

Emerging Strategies in Metalloproteomics

Peter-Leon Hagedoorn

Abstract

The impact of changing metal ion concentrations on infection with a pathogenic microorganism has been established in several cases. To understand the molecular basis of the interplay between metals in the host and the pathogen in infectious diseases, metalloproteomics may prove to be a very powerful approach. Metalloproteomics is the comprehensive analysis of all metal-binding and metal-containing proteins in a biological sample. Strategies in metalloproteomics can be divided into experimental and computer-based approaches.

Samples that have been analyzed using metalloproteomics strategies are cell lysates, subcellular fractions, and recombinantly expressed proteins from structural genomics platforms. Samples can be enriched in metal-binding protein by using immobilized metal ion affinity chromatography (IMAC). Subsequently, metalloproteins have been separated using either 2D electrophoresis or 2D liquid chromatography (LC). Resolved proteins have been further analyzed for metals and protein content. Inductively coupled plasma mass spectrometry (ICP-MS), X-ray absorbance or fluorescence and autoradiography of radioactive metal isotopes have been used to determine the metal content, and different soft ionization mass spectrometry techniques and SDS-PAGE have been used to determine or quantify protein content. Computer-based strategies have been developed to predict metalloproteomes from genomes based on information from literature resources and different genomic and protein structural databases.

Many of the methods presented here are comprehensive in nature. The interplay of computer-based and experimental strategies in metalloproteomics will significantly advance the field in the near future.

Introduction

The term metallome was originally introduced by the Oxford professor R. J. P. Williams to refer to the free metal ion pool in the cell (Williams 2001). Today, the metallome is defined more broadly (even covering nonmetal elementals) and can be viewed as the comprehensive analysis (structure, function, and

interactions) of all metals and metal species in a cell or tissue type (Mounicou et al. 2009). This includes semi-metals (e.g., Se), free metal ions, metal–DNA interactions, and metalloproteins. Metalloproteomics, on the other hand, is the comprehensive analysis of all metal-binding and metal-containing proteins in a biological sample. Thus, it is a subset of the metallome.

Strategies in metalloproteomics can be divided into experimental and computer-based approaches. Experimental approaches usually involve a combination of protein separation and metal analysis techniques (Figure 18.1). Computer-based approaches involve bioinformatics tools to predict, ideally, the whole metalloproteome of an organism or group of organisms based on

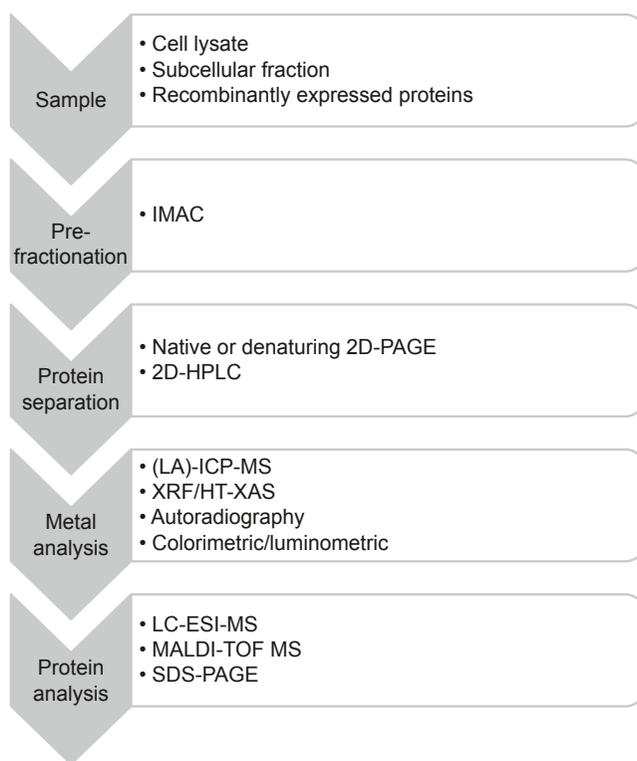


Figure 18.1 Schematic overview of experimental metalloproteomic strategies. IMAC: immobilized metal ion affinity chromatography; 2D-PAGE: two-dimensional polyacrylamide gel electrophoresis; 2D-HPLC: two-dimensional high performance liquid chromatography; (LA)-ICP-MS: (laser ablation) inductively coupled plasma mass spectrometry; XRF: X-ray fluorescence; HT-XAS: high throughput-X-ray absorbance spectroscopy; LC-ESI-MS: liquid chromatography–electrospray ionization–mass spectrometry; MALDI-TOF MS: matrix assisted laser desorption–time of flight mass spectrometry; SDS-PAGE: sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

genomic information, protein structural information, and biochemical information available in databases.

In this chapter, several experimental and computer-based strategies are presented that have emerged in recent years. In addition, relevant biochemical information is presented that has been obtained using these methods, and specific advantages and disadvantages of the different methods are discussed.

Experimental Metalloproteomics

Metal-Binding Protein Fractionation Using Immobilized Metal Ion Affinity Chromatography (IMAC)

IMAC is well known, in the nickel ion bound form, for the isolation of his-tagged proteins. IMAC can be used for other applications as well (Cheung et al. 2012). The method is based on a resin that contains metal ion-chelating groups, such as nitrilotriacetic acid or iminodiacetic acid. Such materials can be charged with different metal ions in a manner that still makes coordination by amino acid residues from metal binding proteins possible. These proteins with high affinity to the chelated metal ions bind and can be eluted using a pH gradient or a competitive ligand, such as EDTA or imidazole. IMAC has been used to enrich metal ion-binding proteins for proteomic investigations. In one instance, Ni-IMAC was used in combination with 2D-PAGE to identify Ni-binding proteins that may be related to metal-specific allergic contact dermatitis (Thierse et al. 2008). Another example is the combination of Ni- and Co-IMAC in combination with mass spectrometry (MS) to identify putative Ni- and Co-binding proteins from *Streptococcus pneumoniae* (Sun et al. 2013).

In a metalloproteomic workflow, IMAC can be used prior to further separation with 2D-PAGE or 2D-LC and subsequent protein identification. The advantage of IMAC is that it is relatively inexpensive and straightforward and can be implemented in high-throughput liquid-handling workflows. Important disadvantages are the occurrence of false positives (e.g., histidine-rich proteins) and false negatives (e.g., metal-containing proteins).

Methods Based on Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

ICP-MS is a powerful metal analysis technique that is used as a standard technique in trace element analysis. The technique is based on the ionization of a sample by introducing it into an ICP, where it vaporizes and is broken down into atoms, of which many are ionized to form singly charged cations. Ions are subsequently transferred via an interface to a mass spectrometer. The interface is necessary to change the ion beam from the atmospheric pressure of the ICP to the high vacuum of the MS. Metals are identified based on their masses and

isotopic distribution. The method is sensitive and has a large dynamic range for many metals and other elements. One inherent problem of ICP-MS is the occurrence of isobaric interferences originating from elements in the sample matrix or the ICP gas, e.g., ^{56}Fe ($m = 55.93494$) and $^{40}\text{Ar}^{16}\text{O}$ ($m = 55.95729$). High-resolution ICP-MS allows better separation of the different metal isotopes and prevents most isobaric interferences. ICP-MS has been used in quantitative proteomics by employing metal-coded affinity tags, which consist of three parts: (a) a metal chelate complex consisting of 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) coordinating a lanthanide, (b) an affinity anchor for purification, and (c) a reactive group for the reaction with amino acids to label proteins (Ahrends et al. 2007). Below, two different strategies are presented for metalloproteomics using ICP-MS.

2D Gel Electrophoresis – Laser Ablation ICP-MS

Laser ablation ICP-MS (LA-ICP-MS) offers the possibility of directly analyzing metalloproteins following 1D or 2D-PAGE. This technique is based on the use of a focused laser beam to ablate part of the sample, which is transferred to the ICP using a continuous argon gas stream; the elemental ions are then analyzed by MS. Becker et al. (2004) used LA-ICP-MS to analyze the yeast mitochondrial proteome for phosphorus, sulfur, copper, zinc, and iron. Samples were separated by Blue Native PAGE in the first dimension and SDS-PAGE in the second. Two gels were produced in parallel: one analyzed for protein phosphorylation sites using MALDI-FTICR-MS; the other analyzed with LA-ICP-MS. This method offers primarily qualitative identification of metalloproteins and can provide element ratio, whereby the S content can be used as a measure of the protein amount.

LA-ICP-MS analysis of the same gel would require ICP-MS measurements of all parts of the gel, even those that do not contain any metals. Furthermore, during the process the gel is destroyed. However, ICP-MS has the advantage that it provides a multi-elemental analysis, and that nonmetal elements, such as sulfur, can be used to estimate the amount of protein.

LA-ICP-MS can be applied to a biological tissue directly and used to detect multiple different metals simultaneously. However, it has the disadvantage of being destructive to the biological sample and has a detection limit in the μM range for copper and zinc.

2D Liquid Chromatography – ICP-MS

A successful metalloproteomic workflow resulted in the finding that protein-folding location is used as a natural strategy to regulate metal ion binding in proteins of the cupin family in the cyanobacterium *Synechocystis* PCC 6803 (Tottey et al. 2008). The workflow was based on protein separation using 2D-LC and subsequent analysis of each fraction with ICP-MS for metals and

SDS-PAGE for proteins. The metal distribution is carefully aligned with the distribution of proteins after SDS-PAGE by using principal component analysis. The advantage of this method is that multiple different metals can be analyzed simultaneously and considerable flexibility in the separation is possible by changing LC techniques. The 2D-LC approach generates a large number of samples that have to be analyzed by ICP-MS and SDS-PAGE.

Cvetkovic et al. (2010) used a combination of 2D-LC, high-throughput tandem mass spectrometry, and ICP-MS to identify the metalloproteome of the thermophilic archaeon *Pyrococcus furiosus*. The organism was found to incorporate more metals in its proteins than anticipated, as Pb-, Mn-, Mo-, U-, and V-containing proteins were found. However, the incorporation of lead and uranium was found to be only substoichiometric; that is, only 0.01 U atom per ferritin (Pf0742) monomer (Cvetkovic et al. 2010). These data, together with a preliminary analysis of the metalloproteomes of *Escherichia coli* and *Sulfolobus solfataricus* by the same approach, led Cvetkovic et al. to conclude that the microbial metalloproteome is largely uncharacterized. The method employed by Cvetkovic et al. is laborious and does not appear to be suitable for the analysis of a large number of different biological samples (e.g., different microbial growth conditions). Multielement detection by ICP-MS is powerful, because it allows for the detection of metals that were not anticipated by the researcher. This method is a good starting point to explore the metalloproteome of a particular (micro-)organism.

Methods Based on Radioactive Metal Isotopes

Radioactive metal isotopes have great potential in metalloproteomics, as they offer superior sensitivity as well as the possibility of imaging metalloproteins and quantifying metal levels. Table 18.1 summarizes suitable radioactive isotopes for metalloproteomics based on their potential for imaging and relatively short half-life.

Table 18.1 Metal radioisotopes with potential for metalloproteomics. The selection criteria are: $t_{1/2}$ between 2 and 100 hours; suitable β^- abundance and energy.

Radionuclide	Half-life time (hours)
^{56}Mn	2.6
^{65}Ni	2.5
^{64}Cu	12.7
^{67}Cu	61.8
^{69}Zn	13.8
^{99}Mo	66.0
^{187}W	23.8

Metal Isotope Native Radio Autography in Gel Electrophoresis (MIRAGE)

MIRAGE is a metalloproteomics technique that involves four steps:

1. Labeling of target proteins with a radioisotope.
2. Separation of intact holoproteins using native isoelectric focusing, followed by Blue Native PAGE in the second dimension.
3. Spot visualization and quantification using autoradiography.
4. Protein identification by tandem mass spectrometry after in-gel trypsin digestion.

Step 1 can be achieved simply by growing a microorganism in a medium containing a radioactive metal isotope. The advantage of MIRAGE is that on a single 2D-PAGE gel, all proteins which contain a particular metal ion can be visualized (Figure 18.2). Furthermore by quantifying the beta emission from the radioisotopes, it is possible to obtain absolute quantities of the metal that is associated to a protein. Obtaining absolute quantities is challenging in conventional proteomic techniques, and the term quantitative proteomics nearly always refers to relative quantities. The main disadvantage of this technique is that one can detect only one type of metal at a time. This is especially interesting for researchers who are interested in a particular metal, rather than all types of metals at the same time. Protein identification by MS/MS typically results in several overlapping proteins for each metal-containing spot. Using the *E. coli* genome annotation, the proteins relevant to the metal under investigation can be selected.

MIRAGE has previously been used to determine the soluble Cu, Fe, and Zn proteome of *E. coli* and the Mo and W proteomes of *P. furiosus*. MIRAGE investigation of the soluble Cu proteome of *E. coli* showed that the multicopper oxidase CueO is the only detectable Cu protein. CueO is a multicopper oxidase responsible for the oxidation of Cu^+ to Cu^{2+} . Scavenging cytoplasmic free copper has been proposed as a possible physiological role of CueO (Sevcenco et al. 2009a). The soluble Fe proteome of *E. coli* was found to be dominated by just three different proteins: 90% of the Fe was associated with superoxide dismutase, ferritin, and bacterioferritin (Sevcenco et al. 2011). The Zn proteome of *E. coli* experiencing Zn stress was dominated by ZnP, a putative Zn storage protein (Sevcenco et al. 2011). MIRAGE investigation of the effect of molybdenum on the W proteome of *P. furiosus* revealed that the organism exhibits a strong preference for W over Mo incorporation in its enzymes, even when the intracellular concentration of molybdenum was higher than tungsten (Sevcenco et al. 2009b; Sevcenco et al. 2010).

The occurrence of numerous overlapping proteins in the spots following native-native 2D-PAGE can be resolved by improving the resolution of the native-native 2D-PAGE; for example, by using a more narrow pH range in the isoelectric focusing step or by employing pre-fractionation using other protein separation techniques. Quantitative data on the protein levels will provide

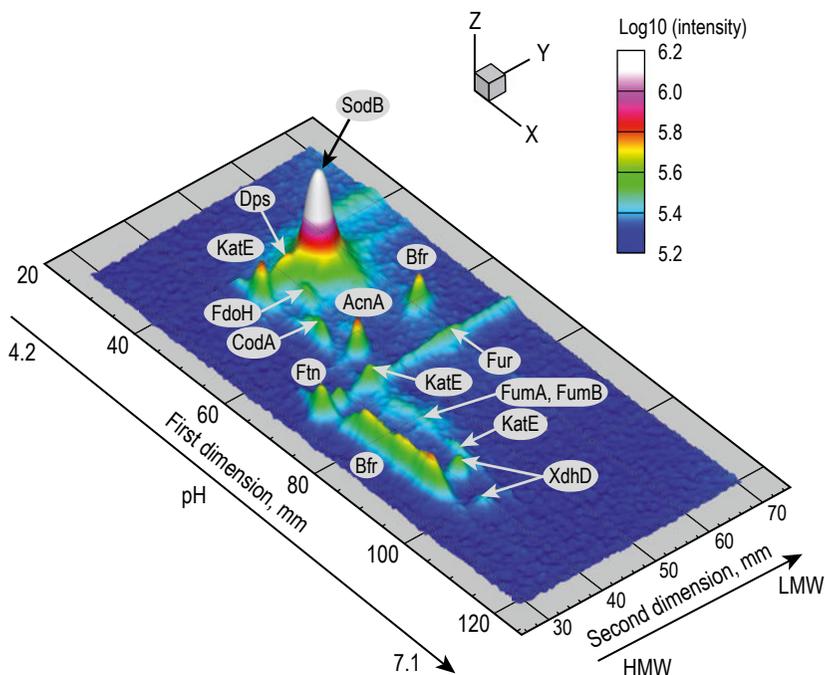


Figure 18.2 3D image of ^{59}Fe -MIRAGE of *E. coli* soluble protein extract (575 μg protein) obtained from cells grown in 6 μM Fe (Sevcenco et al. 2012). AcnA: aconitase A; Bfr: bacterioferritin; CodA: cytosine deaminase; Dps: DNA-binding protein from starved cells; FdoH: formate dehydrogenase-O iron-sulfur cluster binding subunit; Ftn: ferritin; FumA: fumarase A; FumB: fumarase B; Fur: ferric uptake regulator; KatE: hydroperoxidase HPII; SodB: Fe superoxide dismutase; XdhD: xanthine dehydrogenase-like protein Mo and iron-sulfur cluster containing subunit. Reprinted with permission from John Wiley and Sons. Copyright © 1999–2014 John Wiley & Sons, Inc. All rights reserved.

valuable information on the metal/protein stoichiometry. At present, the absolute quantification of proteomes is difficult and laborious, although first studies into this area have recently appeared (Schmidt et al. 2011).

2D-Denaturing, PAGE, Western Blotting, Phosphor Imaging

Radioactive metal isotopes have been reported many times in the literature to identify and isolate metalloproteins. A relatively early approach to identify Zn-binding proteins in a comprehensive manner was published by Katayama et al. (2002). Proteins were separated using denaturing 2D-PAGE and, subsequently, the denatured proteins were transferred to a PVDF membrane using Western blotting. The resulting blot was incubated with a radioactive Zn isotope under

non-denaturing conditions. The rationale behind this approach is that upon refolding of the denatured proteins in the presence of zinc, the metal will be incorporated and the metalloproteins can be identified.

As this method is based on unfolding and refolding, there is a high risk of obtaining false positives as well as false negatives. Many of the Zn-containing proteins identified using MIRAGE were also found by Katayama et al. (2002), indicating that the method is apparently successful. This method does not provide quantitative information on the (natural) distribution of zinc among the proteins in a biological sample.

High-Throughput Colorimetric-Luminometric Metalloprotein Detection

A very different approach has been published by Högbom et al. (2005) and involves an elegant combination of colorimetric and luminometric techniques to determine manganese, iron, cobalt, copper, and nickel or zinc. The procedure is a combination of two luminescence and one colorimetric assays. The protein sample (ca. 10 μ g) was added to urea to denature the protein and release the metal ions. Subsequently luminol, sodium carbonate, and hydrogen peroxide were added, and the first luminescence data were recorded. Metal ions are known to enhance the rate of the reaction of luminol with hydrogen peroxide under alkaline conditions. In the second step, 4-(2-pyridylazo)resorcinol (PAR) was added, and the second luminescence data were recorded. PAR is a metal ion chelator and forms colored complexes with various metal ions. After a two-hour incubation period and centrifugation step to release nitrogen gas bubbles resulting from the luminol reaction, the absorbance was measured at 492 nm. The combination of positive and negative responses on the three different assays allows for the identification of the different metal ions (Figure 18.3).

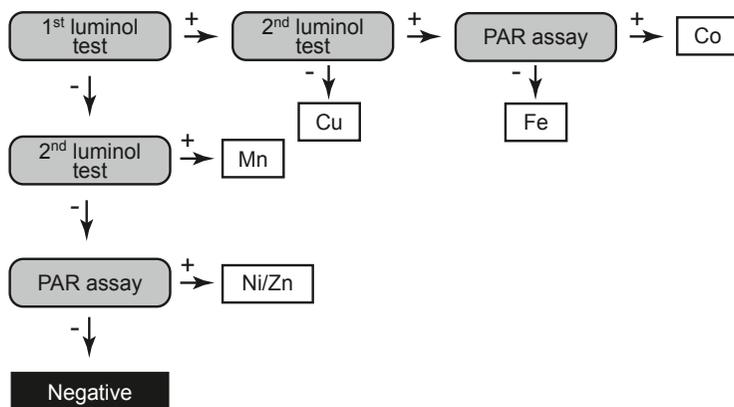


Figure 18.3 Schematic workflow of the luminometric/colorimetric metalloproteomics approach.

This method appeared to be powerful in the qualitative detection of metals in relatively pure protein samples, although it tolerates sub-stoichiometric levels of contaminating metals and can simultaneously measure copper and zinc or nickel. Since the identification of the metal ions in the colorimetric/luminometric method relies on an ingenious combination of positive and negative results in the different assays, mixtures of metal ions will be difficult to analyze. It appears that this approach is most useful in cases where it is known that the sample contains relatively pure proteins and only one major metal is expected. An important advantage of this approach is that it can be implemented in existing high-throughput microplate-based platforms. The method was performed in a 384-well format and took three hours to complete, which, according to Högbom et al. (2005), makes it a true high-throughput method. The sensitivity of the method, however, is orders of magnitude lower than for most other techniques discussed here.

Methods Based on Synchrotron Radiation X-Ray Absorption (XAS) and X-Ray Fluorescence (XRF)

XAS and XRF are based on the property that X-ray photons of a certain energy are able to expel electrons from the 1s electron shell around the nucleus. This hole is subsequently filled by another electron from a higher energy shell under the emission of a fluorescent X-ray photon. The fluorescence signals are highly characteristic of the type of metal. XAS and XRF have been used to structurally characterize metal sites in proteins in techniques such as EXAFS and XANES. SR-XRF has been used to obtain the distribution of certain metals in a single cell (McRae et al. 2009). In recent years, strategies have been developed to use XAS and XRF to perform metalloproteomics (Shi et al. 2005; Shi and Chance 2011). Spatial resolution of Synchrotron X-ray beams has improved to a resolution in the order of 0.1–1 μm with a penetration depth of approximately 1 mm. These metalloproteomic strategies involve performing high-throughput XAS on proteins produced in a structural genomics pipeline and detecting metalloproteins that have been separated using non-denaturing 1D or 2D gel electrophoresis using XRF or XAS.

High-throughput XAS to detect cobalt, nickel, copper, and zinc has been performed on recombinant proteins from *P. furiosus* produced in a structural genomics pipeline (Scott et al. 2005). This structural genomics pipeline involved the cloning, expression, and purification of ca. 2200 gene products from *P. furiosus*. Approximately 3 μl of 0.2–1.0 mM of each protein were required. In a similar approach, 654 proteins (100 μg each) from the New York Structural Genomics Research Consortium were analyzed for Mn, Fe, Co, Ni, Cu, and Zn content (Shi et al. 2005). Over 10% of the proteins were found to contain stoichiometric amounts of one of these metals. Scott et al. (2005) concluded that their method was approximately 95% accurate in predicting the stoichiometric metal content. When analyzing metalloproteins from such

structural genomics pipelines, one must be careful as often his-tags and IMAC are used for high-throughput protein purification, and this may lead to improper metal incorporation or metal loss during purification (Jenney et al. 2005).

A different approach involves the use of XRF to directly scan proteins that have been separated on a non-denaturing electrophoresis gel (Native 2D-PAGE). This method was successfully applied to identify the metal composition of different isoforms of metalloproteins on a native 2D-PAGE (Ortega 2009). In principle, even information of the redox state and the chemical environment of the metal ion can be obtained (Kemner et al. 2005). A similar approach has been used in a metalloproteomic investigation of the response of *E. coli* to Hg^{2+} (Gao et al. 2013). Proteins were separated using a conventional denaturing 2D-PAGE, after which differentially expressed proteins were identified using MS; Hg-containing proteins were found by scanning the gel using XRF. Apparently, Hg binding to these proteins was so tight that they were not disrupted by the extensive exposure of the protein to the denaturing agents (SDS and urea).

XAS and XRF are essentially noninvasive to the protein sample. These techniques, however, are only applicable to isolated pure proteins of high concentration. XAS provides more chemical information on the metal environment than any of the other techniques discussed in this chapter. The main disadvantage of the XRF- and XAS-based techniques is the requirement of a Synchrotron facility. In addition, although XRF is a very sensitive technique and allows the measurement of metals in small sample volumes, even within eukaryotic cells, the concentration of the metal required in the sample is relatively high (Ascone and Strange 2009).

Computer-Based Metalloproteomics

Computer-based methods have already been used in several of the experimental strategies outlined above. Cvetkovic et al. (2010) used the Integrated Resource of Protein Domains and Functional Sites (InterPro database) to link the found metals to the proteins detected by MS. In the MIRAGE experiments, genome annotation was used to select the proteins relevant to the metal under investigation when a mixture of proteins was found to be present in a metal-containing spot (Sevcenco et al. 2011).

The wealth of genomic and protein structural information has not been overlooked by metalloproteomics researchers. The Metal MACiE database¹ provides structural and functional information on the metals involved in enzyme catalysis. Metal MACiE contains only metalloproteins with a known crystal structure and only a selected number of metal ions (Andreini et al. 2008). One typical example of this limitation is that there are no tungsten-containing

¹ http://www.ebi.ac.uk/thornton-srv/databases/Metal_MACiE/home.html (accessed Dec. 2, 2014)

enzymes in the database, whereas in the Protein Data Bank there are at least four known tungsten enzyme crystal structures. More recently the MetalPDB database² has been constructed (Andreini et al. 2012). This database provides structural information about metal sites in proteins and is searchable by metal, by PDB entry, by enzyme or protein name, as well as by other identifiers and keywords. The information that can be retrieved from this database is predominantly structural. The user has to be aware that metal identities and coordination in crystal structures from proteins may contain errors: mismetallation may have occurred during protein production or crystallization, or the metal identity may not have been properly validated by structural biologists.

Several strategies to extract metalloproteomic information from such databases have been developed. One particularly comprehensive approach involves the following steps for one particular metal:

1. Interrogate literature and database resources to identify metalloproteins, transporters, cofactor biosynthesis proteins, and others (e.g., chaperones or regulatory proteins) and compile a set of proteins linked to one metal.
2. Perform BLAST searches of this set of proteins against sequenced genomes, identify homologs, and identify metal-using organisms and metal-containing proteins.
3. Analyze the data to construct metalloproteomes for a particular organism or group of organisms or identify the metalloproteome for a particular metal ion.

The search includes transporters, cofactor biosynthesis proteins, and regulatory proteins, most of which will not be metal containing. By using the strategy outlined above Zhang and Gladyshev (2009) have been able to investigate the metal usage by many different organisms.

Computer-based methods are interesting as they offer predictions that can be challenged experimentally. Furthermore, these methods offer an overview of metal usage in biology, based on our current knowledge. Their main drawback is that they are limited by our current knowledge and understanding of biological metal metabolisms. The strategies still require significant manual involvement and databases need to be curated. Unfortunately, two important metalloproteomics resources—Prosthetic centers and metal ions in protein active sites (PROMISE) and the Metalloprotein Database and Browser (MDB)—have recently been discontinued and no longer exist (Degtyarenko et al. 1998; Castagnetto et al. 2002). On the other hand, comprehensive genome, pathway, and protein databases like PROSITE and BioCyc are searchable for particular metals or metal-related biological functions (Caspi et al. 2010; Sigrist et al. 2012).

² <http://metalweb.cerm.unifi.it> (accessed Dec. 2, 2014)

Concluding Remarks

In a recent perspective article by Lothian et al. (2013), an integrative metalloproteomics strategy was envisioned, combining traditional proteomics and ICP-MS-based metalloproteomics on biological samples that have been enriched in particular stable isotopes for certain metals. In such a workflow, changes in protein level and metal content can be obtained, although no experimental data have yet been presented. Many of the methods presented in this chapter are comprehensive in nature. Some strategies are especially suitable for the systematic analysis of relatively pure recombinant proteins as produced in a structural genomics pipeline (e.g., SR-XRF and the high-throughput colorimetric/luminometric method). Other methods are suited to find and quantify metalloproteins that have been separated on a 2D-PAGE gel or by 2D-LC, such as methods based on LA-ICP-MS and MIRAGE. The first method allows for the detection of many metals simultaneously and requires analysis of large numbers of sample, whereas the second can only measure one (radioactive) metal isotope and provides a direct image of metal distribution and quantification of the metal levels in the different proteins. For the intracellular localization of metals in a single eukaryotic cell, SR-XRF appears to be the only amenable technology. For the multi-elemental imaging of a piece of tissue, LA-ICP-MS may well be the most suitable technique. For the identification and quantification of all proteins that contain a particular metal ion in any biological sample, MIRAGE is perhaps the optimal approach. Finally, in the near future, we can expect the interplay of computer-based and experimental strategies in metalloproteomics to advance the field significantly.