Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 4 Number 2 (2011) pp. 127-133

http://www.ripublication.com/jcib.htm

Generation and Analysis of New Metalloprotein Database

***[1]Arul Mugilan and [2]Ajitha Mohan**

[1]Department of Bioinformatics, Karunya University,
Karunya Nagar, Coimbatore, Tamil Nadu, India
**Corresponding Author E-mail: bioinformaticsmugil@gmail.com*
[2]Department of Biotechnology, Kalasalingam University, Madurai, India
E-mail: aji.sujo@gmail.com

Abstract

Over the last decade, biological database plays a major role to provide the long term memory of computer operations take on a variety of names depending on their structure, contents and amount of data about biological macromolecules. Here, metalloprotein database facilitates budding researchers and scientist to decipher files composed of records, each containing data together with a set of operations for searching, sorting, recombining and to manipulate various information about metalloproteins. Since, the cut-off values plays a valuable role in computer aided drug designing, it is essential to analyze this value. Our aim is to generate a new metalloprotein database with a standard cut off value, which particularly shines as a repository for enormous metalloprotein data which falls under the category of first and second shell ligands. This serves as a platform for three dimensional structural analysis of metalloproteins, molecular modeling and drug designing. Implementation of this new database, leads to retrieval of specific residue from an obscure protein databank .It gains the information about cutoff values by analyzing the interaction between metals and ligands with other sorted elements in metalloproteins.

Keywords: Metelloprotein, database, chi-square value, PDB.

Introduction

Metalloprotein makes a possible life on earth, due to its crucial role in biological pathway and ultimately control the binding activity of metal ions. Metalloprotein depicts protein that contains a metal ion as a cofactor. As a cofactor [1] these metal

ions are critical to the metabolism, structure, and stability of metalloproteins. The presence of this metal ion allows metalloenzymes [2] to perform their various functions effectively, which cannot be easily performed only with the presence of limited set of functional groups found in amino acids alone. In fact, numerous essential biological mechanisms require metal ions and this shows their great biological and medical elucidated significance.

In a metalloprotein the central metal ion is usually coordinated by ligands [3] such as nitrogen, oxygen or sulfur atoms of an amino acid in the polypeptide chain. There are some exceptions from the general structure of amino acids such as Glycine , is not chiral since its side chain is a hydrogen atom and proline, where the hydrogen atom is replaced by a bond to the side chain. As an outstanding element, Cysteine contains thiol group which is consumed from its three dimensional structure. These structural properties were obtained from experimental studies such as X-ray crystallography [4] and NMR spectroscopy.

MDB [5] -Metalloprotein Database contains quantitative information on all the metal-containing sites available from structures in the form of PDB distribution. Their ultimate goal is to design metalloproteins with distance cutoff parameter of 3A, allowing the interactive visualization of geometrical and functional information garnered from the MDB. But the known caveat of this database is site information which includes only first shell ligands, and do not recognize any other shell of the ligands from the cluster sites.

This article describes about our novel attempt of identification of significant ligands in metalloproteins and generation of New Metalloprotein Databases-NMDB with the cutoff distance parameter of about 5A° which includes the recognition of second shell ligands. Ligand patterning based on the atomic co-ordinate Values [6] acts as an essential property in structural analysis. So, our database was generated on the basis of statistical analysis of ligand geomentry and their atomic co-ordinates. Computational Modeling for metalloproteins can be done with the information of coordination sites [7] of metallosystems. Thus, our database can be utilized for creation of innovative attempts in computational modeling

Earlier studies indicate that level of sequential information has been increased repertoire when compared to that of three dimensional counterpart. Any vital information obtained by the analysis of three dimensional structure in terms of sequence [8] will have definite impact on the structure prediction analysis. Thus the valuable sequential information fulfills the thrust of the researchers in the field of secondary structure prediction. Based on this sequential information of proteins, the frequency of occurrence [9,10] can be determined for their geometrical and three dimensional structural information patterns. Based on these ideas, the identification of distribution of significant ligands can be resolved on the basis of the following calculations.

Methodology

As an initial step, 371 Files for 30 metals were retrieved from PDB [10] database. Python Program was executed for recuperation of selected number of ligands from the

whole complex structure of the biomolecule. This sorting process takes place on the basis of co-ordinate values with the distance cutoff parameter of about 5A. Reckoning up on this program results, visualization of all recuperated structures using Rasmol software was done for the identification of nearby ligands which is attached to that of the central metal ion. Then the analysis of first and second shell ligands, based on the visualization of structural geomentric phenomenon was performed.

In order to retrieve the frequency of occurrence for all 20 amino acids in each metal, a new python program was fetched. Further, based on this new program results, chi square value calculations were accomplished from frequency of occurrence. In general chi square value is a quantitative measure used to determine whether a relationship exists between two categorical variables. Here this calculation includes *statistical analysis used to test how well the distribution of a set of observed data matches a theoretical probability distribution. Thus,* chi square analysis acts as an ensue for determining significant ligand for each metal.

Chi Square Value can be predicted from frequency of occurrence by using the following formula:

$$\text{Chi square} = [(\text{observed value} - \text{Expected value})^2 / \text{Expected value} \quad (1)$$

As a consequence, depending on the variation of their frequency of occurrence, *com*parative analysis between various ligands takes place in order to predict their significant role in specific metalloprotein. *Relaying up on this* comparative analysis, favourable & unfavourable ligands were identified for each metal ion. Thus, designation of specific significant ligands from the first shell level as well as second shell level which interacts with that of central metal ions were manipulated. These favourable ligands were found to play an essential role in drug designing process.

With the regular updating of all these valuable information a New Metalloprotein Database -NMPD was generated with the distance cutoff parameter of 5A radius. It leads visceral structural analysis of central metal ion with that of specific ligands present in both first and second shell level. Thus, the recognition of second shell ligands acts as a novelty in this article. This novelty breaks the limitation of site information of MDB-Metalloprotein Database with 3A which do not recognize the second shell level ligands.

Result and Discussion

Our freshly generated New Metalloprotein Database (NMDB) with the radius of 5A explore an intuitive representation of relationship among large group of metalloproteins using manifestation process. Our database acts as a commercial free, non redundant database, which consumes less time by its readily available database about metalloproteins. So our NMPD will be enrolled as a user friendly, key structural informative bank for structural bioinformaticians, Insilico novelist, drug designers and other metalloprotein researchers. It is intercommunicated to fulfill the need of quantitative information about the designing of metal binding site. Our generated database owes a conventional pathway for utilization of significant amino acid which acts as a ligand during computer aided drug designing in metalloproteins..

Further it can illustrate, the influence of various structural factors for understanding structure-function relationship in metalloproteins based on the reduction potential of metal bound protein scaffolds. This takes place in all highlighting areas related to structural analysis in a global scale. Applications in Biotechnology and Bioinformatics field, impart them with functions not found in nature. Relaying up on their capabilities, database can be equipped with properties such as higher stability and greater efficiency. These additional developments will be beneficial to computational and structural biologist in a discriminate way as it acts as a open resource for them to visualize and analyse structural data.

Here, our final result of comparative analysis predicts the specific amino acid which acts as a significant ligand binding group to that of each central metal ions. Chi square value determines the deviation parameter of various amino acids in each metalloprotein. Our comparative analysis, indicates that most of the amino acid shows the value in range of 0.1 to 5.0 but the cysteine alone shows a high positive deviation from them with the values of 14.97277, 22.90322, 2.23231,1.007387, 19.78003, 23.48438, 15.34446, 4.219487, 12.62617, 11.55159, 22.61447 in arium, Calcium, Godolinium, Iodine, Magnesium, Molybdenum, Selenium, Silver, Thallium, Ytterbium, Telurium respectively. Molecular Interaction of Ytterbium with that of Cysteine is shown in Fig:1. Similarly, Glycine was found to be significant in Holmium, Lithium, Magnesium, Terbium.

Molecular Interaction of Holmium with that of Glycine is shown in Fig:2. Concominently, Arginine, Methionine, Asparagine, Glutamine were found to be act as significant ligands in Manganese, Platinium, Silicon and Vanadium metals respectively.

In Figure 1 Pink atom indicates Ytterbium metal and red atom indicates Cysteine. In figure 2 pink atom represents Holmium metal and red atom represents Glycine

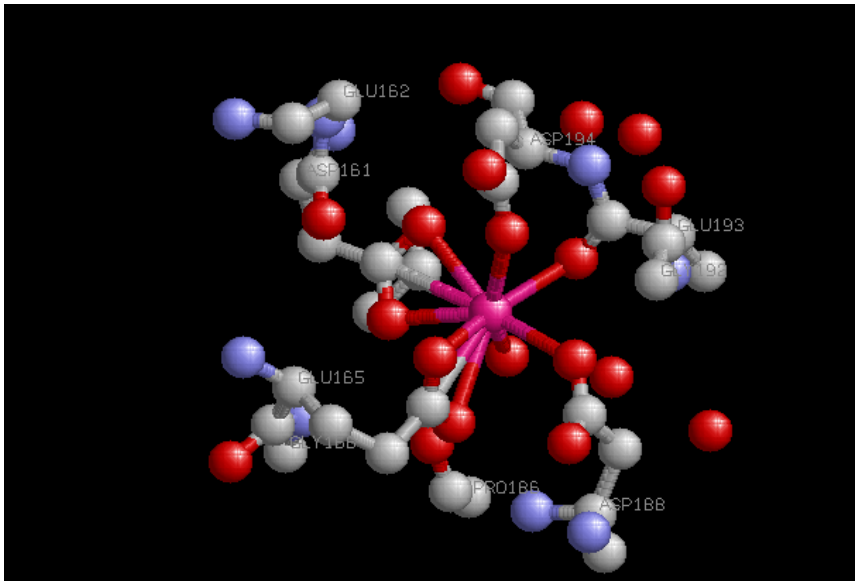


Figure 1: Visualisation of Molecular Interaction between Ytterbium and Cysteine.

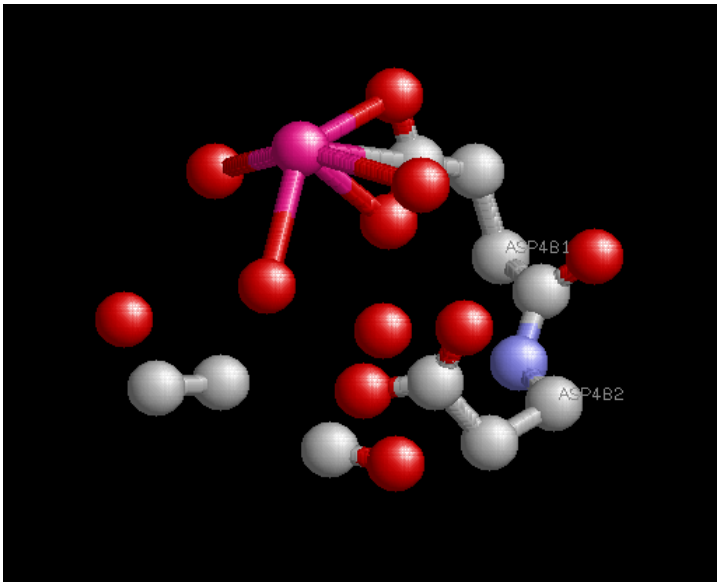
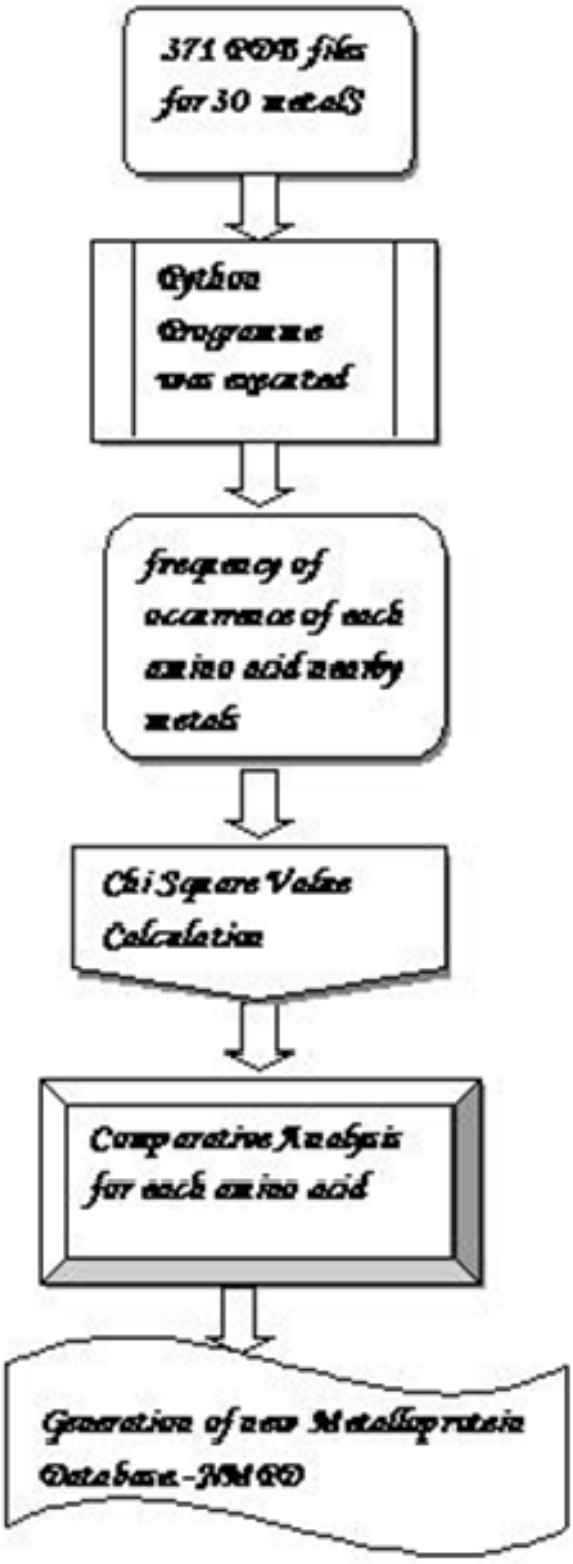


Figure 2: Visualisation of Molecular Interaction between Holmium and Glycine.

In accordance, Cysteine was found to play an outstanding role as it shows significant in high number of metals. This outstanding property of cysteine was depicted due to the presence of Thiol Group. This proves that most of the D-Block elements binds with that of nitrogen or sulphur containing ligands. Rasmol command line and Graphical representation acts as a proof in order to substantiate our results. As a visual treat, rasmol structure gives an additional evidence for generation of our new NMPD database. Thus our database will be very beneficial to innovators and drug designers who are in need of information about the significant ligands which binds to that of specific metal ions. In addition, it uplifts the research work, by indicating the comparison of parametric value of each amino acids in different metals. Our database confines an enormous information about metalloproteins in a descriptive pathway. It also acts as a specialized data mining tool to perform computational analysis of metalloproteins and for predicting its related sorting data.

Conclusion

In emerging fields, due to its outstanding quality NMPD will be implemented in following areas of protein sequence, validation analysis, user database interface development. It can be applied in number of tasks, such as metal-protein complex structure prediction, drug discovery for metalloproteins, and in designing interactions of metalloproteins. Caveats of MDB database were eradicated in a tremendous way due to our new innovative attempts. Extraordinary *properties of this databases are easy to use, runs on many platforms, requires only meager computational resources and it is extremely powerful.* In future, new database will be generated by us for other biomolecules such as nucleic acids with a standard cut off parameter.



Flow Chart for generation of Metelloproteins

References

[1] R. Bittl and P. *Hildebrandt*, 2010, "Protein-cofactor interactions in biological processes", *Chem. Phys. Chem.,* 11(1), pp. 1075-1076

[2] C.L. Davies, E.L. *Dux* and A.K. Duhme-Klair, 2009, "Supramolecular interactions between functional metal complexes and proteins", E.pub, 46(1), pp.10141-10154.

[3] D.J. Mercer and S. Loeb, 2010, "Metal-based anion receptors: an application of second- sphere coordination", J. Chem. Soc. Rev., 39(1), pp. 3612-3620.

[4] A. *Deniaud*, E. Moiseeva, V. Gordeliy and E. Pebay-Peyroula, 2010, "Crystallography of membrane proteins: from crystallization to structure", Methods. Mol. Biol., 654(1), pp. 79-103

[5] R.B. Pauling L. Corey, 1951, "Atomic Coordinates and Structures Factors for Two helical Configurations of Polypeptide Chains", Proc. Natl. Acad. Sci., 37(1), pp. 235-240.

[6] J.A. Ibers and R. H. Holm, 1980, "Modeling coordination sites in metallobiomolecules", Science, 209(1), pp. 223-235.

[7] J.M. Castagnetto, S.W. Hennessy, V.A. Roberts, E.D. Getzoff, J.A. Tainer and M.E. Pique, 2003,"MDB: the Metalloprotein Database and Browser at The Scripps Research Institute", Nucleic. Acids. Res. 30(1), pp. 379-382.

[8] G. D. Fasman and P.Y. Chou, 1974, "Prediction of protein conformation", Bio.Chem., 13(1), pp. 222-245.

[9] K. Veluraja and S.A. Mugilan, "Generation of deviation parameters for amino acid singlets, doublets and triplets from three-dimensional structures of proteins and its implications in secondary structure prediction from amino acid sequences", 2000, J.Bio.Sci., 25(1), pp. 81 – 91.

[10] S. Arul Mugilan, Ajitha, Mary cathrin, Monica kumar and Devi,2010, "In Silico Secondary Structure Prediction Method (KLUSP) using Comparative Analysis", Trends in Bioinformatics,5 (1), pp. 11-19.

[11] U. Hobohm, M. Scharf, R. Schneider and C. Sander, 1992, "Selection of representative protein data sets", Protein Sci., 3(1), pp. 409-417.