Prediction of cation binding sites using structural water

L. A. Uroshlev, V. J. Makeev
Vavilov Institute of General Genetics RAS, *Engelhardt Institute of Molecular Biology RAS;*
leonidurоshlev@gmail.com

Introduction

Bivalent cations are well-known protein cofactors. They are present in such important proteins like calmoduline (Ca^{2+}) or insulin (Zn^{2+}). The ion binding pockets are local, have specific structures, and thus can be recognized as structural motifs. Usually structural templates or sequence motifs are used for identification of bound cations [1,2]. But in many cases these approaches don't work correctly, for example in the case of ion binding site recognition in apoproteins (i.e. proteins with structures obtained without ions). Ion binding pockets can be highly deformed in apoproteins. Also molecules of structural water (tightly bound in the structure) can contribute into ion binding. But that contribution is not taken into account in the template approach.

Method

We used statistical potential method [3] for prediction of most probable position and type of ion binding in a 3D structure. Statistical potentials for ion interacting with all types of protein atoms and structural water oxygens were developed. We built potentials for calcium, magnesium and zinc ions, the most common ions in PDB databank.

Results

Statistical potentials allow calculating pseudoenergy of ion binding in some point. That value shows statistical preference of an ion to be bound in the specifc protein site. Figure 1 compares contribution of different protein atoms and structural water into pseudoenergy of binding for magnesium, calcium and zinc. Water contribution into binding pseudoenergy was more than 70% for magnesium, about 30% for calcium and non important for zinc ions.

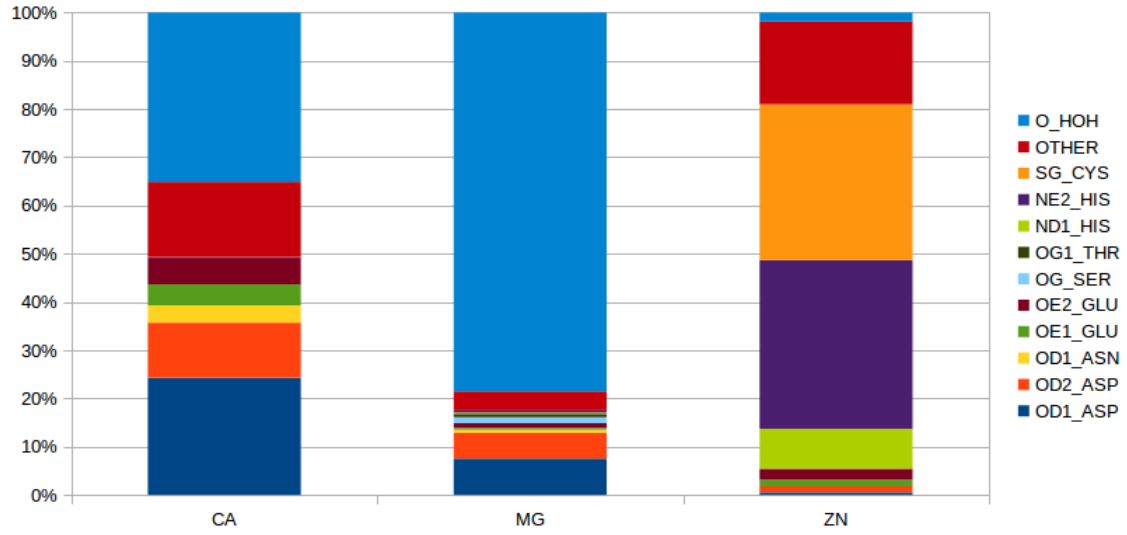

Fig. 1. Contribution of protein atoms of different types and structural water into pseudoenergy of binding for differ ion types. Only small distances (less 3Å) are considered.

To test the power of our approach for prediction of ions bound in apoproteins we selected a subset of non-homological apo-holo pairs from AH-DB database [4] with the help of PDBselect software [5]. All pairs were classified by RMSD of the structural alignment of apo- and holoproteins (denoted as Δ). For $\Delta < 0.5$ A our program predicted correctly 83% pairs, for $0.5 < \Delta < 1$ A the correct prediction was 69% of pairs, and for $\Delta > 1$ A the correct prediciton was 54% of pairs. These values are substantially greater than the correspondent prediction of other tools like CHED [Ref], Findsite [1], and Fold-X [Ref].

References

1. Brylinski, M., & Skolnick, J. (2011). FINDSITE-metal: Integrating evolutionary information and machine learning for structure-based metal-binding site prediction at the proteome level. *Proteins: Structure, Function, and Bioinformatics*, *79*(3), 735-751.

2. Levy, R., Edelman, M., & Sobolev, V. (2009). Prediction of 3D metal binding sites from translated gene sequences based on remote-homology templates. *Proteins: Structure, Function, and Bioinformatics*, *76*(2), 365-374.

3. Rahmanov, S., KULAKOVSKIY, I., Uroshlev, L., & Makeev, V. (2010). Empirical potentials for ion binding in proteins. *Journal of bioinformatics and computational biology*, *8*(03), 427-435.

4. Chang, D. T. H., Yao, T. J., Fan, C. Y., Chiang, C. Y., & Bai, Y. H. (2012). AH-DB: collecting protein structure pairs before and after binding. *Nucleic acids research*, *40*(D1), D472-D478.

5. Griep, S., & Hobohm, U. (2009). PDBselect 1992–2009 and PDBfilter-select. *Nucleic acids research*, gkp786.