

Cop-like operon: Structure and organization in species of the *Lactobacillale* order

ANGÉLICA REYES¹, ANDREA LEIVA¹, VERÓNICA CAMBIAZO¹,
MARCO A MÉNDEZ^{1,2} and MAURICIO GONZÁLEZ¹

¹ Laboratorio de Bioinformática y Expresión Génica. INTA, Universidad de Chile

² Laboratorio de Bioestadística (INTA). Universidad de Chile

ABSTRACT

Copper is an essential and toxic trace metal for bacteria and, therefore, must be tightly regulated in the cell. *Enterococcus hirae* is a broadly studied model for copper homeostasis. The intracellular copper levels in *E. hirae* are regulated by the *cop* operon, which is formed by four genes: *copA* and *copB* that encode ATPases for influx and efflux of copper, respectively; *copZ* that encodes a copper chaperone; and *copY*, a copper responsive repressor. Since the complete genome sequence for *E. hirae* is not available, it is possible that other genes may encode proteins involved in copper homeostasis. Here, we identified a *cop-like* operon in nine species of *Lactobacillale* order with a known genome sequence. All of them always encoded a CopY-like repressor and a copper ATPase. The alignment of the *cop-like* operon promoter region revealed two CopY binding sites, one of which was conserved in all strains, and the second was only present in species of *Streptococcus* genus and *L. johnsonii*. Additional proteins associated to copper metabolism, CutC and Cupredoxin, also were detected. This study allowed for the description of the structure and organization of the *cop* operon and discussion of a phylogenetic hypothesis based on the differences observed in this operon's organization and its regulation in *Lactobacillale* order.

Key terms: *cop* operon, CopY binding site, copper homeostasis, *Enterococcus hirae*, *Lactobacillale* order

INTRODUCTION

Copper (Cu) is an essential micronutrient for all living organisms from bacteria to humans. Cu acts as a cofactor for several enzymes that carry out fundamental biological functions required for growth and development (4). However, excess of Cu is harmful to cells, resulting in cell death by their binding to essential cellular components (20). Due to this duality, Cu levels must be tightly controlled. In general, Cu homeostasis is regulated at three levels: uptake, intracellular distribution, and efflux. This regulation is achieved by the activity of conserved sets of cellular components that participate in these events (13).

Currently, the genomes sequence of multiple bacteria strains has become

available in on-line databases. This information might be useful to identify genes associated to Cu metabolism in microorganisms and to inquire how gene reordering and gene structure divergence had occurred during operon evolution. Within Gram positive bacteria, one of the most extensively studied models for Cu homeostasis is *Enterococcus hirae* (*Lactobacillale* order) (16). The regulation of intracellular Cu in this strain is mediated by the *cop* operon that contains four genes: *copA*, *copB*, *copy*, and *copZ*. The CopA and CopB proteins are heavy metal CPx-type ATPases and show a high sequence identity with the human Menkes and Wilson Cu ATPases (17). In *E. hirae*, CopA is required for uptake and CopB for efflux of Cu (15). In contrast, in other bacteria,

functional studies of CopA orthologs show that this protein is involved in copper efflux (12). The *copZ* gene encodes a chaperone protein for Cu, which receives the metal from CopA and delivers it to CopY (2). Recent studies have showed that higher Cu concentrations (> 0.5mM of Cu) down-regulate CopZ by inducing its proteolysis (11). CopY acts as a Cu-responsive repressor, thus when Cu increases in the media, CopY is released from the DNA, allowing transcription to proceed (2). Since the complete genome sequence of *E. hirae* is not available, we cannot discard that additional components might exist that play roles in Cu homeostasis.

In order to assess whether the mechanism of copper bacterial homeostasis is conserved in other species of the *Lactobacillales* order, we applied a bioinformatic methodology to identify the members of the *cop* operon. In doing so, we compared the sequence of *cop* operon genes to the complete genome sequence of other species of this order. Using this approach, we identified a *cop-like* operon in all the analyzed species. This operon contains a conserved core formed by a CopY-like repressor and a copper ATPase. The *cop-like* operon also contains a similar CopY binding site, suggesting that it also is regulated by Cu. In addition, in the analyzed species, we found candidate genes that might complement the functions of the *cop* operon in *E. hirae* Cu homeostasis.

METHODS

Bacterial strain and sequence searches

The complete genomes of the strains included in this study have been sequenced, and they belong to the *Lactobacillales* order. The sequences of *cop-like* genes were obtained using BLAST (Basic Local Alignment Search Tool) algorithm against the databases at The Institute for Genomic Research (TIGR) (<http://www.tigr.org>). The *E. hirae* *cop* operon sequences used as template are indicated in figure 1. The sequences of cupredoxin (*cuA*) and *cutC* genes were found by searching for the term

copper in the description of annotated genes.

ClustalW analysis

BioEdit software version 5.0.9 (7) was used to perform a multiple sequence alignment, using ClustalW option (19) with Protein weight matrix Blosum 62 and protein default options.

Phylogenetic analysis

We performed a Neighbor-joining analysis using the algorithm of Kimura of two-parameter distances (8). Statistical support of nodes was obtained by bootstrap analysis (1000 pseudoreplicates). All analyses were made using Mega 2.1 software (9). The 16S rRNA of *Escherichia coli* and *Shigella flexneri* were used as outgroup.

CopY binding site

We obtained a sequence 100 bp upstream of *copY-like* gene start codon for 15 strains. We used ClustalW to perform alignment of the promoter region. Pairs of sequences were aligned according to the clade distribution of their 16S rRNA tree (see Fig. 2).

RESULTS AND DISCUSSION

In this work, the term *cop-like* operon was used to denote sequences that appear as a cluster of at least two continuous *cop* genes with orthologs in *E. hirae* that were separated by less than 50 bp and in the same transcriptional orientation. Using this definition, we detected the *cop-like* operon in 14 strains that corresponded to 9 species from the *Lactobacillales* order (Fig 1). The result of our sequence searches indicated the presence of a *copY-like* gene in all the strains, while in 13 strains we detected a *copA-like* gene, and a *copZ-like* gene was found in 5 species. A *copB-like* gene was detected in only one case. The apparent absence of a CopB-like protein (Cu efflux ATPase) in these nine species suggests that this function might be provided by paralogs

of CopA ATPase, such as those found in *E. faecalis*, *S. agalactiae*, *L. plantarum*, and *L. lactis* (data not shown). Alternatively, we can not discard that CopA-like ATPase might have efflux function or have both are participating in uptake and efflux of Cu in the strains that possess only one gene coding for ATPase. Certainly, to clarify this point, functional analysis of CopA-like protein needs to be developed in the different

strains. Further differences from the *E. hirae* *cop* operon were observed in *S. agalactiae*, which exhibited two copies of a *cop*-like operon, one of them containing *copY*-like, *copA*-like and *copZ*-like genes, and the other containing only *copY*-like and *copA*-like genes. In three species (*S. pneumoniae*, *S. mitis*, and *L. johnsonii*), additional open reading frames (ORF) that formed part of the *cop*-like operon were also detected.

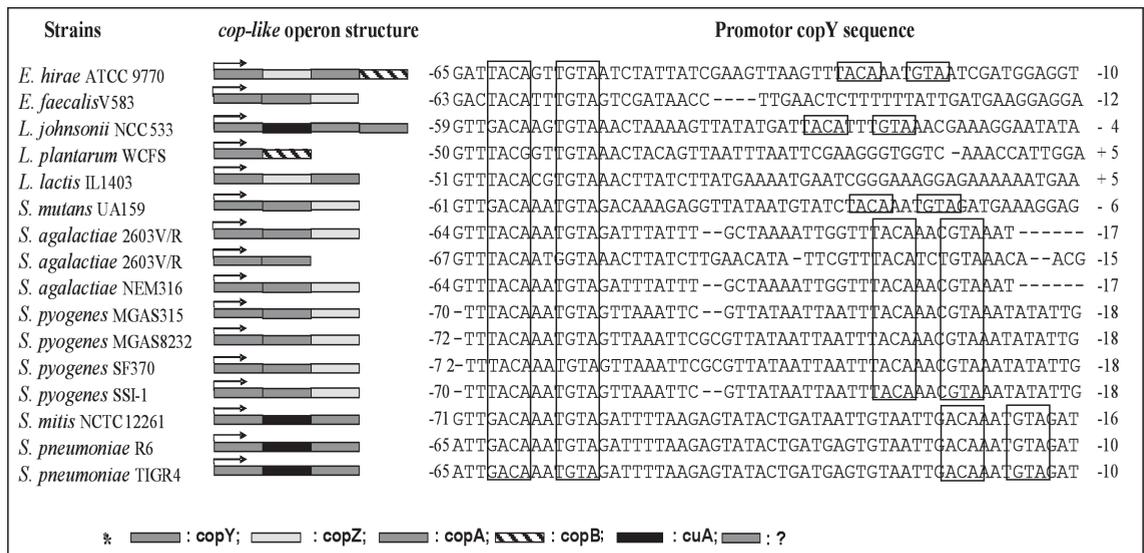


Figure 1. *Cop*-like operon organization and putative CopY binding site in species of the *Lactobacillale* order.

Left side. The figure shows the organization of *cop*-like operon (arrows represent transcription start). *S. agalactiae* 2603V/R showed a second *cop*-like operon. Asterisk (*) indicates the names of genes in the *cop* operon of *E. hirae* and in *cop*-like operon from the other species. Right side. The promoter *copY* sequence containing putative CopY binding sites. Open boxes indicate consensus sequences for the 15 strains (10 species). The numbers at both sides of the promoter *copY* sequence indicate the position of nucleotides with respect to the first codon of CopY.

The Genebank access number of each of the *cop*-like operon genes are indicated from left to right for each strain: *E. hirae*: CAA86835.1, AAA61835.1, AAA61836.1, CAA86836.1; *E. faecalis* V583: AAO80160.1, AAO80161.1, AAO80162.1; *L. johnsonii* NCC 533: AAS09780.1, AAS09781.1, AAS09782.1, AAS09783.1; *L. plantarum* WCFS1: CAD65473.1, CAD65472.1; *L. lactis subsp. lactis* II1403: AAK04930.1, AAK04931.1, AAK04932.1; *S. mutans* UA159: AAN58178.1, AAN58179.1, AAN58180.1; *S. agalactiae* 2603V/R: AAM99290.1, AAM99291.1, AAM99292.1 and AAN00137.1, AAN00136.1; *S. agalactiae* NEM316: CAD46064.1, CAD46065.1, CAD46066.1; *S. pyogenes* MGAS315: AAM80099.1, AAM80098.1, AAM80097.1; *S. pyogenes* MGAS8232: AAL98255.1, AAL98254.1, AAL98253.1; *S. pyogenes* M1 GAS: AAK34463.1, AAK34462.1, AAK34461.1; *S. pyogenes* SSI-1: BAC63470.1, BAC63471.1, BAC63472.1; *S. mitis* NCTC12261: SMT0402, SMT0403, SMT0404; *S. pneumoniae* R6: AAK99443.1, AAK99444.1, AAK99445.1; *S. pneumoniae* TIGR4: AAK74868.1, AAK74869.1, AAK74870.1.

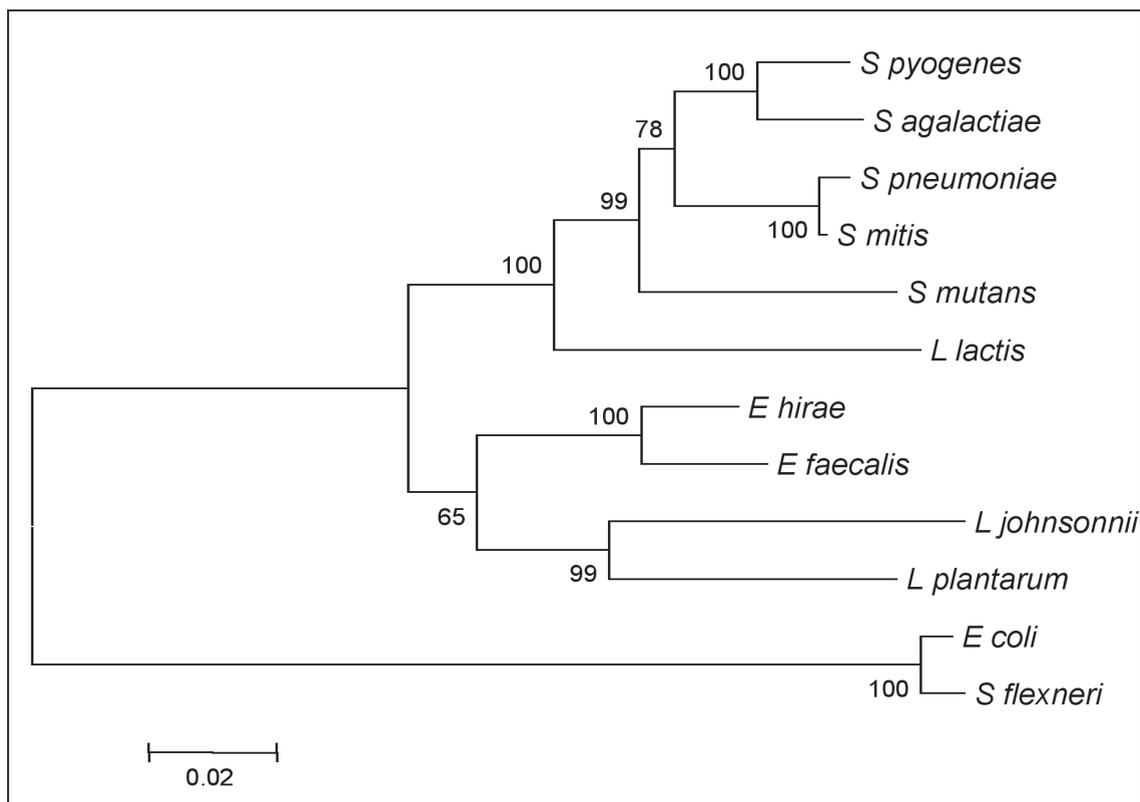


Figure 2. Neighbor-joining tree for the 16S rRNA sequences of *Lactobacillale* species. Bootstrap values are indicated on each node of the tree (1000 pseudoreplicates). Bar indicates number of substitution changes according to Kimura two-parameters distances. The Gene ID access of 16S rRNA sequences are: *E. hirae* AJ554205, *E. faecalis* 1199160, *L. johnsonii* 2742775, *L. plantarum* 1061905, *L. lactis* 1115945, *S. mutans* 2886073, *S. pyogenes* 1066450, *S. agalactiae* 1012735, *S. pneumoniae* 933447, *S. mitis* 40204845.

We compared the organization of the *cop*-like operon in the 14 strains analyzed with that described for *E. hirae* *cop* operon; we found that the *copY*-like gene was at the beginning of the transcriptional units of all the strains. In 9 of them, the *copA*-like gene was down-stream from *copY*-like gene, while *copZ*-like gene either precedes or follows *copA*-like gene. These results indicate that a common position in the operon is found only for *copY*-like and *copA*-like genes among the strains of the *Lactobacillale* order, whereas a higher variability in the position of the *copZ*-like gene was detected. A phylogenetic analysis considering the organization of the operon is presented below.

When we compared the identity of the proteins encoded by the *cop*-like operon with those of *E. hirae*, we found that the ATPases presented high identity scores (42%-51%) with the CopA protein of *E. hirae*, however, they showed low identity with CopB (<36%). An exception was observed for *L. plantarum*: the ATPase encoded in its *cop*-like operon showed 56% sequence identity and 72% similarity with CopB ATPase and a lower identity level with CopA (33%). Similarly, Stentz et al. (2000) (18) showed that in *Lactobacillus sakei*, the single ATPase in the *cop*-like operon (ATKB) has 56.2% identity to the copper efflux CopB ATPase of *E. hirae*. Just as the ATPase found in *L. plantarum*,

ATKB presents the same CPH motif found in CopB instead of the CPC motif usually present in the majority of ATPases (15). The high percentage of identity between *L. plantarum* ATPase and *E. hirae* CopB-like protein suggest that in *L. plantarum*, the ATPase might play a role in efflux.

Among the four components encoded in the *cop-like* operon, CopZ-like protein showed the lowest percentage of identity (24%-42%), and it was detected in 5 of the 9 species (Fig 1). In every case, CopZ-like protein presented an MxCxxC metal binding motif, characteristic of Cu chaperones, at the N-terminus (10). Interestingly, in *L. lactis*, we identified an amino-acidic sequence that exhibits a 42% identity with CopZ of *E. hirae*, which have been annotated as a mercuric reductase (YieF). Considering that in *L. lactis* this gene is localized in a *cop-like* operon between *copY-like* and *copA-like* genes and it exhibited a high identity with the CopZ of *E. hirae*, we propose that this protein corresponds to a Cu chaperone.

Regarding the CopY-like protein detected in the species of *Lactobacillale* order, their percentage of identity with CopY of *E. hirae* ranged from 33 to 49%. In all of the proteins, we detected the CxCxxxxCxC domain at the C-terminus, which is a conserved domain of Cu-response transcription factors, such as MacI and AceI (3). In *E. hirae*, the consensus sequence TACAxXTGTA has been identified as the CopY binding site in the promoter region of the *cop* operon (*cop* box), which occurs two times in this species (14). Using the TACAxXTGTA consensus sequence, we found at least one *cop* box in all of the analyzed genomes (Fig 1), indicating that one of the *E. hirae* CopY binding sites determined by Portmann et al (14) is conserved in *Lactobacillale* species. Sequence alignment shows that 7 of the 8 bp near the -60 position from the start codon of the *copY-like* gene were maintained in all the strains. The other *cop* box described in *E. hirae* was present in all strains of *Streptococcus* genus and in *L. johnsonii*, approximately 22 bp downstream from the first *cop* box. Thus, the presence of two binding sites, described for *E. hirae*

seems to be a conserved feature of *cop-like* operon promoters of *Lactobacillale* species. Considering together the evidence of the conservation of the amino acid sequence of the CopY-like protein and its binding sites on the promoter region, we propose that the regulation of the *cop-like* operon is similar to that described for *E. hirae*.

In addition to the *cop* operon genes identified in *E. hirae*, other genes were detected in the *cop-like* operon after our analysis. In *S. pneumoniae*, *S. mitis*, and *L. johnsonii* we detected a gene encoding a protein whose C-terminal domain was similar to the copper binding site present in cupredoxine (CuA) protein. The cupredoxins, known as blue copper proteins, are small (10–20 kDa) soluble copper proteins whose role is to shuttle electrons from an electron donor to an electron acceptor in bacteria and plants (1, 5). Interestingly, in *L. johnsonii*, we found an extra ORF of unknown function between the *copY-like* and *copA-like* genes. The presence of these additional genes in the *cop-like* operon suggest that their expression is also under the regulation of CopY-like, and that they might complement the functions of *E. hirae* *cop* operon in Cu homeostasis. A component with putative function in copper homeostasis, annotated as *cutC*, was detected at a different chromosomal region in all the bacterial strains, except *L. johnsonii* and *S. mutans*. CutC is a cytoplasmic Cu-binding protein (21) whose amino acidic sequence contains a pattern (M-X-X-M-X-X-X-M) similar to a putative Cu-binding motif, and it has been suggested that CutC could participate in an efflux pathway for Cu (6). It will be interesting to test whether *cutC* is also present in the *E. hirae* genome, since it appears to be conserved among the *Lactobacillale* species.

With the purpose of examining the structure of the *cop* operon from an evolutionary point of view, we performed a phylogenetic analysis using their 16S rRNA sequences of the species analyzed above. The phylogenetic tree shows two major clades, one conformed by species of genus *Streptococcus* and *Lactococcus*, and the other by species of *Lactobacillus* and

Enterococcus genus. Both clades presented terminal nodes with high frequency of occurrence (99-100%) (Fig 2). We observed four different operon organizations among the species that form the first clade. The *cop*-like operon of *S. pyogenes* and *S. agalactiae* exhibited a collinear organization of *copY*-like, *copA*-like and *copZ*-like genes (Fig. 1 and Fig. 2). In *S. pneumoniae* and *S. mitis*, there is an insertion between *copY*-like and *copA*-like genes, which is only present in this group, suggesting that it might be acquired from a common ancestor of this terminal node. *S. mutans* showed exactly the same operon organization as the *S. pyogenes* and *S. agalactiae* group, suggesting that the distribution pattern of *copY*-like, *copA*-like and *copZ*-like genes corresponds to an ancestral configuration of the first clade. In *L. lactis*, we observed a different operon organization (*copY*-like, *copZ*-like and *copA*-like) and only one CopY binding site. Since this species belong to another genus, the evaluation of these structural differences in an evolutionary context will depend on the comparison of other species of the genus. When we compared the *cop* operon in the second clade, we observed that there are practically no similarities in operon organization even within the same genus *Enterococcus*. However, more species of this genus are necessary to confirm the heterogeneity of operon organization within this clade. Regarding the conservation of the CopY binding site we observed that, independently of their relatedness, the majority of the species show two sites, suggesting that this attribute is a conserved character for all the analyzed species.

In summary, our results give insight on the conserved structural features of the *cop* operon among the *Lactobacillale* species and provide evidence for common regulatory mechanisms that coordinate the response to copper exposure in the cell. The operon exhibited a conserved core formed by a CopY-like repressor, a copper ATPase and a CopY binding site. In addition, the organization of genes inside the *cop*-like operon is in agreement with the existence of a common ancestor of the *Streptococcus* genus.

ACKNOWLEDGEMENTS

This work was supported by grants Fondecyt 1050235 and 1030618. AR was a recipient of a CONICYT fellowship.

REFERENCES

- ARNESANO F, BANCIL L, BERTINI I, THOMPSETT AR (2002) Solution structure of CopC: A cupredoxin-like protein involved in copper homeostasis. *Structure (Camb)*. 10: 1337-1347
- COBINE P, WICKRAMASINGHE WA, HARRISON MD, WEBER T, SOLIOZ M DAMERON CT (1999) The *Enterococcus hirae* copper chaperone CopZ delivers copper(I) to the CopY repressor. *FEBS Lett*. 445: 27-30
- COBINE PA, GEORGE GN, JONES CE, WICKRAMASINGHE WA, SOLIOZ M, DAMERON CT (2002). Copper transfer from the Cu(I) chaperone, CopZ, to the repressor, Zn(II)CopY: Metal coordination environments and protein interactions. *Biochemistry* 41: 5822-5829
- CRICHTON RR, PIERRE JL (2001) Old iron, young copper: From Mars to Venus. *Biometals* 14: 99-112
- DE RIENZO F, GABDOULLINE RR, MENZIANI MC, WADE RC (2000) Blue copper proteins: A comparative analysis of their molecular interaction properties. *Protein Sci*. 9: 1439-1454
- GUPTA SD, LEE BT, CAMAKARIS J, WU HC (1995) Identification of cutC and cutF (nlpE) genes involved in copper tolerance in *Escherichia coli*. *J. Bacteriol*. 177: 4207-4215
- HALL TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser* 41: 95-99
- KIMURA M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol*. 16: 111-120
- KUMAR S, TAMURA K, JAKOBSEN IB, NEI M (2001) MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics*. 17: 1244-1245
- LAMB AL, WERNIMONT AK, PUF AHL RA, CULOTTA VC, OHALLORAN TV, ROSENZWEIG AC (1999) Crystal structure of the copper chaperone for superoxide dismutase. *Nat. Struct. Biol*. 6: 724-729
- LU ZH, SOLIOZ M (2001) Copper-induced proteolysis of the CopZ copper chaperone of *Enterococcus hirae*. *J. Biol. Chem*. 276: 47822-47827
- OUTTEN FW, OUTTEN CE, HALE J, O'HALLORAN TV (2000) Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, cueR. *J Biol Chem*. 275: 31024-3102
- PENA MM, LEE J, THIELE DJ (1999) A delicate balance: Homeostatic control of copper uptake and distribution. *J. Nutr*. 129: 1251-1260
- PORTMANN R, MAGNANI D, STOYANOV JV, SCHMECHEL A, MULTHAUP G, SOLIOZ M (2004) Interaction kinetics of the copper-responsive CopY repressor with the cop promoter of *Enterococcus hirae*. *J. Biol. Inorg. Chem*. 9: 396-402
- SOLIOZ M, ODERMATT A (1995) Copper and silver transport by CopB-ATPase in membrane vesicles of *Enterococcus hirae*. *J. Biol. Chem*. 270: 9217-9221
- SOLIOZ M, STOYANOV JV (2003) Copper homeostasis in *Enterococcus hirae*. *FEMS Microbiol. Rev*. 27: 183-195

17. SOLIOZ M, VULPE C (1996) CPx-type ATPases: A class of P-type ATPases that pump heavy metals. *Trends Biochem. Sci.* 21: 237-241
18. STENTZ R, LOIZEL C, MALLERET C, ZAGOREC M (2000) Development of genetic tools for *Lactobacillus sakei*: Disruption of the beta-galactosidase gene and use of lacZ as a reporter gene To study regulation of the putative copper ATPase, AtkB. *Appl Environ Microbiol.* 66: 4272-4278
19. THOMPSON JD, HIGGINS DG, GIBSON TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680
20. URBANSKI NK, BERSEWICZ A (2000) Generation of *OH initiated by interaction of Fe²⁺ and Cu⁺ with dioxygen; comparison with the Fenton chemistry. *Acta Biochim. Pol.* 47: 951-962
21. ZHU DY, ZHU YQ, HUANG RH, XIANG Y, YANG N, LU HX, LI GP, JIN Q, WANG DC (2004) Crystal structure of the copper homeostasis protein (CutCm) from *Shigella flexneri* at 1.7 Å resolution: The first structure of a new sequence family of TIM barrels. *Proteins* 58: 764-768