

# Bioinformatic Survey of S-Layer Proteins in Bifidobacteria

Juan Li<sup>1,2</sup>, Yihao Shen<sup>1</sup>, Yatao Jiang<sup>1</sup>, Le He<sup>1</sup>, Zhongke Sun<sup>1\*</sup>

<sup>1</sup>Institute of Food and Drug Inspection, College of Life Sciences and Agronomy, Zhoukou Normal University, Zhoukou, China

<sup>2</sup>Library of Zhoukou Normal University, Zhoukou, China

Email: \*sunzh@daad-alumni.de

**How to cite this paper:** Li, J., Shen, Y.H., Jiang, Y.T., He, L. and Sun, Z.K. (2018) Bioinformatic Survey of S-Layer Proteins in Bifidobacteria. *Computational Molecular Bioscience*, 8, 68-79.  
<https://doi.org/10.4236/cmb.2018.82003>

**Received:** March 31, 2018

**Accepted:** June 4, 2018

**Published:** June 7, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Surface layer (S-layer) proteins are one of the most commonly observed cell envelope components in both Archaea and Bacteria. It has versatile functions and holds considerable application potential in biotechnology. Bifidobacteria are representative probiotics conferring health promoting properties. However, there is little study of S-layer in bifidobacteria yet. The distribution and characteristics of S-layer in bifidobacteria are unknown. In this study, search for S-layer protein in the identical protein groups in NCBI yielded 49 hits belonging to bifidobacteria. These proteins were annotated as either “S-layer (domain) protein” or “putative S-layer (y) domain protein” that distributed among 26 species of *Bifidobacterium* genus. Multiple alignments suggest S-layer proteins are relatively conservative. Phylogenetic analysis of 24 S-layer (domain) protein sequences groups them into three distinct clusters, with the majority species in Cluster-2. S-layer (domain) protein has a universe motif DUF4381, though its function is unknown. Meanwhile, two other motifs CARDB and EphA2\_TM involved in cell adhesion and cell signaling respectively, presented in most S-layer (domain) protein in bifidobacteria. All S-layer proteins have a typical N-terminal Sec-dependent signal peptide and a C-terminal trans-membrane region. Homological modeling of representative S-layer proteins from each cluster revealed a few unique structural features. All representative S-layer proteins have a plenty of  $\beta$ -meander motif that exclusively composed by  $\beta$ -barrel structural architectures linked together by hairpin loops.

## Keywords

Bifidobacteria, S-layer Domain Protein, Phylogeny, Homology Modeling, Domain

## 1. Introduction

S-layers are one of the most commonly observed prokaryotic cell surface structures. They are composed by two-dimensional arrays of proteinaceous subunits (S-layer proteins), presented in almost every taxonomic group of walled Bacteria and almost universal in Archaea [1]. S-layer proteins are one of the most abundant biopolymers on our planet, as they account for approximately ten percent of cellular proteins in Archaea and Bacteria [2]. They are generally composed of a single molecular species that can assemble on the cell surface into closed regular arrays. Therefore, S-layer can function as protective coats, molecular sieves, molecule and ion traps, promoters for cell adhesion, immunomodulators, surface recognition, because of their repetitive identical physicochemical properties and identical pores in size and morphology [3].

In Gram-positive bacteria, S-layers attached to the rigid peptidoglycan-containing layer. S-layers completely cover the cell surface during all stages of cell growth and division. Chemical and genetically analysis of many S-layers has revealed a similar overall composition [4]. They are generally composed of a single protein or glycoprotein species with molecular masses ranging from 40 to 170 kDa [5]. Most S-layers of bacteria are composed of weakly acidic proteins or glycoproteins, contain 40% - 60% hydrophobic amino acids, and possess few or no sulfur-containing amino acids [6]. The pI values of the proteins range from 4 to 6. However, pIs of the S-layer proteins between 8 and 10 have been determined in Lactobacilli [7]. Comparative studies on S-layer genes of organisms from different taxonomic affiliations revealed that homologies between nonrelated organisms are low despite the fact that their amino acid composition shows no significant difference [8]. Nevertheless, it is obvious that common structural principles must exist in S-layer proteins (e.g. the ability to form inter-subunit bonds and to self-assemble into monomolecular arrays, the formation of hydrophilic pores with low unspecific adsorption, and the interaction with underlying cell envelope components).

Bifidobacteria are generally recognized as safe (GRAS), exerting many beneficial health effects on their host, and have attracted strong interest in the health care and food industries [9]. Although a large amount of knowledge has accumulated on the structure, assembly, chemistry, and genetics of S-layers [10], little data are available about their specific presence in bifidobacteria. In this study, we try to survey the distribution and study the genetics as well as structures of S-layers by bioinformatic approaches.

## 2. Material and Methods

### 2.1. Sequence Search

The S-layer protein sequences of bifidobacteria were searched from NCBI-Identical Protein Groups (IPG) with the key words “S-layer domain protein AND Bifidobacterium” (<https://www.ncbi.nlm.nih.gov/>). The resulting 49 protein sequences annotated as either “S-layer (domain) protein” or “putative S-layer (y) domain

protein” were used for primary analysis. Domain Enhanced Lookup Time Accelerated BLAST (DRLTA-BLAST) conducted a second search with the longest consensus regions as queries when expected threshold was 4.0. The queries of S-layer (domain) protein (P146, YVNFGKGD, 8aa) and putative S-layer (y) domain protein (P277, QLVTWVESHDNYAN, 14aa) were obtained when threshold was set at 100% by local ClustalW multiple alignments [11].

## 2.2. Multiple Alignment and Phylogenetic Analysis

Protein sequences of S-layer (domain) protein and putative S-layer (y) domain protein were then aligned separately by local ClustalW program version 2.0 with the progressive method [12]. Sequences too short or significantly different from the others were removed in the final alignment. Consensus regions were recognized when at least four continuous amino acids are identical and the threshold is 100%. Sequences upward the first or downward the last consensus region were deleted. Remaining sequences were used for the construction of phylogenetic tree by the neighbor-joining method using Protein-Dist program incorporated in BioEdit with 1000 bootstrap replicates (version 6.0) as described elsewhere [13].

## 2.3. Physicochemical Analysis and Motif Scan

Representative sequences, including S-layer proteins of *B. thermophilum* RBL67 (Accession: AGH41482.1), *B. pseudocatenuatum* LMG10505 (Accession: KFI75572.1), and *B. longum* DJO10A (Accession: ACD98337.1), belonging to Clusters 1, 2, and 3, respectively, were analyzed by ProtParam tool at ExPASy (<http://web.expasy.org/protparam/>). The sub-location of S-layer proteins was analyzed by PSORTb v3.0.2 program as well (<http://www.psорт.org/psортb/index.html>). All motifs in S-layer (domain) proteins were screened by MOTIFS program (<http://www.genome.jp/tools/motif/>). Above representative sequence in each cluster was used as example for illustration of conserved and/or unique structural motifs. The database used for the search is Pfam library and the E-value is 1.0 with Profile Hidden Markov Model [14]. The motifs were analyzed by comparison of their structural characteristics annotated in PDB (<https://www.rcsb.org/>).

## 2.4. Analysis of Functional and Structural Regions

Potential signal peptide (SP) sequences of all S-layer protein were analyzed using SignalP Version 4.1 and TATFIND [15] [16]. TMHMM Server predicted potential trans-membrane helices in protein v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). Conserved domains were computed by batch Web CD search Tool with default setting [8].

## 2.5. Structural Modeling of S-Layer (domain) Protein

The sequences of representative S-layer (domain) proteins from each cluster of

the phylogenetic tree were searched for closest homologues in protein data bank (PDB) database using NCBI-BLASTp search program with the algorithm of DELTA-BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp>). Meanwhile, the three sequences were used as template separately for structure prediction. Five homology models were obtained for each sequence from RaptorX server (<http://raptorx.uchicago.edu/StructurePrediction>) [17]. The models were evaluated and the best model was chosen based on P-value as well as score. The superposed structure was visualized using the web-based 3-D structure viewer iCn3D in Ribbon style

(<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>).

Furthermore, homology models of these three proteins were built separately by SWISS-MODEL (<https://swissmodel.expasy.org/>). Finally, QMEANBrane estimated the quality of modeling (<https://swissmodel.expasy.org/qmean/>).

### 3. Results and Discussion

#### 3.1. Distribution of S-Layer (domain) Proteins in *Bifidobacterium*

Search for S-layer protein in the identical protein groups in NCBI yielded 49 hits belonging to bifidobacteria. The sequences that were annotated as either S-layer (domain) protein or putative S-layer (y) domain protein were downloaded. Basic information of these proteins was summarized in **Table 1** and listed according to species. It is clear that S-layer and putative S-layer proteins distributed in only 26 known species and 1 uncharacterized specie of the genus of *Bifidobacterium*. In reference to the most recent bacterial classification and sequence record, no (putative) S-layer homologues could be identified in other 30 species of the genus of *Bifidobacterium*

(<https://www.dsmz.de/fileadmin/Bereiche/ChiefEditors/BacterialNomenclature>).

However, we noticed that there is more than one sequence in many hits that have identical sequence but with different name/annotation from NCBI-IPG. Therefore, we performed the secondary search in the non-redundant protein sequences database with two consensus regions P146 and P277 by DELTA-BLAST, respectively. Results suggest S-layer (domain) proteins are more widely annotated as either “hypothetical protein” or “ABC transporter permease”, while putative S-layer (y) domain protein are mainly annotated as “hypothetical protein” or “alpha amylase” in nearly all species of *Bifidobacterium*. Furthermore, we found that there is more than one (putative) S-layer (domain) protein in several species/strains. For example, *B. scardovii* LMG 21589 has two S-layer (domain) proteins; *B. choerinum* LMG 10510 has two putative S-layer (y) domain proteins; while *B. gallicum* LMG 11596 has one S-layer (domain) protein and as much as 6 putative S-layer (y) domain proteins.

#### 3.2. Conservation and Phylogeny of S-Layer (Domain) Protein

Multiple alignments of 24 S-layer (domain) proteins yielded several consensus regions. The longest consensus region YVNFQKGD was marked as P146. As

**Table 1.** Bifidobacterial S-layer (domain) proteins in NCBI-identical protein groups.

Name	Species	Rep. strain	Accession No.	Size (aa)	Source
S-layer (domain) protein	<i>B. adolescentis</i>	2789STDY5608824	CUN56955.1	390	INSDC
		2789STDY5608862	CUN77683.1	390	INSDC
	<i>B. longum</i>	DJO10A	ACD98337.1	388	INSDC
		BT1	ALE08127.1	388	INSDC
		LMG 21814	KFI71892.1	388	INSDC
		BG7	ALE35760.1	391	INSDC
		LMG 13197	KFI65071.1	388	INSDC
		AH1206	AOP00589.1	388	INSDC
		35624	AOL09953.1	391	INSDC
	<i>B. boum</i>	BBMN68	ADQ01830.1	391	INSDC
		LMG 10736	KFI46217.1	405	INSDC
	<i>B. breve</i>	BR3	ALE13696.1	431	INSDC
	<i>B. catenulatum</i>	LMG 11043	KFI55811.1	378	INSDC
	<i>B. dentium</i>	Bd1	ADB09373.1	407	INSDC
	<i>B. gallinarum</i>	LMG 11586	KFI60864.1	400	INSDC
	<i>B. gallicum</i>	LMG 11593	WP006295244.1		RefSeq
	<i>B. minimum</i>	LMG 11592	KFI72335.1		INSDC
	<i>B. moukalabense</i>	DSM 27321	ETY71086.1	416	INSDC
	<i>B. myosotis</i>	DSM 100196	OZG56886.1		INSDC
	<i>B. pseudocatenulatum</i>	LMG 10505	KFI75572.1	373	INSDC
	<i>B. reuteri</i>	DSM 23975	KFI87931.1	480	INSDC
	<i>B. scardovii</i>	LMG 21589	KFI91418.1	454	INSDC
		LMG 21589	KFI92534.1		INSDC
<i>B. stellenboschense</i>	DSM 23968	KFI98781.1	387	INSDC	
<i>B. stercoris</i>	DSM 24849	KFI96690.1	390	INSDC	
<i>B. subtilis</i>	LMG 11597	KFI98912.1		INSDC	
<i>B. thermophilum</i>	JCM 1207	KFJ07718.1	407	INSDC	
	RBL67	AGH41482.1	405	INSDC	
<i>B. sp.</i>	12_1_47BFAA	EFV36416.1	391	INSDC	
putative S-layer (y) domain protein	<i>B. angulatum</i>	LMG 11039	KFI40943.1		INSDC
	<i>B. animalis</i>	Bl12	AGO52940.1		INSDC
	<i>B. choerinum</i>	LMG 10510	KFI56970.1		INSDC
		LMG 10510	KFI54244.1		INSDC
<i>B. gallicum</i>	LMG 11596	KFI59372.1		INSDC	

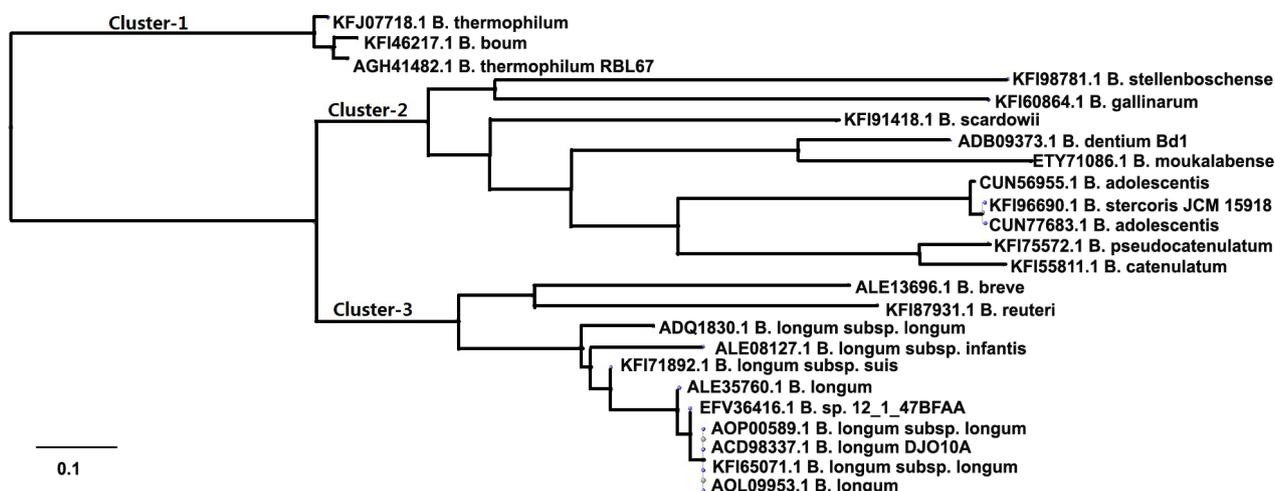
continued

	LMG 11596	KFI59699.1	INSDC
	LMG 11596	KFI59086.1	INSDC
	LMG 11596	KFI59089.1	INSDC
	LMG 11596	WP006295430.1	RefSeq
	LMG 11596	WP006295735.1	RefSeq
<i>B. magnum</i>	LMG 11591	KFI68108.1	INSDC
<i>B. merycicum</i>	LMG 11341	KFI69438.1	INSDC
<i>B. minimum</i>	LMG 11592	KFI73654.1	INSDC
<i>B. pseudolongum</i>	LMG 11569	KFI75530.1	INSDC
	LMG 11569	KFI75529.1	INSDC
	LMG 11569	KFI75058.1	INSDC
	LMG 11571	KFI77915.1	INSDC
<i>B. pseudocatenulatum</i>	LMG 10505	KFI74947.1	INSDC
<i>B. stercoris</i>	DSM 24849	KFI95044.1	INSDC
<i>B. tsurumiense</i>	JCM 13495	KFJ06122.1	INSDC

shown in **Figure S1**, S-layer proteins are quite conserving in 14 different species in *Bifidobacterium* as the general identity reaches 50% (number of identical amino acids 80/150 analyzed). Phylogenetic analysis of 24 S-layer (domain) protein sequences groups them into three distinct clusters, with the majority species in Cluster-2 (**Figure 1**). Similar analysis of putative S-layer (y) domain sequences in ninespecies of Bifidobacteria suggests they are conserved and can be grouped into three clusters as well. We next analyzed the S-layer (domain) sequence clusters according to their habitat and found that most of the sequences in Cluster-2 distributed as endosymbionts in the gastrointestinal tract of human and other animals, while sequences in Cluster-3 is mainly belongs to *B. longum*, which is frequently isolated from human faces (<http://www.bacterio.net/bifidobacterium.html>).

### 3.3. Physiochemical Features and Motifs of S-Layer (Domain) Proteins

ProtParam computation of representative S-layer proteins from each cluster indicates they are stable proteins have high value of aliphatic index and close pI (detail values see **Table 2**). Generally, all of them have much more negatively charged residues than positively charged residues. PSORTb analysis suggests S-layer protein in *B. thermophilum* RBL67 localized in the cytoplasmic membrane (Localization Score 9.87/10). In contrast, S-layer proteins in *B. longum* DJO10A and *B. pseudocatenulatum* LMG 10505 may have multiple localization sites, as their localization scores are 3.33/10 in both cytoplasmic membrane, cell wall, and extracellular, respectively [18].



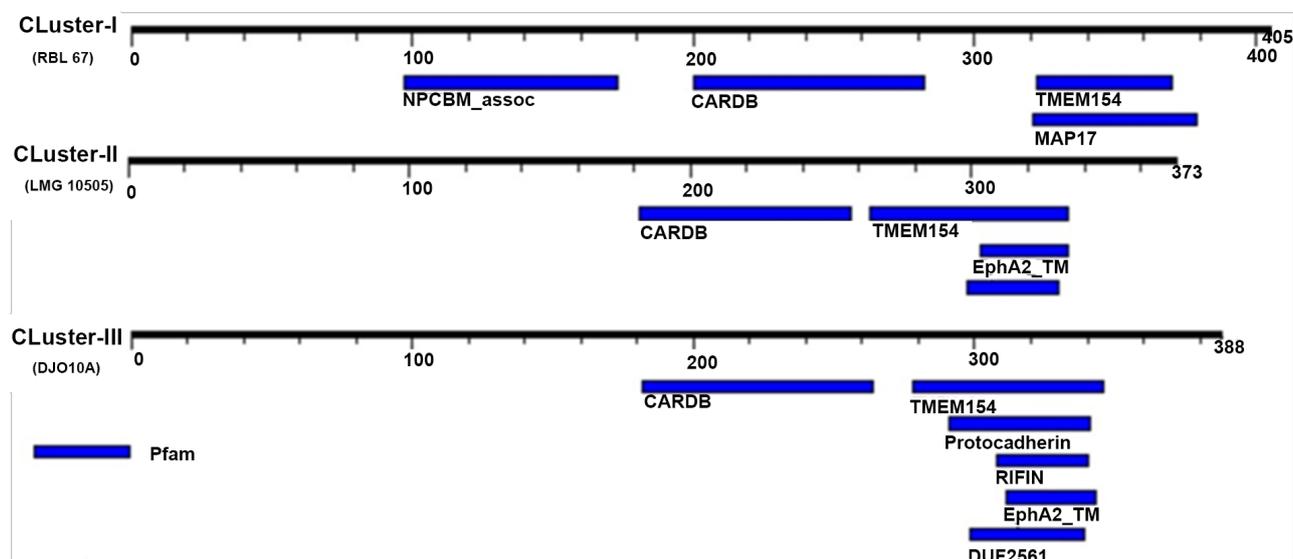
**Figure 1.** Phylogenetic analysis of S-layer domain proteins in Bifidobacteria. Tree of S-layer (domain) proteins was constructed by BioEdit Protein-Dist-Neighbor Phylogenetic Tree with 1000 bootstrap replicates.

**Table 2.** Physicochemical parameters of representative S-layer proteins.

Parameters/Species	<i>B. thermophilum</i> RBL67	<i>B. pseudocatenulatum</i> LMG10505	<i>B. longum</i> DJO10A
No. of amino acids	405	373	388
Molecular weight	42,180.03	38,803.91	40,777.55
Theoretical pI	4.39	4.13	4.41
No. of negatively charged residues	51	49	47
No. of positively charged residues	30	21	25
Instability index	35.84	24.03	28.21
Aliphatic index	80.94	81.07	82.42
GRAVY	-0.17	-0.103	-0.091

No., number; GRAVY, Grand average of hydrophobicity.

By MOTIFS searching of S-layer (domain) sequences extracted from NCBI-IPG, we recognized a plenty of motifs in each sequence. For simplicity, motifs in representative sequences were compared when E-value is 0.01 (Figure 2). Motifs CARDB and TMEM154 are presented on all representative sequences and each S-layer protein have unique motifs. However, there are a large number of motifs when E-value is 1.0. Comparison analysis suggests both common motifs in all the three clusters and unique motifs in each cluster. DUF4381 is the only universe motif among all species of bifidobacteria, though its function is unknown. CARDB is the second motif widely distributed among 23 of 24 sequences. CARDB (PF07705) represents cell adhesion related domain found in bacteria [19]. This structure supports the revealed adhesion function of S-layer protein in bifidobacteria. EphA2\_TM also widely presents in bifidobacteria (22/24), which is the left-handed dimer trans-membrane domain of Ephrin receptors [20]. This binding leads to contact-dependent bidirectional signaling into neighboring cells, which may contribute to the probiotic effects of bifidobacteria by antagonisms. Furthermore, analysis of the structural characteristics of



**Figure 2.** Simplified representation of motifs in representative S-layer (domain) proteins.

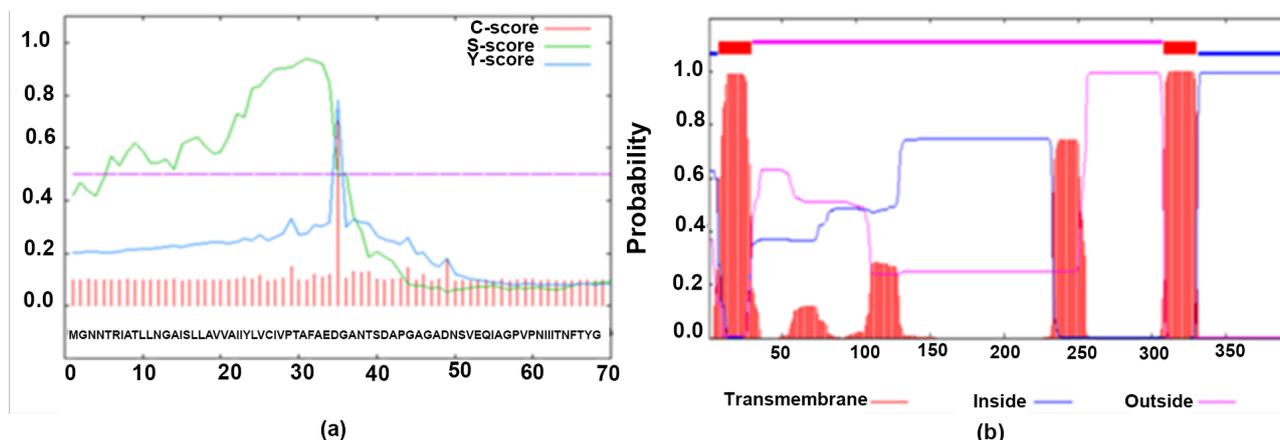
these motifs indicates some important properties of S-layer protein. For example, structural motif corresponding to the first  $\alpha$ -helix of S-layer protein is conserved in all clusters.

### 3.4. Functional and Structural Regions of S-Layer (Domain) Proteins

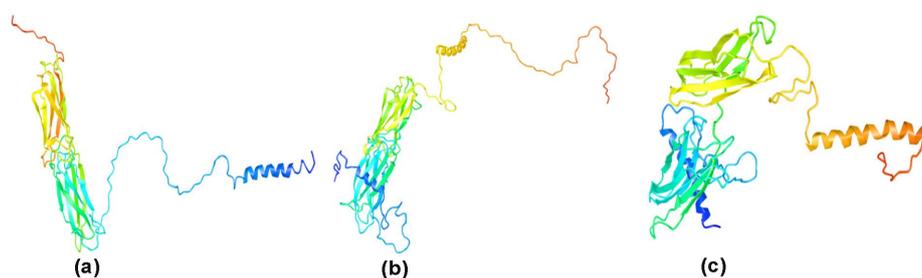
Signal peptide (SP) responses for the direction of protein secretion across cell membrane. S-layer (domain) protein needs such structural element to direct its sub-localization. SignalP-TM prediction with Gram-positive bacteria model indicates most S-layer (domain) proteins, exactly 23 of 24 analyzed, have a potential Sec dependent SP (as represented in **Figure 3(a)**). However, there is no Tat-dependent SP in all sequences analyzed. TMHMM Server v.2.0 prediction suggests all S-layer (domain) protein have trans-membrane (TM) helices in both N-terminal and C-terminal, even the N-terminal TM region is probably signal peptide (**Figure 3(b)**). Conserved domain search suggests COG1361 (PSSMID 224280) and CARDB (PSSMID 285007) super-family domains are presented in most of these sequences, though 7 proteins have no domain hits (**Figure S2**). COG1361 is either an uncharacterized conserved protein domain or S-layer domain involved in cell envelope and outer membrane biogenesis [21] CARDB domain is a cell adhesion related domain widely found in bacteria.

### 3.5. Structural Modeling of Representative S-layer Proteins

Search of PDB database of these three representative sequences by Blastp yielded same results. All sequences have a homological structure model to Chain A of *Vibrio nigripulchritudo* nigrtoxine with 33% identity in 45aa (PDB ID: 5M41). However, this model represents a small partial structure, as only 141aa was included in the model. Therefore, we next generated homology models of representative S-layer (domain) protein sequences from each cluster. Five models



**Figure 3.** Analysis of functional and structural regions in S-layer proteins: (a) a representative signal peptide of S-layer protein (CUN56955.1) predicted by Singal P4.1 using the N-terminal 70 amino acids. The violet line is the default cutoff (score = 0.5). Red line is C-score (raw cleavage site score). Green line is S-score (signal peptide score). Blue line is Y-score, which are a combination of the C-score and the slope of the S-score; (b) TMs of a representative S-layer protein (CUN56955.1) predicted by TMHMM 2.0.



**Figure 4.** Homological structures of representative S-layer proteins in each cluster: a. a structure model of S-layer protein in *B. thermophilum*; (b) a structure model of S-layer protein in (b) *pseudocatenuatum*; (c) a structure model of S-layer protein in *B. longum*. Each model displayed above is the model predicted by RaptorX server with the highest score and the lowest P-value. N-terminus regions are in blue, and C-terminus regions are in red.

were generated for each sequence. The best predicted model was selected and demonstrated in **Figure 4**. Multiple structural alignments of the homology models show that all proteins have a plenty of  $\beta$ -meander motif that are exclusively composed by  $\beta$ -barrel structural architectures linked together by hairpin loops. Furthermore, similar spatial orientation displayed in motifs having same secondary structural elements and all N-terminus regions (shown in blue) have  $\alpha$ -helices. In contrary,  $\alpha$ -helices in C-terminus regions (shown in orange) absent in Cluster-1 (**Figure 4(a)**, *B. thermophilum*). The loop regions between the first  $\alpha$ -helix and  $\beta$ -sheet (from N-terminus, shown in light blue), and the regions after the last  $\beta$ -sheet or  $\alpha$ -helix (shown in red) have obvious structural deviations. The electrostatic surface potential was also analyzed using the predicted structure of S-layer domain protein and observed the presence of a patch of negatively charged potential on both  $\beta$ -sheets and  $\alpha$ -helices.

#### 4. Conclusion

In this study, we investigated the distribution of S-layer domain protein in *Bifi-*

*dobacterium* from a phylogenetic and structural perspective. Phylogenetic analysis on all annotated S-layer protein sequences grouped them into three distinct clusters. (Putative) S-layer (y) domain proteins distributed in less than half species in bifidobacteria, though they have several conserve regions and their longest consensus sequences P146/P227 are common in nearly all species of bifidobacteria. S-layer proteins have different motifs and domains that are either involved in cell envelope and outer membrane biogenesis or related to cell adhesion. Furthermore, all S-layer (domain) proteins have a typical signal peptide sequence and a C-terminal trans-membrane region. Analysis of homological models of representative sequences revealed cluster-specific structural properties of S-layer protein.

### Acknowledgements

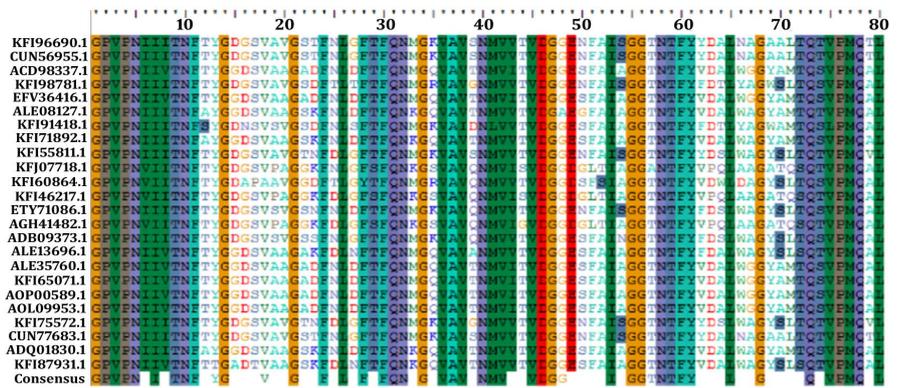
This study was supported by a cooperation grant (No. 172102410055) from the Henan Agency of Science and Technology and a grant of Key Scientific Research Project (No. 17A180017) from the Henan Province Department of Education. The funders had no role in the study design, data collection and analysis, or decision to publish.

### References

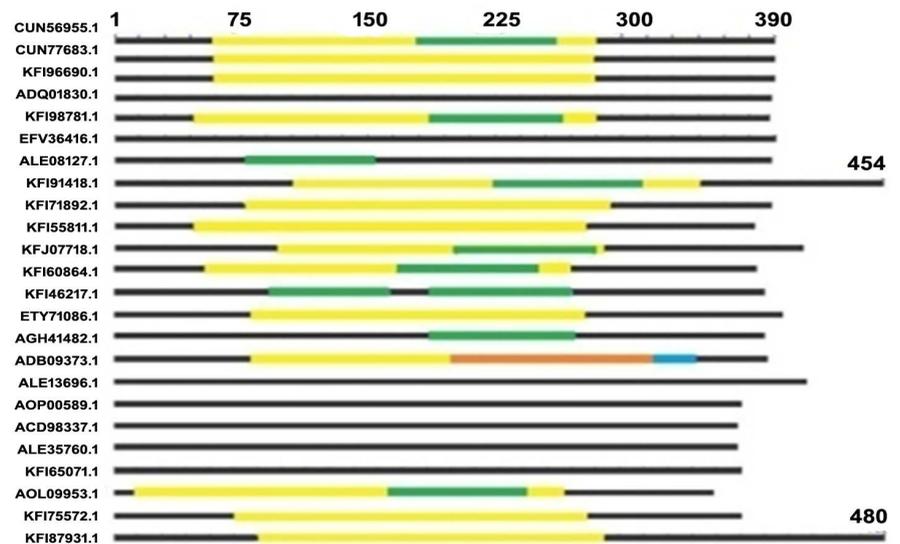
- [1] Sleytr, U.B., Schuster, B., Egelseer, E.M., *et al.* (2014) S-Layers: Principles and Applications. *FEMS Microbiology Reviews*, **38**, 823-864. <https://doi.org/10.1111/1574-6976.12063>
- [2] Pum, D. and Sleytr, U.B. (2014) Reassembly of S-Layer Proteins. *Nanotechnology*, **25**, 312001. <https://doi.org/10.1088/0957-4484/25/31/312001>
- [3] Ladenhauf, E.M., Pum, D., Wastl, D.S., *et al.* (2015) S-Layer Based Biomolecular Imprinting. *RSC Advances*, **5**, 83558-83564. <https://doi.org/10.1039/C5RA14971A>
- [4] Messner, P. (2010) Occurrence, Structure, Chemistry, Genetics, Morphogenesis, and Function of S-Layers. *Prokaryotic Cell Wall Compounds-Structure and Biochemistry*, Chapter 2, Springer, Heidelberg, Germany, 53-109.
- [5] Messner, P., Schaaffer, C. and Kosma, P. (2013) Bacterialcell-Envelope Glycoconjugates. *Advances in Carbohydrate Chemistry and Biochemistry*, **69**, 209-272. <https://doi.org/10.1016/B978-0-12-408093-5.00006-X>
- [6] Kern, J., Wilton, R., Zhang, R., *et al.* (2011) Structure of Surface Layer Homology (slh) Domains from Bacillus Anthracis Surface Array Protein. *Journal of Biological Chemistry*, **286**, 26042-26049. <https://doi.org/10.1074/jbc.M111.248070>
- [7] Johnson, B.R., Hymes, J., Sanozky-Dawes, R., *et al.* (2016) Conserved S-Layer-Associated Proteins Revealed by Exoproteomic Survey of S-Layer-Forming Lactobacilli. *Applied and Environmental Microbiology*, **82**, 134-145. <https://doi.org/10.1128/AEM.01968-15>
- [8] Marchler-Bauer, A., Bo, Y., Han, L., *et al.* (2017) Cdd/Sparcle: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Research*, **45**, D200-D203. <https://doi.org/10.1093/nar/gkw1129>
- [9] Sun, Z., Baur, A., Zhurina, D., *et al.* (2012) Accessing the Inaccessible: Molecular Tools for Bifidobacteria. *Applied and Environmental Microbiology*, **78**, 5035-5042. <https://doi.org/10.1128/AEM.00551-12>

- [10] Hynönen, U. and Palva, A. (2013) Lactobacillus Surface Layer Proteins: Structure, Function and Applications. *Applied Microbiology & Biotechnology*, **97**, 5225-5243. <https://doi.org/10.1007/s00253-013-4962-2>
- [11] Boratyn, G.M., Schäffer, A.A., Agarwala, R., *et al.* (2012) Domain Enhanced Lookup Time Accelerated Blast. *Biology Direct*, **7**, 12. <https://doi.org/10.1186/1745-6150-7-12>
- [12] Larkin, M.A., Blackshields, G., Brown, N.P., *et al.* (2007) Clustal W and Clustal X Version 2.0. *Bioinformatics*, **23**, 2947-2948. <https://doi.org/10.1093/bioinformatics/btm404>
- [13] Sun, Z., Liu, K., Zhang, J., *et al.* (2017) IAA Producing *Bacillus altitudinis* Alleviates Iron Stress in *Triticum aestivum* L. Seedling by Both Bioleaching of Iron and Up-Regulation of Genes Encoding Ferritins. *Plant & Soil*, **419**, 1-11.
- [14] Finn, R.D., Bateman, A., Clements, J., *et al.* (2014) Pfam: The Protein Families Database. *Nucleic Acids Research*, **42**, D222-D230.
- [15] Petersen, T.N., Brunak, S., von Heijne, G., *et al.* (2011) SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nature Methods*, **8**, 785-786. <https://doi.org/10.1038/nmeth.1701>
- [16] Rose, R.W., Brüser, T., Kissinger, J.C., *et al.* (2002) Adaptation of Protein Secretion to Extremely High-Salt Conditions by Extensive Use of the Twin-Arginine Translocation Pathway. *Molecular Microbiology*, **45**, 943-950. <https://doi.org/10.1046/j.1365-2958.2002.03090.x>
- [17] Källberg, M., Wang, H., Wang, S., *et al.* (2012) Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nature Protocols*, **7**, 1511-1522. <https://doi.org/10.1038/nprot.2012.085>
- [18] Yu, N.Y., Wagner, J.R., Laird, M.R., *et al.* (2010) PSORTb 3.0: Improved Protein Subcellular Localization Prediction with Refined Localization Subcategories and Predictive Capabilities for All Prokaryotes, *Bioinformatics*, **26**, 1608-1615.
- [19] Mello, L.V., Chen, X. and Rigden, D.J. (2010) Mining Metagenomic Data for Novel Domains: Bacon, a New Carbohydrate-Binding Module. *FEBS Letters*, **584**, 2421-2426. <https://doi.org/10.1016/j.febslet.2010.04.045>
- [20] Kullander, K. and Klein, R. (2002) Mechanisms and Functions of Eph and Ephrin Signaling. *Nature Reviews Molecular Cell Biology*, **3**, 475-486. <https://doi.org/10.1038/nrm856>
- [21] Auernik, K.S., Cooper, C.R. and Kelly, R.M. (2008) Life in Hot Acid: Pathway Analyses in Extremely Thermoacidophilic Archaea. *Current Opinion in Biotechnology*, **19**, 445-453. <https://doi.org/10.1016/j.copbio.2008.08.001>

### Supplementary Figures



**Figure S1.** Partial representation of 24 S-layer (domain) proteins alignment by ClustalW. Conservation was indicated by consensus amino acids when the threshold is 100%.



**Figure S2.** Schematic presentation of conserved domains in S-layer proteins in different species belonging to *Bifidobacterium*. Regions labeled in yellow are domain COG1363. Regions labeled in green are domain CARDB.