Structures and mechanisms of glycosyl hydrolases

The wealth of information provided by the recent structure determinations of many different glycosyl hydrolases shows that the substrate specificity and the mode of action of these enzymes are governed by exquisite details of their three-dimensional structures rather than by their global fold.

Carbohydrates show wide stereochemical variation and can be assembled in so many different fashions that there are over $10^{12}$ possible isomers for a reducing hexasaccharide [1]. Living organisms take advantage of this diversity by using oligosaccharides and polysaccharides for a multitude of biological functions, from storage and structure to highly specific signalling roles. Selective hydrolysis of glycosidic bonds is therefore crucial for energy uptake, cell wall expansion and degradation, and turnover of signalling molecules. As a consequence of saccharide diversity, there is great variety amongst the enzymes that hydrolyze glycosidic bonds, the O-glycosyl hydrolases (EC 3.2.1.x). Heritable deficiencies in glycosyl hydrolases, for example lactose intolerance [2] and a number of lysosomal storage diseases [3], are among the most frequent genetically based syndromes in man.

Mechanisms

Enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires two critical residues: a proton donor and a nucleophile/base [4,5] (Fig. 1). This hydrolysis occurs via two major mechanisms giving rise to either an overall retention, or an inversion, of anomeric configuration [4]. In both the retaining and the inverting mechanisms, the position of the proton donor is identical, in other words it is within hydrogen-bonding distance of the glycosidic oxygen. In retaining enzymes, the nucleophilic catalytic base is in close vicinity of the sugar anomeric carbon. This base, however, is more distant in inverting enzymes which must accommodate a water molecule between the base and the sugar. This difference results in an average distance between the two catalytic residues of ~5.5 Å in retaining enzymes as opposed to ~10 Å in inverting enzymes [6].

Lysozymes were the first glycosyl hydrolases to have their three-dimensional (3D) structures solved [7,8]. The two catalytic amino acids were identified as aspartate and glutamate residues. In most glycosyl hydrolases studied since, only aspartate and/or glutamate residues have been found to perform catalysis. Recent data, however, suggest that other residues may sometimes be involved in glycosidic bond cleavage. Typical examples are viral neuraminidase and bacterial sialidase, where the transition state is thought to be stabilized with the help of a tyrosine [9,10].

Enzymatic hydrolysis of certain oligosaccharides may also take advantage of the natural chemistry of the substrate. Participation of the substrate C2 acetamido group in catalysis by lyzosymes and some retaining chitinases has been suggested from several underrated sources. Hydrolysis by hen egg white lysozyme (HEWL) of N-acetylchitobioside substrates with either a C2 hydroxyl or C2 acetamido group has been studied in detail [11].
Removal of the C2 acetamido group leads to a reduction in $k_{cat}$ of at least 100-fold, a factor which is similar, if not greater, than the effect of mutation of the proposed catalytic base Asp52. Perhaps the most serious evidence for the role of the acetamido substituent at C2, both in the enhancement of the catalytic rate and the retention of the anomeric configuration during hydrolysis, comes from solution studies on $N$-acetylglucosamine-containing saccharides [12,13]. Hydrolysis of these sugars, in solution, not only goes faster than that of C2 hydroxyl-substituted sugars, but with overall retention of anomeric configuration, suggesting that, 'on enzyme', a catalytic base, such as Asp52 in HEWL, is not necessarily required. These observations may also help to explain the apparent anomaly that some enzymes known to catalyze the cleavage of $\beta_1,4$ bonds adjacent to C2 acetylamido-substituted saccharides, whose structures have recently been determined, appear to lack a suitable catalytic base located in an appropriate position. Soluble lytic transglycosylase [14], goose lysozyme [15], hevamine [16], and others fall into this category.

Glycosyl hydrolases have developed various ways to lower the energy barrier of the hydrolysis reaction, such as substrate distortion into a sofa or half-chair conformation [17,18]. It is believed that protonation of the glycosidic substrate distortion into a sofa or half-chair conformation the energy barrier of the hydrolysis reaction, such as Glycosyl hydrolases have developed various ways to lower the role of the acetamido substituent at C2, both in the enhancement of the catalytic rate and the retention of the anomeric configuration during hydrolysis, comes from solution studies on $N$-acetylglucosamine-containing saccharides [12,13]. Hydrolysis of these sugars, in solution, not only goes faster than that of C2 hydroxyl-substituted sugars, but with overall retention of anomeric configuration, suggesting that, 'on enzyme', a catalytic base, such as Asp52 in HEWL, is not necessarily required. These observations may also help to explain the apparent anomaly that some enzymes known to catalyze the cleavage of $\beta_1,4$ bonds adjacent to C2 acetylamido-substituted saccharides, whose structures have recently been determined, appear to lack a suitable catalytic base located in an appropriate position. Soluble lytic transglycosylase [14], goose lysozyme [15], hevamine [16], and others fall into this category.

Because the 3D structures of proteins are more highly conserved than their sequences, several sequence-based families may have related folds. For instance, a structural similarity was suggested for family 11 xylanases and

### Table 1. Structures and mechanisms in various families of glycosyl hydrolases.

<table>
<thead>
<tr>
<th>Family</th>
<th>Enzyme</th>
<th>Organism</th>
<th>EC number</th>
<th>PDB* code</th>
<th>Mechanism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta$-glucosidase</td>
<td>Trifolium repens</td>
<td>3.2.1.21</td>
<td>-</td>
<td>retaining</td>
<td>[31]</td>
</tr>
<tr>
<td>2</td>
<td>$\beta$-galactosidase</td>
<td>Escherichia coli</td>
<td>3.2.1.23</td>
<td>1BGL</td>
<td>retaining</td>
<td>[32]</td>
</tr>
<tr>
<td>5</td>
<td>endoglucanase A</td>
<td>Clostridium cellulolyticum</td>
<td>3.2.1.4</td>
<td>-</td>
<td>retaining</td>
<td>[33]</td>
</tr>
<tr>
<td>6</td>
<td>cellobiohydrolase II</td>
<td>Trichoderma reesei</td>
<td>3.2.1.91</td>
<td>3CBH</td>
<td>inverting</td>
<td>[28]</td>
</tr>
<tr>
<td>7</td>
<td>endoglucanase</td>
<td>Thermoanospora fusca</td>
<td>3.2.1.2</td>
<td>1TML</td>
<td>inverting</td>
<td>[34]</td>
</tr>
<tr>
<td>8</td>
<td>endoglucanase</td>
<td>Humicola insolens</td>
<td>3.2.1.4</td>
<td>-</td>
<td>retaining</td>
<td>[35]</td>
</tr>
<tr>
<td>9</td>
<td>endoglucanase D</td>
<td>Clostridium thermocellum</td>
<td>3.2.1.4</td>
<td>-</td>
<td>inverting</td>
<td>[36]</td>
</tr>
<tr>
<td>10</td>
<td>xylanase A</td>
<td>Streptomyces lividans</td>
<td>3.2.1.8</td>
<td>1XAS</td>
<td>retaining</td>
<td>[37]</td>
</tr>
<tr>
<td>11</td>
<td>xylanase</td>
<td>Bacillus circulans</td>
<td>3.2.1.8</td>
<td>1BCX</td>
<td>retaining</td>
<td>[38]</td>
</tr>
<tr>
<td>13</td>
<td>$\alpha$-amylase</td>
<td>Aspergillus oryzae</td>
<td>3.2.1.11</td>
<td>6TAA</td>
<td>retaining</td>
<td>[39]</td>
</tr>
<tr>
<td>14</td>
<td>$\beta$-amylase</td>
<td>Glycine max</td>
<td>3.2.1.2</td>
<td>1BTC</td>
<td>inverting</td>
<td>[40]</td>
</tr>
<tr>
<td>15</td>
<td>glucoamylase</td>
<td>Aspergillus awamori</td>
<td>3.2.1.3</td>
<td>3GLY</td>
<td>inverting</td>
<td>[41]</td>
</tr>
<tr>
<td>16</td>
<td>$\beta$-1,3-1,4-glucanase</td>
<td>Bacillus sp.</td>
<td>3.2.1.73</td>
<td>1BYP</td>
<td>retaining</td>
<td>[42]</td>
</tr>
<tr>
<td>17</td>
<td>$\beta$-1,3-1,4-glucanase</td>
<td>Hordeum vulgare</td>
<td>3.2.1.73</td>
<td>1GHR</td>
<td>unknown</td>
<td>[43]</td>
</tr>
<tr>
<td>18</td>
<td>chitinase</td>
<td>Serratia marcescens</td>
<td>3.2.1.14</td>
<td>1CTN</td>
<td>retaining</td>
<td>[44]</td>
</tr>
<tr>
<td>19</td>
<td>chitinase</td>
<td>Hordeum vulgare</td>
<td>3.2.1.14</td>
<td>1BAA</td>
<td>retaining</td>
<td>[45]</td>
</tr>
<tr>
<td>20</td>
<td>chitobiase</td>
<td>Serratia marcescens</td>
<td>3.2.1.52</td>
<td>-</td>
<td>retaining</td>
<td>[46]</td>
</tr>
<tr>
<td>22</td>
<td>lysozyme</td>
<td>Hen egg white</td>
<td>3.2.1.17</td>
<td>1HEL</td>
<td>retaining</td>
<td>[47]</td>
</tr>
<tr>
<td>23</td>
<td>lysozyme</td>
<td>Goose</td>
<td>3.2.1.17</td>
<td>153L</td>
<td>unknown</td>
<td>[48]</td>
</tr>
<tr>
<td>24</td>
<td>lysozyme</td>
<td>Bacteriophage T4</td>
<td>3.2.1.17</td>
<td>1LYD</td>
<td>unknown</td>
<td>[49]</td>
</tr>
<tr>
<td>33</td>
<td>sialidase</td>
<td>Salmonella typhimurium</td>
<td>3.2.1.18</td>
<td>2SIL</td>
<td>retaining</td>
<td>[50]</td>
</tr>
<tr>
<td>34</td>
<td>neuraminidase</td>
<td>Influenza virus B</td>
<td>3.2.1.18</td>
<td>1NSB</td>
<td>retaining</td>
<td>[51]</td>
</tr>
<tr>
<td>45</td>
<td>endoglucanase V</td>
<td>Humicola insolens</td>
<td>3.2.1.4</td>
<td>1ENG</td>
<td>inverting</td>
<td>[52]</td>
</tr>
</tbody>
</table>

*Protein Data Bank. †This family is predicted to have a retaining mechanism [24,25]. ‡Tews, Z Dauter, KS Wilson & CE Vorgias, [abstract 038], 4th European Workshop on Crystallography of Biological Macromolecules, Como, Italy, May 1995.
family 12 cellulases [22], whereas family 7 cellulases have been found to have an arrangement of catalytic residues and a fold similar to those of the β-1,3-glucanases and β-1,3-1,4-glucanases of family 16 [23]. More recently, families 1, 2, 5, 10, 17, 30, 35, 39 and 42 were proposed to have evolved from a common ancestor [24, 25]. All of these family groupings substantiate the strict conservation of the catalytic machinery and mechanism during evolution. A grouping indicating structural similarity for various lysozymes (families 22, 23 and 24) and family 19 plant chitinases has also been proposed [26]. This grouping, however, includes enzymes known to operate with retention of configuration, such as HEWL, and some with a substantially different arrangement of catalytic amino acids, such as family 19 plant chitinases, which are inverting enzymes [27].

Table 1 reports the various glycosyl hydrolases families for which at least one 3D structure has been determined, together with the mechanism of glycosidic bond hydrolysis, in cases where it is known. The substrates for these enzymes are shown in Figure 2. Many glycosyl hydrolases have a modular structure consisting of a catalytic domain and one or more non-catalytic domains, some of which are involved in substrate binding, but most of which have unknown functions. Figure 3 shows the main folds found in the catalytic domains of selected glycosyl hydrolases and, for families 6 and 7, the comparison of the structures of cellobiohydrolases with those of the corresponding endoglucanases (see below).

**Active-site topologies**
Although many protein folds are represented in the 22 families for which a 3D structure is known, the overall topologies of the active sites fall into only three general classes, regardless of whether the enzyme is inverting or retaining. These three topologies (Fig. 4) can, in principle, be built on the same fold, with the same catalytic residues.

**Pocket or crater**
This topology (Fig. 4a) is optimal for the recognition of a saccharide non-reducing extremity and is encountered in monosaccharidases such as β-galactosidase, β-glucosidase, sialidase and neuraminidase, and in exopolysaccharidases such as glucoamylase and β-amylase. Such exopolysaccharidases are adapted to substrates having a large number of available chain ends, such as native starch granules, whose radial structure exposes all the non-reducing chain ends at the surface. On the other hand, these enzymes are not very efficient for fibrous substrates such as native cellulose, which has almost no free chain ends.
Fig. 3. Ribbon representation of the main fold of the catalytic domain in various glycosyl hydrolase families (see Table 1 and text). β strands are shown in cyan and α helices in red. (Figure produced using the program MOLSCRIPT [49].)

Cleft or groove
This ‘open’ structure (Fig. 4b) allows a random binding of several sugar units in polymeric substrates and is commonly found in endo-acting polysaccharidases such as lysozymes, endocellulases, chitinases, α-amylases, xylanases, β-1,3-1,4-glucanases and β-1,3-glucanases.

Tunnel
This topology (Fig. 4c) arises from the previous one when the protein evolves long loops that cover part of the cleft. Found so far only in cellobiohydrolases, the resulting tunnel enables a polysaccharide chain to be threaded through it [28]. A comparison of the cellobiohydrolases in families 6 and 7 with the corresponding endoglucanases is shown in Figure 3. The loops that cause the catalytic centres of cellobiohydrolases to lie within enclosed tunnels can be seen clearly and, for the family 6 enzymes, are also illustrated by the surface representations in Figures 4b and 4c. This topology allows these enzymes to release the product while remaining firmly bound to the polysaccharide chain, thereby creating the conditions for processivity (Fig. 5). It remains unclear at present whether the substrate initially penetrates the active site in an ‘exo’ fashion by one of the two entrances of the tunnel or whether the loops that close the active site can ‘open’ occasionally to allow a random binding followed by the processive action. In either case it should be noted that, depending on the mechanism (inverting or retaining) and the exact position of the cleavage point with respect to the several subsites, the directionality of the enzyme motion along the chain may change. For instance, cellobiohydrolase II of Trichoderma reesei proceeds towards the reducing end of cellulose, whereas the reverse was suggested for cellobiohydrolase I from the same organism [23]. Processivity is probably a key factor for the efficient enzymatic degradation of insoluble microcrystalline cellulose.

Concluding remarks
Orengo et al. [29] have shown that certain protein folds (superfolds) occur more often than others. More precisely, only nine superfolds are sufficient to describe the folding in ~30% of all proteins and it is thought that the total number of protein folds is not more than a few thousand [29,30]. So far, from the 22 families of glycosyl hydrolases for which a 3D structure has been
Fig. 4. The three types of active site found in glycosyl hydrolases. (a) The pocket (glucoamylase from A. awamori). (b) The cleft (endoglucanase E2 from T. fusca). (c) The tunnel (cellulbiohydrolase II from T. reesei). The proposed catalytic residues are shaded in red. (Molecular surface diagrams were prepared using the MOLVIEWER program [M Hartshorn, unpublished program].)

Fig. 5. The mechanism of processivity of cellulbiohydrolases. Once a disaccharide product is liberated (shown as two linked circles), the enzyme remains bound to the polysaccharide chain by several subsites, the ‘lid’ closing the active site and, for retaining enzymes, the glycosyl enzyme. The two empty sites, and perhaps other factors such as loop movements, provide the driving force for enzyme motion along the chain (or chain threading along the enzyme’s active site) by two sugar units. Hydrolysis can then proceed iteratively until enzyme movement is stopped by steric factors, or until the loops that close the active site move and release the polysaccharide chain.
There are presently 52 families in the classification based on sequence similarities (B Henrissat and A Bairoch, unpublished data); 3D structures are known for about 40% of these families. The first enzyme structure to be solved, more than 30 years ago, was that of hen egg white lysozyme, a glycosyl hydrolase. At the present pace of structural investigations, the 3D fold for all of the remaining families could be determined within a few years.\[...

Acknowledgements: We thank those who gave coordinates not yet available from the Protein Data Bank: Tracey Barrett (Mill Hill) for the cyanoergic β-glucosidase of Trichologon repes; Zygmunt and Ulla Derewenda (University of Alberta) for xylanase A from Streptomyces lividans; Valérie Ducros (CNRS Marseille) for endoglucanase A from Clostridium cellulolyticum; Michel Juy (CNRS Marseille) for endoglucanase D from Clostridium thermocellum; and Brian Matthews (University of Oregon) for β-galactosidase from E. coli and for goose lysozyme. We would like to thank Mike Hartshorn (University of York) for help with the MOLVIEWER program and the production of figures. This work has been funded by the BIO2 CT 94 3018 research grant from the European Community.

References
1. Laine, R.A. (1994). A calculation of all possible oligosaccharide iso-
mers both branched and linear yields 1.05 x 1012 structures for a redu-
cing hexasaccharide: the Isomer Barrier to development of single-
method saccharide sequencing or synthesis systems. Glycobio-
logy 4, 759-767.
chem. 60, 257-280.
6. Blakley, G.O. (1994). The crystal structure of an endoglucanase from Cen-

Gideon Davies, Department of Chemistry, University of York, Heslington, York, YO1 5DD, UK and Bernard Henrissat, Centre de Recherches sur les Macromolécules Végétales, CNRS, BP 53, F-38041 Grenoble Cedex 9, France.