Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Multiple rewards from a treasure trove of novel glycoside hydrolase and polysaccharide lyase structures: new folds, mechanistic details, and evolutionary relationships

Shinya Fushinobu[1], Victor D Alves[2] and Pedro M Coutinho[3]

Recent progress in three-dimensional structure analyses of glycoside hydrolases (GHs) and polysaccharide lyases (PLs), the historically relevant enzyme classes involved in the cleavage of glycosidic bonds of carbohydrates and glycoconjugates, is reviewed. To date, about 80% and 95% of the GH and PL families, respectively, have a representative crystal structure. New structures have been determined for enzymes acting on plant cell wall polysaccharides, sphingolipids, blood group antigens, milk oligosaccharides, *N*-glycans, oral biofilms and dietary seaweeds. Some GH enzymes have very unique catalytic residues such as the Asp-His dyad. New methods such as high-speed atomic force microscopy and computational simulation have opened up a path to investigate both the dynamics and the detailed molecular interactions displayed by these enzymes.

Addresses
[1] Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[2] Centro de Investigação Interdisciplinar em Sanidade Animal, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, Lisboa 1300-477, Portugal
[3] Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7257, Marseille Cedex 9 13288, France

Corresponding author: Fushinobu, Shinya (asfushi@mail.ecc.u-tokyo.ac.jp)

Current Opinion in Structural Biology 2013, **23**:652–659

This review comes from a themed issue on **Protein-carbohydrate interactions**

Edited by **Gideon J Davies** and **Bernard Henrissat**

For a complete overview see the Issue and the Editorial

Available online 28th June 2013

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2013.06.001

Introduction

The functions of carbohydrates and glycoconjugates can be categorized into three major areas: firstly, energy storage (e.g. starch and glycogen); secondly, structural framework in rigid (e.g. cellulose and hemicelluloses) or in gel networks (e.g. agar and pectin) and thirdly, other higher order functions such as molecular targeting and cell–cell recognition (e.g. glycoproteins and proteoglycans on the cell surface). The large variety of biological functions of carbohydrates depends on their structural diversity, resulting from natural combinations of large panels of monosaccharides (aldoses, ketoses and their stereoisomers) and of glycosidic bond variants with two anomers (e.g. α-1,4-, α-1,6-, β-1,4-, etc.) that can be interconnected with different degrees of polymerization and branching. Concomitantly with the enormous diversity of carbohydrates currently available in the biosphere, a wide variety of carbohydrate-degrading enzymes have emerged through the molecular evolution of original scaffolds. The Carbohydrate-Active enZymes (CAZy) database (http://www.cazy.org/) endeavored to describe all known CAZy families and to reflect functional and structural knowledge of these enzymes in what gradually became a reference map for researchers exploring the vast area of biological systems related to carbohydrates [1]. The numerical classification system of glycoside hydrolases (GHs), which established the initial 35 families in 1991 [2], was based on their sequence, and hence structure, but independent of their specificity. Similar principles were applied to the classification of polysaccharide lyases (PLs), presented online since 1998 but only recently described [3]. When Henrissat *et al.* summarized the status of the database in 2008, CAZy contained 113 GH and 18 PL families that cleave glycosidic bonds [4]. The expansion of the number of families continued at steady pace, and as of end of March 2013 CAZy describes families up to GH132 and PL22, respectively (Figure 1). Structural characterization methods, protein X-ray crystallography in particular, have become widely used over the last 15 years as structural knowledge on 'CAZymes' is vital for thorough understanding of their action and for molecular engineering in view of their wide range of industrial applications. Although some families have been deleted with time upon identification of functional misassignments, a total of 100 GH and 20 PL families (out of 126 and 21 existing GH and PL families, respectively) have at least one member with a three-dimensional structure available in the PDB. With the advances in structural biology outputs, it is not uncommon to combine the creation of a new family with the determination of its reference structure in a single publication (e.g. GH124 and GH126) [5,6]. In other cases, protein structures issued from structural genomics projects preceded a relevant structural report of the family. In the cases of families GH125 and GH130, several structures derived of genome-identified putative proteins were already available in the PDB when the corresponding new families were established via discovery of the enzymatic activity of homologous proteins [7,8]. In this review, we overview

Figure 1

GH families

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	
	62	63	64	65	66	67	68		70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132								

PL families

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20
21	22																	

Current Opinion in Structural Biology

GH and PL families in the light of representative crystal structures. Families with white background indicate that these families lack representative crystal structures. Families with black background (white characters) indicate that their first structural reports were published after 2008. For underlined families (GH76 and GH130), only structure data from structural genomics projects are available. Open cells (GH21, GH40, GH41, GH60, GH61, GH69 and PL19) indicate deleted families. GH61 was reclassified as family AA9. GH104–GH132 and PL20–PL22 were created after 2008.

the GH and PL families by selecting 'family first' three-dimensional structures reported after 2009. These structures greatly contribute as the fold representatives to fundamental and applied studies of enzymes within the corresponding family and as landmarks to reveal unsuspected evolutionary and mechanistic relationships between families.

Cellulases, hemicellulases and enzymes that degrade plant polysaccharides

Cellulases and related enzymes have always attracted substantial interest as key actors in the production of biofuels and biorefineries from cellulosic biomass. Unsurprisingly, new insights have emerged in regards to recalcitrant substrates such as lignocellulose. One of the more notable discoveries in this area is the finding that GH61 enzymes, known 'cellulase-enhancing factors', are in fact copper-dependent monooxygenases (reviewed elsewhere in this issue). This family was very recently renamed as family 9 of Auxiliary Activities (AA9) as the corresponding enzymes can be seen as auxiliary activities to the cleavage of glycosidic bonds by GHs and PLs [9], and whose implications go much beyond the scope of this review.

Following the analysis of one of the dockerin-containing cellulosomal proteins from *Clostridium thermocellum* associated with an unknown module and upregulated on a crystalline cellulose substrate, Brás *et al.* characterized an inverting *endo*-β-1,4-glucanase as the founder of the new GH124 family [5•]. The GH124 enzyme exhibits a novel α_8 superhelical fold, where seven α-helices encircle a central α-helix that constitutes the hydrophobic core of the protein. The catalytic residue (Glu96) is located at the C-terminus of the central helix. The catalytic module of this enzyme is a globular protein with a flattened face that, contrary to most glucanases and carbohydrate-binding modules (CBMs), does not contain an evident hydrophobic and deep substrate-binding cleft. The complex structure contains two cellotriose molecules bound with numerous direct polar contacts at subsites ranging from −4 to −2 and from +1 to +3 (Figure 2). The central core residues of the active site, including the catalytic acid Glu96, are structurally conserved with those of GH23 family lysozyme G, although there is no candidate for a catalytic base. The missing catalytic base necessary for the direct nucleophilic attack on the water molecule could be replaced by a remote amino acid-bound water chain, in a Grotthuss-type mechanism, or via thiols, acetate, or phosphate small nucleophilic ions.

GH93 is a relatively small family containing retaining *exo*-acting glycoside hydrolases from fungi that degrade linear arabinan. The first structure of a family GH93 member was determined for *exo*-α-1,5-L-arabinanase from the phytopathogenic fungus *Fusarium graminearum* [10]. This family is now a part of clan GH-E, sharing substantial structural similarity with GH33 sialidases/neuraminidases, but differing in the catalytic residues. Known GH93 members present a six-bladed β-propeller fold. In the active site, an α-1,5-L-arabinobiose where each moiety adopts an E_3 conformation is bound to subsites −1 and −2. The catalytic residues (Glu170 and Glu242 acting as nucleophile and acid/base, respectively) have been proposed from structural analysis and site-directed mutagenesis. Tyr337 is present at an equivalent position of the nucleophile Tyr residue of GH33 sialidases. However, Tyr337 to alanine mutation resulted in a reduction of activity of 93%, indicating that