Screening genomes of Gram-positive bacteria for double-glycine-motif-containing peptides

Secreted peptides fulfil major functions in the physiology of eukaryotes as well as prokaryotes. Yet, in many genome-sequencing projects, small peptides either remain un-annotated or are classified as hypothetical open reading frames, without any function associated. Therefore, the identification of signal peptide sequences would help in finding novel peptide genes or in assigning functions to automatically annotated sequences.

In Gram-positive bacteria, the double-glycine (GG) motif plays a key role in many peptide secretion systems involved in quorum sensing and bacteriocin production. Competence-stimulating

peptides and class II bacteriocins, produced by streptococci and lactic acid bacteria, respectively, are generally synthesized as inactive prepeptides containing a conserved GG-type leader sequence. This leader sequence, typically between 15 and 30 aa in length, is recognized and proteolytically removed during secretion by its cognate ABC-transporter, resulting in the release and activation of the peptide. Processed peptides vary in length from 17 to over 80 aa. The GG-type leader sequence is well conserved and possesses the following consensus: LSX_2ELX_2IXGG (Havarstein *et al.*, 1994). Beside this conserved leader sequence, GG-motif-containing peptides lack common sequence similarities. Their cognate transporters contain a specific domain of about 150 aa which is responsible for the proteolytic removal of the GG-type leader peptide and, on the basis of its sequence, has been classified as the Peptidase C39 protein family domain. The Peptidase C39 domain contains two conserved motifs, called the cysteine and the histidine motifs (Havarstein *et al.*, 1995).

Our aim was to detect new GG-motif-containing peptides in the fully sequenced genomes of Gram-positive bacteria. Since many peptides containing such a motif are small, it is likely that many of them may not have been annotated in genome-sequencing projects or have not been recognized as secreted peptides. Therefore, an *in silico* strategy was designed and applied at the nucleotide level. The 45 fully sequenced genomes of Gram-positive bacteria [situation on 15 September 2003; for a complete list, see Dirix *et al.* (2004)] were screened both for the presence of GG-motifs and for Peptidase C39 domains. For the latter screening, a motif was available (http://www.sanger.ac.uk/Software/Pfam; accession number PF03412) (Bateman *et al.*, 2002); for the GG-motif search, a new model was built based on already known GG-motif peptides (Dirix *et al.*, 2004; Michiels *et al.*, 2001). Based on our knowledge of characterized GG-motif-containing peptides, several restrictions on the GG-motif candidate genes were imposed. First, the GG-motif was forced to end with a Gly-Gly or a 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-containing gene. Finally, the length of the leader sequence and the total peptide length were set to a maximum of 30 and 150 aa, respectively. As a result, by using these restrictions, we cannot exclude that some GG-motif peptides were not retrieved during the screening process.

A search for the Peptidase C39 domain in 45 fully sequenced Gram-positive genomes resulted in a total of 29 hits. These hits were found in the genera *Bacillus*, *Clostridium*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Mycoplasma*, *Streptococcus*, *Streptomyces* and *Ureaplasma*, but not in the genus *Bifidobacterium*, *Corynebacterium*, *Deinococcus*, *Listeria*, *Mycobacterium*, *Oceanobacillus*, *Staphylococcus* or *Tropheryma*. Interestingly, all of the screened lactic acid bacteria, with the exception of *Streptococcus agalactiae* (strains 2603V/R and NEM316) and *Bifidobacterium longum* NCC2705, contain a Peptidase C39 domain. In several strains belonging to the genera *Streptococcus* and *Enterococcus*, more than one protein containing the C39 domain was found. Besides two protein hits that are truncated in their Peptidase C39 domain, all hits contain the conserved cysteine and histidine motifs involved in GG-motif recognition and peptidase activity (Havarstein *et al.*, 1995), suggesting that those domains have peptidase activity.

The screening for peptides containing a GG-motif resulted in a total of 48 candidate peptides. Although out of the 45 screened bacterial genomes, only 12 genomes were from lactic acid bacteria, 92 % of all GG-motif-containing hits were found in lactic acid bacteria (of which 80 % belong to streptococcal strains). The size of the peptides ranges from 29 to 126 aa, or in the mature form (i.e. without the leader peptide) from 11 to 103 aa. A list of the possible GG-motif-containing peptides, their cognate transport protein, their length, amino acid context, theoretical pI and molecular mass is given in Table 1. If available, the gene name of the GG-peptide-encoding sequence was taken from the genome annotation data and included in Table 1. Sixty-seven per cent of the candidate peptides have a high glycine content (>10 % Gly), whereas in 63 % of the peptides more than half of the amino acids are hydrophobic. Also, half of the hits have two or more cysteine residues and, in 56 % of the peptides, the theoretical pI is higher than 8. These data are consistent with the properties of previously described GG-motif-containing peptides (Ennahar *et al.*, 2000; Jack *et al.*, 1995). Among the 48 hits, three were not annotated in the corresponding genome sequence project. Seventeen hits, annotated as hypothetical proteins, did not display similarity to any known protein or peptide. The remaining hits are bacteriocins ($n=15$) or bacteriocin homologues ($n=10$), a conserved domain protein ($n=1$), a plantaricin biosynthesis protein ($n=1$) and a phage-related protein ($n=1$).

Physical linkage to one or more possible GG-peptide(s) was obtained in 21 out of the 29 Peptidase C39 domains retrieved. Screening of the lactic acid bacterium *Lactococcus lactis* subsp. *lactis* strain IL1403 revealed the presence of two un-annotated putative GG-peptides. In this strain, LcnC, a peptidase C39 domain containing protein, was shown previously to transport lactococcin A (Bolotin *et al.*, 2001). The bacteriocin lactococcin A is synthesized as a precursor containing a GG-type leader peptide (only produced by some *Lactococcus lactis* strains) and is plasmid-encoded (Holo *et al.*, 1991). Although the *lcnC* and *lcnD* genes are present on the chromosome of strain IL1403, no gene encoding a lactococcin A homologue was found, either on the chromosome or on one of its plasmids. As suggested previously, the LcnCD proteins could secrete compounds other than bacteriocins (Venema *et al.*, 1996). The two putative GG-peptides obtained from this screening are good candidates for substrates of LcnCD. In *Lactobacillus plantarum* WCFS1, six GG-peptides were found, five of which are the plantaricin bacteriocins PlnA, PlnE, PlnF, PlnJ and PlnN, the other is PlnY, annotated as a putative plantaricin biosynthesis protein (Diep *et al.*, 1996; Nissen-Meyer *et al.*, 1993). Although the precursor of bacteriocin PlnK contains a GG-motif in *Lactobacillus plantarum* C11 (Diep *et al.*, 1996), PlnK was not retrieved in this screening. Further analysis learnt that the *plnK* genes from *Lactobacillus plantarum* strains C11 and WCFS1 differ in two

Table 1. List of the possible GG-motif-containing peptides

ID	Organism	Accession no.	Start*	Stop*	Gene	Transporter†	Length‡		Amino acid composition§						pI	Mol. mass (Da)
							GG	Total	Arom. (%)	Charged (%)	Polar (%)	Hydroph. (%)	Cys	Gly		
1	*Bacillus subtilis*	NC_000964	2 271 945	2 271 901	*yolB*	*sunT*	16	57	4·9	21·9	34·1	43·8	0	2·4	8·38	4671·3
2	*Clostridium acetobutylicum*	NC_001988	76 046	76 093	CAP0072‖	CAP0073	15	58	18·6	21·0	39·6	39·6	1	2·3	4·23	4907·3
3	*Enterococcus faecalis*	NC_004671	49 438	49 485	EFB0056‖	EFB0050	26	74	6·3	24·9	33·3	41·7	5	8·3	8·72	5179·0
4	*Lactococcus lactis*	NC_002662	86 271	86 315	NA	*lcnC*	21	41	10·0	10·0	55·0	35·0	0	35	3·67	1985·9
5	*Lactococcus lactis*	NC_002662	88 983	89 027	NA	*lcnC*	17	41	20·8	16·8	29·3	54·2	0	16·7	6·07	2777·1
6	*Lactobacillus plantarum*	NC_004567	366 994	366 947	*plnJ*	*plnG*	21	46	20·0	32·0	24·0	44·0	0	16·0	10·93	2929·3
7	*Lactobacillus plantarum*	NC_004567	368 259	368 306	*plnN*	*plnG*	25	55	20·0	19·9	43·3	36·6	0	13·3	9·70	3369·8
8	*Lactobacillus plantarum*	NC_004567	371 389	371 436	*plnA*	*plnG*	15	41	15·3	23·1	34·5	42·1	0	7·7	10·40	2985·5
9	*Lactobacillus plantarum*	NC_004567	375 965	375 918	*plnF*	*plnG*	18	52	20·6	20·5	23·5	55·9	0	11·8	10·27	3703·1
10	*Lactobacillus plantarum*	NC_004567	376 145	376 098	*plnE*	*plnG*	23	56	12·1	24·2	18·2	57·6	0	18·2	11·57	3545·1
11	*Lactobacillus plantarum*	NC_004567	383 668	383 715	*plnY*	*plnG*	18	29	18·2	27·3	45·5	27·3	0	0·0	8·53	1356·5
12	*Streptococcus mutans*	NC_004350	269 347	269 391	SMU.283‖	SMU.286	20	69	10·2	6·0	22·4	71·2	2	16·3	4·37	4712·4
13	*Streptococcus mutans*	NC_004350	1 776 619	1 776 575	SMU.1882c‖	SMU.1881c/1897	23	117	13·8	10·6	51·2	38·3	0	13·8	4·53	10212·9
14	*Streptococcus mutans*	NC_004350	1 781 366	1 781 319	SMU.1889c	SMU.1881c/1897	23	87	6·3	3·2	23·3	73·3	2	28·1	3·67	5713·3
15	*Streptococcus mutans*	NC_004350	1 781 849	1 781 805	SMU.1892c‖	SMU.1881c/1897	25	61	19·5	25·0	36·2	39·0	0	5·6	11·32	4213·6
16	*Streptococcus mutans*	NC_004350	1 783 593	1 783 549	SMU.1895c	SMU.1881c/1897	23	53	13·3	10·0	19·9	70·0	0	10	8·5	3275·9
17	*Streptococcus mutans*	NC_004350	1 783 889	1 783 845	SMU.1896c	SMU.1881c/1897	18	78	10·0	6·8	23·2	70·0	2	30	8·07	5608·4
18	*Streptococcus mutans*	NC_004350	1 787 899	1 787 855	SMU.1902c‖	SMU.1897	22	47	16·0	32·0	28·0	40·0	0	4	6·3	3040·4
19	*Streptococcus mutans*	NC_004350	1 789 887	1 789 843	SMU.1905c	SMU.1897	22	62	5·0	10·0	15·0	75·0	2	17·5	5·95	3736·3
20	*Streptococcus mutans*	NC_004350	1 790 260	1 790 213	SMU.1906c	SMU.1897	18	65	6·3	10·7	14·9	74·4	1	38·3	4·56	4190·6
21	*Streptococcus mutans*	NC_004350	1 794 717	1 794 670	SMU.1914c	SMU.1897	23	76	9·5	1·9	22·8	75·5	2	32·1	8·05	4777·3
22	*Streptococcus pneumoniae* R6	NC_003098	39 538	39 585	*thmA*	*comA*	18	71	11·3	9·5	26·4	64·3	2	26·4	9·39	5181·8
23	*Streptococcus pneumoniae* R6	NC_003098	117 752	117 796	Spr0109‖	Spr0105¶	23	126	12·6	13·6	28·2	58·2	0	14·6	9·77	10591
24	*Streptococcus pneumoniae* R6	NC_003098	119 712	119 756	Spr0111‖	Spr0105¶	23	123	9·0	12·0	20·0	68·0	0	18	9·98	10061·6
25	*Streptococcus pneumoniae* R6	NC_003098	123 834	123 878	Spr0115‖	Spr0105¶	23	124	9·0	12·0	20·0	68·0	0	18	11·48	10735·3
26	*Streptococcus pneumoniae* R6	NC_003098	472 023	471 979	Spr0465	Spr0469	24	51	14·8	29·6	22·2	48·1	0	3·7	5·57	3236·8
27	*Streptococcus pneumoniae* R6	NC_003098	1 732 074	1 732 118	Spr1765‖	*clyB*	29	74	4·4	13·2	33·2	53·4	2	11·1	5·84	4478
28	*Streptococcus pneumoniae* R6	NC_003098	1 732 293	1 732 337	Spr1766‖	*clyB*	25	62	5·4	24·3	37·8	37·8	2	8·1	9·24	3905·4
29	*Streptococcus pneumoniae* TIGR4	NC_003028	39 895	39 942	SP0041	SP0042	18	71	11·3	9·5	26·4	64·3	2	26·4	9·39	5181·8
30	*Streptococcus pneumoniae* TIGR4	NC_003028	507 823	507 779	SP0528	SP0530	24	42	27·9	22·4	39	39	0	5·6	5·32	2254·5
31	*Streptococcus pneumoniae* TIGR4	NC_003028	511 742	511 786	SP0531	SP0530	18	60	0	2·4	23·8	73·9	2	23·8	8·07	3772·4
32	*Streptococcus pneumoniae* TIGR4	NC_003028	512 406	512 453	SP0532	SP0530	18	84	6	6	25·6	68·2	2	27·3	5·21	6123·0
33	*Streptococcus pneumoniae* TIGR4	NC_003028	512 744	512 791	SP0533	SP0530	18	71	11·3	7·6	26·4	66·1	2	28·3	8·86	5054·6
34	*Streptococcus pneumoniae* TIGR4	NC_003028	515 115	515 159	SP0539	SP0530	18	79	13·1	6·5	24·5	68·9	2	27·9	6·05	5868·6
35	*Streptococcus pneumoniae* TIGR4	NC_003028	515 370	515 414	SP0540	SP0530	18	67	8·1	4	22·4	73·4	2	26·5	8·82	4470·1
36	*Streptococcus pneumoniae* TIGR4	NC_003028	515 832	515 879	SP0541	SP0530	18	44	7·6	19·1	26·8	53·8	2	15·4	4·43	2613·9
37	*Streptococcus pneumoniae* TIGR4	NC_003028	1 850 404	1 850 448	SP1948	SP1953	29	74	4·4	13·2	33·2	53·4	2	11·1	5·84	4478
38	*Streptococcus pneumoniae* TIGR4	NC_003028	1 850 623	1 850 667	SP1949‖	SP1953	25	62	5·4	24·3	37·8	37·8	2	8·1	9·24	3905·4
39	*Streptococcus pyogenes* MGAS315	NC_004070	1 664 879	1 664 835	*salA*	SpyM3_1650	19	41	13·6	18·2	40·8	40·8	3	9·1	5·94	2378·7
40	*Streptococcus pyogenes* MGAS8232	NC_003485	423 034	423 081	SpyM18_0525	SpyM18_0524¶	22	74	15·3	7·6	28·7	63·4	2	25·0	9·50	5167·8

Table 1. cont.

ID	Organism	Accession no.	Start*	Stop*	Gene	Transporter†	Length‡		Amino acid composition§					pI	Mol. mass (Da)
							GG	Total	Arom. (%)	Charged (%)	Polar (%)	Hydroph. (%)	Cys	Gly	
41	*Streptococcus pyogenes* MGAS8232	NC_003485	423 161	423 207	NA	SpyM18_0524¶	21	82	27·5	23·5	33·3	43·2	0	5·9	5·62	6302·1
42	*Streptococcus pyogenes* MGAS8232	NC_003485	424 632	424 679	SpyM18_0528‖	SpyM18_0524¶/0543	18	70	9·6	11·5	30·8	57·7	2	21·2	8·82	5297·0
43	*Streptococcus pyogenes* MGAS8232	NC_003485	430 596	430 552	SpyM18_0540‖	SpyM18_0524¶/0543	24	41	11·8	47·1	5·9	51·6	0	0	10	2042·5
44	*Streptococcus pyogenes* MGAS8232	NC_003485	434 538	434 582	SpyM18_0544	SpyM18_0543	23	75	9·5	3·8	28·7	67·3	2	21·2	8·86	4927·6
45	*Streptococcus pyogenes* MGAS8232	NC_003485	434 763	434 807	SpyM18_0545	SpyM18_0543	15	63	10·5	8·4	18·9	72·9	2	25	8·03	4561·3
46	*Streptococcus pyogenes* SSI-1	NC_004606	1 658 665	1 658 618	SPs1650‖	NA	19	41	13·6	18·2	40·8	40·8	3	9·1	5·94	2378·7
47	*Streptomyces avermitilis*	NC_003155	8 931 651	8 931 604	SAV7495‖	SAV7493	16	64	0	4·2	25	70·9	0	16·7	3·67	4432·0
48	*Streptomyces coelicolor*	NC_003888	796 608	796 652	SCO0753‖	SCO0755	23	71	0	4·2	25	70·9	0	18·8	3·67	4473·1

NA, Not annotated in the genome-sequencing project.

*The position of the GG-motif-coding region on the corresponding replicon is given.

†The gene(s) encoding the transport protein(s) to which the possible peptide is/are genetically linked.

‡Length of the GG-leader sequence (GG) and the total peptide (Total) in amino acids. The first ATG, GTG or TTG upstream of the GG-encoding sequence was arbitrarily chosen as the start codon.

§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. Arom., aromatic residues (Fen, His, Try, Tyr); Charged, (Arg, Asp, Glu, Lys, His); Polar, (Asn, Cys, Gln, Ser, Thr, Trp, Tyr); Hydroph., hydrophobic (Ala, Fen, Gly, Ile, Leu, Val, Met, Pro).

‖Annotated as hypothetical protein without any similarity.

¶Protein containing a truncated Peptidase C39 domain.

nucleotides, corresponding to one amino acid difference in the GG-motif. The 'GG-motif' of the *Lactobacillus plantarum* WCFS1 PlnK ends with a Gly-Asn pair (in contrast to Gly-Gly in *Lactobacillus plantarum* C11), and was therefore not identified in our screening, as only Gly-Gly or Gly-Ala pairs were allowed.

The screened streptococci can be subdivided into the naturally competent (*Streptococcus pneumoniae* and *Streptococcus mutans*) and the non-competent (*Streptococcus agalactiae* and *Streptococcus pyogenes*) species (Havarstein *et al.*, 1997; Li *et al.*, 2001). All screened strains belonging to the competent group have more than one Peptidase C39 domain-containing protein, of which one is the competence-stimulating peptide transporter ComA. Although the pheromone was never found in this screening (because of the 10 kb restriction), many other possible GG-peptides were retrieved. Beside hypothetical proteins, these hits constitute bacteriocins or bacteriocin homologues (de Saizieu *et al.*, 2000). The non-competent group can be further subdivided on the basis of the presence (*Streptococcus pyogenes* strains MGAS315, MGAS8232 and SSI-1) or the absence (*Streptococcus pyogenes* strain M1 GAS and *Streptococcus agalactiae*) of a Peptidase C39 domain. In the *Streptococcus pyogenes* strains containing a Peptidase C39 domain, several putative bacteriocins (de Saizieu *et al.*, 2000) and a putative pheromone (Smoot *et al.*, 2002) were retrieved, including the lantibiotic salivaricin A, which functions both as a bacteriocin and a pheromone (Ross *et al.*, 1993; Upton *et al.*, 2001). In *Enterococcus faecalis* V583, one putative GG-peptide was found, annotated as a hypothetical protein.

In addition to the lactic acid bacteria, the screening also revealed four more GG-motif-encoding genes in the strains *Bacillus subtilis* subsp. *subtilis* str. 168, *Clostridium acetobutylicum* ATCC 824, *Streptomyces avermitilis* MA-4680 and *Streptomyces coelicolor* A3(2), encoding a phage-related protein (*Bacillus subtilis*) (Kunst *et al.*, 1997) and three hypothetical proteins.

Finally, none of the Peptidase C39 transporter genes found in *Mycoplasma* and

Ureaplasma strains, in *Bacillus halodurans* and *Enterococcus faecalis* (the second C39 domain only) is linked to a gene encoding a GG-peptide. The latter two proteins are involved in the transport of mersacidin and cytolysin, respectively. Mersacidin and cytolysin are two lantibiotics that are synthesized as prepropeptides with GG-type leader sequences that differ too much from the GG-motif consensus sequence (mersacidin) or end with a Gly-Ser pair (both precursors from cytolysin) and were therefore not retrieved in this screening (Altena *et al.*, 2000; Gilmore *et al.*, 1994). The same holds true for sublancin 168, a lantibiotic from *Bacillus subtilis*, of which the leader sequence also ends with a Gly-Ser pair (Paik *et al.*, 1998).

To conclude, our screening strategy led to new insights in the distribution of GG-peptide processing and secretion systems among Gram-positive bacteria. The results are not dependent on previous annotations as the screening was performed at the nucleotide level. Interestingly, for all Peptidase C39 domains identified, one or more possible GG-peptide genes were found within the 10 kb limit of the *in silico* screening or in the literature (see also previous paragraph), except for *Mycoplasma* and *Ureaplasma* species. Although the competence-stimulating peptides from *Streptococcus* competent strains have not been retrieved in this analysis, other GG-motif-containing candidates were found. This could imply that cognate transporters are used to secrete multiple compounds with very different functions. More than half of the GG-hits retrieved are bacteriocins or putative bacteriocins, some of them also function as pheromones. More than 40 % of the identified peptide genes were either un-annotated or had not yet been recognized as secreted peptides in the genome-sequencing projects. These peptide genes were detected in the genomes of lactic acid bacteria, but also in the genera *Bacillus*, *Clostridium* and *Streptomyces*. Finally, not all known GG-motif-containing peptides were found in the current analysis. Also, in the case of several Peptidase C39 domains no corresponding peptides were present. This may be due to the 10 kb restriction, but this clearly is not always the case (e.g. for mersacidin, cytolysin and the plantaricin PlnK). On the other hand, the motif of the GG-leader sequence used may be too specific. Therefore, as more GG-containing peptides will be characterized biochemically in the future, including those with a divergent leader sequence, the algorithm could be further refined.

G. Dirix,[1]† P. Monsieurs,[2]† K. Marchal,[2] J. Vanderleyden[1] and J. Michiels[1]

[1]Centre of Microbial and Plant Genetics, K. U. Leuven, Heverlee, Belgium

[2]ESAT-SCD, K. U. Leuven, Heverlee, Belgium

†Both authors made equal contributions to this article.

Correspondence: J. Michiels (Jan.Michiels@agr.kuleuven.ac.be)

Altena, K., Guder, A., Cramer, C. & Bierbaum, G. (2000). Biosynthesis of the lantibiotic mersacidin: organization of a type B lantibiotic gene cluster. *Appl Environ Microbiol* **66**, 2565–2571.

Bateman, A., Birney, E., Cerruti, L. & 7 other authors (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276–280.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D. & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**, 731–753.

de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T. J., Keck, W., Amrein, K. E. & Lange, R. (2000). Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J Bacteriol* **182**, 4696–4703.

Diep, D. B., Havarstein, L. S. & Nes, I. F. (1996). Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. *J Bacteriol* **178**, 4472–4483.

Dirix, G., Monsieurs, P., Dombrecht, B., Daniels, R., Marchal, K., Vanderleyden, J. & Michiels, J. (2004). 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. *Peptides* (in press).

Ennahar, S., Sashihara, T., Sonomoto, K. & Ishizaki, A. (2000). Class IIa bacteriocins: biosynthesis, structure and activity. *FEMS Microbiol Rev* **24**, 85–106.

Gilmore, M. S., Segarra, R. A., Booth, M. C., Bogie, C. P., Hall, L. R. & Clewell, D. B. (1994). Genetic structure of the *Enterococcus faecalis* plasmid pAD1-encoded cytolytic toxin system and its relationship to lantibiotic determinants. *J Bacteriol* **176**, 7335–7344.

Havarstein, L. S., Holo, H. & Nes, I. F. (1994). The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. *Microbiology* **140**, 2383–2389.

Havarstein, L. S., Diep, D. B. & Nes, I. F. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol* **16**, 229–240.

Havarstein, L. S., Hakenbeck, R. & Gaustad, P. (1997). Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J Bacteriol* **179**, 6589–6594.

Holo, H., Nilssen, O. & Nes, I. F. (1991). Lactococcin A, a new bacteriocin from *Lactococcus lactis* subsp. *cremoris*: isolation and characterization of the protein and its gene. *J Bacteriol* **173**, 3879–3887.

Jack, R. W., Tagg, J. R. & Ray, B. (1995). Bacteriocins of gram-positive bacteria. *Microbiol Rev* **59**, 171–200.

Kunst, F., Ogasawara, N., Moszer, I. & 148 other authors (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256.

Li, Y. H., Lau, P. C., Lee, J. H., Ellen, R. P. & Cvitkovitch, D. G. (2001). Natural genetic transformation of *Streptococcus mutans* growing in biofilms. *J Bacteriol* **183**, 897–908.

Michiels, J., Dirix, G., Vanderleyden, J. & Xi, C. (2001). Processing and export of peptide pheromones and bacteriocins in Gram-negative bacteria. *Trends Microbiol* **9**, 164–168.

Nissen-Meyer, J., Larsen, A. G., Sletten, K., Daeschel, M. & Nes, I. F. (1993). Purification and characterization of plantaricin A, a *Lactobacillus plantarum* bacteriocin whose activity depends on the action of two peptides. *J Gen Microbiol* **139**, 1973–1978.

Paik, S. H., Chakicherla, A. & Hansen, J. N. (1998). Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by *Bacillus subtilis* 168. *J Biol Chem* **273**, 23134–23142.

Ross, K. F., Ronson, C. W. & Tagg, J. R. (1993). Isolation and characterization of the lantibiotic salivaricin A and its structural gene *salA* from *Streptococcus salivarius* 20P3. *Appl Environ Microbiol* **59**, 2014–2021.

Smoot, J. C., Barbian, K. D., Van Gompel, J. J. & 15 other authors (2002). Genome sequence and comparative microarray analysis of serotype

M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proc Natl Acad Sci U S A* **99**, 4668–4673.

Upton, M., Tagg, J. R., Wescombe, P. & Jenkinson, H. F. (2001). Intra- and interspecies signaling between *Streptococcus salivarius* and *Streptococcus pyogenes* mediated by SalA and SalA1 lantibiotic peptides. *J Bacteriol* **183**, 3931–3938.

Venema, K., Dost, M. H., Beun, P. A., Haandrikman, A. J., Venema, G. & Kok, J. (1996). The genes for secretion and maturation of lactococcins are located on the chromosome of *Lactococcus lactis* IL1403. *Appl Environ Microbiol* **62**, 1689–1692.

DOI 10.1099/mic.0.27040-0