Relationships between SLH motifs from different glycoside hydrolase families

Richard Zona & Štefan Janeček*

Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava, Slovakia; phone: ++ 421 2 5930 7420, fax: ++ 421 2 5930 7416, e-mail: Stefan.Janecek@savba.sk

Abstract: Many glycoside hydrolases (GH) are very large proteins consisting of catalytic and non-catalytic domains. With regard to the non-catalytic domains, much research has been performed on the carbohydrate-binding modules (CBM), whereas substantially less attention has been paid to the surface layer homology (SLH) domain. The SLH sequences are involved in the attachment of proteins to the underlying cell wall. SLH domains are made of one to three repeats of 50 amino acids among which ten to fifteen residues are conserved. Three amylopullulanases from the α-amylase family GH-13 contain the SLH motifs; each in three copies. Within the CAZy classification, in addition to the α-amylase family GH-13, the typical SLH motifs are present in six other GH families: GH-5, GH-10, GH-16, GH-26, GH-28 and GH-73. Moreover, longer repeated domains which display some resemblance to SLH motifs have been identified in families GH-15 and GH-57. These so-called SLH motif-bearing domains contain two and a half typical SLH motifs. Based on the present sequence comparison data, a short sequence fingerprint, localized in the middle of the SLH motif, constitutes a novel third conserved region in glycoside hydrolase-associated SLH motifs. The evolutionary tree illustrates the relationships among the individual copies of the SLH motifs as well as between the typical SLH motifs and the longer SLH motif-bearing domains. It has been concluded that the evolutionary relationships of the SLH motifs reflect more taxonomy than the enzyme specificity of the catalytic domain to which they are linked.

Key words: SLH motif, glycoside hydrolase, alpha-amylase family.

Abbreviations: CBM, carbohydrate-binding module; GH, glycoside hydrolase; SLH, surface layer homology.

Introduction

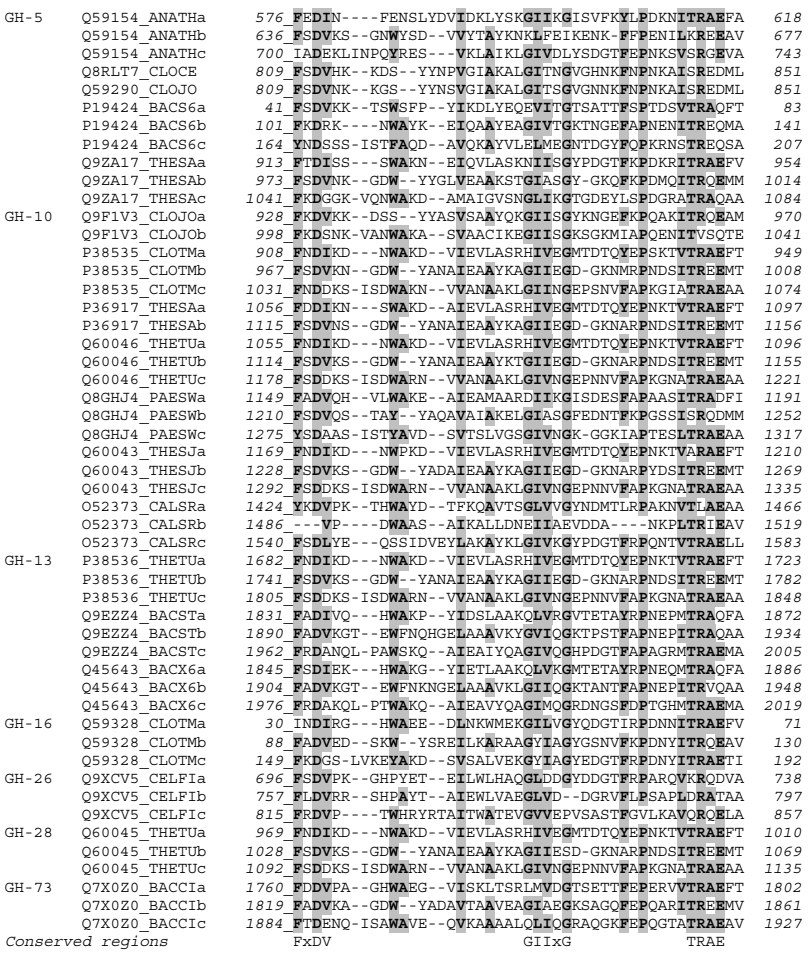
Surface layers (S-layers) from Bacteria and Archaea are built from protein molecules arrayed in a two-dimensional lattice, forming the outermost cell wall layer in many prokaryotes (Engelhardt & Peters, 1998). At the time of the discovery of the S-layers the sequence comparison of S-layers from distantly related bacteria did not reveal strong similarities. Nevertheless, one exceptional similarity was identified between the S-layer sequence, i.e. the N-terminal region of about 200 amino acid residues, of *Thermoanaerobacter kivui* and the N-terminal part of the middle wall protein of *Brevibacillus brevis* (Peters et al., 1989). This similarity was later shown to be a widely conserved motif among bacterial surface proteins and named as the S-layer homology (SLH) domain (Lupas et al., 1994). It was proposed to function as a peptidoglycan-binding structure of proteins to the underlying cell wall (Lupas et al., 1994). Later, SLH domains were shown to be both necessary and sufficient to bind cell walls (Lemaire et al., 1995; Mesnage et al., 1999).

At present the proteins possessing an SLH motif are divided into three groups (Engelhardt & Peters, 1998): (i) group I – S-layer proteins; (ii) group II – extracellular enzymes and proteins mostly involved in polysaccharide degradation; and (iii) group III – outer membrane proteins (Omps), also including some hypothetical proteins. The SLH motifs are located either at the N- or C-terminal end of the protein and the SLH domain consists of one to three SLH motifs (Lupas et al., 1994; Engelhardt & Peters, 1998; Mesnage et al., 2000). A typical SLH motif is a segment of ∼40–50 amino acids with 10–15 conserved residues, the C-terminus being the best conserved (Lupas et al., 1994; Engelhardt & Peters, 1998). According to the Pfam database (Bateman et al., 2002) the SLH module constitutes the family PF00395.

Since the entire sequences of the individual groups of the S-layer proteins do not share common similari-

* Corresponding author

(A) *SLH motifs:*



ties, the SLH domain must be regarded as a modular component that was linked to different proteins during evolution (Engelhardt & Peters, 1998).

Many glycoside hydrolases (GHs) contain SLH motifs (Schwarz et al., 2004). In the frame of the SLH classification they belong to group II. They are very large proteins consisting of catalytic and non-catalytic domains. With regard to the non-catalytic domains, much research has been performed on the carbohydrate binding modules (CBMs; Boraston et al., 2004), whereas a substantially less attention has been paid to and/or has been known for the SLH motifs (Beveridge et al., 1997). The S-layer protein and the three glycoside hydrolases of *Thermoanaerobacterium thermosulfurigenes* EM1 (GH-10 xylanase, GH-13 amylopullulanase and GH-28 polygalacturonase) were most deeply studied (Matuschek et al., 1994; 1996; Brechtel et al., 1999) with the conclusion that the SLH domains present in the S-layer and the enzymes are responsible for the anchoring of both protein types by binding of the SLH domain to the underlying peptidoglycan-containing sacculus (Brechtel et al., 1999). Using the C-terminally truncated forms of that xylanase (i.e. by removing the SLH motifs), Brechtel & Bahl (1999) demonstrated that multiple SLH motifs are necessary for the xylanase attachment to the cell wall.

Three amylopullulanases from the α-amylase family, i.e. the clan GH-H (MacGregor et al., 2001) contain the SLH modules; each in three copies. There are about 30 different enzyme specificities in the α-amylase clan GH-H (Janecek, 2002; Svensson et al., 2002; MacGregor, 2005) but the amylopullulanase is the only one containing the module. Within the all CAZy GH families (Coutinho & Henrissat, 1999), these SLH modules are present in families GH-5, GH-10, GH-16, GH-26, GH-28, and GH-73 in addition to the α-amylase family GH-13. Moreover the SLH-like motifs were found in two more families, in GH-57 (Erra-Pujada et al., 1999) and GH-15 (Mizuno et al., 2004). These SLH-like sequence segments were first described in the primary structure of GH-57 amylopullulanase from *Thermococcus hydrothermalis* and defined as the longer SLH motif-bearing domain containing two and a half typical SLH motifs (Erra-Pujada et al., 1999). Similar to the situation in α-amylase family GH-13, in GH-57 only the amylopullulanases appear to contain SLH-motif-bearing domains (Zona et al., 2004). Based on the three-dimensional structure of the GH-15 glucodextranase, the SLH motif-bearing domain covers

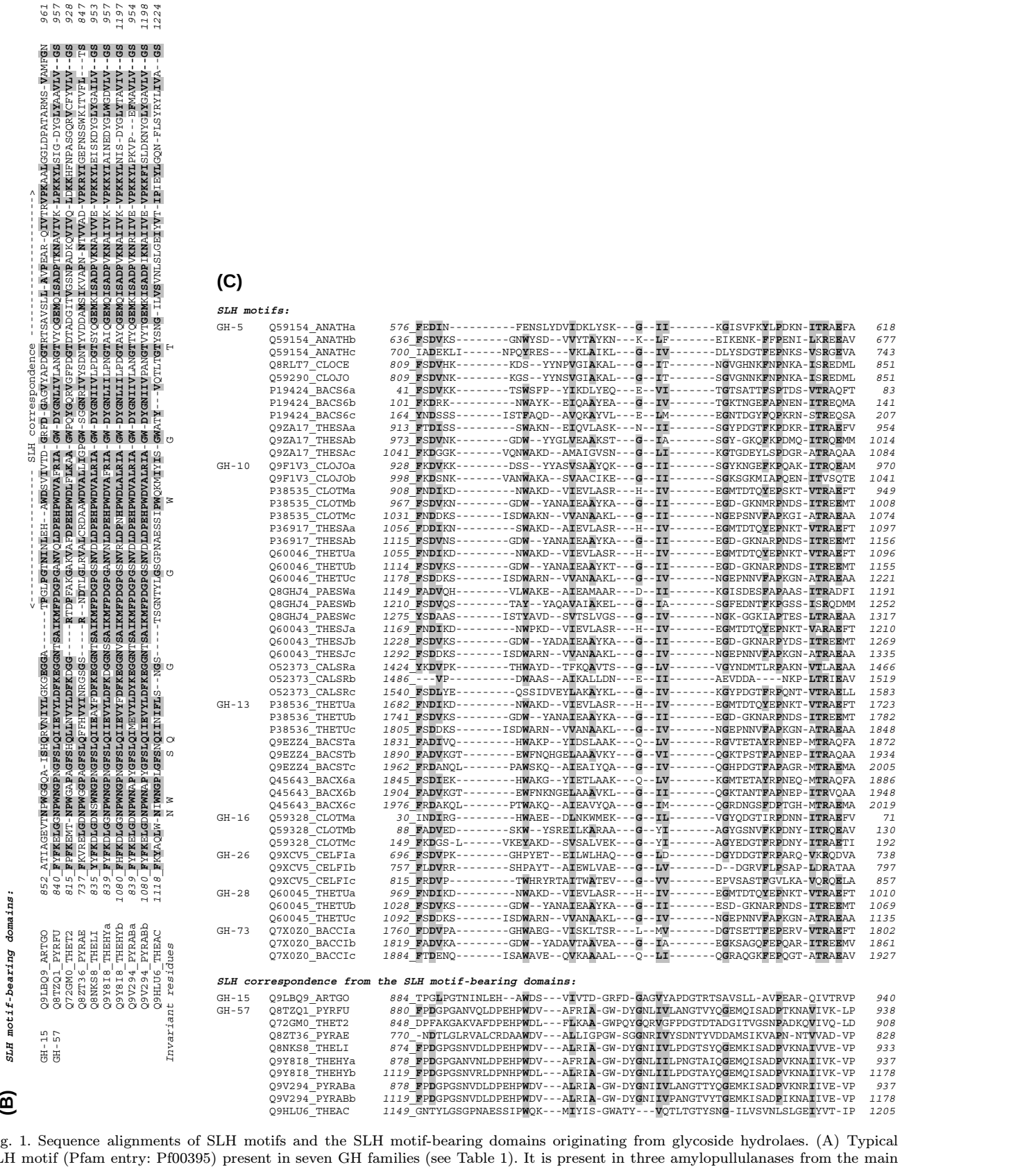


Fig. 1. Sequence alignments of SLH motifs and the SLH motif-bearing domains originating from glycoside hydrolaes. (A) Typical SLH motif (Pfam entry: Pf00395) present in seven GH families (see Table 1). It is present in three amylopullulanases from the main α-amylase family GH-13. The three conserved regions are indicated below the alignment. (B) SLH motif-bearing domains present in GH-15 and GH-57 containing two and a half typical SLH motif. The segment that best resembles the typical SLH motif is indicated as SLH correspondence. (C) Common alignment of SLH and the SLH motif-bearing domains. The second copy from the longer SLH-like motif was used to illustrate the sequence similarity. The abbreviations of enzyme sources are given in Table 1. The sequences are ordered according to increasing sequence length and, in the case of equal lengths, alphabetically. The residues conserved at least at 50% level are highlighted in grey.

Table 1. The various glycoside hydrolases containing the SLH motifs and SLH motif-bearing domains.[a]

Enzyme (hypothetical protein)	EC	Microorganism	Abbreviation	Family	Clan	Domain	Length	Copies	Sequences of SLH
SLH motifs									
Endo-1,4-glucanase	3.2.1.4	*Anaerocellum thermophilum*	Q59154_ANATH	GH-5	GH-A	Bacteria	749	3	576-618, 636-677, 700-743
β-1,4-Glucanase	n.d.	*Clostridium cellulolyticum*	Q8RLT7_CLOCE	GH-5	GH-A	Bacteria	930	1	809-851
Endo-1,4-glucanase	3.2.1.4	*Clostridium josui*	Q59290_CLOJO	GH-5	GH-A	Bacteria	930	1	809-851
Endo-1,4-glucanase	3.2.1.4	*Bacillus* sp. KSM-635	P19424_BACS6	GH-5	GH-A	Bacteria	941	3	41-83, 101-141, 164-207
β-Mannanase	3.2.1.78	*Thermoanaerobacterium polysaccharolyticum*	Q9ZA17_THESA	GH-5	GH-A	Bacteria	1097	3	913-954, 973-1014, 1041-1084
Xylanase	3.2.1.8	*Clostridium josui* FERM P-9684	Q9F1V3_CLOJO	GH-10	GH-A	Bacteria	1050	2	928-970, 998-1041
Xylanase	3.2.1.8	*Clostridium thermocellum* ATCC 27405	P38535_CLOTM	GH-10	GH-A	Bacteria	1087	3	908-949, 967-1008, 1031-1074
Xylanase	3.2.1.8	*Thermoanaerobacterium saccharolyticum*	P36917_THESA	GH-10	GH-A	Bacteria	1157	2	1056-1097, 1115-1156
Xylanase	3.2.1.8	*Thermoanaerobacterium thermosulfurigenes*	Q60046_THETU	GH-10	GH-A	Bacteria	1234	3	1055-1096, 1114-1155, 1178-1221
Xylanase	3.2.1.8	*Paenibacillus* sp. W-61	Q8GHJ4_PAESW	GH-10	GH-A	Bacteria	1326	3	1149-1191, 1210-1252, 1275-1317
Xylanase	3.2.1.8	*Thermoanaerobacterium* sp. JW/SL-YS 485	Q60043_THESJ	GH-10	GH-A	Bacteria	1348	3	1169-1210, 1228-1269, 1292-1335
Xylanase	3.2.1.8	*Caldicellulosiruptor* sp. Rt69B.1	O52373_CALSR	GH-10	GH-A	Bacteria	1595	3	1424-1466, 1486-1519, 1540-1583
Amylopullulanase	3.2.1.1/41	*Thermoanaerobacterium thermosulfurigenes*	P38536_THETU	GH-13	GH-H	Bacteria	1861	3	1682-1723, 1741-1782, 1805-1848
Amylopullulanase	3.2.1.1/41	*Bacillus stearothermophilus* TS-23	Q9EZZ4_BACST	GH-13	GH-H	Bacteria	2018	3	1831-1872, 1890-1934, 1962-2005
Amylopullulanase	3.2.1.1/41	*Bacillus* sp. XAL601	Q45643_BACX6	GH-13	GH-H	Bacteria	2032	3	1845-1886, 1904-1948, 1976-2019
Lichenase	3.2.1.73	*Clostridium thermocellum* DSM1237	Q59328_CLOTM	GH-16	GH-B	Bacteria	1321	3	30-71, 88-130, 149-192
β-Mannanase	3.2.1.78	*Cellulomonas fimi*	Q9XCV5_CELFI	GH-26	GH-A	Bacteria	1010	3	696-738, 757-797, 815-857
Polygalacturonase	3.2.1.15	*Thermoanaerobacterium thermosulfurigenes*	Q60045_THETU	GH-28	GH-N	Bacteria	1148	3	969-1010, 1028-1069, 1092-1135
Endo-β-*N*-acetylglucosaminidase	3.2.1.96	*Bacillus circulans*	Q7X0Z0_BACCI	GH-73	-	Bacteria	1936	3	1760-1802, 1819-1861, 1884-1927
SLH motif-bearing domains									
Glucodextranase	3.2.1.70	*Arthrobacter globiformis* I42	Q9LBQ9_ARTGO	GH-15	GH-L	Bacteria	1048	1	852-961
PF1935 (amylopullulanase)	n.d.	*Pyrococcus furiosus* DSM3638	Q8TZQ1_PYRFU	GH-57	-	Archaea	985	1	840-957
TTC1828	n.d.	*Thermus thermophilus* HB27	Q72GM0_THET2	GH-57	-	Bacteria	994	1	815-928
PAE3454 (pullulanase)	n.d.	*Pyrobaculum aerophilum* IM2	Q8ZT36_PYRAE	GH-57	-	Archaea	999	1	737-847
Amylopullulanase	3.2.1.1/41	*Thermococcus litoralis*	Q8NKS8_THELI	GH-57	-	Archaea	1089	1	835-953
Amylopullulanase	3.2.1.1/41	*Thermococcus hydrothermalis*	Q9Y8I8_THEHY	GH-57	-	Archaea	1337	2	839-957, 1080-1197
PAB0122 (amylopullulanase)	n.d.	*Pyrococcus abyssi* GE5	Q9V294_PYRAB	GH-57	-	Archaea	1362	2	839-954, 1080-1198
TA0129	n.d.	*Thermoplasma acidophilum* DSM1728	Q9HLU6_THEAC	GH-57	-	Archaea	1641	1	1118-1224

[a] The individual copies of the SLH motifs are marked throughout the manuscript as "a", "b" and "c" in the order of appearance in the sequence, e.g. Q59154_ANATHa, Q59154_ANATHb and Q59154_ANATHc, respectively. The first part of the abbreviation is formed by the UniProt Accession No. (e.g. Q59154 for the endo-1,4-glucanase from *Anaerocellum thermophilum*).

several β-strand segments forming thus a substantial part of the C-terminal domain C (Mizuno et al., 2004; 2005).

The aim of this work was to present the evolutionary picture that illustrates the relationships among the individual copies of the SLH motifs originating from a single GH enzyme sequence and/or a GH family, among the SLH motifs derived from the various GH families as well as between the typical SLH motifs and the longer SLH motif-bearing domains.

Background

The enzymes belonging to various GH families involved in the present study are listed in Table 1. To collect the sequences, the CAZy (Coutinho & Henrissat, 1999) and Pfam (Bateman et al., 2002) server and database, respectively, were used:

– CAZy at http://afmb.cnrs-mrs.fr/CAZY/ (July 2004);

– Pfam at http://www.sanger.ac.uk/Software/Pfam/index.shtml (August 2004).

The sequences were retrieved from GenBank (Benson et al., 2004) and UniProt (Apweiler et al., 2004) sequence databases. Two alignments (the typical SLH motifs and the longer SLH motif-bearing domains) were done using the program CLUSTAL W (Thompson et al., 1994) with partial corrections performed manually. The alignment of the typical SLH motifs together with the longer SLH motif-bearing domains was made completely manually.

The method used for building the evolutionary trees was the neighbour-joining method (Saitou & Nei, 1987). The Phylip format tree output was applied (Felsenstein, 1985) using the bootstrapping procedure; the number of bootstrap trials used was 1,000. The trees were drawn with the program TreeView (Page, 1996).

Results and discussion

Sequence comparison

Three sequence alignments are presented: (i) the typical SLH motifs (Fig. 1A); (ii) the SLH motif-bearing domains (Fig. 1B); and (iii) the combination of both motifs (Fig. 1C).

The typical SLH motifs (51 sequences derived from 19 enzymes) were taken from seven GH families. With regard to the α-amylase family, three extremely long GH-13 amylopullulanases ($\sim$ 2,000 residues) possess this motif (Lee et al., 1994; Matuschek et al., 1994; Chen et al., 2001); each in three copies. The motif usually exists in three copies, being rarely found as a single motif or duplicate (Table 1). The length varies around 40 residues. No residue was found to be invariantly conserved, however, a few positions are very well conserved, especially at the N-terminal and the C-terminal end of the motif (Fig. 1A). These two short regions (FxDV and TRAE; Fig. 1A) are considered to be the two conserved sequence regions of the SLH motif (Schwarz et al., 2004). Based on our comparison, a third short segment in the middle of the SLH motif, GIIxG (Fig. 1A), seems to be highly conserved, at least among the SLH motif originating from GH enzymes. It will be possible to give a more detailed view once the comparison of more than 400 copies of the SLH motifs from all of the SLH-containing proteins will have been completed (R. Zona & S. Janecek, in preparation). This will con-

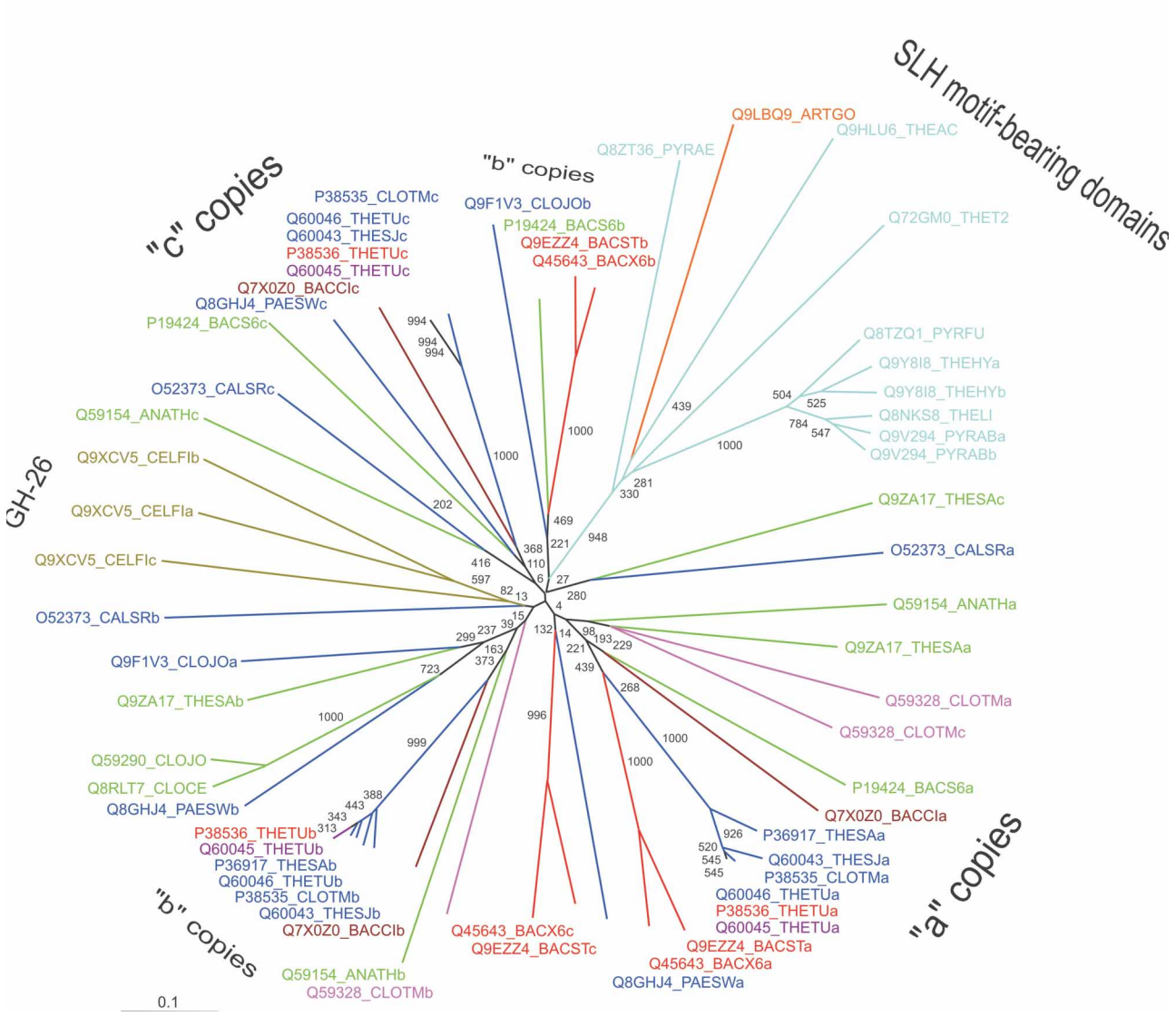


Fig. 2. Evolutionary tree of various glycoside hydrolases containing the SLH motifs and the SLH motif-bearing domains. The tree is based on the alignment shown in Fig. 1C. It was calculated using gapped sequences. Branch lengths are proportional to sequence divergence. The abbreviations of enzyme sources and the colour code are explained in Table 1.

cern the individual groups of the SLH motifs based on taxonomy, protein function and SLH copy-associated evolutionary relatedness.

The SLH motif-bearing domains (10 sequences derived from 8 enzymes) come from the two GH families: GH-15 and GH-57. The lengths of these motifs are approximately 115 residues. The sequence similarity is obviously higher than among the typical SLH motifs; this may be, however, mainly due to narrower spectrum of the enzyme sources (only two specificities) and GH families (9 of 10 sequences being from the GH-57). This resulted in 9 residues found totally conserved throughout the alignment (Fig. 1B). Interestingly, two of the nine are tryptophan and three are glycine. The substantially lesser amount of knowledge on these longer SLH-like sequences (Erra-Pujada et al., 1999; Mizuno et al., 2004; 2005) may also be due to their infrequent occurrence in proteins (cf. Table 1).

The SLH motif-bearing domain was originally defined as two and a half typical SLH motifs (Erra-Pujada et al., 1999). It should be pointed out, however, that the first copy at the N-terminal end and the half at the C-terminal end of the domain exhibit only marginal similarity to the typical SLH motif (cf. Fig. 1A and Fig. 1B). On the other hand, the middle copy possesses clear correspondences with the SLH motif (Fig. 1B).

In order to draw the evolutionary relationships among the individual copies of the SLH motifs and the SLH motif-bearing domains as well as among these two SLH groups, the alignment joining the two motifs together was prepared (Fig. 1C). It is evident that the present-day SLH modules share several common sequence features, however, there are many differences indicating a remote homology only. Since the second copy of the longer SLH motif-bearing domain (two and a half of a typical SLH motif; Fig. 1B) was found to exhibit

the highest similarity to the typical SLH motif, it was taken to show the correspondences in Figure 1C.

Of the two well-accepted conserved sequence regions that are best conserved among the SLH motifs (Schwarz et al., 2004) only the first segment (FxDV) has its clear counterpart in the SLH motif-bearing domains (Fig. 1C). The second segment (TRAE) cannot be identified. It is worth mentioning that the third conserved segment proposed here (GIIxG; Fig. 1C) can be present in the SLH motif-bearing domain, although it is necessary to insert a few gaps to achieve the correspondences. The insertions, however, may reflect the above-mentioned remote homology.

Evolutionary relationships

The evolutionary tree common for both the SLH and SLH motif-bearing domains is shown in Figure 2. One of the expected results is that the longer SLH-like motifs were not scattered among the typical SLH motifs, i.e. each of the two types keeps its own independence.

The position and the branch length of the only representative originating from the family GH-15 (Mizuno et al., 2004; 2005) indicate that its similarity to the rest of the SLH motif-bearing domains is comparable to those found between the motifs originating from the same family GH-57. Within the family GH-57 there are only two amylopullulanases that contain the longer SLH motifs in two copies: from *Thermococcus hydrothermalis* and *Pyrococcus abyssi* (cf. Table 1). The biochemistry of the former amylopullulanase has been studied in a detail (Erra-Pujada, 2001; Chang-Pi-Hin et al., 2002) whereas the latter enzyme is a putative protein deduced from the genome ORF (Cohen et al., 2003). It should be pointed out that the copies ("a" and "b") in both cases share the same branch (Fig. 2).

With regard to the typical (shorter) SLH motifs, four groups were revealed that can be characterized as copy-specific groups, i.e. groups containing the same copies in terms of their appearance in the sequence. The copies marked as "a" and "c", i.e. the first and the third copy of the motif, form their own groups, whereas the copy located in the middle, marked as "b", forms one larger and one smaller group (Fig. 2). The most important observation is that all these groups are formed regardless the GH family from which the SLH motif originates. The only β-mannanase from GH-26 (Stoll et al., 1999) should be of interest since it contains all the three SLH copies that are mutually similar thus forming their own cluster (Fig. 2).

It is not easy to hypothesize about the fact why some amylopullulanases and also some α-D-glucan acting enzymes (or members from various GH families; see Table 1) are preferentially associated with SLH motifs. This fact can be compared with the presence of starch-binding domain mainly of the CBM-20 type in the sequences of amylolytic enzymes. Only 10% of sequences of amylases contain that domain (Janecek & Sevcik, 1999; Janecek et al., 2003; Rodriguez-Sanoja et al., 2005). It might be a consequence of some advantageous evolutionary behavior that is still not fully understood.

It could be concluded that, in general, the evolutionary relationships of the SLH motifs reflect more taxonomy than the enzyme specificity of the catalytic domain to which they are linked. This fact seems to be a more general feature of non-catalytic modules of glycoside hydrolases since also, e.g., the above-mentioned starch-binding domain of the CBM-20 type exhibits similar behaviour (Janecek & Sevcik, 1999; Janecek et al., 2003). A more detailed study taking into account all available SLH motifs, i.e. not only those present in glycoside hydrolases studied here, is in progress.

Acknowledgements

This work was financially supported in part by the VEGA grant number 2/5067/25 from the Slovak Grant Agency for Science.

References

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. & Yeh, L.S. 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. **32:** D115–D119.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. & Sonnhammer, E.L. 2002. The Pfam protein families database. Nucleic Acids Res. **30:** 276–280.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. 2004. GenBank: update. Nucleic Acids Res. **32:** D23–D26.

Beveridge, T.J., Pouwels, P.H., Sara, M., Kotiranta, A., Lounatmaa, K., Kari, K., Kerosuo, E., Haapasalo, M., Egelseer, E.M., Schocher, I., Sleytr, U.B., Morelli, L., Callegari, M.L., Nomellini, J.F., Bingle, W.H., Smit, J., Leibovitz, E., Lemaire, M., Miras, I., Salamitou, S., Beguin, P., Ohayon, H., Gounon, P., Matuschek, M. & Koval, S.F. 1997. Functions of S-layers. FEMS Microbiol. Rev. **20:** 99–149.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. **382:** 769–781.

Brechtel, E. & Bahl, H. 1999. In *Thermoanaerobacterium thermosulfurigenes* EM1 S-layer homology domains do not attach to peptidoglycan. J. Bacteriol. **181:** 5017–5023.

Brechtel, E., Matuschek, M., Hellberg, A., Egelseer, E.M., Schmid, R. & Bahl, H. 1999. Cell wall of *Thermoanaerobacterium thermosulfurigenes* EM1: isolation of its components and attachment of the xylanase XynA. Arch. Microbiol. **171:** 159–165.

Chang-Pi-Hin, F., Erra-Pujada, M., Dauchez, M., Debeire, P., Duchiron, F. &. O'Donohue, M.J. 2002. Expression and characterization of the catalytic domain of an archaeal family 57 pullulanase type II. Biologia, Bratislava **57 (Suppl. 11):** 155–162.

Chen, J.T., Chen, M.C., Chen, L.L. & Chu, W.S. 2001. Structure and expression of an amylopullulanase gene from *Bacillus stearothermophilus* TS-23. Biotechnol. Appl. Biochem. **33:** 189–199.

Cohen, G.N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J.C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol. Microbiol. **47:** 1495–1512.

Coutinho, P. M. & Henrissat, B. 1999. Carbohydrate-Active Enzymes server; http://afmb.cnrs-mrs.fr/CAZY/.

Engelhardt, H. & Peters, J. 1998. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions. J. Struct. Biol. **124:** 276–302.

Erra-Pujada, M., Debeire, P., Duchiron, F. & O'Donohue, M.J. 1999. The type II pullulanase of *Thermococcus hydrothermalis*: molecular characterization of the gene and expression of the catalytic domain. J. Bacteriol. **181:** 3284–3287.

Erra-Pujada, M., Chang-Pi-Hin, F., Debeire, P., Duchiron, F. & O'Donohue, M.J. 2001. Purification and properties of the catalytic domain of the thermostable pullulanase type II from *Thermococcus hydrothermalis*. Biotechnol. Lett. **23:** 1273–1277.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

Janecek, S. 2002. How many conserved sequence regions are there in the α-amylase family? Biologia, Bratislava **57 (Suppl. 11):** 29–41.

Janecek, S. & Sevcik, J. 1999. The evolution of starch-binding domain. FEBS Lett. **456:** 119–125.

Janecek, Š., Svensson, B. & MacGregor, E.A. 2003. Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. **270:** 635–645.

Lee, S.P., Morikawa, M., Takagi, M. & Imanaka, T. 1994. Cloning of the *aapT* gene and characterization of its product, α-amylase-pullulanase (AapT), from thermophilic and alkaliphilic *Bacillus* sp. strain XAL601. Appl. Environ. Microbiol. **60:** 3764–3773.

Lemaire, M., Ohayon, H., Gounon, P., Fujino, T. & Béguin, P. 1995. OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope. J. Bacteriol. **177:** 2451–2459.

Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S. & Baumeister, W. 1994. Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J. Bacteriol. **176:** 1224–1233.

MacGregor, E.A. 2005. An overview of clan GH-H and distantly-related families. Biologia, Bratislava **60 (Suppl. 16):** 5–12.

MacGregor, E.A., Janecek, S. & Svensson, B. 2001. Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim Biophys Acta **1546:** 1–20.

Matuschek, M., Burchhardt, G., Sahm, K. & Bahl, H. 1994. Pullulanase of *Thermoanaerobacterium thermosulfurigenes* EM1 (*Clostridium thermosulfurogenes*): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface. J. Bacteriol. **176:** 3295–3302.

Matuschek, M., Sahm, K., Zibat, A. & Bahl, H. 1996. Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that encode two glycosyl hydrolases with conserved S-layer-like domains. Mol. Gen. Genet. **252:** 493–496.

Mesnage, S., Tosi-Couture, E. & Fouet, A. 1999. Production and cell surface anchoring of functional fusions between the SLH motifs of the *Bacillus anthracis* S-layer proteins and the *Bacillus subtilis* levansucrase. Mol. Microbiol. **31:** 927–936.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M. & Fouet, A. 2000. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J. **19:** 4473–4484.

Mizuno, M., Tonozuka, T., Ichikawa, K., Kamitori, S., Nishikawa, A. & Sakano, Y. 2005. Three-dimensional structure of glucodextranase, a glycoside hydrolase family 15 enzyme. Biologia, Bratislava **60 (Suppl. 16):** 171–177.

Mizuno, M., Tonozuka, T., Suzuki, S., Uotsu-Tomita, R., Kamitori, S., Nishikawa, A. & Sakano, Y. 2004. Structural insight into substrate specificity and function of glucodextranase. J. Biol. Chem. **279:** 10575–10583.

Page, R.D. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Applic. Biosci. **12:** 357–358.

Peters, J., Peters, M., Lottspeich, F. & Baumeister, W. 1989. S-layer protein gene of *Acetogenium kivui*: cloning and expression in *Escherichia coli* and determination of the nucleotide sequence. J. Bacteriol. **171:** 6307–6315.

Rodriguez-Sanoja, R., Oviedo, N. & Sanchez, S. 2005. Microbial starch-binding domain. Curr. Opin. Microbiol. **8:** 260–267.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Schwarz, W.H., Zverlov, V.V. & Bahl, H. 2004. Extracellular glycosyl hydrolases from clostridia. Adv. Appl. Microbiol. **56:** 215–261.

Stoll, D., Stalbrand, H. & Warren, R.A.J. 1999. Mannan-degrading enzymes from *Cellulomonas fimi*. Appl. Environ. Microbiol. **65:** 2598–2605.

Svensson, B., Jensen, M.T., Mori, H., Bak-Jensen, K.S., Bŕnsager, B., Nielsen, P.K., Kramhŕft, B., Prtorius-Ibba, M., Nŕhr, J., Juge, N., Greffe, L., Williamson, G. & Driguez, H. 2002. Fascinating facets of function and structure of amylolytic enzymes of glycoside hydrolase family 13. Biologia, Bratislava **57 (Suppl. 11):** 5–19.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Zona, R., Chang-Pi-Hin, F., O'Donohue, M.J. & Janecek, S. 2004. Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. Eur. J. Biochem. **271:** 2863–2872.

Received February 9, 2005
Accepted April 1, 2005