BioData Mining

Detection of putative new mutacins by bioinformatic analysis using available web tools

Guillaume G Nicolas

Correspondence: guillaume.
nicolas.1@ulaval.ca
Département de Biochimie
Microbiologie et Bioinformatique,
Faculté des Sciences et Génie,
Université Laval, Québec (Québec),
G1K7P4, Canada

Abstract

In order to characterise new bacteriocins produced by *Streptococcus mutans* we perform a complete bioinformatic analyses by scanning the genome sequence of strains UA159 and NN2025. By searching in the adjacent genomic context of the two-component signal transduction system we predicted the existence of many putative new bacteriocins' maturation pathways and some of them were only exclusive to a group of *Streptococcus*. Computational genomic and proteomic analysis combined to predictive functionnal analysis represent an alternative way for rapid identification of new putative bacteriocins as well as new potential antimicrobial drugs compared to the more traditional methods of drugs discovery using antagonism tests.

Findings

The increasing resistance of bacteria to antibiotics motives researches for new antimicrobial compounds [1]. In this way bacteriocins which are small antibacterial ribosomally synthetized peptides produced by bacteria represent promising candidates [2,3]. Bacteriocins acted on sensitive cells by punching pores in their membrane. To date, the bacteriocins produced by Gram positive bacteria are grouped in two major classes [4] but four classes are also proposed [5]. Lantibiotic_class I and non-lantibiotic_class II bacteriocins display great diversity with regard to their structures, modes of action, and genetic determinants [4,6]. Typical bacteriocin biosynthesis operons are usually organised as a cluster of genes comprising the prepropeptide coding gene associated with genes for exportation and maturation (ATP-binding cassette (ABC) transporter and sometimes combined to a specific protease), genes conferring immunity to the inhibitory activity to prevent self-killing and occasionally genes involved in regulation of the production of the bacteriocin [6,3]. The expression of the bacteriocin gene cluster is under the control of a two-component signal transduction system (TCS) composed of an histidine kinase (HK) and its associated response regulator (RR) that are usually part of the cluster. The inducer can be either the bacteriocin itself or a bacteriocin-like peptide [7].

Discovery of new bacteriocins traditionally rest upon functionnal assays based on the inhibition of specific target bacteria. Such method is limited and time-consuming regarding the culture condition for bacteriocin production with the indicator strains used. The growing of genomic data makes the detection of new bacteriocin peptides possible by using an *in silico* screening strategy and precise computational analyses.

Recently many research teams have bring to light existence of new type of bacteriocins using this strategy [8-11]. Furthermore, a very powerful tool for direct discovery of bacteriocins inside genomic data have been recently develop [12]. However, such tool build on well-known bacteriocins characteristics may overlook detection of new type of bacteriocins as bacteriocins detect by Haft methodoly are not found using BAGEL2 [8,9]. Open reading frame detection and identification coding for short peptides including bacteriocin precursors inside genomes is generally recognised as difficult to perform [13].

Our research group is interested in the discovery of new antibacterial compounds produced by *Streptococcus mutans* and named mutacins [14]. Based on the conserved organisation of bacteriocin biosynthesis operon, we screened the genomic context of the HK/RR genes found in the *S. mutans* UA159 genome to detect new putative bacteriocin-encoding genes (GenBank: AE014133) [15]. Following a profound inspection by bioinformatic analysis using available web tools we were able to identify new putative bacteriocin maturation patchways in the *S. mutans* genome.
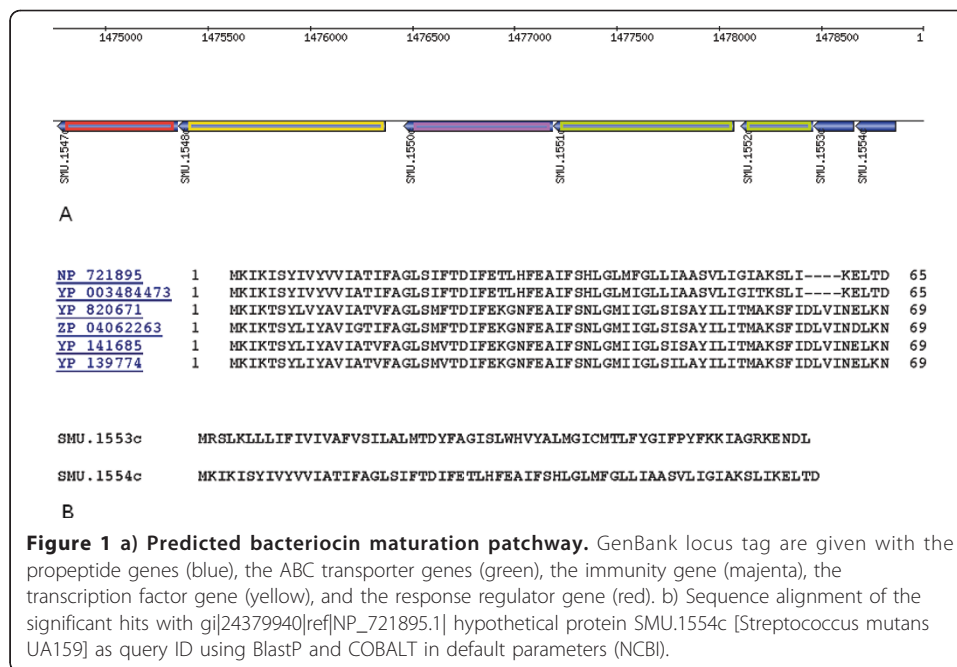
The Microbial Signal Transduction database (MiST, http://mistdb.com) [16] was used to locate the HK/RR genes inside the *S. mutans* genome (Table 1). A set of small ORF encoding small peptides were identified around each TCS. By browsing the genomic context using the Entrez Gene tool from the NCBI http://www.ncbi.nlm.nih.gov/gene we identified a complete set of bare genes able to produce bacteriocins in the vicinity of the SMU.1548c/1547c locus tag (Figure 1a).

Table 1 Two Components Systems found in the *S. mutans* UA159 genome.

HK/RR - Locus tag (NCBI)/gene name	Identifed peptides surrounding the HK/RR	Predicted protein function
SMU.45[#]	SMU.40/41	
SMU.486/487		
SMU.577/576 - lytS/lytR	SMU.571	SMU.572 dehydrogenase/cyclohydrolase
SMU.660/659		
SMU.928/927		
SMU.1009/1008		
SMU.1037c/1038c	SMU.1047c ?	
SMU.1128/1129 - ciaH/ciaR	SMU.1131c	
SMU.1145c/1146c	SMU.1147c	Smu.1148-1150 abc transporter
SMU.1516/1517 covS/covR (vicK/vicR)		
SMU.1548c/1547c	SMU.1553c/1554c	Smu.1550c integrale membrane protein, ...
SMU.1814/1815 - scnK/scnR	SMU.1818c	
SMU.1965c/1964c		
SMU.1916/1917 - comD/comE		
SMU.1924 - gcrR*		

[#] HK uncoupled to a RR

* RR uncoupled to an HK

Figure 1 a) Predicted bacteriocin maturation patchway. GenBank locus tag are given with the propeptide genes (blue), the ABC transporter genes (green), the immunity gene (majenta), the transcription factor gene (yellow), and the response regulator gene (red). b) Sequence alignment of the significant hits with gi|24379940|ref|NP_721895.1| hypothetical protein SMU.1554c [Streptococcus mutans UA159] as query ID using BlastP and COBALT in default parameters (NCBI).

The cluster of genes (location: 1475357-1478860) presents the same genomic organisation than conventional bacteriocin biosynthesis operon with the genes encoding two small peptides (SMU.1554c and SMU.1553c), the ABC transporter genes (SMU.1552c/SMU.1551c), a gene encoding an integrale membrane protein possibly involved in the immunity function (SMU.1550c), and the TCS genes, HK gene (SMU.1548c) and RR gene (SMU.1547c), probably implicated in the regulation of the biosynthesis of the bacteriocin. Furthermore additionnal untypical genes were identified: a methionine aminopeptidase (ampM/SMU.1556c) and a putative acetyltransferase (SMU.1558c) with related function to proteases and scaffoldingproteins.

Putative precursor peptides were analysed for the presence of a signal peptide using Signal-3L http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/[17] and PrediSi http://www.predisi.de/[18] algorithms.

The potential of antimicrobial activity of the putative mature peptides was evaluated using freely web available programs such as APD2 http://aps.unmc.edu/AP/main.php [19] and the AntiBP2 server http://www.imtech.res.in/raghava/antibp/[20]. Similarity with known antimicrobial peptides was retrieved for the query input peptide sequences. SMU.1553c presents similarity with the carnocyclin A peptide [21].

A BlastP analysis [22] of the precursor peptides reveals the strict conservation of these peptides with their genomic context to the *Streptococcus salivarius* group species (Figure 1b).

Upstream genomic coding sequence was analyse to detect putative promoter regions and transcription factor binding sites using the bacterial promoter recognition program BPROM (Softberry inc.) (Figure 2).

Many putative mutacin-encoding genes have been previously predicted using bioinformatic analyses and some of them were functionnally verified using mutational analyses for *S. mutans* UA159 [23]. Inactivation of all putative mutacin genes did not abrogate complete antibacterial activity of the strain, let suggest the existence of an

```
>gi|24378532:1478430-1478860 Streptococcus mutans UA159,
complete genome
 Length of sequence-        431
 Threshold for promoters -  0.20
 Number of predicted promoters -      2
 Promoter Pos:     78 LDF-  6.88
 -10 box at pos.     63 TTTTAAAAT Score    79
 -35 box at pos.     46 TTTCCT     Score    37
 Promoter Pos:    410 LDF-  2.79
 -10 box at pos.    395 TGACAACAT Score    31
 -35 box at pos.    377 CTGCAA     Score    20

 Oligonucleotides from known TF binding sites:

 For promoter at      78:
      fhlA:   TCATTTTC at position      38 Score -   7
       lrp:   ATTTTTTT at position      59 Score -  11
      lexA:   TTTTTTTA at position      60 Score -  16
    rpoD18:   TTTAAAAT at position      64 Score -   7
    rpoD17:   GTGTCATA at position      93 Score -   7
 For promoter at     410:
      carP:   CTGTAAAA at position     359 Score -   7
    rpoD17:   AAAAATAG at position     381 Score -   9
      nagC:   TTTAATTT at position     420 Score -   7
    rpoD18:   TTAATTTT at position     421 Score -   9
```

Figure 2 Report of the BPROM promoter detection software.

other type of inhibitory substance produced which confort the reliability of our hypothesis and findings hither [23].

The group of genes detect by our method predicted the existence of a putative bacteriocin maturation pathway in an exclusive group of *Streptococcus* and reveals its potential to encode for a new type of bacteriocin. It also provides mature hypothesis that may be test by a focused wet lab experiment. Since inactivation of small genes remains difficult to perform, our method study provides a computational evidence for identification of a new putative bacteriocin production. This method can be applied to a large set of short coding sequence with unknown function yet found in the streptococcal genomes [13].

Abbreviations
ABC: ATP-binding cassette; BlastP: Basic Local Alignment Search Tool for protein; ORF: Open Reading frame; TCS: Two-component signal transduction system; HK: Histidine Kinase; RR: response regulator.

Authors' contributions
GGN performed the bioinformatic analyses, interpreted the results, and wrote the manuscript.

Competing interests
The author declares that they have no competing interests.

References
1.   Andersson DI, Hughes D: **Antibiotic resistance and its cost: is it possible to reverse resistance?** *Nat Rev Microbiol* 2010, **8**:260-271.

2. Smith L, Hillman J: **Therapeutic potential of type A(I) lantibiotics, a group of cationic peptide antibiotics.** *Curr Opin Microbiol* 2008, **11**:401-408.
3. Riley MA, Wertz JE: **Bacteriocins: evolution, ecology, and application.** *Annu Rev Microbiol* 2002, **56**:117-137.
4. Nes IF, Diep DB, Holo H: **Bacteriocins diversity in** *Streptococcus* **and** *Enterococcus.* *J Bacteriol* 2008, **189**:1189-1198.
5. Wescombe PA, Heng NC, Burton JP, Chilcott CN, Tagg JR: **Streptococcal bacteriocins and the case for** *Streptococcus salivarius* **as model oral probiotics.** *Future Microbiol* 2009, **4**:819-835.
6. Dufour A, Hindré T, Haras D, le Pennec JP: **The biology of lantibiotics from the lacticin 481 group is coming of age.** *FEMS Microbiol Rev* 2006, **31**:134-167.
7. 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.
8. Haft DH: **A strain-variable bacteriocin in** *Bacillus anthracis* **and** *Bacillus cereus* **with repeated Cys-Xaa-Xaa motifs.** *Biol Direct* 2009, **4**:15.
9. Haft DH, Basu MK, Mitchell DA: **Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family.** *BMC Biology* 2010, **8**:70.
10. Begley M, Cotter PD, Hill C, Ross RP: **Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins.** *Appl Environ Microbiol* 2009, **75**:5451-5460.
11. Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, Wohlrab A, Dorrestein PC, Nizet V, Dixon JE: **Discovery of a widely distributed toxin biosynthetic gene cluster.** *Proc Natl Acad Sci USA* 2008, **105**:5879-5884.
12. de Jong A, van Heel AJ, Kok J, Kuipers OP: **BAGEL2: mining for bacteriocins in genomic data.** *Nucleic Acids Res* 2010, **Web Server.**
13. Ibrahim M, Nicolas P, Bessières P, Bolotin A, Monnet V, Gardan R: **A genome-wide survey of short coding sequences in streptococci.** *Microbiology* 2007, **153**:3631-3644.
14. Nicolas GG, Lavoie MC, LaPointe G: **Molecular genetics, genomics and biochemistry of mutacins.** *Genes, Genomes and Genomics* 2007, **1**:193-208.
15. Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, Primeaux C, Tian R, Kenton S, Jia H, Lin S, Qian Y, Li S, Zhu H, Najar F, Lai H, White J, Roe BA, Ferretti JJ: **Genome sequence of** *Streptococcus mutans* **UA159, a cariogenic dental pathogen.** *Proc Natl Acad Sci USA* 2002, **99**:14434-14439.
16. Galperin MY: **Diversity of structure and function of response regulator output domains.** *Curr Opin Microbiol* 2010, **13**:150-159.
17. Shen HB, Chou KC: **Signal-3L: A 3-layer approach for predicting signal peptides.** *Biochem Biophys Res Commun* 2007, **363**:297-303.
18. Hiller K, Grote A, Scheer M, Münch R, Jahn D: **PrediSi: prediction of signal peptides and their cleavage positions.** *Nucleic Acids Res* 2004, , **32** Web Server: W375-9.
19. Wang G, Li X, Wang Z: **APD2: the updated antimicrobial peptide database and its application in peptide design.** *Nucleic Acids Res* 2009, , **37** Database: D933-937.
20. Lata S, Mishra NK, Raghava GP: **AntiBP2: improved version of antibacterial peptide prediction.** *BMC Bioinformatics* 2010, **11(Suppl 1)**:S19.
21. Martin-Visscher LA, van Belkum MJ, Garneau-Tsodikova S, Whittal RM, Zheng J, McMullen LM, Vederas JC: **Isolation and characterization of carnocyclin a, a novel circular bacteriocin produced by** *Carnobacterium maltaromaticum* **UAL307.** *Appl Environ Microbiol* 2008, **74**:4756-4763.
22. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
23. Hale JD, Ting YT, Jack RW, Tagg JR, Heng NC: **Bacteriocin (mutacin) production by** *Streptococcus mutans* **genome sequence reference strain UA159: elucidation of the antimicrobial repertoire by genetic dissection.** *Appl Environ Microbiol* 2005, **71**:7613-7617.