Crystals grew to maximum dimensions of ~0.1 x 0.1 x 1 mm. *St*FlgJ$_{GH(151-301)}$ and *St*FlgJ$_{GH(151-316)}$ crystalized under identical conditions, albeit in different space groups (C 1 2 1 vs P 2_1 2_1 2_1)

Crystals were harvested by sweeping through a solution containing 22% (w/v) PEG 3350, 350 mM NaI, 15% (v/v) glycerol, before being flash-cooled in liquid nitrogen. X-ray diffraction data were then collected in-house at 93K from individual crystals using a MicroMax-007 HF X-ray source and R-axis 4++ detector (Rigaku). With the assistance of Mr. Ben Bailey-Elkin, a PhD student in the Mark Lab, X-ray data

were integrated using iMosflm and scaled and merged using Aimless within the CCP4 software suit (Collaborative Computational Project, 1994). Iodide substructures were determined using phenix.autosol (Adams et al., 2010). I then constructed all X-ray models using Coot (Emsley & Cowtan, 2004), and refinement was carried out using phenix.refine (Adams et al., 2010).

To assay whether the C-terminus of the *St*FlgJ GH domain autoinhibited catalytic activity in solution, a zymogram activity assays of *St*FlgJ$_{GH(151-316)}$ and *St*FlgJ$_{GH(151-301)}$ were performed in which the purified proteins were run on a 15% acrylamide gel containing 0.2% *M. lysodeikticus* lyophilized cells (Sigma), which provided the PG substrate necessary to detect FlgJ GH activity as previously described (Bublitz et al., 2009). To help ensure the integrity of the proteins, sodium dodecyl sulfate and β-mercaptoethanol were excluded from the gels and the protein samples were not boiled prior to gel loading. After resolving the protein on the gel, the gel was rinsed in water for 30 min followed by an equilibration in 100 mM sodium acetate buffer, pH 4.8, for 30 min. The gel was then stained with methylene blue for 30 min, followed by destaining in water to visualize zones of clearing indicative of PG hydrolysis by *St*FlgJ.

Chapter 3
Results and Discussion

The crystal structure *St*FlgJ$_{GH(151-316)}$ (named for the GH domain sequence of the wild-type protein, residues 151-316), which represents the full-length GH domain of FlgJ, was determined by an in-house SAD-phasing experiment based on 11 iodide ions that had bound to the domain from the mother liquor and acted as anomalous scattering centers using Cu K X-radiation (table 1). Due to the experimental method, a high multiplicity (13.3) was needed to accurately measure the anomalous signal of the iodide atoms as the single-wavelength anomalous dispersion (SAD) method, which can be used to locate the positions of the heavy atoms in the unit cell of the crystal (Abendroth et al., 2011). The high multiplicity of the data raised the value of the R_{merge} statistic, which reflects the error associated with averaging multiple symmetry-related reflections in an X-ray data set (which should be the same but aren't due to experimental error) (Wlodawer et al., 2008). Upon scaling and merging these data, an initial model was built using the Phenix software suit (see experimental procedure). A density modification function modified the electron density map to reduce noise in solvent regions of the map, thus improving the experimental phases. Manual model building continued until R-values (R_{work} and R_{free}) were at levels representative of an accurate structural model built using X-ray data to a resolution the 1.8 Å. An R_{work}

value, which compares a simulated diffraction data set based on the built model to the actual experimental data set gives a decent representation of the quality of the model (Wlodawer et al., 2007). The 18.2% R_{work} obtained for this model was arrived at when all density associated with the protein structure and ordered solvent had been modeled as completely as possible. An R-value value of 18.2% at a resolution of 1.8 Å is an acceptable level of error in the protein X-ray crystallography community, where an R-value (in %) divided by 10 should be close to or less than the resolution limit of the X-ray data. The statistics for the truncation mutant $StFlgJ_{GH(151-301)}$ with an R-value of 17.5% at a resolution of 2.15 Å is also well within the acceptable error. Further, overfitting the experimental data, by modeling water molecules into noise for example, can artificially decrease R_{work} without an actual improvement in model accuracy (Wlodawer et al., 2007). Thus an additional R-factor is used, R_{free} (Brunger, 1992) which is a validation statistic that compares the simulated diffraction data from the model to a subset of the experimental X-ray data that has not been used in model construction or refinement (Wlodawer et al., 2007). A divergence between R-work and R-free values (ie. When R-work decreases but R-free increases) indicates overfitting of the model to the data, and is thus avoided. The results of the refinement of $StFlgJ_{GH(151-316)}$ show only a small difference between the R-values (Table 1), indicating that the data were not overfitted. Similar statistics were obtained for the truncation mutant of $StFlgJ_{GH(151-301)}$,

though the difference in the R factors of this structure were higher, but consistent with ratios observed for a large sampling of protein structures from the Protein Data Bank (Tickle et al.,1998).

The resulting structure had a single molecule comprising the asymmetric unit and the domain was found to adopt an α+β lysozyme-like fold common to other GH73 enzymes, which consists of an α-helical subdomain packed against a β-lobe that forms an extended β-hairpin motif (figure 17A)(Bublitz et al., 2009; Hashimoto et al., 2009; Lipski et al., 2015). The α-helical domain and β-hairpin together form a groove-shaped active site that houses a glutamate general acid/base (E184) and a distinctive β-hairpin, in turn containing a second glutamate residue (E223) previously determined to be required for GH activity (figure 17A) (Herlihey et al., 2014). Interestingly, the active site of *St*FlgJ$_{GH(151-316)}$ was found to be occupied by the C-terminal α-helix of its neighbouring symmetry mate (figure 17B).

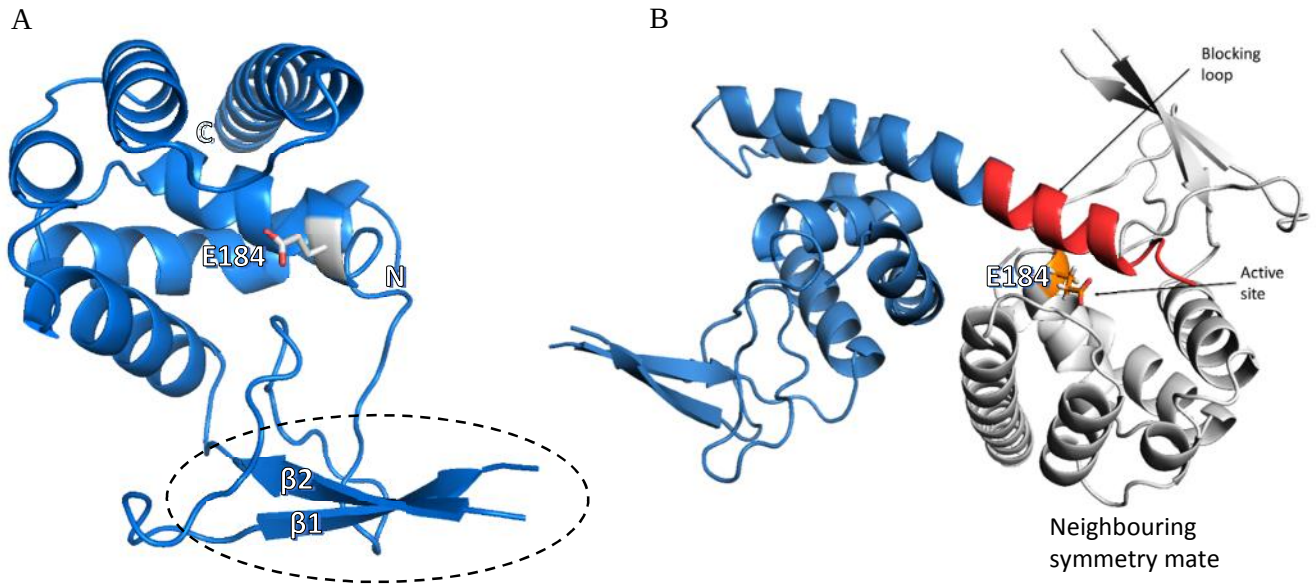


Figure 17. A) The cartoon diagram of $FlgJ_{GH(151-316)}$ GH catalytic domain (blue) is shown with the catalytic glutamate, E184 (grey), within the active site groove. Also highlighted is the β-hairpin subdomain which forms the binding groove. B) All molecules within the crystal structure were found to have their active sites blocked by the C-terminal α-helix (red) of a neighbouring symmetry mate (conserved catalytic residue E184 is depicted in orange). Structural models were generated using PyMol (Delano, 2002)

Table 1.
X-ray Data-collection and refinement statistics

Values in parentheses are for the highest resolution shell

Crystal	*St*FlgJ$_{GH(151-316)}$	*St*FlgJ$_{GH(151-301)}$
Crystal geometry		
Space group	P 2_1 2_1 2_1	C 1 2 1
Unit-cell (Å, °)	a=38.8, b=43.6, c=107.9, α=β=γ=90	a=105.7, b=61.1, c=65.2, α=90, β=106.7, γ=90
Crystallographic data		
Wavelength (Å)	1.54187	1.54187
Resolution Range (Å)	31.50-1.80 (1.86-1.80)	37.47-2.15 (2.22-2.15)
Total observations	235307 (12749)	159127 (13542)
Unique reflections	17650 (997)	21763 (1867)
Multiplicity	13.3 (12.8)	7.3 (7.3)
Completeness (%)	99.8 (98.4)	99.7 (99.0)
Anomalous completeness (%)	99.8 (98.2)	99.8 (99.1)
Mean I/σ(I)	12.9(3.0)	10.3 (2.5)
R_{merge} (%)	11.4 (68.9)	15.1 (82.7)
CC1/2	0.996 (0.959)	0.994 (0.724)
Wilson B-factor (Å^2)	18.6	15.9
Phasing statistics		
Anomalous substructure sites (Iodide)	11	10
Refinement statistics		
R_{work}/R_{free} (%)	18.2/21.6	17.5/25.2
Reflections in test set	900	1096
Protein atoms	1174	3250
Ligands	12	13
Water	150	177
Average *B*-factor (Å^2)		
Protein	29.8	33.8
Ligands	32.3	32.0
Solvent	38.2	33.2
Root mean square deviations		
Bond lengths/angles (Å/°)	0.011/1.27	0.014/1.41
Ramachandran plot		
Favoured/allowed (%)	99.4/0.60	94.3/5.52

A similar *in trans* blockage of the GH domain active site of FlgJ from *Sphingomonas* sp. strain A1 has also been observed (*Sp*FlgJ), where the C-terminus forms an elongated strand that binds into the active site of its neighbour (Hashimoto et al., 2009; Maruyama et al., 2010). This same C-terminal segment, though noted to be a potential heptad repeat of hydrophobic residues used for coiled-coil interaction (Hirano et al., 2001), was reminiscent of the N-terminal autoinhibitory α-helix of the GH73 enzyme Auto from *L. monocytogenes*. Auto was also observed to block *in trans* the GH domain of Auto in a crystal structure of the enzyme and must be proteolytically removed to activate the enzymes (Bublitz et al., 2009). A superposition of the *St*FlgJ$_{GH(151-316)}$ and *Sp*FlgJ structures with Auto revealed that the blocking C- and N-termini of these enzymes are also spatially conserved (figure 18).

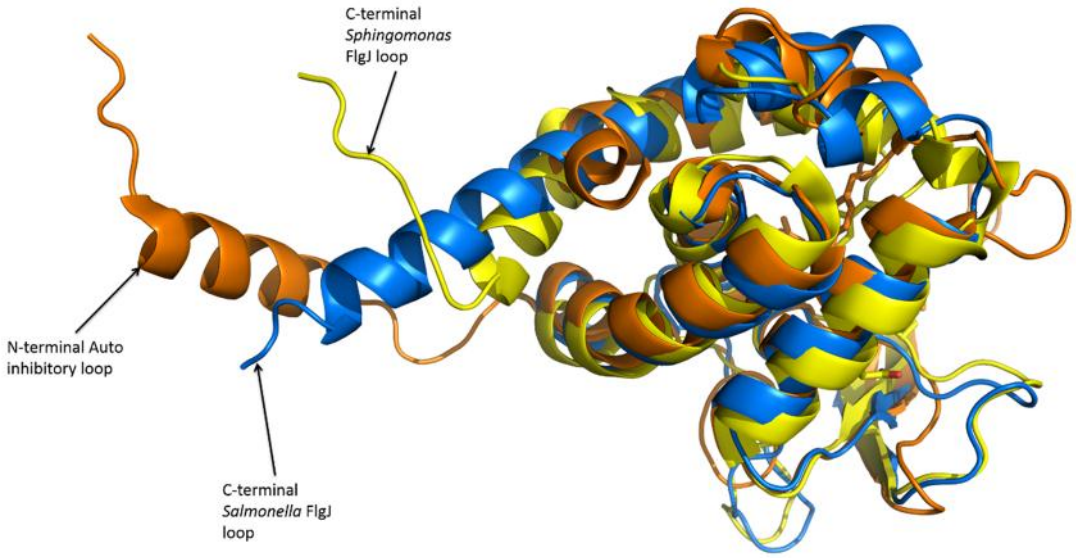


Figure 18. Several superposed and previously characterized GH73 enzymes including *St*FlgJ$_{GH(151-316)}$ from the Mark lab (blue), Auto from the work of Bublitz *et al,* 2009 (orange), and Sphingomonas sp. FlgJ from the work of Hashimoto *et al.,* 2009 (yellow) show a localization of the N- and C-terminal regions to one end of the enzyme. Auto and *Sp*FlgJ share high structural similarity to *St*FlgJ (DALI (Holm and Rosenström, 2010)- Auto: Z-score 15.5; C_α rmsd of 2.2 Å over 133 equivalent positions. FlgJ: Z-score 19.1; C_α rmsd of 3.6 Å over 154 equivalent positions)

Together, these findings suggested that FlgJ may be auto-inhibited by the C-terminal α-helix of the enzyme similarly to the N-terminus of Auto. Moreover, Herlihey *et al.* (2014) recently found the *St*FlgJ GH domain to be susceptible to degradation, resulting in a stable fragment that had lost a decapeptide from its C-terminus. The decapeptide contains the helix that blocks the activate site of the *St*FlgJ GH domain further supporting our hypothesis that the C-terminus may act as an autoinhibitor that can be proteolytically removed. To determine if the C-terminus of *St*FlgJ$_{GH(151-316)}$ acts as an autoinhibitor of the enzyme, zymogram assays of the full-

length *St*FlgJ GH domain, *St*FlgJ$_{GH(151-316)}$, and the truncated form of the domain lacking the last 15 amino of its C-terminus, *St*FlgJ$_{GH(151-301)}$, were carried out. Contrary to our hypothesis, the *St*FlgJ GH domain was found to be fully active with or without the C-terminal α-helix, indicating that the helix does not autoinhibit *St*FlgJ GH activity and is likely not functionally equivalent to the N-terminal inhibitory helix of Auto (figure 19). The C-terminus may instead participate in coiled-coil interactions with other flagellar component proteins as previously proposed (Hirano et al., 2001). Though we did not observe obvious degradation of the domain during expression and purification, the zymogram analysis of FlgJ glycosidase activity showed faint banding indicative of degradation when sodium dodecyl sulphate and β-mercaptoethanol were included in the assay (*data not shown*). The degradation was reduced when SDS and β-mercaptoethanol were removed from zymogram assay (*see methods*).

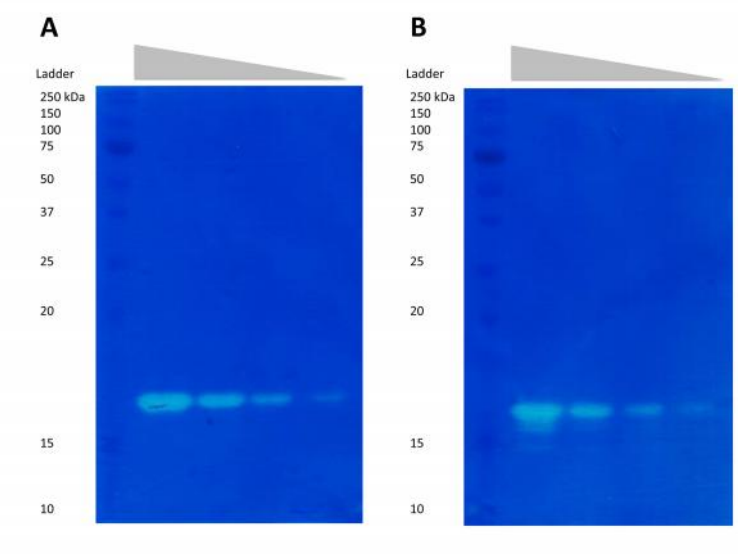


Figure 19 A) Zymogram activity assay reveals the activie of the full-length *St*FlgJ GH domain *St*FlgJ$_{GH(151-316)}$. B) The activity of the C-terminally truncated mutant *St*FlgJ$_{GH(151-301)}$ shows no loss of activity is apparent when the C-terminal α-helix is removed. Slight banding is apparent for *St*FlgJ$_{GH(151-301)}$. For each construct, a total of 3 µg, 1.5 µg, 0.75 µg or 0.375 µg was added to each well.

To gain insight into the active site architecture *St*FlgJ when it is not occluded by the C-terminal a-helix, the crystal structure of the truncated GH domain (*St*FlgJ$_{GH(151-301)}$) was determined as in Chapter 2. As for the full-length domain, the structure of determined by an in-house SAD-phasing experiment where 10 iodide ions had bound the domain from the mother liquor and could be used as anomalous scattering centers. Three copies of the truncated domain comprised the asymmetric unit, which was refined to 2.2 Å resolution. Though the C-terminal α-helix no longer occupied the active site, the β-hairpin motifs of each domain within the asymmetric unit were instead found to partly occlude the active site of neighbouring domains in the crystal structure (figure 20).

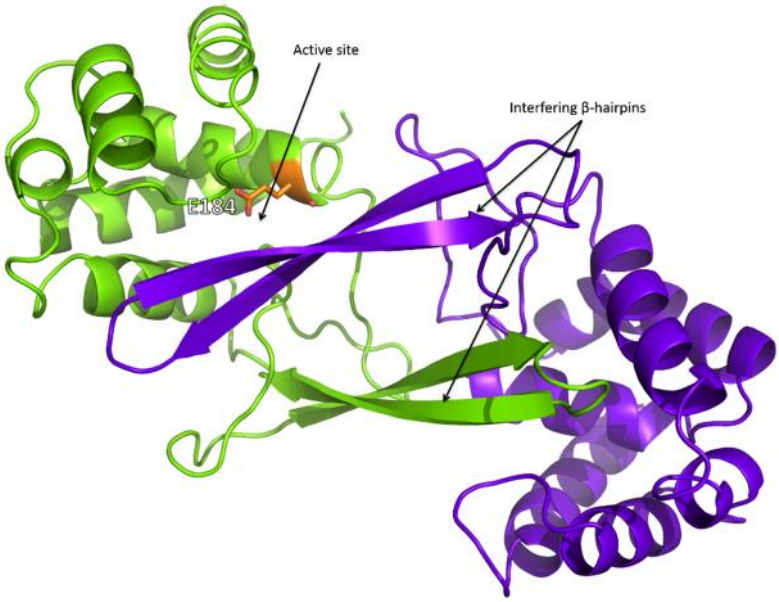


Figure 20. Shown is a cartoon of the modified structure of the truncated mutant, $FlgJ_{GH(151-301)}$ (green) and a neighbouring symmetry mate (purple) within the crystal structure of the mutant. While the catalytically active β-hairpin is highly ordered, it occludes the active site of each copy of the protein in the crystal structure.

While this crystal packing did not provide an unobstructed active site, it did impart significant structural order to the β-hairpin, which is often too flexible to observe in crystals structures of GH73 GH domains that have been obtained (including the full-length *St*FlgJ GH domain reported here, *St*$FlgJ_{GH(151-316)}$) (Bai et al., 2014; Hashimoto et al., 2009; Lipski et al., 2015; Maruyama et al., 2010). The truncated *St*$FlgJ_{GH(151-301)}$ structure reveals that the β-hairpin is capable of moving out from the helical domain considerably further than previously reported for other GH73 enzymes, including Auto and *Sp*FlgJ (figure 21), an implication which might help answer previously asked questions of whether the catalytic mechanism requires large conformational changes of the β-hairpin. The remarkable mobility of the β-hairpin results in a highly flexible active site architecture that likely not only assists in substrate capture but turnover as well since it harbours the putative catalytic nucleophile, Glu223, which may work with

the glutamic acid general acid base on the helical domain to carry out glycosidic bond hydrolysis (Herlihey et al., 2014; Maruyama et al., 2010; Nambu et al., 1999).

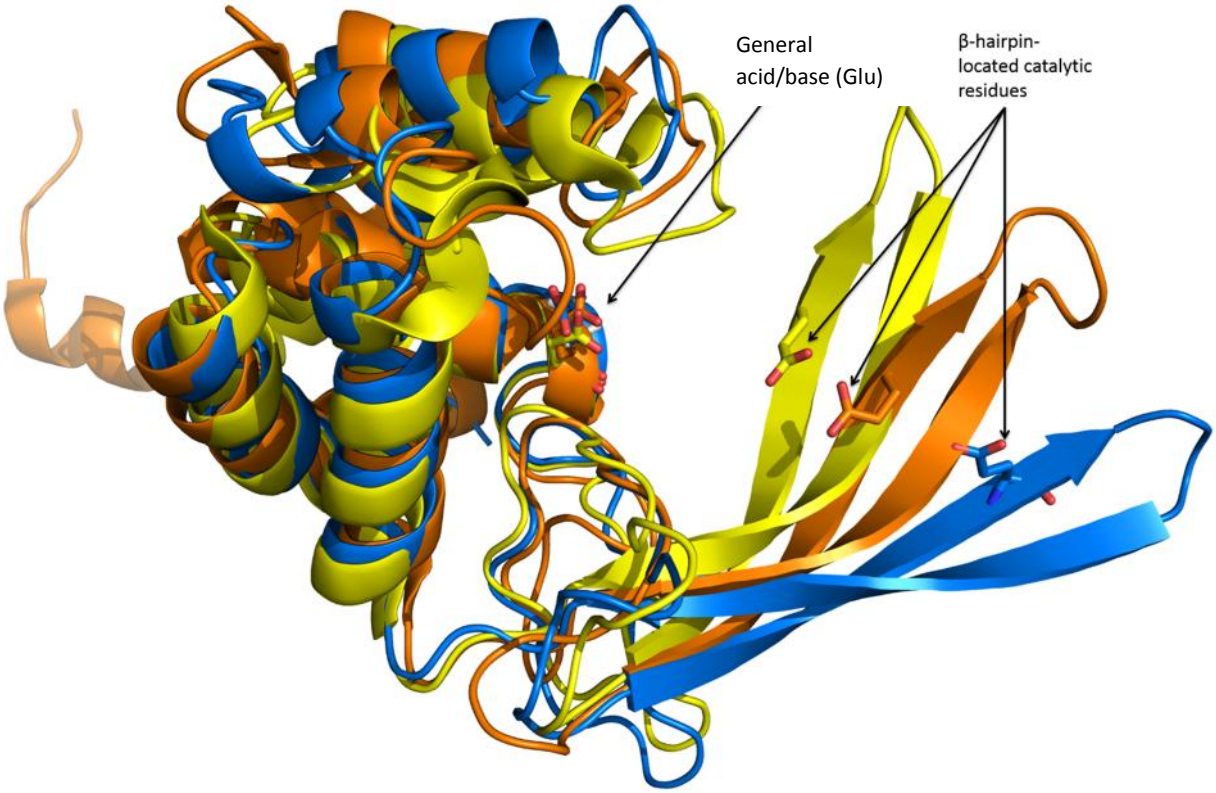


Figure 21. The extent to which the β-hairpin of GH73 enzymes can extend open and closed is shown by a superposition of three GH73 enzymes with ordered hairpins: $FlgJ_{GH(151-301)}$ (blue). Auto from the work of Bublitz *et al.*, 2009 (orange), and *Sp*FlgJ from the work of Maruyama et al. (unpublished, PDB: 3VWO; yellow) (DALI (Holm and Rosenström, 2010)- Auto: Z-score 16.4; C_α rmsd of 3.3 Å over 140 equivalent positions. *Sphingo* FlgJ: Z-score 18.8; C_α rmsd of 2.3 Å over 140 equivalent positions). Also shown is the location of the conserved general acid/base residues as well as the conserved β-hairpin nucleophiles.

FlgJ (Herlihey et al., 2014; Maruyama et al., 2010; Nambu et al., 1999), Auto (Bublitz et al., 2009), and the recently characterized GH73 enzyme from *Thermotoga maritima*, TM0633 (Lipski et al., 2015), all have a conserved glutamate on the flexible β-hairpin that is required for catalysis, which suggests one of three possible catalytic mechanisms occurs in these enzymes: 1) With substrate bound, the β-hairpin moves deep enough into the active site to allow the glutamate to act as a nucleophile, directly attacking the anomeric carbon of GlcNAc to promote glycosidic bond cleavage that proceeds via a glycosyl-enzyme intermediate where anomeric stereochemistry is retained; 2) The hairpin does not move in deep enough to allow the glutamate to directly attack GlcNAc but instead the glutamate activates an intervening water molecule that, in turn, attacks the anomeric carbon of GlcNAc to promote glycoside bond cleavage with inversion of anomeric configuration, or finally 3) The β-hairpin moves in to position the glutamate so that it can help facilitate a substrate assisted catalytic mechanism whereby the 2-acetamido group of GlcNAc acts as the nucleophile to drive a reaction that proceeds via an oxazolinium ion intermediate and generates a product with retained stereochemistry. Current structural and functional studies point towards the classical inverting mechanism (#2 above) as the likely mechanism at play given the large distances (~10 – 20 Å) that have been observed between the glutamate general acid on the helical domain and putative general base on the β-hairpin of these

enzymes. GH enzymes employing an inverting mechanism typically have the general acid and general base ~10.5 Å apart, whereas, those using a retaining mechanism generally proceed through a glycosyl-enzyme intermediate place these residues ~5.5 Å apart (Zechel & Withers, 2000). Though a distance approaching ~5.5 Å has yet to be observed between the catalytic residues of a GH73 enzyme, given the apparent flexibility of the β-hairpin, either mechanism remains possible and stereochemical outcome studies will be needed to rule out the latter. Alternatively, GH18, GH20, GH56, GH84 enzymes employ a second enzyme carboxyl group (separate from the general acid/base) that appears to activate the 2-acetamido group of the substrate to participate in a substrate assisted reaction and help to stabilize the resulting oxazolinium ion intermediate (#3 above) (Aalten et al., 2001; Dennis et al., 2006; Mark et al., 2001; Marković-Housley et al., 2000; Tews et al., 1996). The SAC mechanism of GH85 enzymes uses an asparagine in place of an enzyme carboxylate to help facilitate 2-acetamido group participation (Abbott et al., 2009). While it is possible the catalytic glutamate on the β-hairpin could serve this role, β-*N*-acetylglucosaminidases that engage the 2-acetamido group in this manner typically have ridged active site architectures and a number of additional residues that are needed to correctly orient the carbonyl oxygen of the 2-acetamido within striking distance of the anomeric center of GlcNAc (Mark et al., 2001). These additional requirements for substrate-

assisted catalysis may make this mechanism less likely for these enzymes. Interestingly however, there exists additional groups of GH73 enzymes in which the β-hairpin contains a carboxyl group that is not needed for catalysis, as in *Staphylococcus warneri* M Atl_{WM} (Yokoi et al., 2008), or as for the GH73 enzyme AcmA from *Lactococcus lactis ssp. cremoris* MG1363, the β-hairpin contains a carboxyl group (Glu) that can be exchanged for an amide group (Gln) without complete loss of activity (Inagaki et al., 2009). In addition some GH73 sequences appear to completely lack a suitable acidic residue within the β-hairpin (Herlihey et al., 2014). It thus appears these latter groups may use a substrate assisted catalytic mechanism that does not depend on the use of a second enzymic carboxyl group.

Conclusion

Further structural and functional studies, including the development of straightforward kinetic assays using substrate analogues, will be needed to resolve which catalytic mechanism(s) are employed by GH73 enzymes. Given the most recent observations describing a variable β-hairpin subdomain, only speculations can be made regarding the β-hairpins' role in enzymatic activity. It seems that the lysozyme-like fold is capable of hosting a range of PG hydrolyzing activities as described in Chapter 1, so having family GH73 host two mechanisms would not be too great a

surprise. Indeed, GH families 23 and 97 both host inverting and retaining mechanisms with family GH23 being closely related to GH73 enzymes as both contain a lysozyme-fold (Blackburn & Clarke, 2001; Gloster et al., 2008; Kuroki et al., 1999). At the moment, the task of determining mechanism may seem daunting, given GH73 enzymes pervasiveness across three phyla, within five clusters (Lipski et al., 2015). Each cluster may have its own nuance that may demand a characterization of a representative member given that each has a differing set of residues comprising the β-hairpin (Lipski et al., 2015), whose catalytic importance seem secondary only to the catalytic general acid/base.

With the work that has been done, many groups investigating family GH73 have the groundwork laid and tools at hand to continue probing the active site and to overcome the obstacles inherent in studying PG hydrolyzing enzymes. Those challenges, namely analyzing a complex substrate whose composition is highly variable, and pairing it to the enzymatic activity of the enzyme, may be overcome as the groups studying family GH73 develop novel techniques for targeting these PG-hydrolase specific problems (Bai et al., 2014; Herlihey et al., 2014). Upon successful determination of catalytic mechanism(s), GH73-specific structure-based inhibitor development should follow naturally, whose ultimate goal could be toward therapy of infectious disease. Indeed, certain GH73 virulence factors, such as Auto, are necessary

for things such as entry into host cells (Cabanes et al., 2004), and in instances like these, inhibitor development based on continued structure determination would prove invaluable.

Bibliography

Aalten, D. M. F. van, Komander, D., Synstad, B., Gåseidnes, S., Peter, M. G., & Eijsink, V. G. H. (2001). Structural Insights into the Catalytic Mechanism of a Family 18 Exo-Chitinase. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(16), 8979–8984.

Abbott, D. W., Macauley, M. S., Vocadlo, D. J., & Boraston, A. B. (2009). Streptococcus pneumoniae Endohexosaminidase D, Structural and Mechanistic Insight into Substrate-assisted Catalysis in Family 85 Glycoside Hydrolases. *Journal of Biological Chemistry*, *284*(17), 11676–11689.

Abendroth, J., Gardberg, A., Robinson, J., Christensen, J., Staker, B., Myler, P., Stewart, L., & Edwards, T. (2011). SAD phasing using iodide ions in a high-throughput structural genomics environment. *Journal of Structural and Functional Genomics*, *12*(2), 83–95.

Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, *66*(2), 213–221.

Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., Paris, G., Frasch, A. C., Withers, S. G., & Alzari, P. M. (2004). Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase. *Structure*, *12*(5), 775–784.

Bai, X. H., Chen, H. J., Jiang, Y. L., Wen, Z., Huang, Y., Cheng, W., Li, Q., Qi, L., Zhang, J. R., Chen, Y., & Zhou, C. Z. (2014). Structure of pneumococcal peptidoglycan hydrolase LytB reveals insights into the bacterial cell wall remodeling and pathogenesis. *Journal of Biological Chemistry*, *289*(34), 23403–23416.

Blackburn, N., & Clarke, A. J. (2001). Identification of Four Families of Peptidoglycan Lytic Transglycosylases. *Journal of Molecular Evolution*, *52*(1), 78–84.

Bourgeois, I., Camiade, E., Biswas, R., Courtin, P., Gibert, L., Götz, F., Chapot-chartier, M., Pons, J., & Pestel-caron, M. (2009). Characterization of AtlL, a bifunctional

autolysin of Staphylococcus lugdunensis with N - acetylglucosaminidase and N-acetylmuramoyl-l-alanine amidase activities. *FEMS Microbiology Letters*, *290*(1), 105–113.

Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *355*(6359), 472–475.

Bublitz, M., Polle, L., Holland, C., Heinz, D. W., Nimtz, M., & Schubert, W. (2009). Structural basis for autoinhibition and activation of Auto, a virulence- associated peptidoglycan hydrolase of Listeria monocytogenes. *Molecular Microbiology*, *71*(6), 1509–1522.

Buist, G., Steen, A., Kok, J., & Kuipers, O. P. (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Molecular Microbiology*, *68*(4), 838–847.

Cabanes, D., Dussurget, O., Dehoux, P., & Cossart, P. (2004). Auto, a surface associated autolysin of Listeria monocytogenes required for entry into eukaryotic cells and virulence. *Molecular Microbiology*, *51*(6), 1601–1614.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., & Mornon, J. P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and Molecular Life Sciences*, *53*(8), 621–645.

Camiade, E., Peltier, J., Bourgeois, I., Couture-Tosi, E., Courtin, P., Antunes, A., Chapot-Chartier, M.-P., Dupuy, B., & Pons, J.-L. (2010). Characterization of Acp, a Peptidoglycan Hydrolase of Clostridium perfringens with N-Acetylglucosaminidase Activity That Is Implicated in Cell Separation and Stress-Induced Autolysis. *The Journal of Bacteriology*, *192*(9), 2373.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate- Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*, 233–238.

Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*, *5*(4), 823.

Collaborative Computational Project, N. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D*, *50*(5), 760–763.

Comfort, D. A., Bobrov, K. S., Ivanen, D. R., Shabalin, K. A., Harris, J. M., Kulminskaya, A. A., Brumer, H., & Kelly, R. M. (2007). Biochemical analysis of Thermotoga maritima GH36 alpha- galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry*, *46*(11), 3319.

Davies, G. J., Ducros, V. M. A., Varrot, A., & Zechel, D. L. (2003). Mapping the conformational itinerary of beta- glycosidases by X- ray crystallography. *Biochemical Society Transactions*, *31*(3), 523.

Davies, G. J., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853–859.

Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., & Withers, S. G. (1998). Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry*, *37*(34), 11707.

Davies, G. J., & Sinnott, M. L. (2008). Sorting the diverse: the sequence-based classifications of carbohydrate-active enzymes. *Biochemical Journal*. http://doi.org/10.1042/BJ20080382

Davies, G. J., Wilson, K. S., & Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal*, *321*(2), 557–559.

De Las Rivas, B., Garcia, J. L., Lopez, R., & Garcia, P. (2002). Purification and Polar Localization of Pneumococcal LytB, a Putative Endo-{beta}-N-Acetylglucosaminidase: the Chain-Dispersing Murein Hydrolase. *The Journal of Bacteriology*, *184*(18), 4988.

Delano, W. L. (2002). The PyMOL Molecular Graphics System. http://www.pymol.org.

Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J., Black, G. N., Vocadlo, D. J., & Davies, G. J. (2006). Structure and mechanism of a bacterial {beta}-glucosaminidase having O-GlcNAcase activity. *Nature Structural & Molecular Biology*, *13*(4), 365–371.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Teeri, T. T., & Jones, T. A. (1994). The Three- Dimensional Crystal Structure of the Catalytic Core of Cellobiohydrolase I from Trichoderma reesei. *Science*, *265*(5171), 524–528.

Eckert, C., Lecerf, M., Dubost, L., Arthur, M., & Mesnage, S. (2006). Functional Analysis of AtlA, the Major N- Acetylglucosaminidase of Enterococcus faecalis. *The Journal of Bacteriology*, *188*(24), 8513.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D*, *60*(12-1), 2126-2132.

Foster, S. J. (1995). Molecular characterization and functional analysis of the major autolysin of Staphylococcus aureus 8325/4. *The Journal of Bacteriology*, *177*(19), 5723.

García, P., González, M. P., García, E., López, R., & García, J. L. (1999). LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Molecular Microbiology*, *31*(4), 1275-1277.

Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B., & Davies, G. J. (2008). Divergence of Catalytic Mechanism within a Glycosidase Family Provides Insight into Evolution of Carbohydrate Metabolism by Human Gut Flora. *Chemistry & Biology*, *15*(10), 1058-1067.

Gosink, K. K., Mann, E. R., Guglielmo, C., Tuomanen, E. I., & Masure, H. R. (2000). Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. *Infection and Immunity*, *68*(10), 5690.

Hashimoto, W., Ochiai, A., Momma, K., & Itoh, T. (2009). Crystal structure of the glycosidase family 73 peptidoglycan hydrolase FlgJ. *Biochemical and Biophysical Research Communications*, *381*(1), 16-21.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical Journal*, *280*(2), 309.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P., & Davies, G. (1995). Conserved Catalytic Machinery and the Prediction of a Common Fold for Several Families of Glycosyl Hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(15), 7090-7094.

Henrissat, B., & Davies, G. (1997). Structural and sequence- based classification of glycoside hydrolases. *Current Opinion in Structural Biology*, *7*(5), 637-644.

Herlihey, F. A., Moynihan, P. J., & Clarke, A. J. (2014). The essential protein for bacterial flagella formation FlgJ functions as a beta-N-acetylglucosaminidase. *The Journal of Biological Chemistry*, *289*(45), 31029–31042.

Hirano, T., Minamino, T., & Macnab, R. M. (2001). The role in flagellar rod assembly of the N- terminal domain of Salmonella FlgJ, a flagellum-specific muramidase. *Journal of Molecular Biology*, *312*(2), 359–369.

Holtje, J.-V. (1998). Growth of the Stress- Bearing and Shape- Maintaining Murein Sacculus of Escherichia coli. *Microbiology and Molecular Biology Reviews*, *62*(1), 181.

Inagaki, N., Iguchi, A., Yokoyama, T., Yokoi, K., Ono, Y., Yamakawa, A., Taketo, A., & Kodaira, K.-I. (2009). Molecular properties of the glucosaminidase AcmA from Lactococcus lactis MG1363: Mutational and biochemical analyses. *Gene*, *447*(2), 61–71.

Kempton, J. B., & Withers, S. G. (1992). Mechanism of Agrobacterium .beta.-glucosidase: kinetic studies. *Biochemistry*, *31*(41), 9961–9969.

Knapp, S., Vocadlo, D., Gao, Z., Kirk, B., Lou, J., & Withers, S. G. (1996). NAG-thiazoline, An N-Acetyl-β-hexosaminidase Inhibitor That Implicates Acetamido Participation. *Journal of the American Chemical Society*, *118*(28), 6804–6805.

Koshland, D. E. (1953). Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews*, *28*(4), 416–436.

Kuroki, R., Weaver, L. H., & Matthews, B. W. (1999). Structural Basis of the Conversion of T4 Lysozyme into a Transglycosidase by Reengineering the Active Site. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 8949–8954.

Laine, R. A. (1994). A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*, *4*(6), 759.

Lipski, A., Herve, M., Lombard, V., Nurizzo, D., Mengin-Lecreulx, D., Bourne, Y., & Vincent, F. (2015). Structural and biochemical characterization of the beta-N-

acetylglucosaminidase from Thermotoga maritima: toward rationalization of mechanistic knowledge in the GH73 family. *Glycobiology*, *25*(3), 319–330.

Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W., Gong, C.-X., & Hakomori, S.-I. (2004). O- GlcNAcylation Regulates Phosphorylation of Tau: A Mechanism Involved in Alzheimer's Disease. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(29), 10804–10809.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate- active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*, 490. http://doi.org/10.1093/nar/gkt1178

Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D., & Vocadlo, D. J. (2005). O-GlcNAcase uses substrate- assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired inhibitors. *The Journal of Biological Chemistry*, *280*(27), 25313.

Macleod, A. M., Lindhorst, T., Withers, S. G., & Warren, R. A. (1994). The acid/ base catalyst in the exoglucanase/ xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry*, *33*(20), 6371.

Mark, B. L., Vocadlo, D. J., Knapp, S., Triggs-Raine, B. L., Withers, S. G., & James, M. N. (2001). Crystallographic evidence for substrate- assisted catalysis in a bacterial beta- hexosaminidase. *The Journal of Biological Chemistry*, *276*(13), 10330.

Marković-Housley, Z., Miglierini, G., Soldatova, L., Rizkallah, P. J., Müller, U., & Schirmer, T. (2000). Crystal Structure of Hyaluronidase, a Major Allergen of Bee Venom. *Structure*, *8*(10), 1025–1035.

Maruyama, Y., Ochiai, A., Itoh, T., Mikami, B., Hashimoto, W., & Murata, K. (2010). Mutational studies of the peptidoglycan hydrolase FlgJ of Sphingomonas sp. strain A1. *Journal of Basic Microbiology*, *50*(4), 311–317.

McCarter, J. D., & Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*, *4*(6), 885–892.

Mesnage, S., Dellarole, M., Baxter, N. J., Rouget, J. B., Dimitrov, J. D., Wang, N., Fujimoto, Y., Hounslow, A. M., Lacroix-Desmazes, S., Fukase, K., Foster, S. J., &

Williamson, M. P. (2014). Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nature Communication*, *5*, 4269.

Moens, S., & Vanderleyden, J. (1996). Functions of bacterial flagella. *Critical Reviews in Microbiology*, *22*(2), 67–100.

Nakajima, R., Imanaka, T., & Aiba, S. (1986). Comparison of amino acid sequences of eleven different α-amylases. *Applied Microbiology and Biotechnology*, *23*(5), 355–360.

Nambu, T., Minamino, T., Macnab, R. M., & Kutsukake, K. (1999). Peptidoglycan-Hydrolyzing Activity of the FlgJ Protein, Essential for Flagellar Rod Formation in Salmonella typhimurium. *The Journal of Bacteriology*, *181*(5), 1555.

Namchuk, M. N., & Withers, S. G. (1995). Mechanism of Agrobacterium beta-glucosidase: kinetic analysis of the role of noncovalent enzyme/ substrate interactions. *Biochemistry*, *34*(49), 16194.

Neuhaus, F. C., & Baddiley, J. (2003). A Continuum of Anionic Charge: Structures and Functions of D- Alanyl- Teichoic Acids in Gram- Positive Bacteria. *Microbiology and Molecular Biology Reviews*, *67*(4), 686.

Phillips, D. C. (1967). The Hen Egg-White Lysozyme Molecule. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(3), 483–495.

Raimbaud, E., Buléon, A., Perez, S., & Henrissat, B. (1989). Hydrophobic cluster analysis of the primary sequences of α-amylases. *International Journal of Biological Macromolecules*, *11*(4), 217–225.

Rashid, M. H., Mori, M., & Sekiguchi, J. (1995). Glucosaminidase of Bacillus subtilis: cloning, regulation, primary structure and biochemical characterization. *Microbiology*, *141*(1), 2391.

Schleifer, K. H., & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Reviews*, *36*(4), 407.

Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews*, *90*(7), 1171–1202.

Sinnott, M. L., & Souchard, I. J. (1973). The mechanism of action of beta-galactosidase. Effect of aglycone nature and -deuterium substitution on the hydrolysis of aryl galactosides. *The Biochemical Journal*, *133*(1), 89.

Smith, T. J., Blackman, S. A., & Foster, S. J. (2000). Autolysins of Bacillus subtilis: multiple enzymes with multiple functions. *Microbiology*, *146* (2), 249-262.

Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., & Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of -amylase-related proteins. *Protein Engineering, Design and Selection*, *19*(12), 555-562.

Stoddart, J. F. (1971). *Stereochemistry of carbohydrates*. New York: New York, Wiley-Interscience.

Terwisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B., & Dijkstra, B. W. (1995). Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and x-ray structure of a complex with allosamidin evidence for substrate assisted catalysis. *Biochemistry*, *34*(48), 15619-15623.

Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., & Vorgias, C. E. (1996). Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nature Structural & Molecular Biology*, *3*(7), 638-648.

Tickle, I. J., Laskowski, R. A., & Moss, D. S. (1998). Rfree and the Rfree Ratio. I. Derivation of Expected Values of Cross-Validation Residuals Used in Macromolecular Least-Squares Refinement. *Acta Crystallographica Section D*, *54*(4), 547-557.

Törrönen, A., Kubicek, C. P., & Henrissat, B. (1993). Amino acid sequence similarities between low molecular weight endo-1,4-β-xylanases and family H cellulases revealed by clustering analysis. *FEBS Letters*, *321*(2), 135-139.

Uitdehaag, J. C. M., Mosi, R., Kalk, K. H., van der Veen, B. A., Dijkhuizen, L., Withers, S. G., & Dijkstra, B. W. (1999). X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nature Structural Biology*, *6*(5), 432.

Umezurike, G. M. (1988). The effect of glycerol on the activity of beta-glucosidase from Botryodiplodia theobromae Pat. *Biochemical Journal*, *254*(1), 73–76.

Van Doorslaer, E., Van Opstal, O., Kersters-Hilderson, H., & De Bruyne, C. K. (1984). Kinetic α-deuterium isotope effects for enzymatic and acid hydrolysis of aryl-β-d-glycopyranosides. *Bioorganic Chemistry*, *12*(2), 158–169.

Vocadlo, D. J., & Davies, G. J. (2008). Mechanistic insights into glycosidase chemistry. *Current Opinion in Chemical Biology*, *12*(5), 539–555.

Vocadlo, D. J., Davies, G. J., Laine, R., & Withers, S. G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature*, *412*(6849), 835.

Vollmer, W., Joris, B., Charlier, P., & Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiology Reviews*, *32*, 259–286.

Vosseller, K., Wells, L., Lane, M. D., & Hart, G. W. (2002). Elevated Nucleocytoplasmic Glycosylation by O- GlcNAc Results in Insulin Resistance Associated with Defects in Akt Activation in 3T3-L1 Adipocytes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(8), 5313–5318.

Vötsch, W., & Templin, M. F. (2000). Characterization of a β-N-acetylglucosaminidase ofEscherichia coli and Elucidation of Its Role in Muropeptide Recycling and β-Lactamase Induction . *Journal of Biological Chemistry* , *275* (50), 39032–39038.

Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J., & Withers, S. G. (1994). Changing Enzymic Reaction Mechanisms by Mutagenesis: Conversion of a Retaining Glucosidase to an Inverting Enzyme. *Journal of the American Chemical Society*, *116*(25), 11594–11595.

Watson, J. N., Dookhun, V., Borgford, T. J., & Bennet, A. J. (2003). Mutagenesis of the conserved active- site tyrosine changes a retaining sialidase into an inverting sialidase. *Biochemistry*, *42*(43), 12682.

Watts, A. G., Damager, I., Amaya, M. L., Buschiazzo, A., Alzari, P., Frasch, A. C., & Withers, S. G. (2003). Trypanosoma cruzi trans- sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. *Journal of the American Chemical Society*, *125*(25), 7532.

Weaver, L. H., Grütter, M. G., & Matthews, B. W. (1995). The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "Goose-type" ly'sozymes lack a catalytic aspartate residue. *Journal of Molecular Biology*, *245*(1), 54–68.

Withers, S. G., & Aebersold, R. (1995). Approaches to labeling and identification of active site residues in glycosidases. *Protein Science*, *4*(3), 361–372.

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*, *275*(1), 1–21.

Woodcock, S., Mornon, J. P., & Henrissat, B. (1992). Detection of secondary structure elements in proteins by hydrophobic cluster analysis. *Protein Engineering Design and Selection*, *5*(7), 629–635.

Yang, J., Schenkman, S., & Horenstein, B. A. (2000). Primary 13C and beta- secondary 2H KIEs for trans- sialidase. A snapshot of nucleophilic participation during catalysis. *Biochemistry*, *39*(19), 5902.

Yip, V. L. Y., & Withers, S. G. (2006). Breakdown of oligosaccharides by the process of elimination. *Current Opinion in Chemical Biology*, *10*(2), 147–155.

Yokoi, K., Sugahara, K., Iguchi, A., Nishitani, G., Ikeda, M., Shimada, T., Inagaki, N., Yamakawa, A., Taketo, A., & Kodaira, K.-I. (2008). Molecular properties of the putative autolysin Atl WM encoded by Staphylococcus warneri M: Mutational and biochemical analyses of the amidase and glucosaminidase domains. *Gene*, *416*(1), 66–76.

Zechel, D. L., & Withers, S. G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Accounts of Chemical Research*, *33*(1), 11.