

Title: Analyzing and predicting metal binding sites in proteins

PI: Dr. Andrew C.R. Martin, UCL

Co-I: Prof Steve Perkins, UCL

Aims and objectives

This project will create an automatically maintained database of metal binding sites in the PDB. From analysis of these sites, using machine learning methods, we will develop a new method for predicting metal binding sites. A major novel part of the work will be to examine and create prediction methods for sites which form *between* protein chains. These predictions will be tested through the collaborative part of the project which will be performed with Prof. Steve Perkins.

Biological background

Metal binding is critical to the function of many proteins. Metals can have purely stabilizing roles, or may be central to protein activity, perhaps involved in enzyme activity (Auld, 2001a). An understanding of metal binding may also allow us to engineer metal binding functions into proteins. Williams (1985) suggested that the environment of metals in proteins is quite different from the aqueous environment and this may confer unusual properties on the metal such that it may be better suited to a particular role in structural stability or catalysis. Low resolution structures (e.g. from structural genomics) may not reveal the presence of metals making prediction an important tool.

Some proteins bind metal ions through sites formed from two protein chains (Auld, 2001b). e.g. staphylococcal enterotoxin type A (PDB code 1sxt) binds zinc at an interface where one chain provides two His and one Asp while the second chain provides one His residue. When only the structures of individual chains are available, predictive methods are again important.

Work which has led up to this project

Previous work (Vallee & Auld, 1989, 1990) has shown that so-called 'hard' metals (e.g. Ca^{2+} & Mg^{2+}) tend to be bound by negatively charged sidechains of Asp and Glu (and to a lesser extent Asn, Gln, Ser and Thr) while borderline/soft metals (Zn^{2+} , Cu^{2+} , Fe^{2+}) tend to be bound by Cys and His. Yamashita, et al., (1990) showed that metal binding sites tend to be characterized by high 'hydrophobicity contrast' and more recent methods have combined sequence profiles with structural data (Sodhi et al., 2004). Our work (Gregory, et al., 1993) generated templates for metal binding sites from 28 available metal-binding protein structures and used these templates together with hydrophobicity contrast to search for sites at which a metal binding site might be introduced. Unpublished work (Staunton, Gregory, Martin and Rees) successfully introduced a metal binding site into CDR-L1 of a lysozyme binding antibody which retained binding for lysozyme and was shown to bind metal.

Recently Steve Perkins has shown that Factor H precipitates in the presence of zinc through formation of an inter-chain zinc binding site leading and polymerization of Factor H. A rotation project started to analyze and investigate Zn^{2+} -binding sites and use these to search for partial sites on the surface of Factor H domains. This work has revealed that the definition of templates is rather more complex than the work performed by Gregory et al. (1993).

Plan of Work

1. Methods will be developed to extract metal binding sites automatically from the PDB. As new structures become available, they will be analyzed to identify metal binding sites together with their liganding residues and the structural environment of those residues. All data will be fed into a relational database which will be made available over the web. [3 months]

2. An improved template description will be derived. The assumption in the work of Gregory et al (1993) was that distances between pairs of C α (or C β) atoms of residues which form the binding site followed a normal distribution – i.e. relative backbone orientation was conserved. Our recent work with a much larger dataset has shown this is not the case and different backbone

orientations can lead to suitable sites. Thus templates need to account for this. For each common metal type, we will examine whether templates form clusters or whether there is a true continuum. Templates, or more flexible propensity-based scoring methods, will be derived. [4 months]

3. Using machine learning methods (e.g. neural networks) these templates will be combined with the hydrophobicity contrast function and sequence spacing (Vallee & Auld, 1989) to assess likelihood of a site being a metal binding site. Results will be assessed against data in the PDB. [6 months]

4. The method will be extended to finding inter-chain sites where the structure of the chains is only available individually. The surface of one chain will be scanned to find partial templates (2 or 3 of the 4 coordinating residues). The second chain will then be fitted to the empty sites in the template. Since only 1 or 2 anchor points within the template will be available, it will then be necessary to rotate the structure of the second chain to find an optimum position which avoids clashes and optimizes interactions using a simple docking-type potential. Results will be assessed. [12 months]

5. The methods will be applied to the Factor H problem (and any other appropriate proteins available in the Perkins lab) and predictions will be tested experimentally using appropriate biophysical techniques in collaboration with Steve Perkins. [8 months]

Interdisciplinary Aspects

This is a Bioinformatics project, but will be a continuous collaboration with the Perkins lab who are interested in the results of this work. The project will involve 8 months spent in the Perkins lab to test the results of the predictions.

References

- Auld, DS (2001a) *Biometals* 14: 271-313
- Auld, DS (2001b) In: Bertini, I, Sigel, A, Sigel, H, eds. *Handbook on Metalloproteins*. New York: M. Dekker; pp. 881-969.
- Gregory DS, Martin AC, Cheetham JC, Rees AR. (1993) *Protein Eng.* 6:29-35.
- Sodhi JS, Bryson K, McGuffin LJ, Ward JJ, Wernisch L, Jones DT. (2004) *JMB.* 342:307-320
- Vallee BL, Auld DS. (1989) *FEBS Lett.* 257:138-140.
- Vallee BL, Auld DS. (1990) *Biochemistry* 29:5647-5659.
- Williams, RJP. (1985) *Eur. J. Biochem* 150:231-248
- Yamashita MM, Wesson L, Eisenman G, Eisenberg D. (1990) *PNAS USA.* 87:5648-5652.