Evolution of Regulatory Systems in Bacteria
(Invited Keynote Talk)

Mikhail S. Gelfand[1,2], Alexei E. Kazakov[1], Yuri D. Korostelev[2], Olga N. Laikova[3], Andrei A. Mironov[1,2], Alexandra B. Rakhmaninova[1,2], Dmitry A. Ravcheev[1], Dmitry A. Rodionov[1,4], and Alexei G. Vitreschak[1]

[1] A.A.Kharkevich Institute for Information Transmission Problems, RAS, Bolshoi Karetny pereulok 19, Moscow, 127994, Russia
{gelfand,kazakov,ravcheyev,rodionov,vitreschak}@iitp.ru
[2] Faculty of Bioengineering and Bioinformatics, M.V.Lomonosov Moscow State University, Vorobievy Gory 1-73, Moscow, 119992, Russia
{mironov@bioinf.fbb,abr@belozersky}.msu.ru
[3] Research Institute for Genetics and Selection of Industrial Microorganisms, Pervy Dorozhny proezd 1, Moscow, 127994, Russia
[4] A.A. Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

Abstract. Recent comparative studies indicate surprising flexibility of regulatory systems in bacteria. These systems can be analyzed on several levels, and I plan to consider two of them. At the level of regulon evolution, one can attempt to characterize the evolution of regulon content formed by loss, gain and duplications of regulators and regulated genes, as well as gain and loss of individual regulatory sites and horizontal gene transfer. At the level of transcription factor families, one can study co-evolution of DNA-binding proteins and the motifs they recognize. While this area is not yet ripe for fully automated analysis, the results of systematic comparative studies gradually start to coalesce into an understanding of how bacteria regulatory systems evolve.

Keywords: Comparative genomics, bacteria, regulation of transcription, regulation of translation, transcription factor, binding site, T-box.

1 Introduction

Sequencing of hundreds of bacterial genomes has created a situation when in many taxa we have rather dense and relatively uniform sampling of genomes at varying evolutionary distance from each other. This paves way for careful comparative genomic analysis of regulatory systems and their evolution. Identification of candidate transcription factor binding sites and regulatory RNA structures and analysis of their distribution in related genomes allows one to reconstruct the evolutionary history of regulons, whereas analysis of candidate binding sites for transcription factors forming a structural family creates an opportunity for studying co-evolution of transcription factors and their binding motifs, and hence, elucidation of family-specific protein-DNA interaction code.

I. Măndoiu, G. Narasimhan, and Y. Zhang (Eds.): ISBRA 2009, LNBI 5542, pp. 1–4, 2009.

2 Evolution of Regulons

The list of basic events shaping the regulons includes gain (via duplication and horizontal gene transfer) and loss of regulators, changes of specificity, gain, loss and duplication of regulated genes, shuffling of genes in operons, and gain and loss of individual regulatory sites. While it is not currently possible to estimate the rate of these events, it is clear that these rates are not uniform for different regulons and their life stages.

Further, it is clear that in most cases there are significant overlaps between individual regulons, and it makes more sense to speak of interacting regulatory systems.

Life without FUR: Evolution of Iron Homeostasis in the Alpha-Proteobacteria [1]. One of examples where the evolutionary history could be reconstructed in sufficient detail is the regulation of iron homeostasis in the alpha-proteobacteria. In this case the starting even was change of ligand specificity that transformed the usual iron repressor FUR into manganese-responsive MUR in the common ancestor of the Rhizobiales and Rhodobacteriales. The role of iron regulator was assumed by a distant member of the FUR family, Irr, and this state is conserved in the Bradirhizobiceae. In the Rhizobiaceae, IscR, a regulator of genes involved in the synthesis of the iron-sulfur clusters also was lost. Further on, in the Rhizobiaceae the RirA regulator appeared, and the job of iron-dependent regulation is shared by Irr (mainly responsible for iron storage, Fe-S clusters, heme and iron-dependent enzymes) and RirA (main regulator of iron acquisition and also some Fe-S and iron storage genes). In the Rhodobacteriales, iron acquisition is regulated by an unknown transcription factor binding to the motif CTGActrawtyagTCAG, that is somewhat similar to the binding moitif of IscR; iron storage genes are co-regulated by this factor and Irr, Fe-S synthesis, by IscR and Irr, and iron-dependent enzymes, solely by Irr.

Fatty Acid and Branched-Chain Amino Acid Utilization in the Gamma- and Beta-proteobacteria [2]. A similar reconstruction could be performed for a large system regulating the catabolism of fatty acids (FA) and branched-chain amino acids (ILV) in the gamma- and beta-proteobacteria. This system involves six transcriptional factors from the MerR, TetR, and GntR families binding to eleven distinct DNA motifs. The ILV degradation genes in the gamma- and beta-proteobacteria are regulated mainly by a newly identified regulator from the MerR family (e.g., LiuR in *Pseudomonas aeruginosa*) and, in some beta-proteobacteria, by a TetR-family regulator LiuQ. In addition to the core set of ILV utilization genes, the LiuR regulon in some lineages is expanded to include genes from other metabolic pathways, such as the glyoxylate shunt and glutamate synthase, as well as salt- and alkaline stress response in the *Shewanella* species. The FA degradation genes are controlled by four regulators including FadR in the gamma-proteobacteria, PsrA in the gamma- and beta-proteobacteria, FadP in the beta-proteobacteria, whereas in the alpha-proteobacteria it is regulated by LiuR orthologs. The most parsimonious evolutionary scenario for the ILV and FA regulons seems to be that LiuR and PsrA were likely present in the common ancestor of the gamma- and beta-proteobacteria, and they have been partially or fully substituted by LiuQ and FadP in the *Burkholderiales* and by FadR in some groups of the gamma-proteobacteria.

T-boxes and Regulation of Amino Acid Metabolism in the Firmicutes [3]. T-boxes are regulatory RNA structures that bind to uncharged tRNAs and regulate aminoacyl-tRNA synthetase genes as well as genes encoding amino acid transporters and metabolic enzymes. T-boxes are sufficiently large to retain the phylogenetic signal, at least at short evolutionary distances, and hence it is possible to follow the history of T-box duplications. Further, since the specificity of T-boxes is dictated by the interaction between a well-defined structural element (so-called specifier codon) and the tRNA anticodon, they are an ideal material for studying changes in specificity. One of the most interesting observations is rapid, duplication-driven, lineage-specific expansion of some specific T-box regulon following the loss of previously existing transcription factors.

Regulon Expansion, or how FruR Has Become CRA and Duplicated RbsR Has Become PurR. The fructose repressor FurR, a member of the LacI family, is a standard sugar regulator in most lineages of the gamma-proteobacteria, whereas in *E.coli* it is a well-studied global regulator named CRA (catabolism repressor and activator). Following the fate of known binding sites in the genomes ordered by increasing phylogenetic distance from *E. coli*, one can see that the regulon expansion started with the glycolysis pathway and then extended to some genes of the Krebs cycle and sugar catabolic pathways.

Similarly, the ribose operon regulator RbsR duplicated in the common ancestor of the Enterobacteriales and Vibrionales. The RbsR copy retained the ligand (ribose) specificity and the regulon, but its DNA motif changed somewhat (to A**GCG**AAACGTTT**CGC**T), whereas the other copy retained the DNA motif (A**CGC**AAACGTTT**GCG**T), but has become the purine repressor PurR regulating, in *E. coli*, more than twenty genes from the purine biosynthesis pathway and some adjacent pathways.

3 Co-evolution of Transcription Factors and DNA Motifs They Recognize

As mentioned in the previous section, evolution of regulons is often accompanied by changes in the DNA motifs. To study co-evolution of transcription factors (TFs) and their binding sites systematically, we are doing large-scale comparative genomics analysis of several families of TFs. An outcome of such studies is lists of TFs, each with a set of candidate binding sites.

Several recently developed programs are used to identify correlated positions in proteins and DNA. Indeed, it turns out that when this analysis was applied to the LacI family of TFs, the identified set of correlated positions was consistent with several known X-ray structures of TF-DNA complexes. Notably, however, the set of protein positions correlated with specific nucleotides was not limited to residues in immediate contact with the DNA: in several families this set also included positions situated on the other side of the DNA-binding alpha-helix and forming hydrophobic interactions with the rest of the protein. Further, these studies revealed that the family-specific protein-DNA recognition code is not limited to known universal correlations

(like “arginine binds to guanine”), nor to pairwise correlations. Some of predictions coming from these analyses were recently confirmed in experiment [4].

Acknowledgments. The reported studies were supported by grants from the Howard Hughes Medical Institute (55005610 to M.S.G.), the Russian Fund of Basic Research (08-04-01000 to A.E.K.), and the Russian Academy of Sciences (program «Molecular and Cellular Biology»).

References

1. Rodionov, D.A., Gelfand, M.S., Todd, J.D., Curson, A.R.J., Johnston, A.W.B.: Comparative Reconstruction of Transcriptional Network Controlling Iron and Manganese Homeostasis in Alpha-Proteobacteria. PLoS Comp. Biol. 2, e163 (2006)
2. Kazakov, A.E., Rodionov, D.A., Alm, E., Arkin, A., Dubchak, I., Gelfand, M.S.: Comparative Genomics of Regulation of Fatty Acid and Branched-Chain Amino Acid Utilization in Proteobacteria. J. Bacteriol. 191, 52–64 (2009)
3. Vitreschak, A.G., Mironov, A.A., Lyubetsky, V.A., Gelfand, M.S.: Functional and Evolutionary Analysis of the T-box Regulon in Bacteria. RNA 14, 717–735 (2008)
4. Desai, T., Rodionov, D., Gelfand, M., Alm, E., Rao, C.: Engineering Transcription Factors with Novel DNA-binding Specificity Using Comparative Genomics. Nucleic Acids Res. (in press)