Glycoconjugate Journal 20, 435–447, 2004
© 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

MINI REVIEW

Genetic organization of chromosomal S-layer glycan biosynthesis loci of *Bacillaceae*

René Novotny*, Andreas Pfoestl*, Paul Messner and Christina Schäffer

University of Applied Life Sciences and Natural Resources, Center for NanoBiotechnology, Gregor-Mendel Strasse 33, A-1180 Wien, Austria

S-layer glycoproteins are cell surface glycoconjugates that have been identified in archaea and in bacteria. Usually, S-layer glycoproteins assemble into regular, crystalline arrays covering the entire bacterium. Our research focuses on thermophilic *Bacillaceae*, which are considered a suitable model system for studying bacterial glycosylation. During the past decade, investigations of S-layer glycoproteins dealt with the elucidation of the highly variable glycan structures by a combination of chemical degradation methods and nuclear magnetic resonance spectroscopy. It was only recently that the molecular characterization of the genes governing the formation of the S-layer glycoprotein glycan chains has been initiated. The S-layer glycosylation *(slg)* gene clusters of four of the 11 known S-layer glycan structures from members of the *Bacillaceae* have now been studied. The clusters are ∼16 to ∼25 kb in size and transcribed as polycistronic units. They include nucleotide sugar pathway genes that are arranged as operons, sugar transferase genes, glycan processing genes, and transporter genes. So far, the biochemical functions only of the genes required for nucleotide sugar biosynthesis have been demonstrated experimentally. The presence of insertion sequences and the decrease of the G + C content at the *slg* locus suggest that the investigated organisms have acquired their specific S-layer glycosylation potential by lateral gene transfer. In addition, S-layer protein glycosylation requires the participation of housekeeping genes that map outside the cluster. The gene encoding the respective S-layer target protein is transcribed monocistronically and independently of the *slg* cluster genes. Its chromosomal location is not necessarily in close vicinity to the *slg* gene cluster.
Published in 2004.

Keywords: **bacterial glycosylation, S-layer, glycosylation gene cluster, sugar nucleotides, S-layer gene, glycan biosynthesis**

Introduction to S-layer glycoproteins

In the past decade, prokaryotic glycobiology has been recognized as a very attractive area of research, with high emphasis being on glycoproteins. This is well reflected by the high number of recent review articles on that subject. In general, prokaryotic glycoprotein glycans are highly variable compounds with regard to both composition and structure, comparable to those found in eukaryotic organisms [1–5]. Among them, the most dominant role play glycosylated flagellae and pili from pathogenic bacteria [4,6,7] and archaea [8], as well as glycosylated surface layer (S-layer) proteins of bacteria [9,10].

S-layer glycoproteins constitute the outermost layer of many bacteria from the *Bacillaceae* family [2]. If present, the S-layer glycoproteins are the most abundant cellular proteins, with 20% of the total protein synthesis effort of the cell being devoted to S-layer protein biosynthesis. This ensures a complete coverage of the bacterium with a crystalline, regular S-layer glycoprotein lattice during all stages of the cell growth cycle. The glycosylation degree of S-layer proteins generally varies between two and 10% (w/w), yielding overall apparent molecular masses of the constituting protomers between 45 and 200 kDa on SDS-PA gels. Besides the positive color reaction with the periodic acid-Schiff reagent on SDS-PA gels, the S-layer glycan chains can be visualized directly by electron microscopy of high-pressure frozen and thin-sectioned bacterial cells. They appear as partly collapsed, filiform structures that serve as a coating for the cell. Thus, it is conceivable to assume that they are involved in cell surface phenomena; more precise functions of S-layer glycans from the *Bacillaceae* are not known so far. Usually, bacterial

To whom correspondence should be addressed: Dr. Christina Schäffer, University of Applied Life Sciences and Natural Resources, Center for NanoBiotechnology, Gregor-Mendel Strasse 33, A-1180 Wien, Austria. Tel.: +43-1-47654-2203; Fax: +43-1-4789112; E-mail: christina.schaeffer@boku.ac.at
*Both authors contributed equally to this work.

S-layer glycans are composed of individual repeating units and they comprise in total up to 100 sugars residues. The repeating unit is one of the most variable cell constituents. The variations are (i) in the types of sugars present, which include, besides a wide range of hexoses and *N*-acetylated amino sugars, rare residues such as quinovosamine, *N*-acetyl-D-fucosamine, D-*glycero*-D-*manno*-heptose, D-rhamnose, or D-fucose, (ii) in their arrangement within an either linear or branched repeating unit, and (iii) in the linkage between these units. Usually, these O-antigen-like polysaccharides are linked via a short core saccharide to a rather small number (present analyses indicate between one and two glycosylated amino acids per S-layer protomer) of distinct glycosylation sites on the S-layer protein. Glycosidic linkage units found among members of the *Bacillaceae* include those to serine and threonine (Gal→Thr/Ser, GalNAc→Thr/Ser) and to tyrosine (Glc→Tyr, Gal→Tyr), with the linkage sugar being in the β-configuration. S-layer glycans usually possess a tripartite structure reminiscent of that known from the lipopolysaccharides of Gram-negative bacteria. It comprises an O-antigen-like polysaccharide chain, whose length is clustered around a modal value [11], a core region and the glycosylation site replacing lipid A (details about the structural schemes of S-layer glycoprotein are reviewed in [2]). Despite these possible variations within S-layer O-glycan chains, only one type of glycan has been detected per individual organism so far, but this may be attached to different amino acids of the S-layer protein [11,12]. Please note that in archaea, different types of glycans may be linked to an S-layer protein, but the composition of these archaeal glycan chains is usually different from bacterial S-layer glycans and they do not follow the above mentioned tripartite structural scheme (reviewed [13]). However, within the *Bacillaceae* family, structural diversity may be quite extensive even among closely related organisms, as is exemplified by *Aneurinibacillus thermoaerophilus* strains L420-91^T and DSM 10155 (compare with [14]). Thus, S-layer glycan chains cannot serve as a criterion for taxonomic typing of bacteria. However, they are useful tools for typing strains. It is interesting to note that the S-layer glycans present in fresh isolates may be lost after prolonged cultivation of the bacteria in rich laboratory media (*e.g.*, S-layer glycan deficient variant of *A. thermoaerophilus* DSM 10155/G$^-$ [15]; GenBank accession number of the gene encoding the non-glycosylated S-layer protein SatC, AY422725). This observation supports the assumption that the carbohydrates provide a selection advantage for the bacteria under the competitive conditions of the natural habitat.

Whereas S-layer glycoprotein research has previously focused on structural investigations of S-layer glycan chains from different bacterial species (for a compilation of S-layer glycan structures see [2,16]), recent studies have focussed on the genes governing the S-layer glycan formation [17]. From the preliminary data obtained from selected organisms, an even more interesting picture of S-layer glycoproteins is emerging. The long-term goal of S-layer glycoprotein research is to provide a detailed understanding of the mechanisms underlying the biosynthesis of S-layer glycoprotein glycans and the regulatory events that coordinate S-layer glycan and S-layer protein biosyntheses. This knowledge shall eventually allow the rational design of S-layer glycosylation motives by carbohydrate engineering techniques for various types of nanobiotechnology applications [10,16].

Description of S-layer glycosylation *(slg)* gene clusters

Analysis of S-layer glycosylation on the molecular level was performed on the basis of completely elucidated S-layer glycan structures. Currently, most detailed data are available from the organisms *Aneurinibacillus thermoaerophilus* strains L420-91^T and DSM 10155/G$^+$ and *Geobacillus stearothermophilus* NRS 2004/3a. Recently, we have extended our studies to *Thermoanaerobacterium thermosaccharolyticum* E207-71. Figure 1 shows the wide variations of the S-layer glycan structures among the selected organisms. The most simple S-layer glycan is found in *G. stearothermophilus* NRS 2004/3a, representing a linear poly-L-rhamnan made of trisaccharide repeats [11,18]; *A. thermoaerophilus* DSM 10155/G$^+$ possesses disaccharide repeating units containing the rare constituent D-*glycero*-D-*manno*-heptose [12,19]. Branched repeating units are present in *A. thermoaerophilus* L420-91^T [20,21] and *T. thermosaccharolyticum* E207-71 [22], with the hexasaccharide of the latter organism containing even a furanosidic sugar (Gal*f*). The presence of either L-rhamnose or, less often, D-rhamnose residues is frequently observed in S-layer glycans and is also valid for the investigated organisms (compare with Figure 1).

Based on the knowledge that most sugars are incorporated into growing glycan chains from their nucleotide-activated precursor, we surveyed the literature for what is known about the deoxy-thymidine diphosphate (dTDP)-β-L-rhamnose and the guanosine diphosphate (GDP)-α-D-rhamnose biosynthesis pathways as utilized for LPS assembly. GDP-α-D-rhamnose is synthesized in a two-step reaction catalyzed by the Gmd and Rmd enzymes [23], whereas four enzymes, encoded by the *rml*ABCD genes, act sequentially to synthesize dTDP-β-L-rhamnose (reviewed in [24]). Degenerate oligonucleotide probes targeted to the initial step of the respective pathway were used to localize the respective DNA sequence on the chromosome of the investigated organisms. For *A. thermoaerophilus* L420-91^T, the highly conserved eight-amino acid stretch GILFNHES of the GDP-D-mannose-dehydratase Gmd, which has been found in 16 out of 18 aligned Gmd sequences from database entries, was used for primer design and it gave highly specific signals in Southern hybridization experiments with *Eco*RI-digested DNA of the organism [25]. Further sequence information was obtained from a pBCKS-based construct, which confirmed that the GDP-α-D-rhamnose operon consisted only of the two ORFs Gmd and Rmd also in the S-layer carrying organism. For the identification of genes from

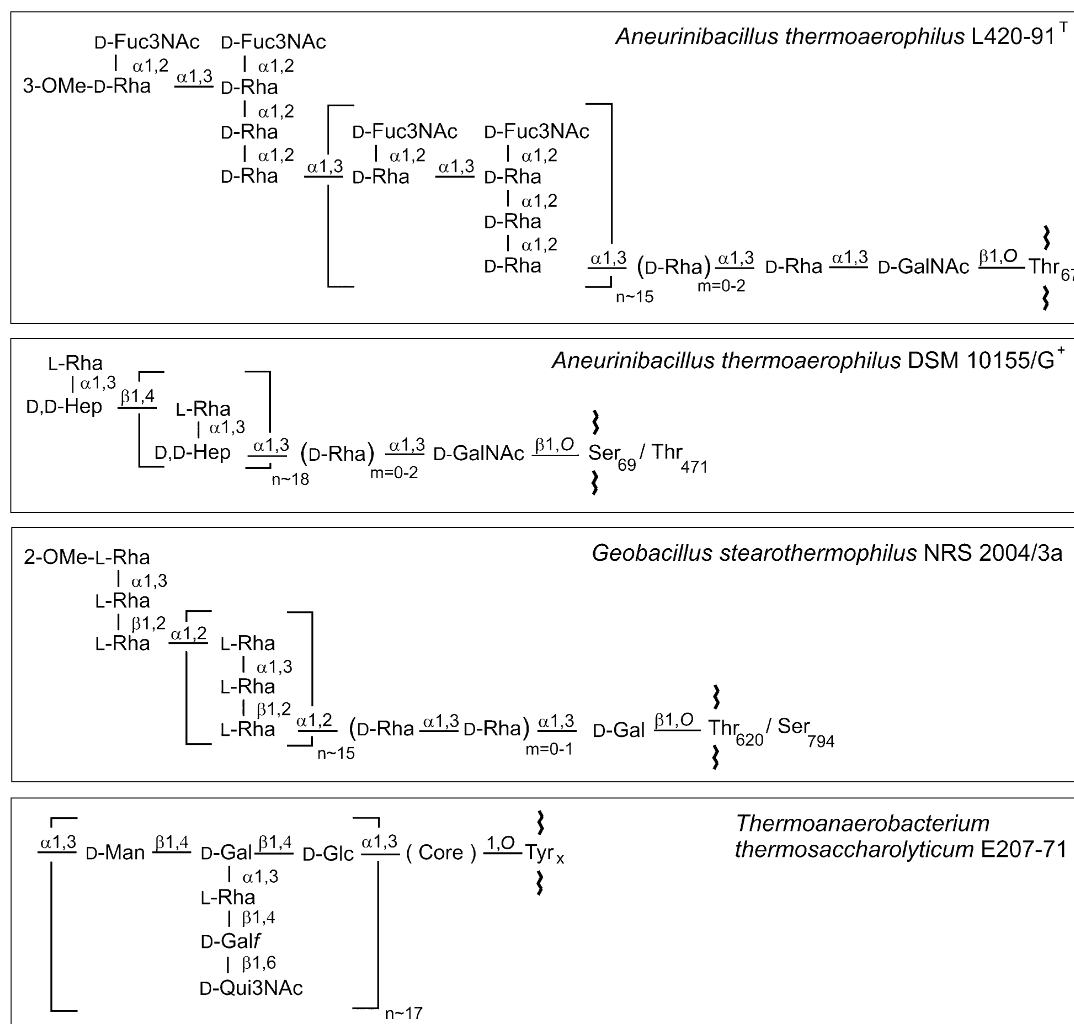

Figure 1. S-layer glycoprotein glycan structures of the investigated organisms. The lower case numbers indicate the position of the glycosylated amino acid on the respective S-layer precursor protein. The curved line symbolizes the S-layer polypeptide. For *T. thermosaccharolyticum* E207-71 the core structure has not yet been determined.

the dTDP-β-L-rhamnose biosynthesis, the highly conserved seven-amino acid stretch YDKPMIY of RmlA, and the six-amino acid stretch TDEVYG of the RmlB protein were used for primer design [17]. All four organisms included in this study gave a positive signal in a PCR approach using the degenerate primers. Sequencing of adjacent regions by chromosome walking [26] confirmed the presence of the complete *rml* operon in *A. thermoaerophilus* DSM 10155/G$^+$, *G. stearothermophilus* NRS 2004/3a, and *T. thermosaccharolyticum* E207-71. In the former organisms the *rml* genes are arranged in the order *rml*ACBD, in agreement with the order already described for other Gram-positive bacteria [27,28], whereas in *T. thermosaccharolyticum* E207-71 the order of the genes is *rml*BADC. This finding indicates that, as in Gram-negative bacteria [29–31], the order of the four genes is quite variable. In addition, the *rml*AB-specific probes gave also a positive result for *A. thermoaerophilus* L420-91^T, where these genes are responsible for the initial steps of the dTDP-α-D-Fuc3NAc biosynthesis [32] (compare with Figure 2).

Sequencing of upstream and downstream regions of the rhamnose biosynthesis operons revealed the presence of chromosomal S-layer glycan biosynthesis (*slg*) gene clusters in *A. thermoaerophilus* strains L420-91^T (GenBank accession number AY442352) and DSM 10155/G$^+$ (AF324836) and in *G. stearothermophilus* NRS 2004/3a (AF328862). For comparative reasons we have also included in this survey the incomplete *slg* gene locus of *T. thermosaccharolyticum* E207-71 (AY422724) (Figure 3). The clusters vary strain-specifically in size due to the differences in the mature S-layer glycan structures (preliminary data indicate DNA regions extending between 16 and 25 kb) and they comprise closely spaced or even overlapping open reading frames (ORFs), all of which are transcribed in the same direction. If present, putative transposases or fragments thereof may be transcribed in the opposite direction

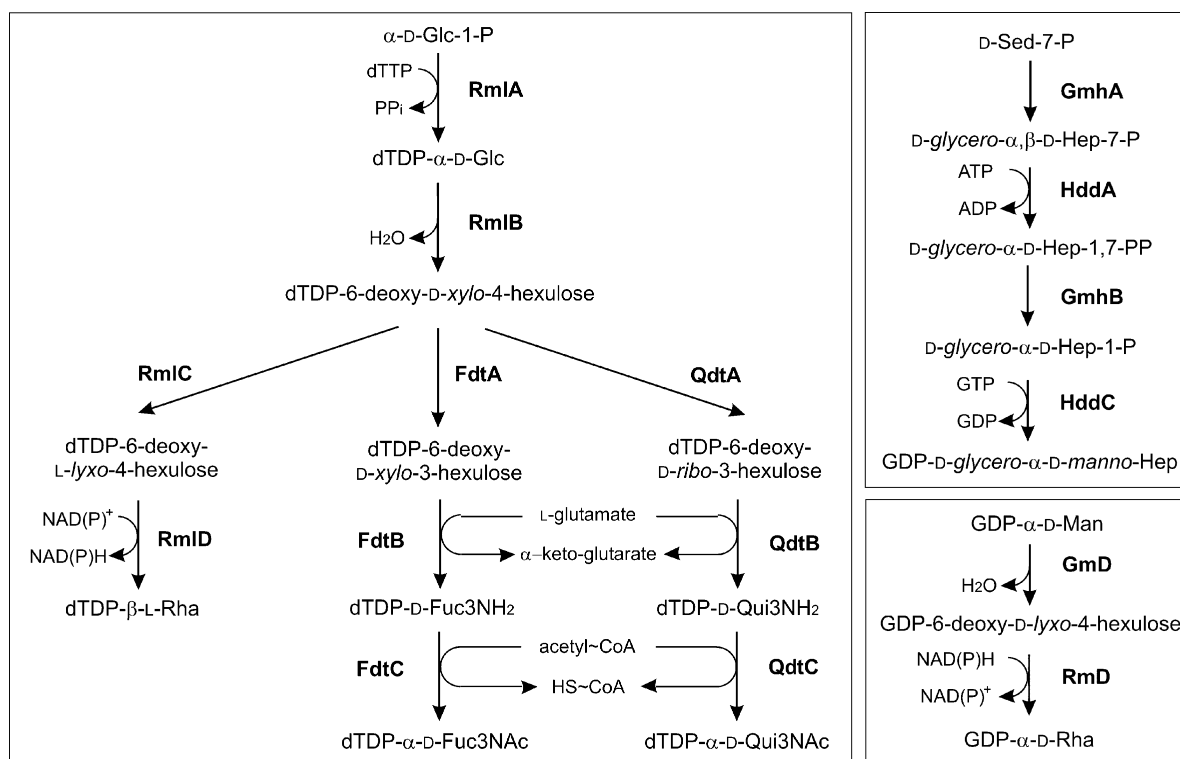

Figure 2. Overview of nucleotide sugar biosynthetic pathways involved in S-layer protein glycosylation. Enzyme names are in boldface. All reaction steps have been demonstrated experimentally in *in vitro* assays using purified enzyme preparations.

(compare with Figures 3a and b). Most of the putative gene products encoded by the assigned ORFs show high homology with proteins involved in the biosynthesis of different bacterial surface polysaccharides, which supports the assumption that these DNA loci indeed encode the S-layer protein glycosylation event in the investigated organisms. Based on these sequence similarities, putative biological functions could be assigned to most of the genes of the *slg* gene clusters (Tables 1a–d).

The *slg* gene clusters include nucleotide sugar pathway genes, which are usually clustered in an operon, glycosyltransferase genes, glycan processing genes (*e.g.*, a putative methyltransferase, which may also be involved in chain termination), and transporter genes. The *A. thermoaerophilus* L420-91^T and *G. stearothermophilus* NRS 2004/3a *slg* gene cluster additionally contains genes involved in the formation of a putative lipid-bound glycan intermediate. From the assigned genes it is evident that none of the *slg* gene clusters encodes the biosynthesis of the nucleotide-activated linkage sugar of the S-layer glycan (UDP-Gal and UDP-GalNAc, respectively). Obviously, these precursors that are synthesized from UDP-Glc and UDP-GlcNAc, respectively, through the activity of the corresponding 4-epimerases GalE [33] and Gne [34], are also involved in housekeeping functions and not duplicated in the *slg* gene cluster. The current impossibility to transform any of the *A. thermoaerophilus* strains and *G. stearothermophilus* NRS 2004/3a has so far impeded the experimental proof for the completeness of the *slg* gene clusters with regard to S-layer glycan-specific information. It is interesting to note that the number of transferase-like genes present in the clusters matches the theoretically required number for building up the respective S-layer glycan, not considering the possibility of multispecificity of single enzymes. This, in combination with the position of putative transposases may be taken as an indication that the sequenced regions indeed encode the complete information for glycosylating the target S-layer protein. However, only knock-out mutants and complementation experiments will eventually give the final proof for this assumption.

From all preliminary data available it is evident that the *slg* gene clusters of *Bacillaceae* are much less organized than the clusters encoding the biosynthesis of other bacterial polysaccharides, such as the LPS O-antigens of Gram-negative bacteria [35] or the exopolysaccharides of lactic acid bacteria [36]. Furthermore, the current sequence information did not allow the identification of specific genes on the chromosome of the *Bacillaceae*, such as the *galF* and *gnd* genes in *Escherichia coli* and *Salmonella enterica* [35] or the *hem*H and *gsk* genes in *Yersinia enterocolitica* [37], between which the *slg* locus is preferentially located. Another difference between LPS and *slg* gene clusters may be the absence of a typical JUMPstart sequence that is found in proximity to probable promoter regions of several polysaccharide gene clusters [38]. At least in the examined upstream region of the *slg* gene clusters, which was extending ~2.5 kb of *A. thermoaerophilus* strain L420-91^T and DSM 10155/G$^+$, and ~4 kb in the case of

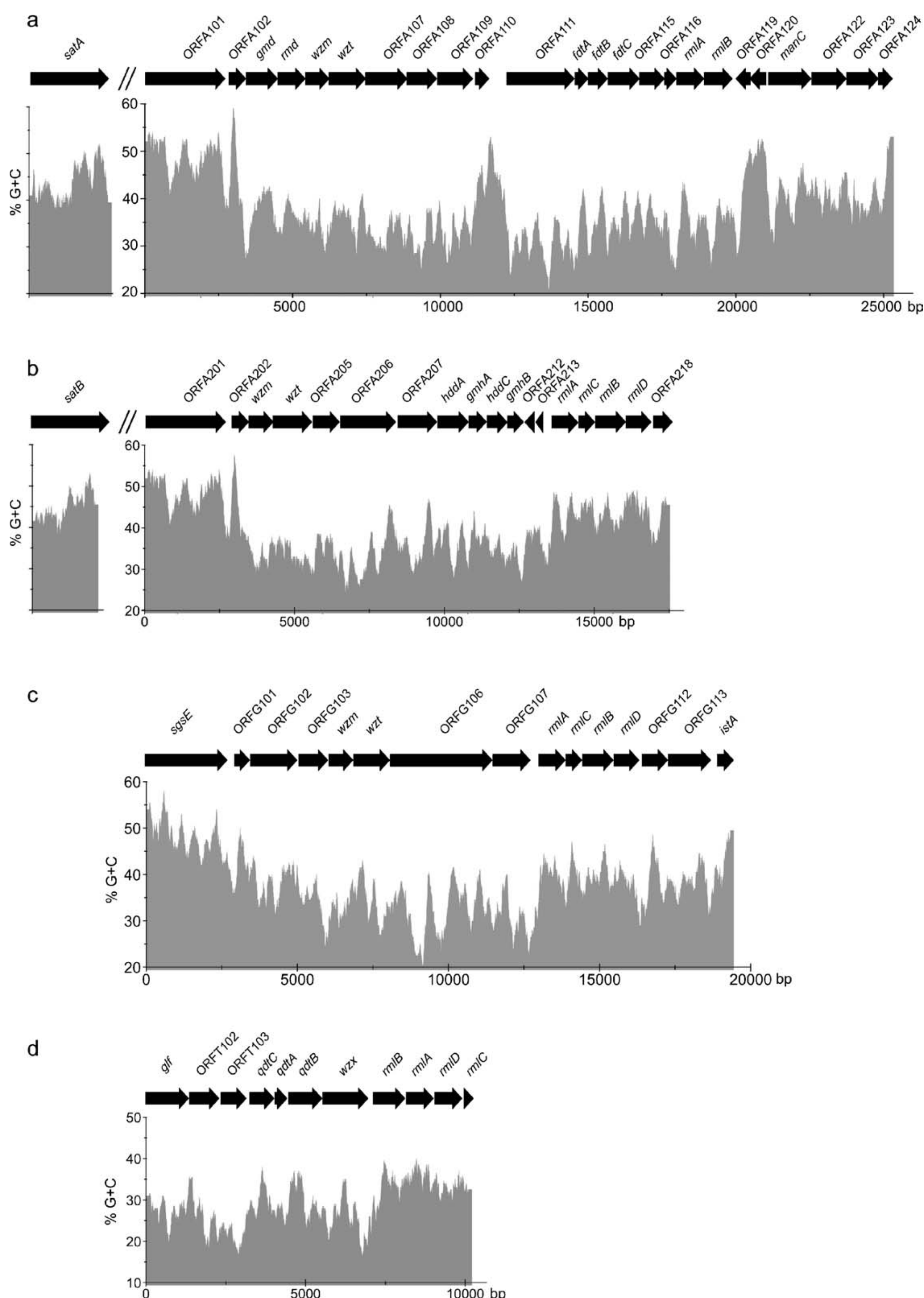

Figure 3. Genetic organization of the chromosomal *slg* gene cluster of (a) *Aneurinibacillus thermoaerophilus* L420-91^{T} (GenBank accession number AY442352), (b) *A. thermoaerophilus* DSM 10155/G^{+} (GenBank accession number AF324836), (c) *Geobacillus stearothermophilus* NRS 2004/3a (GenBank accession number AF328862), and (d) *Thermoanaerobacterium thermosaccharolyticum* E207-71 (incomplete cluster, GenBank accession number AY422724). The corresponding percentage G+C base composition is given below each cluster map. The position of the S-layer genes *satA* and *satB* in relation to the respective *slg* gene cluster is shown on the left side of the figure.

Table 1. Database homologies of the open reading frames contained in the *slg* gene clusters of the investigated thermophilic *Bacillaceae*

Description of ORF					Related proteins					
ORF	*aa*[1]	*MW*[2] *(kDa)*	*G+C (%)*	*TM reg.*[3]	*Name*	*Organism*	*Accession no.*	*aa*[1]	Function	*Identity/ similarity*[4]
(a) Description of the *slg* gene cluster of *Aneurinibacillus thermoaerophilus* L420-91^T (GenBank accession number AY442352)										
ORFA101	883	97.0	48.4	–	Chte2135	*Clostridium thermocellum*	ZP_00061713	710	Hypothetical protein	35/53 (38-213)
					not annotated	*Thermoanaerobacter tengcongensis*	AAM23420	192	S-layer homology domain	28/55 (15-139)
ORFA102	188	21.2	44.0	–	Ddes0174	*Desulfovibrio desulfuricans*	ZP_00128569	176	Hypothetical protein	47/65 (11-152)
					not annotated	*Legionella longbeachae*	AAK00281	192	Unknown function	28/55 (15-139)
Gmd	341	39.0	37.0	–	Gmd	*Pseudomonas aeruginosa*	AAG08828	319	GDP-mannose-4,6-dehydratase	61/75 (2-319)
					LpsA	*Caulobacter vibrioides*	AAC38668	325	GDP-mannose-4,6-dehydratase	63/77 (2-319)
Rmd	309	34.5	35.9	–	Rmd	*Pseudomonas aeruginosa*	AAG08838	304	Oxidoreductase	33/55 (2-306)
					WcbK	*Burkholderia pseudomallei*	AF228583	337	Unknown function	35/54 (2-307)
Wzm	261	30.5	33.2	6	Wzm	*Klebsiella pneumoniae*	BAA28343	261	ABC-2 type transport system integral membrane protein	45/61 (1-261)
					Wzm	*Xanthomonas campestris*	AAK53482	261	ABC-2 type transport system integral membrane protein	36/58 (1-261)
Wzt	408	45.7	35.4	–	Wzt	*Escherichia coli*	BAA28325	404	ABC-2 type transport system ATP-binding protein	55/73 (1-403)
					Wzt	*Pseudomonas aeruginosa*	AAG08835	421	ABC-2 type transport system ATP-binding protein	50/69 (1-404)
ORFA107	467	55.4	32.0	–	not annotated	*Methanosarcina acetivorans*	NP_616119	466	Hypothetical protein	46/63 (242-467)
					CT2121	*Chlorobium tepidum*	AAM73337	241	Putative methyltransferase	30/44 (247-376)
ORFA108	353	41.2	31.0	–	not annotated	*Leptospira interrogans*	AAN47431	349	Putative glycosyl transferase	33/54 (3-347)
					Meth1435	*Methanosarcina barkeri*	ZP_00076850	351	Hypothetical protein	40/64 (3-350)
ORFA109	393	46.4	32.2	1	not annotated	*Methanosarcina acetivorans*	AAM04601	388	Hexosyltransferase	37/57 (4-392)
					not annotated	*Leptospira interrogans*	AAN47430	444	Putative glycosyl transferase	27/46 (3-392)
ORFA110	103	11.94	41.1	–	TpnA	*Ralstonia metallidurans*	CAB50739	411	Putative transposase	33/58 (3-89)
					Chte1727	*Clostridium thermocellum*	ZP_00061317	207	Hypothetical protein	37/63 (2-85)
ORFA111	746	86.2	29.4	10	WxcO	*Xanthomonas campestris*	AAK53474	750	Integral membrane protein	31/49 (55-600)
FdtA	139	16.0	34.8	–	WxcM	*Xanthomonas campestris*	AAM39931	309	Bifunctional acetyl transferase/isomerase	37/60 (4-127)
					WbsB	*Listonella anguillarum*	AAB81630	132	Unknown function	57/74 (1-129)
					WbsB	*Escherichia coli*	AAK60451	134	Unknown function	50/67 (1-133)
FdtC	192	21.0	35.1	–	WxcM	*Xanthomonas campestris*	AAM39931	309	Bifunctional acetyl transferase/isomerase	61/77 (9-158)
					Orf16x2	*Listonella anguillarum*	AAB81631	151	Putative acetyltransferase	59/75 (9-153)
FdtB	363	41.0	35.4	–	orf41x4	*Listonella anguillarum*	AAB81632	371	Putative aminotransferase	62/80 (1-364)
					not annotated	*Campylobacter jejuni*	AAK12954	360	Putative aminotransferase	66/77 (4-346)
ORFA115	310	35.8	34.4	2	YkcC	*Bacillus subtilis*	CAA05569	323	Putative dolichol phosphate mannose synthetase	35/57 (1-309)
					GtrB	*Salmonella enterica*	AAO69903	308	Bactoprenol glucosyltransferase	31/53 (6-305)

ORFA116	119	13.7	25.8	3	not annotated	*Pseudomonas syringae*	AAO54598	132	Putative membrane protein	40/70 (1-114)
					WxcN	*Xanthomonas campestris*	AAM39930	151	Sugar translocase	37/61 (7-117)
RmlA	305	33.8	36.0	–	RmlA	*Aneurinibacillus thermoaerophilus*	AAK27854	296	Glucose 1-phosphate thymidyltransferase	81/90 (1-288)
					RmlA	*Salmonella enterica*	AAG09504	292	Glucose 1-phosphate thymidyltransferase	70/87 (2-288)
RmlB	343	39.7	33.7	–	RmlB	*Xanthomonas campestris*	AAM39937	351	dTDP-glucose 4,6-dehydratase	51/65 (11-339)
					RmlB	*Pseudomonas aeruginosa*	AAG08546	352	dTDP-glucose 4,6-dehydratase	51/65 (9-340)
ORFA119	132	15.7	47.0	–	BH2520	*Bacillus halodurans*	BAB06239	188	Transposase	84/91 (1-132)
ORFA120	172	19.7	47.5	–	BH2521	*Bacillus halodurans*	BAB06240	173	Transposase	58/71 (7-170)
					ORFU	*Bacillus stearothermophilus*	CAA79750	160	Transposase	56/69 (4-159)
ManC	464	52.8	30.1	–	RfbM	*Salmonella enterica*	AAO68480	479	Mannose-1-phosphate guanylyltransferase	29/48 (4-444)
					Tlr0018	*Thermosynechococcus elongatus*	BAC07571	498	Mannose-1-phosphate guanylyltransferase	29/29 (4-443)
ORFA122	389	44.9	30.4	–	WbdB	*Klebsiella pneumoniae*	AAF04384	381	Mannosyltransferase	31/52 (16-388)
					not annotated	*Clostridium tetani*	AAO34916	384	Mannosyltransferase	35/58 (124-389)
ORFA123	377	43.1	27.1	–	WbpZ	*Pseudomonas aeruginosa*	AAG08832	381	Glycosyltransferase	34/54 (1-373)
					WbdC	*Escherichia coli*	BAA28303	274	Mannosyltransferase	39/57 (109-373)
ORFA124 (fragment)	111	12.6	48.3	–	DRA0039	*Deinococcus radiodurans*	AAF12270	343	Mannosyltransferase	49/66 (4-91)
						Pseudomonas syringae	AAO56922	369	Glycosyltransferase	32/50 (4-91)
(b) Description of the *slg* gene cluster of *Aneurinibacillus thermoaerophilus* DSM 10155/G⁺ (GenBank accession number AF324836)										
ORFA201	883	97.0	48.2	1	Chte2135	*Clostridium thermocellum*	ZP_00061713	710	Hypothetical protein	35/53 (38-213)
					not annotated	*Thermoanaerobacter tengcongensis*	AAM23420	408	S-layer homology domain	29/47 (22-293)
ORFA202	188	21.1	44.0	–	Ddes0174	*Desulfovibrio desulfuricans*	ZP_00128569	176	Hypothetical protein	47/65 (11-152)
					not annotated	*Legionella longbeachae*	AAK00281	192	Unknown function	28/55 (15-139)
Wzm	267	31.4	32.3	6	Wzm	*Actinobacillus actinomycetemcomitans*	BAA82536	264	ABC-2 type transport system integral membrane protein	57/79 (5-267)
					Wzm	*Enterococcus faecalis*	AAO81915	264	ABC-2 type transport system integral membrane protein	42/64 (1-267)
Wzt	435	49.4	33.6	–	Wzt	*Pseudomonas syringae*	AAO54604	454	ABC-2 type transport system ATP-binding protein	38/58 (1-426)
					Wzt	*Enterococcus faecalis*	AO81914	405	ABC-2 type transport system ATP-binding protein	44/65 (7-355)
ORFA205	303	34.7	33.8	–	PSPTO1074	*Pseudomonas syringae*	AAO54603	1561	Glycosyltransferase	39/58 (19-247)
					RHMO02586	*Heliobacillus mobilis*	AAN87470	254	Methyltransferase	21/41 (11-235)
ORFA206	618	72.4	32.4	–	XCC2933	*Xanthomonas campestris*	AAM42205	700	Glycosyltransferase	41/61 (442-618)
					XAC3110	*Xanthomonas axonopodis*	AAM37955	732	Glycosyltransferase	38/56 (415-617)
ORFA207	438	50.9	34.8	–	XCC2933	*Xanthomonas campestris*	AAM42205	700	Glycosyltransferase	35/53 (1-433)
					XAC3110	*Xanthomonas axonopodis*	AAM37955	732	Glycosyltransferase	34/54 (1-434)
HddA	341	38.1	36.5	–	WcbL	*Burkholderia mallei*	AAK26467	346	Putative sugar kinase	52/71 (2-341)
					Cj1425c	*Campylobacter jejuni*	CAB73849	339	Putative sugar kinase	51/67 (4-341)
GmhA	198	21.5	36.0	–	GmhA2	*Campylobacter jejuni*	CAB73848	201	Putative phosphoheptose isomerase	65/78 (1-194)
					CAC3054	*Clostridium acetobutylicum*	AAK80994	196	Phosphoheptose isomerase	46/66 (1-193)

(*Continued on next page.*)

Table 1. (*Continued*).

Description of ORF					*Related proteins*					
ORF	*aa*[1]	*MW*[2] *(kDa)*	*G+C (%)*	*TM reg.*[3]	*Name*	*Organism*	*Accession no.*	*aa*[1]	Function	*Identity/ similarity*[4]
HddC	230	26.6	35.1	–	WcbM	*Burkholderia mallei*	AAK26469	230	Putative sugar-phosphate nucleotidyltransferase	39/60 (2-226)
					CAC3056	*Clostridium acetobutylicum*	AAK80996	234	Nucleoside-diphosphate-sugar pyrophosphorylase	45/62 (1-221)
GmhB	179	20.8	31.1	–	WcbN	*Burkholderia mallei*	AAK26470	189	Putative phosphatase	61/74 (1-152)
					VC0908	*Vibrio cholerae*	AAF94070	186	Histidinol phosphatase-related protein	47/64 (1-157)
ORFA212	114	13.7	35.7	–	Trp	*Clostridium cellulovorans*	AAF61311	313	Putative transposase fragment	27/43 (8-114)
ORFA213	78	8.9	38.9	–	TNPACP2	*Citrobacter freundii*	AAN87699	456	Putative transposase fragment	32/58 (24-70)
RmlA	296	32.9	43.0	–	RmlA	*Salmonella enterica*	AAG09524	292	Glucose-1-phosphate thymidylyltransferase	62/74 (2-285)
					RmlA	*Escherichia coli*	AAC75100	293	Glucose-1-phosphate thymidylyltransferase	62/73 (2-289)
RmlC	182	20.7	44.0	–	RmlC	*Vibrio cholerae*	AAO88950	183	dTDP-6-deoxy-D-glucose-3,5-epimerase	53/67 (4-180)
					RmlC	*Salmonella typhimurium*	AAL20998	183	dTDP-6-deoxy-D-glucose-3,5-epimerase	54/66 (4-177)
RmlB	341	39.1	41.4	–	RmlB	*Clostridium acetobutylicum*	AAK80288	351	dTDP-glucose 4,6-dehydratase	60/77 (4-335)
					RmlB	*Xanthomonas campestris*	AAM39937	351	dTDP-glucose 4,6-dehydratase	56/69 (4-325)
RmlD	282	31.7	45.2	–	RmlD	*Bacillus halodurans*	BAB07084	283	dTDP-4-dehydrorhamnose reductase	43/57 (3-278)
					RmlD	*Actinobacillus actinomycetemcomitans*	AAG49405	292	dTDP-4-dehydrorhamnose reductase	43/61 (2-273)
ORFA218 (partial)	295	23.2	41.0	(-)	WbbT	*Yersinia enterocolitica*	CAA79348	312	Glycosyltransferase	43/60 (5-200))
					RgpB	*Lactococcus lactis*	AAK04299	319	Rhamnosyltransferase	32/54 (4-199)
(c) Description of the *slg* gene cluster of *Geobacillus stearothermophilus* NRS 2004/3a+ (GenBank accession number AF328862)										
ORFG101	169	19.9	43	1	AQ_854	*Aquifex aeolicus*	AAC06984	545	TPR-domain containing protein	26/50 (46-165)
					YbgF	*Salmonella typhimurium*	AAL19694	262	TPR-domain containing protein	28/48 (48-150)
ORFG102	526	59.9	37.4	12	WaaL	*Edwardsiella tarda*	AAL01247	377	Putative lipid A core: surface polymer ligase	28/52 (378-466)
					Wzy	*Bacteroides fragilis*	AAD56734	478	Putative polymerase	22/41 (148-366)
ORFG103	324	38.2	33.0	1	RgpB	*Lactococcus lactis*	AAK04299	319	Putative rhamnosyltransferase	33/51 (6-227)
					RgpBc	*Streptococcus mutans*	BAA32090	311	Putative rhamnosyltransferase	29/52 (6-223)
Wzm	268	30.4	33.0	6	Wzm	*Serratia marcescens*	AAC00181	277	ABC-2 type transport system integral membrane protein	45/67 (7-268)
					Wzm	*Klebsiella pneumoniae*	AAN06492	277	ABC-2 type transport system integral membrane protein	43/65 (4-268)
Wzt	409	46.2	33.7	–	Wzt	*Serratia marcescens*	AAC00182	441	ABC-2 type transport system ATP-binding protein	44/62 (2-400)

					Wzt	*Klebsiella pneumoniae*	AAN06493	440	ABC-2 type transport system ATP-binding protein	46/61 (2-369)
ORFG106	1127	132.6	31.8	–	MenG	*Methanosarcina acetivorans*	AAM07690	179	Menaquinone biosynthesis methyltransferase	40/57 (75-141)
					ExpE3	*Sinorhizobium meliloti*	CAB01936	281	Putative methyltransferase	32/54 (62-120)
					Unnamed	*Enterococcus faecalis*	AAO81912	713	Putative glycosyl transferase	60/74 (604-765), 42/62 (863-1060)
					RfbC	*Myxococcus xanthus*	AAB05019	1275	Putative glycosyl transferase	38/59 (604-765), 39/64 (863-1060)
ORFG107	413	48.3	30.0	1	RfbC	*Myxoxoccus xanthus*	AAB05019	1275	Putative glycosyl transferase	30/46 (134-344)
					WbbX	*Yersinia enterocolitica*	CAA79354	421	Unknown	27/42 (137-355)
RmlA	299	33.2	39.6	–	RmlA	*Aneurinibacillus thermoaerophilus*	AAK27854	296	Glucose 1-phosphate thymidyltransferase	75/84 (1-288)
					RmlA	*Enterococcus faecalis*	AAC35920	288	Glucose 1-phosphate thymidyltransferase	69/78 (1-286)
RmlC	183	21.3	38.4	–	RmlC	*Aneurinibacillus thermoaerophilus*	AAL18012	182	dTDP-dehydrorhamnose 3,5-epimerase	74/81 (4-181)
					RmlC	*Bacillus anthracis*	AAP25187	181	dTDP-dehydrorhamnose 3,5-epimerase	62/79 (1-181)
RmlB	342	38.8	38.4	–	RmlB	*Aneurinibacillus thermoaerophilus*	AAL18013	341	dTDP-glucose 4,6-dehydratase	73/84 (3-338)
					RmlB	*Lactobacillus gasseri*	AAL91481	345	dTDP-glucose 4,6-dehydratase	61/74 (3-337)
RmlD	282	31.9	38.5	–	RmlD	*Aneurinibacillus thermoaerophilus*	AAL18014	282	dTDP-dehydrorhamnose reductase	54/72 (1-277)
					RmlD	*Clostridium acetobutylicum*	AAK80271	280	dTDP-dehydrorhamnose reductase	48/66 (1-276)
ORFG112	289	33.4	37.0	1	RfbQ	*Clostridium acetobutylicum*	AAK80268	303	Putative rhamnosyltransferase	29/55 (3-263)
					WbbL	*Mycobacterium smegmatis*	AAF04376	296	Putative rhamnosyltransferase	25/48 (5-255)
ORFG113	471	54.5	37.9	5	RfbP2	*Leptospira interrogans*	AAN49708	473	Putative UDP-glucose lipid carrier transferase	39/61 (4-471)
					WbaP	*Escherichia coli*	AAD21565	476	UDP-Gal::undP Gal-1-P transferase	31/47 (17-471)
					RfaP	*Salmonella enterica*	AAC44096	476	Galactosyl-1-phosphate transferase	31/49 (78-471)
IstA	184	21.7	42.2	–	ORFA	*Geobacillus stearothermophilus*	CAA48045	400	Transposable element	89/90 (1-184)
					MM2694	*Methanosarcina mazei*	AAM32390	229	Transposase	41/65 (1-184)
(d) Description of the incomplete *slg* gene cluster of *Thermoanaerobacterium thermosaccharolyticum* E207-71 (GenBank accession number AY422724)										
Glf	372	44.6	27.1	–	MTH344	*Methanothermobacter thermautotrophicus*	AAB84850	380	UDP-galactopyranose mutase	49/70 (3-362)
					CAC2169	*Clostridium acetobutylicum*	AAK80127	360	UDP-galactopyranose mutase	51/68 (3-362)
ORFT102	321	37.5	26.6	–	MTH343	*Methanothermobacter thermautotrophicus*	AAB84849	328	Rhamnosyltransferase	41/58 (7-296)
					WcfN	*Bacteroides fragilis*	AAK68916	291	Putative glycosyltransferase	29/51 (7-294)
ORFT103	278	33.6	20.7	–	Desu2509	*Desulfitobacterium hafniense*	ZP_00099369	235	Glycosyltransferases	33/56 (1-198)
					XAC1693	*Xanthomonas axonopodis*	AAM36560	316	Glycosyltransferase	28/49 (1-201)
QdtC	265	29.7	29.2	–	FdtC	*Aneurinibacillus thermoaerophilus*	AAO06352	192	dTDP-D-Fucp3N acetylase	39/56 (76-219)

(*Continued on next page.*)

Table 1. (*Continued*).

Description of ORF					Related proteins					
ORF	*aa*[1]	*MW*[2] *(kDa)*	*G+C (%)*	*TM reg.*[3]	*Name*	*Organism*	*Accession no.*	*aa*[1]	Function	*Identity/ similarity*[4]
					TM0759	*Thermotoga maritima*	AAD35841	254	Putative acyltransferase	35/51 (5-221)
QdtA	136	16.0	26.5	–	FdtA	*Aneurinibacillus thermoaerophilus*	AAO06351	139	dTDP-6-deoxy-3,4-keto-hexulose isomerase	55/69 (7-126)
					WbsB	*Escherichia coli*	AAK60451	134	Unknown	48/64 (4-128)
QdtB	365	41.5	29.4	–	FdtB	*Aneurinibacillus thermoaerophilus*	AAO06353	363	dTDP-6-deoxy-D-xylo-hex-3-ulose aminase	60/74 (31-345)
					XCC1671	*Xanthomonas campestris*	AAM40965	380	Aminotransferase	53/72 (31-358)
Wzx	491	55.8	25.3	14	Wzx	*Streptococcus thermophilus*	AAL32506	473	Putative protein	21/41 (1-432)
					CTC02254	*Clostridium tetani*	AAO36736	488	Transporter	22/41 (5-410)
RmlB	351	40.7	32.2	–	RmlB	*Clostridium acetobutylicum*	AAK80288	351	dTDP-D-glucose 4,6-dehydratase	80/91 (1-351)
					RmlB	*Pseudomonas aeruginosa*	CAC82202	352	dTDP-D-glucose 4,6-dehydratase	51/67 (3-349)
RmlA	302	33.6	35.0	–	RmlA	*Vibrio cholerae*	AAO88922	292	Glucose-1-phosphate thymidylyltransferase	64/87 (1-289)
					RmlA	*Actinobacillus actinomycetemcomitans*	BAA82533	290	Glucose-1-phosphate thymidylyltransferase	65/78 (1-289)
RmlD	294	33.5	32.8	–	RmlD	*Aneurinibacillus thermoaerophilus*	AAK76867A	239	dTDP-dehydrorhamnose reductase	44/62 (1-288)
					RmlD	*Clostridium acetobutylicum*	AL18014	288	dTDP-4-keto-L-rhamnose reductase	58/77 (55-289)
RmlC (partial)	(83)	(9.7)	(32.0)	(–)	RmlC	*Clostridium acetobutylicum*	AAK80287	185	dTDP-4-dehydrorhamnose 3,5-epimerase	84/92 (1-83)
					RmlC	*Methanobacterium thermoautotrophicum*	AAB86256	185	dTDP-4-dehydrorhamnose 3,5-epimerase	58/78 (1-83)

[1]Number of amino acid residues.
[2]Calculated molecular masses.
[3]Putative transmembrane regions.
[4]The locations of the homology in the respective ORF are indicated in parentheses.

G. stearothermophilus NRS 2004/3a, no JUMPstart-like sequence could be identified.

As is commonly the case with bacterial polysaccharide biosynthesis gene clusters, the genes in the *slg* gene clusters have a low G+C content, mostly ranging between 30 and 43% for individual genes. This is significantly lower than the G+C content determined for the respective bacterial genome as a whole, which is 46.7 and 46.3% for *A. thermoaerophilus* strains L420-91^{T} and DSM 10155/G^{+}, respectively, and 53.0% for *G. stearothermophilus* NRS 2004/3a. In the case of *G. stearothermophilus* NRS 2004/3a this is also lower than the G+C content of the genes adjacent to the cluster (*sgsE* and *istA*), or lower of that of the fragmentary transposases contained in the *A. thermoaerophilus* clusters (ORFA110, ORFA119, ORFA120; ORFA212) (Figures 3a–c). These observations suggest that the investigated organisms may have recently acquired their S-layer glycosylation machinery by lateral gene transfer [39]. The possible division of the *slg* gene cluster into distinct groups of genes with similar G+C contents, which are not necessarily flanked by transposase sequences, may even indicate different origins of these gene groups.

Comparison of some of the genes of the analysed bacterial *slg* clusters with data base entries for the corresponding genes from eukaryotes, revealed only low overall homologies. For instance, an alignment of the GDP-mannose 4,6 dehydratase Gmd from *A. thermoaerophilus* L420-91^{T} and *Caenorhabditis elegans* (GenBank accession number O45583) revealed an identity/similarity value of 52/66% for the region between amino acid 2 and 318 of the bacterial ORF. For other enzymes homologies exist only for certain protein motifs, *e.g.*, the amino acid stretch between position 129 and 311 of ORFA205 possesses 39/51% identity/similarity with the F32D8.8 protein of *C. elegans* (GenBank accession number Q19960). This indicates that for any protein of interest, a detailed analysis will be required to obtain conclusive comparative data.

In addition to the analyses of the *slg* gene clusters, for three of the investigated organisms, the respective S-layer protein structural gene has been sequenced. S-layer structural genes were given *s*** names, followed by the name of the bacterial species as a two-letter code. Consequently, the genes were designated *satA* for *A. thermoaerophilus* L420-91^{T} (GenBank accession number AY395578), *satB* for *A. thermoaerophilus* DSM 10155/G^{+} (AY395579), and *sgsE* for *G. stearothermophilus* NRS 2004/3a (AF328862). The G+C content of the S-layer genes is typical of housekeeping genes of the host chromosome. Concerning the chromosomal location of the S-layer genes in relation to the *slg* gene cluster, no common pattern seems to exist. *SgsE* of *G. stearothermophilus* NRS 2004/3a is located immediately upstream of the *slg* gene cluster with the intergenic region comprising 214 nucleotides, whereas *satA* and *satB* are located elsewhere on the chromosome. For *G. stearothermophilus* NRS 2004/3a it was demonstrated experimentally that the S-layer gene is transcribed monocistronically and independently of the *slg* gene cluster, which itself represents a polycistronic transcription unit [17]. However, some higher-level co-regulation of the S-layer protein and the S-layer glycan biosyntheses cannot be excluded. The S-layer precursor proteins encoded by the *satA*, *satB*, and *sgsE* genes, respectively, contain a signal sequence of 30 amino acids. Based on the known amino acid sequence it was possible to determine the position of the glycosylation sites on the respective S-layer precursor protein using purified proteolytic cleavage fragments of the S-layer glycoproteins (indicated in Figure 1).

Biosynthesis pathways of nucleosidediphosphate sugars involved in S-layer glycan biosynthesis

Based on the identification of the S-layer glycan specific nucleotide sugar genes in the *slg* gene clusters, the encoded proteins were cloned and overexpressed in *E. coli*. Functional assays were established that eventually led to the characterization of the biosynthesis pathways for dTDP-β-L-Rha [40], dTDP-α-D-Fuc3NAc [32], dTDP-α-D-Qui3NAc (Pfoestl A, Zayni S, Hofinger A, Kosma P, Schäffer C, Messner P, unpublished data), GDP-D-*glycero*-α-D-*manno*-heptose [41], and GDP-α-D-Rha [25] in Gram-positive organisms (Figure 2). In the course of these studies, Gmd from the GDP-α-D-Rha pathway was identified as a novel bifunctional enzyme exhibiting both dehydratase and reductase activities, and FdtA from the dTDP-α-D-Fuc3NAc pathway was the first isomerase described that is capable of synthesizing dTDP-6-deoxy-D-xylohex-3-ulose from dTDP-6-deoxy-D-xylohex-4-ulose. It should be noted that the heptose residue present in the S-layer glycan of *A. thermoaerophilus* L420-91^{T} is synthesized as GDP-D-*glycero*-α-D-*manno*-heptose, whereas ADP-L-*glycero*-β-D-*manno*-heptose is the precursor of the inner core lipopolysaccharide biosynthesis of organisms like *E. coli* or *Salmonella enterica* [42]. Thus, the S-layer protein glycosylation pathway provides a spectrum of rare enzymes that may be used for glycoengineering purposes in heterologous hosts. Furthermore, some of these enzymes from thermophilic S-layer carrying organisms exhibit significantly higher stability at 37°C than the enzymes from *S. enterica* (*e.g.,* most Rml enzymes from *A. thermoaerophilus* DSM 10155/G^{+}) [40]. This advantage could lead to the development of improved high-throughput screening systems for specific sugars [43].

S-layer protein glycosylation in comparison to other glycosylation pathways

As described above, the organizational pattern of the *slg* gene clusters to some extent is reminiscent of the situation found in LPS O-antigen genes [44–46]. Pathway genes for more common sugar precursors are scattered around the chromosome, whereas genes specific to S-layer glycans are in the cluster, glycosyltransferase genes are dispersed throughout the cluster, and genes for nucleotide sugar biosynthesis are usually organized in operons. The occurrence of an ABC-2 type transporter

system and the absence of a putative polymerase in the three fully sequenced *slg* gene clusters suggest that these S-glycan chains are synthesized in a process comparable to the *wzy*-independent pathway of the LPS O-polysaccharide assembly route, which has been described in detail elsewhere [47,48]. On the other hand, the presence of a putative flippase Wzx in the *slg* gene cluster *T. thermosaccharolyticum* E207-71 indicates that some S-layer glycoprotein glycans may also be synthesized via a *wzy*-dependent pathway [47,48]. In analogy to what is known from LPS O-polysaccharide biosynthesis, it may be speculated that this S-layer glycan is assembled via the *wzy*-dependent route because of its complex structure. The analyzed *slg* gene clusters from *A. thermoaerophilus* strains L420-91^T and DSM 10155/G$^+$, *G. stearothermophilus* NRS 2004/3a, and the incomplete cluster from *T. thermosaccharolyticum* E207-71 obviously do not possess a common organizational concept and also the genes encoding similar putative functions show significant differences in the number of membrane spanning domains, which may be taken as an indication of their involvement in a membrane-associated processes. Thus, giving here more details of a putative S-layer glycan biosynthesis pathway, as to how individual steps (*e.g.*, polymerization, chain-length termination) are carried out and at which topological location of the cell, would be unnecessarily speculative.

Other than the data presented here on the *slg* gene clusters of selected *Bacillaceae*, some *G. stearothermophilus* strains possess an *slg* gene cluster on the chromosome, but they do not exhibit a glycosylated S-layer protein under laboratory conditions [17]. Thus it is conceivable to assume that S-layer protein glycosylation is more widespread among bacteria in their natural environment than has been initially assumed. Only detailed molecular biological studies on a number of different organisms will show whether a general S-layer protein glycosylation pathway does exist for *Bacillaceae*. Studying also regulatory aspects of S-layer protein glycosylation will be a challenging task of the future.

Acknowledgments

We acknowledge the previous contributions of Michael Graninger and Bernd Kneidinger to the sequencing work of the *slg* gene clusters. We thank Sonja Zayni and Andrea Scheberl for excellent technical assistance. This work was supported by the Austrian Science Fund, projects P14209-B07 and P15612-B07 (to PM) and Nestec Ltd., project RE-002804.05 (to CS).

References

1 Benz I, Schmidt MA, Never say never again: Protein glycosylation in pathogenic bacteria, *Mol Microbiol* **45**, 267–76 (2002).

2 Messner P, Schäffer C, Prokaryotic glycoproteins. In *Progress in the Chemistry of Organic Natural Products*, edited by Herz W, Falk H, Kirby GW (Springer-Verlag, Wien, 2003), vol. 85, pp. 51–124.

3 Moens S, Non-S-Layer glycoproteins: A review. In *Glycomicrobiology*, edited by Doyle RJ (Kluwer Academic/Plenum Publishers, New York, 2000), pp. 93–125.

4 Szymanski CM, Logan SM, Linton D, Wren BW, *Campylobacter* — a tale of two protein glycosylation systems, *Trends Microbiol* **11**, 233–8 (2003).

5 Upreti RK, Kumar M, Shankar V, Bacterial glycoproteins: Functions, biosynthesis and applications, *Proteomics* **3**, 363–79 (2003).

6 Castric P, Cassels FJ, Carlson RW, Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan, *J Biol Chem* **276**, 26479–85 (2001), Correction, *J Biol Chem* **276**, 36058 (2001).

7 Power PM, Jennings MP, The genetics of glycosylation in Gram-negative bacteria, *FEMS Microbiol Lett* **218**, 211–22 (2003).

8 Thomas NA, Bardy SL, Jarrell KF, The archaeal flagellum: A different kind of prokaryotic motility structure, *FEMS Microbiol Rev* **25**, 147–74 (2001).

9 Sleytr UB, Messner P, Crystalline bacterial cell surface layers (S-layers). In *Desk Encyclopedia of Microbiology*, edited by Schaechter M (Elsevier Science USA, San Diego, 2003), pp. 286–93.

10 Sleytr UB, Sára M, Pum D, Schuster B, Messner P, Schäffer C, Self-assembly protein systems: Microbial S-layers. In *Biopolymers, Polyamides and Complex Proteinaceous Matrices I*, edited by Steinbüchel A, Fahnestock SR (Wiley-VCH, Weinheim, 2002), vol. 7, pp. 285–338.

11 Schäffer C, Wugeditsch T, Kählig H, Scheberl A, Zayni S, Messner P, The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation, *J Biol Chem* **277**, 6230–9 (2002).

12 Wugeditsch T, Zachara NE, Puchberger M, Kosma P, Gooley AA, Messner P, Structural heterogeneity in the core oligosaccharide of the S-layer glycoprotein from *Aneurinibacillus thermoaerophilus* DSM 10155, *Glycobiology* **8**,787–95 (1999).

13 Sumper M, Wieland FT, Bacterial glycoproteins. In *Glycoproteins*, edited by Montreuil J, Vliegenthart JFG, Schachter H (Elsevier, Amsterdam, 1995), pp. 455–73.

14 Meier-Stauffer K, Busse HJ, Rainey FA, Burghardt J, Scheberl A, Hollaus F, Kuen B, Makristathis A, Sleytr UB, Messner P, Description of *Bacillus thermoaerophilus* sp. nov., to include sugar beet isolates and *Bacillus brevis* ATCC 12990, *Int J Syst Bacteriol* **46**, 532–41 (1996).

15 Wugeditsch T, *Strukturanalyse des S-Schichtglykoproteins und Zellwand-Aminozuckerpolymers von Aneurinibacillus thermoaerophilus DSM 10155* (Doctoral thesis, Universität für Bodenkultur Wien, 1998), 238 pages.

16 Schäffer C, Messner P, Surface-layer glycoproteins: An example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology, *Glycobiology* **14**, 31R–42R (2004).

17 Novotny R, Schäffer C, Strauss J, Messner P, S-layer glycan-specific loci on the chromosome of *Geobacillus stearothermophilus* NRS 2004/3a and dTDP-L-rhamnose biosynthesis potential of *Geobacillus stearothermophilus* strains, *Microbiology* **150**, 953–65 (2004).

18 Christian R, Schulz G, Unger FM, Messner P, Küpcü Z, Sleytr UB, Structure of a rhamnan from the surface-layer glycoprotein of *Bacillus stearothermophilus* strain NRS 2004/3a, *Carbohydr Res* **150**, 265–72 (1986).

19 Kosma P, Wugeditsch T, Christian R, Zayni S, Messner P, Glycan structure of a heptose-containing S-layer glycoprotein of *Bacillus thermoaerophilus*, *Glycobiology* **5**, 791–6 (1995). Erratum in: *Glycobiology* **6**, 5 (1996).
20 Kosma P, Neuninger C, Christian R, Schulz G, Messner P, Glycan structure of the S-layer glycoprotein of *Bacillus* sp. L420-91, *Glycoconj J* **1**, 99–107 (1995).
21 Schäffer C, Müller N, Christian R, Graninger M, Wugeditsch T, Scheberl A, Messner P, Complete glycan structure of the S-layer glycoprotein of *Aneurinibacillus thermoaerophilus* GS4-97, *Glycobiology* **4**, 407–14 (1999).
22 Altman E, Schäffer C, Brisson JR, Messner P, Characterization of the glycan structure of a major glycopeptide from the surface layer glycoprotein of *Clostridium thermosaccharolyticum* E207-71, *Eur J Biochem* **229**, 308–15 (1995).
23 Markovitz A, Biosynthesis of guanosine diphosphate D-rhamnose and guanosine diphosphate D-talomethylose from guanosine diphosphate α-mannose, *J Biol Chem* **239**, 2091–8 (1964).
24 Giraud M-F, Naismith JH, The rhamnose pathway, *Curr Opin Struct Biol* **10**, 687–96 (2000).
25 Kneidinger B, Graninger M, Adam G, Puchberger M, Kosma P, Zayni S, Messner P, Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in *Aneurinibacillus thermoaerophilus* L420-91^{T}, *J Biol Chem* **276**, 5577–83 (2001).
26 Kneidinger B, Graninger M, Messner P, Chromosome walking by cloning of distinct PCR fragments, *Biotechniques* **30**, 248–9 (2001).
27 Jiang S-M, Wang L, Reeves PR, Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters, *Infect Immun* **69**, 1244–55 (2001).
28 Tsukioka Y, Yamashita Y, Oho T, Nakano Y, Koga T, Biological function of the dTDP-rhamnose synthesis pathway in *Streptococcus mutans*, *J Bacteriol* **179**, 1126–34 (1997).
29 Mitchison M, Bulach DM, Vinh T, Rajakumar K, Faine S, Adler B, Identification and characterization of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related *rfb* locus in *Leptospira interrogans* serovar Copenhageni, *J Bacteriol* **179**, 1262–7 (1997).
30 Nakano Y, Yoshida Y, Suzuki N, Yamashita Y, Koga T, A gene cluster for the synthesis of serotype d-specific polysaccharide antigen in *Actinobacillus actinomycetemcomitans*, *Biochim Biophys Acta* **1493**, 259–63 (2000).
31 Wang L, Liu D, Reeves PR, C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis, *J Bacteriol* **178**, 2598–604 (1996).
32 Pfoestl A, Hofinger A, Kosma P, Messner P, Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91^{T}, *J Biol Chem* **278**, 26410–7 (2003).
33 Bengoechea JA, Pinta E, Salminen T, Oertelt C, Holst O, Radziejewska-Lebrecht J, Piotrowska-Seget Z, Venho R, Skurnik M, Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O:8, *J Bacteriol* **184**, 4277–87 (2002).
34 Estrela AI, Pooley HM, de Lencastre H, Karamata D, Genetic and biochemical characterization of *Bacillus subtilis* 168 mutants specifically blocked in the synthesis of the teichoic acid poly(3-O-β-D-glucopyranosyl-N-acetylgalactosamine 1-phosphate): *gneA*, a new locus, is associated with UDP-N-acetylglucosamine 4-epimerase activity, *J Gen Microbiol* **137**, 943–50 (1991).
35 Reeves PR, Wang L, Genomic organization of LPS-specific loci, *Curr Top Microbiol Immunol* **264**, 109–35 (2002).
36 Jolly L, Stingele F, Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria, *Int Dairy J* **11**, 733–45 (2001).
37 Skurnik M, Molecular genetics, biochemistry and biological role of *Yersinia* lipopolysaccharide, *Adv Exp Med Biol* **529**,187–97 (2003).
38 Hobbs M, Reeves PR, The JUMPstart sequence: A 39 bp element common to several polysaccharide gene clusters, *Mol Microbiol* **12**, 855–56 (1994).
39 Keenleyside WJ, Whitfield C, Genetics and biosynthesis of lipopolysaccharide O-antigens. In *Endotoxin in Health and Disease*, edited by Brade H, Opal SM, Vogel SN, Morrison DC (Marcel Dekker, New York, Basel, 1999), pp. 331–58.
40 Graninger M, Kneidinger B, Bruno K, Scheberl A, Messner P, Homologs of the Rml enzymes from *Salmonella enterica* are responsible for dTDP-β-L-rhamnose biosynthesis in the gram-positive thermophile *Aneurinibacillus thermoaerophilus* DSM 10155, *Appl Environ Microbiol* **68**, 3708–15 (2002).
41 Kneidinger B, Graninger M, Puchberger M, Kosma P, Messner P, Biosynthesis of nucleotide-activated D-*glycero*-D-*manno*-heptose, *J Biol Chem* **276**, 20935–44 (2001).
42 Kneidinger B, Marolda C, Graninger M, Zamyatina A, McArthur F, Kosma P, Valvano MA, Messner P, Biosynthesis pathway of ADP-L-*glycero*-β-D-*manno*-heptose in *Escherichia coli, J Bacteriol* **184**, 363–69 (2002).
43 Ma Y, Stern RJ, Scherman MS, Vissa VD, Yan W, Jones VC, Zhang F, Franzblau SG, Lewis WH, McNeil MR, Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: Genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose, *Antimicrob Agents Chemother* **45**, 1407–16 (2001).
44 Samuel G, Reeves P, Biosynthesis of O-antigens: Genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly, *Carbohydr Res* **338**, 2503–19 (2003).
45 Whitfield C, Paiment A, Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria, *Carbohydr Res* **338**, 2491–502 (2003).
46 Whitfield C, Roberts IS, Structure, assembly and regulation of expression of capsules in *Escherichia coli*, *Mol Microbiol* **31**, 1307–19 (1999).
47 Raetz CRH, Whitfield C, Lipopolysaccharide endotoxins, *Annu Rev Biochem* **71**, 635–700 (2002).
48 Whitfield C, Biosynthesis of lipopolysaccharide O-antigens, *Trends Microbiol* **3**, 178–85 (1995).

Received 22 December 2003; revised 7 April 2004;
accepted 26 April 2004