Classification of hyper-variable *Corynebacterium glutamicum* surface-layer proteins by sequence analyses and atomic force microscopy

Nicole Hansmeier[a,b], Frank W. Bartels[c], Robert Ros[c], Dario Anselmetti[c], Andreas Tauch[b], Alfred Pühler[a], Jörn Kalinowski[b,*]

[a] *Lehrstuhl für Genetik, Universität Bielefeld, Universitätsstraße 25, D-33615 Bielefeld, Germany*
[b] *Institut für Genomforschung, Universität Bielefeld, Universitätsstraße 25, D-33615, Bielefeld, Germany*
[c] *Lehrstuhl für Biophysik und angewandte Nanowissenschaften, Universität Bielefeld, Universitätsstraße 25, D-33615 Bielefeld, Germany*

Received 15 December 2003; received in revised form 19 February 2004; accepted 19 March 2004

Abstract

The structural S-layer proteins of 28 different *Corynebacterium glutamicum* isolates have been analyzed systematically. Treatment of whole *C. glutamicum* cells with detergents resulted in the isolation of S-layer proteins with different apparent molecular masses, ranging in size from 55 to 66 kDa. The S-layer genes analyzed were characterized by coding regions ranging from 1473 to 1533 nucleotides coding for S-layer proteins with a size of 490–510 amino acids. Using PCR techniques, the corresponding S-layer genes of the 28 *C. glutamicum* isolates were all cloned and sequenced. The deduced amino acid sequences of the S-layer proteins showed identities between 69 and 98% and could be grouped into five phylogenetic classes. Furthermore, sequence analyses indicated that the S-layer proteins of the analyzed *C. glutamicum* isolates exhibit a mosaic structure of highly conserved and highly variable regions. Several conserved regions were assumed to play a key role in the formation of the *C. glutamicum* S-layers. Especially the N-terminal signal peptides and the C-terminal anchor sequences of the S-layer proteins showed a nearly perfect amino acid sequence conservation. Analyses by atomic force microscopy revealed a committed hexagonal structure. Morphological diversity of the *C. glutamicum* S-layers was observed in a class-specific unit cell dimension (ranging from 15.2 to 17.4 nm), which correlates with the sequence similarity-based classification. It could be demonstrated that differences in the primary structure of the S-layer proteins were reflected by the S-layer morphology.

Keywords: *Corynebacterium glutamicum*; Cell envelope; Surface-layer; Hexagonal symmetry; Unit cell dimension; Diversity; S-layer protein classification; Atomic force microscopy

1. Introduction

Paracrystalline cell surface-layers (S-layers) are one of the most common surface structures in *Bacteria* and *Archaea* (Beveridge, 1994; Sleytr and Beveridge, 1999). S-layers are generally composed of single (glyco-)protein species which greatly vary

* Corresponding author. Fax: +49-521-106-5626.
E-mail address: joern.kalinowski@genetik.uni-bielefeld.de (J. Kalinowski).

within their molecular masses from 40 to 200 kDa in different bacterial genera. The S-layer protein can represent up to 15% of the total protein content of the bacterial cell (Boot and Pouwels, 1996). Following transport across the membrane, S-layer proteins assemble into two-dimensional arrays with oblique, square or hexagonal symmetry on the cellular surface (Sleytr and Messner, 1983). Assembly of S-layers is an entropy-driven process in which all information for crystallization into a regular array resides within the monomers. Because of their location, S-layers are involved in the interaction between bacteria and their environment. Therefore, diverse functions have been attributed to the S-layers of individual bacterial species (Beveridge et al., 1997), including protection of the cell from hostile factors, serving as molecular sieves (Sara and Sleytr, 1987) and mediation of attachment to host tissues (Schneitz et al., 1993). Nowadays, a broad spectrum of biotechnological applications involves the use of S-layer lattices in suspension or attached to supports as immobilization matrices for binding monolayers of functional molecules, like antibodies and ligands (Sara and Sleytr, 1996b; Sleytr and Sara, 1997).

Most S-layer proteins are acidic (pI 4–6) and contain about 40% hydrophobic amino acids and very few sulfur-containing amino acids (Sleytr, 1997). They differ markedly in their primary structures. Only very low levels of amino acid similarity have been found between S-layer proteins of different species. From secondary structure data of S-layer proteins known to date it was concluded that S-layer proteins are in general composed of β-sheets with a minor portion of α-helices. How these secondary structure components are organized into tertiary structures is only known in a few instances (Engelhardt and Peters, 1998). An example is the filamentous archaebacterial surface protein of *Staphylothermus marinus* (Peters et al., 1996, 1995). Electron microscopy and more recently also atomic force microscopy (AFM) has been applied to obtain further structural information of S-layers down to the nanometer range (Engel and Müller, 2000; Müller et al., 1999).

Recently, AFM techniques were applied to analyze the S-layer of the Gram-positive bacterium *Corynebacterium glutamicum* ATCC 17965 (Scheuring et al., 2002). *C. glutamicum* is well known for its capacity to produce large quantities of amino acids and it is used for industrial fermentation processes of L-glutamate and L-lysine (Hermann, 2003). It belongs to the *Corynebacterineae*, a supragenic group of bacteria that includes corynebacteria, mycobacteria, nocardia, and rhodococci (Stackebrandt et al., 1997). *C. glutamicum* shares with each other genera of this group a specific cell wall structure mainly characterized by the presence of mycolic acids (reviewed in Bayan et al., 2003). The S-layer of *C. glutamicum* ATCC 17965 is strongly associated with the cell wall and thus also part of the cell envelope.

The S-layer of the *C. glutamicum* ATCC 17965 is formed by the PS2 protein which is encoded by the *cspB* gene (Peyret et al., 1993). The mature PS2 protein has a molecular mass of 52.5 kDa, is devoid of any sulfur-containing amino acid and is characterized by a high content of hydrophobic amino acid residues (Peyret et al., 1993). The PS2 protein of *C. glutamicum* ATCC 17965 exhibits no similarities to any other protein in the EMBL database (Peyret et al., 1993). The hydrophobic C-terminus of the PS2 protein was found to be involved in the attachment of PS2 to the cell wall (Chami et al., 1997). The S-layer of *C. glutamicum* ATCC 17965 is characterized by a hexagonal lattice symmetry (Chami et al., 1995). AFM analyses of S-layer sheets indicate that monomers from hexameric cores are connected to six other cores (Scheuring et al., 2002). Accordingly, the S-layer of *C. glutamicum* ATCC 17965 was classified as an M_6C_3 layer type (Saxton and Baumeister, 1986; Scheuring et al., 2002).

In order to gain a deeper understanding of the *C. glutamicum* S-layer and its structural features, comparative studies of similar S-layer proteins were performed. In this report, we present the cloning and sequencing of 28 *cspB* genes from different *C. glutamicum* isolates. Based on nucleotide sequence and protein analysis coupled with atomic force microscopy, the *C. glutamicum* S-layers were classified into five groups. This classification is a first step in identifying conserved protein regions which are involved in the formation of *C. glutamicum* S-layers.

2. Materials and methods

2.1. Bacterial strains and growth conditions

C. glutamicum isolates used in this study were obtained from the American Type Culture Collection

(ATCC; Manassas, VA) and from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ; Braunschweig, Germany), with the exception of *C. glutamicum* 22220 and 22243 (Takeda et al., 1990) and *C. glutamicum* LP-6 (Sonnen et al., 1991). *Escherichia coli* TOP10 (Invitrogen, Karlsruhe, Germany) and *E. coli* DH5αMCR (Grant et al., 1990) were used in cloning experiments and were routinely grown at 37 °C on solid Antibiotic Medium No. 3 (Oxoid, Wesel, Germany). Selection for the presence of cloning vectors in *E. coli* was performed by the addition of 50 μg ml^{-1} kanamycin. *C. glutamicum* isolates were cultivated in minimal medium 1 (MMYE without yeast extract; Katsumata et al., 1984) at 30 °C. Minimal medium 1 was prepared with 2% acetate and 30 mg l^{-1} protocatechuic acid instead of glucose where required.

2.2. Extraction of S-layer proteins from C. glutamicum

S-layers were extracted according to a protocol of Peyret et al. (1993) with minor modifications. Briefly, 1.2×10^9 *C. glutamicum* cells were harvested by centrifugation at 5000 × *g* for 5 min. The cell pellet was resuspended in 500 μl of 50 mM Tris–HCl buffer (pH 6.8) containing 2% SDS. Subsequently, the cell suspension was incubated with intensive shaking (300 rpm) at room temperature for 1 h and was then centrifuged again at 5000 × *g* for 5 min. This experimental treatment of the cells does not result in the solubilization of the cytoplasmic membrane or in cell lysis (Peyret et al., 1993).

2.3. Separation and identification of PS2 proteins by SDS–PAGE and MALDI-TOF-MS

Proteins were separated by one-dimensional denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) with a 4% stacking gel and a 12.5% resolving gel. Samples were denatured in the presence of 2% SDS and 4% β-mercaptoethanol in 60 mM Tris–HCl buffer (pH 6.8) by heating to 100 °C for 5 min. Molecular masses were determined using the Unstained Broad Range Precision Protein Standard (Bio-Rad Laboratories, Hercules, CA). Following electrophoresis, gels were stained with Coomassie brilliant blue R-250 and G-250 as described by Sambrook et al. (1989) and briefly treated with 7% acetic acid to visualize protein bands.

For the identification of proteins, the protein bands were excised from fresh Coomassie-stained 12.5% SDS–PAGE. The protocols of Hermann et al. (2001) for tryptic digestions and MALDI-TOF-MS analyses were applied to generate peptide mass fingerprints. The Bruker Biflex III MALDI TOF mass spectrometer (Bruker Daltonik, Bremen, Germany) was used to generate mass spectra. Peptide mass fingerprints were compared with in silico-generated tryptic fingerprints using the MASCOT software (Perkins et al., 1999).

2.4. PCR amplification, cloning and sequencing of C. glutamicum S-layer genes

Routine manipulation of DNA followed the protocols of Sambrook et al. (1989). The *cspB* genes of *C. glutamicum* isolates were amplified using combinations of the forward primers pcspB1-4 and reverse primers pcspB5-7 (Table 1). PCR experiments were carried out with a PTC-100 thermocycler (MJ Research, Watertown, MA) and *Pfx* DNA polymerase (Invitrogen). Initial denaturation was conducted at 94 °C for 2 min followed by denaturation for 30 s, annealing for 35 s at a primer-dependent temperature and extension at 68 °C for 90 s. This cycle was repeated 30 times, followed by a final extension step at 68 °C for 4 min. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). The PCR products were cloned into the vector pZERO-2 or into pCR-Blunt II using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). Sequencing of cloned PCR products was performed by IIT Biotech (Bielefeld, Germany). The nucleotide sequences of

Table 1
Oligonucleotides used as PCR primers to amplify *cspB* genes of *C. glutamicum* isolates

Name	Sequence
pcspB1	CCATGTCGTGATCAGCCATT
pcspB2	GATACTGGTACCTATGCGCTTCAGAGCTTC
pcspB3	GATAAGGAATTCGTTCGGTGGCCTAGTGAG
pcspB4	GAAGTCTGACGTTCTGGTTCG
pcspB5	GATACGTCTAGATGGTGCTGTTAGCCAAGG
pcspB6	GGTTGCACGCTCGAGGAAGA
pcspB7	GCCAGTGGCGAGTCATTAAG

Table 2
Characteristics of S-layer proteins and their respective genes of *C. glutamicum* isolates

Number	*C. glutamicum* isolate	Length of *cspB* coding region (bp)	G + C content of *cspB* coding region (%)	Expression level of *cspB* coding region[a]	Nucleotide identity to *cspB* of ATCC 17965[b] (%)	Length of PS2 protein (aa)	Theoretical molecular mass of processed PS2 (kDa)	Theoretical IEP of processed PS2	No. of sulfur-containing amino acids of processed PS2 [Cys+Met]	Length of C-terminal anchor (aa)	Amino acid identity to PS2 of ATCC 17965[b] (%)	GenBank accession number
1	ATCC 13058	1482	55.2	0.22	77.2	493	50.6	4.22	0 + 0	21	70.8	AY524990
2	ATCC 13744	1497	54.2	0.22	75.8	498	51.0	4.21	1 + 0	21	69.2	AY524991
3	ATCC 13745	1491	54.9	0.23	77.2	496	50.7	4.14	0 + 0	21	71.4	AY524992
4	ATCC 14017	1491	55.6	0.24	81.5	496	50.9	4.20	0 + 0	21	77.2	AY524993
5	ATCC 14020	1524	55.5	0.25	84.7	507	51.9	4.20	0 + 0	21	81.0	AY525009
6	ATCC 14067	1476	55.0	0.24	82.0	491	50.5	4.19	0 + 0	21	77.4	AY524994
7	ATCC 14068	1524	55.6	0.26	84.9	507	52.0	4.20	0 + 0	21	81.1	AY525010
8	ATCC 14747	1497	55.2	0.23	75.9	498	50.9	4.22	0 + 0	21	69.2	AY525011
9	ATCC 14751	1491	55.3	0.23	81.0	496	51.0	4.16	1 + 0	21	75.8	AY524995
10	ATCC 14752	1497	55.5	0.26	77.0	498	51.1	4.24	0 + 0	21	70.2	AY524996
11	ATCC 14915	1476	55.3	0.25	79.7	491	50.3	4.17	0 + 0	21	73.9	AY524997
12	ATCC 15243	1491	54.8	0.23	80.8	496	51.2	4.19	0 + 0	21	76.1	AY524998
13	ATCC 15354	1491	55.2	0.24	80.1	496	50.8	4.24	0 + 0	21	76.2	AY524999
14	ATCC 17965	1533	55.5	0.26	100.0	510	52.6	4.24	1 + 0	21	100.0	AY525000
15	ATCC 17966	1533	55.1	0.26	98.6	510	52.5	4.23	0 + 0	21	97.1	AY525001
16	ATCC 19223	1503	54.9	0.25	82.0	500	51.6	4.11	0 + 1	21	76.4	AY525002
17	ATCC 19240	1488	54.9	0.22	77.4	495	50.9	4.14	0 + 1	21	71.6	AY525012
18	ATCC 21341	1491	55.7	0.23	81.0	496	51.2	4.16	0 + 2	21	77.5	AY525003
19	ATCC 21645	1497	54.8	0.22	73.0	498	51.0	4.27	0 + 0	21	68.9	AY525004
20	ATCC 31808	1482	55.1	0.18	77.1	493	50.6	4.20	0 + 0	21	69.5	AY525013
21	ATCC 31830	1473	55.0	0.22	77.2	490	49.9	4.22	0 + 0	14	71.3	AY525007
22	ATCC 31832	1533	55.3	0.27	99.1	510	52.3	4.23	0 + 0	21	98.0	AY525008
23	LP-6	1503	55.0	0.25	82.5	500	51.6	4.12	0 + 0	21	76.8	AY525014
24	DSM 20137	1533	55.6	0.27	99.3	510	52.5	4.23	0 + 0	21	98.4	AY525015
25	DSM 20598	1482	54.8	0.23	81.8	493	50.9	4.23	1 + 0	21	76.6	AY525016
26	DSM 46307	1497	55.4	0.22	76.8	498	51.1	4.14	0 + 0	21	70.4	AY525017
27	22220	1500	54.6	0.24	86.6	499	51.4	4.17	0 + 0	21	80.5	AY525005
28	22243	1527	55.4	0.24	91.3	508	52.5	4.21	0 + 0	21	87.8	AY525006

[a] The expression levels of the *cspB* coding regions were calculated by the Cobias program (McHardy et al., 2004).
[b] The calculation of nucleotide and amino acid sequence identity is based on the *cspB* gene from ATCC 17965 determined in this study. The *cspB* sequence shows 0.7% difference to that published by Peyret et al. (1993).

the *cspB* genes from *C. glutamicum* isolates were deposited in the EMBL database. Accession numbers are listed in Table 2.

2.5. Bioinformatics tools used to analyze nucleotide and amino acid sequences

Nucleotide sequences were assembled with the STADEN software package (Staden, 1996). Database searches were performed with the BLAST programs (Altschul et al., 1997). Comparisons of DNA sequences and protein sequences were done by the SSEARCH program, using the Smith–Waterman algorithm (Smith and Waterman, 1981). Global nucleotide and amino acid sequence alignments were carried out with the CLUSTALX program (Thompson et al., 1997). Phylogenetic trees were calculated using the neighbor-joining method (Saitou and Nei, 1987) integrated in the CLUSTALX software package. Signal peptides were determined with SignalP for Gram-positive bacteria, using the neural network and the Hidden–Markov model methods (Nielsen et al., 1997). Secondary structure predictions were performed by GOR (Garnier et al., 1996). The analysis of hydrophobicity patterns was performed by ProtScale (Kyte and Doolittle, 1982). To measure expression level-dependent features in the synonymous codon usage of the gene sequence *g*, the CoBias program (McHardy et al., 2004) was used with a log-odds ratio scoring matrix reflecting the differences between highly and non-highly expressed genes in the synonymous codon usage of *C. glutamicum*. P(B|*g*) is an estimate of the significance of the observed feature in the gene sequence; the lower the value of P(B|*g*), the more significant are the expression-level dependent features in the codon usage of *g*. The strength of the feature which may be related to the expression level of a gene can be assessed using the values of S(*g*)/L.

2.6. Atomic force microscopy of S-layer proteins

Mica surfaces (Provac AG, Balzers, Liechtenstein) were silanized with aminopropyl-triethoxysilane in an exsiccator (Lyubchenko et al., 1993) and incubated with 5 μl of S-layer solution (∼0.5 mg/ml in 10 mM Tris–HCl, 2% SDS, pH 6.8) for 1 h at room temperature. The sample then was carefully rinsed with 1–2 ml of milliQ deionized water (Millipore) to remove weakly attached proteins and dried under a gentle flow of nitrogen gas. Imaging was performed with non-contact silicon probes (NanoSensors, Wetzlar, Germany) on a Nanoscope IIIa AFM system equipped with a Multimode head and a type E piezoelectric scanner (Multimode, Veeco Instruments, Santa Barbara, CA, USA). The AFM was operated in tapping mode (Möller et al., 1999) at a scan line frequency of 1–2 Hz. Raw AFM images have been processed only for background removal (flattening) by using the microscope manufacturer's image-processing software. Image analysis by Fast Fourier Transforms (FFT) was performed with the WSxM 2.0 SPM software (Nanotec, Madrid, Spain).

3. Results

3.1. Identification of surface-layer proteins of 28 different C. glutamicum isolates

It has been reported previously that the soil isolate *C. glutamicum* ATCC 17965 harbors an hexagonal S-layer (Chami et al., 1997; Peyret et al., 1993), but the existence of S-layer proteins in other *C. glutamicum* strains has not been investigated systematically to date. For this reason, we cultivated 28 different *C. glutamicum* isolates obtained from culture collections in minimal medium 1 supplemented with 2% glucose as sole carbon source and performed S-layer extractions according to the protocol by Peyret et al. (1993) using chaotropic detergents. The extracted S-layers were then visualized in SDS–PAGE (Fig. 1), whereupon the S-layer isolation of *C. glutamicum* ATCC 17965 served as positive control. The S-layer extractions from 26 out of the 28 *C. glutamicum* isolates exhibited one major band in SDS–PAGE (Fig. 1A), whereas S-layer extractions of *C. glutamicum* ATCC 19223 and *C. glutamicum* LP-6 yielded no products under the applied conditions (Fig. 1A; lanes 16 and 23). The apparent molecular masses of the putative S-layer proteins varied between 55 and 66 kDa. After tryptic digestion, MALDI-TOF-MS analyses of the putative S-layer proteins were performed to obtain peptide mass fingerprints, which were used in MASCOT searches against the published amino acid sequence of the PS2 protein from *C. glutamicum* ATCC 17965 (Peyret et al., 1993). Using this method, the

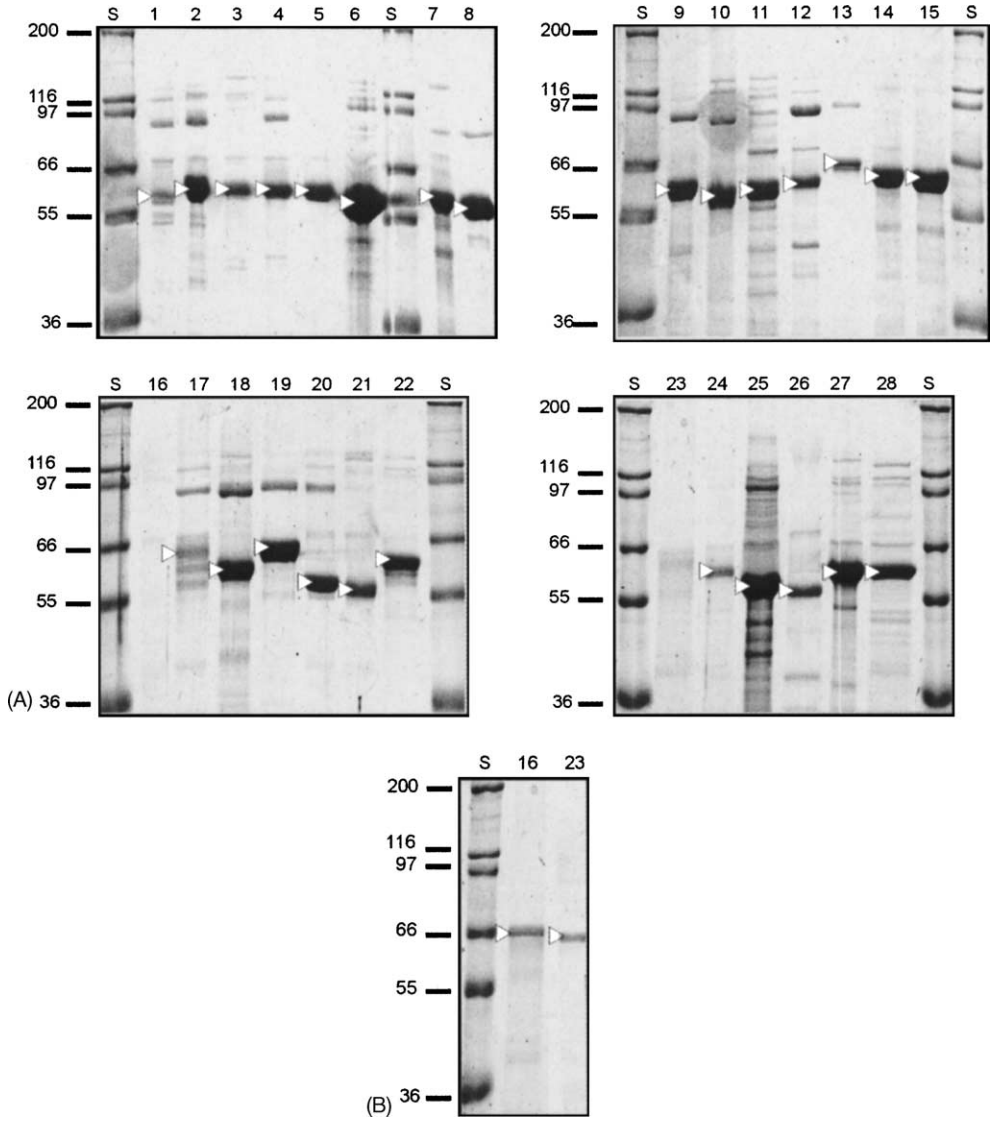

Fig. 1. Coomassie blue-stained SDS–PAGE of cell envelope protein fractions from 28 different *C. glutamicum* isolates. (A) S-layer analysis from *C. glutamicum* isolates cultivated in minimal medium 1 with glucose as sole carbon source. The arrows indicate the S-layer proteins verified by MALDI-TOF-MS analyses. The S-layer isolation of *C. glutamicum* ATCC 17965 (lane 14) served as a reference. (B) S-layer analysis from the two *C. glutamicum* isolates ATCC 19223 and LP-6 grown in minimal medium 1 with acetate as sole carbon source. Key: lane S, molecular weight standard (kDa); lane 1–28, *C. glutamicum* isolates listed in Table 2.

prominent protein bands of the four *C. glutamicum* isolates ATCC 17965, ATCC 17966, DSM 20137, and ATCC 31832 were unequivocally identified as PS2 homologous proteins. Peptide mass fingerprints of the remaining major protein bands revealed neither highly significant hits to the PS2 from *C. glutamicum* ATCC 17965 nor to other protein sequences deposited in public databases.

As shown earlier, the S-layer production of *C. glutamicum* isolates may depend on the carbon source used for cultivation (Soual-Hoebeke et al., 1999). To test whether different carbon sources stimulate

the expression of S-layer proteins in *C. glutamicum* ATCC 19223 and *C. glutamicum* LP-6, we cultivated both strains in minimal medium 1 supplemented with acetate as sole carbon source. Interestingly, the S-layer extractions of both strains showed distinct bands with molecular masses of around 60 kDa in SDS–PAGE (Fig. 1B). This indicates that in both *C. glutamicum* ATCC 19223 and *C. glutamicum* LP-6 the expression of S-layer proteins is stimulated by acetate. SDS–PAGE of S-layer extractions from other *C. glutamicum* isolates revealed no differences in S-layer expression during growth on acetate as carbon source compared to the growth on glucose (data not shown). It is interesting to note that neither the complete genome sequence of *C. glutamicum* ATCC 13032 (Kalinowski et al., 2003; Ikeda and Nakagawa, 2003) nor one of the other corynebacterial genomes (*C. efficiens* (Nishio et al., 2003); *C. diphtheriae* (Cerdeno-Tarraga et al., 2003)) contain a *cspB* gene (data not shown). Based on the observation that only four of the isolated *C. glutamicum* S-layer proteins were identifiable as PS2 proteins with high significance by MASCOT analyses and that the molecular masses of the proteins varied between 55 and 66 kDa, it can be assumed that significant differences in the corresponding genes are to be expected.

3.2. Sequence analysis of S-layer genes isolated from 28 different C. glutamicum isolates

To determine the nucleotide sequences of the S-layer genes from the different *C. glutamicum* isolates, oligonucleotide primers based on the known *cspB* gene sequence of *C. glutamicum* ATCC 17965 were designed (Table 1). The DNA regions used for the design of seven PCR primers were located upstream of the initiation codon, in the central part of the gene and downstream of the termination codon of *cspB* (Fig. 2). Five primer combinations were successfully applied to obtain PCR products for all *C. glutamicum* isolates, indicating the conservation of the primer sequences. These PCR amplifications yielded fragments of a length from ~800 to ~1100 bp. The PCR fragments, named pcspBl-6, pcspB2-6, pcspB3-6, pcspB4-5, and pcspB4-7 (Fig. 2), were cloned in *E. coli* into the vectors pZERO-2 and pCR-Blunt II and completely sequenced. General features deduced from this set of gene sequences are summarized in Table 2. All predicted open reading frames (ORFs) were clearly identified as *cspB* homologs by similarity searches, showing 73–99% nucleotide sequence identity with the *cspB* gene of *C. glutamicum* ATCC 17965. Based on these similarity

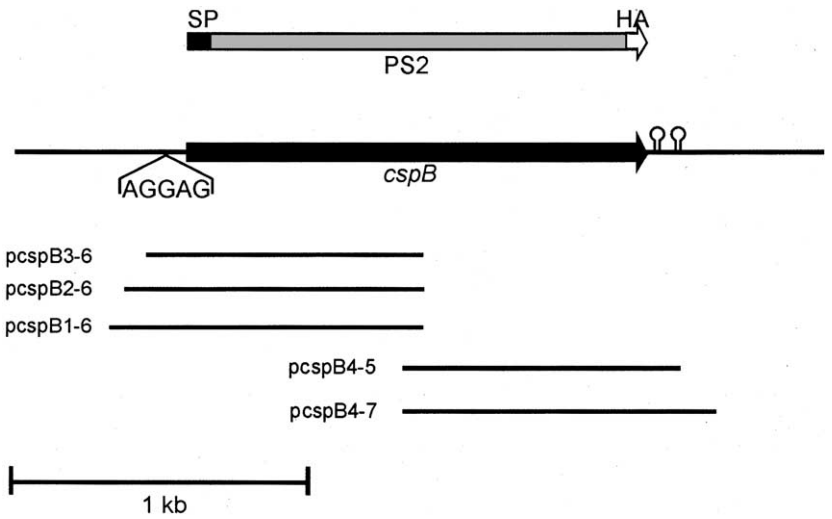

Fig. 2. Schematic representation of the 2.7-kb *cspB* gene region of *C. glutamicum* ATCC 17965. The putative Shine-Dalgarno sequence (AGGAG) and two putative transcriptional terminators are indicated. The positions of amplified PCR products pcspBl-6, pcspB2-6, pcspB3-6, pcspB4-5, and pcspB4-7 used for DNA sequencing are shown. The expressed PS2 proteins contain an N-terminal signal peptide (SP) and a C-terminal hydrophobic anchor sequence (HA) (Peyret et al., 1993; Chami et al., 1995).

data and additional multiple alignments, the *cspB* homologs of the different *C. glutamicum* isolates were designated as *cspB* genes at this point.

Analysis of the 28 sequences revealed one complete ORF characterized by a varying length between 1473 and 1533 nucleotides, an ATG start codon and the termination codon TAA. The coding regions of the *cspB* genes have also very similar G + C contents, ranging from 54.2 to 55.7%. This is very close to the mean G + C content of the *C. glutamicum* ATCC 13032 genome sequence (Kalinowski et al., 2003). Based on the *C. glutamicum* codon usage, predictions of the *cspB* gene expression levels were performed. To measure expression level-dependent features in the synonymous codon usage of a gene sequence, the CoBias program was used with a log-odds ratio scoring matrix reflecting the differences between highly and not-highly expressed genes in the synonymous codon usage of *C. glutamicum* (McHardy et al., 2004). All analyzed *cspB* genes showed very high $S_{Av}(g)$ values (S (*g*)/L > 0.18; Table 2), which are characteristic for highly expressed genes. The significance levels of the analyzed *cspB* genes show very low P(B/*g*) values (P (B/*g*) < $1e^{-38}$), corresponding to high significance of the observed feature. These bioinformatics data, showing very high expression levels of the *cspB* genes from *C. glutamicum*, coincide well with the initial detection of prominent S-layer protein bands in SDS–PAGE (Fig. 1).

At the nucleotide level, sequence alignments revealed that the upstream and downstream region of the 28 analyzed *cspB* genes is highly conserved. All *cspB* genes are preceded by a putative Shine-Dalgarno sequence (AGGAG) located 12 or 13 bp in front of the assigned ATG initiation codon. Two highly conserved putative transcriptional terminators were found in the downstream region of the S-layer genes. The first transcriptional terminator is located 6 bp downstream of the *cspB* stop codon, its stem-loop structure possesses a calculated ΔG of $-68.3\,kJ\,mol^{-1}$. The second transcriptional terminator is located further 33 bp downstream. Variations in the loop size of this terminator were found in the different *C. glutamicum* isolates (data not shown). Downstream of the *cspB* locus, an ORF with no apparently related function to S-layer formation was identified in all *C. glutamicum* isolates in the opposite strand. The amino acid sequences deduced from this ORF are identical and display substantial similarity (64% positives, E-value: $5e^{-43}$) to a Zn-dependent class III alcohol dehydrogenase from *C. glutamicum* ATCC 13032 (CAF21131; Kalinowski et al., 2003).

In contrast to their highly conserved surrounding, the *cspB* coding regions themselves seem to be highly variable. Translation of the 28 *cspB* genes revealed PS2 proteins with lengths ranging from 490 to 510 amino acids. The calculated molecular masses of the processed PS2 homologs vary between 49.9 and 52.6 kDa (Table 2). The deduced PS2 proteins all show a predicted weak acidic isoelectric point between 4.11 and 4.27 resulting from a high content of acidic amino acids (26.1–28.0%) as compared to basic amino acids (7.2–8.3%). Moreover, the PS2 proteins contain around 50% hydrophobic amino acids. Additionally, sequence analysis of all PS2 homologs shows a high content of alanine residues (18.5–19.8%), exceedingly localized in the N- and C-terminal regions of the PS2 protein. As a feature common with other S-layer proteins from different bacteria and PS2 from *C. glutamicum* ATCC 17965 the newly identified S-layer proteins contain only a few sulfur-containing amino acids (Table 2).

In order to verify that the identified S-layer proteins are PS2 homologs, the PS2 protein sequences were digested in silico with trypsin, and the resulting peptides were used to establish a PS2-specific MASCOT database. Using this database it could be confirmed that all of the prominent bands represented in SDS–PAGE (Fig. 1) are indeed PS2 homologs (data not shown). The previously observed diversity of the PS2 proteins in the SDS–PAGE is thus reflected by differences in the nucleotide sequence of the corresponding *cspB* genes.

3.3. Classification of C. glutamicum S-layer proteins based on sequence similarities

Multiple alignments (Thompson et al., 1994) of the S-layer proteins from the analyzed *C. glutamicum* isolates showed highly conserved regions in the amino acid sequence of the respective PS2 proteins (Fig. 3). The N-terminal signal peptide of PS2 having a length of 30 amino acid residues is nearly identical in all isolates. A high degree of conservation was also found in the C-terminal anchor sequence (Fig. 3; Table 2). However, the anchor sequence of the PS2 protein of

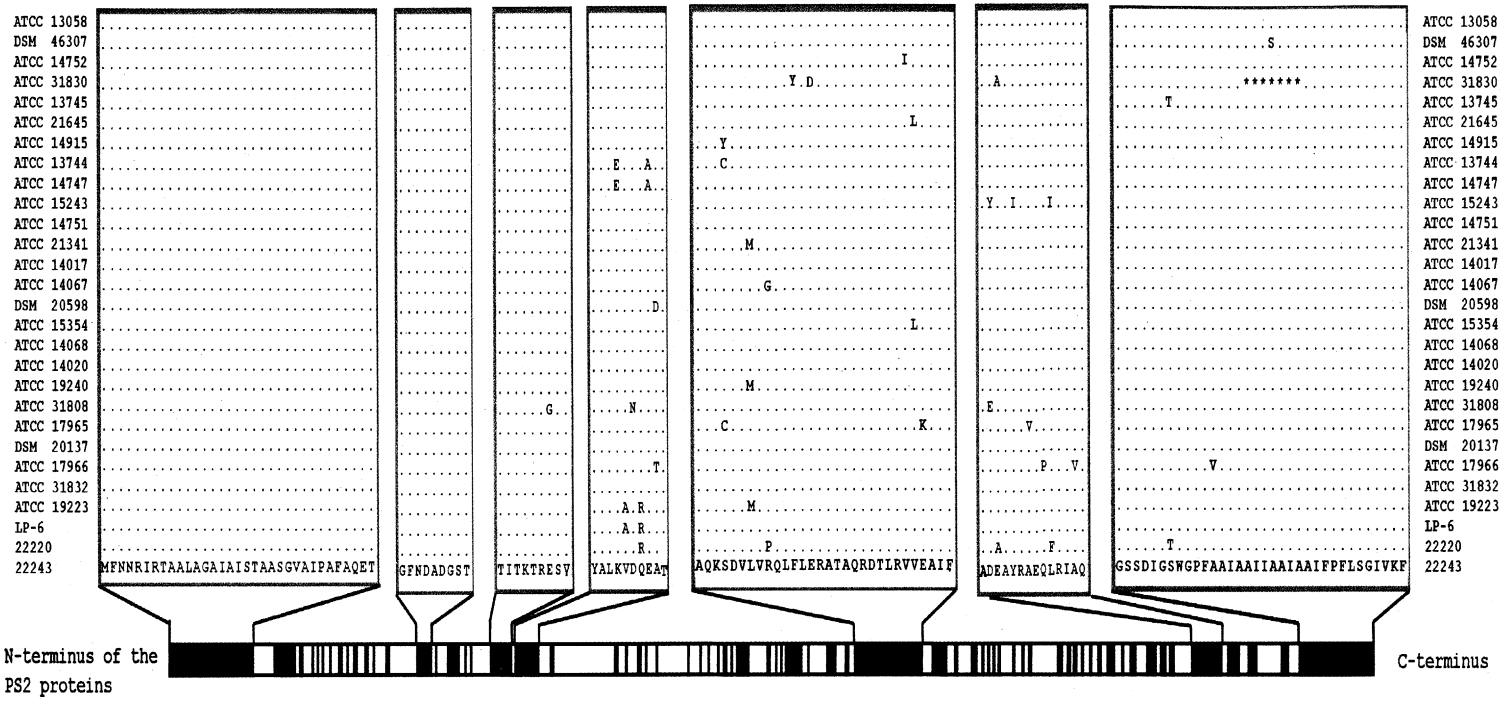

Fig. 3. Conserved amino acid regions of the 28 analyzed *C. glutamicum* S-layer proteins. Highly conserved amino acid regions are schematically indicated with black boxes in a linear PS2 protein sequence. Additionally, alignments of highly conserved regions larger than eight amino acid residues are shown in detail. The amino acid sequence of PS2 from *C. glutamicum* 22243 is used as reference and shown at the bottom. Amino acid differences are depicted, gaps are indicated by asterisks.

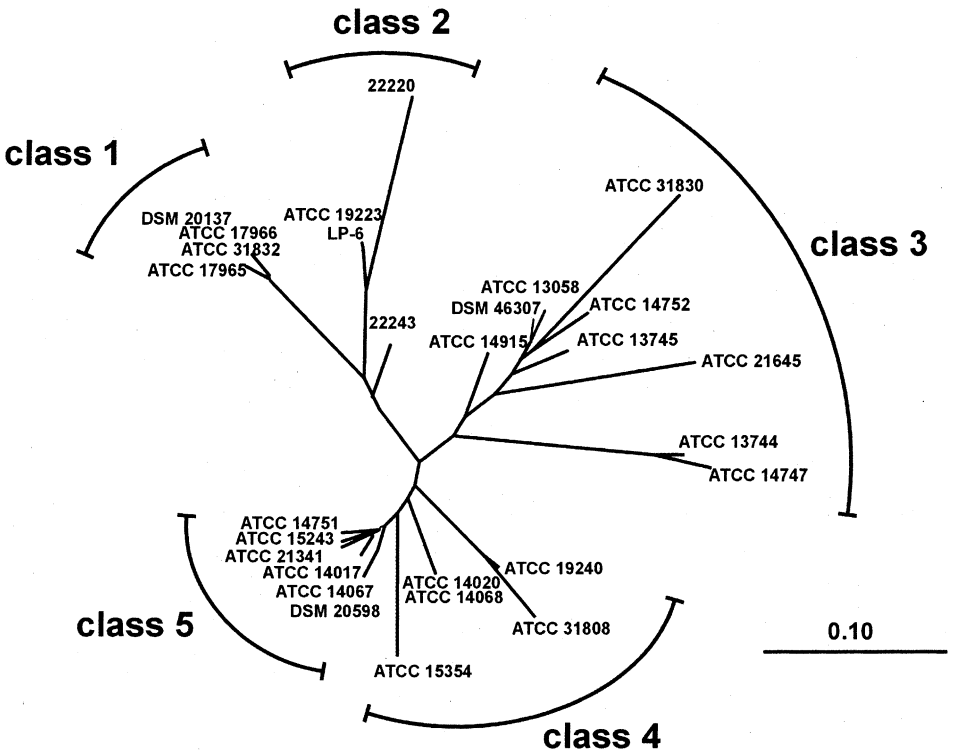

Fig. 4. Dendrogram depicting the relationship among the different PS2 protein sequences from *C. glutamicum* isolates. The diagram was constructed with the CLUSTALX program (Thompson et al., 1994) under usage of the BLOSUM 62 matrix with the neighbour-joining method (version 2.2). The classification of the PS2 proteins was preformed according to their sequence conservation. Scale bar: 0.1% amino acid substitution.

C. glutamicum ATCC 31830 only spans 14 amino acid residues, whereas the anchor sequences of all other PS2 proteins possess a length of 21 amino acids. In addition to these functionally known regions of PS2, several other stretches with nearly 100% identity were identified among the different class members (Fig. 3). The biological function of these conserved protein regions are hitherto unknown.

On the basis of these alignments, the PS2 proteins of the 28 studied *C. glutamicum* isolates were divided into five different classes (Fig. 4). All S-layer proteins, which were initially identified as PS2 proteins by MASCOT searches against the PS2 sequence from *C. glutamicum* ATCC 17965 belong to class 1 (Fig. 4). This is obvious considering the high percentage of amino acid sequence identity between these proteins (Table 2). The S-layer proteins of the *C. glutamicum* isolates ATCC 19223 and LP-6, which were expressed during growth on acetate as sole carbon source, both belong to class 2 (Fig. 4). It has to be noted that the S-layer proteins within the classes 1 and 5 appear to be closely related to each other (>93% amino acid identity), while the S-layer proteins within the classes 2–4 exhibit more differences when compared to other members of the same class (∼80% amino acid identity).

Among the *C. glutamicum* strains the proportional identity of the PS2 proteins amounts from 69 to 98%, which is a high degree of variability for the same protein of the same species. An alignment of representative members of each class showed the pattern of extremely variable and conserved regions within the amino acid sequences of the PS2 proteins (Fig. 5). Most of the highly variable regions were indeed specific for a distinct protein class (data not shown). For further analysis of the *C. glutamicum* PS2 proteins,

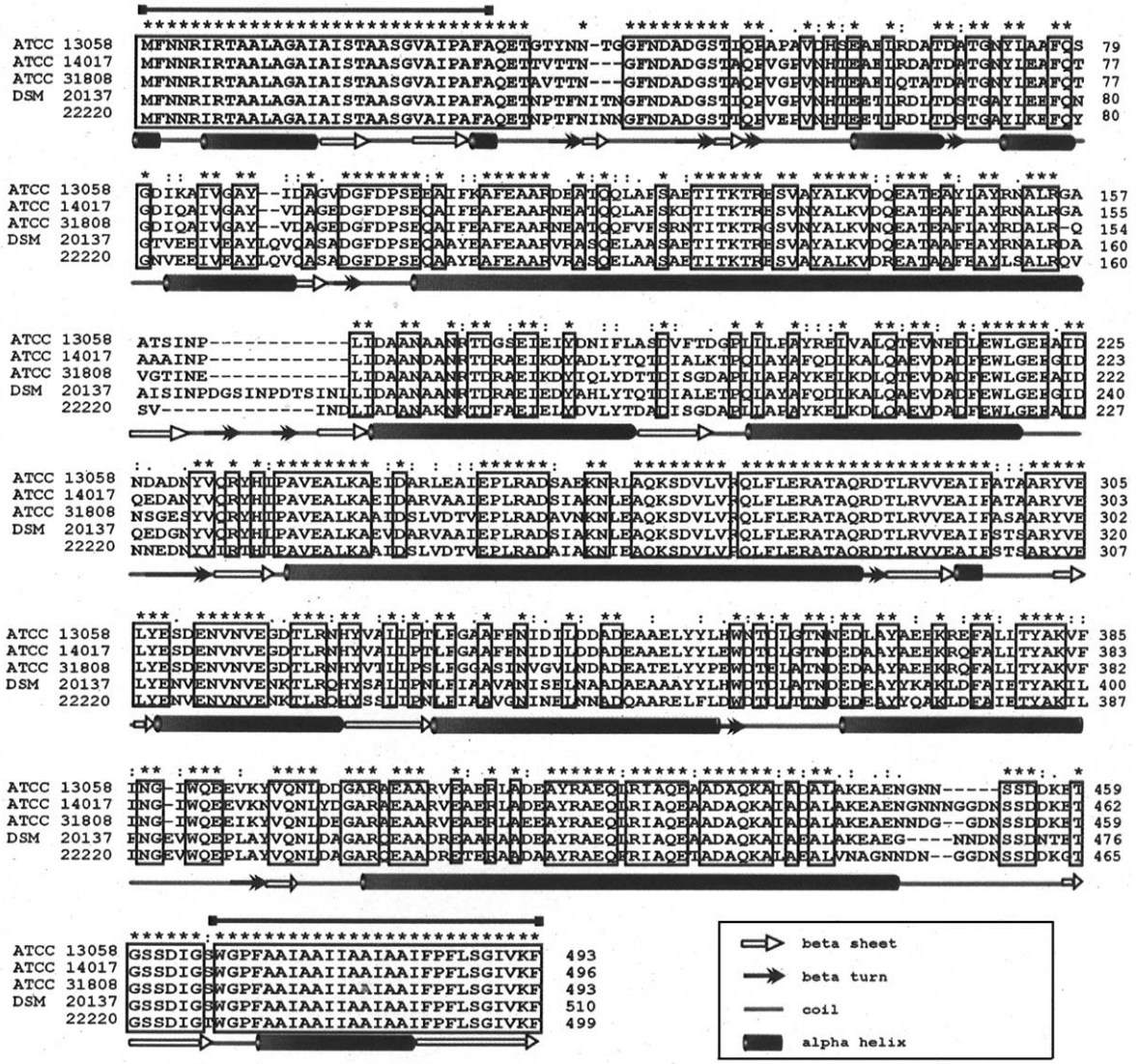

Fig. 5. Alignment of the deduced PS2 amino acid sequences from *C. glutamicum* isolates representing the five different S-layer classes. The alignment was produced with CLUSTALX (Thompson et al., 1994), using the BLOSUM 62 matrix, a gap opening penalty of 10 and a gap extension penalty of 0.1. Amino acid regions with perfect identity are enclosed in boxes and additionally marked with asterisks. The brackets indicate the potential N-terminal signal peptide sequence and the C-terminal hydrophobic anchor sequence. A secondary structure prediction is shown exemplarily for the PS2 proteins of *C. glutamicum* DSM 20137.

the secondary structure prediction tool GOR (Garnier et al., 1996) was applied (Fig. 5). All *C. glutamicum* PS2 proteins turned out to contain a large amount of α-helices. Furthermore, the secondary structure predictions indicate a similar arrangement of secondary structural elements for all *C. glutamicum* PS2 proteins. The PS2 proteins of classes 1–5 appeared to be variable only in the specific length of the secondary structural elements but not in their composition. Hydrophobicity plots of the *C. glutamicum* PS2 proteins calculated according to Kyte and Doolittle (1982) showed that the proteins consist of alternating stretches of hydrophilic and hydrophobic amino acid residues. Interestingly, the highly conserved regions

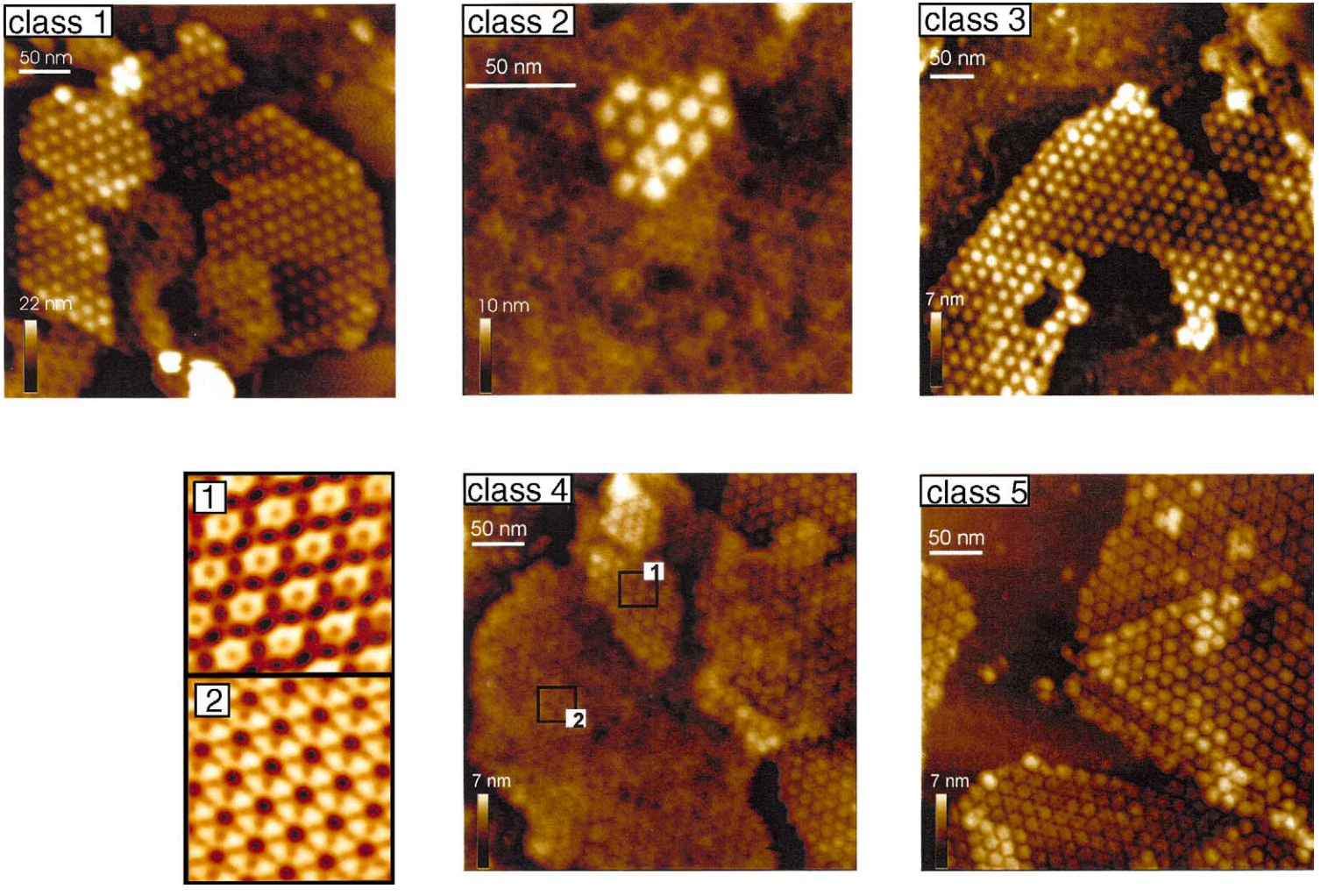

Fig. 6. AFM images of *C. glutamicum* S-layers from five different sequence similarity-based classes adsorbed to silanized mica. All strains display hexagonal lattices with flower-shaped bottom surfaces and triangular top surfaces, but with different unit cell dimensions. Fourier transformations of the bottom (1) and the top (2) sides of the S-layer of *C. glutamicum* ATCC 19240 are shown. AFM-images represent S-layers of each class: class 1 S-layer of ATCC 17966, class 2 S-layer of 22243, class 3 S-layer of ATCC 13058, class 4 S-layer ATCC 19240, and class 5 S-layer of ATCC 14751.

were either extreme hydrophilic or extreme hydrophobic (data not shown). In conclusion, sequence alignments, secondary structure predictions and hydrophobicity determinations revealed that the PS2 proteins from *C. glutamicum* isolates possess highly conserved structural elements, which are alternating hydrophilic and hydrophobic.

3.4. The C. glutamicum S-layers of the five identified classes are characterized by hexagonal structures with a variable unit cell dimension

To assess whether the sequence differences of the *C. glutamicum* PS2 proteins are also reflected by their S-layer morphology, atomic force microscopy (AFM) analyses were performed. PS2 proteins of 28 *C. glutamicum* isolates were extracted from cells with 2% SDS and precipitated by centrifugation to enable imaging with AFM (Fig. 6). The S-layers of the different isolates adsorbed to silanized mica as mono, double or multi layers and were finally imaged in the absence of buffers. Only the S-layers of four *C. glutamicum* isolates, namely ATCC 21341, ATCC 19223, LP-6, and ATCC 14017 displayed no patches of S-layers under these conditions. All other analyzed S-layers from *C. glutamicum* isolates formed hexameric complexes within a hexagonal lattice (Fig. 6). The inner surface of the S-layers exhibits a flower-shaped appearance (Fig. 6.1), whereas the outer surface possesses a triangular surface structure (Fig. 6.2).

Detailed analysis of 24 *C. glutamicum* S-layers showed small but significant differences in their unit cell dimension, specifying the distance between the centers of two PS2 hexamers. The unit cell dimension varies between 15.2 ± 0.25 nm and 17.4 ± 0.2 nm (Fig. 7). The different unit cell dimensions could be a result of the different coil lengths and respective angles connecting the conserved regions of the PS2 proteins. According to these variations of the unit cell dimensions, the S-layers could be divided into five classes. The correlation between the AFM-based classification of the S-layers and the protein sequence-based classification is striking, although PS2 proteins from two isolates (*C. glutamicum* ATCC 31830 and ATCC 13744) do not fit well into this classification (Fig. 7). Therefore, class 3 might need to be further subdivided.

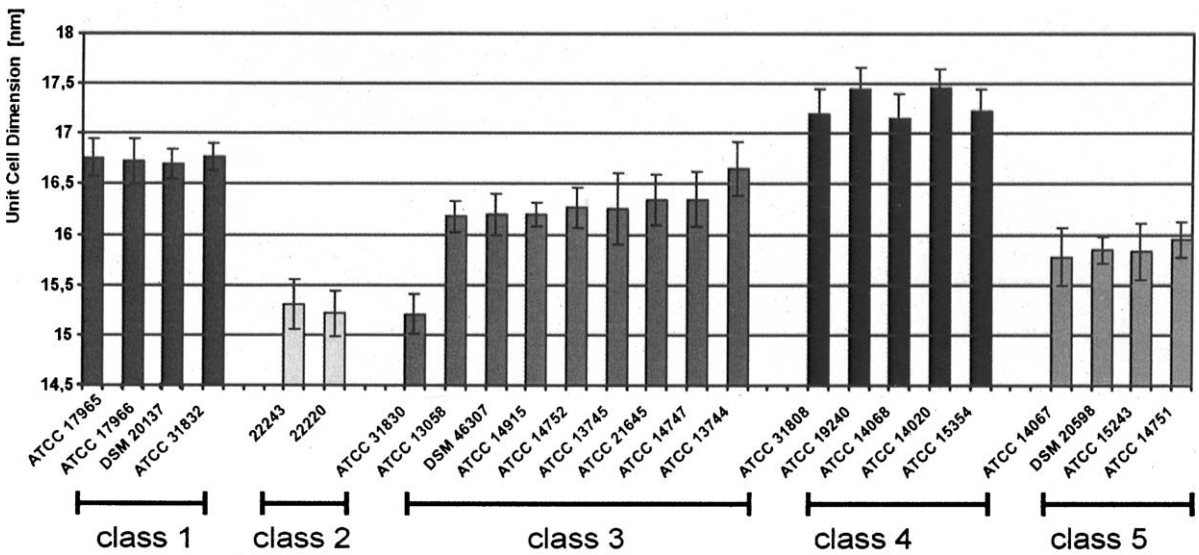

Fig. 7. Measurement of the unit cell dimension (UCD) of imaged S-layers. Each bar represents 180 independent measurements (three dimensions of 20 hexagons originating from three different images).

4. Discussion

4.1. Comparative sequence analyses revealed hyper-variable regions within the C. glutamicum S-layer proteins

In this study, we have sequenced and characterized the *cspB* genes of 28 different *C. glutamicum* isolates in order to elucidate functional elements of the respective S-layers. We observed that all analyzed *C. glutamicum* isolates possess a *cspB* gene, encoding the S-layer protein PS2. It is interesting to note that *cspB* and its downstream ORF, coding for a putative Zn-dependent alcohol dehydrogenase, are therefore only missing in the sequenced *C. glutamicum* ATCC 13032 genome (Ikeda and Nakagawa, 2003; Kalinowski et al., 2003). It is possible that *C. glutamicum* ATCC 13032 previously also possessed an S-layer gene, which was lost during cultivation under continuous laboratory conditions, without demand for the protective functions of an S-layer (Fujita et al., 1997). Therefore, it would be interesting to analyze the putatively lost *cspB* gene region in its entirety and to identify its genomic location in relation to the *C. glutamicum* ATCC 13032 genome.

Sequence analysis of the S-layer genes and bioinformatics prediction of putative signal peptides revealed theoretical molecular masses of the deduced and processed amino acid sequences between 49.9 and 52.6 kDa. However, SDS–PAGE of the extracted *C. glutamicum* S-layer proteins showed varying sizes between 55 and 66 kDa. The reason for these differences between predicted and detected molecular masses of the S-layer proteins in SDS–PAGE might be the result of either a high number of acidic amino acids (Peyret et al., 1993) or of posttranslational modifications, such as glycosylation known from several other bacterial S-layer proteins (Calabi et al., 2001; Schaffer et al., 2001; Sumper et al., 1990). Bacterial S-layer glycans usually consist of long homo- or heteropolysaccharide chains composed of linear or branched repeating units with an overall chain length ranging up to 150 monosaccharide residues (Messner, 1997; Messner et al., 1997). Such type of posttranslational modification might result in large differences between the calculated and the detected molecular masses of S-layer proteins. Whether the S-layer proteins of *C. glutamicum* are posttranslationally modified has to be investigated in future studies.

In general, S-layer proteins have been recognized as belonging to the most abundant cellular proteins. An S-layer can comprise up to 15% of the total protein content of a bacterial cell (Boot and Pouwels, 1996; Messner and Sleytr, 1992; Sleytr, 1997; Smarda et al., 2002).

In this study, we used a new bioinformatics approach based on synonymous codon usage to predict the S-layer gene expression of different *C. glutamicum* isolates. The predicted expression levels of the S-layer genes appear to be comparable to those of other very highly expressed *C. glutamicum* genes, like *eno* coding for the glycolytic enzyme enolase and *rbsE*, which encodes a 30S ribosomal subunit (McHardy et al., 2004). These high expression levels correlate well with the high abundance of the isolated S-layer proteins in SDS–PAGE. Furthermore, the putative promoter regions of the analyzed S-layer genes from *C. glutamicum* were found to be highly conserved, suggesting that the promoter strength is comparable in all isolates.

BLAST searches in protein databases yielded no significant hits for the *C. glutamicum* PS2 proteins. This lack of homology supports earlier conclusions that PS2 proteins are very diverse in their sequences (Peyret et al., 1993). However, the PS2 proteins of the *C. glutamicum* isolates share several features with S-layer proteins of other bacteria. These similarities include a high content of hydrophobic amino acids, a higher portion of acidic amino acids as compared to basic amino acids resulting in a predicted weakly acidic isoelectric point and a very low content of sulfur-containing amino acids (Beveridge, 1994; Sara and Sleytr, 1996a, 2000; Sleytr, 1997).

We applied further bioinformatics tools to find conserved regions within the PS2 amino acid sequence, which could be relevant for the general structure of the *C. glutamicum* S-layer. Only the function of the 30 amino acid long N-terminal signal peptide and of the 21 amino acid-containing hydrophobic C-terminal anchor sequence of *C. glutamicum* S-layers have been described previously (Chami et al., 1997; Peyret et al., 1993). Although these regions appear to be highly conserved, we found that the anchor sequence of *C. glutamicum* ATCC 31830 consists only of 14 amino acids. This observation shows that variations might be possible even in otherwise highly conserved

protein regions of PS2. Additionally, several other conserved regions were newly identified in all analyzed *C. glutamicum* PS2 proteins. It can be assumed that these protein regions play an important role in S-layer formation since all information regarding the assembly process is obviously contained within the PS2 monomer (Beveridge et al., 1997; Sleytr and Messner, 1983). A remarkable outcome of this investigation is the high variability of PS2 proteins from the same species, which is in contrast to the highly conserved nucleotide sequences surrounding the respective *cspB* genes. Such high levels of amino acid sequence variability are especially known from S-layer proteins of pathogenic bacteria, like *Campylobacter fetus* (Tummuru and Blaser, 1993) and *Clostridium difficile* (Calabi and Fairweather, 2002). The variability of S-layer proteins in these species is supposed to result from a selective pressure to escape the immune response of their respective hosts. It is therefore surprising to detect such a variability in a bacterium that is generally regarded as a non-pathogenic soil bacterium.

4.2. C. glutamicum S-layers display a common hexagonal structure with a varying unit cell dimension which correlates to a sequence similarity-based classification scheme

Information on the overall structure and arrangement of the *C. glutamicum* S-layer proteins was deduced from a combination of both secondary structure predictions and morphological analyses by atomic force microscopy. The arrangements of the secondary structural elements of all analyzed PS2 proteins are highly conserved and vary only in the length of these elements. We have shown in this study that the predicted basic secondary structure is very similar in all analyzed S-layer proteins from *C. glutamicum*, whereas the individual amino acid sequences of the PS2 proteins are quite different when one considers that they represent the same structural protein of the same bacterial species. A division into five divergent *C. glutamicum* PS2 protein classes arose from a classification on the basis of protein sequence alignments by CLUSTALX. These differences of the primary structure of the PS2 proteins are also reflected by the morphology of the S-layers. Utilizing AFM imaging, we found out that all *C. glutamicum* S-layers exhibit a hexagonal symmetry with distinct differences in the unit cell dimension. Based on AFM measurements, the *C. glutamicum* S-layers were divided into five classes, the composition of which mostly fits the classification by sequence alignments. Therefore, it can be proposed that the amino acid sequence variations of PS2 correlate with differences in S-layer morphology of *C. glutamicum*.

To our knowledge, this is the first report in which a direct coherence between S-layer sequence information and morphology is pointed out. Systematical structure analyses were already accomplished for S-layer proteins of cyanobacteria, but without knowledge of the corresponding genes or S-layer protein sequences (Smarda et al., 2002). In these analyses, the hexagonal symmetry appears to be most common in cyanobacteria (Rachel et al., 1997; Smarda et al., 2002), which seems to be similar to the observation made in *C. glutamicum*. Additionally, the S-layers of a set of *Microcystis wesenbergii* isolates displayed differences in their unit cell dimension, ranging from 12.9 to 18.6 nm (Rachel et al., 1997; Smajs et al., 1999; Smarda et al., 2002). These differences of the unit cell dimension are in a roughly comparable range to those measured for the different S-layers of the *C. glutamicum* isolates (15.2–17.4 nm). In conclusion, systematical sequence analyses of *cspB* genes and bioinformatics structure predictions of PS2 proteins from a single species provided valuable information regarding a structure-function correlation. Therefore, this report provides first steps towards the identification of PS2 regions necessary for S-layer formation of *C. glutamicum*.

Acknowledgements

We are grateful to the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 613 for financial support. We thank Prof. Dr. G. Leblon (Université Paris-Sud) for supplying plasmid pCGL815, which was used for comparative analyses, A.C. McHardy for performing CoBias analyses and T.-C. Chao for helpful discussions.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and

PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bayan, N., Houssin, C., Chami, M., Leblon, G., 2003. Mycomembrane and S-layer: two important structures of *Corynebacterium glutamicum* cell envelope with promising biotechnology applications. J. Biotechnol. 104, 55–67.

Beveridge, T.J., 1994. Bacterial S-layers. Curr. Opin. Struct. Biol. 4, 204–212.

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

Boot, H.J., Pouwels, P.H., 1996. Expression, secretion and antigenic variation of bacterial S-layer proteins. Mol. Microbiol. 21, 1117–1123.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., Fairweather, N., 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. Mol. Microbiol. 40, 1187–1199.

Calabi, E., Fairweather, N., 2002. Patterns of sequence conservation in the S-layer proteins and related sequences in *Clostridium difficile*. J. Bacteriol. 184, 3886–3897.

Cerdeno-Tarraga, A.M., Efstratiou, A., Dover, L.G., Holden, M.T., Pallen, M., Bentley, S.D., Besra, G.S., Churcher, C., James, K.D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M.A., Rabbinowitch, E., Rutherford, K.M., Thomson, N.R., Unwin, L., Whitehead, S., Barrel, B.G., Parkhill, J., 2003. The complete genome sequence and analysis of *Corynebacterium diphtheria* NCTC 13129. Nucl. Acids Res. 31, 6516–6523.

Chami, M., Bayan, N., Dedieu, J., Leblon, G., Shechter, E., Gulik-Krzywicki, T., 1995. Organization of the outer layers of the cell envelope of *Corynebacterium glutamicum:* a combined freeze-etch electron microscopy and biochemical study. Biol. Cell 83, 219–229.

Chami, M., Bayan, N., Peyret, J.L., Gulik-Krzywicki, T., Leblon, G., Shechter, E., 1997. The S-layer protein of *Corynebacterium glutamicum* is anchored to the cell wall by its C-terminal hydrophobic domain. Mol. Microbiol. 23, 483–492.

Engel, A., Müller, D.J., 2000. Observing single biomolecules at work with the atomic force microscopy. Nat. Struct. Biol. 7, 715–718.

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.

Fujita, M., Moriya, T., Fujimoto, S., Hara, N., Amako, K., 1997. A deletion in the sapA homologue cluster is responsible for the loss of the S-layer in *Campylobacter fetus* strain TK. Arch. Microbiol. 167, 196–201.

Garnier, J., Gibrat, J.F., Robson, B., 1996. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. 266, 540–553.

Grant, S.G., Jessee, J., Bloom, F.R., Hanahan, D., 1990. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. Proc. Natl. Acad. Sci. USA 87, 4645–4649.

Hermann, T., Pfefferle, W., Baumann, C., Busker, E., Schaffer, S., Bott, M., Sahm, H., Dusch, N., Kalinowski, J., Punier, A., Bendt, A.K., Kramer, R., Burkovski, A., 2001. Proteome analysis of *Corynebacterium glutamicum*. Electrophoresis 22, 1712–1723.

Hermann, T., 2003. Industrial production of amino acids by coryneform bacteria. J. Biotechnol. 104, 155–172.

Ikeda, M., Nakagawa, S., 2003. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. Appl. Microbiol. Biotechnol. 62, 99–109.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B.J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Krämer, R., Linke, B., McHardy, A.C., Meyer, F., Möckel, B., Pfefferle, W., Pühler, A., Rey, D.A., Rückert, C., Rupp, O., Sahm, H., Wendisch, V.F., Wiegräbe, I., Tauch, A., 2003. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J. Biotechnol. 104, 5–25.

Katsumata, R., Ozaki, A., Oka, T., Furuya, A., 1984. Protoplast transformation of glutamate-producing bacteria with plasmid DNA. J. Bacteriol. 159, 306–311.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Lyubchenko, Y., Shlyakhtenko, L., Harrington, R., Oden, P., Lindsay, S., 1993. Atomic force microscopy of long DNA: imaging in air and under water. Proc. Natl. Acad. Sci. USA 90, 2137–2140.

McHardy, A.C., Pühler, A., Kalinowski, J., Meyer, F., 2004. Comparing expression level-dependent features in codon usage with protein abundance: an analysis of 'predictive proteomics'. Proteomics 4, 46–58.

Messner, P., 1997. Bacterial glycoproteins. Glycoconj. J. 14, 3–11.

Messner, P., Allmaier, G., Schaffer, C., Wugeditsch, T., Lortal, S., König, H., Niemetz, R., Dorner, M., 1997. Biochemistry of S-layers. FEMS Microbiol. Rev. 20, 25–46.

Messner, P., Sleytr, U.B., 1992. Crystalline bacterial cell-surface layers. Adv. Microb. Physiol. 33, 213–275.

Möller, C., Allen, M., Eilings, V., Engel, A., Müller, D.J., 1999. Tapping-mode atomic force microscopy produces faithful high-resolution images of protein surfaces. Biophys. J. 77, 1150–1158.

Müller, D.J., Baumeister, W., Engel, A., 1999. Controlled unzipping of a bacterial surface layer with atomic force microscopy. Proc. Natl. Acad. Sci. USA 96, 13170–13714.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int. J. Neural. Syst. 8, 581–599.

Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K., Gojobori, T., 2003. Comparative complete genome

sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. Genome Res. 13, 1572–1579.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

Peters, J., Baumeister, W., Lupas, A., 1996. Hyperthermostable surface layer protein tetrabrachion from the archaebacterium *Staphylothermus marinus:* evidence for the presense of a right-handed coiled coil derived from the primary structure. J. Mol. Biol. 257, 1031–1041.

Peters, J., Nitsch, M., Kuhlmorgen, B., Golbik, R., Lupas, A., Kellermann, J., Engelhardt, H., Pfander, J.P., Müller, S., Goldie, K., et al., 1995. Tetrabrachion: a filamentous archaebacterial surface protein assembly of unusual structure and extreme stability. J. Biol. 245, 385–401.

Peyret, J.L., Bayan, N., Joliff, G., Gulik-Krzywicki, T., Mathieu, L., Shechter, E., Leblon, G., 1993. Characterization of the *cspB* gene encoding PS2, an ordered surface-layer protein in *Corynebacterium glutamicum*. Mol. Microbiol. 9, 97–109.

Rachel, R., Pum, D., Smarda, J., Smajs, D., Komrska, J., Krzyzanek, V., Rieger, G., Stetter, K.O., 1997. II. Fine structure of S-layers. FEMS Microbiol. Rev. 20, 13–23.

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

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sara, M., Sleytr, U.B., 1987. Molecular sieving through S-layers of *Bacillus stearothermophilus* strains. J. Bacteriol. 169, 4092–4098.

Sara, M., Sleytr, U.B., 1996a. Crystalline bacterial cell surface layers (S-layers): from cell structure to biomimetics. Prog. Biophys. Mol. Biol. 65, 83–111.

Sara, M., Sleytr, U.B., 1996b. Biotechnology and biomimetic with crystalline bacterial cell surface layers (S-layers). Micron 27, 141–156.

Sara, M., Sleytr, U.B., 2000. S-Layer proteins. J. Bacteriol. 182, 859–868.

Saxton, W.O., Baumeister, W., 1986. Principles of organization in S-layers. J. Mol. Biol. 20, 251–253.

Schaffer, C., Graninger, M., Messner, P., 2001. Prokaryotic glycosylation. Proteomics 1, 248–261.

Scheuring, S., Stahlberg, H., Chami, M., Houssin, C., Rigaud, J.L., Engel, A., 2002. Charting and unzipping the surface layer of *Corynebacterium glutamicum* with the atomic force microscope. Mol. Microbiol. 44, 675–684.

Schneitz, C., Nuotio, L., Lounatma, K., 1993. Adhesion of *Lactobacillus acidophilus* to avian intestinal epithelial cells mediated by the crystalline bacterial cell surface layer (S-layer). J. Appl. Bacteriol. 74, 290–294.

Sleytr, U.B., 1997. Basic and applied S-layer research: an overview. FEMS Microbiol. Rev. 20, 5–12.

Sleytr, U.B., Beveridge, T.J., 1999. Bacterial S-layers. Trends Microbiol. 7, 253–260.

Sleytr, U.B., Messner, P., 1983. Crystalline surface layers on bacteria. Ann. Rev. Microbiol. 37, 311–339.

Sleytr, U.B., Sara, M., 1997. Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications. Trends Biotechnol. 15, 20–26.

Smajs, D., Smarda, J., Krzyzanek, V., 1999. New findings of S-layers among cyanobacteria. Algol. Stud. 94, 317–332.

Smarda, J., Smajs, D., Komrska, J., Krzyzanek, V., 2002. S-layers on cell walls of cyanobacteria. Micron 33, 257–277.

Smith, T.F., Waterman, M.S., 1981. Identification of common molecular subsequences. J. Biol. 147, 195–197.

Sonnen, H., Thierbach, G., Kautz, S., Kalinowski, J., Schneider, J., Puhler, A., Kutzner, H.J., 1991. Characterization of pGAl, a new plasmid from *Corynebacterium glutamicum* LP-6. Gene 107, 69–74.

Soual-Hoebeke, E., de Sousa-D'Auria, C., Chami, M., Baucher, M.F., Guyonvarch, A., Bayan, N., Salim, K., Leblon, G., 1999. S-layer protein production by *Corynebacterium* strains is dependent on the carbon source. Microbiology 145, 3399–3408.

Stackebrandt, E., Rainey, F.A., Ward-Rainey, N.L., 1997. Proposal for a new hierarchic classification system, *Actinobacteria classis* nov. Int. J. Syst. Bacteriol. 47, 479–491.

Staden, R., 1996. The STADEN sequence analysis package. Mol. Biotechnol. 5, 233–241.

Sumper, M., Berg, E., Mengele, R., Strobel, I., 1990. Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. J. Bacteriol. 172, 7111–7118.

Takeda, Y., Fujii, M., Nakajyoh, Y., Nichimura, T., Isshiki, S., 1990. Isolation of a tetracycline resistance plasmid from a glutamate-producing corynebacterium, *Corynebacterium melassecola*. J. Ferment. Bioeng. 70, 177–179.

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

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tummuru, M.K.R., Blaser, M.J., 1993. Rearrangement of *sapA* homologs with conserved and variable regions in *Camphylobacter fetus*. Proc. Natl. Acad. Sci. USA. 90, 7265–7269.