Hypothesis

A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of *all* glycoside hydrolases

W. Nerinckx*, T. Desmet, M. Claeyssens

Laboratory of Biochemistry, Department of Biochemistry, Physiology and Microbiology, Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

Received 19 November 2002; revised 23 December 2002; accepted 5 February 2003

First published online 19 February 2003

Edited by Judit Ovádi

Abstract An in silico survey of the −1 subsite of all known 3D-structures of *O*-glycoside hydrolases containing a suitably positioned ligand has led to the recognition – apparently without exceptions – of a transition state stabilising hydrophobic platform which is complementary to a crucial hydrophobic patch of the ligand. This platform is family-specific and highly conserved. A comprehensive list is given with examples of enzymes belonging to 33 different families. Several typical constellations of platform – protein residues are described.

Key words: Glycoside hydrolase; Hydrophobic platform; Ligand; Transition state

1. Introduction

O-Glycoside hydrolases (EC 3.2.1.x) are abundantly present in nature. On the basis of amino acid sequence similarity, this structurally variable group of enzymes has hitherto been classified into 85 different families [1,2], and observation of fold conservation allowed some families to be grouped in clans [3]. Although the 3D-structures of hundreds of *O*-glycoside hydrolases have already been solved, data are still missing for representatives of more than 40 families. Of prime importance for insight into ligand recognition, substrate specificity and modus operandi, is the availability of structural data of enzymes containing a ligand within the active centre, especially in the crucial −1 subsite since there the carbohydrate's conformation and configuration change during the hydrolysis reaction. For enzymes in more than 30 *O*-glycoside hydrolase families, this has been addressed by the use of specific inhibitors, cryo-techniques, or site-directed mutagenesis. A major breakthrough is the development of 2-deoxy-2-fluoro-derivatives, allowing the covalent trapping of the glycosyl-enzyme intermediate of many retaining hydrolases [4].

With almost no exceptions, *O*-glycoside hydrolases possess at their active centre two mechanistically important carboxyl functions at the subsites' −1/+1 junction [5]: (i) a *syn* or *anti* positioned [6] proton donor close to the oxygen atom of the scissile glycosidic bond; and (ii) a nucleophilic carboxylate residue (retaining hydrolases) or a solvent-nucleophile assisting carboxylate residue (inverting hydrolases). For several decades there have been major controversies in the literature concerning the mechanistic details of a glycoside hydrolysis, more specifically, to what extent stereoelectronic effects are operative; these apparently have ended with a highly authorative article by P. Deslongchamps [7]. The nucleophilic displacement reaction is borderline SN_2–SN_1, in which C_1 and O_5 at the stage of the transition state have an sp^2-hybridised geometry with considerable mutual double bond character. The current understanding of glycosidase mechanisms has recently been thoroughly reviewed [8].

2. Hydrophobic platforms in the −1 subsite of glycoside hydrolases

A pre-arranged complementarity [9] which specifically stabilises the transition state [10] is important in enzyme catalysed reactions. Fig. 1 (top left) shows a D-hexapyranoside in a ground state $^4C_1$ chair conformation, with the C_5-hydroxymethylene functionality occupying an equatorial position. Corresponding transition states should have C_2, C_1, O_5 and C_5 residing within the same plane allowing orbital overlap between a ring-oxygen's lone pair and the scissile bond, as is the case for a $^4H_3$ half-chair, an E_3 envelope and a $^{2,5}B$ boat conformation (respectively Fig. 1 top right, bottom left and bottom right). In this order, these three conformations are neighbours within the pseudorotational sequence of a double bond-containing six-membered ring. Note that the next pseudorotational neighbour, an E_4 envelope, is unlikely as a transition state candidate because of an unfavourable C_3 hydroxyl–C_5 hydroxymethylene syn-diaxial position. Also, a rigorous $^{2,5}B$ boat transition state conformation is unlikely because it contains an eclipsed C_3–C_4 bond and an unfavourable 2,5-flagpole steric interaction. This however can be relieved by a small twist in the direction of a neighbour in the pseudorotational sequence.

It is clear from Fig. 1 that in these putative transition states the C_4–C_5–C_6 hydrophobic patch at the B-face [11] has substantially tilted out of the equatorial plane, relative to the situation in the ground state. As a sialic acid glycosidic ring

*Corresponding author. Fax. (32)-9-264 5332.
E-mail addresses: wim.nerinckx@rug.ac.be (W. Nerinckx), t.desmet@rug.ac.be (T. Desmet), marc.claeyssens@rug.ac.be (M. Claeyssens).

0014-5793/03/$22.00 © 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00148-0

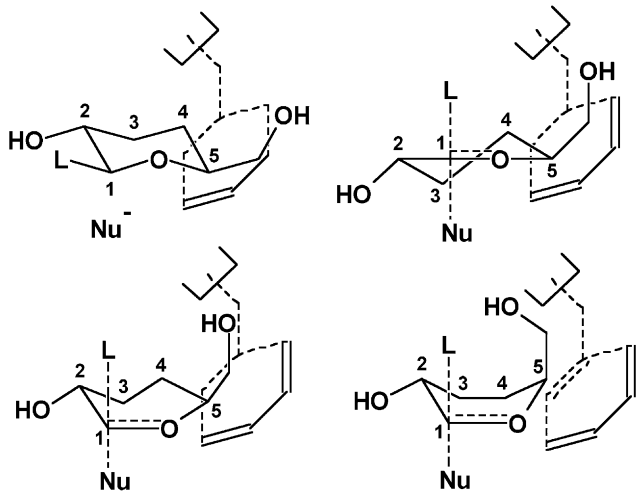

Fig. 1. A β-hexapyranoside in a ground state $^4C_1$ chair (top left), and transition state $^4H_3$ half-chair (top right), E_3 envelope (bottom left) and $^{2,5}B$ boat (bottom right) conformations. Shown with a phenylalanine protein residue as a hydrophobic platform at the carbohydrate's B-face close to the C_5-hydroxymethylene group. L = Leaving group; Nu = nucleophile.

shows a ground state $^2C_5$ conformation (corresponding to $^1C_4$ for penta/hexapyranosides), α-sialidases will probably hydrolyse their substrate through a $^4H_3$ transition state conformation (corresponding to $^3H_4$ for penta/hexapyranosides) [12–14]; its hydrophobic patch now however resides at the ring-C_3-methylene group (see further). Family 11 xylanases have been suggested to operate through a $^{2,5}B$ conformation [15,16]. The same has been proposed for family 6 *Trichoderma reesei* cellobiohydrolase II (Cel6A) [17] and the occurrence of a $B_{2,5}$ transition state has been suggested for family 26 β-mannanases [18]. Both hexapyranoside boat conformations however would also present an unfavourable C_5 hydroxymethylene, respectively C_3 hydroxyl, syn-diaxial interaction versus the leaving group at C_1; the pseudorotational neighbour without this interaction is respectively an E_3 and an ^4E envelope. α(1,2)-Mannosidases from families 38 and 47 probably operate through a $^3H_4$ transition state conformation [19–22] (see further). However, in all these instances a similar tilt of the corresponding hydrophobic patch occurs.

This raises an obvious question: are *all* *O*-glycoside hydrolases equipped with one or more hydrophobic amino acid residues within the −1 subsite so as to be complementary to the carbohydrate's out-of-plane hydrophobic patch at the stage of the transition state (and not or much less so with the hydrophobic patch in the ground state chair situation)? Since van der Waals' contributions to protein/ligand interactions can be quite large, this question is of relevance. Indeed, for interactions between ligand/enzyme lipophilic surfaces, the free energy contribution is estimated to be −0.1 [23] to −0.2 kJ mol^{-1} Å^{-2} [24,25]. Recent experimental evidence even suggests a value of −0.5 kJ mol^{-1} Å^{-2} [26].

An *in silico* survey of the −1 subsites of all as yet known 3D-structures of ligand/enzyme cocrystallysates and/or trapped *O*-glycosyl-enzyme intermediates leads to the following observations. Independently of being retaining or inverting, or α or β *O*-pyranoside-glycoside hydrolases, *all* available structures possess within subsite −1, at approximately the same position i.e. at the B-face and close to the C_5-hydroxymethylene function of the ligand-carbohydrate moiety, a family-specific hydrophobic platform consisting of one (or two) aromatic amino acid residue(s) (Figs. 1 and 2). Only three variations occur with a non-aromatic platform: (i) the family 11 xylanases where, opposite to the xylopyranoside's ring-C_5-methylene function, a non-aromatic hydrophobic protein residue is found; (ii) the family 14 (gluco)amylases containing a discrete non-aromatic hydrophobic system; and (iii) the clan-E sialidases where the ring-C_3-methylene function of the neuraminic acid analogous ligand also has a non-aromatic hydrophobic protein residue in juxtaposition (see further). Subsequent checking with the PFAM sequence-alignment server [27] shows that these residues are highly conserved within each family. Table 1 presents a selection of representatives of 33 different hydrolase families, the relevant PDB files and the nature of the respective hydrophobic platforms. Fig. 2 shows a typical situation in many β(1,4)-glucoside hydrolases.

Although this survey cannot be exhaustive, the presently available 3D-data show no exceptions. When more than one structure with a suitably situated ligand has been published for the same enzyme, the entry with the covalent glycosyl-enzyme intermediate (if available) was preferred for inclusion in Table 1.

It must be emphasised that, in X-ray derived structures, all carbohydrate ligands reside in a ground state chair or in a (skew-)boat conformation which is necessarily a local energy minimum, since it is unlikely that enzymes are capable to impose a catalytically significant strain upon their substrate [28]. In several cases (e.g. [15, 18, 29–31]) a (skew-)boat conformation is observed for the carbohydrate moiety in subsite −1, and the energetic penalty must have been paid for by a number of compensating protein–ligand interactions. Such a conformation belongs to the mechanistic itinerary of a glycoside hydrolysis, i.e. the interconversion of a β-chair glycoside – protonated β-(skew-)boat – transition state – protonated α-chair – protonated α-(skew-)boat – α-chair glycoside (or vice versa) [7]. Apart from mechanistic implications [8], an observed (skew-)boat can resemble the transition state more than a chair conformation, and therefore might fit better in the −1 subsite. It nevertheless still has to be a local minimum – in casu the conformation with the lowest energy when the ligand is surrounded by the enzyme's active centre residues (and not by the solvent). However, an experimentally ob-

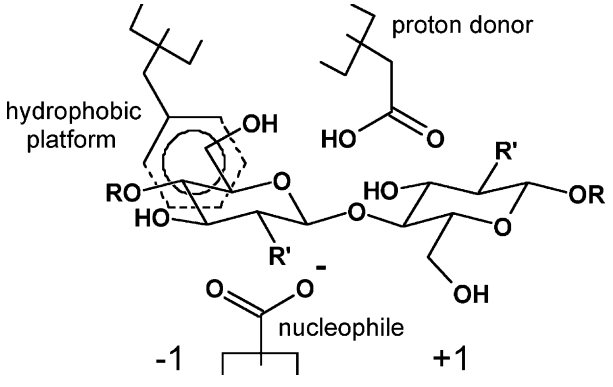

Fig. 2. A typical position of an aromatic hydrophobic platform at the B-face and close to C_5–C_6 of a β-D-hexapyranoside ligand within subsite −1. This situation is observed in many β(1,4)-glycoside hydrolases. R = H or glycopyranosyl, R′ = OH or *N*-acetyl.

Table 1
List of enzyme structures and their hydrophobic platform

Family	Organism	Enzyme (mechanism)[a]	PDB[h]	Platform	Ligand	Shortest distance (Å)
1/A	*B. polymyxa*	1,4-β-glucosidase (r)	1e4i	W398/F414(/Y296)	GluF-glycenz.	3.86(C_5-W_{CD1})/3.68(C_6-F_{CZ})
1/A	*S. alba*	myrosinase (r)	1e73	W457/F473(/Y330)	GluF-glycenz.	3.52(C_5-W_{CE2})/3.56(C_6-F_{CZ})
2/A	*E. coli*	β-galactosidase (r)	1jz0	W568(/Y503)	GalF-glycenz.	3.32(C_5-W_{CE2})
3	*H. vulgare*	1,4-β-glucohydrolase (r)	1iew	F144/L54/M316	GluF-glycenz.	3.95(C_6-F_{CD2})/4.09(C_6-L_{CD1})/3.36(C_6-M_{CE})
5-1/A	*A. cellulolyticus*	1,4-β-endoglucanase (r)	1ece	W319/T328(/Y240)	cellotetraose	4.36(C_4-W_{CE2})/3.99(C_6-T_{CG2})
5-2/A	*B. agaradhaerens*	1,4-β-endoglucanase (r)	1h2j	W262(/Y202)	GluF-glycenz.	3.88(C_5-W_{CD1})
5-3/A	*C. thermocellum*	1,4-β-glucohydrol.[b] (r)	1cen	W313/F320(/Y200)	cellohexaose	3.40(C_5-W_{CD1})/3.64(C_6-F_{CE1})
6	*T. reesei*	1,4-β-exoglucanase (i)	1qk0	Y169	GX-IBn[i]	4.05(C_5-Y_{CZ})
7/B	*F. oxysporum*	1,4-β-endoglucanase (r)	1ovw	Y145	thio-G_5[j]	3.49(C_6-Y_{CE1})
8/M	*C. thermocellum*	1,4-β-exoglucanase[c] (i)	1kwf	W132	cellohexaose	3.55(C_6-W_{CZ3})
9	*T. fusca*	1,4-β-endoglucanase (i)	4tf4	Y429/A426	cellopentaose	4.45(C_6-Y_{CZ})/4.56(C_6-A_{CB})
10/A	*C. fimi*	xylanase (r)	2xyl[h]	W273/W281(/H205)	X_2F-glycenz.	3.87(C_5-$W273_{CE2}$)/3.89(C_5-$W281_{CZ3}$)
11/C	*B. circulans*	xylanase (r)	1bvv	V37	X_2F-glycenz.	3.72(C_5-V_{CG2})
12/C	*S. lividans*	1,4-β-glucanase (r)	2nlr	Y57	G_3F-glycenz.	3.58(C_6-Y_{CB})
13/H	*H. vulgare*	1,4-α-glucanase (r)	1bg9	F143	acarbose (-1/+3)	3.55(C_6-F_{CE1})
14	*B. cereus*	1,4-β-glucoamylase (i)	1b9z	A170/K287/E172	maltose	4.86(C_6-A_{CB})/4.08(C_6-K_{CD})/4.22(C_6-E_{CG})
14	soybean	1,4-β-amylase (i)	1byd	A184/K295/E186	2-deoxymaltose	4.54(C_6-A_{CB})/4.05(C_6-K_{CD})/4.54(C_6-E_{CG})
15/L	*A. awamori*	1,4-α-glucoamylase (i)	1gai	W52	G-H_2-acarbose[k]	3.56(C_6-W_{CD2})
16/B	*B. macerans*	lichenase (r)	1byh	F92[l]	epoxybutyl-G_2[l]	not measurable[l]
18/K	*S. marcescens*	chitinase[d] (r)	1ffr	W539	$(NAG)_6$	3.72(C_5-W_{CD2})
20/K	*S. marcescens*	chitobiase[e] (r)	1c7s	W737/Y669	chitobiose	3.81(C_5-W_{CG})/4.11(C_5-Y_{CE2})
22	*G. gallus*	lysozyme (r)	1h6m	W108	$(Chit)_2$F-glycenz.	3.98(C_6-W_{CD1})
23	*C. atratus*	lysozyme (i)	1lsp	Y147	bulgecin A	3.65(C_5-Y_{CE2})
24/I	BacteriophageT4	lysozyme[f] (i)	148l	F104	glycenz.	3.60(C_6-F_{CD1})
26/A	*C. japonicus*	β-mannanase (r)	1gw1	W360/W379(/Y285)	$(M)_3$F-glycenz.	3.69(C_5-$W360_{CD1}$)/3.47(C_6-$W379_{CH2}$)
27/D	*G. gallus*	α-galactosaminidase (r)	1ktc	W16	NAGal	3.50(C_6-W_{CZ2})
33/E	*B. typhimurium*	α-sialidase (r)	2sim	at C_3: Y342/I38	DD-neuram.[m]	3.76(C_3-Y_{CZ})/3.87(C_3-I_{CD1})
34/E	Influenza virus	α-sialidase (r)	2bat	at C_3: Y406/E119	sialic acid	3.95(C_3-Y_{CZ})/4.79(C_3-E_{CG})
38	*D. melanogaster*	1,2-α-mannosidase (r)	1hxk	W95/Y727	deoxymannojir.	3.62(C_5-W_{CZ2})/3.88(C_6-Y_{CZ})
42/A	*T. thermophilus*	β-galactosidase (r)	1kwk	F350(/H363/Y266)	galactose	3.65(C_4-F_{CZ})
45	*H. insolens*	1,4-β-endogluc.[g] (i)	4eng	Y8[n]	2× cellotriose[n]	not measurable[n]
47	*H. sapiens*	1,2-α-mannosidase (i)	1fo3	F659	kifunensine	3.59(C_5-F_{CD2})
48/M	*C. cellulolyticum*	cellulase (i)	1f9d	W154	2× cellotetraose	3.30(C_5-W_{CD2})
56	*A. mellifera*	hyaluronidase (r)	1fcv	W301/W267	$(hyaluron.)_4$	3.85(C_5-$W301_{CZ3}$)/4.01(C_6-$W267_{CZ3}$)
67	*C. japonicus*	α-glucuronidase (i)	1gql	F327	glucuronic acid	3.65(C_5-F_{CE2})
77/H	*T. aquaticus*	amylomaltase (r)	1esw	W258	acarbose (−3/+1)	4.85(C_6-W_{CG})
83/E	Paramyxovirus	α-sialidase (r)	1e8v	at C_3: Y526/I175	DD-neuram.[m]	3.92(C_3-Y_{CZ})/4.07(C_3-I_{CD1})

[a] r = retaining; i = inverting.
[b] G140Q mutant.
[c] E95Q mutant.
[d] Y390F mutant.
[e] D539A mutant.
[f] T26E/C54T/C97A mutant.
[g] D10N mutant.
[h] Reference in PDB-file 2xyl incorrect, see [39]; all other references as in the respective PDB-file.
[i] *m*-Iodobenzyl β-D-glucopyranosyl-β(1→4)-D-xylopyranoside.
[j] Methyl *S*-β-D-glucopyranosyl-(1→4)-*S*-4-thio-β-D-glucopyranosyl-(1→4)-*S*-4-thio-β-D-glucopyranosyl-(1→4)-*S*-4-thio-β-D-glucopyranosyl-(1→4)-4-thio-α-D-glucopyranoside.
[k] D-Gluco-dehydroacarbose.
[l] Epoxybutyl-trapped glycosyl-enzyme intermediate. Phe92 pinpointed from relative positions of subsite −1 residues.
[m] 2,3-Dehydro-2-deoxy-*N*-acetyl neuraminic acid.
[n] Subsite −1 empty. Tyr8 pinpointed by comparing the subsite −1 residues with the ligand-positions in the adjacent subsites.

served ligand-conformation will be energetically far away from the short-lived high-energy transition state. Therefore, the shortest ligand–protein residue distances in Table 1 can substantially differ from these occurring with the real transition state.

3. Description of the platforms in several families and clans

The amino acid residues that give rise to the hydrophobic platforms are specific for each glycoside hydrolase family and are partly conserved within clans. The situation can however be quite different for non-clan related families and necessitates separate descriptions.

3.1. Clan-A

Clan-A comprises 16 different families that share a common $(\beta/\alpha)_8$ fold and act by a retaining mechanism on various β(1,4)-glycosides. Structures with a suitably positioned ligand are available for members of six families (see Table 1). These enzymes show a 'platform-triad' constellation of protein residues at the B-face near C_5 of the ligand in subsite −1 (Fig. 3): (i) 'left', a highly conserved tryptophan close to the C_4–C_5 hydrophobic patch; (ii) 'top', a not-always-present aromatic residue close to C_6; and (iii) 'right', a conserved tyrosine with its hydroxyl function pointing towards the O_5 ring oxygen (but probably less involved in a hydrophobic interaction). A typical example of the full platform-triad is the left Trp313 /

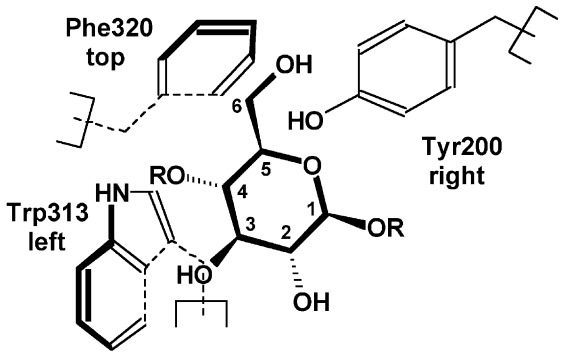

Fig. 3. A typical clan-A Trp/Phe/Tyr platform-triad in the family 5-3 *C. thermocellum* Cel5B β(1,4)-glucanase. R = D-glucopyranosyl.

top Phe320 / right Tyr200 combination in the β(1,4)-glucoside hydrolase Cel5B from *Clostridium thermocellum* (PDB 1cen) belonging to family 5, subfamily 3 [32], as shown in Fig. 3. The same triad is present in the family 1 *Sinapis alba* myrosinase (PDB 1e73) catalysing a thioglycosidic bond hydrolysis and using ascorbate as cofactor [33].

The 'top' residue of the platform-triad varies considerably throughout clan-A, but less so within its families. In family 1, this is phenylalanine or (less frequently) a tryptophan. In the family 2 *Escherichia coli* β-galactosidase example however (PDB 1jz0), a highly organised and conserved Na^+ binding network is situated at the 'top' position, interacting with the C_6-hydroxyl group of the 2-deoxy-2-fluoro-D-galactosyl-enzyme ligand [34]. This family is the only one containing β-glucuronidases, and the metal-ion scaffold is probably nature's solution for an antipole towards the C_6-carboxyl group of a glucuronide. Unfortunately, no 3D-structure of a family 2 glucuronidase with a substrate derivative occupying subsite −1 is available. In families 10 and 26, 'top' is a conserved tryptophan. An interesting variation appears in family 42, where 'left' and 'top' respectively consist of conserved phenylalanine and histidine residues.

Within the family 5 enzymes considerable variation in the 'top' residue occurs. Sequence alignment comparison often shows Phe, Trp, Thr and Val residues; but also Gly, Ala and even Glu can be present. The substrate specificity of this family is quite broad and some members share comparatively closer sequence identity, which has lead to a subdivision into as yet six subfamilies [35–37]. For three of these, structures with a suitable ligand in subsite −1 are available (Table 1). The family 5-1 example has the methyl group of a threonine on 'top'; in 5-2 'top' is absent, while the 5-3 example (Fig. 3) exhibits the more common 'top'-phenylalanine.

3.2. Clan-B

In their recent study of the *Pseudoalteromonas carrageenovora* κ-carrageenase structure and its comparison with other clan-B β-retaining glycoside hydrolases (families 7 and 16), G. Michel and coworkers have observed that, in spite of their wide substrate diversity, all clan-B members feature one aromatic amino acid residue at exactly the position indicated in Fig. 2. They suggest its presence to be "for the correct positioning of the glycoside unit undergoing the nucleophilic substitution". Sequence comparison even indicates that the type of residue is substrate-specific: tryptophan in κ-carrageenases, β-agarases and 1,3-β-glycanases; tyrosine in xyloglucan endotransglycosylases and the family 7 cellulases; and phenylalanine in 1,3–1,4-β-glucan hydrolases (lichenases) [38].

3.3. Family 10 and 11 xylanases

Although their natural substrate lacks C_5-hydroxymethylene groups, an analogous hydrophobic platform is also found within the xylan-degrading enzymes of family 10 and 11. At the B-face of pentapyranosides, a ring-C_5-methylene hydrophobic patch is still present; and again, at the transition state this patch may be recognised by a complementary platform better than it is at the $^4C_1$ chair ground state.

Family 10 xylanases are equipped with a rigorously conserved variant of the clan-A platform-triad in which the 'right' protein residue is now a histidine while the 'left' and 'top' tryptophans appear to be sandwiching the ring-C_5-methylene function of the D-xylopyranosyl moiety. A typical example is the Trp273/Trp281/His205 constellation with the 2-deoxy-2-fluoro-xylobiosyl-enzyme intermediate of the *Cellulomomas fimi* xylanase Xyn10A (CEX; PDB 2xyl) [39], as shown in Fig. 4.

Family 11 xylanases, belonging to clan-C, show a single valine, leucine or isoleucine in juxtaposition to the ligand's ring-C_5-methylene group; e.g. Val37 near the 2-deoxy-2-fluoro-xylobiosyl-enzyme intermediate (with the xylopyranosyl ring at subsite −1 in a $^{2,5}B$ conformation) of the *Bacillus circulans* xylanase example (PDB 1bvv) [15] (Table 1). This clan also comprises family 12 glucoside hydrolases, where the *Streptomyces lividans* 2-deoxy-2-fluoro-cellotriosyl-enzyme intermediate example (PDB 2nlr) [40] shows the usual aromatic residue (Fig. 2), in casu a family-conserved tyrosine, as hydrophobic platform.

3.4. Family 14

Structures containing a ligand within subsite −1 have been published for two members of this family of β-inverting (gluco)amylases (see Table 1) [41,42]. Both reveal an identical system of three highly conserved non-aromatic amino acid residues at the B-face of the ligand, in a similar position as in the above described platform-triad within clan-A: (i) 'left', an alanine with its methyl group pointing towards C_6; (ii) 'top', a lysine with its penultimate methylene close to C_6 (and its amino group hydrogen-bonding to the C_6-hydroxyl

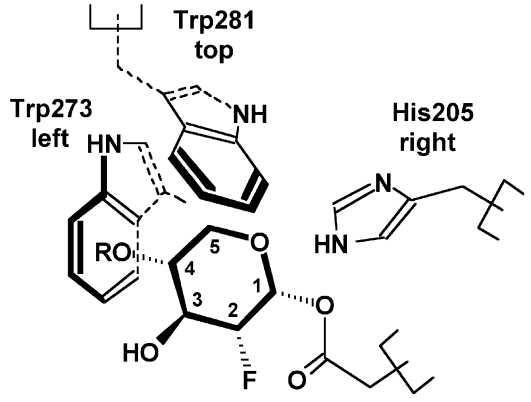

Fig. 4. A typical family 10 clan-A platform-triad variant with a 'right' histidine and with the ligand's ring-C_5-methylene function sandwiched between the 'left' and 'top' tryptophan residues. Shown is the situation with the 2-deoxy-2-fluoro-xylobiosyl-enzyme intermediate of the *C. fimi* xylanase Xyn10A. R = β-D-xylopyranosyl.

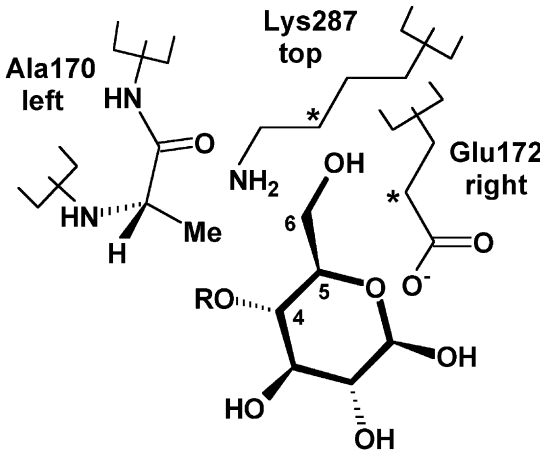

Fig. 5. The family 14 hydrophobic Ala/Lys/Glu platform-triad in *B. cereus* glucoamylase. The methylene groups close to C_6 of the ligand are indicated with an asterisk. R = α-D-glucopyranosyl.

group); and (iii) 'right', a glutamate with its α-methylene near C_6. The situation in the *Bacillus cereus var. myciodes* glucoamylase (PDB 1b9z) [41] is shown in Fig. 5.

3.5. The family 26 mannanases, and family 38 and 47 mannosidases

Although the ground state conformation of a mannopyranoside ring is a $^4C_1$, molecular orbital calculations have shown that a mannopyranosyl cation prefers a $^3H_4$-like geometry over a $^4H_3$ [19]. This suggests that dedicated mannosyl hydrolase families might have evolved to hydrolyse their substrate through a $^3H_4$, ^3E or $B_{2,5}$ transition state.

The family 26 retaining endo-β(1,4)-mannanase A (Man26A) from *Clypeaster japonicus* (formerly *Pseudomonas cellulosa*) belongs to clan-A, its 3D-structure (e.g. PDB 1gw1) [18,43] showing in subsite −1 a clan-A hydrophobic platform-triad variant with left Trp360 / top Trp379 / right Tyr285. Very intriguing for this family is that it also contains β(1,3)-xylanases; indeed, the double left/top tryptophan also occurs as a conserved clan-A platform-variant with the family 10/clan-A β(1,4)-xylanases. Complexes of the Man26A catalytic acid/base E212A mutant with the mechanism-based substrate 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-mannotrioside revealed at the Michaelis complex a $^1S_5$ skew-boat for the mannopyranoside ring at subsite −1 while the glycosylenzyme intermediate shows a $^OS_2$ skew-boat. This has lead to the suggestion that a $B_{2,5}$ would occur at the transition state [18], not withstanding several unfavourable steric interactions as discussed above.

The family 38 retaining *Drosophila* Golgi α-mannosidase II structures in complex with 1- deoxymannojirimycin (PDB 1hxk) and swainsonine (PDB 1hww) [20] show these inhibitors at their B-site stacking to Trp95 while a nearby Tyr727 is in a position similar to the clan-A 'right' triad residue; their C_2 and C_3 hydroxyl groups coordinate with an essential zinc ion. The strong inhibitor capacity of swainsonine (IC_{50} 20 nM) [20] and its neat superimposition on a $^3H_4$ mannopyranosyl cation structure [19] indicates a hydrolysis mechanism passing through a $^3H_4$ transition state conformation.

Of two members of the family 47 inverting α(1,2)-mannosidases, 3D-structures with 1-deoxymannojirimycin and kifunensine occupying subsite −1 have been published. Both the

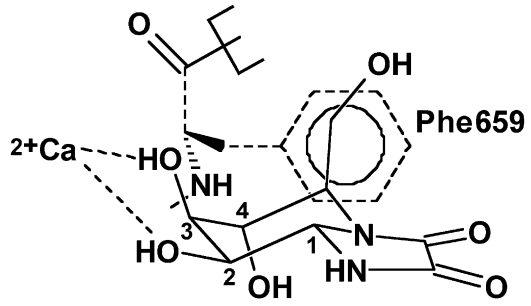

Fig. 6. The family 47 single Phe platform in *H. sapiens* α(1,2)-mannosidase I complexed with kifunensine.

Homo sapiens [21] and the *Penicillium citrinum* [22] enzymes show only one hydrophobic protein residue in the −1 subsite, i.e. a conserved phenyl alanine (respectively Phe659 and Phe468), which is at van der Waals distance with C_4, C_5 and C_6 of the inhibitor; the C_2 and C_3 hydroxyl groups of the inhibitor coordinate with an essential calcium ion. In all cases, the inhibitor was found to reside in a non-standard $^1C_4$ conformation, in which the C_4–C_5–C_6 hydrophobic patch at the B-face is fully exposed to the adjacent phenyl alanine platform. This C_3–C_4–C_5 'all-axial' substituent positioning also suggests the hydrolysis mechanism to pass through a geometrically related $^3H_4$ or ^3E transition state conformation. The situation with the *H. sapiens* α(1,2)-mannosidase I in complex with kifunensine (PDB 1fo3) is shown in Fig. 6.

3.6. Clan-E

Acetylneuraminyl hydrolases (exo-α-sialidases, E.C. 3.2.1.18) are exclusively found in clan-E which contains the families 33, 34 and 83. A C_6-trihydroxypropyl group is present in their substrate, and the sialidases expectedly exhibit a complementary polar and hydrogen bonding network at its surroundings within subsite −1. However, the sialic acid moiety also has a ring-C_3-methylene function (this position corresponds to C_2 for pentapyranosides), and, going from the ground state $^2C_5$ chair conformation (corresponding to $^1C_4$ for pentapyranosides) to the transition state, this group undergoes about the same reorientation as the ring-C_5-methylene of a xylopyranoside substrate. It comes then as almost no surprise that a non-aromatic hydrophobic protein residue is juxtaposed to the ring-C_3-methylene group within enzymes

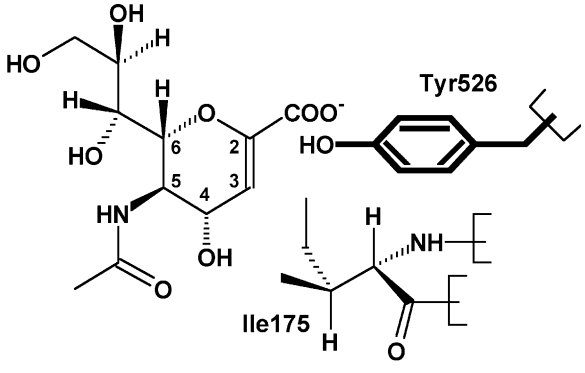

Fig. 7. A Clan-E Ile/Tyr protein residue platform-diad in juxtaposition to C_3 of the ligand. This situation is the family 83 Newcastle disease virus' hemagglutinin neuraminidase in complex with 2,3-dehydro-2-deoxy-*N*-acetyl neuraminic acid.

of clan-E (cfr. family 11 xylanases and their substrate's ring-C_5-methylene group). Within the families 33 and 83, this is a highly conserved isoleucine; the three families also possess in its vicinity a highly conserved tyrosine. A typical situation present in clan-E is the family 83 Newcastle disease virus' hemagglutinin neuraminidase in complex with the inhibitor 2,3-dehydro-2-deoxy-*N*-acetyl neuraminic acid (PDB 1e8v) [12], as shown in Fig. 7.

An interesting variant is the family 34 influenza A virus (human strain, Tokyo/3/67) sialidase (Table 1) in complex with sialic acid (PDB 2bat) [13]. Here, at the expected position for isoleucine, a glutamate is found which is highly conserved within this family. This residue has its α-methylene group close to the ligand's ring-C_3-methylene function while its carboxyl group is pointing away.

4. General conclusion

As evidenced in the present comparative study, peculiarities observed in the active site architecture at subsite −1 of glycoside hydrolases could be exploited for the design of potent inhibitors (including compounds with medical importance) and/or substrates of almost any of these enzymes.

In particular, the recognition of a hydrophobic platform could lead to the presumption that carbohydrate analogues lacking the C_5-hydroxymethylene function and binding at subsite −1 will behave as excellent inhibitors for many D-hexapyranosyl glycoside hydrolases, except where a polar interaction is expected (as with the metal-ion network of family 2, or with the 'top' lysine of family 14 enzymes). Two examples from our own experience are illustrative: (i) family 6 cellobiohydrolases recognise β-D-glucopyranosyl-β(1→4)-D-xylopyranosides as excellent inhibitors [30,44]; and (ii) the recently synthesised 1α-methyl α(1→2)-D-lyxopyranosyl-D-mannobioside is as yet the only known substrate-analogous inhibitor spanning the essential subsites −1/+1 of family 38 and 47 α(1,2)-mannosidases, while the corresponding mannobiose derivative is a substrate [45].

The strategical position of this hydrophobic platform augurs well for the development of pyranoside-substrate analogues where methyl or ethyl groups substitute for the C_5-hydroxymethylene group, which should readily be recognised in subsite −1 by many of these enzymes (exceptions as above). Such compounds are expected to act as good substrates, and this is corroborated by recent literature reports where similar relative preferences for β-D-glucosides, β-D-galactosides and β-D-fucosides have been observed for family 1 β-glycosidases from different organisms; e.g. comparison of the kinetic constants of the *Streptomyces* sp. glucosidase for *p*-nitrophenyl β-D-glucoside (K_m 0.15 mM; k_{cat} 28.4 s^{-1}), *p*-nitrophenyl β-D-galactoside (4-OH axial; K_m 7.3 mM; k_{cat} 118 s^{-1}) and *p*-nitrophenyl β-D-fucoside (4-OH axial and 6-deoxy; K_m 0.14 mM, k_{cat} 37.1 s^{-1}) evidences the compensation of the unfavourable interaction at C_4 (D-galactose) through the increase of hydrophobicity in the C_6-deoxy derivative (D-fucose) [46]. Derivatives of quinovosides (5-deoxy D-glucopyranosides) should therefore be readily recognised as substrates by many glucosidases except where the enzyme's structural features are contra-indicating. The preparation of these potential substrates is underway.

Acknowledgements: Financial support by the Ghent University Research Board (BOF-GOA 12051199 and BOF 01107201) and by FWO-Flanders (3G024997) is acknowledged.

References

[1] Coutinho, P.M. and Henrissat, B. (2002) Carbohydrate-Active enZYmes server at URL: http://afmb.cnrs-mrs.fr/cazy/CAZY/index.html.
[2] Henrissat, B. and Davies, G. (1997) Curr. Opin. Struct. Biol. 7, 637–644.
[3] Henrissat, B. and Bairoch, A. (1996) Biochem. J. 316, 695–696.
[4] Zechel, D. and Withers, S.G. (1999) Mechanisms of glycosyl transfer. In: Comprehensive Natural Products Chemistry (Barton, D., Nakanishi, K. and Poulter, C.D., Eds.), Vol. 5, Chapter 12, 279–314. Elsevier, Amsterdam.
[5] Koshland, D.E. (1953) Biol. Rev. Camb. Philos. Soc. 28, 416–436.
[6] Heightman, T.D. and Vasella, A.T. (1999) Angew. Chem. Ed. Int. Engl. 38, 751–770.
[7] Deslongchamps, P. (1993) Pure Appl. Chem. 65, 1161–1178.
[8] Vasella, A., Davies, G.J. and Böhm, M. (2002) Curr. Opin. Chem. Biol. 6, 619–629.
[9] Kurzinsky, M. (1993) FEBS Lett. 328, 221–224.
[10] Pauling, L. (1946) Chem. Eng. News 24, 1375–1376.
[11] Vyas, N.K. (1991) Curr. Opin. Struct. Biol. 1, 732–740.
[12] Crennell, S., Takimoto, T., Portner, A. and Taylor, G. (2000) Nat. Struct. Biol. 7, 1068–1074.
[13] Varghese, J.N., Mc Kimm-Breschkin, J., Caldwell, J.B., Kortt, A.A. and Colman, P.M. (1992) Proteins Struct. Funct. 14, 327–332.
[14] Crennell, S.J., Garman, E.F., Philippon, C., Vasella, A., Laver, W.G., Vimr, E.R. and Taylor, G.L. (1996) J. Mol. Biol. 259, 264–280.
[15] Sidhu, G., Withers, S.G., Nguyen, N.T., McIntosh, L.P., Ziser, L. and Brayer, G.D. (1999) Biochemistry 38, 5346–5354.
[16] Sabini, E., Wilson, K.S., Danielsen, S., Shülein, M. and Davies, G. (2001) Acta Crystallogr. Sect. D. Biol. Cryst. D57, 1344–1347.
[17] Ruohonen, L., Koivula, A., Wohlfahrt, G., Becker, D., Sinnott, M., Reinikainen, T., Claeyssens, M., Jones, A. and Teeri, T.T. (2002) J. Am. Chem. Soc. 124, 10015–10024.
[18] Ducros, V.M.-A., Zechel, D.L., Murshudov, G.N., Gilbert, H.J., Szabó, L., Stoll, D., Withers, S.G. and Davies, G.J. (2002) Angew. Chem. Int. Ed. Engl. 41, 2824–2827.
[19] Winkler, D.A. and Holan, G. (1989) J. Med. Chem. 32, 2084–2089.
[20] van den Elsen, J.M.H., Kuntz, D.A. and Rose, D.R. (2001) EMBO J. 20, 3008–3017.
[21] Valleé, F., Karaveg, K., Herscovics, A., Moremen, K.W. and Howell, P.L. (2000) J. Biol. Chem. 275, 41287–41298.
[22] Lobsanov, Y.D., Valleé, F., Imberty, A., Yoshida, T., Yip, P., Herscovics, A. and Howell, P.L. (2002) J. Biol. Chem. 277, 5620–5630.
[23] Bohm, H.J. (1998) J. Comput.-Aided Mol. Des. 12, 309–323.
[24] Sharp, K.A., Nicholls, A., Fine, R.F. and Honig, B. (1991) Science 252, 106–109.
[25] Sharp, K.A., Nicholls, A., Friedman, R. and Honig, B. (1991) Biochemistry 30, 9686–9697.
[26] Burkhard, P., Taylor, P. and Walkingshaw, M.D. (2000) J. Mol. Biol. 295, 953–962.
[27] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E. (2002) Nucleic Acids Res. 30, 276–280.
[28] Warshel, A. (1991) Computer Modeling of Chemical Reactions in Enzymes and Solutions. Chapter 6, 155–159 and Chapter 9, 209–211. Wiley and Sons, Inc., New York.
[29] Davies, J.G., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A., Schülein, M. and Withers, S.G. (1998) Biochemistry 37, 11707–11713.
[30] Zou, J.-Y., Kleywegt, G.J., Ståhlberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T.T. and Jones, T.A. (1999) Structure 7, 1035–1045.
[31] Sulzenbacher, G., Davies, G.J. and Schülein, M. (1996) Biochemistry 35, 15280–15287.
[32] Dominguez, R., Souchon, H., Lascombe, M.B. and Alzari, P.M. (1996) J. Mol. Biol. 257, 1042–1051.

[33] Burmeister, W.P., Cottaz, S., Driguez, H., Ori, R., Palmieri, S. and Henrissat, B. (1997) Structure 5, 663–676.
[34] Juers, D.H., Heightman, T.D., Vasella, A., McCarter, J.D., Mackenzie, L., Withers, S.G. and Matthews, B.W. (2001) Biochemistry 40, 14781–14794.
[35] Béguin, P. (1990) Annu. Rev. Microbiol. 44, 219–248.
[36] Wang, Q., Tull, D., Meinke, A., Gilkes, N.R., Warren, R.A.J., Aebersold, R. and Withers, S.G. (1993) J. Biol. Chem. 268, 14096–14102.
[37] Lo Leggio, L., Parry, N.J., Van Beeumen, J., Claeyssens, M., Bhat, M.K. and Pickersgill, R. (1997) Acta Crystallogr. D53, 599–604.
[38] Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B. and Dideberg, O. (2001) Structure 9, 513–525.
[39] White, A., Tull, D., Johns, K., Withers, S.G. and Rose, D.R. (1996) Nat. Struct. Biol. 3, 149–154.
[40] Sulzenbacher, G., Mackenzie, L.F., Wilson, K.S., Withers, S.G., Dupont, C. and Davies, G.J. (1999) Biochemistry 38, 4826–4833.
[41] Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R. and Utsumi, S. (1999) Biochemistry 38, 7050–7061.
[42] Mikami, B., Degano, M., Hehre, E.J. and Sacchettini, J.C. (1994) Biochemistry 33, 7779–7787.
[43] Hogg, D., Woo, E.-J., Bolam, D.N., McKie, V.A., Gilbert, H.J. and Pickersgill, R.W. (2001) J. Biol. Chem. 276, 31186–31192.
[44] Barr, B.K., Wolfgang, D.E., Piens, K., Claeyssens, M. and Wilson, D.B. (1998) Biochemistry 37, 9220.
[45] Desmet, T., Nerinckx, W., Stals, I., Callewaert, N., Contreras, R. and Claeyssens, M. (2002) Anal. Biochem. 307, 361–367.
[46] Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Austin, C., Querol, E., Planas, A. and Pérez-Pons, J.A. (2001) Biochemistry 40, 5975–5982.