AviJMB
Avicenna Journal of Medical Biochemistry
journal homepage: http://AJMB.umsha.ac.ir

Original Article

The arsR gene: A comparative phylogenetic and bioinformatics analysis to study reasons for different sensitivity to arsenic compounds

Maryam Daneshpour[1], Nooshin Shabab[2] and Massoud Saidijam[2*]

[1]*Department of Molecular Medicine and Human Genetics, School of Medicine,Hamadan University of Medical Sciences Hamadan, Iran*
[2]*Research Center for Molecular Medicine, Hamadan University of Medical Sciences, Hamadan, Iran*

*** CORRESPONDENCE**
Massoud Saidijam
Tel: +98 9121324616
Fax: +988118276299
E-mail: sjam110@yahoo.com

ARTICLE HISTORY
Received: 14 July 2013
Revised: 18 August 2013
Accepted: 10 September 2013
Available online: 28 September 2013

Citation:
Daneshpour M, Shabab N, Saidijam M. The arsR gene: A comparative phylogenetic and bioinformatics analysis to study reasons for different sensitivity to arsenic compounds. Avi J Med Biochem. 2013; 1(1): 30-35.

ABSTRACT

Objectives: The overall goal of this study was to carry out a set of comparative analyses of *arsR* gene in plasmid R773 and bacterial chromosome from *Escherichia coli BL-21(DE3).*

Methods: PDB and NCBI databases and Chimera, Mega4, CLC main workbench software and 3D-jigsaw and EMBL-EBI servers were applied to perform this study. By using these software and servers, multiple analyses including determination of residue composition, secondary structure and motifs, 3D structure, conserved regions, etc. were done.

Results: The results suggest that such high sensitivity to arsenic compounds in *ars*-containing plasmid R773 may be related to ArsR protein characteristics such as amino acids composition, secondary and tertiary structure, hydrophobicity, level of interaction with DNA.

Conclusion: Bioinformatics studies could be applied to describe the reason of different sensitivities to Arsenic compounds between *arsR* gene and ArsR protein in plasmid R773 and bacterial chromosome.

Keywords: *ArsR protein, E coli; Arsenic; Plasmid R773; Protein secondary structure; Protein three dimensional structure*

Introduction

Arsenic is one of the toxic metalloids which is widely dispersed in the environment and occurs in two oxidation states, arsenate (As V) and arsenite (As III) that both are toxic and could inhibit many biochemical processes in living organisms [1, 2]. Microorganisms inhabiting arsenic polluted environments have developed kinds of detoxification systems [3]. The *ars* operons which encode arsenic resistance have been found in multicopy plasmids in both gram negative and gram positive bacteria [2]. The plasmid version contains five important genes containing *arsR*, *arsD*, *arsA*,*arsB* and *arsC* respectively which are shown in Fig. 1 with their proteins functions [4, 5]. Several researches based on molecular techniques showed obvious homology between this operon and special sequences on chromosomal DNA from a number of bacterial species [4]. The chromosomal ars operon from *Escherichia coli* has three genes (*arsR, arsB and arsC*)[5]. These results suggest that the chromosomal *ars* operon may be the evolutionary precursor of the plasmid-borne operon [2]. Moreover, the experimental studies indicated that *ars* operon from plasmid R773 is more sensitive to arsenic compounds compared to chromosomal *ars* operon [6]. One of the probable factors for this phenomenon may be *arsR* gene (the first cis-

tron of *ars* operon) and its product, ArsR protein, which is a trans-acting repressor that regulates expression of *ars* operon. ArsR a member of the ArsR/SmtB family, acts as a homodimeric winged helix-turn-helix transcriptional repressor that its dimerization domain locates at 79-92 residues [7]. This protein is specifically binds to its operator/promoter (O/P) DNA binding site in the absence of arsenic [7]. Binding of arsenic ions to the ArsR protein leads to derepression of target family. In this study we will compare *arsR*gene from plasmid R773 and *Escherichia coli BL-21(DE3)* chromosome to explain reasons for different level of sensitivity to arsenic compounds. PDB and NCBI databases and Chimera, Mega4, CLC main workbench software and 3D-jigsaw and EMBL-EBI servers were applied to perform this study. By using these software and servers, multiple analyses including determination of residue composition, secondary structure and motifs, 3D structure, conserved regions were done.

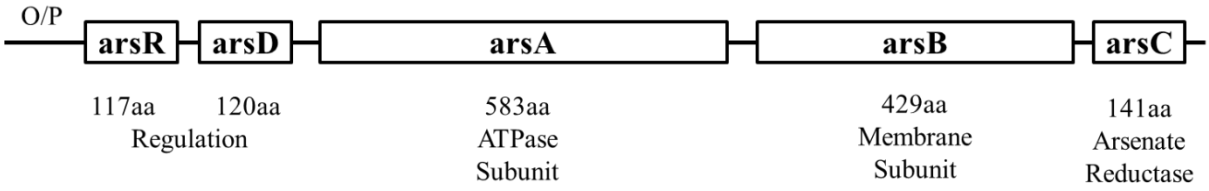

Figure 1. *ars* operon genes and their functions from plasmid R773 (Silver *and Ji*, 1994(ref. 5)).

Material and methods

The *arsR* sequences were achieved from NCBI databases (http://www.ncbi.nlm.nih.gov). For multiple alignments we used CLC main workbench software. In order to phylogenic analyses, construction of evolutionary trees and computation, MEGA4 software was applied [8-10]. Comparison of conserved domains and 3D-structures were done via 3D-jigsaw (http://bmm.cancerresearchuk.org/~3djigsaw), EMBL-EBI (http://www.ebi.ac.uk) servers and Chimera software [11].

Results

In order to get a general conception of phylogenetic relations of *arsR* gene and ArsR protein, MEGA4 software was applied. The evolutionary trees were constructed based on nucleotide sequences of *arsR* gene and amino acid sequences of ArsR protein which showed the expected harmonies (Fig. 2). Amino acid sequences of marked taxa in Fig. 2B were aligned using CLC main work bench (Fig. 3). Regions with more red color are more conserved.

In the next step, the similarity of *arsR* gene and ArsR protein from plasmid R773 and *E.coli BL21* was determined using EMBL-EBI. The results showed 69% similarity between these two genes and also 72.5% identity and 85.5% similarity between their related proteins.

3D-jigsaw server was applied to predict the three dimensional structure of the ArsR proteins. The predicted structures were edited by Chimera software. Based on these analyses, the ArsR proteins have been made from 6 helixes and 2-β sheets (Fig. 4A, Fig. 4B).

These two proteins were compared based on the volumes and some important angles and distances using Chimera software. As it can be seen in Fig. 4C and Fig. 4D, the measured volume of ArsR-R773 (13335) is less than ArsR-BL21 (13929). The angle between Asp41, Gln42 and Ser43 (the amino acids of turn region in HTH motif) and the angle between Val33, Gln42 and Glu56 (the angle between two helixes of HTH region) were measured. The results (Fig. 5) showed that unlike the former angle, which is 110.11° in ArsR-BL21 and 100.49° in ArsR-R773, the later angle was smaller in ArsR-BL21(40.85° in ArsR-BL21 and 42.25° in ArsR-R773). The measured distances which can be seen in Fig. 5E and Fig. 5F, are all larger in ArsR-R773. All the measurements are summarized in Table 1.

In the next part, we compared the arsenic binding sites in two proteins which are shown in Fig. 6A and Fig. 6B. It appears that these sites are more available in ArsR-R773. In Fig. 6C and Fig. 6D the DNA binding sites were compared. It seems that this region, unlike arsenic binding sites, is less available in ArsR-R773 than ArsR-BL21.

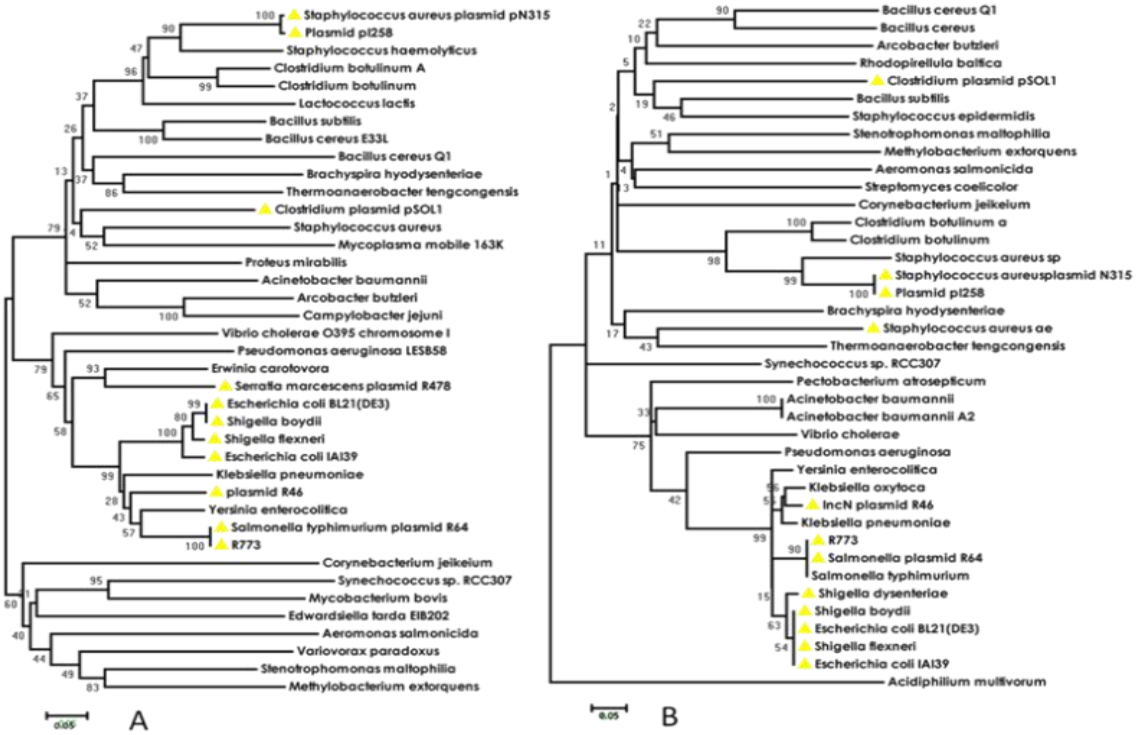

Figure 2. Phylogenetic tree based on nucleotide sequences of arsR genes (A) and based on amino acid sequences of ArsR proteins(B). The numbers shown next to the nodes indicate percent bootstrap values of the 500 replicates. Genetic distances were computed by using p-distance model. The tree was constructed by MEGA4 software.

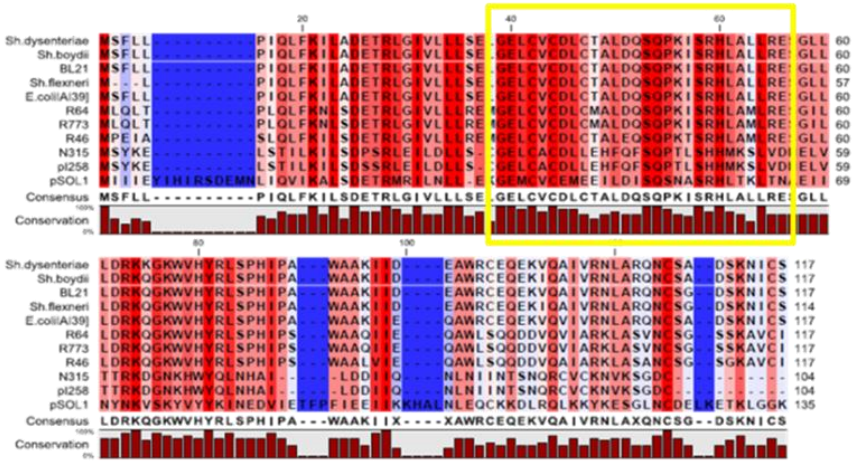

Figure 3. Alignment of amino acid sequences of ArsR protein for marked taxa in Figure 3B. Alignment was done via CLC main workbench software. Regions having more red color are more conserved.

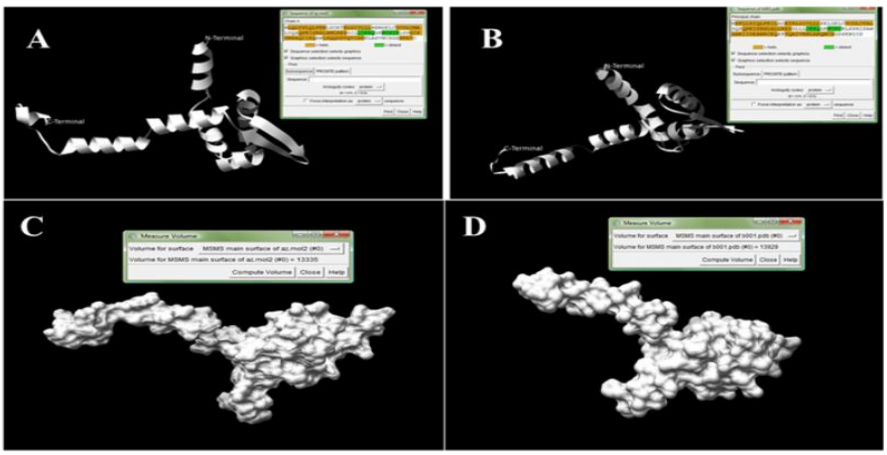

Figure 4. Ribbon forms of three dimensional structures of ArsR-R773 (A) and ArsR-BL21 (B). The volume measurements of ArsR proteins from plasmid R773(C) and *E.coliBL21* (D). Structure prediction and post edition were done via 3D-jigsaw and Chimera.

Table 1. The volumes, angles and distances measurements of the ArsR proteins from plasmid R773 and *E.coli BL21* that were done by using Chimera software

	ArsR-R773		ArsR-BL21	
	measures	figures	measures	Figures
Volume	13335	4C	13929	4D
Angle: Asp 41[1]-Gln 42[1]-Ser 43[1]	100.49 A˚	5A	110.11 A˚	5B
Angle: Val 33[2]-Gln 42[1]-Glu 56[2]	42.25 A˚	5C	40.85 A˚	5D
Distance: Gln 42[1]-Gly 66[3]	23.180 A˚	5E	22.541 A˚	5F
Distance: Gln 42[1]-Lys 67[3]	20.289 A˚	5E	19.794 A˚	5F
Distance: Gln 42[1]-Arg 63[4]	21.241 A˚	5E	20.823 A˚	5F
Distance: Gln 42[1]-His 70[4]	16.417 A˚	5E	15.864 A˚	5F

[1]Asp 41-Gln 42-Ser 43(Amino acids of turn region in HTH motif)
[2]Val 33, Glu 56 (The last amino acids at the ends of helixes in HTH motif)
[3]Gly 66, Lys 67 (Amino acids between two β-sheets)
[4]Arg 63, His 70 (Mddle amino acids of every β-sheets)

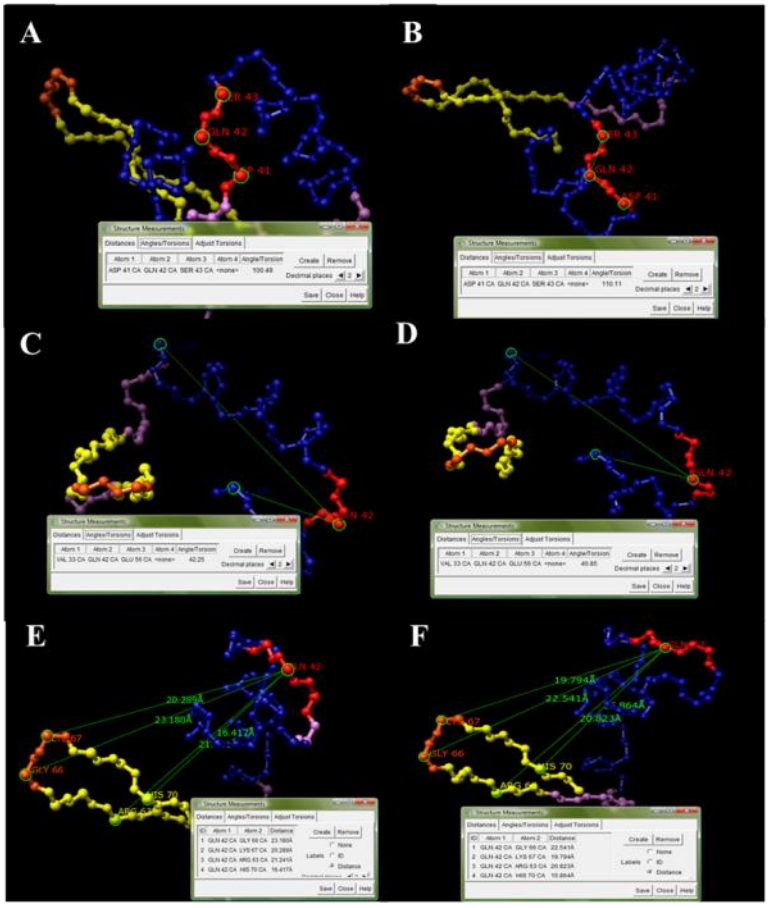

Figure 5. The angles measurements between: Asp 41-Gln 42-Ser 43 for ArsR-R773 (A) and ArsR-BL21(B); between: Val 33-Gln 42-Glu 56 for ArsR-R773(C) and ArsR-BL21(D). The distances measurements for ArsR-R773 (E) and ArsR-BL21(F). Structure prediction and post edition were done via 3D-jigsaw and Chimera.

Discussion

The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. The trees are drawn to scale,

with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4.

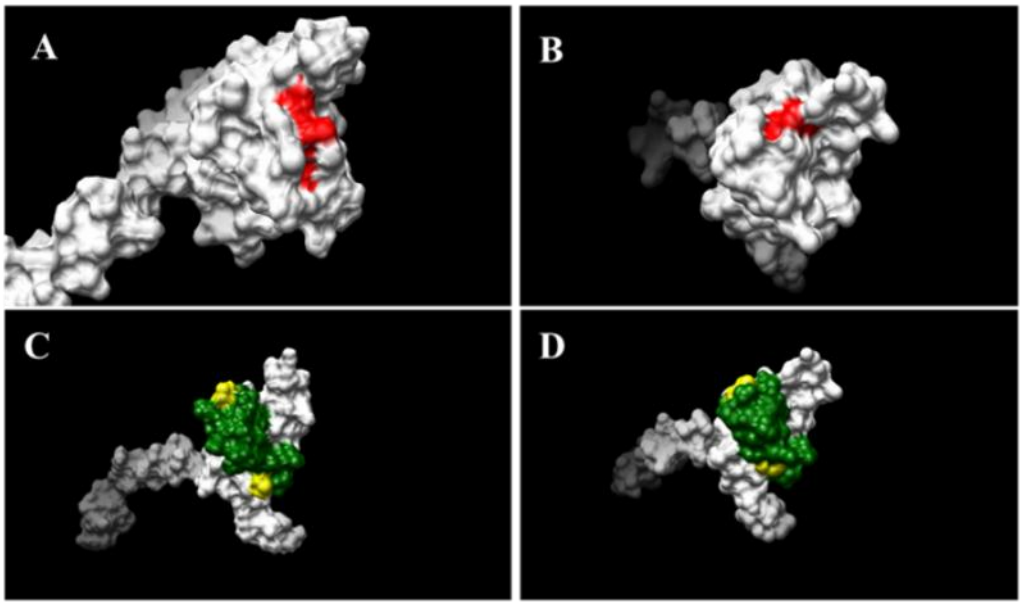

Figure 6. The arsenic binding residues (Cys 32, Cys 34 and Cys 37) are showed in red color in ArsR-R773 (A) and ArsR-BL21 (B). The green regions show the HTH motifs in ArsR-R773 (C) and ArsR-BL21 (D). Amino acid residues 38, 53 are highlighted in yellow color. Structure prediction and post edition were done via 3D-jigsaw and Chimera.

In these figures, some taxa which were more relative to *arsR* gene and ArsR protein from plasmid R773 or which were have been favorable in *ars*operon studies are marked and amino acid sequences of the marked taxa were aligned. As it can be seen in Fig. 3, ArsR sequences have the most similarity in residues from 29 to 57 which are shown in a yellow box. This region plays two important roles in ArsR activity: the three sulfur thiolates of the cysteine residues (cys 32, cys 34 and cys 37) form a very specific binding site for trivalent metalloid (As III)[12]. Binding of arsenite ions to ArsR induce a conformational change leading to dissociation from DNA and hence derepression. On the other hand, this region contains a conserved helix-turn-helix motif that is responsible for DNA binding [12,13].

It was shown a high level of similarity (85.5%) between these two *arsR*, but in comparative view, it can easily be observed that in helix-turn-helix motif of *arsR* from plasmid R773, there are two methionine residues which are absent in chromosomal *arsR*gene from *E.coli BL21*. ArsR protein from *E.coli BL21* has threonine and leusine instead of methionine in sites 38 and 53. Since ArsR is one of the proteins with winged helix topology, it could not be weird that difference in amino acids composition related to HTH motif results in different level of sensitivity expression to arsenic compounds between *arsR* genes from plasmid R773 and *E.coli BL21* through trace on conformation of protein and its reaction with DNA. Obviously, more practical analysis such as directed substitution mutations are necessary to prove our suggestion about the methionine residues role.

In order to compare the ArsR proteins from plasmid R773 and *E.coli BL21*, some measurements have been done using Chimera software and are shown in related figures. In view of the measured volumes, it could be highlighted that the ArsR-R773 is more condense and one of the angle between helixes which are involved in recognition and DNA binding is smaller compared to ArsR-BL21. Therefore we suggest that its DNA binding may be looser because of unavailability of the necessary binding sites on DNA and ArsR protein. However the other measured angles and distances do not support this idea completely. Another contributing factor which could influence the different levels of sensitivity may be the availability of three mentioned cysteine residues involved in arsenite ions bindings. As it can be seen in Fig. 6, it seems that compared to same from BL21, arsenite ions can interact easier with the ArsR protein from plasmid R773 due to space limitation, which could be the reason of faster response to arsenic or response to the lower concentrations of arsenic that are defined as higher sensitivity. In addition, regarding to the alignments, high level of dissimilarity can be observed at the C-terminal of proteins. Hence

it is possible that the residues of this region play a critical role in conformational status of these proteins which is very important in protein-DNA or protein-metal reactions.

Conclusion

The ArsR super family is a member of helix-turn-helix bacterial transcription regulatory proteins that appear to dissociate from DNA in the presence of specific inducible compounds [14]. Overall goal of this study was to carry out a set of comparative analyses of *arsR* gene and ArsR protein in plasmid R773 and bacterial chromosome from *Escherichia coli BL-21(DE3).* Our results suggest that such high sensitivity to arsenic compounds in *ars*-containing plasmid R773 may be due to the related ArsR protein characteristics such as amino acids composition, secondary and tertiary structure, hydrophobicity, level of interaction with DNA etc.

Acknowledgement

This research has been done as MSc thesis and supported by Deputy of Research and Technology, Hamadan University of Medical Sciences.

Conflict of interest

The authors report no conflict of interest.

References

1. Achour-Rokbani A, Cordi A, Poupin P, Bauda P, Billard P. Characterization of the ars gene cluster from extremely arsenic-resistant Microbacterium sp. strain A33. Appl Environ Microbiol 2010; 76:948-55.
2. Diorio C, Cai J, Marmor J, Shinder R, DuBow MS. An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. J Bacteriol 1995; 177:2050-6.
3. Baker-Austin C, Dopson M, Wexler M, et al. Extreme arsenic resistance by the acidophilic archaeon 'Ferroplasmaacidarmanus' Fer1. Extremophiles 2007; 11:425-34.
4. Carlin A, Shi W, Dey S, Rosen BP. The ars operon of Escherichia coli confers arsenical and antimonial resistance. J Bacteriol 1995;177:981-6.
5. Silver S, Ji G. Newer systems for bacterial resistances to toxic heavy metals. Environ Health Perspect 1994;102 3:107-13.
6. Cai J, DuBow MS. Expression of the Escherichia coli chromosomal ars operon. Can J Microbiol 1996 ;42:662-71.
7. Pennella MA, Giedroc DP. Structural determinants of metal selectivity in prokaryotic metal-responsive transcriptional regulators. Biometals 2005;18:413-28.
8. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4:406-25.
9. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 1985; 39: 783-91.
10. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24:1596-9.
11. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J ComputChem 2004; 25:1605-12.
12. Ordonez E, Thiyagarajan S, Cook JD, et al. Evolution of metal (loid) binding sites in transcriptional regulators. J Biol Chem 2008; 283:25706-14.
13. Rosinski JA, Atchley WR. Molecular Evolution of Helix–Turn–Helix Proteins. J Mol Evol 1999; 49:301-9.
14. Gajiwala KS, Burley SK. Winged helix proteins. Current Opinion in Structural Biology 2000; 10:110–6.