# Comparative genomics of bacterial zinc regulons: Enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins

Ekaterina M. Panina*, Andrey A. Mironov, and Mikhail S. Gelfand†

State Scientific Center GosNIIGenetika, 1st Dorozhny Proezd 1, Moscow 113545, Russia

Zinc is an important component of many proteins, but in large concentrations it is poisonous to the cell. Thus its transport is regulated by zinc repressors ZUR of proteobacteria and Gram-positive bacteria from the *Bacillus* group and AdcR of bacteria from the *Streptococcus* group. Comparative computational analysis allowed us to identify binding signals of ZUR repressors GAAATGTTATANTATAA-CATTTC for γ-proteobacteria, GTAATGTAA̲A̲TAACATTAC for the *Agrobacterium* group, GATATGTTA̲T̲AACATATC for the *Rhododoccus* group, TAAATCGTAATN̲ATTACGATTTA for Gram-positive bacteria, and TTAACY̲R̲GTTAA of the streptococcal AdcR repressor. In addition to known transporters and their paralogs, zinc regulons were predicted to contain a candidate component of the ATP binding cassette, *zinT* (*b1995* in *Escherichia coli* and *yrpE* in *Bacillus subtilis*). Candidate AdcR-binding sites were identified upstream of genes encoding pneumococcal histidine triad (PHT) proteins from a number of pathogenic streptococci. Protein functional analysis of this family suggests that PHT proteins are involved in the invasion process. Finally, repression by zinc was predicted for genes encoding a variety of paralogs of ribosomal proteins. The original copies of all these proteins contain zinc-ribbon motifs and thus likely bind zinc, whereas these motifs are destroyed in zinc-regulated paralogs. We suggest that the induction of these paralogs in conditions of zinc starvation leads to their incorporation in a fraction of ribosomes instead of the original ribosomal proteins; the latter are then degraded with subsequent release of some zinc for the utilization by other proteins. Thus we predict a mechanism for maintaining zinc availability for essential enzymes.

Zinc is a component of many proteins, in particular, DNA polymerases, proteases, ribosomal proteins, etc. Thus bacteria must have effective systems of zinc transport. Two such systems, orthologous ATP binding cassette (ABC) transporters ZnuABC (YebLMI) of *Escherichia coli* and AdcABC of *Streptococcus pneumoniae*, the latter known as YcdHI-YceA in *Bacillus subtilis*, have been studied in experiments (1–4). YciABC of *B. subtilis* is a low-affinity transporter (5).

However, despite the importance of zinc, it is toxic at large concentrations, because it competes with other metals for binding to active centers of enzymes. Thus all bacteria tightly regulate zinc transport. In *E. coli* and *B. subtilis*, as well as their closest relatives, transcription of the zinc transporter genes is regulated by ZUR repressors belonging to the FUR family (6) or, in the case of *zosA* of *B. subtilis*, by the PerR repressor from the same family (7). ZUR proteins from Gram-negative and Gram-positive bacteria are distant homologs, with the average identity of 25%, and we will relate to these regulators as nZUR and pZUR, respectively. In *Streptococcus* species, zinc transporter genes are regulated by AdcR, a transcription factor from the MarR family (8).

Several ZUR-binding sites of *E. coli* and *B. subtilis* were identified in experiments (5, 6). Binding sites of AdcR are unknown. We constructed recognition profiles using known nZUR sites of γ-proteobacteria and pZUR sites from *Bacillus* spp. and identified *de novo* signals for nZUR in α-proteobacteria and for AdcR in

*Streptococcus* species. The obtained recognition rules were used to study zinc-dependent regulation in these four groups of genomes.

We described the evolution of the nZUR recognition signal in Gram-negative bacteria. We identified candidate ZUR-binding sites upstream of known and previously uncharacterized zinc transporters; in the latter case the analysis of regulation allowed us to assign zinc specificity to these transport systems. In *Streptococcus* species, we described a unique class of zinc-regulated proteins from the histidine-triad family that could be involved in adhesion and pathogenesis.

An unexpected finding was that zinc controls transcription of some paralogs of ribosomal protein genes, namely L36, L33, L31, and S14. In ref. 9 it was shown that of two or more copies of L36, L33, L31, and S14 proteins, one usually contains a predicted Zn-ribbon motif, whereas in other copies this motif is disrupted. An intriguing correlation between the presence of a Zn ribbon in a ribosomal protein and gene duplication was observed (9). However, the function of paralogous ribosomal proteins was unresolved.

Here we demonstrate the presence of candidate binding sites for zinc repressors (nZUR, pZUR, and AdcR) upstream of genes encoding paralogs of ribosomal proteins L36, L33, L31, and S14 with disrupted Zn ribbons. We suggest that these proteins act as alternative ribosomal proteins replacing the original proteins in zinc-depleted conditions. This decreases the total zinc requirement of ribosomal proteins and thus frees some zinc for the use by other proteins, improving survival during zinc starvation.

## Data and Methods

**Sequence Data.** Complete genome sequences of *E. coli* (*EC*), *Salmonella typhi* (*ST*), *Yersinia pestis* (*YP*), *Vibrio cholerae* (*VC*), *Agrobacterium tumefaciens* (*AT*), *Brucella melitensis* (*BM*), *Sinorhizobium meliloti* (*SM*), *Mesorhizobium loti* (*ML*), *B. subtilis* (*BS*), *Bacillus halodurans* (*HD*), *S. pneumoniae* (*PN*), *Lactococcus lactis* (*LL*), *Streptococcus pyogenes* (*PY*), *Staphylococcus aureus* (*SA*), *Listeria monocytogenes* (*LM*), and *Listeria innocua* (*LI*) were downloaded from GenBank (10). Partially sequenced genomes of *Bacillus stearothermophilus* (*BE*), *Streptococcus mutans* (*MU*), and *Enterococcus faecalis* (*EF*) were extracted from the ERGO database (11). Partially sequenced genomes of *Rhodobacter capsulatus* (*RC*), *Rhodobacter sphaeroides* (*RS*), *Bacillus anthracis* (*BA*), and *Klebsiella pneumoniae* (*KP*) were obtained from the web sites of the University of Chicago (http://rhodo.img.cas.cz), University of Texas Health Science Center (www.rhodobacter.org), Institute for Genomic Research (www.tigr.org), and the Washington University Consortium (http://genome.wustl.edu), respectively.

---

**GAAATGTTATAWTATAACATTTC**



**GTAATGTAATAACATTAC**
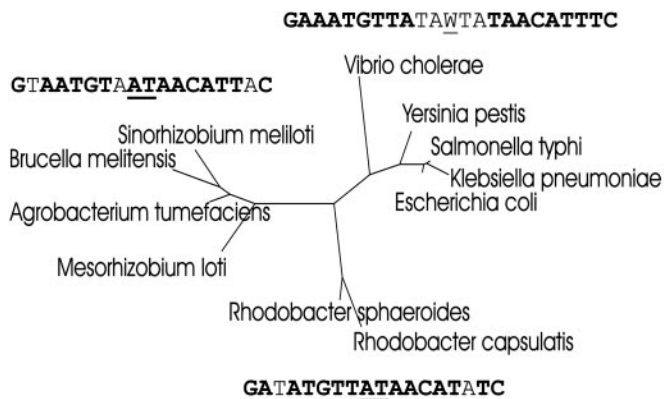
**GATATGTTATAACATATC**

**Fig. 1.** Phylogenetic tree of ZUR repressors from proteobacteria and consensuses of binding sites. Underlines, center of symmetry of inverted palindromes; boldfaced type, positions coinciding among the three signals.

**Identification of Regulatory Signals.** Common palindromic words in the upstream regions of genes likely to be regulated by the same transcription factor were identified by using the SIGNALX program (12). Positional nucleotide weights in the profiles and $Z$ scores of candidate sites were calculated as described in refs. 12 and 13.

All genomes were divided into groups based on the type of Zn-dependent regulator they contain, and genomes from each group were scanned by using the constructed profile for the respective regulator. We set the cutoffs for profiles such that they produced at most 10 candidate sites per genome in the 300-bp regions upstream of the translation start sites. A string of genes transcribed in the same direction was assumed to form a candidate operon if all intergenic spacers were shorter than 80 bp.

**Software.** Genomic analyses (protein similarity searches using the Smith–Waterman algorithm, analysis of orthology, and identification of candidate sites in genomic sequences) were done by using GENOMEEXPLORER (12). Multiple protein alignments were constructed by using CLUSTAL (14). Phylogenetic trees were constructed by using PHYLIP (15). Transmembrane segments in proteins were predicted by TMPRED (16); only scores >1,000 were considered significant in the TMPRED prediction. A protein similarity search was done by using BLAST (17), whereas functional analysis was done by using the COG (18) and InterPro (19) databases.

**Gene Names.** By default, genes in unannotated genomes were given the names of their orthologs in annotated species. Thus, in *KP*, genes were named as in *EC*; in *MU*, genes were named as in *PN*; in *BE*, *BA*, and *EF*, genes were named as in *BS*, except for the zinc ABC transporter genes *adcABC*, which had been experimentally studied in *PN* (2).

## Results

**Zinc Regulation and Transport in Gram-Negative Bacteria.** Orthologs of the nZUR protein from *EC* were found in the genomes of γ-proteobacteria *ST*, *KP*, *YP*, and *VC* and α-proteobacteria *AT*, *BM*, *SM*, *ML*, *RC*, and *RS*. The identity of candidate nZUR proteins with nZUR of *EC* is 35–45% for α-proteobacteria and 50–90% for γ-proteobacteria.

The phylogenic tree of the nZUR proteins from Gram-negative species (Fig. 1) has three major branches: the *Agrobacterium* group (*AT*, *BM*, *SM*, and *ML*), the *Rhodobacter* group (*RC* and *RS*), and the γ-proteobacteria group (*EC*, *ST*, *KP*, *YP*, and *VC*). We analyzed the nZUR-binding signals in all three groups separately.

A 23-bp palindrome GAAATGTTATAWTATAACATTTC was known to serve as an nZUR-binding site upstream of the *znuA*

gene in *EC* (6). We identified similar sites upstream of orthologous *znuA* genes of *ST* and *KP*. A recognition profile was constructed based on these three nZUR-binding sites. The profile is highly selective: only three to five candidate sites per genome score >5.00.

Table 2, which is published as supporting information on the PNAS web site, www.pnas.org, lists candidate nZUR-binding sites identified in the genomes of γ-proteobacteria. First, candidate nZUR sites were found upstream of the *znuA* genes in *YP* and *VC*. Thus, the identified signal is conserved not only within Enterobacteriaceae but also beyond this family in Vibrionaceae. Second, candidate nZUR sites were found upstream of the *b1973* gene in *EC* (new name *zinT*, see below) and its orthologs in *ST* (*STY1858*) and *KP*. No *zinT* orthologs were found in other γ-proteobacteria. Next, a candidate nZUR site was observed upstream of the *KP* operon paralogous to the *znuC-znuB* operon (we will further refer to it as *znuC2-znuB2*). The identity between two pairs of proteins is ≈30%, and *znuC2* and *znuB2* have no orthologs in other genomes. Finally, we identified candidate nZUR sites upstream of the *EC* gene *ykgM* and its orthologs in *ST* (*rpmE2*), *YP* (*YP03134*), and *VC* (*VC0878*). This gene encodes a paralog of the ribosomal protein L31 (see below).

To identify candidate nZUR-binding sites in the *Rhodobacter* species, we selected the upstream regions of the *znuA* orthologs from *RC* and *RS* and the upstream region of the *zinT* ortholog from *RS*. No *zinT* orthologs were found in *RC*. A common 18-bp palindrome GATATGTTATAACATATC with few mismatches was found in these regions (Table 2). A profile was constructed based on these three sites; however, it did not identify any new genes preceded by candidate sites in either *Rhodobacter* species.

A similar procedure was used to identify candidate nZUR-binding sites in the *Agrobacterium* group. The upstream regions of the *znuA* orthologs were taken from *AT*, *BM*, and *SM* in addition to the upstream region of the *zinT* ortholog from *AT*. No *zinT* orthologs were found in three other genomes, and no *znuA* orthologs were found in *ML*. A common 18-bp palindrome GTAATGTNATNACATTAC with few mismatches was found in all analyzed sequences (Table 2). A profile was constructed based on the identified sites. At most two genes per genome have sites scoring >5.00. Two candidate sites were identified by using this profile (Table 2). The first site is located upstream of the *SMc03799* gene (with the new name *zinL*) in *SM*. This gene is 55% identical to the *BMEII0308* gene of *BM* annotated as "low-affinity zinc transport membrane protein" (10). The second site was found upstream of the *mll8315-mll8314-mll8313* operon in *ML*. The Mll8315 and Mll8314 proteins are orthologs of ZnuC and ZnuB from *AT* with the identity of 34% and 30%, respectively. Mll8313 is a hypothetical periplasmic protein that is only 24–28% identical to ZnuA proteins from other species. Because a candidate regulatory site was identified in the upstream region of *mll8315-mll8314-mll8313* and no *znuA* orthologs were found in *ML*, we propose that *mll8313* (with the new name *zinA*) encodes a zinc-transporting periplasmic protein.

**pZUR Regulation in the *Bacillus*, *Staphylococcus*, *Listeria*, and *Enterococcus* Species.** Binding sites of pZUR in the *BS* genome were found upstream of *yciC*, *yciA*, and *ycdH* (4, 5). The pZUR-binding signal is a 23-bp palindrome, TAAATCGTAATNATTACGATTTA. We will further refer to the *ycdH* gene as *adcA*, because it is orthologous to the *adcA* gene of *PN* (4, 8).

A recognition profile was constructed based on several known pZUR-binding sites from *BS* and candidate sites from *HD*, *BE*, and *SA*. Orthologs of the *BS* pZUR protein were found in the genomes of *HD*, *BE*, *BA*, *SA*, *EF*, *LM*, and *LI*. Accordingly, these genomes were screened for candidate pZUR-binding sites (Table 2). Three to six candidate sites per genome score >5.80.

We identified candidate pZUR-binding sites upstream of the *adcA* orthologs in *HD*, *BE*, *LM*, *LI*, *SA*, and *EF*. In *LM*, *LI*, and *EF* as well as in *BS*, the *adcA* gene, encoding a lipoprotein, forms a

candidate operon with genes orthologous to *adcC* and *adcB* of *PN*, the latter two encoding an ATPase and permease components of a zinc transporter, respectively. In *HD*, *BE*, *BA*, and *SA*, the *adcC* and *adcB* genes are located separately from the *adcA* gene. Instead, they form operons with *zur* orthologs. Candidate pZUR-binding sites were found upstream of the latter operons. Further, in *LM* and *LI*, a candidate pZUR-regulated operon was found, which was composed of the *adcC* and *adcB* paralogs and the *zur* gene. In *EF*, the *zur* gene is located separately from other zinc-related genes, and it is preceded by a strong candidate pZUR-binding site. Thus, in *HD*, *BA*, *SA*, and *EF*, pZUR seems to act as an autorepressor, whereas in *BS*, *LM*, and *LI* it does not.

Candidate pZUR-binding sites were found upstream of *yciC* orthologs in *HD*, *BA*, and *SA*. We also found a candidate pZUR-binding site upstream of *BH0366* of *HD*, which encodes a paralog of YciC. Finally, we found a candidate pZUR-binding site upstream of the *s14p-l33p-yciC* candidate operon in *EF*. The first two genes of the operon encode proteins paralogous to ribosomal proteins S14 and L33. More candidate pZUR-binding sites were found upstream of genes encoding paralogs of ribosomal proteins, namely S14 paralogs in *SA*, *EF*, *LM*, and *LI* and S14, L31, and L33 paralogs in *BS* (see below).

Finally, we found candidate pZUR sites upstream of genes encoding hypothetical proteins: the *yciA* genes of *BS* (7) and *HD* and the *zinT* (old name *yrpE*) gene of *BS*. The *yciA* gene is located within the *yciC* locus, which thus has two strong candidate pZUR-binding sites. The detailed analysis of the ZinT structure and function is given below.

YciC is a GTPase of the G3E family (COG0523) (20). All but one member of this COG contain a conserved three-cysteine motif CXCC, where X is I, V, L, or M. This motif has mutated into CVSC in the most distant member of this COG, RV0106 from *Mycobacterium tuberculosis*, whereas it is still conserved in YNR029c from yeast *Saccharomyces cerevisiae*. The G3E family also contains the UreG/HypB group of proteins (INTERPRO IPR002894) containing accessory proteins for incorporation of nickel into urease (21) and hydrogenase (22). Thus it is tempting to suggest that YciC performs a similar function of zinc incorporation into some protein.

YciA belongs to COG1469, containing proteins of unknown function. All members of this COG seem to contain the pattern CPC. . . HXH, where X is L, A, P, Q, N, or R, or variants of this pattern CPH. . . HKH and CPS. . . HPH, suggestive of metal binding.

**AdcR Regulation in the *Streptococcus* Group and Structural Characteristics of Zinc-Regulated Pneumococcal Histidine Triad (PHT) Proteins.** In *PN*, zinc uptake is regulated by the AdcR protein belonging to the MarR family (8). AdcR seems to regulate transcription of the *adcR-adcC-adcB-adcA* operon; however, no AdcR-binding sites have yet been identified. We compared the upstream regions of the *adcR-adcC-adcB-adcA* operon of *PN*, the *adcR-adcC-adcB* operons and the *adcA* genes of *PY* and *MU*. A common 12-bp palindrome, TTAACYRGTTAA, was identified; moreover, in the upstream regions of the *adcA* genes of *PY* and *MU* this palindrome occurred twice. The profile constructed by using these sites selected two similar sites upstream of the *zitR-zitS-zitQ-zitP* operon of *LL* (the *zitR*, *zitQ*, and *zitP* genes are orthologous to the *adcR*, *adcC*, and *adcB* genes of *PN*, respectively; the *zitS* gene is a homolog of the *adcA* gene with 42% identity). One more *adcA* homolog of *LL*, *yndG*, is 64% identical to *adcA*, but this gene is not preceded by a similar palindrome. The constructed profile produces three to eight high-scoring (>5.00) sites per genome (Table 3, which is published as supporting information on the PNAS web site). Because the identified palindromes specifically occur upstream of genes encoding zinc-uptake proteins and some of these genes are known to be AdcR-regulated, we propose that these palindromes are AdcR-binding sites.

Two candidate AdcR-binding sites were found upstream of the *SP1002-SP1003* operon in *PN*. *SP1002* encodes a putative adhesion lipoprotein that is 67% identical to the laminin adhesion protein Lmb from *Streptococcus agalactiae*. We will further refer to SP1002 and its orthologs as Lmb proteins. Notably, it was shown recently that the *lmb* expression in *PY* decreases at high $Zn^{2+}$ ion concentration (23). The product of the *SP1003* gene (also known as *phtD*) belongs to the PHT family of proteins. Proteins of this family are characterized by multiple histidine-triad (HXXHXH) motifs (24). Three other genes of the PHT family present in the genome of *PN* are *phtA*, *phtB*, and *phtE*. They are 65–95% identical to *phtD*. The *phtA*, *phtB*, and *phtE* genes have one, two, and three high-scoring AdcR-binding sites in the upstream regions, respectively. The PhtE protein contains six histidine-triad motifs (HXXHXH), whereas each of three other proteins contains five motifs. For PhtA, PhtB, PhtD, and, to a lesser extent, PhtE, the localization on the bacterial cell surface has been shown by flow cytometry (24).

Two candidate AdcR sites were found upstream of the orthologous *lmb-phtD* operons in the genomes of *PY*, *S. agalactiae* (*AG*), and *Streptococcus equi* (*EQ*). In *S. agalactiae*, the Lmb protein was shown to mediate attachment of *S. agalactiae* to human laminin, which is essential for bacterial colonization of damaged epithelium and translocation of bacteria into the bloodstream (25). All three previously uncharacterized PhtD proteins have strong candidate hydrophobic regions at the N termini identified by the TMPRED algorithm and signal peptidase II motifs (LXXC in PhtD of *EQ* and IXXC in PhtD of *PY* and *AG*). Therefore, we expect them all to be exposed on the cell surface, similar to the PHT proteins of *PN*. Transcriptional coregulation of the *lmb* and *phtD* genes suggests a functional link between these proteins. No orthologs of either *lmb* or *phtD* genes were found in any other available genome.

We also identified an additional AdcR-regulated gene of *PY*, *SPY1361*, encoding a protein with multiple HXXHXH motifs and named it *phtY*. The PhtY protein contains an N-terminal hydrophobic region and a signal peptidase II motif (LXXC), typical for the PHT family. PhtY consists of three domains. The N-terminal domain (≈360 aa) contains four HXXHXH motifs and is 25% identical to the PhtE protein of *PN*. The second domain (≈390 aa) is 30–35% identical to a family of internalins from *LM*. The C-terminal domain consists of 42 aa and is abundant with histidine and aspartic and glutamic acids (HDE), which suggests its strong metal-chelating properties. The internalins of *LM* allow this bacterium to enter eukaryotic cells. All seven members of the listerial internalin family have been shown to share two structural features: an N-terminal leucine-rich-repeat domain (LRR), followed by a conserved interrepeat region (IR) (26). It has been demonstrated that LRR and IR are both necessary and sufficient for the internalin binding to E-cadherin, and this interaction is critical for the internalin-mediated invasion. (27, 28). Both LRR and IR are present in the N terminus of the second domain of PhtY, which implies that PhtY may play a role in the *PY* invasion.

Candidate AdcR sites were also found upstream of the *LL* and *PY* genes encoding paralogs of the ribosomal protein S14 (*rpsN2* and *rs14*, respectively; see below).

***zinT* Genes and Their Products.** In this study, six genomes contained orthologs of *zinT*, namely *BS* (a Gram-positive bacterium from the *Bacillus* group), *EC*, *ST*, and *KP* (γ-proteobacteria), and *AT* and *RS* (α-proteobacteria). Six other genomes (Gram-positive bacteria from the *Streptococcus* group, namely *PN*, *MU*, *PY*, *LL*, and *EF* as well as *SA*) had *zinT* fused to genes encoding zinc-binding lipoproteins (*adcA* orthologs). We will further refer to the zinc-transport lipoprotein domains as ADC and to ZinT domains as ZINT. The fused ADC-ZINT protein of *PN* was shown experimentally to participate in the zinc transport (2). Note that in *LL*, there are two paralogous genes similar to *adcA*, *zitS*, and *yndG*. The YndG protein consists of the ADC and ZINT domains, whereas ZitS consists of the ADC domain only.

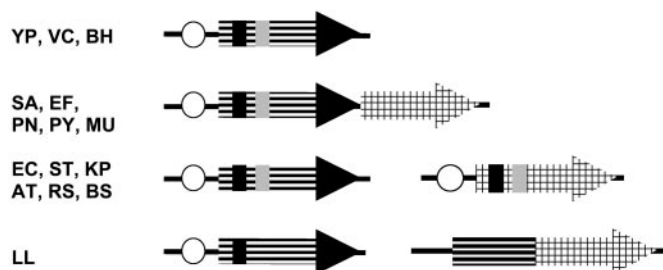Analysis by TMPRED revealed transmembrane helices at the N

**Fig. 2.** ZinT domains and regulation of *zinT* genes. Circles, candidate zinc repressor-binding sites; striped arrows, *adcA/znuA*; hatched arrows, *zinT*; black rectangles, transmembrane segment; gray rectangles, histidine/aspartate/ glutamate-rich segment.

termini of all ADC domains as well as at the N termini of ZINT domains in isolated ZinT proteins (Fig. 2). Being attached to the cell surface by the N-terminal region, a ADC-ZINT protein is totally exposed in the extracellular space, and its ZINT domain in particular, which strongly suggests that isolated ZINT domains bearing an N-terminal transmembrane helix also are extracellular and function on the cell surface. The only exception is the YndG protein of *LL*, which has no predicted transmembrane helix in either the ADC or ZINT domains.

It is well known that histidine is a rare amino acid with a strong propensity to bind metals. HDE-rich regions were observed in both ADC-ZINT proteins and single-domain ZINT proteins (Fig. 2). Notably, both transmembrane helices and HDE-rich regions were observed only in the isolated ZINT domains, whereas in the fused ADC-ZINT proteins as well as in single-domain AdcA proteins, these structural features occur in the ADC domains; the only exception is the YndG protein from *LL*, which has no HDE-rich region in either domain. This shows that the isolated ZINT proteins have acquired both these structural features and zinc-dependent regulation. Thus, these features seem to be critical for the ZinT function.

**Sequence Analysis and Regulation of Ribosomal Proteins.** As shown above, a large number of genes encoding paralogs of ribosomal proteins have been identified under candidate zinc regulation in all analyzed groups of species.

In ref. 8 it was shown that genes encoding four ribosomal proteins, L36, L33, L31, and S14, are each duplicated in several bacterial genomes. If a duplication occurs, the original copies of L36, L33, L31, and S14 contain predicted Zn-ribbon motifs that consist of two pairs of conserved cysteines (in some cases, one cysteine can be replaced by histidine), whereas in the paralogs this motif is usually lost (9). Cysteines in S14 of *Thermus thermophilus* are indeed involved in zinc binding and formation of the Zn-ribbon domain (29).

To this list we add several duplications not considered in ref. 9. The L31 protein is duplicated in *ST* and *KP*. S14 is duplicated in *SA*, *LM*, and *LI*. L33 is duplicated in *LM* and *LI*; it is triplicated in *BS*, *SA*, *PN*, *MU*, and *PY*. Finally, in *EF*, S14 is present in two copies, whereas L33 is present in four.

These duplications fit the above-stated rule: the original proteins contain predicted Zn ribbons, but the paralogs do not. The triplicated L33 proteins display the same cysteine pattern as described in ref. 9 for the L33 triplication in *LL*: one copy has four cysteines (the intact Zn ribbon), another one has three, and the last one has none. Of four L33 copies in *EF*, again one has four cysteines, another one has three, and the remaining two have no cysteine residues.

With the constructed profiles, we identified candidate nZUR-binding sites upstream of the *L31* paralogs (*L31p*) in *EC*, *ST*, *YP*, and *VC* (Table 2). Moreover, in *YP* and *VC*, *L31p* forms a candidate operon with the *L36* paralog (*L36p*); thus both ribosomal protein

**Table 1. Ribosomal proteins L36, L33, L31, and S14 and their paralogs**

| Genomes | L36 | L33 | L31 | S14 |
|---|---|---|---|---|
| *E. coli* | (−) | – | (−)+ | – |
| *S. typhi* | (−) | – | (−)+ | – |
| *K. pneumoniae* | (−) | – | (−)− | – |
| *V. cholerae* | (−)× | – | (−)+ | – |
| *Y. pestis* | (−)× | – | (−)+ | – |
| **B. subtilis** | (−) | (−)+− | (−)+ | (−)+ |
| **S. aureus** | (−) | (−)−− | − | (−)+ |
| **L. monocytogenes** | (−) | (−)− | − | (−)+ |
| **L. innocua** | (−) | (−)− | − | (−)+ |
| **E. faecalis** | (−) | (−)×−− | − | (−)+− |
| **S. pneumoniae** | (−) | (−)−− | − | (−) |
| **S. mutans** | (−) | (−)−− | − | (−) |
| **S. pyogenes** | (−) | (−)−− | − | (−)+ |
| **L. lactis** | (−) | (−)−− | − | (−)+ |

Only genomes with at least one duplication are shown. +, genes with candidate zinc repressor-binding sites in upstream regions; ×, genes in a candidate zinc-controlled operon; −, genes without zinc regulation; (−), proteins with Zn-ribbon motifs. Genomes with duplications identified in this study are shown in bold type.

paralogs in *YP* and *VC* seem to be under the nZUR regulation. Further, we observed candidate pZUR-binding sites upstream of the *L31p* and *L33p* genes in *BS* and upstream of the *S14p* gene in *BS*, *SA*, *EF*, *LM*, and *LI* (Table 2). In *EF*, *S14p* forms a candidate operon with the *L33p* gene and, notably, with the *yciC* gene encoding a component of the low-affinity zinc transporter. As mentioned above, the latter gene was shown to be regulated by pZUR in *BS* (4). Finally, we identified candidate AdcR-binding sites upstream of the *S14p* genes in *PY* and *LL* (Table 3).

Table 1 summarizes the known and new data about duplications of the Zn-ribbon ribosomal proteins in the analyzed groups of bacteria. Thus far all duplications of ribosomal proteins in the analyzed genomes except for the L31 duplication in *KP* and L33 duplication in *Listeria* fit the following rule: if one copy retains a Zn ribbon and the other one lacks it, then the gene encoding the copy without a Zn-ribbon motif is regulated by a Zn-dependent repressor and therefore is induced under Zn-restricted conditions.

The triplication of *L33* in *SA*, *PN*, *MU*, *PY*, and *LL* represent an exception, because none of the copies are regulated. However, one of three copies of *L33* in *BS* seems to be regulated by pZUR. Additionally, in *EF*, where *L33* is present in four copies, one of the no-cysteines copies is located in a zinc-regulated operon.

## Discussion

**Four Regulators That Mediate Bacterial Response to Zinc Starvation and Their Regulons in Various Microbial Species.** Zinc-dependent regulation involves four transcriptional factors in four different groups of bacterial species: nZUR in γ-proteobacteria; orthologous nZUR proteins in α-proteobacteria; pZUR in the *Bacillus* group; and AdcR in the *Streptococcus* group. ZUR-like proteins were observed also in other proteobacteria not considered here, in particular *Pseudomonas aeruginosa*, *Xylella fastidiosa*, and *Neisseria gonorrhoeae* (3). The DNA-binding signal of the nZUR repressor in γ-proteobacteria (GAAATGTTATA**W**TATAACATTTC) is conserved through the Vibrionaceae group. We identified previously unknown DNA-binding signals for the nZUR repressors in two groups of α-proteobacteria, the *Agrobacterium* (GTAATGTA-**AT**AACATTAC) and *Rhodobacter* (GATATGTT**AT**AA-CATATC) groups (Fig. 1). Further, we identified DNA-binding signal for the AdcR regulator in the *Streptococcus* group (TTAA-C**YR**GTTAA). The pZUR-binding signal in the *Bacillus* group is TAAATCGTAAT**N**ATTACGATTTA (5).

The signal of γ-proteobacteria is an inverted repeat of the highly

conserved 9-bp box GAAATGTTA, separated by a 5-bp AT-rich spacer. This box is similar to the symmetric boxes of the nZUR signals from $\alpha$-proteobacteria. In particular, the *Agrobacterium* signal has two differences (T instead of A in position 2 and A instead of T in position 8), and the *Rhodobacter* signal has only one (T instead of A in position 3). Thus it is likely that the DNA-binding surface of the nZUR proteins in these genomes has not evolved much, and the main difference, a larger spacer in $\gamma$-proteobacteria, is caused by a change in the mutual orientation of the two subunits in the nZUR dimer.

We identified a number of genes encoding various components of zinc transporters and predicted that they are regulated by zinc repressors. Among these transporters are orthologs and paralogs of the high-affinity zinc ABC transporters AdcABC/ZnuABC and the low-affinity zinc transporter YciABC. The specificity of transporters is known to evolve rapidly even in orthologous systems. Thus candidate zinc repressor-binding sites upstream of genes encoding such systems allow us to assign zinc specificity to these systems with higher confidence.

Zinc-dependent regulators act as autorepressors in most of the analyzed species. However, the distribution of autoregulation does not fit the phylogenic pattern. Indeed, of two closely related species, one may have an autoregulatory protein, whereas the other may not (e.g., *SM* and *ML* or *BS* and *HD*). At the same time, genes encoding nonautoregulatory proteins always lie isolatedly, whereas genes encoding autoregulators are usually belong to zinc transporter operons. Thus, the autoregulation of zinc repressors seems unessential and is readily lost once the regulator gene separates from the transporter genes. No autoregulation of *zur* was observed in *E. coli* (3), which is in agreement with the absence of candidate pZUR-binding sites upstream of *zur*.

Three previously uncharacterized functional classes of proteins regulated by zinc were identified: ZinT proteins, PHT proteins, and paralogs of ribosomal proteins.

### ZinT Represents a Previously Uncharacterized Type of Zinc-Binding Outer Membrane Components of ABC Transporters in Gram-Positive and Gram-Negative Bacteria. The *zinT* gene is present in bacterial genomes either as an isolated ZINT domain or a part of the ADC-ZINT fusion, where ADC is a zinc-binding component of the zinc ABC transporter. Both isolated ZINT domains and the ADC-ZINT fusions appear to be exposed to the cell surface, contain metal-chelating HDE-rich regions, and be regulated by zinc repressors. The only exception is the fusion in *LL* (the YndG protein), which has lost all three features. Candidate binding sites for zinc repressors upstream of the *zinT* and *adcA* genes strongly suggest that HDE-rich regions in the ZINT and ADC domains serve for zinc binding. These observations lead to the conclusion that ZinT is a previously uncharacterized type of zinc-binding protein, which likely functions as an alternative or additional zinc-chelating component of the zinc ABC transporter. ACD-ZINT fusions suggest that ZinT may associate with the same ATP-binding and permease proteins as AdcA.

We also suggest that the *yndG* gene of *LL* does not encode a functional protein, because it has lost both the regulatory site and the obligatory structural features. Alternatively, it might have changed the function dramatically and is not involved in the zinc acquisition any more.

### The PHT Proteins: A Candidate Family of Adhesins in Streptococci. The PHT protein family is restricted to the genus *Streptococcus*. All members of this family contain four to six histidine-triad motifs (HXXHXH) and seem to be under zinc control either directly, by candidate AdcR-binding sites in the upstream regions, or as members of the AdcR-regulated *lmb-phtD* operons. The metal-chelating property of histidine together with the predicted regulation of PHT proteins by zinc repressors suggest that histidine-triad motifs in these proteins are involved in zinc binding. However, these motifs

differ from HDE-rich regions that function as zinc-adsorption sites of zinc transporters. The latter have no regularity of histidine positions and are abundant with aspartic and glutamic acids in addition to histidine. On the contrary, positions of histidines within triads in the PHT proteins are strongly conserved, and these histidines are surrounded by aromatic amino acids rather than negatively charged ones. This leads to the conclusion that histidine-triad motifs likely play a structural or functional role in the PHT proteins.

The PHT proteins of *PN* are located on the cell surface. The *phtD* gene forms a candidate operon with the *lmb* gene encoding a putative laminin adhesion protein, which is probably involved in the colonization of the human epithelium by streptococci and their subsequent invasion into the bloodstream. The *PY lmb* mutant showed reduced laminin binding, adherence, and internalization into epithelial cells (23). The transcriptional association of the *phtD* and *lmb* genes may indicate a functional link and point to a possible role of the PHT proteins in *Streptococcus* adhesion and invasion.

This hypothesis is supported further by analysis of the AdcR-regulated PhtY protein of *PY*. PhtY consists of three domains: the N-terminal PHT domain, the middle domain that has two characteristic features of internalin proteins of the *Listeria* species (LRR and IR), and the C-terminal HDE-rich domain. The LRR and IR were shown to be both necessary and sufficient for the internalin-mediated invasion of *Listeria* species. Thus, it is likely that the PhtY protein plays a role in the invasion of *PY*. The C-terminal HDE-rich domain of PhtY might be involved in zinc scavenging before zinc incorporation into histidine-triad motifs of the PHT domain.

The first major reservoir for streptococcal infections is the human oral/nasal mucosa. After successful colonization of the nasopharynx mucosa, a small fraction of the streptococci invades the epithelium cells. Invasion is considered a multistage process initiated by adherence. As a result, invasion provides a window for streptococci to reach deeper tissues including the bloodstream (30, 31). Zinc concentration in bronchoalveolar lavages is 5- to 10-fold lower than in the human plasma (32, 33). Thus we suggest the following very speculative scenario. At the initial stage of *Streptococcus* infection in the human nasopharynx, bacteria face zinc-restricted environment and induce expression of the Lmb and PHT proteins, which are likely involved in the adhesion and invasion processes. It takes time for the PHT proteins to archive a functional conformation, because they have to bind zinc ions by the histidine-triad motifs. Indeed, in accordance with this hypothesis, zinc was shown to stimulate the protein synthesis-independent adhesion of *PY* (34). The molecular mechanism of this adhesion still remains unknown, but we expect it to be mediated by the PHT proteins. After streptococci have reached the bloodstream, they face zinc-sufficient conditions, and expression of the PHT proteins and zinc transporters is blocked. This may function as a protective mechanism that allows bacteria to avoid adhesion to macrophages and extermination by the human immune system.

### Paralogs of Ribosomal Proteins Help Bacteria to Survive During Zinc Starvation. Only four ribosomal proteins, L36, L33, L31, and S14, are duplicated in more than one bacterial genome, and the original copies of these proteins contain the Zn-ribbon motif. This motif was suggested to play a role in the ribosomal stability under high temperatures in thermophilic Archaea and bacteria (9), whereas it is not absolutely indispensable under normal conditions. The present study demonstrated that almost all genes encoding the paralogs of L36, L33, L31, and S14 proteins in the analyzed species are likely to be regulated by zinc (Table 1), and all zinc-regulated copies have lost the Zn ribbons in contrast to the original proteins that retain these motifs. Because there are four different proteins and three different regulators involved, this seems to be a general rule rather than a set of coincidences. We propose the following scenario for the L36p, L33p, L31p, and S14p function in bacteria (Fig. 3). In zinc-rich conditions, the *L36p*, *L33p*, *L31p*, and *S14p*
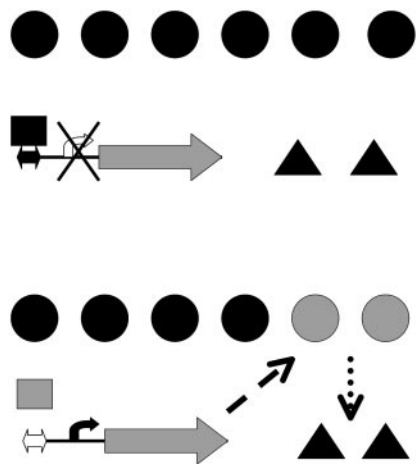
**Fig. 3.** The predicted regulation of ribosomal proteins by zinc. (*Upper*) Zn-rich conditions. The ribosomes contain the protein with functional Zn ribbon (black circles), and transcription of the gene encoding the paralog (gray arrow) is inhibited by the zinc repressor (ZUR or AcdR, black rectangle) bound to the upstream recognition site (double arrow). (*Lower*) Zn-depleted conditions. The paralog gene is expressed (broken arrow), and the protein is incorporated into a fraction of ribosomes (gray circles), thus releasing zinc (dotted arrow) for incorporation into Zn-dependent enzymes (triangles).

genes are repressed by the respective regulators, whereas the original L36, L33, L31, and S14 proteins function in ribosomes. In zinc-depleted conditions, however, the original Zn-ribbon-containing L36, L33, L31, and S14 proteins would use all available zinc, leading to zinc starvation for other cellular proteins that require zinc for their function (indeed, the number of ribosomes is orders of magnitude larger than the number of zinc-containing enzymes such as DNA polymerase, primase, etc.). To prevent this situation, non-Zn-ribbon paralogs are expressed under zinc-restricted conditions. They partially replace the original proteins, which results in releasing some zinc into the cytoplasm. This zinc then can be incorporated into other zinc-requiring proteins.

This situation resembles the correlation between atomic composition and metabolic function of enzymes (35): enzymes involved in assimilation of sulfur and thus switched on in conditions of sulfur starvation are depleted in sulfur-containing amino acids. Similarly, the carbon composition of carbon-assimilation enzymes is less than average. Similarly, zinc-independent paralog proteins are switched on in the conditions of zinc starvation.

GENETICS

1. Patzer, S. I. & Hantke, K. (1998) *Mol. Microbiol.* **28,** 1199–1210.
2. Dintilhac, A., Alloing, G., Granadel, C. & Claverys, J. P. (1997) *Mol. Microbiol.* **25,** 727–739.
3. Hantke, K. (2002) *J. Mol. Microbiol. Biotechnol.* **4,** 217–222.
4. Gaballa, A. & Helmann, J. D. (1998) *J. Bacteriol.* **180,** 5815–5821.
5. Gaballa, A., Wang, T., Ye, R. W. & Helmann, J. D. (2002) *J. Bacteriol.* **184,** 6508–6514.
6. Patzer, S. I. & Hantke, K. J. (2000) *Biol. Chem.* **275,** 24321–24332.
7. Gaballa, A. & Helmann, J. D. (2002) *Mol. Microbiol.* **45,** 997–1005.
8. Dintilhac, A. & Claverys, J. P. (1997) *Res. Microbiol.* **148,** 119–131.
9. Makarova, K. S., Ponomarev, V. A. & Koonin, E. V. (2001) *Genome Biol.* **2,** research0033.
10. Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouelette B. F., Rapp, B. A. & Wheeler, D. L. (1999) *Nucleic Acids Res.* **27,** 12–17.
11. Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., *et al.* (2003) *Nucleic Acids Res.* **31,** 164–171.
12. Mironov, A. A., Vinokurova, N. P. & Gelfand, M. S. (2000) *Mol. Biol.* **34,** 222–231.
13. Panina, E. M., Mironov, A. A. & Gelfand, M. S. (2001) *Nucleic Acids Res.* **29,** 5195–5206.
14. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgons, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
15. Felsenstein, J. (1996) *Methods Enzymol.* **266,** 418–427.
16. Hofmann, K. & Stoffel, W. (1993) *Biol. Chem. Hoppe-Seyler* **374,** 166.
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
18. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. (2000) *Nucleic Acids Res.* **28,** 33–36.
19. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., *et al.* (2003) *Nucleic Acids Res.* **31,** 315–318.
20. Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. (2002) *J. Mol. Biol.* **317,** 41–72.
21. Moncrief, M. B. & Hausinger, R. P. (1997) *J. Bacteriol.* **179,** 4081–4086.
22. Olson, J. W., Maier, R. J. & Fu, C. (1997) *Mol. Microbiol.* **24,** 119–128.
23. Elsner, A., Kreikemeyer, B., Braun-Kiewnick, A., Spellerberg, B., Buttaro, B. A. & Podbielski, A. (2002) *Infect. Immun.* **70,** 4859–4869.
24. Adamou, J. E., Heinrichs, J. H., Erwin, A. L., Walsh, W., Gayle, T., Dormitzer, M., Dagan R., Brewah, Y. A., Barren, P., Lathigra, R., *et al.* (2001) *Infect. Immun.* **69,** 949–958.
25. Spellerberg, B., Rozdzinski, E., Martin, S., Weber-Heynemann, J., Schnitzler, N., Lutticken, R. & Podbielski, A. (1999) *Infect. Immun.* **67,** 871–878.
26. Dramsi, S., Dehoux, P., Lebrun, M., Goossens, P. L. & Cossart, P. (1997) *Infect. Immun.* **65,** 1615–1625.
27. Lecuit, M., Ohayon, H., Braun, L., Mengaud, J. & Cossart, P. (1997) *Infect. Immun.* **65,** 5309–5319.
28. Mengaud, J., Ohayon, H., Gounon, P., Mege, R.-M. & Cossart, P. (1996) *Cell* **84,** 923–932.
29. Tsiboli, P., Triantafillidou, D., Franceschi, F. & Choli-Papadopoulou, T. (1998) *Eur. J. Biochem.* **256,** 136–141.
30. Cue, D., Dombek, P. E. & Cleary, P. (2000) in *Gram-Positive Pathogens*, eds. Fischetti, V. A., Novick, R. P., Ferretti, J. J., Portnoy, D. A. & Rood, J. I. (Am. Soc. Microbiol., Washington, DC), pp. 27–33.
31. Gosink, L. & Tuomanen, E. (2000) in *Gram-Positive Pathogens*, eds. Fischetti, V. A., Novick, R. P., Ferretti, J. J., Portnoy, D. A. & Rood, J. I. (Am. Soc. Microbiol., Washington, DC), pp. 214–224.
32. Harlyk, C., Mccourt, J., Bordin, G., Rodriguez, A. R. & van der Eeckhout, A. (1997) *J. Trace Elem. Med. Biol.* **11,** 137–142.
33. Bunker, V. W., Hinks, L. J., Lawson, M. S. & Clayton, B. E. (1984) *Am. J. Clin. Nutr.* **40,** 1096–1102.
34. Lee, J. Y. & Caparon, M. (1996) *Infect. Immun.* **64,** 413–421.
35. Baudouin-Cornu, P., Surdin-Kerjan, Y., Marliere, P. & Thumas, D. (2001) *Science* **293,** 297–300.