**Full Report of MerR family proteins (CueR & GolS)**

CueR sequence:  
MNISDVAKITGLTSKAIRFYEEKGLVTPPMRSENGYRTYTQQHLNELTLLRQARQVGFNLEESGELVNLFNDPQRHSADVKRRTLEKVAEIERHIEELQSMRDQLLALANACPGDDSADCPIIENLSGCCHHRAG

GolS sequence:   
MNIGKAAKASKVSAKMIRYYEQIGLIPAASRTDSGYRAYTQADVNQLHFIRRARDLGFSVAEISDLLNLWNNQSRQSADVKRLAQTHIDELDRRIQNMQHMAQTLKALIHCCAGDALPDCPILHTLGQPDDSEPEARTGAVLRRPRRHGLAKRL

**EMBOSS NEEDLE between GolS and CueR:**

EMBOSS\_001 1 MNIGKAAKASKVSAKMIRYYEQIGLIPAASRTDSGYRAYTQADVNQLHFI 50

|||...||.:.:::|.||:||:.||:....|:::|||.|||..:|:|..:

EMBOSS\_001 1 MNISDVAKITGLTSKAIRFYEEKGLVTPPMRSENGYRTYTQQHLNELTLL 50

EMBOSS\_001 51 RRARDLGFSVAEISDLLNLWNNQSRQSADVKRLAQTHIDELDRRIQNMQH 100

|:||.:||::.|..:|:||:|:..|.||||||.....:.|::|.|:.:|.

EMBOSS\_001 51 RQARQVGFNLEESGELVNLFNDPQRHSADVKRRTLEKVAEIERHIEELQS 100

EMBOSS\_001 101 MAQTLKALIHCCAGDALPDCPILHTLGQPDDSEPEARTGAVLRRPRRHGL 150

|...|.||.:.|.||...||||:..|.........

EMBOSS\_001 101 MRDQLLALANACPGDDSADCPIIENLSGCCHHRAG--------------- 135

EMBOSS\_001 151 AKRL 154

EMBOSS\_001 136 ---- 135

Note: | is identical, : is different but similar in terms of function, . is low level of similarity or not similar at all, A is DNA binding catalytic residue, A is metal chelation catalytic residue, A is special function suspect residue.

Reason for splitting right at Ser77 because that is extra chelation between metal ion to Ser77 from another monomer.

**DNA binding domain (N-terminus to Ser77):** DNA binding domain is where the N-terminus interacts with the protein at certain **catalytic residues** **CueR.**

***List of DNA catalytic residues for monomer CueR:***

|  |  |  |
| --- | --- | --- |
| Catalytic residues | Interaction portions | DNA portions |
| Ser4 | Backbone O | 4C - DNA backbone O phosphate 1 – chain 1 |
| Backbone N |
| Thr13 | Backbone O | 15T - DNA backbone O phosphate 2 – chain 2 |
| Lys15 | Side chain NΖ (zeta) | 16G – DNA base N7 – chain 2 |
| 17G – DNA base O6 – chain 2 |
| Arg18 | Side chain NE (epsilon) | 5T – DNA backbone O phosphate 2 – chain 1 |
| Side chain NH2 (end) |
| Tyr20 | Side chain OH | 14C – DNA sugar O5’ – chain 2 |
| 14C – DNA backbone O phosphate 2 – chain 2 |
| Asn34 | Side chain ND2 (delta) | 23C – DNA backbone O phosphate 1 – chain 2 |
| Tyr36 | Backbone N | 22T – DNA sugar O4’ – chain 2 |
| Arg37 | Side chain OH | 5T – DNA backbone O phosphate 1 – chain 1 |
| Side chain NH1 | 5T – DNA backbone O phosphate 1 – chain 1 |
| 6T – DNA backbone O phosphate 2 – chain 1 |
| Side chain NH2 |
| Arg54 | Side chain NE | 14C – DNA backbone O phosphate 2 – chain 2 |
| Side chain NH2 | 14C – DNA backbone O phosphate 2 – chain 2 |

So far in all of the catalytic residues are either very similar in terms of function or identical with the exception of the Ser4 is completely different from Gly4. Glycine is a very flexible residue, even though it is hydrophobic but it can be exposed out to the environment or tucked in, my suspect is that it may be exposed but not a catalytic residue.

***Promoter sequence that CueR and GolS bind to:***  
There are 2 binding sites on the promoter of CueR and GolS,  
For GolS: the promoter we’re using is pGolB



For CueR:

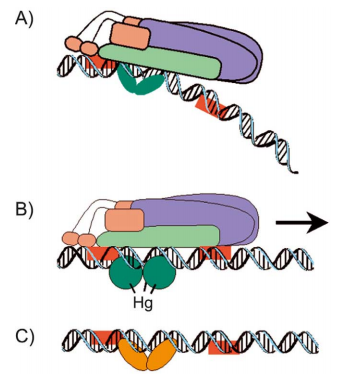
Using pdb file code:

Chain 1 (Chain Y): GACCTTCCAGCAAGGGGAAGGTC  
Chain 2 (Chain X): CTGGAAGGTCGTTCCCCTTCCAG

Note: highlight indicates the DNA binding site where CueR catalytic residues bind to.

**Metal ion binding domain (helix-loop-helix or helix-turn-helix):**

MerR family protein general contains a helix-loop-helix structure where the loop is the metal ion binding site with 2 Cys residues (C112 and C120) conserved among CueR and GolS. These residues are conserved, Cys residues are considered hotspot for metal chelation (good electron donating group – thiol).  
Note: a good check when trying to monitor metal ion binding domain should be a big loop where Cys residues stick out along with the sequence 113-119 in between. In the model, should accound for interaction between Cys residues as they most likely to form disulphide bridges (or not if metal ions were placed there).   
MerR family mostly exist in homodimer form; we suspect that interaction between 1 monomer to another monomer can be enhanced using metal ion chelation. Besides the C112 and C120, there is S77, also highly conserved among CueR and GolS. We expect this residue to play a helping role in securing the chelation and furthermore, stabilize the conformation of dimer.

***How CueR and GolS (MerR family proteins) function transcriptional activation in general:***Without the presence of metal ions (i.e Cu1+ and Au3+), the RNA polymerase cannot run through the DNA for gene expression, There is a certain bent conformation to the DNA (in general – does not mean it should be applicable for all MerR family) that inhibits the transcription of the gene.   
With the presence of metal ions, the protein dimer bound to 2 metal ions and thus create a change in DNA conformation allowing for transcriptional activation of the gene

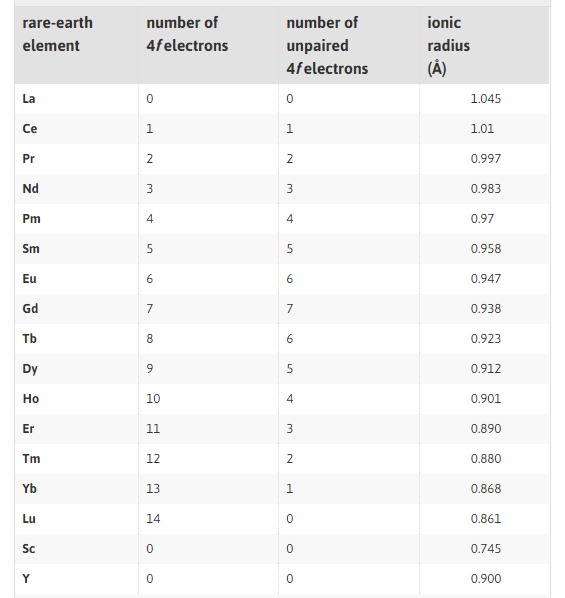
***So far, there has not been any cases where MerR family chelates with non-ionic form of metal (i.e metal (0)), why?***  
Reason: binding site is full of electron donating group (great nucleophile) and most metal(0) are stable and inert to any “reactions”. It requires an electron deficient form of metal to form chelation with Cys-residues – cations provide the best complementation.

***What role does Cys at the metal site do?***  
Reason: Cys has the most nucleophilic side chain among all amino acids (-SH group provides the best electron donation to any potential cations), knocking Cys or switching Cys to a weaker nucleophile like Ser or Thr would not perform a good chelation job to metal ios in general.

***With consideration of metal pocket size calculation, why does size matter?***Reason: As the catalytic residues that play a major role in the metal chelation site do not differ between CueR and GolS, we suspect the difference that makes CueR and GolS specific to different metal ions are pocket size and shape.  
A potential continuous outcome of the project after accomplishment of the proteỉn 3D model is to calculate pocket size of the protein loop to confirm our suspect that size really does matter. That could be a very important discovery if we are planning to modify any MerR proteins to detect other metal ions.

***A special suspect feature of GolS that could have made a big difference in it from CueR?***We discussed that potential Cys residues create a nice hotspot for metal ion docking site. If we paid close attention to the GolS sequence, we observed that upstream of C112, there is another **C111** exist which raise up the question as these 2 Cys are so close. Our suspect is that this C111 residue may play a huge role in metal ion binding since there are only 2 Cys residues for CueR.  
Potential function of C111 is to increase the electron density around our metal chelation site meaning making this pocket filled with lots of lots of electron donating group and enhance a tighter ionic interaction with metal ions. This means as soon as a metal ion docks onto the pocket, our protein is guaranteed to perform a “tight” interaction with this metal ion.  
Note: for the dry lab, take into consideration of the presence of this extra C111, there is a high chance that C111 can influence the interaction of docking.  
GolS sequence contains 154 residue while CueR contains 135 residues. The region downstream of C120 is considered a highly variant region for all MerR proteins. The most interesting thing about this region is the length or the properties of residues determining the shape, size, characteristic of the metal pocket. Given that we have a long variant region after C120, we believe that these end residues have to fold within a restricted area as it is overcrowded and really close to DNA binding domain. Its folding may result different changes in the shape of the metal pocket.   
Note: be aware of the potential interaction between of the C-term region to the DNA binding domain, like random electrostatic, hydrophobic and H-bond interaction may occur.

**Side project:**   
Metal chelation site modification:  
There is a huge investment interest in rare earth metal detection or mining, these are of high government interests. We are also looking into rare earth metal bio-sensing using metalloproteins like MerR families to detect. There are a couple of issues with regards to rare earth metal bio-sensing that we have to address:  
1. Metalloproteins have a low level of detection for metal(0) in general, unless there is a mixture of ionic and solid metal(0) in solution. Rare Earth metal are extremely rare and mostly in ore deposits which means they are covalently form with multiple elements, even rarer to find them in single ionic form.  
2. Rare Earth metal ions (if existed in a survey sample) contains a very large ionic radius compared to Au3+ (0.75 A radius) (mostly >> 0.8 A) => resulting in a huge difference in pocket size (something that requires drastic change in the pocket size structure) but also results in the decrease in specificity.  
3. There has not been a well-studied transcriptional activator metalloprotein that utilize its mechanism on rare earth element: 3D model pdb database is not well establish, keeping in mind most 3D prediction model is not metal ions docking friendly, let alone rare earth metal (which means that we have to create de novo a rare earth metal ion interaction to the proteins).



Toehold Switches (Riboswitches):  
We also consider 3D modelling toehold switches (riboswitches) which are regulated using RNA secondary and tertiary metal-RNA complex structure. We also encountered a couple of challenges with regards to monitor riboswitches in prediction program:  
1. RNA toehold switches contain 4 main bases (A, U, C, G) and modified bases (D, Ψ, I and random methylated bases) which interactions may not constraint within the Watson-Crick basepairing but also take into consideration of Hoogsteen-Loop basepair, non-WC, tetraduplex, etc… We suspect that with the level of complexity of toehold switches to metal ions: there is a contribution of these specially modified bases to the interactions and conformations of switches with metal ions.   
2. PyRosetta is considered not RNA prediction friendly; the program is more accurate on proteins interaction mostly. It requires a completely different RNA prediction program which exist, however, the learning time for a brand new program takes time. Plus, these programs tend to focus more of secondary structure prediction and unsure in terms of tertiary structure.  
3. There are also a lot of considerations we have to look over if we are to model how our RNA regulate translation in general which require 1 pdb model of target RNA and its sequence as well as the ribosomal apparatus interaction with the RNA (given that the toehold switches live directly on the mRNA). If the toehold switches are separate from the mRNA, the interaction could become even more complicated.