PROTEINS: Structure, Function, and Bioinformatics 54:424–432 (2004)

Three Acidic Residues Are at the Active Site of a β-Propeller Architecture in Glycoside Hydrolase Families 32, 43, 62, and 68

Tirso Pons,[1]* Daniil G. Naumoff,[2] Carlos Martínez-Fleites,[1] and Lázaro Hernández[1]
[1]*Centro de Ingeniería Genética y Biotecnología (CIGB), Habana, Cuba*
[2]*State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia*

ABSTRACT **Multiple-sequence alignment of glycoside hydrolase (GH) families 32, 43, 62, and 68 revealed three conserved blocks, each containing an acidic residue at an equivalent position in all the enzymes. A detailed analysis of the site-directed mutations so far performed on invertases (GH32), arabinanases (GH43), and bacterial fructosyltransferases (GH68) indicated a direct implication of the conserved residues Asp/Glu (block I), Asp (block II), and Glu (block III) in substrate binding and hydrolysis. These residues are close in space in the 5-bladed β-propeller fold determined for *Cellvibrio japonicus* α-L-arabinanase Arb43A [Nurizzo et al., Nat Struct Biol 2002;9:665–668] and *Bacillus subtilis* endo-1,5-α-L-arabinanase. A sequence–structure compatibility search using 3D-PSSM, mGenTHREADER, INBGU, and SAM-T02 programs predicted indistinctly the 5-bladed β-propeller fold of Arb43A and the 6-bladed β-propeller fold of sialidase/neuraminidase (GH33, GH34, and GH83) as the most reliable topologies for GH families 32, 62, and 68. We conclude that the identified acidic residues are located at the active site of a β-propeller architecture in GH32, GH43, GH62, and GH68, operating with a canonical reaction mechanism of either inversion (GH43 and likely GH62) or retention (GH32 and GH68) of the anomeric configuration. Also, we propose that the β-propeller architecture accommodates distinct binding sites for the acceptor saccharide in glycosyl transfer reaction. Proteins 2004;54:424–432.**

Key words: catalytic mechanism; functional residues; fructofuranosidases; clan GH-E; clan GH-F; clan GH-J; fold recognition, glycosidases

INTRODUCTION

Glycoside hydrolases (GHs) have been classified into 87 families according to the similarity of their amino acid sequences, which imply both structural and mechanistic relationships.[1,2] Some GH families are grouped at a higher hierarchical level described as clans.[2] Family GH32 comprises sucrose-6-phosphate hydrolases, invertases, inulinases, levanases, eukaryotic fructosyltransferases, and bacterial fructanotransferases; family GH43 includes β-xylosidases, β-xylanases, α-L-arabinases, and α-L-arabinofuranosidases; family GH62 groups some α-L-arabinofuranosidases; and family GH68 includes the majority of bacterial fructosyltransferases and two invertases. Thus, each of these families includes enzymes with a glycofuranosidase activity. Families GH32 and GH68 are combined into the clan GH-J, whereas GH43 and GH62 compose the clan GH-F[3–5] (for updates see the CAZy server at http://afmb.cnrs-mrs.fr/cazy/families.html). On the basis of sequence similarity, these four families compose the furanosidase superfamily,[5] which also includes a family of enzymatically uncharacterized proteins, known as GHLP[5] or COG2152 (http://www.ncbi.nlm.nih.gov/cog/). For the discussion that follows, we refer to the GH families 32, 43, 62, and 68. It should be stressed that not all furanosidases belong to these families. For example, inulin fructotransferase III [IFTaseIII, Enzyme Commission (EC) 2.4.1.93) of *Arthrobacter globiformis* is included into GH91, and its crystal structure revealed a right-handed β-helix fold.[6] Furthermore, α-L-arabinofuranosidase from *Geobacillus stearothermophilus* (AbfA T-6, EC 3.2.1.55) belongs to GH51 and has a catalytic domain with a $(\beta/\alpha)_8$-barrel fold.[7]

Enzymatic hydrolysis of the glycosidic bond occurs via a general acid catalysis that requires at least two critical residues, a proton donor and a nucleophilic base, and involves two major mechanisms giving rise to either an overall retention or inversion of anomeric configuration.[8–10] The proton donor is positioned within hydrogen-bond distance of the glycosidic oxygen in both the retaining (e.g., GH32 and GH68) and the inverting (GH43) enzymes, but the nucleophilic base is more distant than the anomeric carbon in the inverting enzymes that need to accommodate a water molecule between the nucleophilic residue and the sugar. This difference results in an average distance between the two catalytic residues of ~4.5 to 5.5 Å in retaining enzymes as opposed to ~9.0 to 9.5 Å in inverting enzymes.[9,10]

Site-directed mutations on yeast and plant invertases, two bacterial α-L-arabinanases, and various bacterial fruc-

Grant sponsor: International Foundation for Science; Grant number: E/3108-1 (to T. Pons).

*Correspondence to: Tirso Pons Hernández, Centro de Ingeniería Genética y Biotecnología (CIGB), P.O. Box 6162, Habana 10600, Cuba. E-mail: tirso.pons@cigb.edu.cu

Received 4 June 2003; Accepted 6 August 2003

tosyltransferases have revealed a direct implication of Asp and Glu residues in substrate hydrolysis.[11–20] The representative GH43 crystal structures of *Cellvibrio japonicus* α-L-arabinanase Arb43A and *Bacillus subtilis* endo-1,5-α-L-arabinanase show a 5-bladed β-propeller fold with 3 invariant carboxylates close in space at the catalytic center.[14,15] The 6-bladed β-propeller fold of sialidases and neuraminidases is the other example of a β-propeller architecture representative of GH families 33, 34, and 83 (clan GH-E).[21] Crystal structures at high resolution have not been elucidated for members of the GH families 32, 62, and 68. Previous work on the development of methods for the detection of catalytic residues using homology were proposed by Casari et al.,[22] Fetrow and Skolnick,[23] and Aloy et al.[24]

In this article, we have aligned 160 protein sequences of GH families 32, 43, 62, and 68 to identify conserved blocks containing acidic residues. The integrative analysis of our sequence alignments, the sequence–structure compatibility search, and the previous site-directed mutagenesis studies allow us to propose the existence of three key acidic residues located at the active site in a β-propeller architecture for GH families 32, 43, 62, and 68. The three strictly conserved acidic residues operate with the canonical reaction mechanism of inversion (GH43) or retention (GH32 and GH68) of anomeric configurations.

MATERIALS AND METHODS

Protein sequences were retrieved from the SwissProt/TREMBL database. The MACAW program version 2.0.5 Win32i[25] and CLUSTALW service (http://www.ebi.ac.uk/clustalw) were used for multiple-sequence alignments. Sequences of the GHLP family were not analyzed, since they are homologous only to the C-terminal part of the other sequences of the furanosidase superfamily.[5] The MACAW program with a maximum of 30 proteins selected from GH32 (8 sequences), GH43 (8 sequences), GH62 (4 sequences), and GH68 (10 sequences) was used to identify the conserved regions. Block I was revealed by aligning the first 150 amino acids of the selected proteins; blocks II and III were identified using the complete amino acid sequences. The probability of obtaining the observed level of similarity by chance—*P* value—was 10^{-19} or lower (search space $N = 5.457e + 055$ for block I, and $N = 1.539e + 080$ for blocks II and III). Blocks I, II, and III of the total 160 sequences were extracted from CLUSTALW alignments for each GH sequence. The fold-recognition methods 3D-PSSM (http://www.sbg.bio.ic.ac.uk/servers/3dpssm/), INBGU (http://www.cs.bgu.ac.il/~bioinbgu/query.html), mGenTHREADER (http://bioinf.cs.ucl.ac.uk/psiform.html), and SAM-T02 (http://www.soe.ucsc.edu/research/compbio/hmm-apps/t02-query.html) were used to study structural compatibility between β-propeller architectures and the amino acid sequences of GH families 32, 62, and 68. The structural comparison method CE (http://cl.sdsc.edu) was used for superposition of β-propeller crystal structures. The calculation of the distance between the catalytic residues in the crystal structure of Arb43A was done using the WHAT IF program.[26] The complete list of fold-recognition results, structural comparisons, and other additional information is available at http://bio.cigb.edu.cu/~pons/.

RESULTS AND DISCUSSION

Three Key Acidic Residues Are Conserved in GH Families 32, 43, 62, and 68

The comparison of 160 protein sequences of GH families 32, 43, 62, and 68 using MACAW and CLUSTALW revealed three equivalent blocks containing acidic residues in all the enzymes (Fig. 1). Previous site-directed mutagenesis studies have revealed a functional role for these residues in members of the families GH32, GH43, and GH68.[11–20] To our knowledge, the catalytic residues of GH62 enzymes have not been identified.

The conserved residues Asp42 (block I) and Glu223 (block III) (Asp23 and Glu204 in the mature protein) were recognized as the nucleophilic base and the proton donor, respectively, in *Saccharomyces cerevisiae* invertase.[11,12] The same functions were attributed to the corresponding Asp54 (block I) and Glu236 (block III) of invertase from the plant *Chenopodium rubrum*.[13] For this family (GH32), there are no mutagenesis data for the conserved aspartate residue in block II. Concerning family GH43, the separate substitution of the three conserved carboxylates—Asp38 (block I), Asp158 (block II), and Glu221 (block III)—of *C. japonicus* α-L-arabinanase Arb43A inactivated the enzyme or considerably reduced its activity. Asp38 lies 6 Å from the anomeric carbon of the substrate and may function as the nucleophilic base. Glu221 is adjacent to the glycosidic oxygen atom in the site of bond cleavage, and it must be the proton donor. Asp158 is important in catalysis presumably by modulating pKa and orienting the proton donor.[14] Similar results were obtained for *B. subtilis* endo-1,5-α-L-arabinanase. In this case, mutation of Asp44 (block I), Asp163 (block II), and Glu215 (block III) (Asp14, Asp133, and Glu185 in the mature protein) to alanine abolished the enzyme activity, showing that they form the basis of the catalytic center.[15]

In bacterial fructosyltransferases (GH68), the three conserved carboxylates have been independently mutated.[16–19] Site-directed mutagenesis of Asp135 (block I) in *Gluconacetobacter diazotrophicus* levansucrase resulted in an inactive enzyme (Hernández, unpublished data). Similarly, the substitution Asp → Asn in block I of the *Lactobacillus reuteri* fructosyltransferases (Asp272 in inulosucrase and Asp249 in levansucrase) nearly abolished the enzyme activity.[19] The mutation of the equivalent Asp residues in block II of *G. diazotrophicus* levansucrase,[16] *Streptococcus salivarius* fructosyltransferase,[17] *Zymomonas mobilis* levansucrase,[18] and *L. reuteri* inulosucrase and levansucrase[19] either abolished or reduced dramatically sucrose hydrolysis. The substitution of Glu278 (block III) by His in *Z. mobilis* levansucrase reduced 210-fold the k_{cat} for sucrose hydrolysis.[18] *L. reuteri* inulosucrase and levansucrase became almost inactive when Glu in block III was changed to Gln.[19] Visual inspection of the active site in the crystal structure of Arb43A,[14] using the WHAT IF program, shows that blocks I, II, and III are

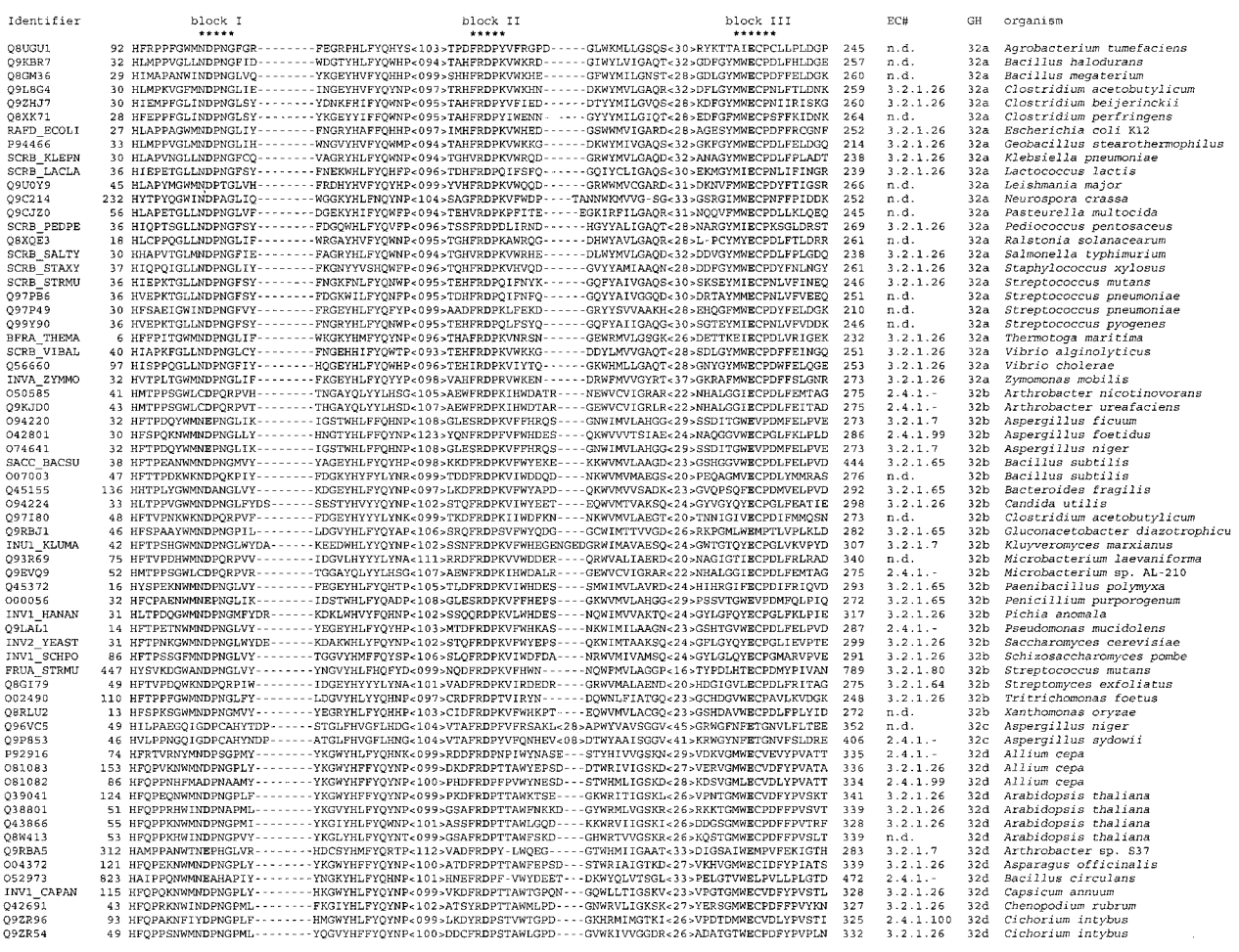

Fig. 1. Multiple-sequence alignment of proteins from GH32, GH43, GH62, and GH68 families. The distance from the protein termini and length of variable spacers are indicated. The sequence identifiers are according to the nomenclature in the SwissProt/TREMBL database. At the top, three conserved blocks (I, II, and III) are indicated by asterisks. EC#, Enzyme Commission number; n.d., not determined. Fructosyltransferases, fructan:fructan 6G-fructosyltransferase, or sucrose fructan 6-fructosyltransferase (2.4.1.-); levansucrase (2.4.1.10); sucrose:sucrose 1-fructosyltransferase (2.4.1.99); fructan:fructan 1-fructosyltransferase (2.4.1.100); inulinase (3.2.1.7); xylanase (3.2.1.8); invertase or sucrose 6-phosphate hydrolase (3.2.1.26); β-xylosidase (3.2.1.37); α-L-arabinofuranosidase (3.2.1.55); levanase (3.2.1.65); fructan β-fructosidase (3.2.1.80); endo-arabinase (3.2.1.99). *These enzymes have the α-L-arabinofuranosidase (EC 3.2.1.55) activity as well. GH: family (GH32, GH43, GH62, and GH68 according to http://afmb.cnrs-mrs.fr/CAZY/families.html) and subfamily (a/b/c/d/e/f/g, according to Naumoff[5] and Naumoff and Livshits[34]) of the sequences.

close in space. Also, the nucleophilic base residues Tyr406+Glu277 in neuraminidase of subtype N9 from an avian influenza virus are in close vicinity to Asp38 (block I) [Fig. 2(A and B)].

Five- and Six-Bladed β-Propeller Topologies Predicted for GH32, GH62, and GH68

We previously predicted a 6-bladed β-propeller catalytic domain for fructofuranosidases included in GH32 and GH68.[28,29] In this work, sequence–structure compatibility searches using 3D-PSSM, INBGU, mGenTHREADER, and SAM-T02 programs showed the 5-bladed β-propeller fold of Arb43A and the 6-bladed β-propeller fold of sialidase/neuraminidase (clan GH-E) as the most reliable topologies for GH families 32, 62, and 68 (Table I). The choice of divergent amino acid sequences [*S. cerevisiae* invertase (GH32), *G. diazotrophicus* levanase (GH32), *C. japonicus* α-L-arabinofuranosidase C (GH62), and *G. diazotrophicus* levansucrase (GH68)] and four prediction methods using different approaches provide confidence in our results. 3D-PSSM, mGenTHREADER, INBGU, and SAM-T02 identified several β-propeller structures with 4-, 5-, 6-, 7-, and 8-bladed topologies among the first hits for each method (Table I). The 5- and 6-bladed topologies of Arb43A and sialidase/neuraminidase are identified with highest frequency and their superposition revealed their high structural similarity (Fig. 2).

A general conclusion can be drawn from inspection of the recently determined three-dimensional (3D) structures, and the sequence–structure compatibility searches

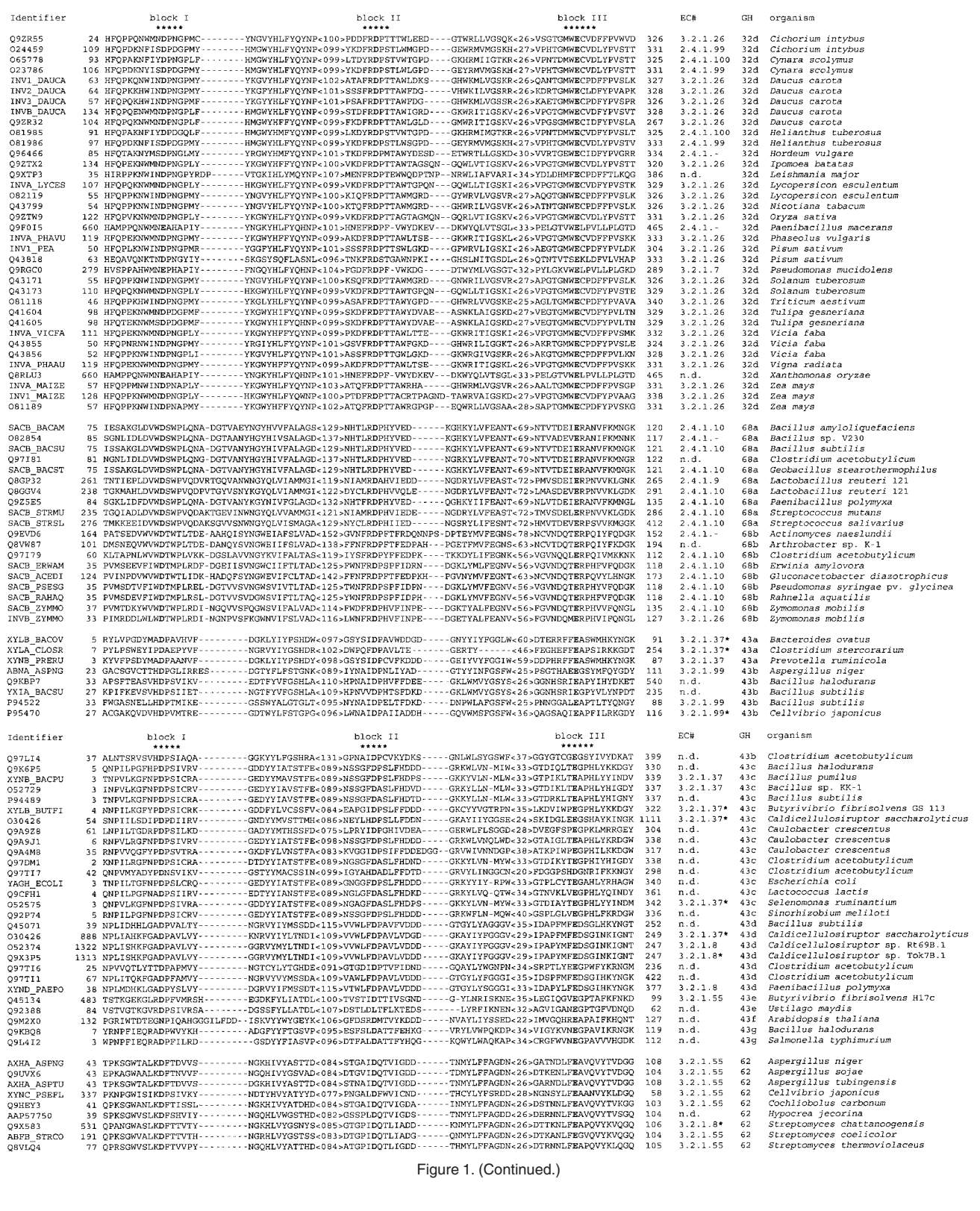

Figure 1. (Continued.)

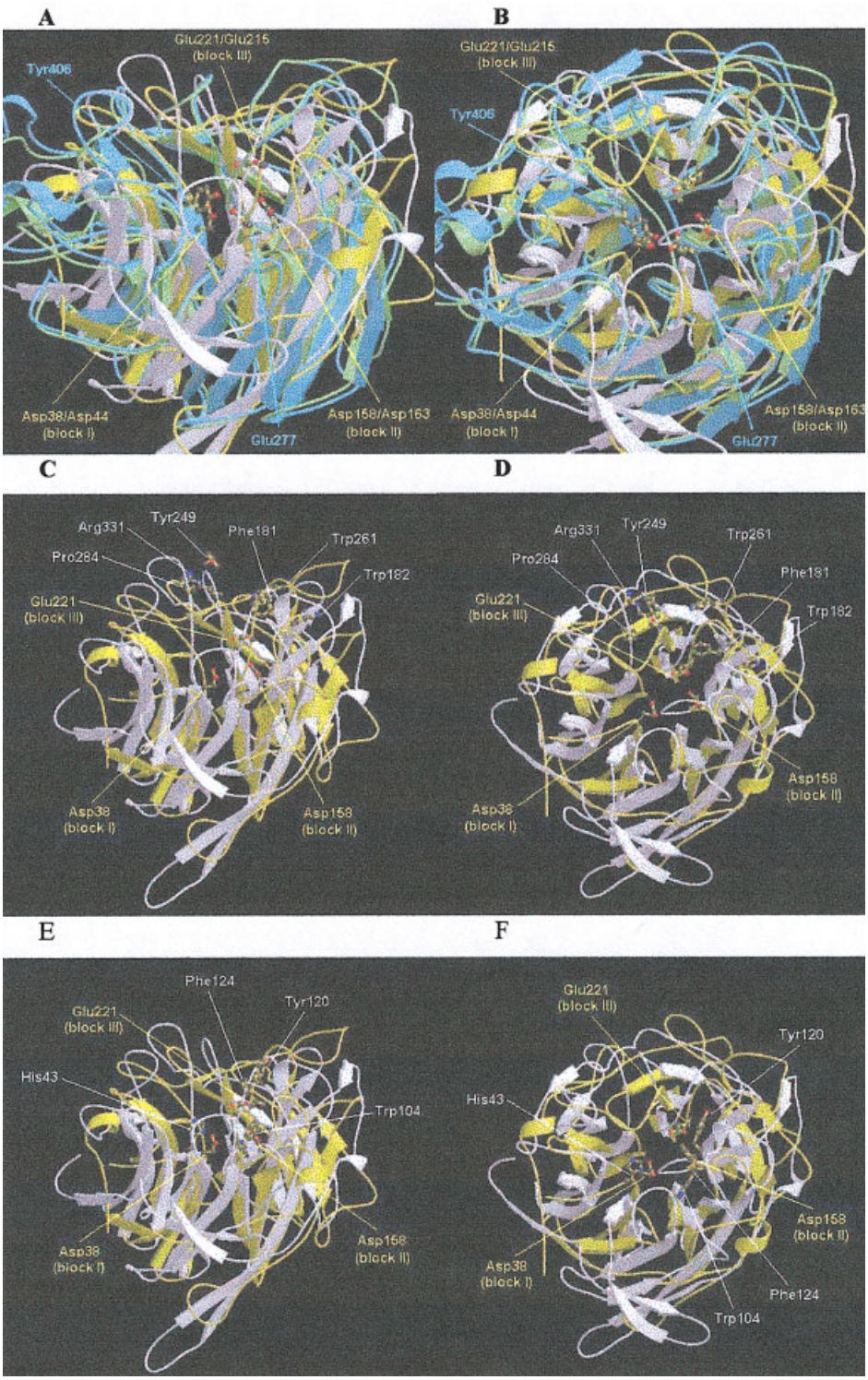

Fig. 2. Structural comparison of *C. japonicus* α-L-arabinanase Arb43A (PDB code: 1gyd) in yellow, influenza B virus neuraminidase (PDB code: 1nscA) in green, neuraminidase of subtype N9 from an avian influenza virus (PDB code: 5nn9) in cyan, and *Trypanosoma rangeli* trans-sialidase (PDB code: 1mz5) in white. Side (**A**) and top (**B**) view represent the catalytic acidic residues Asp38/Asp44 (block I), Asp158/Asp163 (block II), Glu221/Glu215 (block III) in *C. japonicus* and *B. subtilis* arabinanases, and the nucleophilic base Glu277+Tyr406 in neuraminidase N9. Side (**C**) and top (**D**) view represent the functional residues far from the active site: Arg331 from *B. subtilis* levansucrase; Trp261 from *Z. mobilis* levansucrase; Tyr249 and Pro284 from *T. cruzi* trans-sialidase; and Phe181 and Trp182 from *B. subtilis* endo-1,5-α-L-arabinanase; while side (**E**) and top (**F**) view represent the functional residues close to the active site: Tyr120 from *T. cruzi* trans-sialidase; His43, Trp104, and Phe124 from *B. subtilis* endo-1,5-α-L-arabinanase. The figure was prepared using the MOLSCRIPT program of Per Kraulis.[27] Structural similarities observed in the β-propeller topologies and the different functions hosts by this architecture are reviewed.[35]

TABLE I. Fold Recognition Results for GH32, GH62, and GH68 Families Using 3D-PSSM, INBGU, SAM-T02, and mGenTHREADER

	3D-PSSM		mGenTHREADER		INBGU		SAM-T02	
Enzyme	PDB	Score	PDB	Score	PDB	Score	PDB	Score
G. diazotrophicus	1eur	0.845	1gydB	0.602	1gyhA	51.3	1gyhA	0.011
levanase Q9RBJ1	1nscA	1.110	3sil	0.601	1cwvA	8.4	1cruA	0.136
(GH32)	1gyhA	1.410	*1l0qA*	0.595	*1gof*	6.3	1qbiA	0.137
	1erjA	1.470	1c9uA	0.591	*1flgA*	3.7	1c9uA	0.299
	1eut	1.600	*1qksA*	0.583	*1l0qA*	3.6	1dim	0.308
	2sli	1.650	*1gof*	0.541	1eut	2.9	2sil	0.445
	3sil	1.660	1e07A	0.531	1eur	2.4	3sil	0.475
	1gxrA	2.960	2er7E	0.527	*1jjuB*	2.2	1kit	0.479
	1eeeA	3.340	*1a12A*	0.520	*1g72A*	2.2	*1h2xA*	0.618
	1k8kC	3.380	1k0sA	0.517	1c9uA	2.2	1dfcA	0.663
G. diazotrophicus	1gyhA	0.555	1by5A	0.564	1gyhA	95.5	1gyhA	0.0001
levansucrase	1nscA	0.729	1bf2	0.540	*1gof*	8.0	1gydB	0.046
SACB_ACEDI	1dlpA	1.160	*1gof*	0.534	1c9uA	8.0	1gyeB	0.187
(GH68)	*1gof*	1.420	1sat	0.526	1cwvA	5.0	*1al2A*	0.770
	2sli	1.770	2mprA	0.516	1ihmA	4.0	*1qhuA*	0.119
	1dnv	1.900	1fepA	0.514	1k32A	3.3	*1pex*	1.352
	1eur	2.540	1a0tP	0.508	*1gof*	2.8	117jA	1.647
	1qjvA	2.880	1c9uA	0.503	113wA	2.7	1nltA	1.661
	1pooA	3.000	1air	0.499	1gh7A	2.5	1mz5A	1.865
	1bwmA	3.030	1uok	0.490	*1jx5A*	2.2	1axiB	1.983
S. cerevisiae invertase	1gyhA	0.107	1gydB	0.709	1gyhA	37.5	1gyhA	0.156
INV2_YEAST	3sil	0.220	1kit	0.602	*1gof*	9.0	1eur	0.508
(GH32)	1eut	0.227	1by5A	0.562	1cwvA	6.4	1dim	0.563
	1bihA	0.608	1kmoA	0.558	*1jx5A*	4.2	3sil	0.866
	1gog	0.612	*1erjA*	0.526	1dfcA	2.8	2sil	0.875
	1eur	0.633	2sli	0.517	1e1aA	2.8	1eut	0.913
	1cruA	2.530	*1jofA*	0.506	*1k8kC*	2.7	1euu	1.017
	1gof	3.310	*1bpoA*	0.500	1eur	2.5	1dfcA	1.109
	1fnf	3.720	1crzA	0.500	*110qA*	2.2	1htyA	1.262
	2sli	4.670	*1gof*	0.495	1mz5A	2.1	1hxkA	1.331
C. japonicus	1tl2A	0.637	1gydB	0.604	1gyhA	28.9	1gyhA	0.517
arabinofuranosidase	1ilmB	2.030	1tl2A	0.527	1tl2A	13.9	1n1tA	0.930
XYNC_PSEFL	1a65A	2.320	*1qfmA*	0.519	1lpxA	5.4	1mz5A	0.997
(GH62)	1aozA	3.130	1m7xA	0.518	1hwmB	3.5	1l7jA	1.958
	1l7lA	3.450	1bihA	0.506	1dlpA	2.5	*1gen*	2.167
	1hwpB	3.760	*1gof*	0.504	1dlpA	2.5	*1gen*	2.380
	1bxfA	5.180	1jbjA	0.497	1efaA	1.9	1mmuA	2.437
	1bp3B	5.260	*1flgA*	0.493	*1gof*	1.5	1nszA	2.647
	1kacA	5.880	1qinA	0.486	1hcxA	1.4	1tl2A	2.943
	1qclA	5.970	1by5A	0.480	1lrxB	1.2	1ebpA	3.214

Note: The β-propeller topologies are represented by 4-bladed haemopexin (1qhuA), gelatinase A (1gen), and collagenase-3 (1pex); 5-bladed tachylectin-2 (1tl2A) and Arb43A (1gydB, 1gyhA); 6-bladed tricorn protease (1k32A), bacterial sialidase (2sil, 3sil, 1kit, 1eur, 1eut, 1euu, 1dim), influenza B virus neuraminidase (1nscA), leech (2sli) and trypanosomal (1mz5A, 1n1tA) trans-sialidases, diisopropyl fluoro phosphatase (1elaA), phytase (1pooA), TolB protein (1crzA), LRP1 protein (1lpxA), human low-density lipoprotein receptor (1lrxB), and glucose dehydrogenase (1cruA, 1c9uA, 1qbiA); 7-bladed Tup1 protein (1erjA), galactose oxidase (1gof, 1gog), RCC1 protein (1a12A), clathrin heavy chain (1bpoA), prolyl oligopeptidase (1qfmA, 1h2xA), groucho/TLE1 (1gxrA), Arp2/3 complex (1k8kC), quinohemoprotein amine dehydrogenase (1jjuB), protein-binding YVTN (110qA), integrin α-IIB (1jx5), and muconate lactonizing enzyme (1jofA); 8-bladed quinoprotein alcohol dehydrogenase (1eeeA), ethanol dehydrogenase (1flgA), methanol dehydrogenase (1g72A), and cytochrome *cd1* (1qksA). Bold and underlined letters represent the 5- and 6-bladed β-propellers, while italics represent 7- and 8-bladed ones.

presented in this work. The enzymatic function of furanoside hydrolysis has evolved independently on at least three structural frameworks: $(\beta/\alpha)_8$-barrel (GH51), β-propeller (GH43), and right-handed β-helix (GH91) architectures.

Similar Architectures in Retaining and Inverting Enzymes

Calculation of the distances between the oxygen atoms of the carboxyl groups of the nucleophilic base Asp38 (block I) and the proton donor Glu221 (block III) in the

TABLE II. Three Key Acidic Residues of Enzymes Acting on Pyranoside and Furanoside Substrates With a Canonical Reaction Mechanism of Inversion or Retention of the Anomeric Configuration

		Fold	Mechanism	(p.d., n.b., r.p.)	PDB	Ref.
Enzymes acting on pyranoside substrates						
α-glucuronidase A (GlcA67A)	GH67	$(\beta/\alpha)_8$-barrel	Inverting	Glu292, Asp365, Glu393	1GQI	[30]
α-amylase (TAKA-amy)	GH13	$(\beta/\alpha)_8$-barrel	Retaining	Glu230, Asp206, Asp297	2TAA	[31]
trypanosomal sialidases	GH33	β-propeller	Retaining	—, Tyr343+Glu231, —	1MZ5	[32]
Enzymes acting on furanoside substrates						
α-L-arabinanase 43A (Arb43A)	GH43	β-propeller	Inverting	Glu221, Asp38, Asp158	1GYD	[14]
levansucrase (LsdA)	GH68	β-propeller*	Retaining	Glu401, Asp135, Asp309	—	[29]
invertase (Suc2)	GH32	β-propeller*	Retaining	Glu204, Asp23, Asp152	—	[28]
α-L-arabinofuranosidase (AbfA T-6)	GH51	$(\beta/\alpha)_8$-barrel	Retaining	Glu175, Glu294, —	—	[7]

Note: *Predicted architecture based on 6-bladed β-propeller. p.d., proton donor; n.b., nucleophilic base; r.p., residue modulating pKa or the orientation of the proton donor; —, functional residue not identified.

TABLE III. Different Combinations of Nucleophilic Base and Proton Donor for GH Clans With $(\beta/\alpha)_8$ Barrel and β-Propeller Architectures

Clan	Fold	GH families	Nucleophile/base	Proton donor
GH-A	$(\beta/\alpha)_8$-barrel	GH1,GH2,GH5,GH10,GH17,GH26 GH30,GH35,GH39,GH42,GH51 GH53,GH59,GH72,GH79,GH86	Glu	Glu
GH-D	$(\beta/\alpha)_8$-barrel	GH27,GH36	Asp	Asp
GH-H	$(\beta/\alpha)_8$-barrel	GH13,GH70,GH77	Asp	Glu
GH-K	$(\beta/\alpha)_8$-barrel	GH18,GH20	C-2	Glu
GH-E	β-propeller	GH33,GH34,GH83	Tyr+Glu	Not known

Note: All GH clans contain families that retain the anomeric configuration at the cleavage point. C-2 is carbonyl oxygen of C-2 acetamido group of substrate.

crystal structure of Arb43A[14] gives 5.90 Å between oδ1 Asp38 and oϵ2 Glu221, and 8.26 Å between oδ2 Asp38 and oϵ1 Glu221. Both values are in agreement with the structural requirements for inverting (GH43) and retaining (GH32, GH68) enzymes.[9,10] The mechanism of substrate hydrolysis for GH62 is still unknown. We propose the inverting mechanism of hydrolysis for GH62 on the basis of substrate similarity with GH43 family (α-L-arabinofuranosidase activity).

Crystal structures of two *C. japonicus* (*Pseudomonas cellulose*) glycosidases solved by Nurizzo et al.: α-L-arabinanase Arb43A[14] and α-glucuronidase GlcA67A,[30] together with the recently determined 3D structure of *G. stearothermophilus* T-6 α-L-arabinofuranosidase (AbfA T-6) by Hoevel et al.,[7] provide examples of GH families where the substrate, the mechanism of hydrolysis, and even the combination of nucleophilic base and proton donor residues are not conserved respect to GH families or clans, having similar architectures. GH families sharing $(\beta/\alpha)_8$-barrel architectures and three carboxylates at the active site can operate with inverting (GlcA67A, GH67) and retaining (GH-clans A, D, H, and K) hydrolysis of pyranoside and furanoside substrates.[7,30,31] A similar scenario is observed for GH families 32, 33, 43, 62, and 68 acting on pyranoside and furanoside substrates. These enzymes share a β-propeller architecture and operate with inverting (GH43 and GH62) and retaining (GH32, GH33, and GH68) mechanisms of hydrolysis (Tables II and III).

For the enzymes, there is no clear correlation between the Enzyme Commission (EC) classification and the protein architecture. Enzymes with the same EC number may exhibit different folds and vice versa. Indeed, many GH families are polyspecific (containing at least two EC numbers).[2] A small cluster of catalytic residues that can be placed on almost any architecture determines the enzyme function. Figure 2(A and B) show the superposition of the proton donor residue Glu221 in the protein Arb43A and the nucleophilic base Glu277 in neuraminidase N9 [Protein Data Bank (PDB) code: 5nn9). Based on the similar location of the active site in the β-propeller architecture of GH43 and clan GH-E, and the other observations presented in this section (Tables II and III), we suggest that some GH families with similar 3D structures, and acting on pyranoside and furanoside substrates, would operate with a canonical reaction mechanism of inversion/retention of anomeric configuration using three key acidic residues.

Distinct Binding Sites for the Acceptor Saccharide in Glycosyl Transfer Reaction

Some retaining GH catalyzes glycoside transfer reactions in addition to glycosyl hydrolysis, for example, the sialyltransferase reaction of trypanosomal trans-sialidases (GH33) and the fructosyltransferase reaction of bacterial levansucrases (GH68).[1,2] Also, at high concentrations of substrate, the *S. cerevisiae* invertase (GH32) can transfer the fructosyl moiety to primary alcohols such as

methanol and ethanol, and to monosaccharides.[12] According to the sequence–structure compatibility searches presented here, and the published crystal structure, these GH families share a β-propeller architecture.

Previous site-directed mutagenesis studies have revealed a functional role for Tyr120, Tyr249, and Pro284, in the sialyltransferase reaction of *Trypanosoma cruzi* trans-sialidase,[32] Arg331 in the fructosyl transfer reaction of *B. subtilis* levansucrase,[20] and Trp261 in the fructosyl transfer reaction of *Z. mobilis* levansucrase.[33] On the other hand, mutation of His43, Trp104, Phe124, Phe181, and Trp182 (His13, Trp74, Phe94, Phe151, and Trp152 in mature protein) to alanine in the *B. subtilis* endo 1,5-α-L-arabinanase (GH43) affected the efficient binding and hydrolysis of arabinofuranose and long-chain arabinans.[15]

After a structural superposition of the 5-bladed (GH43) and 6-bladed (GH33) β-propellers, the analysis of the location of these aromatic, hydrophobic, and polar residues, revealed that (1) Arg331 and Trp261 are located in different positions far from the active site; (2) Tyr249 and Pro284 are close in space to Arg331; (3) Phe181 and Trp182 are far from the active site in a different position than Tyr249, Trp261, Pro284, and Arg331; (4) His43 is neighbor of the catalytic Asp44 (block I); (5) Trp104 is close in space to the catalytic Glu277 of neuraminidase N9; and (6) Phe124 is in close vicinity to Tyr120, both close to the active site.

Positions far from the active site are involved in sialyl and fructosyl transfer reactions, as well as in the hydrolysis of long-chain glycans, while positions close to the active site are important for the orientation of substrates and the efficient hydrolysis of short glycans [Fig. 2(C–F)]. Thus, based on the structural compatibility between GH68 and the 6-bladed β-propeller of sialidase/neuraminidase, we propose that the β-propeller architecture accommodates distinct binding sites for the acceptor saccharide in glycosyl transfer reaction.

NOTE

The crystal structure of *Bacillus subtilis* levansucrase has been recently published [36]. This structure (PDB codes: 1OYG (apo) and 1PT2 (complex)) reveals a five-bladed β-propeller with three strictly conserved acidic residues at the active site.

ACKNOWLEDGMENTS

We are indebted to Dr. Aurora Pérez-Gramatges for a critical reading of the manuscript.

REFERENCES

1. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 1991;280:309–316.
2. Coutinho PM, Henrissat B. Carbohydrate-active enzymes: an integrated database approach. In: Gilbert HJ, Davies G, Henrissat B, Svensson B, editors. Recent advances in carbohydrate bioengineering. Cambridge, UK: Royal Society of Chemistry; 1999. p 3–12.
3. Naumov DG, Doroshenko VG. β-Fructosidases: a new superfamily of glycosyl hydrolases. Mol Biol 1998;32:761–766.
4. Naumoff DG. Conserved sequence motifs in levansucrases and bifunctional β-xylosidases and α-L-arabinases. FEBS Lett 1999; 448:177–179.
5. Naumoff DG. β-Fructosidase superfamily: homology with some α-L-arabinases and β-D-xylosidases. Proteins 2001;42:66–76.
6. Momma M, Fujimoto Z, Maita N, Haraguchi K, Mizumo H. Crystal structure of *Arthrobacter globiformis* inulin fructotransferase: a right-handed β-helix. Presented at the 5th Carbohydrate Bioengineering Meeting, University Hospital Groningen, Groningen, The Netherlands, April 6–9, 2003.
7. Hoevel K, Shallom D, Niefind K, Belakhov V, Baasov T, Shoham G, Shoham Y, Schomburg D. The first 3D structure of a family 51 α-arabinofuranosidase: crystal structures of Michaelis complexes and the covalent intermediate. Presented at the 5th Carbohydrate Bioengineering Meeting, University Hospital Groningen, Groningen, The Netherlands, April 6–9, 2003.
8. Sinnot ML. Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev 1990;90:1171–1202.
9. McCarter JD, Withers SG. Mechanisms of enzymatic glycoside hydrolysis. Curr Opin Struct Biol 1994;4:885–892.
10. Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure 1995;3:853–859.
11. Reddy VA, Maley F. Identification of an active-site residue in yeast invertase by affinity labeling and site-directed mutagenesis. J Biol Chem 1990;265:10817–10820.
12. Reddy A, Maley F. Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase. J Biol Chem 1996;271:13953–13958.
13. Goetz M, Roitsch T. Identification of amino acids essential for enzymatic activity of plant invertases. J Plant Physiol 2000;157:581–585.
14. Nurizzo D, Turkenburg JP, Charnock SJ, Roberts SM, Dodson EJ, McKie VA, Taylor E, Gilbert HJ, Davies GJ. *Cellvibrio japonicus* α-L-arabinanase 43A has a novel five-blade β-propeller fold. Nat Struct Biol 2002;9:665–668.
15. Proctor MR, Davies GJ, Taylor E, Nurizzo D, Lloyd R, Andrews K, Gilbert HJ. Mutational analysis of the active site of the GH43 endo 1,5-α-L-arabinanase from *Bacillus subtilis*. Presented at the 5th Carbohydrate Bioengineering Meeting, University Hospital Groningen, Groningen, The Netherlands, April 6–9, 2003.
16. Batista FR, Hernández L, Fernández JR, Arrieta J, Menéndez C, Gómez R, Támbara Y, Pons T. Substitution of Asp-309 by Asn in the Arg-Asp-Pro (RDP) motif of *Acetobacter diazotrophicus* levansucrase affects sucrose hydrolysis, but not enzyme specificity. Biochem J 1999;337:503–506.
17. Song DD, Jacques NA. Mutation of aspartic acid residues in the fructosyltransferase of *Streptococcus salivarius* ATCC 25975. Biochem J 1999;344:259–264.
18. Yanase H, Maeda M, Hagiwara E, Yagi H, Taniguchi K, Okamoto K. Identification of functionally important amino acid residues in *Zymomonas mobilis* levansucrase. J Biochem (Tokyo) 2002;132:565–572.
19. Ozimek LK, van Geel-Schutten GH, van der Maarel MJEC. Identification of amino acid residues involved in catalysis in novel fructosyltransferases. Presented at the 5th Carbohydrate Bioengineering Meeting, University Hospital Groningen, Groningen, The Netherlands, April 6–9, 2003.
20. Chambert R, Petit-Glatron MF. Polymerase and hydrolase activities of *Bacillus subtilis* levansucrase can be separately modulated by site-directed mutagenesis. Biochem J 1991;279:35–41.
21. Crennel SJ, Garman EF, Laver WG, Vimr ER, Taylor GL. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. Proc Natl Acad Sci USA 1993;90:9852–9856.
22. Casari G, Sander C, Valencia A. A method to predict functional residues in proteins. Nat Struct Biol 1995;2:171–178.
23. Fetrow JS, Skolnick J. Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T1 ribonucleases. J Mol Biol 1998;281:949–968.
24. Aloy P, Querol E, Aviles FX, Sternberg MJ. Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. J Mol Biol 2001;311:395–408.
25. Schuler GD, Altschul SF, Lipman DJ. A workbench for multiple alignment construction and analysis. Proteins 1991;9:180–190.
26. Vriend G. WHATIF, a molecular modelling and drug design program. J Mol Graph 1990;8:52–56.

27. Kraulis P. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24: 946–950.
28. Pons T, Olmea O, Chinea G, Beldarraín A, Márquez G, Acosta N, Rodríguez L, Valencia A. Structural model for family 32 of glycosyl-hydrolase enzymes. Proteins 1998;33:383–395.
29. Pons T, Hernández L, Batista FR, Chinea G. Prediction of a common β-propeller catalytic domain for fructosyltransferases of different origin and substrate specificity. Protein Sci 2000;9:2285–2291.
30. Nurizzo D, Nagy T, Gilbert HJ, Davies GJ. The structural basis for catalysis and specificity of the *Pseudomonas cellulose* α-glucuronidase, GlcA67A. Structure 2002;10:547–556.
31. Janecek S. α-Amylase family: molecular biology and evolution. Prog Biophys Molec Biol 1997;67:67–97.
32. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, Paris G, Amaya MF, Frasch AC, Alzari PM. Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J 2000;19:16–24.
33. Sangiliyandi D, Gunasekaran P. Polymerase and hydrolase activities of *Zymomonas mobilis* levansucrase separately modulated by in vitro mutagenesis and elevated temperature. Process Biochem 2001;36:543–548.
34. Naumoff DG, Livshits VA. Molecular structure of the *Lactobacillus plantarum* sucrose utilization locus: comparison with *Pediococcus pentosaceus*. Mol Biol 2001;35:15–22.
35. Pons T, Gómez R, Chinea G, Valencia A. Beta-propellers: associated functions and their role in human diseases. Curr Med Chem 2003;10:505–524.
36. Meng G, Fütterer K. Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. Nature Struct Biol 2003;DOI: 10.1038/nsb974.