J. Mol. Microbiol. Biotechnol. (2001) 3(2): 319-324.

JMMB Research Article

Comparative Approach to Analysis of Regulation in Complete Genomes: Multidrug Resistance Systems in Gamma-Proteobacteria

Dmitry A. Rodionov, Mikhail S. Gelfand*, Andrey A. Mironov and Alexandra B. Rakhmaninova

State Scientific Center GosNIIGenetika, Moscow, 113545, Russia

Abstract

Comparative approach is a powerful tool for analysis of gene regulation in bacterial genomes. Here we apply it to analysis of regulation of the multidrug resistance transport (MDRT) systems in enterobacteria *Escherichia coli*, *Salmonella typhi*, *Klebsiella pneumoniae* and *Yersinia pestis*. Comparison of nucleotide sequences upstream of MDRT genes was performed in order to predict new regulatory sites (operators) and identify candidate regulons. Since the regulatory sites diverge slower than the non-coding regions in general, they are visible as strongly conserved islands. This analysis resulted in description of a regulatory network for known and hypothetical MDRT systems and porins. New candidate members of the MarA regulon were detected. Putative binding sites for EmrR and AcrR were suggested. A new hypothetical MarX regulon was described that includes some multidrug transporters and porins.

Introduction

Bacteria posess a large number of membrane proteins involved in transmembrane drug transport. Some of these drug transporters are specific for certain compounds, but the so-called multidrug transporters (MDRT) export various unrelated compounds. There exist both simple multidrug efflux pumps consisting of a single transmembrane protein and more complex transport system spanning the cytoplasmic and outer membranes. The latter systems consist of three components. A transmembrane protein is brought into apposition with an outer membrane porin through a periplasmic linker protein.

The experimental data and protein sequence comparisons show that the genome of *Escherichia coli* encodes about 30 MDRT pumps. Some of them belong to the ABC family of ATP-depending transporters, whereas the others are members of the MFS (Major Facilitator Superfamily), RND (Resistance – Nodulation - cell Division) and SMR (Small Multidrug Resistance) families of secondary transporters which mediate the extrusion of drugs from the cell in a coupled exchange with ions (Paulsen *et al.*, 1998; Saier and Paulsen, 1998).

Little is known about regulation of MDRT genes (Figure 1). Expression of the EmrAB multidrug transport system of *E. coli* (Lomovskaya and Lewis, 1992) is regulated by repressor EmrR (Lomovskaya and Lewis, 1995), encoded by the upstream gene and induced by various uncouplers of oxidative phosphorylation, salicylic acid and carbonyl cyanide m-chlorophenylhydrazone (CCCP). Another MDRT system, AcrAB (Ma *et al.*, 1995), is regulated by the global transcriptional activator MarA and the local repressor protein AcrR encoded by the upstream gene (Ma *et al.*, 1996). The MarA regulon consists of genes involved in multidrug and oxidative stress resistance which are regulated by the transcription factor MarA (Alekshun and Levy, 1997; Alekshun and Levy, 1999). In addition to AcrAB, MarA also positively regulates expression of the outer membrane porine TolC which is required for functioning of the AcrAB efflux pump (Aono *et al.*, 1998; Fralick *et al.*, 1997). Thus in this case all components of the multidrug efflux machinery are co-regulated. Furthermore, MarA downregulates the synthesis of the major porin OmpF through the increased production of the antisense RNA micF. The consequent drop of the outer membrane permeability increases the effect of the efflux machinery. Besides, MarA activates *fumC, fpr, nfo* and other genes possibly involved in response to the oxidative stress. The *marA* gene is co-transcribed with another regulatory gene, *marR,* which blocks expression of the *marRAB* operon in the absence of the inducer (sodium salilycilate and other chemicals containing phenolic rings) (Martin and Rosner, 1995). Finally, MarA auto-activates its own expression when it binds upstream of the *marRAB* promoter (Martin *et al.*, 1996).

Binding sites (operators) of only two transcriptional factors (MarA and MarR) involved in regulation of the multidrug resistance genes are known. Recently the crystal structure for MarA in complex with its cognate DNA-binding site was described (Rhee *et al.*, 1998).

The comparative approach is a powerful tool for analysis of gene regulation in completely sequenced bacterial genomes (Mironov *et al.*, 1999; Gelfand, 1999; Gelfand *et al.*, 2000). However, in the cited studies it has been applied to considerably diverged genomes. At that, occurrences of candidate sites upstream of orthologous genes can be considered as statistically independent events, and thus consistency in distribution of these sites is a powerful sign of their functional relevance.

A different variant of the comparative technique can be applied for analysis of closely related genomes. In this case the genic upstream regions are less diverged and often can be aligned without ambiguities. However, the

Received July 05, 2000; accepted July 05, 2000. *For correspondence. Email misha@imb.imb.ac.ru; Fax. 7-(095)-315-05-01.

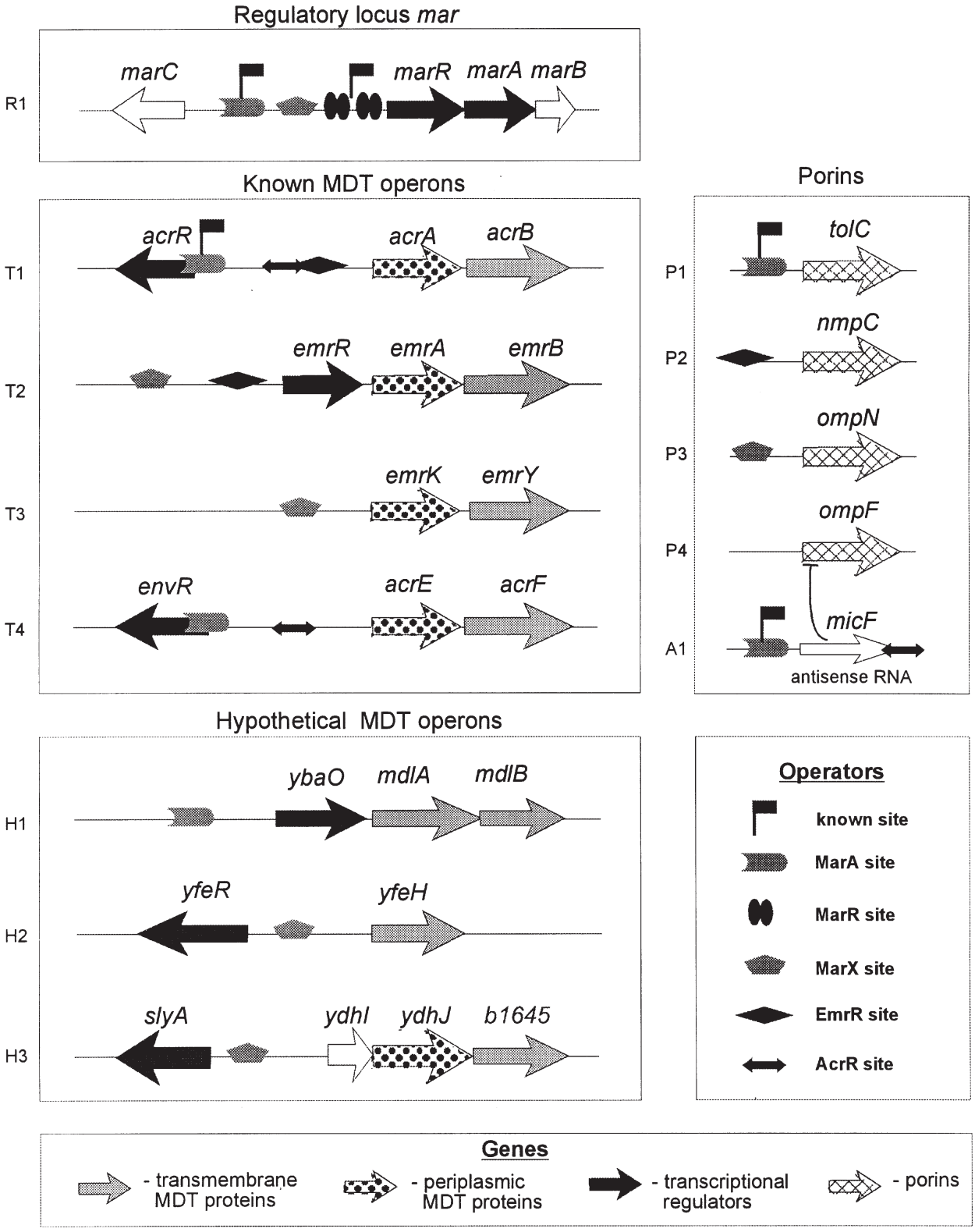

Figure 1. The operon structures and regulatory sites for the multidrug resistance genes in *E. coli.*

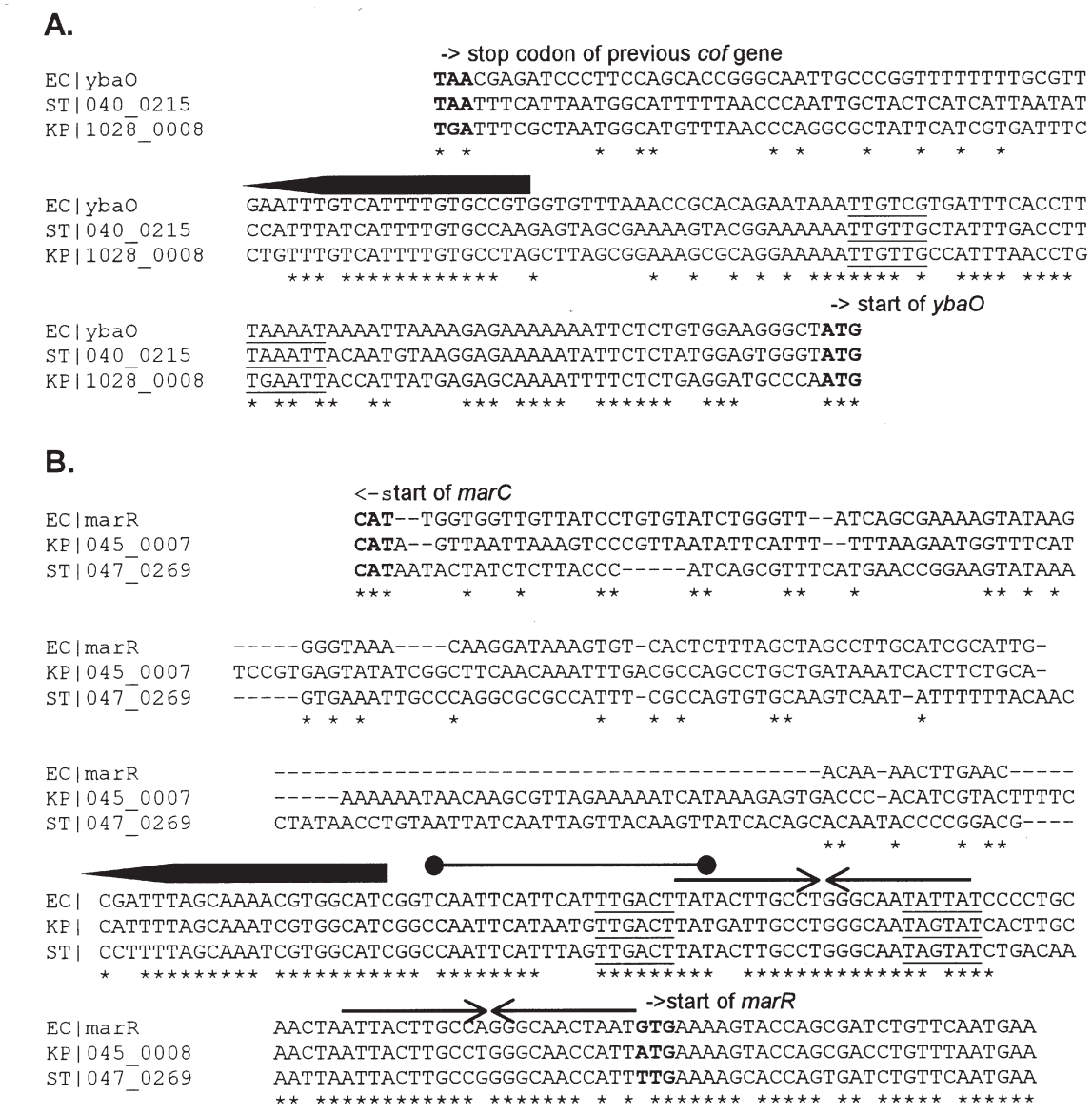

Figure 2. The multiple alignment of upsteam regions of *ybaO-mdlAB* (A) and *marRAB* (B) operons from *E. coli* (EC), *S. typhi* (ST) and *K. pneumoniae* (KP). The promoter –10 and –35 boxes are underlined. The start and stop codons are shown in bold. The assigned regulatory regions:

a) MarR site: palindromic, known
b) MarA site: not palindromic, known
c) MarX site: not palindromic, predicted

regulatory sites diverge slower than the non-coding regions in general, and thus they are visible as strongly conserved islands. This approach, extensively applied for analysis of eukaryotic regulation under the name "phylogenetic footprinting" (Gelfand, 1999; Duret and Bucher, 1997), has been recently shown to be applicable for analysis of prokaryotic sequences as well (Stojanovic *et al.*, 1999). Here we use it to analyze regulation of the multidrug resistance transport systems in enteric bacteria *Escherichia coli*, *Salmonella typhi*, *Klebsiella pneumoniae* and *Yersinia pestis*. We describe candidate binding sites in the less studied genomes, identify new members of known regulons, and find conserved candidate binding sites of unknown regulators.

Results and Discussion

MarA and MarR Regulons

Operon *marRAB* encodes two regulatory proteins: repressor MarR which blocks expression of the *marRAB* operon upon binding to two palindromic operators, and activator MarA which controls expression of several genes involved in multidrug and oxidative stress resistance (Figure 1). Orthologous operons were found only in *S. typhi* and *K. pneumoniae*. The profile was constructed using 7 out of 15 known MarA binding sites (Martin *et al.*, 1999). These sites were conserved in all three genomes. The only exception was the site upstream of *fumC* that is not

Table 1. Known and putative operators of multidrug transporter genes

Gene (operon)	Site position	Site	Site direc-tion	The jack-knife procedure				
				Site *Z*-score			N_{weak}	N_{strong}
A. MarA site:		**AtGGCACraAawrcCAAAyc**		EC	ST	KP		
acrAB/acrR	-160	ATGGCACGAAAAACCAAACA	B	4.84	4.84	#	6	2
fumC	-110	ATGGCACGAAAGACCAAACA	F	4.84	3.68*	4.27	6	2
ybaO-mdlAB	-105	ACGGCACAAAATGACAAATT	B	4.28	3.60	3.82	2	7
tolC	-98	ATGGCACGTAACGCCAACCT	F	4.06	4.18	4.18	0	18
acrEF/envR	-159	ATGGCAAAAAGAACCAAAGC	B	3.82	#	3.37	0	23
marRAB	-157	ATGCCACGTTTTGCTAAATC	B	3.41	3.12	3.43	0	~135
fpr	-93	AAGGCTCAATCGATCAAATC	B	3.40	3.29	2.84	0	~150
micF	-55	ACAGCACTGAATGTCAAAAC	F	2.93	3.03	#	0	~750
nfo	-83	ATCGCATAAACCACTACATC	F	3.15	2.52	#	0	~500
B. MarX site:		**CAATaCATtmAtTTkAytTa**		EC	ST	KP		
marRAB	-133	CAATTCATTCATTTGACTTA	F	5.08	4.77	4.1	2	3
emrRAB	-67	CAATACATTTACTTTATTTG	F	4.77	4.26	3.78	1	4
emrKY	-38	CAATTCATAAATTTCATCTA	F	4.3	#	#	0	6
yfeR/yfeH	-68	CAATAAATCAATTTTATTTC	F	4.14	4.14	4.14	0	14
slyA/ydhIJ-b1645	-80	CAATACATTTGTTTTAGCAA	F	3.98	3.24	4.25	0	22
ompN	-166	CGATACATTCACCAGACTTA	F	3.66	3.72*	4.23*	0	~130
C. EmrR site:		**ATTtrTrAmTgTmkTTACTATA**		EC	ST	YP		
emrRAB	-52	ATTTGTCACTGTCGTTACTATA	F	4.81	4.81	4.18	1	1
nmpC	-163	ATAAATAACAGCCGTTAATATA	F	3.71	3.77*	3.38*	0	~40
acrAB/acrR	-105	ATTTGTGAATGTATGTACCATA	F	3.67	3.07	3.03	0	~80
D. AcrR site:		**TACATACATTtatgAATGTATGTA**		EC	ST	YP		
acrAB/acrR	-112	TACATACATTtgtgAATGTATGTA	P	7.78	7.65	7.9		
micF	+80	AACAgACATTCAgaAATGaATGac	P	5.08	4.6	4.0		
acrEF/envR	-116	TcCtTACATcgAcGAATGataaTt	P	4.28	#	4.25		
E. MarR site:		**AtTacTTGCCngGGCAActAaT**			ST	KP		
marRAB 1	-22	ATTACTTGCCaGGGCAACTAAT	P	7.13	6.89	6.89		
marRAB 2	-57	TAtacTTGCCtgGGCAAtatTA	P	5.02	5.21	5.15		

Known and putative operators of multidrug transporter genes from enterobacteria *E. coli* , EC; *S. typhi*, ST; *K. pneumoniae*, KP; *Y. pestis*, YP. The derived consensus sequences of the operator sites are shown in bold. The site positions are relative to the translation start in *E. coli.* Location of site relative to direction of transcription of the first operon: F – forward (same strand), B – backward (complementary strand), P – palindromic site. N_{weak} : the number of weaker sites from the learning set; N_{strong}: the number of strongers sites in the *E. coli* genome. Both numbers are computed using the jack-knife procedure, see the text for the details.
- no orthologues
* - closely homologous gene with similar function, but not an orthologue

conserved in *S. typhi*, although there was a signal upstream of a *fumC* paralogue (42% identity). The MarA box is an asymmetrical site that is functional in both orientations (Martin *et al.*, 1999).

Scanning of the *E.coli* genome with the contructed profile produced approximately 60 genes with an upstream site scoring 3.5 or higher. However, in only 10 cases the site is conserved in the other two genomes. Two operons with consistent sites encode multidrug transport systems, namely AcrEF and MdlAB (Figure 1). AcrEF is closly homologous to AcrAB (nearly 80% identity) in *E. coli* and obviously has a similar function. The second system is a hypothetical ATP-dependent multidrug efflux pump, encoded by the putative *ybaO-mdlAB* operon. The gene of this operon, *ybaO*, encodes a hypothetical transcription factor from the Lrp family. The found putative MarA binding site corresponds to a strongly conserved region in the multiple alignment of orthologous *ybaO-mdlAB* upstream regions (Figure 2, A).

Using the jack-knife analysis we see that the most significant MarA sites are those upstream of *acrAB, fumC, ybaO-mdlAB, tolC* and *acrEF* operons, whereas the experimentally determined sites upstream of *marRAB, fpr, micF* and *nfo* seem less significant (Table 1, A). In the case of *acrAB* and *acrEF,* the MarA sites are located in the coding region of the upstream regulatory genes *acrR* and *envR*, respectively.

The upstream region of the *marRAB* operon contains also two palindromic binding sites of the MarR repressor. Application of the standard procedure using the MarR box profile produced no additional candidate MarR-regulated genes in *E. coli* (Table1, E). Therefore it is likely that MarR is simply a local repressor of the *marRAB* operon.

Hypothetical MarX Regulon

The upstream region of the *marRAB* operon contains another conserved DNA segment located between the MarA and MarR operators (Figure 2, B). The standard procedure detects similar sites in regulatory regions of some genes encoding multidrug transporters or porins (Figure 1). Since there is no indication as to which transcriptional factor binds to these sites, we call this signal MarX box. The hypothetical MarX regulon of *E. coli* consists of known and hypothetical multidrug tranporters EmrAB, EmrKY, YfeH, YdhIJ-B1645 and a hypothetical outer membrane porin, OmpN. The *emrKY* operon is paralogous

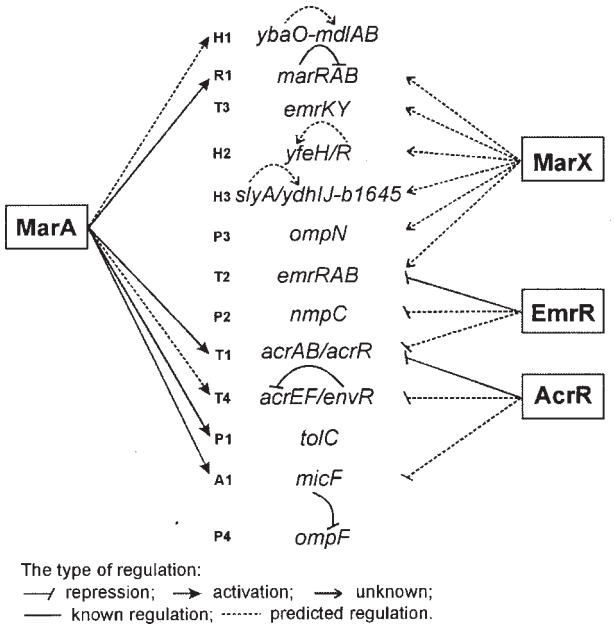

Figure 3. Multidrug resistance regulatory network.

to the known multidrug tranporter operon *emrAB*, and it has no orthologues in related bacterial genomes. The gene *yfeH* encodes a hypothetical transport protein homologous to arsenical resistance proteins and Na/bile acids cotransporters from eukaryotes. The *ydhIJ-b1645* operon is homologous to the fusaric acid resistance system from *Burkholderia cepacia*. It is noteworthy that upstream of all three operons there are genes encoding transcription factors that may serve as local regulators.

The jack-knife analysis confirmes all sites except the one upstream of the *ompN* gene (Table1, B).

Hypothetical EmrR Regulon

The first gene of the *emrRAB* operon encodes transcription factor EmrR that represses transcription of this operon. We have identified a putative regulatory site with imperfect symmetry (EmrR-box). It overlaps with the known promoter of the *emrRAB* operon and is conserved in the genomes of *S. typhi, K. pneumoniae* and *Y. pestis*. Using the standard procedure, we have identified possible members of this new regulon, in particular a multidrug transport system AcrAB and an outer membrane porin NmpC from *E. coli* (Table1, C). The other genomes in this study have no orthologues to NmpC, but there are strong EmrR boxes upstream of other porin genes.

Thus in each case we have observed co-regulation of multidrug systems (EmrAB and AcrAB) and outer membrane porins. This resembles the known co-regulation of *acrAB* and *tolC* by the transcriptional factor MarA.

Hypothetical AcrR Regulon

The upstream region of the *acrAB* operon contains a candidate AcrR operator (Table 1, D). It is a perfect palindrome of 24 bp, well-conserved in ortologous upstream regions and overlapping with the *acrAB* promoter. Similar conserved sites were found upstream of the paralogous operon *acrEF* and within the *micF* gene for the antisense RNA that negatively controls expression of the major porin OmpF (Figure 1).

Thus we have described a complex regulatory network including eight multidrug transport operons, four porin genes and the regulatory locus *marRAB* (Figure 3). The multidrug systems in enterobacteria seem to be regulated by several diverse regulatory systems. Many operons are regulated by several regulators: a global one (MarA, MarX, EmrR, AcrR) and a local one (MarR, EnvR, YbaO, YfeR, SlyA).

Experimental Procedures

Known and putative MDRT of *E.coli* were selected from the database of transport proteins (Paulsen *et al.*, 1998; http://www.biology.ucsd.edu / ~ipaulsen/transport/). The proteins having functional annotation "drug efflux" and, in addition, all hypothetical transmembrane proteins from *E.coli* were screened against the non-redundant database using PSI-BLAST (Altschul *et al.*, 1997) in order to detect additional homologues of the known drug transporters. The resulting set consisted of 50 known or hypothetical MDRT genes.

Known MarA binding sites were collected from the literature (Martin *et al.*, 1999). The search profile was constructed using the alignment of these sites. Positional nucleotide weights in this profile were defined as:

$$W(b,k) = log\,[N(b,k) + 0.5\,] - 0.25\,\Sigma_{i=A,C,G,T} \log\,[\,N(i,k) + 0.5\,],$$

where $N(b,k)$ is the count of nucleotide b at position k. The score of a candidate site was calculated as the sum of respective positional nucleotide weights:

$$Z(b_1 \ldots b_k) = \Sigma_{\kappa=1\ldots\kappa}\, W(b_k,k)$$

where k is a the length of the site. Z-score can be used to assess the significance of an individual site.

In other cases no experimentally defined sites were available and to search for putative regulatory sites the following standard procedure was used:

- Detection of the first gene in the putative operon containing the gene under analysis.
- Alignment of the upstream region of this operon from *E. coli* with upstream regions of the orthologous operons of *S. typhi*, *K. pneumoniae* and *Y. pestis*.
- If necessary, correction of the annotated gene starts.
- Construction of profiles corresponding to highly conserved regions in the upstream regions.
- Scanning of the *E. coli* genome with the constructed profile and identification of candidate sites.
- Check of the site conservation upstream of orthologous operons. If all orthologues in the analyzed genomes have a candidate site with Z-score exceeding 3.5, then we consider this site as a putative regulatory box and place the corresponding operons to a hypothetical regulon.

Protein alignment was done using the Smith-Waterman algorithm implemented in the *GenomeExplorer* program (Mironov *et al.*, 2000). Orthologous proteins were defined as bidirectional best hits (Tatusov *et al.*, 2000). Distant homologues were identified using PSI-BLAST (Altschul *et al.*, 1997). Multiple sequence alignments were constructed using CLUSTALX (Thompson *et al.*, 1997). Site recognition was done using *GenomeExplorer* (Mironov *et al.*, 2000). Preliminary sequence data were obtained from The Institute for Genomic Research WWW site (http://www.tigr.org).

To assess significance of the putative regulatory sites we applied the jack-knife procedure. To do that, we constructed profiles using all candidate sites from a regulon excluding one particular site. Then the Z-score of this site was computed using this "partial" profile. For each candidate site we also computed the number of weaker sites from the learning set and the number of genes in the entire *E. coli* genome with a stronger site.

Acknowledgements

This study was supported by grants from the Merck Genome Research Institute (244), the Russian Fund of Basic Research (99-04-48247 and 00-15-99362), the Russian State Scientific Program "Human Genome" HHMI (55000309) and INTAS (99-1476).

Note Added In Proof

Recently it has been shown that the *ybaO-mdlAB* operon is controlled by Rob (Bennik *et al.*, 2000). Rob and MarA have very similar recognition signals and in many cases bind same sites. Thus this result may be considered as a confirmation of our predictions. At the same time, it should be noted that it is impossible to distinguish MarA and Rob binding sites by sequence analysis done. The set of genes affected by constitutive expression of MarA was identified in (Barbosa and Levy, 2000). This set does not contain the candidate MarA-regulon members from our study. However, since the set from (Barbosa *et al.*, 2000) does not contain several genes known to be regulated by MarA and even co-transcribed with other genes from the set, the cited results do not directly contradict our predictions.

References

Alekshun, M., and Levy, S. 1997. Regulation of chromosomally mediated multiple antibiotic resistance: the *mar* regulon. Antimicrob. Agents and Chemotherapy 41: 2067-75.

Alekshun, M., and Levy, S. 1999. The *mar* regulon: multiple resistance to antibiotics and other toxic chemicals. Trends in Microbiol. 7: 410-413.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-402.

Aono, R., Tsukagoshi, N., and Yamamoto, M. 1998. Involvement of outer membrane protein TolC, a possible member of the *mar-sox* regulon, in maintenance and improvement of organic solvent tolerance of Escherichia coli K-12. J.Bacteriol. 180: 938-44.

Barbosa, T.M, and Levy, S.B. 2000. Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. J. Bacteriol. 182: 3467-74.

Bennik, M.H., Pomposiello, P.J., Thorne, D.F., and Demple, B. 2000. Defining a *rob* regulon in *Escherichia coli* by using transposon mutagenesis. J. Bacteriol. 182: 3794-801.

Duret, L., and Bucher, P. 1997. Searching for regulatory elements in human noncoding sequences. Curr. Opin. Struct. Biol. 7: 399-406.

Fralick, J. 1996. Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of *Escherichia coli*. J.Bacteriol. 178: 5803-5805.

Gelfand, M.S. 1999. Recognition of regulatory sites by genomic comparison. Res. Microbiol. 150: 755-772.

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.

Lomovskaya, O., and Lewis, K. 1992. *Emr*, an *Escherichia coli* locus for multidrug resistance. Proc. Natl Acad. Sci. USA 89: 8938-8942.

Lomovskaya, O., and Lewis, K. 1995. EmrR is a negative regulator of the *Escherichia coli* multidrug resistance pump EmrAB. J. Bacteriol. 177: 2328-2334.

Ma, D., Alberti, M., Lynch, C., Nikaido, H., and Hearst, J.E. 1996. The local repressor AcrR plays a modulating role in the regulation of *acrAB* genes of *Escherichia coli* by global stress signals. Mol. Microbiol. 19: 101-112.

Ma, D., Cook, D., Alberti, M., Pon, N.G., Nikaido, H., and Hearst, J.E. 1995. Genes *acrA* and *acrB* encode a stress-induced efflux system of *Escherichia coli*. Mol. Microbiol. 16: 45-55.

Martin, R., and Rosner, J. 1995. Binding of purified multiple antibiotic-resistance repressor protein (MarR) to *mar* operator sequences. Proc. Natl Acad. Sci. USA 92: 5456-60.

Martin, R.G., Gillette, W.K., Rhee, S., and Rosner, J.L. 1999. Structural requirements for marbox function in transcriptional activation of *mar/sox/rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. Mol. Microbiol. 34: 431-41.

Martin, R.G., Jair, K.W., Wolf, R.E.Jr., and Rosner, J.L. 1996. Autoactivation of the *marRAB* multiple antibiotic resistance operon by the MarA transcriptional activator in *Escherichia coli.* J. Bacteriol. 178: 2216-23.

Mironov, A.A., Koonin, E.V., Roytberg, M.A., and Gelfand, M.S. 1999. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. Nucleic Acids Res. 27: 2981-2989.

Mironov, A.A., Vinokurova, N.P., and Gelfand, M.S. 2000. Software for analysis of bacterial genomes. Mol. Biol. 34: 253-262

Paulsen, I., Sliwinski, M., and Saier, M. 1998. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J. Mol. Biol. 277: 573-592.

Paulsen, I.T., Brown, M.H., and Skurray, R.A. 1996. Proton-dependent multidrug efflux systems. Microbiol Rev. 60: 575-608.

Rhee, S. , Martin, R.G., Rosner, J.L., and Davies, D.R. 1998. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. Proc. Natl Acad. Sci. USA 95: 10413-18.

Saier, M.H., and Paulsen, I.T. 1998. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. FASEB J. 12: 265-274.

Stojanovic, N., Florea, L., Riemer, C., Gumuchio, D., Slightom, J., Goodman, M., Miller, W., and Harrison, R. 1999. Comparison of five methods for finding conserved sequences in multiple alignments of gene regulatory regions. Nucleic Acids Res. 27: 3899-3910.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28: 33-6.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-82.