BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 12 2012, pages 1658–1660
doi:10.1093/bioinformatics/bts246

Structural bioinformatics

Advance Access publication May 3, 2012

FindGeo: a tool for determining metal coordination geometry

Claudia Andreini[1,2,*], Gabriele Cavallaro[1] and Serena Lorenzini[1]

[1]Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy.
[2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Associate Editor: Anna Tramontano

ABSTRACT

Summary: Metals are essential for the structure and function of many proteins and nucleic acids. The geometrical arrangement of the atoms that coordinate a metal in a biological macromolecule is an important determinant of the specificity and role of that metal. At present, however, this information can be retrieved only from the literature, which sometimes contains an improper or incorrect description of the geometry, and often lacks it altogether. Thus, we developed FindGeo to quickly and easily determine the coordination geometry of selected, or all, metals in a given structure. FindGeo works by superimposing the metal-coordinating atoms in the input structure to a library of templates with alternative ideal geometries, which are ranked by RMSD to identify the best geometry assignment.
Availability: FindGeo is freely available as a web service and as a stand-alone program at http://metalweb.cerm.unifi.it/tools/findgeo/.
Contact: andreini@cerm.unifi.it
Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on 6 February, 2012; revised on 27 March, 2012; accepted on 21 April, 2012

1 INTRODUCTION

It is well established that some metals are essential for living organisms (Bertini *et al*., 2006). A major reason for this is that a considerable fraction of proteins are metalloproteins (Bertini *et al*., 2001). Metals can play diverse roles in metalloproteins, in part, because proteins are able to form metal-binding sites that modulate the properties of metals so as to achieve specific functions (Andreini *et al.*, 2009, 2011; Holm *et al.*, 1996). The geometric structure is recognized to be crucial for the specificity and activity of metal sites (Lu *et al.*, 2009; Waldron *et al.*, 2009). Metals are also closely involved in nucleic acid chemistry. Besides being essential to stabilize the phosphate-sugar backbone of DNA and RNA, metals serve crucial functions, e.g. in RNA folding and ribozyme catalysis (Muller, 2010). Nucleic acids can also achieve a remarkable degree of metal selectivity, which depends, among other factors, on coordination geometry (Freisinger and Sigel, 2007). In this scenario, it is perhaps surprising that there is no tool available to determine the coordination geometry of metals in biological macromolecules (or in small complexes) with known structure. We thus present here FindGeo, a tool for this purpose which is freely available both as a web service and as a stand-alone program. At present, information on metal coordination geometry can be retrieved only by exploring the primary literature, which however sometimes contains an improper or incorrect description of the geometry, and often lacks it altogether. Therefore, the use of FindGeo will benefit scientists by: (1) minimizing errors in the assignment of geometries, (2) promoting a uniform terminology and classification of geometries and thus (3) providing a reliable basis for structure–function relationship studies where coordination geometry is a relevant parameter.

*To whom correspondence should be addressed.

2 DESCRIPTION OF THE PROGRAM

FindGeo is written partly in Fortran 77 and partly in Python. A sample workflow illustrating the use of the web version is shown in Figure 1. FindGeo takes as input a PDB file (either found on a local disk or downloaded from the PDB (Berman *et al.*, 2000)), which is searched for metals (any or selected by the user). For each metal, coordinating atoms are identified as those closer than a specified threshold distance (default 2.8 Å) to the metal. Specific elements (default: C and H) can be excluded from the search for coordinating atoms. Each metal site is then compared against a library of structural templates with ideal geometries (Supplementary Table S1 and Figure S1). Depending on the RMSD values obtained after superposition, the various possible geometries are tagged as regular, distorted or irregular. The regular or distorted geometry with the lowest RMSD is taken as the best estimate of the metal coordination geometry. When all the possible geometries are tagged as irregular, the geometry is not assigned. The RMSD-based criteria for tagging geometries are detailed in Supplementary Material Appendix A. On output, FindGeo produces a summary text file and a series of PDB files containing the metal site superimposed to each tested geometry. The structural templates in the library of FindGeo cover the most common geometries for coordination numbers 2–9, as well as geometries that are derived from these by leaving one of the coordinating positions empty. These latter geometries can account for cases where one of the metal ligands has been overlooked (e.g. a missing water molecule) and is not present in the native structure. Prior to superposition, the original coordinates of the metal site are modified so as to set all the metal-coordinating atom distances to 3 Å, as it is in the ideal structural templates. To perform superposition, FindGeo uses a method based on the quaternion parameterization of rotation in the form developed by Kearsley (1989) and implemented by Rupp (http://www.ruppweb.org/xray/comp/superpos.htm). In order for Kearsley's method to be applied, it is necessary that the atoms to be superimposed are specified in advance. In FindGeo, this requirement is overcome by iteratively applying the algorithm for all the possible atom–atom pairings, and selecting the pairing for which the lowest RMSD is obtained.

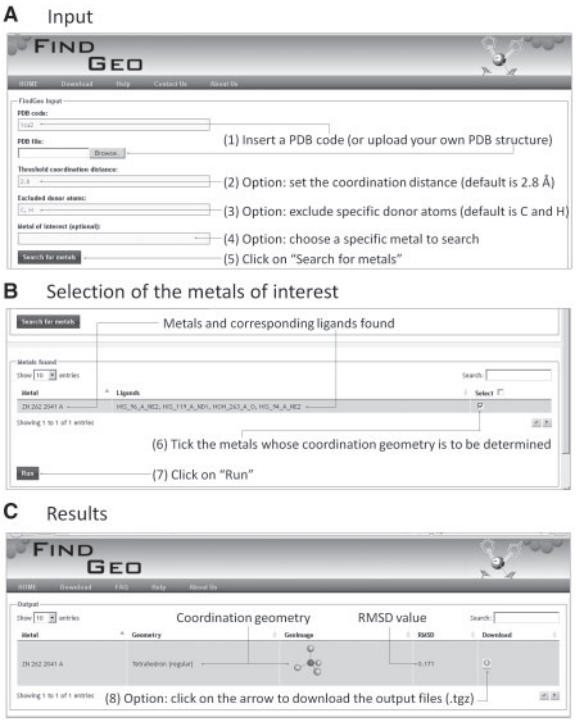

Fig. 1. Sample workflow illustrating the use of FindGeo. In this example, the geometry of a zinc ion in PDB structure 1ca2 is determined by a simple sequence of actions, only four of which (i.e. 1, 5, 6 and 7) are required

3 RESULTS AND DISCUSSION

The performance of FindGeo was evaluated on three different data sets. The first set consisted of 10,800 metal sites (300 for each of the 36 geometries included in the library of FindGeo) that were artificially generated by introducing random distortions in the structural templates with ideal geometries. The second set consisted of 14,342 high-resolution X-ray structures of metal complexes taken from the CSD (Allen, 2002). The third set consisted of 136 metal sites found in protein structures whose geometries were previously analyzed and are available in the literature (Rulisek and Vondrasek, 1998). The construction of these data sets and the results obtained using them as input for FindGeo are detailed in Supplementary Material Appendix B. The results validated the effectiveness and robustness of the approach implemented in FindGeo, which could be successfully applied to determine both the well-defined geometries that are typically observed in small molecules and the less regular geometries found in metalloproteins. In particular, the 26 sites of metalloproteins which were assigned a geometry different from that reported in (Rulisek and Vondrasek, 1998) represent examples of how FindGeo can be used to revise and improve the information available in the literature. Possibly, the most conspicuous among these examples are the assignment of a square pyramidal geometry (*spy*) to a cobalt site reported to be trigonal bipyramidal (PDB 1iab) and, conversely, the assignment of a trigonal bipyramidal geometry (*tbp*) to a nickel site reported to be square pyramidal (PDB 1vkl, chain B). These cases illustrate the discriminating power of FindGeo in challenging cases where two closely related geometries such as *spy* and *tbp* are possible options, and the assignment found in (Rulisek and Vondrasek, 1998) is wrong. A potential pitfall in the use of FindGeo is in the interpretation of the geometries with an empty coordination position. While it is true that such assignments can be, in many cases, a useful hint to identify a missing ligand, users must not be misled to believe that every site whose geometry is described as having a vacant position lacks a ligand (this is indeed the case for 24 out of 37 sites in the metalloprotein data set). It must be kept in mind that ideal geometries are only convenient descriptors of the configuration of the coordinating atoms around the metal, and the fact that the best descriptor for a site is an ideal geometry with an empty position does not necessarily imply that a ligand was overlooked in that site. Also, the geometrical descriptors used in FindGeo should not be confused with descriptors of structural symmetry. For example, a truly octahedral complex with all bond lengths equal and a tetragonally distorted octahedral complex (e.g. due to Jahn–Teller effect) have O_h and D_{4h} symmetry, respectively, but would both be described by FindGeo as octahedral. Still, the output PDB files can be of help to users that wish to investigate such structural details. Possible large scale applications of FindGeo, for which the stand-alone program will be most suitable, include, e.g. the analysis of all metal sites in the PDB or in nucleic acid databases, such as MINAS (Schnabl *et al.*, 2012), and structure–function studies of correlations between coordination geometry and other parameters (e.g. metal function, oxidation state). Also, FindGeo can be used in protein function prediction in conjunction with predictors of metal sites such as CHED (Babor *et al.*, 2008). Finally, FindGeo can be of help in many other fields where the knowledge of coordination geometry is relevant, including, e.g. magnetochemistry and materials science.

ACKNOWLEDGEMENT

We gratefully acknowledge Enrico Morelli for the development of the web version of FindGeo.

Funding: This work was supported by MIUR (Ministero Italiano dell'Università e della Ricerca) through the FIRB project RBFR08WGXT.

Conflict of Interest: none declared.

REFERENCES

Allen,F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B*, **58**, 380–388.

Andreini,C. *et al*. (2009) Structural analysis of metal sites in proteins: non-heme iron sites as a case study. *J. Mol. Biol.*, **388**, 356–380.

Andreini,C. *et al*. (2011) Minimal functional sites allow a classification of zinc sites in proteins. *PLoS One*, **6**, e26325.

Babor,M. *et al*. (2008) Prediction of transition metal-binding sites from apo protein structures. *Proteins*, **70**, 208–217.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Bertini,I. *et al*. (2001) *Handbook on Metalloproteins*. 1st edn. Marcel Dekker, New York. p. 1.

Bertini,I. *et al*. (2006) *Biological Inorganic Chemistry*. University Science Books, Sausalito, California.

Freisinger,E. and Sigel,R.K.O. (2007) From nucleotides to ribozymes: a comparison of their metal ion binding properties. *Coord. Chem. Rev.*, **251**, 1834–1851.

Holm,R.H. *et al*. (1996) Structural and functional aspects of metal sites in biology. *Chem. Rev.*, **96**, 2239–2314.

Kearsley,S.K. (1989) On the orthogonal transformation used for structural comparisons. *Acta Cryst.*, **A45**, 208–210.

Lu,Y. *et al*. (2009) Design of functional metalloproteins. *Nature*, **460**, 855–862.
Muller,J. (2010) Functional metal ions in nucleic acids. *Metallomics*, **2**, 318–327.
Rulisek,L. and Vondrasek,J. (1998) Coordination geometries of selected transition metal ions (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}) in metalloproteins. *J. Inorg. Biochem.*, **71**, 115–127.
Schnabl,J. *et al*. (2012) MINAS: a database of Metal Ions in Nucleic Acids. *Nucleic Acids Res.*, **40**, D434–D438.
Waldron,K.J. *et al*. (2009) Metalloproteins and metal sensing. *Nature*, **460**, 823–830.