Correlations of substitutions predict specific protein-DNA contacts in the MerR family of transcriptional factors

Ilya Zharov,

Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994, peshwalk@gmail.com

Yuriy Korostelev

Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994, Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia, 119991, lan787@yko.name

Correlations of substitutions between the sequences of transcriptional factors (TFs) and their binding sites may be used to predict specific contacts between the protein and DNA residues. Here we apply this method to the group of heavy metal resistance TFs of the bacterial MerR family using the Prot-DNA-Korr program.

MerR family TFs regulate response to different stresses (antimicrobial agents, heavy metals, heat shock, oxidative stress, nitrosative stress) as well as a number of metabolic processes (nitrogen metabolism, carotenoid biosynthesis, degradation of isoprenoids and branched-chain amino acids, curli and biofilm formation) in bacteria. A separate group of this family is responsible for transcriptional activation of heavy metal resistance (HMR) systems that detoxify and export mercury, copper, zinc, cadmium, lead, silver, and gold ions. Experimentally studied TFs from this group include MerR, CueR, HmrR, ZntR, CadR, PbrR, and GolS [1-4].

Most of the MerR family transcriptional activators including the studied TFs bind their palindromic sites between the −35 and −10 promoter boxes. Spacers of the regulated promoters are characterized by the increased length (19-20 bp comparing with average 16-17 bp). Upon binding of the TF the distance between the promoter boxes decreases and they realign for optimal DNA polymerase binding [5]. Unsurprisingly TFs that adopt this mode of regulation share very similar spatial conformation of DNA binding and dimerization domains [5-8, 10]. Their ligand-binding domains differ in sequence, size and structure. 3D-structures of DNA-bound TFs are known for BmrR, MtaN, SoxR, TipAL, GlnR and TnrA [5-11] as

well as DNA-free structures for BmrR, MtaN, CueR, ZntR, NmlR, LMOf2365_2715 and SCO5550 [11-14]. Among them only CueR and ZntR belong to the studied group. GlnR, TnrA and SCO5550 have a different DNA binding and dimer formation modes. Nevertheless these structures show that the conformations of the DNA-binding winged helix-turn-helix (WHTH) domains of the MerR family TFs and therefore their contacts with the half-sites are highly similar.

For the current work we selected proteins from GenBank RefSeq database that contain HTH_CueR, HTH_MerR1, HTH_CadR-PbrR, HTH_CadR-PbrR-like and HTH_HMRTR conserved domains of GenBank CDD (1516 TFs in total). In this study they are referred to CueR, MerR, CadR-PbrR, CadR-PbrR-like and HMRTR subfamilies, respectively Structure-based sequence alignments and phylogenetic trees were built for each subfamily. After filtering out the TFs that impair the alignments as well as nearly identical TFs encoded in the genomes of close strains, 906 TFs remained for further study. Selective positional weighted matrices (PWMs) were built for searching the genomes for the binding sites. The genomes were searched in regions from −400 to +50 bp using the GenomeExplorer package. Only sites located in the long (19-20 bp) spacers of the putative σ70 promoters were retained for the correlation study. 884 sites for 763 TFs were identified using this procedure.

Subfamily	Number of TFs in completely assembled genomes	Number of TFs after filtering	Number of TFs with identified TFBSs	Number of TFBSs
CueR	511	260	238	324
MerR	205	123	105	106
CadR-PbrR	253	193	172	174
CadR-PbrR-like	189	147	100	110
HMRTR	358	183	148	170
Total	1516	906	763	884

Table 1. TF and TFBS statistics

The aligned sequences of the DNA-binding domains and their binding sites were sent to Prot-DNA-Korr program. 32 correlates pairs of positions of these two alignments were identified. Then we searched the literature and the NPIDB database for the experimentally identified contacts of the MerR family TFs with DNA. A pair of positions was marked as interacting if the contact was reported at least once. In total, 36 pairs of positions with side chain to base

interactions were found. 9 pairs appear both contacting and correlated. Given 74 protein and 20 DNA alignment positions (1480 possible pairs), Fisher's exact test gives p-value of 1.96×10^{-8} for this overlap. We mapped several correlated pairs of positions with large (over 50) counts for overrepresented pairs of residues on the general phylogenetic tree of the studied TFs. The same overrepresented pairs 'nucleotide – amino acid' in fixed alignment positions appeared several times independently during the evolution. We reconstructed ancestral sequences of the studied TFs and their binding sites using PAML package to test whether mutations in TFs lead to changes in binding sites. Unfortunately we could not find evidence neither for this order of changes nor for the opposite. Our results show that correlations of substitutions in the sequences of TFs and their binding sites can be used as predictors of specific contacts

This is joint work with Mikhail Gelfand.

1. N.L. Brown, J.V. Stoyanov, S.P. Kidd, and J.L. Hobman (2003) The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.*, **27:** 145-163.

2. J.L. Hobman, J. Wilkie and N.L. Brown (2005) A design for life: prokaryotic metal-binding MerR family regulators. *Biometals*, **18:** 429-436.

3. P.R. Chen and C. He (2008) Selective recognition of metal ions by metalloregulatory proteins. *Curr. Opin. Chem. Biol.*, **12:** 214-221.

4. A.O. Summers (2009) Damage control: regulating defenses against toxic metals and metalloids. *Curr. Opin. Microbiol.*, **12:** 138-144.

5. E.E. Heldwein and R.G. Brennan (2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature*, **409:** 378-382.

6. K.J. Newberry and R.G. Brennan (2004) The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *J. Biol. Chem.*, **279:** 20356-20362.

7. Newberry,K.J. et al. (2008) Structures of BmrR-drug complexes reveal a rigid multidrug binding pocket and transcription activation through tyrosine expulsion. *J. Biol. Chem.*, **283:** 26795-26804.

8. S. Bachas, C. Eginton, D. Gunio and H. Wade (2011) Structural contributions to multidrug recognition in the multidrug resistance (MDR) gene regulator, BmrR. *Proc. Natl. Acad. Sci. U S A*, **108:** 11046-11051.

9. M.A. Schumacher et al. (2015) Structures of regulatory machinery reveal novel molecular mechanisms controlling *B. subtilis* nitrogen homeostasis. *Genes Dev.*, **29:** 451-464.

10. S. Watanabe, A. Kita, K. Kobayashi and K. Miki (2008) Crystal structure of the [2Fe-2S] oxidative-stress sensor SoxR bound to DNA. *Proc. Natl. Acad. Sci. U S A*, **105:** 4121-4126.

11. M. Kumaraswami, K.J. Newberry and R.G. Brennan (2010) Conformational plasticity of the coiled-coil domain of BmrR is required for *bmr* operator binding: the structure of unliganded BmrR. *J. Mol. Biol.*, **398:** 264-275.

12. M.H. Godsey, N.N. Baranova, A.A. Neyfakh, and R.G. Brennan (2001) Crystal structure of MtaN, a global multidrug transporter gene activator. *J. Biol. Chem.*, **276:** 47178-47184.

13 A. Changela et al. (2003) Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. *Science*, **301:** 1383-1387.

14. T. Hayashi et al. (2013) Structural and genomic DNA analysis of a putative transcription factor SCO5550 from *Streptomyces coelicolor* A3(2): regulating the expression of gene *sco5551* as a transcriptional activator with a novel dimer shape. *Biochem. Biophys. Res. Commun.*, **435:** 28-33.