Recognition of metal ions using ProMOL to improve protein function assignment

Venkata Aditya Kovuri[1], Paul A. Craig[2], Herbert J. Bernstein[3]

[1]RIT School of Life Sciences; [2]RIT School of Chemistry and Materials Science; [3]Dowling College, Department of Mathematics and Computer Science

Abstract

ProMOL is a plugin for the PyMOL molecular graphics environment that utilizes the 3-dimensional visualization and measurement capacities of PyMOL to align enzyme active sites with a collection of annotated active sites found in our motif template library. Through this method, ProMOL/PyMOL uses structural alignment to predict enzyme function for proteins of unknown function. The purpose for this initiative can be understood by looking at the number of structures in the Protein Data Bank that are labeled as having "unknown function"; to date, there are 3975 such structures in the Protein Data Bank. ProMOL undergoes constant revision and development from the various team members from RIT and Dowling College. One of these recent developments includes an additional capability of ProMOL to recognize metal ions in active sites, as nearly 40% of all the PDB structures contain at least one metal ion[1]. This is biologically significant as metal ions such as Mn^{+2}, Zn^{+2}, Ca^{+2}, and Fe^{+2} play a vital role in both the structure and function of many different proteins. The addition of metal ion recognition to ProMOL/PyMOL promises to lead us to better function assignments for this collection of protein structures.

Introduction

ProMOL is a bioinformatics tool that is used in the determination of the function of proteins that have a known structure listed in the PDB. This can be categorized as a tool that utilizes the structure based function prediction method.

- Our aim is to successfully predict the function of unknown function protein structures with the aid of this Bioinformatics tool and then to validate the results through in-vitro testing.
- About 40% of all the structures in PDB contain metal ions. In most of these structures, metal ions play an important role in expressing their function.
- To carry out biological function, it is estimated that about 30-40% of proteins need metal ions.

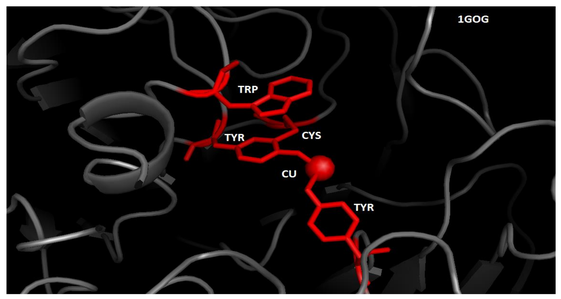

Figure 1: This figure shows the 3-dimensional visualization of the active sites in 1GOG with Copper present in the active sites of the residue.

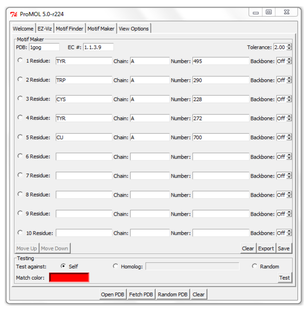

Figure 2: The motif maker interface for taking in the user input parameters for the residues to identify the active sites.

Methods:

ProMOL's incentive for recognizing metal ions in their active sites is to increase the understanding of potential functions of metallo-enzymes and the functionality that these metal ions bring. Proteins with metal ions in their active sites have Magnesium in a larger percentage, when compared to other metal ions such as Ca^{+2}, Zn^{+2}, Mn^{+2}, Fe^{+2},Cu^{+2},Co^{+2}, and Mo^{+}. Each of these have various biological function in proteins; copper helps in the respiratory process, calcium in the hardening of the skeletal system and iron in enhanced brain function. Here, we generate motif templates with metal ions in them and search for interesting alignments. Shown in Figure 4 is a near perfect alignment between 1jms and its homolog 1kdh; this is an example to show how effective ProMOL is at identifying these active sites and corresponding metal ions.
ProMOL[3] software is primarily coded in Python as its parent software PyMOL. Hence, the motif files that are generated through ProMOL are in a Python Script format.

The motifs are created in the motif maker section of ProMOL. The information for the amino acids within the active sites is obtained from the Catalytic Site Atlas (CSA) [4] and the information for the metal ion residues can be obtained from the metalPDB database[1], metal Macie database[2] or the respective PDB file of the protein. The amino acid residues and metal ions in the active sites are loaded as shown in Figure 2 and ProMOL validates these results and visualizes the active sites as shown in Figure 3.

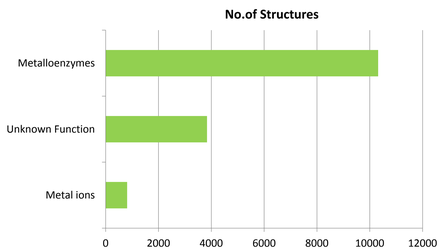

Figure 3: This figure is a graphical representation of the number of structures present in the PDB. The X-axis represents the number of structures and the y-axis represents the category that these structures belong to

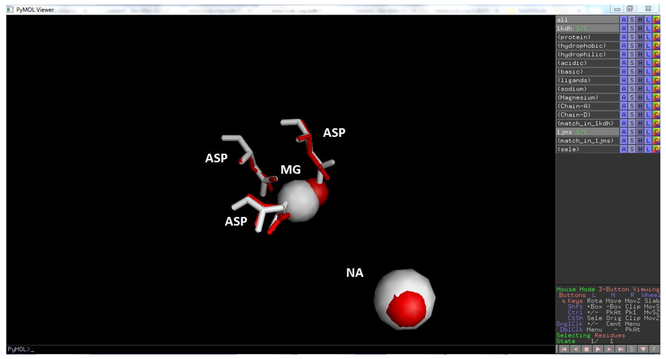

Figure 4: This figure visualizes the alignment between 1JMS and its homolog 1KDH (both are terminal deoxynucleotidyl transferases). We can see the clear overlap between the aspartic acid residues in the two active sites and especially the metal ions, Sodium and

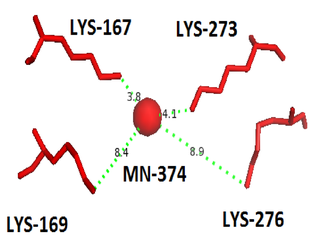

Figure 5: This figure shows the active site of 1MUC, a muconate lactonizing enzyme, with Manganese as an essential part of the active site.

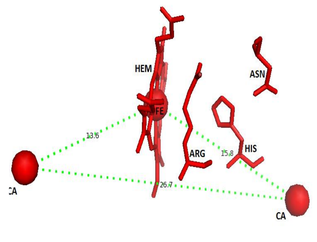

Figure 6: This figure shows the active sites of 7ATJ, a recombinant horseradish peroxidase, containing a Heme group along with Calcium.

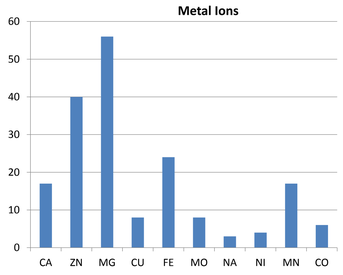

Figure 7: This figure shows a bar graph representation of the diversified distribution of metal ions that are present in the M-set of motif templates.

Results:

The ProMOL Motif library at present consists of "P-set" and "A-set", with A-set containing the motifs that are automatically generated with the help of automated motif generation through ProMOL[5]. The capability to recognize metal ions and other prosthetic groups in ProMOL apart from the 20 Amino acids, enabled us to create a new set of motifs that contain metal ions and other residues such as SF4 (Sulfur tetrafluoride), F3S (FE3-S4 cluster), CLF (FE(8)-S(7) cluster), CFM (FE-MO-S cluster) in their active sites. These motifs combined constitute the "M-set". At present, there are 54 motifs in the M-set and this is expected to increase with ProMOL's capability to recognize other residues. The active site information is obtained from Metal-Macie[2] and Metal PDB[1] databases.

```
'''
FUNC:U_1jms_2_7_7_31
PDB:1jms
EC:2.7.7.31
RESI:asp,na
LOCI:a-434,702;
'''
cmd.select('asp1', 'n. CB&r. asp w. %s of n. NA&r. na'%(d*15.44))
cmd.select('asp2', 'n. CG&r. asp w. %s of n. NA&r. na'%(d*13.93))
cmd.select('asp3', 'n. OD1&r. asp w. %s of n. NA&r. na'%(d*13.53))
cmd.select('asp4', 'n. OD2&r. asp w. %s of n. NA&r. na'%(d*13.24))
cmd.select('asp', 'br. asp1&br. asp2&br. asp3&br. asp4')
cmd.delete('asp1')
cmd.delete('asp2')
cmd.delete('asp3')
cmd.delete('asp4')
cmd.select('na1', 'n. NA&r. na w. %s of n. CB&asp'%(d*15.44))
cmd.select('na2', 'n. NA&r. na w. %s of n. CG&asp'%(d*13.93))
cmd.select('na3', 'n. NA&r. na w. %s of n. OD1&asp'%(d*13.53))
cmd.select('na4', 'n. NA&r. na w. %s of n. OD2&asp'%(d*13.24))
cmd.select('na', 'br. na1&br. na2&br. na3&br. na4')
cmd.delete('na1')
cmd.delete('na2')
cmd.delete('na3')
cmd.delete('na4')
cmd.select('U_1jms_2_7_7_31', 'asp|na')
cmd.delete('asp')
cmd.delete('na')
```

Figure 8: This is an image of a motif file that is generated through ProMOL. When residues are set to be tested in the motif maker interface of ProMOL. The distance from each atom in the residue to the metal ion is calculated to give positional constraints and identify the active site from the visualization window.

Below are a list of PDB IDs selected that contain metal ions in their active sites. These PDB IDs combined form the M-set.

M-SET Motifs:

1AH7	1ALK	1AQ2	1B57	1BG0	1BZY	1C9U
1CA2	1CTT	1D8C	1DII	1DL2	1DO8	1DQS
1F7L	1FCB	1FUA	1G72	1GIM	1GOG	1GSA
1HXQ	1IR3	1ITQ	1J09	1LOK	1MUC	1N20
1O8A	1P7L	1POW	1PVD	1PYM	1Q0N	1QLH
1QUM	1RDD	1RU4	1SML	1SOX	1TRK	1UBY
1UW8	1V25	1ZIO	2AG0	2CPO	2PHK	2QF7
2TOH	5EAT	7ATJ	13PK			

Conclusion & Future Plans:

1. The next step is to generate motifs for all the PDB IDs containing metal ions and run a parallel search through other databases that have information on these metal ion binding sites.
2. We also plan to run the entire PDB database and its homologous entries through this method to obtain more information about the proteins with metal binding sites.
3. Incorporating Automated generation of motifs containing metal ions through ProMOL, which in turn improves the size and range of our library .
4. Testing the M-set against the 3975 structures with unknown function.

References:

[1] Andreini, C., Cavallaro, G., Lorenzini, S., & Rosato, A. (2013). MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic acids research*, *41*(D1), D312-D319.
[2] Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L., & Thornton, J. M. (2009). Metal-MACiE: a database of metals involved in biological catalysis. *Bioinformatics*, *25*(16), 2088-2089.
[3] Lill, M. A., & Danielson, M. L. (2011). Computer-aided drug design platform using PyMOL. *Journal of computer-aided molecular design*, *25*(1), 13-19.
[4] Porter, C. T., Bartlett, G. J., & Thornton, J. M. (2004). The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic acids research*, *32*(suppl 1), D129-D133.
[5] *Osipovitch, M., Kovuri, V. A., Craig, P. A., & Bernstein, H. J.(2014). Automated Protein Structural Motif Generation and Inclusion of Metal Ion Motifs in a Structure-Based Protein Function Prediction Tool ProMOL.(Manuscript in preparation)*

SBEVSL
Structural Biology Extensible Visualization Scripting Language