

A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family

Anastasia N. Nikolskaya and Michael Y. Galperin*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received January 31, 2002; Revised and Accepted March 31, 2002

ABSTRACT

Sequence analysis of bacterial genomes revealed a novel DNA-binding domain. This domain is found in several response regulators of the two-component signal transduction system, such as *Pseudomonas aeruginosa* AlgR, involved in the regulation of alginate biosynthesis and in the pathogenesis of cystic fibrosis; *Clostridium perfringens* VirR, a regulator of virulence factors, and in several regulators of bacteriocin biosynthesis, previously unified in the AgrA/ComE family. Most of the transcriptional regulators that contain this DNA-binding domain are involved in biosynthesis of extracellular polysaccharides, fimbriation, expression of exoproteins, including toxins, and quorum sensing. We refer to it as the LytTR ('litter') domain, after *Bacillus subtilis* LytT and *Staphylococcus aureus* LytR response regulators, involved in regulation of cell autolysis. In addition to response regulators, the LytTR domain is found in combination with MHTT, PAS and other sensor domains.

INTRODUCTION

Response regulators of the microbial two-component signal transduction systems consist of an N-terminal CheY-like receiver domain and a C-terminal output domain, typically a DNA-binding helix–turn–helix (HTH) or winged helix domain (1–3). In response to an environmental stimulus, a phosphoryl group is transferred from the sensor histidine kinase to an Asp residue in the CheY-like receiver domain of the cognate response regulator. Although the CheY-like domains do not bind DNA by themselves, phosphorylation of these domains induces conformational changes that affect the binding of the associated HTH domains to their recognition sites on the chromosomal DNA. Phosphorylation-induced conformational changes in the response regulator molecule have been demonstrated in direct structural studies (4,5).

Recent studies, including analyses of the response regulators encoded in complete bacterial genomes, illuminated a previously unseen diversity of their output domains. The HTH domains were found to fall into two main categories: NarL/FixJ family regulators contain a single HTH motif in the middle of a four-helix bundle (6), whereas regulators of the AraC/XylS family contain two HTH motifs separated by a linker helix and are

usually flanked by two additional α -helices (7). A majority of known response regulators, including the well studied OmpR and PhoB proteins, were found to contain an alternative DNA-binding domain, referred to as 'winged helix' or 'winged HTH' (8). In addition, response regulators can contain other DNA-binding domains (1,9–11), as well as alternative output domains, such as AAA-type ATPase, methyltransferase, GGDEF, EAL and HD-GYP domains, which do not bind DNA (1–3,12). Finally, alternative output domains can be combined with the DNA-binding domains, e.g. in the NtrC family of transcriptional regulators (2,4).

A detailed study of 13 response regulators encoded in the *Streptococcus pneumoniae* genome found a CheY-HTH domain organization in four of them; seven more had an OmpR-like winged helix DNA-binding domain (13). The two remaining response regulators, AgrA and BlpR, were found to have an unusual C-terminal DNA-binding domain, specific for the AgrA/ComE family (13). The importance of this domain was underscored by the observation that BlpR appeared to be the only response regulator essential for growth of *S.pneumoniae* cells (13,14).

Here, we present sequence analysis of this unusual DNA-binding domain, which shows no significant similarity to the HTH or winged-helix domains found in the previously characterized response regulators. Bacterial response regulators containing this novel DNA-binding domain comprise a well defined family of proteins that regulate production of important virulence factors, such as toxins, bacteriocins and extracellular polysaccharides.

MATERIALS AND METHODS

Initial similarity searches against the NCBI non-redundant protein database were performed using the PSI-BLAST program run to convergence with *S.pneumoniae* BlpS (NCBI protein database entry CAC03514.1) or the cytoplasmic portion of *Oligotropha carboxidovorans* CoxC (NCBI protein database entry CAB76400.1) as queries. C-terminal DNA-binding domains of various response regulators identified in these initial searches were used as queries for subsequent PSI-BLAST searches with varying parameters (because of the short size of the query proteins, to improve the sensitivity of the search, the expect value threshold was increased to 100 or more, the word size was decreased to 2, and the inclusion threshold was changed to 0.02) and for non-iterated BLAST searches against finished and unfinished microbial genome

*To whom correspondence should be addressed. Tel: +1 301 435 5910; Fax: +1 301 435 7794; Email: galperin@ncbi.nlm.nih.gov

A Gram-negative response regulators

AlgR_Psaer 142-247 **ISARTRRG**ELI**PLEV**IEFFIAD-HK**VVTLRHQAQGEV**LLDEPK**KA**EDEFG--
 AlgR_Azoto 142-247 **ISARTRRG**ELV**PVDK**AIFFIAD-HK**VVTLRHESGEV**LLDEPK**KA**EDEFG--
 PprA_Pspu 141-246 **ISARTRRG**ELI**PLQV**IYFIAD-HK**VVTLRHETGGEV**LLDEPK**KA**EDEFG--
 MrkE_Klepn 145-250 **INLIKDER**IIVT**SIHDI**YYAEAH-EK**MTFVY**TRRES**FV**MPMN**ITE**CSKLP-
 MrkE_Ecoli 139-244 **INLVKDER**IIVT**PINDI**YYAEAH-EK**MTFVY**TRRES**FV**MPMN**ITE**CSKLP-
 YehT_Ecoli 137-239 **IFGTGHSR**IYLL**QMKD**VAEVSSR-MS**SVYV**THE**2KE**GETELT**RT**IESR----

Gram-positive response regulators

LytR_Staph 141-245 **LPVEIDDK**HMI**KQCN**IIG**GTH**-NG**TTI**HITN**KY**ETEP**INR**YEKRLNP-
 LytT_Bacsu 137-241 **LALSVGES**IIVVD**TKD**IYAGTE-DG**HVVN**VK**FD**HS**YTV**SD**TLV**VEK**KL**PD-
 AgrA_Staph 143-238 **IELKRGSN**SVY**QYD**EV**MF**FESS**2SH**RLIA**HD**NR**QTE**FY**GN**K**Q**SS**Q**ID---
 VirR_Clostr 147-245 **ITIKDKNN**V**KIR**TE**DIL**FL**ET**F-ER**KVI**HT**NS**Q**YV**K**MS**N**K**E**K**ELNN-
 EntR_Enter 146-243 **FVFSIGS**Q**TF**FD**INDI**YF**VEAS**2**PH**RL**SLC**TK**DG**Q**YF**Y**GR**ISE**LE**K**Y**---
 PlnC_Lacto 144-247 **LCYKIGTR**FF**SP**IN**DVIM**L**STN**3**PGS**IR**LTA**KN**KVAD**FP**GN**IN**SE**EN**KY**---
 PlnD_Lacto 144-247 **FNYKLGTR**Y**FS**L**ALD**V**ILL**ST**S3**PG**S**Q**LH**A**INK**V**AE**FP**GN**IN**AL**E**KY**---
 SapR_Lacto 146-247 **FTFSIGS**Q**V**FS**IK**RE**IL**FI**ESS**2**PH**R**VIL**N**K**NG**H**Y**F**Y**GN**IN**D**L**SE**K**Y**---
 BlpR_Spneu 142-243 **FYFKSK**FA**Q**Q**Y**PF**K**EV**Y**LE**TS**2**AHR**V**IL**Y**TK**DR**LE**FT**AS**EE**V**FK**QE**---
 ComE_Spneu 142-246 **EDNYK**GN**D**IK**I**PY**HD**I**LY**ET**T2**SH**K**L**RI**IG**K**NF**AK**E**F**Y**GT**MT**D**IQ**E**K**D**H**T**

Other

BlpS_Spneu 4-109 **MI**L**Q**T**Q**K**T**V**V**K**V**N**ID**I**Y**I**Y**I**Q**T**H**3**AH**V**Q**I**V**T**E**E**AS**E**N**M**L**Q**N**E**S**N**E**Q**C**G--
 Lmo0984_Lm 42-146 **ISVKK**D**GAT**Y**LL**E**PKA**I**LY**F**EAV**-E**SK**I**F**V**Y**T**E**K**E**I**E**I**H**W**K**Y**E**L**E**E**K**F**K**E-
 SP0161_Spn 46-146 **IKG**K**ID**Q**CV**L**VE**I**G**K**I**Q**R**Y**IE**-N**R**K**V**L**AE**T**AS**Q**T**Y**S**I**D**L**R**Y**Q**V**L**K**L**P--
 SA2153_Sau 44-147 **LVG**I**D**K**E**I**H**I**N**V**S**D**V**I**T**E**Q**V**I**-N**K**N**V**T**A**I**T**S**N**Q**K**E**K**L**R**E**Y**E**L**E**K**Q**L**P--
 BH3894_Bha 238-344 **IPAK**V**NE**K**ML**F**D**P**T**E**ID**Y**IESH**-E**G**F**AN**L**H**V**K**G**Q**V**L**C**T**M**T**I**N**E**L**D**GR**L**K**P-
 CoxC_Oligo 290-393 **LE**V**E**K**Y**H**G**S**K**V**I**N**T**A**D**I**F**S**V**H**AN**-A**H**Y**V**V**F**NG**RED**I**F**C**PL**S**IG**E**I**R**L**P**F**-
 CoxH_Oligo 280-385 **LF**E**K**NG**Q**R**CT**M**L**I**S**R**L**F**AV**O**PO**-A**H**Y**T**L**L**D**GE**A**T**W**F**C**PL**S**IS**O**V**A**R**V**L**D**S**-
 RpfD_Xanth 190-294 **FL**V**R**L**G**R**D**FL**V**A**AD**I**E**W**L**O**AS**-G**N**Y**V**N**L**H**R**G**H**D**Y**PL**R**ST**MT**A**E**A**K**L**D**P-
 CC0551_Cor 407-510 **LT**V**AS**A**R**G**V**E**LV**PL**A**E**IL**S**V**V**G**A-D**D**V**EL**I**R**V**2**R**S**L**L**H**A**A**R**E**D**A**T**T**Q**L**P**--



Gram-negative response regulators

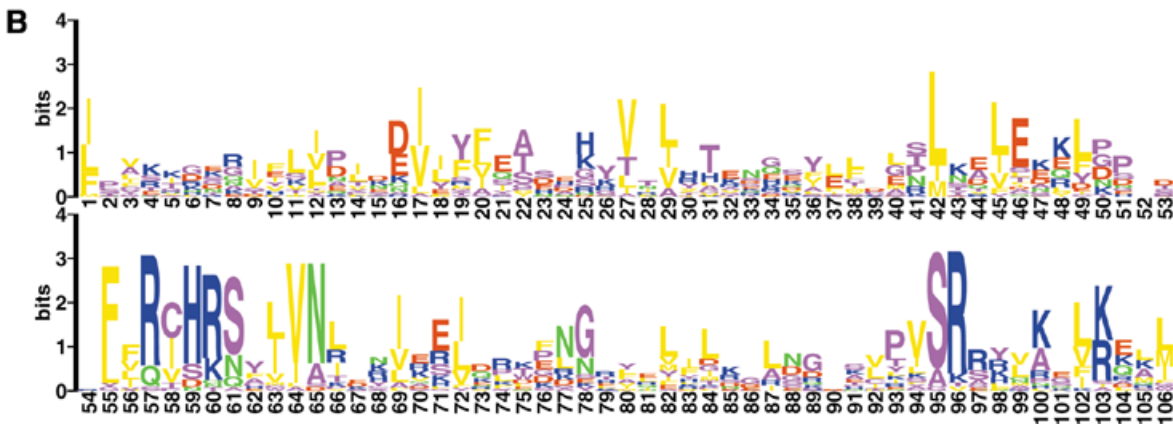
AlgR_Psaer ER**FVR**HR**N**AL**V**ARER**ER**L**Q**RT**PL**G**H**F**Q**L**Y**L**K**G**L**D**G**DA**AT**V**S**RR**H**V**A**G**V**R**R**IM**H**Q 15600454
 AlgR_Azoto DR**FVR**HR**N**AL**V**ARDR**ER**L**Q**RT**PL**G**H**F**Q**L**Y**L**R**G**L**D**G**DA**AT**V**S**RR**H**V**A**G**V**R**K**LM**H**N 3746668
 PprA_Pspu DQ**FVR**HR**N**AL**V**ARER**ER**L**Q**RT**PL**G**H**F**Q**L**F**L**K**G**L**D**G**DA**AT**V**S**RR**H**V**A**G**V**R**K**MM**Q**T 805010
 MrkE_Klepn AH**FFR**HR**S**Y**C**V**N**SK**IRE**IE**P**W**F**N**T**Y**V**I**R**L**R**D**L**E-F**Q**V**P**V**S**R**S**K**V**KE**F**R**Q**LM**N**L WUGSC
 MrkE_Ecoli SH**FFR**HR**S**Y**C**V**N**SK**IRE**IE**P**W**F**N**T**Y**I**I**R**L**K**D**L**D-F**E**V**P**V**S**R**S**K**V**KE**F**R**Q**LM**H**L 1799792
 YehT_Ecoli T**PE**L**R**HR**Q**Y**L**V**N**A**HL**Q**E**I**R**L**E**D**N**G**O**A**E**L**I**L**R**--N**G**L**T**V**P**S**R**R**Y**L**K**S**L**KE**A**I**G**L 13878946

Gram-positive response regulators

LytR_Staph TY**F**I**R**HR**S**Y**I**N**T**K**H**KE**V**Q**Q**W**F**N**T**Y**M**V**I**L**T**--N**G**V**K**V**Q**C**R**S**F**M**K**D**F**K**A**S**I**G**L** 1854577
 LytT_Bacsu SD**F**I**R**HR**S**Y**F**V**N**T**E**Y**KE**I**Q**W**F**N**S**T**Y**N**L**I**M**K--D**G**S**K**I**P**V**S**R**T**Y**A**K**E**L**K**L**L**H**I** 16079944
 AgrA_Staph DR**FFR**HR**S**Y**F**V**N**R**H**N**S**S**I**N**S**KE--R**V**V**Y**F**K**---N**G**E**H**C**Y**A**S**V**R**N**V**K**V**---- 5802562
 VirR_Clostr KG**FFR**HR**S**Y**I**V**N**I**K**I**E**E**K**D**---**Y**L**L**I**---N**K**F**T**L**P**S**K**H**R**M**K**N**L**K**L**R**L**T**S** 498841
 EntR_Enter P**M**L**T**R**S**R**A**C**L**V**N**E**F**N**KE**I**D**F**K**--R**S**L**Y**F**D**-----S**E**L**A**R**N**F**L**G**K**A**Q**K**I**E 4760906
 PlnC_Lacto S**Q**F**F**R**D**K**S**L**V**N**I**D**Y**D**S**Y**D**Y**Q**K**K**E**L**T**M**I**D**N-I**C**S**V**S**Y**R**K**S**R**E**I**N**K**I**L**K**K**--- 1495671
 PlnD_Lacto P**Q**F**F**R**D**K**S**L**V**N**I**N**H**R**S**F**D**Y**KE**K-E**L**L**L**D**G**E**I**R**C**K**A**S**F**R**K**S**R**E**I**N**K**L**R**D**N**--- 1495672
 SapR_Lacto P**Y**L**F**R**N**R**N**C**L**V**N**L**K**N**I**E**I**D**F**K**S**--R**E**I**L**F**G**---S**D**L**S**R**K**F**A**R**G**S**N**Q**L**K**R**A**F**S**Q** 1363369
 BlpR_Spneu P**R**L**L**Q**CHR**S**F**L**I**N**E**A**N**V**H**L**D**K**KE**--K**L**L**F**F**P**---N**G**G**S**C**L**I**A**R**Y**K**V**R**E**V**S**E**A**I**N**K 15900440
 ComE_Spneu Q**R**F**Y**S**H**K**S**F**L**V**N**I**G**N**I**R**E**I**DR**K---N**L**E**I**V**E**Y--E**D**H**R**C**P**I**S**R**L**K**I**R**K**L**D**I**E**K 15902038

Other

BlpS_Spneu E**T**L**M**R**HR**N**C**L**V**N**L**D**K**I**K**S**DF**Q**E**--R**I**L**F**L**E**G**E**-G**Q**Y**A**K**Y**A**R**R**Y**R**E**I**R**Q**K**W**L**K 15900439
 Lmo0984_Lm S**S**F**F**R**S**K**S**M**I**L**N**E**W**E**K**I**A**P**G**F--N**G**K**E**A**K**L**L**--N**R**E**K**V**I**I**S**R**Q**Y**A**K**V**L**R**Q**K**L**N**M 16410386
 SP0161_Spn T**N**F**I**Q**S**O**S**E**I**N**DS**I**SH**L**K**L**T**P--N**G**L**V**E**I**F**L**K--N**E**S**F**T**Y**S**R**R**Y**L**K**I**E**K**E**L 14971628
 SA2153_Sau Q**H**F**I**R**S**K**S**E**I**V**N**K**Y**E**K**L**L**L**E**P--N**G**L**I**R**M**Y**L**K--D**A**H**Y**T**S**S**R**R**Y**L**K**S**I**K**E**R**L**S**I** 15927943
 BH3894_Bha F**G**F**F**R**CHR**S**Y**L**V**N**L**Q**K**V**R**E**V**M**T**W**S**R**N**S**Y**L**I**L**D**T**A**Q**H**D**S**I**PL**S**K**R**L**A**E**L**K**Q**L**I**G**I 10176519
 CoxC_Oligo D**T**F**F**R**CHR**S**Y**I**N**T**H**C**V**A**R**L**K**R**A**G**D**N**G**I**A**E**L**D**S**P--V**R**R**S**V**P**S**R**S**L**P**Q**L**R**E**Q**L**R**E 7159966
 CoxH_Oligo A**N**F**A**Q**CHR**S**H**I**D**L**D**R**E**R**L**V**R**G**T**G**N**G**M**E**A**I**S**K--T**P**Y**K**V**PL**S**R**A**R**A**W**I**Q**Q**L**O**V** 7159970
 RpfD_Xanth A**V**E**V**R**CHR**S**Y**I**V**N**L**G**C**V**S**T**E**P**L**D**A**G**E**A**R**V**HL**S--G**N**G**P**L**P**C**S**R**R**Y**L**A**S**L**R**R**A**A**G**H 7228151
 CC0551_Cor V**S**F**L**R**CHR**S**V**I**A**N**T**H**A**Q**R**L**E**R**D**G**D**--R**W**R**L**H**L**N**E**-G**P**-F**L**P**V**S**R**S**R**P**A**V**R**E**A**L**A**D 16124806



sequences (<http://www.ncbi.nlm.nih.gov/BLAST>). The multiple alignment of the representative sequences, retrieved in these searches, was generated by PSI-BLAST, followed by minimal manual editing. A major part of the LytTR domain (missing the first one or two β -strands) is listed in the ProDom database (<http://prodes.toulouse.inra.fr/prodom/doc/prodom.html>) as domain PD032235. Response regulators containing the LytTR domain comprise COG3279 in the COG database (<http://www.ncbi.nlm.nih.gov/COG/>).

Secondary structure predictions were performed using the full set of programs available at the PredictProtein (<http://cubic.bioc.columbia.edu/predictprotein/>), JPred (<http://jpred.ebi.ac.uk/>) and UCSC HMM (<http://www.cse.ucsc.edu/research/compbio/>) web servers. The pI values were calculated using the Compute pI/Mw tool at the ExPASy web server (<http://www.expasy.org/>).

The consensus sequence of the LytTR-binding site was deduced from a library of the upstream regulatory DNA sequences that have been experimentally demonstrated to bind response regulators with the LytTR domain. This library included AlgR-binding sites RB1, RB2 and RB3 for *Pseudomonas aeruginosa algD* gene (15,16) and the AlgR-binding site of the *algC* gene; sequences upstream of *Lactobacillus plantarum plnABCD*, *plnEFI* and *plnJKLR* operons that bind PlnC and PlnD (17); *Lactobacillus sake sppA*, *sppIP*, *sppT_1* and *sppT_3* probes that bind SppR, PlnC and PlnD (18); probes DR1, DR2 and DR3 with sequences upstream of *S.pneumoniae comAB*, *comCDE* and *orf12* genes that bind ComE (19); and two probes covering the upstream region of *Clostridium perfringens pfoA* gene that binds VirR (20). With this library as an input, the SignalX program (21), a module of the GenomeExplorer program suite (22), was used to generate the profiles of monomeric and tandem LytTR-binding sites and to use these profiles to search for the likely LytTR-binding sites in selected DNA sequences.

RESULTS

LytTR, a novel conserved DNA-binding domain

PSI-BLAST searches of the NCBI non-redundant protein sequence database using the B1pS protein of *S.pneumoniae* as a query converged after five to eight iterations (depending on the search parameters used), retrieving a set of 94 proteins (as of January 1, 2002) with several characteristic sequence motifs (Fig. 1). Most of those hits comprised C-terminal portions of bacterial response regulators, including such well characterized ones as *P.aeruginosa* AlgR, *Staphylococcus aureus* AgrA and LytR, *S.pneumoniae* ComE and *C.perfringens* VirR (Table 1). Since these response regulators all contain N-terminal CheY-like domains, which do not interact directly with DNA, their

C-terminal parts must be responsible for their well documented binding to specific DNA sequences in the upstream regions of target genes (9,10,15–20). Using the C-terminal domains of any of these response regulators as PSI-BLAST queries each time retrieved the same set of 94 proteins, indicating that these domains form a tight family, unrelated to any other DNA-binding domains. Because of a certain confusion with the nomenclature of individual proteins of this family, we refer to this conserved DNA-binding domain found in *Bacillus subtilis* protein LytT and *S.aureus* transcriptional regulator LytR as LytTR ('litter') domain, in line with the designation of the *Staphylococcus lugdunensis* synergistic hemolysin locus as *slush* (23). [In addition to the *S.aureus* LytR protein, the name LytR is also used for another group of transcriptional regulators (e.g. *B.subtilis* LytR), which are unrelated to members of the LytTR family (see COG1316 in the COG database, <http://www.ncbi.nlm.nih.gov/COG/>). Likewise, none of the proteins encoded in the *comE* operon of *B.subtilis* (ComEA, ComEB or ComEC) are related to the streptococcal ComE proteins or belongs to the LytTR family. Finally, the VirR protein from *Agrobacterium tumefaciens* is also unrelated to the LytTR family proteins with the same name. Instead, it belongs to a large family of uncharacterized 'VirR-related' proteins that comprise COG1720 in the COG database and UPF0066 in PROSITE, <http://www.expasy.org/prosite/>.] Accordingly, we refer to transcriptional regulators that contain the LytTR domain (Table 2) as the LytTR family proteins.

Sequence analysis of the LytTR domain indicates that it is most likely composed of four β -strands, followed by two α -helices, a β -strand and one more α -helix (Fig. 1A). The most conserved sequence motif FhRhH[RK][SNQ]hhVN (where 'h' indicates a hydrophobic amino acid residue) is located at the beginning of the second α -helix in the region that could potentially comprise a HTH motif (Fig. 1A and B). However, the variable distance between the two α -helices in different proteins and a lack of sequence conservation in this spacer region (Fig. 1A), argue against its role in DNA binding. In addition, some of the programs used predict, in this region, a short β -strand, while others predict a single long α -helix, which makes the structural assignment in this region somewhat questionable. In any case, the predicted secondary structure of the LytTR domain is not similar to any protein with known three-dimensional structure, indicating that it comprises a novel type of a DNA-binding domain.

The C-terminus of the LytTR domain contains a cluster of Lys and Arg residues (Fig. 1B) that confer strong positive charge to the whole domain, bringing its pI into the 9.0–11.0 range. This positive charge has been previously proposed to play a role in the DNA binding by the response regulators of the AgrA/ComE family (10). However, the LytTR domain appears to bind DNA in a sequence-specific fashion, interacting with direct

Figure 1. (Opposite) Sequence alignment of the LytTR-type DNA-binding domains. (A) LytTR-type domains were aligned on the basis of the PSI-BLAST outputs. The sequence of *K.pneumoniae* MrkE protein [SWISS-PROT accession no. P21649 (32)] has been modified according to the genomic data from the Washington University *K. pneumoniae* Genome Sequencing Project. Protein names, species abbreviations and the GenBank accession nos are as in Tables 1 and 2; the last column shows the protein 'gi' numbers in the NCBI protein database. Conserved amino acid residues that could be involved in DNA binding are shown in bold and colored as follows: blue, Arg, Lys, His; red, Asp, Glu; green, Asn and Gln; purple, other conserved residues. Yellow shading indicates conserved uncharged amino acid residues. The predicted secondary structure is a combination of the assignments produced by the PhD (33) and SAM-T99 (34) programs. (B) Consensus sequence of the LytTR domain, drawn from a multiple alignment of the 96 LytTR domains using the SeqLogo program (35) in the WWW-based implementation by S. E. Brenner (<http://www.bio.cam.ac.uk/seqlogo>). The first residue corresponds to Leu127 of the *P.aeruginosa* AlgR protein. The height of each letter indicates the degree of its conservation, the total height of each column reflects the statistical importance of the given position. The coloring is as in (A).

Table 1. Properties of the response regulators of the LytTR family^a

Organism	Protein name	GenBank accession no.	Regulated processes (genes/operons)
Biosynthesis of extracellular polysaccharides			
<i>Pseudomonas aeruginosa</i>	AlgR	AE004938	Alginate biosynthesis (<i>algC</i> , <i>algD</i>), twitching motility
<i>Azotobacter vinelandii</i>	AlgR	AF077237	Alginate biosynthesis, encystment
<i>Pseudomonas syringae</i>	AlgR1	AF131199	Alginate biosynthesis
Fimbrial expression			
<i>Klebsiella pneumoniae</i>	MrkE	M55912	Expression of type 3 fimbriae
Peptidoglycan turnover			
<i>Staphylococcus aureus</i>	LytR	L42945	Autolysis (<i>lrgAB</i>)
Virulence			
<i>Clostridium perfringens</i>	VirR	D14877	Production of perfringolysin O (<i>pfoA</i>), collagenase, hemagglutinin
Quorum-sensing systems involving autoinducing peptides			
Virulence			
<i>Staphylococcus aureus</i>	AgrA	M21854	Expression of toxins, hemolysins, staphylo-kinase, other secreted proteins (<i>vir</i> factors)
<i>Staphylococcus epidermidis</i>	AgrA	Z49220	Expression of exoproteins (<i>vir</i> factors)
Competence			
<i>Streptococcus pneumoniae</i>	ComE	U76218	Competence (<i>comCDE</i> , <i>comAB</i> , <i>natAB</i>) ^b
<i>Streptococcus gordonii</i>	ComE	X98109	Competence
<i>Streptococcus mitis</i>	ComE	AJ000871	Competence
Biosynthesis of bacteriocins			
<i>Carnobacterium piscicola</i>	CbaR	AF207838	Carnobacteriocin A production
<i>Enterococcus faecium</i>	EntR	AF099088	Enterocin A production
<i>Lactobacillus plantarum</i>	PlnC	X94434	Plantaricin A production (<i>plnABCD</i> , <i>plnEFI</i> , <i>plnJKLR</i>)
	PlnD	X94434	
<i>Lactobacillus plantarum</i>	PlsR	Y15127	Plantaricin S production
<i>Lactobacillus sakei</i>	SppR	Z48542	Sakacin P production (<i>sppA</i> , <i>sppIPKR</i> , <i>sppTE</i>)
<i>Lactobacillus sakei</i>	SapR	Z46867	Sakacin A production (<i>sapA</i> , <i>sapK</i>)
<i>Staphylococcus warneri</i>	ORF1	AB034941	Nukacin ISK-1 production
<i>Streptococcus pneumoniae</i>	BlpR (RR13)	AJ276410	Bacteriocin-like peptide production, essential for survival

^aAn expanded version of this table, which includes additional proteins, accession numbers and references, is available as Supplementary Material.

^bSequence analysis of the *S.pneumoniae* orf12 operon showed that these genes are homologs of *B.subtilis* *natAB* genes, encoding a Na⁺-transporting ATPase (data not shown).

repeats of 10–12 bp with a relatively well defined consensus sequence (see below). In addition, there are several exceptions, including the *B.subtilis* LytT protein whose LytTR domain has

a *pI* of 6.5. These observations argue against direct involvement of those Lys and Arg residues in DNA binding by the LytTR domain.

Table 2. Domain organization of various LytTR-containing proteins

Domain organization ^a	Total no. of proteins ^b	Examples [organism, protein name (accession no.)] ^c
Cytoplasmic		
CheY-LytTR	76	<i>Escherichia coli</i> MrkE (BAA16251), YehT (P33356), <i>Bacillus subtilis</i> LytT (CAB14852), those listed in Table 1
LytTR only	2	<i>Streptococcus pneumoniae</i> BlpS (AAK74683), <i>Pseudomonas</i> phage D3 Orf50 (AAF80809)
LMO-LytTR	5	<i>Lactococcus lactis</i> L121252 (AAK04633), <i>Listeria innocua</i> Lin0983 (CAC96214), <i>Listeria monocytogenes</i> Lmo0984 (CAC99062), <i>Staphylococcus aureus</i> SA2153 (BAB43455), <i>Streptococcus pneumoniae</i> SP0161 (AAK74343), <i>Streptococcus pneumoniae</i> SP1915 (AAK75983)
ABC-LytTR	1	<i>Bacillus halodurans</i> BH3894 (AP001520)
PAS-LytTR	–	<i>Burkholderia cepacia</i> ^d , <i>Geobacter sulfurreducens</i> ^d
Membrane-bound		
MHYT-LytTR	2	<i>Oligotropha carboxidovorans</i> CoxC (CAB76246), CoxH (CAB76400)
MHYE-LytTR	3	<i>Xanthomonas campestris</i> RpfD (CAB77034), <i>Caulobacter crescentus</i> CC1610 (AAK23589), <i>Mesorhizobium loti</i> mll0891 (BAB48381)
3TM-LytTR	1	<i>Caulobacter crescentus</i> CC0295 (AAK22282)
4TM-LytTR	3	<i>Agrobacterium tumefaciens</i> AGR_C_3954p (AAK87922), <i>Caulobacter crescentus</i> CC0330 (AAK22317), CC3036 (AAK24998)
8TM-LytTR	1	<i>Caulobacter crescentus</i> CC0551 (AAK22538)

^aABC, the ATP-binding domain of ABC-type transporters (ATP-binding cassette); LMO, an uncharacterized 40 amino acid domain with a coiled-coil region (Fig. S2A in the Supplementary Material); MHYE, a conserved membrane-associated domain (Fig. S2B in the Supplementary Material); 3TM, 4TM and 8TM, uncharacterized domains with, respectively, three, four and eight predicted transmembrane segments. PAS and MHYT domains have been described earlier (25,36).

^bThe total number of proteins with the given domain organization in the NCBI protein database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>) as of January 1, 2002. Many of those proteins are copies of the same protein sequenced from different species or even different strains of the same species.

^cAccession numbers for the NCBI, EBI and DDBJ protein databases.

^dUnfinished genome sequences from the DOE Joint Genome Institute (*B.cepacia*) and The Institute for Genomic Research (*G.sulfurreducens*).

Alternative domain combinations involving LytTR domain

In addition to numerous response regulators with CheY-LytTR domain organization, the LytTR domain has been found in other contexts (Table 2). The deep-water marine bacterium *Bacillus halodurans* encodes a unique protein, combining an ATP-binding domain of ABC-type transporter family with a C-terminal LytTR domain. Such a domain combination has not been seen elsewhere and the regulatory signal transmitted by this protein remains enigmatic.

In CoxC and CoxH proteins from the CO-metabolizing bacterium *Oligotropha carboxidovorans* (24), the LytTR domain can be found in association with MHYT, a recently described sensor domain, which contains six transmembrane segments carrying conserved His and Met residues (25). The *coxC* and *coxH* genes flank the operon encoding the CO dehydrogenase and are involved in transcriptional regulation of the autotrophic metabolism in *O.carboxidovorans* (24). A LytTR domain-containing protein is also encoded upstream of the CO

dehydrogenase-encoding genes in *Hydrogenophaga pseudo-flava* (data not shown).

Several Gram-positive bacteria combine the LytTR domain with a short (~40 amino acids) poorly conserved N-terminal domain (see Table 2, and Fig. S2A in the Supplementary Material), dubbed LMO ('elmo'). These proteins are likely to be transcriptional regulators whose binding to DNA is controlled by protein-protein interactions involving the LMO domain.

The RpfD protein from the plant pathogen *Xanthomonas campestris* (26) combines the LytTR domain with an N-terminal membrane-bound domain that consists of three predicted transmembrane segments with conserved Glu, Asp, His and Tyr residues (see Fig. S2B in the Supplementary Material). The membrane topology of this domain is somewhat similar to that of the MHYT domain (25), suggesting that it, too, might be membrane-bound and metal-binding. We refer to it as the MHYE domain. Although the *rpfD* gene is located in the *rpf* (regulation of pathogenicity factors) locus of

X.campestris, mutants with transposon insertion in the *rpfD* gene did not exhibit any discernible phenotype (26), just as MHYT-less mutants of *B.subtilis* (25). While the function of this protein still remains unknown, it is worth noting that very similar proteins are encoded in the genomes of *Caulobacter crescentus* and *Mesorhizobium loti* (Table 2).

The binding site for the LytTR domain

Studies of the DNA binding by response regulators of the AgrA family established that these proteins bind to imperfect direct repeats of the sequence pattern [TA][AC][CA]GTTN[AG][TG] separated by a 12–13 bp spacer (10,14,17–19). Remarkably similar binding sites were also identified in front of AlgR- and VirR-regulated promoters (15,20). Using an alignment of the experimentally characterized LytTR-binding sites in a profile-based search with the SignalX program (21) identified potential LytTR-binding sites in the upstream region of the *S.aureus* *lrgAB* operon (27) and in front of *coxC*, *coxD* and *coxH* genes (24) of the *O.carboxidovorans* *cox* gene cluster (data not shown). One of the most interesting findings from these searches was the prediction of the fourth LytTR-binding site in the near-upstream region of the *P.aeruginosa* *algD* gene. This site (–71 TCCGTTTGA –79) is located next to the previously identified RB3 AlgR-binding site (16) and can actually be seen as the second AlgR-protected DNA region on figure 3 in Mohr *et al.* (16). The adjacency of these two AlgR-binding sites suggest that they might function together in binding an AlgR dimer in a manner similar to that of other LytTR-containing response regulators. A detailed analysis of the potential LytTR-binding sites in completely sequenced bacterial genomes will be reported elsewhere.

DISCUSSION

The LytTR family of response regulators

Involvement of the transcriptional regulators that contain the LytTR domain in the pathogenesis of cystic fibrosis (*P.aeruginosa* AlgR), toxic shock syndrome (*S.aureus* AgrA), gas gangrene (*C.perfringens* VirR) and other deadly diseases has ensured sustained interest in their structure and function. However, a detailed analysis of these proteins has been hampered by the remarkable sequence conservation of their N-terminal CheY domain, which largely masked (dis)similarities of the DNA-binding domains of various response regulators. When Shimizu *et al.* (28) sequenced *C.perfringens* *virR* gene, they noticed the similarity of its product to MrkE, AlgR and AgrA response regulators, but paid most attention to their common N-terminal CheY domains. Likewise, Brunskill and Bayles (29) aligned *S.aureus* LytR with YehT, AgrA, AlgR and MrkE, but limited that alignment to the CheY portion of the protein. Nevertheless, several previous studies reported the lack of sequence similarity of the LytTR domain to the HTH or winged-helix DNA-binding domains and succeeded in assigning LytTR-containing response regulators to two families. On one hand, Whitchurch *et al.* (30) linked AlgR with YehT and LytR in a new family of response regulators and Mizuno (11) identified *Escherichia coli* YehT and MrkE (o244) as response regulators with unique output domains. On the other hand, Pestova *et al.* (9) identified the *S.pneumoniae* ComE response regulator and demonstrated its similarity to the

quorum-sensing regulators AgrA, CbnR, PlnC and PlnD and to the *C.perfringens* VirR. Likewise, Grebe and Stock (1) assigned AlgR, LytR, LytT and MrkE to the RC3 group of response regulators with the LytR-type output domain; in contrast, AgrA, CbnR, ComE, PlnC, PlnD, PlsR and SppR were assigned to the RD group of response regulators with the ComE-type output domain. Here, we have unified the AlgR/LytR and AgrA/ComE families into one and identified other proteins with the LytTR domain that are encoded in complete microbial genomes.

Phylogenetic distribution of the LytTR domain

Although transcriptional regulators with the LytTR domain are widely distributed among the bacteria of the *Bacillus/Clostridium* group (low G+C Gram-positive bacteria) and are found, albeit usually in just one copy, in many proteobacteria, they are conspicuously missing in other groups of Gram-negative bacteria, in archaea and in eukaryotes (see Table S3 in the Supplementary Material). Although the LytTR domain is missing in *Mycobacterium tuberculosis* and *Mycobacterium leprae*, response regulators with this domain are encoded in other actinobacteria (high G+C Gram-positive bacteria), such as *Streptomyces coelicolor* and *Rhodococcus ruber*, as well as in the unfinished genome of *Mycobacterium smegmatis* (data not shown). This skewed phylogenetic distribution generally correlates with the genome size, as only those proteobacteria with genomes >2 MB (i.e. encoding at least 2000 proteins) encode a LytTR domain. Thus, although LytTR-containing transcriptional regulators play important roles in regulating virulence factors, this domain appears to be more typical for free-living bacteria and non-obligate parasites that tend to have relatively large genomes. This correlates with the involvement of LytTR-containing transcriptional regulators in complex behavioral responses, such as quorum sensing (9,18).

It is worth noting that in α -proteobacteria, the LytTR domain is found in combinations with various membrane-bound domains, including all five instances of this domain in *C.crescentus* (Table 2). Transcriptional regulation mediated by these proteins should occur only when the corresponding target genes are positioned in close vicinity to the cell membrane and may account for the temporary expression of those genes, which could be particularly important for the complex cell cycle of *C.crescentus*.

Is LytTR a HTH domain?

As discussed above, the LytTR domain consists of three (predicted) α -helices and four (predicted) β -strands with a potential additional short β -strand between the first and second α -helices (Fig. 1A). Remarkably, some of the available structure prediction programs predict that the conserved F[FYVL][RQ][CIV] motif forms an additional β -strand, while others predict that these residues belong to very long α -helix, extending into the α -helix shown in Figure 1. Thus, although different programs disagree on the exact structure of this region, they all consider a loop (or turn) structure in that region unlikely. A possibility of helix-to-strand conversion has been reported in the DNA-binding region of the transcriptional regulator Fis (31). In any case, this region has no apparent similarity to the turn regions of the classical HTH motifs and can be confidently predicted to comprise a novel type of DNA-binding domain. Of course, all secondary structure

assignments discussed above are just predictions that need to be verified by determining the three-dimensional structure of the LytTR domain (currently in progress). Knowledge of the LytTR structure should also help in refining its binding site, shedding some light on the range of cellular functions regulated by this remarkable system.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Eugene Koonin and Igor Rogozin for helpful suggestions, Mikhail Gelfand and Andrey Mironov for letting us use an updated version of their SignalX program, and Chester Price for critically reading the manuscript. We acknowledge the availability of unfinished genome sequences from the Institute for Genomic Research (*Geobacter sulfurreducens*), Washington University (*Klebsiella pneumoniae*) and the DOE Joint Genome Institute (*Burkholderia cepacia*).

REFERENCES

- Grebe, T.W. and Stock, J.B. (1999) The histidine protein kinase superfamily. *Adv. Microb. Physiol.*, **41**, 139–227.
- Stock, A.M., Robinson, V.L. and Goudreau, P.N. (2000) Two-component signal transduction. *Annu. Rev. Biochem.*, **69**, 183–215.
- West, A.H. and Stock, A.M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.*, **26**, 369–376.
- Lee, J., Owens, J.T., Hwang, I., Meares, C. and Kustu, S. (2000) Phosphorylation-induced signal propagation in the response regulator NtrC. *J. Bacteriol.*, **182**, 5188–5195.
- Lewis, R.J., Scott, D.J., Brannigan, J.A., Ladds, J.C., Cervin, M.A., Spiegelman, G.B., Hoggett, J.G., Barak, I. and Wilkinson, A.J. (2002) Dimer formation and transcription activation in the sporulation response regulator Spo0A. *J. Mol. Biol.*, **316**, 235–245.
- Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P. and Dickerson, R.E. (1996) Structure of the *Escherichia coli* response regulator NarL. *Biochemistry*, **35**, 11053–11056.
- Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J.L. (1997) Arac/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.*, **61**, 393–410.
- Martinez-Hackert, E. and Stock, A.M. (1997) Structural relationships in the OmpR family of winged-helix transcription factors. *J. Mol. Biol.*, **269**, 301–312.
- Pestova, E.V., Havarstein, L.S. and Morrison, D.A. (1996) Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol. Microbiol.*, **21**, 853–862.
- Diep, D.B., Havarstein, L.S. and Nes, I.F. (1996) Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. *J. Bacteriol.*, **178**, 4472–4483.
- Mizuno, T. (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.*, **4**, 161–168.
- Galperin, M.Y., Nikolskaya, A.N. and Koonin, E.V. (2001) Novel domains of the prokaryotic two-component signal transduction system. *FEMS Microbiol. Lett.*, **203**, 11–21.
- Lange, R., Wagner, C., de Saizieu, A., Flint, N., Molnos, J., Stieger, M., Caspers, P., Kamber, M., Keck, W. and Amrein, K.E. (1999) Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*. *Gene*, **237**, 223–234.
- de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., Keck, W., Amrein, K.E. and Lange, R. (2000) Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J. Bacteriol.*, **182**, 4696–4703.
- Mohr, C.D., Hibler, N.S. and Deretic, V. (1991) AlgR, a response regulator controlling mucoidy in *Pseudomonas aeruginosa*, binds to the FUS sites of the *algD* promoter located unusually far upstream from the mRNA start site. *J. Bacteriol.*, **173**, 5136–5143.
- Mohr, C.D., Leveau, J.H., Krieg, D.P., Hibler, N.S. and Deretic, V. (1992) AlgR-binding sites within the *algD* promoter make up a set of inverted repeats separated by a large intervening segment of DNA. *J. Bacteriol.*, **174**, 6624–6633.
- Risoen, P.A., Havarstein, L.S., Diep, D.B. and Nes, I.F. (1998) Identification of the DNA-binding sites for two response regulators involved in control of bacteriocin synthesis in *Lactobacillus plantarum* C11. *Mol. Gen. Genet.*, **259**, 224–232.
- Risoen, P.A., Brurberg, M.B., Eijsink, V.G. and Nes, I.F. (2000) Functional analysis of promoters involved in quorum sensing-based regulation of bacteriocin production in *Lactobacillus*. *Mol. Microbiol.*, **37**, 619–628.
- Ween, O., Gaustad, P. and Havarstein, L.S. (1999) Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol. Microbiol.*, **33**, 817–827.
- Cheung, J.K. and Rood, J.I. (2000) The VirR response regulator from *Clostridium perfringens* binds independently to two imperfect direct repeats located upstream of the *pfoA* promoter. *J. Bacteriol.*, **182**, 57–66.
- Gelfand, M.S., Koonin, E.V. and Mironov, A.A. (2000) Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res.*, **28**, 695–705.
- Mironov, A.A., Vinokurova, N.P. and Gelfand, M.S. (2000) Software for analysis of bacterial genomes. *Mol. Biol.*, **34**, 222–231.
- Donvito, B., Etienne, J., Denoroy, L., Greenland, T., Benito, Y. and Vandenesch, F. (1997) Synergistic hemolytic activity of *Staphylococcus lugdunensis* is mediated by three peptides encoded by a non-agr genetic locus. *Infect. Immun.*, **65**, 95–100.
- Santiago, B., Schubel, U., Egelseer, C. and Meyer, O. (1999) Sequence analysis, characterization and CO-specific transcription of the *cox* gene cluster on the megaplasmid pHCG3 of *Oligotropha carboxidovorans*. *Gene*, **236**, 115–124.
- Galperin, M.Y., Gaidenko, T.A., Mulkidjanian, A.Y., Nakano, M. and Price, C.W. (2001) MHYT, a new integral membrane sensor domain. *FEMS Microbiol. Lett.*, **205**, 17–23.
- Dow, J.M., Feng, J.X., Barber, C.E., Tang, J.L. and Daniels, M.J. (2000) Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*. *Microbiology*, **146**, 885–891.
- Brunskill, E.W. and Bayles, K.W. (1996) Identification of LytSR-regulated genes from *Staphylococcus aureus*. *J. Bacteriol.*, **178**, 5810–5812.
- Shimizu, T., Ba-Thein, W., Tamaki, M. and Hayashi, H. (1994) The *virR* gene, a member of a class of two-component response regulators, regulates the production of perfringolysin O, collagenase and hemagglutinin in *Clostridium perfringens*. *J. Bacteriol.*, **176**, 1616–1623.
- Brunskill, E.W. and Bayles, K.W. (1996) Identification and molecular characterization of a putative regulatory locus that affects autolysis in *Staphylococcus aureus*. *J. Bacteriol.*, **178**, 611–618.
- Whitchurch, C.B., Alm, R.A. and Mattick, J.S. (1996) The alginate regulator AlgR and an associated sensor FimS are required for twitching motility in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA*, **93**, 9839–9843.
- Yang, W.Z., Ko, T.P., Corselli, L., Johnson, R.C. and Yuan, H.S. (1998) Conversion of a beta-strand to an alpha-helix induced by a single-site mutation observed in the crystal structure of Fis mutant Pro26Ala. *Protein Sci.*, **7**, 1875–1883.
- Allen, B.L., Gerlach, G.F. and Clegg, S. (1991) Nucleotide sequence and functions of *mrk* determinants necessary for expression of type 3 fimbriae in *Klebsiella pneumoniae*. *J. Bacteriol.*, **173**, 916–920.
- Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.*, **232**, 584–599.
- Karplus, K. and Hu, B. (2001) Evaluation of protein multiple alignments by SAM-T99 using the BALiBASE multiple alignment test set. *Bioinformatics*, **17**, 713–720.
- Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
- Taylor, B.L. and Zhulin, I.B. (1999) PAS domains: internal sensors of oxygen, redox potential and light. *Microbiol. Mol. Biol. Rev.*, **63**, 479–506.