Available online at www.sciencedirect.com
SCIENCE DIRECT®
PEPTIDES
Peptides 25 (2004) 1425–1440
www.elsevier.com/locate/peptides

Review

Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters

G. Dirix[a], P. Monsieurs[b], B. Dombrecht[a], R. Daniels[a], K. Marchal[b], J. Vanderleyden[a], J. Michiels[a,*]

[a] *Centre of Microbial and Plant Genetics, Kasteelpark Arenberg 20, B-3001 Heverlee, Belgium*
[b] *ESAT-SCD, K.U. Leuven, Heverlee, Belgium*

Received 15 July 2003; accepted 31 October 2003

Abstract

Quorum sensing (QS) in Gram-negative bacteria is generally assumed to be mediated by *N*-acyl-homoserine lactone molecules while Gram-positive bacteria make use of signaling peptides. We analyzed the occurrence in Gram-negative bacteria of peptides and transporters that are involved in quorum sensing in Gram-positive bacteria. Many class II bacteriocins and inducing factors produced by lactic acid bacteria (LAB) and competence stimulating peptides (CSPs) synthesized by streptococci are processed by their cognate ABC-transporters during their secretion. During transport, a conserved leader sequence, termed the double-glycine motif (GG-motif), is cleaved off by the N-terminal domain of the transporter, which belongs to the Peptidase C39 protein family. Several peptides containing a GG-motif were recently described in Gram-negative bacteria (Trends Microbiol 2001;9:164–8). To screen for additional putative GG-motif containing peptides, an in silico strategy based on MEME, HMMER2.2 and Wise2 was designed. Using a curated training set, a motif model of the leader peptide was built and used to screen over 120 fully sequenced bacterial genomes. The screening methodology was applied at the nucleotide level as probably many small peptide genes have not been annotated and may be absent from the non-redundant databases. It was found that 33% of the screened genomes of Gram-negative bacteria contained one or more transporters carrying a Peptidase C39 domain, compared to 44% of the genomes of Gram-positive bacteria. The transporters can be subdivided into four classes on the basis of their domain organization. Genes coding for putative peptides containing 23–142 amino acids and a GG-motif were found in close association with genes coding for Peptidase C39 domain containing proteins. These peptides show structural similarity to bacteriocins and peptide pheromones of Gram-positive bacteria. The possibility of signal transduction based on peptide signaling in Gram-negative bacteria is discussed.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Quorum sensing; Peptide pheromone; Signal transduction; Membrane transport; Double-glycine motif; Peptidase C39 domain; ABC-transporters; Protein secretion; Genome-wide screening; HSL; Autoinducer; Bacteriocin

Contents

1. Quorum sensing 1426
   1.1. Quorum sensing in Gram-negative bacteria 1426
   1.2. Quorum sensing in Gram-positive bacteria 1426
2. Bacteriocins 1427
3. Peptides containing a double-glycine motif 1427

* Corresponding author. Tel.: +32 16 321631; fax: +32 16 321966.
E-mail address: Jan.Michiels@agr.kuleuven.ac.be (J. Michiels).

0196-9781/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2003.10.028

4. Genome wide in silico screening for peptides containing double-glycine motif leader sequences and their cognate transporters 1428
4.1. Overview 1428
4.2. Peptidase C39 domain search 1428
4.3. Search for genes encoding double-glycine motif containing peptides 1433
5. Can peptides be involved in signal transduction in Gram-negative bacteria? 1436
References 1438

1. Quorum sensing

Since some of the possible double-glycine motif (GG-motif) containing peptides show similarity to peptide pheromones, a brief overview of quorum sensing (QS) in Gram-negative and Gram-positive bacteria will be presented. Many bacterial processes are regulated in a population density dependent manner. Individual bacterial cells produce and secrete small amounts of a signaling molecule (often called 'autoinducer' or recently quormon), the extracellular concentration of which is related to the population density of the producing organism. When a certain threshold of the signaling molecule is reached, the bacteria sense and respond to this, resulting in a process resembling multicellular behavior. Cell–cell signaling in response to the population density is called 'quorum sensing' (QS). The structures of the signaling molecules differ among Gram-negative and Gram-positive bacteria.

1.1. Quorum sensing in Gram-negative bacteria

The signal molecules involved in QS in Gram-negative bacteria may vary from *N*-acyl-homoserine lactone (AHL) signals in different bacterial species to 3-hydroxypalmitic acid methyl ester of *Ralstonia solanacearum*, AI-2 from *Vibrio harveyi*, bradyoxetin from *Bradyrhizobium japonicum* and 2-heptyl-3-hydroxy-4-quinolone from *Pseudomonas aeruginosa* (for an overview see [37,50,68]).

AHLs are synthesized in Gram-negative bacteria at a low basal level by AHL synthases, generally homologous to the *Vibrio fischeri* LuxI-type protein (for an overview on the molecular biology and genetics of QS with AHLs see [20]). AHLs may be saturated or unsaturated and mainly vary with respect to the length (4–18 carbons) and the substituent (H, O or OH) at the third carbon of the acyl-side chain [39,55]. Newly synthesized AHL signals rapidly move from the cell either by passive diffusion, as observed for 3O,C_6-HSL in *V. fischeri* and *Escherichia coli* cells [29], or by a combination of diffusion and active efflux of AHLs with longer acyl-side chains as in *P. aeruginosa* [49]. At least in *V. fischeri*, the signal freely diffuses back into the cells such that its intracellular concentration also rises as a function of the increase in bacterial population. At a threshold level, the AHL interacts with a transcription factor, homologous to the LuxR protein in *V. fischeri*. This enables the transcription factor–AHL complex to bind to specific promoter sequences in the genome and to induce gene expression resulting in conjugal plasmid transfer, bioluminescence, exopolysaccharide synthesis, biofilm formation, antibiotic synthesis, or virulence (reviewed in [68]). Often, the regulated genes are crucial to the colonization or infection of eukaryotic hosts [48].

In addition to AHLs, other quorum sensing molecules are active in Gram-negative bacteria. For example, besides the production of an AHL, *V. harveyi* also produces AI-2. AI-2 is involved in the regulation of bioluminescence. Production occurs via LuxS, an enzyme involved in the methyl cycle of some Gram-positive and Gram-negative bacteria [56]. Recently, the structure for this universal signal, AI-2, produced by both Gram-negative and Gram-positive bacteria, was predicted to be a cyclic borate diester (furanosyl borate diester) [11]. The notion that the primary role of AI-2 based signaling in other bacteria may not be in cell-to-cell communication has been proposed recently [69]. Nevertheless, the large number and range of genes that appear to be controlled by LuxS would suggest relevance beyond the activated methyl cycle [41]. An alternative is found in the enterohemorrhagic *E. coli* strain serotype O157:H7, where not AI-2 but another autoinducer (AI-3), whose synthesis depends on the presence of LuxS, is involved in bacterial signaling [59].

1.2. Quorum sensing in Gram-positive bacteria

In Gram-positive bacteria, QS induced processes include virulence in *Staphylococcus aureus* [47] and *Enterococcus faecalis* [53], genetic competence in *Streptococcus pneumoniae* [12] and *Bacillus subtilis* [63], and the production of antimicrobial peptides (AMPs) in many lactic acid bacteria (LAB) [32]. The signaling molecules involved consist of small post-translationally processed peptides. The different QS systems differ with respect to the secretion and the sensing of the autoinducer, and the regulatory cascade that results from the perception of the autoinducer [61].

The most common mechanism of QS in Gram-positive bacteria consists, besides the autoinducer, of an ATP-binding-cassette (ABC) transporter for secretion of the peptide and a two-component system for sensing of the autoinducer concentration. Exceptions to this mechanism are the *S. aureus agr*- and the *E. faecalis fsr*-systems, both involved in virulence, which use a non-ABC transporter to secrete the autoinducer, and the *B. subtilis phr*-system, involved in ge-

netic competence and sporulation, in which the peptide, after a Sec-dependent secretion and extracellular processing, is imported into the cytoplasm by an oligopeptide transport system, where it interacts with intracellular receptors [61]. In what follows, only the mechanism containing the ABC-transporter and the two-component system will be discussed briefly.

The autoinducer is continuously produced at low level and secreted by the ABC-transporter. The latter protein is assisted by an accessory protein, which is needed for the secretion, but of which the exact function is still unknown. When a threshold level of autoinducer concentration is reached, the histidine kinase, the membrane-bound sensor part of the two-component system, is auto-phosphorylated. The phosphate is then transferred to the response regulator, a cytoplasmic DNA-binding protein, which subsequently activates transcription of the autoinducer structural gene, the regulatory genes and the transporter genes. Depending on the QS system, additional genes are cotranscribed resulting in the development of competence, the production of antimicrobial compounds or the modification of the autoinducing peptide [32]. During secretion, the autoinducing peptides become activated by cleavage of their leader sequence. The precursor peptides of the autoinducers involved in the competence-development of *S. pneumoniae*, called the competence stimulating peptides (CSPs), and in the AMP-production in many lactic acid bacteria (see below), contain a GG-motif leader sequence that is removed by their dedicated ABC-transporter concomitant with export [25,51].

As is the case in Gram-negative bacteria, the AI-2 molecule is also present in several Gram-positive species (see Section 1.1). Recently, AI-2-like molecules have also been detected in the cell-free culture fluid of several lactic acid bacteria (LAB), including *Lactobacillus rhamnosus* GG [14].

2. Bacteriocins

Bacteriocins are ribosomally synthesized peptides or proteins, produced by bacteria, that have antimicrobial activity and may be important weapons for microbial competition. LAB in particular produce many bacteriocins in a population density dependent manner. These antimicrobial peptides are classified on the basis of their primary structure [27,30]. Class I contains small (<5 kDa), extensively post-translationally modified bacteriocins carrying characteristic lanthionine rings and are called the lantibiotics. The bacteriocins of class II are not modified (except for disulfide bridges), cationic (a common feature of all LAB bacteriocins), heat stable and have a molecular mass less than 10 kDa (with the exception of class IIc members). This class is further subdivided into three subclasses. Class IIa bacteriocins show strong antilisterial activity and are characterized by the occurence of the YGNGVXaaC motif in their N-terminal part [19]. The activity of class IIb bacteriocins depends on the complementary activity of two peptides. Class IIc contains the remaining non-lantibiotic class II bacteriocins. Class III bacteriocins are larger, heat labile proteins, with sizes exceeding 30 kDa. Class I and II AMPs are secreted and processed during secretion, in a way similar to their respective autoinducers. In general, the leader sequence of the lantibiotics' presequence is removed by a specific leader peptidase, which is encoded by a gene that is also autoregulated by the QS system [40]. In contrast, most of the prepeptides belonging to the class II AMPs (and also some lantibiotics), contain a GG-motif leader sequence that is recognized and removed by their cognate ABC-transporter [45,54]. The class II AMP production systems are genetically organized in three operons, containing either the AMP structural gene and the corresponding immunity gene, the processing and secretion genes, or the autoinducer precursor gene with the genes coding for the two-component regulatory system. The autoinducer is generally a secreted bacteriocin-like peptide. The killing action of the class II AMPs appears to be caused mainly by pore formation in the membranes of susceptible organisms [19].

The bacteriocins produced by Gram-negative bacteria (many members of the Enterobacteriaceae family) are divided in colicins, ranging in size from 25 to 80 kDa, and microcins, which have a size less than 10 kDa. The microcins can further be separated into class I microcins, with a size less than 5 kDa and with post-translational modifications, and class II microcins, which range in size between 7 and 10 kDa and possess no modifications [52]. Both classes also differ in their mode of action, namely class I microcins have specific intracellular targets whereas class II peptides interact with bacterial membranes. Besides similarities in amino acid composition and hydrophobicity, many class II microcins and class II bacteriocins produced by LAB are synthesized as precursors containing a GG-motif leader sequence. However, in contrast to class IIa bacteriocins, class II microcins, except Mcc 24, are negatively charged [52]. Also, all class IIa bacteriocins have at least one disulfide bridge which stabilizes the structure of the peptide whereas microcins, except Mcc V, have none.

3. Peptides containing a double-glycine motif

As mentioned above, the competence stimulating peptide of *S. pneumoniae*, most class II (and some class I) bacteriocins produced by LAB as well as their respective autoinducing peptides, and class II microcins, are synthesized as prepeptides containing a GG-motif leader sequence [45,51]. The consensus sequence of this GG-motif was proposed as LSX_2ELX_2IXGG, where X can be any amino acid [25]. Leader peptides of class IIa vary in length between 18 and 27 amino acid residues [19]. It was shown that the presence of a GG-motif containing peptide is clearly related to the presence of a dedicated ABC-transporter, which contains a specific N-terminal extension of about 150 amino acids

that is not found in other ABC-transporters [24]. The remaining part of this dedicated transport protein consists of a central hydrophobic integral domain and a C-terminal cytoplasmic ATPase domain, both domains being common to many other ABC-transporters. In the N-terminal proteolytic domain, two conserved motifs can be distinguished, called the cysteine and the histidine motifs (C/H motifs), with consensus sequences $QX_4(D/E)CX_2AX_3MX_4(Y/F)GX_4(I/L)$ and $H(Y/F)(Y/V)VX_{10}(I/L)XDP$, respectively. These motifs are responsible for recognition and cleavage of the GG-motif containing peptides [24,65].

By using the defined consensus sequences, we previously performed a search to find similar peptides and ABC-transporters in Gram-negative species [43]. A total of 12 peptides, most of them hypothetical, were found. On the basis of multiple alignments of these peptides, the consensus sequence of the GG-motif was extended for Gram-negative bacteria to $M(R/K)ELX_3E(I/L)X_2(I/V)XG(G/A)$. Seven ABC-transporters from Gram-negative species containing the N-terminal extension were retrieved from the search and the consensus sequence of the C/H motifs was adapted to $QX_4(D/E)C(G/A)XAXLX_2(I/V)X_4GX_4(I/L)X_2LR$ and $H(Y/F)(Y/V)V(L/V)X_9(I/L/V)XDP$, respectively [43].

4. Genome wide in silico screening for peptides containing double-glycine motif leader sequences and their cognate transporters

4.1. Overview

So far, searches for GG-motif leader sequences were performed at the amino acid level. In case of a search for small peptides, the major drawback of using protein databases is the dependence on the correct annotation of DNA sequences. Small open reading frames are not always annotated and consequently, the corresponding peptides will not be present in the protein databases. Therefore, a DNA based search strategy is needed.

The Wise2 program (http://www.ebi.ac.uk/Wise2) translates DNA sequences in the six reading frames and compares the translations with a Hidden Markov Model (HMM) [7]. Wise2 was used to screen all fully sequenced bacterial genomes present in the NCBI genome database (situation on 15 September 2003) for GG-motifs and for Peptidase C39 domains. Besides the chromosomes, also plasmids were taken into the search, but only those plasmids that belong to a fully sequenced strain. A list of all screened replicons is given in Table 1. For the Peptidase C39 domain search, the HMM describing this protein family domain was obtained from the Pfam database (http://www.sanger.ac.uk/Software/Pfam; accession number PF03412) [5]. For the peptide search, two HMMs were built, describing the GG-motif based on peptides from either Gram-negative or Gram-positive bacteria. A schematic overview of the screening strategy is shown in Fig. 1.

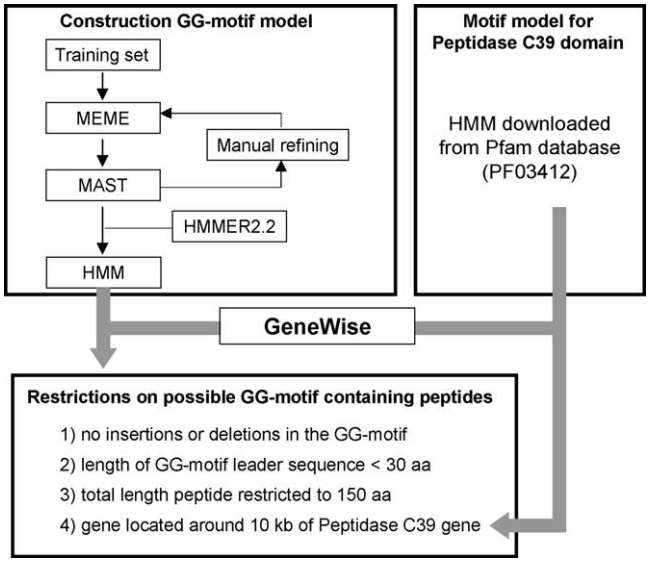


Fig. 1. Schematic overview of the screening strategy used.

4.2. Peptidase C39 domain search

By using the Wise2-software in combination with the HMM describing the Peptidase C39 family (PF03412), all fully sequenced bacterial genomes (Table 1) were screened. The search resulted in 78 candidate Peptidase C39 domain encoding loci, on which a blastx analysis was performed to retrieve the complete protein sequence (http://www.ncbi.nlm.nih.gov/BLAST) [1]. For three hits no correct annotation was found, therefore the putative protein sequences were derived from the corresponding genome sequences. Information about the distribution of the Peptidase C39 hits among the different screened genomes is given in Table 2. Although originally discovered in Gram-positive bacteria, the domain is clearly present in many Gram-negative bacteria. 26 out of 79 Gram-negative bacterial strains (33%) contained one or more Peptidase C39 domain, compared to 20 out of 45 (44%) for Gram-positive bacteria. The genomes of 10 out of the 26 Gram-negative bacteria containing a Peptidase C39 domain (38%), have more than one Peptidase C39 domain protein encoding gene, whereas for Gram-positive bacteria this is 5 out of 20 (25%).

The full sequences of all proteins that resulted from the Peptidase C39 domain search were screened against the Pfam database for the presence of additional protein family domains. Based on the position of the Peptidase C39 domain in the polypeptide, four classes can be distinguished (Fig. 2). Class A represents proteins containing a Peptidase C39 domain (PF03412), an ABC transporter transmembrane region (ABC_membrane: PF00664) and an ATP-binding cassette (ABC_tran: PF00005). The length of class A proteins ranges from 680 to 853 amino acids. Members of class B only contain a Peptidase C39 domain and vary in length from 160 to 270 amino acids. Class C proteins have both the Peptidase C39 and the ABC_tran domain but lack the ABC_membrane

Table 1
List of replicons screened for the presence of Peptidase C39 domains and GG-motifs

	Organism	Acc.Nr. chrom.*	Acc.Nr. plasmids
1	*Agrobacterium tumefaciens str. C58*, (Cereon)	NC_003062, NC_003063	NC_003064 NC_003065
2	*Agrobacterium tumefaciens str. C58*, (U. Washington)	NC_003304, NC_003305	NC_003306 NC_003308
3	*Aquifex aeolicus*	NC_000918	NC_001880
4	*Bacillus anthracis str. A2012*	NC_003995	NC_003980 NC_003981
5	*Bacillus anthracis str. Ames*	NC_003997	
6	*Bacillus cereus*	NC_004722	NC_004721
7	*Bacillus halodurans*	NC_002570	
8	*Bacillus subtilis*	NC_000964	
9	*Bacteroides thetaiotaomicron*	NC_004663	
10	*Bifidobacterium longum*	NC_004307	NC_004943
11	*Bordetella bronchiseptica*	NC_002927	
12	*Bordetella parapertussis*	NC_002928	
13	*Bordetella pertussis*	NC_002929	
14	*Borrelia burgdorferi*	NC_001318	NC_000948, NC_000950, NC_000949, NC_000951, NC_000952, NC_000954, NC_000953, NC_000955, NC_000956, NC_001849, NC_000957, NC_001850, NC_001851, NC_001853, NC_001852, NC_001854, NC_001855, NC_001857, NC_001856, NC_001903 NC_001904
15	*Bradyrhizobium japonicum*	NC_004463	
16	*Brucella melitensis*	NC_003317, NC_003318	
17	*Brucella suis*	NC_004310, NC_004311	
18	*Buchnera aphidicola str. APS*	NC_002528	NC_002252 NC_002253
19	*Buchnera aphidicola str. Bp*	NC_004545	
20	*Buchnera aphidicola str. Sg*	NC_004061	
21	*Campylobacter jejuni*	NC_002163	
22	*Candidatus Blochmannia floridanus*	NC_005061	
23	*Caulobacter crescentus*	NC_002696	
24	*Chlamydia muridarum*	NC_002620	NC_002182
25	*Chlamydia trachomatis*	NC_000117	
26	*Chlamydophila caviae*	NC_003361	NC_004720
27	*Chlamydophila pneumoniae AR39*	NC_002179	
28	*Chlamydophila pneumoniae CWL029*	NC_000922	
29	*Chlamydophila pneumoniae J138*	NC_002491	
30	*Chlamydophila pneumoniae TW-183*	NC_005043	
31	*Chlorobium tepidum*	NC_002932	
32	*Chromobacterium violaceum*	NC_005085	
33	*Clostridium acetobutylicum*	NC_003030	NC_001988
34	*Clostridium perfringens*	NC_003366	
35	*Clostridium tetani*	NC_004557	
36	*Corynebacterium efficiens*	NC_004369	
37	*Corynebacterium glutamicum*	NC_003450	
38	*Coxiella burnetii*	NC_002971	NC_004704
39	*Deinococcus radiodurans*	NC_001263, NC_001264	NC_000958
40	*Enterococcus faecalis*	NC_004668, NC_004670	NC_004671
41	*Escherichia coli CFT073*	NC_004431	
42	*Escherichia coli K12*	NC_000913	
43	*Escherichia coli O157:H7*	NC_002695	NC_002127 NC_002128
44	*Escherichia coli O157:H7 EDL933*	NC_002655	
45	*Fusobacterium nucleatum*	NC_003454	
46	*Haemophilus ducreyi*	NC_002940	
47	*Haemophilus influenzae*	NC_000907	
48	*Helicobacter hepaticus*	NC_004917	
49	*Helicobacter pylori 26695*	NC_000915	
50	*Helicobacter pylori J99*	NC_000921	
51	*Lactobacillus plantarum*	NC_004567	
52	*Lactococcus lactis*	NC_002662	NC_002502, NC_004955 NC_004922, NC_004966
53	*Leptospira interrogans*	NC_004342, NC_004343	
54	*Listeria innocua*	NC_003212	NC_003383
55	*Listeria monocytogenes*	NC_003210	
56	*Mesorhizobium loti*	NC_002678	NC_002679 NC_002682
57	*Mycobacterium bovis*	NC_002945	
58	*Mycobacterium leprae*	NC_002677	

Table 1 (*Continued*)

	Organism	Acc.Nr. chrom.*	Acc.Nr. plasmids
59	*Mycobacterium tuberculosis CDC1551*	NC_002755	
60	*Mycobacterium tuberculosis H37Rv*	NC_000962	
61	*Mycoplasma gallisepticum*	NC_004829	
62	*Mycoplasma genitalium*	NC_000908	
63	*Mycoplasma penetrans*	NC_004432	
64	*Mycoplasma pneumoniae*	NC_000912	
65	*Mycoplasma pulmonis*	NC_002771	
66	*Neisseria meningitidis MC58*	NC_003112	
67	*Neisseria meningitidis Z2491*	NC_003116	
68	*Nitrosomonas europaea*	NC_004757	
69	*Nostoc* sp.	NC_003272	NC_003240, NC_003267, NC_003241, NC_003270, NC_003273 NC_003276
70	*Oceanobacillus iheyensis*	NC_004193	
71	*Pasteurella multocida*	NC_002663	NC_001774, NC_004772 NC_004771
72	*Pirellula* sp.	NC_005027	
73	*Porphyromonas gingivalis*	NC_002950	
74	*Prochlorococcus marinus str. MIT9313*	NC_005071	
75	*Prochlorococcus marinus str. CCMP1375*	NC_005042	
76	*Prochlorococcus marinus str. CCMP1378*	NC_005072	
77	*Pseudomonas aeruginosa*	NC_002516	
78	*Pseudomonas putida*	NC_002947	
79	*Pseudomonas syringae*	NC_004578	NC_004632 NC_004633
80	*Ralstonia solanacearum*	NC_003295	NC_003296 NC_001399
81	*Rickettsia conorii*	NC_003103	
82	*Rickettsia prowazekii*	NC_000963	
83	*Salmonella enterica sv. Typhi*	NC_003198	NC_003384 NC_003385
84	*Salmonella enterica sv. Typhi Ty2*	NC_004631	
85	*Salmonella typhimurium*	NC_003197	NC_003277
86	*Shewanella oneidensis*	NC_004347	NC_004349
87	*Shigella flexneri str. 2457T*	NC_004741	
88	*Shigella flexneri str. 301*	NC_004337	
89	*Sinorhizobium meliloti*	NC_003047	NC_003037 NC_003078 NC_004965
90	*Staphylococcus aureus Mu50*	NC_002758	NC_002774
91	*Staphylococcus aureus MW2*	NC_003923	
92	*Staphylococcus aureus N315*	NC_002745	NC_003140
93	*Staphylococcus epidermidis*	NC_004461	NC_005003, NC_005005, NC_00504, NC_00506, NC_005007 NC_00508
94	*Streptococcus agalactiae 2603V/R*	NC_004116	
95	*Streptococcus agalactiae NEM316*	NC_004368	
96	*Streptococcus mutans*	NC_004350	
97	*Streptococcus pneumoniae R6*	NC_003098	
98	*Streptococcus pneumoniae TIGR4*	NC_003028	
99	*Streptococcus pyogenes M1 GAS*	NC_002737	
100	*Streptococcus pyogenes MGAS315*	NC_004070	
101	*Streptococcus pyogenes MGAS8232*	NC_003485	
102	*Streptococcus pyogenes SSI-1*	NC_004606	
103	*Streptomyces avermitilis*	NC_003155	NC_004719
104	*Streptomyces coelicolor*	NC_003888	NC_003903 NC_003904
105	*Synechococcus* sp.	NC_005070	
106	*Synechocystis* sp.	NC_000911	NC_004967
107	*Thermoanaerobacter tengcongensis*	NC_003869	
108	*Thermosynechococcus elongates*	NC_004113	
109	*Thermotoga maritima*	NC_000853	
110	*Treponema pallidum*	NC_000919	
111	*Tropheryma whipplei TW08/27*	NC_004551	
112	*Tropheryma whipplei str.Twist*	NC_004572	
113	*Ureaplasma urealyticum*	NC_002162	
114	*Vibrio cholerae*	NC_002505, NC_002506	NC_004982
115	*Vibrio parahaemolyticus*	NC_004603, NC_004605	
116	*Vibrio vulnificus*	NC_004459, NC_004460	
117	*Wigglesworthia glossinidia*	NC_004344	
118	*Wolinella succinogenes*	NC_005090	
119	*Xanthomonas axonopodis*	NC_003919	NC_003921 NC_003922

Table 1 (*Continued*)

	Organism	Acc.Nr. chrom.*	Acc.Nr. plasmids	
120	*Xanthomonas campestris*	NC_003902		
121	*Xylella fastidiosa 9a5c*	NC_002488	NC_002489	NC_002490
122	*Xylella fastidiosa Temecula1*	NC_004556	NC_004554	
123	*Yersinia pestis CO92*	NC_003143	NC_003131, NC_003134	NC_003132
124	*Yersinia pestis KIM*	NC_004088	NC_004835, NC_004837, NC_004839	NC_004836, NC_004838

* Accession numbers of the chromosomes.

Table 2
Distribution of the Peptidase C39 hits amongst Gram-negative and Gram-positive bacteria

	Gram−	Gram+	Total
Number of strains screened	79 (64%)	45 (36%)	124
Number of hits	49 (63%)	29 (37%)	78
Number of strains with one or more hits	26 (57%)	20 (43%)	46
Number of strains with more than 1 hit	10 (67%)	5 (33%)	15

Percentages are calculated with respect to Gram-negative and Gram-positive strains or hits.

domain. The six class C members vary in length from 658 to 740 amino acids and five of them belong to the *Mycoplasma* or the *Ureaplasma* genus. Finally the two proteins belonging to class D are both from *Nostoc* sp. They contain the same protein family domains as class A members but have an extra N-terminal fragment of about 300 amino acids. Three hits contained a truncated Peptidase C39 domain and consequently, these hits were not taken into the classification. Interestingly, when we compare the genomic organization, most genes coding for a Peptidase C39 domain have one or more genes in close proximity that code for a membrane fusion protein or for other proteins related to membrane transport indicating that they are likely involved in protein/peptide secretion. A list of all Peptidase C39 domain hits that resulted from our search is given in Table 3.

For every Peptidase C39 domain, the presence of the C/H motifs, involved in GG-motif recognition and peptidase activity [24], was manually checked (for consensus sequences see Section 3). With the exception of Bbr_A1, Bpa_A1, Bpe_A1, Bja_A1, Cvi_A3, Eco_A1, Ecp_A1, Nme_A1, Nmf_A1, Ppu_A1, Rso_A1, Xfa_A2 and Xfb_A2, which lack both conserved motifs, and the three hits truncated in the Peptidase C39 domain, which only contain the cysteine motif, all Peptidase C39 domain hits contain the C/H motifs, suggesting that those domains have peptidase activity. For Nme_B2 and Nmf_B1, which have an identical protein sequence, the cysteine motif was only partially present, but after translation of the region flanking the 5′-end of both genes, the full motif was found, suggesting that both genes are not correctly annotated. Since the presence of both conserved motifs is found in proteins belonging to all different classes, peptidase activity seems to be related to the Peptidase C39 domain regardless of the protein context. The 13 hits lacking the C/H motifs all belong to Gram-negative bacteria and constitute ABC transporters that secrete toxins of the hemolysin family. These hits cluster together in the phylogenetic analysis described below (also see Fig. 3). It was already shown that hemolysins do not contain leader peptides and are not proteolytically processed during export and that their cognate transporters lack the C/H motifs in their N-terminal domain [24,67]. However, the hits Ccr_A1, Efa_B1, Mpn_C1, Spn_A1 and Ssq_A1, also annotated as proteins transporting hemolysin-like toxins, contain the conserved C/H motifs.

A phylogenetic analysis based on the Peptidase C39 domain sequences was performed (Fig. 3). The phylogeny shows a clear distinction between domains from Gram-negative and Gram-positive bacteria. Interestingly, the cyanobacterial domains (Cyano A and Cyano D) cluster out separately from other Gram-negatives. In addition, the different classes of Peptidase C39 domains form distinct clusters, although the phylogenetic analysis was performed on the domain sequences only, regardless of their context in the complete polypeptide. Particularly in Gram-negatives, the class B domains (Gram-B) are visibly distinguishable from the class A domains of the classic ABC transporters (Gram-A). Here, the individual Peptidase C39 domain (class B) might thus be functional in the absence of the rest of the ABC transporter polypeptide. Contrary, in Gram-positive bacteria, the A and B type domains are closely related, forming two clusters (Gram+ AB). Remarkably, all Gram-positive B domain encoding loci are flanked by a separate ABC transporter gene, with the exception of Efa_B1 (data not shown). Apparently,

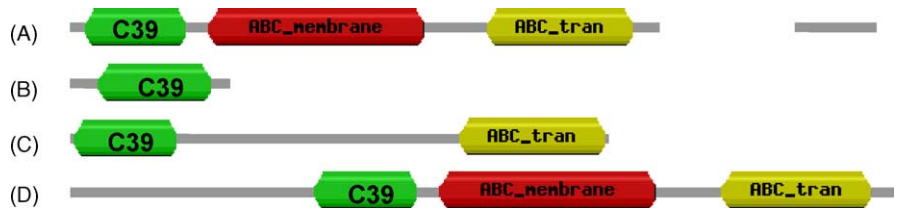


Fig. 2. Schematic representation of the four classes of proteins containing a Peptidase C39 domain (C39; PF03412). The different classes are described in the text. The protein members belonging to the different classes are shown in Table 3. The ABC_membrane domain (PF00664) is an ABC transporter transmembrane region, the ABC_tran domain (PF00005) an ATP-binding cassette. The bar represents the length of 100 amino acids.

Table 3
List of the Peptidase C39 domain containing proteins

ID	Name[a]	Organism	Acc.Nr.[b]	Gene	Length	Description
1	Bha_A1	*Bacillus halodurans*	NC_002570	BH0451	724	Lantibiotic mersacidin transporter
2	Bsu_A1	*Bacillus subtilis*	NC_000964	sunT	705	Lantibiotic sublancin 168 transporter
3	Bth_A1	*Bacteroides thetaiotaomicron*	NC_004663	BT4288	741	ABC transporter
4	Bbr_A1	*Bordetella bronchiseptica*	NC_002927	cyaB	735	Cyclolysin secretion ATP-binding protein
5	Bpa_A1	*Bordetella parapertussis*	NC_002928	cyaB	735	Cyclolysin secretion ATP-binding protein
6	Bpe_A1	*Bordetella pertussis*	NC_002929	cyaB	712	Cyclolysin secretion ATP-binding protein
7	Bja_A1	*Bradyrhizobium japonicum*	NC_004463	bll6293	835	HlyB/MsbA family ABC transporter
8	Cje_B1	*Campylobacter jejuni*	NC_002163	Cj0058	199	Putative periplasmic protein
9	Ccr_A1	*Caulobacter crescentus*	NC_002696	CC0684	726	HlyB/MsbA family ABC transporter
10	Cvi_A1	*Chromobacterium violaceum*	NC_005085	CV0068	727	Colicin V secretion ATP-binding protein
11	Cvi_A2	*Chromobacterium violaceum*	NC_005085	CV4400	706	Colicin V secretion ABC transporter
12	Cvi_A3	*Chromobacterium violaceum*	NC_005085	hlyB	709	Hemolysin B
13	Cvi_B1	*Chromobacterium violaceum*	NC_005085	CV0686	234	Conserved hypothetical protein
14	Cac_A1	*Clostridium acetobutylicum*	NC_001988	CAP0073	731	ABC transporter
15	Efa_A1	*Enterococcus faecalis*	NC_004671	EFB0050	700	Toxin ABC transporter
16	Efa_B1	*Enterococcus faecalis*	NC_004668	EF0528	270	Cytolysin B transporter, truncated
17	Eco_A1	*Escherichia coli CFT073*	NC_004431	hlyB	707	Hemolysin B
18	Eco_A2	*Escherichia coli CFT073*	NC_004431	mchF	704	Probable microcin H47 transporter
19	Ecp_A1	*Escherichia coli O157:H7*	NC_002128	hlyB	706	Hemolysin B
20	Hhe_B1	*Helicobacter hepaticus*	NC_004917	HH0187	228	Conserved hypothetical protein
21	Lpl_A1	*Lactobacillus plantarum*	NC_004567	plnG	716	Bacteriocin ABC transporter
22	Lla_A1	*Lactococcus lactis*	NC_002662	lcnC	715	Lactococcin A ABC transporter
23	Mga_C1	*Mycoplasma gallisepticum*	NC_004829	sunT	667	Bacteriocin/lantibiotic ABC exporter
24	Mge_C1	*Mycoplasma genitalium*	NC_000908	MG390	660	ABC transporter
25	Mpe_C1	*Mycoplasma penetrans*	NC_004432	MYPE7470	691	ABC transporter
26	Mpn_C1	*Mycoplasma pneumoniae*	NC_000912	lcnDR3	660	Hemolysin ABC exporter
27	Mpu_A1	*Mycoplasma pulmonis*	NC_002771	MYPU_3760	680	ABC transporter
28	Nme_A1	*Neisseria meningitidis MC58*	NC_003112	NMB1400	742	ABC transporter
29	Nme_B1	*Neisseria meningitidis MC58*	NC_003112	NMB0098	165	Putative ABC transporter, truncated
30	Nme_B2	*Neisseria meningitidis MC58*	NC_003112	NMB0103	164	Putative bacteriocin resistance protein
31	Nme_B3	*Neisseria meningitidis MC58*	NC_003112	NMB0582	180	Putative bacteriocin resistance protein
32	Nme_B4	*Neisseria meningitidis MC58*	NC_003112	NMB0855	218	Putative bacteriocin resistance protein
33	Nmf_A1	*Neisseria meningitidis Z2491*	NC_003116	NMA1620	742	Putative cytolysin transporter
34	Nmf_B1	*Neisseria meningitidis Z2491*	NC_003116	NMA0173	164	Hypothetical protein
35	Nmf_B2	*Neisseria meningitidis Z2491*	NC_003116	NMA1066	218	Putative periplasmic protein
36	–	*Neisseria meningitidis Z2491*	NC_003116	#	63	Truncated protein
37	Neu_B1	*Nitrosomonas europaea*	NC_004757	NE0789	231	Conserved hypothetical protein
38	Nsp_A1	*Nostoc* sp.	NC_003272	all2021	712	ABC transporter
39	Nsp_A2	*Nostoc* sp.	NC_003272	alr1201	722	ABC transporter
40	Nsp_A3	*Nostoc* sp.	NC_003272	alr5147	716	ABC transporter
41	Nsp_A4	*Nostoc* sp.	NC_003276	alr7014	714	ABC transporter
42	Nsp_A5	*Nostoc* sp.	NC_003276	alr7295	736	ABC transporter
43	Nsp_D1	*Nostoc* sp.	NC_003272	alr1927	1011	ABC transporter
44	Nsp_D2	*Nostoc* sp.	NC_003272	hetC	1044	Heterocyst differentiation protein
45	Pma_A1	*Prochlorococcus marinus*	NC_005071	PMT0978	738	Multidrug efflux family ABC transporter
46	Pae_A1	*Pseudomonas aeruginosa*	NC_002516	PA4143	719	Probable toxin transporter
47	Pae_B1	*Pseudomonas aeruginosa*	NC_002516	PA1762	253	Hypothetical protein
48	Pae_B2	*Pseudomonas aeruginosa*	NC_002516	PA1953	226	Hypothetical protein
49	Ppu_A1	*Pseudomonas putida*	NC_002947	PP4927	731	HlyB family ABC transporter
50	Ppu_B1	*Pseudomonas putida*	NC_002947	PP2855	226	Conserved hypothetical protein
51	Rso_A1	*Ralstonia solanacearum*	NC_003296	cyaB	725	Probable cyclolysin-type ABC transporter
52	Rso_B1	*Ralstonia solanacearum*	NC_003295	RSc0445	234	Hypothetical signal peptide protein
53	Son_A1	*Shewanella oneidensis*	NC_004349	SOA0049	725	Putative toxin secretion ABC transporter
54	Smu_A1	*Streptococcus mutans*	NC_004350	SMU.1881c	763	Putative ABC transporter
55	Smu_A2	*Streptococcus mutans*	NC_004350	SMU.286	760	Putative ABC transporter
56	Smu_B1	*Streptococcus mutans*	NC_004350	SMU.1897	160	Putative ABC transporter
57	Spn_A1	*Streptococcus pneumoniae R6*	NC_003098	clyB	702	Hemolysin ABC transporter
58	Spn_A2	*Streptococcus pneumoniae R6*	NC_003098	comA	717	ABC transporter ComA
59	Spn_B1	*Streptococcus pneumoniae R6*	NC_003098	spr0469	197	Conserved hypothetical protein, truncation
60	–	*Streptococcus pneumoniae R6*	NC_003098	spr0105	77	Truncated protein
61	Spo_A1	*Streptococcus pneumoniae TIGR4*	NC_003028	SP0042	717	Competence factor transporter ComA
62	Spo_A2	*Streptococcus pneumoniae TIGR4*	NC_003028	SP1953	698	Toxin secretion ABC transporter
63	Spo_B1	*Streptococcus pneumoniae TIGR4*	NC_003028	SP0530	195	Putative ABC transporter, truncated

Table 3 (*Continued*)

ID	Name[a]	Organism	Acc.Nr.[b]	Gene	Length	Description
64	Spy_B1	*Streptococcus pyogenes MGAS315*	NC_004070	SpyM3_1650	214	Putative salivaricin A modification enzyme
65	Spz_A1	*Streptococcus pyogenes MGAS8232*	NC_003485	spyM18_0543	717	Putative ABC transporter
66	–	*Streptococcus pyogenes MGAS8232*	NC_003485	spyM18_0524	78	Truncated protein
67	Spa_B1	*Streptococcus pyogenes SSI-1*	NC_004606	#	222	Similar to SalT of S. Salivarius, truncated
68	Sav_A1	*Streptomyces avermitilis*	NC_003155	SAV7493	749	Putative ABC transporter
69	Sco_A1	*Streptomyces coelicolor*	NC_003888	SCO0755	740	Putative ABC transporter
70	Ssp_C1	*Synechococcus* sp.	NC_005070	SYNW2111	740	Putative ABC transporter
71	Ssq_A1	*Synechocystis* sp.	NC_000911	hlyB	733	Hemolysin ABC transporter
72	Uur_C1	*Ureaplasma urealyticum*	NC_002162	ABC-2	658	ABC transporter
73	Xfa_A1	*Xylella fastidiosa 9a5c*	NC_002488	XF1220	707	Colicin V secretion ABC transporter
74	Xfa_A2	*Xylella fastidiosa 9a5c*	NC_002488	XF2397	720	Toxin secretion ABC transporter
75	Xfb_A1	*Xylella fastidiosa Temecula1*	NC_004556	cvaB	707	Colicin V secretion ABC transporter
76	Xfb_A2	*Xylella fastidiosa Temecula1*	NC_004556	hlyB	720	Toxin secretion ABC transporter
77	Ype_A1[c]	*Yersinia pestis CO92*	NC_003143	#	301	Putative secretion ATPase, truncated
78	Ypf_A1	*Yersinia pestis KIM*	NC_004088	y0286	704	Putative ABC transporter

Not annotated.

[a] Consists of a three-letter abbreviation of the species followed by the class to which the protein belongs and a number to differentiate between multiple transporters within the same strain.

[b] Accession number of the genome; /protein containing a truncated Peptidase C39 domain.

[c] Protein containing a truncated ABC_membrane domain (PF:000664).

Gram-positive class B domains do not exist in a genetic context independent of an ABC transporter although they are encoded by a separate gene. The aberrant class C Peptidase C39 domains (Gram+ C), with exception of Scp_C1, evolved independently of other Gram-positive domains. Summarizing, a clear distinction of Peptidase C39 domains can be made in function of bacterial lineages and/or protein context, suggesting a common (and more ancient) origin. This is supported by the presence of Peptidase C39 domains in some archeae. However, horizontal transfer as a mechanism for obtaining Peptidase C39 domains cannot be ruled out, illustrated by the separation of the two Gram-A clusters by some cyanobacterial and Gram-positive domains. Besides plasmid transfer (7 out of the 78 hits are plasmid encoded), horizontal gene transfer could also be mediated through bacteriophages. For example, the bacteriophage SPBc2 contains a Peptidase C39 domain encoding gene, *sunT*.

4.3. Search for genes encoding double-glycine motif containing peptides

For the GG-motif search, two HMMs were built, based on peptides from either Gram-negative or Gram-positive bacteria. A Gram-positive training set was obtained from 31 already known GG-motif containing peptides. Since only a limited number of GG-motif sequences from Gram-negative bacteria was available [43], additional possible GG-motif containing peptides were first obtained via an iterative MEME/MAST approach.

Based on the 12 possible GG-motif containing peptides from Gram-negative bacteria obtained from a previous search [43], a probabilistic model of the GG-motif was constructed using MEME (Multiple EM for Motif Elicitation; http://meme.sdsc.edu/meme/website/meme.html) [3]. Only the N-terminal leader sequences of those peptides were taken as input for MEME. The MEME output (i.e. a motif model describing the GG-motif) was used to screen all peptides from the non-redundant database from NCBI using MAST (Motif Alignment and Search Tool; http://meme.sdsc.edu/meme/website/mast.html) [4]. Hits with an *E*-value less than 10 were retained for further analysis. Subsequently, hits were manually refined by retaining only those peptides from Gram-negative bacteria that did not have any gap or insertion in their GG-motif and that ended with the amino acid pairs Gly–Gly or Gly–Ala (since all GG-motifs described so far end with such a pair). The curated set of peptides obtained was used as a training set for a new MEME/MAST iteration using the same parameter settings as described above. The resulting output was manipulated in a way similar to the first iteration. Based on the second MEME/MAST cycle output, a curated training set of 38 possible GG-motifs from Gram-negative bacteria was obtained. Since most members of the training set had a length of 15 amino acids, only the 15 C-terminal amino acids from members containing longer sequences were taken into the training set (Fig. 4).

The Gram-positive and Gram-negative training sets were used to build HMMs by using the HMMER2.2 software (http://www.hmmer.wustl.edu) [18]. Using the Wise2 program, the GG-motif search was performed with both HMMs on the set of DNA sequences given in Table 1.

Because both HMMs are built on sequences of only 15 amino acids, a number of restrictions were introduced, based on the current knowledge of already known GG-motif containing peptides. Obviously, by using these restrictions we cannot exclude that some GG-motif peptides are not recovered in this search. In a first step, no insertions or gaps were allowed in the GG-motif and the motif had to end with a Gly–Gly or Gly–Ala pair. Secondly, only those peptides were selected from which the coding region was located less than 10 kb from the coding region of a Peptidase C39 domain. This restriction is based on the observation that in many GG-

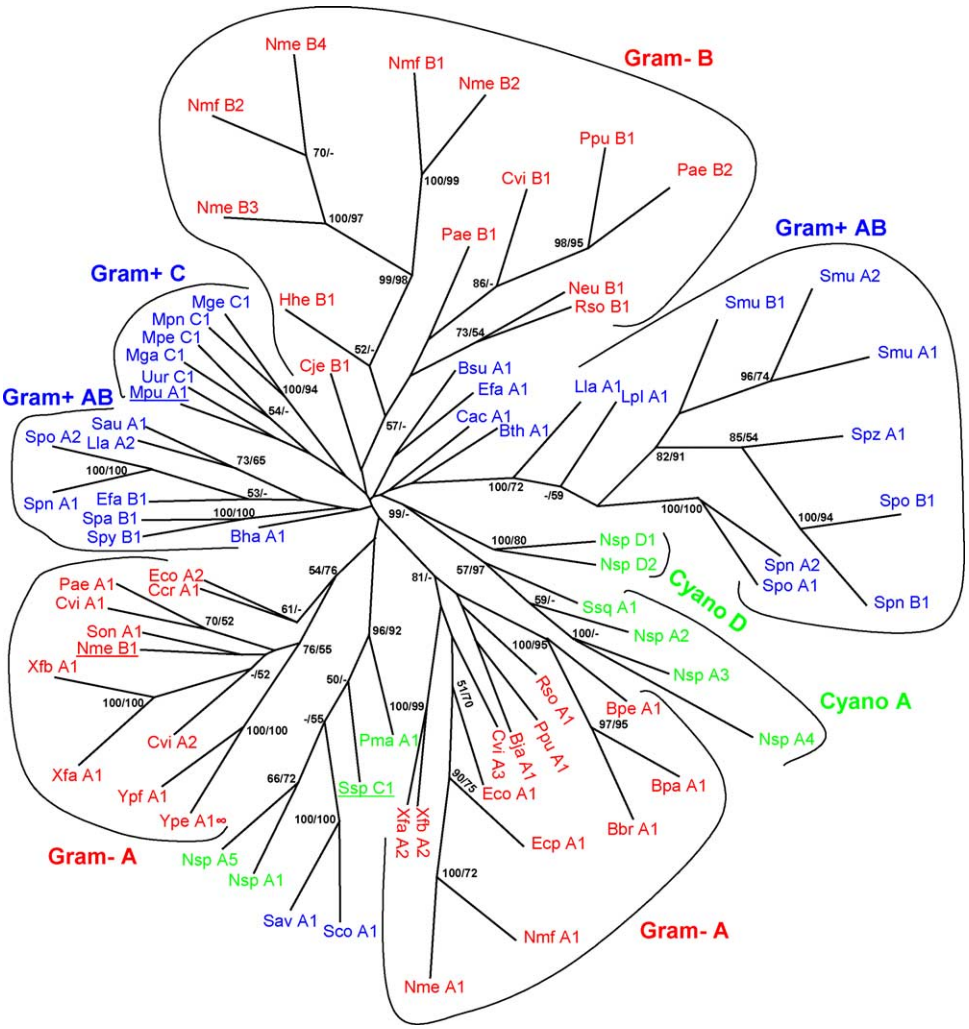


Fig. 3. Phylogenetic analysis of the Peptidase C39 domain. All polypeptides (Table 3), except those with a truncated C39 domain, were aligned to the Peptidase C39 protein family using hmmalign of the HMMER software package [18]. The tree shown here is the most parsimonious tree, obtained from 1000 bootstrapped samples (Phylip package 3.6; http://evolution.genetics.washington.edu/phylip.html). In addition, neighbour-joining trees were constructed from 1000 bootstrapped samples (Treecon for Windows 1.3b) [66]. Distances were calculated with the Poisson Correction and Kimura method, ignoring insertions and deletions. Similar clustering patterns were observed for all neighbour-joining trees (not shown). Bootstrap values above 50% are given (most parsimonious/neighbour-joining). Domains from Gram-negative bacteria are shown in red, those from Gram-positives in blue. As the cyanobacterial domains clustered out differently from the rest of the Gram-negatives, they are shown in green. Underlined domains belong to a different class than the rest of the cluster.

motif peptide systems, the structural gene is clustered with the genes coding for the secretion, the processing and the sensing machinery (in Gram-positive as well in Gram-negative bacteria) [32,43]. As a third restriction, the length of the leader sequence and the total peptide length was set to a maximum of 30 and 150 amino acids respectively. Therefore, for every GG-motif hit sequence an open reading frame was built, starting with one of the three possible start codons ATG, GTG or TTG. Finally the remaining hits were blasted against the non-redundant database using blastx and if a perfect match with a non-hypothetical protein (different from an already known GG-motif containing peptide) was found, the hit was removed. In what follows, only peptides belonging to Gram-negative bacteria will be presented and discussed.

Initially, the search was performed with the HMM derived from the Gram-negative training set. Taking into account the restrictions described above, 38 possible GG-motif containing peptides were retrieved. The search with the HMM derived from the Gram-positive training set resulted in 20 additional candidate peptides, so 58 possible GG-motif containing peptides from Gram-negative bacteria were found. The size of the peptides was 23–142 amino acids, or in the mature form (i.e. without the leader peptide) 5–124 amino acids. From all candidate peptides, the mature part was analyzed with the ProtParam tool (http://us.expasy.org/tools/protparam.html) to calculate various physical and chemical parameters. A list of the possible GG-motif containing peptides in Gram-negative bacteria, their cognate transport protein, the amino acid context, the theoretical p*I* and the molecular weight of the peptides is given in Table 4. Most peptides have a high glycine content (more than 10% glycine for 60% of the peptides) and an el-

Table 4
List of the possible GG-motif containing peptides from Gram-negative bacteria

ID	Organism	Acc. Nr.	DNA start[a]	DNA stop[a]	Gene	Transporter[b]	Length[c]		Amino acid composition[d]						p*I*	M_W[e]
							GG	Total	Arom.	Charged	Polar	Hydroph.	Cys	Gly		
1	*Caulobacter crescentus*	NC_002696	747957	747913	#	Ccr_A1	15	84	7.1	10	12.9	76.7	0	24.6	6.48	6243
2	*Caulobacter crescentus*	NC_002696	748227	748183	#	Ccr_A1	19	83	6.3	9.5	7.9	82.8	0	25	8.75	5800.7
3	*Caulobacter crescentus*	NC_002696	754303	754347	CC0688	Ccr_A1	19	142	7.2	13	14.5	72.3	0	14.6	9.81	12461.7
4	*Caulobacter crescentus*	NC_002696	757113	757066	#	Ccr_A1	17	28	0	36.4	45.5	18.2	1	0	9.5	1223.3
5	*Caulobacter crescentus*	NC_002696	758792	758839	CC0692	Ccr_A1	20	50	6.6	16.5	20	63.3	0	3.3	6.75	3234.8
6	*Chromobacterium violaceum*	NC_005085	81944	81991	CV0071	Cvi_A1	27	50	8.7	12.9	56.4	30.3	2	13	8.03	2476.6
7	*Chromobacterium violaceum*	NC_005085	4746267	4746220	CV4401	Cvi_A2	17	68	7.9	7.8	17.7	74.6	2	33.3	6.12	4640.2
8	*Chromobacterium violaceum*	NC_005085	709116	709163	CV0683	Cvi_B1	30	109	1.3	8.9	27.8	63.3	0	17.7	6.43	7678.7
9	*Chromobacterium violaceum*	NC_005085	709170	709217	CV0683	Cvi_B1	23	84	1.6	9.7	26.2	64	0	18	9.49	5921.7
10	*Escherichia coli CFT073*	NC_004431	1176822	1176866	mchB	Eco_A2	15	75	0	0	23.4	76.7	0	26.7	5.52	4864.3
11	*Escherichia coli CFT073*	NC_004431	1183119	1183163	#	Eco_A2	15	92	7.8	9.1	36.4	54.6	0	19.5	4.66	7282.8
12	*Neisseria meningitidis MC58*	NC_003112	98911	98958	NMB0086	Nme_B1	21	83	8.1	35.6	33.8	30.6	2	3.2	6.45	7411.2
13	*Neisseria meningitidis MC58*	NC_003112	104406	104450	NMB0091	Nme_B1	15	78	15.9	17.5	28.6	54.1	0	17.5	10	6712.6
14	*Neisseria meningitidis MC58*	NC_003112	610925	610972	#	Nme_B3	15	52	10.8	24.3	29.7	45.9	1	5.4	10.05	4076.6
15	*Neisseria meningitidis MC58*	NC_003112	883878	883834	NMB0861	Nme_B4	17	141	13.6	22.5	21.7	55.6	0	16.9	11.44	13281.2
16	*Neisseria meningitidis MC58*	NC_003112	885348	885304	NMB0865	Nme_B4	17	117	8	12	27	61	0	16	12.31	9932.4
17	*Neisseria meningitidis Z2491*	NC_003116	154396	154349	NMA0169	Nmf_B1	23	101	15.4	30.7	11.6	57.7	1	11.5	5.89	8397.3
18	*Neisseria meningitidis Z2491*	NC_003116	1027673	1027629	NMA1073	Nmf_B2	17	141	13.6	22.5	21.7	55.6	0	16.9	11.44	13281.2
19	*Neisseria meningitidis Z2491*	NC_003116	1029700	1029656	NMA1078	Nmf_B2	17	100	8.4	15.6	28.8	55.3	0	18.1	12.4	8657.9
20	*Nitrosomonas europaea*	NC_004757	859268	859321	NE0790	Neu_B1	19	34	13.3	13.4	20	66.7	0	6.7	6	1689.9
21	*Nitrosomonas europaea*	NC_004757	851653	851606	NE0784	Neu_B1	24	37	15.4	0	30.8	69.3	0	0	5.57	1404.6
22	*Nostoc* sp.	NC_003272	2422705	2422661	asl2024	Nsp_A1	23	37	7.1	28.5	21.4	49.9	0	7.1	9.99	1548.8
23	*Nostoc* sp.	NC_003272	1418171	1418215	asr1202	Nsp_A2	21	58	5.4	35.1	29.7	35.1	0	13.5	8.58	4142.5
24	*Nostoc* sp.	NC_003272	1418704	1418751	asr1203	Nsp_A2	25	95	7.1	18.5	34.2	47.2	1	30	9.77	7289.8
25	*Nostoc* sp.	NC_003272	1419515	1419559	asr1204	Nsp_A2	25	68	34.9	25.6	23.3	51.2	0	32.6	12.6	5198.8
26	*Nostoc* sp.	NC_003272	6136580	6136624	asr5139	Nsp_A3	20	80	10	13.3	36.6	50	0	11.7	4.02	6120.6
27	*Nostoc* sp.	NC_003272	6136990	6137034	alr5140	Nsp_A3	20	78	6.9	12	36.2	51.6	0	10.3	4.11	5835.3
28	*Nostoc* sp.	NC_003272	6137394	6137438	alr5141	Nsp_A3	20	78	6.9	12.1	34.5	53.5	0	12.1	3.71	5765.2
29	*Nostoc* sp.	NC_003272	6137793	6137837	asr5142	Nsp_A3	20	80	10	16.6	38.4	45	0	10	4.53	6443
30	*Nostoc* sp.	NC_003272	6138151	6138195	alr5143	Nsp_A3	27	91	15.6	15.6	49.9	34.4	0	1.6	4.14	7268.9
31	*Nostoc* sp.	NC_003272	6139056	6139100	asr5144	Nsp_A3	20	78	22.3	29.3	24	46.4	0	12.1	7.96	6684.4
32	*Nostoc* sp.	NC_003272	6139788	6139832	asr5145	Nsp_A3	24	85	16.4	19.7	27.9	52.4	0	1.6	4.6	6999.9
33	*Nostoc* sp.	NC_003272	6140355	6140399	asr5146	Nsp_A3	23	84	22.8	22.9	32.7	44.3	0	6.6	4.7	7225
34	*Nostoc* sp.	NC_003276	13799	13846	#	Nsp_A4	15	31	18.6	24.9	43.7	31.2	0	0	8.49	1990.2
35	*Prochlorococcus marinus*	NC_005071	1048057	1048104	#	Pma_A1	15	23	0	12.5	50	37.5	0	12.5	9.75	828.8
36	*Prochlorococcus marinus*	NC_005071	1059298	1059348	#	Pma_A1	22	27	0	20	40	40	0	0	9.75	572.6
37	*Pseudomonas aeruginosa*	NC_002516	4629958	4630002	PA4139	Pae_A1	20	99	12.7	17.8	25.4	57	0	13.9	4.59	8259.3
38	*Pseudomonas aeruginosa*	NC_002516	4632476	4632520	PA4141	Pae_A1	15	99	8.4	8.4	20.3	71.5	0	26.2	4.78	8079.2
39	*Pseudomonas aeruginosa*	NC_002516	4632632	4632679	PA4141	Pae_A1	16	48	12.4	15.5	31.1	53	0	18.8	4.68	3404.7
40	*Pseudomonas aeruginosa*	NC_002516	1912484	1912528	PA1768	Pae_B1	24	88	11	21.9	12.5	65.7	0	18.8	4.21	6575.2
41	*Pseudomonas aeruginosa*	NC_002516	2137508	2137555	PA1951	Pae_B2	15	49	14.7	20.5	47.1	32.3	5	8.8	9.59	3938.5
42	*Pseudomonas aeruginosa*	NC_002516	2141561	2141608	#	Pae_B2	21	77	5.4	19.7	25.1	55.3	2	10.7	9.68	5657.3
43	*Ralstonia solanacearum*	NC_003295	472941	472988	RSc0444	Rso_B1	19	87	4.4	8.8	38.3	52.9	0	7.4	5.71	7054.8
44	*Shewanella oneidensis*	NC_004349	38530	38486	#	Son_A1	15	67	11.5	3.8	19.1	77	2	26.9	5.82	4702.3
45	*Shewanella oneidensis*	NC_004349	46897	46941	SOA0057	Son_A1	25	60	14.3	20	25.7	54.3	0	5.7	11.54	4095.8
46	*Synechocystis* sp.	NC_000911	3449398	3449354	#	Ssq_A1	29	39	20	10	30	60	0	0	8.47	1184.4
47	*Xylella fastidiosa 9a5c*	NC_002488	1166571	1166615	XF1217	Xfa_A1	15	79	14.1	11.1	34.3	54.6	0	10.9	9.52	6981.9
48	*Xylella fastidiosa 9a5c*	NC_002488	1166975	1167019	XF1218	Xfa_A1	15	96	13.6	7.3	38.2	54.3	0	13.6	8.5	8465.5
49	*Xylella fastidiosa 9a5c*	NC_002488	1167500	1167544	XF1219	Xfa_A1	15	79	20.4	12.5	40.6	47	0	3.1	8.44	7327.2
50	*Xylella fastidiosa 9a5c*	NC_002488	1167885	1167929	#	Xfa_A1	15	66	4	6	25.6	68.6	3	7.8	4.35	4944.9
51	*Xylella fastidiosa 9a5c*	NC_002488	2276882	2276838	XF2402	Xfa_A2	15	27	8.3	41.5	16.6	41.7	0	0	6.76	1450.7
52	*Xylella fastidiosa Temecula1*	NC_004556	593159	593203	PD0497	Xfb_A1	15	79	12.5	9.5	37.4	53.1	0	10.9	9.52	7003
53	*Xylella fastidiosa Temecula1*	NC_004556	593561	593605	PD0498	Xfb_A1	15	96	13.6	8.7	37	54.3	0	13.6	9.52	8436.5
54	*Xylella fastidiosa Temecula1*	NC_004556	594077	594121	#	Xfb_A1	15	77	20.9	12.8	35.5	51.6	0	4.8	8.44	6915.7
55	*Xylella fastidiosa Temecula1*	NC_004556	594460	594504	#	Xfb_A1	15	66	2	6	27.5	66.6	4	7.8	4.35	4942.9
56	*Yersinia pestis CO92*	NC_003143	105395	105442	#	Ype_A1[d]	22	45	26	47.7	30.4	21.7	0	0	10.17	2787.1
57	*Yersinia pestis CO92*	NC_003143	109863	109907	YPO0100	Ype_A1[d]	15	79	7.8	12.5	20.4	67.2	0	20.3	6.16	6154.9
58	*Yersinia pestis KIM*	NC_004088	309522	309566	y0288	Ypf_A1	15	79	7.8	12.5	20.4	67.2	0	20.3	6.16	6154.9

Arom., aromatic residues (Phe, His, Try, Tyr); Charged: (Arg, Asp, Glu, Lys, His); Polar: (Asn, Cys, Gln, Ser, Thr, Trp, Tyr); Hydroph.: hydrophobic (Ala, Phe, Gly, Ile, Leu, Val, Met, Pro); # not annotated.

[a] The position of the GG-motif coding region on the corresponding replicon is given.

[b] The name of the transport protein to which the possible peptide is genetically linked (<10 kb).

[c] Length of the GG-leader sequence (GG) and the total peptide (Total) in amino acids.

[d] The amino acid composition of the mature peptide is given; all values are percentages with the exception of the Cys-column, where the number of cysteine residues is given.

[e] Molecular weight in Dalton.

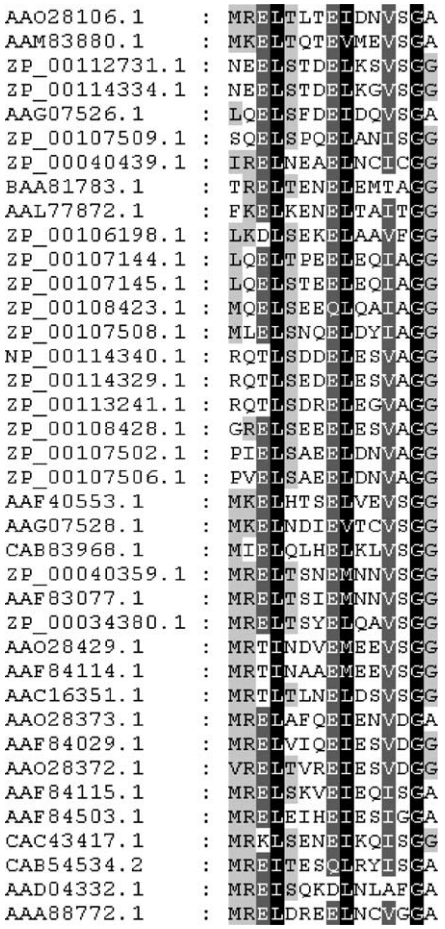


Fig. 4. Multiple alignment of the 38 possible GG-motif containing peptides from Gram-negative bacteria obtained from the MEME/MAST searches. Of those peptides, only the amino acids belonging to the GG-leader sequence are shown. Identical or similar amino acids that are found in all sequences are shaded in black. Dark grey shades are used for amino acids that are present in at least 80% of the sequences, light grey shades when they are present in more than 60% of the sequences.

evated composition of hydrophobic residus. Also, most of the possible peptides have no or only one cysteine residue. For 33% of the peptides, the theoretical p*I* was below 6, for 50% the p*I* was higher than 8. As mentioned in Section 4.2, 13 from the 49 Peptidase C39 domain containing proteins from Gram-negative bacteria do not have the conserved C/H motifs, and as a consequence, no GG-motif coding genes should be found next to their coding region. For all transporters this is the case, except for Xfa_A2, for which one GG-motif hit was retrieved. This could be a false positive hit. Concerning the 36 remaining transporters, 27 of them have at least one possible GG-motif containing peptide. For the nine other transporters, no possible GG-motif hits were observed. This could be due to the restrictions that were applied in this search.

To conclude, many possible GG-motif containing peptides that resulted from our search in the genomes of Gram-negative bacteria, show structural similarity with bacteriocins and peptide hormones.

5. Can peptides be involved in signal transduction in Gram-negative bacteria?

The population density dependent regulation of the production of class I and class II AMPs seems to occur through different mechanisms. Autoregulation of class I bacteriocins, such as nisin from *Lactococcus lactis* or subtilin from *B. subtilis*, involves the bacteriocins themselves, whereas class II bacteriocins may have a dedicated bacteriocin-like autoinducer. This autoinducer is a small cationic peptide either with or without antimicrobial activity, and synthesized with a GG-motif leader peptide, as the class II AMP precursors. As it is the case for the class I bacteriocin nisin, expression of some class II AMPs such as the carnobacteriocin B2, is controlled by the bacteriocins themselves. Besides this functional resemblance, autoinducers and AMPs are also structurally related. Class II bacteriocins are usually between 30 and 60 amino acids in length, have a net positive charge between 3 and 6 at pH 7.0 with isoelectric points above 8.5. Furthermore, they have a high content of non-polar amino acids (>30%) and small amino acids such as glycine, suggested to provide a high degree of conformational freedom to the peptide [19,27]. Similarly, peptide pheromones are cationic with a net charge between 3 and 6 at neutral pH with a p$I > 9$, are between 19 and 26 amino acid residues in length and have a high percentage of hydrophobic residues [31].

In our screening, the peptides identified constitute good candidates for bacteriocins and/or peptide pheromones. All peptides possess a GG-motif leader sequence and their corresponding genes are found in close proximity to a dedicated transporter. In addition, also at the structural level the peptides have physical characteristics (length, amino acid composition, p*I*, see Section 4.3) which correspond either to bacteriocins such as class II AMPs or microcins, or to peptide pheromones. Both of these types of peptides have been shown to play a role in QS in Gram-positive bacteria.

In case that peptides possess a signaling function in Gram-negative bacteria, the question can be raised as how these molecules are sensed by the cells. In general, being surrounded by two concentric lipid bilayer membranes, these peptides should either be sensed at the outer membrane, the periplasm or the inner membrane, or should be active intracellularly.

In Gram-positive bacteria, sensing at the cytoplasmic membrane generally occurs through a two-component signal-transduction system consisting of a membrane-located sensor histidine kinase and a cytoplasmic response regulator. When phosphorylated, the latter protein acts as a transcriptional regulator, altering the expression of target genes, among whom for example the gene coding for the autoinducing peptide. Two-component regulators control competence in *S. pneumoniae* (ComD/ComE) and *B. subtilis* (ComP/ComA), virulence in *S. aureus* (AgrC/AgrA) and the production of class II AMPs in lactic acid bacteria. [17,44]. These two-component regulators belong to at least two different subfamilies, the ComD/ComE subfamily, to which for example the *B. sub-*

tilis ComP/ComA and *S. aureus* AgrC/AgrA systems belong, and the EnvZ/OmpR subfamily, containing the CvnK/CvnR of *Carnobacterium divergens*, NisK/NisR of *L. lactis* and SpaK/SpaR of *B. subtilis* systems. Two-component regulatory systems are ubiquitous in most bacterial species and are important for the bacteria to rapidly generate the appropriate physiological response to changing environmental conditions [60]. It is therefore not unlikely that also Gram-negative bacteria could possess two-component regulators able to detect peptides. For example, in several pathogenic bacteria, such as *Salmonella typhimurium* and *P. aeruginosa*, cationic antimicrobial peptides can activate the expression of resistance genes [2,42]. In *S. typhimurium*, it was suggested that the two-component regulatory system PhoP–PhoQ could directly sense the presence of these peptides [2]. In addition, also other systems could be operative. For example, more recently, a new peptide detection system controlling the production of cytolysin was reported in *E. faecalis* [23]. The products of the *cylR1* and *cylR2* genes are predicted to be transmembrane and cytoplasmic proteins respectively. Binding of one of the cytolysin subunits to the CylR1–CylR2 complex relieves DNA-repression and results in increased transcription of the cytolysin biosynthetic genes, resulting in autoinduction. It was also recently proposed that the PTS permease of the mannose family in *Listeria monocytogenes* might function as a receptor for class IIa bacteriocins [13]. PTS systems have also been reported in many Gram-negative bacteria.

Intracellular functioning peptides need to pass the cytoplasmic membrane to interact with their targets. In *B. subtilis*, unmodified pentapeptides (the Phr peptides) such as CSF inhibit the activity of phosphatases and promote competence development and sporulation. The control of plasmid transfer in *E. faecalis* is exerted through small, unmodified mating pheromones consisting of seven to eight amino acids [35]. In both Gram-positive species, the peptides are transported into the cell by oligopeptide permeases (Opp) [36,58,62]. In general, Opp systems consist of five subunits: a peptide binding protein [62], which is anchored to the cytoplasmic membrane in Gram-positive bacteria or is periplasmic in Gram-negative bacteria and the translocator component consisting of two transmembrane subunits (OppB and OppC) and two intracellular membrane-associated ATPases (OppD and OppE) [16,26]. Affinity and specificity of the periplasmic binding proteins greatly differs. For instance, the receptor protein OppA from *S. typhimurium* binds peptides of two to five amino acid residues independently of their sequence with dissociation constants in the micromolar range [62]. The Opp system of *L. lactis* transports peptides from four to at least 18 residues with relatively little selectivity [15]. These transporters import peptides mainly for nutritional purposes. In contrast, pheromone binding to the receptor proteins in *E. faecalis* is much more specific. For example, the cCF10 mating pheromone of the conjugative plasmid pCF10 is likely detected at picomolar concentrations by the PrgZ protein, which is similar to OppA [36] and subsequently internalized by the OppB and OppC proteins. Interestingly, cCF10 can also be bound by *E. faecalis* OppA, albeit at higher concentrations of the pheromone and in competition with unrelated peptides. As Opp systems are ubiquitous in Gram-negative bacteria, it is not unlikely that short or even relatively long peptides could be imported in the cytoplasm by this system. The concentration limit of the peptide to be detected would depend on the affinity of the receptor protein. It is interesting to note that in *E. coli*, the dipeptide binding protein DppA, possessing strong structural similarity to OppA, besides being part of the dipeptide permease system, is also involved in taxis towards peptides by communicating with methyl-accepting chemotaxis proteins [38]. Finally, it is not unlikely that, at certain concentrations, signaling peptides may enter the cell by diffusion across the cytoplasmic membrane. For example, amphipathic α-helical AMPs are short, cationic and known to interact with and insert into membranes of both Gram-positive and Gram-negative bacteria [64]. At high concentrations, these peptides lead to membrane permeabilization and subsequently death of the bacteria. It is, however, possible that at low concentrations, the peptides insert into the membranes and are released into the cytoplasm of the bacteria where they could fulfil a signaling function. In this respect, it is interesting to note that *P. aeruginosa* is freely permeable to C_4-HSL and *V. fischeri* or *E. coli* to 3O,C_6-HSL [29,49]. However, the efflux of the longer 3O,C_{12}-HSL in *P. aeruginosa* depends on the presence of an active MexAB-OprM pump, containing inner and outer membrane components as well as a membrane fusion protein. It is suggested that this mechanism is similar to the transport of amphipatic antibiotics [49].

In contrast to Gram-positive bacteria, Gram-negative bacteria possess a relatively impermeable outer membrane protecting the bacteria against noxious molecules from the environment such as antibiotics. In addition, proteins embedded within the outer membrane ensure the import of nutrients into the cell, signal transduction and protein secretion. A major class of outer membrane proteins are the porins, integral membrane proteins which form β-barrels containing between 8 and 22 β-strands [57]. The medium size porins mediate the uptake of nutrients across the outer membrane of Gram-negative bacteria. These co-called general porin channels form relatively non-specific pores allowing the passage of small hydrophylic molecules, with an exclusion limit of about 600 Da [46]. The pore size in *P. aeruginosa* is even larger and allows the transport of molecules larger than the size of a tetrasaccharide [6]. These general systems should therefore be able to catalyse the uptake of small oligopeptides.

Examples of the larger barrels in Gram-negative bacteria are the TonB-dependent iron siderophore transporters FhuA and FepA. Interestingly, sensing of the signal occurs at the cell surface and is transduced to the cytoplasm [9]. Colicin B, a member of the group B colicins and a channel forming colicin, probably crosses the outer membrane bilayer through the FepA pore [28]. Colicin M, which causes the lysis of target cells, requires the FhuA protein. The uptake of colicin M is

energized by the interaction between FhuA and TonB [8,10]. Of particular interest are the TonB-dependent microcins, a class of small peptide colicins with a size ranging from 21 to 88 amino acids including colicin V, E492 and microcin 24 [10]. The latter peptides are secreted through the specialized ABC-type transporter concomitant with the removal of a GG-motif leader peptide. Export requires the outer membrane TolC protein. The binding and uptake of these peptides through the outer membrane involves the siderophore receptors Cir, FepA or FhuA in a TonB-dependent manner [34].

The energy-driven export of proteins such as α-hemolysin and peptides by ABC transporters makes use of the TolC protein. This protein is trimeric and forms a structure composed of a β-barrel consisting of 12 β-strands in the outer membrane and a contiguous α-helical barrel protruding through the periplasm, most likely contacting the ABC transporter [33]. This pore can handle proteins up to the size of 200 kDa implicating that, because of the limited size of the TolC channel, these proteins are likely unfolded and present as long thin molecules during transport. Interestingly, this system has also been 'parasitized' by the highly efficient bactericidal colicins. In case of colicin E1, it is proposed that, after binding of the colicin to its receptor and partial unfolding of the N-terminal domain of the toxin, this domain is transferred through the TolC channel to the inner membrane side of the periplasm [22]. In this way, TolC allows the passage of peptides from the cell exterior into the periplasm. As it is the case for the cytoplasmic membrane, signaling peptides with properties reminiscent of AMPs could also integrate and traverse the outer membrane [64].

The question of whether Gram-negative bacteria could sense peptide signals is an intriguing one. At the cytoplasmic membrane, two-component signal transduction systems or oligopeptide permeases, two main components of the QS system in Gram-positive bacteria, are conserved in many Gram-negative bacteria. The PhoP–PhoQ two-component system was recently suggested to sense directly the presence of cationic antimicrobial peptides in *S. typhimurium* [2]. A major difference resides in the outer membrane. The transport of small peptides across the outer membrane may occur through porins. Several systems that could be involved in the transport of larger peptides include the TolC protein and iron siderophore transporters. In this respect, much can be learned from the uptake mechanisms of bacteriocins such as the well-studied colicins and microcins. The focus of this article was put on GG-peptides and their dedicated transporters because of their conserved motifs. However, alternative systems transporting or generating small peptides may very well be present as well. For example, it is assumed that a QS peptide pheromone in the Gram-negative bacterium *Providencia stuartii* is produced through the activation of a signal peptide by regulated intramembrane proteolysis [21]. Likewise, the eight amino acids peptide pheromone cAD1 from *E. faecalis* is also generated through proteolysis of a transmembrane precursor protein [61]. Peptides as signaling molecules have been described in animals, plants and bacteria and could therefore constitute an ancient form of intercellular communication.

References

[1] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.

[2] Bader MW, Navarre WW, Shiau W, Nikaido H, Frye JG, McClelland M, et al. Regulation of *Salmonella typhimurium* virulence gene expression by cationic antimicrobial peptides. Mol Microbiol 2003;50:219–30.

[3] Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 1994;2:28–36.

[4] Bailey TL, Gribskov M. Combining evidence using *P*-values: application to sequence homology searches. Bioinformatics 1998;14:48–54.

[5] Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The Pfam protein families database. Nucleic Acids Res 2002;30:276–80.

[6] Bellido F, Martin NL, Siehnel RJ, Hancock RE. Reevaluation, using intact cells, of the exclusion limit and role of porin OprF in *Pseudomonas aeruginosa* outer membrane permeability. J Bacteriol 1992;174:5196–203.

[7] Birney E, Durbin R. Using GeneWise in the *Drosophila* annotation experiment. Genome Res 2000;10:547–8.

[8] Braun M, Killmann H, Braun V. The beta-barrel domain of FhuADelta5-160 is sufficient for TonB-dependent FhuA activities of *Escherichia coli*. Mol Microbiol 1999;33:1037–49.

[9] Braun V. Surface signaling: novel transcription initiation mechanism starting from the cell surface. Arch Microbiol 1997;167:325–31.

[10] Braun V, Patzer SI, Hantke K. Ton-dependent colicins and microcins: modular design and evolution. Biochimie 2002;84:365–80.

[11] Chen X, Schauder S, Potier N, Van Dorsselaer A, Pelczer I, Bassler BL, et al. Structural identification of a bacterial quorum-sensing signal containing boron. Nature 2002;415:545–9.

[12] Cheng Q, Campbell EA, Naughton AM, Johnson S, Masure HR. The *com* locus controls genetic transformation in *Streptococcus pneumoniae*. Mol Microbiol 1997;23:683–92.

[13] Dalet K, Cenatiempo Y, Cossart P, Hechard Y. A sigma(54)-dependent PTS permease of the mannose family is responsible for sensitivity of *Listeria monocytogenes* to mesentericin Y105. Microbiology 2001;147:3263–9.

[14] DeKeersmaecker SCJ, Vanderleyden J. Constraints on detection of autoinducer-2 (AI-2) signalling molecules using *Vibrio harveyi* as a reporter. Microbiology 2003;149:1953–6.

[15] Detmers FJ, Kunji ER, Lanfermeijer FC, Poolman B, Konings WN. Kinetics and specificity of peptide uptake by the oligopeptide transport system of *Lactococcus lactis*. Biochemistry 1998;37:16671–9.

[16] Detmers FJ, Lanfermeijer FC, Poolman B. Peptides and ATP binding cassette peptide transporters. Res Microbiol 2001;152:245–58.

[17] Dunny GM, Leonard BA. Cell–cell communication in gram-positive bacteria. Annu Rev Microbiol 1997;51:527–64.

[18] Eddy SR. Profile hidden Markov models. Bioinformatics 1998; 14:755–63.

[19] Ennahar S, Sashihara T, Sonomoto K, Ishizaki A. Class IIa bacteriocins: biosynthesis, structure and activity. FEMS Microbiol Rev 2000;24:85–106.

[20] Fuqua C, Parsek MR, Greenberg EP. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. Annu Rev Genet 2001;35:439–68.

[21] Gallio M, Sturgill G, Rather P, Kylsten P. A conserved mechanism for extracellular signaling in eukaryotes and prokaryotes. Proc Natl Acad Sci USA 2002;99:12208–13.

[22] Griko YV, Zakharov SD, Cramer WA. Structural stability and domain organization of colicin E1. J Mol Biol 2000;302:941–53.
[23] Haas W, Shepard BD, Gilmore MS. Two-component regulator of *Enterococcus faecalis* cytolysin responds to quorum-sensing autoinduction. Nature 2002;415:84–7.
[24] Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol 1995;16:229–40.
[25] Havarstein LS, Holo H, Nes IF. The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. Microbiology 1994;140:2383–9.
[26] Higgins CF. ABC transporters: physiology, structure and mechanism—an overview. Res Microbiol 2001;152:205–10.
[27] Jack RW, Tagg JR, Ray B. Bacteriocins of gram-positive bacteria. Microbiol Rev 1995;59:171–200.
[28] Jiang X, Payne MA, Cao Z, Foster SB, Feix JB, Newton SM, et al. Ligand-specific opening of a gated-porin channel in the outer membrane of living bacteria. Science 1997;276:1261–4.
[29] Kaplan HB, Greenberg EP. Diffusion of autoinducer is involved in regulation of the *Vibrio fischeri* luminescence system. J Bacteriol 1985;163:1210–4.
[30] Klaenhammer TR. Genetics of bacteriocins produced by lactic acid bacteria. FEMS Microbiol Rev 1993;12:39–85.
[31] Kleerebezem M, Quadri LE. Peptide pheromone-dependent regulation of antimicrobial peptide production in Gram-positive bacteria: a case of multicellular behavior. Peptides 2001;22:1579–96.
[32] Kleerebezem M, Quadri LE, Kuipers OP, de Vos WM. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. Mol Microbiol 1997;24:895–904.
[33] Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature 2000;405:914–9.
[34] Lagos R, Baeza M, Corsini G, Hetz C, Strahsburger E, Castillo JA, et al. Structure, organization and characterization of the gene cluster involved in the production of microcin E492, a channel-forming bacteriocin. Mol Microbiol 2001;42:229–43.
[35] Lazazzera BA. The intracellular function of extracellular signaling peptides. Peptides 2001;22:1519–27.
[36] Leonard BA, Podbielski A, Hedberg PJ, Dunny GM. *Enterococcus faecalis* pheromone binding protein, PrgZ, recruits a chromosomal oligopeptide permease system to import sex pheromone cCF10 for induction of conjugation. Proc Natl Acad Sci USA 1996;93:260–4.
[37] Loh J, Carlson RW, York WS, Stacey G. Bradyoxetin, a unique chemical signal involved in symbiotic gene regulation. Proc Natl Acad Sci USA 2002;99:14446–51.
[38] Manson MD, Blank V, Brade G, Higgins CF. Peptide chemotaxis in *E. coli* involves the Tap signal transducer and the dipeptide permease. Nature 1986;321:253–6.
[39] Marketon MM, Gronquist MR, Eberhard A, Gonzalez JE. Characterization of the *Sinorhizobium meliloti sinR*/*sinI* locus and the production of novel *N*-acyl homoserine lactones. J Bacteriol 2002;184:5686–95.
[40] McAuliffe O, Ross RP, Hill C. Lantibiotics: structure, biosynthesis and mode of action. FEMS Microbiol Rev 2001;25:285–308.
[41] McNab R, Lamont RJ. Microbial dinner-party conversations: the role of LuxS in interspecies communication. J Med Microbiol 2003;52:541–5.
[42] McPhee JB, Lewenza S, Hancock RE. Cationic antimicrobial peptides activate a two-component regulatory system, PmrA-PmrB, that regulates resistance to polymyxin B and cationic antimicrobial peptides in *Pseudomonas aeruginosa*. Mol Microbiol 2003;50:205–17.
[43] Michiels J, Dirix G, Vanderleyden J, Xi C. Processing and export of peptide pheromones and bacteriocins in Gram-negative bacteria. Trends Microbiol 2001;9:164–8.
[44] Miller MB, Bassler BL. Quorum sensing in bacteria. Annu Rev Microbiol 2001;55:165–99.
[45] Nes IF, Diep DB, Havarstein LS, Brurberg MB, Eijsink V, Holo H. Biosynthesis of bacteriocins in lactic acid bacteria. Antonie Van Leeuwenhoek 1996;70:113–28.
[46] Nikaido H, Rosenberg EY. Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of *Escherichia coli*. J Gen Physiol 1981;77:121–35.
[47] Novick RP, Muir TW. Virulence gene regulation by peptides in staphylococci and other Gram-positive bacteria. Curr Opin Microbiol 1999;2:40–5.
[48] Parsek MR, Greenberg EP. Acyl-homoserine lactone quorum sensing in gram-negative bacteria: a signaling mechanism involved in associations with higher organisms. Proc Natl Acad Sci USA 2000;97:8789–93.
[49] Pearson JP, Van Delden C, Iglewski BH. Active efflux and diffusion are involved in transport of *Pseudomonas aeruginosa* cell-to-cell signals. J Bacteriol 1999;181:1203–10.
[50] Pesci EC, Milbank JB, Pearson JP, McKnight S, Kende AS, Greenberg EP, et al. Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA 1999;96:11229–34.
[51] Pestova EV, Havarstein LS, Morrison DA. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an autoinduced peptide pheromone and a two-component regulatory system. Mol Microbiol 1996;21:853–62.
[52] Pons AM, Lanneluc I, Cottenceau G, Sable S. New developments in non-post translationally modified microcins. Biochimie 2002;84:531–7.
[53] Qin X, Singh KV, Weinstock GM, Murray BE. Characterization of *fsr*, a regulator controlling expression of gelatinase and serine protease in *Enterococcus faecalis* OG1RF. J Bacteriol 2001;183:3372–82.
[54] Sahl HG, Bierbaum G. Lantibiotics: biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. Annu Rev Microbiol 1998;52:41–79.
[55] Schaefer AL, Taylor TA, Beatty JT, Greenberg EP. Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. J Bacteriol 2002;184:6515–21.
[56] Schauder S, Shokat K, Surette MG, Bassler BL. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. Mol Microbiol 2001;41:463–76.
[57] Schulz GE. The structure of bacterial outer membrane proteins. Biochim Biophys Acta 2002;1565:308–17.
[58] Solomon JM, Magnuson R, Srivastava A, Grossman AD. Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. Genes Dev 1995;9:547–58.
[59] Sperandio V, Torres AG, Giron JA, Kaper JB. Quorum sensing is a global regulatory mechanism in enterohemorrhagic *Escherichia coli* O157: H7. J Bacteriol 2001;183:5187–97.
[60] Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annu Rev Biochem 2000;69:183–215.
[61] Sturme MH, Kleerebezem M, Nakayama J, Akkermans AD, Vaugha EE, de Vos WM. Cell to cell communication by autoinducing peptides in gram-positive bacteria. Antonie Van Leeuwenhoek 2002;81:233–43.
[62] Tame JR, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF, et al. The structural basis of sequence-independent peptide binding by OppA protein. Science 1994;264:1578–81.
[63] Tortosa P, Dubnau D. Competence for transformation: a matter of taste. Curr Opin Microbiol 1999;2:588–92.
[64] Tossi A, Sandri L, Giangaspero A. Amphipathic, alpha-helical antimicrobial peptides. Biopolymers 2000;55:4–30.
[65] van Belkum MJ, Worobo RW, Stiles ME. Double-glycine-type leader peptides direct secretion of bacteriocins by ABC transporters: colicin V secretion in *Lactococcus lactis*. Mol Microbiol 1997;23:1293–301.

[66] Van de Peer Y, De Wachter R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 1994;10:569–70.
[67] Wandersman C. Protein and peptide secretion by ABC exporters. Res Microbiol 1998;149:163–70.
[68] Whitehead NA, Barnard AM, Slater H, Simpson NJ, Salmond GP. Quorum-sensing in Gram-negative bacteria. FEMS Microbiol Rev 2001;25:365–404.
[69] Winzer K, Hardie KR, Williams P. Bacterial cell-to-cell communication: sorry, cannot talk now—gone to lunch! Curr Opin Microbiol 2002;5:216–22.