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

Subject category: Comment

Dirix, G.[1,*], Monsieurs, P.[2,*], Marchal, K.[2], Vanderleyden, J.[1] & Michiels, J.[1]

[1]Centre of Microbial and Plant Genetics, K.U.Leuven, Heverlee, Belgium
[2]ESAT-SCD, K.U.Leuven, Heverlee, Belgium
*Both authors equally contributed to this paper

Corresponding author:
Jan Michiels
Centre of Microbial and Plant Genetics
Kasteelpark Arenberg 20
B-3001 Heverlee
Belgium
Tel.: ++32 (0)16 321631
Fax: ++32 (0)16 321966
Jan.Michiels@agr.kuleuven.ac.be

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 (CSPs) and class II bacteriocins, produced by streptococci and lactic acid bacteria (LAB) respectively, are generally synthesized as inactive prepeptides containing a conserved GG-type leader sequence. This leader sequence is recognized and proteolytically removed by its cognate ABC-transporter during secretion, resulting in the release and subsequent activation of the peptide. The following consensus sequence of the GG-motif was proposed: LSX_2ELX_2IXGG (Havarstein et al., 1994). The cognate transporters generally contain three domains. The central transmembrane and the C-terminal ATPase domain are found in other ABC-transporters, while the N-terminally located domain of about 150 amino acids is specific. The latter domain 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 (www.sanger.ac.uk/Software/Pfam; accession number PF03412) (Bateman *et al*., 2002). The Peptidase C39 domain contains two conserved motifs, 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 (Havarstein *et al*., 1995).

Since many quorum sensing and bacteriocin peptides containing a GG-type leader sequence are small, likely many of them may not have been annotated in genome sequencing projects. Therefore, an *in silico* strategy was designed and applied at the nucleotide level to identify novel peptides. 45 fully sequenced genomes of Gram-positive bacteria (situation on September 15th, 2003; for a complete list see Dirix *et al*. (2004)) were screened for the presence of GG-motifs and Peptidase C39 domains by using the Wise2 package. Wise2 (www.ebi.ac.uk/Wise2) translates the bacterial genomes in the six reading frames and compares the translations with a specified Hidden Markov Model (HMM) (Birney & Durbin, 2000). For the Peptidase C39 domain search, the corresponding HMM was obtained from the Pfam database (www.sanger.ac.uk/Software/Pfam; accession number PF03412) (Bateman *et al*., 2002). For the GG-motif search, two HMMs were built by using the HMMER2.2 software (http://hmmer.wustl.edu) on two curated training sets (Eddy, 1998). One training set is based on already known GG-motif peptides from Gram-positive bacteria, the other is based on possible GG-motif peptides from Gram-negative bacteria (Dirix *et al*.,

2004; Michiels *et al*., 2001). Because both HMMs are built on small sequences, some restrictions were introduced in our search, based on the knowledge of already known GG-motif containing peptides. No insertions or gaps were allowed in the GG-motif and the motif was forced 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-motif peptide systems, the structural gene is clustered with the genes coding for the secretion, the processing and/or the sensing machinery (in Gram-positive as well in Gram-negative bacteria) (Kleerebezem *et al*., 1997; Michiels *et al*., 2001). Thirdly, the length of the leader sequence and the total peptide length were set to a maximum of 30 and 150 amino acids respectively. 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. By using these restrictions we cannot exclude that some GG-motif peptides are lost throughout the screening process.

The Peptidase C39 domain screening of 45 fully sequenced Gram-positive genomes resulted in a total of 29 hits. Hits are found in the genera *Bacillus*, *Clostridium*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Mycoplasma, Streptococcus, Streptomyces* and *Ureaplasma*, but not in *Bifidobacterium*, *Corynebacterium*, *Deinococcus*, *Listeria*, *Mycobacterium*, *Oceanobacillus*, *Staphylococcus* or *Tropheryma*. Interestingly, all screened LAB, with the exception of *Streptococcus agalactiae* (strains 2603V/R and NEM316) and *Bifidobacterium longum* NCC2705, contain a Peptidase C39 domain. In some strains belonging to the *Streptococcus* or the *Enterococcus* genus, more than one domain was found. Besides two 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 dedicated transporters were recently reclassified into four classes based on their domain organization (Dirix *et al*., 2004). Members of class A have a Peptidase C39, a transmembrane and an ATPase domain, from N- to C-terminus respectively. Class B proteins only contain the Peptidase C39 domain. Class C and class D resemble class A, but lack a transmembrane domain (class C) or have an extra N-terminal extension (class D). With the exception of class D, the Gram-positive hits are spread amongst all classes.

The GG-motif screening resulted in a total of 48 possible GG-motif containing peptides. Although from the 45 screened bacterial genomes only 12 constituted LAB genomes, 92% of all GG-motif containing hits are retrieved from LAB genomes, of which 80% belong to streptococcal strains. The size of the peptides ranges from 29 to 126 amino acids, or in the mature form (i.e. without the leader peptide) from 11 to 103 amino acids. From all candidate peptides, the mature part was analyzed with the ProtParam tool (http://us.expasy.org/tools/protparam.html). A list of the possible GG-motif containing peptides, their cognate transport protein, their length, amino acid context, theoretical pI and molecular weight is given in Table 1. If applicable, the name of every GG-hit coding sequence was taken from the genome annotation and added to Table 1. 67% of the candidate peptides have high glycine content (more than 10% glycine) whereas for 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 for 56% of the peptides, the theoretical pI is higher than 8. These data are consistent with the properties of bacteriocins and CSPs (Ennahar *et al*., 2000; Jack *et al*., 1995). Of the 48 possible hits, three weren't annotated by the corresponding genome sequence project. For 17 hits, annotated as hypothetical proteins, no similarity with already known proteins or peptides was found. The remaining hits constitute bacteriocins or bacteriocin homologues (26), a conserved domain protein (1) and a plantaricin biosynthesis protein (1).

For 21 out of the 29 found Peptidase C39 domains, physical linkage to one or more possible GG-peptide(s) was obtained. Screening of the LAB *Lactococcus lactis* subsp. *lactis* strain IL1403 revealed the presence of two un-annotated putative GG-peptides. In this strain, the Peptidase C39 domain containing protein constitutes LcnC, the ABC-transporter of 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 *L. 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 (Venema *et al*., 1996) or on one of its plasmids. As speculated by the authors, the LcnCD proteins could secrete compounds other than bacteriocins. The two putative GG-peptides obtained from this screening could be possible candidates. In *Lactobacillus plantarum* WCFS1, six possible GG-peptides were found, five of which are the plantaricin bacteriocins PlnA, PlnE, PlnF, PlnJ and PlnN, the other is PlnY,

105 annotated as a putative plantaricin biosynthesis protein (Diep *et al*., 1996; Nissen-Meyer *et al*., 1993).
106 Although the precursor of bacteriocin PlnK contains a GG-motif in *L. plantarum* C11 (Diep *et al*., 1996),
107 PlnK was not retrieved in this screening. Further analysis learnt that the *plnK* genes from *L. plantarum* strain
108 C11 and WCFS1 differ in two nucleotides, corresponding to one amino acid difference in the GG-motif. The
109 'GG-motif' of the *L. plantarum* WCFS1 PlnK ends with a Gly-Asn pair (in contrast to Gly-Gly in *L.*
110 *plantarum* C11), and was therefore excluded from our screening as only Gly-Gly or Gly-Ala pairs were
111 allowed. The screened streptococci can be subdivided into the naturally competent (*Streptococcus*
112 *pneumoniae* and *Streptococcus mutans*) and the non-competent species (*Streptococcus agalactiae* and
113 *Streptococcus pyogenes*) (Havarstein *et al*., 1997; Li *et al*., 2001). All screened strains of the competent
114 group have more than one Peptidase C39 domain containing protein, of which one is the CSP transporter
115 ComA. Although the CSP itself is never found in this screening (because of the 10 kb restriction), many
116 other possible GG-peptides are retrieved. Besides hypothetical proteins, these hits constitute bacteriocins or
117 bacteriocin homologues (de Saizieu *et al*., 2000). The non-competent group can be further subdivided on the
118 basis of the presence (*S. pyogenes* strains MGAS315, MGAS8232 and SSI-1) or the absence (*S. pyogenes*
119 strain M1 GAS and *S. agalactiae*) of a Peptidase C39 domain. In the *S. pyogenes* strains containing a
120 Peptidase C39 domain, several putative bacteriocins (de Saizieu *et al*., 2000) and a putative pheromone
121 (Smoot *et al*., 2002) were retrieved, also including the lantibiotic salivaricin A, which functions both as a
122 bacteriocin and a pheromone (Ross *et al*., 1993; Upton *et al*., 2001). In *Enterococcus faecalis* V583, one
123 putative GG-peptide was found, annotated as a hypothetical protein.
124 Besides hits in LAB, the screening revealed four more GG-motif encoding genes in the strains *Bacillus*
125 *subtilis* subsp. *subtilis* str. 168, *Clostridium acetobutylicum* ATCC824, *Streptomyces avermitilis* MA-4680
126 and *Streptomyces coelicolor* A3(2), coding for a phage-related protein (*B. subtilis*) (Kunst *et al*., 1997) and
127 for three hypothetical proteins.
128
129 Finally, no GG-hit was found for all Peptidase C39 domains found in *Mycoplasma* and *Ureaplasma* strains,
130 for the C39 domain of *Bacillus halodurans* and for the second C39 domain of *E. faecalis*. The latter two are
131 part of proteins involved in the transport of mersacidin and cytolysin respectively. Mersacidin and cytolysin
132 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 in 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 *B. 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 secreting and processing systems amongst Gram-positive bacteria. Interestingly, for all Peptidase C39 domains, one or more possible GG-hits were found within the 20 kb limit of the screening, except for all domains belonging to *Mycoplasma* and *Ureaplasma* species. More than half of the GG-hits retrieved are bacteriocins or putative bacteriocins, some of them also act as a pheromone. Besides already known GG-motif containing peptides, several new possible GG-motif containing peptides were retrieved by screening at the nucleotide level, not only in LAB, but also in the genera *Bacillus*, *Clostridium* and *Streptomyces*.

Reference List

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., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276-280.

Birney, E. & Durbin, R. (2000). Using GeneWise in the *Drosophila* annotation experiment. *Genome Res* **10**, 547-548.

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. 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.

Eddy, S.R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755-763.

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., 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.

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.

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.

Kleerebezem, M., Quadri, L. E., Kuipers, O. P. & de Vos, W. M. (1997). Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Mol Microbiol* **24**, 895-904.

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 USA* **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.

Table 1
List of the possible GG-motif containing peptides

							Length[‡]		Amino acid composition[§]							
ID	Organism	Acc.Nr.	Start[*]	Stop[*]	Gene	Transporter[†]	GG	Total	% arom.	% charged	% polar	% hydroph.	Cys	Gly	pI	M.W.[‖]
1	B. subtilis	NC_000964	2271945	2271901	yolB	sunT	16	57	4.9	21.9	34.1	43.8	0	2.4	8.38	4671.3
2	C. acetobutylicum	NC_001988	76046	76093	CAP0072	CAP0073	15	58	18.6	21.0	39.6	39.6	1	2.3	4.23	4907.3
3	E. faecalis	NC_004671	49438	49485	EFB0056	EFB0050	26	74	6.3	24.9	33.3	41.7	5	8.3	8.72	5179.0
4	L. lactis	NC_002662	86271	86315	/	lcnC	21	41	10.0	10.0	55.0	35.0	0	35	3.67	1985.9
5	L. lactis	NC_002662	88983	89027	/	lcnC	17	41	20.8	16.8	29.3	54.2	0	16.7	6.07	2777.1
6	L. plantarum	NC_004567	366994	366947	plnJ	plnG	21	46	20.0	32.0	24.0	44.0	0	16.0	10.93	2929.3
7	L. plantarum	NC_004567	368259	368306	plnN	plnG	25	55	20.0	19.9	43.3	36.6	0	13.3	9.70	3369.8
8	L. plantarum	NC_004567	371389	371436	plnA	plnG	15	41	15.3	23.1	34.5	42.1	0	7.7	10.40	2985.5
9	L. plantarum	NC_004567	375965	375918	plnF	plnG	18	52	20.6	20.5	23.5	55.9	0	11.8	10.27	3703.1
10	L. plantarum	NC_004567	376145	376098	plnE	plnG	23	56	12.1	24.2	18.2	57.6	0	18.2	11.57	3545.1
11	L. plantarum	NC_004567	383668	383715	plnY	plnG	18	29	18.2	27.3	45.5	27.3	0	0.0	8.53	1356.5
12	S. mutans	NC_004350	269347	269391	SMU.283	SMU.286	20	69	10.2	6.0	22.4	71.2	2	16.3	4.37	4712.4
13	S. mutans	NC_004350	1776619	1776575	SMU.1882c	SMU.1881c/1897	23	117	13.8	10.6	51.2	38.3	0	13.8	4.53	10212.9
14	S. mutans	NC_004350	1781366	1781319	SMU.1889c	SMU.1881c/1897	23	87	6.3	3.2	23.3	73.3	2	28.1	3.67	5713.3
15	S. mutans	NC_004350	1781849	1781805	SMU.1892c	SMU.1881c/1897	25	61	19.5	25.0	36.2	39.0	0	5.6	11.32	4213.6
16	S. mutans	NC_004350	1783593	1783549	SMU.1895c	SMU.1881c/1897	23	53	13.3	10.0	19.9	70.0	0	10	8.5	3275.9
17	S. mutans	NC_004350	1783889	1783845	SMU.1896c	SMU.1881c/1897	18	78	10.0	6.8	23.2	70.0	2	30	8.07	5608.4
18	S. mutans	NC_004350	1787899	1787855	SMU.1902c	SMU.1897	22	47	16.0	32.0	28.0	40.0	0	4	6.3	3040.4
19	S. mutans	NC_004350	1789887	1789843	SMU.1905c	SMU.1897	22	62	5.0	10.0	15.0	75.0	2	17.5	5.95	3736.3
20	S. mutans	NC_004350	1790260	1790213	SMU.1906c	SMU.1897	18	65	6.3	10.7	14.9	74.4	1	38.3	4.56	4190.6
21	S. mutans	NC_004350	1794717	1794670	SMU.1914c	SMU.1897	23	76	9.5	1.9	22.8	75.5	2	32.1	8.05	4777.3
22	S. pneumoniae R6	NC_003098	39538	39585	thmA	comA	18	71	11.3	9.5	26.4	64.3	2	26.4	9.39	5181.8
23	S. pneumoniae R6	NC_003098	117752	117796	Spr0109	Spr0105[¶]	23	126	12.6	13.6	28.2	58.2	0	14.6	9.77	10591
24	S. pneumoniae R6	NC_003098	119712	119756	Spr0111	Spr0105[¶]	23	123	9.0	12.0	20.0	68.0	0	18	9.98	10061.6
25	S. pneumoniae R6	NC_003098	123834	123878	Spr0115	Spr0105[¶]	23	124	9.0	12.0	20.0	68.0	0	18	11.48	10735.3
26	S. pneumoniae R6	NC_003098	472023	471979	Spr0465	Spr0469	24	51	14.8	29.6	22.2	48.1	0	3.7	5.57	3236.8
27	S. pneumoniae R6	NC_003098	1732074	1732118	Spr1765	clyB	29	74	4.4	13.2	33.2	53.4	2	11.1	5.84	4478
28	S. pneumoniae R6	NC_003098	1732293	1732337	Spr1766	clyB	25	62	5.4	24.3	37.8	37.8	2	8.1	9.24	3905.4
29	S. pneumoniae TIGR4	NC_003028	39895	39942	SP0041	SP0042	18	71	11.3	9.5	26.4	64.3	2	26.4	9.39	5181.8
30	S. pneumoniae TIGR4	NC_003028	507823	507779	SP0528	SP0530	24	42	27.9	22.4	39	39	0	5.6	5.32	2254.5
31	S. pneumoniae TIGR4	NC_003028	511742	511786	SP0531	SP0530	18	60	0	2.4	23.8	73.9	2	23.8	8.07	3772.4
32	S. pneumoniae TIGR4	NC_003028	512406	512453	SP0532	SP0530	18	84	6	6	25.6	68.2	2	27.3	5.21	6123.0
33	S. pneumoniae TIGR4	NC_003028	512744	512791	SP0533	SP0530	18	71	11.3	7.6	26.4	66.1	2	28.3	8.86	5054.6
34	S. pneumoniae TIGR4	NC_003028	515115	515159	SP0539	SP0530	18	79	13.1	6.5	24.5	68.9	2	27.9	6.05	5868.6
35	S. pneumoniae TIGR4	NC_003028	515370	515414	SP0540	SP0530	18	67	8.1	4	22.4	73.4	2	26.5	8.82	4470.1
36	S. pneumoniae TIGR4	NC_003028	515832	515879	SP0541	SP0530	18	44	7.6	19.1	26.8	53.8	2	15.4	4.43	2613.9
37	S. pneumoniae TIGR4	NC_003028	1850404	1850448	SP1948	SP1953	29	74	4.4	13.2	33.2	53.4	2	11.1	5.84	4478
38	S. pneumoniae TIGR4	NC_003028	1850623	1850667	SP1949	SP1953	25	62	5.4	24.3	37.8	37.8	2	8.1	9.24	3905.4
39	S. pyogenes MGAS315	NC_004070	1664879	1664835	salA	SpyM3_1650	19	41	13.6	18.2	40.8	40.8	3	9.1	5.94	2378.7
40	S. pyogenes MGAS8232	NC_003485	423034	423081	SpyM18_0525	SpyM18_0524[¶]	22	74	15.3	7.6	28.7	63.4	2	25.0	9.50	5167.8
41	S. pyogenes MGAS8232	NC_003485	423161	423207	/	SpyM18_0524[¶]	21	82	27.5	23.5	33.3	43.2	0	5.9	5.62	6302.1

42	S. pyogenes MGAS8232	NC_003485	424632	424679	SpyM18_0528	SpyM18_0524[¶]/0543	18	70	9.6	11.5	30.8	57.7	2	21.2	8.82	5297.0
43	S. pyogenes MGAS8232	NC_003485	430596	430552	SpyM18_0540	SpyM18_0524[¶]/0543	24	41	11.8	47.1	5.9	51.6	0	0	10	2042.5
44	S. pyogenes MGAS8232	NC_003485	434538	434582	SpyM18_0544	SpyM18_0543	23	75	9.5	3.8	28.7	67.3	2	21.2	8.86	4927.6
45	S. pyogenes MGAS8232	NC_003485	434763	434807	SpyM18_0545	SpyM18_0543	15	63	10.5	8.4	18.9	72.9	2	25	8.03	4561.3
46	S. pyogenes SSI-1	NC_004606	1658665	1658618	SPs1650	/	19	41	13.6	18.2	40.8	40.8	3	9.1	5.94	2378.7
47	S. avermitilis	NC_003155	8931651	8931604	SAV7495	SAV7493	16	64	0	4.2	25	70.9	0	16.7	3.67	4432.0
48	S. coelicolor	NC_003888	796608	796652	SCO0753	SCO0755	23	71	0	4.2	25	70.9	0	18.8	3.67	4473.1

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

[†]The gene(s) coding for 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 double glycine 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 residus (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)

[‖]molecular weight in Dalton

/not annotated

[¶]protein containing a truncated Peptidase C39 domain