Conserved sequence motifs in levansucrases and bifunctional β-xylosidases and α-L-arabinases

Daniil G. Naumoff*

State Institute for Genetics and Selection of Industrial Microorganisms, I-Dorozhny proezd, 1, Moscow 113545, Russia

Received 25 February 1999

Abstract Comparison of the amino acid sequences of two families of glycosyl hydrolases reveals that they are related in a region in the central part of the sequences. One of these families (GH family 68) includes levansucrases and the other one (glycosyl hydrolase family 43) includes bifunctional β-xylosidases and α-L-arabinofuranosidases. The similarity of the primary structure of proteins from these families allows us to consider the invariant glutamate residue as a component of their active center. It is shown for the first time that glycosyl hydrolases recognizing different glycofuranoside residues can have a common sequence motif.

Key words: Glycosyl hydrolase; β-Fructosidase superfamily; Levansucrase; β-Xylosidase; Multiple sequence alignment

1. Introduction

Levansucrases are a family of bacterial enzymes with a β-fructofuranosyl fructotransferase activity (EC 2.4.1.10). They can produce different types of D-fructose poly- and oligomers (levans etc.) from sucrose. The production process often takes place on the early steps of the plant colonization by various symbiotic and phytopathogenic bacteria [1,2]. As well as polyglucans, fructose polymers are probably implicated in the caries process in humans [3,4]. In addition to the fructosyl transferase activity, levansucrases can hydrolyze sucrose and various fructose polymers (levans, inulin) into monosaccharides. So they also have β-fructofuranosyl fructohydrolase activities [5]. This allows us to consider levansucrases as a family of glycosyl hydrolases (GHs) (family 68) [6,7]. By sequence similarity, this family also includes an extracellular invertase of *Zymomonas mobilis*, which has no fructosyl transferase activity [5–7].

Recently, we have shown that levansucrases have a significant sequence similarity with enzymes of the GH family 32 (sucrases and fructanases) [5]. Enzymes of the two protein families have the same molecular mechanism of hydrolyzing reaction: a double displacement with an overall retention of the anomeric configuration of the fructosyl residue [6,7]. Due to these facts, the two families of GHs have been grouped into the β-fructosidase superfamily [5]. Comparison of the most conserved sites of sequences of the two families of β-fructosidases allowed us to propose several possible components of the active center.

In this work, we compare sequences of levansucrases and GHs from family 43. Enzymes of the latter family usually have β-xylosidase (EC 3.2.1.37) and α-L-arabinofuranosidase (EC 3.2.1.55) activities [6–9]. These enzymes participate in the degradation of plant cell wall material in the rumina of mammals by anaerobic microorganisms [9].

2. Materials and methods

Protein sequences were retrieved from the current sequence databases. GHs compared in this work are listed in Table 1. The alignment of proteins was generated by using the PSI-BLAST program [10]. The program MACAW [11] was used to estimate the probabilities of the independent appearance of the regions of local similarity in the two protein families.

3. Results and discussion

An alignment by using the PSI-BLAST and MACAW programs shows that GHs from families 43 and 68 are related in a region in the central part of the sequences. In this region, there are two highly conserved blocks (Fig. 1) with a probability of obtaining the observed level of similarity by chance (*P* value) below 10^{-19} (search space $N = 2.7 \times 10^{86}$) for each of them. This low *P* value results in part from the high similarity between the members of each family. However, for the pairwise comparison of AF005383 of *Caldicellulosiruptor saccharolyticus* and Q43998 of *Acetobacter diazotrophicus* in the first block (sequence E-g/r-s/p-h/q-a/v-Y-k/l-i/h-N-G-K-Y-Y-I-t/f-T-I), the *P* value is 1.7×10^{-5} ($N = 289$), showing that the relationship is authentic.

The first conserved block partly overlaps with known conserved regions of both GH family 43 (the second conserved region according to Sakamoto et al. [8]) and levansucrases (conserved region VII according to Arrieta et al. [12]). This block includes a glutamate residue invariant in both families and a conserved cluster of tyrosine residues (Fig. 1). The sequence between these two elements is highly conserved and corresponds to the pattern: E-x-(SPA)-x(1,2)-(MILAVFY)-x(3)-(NDGA)-(DGNTE)-x-(YWT)-(YV)-(LFYI). It should be noted that the corresponding site of GH family 32 is also conserved (region VI according to Bezzate et al. [1]) and includes the invariant glutamate residue. In the case of *Saccharomyces cerevisiae* invertase, this Glu was shown to serve as a proton donor in the hydrolysis of the glycosidic bond [13]. However, only one (invertase of *Tritrichomonas foetus*) out of more than 100 sequences in GH family 32 follows the pattern described above. This is a consequence of missing the tyrosine cluster in enzymes of the family. However, the similarity of structures of the first conserved block and the corresponding site of sucrases and fructanases allows us to suppose that the invariant Glu residue is

*Corresponding author. Fax: (7) (095) 315 05 01.
E-mail: lab2@vnigen.msk.su

PII: S0014-5793(99)00369-5

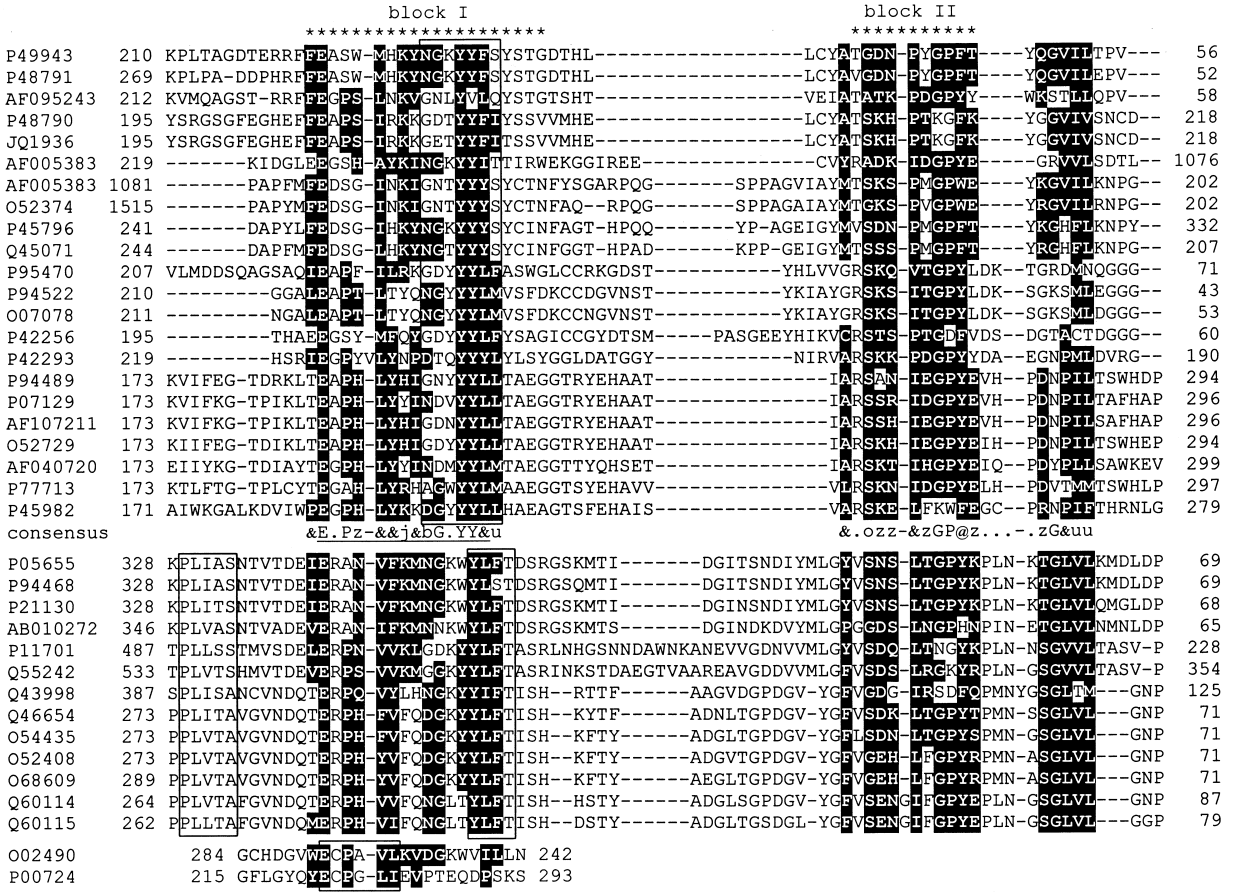

Fig. 1. Multiple alignment of the region of similarity in GH families 43 and 68. The distance from the protein termini is indicated. At the top of the figure, the two conserved blocks described in the paper are indicated by asterisks. The consensus shows residues, which are conserved in the majority of sequences of both families: u indicates a bulky hydrophobic residue (I, L, V, F or M); @ indicates an aromatic residue (F, Y or W); & indicates a hydrophobic residue (I, L, V, F, M, Y, W, A, C or P); j indicates a positively charged residue (K, R or H); b indicates N or D; o indicates a small residue (G, S or A); z indicates a polar residue (D, E, N, Q, H, K, R, S or T). The residues that are conform to the consensus are highlighted. In the consensus, the sequence corresponding to the pattern of GH families 43 and 68 (see text) is underlined. The regions known to be conserved in GH families 43 [8], 68 [12] and 32 [1] are boxed. The sequences are from the current sequence databases with the accession number for each sequence indicated in the most left column (for origin of the sequences see Table 1).

Table 1
Glycosyl hydrolases analyzed in this work

Family number[a]	Enzyme	Organism	Length	GenBank accession number	SWISS-PROT accession number
43	β-Xylosidase and arabinofuranosidase	*Bacteroides ovatus*	325	U04957	P49943
43	β-Xylosidase	*Prevotella ruminicola*	319	Z49241	P48791
43	β-Xylosidase	*Cochliobolus carbonum*	328	AF095243	
43	β-Xylosidase and arabinofuranosidase	*Clostridium stercorarium*	473	D13268	P48790
43	β-Xylosidase and arabinofuranosidase	*Clostridium stercorarium*	473	JQ1936	
43[b]	β-xylosidase and arabinofuranosidase	*Caldicellulosiruptor saccharolyticus*	1347	AF005383	
43/10[c]	β-Xylanase	*Caldicellulosiruptor sp.* Rt69B.1	1779	AF036924	O52374
43	β-Xylanase	*Paenibacillus polymyxa*	635	X57094	P45796
43	β-Xylanase	*Bacillus subtilis*	513	Z99113	Q45071
43	α-L-Arabinase	*Pseudomonas fluorescens*	347	Y10458	P95470
43	α-L-Arabinase	*Bacillus subtilis*	313	Z75208	P94522
43	α-L-Arabinase	*Bacillus subtilis*	324	D85132	O07078
43	α-L-Arabinase	*Aspergillus niger*	321	L23430	P42256
43	ORF	*Bacillus subtilis*	469	D31856	P42293
43	β-Xylosidase	*Bacillus subtilis*	533	U66480	P94489
43	β-Xylosidase	*Bacillus pumilus* IPO	535	X05793	P07129
43	β-Xylosidase	*Bacillus pumilus* PLS	535	AF107211	
43	β-Xylosidase	*Bacillus sp.* KK-1	533	AF045479	O52729
43	β-Xylosidase and arabinofuranosidase	*Selenomonas ruminantium* GA192	583	AF040720	O52575

Table 1 (*continued*)
Glycosyl hydrolases analyzed in this work

Family number[a]	Enzyme	Organism	Length	GenBank accession number	SWISS-PROT accession number
43	β-Xylosidase	*Escherichia coli*	536	U70214	P77713
43	β-Xylosidase and arabinofuranosidase	*Butyrivibrio fibrisolvens*	517	M55537	P45982
68	Levansucrase	*Bacillus subtilis*	473	Z99121	P05655
68	Levansucrase	*Bacillus stearothermophilus*	473	U34874	P94468
68	Levansucrase	*Bacillus amyloliquefaciens*	472	X52988	P21130
68	Levansucrase	*Bacillus sp.* V230	487	AB010272	
68	Levansucrase	*Streptococcus mutans*	797	M18954	P11701
68	Levansucrase	*Streptococcus salivarius*	969	L08445	Q55242
68	Levansucrase	*Acetobacter diazotrophicus*	584	L41732	Q43998
68	Levansucrase	*Erwinia amylovora*	415	X75079	Q46654
68	Levansucrase	*Rahnella aquatilis*	415	U91484	O54435
68	Levansucrase	*Pseudomonas syringae pv. glycinea*	415	AF037443	O52408
68	Levansucrase	*Pseudomonas syringae pv. phaseolicola*	431	AF052289	O68609
68	Levansucrase	*Zymomonas mobilis*	423	L33402	Q60114
68	Invertase	*Zymomonas mobilis*	413	L33403	Q60115
32	Invertase	*Tritrichomonas foetus*	550	U66071	O02490
32	Invertase	*Saccharomyces cerevisiae*	532	Z46921	P00724

[a]According to the classification of GHs [6,7].
[b]The protein consists of two homologous domains (see Fig. 1).
[c]The C-terminal domain belongs to family 43 and the N-terminal domain belongs to family 10 [6,7].

a component of the active center in enzymes of GH families 43 and 68. The conserved tyrosine cluster and preceding glycine residue (Fig. 1) also can be essential for the enzymatic activities.

The similarity found between the sequences of enzymes with β-D-fructofuranosidase and α-L-arabinofuranosidase activities is the first case showing that GHs recognizing different glycofuranoside residues can have a common sequence motif.

Acknowledgements: I thank Dr Eugene Koonin (NCBI) for a helpful discussion of the problem. This work was supported in part by the Russian Foundation for Basic Research (Grants 97-04-48818 and 96-15-97779).

References

[1] Bezzate, S., Steinmetz, M. and Aymerich, S. (1994) J. Bacteriol. 176, 2177–2183.
[2] Geier, G. and Geider, K. (1993) Physiol. Mol. Plant Pathol. 42, 387–404.
[3] Burne, R.A. and Penders, J.E.C. (1992) Infect. Immun. 60, 4621–4632.
[4] Shiroza, T. and Kuramitsu, H.K. (1988) J. Bacteriol. 170, 810–816.
[5] Naumov, D.G. and Doroshenko, V.G. (1998) Mol. Biol. Engl. Tr. 32, 761–766.
[6] Coutinho, P.M. (1999) Glycoside Hydrolase Family Server (http://afmb.cnrs-mrs.fr/–pedro/CAZY/ghf.html).
[7] Bairoch, A. (1998) SWISS-PROT Protein Sequence Data Bank (http://expasy.hcuge.ch/cgi-bin/lists?glycosid.txt).
[8] Sakamoto, T., Yamada, M., Kawasaki, H. and Sakai, T. (1997) Eur. J. Biochem. 245, 708–714.
[9] Gasparic, A., Martin, J., Daniel, A.S. and Flint, H.J. (1995) Appl. Environ. Microbiol. 61, 2958–2964.
[10] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[11] Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) Proteins Struct. Funct. Genet. 9, 180–190.
[12] Arrieta, J., Hernandez, L., Coego, A., Suarez, V., Balmori, E., Menendez, C., Petit-Glatron, M.-F., Chambert, R. and Selman-Housein, G. (1996) Microbiology 142, 1077–1085.
[13] Reddy, A. and Maley, F. (1996) J. Biol. Chem. 271, 13953–13958.