

Chapter 16

Comparative Genomics Analysis of the Metallomes

Vadim N. Gladyshev and Yan Zhang

Contents

| | |
|---|-----|
| ABSTRACT..... | 530 |
| 1 INTRODUCTION..... | 530 |
| 2 GENERAL APPROACHES TO COMPARATIVE GENOMICS OF METAL UTILIZATION..... | 532 |
| 2.1 Identification of Metal Utilization Traits and Metalloprotein Families..... | 532 |
| 2.2 Identification of Metal-Related Orthologs in Genomic Databases..... | 534 |
| 2.3 Comparative Analyses of Metal Utilization and Interaction..... | 534 |
| 3 MOLYBDENUM..... | 535 |
| 3.1 Molybdenum Transport and Molybdenum Cofactor Biosynthesis..... | 535 |
| 3.2 Molybdoenzymes..... | 538 |
| 3.3 Comparative Genomics of Molybdenum Utilization..... | 540 |
| 4 COPPER..... | 543 |
| 4.1 Overview of Copper Trafficking and Homeostasis..... | 543 |
| 4.2 Cuproproteins..... | 547 |
| 4.3 Comparative Genomics of Copper Utilization..... | 551 |
| 5 NICKEL AND COBALT..... | 553 |
| 5.1 Nickel and Cobalt Uptake..... | 553 |
| 5.2 Nickel-Dependent Proteins..... | 556 |

V.N. Gladyshev (✉)

Division of Genetics, Department of Medicine,
Brigham and Women's Hospital and Harvard Medical School,
New Research Building, 77 Av. Pasteur, Boston, MA 02115, USA
e-mail: vgladyshev@rics.bwh.harvard.edu

Y. Zhang

Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences,
Chinese Academy of Sciences, University of Chinese Academy of Sciences,
294 Tai Yuan Road, Shanghai 200031, China
e-mail: yanzhang01@sibs.ac.cn

| | | |
|-------|---|-----|
| 5.3 | Cobalt-Dependent Proteins | 557 |
| 5.3.1 | Vitamin B ₁₂ Uptake and Biosynthesis | 557 |
| 5.3.2 | Vitamin B ₁₂ -Dependent Proteins | 558 |
| 5.3.3 | Non-Corrin Cobalt-Binding Proteins | 562 |
| 5.4 | Comparative Genomics of Nickel, Cobalt, and Vitamin B ₁₂ Utilization | 562 |
| 6 | COMPARATIVE GENOMICS OF OTHER METALS | 565 |
| 6.1 | Comparative Genomics of Zinc-Dependent Metalloproteomes | 565 |
| 6.2 | Advances in Comparative Genomics of Other Metals | 567 |
| 7 | COMPARATIVE GENOMICS OF METAL DEPENDENCY IN BIOLOGY | 568 |
| 8 | CONCLUDING REMARKS | 572 |
| | ABBREVIATIONS AND DEFINITIONS | 572 |
| | ACKNOWLEDGMENTS | 574 |
| | REFERENCES | 574 |

Abstract Biological trace metals are needed in small quantities, but used by all living organisms. They are employed in key cellular functions in a variety of biological processes, resulting in the various degree of dependence of organisms on metals. Most effort in the field has been placed on experimental studies of metal utilization pathways and metal-dependent proteins. On the other hand, systemic level analyses of metalloproteomes (or metallomes) have been limited for most metals. In this chapter, we focus on the recent advances in comparative genomics, which provides many insights into evolution and function of metal utilization. These studies suggested that iron and zinc are widely used in biology (presumably by all organisms), whereas some other metals such as copper, molybdenum, nickel, and cobalt, show scattered occurrence in various groups of organisms. For these metals, most user proteins are well characterized and their dependence on a specific element is evolutionarily conserved. We also discuss evolutionary dynamics of the dependence of user proteins on different metals. Overall, comparative genomics analysis of metallomes provides a foundation for the systemic level understanding of metal utilization as well as for investigating the general features, functions, and evolutionary dynamics of metal use in the three domains of life.

Keywords bioinformatics • comparative genomics • metal • metallome • metalloproteome

Please cite as: *Met. Ions Life Sci.* 12 (2013) 529–580

1 Introduction

Metallomes refer to the complete set of metalloproteins, metalloenzymes and other metal-containing biomolecules that organisms utilize [1]. The study of metallomes, often referred to as metallomics, is a new scientific field that includes high-throughput studies on metals and integrates the research on these elements to obtain systems level understanding of their use in biology. Some biometals, such as sodium, potassium, and calcium, are needed in large amounts, but the majority of these elements

belong to the group of trace elements (also called micronutrients). Although these metals are required in small quantities, they still may be essential for optimal growth, development and metabolic functions of living organisms [2,3]. These trace metals include zinc, iron, copper, manganese, molybdenum, tungsten, nickel, cobalt, chromium, vanadium, and possibly several other metals. They function in widely different ways. Some are essential components of enzymes where they directly interact with substrates and often facilitate their conversion to products; some donate or accept electrons in reactions of reduction and oxidation; some structurally stabilize biological molecules; and some control biological processes by facilitating the binding of molecules to receptor sites on cell membranes [4,5]. Their deficiency or mutations in genes that handle these metals often result in abnormal development, metabolic abnormalities, or even death.

Among trace metals, Zn and Fe appear to be used by all or almost all organisms [6–8]. The utilization of other trace metals, including Cu, Mn, Mo, Ni, and Co, is more scattered. Since all these metals play important roles in cells, the ability of the cell to tightly control their homeostasis is very important; the key processes relate to uptake, storage, excretion, and utilization of metals [9]. High-affinity import systems have been characterized for most biometals in both prokaryotes and eukaryotes. In bacteria, the ATP-binding cassette (ABC) transporters are the most frequently used uptake systems, e.g., ZnuABC for Zn, MntABC for Mn, ModABC for Mo, and NikABCDE for Ni [10–13]. Non-ABC transporters were also reported, e.g., ZupT and ZIP for Zn, MntH for Mn and Fe, CtaA and Ctr1 for Cu, and NiCoT for Ni and Co [14–19]. In addition, some metal ions could be transported via unspecific cation channels, although the efficiency of such processes may be low [20,21]. Excessive uptake of certain metals through either specific or unspecific pathways may result in metal overload and toxicity. Thus, storage of metals in inactive sites or forms and excretion/export systems represent essential mechanisms that prevent accumulation of inappropriate amounts of reactive trace metals in the cell (e.g., metallothioneins for heavy metal binding/detoxification, CopA/ATP7A for Cu export and ZnT for Zn export [22–25]). Besides detoxification, release of a metal ion from a storage site may be important under conditions of metal deficiency. Moreover, the use of some metals may be dependent on other metals. For example, excessive Zn can induce signs of Cu deficiency [26]. It is clear that homeostasis of metals within the cell should be carefully maintained by mechanisms regulating their uptake, storage, and removal in order to provide sufficient levels while preventing accumulation to toxic levels.

Most metals are directly incorporated into their cognate sites in proteins, but some have to become part of prosthetic groups, cofactors, or complexes prior to insertion of these moieties into target proteins. For example, Mo and Co are the main functionalities in molybdopterin (or Mo cofactor, Moco) and cobalamin (vitamin B₁₂), respectively [27,28]. Another interesting feature is that the number of metalloprotein families varies greatly depending on which metal is used. For instance, over 300 protein families require Zn for proper function [29], whereas less than 10 protein families are known to be dependent on Ni [30].

In the past decade, dramatic advances in genomics have provided an opportunity to investigate the occurrence and evolutionary dynamics of pathways that an organism utilizes, including metal utilization. Computational and comparative analyses

of protein sequences and structures on a genomic scale revealed a significant number of proteins that may bind metals. Thus, identification of all or almost all metalloproteins in genomic databases can greatly assist in our understanding of utilization and function of metals in biology. However, due to the lack of reliable approaches, it is currently not possible to identify complete sets of metalloproteins in organisms. In recent years, several comparative and functional genomic analyses have been carried out for certain trace metals, including Zn, Ni, Co, Cu, and Mo [31–39]. These studies improved our understanding of current use and evolutionary trends in the utilization of these metals in organisms in the three domains of life. In this chapter, we focus on the use of several trace metals from the perspective of comparative genomics. Studies on their utilization may provide important information with regard to fundamental issues of function of these metals.

2 General Approaches to Comparative Genomics of Metal Utilization

Comparative genomics is an exciting new field of biological research, which scrutinizes genome sequences and structures of multiple organisms to identify similarities and differences [40–42]. This information provides a powerful tool for studying evolutionary changes, helping to identify genes that are conserved among species, as well as genes that give each organism its unique characteristics. It also helps scientists to better understand the pathways and other biological processes, including trace metals, in currently living organisms. Using methods of comparative genomics, it is now possible to compile metal-dependent pathways and proteins that an organism uses.

Unfortunately, a precise approach has not been developed for the identification of metalloproteins, partially because of overlapping features for different metals or the uncertainty of metal-binding residues. However, studies on sequence and structural properties of known metalloproteins and their metal-binding ligands resulted in the development of a large number of metal-binding motifs, which can help identify additional metal-binding proteins. Furthermore, searches for metal utilization traits can be assisted with the analyses of factors involved in metal transport or biosynthesis of metal-containing cofactors. The procedure of comparative genomics of biometals may briefly include three steps (Figure 1).

2.1 Identification of Metal Utilization Traits and Metalloprotein Families

The metal utilization trait refers to the occurrence of at least one protein that utilizes this metal. Thus, the first step, which also offers the most important evidence, should be to identify all known metalloproteins for corresponding metals. Based on sequence and structural signatures of known metalloprotein families, several

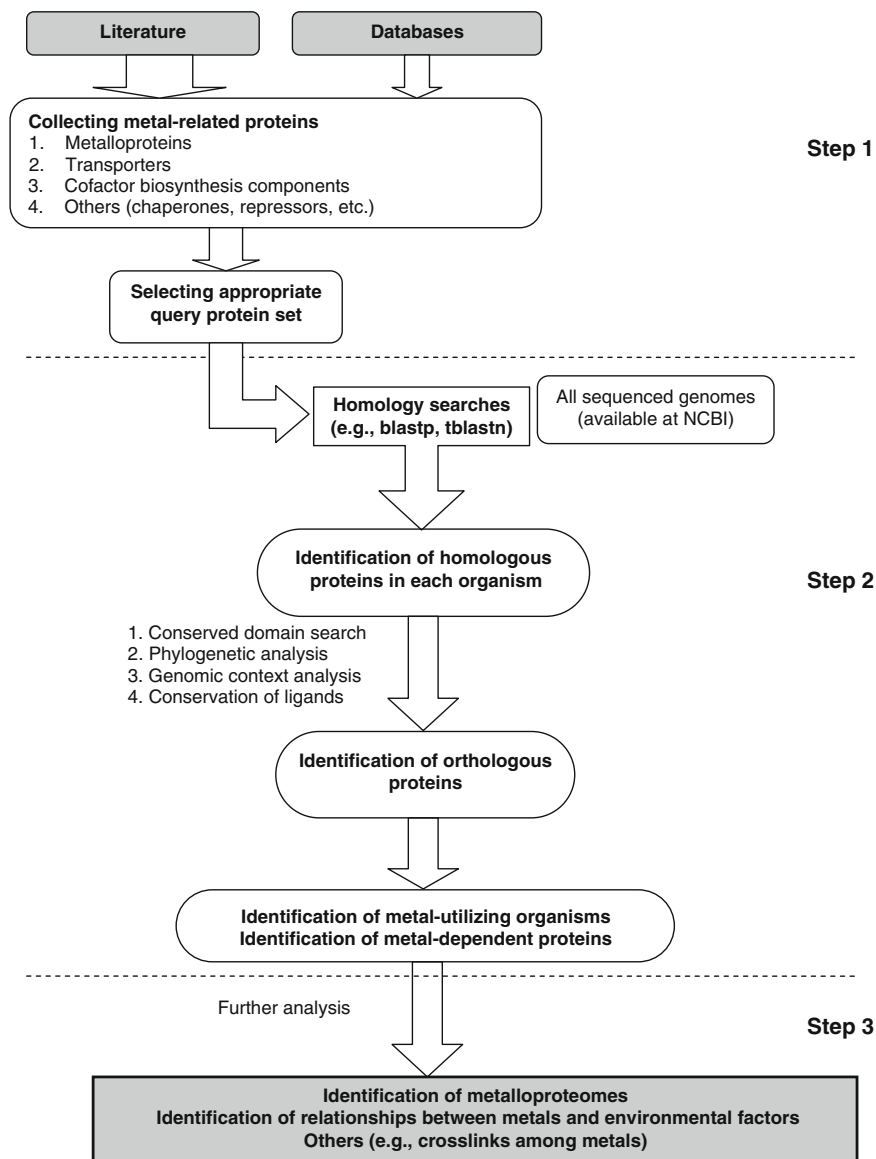


Figure 1 Schematic diagram for comparative genomics analyses of metal utilization in biology. This process can be divided into three major steps. Details are discussed in the text.

databases and bioinformatics tools have been developed to browse and/or predict metalloproteomes, e.g., Pfam, PROSITE, PRINTS, ProDom, COG, MDB, and dbTEU [43–48]. Some of these tools contain metal-binding sequence motifs or patterns, whereas others use position-specific scoring matrices or profiles to describe similarity among different proteins. These resources do not include all metal-binding

motifs and could only help identify a partial set of metalloproteins. Moreover, some metalloproteins may use different metals with the same ligands based on metal availability, protein folding location, and other factors [49].

Except for metalloproteins, occurrence of genes involved in high-affinity metal transport, metal-containing cofactor biosynthesis, and other processes (such as metal chaperones and repressors) may provide additional information regarding metal utilization and should be analyzed in parallel. Thus, a metal utilization trait could also be verified by the presence of high-affinity transporters and/or cofactor biosynthesis pathways.

2.2 Identification of Metal-Related Orthologs in Genomic Databases

The second step may be to identify orthologs of selected proteins in the sequenced genomes of all organisms. The list of currently sequenced organisms from the three domains of life is available at NCBI's website (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). As of January 2012, approximately 2,500 species were included.

In order to identify orthologs of query proteins, a set of sequences obtained in the above step can be used as initial seeds to search for homologous sequences in various organisms via a suite of BLAST programs (such as BLASTP, TBLASTN and PSI-BLAST) [50,51]. Orthologous proteins could be further defined using other approaches, such as conserved domain (COG/Pfam) searches, bidirectional best hits, genomic context analysis, and phylogenetic analysis. Conservation of metal-binding residues in the orthologs should also be analyzed to assess the ability to bind metal. Occurrence of Moco and vitamin B₁₂ biosynthesis could be verified by the presence of components involved in their pathways (see Sections 3.1 and 5.3). Finally, the presence of the utilization trait of a metal M in an organism could be verified by the requirement for occurrence of at least one M-specific transporter, or M-containing cofactor biosynthesis pathway, or at least one M-dependent metalloprotein.

It should be noted that only proteins strictly specific for a particular metal must be selected by this approach, which may result in incomplete analysis of metal utilization in some organisms. However, regarding the metals discussed in this chapter, most of the corresponding metalloproteins are strictly dependent on their primary metal. Therefore, comparative genomics approaches may indeed reveal a general picture of utilization of these metals in organisms.

2.3 Comparative Analyses of Metal Utilization and Interaction

Comparative analyses of metal utilization, function and evolution are among the most important goals of the metal biology field, which may greatly improve our

understanding of mechanisms and evolutionary dynamics of metals used in various organisms, clades, or kingdoms. Based on the results derived from previous steps, additional questions could be addressed, such as the relationship between metal utilization and environmental factors, composition and function of metalloproteomes, and interactions or other features of metal utilization. In the following sections, we will focus on several metals and discuss recent contributions on comparative genomics of their utilization.

3 Molybdenum

Molybdenum plays a critical role in several pathways and functions as a catalytic component of Mo-dependent enzymes (molybdoenzymes) that are essential for nearly all living organisms, including animals, plants, fungi, and bacteria. These molybdoenzymes catalyze oxo-transfer reactions in the metabolism of carbon, nitrogen, and sulfur compounds [27,52]. With the exception of the Fe-Mo cofactor in nitrogenase [53], all molybdoenzymes use this metal in the form of Moco, which consists of Mo coordinated to an organic tricyclic pyranopterin moiety, referred to as molybdopterin [27,54]. Some thermophilic archaea utilize W that is also coordinated by pyranopterin (Wco) [52,55]. In addition, W can be selectively transported into prokaryotic cells by certain transporters [56] and is an essential element for enzymes within the aldehyde:ferredoxin oxidoreductase family [52,57]. Due to the physical and chemical similarities between Mo and W, it is often impossible to distinguish the utilization of these two elements based on sequence analysis. In this chapter, the term Moco refers to the cofactor form of both metals (unless there is specific mention of the metal involved).

3.1 *Molybdenum Transport and Molybdenum Cofactor Biosynthesis*

Identification of Mo (or W) transporters and the Moco biosynthesis pathway are essential for characterization of the Mo utilization trait. In bacteria, the first identified Mo transporter was the ModABC transport system, which consists of ModA (molybdate-binding protein), ModB (membrane integral channel protein), and ModC (cytoplasmic ATPase) [12,58]. In *Escherichia coli*, the *modABC* operon is regulated by ModE repressor, which may sense intracellular levels of Mo and bind the promoter region of *modA* [59]. *E. coli* ModE is composed of an N-terminal DNA-binding domain (ModE_N) and a C-terminal molybdate-binding domain, which contains a tandem repeat of the Mo-binding protein (Mop) domain, or named Di-Mop domain. The ModABC-ModE systems are widely distributed in organisms, but are not ubiquitous, and variations of ModE were also observed in some Mo-utilizing organisms [38,60,61]. Two additional Mo/W ABC transport systems

with different substrate affinity, WtpABC (both Mo and W) and TupABC (W-specific), were reported [56,62]. Both transport systems exhibit low similarity to ModABC [38]. In *Campylobacter jejuni*, a ModE-like protein which lacks the Mop domain was recently reported to repress both ModABC (in the presence of either Mo or W) and TupABC (in the presence of W) systems [63]. However, the regulation of these two transporters is still unclear. Very recently, a member of a universal permease family, PerO, was found to import molybdate and other oxyanions in *Rhodobacter capsulatus*, which is the first reported bacterial molybdate transporter outside the ABC transporter family [64].

In contrast to the well-characterized Mo uptake transport in prokaryotes, information on Mo transport in eukaryotes is limited. In 2007, a high-affinity molybdate transport system, MOT1, which belongs to the sulfate transporter superfamily, was first characterized in *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* [65,66]. The *A. thaliana* MOT1 is strongly expressed in the roots and is localized to the mitochondria instead of the plasma membrane of root cells [67]. Recently, a novel Mo transporter family, MOT2, was identified in both *C. reinhardtii* and animals including humans, which opens a new way towards the understanding of molybdate transport in animals [68].

Moco is synthesized by an evolutionarily conserved multi-step pathway in all three domains of life [52,54,69]. The overall process includes (i) conversion of a guanosine derivative, most likely GTP, into cyclic pyranopterin monophosphate (cPMP, or precursor Z); (ii) transformation of cPMP into molybdopterin; (iii) metal incorporation into the apo-cofactor; and (iv) maturation to an active cofactor in some organisms, e.g., formation of a dinucleotide form (molybdopterin guanine dinucleotide, MGD) or substitution of a terminal oxygen ligand of Moco with a sulfur ligand. In *E. coli*, proteins required for Moco biosynthesis and regulation are encoded in the *moa-mog* operons (Figure 2a) [27,54]. In eukaryotes, at least six proteins (named Cnx1-3 and Cnx5-7 in plants) are involved in Moco biosynthesis (Figure 2b), which are homologous to their counterparts in bacteria [54,69–71]. Thus, the *moa-mog* genes and *cnx* genes could be used for identification of Moco biosynthesis pathways in prokaryotes and eukaryotes, respectively. In addition, a Moco sulfurylase, catalyzing the generation of the sulfurylated form of Moco that is needed for activation of the xanthine oxidase family of proteins such as xanthine dehydrogenase and aldehyde oxidase, has been identified in plants and humans [72,73]. A recent study also revealed that, in *A. thaliana*, the first step of Moco biosynthesis is localized in the mitochondrial matrix, and a mitochondrial ABC transporter ATM3 (previously implicated in the maturation of extramitochondrial Fe-S proteins) has a crucial role in Moco biosynthesis by transporting cPMP [74].

Considering that Moco is highly unstable [69], after synthesis, it should be either transferred immediately to the molybdoenzymes or bound to a storage/carrier protein until further insertion. In bacteria, many molybdoenzymes have known chaperones, such as NarJ for nitrate reductase and DmsD for dimethylsulfoxide reductase, which can bind Moco and assist in cofactor incorporation [75–77]. In contrast, little is known about Moco storage in eukaryotes. Recently, a Moco carrier protein (MCP) has been identified in *C. reinhardtii* [78]. MCP belongs to the lysine decarboxylase

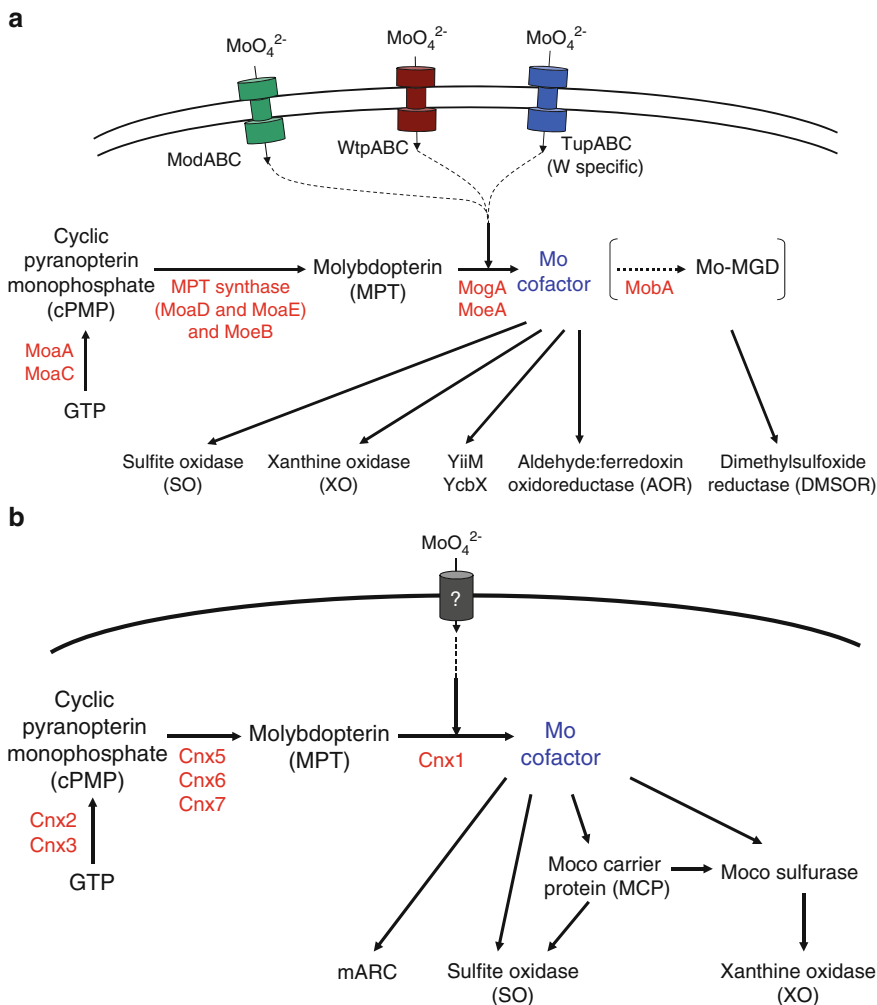


Figure 2 Biosynthesis of molybdenum cofactor. The pathway of Moco synthesis can be divided into three or four steps. **(a)** Biosynthesis of the Mo cofactor in prokaryotes. **(b)** Biosynthesis of the Mo cofactor in eukaryotes. The proteins from *E. coli* and *A. thaliana* catalyzing the respective steps are depicted and their names are given. MGD, molybdopterin guanine dinucleotide.

family and could bind Moco with high affinity. In addition, several homologous Moco-binding proteins (MoBP), which also belong to the lysine decarboxylase family, were discovered in land plants that might be involved in the cellular distribution of Moco [54,79]. However, the mechanism of Moco protection, storage and transfer in mammals is still unclear.

3.2 *Molybdoenzymes*

On the basis of cofactor composition and catalytic function, molybdoenzymes could be divided into two groups: (i) Mo-dependent nitrogenase that contains an Fe-Mo cofactor in the active site, and (ii) all other molybdoenzymes that bind Moco. Table 1 includes known molybdoenzymes.

Nitrogenase is required for biological nitrogen fixation, which is an essential step in the nitrogen cycle in the biosphere. There are four known types of nitrogenases,

Table 1 Mo-dependent proteins

| Group | Family | Protein |
|-----------------------|--|---|
| Moco-binding proteins | Sulfite oxidase | Sulfite oxidase |
| | | Nitrate reductase (assimilatory) |
| | Xanthine oxidase | Xanthine oxidase |
| | | Xanthine dehydrogenase |
| | | Aldehyde oxidase |
| | | Aldehyde oxidoreductase |
| | | 4-Hydroxybenzoyl-CoA reductase |
| | | CO dehydrogenase |
| | | Quinoline 2-oxidoreductase |
| | | Isoquinoline 1-oxidoreductase |
| | | Quinoline 4-carboxylate-2-oxidoreductase |
| | | Quinaldine 4-oxidoreductase |
| | | Quinaldic acid 4-oxidoreductase |
| | | Nicotinic acid hydroxylase |
| | | 6-Hydroxynicotinate hydroxylase |
| | | Nicotine dehydrogenase |
| | | Picolinate hydroxylase |
| | | Pyridoxal oxidase |
| | | Nicotinate hydroxylase |
| | Dimethylsulfoxide reductase | Dimethylsulfoxide reductase |
| | | Biotin sulfoxide reductase |
| | | Trimethylamine-N-oxide reductase |
| | | Nitrate reductase (dissimilatory) |
| | | Formate dehydrogenase |
| | | Formylmethanofuran dehydrogenase |
| | | Polysulfide/thiosulfate/arsenate reductase |
| | | Arsenite oxidase |
| | | Pyrogallol-phloroglucinol transhydroxylase |
| | | Aldehyde:ferredoxin oxidoreductase (W-containing) |
| | Formaldehyde ferredoxin oxidoreductase | |
| | Glyceraldehyde-3-phosphate ferredoxin oxidoreductase | |
| | Carboxylic acid reductase | |
| | Hydroxycarboxylate viologen oxidoreductase | |
| MOSC-containing | Aldehyde dehydrogenase | |
| | mARC/YcbX | |
| Fe-Mo-binding protein | Nitrogenase | YiiM |
| | | Nitrogenase |

each with a different combination of metals in the active site [80,81]. The most abundant and widely studied is the Fe-Mo-dependent nitrogenase, which contains MoFe₃S₃ and Fe₄S₃ cuboidal subunits triply joined by three bridging sulfurs.

The second group of proteins which utilize Moco as cofactor contains sulfite oxidase (SO), xanthine oxidase (XO), dimethylsulfoxide reductase (DMSOR), and aldehyde:ferredoxin oxidoreductase (AOR, mostly W-containing) families as well as the novel Moco-binding proteins. Each family includes a variety of subfamilies based on sequence similarity, spectroscopic properties and substrate preferences (Table 1). Compared to prokaryotes, which contain diverse members belonging to the four major families, eukaryotes only have four typical molybdoenzymes, including nitrate reductase (NR) and SO (members of the SO family), as well as xanthine dehydrogenase (XDH) and aldehyde oxidase (AO) (members of the XO family) [54,69].

Members of the SO family generally catalyze net oxygen atom transfer to or from a heteroatom lone electron pair rather than hydroxylation of a carbon center [82]. Typical enzymes belonging to this family include SO and assimilatory NR. SO is mainly found in eukaryotes and is located in the mitochondrial intermembrane space where it catalyzes the oxidation of sulfite to sulfate, the final step in the oxidative degradation of sulfur-containing amino acids [83]. The assimilatory NR catalyzes the reduction of nitrate to nitrite and is responsible for the first step in the uptake and utilization of nitrate [69]. So far, this enzyme has only been found in autotrophic organisms, such as plants and fungi.

The XO family contains the largest and most diverse Moco-containing enzymes (Table 1). Members of this protein family catalyze oxidative hydroxylation of a wide range of aldehydes and aromatic heterocycles [69]. The major enzymes of this family include AO (catalyzes the oxidation of a variety of aromatic and nonaromatic heterocycles and aldehydes), XDH (a key enzyme of purine degradation that oxidizes hypoxanthine to xanthine and xanthine to uric acid), and a variety of bacterial enzymes such as aldehyde oxidoreductase and 4-hydroxybenzoyl-CoA reductase.

The DMSOR family consists of a number of Moco-binding enzymes, all from bacterial and archaeal sources, that bind a Mo-MGD cofactor consisting of one Mo atom complexed by two MGD molecules [69,84]. Some of these enzymes possess Mo as their sole redox-active center [85]. They are very diverse in reaction, function, and structure [86]. Most of these enzymes function as terminal reductases under anaerobic conditions where their respective cofactors serve as terminal electron acceptors in respiratory metabolism. DMSOR is found in bacteria and catalyzes reductive deoxygenation of dimethyl sulfoxide to dimethyl sulfide. It is a periplasmic single-subunit protein in some bacteria (such as *Rhodobacter sphaeroides*), whereas a membrane-bound protein composed of three subunits (Moco-containing, four [4Fe-4S] cluster-containing and transmembrane subunits) in some other bacteria (such as *E. coli*) [87,88]. Another widespread member of the DMSOR family is formate dehydrogenase, which catalyzes the oxidation of formate to bicarbonate and is also a selenocysteine-containing enzyme in many organisms [89]. Other members include dissimilatory NR, trimethylamine-N-oxide reductase and several additional enzymes exhibiting substantial sequence homology [69,84,86].

AOR catalyzes the interconversion of aldehydes and carboxylates and was the first member of the AOR family to be structurally characterized as a protein containing a Wco cofactor that is analogous to the Moco [90]. Other members include formaldehyde ferredoxin oxidoreductase, glyceraldehyde-3-phosphate ferredoxin oxidoreductase, and carboxylic acid reductase.

In addition to the four major molybdoenzyme families, novel Moco-binding proteins were recently identified in both mammals and *E. coli* [91,92]. In mammals, a Moco-dependent protein was found in the outer mitochondrial membrane and named mitochondrial amidoxime-reducing component (mARC). mARC binds a Moco that carries neither a terminal sulfur ligand like XO nor a covalently bound cysteine (Cys) residue like SO, suggesting that these proteins represent a new family of molybdoenzymes [93]. Recent studies have shown that human mARC proteins may catalyze the N-reduction of a variety of N-hydroxylated substrates such as N-hydroxy-cytosine and N- ω -hydroxy-L-arginine albeit with different specificities [93–95]. So, mARC and its N-reductive enzyme system plays a major role in drug metabolism [95].

3.3 Comparative Genomics of Molybdenum Utilization

Although Mo is an important transition metal, comprehensive analyses of the occurrence and evolutionary trends in its utilization, which could greatly benefit our understanding of Mo and its evolutionary dynamics, have been limited. In recent years, following the availability of a large number of sequenced genomes, several comparative genomics studies have been carried out to investigate the phylogeny of Mo utilization in prokaryotes and eukaryotes at the level of Mo transport, the Moco biosynthesis, and molybdoenzymes [38,39,96,97]. These studies provided a first glance on the Mo utilization in the three domains of life and showed its widespread occurrence, yet limited use of this metal in individual organisms.

First, a wide distribution of genes encoding Mo transport systems, the Moco biosynthesis pathway, and Mo-containing proteins was found in sequenced genomes, and almost all Mo-utilizing organisms contained both Moco biosynthesis proteins and at least one known molybdoenzyme [38,96]. In bacteria and archaea, Mo was utilized by almost all phyla (except *Mollicutes* and *Chlamydiae*), suggesting that Mo utilization is an ancient and essential trait that is common to essentially all organisms in these two domains. In eukaryotes, Mo is used by all animals, land plants, algae, certain fungi, and stramenopiles; however, parasites, yeasts (saccharomycotina and schizosaccharomycetes) and free-living ciliates lack the Mo utilization trait. It is possible that many protozoa, especially parasites, lost the ability to utilize Mo.

Comparative analyses of Mo/W transport systems in sequenced prokaryotes revealed that Mo/W transporters are often present in single copies [39,96,97]. Among them, ModABC is the most common Mo transporter, which is present in approximately 90% of Mo-utilizing bacteria. The occurrence of the other two transporters, WtpABC and TupABC, is much more restricted, especially WtpABC,

which is only detected in 3% of Mo-utilizing bacteria. On the other hand, WtpABC is the most frequently used transporter in archaea. It appeared that WtpABC is mainly an archaeal Mo/W transporter, whereas ModABC functions predominantly in bacteria. The full length-type ModE regulation of ModABC transporters only occurred in less than 30% of Mo-utilizing organisms, suggesting the presence of novel or unspecific regulatory pathways for molybdate uptake in many other organisms such as Gram-positive bacteria and cyanobacteria [38,97]. On the other hand, individual ModE_N and/or Mop/Di-Mop proteins, and novel domain fusions for either ModE_N or Mop were observed in a variety of organisms that lack full-length ModE, indicating complexity of ModE-related regulation. Genomic context analyses of these ModE-related variations suggested potential correlations with ModABC transporters as most of these genes are close to or are even in the same operon with *modABC* [38,96]. It was previously thought that a separate ModE_N and Mop/Di-Mop proteins together may have a function similar to that of the full-length ModE [98]. In eukaryotes, MOT1 was detected in some Mo-utilizing organisms such as land plants, green algae, and stramenopiles, whereas the recently identified MOT2 appeared to have a wider distribution in algae, land plants, and animals [68,99]. Thus, MOT1 and MOT2 proteins may play key roles in Mo transport in eukaryotes although additional unknown Mo transport systems may be also present.

The majority of known proteins involved in Moco biosynthesis pathways could be detected in essentially all Mo-utilizing organisms. However, a very small number of prokaryotes, which contain homologs of molybdoenzymes, lack genes for either Moco biosynthesis components or Mo/W transporters [96]. It is possible that Moco is dispensable for the molybdoprotein homologs in these organisms. Nevertheless, there is a very good correspondence between occurrence of the Moco biosynthesis trait and Moco utilization in the three domains of life.

Comparative genomics of molybdoenzymes also showed complexity in their evolutionary trajectories. In bacteria, DMSOR, SO, and XO families were widespread, especially DMSOR whose members (mostly DMSOR, dissimilatory NR, and formate dehydrogenase) were detected in more than 90% Mo-utilizing organisms [96,97]. In contrast, the W-containing AOR family was only detected in ~15% Mo/W-utilizing organisms. In archaea, DMSOR was also the most abundant molybdoenzyme family (more than 95% Mo-utilizing organisms). Interestingly, members of the AOR family had a much higher occurrence in archaea (~70%). The FeMo-utilizing molybdoenzyme, nitrogenase, was detected in ~20% of Mo-utilizing bacteria (almost all also used Moco) and methanogenic archaea. Further investigation of the predicted molybdoenzyme set (molybdoproteome) of each organism revealed that proteobacteria have larger molybdoproteomes than other organisms [96]. *Desulfitobacterium hafniense*, a dehalorespiring bacterium, was found to have the largest known molybdoproteome in prokaryotes (63 molybdoproteins, 95% of which are members of the DMSOR family). In eukaryotes, almost all Mo-utilizing organisms had SO and XO families. Land plants possessed the largest molybdoproteomes in eukaryotes (10-11 molybdoproteins). On the other hand, all sequenced saccharomycotina (e.g., *Saccharomyces cerevisiae*) and schizosaccharomycetes (e.g., *Schizosaccharomyces pombe*) had neither known molybdoenzymes nor Moco

biosynthesis proteins. It was previously reported that a small number of unsequenced yeast species, such as *Candida nitratophila* and *Pichia angusta*, may utilize Mo-containing assimilatory NR [39,100], but the fact that both homologs of this protein and the Moco biosynthesis pathway are absent in all currently sequenced yeast genomes suggests the loss of Mo utilization in these organisms.

Recent studies on the new mammalian Moco-binding mARC protein revealed that it consists of two conserved domains: N-terminal MOSC_N (pfam03476) and C-terminal MOSC (pfam03473) domains, which are also present in Moco sulfurases [97]. The MOSC domain of eukaryotic Moco sulfurase is involved in Moco binding with high affinity and its Moco carries a terminal sulfur ligand due to the catalytic activity of pyridoxal-5'-phosphate-dependent NifS-like domain [101]. The function of the MOSC_N domain is unknown; however, it is predicted to adopt a β -barrel fold. Two additional Moco-dependent proteins, YcbX and YiiM, were characterized in *E. coli*, which may represent novel enzymatic activities involved in the detoxification pathway of N-hydroxylated base analogs [92]. Both proteins contain the MOSC and additional domains (Figure 3). Bioinformatics analyses showed that *E. coli* YcbX and mammalian mARC proteins could be considered as orthologs and are members of the same family (mARC/YcbX family). On the other hand, no significant sequence similarity could be detected between YiiM and mARC/YcbX, suggesting that they belong to different families within the MOSC superfamily [97]. Further analysis revealed that, in bacteria, both mARC/YcbX and YiiM were widespread but only detected in Moco-utilizing organisms. In contrast, the occurrence of these two families in archaea is limited, i.e., only organisms belonging to *Euryarchaeota/Halobacteriales* had mARC/YcbX proteins. In eukaryotes, mARC proteins were detected in more than 95% Mo-utilizing organisms, suggesting a wide distribution of this novel molybdoenzyme family. In addition, a novel group of

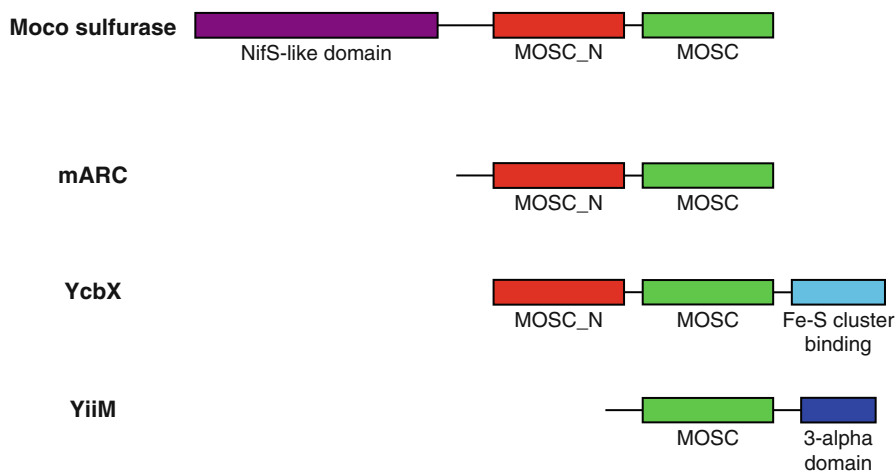


Figure 3 Domain organizations of Moco sulfurase and novel Moco-containing proteins. Distinct domains are shown by different colors. MOSC, C-terminal domain of the eukaryotic Moco sulfurase.

MOSC-containing proteins (designated MOSC-like), which form a separate branch in the MOSC family and might serve as chaperones involved in Moco transfer or storage, was identified in some prokaryotes [97]. In general, these studies suggested complexity and diverse roles of the MOSC superfamily, whose proteins may be (i) involved in the Moco modification pathway (Moco sulfurase); (ii) new molybdoenzymes (mammalian mARC, *E. coli* YcbX and YiiM); (iii) potential Moco chaperones (MOSC-like); and (iv) involved in other functions. Further experiments are needed to better understand the functions of MOSC-containing proteins.

An interesting link between Mo and selenium was also observed as a major member of the DMSOR family, formate dehydrogenase α subunit, is also a selenocysteine-containing protein that may be responsible for maintaining the selenocysteine utilization trait in sequenced prokaryotes [96,97]. Thus, the selenocysteine utilization trait depends on the Mo utilization trait in prokaryotes, most likely because of formate dehydrogenase, which is not only a widespread molybdoenzyme but is also the major user of Se in prokaryotes. In addition, some environmental conditions and other factors may affect Mo utilization and molybdoenzyme families. For example, the majority of intracellular parasites and symbionts lost the ability to utilize Mo, whereas more than 80% of extracellular symbionts utilize the metal [96]. Organisms possessing W-containing AOR proteins appear to favor an anaerobic environment, whereas organisms containing SO or XO proteins favor aerobic conditions. On the other hand, organisms possessing nitrogenase favor both anaerobic and relatively warm conditions [96,97]. These findings suggest that although being dependent on the same processes, such as Mo availability and Moco synthesis, different Mo enzymes are subject to independent and dynamic evolutionary processes. However, no significant correlation was observed between various factors examined and the size of molybdoproteomes.

4 Copper

Copper is an essential micronutrient that serves as an important cofactor for proteins and enzymes that carry out several fundamental biological functions [102]. On the other hand, Cu is highly toxic in the free form because of its ability to produce radicals by cycling between Cu(I) and Cu(II) species [103]. It is important for Cu-utilizing organisms to obtain sufficient levels of Cu ion to meet their needs while tightly controlling intracellular Cu concentration.

4.1 Overview of Copper Trafficking and Homeostasis

Cellular Cu trafficking processes are required for correct utilization of this element in biochemical processes and for limiting Cu toxicity. Cu import mainly needs the coordinate function of proteins with metal-binding domains, whereas detoxification

mechanisms include the binding of Cu to specific proteins (e.g., metallothioneins) and its transfer into cell compartments such as periplasmic space [104].

In prokaryotes, the mechanisms involved in Cu transport and homeostasis are not completely understood. No specific Cu import system has been identified in the majority of bacteria, possibly reflecting no cytosolic requirement for Cu, and the mechanism of Cu entry is largely unknown [105]. To date, Cu trafficking in bacteria is best described in *E. coli* and in *Enterococcus hirae* [106,107]. The most relevant Cu homeostatic systems in *E. coli* are shown in Figure 4a. Several Cu-related transport and resistant proteins have been characterized in a variety of organisms, including CopA/PacS, CusCFBA, CutC, PcoABCDRSE [105–107].

In *E. coli*, the Cu(I)-translocating P-type ATPase CopA is the major component of Cu homeostasis and serves as an exporter for removing Cu(I) from the cytoplasm [105,106]. CopA proteins belong to a superfamily that is involved in transport of transition or heavy metal ions (including Zn, Cd, Ag, Pb, and Co) across membranes [108]. Two Cys residues in a Cys-Pro-Cys motif located in the middle of CopA are needed for CopA function [109]. PacS is a CopA homolog in cyanobacteria and may be involved in Cu homeostasis crucial to the photosynthetic thylakoid function [110]. CtaA, another CopA homolog identified in cyanobacteria, was suggested to be involved in Cu import from the periplasm [110,111]. Both CtaA and PacS are required for Cu transport into the thylakoid [110]. In *E. hirae*, two CopA homologs, CopA and CopB, were identified. The former may be involved in Cu uptake, whereas the latter functions as an exporter of Cu ion [107]. In *E. hirae* and many other organisms a Cu chaperone, CopZ, functions as part of a complex cellular machinery for Cu trafficking and detoxification [112]. A role for *E. hirae* CopZ in routing Cu to the cytoplasmic Cu sensor CopY, to alleviate CopY-mediated repression of the *copYZAB* operon, has also been reported [113]. CopZ homologs are also found in eukaryotes, called Atx1, which interact with the Golgi P-type ATPase Cu transporter ATP7 [114]. The CopZ/Atx1 proteins adopt a very similar structure to the amino-terminal metal-binding domains of the PIB- type ATPases (a subgroup of P-type ATPases that transports transition metals between different compartments of the cell) with typical $\beta\alpha\beta\beta\alpha\beta$ ferredoxin-like folding and a GXXCXXC metal-binding motif present on a flexible solvent-exposed loop [115]. Recently, a Cu-binding metallothionein, MymT, was found in several pathogenic mycobacteria, which may also serve as a chaperone involved in CopA-related Cu(I) detoxification [116].

Gram-negative bacteria contain another Cu efflux system, the CusCFBA system, which includes CusA (the inner membrane pump), CusB (the periplasmic protein), CusC (the outer membrane protein forming a channel bridging the periplasmic space) and CusF (a small periplasmic protein that binds a single Cu(I) ion and interacts with both CusC and CusB) [117]. The recent elucidation of the structure of the Cu(I) bound form of CusF has revealed a new metal recognition site in which Cu(I) is tetragonally displaced from a Met_2His ligand plane toward a conserved tryptophan which involves cation- π interactions [118]. In *E. coli*, genes encoding the four-part Cus complex were present in one *cus* operon, which is only required under conditions of extreme Cu stress and is particularly important under anaerobic conditions [117]. These genes are induced in response to elevated Cu by the CusRS

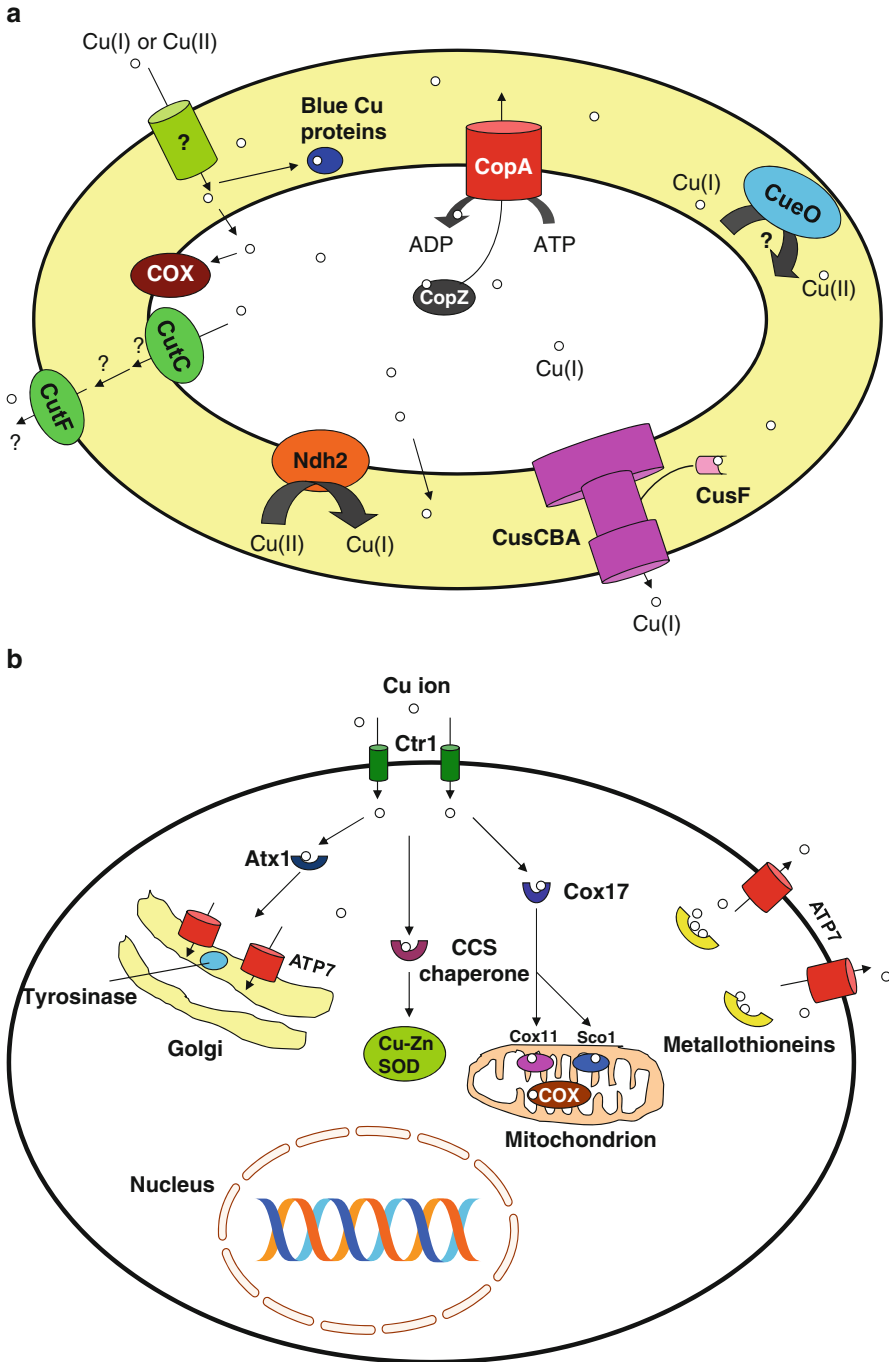


Figure 4 Schematic view of Cu homeostasis. **(a)** Cu homeostasis in *E. coli*. CopA, the Cu(I)-translocating P-type ATPase; CusCFBA, the four-component Cu efflux system; Ndh 2, a cupric reductase; CueO, a multicopper oxidase; CutC and CutF, two proteins involved in Cu efflux and/or homeostasis; CopZ, a Cu chaperone involved in Cu export; COX, cytochrome *c* oxidase. **(b)** Cu homeostasis in *Drosophila melanogaster*. Atx1, CCS, and Cox17, Cu chaperones involved in various pathways; Ctr1, eukaryotic Cu importer; ATP7, eukaryotic Cu exporter (also involved in Cu transport to Golgi); COX11 and Sco1, two proteins involved in cytochrome *c* oxidase assembly; Cu-Zn SOD, Cu-Zn superoxide dismutase.

two-component sensory system, which typically monitors stress at the cell envelope and is thought to respond to Cu(I). It has been shown that CusA and CusB are essential for Cu resistance, and CusC and CusF are required for full resistance [119]. A role for the Cus system in providing Ag resistance has also been shown [119].

Some *E. coli* strains harbor an additional plasmid-borne Pco system which involves seven genes, *pcoABCDRSE*, that confer Cu resistance. The mechanism of Cu detoxification provided by this system is largely unknown but includes the multicopper oxidase PcoA and its putative partner PcoC, both of which are exported to the periplasm, PcoD that is thought to transport Cu across the cytoplasmic membrane, PcoB that is a predicted outer membrane protein and PcoE, a periplasmic protein that binds Cu [106]. PcoRS form a two-component regulator required for Cu-inducible expression of *pco*. A number of other proteins have also been linked to Cu resistance in *E. coli*, including the products of *cutABCDEF* genes, which were identified based on the preliminary characterization of Cu-sensitive mutants [120]. However, few of these genes have been directly linked to Cu metabolism, transport, or regulation. Previous studies implicated CutC in Cu efflux, suggesting a role for CutC in intracellular trafficking of Cu(I) [120]. CutC homologs have also been characterized in eukaryotes, including humans. Recently, a crystal structure of human CutC was reported, suggesting that this protein may function as an enzyme with Cu(I) as a cofactor rather than a Cu transporter and that the potential Cu(I)-binding site consists of two Cys residues and other conserved residues in the vicinity [121].

A general scheme of Cu homeostasis in eukaryotes (using *Drosophila melanogaster* as a representative organism) is shown in Figure 4b. In eukaryotes, Cu is acquired by the high-affinity Cu transporter (Ctr) family proteins [18]. Members of the Ctr family possess an N-terminal extracellular Met-rich domain which is important for the acquisition of Cu(I) ions [122]. Different organisms may possess multiple Ctr proteins located in different biological membranes. *S. cerevisiae* has three Ctr proteins, yCtr1-yCtr3 [123]. yCtr1 and yCtr3 are located in the plasma membrane, whereas yCtr2 is localized in the vacuolar membrane and imports Cu from the vacuole to the cytoplasm [124]. *D. melanogaster* also has three *ctr1* genes (*ctr1A*, *ctr1B*, and *ctr1C*). Ctr1A is located in the plasma membrane and is the major Cu transporter during growth and development. Ctr1B also localizes to the plasma membrane and is not essential for development unless flies are severely Cu-deficient or are subjected to Cu toxicity [125]. Ctr1C is mainly expressed in male gonads and functions as a Cu importer in the male germline, specifically in maturing spermatocytes and mature sperm [126]. Humans contain two Ctr proteins (hCtr1 and hCtr2). hCtr1 is the main Cu importer, which is located predominantly at the plasma membrane, but may also be present in vesicular compartments [127]. hCtr2 was found exclusively to late endosomes and lysosomes and may be involved in the delivery of Cu ions to the cytosol [128].

Cu export in eukaryotes is mediated by an important category of ATP-dependent transporters, the ATP7 family, which is homologous to bacterial CopA proteins [129]. In mammals, there are two isoforms: ATP7A and ATP7B [130]. ATP7A is expressed in the intestinal epithelium as well as most other tissues (such as brain and heart) except the liver, which is required for transport of Cu into the trans-Golgi

network for biosynthesis of several secreted cuproenzymes and for basolateral efflux of Cu in the intestine and other cells [131]. ATP7B is mainly expressed in the liver and is needed for Cu metalation of ceruloplasmin and biliary Cu excretion [131]. *D. melanogaster* has a sole ATP7 protein (named DmATP7), which is required for *in vivo* Cu distribution [132]. Yeasts also have an ATP7 ortholog Ccc2, which is located in the trans-Golgi membrane, obtains Cu from Atx1 and transfers it to several secreted proteins [133].

4.2 Cuproproteins

Copper plays important roles in electron transfer, oxidation of organic substrates and metals, dismutation of superoxide, monooxygenation, transport of dioxygen and iron, and several other processes. So far it has not been possible to identify all Cu-binding proteins (cuproproteins) in any organism using bioinformatics approaches. This chapter only focuses on strictly Cu-dependent protein families which have been used for comparative genomics of Cu utilization in recent studies.

To date, a number of Cu-containing proteins have been characterized (a list is shown in Table 2). Cu centers in proteins could be divided into three types based on spectroscopic and structural properties. Type 1 Cu (also called blue Cu) shows intense absorption at around 600 nm and narrow hyperfine splittings in the electron paramagnetic resonance (EPR) spectroscopy, type 2 Cu does not give strong absorption at 600–700 nm and shows hyperfine splittings of the normal magnitude in the EPR spectrum, whereas type 3 Cu could be detected by neither strong absorption nor EPR studies. Structurally, the Cu atom of a typical type 1 site is coordinated by

Table 2 Cu-dependent proteins

| Prokaryotes | Eukaryotes |
|--|---|
| Plastocyanin family (including plastocyanin, amicyanin, pseudoazurin, halocyanin, etc.) | Plastocyanin family |
| Azurin family (including azurin and auracyanin) | Plantacyanin family (including plantacyanin, umecyanin, mavicyanin, stellacyanin, etc.) |
| Rusticyanin | Cytochrome <i>c</i> oxidase subunit I |
| Nitrosocyanin | Cytochrome <i>c</i> oxidase subunit II |
| Cytochrome <i>c</i> oxidase subunit I | Cu-Zn superoxide dismutase |
| Cytochrome <i>c</i> oxidase subunit II | Copper amine oxidase |
| Nitrous oxide reductase | Peptidylglycine R-hydroxylating monooxygenase |
| NADH dehydrogenase 2 | Dopamine β -monooxygenase |
| Cu-Zn superoxide dismutase | Multicopper oxidases (including laccase, Fet3p, hephaestin, ceruloplasmin, ascorbate oxidase, etc.) |
| Copper amine oxidase | |
| Particulate methane monooxygenase | Tyrosinase (or polyphenol oxidase) |
| Multicopper oxidases (including nitrite reductase, CueO, CotA, laccase, bilirubin oxidase, phenoxazinone synthase, etc.) | Hemocyanin |
| | Cnx1G |
| Tyrosinase | Galactose oxidase |

a Cys and two His residues in a trigonal planar arrangement. Often the thioether of a methionine coordinates axially, distorting the geometry towards tetrahedral. Most type 2 sites are three to four coordinate and one or more of the Cu ligands are the imidazole side chains of His residues. The coordination sphere may be completed by methionine, glutamate, glutamine or tyrosine. Type 3 sites consist of two antiferromagnetically coupled Cu atoms bridged by molecular oxygen or a hydroxyl. Some Cu-dependent proteins, such as multicopper oxidases (MCOs), may contain multiple Cu centers.

Blue Cu proteins (also named cupredoxins) are a group of relatively small proteins containing a single type 1 Cu center. They function in electron transfer in the respiratory and photosynthetic chains of many bacteria and plants [134,135]. These proteins include plastocyanin, azurin, pseudoazurin, amicyanin, rusticyanin, auracyanin, plantacyanin, and some other proteins. Plastocyanin is the best studied blue Cu protein which shuttles electrons from cytochrome b6/f to photosystem I. Crystal and NMR solution structures of several plastocyanins have revealed that this protein has an eight-stranded Greek-key β -barrel fold and contains a type 1 Cu atom coordinated by two histidines, one Cys and one methionine. The red Cu protein, nitrosocyanin, is a variant of the blue Cu protein, whose Cu site is the only known blue Cu-related site with an exogenous water molecule bound to Cu [135,136].

A type 1 Cu center is also detected in several larger enzymes, such as nitrite reductase (NiR) that catalyzes the reduction of nitrite to nitric oxide, and MCOs that function in intramolecular electron transfer. MCOs include a large number of proteins, such as laccase, ascorbate oxidase, CueO, PcoA, EpoA, dihydrogeodine oxidase, hephaestin, ceruloplasmin, phenoxazinone synthase, Fet3p, etc. [135,137]. Most MCOs have four Cu centers: a type 1 Cu and a mixed Cu center containing a type 2 and two type 3 Cu atoms. These MCOs catalyze the oxidation of small molecules and cations with the concomitant four-electron reduction of oxygen to water. Some MCOs such as mammalian ceruloplasmin and yeast Fet3p are ferroxidases, whereas laccases derive electrons from the oxidation of phenolic compounds.

Two additional Cu-dependent proteins, cytochrome *c* oxidase (COX) and nitrous oxide reductase (N_2OR), have a binuclear Cu center, named Cu_A , which is a variant of type 1 Cu. Cytochrome oxidase family members act as the terminal enzymes in respiratory chains. The two major subgroups of this family include COX and quinol oxidase [138]. Both classes have several catalytic subunits, and subunit I contains two heme centers: the first (heme a) acts as an electron input device to the second, and the second (heme a_3) is a part of a binuclear center containing Cu_B . However, there are significant differences of subunit II between the two subgroups. COX subunit II contains the Cu center Cu_A with 2 Cu atoms, which might be the immediate electron acceptor from cytochrome *c*, whereas quinol oxidase subunit II lost the Cu_A center [139]. Three subtypes (aa3, ba3, and cbb3) of both COX and quinol oxidase have been reported [138]. Characterizing all these subtypes and distinguishing the Cu-dependent COX subunit II from the Cu-independent quinol oxidase subunit II is important for correct description of Cu utilization. N_2OR transforms nitrous oxide to dinitrogen and carries six Cu atoms. Two are arranged in the binuclear Cu_A site similar to that of COX, and four make up the sulfide-bridged Cu cluster (Cu_2 catalytic center). The crystal structure of *Pseudomonas nautica* N_2OR revealed

that the Cu_z center belongs to a new type of metal cluster in which the four Cu ions are bound by seven histidine residues [140].

The type 2 Cu-containing proteins include Cu-Zn superoxide dismutase (Cu-Zn SOD), Cu amine oxidase (CuAO), peptidylglycine R-hydroxylating monooxygenase (PHM), and dopamine β -monooxygenase (DBM) [141].

Cu-Zn SOD is widespread in both eukaryotes and prokaryotes. Most studies have focused on the enzymes from eukaryotic sources, such as yeast and human. It has been found that, in the oxidized Cu-Zn SOD, Cu is coordinated by four histidine residues.

CuAO belongs to a larger group of amine oxidases that catalyze deamination of amines with concomitant reduction of oxygen to hydrogen peroxide. These enzymes are found in a large variety of organisms, from microbes to mammals. In bacteria, CuAOs have important roles in providing carbon or nitrogen sources when primary amines are available. In mammals, CuAOs are found in various tissues, including placenta, blood, muscle, and endothelium. It has been reported that increased CuAO expression in humans might be a marker of several diseases including cancer, diabetes and liver cirrhosis [142]. Crystal structures of CuAO from different organisms showed that the Cu atom is coordinated by three histidines and two water molecules [143].

In peptidylglycine R-hydroxylating monooxygenase (PHM) and dopamine β -monooxygenase (DBM), two distinct Cu sites are used to split oxygen, which then serves as the source of OH in the hydroxylation of their respective substrates [141]. Both enzymes are mainly detected in metazoa, and their functions are well established. PHM is one of two domains in peptidylglycine R-amidating monooxygenase (PAM), which is essential for the activation of a variety of hormones by R-amidation, thereby improving hormone-receptor affinity. DBM catalyzes a similar reaction to PHM; however, the hydroxylation of dopamine is at the β -carbon. Sequence analysis revealed that DBM is homologous to PHM, suggesting that they may have evolved from a common ancestor [144].

Other Cu-dependent proteins include NADH dehydrogenase 2 (Ndh2), tyrosinase, hemocyanin, particulate methane monooxygenase (pMMO), Cnx1G, and galactose oxidase (GAO): (1) The Cu(II)-reductase Ndh2 from *E. coli*, which contributes to antioxidant function and Cu homeostasis, is a membrane-bound reductase that diminishes the susceptibility of the respiratory chain to damaging effects caused by Cu and hydroperoxides. (2) Tyrosinases (or catechol/polyphenol oxidases) are ubiquitously distributed in all domains of life. They are essential for pigmentation and are important factors in wound healing and primary immune response. The active site is a type 3 Cu center consisting of two Cu ions, each coordinated by three histidine residues. (3) Hemocyanin is also a type 3 Cu protein family and occurs in the hemolymph of some species in arthropoda and mollusca. These proteins are extracellular oxygen carriers that are responsible for the precise oxygen delivery from the respiratory organs to tissues. (4) pMMO is a membrane-bound Cu-containing enzyme that oxidizes methane to methanol in methanotrophic bacteria. The crystal structure of *Methylococcus capsulatus* pMMO reveals the composition and location of three metal centers, which provides new insight into the molecular details of biological methane oxidation [145]. (5) Cnx1G is the G domain of Cnx1 that is involved in catalyzing the insertion of Mo into molybdopterin (see Section 3.1 and Figure 2b). Identification of the Cu bound to the molybdopterin dithiolate sulfurs

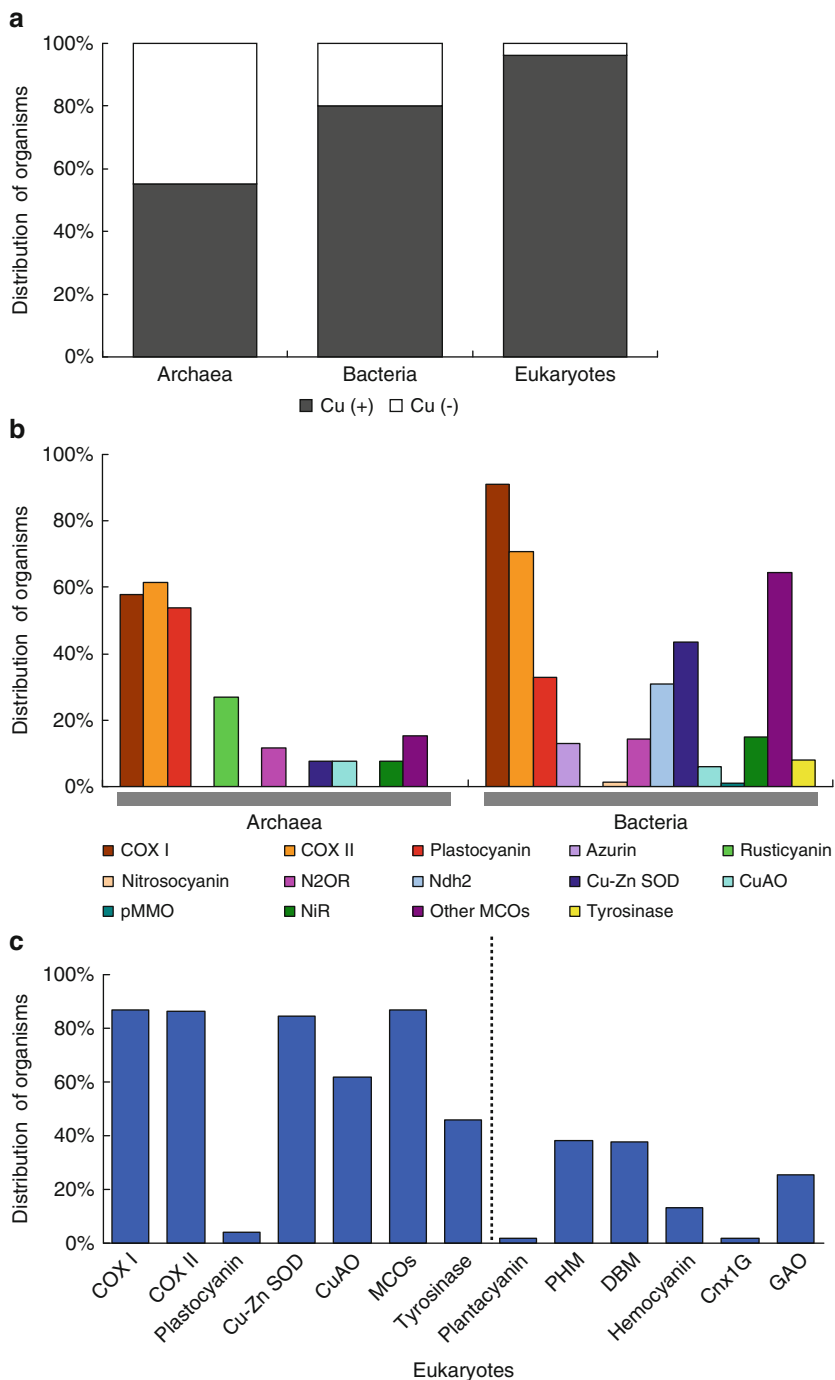


Figure 5 Occurrence of Cu utilization in the three domains of life. **(a)** Proportion of Cu-utilizing organisms among organisms with sequenced genomes. All organisms were classified into two groups: Cu (+), i.e., containing the Cu utilization trait; Cu (-), i.e., lacking the Cu utilization trait.

in CnxIG structures provides an important link between Mo and Cu utilization [146]. (6) GAO contains a single Cu ion and an amino acid-derived cofactor. The enzyme has been well studied, which has provided insights into the catalytic mechanism of this enzyme. One of the most interesting features of the enzyme is the posttranslational generation of an organic cofactor from its active-site amino acid residues, one of which might be also one of the Cu ligands [147].

4.3 Comparative Genomics of Copper Utilization

In recent years, several comparative genomics studies have been carried out to identify the Cu utilization trait and Cu-binding proteins in organisms [148–150]. For example, one study developed a computational approach based on conserved metal-binding patterns of metalloproteins in the PDB to search for new metalloproteins, and applied this method for Cu [148]. A set of Cu-binding patterns was obtained for all Cu-binding proteins in the PDB and then combined with the primary sequences of corresponding metalloproteins to identify all cuproproteins by homology searches. This procedure retrieved a significant number of false positive metalloproteins. To solve this problem, additional searches integrated with domain recognition methods were carried out, which showed better sensitivity and selectivity [34]. Based on this modified approach, the occurrence of Cu-binding proteins in 57 completely sequenced genomes in prokaryotes and eukaryotes was further examined [150]. The size of the Cu proteome is generally less than 1% of the total proteome. The number of putative Cu-binding proteins did not correlate with the size of the proteome, which is different from several other metals such as Zn [34]. Functional analysis of Cu-binding proteins revealed that these proteins are likely to be part of a network which may represent an ancient core that is crucial for Cu homeostasis. The speciation of prokaryotic organisms appeared to only slightly affect this ancestral Cu proteome, whereas eukaryotes may have expanded their ancestral repertoires of Cu proteins, by evolving new Cu domains and reusing old domains for new functions.

Much more comprehensive comparative genomics analyses of sequenced prokaryotes and eukaryotes were then carried out to yield a clearer view of Cu utilization in all three domains of life [37,96]. Using the strategy introduced in Section 2, occurrence of the Cu utilization trait, Cu transporters and strictly Cu-dependent proteins was examined. The distribution of Cu-utilizing organisms and Cu-dependent proteins is shown in Figure 5. Consistent with previous observations,

← **Figure 5** (continued) (b) Occurrence of Cu-dependent proteins in Co-utilizing prokaryotes. (c) Occurrence of Cu-dependent proteins in Cu-utilizing eukaryotes. Protein families on the left side of the dotted line have Cu-containing homologs in bacteria whereas others were only found in eukaryotes. COX I, cytochrome *c* oxidase subunit I; COX II, cytochrome *c* oxidase subunit II; N₂OR, nitrous oxide reductase; Ndh2, NADH dehydrogenase 2; Cu-Zn SOD, Cu-Zn superoxide dismutase; CuAO, Cu amine oxidase; pMMO, particulate methane monooxygenase; NiR, nitrite reductase; MCOs, multi-Cu oxidases; PHM, peptidylglycine R-hydroxylating monooxygenase; DBM, dopamine-monoxygenase; GAO, galactose oxidase.

Cu is widely used by bacteria, i.e., approximately 80% of analyzed organisms were found to be Cu-utilizing. However, all or almost all organisms in some bacterial phyla (such as *Thermotogae*, *Mollicutes*, *Chlamydiae*, and *Spirochaetes*) lacked known Cu-dependent proteins.

In archaea, only half of organisms appeared to utilize Cu (Figure 5a). Analysis of Cu transporters revealed that they had somewhat different patterns of occurrence than Cu-dependent proteins. First, CopA appeared to be the most widespread Cu exporter in bacteria and was the only Cu transporter detected in archaea. Occurrence of other transporters was relatively limited, especially the Cus system that was only detected in Gram-negative bacteria [96]. Second, many organisms, including those that lack known Cu-dependent proteins, had Cu exporters. These data suggested that the pathways of Cu utilization and detoxification are independent and that many organisms likely protect themselves against Cu ions that inadvertently enter the cell. Although occurrence of Cu transporters in prokaryotes may not provide sufficient information about Cu utilization, it may be important for understanding of Cu homeostasis. Third, some organisms were found to have multiple copies of certain Cu transporters. The highest number of Cu transporters in bacteria was observed in *Acidovorax sp.* JS42 and *Ralstonia pickettii* (10 and 9 Cu exporters, respectively), both of which were isolated from highly contaminated environments [96]. It is possible that these species need more efficient mechanisms to maintain cellular Cu homeostasis or protect against this metal.

Further studies on Cu utilization in eukaryotes revealed that almost all sequenced organisms utilized Cu, suggesting a uniformly essential nature of this metal in this domain of life (Figure 5a). In eukaryotes, the occurrence of Cu importer Ctr1 and exporter ATP7 was consistent with that of the Cu utilization trait (Ctr1 was detected in more than 90% of Cu-utilizing organisms and ATP7 in all Cu-utilizing organisms). Interestingly, the majority of organisms had 1–3 *ctr1* genes, but *Caenorhabditis* species (nematodes) possessed additional *ctr1* genes, especially *C. elegans* that had 11 such genes, suggesting unknown complexities in Cu uptake and trafficking in these organisms [96]. It is possible that these Ctr1 proteins are located in various membranes (i.e., plasma or organellar membrane) and/or cell types. The occurrence of Cu exporters varied from one to six genes, with three *Phytophthora* species, which are crop plant pathogens belonging to the genus *Oomycetes*, having the highest numbers of Cu exporters (i.e., *Phytophthora infestans* possessed six ATP7 proteins).

Among Cu-dependent proteins, COX I and COX II were the most frequently used Cu-binding proteins in bacteria and archaea (Figure 5b). Other Cu-binding proteins, such as Cu-Zn SOD, plastocyanin, and a variety of MCOs were also found in many prokaryotes. In contrast, the occurrence of pMMO, nitrosocyanin, CuAO, and tyrosinase appeared to be very limited. In addition, some bacterial Cu-dependent protein families, including azurin, nitrosocyanin, Ndh2, pMMO, and tyrosinase were absent in archaea, whereas a blue Cu protein, rusticyanin, was only detected in archaea. Investigation of the cuproproteomes (the whole set of Cu-dependent proteins) suggested that large cuproproteomes were mainly observed in proteobacteria, especially in *Alphaproteobacteria/Rhizobiaceae* among which two *Sinorhizobium* species (*S. medicae* and *S. meliloti*) contained the largest bacterial cuproproteomes

(22 Cu-dependent proteins, half were COX I and COX II proteins). In archaea, large cuproproteomes were mainly found in *Euryarchaeota/Halobacteriales*, including *Haloarcula marismortui* that had the largest prokaryotic cuproproteome (25 Cu-dependent proteins; half are plastocyanin homologs). Thus, although bacteria and archaea have similar Cu-dependent protein families, occurrence of these proteins was mostly different [96].

Homologs of almost half of the prokaryotic Cu-dependent proteins could not be found in eukaryotes. On the other hand, novel Cu-binding proteins evolved in eukaryotes, such as plantacyanin, PHM, hemocyanin, and GAO (Figure 5c). Analysis of the occurrence of eukaryotic Cu-dependent proteins revealed that, similar to prokaryotes, MCOs, COX I, COX II, and Cu-Zn SOD were the most abundant Cu-dependent proteins. Further analysis of eukaryotic cuproproteomes showed that land plants possessed the largest cuproproteomes (62 and 78 proteins in *A. thaliana* and *Oryza sativa*, respectively [96]). Most of these proteins belonged to plantacyanin, CuAO, and MCO families.

One interesting finding was that organisms living in oxygen-rich environments utilized Cu, whereas the majority of anaerobic organisms did not [37,39]. In addition, among Cu users, cuproproteomes of aerobic organisms were generally larger than those of anaerobic organisms. These data are consistent with the idea that proteins evolved to utilize Cu following the oxygenation of the Earth [151]. In other words, the use of Cu is strongly linked to the use of molecular oxygen.

5 Nickel and Cobalt

Nickel is an essential component of several metalloenzymes involved in energy and nitrogen metabolism, whereas Co is mainly found in the corrin ring of vitamin B₁₂ (also known as cobalamin), a cofactor involved in methyl group transfer and rearrangement reactions, but also occurs in a few non-corrin Co-containing enzymes, such as methionine aminopeptidase from *Salmonella typhimurium* and prolidase from *Pyrococcus furiosus* [152,153]. The list of known Ni- and B₁₂-dependent proteins is shown in Table 3.

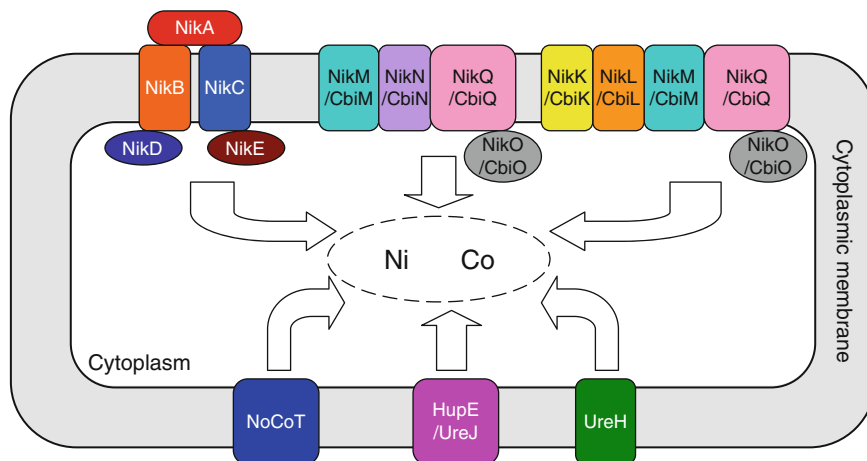
5.1 Nickel and Cobalt Uptake

In prokaryotes, these transition metals use similar transport systems. Thus, identification of substrate preference of members of each transporter family is important for comparative genomics of Ni and Co utilization. A schematic representation of known Ni/Co transport systems is shown in Figure 6.

In bacteria, Ni and Co uptake is mediated by ABC systems and several secondary transporters [13]. The well-studied ABC-type Ni transporter system, NikABCDE, belongs to a large family of ABC transporters (nickel/peptide/opine transporter

Table 3 Ni- and Co(B₁₂)-dependent proteins

| Ni-dependent proteins | Co(B ₁₂)-dependent proteins |
|--|--|
| Urease | Adenosylcobalamin-dependent isomerase family |
| Ni-Fe hydrogenase | Methylmalonyl-CoA mutase |
| Carbon monoxide dehydrogenase | Isobutyryl-CoA mutase |
| Acetyl-coenzyme A decarboxylase/synthase | Ethylmalonyl-CoA mutase |
| Superoxide dismutase SodN | Glutamate mutase |
| Methyl-coenzyme M reductase (using F ₄₃₀ as a cofactor) | Methyleneglutarate mutase |
| | D-lysine 5,6-aminomutase |
| | Diol dehydratase |
| | Glycerol dehydratase |
| | Ethanolamine ammonia lyase |
| | B ₁₂ -dependent ribonucleotide reductase |
| | Methylcobalamin-dependent methyltransferase family |
| | Methionine synthase (MetH) |
| | Other methyltransferases: Mta, Mtm, Mtb, Mtt, Mts, and Mtv |
| | B ₁₂ -dependent reductive dehalogenase family |
| | CprA |

**Figure 6** Schematic representation of Ni/Co transport systems. The Ni/Co transport systems include NikABCDE, Nik/CbiMNQO, Nik/CbiKMLQO, NiCoT, HupE/UreJ, and UreH.

family, PepT). This multi-component system is composed of a periplasmic Ni-binding protein (NikA), two integral membrane proteins (NikB and NikC) and two ABC proteins (NikD and NikE). NikA may also bind divalent Co, Cu, and Fe with at least 10-fold lower affinity [154]. In addition, NikA could bind heme in *E. coli*, suggesting an additional function independent of Ni transport [155]. To date, residues involved in Ni binding for NikA have not been well characterized and

conflicting results were reported by various groups. For example, it was suggested that *E. coli* NikA binds a natural metallophore containing three carboxylate functions that coordinate a Ni ion via four residues (Tyr402, Arg137, Arg97, and His416), and that His416 (the only direct metal-protein contact) of NikA is essential for Ni uptake in *E. coli* [156]. It was also reported that Ni binds *E. coli* NikA without chelators and is coordinated by two histidine residues (His56 and His442) at a position distant from the previously characterized binding site [157]. In any event, the presence of the majority of these residues could be used to help predict NikA orthologs from Ni-unrelated homologs. Distantly related Ni ABC transporters were also identified in *Yersinia* (named YntABCDE), highlighting diversity of Ni ABC-type transporters in bacteria.

An additional ABC-like transport system, Cbi/NikMNQO, is often encoded next to the B₁₂ biosynthesis or urease (a major Ni-dependent enzyme) genes in some bacterial genomes, and was shown to mediate Co and Ni uptake respectively, although the metal-binding ligands are unclear [158]. Comparison of the cbi/nikMNQO operon structures and occurrence of each component revealed that M, Q, and O gene products are universal components. Although the transmembrane proteins CbiN (Co uptake) and NikN (Ni uptake) have no significant homology, they might have the same topology with two transmembrane segments [36]. Two additional proteins, NikK and NikL, were also proposed to be involved in Ni uptake when NikN is absent, and form an alternative NikKMLQO system [36].

Secondary Ni/Co transporters include NiCoT (also designated HoxN, HupN, NicT, NixA or NhfF in different organisms), UreH, and HupE/UreJ [159]. NiCoTs are a family of proteins with eight transmembrane segments. They are widespread among bacteria and found in several thermoacidophilic archaea and certain fungi including *S. pombe* and *Neurospora crassa*. Various subtypes of NiCoTs have different ion preferences ranging from strict selectivity for Ni to unbiased transport of both ions to strong preference for Co. In many cases, the preference for a particular metal correlated with the genomic location of NiCoT genes, which are adjacent to the genes for Ni or Co (or B₁₂ biosynthesis) enzymes [19,160]. UreH and HupE/UreJ are putative secondary transporters, and certain members of these families have recently been shown to mediate Ni transport [159–161]. Homologs of UreH were also detected in plants. In addition, several new types of candidate Co transporters were predicted, such as CbtAB, CbtC, CbtD, CbtE, CbtF, CbtG, and CbtX [61,160]. The distribution of these candidate transporters is quite limited.

In eukaryotes, a subfamily of cation-efflux family members (TgMTP1) was found to account for the enhanced ability of Ni hyperaccumulation in higher plants [162]. To date, no high-affinity Co uptake system has been reported in eukaryotes; however, some suppressors of Co toxicity, such as COT1 and GRR1 in *S. cerevisiae*, were characterized, which may play an important role in metal homeostasis by decreasing the cytoplasmic concentration of metal ions including Co and Zn [163].

In many bacteria including *E. coli*, a Ni repressor gene, *nikR*, is located immediately next to its target, the *nikABCDE* operon. NikR-dependent regulation was also predicted for some other Ni transporters, such as NikMNQO and NiCoT, as well as Ni-dependent enzymes such as Ni-Fe hydrogenase [36,164]. These NikRs regulate

the transcription of target genes in response to Ni ion concentrations, utilizing a combination of allostery and coordination geometry. The presence of a NikR-binding site that contains an inverted repeat sequence and is always located upstream of Ni-associated genes may help identify NikR-related regulation [36].

5.2 Nickel-Dependent Proteins

The characterization of Ni in several enzymes has created an active field exploring the biochemistry of this metal (see Table 3). In prokaryotes, the major strictly Ni-dependent enzymes include urease, Ni-Fe hydrogenase, carbon monoxide dehydrogenase (CODH), acetyl-coenzyme A decarbonylase/synthase (CODH/ACS), superoxide dismutase SodN, and methyl-coenzyme M reductase (MCR). In addition, some homologs of Ni-binding proteins appear to bind other metals. For example, glyoxalase I (GlxI) binds Ni in *E. coli*, *P. aeruginosa*, and human parasites *Leishmania* (e.g., *L. major*) and *Trypanosoma* (e.g., *T. cruzi*), but it binds Zn in *P. putida*, yeast, and human [30,165]. Thus, such proteins could not be used for comparative genomics of Ni utilization because of the uncertainty of the metals they bind in different organisms. In eukaryotes, urease is the only known Ni-dependent enzyme. Additional candidate Ni-containing compounds or proteins have also been described in different organisms [166].

Urease is the first characterized Ni-containing protein that has been found in bacteria, fungi, and plants. It catalyzes the hydrolysis of urea to carbon dioxide and ammonia. In plants, urease is a hexamer of identical chains; whereas in bacteria, it consists of either two or three different subunits [167]. The Ni active site appears to be particularly conserved, as two Ni atoms are associated with each active site of the respective enzymes based on the crystal structures [168].

Hydrogenase catalyzes the reversible reaction of dihydrogen. Based on the metal content and subunit composition of the enzymes, three classes of hydrogenases have been identified: Fe-Fe hydrogenase, Ni-Fe hydrogenase, and hydrogenases that use neither Fe nor Ni [169]. The most studied class comprises the Ni-Fe hydrogenases, which are mainly utilized for hydrogen oxidation. Crystal structures of several Ni-Fe hydrogenases have been identified [170]. One class is composed of two subunits: the large subunit contains the Ni-Fe active site, and the small subunit that contains an Fe-S cluster is used in electron transfer from the large subunit. Other Ni-Fe hydrogenases are tetramers and integral membrane proteins. Two motifs have been proposed to be involved in the ligation of Ni: the N-terminal RxCGxC and the C-terminal DPCxxC. Similar motifs have been found in the sub-class of Ni-Fe-Se hydrogenases which contain a selenocysteine instead of a Cys bound to the Ni atom [171].

Ni-containing CODHs are the biological catalysts for reversible oxidation of CO to CO₂, with water as the source of oxygen. Members of the CODH family have been characterized from archaea and bacteria. The active site of CODH, designated cluster C, is a complex Ni-, Fe-, and S-containing metal center [172]. The recently published high-resolution structure of CODH from *Carboxydotherrmus*

hydrogenoformans in three states demonstrated the mechanism of CO oxidation and CO₂ reduction at the Ni-Fe site of cluster C [173].

CODHs in acetogenic bacteria (anaerobes that can grow autotrophically on the greenhouse gas CO₂) and methanogenic archaea are bifunctional enzymes that perform both the reversible CO-oxidation reaction and the synthesis or degradation of acetyl-coenzyme A (CoA) and are therefore designated CODH/ACS. Both catalytic sites for the individual reactions require Ni for catalysis and are positioned at different sites [174].

MCR is responsible for all biologically produced methane on earth, which catalyzes the final step in the biological synthesis of methane in methanogenic archaea. In contrast to other Ni-dependent proteins, this enzyme contains Ni in a tetrapyrrolic structure known as coenzyme F₄₃₀, which is found exclusively in methanogens [175]. MCR homologs in some uncultured methanotrophic archaea are involved in anaerobic oxidation of methane in marine sediments. Differences between the highly similar structures of these MCR homologs and methanogenic MCR include a F₄₃₀ modification, a Cys-rich patch and an altered post-translational amino acid modification pattern, which may tune the enzymes for their functions in different biological contexts [176].

SODs are important antioxidant enzymes protecting against superoxide toxicity. Various SODs have been characterized that use Fe/Mn, Cu-Zn (see Sections 4.2), or Ni cofactors to carry out the disproportionation of superoxide. The Ni-containing SOD is a product of the *sodN* gene, which encodes a protein with an N-terminal extension that is removed in the mature enzyme. The crystal structure of the active Ni-bound enzyme from *Streptomyces coelicolor* identified a novel SOD fold and the Ni active site. A nine-residue structural motif (His-Cys-X-X-Pro-Cys-Gly-X-Tyr) provides almost all interactions essential for metal binding and catalysis, and thus may be diagnostic of other SodNs [177].

5.3 Cobalt-Dependent Proteins

Although Co is less frequently encountered in metalloenzymes than the other first-row transition metals (e.g., Fe, Cu and Zn), it is nevertheless an essential cofactor in vitamin B₁₂-dependent enzymes. Vitamin B₁₂, also known as cobalamin, is a group of closely related polypyrrole compounds such as cyanocobalamin, methylcobalamin, and deoxyadenosyl cobalamin. They are required for the metabolism of many prokaryotic and eukaryotic organisms.

5.3.1 Vitamin B₁₂ Uptake and Biosynthesis

Vitamin B₁₂ uptake is critical for B₁₂-utilizing organisms that cannot synthesize the coenzyme *de novo*. To date, the only known transport system for B₁₂ in prokaryotes is the BtuFCD system, which includes a periplasmic B₁₂-binding protein BtuF and

two ABC transport subunits BtuC and BtuD [178]. The BtuFCD system belongs to a large superfamily involved in the uptake of Fe, siderophores, and heme. In Gram-negative bacteria, a TonB-dependent outer membrane receptor BtuB is also involved in B₁₂ uptake and forms a complex with BtuFCD [179]. Mammals have developed a complex system for internalization of this vitamin from the diet. Three binding proteins (haptocorrin, intrinsic factor, transcobalamin) and several specific receptors are involved in the process of intestinal absorption, plasma transport, and cellular uptake [180]. However, the mechanism of B₁₂ uptake in other eukaryotes, such as algae, is unclear, although many algae are rich in vitamin B₁₂. It was suggested that algae acquire vitamin B₁₂ through a symbiotic relationship with bacteria [181].

In microorganisms that synthesize vitamin B₁₂, it is produced via two alternative routes: oxygen-dependent (aerobic, or “late Co insertion”) and oxygen-independent (anaerobic, or “early Co insertion”) pathways that differ mainly in the early stages (Figure 7). The aerobic pathway incorporates molecular oxygen into the macrocycle as a prerequisite to ring contraction. The intermediates of the aerobic route from uroporphyrinogen III (uro’gen III) to adenosylcobalamin and more than 20 genes involved in these processes (*cobA-cobW*) have been identified in several bacteria such as *P. denitrificans*. The anaerobic pathway, which was partially resolved in some organisms, such as *S. typhimurium* and *Bacillus megaterium*, takes advantage of the chelated Co ion, in the absence of oxygen, to support ring contraction. It has been suggested that the anaerobic and aerobic routes contain several pathway-specific enzymes [160]. For example, CbiD, CbiG, and CbiK appear to be specific to the anaerobic route of *S. typhimurium*, whereas CobE, CobF, CobG, CobN, CobS, CobT, and CobW are unique to the aerobic pathway of *P. denitrificans*. Besides, an adenosyltransferase that catalyzes the final step in the assimilation of vitamin B₁₂ was found to directly transfer the cofactor to a B₁₂-dependent methylmalonyl-CoA mutase in *Methylobacterium extorquens*, suggesting that the strategy of using the final enzyme in an assimilation pathway for tailoring a cofactor and delivering it to a dependent enzyme may also be general for cofactor trafficking [182]. Recently, it was reported that this process is gated by a small G protein, MeaB. While the GTP-binding energy is needed for the editing function; that is, to discriminate between active and inactive cofactor forms, the chemical energy of GTP hydrolysis is required for gating cofactor transfer [183].

5.3.2 Vitamin B₁₂-Dependent Proteins

Considering that vitamin B₁₂ is the major form of Co utilization and that B₁₂-binding proteins are strictly dependent on this cofactor, identification of all B₁₂-dependent enzymes is extremely important for comparative genomics of Co utilization. To date, vitamin B₁₂ is mainly present in three classes of enzymes in prokaryotes (classified based on different chemical features of the cofactor): adenosylcobalamin-dependent isomerase, methylcobalamin-dependent methyltransferase, and B₁₂-dependent reductive dehalogenase [184,185]. These classes can be further divided into subclasses based on sequence similarity and reactions they catalyze (see Table 3).

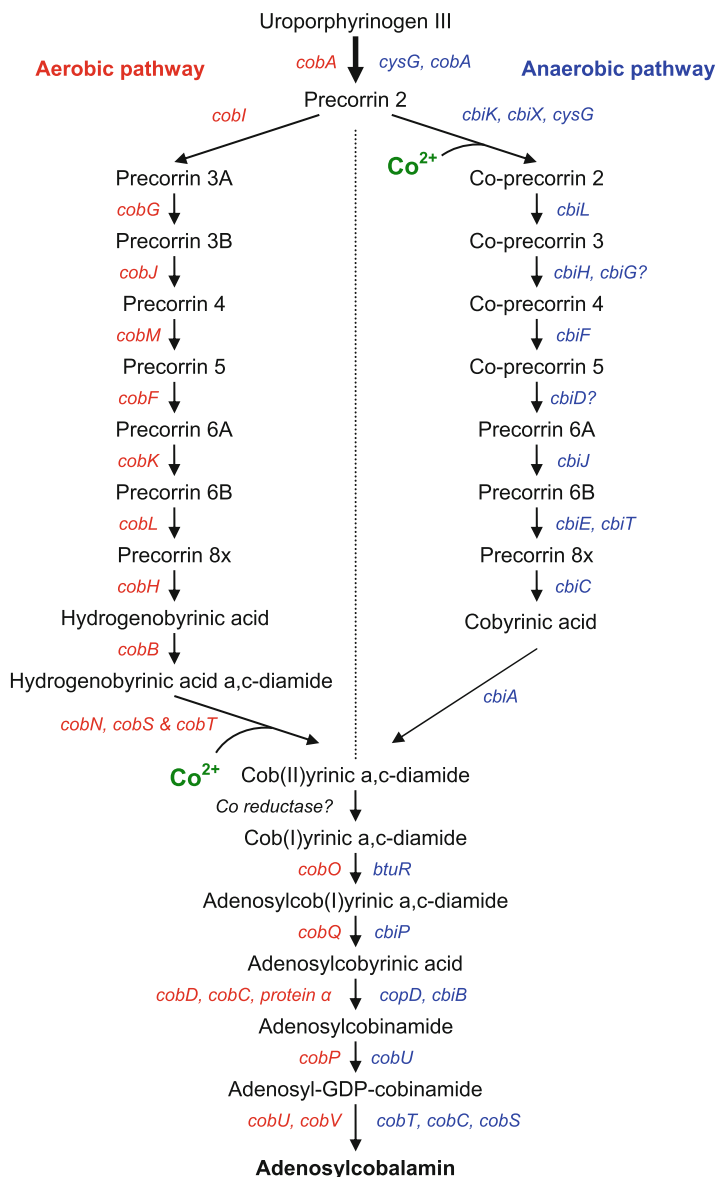


Figure 7 Biosynthetic pathways for vitamin B₁₂ in bacteria. Genes involved in aerobic and anaerobic pathways are shown in red and blue, respectively.

Adenosylcobalamin-dependent isomerases are the largest family of B₁₂-dependent enzymes and are mainly found in bacteria, where they catalyze a variety of chemically difficult 1,2-rearrangements that proceed through a mechanism involving free radical intermediates [186]. Subclasses of this family include methylmalonyl-CoA

mutase (MCM), isobutyryl-CoA mutase (ICM), ethylmalonyl-CoA mutase (ECM), glutamate mutase (GM), methyleneglutarate mutase (MGM), D-lysine 5,6-aminomutase (5,6-LAM), diol/glycerol dehydratase (DDH/GDH), ethanolamine ammonia lyase (EAL), and B₁₂-dependent ribonucleotide reductase (RNR II).

MCM is the only B₁₂-dependent isomerase that is present in both bacteria and animals. It catalyzes the isomerization of methylmalonyl-CoA to succinyl-CoA in the pathway that converts catabolites of odd-chain fatty acids, branched-chain amino acids, and cholesterol to a key intermediary metabolite [187]. In many organisms, such as *S. cinnamomensis*, it consists of two subunits, MutA and MutB, which show high sequence similarities to MCMs from other bacteria and mammals [188]. The crystal structure of MCM from *Propionibacterium shermanii* revealed the coordination of Co in coenzyme B₁₂ by the histidine in the DXHXXG motif within the C-terminal cobalamin-binding domain [189].

ICM and ECM are homologs of MCM with different functions. ICM catalyzes the reversible rearrangement of isobutyryl-CoA to n-butyryl-CoA. It has been detected in a variety of aerobic and anaerobic bacteria, where it appears to play a key role in valine and fatty acid catabolism as well as in the production of fatty acid-CoA thioester building blocks for polyketide antibiotic biosynthesis. In *S. cinnamomensis*, this mutase was found to comprise a large subunit of IcmA and a small subunit IcmB [190]. IcmB shows high sequence similarity to the cobalamin-binding domains of other B₁₂-containing enzymes such as B₁₂-dependent methionine synthase, including the conserved DXHXXG cobalamin-binding motif, suggesting that IcmB has taken on the role of a separate cobalamin-binding domain in ICM. ECM is involved in the central reaction of the ethylmalonyl-CoA pathway and catalyzes the transformation of ethylmalonyl-CoA to methylsuccinyl-CoA in combination with a second enzyme that was identified as promiscuous ethylmalonyl-CoA/methylmalonyl-CoA epimerase. Although ECM showed significant sequence similarity to MCM and ICM from the same organism, sequence analysis revealed that this enzyme is distinct from MCM as well as ICM, and defines a new subfamily of coenzyme B₁₂-dependent acyl-CoA mutases [191].

B₁₂-dependent GM catalyzes a most unusual carbon skeleton rearrangement involving the isomerization of L-glutamate to L-threo-methylaspartate, a reaction that is without precedent in organic chemistry. The active enzyme consists of two subunits (designated GImE and GImS in *Clostridium cochlearium*) as an $\alpha 2\beta 2$ tetramer, whose assembly is mediated by coenzyme B₁₂. The smaller of the protein components, GImS, is similar to the B₁₂-binding domain of MCM and has been shown to be the B₁₂-binding subunit [192].

B₁₂-dependent MGM from the strict anaerobe *Eubacterium barkeri* catalyzes the equilibration of 2-methyleneglutarate with (*R*)-3-methylitaconate. This enzyme also contains the highly conserved DXHXXG(X)(41)GG motif, which is critical for B₁₂ binding [193].

5,6-LAM is an adenosylcobalamin and pyridoxal-5'-phosphate-dependent enzyme that catalyzes a 1,2 rearrangement of the terminal amino group of D-lysine and of L- β -lysine. The crystal structure of a substrate-free form of 5,6-LAM from *C. sticklandii* revealed that a Rossmann domain covalently binds pyridoxal-5'-phosphate and positions it into the putative active site of a neighboring triosephosphate

isomerase barrel domain, while simultaneously positioning the other cofactor, adenosylcobalamin, approximately 25 Å from the active site [194]. Thus, this structure features a locking mechanism to keep the adenosylcobalamin out of the active site and prevent radical generation in the absence of substrate.

B₁₂-dependent GDH and DDH are homologous isofunctional enzymes that catalyze the elimination of water from glycerol and 1,2-propanediol to the corresponding aldehyde via a B₁₂-dependent radical mechanism. The crystal structure of the substrate-free form of GDH in complex with cobalamin has been determined, whose overall fold and subunit assembly closely resemble those of DDH. Structural analysis of the locations of conserved residues among various GDH and DDH sequences helped identify residues important for substrate preference and specificity of protein-protein interactions [195].

EAL catalyzes the deamination of ethanolamine and 2-aminopropanol. Computational modeling of EAL from *S. typhimurium* revealed that this enzyme may have a similar TIM-barrel fold as DDH and GDH [196].

RNR catalyzes a rate-limiting reaction in DNA synthesis by converting ribonucleotides to deoxyribonucleotides. To date, three major classes of RNR have been discovered that depend on different metal cofactors for the catalytic activity: class I RNRs contain a diiron-oxygen cluster, class II contain vitamin B₁₂, and class III use an Fe-S cluster coupled to S-adenosylmethionine [197]. RNR II enzymes are mainly found in bacteria, and also in some of their phages. They utilize an adenosylcobalamin cofactor that interacts directly with an active Cys residue to form the reactive Cys radical needed for ribonucleotide reduction [198].

The methylcobalamin-dependent methyltransferases play important roles in amino acid metabolism in a variety of organisms, including mammals, as well as in carbon metabolism and CO₂ fixation in anaerobic microbes. These methyltransferases could be divided into two subclasses: one subclass binds simple substrates such as methanol (MtaB), methylated amines (MttB, MtbB, MtmB), methylated thiols (MtsB), methoxylated aromatics (MtvB), and methylated heavy metals, while the other, such as methionine synthase (MetH), catalyzes methyl transfer from methyltetrahydrofolate and the methanogenic analog methyltetrahydromethanopterin [184].

Methionine synthase (MetH) catalyzes the transfer of a methyl group from N₅-methyltetrahydrofolate to homocysteine, producing tetrahydrofolate and methionine [199]. This enzyme is the most extensively studied B₁₂-dependent methyltransferase, which is widespread in all three domains of life. It is a modular enzyme with distinct regions for binding homocysteine, methyltetrahydrofolate, B₁₂, and adenosylmethionine. The B₁₂ domain in its different oxidation states may interact with each of the other three domains. The crystal structure of a B₁₂-containing fragment of MetH from *E. coli*, which was the first structure of a protein-bound B₁₂, revealed that the histidine residue in the DXHXXG motif is the Co ligand and is part of a catalytic quartet, Co-His759-Asp757-Ser810, that modulates the reactivity of the B₁₂ prosthetic group in MetH [200].

Other B₁₂-dependent methyltransferases are designated as Mtx, where x denotes the methyl donor (e.g., a, methanol; v, vanillate; m, methylamine; b, dimethylamine; t, trimethylamine; and s, dimethylsulfide). These methyltransferases consist of three components (Mt_A, Mt_B, and Mt_C) [184]. Each component is found on a

different polypeptide or domain. Mt_A methylates coenzyme M (CoM), Mt_B methylates the corrinoid protein, and Mt_C is the corrinoid protein containing B₁₂. These methyltransferases are essential in energy metabolism and in cell carbon synthesis in anaerobic microbes such as methanogenic archaea and acetogenic bacteria [201]. In addition, methyltetrahydromethanopterin:CoM methyltransferase (Mtr), which contains eight subunits (MtrA-H), was found to utilize a histidine as the ligand to the cobalamin in MtrA [202].

B₁₂-dependent reductive dehalogenases CprA play an important role in the detoxification of aromatic and aliphatic chlorinated organics in anaerobic microbes. Most of these enzymes also contain Fe-S clusters [203]. The role of B₁₂ in CprA appears to be significantly different from those of the B₁₂-dependent isomerases and methyltransferases. However, the reaction mechanism of dehalogenases remains unclear.

In eukaryotes, only three B₁₂-dependent enzymes, MetH, MCM, and RNR II, have been identified, implying that Co utilization is quite restricted in this domain of life.

5.3.3 Non-Corrin Cobalt-Binding Proteins

A few proteins containing non-corrin-Co have been reported in different organisms, including methionine aminopeptidase (from *S. typhimurium*), prolidase (from *P. furiosus*), nitrile hydratase (from *Rhodococcus rhodochrous*), methylmalonyl-CoA carboxytransferase (from *P. shermanii*), aldehyde decarbonylase (from *Botryococcus braunii*), glucose isomerase (from *S. albus*), and several other proteins [153]. However, all of these enzymes are not strictly Co-dependent and may use other metals (such as Fe, Zn, and Mn) in place of Co. Thus, it is difficult to identify the metal specificity of these enzymes by computational analysis. To date, only nitrile hydratase (NHase) was suggested to have different active site motifs for Co- and Fe-binding forms [204].

5.4 Comparative Genomics of Nickel, Cobalt, and Vitamin B₁₂ Utilization

As mentioned above, Ni and Co are essential cofactors in several enzymes. Ni is used in several metalloenzymes involved in energy and nitrogen metabolism, detoxification processes, pathogenesis, enzyme inactivation, and lipid peroxidation, whereas Co is primarily found in the corrin ring of coenzyme B₁₂ that plays important roles in several biological systems. In recent years, several comparative genomics studies have been carried out to investigate Ni and Co utilization traits.

An early study examined Ni and Co transport systems in about 200 microbial genomes and demonstrated a complex and mosaic utilization of both metals in prokaryotes [36]. Two computational approaches were used for functional prediction of proteins involved in Ni or Co uptake: (i) analysis of the genomic locations of

genes encoding Ni/Co transporters; and (ii) identification of regulatory signals, such as NikR-dependent regulation through the NikR-binding signal, and B₁₂ riboswitches that regulate many of the candidate Co transporters in bacteria. This study revealed that the Ni/Co transporter genes are often colocalized with the genes for Ni-dependent and B₁₂ biosynthesis proteins. Different families of Ni/Co transporters showed a mosaic distribution in those organisms, and the Cbi/NikMNQO system (including the Cbi/NikKMLQO system) appeared to be the most widespread group of microbial transporters for the two metal ions.

A separate computational analysis of B₁₂ metabolism and regulation also provided important information regarding B₁₂ utilization in prokaryotes [160]. Using comparative analysis of gene regulation, positional clustering of genes, and phylogenetic profiling, the B₁₂ biosynthesis and regulation was described in a variety of prokaryotes. The B₁₂ riboswitch was found to be widely distributed in the regions upstream of B₁₂ biosynthetic and transport genes. In addition, by searching for candidate B₁₂-regulated genes, several new types of candidate Co transporters and new proteins associated with the B₁₂ biosynthesis pathway, such as certain chelatas and methyltransferases, were identified. The B₁₂ transporters, BtuFCD, appeared to be widely distributed in prokaryotes and some of them were B₁₂-regulated. However, it is difficult to selectively identify BtuFCDs among other highly similar transport systems (such as Fe/heme or siderophore transporters) in the majority of sequenced organisms. Furthermore, the B₁₂ element was predicted to regulate B₁₂-independent MetH and RNR isozymes in bacteria that also have corresponding B₁₂-dependent isozymes.

Recently, two much more extensive comparative analyses involving more than 700 organisms in all three domains of life have been carried out, which have provided additional important information regarding Ni and Co utilization traits [96,205]. Only strictly Ni-dependent metalloproteins and B₁₂-binding enzymes were used for comparative genomics of Ni and Co, respectively. Occurrence of the Ni/Co-utilization trait and Ni- or B₁₂-dependent proteins is shown in Figure 8. The distribution and dynamics of Ni and Co utilization were analyzed at the level of both transporters and metalloproteomes. These analyses revealed that both metals are widely used by prokaryotes; however, analyses of occurrence of Ni/Co transporters and metalloenzymes showed great diversity among bacteria and archaea. Urease and B₁₂-dependent MetH were the most widespread Ni- and Co-containing proteins, respectively, in bacteria. In contrast, Ni-Fe hydrogenase and RNR II were the most widespread Ni and Co users in archaea where urease and MetH were very rare or even absent. Further analyses of Ni- or Co-dependent metalloproteomes revealed that, except for deltaproteobacteria and several *Methanosarcina* species, most prokaryotes contained small Ni- and Co-dependent metalloproteomes (1–4 proteins). The largest Ni-dependent metalloproteome was observed in *Deltaproteobacterium* MLMS-1 (16 Ni-binding proteins) and the largest B₁₂-dependent metalloproteome in *Dehalococcoides* sp. CBDB1 (35 B₁₂-binding proteins). Further analysis of Ni and Co utilization based on different habitats, environments, and other factors revealed that, similar to Mo utilization, host-associated organisms (particularly obligate intracellular parasites and endosymbionts) have a tendency for reduced Ni or Co utilization.

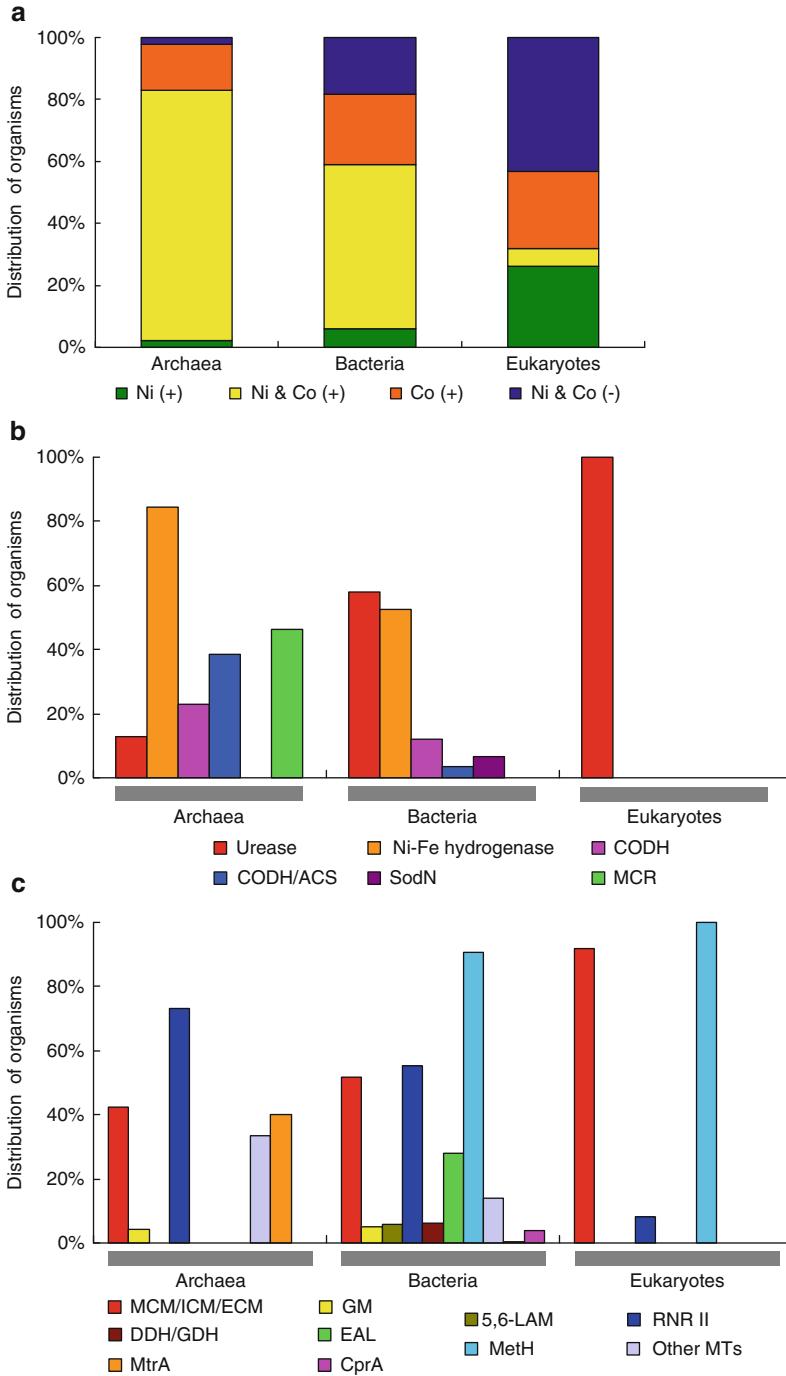


Figure 8 Occurrence of Ni and Co utilization in the three domains of life. **(a)** Distribution of Ni-/Co-utilizing organisms among those with completely sequenced genomes. All organisms were classified into four groups: Ni (+), i.e., containing the Ni utilization trait only; Ni & Co (+), i.e.,

Investigation of Ni and Co utilization in eukaryotes provided a somewhat different Ni and Co utilization trend. Indeed, the use of these two metals is much more restricted in eukaryotes, with regard to both the organisms that utilize Ni/Co and the number of Ni transporters and Ni/B₁₂-dependent protein families. Surprisingly, very few of these organisms utilize both metals. The Ni-utilizing eukaryotes are mostly fungi (except saccharomycotina) and plants, whereas most B₁₂-utilizing organisms are animals. The NiCoT transporter family is the most widely used eukaryotic Ni transporter. Urease and MetH were the most common eukaryotic Ni- and B₁₂-dependent enzymes, respectively. Analysis of Ni- and Co-dependent metalloproteomes in eukaryotes did not reveal organisms that contained many such proteins. Only single copies of urease and 1–3 B₁₂-dependent proteins were detected in various organisms. In contrast to the majority of unicellular organisms that lack B₁₂ utilization, *Dictyostelium discoideum* and *Phytophthora* species contained all three known eukaryotic B₁₂-dependent proteins: MetH, MCM, and RNR II.

6 Comparative Genomics of Other Metals

Besides the metals discussed above, comparative genomics studies have also been carried out for additional metals, such as Zn and Fe, but widespread use of these elements across living organisms makes comparative analyses more challenging. In the following two sections, we briefly introduce recent advances on these metals.

6.1 Comparative Genomics of Zinc-Dependent Metalloproteomes

Zn is thought to be essential for all organisms and was suggested to be a key element in the origin of life [206]. It is found in a great variety of enzymes, structural proteins, transcription factors, and ribosomal proteins. We refer to other articles for review on Zn uptake, storage, homeostasis, and user proteins [6,15,20,25,29] and only focus here on comparative genomics of Zn-dependent metalloproteomes.

←

Figure 8 (continued) containing Ni and Co utilization traits; Co (+), i.e., containing the Co utilization trait only; Ni & Co (-), i.e., containing neither Ni nor Co utilization traits. (b) Distribution of organisms containing different Ni-dependent proteins in Ni-utilizing organisms. (c) Occurrence of B₁₂-dependent proteins in Co-utilizing organisms. CODH, carbon monoxide dehydrogenase; CODH/ACS, acetyl-coenzyme A decarbonylase/synthase; SodN, Ni-containing superoxide dismutase; MCR, methyl-coenzyme M reductase; MCM, methylmalonyl-CoA mutase; ICM, isobutyryl-CoA mutase; ECM, ethylmalonyl-CoA mutase. The latter three subfamilies are quite similar and are combined into one group. GM, glutamate mutase; 5,6-LAM, D-lysine 5,6-aminomutase; RNR II, B₁₂-dependent ribonucleotide reductase; DDH/GDH, diol/glycerol dehydratase; EAL, ethanolamine ammonia lyase; MetH, methionine synthase; Other MTs, various B₁₂-dependent methyltransferases such as Mta, Mtm, Mtb, Mtt, Mts, and Mtv systems; MtrA, methyltetrahydro-methanopterin:CoM methyltransferase subunit A; CprA, B₁₂-dependent reductive dehalogenase.

An early computational search was carried out for Zn proteomes in representative organisms from the three domains of life including humans [34]. A list of known Zn-binding protein domains and of known Zn-binding sequence motifs (Zn-binding patterns) were compiled and then used jointly to analyze the proteome of 57 different organisms (40 bacteria, 12 archaea, and 5 eukaryotes) to obtain an overview of Zn usage by prokaryotic and eukaryotic organisms. It was found that Zn proteins are widespread in living organisms. Within each domain of life, the number of Zn-containing proteins in an organism correlates with the proteome size. Prokaryotes, on average, have a lower proportion of Zn proteins ($6.0\% \pm 0.2\%$ of the entire proteome in archaea and $4.9\% \pm 0.1\%$ in bacteria) than eukaryotic organisms. Interestingly, it was observed that the proteome of the hyperthermophilic prokaryotes is enriched in putative Zn-binding proteins, which may be due to an increased use of Zn to enhance the structural stability of proteins by organisms living at higher temperatures. Eukaryotic proteomes are much richer in putative Zn-binding proteins. On average, the Zn-proteome constitutes $8.8\% \pm 0.4\%$ of the eukaryotic proteome (about 10% in humans). Approximately two-thirds of prokaryotic Zn proteins have homologs in eukaryotes. On the other hand, three-quarters of eukaryotic Zn proteomes comprise proteins encoded only in eukaryotes, suggesting that they are relatively more recent.

There is also a functional diversification of eukaryotic and prokaryotic Zn proteomes. Prokaryotes use Zn proteins mostly to perform enzymatic catalysis, whereas in eukaryotes the Zn proteome is almost equally involved in performing catalysis and in regulating DNA transcription. Such a broad difference in function has a correspondence with the organization of the Zn-binding patterns. The Zn-binding patterns containing four ligands are associated with structural sites where Zn contributes to the stability of the protein structure, whereas the patterns containing three ligands are associated with catalytic sites (the fourth ligand is often water) where Zn actively participates in the reaction mechanism of the enzyme [207]. In addition, identity of the amino acids in the pattern is also quite different among structural and catalytic sites. For example, approximately 2800 Zn-binding proteins were found in humans, 97% having a structural Zn site with at least one Cys ligand, and 40% having four Cys ligands [208]. On the other hand, nearly one-third of human proteins with a three-ligand Zn-binding pattern have a pattern with three histidines. Together, four- and three-ligand patterns account for approximately 96% of all human Zn proteins.

As noted above, the majority of known Zn-dependent enzymes could be detected in both prokaryotes and eukaryotes, implying that Zn has been exploited in the catalytic site of enzymes prior to the split of the three domains of life. On the other hand, Zn-binding transcription factors are almost exclusively a prerogative of eukaryotes. These proteins normally contain Zn-finger domains, which are much rarer in bacteria and archaea. This observation suggests that Zn-binding transcription factors have evolved to meet the need of higher organisms to regulate more complex processes such as cell compartmentalization and differentiation. In addition, transcription factors bind Zn in very similar sites, most often composed by cysteine and histidine and organized in the same structure. The conservation of Zn-finger binding sites could be associated with their more recent origin, whereas the differentiation

of the catalytic Zn-binding sites could be the result of evolutionary processes that resulted in the development of different enzymatic reactions targeting different physiological substrates [209]. Using a similar approach, a recent comparative study on Zn-finger proteins and Zn hydrolytic enzymes in 821 organisms from the three domains of life revealed that there is a correlation in the changes during evolution related to environment [210].

6.2 *Advances in Comparative Genomics of Other Metals*

Iron is essential for life and is the most abundant transition metal ion in living organisms. In cells, Fe is normally found in the +2 and/or +3 oxidation states. Besides Fe ions, proteins can bind a range of Fe-containing cofactors, such as heme or Fe-S clusters. However, some of their close analogs may bind other metals in certain organisms or under certain conditions. For example, an acidophilic archaeon, *Ferroplasma acidiphilum*, was recently identified to possess a very large number of Fe-binding proteins (~86% of the entire protein repertoire), including many metalloproteins that bind different metals (such as Zn and Mn) in other organisms and proteins that are not even known to bind metal [211]. On the other hand, organisms that survive under Fe limitation have also been reported although it is unclear if they do not use this metal under Fe-sufficient conditions [212,213].

Although it is difficult to study the complete Fe-dependent metalloproteomes through computational approaches, a preliminary comparative genomics study has been carried out for the proteome-level analyses of the occurrence of non-heme Fe proteins in a selected number of prokaryotes and eukaryotes [35]. In this work, a similar bioinformatic approach based on the use of non-heme Fe-binding patterns in combination with the analysis of the occurrence of protein domains known to bind non-heme Fe was applied. In contrast to what was observed for Zn, the Fe proteome constitutes a higher fraction of the proteome in archaea (on average 7.1% \pm 2.1%) than in bacteria (3.9% \pm 1.6%) and eukaryotes (1.1% \pm 0.4%). The majority of these proteins have homologs in all three domains of life (~90% of the total), suggesting that extant organisms have inherited the large majority of their Fe proteome from the last common ancestor. The majority of non-heme Fe sites are found in proteins involved in electron transfer or in enzymes performing oxidoreductase functions, which is consistent with the fact that Fe is the most used metal ion in redox catalysis [214]. In addition, Fe is the metal ion with the largest variety of binding sites in proteins, including several types of Fe-S clusters and heme cofactors. This may be due to the necessity to use different chemical environments to modulate the reduction potential of Fe and thus its reactivity. Fe-S clusters are the cofactors of about 40% of non-heme Fe proteins retrieved, and their binding patterns are most often composed of cysteine residues. It is worth noting that cysteine is conversely an uncommon ligand for all the other non-heme Fe sites, where histidine is the most widespread ligand [209].

Heme is the prosthetic group of many proteins that carry out a variety of key biological functions, including oxygen transport and sensing, enzyme catalysis, and electron transfer. The utilization of heme requires a relatively complex machinery for its biosynthesis, insertion into heme-containing proteins, and uptake from external sources. The ability to bind this cofactor is strongly influenced by the interaction of the polypeptide chain with the porphyrin moiety. A recent genome-based analysis of proteins specifically involved in the processes of heme biosynthesis and uptake in 474 prokaryotes revealed that different systems exist in organisms belonging to various branches of the tree of life [215]. Some organisms (14%) presumably cannot perform either of the two processes (14%), some (40%) can perform only one of them, and some (46%) can perform both. Among these organisms, many Gram-positive pathogens support heme uptake from the host, suggesting that this process can be a potential target for wide-spectrum antibiotics. Further inspection of the sequences and structural models for two key domains in the heme uptake pathway suggested that there are possible alternative modes of heme binding. In the future, it would be useful to use computational and comparative approaches for the analysis of the variability of additional heme-binding modes and heme-dependent proteomes in organisms from the three domains of life.

The methods described in this chapter may be, in theory, applied to the study of other metals, such as Mn and Cr, and to the characterization of the corresponding metalloproteins. However, so far it has been difficult to identify complete metalloproteomes for most such metals because of either the limited knowledge of their metabolism or unspecificity/uncertainty of metal-binding ligands. A recent study that was based on liquid chromatography, high-throughput tandem mass spectrometry, and inductively coupled plasma mass spectrometry analyses showed that metalloproteomes are much more extensive and diverse than previously recognized [216]. Based on this study, a computational approach was also developed to predict a number of candidate novel metalloproteins for different metals [217]. Further efforts are needed for experimental verification of these proteins as well as identification of additional metal-binding proteins and metal-binding motifs/patterns.

7 Comparative Genomics of Metal Dependency in Biology

From the many studies on the functions of metals, metalloproteins emerged as one of the most diverse sets of proteins [218]. Some protein families are strictly dependent on certain metals (e.g., Cu in COX I, Ni in urease, and Mo in DMSOR), whereas other families have both metal-dependent and metal-independent forms or evolved to use alternative metals (e.g., GlxI binds Ni in *E. coli* but Zn in humans) [30,165]. In addition, initial genome-wide studies identified a significant number of proteins that bind metals and suggested that information on the occurrence of metal-dependent/metal-independent members of protein families may help better understand the utilization of these micronutrients [219,220]. These observations highlight complex evolutionary dynamics of the dependence of proteins on metals. In this section, we

focus on recent advances in comparative studies of metals, discuss metalloprotein families containing metal-dependent and metal-independent forms, and explore the evolutionary dynamics of the metal dependence of these families.

Comparative genomics studies on Zn metalloproteins have been discussed in Section 6.1. Taking advantage of the abundance of Zn-binding protein families, their Zn dependence could be systematically analyzed. An early comparative study was carried out to investigate the evolutionary history of ribosomal proteins that are present in all genomes and are generally well conserved [221]. Members of each examined ribosomal protein family were extracted from approximately 40 genomes (mostly bacteria). Several ribosomal proteins, such as S14, S18, L31-L33, and L36, all of which bind Zn via four conserved Cys or His residues, also have a Zn-independent form, in which these metal-binding residues have been partially or completely lost. In addition, genomes containing multiple copies of these ribosomal proteins encoded both Zn-dependent and Zn-independent forms. Further analyses revealed that, in most cases, a duplication of an ancestral Zn-dependent form had occurred early during evolution, with subsequent alternative loss of Zn-dependent and Zn-independent forms in different lineages. Another comparative genomics study that analyzed Zur (a repressor of Zn transport) regulation in bacteria has found that Zn repression was also detected in some paralogs of ribosomal protein genes (such as L36, L33, and S14) in which Zn-binding residues were disrupted [31]. This observation suggested a potential mechanism for maintaining Zn availability under Zn-restricted conditions, as these non-Zn-binding paralogs were expressed to partially replace the Zn-binding proteins, thereby freeing up some Zn for the Zn-binding proteins. Similar situations were also reported in another comparative study in which a subset of proteins in the diverse COG0523 family of putative metal chaperones were found to play a predominant role in the response to Zn limitation based on the presence of the corresponding COG0523-encoding genes downstream from putative Zur-binding sites in many bacterial genomes [222].

Very recently, all Zn-containing proteins with defined ligands in the PDB dataset have been analyzed [223]. Approximately 20% of the Zn protein families in the PDB had a significant number of Zn-independent forms, e.g., methionine-R-sulfoxide reductase, several tRNA synthetases, ribosomal proteins, and various subunits of DNA polymerase III. Further analysis of the predicted Zn dependence of a variety of Zn protein families in hundreds of sequenced bacterial genomes revealed that the majority of organisms containing these proteins possessed only single forms. Phylogenetic analyses suggested a role of both vertical inheritance and horizontal gene transfer in shaping the evolution of Zn protein families, which is consistent with previous findings [221]. Overall, all these results suggested a general picture of evolution of Zn utilization: many Zn protein families are strictly dependent on Zn, whereas some families yielded Zn-independent forms by disrupting ligands to Zn ion. These Zn-independent forms may have already been present in the ancestors of prokaryotes and eukaryotes and are widespread in currently living organisms.

As noted above, several comparative genomics studies have been carried out to analyze Cu utilization and to identify cuproproteins and cuproproteomes in organisms [37,96,148,150]. These studies also represent a new resource for studying the

Cu dependence of cuproprotein families. Besides Cu-dependent COX II and Cu-independent quinol oxidase subunit II (see Section 4.2), loss of Cu ligands (cysteine, histidine, and sometimes methionine) has been observed in members of several other cuproprotein families, which was mostly accompanied by changes in the function of these proteins.

As mentioned in Section 4.2, tyrosinases contain a type 3 Cu center and are distributed in all three domains of life. The active site of tyrosinase is characterized by a pair of antiferromagnetically coupled Cu ions (CuA and CuB), each coordinated by three histidine residues. In animals, two additional tyrosinase-related proteins (TRP1 and TRP2) that display significant homology to tyrosinase and that originated by the duplication of the ancestral tyrosinase gene were also detected. TRP1 is a 5,6-dihydroxyindole-2-carboxylic acid oxidase, and TRP2 is a dopachrome tautomerase, both of which are involved in melanin biogenesis [224]. TRP1 and TRP2 have the same six metal-binding histidine residues as tyrosinase. However, whereas TRP2 binds Zn, TRP1 binds an unknown metal that is not Cu, Zn, or Fe [225]. Thus, the specific binding of different metals by these proteins may be responsible for their distinct catalytic functions in melanogenesis. Although it is unclear how to distinguish Cu-dependent and Cu-independent forms in the tyrosinase family, an additional conserved histidine is located immediately upstream of one of three histidine residues in the CuB site in all examined tyrosinases, but it is replaced by Leu in all TRP1 and TRP2 proteins (Figure 9a). It has been shown that this histidine is essential for Cu binding in tyrosinase from *Aspergillus oryzae* [226]. Thus, it may play an important role in binding CuB and may help identify Cu-dependent tyrosinases in sequence databases. However, a recent study suggested the ability of the CuA site in mouse TRP1 to bind Cu and sustain the typical tyrosinase enzymatic activities [227]. It is possible that tyrosinase may acquire Cu inefficiently and subsequently lose it within the trans-Golgi network of melanocytes and then be reloaded with Cu within melanosomes to catalyze melanin synthesis [228]. This scheme would be consistent with the presence of an inactive “Cu-independent” form of Cu-dependent tyrosinase, suggesting an exquisite and complex control of tyrosinase activity.

Hemocyanin also belongs to the type 3 cuproprotein family and uses six histidine residues to bind two Cu ions. Members of the hemocyanin family have been detected only in the hemolymph of animals in the phyla Arthropoda and Mollusca. All sequenced Arthropoda contain both Cu-dependent and Cu-independent forms of proteins of this family, suggesting that they may have co-occurred in the ancestor of modern arthropods [223]. These Cu-independent hemocyanin-derived proteins (designated hexamerins) were previously thought to have lost the ability to bind Cu and transport oxygen; instead, they became storage proteins to serve as sources of amino acids during metamorphosis [229]. Similar to tyrosinase, an additional histidine was found immediately upstream of one of three histidine residues in the first Cu site in all Cu-dependent hemocyanins, but it was absent in all Cu-independent hexamerins (Figure 9b), implying that this additional histidine may be involved in Cu binding in hemocyanins.

Cu dependence is more conserved than Zn dependence, probably because of the prevalence of catalytic functions of Cu in cuproproteins, whereas in many Zn-containing proteins, Zn plays more subtle roles, such as structural integrity, that

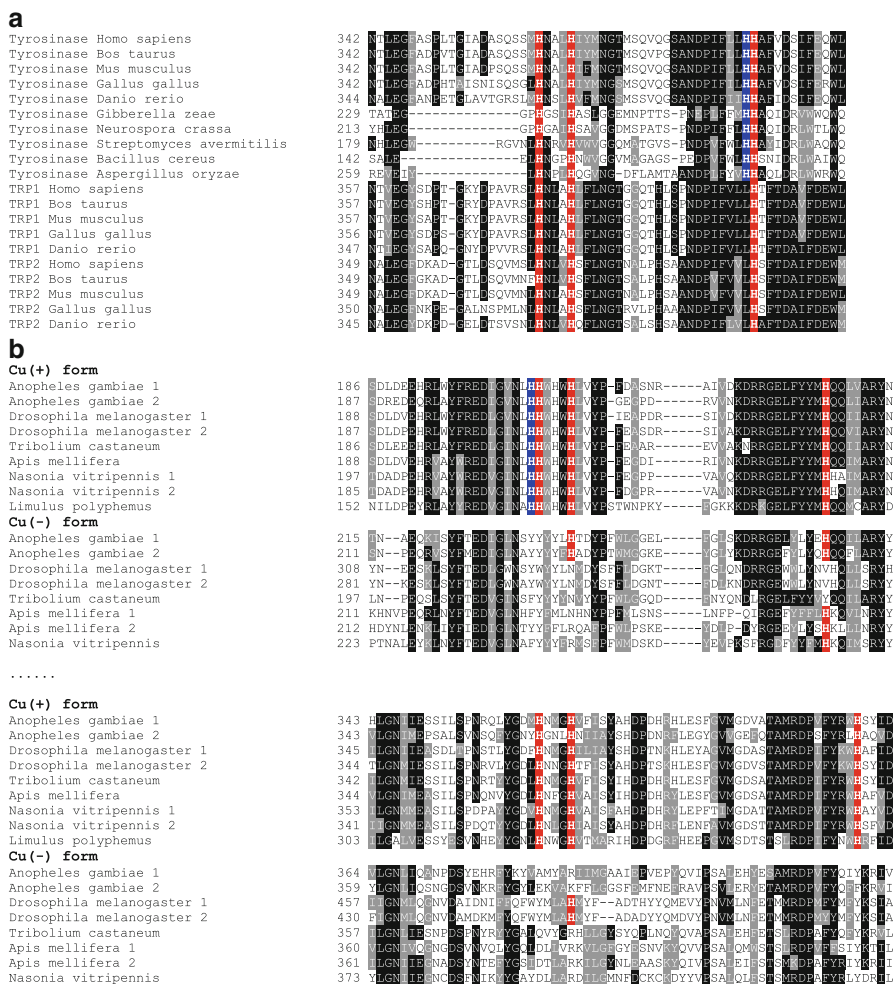


Figure 9 Multiple alignment of Cu-dependent and Cu-independent members of two cuproprotein families. **(a)** The tyrosinase family. Only CuB-binding sites are shown. Three His residues involved in metal binding are shown in a red background. A conserved His that is detected only in tyrosinases and that might also be involved in Cu binding is highlighted in blue. **(b)** The hemocyanin family. Only sequences corresponding to two Cu-binding sites are shown. The six His residues involved in Cu binding are shown in red background. An additional conserved His that is only detected in the Cu-dependent form is shown in blue.

could also be achieved by other means. Interestingly, a recent study revealed that, in the cyanobacterium *Synechocystis PCC 6803*, Mn is utilized by a cupin protein that folds in the cytoplasm, whereas Cu is utilized by a similar protein that folds in the periplasm [230]. This study offered a mechanism whereby the compartment in which a protein folds overrides its binding preference to control its metal content.

With regard to other metals, previous comparative studies have shown that molybdoenzymes are strictly dependent on Mo. Similarly, the majority of Ni-dependent proteins are strictly dependent on this metal although additional Ni-binding proteins were reported in some organisms (such as GlxI). All vitamin B₁₂-dependent proteins contain a B₁₂-binding domain and are strictly dependent on this coenzyme. It is very difficult to predict Co-dependent and Co-independent forms of non-corrin Co-binding protein families solely based on computational approaches. To date, only NHase is known to have different active site motifs for Co- and Fe-binding forms [231,232]. It was previously found that only several sequenced organisms (~4%) contain NHases, and most of them are Co-binding proteins. The Fe-containing NHases might have evolved from Co-binding NHases [205].

8 Concluding Remarks

Comparative genomics provides a powerful tool for studying metal utilization and its evolutionary trends. The majority of these studies used strategies based on either identification of metalloproteomes using metal-binding motifs/patterns or investigation of metal utilization traits (e.g. transporters, regulators, cofactor biosynthesis pathways, and metal-dependent proteins). Currently, it is difficult to identify complete metalloproteomes for most metals as there is no reliable method to predict all metal-binding proteins. Nevertheless, comparative genomics studies have provided significant advances in unraveling the general principles of utilization of metals across the three domains of life.

In this chapter, we discussed how comparative genomics can be used to analyze the function and evolution of metal utilization. We described recent studies that used computational studies, especially comparative genomics analyses, to better understand the utilization of several major transition metals and to offer new avenues for further experimental analyses. These studies offered new insights and helped understand the evolutionary dynamics of metal dependence in proteins and organisms. It may be expected that, in the next few years, with the increased availability of sequenced genomes and improved tools for their analyses, comparative genomics will play a significant role in better understanding of metal utilization and evolution.

Abbreviations and Definitions

| | |
|---------|------------------------------------|
| 5,6-LAM | D-lysine 5,6-aminomutase |
| ABC | ATP-binding cassette |
| AO | aldehyde oxidase |
| AOR | aldehyde:ferredoxin oxidoreductase |
| CoA | coenzyme A |
| CODH | carbon monoxide dehydrogenase |

| | |
|-------------------|---|
| CODH/ACS | acetyl-coenzyme A decarboxylase/synthase |
| CoM | coenzyme M |
| COX | cytochrome c oxidase |
| cPMP | cyclic pyranopterin monophosphate |
| Ctr | Cu transporter |
| CuAO | Cu amine oxidase |
| Cu-Zn SOD | Cu-Zn superoxide dismutase |
| Cys | cysteine |
| DBM | dopamine β -monooxygenase |
| DDH/GDH | diol/glycerol dehydratase |
| DMSOR | dimethylsulfoxide reductase |
| EAL | ethanolamine ammonia lyase |
| ECM | ethylmalonyl-CoA mutase |
| EPR | electron paramagnetic resonance |
| GAO | galactose oxidase |
| GlxI | glyoxalase I |
| GM | glutamate mutase |
| GTP | guanosine 5'-triphosphate |
| ICM | isobutyryl-CoA mutase |
| L-dopa | L-dihydroxyphenylalanine |
| mARC | mitochondrial amidoxime-reducing component |
| MCM | methylmalonyl-CoA mutase |
| MCO | multicopper oxidase |
| MCP | Moco carrier protein |
| MCR | methyl-coenzyme M reductase |
| MetH | methionine synthase |
| MGD | molybdopterin guanine dinucleotide |
| MGM | methyleneglutarate mutase |
| MoBP | Moco-binding protein |
| Moco | Mo cofactor |
| Mop | Mo-binding protein |
| MPT | molybdopterin |
| Mtr | methyltetrahydromethanopterin:CoM methyltransferase |
| N ₂ OR | nitrous oxide reductase |
| NADH | nicotinamide adenine dinucleotide (reduced) |
| NCBI | National Center for Biotechnology Information |
| Ndh2 | NADH dehydrogenase 2 |
| NHase | nitrile hydratase |
| NiR | nitrite reductase |
| NMR | nuclear magnetic resonance |
| NR | nitrate reductase |
| PAM | peptidylglycine R-amidating monooxygenase |
| PDB | Protein Data Bank |
| PHM | peptidylglycine R-hydroxylating monooxygenase |
| pMMO | particulate methane monooxygenase |

| | |
|--------|---|
| RNR II | B ₁₂ -dependent ribonucleotide reductase |
| SO | sulfite oxidase |
| SOD | superoxide dismutase |
| SodN | Ni-containing superoxide dismutase |
| TRP | tyrosinase-related proteins |
| Wco | tunftsens cofactor |
| XDH | xanthine dehydrogenase |
| XO | xanthine oxidase |

Acknowledgments This work was supported, in whole or in part, by National Institutes of Health Grants GM061603 and CA080946 (to V. N. G.), and by the National Natural Science Foundation of China under NO. 31171233 (to Y. Z.).

References

1. H. Haraguchi, *J. Anal. At. Spectrom.* **2004**, *19*, 5–14.
2. S. B. Goldhaber, *Regul. Toxicol. Pharmacol.* **2003**, *38*, 232–242.
3. D. P. Cuthbertson, *Bibl. Nutr. Dieta.* **1973**, *18*, 65–91.
4. W. Mertz, *Biol. Trace Elem. Res.* **1998**, *66*, 185–191.
5. A. Van Gossum, J. Neve, *Curr. Opin. Clin. Nutr. Metab. Care* **1998**, *1*, 499–507.
6. J. C. King, *Am. J. Clin. Nutr.* **2011**, *94*, 679S–684S.
7. D. Galaris, K. Pantopoulos, *Crit. Rev. Clin. Lab. Sci.* **2008**, *45*, 1–23.
8. J. E. Posey, F. C. Gherardini, *Science* **2000**, *288*, 1651–1653.
9. L. A. Ba, M. Doering, T. Burkholz, C. Jacob, *Metallomics* **2009**, *1*, 292–311.
10. S. I. Patzer, K. Hantke, *Mol. Microbiol.* **1998**, *28*, 1199–1210.
11. V. V. Bartsevich, H. B. Pakrasi, *J. Biol. Chem.* **1996**, *271*, 26057–26061.
12. J. A. Maupin-Furlow, J. K. Rosentel, J. H. Lee, U. Deppenmeier, R. P. Gunsalus, K. T. Shanmugam, *J. Bacteriol.* **1995**, *177*, 4851–4856.
13. T. Eitinger, M. A. Mandrand-Berthelot, *Arch. Microbiol.* **2000**, *173*, 1–9.
14. G. Grass, S. Franke, N. Taudte, D. H. Nies, L. M. Kucharski, M. E. Maguire, C. Rensing, *J. Bacteriol.* **2005**, *187*, 1604–1611.
15. J. P. Liuzzi, R. J. Cousins, *Annu. Rev. Nutr.* **2004**, *24*, 151–172.
16. H. Makui, E. Roig, S. T. Cole, J. D. Helmann, P. Gros, M. F. Cellier, *Mol. Microbiol.* **2000**, *35*, 1065–1078.
17. L. T. Phung, G. Ajlani, R. Haselkorn, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9651–9654.
18. A. Dancis, D. S. Yuan, D. Haile, C. Askwith, D. Eide, C. Moehle, J. Kaplan, R. D. Klausner, *Cell* **1994**, *76*, 393–402.
19. O. Degen, T. Eitinger, *J. Bacteriol.* **2002**, *184*, 3569–3577.
20. M. L. Ackland, H. J. McArdle, *Biomaterials* **1996**, *9*, 29–37.
21. M. E. Maguire, *Methods Mol. Biol.* **2007**, *394*, 289–305.
22. C. Hogstrand, C. Haux, *Comp. Biochem. Physiol. C.* **1991**, *100*, 137–141.
23. D. A. Cooksey, *FEMS Microbiol. Rev.* **1994**, *14*, 381–386.
24. P. Gourdon, X. Y. Liu, T. Skjørringe, J. P. Morth, L. B. Møller, B. P. Pedersen, P. Nissen, *Nature* **2011**, *475*, 59–64.
25. R. J. Cousins, R. J. McMahon, *J. Nutr.* **2000**, *130*, 1384S–1387S.
26. I. Bremner, J. H. Beattie, *Proc. Nutr. Soc.* **1995**, *54*, 489–499.
27. G. Schwarz, R. R. Mendel, *Annu. Rev. Plant Biol.* **2006**, *57*, 623–647

28. B. Kräutler, *Biochem. Soc. Trans.* **2005**, *33*, 806–810.
29. J. E. Coleman, *Annu. Rev. Biochem.* **1992**, *61*, 897–946.
30. R. K. Watt, P. W. Ludden, *Cell. Mol. Life Sci.* **1999**, *56*, 604–625.
31. E. M. Panina, A. A. Mironov, M. S. Gelfand, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9912–9917.
32. D. A. Rodionov, M. S. Gelfand, J. D. Todd, A. R. Curson, A. W. Johnston, *PLoS Comput. Biol.* **2006**, *2*, e163.
33. E. A. Permina, A. E. Kazakov, O. V. Kalinina, M. S. Gelfand, *BMC Microbiol.* **2006**, *6*, 49.
34. C. Andreini, L. Banci, I. Bertini, A. Rosato, *J. Proteome Res.* **2006**, *5*, 3173–3178.
35. C. Andreini, L. Banci, I. Bertini, S. Elmi, A. Rosato, *Proteins* **2007**, *67*, 317–324.
36. D. A. Rodionov, P. Hebbeln, M. S. Gelfand, T. Eitinger, *J. Bacteriol.* **2006**, *188*, 317–327.
37. P. G. Ridge, Y. Zhang, V. N. Gladyshev, *PLoS ONE* **2008**, *3*, e1378.
38. Y. Zhang, V. N. Gladyshev, *J. Mol. Biol.* **2008**, *379*, 881–899.
39. Y. Zhang, V. N. Gladyshev, *Chem. Rev.* **2009**, *109*, 4828–4861.
40. R. C. Hardison, *PLoS Biol.* **2003**, *1*, e58.
41. A. Ureta-Vidal, L. Ettwiller, E. Birney, *Nat. Rev. Genet.* **2003**, *4*, 251–262.
42. E. H. Margulies, E. Birney, *Nat. Rev. Genet.* **2008**, *9*, 303–313.
43. R. D. Finn, J. Mistry, J. Tate, P. Coghill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman, *Nucleic Acids Res.* **2010**, *38*, D211–D222.
44. C. J. Sigrist, L. Cerutti, E. de Castro, P. S. Langendijk-Genevaux, V. Bulliard, A. Bairoch, N. Hulo, *Nucleic Acids Res.* **2010**, *38*, D161–D166.
45. F. Servant, C. Bru, S. Carrère, E. Courcelle, J. Gouzy, D. Peyruc, D. Kahn, *Brief Bioinform.* **2002**, *3*, 246–251.
46. R. L. Tatusov, M. Y. Galperin, D. A. Natale, E. V. Koonin, *Nucleic Acids Res.* **2000**, *28*, 33–36.
47. E. Migliavacca, A. A. Adzhubei, M. C. Peitsch, *Bioinformatics* **2001**, *17*, 1047–1052.
48. Y. Zhang, V. N. Gladyshev, *Bioinformatics* **2010**, *26*, 700–702.
49. K. J. Waldron, J. C. Rutherford, D. Ford, N. J. Robinson, *Nature* **2009**, *460*, 823–830.
50. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410.
51. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
52. R. Hille, *Trends Biochem. Sci.* **2002**, *27*, 360–367.
53. R. M. Allen, R. Chatterjee, M. S. Madden, P. W. Ludden, V. K. Shah, *Crit. Rev. Biotechnol.* **1994**, *14*, 225–249.
54. G. Schwarz, R. R. Mendel, M. W. Ribbe, *Nature* **2009**, *460*, 839–847.
55. M. K. Chan, S. Mukund, A. Kletzin, M. W. Adams, D. C. Rees, *Science* **1995**, *267*, 1463–1469.
56. L. E. Bevers, P. L. Hagedoorn, G. C. Krijger, W. R. Hagen, *J. Bacteriol.* **2006**, *188*, 6498–6505.
57. J. R. Andreasen, K. Makdessi, *Ann. N. Y. Acad. Sci.* **2008**, *1125*, 215–229.
58. R. N. Pau, D. M. Lawson, *Met. Ions Biol. Syst.* **2002**, *39*, 31–74.
59. A. M. Grunden, R. M. Ray, J. K. Rosentel, F. G. Healy, K. T. Shanmugam, *J. Bacteriol.* **1996**, *178*, 735–744.
60. D. J. Studholme, R. N. Pau, *BMC Microbiol.* **2003**, *3*, 24.
61. D. A. Rodionov, I. Dubchak, A. Arkin, E. Alm, M. S. Gelfand, *Genome Biol.* **2004**, *5*, R90.
62. K. Makdessi, J. R. Andreasen, A. Pich, *J. Biol. Chem.* **2001**, *276*, 24557–24564.
63. M. E. Taveirne, M. L. Sikes, J. W. Olson, *Mol. Microbiol.* **2009**, *74*, 758–771.
64. J. Gisin, A. Müller, Y. Pfänder, S. Leimkühler, F. Narberhaus, B. Masepohl, *J. Bacteriol.* **2010**, *192*, 5943–5952.
65. H. Tomatsu, J. Takano, H. Takahashi, A. Watanabe-Takahashi, N. Shibagaki, T. Fujiwara, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18807–18812.

66. M. Tejada-Jiménez, A. Llamas, E. Sanz-Luque, A. Galván, E. Fernández, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20126–20130.
67. I. Baxter, B. Muthukumar, H. C. Park, P. Buchner, B. Lahner, J. Danku, K. Zhao, J. Lee, M. J. Hawkesford, M. L. Guerinot, D. E. Salt, *PLoS Genet.* **2008**, *4*, e1000004.
68. M. Tejada-Jiménez, A. Galván, E. Fernández, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6420–6425.
69. R. R. Mendel, *Biofactors* **2009**, *35*, 429–434.
70. R. R. Mendel, R. Hänsch, *J. Exp. Bot.* **2002**, *53*, 1689–1698.
71. J. Reiss, J. L. Johnson, *Hum. Mutat.* **2003**, *21*, 569–576.
72. F. Bittner, M. Oreb, R. R. Mendel, *J. Biol. Chem.* **2001**, *276*, 40381–40384.
73. K. Ichida, T. Matsumura, R. Sakuma, T. Hosoya, T. Nishino, *Biochem. Biophys. Res. Commun.* **2001**, *282*, 1194–1200.
74. J. Teschner, N. Lachmann, J. Schulze, M. Geisler, K. Selbach, J. Santamaria-Araujo, J. Balk, R. R. Mendel, F. Bittner, *Plant Cell* **2010**, *22*, 468–480.
75. F. Sargent, *Microbiology* **2007**, *153*, 633–651.
76. A. Vergnes, J. Pommier, R. Toci, F. Blasco, G. Giordano, A. Magalon, *J. Biol. Chem.* **2006**, *281*, 2170–2176.
77. N. Ray, J. Oates, R. J. Turner, C. Robinson, *FEBS Lett.* **2003**, *534*, 156–160.
78. K. Fischer, A. Llamas, M. Tejada-Jimenez, N. Schrader, J. Kuper, F. S. Ataya, A. Galvan, R. R. Mendel, E. Fernandez, G. Schwarz, *J. Biol. Chem.* **2006**, *281*, 30186–30194.
79. T. Kruse, C. Gehl, M. Geisler, M. Lehrke, P. Ringel, S. Hallier, R. Hänsch, R. R. Mendel, *J. Biol. Chem.* **2010**, *285*, 6623–6635.
80. L. C. Seefeldt, B. M. Hoffman, D. R. Dean, *Annu. Rev. Biochem.* **2009**, *78*, 701–722.
81. B. E. Smith, R. R. Eady, *Eur. J. Biochem.* **1992**, *205*, 1–15.
82. G. J. Workun, K. Moquin, R. A. Rothery, J. H. Weiner, *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 228–248.
83. C. Feng, G. Tollin, J. H. Enemark, *Biochim. Biophys. Acta* **2007**, *1774*, 527–539.
84. M. J. Romão, *Dalton Trans.* **2009**, *21*, 4053–4068.
85. R. Hille, *Chem. Rev.* **1996**, *96*, 2757–2816.
86. R. A. Rothery, G. J. Workun, J. H. Weiner, *Biochim. Biophys. Acta* **2008**, *1778*, 1897–1929.
87. T. Satoh, F.N. Kurihara, *J. Biochem.* **1987**, *102*, 191–197.
88. J. H. Weiner, R. A. Rothery, D. Sambasivarao, C. A. Trieber, *Biochim. Biophys. Acta* **1992**, *1102*, 1–18.
89. M. Jormakka, B. Byrne, S. Iwata, *Curr. Opin. Struct. Biol.* **2003**, *13*, 418–423.
90. M. K. Chan, S. Mukund, A. Kletzin, M. W. Adams, D. C. Rees, *Science* **1995**, *267*, 1463–1469.
91. A. Havemeyer, F. Bittner, S. Wollers, R. Mendel, T. Kunze, B. Clement, *J. Biol. Chem.* **2006**, *281*, 34796–34802.
92. S. G. Kozmin, P. Leroy, Y. I. Pavlov, R. M. Schaaper, *Mol. Microbiol.* **2008**, *68*, 51–65.
93. B. Wahl, D. Reichmann, D. Niks, N. Krompholz, A. Havemeyer, B. Clement, T. Messerschmidt, M. Rothkegel, H. Biester, R. Hille, R. R. Mendel, F. Bittner, *J. Biol. Chem.* **2010**, *285*, 37847–37859.
94. J. Kotthaus, B. Wahl, A. Havemeyer, J. Kotthaus, D. Schade, D. Garbe-Schönberg, R. Mendel, F. Bittner, B. Clement, *Biochem. J.* **2011**, *433*, 383–391.
95. A. Havemeyer, J. Lang, B. Clement, *Drug Metab. Rev.* **2011**, *43*, 524–539.
96. Y. Zhang, V. N. Gladyshev, *J. Biol. Chem.* **2010**, *285*, 3393–3405.
97. Y. Zhang, S. Rump, V. N. Gladyshev, *Coord. Chem. Rev.* **2011**, *255*, 1206–1217.
98. P. M. McNicholas, M. M. Mazzotta, S. A. Rech, R. P. Gunsalus, *J. Bacteriol.* **1998**, *180*, 4638–4643.
99. A. Gasber, S. Klauemann, O. Trentmann, A. Trampczynska, S. Clemens, S. Schneider, N. Sauer, I. Feifer, F. Bittner, R. R. Mendel, H. E. Neuhaus, *Plant Biol. (Stuttg.)* **2011**, *13*, 710–718.
100. K. Fischer, G. G. Barbier, H. J. Hecht, R. R. Mendel, W. H. Campbell, G. Schwarz, *Plant Cell* **2005**, *17*, 1167–1179.

101. S. Wollers, T. Heidenreich, M. Zarepour, D. Zachmann, C. Kraft, Y. Zhao, R. R. Mendel, F. Bittner, *J. Biol. Chem.* **2008**, 283, 9642–9650.
102. M. M. Peña, J. Lee, D. J. Thiele, *J. Nutr.* **1999**, 129, 1251–1260.
103. L. M. Gaetke, C. K. Chow, *Toxicology* **2003**, 189, 147–163.
104. B. P. Rosen, *Comp. Biochem. Physiol. A: Mol. Integr. Physiol.* **2002**, 133, 689–693.
105. D. Osman, J. S. Cavet, *Adv. Appl. Microbiol.* **2008**, 65, 217–247.
106. C. Rensing, G. Grass, *FEMS Microbiol. Rev.* **2003**, 27, 197–213.
107. M. Solioz, J. V. Stoyanov, *FEMS Microbiol. Rev.* **2003**, 27, 183–195.
108. M. Solioz M, C. Vulpe, *Trends Biochem. Sci.* **1996**, 21, 237–241.
109. B. Fan, B. P. Rosen, *J. Biol. Chem.* **2002**, 277, 46987–46992.
110. L. Banci, I. Bertini, S. Ciofi-Baffoni, N. G. Kandias, N. J. Robinson, G. A. Spyroulias, X. C. Su, S. Tottey, M. Vanarotti, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, 103, 8320–8325.
111. S. Tottey, S. A. Rondet, G. P. Borrelly, P. J. Robinson, P. R. Rich, N. J. Robinson, *J. Biol. Chem.* **2002**, 277, 5490–5497.
112. L. Zhou, C. Singleton, N. E. Le Brun, *Biochem. J.* **2008**, 413, 459–465.
113. P. Cobine, W. A. Wickramasinghe, M. D. Harrison, T. Weber, M. Solioz, C. T. Dameron, *FEBS Lett.* **1999**, 445, 27–30.
114. A. C. Rosenzweig, T. V. O'Halloran, *Curr. Opin. Chem. Biol.* **2000**, 4, 140–147.
115. L. Banci, I. Bertini, S. Ciofi-Baffoni, X. C. Su, G. P. Borrelly, N. J. Robinson, *J. Biol. Chem.* **2004**, 279, 27502–27510.
116. B. Gold, H. Deng, R. Bryk, D. Vargas, D. Eliezer, J. Roberts, X. Jiang, C. Nathan, *Nat. Chem. Biol.* **2008**, 4, 609–616.
117. F. W. Outten, D. L. Huffman, J. A. Hale, T. V. O'Halloran, *J. Biol. Chem.* **2001**, 276, 30670–30677.
118. Y. Xue, A. V. Davis, G. Balakrishnan, J. P. Stasser, B. M. Staehlin, P. Focia, T. G. Spiro, J. E. Penner-Hahn, T. V. O'Halloran, *Nat. Chem. Biol.* **2008**, 4, 107–109.
119. S. Franke, G. Grass, C. Rensing, D. H. Nies, *J. Bacteriol.* **2003**, 185, 3804–3812.
120. S. D. Gupta, B. T. Lee, J. Camakaris, H. C. Wu, *J. Bacteriol.* **1995**, 177, 4207–4215.
121. Y. Li, J. Du, P. Zhang, L. Ding, *J. Struct. Biol.* **2010**, 169, 399–405.
122. Y. Guo, K. Smith, J. Lee, D. J. Thiele, M. J. Petris, *J. Biol. Chem.* **2004**, 279, 17428–17433.
123. E. M. Rees, D. J. Thiele, *Curr. Opin. Microbiol.* **2004**, 7, 175–184.
124. E. M. Rees, J. Lee, D. J. Thiele, *J. Biol. Chem.* **2004**, 279, 54221–54229.
125. M. L. Turski, D. J. Thiele, *J. Biol. Chem.* **2007**, 282, 24017–24026.
126. D. Steiger, M. Fetchko, A. Vardanyan, L. Atanesyan, K. Steiner, M. L. Turski, D. J. Thiele, O. Georgiev, W. Schaffner, *J. Biol. Chem.* **2010**, 285, 17089–17097.
127. A. E. Klomp, B. B. Tops, I. E. Van Denberg, R. Berger, L. W. Klomp, *Biochem. J.* **2002**, 364, 497–505.
128. P. V. van den Berghe, D. E. Folmer, H. E. Malingré, E. van Beurden, A. E. Klomp, B. van de Sluis, M. Merckx, R. Berger, L. W. Klomp, *Biochem. J.* **2007**, 407, 49–59.
129. P. C. Bull, G. R. Thomas, J. M. Rommens, J. R. Forbes, D. W. Cox, *Nat. Genet.* **1993**, 5, 327–337.
130. R. Linz, S. Lutsenko, *J. Bioenerg. Biomembr.* **2007**, 39, 403–407.
131. J. R. Prohaska, *Am. J. Clin. Nutr.* **2008**, 88, 826S–829S.
132. K. Balamurugan, D. Egli, H. Hua, R. Rajaram, G. Seisenbacher, O. Georgiev, W. Schaffner, *EMBO J.* **2007**, 26, 1035–1044.
133. C. Singleton, N. E. Le Brun, *Biometals* **2007**, 20, 275–289.
134. A. M. Nersissian, E. L. Shipp, *Adv. Protein Chem.* **2002**, 60, 271–340.
135. T. Sakurai, K. Kataoka, *Cell. Mol. Life Sci.* **2007**, 64, 2642–2656.
136. L. Basumallick, R. Sarangi, S. DeBeer George, B. Elmore, A. B. Hooper, B. Hedman, K. O. Hodgson, E. I. Solomon, *J. Am. Chem. Soc.* **2005**, 127, 3531–3544.
137. K. Nakamura, N. Go, *Cell. Mol. Life Sci.* **2005**, 62, 2050–2066.
138. S. M. Musser, S. I. Chan, *J. Mol. Evol.* **1998**, 46, 508–520.
139. L. Qin, C. Hiser, A. Mulichak, R. M. Garavito, S. Ferguson-Miller, *Proc. Natl. Acad. Sci. USA* **2006**, 103, 16117–16122.

140. K. Brown, M. Tegoni, M. Prudêncio, A. S. Pereira, S. Besson, J. J. Moura, I. Moura, C. Cambillau, *Nat. Struct. Biol.* **2000**, *7*, 191–195.
141. I. S. MacPherson, M. E. Murphy, *Cell. Mol. Life Sci.* **2007**, *64*, 2887–2899.
142. F. Boomsma, H. Hut, U. Bagghoe, A. van der Houwen, A. van den Meiracker, *Med. Sci. Monit.* **2005**, *11*, RA122–126.
143. E. Jakobsson, J. Nilsson, D. Ogg, G. J. Kleywegt, *Acta Crystallogr. D. Biol. Crystallogr.* **2005**, *61*, 1550–1562.
144. J. P. Klinman, *J. Biol. Chem.* **2006**, *281*, 3013–3016.
145. R. L. Lieberman, A. C. Rosenzweig, *Nature* **2005**, *434*, 177–182.
146. J. Kuper, A. Llamas, H. J. Hecht, R. R. Mendel, G. Schwarz, *Nature* **2004**, *430*, 803–806.
147. S. Firbank, M. Rogers, R. H. Guerrero, D. M. Dooley, M. A. Halcrow, S. E. Phillips, P. F. Knowles, M. J. McPherson, *Biochem. Soc. Symp.* **2004**, *71*, 15–25.
148. C. Andreini, I. Bertini, A. Rosato, *Bioinformatics* **2004**, *20*, 1373–1380.
149. M. J. Ellis, J. G. Grossmann, R. R. Eady, S. S. Hasnain, *J. Biol. Inorg. Chem.* **2007**, *12*, 1119–1127.
150. C. Andreini, L. Banci, I. Bertini, A. Rosato, *J. Proteome Res.* **2008**, *7*, 209–216.
151. E. I. Ochiai, *Biosystems* **1983**, *16*, 81–86.
152. R. K. Watt, P. W. Ludden, *Cell. Mol. Life Sci.* **1999**, *56*, 604–625.
153. M. Kobayashi, S. Shimizu, *Eur. J. Biochem.* **1999**, *261*, 1–9.
154. K. de Pina, C. Navarro, L. McWalter, D. H. Boxer, N. C. Price, S. M. Kelly, M. A. Mandrand-Berthelot, L. F. Wu, *Eur. J. Biochem.* **1995**, *227*, 857–865.
155. M. Shepherd, M. D. Heath, R. K. Poole, *Biochemistry* **2007**, *46*, 5030–5037.
156. C. Cavazza, L. Martin, E. Laffly, H. Lebrette, M. V. Cherrier, L. Zeppieri, P. Richaud, M. Carrière, J. C. Fontecilla-Camps, *FEBS Lett.* **2011**, *585*, 711–715.
157. C. Addy, M. Ohara, F. Kawai, A. Kidera, M. Ikeguchi, S. Fuchigami, M. Osawa, I. Shimada, S. Y. Park, J. R. Tame, J. G. Heddle, *Acta Crystallogr. D. Biol. Crystallogr.* **2007**, *63*, 221–229.
158. J. R. Roth, J. G. Lawrence, M. Rubenfield, S. Kieffer-Higgins, G. M. Church, *J. Bacteriol.* **1993**, *175*, 3303–3316.
159. T. Eitinger, J. Suhr, L. Moore, J. A. Smith, *Biometals* **2005**, *18*, 399–405.
160. D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, *J. Biol. Chem.* **2003**, *278*, 41148–41159.
161. B. Brito, R. I. Prieto, E. Cabrera, M. A. Mandrand-Berthelot, J. Imperial, T. Ruiz-Argüeso, J. M. Palacios, *J. Bacteriol.* **2010**, *192*, 925–935.
162. M. W. Persans, K. Nieman, D. E. Salt, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9995–10000.
163. D. S. Conklin, J. A. McMaster, M. R. Culbertson, C. Kung, *Mol. Cell. Biol.* **1992**, *12*, 3678–3688.
164. N. S. Dosanjh, S. L. Michel, *Curr. Opin. Chem. Biol.* **2006**, *10*, 123–130.
165. N. Sukdeo, S. L. Clugston, E. Daub, J. F. Honek, *Biochem. J.* **2004**, *384*, 111–117.
166. E. Denkhaus, K. Salmikow, *Crit. Rev. Oncol. Hematol.* **2002**, *42*, 35–56.
167. H. L. Mobley, R. P. Hausinger, *Microbiol. Rev.* **1989**, *53*, 85–108.
168. C. Follmer, *Phytochemistry* **2008**, *69*, 18–28.
169. A. Böck, P. W. King, M. Blokesch, M. C. Posewitz, *Adv. Microb. Physiol.* **2006**, *51*, 1–71.
170. H. Ogata, W. Lubitz, Y. Higuchi, *Dalton Trans.* **2009**, *37*, 7577–7587.
171. M. K. Eidsness, R. A. Scott, B. C. Prickril, D. V. DerVartanian, J. Legall, I. Moura, J. J. Moura, H. D. Jr Peck, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 147–151.
172. P. A. Lindahl, *Biochemistry* **2002**, *41*, 2097–2105.
173. J. H. Jeoung, H. Dobbek, *Science* **2007**, *318*, 1461–1464.
174. C. L. Drennan, T. I. Doukov, S. W. Ragsdale, *J. Biol. Inorg. Chem.* **2004**, *9*, 511–515.
175. M. Dey, X. Li, Y. Zhou, S. W. Ragsdale, *Met. Ions Life Sci.* **2010**, *7*, 71–110.
176. S. Shima, M. Krueger, T. Weinert, U. Demmer, J. Kahnt, R. K. Thauer, U. Ermler, *Nature* **2011**, *481*, 98–101.
177. D. P. Barondeau, C. J. Kassmann, C. K. Bruns, J. A. Tainer, E. D. Getzoff, *Biochemistry* **2004**, *43*, 8038–8047.

178. E. L. Borths, K. P. Locher, A. T. Lee, D. C. Rees, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16642–16647.
179. N. Cadieux, P. G. Phan, D. S. Cafiso, R. J. Kadner, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10688–10693.
180. J. Wuerges, S. Geremia, S. N. Fedosov, L. Randaccio, *IUBMB Life* **2007**, *59*, 722–729.
181. M. T. Croft, A. D. Lawrence, E. Raux-Deery, M. J. Warren, A. G. Smith, *Nature* **2005**, *438*, 90–93.
182. D. Padovani, T. Labunska, B. A. Palfey, D. P. Ballou, R. Banerjee, *Nat. Chem. Biol.* **2008**, *4*, 194–196.
183. D. Padovani, R. Banerjee, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21567–21572.
184. R. Banerjee, S. W. Ragsdale, *Annu. Rev. Biochem.* **2003**, *72*, 209–247.
185. E. N. Marsh, *Essays Biochem.* **1999**, *34*, 139–154.
186. E. N. Marsh, C. L. Drennan, *Curr. Opin. Chem. Biol.* **2001**, *5*, 499–505.
187. R. Banerjee, M. Vlasie, *Biochem. Soc. Trans.* **2002**, *30*, 621–624.
188. A. Birch, A. Leiser, J. A. Robinson, *J. Bacteriol.* **1993**, *175*, 3511–3519.
189. F. Mancia, N. H. Keep, A. Nakagawa, P. F. Leadlay, S. McSweeney, B. Rasmussen, P. Bösecke, O. Diat, P. R. Evans, *Structure* **1996**, *4*, 339–350.
190. A. Ratnatilleke, J. W. Vrijbloed, J. A. Robinson, *J. Biol. Chem.* **1999**, *274*, 31679–31685.
191. T. J. Erb, J. Rétey, G. Fuchs, B. E. Alber, *J. Biol. Chem.* **2008**, *283*, 32283–32293.
192. B. Hoffmann, R. Konrat, H. Bothe, W. Buckel, B. Kräutler, *Eur. J. Biochem.* **1999**, *263*, 178–188.
193. A. J. Pierik, D. Ciceri, R. F. Lopez, F. Kroll, G. Bröker, B. Beatrix, W. Buckel, B. T. Golding, *Biochemistry* **2005**, *44*, 10541–10551.
194. F. Berkovitch, E. Behshad, K. H. Tang, E. A. Enns, P. A. Frey, C. L. Drennan, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15870–15875.
195. D. I. Liao, G. Dotson, I. Jr. Turner, L. Reiss, M. Emptage, *J. Inorg. Biochem.* **2003**, *93*, 84–91.
196. L. Sun, K. Warncke, *Proteins* **2006**, *64*, 308–319.
197. M. Kolberg, K. R. Strand, P. Graff, K. K. Andersson, *Biochim. Biophys. Acta* **2004**, *1699*, 1–34.
198. M. D. Sintchak, G. Arjara, B. A. Kellogg, J. Stubbe, C. L. Drennan, *Nat. Struct. Biol.* **2002**, *9*, 293–300.
199. R. V. Banerjee, R. G. Matthews, *FASEB J.* **1990**, *4*, 1450–1459.
200. C. L. Drennan, S. Huang, J. T. Drummond, R. G. Matthews, M. L. Lidwig, *Science* **1994**, *266*, 1669–1674.
201. C. H. Hagemeyer, M. Krer, R. K. Thauer, E. Warkentin, U. Ermler, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18917–18922.
202. K. Sauer, R. K. Thauer, *FEBS Lett.* **1998**, *436*, 401–402.
203. J. Krasotkina, T. Walters, K. A. Maruya, S. W. Ragsdale, *J. Biol. Chem.* **2001**, *276*, 40991–40997.
204. A. Banerjee, R. Sharma, U. C. Banerjee, *Appl. Microbiol. Biotechnol.* **2002**, *60*, 33–44.
205. Y. Zhang, D. A. Rodionov, M. S. Gelfand, V. N. Gladyshev, *BMC Genomics* **2009**, *10*, 78.
206. A. Y. Mulkidjanian, *Biol. Direct.* **2009**, *4*, 26.
207. B. L. Vallee, D. S. Auld, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 220–224.
208. C. Andreini, L. Banci, I. Bertini, A. Rosato, *J. Proteome Res.* **2006**, *5*, 196–201.
209. C. Andreini, I. Bertini, A. Rosato, *Acc. Chem. Res.* **2009**, *42*, 1471–1479.
210. L. Decaria, I. Bertini, R. J. Williams, *Metallomics* **2010**, *2*, 706–709.
211. M. Ferrer, O. V. Golyshina, A. Belouqui, P. N. Golyshin, K. N. Timmis, *Nature* **2007**, *445*, 91–94.
212. J. E. Posey, F. C. Gherardini, *Science* **2000**, *288*, 1651–1653.
213. M. Imbert, R. Blondeau, *Curr. Microbiol.* **1998**, *37*, 64–66.
214. C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, *J. Biol. Inorg. Chem.* **2008**, *13*, 1205–1218.
215. G. Cavallaro, L. Decaria, A. Rosato, *J. Proteome Res.* **2008**, *7*, 4946–4954.

216. A. Cvetkovic, A. L. Menon, M. P. Thorgersen, J. W. Scott, F. L. 2nd Poole, F. E. Jr Jenney, W. A. Lancaster, J. L. Praissman, S. Shanmukh, B. J. Vaccaro, S. A. Trauger, E. Kalisiak, J. V. Apon, G. Siuzdak, S. M. Yannone, J. A. Tainer, M. W. Adams, *Nature* **2010**, *466*, 779–782.
217. W. A. Lancaster, J. L. Praissman, F. L. 2nd Poole, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jr Jenney, M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, M. W. Adams, *BMC Bioinformatics* **2011**, *12*, 64.
218. K. Degtyarenko, *Bioinformatics* **2000**, *16*, 851–864.
219. W. Shi, C. Zhan, A. Ignatov, B. A. Manjasetty, N. Marinkovic, M. Sullivan, R. Huang, M. R. Chance, *Structure* **2005**, *13*, 1473–1486.
220. W. Shi, M. R. Chance, *Cell. Mol. Life Sci.* **2008**, *65*, 3040–3048.
221. K. S. Makarova, V. A. Ponomarev, E. V. Koonin, *Genome Biol.* **2001**, *2*, RESEARCH 0033.
222. C. E. Haas, D. A. Rodionov, J. Kropat, D. Malasarn, S. S. Merchant, V. de Crécy–Lagard, *BMC Genomics* **2009**, *10*, 470.
223. Y. Zhang, V. N. Gladyshev, *J. Biol. Chem.* **2011**, *286*, 23623–23629.
224. R. A. Sturm, N. F. Box, M. Ramsay, *Bioessays* **1998**, *20*, 712–721.
225. M. Furumura, F. Solano, N. Matsunaga, C. Sakai, R. A. Spritz, V. J. Hearing, *Biochem. Biophys. Res. Commun.* **1998**, *242*, 579–585.
226. M. Nakamura, T. Nakajima, Y. Ohba, S. Yamauchi, B. R. Lee, E. Ichishima, *Biochem. J.* **2000**, *350*, 537–545.
227. C. Olivares, F. Solano, *Pigment Cell Melanoma Res.* **2009**, *22*, 750–760.
228. S. R. Setty, D. Tenza, E. V. Sviderskaya, D. C. Bennett, G. Raposo, M. S. Marks, *Nature* **2008**, *454*, 1142–1146.
229. J. R. Martins, F. M. Nunes, A. S. Cristino, Z. L. Simões, M. M. Bitondi, *BMC Mol. Biol.* **2010**, *11*, 23.
230. S. Tottey, K. J. Waldron, S. J. Firbank, B. Reale, C. Bessant, K. Sato, T. R. Cheek, J. Gray, M. J. Banfield, C. Dennison, N. J. Robinson, *Nature* **2008**, *455*, 1138–1142.
231. A. Banerjee, R. Sharma, U. C. Banerjee, *Appl. Microbiol. Biotechnol.* **2002**, *60*, 33–44.
232. A. Miyanaga, S. Fushinobu, K. Ito, T. Wakagi, *Biochem. Biophys. Res. Commun.* **2001**, *288*, 1169–1174.