Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochimica et Biophysica Acta 1745 (2005) 223 – 253
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
http://www.elsevier.com/locate/bba

Genomic analysis of the protein secretion systems in *Clostridium acetobutylicum* ATCC 824

Mickaël Desvaux[a,b,*], Arshad Khan[a], Anthony Scott-Tucker[a], Roy R. Chaudhuri[a], Mark J. Pallen[a], Ian R. Henderson[a]

[a]*The Institute for Biomedical Research (IBR), The University of Birmingham-The Medical School, Division of Immunity and Infection, Bacterial Pathogenesis and Genomics Unit, Edgbaston, Birmingham B15 2TT, UK*
[b]*Institut National de la Recherche Agronomique (INRA), Centre de Recherches de Clermont-Ferrand-Theix, Unité de Recherche 370, Equipe Microbiologie, F-63122 Saint-Genès Champanelle, France*

Received 28 January 2005; received in revised form 20 April 2005; accepted 20 April 2005
Available online 23 May 2005

Abstract

Consistent information about protein secretion in Gram-positive bacteria is essentially restricted to the model organism *Bacillus subtilis*. Among genome-sequenced clostridia, *Clostridium acetobutylicum* has been the most extensively studied from a physiological point of view and is the organism for which the largest variety of molecular biology tools have been developed. Following in silico analyses, both secreted proteins and protein secretion systems were identified. The Tat (Twin arginine translocation; TC #2.A.64) pathway and ABC (ATP binding cassette) protein exporters (TC #3.A.1.) could not be identified, but the Sec (secretion) pathway (TC #3.A.5) appears to be used prevalently. Similarly, a flagella export apparatus (FEA; TC #3.A.6.), holins (TC #1.E.), and an ESAT-6/WXG100 (early secreted antigen target of 6 kDa/ proteins with a WXG motif of ~100 residues) secretion system were identified. Here, we report for the first time the identification of a fimbrilin protein exporter (FPE; TC #3.A.14) and a Tad (tight adherence) export apparatus in *C. acetobutylicum*. This investigation highlights the potential use of this saprophytic bacterium in biotechnological and biomedical applications as well as a model organism for studying protein secretion in pathogenic Gram-positive bacteria.

Keywords: *Clostridium*; Protein secretion system; Virulence factor; Gram-positive bacteria; Secretome

1. Introduction

The genus *Clostridium* includes bacteria which are obligately anaerobic, form endospores, are able to carry out dissimilatory sulfate reduction, and possess a Gram-positive type of wall structure [1,2]. In a recently proposed hierarchical structure for clostridia, *Clostridium acetobutylicum* was classified in family 1, corresponding to the *Clostridiaceae* family, and genus 1, corresponding to group I of the previous Johnson and Francis classification [3]. *C. acetobutylicum* ATCC 824 was originally isolated from garden soil in 1924 [4]. This strain is one the best-studied solventogenic clostridia and is closely related to the historical Weizmann strain used to develop an industrial starch-based acetone and butanol fermentation process [5]. *C. acetobutylicum* ATCC 824 possesses a 3.9-Mb chromosome and harbours pSOL1, a megaplasmid of 192 kb coding for genes involved in solvent formation; loss of pSOL1 leads to a loss of solventogenicity of the strain [6–8]. In addition to *C. acetobutylicum* ATCC 824 [9], the genome sequence has been determined for two other clostridial genomes, *Clostridium perfringens* 13 and *Clos-*

* Corresponding author. The Institute for Biomedical Research (IBR), The University of Birmingham-The Medical School, Division of Immunity and Infection, Bacterial Pathogenesis and Genomics Unit, Edgbaston, Birmingham B15 2TT, UK. Tel.: +33 4 73 62 47 23; fax: +33 4 73 62 42 68.
E-mail addresses: m.desvaux@bham.ac.uk, mdesvaux@sancy.clermont.inra.fr (M. Desvaux).

0167-4889/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamcr.2005.04.006

tridium tetani E88 [10,11]. The sequencing and annotation of 4 other clostridial genome, i.e., *Clostridium botulinum* ATCC 3502, *Clostridium difficile* 630, *C. perfringens* ATCC 13124, and *Clostridium thermocellum* ATCC 27405, are currently in progress. Among genome-sequenced clostridia, *C. acetobutylicum* ATCC 824 is undoubtedly the one which has been the most extensively studied from a physiological point of view and is the one for which the largest variety of molecular biology tools have been developed [9,12–15]. Thus, this bacterium appears to be an excellent model for studying the biology of clostridia and Gram-positive bacteria in general.

The availability of the complete genome of *C. acetobutylicum* allows us to focus on its secretome, i.e., the population of the gene products that are translocated through the cytoplasmic membrane to be either displayed on the bacterial cell surface, secreted in the extracellular milieu, or even secreted into a host cell [16]. Protein secretion systems in Gram-positive bacteria have attracted a lesser amount of attention than those of the Gram-negative bacteria. Indeed, consistent information about protein secretion in Gram-positive bacteria is essentially restricted to *Bacillus subtilis*, the paradigm of Gram-positive bacterial biology [17–20], and to scattered information arising from the study of Gram-positive pathogenic bacteria [17,18,20–23]. Protein secretion in clostridia has not been investigated thoroughly, and as a result, relatively little is known about the proteins and secretion systems involved [24].

In Gram-negative bacteria, five major systems, numbered from I to V and the chaperone/usher pathway, are currently recognized as being involved in the secretion of proteins from the cytoplasm to the extracellular milieu [25–28]. This classification is restricted to Gram-negative bacteria, where the cell envelope is composed of the cytoplasmic membrane and the outer membrane. Since Gram-positive bacteria possess only one biological membrane, i.e., the cytoplasmic membrane, the classification of protein secretion pathways is different. Thus, in Gram-negative bacteria, protein translocation through the cytoplasmic membrane corresponds to export into the periplasm, whereas in Gram-positive bacteria, it permits the secretion of proteins into the extracellular milieu [29–40]. Five major protein secretion systems are currently recognized in Gram-positive bacteria: the Sec (secretory) pathway, the Tat (twin arginine translocation) pathway, the ABC (ATP binding cassette) transporters, the fimbrilin protein exporter (FPE) system, and more recently, the ESAT-6/WXG100 (early secreted antigen target of 6 kDa/proteins with a WXG motif of ~100 residues) secretion system [17,18,23,41,42]. It is clear from studies of Gram-negative bacteria that not all the secretion pathways are systematically present in a single organism and that the extent to which each pathway is used varies from one organism to another. It seems plausible that a similar scenario will be found for Gram-positive bacteria.

By using a variety of bioinformatics tools, the protein secretion systems and secreted proteins of *C. acetobutylicum* ATCC 824 were identified and characterized. Such comprehensive analyses not only provide insight into the portion of the proteome dedicated to bacterial secretion but also the physiology of this organism, as well as the possible biotechnological applications. This investigation is a prelude to in vivo study of the proteins secreted by *C. acetobutylicum* and highlights the potential use of this saprophytic bacterium as a model organism for exploring protein secretion in pathogenic Gram-positive bacteria.

2. Materials and methods

2.1. Computer methods

Prior to bioinformatic analyses, the complete genome, coding sequences, and annotation files for *C. acetobutylicum* ATCC 823 were downloaded from GenBank (ftp://ftp.ncbi.nih.gov/genbank/Bacteria/Clostridium_acetobutylicum/). The chromosome of *C. acetobutylicum* possesses a total of 3672 potential coding sequences (CDS), and the megaplasmid pSOL1 encodes 176 CDS. Each CDS was screened for the capacity to encode a secreted protein or a component of a secretion apparatus. Proteins are identified by their GenBank index (GI) number.

BLAST and PSI-BLAST [43,44] were performed at ViruloGenome (http://www.vge.ac.uk/), the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/BLAST/), and ClostriDB (http://clostri.bham.ac.uk/) [45]. The presence of conserved domains and/or motifs was identified using Pfam (http://www.sanger.ac.uk/Software/Pfam/) [46], CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [47], ScanProsite (http://ca.expasy.org/tools/scanprosite/) [48], SMART (http://smart.embl-heidelberg.de/) [49], and/or InterProScan (http://www.ebi.ac.uk/InterProScan/) [50]. The level of homology between two protein sequences was determined using BLAST 2 sequences using the matrix BLOSUM62 and default parameters, except that filter was turned off [51]. Sequence alignments were performed using ClustalW [52], with minor manual refinement using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

All *C. acetobutylicum* predicted CDS were analysed and classified using the Transporter Classification Database (TC-DB) (http://www.biology.ucsd.edu/~msaier/transport/) as previously described [53,54]. This analysis was completed by combining results from TransportDB [55], a relational database describing the predicted cytoplasmic membrane transport proteins in bacteria (http://66.93.129.133/transporter/wb/index2.html). The identification of a FPE system was based on PSI-BLAST and BLAST searches from ViruloGenome and ClostriDB using ComGA (GI: 1303877), ComGB (GI: 1303878), and ComC (GI: 98253) from *B. subtilis* as queries. In a similar fashion, the ESAT-6/WXG100 secretion system was identified by using the previously described ESAT-6 from *Mycobacterium*

tuberculosis (GI: 38490389) and YukA from *B. subtilis* (GI: 7474744) [56]. While in *M. tuberculosis* and *Staphylococcus aureus*, the involvement of YukA homologues in ESAT-6 proteins translocation has been demonstrated, there is no direct experimental evidence in *B. subtilis* [42,57].

All predicted protein-coding genes were analysed using SignalP (http://www.cbs.dtu.dk/services/SignalP/), to detect the presence of signal peptides [58]. In these analyses, both the neural network and the hidden Markov model were used to examine the N-terminal part of each protein sequence. Sequences were sequentially analysed without truncation and with truncation set at 35, 70, and 140 amino acid (a.a.) residues (SignalP score >0.6). This analysis was completed by combining results from PSORT and PSORTb predicting the translocation and cellular localization of a protein (http://psort.ims.u-tokyo.ac.jp/) [59,60]. Signal sequence topology was predicted by combining the results from Kyte and Doolittle, as well as Eisenberg scale-mean hydrophobicity profiles using BioEdit, SignalP, and TMpred.

For the identification of lipoprotein patterns, searches were carried out using ScanProsite for the lipoprotein lipid attachment site (PS00013) (http://www.expasy.org) and G+LPP pattern [61]. Moreover, these results were combined with those from servers dedicated to the recognition of lipoproteins, i.e., DOLOP [62], PSORT, and LipoP (http://www.cbs.dtu.dk/services/LipoP/) [63]. The criteria used for the exclusion of sequences as false-positives unlikely to encode lipoproteins involved TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) [64], SignalP, DAS (http://www.sbc.su.se/~miklos/DAS) [65], and TopPred2 (http://bioweb.pasteur.fr/seanal/interfaces/toppred.html) [66] and have been described previously [61].

From SignalP or PSORT searches, proteins bearing a Type 4 prepilin signal peptide were identified by the presence of a consensus (K/R)G(F/Y) motif present between the n- and h-domains of the signal peptide, the cleavage occurring between G and (F/Y), and a glutamine residue present at position +5 of the cleavage site [17,67]. Some variability in the a.a. present at position +5 was also taken into account [17,39,68].

3. Results

3.1. Absent protein secretion systems

A recent genomic survey of prokaryotic protein secretion systems suggests that the Tat pathway, instead of the Sec pathway, is used predominantly by certain bacteria and archaea [69]. However, neither homologues of the TatABCE translocon nor putative Tat substrates could be identified in *C. acetobutylicum* by ViruloGenome PSI-BLAST and TC-DB searches, indicating the absence of a Tat protein secretion pathway (TC #2.A.64) in this organism. Furthermore, this pathway appears to be absent from other sequenced clostridial genomes, and this could be a general feature of the genus *Clostridium* [10].

The ABC superfamily contains both uptake and efflux transport systems for a large variety of substrates, e.g., glucides, proteins, and ions. So far, ABC transporters for proteins have been identified only in Gram-negative bacteria; in Gram-positive bacteria, only ABC transporters for peptides have been identified [70]. As expected, no protein exporter could be identified in *C. acetobutylicum*; however, TC-DB searches revealed the presence of several peptide exporters belonging to the peptide-1 exporter (Pep1E) family (TC #3.A.1.111), the peptide-2 exporter (Pep2E) family (TC #3.A.1.112), the peptide-4 exporter (Pep4E) family (TC #3.A.1.123), the 3-component peptide-5 exporter (Pep5E) family (TC #3.A.1.124), or the β-Exotoxin I Exporter (βETE) family (TC #3.A.1.126) (see Supplementary material in Appendix A).

It is worth mentioning that an ABC transporter, comprising an ATPase subunit (GI: 15893536) and an integral membrane protein (GI: 15893537), was identified as a member of the lipoprotein translocase (LPT) family (TC #3.A.1.125) (see Supplementary material in Appendix A). In *E. coli*, the LPT differs from other ABC transporters, as it is not involved in transmembrane transport of substrates but permits the localization of lipoproteins to the outer membrane [71]. So far, this transporter family has only been studied in Gram-negative bacteria, and its function in Gram-positive bacteria remains obscure.

In Gram-negative bacteria, members of the family of large conductance mechanosensitive ion channel (MscL; TC #1.A.22) permit the release of small proteins such as thioredoxin during osmotic downshift [72,73]. However, no homologue of this particular protein secretion system could be identified from the genome of *C. acetobutylicum*. Interestingly, MscL homologues could be identified in some other clostridial species, i.e., *C. histolyticum* (GI: 7674133), *C. perfringens* (GI: 20141572), and *C. tetani* (GI: 28210182). Experimental evidence for MscL as a protein secretion pathway in Gram-positive bacteria remains to be shown.

Despite the absence of the systems described above, a Sec system, a flagella export apparatus (FEA), a fimbrial protein export (FPE) apparatus, the tight adherence (Tad) export system, the holins, and an ESAT-6/WX100 secretion system were identified. Table 1 summarizes the protein secretion systems found in *C. acetobutylicum* (see also Supplementary material in Appendix A).

3.2. The Sec system (TC #3.A.5.)

The presence, remarkable conservation, and essential nature of the Sec translocon has given rise to the notion of a general protein secretion mechanism with *E. coli* as the bacterial paradigm [74]. The heterotrimeric SecYEG complex is the central component of the Sec apparatus; it forms a protein-conducting channel through the cytoplasmic

Table 1
Components of the protein secretion pathways present in *C. acetobutylicum*

Systems	Components	TC	GI	Gene location	Cellular location	Substrates[a]
1. The Sec system		#3.A.5.				602
Transmembrane component	SecY	#3.A.5	15896363	Chromosome	Membrane	
	SecE	#3.A.5	15896398	Chromosome	Membrane	
	SecG	#3.A.5	15894002	Chromosome	Membrane	
ATPase	SecA		15896101	Chromosome	Peripheral	
	SecA (fragment)		15896773	Chromosome	Peripheral	
Auxiliary protein	SecF	#2.A.6.4.1.	15895545	Chromosome	Membrane	
	SecD	#2.A.6.4.1.	15895546	Chromosome	Membrane	
	YajC	#9.B.18.	15895549	Chromosome	Membrane	
	SpoIIIJ	#2.A.9.3.2.	15896967	Chromosome	Membrane	
	FtsY	#3.A.5	15895029	Chromosome	Cytoplasm	
	Ffh	#3.A.5	15895031	Chromosome	Cytoplasm	
Signal peptidase	SPase I		15895037	Chromosome	Membrane	
			15895904	Chromosome	Membrane	
	SPase II		15893593	Chromosome	Membrane	
			15895384	Chromosome	Membrane	
2. Flagella export apparatus		#3.A.6.1.				18
Transmembrane component	FlhA		15895416	Chromosome	Membrane	
	FlhB-FliR		15895417	Chromosome	Membrane	
	FlhB		15893927	Chromosome	Membrane	
	FliP		15895419	Chromosome	Membrane	
	FliQ		15895418	Chromosome	Membrane	
ATPase	FliI		15895428	Chromosome	Cytoplasm	
ATPase regulator	FliH		15895429	Chromosome	Cytoplasm	
Chaperone	FliJ		15895427	Chromosome	Cytoplasm	
	FliS		15895474	Chromosome	Cytoplasm	
3. The FPE system		#3.A.14.				7
ATPase	ComGA homologue		15895375	Chromosome	Membrane	
Transmembrane component	ComGB homologue		15895374	Chromosome	Membrane	
Type IV prepilin peptidase	ComC homologue		15896039	Chromosome	Membrane	
4. The Tad export apparatus						?
ATPase	TadA homologue		15895251	Chromosome	Peripheral	
Transmembrane component	TadB homologue		15895250	Chromosome	Membrane	
Transmembrane component	TadC homologue		15895249	Chromosome	Membrane	
Unknown function	TadZ homologue		15895252	Chromosome	Cytoplasm	
5. The holins		#1.E.				
Holins	LrgA homologue	#1.E.14.	15893901	Chromosome	Membrane	3
	LrgA homologue	#1.E.14.	15004877	Megaplasmid	Membrane	?
	TcdE homologue	#1.E.19.	15895117	Chromosome	Membrane	1
6. The ESAT-6/WXG100 secretion pathway						10
Membrane-bound ATPase transporter	YukA homologue		15893337	Chromosome	Membrane	
	YukA homologue		15893699	Chromosome	Membrane	
	YukA homologue		15896940	Chromosome	Membrane	

[a] Details of the substrates related to each secretion pathway are discussed in the corresponding paragraph.

membrane [75]. The SecYEG complex is ubiquitous in all domains of life (termed Sec61αγβ in eukaryotes) [76], and as expected, genes encoding each component were identified in *C. acetobutylicum* (Table 1).

In addition to the SecYEG complex, several other proteinaceous components are required for Sec-dependent secretion. The cytosolic ATPase SecA, which provides the driving force for protein translocation, has only been identified in bacteria and chloroplasts. In *C. acetobutylicum*, besides a full-length SecA homologue (839 a.a. long), a small SecA fragment is also present (166 amino acids long) and is presumably non-functional (Table 1). No duplication of SecY is observed; in *Streptococcus gordonii*, the duplication of SecA and SecY constitutes a specialized system for the transport of large serine-rich proteins involved in pathogenesis [77].

In *E. coli*, three auxiliary proteins, i.e., SecD, SecF, and YajC, form a transmembrane complex loosely associated with SecYEG, which contributes to the efficiency of protein translocation [78]. By contrast, in *B. subtilis*, SecD and SecF are fused into a single protein [79]. TC-DB, Pfam, and PSI-BLAST searches indicate that a gene previously annotated as encoding a protein homologous to SecD/SecF protein exporters (GI: 15896672) is most certainly related to the Gram-positive bacterial hydrophobe/amphiphile efflux-2 family (HAE2) (TC #2.A.6.5) of the drug exporters of the resistance-nodulation-cell division (RND) superfamily (TC #2.A.6.) and has no connection to the Sec apparatus.

However, similar searches identified two open reading frames in *C. acetobutylicum* encoding separate SecD and SecF proteins, as well as YajC (Table 1).

In *E. coli*, the integral membrane protein YidC associates with the SecYEGDF–YajC complex and permits cytoplasmic membrane insertion of polytopic membrane proteins [80]. YidC has also been shown to catalyse membrane protein insertion in a Sec-independent manner [81,82], leading some authors to suggest YidC be classified as an alternative inner membrane translocation pathway [83], even though YidC has not been implicated in the translocation of proteins across the cytoplasmic membrane [84]. Two homologues of YidC are found in *B. subtilis*, i.e., SpoIIIJ and YqjG [17,18], however, only a SpoIIIJ-like protein (GI: 15896967) could be identified in *C. acetobutylicum*.

To allow proteins destined for translocation across the cytoplasmic membrane to interact with the Sec translocon, bacteria have developed several distinct protein targeting pathways. In Gram-negative bacteria, SecB and the signal recognition particle (SRP) pathway are considered to be two distinct protein-targeting pathways converging on the Sec translocon [85]. The SRP pathway was considered a specific mechanism for the integration of inner membrane proteins [86,87]; however, it was recently shown that the SRP-dependent pathway is involved in the secretion of some extracellular proteins in Gram-negative bacteria [88]. The SRP pathway requires the action of the SRP (Ffh/SRP54) and the SRP receptor (FtsY/SRP receptor-α family), which, like the SecYEG translocon, are ubiquitous in all domains of life [76]. Thus, it was not surprising to find Ffh and FtsY in *C. acetobutylicum*.

In Gram-negative bacteria, SecB possesses the dual function of preventing folding and targeting the protein to be secreted to the Sec translocon. As in all Gram-positive bacteria sequenced so far, the SecB chaperone is absent from *C. acetobutylicum* [18]. However, in *E. coli*, it was recently shown that SecB should be considered as only one of a plethora of molecular chaperones available to newly synthesized polypeptides before their translocation through the cytoplasmic membrane [87,89]. Indeed, in *B. subtilis*, the chaperone function has been attributed to CsaA (GI: 2619042); however, no homologue of CsaA could be found in *C. acetobutylicum*.

After Sec-dependent translocation through the cytoplasmic membrane, the signal sequences of most proteins are removed by the action of a signal peptidase (SPase). Membrane-bound SPase I and II permit the cleavage of the signal sequence from the precursor proteins and lipoproteins, respectively [18]. While in most Gram-negative bacteria, one SPase I and one SPase II are expressed, in *B. subtilis*, five SPase I and one SPase II have been identified [18,90]. In *C. acetobutylicum*, two homologues of each SPase I and II have been identified (Table 1). The reason for this redundancy is unclear, but in *B. subtilis*, differences in substrate specificity have been observed.

3.2.1. The signal sequences of the Sec-dependent secreted proteins

All proteins targeted to the Sec translocon possess an N-terminal signal peptide. The signal sequence is composed of three domains: (i) a positively charged amino terminus or n-domain, (ii) a hydrophobic core region or h-domain, and (iii) a consensus signal peptidase recognition site, also called c-domain [39].

Investigations with SignalP and PSORT of all *C. acetobutylicum* CDS identified 491 proteins, 37 of which are encoded on the plasmid pSol1, as potential substrates of SPase I (Table 2). The signal peptides of these proteins have lengths varying from 16 to 53 residues, with an average of 32 residues (Fig. 1). The length of the n-domain ranges from 2 to 31 residues, with an average of 10 residues but with a favored length of 7 (Fig. 1). These signal sequences contain, on average, 2–3 positively charged to a.a., i.e., lysine or arginine, but some could contain up to 14 (Table 2). Like *B. subtilis*, the h-domain has an average of 19 residues, with a favored length of 17–18 residues (Fig. 1) [17]. Interestingly, more than 50% of the predicted signal peptides possess at least one helix-breaking residue in the h-domain (Table 2). When present at the end of the h-domain, these residues facilitate the cleavage by SPase [17].

Because of steric interference with SPase I, a.a. tolerated at positions −3 and −1 of the signal peptide, with respect to the N-terminus of the mature exported protein, are generally small and uncharged [17]. In *B. subtilis*, the sequence consensus at the SPase I cleavage site, i.e. (A/V)-(S/F/K)-A at positions −3 to −1, is not fundamentally different from the consensus found in *C. acetobutylicum*, i.e., (V/A/I)-(K/F/Y)-A (Fig. 2); the consensus are given when the frequency of an a.a. at a given position is higher than 10%. Except for cysteine, all residues appear to be allowed at position +1 of the cleavage site, though an alanine residue is most abundant at this position (Fig. 2). In *E. coli*, though, it has been shown that proline at position +1 would effectively inhibit SPase I activity; therefore, its presence at this position seems unlikely, and such signal peptidase cleavage sites would be mispredicted [91,92].

By merging the results of ScanProsite, G+LPP, DOLOP, and LipoP together, 111 probable lipoproteins were identified in *C. acetobutylicum* (Table 2), 4 of them encoded on pSol1. The signal sequences of lipoproteins in *C. acetobutylicum*, like in *B. subtilis* [17], (i) are shorter, with an average of 21 a.a., (ii) possess shorter n- and h-domains, with an average of 6 and 12, respectively, and (iii) have more conserved structural features than the signal peptides of nonlipoproteins (Figs. 1 and 2). Once again, the lipobox motif in *B. subtilis*, i.e., L-(A/S/T)-(A/G)-C [17], is not fundamentally different from the consensus found in *C. acetobutylicum*, i.e., (L/F/I)-(T/S/V/A)-(G/A/S)-C (Fig. 2). The invariant cysteine at position + 1 is lipid-modified before the signal peptide is removed by SPase II.

Table 2
Substrates of the Sec translocon in *C. acetobutylicum*

GI	Function	Signal peptide	Gene location	SPase
Metabolism of carbohydates				
15893386	Xylanase	**MRSKQKRKLKS**VFIVVIALIVLLSAVIA AV	C	I
15893559	L–lactate dehydrogenase	**MKKNTK**ISVIGAGFVGSSTVFA LM	C	I
15893624	β–mannanase	**MTLKNPKKLF**LSIFAVLYIISAANPTKVSA FF	C	I
15893648	Galactoside symporter	**MKEKIKLSN**ILGYSSINFLGSGAQGLMSAWLLFFYSSVCGINSVKA AS	C	I
15893675	PTS, cellobiose–specific component BII	**MVR**ILLFCSAGMSTSMLVSKMKKA AE	C	I
15893719	Sugar permease	**MIKKIK**ISGIYGLNIILGLITIVPLIYALCVSLMPQDQIFS FP	C	I
15893822	PTS, arbutin–like BC component	**MMQKIQR**FGAAMFVPVLFFAFYGMVVG FS	C	I
15893828	β–mannanase	**MRRNNDIEQKKK**ILSAMLVSVLLVSPKINSKV YA	C	I
15893829	β–mannanase	**MTFNKRKRT**IISLLATTLILSPLVTNKVYA VP	C	I
15893830	β–mannanase	**MILNKRKK**VLISMVTMVGILSPLANSKVYA IP	C	I
15893851	Cellulase CelE	**MLRRK**LLSMIVAASLVVGVGFSNICYA KP	C	I
15893860	PTS enzyme II, ABC component	**MGNK**IFAVLQKIGKSLMLPVSVLPA AG	C	I
15893864	Pectate lyase	**MLYLMKLEGGVIMSKKRPSK**LIALCIIAATASMSVFNSNFVKA DS	C	I
15893866	Endo–arabinase	**MKNKFKK**IISITLGLSLTFVTQLNLIPFGNEVTAFA AA	C	I
15893950	Sugar ABC transporter	**MKISKKR**LAIIGVVIVIIAAIGMYPVFS KK	C	I
15893952	Sugar–binding periplasmic protein	**MKMSKKR**LAIIGVVIVIIVAIGMYPVFS KK	C	I
15893954	Sugar permease	**MKRKNK**TLIFFEYTILVIFAIAAIFPVIYMISSS FM	C	I
15893992	Sugar ABC transporter	**MKKDNKIVSGIVNALK**SLAFPMASVLVSLFVAVFFVMWSKGYA IT	C	I
15893993	Sugar ABC transporter	**MDK**ILLIIAIAAATFRMATPLIFT AI	C	I
15893994	Endo–1,4–β–glucanase	**MKRNRMK**ILNLPIFFVVLMFLLSNVGIIRVKA NN	C	I
15894023	Glucanotransferase	**MKKK**LIIVVMIFVVIFSILINLRIR IK	C	I
15894091	Pectate lyase	**MGRRFSKR**LLVAILASSMAISATFSANA ST	C	I
15894099	Pectate lyase	**MLKQTKRRLSK**AVVSAVILSSVSFYGMSSSIKA NA	C	I
15894112	Endoglucanase	**MFRRLFSKVK**ILFLALTLLGTYLTLNLNTVVA TN	C	I
15894113	Endoglucanase	**MFKNLFSKAK**ISLLTFAFLGTSLSGAYSACA AT	C	I
15894197	Cellulosomal scaffolding protein	**MKKRN**IAILGMFMMTASLGISKNHVFA DT	C	I
15894198	Endoglucanase	**MLKISKNFKK**IMAVALTSTVIFGSLSGLLTNKVAA AT	C	I
15894200	Endoglucanase	**MVKKRLKS**LIAVFITTAMMGGFMNLTDGKVMA DE	C	I
15894202	Endoglucanase	**MKRNK**ILVSALAIAFSSMVLTPVSG LK	C	I
15894203	Endoglucanase	**MKK**LLATLVTVFMICMQITPVQA AK	C	I
15894204	Endoglucanase	**MKK**LVACLTMVLLLGTQIKPLTITA DT	C	I
15894205	Endoglucanase	**MKKLK**VISTVVAGVFLSTLVSFSSITKVKA AD	C	I
15894247	Xylanase	**MKKAKEK**YLIFILIFLLLPINAKA ES	C	I
15894324	Xylanase	**MKTNGISLILKRS**AMILSIALIVLG AT	C	I
15894360	β–glucosidase	**MNKNLKRN**LAVLVSSTFLVSTTINLTSSKPIKA EN	C	I
15894610	Short–chain alcohol dehydrogenase	**MKLPFNVNLKNK**VAVVTGGTGVLGSSWVNALA EC	C	I
15894638	Xylanase	**MKIKDSRFKVKN**IVLVVLTLILLCGFYIIVREG IY	C	I
15894732	Ribose ABC transporter	**MNKNYKK**IVIFILLIFTVVFILINLKVLEG NG	C	I
15894741	Levanase/invertase	**MRLKQ**IAVLLLSSIVIVTGNNWVKA NT	C	I
15894808	Sugar–proton symporter	**MKIFRKD**FFNFGGLFFFYFLIWAVVLTFLPMWLTDVAHLDA SK	C	I
15894834	Chitinase	**MSTRKK**LLTFLAVMLVIIFNGTLAFA KP	C	I
15894839	Glycosyltransferases	**MKLKKHKSNEK**LLICCTVIFTIIYLVWRITCTLPLGGTIVSLVA AI	C	I
15895048	Levansucrase	**MNKLKIVKC**ILIGSMICSGIITQQTFA ST	C	I
15895049	Levanase	**MKRKKLK**LLTILMIFTVIILIVLLSLYIQNNSTI KS	C	I
15895050	Levansucrase	**MKTRKTYK**MISSLMVILAILTIPFLILRHNTGYT SI	C	I
15895239	Pectate lyase	**MNKK**ILSLLTSVAIISSVASFGVQA KS	C	I
15895518	UDP–glucose pyrophosphorylase	**MSIRK**AVIPAAGLGTRFLPATKAQP KE	C	I
15895520	β–glucosidase fused to β–amylase	**MYSGKKKVWKK**TLYYFVAAALTLNTVPEIIRPVSAKA AP	C	I
15895583	Xylanase	**MLCIMEKKKK**YITVFMIIIFLASIVGFSFFNYNEKKA NY	C	I
15895641	Oligopeptide ABC–type transporter	**MKRNIGFYKIVYSRLKRDRR**ALLGLLILFIIIAFAFLGPFLS KY	C	I
15895642	Oligopeptide ABC–type transporter	**MIRYIVKR**LFFMIPIIIITSFIIYMIFTLAPGDIA TS	C	I
15895649	Xylanase	**MHVIYARKVCVKK**LFFPIFLLIFCVGIF MY	C	I
15895779	β–galactosidase	**MNNNFKK**VLMCLLGVGIYISAVTFPSKIYA DT	C	I
15895818	Endoglucanase	**MKSKISK**LLIVAAITMASVGWGALKVGA TT	C	I
15895830	β–galactosidase	**MLKKLIKR**ISLMLLGTGLIINSAFFSISTSANTFANG AD	C	I
15896062	Endoglucanase	**MLRKKIKT**IIYSSLLLGAFVSATVLPVNTFA AS	C	I
15896065	Glucoamylase	**MKTKSIMQK**VALALLSFTFSTSIVGITPSRTMA EE	C	I
15896144	Fusion of β–glucosidase and glycosidase	**MYNSSKTRIWKRT**LYCLVAAAVVVSAMPQALNINVKA DT	C	I
15896209	Galactitol/fructose PTS IIC component	**MNGIVEFANAVFK**PLINLGAAPMMFIVLSVLAILMRVKVSKA IE	C	I

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Metabolism of carbohydates				
15896261	Xylanase	**MKKR**FILILVVILVSVIVLPNKVAA VN	C	I
15896628	Pectate lyase	**MKYLTKK**LFLISTFMGISLASVFSLATTAYA SS	C	I
15896677	α–arabinofuranosidase	**MKKR**YLTILMSTFLTTGAVLSYTSLSAKA AP	C	I
15896708	Endoglucanase with S–layer homologuey	**MRNKKR**ITSLVTGLAMLFTCAVGNTSLKVHA DA	C	I
15896839	Gluconate permease	**MPLLIVVIGVALLLLLMIKFK**VNGFISLILVALVVGIAEGMNPAKA VS	C	I
15896916	Polygalacturonase	**MLGKYK**LQAILAATALLFNVLFVNVNGKIVRA AD	C	I
15004733	Permease MDR–related	**MKMWKKN**LIVCWFGMFITSIGMSQI AP	P	I
15004757	Xylanase	**MLKSKLSK**ICTGVLALGLALSISGVGTFKA AM	P	I
15004758	Xylanase/chitin deacetylase	**MKK**LLTVILILTLLSIPYSVKSA KA	P	I
15004760	Pectate lyase	**MKK**IVCLVASAALSLSLFSLPQLASSVNA AS	P	I
15004775	Xylanase	**MLKRK**LILGLATTLLFMQTATITNVTALA SS	P	I
15004801	Amylase	**MKRIFSLRKS**LLAVVVTLTAFSGSLFYPISNR FQ	P	I
15004817	β–xylosidase	**MLKKRKFSFKKL**GCIAVLICMAVA AK	P	I
15004818	Xylanase	**MKK**ASYIMAFSLAFALVSPVACH AD	P	I
15004819	Xylanase	**MLKSKLAK**ICTGVLALGLALSISGVGTAKA AM	P	I
15004820	β–xylosidase	**MKKR**FISSTLILTTILSLSGAFQSKAFA AT	P	I
15004821	Xylanase	**MKNKIKK**IVLSVVTFSMICLPFTDGFTSIKA AS	P	I
15004822	Xylanase	**MNIKLKRT**LISLVAFSMTCLPFVGTGSSVKA AS	P	I
15004823	Xylanase	**MFKKSN**AFLLTFALICGIGFA KS	P	I
15004827	Arabinose efflux permease	**MNTTINNSNLK**YNNLILILGLAGFVSA AD	P	I
15004831	Arabinose efflux permease	**MEETNKTS**YWPMMIAIFLGSFLVSLGTSTINLALPFLMKSFS AN	P	I
15004832	Glycogen–binding regulatory phosphatase	**MKLNKSAGK**IITLIAFVAAFFIINIGVVPNAKA DT	P	I
15004834	Arabinose efflux permease	**MTQMNSRKKS**IIASLMVAMFLGAIEGTVVTTA MP	P	I
15004871	Amylase	**MSKRSKLLKRR**MLSLSVICVLIGYGPVFNPVRSQ AK	P	I
15896619	Xylanase	**MNKPNERKK**ILIKRSILVG CT	C	II
15893727	Xylanase	**MSKFKK**ILLIIITIIVIAS CL	C	II
15894630	Periplasmic sugar–binding protein	**MKKTSR**ILIIFAATCFLLTA CA	C	II
15894735	Sugar–binding periplasmic protein	**MNFKKHNK**FIIAAILIVIIAGGSFL CI	C	II
15896165	Sugar–binding periplasmic protein	**MKRFLKK**IFVAVMCSSMVFYVSA CT	C	II
15894201	Cellulosome integrating protein	**MKKK**GIITAVILSLLVIGTVG CK	C	II
15895662	Xylanase	**MIRNKRKLMKKR**AVFVCSIALIVILGT CI	C	II
15896269	Xylanase	**MNNEKQKKRRK**LLIMRSIIVFSIVII CW	C	II
Metabolism of lipids and fatty acids				
15893827	GDSL lipase	**MRKK**FSFLLIIFMMLLNVNFTVKA AS	C	I
15893964	Phosphatidylserine synthase	**MAKSS**VANIFTFSNLSCGLLSLIMTLAAS ST	C	I
15894057	Glycerol uptake facilitator protein	**MHYSLAVK**LVCEMLGTAILVLFGNGAVA NV	C	I
15894643	Phospholipase D	**MNRKK**LTCIFMGITVFFLGIGIFELIKSTAKA NA	C	I
15894935	Phospholipase D	**MKLGNSK**YLLGIMLIVIICIIVFTRGDS DN	C	I
15895068	CDP–diglyceride synthetase	**MSLNKR**YLGAVILAPLLIILFLGGIYLKIFVFA LS	C	I
15895090	Phosphatidylglycerophosphate synthase	**MNLANKLTLMR**IFLVPVFLIFIVVRQIPYGRSIATA IF	C	I
15895514	Lysophospholipase	**MRKLSKK**VFLIFAVTTIFILGNSAWD RT	C	I
15895877	GDSL–like Lipase	**MGNGYKSDKK**MTVTVLVLLIITITVIISGIIHD KK	C	I
15896276	GDSL–like lipase/acylhydrolase	**MLKK**LTIGFVVIAAGIVLLLSIFGIKEA AS	C	I
15896559	Phospholipase D	**MRFFIKFIFHR**ATFVVISALIQLFVLIFVII KF	C	I
15004719	GLPQ related phosphodiesterase	**MKRANYK**ILYAVMVVLIFGLLFSGSNNIKA DT	P	I
15893720	Periplasmic glycerol–3–P–binding protein	**MKKTSR**ILIIFAATCFLLTA CA	C	II
Metabolism of proteins, peptides and amino acids				
15893373	Endopetidase	**MKKSD**LFFIFISLAFTVIFLNC TS	C	I
15893470	Oligopeptide transport permease protein	**MLRYIVKR**LLQMIPILIGVSIVIFIIFSLAPGDIV DN	C	I
15893504	BioY protein precursor	**MKFKTRD**LVLIPLFTAMTIIGA FI	C	I
15893593	Signal peptidase II	**MITVVGVRFKKAGK**IYYFSPGDLNINKGDDVIVET AR	C	I
15893790	Carboxyl–terminal protease	**MDNKKKK**WIVITVAIVVVTNIASLFLGGRFLVFA GN	C	I
15893847	Chaperone with dependent protease activity	**MRKK**ISVLFLLCISVMIFFIKLRSA PI	C	I
15893913	Periplasmic aspartyl protease	**MNSYVKR**MLTISTAVFITMGSFASIVHA EP	C	I
15894033	Metal–dependent protease	**MSKTK**ILGLLTAIAITFNATTNLTIK AY	C	I
15894126	ABC–type spermidine transport system	**MKKHRRS**ISQYPYFLWSILFILVPIFLVIYFSLTSSD ST	C	I
15894139	Amino acid permease	**MANSNLKKDN**LSIIETIALSVAIIAPTAAMSINVSLM SL	C	I
15894242	Zn protease	**MKYLKNKH**FIVPAFIFLCSLTLFASSLVSINKTQY KM	C	I
15894888	Branched–chain amino acid permease	**MKELSKKN**TFFVGLLLFSMFFGA GN	C	I
15895037	Signal peptidase I	**MGKSLIEFGKS**IIIAIIVAVIIIMFVFETVSV DG	C	I

(continued on next page)

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Metabolism of proteins, peptides and amino acids				
15895095	Aspartate aminotransferase	**MNFSKK**AGQIAASITLEITAKA DE	C	I
15895250	TadB homologue	**MKLLK**IINLILILSLLCISYILSKR LI	C	I
15895545	Translocase subunit SecF	**MLKVIKHTK**VWFGISLTIIFLGIAFMIYR GV	C	I
15895546	Translocase subunits SecD	**MKKKGRST**IFFIISVLIIAGLAYLGSTGLNV GD	C	I
15895558	Metalloprotease	**MGKK**ILSAFIVLTSVMLYSAPIYA EK	C	I
15895675	Transglutaminase–like enzyme	**MKTN**PVTLILVFLFLFPIARGFLFK FS	C	I
15895782	Metalloprotease	**MKKK**LLSLVLTTAVVSSISGFSSVSA QT	C	I
15895783	Metalloprotease	**MKCK**LLSTILTTIILSTLGSFNNVLA SQ	C	I
15895904	Signal peptidase I	**MGKENEESKSLGQNVKE**YAILLVVALGIAVIFRTFVFARA NV	C	I
15896098	Aminopeptidase	**MKIK**LMLIPLILFASIFSFTFY KY	C	I
15896429	Oligopeptide ABC transporter	**MFLFKNKNIKTSNKSLSKQTLKRILKNK**LAIMSIIFLISILIFNFLGPFFS PY	C	I
15896449	Metalloprotease	**MKK**ISSATVWIVVFLVVIFTILMLVQGGKA TN	C	I
15896462	Parvulin–like peptidyl–prolyl isomerase	**MVKSKLNTCKKTKYKKRGIKMKSAKQ**IATALLVGMFTFSA VG	C	I
15896465	Periplasmic trypsin–like serine protease	**MNNSNGNEGNINFRSRNKLRLNKVLK**FISFITISALSGA IT	C	I
15896590	Amino acid transporter	**MTSKKEQKNELKRTLTARH**MNMIAIGGSVGTGLFFA SG	C	I
15896865	Oligopeptide ABC transporter	**MLKYTLKR**FVYMIITIWIVITITFVMM HS	C	I
15896872	Oligopeptide ABC transporter	**MFRYIIKR**FIASIVTLWIVVTFTFLLAHA IP	C	I
15896877	Oligopeptide ABC transporter	**MFKYIIKR**FIASIVTLWLVVTFTFLLAHA IP	C	I
15004769	Metalloprotease	**MAKIIKH**LINTTIAAAIVLTIGTGVHA AP	P	I
15893671	Periplasmic amino acid–binding protein	**MKKSVKK**IMAVILTAVIGVGLLAG CS	C	II
15894124	Periplasmic spermidine–binding protein	**MKNIKK**IIALAATVILTCSSLTA CN	C	II
15894753	Periplasmic Pro/Gly–binding protein	**MKKSIKR**IMVTALAIVFAVSLTA CA	C	II
15895643	Periplasmic binding protein	**MKK**WHMMSLIASANVMIFLTA CG	C	II
15895925	Periplasmic peptide–binding protein	**MKNFIR**VCIAFMLVCLILTG CV	C	II
15896427	Periplasmic oligopeptide–binding protein	**MKRK**FITVVIPSIIIASLLSS CT	C	II
15896568	Periplasmic amino acid binding protein	**MKRK**GIFASAVLALTLVIGITG CS	C	II
15896866	Periplasmic oligopeptide–binding protein	**MKSKK**LLSVILSAAVISTVALSG CG	C	II
15896868	Periplasmic oligopeptide–binding protein	**MLKRKLTK**LSAVLVSAALASSLLAG CG	C	II
15893469	Periplasmic oligopeptide–binding protein	**MKRK**VLAIFVITAMFSTILFTG CG	C	II
15893955	Sugar–binding periplasmic protein	**MKNKKNK**VVTISIIVFVGIVII CT	C	II
15894167	Periplasmic amino acid–binding protein	**MKKK**LITLSMALLMSVGIFTG CT	C	II
15895803	Reductase/isomerase/elongation factor	**MKK**IILAALLVIMSISFTA CG	C	II
15895810	Reductase/isomerase/elongation factor	**MNKR**ILLAMVVVIISFALMS CT	C	II
15895982	Predicted hydrolase	**MRKSSKIKR**IVLAVVLVIFIFAN CV	C	II
15896854	Periplasmic amino acid binding protein	**MLKK**IFSLSVVLVLTLAIFTG CT	C	II
15896904	Periplasmic sugar–binding protein	**MGWLKK**LCLFSIITCIIFSTVG CV	C	II
Metabolism of nucleotides and nucleic acids				
15894130	Barnase	**MKKKN**FAAICFTLVFMIFGAFLNNVVCVQA KT	C	I
15894550	Competence ComEA protein	**MKNKK**AVVGSITIIFIFVVLIFIG YE	C	I
15894553	ComEC competence related protein	**MKR**PLVFYFVSIILGILTAALLKIDILIGAVIA AS	C	I
15895219	Phage related transcriptional regulator	**MDNRQR**IITFKILRLASGLSAERVAAA LS	C	I
15895240	Plasmid transfer factor TraK	**MHKKFNYR**LIICIFILHTLICIYLLTPLIVIR SE	C	I
15894883	Ketopantoate reductase	**MKRIRN**ISVVGMGAVG CA	C	II
Cell wall, envelope biogenesis				
15893327	Cell–wall hydrolase	**MVNIKKNT**ALVAAFIITATTLTNVKAYA VT	C	I
15893497	Sortase	**MKK**LNIIAATLISSGVILIGFTLGA KY	C	I
15893506	Protein containing LysM motif repeat	**MSKNKLR**FILFLICVFIVGSGIVYY DQ	C	I
15893600	Cell wall–associated hydrolase	**MYSLFLWASVKIFTMESRGR**YVVHKRILSTVVAFGIVASMGSTSVFA AP	C	I
15893683	Peptidoglycan–binding domain	**MRNLKMK**ALSLGIMSTVLISSTAFA AT	C	I
15893789	Cell division protein	**MRVSTLK**LFFIDALKSLKRNKTISTAAA AT	C	I
15893797	Penicillin–binding protein	**MNNISSNIKK**VLFIFLVIFFVTISYI TY	C	I
15894034	Protein with ErfK domain	**MDTTETEKKKSKRNK**IIIGAAAAFCILLVIYLGMA RY	C	I
15894280	Penicilin binding protein	**MKRK**LLSILLSGILSLSISSSALA AV	C	I
15894385	Murein transglycosylase	**MKAK**IFIFFIIVIALFLALNTRNILKH FY	C	I
15894393	Autolysin	**MKNKNR**IVTLLTAVLVISAVIFMPLNIHA QI	C	I
15894515	Lytic murein transglycosylase	**MIKSKFK**FLVFAAALAFVGMGKNAVHA AT	C	I
15894526	Shape–determining protein MreC	**MKLFKNR**LTVTVIVLSVSFLLLISYSAKR KK	C	I
15894528	Penicillin–binding protein	**MKKAKLKKIFNR**FNALMVVVLLVFSGIIAQLINLQIL QT	C	I
15894529	Penicillin–binding protein	**MKKSAKFKKIFNR**VNVLKVVVLIVFIGITVQLVNLQILQNA NY	C	I
15894530	Penicillin–binding protein	**MKKKVKIKNIFNR**LNVLRVTVLIVFFGVTA QL	C	I

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Cell wall, envelope biogenesis				
15894533	Cell division protein	**MILNKLKIDKRTLKQ**IDFLIFLIPIVIAIFSSLNIYSA VH	C	I
15894546	Cell envelope transcriptional regulator LytR	**MYFNKKR**VIIVCCFVLLCIAIGLGSFFYVSL FS	C	I
15894549	D–alanyl–D–alanine carboxypeptidase	**MRK**LIIFTLTVMFLFSTININIYA KV	C	I
15894595	Peptidoglycan–binding domain protein	**MKNLKLK**ALSLGIVSTVLISSSAFA AI	C	I
15894804	Amidase	**MKKTLK**ILAGFVAVVIIIALGIFTSRK LY	C	I
15894965	Penicillin–binding protein	**MERTRIKC**ILIFLGIFAVFAFLCTRLYLIMNVYS KD	C	I
15895151	N–acetylmuramidase	**MKNR**IETLLVAFTFVSATFFIPLNVHA TI	C	I
15895315	Murein hydrolase	**MKRSKGLGSMFKNMAKNKGKKVLKALIKK**IMFPWGLIFCLFIFIVIFA IS	C	I
15895327	D–alanyl–D–alanine carboxypeptidase	**MKIYRS**IVLALIVCILTPATFAKA DK	C	I
115895333	D–alanyl–D–alanine carboxypeptidase	**MKKR**GIAIGAMALLLLFLQSLFLPFSVKA FG	C	I
15895394	Cell division septal protein divIB/FtsQ	**MKEENEYIVKRRKKRRRKR**ITIFLFLLICILVTLCLKLPYFNIKY IN	C	I
15895568	Carboxypeptidase MrcB	**MSDNTKTNSRNKSVKRTKKVKKKKKFGFFKK**LFTILFCLFILLSVAASGVIFA IV	C	I
15895581	Exporter of O–antigen and teichoic acid	**MNKVAKN**FFSMGFSTLVSQFLTFFTGTYA AH	C	I
15895592	Cell wall hydrolase	**MTKK**MNIIFAMFVLMFISIFSSKVNA QT	C	I
15895602	UTP–glucose–1–P uridylyltransferase	**MRVKK**AIIPAAGLGTRFLPATKAQP KE	C	I
15895682	Protein with ErfK domain	**MKISKFQK**IIFVLILILIIEIILVIQFYKM SP	C	I
15895921	Cell–wall hydrolase	**MKNTKK**IISLLFFLVAFVIITPSISSKA DN	C	I
15895953	Metallo–dependent hydrolase	**MKNLKLK**ALSLGIVSTVLISSSAFA AT	C	I
15895979	RCC1 repeats protein	**MNFKK**IFIISTVLGISISLANA NN	C	I
15896129	Undecaprenyl–phosphate	**MEKK**YIMFIIAMAISFLTTPIVRKFA KH	C	I
15896196	Cell wall hydrolase	**MKK**LFTNLLLLCSIVFLGALLNGHNVQA SD	C	I
15896197	Cell wall hydrolase	**MKK**IFTNVLLVGFLVFISAILYGHSVQA SD	C	I
15896297	Cell envelope transcriptional regulator LytR	**MKTKDKKKFGFWKK**FFIVFFILIFGIGA SG	C	I
15896314	Cell envelope transcriptional regulator LytR	**MKSKKKFGFLSK**FFITILILIVVIGVTA GG	C	I
15896343	Amidase	**MLKRDNFFKVFIK**VFSVFIAVTFVFSVMPNNVNA KT	C	I
15896541	D–alanyl–D–alanine carboxypeptidase	**MKK**IILALIIISVISASLFIYKNNQ PK	C	I
15004780	Penicillin binding protein	**MLEKRSKKINIIKR**LFLYILVFAAVFGGIVYLKYA PT	P	I
15893309	Protein with ErfK domain	**MIKKEILK**FLLSSLISSIVA CS	C	II
15893601	Cell wall–associated hydrolase	**MHRK**FLTALIALGITVS CS	C	II
15893754	Serine protease domain	**MKTHK**LLSIVLASVLLTSSFTG CT	C	II
15895098	Protein with ErfK domain	**MFKQKK**IIFIAIIVIFIFSG CS	C	II
15896915	Penicillin–binding protein	**MKRK**YLLLLTLLLTISFTFG CS	C	II
15004762	Rare lipoprotein A	**MKKAKS**FVCLMTLL CV	P	II
Defense mechanisms				
15893949	Multidrug–resistance ABC transporter	**MKYSNINKK**LGSFLSSSKNIIKTVKLLWKSSSLS FI	C	I
15894233	Metallo β–lactamase	**MRNNIKK**LSLIFCFIFTIMFLSAWSKVGNVNVMA AN	C	I
15896063	β–lactamase	**MKMKKC**LVSCLLVALISIPPMKVFA AN	C	I
15896672	HAE2 family protein	**MAKLLYNLGSWVYKRPKR**VIVAWVVFLAILGVAVLNTGIKFS DN	C	I
15896843	Multidrug–resistance ABC transporter	**MNLLKITKRDFINTAKN**PTLLLCNTAFALLLIGIFGFITKNSYG NT	C	I
15896844	Multidrug–resistance ABC transporter	**MIIINIFKNNLSRLLCKK**AILLTSIIFIPLMIIGGVYF TS	C	I
15895880	Metallo β–lactamase	**MNSLKRLLKC**VITLVCILILISG CS	C	II
Sporulation				
15893621	Sporulation penicillin–binding protein	**MKKKRKNPRLVLGKK**LLTVLFIFSIVFSALIFRLFFIMFVHS DT	C	I
15893887	Spore germination protein	**MKINNK**LLFIPVILIFIAVFIGGKGE LV	C	I
15893974	Spore cortex–lytic enzyme prepeptide	**MKGILKYKRS**IVLSLAVLISYLSIYVWLLPYNNAVN AI	C	I
15894303	SpoVB	**MKKQSTVK**GFAVLSIGTMISKVLSLVYVPLLTRILGGA EP	C	I
15894558	Stage II sporulation protein P	**MVVKVINKKQDKFYFK**LGMFILIFLILGVSLLNIIGINV DK	C	I
15894805	Spore germination protein GerKB	**MIHLSKHQ**LFTLTFIFQIGSTPLFALGIEA EQ	C	I
15895342	Stage IV sporulation protein B	**MKNKK**INIVYYILTPVILVVFLAYN KV	C	I
15895357	Stage III sporulation protein AG	**MNFKKFLSNLKFDSKDEKTKKS**TVTNIVILGLIGILLIITA DF	C	I
15895359	Stage III sporulation protein AE	**MKK**VILMFVVILILFPVTVQA YD	C	I
15895395	Stage V sporulation protein E	**MSKPKLLQK**VEVDFVLFATVLLLVAIGVIMIYSA SS	C	I
15895399	Stage V sporulation protein D	**MPKRILVDKATIRNR**IFIVFLILTCVIFLLCYKLFNVMVARS SR	C	I
15896115	Sporulation protein SpoIID	**MVKK**IFMGLAVTIIFIFSVSIFVGG IG	C	I
15896151	Stage II sporulation protein R	**MKKK**IFIVIILAIFYITFMSMNYVLA YS	C	I
15896332	Spore–cortex–lytic enzyme	**MKK**FIWAIILFTLSWCCLPNTCSA QI	C	I
15896545	Spore germination protein GerKB	**MEKKICSKQ**MFSLIVLCVLGSSIIFFLNPEAEQN AW	C	I
15004853	Stage V sporulation protein D	**MKRSSVNKKIQLSKK**LFHVFFFLVAVFFLIIFKLA YI	P	I
15893933	Spore germination protein	**MRKK**ITTICSIIMCCVMLTG CF	C	II
15895356	SpoIIIAH	**MNKKQ**AVIIVTLMVLIV CA	C	II

(continued on next page)

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Sporulation				
15895771	SpoIID-like protein	MQRKSLKFVSMFTTAALTFSY CS	C	II
15896546	GerKC	MRKLISIFLILCTFLLCG CS	C	II
15004723	GerKC	MKSCRKKFLKIISCILIAVNLTS CI	P	II
Ion transport and efflux				
15893399	Adenylylsulfate kinase	MNNKKSTNVVWQETKIKRQNREKMLKQKGAVLWFTGLSGSGKSTVASA LE	C	I
15893572	Molybdate transport system	MIIKKNHFDLDKFFICFFSIFVIIYLVILCFPIVSVVKY SG	C	I
15893661	Cation efflux system protein	MNRLTKFSLKNMFVIFLIIILITVGG LY	C	I
15893970	Ammonium transporter NrgA	MNFADTTFVFLSTVLVMIMTPALA MF	C	I
15893973	Putative Mn transporter	MKKKFSRLLFIFSIVGPGLITANA GN	C	I
15894075	Ferrichrome transport permease	MIIEKKREMKHRLVLLIIILLTVLTIVISLGIG DY	C	I
15894379	Cation efflux system protein	MKKINSAVLSIASNSMLIIFKLTAGI TM	C	I
15894983	Phosphate permease	MEKRSFLKKLKTEYVGRGFATICGVFIVILTLSIIFFIA SK	C	I
15895707	Hemin permease	MNLNLTKNMKYILLLILLFIVLIFTFIFSIST GS	C	I
15896572	Mn transporter	MKKKNKLLFILGIIGPGLITVNA GN	C	I
15896786	Na+ ABC transporter	MKNTTMIVFKKELKDIFRDRKTIIFSILIPMLLLPLISFFMSSVVNK SE	C	I
15896912	K+ transporting ATPase	MMKYFKSALRLGIVLIIICGLIYPLFITAVGQ TV	C	I
15004807	Heavy metal resistance membrane protein	MNKTSKLFWLDRYLTIWIFTAMA IG	P	I
15893573	Molybdate-binding periplasmic protein	MNMKKIITFLSISAFLVSCLSG CT	C	II
15894273	Periplasmic metal ion -binding protein	MKKKFLSVLLSGAVLIGALTG CS	C	II
15894679	Periplasmic nitrate-binding protein	MKKKKILSLISAVIMTSAIITG CT	C	II
15894871	Zn-binding lipoprotein related	MIKKFFSLIAVLIGLVVLAS CS	C	II
15895706	Ferrichrome-binding periplasmic protein	MYTKTKASISALLILITITLSS CS	C	II
15896091	Zn-binding lipoprotein	MKKFLCFIMSVGLILSLVG CG	C	II
15896515	Permease, cation efflux pump	MRNKTIMAFFIGIACFSLTA CT	C	II
15893436	Permease, cation efflux pump	MNKSKLAIAAVLVCATSLSVG CG	C	II
15893610	Permease, cation efflux pump	MNKKIINIIIGVLAV CV	C	II
15893908	Periplasmic nitrate -binding protein	MKKAYFKLLTLFTLLLCITFTG CS	C	II
15894077	Ferrichrome-binding periplasmic protein	MKKILISIMIIMMIFGSA CS	C	II
15894982	Periplasmic phosphate-binding protein	MKSKKIKLMSVAITMTIMAGLFVG CG	C	II
15895258	Ferrichrome-binding periplasmic protein	MIKIIKKTTVIAMTIAMVTTVFSG CS	C	II
Motility and signal transduction				
15893397	Methyl-accepting chemotaxis protein	MRRDSMGFIRNFTIRSKLLSLLLINVLLIVIVVLFMIEIRA EY	C	I
15893404	Sulfate transporter permease	MQKRSTTVTGIFRRVLFYIILIA IW	C	I
15893517	Sensory transduction histidine kinase	MSFIRYLKDNFRFILLYVLLGAFVSA FT	C	I
15893542	Methyl-accepting chemotaxis protein	MNKRVIGIKGKVLISSMLVLLLTLVSISGIVIYQVNSK AY	C	I
15893582	Sensory transduction histidine kinase	MKRKSSNYIKITSLFTKSIVMIFVIMFFVDMLIMIKVTEYR PT	C	I
15893609	Sensory transduction histidine kinase	MSIKTKLVFSNIAMCLIPLILCLITLFGLNNLYSR KL	C	I
15893724	Methyl-accepting chemotaxis protein	MSSKSNSKISSDLKTKFVLCSLLPSTIAMCIYGA II	C	I
15893734	Methyl-accepting chemotaxis protein	MGKKGFKNLKVSSGIIILFVLAVFSSVLIG AV	C	I
15893795	FHA-domain containing secreted protein	MVSVYNISSLAVIGASLSNMSTIFKFVIIA IV	C	I
15893815	Sensory transduction histidine kinase	MLRNREFKKFAILISLIATATVTIGFVINMSAGILTITSSA AF	C	I
15893832	Methyl-accepting chemotaxis protein	MKLMDWKVSKKVGVCFLCIIACFTIIVLIDIFS MK	C	I
15893833	Methyl-accepting chemotaxis protein	MKSGGIKANLSLLVIFMLIFAVGLGGFSFYSLNNIKS ES	C	I
15893855	Histidine kinase	MKKIFKPFLKLYSFSKKKVVKSIRLELVVTFGICLLAAFILGSWY TG	C	I
15893927	Flagellar biosynthesis related protein	MAKERKKAAALKYDLGYEAPIVTA VG	C	I
15893942	Sensory transduction histidine kinase	MLKIKSRIALLYSMLTIALIIIFIISFCL IL	C	I
15894092	Methyl-accepting chemotaxis protein	MNEKSGVNFRSIRVKLVAILLLLCLIPIMISGVYNYYKA SE	C	I
15894102	Methyl-accepting chemotaxis protein	MVNSNLLNKTEKKMNSFLFTLSLSMPVIAFI FV	C	I
15894151	Histidine kinase	MDNKKLFVFIKYFTLVFMIVGIIFG GK	C	I
15894196	Methyl-accepting chemotaxis protein	MVKGWRKLSKSIRSKLIITISLVCMIPIIIFGA IS	C	I
15894631	Chemotaxis sensory transducer protein	MNRANTDSNKARRSLRVKLITILVGIIIPMTFLSIGTFAIM KS	C	I
15894666	Chemotaxis sensory transducer protein	MRIFNDMRIFKKLILVFGVISIFTIIVGIVGL TS	C	I
15894831	Sensory transduction histidine kinase	MRSIRREVSIILILCTISGVILSALFVNIAMN AT	C	I
15894878	Chemotaxis sensory transducer protein	MSIKKKLSIVIIALVIFSTVAVGA AI	C	I
15894978	Sensory histidine kinase	MKKKLVLYNLGILIITLIMITVLFI KI	C	I
15895416	Flagellar biosynthesis protein FlhA	MEEGRRLKFNVKDNRDVIVSLAIVGIVIMMIVPLPA KV	C	I
15895418	Flagellar biosynthesis protein FliQ	MSENMIIKIMKDAITTGLIVSAPMLIVSIIIGLFISIFQA TT	C	I
15895419	Flagellar biosynthesis protein FliP	MKKKSFIMLAVFLVFLAFLGARAHA AP	C	I
15895421	Flagellar protein FliL	MSEKKKDKGEKGGGKSKIIIILLMVIIVVFIAAALAYFFVFS KK	C	I
15895431	Flagellar basal body M-ring protein FliF	MKKLLESLKNLRNKWGEISKRKKIVFLIIVVAIIASLVVFLVST TR	C	I

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Motility and signal transduction				
15895699	Histidine kinase	**MKIKGSITTK**LFVITLSFYVVFILGMLVFQSAFFG KF	C	I
15895809	Methyl–accepting chemotaxis protein	**MQFKRASLNSIRSK**LIISLISICVIPLLIIGTFSYNYA KS	C	I
15895837	GGDEF domain containing protein	**MEDKKKFILK**VFLIFVVSVFLICIFWYS II	C	I
15895875	Methyl–accepting chemotaxis protein	**MKSIKGK**ILLMIETLIIVVIVGMGILAGMLSSNALR NN	C	I
15895986	Protein containing SH3 domain	**MRKSK**LTFIATFLILLFLSMPYKAHA AS	C	I
15895987	Transduction histidine kinase	**MKN**IWMVTTFLFLFILVMLIFYH LY	C	I
15896003	Methyl–accepting chemotaxis protein	**MKLMGKMKVGTK**LVSSFLIIMVFLIIVGITA LG	C	I
15896019	Methyl–accepting chemotaxis protein	**MKRVTDLKIFTK**LIIAFIFVAVISTIAGVVGV LK	C	I
15896020	Sensory transducer protein	**MKSMASPKIFTK**LILSFIFVAVISAFA GI	C	I
15896029	Methyl–accepting chemotaxis protein	**MSISKKAKVKKSLGIKTS**LVLSFSVIIFITMSILSFMIYTKS KK	C	I
15896058	Methyl–accepting chemotaxis protein	**MKLKTK**LIIMITIMIVTSAILCLIGIIGVYS IT	C	I
15896076	Methyl–accepting chemotaxis protein	**MKVRSIRTRT**LVSVLPVIIVILMVLVSISYY FS	C	I
15896466	Histidine kinase	**MRKGIFSK**MIATYTGIIAISFVVTSAVLSFWFQNFYFS QK	C	I
15896595	Methyl–accepting chemotaxis protein	**MELLKKTKEHISFK**IILPIVICCILSIVITAVVC IQ	C	I
15896638	Methyl–accepting chemotaxis protein	**MFRGIKNVK**ISNIIIFMGVFSILLTSIVGFLGFNNMQK IN	C	I
15896714	Methyl–accepting chemotaxis protein	**MKKMKLNNIKLSVK**MIFTLIIPVASLILITGIFVRY ID	C	I
15896748	Methyl–accepting chemotaxis protein	**MIRLFKNLKVVHK**LILCFGIIITVIIMNGIMSYEKM TQ	C	I
15896753	Histidine kinase	**MKLRIK**LPLLFLVMFIVFMVSIGIYLKFIFA LY	C	I
15896895	Sensory histidine kinase	**MKIFKIKTITMR**IWMTFTVMILIIVCSISLLYISAFR AF	C	I
15896920	Methyl–accepting chemotaxis protein	**MKRIFYPGIRLMNKLKYSKK**FLLIFMVFLVPLVAILTFLINQLNNDVTV SD	C	I
15004718	Protein tyrosine phosphatase II superfamily	**MHKKTK**LICIGILATAIVTISIYN FD	P	I
15894857	Methyl–accepting chemotaxis protein	**MSFIKN**LKVRTKLAA CF	C	II
15895847	Protein containing GGDEF domain	**MEKKREKS**FLFKVFIIFMLSLISIAV CL	C	II
15896747	Methyl–accepting chemotaxis protein	**MKTIKKK**LIVAFILIILVPM CT	C	II
Virulence factors				
15893694	Protein with fibronectin type III domain	**MNYKK**ITSVILSALIISAAFSYHSVKA AD	C	I
15894169	Hemolysin III homologue	**MFSKLKD**PVSGITHLVGAVLSIIALVCMVHHSIVA DS	C	I
15894206	Sialidase	**MNKR**IVSMVAGLSIIFTTGFVTHISAAA NK	C	I
15894234	Hemolysin III homologue	**MEEDSFYTKGEE**IANAITHLVGAALAIAAIVILVVFA AR	C	I
15894364	Clostridial enterotoxin	**MLKNK**ITLLLSSIYILGSISTPTLA SE	C	I
15895758	Phospholipase C	**MRIKNKC**LVLLIPLGISIMFDIRVKA FQ	C	I
15896298	Virulence factor MviN	**MAKDKR**ISLLKVTSMVIIINLLSKITGFIRDFITA SK	C	I
15896336	Protein containing cell–adhesion domain	**MKTKR**IFSTLIIAAFVFSFTTTTNNVSA DT	C	I
15896517	Homologue internalin A, cell adhesion domains	**MKKLKCKS**LLSALACLTLLFTIIPVNA YK	C	I
15896518	Protein containing cell adhesion domains	**MKKLKTKH**LITALACLVSLSTFAYTNTCKA AA	C	I
15896519	Protein containing cell adhesion domains	**MNRFKIK**YLLATLACFVSFSAFTCTSSVKA DT	C	I
15896520	Protein containing cell adhesion domains	**MNRFKIK**YLLATLACFVSFSAFTCTASVKA DT	C	I
15896523	Protein containing cell adhesion domains	**MKKLKS**MCSPIIFITFLMFFNFEGLNIKCK AD	C	I
15896524	Homologue internalin A, cell adhesion domains	**MRRLKTK**FLFTTLACFVSFSIFSSTNTYKA DT	C	I
15896525	Homologue internalin A, cell adhesion domains	**MRELKFKH**LLSTLVCFISFSIFIPLNTYKA DT	C	I
15004707	Transglutaminase with cell–adhesion domain	**MKKKN**VVLSIVLALVFILFYSGSTLKA DS	P	I
15004708	Transglutaminase with cell–adhesion domain	**MKKN**VLLTIVLIFMLVISYSGG DF	P	I
15004851	Phospholipase C	**MKNVFKK**ITTAIIASSLTLTFSFTTFA EA	P	I
15004863	Protein containing cell adhesion domains	**MKKKLKS**VLIWFLIFTFNLSLGSFREVFA DN	P	I
15894668	Cell–adhesion protein with ChW–repeats	**MRKALK**FTFTIGVFFLLL CA	C	II
15895377	Protein containing cell adhesion domains	**MKNKFK**VSLVVFLIVILG CF	C	II
15896337	Protein containing cell adhesion domains	**MNKRIK**VFFLVFLIVILG CF	C	II
Unknown function				
15893308	Uncharacterized conserved protein	**MKKK**VTIVTIILFLLVIVGSFG KV	C	I
15893322	Membrane protein	**MNKRVVK**VEYIFMAFALFFGVLFA VI	C	I
15893372	Putative permease	**MKSKTIKSN**IILLITALIWGLAFT AQ	C	I
15893434	Predicted permease	**MWPLISNELYKIIKKKK**FWITFIVFFALFLAYSIVQY SS	C	I
15893435	Predicted permease	**MYRLITNELQKIFKRRK**VLVVSIVFIAIILFFSLVQY KE	C	I
15893445	Hypothetical protein	**MKIKVQKKT**FFIIMFLCILIGFFA AY	C	I
15893459	Predicted permease	**MGAYIFAFKMLKKNTKKN**LIYLISIIFTMAIVFNLLNVINNTNFFA AE	C	I
15893465	Predicted permease	**MHILSMIKSQLKLLFNNR**LAVFAIILAPILLTFLVSYSS SS	C	I
15893486	Uncharacterized conserved protein	**MKDKIELRTDK**LLLSLILVLSAILNFA NI	C	I
15893491	Uncharacterized membrane protein	**MSKKELLKR**YIFFLIGLFVNALGVSFITKA KL	C	I
15893498	Predicted phosphohydrolase	**MISLKSRKK**LIAIAVGVTVIMSSATALGISTFA NS	C	I
15893499	Uncharacterized membrane protein	**MNSRFKGIR**IIIFSAIAIFFLAFLYIFNITHPVKFS YE	C	I

(continued on next page)

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Unknown function				
15893500	Hypothetical protein	MKRKFILGIIAVAVIAVGGYMYTNS QK	C	I
15893519	Predicted permease	MTLCKMALRNVKSNFRNFWAYFLSTSFSVFVIYIFIAIA YN	C	I
15893529	Predicted permease	MRILHIAYYIFKKEFKSSIVMMIIFPIALIAILGNSLK SD	C	I
15893530	Predicted permease	MNIILFGIKRLMNKKRMVIFVYLLPLIFSIFVMNISTSKI KV	C	I
15893534	Predicted permease	MNVFYIAINTIKSNISDKKTLIMMLLMPIVIILILGNALK SV	C	I
15893535	Predicted permease	MTIFFNNIKRIFRKKANIIVMFLFPILFISIISSISSGGSG FT	C	I
15893536	Predicted permease	MLEFRIASKFLRSSKGQTILIILGIAIGVSVQVFIGTLIQGLQKSLV DK	C	I
15893556	Predicted membrane protein	MLNFIKYETKTVYREYLLVLLIAVLANVALMTRVGA WD	C	I
15893578	Uncharacterized membrane protein	MCKLIKCELLKLKSMLGAILAFFFPLLLVSIGLLNIYTGSV KI	C	I
15893612	Predicted permease	MQLTRLIKMSIQSIWANKVRSFLTMLGIIIGIASVIVLVGI GQ	C	I
15893623	Uncharacterized membrane protein	MSNKNMFKTRQLTVIGLLSAISIVLGLTGYGFIPLPIAKA TI	C	I
15893632	Predicted membrane protein	MLLRLMPSMYKIIRKTLALIG IE	C	I
15893640	Predicted membrane protein	MKFNLINVIKNIAKSVILGIVVSYIFLLVFSNFINSGDA VG	C	I
15893645	Uncharacterized membrane protein	METLKKNWTKFILSFIMLVLAIICVYTLSTYKSSEKN TG	C	I
15893665	ABC transporter permease	MKYKKVLSNSLKINFFSYVGFIISSSFAIMITFIYS TV	C	I
15893710	Predicted membrane protein	MKKRINAIITILAFIFLGLYVSKW SI	C	I
15893711	Conserved membrane protein	MNKKLSFGQRYSKLRSGYKIVICFLLIMIAIILFEILTHGYK RN	C	I
15893712	Uncharacterized membrane protein	MSHDTKKLARAALLLAIAIVFQFIGKN FP	C	I
15893743	Permease	MNKNFRNMLIAFMGAFFLSLFSNTLSPFITTIKN TY	C	I
15893745	ABC transporter permease	MNISKILALRYIKRQKRRTIITILGIVLSVALFTAIGTMLASLQHY QT	C	I
15893782	Uncharacterized probably secreted protein	MEKGKKVFIPIVTAVFLILVLAIFMIYYFINR EH	C	I
15893787	Uncharacterized conserved protein	MKKFKEYLYITLGFLIVAASVKFFFEPNNIA GG	C	I
15893803	Uncharacterized conserved protein	MKFIDWIKPGIKLKRWIMLGGMGVLFISFALA EL	C	I
15893817	Predicted permease	MNVFNKVALQGMKKSRTRTIVTVIGVILSVTLITGVTTFG IS	C	I
15893863	Uncharacterized protein	MKVNVKKIVLYLVGICLISYGIAMA LS	C	I
15893889	Predicted membrane protein	MNDGRMLGLFTFIIMIISIISQIIALSA SQ	C	I
15893935	Predicted membrane protein	MSKTQSLTRGSIYAAVCIVFIYLSTIIPAGRLSFLA AA	C	I
15893956	Predicted membrane protein	MNIIKLTKVRLIIGVVVLLTLIIASNYLNFRV ND	C	I
15893958	Hypothetical protein	MKNIVVAVSLFVVILLLSFFSIRY LN	C	I
15893965	Uncharacterized conserved protein	MKDKTKKRLIYAIIALIVIIAIFVVVIVKR QY	C	I
15893979	Uncharacterized conserved protein	MKKVVIGAFSSAIIIILIVLSSIYIKVKK WD	C	I
15894005	Predicted membrane protein	MVYSNIYFGGIELKKVRTKLSLIVSAAFTLSFIGILPPQLMA KT	C	I
15894020	Predicted membrane protein	MNIKRNIRIMLTIVLCIAVFFQILRLNFVMNFF KN	C	I
15894029	Uncharacterized protein	MKFKKTFLSLALTFILSQSYSSVLL AR	C	I
15894044	Predicted membrane protein	MKKFNNFIRKTLIIALMVLLISGQKVAVLA RA	C	I
15894076	Predicted permease	MDKTNSITNFSSKLILSVIILIIMFILSLILGA EK	C	I
15894110	Predicted membrane protein	MKKTFFKNLFKSISTTLSRFLSILIIIAVGVSFYVGVRA SS	C	I
15894116	Uncharacterized membrane protein	MKKIKNKKKIFKVILFGLFIIYIIALFKIILFKS IS	C	I
15894122	Predicted integral membrane protein	MRTIWLVFIANFKRAYSNKKKFFINLLVPVAAIILAIVVNYVSS PS	C	I
15894135	Uncharacterized conserved protein	MTKKISSIENNAKLNFLMQNLMIIVGSLIVALGFNLFLSPSKI AS	C	I
15894155	Predicted membrane protein	MVNFFVIVLANCRRYLKNFRYTASMIVVPILLTGFVTFFNLGNFGTDA DD	C	I
15894162	Predicted permease	MTNENIVDSNLRKKGLIFAVLASTLWGISGIAAQY LF	C	I
15894177	Uncharacterized membrane protein	MLPLKLKIIPSNKYKQYFLVFIALLSLLLLGVYIVNS NE	C	I
15894211	Predicted membrane protein	MIRFQKRMAIFFLTISILIICGDIEK II	C	I
15894239	Hypothetical protein	MLKNFIFKNKKPLIAAFLVILIIFYIDHRINL NK	C	I
15894243	Hypothetical protein	MKSLKGVTTGAVIGAVTGMIIAPQLSKSA RK	C	I
15894253	Predicted membrane protein	MVNEMRIFERNYNMKIKINNLKKKGLVVYILLILLFRLILATGLRVWA FG	C	I
15894254	Predicted membrane protein	MNTKMHNKKKRIFSIYILLILLIRLFLSTGVRIWA FG	C	I
15894281	Predicted membrane protein	MNVTTISLKYFFFALTIIGALLSFYYRSLVSK TS	C	I
15894282	Predicted membrane protein	MKNKLLNASLAFFISFILGPILLY VI	C	I
15894297	Predicted phosphohydrolase	MKKRKTFLSLLIAAVATFGIYSATA IK	C	I
15894305	Predicted membrane protein	MKKILKCIIFIVFVFA AI	C	I
15894329	Predicted membrane protein	MHMAFSIKTLAILFLLMIAVMSVYNFTKRFFI SK	C	I
15894345	Predicted membrane protein	MKKLLENYKVMISLISISLFILVKNTGQA NF	C	I
15894359	Predicted permease	MSKSTKVDLVFLFFTAIWGLSFPLTKNVLGYTSIFA FL	C	I
15894363	Predicted phosphohydrolase	MKKSKKLLKTLTATICTMGIICATSITVNA QY	C	I
15894365	Uncharacterized protein	MSRIKNILLVFSISVTTSAVLLSKPAIKAYA AD	C	I
15894366	Uncharacterized protein	MNNIKKTLLIASVLALGTKMGMDLNAVKAYA AE	C	I
15894368	Membrane–associated metal–binding protein	MTGKTHISIGLMSSAAVLSVNNPTPENIA VG	C	I
15894380	Uncharacterized protein	MNKNNKYKATMLLAIFAGGYFISSA FR	C	I
15894463	Predicted membrane protein	MKKSGCLGCFVPIFLVILNISIGA WS	C	I
15894489	Predicted membrane protein	MIKKFSLLVISLFFASLIAVSISA RG	C	I

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Unknown function				
15894521	Hypothetical protein	MRYTRYDLKHRKSSSIPFLMCLTIILALSFFIGTFIF KT	C	I
15894559	Predicted membrane protein	MIKRYKIIFCMMVFFGVIIFIILNNIEL NR	C	I
15894574	Membrane protein with hydrolase domain	MKIVQLMKKDKSIKVMTFLFAFIIIYVILLTSV DT	C	I
15894594	Hypothetical protein	MKFNKITSLLIAFSIIGTSTLSVSA TT	C	I
15894605	Uncharacterized conserved protein	MKRKYSHVFTFINKILLMVLGSVIA SV	C	I
15894645	Predicted membrane protein	MKSGSNSKNQIKWMDSFTKTLLAVMFVILIFIFVSGKYM EE	C	I
15894681	Unchracterized conserved protein	MPTLTFKLIFSYRIFIYLLVFVLFPILLFY NS	C	I
15894779	Predicted membrane protein	MVSFLTPNYKLLSVILIISSIFIPPTRFIS FT	C	I
15894783	Hypothetical protein	MSIKRISMFLLAAVVVGSASLSIY EN	C	I
15894798	Hypothetical protein	MRRFFVYFIPTALLFLFIFIMLSGNLLKNPIGK ET	C	I
15894810	Protein containing ChW–repeats	MKNKKIVGFIAAAAVAFGLGLPSNSHFNA DT	C	I
15894811	Hypothetical protein	MMIKKLEYRKKALFIIILSIFIVSTIEFYVFA MN	C	I
15894836	ABC–type transport system	MGKISSLFKLRPFVKKYRGIFVLGILGMIISSIVQTPVPY II	C	I
15894840	Predicted membrane protein	MAFNKVFRKLLVSFSVIFFITGSSTFA RQ	C	I
15894843	Predicted membrane protein	MKTNMKRSKEKSTSGIFGKKFITSIIFIIFLISFSA EN	C	I
15894894	Hypothetical protein	MKTNILHKATVMLLITLTLSLSISTIFT KN	C	I
15894962	Uncharacterized protein	MKVQTILKKKVYIFLFLILVLICASA AY	C	I
15894979	Hypothetical protein	MNFTLMKKMKKEKRVSIAYRIAFIVTGLIISIMAAA DG	C	I
15895045	Uncharacterized conserved protein	MKNHKKLITLSILLLILISLIAVYNIKKY TK	C	I
15895051	Predicted membrane protein	MKIKGIIAISMSILLIGSISSVVKA SS	C	I
15895136	Flavodoxin oxidoreductase	MKVIIIGAGWAGAAAAITAKKAGA DV	C	I
15895137	Hypothetical protein	MNKRRLLVLTLTLAITFPTAVFA KG	C	I
15895143	Uncharacterized protein	MLNKTGKIILGAVAIILILLLGVAIGSKK GN	C	I
15895182	Predicted membrane protein	MKRKRLIVKRLSIISSITVSVFGILLILYGLGVINCY DP	C	I
15895233	Predicted phosphohydrolase	MKVKKRVIKTNVVLFFEFIFLLNMISGA DF	C	I
15895238	Surface–layer related glycoprotein	MKINKILMVSIIFIFIFNLNICVIVK AD	C	I
15895244	Hypothetical secreted protein	MKKFSILNGRNKFTKRAKIKSNKMKLLIGAMGLILCCELLVGVVYK LH	C	I
15895246	Hypothetical protein	MQISKLKGDASIAVLSLFTVILMG LC	C	I
15895247	Hypothetical secreted protein	MDSVINKRKAGGVQIVVATLLGVLIMSSVFIYF IS	C	I
15895253	Predicted membrane protein	MNFVTKITKNHIINTIIIAIAAIILLVSLSINPIIG KH	C	I
15895273	Predicted permease	MRILIKYVFKSMWEKKLRSLLVIFAIAVSVSLFYSSLSV KN	C	I
15895296	Predicted flavodoxin	MLKKIILIIGIILILLIIVMVLMFSRLNKPRQA NQ	C	I
15895314	Hypothetical protein	MHGKLKVVILASVMFSIFLLGTIFSTLEFVSE NN	C	I
15895320	Predicted membrane protein	MKKSKVLGFVSSVFSLLFMKVRVYA DS	C	I
15895366	Hypothetical protein	MRNKMLIFSVLICLLFFLSMLNIEKY NE	C	I
15895376	Predicted membrane protein	MKNIKILLFNILNLIIISIMLLTVYL KN	C	I
15895391	Conserved membrane protein	MKNNEASIFIFIASVFLGVLISSNMSFGNKEK KL	C	I
15895393	Conserved membrane protein	MKKFGAQIGVAFVCCILGFMIAHQLKIVSFQ DT	C	I
15895498	Uncharacterized protein	MKKIYSICVAVFILTSTFGGSVEA QK	C	I
15895519	Uncharacterized membrane protein	MKKVKITKEGIALCIIVLISAVLNFFNLSIEGYGNAYY AS	C	I
15895523	Predicted permease	MNILTKFITKSILEKKSRAFLIIFSILISTTLLVA SL	C	I
15895548	Predicted membrane protein	MEKSSKIYVSAAGGVLRAIFVTLIVILIFSFVSTKISFS EG	C	I
15895561	Hypothetical protein	MKNLKLKALSLGIISTLLISSSSFA AT	C	I
15895567	Uncharacterized protein	MGYIFSKKTKVIIIIILTFMLVLFNSVIYA FD	C	I
15895590	Predicted membrane protein	MIKLKNLMNKIIKFRYLIAFIIFILCVASDFNF SS	C	I
15895658	Uncharacterized ABC transporter	MKKGTLFRLASYVGKYKMYMFISLIFALISNVLIA FM	C	I
15895669	Predicted membrane protein	MKKTRDKLKKFFYKYKYDFLCILVLSSLTILILHPLYKSGHVVFS DI	C	I
15895679	Predicted membrane protein	MNKRSTLRILISNALIAAVYA VL	C	I
15895684	Uncharacterized membrane protein	MNKTNTKFLVTTALFIAISIIIKAFG IS	C	I
15895687	Predicted enzyme with TIM–barrel fold	MIKRRTGVKVFLGLCSIVLIGILILSNVSSTVGA YE	C	I
15895713	Flavodoxin dehydrogenase	MKSTKILILGAGPAGFSAAKA AL	C	I
15895720	Hypothetical protein	MKKRILVTGIIMLIAAAAILLFSRYKSYA LV	C	I
15895728	Hypothetical protein	MRKSILAIVLILLSLIGYSKG KV	C	I
15895743	Predicted membrane protein	MDFKLFKDIKTRYLFIAMTAVTIVSGILLAFINVLFN IS	C	I
15895744	Hypothetical protein	MKEVKFMLRKALAIVLFSMIICIIVGWNFNRVNERA EF	C	I
15895749	Predicted membrane protein	MRIIGLRTLKTALGAAIAIIIAEFLGLKYA AA	C	I
15895757	Predicted membrane protein	MVLRKLKKVLGKLSLIIAVLSIILILNVFLKF IP	C	I
15895762	Hypothetical protein	MNRKKKIALAAVFLTTIIVLLSFNQVKS QE	C	I
15895772	Predicted membrane protein	MLKMKRWKLNLIAAAAFAIA AV	C	I
15895776	Predicted membrane protein	MFKKIINLISIIIITAYIFVTIYA FY	C	I
15895785	Hypothetical protein	MKKRLLIIFFSVIIICSSCSTAYA AA	C	I
15895795	Protein containing ChW–repeats	MRSNNIVTITFIAVIIGISMLCFKPSVAYA GA	C	I

(continued on next page)

Table 2 (*continued*)

GI	Function	Signal peptide		Gene location	SPase
Unknown function					
15895796	Protein containing ChW-repeat	**MKKKK**IMALAMAAALTGASVINICSRGVKA	ES	C	I
15895815	Uncharacterized membrane protein	**MNR**GTTLFLKLAVIFIGIPILALC	VF	C	I
15895840	Hypothetical protein	**MKRILKR**FFSLFIAALILLPLSVNA	AS	C	I
15895843	Uncharacterized membrane protein	**MKKIFKIFLRDLKSIKRS**PSAIGMIVGLCFLPSLYA	WI	C	I
15895844	Protein containing ChW-repeats	**MKSNK**IIGFIAAAAFVFILGMPTDKSYK	AD	C	I
15895846	Predicted membrane protein	**MENLKKKFNRYGSIIK**YLSYSIFVTVIDIAAVWVIMSFLNVGIVYA	NT	C	I
15895851	Hypothetical protein	**MKKVIK**FLTIMFLVIEVIIVRPYTVRA	ES	C	I
15895856	Hypothetical protein	**MRKK**IITFAIFTLFVVVGLIPSKAAFA	EV	C	I
15895876	Hypothetical protein	**MKSKMK**IIFSITMFFALVSVGVIKLNKENAAAYA	AK	C	I
15895887	Hypothetical protein	**MKKLRN**VIICVLISIIIQSGLLYYMNAYM	FK	C	I
15895888	Uncharacterized protein	**MSKKSRKADKRTRYKKKKSIVKT**IILFLAFEIIFTGVTVVPYS	LY	C	I
15895891	Uncharacterized protein	**MYSNNRNIKNSRGKKHRKNLKRKKTSKIK**LFIYFLMFQFLFGIGTAPW	IV	C	I
15895893	Hypothetical protein	**MKFKIFFLRLK**HIYYIILAIVFIVLFTIFLVSRKS	IE	C	I
15895901	Hypothetical protein	**MAKPSIFSSDYEKQQRKRKQRRN**ISIFLLICVIFLVLVYLGFKTNLKN	VK	C	I
15895906	Uncharacterized membrane protein	**MEQNIKIDKSSKELRKMARQQLKGK**WGASALILLLLFLISLSFA	IP	C	I
15895930	Predicted membrane protein	**MVKKT**VECLGWVFVSLTTLALIFTVLTTYQL	VT	C	I
15895962	Uncharacterized membrane protein	**MLK**LSNFLQVLLSILVFIFIFFIIVKFTVA	FK	C	I
15895988	Predicted permease	**MKIMNEYTYNQIKKNKRHT**ISILVAITIASALLCSLCIFIYSMWSA	KV	C	I
15896009	Uncharacterized membrane protein	**MIKKIYGSRLQTRQ**ITMVGLLFAVTIVLGATGLGFLPIPPFKT	TI	C	I
15896014	Uncharacterized protein	**MKLRSIVSK**LLVLFILTYVVSAMPLKAKA	DA	C	I
15896018	Polyferredoxin	**MIKKISKLQIAR**FISQLIFLFLLPGIFVLA	FS	C	I
15896022	Uncharacterized protein	**MVKK**LVIAAIGVLVSIGLVVGVRY	FV	C	I
15896037	Predicted permease	**MNSKKIFTDKK**LVVLFASICCFLWGSAYPG	VK	C	I
15896096	Conserved membrane protein	**MKNSKLSTLIK**ISLLSVIGFILMFIEVPLPIFP	GF	C	I
15896114	Hypothetical protein	**MNNDWKNNKSNKQSRVMDFLKRN**AFYIVLFLCLCIIGA	VS	C	I
15896126	Predicted membrane protein	**MNLNIKRMLK**VVTLYDAIIAAIVSVILLFAANYKISLIVIIGIFSA	IF	C	I
15896137	Predicted membrane protein	**MNLDKAIKKQNS**YDSLFILFMCFIFCALPVVLV	HS	C	I
15896149	Uncharacterized protein	**MKSSAKR**IAYTLAVSLIVVFSTSYA	IL	C	I
15896181	Predicted membrane protein	**MDNFSFSKSIKT**ISQSPSAIATLVGVLILIAVLLKVKKIK	FT	C	I
15896183	Uncharacterized membrane protein	**MIYMAFLHLKRILVQNKR**GFIITILFPTIVVSLVAFFSNGVSSNTTINIA	VV	C	I
15896191	Hypothetical protein	**MSKIFKTKT**LITFLTLISLIFVGGNKILA	LS	C	I
15896228	Hypothetical protein	**MMQMSFLNKDIFRN**IAILGIVILAALLFIKLIPVIIIGGLA	IW	C	I
15896236	Hypothetical protein	**MKFILVDGKK**LGAIIIVVGLMLVLFGIGNNFND	KI	C	I
15896275	Hypothetical protein	**MLIMGFLKNKK**VVISLIIIICLLFLSIISYLSFKYVKA	AN	C	I
15896277	Conserved membrane protein	**MIKLIKNEIIKMLFKKK**VILIMAIILIEACAFA	YG	C	I
15896284	Uncharacterized protein	**MNKR**VFLSVITMLAIIITFNNNVGNALA	AQ	C	I
15896325	Uncharacterized protein	**MKK**IFLILLCILCFSFIFYNSSQN	AN	C	I
15896329	Uncharacterized protein	**MEKKTKQEIVIR**IMCFIAAFCLWLYISNY	EN	C	I
15896416l	Predicted membrane protein	**MTKK**ISTFTSLIFILVICIIGFIKFRTIFMNVLVIPA	KI	C	I
15896486	Uncharacterized membrane protein	**MIWGIFMIKNFKT**ILIIANVFIGTIVGA	GL	C	I
15896491	Hypothetical protein	**MNGNFKNRR**FMYSIGSVILIVLVAAVGFM	KM	C	I
15896510	Predicted membrane protein	**MIFLK**VFLMILATILIILLVCLLAKV	KY	C	I
15896513	Uncharacterized membrane protein	**MNFQRFKAIVKKEVIQLKRDR**ASFGIAIMMPIVMIFLFGYA	VN	C	I
15896527	ABC-type transport system	**MIKLLKKLK**PFWISVIALLFLAFFQSMA	DL	C	I
15896532	Predicted membrane protein	**MGEIMNRESKK**FYLFSGIIAFAICVIWVIFVKTA	PY	C	I
15896563	Predicted secreted protein	**MLSISKQYKIR**FTILILIILIISPSFNAVA	AL	C	I
15896600	Hypothetical protein	**MNNNLEETKRLNAKS**ASQKSFASSSSFKNA	AS	C	I
15896611	Predicted permease	**MSKDKKR**IIALAIFLLGIFMGA	ID	C	I
15896630	Possible S-layer protein	**MSRKDKR**VVSFLVMFFMIAGFVVNMKDIANA	SP	C	I
15896644	Predicted membrane protein	**MNIINLFKNNIKRILKQK**AIIIIAVIVVPIMICVAILFSTKAEM	KT	C	I
15896656	ABC-type transport system	**MMKKNNAPKGSPLKTLKR**LLLYVFTQYKLLFFVVIFTIFISAISNVVG	TL	C	I
15896700	Hypothetical protein	**MRRRTAIIK**GSLISAIIVIVFFGVFQIYKS	YF	C	I
15896793	Probable S-layer protein	**MRKTK**MSIFGVMVFTCMIFVVGLNASKLKA	QQ	C	I
15896796	ABC-type transport system	**MAIILKFMLKNIKEKKLRT**FLVVFSIMASAALYFA	SS	C	I
15896800	Uncharacterized secreted protein	**MKK**IMVSIALLLTIILGLSFNVKA	SE	C	I
15896816	Hypothetical protein	**MKIGNKK**VILPVVFLFIIAIYVGFNVFR	YE	C	I
15896817	Predicted permease	**MKLSDCIKMAFSDLNKRKFRT**VLTSFGIAIGTLLVILMGGF	GE	C	I
15896818	Predicted permease	**MKFKDGLRMAGRDLTRRKGRT**FWTSLAIAVGTMLVVTLVSLGTSGEKM	AM	C	I
15896846	Hypothetical protein	**MKNLKKK**LLSAGVILSVTLGASLPVLA	NN	C	I
15896715	Uncharacterized membrane protein	**MKEKKS**FILLWGGYIIGFIVLLLALNAYEGA	KY	C	I
15896907	Hypothetical protein	**MLKKRTK**INIPIIIIVLLIVWILYSIRPVEQI	SL	C	I
15896908	Uncharacterized conserved protein	**MKRNFMKSIIETLSFNKR**AISIIIGLCIIFTGSIFISSYHKSR	NI	C	I

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Unknown function				
15896911	Uncharacterized protein	MRKFTFVVVLLFIVGSISYYLFRSL RE	C	I
15896925	Predicted membrane protein	MGNMNSRVKKLLFLVLCSFIIIGYLSVNHISVLA SY	C	I
15896928	Uncharacterized membrane protein	MRNKTKMTNSKSFSEKFDKKMKKSRLIESGVILFLLVCVAGL TF	C	I
15896943	Hypothetical protein	MLKASVSNKARCYFWWHLLLKSGLLLFICMVLGIVFLASPKCFA VD	C	I
15004788	Hypothetical protein	MTIKCISILMEMNIMNKKIFFGSFTIIAFLVTTMAVY KN	P	I
15004789	Hypothetical protein	MKKEKKLMIALCVIPLVVLALIVLVLPDQIPLHFN YK	P	I
15004805	Membrane protein	MFLKGLMKLNKLLNKSYLSIVFWALILTVILFTMPDMGKL VR	P	I
15004815	Hypothetical protein	MRKKLLSLVMTGAIALTLIPAVNSSA AK	P	I
15004826	Hypothetical protein	MKIHLKRLIPIICFFIICLVIFLKFIYSNSPEKA VS	P	I
15004838	Putative oxidoreductase	MKIIKTINMGIIGCSTILPRAIIGPAKA ID	P	I
15004855	Hypothetical protein	MKKKLFLVSIVLILGIASTTMYLKHKSKA KV	P	I
15004850	Membrane protein	MKHYMIVFLRSAISYIILLLFTRFMG KK	P	I
15893328	Hypothetical protein	MHNGKLIKSILISLFVIFIVG CG	C	II
15893346	Hypothetical protein	MKKNKLKILAAALGVTIIVGLTG CN	C	II
15893347	Hypothetical protein	MKKNRIKIIVAILGVVTVVTLAG CD	C	II
15893502	Predicted membrane protein	MKNIRKSAVILIILLLVPINLIG CK	C	II
15893512	Hypothetical protein	MKKMVLISLAAASLITLSA CT	C	II
15893951	Hypothetical protein	MSKRLKKNLIITGLIIAICTVVSG CS	C	II
15894157	Hypothetical protein	MKFKKGIIIIFTFLIIGMLSA CD	C	II
15894275	Predicted membrane protein	MTKNNILNNKITKVILFRICFIIAITTALV CG	C	II
15894522	Predicted membrane protein	MNGKTFKNVLFLSLIG CV	C	II
15895339	Hypothetical protein	MIIRILIVIATIFVAVIFLGIIG CI	C	II
15896071	Hypothetical protein	MKNIKTKITAIILGIFTVVSLAG CS	C	II
15896077	TPR–repeat–containing protein	MINYKKKIAALMLLATCLSG CS	C	II
15896250	TPR–repeat–containing protein	MRKNYTIYASIIFLITTLILSG CS	C	II
15896952	Hypothetical protein	MKKIFSFIAAISVSASLVG CG	C	II
15893351	Hypothetical protein	MLKMNRVFIIFIFVLLAVG CS	C	II
15893367	Hypothetical protein	MNKKILLTTLIGASIFIMAG CS	C	II
15893441	ABC transporter permease component	MMWNLIKNEFIKLRYKRKYIICTIVFALICIGF CA	C	II
15893501	Predicted membrane protein	MKNMKKGVVMLLMLLLVSTMLMG CK	C	II
15893679	Hypothetical protein	MKEKKIKILCSIAIMFMILLSVG CS	C	II
15893777	Hypothetical protein	MKGFWFKIVLIFVVFLGVFMLFG CN	C	II
15893990	Med/BMP family	MIKKKTIAILTTVMIVAGLFAG CS	C	II
15894093	Hypothetical protein	MKKITIICLSLFLIMLPA CT	C	II
15894115	Hypothetical protein	MKTIKYLKSNLPSFFKKLTILSPLFIFLLTG CN	C	II
15894129	Hypothetical protein	MKKSIILIFLLILICAAVFTA CQ	C	II
15894352	Hypothetical protein	MMVIIQGKQMKKIVLILICIFTFSLVG CN	C	II
15894378	Hypothetical protein	MNLSKLVIIIKAKKITFLILLVIFCTSFSG CN	C	II
15894466	Hypothetical protein	MTKRARVGLFGLCVISTL CL	C	II
15894598	Hypothetical protein	MKKKVYKNRLALFLSAIIILTSFVG CS	C	II
15895138	Hypothetical protein	MNKKILAMLILLSITFSA CT	C	II
15895227	Hypothetical protein	MSLLKKKSSKKLLALTCIISIIVTVFAG CG	C	II
15895245	Hypothetical protein	MNFRKSNIKILISTVLMSCAAFG CL	C	II
15895256	Hypothetical protein	MKKYKIINKFVTIILALFIPLFAFSG CS	C	II
15895554	Uncharacterized protein	MTLKKKVILSVVALLVTFIFSF CI	C	II
15895719	Hypothetical protein	MDKLKRKICICMSLAVVMFLCG CR	C	II
15895726	Hypothetical protein	MKKGKILSAILIALVAAVLGG CS	C	II
15895727	Hypothetical protein	MKKRRILSALLIVSIVAVLGG CV	C	II
15895798	Uncharacterized conserved protein	MLKKYLKFISLSLVIALVTVG CS	C	II
15896056	TPR–repeat–containing protein	MLRKIIFIIKIMSIFMLLLSLLG CT	C	II
15896117	Predicted membrane protein	MKYNKLIFCFALLMFISLIFI CG	C	II
15896235	Hypothetical protein	MKHRRLLGFFLISIMIFSMITFTG CQ	C	II
15896290	Hypothetical protein	MKKIVLIVTLVIFFLATG CK	C	II
15896413	Hypothetical protein	MKRKLILIAVVIMFMFTG CS	C	II
15896485	Predicted membrane protein	MKRFFSYICILLIVFTSTS CT	C	II
15896779	Predicted membrane protein	MKKKGIIILILASLSLTFSS CT	C	II
15896794	Uncharacterized conserved protein	MKKIKMVSIMALTGLMIFAAG CS	C	II
15896847	Hypothetical protein	MKKKILILLVILILILTSVG CK	C	II
15896924	Hypothetical protein	MYNKKALMSLIIIFIVIFSG CN	C	II
15894596	Predicted membrane protein	MKKVIGIISIVLFVLISFQS CA	C	II
15894812	Predicted membrane protein	MVINKKIKRTMLQNKSQYIGSLALIIIS CM	C	II
15896268	Uncharacterized conserved protein	MGNKIFNLIVVVLCA CI	C	II

(continued on next page)

Table 2 (*continued*)

GI	Function	Signal peptide	Gene location	SPase
Unknown function				
15896331	Hypothetical protein	**MKKNFFNKKT**ALLMVTSILS CI	C	II
15896571	Predicted membrane protein	**MKSN**IMNGILSIMIG CI	C	II
15896696	Uncharacterized conserved protein	**MKISTKD**LVLCALFAAIS CV	C	II
15004761	S–layer glycoprotein	**MEYSTVQFIGKINILLKK**LIIAVNYKKLIAIISFSPILLVG CN	P	II
15004859	Hypothetical protein	**MYKK**LLVGALILLNVVFYG CS	P	II

[a] The positively charged n-domain is indicated in bold letters, the hydrophobic h-domain is shaded grey, and the c-domain is underlined. Potential cleavage site is indicated with a gap in the a.a. sequence. I and II stand for SPase I and SPase II, respectively. U stands for uncleavable signal sequence. C stands for *C. acetobutylicum* chromosome, while M refers to the megaplasmid pSol1.

3.2.2. Cell-surface display of Sec-translocon substrates

Once translocated through the cytoplasmic membrane, the signal peptide is cleaved from the precursor protein, and the mature protein can either be released into the extracellular medium or remain in contact with the cell envelope. In Gram-positive bacteria, 5 major types of surface proteins are currently recognized: (i) proteins anchored to the cytoplasmic membrane by hydrophobic transmembrane domain(s), (ii) lipoproteins which are covalently attached by their N-terminus to long-chain fatty acids of the cytoplasmic membrane after cleavage by SPase II, (iii) proteins binding to component(s) of the cell wall, (iv) proteins attached to the cell surface by S-layer homology (SLH) domains, and (v) proteins covalently anchored to the cell wall and possessing a LPXTG motif [93,94].

A number of membrane-associated proteins could be identified in *C. acetobutylicum* (Table 2), including various permeases, PTS components, ABC-transport system components, enzymes (such as sortase and peptidases), protein secretion systems components (such as the Sec translocon and flagella export apparatus), signal transduction system components, and adhesins, as well as proteins with unknown functions. Despite the prediction of a cleavage site for all of them, it cannot be ruled out that for some of these, the signal sequence remains uncleaved and allows the anchoring of the protein to the cytoplasmic membrane. Putative lipoproteins are also listed in Table 2.

Several domains have been shown to be involved in non-covalent attachment to the component of the cell wall: (i) the repetitive hydrophobic GW modules which contain highly conserved Gly-Trp dipeptide and bind lipoteichoic acids [95], (ii) the choline binding domain, also called cell-wall binding domain of Type 1, involved in the specific recognition of the choline-containing cell wall teichoic acids [96], and (iii) the cell-wall binding domain of Type 2 [97]. It is worth mentioning that p60 of *Listeria monocytogenes* (GI: 16802625) is also known to be non-covalently attached to the cell wall [98,99], however, domains permitting attachment to the cell wall have not been yet characterized. In *C. acetobutylicum*, proteins possessing GW modules similar to internalin B (GI: 16802478) or autolysin Ami (GI: 16804596) could not be identified following BLAST and PSI-BLAST searches. Similarly, from Pfam, searches no protein bearing cell-wall binding domain of Type 2 (PF04122), as in Cwp66 from *C. difficile* (GI: 11066029), could be detected.

3.2.2.1. Choline-binding proteins. Choline-binding domains are found in the autolysin LytA and antigen PspA from *Streptococcus pneumoniae* [96]. These 20-a.a.-residue-long domains are characterized by conserved aromatic residues and glycines are found in multiple tandem copies. They are suggested to be responsible for the specific recognition of choline residues of teichoic and lipoteichoic acids.

While several proteins bearing such a domain were previously identified in *C. acetobutylicum* NCIB 8052 [100], Pfam searches revealed that only one protein (GI: 15894364) displays such domains (PF01473) in *C. acetobutylicum* ATCC 824. This protein, annotated for some unclear reasons as an uncharacterized protein related to enterotoxins of other *Clostridiales*, bears a higher number (more than 50 times) of repeated choline binding domains (*E*-values ranging from 1.0×10^{-2} to 4.9×10^{-6}).

3.2.2.2. ChW proteins. Previous analyses of the *C. acetobutylicum* genome revealed the presence of a putative extracellular molecular complex [9]. All the proteins involved in this system possess ChW (Clostridial hydrophobic domain with a conserved W residue) repeats. This repetitive domain can be found several times along the protein sequence and was proposed to function in either substrate-binding or protein/protein interaction [9]. Pfam searches identified 20 proteins possessing such ChW repeats (PF07538) in *C. acetobutylicum* (Table 3). Two of them (GI:15004707 and 15004708) are encoded on the plasmid pSol1. While no function could be attributed to four of them (GI: 15894810, 15895795, 15895796, and 15895844), additional domains related to adhesin, internalin, glycosyl hydrolase, or protease could be detected for the remainder. Therefore, this system was proposed to be involved in the degradation of polymers (glucide, protein) and interaction with the environment (substrate, cells) [9]. No experimental data is yet available for the formation of

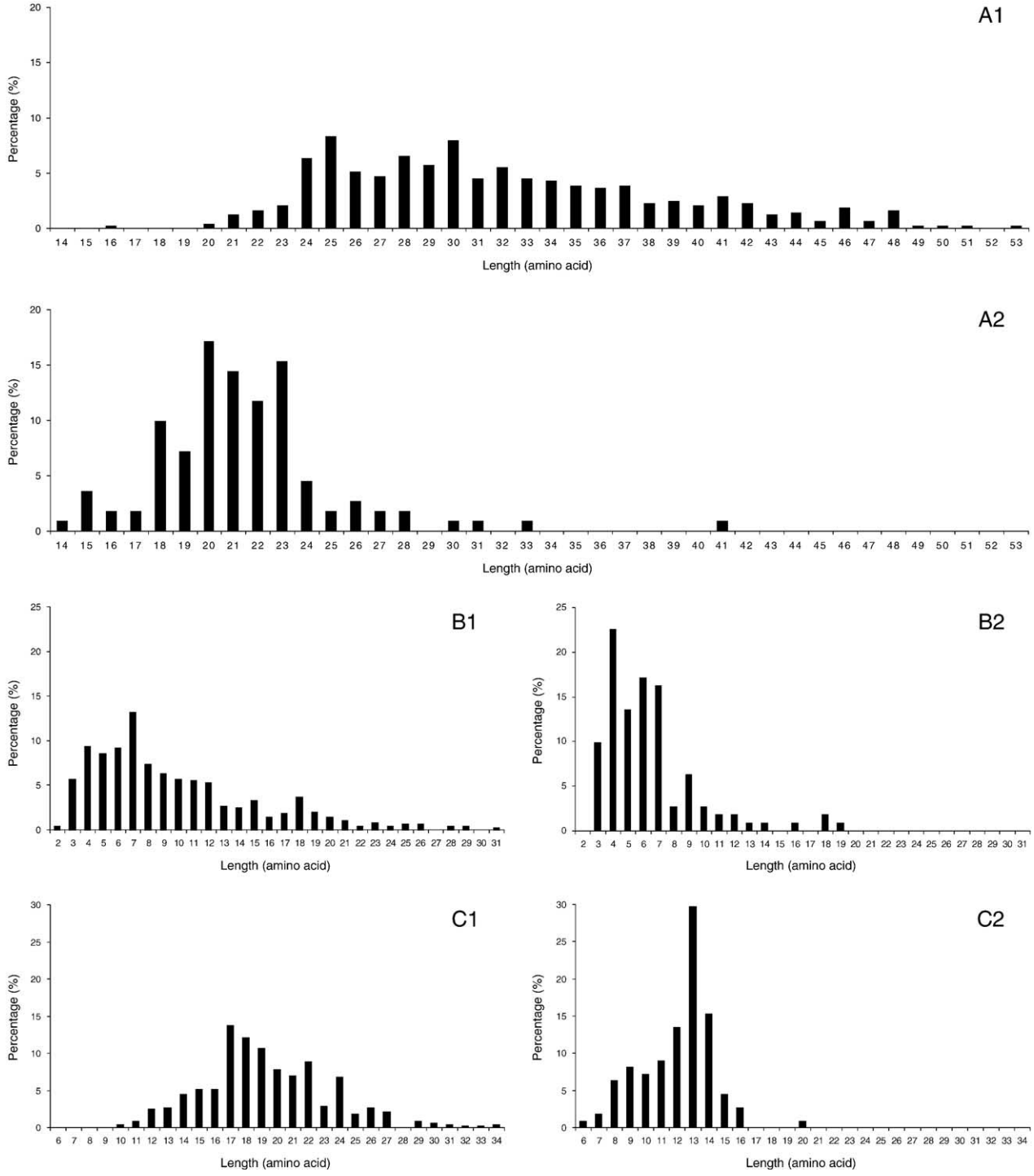

Fig. 1. Features of signal peptides of proteins predicted to be secreted via the Sec-dependent pathway. (A) Length distribution of the complete signal peptides. (B) Length distribution of the n-domains. (C) Length distribution of the h-domains. (1) Secretory signal peptides substrate of SPase I. (2) Lipoprotein signal peptides substrate of SPase II. Distributions are expressed as a percentage of the total number of predicted secretory or lipoprotein signal peptides.

this putative molecular complex on the clostridial cell surface. In ChW proteins, within the repetitive module, a glycine is systematically present before the conserved Trp [9]. Interestingly, in *L. monocytogenes*, a family of surface proteins containing a highly conserved Gly-Trp dipeptide motif, and hence called GW proteins, interact with lipoteichoic acids of the cell wall through these hydrophobic repetitive modules [95].

3.2.2.3. Cellulosome. It has previously been shown that *C. acetobutylicum* possesses a cellulosome which is an extracellular enzymatic complex of high molecular weight.

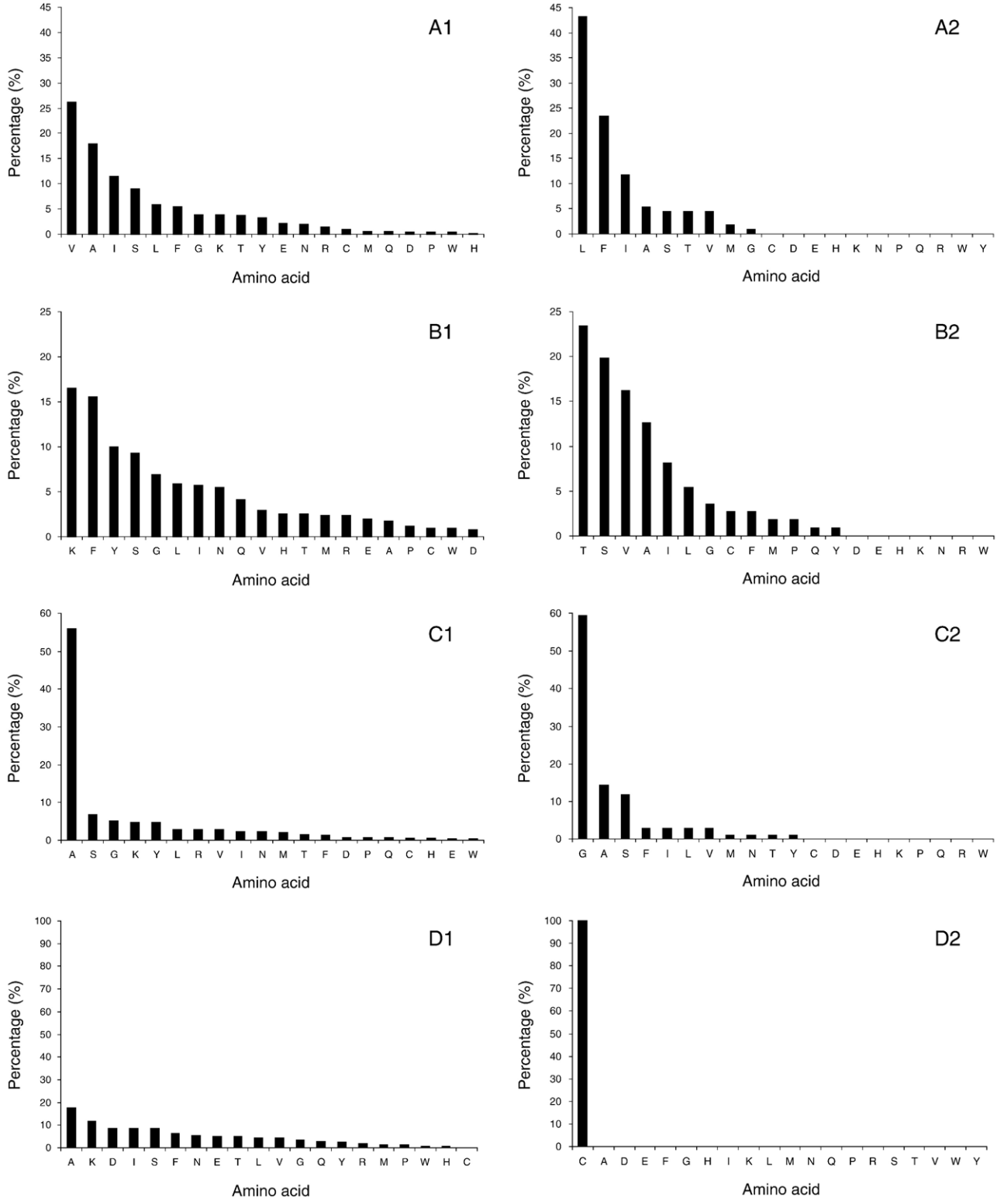

Fig. 2. Features of predicted Sec-dependent signal sequence c-domains. (A) Frequency of a.a. at position −3 of the cleavage site. (B) Frequency of a.a. at position −2 of the cleavage site. (C) Frequency of a.a. at position −1 of the cleavage site. (D) Frequency of a.a. at position +1 of the cleavage site. (1) Secretory signal peptides substrate of SPase I. (2) Lipoprotein signal peptides substrate of SPase II. Frequencies are expressed as the percentage of the total number of predicted secretory or lipoprotein signal peptides.

This complex, normally dedicated to efficient degradation of various plant cell wall materials, seems inactive in *C. acetobutylicum* [101]. Several cellulosomal components could be identified among the secreted proteins of *C. acetobutylicum* (Tables 2 and 4). Cellulosomes are organized around a specialized scaffolding protein, called CipA in

Table 3
C. acetobutylicum proteins containing ChW repeats

GI	Annotation
15894810	Protein containing ChW repeats
15895795	Protein containing ChW repeats
15895796	Protein containing ChW repeats
15895844	Protein containing ChW repeats
15893828	β-Mannanase ManB containing ChW repeats
15893829	β-Mannanase ManB containing ChW repeats
15893830	β-Mannanase ManB containing ChW repeats
15895558	Extracellular neutral metalloprotease, NPRE, fused to ChW repeats
15895592	Possible cell wall hydrolase containing *N*-acetylglucosaminidase domain and ChW repeats
15894668	Protein containing ChW repeats and cell-adhesion domain
15895634	Uncharacterized protein containing predicted cell adhesion domain and ChW repeats
15896523	Uncharacterized protein containing predicted cell adhesion domain and ChW repeats
15896517	Possible surface protein, responsible for cell interaction; contains cell adhesion domain and ChW repeats
15896518	Possible surface protein, responsible for cell interaction; contains cell adhesion domain and ChW repeats
15896519	Possible surface protein, responsible for cell interaction; contains cell adhesion domain and ChW repeats
15896520	Possible surface protein, responsible for cell interaction; contains cell adhesion domain and ChW repeats
15896524	Possible surface protein, responsible for cell interaction; contains cell adhesion domain and ChW repeats
15896525	Possible surface protein, responsible for cell interaction; contains cell adhesion domain and ChW repeats
15004707	Transglutaminase-like predicted protease domain fused to ChW repeats and cell-adhesion domain
15004708	Transglutaminase-like predicted protease domain fused to ChW repeats and cell-adhesion domain

C. acetobutylicum (GI: 15894197); different catalytic components bind to CipA by cohesin/dockerin domain recognition, which is the main characteristic of components belonging to a cellulosome [102]. The cohesin/dockerin interaction is calcium dependent and species specific [103,104]. CipA bears five cohesin domains, and ten proteins possessing a C-terminal dockerin domain could be identified in *C. acetobutylicum* (Table 4); a signal sequence could be predicted for all but one (GI: 15894199). With the exception of two (GI: 15893851 and 15896708), all the cellulosome components are present in the *cel* locus [105].

While it is clear that the proteinaceous components of the cellulosome are scaffolded together on CipA, the molecular mechanisms allowing extracellular assembly of the cellulosome and anchoring to the cell surface remain elusive. It has been speculated that the gene product of *orfXp*, which is present in the *C. acetobutylicum cel* cluster (GI: 15894201) and encodes a lipoprotein with a cohesin domain (Table 1), could be involved in this process [106]. It has recently been demonstrated in *C. cellulovorans* that EngE, a cellulosomal cellulase, anchors the cellulosome by integrating into the cell wall layer via its N-terminal SLH domain and by cohesin/dockerin interaction with its cognate cellulosomal scaffolding protein CbpA, a homologue of the *C. acetobutylicum* CipA [107,108]. A protein homologous to EngE and annotated as a cellulosomal endoglucanase was identified in *C. acetobutylicum* (GI: 15896708). Pfam and ScanProsite searches did not reveal the presence of SLH domains (i.e. PF00395 or PS01072) in other protein of *C. acetobutylicum*.

3.2.2.4. LPXTG proteins. In Gram-positive bacteria, a range of proteins are anchored to the bacterial cell surface by transpeptidase sortases, which cleave and further covalently link them to peptidoglycan via penta-glycine cross-bridges [109]. The protein substrates of this enzyme harbour a conserved LPXTG sortase-cleavage site followed by a transmembrane-spanning hydrophobic domain and by a hydrophilic charged domain at the C-terminus. In various clostridia, sortase-like proteins have been identified, but none of them have been characterized [110]. While in some bacterial genomes several genes coding for sortase-like proteins have been found, *C. acetobutylicum* possesses only one sortase (GI: 15893497). However, as reported in *Streptococcus pyogenes* and *S. aureus*, the presence of additional enzymes nonribosomally synthesized and called LPXTGases cannot be excluded [111]. Only 1 potential substrate, with its gene adjacent to the sortase gene, has been reported so far in *C. acetobutylicum* [110]. A search for additional potential substrates using iterative PSI-BLAST, SignalP, and Pfam searches identified three other putative sortase substrates in *C. acetobutylicum* (Table 5). They all possess a signal sequence of class 1 or 2 (Table 1), a plausible LPXTG-like motif, followed by a membrane-spanning hydrophobic domain and a charged C-terminal tail. For one of them, the conserved motif differs by a single residue, i.e., LPXSG (GI: 15896849); such variation of the LPXTG-like motif from the canonical sequence has already been reported in the literature [109,110,112]. The SpoIID-like protein (GI: 15895771) as a substrate for sortase is more dubious since (i) two LPXTG motifs are present in the C-terminal part of the protein, which both overlap a putative peptidoglycan binding domain (Pfam 01471, E-value$=3.0\times10^{-14}$), (ii) the transmembrane segment, present after the first LPXTG motif occurring from the C-terminus is rather short and exhibits a charged lysine residue (Table 5).

3.2.2.5. Ricin-B proteins. A putative system involved in xylan degradation was recently speculated to be present in *C. acetobutylicum* [9]. The 6 enzymes involved in this putative complex bear a conserved C-terminal ricin-B domain, also called CBM 13 (Carbohydrate-Binding Module Family 13), involved in carbohydrate binding (Table 6); four genes encoding a ricin domain are detected on pSol1, two of them are encoded on the *C. acetobutylicum* chromosome. Experimental evidence for the formation of a cell surface complex is not yet available.

Table 4
Cellulosome components in *C. acetobutylicum*

GI	Annotation	Dockerin domain[a]
		* * * * *##* * * * * *##*
15894197	Scaffolding protein, CipA	VKYGDINDDGVVNGRDIMVLTQYIAG KPVTIDTAAADVNGDGTVNGRDLMILRQYVSGKVTSFPVESNK
15893851	Cellulase, CelE orthologue	ATPGDVNGDGVVNGRDLMELRQYLAGKLDASKINLAAADVNNDGVVNGRDIMELTKLIAK
15894198	Processive endoglucanase family 48, Cel48F orthologue	VQKGDVNGDGAVNGRDVMELRKYISG QITTINKDVADINGDGVINGRDLVELIKMISNNQ
15894199	Non–processive endoglucanase family 5, Cel5A homologue	VVVGDINGDGEVNGRDLMELRKYIAG KTTDINKDSADINGDGVINGRDLMELIRKISSVNP
15894200	Non–processive endoglucanase family 9, Cel9G orthologue	SVKGDVNGDGVVNGRDLMELRKYLAG STSNIDLNAADLNNDGAVNGRDLLELIKLVFEK
15894202	Endoglucanase A, endo–1,4–beta–glucanase	VKTGDVNGDNTVDGRDLMELRKYLAG NNPNINLKAADLNGDGIVNGRDLVILIKLISGKQ
15894203	Non–processive endoglucanase family 9, Cel9G orthologue	TILGDLNDDGVVNGRDIVMMRQYLAG KTVSGIDKNAILINGDGAVNGRDLMELIKKVSNN
15894204	Cellulose–binding endoglucanase family 9; Cel9L orthologue	GLKGDVNGDGVINGRDLMVLRQYLAG QSVNINKANTDVNGDGVVNGRDLMEIVKIVMSK
15894205	Non–processive endoglucanase family 5, mannase A orthologue	VTLGDVNDDGIINGRDVMVLTQYIAG KDVTIDKKAADVNGDGTVNGRDLMTLRQYISGKITSFSSTK
15894206	Sialidase with several ASP–boxes	ALLGDVDGNGVVDSLDITLIRSYLAG KAVTINKTSADVNNDGRVSGADLMALMKKIAKGA
15896708	Endoglucanase family 5 with S–layer homologuey	

[a] Dockerin domains of the catalytic components of the cellulosome which bind to the cellulosome scaffolding protein CipA. The symbol * indicates residues involved in calcium binding, and # indicates residues involved in species-specific cohesin/dockerin interaction.

3.3. The flagella export apparatus (FEA; TC #3.A.6.)

The bacterial flagellum can be subdivided into 5 major parts: (i) the flagella motor/switch, which operates and controls a rotary mechanism; (ii) the basal body, a passive structure which receives torque from the motor and transmits it to the hook and then to the filament; (iii) the hook and junction proteins, which function as a universal joint; (iv) the flagellar filament and its distal cap, which function like a propeller; and (v) the flagella export apparatus (FEA), which permits the translocation of some flagellar components prior to their assembly [113,114]. Based on studies of Gram-negative bacteria, the FEA is composed of the transmembrane components FlhA, FlhB, FliO, FliP, FliQ, and FliR, the chaperones FliJ, FlgN, FliS, and FliT, the ATPase FliI, and its regulator FliH [113–115]. FliI would permit the coupling of ATP hydrolysis to protein export and seems negatively regulated by FliH [116]. Except for FlhB, which seems to gate the export pathway and determines the substrates transported, the function of the remaining transmembrane proteins in the transport machinery complex is still unclear [113,114]. In Gram-negative bacteria, the FEA is related to the Type III secretion system [30,117–119]. As already pointed out, because of the presence of an outer membrane which necessitates additional protein structure(s) to complete the secretion, the Type III system terminology is restricted to Gram-negative bacteria. In contrast to Gram-negative bacteria [119–121], in Gram-positive bacteria, the export of proteins not involved in the flagellum morphogenesis has been reported only in *Bacillus thuringiensis* [122]. Proteins secreted through the FEA possess no cleavable signal peptide. From the study of the Type III secretion system in Gram-negative bacteria, conflicting evidence exists about (i) secondary or tertiary protein structure recognition, (ii) the involvement of general and/or specific chaperones targeting the protein prior to export, and (iii) cotranslational export involving mRNA recognition [113]. As a consequence, no conserved recognition signal for flagellar protein export has been reported yet (Table 6).

Bacterial motility has been experimentally demonstrated in *C. acetobutylicum* [123–125], and genes involved in flagellar assembly are essentially present in two clusters on *C. acetobutylicum* chromosome [9,126]. Compared to Gram-negative bacteria, most of the FEA components could be identified in *C. acetobutylicum* (Table 1). As in *B. subtilis*, FliO also appears to be missing from the *C. acetobutylicum* FEA [90]. Interestingly, as well as a FlhB homologue (GI: 15893927), an FlhB–FliR fusion protein (GI: 15895417) is also present, but no FliR homologue could be found. While the general chaperone FliJ (GI: 15895427) and the FliC-specific chaperone FliS (GI: 15895474) could be identified (Table 1), neither homologues to the FliD-specific chaperone, FliT, nor the FlgK- and FlgL-specific chaperone, FlgN, could be found in *C. acetobutylicum*. In *B. subtilis*, FlgN is also absent, but a

Table 5
Substrates of the sortase in *C. acetobutylicum*

GI	Function	Cleavage site	C-terminal domain	
			Transmembrane segment	Charged tail
15893498	Phosphohydrolase, Icc family	**LPKTG**E	FFDATMLLSIALICLASGAILIFV	NKKKSSPTK
15893644	2,3-Cyclic-nucleotide 2′phosphodiesterase	**LPKTG**S	MIDSTVLLIIGTLLLLLGLAFIIW	NKFKNKQKSVQ
15895771	SpoIID-like protein	**LPVTG**N	VDAATLKVINNMLI	NKPETKNLIF
15896849	Membrane protein	**LPKSG**G	ILINLWSLFIVFVIILFFTFLIF	KNILRILKKYQPIELIRGA

FliT homologue had been identified [90]. By analogy with proteins known to be secreted through the FEA in Gram-negative bacteria [113,127], 16 proteins all involved in flagellar morphogenesis could be identified as substrates of this system in *C. acetobutylicum* (Table 7).

3.4. The fimbrilin protein exporter (FPE; TC #3.A.14)

In *B. subtilis*, the components of the FPE systems are encoded by the *comG* locus, which consists of seven CDS (*comGA-GG*), and *comC*, which is located elsewhere on the chromosome. Because proteins encoded by the *comG* operon and *comC* of Gram-positive bacteria resemble proteins found in Type II secretion system (TTSS), the Type 4 pilus (Tfp) assembly apparatus, and Type IV secretion system (TFSS) of Gram-negative bacteria, they have also been collectively called PSTC (pilus/secretion/twitching motility/competence) [128,129]; the mechanism involved has been suggested to be ancient and predating the divergence of Gram-positive and Gram-negative bacteria. In *B. subtilis*, which is the only Gram-positive bacterium where the FPE has been experimentally investigated, the FPE is part of the Com (*com*petence development) pathway, which allows the internalization of exogenous DNA [17]. The Com pathway involves (i) the FPE, which may form a channel across the cell wall, permitting incoming DNA to access (ii) the bacterial competence-related DNA transformation transporter (TC #3.A.11), which permits bacterial DNA uptake across the cytoplasmic membrane [67,128,130]. In the FPE, the ATPase ComGA is homologous to PulE of the TTSS, PilB of the Tfp assembly apparatus, and VirB11 of the TFSS in Gram-negative bacteria [129]. The integral membrane protein ComGB is homologous to PulF of the TTSS and PilC of the TFP [129]. ComGF is required for the binding of DNA to the cell surface but has no known orthologue [128]. The prepilins ComGC, ComGD, ComGE, and ComGG exhibit similarities to Type 4 prepilins [131]. ComC is a prepilin-specific SPase, which cleaves the prepilins between the n- and h-domains of the signal peptide [17]; those signal sequences belong to the class 3.

TC-DB searches of the *C. acetobutylicum* genome revealed that a protein, initially annotated as a general secretion pathway ATPase (GI: 15895375), is homologous to ComGA, VirB11, PulE, and PilB (Table 1; see also Supplementary material in Appendix A) [129,132]. A domain characteristic of the ATPase components of the TTSS and TFSS could also be identified (PF00437; *E*-value $=5.2\times10^{-101}$). Next to its gene, TC-DB searches revealed the presence of a second CDS (GI: 15895374) coding for a protein homologous to ComGB, PulF, and PilC (Table 1; see also Supplementary material in Appendix A)

Table 6
C. acetobutylicum proteins encoded on pSol1 and bearing C-terminal ricin-B domain

GI	Annotation
15004775	Possible xylan degradation enzyme (alpha/beta hydrolase domain and ricin-B-like domain)
15004821	Possible xylan degradation enzyme (glycosyl hydrolase family 30-like domain and ricin-B-like domain)
15004822	Possible xylan degradation enzyme (glycosyl hydrolase family 30-like domain and ricin-B-like domain)
15004823	Possible xylan degradation enzyme (glycosyl hydrolase family 43-like domain, cellulose-binding domain and ricin-B-like domain)
15893994	Endo-1,4-beta glucanase (fused to two ricin-B-like domains)
15893866	Endo-arabinase related enzyme (family 43 glycosyl hydrolase domain and ricin-B-like domain)

Table 7
C. acetobutylicum proteins secreted through the flagellar export apparatus

Homologue	GI	Function
Rod		
FliE	15895432	Integral membrane ring (MS ring) junction protein
FlgB	15895434	Basal-body rod protein, transmission shaft
FlgC	15895433	Basal-body rod protein, transmission shaft
FlgF	15895409	Basal-body rod protein, transmission shaft
FlgG	15895408	Distal rod protein, transmission shaft
Hook		
FlgD	15895422	Hook capping protein, off during hook/filament transition
FlgE	15895423	Hook protein
FliK	15895426	Hook length control protein
FlgK	15895480	First hook filament junction protein
FlgL	15895471	Second hook filament junction protein
	15895424	Flagellar hook associated protein
Filament		
FliD	15895473	Filament capping protein
FliC	15894833	Flagellin
	15895436	Flagellin
	15895476	Flagellin
	15895479	Flagellin
	15895413	Flagellin
FlgM	15895482	Antiflagellar σ factor

and predicted to possess three membrane spanning segments by TMpred. Looking for a ComC-like protein, a Type 4 signal peptidase (GI: 15896039) was located elsewhere on the chromosome (Table 1; see also Supplementary material in Appendix A). No homologue to ComGF could be identified. When searching for substrates of the FPE, four proteins predicted to possess Type 4 prepilin signal peptides (GI: 15895370-15895373) were encoded by CDS located downstream of *comGA* and *comGB* (Table 8). An additional putative Type 4 prepilin (15894276) is encoded elsewhere on the chromosome (Table 8). These proteins were originally annotated as prepilin peptidases, hypothetical proteins, or as belonging to general secretory pathway; PSI-BLAST searches indicated that they were all homologous to fimbrial proteins/pilin precursors (*E*-values $\leq 4.0\times10^{-24}$ after 2 iterations). As in *B. subtilis*, the coding regions of the first four CDS present downstream of *comGB* overlap [67]; it was suggested that such an arrangement allows the translational coupling of the CDS involved to ensure that the synthesis of their cognate proteins is closely coordinated. Therefore, it seems that a complete set of proteins related to FPE is present in *C. acetobutylicum* (Table 1).

3.5. The tight adherence (Tad) export apparatus

The Tad system, encoded by the *tadZABCDEFG* locus, is a newly characterized secretion pathway firstly described in *Actinobacillus actinomycetemcomitans* and that allows the secretion and assembly of Flp pili, which mediate the tight adhesion of bacteria to surfaces and are essential for colonization and pathogenesis [133,134]. From a genomic survey, it appears that *tad*-related loci are widespread among bacteria and have undergone extensive gene shuffling and horizontal gene transfer [135]. To date, a Tad locus has been reported in four Gram-positive bacteria, namely *Corynebacterium diphteriae*, *M. tuberculosis*, *M. bovis*, and *Streptomyces coelicolor* [135]. As the Flp pili do not have to cross an outer membrane in Gram-positive bacteria, the genes necessary for pilin secretion across the outer membrane are absent. Thus, in Gram-positive bacteria, the *tad* locus is shorter than in Gram-negative bacteria and appears only to encode TadZ, TadC, TadB, and TadA [135]. TadA is an ATPase localized at the periphery of the cytoplasmic membrane [136]. TadB and TadC are integral membrane proteins that may form homo- or hetero-oligomeric structures in association with TadA and thus constitute the secretion apparatus [129]. No function has yet been attributed to TadZ, but it appears to be an essential component of the Tad system, with no homologue in any other known bacterial protein secretion system [135,137]. The Flp prepilins secreted through this system share homology with other Type 4 prepilins [138,139]. However, they belong to a distinct subfamily and are distinguished by (i) a relative small size(<90 a.a.), (ii) a shorter carboxy-terminal domain, and (iii) an invariant tyrosine residue immediately following the conserved glutamate conserved tyrosine residue [138,139]. *flp*-like genes are closely linked to almost every *tad* locus identified so far.

In. *C. acetobutylicum*, Pfam searches identified a protein (GI: 15895251) with an ATPase domain (PF00437; *E*-value$=1.4\times10^{-9}$) which was previously predicted to be involved in pili biogenesis. A CDD search of this protein revealed a domain (COG4962; *E*-value=3.0×10^{-88}) related to the Flp pilus assembly protein ATPase CpaF (GI: 15155107), a TadA homologue from *Agrobacterium tumefaciens* [135,137]; PSI-BLAST searches clearly indicated that this protein is homologous to TadA of *A. actinomycetemcomitans* (GI: 32452626; *E*-value$=2.0\times10^{-23}$). PSI-BLAST searches also revealed that two proteins (GI: 15895250 and 15895249) encoded downstream of this gene were homologous to TadB and TadC, respectively (after 1 iteration, *E*-value$=3.0\times10^{-88}$ and 4.0×10^{-53}, respectively); CDD searches identified that 15895249 possesses a COG2064 domain related to the Flp pilus assembly protein TadC (*E*-value$=3.0\times10^{-10}$). PSI-BLAST searches revealed the CDS located upstream of *tadA* (GI: 15895252) is homologous to CpaE from *Vibrio vulnificus* (GI: 37680838; *E*-value$=2.0\times10^{-88}$ after 3 iterations), i.e. a TadZ homologue [135,137]. Therefore, in *C. acetobutylicum*, a *tad* operon seems to be present (Table 1) [135]. However, taking into account the features of Flp prepilin, searches at a.a. and nucleotide levels in proximity of the *tad* operon, as well as within the whole genome, did not reveal the presence of homologue(s) to Flp prepilin(s). No prepilin other than the ones already described for the FPE system could be identified (Table 8).

Table 8
Predicted signal sequences substrates of the Type 4 prepilin signal peptidase in *C. acetobutylicum*

GI	Annotation	Signal peptide[a]
15895370	Hypothetical protein	**MQFYKVKKG** FTLIEMVAASAIFCVFTVFAISVLFSLINGYKKDMKSNNDQAN
15895371	Hypothetical protein	**MWKLRNKKG** FTLIEVMCSFSIFSILFLFAVNLKVDELKMGKINDDIQNYTYY
15895372	Prepilin peptidase	**MEEELLMTSRQSKKG** YILIELLCTIAIMLILCSVIAISFKSYKDIKNGIEVKYVNNEM
15895373	General secretion pathway protein, pilin family	**MMKKRG** FTLIELIISMSIIAILGAILVPNIYSYIRRANNEKAKDMAALV
15894276	General secretion family related protein	**MKNRINNKG** FTLIELIIVIAILAILAAILVPSISAYKIKAEKSNIQASARTL

[a] The positively charged n-domain is indicated in bold letters and the hydrophobic h-domain is shaded grey. The recognition sequence by the prepilin peptidase is underlined. Predicted cleavage site is indicated with a gap in the a.a. sequence.

3.6. The holins (TC #1.E.)

Holins are small membrane proteins whose function is mainly associated with the collapse of the membrane potential and permeabilization of the membrane [140]. Concomitant to the holin-mediated permeabilization event, a protein is secreted across the cytoplasmic membrane. Holins are also involved in the secretion/activation of enzymes with muralytic activities, which hydrolyze the cell wall polymer as a prelude to cell lysis [141]. The primary function of holins then appears to be the transport of specific lytic enzymes lacking N-terminal signal sequences. It is important to emphasize that the structure of holins is still uncertain, but they are believed to be present as homo-oligomeric complexes that form pores through the cytoplasmic membrane, which provide a passive, energy-independent transport function [140]. Twenty-one distinct families of holin proteins are currently recognized in TC-DB. In *C. acetobutylicum*, three holin-like proteins with no detectable N-terminal signal peptide were found (Table 1).

The first holin is a protein of 117 a.a., lacking putative signal sequence, and previously annotated as a predicted membrane protein of the YohJ family (GI: 15893901). However, it shows 32% identity and 60% similarity to LrgA, a holin originally identified in *S. aureus* (GI: 1575025), and harbours the LrgA family domain (PF03788; *E*-value = 1.3×10^{-33}), thus it should be classified as TC #1.E.14 (Table 1). As with LrgA, and as predicted by TMpred, this homologue possesses 4 putative TMSs. This protein family acts as murein hydrolase exporters involved in programmed cell death, a process that is analogous to apoptosis in eukaryotes [141]. As in *S. aureus*, this LrgA homologue clusters with its potential substrate, originally annotated as a predicted membrane protein of the YohK family (GI: 15893900), a potential membrane associated protein homologous to LrgB (GI: 1575026) with 28% identity and 47% similarity [142]. In *S. aureus*, an additional operon, designated *cidAB*, where the gene products CidA and CidB share 23% and 32% identity with LrgA and LrgB, respectively, functions in a way dramatically opposed to the *lrgAB* operon [143]. While *lrgAB* genes confer negative control on extracellular murein hydrolase activity, *cidAB* genes activate autolysis. Therefore, CidA would act as a holin while LrgA would act as an antiholin. In *C. acetobutylicum*, homologues of CidA and/or CidB could not be identified on the chromosome. However, a second putative holin encoded on the megaplasmid pSol1 and originally annotated as a membrane protein (GI: 15004877) is homologous to LrgA (41% identity and 55% similarity), but this gene is not part of an operon with a *lrgB* or *cidB* homologue (Table 1). PSI-BLAST searches revealed the presence of three probable autolysins lacking putative signal sequences which could be substrate for these holins (GI: 15893844, 15895879, and 15895935). One of them, annotated as an autolytic lysozyme (GI: 15893844), has already been reported as being involved in *C. acetobutylicum* autolysis [144].

From a Pfam search, a protein originally annotated as an uncharacterized protein, *B. subtilis* YtkC orthologue, and related to the regulatory protein UtxA (GI: 15895117) possesses a holin domain related to TcdE (PF05105; *E*-value = 1.7×10^{-49}) (Table 1). In *C. difficile*, *tcdE* is located between the genes *tcdA* and *tcdB*, which encode two large toxins [145]. The holin TcdE then permits the release of TcdA and TcdB into the extracellular medium. In *C. acetobutylicum*, no genes coding for toxins could be found in proximity to the gene coding for the TcdE-like protein. However, PSI-BLAST, TC-DB, and Pfam searches revealed the presence of a gene encoding a protein lacking a potential signal sequence (GI: 15893630) and possessing a clostridial binary toxin A domain (PF03496; *E*-value = 2.1×10^{-74}), and exhibiting 27% identity and 42% similarity with the actin-specific ADP-ribosylating binary toxin genes CtdA from *C. difficile* (GI: 8926247). This constitutes the first report of such an exotoxin in *C. acetobutylicum*.

3.7. The ESAT-6/WXG100 secretion system

ESAT-6 family genes encode small proteins which were originally identified in *M. tuberculosis* [146]. These proteins are potent T-cell antigens of fundamental importance in bacterial virulence [147]. ESAT-6 proteins are secreted into the extracellular milieu even though a typical Sec-dependent signal sequence is lacking. A genome sequence survey has permitted the identification of distant homologues of ESAT-6 in both high and low G+C Gram-positive bacteria [56]. The secretion of ESAT-6-like proteins in low G+C Gram-positive bacteria was recently demonstrated in *S. aureus* [57]. In *C. acetobutylicum*, nine ESAT-6 proteins were originally reported [9,41]. Using BLAST, PSI-BLAST, and Pfam searches, one other putative ESAT-6 protein was identified in *C. acetobutylicum* (Fig. 3). Altogether, ten potential ESAT-6/WXG100 proteins were identified, which all possess a conserved WXG motif, with a protein length of approximately 100 residues, and with no signal sequence identified using SignalP, PSORT, or searching for lipoproteins. No N-terminal signal peptide could be detected. While no function could be attributed to the ESAT-6/WXG100 proteins, they all share the Pfam domain PF06013 referring to proteins of 100 residues with WXG motif and could be hypothesised to play a role in bacterial virulence [57,148]. The corresponding genes are present in four clusters, three of which also contain YukA homologues (Table 1).

From a genomic survey, it was revealed that ESAT-6 genes cluster systematically with large membrane-bound ATPases similar to YukA from *B. subtilis* [41]. In *M. tuberculosis* and *S. aureus*, it was demonstrated that YukA-like proteins are involved in the translocation of ESAT-6-like proteins and, thus, would form a novel Gram-positive secretion system [23,42,57]. No TC number has

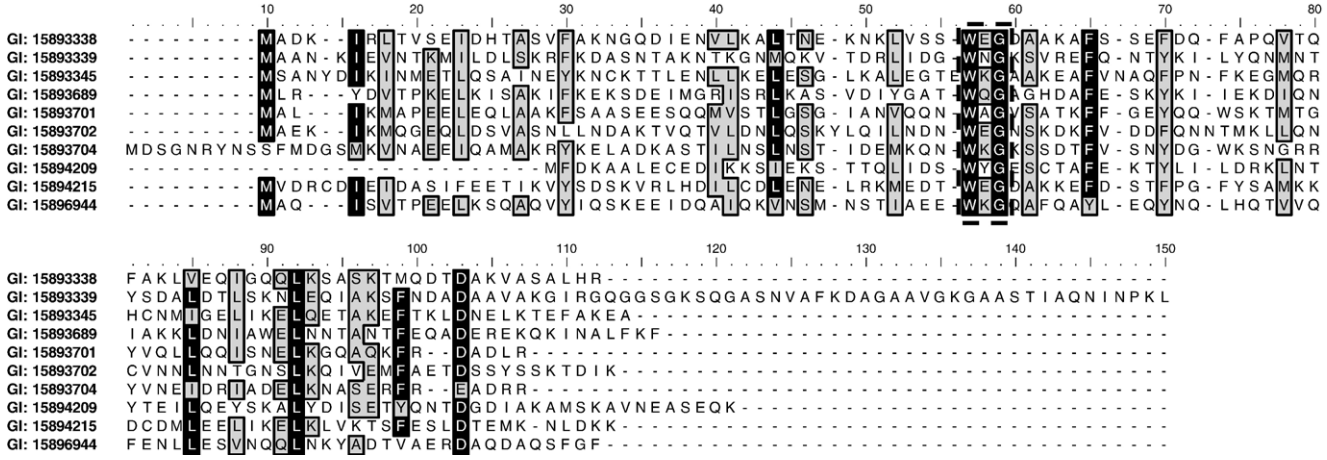

Fig. 3. Sequence alignment of ESAT-6/WXG100 proteins present in *C. acetobutylicum*. The WXG motif is squared. Sequence alignments were performed using ClustalW, with minor manual refinement using BioEdit.

yet been attributed to this novel transport system. At the C-terminus of all YukA homologues, an AAA-ATPase domain (IPR003593),which is also present in Gram-negative Type II and Type IV secretion systems, is found and has been suggested to supply energy for ESAT-6/WXG100 protein secretion [41]. Interestingly, in YukA-like proteins from *C. acetobutylicum*, a forkhead-associated (FHA) domain (PF00498; *E*-value $\leq 5.8\times10^{-9}$) normally involved in phosphopeptide recognition is also present at the N-proximal end. Generally, the presence of such a domain strongly suggests interaction with a protein partner in a process regulated by reversible phosphorylation [149,150]. Signalling interactions between proteins encoded within the WXG100 clusters of *C. acetobutylicum* have been speculated [151].

4. Discussion

Among low G+C Gram-positive bacteria, protein secretion in *C. acetobutylicum* could clearly be investigated as an alternative model to *B. subtilis*. One of the most striking differences is the apparent absence of a Tat pathway. It should be noted, however, that until experimental evidence becomes available, the absence of homologues in any bacterial genome does not necessarily indicate that functional analogues of such proteins are absent; this requires more experimental investigation to be confirmed. At least five protein secretion pathways alternative to the Sec-dependent pathway seem to be present for the transport of specific proteins in *C. acetobutylicum*. This computational analysis is a prelude to in vivo study of the proteins secreted by *C. acetobutylicum*; such proteomic investigations of protein secretion in *C. acetobutylicum* may also reveal the presence of secreted proteins that could not be predicted by genome sequence analyses [152–154]. Alternatively, the number of truly secreted proteins, i.e., in the extracellular milieu, could be overestimated by in silico approaches, especially in clostridia, because of the lack of a reliable data set for the refinement of signal peptide predictions. For example, several proteins predicted to possess cleavable signal peptides such as SPases, SecD, or SecF are probably not cleaved. This also stresses the need for further experimental work to validate or invalidate such bioinformatic analyses. Studies published so far on the clostridial secretome appear to indicate a high discrepancy in the number of secreted proteins in various clostridial species, from 15 in *C. difficile* to over 200 in *C. perfringens* [155,156]. From this bioinformatic investigation, we could expect an even higher number of proteins to be secreted by *C. acetobutylicum*.

Surprisingly, from this study, a number of potential virulence factors secreted either via the Sec-translocon, the holins, or the ESAT-6/WXG100 secretion system have been reported. The presence of such virulence factors has never been emphasized by previous investigations and is quite unexpected for a saprophytic environmental bacterium commonly considered as non-pathogenic to animals or plants [9,11]. This aspect of *C. acetobutylicum* physiology, which has never been investigated previously, undoubtedly requires further in-depth investigations to assay the expression of pili and virulence factors, as well as their functions and activities. The lack of investigations in this direction in *C. acetobutylicum* certainly results from its highly recognized status as a biotechnological microorganism [1,2].

With 602 proteins identified as potential substrates of the Sec translocon, this apparatus appears to be the primary secretion pathway used in *C. acetobutylicum* (Tables 1 and 2). The majority of these proteins have no clear identified function (44%). The second largest group of secreted proteins is dedicated to carbohydrate metabolism (15%), which is highly relevant to the lifestyle of this soil microorganism. Each of the other categories of secreted proteins represents less than 9% of the total number of proteins secreted via the Sec translocon in *C. acetobutylicum*. The number of secreted proteins related to

bacterial virulence has also been identified, including phospholipase C, virulence factor MviN, hemolysins, and adhesins (Table 2). Compared to *B. subtilis* [20,153], in *C. acetobutylicum*, the number of proteins secreted via the Sec translocon for which no function has been attributed is high. Alone, this stresses the necessity to carry out proteomic analysis of this microorganism in order to confirm the secretion of these proteins and to further characterize their function and regulation [157]. With the exception of SecB, all components of the Sec system could be identified in *C. acetobutylicum*. In *S. pyogenes*, it was recently shown that protein secretion through the Sec pathway occurs at a distinct microdomain of the cytoplasmic membrane dedicated to protein export, i.e. the exportal [158]; it was suggested that such a subcellular organization may represent a paradigm in Gram-positive bacteria. The absence of SecB seems a general feature of Gram-positive bacteria [17,20]. Interestingly, in *E. coli*, when proline or glycine is present in the middle of the α-helical h-domain of signal sequences of Sec-dependent secreted proteins, these residues would promote the formation of a hairpin-like structure that would permit the complete insertion of the signal peptide into the cytoplasmic membrane [20]. The presence of these helix-breaking residues in the h-domain of signal sequences also appears to be an important parameter discriminating between SRP-targeting or SecB-targeting pathways [159]. From a recent study in *E. coli*, it appeared that some secreted proteins could be translocated via the Sec translocon in an SRP- and SecB-independent manner, which would involve either the requirement of other unidentified chaperone(s), targeting factor(s), or no accessory factor at all [87].

Concerning protein display on the bacterial cell surface, one of the most unexpected findings from genome analysis of non-cellulolytic *C. acetobutylicum* was the presence of a cluster of cellulosomal genes closely related to *C. cellulolyticum* [9,160,161]. While these genes were originally described as cryptic, i.e. possessing frameshifts and/or disabled promoters [162], more recent investigation revealed that a >665 kDa cellulosome was expressed by *C. acetobutylicum* but was inactive against crystalline cellulose [105].

Altogether, four potential protein substrates for sortase were identified. In comparison to other bacterial genomes, the number of sortase-like proteins and LPXTG-proteins is quite low in *C. acetobutylicum*. Interestingly, in Gram-positive bacteria, most of the surface proteins anchored by a sortase mechanism have a role in virulence, and the number of sortase substrates tend to be higher in pathogenic bacteria [110,163]. Two other speculative cell surface complexes have been proposed in *C. acetobutylicum*, i.e., ricin-B and ChW protein-based complexes [9]. Because of the presence of a highly conserved Gly-Trp dipeptide motif [9], it is tempting to speculate that ChW proteins could be GW-related proteins. However, while repeats similar to ChW proteins have been detected only in proteins from *S. coelicolor* [9], the presence of GW modules have been reported only in proteins from *Listeria* and *Staphylococcus* spp. [95]. In *S. pneumoniae*, the choline-binding proteins LytA or PspA contain short repetitive modules with conserved GW peptides, which allow binding to lipoteichoic acids, but due to weak similarity are not currently considered to be GW proteins [95]. Additionally, GW domains of *L. monocytogenes* have been shown to be structurally and evolutionary related to SH3b domains, present in p60 for example, but are considered unlikely to be functionally similar [164]. Also, further investigations are necessary to (i) experimentally assay the binding of ChW proteins to lipoteichoic acids, and (ii) clarify the phylogenetic relationships between proteins containing repetitive GW modules involved in binding to lipoteichoic acids.

In Gram-negative bacteria, pili are secreted and assembled by different pathways, namely, (i) the Tfp assembly apparatus, (ii) the Type III protein secretion pathway for Hrp pili, (iii) the TFSS for the T and F pili, (iv) the chaperone/usher pathway for the Type 1 and P pili, and (v) the recently discovered Tad system for the Flp pili [135,165]. It should be stressed that in Gram-positive bacteria, the denominations of the Types II, III, and IV and chaperone/usher secretion pathways are irrelevant. To date, assembly mechanisms involving (i) the Sec-translocon and sortases [166] and (ii) the Tad system [135] seem the only ones potentially involved in pili formation in Gram-positive bacteria. Since among the sortase-substrates identified in *C. acetobutylicum*, none share similarity with pilin proteins [166], only the Tad system may be involved in pili formation in this microorganism. The fact that *C. acetobutylicum* is phylogenetically distant from γ- and α-*proteobacteria*, as well as *Actinobacteria*, which are the only bacterial classes where the *tad* loci have been so far reported, could explain the failure to identify Flp pilins in this microorganism. In *B. subtilis*, despite being related to the Tfp, the FPE is not involved in the formation of pili per se [167,168]. Besides, the involvement of FPE in the formation of proper pili in other Gram-positive bacterial species has never been investigated. Interestingly, in *C. acetobutylicum*, a protein annotated as a PilT ATPase, involved in pili biogenesis (GI: 15894967), shares 46% (over 158 out of 338 a.a.: 158/338) identity and 64% (222/338) similarity with PilT from *Pseudomonas aeruginosa* (GI: 77679). In Gram-negative bacteria, besides being involved in the twitching motility of Tfp [169], PilT has been demonstrated as essential to the DNA uptake phase of competence for natural transformation [170]. The involvement of this PilT homologue in the transformation competence development of *C. acetobutylicum* requires further investigation. Importantly, while the presence and the role of fimbriae have been reported to play a part in the virulence of some pathogenic clostridia [171], these aspects have never been investigated in *C. acetobutylicum*. Similarly, flagellar assembly has not been thoroughly investigated in Gram-positive bacteria, including clostridia [172],

and contrary to Gram-negative bacteria, secretion through this apparatus of proteins different from those involved in flagellar assembly has been reported only once [122].

Such an in silico approach should encourage in vivo investigations of protein secretion in *Clostridia* as well as in other Gram-positive bacteria; this should reveal further information and permit the use of organisms other than *B. subtilis* as models of Gram-positive bacterial protein secretion. Compared to pathogenic Gram-positive bacteria, especially within the genus *Clostridium*, *Corynebacterium*, or *Mycobacterium*, the investigation of protein secretion mechanisms, such as Tad, holins, and especially the speculative ESAT-6/WXG100 system, would be facilitated in this more tractable bacterium for which a large number of molecular biology tools are available [173,174]. Moreover, the understanding of protein secretion systems in *C. acetobutylicum*, which has been long used in industrial bioprocesses, could give rise to some interesting biotechnological and/or biomedical applications [1,2].

In industry, *C. acetobutylicum* has essentially been used for acetone–butanol production, but it also produces other metabolites of high industrial interest, notably ethanol, acetate, and hydrogen. Worldwide, several research teams currently focus on the production of these high-value products from inexpensive renewable resources, such as lignocellulosic biomass. While the metabolism of this bacterium has been thoroughly investigated at both biochemical and molecular levels, few investigations have been dedicated to protein secretion. Knowledge of prevalent protein secretion systems and signal sequence features, such as the length and/or preferred a.a. at certain positions, is of great value for any attempt to efficiently express heterologous proteins in *C. acetobutylicum*. Protein display on the bacterial cell surface or secretion into the extracellular medium is of key importance in the capacity of a bacterium to interact and adapt to its surroundings, either by sensing changes in environmental conditions, harbouring enzymatic activities, mediators of motility, adhesion factors, or virulence factors. Since the capacity of *C. acetobutylicum* to degrade biomass results primarily from its capacity to secrete degradative enzymes, knowledge of protein secretion should not be underestimated for such attempts. In this respect, the cellulosome is of great interest because of the synergistic effects of the cellulosomal enzymes [160,175]. Indeed, it may be possible to engineer cellulosomes containing defined, heterologous, or engineered proteins which could improve plant cell wall degradation. From such cellulosomes, artificial metabolic pathways could also be developed combining various catalytic activities; this opens the way to a wide range of biotechnological applications. Still, the mechanisms of surface display and assembly of the different cellulosome components have not been addressed yet in the literature and would require further research to reach such a goal. In this respect, investigations of other macromolecular structures potentially present on the bacterial cell surface, such as ricin-B or ChW protein-based complex, should be carried out. From an industrial point of view, it should also be mentioned that control of bacterial lysis is of key interest especially in processes involving the use of bioreactors. This necessitates a better understanding of secretion systems, i.e., Sec and holins secretion pathways, and the regulation mechanisms involved in cell death. It should be stressed that lytic transglucosylases are involved in the autolysis of bacteria but also in the assembly of macromolecular transport systems such as pili [176].

From a biomedical point of view, several clostridial hydrolytic enzymes and toxins of interest have already been discussed in the literature, such as collagenase, which could be used as an anchoring unit in engineered proteins, or botulinum toxin, which is already used in the treatment of a myriad human neuromuscular disorders [177–179]. A very remarkable feature of clostridia relies on the capacity of their spores to specifically and selectively target tumours [180,181]. The predilection of clostridial spores to germinate in hypoxic/necrotic regions of tumours is likely due to the poor vascularisation of solid tumours, which, in turn, creates favorable anaerobic conditions for the growth of clostridia, causing lysis and degradation of necrotic tissue [182]. Importantly, contrary to live bacteria that could be toxic when injected intravenously, bacterial spores are nontoxigenic to normal animals. However, the sole injection of clostridial spores from various species has had limited clinical success. Alternative strategies have been recently developed, namely (i) combination bacteriolytic therapy (COBALT), where clostridial spores are administrated together with conventional chemotherapeutic drugs [183], and (ii) clostridial-directed enzyme prodrug therapy (CDEPT), where clostridia are used as drug delivery systems [180]. One pledging approach in CDEPT is the antibody-directed enzyme prodrug therapy (ADEPT), where an enzyme is linked either chemically or genetically to tumour-targeting antibody [181]. A combination of these different strategies should bring new hope for future cancer treatment. The use of *C. acetobutylicum* in oncolysis and as a tumour-delivery system also seems a promising approach [184–186]. The advantages of *C. acetobutylicum* over other clostridial species are that (i) neither the live bacteria nor clostridial spores are toxigenic, (ii) it is the only non-pathogenic *Clostridium* sequenced so far, (iii) a large range of molecular biology tools are readily available to genetically manipulate this bacterium, and (iv) this bacterium expresses a cellulosome which could be engineered to express artificial metabolic pathways of interest in cancer treatment. While the aspects of alternative therapeutic strategies have been investigated, little is known about the mechanisms permitting the specific targeting of tumours by *Clostridia* [187]. In fact, it can be legitimately hypothesised that motility, adhesion factors, and/or secreted enzymes play an important role in the colonization and digestion of the tumours. Also, in order to develop efficient anti-cancer therapies based on the use of clostridia, knowledge of protein secretion systems, effector molecules, and regulatory

mechanisms involved during oncolysis is certainly as important as the development of new drugs and/or delivery systems. An intriguing observation is the presence of proteins possessing cell adhesion domains together with the presence of a Tad system in *C. acetobutylicum*. These could play essential roles in colonization and maybe pathogenesis [134].

While this investigation provides the first overview of protein secretion systems present in a *Clostridium*, it also raises a number of questions which would undoubtedly necessitate further investigations: How many proteins are really secreted by *C. acetobutylicum*? How is their secretion regulated? Are the FPE, Tad, holins, and ESAT-6/WXG100 pathways functional in *C. acetobutylicum*? What are their physiological roles and secretion mechanisms? Do extracellular macromolecular structures such as ricin-B and ChW proteins-based complexes exist? What are their functions? How do these macromolecular complexes, including the cellulosome, assemble on the bacterial cell surface? Are the potential virulence factors (i.e. pili, toxins, adhesins, mediators of motility) of *C. acetobutylicum* expressed and functional? How are they related to the pathogenicity of this microorganism towards plant and/or mammalian cells?

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC), in part, with the ViruloGenome projects to Prof. Mark J. Pallen, and grants 81/D14955 and 81/P14130 to Dr. Ian R. Henderson. Dr. Mickaël Desvaux is an Honorary Member of the University of Birmingham.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbamcr.2005.04.006.

References

[1] D.R. Woods, The Clostridia and Biotechnology, Butterworth-Heinemann, Boston, MA, 1993.
[2] H. Bahl, P. Durre, Clostridia: Biotechnology and Medical Applications, Wiley-VHC, Weinheim, NY, 2001.
[3] N.D. Collins, P.A. Lawson, A. Willems, J.J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, J.A.E. Farrow, The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations, Int. J. Syst. Bacteriol. 44 (1994) 812–826.
[4] E.R. Weyer, L.F. Rettger, A comparative study of six different strains of the organism commonly concerned in large-scale production of butyl alcohol and acetone by the biological process, J. Bacteriol. 14 (1927) 399–424.
[5] B. Dixon, Chaim Weizmann: from fermentation chemist to president of Israel, Anaerobe 2 (1996) 195–196.
[6] E. Cornillot, P. Soucaille, Solvent forming genes in *Clostridia*, Nature 380 (1996) 489.
[7] E. Cornillot, R.V. Nair, E.T. Papoutsakis, P. Soucaille, The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degenaration of the strain, J. Bacteriol. 179 (1997) 5442–5447.
[8] E. Cornillot, C. Croux, P. Soucaille, Physical and genetic map of the *Clostridium acetobutylicum* ATCC 824 chromosome, J. Bacteriol. 179 (1997) 5442–5447.
[9] J. Nölling, G. Breton, M.V. Omelchenko, K.S. Makarova, Q. Zeng, R. Gibson, H.M. Lee, J. Dubois, D. Qiu, J. Hitti, Y.I. Wolf, R.L. Tatusov, F. Sabathé, L. Doucette-Stamm, P. Soucaille, M.J. Daly, G.N. Bennett, E.V. Koonin, D.R. Smith, Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*, J. Bacteriol. 183 (2001) 4823–4838.
[10] H. Brüggemann, S. Bäumer, W.F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne, G. Gottschalk, The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 1316–1321.
[11] T. Shimizu, K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara, H. Hayashi, Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 996–1001.
[12] P. Durre, M. Bohringer, S. Nakotte, S. Schaffer, K. Thormann, B. Zickner, Transcriptional regulation of solventogenesis in *Clostridium acetobutylicum*, J. Mol. Microbiol. Biotechnol. 4 (2002) 295–300.
[13] W.J. Mitchell, Physiology of carbohydrate to solvent conversion by clostridia, Adv. Microb. Physiol. 39 (1998) 31–130.
[14] E.T. Papoutsakis, G.N. Bennett, Molecular regulation and metabolic engineering of solvent production by *Clostridium acetobutylicum*, Bioprocess. Technol. 24 (1999) 253–279.
[15] L. Girbal, P. Soucaille, Regulation of solvent production in *Clostridium acetobutylicum*, Trends Biotechnol. 16 (1998) 11–16.
[16] D. Greenbaum, N.M. Luscombe, R. Jansen, J. Qian, M. Gerstein, Interrelating different types of genomic data: from proteome to secretome: Oming in on function, Genome Res. 11 (2001) 1463–1468.
[17] H. Tjalsma, A. Bolhuis, J.D. Jongbloed, S. Bron, J.M. van Dijl, Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome, Microbiol. Mol. Biol. Rev. 64 (2000) 515–547.
[18] K.H.M. van Wely, J. Swaving, R. Freudl, A.J.M. Driessen, Translocation of proteins across the cell envelope of Gram-positive bacteria, FEMS Microbiol. Rev. 25 (2001) 437–454.
[19] M. Simonen, I. Palva, Protein secretion in *Bacillus* species, Microbiol. Rev. 57 (1993) 109–137.
[20] H. Tjalsma, H. Antelmann, J.D. Jongbloed, P.G. Braun, E. Darmon, R. Dorenbos, J.Y. Dubois, H. Westers, G. Zanen, W.J. Quax, O.P. Kuipers, S. Bron, M. Hecker, J.M. Van Dijl, Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome, Microbiol. Mol. Biol. Rev. 68 (2004) 207–233.
[21] M. Hecker, A proteomic view of cell physiology of *Bacillus subtilis*—Bringing the genome sequence to life, Adv. Biochem. Eng. Biotechnol. 83 (2003) 57–92.
[22] M.R. Sharipova, Late stages of protein secretion in bacilli, Biochemistry (Mosc) 67 (2002) 1207–1216.
[23] M.J. Pallen, R.R. Chaudhuri, I.R. Henderson, Genomic analysis of secretion systems, Curr. Opin. Microbiol. 6 (2003) 519–527.
[24] H. Brüggemann, G. Gottschalk, Insights in metabolism and toxin production from the complete genome sequence of *Clostridium tetani*, Anaerobe 10 (2004) 53–68.
[25] G.P. Salmond, P.J. Reeves, Membrane traffic wardens and protein secretion in Gram-negative bacteria, Trends Biochem. Sci. 18 (1993) 7–12.
[26] D.G. Thanassi, Ushers and secretins: channels for the secretion of folded proteins across the bacterial outer membrane, J. Mol. Microbiol. Biotechnol. 4 (2002) 11–20.

[27] I.R. Henderson, J.P. Nataro, J.B. Kaper, T.F. Meyer, S.K. Farrand, D.L. Burns, B.B. Finlay, J.W. St Geme III, Renaming protein secretion in the Gram-negative bacteria, Trends Microbiol. 8 (2000) 352.
[28] M. Desvaux, N.J. Parham, I.R. Henderson, Le système de sécrétion de Type V chez les bactéries Gram-négatives, Biofutur 237 (2003) 34–37.
[29] I.R. Henderson, F. Navarro-Garcia, J.P. Nataro, The great escape: structure and function of the autotransporter proteins, Trends Microbiol. 6 (1998) 370–378.
[30] A. Blocker, K. Komoriya, S. Aizawa, Type III secretion systems and bacterial flagella: insights into their function from structural similarities, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 3027–3030.
[31] J.A. Sexton, J.P. Vogel, Type IVb secretion by intracellular pathogens, Traffic 3 (2002) 178–185.
[32] C. Andersen, Channel-tunnels: outer membrane components of Type I secretion systems and multidrug efflux pumps of Gram-negative bacteria, Rev. Physiol. Biochem. Pharmacol. 147 (2003) 122–165.
[33] D. Buttner, U. Bonas, Port of entry-the Type III secretion translocon, Trends Microbiol. 10 (2002) 186–192.
[34] M. Sandkvist, Type II secretion and pathogenesis, Infect. Immun. 69 (2001) 3523–3535.
[35] P.J. Christie, Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines, Mol. Microbiol. 40 (2001) 294–305.
[36] C. Stathopoulos, D.R. Hendrixson, D.G. Thanassi, S.J. Hultgren, J.W. St Geme III, R. Curtiss III, Secretion of virulence determinants by the general secretory pathway in Gram-negative pathogens: an evolving story, Microbes Infect. 2 (2000) 1061–1072.
[37] D.G. Thanassi, S.J. Hultgren, Multiple pathways allow protein secretion across the bacterial outer membrane, Curr. Opin. Cell. Biol. 12 (2000) 420–430.
[38] M. Desvaux, N.J. Parham, I.R. Henderson, The autotransporter secretion system, Res. Microbiol. 155 (2004) 53–60.
[39] P. Fekkes, A.J. Driessen, Protein targeting to the bacterial cytoplasmic membrane, Microbiol. Mol. Biol. Rev. 63 (1999) 161–173.
[40] B.C. Berks, F. Sargent, T. Palmer, The Tat protein export pathway, Mol. Microbiol. 35 (2000) 260–274.
[41] M.J. Pallen, The ESAT-6/WXG100 superfamily—And a new Gram-positive secretion system? Trends Microbiol. 10 (2002) 209–212.
[42] K.M. Guinn, M.J. Hickey, S.K. Mathur, K.L. Zakel, J.E. Grotzke, D.M. Lewinsohn, S. Smith, D.R. Sherman, Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*, Mol. Microbiol. 51 (2004) 359–370.
[43] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[44] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[45] R.R. Chaudhuri, A.M. Khan, M.J. Pallen, coliBASE: an online database for *Escherichia coli*, *Shigella* and *Salmonella* comparative genomics, Nucleic Acids Res. 32 (2004) D296–D299.
[46] E.L. Soohammer, S.R. Eddy, R. Durbin, Pfam: a comprehensive database of protein domain families based on seed alignments, Proteins 28 (1997) 405–420.
[47] A. Marchler-Bauer, J.B. Anderson, C. DeWeese-Scott, N.D. Fedorova, L.Y. Geer, S. He, D.I. Hurwitz, J.D. Jackson, A.R. Jacobs, C.J. Lanczycki, C.A. Liebert, C. Liu, T. Madej, G.H. Marchler, R. Mazumder, A.N. Nikolskaya, A.R. Panchenko, B.S. Rao, B.A. Shoemaker, V. Simonyan, J.S. Song, R.A. Thiessen, S. Vasudevan, Y. Wang, R.A. Yamashita, J.J. Yin, S.H. Bryant, CDD: a curated Entrez database of conserved domain alignments, Nucl. Acids Res. 31 (2003) 383–387.
[48] A. Gattiker, E. Gasteiger, A. Bairoch, ScanProsite: a reference implementation of a PROSITE scanning tool, Appl. Bioinformatics 1 (2002) 107–108.
[49] I. Letunic, R.R. Copley, S. Schmidt, F.D. Ciccarelli, T. Doerks, J. Schultz, C.P. Ponting, P. Bork, SMART 4.0: towards genomic data integration, Nucleic Acids Res. 32 (2004) D142–D144.
[50] N.J. Mulder, R. Apweiler, T.K. Attwood, A. Bairoch, D. Barrell, A. Bateman, D. Binns, M. Biswas, P. Bradley, P. Bork, P. Bucher, R.R. Copley, E. Courcelle, U. Das, R. Durbin, L. Falquet, W. Fleischmann, S. Griffiths-Jones, D. Haft, N. Harte, N. Hulo, D. Kahn, A. Kanapin, M. Krestyaninova, R. Lopez, I. Letunic, D. Lonsdale, V. Silventoinen, S.E. Orchard, M. Pagni, D. Peyruc, C.P. Ponting, J.D. Selengut, F. Servant, C.J.A. Sigrist, R. Vaughan, E.M. Zdobnov, The InterPro Database, 2003 brings increased coverage and new features, Nucleic Acids Res. 31 (2003) 315–318.
[51] T.A. Tatusova, T.L. Madden, Blast 2 sequences—A new tool for comparing protein and nucleotide sequences, FEMS Microbiol. Lett. 174 (1999) 247–250.
[52] D. Higgins, J. Thompson, T. Gibson, J.D. Thompson, D.G. Higgins, T.J. Gibson, 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 (1994) 4673–4680.
[53] Q. Ma, Y. Zhai, J.C. Schneider, T.M. Ramseir, M.H. Saier Jr., Protein secretion systems of *Pseudomonas aeruginosa* and *P. fluorescens*, Biochim. Biophys. Acta, Biomembr. 1611 (2003) 223–233.
[54] W. Busch, M.H. Saier Jr., The transporter classification (TC) system, Crit. Rev. Biochem. Mol. Biol. 37 (2002) 287–337.
[55] Q. Ren, K.H. Kang, I.T. Paulsen, TransportDB: a relational database of cellular membrane transport systems, Nucleic Acids Res. 32 (2004) D284–D288.
[56] M.J. Pallen, The ESAT-6/WXG100 superfamily—And a new Gram-positive secretion system? Trends Microbiol. 10 (2002) 209–212.
[57] M.L. Burts, W.A. Williams, K. DeBord, D.M. Missiakas, EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 1169–1174.
[58] H. Nielsen, S. Brunak, G. von Heijne, Machine learning approaches for the prediction of signal peptides and other protein sorting signals, Protein Eng. 12 (1999) 3–9.
[59] K. Nakai, P. Horton, PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization, Trends Biochem. Sci. 24 (1999) 34–36.
[60] H. Bannai, Y. Tamada, O. Maruyama, K. Nakai, S. Miyano, Extensive feature detection of N-terminal protein sorting signals, Bioinformatics 18 (2002) 298–305.
[61] I. Sutcliffe, D.J. Harrington, Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes, Microbiology 148 (2002) 2065–2077.
[62] M.M. Babu, K. Sankaran, DOLOP—Database of bacterial lipoproteins, Bioinformatics 18 (2002) 641–643.
[63] A.S. Juncker, H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen, A. Krogh, Prediction of lipoprotein signal peptides in Gram-negative bacteria, Protein Sci. 12 (2003) 1652–1662.
[64] K. Hofmann, W. Stoffel, TMbase—A database of membrane spanning protein segments, Biol. Chem. Hoppe-Seyler 374 (1993) 166.
[65] M. Cserzo, E. Wallin, I. Simon, G. von Heijne, A. Elofsson, Prediction of transmembrane α-helices in prokaryotic membrane proteins: the dense alignment surface method, Protein Eng. 10 (1997) 673–676.
[66] M.G. Claros, G. von Heijne, TopPred II: an improved software for membrane protein structure predictions, Comput. Appl. Biosci. 10 (1994) 685–686.
[67] D. Dubnau, Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type 4 pilin-like proteins—A review, Gene 192 (1997) 191–198.
[68] S.V. Albers, A.J. Driessen, Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey, Arch. Microbiol. 177 (2002) 209–216.

[69] K. Dilks, R.W. Rose, E. Hartmann, M. Pohlschröder, Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey, J. Bacteriol. 185 (2003) 1478–1483.
[70] E. Dassa, P. Bouige, The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms, Res. Microbiol. 152 (2001) 211–229.
[71] T. Yakushi, K. Masuda, S.I. Narita, S.I. Matsuyama, H. Tokuda, New ABC transporter mediating the detachment of lipid-modified proteins from membranes, Nat. Cell. Biol. 2 (2000) 212–218.
[72] B. Ajouz, A. Berrier, A. Garrigues, M. Besnard, A. Ghazi, Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells, J. Biol. Chem. 273 (1998) 26670–26674.
[73] C.D. Pivetti, M.R. Yen, S. Miller, W. Busch, Y.H. Tseng, I.R. Booth, M.H. Saier Jr., Two families of mechanosensitive channel proteins, Microbiol. Mol. Biol. Rev. 67 (2003) 66–85.
[74] M. Desvaux, N.J. Parham, A. Scott-Tucker, I.R. Henderson, The general secretory pathway: a general misnomer? Trends Microbiol. 12 (2004) 306–309.
[75] A.J. Driessen, E.H. Manting, C. van der Does, The structural basis of protein targeting and translocation in bacteria, Nature 8 (2001) 492–498.
[76] T.B. Cao, M.H. Saier Jr., The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions, Biochim. Biophys. Acta, Biomembr. 1609 (2003) 115–125.
[77] B.A. Bensing, P.M. Sullam, An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets, Mol. Microbiol. 44 (2002) 1081–1094.
[78] J. De Keyzer, C. Van der Does, A.J. Driessen, The bacterial translocase: a dynamic protein channel complex, Cell. Mol. Life Sci. 60 (2003) 2034–2052.
[79] A. Bolhuis, C.P. Broekhuizen, A. Sorokin, M. van Roosmalen, G. Venema, S. Bron, W.J. Quax, J.M. van Dijl, SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of proteins, J. Biol. Chem. 274 (1998) 21217–21224.
[80] J.C. Samuelson, M. Chen, M. Jiang, I. Moller, M. Wiedmann, A. Kuhn, G.J. Phillips, R.E. Dalbey, YidC mediates membrane protein insertion in bacteria, Nature 406 (2000) 637–641.
[81] M. Chen, K. Xie, F. Jiang, L. Yi, R.E. Dalbey, YidC, a newly defined evolutionarily conserved protein, mediates membrane protein assembly in bacteria, Biol. Chem. 383 (2002) 1565–1572.
[82] M. van der Laan, P. Bechtluft, S. Kol, N. Nouwen, A.J. Driessen, F_1F_0 ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis, J. Cell. Biol. 165 (2004) 213–222.
[83] M.R. Yen, C.R. Peabody, S.M. Partovi, Y. Zhai, Y.H. Tseng, M.H. Saier, Protein-translocating outer membrane porins of Gram-negative bacteria, Biochim. Biophys. Acta, Biomembr. 1562 (2002) 6–31.
[84] L. Froderberg, E. Houben, J.C. Samuelson, M. Chen, S.K. Park, G.J. Phillips, R. Dalbey, J. Luirink, J.W. De Gier, Versatility of inner membrane protein biogenesis in *Escherichia coli*, Mol. Microbiol. 47 (2003) 1015–1027.
[85] Q.A. Valent, P.A. Scotti, S. High, J.W. de Gier, G. von Heijne, G. Lentzen, W. Wintermeyer, B. Oudega, J. Luirink, The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon, EMBO J. 17 (1998) 2504–2512.
[86] A.P. Pugsley, The complete general secretory pathway in Gram-negative bacteria, Microbiol. Rev. 57 (1993) 50–108.
[87] D. Beha, S. Deitermann, M. Muller, H.G. Koch, Export of β-lactamase is independent of the signal recognition particle, J. Biol. Chem. 278 (2003) 22161–22167.
[88] R. Sijbrandi, M.L. Urbanus, C.M. ten Hagen-Jongman, H.D. Bernstein, B. Oudega, B.R. Otto, J. Luirink, Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular *Escherichia coli* protein, J. Biol. Chem. 278 (2003) 4654–4659.
[89] L.L. Randall, S.J. Hardy, SecB, one small chaperone in the complex milieu of the cell, Cell. Mol. Life Sci. 59 (2002) 1617–1623.
[90] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, N.J. Cummings, R.A. Daniel, F. Denizot, K.M. Devine, A. Düsterhöft, S.D. Ehrlich, P.T. Emmerson, K.D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S.Y. Ghim, P. Glaser, A. Goffeau, E.J. Golightly, G. Grandi, G. Guiseppi, B.J. Guy, K. Haga, J. Haiech, C.R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M.F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S.M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R.P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S.H. Park, V. Parro, T.M. Pohl, D. Portetelle, S. Porwollik, A.M. Prescott, E. Presecan, P. Pujic, B. Purnelle, et al., The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, Nature 390 (1997) 249–256.
[91] G.A. Barkocy-Gallagher, P.J. Bassford Jr., Synthesis of precursor maltose-binding protein with proline in the +1 position of the cleavage site interferes with the activity of *Escherichia coli* signal peptidase I in vivo, J. Biol. Chem. 267 (1992) 1231–1238.
[92] I. Nilsson, G. von Heijne, A signal peptide with a proline next to the cleavage site inhibits leader peptidase when present in a sec-independent protein, FEBS Lett. 299 (1992) 243–246.
[93] M. Sara, U.B. Sleytr, S-layer proteins, J. Bacteriol. 182 (2000) 859–868.
[94] O. Dussurget, J. Pizarro-Cerda, P. Cossart, Molecular determinants of *Listeria monocytogenes* virulence, Annu. Rev. Microbiol. 58 (2004) 587–610.
[95] D. Cabanes, P. Dehoux, O. Dussurget, L. Frangeul, P. Cossart, Surface proteins and the pathogenic potential of *Listeria monocytogenes*, Trends Microbiol. (2002) 238–245.
[96] E. Garcia, J.L. Garcia, P. Garcia, A. Arraras, J.M. Sanchez-Puelles, R. Lopez, Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 914–918.
[97] A.J. Waligora, C. Hennequin, P. Mullany, P. Bourlioux, A. Collignon, T. Karjalainen, Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties, Infect. Immun. 69 (2001) 2144–2153.
[98] M. Kuhn, W. Goebel, Identification of an extracellular protein of *Listeria monocytogenes* possibly involved in intracellular uptake by mammalian cells, Infect. Immun. 57 (1989) 55–61.
[99] M.D. Wuenscher, S. Kohler, A. Bubert, U. Gerike, W. Goebel, The *iap* gene of *Listeria monocytogenes* is essential for cell viability, and its gene product, p60, has bacteriolytic activity, J. Bacteriol. 175 (1993) 3491–3501.
[100] A.R. Sanchez-Beato, C. Ronda, J.L. Garcia, Tracking the evolution of the bacterial choline-binding domain: molecular characterization of the *Clostridium acetobutylicum* NCIB 8052 cspA gene, J. Bacteriol. 177 (1995) 1098–1103.
[101] R.H. Doi, A. Kosugi, K. Murashima, Y. Tamaru, S.O. Han, Cellulosomes from mesophilic bacteria, J. Bacteriol. 185 (2003) 5907–5914.
[102] Y. Shoham, R. Lamed, E.A. Bayer, The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides, Trends Microbiol. 7 (1999) 275–281.
[103] A. Mechaly, S. Yaron, R. Lamed, H.P. Fierobe, A. Bélaïch, J.P. Bélaïch, Y. Shoham, E.A. Bayer, Cohesin–dockerin recognition in cellulosome assembly: experiment versus hypothesis, Proteins 39 (2000) 170–177.

[104] A.L. Carvalho, F.M.V. Dias, J.A.M. Prates, T. Nagy, H.J. Gilbert, G.J. Davies, L.M.A. Ferreira, M.J. Romao, C.M.G.A. Fontes, Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 13809–13814.
[105] F. Sabathé, A. Bélaïch, P. Soucaille, Characterisation of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum*, FEMS Microbiol. Lett. 217 (2002) 15–22.
[106] S. Pagès, A. Bélaïch, H.P. Fierobe, C. Tardif, C. Gaudin, J.P. Bélaïch, Sequence analysis of scaffolding protein CipC and ORFXp, a new cohesin-containing protein in *Clostridium cellulolyticum*: comparison of various cohesin domains and subcellular localization of ORFXp, J. Bacteriol. 181 (1999) 1801–1810.
[107] Y. Tamaru, R.H. Doi, Three surface layer homology domains at the N-terminus of the *Clostridium cellulovorans* major cellulosomal subunit EngE, J. Bacteriol. 181 (1999) 3270–3276.
[108] A. Kosugi, K. Murashima, Y. Tamaru, R.H. Doi, Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE, J. Bacteriol. 184 (2002) 884–888.
[109] W.W. Navarre, O. Schneewind, Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope, Microbiol. Mol. Biol. Rev. 63 (1999) 174–229.
[110] M.J. Pallen, A.C. Lam, M. Antonio, K. Dunbar, An embarrassment of sortase—A richness of substrates? Trends Microbiol. 9 (2001) 97–101.
[111] S.G. Lee, V. Pancholi, V.A. Fischetti, Characterization of unique glycosylated anchor endopeptidase that cleaves the LPXTG sequence motif of cell surface proteins of Gram-positive bacteria, J. Biol. Chem. 277 (2002) 46912–46922.
[112] D. Comfort, R.T. Clubb, A comparative genome analysis identifies distinct sorting pathways in Gram-positive bacteria, Infect. Immun. 72 (2004) 2710–2722.
[113] R.M. Macnab, How bacteria assemble flagella, Annu. Rev. Microbiol. 57 (2003) 77–100.
[114] R. Macnab, Type III flagellar protein export and flagellar assembly, Biochim. Biophys. Acta, Mol. Cell Res. 1694 (2004) 207–217.
[115] T. Iino, Y. Komeda, K. Kutsukake, R.M. Macnab, P. Matsumura, New unified nomenclature for the flagellar genes of *Escherichia coli* and *Salmonella typhimurium*, Microbiol. Rev. 52 (1988) 533–535.
[116] S.L. Bardy, Y.M. Sandy, K.F. Jarrell, Prokaryotic motility structures, Microbiology 149 (2003) 295–304.
[117] V.T. Lee, O.J. Schneewind, Yop fusions to tightly folded protein domains and their effects on *Yersinia enterocolitica* Type III secretion, J. Bacteriol. 184 (2002) 3740–3745.
[118] Q. Jin, S.Y. He, Role of the Hrp pilus in Type III protein secretion in *Pseudomonas syringae*, Science 294 (2001) 2556–2558.
[119] G.M. Young, D.H. Schmiel, V.L. Miller, A new pathway for the secretion of virulence factors by bacteria, the flagellar export apparatus functions as a protein-secretion system, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 6456–6461.
[120] M.E. Konkel, J.D. Klena, V. Rivera-Amill, M.R. Monteville, D. Biswas, B. Raphael, J. Mickelson, Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus, J. Bacteriol. 186 (2004) 3296–3303.
[121] D. Gygi, M.J. Bailey, C. Allison, C. Hughes, Requirement for FlhA in flagella assembly and swarm-cell differentiation by *Proteus mirabilis*, Mol. Microbiol. 15 (1995) 761–769.
[122] E. Ghelardi, F. Celandroni, S. Salvetti, D.J. Beecher, M. Gominet, D. Lereclus, A.C. Wong, S. Senesi, Requirement of flhA for swarming differentiation, flagellin export, and secretion of virulence-associated proteins in *Bacillus thuringiensis*, J. Bacteriol. 184 (2002) 6424–6433.
[123] N.A. Gutierrez, I.S. Maddox, Role of chemotaxis in solvent production by *Clostridium acetobutylicum*, Appl. Environ. Microbiol. 53 (1987) 1924–1927.
[124] M.J. Spivey, The acetone/butanol/ethanol fermentation, Proc. Biochem. 13 (1978) 2–5.
[125] C.A. Tomas, N.E. Welker, E.T. Papoutsakis, Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program, Appl. Environ. Microbiol. 69 (2003) 4951–4965.
[126] C.A. Tomas, K.V. Alsaker, H.P. Bonarius, W.T. Hendriksen, H. Yang, J.A. Beamish, C.J. Paredes, E.T. Papoutsakis, DNA array-based transcriptional analysis of asporogenous, nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5, J. Bacteriol. 185 (2003) 4539–4547.
[127] S.I. Aizawa, Bacterial flagella and Type III secretion systems, FEMS Microbiol. Lett. 202 (2001) 157–164.
[128] D. Dubnau, DNA uptake in bacteria, Annu. Rev. Microbiol. 53 (1999) 217–244.
[129] C.R. Peabody, Y.J. Chung, M.R. Yen, D. Vidal-Ingigliardi, A.P. Pugsley, M.H. Saier Jr., Type II protein secretion and its relationship to bacterial Type 4 pili and archaeal flagella, Microbiology 149 (2003) 3051–3072.
[130] D. Dubnau, R. Provvedi, Internalizing DNA, Res. Microbiol. 151 (2000) 475–480.
[131] Y.S. Chung, F. Breidt, D. Dubnau, Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of *Bacillus subtilis*, Mol. Microbiol. 29 (1998) 905–913.
[132] M. Sandkvist, Biology of Type II secretion, Mol. Microbiol. 40 (2001) 271–283.
[133] S.C. Kachlany, P.J. Planet, M.K. Bhattacharjee, E. Kollia, R. DeSalle, D.H. Fine, D.H. Figurski, Nonspecific adherence by *Actinobacillus actinomycetemcomitans* requires genes widespread in *Bacteria* and *Archaea*, J. Bacteriol. 182 (2000) 6169–6176.
[134] H.C. Schreiner, K. Sinatra, J.B. Kaplan, D. Furgang, S.C. Kachlany, P.J. Planet, B.A. Perez, D.H. Figurski, D.H. Fine, Tight-adherence genes of *Actinobacillus actinomycetemcomitans* are required for virulence in a rat model, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 7295–7300.
[135] P.J. Planet, S.C. Kachlany, D.H. Fine, R. DeSalle, D.H. Figurski, The widespread colonization island of *Actinobacillus actinomycetemcomitans*, Nat. Genet. 34 (2003) 193–198.
[136] M.K. Bhattacharjee, S.C. Kachlany, D.H. Fine, D.H. Figurski, Nonspecific adherence and fibril biogenesis by *Actinobacillus actinomycetemcomitans*, J. Bacteriol. 183 (2001) 5927–5936.
[137] J.M. Skerker, L. Shapiro, Identification and cell cycle control of a novel system in *Caulobacter cresentus*, EMBO J. 19 (2000) 3223–3234.
[138] S.C. Kachlany, P.J. Planet, R. Desalle, D.H. Fine, D.H. Figurski, J.B. Kaplan, *flp-1*, the first representative of a new pilin gene subfamily, is required for non-specific adherence of *Actinobacillus actinomycetemcomitans*, Mol. Microbiol. 40 (2001) 542–554.
[139] S.C. Kachlany, P.J. Planet, R. DeSalle, D.H. Fine, D.H. Figurski, Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum, Trends Microbiol. 9 (2001) 429–437.
[140] I.N. Wang, D.L. Smith, R. Young, Holins: the protein clocks of bacteriophage infections, Annu. Rev. Microbiol. 54 (2000) 799–825.
[141] K.W. Bayles, Are the molecular strategies that control apoptosis conserved in bacteria? Trends Microbiol. 11 (2003) 306–311.
[142] K.H. Groicher, B.A. Firek, D.F. Fujimoto, K.W. Bayles, The *Staphylococcus aureus lrgAB* operon modulates murein hydrolase activity and penicillin tolerance, J. Bacteriol. 182 (2000) 1794–1801.
[143] K.C. Rice, B.A. Firek, J.B. Nelson, S.J. Yang, T.G. Patton, K.W. Bayles, The *Staphylococcus aureus cidAB* operon: evaluation of its role in regulation of murein hydrolase activity and penicillin tolerance, J. Bacteriol. 185 (2003) 2635–2643.
[144] C. Croux, B. Canard, G. Goma, P. Soucaille, Autolysis of *Clostridium acetobutylicum* ATCC824, J. Gen. Microbiol. 138 (1992) 861–869.

[145] K.S. Tan, B.Y. Wee, K.P. Song, Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*, J. Med. Microbiol. 50 (2001) 613–619.
[146] N.C. Gey Van Pittius, J. Gamieldien, W. Hide, G.D. Brown, R.J. Siezen, A.D. Beyers, The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria, Genome Biol. 2 (2001) 0044.1–0044.18.
[147] A.L. Sørensen, S. Nagai, G. Houen, P. Andersen, A.B. Andersen, Purification and characterization of a low-molecular-mass t-cell antigen secreted by *Mycobacterium tuberculosis*, Infect. Immun. 63 (1995) 1710–1717.
[148] T. Hsu, S.M. Hingley-Wilson, B. Chen, M. Chen, A.Z. Dai, P.M. Morin, C.B. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R.G. Russell, S.C. Derrick, F.M. Collins, S.L. Morris, C.H. King, W.R. Jacobs Jr., The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 12420–12425.
[149] D. Durocher, S.P. Jackson, The FHA domain, FEBS Lett. 513 (2002) 58–66.
[150] D. Durocher, Bacterial signal transduction: a FHAscinating glimpse at the origins of phospho-dependent signal transduction, Trends Microbiol. 11 (2003) 67–68.
[151] M. Pallen, R. Chaudhuri, A. Khan, Bacterial FHA domains: neglected players in the phospho-threonine signalling game? Trends Microbiol. 11 (2002) 556–563.
[152] G. Schneider, How many potentially secreted proteins are contained in a bacterial genome? Gene 237 (1999) 113–121.
[153] H. Antelmann, H. Tjalsma, B. Voigt, S. Ohlmeier, S. Bron, J.M. van Dijl, M. Hecker, A proteomic view on genome-based signal peptide predictions, Genome Res. 11 (2001) 1484–1502.
[154] A. Danchin, Genomes and evolution, Curr. Issues Mol. Biol. 5 (2003) 37–42.
[155] K. Mukherjee, S. Karlsson, L.G. Burman, T. Akerlund, Proteins released during high toxin production in *Clostridium difficile*, Microbiology 148 (2002) 2245–2253.
[156] T. Shimizu, K. Shima, K.I. Yoshino, K. Yonezawa, T. Shimizu, H. Hayashi, Proteome and transciptome analysis of the virulence genes regulated by the VirR/VirS system in *Clostridium perfringens*, J. Bacteriol. 184 (2002) 2587–2594.
[157] D.M. Downs, Genomics and bacterial metabolism, Curr. Issues Mol. Biol. 5 (2003) 17–25.
[158] J. Rosch, M. Caparon, A microdomain for protein secretion in Gram-positive bacteria, Science 304 (2004) 1513–1515.
[159] H. Adams, P.A. Scotti, H. De Cock, J. Luirink, J. Tommassen, The presence of a helix breaker in the hydrophobic core of signal sequences of secretory proteins prevents recognition by the signal-recognition particle in *Escherichia coli*, Eur. J. Biochem. 269 (2002) 5564–5571.
[160] M. Desvaux, The cellulosome of *Clostridium cellulolyticum*, Enzyme Microb. Technol. (2005), doi:10.1016/j.enzmictec.2004.04.025.
[161] M. Desvaux, *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia, FEMS Microbiol. Rev. (2005), doi:10.1016/j.femsre.2004.11.003.
[162] W.H. Schwarz, The cellulosome and cellulose degradation by anaerobic bacteria, Appl. Microbiol. Biotechnol. 56 (2001) 634–649.
[163] G.K. Paterson, T.J. Mitchell, The biology of Gram-positive sortase enzymes, Trends Microbiol. 12 (2004) 89–95.
[164] M. Marino, M. Banerjee, R. Jonquieres, P. Cossart, P. Ghosh, GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands, EMBO J. 21 (2002) 5623–5634.
[165] R. Koebnik, The role of bacterial pili in protein and DNA translocation, Trends Microbiol. 9 (2001) 586–590.
[166] H. Ton-That, O. Schneewind, Assembly of pili in Gram-positive bacteria, Trends Microbiol. 12 (2004) 228–234.
[167] I. Chen, D. Dubnau, DNA transport during transformation, Front. Biosci. 8 (2003) s544–s556.
[168] I. Chen, D. Dubnau, DNA uptake during bacterial transformation, Nat. Rev., Microbiol. 2 (2004) 241–249.
[169] J.S. Mattick, Type 4 pili and twitching motility, Annu. Rev. Microbiol. 56 (2002) 289–314.
[170] M. Koomey, Competence for natural transformation in *Neisseria gonorrhoeae*: a model system for studies of horizontal gene transfer, APMIS 84 (1998) 56–61.
[171] S.P. Borriello, Pathogenesis of *Clostridium difficile* infection, J. Antimicrob. Chemother. 41 (1998) 13–19.
[172] M. Lyristis, Z.L. Boynton, D. Petersen, Z. Kan, G.N. Bennett, F.B. Rudolph, Cloning, sequencing, and characterization of the gene encoding flagellin, *flaC*, and the post-translational modification of flagellin, FlaC, from *Clostridium acetobutylicum* ATCC824, Anaerobe 6 (2000) 69–79.
[173] M. Holden, L. Crossman, A. Cerdeno-Taraga, J. Parkhill, Pathogenomics of non-pathogens, Nat. Rev., Microbiol. 2 (2004) 91.
[174] K.R. Sakharkar, M.K. Sakharkar, V.T. Chow, A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*, In Silico Biol. 4 (2004) 0028.
[175] R.H. Doi, A. Kosugi, Cellulosomes: plant-cell-wall-degrading enzymes complexes, Nat. Rev., Microbiol. 2 (2004) 541–551.
[176] G. Koraimann, Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria, Cell. Mol. Life Sci. 60 (2003) 2371–2388.
[177] O. Matsushita, A. Okabe, Clostridial hydrolytic enzymes degrading extracellular components, Toxicon 39 (2001) 1769–1780.
[178] E.A. Johnson, Clostridial toxins as therapeutic agents: benefits of nature's most toxic proteins, Annu. Rev. Microbiol. 53 (1999) 551–575.
[179] K.R. Kessler, R. Benecke, Botulinum toxin: from poison to remedy, Neurotoxicology 18 (1997) 761–770.
[180] N.P. Minton, M.L. Mauchline, M.J. Lemmon, J.K. Brehm, M. Fox, N.P. Michael, A. Giaccia, J.M. Brown, Chemotherapeutic tumor targeting using clostridial spores, FEMS Microbiol. Rev. 17 (1995) 357–364.
[181] N.P. Minton, *Clostridia* in cancer therapy, Nat. Rev., Microbiol. 1 (2003) 237–242.
[182] P. Lambina, J. Theysb, W. Landuyta, P. Rijkenc, A. Van der Kogelc, E. Van der Schuerena, R. Hodgkissd, J. Fowlera, S. Nuytsb, E. De Bruijna, L. Van Mellaertb, J. Anné, Colonisation of *Clostridium* in the body is restricted to hypoxic and necrotic areas of tumours, Anaerobe 4 (1998) 183–188.
[183] L.H. Dang, C. Bettegowda, D.L. Huso, K.W. Kinzler, B. Vogelstein, Combination bacteriolytic therapy for the treatment of experimental tumors, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 15155–15160.
[184] J. Theys, A.W. Landuyt, S. Nuyts, L. Van Mellaert, P. Lambin, J. Anne, W. Landuyt, *Clostridium* as a tumor-specific delivery system of therapeutic proteins, Cancer Detect. Prev. 25 (2001) 548–557.
[185] S. Nuyts, L. Van Mellaert, J. Theys, W. Landuyt, P. Lambin, J. Anne, *Clostridium* spores for tumor-specific drug delivery, Anticancer Drugs 13 (2002) 115–125.
[186] J. Theys, S. Nuyts, W. Landuyt, L. Van Mellaert, C. Dillen, M. Bohringer, P. Durre, P. Lambin, J. Anne, Stable *Escherichia coli*–*Clostridium acetobutylicum* shuttle vector for secretion of murine tumor necrosis factor alpha, Appl. Environ. Microbiol. 65 (1999) 4295–4300.
[187] R.K. Jain, N.S. Forbes, Can engineered bacteria help control cancer? Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 14748–14780.