Many metal ions are known to be essential for life (Frausto da Silva and Williams 2001). Plants are no exception. For instance, many metal-binding proteins are involved in important biological functions of plants, such as photosynthesis (Rhee et al. 1998), nitrogen fixation (Marschner 1986), and degradation of urea (Polacco and Holland 1977). The most important role of the metal association with proteins is as a cofactor of enzymes. The functional and structural aspects of metal-associated enzymes have been extensively studied in the field of bioinorganic chemistry (Andreini et al. 2008). In addition to the role they have with enzymes, metal ions are used as signal mediators with metal-sensing proteins (O’Halloran 1993; Ukaegbu et al. 2006). Stabilization of protein structures is another important function of metal ions (Christianson 1991). Some metal-binding proteins act as metal chaperones in the transport of metal ions (Rae 1999). Approximately one-third of all structurally characterized proteins contain metal ions (Jernigan et al. 1994). The recent emergence of terms such as metallome and metallomics (Williams 2001) is indicative of the increasing efforts devoted toward the study of metalloproteins.

With the rapid progress in structural genomics, an increasing amount of structural information about metal-binding proteins has been accumulated. However, we believe that no organized compilation of such information is publicly available. Therefore, we initiated the MetalMine project. As confirmation of our research, the need for an up-to-date database for metalloproteins has been noted (e.g., Andreini et al. 2009b).

Our database is not the first one proposed. Researchers at Scripps Research Institute created a similar database called the Metalloprotein Database and Browser (MDB) (Castagnetto et al. 2002). This database allows an interactive search based on structural information. Even though the MDB includes all the metal-binding sites in PDB as of 2003, it lacks classification of the metal-binding sites according to biological function. A database called PROMISE (Degtyarenko et al. 1999) provides structural, functional, and bibliographic information for an organized classification of metalloproteins; however, this database has not been updated since 1999, and the number of entries is extremely limited. For example, the number of mononuclear iron binding sites listed in PROMISE is 13, whereas MetalMine has 64 entries. Metal-MACiE is another database that provides information about the catalytic mechanism of metalloenzymes (Andreini et al. 2009a). Metal-MACiE is specific to the metal-binding sites with enzymatic activity, and contained 136 entries as of October 2009. In addition to these web based resources, some books...
describe the properties of a wide variety of metalloproteins, such as the compilation edited by Messerschmidt et al. (Messerschmidt et al. 2001; Messerschmidt et al. 2003). All of these are valuable resources for the community of biological inorganic chemistry. Nevertheless, we believe that most researchers would find MetalMine convenient as an up-to-date database of metalloproteins. For example, a comparison of homologous metal-binding sites would provide insights into the functional and structural evolution of metalloproteins (Nakamura et al. 2005); MetalMine can provide many sources of information for this type of analysis.

Data collection and classification of metal-binding sites

To create MetalMine, the binding sites of metal ions in protein structures in PDB were searched automatically by using a computer program. In this automated search, a metal-binding site is defined as a cluster of metal ions coordinated by several protein amino acid residues, with heteromolecules such as cofactors, substrates, and water. A cluster of metal ions is defined as a single linkage cluster with a distance threshold of 7.0 Å, and coordination to a metal ion is defined as the interatomic distance between a metal ion and a heteroatom (nitrogen, oxygen, or sulfur) closer than 2.9 Å (3.2 Å in the case of sulfur).

Metal-binding sites were then classified according to the structural domain to which the metal-coordinating residues belonged. We used the Structural Classification of Proteins (SCOP) for definition of the structural domains (Murzin et al. 1995). We define homologous metal-binding sites at the corresponding position of the same structural fold family/superfamily as a metal-binding site family/superfamily. For example, the blue-copper sites of azurin and plastocyanin belong to the same metal-binding site family, and the blue-copper site of laccase (EC.1.10.3.2), one of the multi-copper oxidases, belongs to the same metal-binding site superfamily with them. Some proteins contain more than one metal-binding sites at different positions in a SCOP fold region. These sites are distinguished as different metal-binding sites. For example, a laccase has two separate metal-binding sites: a blue-copper site and an inter-domain site. Occasionally, a metal-binding site may consist of amino acid residues from two different structural folds. This implies that the metal-binding site is located at the boundary of two domains in a single chain, or at the boundary of two protein subunits. In such a case, the SCOP IDs of both fold families are listed.

Finally, each group of metal-binding sites was manually examined. During this manual curation process, artificial metal-binding sites were excluded. Although we prepared a GUI program to facilitate manual curation, this is a very tedious and time-consuming process. One of the difficulties is that it is occasionally difficult to determine whether a metal ion coordinated in a crystal structure is actually a native ion used in physiological situations, even if related literature reports enzymatic activity. In MetalMine, entries with this type of ambiguity are listed with a brief note indicating the ambiguity. Nevertheless, we believe that MetalMine contains the largest number of reliable entries among publicly available databases of metalloproteins.

Database structure

The web pages are designed to be simple and self-explanatory. Each page contains a top menu and a sidebar. From the top menu, a user can access pages for a description of the database, a BLAST search, and tutorial. The top page includes a text field and a search button to enable MetalMine to be searched using a PDB ID as a query. From the sidebar, a user can browse the information of metal ions contained in MetalMine.

When one of the metal ions listed in the sidebar is selected, a table of metal-binding sites is shown (Figure 1). In the table are columns that list the names used to specify the metal-binding site, SCOP IDs of residues involved in the metal-binding site, the type and number of metal ions and residues, the representative PDB IDs, number of PDB files found for this site, total number of sites found in the entire PDB, links to Wikipedia and NiceZyme entries at ExPASy (Gasteiger et al. 2003) when available, and other remarks. Clicking on the name of a site leads to a list of all instances of the metal-binding site. This list consists of columns that state the PDB IDs, the residues (one-letter amino acid code followed by the residue number and chain ID), and the heteromolecules such as cofactors and ligands. In addition, a Jmol (Herraez 2006) window opens to display the local metal-binding structure when an instance of the metal-binding site is specified by clicking on the PDB ID (Figure 2). On the right-hand side of the Jmol window, more options are available for inspecting the metal-binding sites, such as a button for displaying the second layer of coordinating residues.

Search by amino acid sequence

In MetalMine, an amino acid sequence search can be performed using BLAST (Altschul et al. 1990). A set of the amino acid sequences from the PDB structures included in MetalMine is prepared for the BLAST search. Given an amino acid sequence as a query, BLAST performs a sequence search of the database for a match with the metal-coordinating residues in MetalMine. Currently, only the exact matches (i.e., a match with the same residues at the corresponding
Figure 1. Screenshot of the MetalMine table for the cobalt ion.

Figure 2. Screenshot of the display of a metal-binding site structure using Jmol in MetalMine.

Figure 3. Screenshot of a BLAST search result from MetalMine.
alignment position) are displayed. Poor matches (i.e., the E-value is greater than 0.0001) are distinguished as low-reliability matches. The output of a BLAST search (Figure 3) includes the amino acid sequence used as the query, the hit residue highlighted with magenta or light cyan for a regular or low-reliability match, respectively, followed by a list of matching residues with links to the metal-binding sites contained in MetalMine.

The BLAST search function in MetalMine can be used as a tool to predict metal-binding residues in amino acid sequences. There have been several prediction methods based on structural information, such as the threading model (Sodhi et al. 2004; Goyal et al. 2008) and the force field model (Schymkowitz et al. 2005). An empirical method based on the comparison of holo-apo pairs of known metal-binding sites appears to be one of the most practical approaches currently available because of the quality of the results (Babor et al. 2008). An ambitious de novo approach based on a sequence using a machine-learning method has also been developed (Lippi et al. 2008), although its prediction ability appears to be limited compared with the structure based predictions (Punta et al. 2008). Searches for a metal-binding sequence fingerprint, such as those in PROSITE (Hulo et al. 2007) and PROMISE (Degtyarenko et al. 1999), are based on sequence homology and are similar to the BLAST search in MetalMine. The application of a sequence fingerprint is usually limited to a relatively short sequence region. In contrast, the BLAST search can detect sequence matches for metal-binding residues spread over the entire sequence. Although the homology-based approach may not lead to the discovery of new metal-binding sites, the prediction results are generally more reliable, as in the case of the protein folding prediction (Zhang 2008).

Conclusion and future work

Plants maintain a unique relationship with metal ions. Because they are stationary, plants cannot leave environments with toxic metal ions such as cadmium or arsenic, and cannot seek environments rich with the necessary metal ions. This has led plants to develop a number of ways to survive with or without different kinds of metal ions. Bioremediation utilizes such relationship between metal ions and plants for the removal of soil contaminants (Salt et al. 1995; Clemens et al. 2002). One goal of this project is to contribute to the development of bioremediation through a functional analysis of metal-binding proteins.

As of October 2009, MetalMine contained 412 entries of metal-binding sites (14, 57, 76, 25, 14, 1, 4, 88 and 133 entries for cobalt, copper, manganese, molybdenum, nickel, vanadium, tungsten, non-heme iron and iron-sulfur clusters, respectively), and 7,917 instances of binding site structures from the PDB database. Information about metal-binding sites for heme iron and zinc is currently being collated to complete the coverage of major essential transition metal ions. Currently, a new metal-binding site entry can only be added manually. However, we are developing programs to make the update process more feasible. As a result, future additions of instances to the existing metal-binding sites, based on the appearance of new PDB entries, will be automated. In addition, because SCOP annotation is often not available for new PDB entries, we are developing tools to handle such entries.

Performance and availability

To minimize the execution time, a BLAST search is performed for a set of sequences that include at least one metal-binding site. A metal-binding site is visualized using Jmol, which requires a web browser that supports Java applets.

The MetalMine database is available at http://metalmine.naist.jp. We encourage users to send suggestions and comments to kensuke-nm@is.naist.jp.

Acknowledgements

The authors thank Professor Takeshi Kawabata at Nara Institute of Science and Technology for his numerous useful comments and suggestions. The BLAST search capability was conceived during the course of discussions with him. This work was supported by a KAKENHI (19500255) Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (JSPS).

References


