Study of metal coordination sphere parameters derived from the open access Crystallography Open Database

Antanas Vaitkus[a], Andrius Merkys[a] and Saulius Gražulis[a,b]

[a]Vilnius University Institute of Biotechnology, Saulėtekio av. 7, LT-10257 Vilnius, Lithuania; [b]Vilnius University Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225 Vilnius, Lithuania

Abstract

Metal coordination sites in proteins have been a subject of research for several decades now. Numerous works with the interest in specific chemical elements as well as wider scope databases have been published both on-line and off-line. Usually one of the two approaches is taken when compiling the datasets – the results are either calculated directly from the protein structures or they are derived from the structures of small molecules. In most cases the same data processing algorithms can be applied with minimal modifications; however, it is the data source that plays a far greater role than it appears at first glance.

The most popular protein structure data source for this kind of calculations is the Protein Data Bank (PDB) [1] database. Among the benefits of using the PDB is its open access nature that allows any interested parties to check and reproduce the published results. However, the PDB is not drawback-free either; for example, grouping structures by the type of coordinating metal often results in unsatisfactory sample sizes.

The use of small molecule structure data sources presents an almost mirror image of the pros and cons – the number of high resolution structures containing metal coordination spheres is much greater, but all of the data sources used in this type of research so far are proprietary databases and as a result makes the research unreproducible without acquiring the database license.

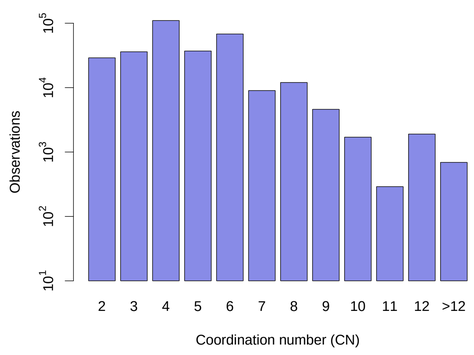
To overcome this limitation the open access Crystallography Open Database (COD) [2] was used as the source to automatically compile a set of metal coordination sphere parameters. The entire COD was scanned for structures containing metal coordination spheres and the selected entries were further processed by calculating coordination parameters such as the coordination number, ligand-metal-ligand angles and metal-ligand distances, as well as comparing the configuration of the coordination sphere with a list of idealised geometry templates.

Currently, the gathered results are only accessible via the MySQL database interface, but an interactive website is being developed. The website will allow the coordination sphere information to be viewed grouped by a combination of various parameters like the chemical type of the coordinating metal, the chemical type of the coordinated ligands and the best fitting idealised geometry. An option to produce a normal distribution mixture model of the examined interatomic distances and angles will also be provided.

Applications

- Detection of unusual features in model geometry;
- Refinement of crystal structures.

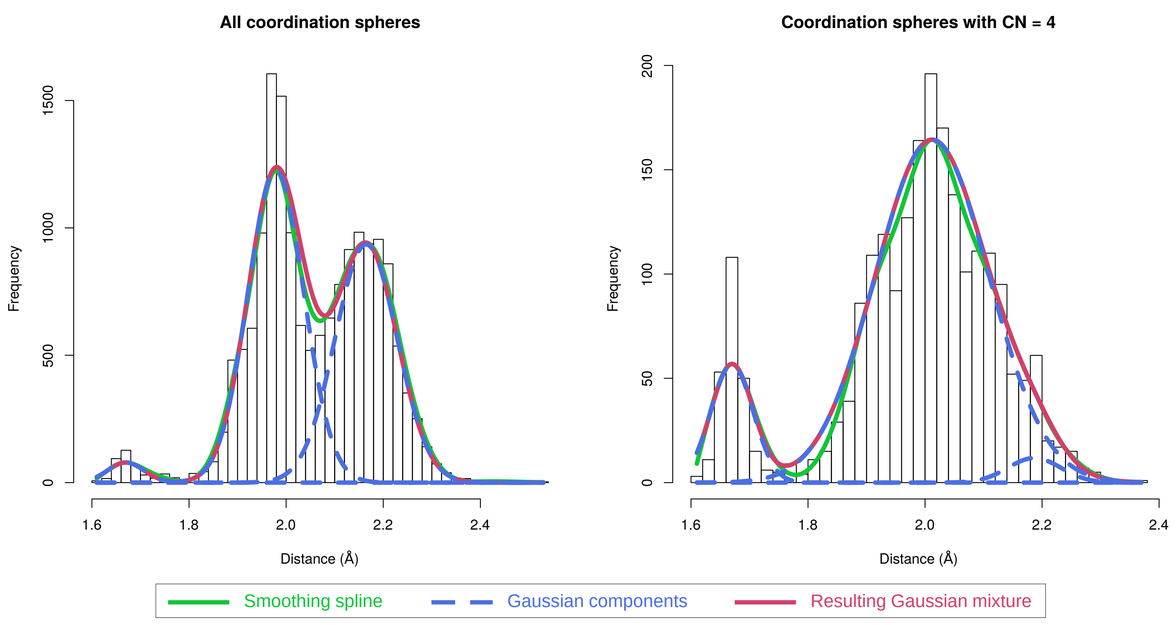
Coordination spheres in COD



- 316 000 coordination spheres were detected in total;
- 78 different coordinating elements were observed;
- Uncommon geometries were spotted among the observations with highest coordination numbers.

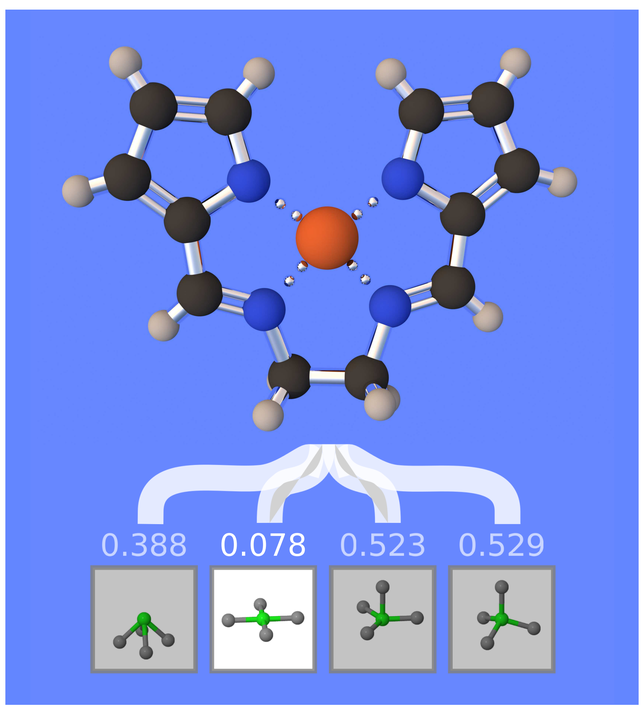
Fe–N distances in COD

Automatically generated normal distribution mixture models can aid in structure validation and identification of the coordinating metal.



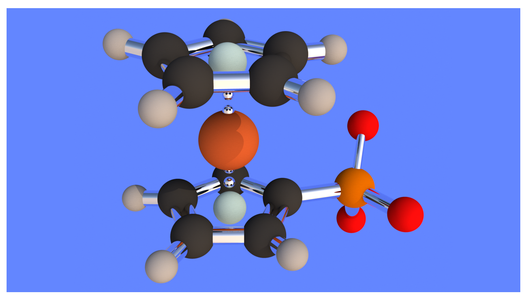
Detection of coordination geometry

- The number of coordinated atoms is estimated;
- The lowest RMSD values are found for each of the idealised geometry models;
- The best fitting model is chosen based on the overall lowest value of RMSD.



The detection of Fe-N(4) square planar coordination sphere. COD ID 2210547.

Hapticity

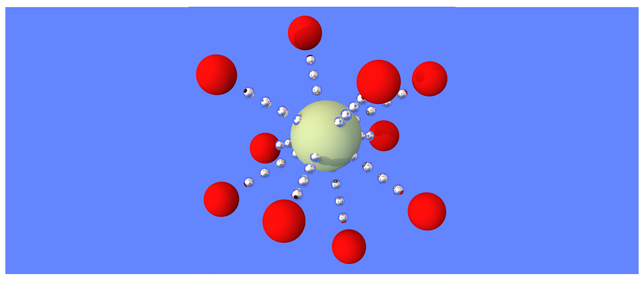


Two dummy atoms are used to represent the coordination sphere in a ferrocene derivative. COD ID 7008471.

- Several most common complex ligands participating in hapticity were succesfully identified;
- These ligands were represented by a calculated dummy atom when fitting idealized geometries;
- Over 7400 such coordination spheres were detected in total.

Causes of extremely high neighbour counts

- Interesting uncommon coordination spheres;
- Hapticity mistaken for denticity;
- Guest molecules.



Pr-O(10) coordination sphere with all non-coordinated atoms removed. COD ID 4338465.

Conclusions

- The COD is a valuable data source containing a great number and variety of metal coordination spheres;
- The collected coordination sphere information can be accessed via the MySQL interface by issuing the following command:

  ```
  mysql -u cod_reader -h www.crystallography.net metal_geometries
  ```

- A web interface to the metal coordination sphere parameter database will soon be available on the COD website (www.crystallography.net).

Bibliography

[1] Berman et al. The protein data bank. *Nucleic Acids Research*, 28:235–242, 2000.

[2] Gražulis et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research*, 40(D1):D420–D427, Jan 2012.

The authors would like to acknowledge Openmolecules.org for providing the experimental unpublished MoleculeViewer v0.8 used in the generation of photorealistic images of the molecules.

Antanas Vaitkus has no conflict of interest.
Andrius Merkys has no conflict of interest.
Saulius Gražulis is a member of the COD Advisory Board.

On-line version of the poster:
http://j.mp/2aY25Lp

antanas.vaitkus90@gmail.com