1 INTRODUCTION

The glycosidic bond is one of the most stable bonds in nature. With an estimated half life of around 5 million years (Wolfenden et al., 1998), the enzymes which catalyze the hydrolysis of this bond may be considered among the most powerful enzyme catalysts known. This, combined with the diversity of glycosidic bonds which may form between different sugar moieties in the order of 10^{12} possible hexasaccharide isomers (Laine, 1994), creates a large and structurally diverse group of enzymes. These enzymes play central roles in diseases such as diabetes, and industrial processes like food technology. This ensures that the glycosyl hydrolases are a medically relevant, industrially important and scientifically interesting group of enzymes.

1.1.1 Classification of Glycosyl Hydrolases

Nearly 3000 glycosyl hydrolases have currently been identified, while ORFs encoding putative ones registered in the Carbohydrate-Active Enzymes Database (Coutinho and Henrissat, 1999) bring the number to over 5500. The traditional method of classification, the Enzyme Classification (E. C.) nomenclature system, relies on assigning enzymes by substrate specificity. Glycosyl hydrolases, however, are virtually impossible to segregate by this system as many have multiple substrate specificities which may only partially overlap with others. The solution to this dilemma was the development of a system of classification based on sequence similarity (Henrissat, 1991). This system has currently subdivided this large group into 90 families, however the number is still growing.

1.1.2 Mechanisms of Action of Glycosyl Hydrolases

The glucosidases are a group within the broad term of glycosyl hydrolases which cleave glycosidic bonds to liberate a glucose molecule from an oligosaccharide. The α-glucosidases specifically hydrolyse the α-glucopyranosidic bond, thereby releasing an α-D-glucose from the non-reducing end of a sugar (see Figure 1.1).

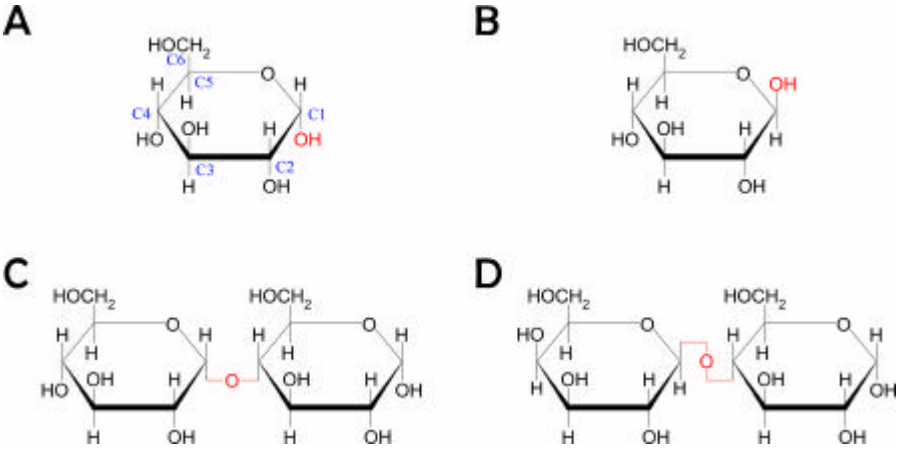

Figure 1.1 Haworth structures of several common sugars displaying different configurations

A. α-D-Glucose with C atoms labeled in *blue*. **B.** β-D-Glucose. **C.** Maltose (α-D-Glucopyranosyl-(1-4)- α-D-glucopyranose). **D.** Lactose (β-D-Galactopyranosyl-(1-4)-α-D-glucopyranose).

The general hydrolysis reaction has been well characterised and involves the concerted action of a general base and a general acid on the C1 carbon atom of the glycosidic bond. In glucosidases, both the general base and acid catalyst are conserved acidic residues (Davies and Henrissat, 1995; Zechel and Withers, 2001).

The hydrolysis reaction has two possible outcomes: either the inversion or the retention of the stereochemistry around the anomeric C1 carbon. The former consists of a single-displacement mechanism involving an oxocarbenium ion-like transition state (Zechel, 2000), where the general base deprotonates a water molecule which then attacks the C1 carbon atom of the glycosidic bond. The general base is aided in cleaving the glycosidic bond by a general acid catalyst which protonates the leaving group, the liberated glucose molecule (Figure 1.2, A.). As a consequence of the inclusion of an activated water located between the general base and the anomeric C1 atom, the catalytic carboxyl groups in the protein are generally about 10.5Å apart (McCarter and Withers, 1994; Wang et al., 1994).

In contrast, the retention mechanism involves a double displacement mechanism which includes the formation of a covalently bound glycosyl enzyme intermediate further hydrolysed through oxocarbenium-ion-like transition states (Figure 1.2, B.). Unlike the inversion mechanism, this is performed by a general base diresctly attacking the C1 atom, without the presence of a water molecule. This therefore requires a closer proximity of the two active carboxyl groups, generally around 5Å. Examples of the active site of both an inverting- and retaining-glucosidase can be seen in Figure 1.3. It shows that the substrate

and catalytic water molecules are very specifically orientated via multiple polar interactions with the active site residues of the respective glucosidases.

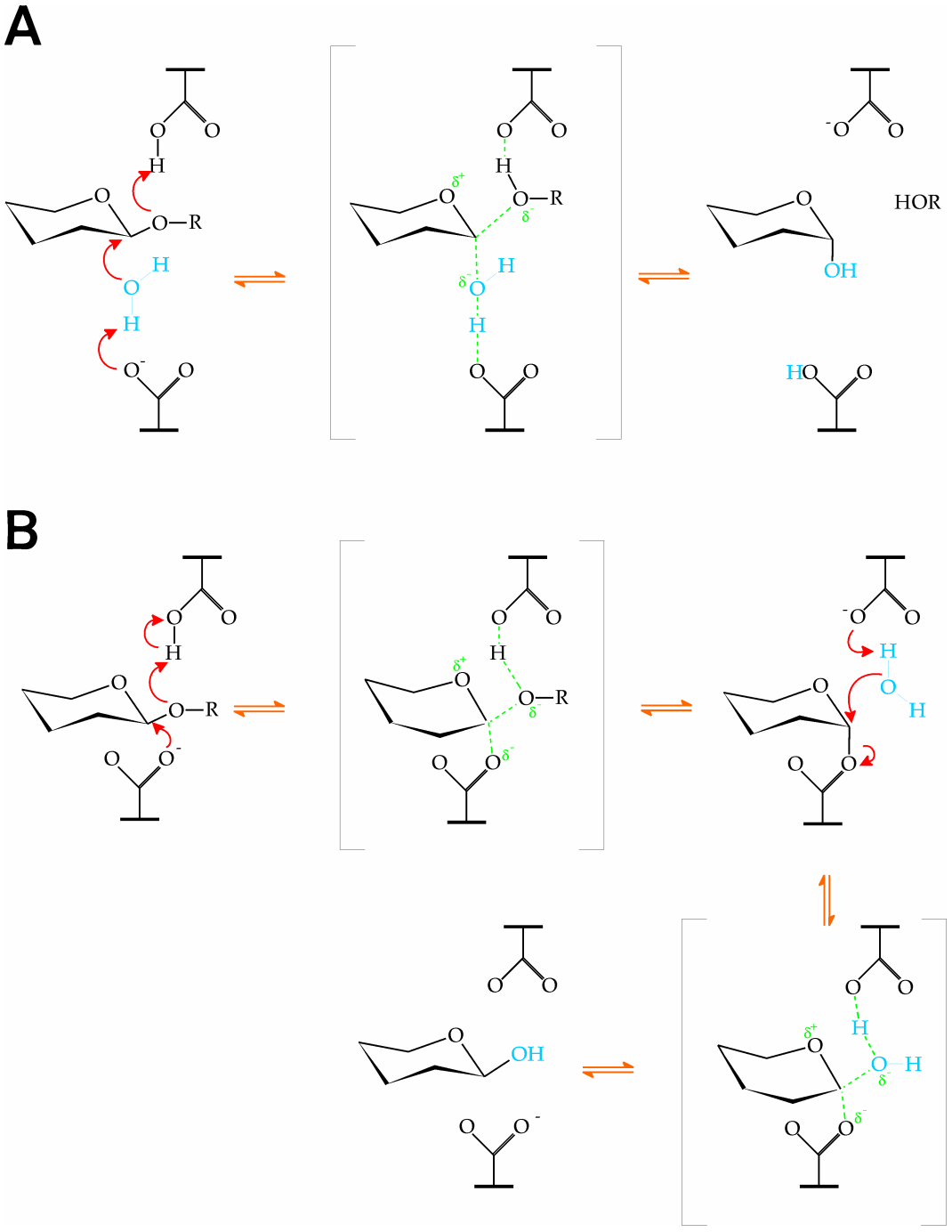

Figure 1.2 Mechanisms of action of glycosyl hydrolases

A. Inverting mechanism, **B**. Retaining mechanism.

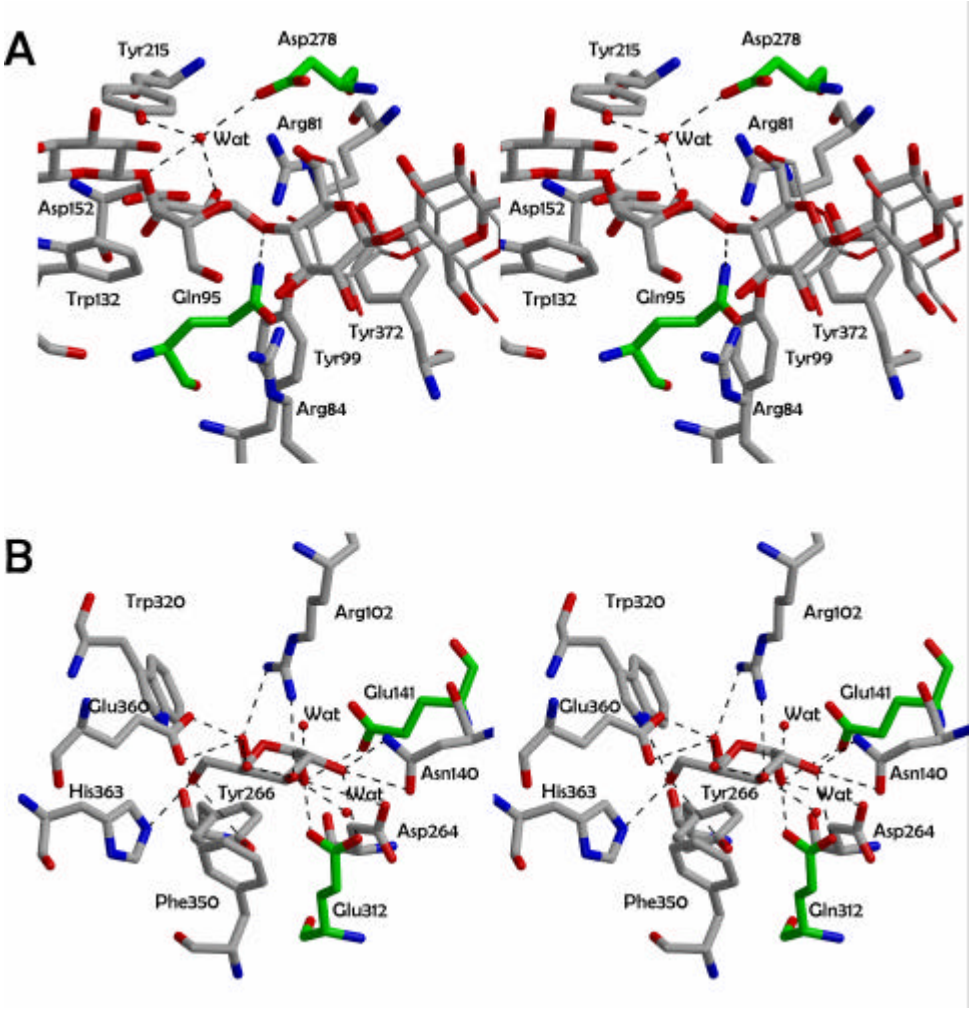

Figure 1.3 Stereo representation of active sites of representative glucosidases

A. The active site of the inverting endoglucanase CelA from *Clostridium thermocellum*, in complex with cellopentaose [pdb code 1KWF (Guerin et al., 2002)]. Catalytic residues are marked in *green*. In this high resolution structure, two alternate conformations for the ligand could be modeled, and interpreted as substrate and product. The product is shown with thinner lines. **B.** Active site of the retaining β-galactosidase from *Thermus thermophilus*, showing a bound galactose [pdb code 1KWK, (Hidaka et al., 2002)]. Catalytic residues are marked in *green*.

1.1.3 Structures of Glycosyl Hydrolases

The first three-dimensional structure of a glucosidase determined was that of hen egg-white lysozyme (Blake et al., 1965). Since this ground-breaking research, glucosidase structures have been the focus of intense research and the additional information afforded from structures of many glycosyl hydrolase families has allowed for further classification. Families whose structures are known may now be grouped into clans based on structural similarity (Henrissat and Bairoch, 1996). This term has been chosen to avoid the idea of an evolutionary relationship implied by the more commonly used superfamily definition. So far 13 clans have been assigned, comprised of 40 of the known 90 families (Table 1-1 and Figure 1.4).

Table 1-1 Structural clans of the glycosyl hydrolases

Clan	Fold *	Families	Mechanism
GH-A	$(\beta/\alpha)_8$ (TIM Barrel)	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51, 53, 59, 72, 79, 86	retaining
GH-B	β-jelly roll	7, 16	retaining
GH-C	β-jelly roll	11, 12	retaining
GH-D	$(\beta/\alpha)_8$ (TIM Barrel)	27, 36	retaining
GH-E	β-propeller	33, 34, 83	retaining
GH-F	β-propeller	43, 62	inverting
GH-G	-	37, 63	inverting
GH-H	$(\beta/\alpha)_8$ (TIM Barrel)	13, 70, 77	retaining
GH-I	α + β	24, 46, 80	inverting
GH-J	-	32, 68	retaining
GH-K	$(\beta/\alpha)_8$ (TIM Barrel)	18, 20	retaining
GH-L	$(\alpha/\alpha)_6$	15, 65	inverting
GH-M	$(\alpha/\alpha)_6$	8, 48	inverting

* Fold assignment from the CAZY database (Coutinho and Henrissat, 1999)

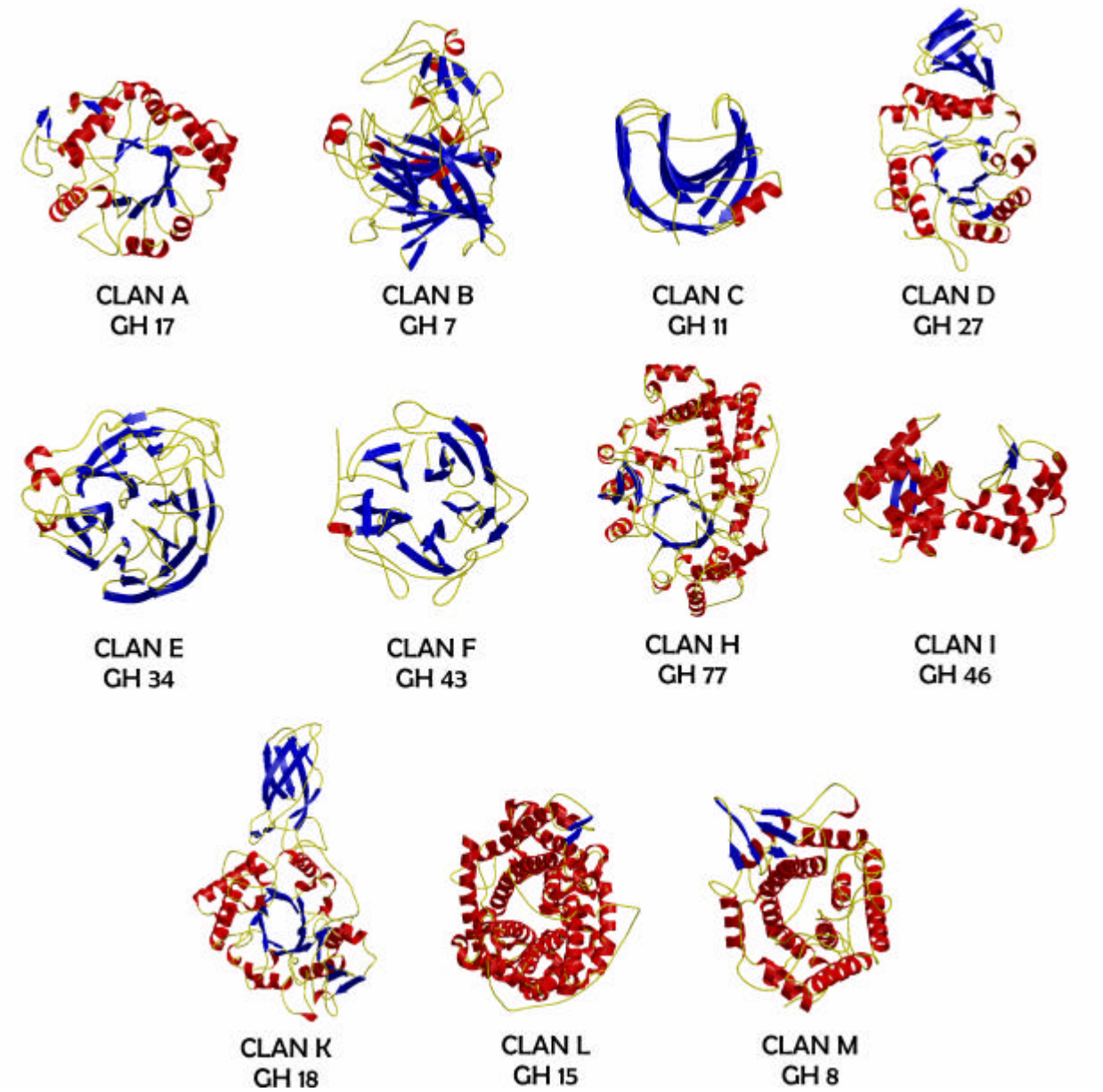

Figure 1.4 Representative members of glycosyl hydrolase clans

Ribbon representations, with helices in *red*, β strands in *blue* and loops in *yellow*. **Clan A**: licheninase, PDB code: 1GHR (Varghese et al., 1994); **Clan B**: cellobiohydrolase I, 1CEL (Divne et al., 1994); **Clan C**: endo-β-1,4-xylanase, 1BCX (Wakarchuck et al., 1994); **Clan D**: α-N-acetylgalactosaminidase, 1KTB (Garman et al., 2002); **Clan E**: neuraminidase, 1L7G (Smith et al., 2002); **Clan F**: endo-1,5-α-L-arabinosidase, 1GYD (Nurizzo et al., 2002); **Clan H**: amylomaltase, 1CWY (Przylas et al., 2000b); **Clan I**: chitosanase, 1CHK (Marcotte et al., 1996); **Clan K**: chitinase A, 1CTN (Perrakis et al., 1994); **Clan L**: glucoamylase, 3GLY (Aleshin et al., 1994); **Clan M**: xylanase, 1H12 (Van Petegem et al., 2003)

Several well-known fold types are represented by the glycosyl hydrolase clans. Clan GH-A, -D, -H, and –K, all contain the classic $(\alpha/\beta)_8$ barrel, also known as the TIM-barrel from the first enzyme in which it was identified, triose phosphate isomerase (Banner et al., 1975). In these, 8 α-helices and 8 β-strands create a barrel-like structure, the core of which is formed by a barrel-shaped parallel β-sheet and surrounded by the α-helices (see Figure 1.4). The active site is formed by the loops which extend from the C-terminal end of the β-strands that form the barrel.

The β-jelly roll of clans GH-B and GH-C is based on the greek key motif which forms the basis of many more complex folding patterns. This motif consists of four β-strands forming an antiparallel sheet where the fourth strand is at the start of the sheet antiparallel to the first, rather than the third. In a β-jelly roll, two of these motifs are joined together. An overview of this and the other protein folds discussed here can be found in Brandon and Tooze (1999).

The third fold type exhibited by glucosidases is the β-propeller of GH-E and GH-F. In these, individual β-sheet domains are arranged in a similar fashion around a core to produce a propeller-like structure. This propeller may have between five and eight blades and the active site is located on the loops that circle the central core and are located on one side of the propeller, in a way homologous to the α/β barrels.

The final fold is epitomised by members of GH-L and GH-M, and is known as the $(\alpha/\alpha)_6$, or the α/α toroid fold (Brandon and Tooze, 1999). In the α/α toroid fold up to seven α-hairpins are arranged in a closed circular array which also forms a barrel. However, unlike the previously discussed folds, the central core is formed by α-helices.

Despite the variety of protein folds which are represented by the already determined glucosidase structures, all structures, regardless of mechanism, belong to one of only three general morphological classes (Figure 1.5). The first class is the pocket. Members of this class are often exo-acting hydrolases, such as β-amylase, or monosaccharidases like β-galactosidase. A pocket is capable of recognising both short substrates and those with large numbers of free saccharide non-reducing ends. The second general class displays a more open groove on the surface of the protein, which allows for several sugar subunits of a polysaccharide to bind. This is typical for endo-acting hydrolases and includes lysozyme and α-amylase, among others. The third class is the smallest and is created by the

covering of an open groove by a long loop. This creates a tunnel through the protein, a morphology found in the cellobiohydrolases (Davies and Henrissat, 1995) and κ-carrageenases (Michel et al., 2001), whose long polysaccharide substrates are suggested to be threaded through it (Rouvinen et al., 1990; Varrot et al., 2003).

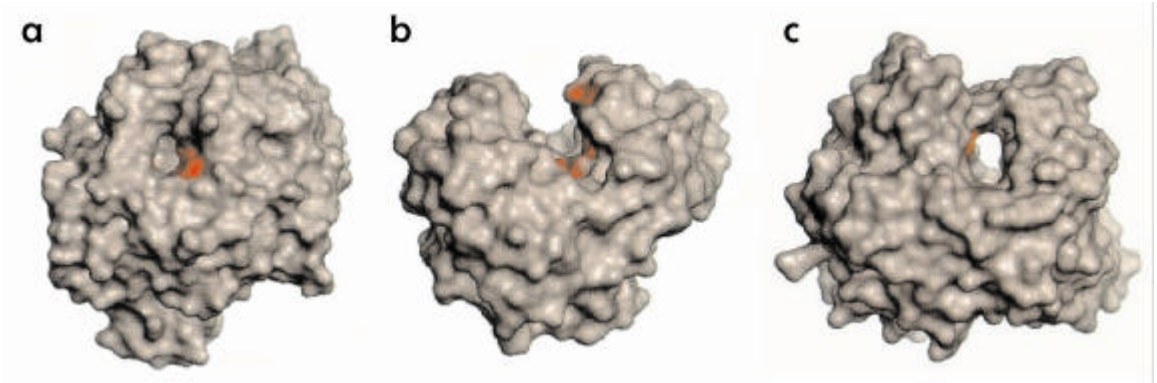

Figure 1.5 Overview of general glycosyl hydrolase morphology

a. Pocket. **b.** Groove. **c.** Tunnel. Active site residues are shaded in red. Adapted from (Davies and Henrissat, 1995).

1.1.4 Family 4 Glycosyl Hydrolases (GH4)

Glycosyl hydrolase family 4 (GH4) is a unique family of glucosidases which exhibit a requirement for unusual cofactors and activation conditions, pointing towards a novel type of reaction mechanism. All except one of the GH4 enzymes investigated so far show a prerequisite for NAD^+ and a divalent cation, and some enzymes also require the addition of a reducing agent such as DTT or β-mercaptoethanol for activity (Raasch et al., 2000; Robrish et al., 1994; Suresh et al., 2002; Thompson et al., 1995; Thompson et al., 1998). The requirement of NAD^+ for the hydrolysis reaction of a glycosidic bond has so far only been described for the GH4 enzymes.

The role of NAD^+ in the mechanism of these enzymes is not yet understood. A sequence pattern at the N-terminus of all GH4 enzymes is related to the 'fingerprint' motif found in many of the classical NAD-binding enzymes (Thompson et al., 1998). Mutations in this GXGS motif demonstrated that it is indeed important for NAD^+ binding and enzymatic activity of *Thermotoga maritima* AglA (Raasch et al., 2002). Though spectroscopic measurements did not reveal the formation of NADH as a product or long-lived intermediate of the catalytic reaction for several GH4 members (Raasch et al., 2002; Thompson et al., 1998; Thompson et al., 1999), a redox reaction during the hydrolysis reaction cannot be excluded.

The GH4 family comprises enzymes with various reaction specificities, including α-glucosidases [E.C. 3.2.1.20; *T. maritima* and *T. neopolitana* AglA (Raasch et al., 2000, (Raasch, 2001)], α-galactosidases [E.C. 3.2.1.22; *Escherichia coli* MelA (Liljestrom and Liljestrom, 1987)], 6-phospho-α-glucosidases [E.C. 3.2.1.122; *Bacillus subtilis* GlvA (Thompson et al., 1998), *Fusobacterium mortiferum* MalH (Thompson et al., 1995) and *Klebsiella pneumoniae* AglB (Thompson et al., 2001)], 6-phospho-β-glucosidases [E.C. 3.2.1.86; *E. coli* CelF (Thompson et al., 1999)] and an α-D-glucuronidase [E.C 3.2.1.139; *T. maritima* Agu (Suresh et al., 2002)]. The GH4 family is unique in being the only family so far identified to contain enzymes which are specific for α-glycosidic bonds as well as those that display β-specificity during cleavage.

Most of the GH4 enzymes so far characterized were uncovered by genetic characterization of the sugar uptake system of the relevant bacterium. The phosphoenol pyruvate-dependent sugar phosphotransferase system (PEP:PTS) was first described for

E.coli (Kundig et al., 1964) and has subsequently been established as the primary mechanism for the accumulation of sugars in a number of both Gram positive and Gram negative bacteria (Meadow et al., 1990; Reizer et al., 1988). The system is comprised of both a membrane-localized and a cytoplasmic protein. These together catalyse the simultaneous phosphorylation and translocation of sugar molecules across the cytoplasmic membrane.

Several members of GH4 have been investigated in more detail. GlvA, from *B. subtilis*, is a 50.5 kDa protein with 6-phospho-α-glucosidase activity capable of cleaving 1,4- linkages (Thompson et al., 1998). It is active as a homotetramer which forms from inactive homodimers in the presence of Mn^{2+}. The enzyme activity is dependant on the presence of NAD^+, but it also shows some activity in the presence of NADH. Directed mutagenesis studies have suggested several putative catalytic residues. Similarly, AglB from *K. pneumoniae* is a 50kDa NAD^+ and metal-ion dependant 6-phospho-α-glucosidase. However AglB is a catalytically active homodimer and investigations have shown it capable of cleaving a large variety of glycosidic linkages (1,1-, 1,3-, 1,4-, 1,6-), enabling the organism to survive on unconventional food sources. Another GH4 enzyme, CelF of *E.coli*, displays the tetrameric quaternary structure of GlvA, but independently of metal-ion presence and it cleaves only phospho-β-glucosides. Enzyme assays clearly indicated that activity was NAD^+ dependant, showing no hydrolysis in the presence of NADH, NADP or NADPH, and also determining that the substrate specificity was similarly broad as that of AglB. In a recent paper, the GH4 protein Agu from *T. maritima*, previously classified by sequence similarity as an α-glucosidase, has been shown to be an α-glucuronidase, thus belonging to a group of enzymes which are generally assigned to family GH67. Like the other GH4 enzymes characterized, Agu requires Mn^{2+} and reducing conditions for activity. However, unusually for a GH4 family member, this enzyme did not display a dependence on added NAD^+.

Until now, no GH4 enzyme has been structurally charactersied. Without the information afforded from a crystal structure, the uniqueness of the cofactor and condition requirements cannot be satisfactorily explained.

1.1.5 AglA

The focus of this work is the crystal structure determination of the family 4 α-glucosidase A, AglA, from the thermophilic bacterium *Thermotoga maritima*. AglA is a 480 amino acid protein with a molecular weight of 55 kDa which was first identified in the characterisation of the maltodextrin utilization gene cluster in *T. maritima* (Bibel et al., 1998) and classified as a GH4 glucosidase by sequence comparison. Heterologously expressed in *E. coli* and purified by Raasch and colleagues (2000), the purified enzyme showed no activity. The following biochemical characterisation was performed in the same study.

Based on the known requirements for GH4 enzymes, the dependence of AglA on NAD^+, metal ions and reducing conditions was investigated and confirmed. Figure 1.6 A. shows that activation of purified AglA was only possible when all three cofactors were present. A more detailed analysis of the metal dependence (Figure 1.6 B) showed that Mn^{2+} has an optimal concentration of 1mM to stimulate AglA activity and a 50% activation at a concentration of 0.01mM. Investigations of other possible divalent metal ions (results not shown) demonstrated that Mn^{2+} is the best metal ion cofactor for AglA, with Co^{2+} showing half, and Ni^{2+} a quarter, of the relative activity. Mg^{2+}, Ca^{2+} and Zn^{2+} were all very poor cofactors. Interestingly, the dimeric existence of AglA was confirmed even in the absence of a metal ion demonstrating that, unlike GlvA, the oligomerisation state is not metal ion dependant.

The cofactor NAD^+ was investigated along with a number of other related substances, including riboflavin, flavin mononucleotide, ATP, biotin, $NADP^+$, NADH, NADPH, however only NAD^+ showed enzyme stimulation (results not shown). An optimal concentration of 0.9mM was determined (Figure 1.6 C) for NAD^+, in the presence of other essential cofactors.

The final requirement of a GH4 enzyme, that of reducing conditions, was analysed in assays containing increasing amounts of DTT, β-mercaptoethanol or L-cysteine (Figure 1.6 D). β-mercaptoethanol was the best stimulant, with 100% relative activity at 600mM, while DTT at 50mM produced 75% relative activity. L-cysteine poorly activated AglA, to only 20% of the activity from β-mercaptoethanol.

Environmental factors influencing the activity of AglA were also investigated. The optimal pH was pH 7.5 (Figure 1.6 E), while the optimum temperature (Figure 1.6 F) was

determined as 90°C, slightly higher than the growth optimum of the organism (80°C), yet still within the growth range of 55°C to 90°C (Huber et al., 1986).

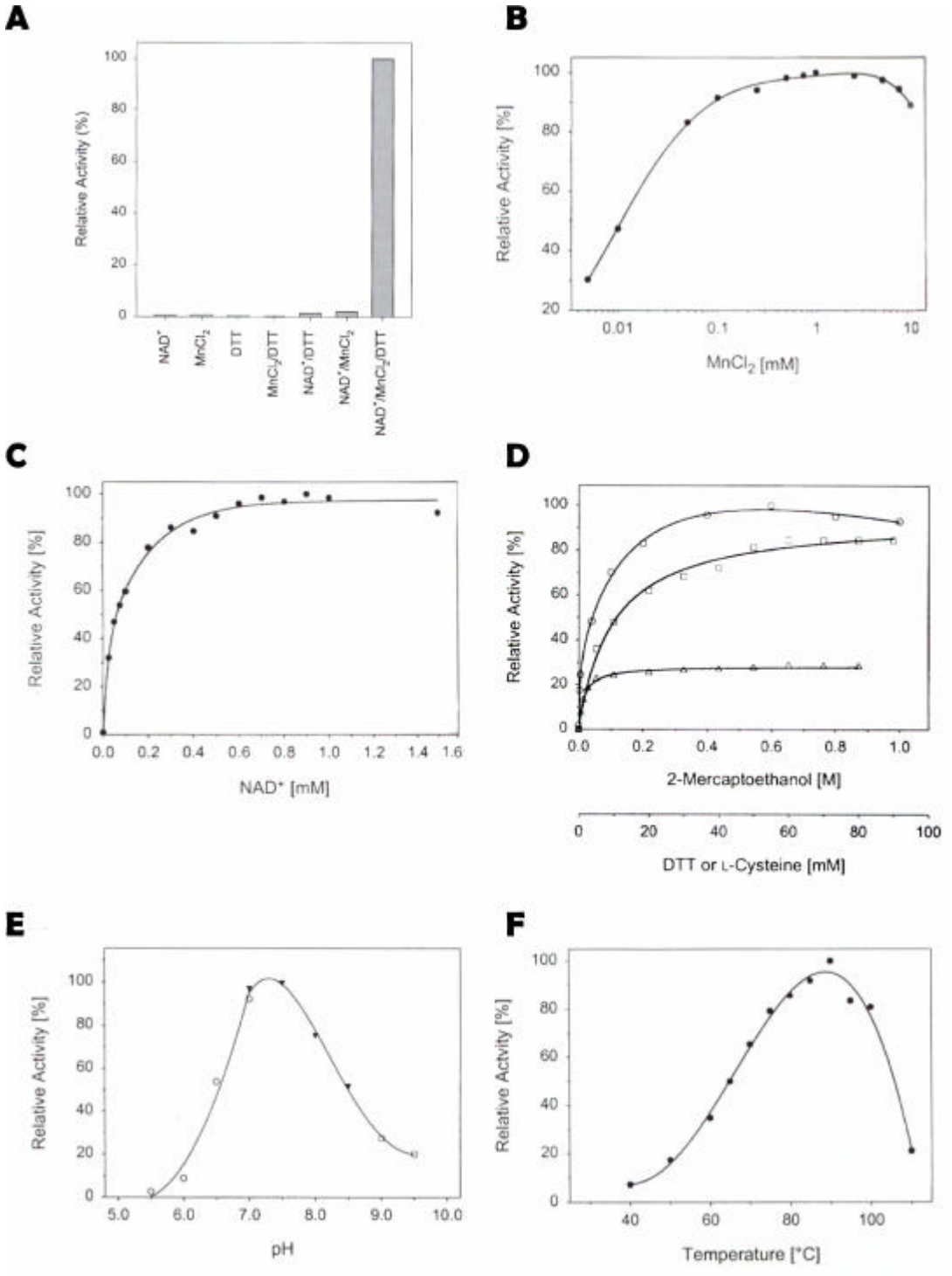

Figure 1.6 Biochemical characterisation of AglA

All assays were done with purified AglA and concentrations of 0.9mM NAD^+, 1mM $MnCl_2$ and 50mM DTT where relevant **A.** Effect of the addition of 0.9mM NAD^+, 1mM $MnCl_2$ and 50mM DTT on the activity of purified AglA. **B.** Effect of Mn^{2+} concentration. **C.** Effect of NAD^+ concentration. **D.** Effect of DTT (*squares*), β-mercaptoethanol (*circles*) or L-cysteine (*triangles*). **E.** pH dependence of AglA activity. **F.** Temperature dependence of AglA activity. All graphs taken from Raasch et. al., 2000.

The substrate specificity of AglA was characterised with a number of substrates displaying different configurations (Table 1-2). AglA demonstrated that it is not discerning for the configuration of the glycosidic bond, being capable of cleaving 1,1-, 1,3-, 1,4-, and 1,6- bonds. However it is specific for the anomeric configuration of the glycone residue, cleaving only α- and not β-glycosidic linkages. AglA also demonstrated specificity for only short oligosaccharides of up to four sugar moieties, and was incapable of cleaving more complex substrates such as starch and chitin.

Table 1-2 Substrate Specificity of AglA

Substrates	Nonsubstrates
?-nitrophenyl-α-Glucoside	α-, β-, γ-Cyclodextrin
Methyl-α-D-glucoside	Soluble Starch
4-*O*-Methylumbelliferyl-α-D-glucoside	Pullulan
n-Dodecyl-β-D-maltoside	Glycogen
Maltose [*O*-α-D-glucopyranosyl-(1→4)-D-glucose]	Mannan
Maltotriose	Laminarin
Maltotetrose	Chitin
Isomaltose [*O*-α-D-glucopyranosyl-(1→6)-D-glucose]	
Panose [*O*-α-D-glucopyranosyl-(1→6)-D-glucopyranosyl-(1→4)-α-D-glucose]	
Trehalose[*O*-α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside]	
Melibiose [*O*-α-D-galactopyranosyl-(1→6)-D-glucose]	
Sucrose [*O*-α-D-glucopyranosyl-(1→2)-β-D-fructofuranoside]	
Raffinose [*O*-α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→2)-β-D-fructofuranoside]	
Turanose [*O*-α-D-glucopyranosyl-(1→3)-β-D-fructofuranose]	
Melizitose [*O*-α-D-glucopyranosyl-(1→3)-*O*-β-D-fructofuranosyl-(2→1)-α-D-glucopyranoside]	

Taken from (Raasch et al., 2000)

1.1.6 Aim of This Work

Glycosyl hydrolases are a large group of sugar cleaving enzymes which play important roles in many diseases and industrial processes. Many of these enzymes have been well characterized by structural and biochemical studies. They all share a similar mechanism of action consisting of an attack on the C1 carbon of the glycosidic bond involving a general acid and a general base in the enzyme active site. Enzymes of glycosyl hydrolase family 4 (GH4), however, have been shown to be unique among the glycosyl hydrolases for their cofactor requirements and conditions for activity. All GH4 members studied require divalent metal ions and NAD^+ as cofactors, but also require strongly reducing conditions to be active.

The mode of action of these enzymes and in particular the role of the redox cofactor is currently unclear. No structures of any GH4 glucosidase have yet been determined. The aim of this work was to determine the X-ray crystal structure of the GH4 enzyme α-glucosidase A from *Thermotoga maritima*. Only the three-dimensional protein structure can form the basis for an understanding of the enzyme function on a molecular basis. To this end, a crystallographic analysis including protein crystallization, solution of the phase problem, model building and structure refinement was to be carried out. The novel protein structure should then be analyzed and compared to known glycosyl hydrolase structures in order to establish the relationship of GH4 to previously defined structural clans.

In addition, co-crystallisation experiments with NAD^+ and divalent metal ions should be performed to study the binding of the known cofactors. Important information on the catalytic mechanism can often be obtained by the determination of co-crystal structures with substrates, products or inhibitors. A further aim was therefore to obtain such structures by crystal soaking or cocrystallisation with maltose. The resulting substrate- and/or cofactor-bound structures may be analysed and the binding modes of these characterised. This structural information may then aid in the understanding of the molecular basis for the unusual co-factor requirement and conditions of activation of AglA and other GH4 enzymes.