Arginine and Lysine interactions with π residues in metalloproteins

Parimelzaghan Anitha, Vaideeswaran Sivasakthi, Pandian Lavanya, Susmita Bag, Kalavathi Murugan Kumar, Anand Anbarasu & Sudha Ramaiah*

Bioinformatics Division, School of Biosciences & Technology, VIT University, Vellore-632014, India; Sudha Ramaiah – Email: sudhaanand@vit.ac.in; Phone: +91-416-2202556/2574/2608; Fax: +91-416-2243092; *Corresponding author

Abstract:
Metalloproteins have many different functions in cells such as enzymes; signal transduction, transport and storage proteins. About one third of all proteins require metals to carry out their functions. In the present study we have analyzed the roles played by Arg and Lys (cationic side chains) interactions with π (Phe, Tyr or Trp) residues and their role in the structural stability of metalloproteins. These interactions might play an important role in the global conformational stability in metalloproteins. In spite of its lower natural occurrence (1.76%) the number of Trp residues involved in energetically significant interactions is higher in metalloproteins.

Key words: Metalloproteins, π interactions, Arginine, Lysine, sequential distance

Background:

Metalloproteins are the most extensively studied class of bio-macromolecules in the field of biological inorganic chemistry [1] and in environment protection [2]. In the field of protein design and engineering, metalloproteins make particularly attractive targets [3]. There are many reasons for this, including the exciting possibility of engineering protein-based molecules with useful catalytic, electronic or optoelectronic properties. Another particularly important reason for their popularity as targets is that metal binding makes available plenty of energy, which helps to compensate for the deficiencies in our ability to design sequences that fold into well-packed, stable structures. Additionally, the abundant spectroscopic methods available for probing metal sites allow relatively easy assessment of success or failure in generating new metal-binding sites. Thus, metal-binding sites have been built into existing protein scaffolds or into scaffolds that have been designed *de novo*, usually with computational screening of structures for appropriate ligand positions [4].

Weak attractive forces are important in determining the three dimensional structure of proteins. Electrostatic interactions, van der Waals force and hydrogen bonds play important roles in folding a protein and establishing its final structure [5]. Arg/Lys interactions with π residues are increasingly recognized as an important non-covalent binding interaction relevant to structural biology [6-8]. Arg/Lys interactions with π residues are found to be common among structures in the Protein Data Bank (PDB) [9], and it is clearly demonstrated that, when a cationic side chain is near a π side chain the geometry will be biased towards one that would experience an energetically significant Arg/Lys-π interaction [6]. Over one-fourth of all Trp in proteins experience an energetically favorable interaction with cationic side chains [6]. A number of studies have established a role for these interactions in biological recognition [10, 11], enhancement of the stability of thermophilic proteins [12, 13], folding of polypeptides [14, 15] and the stability of membrane proteins [16, 17]. Thus considering the above findings we thought it would be useful to investigate the role of Arg/Lys-π interactions in metalloproteins and we are not aware of any bioinformatics approaches to study these interactions in metalloproteins. Ours results from the present study will be useful for researchers in structural biology and bio-chemistry.

BIOINFORMATION

Methodology:

Dataset

All available crystal structures of metalloproteins from PDB **[9]** were taken for the present study. The selection criteria for a metalloprotein to be included in the dataset were based on the following criteria. (i) The protein molecule should be in complex with a metal only [Metalloproteins in complex with nucleic acids, sugars and fatty acids are excluded]; (ii) The sequence identity among the proteins in the dataset was less than 40%; (ii) Three dimensional structures of these proteins have been solved with ≤2.5 **A°**.

Based on the above selection criteria we obtained a non-redundant set of 35 x-ray resolved structures out of 1096 structures and the PDB ID's of the metalloproteins included in the dataset are as follows:
1AWD_A, 1B0L_A, 1D4A_A, 1DLM_A, 1DXQ_A, 1HFX_A, ILNH_A, 1OI8_A, 1P9R_A, 1RRH_A, 1UNF_X, 1VEW_A, 1WZD_A, 1XM5_A, 1Y67_A, 1YGE_A, 1Z6R_A, 2AFM_A, 2CF4_A, 2DD5_C, 2E46_A, 2IF6_A, 2IQ6_A, 2IVN_A, 2NXF_A, 2P18_A, 2RDV_A, 2VW7_A, 2Z68_A, 2ZOW_A, 3D19_A, 3EXM_A, 3FWI_A, 3HIP_A and 3PCG_M.

Arg/Lys-π interactions

The number of Arg and Lys interactions with π residues in each metalloprotein in the dataset was computed by the program CaPTURE (cation pi trends using realistic electrostatics) **[8]**. All metalloprotein complexes that had energetically significant interactions (interaction energy ≤2 kcal/mol) were selected for the bioinformatics analysis.

Secondary structure preferences

Secondary structure types were assigned by dictionary of protein secondary structure (DSSP) **[18]**, and are denoted using letters: H for helix, T for turn, and S for strand **[19, 20]**. Secondary structure preferences for the interacting Arg/Lys-π residues were obtained from DSSP.

Solvent accessibility patterns

The physical representation of water molecules in direct contact with the protein or with a particular part of the protein was termed solvent accessibility. In the geometrical representation it was the surface described by all possible positions of a water molecule in touching contact with protein atoms. The mathematical calculation of solvent accessibility was done by integrating a step function f over all points x on the surface of a sphere of a radius r (atom) + r (water) around atom i. $f = 1$ if a water sphere centered at x (by definition in contact with atom i) does not intersect with any other protein atom; otherwise $f=0$ **[18]**. We carried a systematic analysis of the solvent accessibility patterns for Arg and Lys residues involved in interactions with π residues in metalloproteins using DSSP **[18]**.

Stabilization centers

Amino acid residues that might be responsible for the prevention of the decay in the folded protein structure were termed stabilization centers **[21]**. For predicting the stabilization centers, we used profiles extracted from multiple alignments as input to the network **[21]**. The alignments were taken from the HSSP data bank **[22]**. For each residue the frequency of occurrence was computed for the 20 amino acids at each position in the alignments; thus, the input group contained 20

real values reflecting the statistics on amino acid occurrences at the given sequential position **[21]**. There was one additional input unit for the conservation weight for each residue that reflected the conservation of the given position in the alignment. These weights were also included in the HSSP files. The teaching and the training procedure was similar to the one applied in case of single sequences **[21]**. To estimate the significance of the calculated amino acid composition of the set of residues involved in long range interactions and in the stabilization centers, standard deviations were calculated in the following way: datasets were randomized 1000 times and distributions were calculated from all cases. The standard deviation was derived from the resulting Gaussian-like distribution **[21]**.

Sequence conservation

For computing conservation score the following methodology was adopted: (i) The amino acid sequence was extracted from the PDB **[9]** file; (ii) Homologous sequences in the SWISS-PROT database **[23]** were searched and collected using PSI-BLAST **[24]**; (iii) A multiple sequence alignment (MSA) of these sequences was constructed using CLUSTAL W **[25]**; (iv) A phylogenetic tree was re-constructed based on the MSA, using the neighbor-joining algorithm **[26]** as implemented in the Rate4Site program **[27]**; (v) Position-specific conservation scores were computed using the empirical Bayesian **[28]** or maximum-likelihood **[27]** algorithms. The continuous conservation scores were divided into a discrete scale of 9 grades. Grade 1 contained the most variable positions; grade 5 contained intermediately conserved positions; and grade 9 contained the most conserved positions **[29, 30]**.

Sequence separation between residues involved in Arg/Lys-π interactions

For a given residue, the comparison of the surrounding residue was analyzed in terms of the location at the sequence level. The contribution from <± 4 were treated as short-range contacts, >± 4 to <± 20 as medium-range contacts and >± 20 were treated as long range contacts **[31-33]**. This classification enabled us to evaluate the contribution of short, medium and long-range contacts in the formation of cation-π interactions. This classification also provides clues so as to understand the importance of these non-covalent interactions in the structural stability of secondary structural elements and their importance in the local and global conformational stability.

Metal binding sites

The metal binding sites within the metalloprotein complex was visualized using the LIGPLOT program **[34]**. Using the program we generated schematic 2-D representations of protein-metal complexes from standard PDB file input.

Results:

Arg and Lys interactions with π residues

Arg and Lys residues involved in interactions with π residues are computed for all the interacting pairs and there are a total of 10451 amino acid residues and 145 energetically significant cation-π interactions in the data set studied. Hence there is an average of one Arg/Lys interaction with π residues for every 72 residues and an average of three interactions per metalloprotein in the data set. The occurrences of Arg/Lys and π residues in the metalloproteins studied are presented in **Table 1 (see**

supplementary material). Arg/Lys residues account for 10% among the naturally occurring amino acids. In the acceptor π-group Trp residues are higher than Phe and Tyr in the metalloproteins studied. The number of interactions involving Phe and Tyr residues is almost similar. The Arg-π interacting pairs in lactoferrin [PDB ID 1B0L_A] are shown in **(Figure 1)**. In terms of pair wise interactions the preferences are similar to the above discussed patterns and the highest pairing is with the Arg-Trp pairs. The percentages of Arg/Lys π interacting pairs in metalloproteins are 16, 24, 27, 11, 09 and 13 respectively for Arg-Tyr, Arg-Phe, Arg-Trp, Lys-Tyr, Lys-Phe and Lys-Trp. It is interesting to note that, even though the percentage natural occurrence of Trp in metalloproteins is very minimal when compared with the other two π residues as represented in **Table 1 (see supplementary material)**; there is a significant number of Arg-Trp and Lys-Trp interacting pairs. The higher number of energetically significant cation-π interactions with Trp residues in spite of its lower occurrence in metalloproteins is an important observation in the present study and thus Trp residues might play an important role in the structural stability of metalloproteins through Arg/Lys interactions with π residues.

The average energetic contributions from each Arg/Lys interaction with π residue pairs in metalloproteins and the energies for each interaction in all protein structures studied is presented in **Table 1 (see supplementary material)**. The energetic contribution from Arg-Trp is higher when compared with other interacting pairs and the energetic contribution from Lys-Tyr is minimal. Based on these observations; we suggest that, in the context of energetic contribution, the role of Arg-Trp pairs might be important. The highest cation-π energetic contribution [-10.27 Kcal/mol] is with the Lys 141 and Trp 215 (PDB ID 1DXQ_A) and the lowest [-2.28 Kcal/mol] is with Lys 170 and Trp 91 (PDB ID 2IQ6_A). The average Arg/Lys interaction energy with π residues is around -4.67 Kcal/mol and there is an average of -18.92 Kcal/mol Arg/Lys interaction energy with π residues per protein in metalloproteins.
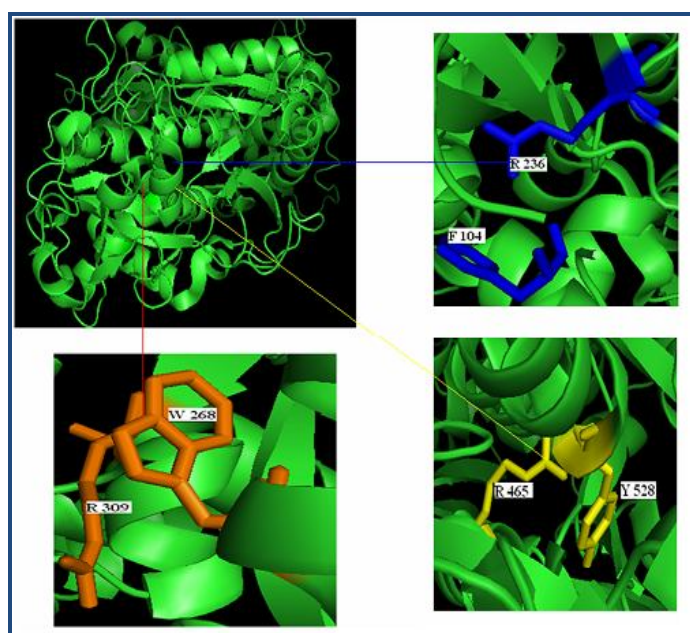

Figure 1: Arg interaction with Phe, Tyr and Trp in Lactoferrin (PDB ID 1BOL_A)

Secondary structure preferences
The secondary structure preferences are obtained from DSSP. The cation-π interacting residues are found to stabilize both the regular and non-regular secondary structural elements in metalloproteins. The helices are predominantly stabilized by Arg interactions with Tyr and Trp, while the coils and sheets are stabilized by Lys interactions with Phe. Hence, the preference of an amino acid to form Arg/Lys interaction with π residues in particular secondary structure is not the same as the preference of the amino acid for a particular secondary structure [35]. From our results in metalloproteins we assume that the stabilization patterns of these regular and non-regular secondary structures are independent of amino acid class.

Solvent accessibility patterns
We carried a systematic analysis of the solvent accessibility patterns for the Arg and Lys residues along with the interacting π residues in metalloproteins using DSSP. Solvent accessibility has been divided into three categories, buried, partially buried and exposed for different ranges of solvent accessibility values; <20, 20–50 and >50, respectively [36, 37]. Our results suggest that most of cationic residues are in the solvent exposed regions, while the majority of π residues prefer to be in the buried regions. These results are different when compared with the results from nucleic acid binding proteins [37]. From our results on the solvent accessibility patterns in metalloproteins we suggest that these interactions might stabilize the interface between the core and terminus in metalloproteins.

Stabilization centers
The stabilization centers for Arg/Lys-π interacting residues are studied. The percentages of cation-π interacting residues with located stabilization centers in metalloproteins are 30, 34, 23, 26 and 31 respectively for Arg, Lys, Phe, Tyr and Trp. From the results observed, we infer that the Arg/Lys-π interacting residues might contribute additional stability to metalloproteins.

Sequence conservation
The conservation scores of Arg/Lys-π interacting residues in each metalloproteins are studied. The percentages of cation-π interacting residues above the cut off conservation scores of 6 are 55, 43, 44, 65 and 54 respectively for Arg, Lys, Phe, Tyr and Trp. From our results we assume that the majority of the residues involved in Arg/Lys-π interactions are evolutionarily conserved and might have a significant contribution towards the stability of metalloproteins.

Sequence separation between residues involved in Arg/Lys-π interactions
We computed the sequential separation between Arg/Lys-π interacting pair in metalloproteins to determine the role of these interactions in the protein secondary and tertiary structures. Majority of the Arg/Lys-π interacting pairs are in long-range contacts and thus these interactions might contribute significantly to the stabilization of the native structure of the protein molecule and might help in maintaining the optimal conformation during binding of this class of proteins to metals. Hence any structural stability studies on the native protein molecule in metalloproteins should also take Arg/Lys-π interactions into consideration along with hydrogen bonds and other stabilizing interactions.

Metal binding sites

We generated schematic 2-D representations of protein-metal complexes from standard PDB file input using LIGPLOT program. The output is a postscript file giving a simple and informative representation of the intermolecular interactions and their strengths, including hydrogen bonds, hydrophobic interactions and atom accessibilities [34]. From our analysis on metal binding patterns in the dataset, we find that the residues involved in metal binding are amino acids that are not involved in Arg/Lys- π interactions as. But few metal binding pockets are stabilized by Arg/Lys-π interactions. The Arg-Tyr stabilizing residues (R 38-Y 35) in Ferredoxin [PDB ID 1AWD_A] along with metal binding residues and the amino acid residues located in metal binding pockets. We could not locate amino acids in metal binding pockets in most structures. Based on our results, we suggest that Arg/Lys interactions with π residues may not contribute to functional specificity (binding of metals) in metalloproteins.

Discussion:

From our results presented we infer that there is an increased preference for Trp over Tyr and Phe in energetically significant Arg/Lys-π interactions. This might be due to the fact that, the larger volume of Trp allows it to contact a greater number of cations relative to Phe or Tyr [8]. In terms of energetically significant Arg/Lys interactions with π residues in the study group, the energies from Arg interactions are higher when compared with Lys interactions with π residues. This phenomenon may be due to the fact that the side chain of Arg is larger and less well water-solvated than that of Lys, it likely benefits from better van der Waals interactions with the aromatic ring [8]. In addition, the side chain of Arg may still donate several hydrogen bonds while simultaneously binding to an aromatic ring, whereas Lys would typically have to relinquish hydrogen bonds to bind to an aromatic [38]. There is an average of one energetically significant Arg/Lys-π interaction for every 84 amino acids in the metalloproteins investigated. These results are comparable to the observations in glycoproteins and tumor necrosis factors while they differ with the observations with RNA binding proteins, lipid binding proteins, cell adhesion molecules and interleukins [39-41]. To understand the interactions that confer secondary structural conformational stability in proteins it is important to know the conformational preferences of amino acids.

The structural preferences of amino acids were introduced and calculated a long time ago, and it was known that different amino acids have distinct preferences for the adoption of helical, strand and turn conformation [42-44]. Although much were known about secondary and tertiary protein structure and folding, the process of folding is not understood completely. The molecular mechanism of protein self-assembly is still an open question [45]. It is believed that the energetics of side chain interactions dominate protein folding processes. However, it was shown that secondary structure can determine native protein conformation, devoid of side chains [46]. Recently, a backbone-based theory of protein folding was proposed, where the protein folding mechanism is based on backbone hydrogen bonding [47], while α-helix and β-sheet propensities are closely connected with the energetics of peptide H-bonds [48]. Hence, we thought it would be useful to study the secondary structure preferences of amino acid

residues involved in cation-π interactions so that the importance of these interactions in the structural stability with respect to global and local conformations can be clearly understood. The secondary structure preferences of Arg/Lys-π interacting residues suggest that the occurrences of amino acids in different secondary structural elements are not usually based on their physico-chemical properties and it varies depending on the protein class. The secondary structural preference results obtained in the present study are similar to the results observed with RNA binding proteins and they are not comparable with results of other conjugated proteins [37, 39].

An interesting question concerns the location of cation-π interactions within protein structures. Cationic residues generally prefer to be on the surface of proteins whereas aromatic amino acids prefer to remain in the hydrophobic core. Because a cation-π interaction contains both a cation and an aromatic, it is not clear whether the interacting pairs should prefer to be located on the surfaces of proteins or in the cores [8]. Solvent accessibility determines the importance of local versus non-local interactions along the protein sequence [49, 50]. Solvent accessibility is also an important parameter in determining the structural stability of a protein molecule [50]. For more than three decades, experimenters and theorists have tried to understand the kind of interactions that govern first stages of protein folding and lead to the formation of folding intermediates, and the forces that maintain protein stability [51]. The main questions concern the relative importance of hydrophobic versus more specific interactions, and of local versus non-local interactions along the sequence. It is indeed possible to detect stability changes caused by mutation, in the folded and transition states, by measuring and comparing the changes in unfolding and activation free energies. Presently there are a lot of experimental data on folding free energy changes upon mutation obtained by site-directed mutagenesis experiments, but only a few theoretical methods have been developed to predict such stability changes. Some of these methods are based on detailed atomic models and others on rougher descriptions of protein structure [52].

Their performances are in general, evaluated by comparing the calculated folding free energies to the measured ones and are reasonably good. In most studies, the mutated residues are buried in the protein core; since hydrophobic interactions dominate in these regions, the energetic criteria obviously involve hydrophobicity. In the few reported studies analyzing mutations of solvent accessible residues, the stability changes are correlated with statistical propensities of single amino acids to be in α-helices or β-strands [53], or with distance-dependent residue-residue potentials [49-53]. Thus, ours results on solvent accessibility would be meaningful to analyze the preferences of residues that are involved in these interactions. Since, the interacting residues are present in both the solvent exposed and buried states these interactions might stabilize the intermediate region between the terminus and the core regions within the metalloprotein molecule. It is also noteworthy to discuss here that the solvent accessibility patterns observed in the present study are not comparable with observations from nucleic acid binding proteins which also exist as metal-protein complexes [37]. Our observations with metalloproteins in the present study suggest that Arg/Lys-π interacting pairs might be contributing to the global conformational stability of these proteins as most

of the interacting pairs are in long-range contacts. These observations are comparable to the results in conjugated proteins [39] and interleukins [40]. The degree to which an amino acid position is recessive to substitutions is strongly dependent on its structural and functional importance. An amino acid that plays an essential role, e.g. in enzymatic catalysis, are likely to remain unaltered in spite of the random evolutionary drift. Hence, the level of evolutionary conservation was used often as indicator for the importance of the position in maintaining the protein's structure and or function. Conservation score is a useful parameter for the identification of conserved residues in a protein sequence [29, 30]. The conservation patterns in the present study indicate that more than half of the residues involved in these interactions are evolutionarily conserved.

This suggests that apart from amino acid residues responsible for specificity; Arg, Lys and π residues also are not prone to changes due to the evolutionary process. Stabilization centers can be defined as clusters of residues that are involved in medium or long range interactions [21, 54]. Any residue is considered part of stabilization center if it is involved in medium or long range interactions and if two supporting residues could be selected from both of their flanking tetra peptides, which together with the central residues form at least seven out of the nine possible contacts [54]. A significant percentage of Arg/Lys-π interacting residues also are located as stabilization centers and thus might provide additional stability to these proteins. On the whole, from the results presented we infer that Arg/Lys-π interactions might play an important role in the global structural stability of metalloproteins. To conclude, our results on the bioinformatics analysis of Arg/Lys-π interactions in metalloproteins will be useful for further structural studies on these classes of proteins and also make a strong case to consider these interactions to be considered along with hydrogen bonds, hydrophobic interactions and other covalent interactions in structural studies.

References:
[1] Rulisek L & Vondrasek J, *J Inorg Biochem.* 1998 **71**: 115 [PMID: 9833317]
[2] Sousa C *et al. Nature Biotech.* 1996 **14**: 1017 [PMID: 9631043]
[3] Barker PD, *Curr Opin Struct Biol.* 2003 **13**: 490 [PMID: 12948779]
[4] Lu Y *et al. Chem Rev.* 2001 **101**: 3047 [PMID: 11710062]
[5] Takahashi O *et al. Chem Rev.* 2010 **110**: 6049 [PMID: 20550180]
[6] Dougherty DA, *Science.* 1996 **271**: 163 [PMID: 8539615]
[7] Ma JC & Dougherty DA, *Chem Rev.* 1997 **97**: 1303 [PMID: 11851453]
[8] Gallivan JP & Dougherty DA, *Proc Natl Acad Sci USA.* 1999 **96**: 9459 [PMID: 10449714]
[9] Berman HM *et al. Nucleic Acids Res.* 2000 **28**: 235 [PMID: 10592235]
[10] Sussman JL *et al. Science.* 1991 **253**: 872 [PMID: 1678899]
[11] Zhong W *et al. Proc. Natl Acad Sci USA.* 1998 **95**: 12088 [PMID: 9770444]
[12] Chakravarty S & Varadarajan R, *Biochemistry.* 2002 **41**: 8152 [PMID: 12069608]
[13] Gromiha MM *et al. Prep Biochem Biotech.* 2002 **32**: 355 [PMID: 12455828]
[14] Shi Z *et al. J Am Chem Soc.* 2002 **124**: 3284 [PMID: 11916412]
[15] Burghardt TP *et al. Biopolymers.* 2002 **63**: 261 [PMID: 11807753]
[16] Mulhern TD *et al. J Mol Biol.* 2000 **297**: 989 [PMID: 10736232]
[17] Gromiha MM, *Biophys Chem.* 2003 **103**: 251 [PMID: 12727287]
[18] Kabsch W & Sander C, *Biopolymers.* 1983 **22**: 2577 [PMID: 6667333]
[19] Rost B, *J Stru Biol.* 2001 **134**: 204 [PMID: 11551180]
[20] Kloczkowski A *et al. Proteins.* 2002 **49**: 154 [PMID: 12210997]
[21] Dosztányi Z *et al. J Mol Biol.* 1997 **272**: 597 [PMID: 9325115]
[22] Sander C & Schneider R, *Prot Struct Funct Genet.* 1991 **9**: 56 [PMID: 2017436]
[23] Boeckman B *et al. Nucleic acids Res.* 2003 **31**: 365 [PMID: 12520024]
[24] Altschul SF *et al. Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
[25] Thompson JD *et al. Nucleic Acids Res.* 1994 **22**: 4673 [PMID: 7984417]
[26] Saitou N & Nei M, *Mol Biol Evol.* 1987 **4**: 406 [PMID: 3447015]
[27] Pupko T *et al. Bioinformatics.* 2002 **18**: S71 [PMID: 12169533]
[28] Mayrose I *et al. Mol Biol Evol.* 2004 **21**: 1781 [PMID: 15201400]
[29] Glaser F *et al. Bioinformatics.* 2003 **19**: 163 [PMID: 12499312]
[30] Landau M *et al. Nucleic Acids Res.* 2005 **33**: W299 [PMID: 15980475]
[31] Gromiha MM & Selvaraj S, *J Biol Phys.* 1997 **23**: 151
[32] Gromiha MM *et al. Int. J Biol Macromol.* 2004 **34**: 203 [PMID: 15225993]
[33] Selvaraj S & Gromiha MM, *Biophys J.* 2003 **84**: 1919 [PMID: 12609894]
[34] Wallace AC *et al. Protein Eng.* 1995 **8**: 127 [PMID: 7630882]
[35] Malkov SN *et al. J Mol Model.* 2008 **14**: 769 [PMID: 18504624]
[36] Gromiha MM *et al. Protein Eng.* 1999 **12**: 549 [PMID: 10436080]
[37] Anbarasu A *et al. Int J Biol Macromol.* 2007 **40**: 479 [PMID: 17197018]
[38] Mitchell JB *et al. J Mol Biol.* 1994 **239**: 315 [PMID: 8196060]
[39] Anbarasu A & Sethumadhavan R, *J Theor Biol.* 2007 **247**: 346 [PMID: 17451749]
[40] Anbarasu A *et al. Cytokine.* 2006 **35**: 263 [PMID: 17055289]
[41] Sophiya K & Anbarasu A, *Protoplasma.* 2011 **248**: 673 [PMID: 20978808]
[42] Chou PY & Fasman GD, *Biochemistry.* 1974 **13**: 211 [PMID: 4358939]
[43] Chou PY & Fasman GD, *Adv Enzymol Relat Areas Mol Biol.* 1978 **47**: 45 [PMID: 364941]
[44] Padmanabhan S *et al. Nature.* 1990 **344**: 268 [PMID: 2314462]
[45] Fitzkee N *et al. Trends Biochem Sci.* 2005 **30**: 73 [PMID: 15691652]
[46] Fleming PJ *et al. Prot Sci. 2*006 **15**: 1829 [PMID:16823044]

BIOINFORMATION

[47] Rose GD *et al. PNAS USA.* 2006 **103**: 16623 [PMID: 17075053]

[48] Baldwin RL, *J Mol Biol.* 2007 **371**: 283 [PMID: 17582437]

[49] Gilis D & Rooman M, *J Mol Biol.* 1996 **257**: 1112 [PMID: 8632471]

[50] Gilis D & Rooman M, *J Mol Biol.* 1997 **272:** 276 [PMID: 9299354]

[51] Munoz V & Serrano L, *Fold Des.* 1996 **1**: 71 [PMID: 9079385]

[52] Sippl MJ. *Curr Opin Struct Biol.* 1995 **5**: 229 [PMID: 7648326]

[53] Munoz V & Serrano L, *Proteins.* 1994 **20**: 301 [PMID: 7731949]

[54] Dosztányi Z *et al. Bioinformatics.* 2003 **19**: 899 [PMID: 12724305]

BIOINFORMATION

Supplementary material:

Table 1: Occurrence of Arg/Lys-π residues in metalloproteins

PDB ID	%R /K	%F	%Y	%W
1AWD_A	5.4	2.1	5.3	0.0
1B0L_A	12.9	4.3	3	1.4
1D4A_A	12.1	6.2	3.7	2.2
1DLM_A	8.4	3.6	4.2	0.6
1DXQ_A	11.4	5.9	3.7	2.2
1HFX_A	10.5	2.4	4.1	2.4
1LNH_A	11.4	4.4	3.7	1.8
1OI8_A	11.2	3.4	4.0	1.3
1P9R_A	13.3	2.5	1.1	0.0
1RRH_A	11.3	4.4	3.6	1.8
1UNF_X	10.7	5.6	4.2	3.7
1VEW_A	11.2	5.4	3.4	2.9
1WZD_A	11.5	4.3	3.8	1.0
1XM5_A	4.6	2.6	2.6	1.3
1Y67_A	8.8	3.8	4.3	2.9
1YGE_A	10.2	3.6	4.4	1.7
1Z6R_A	8.0	2.1	2.1	0.5
2AFM_A	7.8	4.6	3.7	2.2
2CF4_A	10.1	4.4	3.5	0.9
2DD5_C	12.1	2.9	3.9	2.7
2E46_A	8.3	3.2	1.9	0.0
2IF6_A	13.8	3.6	5	2.8
2IQ6_A	4.8	3.4	4.4	1.7
2IVN_A	12.7	4.0	3.4	0.3
2NXF_A	9.3	4.2	3.5	1.6
2P18_A	9.2	5.7	4.2	0.4
2RDV_A	7.7	3.8	5.8	1.9
2VW7_A	9.4	3.3	3.3	1.5
2Z68_A	11.5	4.3	3.8	1.0
2ZOW_A	9.2	2.6	0.7	0.0
3D19_A	13.1	7.7	1.9	0.8
3EXM_A	11.9	2.9	1.4	8.1
3FW1_A	9.6	6.1	3.5	2.6
3HIP_A	8.6	2.5	1.2	3.7
3PCG_M	9.7	4.6	3.5	1.6
Mean	10.05	4.01	3.42	1.76

R- Arginine; K-Lysine; F-Phenylalanine; Y-Tyrosine; W-Tryptophan