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 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 twodimensional 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 Slayer 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 & PE-TERS, 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 (Lu-PAS et al., 1994; ENGELHARDT & PETERS, 1998; MES-NAGE et al., 2000). A typical SLH motif is a segment of \sim 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:

GH-5	Q59154 ANATHa	576	FEDINFENSLYDVIDKLYSKGIIKGISVFKYLPDKNITRAEFA	618
	059154 ANATHb	636	FSDVKSGNWYSDVVYTAYKNKLFEIKENK-FFPENILKREEAV	677
	Q59154 ANATHC	700	IADEKLINPQYRESVKLAIKLGIVDLYSDGTFEPNKSVSRGEVA	743
	Q8RLT7 CLOCE	809	FSDVHKKDSYYNPVGIAKALGITNGVGHNKFNPNKAISREDML	851
	Q59290 CLOJO	809	FSDVNKKGSYYNSVGIAKALGITSGVGNNKFNPNKAISREDML	851
	P19424 BACS6a	41	FSDVKKTSWSFPYIKDLYEQEVITGTSATTFSPTDSVTRAQFT	83
	P19424_BACS6b	101_	FKDRKNWAYKEIQAAYEAGIVTGKTNGEFAPNENITREQMA	141
	P19424_BACS6c	164	YNDSSS-ISTFAQDAVQKAYVLELMEGNTDGYFQPKRNSTREQSA	207
	Q9ZA17_THESAa	913_	FTDISSSWAKNEIQVLASKNIISGYPDGTFKPDKRITRAEFV	954
	Q9ZA17_THESAb	973_	FSDVNKGDWYYGLVEAAKSTGIASGY-GKQFKPDMQITRQEMM	1014
	Q9ZA17_THESAc	1041	FKDGGK-VQNWAKDAMAIGVSNGLIKGTGDEYLSPDGRATRAQAA	1084
GH-10	Q9F1V3_CLOJOa	928_	FKDVKKDSSYYASVSAAYQKGIISGYKNGEFKPQAKITRQEAM	970
	Q9F1V3_CLOJOb	998_	FKDSNK-VANWAKASVAACIKEGIISGKSGKMIAPQENITVSQTE	1041
	P38535_CLOTMa	908_	FNDIKDNWAKDVIEVLASRHIVEGMTDTQYEPSKTVTRAEFT	949
	P38535_CLOTMb	967_	FSDVKNGDWYANAIEAAYKAGIIEGD-GKNMRPNDSITREEMT	1008
	P38535_CLOTMc	1031_	FNDDKS-ISDWAKNVVANAAKLGIINGEPSNVFAPKGIATRAEAA	1074
	P36917_THESAa	1056_	FDDIKNSWAKDAIEVLASRHIVEGMTDTQYEPNKTVTRAEFT	1097
	P36917_THESAb	1115_	FSDVNSGDWYANAIEAAYKAGIIEGD-GKNARPNDSITREEMT	1156
	Q60046_THETUa	1055_	FNDIKDNWAKDVIEVLASRHIVEGMTDTQYEPNKTVTRAEFT	1096
	Q60046_THETUb	1114_	FSDVKSGDWYANAIEAAYKTGIIEGD-GKNARPNDSITREEMT	1155
	Q60046_THETUC	1178_	FSDDKS-ISDWARNVVANAAKLGIVNGEPNNVFAPKGNATRAEAA	1221
	Q8GHJ4_PAESWa	1149_	FADVQHVLWAKEAIEAMAARDIIKGISDESFAPAASITRADFI	1191
	Q8GHJ4_PAESWb	1210_	FSDVQSTAYYAQAVAIAKELGIASGFEDNTFKPGSSISRQDMM	1252
	Q8GHJ4_PAESWc	1275_	YSDAAS-ISTYAVDSVTSLVGSGIVNGK-GGKIAPTESLTRAEAA	1317
	Q60043_THESJa	1169_	FNDIKDNWPKDVIEVLASRHIVEGMTDTQYEPNKTVARAEFT	1210
	Q60043_THESJb	1228_	FSDVKSGDWYADAIEAAYKAGIIEGD-GKNARPYDSITREEMT	1269
	Q60043_THESJC	1292_	FSDDKS-ISDWARNVVANAAKLGIVNGEPNNVFAPKGNATRAEAA	1335
	052373_CALSRa	1424_	YKDVPKTHWAYDTFKQAVTSGLVVGYNDMTLRPAKNVTLAEAA	1466
	052373_CALSRb	1486_	VPDWAASAIKALLDNEIIAEVDDANKPLTRIEAV	1519
	052373_CALSRC	1540_	FSDLYEQSSIDVEYLAKAYKLGIVKGYPDGTFRPQNTVTRAELL	1583
GH-13	P38536_THETUa	1682_	FNDIKDNWAKDVIEVLASRHIVEGMTDTQYEPNKTVTRAEFT	1723
	P38536_THETUb	1741_	FSDVKSGDWYANAIEAAYKAGIIEGD-GKNARPNDSITREEMT	1782
	P38536_THETUC	1805_	FSDDKS-ISDWARNVVANAAKLGIVNGEPNNVFAPKGNATRAEAA	1848
	Q9EZZ4_BACSTa	1831_	FADIVQHWAKPYIDSLAAKQLVRGVTETAYRPNEPMTRAQFA	1872
	Q9EZZ4_BACSTD	1890_	FADVKGTEWFNQHGELAAAVKYGVIQGKTPSTFAPNEPITRAQAA	1934
	Q9EZZ4_BACSTC	1962_	FRDANQL-PAWSKQAIEAIYQAGIVQGHPDGTFAPAGRMTRAEMA	2005
	Q45643_BACA6a	1845_		1886
	Q45643_BACX6D	1904_		1948
CII 1 C	Q45643_BACA6C	1976_	FRDARQL-PIWARQAIEAVIQAGIMQGRDNGSFDPIGHMIRAEMA	2019
GH-16	Q59328_CLOIMa	30_	INDIRGHWAREDUNKWMERGILVGIQUGIIRPDNNITRAEFV	120
	Q59328_CLOIMD	88_	FADVEDSKWISKEILKARAAGIIAGIGSNVFRPDNIITRQEAV	130
au ac	Q59328_CLOIMC	149_		192
GH-26	Q9ACV5_CELFIA	696_ 757		738
	Q9XCV5_CELFID	/5/_		057
GH=28	OG0045 THETTS	960		1010
G11=20	OCODAE TUETTA	1020	PROTECT	1060
	060045 THETUC	1028	RODKS-TOWARN-TIMIALDANIAAKICTUNGEDNNUFADKGMATOADAA	1135
CU-72	07X070 PACCTa	1760		1002
GU= 12	07X0Z0 BACCIA	1810	FDEVER - COW VADAWTAAVFACTAFCKSACOFFDOADT TO FEMU	1861
	07X0Z0_BACCID	1884	FTDENO-ISAWAVEOVKAAAALOLTOGRAOGKEEDOGTATDAEAV	1927
Concern	ed regions	1004	EVDN CITYC TDAE	1721

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 (BEV-ERIDGE 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; BRECH-TEL 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 amylopullulaneses 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

	<i>0000000000000000000000000000000000000</i>							
	00000000000000000000000000000000000000							
	INTA NTA NTA NTA NTA NTA NTA NTA							
	KAA CFY VGA VGA KTA KYL							
	ACC SGC SGC SGC SGC SGC SGC SGC SGC SGC S							
	C C C C C C C C C C C C C C C C C C C							
	KA KK KK KK KK IE							
	~ T T T T T T T T T T T T T T T T T T T							
	L A A A A A A A A A A A A A A A A A A A							
	NUL NUL NKN NKN NKN SLG							
		$\langle \mathbf{O} \rangle$						
	L - A L C C C C C C C C C C C C C C C C C C C	(C)						
	LIS NWK	SLH mo	tifs:					
	- TGGI TGGI NG- CGI NG- CGI NG	GH-5	Q59154 ANATHa	576 FEDIN	FENSLYDVIDKLYSK	GII	KGISVFKYLPDKN-	itr a e fa
	TTYY TTYY TTYY TTYY TYYY		Q59154_ANATHb	636 _ FS DV KS	GNWYSDVVYTAYKN	-KLF	EIKENK-FFPENI-I	lkreeav
	D D D D D D D D D D D D D D D D D D D		Q59154_ANATHC OSBLT7_CLOCE	700_IADEKLI	NPQYRESVKLAIKL	GIV	DLYSDGTFEPNKS-V	VSRGEVA
	ULP DULP DULP DULP DULP DULP DULP DULP D		Q59290 CLOJO	809 FSDVNK	·KGSYYNSVGIAKAL	GIT	SGVGNNKFNPNKA-	ISREDML
	AGU AGU ALL ALL ALL ALL ALL ALL ALL V		P19424_BACS6a	41 _F S DV KK	TSWSFPYIKDLYEQ	E VI	TGTSATTFSPTDS-V	VTR AQFT
	- L C C C C C C C C C C C C C C C C C C		P19424_BACS6b P19424_BACS6c	101_FKDRK 164_YNDSSS	·ISTFAODAVOKAYEA	GIV	EGNTDGYFOPKEN-	ITREQMA STREOSA
			Q9ZA17_THESAa	913_FTDISS	SWAKNEIQVLASK	-N II	SGYPDGTFKPDKR-I	ITRAEFV
	H		Q9ZA17_THESAb	973_FSDVNK	GDWYYGLVEAAKST	GIA	SGY-GKQFKPDMQ-I	ITRQEMM
	V RLA RLA RLA RLA RLA RLA VIS	GH-10	Q9ZAI7_IHESAC Q9F1V3 CLOJOa	928 FKDGGK	DSSYYASVSAAYOK	GLI	SGYKNGEFKPOAK-	ATRAQAA ITROEAM
	AFL AFL AFL AFL AFL AFL AFL AFL AFL AFL		Q9F1V3_CLOJOb	998_ F K D SNK	VANWAKASVAACIKE	GII	SGKSGKMIAPQEN-I	it vŝqte
			P38535_CLOTMa P38535_CLOTMb	908_FNDIKD	NWAKDVIEVLASR	-HIV	EGMTDTQYEPSKT-V	VTRAEFT ITREEMT
			P38535_CLOTMc	1031_FNDDKS	ISDWAKNVVANAAKL	GII	NGEPSNVFAPKGI-A	ATRAEAA
			P36917_THESAa	1056_FDDIKN	SWAKDAIEVLASR	-HIV	EGMTDTQYEPNKT-V	VTRAEFT
			060046 THETUA	1055 FNDIKD	·GDWYANAIEAAYKA ·VIEVLASR	GII HIV	EGD-GKNARPNDS-I	VTRAEFT
	- HERRY AND		Q60046_THETUb	1114 _ FS DV KS	GDWYANAIEAAYKT		EGD-GKNARPNDS-I	ITREEMT
			Q60046_THETUC	1178_FSDDKS	ISDWARNVVANAAKL	GIV	NGEPNNVFAPKGN-A	ATRAEAA
			Q8GHJ4_PAESWa Q8GHJ4 PAESWb	1210 FSDVQH	TAYYAQAVAI A KEL		SGFEDNTFKPGSS-I	ISRODMM
	T R T R T R T R T R T R T R T R T R T R		Q8GHJ4_PAESWc	1275 YSDAAS	ISTYAVDSVTSLVGS	GIV	NGK-GGKIAPTES-I	LTRAEAA
	SAJ SAJ SAJ SAJ SAJ SAJ SAJ		Q60043_THESJa O60043_THESJb	1169_FNDIKD	OWPKDVIEVLASR	н IV G тт	EGMTDTQYEPNKT-V	VARAEFT ITREEMT
			Q60043_THESJc	1292_FSDDKS	ISDWARNVVANAAKL	GIV	NGEPNNVFAPKGN-A	A TR A E AA
	C C C C C C C C C C C C C C C C C C C		052373_CALSRa	1424_YKDVPK	THWAYDTFKQAVTS	GLV	VGYNDMTLRPAKN-V	VTLAEAA
	DFN DFN DFN DFN CFN CFN CFN CFN CFN CFN CFN CFN CFN C		052373 CALSRC	1540 FSDLYE	OSSIDVEYLAKAYKL	GIV	KGYPDGTFRPONT-	VTRAELL
		GH-13	P38536_THETUa	1682 _FNDI KD	NWAKDVIEVLASR	-HIV	EGMTDTQYEPNKT-V	VTRAEFT
	VN: TEV TEV TEV TEV TEV TEV TEV		P38536_THETUb P38536_THETUC	1741_FSDVKS	GDWYANAIEAAYKA ISDWARNVVANAAKI	GII GTV	EGD-GKNARPNDS-	ITREEMT Atraeaa
			Q9EZZ4_BACSTa	1831_FADIVQ	HWAKPYIDSLAAK	Q LV	RGVTETAYRPNEP-N	M TR AQFA
			Q9EZZ4_BACSTb	1890_FADVKGT	EWFNQHGELAAAVKY	GVI	QGKTPSTFAPNEP-I	ITRAQAA
			Q9E224_BACSIC 045643 BACX6a	1962_FRDANQL 1845 FSDIEK	·HWAKGYIETLAAK	OLV	KGMTETAYRPNEO-N	M TRAE MA M TR AOFA
	VCG VNG VNG VNG VNG VNG VNG VNG VNG VNG VN		Q45643_BACX6b	1904 _F A DV KGT	EWFNKNGELAAAVKL	ĜII	QGKTANTFAPNEP-I	itr võaa
	Idni Idni Idni Idni Idni Idni Idni Idni	GH-16	Q45643_BACX6c	1976_FRDAKQL	PTWAKQAIEAVYQA	GIM	QGRDNGSFDPTGH-N	MTRAEMA ITRAEEV
		011 10	Q59328_CLOTMb	88_FADVED	SKWYSREILKARAA	GYI	AGYGSNVFKPDNY-I	ITRQEAV
	NE NE NE NE NE NE NE NE NE NE		Q59328_CLOTMc	149 F K D GS-L	VKE Y AKDS V SALVEK		AGYEDGTFRPDNY-I	ITRAETI
	АТЛ 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	GH-26	Q9XCV5_CELFIa O9XCV5_CELFIb	696_FSDVPK 757 FLDVRR	·GHPYETEILWLHAQ ·SHPAYTAIEWLVAE	GLD	DGYDDGTFRPARQ-V	VKRQDVA Ldrataa
			Q9XCV5_CELFIC	815_ F R DV P	TWHRYRTAITWATEV	GVV	EPVSASTFGVLKA-V	VQRQELA
: 51	8887880000 11 11 18	GH-28	Q60045_THETUa	969_FNDIKD	NWAKDVIEVLASR	-HIV	EGMTDTQYEPNKT-V	VTRAEFT
air			Q60045_IHEIUD O60045_THETUC	1028_FSDVKS 1092 FSDDKS	ISDWARNVVANAAKL	GII	NGEPNNVFAPKGN-A	ATRAEAA
lon	K B B B B B B B B B B B B B B B B B B B	GH-73	Q7X0Z0_BACCIa	1760 _F D DV PA	GHWAEGVISKLTSR	LM V	DGTSETTFEPERV-V	VTRAEFT
ь Бл	YRI HEH HEH HEH HEH HEZ VR/ C		Q7X0Z0_BACCIb	1819_FADVKA	GDWYADAVTAAVEA	GIA	EGKSAGQFEPQAR-I	ITREEMV Atraeav
rin	онооваваар 1 4. П.		z.molo_bheere		TOTINITY OF CONSTRUCT	×	SerarSour 1: 501-1	
Jea.	E C C C C C C C C C C C C C C C C C C C	SLH CO	rrespondence from	the SLH motif-be	earing domains:		_	
f-L	0000000000000000000000000000000000000	GH-15	Q9LBQ9_ARTGO	884_TPGLPGTNIN	ILEHAWDSVIVTD-GRFI	-GAGVYAPE	GTRTSAVSLL-AVPEAR-(QIVTRVP
ti	e e	GH-27	Q8TZQ1_PYRFU Q72GM0 THET2	880_FPDGPGANVQ 848 DPFAKGAKVA	LUPEHPWDVAFRIA-GW-I AFDPEHPWDLFLKAA-GWP0	DY G OR V GFPD	GIVIQGEMQISADPTKNA GTDTADGITVGSNPADKOV	VIVO-LD
ŭ	15 57 ari		Q8ZT36_PYRAE	770NDTLGLRVA	ALCRDAAWDVALLIGPGW-S	GGGNRIVYSI	NTYVDDAMSIKVAPN-NT	VAD-VP
SLF	- HB - HB		Q8NKS8_THELI	874 FPDGPGSNVI	DEPEHPWDVALRIA-GW-I	DYGNIIVLPE	GTSYQGEMKISADPVKNA	IVVE-VP
-	'		Q9Y8I8_THEHYb	1119_FPDGPGSNVF	RLDPNHPWDLALRIA-GW-I	DYGNLIILPR	GTAYQGEMQISADPVKNA	IVK-VP
-			Q9V294_PYRABa	878_FPDGPGSNVI	DLDPEHPWDVALRIA-GW-I	DY G NI IV LAN	IGTTYQ G EMKISAD P VKNR	IIVE-VP
<u>n</u>			Q9V294_PYRABD Q9HLU6 THEAC	1119_FPDGPGSNVI 1149 GNTYLGSGPN	JLDPEHPWDVALRIA-GW-I NAESSIPWQKMIYIS-GWAT	TYVQTLI	GIVITGEMKISADPIKNA. GTYSNG-ILVSVNLSLGEI	IIVE-VP IYVT-IP

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.

618

677

743

851

851

83

141

207

954

970

949

1041

1008

1074

1097

1156

1096

1155

1221

1191

1252

1317

1210

1269

1335

1466

1519

1583

1723

1782

1848

1872

1934

2005

1886

1948

2019

71

130

192

738

797

857

1010

1069

1135

1802

1861

1927

940 938

908

828

933

937

937 1178

1178

1205

1014 1084 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 S	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
B-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	059290 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	09F1V3 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 142	O9LBO9 ARTGO	GH-15	GH-L	Bacteria	1048	1	852-961		
PF1935 (amylopullulanase)	n.d.	Pvrococcus furiosus DSM3638	O8TZO1 PYRFU	GH-57	1.1	Archaea	985	1	840-957		
TTC1828	n.d.	Thermus thermophilus HB27	072GM0 THET2	GH-57	1.1	Bacteria	994	1	815-928		
PAE3454 (pullulanase)	n.d.	Pvrobaculum aerophilum IM2	O8ZT36 PYRAE	GH-57	1.1	Archaea	999	1	737-847		
Amylopullulanase	3.2.1.1/41	Thermococcus litoralis	08NKS8 THELI	GH-57	1.1	Archaea	1089	1	835-953		
Amylopullulanase	3.2.1.1/41	Thermococcus hydrothermalis	Q9Y8I8 THEHY	GH-57	1.1	Archaea	1337	2	839-957,	1080-1197	
PAB0122 (amylopullulanase)	n.d.	Pyrococcus abyssi GE5	Q9V294 PYRAB	GH-57	1.1	Archaea	1362	2	839-954,	1080-1198	
TA0129	n.d.	Thermoplasma acidophilum DSM1728	Q9HLU6_THEAC	GH-57		Archaea	1641	ī	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 (BEN-SON 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 ap-

plied (FELSENSTEIN, 1985) using the bootstrapping procedure; the number of bootstrap trials used was 1,000. The trees were drawn with the program Tree-View (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 starchbinding 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 abovementioned 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., ETWIL-LER, 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., SALAMI-TOU, S., BEGUIN, P., OHAYON, H., GOUNON, P., MA-TUSCHEK, 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 *Bacil*lus 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., QUEREL-LOU, J., RIPP, R., THIERRY, J.C., VAN DER OOST, J., WEIS-SENBACH, 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., DUCHI-RON, 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 thermo*sulfurigenes 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. Mannandegrading enzymes from *Cellulomonas fimi*. Appl. Environ. Microbiol. **65**: 2598–2605.
- SVENSSON, B., JENSEN, M.T., MORI, H., BAK-JENSEN, K.S., BRNSAGER, B., NIELSEN, P.K., KRAMHRFT, B., PRTORIUS-IBBA, M., NRHR, 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. CLUS-TAL 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